جامعة سيدي محمد بن عبد الله بفاس
UNIVERSITÉ SIDI MOHAMED BEN ABDELLAH DE FES

كلية العلوم والتقنيات فاس
FACULTÉ DES SCIENCES ET TECHNIQUES DE FÈS

N° d'ordre : 22/2024

Centre des Etudes Doctorales : **Sciences et Techniques et Sciences Médicales**
Formation Doctorale : **Sciences et Génie de la Matière, de la Terre et de la Vie**
Discipline : **Chimie**
Spécialité* : **Chimie physique et modélisation**
Laboratoire : **Laboratoire d'ingénierie, systèmes et applications (LISA)**

Thèse pour l'obtention de DOCTORAT

Conception de médicaments assistée par ordinateur : Applications anticancéreuses et antivirales

Préparée par :

Mr. OSSAMA DAOUI

Soutenue le 27 avril 2024 devant le jury :

NOM ET PRÉNOM	GRADE	ETABLISSEMENT	QUALITÉ
Adil Touimi Benjelloun	PES	Faculté des Sciences Dhar El Mahraz, Fès	Président
Mohammed BOUACHRINE	PES	Faculté des Sciences de Meknès	Rapporteur
Hamid TOUFIK	PES	Faculté Polydisciplinaire de Taza	Rapporteur
Fouad KHALIL	PES	Faculté des Sciences et Techniques de Fès	Rapporteur
Samir CHTITA	PH	Faculté des Sciences Ben M'sik de Casablanca	Examinateur
Abdelkbir ERROUGUI	PH	Faculté des Sciences Ben M'sik de Casablanca	Examinateur
Hayat BENMOUSSA	PH	Cité d'Innovation de Fès	Examinatrice
Abdelmoula EL ABBOUCHI	PA	Université Euro-Méditerranéenne de Fès	Invité
Souad EL KHATTABI	PH	Ecole Nationale des Sciences Appliquées de Fès	Directrice de thèse

À mes parents, mes frères, et ma sœur

REMERCIEMENTS

J’aimerais remercier ma directrice de recherche, Mme, Pr. Souad EL KHATTABI, pour toutes ses contributions précieuses lors des discussions, conseils, orientations, partages de connaissances, échanges d'idées durant la période de préparation de ma thèse. En effet, ma directrice de recherches a pu me permettre de pousser mes efforts à fond en vue d'atteindre des finalités que je ne pensais pas possibles au départ. Je tiens également à exprimer mes profonds remerciements à Mr, le Pr. Samir CHTITA pour ses orientations, sa collaboration, sa confiance, son soutien et l'intérêt constant au cours de la réalisation de cette thèse, sans compter ses propres encouragements et conseils. De plus, j’aimerais remercier tous les membres de jury : Pr. Adil Touimi Benjelloun de FSDM-Fès, Pr. Mohammed BOUACHRINE de FS-Meknès, Pr. Hamid TOUFIK de FP-Taza, Pr. Fouad KHALIL de FST-Fès, Pr. Samir CHTITA de FSBM-Casablanca, Pr. Abdelkbir ERROUGUI de FSBM-Casablanca, Pr. Hayat BENMOUSSA de CI-USMBA-Fès, Pr. Abdelmoula EL ABBOUCHI de UEM-Fès pour avoir accepté l'évaluation du présent travail. Je tiens également à remercier à la fois le président et tous les examinateurs. Finalement, je tiens également à remercier mes proches et mes amis qui m'ont soutenu et motivé pendant la préparation de la présente thèse.

TABLE DES MATIÈRES

REMERCIEMENTS i
TABLE DES MATIÈRES ii
LISTE DES FIGURES x
LISTE DES TABLEAUX xvi
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES xix
RÉSUMÉ DE LA THÈSE 1
ABSTRACT 2
الملخص 3
INTRODUCTION GÉNÉRALE 4
1 Contexte, défis majeurs et pistes de solution 4
2 Approches proposées 5
3 Cadre spécifique des applications 7
4 Plan du manuscrit 8
CHAPITRE I : APERÇU DU PROCESSUS DE DÉCOUVERTE DE NOUVEAUX MÉDICAMENTS, DE LA RECHERCHE À LA MISE SUR LE MARCHÉ 10
1. Introduction et problématiques 10
2 Les méthodes thérapeutiques et leurs particularités 11
2.1 Traitement chirurgical 11
2.2 Traitement radiologique 11
2.3 Traitements médicamenteux 12
2.3.1 Les médicaments cytotoxiques 12
2.3.2 La thérapeutique endocrine 12
2.3.3 Vaccins thérapeutiques 12
2.3.4 Les thérapies ciblées 12
3 Variantes des médicaments fondamentaux 14
4 Inhibition enzymatique réversible et irréversible 16
4.1 Mode d'action et mécanismes de liaison 16
4.2 Caractérisation des inhibiteurs 18
5 Découverte de médicaments, de la recherche à la mise en œuvre 19
5.1 Évolution de la découverte de médicaments 19
5.2 Processus de découverte de médicaments 20
5.3 Découverte précoce de médicaments 21
5.3.1 Identification et validation de cible 21
5.3.2 Criblage à haut débit 21
5.3.3 Identification des Hits 22
5.3.4 Développement d'essais et criblage 22
5.3.5 Hit-To-Lead (H2L) 23
5.3.6 Génération et optimisation des leads (Lead optimization) 23
5.3.7 Essais in vitro et in vivo 23
5.3.8 Phase préclinique 23
5.4 Phase clinique 23

5.5 Autorisation de mise sur le marché 24
5.6 Contraintes et défaillances 25
6 Approche de la Conception de Médicaments Assistée par Ordinateur (CADD) : Classification, Utilisations, Limitations et Avenir 25
6.1 Progrès du CADD 25
6.2 Approches de CADD 26
6.3 Applications et principe des approches CADD 27
6.4 Classification des approches de CADD 27
6.4.1 Conception de médicaments basée sur la structure (SBDD) 28
6.4.2 Conception de médicaments basée sur le ligand (LBDD) 29
7 Application des Techniques de Machine Learning dans la Conception et la Découverte de Médicaments : Méthodes et Applications 30
7.1 Apprentissage supervisé 31
7.2 Apprentissage non superviser 31
7.3 Apprentissage semi-supervisé 31
7.4 Apprentissage par renforcement 32
7.5 Apprentissage profond 32
8 Intégration de l'Intelligence Artificielle dans la conception et la découverte de médicaments : Applications et Potentiel 32
8.1 Applications de CADD, ML, IA dans la découverte de médicaments 33
8.2 Grands défis de CADD, ML, IA et perspectives 34
8.3 Limites de CADD et perspectives 34
9 Approches de ciblage moléculaire employées dans ce travail de recherche 36
9.1 Optimisation de nouvelles voies de ciblage moléculaire anticancéreux 37
9.1.1 Ciblage moléculaire du récepteur à tyrosine kinase c-Met 39
9.1.2 Ciblage moléculaire du récepteur du facteur de croissance épidermique (EGFR) 41
9.1.3 Ciblage moléculaire de la Protéine-1 de translocation de lymphome du tissu lymphoïde associé aux muqueuses (MALT1) 44
9.2 Optimisation d'une nouvelle voie de ciblage moléculaire antiviral contre le SARS-CoV-2 45
10 Conclusion 48
CHAPITRE II : FUSION DE LA MODÉLISATION BIMOLÉCULAIRE DANS LA STRATÉGIE HYBRIDE DE CADD : MÉTHODOLOGIES, IMPLÉMENTATION ET APPLICATIONS THÉRAPEUTIQUES CIBLÉES 50
1 Introduction 50
2 Ajustement des bases de données moléculaires ciblées 50
2.1 Source, homogénéité et qualité de donnée 50
2.2 Classification et critères de sélection des chimiothèques étudiées dans ce travail 52
2.2.1 Homogénéité des chimiothèques 52
2.2.2 Critères de sélection des chimiothèques 52
3 Stratégie globale des modèles de CADD hybrides proposés dans ce travail 53
3.1 Analyse des données par voie bio-informatique et biostatistique 55
3.2 Contexte d'application 55
4 Modélisation biomoléculaire basée sur l’approches de QSAR multidimensionnelles 56

4.1 Principes fondamentaux des méthodes QSAR 56
4.2 Principes de l'OCDE relatifs à la validité des modèles QSAR/QSPR 57
4.3 Application et extension des approches QSAR 57
4.4 Classification des Modèles QSAR par dimension 58
4.5 Méthodologie de modélisation QSAR appliquée dans cette thèse 59
4.5.1 Méthodologie 2D-QSAR 60
4.5.2 Méthodologie 3D-QSAR 62
4.6 Descripteurs moléculaires pour la modélisation QSAR 63
4.6.1 Descripteurs moléculaires exploités dans le cadre de cette thèse 64
4.6.2 Classification par dimension des descripteurs moléculaires appliqués 65
4.6.3 Techniques et Outils pour transformer la présentation graphique en descripteurs moléculaires multidimensionnels 69
4.7 Base de données de descripteurs moléculaires pour la modélisation 2D-QSAR 72
4.7.1 Descripteurs moléculaires basés sur l'optimisation 3D 72
4.7.2 Descripteurs moléculaires basés sur l'optimisation MM2 74
4.7.3 Descripteurs moléculaires basés sur les calculs DFT 74
4.8 Base de Données de Descripteurs Moléculaires 4D pour la Modélisation 3D-QSAR 76
4.8.1 Champ de force Tripos et méthodologies associées 76
4.8.2 Méthodes de calcul des charges pour les structures 3D des molécules 77
4.8.3 Analyse conformationnelle des molécules bioactives 78
4.8.4 Descripteurs moléculaires identifiés par les méthodes de CoMFA et de CoMSIA 78
4.9 Optimisation de la dimensionnalité des bases de données de descripteurs moléculaires générés 79
4.9.1 Analyse en composantes principales 79
4.9.2 Classification ou partitionnement des données 81
4.10 Modélisation 2D/3D-QSAR à l'aide des analyses biostatistiques prédictives 82
4.10.1 Stratégie globale pour la mise en œuvre 82
4.10.2 Modélisation 2D-QSAR à l'aide des méthodes de Régression Multiple 84
4.10.3 Régression Linéaire Multiples (RLM) 85
4.10.4 Régression Non Linéaire Multiples (RNLM) 85
4.10.5 Modélisation 3D-QSAR à l'aide des méthodes dépendantes de l'alignement moléculaire 90
5 Critères de la validation des modèles 2D/3D-QSAR 96
5.1 Validation interne 96
5.2 Validation croisée 97
5.3 Validation externe 97
5.4 Test de Y-randomisation 97
5.5 Domaine d'application (DA) 98
5.6 Critères statistiques pour l'évaluation de la performance des modèles QSAR 98
5.7 Exploitation de modèles 2D/3D-QSAR pour le criblage de chimiothèques et l'optimisation de composés têtes de série 100

5.8 Méthodes employées pour la conception et l'optimisation informatique de nouvelles molécules bioactives 100
5.8.1 Identification des sites pharmacophores 100
5.8.2 Sélection des groupes fonctionnels optimaux pour la conception de nouveaux ligands 102
5.8.3 Sélection et identification de molécules candidates en tant qu'inhibiteurs spécifiques par criblage 102
6 Approche employée pour la sélection des ligands candidats dans le processus de conception et de découverte de médicaments par modélisation informatique 102
6.1 Modélisation pharmacocinétique et pharmacodynamique (PK/PD) 103
6.1.1 Contexte et principe 103
6.1.2 Rôle des modèles Drug-like et ADME-Tox in silico 104
6.1.3 Mise en œuvre des modèles de type Drug-like et ADME-Tox in silico 105
6.2 Stratégie de criblage virtuel basé sur le Docking moléculaire 107
6.2.1 Principe de Docking Moléculaire 108
6.2.2 Représentation du système macromoléculaire (protéine/enzyme) 108
6.2.3 Fonctions de score 108
6.2.4 Estimation du score des poses 109
6.2.5 Méthodes de Docking moléculaire adaptées en vue de détecter les conformations optimales protéine-ligand 109
6.2.6 Limites de la stratégie de Docking moléculaire 110
6.2.7 Perspectives pour relever certains défis de la stratégie du Docking moléculaire 111
6.2.8 Modèle du Docking Moléculaire mis en œuvre dans cette thèse 111
6.3 Méthode adoptée pour la validation du modèle du Docking moléculaire 113
6.3.1 Outils de Docking protéine-ligand utilisés 114
6.3.2 Stratégies des Analyses post-Docking appliqué 115
6.3.3 Préparation de systèmes macromoléculaires protéine-ligand 115
6.3.4 Champs de force exploités dans l'optimisation des systèmes protéine-ligand 115
6.3.5 Optimisation énergétique des systèmes protéine-ligand 116
6.3.6 Optimisation du solvant 117
6.4 Estimation des affinités de liaison du ligand via MM-GBSA et MM-PBSA 117
6.4.1 Outils de mise en œuvre 117
6.4.2 Mise en œuvre 118
6.5 Simulations de la dynamique moléculaire (DM) 118
6.5.1 Définition et principe fondamental des simulations de la DM 119
6.5.2 Aspects techniques 121
6.5.3 Économiser le temps de la simulation MD 123
6.5.4 Pas de temps (Δt) 124
6.5.5 Stratégie de la DM mise en œuvre adoptée dans ce travail 125
6.5.6 Applications de la dynamique moléculaire 128
7 Conclusion 128
CHAPITRE III : MODÉLISATION D'INHIBITEURS NON-COVALENTS BASEE SUR DES MOLÉCULES BIOACTIVES DE PRODUITS SYNTHÉTIQUES POUR L'INHIBITION DE LA

PROLIFÉRATION CELLULAIRE CANCÉREUSE : CIBLAGE MOLÉCULAIRE DU RÉCEPTEUR TYROSINE KINASE DE c-Met 130
Introduction 130
A. Identification in silico de Nouvelles Molécules Bioactives Dérivées de Dimédone, Basées sur la Structure de 4,5,6,7-Tétrahydrobenzo[D]-Thiazol-2-Yl, en tant qu'Inhibiteurs Potentiels du Récepteur Tyrosine Kinase c-Met pour la Thérapie Ciblée contre le Cancer 132
1 Introduction 132
2 Stratégie de conception de médicaments assistée par ordinateur développée et appliquée 133
3 Résultats et discussion 136
3.1 Analyses Structure-Activité par 2D-QSAR 136
3.1.1 Modélisation 2D-QSAR linéaire et non linéaire basée sur des descripteurs moléculaires 0D-3D 136
3.1.2 Criblage in silico basé sur les modèles 2D-QSAR 139
3.2 Criblage virtuel basé sur la modélisation Drug-like 140
3.3 Analyse des interactions entre cMet et l'inhibiteur et évaluation des scores d'énergie de liaison par Docking Moléculaire 141
3.3.1 Validation du modèle 3D de la protéine ciblée par la procédure de re-Docking 141
3.3.2 Découverte de potentiels inhibiteurs de c-Met basés sur le 4,5,6,7-Tétrahydrobenzo[D]-Thiazol-2-Yl 143
3.4 Profil Pharmacocinétique et Pharmacodynamique des inhibiteurs de c-Met identifiés in silico 146
4 Conclusion 147
.B Criblage in silico, Modélisation et Optimisation d'Inhibiteurs de Type II du Récepteur Tyrosine Kinase c-Met à Base de Phénoxypyridine pour la Thérapie Ciblée Anticancéreuse 149
1 Introduction 149
2 Stratégie de conception de médicaments assistée par ordinateur développée et appliquée 151
3 Résultats et discussion 152
3.1 Analyses structure-activité basée sur la modélisation 3D-QSAR 152
3.1.1 Exploration et modélisation in silico à travers l'orientation des modèles de CoMFA et CoMSIA 152
3.1.2 Optimisation de nouveaux dérivés de la Phénoxypyridine comme inhibiteurs optimaux de c-Met 156
3.2 Exploration in silico fondée sur les prédictions Drug-like et ADME-Tox 158
3.2.1 Analyse des propriétés de biodisponibilité orale 158
3.2.2 Analyse pharmacocinétique et pharmacodynamique par la modélisation ADMET-Tox. 159
3.3 Docking Moléculaire : Identification de candidats inhibiteurs de c-Met 161
3.3.1 Validation du protocole du Docking Moléculaire 161
3.3.2 Caractérisation des conformations des dérivés de 4-phénoxypyridine adaptées à l'inhibition de c-Met 163
3.4 Validation rétrospective des résultats à l'aide des simulations de la Dynamique Moléculaire... 165
4 Conclusion 169
C. Criblage, Modélisation et Optimisation de Nouvelles Molécules Bioactives Dérivées de Cyclohexane-1,3-dione en tant qu'Agents Thérapeutiques Potentiels contre le Cancer du Poumon Non à Petites Cellules (CPNPC) en Ciblant le Récepteur Tyrosine Kinase c-Met 170

1 Introduction 170
2 Stratégie de conception de médicaments assistée par ordinateur développée et appliquée 171
3 Résultats et discussion 174
3.1 Analyses structure-activité par une démarche hybride combine 2D-QSAR et cartographie structurelle en espace 3D 174
3.1.1 Modélisation Linéaire et Non Linéaire via 2D-QSAR 174
3.1.2 Sélection de la conformation bioactive prototype adaptée pour la conception de médicaments à partir du ligand 177
3.2 Caractérisation 3D des propriétés optimales de la structure du ligand prototype 178
3.2.1 Cartographie 3D des contours du potentiel moléculaire électrostatique 178
3.2.2 Analyse de la compatibilité entre la surface c-Met et l'inhibiteur par Docking Moléculaire 179
3.3 Concevoir, optimisation et criblage de nouveaux dérivés de Cyclohexane-1,3-dione 182
3.4 Sélection des candidats H2L à travers des prédictions de Drug-like et d'ADME-Tox 185
3.5 Sélection des conformations optimales des dérivés de Cyclohexane-1,3-dione compatibles en tant qu'inhibiteurs de c-Met 186
3.6 Analyse des Interactions entre cMet et l'Inhibiteur via le Docking Moléculaire et l'Approche Prime MM-GBSA 187
3.7 Vérification rétrospective des résultats à travers des Simulations de Dynamique Moléculaire 187
3.7.1 Analyses RMSD et RMSF 192
3.7.2 Dynamique des contacts Protéine-ligand 193
3.7.3 Profil des propriétés dynamiques des inhibiteurs de c-Met proposés 193
3.7.4 Analyse des propriétés thermodynamiques des systèmes cMet-inhibiteur 193
4 Conclusion 195
CHAPITRE IV : EXPLORATION DES PHYTOCHIMIQUES POUR L'INHIBITION NON-COVALENTE DE LA PROLIFÉRATION CELLULAIRE CANCEREUSE : CIBLAGE MOLECULAIRE DU RÉCEPTEUR TYROSINE KINASE DE L'EGFR 197
Introduction 197
A. Modélisation assistée par ordinateur des composés dérivés de Magnolia officinalis pour le ciblage de l'EGFR en vue du traitement du cancer du poumon non à petites cellules 200
1 Introduction 200
2 Stratégie de conception de médicaments assistée par ordinateur développée et appliquée 201
3 Résultats et discussion 204
3.1 Analyses structure-activité basée sur 3D-QSAR 204
3.1.1 Caractérisation biomoléculaire d'inhibiteurs de l'EGFR par CoMFA et CoMSIA 204
3.1.2 Mécanisme d'affinité et de compatibilité des dérivés de Magnolol avec l'EGFR-TKD 208
3.1.3 Analyse biomoléculaire des dérivés du Magnolol via MolCAD 211
3.1.4 Design in silico de nouveaux dérivés du Magnolol à activité anticancéreuse 212
3.2 Explorer le potentiel thérapeutique des dérivés du Magnolol : Analyse in silico des propriétés Drug-like et ADME-Tox 214
3.3 Recherche d'inhibiteurs de l'EGFR-TKD par simulation de Docking Moléculaire 215
3.4 Affiner la sélection des inhibiteurs de l'EGFR-TKD par Prime/MM-GBSA 218
4 Conclusion 221

B. Explorer le Potentiel Anticancéreux des Phytochimiques du Cannabis : Une étude in silico 222
1 Introduction 222
2 Stratégie de conception de médicaments assistée par ordinateur développée et appliquée 224
3 Résultats et discussion 226
3.1 Criblage virtuel des phytochimiques de Cannabinoïdes et de Terpènes de Cannabis..... 226
3.1.1 Découverte de substances actives à l'aide du Docking Moléculaire 226
3.1.2 Prédiction de l'affinité des substances phytochimiques dérivées du Cannabis vis-à-vis le site actif d'EGFR-TKD..... 227
3.2 Prédiction des propriétés pharmacologiques et des risques toxicologiques des inhibiteurs potentiels de l'EGFR-TKD 228
3.3 Caractérisation du comportement des médicaments et de leur toxicité potentielle par modélisation ADME-Tox..... 229
3.4 Évaluation de la stabilité conformationnelle des complexes EGFR-TKIs 230
3.4.1 Étude des modes de compatibilité, d'affinité et d'interaction Protéine-Ligand 230
3.4.2 Découverte de candidats Lead Optimization (LO) à l'aide de simulations de Docking moléculaire et MM-GBSA 234
3.5 Analyse post-Docking des complexes EGFR-TKIs par simulations de Dynamique Moléculaire 237
3.5.1 Investigation dynamique des modifications de RMSD et RMSF 237
3.5.2 Caractérisation des interactions Protéine-Ligand au cours de la simulation de DM 238
3.5.3 Caractérisation temporelle des conformations des ligands 239
3.5.4 Investigation des influences des propriétés thermodynamiques 241
4 Conclusion..... 244
CHAPITRE V : MODÉLISATION DE LIGANDS BIOACTIFS POUR LUTTER SPÉCIFIQUEMENT CONTRE LE LYMPHOME ET LES INFECTIONS VIRALES SARS-CoV-2 À TRAVERS LE CIBLAGE MOLÉCULAIRE RÉVERSIBLE DES PROTÉASES MALT1 ET $3CL^{PRO}$..... 245
Introduction 245
A. Exploration, modélisation et optimisation de nouvelles petites molécules pyridiniques en tant que candidats inhibiteurs allostériques de la protéase MALT1 pour la thérapie ciblée du lymphome 246
1 Introduction..... 246
2 Stratégie de conception de médicaments assistée par ordinateur développée et appliquée 248
3 Résultats et discussion 251
3.1 Analyse structure-activité par modélisation 2D-QSAR..... 251
3.1.1 Modèle 2D-QSAR linéaire 251
3.1.2 Décryptage des corrélations entre les descripteurs moléculaires et la puissance d'inhibition..... 252
3.1.3 Évaluation de la validité interne et externe du modèle RLM/2D-QSAR..... 252
3.2 Étude des relations structure-activité par modélisation 3D-QSAR 253
3.2.1 Évaluation statistique des modèles CoMFA et CoMSIA..... 254
3.2.2 Corrélation des cartes de contour avec les données d'activité biologique..... 255
3.3 Identification des propriétés 2D et 3D essentielles pour l'activité optimale des dérivés pyridiniques 257
3.4 Identification des modes de liaison Protéine-ligand et des interactions clés par Docking Moléculaire 257
3.4.1 Évaluation de la robustesse du modèle de Docking moléculaire par re-Docking 258

3.4.2 Évaluer l'affinité des dérivés de pyridine avec la poche allostérique de MALT1 259
3.5 Conception et optimisation d'inhibiteurs allostériques de MALT1 260
3.5.1 Conception computationnelle de nouveaux dérivés de pyridine 262
3.5.2 Identification des composés de tête de la série "hits" 262
3.6 Évaluation in silico des propriétés biodisponibilité, pharmacocinétique et pharmacocinétique de molécules candidates comme H2L 264
3.6.1 Prédiction in silico basée sur le modèle "Drug-like" 264
3.6.2 Évaluations ADME-Tox in silico 265
3.7 Analyse des interactions entre les systèmes biomoléculaires MALT et les inhibiteurs allostériques candidats 266
3.7.1 Simulations de Docking Moléculaire et calculs Prime/MM-GBSA 267
3.7.2 Analyses post-Docking via simulations de dynamique moléculaire et calculs Thermal/MM-GBSA 268
3.7.3 Analyse des effets des propriétés thermodynamiques 273
4 Conclusion 274
B. Modélisation d'inhibiteurs non-covalents spécifiques de la protéase 3CL^{pro} basés sur le 9,10-Dihydrophénanthrène pour la thérapie ciblée contre le SARS-CoV-2 (COVID-19) 276
1 Introduction 276
2 Stratégie de conception de médicaments assistée par ordinateur développée et appliquée 279
3 Résultats et discussion 281
3.1 Exploration des relations structure-activité par 3D-QSAR 281
3.1.1 Modélisation PLS avec des descripteurs moléculaires CoMFA et CoMSIA 281
3.1.2 Interprétation des cartes de contours 3D issues des modèles CoMFA et CoMSIA 283
3.2 Modélisation générative basée sur la forme pour la conception de nouveaux médicaments 284
3.3 Analyse des profils de compatibilité et d'affinité Protéine-ligand 287
3.3.1 Criblage virtuel basé sur le Docking moléculaire 287
3.3.2 Caractérisation de la sélectivité des inhibiteurs non-covalents de 3CL^{pro} obtenus par optimisation 288
3.3.3 Sélection d'inhibiteurs prometteurs de la 3CL^{pro} basée sur l'analyse des interactions non-covalentes 292
3.4 Identification de ligands potentiels H2L par analyse des propriétés ADME-Tox in silico 296
3.5 Affinement des scores du criblage virtuel par le modèle Prime/MM-GBSA 296
3.6 Évaluation de la stabilité des interactions Protéine-Ligand par simulations de dynamique moléculaire 299
3.6.1 Analyses de RMSD et de RMSF 299
3.6.2 Caractérisation des interactions non-covalentes dynamiques dans les complexes 3CL^{pro}-DHIs 300
3.6.3 Exploration des propriétés dynamiques des ligands DHIs au sein des complexes 3CL^{pro}. 302
3.6.4 Caractérisation thermodynamique des complexes Protéine-Ligand 303
4 Conclusion 307
CONCLUSION GÉNÉRALE ET PERSPECTIVES 308
RÉFÉRENCES 314

LISTE DES FIGURES

Figure I.1. Modèle " clé-serrure" du fonctionnement des enzymes. 13
Figure I.2. Mécanisme d'inhibition de la croissance cellulaire pathologique par la thérapie ciblée. 13
Figure I.3. Schéma illustrant les mécanismes d'interaction entre le récepteur biologique et l'inhibiteur : Covalent (a) et Non-Covalent (b) 16
Figure I.4. Interactions non covalentes, (a) interactions de Van der Waals ; (b) liaison hydrogène ; (c) liaison ionique. 17
Figure I.5. Évolution des principaux événements de la découverte de médicaments. 20
Figure I.6. La chronologie de la découverte d'un seul médicament. 20
Figure I.7. Exploration précoce de médicaments. 21
Figure I.8. Illustration schématique du rôle de certaines organes du corps humain dans le processus de l'ADME des médicaments. 22
Figure I.9. Principales phases du processus de diagnostic clinique des médicaments. 24
Figure I.10. Techniques computationnelles appliquées à la conception et au développement de médicaments. 27
Figure I.11. Classification des méthodes de conception de médicament assistée par ordinateur (CADD). 28
Figure I.12. Modèle de mécanisme moléculaire Schéma des voies de signalisation en aval du HGF/c-Met et de la médiation entre c-Met et d'autres voies de signaux de transduction cellulaires [117]. 40
Figure II.1. Schéma général de la stratégie hybride de CADD conçue et appliquée dans cette thèse afin de procéder à la conception et à l'optimisation computationnelles de nouveaux médicaments.54
Figure II.2. Méthodologie générale appliquée pour la modélisation 2D-QSAR. 62
Figure II.3. Méthodologie générale appliquée pour la modélisation 3D-QSAR. 63
Figure II.4. Représentation graphique des différentes représentations moléculaires d'une même structure (un dérivé de la pyridine, à titre d'exemple). 65
Figure II.5. Exemple de quelques descripteurs moléculaires 0D calculés à partir de la formule chimique C5H5N de la pyridine. 66
Figure II.6. Illustration de la sous-structure par empreintes digitales de la molécule de N-(3-chloro-4-hydroxyphényl) pipéridine-4-carboxamide ; les sous-structures moléculaires représentées par les bits présents dans la molécule (entourés) sont définies à 1 et celles qui sont en absence sont définies à 0. 66
Figure II.7. Exemple de matrice d'adjacence d'un graphe (2,2,3-triméthylbutane). 67
Figure II.8. Exemple de matrice de distance d'un graphe (2,2,3-triméthylbutane). 68
Figure II. 9. Exemple de représentation de type Z-matrix de la structure du 2-aminoéthan-1-ol. 68
Figure II.10. Schéma des méthodes de modélisation prédictives biostatistiques appliquées dans nos travaux pour développer des modèles 2D/3D-QSAR. 84
Figure II.11. (a) Le neurone biologique, dans ce modèle les dendrites des réseaux neuronaux biologiques symbolisent les entrées dans les réseaux neuronaux artificiels (ANNs), le noyau cellulaire représente les nœuds, les synapses représentent les poids et l'axone représente la sortie ; (b) Modèle de neurones McCulloch-Pitts. 87
Figure II.12. Architecture du modèle de réseau FFNN. 88
Figure II.13. (a) Molécule de référence, (b) Cœur commun, (c) Base de données alignée. 92
Figure II.14. Cartes de contours générées par le modèle CoMFA, (a) champ d'interactions stériques (vert = favorable/jaune = défavorable), (b) champ d'interactions électrostatiques (bleu = favorable/rouge = défavorable). 93
Figure II.15. Cartes de contours générées par le modèle CoMSIA, (a) Interactions de champ stérique (vert = favorable/jaune = défavorable), (b) Interactions de champ électrostatique (bleu = favorable/rouge = défavorable), (c) Interactions de champ hydrophobe (jaune = favorable / blanc = défavorable), (d) Interactions de champ de liaisons hydrogène donneur (cyan = favorable / violet = défavorable), (e) Interactions de champ de liaisons hydrogène accepteur (magenta = favorable / rouge = défavorable). 94
Figure II.16. La méthodologie globale de CoMSIA /CoMSIA appliquée dans nos travaux. 95
Figure II.17. Schéma de la méthodologie implémentée dans nos travaux en vue d'identifier des sites pharmacophores favorables pour rationaliser la conception de nouvelles molécules. 101
Figure II.18. Modèles de criblage virtuel que nous avons mis en œuvre pour identifier des agents candidats au développement d'un médicament. 103
Figure II.19. Critères du type " Drug-like " à respecter pour sélectionner les candidats H2L dans le cadre de ce travail. 105

Figure II.20. Critères ADME-Tox impliqués dans le criblage des candidats médicaments dans nos travaux. 106
Figure II.21. Interface de l'outil "OSIRIS Property Explorer" destiné à la prédiction des risques de toxicité des petites molécules. 107
Figure II.22. Stratégie globale du modèle de criblage virtuel basé sur le Docking moléculaire appliqué dans nos études. 112
Figure II.23. L'approche du Docking moléculaire utilisée par le logiciel AutoDock. 113
Figure II.24. Présentation du squelette protéique via deux modèles : Atomistiques et gros grains (illustration en haut), modèles de paysage énergétique (illustration en bas). 116
Figure II.25. Visualisation 2D des conditions aux limites périodiques (cubiques). La boîte centrale apparaît encadrée en vert ; la particule rouge se trouve à la limite de la boîte. Les flèches indiquent le mouvement de la particule au cours du pas de temps successif. 123
Figure II.26. Schéma relatif à la liste de voisins (Neighbor list) : Pendant ce temps, la particule 1 interagit avec les particules 2, 3, 7 et 9, tandis que les particules 4 et 8 apparaissent dans la liste des voisins de la particule 1, bien qu'elles n'interagissent pas avec elle. 123
Figure II.27. Liste des cellules : Pour calculer la force d'une particule située dans la cellule i, seules les particules situées dans les cellules voisines (en violet) doivent être prises en considération. .. 124
Figure II.28. Adaptation des différentes résolutions de la modélisation de la dynamique moléculaire aux échelles de temps et aux tailles (longueurs) des systèmes [373]. 125
Figure II.29. Algorithme de base appliqué aux simulations de dynamique moléculaire. 126
Figure II.30. Principales phases des simulations de la DM. 126
Figure II. 31. Étapes générales de l'exécution d'une simulation de DM par GROMACS [404]. 127
Figure II.32. Étapes générales de l'exécution d'une simulation de DM par Desmond. 127
Figure III.1. Dimédone (a) Cyclohexane-1,3-dione (b), 4-Phénoxypyridine (c): échafaudages candidats pour la conception de nouveaux agents inhibiteurs de la tyrosine kinase c-Met pour la thérapie ciblée contre le cancer à base des petites molécules biologiquement actives synthétisées. 131
Figure III.A1. Structures de Zopolrestat (a) et de N-(6-substitué-1,3-benzothiazol-2-yl)-Benzenesulfonamide (b). 132
Figure III.A2. Structure moléculaire de l'inhibiteur Crizotinib (PF-02341066) avec une IC_{50} de 8 nM sur c-Met [421]. 132
Figure III.A3. Stratégie de CADD proposée et appliquée *in silico* pour identifier de nouveaux inhibiteurs non covalents de la protéine c-Met à base de 4,5,6,7-tétrahydrobenzo [D]-Thiazol-2-Y en vue d'une thérapie ciblée contre le cancer. 134
Figure III.A4. Les pIC_{50} observés *in vitro* versus les pIC_{50} prédits *in silico* par les modèles RLM-QSAR (a), RNLM-QSAR (b), ANN-QSAR (c) 138
Figure III.A5. Tracé de William des résidus normalisés en fonction de l'effet de levier pour le modèle RLM-QSAR original. 138
Figure III.A6. Visualisations de la poche active de la protéine c-Met dans le complexe natif 2WGJ (cMet-crizotinb) identifiée via les outils PoseView (a) et BIOVIA Discovery Studio (b). 142
Figure III.A7. Alignement de l'inhibiteur de référence Crizotinib original (couleur verte) avec le Crizotinib re-Docké (couleur marron) dans la poche active de c-Met. 143
Figure III.A8. Visualisations 2D (à droite) et 3D (à gauche) des conformations des ligands L5 (a), L16 (b) et L46 (c) complexés avec c-Met. 144
Figure III.B1. Stratégie de CADD proposée et appliquée *in silico* pour concevoir et identifier de nouveaux inhibiteurs non covalents de type II de la protéine c-Met à base de 4-phénoxypyridine en vue d'une thérapie ciblée contre le cancer. 152
Figure III.B2. (a) structure 3D de la conformation moléculaire prototype T14 (pIC_{50}=7.92), (b) unité de base commune, (c) base de données des dérivés de 4-phénoxypyridine alignée. 152
Figure III.B3. (a, b) Valeurs des pIC_{50} (*in vitro*) par rapport à celles prédites par les modèles 3D-QSAR (*in silico*), (c, d) Cartes de contours générées à partir des descripteurs CoMFA/SE et CoMSIA/SEHD utilisés par le modèle 3D-QSAR. 155
Figure III.B4. Caractéristiques des sites de pharmacopées favorables à l'optimisation de l'activité inhibitrice des dérivés de la 4-phénoxypyridine vis-à-vis de l'activité enzymatique de la protéine c-Met. 156
Figure III.B5. Voies de réactions rétro-synthétiques des inhibiteurs de c-Met proposés, (a) : ensembles I et II (P27-PP56), (b) : ensemble III (P57-P72). 157
Figure III.B6. (a) Alignement de l'ensemble de 46 (P27-P72) nouveaux dérivés de 4-phénoxypyridine par rapport à la structure de référence T14. (b) Ensemble I (P27-P41) aligné avec T14. (c) Ensemble II (P42-P56) aligné avec T14. (d) Ensemble III aligné avec T14 (P57-P72). 157

Figure III.B7. Dérivés de 4-phénoxypyridine (H2L) candidats en tant que molécules médicamenteuses potentielles *in silico*. 161
Figure III.B8. Visualisation 3D de l'alignement du complexe natif 3LQ8 et du complexe prototypique cMet-Foretinib obtenu par re-Docking moléculaire. 162
Figure III.B9. (a) Échelles de temps de l'écart quadratique moyen (RMSD) des atomes du squelette de la protéine cMet libre et complexée avec les ligands P55 (a), T14 (b) et Foretinib (c). 165
Figure III.B10. (a) Échelles de temps de la fluctuation quadratique moyenne (RMSF) des atomes Cα du squelette de la protéine cMet libre et complexée avec les ligands P55 (a), T14 (b) et Foretinib (c). 168
Figure III.B11. Rayon de giration (Rg) (a), Nombre de liaisons hydrogènes établies (b) et Surface accessible au solvant (c) des systèmes cMet non complexes et complexés (cMet-P55, cMet-T14, et cMet-Foretinib. 168
Figure III.C1. Stratégie de CADD hybride proposée et appliquée pour le criblage, la caractérisation et l'optimisation de nouvelles molécules de dérivés de cyclohexane-1,3-dione, candidates à la conception d'un nouveau médicament spécifique contre le CPNPC par le ciblage de c-Met.... 172
Figure III.C2. Valeurs des pIC_{50} des dérivés du cyclohexane-1,3-dione observées *in vitro* comparativement à celles prédites *in silco* à l'aide du modèle RLM-QSAR ; (b) Diagramme de William des résidus standard en fonction de l'effet de levier du modèle RLM-QSAR ; (c) Architecture du modèle non linéaire du RNA-QSAR ; (d) Valeurs des pIC_{50} des dérivés du cyclohexane-1,3-dione observées *in vitro* comparativement à celles prédites *in silco* à l'aide du modèle RNA-QSAR ; (e) Coefficients normalisés de chaque descripteur moléculaire choisi par le modèle RLM-QSAR. 176
Figure III.C3. Cartographie du potentiel électrostatique moléculaire (MEP) et orbitales moléculaires HOMO/LUMO pour le composé optimal 6d. 179
Figure III.C4. Modèle 3D des complexes cMet-Foretinib (3LQ8) et celui modélisé par le re-Docking (EL=-6.8 kcal/mol; RMSD=-0.2292 Å). 180
Figure III.C5. (a) Poses des ligands 6d (vert) et Foritinib (jaune) dans la poche active de c-Met, (b) Représentations 3D et 2D des interactions entre le ligand 6d et le site actif de liaison identifié (EL=-7.4 kcal/mol). 181
Figure III.C 6. Propriétés structurelles clés du ligand 6d favorables à l'optimisation de nouveaux inhibiteurs de c-Met à base de cyclohexan-1,3-dione en vue d'une thérapie future contre le CPNPC. 182
Figure III.C7. Voies de synthèse adoptées pour concevoir de nouveaux dérivés de cyclohexan-1,3-dione avec la base (6d). 182
Figure III.C8. Représentation graphique de la surface (a) aromatique, (b) de liaison hydrogène, (c) de charge, (d) d'hydrophobie, (e) d'ionisation, (f) d'accessibilité au solvant du c-Met autour de 6d. 183
Figure III.C9. Visualisations 2D des interactions identifiées dans les complexes cMet-H06, cMet-H12, cMet-H30 et cMet-H36 par docking moléculaire (à gauche) et par calculs Prime MM-GBSA (à droite). 190
Figure III.C10. Trajectoires temporelles de RMSD (a) et de RMSF (b) des atomes du squelette des systèmes de c-Met à l'état libre et complexés avec H06, H12, H30, H36 et Foretinib. RMSF (c) trajectoires temporelles des inhibiteurs proposés H06, H12, H30, H36 et Foretinib dans la poche active de c-Met. 192
Figure III.C11. Diagrammes des interactions cMet-ligand générés par des simulations de dynamique moléculaire au cours de 100 ns. 194
Figure III.C12. Profil des propriétés dynamiques des structures des ligands (H06, H12, H30, H36) avec la structure du c-Met pendent 100 ns de simulation de la DM. 194
Figure III.C13. Qualité du protocole de dynamique moléculaire mis en œuvre à travers l'évaluation des propriétés thermodynamiques des systèmes cMet-inhibiteur. 195
Figure IV.1. *Magnolia officinalis* (a), *Cannabis sativa L.* (b) : phytochimiques candidats pour le développement d'une nouvelle génération d'inhibiteurs EGFR-TKD comme stratégie de thérapie ciblée contre le cancer. 200
Figure IV. A1. Stratégie de CADD hybride proposée et appliquée pour le criblage, la caractérisation et l'optimisation de nouvelles molécules à base de Magnolol, candidates à la conception d'un nouveau médicament anti-CPNPC ciblant spécifiquement l'EGFR-TKD. 202
Figure IV. A2. (a) Base de données de dérivés de Magnolol alignés, (b) Structure de la molécule de référence "30", c) Unité de base commune. 204
Figure IV. A3. (a) Distribution des valeurs pIC_{50} obtenues *in vitro* pour les dérivés du Magnolol par rapport à celles prédites par le modèle 3D-QSAR *in silico*. (b) Diagrammes de Williams des

domaines d'applicabilité du modèle 3D-QSAR élaboré à l'aide des descripteurs CoMFA et CoMSIA. 206
Figure IV. A4. Cartes de contours générées à partir des analyses CoMFA (a) et CoMSIA (b-f) des interactions stériques, électrostatiques, hydrophobiques, donneurs de liaisons H, accepteurs de liaisons H (favorables et défavorables) pour optimiser l'activité biologique inhibitrice de la structure de référence des dérivés de Magnolol examinés contre le cancer du poumon non à petites cellules (CPNPC)..... 207
Figure IV. A5. Visualisation des modes de contacts EGFR-Erlotinib dans le complexe original 1M17 et le complexe remodelé par re-Docking moléculaire. (a) visualisation en 3D de la disposition de l'alignement du Foretinib dans le complexe original (couleur noire) et le complexe remodelé (couleur jaune) à l'intérieur de la poche active d'EGFR-TKD. (b) visualisation en 2D des interactions entre l'Erlotinib redocké dans la cavité active de l'EGFR-TKD. 209
Figure IV. A6. Visualisation 2D des interactions entre le dérivé de Magnolol de référence #30 et la poche active d'EGFR-TKD (énergie de liaison estimée à -7.9 kcal/mol). 210
Figure IV. A7. Visualisation MolCAD de la surface moléculaire du ligand bioactif #30 à l'intérieur d'une poche active de l'EGFR-TKD. (a) (CD) surface de la profondeur de la cavité potentielle dans le complexe EGFR-30, la rampe de couleur de la profondeur de la cavité varie du bleu (valeurs de faible profondeur = à l'extérieur de la poche) à l'orange (valeurs de forte profondeur = cavités profondes à l'intérieur de la poche). (b) (LP) surface du lipophile potentiel dans le complexe EGFR-30, la rampe de couleur pour LP varie du brun (la région hydrophobe la plus élevée de la surface) au bleu (la région hydrophile la plus élevée). (c) (HB) liaison hydrogène au niveau du complexe EGFR-30, la rampe de couleur pour HB varie du rouge (donneurs d'hydrogène) au bleu (accepteurs d'hydrogène). (d) (EP) potentiel électrostatique de surface au niveau du complexe EGFR-30, la rampe de couleur pour EP varie du rouge (région la plus positive) au bleu (région la plus négative). 212
Figure IV. A8. Propriétés bio-structurales optimales favorables à l'amélioration de l'activité biologique inhibitrice et affinité de liaison au poche active de l'EGFT-TKD des dérivés du Magnolol pour lutter contre le CPNPC. 213
Figure IV. A9. Représentation en 2D des interactions non covalentes observées dans le complexe EGFR-X10. 217
Figure IV. A10. Interactions non covalentes protéine-ligand régénérées après avoir soumis les complexes EGFR-X10, EGFR-30 et EGFR-Erlotinib aux simulations de MM-GBSA..... 219
Figure IV. B1. La stratégie CADD implémentée dans cette étude pour identifier de nouveaux agents candidats comme EGFR-TKIs basés sur des substances phytochimiques bioactives de type cannabinoïdes et terpènes pour la thérapie ciblée contre le cancer. 226
Figure IV. B2. Positions conformationnelles de l'inhibiteur Erlotinib original (jaune) et re-Docké (noir) au sein de la poche active flexible d'EGFR-TKD. 227
Figure IV. B3. Prédictions de modèle de Docking moléculaire flexible appliquées aux nouveaux EGFR-TKIs basés sur les phytochimiques de type Cannabinoïdes et Terpènes. 232
Figure IV. B4. Prédictions du modèle Prime/MM-GBSA appliquées aux complexes EGFR-TKIs générés. 233
Figure IV. B5. Interactions EGFR-C11. (a) Prédictions de Docking moléculaire flexible. (b) Prédictions de Prime/MM-GBSA. Les acides aminés entourés d'un cercle rouge indiquent un site de liaison actif de référence, les acides aminés entourés d'un cercle bleu indiquent de nouveaux sites actifs de liaison. 235
Figure IV. B6. Interactions EGFR-T15. (a) Prédictions de Docking moléculaire flexible. (b) Prédictions de Prime/MM-GBSA. Les acides aminés entourés d'un cercle rouge indiquent un site de liaison actif de référence, les acides aminés entourés d'un cercle bleu indiquent de nouveaux sites actifs de liaison. 235
Figure IV. B7. Interactions EGFR-Erlotinib. (a) Prédictions de Docking moléculaire flexible. (b) Prédictions de Prime/MM-GBSA. Les acides aminés entourés d'un cercle rouge indiquent un site de liaison actif de référence, les acides aminés entourés d'un cercle bleu indiquent de nouveaux sites actifs de liaison. 236
Figure IV. B8. Interactions EGFR-Tamoxifen. (a) Prédictions de Docking moléculaire flexible. (b) Prédictions de Prime/MM-GBSA. Les acides aminés entourés d'un cercle rouge indiquent un site de liaison actif de référence, les acides aminés entourés d'un cercle bleu indiquent de nouveaux sites actifs de liaison. 237
Figure IV. B9. Tracés temporels des simulations de la DM pour les systèmes d'EGFR libres et complexés avec les ligands C11, T15, Tamoxifen, et Erlotinib. (a) RMSD des squelettes C-alpha

des systèmes EGFR, (b) RMSF des chaînes latérales des systèmes EGFR, (c) RMSF des ligands complexés avec l'EGFR. 238
Figure IV. B10. Diagrammes détaillés de la distribution des interactions locales entre les ligands C11 (THCA), T15 (BCP), le Tamoxifen et l'Erlotinib dans la poche active de l'EGFR pendant 200ns de simulations de la DM. 240
Figure IV. B11. Tracés des variations des propriétés des ligands tout au long de la trajectoire de la simulation de la dynamique moléculaire (0 à 200nsec). 242
Figure IV. B12. Propriétés thermodynamiques en fonction du temps de l'énergie totale, de l'énergie potentielle, de la température, de la pression et du volume pour l'EGFR libre (a) et l'EGFR complexé avec les substances C11 (b), T15 (c), Tamoxifen (d) et Erlotinib (e). 243
Figure V. A1. Conception de nouveaux inhibiteurs allostériques de MALT1 pour la thérapie ciblée contre le lymphome, utilisant un échafaudage moléculaire pyridinique. 247
Figure V. A2. Les échafaudages à base de pyridine de 40 petites molécules étudiées. 248
Figure V. A3. Stratégie de CADD hybride proposée et appliquée pour la conception, criblage, la caractérisation et l'optimisation de nouvelles inhibiteurs allostériques de la protéase MALT1 à base de Pyridine, candidates à la conception d'un nouveau médicament Anti-Lymphome. 249
Figure V. A4. (a) Distribution des valeurs pIC_{50} observées *in viro* et prédites *in silico*. (b) Répartition des degrés des sept descripteurs moléculaires dans le modèle de RLM/2D-QSAR par les coefficients normalisés. 252
Figure V. A5. Domaine d'applicabilité du modèle RLM/2D-QSAR défini par le diagramme de Williams. 253
Figure V. A6. (a) Base de données alignée de petites molécules à base de pyridine, (b) conformation de référence du ligand bioactif D40, (c) sous- structure commune. 253
Figure V. A7. Distribution des valeurs pIC_{50} observées *in vitro* et prédites par le modèle 3D-QSAR *in silico* qui est basé sur les descripteurs des champs moléculaires de CoMFA et CoMSIA. 255
Figure V. A8. Cartes des contours stériques (a) et électrostatiques (b) générées par le modèle 3D-QSRA/CoMFA. 256
Figure V. A9. (a) stérique, (b) électrostatique, (c) hydrophobique, (d) donneur de liaison hydrogène, (e) accepteur de liaison hydrogène : cartes des contours fournies par le modèle 3D-QSAR/CoMSIA. 256
Figure V. A10. Propriétés structurelles optimales favorables à la génération de nouvelles petites molécules à base de pyridine ayant une activité inhibitrice et une affinité vis-à-vis de la poche allostérique active de la MALT1. 257
Figure V. A11. Positions des conformations de l'inhibiteur allostérique de référence Thioridazine (TH) (original/vert, re-docké/rouge) dans la poche allostérique active de MALT1. 258
Figure V. A12. Pose optimale de la conformation de la petite molécule à base de pyridine (D40) au sein du complexe MALT1-D40. 259
Figure V. A13. Représentation 3D des cavités de surface du système MALT1-D40. (a) Surface aromatique, (b) Surface d'accessibilité au solvant de MALT1 vis-à-vis de D40, (c) Surface des liaisons hydrogène, (d) Surface de charge, (e) Surface d'ionisation, (f) Surface hydrophobique. 261
Figure V. A14. Schéma des substitutions proposées pour optimiser de nouveaux inhibiteurs allostériques de MALT1 à base de pyridine. 262
Figure V. A15. Graphique de AlogP98 vs 2D-PSA pour les quatre nouveaux composés optimisés à base de pyridine. 265
Figure V. A16. Caractérisations visuelles des profils de compatibilité/affinité des systèmes MALT1-P2 et MALT1-P3 modélisés par docking moléculaire et calculs de Prime/MM-GBSA. 267
Figure V. A17. (a) RMSD (b) RMSF des systèmes : (MALT1 free, MAL1-P2, MALT1-P3 et MALT1-Thioridazine), (c) RMSF timeline des ligands (P2, P3 et Thioridazine) tout au long des 100ns de simulations de la dynamique moléculaire. 269
Figure V. A18. Propriétés dynamiques des nouveaux inhibiteurs allostériques optimisés de MALT1 P2 (a) et P3 (b), et de l'inhibiteur standard Thioridazine (c). 270
Figure V. A19. Rapports de diagrammes protéine-ligand obtenus après 100 ns de simulations de dynamique moléculaire pour les complexes MALT1-P2 (a), MALT1-P3 (b) et MALT1-Thioridazine (c). 270
Figure V. A20. Représentation graphique du modèle Thermal/MM-GBSA concernant les scores d'énergie libre de liaison mesurés pendant 100ns de simulation de dynamique moléculaire. 272
Figure V. A21. Caractéristiques thermodynamiques des inhibiteurs de MALT1 examinés, (a) MALT1_free, (b) MALT1-P2, (c) MALT1-P3, (d) MALT1-Thioridazine. 274

Figure V. B1. Récents inhibiteurs covalents (a) et non-covalents (b) de 3CL^{pro} destinés au SARS-CoV-2 dans le cadre de la recherche d'un médicament anti-COVID19. 277
Figure V. B2. 9,10-Dihydrophénanthrène : échafaudage candidat pour le développement de nouveaux inhibiteurs de SRAS-CoV-2-3CL^{pro} pour la thérapie de COVID-19. 279
Figure V. B3. Stratégie de CADD hybride proposée et appliquée pour la conception, criblage, la caractérisation et l'optimisation de nouveaux inhibiteurs non-covalents de SARS-CoV-2 3CL^{pro} à base de 9,10-Dihydrophénanthrène, candidates à la conception d'un nouveau médicament Anti-COVID19. 281
Figure V. B4. (a) conformation de référence D40, (b) base de sous-structure commune, (c) base de données alignée. 282
Figure V. B5. Cartes de contours 3D générées par le modèle CoMFA/SE, (a) interactions de champ stérique (vert = 80% favorable/jaune = 20% défavorable), (b) interactions de champ électrostatique (bleu = 80% favorable/rouge = 20% défavorable). (**c**) pIC_{50} observés vs pIC_{50} prédits. 283
Figure V. B6. Cartes de contours 3D générées par le modèle CoMSIA, (a) interactions de champ stérique (vert = favorable/jaune = défavorable), (b) interactions de champ électrostatique (bleu = 80% favorable/rouge = 20% défavorable), (c) interactions de champ hydrophobe (jaune = 80% favorable / blanc = 20% défavorable), (d) interactions de champ de liaison hydrogène-donneur (cyan = 80% favorable / violet = 20% défavorable), (e) interactions de champ de liaison hydrogène-accepteur (magenta = 80% favorable / rouge= 20% défavorable). (f) pIC_{50} observés versus pIC_{50} prédits. 283
Figure V. B7. Caractérisation des propriétés structurelles favorables à la conception et à l'optimisation de nouveaux inhibiteurs du SARS-CoV-2 3CL^{pro} par la structure (D40). 285
Figure V. B8. Visualisation en 3D des séquences d'alignement moléculaire entre la molécule de référence (D40) et ses nouvelles dérivées (DH1-DH96). 285
Figure V. B9. Nouveaux inhibiteurs (DH) proposés pour le SARS-CoV-2 3CL^{pro} générés à partir du composé D40 et identifiés par le criblage in *silico* de 3D-QSAR et de Drug-like. 286
Figure V. B10. Poses de re-Docking par superposition 3D/2D de l'inhibiteur covalent N3 original (noir) et re-Docké (jaune) dans la poche active de la 3CL^{pro} (complexe 6LU7). 289
Figure V. B11. Écart quadratique moyen entre les positions atomiques de la protéine 3CL^{pro} et de l'inhibiteur N3 dans le complexe 6LU7. (b) Visualisation 2D des contacts établis entre 3CL^{pro} et N3 qui persistent au-delà de 30 % dans la simulation de la MD. (c) Histogramme des contacts entre 3CL^{pro} et N3 pendant 100 ns de la simulation de la MD. 290
Figure V. B12. Visualisation 3D des poses du Docking moléculaire des meilleures conformations des vingt nouveaux composés DH optimisés, D40, N3 et DSF dans la poche active du SARS-CoV-2 3CL^{pro}. 291
Figure V. B13. Histogramme des énergies libres de liaison recalculées par le modèle Prime/MM-GBSA pour les systèmes 3CL^{pro}-inhibiteurs. 297
Figure V. B14. Diagrammes temporels des variations RMSD (a) et RMSF (b) correspondant aux atomes du squelette des systèmes 3CL^{pro}_free, 3CL^{pro}-DH8, 3CL^{pro}-DH23, 3CL^{pro}-DH76, et 3CL^{pro}-D40. (c) Diagrammes temporels RMSF des atomes ligands DH8, DH23, DH76, et D40 à l'intérieur de la poche active de 3CL^{pro}. 300
Figure V. B15. Tracés temporels en pourcentage des contacts protéine-ligand (PL) enregistrés au cours des 100ns de simulations de dynamique moléculaire effectuées sur les systèmes biomoléculaires 3CL^{pro}-DH8, 3CL^{pro}-DH23, 3CL^{pro}-DH76, et 3CL^{pro}-D40. 301
Figure V. B16. Visualisation 2D des contacts enregistrés au niveau de la poche active de la 3CL^{pro} et des DHIs optimisées (DH8, DH23, DH76), ainsi que de la référence D40 pendant 100ns de simulations de dynamique moléculaire. 301
Figure V. B17. Diagrammes temporels de l'évolution des propriétés dynamiques des ligands DH8, DH23, DH76 et D40 examinés dans la poche active de la 3CL^{pro} au cours des 100ns de trajectoire des simulations de dynamique moléculaire. 305
Figure V. B18. Diagrammes des propriétés thermodynamiques enregistrées au cours de 100ns des simulations de la dynamique moléculaire appliquées aux systèmes biomoléculaires de 3CL^{pro} (3CL^{pro}_free, 3CL^{pro}-DH8, 3CL^{pro}-DH23, 3CL^{pro}-DH76, et 3CL^{pro}-D40). 306

LISTE DES TABLEAUX

Tableau I.1. Comparaison Approfondie : Produits Biologiques vs Petites Molécules [27]. 15
Tableau I. 2. Exemples de succès de CADD, d'IA et de ML [72], [81], [82], [83] 33
Tableau II.1. Exemples des bases de données virtuelles (chimiothèques). 51
Tableau II.2. Classification de la modélisation QSAR par dimensions. 59
Tableau II.3. Méthodes de calcul et champs de force les plus utilisés. 70
Tableau II.4. Outils bio-informatique et méthodes d'optimisation géométrique et énergétique utilisés dans nos travaux pour générer des descripteurs moléculaires. 72
Tableau II.5. Descripteurs moléculaires calculés dans nos travaux à l'aide d'ACD/ChemSketch sur la base de la méthode d'optimisation 3D. 72
Tableau II.6. Descripteurs moléculaires calculés dans nos travaux par Chem3D sur la base de la méthode d'optimisation MM2. 74
Tableau II.7. Les descripteurs moléculaires électroniques et thermodynamiques résultent aux calculs DFT menés dans nos travaux sur la base de la méthode B3LYP/6-31G(d,p) de Gaussian 09W. 75
Tableau II.8. Outils de préparation des données d'entrée et de mise en œuvre des modèles QSAR utilisés dans nos travaux. 84
Tableau II.9. Classification des techniques 3D-QSAR selon des critères spécifiques[235]. 91
Tableau II.10. Critères statistiques utilisés dans cette thèse pour la validation du modèle 2D/3D-QSAR. 98
Tableau II.11. Exemples de fonctions de score et de leurs classes correspondantes employées par certains outils de Docking moléculaire. 109
Tableau II.12. Logiciels de Docking protéine-ligand employés dans le cadre de cette thèse. 114
Tableau II.13. Répertoire (non exhaustif) des logiciels de dynamique moléculaire les plus connus. 127
Tableau III.A1. Données de structure-activité disponibles pour les dérivés de 4,5,6,7-Tétrahydrobenzo[D]-Thiazol-2-Yl issus de Dimédone. 135
Tableau III.A2. Critères statistiques pour la validation des modèles 2D-QSAR. 137
Tableau III.A3. Les molécules basées sur le 4,5,6,7-Tétrahydrobenzo[D]-thiazol-2-Yl candidats pour inhiber la protéine c-Met plus efficacement que le Crizotinib. 139
Tableau III.A4. Prédiction des propriétés de type Drug-like. 140
Tableau III.A5. Interactions protéine-ligand, énergies libres de liaison estimées par Docking moléculaire dans les complexes de c-Met. 145
Tableau III.A6. Propriétés ADME-Tox prédites *in silico* pour le composé "hit" (16) et le médicament standard Crizotinib (indiqué par Criz.). 146
Tableau III.B1. Données structure-activité observées *in vitro* pour les 26 dérivés de 4-phénoxypyridine vis-à-vis la protéine c-Met (T1-T26). 150
Tableau III.B2. Résultats des analyses PLS pour les modèles CoMFA et coMSIA générés. 153
Tableau III.B3. Les valeurs de pIC_{50} des dérivés de la 4-phénoxypyridine observés *in vitro* par rapport à ceux prédits par les descripteurs CoMFA et CoMSIA *in silico*. 154
Tableau III.B4. Test de Y-randomisation. 155
Tableau III.B5. Les valeurs de pIC_{50} calculées *in silico* pour la série de dérivés de 4-phénoxypyridine conçus dans cette étude sur la base des modèles CoMFA/SE et CoMSIA/SEHD. 158
Tableau III.B6. Profil "Drug-like" des dérivés de la 4-phénoxypyridine proposés et du prototype (T14) et du médicament standard Foretinib. 159
Tableau III.B7. Propriétés ADME-Tox prédites *in silico* pour de nouvelles molécules à base de 4-phénoxypyridine conçues à partir de la structure de base T14 et du médicament standard Foretinib. 160
Tableau III.B8. Sorties du Docking moléculaire en termes d'énergie de liaison et de constante d'inhibition estimée (Ki) des inhibiteurs de c-Met développés. 163
Tableau III.B9. Interactions protéine-ligand identifiées par le modèle de Docking moléculaire. 166
Tableau III.B10**.** Suite du tableau. 167
Tableau III.C1. Structures moléculaires de dérivés de cyclohexane-1,3-dione optimisées par la méthode de la DFT conceptuelle avec informations *in vitro* sur leur activité inhibitrice « pIC_{50} » vis-à-vis de la lignée cellulaire H460 du cancer du poumon non à petites cellules (CPNPC). 173
Tableau III.C2. Criblage *in silico* basé sur un modèle de type Drug-like. 177

Tableau III.C3. Structures-activités de cyclohexane-1,3-dione identifiées comme candidates à la conception de nouveaux médicaments. 178
Tableau III.C4. Structure-activité optimisée *in silico* de nouveaux dérivés de cyclohexan-1,3-dione en tant qu'inhibiteurs prometteurs de la protéine c-Met anti-CPNPC. 184
Tableau III.C5. (a) Profil de type Drug-like appliqué à des nouveaux dérivés optimisés de la cyclohexane-1,3-dione. 185
Tableau III.C6. Propriétés ADME-Tox prédites *in silico*. 188
Tableau III.C7. Résultats du criblage virtuel basé sur le Docking moléculaire de candidats inhibiteurs de c-Met. 189
Tableau III.C8. Bilan des simulations de Docking moléculaire et du calcul de l'énergie libre de liaison par le modèle Prime/MM-GBSA. 191
Tableau III.C9. Score moyen des propriétés thermodynamiques des systèmes examinés. 193
Tableau IV. A1. Données structure-activité des dérivés de Magnolol basés sur la plante médicinale *Magnolia officinalis* contre les lignées cellulaires du CPNPC obtenues par des essais *in vitro*.203
Tableau IV. A2. Analyse statistique selon la méthode PLS. 205
Tableau IV. A3. Résultats du test de Y-randomisation. 206
Tableau IV. A4. Activité biologique inhibitrice pIC_{50} prédite *in silico* contre le CPNPC de la série de dérivés du Magnolol optimisés dans cette étude. 214
Tableau IV. A5. Profil des propriétés de type drug-like des substances candidates. 214
Tableau IV. A6. Profil ADME-Tox des dérivés de Magnolol proposés comme H2L computationnel. . 215
Tableau IV. A7. Résumé des résultats d'un criblage virtuel basé sur le Docking moléculaire des inhibiteurs spécifiques de l'EGFR. 216
Tableau IV. A8. Principaux résultats des calculs Prime/MM-GBSA. 218
Tableau IV. B1. Chimiothèque des substances phytochimiques à base de Terpènes (T1-T38) et de Cannabinoïdes (C1-C12) extraits de *Cannabis sativa L.* 225
Tableau IV. B2. Compilation des affinités de liaison des candidats inhibiteurs de l'EGFR-TKD. 228
Tableau IV. B3. Modèle de type Drug-like et risques de toxicité associés aux substances criblées.. 229
Tableau IV. B4. Analyse *in silico* des propriétés ADME-Tox des substances phytochimiques candidates comme H2L. 231
Tableau IV. B5. Valeurs moyennes estimées pour les propriétés de l'énergie totale (E), de l'énergie potentielle (EP), de la température (T), de la pression (P) et du volume (V) pour les systèmes de protéines EGFR examinés. 241
Tableau V. A1. Informations sur la structure et l'activité obtenues *in vitro* pour les petites molécules à base de pyridine vis-à-vis de l'activité enzymatique de la MALT1. 250
Tableau V. A2. Test de Y-randomisation appliqué au modèle RLM/2D-QSAR. 253
Tableau V. A3. Bilan statistique des analyses PLS de CoMFA et CoMSIA. 255
Tableau V. A4. Test de randomisation Y réalisé sur le modèle 3D-QSAR (CoMFA et CoMSIA). 255
Tableau V. A5. Profil de compatibilité/affinité protéine-ligand dans les complexes MALT1-D40 et MALT1-TH. 260
Tableau V. A6. Les valeurs pIC_{50} observées *in vitro* et prédites *in silico* pour les petites molécules à base de pyridine obtenues par la synthèse (A1 à D40) et celles optimisées dans le cadre de cette étude (P1 à P4). 263
Tableau V. A7. Modèle de type Drug-like pour évaluer les molécules optimisées du point de vue de leurs biodisponibilités. 264
Tableau V. A8. Évaluation des propriétés pharmacocinétiques/pharmacodynamiques *in silico* des molécules candidates comme H2L via le modèle ADMT-Tox. 265
Tableau V. A9. Mécanismes d'interactions non covalentes protéine-ligand des complexes MALT1-P1 et MALT1-P2 modélisés dans cette étude vis-à-vis des complexes de référence MALT1-TH et MALT1-D40. 268
Tableau V. A10. Aperçu général des contacts protéine-ligand observés lors des simulations de dynamique moléculaire comparé aux prédictions du modèle de Docking moléculaire mis en œuvre sur les complexes MALT1-P2, MALT1-P3, et MALT1-Thioridazine. 271
Tableau V. A11. Moyennes des énergies libres de liaison estimées par le modèle Thermal/MM-GBSA pendant 100ns de dynamique moléculaire pour les complexes MALT1-P2, MALT1-P3 et MALT1-Thioridazine. 272

Tableau V. A12. Valeurs moyennes des propriétés thermodynamiques calculées à partir des trajectoires de 100ns simulées par la DM pour les systèmes biomoléculaires de MALT (MALT_free, MALT1-P2, MALT1-P3, et MALT1-Thioridazine). 274
Tableau V. B1. Activité inhibitrice *in vitro* des dérivés du 9,10-dihydrophénanthrène (D1-D46) vis-à-vis de la $3CL^{pro}$ du SARS-CoV-2. 280
Tableau V. B2. Bilan des performances en matière de signification statistique du modèle 3D-QSAR de type CoMFA et CoMSIA..... 282
Tableau V. B3. Prédiction *in silico* des pIC_{50} de 47 nouveaux dérivés de 9,10-dihydrophénanthrène optimisés sur la base de la structure de référence D40 contre le SARS-CoV-2 $3CL^{pro}$..... 286
Tableau V. B4. Score des énergies de liaison obtenues par le modèle de Docking moléculaire appliqué aux candidats inhibiteurs de $3CL^{pro}$ et aux références. 291
Tableau V. B5. Les profils d'interactions non-covalentes protéine-ligand prédits par la simulation de docking moléculaire flexible. 292
Tableau V. B6. Profil d'ADME-Tox des candidats H2L. 298
Tableau V. B7. Re-scores par Prime/MM-GBSA des énergies libres de liaison des nouveaux $3CL^{pro}$-DHIs optimisés par rapport aux systèmes biomoléculaires de référence $3CL^{pro}$-N3 et $3CL^{pro}$-D40. 299
Tableau V. B8. Contacts $3CL^{pro}$-DHIs enregistrés sur les trajectoires de 100ns des simulations de dynamique moléculaire appliquées aux échantillons de $3CL^{pro}$-DH8, $3CL^{pro}$-DH23, $3CL^{pro}$-DH76 et $3CL^{pro}$-D40 de type macromoléculaire. 302
Tableau V. B9. Valeurs moyennes des propriétés thermodynamiques enregistrées pendant 100ns de simulations de dynamique moléculaire pour les systèmes biomoléculaires de $3CL^{pro}$. 304

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

Symboles/Abréviations	Significations
CADD	Conception de médicaments assistée par ordinateur
COVID-19	Maladie à coronavirus 2019
LBDD	Conception de médicaments basée sur le ligand
SBDD	Conception de médicaments basée sur la structure
ML	Apprentissage automatique
IA	Intelligence Artificielle
ARN	Acide ribonucléique
ADN	Acide désoxyribonucléique
0D	Zéro dimension
1D	Unidimensionnel
2D	Bidimensionnel
3D	Tridimensionnel
4D	Quadridimensionnel
5D	Quintidimensionnel
6D	Six-dimensionnel
QSAR	Relation quantitative structure-activité
QSPR	Relation quantitative structure-propriété
RLM	Régression linéaire multiple
RNLM	Régression non linéaire multiple
PLS	Régression par moindres carrés partiels
ANN	Réseau de neurones artificiels
RNA	Réseau de neurones artificiels
ACP	Analyse en composantes principales
K-means	Classification par k-moyennes
UV	Classification univariée
Drug-like	Similaire à un médicament
ADME-Tox	Absorption, Distribution, Métabolisme, Excrétion et Toxicité
FDA	Administration des aliments et des médicaments
EGFR-TKI	Inhibiteur de la tyrosine kinase du récepteur du facteur de croissance épidermique
IC_{50}	Concentration inhibitrice médiane
K_i	Constante d'affinité/inhibition
HTS	Criblage à haut débit
HCS	Criblage à haut contenu
H2L	Hit-To-Lead
LO	Lead Optimisation
CDER	Centre d'évaluation et de recherche sur les médicaments
HTVS	Criblage Virtuel à Haut Débit
RCSB PDB	Banque de données sur les protéines
IRMf	Imagerie par résonance magnétique fonctionnelle
TEP	Tomographie par émission de positrons
CRISPR/Cas9	Répétitions palindromiques groupées et régulièrement intercalées
GWAS	Etude d'association à l'échelle du génome
TKD	Domaine tyrosine kinase
c-Met	Transition mésenchymateuse et épithéliale basée sur la cystéine
EGFR	Récepteur du facteur de croissance épidermique
MALT1	Protéine de translocation du lymphome du tissu lymphoïde associé aux muqueuses 1
SARS-CoV-2	Coronavirus du syndrome respiratoire aigu sévère 2
$3CL^{pro}$	Protéase de type 3-chymotrypsine
COVID-19	Maladie à coronavirus 2019
GLOBOCAN	Observatoire mondial du cancer
CPNPC	Cancer du poumon à petites cellules
HGF	Facteur de croissance hépatocytaire

PI3K/AKT	Phosphoinositide 3-kinase/ Protéine kinase B
Ras/MAPK	Protéine kinase activée par Ras/mitogène
JAK/STAT	Janus kinase/transducteurs de signaux et activateurs de transcription
SAR	Relation structure-activité
MM-GBSA	Mécanique Moléculaire Généralisée Surface de Born
MM-PBSA	Mécanique Moléculaire de la Zone de Surface de Poisson-Boltzmann
CoMFA	Analyse comparative des champs moléculaires
CoMSIA	Analyse Comparative des Indices de Similitude Moléculaire
DFT	Théorie de la fonctionnelle de la densité
3D-MoRSE	Représentation moléculaire des structures basée sur la diffraction électronique
MM	Mécanique Moléculaire
MQ	Mécanique Qantique
HF	Hartree-Fock
MMFF	Champ de force moléculaire de Merck
AMBER	Assistance à la construction du modèle et raffinement de l'énergie
OPLS	Optimisation des potentiels pour la simulation des liquides
CHARMM	Chimie à Harvard Mécanique macromoléculaire
GROMOS	Groningen Simulation moléculaire
UFF	Champ de force universel
DREIDING	Champ de force de Dreiding
MOPAC	Package des orbitales moléculaires
CI	Interaction de configuration
CADD	Conception de médicaments assistée par ordinateur
COVID-19	Maladie à coronavirus 2019
LBDD	Conception de médicaments basée sur le ligand
SBDD	Conception de médicaments basée sur la structure
ML	Apprentissage automatique
IA	Intelligence Artificielle
ARN	Acide ribonucléique
ADN	Acide désoxyribonucléique
CI	Interaction de configuration
B3LYP	Lee-Yang-Parr à trois paramètres
DA	Domaine d'applicabilité
PK/PD	Pharmacocinétique et pharmacodynamique
DM	Dynamique moléculaire
RMSD	Écart quadratique moyen
RMSF	Fluctuation quadratique moyenne
LDA	Approximation de la densité locale
GGA	Approximation du gradient généralisé
HOMO	Orbital moléculaire le plus occupé
LUMO	Orbite moléculaire la moins occupée
PEM	Potentiel électrostatique moléculaire
RMN	Résonance magnétique nucléaire
DBSCAN	Classification spatiale basée sur la densité des applications avec du bruit

RÉSUMÉ DE LA THÈSE

La recherche incessante d'agents thérapeutiques innovants pour combattre les maladies invalidantes reste un pilier de l'industrie pharmaceutique. Malgré des investissements substantiels, des délais de développement prolongés et des taux d'échec élevés, la quête de médicaments efficaces se poursuit. Dans ce contexte, cette thèse présente une approche novatrice qui exploite la puissance de la conception de médicaments assistée par ordinateur (CADD) pour révolutionner le processus de découverte de médicaments, minimisant les risques d'échec et accélérant le développement de thérapies salvatrices. Au cœur de cette thèse se trouve l'identification méticuleuse, le criblage et la modélisation précise des propriétés thérapeutiques des ligands bioactifs, améliorant leurs perspectives en tant que médicaments potentiels. Pour y parvenir, nous avons utilisé des techniques de pointe de modélisation biostatistique et bio-informatique, notamment 2D/3D-QSAR, les réseaux de neurones artificiels (RNA), la théorie fonctionnelle de la densité (DFT), et l'analyse comparative de champ moléculaire et les indices de similarité (CoMFA & CoMSIA). Ces approches sophistiquées ont permis la génération de modèles prédictifs hautement précis qui décrivent méticuleusement les relations structure-activité des ligands criblés. De plus, nous avons mis en œuvre des simulations biomoléculaires avancées utilisant des techniques de Docking Moléculaire semi-flexible et/ou flexible, MolCAD, MM-GBSA et la Dynamique Moléculaire. Ces simulations ont fourni une large gamme de données sur la stabilité et la compatibilité des systèmes macromoléculaires protéine-ligand, mettant en lumière les interactions complexes entre ces entités. De plus, nous avons appliqué rigoureusement des modèles de criblage virtuel de type « Drug-like » et « ADME-Tox » pour évaluer méticuleusement les aspects de biodisponibilité, de pharmacocinétique et de pharmacodynamie des ligands candidats, garantissant leur adéquation pour un développement ultérieur. Pour démontrer l'efficacité de notre approche hybride CADD proposée, nous avons ciblé certaines des maladies les plus dévastatrices, notamment le cancer (poumon, sein, lymphome) et les infections virales (SARS-CoV-2). Dans le domaine de la thérapie ciblée contre le cancer, nous avons identifié et optimisé avec succès de nouvelles générations d'inhibiteurs non covalents ciblant spécifiquement le récepteur tyrosine-kinase c-Met, en utilisant des dérivés synthétiques de Dimédone, de 4-phénoxypyridine et de Cyclohexanedione-1,3. De plus, nous avons développé de nouveaux inhibiteurs non covalents ciblant la protéase MALT1 à base de dérivés de Pyridine. En outre, nous avons identifié et optimisé de nouveaux inhibiteurs non covalents ciblant spécifiquement le récepteur tyrosine-kinase d'EGFR, en exploitant la puissance des substances phytochimiques dérivées de *Magnolia officinalis* et de Cannabis sativa L. Dans le contexte de la thérapie antivirale ciblée spécifiquement liée à la pandémie de SRAS-CoV-2 (COVID-19), nous avons optimisé une nouvelle génération d'inhibiteurs non covalents ciblant spécifiquement l'enzyme $3CL^{pro}$, en utilisant des dérivés du 9,10-Dihydrophénanthrène. Le succès retentissant de nos efforts démontre sans équivoque l'efficacité et la rationalité remarquables de l'approche hybride CADD adoptée dans cette thèse. Cette approche s'est avérée être un outil précieux pour le criblage virtuel et l'optimisation *in silico* des ligands candidats bioactifs, ouvrant la voie à la conception et au développement de médicaments à un stade précoce.

Mots clés : CADD, *in silico*, 2D/3D-QSAR, DFT, RNA, MolCAD, MM-GBSA, Docking moléculaire, Dynamique moléculaire, Drug-like/ADME-Tox, Inhibiteurs non-covalents, Anticancer, Antivirale.

ABSTRACT

The relentless pursuit of novel therapeutic agents to combat debilitating diseases remains a cornerstone of the pharmaceutical industry. Despite substantial investments, protracted development timelines, and high failure rates, the quest for effective medications continues. In this context, this thesis presents a novel approach that leverages the power of computer-aided drug design (CADD) to revolutionize the drug discovery process, minimizing the risk of failure and accelerating the development of life-saving therapies. At the heart of this thesis lies the meticulous identification, screening, and precise modeling of the therapeutic properties of bioactive ligands, enhancing their prospects as potential drugs. To achieve this, we have employed cutting-edge biostatistical and bioinformatics modeling techniques, including 2D/3D-QSAR, artificial neural networks (ANNs), density functional theory (DFT), and comparative molecular field analysis and similarity indices (CoMFA & CoMSIA). These sophisticated approaches have enabled the generation of highly accurate predictive models that meticulously describe the structure-activity relationships of the screened ligands. Delving further, we have implemented advanced biomolecular simulations utilizing semi-flexible and/or flexible Molecular Docking, MolCAD, MM-GBSA, and Molecular Dynamics techniques. These simulations have provided comprehensive insights into the stability and compatibility of protein-ligand macromolecular systems, shedding light on the intricate interactions between these entities. Additionally, we have rigorously applied virtual screening models of Drug-like and ADME-Tox types to meticulously evaluate the bioavailability, pharmacokinetic, and pharmacodynamic aspects of candidate ligands, ensuring their suitability for further development. To demonstrate the efficacy of our proposed CADD hybrid approach, we have targeted some of the most devastating diseases, including cancer (lung, breast, lymphoma) and viral infections (SARS-CoV-2). In the realm of targeted cancer therapy, we have successfully identified and optimized new generations of non-covalent inhibitors specifically targeting the c-Met tyrosine kinase receptor, utilizing synthetic derivatives of Dimedone, 4-phenoxypyridine, and Cyclohexane-1,3-dione. Furthermore, we have developed novel non-covalent inhibitors targeting the MALT1 protease based on Pyridine derivatives. Additionally, we have identified and optimized new non-covalent inhibitors specifically targeting the EGFR tyrosine kinase receptor, harnessing the power of phytochemicals derived from Magnolia officinalis and Cannabis sativa L. In the context of targeted antiviral therapy specifically related to the SARS-CoV-2 (COVID-19) pandemic, we have optimized a new generation of non-covalent inhibitors specifically targeting the $3CL^{pro}$ enzyme, utilizing 9,10-Dihydrophenanthrene derivatives. The resounding success of our endeavors unequivocally demonstrates the remarkable effectiveness and rationality of the CADD hybrid approach adopted in this thesis. This approach has proven to be an invaluable tool for virtual screening and *in silico* optimization of bioactive candidate ligands, paving the way for early-stage drug design and development.

Keywords: CADD, *in silico*, 2D/3D-QSAR, DFT, ANN, MolCAD, MM-GBSA, Molecular Docking, Molecular Dynamics, Drug-like/ADME-Tox, Non-covalent inhibitors, Anticancer, Antiviral.

الملخص

يظل البحث الدؤوب عن العوامل العلاجية المبتكرة لمكافحة الأمراض المُنهكة ركيزة أساسية لصناعة الأدوية. على الرغم من الاستثمارات الجسيمة، والفترات الطويلة لتطوير الأدوية، ومعدلات الفشل العالية، إلا أن البحث عن أدوية فعالة لا يزال مستمراً. في هذا السياق، تقدم هذه الأطروحة نهجًا مبتكرًا يستغل قوة تصميم الأدوية بمساعدة الحاسوب لإحداث ثورة في عملية اكتشاف الأدوية، مما يقلل من أخطار الفشل ويُسرّع تطوير العلاجات المنقذة للحياة. في قلب هذه الأطروحة تكمن عملية تحديد واختبار ونمذجة دقيقة لخصائص الروابط الحيوية العلاجية (ligands)، مما يُحسّن من آفاقها كأدوية مُحتملة. لتحقيق ذلك، استخدمنا تقنيات متطورة لنمذجة الأحياء الإحصائية والمعلوماتية الحيوية، بما في ذلك 2D/3D-QSAR، وشبكات العصبونية الاصطناعية (ANN)، ونظرية الكثافة الوظيفية (DFT)، وتحليل المجال الجزيئي المقارن ومؤشرات التشابه (CoMFA & CoMSIA). لقد أتاحت لنا هذه الأساليب المُتطورة إنشاء نماذج تنبؤية عالية الدقة تصف بدقة علاقات البنية-الفعالية للربيطة المُفحوصة. علاوة على ذلك، قمنا بتنفيذ محاكاة جزيئية متقدمة باستخدام تقنيات الإرساء الجزيئي (Molecular Docking) شبه المرن و/أو المرن، و MolCAD، و MM-GBSA، والديناميات الجزيئة. لقد قدمت لنا هذه المحاكاة معلومات واسعة النطاق حول أنماط استقرار وتوافق أنظمة الماكروموليكولات من نوع بروتين-ربيطة (Protein-Ligand)، وسلّطت الضوء على التفاعلات المعقدة بين هذه الكيانات. بالإضافة إلى ذلك، قمنا بتطبيق نماذج الفرز الافتراضي الصارمة من نوع "Drug-like" و "ADME-Tox" لتقييم جوانب التوافر البيولوجي عن طريق الفم، والحراك الدوائي، والديناميكي للروابط الحيوية المرشحة للتصميم الدوائي، ذلك لضمان ملاءمتها لمزيد من التطوير. لتوضيح كفاءة نهجنا الهجين ل CADD المقترح، استهدفنا بعض أكثر الأمراض فتكًا، بما في ذلك السرطان (الرئة، الثدي، الليمفوما) والعدوى الفيروسية (SARS-CoV-2). في مجال العلاج المُستهدف للسرطان، قمنا بنجاح بتحديد وتحسين أجيال جديدة من مُثبّطات غير تساهمية تستهدف خصيصًا مستقبل التيروزين كيناز من نوع c-Met، ذلك باستخدام مشتقات اصطناعية من Dimedone، و4-Phenoxypyridine، و Cyclohexane-1,3-dione. بالإضافة إلى ذلك، قمنا بتطوير مُثبّطات جديدة غير تساهمية تستهدف بروتياز MALT1 بناءً على مشتقات Pyridine. علاوة على ذلك، قمنا بتحديد وتحسين مُثبّطات جديدة غير تساهمية تستهدف خصيصًا مستقبل التيروزين كيناز من نوع EGFR، مستغلين الفعالية البيولوجية للمواد الكيميائية النباتية المُشتقة من *Magnolia officinalis* و *Cannabis sativa L*. في سياق العلاج المُستهدف المضاد للفيروسات المرتبط بشكل خاص بجائحة SARS-CoV-2 (COVID-19)، قمنا بتحسين جيل جديد من مُثبّطات غير تساهمية تستهدف خصيصًا إنزيم $3CL^{pro}$ باستخدام مشتقات Dihydrophenanthrene. يُظهر النجاح الباهر لمساعينا بشكل قاطع النجاعة والمنطقية الملحوظة لنهج CADD الهجين المُعتمد في هذه الأطروحة. لقد أثبتت هذه الطريقة أنها أداة قيّمة للفرز الافتراضي في السيليكو (*in silico*) وللتحسين العقلاني للروابط الحيوية (ligands) المرشحة لتصميم الدوائي، مما يمهد الطريق لتطوير الأدوية واكتشافها في مرحلة مبكرة.

الكلمات المفتاحية: CADD، *in silico*، 2D/3D-QSAR، DFT، ANN، MolCAD، MM-GBSA، الإرساء الجزيئي، الديناميكيات الجزيئية، Drug-like/ADME-Tox، المثبطات غير التساهمية، مضادات السرطان، مضادات الفيروسات.

INTRODUCTION GÉNÉRALE

Le contexte général des travaux présentés dans ce rapport de thèse s'inscrit dans le domaine de la modélisation et de la simulation biomoléculaires basées sur des approches computationnelles rationnelles pour prévoir et concevoir de nouveaux médicaments potentiels. Les contributions proposées dans cette thèse abordent de manière générale les principaux défis et difficultés auxquels est confrontée la recherche pharmaceutique. Parmi ces défis, l'un des plus importants est principalement d'ordre scientifique, mais aussi financier et stratégique. Dans une tentative de couvrir le plus large possible ces défis, nous avons décidé dans cette thèse de renforcer nos efforts pour élaborer des méthodologies de conception et de découverte rationnelle de nouveaux médicaments par ordinateur, qui simulent le processus traditionnel de conception de médicaments dépendant entièrement de la démarche à caractère expérimental concrète. Cela est susceptible de contribuer aux gains de délais de développement et d'investissements financiers, ainsi qu'à la limitation des échecs liés à la découverte et à la conception de nouveaux médicaments, depuis les stades préliminaires aux stades cliniques, et même au-delà. À cet égard, après avoir rationalisé les approches computationnelles et les techniques de modélisation biomoléculaire susceptibles d'être exploitées pour construire des méthodologies hybrides de conception de médicaments assistée par ordinateur ("Computer-aided drug design", CADD), nous avons appliqué ces démarches pour mener diverses études en phase d'expérimentation virtuelle *in silico*. Le cancer (poumon, sein, lymphome) et les infections virales (SARS-CoV-2) constituent les pathologies fatales les plus fréquentes que nous avons considérées comme des modèles pour la mise en œuvre des méthodologies hybrides de CADD proposées dans le cadre de cette thèse. Dans cette introduction, tout d'abord nous présentons le cadre par lequel nous nous situons, et pas à pas nous balayons grossièrement les circonstances des problèmes qui ont nécessité l'intervention de stratégies CADD hybride apte à couvrir une part intéressante de ces problèmes. Enfin, nous présentons le plan général du contenu de cette thèse.

1 Contexte, défis majeurs et pistes de solution

Le développement de nouveaux médicaments représente un défi majeur pour la recherche pharmaceutique, en raison du temps, du risque élevé et des coûts importants impliqués. Malgré l'introduction de stratégies expérimentales innovantes, l'industrie pharmaceutique continue de faire face à un taux d'échec élevé, estimé à 90 % tout au long du processus de développement. Pour améliorer ces résultats et minimiser les risques, il est crucial d'adopter des approches innovantes telles que la CADD. Cette thèse explore l'utilisation de techniques de CADD basées sur le criblage biochimique et l'intelligence artificielle pour cibler les propriétés pharmacologiques, augmentant ainsi les chances de succès des médicaments jusqu'à leur phase avancée. En utilisant des petites molécules organiques issues de chimiothèques de qualité, l'approche de criblage contribue à identifier des candidats médicamenteux plus prometteurs. Les retombées de cette approche enrichissent les connaissances en pharmacologie et le développement de nouveaux médicaments. La CADD, qui comprend la compréhension et la prédiction des propriétés 3D des cibles macromoléculaires, a toujours

été un pilier de la découverte et du développement de médicaments. Ces techniques offrent une vision précise des processus biologiques, des interactions moléculaires et de l'affinité des molécules, stimulant ainsi l'innovation en recherche médicale. Malgré leur potentiel, il n'existe pas encore de stratégie unifiée pour intégrer efficacement les techniques informatiques de CADD dans le complexe processus de développement de médicaments. L'un des objectifs de cette thèse est donc de proposer une stratégie hybride de CADD, axée sur la conception de médicaments candidats à partir de ligands bioactifs. Cette approche intègre diverses techniques de conception biomoléculaire computationnelle et d'analyses biophysiques, permettant une conception, une modélisation et une optimisation précoces des médicaments. La stratégie hybride inclut des méthodes statistiques et physico-chimiques/mécaniques telles que les QSAR multidimensionnels, le docking moléculaire hybride, les simulations moléculaires, ainsi que des modèles Drug-Like et ADME/Tox. L'intégration de ces techniques dans une approche de découverte de médicaments virtuelle unifiée est cruciale pour prédire les propriétés pharmacocinétiques et pharmacodynamiques, ainsi que les mécanismes d'action des médicaments en développement. Cette méthodologie vise à rendre la conception de médicaments basée sur les modèles CADD proposés dans cette thèse plus accessible, fiable et rapide, tout en contribuant positivement au cycle de développement des médicaments. L'accent est mis sur l'exploitation de la diversité structurelle et des activités biologiques des molécules pour évaluer leurs interactions avec les récepteurs ciblés, fournissant ainsi une base solide pour la découverte de nouveaux médicaments thérapeutiques.

2 Approches proposées

Afin d'avancer pas à pas dans la mise en œuvre de l'approche de CADD hybride proposée dans ce travail, nous avons pris l'initiative d'apporter les aspects suivants :

- Personnalisation des stratégies hybrides de CADD développées dans ce travail pour les adapter aux conditions réelles et les assister.

- Pour rendre les stratégies hybrides de CADD de ce travail plus souples et plus rationnelles lors de leur utilisation pour le criblage des bases de données moléculaires, qu'il s'agisse ou non de l'association avec leur activité biologique testée *in vitro* ou *in vivo*, nous développons des modèles de criblage virtuel *in silico*. Ces modèles se fondent sur les conformations bioactives du ligand par rapport à une pathologie spécifique et/ou sur la structure 3D du récepteur biologique impliqué dans cette pathologie. De plus, dans toutes nos études, nous avons inclus au moins un médicament standard déjà approuvé et commercialisé pour évaluer la qualité des résultats générés *in silico* par nos modèles CADD.

- Décrypter la relation structure-activité des bases de données chimiothèques étudiées ; traduire cette relation sur sa dimension quantitative qui permet de passer au niveau du graphe à la représentation codifiée sous la forme de descripteurs moléculaires à plusieurs dimensions (de 0D à 4D). À cet égard, nous exploitons les méthodes computationnelles avancées liées à la mécanique moléculaire et à la mécanique quantique pour générer ces descripteurs moléculaires. Parallèlement, nous exploitons des méthodes de modélisation statistique, d'apprentissage automatique et

d'intelligence artificielle pour modéliser et optimiser la relation structure-activité des molécules considérées, conformément à des descripteurs moléculaires hybrides (0D-3D et/ou 4D). Nous faisons tout cela dans le cadre de l'approche de la conception de médicaments basée sur le ligand (ligand-based drug design, LBDD) à l'aide des méthodes 2D-QSAR et/ou 3D-QSAR.

- Les modèles QSAR multidimensionnels (mD-QSAR) sont soumis à des tests de validation interne et externe plus rigoureux et approfondis, conformément aux principes de l'OCDE (Organisation de coopération et de développement économiques) pour la validation et l'acceptabilité de ces modèles. L'utilisation de ces modèles dans le criblage et la sélection de molécules candidates pour la conception de nouveaux médicaments, ainsi que leur caractérisation rationnelle en vue de créer de nouvelles têtes de série tout en prédisant leurs activités biologiques *in silico*, renforce la perspective d'un développement rapide et efficace de nombreuses molécules thérapeutiques candidates, économiquement avantageuses, dans le domaine de la santé publique et des pathologies complexes.

- Nous complétons les sorties des modèles mD-QSAR en caractérisant les propriétés favorables à l'optimisation de la relation structure-activité du ligand. Nous proposons des visualisations détaillées de la structure 3D du récepteur biologique ciblé pour la thérapie moléculaire, utilisant des cartes de contour 3D pour représenter les interactions potentielles entre le ligand et le récepteur. Ces visualisations, basées sur les simulations du potentiel électrostatique moléculaire (PEM) générées par la méthode de la mécanique quantique, ainsi que sur les simulations de Docking moléculaire, permettent de sélectionner les sites pharmacophores adaptés à la stratégie thérapeutique. Cette approche évite une conception aléatoire des molécules en tenant compte des critères spécifiques du site et des caractéristiques souhaitées.

- Dans cette thèse, nous sélectionnons les groupements fonctionnels pour optimiser les relations structure-activité et structure-compatibilité selon deux approches. La première utilise l'intuition chimique humaine, guidée par les modèles QSAR développés. La seconde fait appel à des outils d'intelligence artificielle basés sur les réseaux neuronaux pour décoder les formes, offrant ainsi une approche automatisée.

- En utilisant une approche combinant des données *in vitro*/*in vivo* avec des optimisations *in silico*, nous identifions les molécules les plus actives et fiables sur le plan biologique *in silico*. Cette sélection rigoureuse est basée sur des profils de biodisponibilité ainsi que des propriétés pharmacocinétiques et pharmacodynamiques similaires à celles de médicaments existants. Notre objectif est d'identifier des molécules H2L candidates présentant une activité biologique élevée, une sécurité optimale et de fortes chances de succès dans le processus de conception et de découverte de nouveaux médicaments avancés. Pour cela, nous avons utilisé des modèles d'apprentissage automatique et d'intelligence artificielle basés sur les structures moléculaires SMILES pour prédire ces propriétés pharmacocinétiques et pharmacodynamiques. Dans ce cadre, nos études ont abouti à la création de modèles de criblage robustes et diversifiés basés sur la modélisation Drug-like, ADME-Tox et/ou le modèle de risque de toxicité OSIRIS.

- Après avoir identifié les molécules candidates H2L *in silico*, nous évaluons leur comportement structurel et leur compatibilité/affinité avec le site actif du récepteur biologique ciblé. Pour cela, nous estimons divers paramètres biophysiques à l'aide de méthodes bio-informatiques, tels que l'énergie de liaison, les constantes d'inhibition nécessaires pour inhiber 50 % de l'activité enzymatique, ainsi que les profils des interactions récepteur-ligand. Ces analyses sont réalisées via des simulations de Docking moléculaire hybrides (semi-flexibles et/ou flexibles). Ces mesures visent à filtrer les ligands candidats pour optimiser les têtes de série dans le modèle virtuel, créant ainsi une base de données robuste pour la découverte de nouveaux agents moléculaires. Cela ouvre la voie au développement de nouvelles générations d'inhibiteurs enzymatiques dans le cadre de la conception de médicaments innovants.

- L'approche du Docking moléculaire est essentielle dans la conception de médicaments, mais elle peut avoir des limitations telles que la prédiction de poses instables ou incorrectes du ligand par rapport au site actif. Pour surmonter ces défis, nous utilisons des simulations de mécanique moléculaire (MM) et de dynamique moléculaire (DM) pour purifier les poses de Docking. De plus, nous appliquons les calculs d'énergie de liaison libre avec le modèle MM-GBSA pour valider *in silico* les résultats de notre stratégie CADD hybride, afin d'éviter les faux positifs et négatifs associés aux fonctions de score du Docking.

3 Cadre spécifique des applications

Les biomolécules telles que les protéines et les petites molécules sont d'une grande importance pour le contrôle des rôles biologiques, fonctionnels et structurels d'un organisme. Par conséquent, les petites molécules bioactives peuvent jouer un rôle crucial dans l'inhibition et le contrôle de l'activité enzymatique anormale des protéines. Récemment, la recherche de nouvelles substances bioactives synthétiques et/ou phytochimiques pour leur potentiel d'inhibition des protéines et la recherche de nouvelles molécules bioactives à utiliser comme médicaments sont devenues essentielles dans la lutte contre les maladies incurables. À cette fin, dans cette thèse, nous soulignons nos intérêts pour les méthodes de modélisation biomoléculaire assistée par ordinateur dans une stratégie CADD hybride dans le cadre de la conception de nouveaux agents thérapeutiques potentiels pour des applications thérapeutiques. Dans cette optique, la mise en œuvre de différentes applications de la stratégie de CADD hybride proposée dans cette thèse nous a permis d'approcher la réponse à la problématique fondamentale suivante : Dans quelle large mesure les petites molécules sont-elles adaptées à la thérapie ciblée via le ciblage moléculaire de récepteurs biologiques ?

Afin de répondre à cette interrogation, dans cette thèse nous visons à étudier et à évaluer à l'échelle *in silico* le pouvoir et le mode d'action de petites molécules bioactives issues de différents échafaudages chimiques pour inhiber l'activité enzymatique anormale de plusieurs cibles biologiques (protéines, enzymes, protéases, etc.) impliqués dans le développement de lignées cellulaires pathologiques de nature cancéreuse et d'infections virales à titre d'exemples d'applications. Dans ce cadre, nous nous focalisons sur l'optimisation computationnelle de nouvelles classes d'inhibiteurs réversibles, non-covalents, spécifiques pour les récepteurs biologiques de c-Met (cysteine based

mesenchymal epithelial transition), EGFR-TKD (epidermal growth factor receptor-tyrosine kinase domaine), MALT1 (Mucosa-associated lymphoid tissue lymphoma translocation protein 1), et SARS-CoV-3CL^{pro} (severe acute respiratory syndrome coronavirus-2- 3-chymotrypsin like protease) impliqués dans le développement des différentes variétés de cancers et d'infections virales les plus graves, les plus mortelles, dont le cancer du poumon non à petites cellules, le cancer du sein, le lymphome, et le Covid-19, respectivement. La finalité consiste à favoriser la découverte précoce de nouveaux agents moléculaires, candidats à la conception de nouveaux médicaments efficaces, biologiquement très actifs, sûrs et faciles à mettre en place.

4 Plan du manuscrit

En plus d'une introduction générale et d'un résumé, le manuscrit de cette thèse comprend cinq chapitres organisés comme suit :

- Le premier chapitre de ce travail comprend une rapide introduction sur le contexte de ce travail et aborde les aspects liés aux objectifs souhaités. Ensuite, nous décrirons les grands concepts bibliographiques concernant les méthodes de traitement les plus répandues dans le monde du traitement des maladies, y compris la thérapie ciblée par médicaments à base de petits molécules bioactives (synthétiques et naturelles), qui consiste le point focal de notre intérêt dans cette thèse. Après cela, nous présentons les étapes les plus importantes par lesquelles passe la conception/découverte d'un nouveau médicament dans le pipeline de conception de médicaments basé sur des approches concrètes, et nous abordons les défis les plus importants auxquels il est confronté avec des approches de solutions suggérées. Parmi les solutions adoptées, les approches de conception de médicaments assistée par ordinateur (CADD) comportent diverses techniques, méthodologies et approches, à la fois traditionnelles et modernes, qui servent principalement à améliorer la rentabilité de chaque étape du processus de découverte et de conception de médicaments. De cette manière, nous exposons l'importance et le rôle important des approches de CADD, leurs succès, leurs classifications, leurs défis et les pistes de solutions proposées pour surmonter le plus possible ces défis. Après cela, nous illustrons la description des protéines/enzymes biomoléculaires ciblées en tant que modèles d'étude et d'application au cours de la préparation de cette thèse. Où nous avons concentré notre intérêt à contribuer à aborder la question du développement de nouvelles pistes thérapeutiques contre un échantillon de pathologies incurables, à savoir celles les plus fatales provoquées par le cancer et les infections virales (COVID-19) en tant que modèles.
- Le deuxième chapitre est consacré aux détails liés notamment aux méthodes et techniques de modélisation computationnelle, bio-informatique et biostatistique que nous avons adoptées dans le cadre de cette thèse en vue d'élaborer et d'appliquer avec succès des stratégies hybrides de CADD pour le criblage, la conception/l'optimisation et la sélection d'agents moléculaires candidats à la conception de nouveaux médicaments.
- Les chapitres III, IV et V sont dédiés aux applications *in silico* de l'approche CADD hybride destinée notamment pour étudier, évaluer et optimiser des bases de données des molécules biologiquement

actives, tout en respectant les critères de diversité structurelle et d'activité biologique, dans la perspective de développer de nouveaux candidats ligands bioactifs ayant des potentialités thérapeutiques. À cet égard, nous répartirons les résultats obtenus comme suit :

- Dans le troisième chapitre, nous présentons des contributions relatives à l'inhibition de la prolifération des cellules cancéreuses par le biais du ciblage moléculaire de la protéine c-Met par de petites molécules bioactives synthétisées sur la base du Dimédone, 4-phénoxypyridinine et de la Cyclohexane-1,3-dione.
- Dans le quatrième chapitre, nous présentons des contributions relatives à l'inhibition de la prolifération des cellules cancéreuses par le ciblage moléculaire de la protéine EGFR-TKD par des substances phytochimiques basées sur les plantes naturelles de *Magnolia officinalis* et *Cannbis sativa L*.
- Dans le cinquième et dernier chapitre, nous présentons des contributions relatives à l'inhibition de la prolifération des cellules cancéreuses et virales grâce au ciblage moléculaire des protéases MALT1 et $3CL^{pro}$ par de petites molécules bioactives synthétisées à partir de la pyridine et du 9,10-dihydrophénanthrène, respectivement. Enfin, nous terminons ce travail par une conclusion générale et des perspectives.

CHAPITRE I : APERÇU DU PROCESSUS DE DÉCOUVERTE DE NOUVEAUX MÉDICAMENTS, DE LA RECHERCHE À LA MISE SUR LE MARCHÉ

1. Introduction et problématiques

La bonne santé de l'être humain est l'un des premiers critères qui contribuent au bien-être psychique et social, ainsi qu'à la productivité économique des peuples. Cependant, garantir notre bonne santé tout au long de la vie est une chose surréaliste et difficile à contrôler. Cependant, il existe une tendance générale à l'amélioration, qui est principalement due à l'amélioration du milieu de vie et à la mise à disposition de médicaments efficaces. Les développements survenus durant les cinq dernières décennies ont amené à améliorer les conditions de soins de santé et à mettre à disposition les meilleurs médicaments afin de réduire le nombre de décès causés par diverses maladies. À cet égard, l'un des développements les plus impressionnants du XXI^{e} siècle est le déchiffrage du génome humain. Cette évolution est aujourd'hui liée à un énorme progrès concernant notre aptitude à simuler et à comprendre les cibles médicamenteuses associées à différentes maladies, par le biais de calculs en temps réel effectués à l'aide de superordinateurs et de services de nuage informatique (en anglais, cloud computing) [1], [2]. Cela a été confirmé, en avril 2020, par le projet cloud computing visant à comprendre les cibles médicamenteuses associées à COVID-19. Le développement actuel des performances technologiques était hors de portée des savants il y a quelques années, ce qui rendait le processus de découverte de médicaments très lent et coûteux. Par conséquent, dans le but de gagner du temps et de réduire le coût associé à la conception de nouveaux médicaments, l'industrie pharmaceutique s'est orientée vers le développement de nouvelles voies pour enrichir et rendre le processus traditionnel de découverte de médicaments plus satisfaisant, ceci grâce à la mise en place des techniques d'ingénierie biomoléculaire basées sur différentes disciplines de modélisation, y compris le calcul, la bio-informatique et la biostatistique, etc. Dans ce contexte, de nombreuses techniques de modélisation moléculaire sont développées et leurs applications dans le design moléculaire ont été mises en œuvre dans plusieurs secteurs industriels (pharmaceutique, photovoltaïque, agriculture, etc.). Les méthodes de modélisation moléculaire et de simulation biophysique appliquées à la conception de médicaments peuvent être classées dans la catégorie des approches de conception de médicaments assistée par ordinateur (en anglais, Computer Aided Drug Design), ce que l'on appelle l'approche CADD. Les approches de CADD se répartissent en deux grandes classes : (1) la conception de médicaments basée sur la structure (SBDD) et (2) la conception de médicaments basée sur le ligand (LBDD). Ces deux catégories ont des avantages et des limites, mais elles convergent toutes vers un but commun, à savoir fournir des solutions et des prédictions pharmacopées sur les échantillons biomoléculaires à chaque étape de la découverte de médicaments. Bien que les méthodes CADD soient largement appliquées à la conception de médicaments, il n'existe pas de méthodologie de modélisation rationnelle applicable partout. Par conséquent, la question de la mise au point et de la diversification des stratégies permettant de mener à bien ces méthodes reste

ouverte et nécessite davantage sur le plan méthodologique. En ce qui concerne l'aspect méthodologique, nous avons travaillé sur l'élaboration de stratégies de CADD facilement applicables et transparentes qui permettent de procéder rapidement au criblage rationnel (screening) de molécules candidates à la découverte de nouveaux médicaments. Parallèlement, nous travaillons sur l'application de ces méthodologies dans le cadre d'études préliminaires menées dans le cadre du laboratoire interne (In-house laboratory testing ou *in silico*) afin d'optimiser et de concevoir un large répertoire de nouveaux médicaments computationnels pouvant être proposés pour lutter contre de nombreuses pathologies incurables, en particulier celles liées à diverses formes de cancer et d'infections virales. Dans ce chapitre, nous commençons par un aperçu conceptuel sur les différents types et stratégies thérapeutiques appliquées dans le domaine des soins, à savoir : la chirurgie, la radiothérapie et la thérapie médicamenteuse. Ensuite, nous abordons, de manière plus exhaustive, les médicaments thérapeutiques basés sur des molécules bioactives dans le cadre de la thérapie ciblée, et nous décrivons les phases principales de la conception et du développement d'un nouveau médicament selon la voie classique ainsi que les limites et les défis de cette voie, et nous concluons par les contributions idéales et les progrès des méthodes CADD, Machine Learning (ML) et Intelligence Artificielle (IA) au développement rapide de nouveaux médicaments.

2 Les méthodes thérapeutiques et leurs particularités

En médecine, on appelle "traitement thérapeutique", ou simplement "traitement", l'ensemble des démarches mises en place par un professionnel de santé pour aider un patient à guérir d'une pathologie et à se soulager de ses symptômes. Ces traitements peuvent être classés en trois grandes catégories: la chirurgie, la radiothérapie et les médicaments [3].

2.1 Traitement chirurgical

La chirurgie est une forme de traitement local qui consiste à effectuer des procédures médicales sur un corps vivant utilisant des outils spécialisés [4]. Les techniques de chirurgie modernes sont variées et peuvent être utilisées pour corriger des traumatismes, traiter des infections, enlever des tumeurs, réduire les effets des hormones excessives, modifier des déformations et remplacer des organes par des transplantations.

2.2 Traitement radiologique

Parmi les applications les plus importantes de la radiothérapie, on peut citer :

- Pour traiter avant et après la chirurgie certains cancers avancés localement ; cette méthode est généralement appliquée en association avec la chimiothérapie.
- Également pour réduire les symptômes de certaines tumeurs très avancées, ayant subi ou non une intervention chirurgicale.
- Il est utilisé également pour traiter les douleurs qui ne sont pas soulagées par les médicaments antidouleur habituels ou pour contrôler les hémorragies localisées au niveau des tumeurs [5].

2.3 Traitements médicamenteux

En général, le terme chimiothérapie (chimio = chimique, thérapie = traitement) exprime l'ensemble des traitements médicamenteux [6]. Cependant, la chimiothérapie peut être subdivisée en 4 grandes catégories : Médicaments cytotoxiques, Hormonothérapie, Vaccins, et Thérapies ciblées [7].

2.3.1 Les médicaments cytotoxiques

Les médicaments cytotoxiques peuvent détruire les cellules qui sont en croissance ou en division. Ils sont principalement utilisés en chimiothérapie contre le cancer. De plus, ces médicaments sont aussi employés pour traiter certaines maladies cutanées telles que le psoriasis, le rhumatisme articulaire aigu et la polyarthrite rhumatoïde, ainsi que des problèmes musculaires résistants aux stéroïdes. Malheureusement, ces traitements peuvent causer des effets secondaires désagréables, tels que la fatigue, la perte de cheveux, des nausées et des vomissements graves [3].

2.3.2 La thérapeutique endocrine

Les traitements endocriniens consistent en des médicaments qui modifient la production et la fonction des hormones naturelles dans le corps [8]. L'utilisation de ce type de traitement est souvent destinée à combattre le cancer. Par exemple, le cancer de la prostate est causé par la testostérone et certains types de cancer du sein sont provoqués par l'œstrogène. Ces traitements endocriniens bloquent alors la production de ces hormones, ce qui les empêche d'alimenter la croissance des cellules cancéreuses.

2.3.3 Vaccins thérapeutiques

Les vaccins thérapeutiques sont conçus pour stimuler le système immunitaire afin de lutter contre les agents pathogènes [9]. La vaccination peut bloquer la croissance des cellules pathogènes et réduire les symptômes de certaines maladies. La plupart des personnes vaccinées sont protégées contre les maladies pour lesquelles elles ont été vaccinées. Cependant, certains individus ne peuvent pas être vaccinés en raison de problèmes de santé ou d'allergies aux composants des vaccins. Pour ces personnes, l'immunité collective est la meilleure protection. Plus le nombre de personnes vaccinées est élevé, plus le risque que des personnes non vaccinées soient exposées à des agents pathogènes est faible.

2.3.4 Les thérapies ciblées

Le but des thérapies ciblées est d'inhiber le mécanisme de croissance des cellules pathologiques, telles que les cellules cancéreuses et les lignées cellulaires de Sras-Cov-2, résultant de l'expression anormale ou de l'activité enzymatique de certaines récepteurs biologiques (biomarqueurs) [10], [11]. Ainsi, ces thérapies sont spécifiques et visent à cibler un récepteur bien défini, par exemple des protéines, des enzymes ou une voie de signalisation biologique, et sont différentes des chimiothérapies (médicaments cytotoxiques, thérapeutique endocrine et vaccins thérapeutiques). La plupart des enzymes sont des protéines qui sont générées par la traduction des informations de l'ADN à travers

les molécules d'ARN [12]. Les enzymes favorisent les réactions chimiques nécessaires à la bonne santé et à la vie de tous les individus [13]. Par exemple, ils remplissent une fonction essentielle dans le métabolisme (transformation des aliments en énergie), en accélérant les réactions chimiques dans les cellules grâce à la liaison entre le site actif d'une enzyme avec une autre substance appelée substrat (**Figure I.1**) [14].

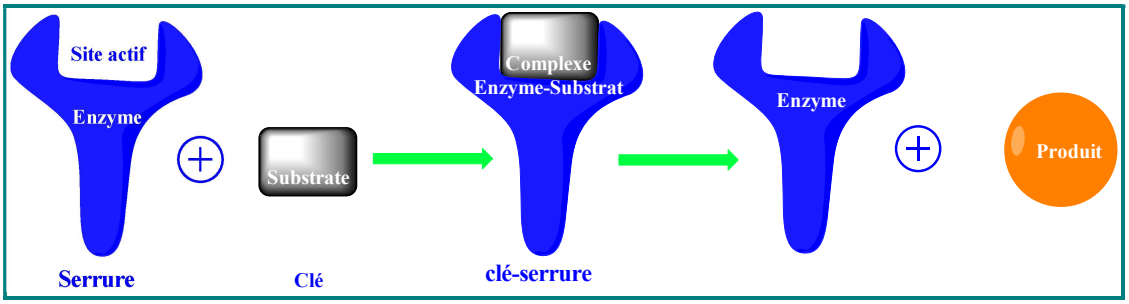


Figure I.1. Modèle " clé-serrure" du fonctionnement des enzymes.

Cependant, tout dysfonctionnement dans l'interaction "clé et serrure" peut entraîner un excès de produit ou ce que l'on appelle une hyperactivité enzymatique, ce qui ne permet pas à l'enzyme d'effectuer son action comme souhaité [15]. Cela peut se traduire par des symptômes que le système immunitaire a du mal à surmonter. Dans ce contexte, la thérapie ciblée permet de moduler l'action des enzymes et d'inhiber leur hyperactivité enzymatique (**Figure I.2**). Ce mécanisme consiste à cibler le site actif de l'enzyme en utilisant des agents bioactifs (appelés inhibiteurs) qui sont capables de se lier à l'enzyme, et donc d'inhiber son activité enzymatique [16], [17]. Par conséquent, la croissance des lignées cellulaires indésirables (par exemple, les tumeurs) est inhibée.

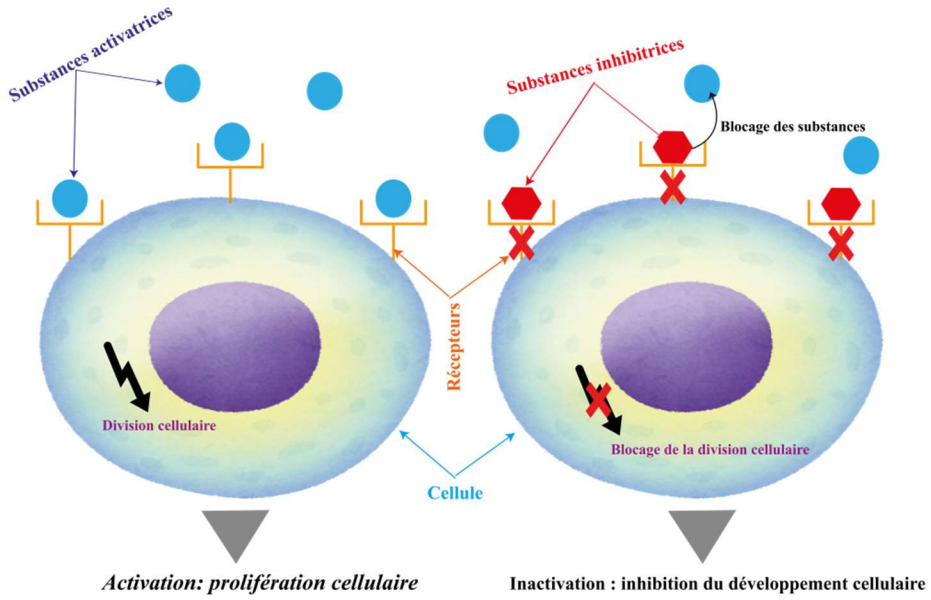


Figure I.2. Mécanisme d'inhibition de la croissance cellulaire pathologique par la thérapie ciblée.

La chirurgie, la radiothérapie et les traitements médicamenteux sont trois méthodes de base pour traiter de nombreuses maladies, dont le cancer est le modèle le plus marquant, et ces méthodes peuvent être utilisées seules ou en association pour éliminer les tumeurs [3]. Contrairement à la chirurgie et à la radiothérapie, les traitements médicamenteux basés sur les thérapies ciblées ont la capacité d'atteindre pratiquement toutes les régions du corps et d'agir systématiquement pour inhiber le développement et la propagation de cellules pathologiques [3]. Cela, même lorsqu'il est difficile pour

la radiologie ou la chirurgie de constater et de surveiller la propagation de cellules pathologiques. Néanmoins, certains problèmes sont associés aux thérapies ciblées. Par exemple, l'inhibition de certaines cibles peut ne pas être suffisante, étant donné le potentiel de mutagénicité des protéines ou des voies de signalisation cibles. Par conséquent, il peut se produire une résistance au médicament appliqué et l'incapacité de ce dernier à atteindre l'objectif thérapeutique souhaité. Bien que la thérapie ciblée pose certains problèmes, les principes fondamentaux de ce type de thérapie restent clairement applicables, c'est la raison pour laquelle nous devons l'utiliser en tant que point central pour la recherche et l'amélioration du protocole de conception de nouveaux médicaments destinés à la lutte contre les maladies réfractaires associées à des dysfonctionnements enzymatiques dans l'organisme. Par conséquent, pour assurer le bon fonctionnement des différents systèmes de l'organisme, il est parfois nécessaire de ralentir la fonction enzymatique. Par exemple, si une enzyme produit une quantité trop élevée d'un produit (lignées cellulaires), l'organisme a besoin d'un moyen de réduire ou d'arrêter la production. Les agents les plus importants qui peuvent jouer un rôle dans la limitation des niveaux d'activité enzymatique des protéines sont les inhibiteurs (inhibiteurs réversible ou irréversibles) [18], [19], [20]. Par conséquent, la nécessité de développer la recherche pour la découverte et l'amélioration d'inhibiteurs moléculaires plus stables et compatibles avec les protéines/enzymes cibles à des fins thérapeutiques est fortement recommandée. Cela nous a incités à focaliser notre intérêt dans cette thèse sur l'examen approfondi des caractéristiques des molécules biologiquement actives susceptibles de devenir des candidats inhibiteurs spécifiques et appropriés pour la thérapie ciblée.

3 Variantes des médicaments fondamentaux

Les médicaments à base de produits biologiques et de petites molécules représentent les principales catégories de médicaments développés pour lutter contre les maladies [4]. Les médicaments biologiques désignent les substances qui sont dérivées de cellules vivantes ou par des processus biologiques. Les médicaments biologiques peuvent consister en des substances génétiques, des protéines, des anticorps, des acides nucléiques, des cellules ou des tissus destinés à la transplantation [22]. Parmi les exemples de médicaments biologiques, il y a les hormones, les vaccins, les produits sanguins, les facteurs de croissance, l'insuline, etc. [23]. D'autre part, les médicaments qui peuvent être accessibles par voie chimique (synthèse combinatoire ou extraction à partir de substances naturelles telles que les plantes) sont appelés petites molécules [24]. Il s'agit par exemple de l'Aspirine, du Felbamate, de la Varénicline, de la Procaïne, de la Pénicilline, du Paracétamol, de l'Esoméprazole , etc. [25, p. 17]. Plus de 90% des médicaments disponibles actuellement sur le marché sont issus de petites molécules [21]. Par conséquent, la conception de médicaments basée sur les petites molécules est très demandée par rapport aux médicaments biologiques [26]. En termes techniques (**Tableau I.1**), les médicaments à base biologique et ceux à base de petites molécules présentent de nombreuses différences en termes de taille, de processus de mise à disposition, de caractérisation, de stabilité, de compatibilité immunologique, etc. [21].

Tableau I.1. Comparaison Approfondie : Produits Biologiques vs Petites Molécules [27].

Biologiques	Petites molécules
Poids moléculaire grand	Poids moléculaire petit
Obtenu par la culture de cellules vivantes	Obtenus par des processus chimiques
Structure hétérogène et complexe	Structure moléculaire bien définie
Fortement dépendant de l'opération	Principalement indépendant de l'opération
Absence de caractérisation exacte	Complètement caractérisable
Instable	Stable
Immunogène	Non immunogène

En termes de mode d'administration, étant donné la petite taille des petites molécules (20 à 100 atomes) qui assurent la stabilité typique de ces molécules, l'administration de ces médicaments se fait principalement par voie orale. Alors que la grande taille et la structure des médicaments biologiques (200 à 50 000 atomes) entraînent l'instabilité de ces molécules, ce qui conduit à leur décomposition dans le système digestif. Par conséquent, les médicaments biologiques sont souvent administrés par injection ou perfusion [27]. En termes de caractérisation structurelle, la complexité structurelle des médicaments biologiques rend la caractérisation difficile ; il est donc difficile de prévoir les effets cliniques et la réponse immunitaire des patients lors de la prise du médicament [27]. Par contre, les petites molécules possèdent des structures simples et caractérisables qui ne déclenchent pas de réponse immunitaire [27]. En général, les petites molécules présentent de nombreux avantages par rapport aux substances biologiques, nous citons par exemple la facilité d'administration des médicaments, les médicaments biologiques sont beaucoup plus chers à développer et donc beaucoup plus chers que les médicaments à base de petites molécules pour le consommateur. Par conséquent, il est fort probable que les médicaments à base de petites molécules resteront importants et utiles dans le développement et la découverte de médicaments thérapeutiques peu coûteux et plus efficaces. En dépit des bénéfices qui distinguent les petites molécules des substances biologiques, le processus de développement d'un nouveau médicament reste un projet complexe, risqué, coûteux et long qui peut prendre plus de 15 ans et plus d'un milliard de dollars US$ [28]. Mais, aujourd'hui, ce scénario a évolué avec l'émergence de nouvelles approches visant à améliorer l'efficacité du processus de découverte de médicaments, et les techniques computationnelles sont devenues un volet essentiel de la plupart des projets de découverte de médicaments. Les techniques de la CADD sont largement utilisées dans de nombreux efforts de découverte, depuis la recherche de résultats jusqu'à l'optimisation des candidats médicaments [29]. Par conséquent, la démarche visant à découvrir de nouvelles petites molécules adaptées comme médicaments à l'aide de méthodes CADD combinées à des méthodes d'exploration de données (en anglais, Data mining) basées sur les techniques d'apprentissage automatique et de l'intelligence artificielle peut contribuer fortement à améliorer davantage l'emplacement de ces molécules dans le pipeline de conception rationnelle de nouveaux médicaments [30]. Dans ce contexte, nos activités de recherche menées dans le cadre de cette thèse se concentrent sur l'élaboration de stratégies de CADD hybrides applicables, dans lesquelles les méthodes, approches et outils computationnels sophistiqués de modélisation biomoléculaire, bio-informatique et biostatistique, tels que les études de relations quantitatives structure-activité en 2D et 3D par les méthodes d'apprentissage automatique comme les régressions linéaire et non linéaire (RLM et RNLM),

régression par les moindres carrés partiels (PLS, en anglais « Partial least-squares »), réseau de neurones artificiels (ANN, en anglais «Artificial neural network »), analyse en composantes principales (ACP, en anglais « Principal component analysis », les algorithmes de classification k-moyennes (k-means clustering) et univariée (UV Clustering), etc ; des études pharmacocinétiques et pharmacodynamiques *in silico* (modélisation Drug-like et ADME-Tox); des modèles d'affinité et de stabilité des complexes Protéine-inhibiteur (Récepteur-ligand) géants simulés via des techniques de Docking moléculaire, de simulation de la mécanique moléculaire de la surface de Born généralisée (MM-GBSA), et de dynamique moléculaire, etc. Dans le cadre de cette thèse, l'intégration de ces méthodes mixtes dans une stratégie CADD structurée, compréhensive et rapidement applicable au processus précoce de découverte de nouveaux médicaments constitue l'une des principales bases pour le criblage, la conception, la prédiction et la modélisation de l'activité biologique inhibitrice de nouvelles petites molécules qui peuvent interagir avec la cible (protéine, enzyme) et produire leur effet thérapeutique recherché.

4 Inhibition enzymatique réversible et irréversible

4.1 Mode d'action et mécanismes de liaison

La principale différence entre l'inhibition enzymatique irréversible et réversible est le type d'interactions (liaison) entre l'enzyme et l'inhibiteur (**Figure I.3**) [31].

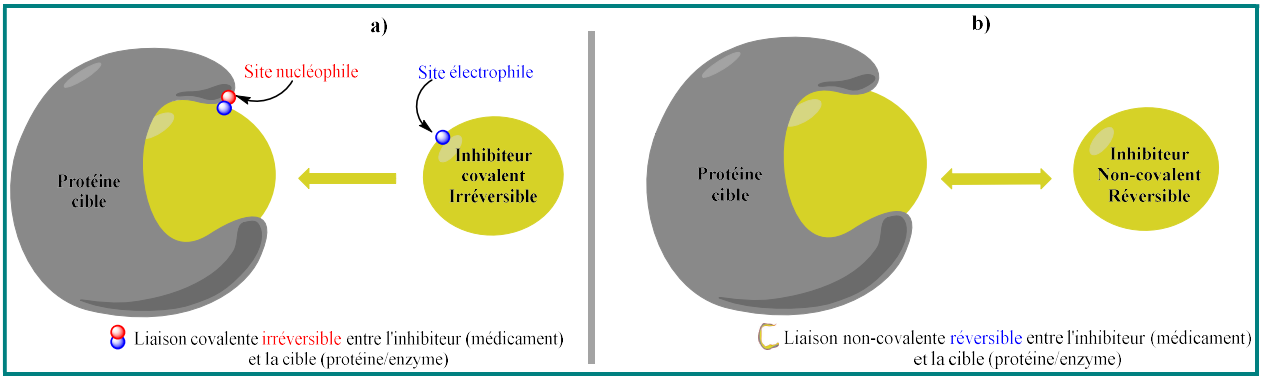


Figure I.3. Schéma illustrant les mécanismes d'interaction entre le récepteur biologique et l'inhibiteur : Covalent (a) et Non-Covalent (b)

L'inhibition enzymatique irréversible permet de bloquer (inhiber) l'activité enzymatique de la protéine cible par l'inhibition covalente du site actif grâce à la présence d'une liaison covalente forte entre l'enzyme et l'inhibiteur (**Figure I.3a**) [32]. Parmi les inhibiteurs covalents irréversibles approuvés par l'Agence fédérale américaine des produits alimentaires et médicamenteux (FDA, en anglais : «Food and Drug Administration») pour usage thérapeutique entre 1988 et 2021 il y a : Pénicilline, Aspirine, Fluorouracil, Oméprazole, Phosfomycine, Clopidogrel, Bortezomib, Boprevir, Carfilzomib, Ibrutinib, Afatinib, Osimertinib, Foxlotor, Suturasper, Nermatril [33], [34]. Les liaisons covalentes sont généralement formées par l'interaction entre les substances nucléophiles que sont la Cystéine, Sérine, Thréonine ou Lysine, avec des groupes fonctionnels réactifs (électrophiles) tels que les groupes hydroxyle, époxy ou carbonyle, ce qui entraîne la formation de liaisons covalentes entre l'inhibiteur et l'enzyme (**Figure I.3b**) [35]. Environ 30 % des médicaments commercialisés qui ciblent les enzymes

agissent comme des inhibiteurs covalents, mais malgré cela, l'incorporation de groupements réactifs n'était pas souhaitable dans la conception de médicaments jusqu'à récemment [36]. Cela est dû à l'activité biologique non identifiée/ non ciblée, qui est la raison majeure pour laquelle on évite d'incorporer des groupements réactifs dans les nouveaux médicaments candidats [37].

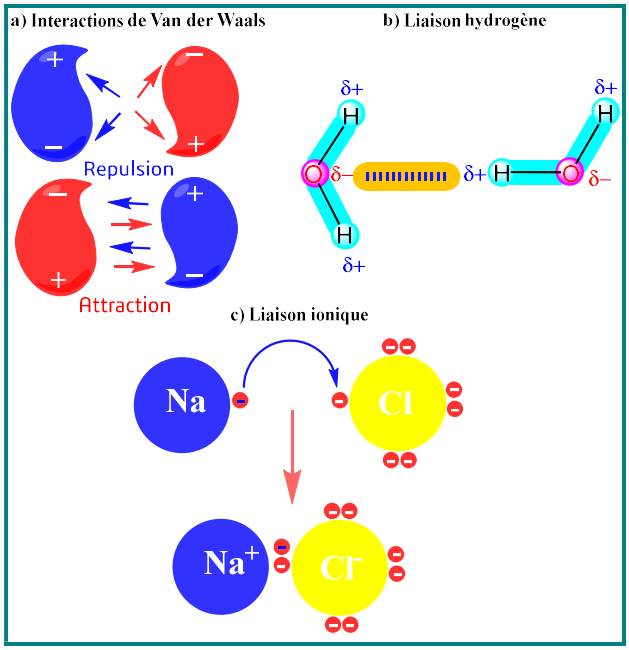


Figure I.4. Interactions non covalentes, (a) interactions de Van der Waals ; (b) liaison hydrogène ; (c) liaison ionique.

En revanche, l'inhibition réversible de l'enzyme est obtenue par le mécanisme d'inhibition non-covalente du site actif, en optimisant la forme et les interfaces non-covalentes (liaisons hydrogène, ponts salins, interactions électrostatiques ioniques, interactions de van der Waals, interactions hydrophobes, etc.) entre l'inhibiteur et les acides aminés actifs situés au site de liaison spécifique (**Figure I.4**). Il existe également une large gamme de médicaments approuvés à base de petites molécules qui agissent comme des inhibiteurs non-covalents réversibles, par exemple les inhibiteurs de type tyrosine kinase du récepteur du facteur de croissance épidermique (EGFR-TKI) tels que le Géfitinib [38], Erlotinib [39], Crizotinib [40], Foretinib [41], etc. Grâce aux interactions non-covalentes faibles entre l'inhibiteur et l'enzyme, l'enzyme est temporairement inhibée, ce qui permet la dissociation rapide du complexe enzyme-inhibiteur. Ainsi, la restauration de la fonction enzymatique et l'élimination de l'inhibiteur sont faciles après avoir atteint l'objectif thérapeutique [42]. Par contre, l'inhibition irréversible de l'enzyme entraîne une inhibition persistante de la fonction enzymatique, ce qui fait que la substance inhibitrice de l'enzyme se décompose très difficilement, ce qui peut entraîner de nombreux effets secondaires du fait que la fonction de l'enzyme n'est pas restaurée [43]. Pour ces raisons, les inhibiteurs réversibles non-covalents peuvent exhiber de meilleures propriétés pharmacocinétiques et pharmacodynamiques que celles observées pour les inhibiteurs irréversibles. Dans le cadre de ce travail, nous mettons l'accent sur l'évaluation des interactions non covalentes de petites molécules bioactives vis-à-vis les protéines ciblées. Ceci dans le but de modéliser et de sélectionner de nouvelles molécules candidates comme inhibiteurs réversibles, non covalents des récepteurs ciblés.

4.2 Caractérisation des inhibiteurs

Tout projet de découverte de médicaments réussi repose sur la caractérisation des inhibiteurs [44]. Compte tenu du vaste nombre de paramètres physico-chimiques et biologiques qui doivent être examinés, il convient de déterminer quels sont les paramètres les plus importants pour atteindre le meilleur équilibre entre efficacité et productivité [45]. La constante d'affinité/inhibition (K_i) et la concentration inhibitrice médiane (IC_{50}) sont généralement déterminées en laboratoire comme les principaux critères descriptifs des inhibiteurs adoptés dans les projets de découverte de médicaments. Le paramètre K_i correspond à la constante de dissociation qui décrit l'affinité de liaison entre l'inhibiteur et l'enzyme (la concentration libre de l'inhibiteur nécessaire pour atteindre 50% de saturation de l'enzyme), tandis que la valeur IC_{50} correspond à la concentration de l'inhibiteur nécessaire pour réduire l'activité enzymatique à la moitié de la valeur non inhibée (la concentration totale de l'inhibiteur nécessaire pour atteindre 50% d'inhibition) [46]. K_i est une mesure de l'affinité de l'inhibiteur pour l'enzyme cible, et que ce paramètre est indépendant de la concentration de l'enzyme (il peut dépendre de la concentration du substrat en raison des différents mécanismes d'inhibition) [45]. Alors que la valeur IC_{50} est basée sur la concentration d'enzyme pour décrire l'activité inhibitrice d'une substance spécifique. Les deux paramètres K_i et IC_{50} peuvent être utilisés comme indicateurs quantitatifs pour évaluer le pouvoir de l'inhibiteur. Pour un équilibre d'inhibition de 50%, les paramètres K_i et IC_{50} peuvent être corrélés selon la relation suivante : $T[I]=B[I]+L[I]$; où $T[I]$ est la concentration totale de l'inhibiteur, c'est-à-dire IC_{50}; $B[I]$ est la concentration de l'inhibiteur lié à l'enzyme dans le complexe enzyme-inhibiteur, qui est égale à 50% de la concentration totale de l'enzyme (c'est-à-dire $\frac{[E]}{2}$) ; et $L[I]$ est la concentration de l'inhibiteur libre, qui correspond à K_i. Par conséquent, $IC_{50}=\frac{[E]}{2}+K_i$ (IC_{50} se trouve systématiquement supérieure à K_i). Dans ce contexte, la méthode de Cheng-Prussoff définie par **Eq. (I.1)** permet de convertir les valeurs de IC_{50} en un K_i plus exhaustif pour effectuer la comparaison des données [47].

$$K_i=\frac{IC_{50}}{\left(1+\frac{S}{K_m}\right)} \qquad (I.1)$$

où, S est la concentration de substrat dans laquelle l'étude est réalisée et Km est la constante de Michaelis de substrat vis-à-vis de l'enzyme.

Sur la base des valeurs de K_i et IC_{50} obtenues *in vitro*, le processus d'optimisation des potentiels candidats inhibiteurs consiste à améliorer ces paramètres. Principalement, ces paramètres sont significatifs pour les inhibiteurs qui peuvent être rapidement réversibles (la liaison et la dissociation étant plus rapides que le temps de mesure de l'activité) [48]. À la différence des inhibiteurs covalents qui se lient lentement et n'atteignent pas l'équilibre avant la mesure de l'activité. Par conséquent, les inhibiteurs covalents sont ignorés si le temps de préincubation est trop court afin d'atteindre l'équilibre [45]. En raison de la grande importance de ces paramètres dans la caractérisation et l'amélioration de l'activité inhibitrice des molécules (biologiques, volumineuses et petites) désignées comme

médicaments et dans la détermination de leurs sélectivités et caractères, la recherche doit être approfondie pour développer de nouvelles méthodes alternatives facilitant l'évaluation de ces paramètres. Dans ce contexte, les méthodes de modélisation moléculaire assistée par ordinateur et de simulation biophysique peuvent être utilisées pour prédire ces paramètres avant de procéder à la synthèse et au test de l'activité biologique des molécules candidates aux médicaments. À cet égard, nous proposons dans ce projet de travail des stratégies hybrides basées sur l'intégration de techniques de CADD avec ML et IA pour améliorer, prédire et cribler l'activité biologique inhibitrice (IC_{50}), l'affinité protéine-inhibiteur, la biodisponibilité (Drug-like), les propriétés pharmacocinétiques et pharmacodynamiques (ADME-Tox) de petites molécules candidates en tant qu'inhibiteurs spécifiques non-covalents prometteuses pour le ciblage thérapeutique.

5 Découverte de médicaments, de la recherche à la mise en œuvre

La recherche scientifique repose sur la compréhension du mécanisme d'action des phénomènes naturels dans les objets non vivants et dans les organismes vivants, ainsi que sur l'utilisation des connaissances qui en résultent pour améliorer et développer la vie des organismes vivants, ainsi que pour préserver les objets non vivants. La recherche scientifique peut être classée en deux volets. La recherche fondamentale, qui vise à comprendre la nature des choses et à découvrir le mécanisme de leur fonctionnement et leur relation avec d'autres objets. Et la recherche appliquée, qui vise à appliquer les résultats de la recherche fondamentale et à les adapter à la réalité productive et commerciale. La recherche fondamentale est donc la clé de voûte des découvertes scientifiques et de produits de toutes sortes en médecine et en ingénierie. Chaque produit fait l'objet de nombreuses recherches. L'un des principaux produits de recherche pour l'homme est le médicament.

5.1 Évolution de la découverte de médicaments

La chronologie de la découverte de médicaments à notre époque peut être répartie en trois grandes périodes (**Figure I.5**). La première période notable remonte au XIX^{e} siècle, lorsque la base de la découverte de médicaments reposait sur le hasard des chimistes médicinaux. La deuxième période a commencé à partir du début du XX^{e} siècle, lorsque de nouvelles compositions médicamenteuses ont été découvertes, contribuant à une nouvelle ère de découverte et de développement des antibiotiques. Au cours de cette période, en fonction des structures moléculaires identifiées, et grâce au développement de nouvelles méthodes de simulation assistée par ordinateur telles que la modélisation moléculaire et le criblage virtuel à haut débit, des progrès rapides ont été réalisés dans la découverte de médicaments à la fin du XX^{e} siècle. Une révolution scientifique majeure a également été provoquée à cette époque par la naissance de la technologie de l'ADN, grâce à laquelle il est devenu possible de développer des médicaments potentiels en thérapie ciblée. Le XXI^{e} siècle a marqué le début de la troisième période, marquée par l'émergence de la biotechnologie "Omics" [49], qui a commencé par une accélération du processus de découverte de médicaments biopharmaceutiques approuvés par la FDA pour un usage thérapeutique. L'ensemble des biotechnologies " Omics" vise à analyser la forme, la fonction et la structure complète d'une fonction biologique spécifique à différents niveaux, notamment l'étude du génome, le niveau d'expression des gènes, le niveau des protéines des cellules ainsi que le

niveau du métabolisme. Ces techniques sont utilisées pour comprendre les systèmes biologiques complexes et pour révéler les causes moléculaires qui motivent les phénotypes cellulaires complexes qui apparaissent, par exemple, sur les cellules cancéreuses [50].

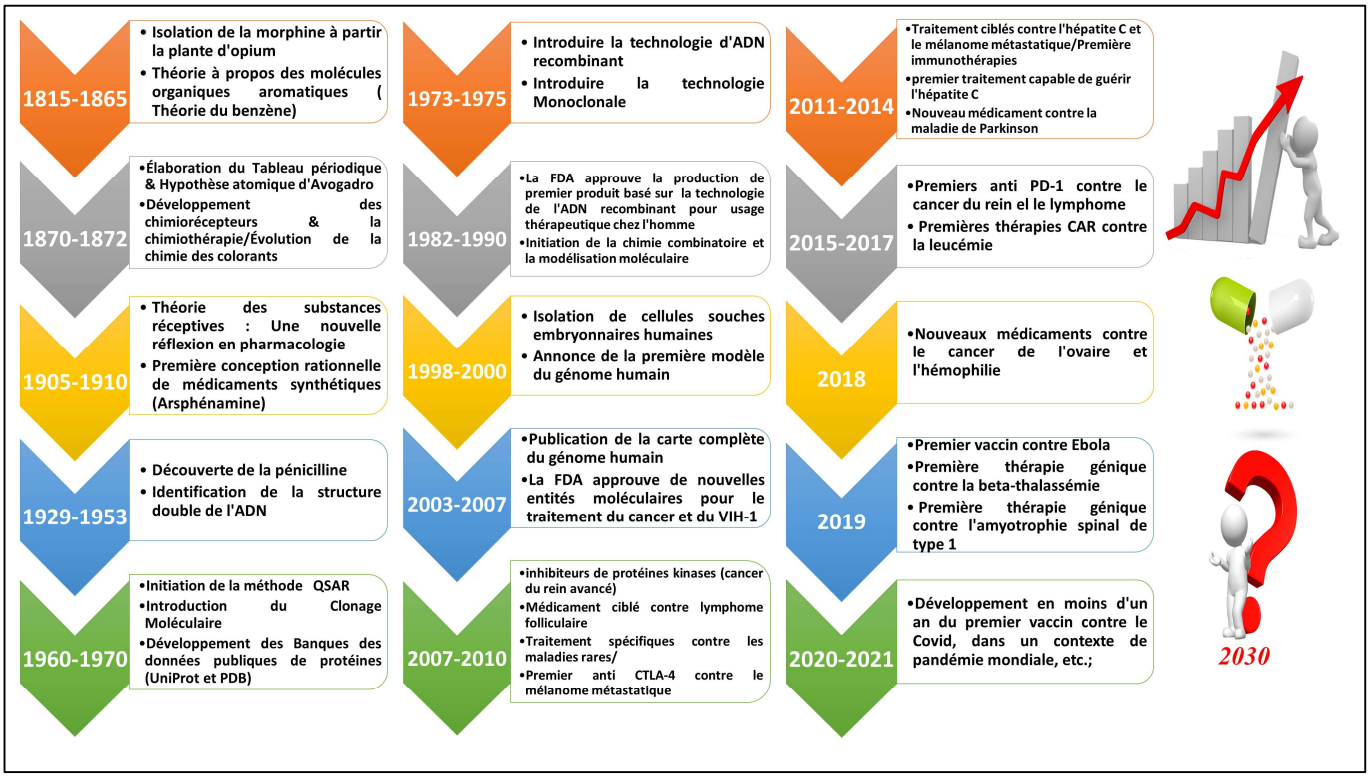


Figure I.5. Évolution des principaux événements de la découverte de médicaments.

5.2 Processus de découverte de médicaments

La découverte de médicaments implique de nombreuses étapes et processus différents, de la conception à la commercialisation du médicament en passant par son développement. Dans le cadre de cette description, nous présentons brièvement le processus de découverte de médicaments, à savoir son objet, ses quatre principales étapes, les premières étapes de la découverte de médicaments, les phases précliniques et cliniques (**Figure I.6**).

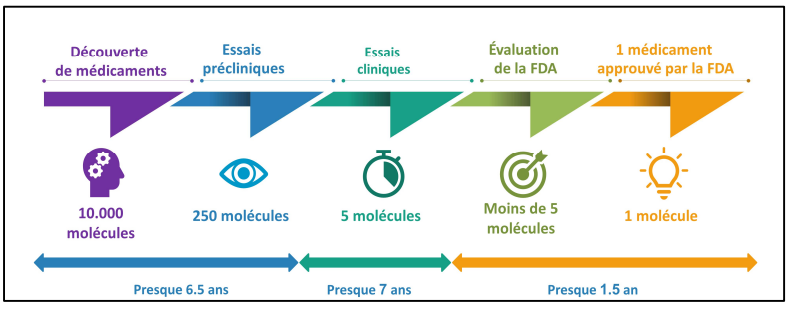


Figure I.6. La chronologie de la découverte d'un seul médicament.

Le processus de découverte de médicaments est un processus très long et coûteux qui peut prendre jusqu'à 15 ans. Le processus précoce de découverte de médicaments commence généralement par une évaluation des composés ayant une activité biologique potentielle et un effet thérapeutique sur la maladie cible. Après avoir identifié les molécules actives, elles sont soumises à des tests de sécurité et d'efficacité. En général, parmi les 5 000 à 10 000 candidats médicaments, seuls

250 passent avec succès les tests précliniques. Le passage de la découverte à la commercialisation d'un médicament est un processus très coûteux. Selon une étude menée par le Tufts Center le centre Tufts pour l'étude du développement des médicaments (Tufts Center for the Study of Drug Development), le coût du parcours de découverte d'un nouveau médicament est estimé à environ 2.6 à 3 milliards de dollars US$ pour chaque nouveau médicament [51].

5.3 Découverte précoce de médicaments

Cette phase inclut plusieurs procédures et tests de nature différente. Les collaborateurs renforcent leur coopération afin de définir de nouvelles pistes de réflexion sur la maladie cible, dans le but d'identifier et d'améliorer un médicament candidat permettant de traiter ou d'inverser les effets de la maladie. En gros, les molécules tête de série (leaders) doivent avoir un effet désirable sur la cible biologique impliquée dans la maladie, dans l'espoir de la traiter. À ce stade, la recherche est menée en laboratoire à l'aide de plateformes *in silico*, d'essais biochimiques, de cultures cellulaires et de divers modèles animaux. Les sous-processus avancés les plus importants dont découle la phase de découverte précoce de médicaments peuvent être classés en 7 étapes : Identification et validation de la cible, criblage à haut débit ou criblage à haut contenu, identification des hits, Développement d'essais et criblage, Hit-To-Lead (H2L), Génération et optimisation des leads, et tests *in vivo* et *in vitro* (**Figure I.7**). Toutes ces activités nécessitent des systèmes de traitement des données avancés, l'utilisation de robots et de dispositifs de manipulation des fluides, ainsi que des installations coûteuses dotées des technologies de pointe nécessaires.

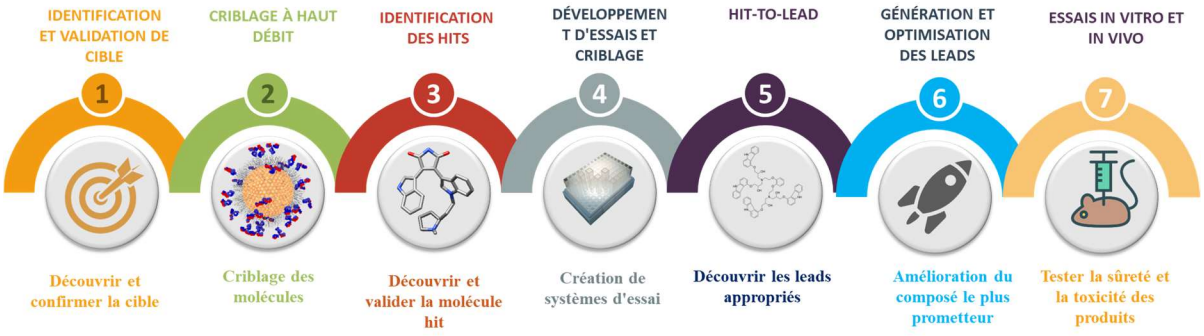


Figure I.7. Exploration précoce de médicaments.

5.3.1 Identification et validation de cible

Une compréhension correcte et précise des origines de la maladie est la clé principale de la conception d'un médicament. Une cible biologique se prête à un ciblage thérapeutique lorsqu'une molécule "hit" a la capacité de modifier son activité biologique. Les cibles potentielles peuvent être trouvées grâce aux bases de données disponibles et aux publications.

5.3.2 Criblage à haut débit

Après avoir identifié et validé la cible potentielle, le criblage à haut débit (High Throughput Screening, HTS) et le criblage à haut contenu (High Content Screening, HCS) de la base de données

de molécules sont lancés pour atteindre la cible thérapeutique identifiée. Ceci a pour but de fournir des données sur l'activité et la sélectivité de chaque molécule vis-à-vis de la cible.

5.3.3 Identification des Hits

Cette étape est basée sur les résultats du criblage à haut débit, les candidats molécules qui interagissent avec la cible et provoquent l'effet thérapeutique désiré sont retenus. Ces molécules candidates sont appelées "Hits".

5.3.4 Développement d'essais et criblage

Après l'identification des molécules les plus prometteuses pour la mise au point thérapeutique, des essais *in vitro* sont réalisés au préalable, pour déterminer principalement les paramètres pharmacocinétiques & pharmacodynamiques ADME-Tox (absorption, distribution, métabolisme et excrétion) et évaluer la toxicité en phase préliminaire. ADME-Tox est un concept clé en pharmacologie et en développement de médicaments. Il englobe les processus impliqués dans la répartition d'un composé pharmaceutique à l'intérieur d'un organisme (**Figure I.8**). Ces processus ont un impact significatif sur les niveaux et la cinétique de l'exposition du médicament aux tissus, ce qui influe sur son efficacité en tant que médicament. Les études ADME-Tox cherchent à comprendre comment un composé chimique, notamment un médicament, est traité par un organisme vivant. Les tests de toxicologie sont souvent intégrés à ces études, donnant ainsi l'acronyme ADME [52]. Lors du développement de médicaments, il est essentiel que les chercheurs évaluent la sécurité et la toxicité potentielle des médicaments dans le corps. Ils réalisent donc des études sur le métabolisme, la pharmacocinétique et la toxicologie des médicaments, comme les études ADME-Tox et les tests de toxicité, afin d'examiner en détail l'activité des médicaments dans l'organisme.

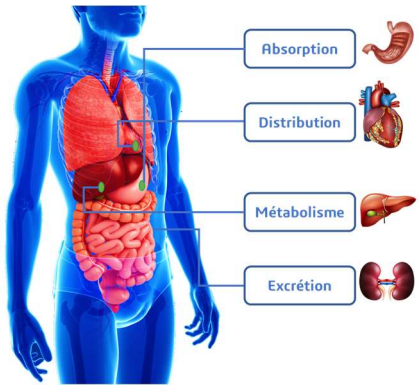


Figure I.8. Illustration schématique du rôle de certaines organes du corps humain dans le processus de l'ADME des médicaments.

Voici une brève description de chaque terme dans ADME-Tox [52]:

- Absorption : L'absorption désigne le mécanisme par lequel un médicament passe du point d'administration à la circulation sanguine.
- Distribution : La distribution désigne le transport d'un médicament dans l'organisme jusqu'à sa cible ou son site d'action prévu.

- Métabolisme : Le métabolisme implique la décomposition d'un médicament en molécules plus petites, généralement facilitée par des enzymes dans le foie ou d'autres organes (les reins, la peau ou l'appareil digestif).
- Excrétion : L'excrétion est le processus par lequel un médicament ou ses métabolites sont éliminés de l'organisme, généralement par l'urine ou les fèces. D'autres méthodes d'excrétion existent, notamment par la peau, les poumons ou la sueur.
- Toxicité : Dans le cadre de l'ADMET, la toxicité désigne les effets nocifs ou les réactions indésirables d'une substance sur les organismes vivants. Cette évaluation est cruciale pour déterminer la sécurité d'une substance et prendre des décisions concernant les futurs développements ou l'utilisation de celle-ci.

5.3.5 Hit-To-Lead (H2L)

Cette étape vise à réduire le nombre de molécules candidates à un nombre plus limité de candidats susceptibles de progresser le long de la voie de conception du médicament jusqu'à un médicament efficace cliniquement. Au cours de cette étape, des méthodes de laboratoire et bio-informatiques sont adoptées pour identifier les molécules candidates qui présentent des propriétés suffisantes pour examiner leur activité dans des modèles *in vivo*.

5.3.6 Génération et optimisation des leads (Lead optimization)

Après la génération d'agents candidats par la voie du H2L, les composés candidats sont ensuite optimisés par amélioration de leur activité biologique, diminution de leur toxicité potentielle ou augmentation de leur capacité d'absorption.

5.3.7 Essais in vitro et in vivo

Une fois passées toutes les phases précoces de découverte de médicaments, il faut tester des molécules médicamenteuses dans l'environnement de cellules vivantes pour vérifier leur effet toxique potentiel et leur fiabilité. Au cours de cette étape, des cultures cellulaires ou des modèles animaux sont adoptés pour des tests *in vitro* et *in vivo*.

5.3.8 Phase préclinique

Avant de passer aux étapes cliniques, il faut mener davantage d'études pour fournir des preuves suffisantes sur les risques et les effets des substances identifiées lors de la découverte précoce de médicaments, avant de procéder à des essais cliniques sur l'homme. Il est également utile de fournir des quantités adéquates de nouvelles substances et de calculer les doses appropriées pour les essais sur l'homme au cours des études cliniques.

5.4 Phase clinique

Au cours de cette étape, il faut d'abord produire la substance active dans de bonnes conditions, selon les directives des bonnes pratiques de fabrication. Ensuite, les essais cliniques sont repris sur des individus, en quatre phases (**Figure I.9**).

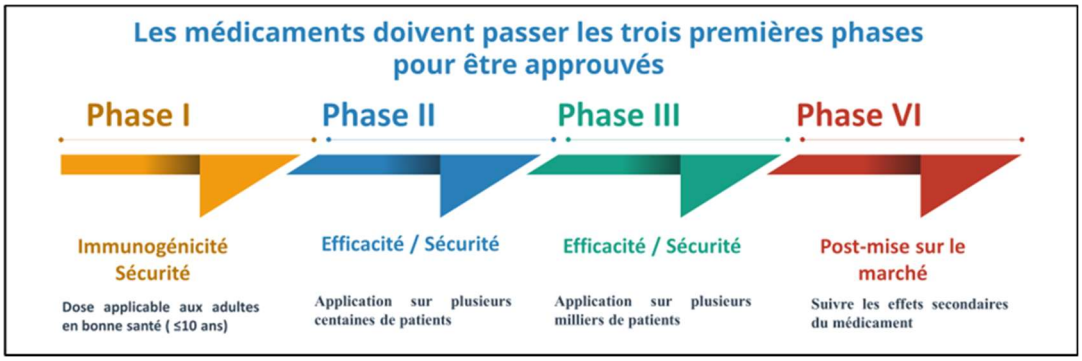


Figure I.9. Principales phases du processus de diagnostic clinique des médicaments.

Au cours de la phase I, la tolérance, l'efficacité et la sûreté du médicament candidat sont testées sur un groupe de personnes en bonne santé (20 à 80 personnes). L'objectif de cette phase est de déterminer la dose maximale admissible du médicament, d'examiner comment l'organisme réagit au médicament et de vérifier les effets secondaires. Dans la phase II, le médicament candidat est testé sur un groupe de patients (100 à 500). L'objectif de cette étape est de trouver la dose optimale à administrer comme médicament chez l'homme. La phase III consiste à tester le médicament candidat sur un grand groupe de patients (1000 à 3000 participants). L'objectif de cette étape est de s'assurer que le médicament est efficace et sûr chez de nombreux patients, qu'il n'a pas d'effets secondaires fréquents ou graves et que le nouveau médicament présente des avantages par rapport aux médicaments existants qui ciblent la même maladie. Les interactions possibles avec d'autres médicaments sont également testées. Les essais de phases II et III sont généralement appelés des études contrôlées, dans lesquelles certains patients sont traités avec le nouveau médicament et un autre groupe subit le traitement standard existant. La réussite de la phase III autorise le fabricant à demander l'autorisation officielle de mettre le médicament sur le marché. Immédiatement après cette étape vient la phase IV, connue sous le nom d'essais de surveillance post-commercialisation. Cette étape permet de collecter des données plus exhaustives sur l'efficacité et la sûreté du nouveau médicament, d'évaluer les effets à long terme du médicament et d'enregistrer et d'éviter les effets indésirables.

5.5 Autorisation de mise sur le marché

Lorsque la substance active répond à la majorité des essais cliniques et que les données nécessaires sont collectées et analysées. Le produit est ensuite soumis aux autorités responsables pour être examiné et évalué. Avant de vendre le médicament ou le vaccin développé, il faut obtenir une autorisation auprès des autorités chargées de la santé. Finalement, une seule molécule parmi le grand nombre de molécules testées est approuvée en tant que médicament. Après la vérification des données relatives à l'efficacité et à la sécurité du médicament par le CDER (Centre d'évaluation et de recherche sur les médicaments), la FDA procède à une analyse des maladies cibles et des traitements disponibles, ainsi qu'à une évaluation des avantages et des risques à partir des données cliniques et des stratégies de gestion des risques. Après avoir démontré qu'un médicament présente plus d'avantages que de risques ou connus, le médicament entre dans le processus d'approbation réglementaire par la FDA.

5.6 Contraintes et défaillances

La recherche pharmaceutique en vue de la découverte et du développement de nouveaux médicaments est confrontée à de nombreux défis, principalement scientifiques, mais aussi financiers et stratégiques [53]. Le délai considérable de plus de 10 à 15 ans, le risque élevé et le coût de plus de 1-3 milliards de dollars pour chaque nouveau médicament approuvé pour des essais cliniques sont des défis majeurs de la méthodologie classique de développement de médicaments (**Figure I.9**). De plus, le taux d'échec de la découverte et de la conception de médicament, du stade préclinique aux phases cliniques et à leur approbation, dépasse 90% [54]. La synthèse des analyses des données relatives aux essais cliniques réalisés dans la période 2010-2017 indique que les échecs cliniques dans le développement des médicaments sont liés à quatre causes : l'absence de besoin commercial, la mauvaise qualité de la stratégie (10 %), les propriétés pharmacologiques (10 % à 15 %), la toxicité (30 %) et l'absence d'efficacité thérapeutique (40 % à 50 %) [5,6]. Ce taux d'échec constamment élevé soulève plusieurs questions problématiques : Pourquoi 90 % des essais cliniques de développement de médicaments ne réussissent pas malgré la mise en œuvre de multiples stratégies expérimentales réussies au cours des dernières décennies ? Y a-t-il des stratégies appliquées pour la conception rationnelle de médicaments négligées ? Comment pouvons-nous améliorer le taux de réussite du développement de médicaments ?

Actuellement, grâce au développement de techniques bio-informatiques et de méthodes analytiques basées sur l'intelligence artificielle (IA) [57], il devient possible d'améliorer l'efficacité du processus de découverte de médicaments et d'augmenter les chances de succès d'un médicament jusqu'au stade le plus avancé. Pour cela, il faut élaborer des modèles prédictifs capables de traiter de manière plus efficace et plus rentable les principaux facteurs liés au succès ou à l'échec d'un médicament, mais aussi de prévoir les besoins des patients et les tendances du marché. Dans ce contexte, l'utilisation de techniques de Conception de Médicaments Assistée par Ordinateur (CADD) est devenue une démarche primordiale dans la plupart des projets de découverte/développement de médicaments [58], [59]. Par conséquent, les méthodes CADD peuvent aider à augmenter la proportion de réussite des médicaments et à éviter les scénarios défavorables. Dans ce contexte, nous présentons dans la sous-section suivante le contexte historique du développement de l'approche CADD, ainsi que ses différentes applications, catégories, avantages et limites.

6 Approche de la Conception de Médicaments Assistée par Ordinateur (CADD) : Classification, Utilisations, Limitations et Avenir

6.1 Progrès du CADD

Au fil des ans, coïncidant avec le développement des techniques de criblage à haut débit (HTS) utilisées pour rechercher des composés qui provoquent une réponse biologique souhaitée, le processus de découverte et de développement de médicaments est accéléré et facilité par le criblage d'une grande diversité de composés dans un laps de temps considérablement court. Mais les taux de réussite sont souvent très faibles, en raison d'échecs dans l'optimisation des molécules têtes de série ou dans les tests précliniques/cliniques. Le taux d'échec aux derniers stades de développement est

estimé à environ 40-60 % en raison de l'absorption, la distribution, le métabolisme, l'hypo-excrétion et la toxicité (ADME-Tox) [3]. Tous ces problèmes rendent l'utilisation des technologies de criblage à haut débit (HTS) très limitée. Par conséquent, tous ces problèmes soulignent la nécessité de développer des stratégies alternatives qui peuvent aider à améliorer les taux de réussite et à réduire les coûts et le temps dans l'ensemble du processus de découverte et de développement des médicaments. À cet égard, les progrès de la cristallographie aux rayons X et de la puissance de calcul ont permis d'améliorer rapidement les progrès de l'application des méthodes de CADD dans l'industrie pharmaceutique. L'application des approches CADD a permis de développer de nombreux médicaments commercialisés qui ont été découverts jusqu'à présent, à savoir : Saquinavir, Nelfinavir, Norfloxacin, Zanamivir, Amprenavir, Zolmitriptan, Dorzolamide [60], ce qui permet de réduire le coût de la découverte et du développement de médicaments de plus de 50% par rapport aux anciennes stratégies basées sur le criblage d'un grand nombre de composés par les tests expérimentaux [61]. Ce qui fait que les stratégies expérimentales prennent beaucoup de temps et des investissements à grande échelle. Par conséquent, l'application des approches CADD dans le criblage des composés chimiques dans la phase quasi initiale du processus de sélection réduisent de manière considérable le nombre de composés qui à testet *in vitro* et *in vivo*. Cela permet forcément de faciliter la découverte de nouveaux médicaments plus rentables et moins chers.

6.2 Approches de CADD

À ce jour, les approches de CADD jouent un rôle crucial dans les phases de conception précoce, d'optimisation et de repositionnement des nouveaux médicaments. L'une des principales stratégies utilisées en CADD est le criblage de chimiothèques de molécules virtuelles, également appelé Criblage Virtuel à Haut Débit (High-Throughput Virtual Screening, HTVS). Comparativement au criblage à haut débit (High-throughput screening, HTS) traditionnel basé sur la technologie expérimentale, qui demande plus de travail et de ressources et ne permet pas de comprendre le mécanisme moléculaire à l'origine de l'activité des molécules candidates, le HTVS a une charge de travail nettement inférieure. Le procédé HTVS utilise des algorithmes de recherche plus variés, qui reposent sur la recherche dans les bases de données disponibles de ligands et de structures des protéines ciblées [62]. En raison de la transparence des stratégies HTVS appliquées à des bases de données virtuelles massives, de la nécessité d'effectuer des tests expérimentaux et des évaluations des composés principaux sélectionnés, et du fait que la gamme de propriétés des candidats médicaments est loin d'être aléatoire. Tous ces points ont conduit à favoriser l'examen des bases de données qui contiennent des données limitées sur le nombre d'éléments chimiques, la qualité et l'activité biologique envers une cible spécifique (approche basée sur le ciblage moléculaire) [63]. Par conséquent, une approche fondée sur le criblage de bases de données contenant un nombre limité de molécules bioactives vis-à-vis d'une cible thérapeutique spécifique est susceptible de fournir des résultats plus précis et plus rapides à la conception rationnelle et à la découverte de nouveaux médicaments. Sur la base de ces aspects, nous nous sommes focalisés, dans les différentes études réalisées dans cette thèse, sur le criblage de bases de données moléculaire (chimiothèques) avec un ensemble limité de petites molécules issues de

produits synthétiques ou naturels et évaluées pour leur activité biologique *in vitro* en tant que premier pas vers l'optimisation et la conception rationnelle de nouveaux médicaments computationnels.

6.3 Applications et principe des approches CADD

Les approches de la CADD s'appliquent à un ensemble de techniques liées principalement aux premières étapes de la découverte de médicaments, à l'identification de candidats médicaments et à l'amélioration de leur efficacité clinique. Les notions de chimie computationnelle, de modélisation moléculaire, de conception moléculaire et de conception rationnelle de médicaments sont toutes des aspects de CADD. Les méthodes CADD aident à accélérer le processus d'identification des agents médicamenteux candidats à tester sur des modèles animaux et à les faire passer aux essais cliniques dans un délai court et peu coûteux [61]. Actuellement, les outils de la CADD sont appliqués à presque toutes les étapes de la découverte de médicaments (**Figure I.10**). Dans le processus de découverte et de développement de nouveaux médicaments, les méthodes CADD sont souvent utilisées à deux fins principales : (i) le criblage virtuel des bibliothèques de molécules disponibles et la sélection de molécules lead candidates pour les essais, sur la base de critères d'activité biologique prévue, d'affinité et de compatibilité avec la cible biologique, et de propriétés pharmacocinétiques (ADME-Tox) ; (ii) Conception et amélioration de nouveaux composés et utilisation de ceux-ci pour les comparer avec les données des tests biologiques. Si un composé passe toutes les étapes du criblage stratégique, il peut devenir un médicament candidat pour les essais cliniques suivants.

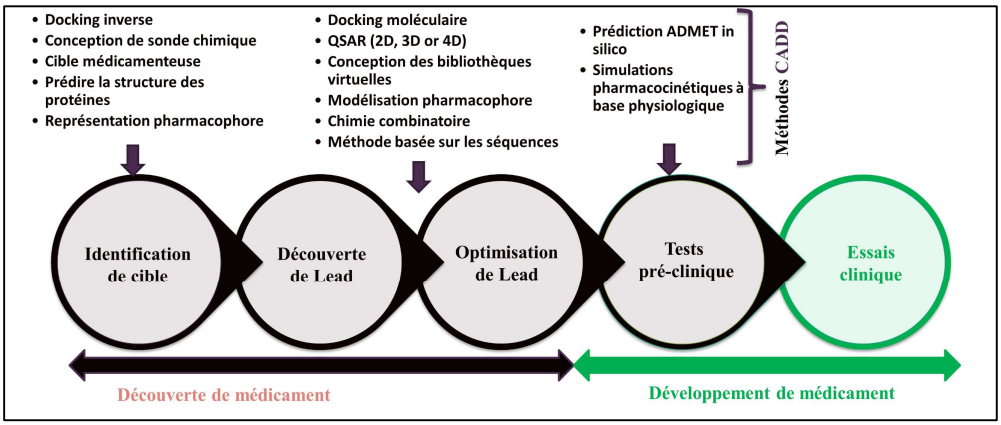


Figure I.10. Techniques computationnelles appliquées à la conception et au développement de médicaments**.**

6.4 Classification des approches de CADD

Les approches classiques du CADD peuvent être largement classées en deux catégories en fonction de la disponibilité de la structure 3D de la cible (enzyme/protéine) :(i) la conception de médicaments basée sur le ligand (LBDD) et (ii) la conception de médicaments basée sur la structure (SBDD). L'approche SBDD repose notamment sur les techniques suivantes : Criblage virtuel à l'aide du Docking moléculaire, modélisation pharmacologique basée sur la structure et conception de novo de médicaments, tandis que l'approche LBDD comprend le criblage basé sur la similarité, la modélisation quantitative structure-activité (QSAR), la modélisation pharmacopée basée sur le ligand, et le saut d'échafaudage « scaffold-Hopping » (**Figure I.11**).

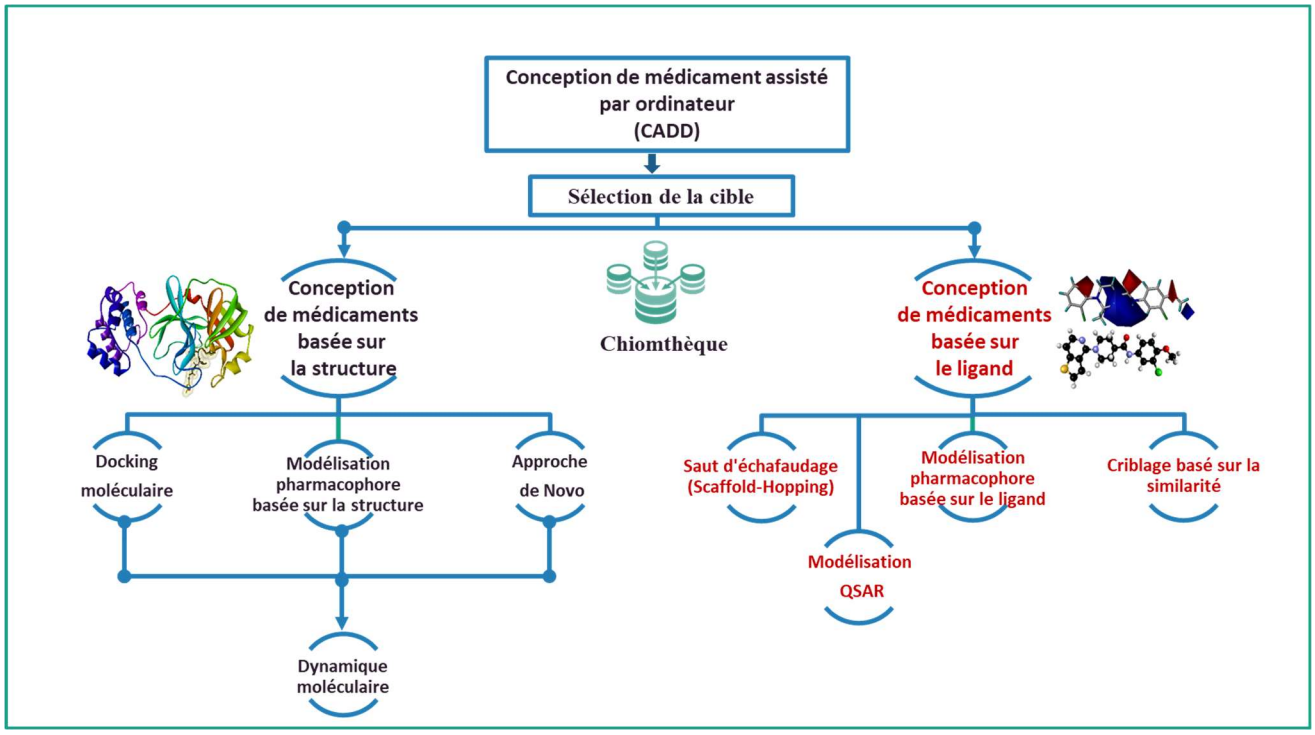


Figure I.11. Classification des méthodes de conception de médicament assistée par ordinateur (CADD).

6.4.1 Conception de médicaments basée sur la structure (SBDD)

La structure de la cible thérapeutique offre un avantage considérable dans le processus de découverte de médicaments par la stratégie de SBDD. La base de données des protéines (PDB) disponible sur RCSB PDB (http://www.rcsb.org/pdb) est l'une des ressources les plus utilisées pour identifier les structures 3D des cibles biologiques. La version actuelle du PDB contient plus de 197848 structures, identifiées avec haute précision par la cristallographie aux rayons X (X-ray), la spectroscopie de résonance magnétique nucléaire (RMN) et la spectroscopie de microscopie électronique (ME) [64]. Dans le cas où la structure 3D de la cible biologique n'est pas déterminée expérimentalement et n'est pas encore répertoriée dans la base de données PDB, elle peut être modélisée computationnellement. Il s'agit de construire un modèle approximatif de la protéine cible sur la base des structures 3D connues de protéines apparentées, via la technique de modélisation par homologie [65] .Sur la base de la structure de la cible biologique connue ou modélisée, la simulation du Docking moléculaire du ligand (généralement une molécule ou une entité) dans la poche active de la cible (le site de liaison actif) peut être appliqué. Le criblage virtuel par la simulation de Docking moléculaire assiste à prédire les conformations et les modèles de liaison protéine-ligand les plus appropriés et compatibles avec la cible biologique grâce à l'évaluation du niveau d'affinité entre la protéine et le ligand. En outre, les interactions entre le ligand co-cristallisé (médicament standard) et le site active de la protéine peuvent être traduites dans un modèle pharmacologique basé sur la structure et pouvant être employé pour cribler une large base de données de ligands actives qui sont compatibles avec la poche de liaison du récepteur. L'approche « de novo » basée sur la structure permet de générer de nouveaux ligands au sein de la poche active de la protéine par simple insertion de petits radicaux (atomes individuels ou fragments moléculaires) pas à pas [66]. Bien que l'approche de novo existe depuis plus de vingt ans, sa contribution aux projets de découverte de médicaments ne cesse de progresser en raison de sa facilité d'application. Néanmoins, le Docking moléculaire est un outil plus important dans le criblage virtuel que

d'autres techniques SBDD telles que De Novo, car il est capable de prédire l'affinité de liaison d'un composé à son récepteur cible. Ce type d'information est inestimable dans la découverte et le développement de nouveaux médicaments, car il permet aux chercheurs d'identifier et de classifier rapidement les candidats médicaments potentiels. En outre, le Docking moléculaire peut également être utilisé pour évaluer la précision des modèles de Novo. En comparant l'affinité de liaison prédite d'un composé avec l'affinité de liaison réelle, les scientifiques peuvent déterminer la précision de leurs modèles de De Novo et faire les ajustements nécessaires pour améliorer le pouvoir prédictif de leurs modèles. En général, le Docking moléculaire constitue un outil très utile dans le cadre du criblage virtuel au sein de la SBDD, car il offre une méthodologie fiable pour découvrir rapidement des candidats médicaments potentiels et évaluer la fiabilité d'autres modèles de criblage proposés tels que les modèles de novo [66]. Dans le cadre des stratégies de conception/optimisation rationnelle de nouveaux médicaments adoptées dans ce travail, nous avons adopté l'approche de Docking moléculaire comme entrée pour évaluer l'affinité/compatibilité des candidats médicaments avec les récepteurs biologiques ciblés (récepteurs de type protéine/enzyme). En tant que point de vérification des résultats de Docking moléculaire, il est indispensable d'évaluer la stabilité dynamique et le changement structurel des complexes biomarqueur-médicament par des méthodes biophysiques expérimentales (telles que des méthodes d'analyse spectrographique comme les UV, les rayons X) ou des simulations biophysiques avancées (telles que des simulations de dynamique moléculaire) [67]. Dans le contexte du post-Docking moléculaire mis en œuvre dans le cadre de cette thèse, nous avons combiné la simulation de la dynamique moléculaire et les méthodes de la mécanique moléculaire avec solvatation généralisée des naissances et des surfaces (Molecular mechanics with generalised Born and surface area solvation, MM/GBSA) comme procédure de validation permettant de validation biophysique computationnelle des stratégies CADD hybrides que nous proposons, ainsi que des résultats obtenus par voie computationnelle. Les simulations de dynamique moléculaire (MD) sont importantes dans les procédures post-Docking car elles peuvent fournir des informations sur le comportement dynamique d'un complexe protéine-ligand et sur l'effet de la liaison du ligand sur la protéine. Les simulations MD peuvent également aider à identifier des points faibles dans le complexe qui peuvent être étudiés plus en détail. Des méthodes de mécanique moléculaire telles que MM-GBSA peuvent être utilisées pour calculer l'énergie libre de liaison du complexe, ce qui peut fournir des informations importantes sur la stabilité relative du ligand et sa capacité à se lier au récepteur. Ces méthodes peuvent également être utilisées pour comparer différents ligands et déterminer lequel est le plus favorable à la liaison. Dans l'ensemble, les méthodes MD et de MM peuvent fournir des informations précieuses sur la stabilité, la spécificité et l'affinité d'un complexe protéine-ligand et peuvent aider à orienter les études ultérieures.

6.4.2 Conception de médicaments basée sur le ligand (LBDD)

Tandis que les méthodes LBDD ne nécessitent pas d'informations structurelles sur la protéine thérapeutique ciblée. Les méthodes LBDD reposent principalement sur la connaissance des activités biologiques observées *in vitro* de la base de données moléculaire analysée. C'est ce que permet de construire des modèles mathématiques capables de prédire l'activité biologique de nouvelles molécules à l'aide de techniques de ML, de chimio-informatique/bio-informatique et d'autres techniques d'IA [68].

Les méthodes QSAR multidimensionnelles (2D et 3D) constituent l'une des techniques les plus intéressantes employées dans la démarche de LBDD, à savoir l'étude de la relation structure-activité des molécules bioactive. Les méthodes QSAR permettent de modéliser la relation structure-activité grâce à l'identification des propriétés structurelles (descripteurs moléculaire) les plus significatives qui affectent positivement ou négativement l'activité biologique des molécules sans avoir besoin de la structure du récepteur protéique cible. Par conséquent, en utilisant les informations fournies par les modèles de QSAR, des caractérisations structurelles optimales pour la modélisation de l'activité biologique de nouvelles molécules peuvent être générées et améliorées avant la synthèse et/ou le test *in vitro*. Ainsi, des molécules de substitution peuvent être conçues et leurs activités biologiques prédites *in silico* à l'aide de modèles QSAR élaborés par les méthodes de Machine Learning (ML) et validés par la modélisation statistique [69]. L'approche de « criblage basée sur la similarité » utilise les informations relatives à la similarité des propriétés chimiques ou physicochimiques des composés [70]. Sur la base des valeurs de similarité moléculaire calculées, les composés sont considérés comme ayant une activité biologique similaire à celle des molécules de référence connues (généralement les médicaments). L'approche du criblage basée sur la similarité ne tient pas compte de l'activité biologique des ligands connus, qui est un bloc de base fondamental pour les techniques QSAR et de modélisation pharmacophore basée sur le ligand. Concernant la technique du « Scaffold Hopping », la structure chimique du ligand original est substituée par une nouvelle structure chimique, tout en maintenant, ou même en améliorant, l'activité biologique et les profils pharmacologiques du médicament original [71]. Le protocole de Scaffold Hopping est principalement basé sur deux approches, à savoir : (i) la recherche dans les bases de données moléculaires de médicaments disponibles ou hypothétiques pour trouver un nouveau chimio-type et non un Scaffold (échafaudage) spécifique. Ou (ii) la substitution de la structure de base du ligand original par un autre motif chimique différent, tout en conservant des interactions ligand-récepteur similaires avec des groupes latéraux de ligand original.

7 Application des Techniques de Machine Learning dans la Conception et la Découverte de Médicaments : Méthodes et Applications

L'apprentissage automatique (ML) est une branche de l'intelligence artificielle et de l'informatique qui utilise des données et des algorithmes pour simuler l'apprentissage humain, ce qui permet d'augmenter la précision [72]. C'est un domaine de science des données qui utilise des approches statistiques et des algorithmes entraînés pour générer des classifications ou des prédictions qui révèlent des informations critiques dans l'exploration des données. Il est catégorisé en cinq groupes selon les méthodologies et les techniques d'apprentissage : Supervisé, Non supervisé, Semi-supervisé, Apprentissage par renforcement et Apprentissage profond [73]. L'apprentissage supervisé utilise des jeux de données étiquetés pour entraîner des machines à prédire des résultats, tandis que l'apprentissage non supervisé entraîne des machines à l'aide de jeux de données non étiquetés sans supervision. L'apprentissage semi-supervisé combine des jeux de données étiquetés et non étiquetés pendant la phase d'entraînement, et l'apprentissage par renforcement entraîne des modèles d'apprentissage automatique à effectuer une série de décisions. Les techniques d'apprentissage automatique sont utilisées pour améliorer la prise de décision dans les données pharmaceutiques sur

diverses applications, notamment l'analyse QSAR, la découverte de composés actifs et la conception de médicaments de novo, ce qui permet d'obtenir des résultats plus précis [74]. L'apprentissage profond (deep learning) est devenu une base de plus en plus significative et productive pour les approches informatiques dans le développement de médicaments [75].

7.1 Apprentissage supervisé

L'apprentissage supervisé est une technique dans laquelle un jeu de données étiqueté est utilisé pour former des machines et faire des prédictions sur la base de la formation. Ce type d'apprentissage machine est subdivisé en deux catégories : Classification et Régression. Les algorithmes de classification sont utilisés pour prédire les catégories d'un jeu de données et résoudre des problèmes de classification. Des exemples de ces algorithmes sont l'algorithme Random Forest, l'algorithme de l'arbre de décision, l'algorithme de régression logistique et l'algorithme de machine à vecteurs de support. Les algorithmes de régression sont utilisés pour prédire des variables de sortie continues et aborder des problèmes de régression pour lesquels les variables d'entrée et de sortie ont une relation linéaire. Des exemples de ces algorithmes sont l'algorithme de l'arbre de décision et la régression Lasso, l'algorithme de régression multivariée l'algorithme de régression linéaire simple/multiple.

7.2 Apprentissage non superviser

L'apprentissage non superviser est une technique dans laquelle les machines sont formées avec un jeu de données non étiqueté sans aucune supervision. Le principal objectif de ce type d'apprentissage machine est de regrouper ou de catégoriser le jeu de données non trié selon les similarités, les modèles et les différences. Il est divisé en deux catégories : le regroupement et l'association. Les algorithmes de regroupement sont utilisés pour identifier les groupes inhérents dans les données et regrouper les objets dans le même groupe en fonction des similarités entre eux. Des exemples d'algorithmes de regroupement sont l'algorithme de regroupement K-Means, l'algorithme de Univariate Clustering, l'algorithme Mean-shift, l'algorithme DBSCAN (Density-Based Spatial Clustering of Applications with Noise), l'analyse en composantes principales et l'analyse en composantes indépendantes. Les algorithmes d'association sont utilisés pour découvrir les relations entre les variables dans un grand jeu de données et les mapper en conséquence pour maximiser les profits. Des exemples d'algorithmes d'apprentissage des règles d'association sont l'algorithme Apriori, Eclat et l'algorithme FP-growth.

7.3 Apprentissage semi-supervisé

L'apprentissage machine semi-supervisé est une technique qui combine des jeux de données étiquetés et non étiquetés pendant la phase d'entraînement. Son objectif principal est de tirer le meilleur parti de toutes les données disponibles plutôt que des données étiquetées, comme c'est le cas avec l'apprentissage supervisé. L'apprentissage non supervisé est utilisé pour regrouper d'abord des données similaires, ce qui aide à étiqueter les données non étiquetées en données étiquetées car l'acquisition de données étiquetées est plus coûteuse que l'acquisition de données non étiquetées.

7.4 Apprentissage par renforcement

L'apprentissage par renforcement est un type d'apprentissage machine dans lequel les modèles sont formés pour prendre une série de décisions. Ce type d'apprentissage est utilisé dans un environnement incertain et potentiellement complexe et l'ordinateur utilise l'essai et l'erreur pour trouver une solution au problème. L'apprentissage par renforcement est divisé en deux types : l'apprentissage par renforcement positif, qui ajoute quelque chose pour augmenter la probabilité que le comportement spécifié se reproduise, et l'apprentissage par renforcement négatif, qui a l'effet inverse en évitant la situation défavorable pour renforcer le comportement de l'agent.

7.5 Apprentissage profond

L'apprentissage profond est une branche de la machine learning qui utilise plusieurs couches de transformations non linéaires pour extraire des données [63]. Des réseaux neuronaux artificiels (ANN) inspirés du cerveau humain sont employés dans le « deep learning », permettant aux données d'être diffusées et partagées entre les différentes parties du réseau. Le deep learning supervisé et le deep learning non supervisé sont les deux principaux types de techniques de deep learning [64]. Les réseaux neuronaux artificiels (ANN), les réseaux neuronaux convolutifs (CNN) et les réseaux neuronaux récurrents (RNN) sont des algorithmes de deep learning supervisé. Chacun de ces algorithmes fonctionne différemment, mais ils impliquent tous la manipulation de poids et de biais pour réduire les erreurs et faire des prédictions précises sur des données non vues. L'apprentissage profond (deep learning) a servi à prédire les cibles, les hors-cibles et les conséquences dangereuses potentielles de divers produits chimiques.

8 Intégration de l'Intelligence Artificielle dans la conception et la découverte de médicaments : Applications et Potentiel

La technologie de l'IA permet de simuler l'intelligence humaine et sert à diverses applications dans le domaine des soins de santé, telles que la prédiction de la gravité d'une maladie et de l'efficacité d'un traitement, l'ajustement du traitement, l'identification de composés ciblés et l'optimisation de la conception de la structure des médicaments, ainsi que l'aide à la prédiction de la structure 3D d'une protéine ciblée, l'activité des médicaments et la conception de novo de médicaments [72], [76], [77]. En général, les méthodes heuristiques, les machines à vecteurs de support, les réseaux neuronaux artificiels, le processus de décision de Markov et le traitement du langage naturel sont les techniques d'IA fondamentales utilisées dans la découverte de médicaments [72]. Les méthodes heuristiques sont utilisées pour réduire le temps consacré au processus de reconstruction, tandis que les machines à vecteurs de support peuvent être utilisées pour la classification, la régression et la détection des valeurs aberrantes [75], [78]. Les réseaux neuronaux artificiels sont des dispositifs de traitement modelés approximativement sur la structure neuronale du cerveau et peuvent être utilisés pour prédire les propriétés ADMET, élaborations des modèles QSAR, construire des caractéristiques de surface moléculaire et effectuer un criblage virtuel [79]. Le processus de décision de Markov et le traitement du langage naturel sont utilisés pour prendre des décisions et traiter le langage, respectivement [80].

8.1 Applications de CADD, ML, IA dans la découverte de médicaments

Dans l'ensemble, grâce aux applications des technologies CADD, ML, AI, les chercheurs ont pu découvrir de nouveaux médicaments plus efficaces et plus sûrs, et réduire considérablement les coûts et le temps associés à la découverte de nouveaux médicaments. Le **Tableau I.2** énumère les plus récents exemples de succès enregistrés par les applications de CADD, d'IA et de ML dans le développement et la découverte de médicaments durant les dernières décennies du XXIe siècle.

Tableau I. 2. Exemples de succès de CADD, d'IA et de ML [72], [81], [82], [83]

Année	Exemple de succès	Domaine d'application	Année	Exemple de succès	Domaine d'application
2011-2013	Découverte du médicament Peroxisome	Dyslipidémie diabétique	2013-2017	Découverte du médicament Vaborbactam	Infections bactériennes.
	Découverte du médicament Axitinib	Nouvelle stratégie thérapeutique ciblant le carcinome avancé des cellules rénales.		Découverte du médicament Lifitegrast	Maladie des yeux secs.
	Découverte du médicament Crizotinib	Thérapie ciblée du cancer du poumon non à petites cellules (NSCLC)		Découverte du médicament Dolutegravir	Diminution de la proportion de VIH dans le sang et augmentation du nombre de cellules immunitaires.
2017-2020	Découverte du médicament Erdafitinib	Carcinome urothélial	2021-2023	Découverte d'inhibiteur N3	Inhibiteur du SARS-CoV-2
	Découverte du médicament Abemaciclib	Thérapie médicamenteuse contre le cancer du sein.		Découverte du médicament Tirabrutinib (nom commercial Velexbru).	Traitement du lymphome à cellules B et de la macroglobulinémie de Waldenström.
	Découverte du médicament Acalabrutinib	Thérapie ciblée contre le cancer.		Découverte du médicament Obeticholic acid (OCA)	Traitement de la maladie hépatique chronique de type cholangite biliaire primaire (CBP).
	Découverte du médicament Rivaroxaban	Thrombose veineuse profonde (TVP)			

Au-delà de ces énormes succès, la machine learning / deep learning et l'intelligence artificielle ont fortement contribué au renforcement des modèles prédictifs de haute performance utilisés pour analyser les activité biologique, propriétés de biodisponibilité et de pharmacocinétique / pharmacodynamique des molécules médicamenteuses candidates. Dans ce contexte, plusieurs outils en ligne aident à prédire rapidement et avec précision les profils de biodisponibilité et d'ADME-Tox des molécules qui peuvent être utilisées comme médicaments. Parmi ces outils, on retrouve (http://www.swissadme.ch/), (https://biosig.lab.uq.edu.au/pkcsm/) et (https://preadmet.qsarhub.com/), (https://www.molinspiration.com/), (https://playmolecule.com/Glimpse/) les plus populaires et

largement appliqués au cadre de CADD. Ces outils, développés grâce aux techniques d'apprentissage automatique et d'intelligence artificielle, aident les chercheurs à prévoir les propriétés et le comportement des médicaments avec précision. Ces outils sont inestimables pour la réalisation des études post pharmacocinétiques, le développement des médicaments, le criblage virtuel et la détection des molécules candidates à un usage pharmaceutique, car ils offrent la possibilité de procéder à une analyse plus détaillée et prédictive concernant les propriétés et le mode d'action des médicaments. Grâce à ces outils, les chercheurs peuvent comprendre de manière plus précise et plus rapide les propriétés d'un médicament donné et prendre des décisions éclairées sur la formulation, le dosage et l'administration du médicament. Par l'analyse détaillée du comportement d'un médicament dans l'organisme, ces outils peuvent contribuer à réduire le risque d'effets indésirables et à garantir qu'un médicament est utilisé de manière sûre et efficace. En outre, ces outils peuvent contribuer à accélérer le développement des médicaments, à réduire les coûts et à améliorer le taux de réussite des essais cliniques. Les aptitudes de ces outils pour analyser rapidement les données appliquées nous ont incités dans cette thèse à les exploiter pour modéliser les profils ADME-Tox et identifier des molécules médicamenteuses candidates.

8.2 Grands défis de CADD, ML, IA et perspectives

Les applications de l'IA et des outils de ML dans la conception de nouveaux médicaments sont plus avantageux que les méthodes traditionnelles de CADD, notamment lorsqu'il s'agit de traiter des bases de données volumineuses [84]. Cependant, la sélection de données appropriées et de haute qualité à partir de bases de données hétérogènes accessibles au public est une tâche très difficile. L'obstacle de la mauvaise adaptation due à l'absence de base de données homogène et de qualité contrôlée est un problème persistant lors de la mise en place de modèles basés sur la ML et l'IA. Pour surmonter ces difficultés, il est nécessaire de développer et de mettre en œuvre de nouveaux schémas de modélisation assistée par ordinateur (CADD) capables de générer des modèles applicables et plus rapides dans le cadre de la modélisation et du traitement rationnel des données. Dans ce contexte, durant la préparation de cette thèse, nous avons voulu accomplir la plus grande proportion possible des défis mentionnés par le biais de la mise en œuvre d'une stratégie hybride et pratique de CADD combinant les approches des modèles (LBDD et SBDD) soutenues par des techniques de ML et d'AI dans le criblage de chimiothèques homogènes de haute qualité sélectionnées à partir des littératures.

8.3 Limites de CADD et perspectives

Même avec le succès des applications des méthodes CADD dans la découverte avancée de médicaments, ces méthodes présentent de nombreuses limites. Par exemple, certaines molécules conçues *in silico* sur la base d'approches de chimie computationnelle et théoriques sont parfois incompatibles avec les systèmes biologiques réels [85]. À cause des propriétés pharmacocinétiques et/ou pharmacodynamiques inadéquates des molécules, moins de 40% des médicaments candidats passent la phase I des essais cliniques [86]. De plus, le manque ou l'absence de données expérimentales concernant les propriétés ADME-Tox des molécules candidates entraîne plusieurs échecs rapportés [19,20]. Par ailleurs, l'application des méthodes CADD basées sur des algorithmes

prédéfinis ayant des limites différentes, ces algorithmes doivent être mis à jour régulièrement afin d'améliorer la précision des calculs et de fournir de nouveaux médicaments. Aussi, les résultats fournis par l'approche CADD doivent être validés dans des systèmes biologiques réels. Afin de surmonter certaines des limites du CADD et d'améliorer sa précision lors de la modélisation des nouveaux candidats médicaments, une stratégie robuste doit être mise en place à l'aide de bases de données fiables (ex : bases de données PDB, données d'activité biologique évaluées *in vitro* ou *in vivo*, bases de données CSD (Cambridge Structure Data). De plus, la pertinence des méthodes computationnelles utilisées et la validité des hypothèses prévues doivent être contrôlées rétrospectivement vis-à-vis des données publiées dans la littérature. Les futurs développements du CADD pourraient être favorisés par l'adoption de nouvelles démarches expérimentales visant à développer de nouvelles cibles thérapeutiques contre des maladies complexes. Ceci pourrait être accompli par l'exploitation des résultats des recherches biomédicales actuelles, telles que les études d'association pangénomique (en anglais genome-wide association study, GWAS) [89], les techniques d'organes sur puce [90], l'imagerie par résonance magnétique fonctionnelle (IRMf) [91], la Tomographie par émission de positrons (TEP) et la technologie CRISPR/Cas9 [92], [93]. Jusqu'à présent, la mise en œuvre des méthodes SBDD indépendamment des méthodes LBDD est insuffisante pour caractériser la relation structure-activité, la structure-propriété, les profils de biodisponibilité, les propriétés pharmacocinétiques et pharmacodynamiques des molécules étudiées. Par conséquent, la combinaison de la cible biologique et des ligands bioactifs avec la mise en œuvre d'une stratégie hybride combinant les approches SBDD et LBDD, assistées par les techniques d'intelligence artificielle et de ML, peut fournir des solutions très fiables et claires pour rationaliser le criblage, l'optimisation et la sélection des nouvelles substances candidates, clés pour la conception et le développement de nouvelles stratégies thérapeutiques contre les maladies incurables. Par le biais de cette approche hybride que nous proposons dans cette thèse, il est possible de limiter la marge d'échec et de renforcer le succès de l'implantation cliniques de molécules médicamenteuses et pourquoi pas de mise sur le marché. Dans ce contexte, nous proposons et implémentons dans ce travail une stratégie hybride de CADD basée sur l'application des approches LBDD et SBDD renforcées par les techniques de ML/IA et les simulations biophysiques avancées dans le criblage, la modélisation et l'optimisation de l'activité biologique et de la compatibilité structurelle des molécules bioactives candidates. Cette stratégie permet de rendre les molécules bioactives plus favorables à la conception de médicaments, plus sûres et plus rentables dans le cadre de l'application thérapeutique. Ces différents volets sont abordés dans cette thèse, afin de mettre en œuvre une stratégie hybride plus exhaustive qui intègre des méthodes CADD basées sur le ligand et la structure dans la conception rationnelle de médicaments à base de petites molécules biologiquement actives. Nous discutons, dans la section à venir, de la définition des biomarqueurs retenus en tant que modèle pour appliquer la stratégie CADD hybride à des fins de recherche de nouvelles substances médicales contre le cancer et les infections virales, avant de déterminer les méthodes et outils à mettre en œuvre pour implémenter cette stratégie.

9 Approches de ciblage moléculaire employées dans ce travail de recherche

Dans le cadre de cette thèse, nous visons à étudier et à évaluer par voie computationnelle le pouvoir inhibiteur de petites molécules bioactives vis-à-vis des activités enzymatiques anormales de cibles biologiques impliquées dans le développement de lignées cellulaires pathologiques de nature cancéreuse et virales en tant que modèles de recherche. Dans ce cadre, nous nous focalisons sur l'optimisation computationnelle de nouvelles classes d'inhibiteurs réversibles, non-covalents, spécifiques pour les récepteurs biologiques de c-Met, EGFR-TKD, MALT1, et SARS-CoV-2-($3CL^{pro}$) impliqués dans le développement des différentes variétés de cancers et d'infections virales les plus graves, les plus mortelles, dont le cancer du poumon non à petites cellules, le cancer du sein, le lymphome, et le Covid-19, respectivement. La finalité consiste à favoriser la découverte précoce de nouveaux agents moléculaires, candidats à la conception de nouveaux médicaments efficaces, biologiquement très actifs, sûrs et faciles à atteindre. Dans les dernières années, les biomolécules générées par les cellules vivantes ont suscité un grand intérêt et une forte demande en raison de leur importance chimique, biologique, pharmaceutique et industrielle [94]. Ces biomolécules varient selon leur volume (géant/petit) ainsi que selon leur structure et leur source de production (à l'intérieur ou à l'extérieur de l'organisme) [94]. Par ailleurs, les biomolécules se composent principalement des éléments carbones, azote, oxygène, phosphore et hydrogène, qui constituent environ 96 % de l'espace biomoléculaire, tandis que d'autres minéraux tels que l'or, le magnésium, le zinc et le manganèse sont présents à l'état de traces [95]. Les macromolécules, notamment les enzymes, les protéines, les acides nucléiques, les lipides, les glucides et les petites molécules, sont toutes considérées comme des biomolécules d'une grande importance pour le contrôle des rôles structurels, fonctionnels et biologiques d'un organisme. Tout déséquilibre dans ces rôles peut être à l'origine de nombreuses maladies qui peuvent résulter principalement de l'expression anormale de l'activité enzymatique des protéines. Dans ce contexte, les petites molécules bioactives jouent un rôle important dans l'inhibition et le contrôle de l'activité enzymatique anormale des protéines, ce qui a accru la recherche de molécules bioactives pouvant être utilisées comme inhibiteurs médicamenteux potentiels. Ainsi, la modélisation de nouvelles inhibiteurs basées sur des petites molécules à potentiel activité inhibitrice est devenue cruciale pour fournir un traitement approprié à de nombreuses maladies incurables [21], [26]. Bien que les activités biologiques de nombreuses petites molécules aient été découvertes et évaluées *in vitro* et *in vivo*. Néanmoins, parmi environ 10000 molécules obtenues par voie synthétique, une seule molécule réussit à passer avec succès les tests cliniques avant d'être commercialisée ou rejetée pour diverses raisons. Parmi les raisons les plus importantes qui peuvent conduire au rejet de l'utilisation de molécules médicamenteuses et de leurs tests cliniques, nous trouvons : 1) leurs propriétés structurelles ne sont pas compatibles avec les caractéristiques de biodisponibilité similaires du médicament, 2) la pharmacocinétique appropriée n'est pas atteinte, 3) l'apparition d'effets secondaires indésirables, 4) la concentration de la dose est fatale, 5) les risques toxiques, 6) l'incapacité à atteindre l'objectif thérapeutique et à atteindre la réponse biologique. Par conséquent, la mise en œuvre de tests expérimentaux *in vitro* dans le cadre d'une stratégie de criblage traditionnelle conduit fréquemment à

une marge d'erreur plus importante vis-à-vis des effets thérapeutiques *in vivo/clinique* souhaités. Par conséquent, le taux de réussite des molécules filtrées en tant qu'agents médicinaux est relativement limité. Donc, pour éviter de telles difficultés potentielles qui pourraient perturber le processus de découverte de médicaments innovants, il est nécessaire de procéder à des tests *in silico* à chaque étape avant de passer aux tests *in vitro*/*in vivo* [96], [97]. De cette façon, les déséquilibres potentiels peuvent être explorés, adaptés et guidés vers le renforcement de la base de données des molécules (chimiothèques) par le design et la synthèse de nouvelles molécules qui ont plus de chances de réussir et d'atteindre les phases d'essais cliniques. De la sorte, les déséquilibres potentiels peuvent être explorés et adaptés en vue d'obtenir de nouvelles molécules plus susceptibles de devenir des agents médicamenteux. Dans ce contexte nous avons concentré dans cette thèse sur l'utilisation de la modélisation biomoléculaire et de l'adaptation structurelle comme outils scientifiques pour développer des agents pharmacologiques à applications thérapeutiques ciblées. Nos travaux de recherche sont basés sur des molécules synthétiques et des substances phytochimiques bioactives à des fins thérapeutiques spécifiques. Grâce à la diversité structurelle et les activités biologiques de ces molécules, nous espérons contribuer au développement de nouveaux médicaments destinés à lutter contre les maladies incurables provoquées par l'activité enzymatique anormale de certaines protéines spécifiques. À cet égard, nous abordons les avantages et les défis liés à la découverte de nouveaux agents inhibiteurs pour la thérapie ciblée contre le cancer et les infections virales en tant que modèles de recherche et d'application de la stratégie CADD que nous proposons dans cette thèse.

9.1 Optimisation de nouvelles voies de ciblage moléculaire anticancéreux

En 2020, le Centre international de recherche sur le cancer publié les estimations mondiales de l'incidence du cancer et de la mortalité liée au cancer selon GLOBOCAN 2020 [98]. Près de 10.0 millions de décès par cancer et 19,3 millions de nouveaux cas de la maladie ont été déclarés. Les types de cancer les plus répandus étaient le cancer du poumon et le cancer du sein féminin dans les nouveaux cas. En tête de liste, environ 2,3 millions de nouveaux cas (11,7 %) de cancer du sein féminin; suivis par le cancer du poumon (11,4 %), le cancer colorectal (10,0 %), le cancer de la prostate (7,3 %) et le cancer de l'estomac (5.6 %) [99]. Ainsi que d'autres types de cancer, comme le lymphome, responsable de 544000 nouveaux cas et de 260000 décès en 2020 [100]. Au vu de ces statistiques alarmantes, le cancer du poumon reste le plus meurtrier, avec un nombre de décès estimé à 1.8 million (18%), suivi du cancer colorectal (9,4%), du foie (8,3%), de l'estomac (7,7%) et du cancer du sein féminin (6,9%). Parallèlement à ces indicateurs terribles, suite aux évolutions démographiques et à l'augmentation des facteurs de risque associés à la mondialisation et à la progression de l'économie, la charge mondiale du cancer devrait atteindre 28.4 millions de cas en 2040, ce qui correspond à une augmentation de 47 % par rapport aux chiffres de 2020 [98]. Malgré les efforts étendus pour mettre en place une infrastructure durable pour la mise en œuvre de moyens de prévention contre le cancer et la prise en charge des patients atteints de cette maladie, l'oncologie est l'un des domaines où le nombre de médicaments ayant atteint le stade des essais cliniques est le plus faible par rapport à d'autres maladies telles que les vaisseaux sanguins et le cœur [3]. Le taux de réussite global des agents anticancéreux ne dépasse pas 10 % dans le développement clinique [101]. Par ailleurs, en raison des

progrès technologiques et de la modernisation des méthodes adoptées dans le cadre du développement et de la découverte de médicaments, le taux d'approbation de nouvelles molécules pour le traitement du cancer n'a cessé de diminuer au cours de la dernière décennie. La situation actuelle de la découverte et du développement de médicaments anticancéreux est moins prometteuse et moins prospective en raison de la complexité des approches de découverte de médicaments. La recherche d'agents cytotoxiques inhibant le processus de prolifération rapide des lignées cellulaires cancéreuses constituait le principal objectif de la mise au point de médicaments anticancéreux jusqu'aux années 1990 [102]. Depuis lors, avec l'avancement des méthodes de biologie moléculaire et la compréhension croissante du cancer au niveau moléculaire, l'ère de la "thérapie ciblée" est arrivée à la conception et au développement de médicaments anticancéreux [103]. La thérapie ciblée par petites molécules gagne en popularité en raison de ses avantages en termes d'efficacité et de sûreté par rapport aux médicaments de chimiothérapie traditionnels [104]. Et ce, depuis que le premier inhibiteur de la tyrosine kinase, l'imatinib, a été autorisé par la FDA américaine à entrer sur le marché en 2001 [105]. Depuis lors, plus de 89 médicaments antitumoraux à base de petites molécules ont été approuvés par la Food and Drug Administration américaine et l'Administration nationale des produits (NMPA) en Chine jusqu'en 2020 [104]. Malgré les progrès réalisés dans le domaine des médicaments anticancéreux à base de petites molécules, de nombreux défis persistent, notamment une faible réponse, une résistance aux médicaments résultant de mutations génétiques et la toxicité associée. Afin de maximiser la découverte de nouveaux médicaments anticancéreux à base de molécules de petite dimension, nous visons à développer et à optimiser leur structure pour les rendre plus actifs, plus compatibles et plus sûrs dans le cadre d'une thérapie ciblée contre le cancer. Nous nous appuyons sur la stratégie de diversité structurelle des petites molécules pour inhiber/réguler l'activité enzymatique des protéines qui jouent un rôle important dans le développement de divers types de cancer, notamment le cancer du poumon à petites cellules (CPNPC), le cancer de sein et le lymphome, en tant que modèles pour la modélisation et la conception médicamenteuse anticancer. Dans cet égard, les récepteurs de type tyrosine kinase, notamment l'EGFR et c-Met, jouent un rôle important dans la thérapie ciblée contre le cancer [106]. Ces récepteurs sont impliqués dans la signalisation cellulaire et leur activation peut entraîner la prolifération, l'invasion et la résistance aux médicaments. L'EGFR est une protéine membranaire qui se lie aux composés épidermiques et à certains facteurs de croissance et qui est impliquée dans la signalisation du facteur de croissance épidermique (EGF). C'est l'un des principaux récepteurs cibles pour les thérapies ciblées contre le cancer. Des médicaments ciblant spécifiquement ce récepteur sont utilisés dans le traitement de certains types de cancer, notamment le cancer du poumon, du sein et de la tête et du cou [107]. Le c-Met est un autre récepteur de type tyrosine kinase qui est impliqué dans la signalisation des facteurs de croissance et qui est également ciblé par des thérapies ciblées contre le cancer. Il est impliqué dans le développement et la progression de certains types de cancer, notamment le cancer du rein, du foie et de l'ovaire, etc. Cibler ces récepteurs de type tyrosine kinase dans le cadre de la thérapie ciblée contre le cancer peut aider à bloquer l'activation des facteurs de croissance et à inhiber la prolifération et l'invasion des cellules cancéreuses. Cela peut également aider à réduire la résistance aux médicaments et à améliorer les résultats thérapeutiques pour les patients atteints de cancer [108], [109]. D'autre part, la protéase MALT1 est

une protéine qui est impliquée dans le développement et la progression du lymphome [110]. Elle est associée à la signalisation des récepteurs EGFR-TKD et c-Met et peut en moduler l'activation [111]. Des études ont montré que des mutations de la protéase MALT1 et des anomalies de la signalisation EGFR-TKD et c-Met sont associées à un plus grand risque de lymphome et à des stades plus avancés de la maladie [108], [112]. Par conséquent, la protéase MALT1 et la signalisation EGFR-TKD et c-Met pourraient être des cibles importantes à considérer pour le développement de thérapies ciblées contre le lymphome. La proximité et la compatibilité des récepteurs protéiques sont étroitement liées au développement et à la prolifération des cancers les plus mortels, qui tuent des millions de personnes chaque année malgré les efforts constants pour les éradiquer ou au moins les transformer en maladies chroniques. Notre travail vise à contribuer à la découverte de nouveaux agents moléculaires qui pourraient interagir/interférer avec les récepteurs tyrosine kinase EGFR, c-Met et MALT1 pour inhiber/moduler leur activité enzymatique et limiter la génération de lignées cellulaires cancéreuses. Notre objectif est de proposer un répertoire des médicaments basés sur la technologie computationnelle qui pourraient être utilisés pour réduire la létalité et les effets secondaires du cancer et améliorer la santé publique.

9.1.1 Ciblage moléculaire du récepteur à tyrosine kinase c-Met

Le facteur de transition mésenchymateuse-épithéliale c-Met est un récepteur tyrosine kinase qui appartient à la famille MET, il se trouve notamment à la surface de diverses cellules [113]. Son coordinateur (ligand) principal est le facteur de croissance des hépatocytes (HGF en anglais Hepatocyte growth factor). La fixation de l'HGF sur c-Met déclenche la génération d'une cascade de signaux intracellulaires qui interviennent dans l'embryogenèse, la régénération des tissus et la formation des nerfs et des muscles, lesquels sont contrôlés par le suppresseur de tumeurs p53 [114, p. 53]. Par conséquent, c-Met joue un rôle essentiel dans les fonctions biologiques normales des êtres humains et des cellules normales. Cependant, l'activation anormale et la surexpression de c-Met favorisent le développement et la progression des tumeurs par le biais de la stimulation de PI3K/AKT, Ras/MAPK, JAK/STAT, Wnt/β-caténine et d'autres voies de signalisation [115], [116] (**Figure I.12**). L'activité enzymatique anormale et la surexpression de c-Met favorisent le développement de nombreux cancers humains tels que les cancers du poumon, du sein, du pancréas, du foie, du système nerveux, de l'ovaire, de la prostate et de l'estomac, ainsi que le mauvais pronostic précoce de ces pathologies [113]. Certaines études ont montré que l'activation anormale de c-Met contribue fortement à la résistance aux thérapies ciblées par les inhibiteurs de tyrosine kinase conventionnels et les médicaments ordinaires qui agissent contre les voies de signalisation liées à c-Met, comme l'Onartuzumab, le Tivantinib, le Crizotinib, le Cabozantinib, le Foretinib, le Tivantinib, l'Erlotinib, etc. [113]. La réponse anormale du récepteur c-Met peut contribuer à l'accroissement des difficultés liées au traitement des tumeurs et à l'apparition de résistances aux médicaments chez les patients. Par conséquent, il est important de rechercher des agents pharmacologiques plus puissants, plus stables et mieux adaptés avec la structure de c-Met que les médicaments généralement utilisés pour lutter contre le cancer. La découverte d'une nouvelle génération d'inhibiteurs ciblés spécifiquement l'activité anormale de la protéine c-Met pourrait s'avérer très bénéfique pour le traitement des tumeurs et

améliorer les résultats thérapeutiques. C'est pourquoi, pour notre part, nous avons contribué à cette orientation en menant des études computationnelles à grande échelle dans le but d'optimiser de nouvelles molécules candidates en tant qu'inhibiteurs spécifiques potentiels de c-Met ayant une activité, sélectivité et sûreté optimales en vue des thérapies ciblées contre le cancer.

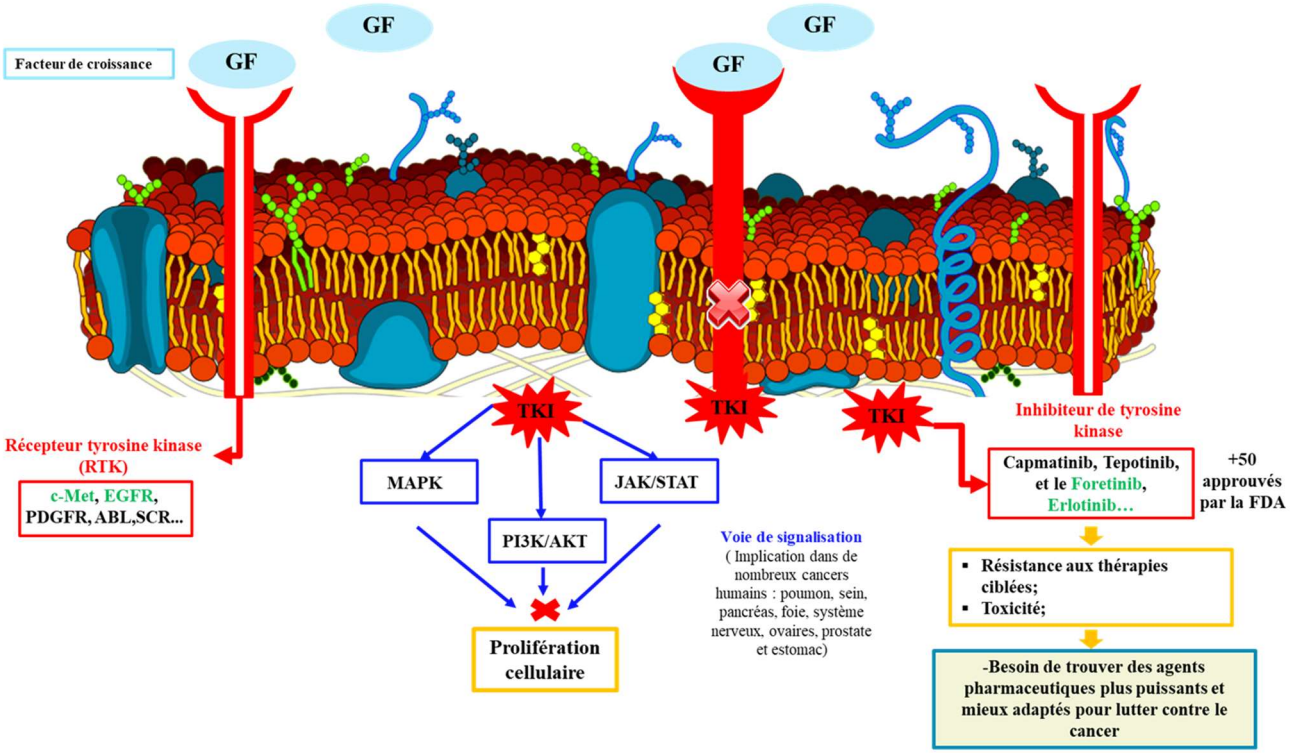


Figure I.12. Modèle de mécanisme moléculaire Schéma des voies de signalisation en aval du HGF/c-Met et de la médiation entre c-Met et d'autres voies de signaux de transduction cellulaires [117].

a) Voie de signalisation de c-Met

Les étapes principales du mécanisme de signalisation de c-Met peuvent être synthétisées en sept phases principales [118], [119]:

- Activation du récepteur c-Met : Le récepteur c-Met est activé lorsqu'il se lie à son ligand, à savoir le facteur de croissance des hépatocytes (HGF).
- Phosphorylation du récepteur c-Met : Lors de la liaison au HGF, le récepteur c-Met est phosphorylé sur ses résidus tyrosine.
- Activation de la voie PI3K/AKT : La phosphorylation du récepteur C-Met active la voie PI3K/Akt, ce qui entraîne une augmentation de la prolifération, de la survie et de la motilité des cellules.
- Activation de la voie RAS/MAPK : L'activation de la voie PI3K/AKT conduit également à l'activation de la voie RAS/MAPK, qui est responsable de la promotion de la migration et de l'invasion cellulaire.
- Activation de la voie STAT3 : L'activation de la voie RAS/MAPK conduit également à l'activation de la voie Stat3, qui est responsable de la promotion de la prolifération et de la survie des cellules.
- Activation de la Phospholipase C-γ : L'activation de la voie PI3K/AKT conduit à l'activation de la voie de la phospholipase C-γ (PLC-γ), qui est responsable de la production d'inositol trisphosphate (IP3) et de diacylglycérol (DAG), qui à leur tour activent la voie de la protéine kinase C (PKC).

- Activation de la voie de la protéine kinase C (PKC) : L'activation de la voie PKC conduit à la phosphorylation de nombreuses protéines effectrices, telles que la Paxilline, la kinase d'adhésion focale (FAK) et la kinase SRC, qui sont responsables de la médiation de la migration et de l'invasion cellulaire.

La mutation du gène c-Met après la liaison de ce dernier au facteur de croissance épidermique (EGF) entraîne l'accumulation de la protéine c-Met anormale, activant ainsi les voies de signalisation en aval [120]. Par conséquent, la prolifération aléatoire des cellules et le développement des tumeurs sont augmentés.

b) Mode d'inhibition de c-Met

Le mode d'inhibition de c-Met est un mécanisme qui permet de réguler la voie de signalisation de c-Met et d'inhiber son activité. Il inclut l'utilisation de petits inhibiteurs spécifiques de c-Met, qui se lient à la kinase c-Met et l'empêchent de se lier à son substrat et de transmettre des signaux. Ces inhibiteurs sont capables de bloquer la voie de signalisation de c-Met et peuvent ainsi jouer un rôle important dans la prévention et le traitement de cancer [121], [122]. À ce jour, plusieurs inhibiteurs de c-Met ont été développés et testés dans des essais cliniques contre les cancers non à petites cellules du poumon, du foie, du sein et d'autres [117]. Les inhibiteurs de la kinase à petite molécule peuvent bloquer la phosphorylation du domaine catalytique du récepteur par un antagonisme compétitif ou non compétitif du site de liaison d'ATP, empêchant ainsi le recrutement des transducteurs et médiateurs de signal et entravant ainsi la transduction du signal en aval. Les agents anti-c-Met peuvent être classés en trois types : les inhibiteurs sélectifs de la tyrosine kinase de c-Met (TKIs), les TKIs multi-cibles comprenant c-Met et les anticorps monoclonaux contre l'HGF ou c-Met [123]. Plusieurs essais cliniques récents ont indiqué qu'une seule thérapie médicamenteuse facilite le développement de la résistance aux médicaments [113]. Par conséquent, la recherche sur l'utilisation des inhibiteurs de c-Met en association avec d'autres médicaments, tels que les EGFR-TKI, accélérera le développement d'alternatives thérapeutiques efficaces [113]. Ceci a été confirmé par une étude effectuée sur un modèle animal de CPNPS (Cancer du poumon non à petites cellules) où l'injection de SU1274 a permis de vaincre la résistance à l'Erlotinib [124]. De plus, il a été découvert que le Cabozantinib peut surmonter la résistance au Crizotinib. Par conséquent, l'utilisation conjointe de plusieurs médicaments pour prévenir la résistance à un seul médicament pourrait devenir une priorité majeure pour les chercheurs en vue de développer de nouveaux inhibiteurs de c-Met. À la lumière de ces considérations, nous avons opté dans cette partie de la thèse de rechercher de nouveau médicaments anticancéreux, diversifiés et ayant des propriétés pharmacologiques intéressantes pour la thérapie anticancéreuse.

9.1.2 Ciblage moléculaire du récepteur du facteur de croissance épidermique (EGFR)

Le récepteur du facteur de croissance épidermique (EGFR) est un récepteur situé à la surface des cellules qui appartient à la famille des récepteurs ErbB; c'est un type de domaine des tyrosines kinases[125] . Après sa liaison avec le ligand, l'EGFR se transforme d'un monomère inactif en un homodimère actif [126]. La dimérisation de l'EGFR permet de catalyser l'activité de la protéine tyrosine kinase intrinsèque qui entraîne l'autophosphorylation des résidus tyrosine C-terminal [127].

L'autophosphorylation de l'EGFR active des cascades de signalisation en aval telles que les voies Ras/MAPK, PLCγ1/PKC, PI(3)kinase/Akt et STAT, ce qui conduit finalement à la production d'ADN, et à la prolifération et la progression du cycle cellulaire [128]. Certaines études ont confirmé que la transduction du signal entre la voie c-Met et l'EGFR est fortement associée au développement de nombreux types de cancer, notamment le cancer du poumon non à petites cellules, cancer de la tête, cancer du nez, cancer du sein et autres [129], [130]. Il a également confirmé que 70% des mutations actives de l'EGFR dans le cancer du poumon non à petites cellules sont associées à une réponse positive aux inhibiteurs de l'EGFR par le biais de l'Erlotinib et du Gefitinib [129]. Cependant, la plupart des tumeurs qui répondent aux inhibiteurs de l'EGFR développent une résistance acquise [131]. Dans ce contexte, une activité enzymatique c-Met excessive a été associée à une résistance primaire et acquise aux inhibiteurs de l'EGFR chez les patients atteints de NSCLC [132]. Ainsi, la résistance aux médicaments contre le CPNPC est causée par l'activation simultanée de c-Met et de l'EGFR, tandis que leur inhibition renforce l'effet inhibiteur sur la tumeur. Ainsi, l'inhibition simultanée de c-Met et d'EGFR peut permettre de surmonter la résistance aux inhibiteurs utilisés pour traiter le CPNPC et le cancer du sein [109], [133]. L'erlotinib est un inhibiteur de tyrosine-kinase de première génération utilisé comme traitement standard contre le cancer du poumon, mais des cas de résistance ont été signalés. Lapatinib, Neratinib et Pyrotinib ont été utilisés comme EGFR-TKIs pour le traitement du cancer du sein, mais ne sont pas unilatéraux et doivent être associés à d'autres médicaments [134]. Tamoxifène a été utilisé pour le traitement du cancer du sein depuis des décennies, mais une résistance au médicament a été observée. Des études ont montré que l'expression élevée d'ErbB est associée à une invasion tumorale accrue, une métastase, une résistance à la chimiothérapie et un mauvais pronostic [109]. Dans ce contexte, des recherches supplémentaires sont nécessaires pour explorer le repositionnement du Tamoxifène comme EGFR-TKI et le potentiel pour inhiber les lignées cellulaires cancéreuses résultant d'une surexpression de l'EGFR. En vue de ces scénarios complexes et conformément aux statistiques effrayantes de décès par cancer du sein et du poumon dans le monde entier, qui sont rapportés pour représenter plus de 50 % de tous les décès [135]. Le taux de mortalité alarmant de ces deux types de cancer et l'absence de médicaments efficaces, sûrs et stables nous ont incités à nous concentrer sur eux comme prototypes pour explorer des agents thérapeutiques potentiels contre eux en inhibant l'activité enzymatique d'EGFR/c-Met. Ceci est fait en concevant / découvrant / améliorant de nouveaux inhibiteurs spécifiques pour ces récepteurs protéiques plus adaptés à l'utilisation pharmaceutique par rapport aux médicaments courants tels que Erlotinib, et Tamoxifène en tant qu'exemple.

a) Voie de signalisation de EGFR-TKD

Le mécanisme de signalisation de l'EGFR-TKD est un processus complexe impliquant plusieurs protéines et molécules. Elle est déclenchée par la liaison d'un ligand à un récepteur de la famille des récepteurs à tyrosine kinase (EGFR), ce qui entraîne la phosphorylation de la tyrosine sur le domaine intracellulaire du récepteur. Cette phosphorylation active le récepteur et permet la formation d'un complexe multiprotéique qui comprend des protéines adaptatrices et des kinases intracellulaires [108]. Les protéines adaptatrices, telles que les protéines de liaison à l'EGF (GRB2 et SOS), permettent la

liaison entre la protéine RAS et son activateur GTP (en anglais, Guanosine triphosphate), ce qui entraîne l'activation de la cascade MAPK/ERK. Les kinases intracellulaires, telles que les protéines PI3K et AKT, activent également des cascades de signalisation qui entraînent des réponses cellulaires telles que la prolifération, la survie et l'invasion des cellules cancéreuses [136]. Par conséquent, l'activation de la voie EGFR-TKD entraîne l'activation de plusieurs cascades de signalisation qui contribuent à la croissance et à la progression des cellules cancéreuses.

b) Mode d'inhibition de EGFR-TKD

L'inhibition de l'EGFR-TKD désigne l'inhibition de l'activité de la tyrosine kinase de la protéine du récepteur du facteur de croissance épidermique (EGFR). L'EGFR est une protéine transmembranaire qui joue un rôle essentiel dans la prolifération et la survie des cellules. Lorsqu'il est activé par ses ligands, l'EGFR subit un changement au niveau de sa conformation qui amène son domaine tyrosine kinase intracellulaire (TKD) à proximité d'autres protéines intracellulaires, ce qui lui permet de les phosphoryler et de les activer. Les inhibiteurs de l'EGFR-TKD fonctionnent en établissant des liaisons avec le TKD et en inhibant sa capacité à phosphoryler d'autres protéines. Cela empêche à la fois les voies de signalisation en aval qui conduisent à la prolifération et à la survie des cellules, et peut conduire à la mort cellulaire. Il existe plusieurs types d'inhibiteurs de l'EGFR-TKD, notamment les inhibiteurs à petites molécules et les anticorps monoclonaux. Les inhibiteurs de l'EGFR-TKD sont des médicaments qui bloquent la kinase intracellulaire de l'EGFR, empêchant ainsi la phosphorylation de ses protéines cibles et l'activation des voies de signalisation. Il existe différents types d'inhibiteurs de l'EGFR-TKD, tels que les inhibiteurs de la petite molécule, tels que le Gefitinib et l'Erlotinib qui se lient directement à la kinase intracellulaire de l'EGFR, et les anticorps monoclonaux, tels que le Cetuximab et le Panitumumab, qui se lient à l'extérieur de la cellule à la région extracellulaire de l'EGFR et empêchent son activation. Les études cliniques ont montré que les inhibiteurs de l'EGFR-TKD ont une efficacité significative dans le traitement du cancer du poumon non à petites cellules (NSCLC) chez des patients présentant des mutations spécifiques de l'EGFR (tels que les mutations de la tyrosine kinase exon 19 ou exon 21) [137], [138]. Les inhibiteurs de l'EGFR-TKD sont également utilisés en combinaison avec d'autres thérapies pour augmenter leur efficacité [139]. Cependant, il est important de noter que l'utilisation des inhibiteurs de l'EGFR-TKD peut également entraîner des effets secondaires tels que des réactions cutanées et de la diarrhée [139]. De plus, certains patients développent une résistance aux inhibiteurs de l'EGFR-TKD, ce qui nécessite l'utilisation de thérapies alternatives [140]. En résumé, l'inhibition de c-Met/EGFR-TKD est un mécanisme efficace pour traiter certains types de cancer, notamment le cancer du poumon non à petites cellules et le cancer de sein, mais il est important de prendre en compte les effets secondaires potentiels et les possibilités de résistance. Dans ce contexte, nous envisageons dans ce travail de concevoir de nouveaux inhibiteurs de c-Met et d'EGFR-TKD qui soient efficaces, moins toxiques et qui présentent des propriétés pharmacocinétiques et pharmacodynamiques adaptées au développement d'un modèle thérapeutique contre le CPNPC et le cancer de sein basé sur des petites molécules bioactives.

9.1.3 Ciblage moléculaire de la Protéine-1 de translocation de lymphome du tissu lymphoïde associé aux muqueuses (MALT1)

Les lymphomes à grandes cellules B sont un type courant de lymphome non hodgkinien, représentant 30% des cas dans le monde avec 150 000 nouveaux cas par an [141]. Les symptômes comprennent une lymphadénopathie progressive et une maladie extranodale. Le traitement R-CHOP peut guérir 60 % des patients, mais les patients en échec thérapeutique ont de mauvais résultats [142]. Les progrès réalisés dans la compréhension de la maladie ont permis d'améliorer sa caractérisation et d'élaborer de nouvelles approches thérapeutiques. Le sous-type de lymphome de grandes cellules B activé (ABC) du lymphome de grandes cellules B diffus (DLBCL) est agressif et associé à un pronostic défavorable en cas de rechute ou de réfractaire [110]. L'ABC-DLBCL est caractérisé par une activation chronique du signal NF-κB, dépendante du complexe CARD11-BCL10-MALT1 (CBM) [143]. MALT1 est un acteur clé du complexe CBM et active des facteurs de transcription tels que le NF-κB canonique et l'AP-1 via des fonctions de protéase et d'échafaudage distinctes [110]. De plus, en raison de ses diverses fonctions biologiques impliquées dans l'immunité, l'inflammation et le cancer, MALT1 est devenu une cible thérapeutique prometteuse dans les maladies liées au lymphome B-cellule et T-cellule [144]. Il y a donc un intérêt croissant pour cibler thérapeutiquement MALT1 pour les maladies des lymphomes. Dans ce contexte, des études ont confirmé que l'expression anormale de l'activité enzymatique de MALT1 provoque le développement de plusieurs pathologies, notamment des maladies auto-immunes, des tumeurs solides et des lymphomes [145], [146]. Ainsi, l'inhibition de l'activité enzymatique de la protéase MALT1 serait bénéfique dans la lutte contre le lymphome comme modèle.

a) Voie de signalisation de MALT1

Le protéase MALT1 (Protéine de Translocation de Lymphome Tissulaire Associé à la Muqueuse 1) joue un rôle crucial dans la régulation des réponses immunitaires et est impliqué dans l'activation de la voie du facteur nucléaire κB (NF-κB), qui régule l'expression de gènes impliqués dans l'inflammation et la survie cellulaire [147]. Lorsqu'un récepteur antigénique sur une cellule T ou B est stimulé, MALT1 est activé et clive ses substrats, entraînant l'activation de plusieurs voies de signalisation en aval [148]. MALT1 agit comme une protéine de cadre pour recruter d'autres protéines de signalisation sur le site de stimulation du récepteur antigénique, formant un complexe de signalisation qui comprend la famille d'adaptateurs TNF (receptor-associated factor, TRAF), le complexe IKK (kinase IκB) et la voie MAPK (protéine kinase activée par les mitogènes) [149]. L'activation du complexe IKK entraîne la phosphorylation et la dégradation de la protéine inhibitrice IκB, libérant la forme active de NF-κB pour entrer dans le noyau et réguler l'expression de gènes cibles. Cette activation de NF-κB est cruciale pour l'induction de gènes pro-inflammatoires, la survie cellulaire et la régulation des réponses des cellules T et B [150]. MALT1 interagit également avec d'autres protéines de signalisation, telles que caspase-8, pour réguler l'apoptose, la survie cellulaire et l'activation du système immunitaire [151]. En résumé, l'activation du MALT1 entraîne l'activation de la voie NF-κB, ce qui conduit à la production de cytokines, à la prolifération cellulaire et à la survie. La dérégulation du MALT1 est impliquée dans plusieurs pathologies immunitaires et inflammatoires, notamment les maladies auto-immunes et le

cancer[152]. Il est nécessaire de mener davantage la recherche pour comprendre les diverses fonctions et la régulation de MALT1 dans le système immunitaire et son lien avec le développement de pathologies cancéreuses.

b) Mode d'inhibition de MALT1

Il est possible d'inhiber l'activité enzymatique de la protéase MALT1 selon des modes de régulation multiples et différents, parmi lesquels les plus importants sont les suivants :

- Inhibition de la protéase : Les inhibiteurs de la protéase MALT1 basés sur des petites molécules biologiquement actives, ou des molécules peptidiques contenant des têtes réactives, peuvent se lier directement à la protéine MALT1 et inhiber son activité enzymatique. Il existe deux classes de ce type d'inhibiteurs [153], [154]: les inhibiteurs allostériques tels que les dérivés de la phénothiazine (mépazine, promazine et thioridazine), et les inhibiteurs covalents tels que Z-VRPR-fmk[155, p. 9] et MI-2 [156] (aryl triazoles).
- Inhibition de l'ubiquitination : MALT1 est régulé par l'ubiquitination, un processus par lequel les protéines sont ciblées pour dégradation. Les inhibiteurs de MALT1 peuvent cibler ce processus pour stabiliser MALT1 et inhiber son activité. Par exemple, des études ont montré qu'un inhibiteur de petite molécule appelé LCL161 peut inhiber l'activité de MALT1 en bloquant son ubiquitination [157].
- Interférence avec les interactions protéine-protéine : La fonction de MALT1 dépend de son interaction avec d'autres protéines [158]. Les inhibiteurs de MALT1 peuvent interférer avec ces interactions pour perturber l'activité de MALT1 [159]. Par exemple, des études ont montré que le composé PRI-724 peut perturber l'interaction entre MALT1 et son partenaire de liaison PAR-1 pour inhiber le signalement NF-κB [160].

Au-delà des mécanismes indiqués ci-dessus, il existe également d'autres modes d'inhibition de MALT1, comme l'inhibition de son expression génétique, l'inhibition de la maturation et de la localisation des cellules, ainsi que l'inhibition des interactions avec d'autres protéines associées à la signalisation [161].

9.2 Optimisation d'une nouvelle voie de ciblage moléculaire antiviral contre le SARS-CoV-2

Le virus Covid-19 désigne une pathologie infectieuse due à un nouveau coronavirus (SRAS-CoV-2) qui a été identifié en 2019 pour la première fois à Wuhan, en Chine [162]. Les symptômes du Covid-19 sont notamment la fièvre, la toux, l'essoufflement, les frissons et les courbatures. Dans les cas plus sévères, le virus peut présenter une dépression, un syndrome respiratoire aigu, des insuffisances rénales, et même conduire jusqu'à un décès [163]. La meilleure façon de prévenir l'infection est de pratiquer la distanciation sociale, porter un masque, l'hygiène des mains et éviter le contact avec les personnes affectées. À ce jour, le monde entier connaît encore l'épidémie de COVID-19 causée par le SRAS-CoV-2 (Coronavirus 2 du Syndrome Respiratoire Aigu Sévère), sans qu'un véritable traitement capable de lutter efficacement contre ce type de virus soit disponible. Selon les chiffres déclarés par l'OMS (Organisation mondiale de la Santé), plus de 649 millions de cas confirmés et plus de 6,6 millions de décès ont été signalés dans le monde jusqu'en décembre 2022 [164]. En raison du caractère

dangereux de ce virus et de ses effets néfastes sur la santé et le déroulement habituel de la vie, de nombreuses mesures de précaution rigoureuses ont été appliquées afin de prévenir la transmission et la propagation rapide du COVID-19 entre les populations. Parallèlement, les chercheurs du monde entier ont déployé de nombreux efforts pour développer un médicament ou un vaccin capable de combattre cette infection. Dans cette optique, les activités continuent et approfondies menées par les chercheurs ont conduit à la proposition de nombreux vaccins et médicaments, dont certains ont été approuvés par la FDA et d'autres sont encore en phase d'évaluation clinique [165]. Cependant, il n'existe aucun médicament/vaccin antiviral approuvé cliniquement et disponible spécifiquement pour la lutte contre le COVID-19 [166]. La plupart de ces vaccins/médicaments qui ont été proposés et appliqués sont soit des médicaments conçus à l'origine pour d'autres maladies, soit conçus en ab initio sur la base d'une description de médicaments basée sur la structure [165], [167]. Bien que la proposition de plusieurs antiviraux tels que le Remdesivir, la Dexaméthasone et le Tocilizumab pour le traitement du COVID-19 ait été faite, leur application est limitée en raison de leur coût et de leurs effets secondaires potentiels [168]. Par conséquent, les vaccins apparaissent comme le moyen le plus efficace de se protéger contre ce virus. Quatre vaccins ont été approuvés par la FDA pour être utilisés en cas d'urgence : Pfizer, Moderna, AstraZeneca et Novavax [169]. Ces vaccins se sont avérés sûrs et efficaces pour prévenir le COVID-19, bien que certains effets secondaires puissent survenir en raison de l'inadéquation de ces vaccins aux cas de maladies chroniques comme celles liées à la cardiologie ou à la faiblesse de la réponse immunitaire [170]. Bien que des traitements suggérés soient disponibles à cet égard, les mutations génétiques des protéines gérées par le virus rendent le SRAS-CoV-2 durement dangereux et répandu [171]. Cela limite également l'efficacité des médicaments / vaccins proposés pour combattre le virus COVID19. La nature du processus de développement viral du SRAS-CoV-2 entraîne la génération de nouvelles mutations clonées à partir de celui-ci, de sorte que ces mutations n'ont que peu d'effet sur les caractéristiques du COVID-19 d'origine, à savoir la vitesse de propagation et le niveau de dangerosité [172]. Par conséquent, les mêmes médicaments/vaccins utilisés pour le SRAS-CoV-2 natif peuvent être adoptés pour combattre ses mutations génétiques telles que les variantes alpha (B.1.1.7), bêta (B.1.351), gamma (P.1) et delta (B.1.617.2) [173]. Cependant, plusieurs études ont confirmé que les personnes vaccinées sont toujours exposées au risque de réinfection par le virus COVID-19. Tous ces aspects réclament la nécessité d'un engagement collectif et l'intensification de futures études pour la mise au point et le développement d'un nouveau médicament/vaccin plus efficace et spécifique contre le SRAS-CoV-2 et ses variantes potentielles. À cet égard, à l'aide de ressources informatiques avancées et des méthodes bio-informatiques ultramodernes, la séquence du virus COVID19 a été codée et sa structure génétique et génomique a été découverte. On a découvert que le génome du SRAS-CoV-2 code pour vingt protéines différentes, dont de nombreuses protéines non structurelles, notamment la principale protéase du coronavirus comme $3CL^{pro}$ ou M^{pro}(3C-like proteinase ou main protease), RdRp (RNA-dependent RNA polymerase), PL^{pro} (Papain-like protease), et une protéine structurelle principale, à savoir la spike (S) ; l'enveloppe (E) ; la membrane (M) ; la nucléocapside (N) [166], [174]. Cette découverte concrète a permis de développer et de proposer différentes classes de médicaments potentiels capables de lutter contre le COVID-19 par l'inhibition de l'activité enzymatique de protéines du SRAS-CoV-2 identifiées.

Cependant, malgré tous ces efforts, le 9 novembre 2021, une nouvelle variante du SRAS-CoV-2 a été découverte, appelée Omicron, dont la plupart des mutations sont associées à la protéine S[175]. Cela confirme que le fait de cibler des protéines structurelles comme la protéine S ne constitue peut-être pas une solution plus efficace pour lutter contre le COVID-19 [174]. En revanche, la possibilité d'une mutation de la $3CL^{pro}$ reste très faible, ce qui pourrait faire de l'inhibition de l'activité enzymatique de cette dernière une solution plus adaptée au traitement du COVID-19 [174]. Dans cette perspective, nous nous attachons dans cette partie de la thèse à contribuer à la découverte de nouveaux inhibiteurs antiviraux portant sur l'activité enzymatique de la SARS-CoV-2 $3CL^{pro}$ en vue de combattre l'infection due au COVID-19.

a) Ciblage moléculaire de la protéase $3CL^{pro}$

L'enzyme $3CL^{pro}$, également connue sous le nom de protéase à 3-chymotrypsine, est une enzyme présente dans le virus SARS-CoV-2 [174]. Il est responsable du clivage et de l'activation des grandes protéines virales [176]. Cet enzyme est essentiel pour la réplication et la propagation du virus, et est donc une cible pour le développement d'inhibiteurs pour traiter le COVID19. L'enzyme $3CL^{pro}$ a été trouvée être la protéase la plus importante et la plus conservée de SARS-CoV-2. De plus, la séquence de cet enzyme a évolué au fil du temps en réponse à l'environnement changeant [177]. On pense que ces mutations augmentent l'activité de l'enzyme et sa résistance aux inhibiteurs existants. En conséquence, de nouveaux inhibiteurs doivent être développés pour cibler les formes mutées de l'enzyme, afin de combattre efficacement le virus. En conclusion, l'enzyme $3CL^{pro}$ est une composante essentielle du virus SARS-CoV-2 et a été trouvée être hautement conservée et ayant évolué en réponse aux changements environnementaux. Cela a entraîné le besoin de développer de nouveaux inhibiteurs afin de combattre efficacement des virus tels que le COVID19.

b) Voie de signalisation de $3CL^{pro}$

Le processus de signalisation $3CL^{pro}$ est au cœur de l'évolution et de la réplication du SARS-CoV-2. Il commence par la liaison de l'enzyme $3CL^{pro}$ à la protéine virale [178], [179]. Cet événement induit un changement de conformation de l'enzyme $3CL^{pro}$ qui active son site actif et permet de découper les protéines virales en fragments plus petits. Ces protéines sont alors reconnues et liées par des protéines cellulaires, ce qui leur permet d'être intégrées à la machinerie de réplication du virus. Par la suite, de nouvelles particules virales sont libérées pour se propager et potentiellement infecter d'autres cellules. L'inhibition de la voie $3CL^{pro}$ offre une opportunité prometteuse pour le contrôle de SARS-CoV-2 [180]. En effet, l'inhibition de cette voie empêcherait le virus de se répliquer et de se propager, ce qui contribuerait à réduire le nombre d'infections. De plus, cette inhibition pourrait diminuer la gravité des symptômes associés à SARS-CoV-2, tels que la fièvre, la toux et la difficulté à respirer. Il s'agit d'une cible thérapeutique potentielle pour le traitement des infections à SARS-CoV-2. Dans ce contexte, nous comptons, à travers cette partie du travail, développer de nouveaux inhibiteurs enzymatiques de $3CL^{pro}$ à partir de petites molécules ayant des propriétés structurelles optimales pour la conception de médicaments contre le COVID19.

c) Mode d'inhibition de $3CL^{pro}$

L'enzyme $3CL^{pro}$ peut être inhibé par deux principaux types d'interactions enzymes-inhibiteurs : l'inhibition covalente et l'inhibition non covalente. L'inhibition covalente implique la modification covalente directe de l'enzyme par une molécule, telle qu'un analogique du substrat, qui se lie à la zone active de l'enzyme et bloque son activité. Des exemples d'inhibiteurs covalents de $3CL^{pro}$ comprennent l'acide acétylsalicylique, la sulfanilamide et la benzamidine. L'inhibition non-covalente implique la liaison d'une molécule à une enzyme sans modification covalente. Des exemples d'inhibiteurs non covalents de $3CL^{pro}$ comprennent des inhibiteurs compétitifs tels que la pyridoxal-phosphate, des inhibiteurs non compétitifs tels que l'oxyde de phénylarsine et des inhibiteurs allostériques tels que le NADPH (Nicotinamide adénine dinucléotide phosphate) [174]. Les inhibiteurs non covalents de $3CL^{pro}$ sont importants car ils offrent une façon réversible et spécifique d'inhiber l'activité enzymatique. Contrairement aux inhibiteurs covalents, qui forment une liaison covalente permanente avec l'enzyme et inhibent ainsi de manière irréversible son activité, les inhibiteurs non covalents peuvent se lier de manière réversible à l'enzyme et peuvent être facilement retirés lorsqu'ils ne sont plus nécessaires. Cela les rend idéaux pour une utilisation dans les traitements médicaux, car ils peuvent être utilisés pour contrôler précisément l'activité des enzymes dans le corps. De plus, les inhibiteurs non covalents se lient souvent plus spécifiquement que les inhibiteurs covalents, ce qui les rend plus utiles pour cibler des enzymes particulières. Dans ce volet de la thèse, nous nous focalisons sur la modélisation d'inhibiteurs non covalents de la $3CL^{pro}$ basés sur de petites molécules biologiquement actives. Du fait de leurs propriétés distinctives vis-à-vis de leurs homologues covalents.

10 Conclusion

Dans ce chapitre, nous avons minutieusement examiné les traitements de médecine courante, en mettant particulièrement l'accent sur la stratégie de thérapie ciblée qui nous intéresse spécifiquement. Nous avons ainsi exposé les principes fondamentaux sous-tendant la conception et la découverte de médicaments qui favorisent l'utilisation des techniques de Conception de Médicaments Assistée par Ordinateur (CADD). Ces techniques se révèlent particulièrement précieuses pour le criblage des bases de données de molécules, la caractérisation et l'analyse des phénomènes et mécanismes biomoléculaires, ainsi que pour la prédiction et l'analyse de la biodisponibilité, des propriétés pharmacocinétiques et pharmacodynamiques des molécules. Leur utilité est indéniable, offrant des avantages notables en termes de rapidité et d'efficacité économique. Cependant, leur pertinence dépend du processus de recherche en cours, des objectifs visés, ainsi que de la disponibilité des données et des informations sur le récepteur biologique étudié. De plus, la stratégie de thérapie ciblée que nous proposons dans cette thèse poursuit deux objectifs distincts : la conception d'inhibiteurs spécifiques du récepteur tyrosine kinase de la protéine c-Met et du récepteur tyrosine kinase du facteur de croissance épidermique EGFR pour le traitement du cancer, ainsi que l'inhibition allostérique de la protéase MALT1 et l'inhibition non-covalente de la protéase $3CL^{pro}$ pour lutter contre l'infection virale du COVID-19. Pour atteindre ces objectifs, nous nous efforçons d'améliorer et d'intégrer les approches traditionnelles de CADD avec les techniques d'apprentissage automatique, d'intelligence artificielle et de simulation bio-informatique. Dans cette optique, ces stratégies combinent l'utilisation de molécules bioactives et de composés issus de plantes naturelles, agissant à la fois comme des modèles

structurels pour la conception et l'optimisation de nouveaux agents thérapeutiques potentiellement efficaces contre le cancer et les infections virales. Dans le prochain chapitre, nous approfondirons les méthodes de modélisation biostatistique et bio-informatique que nous avons utilisée pour développer un modèle de CADD hybride. L'objectif principal de cette démarche est de rationaliser la conception de nouveaux médicaments grâce à des approches computationnelles avancées.

CHAPITRE II : FUSION DE LA MODÉLISATION BIMOLÉCULAIRE DANS LA STRATÉGIE HYBRIDE DE CADD : MÉTHODOLOGIES, IMPLÉMENTATION ET APPLICATIONS THÉRAPEUTIQUES CIBLÉES

1 Introduction

Les méthodes de CADD sont largement utilisées dans la découverte et l'optimisation de nouveaux médicaments, mais il n'existe pas de stratégie uniforme applicable à tous les cas. Ces techniques peuvent être utilisées pour aider les chercheurs à chaque étape du développement et de la découverte d'un médicament. Afin de rationaliser le processus de développement des médicaments et d'atteindre des stratégies thérapeutiques sûres et efficaces, nous proposons et implémentons des processus systématiques hybrides de CADD pour optimiser à la fois les bases de données moléculaires et leur criblage virtuel. Ces processus peuvent être utilisés pour sélectionner des candidats médicaments avant de passer aux essais de validation préclinique et clinique. Pour développer cette démarche de criblage hybride qui simule le processus de conception et de découverte d'un nouveau médicament, nous combinons des méthodes de LBDD, d'intelligence artificielle (IA) et de machine Learning (ML) avec des méthodes de SBDD et des simulations biophysique. Dans ce chapitre, nous détaillerons la méthodologie hybride CADD que nous avons appliquée dans nos travaux pour repérer des médicaments candidats et aborder divers aspects du processus de découverte de médicaments. Nous présenterons également une description détaillée des outils de simulation et de modélisation biomoléculaire employés et comment les appliquer.

2 Ajustement des bases de données moléculaires ciblées

2.1 Source, homogénéité et qualité de donnée

Traditionnellement, pour obtenir une molécule "hit/ tête active" contre n'importe quelle cible dans un pipeline de découverte de médicaments, des données sur l'activité biologique et la structure des composés qui peuvent agir comme inhibiteur ou activateur sont nécessaires pour la conception rationnelle d'un médicament. La méthode de criblage virtuel à haut débit (HTVS) utilise des bases de données chimiques stockées dans des chimiothèques virtuelles contenant des millions de composés ($>10^6$) pour rechercher des composés potentiels à synthétiser [181]. Il y a de nombreuses chimiothèques disponibles en version gratuite ou commerciale contenant des collections de substances bioactives et des produits naturels [95]. Le **Tableau II.1** présente certains exemples de bases de données disponibles gratuitement et commercialement, à partir desquelles il est possible de consulter des données sur la structure et l'activité biologique de plusieurs composés. Ces bases de données hétérogènes contiennent les structures de molécules et leurs activités vis-à-vis des évaluations biologiques réalisées précédemment (littérature et/ou brevets) [182]. Le criblage virtuel de ces bases de données aide à identifier de nouvelles structures chimiques adaptées aux nouvelles cibles

biologiques ; il est également utile pour fournir des informations sur la corrélation entre un composé chimique et ses diverses fonctions biologiques possibles. Cependant, ces propositions exigent forcément la réalisation et la mise en œuvre à la fois de la synthèse rétrospective et des évaluations et essais biologiques parallèles. Cela peut poser divers problèmes, notamment : (i) l'obtention de produits impurs (mélange racémique) lors de la synthèse, (ii) des difficultés pour identifier la contribution de chaque homologue (mélange équimolaire des deux énantiomères) à l'activité prospective observée, (iii) l'absence d'activité biologique contre la cible biologique visée, etc. [94]. D'autre part, il existe d'autres limitations liées aux algorithmes de Docking moléculaire utilisés dans le criblage virtuel classique, qui fonctionnent généralement à l'échelle de millions de molécules seulement [183]. Par conséquent, la création éventuelle de chimiothèques volumineuses contenant plus de 100 milliards de molécules constituera un véritable défi pour les méthodes actuelles de criblage virtuel [184].

Tableau II.1. Exemples des bases de données virtuelles (chimiothèques).

Classe de substances	Chimiothèque	Nombre de Composés	Accès	Site web
Bioactive	DrugBank	>500 000	Gratuit	Lien
	NR-DBIND	>15 116		Lien
	DrugCentral	>500 000		Lien
	GLASS	>342,539		Lien
	CARLSBAD	~1,50 millions		Lien
	ChEMBL	~1,97 millions		Lien
	ZINC15	~750 millions		Lien
	PubChem	~113 millions		Lien
Produits naturels	The Natural Products Atlas	>25 500		Lien
	ANPDB	>6 515		Lien
	AfroDB	>1 000		Lien
	AntiBase	>53 000	Commercial	Lien
	Dictionary of Natural Products	>340 000		Lien
	Reaxys	>220,000		Lien
	MarinLit	>39500		Lien
Hybride	Maybridge	>30 000		Lien
	ChemSpider	115 millions		Lien
	Life Chemicals	>490 000		Lien
	Specs	>350 000		Lien
	Asinex	>260 000		Lien
	Enamine	~2,7 millions		Lien
	ChemDiv	~1,6 millions		Lien

En outre, la génération rapide de résultats entraîne de nombreuses hypothèses et estimations qui réduisent la précision des algorithmes adoptés dans le criblage virtuel rapide. De l'autre côté de l'esprit, les approches informatiques lentes telles que les simulateurs de dynamique moléculaire et la mécanique moléculaire fournissent des modèles plus précis et plus réalistes pour interpréter les interactions du niveau des complexes récepteur-ligand, mais elles sont plus lentes et nécessitent un ordinateur sophistiqué. Les approches rapides et lentes sont toutes deux utiles et permettent de résoudre différents aspects aux premiers stades de la découverte de médicaments, mais la sélection de la base de données moléculaire reste l'étape la plus importante avant de commencer toute évaluation *in silico*, *in vitro*, et *in vivo*.

2.2 Classification et critères de sélection des chimiothèques étudiées dans ce travail

2.2.1 Homogénéité des chimiothèques

Sans égard à l'importance des données hétérogènes disponibles dans les chimiothèques virtuelles dans la conception et le repositionnement des molécules et des médicaments, il est important de sélectionner des données de haute qualité et de faible incertitude afin de réduire les barres d'erreurs expérimentales potentielles lors de la conception des médicaments [94], [185]. Plus précisément, l'homogénéité des données en termes des structures moléculaires et leurs activités biologiques vis-à-vis de la cible biologique visée (biomarqueurs) est primordiale pour analyser logiquement les relations structure-activité et procéder notamment à des criblages *in silico*, *in vitro*, et *in vivo*. Par conséquent, afin de limiter éventuellement la marge d'erreur lors du criblage (screening) de molécules ciblées, il est crucial de répertorier des données homogènes obtenues dans des conditions de laboratoire et des essais identiques lors de l'évaluation des propriétés/activités des molécules en présence d'un médicament de référence. Sur ces bases, dans cette thèse, nous utilisons des chimiothèques homogènes de petites molécules biologiquement actives (Synthétisé ou isolé à partir des plantes), qui partagent des structures communes, pour commencer le criblage virtuel et la modélisation biomoléculaire afin d'identifier, de concevoir, et d'optimiser des nouveaux médicaments computationnels.

2.2.2 Critères de sélection des chimiothèques

Le schéma CADD hybride proposé dans cette thèse vise à faciliter le criblage computationnel de chimiothèques homogènes avec une grande précision et une grande fiabilité, quelle que soit leur dimension. En accord avec l'approche de la thérapie ciblée, qui est généralement basée sur le mécanisme d'interaction/affinité entre la conformation bioactive du ligand et le récepteur biologique ciblé. Ce mécanisme repose donc sur deux critères fondamentaux : (1) un critère structure-activité pour les ligands et (2) le critère de la structure du récepteur biologique ciblé. Ainsi, la disponibilité de ces deux critères permet de structurer une stratégie de CADD cohérente. En revanche, l'absence d'un critère et la présence de l'autre nécessitent des méthodologies spécifiques pour rationaliser la mise en œuvre du concept de CADD. Dans ce contexte, l'intervention des méthodologies LBDD et SBDD sert de médiateur à la construction et au développement de stratégies logistiques intermédiaires en CADD. En dépit de cela, la grande hétérogénéité des méthodes et des outils bio-informatiques utilisés dans ces approches crée une certaine confusion dans le choix de celles qui sont les plus appropriées en termes de performances, de robustesse et de fiabilité pour produire des prédictions et des hypothèses en termes de pharmacopée. Pour cela, dans cette phase nous proposons des stratégies de CADD hybrides adaptées en fonction de la présence ou de l'absence de la structure d'activité des ligands et de la structure du récepteur biologique impliqué dans la pathologie ciblée.

2.2.2.1 Critères basés sur la structure et d'activité du ligand

La méthodologie CADD hybride proposée dans ce cadre vise à faciliter le criblage numérique de chimiothèques contenant des informations relatives à la structure et à l'activité biologique des ligands. Dans cette optique, l'exploration des relations structure-activité (SAR) des molécules ciblées peut contribuer considérablement à l'optimisation des structures moléculaires candidates pour la conception et le développement de médicaments *in silico*. Pour atteindre cet aspect, dans cette thèse, nous portons nos réflexions sur l'application des techniques de QSAR multidimensionnelles comme support pour explorer et décoder avec précision les corrélations structure-activité. La diversité structurelle et les variations d'activité bibliologique observées *in vitro* pour les ligands sont utilisées comme critère de base pour sélectionner des bases de données de molécules cibles et développer des modèles QSAR entraînés à l'aide de techniques d'analyse de données telles que la RLM, PLS, RNLM, ANN, PCA, K-Means, etc. Les modèles QSAR peuvent ensuite prédire des valeurs d'activité IC_{50} précises à partir de faibles valeurs observées *in vitro* (µM/nM), en exprimant les données selon l'échelle logarithmique pIC_{50} (-Log IC_{50}) [94]. Cette approche basée sur LBDD vise à rationaliser le processus de découverte et de conception de nouvelles molécules à partir des données disponibles et validées expérimentalement concernant la structure et l'activité de bases de données moléculaires des ligands spécifiques. Ceci après avoir décodé les informations retenues au niveau structurel en informations à caractère numérique/graphique susceptibles d'être interprétées et modalisées.

2.2.2.2 Critère basé uniquement sur la structure du ligand et du récepteur

Pour examiner une base de données moléculaires *in silico* dans laquelle des informations concernant la structure 3D de la cible biologique et des molécules sont disponibles et les activités biologiques observées *in vitro* ne sont pas disponibles, il convient dans ce cas d'orienter la démarche vers l'application de techniques basées sur l'approche SBDD. Dans cette optique, nous avons intégré dans cette thèse l'approche du Docking moléculaire comme base de départ pour l'application de la stratégie SBDD.

3 Stratégie globale des modèles de CADD hybrides proposés dans ce travail

La **Figure II.1** illustre la stratégie globale de conception de médicaments assisté par l'ordinateur adoptée dans le cadre de préparation de cette thèse.

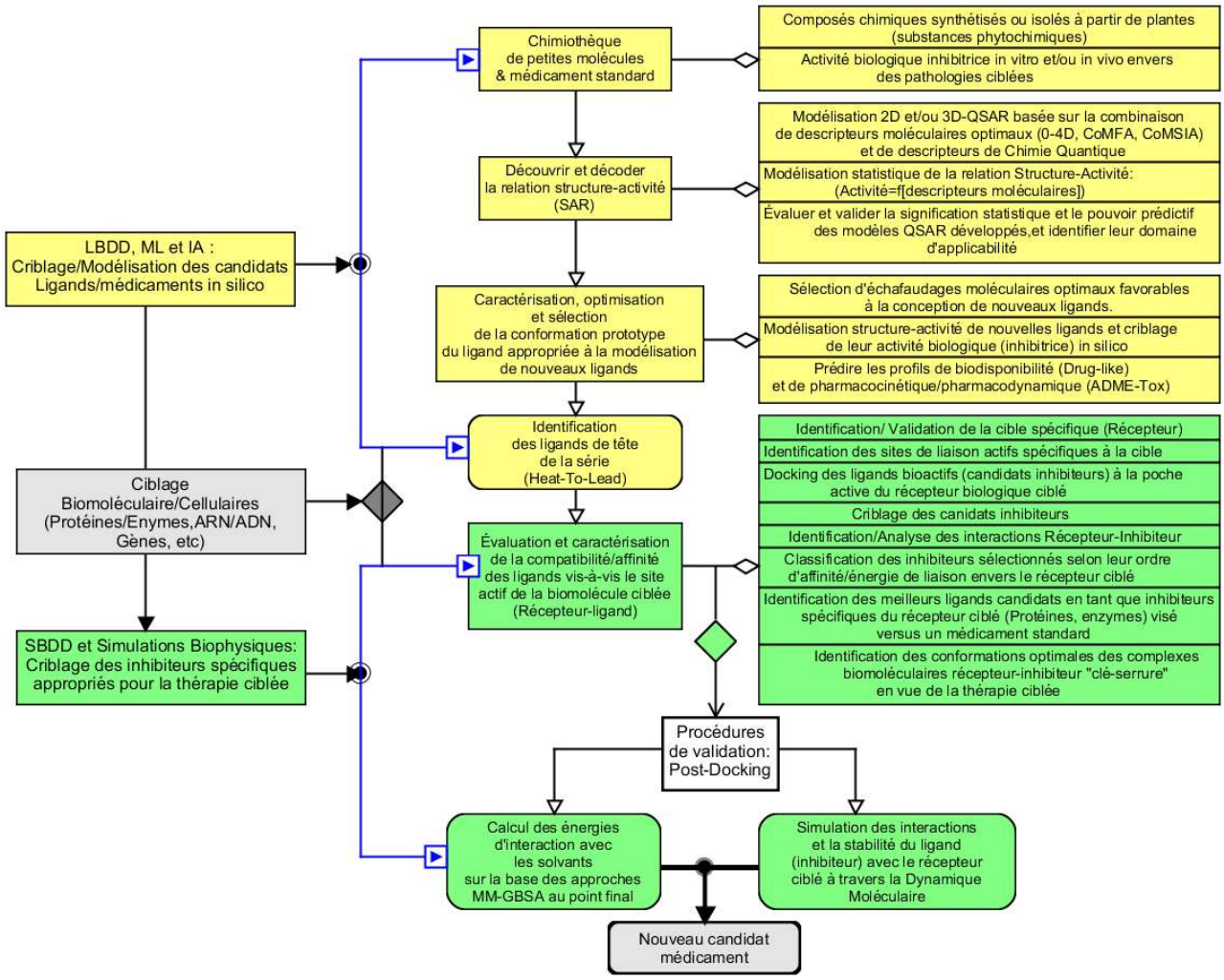


Figure II.1. Schéma général de la stratégie hybride de CADD conçue et appliquée dans cette thèse afin de procéder à la conception et à l'optimisation computationnelles de nouveaux médicaments.

La sélection de bases des chimiothèques fondées sur des critères de structure-activité va rendre le processus de criblage virtuel par la combinaison de LBDD et SBDD plus rationnel, ce qui conduira à de nouvelles biomolécules présentant une forte activité biologique et des propriétés structurelles favorables à la conception de médicaments et à leur application dans des thérapies ciblées contre des maladies ciblées. L'adoption du critère d'homogénéité structurelle en présence ou en l'absence d'activité biologique des chimiothèques servent de point de départ à la mise en œuvre de la méthodologie CADD hybride mise en œuvre dans cette thèse selon chaque contexte. D'autres aspects de la méthodologie hybride CADD que nous avons proposée servent aussi à évaluer *in silico* les propriétés ADME-Tox et Drug-like des composés originales et conçus, ainsi qu'à évaluer la compatibilité et la stabilité des cibles biologiques vis-à-vis des ligands candidats à la découverte de nouveaux médicaments. Dans cette optique, nous appliquons une démarche sophistiquée de criblage virtuel basée sur le Docking moléculaire et sur des simulations biophysiques avancées. Le reste de ce chapitre décrit les méthodes utilisées pour développer cette stratégie de CADD hybride, incluant des approches QSAR multidimensionnelles basées sur des descripteurs moléculaires hybrides, des techniques de modélisation bio-informatique et biostatistique, la modélisation pharmacocinétique/pharmacodynamique, le criblage virtuel basé sur le Docking moléculaire, et des

procédures de validation après Docking (post-Docking) telles que des simulations MM-GBSA et de la Dynamique Moléculaire.

3.1 Analyse des données par voie bio-informatique et biostatistique

La bio-informatique et la biostatistique sont deux disciplines interdisciplinaires qui jouent un rôle essentiel dans la recherche biomédicale actuelle. Bien qu'elles aient des similitudes, elles ont des spécialités et des méthodes distinctes. La bio-informatique est le processus d'utilisation d'algorithmes et de logiciels informatiques pour analyser, manipuler et interpréter des données biologiques [186]. Cela peut inclure l'analyse de fragments d'ADN, de protéines, de profils d'expression génétique et de voies métaboliques. La bio-informatique est également importante pour développer et découvrir des médicaments, vaccins et autres méthodes de diagnostic [187]. D'autre part, la biostatistique utilise des méthodes statistiques pour examiner les données biologiques et médicales [188]. Cela peut inclure la conception et l'analyse d'expériences et d'études prospectives pour rechercher des modèles et des relations dans les bases de données. La biostatistique est importante pour le développement, l'évaluation et le contrôle des traitements médicaux, ainsi que pour l'évaluation et la prédiction de la tolérance et de l'efficacité des médicaments approuvés ou en cours de développement, et pour identifier les risques potentiels de toxicité des agents pharmacologiques proposés [189], [190]. Ces deux domaines nécessitent une solide connaissance en biologie, mathématiques et informatique pour être adéquatement utilisés. Ils sont de plus en plus importants avec l'essor des données massives et la médecine de précision, car ils permettent aux chercheurs de découvrir des informations à partir d'ensembles de données biologiques et de prendre des décisions et des hypothèses rationnelles sur la base de leur analyse précise.

3.2 Contexte d'application

Dans ce contexte, les techniques d'apprentissage automatique et d'intelligence artificielle sont de plus en plus utilisées dans les analyses bio-informatiques et biostatistiques pour améliorer la précision et l'efficacité de diverses tâches d'analyse [191]. Parmi ces techniques, les algorithmes d'apprentissage supervisé et non supervisé sont utilisés pour découvrir des modèles et des corrélations dans les données et pour développer des modèles mathématiques prédictifs [192]. Les réseaux neuronaux peuvent être utilisés pour des tâches de classification et de prédiction, comme la détection de maladies à partir d'images médicales ou de données numériques [193]. Les machines à vecteurs de support peuvent être utilisées pour collecter et classer des données biologiques [194]. L'apprentissage profond est appliqué à la prédiction de la structure des protéines et à un large éventail d'autres applications [195]. Par conséquent, l'utilisation de ces méthodes contribue fortement à l'optimisation et à la conception de nouveaux médicaments, ainsi qu'à l'amélioration des normes d'analyse biostatistique. C'est ce qui nous a amené, lors de la préparation de cette thèse, à adopter une démarche basée sur la bio-informatique et la biostatistique pour élaborer et proposer la voie hybride de la CADD. Ceci pour optimiser davantage le processus précoce de conception et de découverte de nouveaux médicaments candidats.

4 Modélisation biomoléculaire basée sur l'approches de QSAR multidimensionnelles

Les méthodes QSAR basées sur la biostatistique peuvent être considérées comme un raccourci logique du processus de découverte de médicaments, car ces méthodes permettent de réduire considérablement le gaspillage de l'approche traditionnelle, qui implique des synthèses complètes et des essais expérimentaux sur les composés [196]. L'objectif principal des méthodes QSAR multidimensionnelles est de renforcer la possibilité de découvrir des composés actifs à partir d'un ensemble de composés synthétisés et de minimiser les efforts en termes de temps et de ressources afin que la découverte soit la plus fructueuse possible [197]. La stratégie de relation quantitative structure-activité (QSAR) associe l'activité biologique d'un ligand à sa structure moléculaire. En d'autres termes, les changements apportés à la structure d'une molécule peuvent avoir un impact positif ou négatif sur son activité [95].

4.1 Principes fondamentaux des méthodes QSAR

Le concept QSAR repose fondamentalement sur le modèle mathématique de Hansch défini par l'**Eq.(II.1)**, également connu en tant qu'analyse de Hansch, qui permet de prédire l'activité biologique d'un composé chimique sur la base de ses propriétés physiques et chimiques [198], [199]. L'équation de Hansch a été développée par Corwin Hansch et ses collaborateurs dans les années 1960 et a été largement utilisée en chimie médicinale, en toxicologie chimique et en chimie de l'environnement. L'équation de Hansch relie le logarithme de l'activité biologique d'un composé (comme sa capacité de liaison avec un récepteur ou une enzyme) au logarithme du coefficient de partage (P), qui est une mesure de son hydrophobicité, et au logarithme de ses propriétés électroniques et stériques (σ et π, respectivement) [200], [201]:

$$\mathrm{Log}\left(\frac{1}{IC_{50}}\right) = a_0 + a_1 \mathrm{LogP} + a_2\sigma + a_3\pi \qquad \text{(II.1)}$$

où IC_{50} est la concentration nécessaire pour inhiber 50 % de l'activité biologique enzymatique de la cible, a_0 est une constante non physico-chimique, et a_1, a_2 et a_3 sont les coefficients de régression correspondant à la contribution de chaque propriété physico-chimique sur l'activité biologique. Des méthodes telles que FreeWilson et Topliss ont été développées pour optimiser les propriétés structurelles des molécules en utilisant l'analyse de Hansch, qui repose sur l'hypothèse que les substituants de la structure moléculaire originale contribuent à son activité biologique [202]. L'équation de Hansch a été modifiée pour prendre en compte d'autres facteurs tels que la liaison hydrogène, l'ionisation et la conformation, afin d'améliorer son pouvoir prédictif dans la recherche de médicaments. Cette méthode a été combinée avec d'autres approches computationnelles telles que le Docking Moléculaire et l'analyse QSAR multidimensionnelle pour améliorer la conception de nouveaux composés dotés de propriétés biologiques souhaitables. L'analyse de Hansch est un outil important pour la découverte de médicaments et l'évaluation des risques chimiques, car elle aide à mieux comprendre les relations structure-activité des molécules complexes.

4.2 Principes de l'OCDE relatifs à la validité des modèles QSAR/QSPR

Conformément à la progression des méthodes QSAR/QSPR, l'Organisation de Coopération et de Développement Économiques (OCDE) a défini six principes pour valider les modèles QSAR/QSPR et évaluer leur applicabilité dans le cadre du plan expérimental [203], [204]. A cet égard, les six principes de l'OCDE sont les suivants : Adéquation, Robustesse, Précision, Fiabilité, Transparence et Prévisibilité [205]. C'est ce que nous visons à atteindre grâce à la méthodologie QSAR que nous avons adoptée dans le cadre de cette thèse. (1) L'adéquation désigne la capacité du modèle à décrire les phénomènes biologiques ou chimiques liés aux données sur lesquelles il est basé**.** (2) La robustesse désigne la capacité du modèle à être facilement reproductible et à tenir compte des variations des données. (3) La précision désigne la capacité du modèle à prédire correctement le résultat pour un ensemble spécifique de variables d'entrée. (4) La fiabilité désigne la capacité du modèle à prédire le résultat correct pour différents ensembles de variables d'entrée. (5) La transparence désigne la capacité à interpréter le modèle et la traçabilité de ses prédictions. (6) La prédictibilité est la capacité du modèle à prédire le résultat même lorsque de nouvelles données sont introduites.

4.3 Application et extension des approches QSAR

Les approches QSAR sont largement appliqués dans le cadre de la démarche LBDD [206]. La modélisation QSAR repose sur le principe que l'activité biologique d'une molécule peut être mesurée à partir de ses descripteurs moléculaires tels que la taille, la forme, la polarisabilité, la capacité de créer des liaisons hydrogène et les propriétés électroniques, etc. ou des caractéristiques moléculaires telles que la couleur, l'odeur, le point de fusion, le point d'ébullition, la solubilité et la réactivité, etc. [202]. Pour générer la relation structure-activité (SAR) et la traduire en un modèle mathématique (Activité=f(descripteurs/caractéristiques)), il est nécessaire de collecter un vaste ensemble de données pour l'apprentissage [94]. Sur la base des données de structure et d'activité biologique, des modèles QSAR sont développés en utilisant l'activité biologique de ligands connus afin de concevoir de nouveaux composés et de prédire leurs activités. La régression linéaire et non linéaire multiple (RLM et RNLM), la régression en composantes principales (ACP), la régression des moindres carrés partiels (PLS), la régression de type ML, les réseaux neuronaux artificiels (RNA), etc. sont autant de techniques servant à développer des modèles QSAR. La principale différence entre les modèles QSAR et le modèle pharmacophore est que le QSAR ne considère pas seulement les caractéristiques du ligand, mais aussi les caractéristiques associées à l'activité biologique du ligand et de son récepteur ciblé. De ce fait, l'application des méthodologies QSAR au criblage des bases de données de molécules bioactives, la conception de nouvelles molécules et l'amélioration de leurs activités biologiques et leurs sélectivités vis-à-vis d'une cible biologique spécifique offrent des perspectives plus rationnelles pour la conception et la découverte d'un nouveau médicament. Les études QSAR offrent plusieurs avantages qui en font un véritable outil de recherche de nouveaux médicaments. Grâce à leurs prédictions précises de l'activité biologique, de la toxicité et des caractéristiques structurelles d'un composé, elles peuvent aider à prendre des décisions sur l'utilisation sûre d'un composé et guider la conception de nouveaux composés avant de procéder à la synthèse et à l'évaluation *in vitro* ou *in vivo* [207]. En outre,

ces études peuvent être utilisées pour analyser rapidement et efficacement un grand nombre de composés, ce qui permet de gagner du temps et de réduire les coûts. Les études QSAR peuvent également donner un aperçu sur les mécanismes moléculaires de l'action pharmacologique, ce qui contribue à la mise au point de nouveaux médicaments.

4.4 Classification des Modèles QSAR par dimension

La modélisation QSAR (Relation quantitative structure-activité), QSPR (Relation quantitative structure-propriété), QSTR (Relations quantitatives structure-toxicité), etc. est utilisée pour évaluer et prédire les propriétés d'un ligand en fonction de ses propriétés structurelles. Les méthodes QSAR sont classées en fonction du nombre et du type de descripteurs utilisés (0D à 6D) [208], ainsi que des critères d'évaluation tels que l'activité, la propriété, la toxicité, etc. Ces méthodes sont appliquées à des domaines spécifiques, tels que l'industrie pharmaceutique pour QSAR, l'industrie pharmaceutique et énergétique pour QSPR, et le développement de pesticides pour QSTR. Dans le cadre de nos efforts de recherche axés sur la découverte de nouveaux médicaments pour lutter contre le cancer et les infections virales induites par COVID-19 comme modèles d'application, nous utilisons l'approche QSAR pour cribler des chimiothèques, modéliser et concevoir de nouveaux médicaments prometteurs. Le **Tableau II.2** décrit les approches QSAR selon leurs dimensions. Les méthodes de modélisation 2D-QSAR et 3D-QSAR sont les plus appropriées pour la conception de médicaments basée sur le ligand [209]. La modélisation 2D-QSAR prend en compte des caractéristiques géométriques et topologiques, mais pas la conformation 3D de la structure moléculaire. La modélisation 3D-QSAR intègre des méthodes plus complexes, ce qui lui permet de capturer les interactions subtiles entre les atomes dans l'espace tridimensionnel [29]. Les modèles 2D-QSAR sont simples à construire et à interpréter, tandis que les modèles 3D-QSAR sont plus précis dans la prédiction de l'activité. Les méthodes 4D, 5D et 6D-QSAR sont plus complexes et plus susceptibles de provoquer un mauvais ajustement et une surinterprétation des données, et sont donc moins utilisées [210]. Dans le cadre de nos travaux, les approches 2D et 3D-QSAR sont adoptées pour explorer et décoder la corrélation structure-activité des ligands bioactifs et pour appliquer l'approche LBDD.

Tableau II.2. Classification de la modélisation QSAR par dimensions.

Dimension	Approche	Cible biologique	Référence
0D-QSAR	Un seul descripteur tel que le poids moléculaire, le LogP, la liaison hydrogène, le nombre de liaisons rotatives, etc. est utilisé pour construire un modèle mathématique qui peut être utilisé pour prédire l'activité biologique d'une molécule spécifique.	No	[211]
1D-QSAR	Contrairement à la modélisation 0D-QSAR, 1D-QSAR est basée sur l'utilisation de plusieurs descripteurs (poids moléculaire, indices topologiques et descripteurs de groupes fonctionnels, etc.) pour représenter la structure chimique d'une molécule en 1D et pour élaborer un modèle mathématique capable de prédire l'activité biologique d'une molécule donnée.	No	[209]
2D-QSAR	Utiliser une représentation bidimensionnelle d'une molécule pour saisir ses attributs/propriétés structurels (descripteurs moléculaires) tels que les coordonnées atomiques, les longueurs de liaison, les propriétés électroniques/physico-chimiques. Ces descripteurs sont utilisés pour générer une carte topologique en 2D de la molécule. Cette carte est ensuite utilisée pour construire un modèle mathématique qui peut être utilisé pour prédire l'activité biologique d'une molécule donnée.	No	[212]
3D-QSAR	Cette modélisation repose sur l'utilisation d'une combinaison de descripteurs moléculaires 3D et de méthodes statistiques pour corréler l'activité biologique d'un composé avec sa structure 3D. Elle sert à comprendre la base moléculaire de l'activité biologique, à prédire l'activité biologique de nouvelles molécules et à rationaliser leurs propriétés structurelles favorables à l'affinité et à l'interaction avec la cible biologique (protéine/enzyme).	Possible	[213]
4D-QSAR	4D-QSAR est une extension de la modélisation 3D-QSAR, dans laquelle le temps de simulation est prolongé et les ligands sont représentés comme un ensemble de configurations, ce qui permet de représenter de manière plus approfondie la nature dynamique des effets biologiques.	Oui	[214]
5D-QSAR	Extension du 4D-QSAR reposant sur la mise en place et l'analyse de la relation entre l'activité biologique et divers descripteurs tels que la forme moléculaire, le potentiel électrostatique, l'hydrophobicité, les propriétés des donneurs et des accepteurs de liaisons hydrogène et d'autres propriétés. Sur la base de ce mélange de descripteurs 5D, une équation mathématique est générée pour prédire l'activité biologique de la molécule.	Oui	[215]
6D-QSAR	Extension du 5D-QSAR avec intégration de différents paramètres de solvatation. Les modèles 6D-QSAR générés peuvent être appliqués pour identifier de nouvelles molécules dont l'activité biologique est similaire à celle des originales.	Oui	[216]

4.5 Méthodologie de modélisation QSAR appliquée dans cette thèse

Les techniques de QSAR (Relation Quantitative Structure-Activité) allant de 0D à 6D ont été fondamentales pour aider à la conception et à la découverte de médicaments basées sur des structures moléculaires et leur corrélation avec la réponse biologique. Chaque dimension a été développée pour surmonter les limites des dimensions précédentes, en fonction de l'objectif de l'application [217]. 2D-QSAR et la 3D-QSAR sont les techniques les plus couramment utilisées pour l'optimisation des structures moléculaires et l'identification de nouvelles molécules candidates pour le développement de médicaments. Alors que la 2D-QSAR ne nécessite pas l'alignement des structures des molécules et est plus rapide que la 3D-QSAR, cette dernière est plus précise en termes de paramètres biologiques des molécules. La modélisation 2D-QSAR présente l'avantage d'une évaluation rapide de l'activité/toxicité et la biodisponibilité de nombreux composés, en raison de certaines propriétés ADME-Tox étant liées aux propriétés globales de la molécule. Pendant ce temps, la 3D-QSAR est plus efficace

pour prédire les interactions des petites molécules avec les récepteurs biologique (biomarqueurs) et fournir des informations plus précises sur la structure des molécules et leur activité biologique. Pour cette raison, des investigations 2D-QSAR et 3D-QSAR sont menées dans cette thèse pour générer des modèles pharmacophores fiables dans les espaces 2D et 3D, fournissant ainsi un aperçu des propriétés structurales essentielles nécessaires pour obtenir une forte activité biologique contre une cible thérapeutique ciblée. Dans le cadre de la préparation de cette thèse, tous les modèles QSAR que nous avons développés se sont renforcés en termes de précision et de pouvoir prédictif grâce au couplage avec d'autres analyses *in silico* basées sur la modélisation des type (Drug-like, ADME-Tox, Docking moléculaire, MM-GBSA, et la Dynamique moléculaire). Par la suite, nous abordons les méthodologies, les techniques et les outils que nous avons adoptés pour construire des modèles 2D-QSAR et 3D-QSAR afin de caractériser la relation structure-activité des molécules. Ensuite, étape par étape, nous abordons chaque phase de la construction, de l'entraînement et du test de ces modèles. Pour ce faire, nous décrivons les descripteurs moléculaires sur lesquels nous avons basé pour modéliser la relation structure-activité à traverse des études 2D et 3D-QSAR, ainsi que les méthodes de calcul et de sélection des descripteurs moléculaires candidats pour le développement des modèles QSAR, les méthodes de division de la base de données générée et les méthodes statistiques utilisées pour valider la performance et la fiabilité des modèles QSAR proposés.

4.5.1 Méthodologie 2D-QSAR

La **Figure II.2** présente un schéma descriptif de la méthodologie générale de la modélisation 2D-QSAR que nous avons adoptée dans le cadre de la préparation de cette thèse.

La première étape de la mise en œuvre du modèle 2D-QSAR est la sélection des données relatives à la structure et à l'activité des molécules ciblées issues expérimentalement. Idéalement, les données doivent répondre aux critères de diversité structurelle et de différence d'activité biologique entre les molécules étudiées. Par ailleurs, plus le nombre de données est élevé, plus les résultats sont bénéfiques, mais il est possible de commencer raisonnablement par plus de 20 échantillons. Une fois les données structure-activité collectées, nous devons les raffiner. Pour cela, nous excluons les échantillons qui ne sont pas biologiquement actifs vis-à-vis de la cible biologique visée et nous ne conservons que les molécules ayant une activité biologique vis-à-vis de la cible visée. Puis, nous dessinons et optimisons soigneusement les géométries et les énergies des structures 2D de chaque molécule. Ceci en vue de définir les structures les plus stables et les plus appropriées pour extraire les informations chimiques encodées dans les représentations symboliques des molécules (descripteurs moléculaires). Après avoir transformé les données structure-activité en données descripteurs-activité, c'est au tour de l'optimisation de cette nouvelle base de données numérique. Pour ce faire, les données numériques sont analysées par le biais de l'analyse en composantes principales (ACP). Il s'agit de réduire la taille des données massives des descripteurs moléculaires calculés à une dimension raisonnable à partir de laquelle les descripteurs candidats pour décrire l'activité ou la propriété étudiée peuvent être identifiés. Une fois que le nombre de descripteurs moléculaires calculés est réduit, nous préparons la nouvelle base de données pour développer le modèle 2D-QSAR. Pour ce faire, nous divisons ces données en deux sous-ensembles, l'ensemble d'entraînement (comprenant 4/5 de la

totalité des échantillons moléculaires) et l'ensemble de test (contenant 1/5 de la totalité des échantillons moléculaires). Au cours de l'étape de division de la base de données dans tous nos travaux, nous avons adopté les techniques de classification K-means ou univarié (UV) plutôt que d'utiliser une division aléatoire vague. L'ensemble d'entraînement est utilisé pour développer le modèle 2D-QSAR et le mettre en œuvre pour sélectionner les descripteurs moléculaires les plus appropriés pour décrire l'activité ou à la propriété biologique étudiée. Alors que l'ensemble de test est utilisé dans le cadre du test externe relatif au pouvoir prédictif du modèle QSAR sélectionné. Nous appliquons l'algorithme de Régression Linéaire Multiple (RLM) sur l'ensemble d'entraînement plusieurs fois, jusqu'à ce que nous obtenions un modèle qui a une signification statistique fiable et qui est capable de corréler l'activité biologique (variable dépendante) à un minimum de descripteurs moléculaires (variables indépendantes) qui peuvent expliquer la relation entre la structure et l'activité.

Pour optimiser la pertinence des descripteurs moléculaires sélectionnés par le biais du modèle linéaire multiple en vue d'interpréter et de modéliser la relation structure-activité, nous explorons la relation non linéaire qui relie la structure et l'activité des molécules étudiées. Pour ce faire, nous développons d'autres modèles 2D-QSAR par le biais de la Régression Non Linéaire Multiple (RNLM) et/ou de Réseaux de Neurones Artificiels (RNA ou ANN). Cette démarche a pour but de valider davantage la pertinence des descripteurs moléculaires sélectionnés par le biais du modèle linéaire dans l'interprétation de l'activité biologique des molécules étudiées. Également, l'analyse des corrélations linéaires et non linéaires entre les descripteurs moléculaires sélectionnés et l'activité biologique évaluée nous offre une vision plus profonde et plus complète pour décrire la relation structure-activité. Par conséquent, nous pouvons identifier le(s) modèle(s) QSAR le(s) plus approprié(s) pour faciliter le design de nouvelles molécules et/ou le criblage virtuel des bases de données disponibles.

Pour confirmer la validité des modèles 2D-QSAR (linéaires et non linéaires) proposés afin d'établir la corrélation entre les descripteurs moléculaires et l'activité biologique des molécules, nous appliquons à cet égard une série d'évaluations statistiques sur la base de tests (par exemple, principes de l'OCDE pour la validation des modèles QSAR, domaine d'applicabilité, test de la Y-Randomisation, etc). Si les modèles 2D-QSAR développés passent avec succès tous les tests, ils peuvent être utilisés pour comprendre les mécanismes et les effets des propriétés structurelles inhérentes à la structure 2D des molécules sur leur activité étudiée. Ces modèles peuvent également être exploités pour cribler des bases de données d'autres molécules exogènes qui portent le même échafaudage commun que les molécules originales, rationalisant ainsi la conception de nouvelles molécules et prédisant leurs activités biologiques avant même leur synthèse. Si ces modèles ne passent pas les tests internes, ils doivent se référer aux données des descripteurs moléculaires et y apporter de nouvelles optimisations. Ceci est basé sur l'intuition chimique et statistique, qui peut principalement inclure les estimations suivantes : (1) augmenter le nombre de cycles d'entraînement du modèle linéaire original, (2) augmenter le nombre de descripteurs moléculaires ou les réduire, (3) procéder à de nouvelles divisions pour les données générées, (4) changer l'ensemble d'entraînement et de test dans la construction de nouveaux modèles.

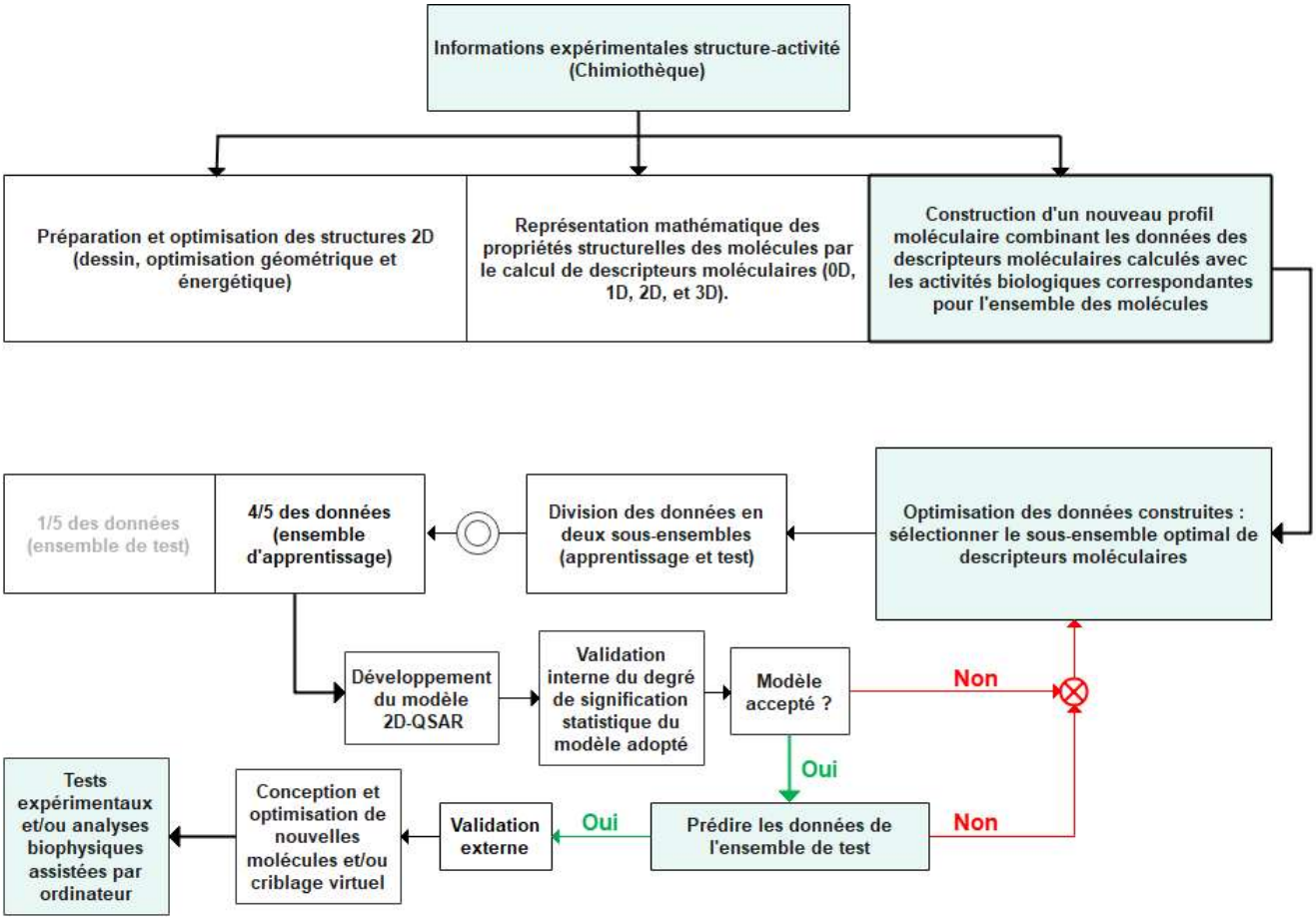


Figure II.2. Méthodologie générale appliquée pour la modélisation 2D-QSAR.

4.5.2 *Méthodologie 3D-QSAR*

La **Figure II.3** présente la méthodologie de la modélisation 3D-QSAR appliquée dans le cadre de cette thèse. Similaire à d'autres méthodologies de QSAR, 3D-QSAR commence par le traitement d'une série de molécules ayant des structures et des activités biologiques bien définies, où la première étape est l'alignement des structures moléculaires 3D à l'aide d'algorithmes d'alignement. Cette étape est suivie par le calcul du champ moléculaire autour de chaque molécule de l'ensemble d'apprentissage. Comme pour les étapes de développement du modèle 2D-QSAR décrites dans la partie précédente, des descripteurs moléculaires 4D (ou des descripteurs basés sur un réseau) sont utilisés pour caractériser les structures 3D des molécules lors de la construction du modèle 3D-QSAR. Ensuite, les points de champ calculés sont combinés pour prédire l'activité biologique par le biais des algorithmes statistiques. Pour ce faire, les valeurs calculées des descripteurs de champ sont corrélées avec les réponses biologiques grâce à l'utilisation d'une puissante méthode de régression linéaire telle que la méthode des moindres carrés partiels (PLS). Après avoir évalué la signification statistique interne du modèle 3D-QSAR développé, les calculs du champ moléculaire générés à partir de ce modèle sont appliqués pour prédire l'activité biologique des molécules de l'ensemble de test. Puis, le pouvoir prédictif externe du modèle développé est évalué. Pour renforcer la validité et la fiabilité de la méthodologie de modélisation 3D-QSAR utilisée dans cette thèse, nous avons combiné les approches de l'analyse comparative des champs moléculaires (CoMFA) avec l'analyse comparative des indices de similarité moléculaire (CoMSIA) pour sélectionner les meilleurs modèles susceptibles de couvrir

avec précision la relation structure-activité. Après la vérification de la signification statistique et du pouvoir prédictif des modèles 3D-QSAR générés par la technique PLS, les potentiels de champ moléculaire favorables et défavorables de la réponse biologique sont déterminés et représentés. Ces potentiels sont représentés par des cartes de contour figurant sur la structure de la molécule de référence ayant l'activité biologique la plus élevée observée *in vitro* dans la série de molécules étudiées. Les analyses CoMFA permettent de déterminer les potentiels électrostatiques (E) et stériques (S), tandis que les analyses CoMSIA décodent les potentiels électrostatiques (E), stériques (S) et hydrophobiques (H), ainsi que les liaisons hydrogène donneur et accepteur (LHD et LHA). Après avoir sélectionné les modèles 3D-QSAR établis sur la base des approches CoMFA et CoMSIA et identifié les champs potentiels favorables à l'amélioration/optimisation de l'activité biologique des molécules étudiées, ces modèles et leurs hypothèses pharmacophores sont appliqués à la conception de nouvelles molécules et à la prédiction de leurs activités biologiques et/ou au criblage d'autres bases de données moléculaires similaires à la base de données originale.

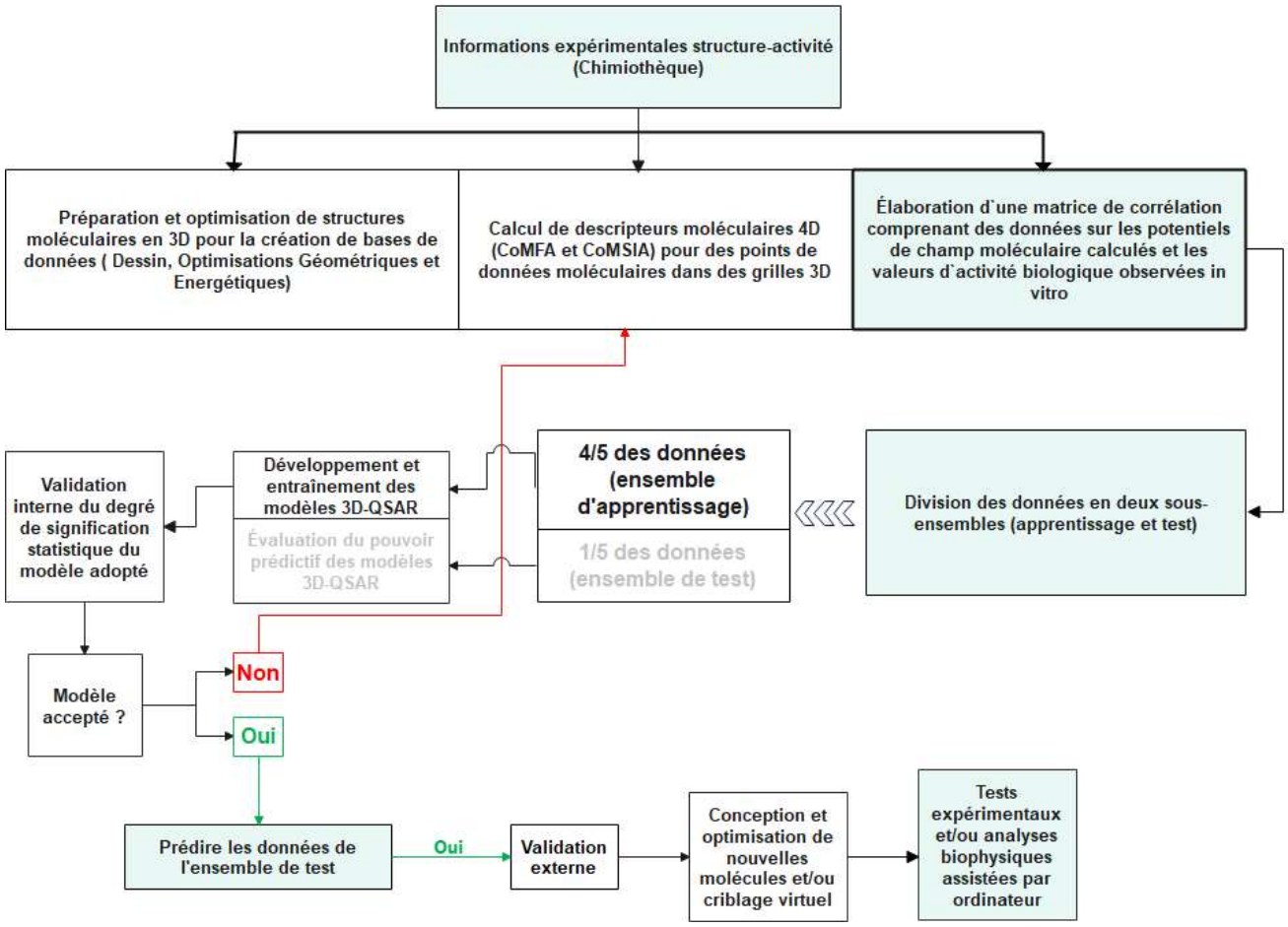


Figure II.3. Méthodologie générale appliquée pour la modélisation 3D-QSAR.

4.6 Descripteurs moléculaires pour la modélisation QSAR

L'application de méthodes d'apprentissage automatique à la modélisation QSAR/QSPR/QSTR joue un rôle important dans plusieurs secteurs de la chimie pharmaceutique, de la science des matériaux et de la toxicologie. Ces méthodes peuvent être utilisées pour prédire les propriétés des molécules et/ou concevoir de nouvelles molécules avec des propriétés souhaitées avant de passer aux étapes de la synthèse et des tests expérimentaux. Pour prédire les propriétés des molécules, les

structures moléculaires sont transformées en valeurs appelées descripteurs moléculaires qui peuvent être utilisés par les algorithmes d'apprentissage automatique [218]. Ces descripteurs moléculaires peuvent être représentés par un nombre, un vecteur, une matrice ou des données d'un autre format, telles que des chaînes de caractères " SMILES string "[219]. Le point fort des descripteurs moléculaires réside dans le fait qu'ils traduisent la constitution d'une molécule en un ensemble de valeurs quantitatives qui codent des aspects spécifiques de la structure moléculaire. La nature quantitative de ces descripteurs moléculaires permet de relier les informations théoriques issues de la structure moléculaire à certaines données expérimentales relatives au comportement de la molécule (propriété biologique/physicochimique). Ces descripteurs moléculaires sont donc devenus le support de plusieurs applications en bio-informatique et la modélisation QSAR/QSPR dans une perspective de descripteurs moléculaires, conformément au principe défini par **l'Eq (II.2)** [220].

$$\phi = f(x_1, x_2, x_3, \ldots, x_p) \qquad \text{(II.2)}$$

où Φ représente l'activité biologique/propriété physico-chimique de la molécule, liée à certaines propriétés structurelles, codées dans les descripteurs moléculaires p (x_1, ... x_p).

Dans cette thèse, nous utilisons les descripteurs moléculaires pour développer et entraîner des modèles d'apprentissage automatique basés sur l'approche QSAR afin de prédire l'activité biologique des molécules étudiées, de les cribler et de sélectionner les molécules leaders (les têtes de série les plus actives).

4.6.1 Descripteurs moléculaires exploités dans le cadre de cette thèse

La modélisation chimio-informatique qui repose sur les techniques de QSAR/QSPR consiste à déterminer la relation entre la structure moléculaire d'une molécule avec un point final spécifique (activité, propriété, toxicité...) [221]. Pour appliquer cette modélisation, la structure moléculaire doit être définie par un ensemble de paramètres quantitative (descripteurs moléculaires). Il existe de nombreuses représentations mathématiques de la structure moléculaire, à partir desquelles les paramètres quantitatifs des molécules peuvent être obtenus soit expérimentalement, soit théoriquement [222]. Les descripteurs moléculaires mesurés expérimentalement désignent les propriétés physiques et chimiques qui sont mesurées à l'aide d'une procédure expérimentale spécifique [223]. Tandis que les descripteurs théoriques sont calculés à partir d'algorithmes chimio-informatiques appliqués à une représentation moléculaire spécifique [222]. La classification des descripteurs moléculaires théoriques la plus courante consiste à les répartir en cinq catégories distinctes selon les dimensions (0D, 1D, 2D, 3D et 4D). La **Figure II.4** présente un exemple graphique de différentes représentations moléculaires de la même structure (exemple d'un dérivé de la pyridine) à partir desquelles les descripteurs moléculaires 0D, 1D, 2D, 3D et 4D peuvent être générés.

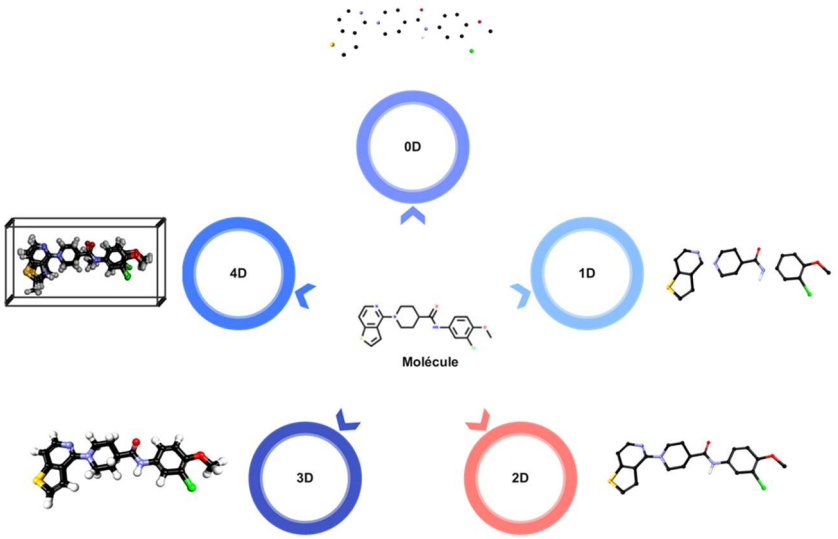


Figure II.4. Représentation graphique des différentes représentations moléculaires d'une même structure (un dérivé de la pyridine, à titre d'exemple).

Les molécules sont des entités superficiellement réelles [224]; il est donc essentiel de les représenter à un certain niveau d'abstraction, par le biais des procédures conventionnelles et des règles de base. Le choix de la méthode utilisée pour produire la représentation symbolique de cette abstraction a une incidence directe sur la quantité et le type d'informations chimiques contenues dans la structure de la molécule et, par conséquent, sur le type de descripteurs disponibles pour le calcul. Dans cette perspective, il est nécessaire de déterminer le niveau optimal de la complexité pour traiter le problème analysé. Par conséquent, du point de vue conceptuel, l'utilisation de représentations moléculaires de différents types de complexité, c'est-à-dire "la multi- dimensionnalité" (de 0D à 4D) (**Figure II.4**), peut fournir la plus grande quantité de données intrinsèques à la structure de la molécule. Ainsi, l'interprétation de l'activité biologique / des propriétés physico-chimiques basée sur les différentes données hybrides extraites de chaque dimension peuvent fournir des modèles prédictifs plus fiables et plus efficaces que les modèles basés sur chaque dimension séparément. Sur la base de ce contexte, nous avons adopté dans cette thèse une approche hybride dans la construction de modèles QSAR multidimensionnels en adoptant le mélange de descripteurs 0D, 1D, 2D et 3D dans le développement de modèles 2D-QSAR d'une part, et de descripteurs 4D basés sur CoMFA (Analyse Comparative des Champs Moléculaires) et CoMSIA (Analyse Comparative des Indices de Similitude Moléculaire) dans le développement de modèles 3D-QSAR d'autre part.

4.6.2 Classification par dimension des descripteurs moléculaires appliqués

D'un point de vue graphique, les représentations moléculaires de différents types (dimensions) et leur lien avec les descripteurs moléculaires générés à partir de chaque dimension peuvent être définis par les termes catégoriels suivants :

4.6.2.1 Descripteurs zéro-dimensionnels (0D)

La représentation la plus simple d'une molécule est sa formule (liste des différents éléments chimiques et de leur présence). La pyridine, par exemple, est représentée par C5H5N (**Figure II.5**).

Les descripteurs de ce type sont indépendants de toute connaissance sur la structure moléculaire ou la connectivité des atomes. En tant que tels, les descripteurs dérivés de la formule chimique sont connus sous le nom de descripteurs 0D et représentent des propriétés globales, telles que la masse exacte, le poids moléculaire, le m/z et l'analyse élémentaire [223]. Les propriétés constitutionnelles telles que le volume molaire, l'indice de réfraction, la densité et la polarisabilité entrent généralement dans la classe des descripteurs moléculaires 0D. Ces descripteurs sont faciles à obtenir, mais ne contiennent pas beaucoup d'informations sur les caractéristiques structurelles des molécules. C'est pourquoi ils sont souvent utilisés en conjonction avec d'autres descripteurs (par exemple, les descripteurs moléculaires topologiques, physico-chimiques et chimiques quantiques) afin d'obtenir une compréhension plus complète de la molécule.

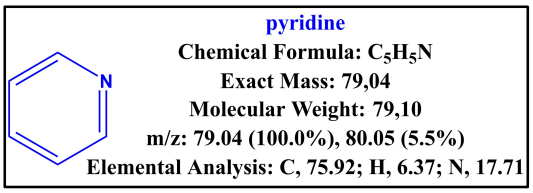


Figure II.5. Exemple de quelques descripteurs moléculaires 0D calculés à partir de la formule chimique C5H5N de la pyridine.

4.6.2.2 Descripteurs unidimensionnels (1D)

La représentation des molécules dans une structure 1D est basée sur la présentation de leurs sous-structures, telles que les groupes fonctionnels, les fragments moléculaires et les substituants [220]. Cette représentation ne nécessite pas une connaissance approfondie de la structure moléculaire. Par conséquent, ces descripteurs sont principalement définis de manière binaire (1 pour la présence ou 0 pour l'absence) d'une sous-structure (**Figure II.6**). Les descripteurs 1D les plus courants sont les empreintes digitales " fingerprints " [225], telles que : Nombre d'atomes, Nombre de liaisons rotatives, Nombre de liaisons hydrogène (accepteurs et donneurs), Nombre de groupes lipophiles, Nombre de groupes hydrophiles, Nombre de centres chiraux, Nombre d'atomes hybridés en sp^3, etc.

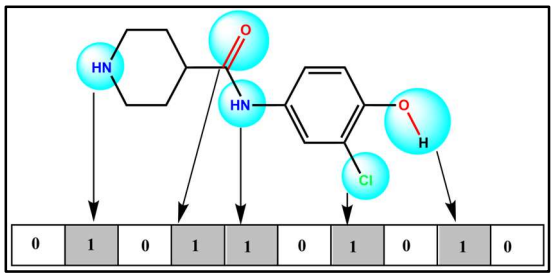


Figure II.6. Illustration de la sous-structure par empreintes digitales de la molécule de N-(3-chloro-4-hydroxyphényl) pipéridine-4-carboxamide ; les sous-structures moléculaires représentées par les bits présents dans la molécule (entourés) sont définies à 1 et celles qui sont en absence sont définies à 0.

4.6.2.3 Descripteurs bidimensionnels (2D)

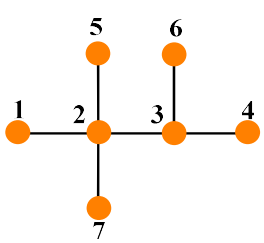
Dans la représentation 2D, les liaisons et les atomes d'une molécule sont considérés comme les arêtes et les sommets d'un graphe, respectivement. Ce type de descripteur intègre des propriétés topologiques telles que la taille, la forme, la symétrie, la ramification et la cyclicité [220]. En outre, les propriétés chimiques des atomes, telles que la masse et la polarisabilité, et la présence de

donneurs/accepteurs de liaisons hydrogène sont prises en compte. La plupart des descripteurs de cette classe sont dérivés de la matrice d'adjacence [222], de la matrice de Coulomb ou de la matrice de distance [226], [227].

Matrice d'adjacence (a_{ij}) : Une matrice d'adjacence est une matrice carrée utilisée pour représenter un graphe défini par l'**Eq. (II.3)**.

$$a_{ij} = \begin{cases} 1 \text{ si les atomes i et j sont connect} \\ 0 \text{ si les atomes i et j ne sont pas connect} \end{cases} \qquad \text{(II.3)}$$

Les éléments de la matrice indiquent si les paires de sommets sont adjacentes ou non dans le graphe. Cette matrice peut être utilisée pour calculer des descripteurs moléculaires bidimensionnels, tels que la forme, la taille et la connectivité des molécules. Pour les descripteurs moléculaires, chaque atome de la molécule est représenté par un sommet, et les arêtes du graphe représentent les liaisons chimiques entre les atomes. Une arête entre deux sommets indique que les deux atomes sont reliés par une liaison chimique. La matrice d'adjacence peut ensuite être utilisée pour calculer le nombre d'atomes dans la molécule, le nombre de liaisons distinctes, le nombre de différents types de liaisons et le nombre de différents types d'anneaux dans la molécule (**Figure II.7**).



	1	2	3	4	5	6	7	δ_i
1	0	1	0	0	0	0	0	1
2	1	0	1	0	1	0	1	4
3	0	1	0	1	0	1	0	3
4	0	0	1	0	0	0	0	1
5	0	1	0	0	0	0	0	1
6	0	0	1	0	0	0	0	1
7	0	1	0	0	0	0	0	1

δ_i: Somme des lignes de la matrice d'adjacence des sommets

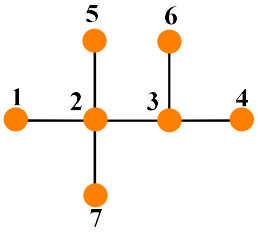
Figure II.7. Exemple de matrice d'adjacence d'un graphe (2,2,3-triméthylbutane).

Matrice de Coulomb : Une matrice de Coulomb est une matrice utilisée pour représenter les interactions entre les atomes d'une molécule. Elle est calculée en additionnant les interactions de Coulomb entre chaque paire d'atomes, ce qui permet de calculer des descripteurs moléculaires bidimensionnels tels que la structure électronique, le moment dipolaire et les charges atomiques. Les éléments de la matrice sont calculés comme la somme des interactions de Coulomb entre chaque paire d'atomes. Par exemple, l'élément de la ligne i et de la colonne j de la matrice est donné par la somme des interactions de Coulomb entre l'atome i et l'atome j , comme le montre l'**Eq. (II.4).** La matrice peut ensuite être utilisée pour calculer le nombre d'atomes dans la molécule, le nombre de liaisons distinctes, le nombre de différents types de liaisons et le nombre de différents types d'anneaux dans la molécule.

$$\left[Couloumb\right]_{ij} = \begin{cases} \frac{1}{2} \times Z_i^{2.4} & \text{si i=j} \\ \frac{Z_i Z_j}{r_{ij}} & \text{si i} \neq \text{j} \end{cases} \qquad \text{(II.4)}$$

où Z_i est le numéro atomique des atomes i et j ; r_{ij} est la distance interatomique entre les atomes i et j.

Matrice de distance (d_{ij}): Une matrice de distance est une matrice carrée utilisée pour représenter les distances entre les atomes d'une molécule (**Figure II.8**). Elle est calculée en calculant la distance euclidienne entre chaque paire d'atomes, qui peut être utilisée pour calculer des descripteurs moléculaires 2D tels que la forme, la taille et la connectivité des molécules. Les éléments de la matrice sont calculés comme la distance euclidienne entre chaque paire d'atomes. Par exemple, l'élément de la ligne i et de la colonne j de la matrice est donné par la distance euclidienne entre l'atome i et l'atome j. La matrice peut ensuite être utilisée pour calculer le nombre d'atomes dans la molécule, le nombre de liaisons distinctes, le nombre de différents types de liaisons et le nombre de différents types d'anneaux dans la molécule.



	1	2	3	4	5	6	7	θ_i
1	0	1	2	3	2	3	2	13
2	1	0	1	2	1	2	1	8
3	2	1	0	1	2	1	2	9
4	3	2	1	0	3	2	3	14
5	2	1	2	3	0	3	2	13
6	3	2	1	2	3	0	3	14
7	2	1	2	3	2	3	0	13

θ_i:Somme totale des lignes de la matrice de distance entre les sommets

Figure II.8. Exemple de matrice de distance d'un graphe (2,2,3-triméthylbutane).

4.6.2.4 *Descripteurs tridimensionnels (3D)*

Les descripteurs 3D fournissent de nombreuses informations sur une molécule en tenant compte de sa nature, de sa connectivité et de sa configuration spatiale. Cette classe de descripteurs est généralement utilisée pour modéliser les propriétés biologiques et pharmaceutiques, mais il existe des problèmes potentiels liés au processus d'optimisation. Parmi les descripteurs 3D les plus connus est la matrice moléculaire, qui utilise les coordonnées cartésiennes (x, y, z) pour différencier les isomères, mais leur calcul peut prendre beaucoup de temps [228]. Il existe également un autre type de descripteurs appelés 3D-MoRSE (Molecular Representation of Structures based on Electronic diffraction) qui sont des descripteurs moléculaires calculés à partir de la matrice moléculaire [229]. Dans cette classe de descripteurs moléculaires (3D), nous utilisons des descripteurs générés par des algorithmes prédéfinis dans le programme Chem3D, ainsi que des descripteurs moléculaires générés à partir de calculs DFT basés sur la Z-matrice «également nommé Z-matrix » [230]. La Z-matrice permet de représenter les molécules en fonction de leurs coordonnées atomiques. Elle est composée de lignes qui fournissent le symbole atomique, le numéro de l'atome et des paramètres qui spécifient la distance, l'angle et l'angle dièdre entre chaque atome (**Figure II.9**).

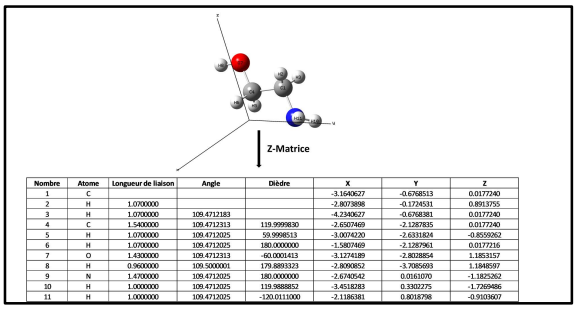


Figure II. 9. Exemple de représentation de type Z-matrix de la structure du 2-aminoéthan-1-ol.

Grâce à la Z-matrice, il est possible de calculer des descripteurs moléculaires 3D sur la base des distances, des angles et des angles dièdres entre chaque atome. Ces valeurs sont ensuite utilisées comme paramètres d'entrée pour une méthode de calcul telle que la théorie de la fonctionnelle de la densité (DFT), qui peut calculer une variété de propriétés telles que l'énergie, les moments dipolaires et les densités de charge. Les propriétés calculées peuvent ensuite être utilisées pour développer une description tridimensionnelle de la molécule.

4.6.2.5 Descripteurs quadridimensionnels (4D)

Les descripteurs basés sur une grille, couramment référés comme descripteurs 4D, prennent en considération non seulement la géométrie moléculaire, mais introduisent également une quatrième dimension. Cette dernière concerne généralement l'interaction entre la molécule et le site actif du récepteur ou les différents états de conformation de la molécule. Analyse Comparative des Champs Moléculaires (CoMFA) et l'Analyse Comparative des Indices de Similitude Moléculaire (CoMSIA) sont deux modèles de descripteurs 4D [94], [231]. Bien que ces descripteurs fournissent plus d'informations que les descripteurs 3D, ils sont plus difficiles à obtenir et peuvent donc nécessiter plus de temps et de ressources. Pour générer une représentation basée sur une grille, les molécules doivent être placées dans une grille 3D de plusieurs milliers de points régulièrement espacés, puis cartographiées à l'aide d'une sonde (stérique, électrostatique, hydrophile, etc.). Malgré la difficulté d'obtenir des descripteurs 4D, leur avantage est qu'ils peuvent toujours fournir des valeurs dissemblables pour des molécules structurellement différentes. C'est l'une des raisons pour lesquelles ces descripteurs moléculaires fournissent une caractérisation structurelle plus approfondie et une meilleure rationalisation des propriétés structurelles favorables/défavorables pour la modélisation de l'activité biologique/propriété physico-chimique des molécules.

4.6.3 Techniques et Outils pour transformer la présentation graphique en descripteurs moléculaires multidimensionnels

Les descripteurs moléculaires sont utilisés pour prédire les propriétés et les caractéristiques des molécules [223]. Il est donc essentiel de préparer et d'optimiser les structures moléculaires avant de calculer les descripteurs moléculaires. Les structures moléculaires optimales fournissent une représentation précise et fiable de la molécule qui reflète fidèlement ses propriétés physiques et chimiques [232]. Ceci est particulièrement important lors du calcul des descripteurs moléculaires, car les valeurs obtenues dépendent de la structure de la molécule. En outre, l'optimisation de la structure moléculaire peut contribuer à réduire le temps de calcul et le coût du calcul des descripteurs [233]. Des structures mal optimisées ou imprécises peuvent conduire à des résultats inexacts ou incohérents, ce qui peut affecter la fiabilité des descripteurs. Il est donc important de préparer et d'optimiser les structures moléculaires avant de calculer les descripteurs moléculaires afin de garantir la fiabilité et la précision des résultats.

4.6.3.1 Approches Empiriques et Semi-Empiriques utilisant la Mécanique Moléculaire et Quantique (MM/MQ)

Pour obtenir des résultats précis et optimaux avec des structures moléculaires, il est important de se familiariser avec des outils de préparation et d'optimisation. Ces outils peuvent être des logiciels ou des applications informatiques qui peuvent être utilisés pour ajuster la taille et la forme des molécules, ainsi que pour ajouter des atomes aux molécules existantes. Ils permettent d'améliorer la stabilité et la force des molécules, et de réduire leurs coûts et leur temps de traitement. En comprenant et en utilisant ces outils, on peut optimiser les structures moléculaires et obtenir des résultats précis et fiables. Les méthodes d'optimisation moléculaire ont une grande importance pour la recherche des structures énergétiques minimales des molécules. L'optimisation moléculaire est une méthode computationnelle utilisée pour affiner les structures de molécules, telles que les composés organiques, les peptides et les protéines [234]. L'optimisation moléculaire peut être exécutée à l'aide des méthodes empiriques de la mécanique moléculaire (MM) ou des méthodes semi-empiriques (*ab Initio*) de la mécanique quantique (MQ) [235]. La recherche des structures d'énergie minimale nous permet de mieux comprendre le comportement de la molécule et la manière dont elle peut interagir avec d'autres molécules.

4.6.3.2 Champs de force et méthodes de calcul exploités

Les champs de force sont utilisés pour décrire les interactions entre les atomes et nous aident donc à comprendre comment la molécule interagira avec d'autres molécules [236]. Les champs de force peuvent également être utilisés pour générer des structures 3D précises de molécules, qui peuvent être utilisées dans des applications QSAR pour générer des descripteurs moléculaires qui peuvent être utilisés pour prédire l'activité des molécules. Par exemple, les champs de force peuvent être utilisés pour calculer le nombre de liaisons hydrogène, les distances entre les atomes et d'autres propriétés qui peuvent être utilisées comme descripteurs dans les applications QSAR. Le **Tableau II.3** présente certaines méthodes qui peuvent être utilisées pour optimiser les structures moléculaires, ainsi que les champs de force utilisés dans chaque méthode.

Tableau II.3. Méthodes de calcul et champs de force les plus utilisés.

Méthode de calcul	Définition	Champ de force
Mécanique Moléculaire (MM)	En MM, des fonctions d'énergie potentielle sont utilisées pour modéliser l'interaction entre les atomes de la molécule et les forces qui agissent sur eux.	Optimisation 3D, MM2/MM3/MM4, Tripos, AMBER/Kollman, MMFF94/MMFF94s, OPLS, CHARMM, GROMOS, UFF, DREIDING FF, MOPAC, etc.
Mécanique Quantique (MQ)	La Mécanique Quantique (MQ) est une autre branche de la chimie computationnelle qui utilise l'équation de Schrödinger pour modéliser avec précision le comportement des électrons dans une molécule. Cette approche est beaucoup plus coûteuse en termes de calcul que la MM, mais elle permet de prédire avec plus de précision les dimensions, les angles et les énergies des liaisons.	Hartree-Fock (HF), Théorie de la fonctionnelle de la densité (DFT), Interaction de Configuration (CI), Kohn-Sham, Théorie des perturbations de Møller-Plesset, Théorie de la réponse linéaire, Théorie des perturbations à références multiples, etc.

Lors du choix d'une méthode d'optimisation moléculaire et d'un champ de force pour l'optimisation structurelle, l'utilisateur doit évaluer la précision souhaitée, la complexité du système et les ressources de calcul disponibles. Les champs de force empiriques de la MM et semi-empirique de la MQ (*ab initio*) peuvent être adaptés pour les petites molécules organiques, tandis que les champs de force à tous les atomes peuvent être plus appropriés pour les systèmes plus grands tels que les protéines ou les acides nucléiques [237]. Les méthodes MM/MQ peuvent être utilisées en combinaison pour obtenir des résultats plus précis. C'est ce que nous avons réalisé dans le cadre de nos travaux afin d'optimiser les structures moléculaires des bases de données de petites molécules et de calculer les descripteurs moléculaires appropriés pour élaborer des modèles QSAR avec un haut degré de fiabilité et de performance.

4.6.3.3 Approche et outils pour la génération des descripteurs moléculaires multidimensionnels

Le **Tableau II.4** présente les outils et les méthodes que nous avons adoptés dans nos travaux pour préparer et optimiser les structures moléculaires 2D et 3D utilisées comme base pour calculer les descripteurs moléculaires de 0D/1D (Constitutionnel), 2D (Topologique), 3D (Physico-chimique & Descripteurs quantiques de la réactivité chimique), et 4D (CoMFA et CoMSIA). Au-delà des outils et des méthodes que nous avons appliqués pour générer des descripteurs moléculaires 0D-4D dans cette thèse, il existe de nombreux autres outils et serveurs disponibles sur Internet, tels que Schrödinger software (https://www.schrodinger.com/kb/1439), BIOVIA Discovery Studio (https://discover.3ds.com/), Molecular Operating Environment (MOE : https://www.chemcomp.com/Products.htm), alvaDesc (https://www.alvascience.com/alvadesc/), ADMET predictor (https://www.simulations-plus.com/software/admetpredictor/), ADRIANA.Code (http://www.akosgmbh.eu/molnet/adriana.htm), CDK (http://cdk.github.io/), CODESSA (http://www.codessa-pro.com/), DRAGON (http://www.talete.mi.it/), JOELib (http://www.ra.cs.uni-tuebingen.de/), MOLGEN-QSPR (http://www.edusoft-lc.com/molconn), PADEL (http://www.padel.nus.edu.sg/), OpenBabel (http://www.openbabel.org/), ChemDes (http://www.scbdd.com/chemdes), CORAL-QSAR/QSPR (http://www.insilico.eu/coral/), PyBioMed (https://pybiomed.readthedocs.io/en/latest/index.html), PaDELPy (https://github.com/ecrl/padelpy), etc. Dans le cadre de ce travail, afin de construire des modèles QSAR basés sur des données fiables, nous utilisons des descripteurs moléculaires multidimensionnels calculés par les outils professionnels et puissants que sont ChemSketch, ChemOffice, Gaussian 09W et SYBYL-X 2.1.1, développés à des fins de modélisation moléculaire sophistiquée. Ci-dessous nous indiquons de manière générale quelques raisons qui nous ont poussés à exploiter des outils professionnels pour le calcul des descripteurs moléculaires et à ne pas utiliser les outils open source et les interfaces de serveurs web en ligne :

- Les outils professionnels offrent des fonctionnalités d'optimisation 3D plus avancées et plus précises que les serveurs web et les algorithmes basées sur Python.
- Ces outils incluent des algorithmes de calcul de descripteurs moléculaires très avancés qui peuvent être utilisés pour calculer des descripteurs moléculaires plus précis.

- Ces outils offrent des options d'optimisation avancées telles que la méthode B3LYP hybride Gaussian 09W qui peut être utilisée pour obtenir des résultats plus précis et plus fiables.
- Ces outils offrent des fonctionnalités qui permettent aux utilisateurs de visualiser et de modéliser des molécules en 3D, ce qui est très utile pour l'analyse et la conception moléculaires.

De plus, les outils professionnels sont faciles à utiliser et à comprendre, ce qui permet aux utilisateurs de gagner du temps et de réaliser des tâches avec une plus grande précision.

Tableau II.4. Outils bio-informatique et méthodes d'optimisation géométrique et énergétique utilisés dans nos travaux pour générer des descripteurs moléculaires.

Méthode de calcul	Méthodes d'optimisation	Outils	Descripteurs moléculaires calculés	Modèle QSAR
MM	Optimisation 3D disponible dans ChemSketch [238].	*ACD/ChemSketch*	0D-3D	2D-QSAR
	Méthode de champ de force MM2 disponible dans Chem3D [239, p. 2].	*ChemOffice*		
MQ	Méthode B3LYP hybride de Lee-Yang-Parr Becke à trois paramètres utilisant la base 6-31G (d, p) disponible dans Gaussian 09W [240].	*GaussView 6.0*		
MM	Champ de force Tripos disponible dans SYBYL-X 2.1.1[241] .	*SYBYL-X 2.1.1*	4D	3D-QSAR

4.7 Base de données de descripteurs moléculaires pour la modélisation 2D-QSAR

4.7.1 Descripteurs moléculaires basés sur l'optimisation 3D

ACD/ChemSketch offre des fonctions d'optimisation 3D avec lesquelles les utilisateurs peuvent élaborer des modèles moléculaires 3D plus précis et plus complexes que les modèles 2D représentés. Les molécules peuvent être facilement optimisées en utilisant une énergie minimale et des forces intermoléculaires [242]. L'optimisation 3D dans ChemSketch permet également à l'utilisateur d'ajuster les paramètres des molécules et d'utiliser des algorithmes d'optimisation spécifiques pour chaque type de molécule (organiques, organométalliques, polymères, ...). Ces algorithmes adaptés à chaque type de molécule aident à réduire le temps de calcul et à obtenir des résultats plus précis. Le **Tableau II.5** présente les quinze descripteurs moléculaires que nous avons collectés après avoir optimisé les structures moléculaires par la méthode d'optimisation 3D à l'aide du logiciel ACD/ChemSketch.

Tableau II.5. Descripteurs moléculaires calculés dans nos travaux à l'aide d'ACD/ChemSketch sur la base de la méthode d'optimisation 3D.

Descripteur moléculaire	Catégories/ Nombre	Définition
Poids de formule (PF)	17 descripteurs moléculaires mixtes	Le poids de la formule est la somme des poids atomiques de chaque élément du composé chimique [243].
Composition chimique (X%)		La composition chimique est une évaluation de la composition chimique d'une substance. Elle est utilisée pour déterminer la quantité exacte de chaque élément présent dans la substance, ainsi que les proportions relatives de chaque élément [244].

Descripteur moléculaire	Catégories/ Nombre	Définition
Réfractivité molaire (RM)		La réfractivité molaire est la capacité d'une substance à modifier la direction de propagation de la lumière à cause des interactions entre ses molécules et les photons. Elle est un indicateur de la force de ces interactions [245].
Volume molaire (VM)		Le volume molaire représente la quantité d'espace occupée par une mole d'une substance, ce qui correspond à 22,4 litres de gaz à une température et une pression standard [246]. Il sert à déterminer les propriétés physiques et chimiques des matériaux et à construire des modèles précis de leur structure.
Parachor (Pc)		Le parachor est une unité de mesure qui permet de calculer la tension superficielle relative d'un composé [247]. Cette valeur est obtenue en divisant la tension superficielle d'un liquide pur par le poids moléculaire du composé. Elle peut être utilisée pour comparer la tension superficielle de différents composés et pour déterminer leur solubilité dans un solvant spécifique.
Indice de réfraction (n)		L'indice de réfraction (n) est une valeur qui exprime la courbure de la lumière lorsqu'elle traverse un milieu [248]. La vitesse de la lumière diminue à mesure que l'indice de réfraction augmente, et cette valeur est calculée en divisant la vitesse de la lumière dans le vide (c) par la vitesse de la lumière dans le milieu (v).
Tension superficielle (γ)		La tension superficielle est une caractéristique des liquides qui leur permet de résister à des forces extérieures [249]. Elle se forme lorsque les molécules à la surface du liquide sont liées les unes aux autres, créant ainsi une couche qui empêche les molécules de se déplacer. Cette propriété joue un rôle essentiel dans diverses applications, notamment la formation de gouttes et de bulles, le mouvement des liquides dans les capillaires et la capacité des insectes à marcher sur l'eau.
Densité (ou densité relative) (DR)		La densité d'une substance est le rapport entre sa masse et le volume qu'elle occupe [250]. Sa densité relative, ou gravité spécifique, est le rapport entre sa densité et celle d'une substance de référence, généralement l'eau.
Constante diélectrique (K)		La constante diélectrique, également appelée permittivité, est une caractéristique sans dimension qui permet de mesurer la capacité d'un matériau à être polarisé lorsqu'un champ électrique est appliqué [251]. Elle est calculée en divisant la densité du flux électrique par l'intensité du champ électrique à travers un matériau.
Polarisabilité (αe)		La polarisabilité d'une molécule est sa capacité à être affectée par un champ électrique externe, mesurée en unités de cm3 [252]. Elle est influencée par la taille, la forme et la composition chimique de la molécule.
Anneau à double liaison équivalent (Ring Double Bond Equivalents) (RDBE)		Le degré d'insaturation d'une molécule est exprimé par son Indice de Double Liaison (EBRD) [253]. Cette mesure est obtenue en divisant le nombre total de doubles liaisons dans la molécule par le nombre d'atomes de carbone. Plus, l'EBRD est élevé, plus la molécule est réactive.
Masse mono-isotopique (MMI)		La masse mono-isotopique est la somme des masses des atomes qui composent une molécule, calculée en considérant leur isotope le plus abondant [254]. On l'utilise pour déterminer la masse moléculaire réelle d'un composé plutôt que sa masse moyenne. Elle est aussi connue sous le nom de masse atomique ou de masse molaire.
Masse nominale (MN)		La masse nominale d'un composé est la masse correspondant à une molécule du composé, calculée à partir des poids atomiques de ses atomes individuels [255].
Masse Moyenne (MM)		La masse moyenne d'une molécule est un nombre qui représente la somme des masses atomiques des atomes qui composent la molécule, divisée par le nombre d'atomes qui composent la molécule [256].
Spectrométrie de masse par photo-ionisation (PPI) :		La spectrométrie de masse PPI produit quatre types d'ions : M+, M-, [M+H]+ et [M-H]+. M+ et M- sont les ions moléculaires, [M+H]+ est l'ion moléculaire protoné et [M-H]+ est l'ion moléculaire déprotoné.

Descripteur moléculaire	Catégories/ Nombre	Définition
M^+, M^-, $[M+H]^+$, $[M+H]^-$ [257]		M+ et M- : Ces ions moléculaires sont formés lorsque la molécule est ionisée par un champ électrique élevé. [M+H]+ : Cet ion moléculaire protoné est formé par la réaction de transfert de proton entre l'ion moléculaire M+ et un proton (H+). [M-H]+ : Cet ion moléculaire déprotoné est formé par la réaction de transfert de protons entre l'ion moléculaire M- et un proton (H+).

4.7.2 Descripteurs moléculaires basés sur l'optimisation MM2

La méthode de champ de force MM2 est une méthode classique de champ de force utilisée dans les simulations de mécanique moléculaire [258]. La méthode du champ de force MM2 a été développée en 1985 par Alan J. Rappe et al. [259], [260]. Le modèle MM2 est basé sur une approche empirique qui utilise une combinaison de paramètres d'étirement de liaison, de flexion d'angle, d'angle de torsion et de van der Waals pour décrire les interactions entre les atomes. Les simulations du champ de force MM2 sont utilisées pour prédire les propriétés physiques et chimiques des molécules. Les simulations permettent de déterminer l'énergie d'une molécule, qui peut ensuite être utilisée pour prédire diverses propriétés telles que le point d'ébullition, le point de fusion, la solubilité, la réactivité et la stabilité. Le champ de force MM2 est également utilisé pour simuler le comportement conformationnel des molécules, comme les rotations des chaînes latérales et les angles dièdres. Le modèle MM2 est largement utilisé dans les simulations de petites molécules et de protéines à l'aide du logiciel Chem3D. Grâce à l'optimisation énergétique et géométrique des molécules réalisée sur le modèle MM2 par le logiciel Chem3D, nous avons pu calculer 41 descripteurs moléculaires hybrides (Constitutionnel et physico-chimique et Topologique) pour chaque structure moléculaire (**Tableau II.6**).

Tableau II.6. Descripteurs moléculaires calculés dans nos travaux par Chem3D sur la base de la méthode d'optimisation MM2.

Catégorie	Nombre	Descripteur moléculaire
Constitutionnel, physico-chimique, et Géométrique (0D-3D).	27	Énergie élastique (SE), Énergie de flexion (BE), Énergie élastique-fléxion (SBE), Énergie de torsion (E_{Tor}), Énergie Non-1,4 van-der-Waals (E_{NVDW}), Énergie 1,4 van-der-Waals (E_{VDW}), Énergie totale (E_T), Charge formelle (FC), Zone accessible de Connolly (CAA), Zone moléculaire de Connolly (CMA), Volume de solvant Connolly exclu (CSVE), Masse exacte (EM), Masse, Masse moléculaire (MM), Nombre de liaisons Hydrogène Accepteurs (NHBA), Nombre d'accepteurs de liaisons hydrogène (NHBA), Nombre de donneurs de liaisons hydrogène (NHBD), Ovalité, Moment principal, Analyse élémentaire (%H, %C), Spectre de masse (m/z), Formule moléculaire, Formule moléculaire en HTML, Réfractivité moléculaire, Coefficient de partage octanol-eau (LogP), Solubilité dans l'eau (logS), Acidité moléculaire (pKa), Indice de Balaban (BI), Nombre de groupes.
Topologique (2D)	14	Indice de Balaban (BI), Nombre de blocs (CC), Index de topologie moléculaire (MTI), Nombre de liaisons rotatives (NRB), Zone de Surface Polaire (PSA), Rayon (r), Attribut de forme, Coefficient de forme, Somme des Degrés, Somme des Degrés de Valence, Diamètre Topologique, Connectivité Totale, Connectivité Total de Valence, Indice de Wiener.

4.7.3 Descripteurs moléculaires basés sur les calculs DFT

La méthode hybride Lee-Yang-Parr à trois paramètres (B3LYP) combine la fonctionnelle d'échange de Becke-88 et la fonctionnelle de corrélation Lee-Yang-Parr, dans le but de réduire le coût

de calcul associé aux calculs réalisés sur la base de la Théorie de la Fonctionnelle de la Densité (DFT, sigle pour Density Functional Theory) [261]. La méthode B3LYP permet de réaliser des calculs énergétiques et géométriques de petites molécules à l'aide d'ensembles de bases tels que le 6-31G (d, p) qui est appliqué dans le cadre de nos travaux. La méthode B3LYP/6-31G(d,p) est bien adaptée au calcul des descripteurs de la réactivité moléculaire, car elle est capable de décrire avec précision la géométrie et les caractéristiques énergétiques des petites molécules [262]. Ce procédé est très utile dans les études 2D-QSAR car il permet de calculer les descripteurs moléculaires électronique tels que les charges atomiques, les moments dipolaires et les polarisations, Énergies des orbitales moléculaires de frontière (Gap énergétique, Energie totale, etc.), qui sont essentiels pour comprendre comment les molécules interagissent mutuellement [263]. De plus, la méthode B3LYP/6-31G(d,p) s'est avérée être une solution optimale pour le calcul des propriétés thermodynamiques telles que l' énergie électronique, l'énergie du point zéro, l'énergie thermique, la capacité thermique, l'énergie libre de Gibbs, etc. Le **Tableau II.7** rassemble les descripteurs moléculaires issus de l'optimisation énergétique et géométrique des molécules avec la méthode de calcul B3LYP/6-31G(d,p).

Tableau II.7. Les descripteurs moléculaires électroniques et thermodynamiques résultent aux calculs DFT menés dans nos travaux sur la base de la méthode B3LYP/6-31G(d,p) de Gaussian 09W.

Catégorie/Nombre	**Descripteurs moléculaire**	
	Électronique	**Thermodynamique**
24 descripteurs de chimie quantique basés sur la DFT	Énergie électronique totale E (RB3LYP), Moment dipolaire (DM), Polarisabilité (α), Énergie Total (E_T), Énergies de l'orbitale moléculaire la plus haute occupée (E_{HOMO}) et de l'orbitale moléculaire la plus basse inoccupée (E_{LUMO}), Bande interdite (E_{gap}), Electronégativité (χ), Potentiel chimique (μ), Dureté chimique (η), Mollesse (S), L'indice d'électrophilie (w).	Énergie électronique (EE), Correction de l'énergie du point zéro (ZPE), Correction thermique de l'énergie (ΔE), Correction thermique de l'enthalpie (ΔH), Correction thermique de l'énergie libre (ΔG), EE + Énergie du point zéro, EE + Correction de l'énergie thermique, EE + Correction de l'enthalpie thermique, EE + Correction de l'énergie libre thermique, ΔE (thermique), Capacité thermique (Cv), Entropie (ΔS).

Sans doute la combinaison de descripteurs moléculaires multidimensionnels (0D-3D) peut-elle conduire au développement de modèles 2D-QSAR plus fiables et plus faciles pour simuler et supporter des modèles expérimentaux qui ne dépendent pas de la structure protéique des récepteurs cibles pendant l'évaluation biologique. Cela peut être considéré comme la limite majeure des modèles 2D-QSAR. En effet, les descripteurs 0D-3D ne peuvent pas expliquer la relation quantitative structure-activité qui résulte des interactions médicament-récepteur. Par conséquent, l'interpolation des descripteurs moléculaires basés sur le champ de force 4D dans la modélisation QSAR sert de clé pour interpréter cette relation dans l'espace 3D. Cependant, la modélisation moléculaire basée sur l'approche 2D-QSAR n'est pas moins importante par rapport au 3D-QSAR dans le processus de conception de médicaments basé sur l'approche LBDD. Ceci lorsque ces approches sont combinées avec d'autres modèles computationnels plus exhaustifs et plus fiables tels que les tests ADME-Tox *in silico*, les simulations de Docking moléculaire et la dynamique moléculaire. C'est ce que nous avons visé à atteindre par le biais de nos travaux réalisés dans le cadre de cette thèse, à savoir la mise en place d'une stratégie de CADD hybride.

4.8 Base de Données de Descripteurs Moléculaires 4D pour la Modélisation 3D-QSAR

Les descripteurs moléculaires quadridimensionnels (4D) se traduisent par des échanges énergétiques entre les molécules insérées dans une grille de mesure et dans une sonde [264]. Une fois ces descripteurs moléculaires avancés calculés, il est possible d'encoder les types de champs de force affectant les propriétés/activités moléculaires dans l'espace 3D, ce qui peut être réalisé par le biais d'études 3D-QSAR. À cet égard, plusieurs techniques et méthodes ont été développées et optimisées pour permettre le calcul de descripteurs moléculaires 4D. Les champs de force Tripos constituent la plus importante de ces méthodes. Nous les avons adoptés dans nos travaux en tant que base pour générer des descripteurs moléculaires de champ 4D lors de la construction de modèles 3D-QSAR.

4.8.1 Champ de force Tripos et méthodologies associées

En générale, le champ de force peut être défini comme une combinaison computationnelle des aspects liés aux coordonnées et aux paramètres qui correspondent empiriquement à l'énergie potentielle de surface d'une molécule ainsi qu'aux propriétés structurelles de cette molécule (telles que les distances entre les atomes, les angles de liaison, les angles dièdres, les longueurs de liaison, etc.). Les forces stériques, électrostatiques, hydrophobiques et de liaison hydrogène sont parmi les forces les plus importantes observées dans les interactions médicament-récepteur [265]. Le calcul de ces forces et leur décodage en descripteurs moléculaires (4D) fourniront donc des données riches nous permettant d'interpréter la relation structure-activité dans l'espace 3D. Ceci peut être réalisé par des calculs de champ de force adaptés à cette fin, notamment le champ de force Tripos. Le champ de force Tripos est un champ de force connu développé en 1985 par Tripos Inc [241]. et est utilisé dans l'optimisation 3D de petites molécules et dans le développement de modèles 3D-QSAR basés sur les approches de CoMFA et CoMSIA. Le champ de force Tripos est un modèle de paramétrage moléculaire qui utilise des termes de liaison, d'angle et de dièdre pour représenter la structure moléculaire, et qui est utilisé pour calculer numériquement l'énergie de la molécule [266]. Cette information est ensuite utilisée pour comprendre la relation structure-activité (SAR) de la molécule. Le champ de force Tripos est important car il identifie la conformation la plus stable d'une molécule, ce qui permet d'optimiser la structure de la molécule. De plus, le champ de force Tripos fournit les descripteurs de champs basés sur les approches de CoMFA et CoMSIA, qui sont également exploités pour élaborer des modèles 3D-QSAR [235]. Ces modèles peuvent ensuite être utilisés pour prédire l'activité de nouvelles molécules et pour concevoir de nouvelles molécules ayant les propriétés souhaitées. L'énergie "E" d'une molécule dérivée du champ de force de Tripos est donnée par la somme des contributions énergétiques. Ces contributions sont représentées comme étant la somme de toutes les liaisons, angles de liaison, angles de torsion et interactions non liées entre les atomes non liés ou un co-atome (c.-à-d. 1, 4-interactions et plus), comme le montre **l'Eq. (II.5)** [241].

$$E_{TFF} = \sum E_{str} + \sum E_{bend} + \sum E_{oop} + \sum E_{tors} + \sum E_{vdw} + \left[\sum E_{elec} + \sum E_{dist_c} + \sum E_{ang_c} + \sum E_{tor_c} + \sum E_{range_c} + \sum E_{multi} + \sum E_{fieldfit}\right] \quad \text{(II.5)}$$

où E_{str} (terme d'énergie d'étirement de la liaison): Énergie d'une liaison étirée ou compressée par rapport à sa longueur de liaison originale ; E_{bend} (Angle de flexion de l'énergie) : Énergie de flexion des angles de liaison par rapport à leurs valeurs naturelles *;* E_{oop} (terme d'énergie de flexion hors du plan) : Énergie de flexion des atomes planaires hors du plan ; E_{tors} (terme d'énergie de torsion) : Énergie de torsion due à la torsion des liaisons ; E_{vdw} (terme d'énergie de van der Waals) : Énergie due aux interactions non liées de van der Waals ; E_{ele} (terme d'énergie électrostatique) : Énergie due aux interactions électrostatiques ; E_{dist_c} (terme d'énergie lié aux contraintes de distance) : Énergie associée aux contraintes de distance ; E_{ang_c} (terme d'énergie lié à la contrainte d'angle) : Énergie associée aux contraintes d'angle ; E_{tor_c} (terme d'énergie lié à la contrainte de torsion) : Énergie associée aux contraintes d'angle de torsion ; E_{range_c} (terme d'énergie lié aux contraintes de zone étendue) : Énergie associée aux contraintes liées à l'étendue ; E_{multi} (terme énergétique multifit) : Énergie associée aux interactions intra et intermoléculaires d'une molécule ; E_{field} fit : Énergie associée à l'ajustement du champ.

4.8.2 Méthodes de calcul des charges pour les structures 3D des molécules

Le calcul des charges électrostatiques sur les atomes d'une molécule est important pour comprendre la relation entre la structure et l'activité de la molécule [267]. Les charges ponctuelles sont utilisées pour décrire les interactions électrostatiques entre les atomes et les molécules d'un système moléculaire (comme les protéines et les enzymes,…) [268]. Par le calcul des charges sur chaque atome d'une molécule, il est possible de déterminer l'énergie de la molécule, qui peut alors être utilisée pour comprendre sa réactivité vis-à-vis le récepteur protéique ciblé. En outre, les charges ponctuelles peuvent être utilisées pour calculer le potentiel électrostatique d'une molécule, ce qui permet de déterminer la conformation la plus stable d'une molécule. Ces informations sont importantes pour optimiser la structure 3D d'une molécule et pour le développement de modèles QSAR. Dans cette thèse, en utilisant le logiciel SYBYL-X 2.1.1, en vue du calcul des descripteurs moléculaires 4D sur la base des approches CoMFA et CoMSIA, les charges formelles de chaque atome d'une molécule sont calculées en fonction du nombre de valences et de liaisons libres attribuées à l'atome. Les ions métalliques, auxquels on attribue une charge formelle initiale de zéro, font l'objet d'une attention particulière. Plusieurs méthodes de calcul de la charge des petites molécules sont disponibles dans le logiciel SYBYL. Les charges de Gasteiger-Hueckel sont la méthode la plus utilisée pour calculer la charge des petites molécules dans la modélisation 3D-QSAR via les techniques CoMFA et CoMSIA en raison de leur précision. Dans cette méthode, les méthodes Gasteiger-Marsili et Hückel sont combinées pour calculer les composantes σ et π de la charge atomique dans un système π, respectivement. La charge totale est obtenue par la sommation des charges issues dans les deux méthodes, les charges Hückel servant de base au calcul de la charge atomique [269], [270], [271]. D'autres méthodes, telles que les charges Del Re, les charges Gasteiger et Marsili, les charges Hueckel et les charges Pullman, sont également disponibles dans le logiciel de modélisation moléculaire SYBYL [269], [272].

4.8.3 Analyse conformationnelle des molécules bioactives

L'analyse conformationnelle des composés est une étape très importante pour générer des conformations structurelles 3D adaptées à la modélisation 3D-QSAR. Afin de s'assurer qu'une analyse conformationnelle appropriée est fournie pour les composés à analyser, l'une des méthodes de recherche conformationnelle doit être appliquée (comme la recherche systématique, Monte Carlo, la recherche aléatoire, etc.) [273]. Toutes ces méthodes conduisent à la recherche de la structure géométrique 3D la plus stable de la molécule. À cette fin, nous avons appliqué, à l'aide du logiciel SYBYL-X2.1.1, un algorithme de recherche aléatoire qui génère un ensemble de conformations par la modification arbitraire et répétitive des coordonnées cartésiennes (x, y, z) ou internes (longueurs de liaison, angles de liaison et angles de torsion/dièdre) de la structure géométrique de la particule d'origine.

4.8.4 Descripteurs moléculaires identifiés par les méthodes de CoMFA et de CoMSIA

4.8.4.1 Analyse comparative des champs moléculaires (CoMFA)

La méthode CoMFA, qui a été la première technique utilisée dans le contexte du 3D-QSAR, est basée sur le calcul d'un champ stérique à l'aide de l'énergie potentielle de Lennard-Jones (LJ) 6-12 ou de van der Waals et d'un champ électrostatique (coulombien) entre et la sonde Csp^3 [274]. Ce champ a dû être limité en raison de l'augmentation brutale de l'énergie stérique à de courtes distances autour du vdWS. En raison de sa dureté, les résultats du CoMFA étaient très sensibles à l'orientation relative et à l'alignement précis des molécules, ainsi qu'aux conformères sélectionnés. En utilisant les potentiels de Lennard-Jones et de Coulomb, CoMFA est confronté à des anomalies au niveau des positions atomiques. Pour éviter cela, des valeurs limites des énergies stérique et électrostatique sont fixées à 30 kcal/mol [235].

4.8.4.2 Analyse comparative des indices de similitude moléculaire (CoMSIA)

Plus tard, d'autres champs stériques ont été développés, tels que ceux utilisés dans CoMSIA, qui remplacent la fonction de distance de l'énergie de LJ (6-12) par des fonctions de type Gaussien [275]. Ces champs stériques ne nécessitent pas de seuils de coupure et sont moins sensibles à la conformation et à la superposition des ligands. Dans le calcul des champs potentiels stériques et électrostatiques pour CoMFA, les termes de Lennard-Jones 6-12 et de Columbic sont appliqués respectivement avec une constante diélectrique dépendante de la distance de 1,0. Le remplacement de la fonction de distance 6-12 de l'énergie LJ par des fonctions de type Gaussien a permis de dériver cinq champs de CoMSIA, à savoir Stérique (S), Électrostatique (E), Hydrophobique (H), Donneur de liaison hydrogène (D) et Accepteur de liaison hydrogène (A), sur la base de l'atome de la sonde. Les indices stériques sont générés à partir de la troisième puissance des rayons atomiques, les descripteurs électrostatiques à partir des charges partielles atomiques, les champs hydrophobes à partir des paramètres basés sur les atomes, et les indices de donneur et d'accepteur de liaison

hydrogène sont calculés à l'aide d'une méthode basée sur des règles à partir des valeurs expérimentales [276].

4.9 Optimisation de la dimensionnalité des bases de données de descripteurs moléculaires générés

Le but de cette procédure est de réduire le nombre de variantes (c'est-à-dire le nombre de descripteurs moléculaires) par l'élimination des informations redondantes et non pertinentes. Plusieurs approches ont été proposées à cette fin durant ces dernières années [277], [278]. Parmi lesquelles la plus simple et la plus courante des procédures appliquées aux descripteurs moléculaires : (a) suppression des descripteurs ayant des valeurs manquantes ; (b) filtre à faible variance ; (c) forte corrélation. Suppression des descripteurs ayant des valeurs manquantes : Pour réduire les erreurs dans l'analyse des données et les algorithmes de modélisation, la suppression des descripteurs dont les valeurs manquent peut s'avérer nécessaire. Filtre à faible variance : Ceci est particulièrement utile lorsque l'ensemble de données est plus grand que le nombre de descripteurs et que le nombre de molécules avec des valeurs manquantes est faible. De plus, les descripteurs ayant une faible variance peuvent être éliminés, car ils contiennent peu d'informations sur les molécules analysées. Forte corrélation : Il convient de supprimer les descripteurs dont les valeurs sont fortement corrélées à celles d'autres descripteurs, car ils ne sont pas susceptibles d'apporter beaucoup d'informations nouvelles à l'ensemble existant de caractéristiques d'entrée. La réduction du nombre de descripteurs moléculaires dans la modélisation QSAR peut être bénéfique à plusieurs égards. Elle permet tout d'abord d'accroître la précision du modèle en réduisant sa complexité et en le rendant plus facile à interpréter. Deuxièmement, elle permet d'améliorer la performance du modèle, car un nombre réduit de descripteurs nécessite moins de temps de calcul. De plus, la réduction du nombre de descripteurs permet de limiter l'ajustement excessif du modèle, car un nombre réduit de descripteurs réduit le risque de capturer des relations non significatives dans les données. Il existe aujourd'hui plusieurs approches employées pour réduire la dimension des données, dont les plus utiles sont les suivantes : (a) l'analyse en composantes principales (ACP) [279], qui permet d'analyser des données multidimensionnelles à partir de variables quantitatives et de bien saisir la variable/corrélation entre les descripteurs. (b) Les techniques de classification/division (par exemple, partitionnement k-moyennes [280], Répartition univariée [281], et la classification par l'algorithme de la diversité de l'optimisme [282], etc.). Dans nos travaux, nous avons appliqué la méthode ACP pour réduire le nombre de descripteurs moléculaires possibles pour la construction de modèles QSAR. Pour diviser les données en deux sous-ensembles (apprentissage et test), nous avons utilisé partitionnement k-moyennes (K-means) ou univariée (Univariate Clustering) pour la modélisation des modèles 2D/3D-QSAR, et la classification par l'algorithme de la diversité de l'optimisme (OptiSim Diversity Algorithm) pour la modélisation 3D-QSAR en singulier.

4.9.1 Analyse en composantes principales

a) Définition

L'analyse en composantes principales (ACP) est l'une des méthodes statistiques d'exploration de données (*data mining*) les plus populaires [283]. Il s'agit d'une technique de réduction de la dimensionnalité appliquée pour réduire le nombre de variables dans un ensemble de données tout en conservant les informations les plus importantes. Cette méthode permet de trouver des modèles à partir des données et d'identifier les relations entre les variables. De plus, les données sont souvent utilisées pour réduire la dimensionnalité des grands ensembles de données et les rendre plus faciles à analyser. L'ACP peut également être exploitée pour visualiser les données et mettre en évidence des différences qui ne sont pas forcément visibles dans les données brutes.

b) Principe et la stratégie de la mise en œuvre

L'analyse en composantes principales (ACP) est une méthode statistique multivariée couramment utilisée en biostatistique, marketing, sociologie et d'autres domaines. Elle permet de projeter un ensemble de données initial composé de p variables (dimensions) dans un espace à k dimensions (où k < p) tout en préservant un maximum d'informations [284]. L'ACP trouve son utilité dans l'exploration de données complexes, permettant :

- L'étude et la visualisation des corrélations entre variables : L'ACP permet d'identifier les variables fortement corrélées et de les regrouper, simplifiant ainsi l'analyse.
- L'obtention de facteurs non corrélés pour les méthodes de modélisation : En réduisant la dimensionnalité des données, l'ACP fournit des facteurs non corrélés pouvant être utilisés dans des modèles statistiques ultérieurs.
- La visualisation des observations dans un espace à 2 ou 3 dimensions : Cette projection facilite l'identification de groupes homogènes ou d'observations atypiques au sein du jeu de données [285].

Sélection des descripteurs moléculaires pertinents pour la modélisation QSAR/QSPR :

Avant de procéder à l'ACP, il est crucial de sélectionner les descripteurs moléculaires les plus pertinents pour la construction du modèle QSAR/QSPR. Cette étape implique :

- Analyse de la contribution des descripteurs moléculaires : Évaluer la contribution individuelle de chaque descripteur moléculaire (variable indépendante) à la variance globale de l'ensemble de données. Les descripteurs peu importants sont éliminés à ce stade.
- Calcul de la matrice de corrélation des descripteurs : Calculer le coefficient de corrélation de Pearson (R) pour chaque paire de descripteurs restants. Les descripteurs présentant une forte corrélation (R > 0,9) sont éliminés afin d'éviter la multi colinéarité.
- Réduction dimensionnelle par ACP : Appliquer l'ACP pour réduire le nombre de descripteurs à un ensemble plus petit de composantes orthogonales capturant la majeure partie de la variance dans l'ensemble de données initial.
- Identification des composantes explicatives : Identifier les composantes ACP qui expliquent la plus grande partie de la variance (généralement > 60 %) dans l'ensemble de données. Ces composantes clés serviront à sélectionner l'ensemble final de descripteurs moléculaires candidats à corréler avec l'activité biologique ou la propriété physico-chimique (variable dépendante) dans le modèle QSAR/QSPR à développer.

4.9.2 Classification ou partitionnement des données

La classification ou le partitionnement des données moléculaires pour la modélisation QSAR revêt une importance cruciale pour garantir l'efficacité et la précision des prédictions. En divisant les données en sous-ensembles bien définis, nous pouvons plus facilement identifier les modèles et les relations, ce qui permet d'éviter le surajustement du modèle et de favoriser sa généralisation. Le partitionnement facilite également la validation interne et externe, essentielle pour évaluer les performances du modèle. De plus, il contribue à réduire la complexité informatique de la modélisation en divisant les données en sous-ensembles plus petits, ce qui permet une modélisation plus rapide et plus efficace. Globalement, il existe deux grandes techniques de partitionnement des données : l'apprentissage supervisé et l'apprentissage non supervisé[286]. L'apprentissage supervisé utilise des données étiquetées pour former un algorithme qui peut ensuite classifier de nouvelles données, tandis que l'apprentissage non supervisé cherche des modèles et des relations dans les données sans étiquettes ni catégories prédéfinies. Les exemples d'algorithmes d'apprentissage supervisé incluent les arbres de décision, les machines à vecteurs de support et les réseaux neuronaux, tandis que les exemples d'apprentissage non supervisé comprennent le partitionnement k-moyennes, la classification univariée, la classification hiérarchique, l'analyse en composantes principales et la classification par l'algorithme de la diversité de l'optimisme.

4.9.2.1 Partitionnement k-moyennes (k-means)

La méthode k-moyennes (ou k-means en anglais) est conçue pour diviser les données à deux voies et à deux modes (N objets et P variables) en K classes (C_1, C_2,...,C_K). La matrice de données est notée $X_{N\times P} = \{x_{ij}\}_{N\times P}$ et l'objectif de cette méthode est de construire des classes avec la plus petite distance euclidienne au carré entre le vecteur de ligne pour n'importe quel objet et le vecteur centroïde de sa classe respective [287]. Le centroïde de la classe C_k est calculé en prenant la moyenne de toutes les valeurs de chaque variable pour les objets de la classe. La méthode des k-moyennes est un processus itératif qui trouve la solution optimale correspondant au critère sélectionné lorsqu'il est exécuté plusieurs fois. Elle commence par assigner le centre des k classes à N objets et par calculer les distances entre les objets et les centres. Ensuite, les centres sont redéfinis et les objets réaffectés en fonction de leurs distances par rapport aux nouveaux centres. Les critères de classification pour le k-moyennes sont les suivants : la trace W, les matrices de classification SSCP groupées, le déterminant de W, les matrices de covariance internes groupées, le lambda de Wilks et Trace(W) /Médiane. Le logiciel XLSTAT propose de faire évoluer à la fois l'inertie intra-classe et le score de similarité afin de déterminer le nombre optimal de classes [288]. Les résultats sont présentés sous forme de tableaux et de graphiques. Après avoir introduit la matrice de données des composés de l'échantillon (N), ainsi que les descripteurs moléculaires (P) calculés pour ces composés, et après avoir effectué une analyse K-means, nous parvenons à la division optimale de ces données dans les groupes souhaités (selon la division 4/5 pour l'ensemble d'apprentissage et 1/5 pour l'ensemble de test). Parmi les éléments de chaque classe, nous choisissons un échantillon au hasard pour former l'ensemble de test, tandis que les éléments restants sont considérés comme l'ensemble d'apprentissage.

4.9.2.2 Partitionnement univarié (UV)

Lorsqu'il n'y a qu'une seule variable ou caractéristique, ou que les données sont peu nombreuses, le modèle univarié peut s'avérer un meilleur choix que le modèle k-means. En outre, si les données ne sont pas uniformément distribuées, la méthode de répartition univariée est préférable, car k-means repose sur l'existence d'une distribution uniforme. Le partitionnement univarié consiste à regrouper N observations unidimensionnelles (décrites par une seule variable quantitative) en k classes homogènes pour maximiser l'homogénéité des classes. Pour ce faire, nous cherchons à minimiser la somme des variances intra-classe. L'algorithme utilisé est très rapide et s'appuie sur la méthode proposée par W.D. Fisher (1958), qui peut être vue comme une discrétisation d'une variable quantitative en une variable ordinale [289]. Dans nos travaux, lorsque k-means ne peut pas assurer une partition appropriée pour élaborer des modèles 2D-QSAR ayant une signification statistique et un bon pouvoir prédictif, nous avons adopté la technique du partitionnement univarié comme solution alternative.

4.9.2.3 L'algorithme de sélection par la diversité

L'algorithme de sélection de la diversité implique une série d'étapes [282], [290]. À chaque étape : un sous-échantillon aléatoire de composés est choisi (K). La taille du sous-échantillon est spécifiée par vous. Notez que plus la taille du sous-échantillon est grande, plus le temps d'exécution est long et plus le sous-ensemble résultant est diversifié et moins représentatif. Avant qu'un composé soit accepté dans un sous-échantillon, il est comparé aux composés qui ont déjà été sélectionnés dans d'autres sous-échantillons lors des étapes précédentes. Tout composé dans le sous-échantillon qui est plus similaire (ou plus proche) qu'une valeur seuil à un des composés déjà sélectionnés est remplacé par un composé qui n'est pas. Ceci continue jusqu'à ce que la taille du sous-échantillon (K) soit atteinte. Du sous-échantillon, le composé qui est le moins similaire (ou le plus éloigné) de ceux déjà sélectionnés lors des étapes précédentes est choisi pour être ajouté au jeu de sélection. Le processus se termine lorsque le nombre demandé de composés a été sélectionné ou lorsque tous les composés non sélectionnés sont trop similaires (ou trop proches) à ceux qui ont été sélectionnés lors des étapes précédentes. Dans le cadre de nos travaux, cette méthode a été utilisée pour diviser de manière non aléatoire les données moléculaires en vue de construire des modèles 3D-QSAR fondés sur les descripteurs de champ CoMFA et CoMSIA. Cela est possible lorsque l'on part de la modélisation 3D-QSAR sans passer par la modélisation 2D-QSAR.

4.10 Modélisation 2D/3D-QSAR à l'aide des analyses biostatistiques prédictives

4.10.1 Stratégie globale pour la mise en œuvre

Après avoir appliqué des techniques de statistiques descriptives (ACP, K-means, UV et l'algorithme de sélection par la diversité) lors de la description et de la récapitulation des données et de la mise en évidence avec précision les schémas et les évolutions de l'ensemble des données. Les statistiques décisionnelles (également appelées statistiques prédictives) doivent alors être utilisées pour modéliser statistiquement les données quantitatives collectées et formuler des conclusions et des décisions relatives à la relation structure-activité fondée sur les informations stockées. La statistique décisionnelle

(prédictive) est utilisée pour tirer des conclusions sur les données à partir des informations (structure-activité) collectées [291], telles que les valeurs de l'activité biologique inhibitrice (IC_{50}, K_i, …etc.) ainsi que les descripteurs moléculaires calculés pour chaque molécule dans les espaces multidimensionnels de 0D-4D. Dans tous les travaux de 2D/3D-QSAR que nous avons réalisés dans cette thèse, nous avons utilisé des outils bio-informatiques comme XLSTAT, MATLAB et SYBYL-X2.1.1 afin de préparer et d'exploiter les données d'entrée nécessaires au développement de modèles 2D/3D-QSAR de haut pouvoir prédictif et statistiquement significatifs (**Tableau II.8**). L'analyse prédictive consiste à combiner des algorithmes, des données et l'apprentissage automatique afin de prévoir les exigences et les résultats futurs. Grâce à des modèles d'analyse sophistiqués, les différentes entités peuvent optimiser leur usage des ressources et des données afin de développer une projection optimale de la manière dont la variable de réponse est liée aux variables indépendantes. L'utilisation de cette relation aide les utilisateurs à optimiser la réponse au sein des variables indépendantes selon leurs besoins. Dans ce contexte, l'industrie pharmaceutique s'est largement inspirée de cette approche pour prédire et modéliser la connaissance de SAR dans ses projets, notamment dans la conception et la découverte de médicaments. Plus spécifiquement, l'utilisation de ce type de modélisation biostatistique dans le développement de modèles mathématiques de type QSAR. En effet, ce type de statistiques facilite la prise de décision informée sur les caractéristiques/propriétés structurelles des composés les plus aptes à avoir un effet favorable ou défavorable en vue de l'activité biologique souhaitée ainsi que sur le choix des meilleures substances actives. Dans cette thèse, nous avons adopté différentes méthodes et techniques biostatistiques pour modéliser quantitativement la relation structure-activité (SAR) et pour interpréter l'activité biologique des molécules basées sur des descripteurs moléculaires hybrides (0D-4D). À cette fin, nous avons appliqué des modèles mathématiques linéaires et non linéaires pour la modélisation 2D-QSAR et des méthodes dépendant de l'alignement pour la modélisation 3D-QSAR. Pour développer des modèles QSAR linéaires, nous avons utilisé des techniques de régression linéaire multiple (MLR) et de moindres carrés partiels (PLS). Les techniques non linéaires basées sur les algorithmes de régression non linéaire (MNLR) et les réseaux neuronaux artificiels (RNA/ANN) sont adoptés comme procédés de renforcement et d'amélioration du pouvoir prédictif des modèles 2D-QSAR proposés. Parallèlement, nous avons utilisé la méthode d'alignement moléculaire pour développer des modèles 3D-QSAR sur la base d'une méthodologie hybride fondée sur les approches de l'analyse comparative des champs moléculaires (CoMFA) et de l'analyse comparative de l'indice de similarité moléculaire (CoMSIA). La méthode des moindres carrés partiels (PLS) est utilisée pour explorer la corrélation linéaire entre les champs CoMFA et CoMSIA et les valeurs d'activité biologique (pIC_{50}).

Tableau II.8. Outils de préparation des données d'entrée et de mise en œuvre des modèles QSAR utilisés dans nos travaux.

Objectif	Méthode	Application	Outil
Réduction de dimensionalité	ACP	2D et 3D-QSAR	XLSTAT [292]
Classification/Partitionnement des données	Partitionnement k-moyennes (k-means)	Combinaison des méthodes 2D-QSAR et 3D-QSAR	
	Partitionnement univarié (UV)	Combinaison des méthodes 2D-QSAR et 3D-QSAR	
	L'algorithme de sélection par la diversité	3D-QSAR	SYBYL X2.1.1 [293]
Statistiques prédictives	RLM	2D-QSAR	XLSTAT
	RNLM		
	RNA		MATLAB
	PLS	3D-QSAR	SYBYL X2.1.1

La **Figure II.10** résume les stratégies de prévision biostatistique que nous avons adoptées pour développer les modèles 2D/3D-QSAR élaborés dans le cadre de cette thèse.

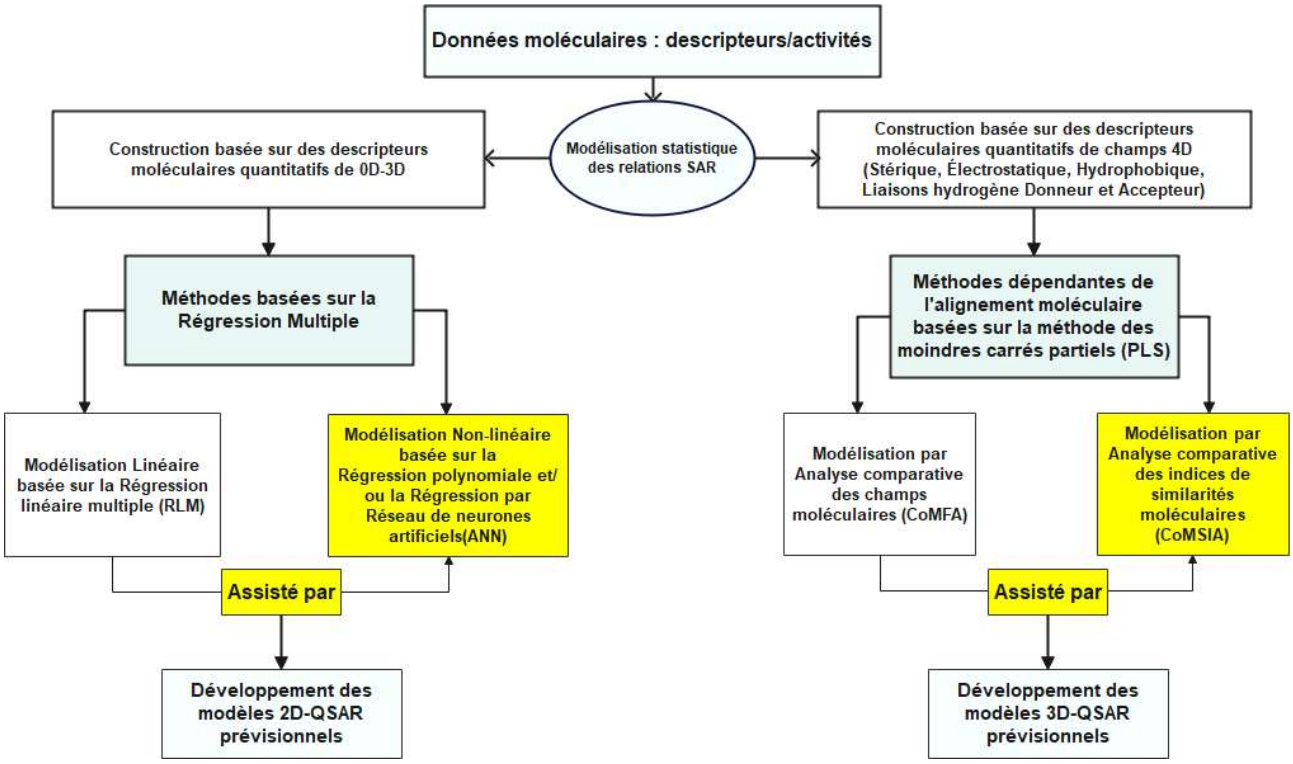


Figure II.10. Schéma des méthodes de modélisation prédictives biostatistiques appliquées dans nos travaux pour développer des modèles 2D/3D-QSAR.

4.10.2 Modélisation 2D-QSAR à l'aide des méthodes de Régression Multiple

Après la simplification computationnelle de la structure des molécules en descripteurs moléculaires multidimensionnels (de 0D à 3D), on passe à l'étape de la modélisation statistique de la relation quantitative structure-activité en 2D. Cette étape vise à identifier les caractéristiques/propriétés structurelles (descripteurs moléculaires) ayant une relation plus forte avec l'activité biologique des molécules examinées. L'identification de ces caractéristiques facilite l'obtention d'une vue statistique rationnelle rapide et plus compréhensible sur les descripteurs moléculaires favorables et défavorables, en vue d'améliorer l'activité biologique des nouvelles molécules sur la base de leur structure. Les techniques les plus couramment utilisées pour développer des modèles 2D-QSAR sont la Régression

Linéaire Multiples (RLM) et la Régression Non Linéaire Multiples (RNLM), ainsi que les Réseaux de Neurones Artificiels (RNA ou ANN). Les modèles logiques basés sur la régression linéaire multiple servent à analyser la corrélation linéaire entre plusieurs descripteurs moléculaires du composé et son activité biologique, tandis que la régression non linéaire multiple permet d'analyser la corrélation non linéaire entre les deux. D'autre part, les RNA sont utilisés pour développer des modèles plus complexes capables d'apprendre à partir des données et de reconnaitre des modèles. Les modèles 2D-QSAR basés sur les techniques MLR, MNLR et RNA peuvent être utilisés pour prédire avec précision la activité biologique d'un composé sur la base de ses caractéristiques structurelles [294].

4.10.3 Régression Linéaire Multiples (RLM)

Les algorithmes de régression linéaire multiple permettent de prédire une variable dépendante en fonction de nombreuses autres variables indépendantes, sur la base d'une relation linéaire établie par l'algorithme d'apprentissage utilisé à cette fin. Parmi les nombreux algorithmes développés à cette fin, on trouve la méthode des moindres carrés ordinaires (MCO), la descente de gradient, la descente de gradient stochastique (SGD), l'élimination à rebours, ...etc. [295]. Les principaux points communs de ces algorithmes sont leur objectif commun de minimiser une fonction de coût et de trouver les paramètres les plus optimaux pour un modèle linéaire donné. De plus, ils peuvent tous être utilisés pour construire des modèles linéaires mathématiques et implémenter la technique de régression linéaire multiple. Les techniques de la RLM sont largement appliquées à la modélisation QSAR et à la sélection de descripteurs moléculaires appropriés (variables indépendantes) destinés à modéliser l'activité biologique des molécules (variables dépendantes) selon la corrélation décrite dans **1'Eq. (II.6)**.

$$' = a_{} + \sum_{i=1}^{n} a_i X_i \qquad \text{(II.6)}$$

où Y correspond à la variable dépendante (activité biologique à prédire), X_i désigne les variables indépendantes (descripteurs moléculaires multiples), n est le nombre de descripteurs moléculaires ; a_0 désigne la constante du modèle de RLM; a_i indique les coefficients associés à chaque descripteur dans l'équation du modèle RLM-QSAR. Dans nos travaux relatifs à la modélisation 2D-QSAR via MLR, nous avons appliqué l'algorithme de régression basé sur les Moindres Carrés Ordinaires (MCO) dans le cadre de l'apprentissage supervisé réalisé à l'aide du logiciel XSLSTAT 2019.

4.10.4 Régression Non Linéaire Multiples (RNLM)

La régression non linéaire multiple est une méthode utile pour modéliser les relations non linéaires entre des variables indépendantes et une variable dépendante, avec des modèles tels que les exponentiels, logarithmiques, polynomiaux et les Réseaux de Neurones Artificiels (RNA) [100], [294]. Les modèles exponentiels, logarithmiques, polynomiaux et RNA sont tous utilisés lorsque la relation entre les variables indépendantes et la variable dépendante suit une forme spécifique. Ces modèles sont particulièrement adaptés lorsque les variables indépendantes présentent une large gamme de valeurs. Ces modèles sont utiles dans un large éventail d'applications, notamment la modélisation

QSAR, en offrant une meilleure compréhension des relations entre les variables et une plus grande précision des prédictions. Les modèles linéaires et non linéaires peuvent être combinés pour obtenir des résultats prédictifs améliorés. Dans ce contexte, nous avons inclus dans nos travaux réalisés dans le cadre de cette thèse les approches basées sur les modèles polynomiaux et RNA dans la construction de modèles QSAR non linéaires, en raison des considérations suivantes :

- Les modèles polynomiaux et RNA sont préférés aux modèles exponentiels et logarithmiques pour la construction de modèles QSAR non linéaires car ils sont plus flexibles et plus généralement applicables.
- Les modèles polynomiaux permettent un plus grand degré de complexité, ce qui est important pour décrire avec précision les relations non linéaires complexes entre les descripteurs et l'activité.
- Les modèles RNA sont également plus flexibles et permettent de modéliser avec précision des relations non linéaires complexes.
- Ils présentent également l'avantage de pouvoir être entraînés sur de grands ensembles de données, ce qui peut améliorer la précision et la généralisation du modèle.

En revanche, les modèles exponentiels et logarithmiques sont limités dans leur complexité et sont donc moins adaptés à la modélisation précise des relations non linéaires [296]. Dans ce contexte, nous utilisons le modèle mathématique polynomial du second ordre pour élaborer le modèle RNLM-QSAR, sur la base des descripteurs identifiés par le modèle RLM. En utilisant la modélisation des données par régression non linéaire multiple intégrée dans le package de XLSTAT, la représentation numérique des relations non linéaires entre les descripteurs moléculaires et l'activité biologique peut être modélisée conformément au modèle polynomial du second ordre montré par l'**Eq. (II.7).**

$$Y = a_0 + \sum_{i=1}^{n} a_i \times X_i + b_i \times X_i^2 \qquad \text{(II.7)}$$

où Y représente la variable dépendante (l'activité biologique à évaluer); X_i sont les variables indépendantes (les descripteurs moléculaires); *n* est le nombre de descripteurs moléculaires; a_0 désigne la constante de l'équation du modèle, a_i et b_i sont les coefficients associés à chaque descripteur dans l'équation du modèle RNLM-QSAR polynomial.

4.10.4.1 Réseaux de Neurones Artificiels (ANNs ou RNA)

Le sigle ANNs en anglais signifie "Artificial Neural Networks" ("Réseaux de neurones Artificiels"). ANNs ou RNA sont une sorte de modèle d'apprentissage automatique qui s'inspire de la structure et de la fonction biologiques du cerveau humain [297] (**Figure II.11a).** L'idée des réseaux de neurones artificiels commence en 1943 [298], lorsque Warren McCulloch et Walter Pitts ont proposé un modèle mathématique de neurone, connu sous le nom de neurone McCulloch-Pitts (**II.11b**). L'idée consistant à simuler l'intelligence humaine à l'aide du modèle ANN commence à germer en 1943. Le modèle de neurone McCulloch-Pitts calcule la somme pondérée des entrées, la compare à une valeur seuil et émet un 1 ou un 0 selon que le seuil est dépassé ou non [299]. Le neurone McCulloch-Pitts est souvent

utilisé comme élément de base des réseaux neuronaux artificiels et son développement a servi de fondement à l'élaboration de modèles de réseaux neuronaux plus sophistiqués.

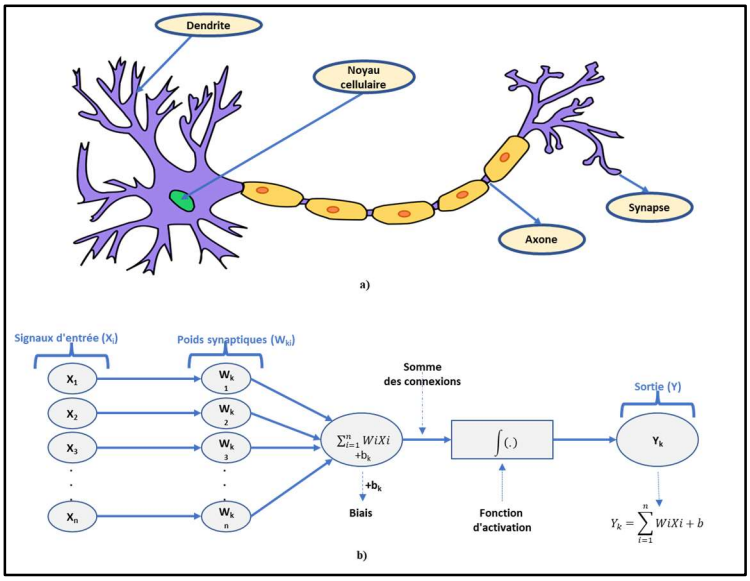


Figure II.11. (a) Le neurone biologique, dans ce modèle les dendrites des réseaux neuronaux biologiques symbolisent les entrées dans les réseaux neuronaux artificiels (ANNs), le noyau cellulaire représente les nœuds, les synapses représentent les poids et l'axone représente la sortie ; (b) Modèle de neurones McCulloch-Pitts.

a) Historique du développement de ANNs [94]

En 1958, Rosenblatt a développé le Perceptron, le premier réseau neuronal artificiel capable d'apprendre par l'expérience. Le modèle du Perceptron comprend une couche d'entrée et une couche de décision. En 1960, Widrow et Hoff ont proposé le modèle Adaline, qui est devenu la base des réseaux neuronaux multicouches. Cependant, en 1969, Minsky et Papert ont révélé les limites des réseaux neuronaux à une seule couche, notamment leur incapacité à traiter des problèmes non linéaires. En 1982, Hopfield a proposé les neurones de type associatif ; la même année, Werbos a développé l'algorithme de rétropropagation, qui fournit un processus d'entraînement des couches cachées des réseaux multicouches de type Perceptron. En 1984, Kohonen a introduit l'apprentissage non supervisé à l'aide de cartes auto-organisatrices ; par la suite, les machines de Boltzmann ont été découvertes un an plus tard. Puis, en 1989, Moody et Darken ont proposé les réseaux à base radiale (Radial Basis Function).

b) Classification de modèles ANNs selon leur application

De nos jours, conformément aux développements des méthodes de programmation, de l'ingénierie informatique et de la modélisation computationnelle basée sur la technologie informatisée, davantage d'améliorations sont apportées aux modèles traditionnels d'ARN pour les rendre plus performants et plus adaptés à l'analyse des phénomènes naturels (biologiques, physiologie du cerveau, ...) et des problèmes complexes. Les applications les plus importantes des ANNs peuvent être récapitulées comme suit : (1) statistiques (analyse et classification des données, construction de modèles prédictifs) ; (2) ingénierie robotique (contrôle et guidage de robots ou de véhicules autonomes) ; (3) reconnaissance et traitement d'images et de formes ; (4) traitement du signal et simulation de l'apprentissage automatique. Dans ce contexte, les réseaux neuronaux artificiels sont classés en

plusieurs catégories selon l'application visée, notamment le réseau de neurones à propagation avant (Feedforward neural network "FFNN") [300], les réseaux neuronaux récurrents et les réseaux neuronaux convolutifs se révèlent les modèles les plus simples et les plus appropriés pour la modélisation statistique. Le réseau FFNN est un outil puissant pour résoudre un large éventail de problèmes, tels que la reconnaissance d'images et de la parole, le traitement du langage naturel et les tâches de prédiction et d'analyse des données. Sa simplicité et sa polyvalence en ont fait l'un des types de réseaux neuronaux les plus utilisés dans les applications pratiques. Ils se composent d'une couche d'entrée, d'une ou plusieurs couches cachées et d'une couche de sortie. Les réseaux neuronaux récurrents sont conçus pour traiter des blocs de données, tels que des données linguistiques ou des séries temporelles. Les réseaux neuronaux convolutifs sont spécialisés dans le traitement des images et d'autres types de données à structure spatiale. Dans cette thèse, nous avons appliqué la méthode des réseaux de neurones « Feedforward » (**Figure II.12**) pour élaborer des modèles ANN-QSAR non linéaires afin d'améliorer la performance des modèles de RLM-QSAR. Ceci est en accord avec la pertinence de l'approche KKNN avec l'analyse et la modélisation statistique des données.

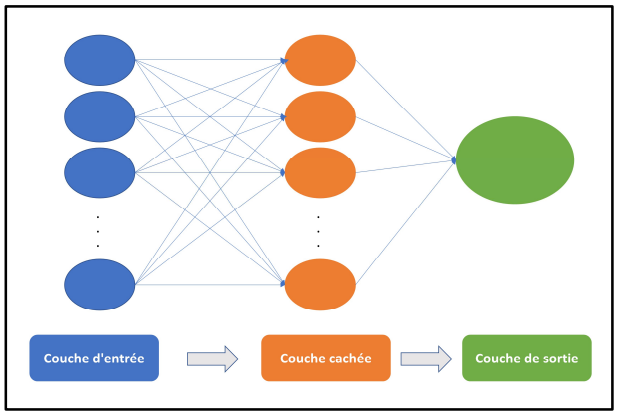


Figure II.12. Architecture du modèle de réseau FFNN.

4.10.4.2 Réseau de neurones à propagation avant (FFNN)

a) Principe

Le réseau neuronal feedforward (FFNN) est un type de réseau neuronal artificiel dans lequel le flux d'informations circule dans une seule direction, de la couche d'entrée à la couche de sortie. Le terme "feedforward" fait référence au fait que la sortie de chaque couche sert d'entrée à la couche suivante sans aucune boucle de rétroaction. Le FFNN se compose de trois types de couches : la couche d'entrée, une ou plusieurs couches cachées et la couche de sortie (**Figure II.12**). Chaque couche est composée d'un ensemble de neurones ou de nœuds qui effectuent une opération mathématique simple sur leurs entrées et génèrent une sortie. Les neurones de la couche d'entrée reçoivent les données d'entrée et les transmettent aux neurones de la première couche cachée. Chaque neurone de la couche cachée reçoit des entrées de tous les neurones de la couche précédente et applique une fonction d'activation non linéaire à la somme pondérée de ses entrées [301]. La sortie de la dernière couche cachée est ensuite transmise à la couche de sortie, où chaque neurone applique une autre fonction d'activation à la somme pondérée de ses entrées pour produire la sortie finale du réseau. Ici, dans nos travaux pour entraîner et tester le modèle ANN-QSAR fondé sur la méthode

FFNN, nous avons utilisé la fonction d'entraînement de Levenberg-Marquardt dans la couche d'entrée, la fonction de transfert sigmoïdale dans la couche cachée et la fonction de transfert linéaire dans la couche de sortie. Par la suite, le modèle FFNN est formé à l'aide d'un processus appelé rétropropagation, au cours duquel le réseau ajuste ses poids afin de minimiser la différence entre la sortie prédite ($Y_{calculé}$) et la sortie réelle ($Y_{observé}$) [302]. Pendant la formation, les poids des connexions (W_i) entre les neurones sont ajustés en propageant itérativement l'erreur à travers le réseau, en utilisant la règle de la chaîne de calcul pour calculer la contribution de chaque neurone à l'erreur.

b) Mise en œuvre

Pour développer les modèles ANN-QSAR basés sur l'approche FFNN dans nos travaux, on a utilisé le logiciel MATLAB 2015a, avec les étapes suivantes :

- Collectez les données d'entrée (descripteurs moléculaires calculés) et de sortie (données expérimentales sur l'activité biologique) pour l'entraînement du réseau.
- Créez le modèle FFNN à l'aide de la fonction "feedforwardnet" dans MATLAB. Cette fonction permet de spécifier le nombre de couches cachées et le nombre de neurones dans chaque couche. Ici, il convient de noter que la détermination du nombre de neurones cachés joue un rôle important dans la définition de l'architecture optimale du réseau. Le meilleur nombre de couches cachées conduit au meilleur modèle, certains auteurs ayant suggéré le paramètre **ρ** ($\frac{\text{Nombre total des données de l'ensemble d'entraînement}}{\text{Somme de nombre de connexions}}$) comme critère de choix du nombre de couches cachées utilisées lors de l'élaboration du modèle ANN-QSAR. Dès lors, il est recommandé que la valeur de ρ reste dans la plage des valeurs comprises entre 1 et 3, ceci afin d'éviter tout risque de surajustement ou de sous-ajustement [303]. Pour calculer la valeur de ρ, la relation simplifiée dans l`**Eq (II.8)** est appliquée.

$$\rho = \frac{\sum_{i=1}^{N} W_i}{\sum (CE \times CC) + (CC \times CS) + CC + CS} \qquad \text{(II.8)}$$

où, Wi est la somme du nombre de poids des éléments de l'ensemble d'entraînement (N); CE (couche d'entrée) ; CC (couche cachée) ; CS (couche de sortie).

- Définissez les paramètres d'entraînement pour le réseau, tels que la méthode d'optimisation, le nombre maximal d'itérations, le critère d'arrêt, etc.
- Divisez les données d'entrée et de sortie en ensembles d'entraînement (70%), de validation (15%) et de test (15%) à l'aide de la fonction "divideblock".
- Entraînez le réseau en utilisant la fonction "train" avec les données d'entraînement. Cette fonction ajuste les poids des neurones du réseau en fonction des données d'entrée et de sortie.
- Évaluez les performances du réseau sur l'ensemble de validation à l'aide de la fonction "sim".
- Utilisez le réseau entraîné pour prédire les sorties correspondantes aux données d'entrée de l'ensemble de test à l'aide de la fonction "sim".
- Analysez les performances du réseau en comparant les sorties prédites aux sorties réelles.

En résumé, pour implémenter un modèle FFNN-QSAR dans MATLAB, il faut créer le modèle, définir les paramètres d'entraînement (les molécules choisies pour entraîner le modèle ANN-QSAR), diviser les données en ensembles d'entraînement, de validation et de test, entraîner le réseau, évaluer les performances sur l'ensemble de validation et tester le réseau sur l'ensemble de test.

4.10.5 Modélisation 3D-QSAR à l'aide des méthodes dépendantes de l'alignement moléculaire

Les études QSAR (Quantitative Structure-Activity Relationship) visent à expliquer les raisons des différences de réponse biologique observées entre les composés d'une série en les mettant en corrélation avec leurs propriétés structurelles. Les techniques QSAR traditionnelles établissent une corrélation entre les réponses biologiques et les propriétés atomiques, de groupe ou moléculaires ou certaines caractéristiques structurelles. Toutefois, elles ne tiennent pas compte des structures tridimensionnelles (3D) des molécules, ce qui limite leur utilité pour la conception de nouvelles molécules. La modélisation 3D-QSAR est apparu comme une adaptation artificielle des techniques traditionnelles, utilisant des outils chimio informatiques pour corréler les propriétés macroscopiques de la cible avec les descripteurs calculés sur la base du modèle atomique, dérivés des propriétés 3D des ligands [235]. L'application de la méthode 3D-QSAR rend la conception rationnelle de produits pharmaceutiques et agrochimiques plus sophistiqués, le criblage de bases de données moléculaires et la génération d'hypothèses pharmacopées liées à la présence ou à l'absence d'activité biologique des molécules. Cette technique peut également être appliquée à la modélisation des propriétés optoélectroniques des molécules. Toutes ces applications citées ont été confirmées dans le cadre de nos travaux menés pendant la préparation de cette thèse, par l'intégration de la méthode 3D-QSAR dans le cadre de la stratégie de modélisation rationnelle et de conception biomoléculaire proposée ici [217], [304], [305].

4.10.5.1 Principe et classification des techniques de 3D-QSAR

La méthodologie 3D-QSAR est plus complexe et nécessite plus de calculs que les approches 2D-QSAR. Elle comporte plusieurs étapes pour obtenir des descripteurs numériques à partir des structures 3D des composés, notamment la détermination de la conformation biologiquement active grâce à des méthodes expérimentales ou théoriques, la génération d'un alignement de conformères dans l'espace 3D, la recherche computationnelle de descripteurs dans l'espace et la corrélation de ces descripteurs avec les réponses biologiques observées expérimentalement. Les modèles QSAR ne remplacent pas les méthodes expérimentales, mais constituent plutôt un outil pour compléter ces dernières. Les chercheurs en QSAR s'efforcent de développer des modèles qui représentent fidèlement le modèle réel. Les techniques 3D-QSAR reposent sur certaines hypothèses de base, notamment que la conformation biologiquement quasi active d'un composé est la même que la conformation dans le système biologique et que les composés étudiés ont des mécanismes de liaison similaires [306]. La classification des méthodes 3D-QSAR est basée sur plusieurs critères, comme le montre le **Tableau II.9** [235]. Dans le cadre de nos travaux menés dans cette thèse, nous avons adopté les techniques

CoMFA et CoMSIA les plus courantes, basées sur le critère d'alignement dépendant pour générer des modèles 3D-QSAR.

Tableau II.9. Classification des techniques 3D-QSAR selon des critères spécifiques[235].

Critères de classification	Méthodes	Techniques
Modélisation intermoléculaire	Basé sur le ligand	CoMFA, CoMSIA, MSA, RSA, Compass, GERM, CoMMA, SoMFA
	Basé sur le récepteur	AFMoC, HIFA
Modélisation chimiométrique	Linéaire	CoMFA, CoMSIA, AFMoC, …
	Non-linéaire	Compass,
Critère d'alignement	Dépend de l'alignement	CoMFA, CoMSIA, MSA, …
	Indépendant de l'alignement	Compass, CoMMA, HQSAR, WHIM, GRIND, VolSurf, CoSA,…

4.10.5.2 Méthodes dépendant de l'alignement

L'un des facteurs les plus importants influençant la qualité des modèles 3D-QSAR est la qualité de l'alignement moléculaire. SYBYL-X.2.1.1 propose plusieurs stratégies d'alignement moléculaire en fonction des caractéristiques de chaque base de données. Parmi ces stratégies, on trouve l'alignement par sous-structure (Distillation rigide, Alignement rigide), l'alignement par pharmacophore (par exemple, DISCOtech, GALAHAD, et GASP) ou l'alignement par Docking (Surflex-Dock) [307]. Il existe également d'autres options d'alignement si l'une des stratégies mentionnées ne fonctionne pas (Fonctionnalité Fit Atoms, Fonctionnalité Match Atoms, Multifit s'adapte à l'atome, Surflex-Sim). Dans nos travaux, nous considérons la sous- structure comme la méthode la plus simple, la plus transparente et la plus proche du contexte de nos préoccupations, à savoir la modélisation de l'activité biologique de nouvelles molécules ayant une structure de base commune.

4.10.5.3 L'alignement par sous-structure

Cette méthode repose sur le principe que les molécules partagent un cœur commun d'atomes. Ce cœur se superpose à chaque molécule de l'ensemble de données. Par conséquent, les atomes qui ne se superposent pas doivent agir de manière raisonnable pour se rapprocher et s'aligner sur la molécule de référence (la plus active). Une solution simple consiste donc à relaxer les liaisons latérales pour les amener dans leur conformation complètement étendue. Ceci peut être atteint par l'application de la propriété de distillation rigide (Distill Rigid) à la base de données qui doit être alignée. L'alignement basée sur la distillation rigide permet d'ajuster les angles de torsion des liaisons rotatives dans le noyau commun de chaque molécule pour qu'ils correspondent à ceux du composé de référence. Contrairement à la fonction d'alignement rigide, qui est plus stricte en ce qui concerne ce genre de caractéristiques. C'est pourquoi nous avons adopté la distillation rigide pour aligner les molécules de la base de données moléculaires cible sur les molécules de référence les plus actives figurant dans chaque série. La **Figure II.13** présente un exemple représentatif de l'alignement moléculaire d'une base de données de molécules effectué par la méthode de distillation rigide dans le cadre de l'un de nos travaux.

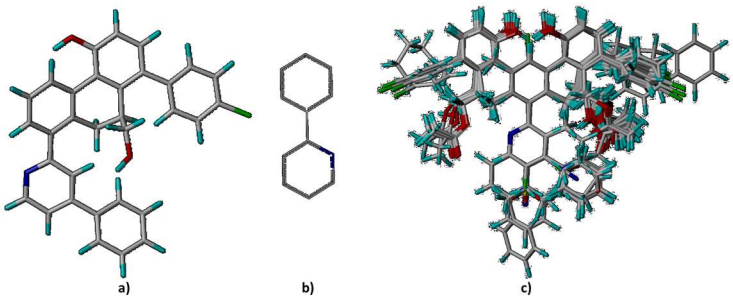


Figure II.13. (a) Molécule de référence, (b) Cœur commun, (c) Base de données alignée.

Après avoir terminé l'alignement de la base de données avec le composé de référence, la base de données ajustée est ensuite divisée en deux ensembles : un ensemble d'entraînement (80 %) et un ensemble de test (20 %), conformément à la méthodologie décrite au section (**4.3.2**). Puis, sur la base des approches CoMFA et CoMSIA, nous procédons à la construction d'un modèle 3D-QSAR.

4.10.5.4 Analyse Comparative du Champ Moléculaire (CoMFA)

a) Principe

L'analyse comparative des champs moléculaires (CoMFA) est une méthode basée sur les champs moléculaires, basée sur les ligands et dépendante de l'alignement, développée par Cramer et al [213]. Dans le but d'établir une relation quantitative entre les structures moléculaires et leurs propriétés en termes de réponse biologique.

b) Mise en œuvre

L'approche de CoMFA repose principalement sur les propriétés des ligands, telles que les interactions stériques et électrostatiques, et sur leur impact sur les interactions récepteur-ligand. Dans le cadre du processus CoMFA, les ligands alignés sont placés dans une grille énergétique et l'énergie à chaque point du réseau est calculée à l'aide d'une sonde appropriée (la dimension usuelle de l'espacement de la grille est de 2 Å ; la dimension typique de la boîte de la grille est de 3-4 Å) [235]. La sonde peut être une petite molécule (H_2O, par exemple) ou un groupement fonctionnel. Le calcul des énergies électrostatiques se fait à l'aide d'une sonde H^+, tandis que le calcul des énergies stériques se fait à l'aide d'un atome de carbone hybridé sp^3 ayant un rayon effectif de 1,53 Å et une charge de 1,0. Les valeurs de ces champs de force calculées servent de descripteurs 4D dans le cadre du développement du modèle 3D-QSAR à l'aide d'une méthode de régression linéaire robuste telle que la méthode des moindres carrés partiels (PLS) que nous avons utilisée dans la modélisation 3D-QSAR dans nos travaux. Les résultats générés par la méthode PLS permettent de déterminer les potentiels stériques et électrostatiques favorables et défavorables associés à la réponse biologique évaluée [308]. La **Figure II.14** présente un exemple illustratif des potentiels stériques et électrostatiques qui ont été définis sur la structure de la molécule de référence observée dans l'un de nos travaux [309].

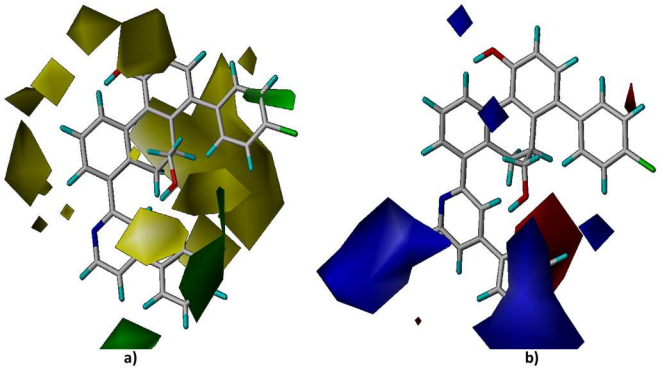


Figure II.14. Cartes de contours générées par le modèle CoMFA, (a) champ d'interactions stériques (vert = favorable/jaune = défavorable), (b) champ d'interactions électrostatiques (bleu = favorable/rouge = défavorable).

4.10.5.5 Analyse Comparative des Indices de Similarité Moléculaire (CoMSIA)

a) Principe

L'approche CoMSIA est une version modifiée de CoMFA qui utilise des potentiels Gaussiens pour analyser le potentiel stérique et électrostatique appliqué à la conception de nouveaux médicaments. CoMSIA calcule également la similarité moléculaire pour résoudre les problèmes de sensibilité à l'alignement et inclut un terme entropique moléculaire dépendant du solvant pour tenir compte des caractéristiques de nature hydrophobique.

b) Mise en œuvre

Pour analyser un ensemble de molécules, une sonde commune est placée et les similarités à chaque point de grille sont calculées pour les propriétés stériques (S), électrostatiques (E), hydrophobique (H) et de liaison hydrogène donneur et accepteur (D et A), ces dernières devant être corrélées avec la réponse biologique. La méthodologie appliquée pour générer des modèles 3D-QSAR basés sur les descripteurs CoMSIA est la même que celle basée sur les descripteurs CoMFA. Avec certaines adaptations nécessaires : calcul de l'interaction moléculaire sur la base des cinq propriétés physico-chimiques (SEHDA) en utilisant un atome sonde commun avec un rayon de 1 A˚, une charge de 1, une hydrophobie de 1 et des propriétés de donneur et d'accepteur de liaisons hydrogène de 1. Puis, on utilise l'approche PLS pour générer des modèles 3D-QSAR à l'aide du coefficient de similarité (CoMSIA) comme variables indépendantes et de la réponse biologique (pIC_{50}) comme variable dépendante. Cette analyse permet d'identifier les potentiels SEHDA favorables et défavorables associés à l'activité biologique pIC_{50} des molécules. La **Figure II.15** propose un exemple illustratif des potentiels SEHDA identifiés sur la structure de la molécule de référence étudiée dans le cadre de l'un de nos travaux [309].

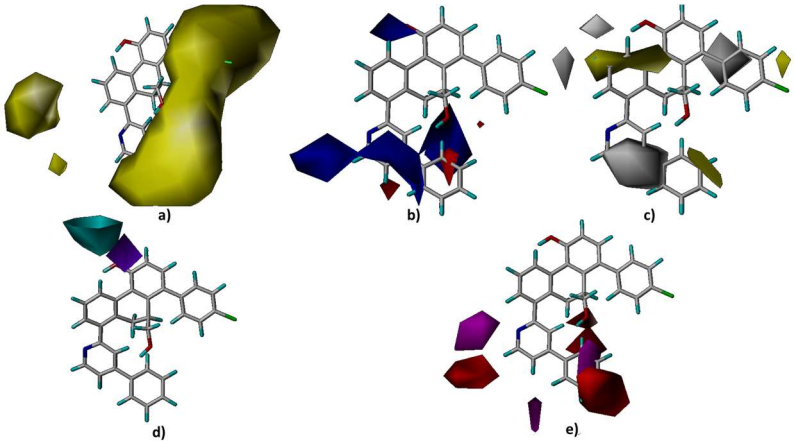


Figure II.15. Cartes de contours générées par le modèle CoMSIA, (a) Interactions de champ stérique (vert = favorable/jaune = défavorable), (b) Interactions de champ électrostatique (bleu = favorable/rouge = défavorable), (c) Interactions de champ hydrophobe (jaune = favorable / blanc = défavorable), (d) Interactions de champ de liaisons hydrogène donneur (cyan = favorable / violet = défavorable), (e) Interactions de champ de liaisons hydrogène accepteur (magenta = favorable / rouge = défavorable).

4.10.5.6 Avantages et limites de CoMFA et CoMSIA

Les approches de CoMFA et CoMSIA sont couramment utilisés dans la conception et le développement de nouveaux médicaments ; pour prédire les propriétés de nouveaux ligands et optimiser leur structure en fonction de réponses biologiques ciblées. CoMFA peut fournir des informations sur les caractéristiques structurelles 3D des ligands qui sont cruciales pour l'activité biologique, mais il suppose que la relation entre la structure et l'activité est uniquement basée sur les interactions stériques et électrostatiques, ce qui peut ne pas être le cas pour tous les ligands. D'autre part, CoMSIA peut intégrer dans l'analyse des propriétés moléculaires plus importantes telles que l'hydrophobicité et la liaison hydrogène, ce qui peut améliorer sa précision, mais elle peut être coûteuse en termes de calcul et de temps, et sa précision dépend de la qualité des données utilisées pour l'analyse. Les deux méthodes ont leurs propres avantages et obstacles qui doivent être considérés pour leur application optimale dans le cadre de l'optimisation des ligands [310]. Par conséquent, en vue de proposer un modèle 3D-QSAR plus rationnel offrant une précision et une fidélité optimales dans le cadre du processus de conception et d'optimisation d'un nouveau médicament, il est préférable de compléter les analyses CoMFA par des analyses CoMSIA. Cette démarche peut contribuer à dépasser les principaux défis posés par les analyses CoMFA. Dans ce contexte, notre stratégie de construction d'un modèle 3D-QSAR robuste et fiable est née de la combinaison entre les analyses de CoMFA et CoMSIA.

4.10.5.7 Étapes globales de la modélisation CoMFA/CoMSIA

Dans nos travaux, toutes les procédures liées au dessin et à la préparation des structures 3D des molécules, l'optimisation géométrique et énergétique, le calcul des charges atomiques partielles, l'analyse conformationnelle des composés, alignement et division de la base de données moléculaires, calcul des descripteurs CoMFA et CoMSIA et les corréler avec l'activité biologique observée expérimentalement (pIC_{50}) par le biais d'analyses PLS, validation interne et externe des modèles 3D-QSAR sélectionnés, et prédiction de l'activité biologique des nouvelles molécules bioactives modélisées ont été effectuées à l'aide du logiciel SYBYL-X2.1.1 basé sur le champ de force standard

Tripos Powell. Certains paramètres sont ajustés manuellement, comme le nombre minimum de cycles d'optimisation (généralement varié entre 100 et 2000 cycles dans nos travaux) ; la valeur de la moyenne quadratique (0.01<RMS<0.05 kcal/mol) adoptée comme critère dans la fonction diélectrique pour le calcul des charges atomiques partielles par la méthode de Gasteiger-Huckel. Le diagramme présenté dans la **Figure II.16** illustre le résumé des phases principales de la stratégie de modélisation 3D-QSAR que nous avons adoptée dans nos travaux en vue de proposer des modèles prédictifs *in silico* basés sur les approches CoMFA/CoMSIA susceptibles de simuler des modèles de criblage réels basés sur les évaluations biologiques réalisées *in vitro*.

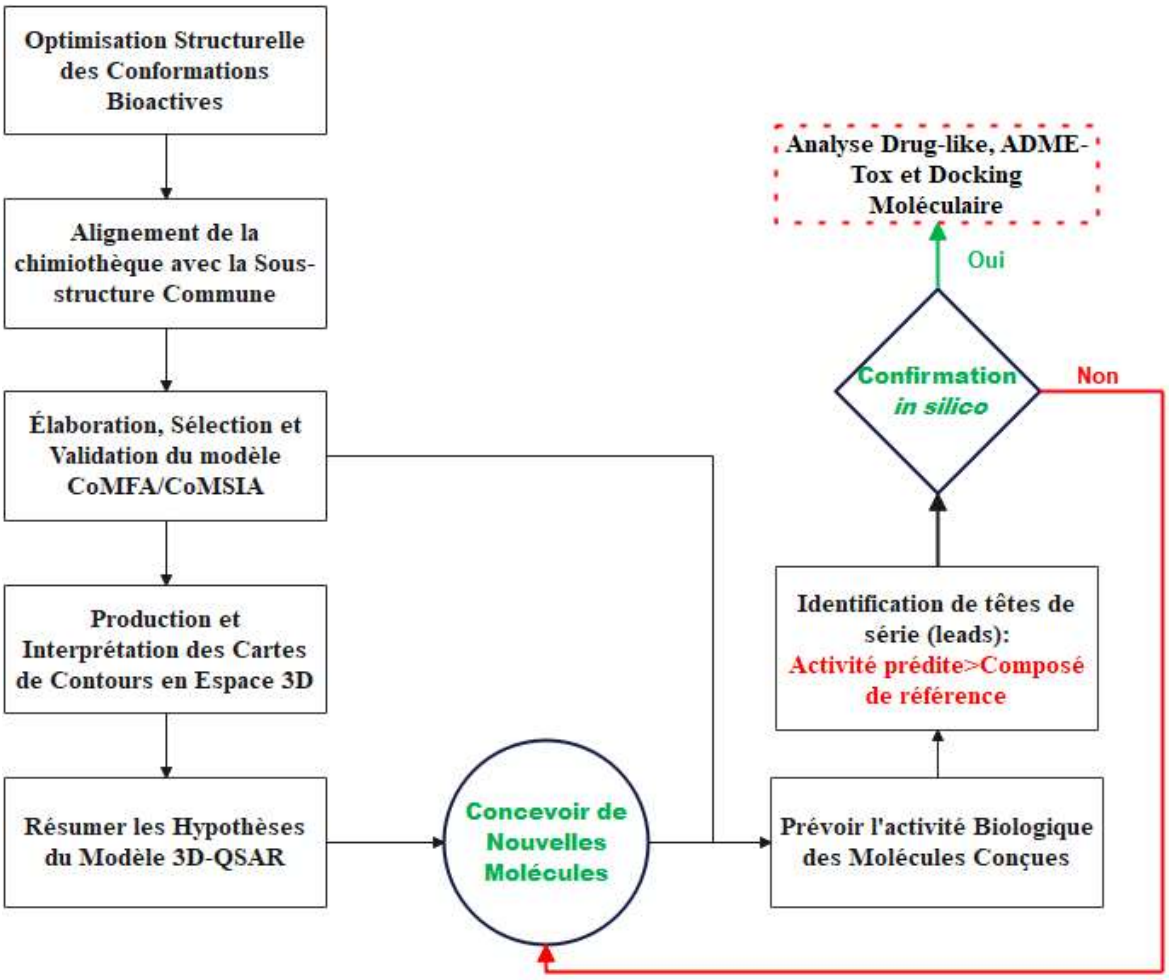


Figure II.16. La méthodologie globale de CoMSIA /CoMSIA appliquée dans nos travaux.

En vue de la validation des modèles 3D-QSAR, le nombre idéal de composants (N) et le coefficient de corrélation (Q^2l_{oocv}) sont calculés au cours des analyses PLS de CoMFA et CoMSIA, grâce à la validation croisée de type "leave-one-out". Après avoir accepté le nombre de composantes idéal (N) et le coefficient de corrélation de validation croisée (Q^2), les paramètres statistiques sont calculés, notamment le coefficient de détermination (R^2), le coefficient de détermination par Bootstrapping (R^2_{bs}), l'erreur standard estimée (SEE), le test de Fischer (F-value), les contributions des descripteurs de CoMFA (stérique et électrostatique) et de CoMSIA (Stérique, Électrostatique, Hydrophobique, Donneur de liaison hydrogène et Accepteur de liaison hydrogène) exprimées en fraction. Le pouvoir prédictif externe des modèles 3D-QSAR développés est ensuite testé en appliquant les modèles proposés à l'ensemble de test et en calculant le coefficient de détermination (R^2_{test}t). En plus de ces évaluations,

d'autres tests statistiques peuvent être appliqués, tels que le test de Y-randomisation et/ou le domaine d'applicabilité, ceci afin d'étayer la fiabilité et la performance des modèles 3D-QSAR proposés.

4.10.5.8 Visualisation et interprétation des résultats CoMFA/CoMSIA

Les résultats des analyses PLS peuvent être présentés sous forme de cartes de contours résultant du produit standard des coefficients PLS et de la déviation standard (coefficient SD*). Cela rend l'interprétation des modèles CoMFA et CoMSIA visuellement facile et rapide lors de l'analyse. Cela aide considérablement à identifier les zones critiques dans l'espace 3D (80% de caractéristiques favorables et 20% de caractéristiques défavorables) autour des composés où les modifications structurelles spécifiques diffèrent de manière significative, ainsi que leur impact sur l'activité biologique. Outre les cartes de contours, les analyses PLS des modèles CoMFA et CoMSIA présentent la distribution des valeurs d'activité biologique observées expérimentalement (vraies) en termes de celles prédites par les modèles 3D-QSAR développés. Ensuite, la validité et la fiabilité des modèles QSAR proposés doivent être vérifiées, avant de passer aux phases de post-génération des hypothèses pharmacophores liées à la rationalisation des propriétés physico-chimiques favorables pour améliorer l'activité biologique des molécules examinées, à la conception de nouvelles molécules et prédire leurs activités biologiques, à la sélection de nouvelles molécules de tête (leader) et à la prédiction *in silico* de leurs propriétés de la biodisponibilité (Drug-like), pharmacocinétique et pharmacodynamie (ADME-Tox).

5 Critères de la validation des modèles 2D/3D-QSAR

Dans cette section, nous abordons l'importance de la validation des modèles QSAR afin de confirmer leur fidélité et leur précision dans la prédiction de l'activité ou de la propriété souhaitée. Il existe plusieurs procédures de validation, notamment la validation interne, externe et croisée [311], [312]. Il est également important d'évaluer le domaine d'applicabilité des modèles QSAR, ce qui peut être fait à l'aide de méthodes telles que l'analyse de l'effet de levier, l'analyse de la distance et l'analyse de la similarité [313]. D'autres critères de validation, tels que le test de Y-randomisation, peuvent être appliqués pour conforter la stabilité du modèle QSAR proposé [314].

5.1 Validation interne

Plusieurs paramètres statistiques peuvent être utilisés dans le cadre de la validation interne des modèles QSAR, notamment le coefficient de corrélation (R), l'erreur quadratique moyenne (MSE), l'erreur standard de l'estimation (SEE) et le coefficient de détermination (R^2). Les modèles doivent avoir un coefficient de corrélation élevé (R>0.9), une MSE et une SEE faibles et un coefficient de détermination élevé pour être considérés comme des modèles fiables. En outre, dans la modélisation QSAR, le p-value et F-test sont des mesures statistiques utilisées pour évaluer la validation d'un modèle. Un faible p-value (par exemple, moins de 0,05) indique que l'hypothèse nulle peut être rejetée. En revanche, le F-test permet de comparer la variance de deux ensembles de données afin de déterminer s'ils sont significativement différents, si la valeur du F est élevée, cela signifie que le modèle est significativement meilleur que le modèle nul. Pour valider les modèles QSAR, il est important

d'utiliser conjointement les deux mesures (p-value et F-test) et de les combiner à d'autres techniques de validation pour assurer la robustesse du modèle.

5.2 Validation croisée

La méthode LOOCV (Leave-One-Out Cross-Validation) est couramment utilisée pour valider les modèles QSAR [315]. Cette méthode consiste à construire un modèle à partir des données restantes après avoir retiré un élément de l'ensemble de données, puis à utiliser ce modèle pour prédire l'activité de l'élément exclu. Ce processus est répété pour chaque élément de données afin d'évaluer la performance du modèle. Le processus à suivre pour valider le modèle consiste à diviser les données en ensembles d'apprentissage et de validation, à construire le modèle QSAR, à prédire l'activité de l'échantillon exclu, à calculer l'erreur de prédiction, à répéter le processus pour tous les échantillons et à évaluer la performance du modèle à l'aide d'une valeur seuil prédéterminée. La qualité de la prédiction du modèle QSAR est mesurée à l'aide du coefficient de validation croisée (Q^2) qui doit être égal ou supérieur à 0.5 pour que le modèle soit performant. Enfin, le modèle QSAR validé est utilisé pour prédire l'activité biologique des molécules de l'ensemble de test afin d'évaluer son pouvoir prédictif par le biais d'une validation externe.

5.3 Validation externe

La validation externe est une méthode fondamentale pour évaluer la performance d'un modèle QSAR. Cette méthode permet d'évaluer la performance d'un modèle QSAR (relation quantitative structure-activité). A cet effet, le coefficient de détermination (R^2) est généralement considéré comme un indicateur clé de la qualité d'un modèle QSAR, mais il doit être calculé à la fois pour les données d'apprentissage (R^2_{train}) et pour les données de validation (R^2_{test}). Le R^2_{test} est un paramètre important de la validation externe qui mesure la précision de la prédiction du modèle sur les données de l'ensemble de test qui n'ont pas été utilisées à l'entraînement du modèle. Ce paramètre indique la capacité du modèle à se généraliser à de nouvelles données et à prédire avec précision la réponse de ces données. Si la valeur de R^2_{test} est supérieure à 0.6, cela signifie que le modèle est capable de s'adapter à de nouvelles données et qu'il a un fort pouvoir prédictif.

5.4 Test de Y-randomisation

Afin de garantir que les modèles QSAR développés sont suffisamment robustes, le test de randomisation Y est appliqué pour exclure la possibilité que les modèles soient robustes en raison du hasard. Dans ce test, les valeurs d'activité biologique des composés sont distribuées de manière aléatoire aux molécules, tandis que les valeurs des descripteurs moléculaires restent stables. Les données aléatoires générées sont utilisées pour développer de nouveaux modèles QSAR randomisés. En vue de préserver la stabilité et la validité du modèle QSAR original, le coefficient de détermination (R^2_r) et le coefficient quadratique de validation croisée (Q^2_r) des modèles aléatoires doivent être inférieurs au coefficient de détermination (R^2) et au coefficient quadratique de validation croisée (Q^2) du modèle QSAR original (non randomisé). La différence entre les valeurs du coefficient de détermination du modèle randomisé (R_r^2) et le coefficient de détermination du modèle original (R^2) est

exprimée comme la valeur du paramètre $^{c}R^{2}_{p}$. L'évaluation de la valeur de $^{c}R^{2}_{p}$ permet de réduire la différence entre les valeurs des coefficients de détermination du modèle original (R^2) et du modèle randomisé (R^2_r). Pour que la qualité du modèle testé soit confirmée, le seuil de $^{c}R^{2}_{p}$ doit dépasser 0.5.

5.5 Domaine d'application (DA)

Le domaine d'applicabilité (DA) est un concept associé à la fiabilité des prédictions des modèles QSAR/QSPR et à la possibilité de refuser toute mauvaise interprétation de leurs résultats. Le DA est défini comme la gamme dans laquelle les structures chimiques et les réponses sont prédites par le modèle avec un niveau de confiance donné [316]. La fiabilité des prédictions du modèle est déterminée par sa validation, qui permet de vérifier sa capacité à prédire des données précises pour de nouveaux composés. L'un des principaux défis de la modélisation QSAR est l'estimation de la confiance accordée aux prédictions, qui est liée à la définition du domaine d'applicabilité du modèle [317]. Le diagramme de Williams (Williams plot) est une méthode très utile pour évaluer la fiabilité d'un modèle QSAR [318]. Cette méthode permet de présenter les résidus standardisés de prédiction (r_i) et les valeurs d'effet de levier (h_i) pour chaque échantillon (i), ce qui permet d'identifier les réponses aberrantes et les molécules influentes du point de vue structurel. Les valeurs aberrantes sont définies comme des molécules dont les résidus standard (r_i) dépassent le niveau de (-2 et +2) à (-3 et +3). D'autre part, les molécules structurellement affectées ont des valeurs de levier (h_i) supérieures au seuil de levier (h^*), qui est généralement 3 fois la valeur moyenne de levier. En effet, quand ($hi < h^*$), la probabilité de concordance entre les valeurs prédites et observées est élevée. Cependant, les molécules qui présentent ($h_i > h^*$) sont considérées comme structurellement éloignées des autres molécules de l'ensemble d'entraînement ; par conséquent, elles sont considérées comme en dehors de la DA du modèle.

5.6 Critères statistiques pour l'évaluation de la performance des modèles QSAR

Dans cette thèse, la validation des modèles 2D et 3D-QSAR multidimensionnelles est effectuée en accord avec les principes de l'OCDE ainsi qu'avec les critères de Golbraikh et Tropsha pour la validation des modèles QSAR [204], [315]. Pour renforcer la fiabilité de ces modèles développés dans le cadre de nos travaux, nous avons mis en place des tests complémentaires, tels que la validation croisée de type "LOOCV", le test de Y-randomisation et la détermination du domaine d'applicabilité (DA) des modèles QSAR que nous avons élaboré [314], [319], [320]. Le **Tableau II.10** rassemble les outils et les paramètres statistiques utilisés pour évaluer la signification statistique et le pouvoir prédictif des modèles QSAR développés dans le cadre de nos travaux.

Tableau II.10. Critères statistiques utilisés dans cette thèse pour la validation du modèle 2D/3D-QSAR.

Validation	Outil de calcul	Critères d'évaluation	Formule de calcul utilisée	Seuil d'acceptation
Interne	Microsoft Excel [321, p. 365],et XLSTAT pour les modèles	Coefficient de détermination interne (R^2)	$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_{obs} - Y_{calc})^2}{\sum_{i=1}^{n}(Y_{obs} - \overline{Y}_{calc})^2}$	> 0.6

Validation	Outil de calcul	Critères d'évaluation	Formule de calcul utilisée	Seuil d'acceptation
	2D-QSAR. SYBYL-X2.1.1 pour les modèles 3D-QSAR	Coefficient de détermination ajusté (R^2_{Adj})	$R^2_{adj} = \frac{(n-1)\times R^2 - p}{n-(p+1)}$	
		Erreur quadratique moyenne (MSE)	$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_{Obs} - Y_{Calc})^2$	Les valeurs plus faibles sont optimales
		Erreur standard de l'estimation (SEE)	$SEE = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-(p+1)}}$	
		Test de Fisher (Ftest)	$F_{test} = \frac{\sum(Y_{Calc} - \overline{Y_{Calc}})^2}{\sum(Y_{Obs} - Y_{Calc})^2} \times \frac{n-(p+1)}{p}$	Plus la valeur est >0,3, plus la qualité est élevée
		Hypothèse nulle (p-value)	Méthodes de convergence et/ou de Monte Carlo disponibles dans le logiciel XLSTAT.	La valeur p doit être très faible, idéalement p-value≤ 0,05
Croisée	Microsoft Excel	Coefficient quadratique de validation croisée (Q^2ou R^2_{cv})	$Q^2_{cv} = 1 - \frac{\sum(Y_{Calc} - Y_{Obs})^2}{\sum(Y_{Obs} - \overline{Y_{Obs}})^2}$	> 0.5
Externe	Microsoft Excel	Coefficient de détermination externe (R^2_{test})	$R^2_{test} = 1 - \frac{\sum(Y_{Calc}(test) - Y_{Obs}(test))^2}{\sum(Y_{Obs}(test) - \overline{Y_{Obs}}(train))^2}$	> 0.6
Y-randomisation	Microsoft Excel pour les modèles 2D-QSAR, et SYBYL-X2.1.1.1 pour les modèles 3D-QSAR	Coefficient de détermination (R^2_{Rand})	R^2_{Rand} moyen des itérations aléatoires (i fixé à 50 pour le 2D-QSAR et à 5 pour le 3D-QSAR).	< $R^2_{original}$
	Microsoft Excel, et/ou MLR Y-Randomization Test 1.2 [322].	Coefficient quadratique de validation croisée (Q^2_{Rand})	Q^2_{Rand} est la moyenne des valeurs Q^2_{loocv} calculées après plusieurs itérations randomisées (i fixé à 50 pour le 2D-QSAR et à 5 pour le 3D-QSAR).	Q^2_{Rand} < $Q^2_{Orignal}$
		Coefficient de corrélation aléatoire quadratique ($^cR^2_p$).	$^cR^2_p = R_{original} \times \sqrt{R^2_{original} - R^2_{random}}$	> 0.5
Domaine d'applicabilité (DA)	Applicability Domain toolbox for MATLAB	Valeurs de l'effet de levier (h_i)	$h_i = X_i^T (X^T X)^{-1} X_i$	$h^*<h$
		Seuil de l'effet de levier (h^*)	$h^* = 3 \times \frac{(k+1)}{n}$	

Validation	Outil de calcul	Critères d'évaluation	Formule de calcul utilisée	Seuil d'acceptation
	version 1.0 [323]	Résidu standardisé (r_i)	$r_i = \frac{Y_{calc} - Y_{Obs}}{\sqrt{\frac{\sum_{i=1}^{n} (Y_{Calc} - Y_{Obs})^2}{n}}}$	$-3 < r_i < +3$

Yobs représente les valeurs de pIC_{50} observées expérimentalement ; $\mathbf{Y_{calc}}$ représente les valeurs de pIC_{50} calculées par le modèle QSAR ; $\overline{Ycalc}$ représente la moyenne des valeurs de pIC_{50} (observées/prédites) ; **n-(p +1)** degrés de liberté ; **n** représente le nombre d'éléments dans l'ensemble d'entraînement, **p**/ou **k** signifie le nombre de descripteurs moléculaires sélectionnés par le modèle QSAR ; **LOOCV** signifie (Leave-One-Out Cross-Validation), il s'agit d'une technique de validation croisée utilisée dans le cadre de l'apprentissage automatique pour évaluer les performances d'un modèle ; **Rand** indique Randomisation, ***h*** est l'effet de levier du composé, ***h**** est le levier d'alerte du même composé. **X** est la matrice du descripteur, $\mathbf{X^T}$ est la transposée de **X**.

5.7 Exploitation de modèles 2D/3D-QSAR pour le criblage de chimiothèques et l'optimisation de composés têtes de série

Après avoir complété les phases de développement et de validation des modèles 2D et/ou 3D-QSAR, la phase importante de l'application de ces modèles mathématiques commence. Dans cette thèse, les principales applications de ces modèles consistent principalement à décrypter la relation structure-activité des molécules étudiées, à faciliter le criblage de chimiothèques de molécules biologiquement actives, à rationaliser la conception de nouvelles molécules, à optimiser leur activité biologique et à la prédire par simulation numérique avant la synthèse et la mise en œuvre des essais *in vitro* et *in vivo.* Dans cette thèse, la contribution à la rationalisation la modélisation de nouvelles petites molécules candidates à la conception de nouveaux médicaments est l'élément central qui unit les diverses applications de la modélisation QSAR que nous avons accomplie.

5.8 Méthodes employées pour la conception et l'optimisation informatique de nouvelles molécules bioactives

Afin de concevoir de nouvelles molécules biologiquement plus actives et compatibles avec la structure du biomarqueur ciblé comparativement à la conformation bioactive du ligand prototype (la structure moléculaire présentant l'activité biologique optimale) et au médicament standard choisi comme contrôle, nous avons procédé à des modifications structurelles et à des substitutions favorables pour la modélisation biomoléculaire basée sur les orientations fournies par les modèles 2D et/ou 3D-QSAR développés. Conformément à cette démarche, nous comptons principalement sur l'identification de sites pharmacophores favorables pour parvenir à la modélisation biomoléculaire visée, ainsi que sur la sélection de groupes fonctionnels appropriés pour la procédure de conception de nouvelles molécules.

5.8.1 Identification des sites pharmacophores

Dans le cadre de nos travaux de modélisation 2D/3D-QSAR hybride, nous avons identifié les sites pharmacophores localisés au niveau du ligand de référence via les orientations des cartes de contours générées par les analyses CoMFA et CoMSIA puis par des simulations de Docking Moléculaire [217], [304], [309], [324], tout cela d'une part. D'autre part, à l'aide de simulations de potentiel électrostatique

moléculaire (PEM) et du Docking Moléculaire, nous avons identifié les sites pharmacophores susceptibles pour la modélisation biomoléculaire basée sur les modèles 2D-QSAR [100]. Dans ce contexte, nous identifions les sites pharmacophores sur la structure de référence en respectant le critère de compatibilité de la structure du ligand (médicament candidat) avec la structure de la protéine cible. Sur cette base, l'identification des sites de réactivité favorables aux attaques électrophiles et nucléophiles ainsi qu'aux interactions non-covalentes protéine-médicament favorisées par le ligand de référence sera extrêmement utile pour rationaliser la conception de nouvelles molécules compatibles avec une cible biologique souhaitée dans les phases préliminaires de la conception et de la découverte de médicaments (**Figure II.17**).

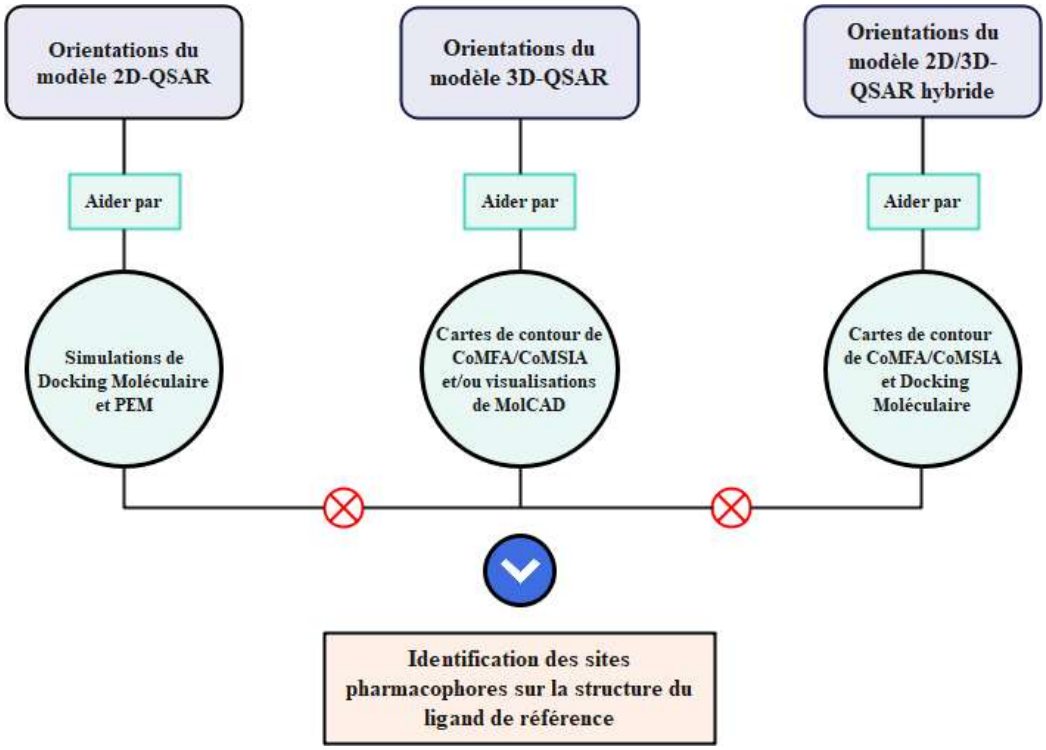


Figure II.17. Schéma de la méthodologie implémentée dans nos travaux en vue d'identifier des sites pharmacophores favorables pour rationaliser la conception de nouvelles molécules.

A ce stade, sur la base de l'analyse visuelle des cartes de contours générées par les calculs (PEM) réalisés grâce à l'approche DFT [100], nous pouvons identifier les sites d'interaction favorables et défavorables pour les interactions électrostatiques potentielles avec la cible protéique. D'autre part, la simulation de Docking Moléculaire permet de générer des visualisations 3D des interactions entre le ligand de référence et le récepteur cible. Grâce à la visualisation de la surface moléculaire du système protéine-ligand, il est possible d'identifier les sites favorables pour explorer les interactions non covalentes au sein des récepteurs protéiques (interactions hydrophobiques, électrostatiques, liaisons hydrogène conventionnelles, interactions ioniques, ...etc. Dans le cadre de cette thèse, après avoir procédé à la simulation du Docking Moléculaire du ligand de référence à l'intérieur de la poche active du récepteur cible, nous avons généré des visualisations 3D par Discovery Studio 2016 et/ou des visualisations MolCAD (en anglais Molecular Computer-Aided Design) avec SYBYL-X2.1.1 [324]. Dans cette thèse, nous avons appliqué la présente stratégie à l'identification de sites pharmacophores cibles à la modélisation, afin d'éviter l'improvisation et le choix aléatoire des sites lors de la proposition de nouvelles molécules.

5.8.2 Sélection des groupes fonctionnels optimaux pour la conception de nouveaux ligands

Concernant la sélection des groupements fonctionnels que nous avons substitués par d'autres ou additionnés aux sites pharmacophores identifiés sur la molécule de référence, cette procédure est réalisée dans la plupart de nos travaux, soit par le biais de l'intuition chimique humaine, ou par l'intervention des outils d'intelligence artificielle. Dans le cas d'une sélection manuelle, nous nous basons principalement sur les relations SAR décodées par des modèles QSAR et sur nos intuitions chimiques relatives à la possibilité et à la facilité de rétro-synthèse des nouvelles molécules proposées. Dans le cas de l'approche automatisée, nous avons appliqué l'approche de décodage des réseaux neuronaux basée sur la forme (shape-based neural network decoding) disponible sur le serveur Internet LigDream toolkit (https://playmolecule.org/LigDream/) [325]. Le modèle de la sélection automatique généré à l'aide de l'outil LigDream permet de conçu un maximum de 100 nouvelles molécules comportant des groupes fonctionnels qui présentent des propriétés similaires à celles des médicaments et qui couvrent de nouveaux sites dans la structure de la molécule de référence. Après avoir généré les nouvelles molécules automatiquement, nous sélectionnons parmi elles celles qui sont adaptées aux recommandations et aux prédictions des modèles QSAR développés [309].

5.8.3 Sélection et identification de molécules candidates en tant qu'inhibiteurs spécifiques par criblage

Après avoir achevé la méthodologie de conception rationnelle de nouvelles molécules suivie dans nos travaux, nous dérivons de nouvelles molécules à partir de la structure de la molécule de référence et optimisons leurs structures géométriques et énergétiques selon la même méthodologie mise en œuvre sur les données originales. Ensuite, nous appliquons les modèles QSAR entraînés automatiquement pour prédire leurs activités biologiques (pIC_{50}) vis-à-vis de la cible biologique visée dans le sens de l'inhibition de son activité enzymatique. Sur la base des prédictions des modèles QSAR, nous comparons les données des activités biologiques prédites des nouvelles molécules avec celles observées *in vitro* (dans la chimiothèque choisie). Ensuite, nous classons les nouvelles données (structures/activités) de la plus forte activité prévue à la plus faible. Enfin, nous sélectionnons les nouvelles molécules dont l'activité biologique prédite est supérieure à celle de référence ($préd_pIC_{50}$>Obs_pIC_{50}) et nous excluons les molécules inactives ou moins actives.

6 Approche employée pour la sélection des ligands candidats dans le processus de conception et de découverte de médicaments par modélisation informatique

L'étape d'identification et de sélection de petites molécules ("Hits" en anglais) par la voie computationnelle ou expérimentale est très importante dans le tube de conception du médicament. En effet, de nouveaux "Hits" sont susceptibles de conduire à la découverte de nouveaux inhibiteurs spécifiques d'une cible biologique particulière, telle qu'une protéine ou une enzyme impliquée dans une pathologie spécifique. Par conséquent, la possibilité d'obtenir un nouveau médicament est renforcée. Pour garantir la fiabilité de la sélection, les nouvelles molécules "Hits" doivent être soumises à

davantage de protocoles de criblage, dans le but de les filtrer selon les critères liés à leurs propriétés pharmaceutiques (PK), pharmacodynamiques (PD), sélectives, affinité et compatibilité avec le récepteur ciblé. Cette procédure très importante pour la découverte précoce de médicaments, basée sur des modèles réels de criblage à haut débit (HTS), se réfère à l'expression H2L (Hit to Lead). Grâce au volet computationnel (*in silico*) basé sur l'approche CADD et à ses avantages en termes d'accélération du processus décisionnel, de réduction des coûts et d'augmentation des chances de succès [326], nous avons mis en œuvre dans cette thèse une stratégie computationnelle virtuelle *in silico* visant à simuler le modèle réel de criblage à haut débit (HTS). Ceci afin de cribler et d'identifier rapidement les molécules (H2L) candidates, susceptibles de favoriser la conception de nouveaux médicaments sûrs et efficaces ("Lead Optimisation"). Le diagramme illustré dans la **Figure II.18** présente le modèle virtuel général utilisé pour le criblage du H2L, qui repose sur les méthodes de CADD hybride appliquées dans cette thèse.



Figure II.18. Modèles de criblage virtuel que nous avons mis en œuvre pour identifier des agents candidats au développement d'un médicament.

Dans les sections suivantes, nous approchons des aspects consacrés aux stratégies que nous avons appliquées pour satisfaire les critères visés.

6.1 Modélisation pharmacocinétique et pharmacodynamique (PK/PD)

6.1.1 Contexte et principe

Au cours de la dernière décennie, les outils expérimentaux *in vitro* et les modèles *in silico* ont été de plus en plus utilisés pour évaluer les profils ADME-Tox des composés lors de la découverte de médicaments. Ces approches visent à renforcer les taux de réussite et à sélectionner de meilleurs candidats pour le développement [327]. Dans ce sens, la combinaison de méthodes de criblage

expérimentales et virtuelles est essentielle pour assurer le suivi efficace du profil ADME-Tox dans le processus de découverte [328]. Dans ce contexte, la combinaison de méthodes de criblage expérimentales et computationnelles est indispensable pour assurer un suivi adéquat du profil ADME-Tox dans le processus précoce de conception et d'évaluation des médicaments. De ce fait, l'application de méthodes de modélisation *in silico* pharmacocinétique et pharmacodynamique (PK/PD) en vue de prédire les profils ADME-Tox est un aspect critique de la mise au point d'un nouveau médicament [329]. Grâce à la modélisation PK/PD *in silico*, les essais expérimentaux *in vitro* et *in vivo* peuvent être orientés, ce qui permet de gagner du temps et de réduire considérablement le nombre des tâches et des ressources. En général, la modélisation PK permet de décrire la réponse de l'organisme vis-à-vis d'un médicament en termes de capacité d'absorption (A), de distribution (D), de métabolisme (M) et d'excrétion (E) [330]. Par ailleurs, la modélisation PD permet de décrire la dynamique de l'action d'un médicament sur l'organisme en reliant la concentration de ce dernier à une mesure d'activité/de sûreté ou du risque de toxicité (Tox) [330].

6.1.2 Rôle des modèles Drug-like et ADME-Tox in silico

Dans les phases précoces de la découverte de médicaments, les modèles *in silico* ADME-Tox sont utilisés pour détecter les composés présentant des propriétés défavorables et les éliminer. Cependant, il existe des risques de faux positifs et de faux négatifs. Les faux positifs peuvent être traités par des essais ultérieurs, tandis que les faux négatifs sont préoccupants car ils sont susceptibles de négliger certains composés susceptibles de constituer une nouvelle piste thérapeutique. En raison du grand nombre de composés examinés, la probabilité de sélectionner des composés présentant des caractéristiques favorables est plus élevée. On utilise également ADME-Tox pour mettre en évidence les faiblesses des composés sélectionnés en vue d'une stratégie d'optimisation adaptée. A cet égard, le succès de cette approche dépend de la diversité des composés disponibles, ce qui réduit les risques de faux-négatifs. En général, les médicaments candidats de haute qualité ont un pouvoir thérapeutique suffisant pour atteindre la cible appropriée lorsqu'ils présentent les propriétés ADME-Tox souhaitées [331]. D'autre part, les tests expérimentaux dans une véritable stratégie de criblage peuvent entraîner une marge d'erreur importante entre les évaluations *in vitro* et/ou *in vivo*, ce qui réduit le taux de réussite des molécules en tant qu'agents médicinaux. Les principaux problèmes comprennent l'incompatibilité des propriétés structurelles de la molécule avec la biodisponibilité du médicament, la pharmacocinétique inadéquate, les effets secondaires indésirables, la concentration de dose létale, le risque de toxicité et l'incapacité à atteindre une cible thérapeutique et une réponse biologique. Pour éviter ces problèmes potentiels et faciliter le processus de découverte de médicaments, il est essentiel d'accomplir un criblage virtuel *in silico* pour prédire les profils ADME-Tox des molécules candidates avant de procéder à des investigations plus approfondies. Dans cette thèse, l'accent est mis sur l'évaluation de la compatibilité des propriétés médicamenteuses des molécules candidates avec la biodisponibilité orale et les paramètres pharmacocinétiques et pharmacodynamiques.

6.1.3 Mise en œuvre des modèles de type Drug-like et ADME-Tox in silico

De nombreux outils et logiciels sont disponibles pour cribler et filtrer les molécules en fonction de leur profil ADME-Tox prédit. Dans le cadre de nos travaux, nous élaborons des modèles PK/PD à l'aide des outils en ligne SwissADME et pkCSM, les plus populaires. SwissADME fournit des prédictions et des informations précieuses sur les propriétés de similitude avec les médicaments et les propriétés ADME des petites molécules, qui sont des aspects critiques dans la découverte et le développement de médicaments [332]. En revanche, la plateforme d'apprentissage automatique pkCSM permet de prédire les propriétés PK/PD des petites molécules, sur la base de modèles de distance/pharmacocinétique codés sous forme de signatures graphiques [333]. Au préalable, nous convertissons la structure des petites molécules de leur présentation graphique en une présentation de type SMILES, compatible avec les plates-formes SwissADME et pkCSM. Après avoir défini les formules SMILES pour les plates-formes, diverses propriétés sont calculées à partir desquelles des modèles de type "Drug-like" et ADME-Tox peuvent être développés en vue de cribler et de filtrer les molécules. Les calculs de SwissADME des propriétés physicochimiques, de la lipophilie, de la solubilité, de la pharmacocinétique et de certains paramètres de chimie médicinale (accessibilité synthétique) permettent de construire des profils "Drug-likeness ou Drug-like". Dans nos travaux, nous utilisons le profil Drug-like défini par Lipinski [334] et soutenu par les règles de Ghose [335], Veber [336], Egan et Muegge [337] pour évaluer le degré de biodisponibilité, d'absorption et de perméation du médicament candidat. La **Figure II.19** résume les règles les plus importantes que nous avons adoptées pour exclure les molécules qui ne sont pas conformes aux critères de bon candidat-médicament.

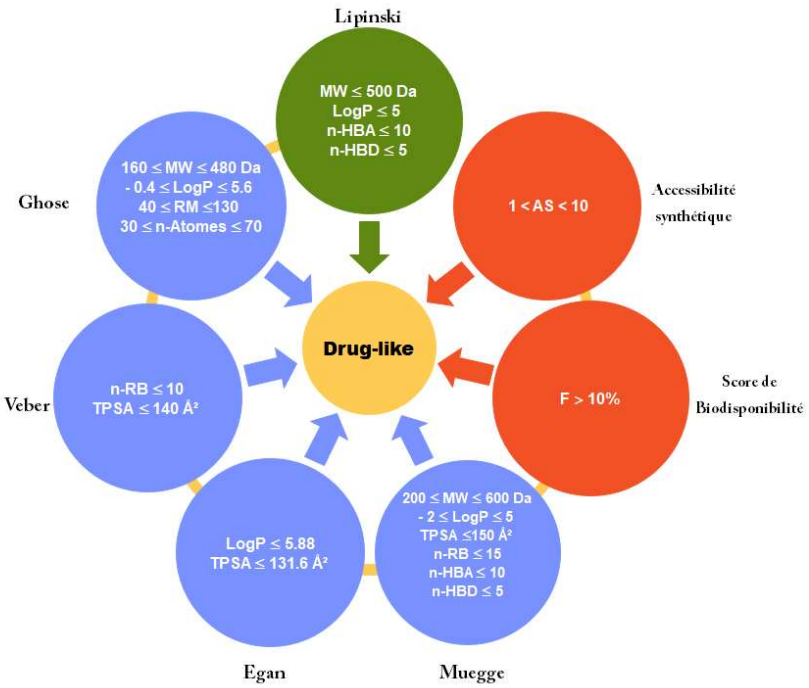


Figure II.19. Critères du type " Drug-like " à respecter pour sélectionner les candidats H2L dans le cadre de ce travail.

Le modèle "Drug-like" est mis en œuvre pour sélectionner les molécules sur la base de leurs propriétés de biodisponibilité et de leurs flexibilités en termes d'interaction avec la cible. Avec MW : Poids moléculaire ; RM : Réfractivité molaire, LogP : Logarithme du coefficient de partage entre l'octanol

et l'eau ; n-HBA : Nombre d'accepteurs de liaisons hydrogène ; n-HBD : Nombre de donneurs de liaisons hydrogène ; n-RB : Nombre de liaisons rotatives ; TPSA : Surface polaire topologique ; n-Atomes : Nombre d'atomes dans la molécule ; F : Indice de biodisponibilité ; AS : Accessibilité synthétique. Dans cette phase du protocole de criblage virtuel, nous utilisons les sorties du modèle "drug-like" développé pour sélectionner et identifier les molécules qui répondent aux conditions de biodisponibilité. Nous excluons toutes les molécules qui violent au moins deux règles de la règle des cinq de Lipinski (RO5) et al [338]. De plus, nous intégrons le score de l'indice d'accessibilité synthétique (AS) afin de prédire le degré de facilité (AS proche de 1) ou de complexité (AS proche de 10) de l'accessibilité de la molécule étudiée par voie synthétique. Après avoir identifié les molécules qui ont passé avec succès la phase de filtrage basée sur le modèle "Drug-like", nous les avons soumises à une évaluation ADME-Tox très stricte. Cette démarche vise à renforcer les études et les évaluations prédictives qui alimentent le criblage des structures moléculaires candidates à la conception et à la découverte de nouveaux médicaments. Dans ce contexte, nous adoptons les prédictions ADME-Tox générées par le modèle pKCSM comme critères d'évaluation du profil pharmacocinétique et pharmacodynamique des molécules objet du criblage virtuel. Les résultats des sorties du modèle pkCSM sont interprétés à la lumière de la théorie proposée par le modèle concernant les paramètres ADME-Tox prédits, accessibles via le lien https://biosig.lab.uq.edu.au/pkcsm/theory. La **Figure II.20** récapitule les principaux paramètres utilisés dans la construction du modèle ADME-Tox, que nous avons adopté dans nos travaux lors du criblage des molécules médicamenteuses candidates.



Figure II.20. Critères ADME-Tox impliqués dans le criblage des candidats médicaments dans nos travaux.

Pour renforcer le processus de filtrage et réduire la marge d'erreur que le modèle pkCSM peut produire, nous avons suggéré dans certains de nos travaux d'ajouter d'autres modèles de filtrage. Ces modèles de calcul inclus le modèle "ADMET_AlogP98 vs. ADMET_PSA_2D"[217] et les calculs d'OSIRIS [339]. À l'aide du package de Discovery Studio 2016, nous développons le modèle ADMET_AlogP98 vs. ADMET_PSA_2D sur la base des descripteurs logP98 et PSA. ADMET_AlogP98

permet d'estimer le coefficient de partage (logP), qui indique l'hydrophobie d'une molécule et aide à évaluer son absorption et sa distribution. Le modèle ADMET_PSA_2D prédit la surface polaire bidimensionnelle (PSA), qui reflète les propriétés de solubilité et de perméabilité d'une molécule. Ce modèle nous permet de générer une visualisation par éclipse des intervalles de confiance à 99 % et à 95 % des régions dans lesquelles les molécules présentent une bonne absorption dans l'intestin et passent dans la barrière hémato-encéphalique (BHE) [340]. Alors que les calculs d'OSIRIS permet de produire des profils sur les risques de toxicité moléculaire susceptibles de résulter de radicaux indésirables dans la structure des molécules codées par SMILES, tels que les risques associés à la mutagénicité, à la cancérogénicité, à la reproductibilité et à l'irritabilité [341]. Les résultats de la prédiction d'OSIRIS sont codés en couleur pour évaluation. Le rouge indique un risque élevé d'effets indésirables, tandis que le vert indique un comportement conforme au médicament (**Figure II.21**).

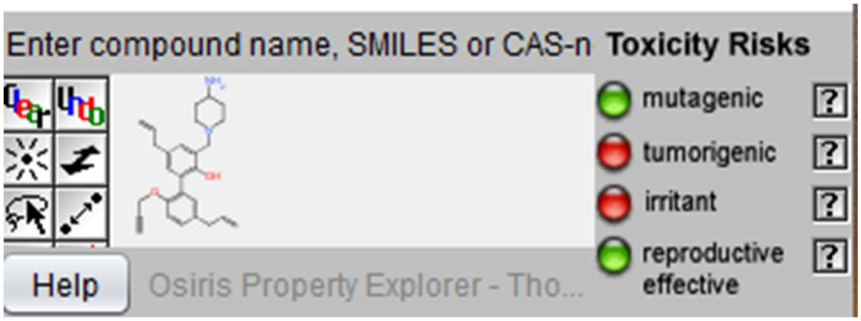


Figure II.21. Interface de l'outil "OSIRIS Property Explorer" destiné à la prédiction des risques de toxicité des petites molécules.

6.2 Stratégie de criblage virtuel basé sur le Docking moléculaire

Après avoir prédit les profils PK/PD *in silico* et identifié les ligands qui présentent des propriétés AMDE-Tox adaptées à la conception de médicaments, vient l'étape de l'évaluation de la compatibilité de ces molécules avec la cible thérapeutique (protéine, enzyme...). Pour ce faire, on prédit les modèles d'interaction protéine-ligand, on estime les énergies de liaison/affinité, la constante d'inhibition (Ki) et d'autres paramètres qui contribuent à la bonne compréhension des agents candidats en tant qu'inhibiteurs de la cible biologique impliquée dans la maladie ciblée. Les interactions biomoléculaires peuvent être classées comme intramoléculaires (au sein des biomolécules) ou intermoléculaires (entre les biomolécules) [95]. Ces interactions jouent un rôle crucial dans la stabilité, la fonction et la signalisation des biomolécules. Différents types d'interactions, tels que les liaisons covalentes, les liaisons hydrogène, les interactions de van der Waals, les ponts salins et les interactions électrostatiques et hydrophobiques, contribuent à ces processus [342]. La compréhension et la modélisation de ces interactions sont importantes dans divers domaines tels que la conception de médicaments, les nanotechnologies, les biocapteurs et la conception de biomarqueurs. Toutefois, l'étude de ces interactions au niveau atomique est complexe et nécessite des techniques avancées. Les méthodes expérimentales traditionnelles, telles que la spectrométrie de masse et la résonance magnétique nucléaire (RMN), présentent plusieurs désavantages, notamment en termes de coût, de temps et d'incohérence des résultats en raison des conditions de laboratoire variables [343], [344]. Pour surmonter ces limitations, des méthodes computationnelles ont été adoptées, en particulier l'approche de Docking moléculaire [345]. Cette technique de simulation permet d'analyser des données

d'interaction à grande échelle et fournit des informations précises sur les interactions protéine-médicament. Elle aide à évaluer les paramètres de compatibilité et de stabilité entre les protéines et les médicaments, facilitant ainsi la compréhension des interactions biomoléculaires. La polyvalence du Docking moléculaire le rend un outil de premier plan dans l'approche de SBDD [183].

6.2.1 Principe de Docking Moléculaire

Avant de décrire la méthodologie du Docking moléculaire que nous avons appliquée dans nos travaux, nous commençons au préalable par quelques informations sur la manière dont les grands systèmes biomoléculaires (tels que les protéines) sont représentés par le biais du Docking moléculaire. Nous abordons ensuite la classification des algorithmes utilisés pour réaliser le Docking moléculaire conformément aux outils bio-informatiques adaptés à cet usage. Ensuite, nous indiquons comment interpréter les résultats de Docking. Enfin, nous présentons brièvement la méthodologie du Docking moléculaire mise en œuvre dans ce travail et les outils bio-informatiques impliqués dans la mise en œuvre de cette stratégie. Le Docking Moléculaire est une technique computationnelle appliquée à la modélisation moléculaire pour prédire/simuler la façon dont les biomolécules se lient entre elles. Elle est essentielle pour comprendre les processus biologiques et développer des médicaments. Le problème du Docking Moléculaire consiste à trouver la meilleure orientation pour qu'un ligand se lie à une molécule cible, en tenant compte de la nature dynamique des deux [95]. Des algorithmes de Docking flexibles ont été développés pour intégrer la flexibilité du ligand et la conformation du récepteur [346]. Le processus nécessite les structures 3D du récepteur et des ligands, obtenues par des méthodes expérimentales ou computationnelles. Les algorithmes de recherche conformationnelle permettent d'échantillonner les conformations de ligands stables, et les fonctions de score évaluent les énergies d'interaction pour classer les poses de Docking sur la base de leurs affinités en termes de liaison [347]. Les fonctions de score spécifiques utilisées peuvent varier en fonction des paramètres de base utilisés par le logiciel de Docking [348].

6.2.2 Représentation du système macromoléculaire (protéine/enzyme)

La plupart des techniques de Docking moléculaire reposent sur des modèles de mécanique et dynamique moléculaire (MM et DM) qui représentent de manière simplifiée le squelette central des chaînes de protéines [94]. Ces modèles moléculaires permettent de simplifier la structure des protéines en atomes, où chaque atome est représenté par un point sur lequel ses propriétés sont calculées.

6.2.3 Fonctions de score

Différentes fonctions de score et leurs catégories correspondantes (basées sur la physique, empiriques, basées sur la connaissance statistiques, apprentissage automatique (ML)) sont utilisées pour évaluer l'affinité entre le ligand et le récepteur [95]. Ces fonctions de score visent à estimer l'énergie et la stabilité du complexe (Récepteur-ligand). Le **Tableau II.11** donne un bref récapitulatif des fonctions de score et de leurs classes correspondantes utilisées par certains outils de Docking moléculaire.

Tableau II.11. Exemples de fonctions de score et de leurs classes correspondantes employées par certains outils de Docking moléculaire.

Outils	Fonctions de score	Classe
AutoDock 4 [349], [350]	Fonction d'énergie libre empirique (	Combinaison du calcul basé sur le champ de force et de l'énergie empirique, incluant les interactions de van der Waals, la liaison hydrogène, l'électrostatique et les énergies de désolvatation.
AutoDock Vina [351]	Fonction de score empirique avec un algorithme d'optimisation avancé	Fonction de score empirique utilisant les interactions de van der Waals, la liaison hydrogène et l'électrostatique. Cette fonction repose sur un algorithme optimisé qui permet d'explorer rapidement l'espace conformationnel.
Schrödinger Maestro [352]	Prime MM-GBSA, GlideScore et autres fonctions de score disponibles	Combiner les champs de force basés sur la physique avec les scores basés sur l'empirisme.
Biovia Discovery Studio [353], [354]	CDOCKER, LigScore, et autres	Combiner des champs de force basés sur la physique, des termes énergétiques empiriques et des potentiels basés sur la connaissance.
PyRx [355]	AutoDock Vina, Smina, et autres	PyRX applique diverses fonctions de score disponibles dans AutoDock Vina et Smina.
SYBYL [356]	ChemScore, D-Score, G-Score, et autres	Intégrer des termes énergétiques empiriques, des interactions de van der Waals, des liaisons hydrogène et d'autres critères pour estimer l'affinité de la liaison.

6.2.4 Estimation du score des poses

L'évaluation et la classification des conformations des ligands dans le site actif d'une protéine sont des tâches critiques dans le processus de Docking moléculaire. Cette évaluation dépend essentiellement d'une fonction de score qui doit être suffisamment fiable pour distinguer entre les différentes conformations de ligands générées dans le site actif ciblé et les classer. Une fonction de score est considérée efficace et précise si elle est capable de distinguer les composés actifs des composés inactifs sur la base de leur affinité ou de leur énergie de liaison vis-à-vis de la protéine, ce qui conduit ensuite à leur classification.

6.2.5 Méthodes de Docking moléculaire adaptées en vue de détecter les conformations optimales protéine-ligand

Dans le cadre du processus de Docking moléculaire, diverses stratégies sont adoptées pour prédire l'orientation optimaux des ligands, le profil des interactions et l'affinité entre la protéine et le ligand. Les approches aveugle "blind", rigide, semi-flexible et flexible sont les démarches les plus utilisées dans le processus de criblage virtuel basé sur le Docking moléculaire [357]. La méthode dite "blind Docking" consiste à analyser l'ensemble de la surface de la protéine pour identifier les sites de liaison potentiels [358]. Le Docking rigide repose sur l'hypothèse que le ligand et la protéine sont tous deux rigides, tandis que le Docking semi-flexible autorise la flexibilité d'un composant tout en maintenant la rigidité de l'autre. Enfin, la méthode dite "flexible" prend en compte la flexibilité du ligand et de la protéine. Chaque stratégie a des avantages et des limites : la méthode "Blind Docking" est utile lorsque le site de liaison (site actif) est inconnu, la méthode "Rigide Docking" est efficace sur le plan computationnel, la méthode "Semi-Flexible Docking" est un équilibre entre flexibilité et efficacité, et la méthode "Flexible Docking" offre la modélisation la plus complète au prix d'une augmentation des ressources computationnelles. Le choix de l'approche dépend des besoins spécifiques de la recherche

(comparaison des méthodes, évaluation, criblage des données, etc.), ainsi que sur les ressources et les connaissances disponibles. Concernant les résultats de chaque méthode, il y aura forcément une différence significative, même si le même système (protéine-ligand) est étudié avec le même outil de Docking. D'autre part, il existe également d'autres facteurs qui contribuent fortement à la différence entre les résultats et les perspectives des simulations de Docking moléculaire mises en œuvre dans le cadre de l'étude d'un même système (protéine-ligand), même si c'est de la même méthode. C'est ce que nous avons noté dans certaines de nos études lors du re-Docking des médicaments Foretinib et Erlotinib sur les récepteurs c-Met et EGFR-TKD, respectivement [100], [324], [359], [360]. Nous avons remarqué que les énergies de liaison de ces deux médicaments à leurs récepteurs protéiques changeaient entre les deux études en raison de légères modifications apportées aux paramètres d'insertion, aux méthodes d'optimisation structurelle et énergétique des ligands et du récepteur, aux versions du même logiciel de Docking moléculaire et à la flexibilité des ligands et/ou de la poche active spécifique. A cet égard, dans l'aspect méthodologique appliquée dans de nos travaux de recherche, nous avons utilisé notamment les méthodes semi-flexibles et/ou flexibles ou les deux à la fois pour mettre en œuvre le Docking moléculaire. Quant à l'évaluation de la validité de la partie technique de la méthodologie mise en œuvre dans nos travaux, nous avons fait appel à deux grandes notions : (1) connaître et identifier soigneusement le site actif du récepteur ciblé et (2) s'assurer que les ligands étudiés ne divergent pas des sites de liaison définis.

6.2.6 Limites de la stratégie de Docking moléculaire

Même si le criblage virtuel basé sur le Docking moléculaire est extrêmement utile dans l'identification et l'optimisation des molécules médicamenteuses candidates et dans le repositionnement des médicaments, il n'est pas exempt de lacunes et de limites [361], [362], [363]. En général, les limites de ce procédé peuvent être définies en deux volets, un volet technique lié à des logiciels, et un volet méthodologique lié aux algorithmes développés à cet égard. Les points suivants récapitulent les limites les plus significatives relevées :

- La diversité des logiciels/protocoles, des fonctions de classement et des approches de recherche conformationnelle adoptés dans les procédures de Docking moléculaire.
- Difficulté de choisir et de favoriser un outil informatique approprié en termes absolus, ou de dégager des conclusions générales sur son pouvoir prédictif.
- Limitations des algorithmes de Docking et des fonctions de score dans la production de prédictions pour tous les types de systèmes biomoléculaires tels que les grands complexes protéine/peptide, les protéines contenant des métaux, les interactions covalentes, les molécules fortement chargées ou polaires, et les protéines très flexibles. Cela a conduit au développement de nouvelles stratégies de Docking et des fonctions de score spécifiques qui se limitent à certaines classes de systèmes biomoléculaires.
- Les limites des algorithmes de Docking moléculaire incluent les défis de la sélection des conformations, la négligence de la flexibilité totale des protéines et des ligands (dans le cas de Docking rigide), la négligence des effets des solvants, etc.

- En outre, le traitement des molécules d'eau et la prédiction précise du site de liaison peuvent constituer des défis lors de la mise en œuvre de stratégies de Docking.
- La validation expérimentale ou computationnelle avec des techniques complémentaires est souvent nécessaire pour soutenir et valider les résultats de Docking.

6.2.7 Perspectives pour relever certains défis de la stratégie du Docking moléculaire

Il ne fait pas de doute que les perspectives actuelles et futures offertes par les techniques de Docking moléculaire sont une source d'innovation et d'aide à la conception de nouveaux médicaments. Afin d'améliorer l'adéquation des outils de Docking, il est recommandé de développer en priorité de nouvelles fonctions de score permettant de simuler la plupart des phénomènes, sans entraîner un surcoût trop important au niveau des calculs. D'autre part, il faudrait chercher la possibilité d'intégrer des fonctions de score qui tiennent compte de la flexibilité totale de la protéine et des effets de l'entropie et du solvant dans les outils de Docking les plus utilisés et les plus fiables. En intégrant certaines méthodes de simulation biophysique computationnelle, telles que les simulations de dynamique moléculaire ou les méthodes de mécanique moléculaire, aux prédictions des modèles de Docking moléculaire disponibles, il est possible de surmonter un grand nombre des limitations susmentionnées. Cependant, les prédictions basées uniquement sur des stratégies de Docking moléculaire sans investigations post-Docking sont très limitées. Par conséquent, il serait très utile que les spécialistes de la bio-informatique se concentrent sur le développement de nouveaux outils de simulation, tels que des codes open-source, qui relèvent les principaux défis auxquels sont confrontés les logiciels conventionnels utilisés dans le criblage virtuel basé sur le Docking moléculaire.

6.2.8 Modèle du Docking Moléculaire mis en œuvre dans cette thèse

A travers nos convictions dans cette thèse, on justifie la stratégie de conception rationnelle d'un nouveau médicament basée sur une approche hybride des méthodes de CADD. Compte tenu des défis et des spécificités présentés par l'approche de Docking moléculaire à ce stade, indépendamment de sa pertinence, cette stratégie doit être justifiée en répondant aux questions " Que voulons-nous atteindre, pendant combien de temps, à quel moment et avec quelle précision ?". La réponse à ces interrogations est la clé de la stratégie de criblage virtuel basée sur le Docking moléculaire et des évaluations post-Docking mises en œuvre dans le cadre de cette thèse. Ici, notre objectif principal est de filtrer rapidement les données d'entrée des molécules et de les classer en fonction de leurs caractéristiques de compatibilité avec la structure des récepteurs protéiniques/enzymatiques identifiés expérimentalement et ciblés dans le cadre de chaque étude. Pour ce faire, nous procédons à des comparaisons rationnelles et concrètes entre nos données arrimées et les médicaments utilisés à titre de co-ligands réels cristallisés dans la poche active du récepteur ciblé. Ces médicaments sont utilisés ici comme références pour valider et guider nos études étape par étape. Tout cela dans le but de générer des hypothèses précises et de haute qualité sur les conformations bioactives des ligands (inhibiteurs de protéines/enzymes) qui sont criblées en tant que nouveaux médicaments candidats. Le schéma présenté dans la **Figure II.22** résume les principales phases exécutées dans le cadre du

protocole de Docking moléculaire mis en œuvre dans la majorité de nos études publiées dans le cadre de cette thèse. Le modèle de criblage virtuel basé sur le Docking moléculaire mis en œuvre dans nos travaux a comporté trois grandes phases. La première consiste à rationaliser l'application du modèle afin qu'il soit compatible avec le contexte de l'étude envisagée. La deuxième phase consiste à valider le modèle de Docking moléculaire à appliquer. Troisièmement, implémentation du modèle de Docking moléculaire pour cribler les ligands et les filtrer selon leurs affinités et compatibilités avec le récepteur ciblé. Dans les sous-sections suivantes, nous décrivons les phases les plus intéressantes que nous avons suivies pour la mise en œuvre du processus de Docking moléculaire appliqué dans cette thèse.

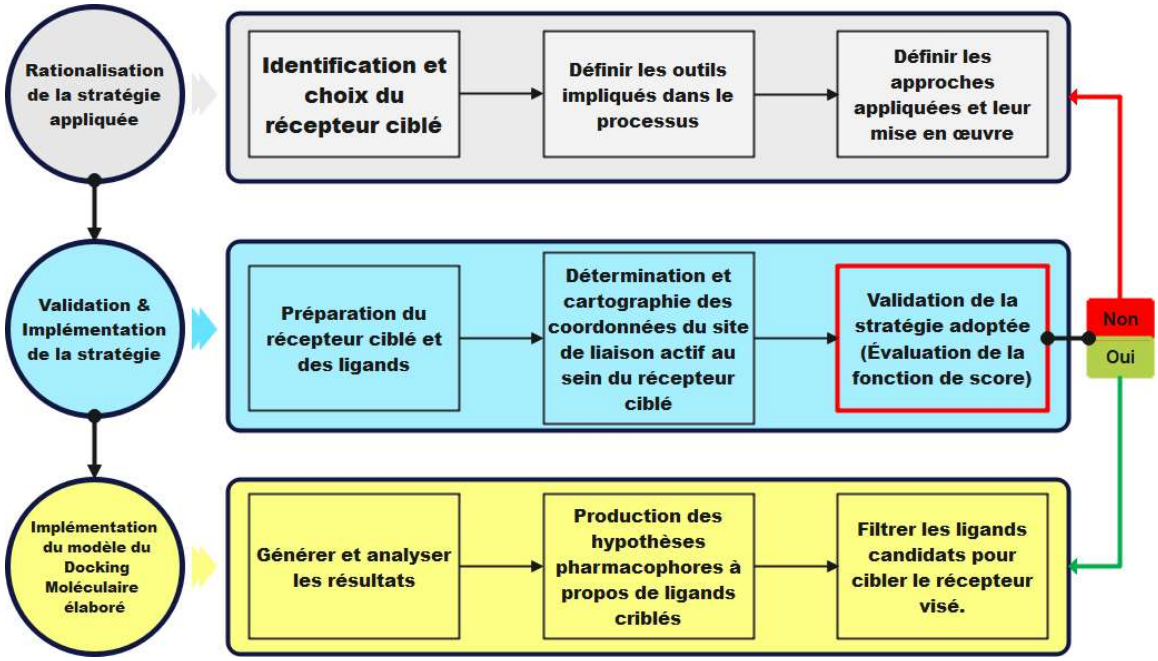


Figure II.22. Stratégie globale du modèle de criblage virtuel basé sur le Docking moléculaire appliqué dans nos études.

6.2.8.1 *Identification de la cible biologique*

Nous identifions les cibles biologiques/récepteurs impliqués dans une maladie ciblée à partir des publications, puis nous obtenons leur structure 3D à partir de RCSB PDB. Une fois que la structure est choisie, nous effectuons des études CADD *in silico* en fonction de plusieurs paramètres, tels que la qualité cristallographique, le procédé expérimental, l'organisme à l'origine et la présence de ligands de référence cristallisés au niveau du site de liaison actif (poche).

6.2.8.2 *Adaptation des outils, approches et algorithmes appropriés*

Pour rendre notre stratégie du Docking moléculaire pertinents, nous avons utilisé les packages d'Autodock (AutoDock4 et AutoDock Vina), qui ont été développés pour surmonter certaines limitations des versions précédentes d'AutoDock [364]. AutoDock4 comprend un champ de force semi-empirique avec des corrections basées sur les charges, un modèle de solubilité avec des termes relatifs à tous les types d'atomes, et un modèle optimisé de l'orientation des liaisons hydrogènes [364]. Autodock Vina optimise également la précision des prédictions de position et peut utiliser plusieurs threads pour des calculs parallèles sur des machines multi-cœurs [351]. La **Figure II.23** montre l'approche d'AutoDock pour estimer l'énergie libre de liaison - calculée en deux étapes : évaluation de l'énergie intramoléculaire et intermoléculaire pour passer de l'état non lié à la conformation liée du complexe protéine-ligand

[364]. L'énergie libre de liaison ($\Delta G = \Delta H - T\Delta S$) est calculée sur la base de la combinaison des contributions enthalpique et entropique : ΔH et TΔS respectivement [365]. Le terme enthalpique favorable est dû aux interactions et à la formation de liaisons intermoléculaires, tandis que le terme entropique tend à augmenter à cause de la diminution des degrés de liberté de torsion. La valeur négative de ΔG indique que la liaison est favorable du point de vue énergétique. L'estimation de l'énergie de liaison libre ΔG définie par l'**Eq**. **(II.9)** inclut six évaluations par paire (V) et une estimation de l'entropie harmonique manquante sur la liaison (ΔS_{conf}) [364]. Les termes atomiques par paire intègrent des évaluations de la dispersion/répulsion, de la liaison hydrogène, des propriétés électrostatiques et de la solubilité au niveau atomique.

$$\Delta G = (V_{lié}^{L-L} - V_{non-lié}^{L-L}) + (V_{lié}^{P-P} - V_{non-lié}^{P-P}) + (V_{lié}^{P-L} - V_{non-lié}^{P-L} + \Delta G_{conf}) \qquad \text{(II.9)}$$

Dans **l'Eq. (II.9)**, *P* désigne la protéine, *L* désigne le ligand, *V* sont les évaluations par paire indiquées précédemment, et ΔS_{conf} désigne la perte d'entropie conformationnelle lors de la liaison. Dans le cadre de nos travaux, nous avons utilisé un algorithme génétique lamarckien (AGL), une fonction d'énergie libre intégrée dans AutoDock 4, l'algorithme de Broyden-Fletcher-Goldfarb-Shanno (BFGS) et la recherche locale dans AutoDock Vina pour estimer les énergies de liaison et d'affinité récepteur-ligand. Ces résultats permettent de classer les ligands en fonction de leur énergie de liaison/affinité et de calculer les constantes d'inhibition théoriques correspondantes (K_i). Avec ($K_i = e^{\frac{\Delta G}{R \times T}}$), où ΔG, R et T sont l'énergie de liaison du Docking, la constante des gaz (1.985×10^{-3} kcal $K^{-1} \times mol^{-1}$) et la température ambiante (298.15 K), respectivement.

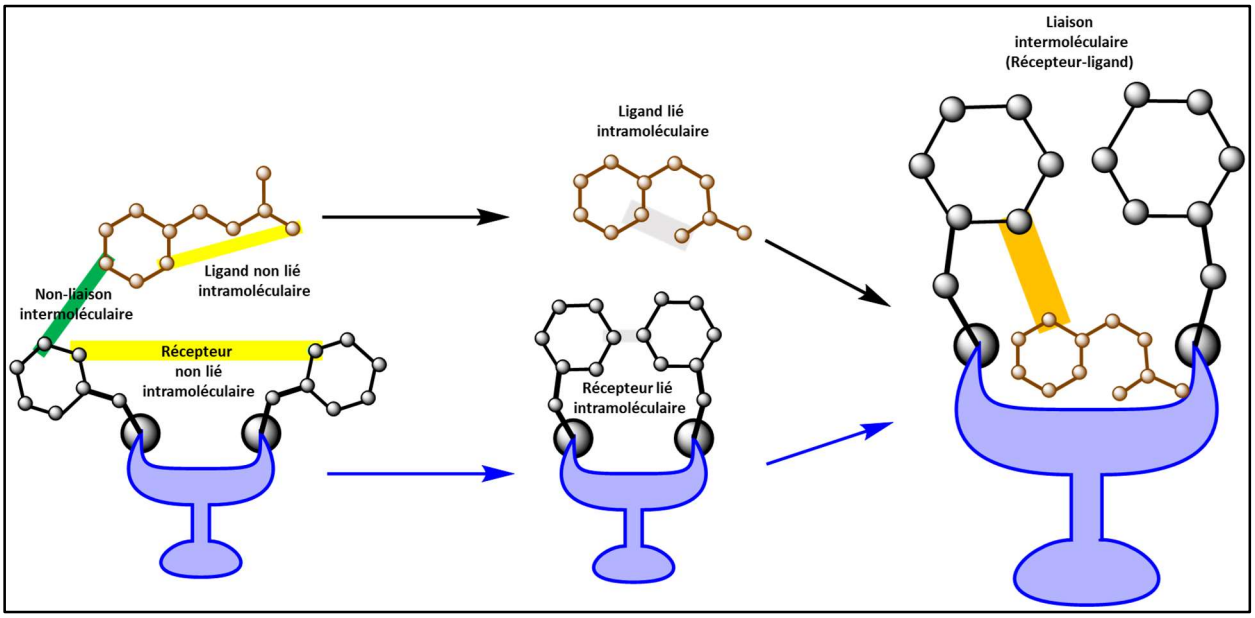


Figure II.23. L'approche du Docking moléculaire utilisée par le logiciel AutoDock.

6.3 Méthode adoptée pour la validation du modèle du Docking moléculaire

Après avoir défini la stratégie et les méthodes impliquées dans ce processus, il s'agit de valider la pertinence de ce protocole de Docking moléculaire. Cette étape du protocole de Docking moléculaire consiste à évaluer la fonction de score appliquée pour élaborer le modèle proposé. On utilise pour cela

un complexe protéine-ligand observé expérimentalement comme situation de contrôle. Une bonne position d'alignement est obtenue lorsque l'écart quadratique moyen (RMSD) entre les ligands redocké et expérimental (co-cristallisé) est faible (RMSD ≤ 2 Å), calculé à partir de la distance moyenne entre les atomes des deux entités [364]. Dans ce cadre, nous avons choisi d'utiliser le re-Docking moléculaire et/ou la dynamique moléculaire pour valider la précision des modèles décrits basés sur le Docking moléculaire.

6.3.1 Outils de Docking protéine-ligand utilisés

Dans cette thèse, plusieurs outils ont été utilisés dans l'élaboration du modèle de Docking moléculaire, conformément aux étapes et aux besoins fixés. Le **Tableau II.12** indique les outils / logiciels utilisés lors du processus de Docking moléculaire décrit dans la **Figure II.22**. Après avoir généré les résultats, la phase finale de cette procédure est l'analyse des données et la formulation d'hypothèses pharmaceutiques sur les ligands candidats en tant qu'agents capables de cibler le récepteur visé. Au cours de cette phase, nous examinons visuellement les profils d'interactions protéine-ligand et les énergies de liaison prévues pour filtrer les ligands. À cette fin, nous comparons les résultats obtenus avec la version expérimentale du modèle protéine-ligand qui a été adoptée comme référence.

Tableau II.12. Logiciels de Docking protéine-ligand employés dans le cadre de cette thèse.

Étape	Logiciel/outil	Rôle
Identification et choix du récepteur ciblé	RCSB PDB (https://www.rcsb.org/)	Obtenir la structure 3D du récepteur.
Préparation protéine	MOE, MGLTools, et Discovery Studio 2016	La structure protéique est simplifiée et l'espace de recherche est réduit par l'élimination de l'eau et d'autres éléments non protéiques. Les atomes manquants sont ajoutés par une modélisation moléculaire basée sur la structure connue de la protéine. Une fois les atomes ajoutés, des charges et des types atomiques appropriés sont assignés.
Préparation des ligands	ChemSketch, ChemOffice, GaussView 6.0, Gaussian 09, SYBYL-X2.1.1, et MGLTools, OpenBabel	Pour garantir une représentation correcte de sa géométrie et des interactions potentiels, il est nécessaire de rajouter des hydrogènes et d'assigner les charges partielles et les types atomiques appropriés aux atomes du ligand. De plus, une multi-optimisation de la structure des ligands est requise pour évaluer l'équilibre du système et vérifier l'état des atomes d'hydrogène et des protons de liaison dans un environnement aqueux similaire à celui des protéines.
Identification du site actif	Discovery Studio 2016 et ProteinsPlus [366]	Identification des résidus des acides aminés actifs dans la poche réactive du récepteur biologique à cibler.
Définir l'espace de recherche	MGLTools/ AutoGrid4	Fixer la dimension et la position du centre de la grille de docking moléculaire.
Implémentation du docking moléculaire	AutoDock 4 et/ou AutoDock Vina	Implémentation du docking des ligands dans la poche active du récepteur à cibler.
Évaluation de la fonction de score	MGLTools, Chimera 1.15, et DockRMSD	Évaluation du niveau d'alignement (RMSD) entre le ligand original (Co-cristallisé) dans le complexe obtenu expérimentalement avec le même ligand remodelé et re-docké par la stratégie de docking moléculaire appliquée.

Générer des résultats	Discovery Studio 2016, Suite du logiciel MGLTools	Générer et visualiser des résultats relatifs aux modes de liaison et aux interactions prédites au niveau des complexes récepteur-ligand.

6.3.2 Stratégies des Analyses post-Docking appliqué

Bien que le Docking moléculaire occupe toujours une part importante dans le processus de conception et de découverte des médicaments, il présente des limites considérables, notamment celles de supposer que le site de liaison de la protéine est connu et limité en raison des contraintes imposées par la protéine, et également de ne pas prendre en compte les effets des molécules d'eau comme de solvant dans la génération des profils d'interaction et des énergies de liaison. Cela conduit souvent à des prédictions de poses de Docking protéine-ligand instables ou parfois à des écarts considérables du ligand par rapport au site de la poche active ciblée. Par conséquent, les poses de Docking prédites doivent être purifiées afin de valider la stabilité et la compatibilité structurelle des complexes générés. Cela peut se faire grâce à des simulations de mécanique moléculaire (MM) et de dynamique moléculaire (MD) lentes et plus avancées [367], [368], [369]. Parallèlement, les fonctions de score de Docking moléculaire mènent fréquemment à des prédictions de type faux positif ou faux négatif en raison des différentes approximations. Pour pallier ces lacunes, des méthodes de calcul plus avancées, telles que MM-PBSA, MM-GBSA, la perturbation de l'énergie libre (FEP), l'intégration thermodynamique (TI) et l'énergie d'interaction linéaire (LIE), peuvent être appliquées afin d'obtenir des résultats plus précis, même si leur exécution prend plus de temps [370]. Parmi les possibilités à exploiter pour améliorer les résultats de la stratégie de CADD hybride mise en œuvre dans nos travaux, nous avons intégré les calculs de MM-GBSA et/ou de Dynamique Moléculaire (MD) en tant que démarche ultime post-Docking.

6.3.3 Préparation de systèmes macromoléculaires protéine-ligand

Avant de procéder à la mise en œuvre de la procédure de post-Docking, les systèmes macromoléculaire (Protéine-ligand) doivent être préparés en optimisant leurs structures moléculaires et énergétiques. À ce stade, la spécificité du système à examiner et à analyser est le critère de sélection de la méthode d'optimisation appropriée. Les méthodes utilisées pour optimiser les systèmes moléculaires vont des petites molécules avec un nombre limité d'atomes aux molécules géantes telles que les protéines, qui servent actuellement comme modèle. Concernant notre démarche de post-Docking qui repose sur des systèmes biomoléculaires géants (complexes protéine-ligand), nous utilisons des champs de force et des méthodes d'optimisation compatibles avec ces systèmes en guise de phase préparatoire.

6.3.4 Champs de force exploités dans l'optimisation des systèmes protéine-ligand

Les champs de force à résolution atomistique (par exemple, United-Atoms et All-atoms) sont utilisés pour les systèmes comportant des molécules dont les atomes et les liaisons sont localisés avec précision, comme dans les protéines et les enzymes [371]. Parmi les champs de force à résolution atomistique, on peut citer les champs de force de la famille AMBER, CHARMM et OPLS. Par ailleurs,

les champs de force dits " gros grains / ou coarse-grained " sont utilisés pour les systèmes ne comportant pas la localisation exacte des atomes et des liaisons, tels que les simulations à méso-échelle [372]. Les champs de force MARTINI et modèle de polymère MARTINI sont deux exemples des champs de force utilisés à gros grains [372]. Ces champs de force permettent des simulations plus faciles et plus rapides, mais ils reposent sur certaines hypothèses qui peuvent entraîner des insuffisances dans la simulation. La **Figure II.24** présente une illustration du squelette d'une protéine à l'aide d'un modèle atomistiques « All-atoms » par rapport à des modèles à gros grains, ainsi que des paysages énergétiques [373].

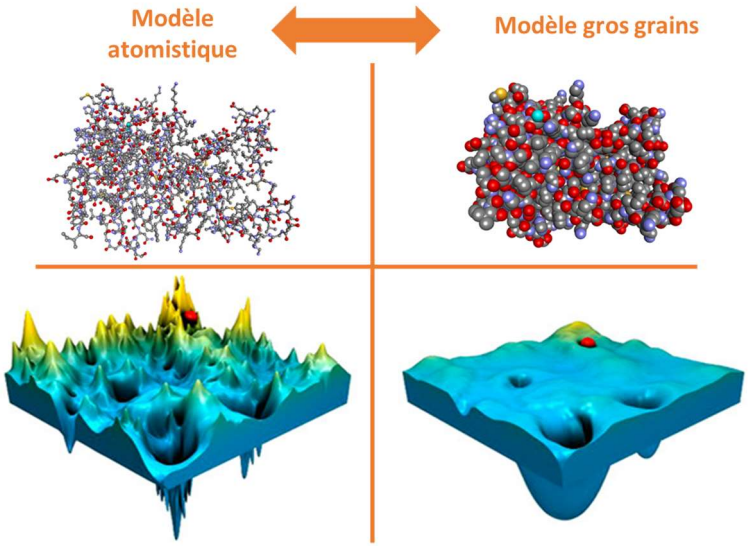


Figure II.24. Présentation du squelette protéique via deux modèles : Atomistiques et gros grains (illustration en haut), modèles de paysage énergétique (illustration en bas).

Généralement, les champs de force basés sur le modèle atomistique permettent de prédire avec précision les propriétés des systèmes en dépit de la complexité de l'intensité de calcul par rapport aux champs de force à gros grains [373]. Ces derniers sont moins intensifs en calcul et modélisent le système à une échelle macroscopique plus grande grâce à l'utilisation d'un moins grand nombre d'atomes pour définir chaque résidu. Par conséquent, la simulation est plus rapide, mais elle n'a pas la précision fournie par le modèle atomistique. À cet égard, les champs de force les plus utilisés pour étudier la dynamique des biomolécules sont CHARMM, AMBER, OPLS et GROMOS [374]. Malgré le développement de ces champs de force au cours de la dernière décennie de façon la plus complète possible, le choix se porte souvent sur les champs de force CHARMM, AMBER et OPLS pour les systèmes protéiques. Tandis que le champ de force GROMOS est souvent choisi pour étudier les lipides et les membranes.

6.3.5 Optimisation énergétique des systèmes protéine-ligand

La nécessité d'optimiser l'énergie des systèmes protéine-ligand est liée au fait qu'après le Docking, le complexe protéine-ligand peut ne pas avoir atteint une conformation stable, de type natif, avec l'énergie la plus basse possible. Pour optimiser l'énergie du complexe, divers algorithmes et techniques tels que la dynamique moléculaire, la méthode de Monte Carlo, l'optimisation de la distance géométrique et les fonctions de score basées sur la connaissance (knowledge-based scoring functions)

sont appliqués pour raffiner et optimiser les poses de Docking et évaluer avec précision l'affinité de la liaison [375]. Cela représente un moyen précieux de limiter les faux positifs et de renforcer la probabilité de découvrir de véritables liaisons entre la protéine et les ligands examinés.

6.3.6 Optimisation du solvant

Une grande partie des systèmes moléculaires se développent dans un solvant, qui est généralement l'eau. Par conséquent, il est tout à fait évident que les systèmes sont également solvatés dans l'eau lors des simulations biophysiques réalisées sur la base d'approches computationnelles ou expérimentales. Le solvant est donc un paramètre indispensable qui doit être pris en compte en raison de son effet/interaction direct avec la biomolécule étudiée. La modélisation d'une solution peut être réalisée en construisant une "boîte" contenant une unité de soluté au milieu de multiples molécules de solvant. De cette manière, on évite de simuler un système complètement isolé dans l'espace, ce qui conduirait à l'évaporation d'atomes situés au bord du système à 300 K, et donc à la perturbation du système (effet de bord). Pour éviter cela, le système doit être représenté comme un système périodique connecté sur les six bords, où une molécule qui sort d'un bord entre dans le système de l'autre bord. Cette représentation est connue sous le nom des conditions aux limites périodiques (CLP) (Periodic Boundary Condition, ou PBC) [376]. Cela permet de simuler un nombre limité de molécules de solvant tout en éliminant les effets de bord. Pour minimiser l'énergie d'un système solvant-soluté, il existe divers modèles de solvants que l'on peut classer dans trois catégories : flexibles, polarisables et rigides [377]. Parmi ces modèles, SPC (simple point charge) et TIP3P (Transferable Intermolecular Potential 3 Points) sont les plus utilisés [378]. Ils attribuent une charge ponctuelle à chacun des trois atomes représentés, l'atome d'oxygène étant spécifié par les paramètres de Lennard-Jones.

6.4 Estimation des affinités de liaison du ligand via MM-GBSA et MM-PBSA

Pour évaluer les énergies libres de liaison entre deux biomolécules, les méthodes de mécanique moléculaire de surface de Poisson-Boltzmann (MM-PBSA) et de mécanique moléculaire de surface généralisée (MM-GBSA) sont généralement associées aux résultats des simulations de DM des ensembles de structures complexes (Protéine-ligand) extraites à partir des trajectoires générées par la simulation de DM. Les énergies libres de liaison estimées par les méthodes MM-PB(GB)SA sont calculées à partir du modèle défini par **l'Eq. (II.10)** [3].

$$\Delta G_{bind} = \Delta G_{Complexe} - \left(\Delta G_{Protéine} + \Delta G_{Ligand}\right) \qquad \textbf{(II.10)}$$

6.4.1 Outils de mise en œuvre

Différents outils/packages destinés dans le cadre des méthodes MM-PBSA et MM-GASA permettent d'évaluer ΔG_{bind} de complexes obtenus après le Docking moléculaire, tels que NAMD, AmberTools, Gromacs, Schrödinger, et d'autres [379], [380]. En ce sens, dans nos travaux, nous avons utilisé le package de logiciel Schrödinger 2020-3 pour évaluer (ΔG_{bind}) et valider les hypothèses générées à partir de la Docking moléculaire, notamment en ce qui concerne l'affinité et la stabilité des interactions protéine-ligand prédites par le docking moléculaire. Dans nos travaux, en utilisant le

modèle Prime MM-GBSA et le script Thermal_MM-GBSA implémenté dans Schrödinger 2020-3 afin de calculer ΔG_{bind}. Ce dernier est obtenu par la somme d'un ensemble de paramètres de liaison libres tels que l'énergie de liaison hydrogène (ΔG_{bind} H-bond), l'énergie de van der Waals (ΔG_{bind} vdW), l'énergie covalente (ΔG_{bind} covalent), l'énergie coulombienne (ΔG_{bind} Coulomb), et l'énergie lipophile (ΔG_{bind} Lipo), l'énergie de solvatation électrostatique de Born généralisée (ΔG_{bind} Solv_GB) et la correction des interactions π-π (ΔG_{bind} pi-pi packing) [217].

6.4.2 Mise en œuvre

Dans le cadre de nos travaux, La méthode Prime MM-GBSA est utilisée pour estimer les énergies libres de liaison des complexes protéine-ligand. Elle combine des calculs de mécanique moléculaire (MM) avec le modèle de solvatation de Born généralisé (GB) et le terme de surface (SA). La méthode MM-GBSA thermique complète la première en incluant les effets de la température et les mouvements moléculaires par le biais de simulations de dynamique moléculaire (MD) afin de capturer les variations conformationnelles et les contributions entropiques à l'énergie libre. Dans ce cadre, nous avons utilisé Prepwizard (Protein Preparation Wizard) installé sur l'interface Maestro 12.5, y inclus dans le package Schrödinger [381], à un pH de 7 ± 2. Les systèmes protéine-ligand sont préparés selon le champ de force OPLS3e et le modèle de solvant VSGB 2.1 [382]. Pendant cette préparation, des atomes d'hydrogène ont été ajoutés à la structure 3D des complexes, les liaisons disulfures manquantes ont été formées, des états d'ionisation/tautomères ont été générés pour toutes les parties (chaînes latérales d'acides aminés et ligands), les liaisons hydrogènent et les optimisations fonctionnelles basées sur le système ont été réalisées à l'aide du champ de force OPLS3e (Optimized Potential for Liquid Simulations 3e) [383]. En outre, la fonction Epik (Empirical pKa Prediction) a été incorporée pendant la préparation du système afin de fournir une résolution appropriée pour les états de protonation des ligands pendant les étapes de découverte des médicaments [384]. En effet, le pKa des groupes fonctionnels joue un rôle important dans la définition du profil pharmacocinétique et du comportement dynamique structurel des groupes fonctionnels qui caractérisent le médicament. Le module Epik que nous avons utilisé nous a permis de prédire de manière fiable l'état protoné des ligands, ce qui nous a permis de comprendre les interactions spécifiques qui permettent aux ligands de se lier à la protéine ciblée. Cependant, dans le criblage virtuel des chimiothèques, nous constatons qu'il existe plusieurs outils qui fournissent des structures rigides connues contenant tous les groupes fonctionnels, mais qui ne prennent pas en compte les états de protonation et de tautomérie des ligands.

6.5 Simulations de la dynamique moléculaire (DM)

Grâce à la mécanique moléculaire, il est possible d'estimer la position des atomes dans une molécule et la conformation d'énergie minimale correspondante. Cette approche ne tient pas compte des vibrations ou autres mouvements qui se produisent à des températures supérieures à 0K. C'est pourquoi la dynamique moléculaire est appliquée pour examiner la dynamique moléculaire, l'entropie et les effets liés aux changements de température [258]. La dynamique moléculaire simule le comportement dynamique des molécules en attribuant des vitesses aux atomes du système et permet d'explorer plus précisément la surface d'énergie potentielle. La flexibilité des structures présentes dans

la boîte de simulation nous permet de simuler des mouvements intramoléculaires et de suivre les transformations de la conformation sous l'effet du champ de force et des vitesses atomiques, quand les lois de Newton sur le mouvement sont appliquées [377].

6.5.1 Définition et principe fondamental des simulations de la DM

La dynamique moléculaire est une méthode computationnelle utilisée pour simuler le mouvement et les interactions des molécules à l'échelle atomique ou moléculaire [385]. Les simulations DM utilisent la mécanique newtonienne pour prédire les trajectoires des molécules sur la base de leurs forces d'attraction et de répulsion pendant des pas de temps très courts, ce qui permet de comprendre le comportement du système à long terme [386]. La simulation de la dynamique moléculaire peut être appliquée à la science des matériaux, ainsi qu'à la biophysique et à la biochimie pour raffiner les structures 3D des protéines et d'autres macromolécules. Il est possible de déterminer les propriétés (ou les caractéristiques) des atomes ou des particules d'un système moléculaire donné par l'étude de la trajectoire de chaque atome ou particule au cours du temps. Pour ce faire, on applique les lois de la mécanique classique à tous les atomes de ce système simultanément et on résout les équations classiques du mouvement :

$$\vec{F} = m_i \vec{a}_i \qquad \text{(II.11)}$$

Selon l'**Eq. (II.11)**, $\vec{F}$ exprime la somme des forces exercées sur l'atome i, m_i la masse de l'atome et $\vec{a}_i$ son accélération.

À l'aide d'un ordinateur sophistiqué, les équations différentielles de mouvement du modèle newtonien peuvent être résolues, ce qui permet de générer une trajectoire de simulation de DM. Le mouvement d'un atome est donc déterminé par l'ensemble des forces exercées sur celui-ci par les atomes voisins, en fonction du temps. Donc cette mesure dépend du placement r_i de l'atome au pas de temps t. La vitesse v_i et la position $r_i$de chaque atome sont liées à l'énergie cinétique et à la température du système, conformément au modèle de distribution de Maxwell-Boltzmann dans l'**Eq. (II.12)** [386].

$$F(v)dv = \left(\frac{m}{2\pi KT}\right)^{\frac{3}{2}} \left(e^{-\frac{mv^2}{2KT}}\right) 4\pi n v^2 dv \qquad \textbf{(II.12)}$$

où, n représente le nombre moyen des molécules par unité de volume; K est la constante de Boltzmann, égale à 1.3806452×10^{-23} Joules/Kelvin; m est la masse d'une molécule; T est la température du système. La difficulté ici est donc de pouvoir simuler la position de l'atome au temps t + Δt à partir sa position dans le temps t. Ce point peut être abordé plus avant par l'évaluation de la force totale agissant sur chaque particule (F_n). Cela peut se faire à l'aide du champ de force de la mécanique moléculaire grâce à la fonction potentielle V_r décrit dans l'**Eq. (II.13)** [3].

$$V_r = \sum_{étirement} \frac{k_b}{2}(l_i - l_{i,0}) + \sum_{angles} \frac{k_\theta}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2}\left[1 + \cos(n\omega - \gamma)\right]$$
$$+\sum_{i=1}^{N}\sum_{j=i+1}^{N} 4\varepsilon_{ij}\left(\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right) \quad \textbf{(II.13)}$$

L'**Eq. (II.13)** indique clairement que le total V_r comprend le total des interactions liante (étirement de la liaison, flexion des angles et rotations/torsions de la liaison) ainsi que les interactions non-liante (interactions de van der Waals et interactions électrostatiques). L'étirement de la liaison et la flexion de l'angle sont exprimés par les fonctions harmoniques autour de la longueur de liaison de référence $l_{i,0}$, et de l'angle de liaison de référence $\theta_{i,0}$, respectivement, où tous les autres termes du champ de force sont fixés à zéro. k_b et k_θ sont les constantes de force pour l'étirement de la liaison et la flexion de l'angle, respectivement. L'angle de torsion est exprimé par une fonction qui dépend des cosinus pour décrire et calculer l'angle dièdre entre quatre atomes liés. Dans l'expression de la torsion, V_n exprime la hauteur de la barrière, n le nombre de minimums dans la fonction lorsque la liaison est tournée jusqu'à 360°, ω l'angle de torsion et γ le facteur de phase qui spécifie l'emplacement des minimums. Les interactions non liantes, indiquées dans la partie basse de l'équation, correspondent à la somme des interactions de van der Waals, définies par le potentiel de Lennard Jones 12-6 avec les paramètres "ε : profondeur du puits de potentiel ", "σ : distance intermoléculaire limitée où la valeur du potentiel est nulle" et "r : distance entre les molécules", et des interactions électrostatiques, calculées comme une somme de paires de charges ponctuelles "q", en utilisant la loi de Coulomb. Cette sommation est généralement obtenue par la méthode de sommation d'Ewald, ou sa prolongation, la sommation d'Ewald à maille particulaire ("particle mesh Ewald", PME), qui permet d'obtenir une convergence plus rapide [387], [388]. A la suite de cela, la force totale F sera calculée en tant que gradient négatif de la fonction d'énergie potentielle V_r, avant d'être utilisée pour calculer l'accélération des particules ai(t) conformément à **l'Eq. (II.14 et II.15)**.

$$F_{ri} = -\frac{dV_r}{dr_i} \quad \text{(II.14)}$$

$$a_i(t) = \frac{d^2 r_i}{dt^2} = \frac{F_{r_i}}{m_i} = \frac{dv_i}{dt} \quad \text{(II.15)}$$

Parallèlement aux positions (r_i) et aux vitesses (v_i) obtenues, on utilise l'accélération pour calculer les positions et les vitesses lors des étapes successives de la simulation. À l'aide des algorithmes qui intègrent les équations du mouvement, il est possible d'accélérer le processus de mise en œuvre et de simulation ; il est également possible de prédire le comportement général du système, mais pas de prédire avec précision la trajectoire d'un atome en particulier. Ceci est dû à la nature chaotique du mouvement de plusieurs corps à la fois, qui rend la prédiction exacte de la trajectoire d'un atome particulier impossible. Par conséquent, le fait d'intégrer ces algorithmes permet d'obtenir les plus grands pas de temps d'intégration possibles Δt tout en maintenant l'énergie du système quasi stable ($\frac{\Delta E}{E} \approx 10^{-4}$) qui constitue la norme principale de convergence [386]. L'algorithme de Verlet dérivé du

développement de Taylor (ri (t)) fait partie des algorithmes les plus simples et les plus utilisés décrit dans l'**Eq. (II.16)** [389], [390]:

$$\vec{r}_i(t+\Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t+\Delta t) + \Delta t^2 . \vec{a}_i(t)$$
$$\vec{v}_i(t) = \frac{\vec{r}_i(t+\Delta t) - \vec{r}_i(t-\Delta t)}{2\Delta t} \quad \text{(II.16)}$$

À partir de **l'Eq. (II.16)**, on peut voir que l'algorithme de Verlet donne des positions à l'instant (t + Δt) et des vitesses à l'instant (t). Pour résoudre ce problème, plusieurs modèles plus sophistiqués ont été développés, notamment "leapfrog" [391] (**équation II.17**) ou "Verlet vitesse" (**équation II.18**) [392], dans lesquels les vitesses sont calculées à la moitié du pas d'intégration.

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + \Delta t \ . \vec{v}_i(t+\frac{\Delta t}{2})$$
$$\vec{v}_i(t+\frac{\Delta t}{2}) = \vec{v}_i\left(t-\frac{\Delta t}{2}\right) + \Delta t \ . \vec{a}i(t) \quad \textbf{(II.17)}$$

$$\vec{r}_i = (t+\Delta t) = \vec{r}_i(t) + \Delta t \ . \vec{v}_i(t) + \frac{1}{2}\Delta t^2 . \vec{a}i(t)$$
$$\vec{v}_i(t+\Delta t) = \vec{v}_i(t) + \frac{1}{2}\Delta t.\left[\vec{a}i(t) + \vec{a}i(t+\Delta t)\right] \quad \textbf{(II.18)}$$

Grâce à ces algorithmes, de nouvelles vitesses sont calculées après avoir calculé de nouvelles positions et forces tout au long de la simulation de DM. Ce calcul est effectué par étapes séparées par des intervalles de temps Δt très courts (1-2 femtosecondes, 10^{-15} s) qui sont bien inférieurs à la période du mouvement le plus rapide du système (comme les vibrations des liaisons) [386]. De tels calculs sont effectués sur des échelles de temps de centaines de nanosecondes (10^{-9} s), ou même de microsecondes (10^{-6} s), ce qui permet de générer des trajectoires atomiques qui représentent une série d'alignements moléculaires en fonction du temps.

6.5.2 Aspects techniques

6.5.2.1 Ensembles thermodynamiques

Les ensembles thermodynamiques sont des outils utilisés en dynamique moléculaire pour contrôler et simuler des variables telles que la température, la pression, le potentiel chimique et le nombre de particules. Ils sont classés selon leurs constantes (NVE, NVT, NPT, µVT) et permettent d'étudier des systèmes en équilibre thermodynamique, les transitions de phase et les propriétés thermodynamiques et cinétiques [393], [394]. L'ensemble micro-canonique (NVE) : Dans cet ensemble, le nombre de particules (N), le volume (V) et l'énergie totale (E) du système sont conservés. Il permet de simuler un système isolé, sans échange de particules, de chaleur ou de travail avec l'environnement. L'ensemble canonique (NVT) : Cet ensemble maintient la température (T), le volume (V) et le nombre de particules (N) constants. La simulation dans cet ensemble permet d'étudier des systèmes en contact thermique avec un thermostat (totaux ou locaux) [395], c'est-à-dire avec une source de chaleur externe. L'ensemble isobare-isotherme (NPT) : Dans cet ensemble, la température (T), la pression (P), le volume (V) et le nombre de particules (N) sont conservés. Il permet de simuler des systèmes en contact

thermique avec un thermostat et en contact avec un barostat (Andersen, Parrinello-Rahman, Berendsen) [396], qui maintient une pression constante. L'ensemble grand canonique (μVT): Cet ensemble maintient la température (T), le potentiel chimique (μ), le volume (V) et le nombre de particules (N) constants. Il est utilisé pour simuler des systèmes en contact avec un thermostat et un réservoir de particules, ce qui permet d'étudier des phénomènes tels que l'adsorption ou la désorption.

6.5.2.2 Conditions initiales

Les simulations de dynamique moléculaire nécessitent une sélection minutieuse des conditions initiales, telles que les positions atomiques, les vitesses, la température, la taille de la boîte et les conditions aux limites, les paramètres du champ de force et le pas de temps [258]. Ces conditions définissent l'état du système et doivent être déterminées en fonction du système et des objectifs de la simulation. Cela permet de s'assurer que les résultats produits par la simulation sont précis et significatifs. Ces conditions initiales ne devraient pas affecter le résultat de la simulation et doivent donc être choisies en fonction de la structure que l'on souhaite étudier. Pour faciliter et accélérer le processus d'équilibrage, il est préférable de choisir comme point de départ les données relatives à la condition préalable du système (solide, liquide, gazeux, etc.) à l'état proche de l'état initial, qui ont été obtenues par voie expérimentale ou computationnel.

6.5.2.3 Conditions aux limites

Au-delà des conditions initiales, il est également très important de définir les limites de la boîte de simulation qui contient le système examiné dans le milieu favorable [376]. Ceci est dû au fait que si la grille de simulation est complètement fermée ou ouverte, les comportements simulés ne seront pas réalistes et ne seront pas possibles. Dans le cas d'un système entièrement fermé, les particules/molécules du bord de la boîte seront soumises à des effets de surface virtuels, il faut donc définir une distance virtuelle entre la frontière du système et celle de la boîte de simulation. Concernant le système complètement ouvert, la densité des molécules en principe stables peut varier au cours de la simulation, en raison de leur propagation à l'infini dans l'espace de simulation. Pour résoudre ces obstacles, la solution la plus adaptée consiste à fixer des conditions aux limites périodiques (CLP) (**Figure II.25**) [377], [386]. Généralement, pour simuler un système moléculaire avec un nombre de N atomes dans un environnement spécifique (par exemple aqueux), ce système moléculaire fait partie au sein d'une boîte de dimensions Lx, Ly et Lz. Des conditions périodiques sont souvent appliquées aux bords de la boîte dans laquelle se trouve le système. Lorsque des conditions aux limites périodiques (CLP) sont appliquées, une image virtuelle de la boîte de simulation est infiniment bouclée autour d'une boîte centrale, dans toutes les directions XYZ. Dans ces conditions, chaque fois que l'atome quitte la boîte d'un côté, il est repoussé dans la boîte de l'autre côté à la même vitesse. Par conséquent, l'élan global du système dont la dynamique doit être examinée et évaluée est préservé. Il existe plusieurs modèles géométriques de systèmes volumineux utilisant les CLP qui peuvent limiter la taille des systèmes (boîtes cubiques, rectangulaires, dodécaèdres rhombiques, et tronqués) [376].

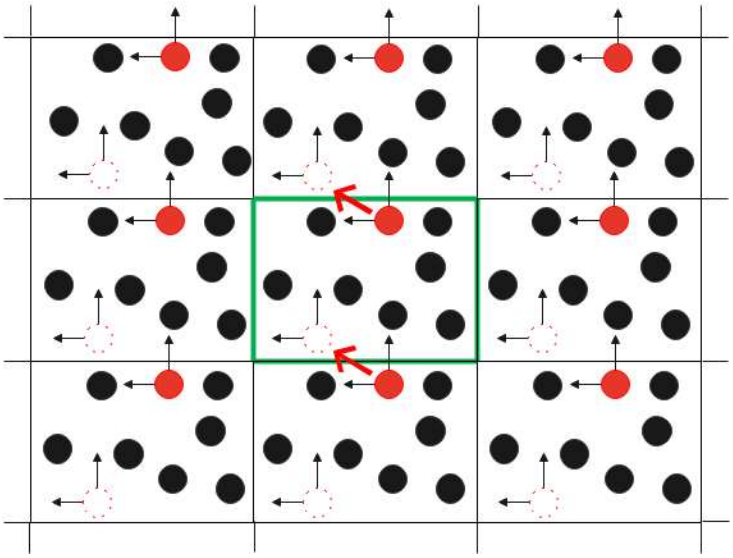


Figure II.25. Visualisation 2D des conditions aux limites périodiques (cubiques). La boîte centrale apparaît encadrée en vert ; la particule rouge se trouve à la limite de la boîte. Les flèches indiquent le mouvement de la particule au cours du pas de temps successif.

6.5.3 *Économiser le temps de la simulation MD*

Les simulations MD peuvent prendre beaucoup de temps en raison du calcul des forces résultant des interactions entre N particules. Ces interactions se traduisent par N(N-1)/2 forces entre paires et peuvent être simplifiées par l'utilisation de méthodes de coupure (cut-off) [376]. Les listes de voisins (Neighbor lists) et les listes de cellules (cell-lists) sont des techniques d'optimisation utilisées pour réduire le temps de calcul des interactions non-covalentes [376]. *Verlet Neighbor List :* Les listes de Verlet désignent une méthode de calcul plus efficace des forces entre les particules d'un système. Cette méthode consiste à introduire une deuxième coupure, appelée peau de Verlet r_v, qui est plus grande que la coupure standard à court terme r_{cu}. Cela permet de réduire le calcul de force uniquement pour les particules qui se trouvent à l'intérieur de r_v. Pour chacune des particules, une liste de Verlet contenant les voisins à l'intérieur de r_v est créée (**Figure II.26**). Cette liste est ensuite actualisée lorsque le déplacement maximal est supérieur à $|r_v-r_{cut}|$. La valeur optimale de r_v dépend de la densité du système.

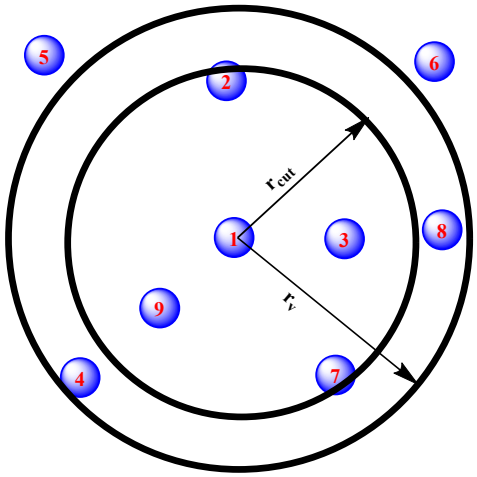


Figure II.26. Schéma relatif à la liste de voisins (Neighbor list) : Pendant ce temps, la particule 1 interagit avec les particules 2, 3, 7 et 9, tandis que les particules 4 et 8 apparaissent dans la liste des voisins de la particule 1, bien qu'elles n'interagissent pas avec elle.

Cell-List : Les listes de cellules (Cell Lists) servent à accélérer les simulations pour gagner du temps ; la boîte de simulation est divisée en cellules de taille équilatérale plus grande que r_{cut}, et pour le calcul des interactions de chaque particule avec les autres, seules les cellules contenant la particule et les cellules voisines doivent être prises en compte (**Figure II.27**). Cette approche élimine les calculs coûteux de coupure de distance et permet d'accélérer considérablement le processus.

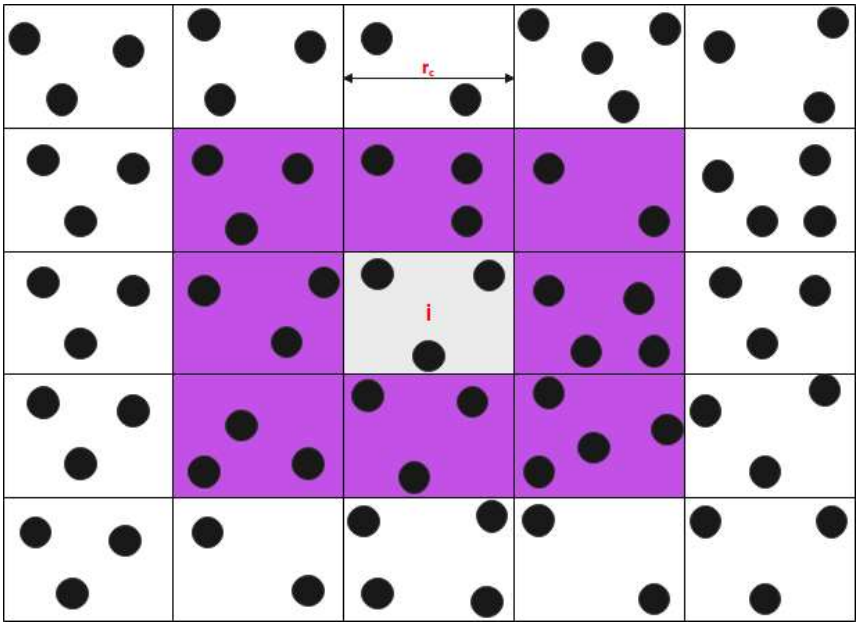


Figure II.27. Liste des cellules : Pour calculer la force d'une particule située dans la cellule i, seules les particules situées dans les cellules voisines (en violet) doivent être prises en considération.

Les contraintes : Les techniques de contrainte sont fréquemment utilisées pour réduire le temps de simulation de DM en considérant les molécules comme des unités rigides ou semi-rigides, avec des longueurs de liaison fixes (et parfois des angles de liaison). Ces techniques peuvent être classées en trois catégories en fonction des multiplicateurs de Lagrange, des systèmes de coordonnées internes distincts et des forces de contrainte explicites [377]. Les algorithmes de contrainte les plus populaires utilisés pour les simulations biomoléculaires sont les algorithmes de contrainte itératifs (tels que SHAKE, RATTLE et Linear Constraint Solver), ainsi que les algorithmes de contrainte non itératifs tels que les contraintes d'évolution gaussienne [376].

6.5.4 Pas de temps (Δt)

Dans la dynamique moléculaire, le "Pas de temps" correspond au temps consacré à la simulation. Des pas de temps plus courts permettront d'obtenir des résultats plus précis car des détails supplémentaires sont alors capturés. Toutefois, un pas de temps trop long peut engendrer une perte d'informations et des résultats inexacts. La notion de "pas de temps" désigne le temps qui sépare chaque mouvement entre les atomes. Du point de vue technique, lorsque Δt est très court (1-2 fs), on peut s'approcher de la solution exacte des équations différentielles, mais si Δt est très long, la simulation devient instable en raison de la forte augmentation de l'énergie du système [386].

6.5.5 Stratégie de la DM mise en œuvre adoptée dans ce travail

6.5.5.1 Choix de la méthode de DM appropriée

Généralement, la méthode appropriée peut être choisie pour simuler la dynamique moléculaire du système considéré en termes de temps selon le domaine d'application (**Figure II.28**). Sur cette base, on peut distinguer quatre méthodes de dynamique moléculaire, à savoir la dynamique moléculaire quantique, la dynamique moléculaire à tout-atome, la dynamique moléculaire à gros grains et la dynamique moléculaire à méso-échelle [373]. La **Figure II.28** indique les limites approximatives de ces catégories de DM et montre également comment ces gammes de résolutions peuvent être complétées par la combinaison de différentes méthodes de modélisation dans des schémas multi-échelles. La dynamique moléculaire quantique (quantum MD) est une technique informatique utilisée pour simuler la dynamique des systèmes au niveau de la mécanique quantique. La dynamique moléculaire à tout-atome (all-atom MD) simule les molécules au niveau atomique, tandis que la dynamique moléculaire à gros grains (coarse-grained MD) réduit le nombre d'atomes afin que les simulations puissent être effectuées sur des systèmes biomoléculaires de taille plus élevée et sur des échelles de temps plus longues. La dynamique moléculaire à méso-échelle (mesoscale MD) relie les simulations atomistiques et macroscopiques et permet de mieux comprendre la dynamique des grands systèmes biomoléculaires. Les trois types de simulations de DM sont utilisés pour étudier les propriétés et les processus moléculaires, les approches quantiques et tout-atome se concentrant principalement sur les petits systèmes, et la DM à gros grain et à méso-échelle étudiant les systèmes plus grands. La DM quantique et la DM tous atomes sont basées respectivement sur la mécanique classique et la mécanique quantique, tandis que les simulations à gros grain et à méso-échelle nécessitent l'utilisation de modèles simplifiés et de potentiels approximatifs pour capturer les interactions pertinentes dans le système. Chaque méthode a ses propres avantages et limites, et c'est en choisissant soigneusement la méthode la plus appropriée que l'on obtiendra les résultats les plus précis et les plus utiles [377].

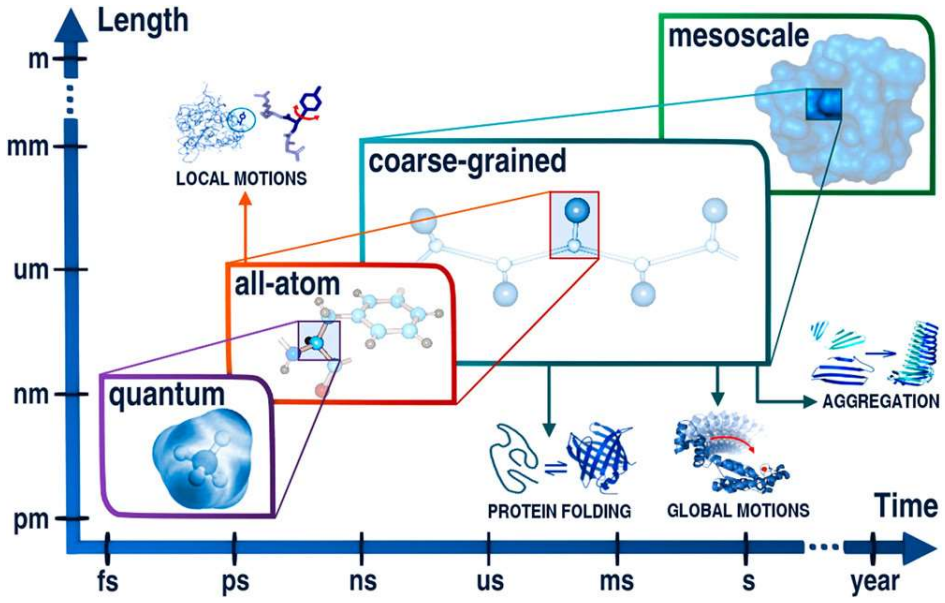


Figure II.28. Adaptation des différentes résolutions de la modélisation de la dynamique moléculaire aux échelles de temps et aux tailles (longueurs) des systèmes [373].

6.5.5.2 *Étapes de la dynamique moléculaire*

Du point de vue algorithmique [397], on peut résumer l'algorithme de base de la simulation de la dynamique moléculaire dans le schéma indiqué par la **Figure II.29**.

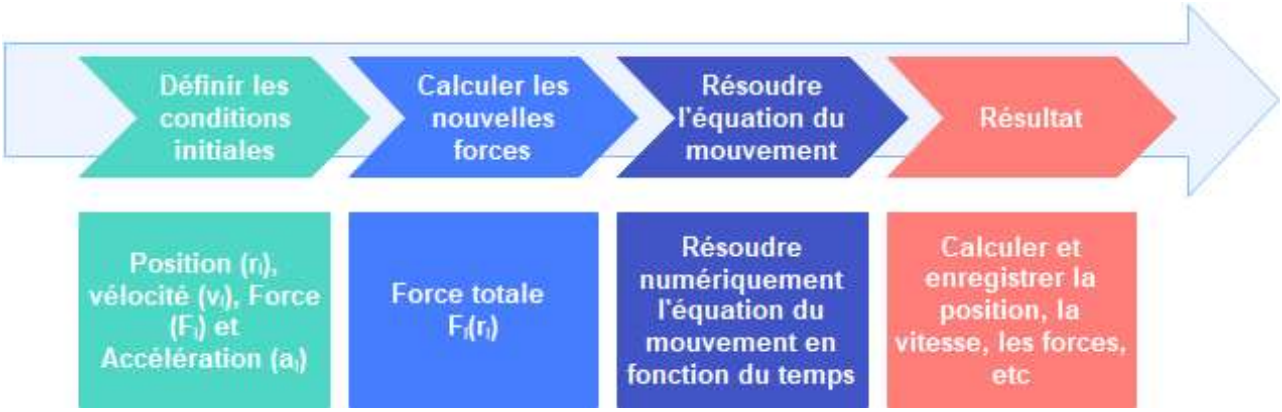


Figure II.29. Algorithme de base appliqué aux simulations de dynamique moléculaire.

Du point de vue des simulations, toute la simulation de la MD est basée sur trois phases principales comprenant respectivement la thermisation du système moléculaire, l`équilibrage et la production (**Figure II.30**).

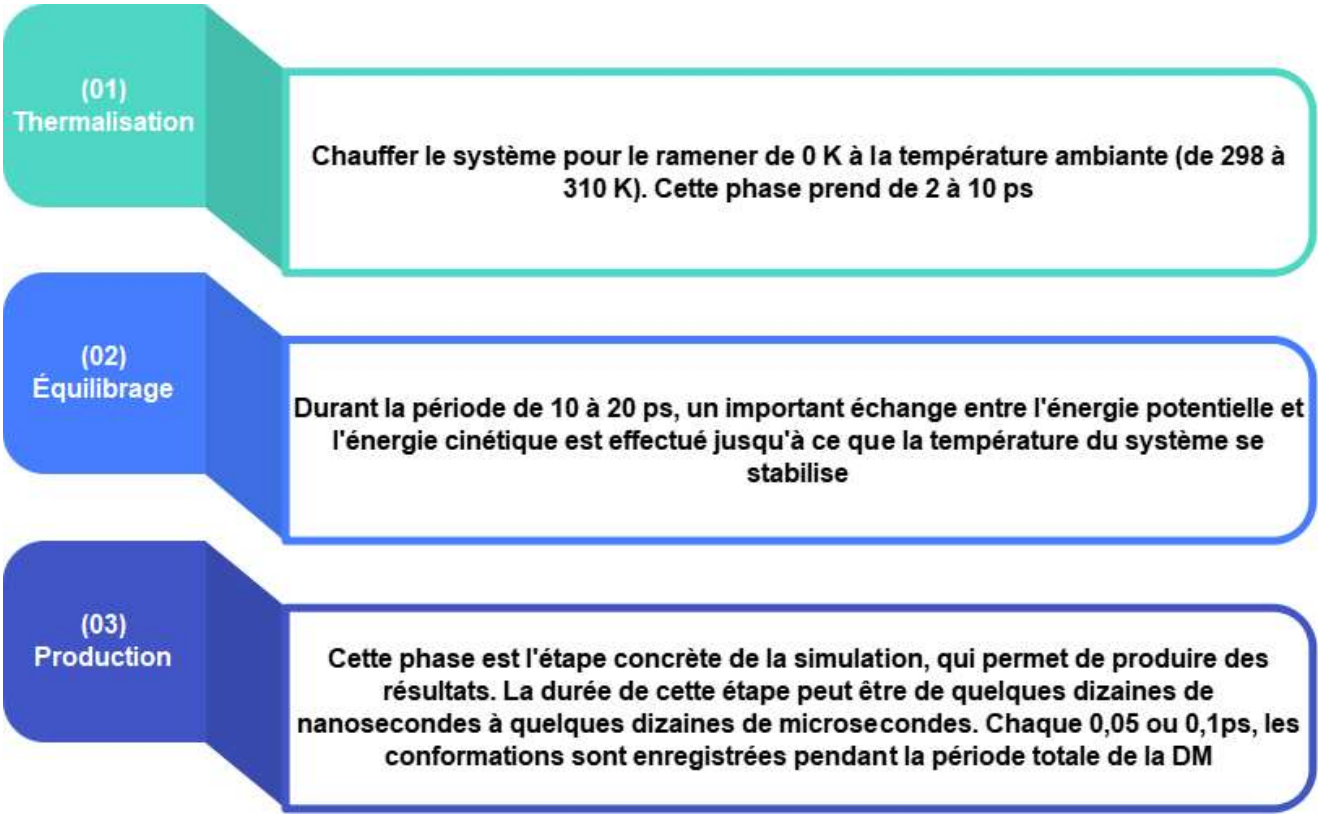


Figure II.30. Principales phases des simulations de la DM.

6.5.5.3 *Logiciel pour les simulations Biomoléculaires*

Grâce aux progrès de la technologie informatique en général et des outils de simulation biomoléculaire avancés en particulier, les unités de traitement graphique (GPU) ont été adoptées récemment pour effectuer des simulations de la DM [398], [399]. À l'origine, les unités de traitement graphique (GPU) ont été développées pour traiter les graphiques à haute résolution, tels que les jeux vidéo ; cependant, il existe de plus en plus de matériel et de logiciels GPU pour diverses applications

informatiques. Comme les simulations MD nécessitent des calculs coûteux, l'utilisation des GPU pour ces calculs a fait l'objet d'un grand intérêt. Le **Tableau II.13** indique certains des principaux packages développés pour mettre en œuvre des simulations MD qui prennent en compte l'exploitation des GPU. Dans le cadre de cette thèse, nous avons effectué des simulations de DM de systèmes biomoléculaires à l'aide des packages Schrödinger (Desmond) et Gromacs. Dans nos travaux, on a souvent mis en œuvre la dynamique moléculaire à la fin des procédures de docking pendant des périodes temporelles (100-200 ns), ce qui permet d'obtenir un échantillonnage réaliste de l'espace d'alignement conformationnel.

Tableau II.13. Répertoire (non exhaustif) des logiciels de dynamique moléculaire les plus connus.

Package	Référence
Desmond Schrödinger software https://www.schrodinger.com/products/desmond	[400]
Gromacs (GROningen Machine for Chemical Simulations) https://www.gromacs.org/	[401]
Namd (Nanoscale Molecular Dynamics) https://www.ks.uiuc.edu/Research/namd/	[402]
Amber (Assisted Model Building and Energy Refinement) https://ambermd.org/	[403]

6.5.5.4 *Étapes générales d'exécution des simulations MD*

Les **Figures II.31** et **II.32** présentent les étapes générales de mise en œuvre des simulations MD des systèmes (complexes) protéine-ligand par Desmond et GROMACS, respectivement systématisées dans nos travaux.



Figure II. 31. Étapes générales de l'exécution d'une simulation de DM par GROMACS [404].

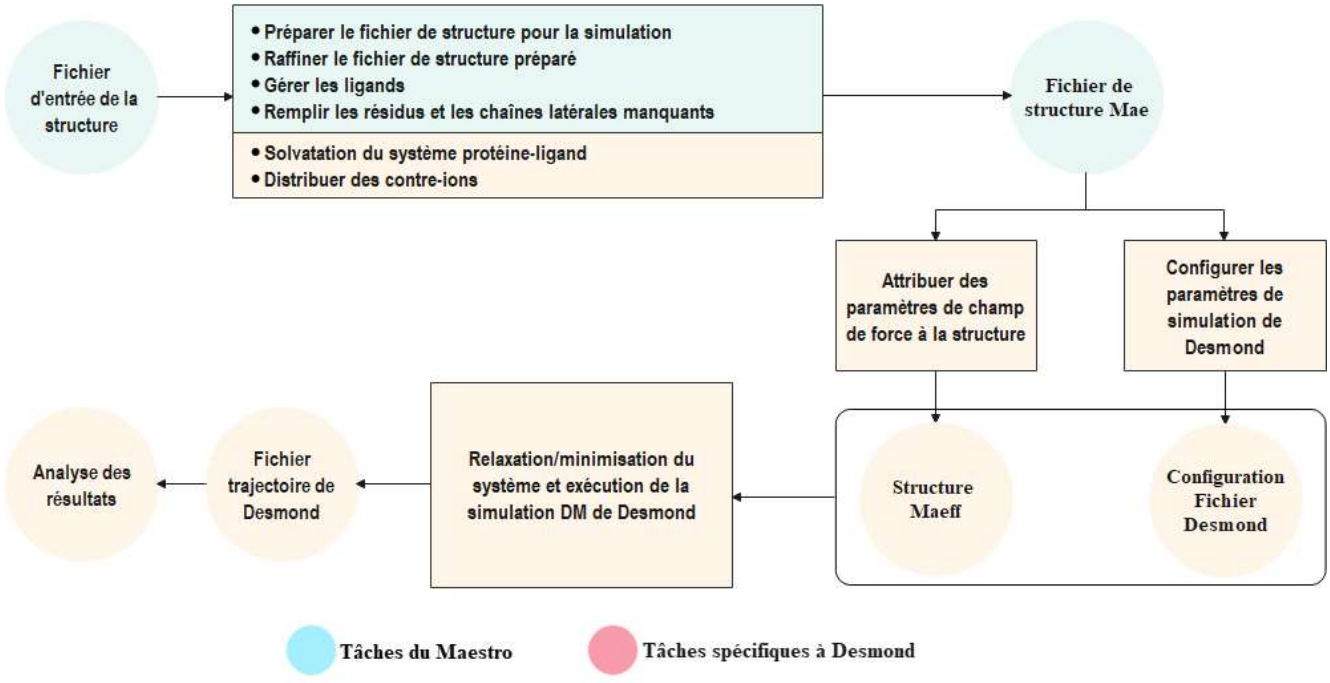


Figure II.32. Étapes générales de l'exécution d'une simulation de DM par Desmond.

6.5.6 Applications de la dynamique moléculaire

Les méthodes de simulation de MD, associées à la disponibilité de centres de calcul à haute performance (HPC), bénéficient aujourd'hui de multiples applications, notamment dans le cadre de l'étude et de la simulation des structures de systèmes protéiques géants (plus de 500 000 atomes) [405]. Les simulations de dynamique moléculaire constituent une méthode puissante sur laquelle s'appuyer pour informer la conception de nouveaux médicaments, en fournissant des informations précises sur la liaison entre le médicament et la cible, sur le comportement de la cible suite à l'administration du médicament, et sur la capacité du médicament à remplir ses fonctions [406]. On utilise la dynamique moléculaire pour étudier la stabilité du complexe récepteur-ligand obtenu par la stratégie du Docking moléculaire ainsi que pour l'optimisation *in silico* des ligands [347]. De plus, l'approche « relaxed-complex scheme » utilise la dynamique atomistique sur le récepteur pour décrire sa flexibilité et ses différentes conformations, dans lesquelles on introduit les ligands [406]. Ces simulations peuvent nous aider à identifier les conformationnels les plus stables et à cerner les interactions à l'origine de la liaison.

7 Conclusion

Les techniques de modélisation biomoléculaire basées notamment (CADD) revêtent une grande importance dans le cadre de la découverte et du développement de nouveaux médicaments, car elles permettent de comprendre la structure 3D des systèmes macromoléculaire de type protéine-ligand cibles, ce qui n'est pas possible avec les méthodes expérimentales traditionnelles. Cela permet de prédire les interactions moléculaires, de concevoir des médicaments plus puissants et plus sélectifs, d'identifier de nouvelles indications et d'exploiter de nouveaux espaces chimiques. En outre, les techniques de modélisation basées sur les méthodes CADD peuvent être utilisées pour optimiser les candidats têtes de série dans le processus d'optimisation de ces dernières, ce qui permet de réduire considérablement le temps et les investissements consacrés à la mise au point de nouveaux médicaments. L'absence d'une stratégie unifiée pour l'application des différentes techniques de CADD est un problème majeur pour la recherche appliquée dans le domaine pharmaceutique. Il existe une multitude d'approches pour modéliser et prédire les effets biologiques des molécules, mais aucune approche globale n'est universellement acceptée. Cette situation se traduit en fin de compte par des processus longs et complexes de développement et d'essai de substances médicamenteuses, qui sont coûteux et conduisent souvent à la mise au point de composés inefficaces. Le manque de stratégie unifiée bloque en outre la traduction des connaissances scientifiques acquises à partir de ces modèles en applications concrètes pour la découverte de médicaments. Par conséquent, la demande d'une stratégie intégrée de CADD, adaptable à différentes applications, devient urgente. Grâce à la mise en place d'une stratégie hybride de CADD, il est possible de rationaliser plus facilement la conception de nouvelles molécules et leur évaluation *in silico*. Cela pourrait permettre un développement plus rapide des nouveaux médicaments et une compréhension plus approfondie de leur action, tout en réduisant les coûts globaux. C'est ce que nous avons visé grâce à l'intégration des approches CADD, dans ses deux volets, LBDD et SBDD, dans le criblage et la sélection de petites molécules bioactives et la

rationalisation de la conception de nouveaux médicaments. À cet égard, au niveau de cette thèse, nous appliquons cette approche computationnelle hybride à l'étude de différentes données liées aux systèmes moléculaires à base de petites molécules de produits synthétiques ou naturels candidates à la conception de nouveaux médicaments contre plusieurs types du cancer (comme le cancer du sein, du poumon, des lymphomes, etc.) et des virus (comme le SARS-CoV-2). Dans les chapitres restants, nous présentons la synthèse des résultats et des discussions les plus récents que nous avons acquis et qui ont été publiés sous la forme d'articles de recherche pendant la préparation de cette thèse.

CHAPITRE III : MODÉLISATION D'INHIBITEURS NON-COVALENTS BASEE SUR DES MOLÉCULES BIOACTIVES DE PRODUITS SYNTHÉTIQUES POUR L'INHIBITION DE LA PROLIFÉRATION CELLULAIRE CANCÉREUSE : CIBLAGE MOLÉCULAIRE DU RÉCEPTEUR TYROSINE KINASE DE c-Met

Introduction

Dans cette étude, nous visons à développer de nouvelles générations d'inhibiteurs de tyrosine kinase (ITK) basés sur le ciblage moléculaire de la protéine c-Met par de petites molécules synthétiques, adapté aux besoins de la thérapie ciblée contre le cance. Dans ce contexte, une récente étude de synthèse et d'évaluation biologique anticancer réalisée *in vitro* par Mohareb et al [407]. a révélé qu'une série de 48 petites molécules biologiquement active des dérivés de 4,5,6,7-Tétrahydrobenzo[D]-Thiazol-2-Yl synthétisées à base de la Dimédone présentait une forte activité biologique inhibitrice vis-à-vis de la protéine c-Met comparativement au médicament standard Crizotineb. Cependant, la partie relative aux propriétés de type médicament, à la pharmacocinétique et à la pharmacodynamique de ces molécules, ainsi que les mécanismes d'interactions potentielles entre ces molécules et le récepteur c-Met, n'ont pas été abordés. Cela nous a incités à mener une étude *in silico* à grande échelle sur ces molécules afin de sélectionner des molécules leaders potentielles à utiliser pour la conception médicamenteuse dans le cadre d'une thérapie ciblée contre le cancer. Dans le même esprit, une autre récente étude menée *in vitro* par JuLiu et al [408]. a montré que 26 petites molécules dérivées de la 4-Phénoxypyridine ont une activité biologique inhibitrice intéressante vis-à-vis de c-Met. Cependant, ces molécules n'ont pas atteint l'activité biologique inhibitrice démontrée par le médicament standard d'Foretinib, lequel a été utilisé comme contrôle dans l'évaluation biologique réalisée *in vitro*. De plus, au cours de cette étude, la pharmacocinétique, la pharmacodynamique, les mécanismes combinatoires et les interactions protéine-médicament de ces molécules n'ont pas été estimés et approfondis. Toutes ces raisons nous ont incités à mener davantage d'études computationnelles *in silico* pour améliorer l'activité biologique et les propriétés structurelles de ces petites molécules afin qu'elles deviennent de nouveaux agents potentiels pour l'utilisation thérapeutique contre le cancer à travers l'inhibition de l'activité enzymatique de la protéine c-Met. Parallèlement à ces deux études, nous avons mené une autre étude parallèle pour étayer cette idée et accorder une place particulière au cancer du poumon non à petites cellules en tant que modèle pour le traitement par des inhibiteurs de c-Met basés sur de petites molécules biologiquement actives. À cette fin, nous avons criblé *in silico* une base de données moléculaires contenant 40 petites molécules hétérocycliques portant l'échafaudage Cyclohexane-1,3-dione ayant démontré leurs activités biologiques inhibitrices contre les lignées cellulaires de cancer du poumon non à petites cellules *in vitro* [409]. Cette étude met en évidence l'importance de ces molécules comme stratégie thérapeutique prometteuse contre le cancer. Même si l'activité biologique de ces molécules a été observée *in vitro* contre plusieurs lignées

cancéreuses, dont le CPNPC et le cancer du sein, cela n'est pas suffisant pour confirmer le placement de ces molécules dans le pipeline de conception et la découverte de nouveaux médicaments. Par conséquent, le criblage *in silico* de ces molécules est hautement nécessaire, afin d'optimiser la structure moléculaire et l'activité biologique de ces molécules pour qu'elles deviennent des agents plus candidats pour leur utilisation thérapeutique contre le cancer par le ciblage des récepteurs tyrosine kinases. C'est ce que nous avons visé à atteindre dans cette partie de ce travail, en accumulant notre attention sur la modélisation de ces molécules pour qu'elles deviennent plus compatibles avec le récepteur c-Met, qui est impliqués dans de nombreux cancers, en plus du cancer du poumon non à petites cellules qui est ciblé comme modèle dans cette partie du travail. La **Figure III.1** présente un schéma des échafaudages de petites molécules bioactives que nous avons adoptés dans nos études comme structures de base pour le développement de nouvelles classes d'inhibiteurs de c-Met pour une application dans la thérapie ciblée contre le cancer. À cet égard, nous utilisons les deux médicaments Crizotinib et Foretinib comme références pour la quantification, la comparaison et la formulation de conclusions lors de la conception rationnelle de nouveaux médicaments potentiels pour le traitement du cancer par le biais de l'inhibition de l'activité enzymatique de la protéine c-Met.

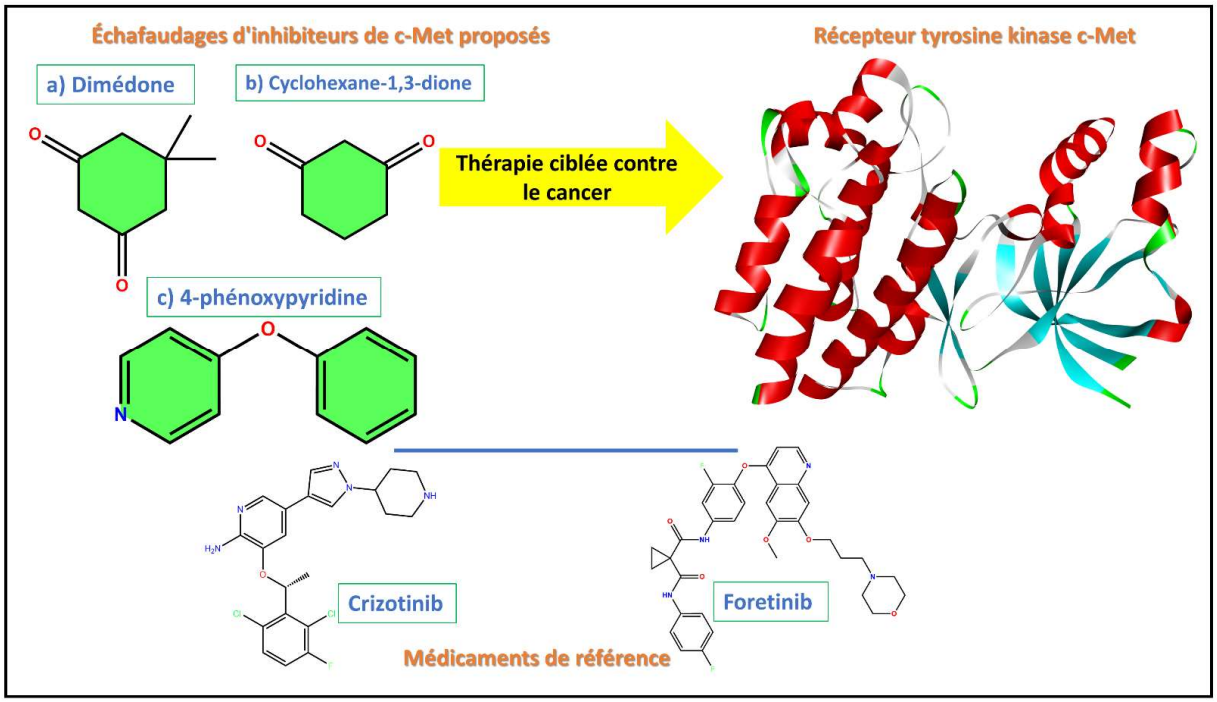


Figure III.1. Dimédone (a) Cyclohexane-1,3-dione (b), 4-Phénoxypyridine (c): échafaudages candidats pour la conception de nouveaux agents inhibiteurs de la tyrosine kinase c-Met pour la thérapie ciblée contre le cancer à base des petites molécules biologiquement actives synthétisées.

Dans la suite de ce chapitre, nous présentons les principaux résultats et conclusions auxquels nous sommes parvenus grâce aux trois études que nous avons menées à cet égard. Dans les sections A, B et C, nous présentons successivement le contenu de chaque étude.

A. Identification *in silico* de Nouvelles Molécules Bioactives Dérivées de Dimédone, Basées sur la Structure de 4,5,6,7-Tétrahydrobenzo[D]-Thiazol-2-Yl, en tant qu'Inhibiteurs Potentiels du Récepteur Tyrosine Kinase c-Met pour la Thérapie Ciblée contre le Cancer

1 Introduction

Récemment, les 1,3-thiazoles hétérocycliques aromatiques ont fait l'objet d'un grand intérêt dans le domaine de la chimie médicinale [410], grâce à leurs propriétés multiples et impressionnantes : anti-inflammatoires [411], anticonvulsifs [412], insecticides [413], antioxydants, antitumoraux, antihypertenseurs, pesticides [414], et activité antidiabétique [415], pour ne citer que ceux-là. Les formulations de 1,3-thiazole S3U937 (**Figure III.A1a**) [416] et S8A375 [417] (**Figure III.A1b**), mentionnées en particulier, ont démontré des effets préventifs et thérapeutiques contre différents types de cancer [418] (**Figure III.A1**).

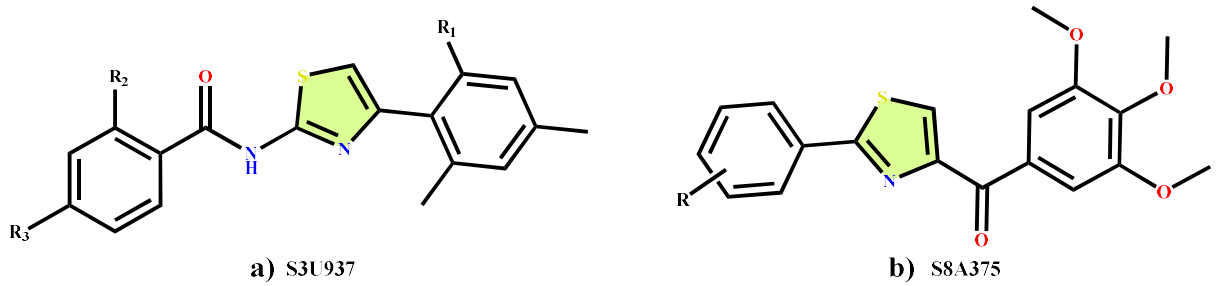


Figure III.A1. Structures de Zopolrestat (a) et de N-(6-substitué-1,3-benzothiazol-2-yl)-Benzenesulfonamide (b).

Le récepteur tyrosine kinase c-Met présente une grande importance en tant que cible thérapeutique anticancéreuse [419]. Une nouvelle série de molécules appelée 2-amino-5-aryl-3-benzyloxypyridines a fourni l'inhibiteur de c-Met connu sous le nom de crizotinib (PF-02341066) (**Figure III.A2**) [420], qui a démontré une forte activité inhibitrice de la kinase tyrosine c-Met in vitro et *in vivo*, et qui présente également des propriétés pharmaceutiques très convaincantes pour le traitement du cancer [421]. Le Crizotinib est capable de bloquer la kinase c-Met et déstabiliser sa voie de signalisation [40]. Cet inhibiteur peut donc être utilisé comme référence dans le développement de nouveaux agents ciblant la kinase c-Met pour le traitement des tumeurs malignes.

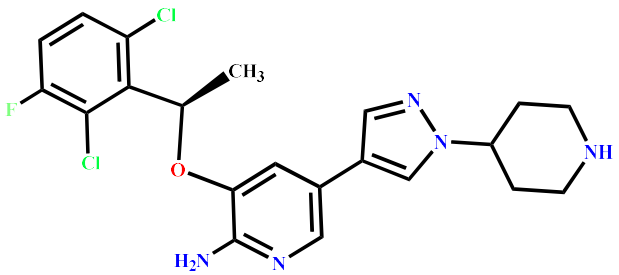


Figure III.A2. Structure moléculaire de l'inhibiteur Crizotinib (PF-02341066) avec une IC_{50} de 8 nM sur c-Met [421].

À ce jour, le Crizotinib est largement utilisé en tant qu'inhibiteur de c-Met. Cependant, le taux de réponse au Crizotinib est faible (30-40%), certains patients développent une résistance et le Crizotinib peut entraîner de graves effets secondaires [420], [422] . Il est donc urgent de concevoir de nouveaux

inhibiteurs de c-Met qui permettent d'obtenir un taux de réponse plus élevé, d'échapper à la résistance et d'avoir moins d'effets secondaires. Ces défauts soulignent la nécessité de concevoir de nouveaux inhibiteurs de c-Met plus efficaces, présentant moins d'effets secondaires et capables d'échapper à la résistance. Dans ce contexte, une récente étude menée par Mohareb et al. [407] a révélé *in vitro* la forte activité biologique inhibitrice de nouvelles petites molécules organiques dérivées du 4,5,6,7-tétrahydrobenzo-[D]-Thiazol-2-Y portant l'échafaudage Dimédone contre l'activité enzymatique du c-Met. La série de 48 molécules optimisées *in vitro* a montré des activités biologiques en termes de (pIC_{50}) meilleures que le Crizotinib, qui a été adopté comme médicament standard lors d'un contrôle positif (**Tableau III.A1**). Pour suivre les conclusions de cette étude et progresser vers le développement de nouveaux inhibiteurs de c-Met pour une thérapie ciblée contre le cancer, il faut suivre des processus de synthèse très longs, des tests de criblage réels complexes et coûteux. Par conséquent, l'utilisation des méthodes de modélisation moléculaire en général dans la recherche liée à l'industrie pharmaceutique et des approches CADD en particulier reste une solution auxiliaire à ces difficultés. À cet égard, étant donné la disponibilité du critère structure-activité de 48 dérivés de 4,5,6,7-tétrahydrobenzo[D]-Thiazol-2-Y et disponibilité de la structure 3D de la protéine c-Met cristallisée avec le médicament Crizotinib (code PDB : 2WGJ) [421], il serait très bénéfique d'appliquer l'approche CADD hybride. En rationalisant la recherche dans les données moléculaires disponibles (chimiothèque) en vue du choix de la meilleure structure moléculaire pour la conception d'un nouveau médicament susceptible de lutter contre le cancer. Dans cette partie du travail, à l'aide de la modélisation QSAR basée sur des descripteurs moléculaires hybrides (0D-3D), de la modélisation "Drug-like" et de l'ADME-Tox, des analyses de Docking moléculaire, nous criblons la série de molécules ciblées pour identifier les inhibiteurs de c-Met candidats à la thérapie ciblée contre le cancer.

2 Stratégie de conception de médicaments assistée par ordinateur développée et appliquée

Le schéma présenté à la **Figure III.A3** expose un aperçu synthétique du modèle CADD hybride que nous avons appliqué *in silico* dans cette étude pour identifier les inhibiteurs potentiels de c-Met destinés à la thérapie anticancéreuse sur la base molécules bioactive des dérivés de 4,5,6,7-Tétrahydrobenzo[D]-Thiazol-2-Yl, tout en prenant l'inhibiteur Crizotinib comme base de comparaison. Tous les détails méthodologiques et techniques relatifs aux différentes techniques utilisées dans cette stratégie et à leur implémentation sont détaillés dans le deuxième chapitre de cette thèse ainsi que dans la version électronique de l'article que nous avons publié sur ce sujet [294].

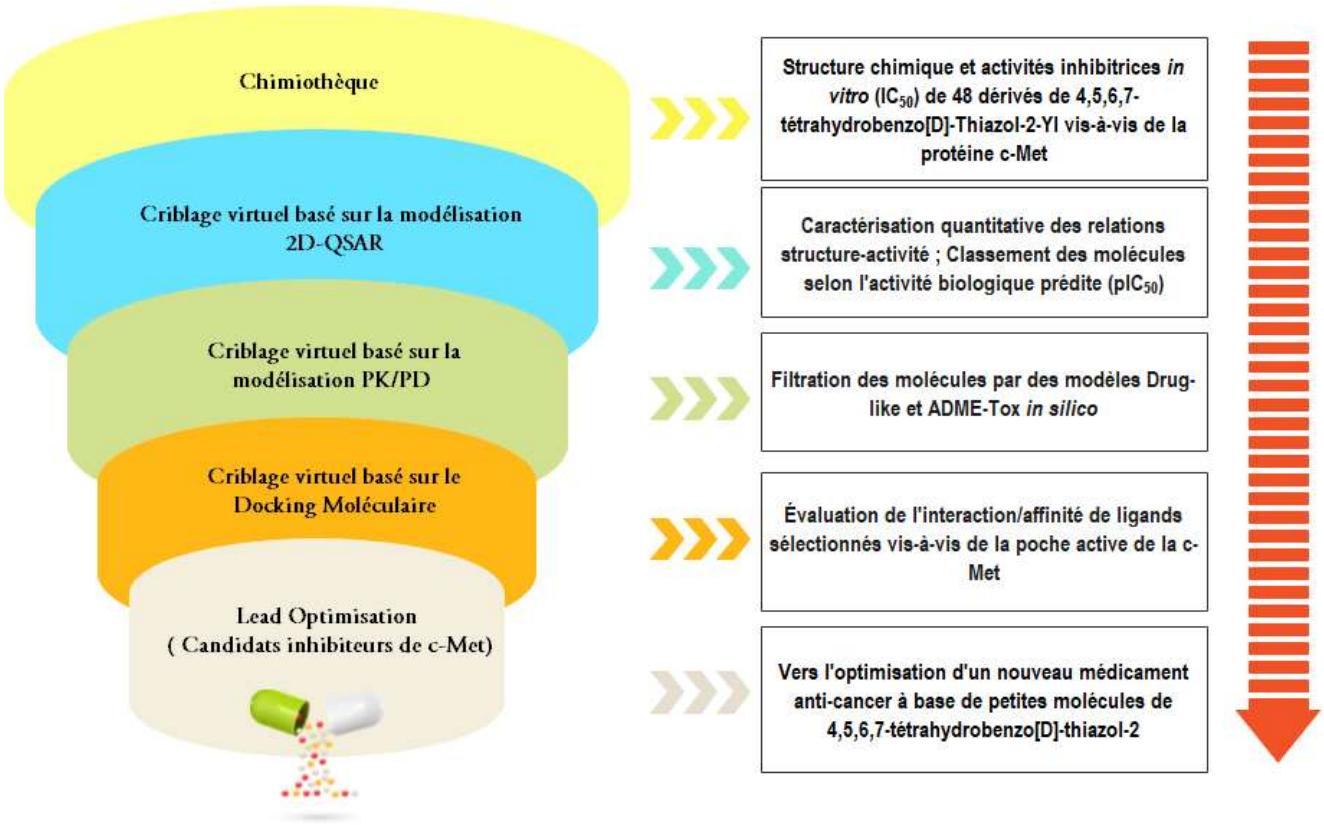
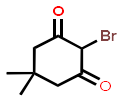
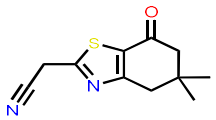
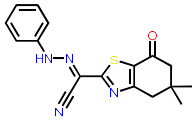
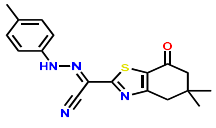
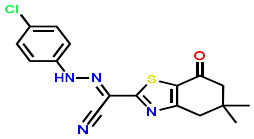
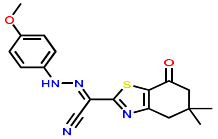
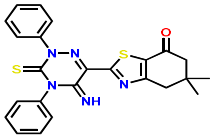
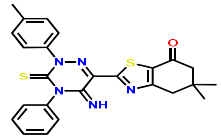
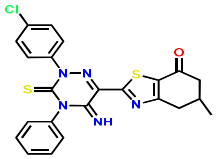
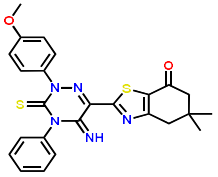
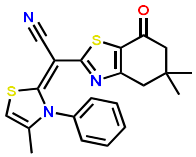
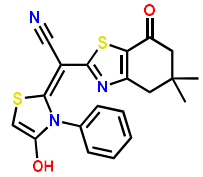
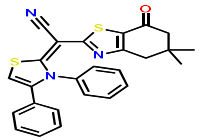
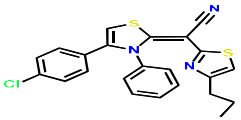
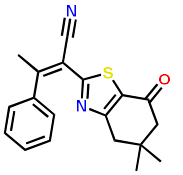
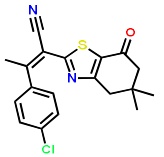
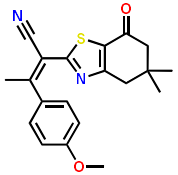
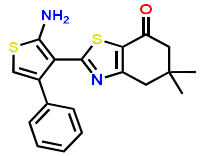
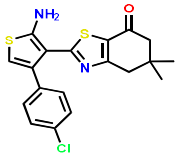
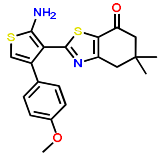
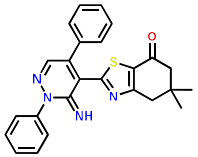
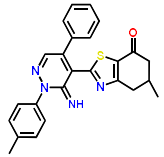
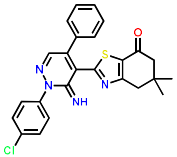
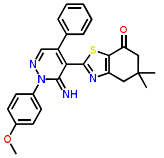
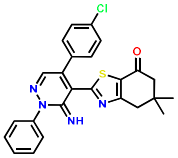
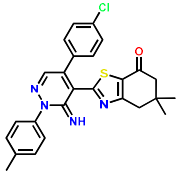
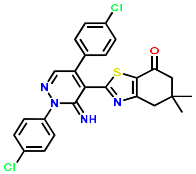
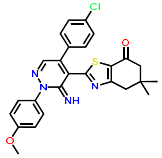
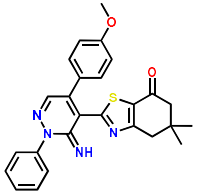
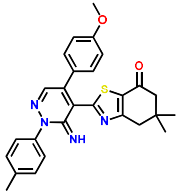
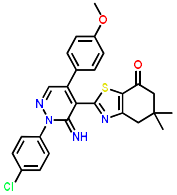
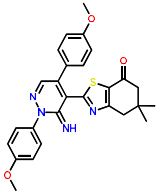
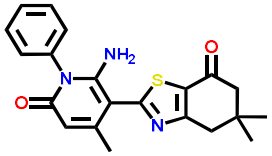
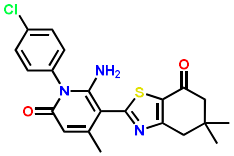
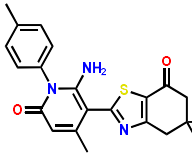
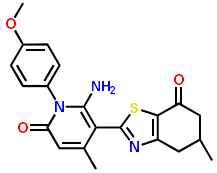
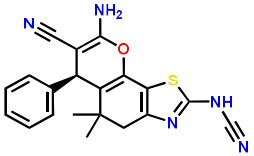
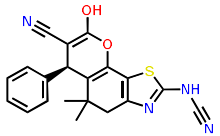
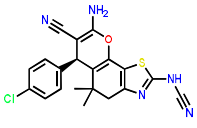
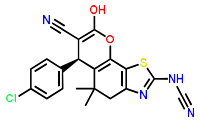


Figure III.A3. Stratégie de CADD proposée et appliquée *in silico* pour identifier de nouveaux inhibiteurs non covalents de la protéine c-Met à base de 4,5,6,7-tétrahydrobenzo [D]-Thiazol-2-Y en vue d'une thérapie ciblée contre le cancer.

Tableau III.A1. Données de structure-activité disponibles pour les dérivés de 4,5,6,7-Tétrahydrobenzo[D]-Thiazol-2-Yl issus de Dimédone.

							
01 pIC_{50}=8.87	02 pIC_{50}=4.64	03 pIC_{50}=8.28	04 pIC_{50}=8.28	05 pIC_{50}=9.49	06 pIC_{50}=8.90	07 pIC_{50}=8.33	08 pIC_{50}=8.11
							
09 pIC_{50}=10.10	10 pIC_{50}=8.90	11 pIC_{50}=8.44	12 pIC_{50}=9.38	13 pIC_{50}=8.88	14 pIC_{50}=9.64	15 pIC_{50}=8.20	16 pIC_{50}=9.59
							
17 pIC_{50}=8.62	18 pIC_{50}=8.08	19 pIC_{50}=9.32	20 pIC_{50}=8.62	21 pIC_{50}=8.28	22 pIC_{50}=8.46	23 pIC_{50}=8.34	24 pIC_{50}=8.49
							
25 pIC_{50}=8.94	26 pIC_{50}=8.46	27 pIC_{50}=9.55	28 pIC_{50}=8.86	29 pIC_{50}=8.37	30 pIC_{50}=8.08	31 pIC_{50}=8.88	32 pIC_{50}=8.42
							
33 pIC_{50}=8.42	34 pIC_{50}=8.97	35 pIC_{50}=8.25	36 pIC_{50}=8.73	37 pIC_{50}=8.14	38 pIC_{50}=8.55	39 pIC_{50}=8.99	40 pIC_{50}=9.55
41 pIC_{50}=8.47	42 pIC_{50}=8.85	43 pIC_{50}=8.20	44 pIC_{50}=9.09	45 pIC_{50}=9.24	46 pIC_{50}=9.49	47 pIC_{50}=8.19	48 pIC_{50}=8.68

3 Résultats et discussion

Dans cette section, nous présentons et discutons les principaux résultats générés dans le cadre de cette étude. Les détails associés à ces résultats peuvent être consultés dans le papier original de notre publication (Daoui et al. [294]). Après avoir optimisé les structures moléculaires de la chimiothèque examinée et généré des descripteurs moléculaires hybrides 0D-3D (constitutionnels, topologiques, physico-chimiques et électroniques de réactivité moléculaire issus de la DFT conceptuelle), les nouvelles données ont été soumises à des analyses de type ACP. Grâce à la matrice de corrélation obtenue, nous avons constaté que les sept descripteurs moléculaires EVDW (Énergie de Van der Waals), LogP (lipophilie), αe (Polarisabilité), Pc (Parachor), MW (Poids moléculaire), MV (Volume moléculaire), et η (Dureté chimique) sont les moins corrélés entre eux. Par conséquent, ces descripteurs moléculaires furent définis comme données d'entrée pour le développement de modèles 2D-QSAR linéaires et non linéaires à l'aide des techniques de RLM, de RNLM et de RNA. Ensuite, en appliquant la méthode de classification K-means, les molécules (01, 03, 04,05, 07, 09, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 40, 41, 43, 44, 45, 46, 47, et 48) sont sélectionnées comme éléments pour développer et entraîner les modèles 2D-QSAR, tandis que les molécules (02, 06, 08, 14, 22, 23, 24, 26, 38 et 42) sont sélectionnées comme éléments de l'ensemble de test.

3.1 Analyses Structure-Activité par 2D-QSAR

3.1.1 Modélisation 2D-QSAR linéaire et non linéaire basée sur des descripteurs moléculaires 0D-3D

Grâce à la technique d'apprentissage automatique de la RLM, nous avons identifié la meilleure relation linéaire reliant l'activité biologique pIC_{50} observée *in vitro* des dérivés de 4,5,6,7-tétrahydrobenzo[D]-Thiazol-2-Y à leur structure moléculaire en espace 2D, indiquée par les sept descripteurs moléculaires, comme mentionné dans **l'Eq. (III.A1)**.

$$\begin{aligned}(pIC_{50})_{RLM} &= 2.349 - 1.049 \times EVDW + 0.233 \times LogP \\ &-0.194 \times \alpha e + 5.4 \times 10^{-3} \times Pc + 2.56 \times 10^{-2} \times MW \\ &- 8.8 \times 10^{-3} \times MV + 0.954 \times \eta\end{aligned} \qquad \text{(III.A1)}$$

Grâce aux coefficients normalisés des descripteurs moléculaires indiqués dans l'**Eq.** (**III.A1)**, nous pouvons noter que l'activité biologique pIC_{50} des molécules examinées augmente avec l'augmentation de LogP, Pc, MW et η, tandis que le niveau d'activité biologique est influencé défavorablement par l'augmentation de EVDW, αe, et MV. Avant de confirmer cette hypothèse, le modèle linéaire RLM-QSAR proposé dans cette étude doit d'abord être validé. À cette fin, un ensemble de critères statistiques ont été adoptés pour évaluer la validité du modèle proposé. À cet égard, le modèle proposé a fait l'objet du test de validation interne et externe, validation croisée, du test de Y-randomisation, et le domaine d'applicabilité. Ensuite, on a élaboré des modèles 2D-QSAR non linéaires via les techniques de la RNLM (l'**Eq III.A2**) et de RNA afin de confirmer la validité de la sélection des sept descripteurs moléculaires comme paramètres prédictifs de l'activité biologique des dérivés de 4,5,6,7-tétrahydrobenzo[D]-Thiazol-2-Y.

$$(pIC_{50})_{MNLR} = 2.295 - 1.866 \times EVDW + 0.688 \times EVDW^2 + 0.393 \times LogP - 1.61 \times 10^{-2} \times LogP^2 - 0.124 \times \alpha e - 1.5 \times 10^{-3} \times \alpha e^2 + 5.7 \times 10^{-3} \times Pc - 2.729 \times 10^{-7} \times Pc^2 - 1.66 \times 10^{-2} \times MW + 1.63 \times 10^{-5} \times MW^2 + 6.5 \times 10^{-3} \times MV - 3 \times 10^{-5} \times MV^2 + 0.345 \times \eta + 0.064 \times \eta^2 \quad \text{(III.A2)}$$

Le **Tableau III.A2** présente une partie des paramètres statistiques les plus importants sur lesquels nous avons porté notre jugement lors des phases de sélection et de validation des modèles 2D-QSAR proposés dans le cadre de cette étude. Grâce aux données présentées dans le **Tableau III.A2**, les valeurs des coefficients de détermination (R^2, Q^2_{cv} et R^2_{test}) supérieures au seuil de 0.6, les faibles valeurs de MSE, ainsi que l'indice 1<ρ<3 du modèle RNA-QSAR indiquent la signification statistique et la fiabilité prédictive considérables des modèles RLM-QSAR, RNLM-QSAR, et RNA-QSAR proposés dans le cadre de cette étude.

Tableau III.A2. Critères statistiques pour la validation des modèles 2D-QSAR.

Critères de validation	Critères statistiques	Score du modèle			Seuil	Validation
		RLM	RNLM	RNA		
Interne	R^2	0.81	0.82	0.84	> 0.6	Oui
	Q^2_{cvv}	0.62	0.64	0.78	>0.5	Oui
	ρ	-	-	1.027	1< ρ <3	Oui
	MSE	0.060	0.070	0.062	Faible valeur	Oui
Externe	R^2_{test}	0.67	0.69	0.64	> 0.6	Oui
	MSE	0.07	0.13	0.11	Faible valeur	Oui

La forte corrélation entre l'activité biologique pIC_{50} et les propriétés constitutionnelles (MW), physico-chimiques (LogP, αe, Pc), géométriques (EVDW, MV) et quantiques (η) peut être confirmée par les distributions logiques de l'activité pIC_{50} observées *in vitro* vs. pIC_{50} prédites par les modèles 2D-QSAR (**Figure III.A4**).

La **Figure III.A4** révèle la distribution régulière des valeurs pIC_{50} observées *in vitro* et prédites *in silico*. Cela signifie que les prédictions des modèles 2D-QSAR développés dans notre étude sont en accord avec les évaluations obtenues via le modèle expérimental *in vitro*. Ceci confirme la faisabilité de l'exploitation de ces modèles mathématiques pour rationaliser la conception/optimisation de nouvelles molécules et pour prédire leurs activités biologiques avant la synthèse. De plus, ces modèles peuvent être exploités pour filtrer les meilleures structures moléculaires candidates pour les placer dans le pipeline de la conception rationnelle d'un nouveau médicament. Pour ce faire, on compare les activités biologiques prédites des molécules à celles observées pour la molécule médicamenteuse choisie comme référence (ici, l'inhibiteur Crizotinib). Toutefois, avant cela, d'autres vérifications doivent être engagées pour confirmer la stabilité de la relation structure-activité dégagée par le modèle 2D-QSAR original (ici, le modèle RLM-QSAR). À cette fin, dans cette étude, nous avons soumis le modèle RLM-QSAR à un test de Y-randomisation pour rejeter ou confirmer l'hypothèse de corrélation aléatoire entre la structure et l'activité biologique. À cet égard, après cinquante reconstructions aléatoires des modèles RLM-QSAR, il s'est avéré que tous les coefficients de détermination R^2 (valeur moyenne de 0.2) et Q^2_{cv} (valeur moyenne de -0.49) pour les modèles randomisés étaient inférieurs à ceux obtenus

dans le modèle original. En outre, le paramètre ${}^{c}R_{p}^{2}$ (0.7 > 0.5) indique que la relation structure-activité décodée par le modèle RLM-QSAR est stable et qu'il est très improbable de la trouver de manière arbitraire.

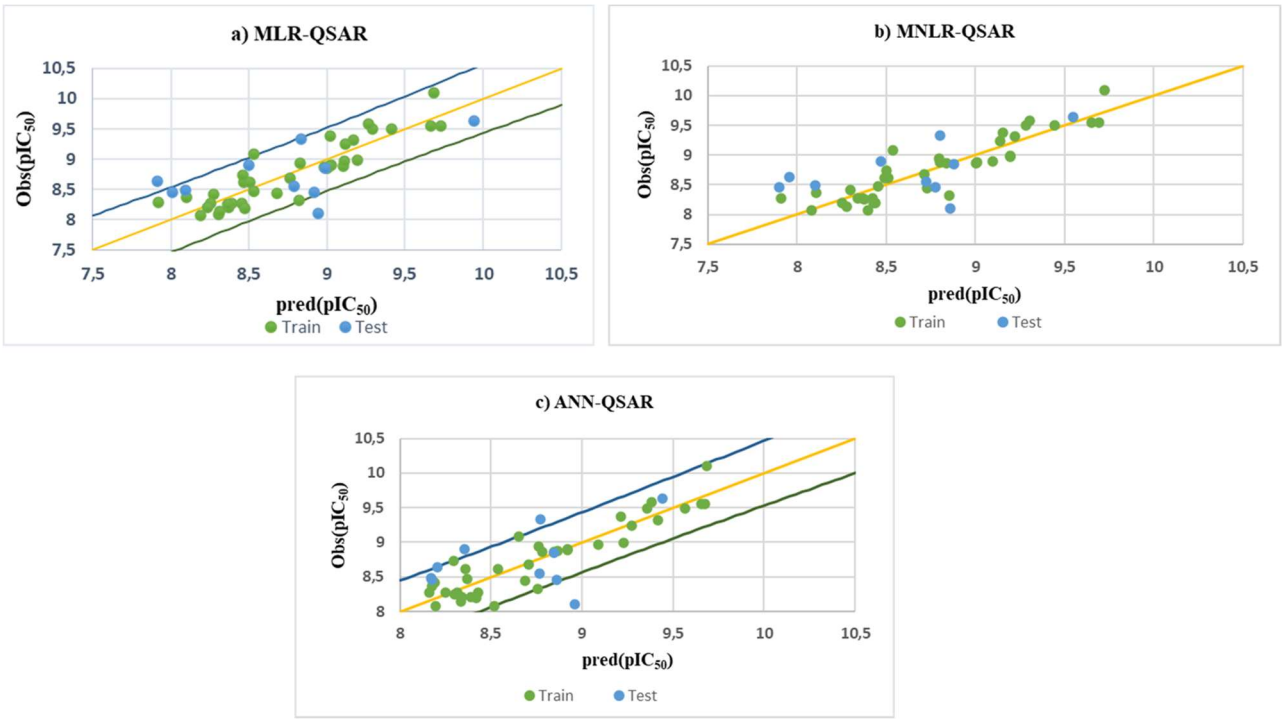


Figure III.A4. Les pIC$_{50}$ observés *in vitro* versus les pIC$_{50}$ prédits *in silico* par les modèles RLM-QSAR (a), RNLM-QSAR (b), ANN-QSAR (c)

Après avoir constaté que le modèle 2D-QSAR construit dans cette étude selon les principes de l'OCDE est valide, il reste à définir le domaine d'applicabilité de ce modèle. Ceci afin d'identifier les données moléculaires inappropriées qui pourraient affecter les prédictions du modèle développé. Sur cette base, nous avons défini le domaine d'applicabilité du modèle RLM-QSAR original, comme le montre le tracé de Williams présenté dans la **Figure III.A5**.

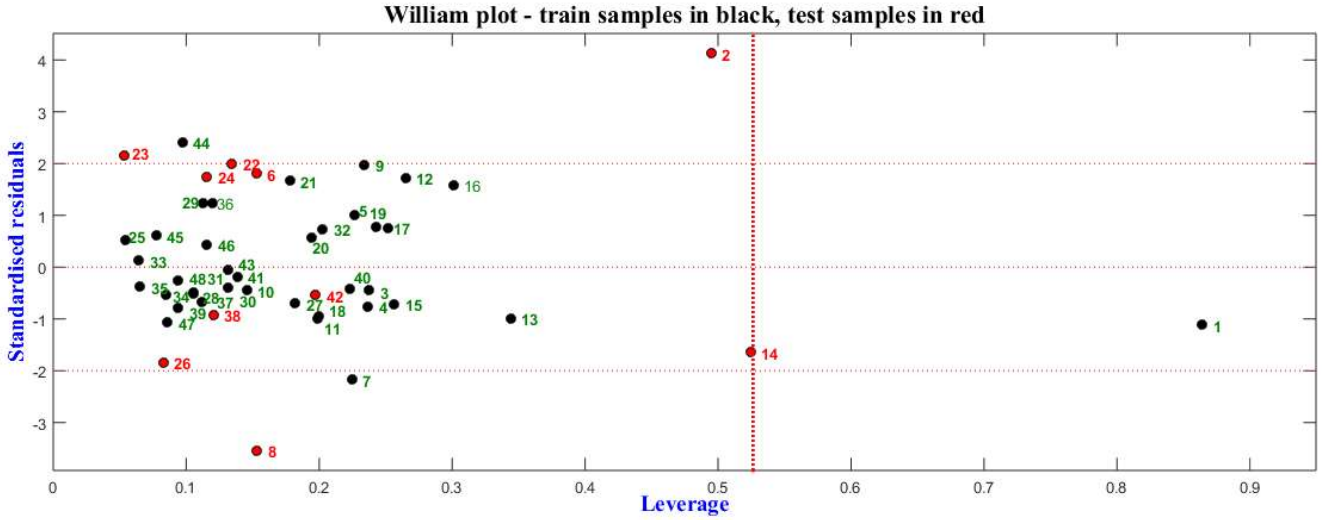


Figure III.A5. Tracé de William des résidus normalisés en fonction de l'effet de levier pour le modèle RLM-QSAR original.

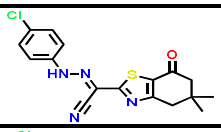
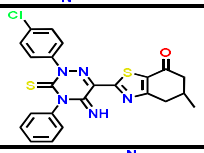
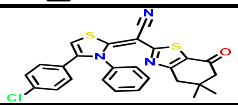
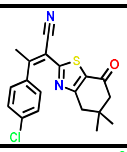
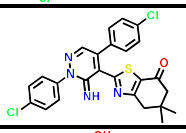
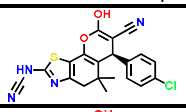
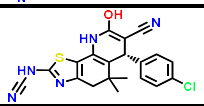
La **Figure III.A5** du tracé de Williams montre le domaine d'applicabilité, caractérisé par une valeur seuil de l'effet de levier (h^*) de 0.52. Les molécules situées dans le domaine à gauche du seuil h^*= 0.52

sont prédites correctement par le modèle RLM-QSAR, tandis que celles situées en dehors du domaine ne le sont pas. Les molécules 2 et 8 ont la déviation standard en dehors de la zone de ± 2.5, mais restent dans le domaine d'applicabilité. En revanche, la molécule 1 de l'ensemble d'apprentissage se situe en dehors du domaine d'applicabilité et est exclue parmi les candidats inhibiteurs de c-Met. D'autre part, les molécules situées à proximité (h^*<0.52) seront gardées en tant qu'inhibiteurs de c-Met candidats.

3.1.2 Criblage in silico basé sur les modèles 2D-QSAR

Après avoir validé le modèle 2D-QSAR développé dans cette étude, nous pouvons en toute confiance exploiter ses prédictions pour sélectionner des molécules candidates en tant que nouveaux inhibiteurs puissants et sélectifs de c-Met par rapport au Crizotinib. D'après les données observées *in vitro* et les prédictions du modèle 2D-QSAR développé à l'aide des techniques RLM, RNLM et RNA *in silico*, les molécules 5, 9, 14, 16, 27, 40 et 46 présentent des activités biologiques inhibitrices vis-à-vis de la protéine c-Met plus fortes que celles observées pour le médicament de référence, le Crizotinib (**Tableau III.A3**).

Tableau III.A3. Les molécules basées sur le 4,5,6,7-Tétrahydrobenzo[D]-thiazol-2-Yl candidats pour inhiber la protéine c-Met plus efficacement que le Crizotinib.

Composé	Structure	pIC_{50} observé	pIC_{50} prédit		
			RLM	RNLM	RNA
5		9.49	9.27	9.28	9.33
9		10.10	9.66	9.72	9.66
14		9.64	9.92	9.55	9.42
16		9.59	9.25	9.31	9.36
27		9.55	9.71	9.69	9.66
40		9.55	9.64	9.65	9.64
46		9.49	9.39	9.44	9.54
pIC_{50} (Crizotinib)=8.09					

Après avoir identifié des molécules candidates comme inhibiteurs sélectifs prometteurs de c-Met, il s'agit d'aborder le comportement pharmacocinétique et pharmacodynamique de ces molécules dans l'organisme humain. Cette étape est l'une des plus importantes dans le processus de conception rationnelle d'un nouveau médicament. À cet égard, nous avons soumis les sept molécules restantes issues du criblage virtuel basé sur l'étude QSAR à un criblage *in silico* à l'aide d'un modèle "Drug-like"

dans un premier temps. Pour mémoire, les évaluations de type Drug-like et ADME-Tox peuvent être réalisées respectivement ou avancer/retarder l'une par rapport à l'autre par le biais de la méthodologie CADD hybride appliquée dans nos travaux. Ceci est dû à la flexibilité de ces modèles prédictifs en termes d'application dans le criblage virtuel. Dans cette étude, nous avons mené le criblage *in silico* du type Drug-like avant l'analyse de Docking moléculaire et le déplacement de la modélisation ADME-Tox comme dernière étape du criblage.

3.2 Criblage virtuel basé sur la modélisation Drug-like

À l'aide du serveur en ligne SwissADME, nous avons construit un modèle de type Drug-like pour prédire les propriétés de biodisponibilité des sept molécules candidates. Le **Tableau III.A4** présente les prédictions computationnelles obtenues pour les propriétés similaires à celles des médicaments que nous avons étudiés, ainsi que les critères/règles utilisés comme références pour l'évaluation. Le **Tableau III.A4** indique que les composés 5, 16 et 46 répondent aux règles de Lipinski, Veber et Egan, et ne présentent donc aucun problème lié à la biodisponibilité orale. Les composés 5, 16 et 46 présentent des capacités d'absorption élevées, tandis que le composé 40 présente une faible capacité d'absorption. Les composés 9, 14 et 27 s'écartent du niveau d'affinité lipidique élevé (LogP > 5) prévu par Lipinski, Veber et Egan, ce qui se traduit par une transformation métabolique plus lourde, une solubilité plus faible et une absorption orale médiocre, de sorte que ces composés ne seront pas pris en considération pour les besoins thérapeutiques.

Tableau III.A4. Prédiction des propriétés de type Drug-like.

Propriétés		ABS	TPSA	n-RB	MW	LogP	n-HBA	n-HBD	Violations			SA
									Lipinski	Veber	Egan	
Règle		-	≤140 Å^2	≤10	<500	≤5	<10	<5	≤1			0<SA<10
Composé	5	Élevé	106.38	3	358.85	3.46	4	1	0	0	0	3.47
	9	Faible	136.89	3	494.03	5.61	4	1	1	1	1	4.18
	14	Faible	115.16	3	490.04	7.51	3	0	1	1	1	4.14
	16	Élevé	81.99	2	356.87	4.32	3	0	0	0	0	3.46
	27	Faible	99.87	3	495.42	5.71	4	1	1	1	1	3.97
	40	Faible	118.17	2	409.89	2.37	5	1	0	0	0	4.26
	46	Élevé	120.97	2	408.90	2.96	4	2	0	0	0	4.32
Abréviations		**ABS** : Absorption, **TPSA** : Surface Polaire Topologique, **n-RB** : Nombre de Liaisons Rotatives, **MW** : Poids Moléculaire, **LogP** : logarithme du coefficient de partage du composé entre n-octanol et l'eau, **n-HBA** : nombre d'accepteurs de liaisons hydrogène, **n-HBD** : nombre de donneurs de liaisons hydrogène, **SA** : Accessibilité Synthétique										

Les composés 5, 16 et 46 présentent des propriétés similaires à celles des médicaments et ont un potentiel de synthèse facile (valeurs S.A. comprises entre 3.46 et 4.32). En outre, ces molécules ont des valeurs TPSA inférieures à 140 Å^2 et un nombre de liaisons rotatives (n-ROTB) inférieur à 10, ce qui leur permet d'interagir de manière flexible avec le récepteur c-Met. Par conséquent, les composés **5**, **16** et **46** peuvent être sélectionnés comme les meilleurs candidats pour inhiber le récepteur c-Met et présentent un bon profil de type Drug-like aux fins des usages pharmaceutiques. Dans la phase suivante, nous abordons l'identification des conformations des ligands **L5**, **L16** et **L46** et leur compatibilité structurelle et énergétique avec la poche active de la protéine c-Met où interagit l'inhibiteur Crizotinib. Ceci afin de sélectionner une structure moléculaire de dérivés du 4,5,6,7-

Tétrahydrobenzo[D]-Thiazol-2-Yl qui peut servir comme modèle structurel (tête de la série) pour optimiser de nouveaux inhibiteurs spécifiques de la protéine c-Met à base de petites molécules. Pour réaliser cette étape de criblage dans cette étude, nous avons docké les ligands 5, 16 et 46 à l'intérieur de la poche active de la protéine c-Met, en suivant la méthodologie de docking moléculaire détaillée dans le deuxième chapitre de cette thèse et dans notre article publié dans ce contexte (Daoui et al. [294]).

3.3 Analyse des interactions entre cMet et l'inhibiteur et évaluation des scores d'énergie de liaison par Docking Moléculaire

3.3.1 Validation du modèle 3D de la protéine ciblée par la procédure de re-Docking

Avant d'effectuer des simulations de Docking moléculaire, nous identifions d'abord le site actif de la protéine c-Met qui regroupe les résidus d'acides aminés de référence avec lesquels l'inhibiteur Crizotinib interagit au sein du complexe cMet-Crizotinib (PDB code : 2WGJ). Dans la présente étude, les résidus d'acides aminés identifiés comme référence pour évaluer l'affinité/compatibilité des ligands 5, 16 et 46 sont la Tyrosine (Tyr1159, Tyr1230), la Méthionine (Met1160 et Met1211), l'Alanine (Ala1108, Ala1226, Ala1221), la Proline (Pro1158), et la Valine (Val1092), la Leucine (Ala1108, Ala1226, Ala1221), et la Leucine (Ala1092). (Val1092), Leucine (Leu1157, Leu1140), Acide aspartique (Asp1222), Isoleucine (Ile1084) (**Figure III.A6**). Après avoir identifié la poche active de c-Met et identifié les résidus d'acides aminés de référence dans son inhibition, il s'agit de caractériser ce site et de le définir au niveau de la structure de c-Met cible. Il s'agit maintenant de caractériser ce site et de le définir au niveau de la structure de la c-Met ciblée, en vue de pouvoir y arrimer (docker) les têtes de série (ligands 5, 16 et 46). A cet égard, dans notre présente étude, le site d'insertion est défini comme suit : 60*60*60 $Å^3$ (points dans la dimension x-y-z), 0.375 Å (espacement entre les points), x = 21.70 Å, y = 83.73 Å, z = 4.28 Å (centres x, y, z).

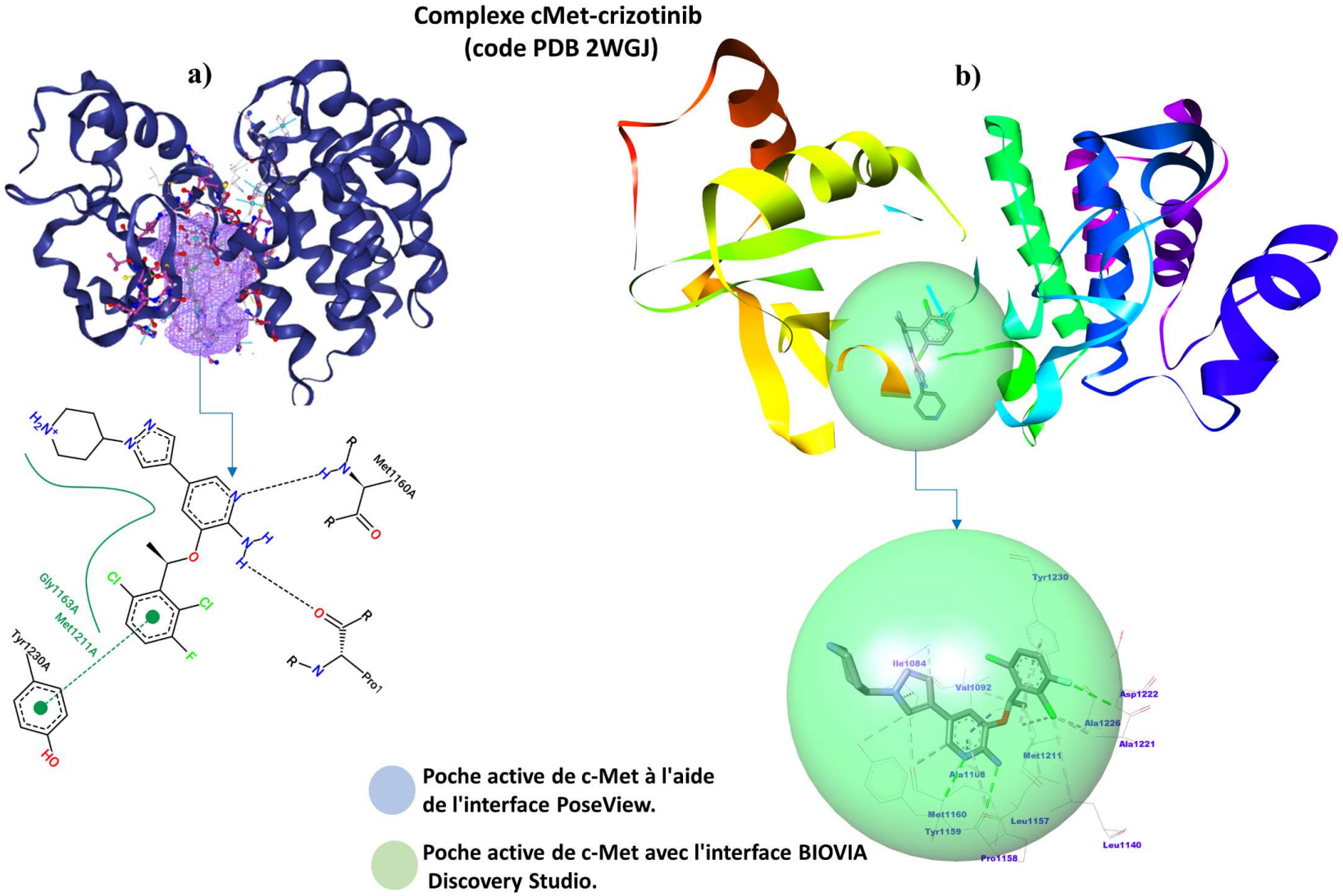


Figure III.A6. Visualisations de la poche active de la protéine c-Met dans le complexe natif 2WGJ (cMet-crizotinb) identifiée via les outils PoseView (a) et BIOVIA Discovery Studio (b).

Après avoir soigneusement préparé les deux parties principales du processus de Docking moléculaire (le récepteur et les ligands), il faut valider la précision et l'efficacité des outils utilisés pour réaliser cette procédure. Dans tous nos travaux réalisés dans le cadre de cette thèse, nous réalisons cette étape importante en mettant en œuvre le protocole de re-Docking. Dans la présente étude, nous avons procédé au re-Docking de l'inhibiteur Crizotinib dans la poche active de c-Met afin de vérifier l'efficacité du package du logiciel AutoDockTools-1.5.6 dans la réalisation de cette procédure et la production de prédictions fidèles. La **Figure III.A7** présente la position superposée de la conformation du Crizotinib re-Dockée (couleur marron) avec la conformation originale du Crizotinib cristallisé dans le complexe 2WGJ (couleur verte). Le calcul de la distance de pose (RMSD) a montré une valeur de 1.428 Å. Cette valeur inférieure à 2 Å reflète la validité et la fiabilité de la méthode de Docking moléculaire choisie pour cette étude.

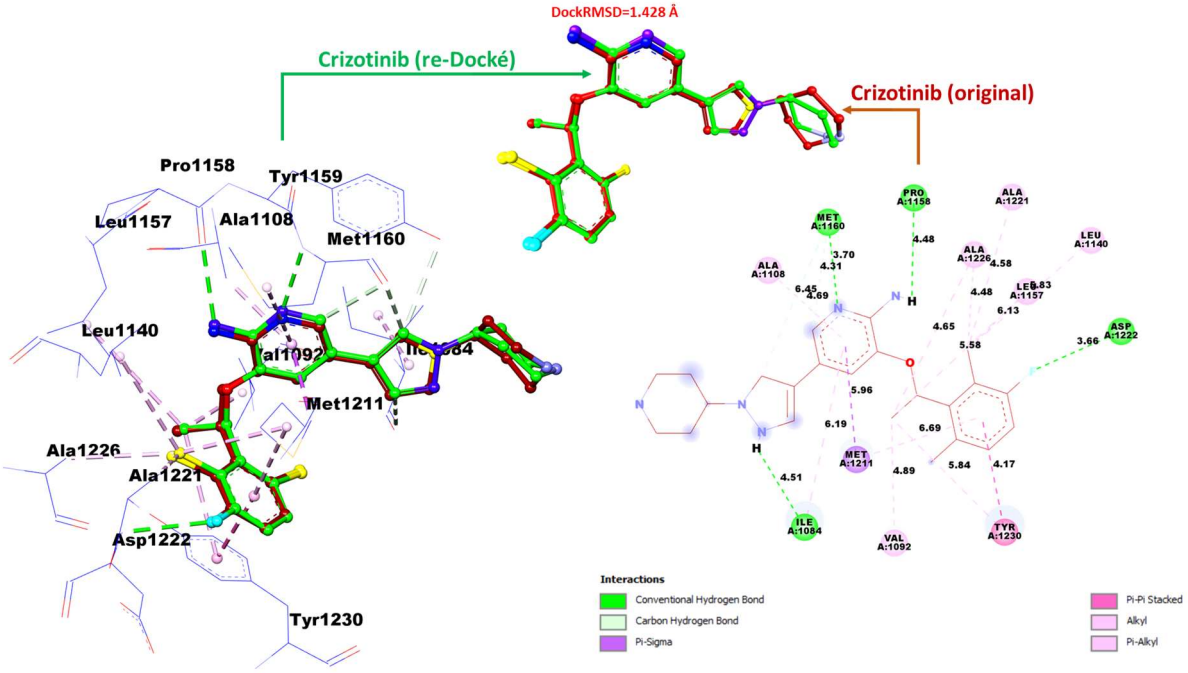


Figure III.A7. Alignement de l'inhibiteur de référence Crizotinib original (couleur verte) avec le Crizotinib re-Docké (couleur marron) dans la poche active de c-Met.

Au-delà de la valeur RMSD calculée qui est inférieure à 2 Å, nous avons également procédé à une inspection visuelle pour valider l'efficacité du modèle de Docking proposé dans le cadre de cette étude. À cet égard, la **Figure III.A7** confirme que le Crizotinib redocké interagit avec le même résidu d'acide aminé de référence que l'inhibiteur dans le complexe 2WGJ d'origine. Cela indique que les étapes méthodologiques et techniques que nous avons mises en œuvre dans cette procédure sont tout à fait appropriées pour générer des prédictions correctes sur les interactions des ligands **5**, **16** et **46** dans la poche active de c-Met.

3.3.2 Découverte de potentiels inhibiteurs de c-Met basés sur le 4,5,6,7-Tétrahydrobenzo[D]-Thiazol-2-Yl

Après avoir validé avec succès le protocole de Docking moléculaire, nous exécutons le Docking des ligands **5**, **16** et **46** et comparons leur comportement en termes d'interaction avec l'inhibiteur de référence Crizotinib dans la poche active de la protéine c-Met. La **Figure III.A8** présente un aperçu en

2D et 3D des conformations des ligands les plus stables détectés après leur Docking vis-à-vis du récepteur de c-Met.

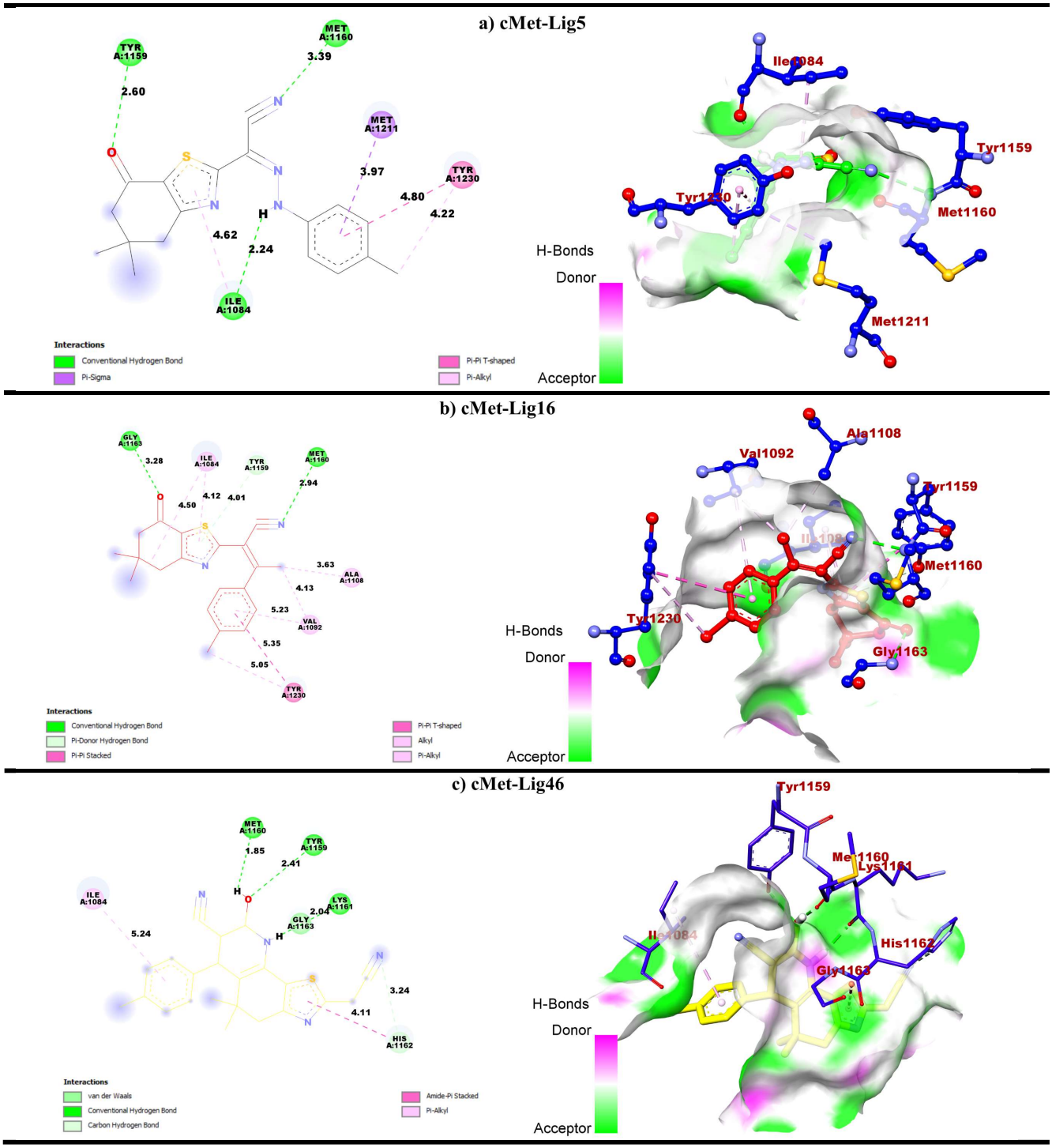


Figure III.A8. Visualisations 2D (à droite) et 3D (à gauche) des conformations des ligands L5 (a), L16 (b) et L46 (c) complexés avec c-Met.

Le **Tableau III.A5** précise les types d'interactions protéine-ligand identifiées dans les complexes cMet-L5, cMet-L16, cMet-L46, et cMet-crizotinib ainsi que l'énergie libre de liaison estimée dans chaque complexe. D'après les résultats présentés dans la **Figure III.A8** et les données présentées dans le **Tableau III.A5**, il est évident que les ligands L5, L16, et L46 sont dockés avec succès dans la poche active de c-Met. Cela peut être confirmé par les fortes interactions non-covalentes de type hydrophobique et les liaisons hydrogène conventionnelles qui ont eu lieu entre ces ligands et un ensemble de résidus d'acides aminés identifiés comme étant des références. En outre, les énergies de

liaison des ligands L5, L16 et L46 envers le c-Met sont estimées à -7,08 kcal/mol, -8.27 kcal/mol et -7.40 kcal/mol, tandis que le médicament de référence Crizotinib présente une énergie de liaison de -7.02 kcal/mol. Cela indique que les structures des petites molécules actives L5, L16 et L46 sont susceptibles de présenter une meilleure action thérapeutique, une meilleure stabilité et une meilleure compatibilité avec la structure de la protéine c-Met que le médicament de référence, le Crizotinib.

Tableau III.A5. Interactions protéine-ligand, énergies libres de liaison estimées par Docking moléculaire dans les complexes de c-Met.

Ligands	Complexe	Énergie de liaison (kcal/mol)	Interaction par liaison hydrogène	Interaction de type hydrophobique
L_{ref}	cMet-crizotinib	-7.02	**Met1160, Pro1158, Asp1222, Ile1084**	**Ala1108, Ala1221, Ala1226, Leu1157, Leu1140, Tyr1230, Val1092, Met1211**
L5	cMet-L5	-7.08	Tyr1159, **Met1160**, **Ile1084**	**Met1211**, **Tyr1230**
L16	cMet-L16	-8.27	Gly1163, **Met1160**, Tyr1159	**Ile1084**, **Ala1108**, **Val1092**, **Tyr1230**
L46	cMet-L46	-7.40	**Met1160**, Tyr1159, Lys1161, His1162	**Ile1084**, His1162

En **gras** : les résidus d'acides aminés de référence

En ce qui concerne le nombre d'interactions non-covalentes qui ont eu lieu entre les conformations actives L5, L16 et L46 et les résidus d'acides aminés de référence, nous pouvons noter que L16 a interagi avec le plus grand nombre de résidus d'acides aminés de référence (cinq) qui sont Met1160, Ile1084, Ala1108, Val1092, et Tyr1230. Les ligands L5 et L46 interagissent avec quatre et deux de ces résidus de référence (Met1160, Ile1084) et (Met1211, Tyr1230), respectivement. D'après les résultats des simulations de Docking moléculaire, il est clair que les conformations des ligands L5, L16 et L46 peuvent être considérées comme les structures de tête de la série de dérivés 5,6,7-Tétrahydrobenzo[D]-Thiazol-2-Yl étudiés. Ils peuvent donc être sélectionnés comme inhibiteurs spécifiques potentiels du récepteur tyrosine kinase c-Met. En raison de la distinction du ligand L16 de ses homologues L5, et L16, en termes d'activité biologique pIC_{50} observée *in vitro* et de ses propriétés similaires à celles d'un médicament, ainsi qu'en termes d'énergie de liaison à c-Met et du nombre de ses interactions avec les résidus d'acide aminé de référence, ce dernier peut être considéré comme une référence pour la conception davantage de nouvelles molécules portant l'échafaudage 4,5,6,7-Tétrahydrobenzo[D]-Thiazol-2-Yl. Par la suite, les modèles CADD hybrides développés dans cette étude peuvent être exploités pour prédire et évaluer l'activité biologique inhibitrice optimale et les profils de compatibilité des nouvelles molécules candidates afin de construire une nouvelle génération d'inhibiteurs non covalents de c-Met destinés à être appliqués dans la thérapie ciblée contre le cancer. Mais avant d'oser ces hypothèses qui ont été formulées à travers les modèles computationnels que nous avons obtenus dans cette étude. Tout d'abord, des recherches computationnelles plus approfondies devraient être menées sur la structure de la molécule L16 sélectionnée dans le cadre de cette étude comme étant un candidat de premier choix destiné pour devenir un inhibiteur sélectif puissant de c-Met à des fins thérapeutiques contre le cancer. Il convient également de définir ses propriétés pharmacocinétiques et pharmacodynamiques *in silico* par le biais d'une modélisation ADME-Tox et de procéder à des simulations de post-Docking telles que des calculs de MM-GBSA et de

dynamique moléculaire. Dans cette étude, nous avons ouvert la voie à de futures études sur les ligands L16 par le biais de la comparaison de leurs propriétés ADME-Tox avec le profil du médicament de référence Crizotinib.

3.4 Profil Pharmacocinétique et Pharmacodynamique des inhibiteurs de c-Met identifiés *in silico*

Dans la présente étude, la molécule #16 (L16) et le médicament Crizotinib font l'objet d'une analyse *in silico* avec l'outil pkCSM pour prédire leurs propriétés ADME-Tox (Absorption, Distribution, Métabolisme, Excrétion, Toxicité). Les résultats de la prédiction sont présentés dans le **Tableau III.A6**, ce qui permet de comparer les propriétés des deux échantillons (L16 et Crizotinib).À la suite des résultats ADME-Tox présentés dans le tableau 7, nous pouvons constater que :

Le composé #16 présente un taux d'absorption dans l'intestin humain nettement supérieur à celui du Crizotinib, avec un pourcentage de plus de 94 %. Cela indique qu'il aura un meilleur pouvoir d'absorption que le Crizotinib, dont le pouvoir d'absorption était de 91.38 %. Cela suggère que le composé 16 sera fortement administré dans l'intestin humain.

Tableau III.A6. Propriétés ADME-Tox prédites *in silico* pour le composé "hit" (16) et le médicament standard Crizotinib (indiqué par Criz.).

	Propriétés ADME-Tox												
	Absorption	Distribution			Métabolisme							Excrétion	Toxicité
Modèle	Absorption intestinale (Humain)	VDss (humain)	Perméabilité (BBB)	Perméabilité (CNS)	CYP							Clairance totale	Toxicité d'AMES
					Substrat		Inhibiteur						
					2D6	3A4	1A2	2C19	2C9	2D6	3A4		
Unité	Numérique (%)	Numérique (Log L kg^{-1})	Numérique (Log BB)	Numérique (Log PS)	Catégorique (oui/non)							Numérique (log mL min^{-1} kg^{-1})	Catégorique (Oui/non)
Données prédites													
#16	94.70	0.474	0.396	-1.438	**Non**	**Oui**	**Oui**	**Oui**	**Oui**	Non	**Oui**	0.113	No
Criz.	91.38	1.315	-1.016	-3.005	Non	**Oui**	Non	**Oui**	Non	Non	**Oui**	0.583	Non

Selon les résultats obtenus, le VDss et le LogBB de la molécule #16 sont plus élevés que ceux du Crizotinib, ce qui la rend plus apte à la distribution, et sa valeur LogPS indique même son potentiel de pénétration dans le SNC. Dans l'ensemble, les indices de distribution rapportés pour la molécule 16 montrent qu'elle est plus susceptible d'être distribuée efficacement que le Crizotinib. Le cytochrome (CYP450) est une enzyme importante pour la désintoxication que l'on trouve dans tous les tissus de notre corps. Elle est responsable du métabolisme de la plupart des médicaments et peut être inhibée ou activée par des composés [423]. Les CYP1, CYP2, CYP3 et CYP4 sont principalement responsables du métabolisme des médicaments à l'aide des isoformes 2D6 et 3A4 [423]. Une étude a montré que le métabolisme du Crizotinib peut être affecté par une augmentation de l'activité enzymatique du CYP3A4, qui peut inhiber l'activité du médicament [424]. Afin de prédire l'effet métabolique du CYP3A4 sur un candidat médicament, l'effet du composé 16 sur le CYP3A4 doit être évalué. Les résultats du métabolisme du composé 16 montrent qu'il peut être à la fois un substrat et un inhibiteur du CYP3A4, ce qui signifie qu'il peut atteindre sa cible thérapeutique avant d'être transformé et éliminé. L'indice de clairance du médicament en termes de propriétés d'excrétion de la molécule 16

et du Crizotinib est analysé afin d'évaluer la stabilité potentielle de la molécule 16 en tant que molécule médicamenteuse et sa capacité à rester plus longtemps dans le corps avant d'être éliminée. Les valeurs prédites de l'indice de clairance totale ont montré que la molécule 16 a un indice de clairance totale de 0.113, ce qui est inférieur à l'indice de clairance totale du Crizotinib (0.583), ce qui suggère que la molécule 16 aura une persistance plus longue dans l'organisme que le Crizotinib. Cela signifie également que la molécule 16 pourrait être plus active à des doses plus faibles que le Crizotinib, car elle restera plus longtemps dans l'organisme avant d'être éliminée. Les risques de toxicité de deux molécules, 16 et Crizotinib, sont évalués à l'aide du test AMES. Ce test a révélé que la molécule 16 n'était pas toxique, alors que la molécule Crizotinib présentait des caractéristiques toxiques. Grâce à ce criblage virtuel basé sur le modèle ADME-Tox *in silico*, il est évident que les propriétés structurelles de la molécule 16 sont plus compatibles du point de vue de la conception médicamenteuse que celles de la molécule de référence, le Crizotinib. Cela confirme en outre les faibles propriétés PK/PD du Crizotinib rapportées dans plusieurs études. Ces propriétés peuvent être à l'origine de la résistance qui se développe chez les patients cancéreux soumis à une chimiothérapie impliquant le Crizotinib. Par conséquent, l'optimisation de nouveaux médicaments basés sur la structure du ligand **16** peut être très utile pour parvenir à une nouvelle stratégie thérapeutique efficace contre le cancer. Cela se fait par l'application isolée des nouveaux inhibiteurs de c-Met ou en combinaison avec les anciens inhibiteurs comme le Crizotinib en tant que modèle.

4 Conclusion

L'objectif principal de cette étude est d'explorer *in silico* une série de petites molécules biologiquement actives dérivées de 4,5,6,7-Tétrahydrobenzo[D]-thiazol-2-Yl vis-à-vis du récepteur de tyrosine kinase c-Met. Cette démarche s'inscrit dans le cadre des applications du modèle CADD hybride pour la conception rationnelle et la recherche de nouveaux médicaments anticancéreux. Dans la première phase de l'étude, nous avons décrypté la relation structure-activité des molécules étudiées en nous basant sur une approche QSAR établie utilisant des descripteurs hybrides topologiques, physico-chimiques et de réactivité basée sur la DFT (0D-3D). À cette fin, nous avons utilisé plusieurs méthodes de modélisation statistique avancées pour construire un modèle 2D-QSAR robuste, incluant l'ACP, les k-moyennes, la RLM, la RNLM, le RNA (ANN), la validation croisée, le domaine d'applicabilité, la Y-randomisation, entre autres. Ce processus nous a permis de développer un modèle 2D-QSAR stable, caractérisant l'activité biologique des molécules dérivées du 4,5,6,7-Tétrahydrobenzo[D]-thiazol-2-Yl à l'aide de sept descripteurs moléculaires hybrides. Ensuite, nous avons élaboré un modèle de type Drug-like pour soutenir le modèle 2D-QSAR dans la sélection des molécules candidates les plus prometteuses pour la conception de nouveaux médicaments. Les molécules sélectionnées, notamment les molécules 5, 16 et 46, ont été identifiées comme les plus actives biologiquement, et donc appropriées pour le développement d'inhibiteurs spécifiques de la protéine c-Met. Ces molécules ont ensuite été soumises à des simulations de Docking moléculaire et à une modélisation ADME-Tox afin d'évaluer leur compatibilité avec la structure de la protéine c-Met et leur profil pharmacocinétique et pharmacodynamique. Les résultats du Docking moléculaire ont démontré que les ligands L5, L16 et L46 présentaient une forte affinité et compatibilité avec la protéine c-Met par rapport au médicament

de référence, le Crizotinib. Ces résultats ont été confirmés par une série d'interactions non-covalentes entre les ligands et les acides aminés de la protéine c-Met. Parmi les trois molécules étudiées, la molécule L16 s'est avérée particulièrement prometteuse dans toutes les évaluations menées jusqu'à présent. Les prédictions ADME-Tox *in silico* ont également confirmé que cette molécule présentait des propriétés pharmacocinétiques et pharmacodynamiques favorables par rapport au Crizotinib. En conclusion, la molécule L16, dérivée de Dimédone portant le 4,5,6,7-Tétrahydrobenzo[D]-thiazol-2-Yl, émerge comme un candidat privilégié pour le développement d'une nouvelle génération d'inhibiteurs spécifiques de c-Met pour la thérapie ciblée du cancer. Les résultats obtenus dans cette étude peuvent servir de base pour la conception et l'évaluation de nouveaux dérivés du 4,5,6,7-Tétrahydrobenzo[D]-thiazol-2-Yl par des approches CADD similaires. Il convient également de souligner que l'importance de c-Met dans de nombreuses maladies, notamment le cancer, suggère que davantage d'attention devrait être accordée à ce récepteur pour développer des stratégies d'inhibition ou de modulation de son activité enzymatique. Dans la section suivante de notre recherche, nous poursuivrons nos investigations en utilisant des approches hybrides CADD et des simulations biophysiques pour modéliser une nouvelle génération d'inhibiteurs spécifiques de c-Met, cette fois basés sur des petites molécules à base de 4-Phénoxypyridine.

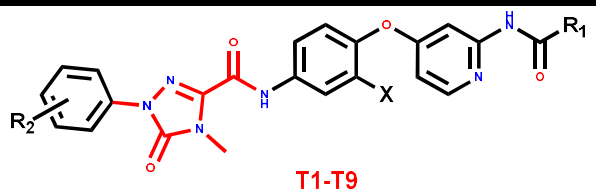
B. Criblage *in silico*, Modélisation et Optimisation d'Inhibiteurs de Type II du Récepteur Tyrosine Kinase c-Met à Base de Phénoxypyridine pour la Thérapie Ciblée Anticancéreuse

1 Introduction

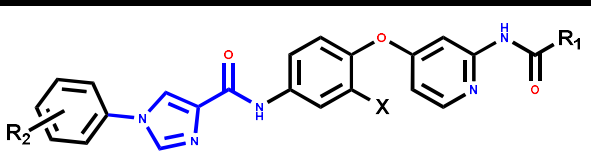
Conformément à nos objectifs généraux visant à étendre les applications des méthodologies CADD hybrides dans la conception et la découverte de médicaments *in silico*, cette étude est venue comme un progrès dans l'un de nos objectifs de modélisation et de conception de nouveaux inhibiteurs du récepteur c-Met de la tyrosine kinase, basés sur de petites molécules bioactives. Depuis la découverte de la structure moléculaire de l'ADN par James Watson, des ensembles de gènes contenus dans l'ADN sont connus pour informer la cellule d'un processus de production de protéines [425]. L'étude des protéines et des mécanismes d'interaction entre elles s'est avérée très pertinente pour la recherche de thérapies ciblant des maladies incurables comme le cancer. En particulier, l'inhibition des récepteurs de protéines RTK, tels que c-Met, a permis de contrôler la destruction des voies de signalisation impliquées dans la croissance et la génération des cellules cancéreuses [122]. Dès lors, plusieurs inhibiteurs à base de petites molécules ont été proposées pour être utilisées comme agents thérapeutiques dans différents types de cancers, avec succès. Les inhibiteurs de c-Met à base de petites molécules peuvent être divisés en deux familles (Type I et Type II) sur la base de leurs propriétés structurelles [426]. Les inhibiteurs de Type I se sont révélés plus facilement sensibles à la résistance due à la mutation des résidus d'acides aminés de la poche de c-Met, tandis que les inhibiteurs de Type II ont tendance à avoir une meilleure activité pour surmonter de telles mutations en raison des interactions multiples entre les inhibiteurs et la poche de c-Met [408], [427]. Des médicaments ont déjà été commercialisés, comme le Cabozantinib [428], et d'autres, comme le Tivantinib [429], le Crizotinib [430], le Foretinib [431], l'Onartuzumab [432], le BMS-777607 [433], l'Altiratinib [434] et le NPS-1034 [435], ont tous été soumis à des essais cliniques avec succès. Même si ces médicaments ont eu des effets bénéfiques qui se sont manifestés en préservant la vie du patient plus longtemps, leurs effets négatifs se manifestent généralement par de multiples effets secondaires sur les patients pendant la période de traitement et par l'absence d'une réponse positive rapide et immédiate sur les sites cancéreux [436], [437]. Cependant, de nouveaux agents ayant une activité biologique optimale restent encore à développer afin de bien inhiber l'activité enzymatique de c-Met et de découvrir de nouveaux médicaments anticancéreux peu coûteux, à faible toxicité et très efficace. Dans ce contexte, afin d'optimiser et d'élargir la liste des agents moléculaires candidats à la thérapie ciblée contre le cancer via l'inhibition de c-Met, nous avons arrêté sur une étude réalisée *in vitro* par (JuLiu et al.) [408]. Cette étude a synthétisé une série de 26 dérivés de la 4-phénoxypyridine portant des moitiés imidazole-4-carboxamide (**T1-T9**) et 1,2,4-triazole-3-carboxamide (**T10-T26**), qui ont présenté une activité inhibitrice de c-Met (**Tableau III.B1**). Des évaluations de l'activité biologique inhibitrice *in vitro* (IC_{50}) en présence de Foretinib comme référence ont indiqué que les dérivés de 4-phénoxypyridine peuvent être considérés comme de futurs candidats pour développer une nouvelle génération d'inhibiteurs de c-Met de type II appliqués à la thérapie anticancéreuse [408]. Malgré l'importance des évaluations biologiques

et de la caractérisation structurelle des molécules obtenues le modèle de criblage réalisé *in vitro*, ces évaluations ne sont pas suffisantes pour positionner ces molécules dans le pipeline de la conception et de la découverte de médicaments. Par conséquent, les résultats rapportés doivent être soumis à davantage d'évaluations/investigations expérimentales et/ou computationnelles, chacun en fonction de sa spécialité et de son savoir dans ce domaine (conception et découverte de médicaments). Pour notre part, nous avons décidé de contribuer à faire progresser les résultats de cette étude en les soumettant à des évaluations approfondies *in silico*. Parmi les lacunes scientifiques qui nous ont conduits à viser cette étude et à nous intéresser à développer ses résultats, nous mentionnons en particulier : (1) la nouveauté de l'étude rapportée ; (2) l'absence de décodage quantitatif de la relation structure-activité pour ces molécules, malgré leur importance critique pour rationaliser la conception et l'optimisation de nouvelles molécules; (3) l'absence du rapport du type Drug-like et ADME-Tox; (3) l'absence de critères d'affinité et de compatibilité des molécules avec la structure de la protéine c-Met ciblée. Toutes ces lacunes nous ont incités à appliquer notre stratégie CADD hybride afin de les réduire et de rationaliser le choix des meilleurs ligands bioactifs contenant la 4-phénoxypyridine candidats comme nouveaux inhibiteurs de type II de la protéine c-Met.

Tableau III.B1. Données structure-activité observées *in vitro* pour les 26 dérivés de 4-phénoxypyridine vis-à-vis la protéine c-Met (T1-T26).

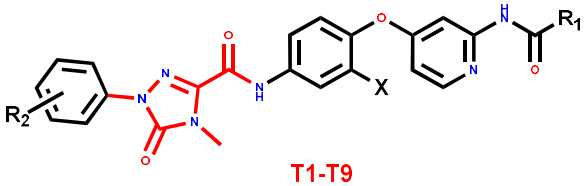


T1-T9

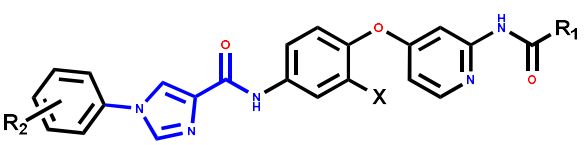


T10-T26

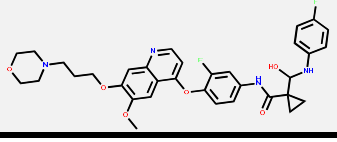
Composé	R_1	R_2	X	pIC_{50}
T01		2-F	F	6.09
T02		2-Br	F	5.66
T03		2-OCH3	F	5.43
T04		3-F	F	5.62
T05		4-Cl	F	6.01
T06		4-OCHF2	F	5.18
T07		2,4-diCl	F	5.87
T08		2,6-diCH3	F	5.23
T09		2-F	H	5.61
T10		2-F	F	6.82
T11		2-Cl	F	7.04
T12		2-Br	F	6.48
T13		2-OCF3	F	6.08
T14		4-F	F	7.92
T15		4-Cl	F	7.19
T16		4-CF3	F	6.39
T17		4-OCH3	F	5.23
T18		4-CH3	F	5.55
T19		2-Br-4-F	F	6.27
T20		2,6-diF	F	6.43
T21		2,6-diCH3	F	5.43



T1-T9



T10-T26

Composé	R_1	R_2	X	pIC_{50}
T22		2-F	F	5.45
T23		4-F	F	5.90
T24		4-OCH3	F	5.00
T25		4-F	H	5.20
T26		4-F	H	5.06
Foretinib				8.68

Foretinib utilisé comme référence de contrôle positif *in vitro*

2 Stratégie de conception de médicaments assistée par ordinateur développée et appliquée

Au-delà des résultats positifs obtenus par la stratégie hybride CADD que nous avons appliquée au criblage de la série de dérivés 4,5,6,7-Tétrahydrobenzo[D]-Thiazol-2-Yl pour identifier les candidats inhibiteurs de c-Met en vue de développer de nouveaux médicaments contre le cancer dans l'étude précédente, la présente étude a ajouté des protocoles de conception/optimisation (modélisation) de nouvelles molécules de dérivés de 4-phénoxypyridine. Cette démarche vise à renforcer la base de données étudiée et à tirer davantage de conclusions et d'hypothèses pharmacologiques liées à ce type de petites molécules biologiquement actives. Le schéma proposé dans la **Figure III.B1** résume le modèle CADD hybride adopté dans cette étude pour examiner, concevoir et sélectionner les meilleures structures moléculaires candidates en tant qu'inhibiteurs de type II de c-Met destinés pour la thérapie ciblée contre le cancer. Dans le cadre de cette étude, l'inhibiteur Foretinib est adopté comme modèle de référence aux fins de contrôle des résultats obtenus. Nous avons discuté de toutes les considérations méthodologiques et techniques liées aux différentes méthodes que nous avons utilisées pour cette stratégie et à leurs protocoles de mise en œuvre dans le deuxième chapitre de cette thèse, ainsi que dans la version électronique de l'article que nous avons publié à ce sujet (Daoui et al. [359]).

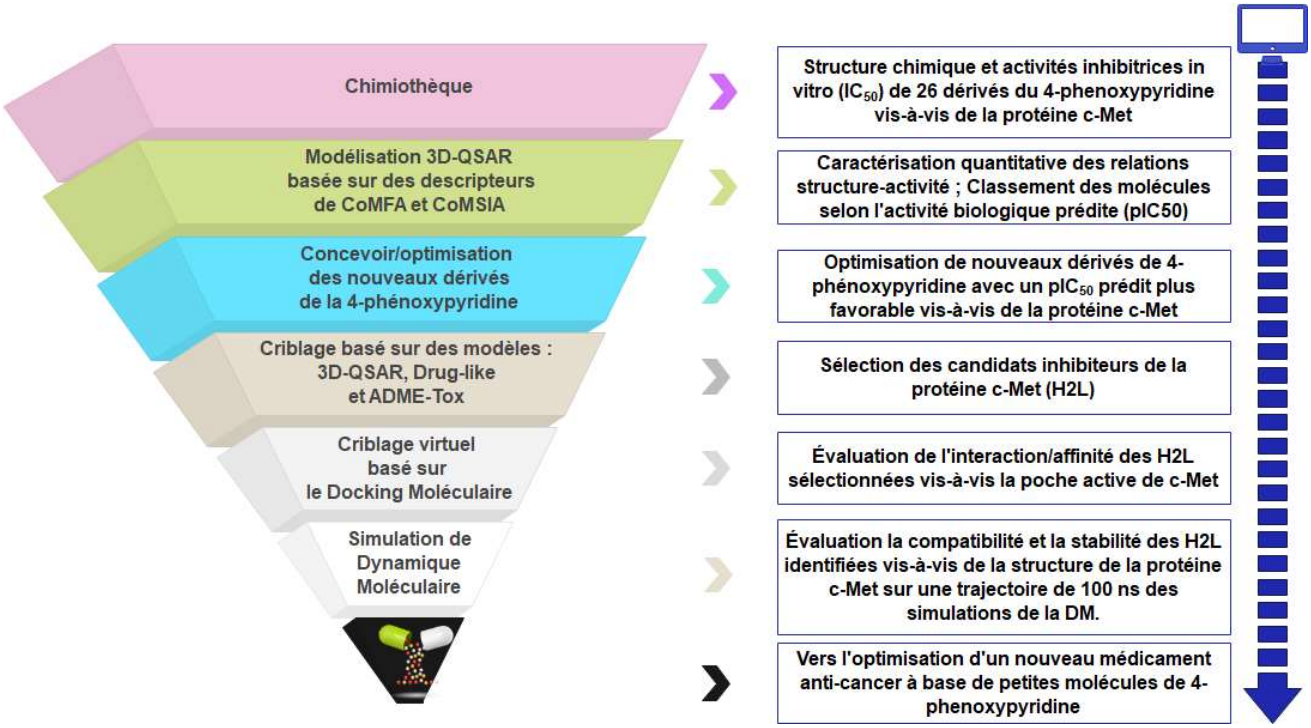


Figure III.B1. Stratégie de CADD proposée et appliquée *in silico* pour concevoir et identifier de nouveaux inhibiteurs non covalents de type II de la protéine c-Met à base de 4-phénoxypyridine en vue d'une thérapie ciblée contre le cancer.

3 Résultats et discussion

3.1 Analyses structure-activité basée sur la modélisation 3D-QSAR

3.1.1 Exploration et modélisation in silico à travers l'orientation des modèles de CoMFA et CoMSIA

Grâce à l'alignement des structures 3D des dérivés de la 4-phénoxypyridine sur le composé prototype **T14** (conformation la plus active) (**Figure III.B2**), la base de données de 26 composés est divisée en deux ensembles ("entraînement" et "test"). Cette division était réalisée sur la base de la diversité structurelle et des variations de pIC_{50}. La distribution optimale obtenue, adaptée à la construction de modèles 3D-QSAR, se présente comme suit : (T1, T2, T3, T4, T5, T6, T7, T8, T10, T12, T13, T14, T16, T17, T18, T20, T21, T24, T25, T26) pour élaborer des modèles CoMFA et CoMSIA et (T11, T15, T19, T22, T23) pour tester leur fiabilité statistique et prédictive. Avec la distribution optimale obtenue, des modèles 3D-QSAR sont développés afin d'identifier les descripteurs de champ stérique (S), électrostatique (E), hydrophobe (H), donneur de liaisons hydrogène (D) et accepteur de liaisons hydrogène (A) qui ont influencé le pIC_{50} des dérivés de 4-phénoxypyridine criblés.

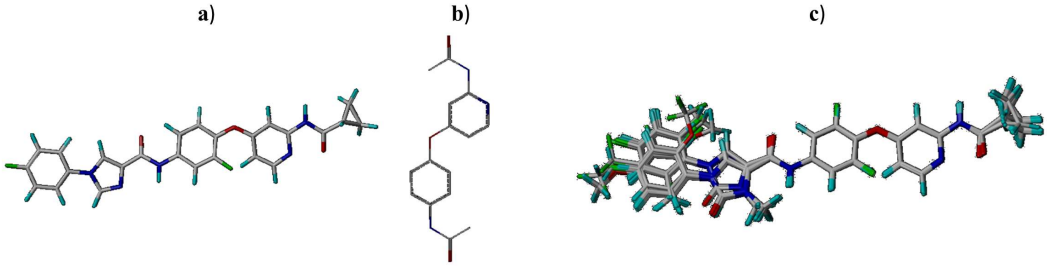


Figure III.B2. (a) structure 3D de la conformation moléculaire prototype T14 (pIC_{50}=7.92), (b) unité de base commune, (c) base de données des dérivés de 4-phénoxypyridine alignée.

La relation logique entre les descripteurs moléculaires 4D de CoMFA et CoMSIA calculées et l'activité biologique inhibitrice pIC_{50} est généré à l'aide de l'algorithme des moindres carrés partiels (PLS). Le **Tableau III.B2** présente les analyses PLS des modèles CoMFA et CoMSIA générés dans cette étude.

Tableau III.B2. Résultats des analyses PLS pour les modèles CoMFA et coMSIA générés.

Modèle	Les seuils de performance des modèles développés selon les seuils d'acceptation						Contribution des descripteurs Fractions				
	Q^2	R^2	SEE	F	N	R^2_{pred}					
	>0.5	>0.6	Faible	Élevé	3<N<6	>0.6	S	E	H	D	A
^{0}CoMFA/SE	**0.733**	**0.96**	**0.157**	**64.77**	**3**	**0.94**	**0.475**	**0.525**	**-**	**-**	**-**
^{1}CoMSIA/SEA	0.485	0.90	0.203	37.75	4	0.77	0.133	0.526	-	-	0.342
^{2}CoMSIA/SEH	0.527	0.95	0.176	51.03	3	0.88	0.539	0.303	0.158	-	-
^{3}CoMSIA/SED	0.591	0.95	0.178	49.71	3	0.90	0.702	0.238	-	0.059	-
^{4}CoMSIA/SHD	0.572	0.84	0.330	12.86	5	0.74	0.292	-	0.541	0.167	-
^{5}CoMSIA/SHA	0.523	0.81	0.361	10.42	5	0.73	0.210	-	0.390	-	0.400
^{6}CoMSIA/SDA	0.368	0.74	0.432	6.64	6	0.69	0.241	-	-	0.212	0.547
^{7}CoMSIA/EHA	0.473	0.93	0.214	34.01	5	0.76	-	0.419	0.244	-	0.337
^{8}CoMSIA/EHD	0.456	0.88	0.185	46.17	6	0.74	-	0.588	0.368	0.044	-
^{9}CoMSIA/EDA	0.524	0.93	0.220	32.05	6	0.71	-	0.529	-	0.055	0.416
^{10}CoMSIA/HDA	0.368	0.74	0.423	6.93	5	0.63	-	-	0.407	0.133	0.460
^{11}CoMSIA/EHDA	0.496	0.91	0.242	25.88	6	0.75	-	0.407	0.225	0.052	0.315
^{12}CoMSIA/SHDA	0.521	0.80	0.377	9.63	6	0.71	0.095	-	0.272	0.157	0.746
^{13}CoMSIA/SEDA	0.577	0.93	0.209	35.55	6	0.77	0.120	0.491	-	0.069	0.320
^{14}CoMSIA/SEHA	0.616	0.94	0.204	37.45	6	0.79	0.106	0.383	0.211	-	0.300
^{15}CoMSIA/SEHD	**0.716**	**0.95**	**0.181**	**48.15**	**4**	**0.93**	**0.155**	**0.499**	**0.299**	**0.047**	**-**
^{16}CoMSIA/SEHDA	0.554	0.92	0.222	29.79	5	0.86	0.098	0.361	0.199	0.062	0.280

Q^2 est le coefficient de corrélation de validation croisée, **R^2** est le coefficient de corrélation non validé, **SEE** est l'erreur standard d'estimation, F est la valeur du F-test, **N** est le nombre optimal de composantes déterminé par PLS-loocv , **R^2_{pred}** est le coefficient de corrélation prédictif obtenu par validation externe. Le terme "**Fraction**" désigne les contributions des champs S, E dans le modèle CoMFA et S, E, H, D, A dans le modèle CoMSIA.

Les valeurs pIC_{50} observées *in vitro* et celles prédites *in silico* par les modèles 3D-QSAR que nous avons développés à l'aide des descripteurs 4D (CoMFA et CoMSIA) sont reportées dans le **Tableau III.B3**. Les données présentées dans les **Tableaux** (**II.2** et **II.3**) nous ont permis d'identifier les descripteurs moléculaires 4D qui sont favorables/défavorables pour optimiser l'activité biologique inhibitrice (pIC_{50}) des dérivés de 4-phénoxypyridine envers la protéine c-Met. Conformément au **Tableau III.B2**, les paramètres de PLS tels que Q^2, R^2, F, N, SEE, R^2 et les fractions des descripteurs indiquent le degré de stabilité et la signification statistique des modèles 3D-QSAR de CoMFA et de CoMSIA développées. Parmi les dix-sept modèles développés, nous pouvons constater que le modèle CoMFA/SE présente une forte corrélation de 0.96 entre l'activité biologique inhibitrice pIC_{50} et les propriétés stériques (47.5 %) et électrostatiques (52.5 %) CoMFA des dérivés de la 4-phénoxypyridine. Le modèle CoMSIA/SEHD présente également une forte corrélation avec le pIC_{50} et les propriétés stériques (15.5 %), électrostatiques (49.9 %), hydrophobes (29.9 %) et donneuses de liaisons H (4.7 %), avec une valeur R^2 de 0.95. La forte concordance entre les valeurs expérimentales de pIC_{50} et celles prédites par les modèles 3D-QSAR est confirmée par les données présentées dans le **Tableau III.B3** recueillies dans le cadre de ce travail. Cette convergence s'explique par les faibles valeurs de SEE enregistrées (0.157 et 0.181) pour les deux modèles optimaux CoMFA/SE et CoMSIA/SEHD, respectivement. Les indicateurs de validation interne et externe des modèles CoMFA/SE (R^2=0.96, Q^2=0.68, SEE=0.157, F=64.77, R^2_{pred}=0.94) et CoMSIA/SEHD (R^2=0.95, Q^2=0.63, SEE=0.181, F=48.15, R^2_{pred}=0.93) indiquent que ces modèles sont très performants. Par conséquent, ces modèles

peuvent être appliqués au criblage virtuel et à la rationalisation de la conception de nouvelles molécules à base de 4-phénoxypyridine *in silico*.

Tableau III.B3. Les valeurs de pIC_{50} des dérivés de la 4-phénoxypyridine observés *in vitro* par rapport à ceux prédits par les descripteurs CoMFA et CoMSIA *in silico*.

c-Met (pIC_{50})	*In vitro*	3D-QSAR *in silico*																
		CoMFA	CoMSIA															
		(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	**(15)**	(16)
T01	6.09	5.94	6.12	6.13	6.13	6.02	5.84	6.14	6.03	6.77	6.19	5.72	6.05	5.84	5.18	6.02	6.14	5.68
T02	5.66	5.63	5.65	5.53	5.75	5.55	5.63	5.52	5.70	5.35	5.56	5.30	5.65	5.43	4.87	5.71	5.51	5.81
T03	5.43	5.49	5.46	5.40	5.31	5.45	5.52	5.30	5.53	5.34	5.56	5.93	5.60	5.81	4.87	5.53	5.39	5.50
T04	5.62	5.75	5.63	5.71	5.63	5.78	5.87	6.08	5.60	5.71	5.64	5.69	5.63	5.78	5.36	5.66	5.72	5.64
T05	6.01	5.95	6.07	6.32	6.18	6.44	6.37	6.08	6.13	6.27	6.08	6.04	6.08	6.07	5.24	6.17	6.31	6.13
T06	5.18	5.28	4.90	4.95	5.03	4.77	4.94	5.01	4.94	4.99	4.94	5.26	4.95	4.98	5.14	4.88	4.95	4.99
T07	5.87	5.87	6.23	5.88	6.12	5.84	5.94	5.64	6.24	5.93	6.28	6.17	6.29	6.12	5.51	6.15	5.89	6.25
T08	5.23	5.05	5.19	5.23	5.12	5.39	5.04	5.12	5.06	5.14	5.08	5.05	4.97	5.09	5.42	5.14	5.23	5.09
T09	5.61	5.74	5.47	5.51	5.40	5.47	5.57	5.64	5.53	5.55	5.40	5.63	5.55	5.67	5.65	5.50	5.48	5.52
T10	6.82	6.92	6.89	6.85	6.82	6.97	6.90	6.86	6.86	6.88	6.81	6.94	6.87	7.02	5.47	6.89	6.88	6.74
T12	6.48	6.64	6.61	6.46	6.49	6.74	6.61	6.85	6.50	6.49	6.52	6.43	6.42	6.52	5.63	6.49	6.45	6.48
T13	6.08	5.77	6.03	6.00	6.00	6.01	5.91	6.08	5.93	5.86	5.99	6.24	5.93	6.10	5.70	5.97	5.98	6.07
T14	**7.92**	**7.77**	**7.53**	**7.65**	**7.65**	**7.12**	**7.11**	**6.93**	**7.57**	**7.42**	**7.50**	**6.95**	**7.45**	**7.02**	**6.24**	**7.58**	**7.63**	**7.36**
T16	6.39	6.41	6.44	6.49	6.43	6.13	6.31	6.37	6.35	6.43	6.56	6.11	6.46	6.18	6.06	6.33	6.46	6.38
T17	5.23	5.15	5.22	5.10	5.09	5.40	5.16	5.37	5.24	5.09	5.18	5.30	5.30	5.31	5.98	5.22	5.11	5.28
T18	5.55	5.67	5.71	5.69	5.60	5.92	6.31	6.07	5.82	5.82	5.77	6.34	5.78	6.29	6.15	5.79	5.69	5.91
T20	6.43	6.39	6.31	6.38	6.41	6.55	6.35	5.92	6.46	6.43	6.33	6.50	6.56	6.46	6.55	6.45	6.43	6.48
T21	5.43	5.53	5.50	5.53	5.58	5.25	5.33	5.35	5.45	5.54	5.56	5.35	5.33	5.25	6.34	5.47	5.52	5.38
T24	5.00	4.97	5.18	5.06	5.06	5.10	4.99	4.77	5.26	5.08	5.15	5.26	5.31	5.15	6.57	5.22	5.07	4.97
T25	5.2	5.40	5.34	5.50	5.53	5.43	5.58	5.85	5.30	5.44	5.28	5.40	5.28	5.35	6.88	5.36	5.51	5.51
T26	5.06	4.98	4.81	4.93	4.96	4.99	5.02	5.35	4.79	4.97	4.92	4.60	4.84	4.86	7.50	4.78	4.94	4.58
T11*	7.04	6.82	6.06	6.87	6.57	6.84	6.75	6.90	6.61	6.59	6.61	6.69	6.57	6.75	6.66	6.63	6.58	6.58
T15*	7.19	6.92	6.00	7.20	6.97	7.33	7.39	6.83	7.14	7.20	6.90	7.26	7.06	7.25	6.91	7.17	7.19	7.12
T19*	6.27	6.54	5.94	6.16	6.47	5.68	5.81	5.65	6.19	6.04	6.24	5.67	6.20	5.70	6.27	6.19	6.08	6.30
T22*	5.45	5.82	5.75	6.27	6.13	6.47	6.38	6.24	6.35	6.35	6.48	6.73	6.40	6.48	6.38	6.32	5.29	5.73
T23*	5.9	5.82	5.73	6.62	6.13	6.64	6.69	6.48	7.23	7.17	7.41	6.53	7.15	6.59	7.15	7.13	6.09	6.46

*test set ; **T14** prototype (Template)

Pour examiner plus avant la stabilité de la relation quantitative structure-activité modélisée par CoMFA/SE et CoMSIA/SEHD pour les dérivés de la 4-phénoxypyridine dans cette étude, nous avons soumis le modèle 3D-QSAR proposé à un test de Y-randomisation pour cinq itérations randomisées (**Tableau III.B4**). D'après les résultats du test de Y-randomisation, les paramètres ($R^2_{yrand}<R^2_{original}$, $Q^2_{yrand}<Q^2_{original}$, et ${}^{C}R^2_{p}>0.5$) indiquent que les modèles originaux CoMFA et CoMSIA/SEHD sont très stables par rapport aux élaborés aléatoirement. Brièvement, tous les paramètres PLS statistiques obtenus à partir des analyses CoMFA et CoMSIA indiquent une forte aptitude des descripteurs moléculaires CoMFA/SE et CoMSIA/SEHD à décrire parfaitement la relation structure-activité pour le dérivé de la 4-phénoxypyridine. Après avoir validé la pertinence du modèle 3D-QSAR proposé, basé sur les descripteurs CoMFA/SE et CoMSIA/SEHD, l'étape de génération d'hypothèses pharmacopées dans l'espace 3D s'ensuit. A ce stade, nous décodons les propriétés de champ favorables et défavorables de CoMFA/SE (**Figure III.B3(c)**) et CoMSIA/SEHD (**Figure III.B3(d)**) dans la perspective d'améliorer l'activité biologique inhibitrice des dérivés de la 4-phénoxypyridine vis-à-vis du récepteur de la tyrosine kinase c-Met.

Tableau III.B4. Test de Y-randomisation.

		CoMFA			CoMSIA/SEHD		
Modèle		Q^2_{yrand}	R^2_{yrand}	$^cR^2_p$	Q^2_{yrand}	R^2_{yrand}	$^cR^2_p$
Original		**0.680**	**0.960**	-	**0.630**	**0.950**	-
Aléatoire	I1	0.440	0.517	0.652	-0.662	0.384	0.730
	I2	0.113	0.374	0.750	-0.587	0.453	0.684
	I3	-0.215	0.438	0.708	-0.225	0.528	0.630
	I4	-0.108	0.487	0.674	0.208	0.638	0.542
	I5	-0.216	0.335	0.775	-0.419	0.376	0.735
Moyenne		0.003	0.430	**0.712**	-0.337	0.476	**0.664**

La **Figure III.B3(a,b)** montre la distribution idéale des valeurs pIC_{50} observées *in vitro* par rapport aux valeurs prédites par le modèle 3D-QSAR basé sur les descripteurs CoMFA/SE et CoMSIA/SEHD développés dans cette étude. La **Figure III.B3(c,d)** présente des visualisations 3D des cartes de contour de CoMFA/SE et CoMSIA/SEHD sur la conformation 3D bioactive **T14**.

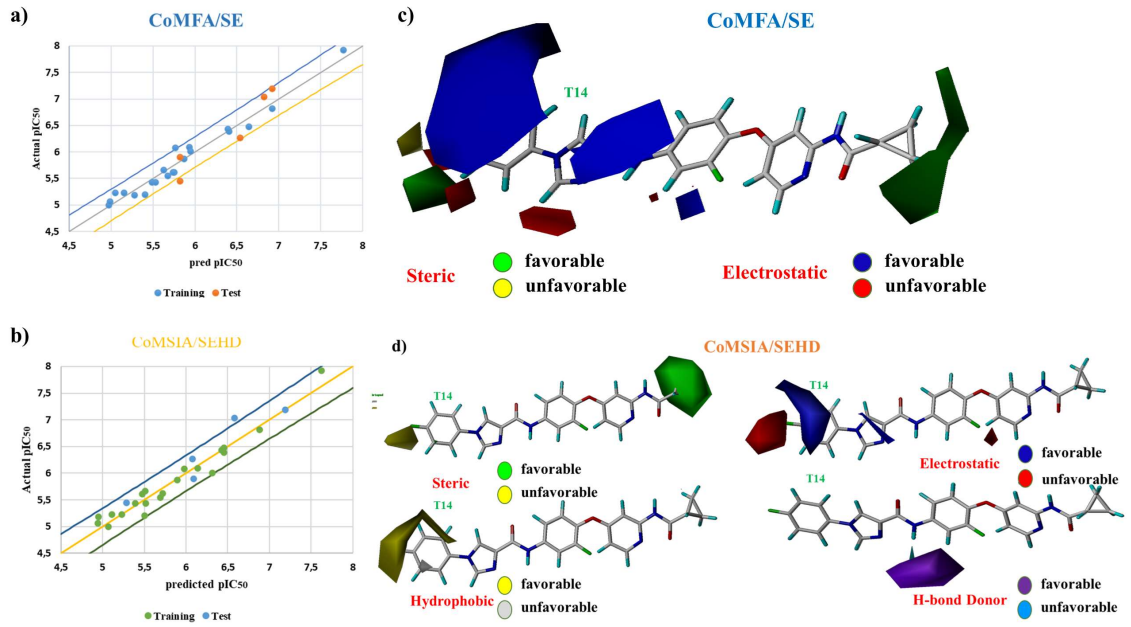


Figure III.B3. (a, b) Valeurs des pIC_{50} (*in vitro*) par rapport à celles prédites par les modèles 3D-QSAR (*in silico*), (c, d) Cartes de contours générées à partir des descripteurs CoMFA/SE et CoMSIA/SEHD utilisés par le modèle 3D-QSAR.

À partir de la **Figure III.B3(c,d)**, nous pouvons constater que les contours verts contribuent significativement plus que les contours jaunes à la structure de la molécule T14 aux sites cyclopropyle (**R1**) et 1-(4-fluorophényle) (**R2**) respectivement, ce qui indique que le renforcement de l'activité biologique des dérivés de la 4-phénoxypyridine à base d'imidazole-4-carboxamide vis-à-vis de l'activité enzymatique de la c-Met est possible par la présence de grands radicaux aux régions **R1** et **R2**. Cette hypothèse peut être confirmée par les données obtenues *in vitro* (**Tableau III.B1**), qui montrent que la présence du groupe (cyclopropyle) dans T14 donne un pIC_{50} = 7.92, tandis que le remplacement de ce groupe fonctionnel par le méthyle (T23) entraîne une diminution significative de l'activité (pIC_{50} = 5.90). De même, la substitution du groupe fonctionnel 1- (4-fluorophényle) en **T14** par un atome de chlore en **T15** a entraîné une diminution du pouvoir bioactif (pIC_{50} = 7.19). Nous pouvons également remarquer que la substitution d'atomes non halogénés dans le site para de l'anneau phényle (**R2**) a également

entraîné une diminution de l'activité biologique des dérivés de la 4-phénoxypyridine. Ensuite, les propriétés électrostatiques des molécules étudiées indiquent que les contours bleus sont dominants près de **R2** et des motifs imidazole-4-carboxamide, ce qui implique que l'ajout d'un groupe fonctionnel donneur d'électrons (+I) peut améliorer l'activité biologique de ces molécules. En outre, le contour rouge près de l'atome de fluor (**R2**) implique que l'insertion de groupements électro-attractifs (-I) dans le site **R2** peut favoriser l'inhibition de l'activité enzymatique de la c-Met par les dérivées de 4-phenoxypyridine. Cela peut être confirmé également par le **Tableau III.B1**, dans lequel la substitution de l'atome de fluor dans le site R2 par le groupement (OCH3) conduit à une diminution du pIC_{50} dans le composé T24 (5.00) par rapport au composé T23 (pIC_{50}=5.90). En outre, le contour jaune près de R2 est dominant par rapport au contour blanc, ce qui indique que le groupe 1-(4-fluorophényle) a des propriétés hydrophobes qui empêchent la liaison entre les molécules d'eau et les sites actifs dans la poche protéique de la protéine c-Met. Enfin, la prédominance du contour violet sur le cyan à proximité du groupe carboxamide suggère que l'ajout d'un groupement donneur de liaisons hydrogène à l'imidazole-4-carboxamide est favorable au renforcement des interactions des ligands biactifs au sein de la poche active de la protéine c-Met et donc à l'amélioration de l'activité biologique des molécules examinées.

3.1.2 Optimisation de nouveaux dérivés de la Phénoxypyridine comme inhibiteurs optimaux de c-Met

La **Figure III.B4** synthétise les principaux sites de pharmacopée et les propriétés 4D du champ de force des dérivés de la 4-phénoxypyridine favorables à l'amélioration de leur activité biologique contre la prolifération de la lignée cellulaire cancéreuse provoquée par l'activité enzymatique de c-Met, déterminés dans le cadre de cette étude.

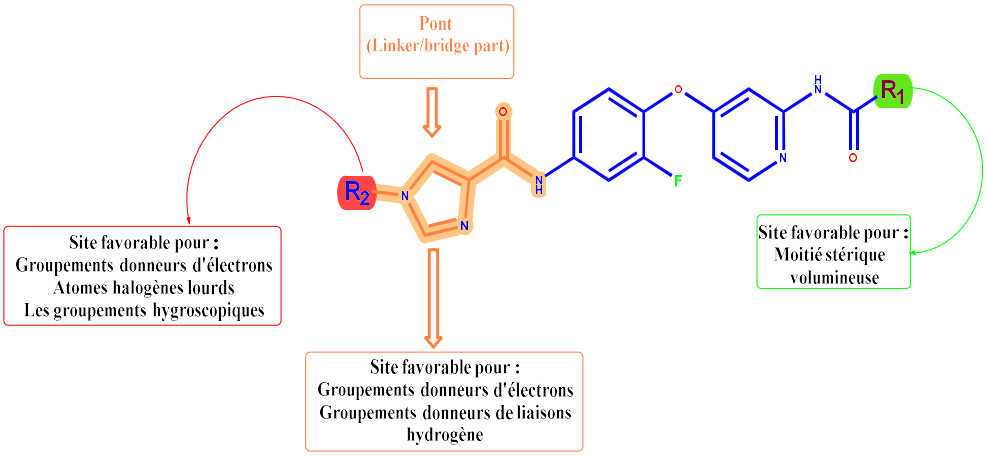


Figure III.B4. Caractéristiques des sites de pharmacopées favorables à l'optimisation de l'activité inhibitrice des dérivés de la 4-phénoxypyridine vis-à-vis de l'activité enzymatique de la protéine c-Met.

Sur la base des résultats des analyses de cartes de contour du modèle 3D-QSAR, nous avons constaté que les dérivés de la 4-phénoxypyridine contenant une partie de liaison de l'imidazole-4-carboxamide présentaient une activité inhibitrice plus forte vis-à-vis de la c-Met. Les caractérisations structurelles optimales obtenues nous ont aidés à concevoir rétrospectivement trois nouvelles séries de 46 dérivés de la 4-phénoxypyridine et à tester leurs activités biologiques inhibitrices vis-à-vis de la

protéine c-Met dans le cadre d'une étude *in silico*. La **Figure III.B5** présente la voie de rétro-synthèse proposée pour désigner les trois groupes I (P27-P41), II (P42-P56) et III (P57-P72) de dérivés de la 4-phénoxypyridine candidats au criblage *in silico* en tant que nouveaux inhibiteurs spécifiques de type II de la protéine c-Met.

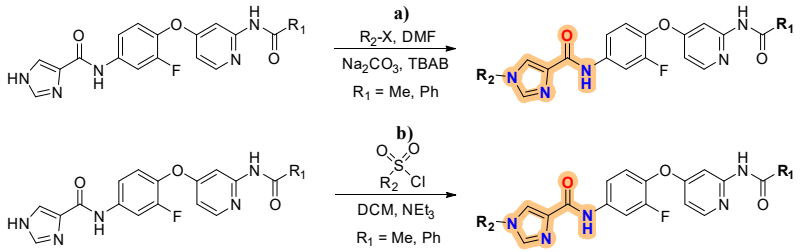


Figure III.B5. Voies de réactions rétro-synthétiques des inhibiteurs de c-Met proposés, (a) : ensembles I et II (P27-PP56), (b) : ensemble III (P57-P72).

Les données structurelles des molécules dans l'espace 2D pour les nouvelles molécules conçues (P27-P72) sont disponibles dans le fichier électronique des matériels supplémentaires de notre article publié. Pour prévoir l'activité biologique inhibitrice de la nouvelle série conçue (I-III) vis-à-vis de c-Met, les structures 3D de ces composés (P27-P72) sont préparées dans les mêmes conditions et méthodologies appliquées à la construction du modèle 3D-QSAR. La **Figure III.B6** représente les conformations des trois nouveaux ensembles après leur alignement avec la molécule de référence T14.

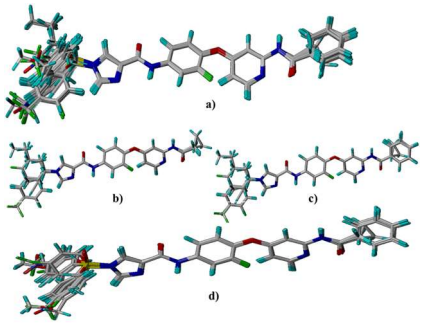


Figure III.B6. (a) Alignement de l'ensemble de 46 (P27-P72) nouveaux dérivés de 4-phénoxypyridine par rapport à la structure de référence T14. (b) Ensemble I (P27-P41) aligné avec T14. (c) Ensemble II (P42-P56) aligné avec T14. (d) Ensemble III aligné avec T14 (P57-P72).

Le **Tableau III.B5** résume les prédictions du modèle 3D-QSAR (CoMFA/SE et CoMSIA/SEHD) associées aux activités biologiques inhibitrices (pIC_{50}) des nouveaux ensembles des inhibiteurs de la c-Met. Les prédictions des modèles 3D-QSAR présentées dans le **Tableau III.B5** indiquent que l'activité biologique inhibitrice des nouvelles molécules proposées (P27-P72) est considérablement améliorée par rapport à celle observée *in vitro* (T1-T26). Les valeurs moyennes ($\overline{pIC_{50}}$) des inhibiteurs de c-Met proposés (ensembles I, II et III) ont amélioré de +2.46 (($\overline{IC_{50}}$)/-3.45 nM) par rapport à ceux synthétisés sur la base de la 4-phénoxypyridine (T1-T26). Les molécules P30 (pIC_{50}=8.114/CoMFA et pIC_{50}=8.130/CoMSIA) et P38 (pIC_{50}=8.028/CoMFA et pIC_{50}=7.967/CoMSIA) de l'ensemble I ont montré des pIC_{50} plus élevés que ceux observés pour la molécule de référence T14 (pIC_{50}=7.92). Dans l'ensemble III, les molécules P70 (pIC_{50}=7.829/CoMFA et pIC_{50}=7.802/CoMSIA) et P72 (pIC_{50}=7.798/CoMFA et pIC_{50}=7.893/CoMSIA) ont montré un pIC_{50} très proche de celui de la molécule de référence T14. Il ressort que la totalité des molécules de l'ensemble II ont montré une activité biologique nettement plus élevée que celle observée pour la molécule de référence T14. Cela signifie

que les composés de l'ensemble II peuvent être biologiquement plus actifs en termes d'inhibition de l'activité enzymatique c-Met que les molécules synthétisées *in vitro* (T1-T26). L'augmentation remarquable des valeurs des activités biologiques inhibitrices des molécules proposées peut être expliquée par le fait que l'activité biologique des dérivés de la 4-phénoxypyridine est fortement liée aux propriétés stériques, électrostatiques, hydrophobes et donneuses de liaisons hydrogène de ce type de molécules. Par conséquent, ces propriétés peuvent être considérées comme des caractéristiques structurelles optimales pour la conception de nouveaux inhibiteurs de c-Met de classe II à base de 4-phénoxypyridine. Par conséquent, les prédictions du modèle 3D-QSAR (CoMFA/SE et CoMSIA/SEHD) *in silico* des activités biologiques inhibitrices pIC_{50} des nouvelles molécules proposées ont abouti à la sélection de dix-neuf molécules en tant que têtes des séries I, II et II ayant une activité plus forte que celle de la molécule de référence **T14**.

Tableau III.B5. Les valeurs de pIC_{50} calculées *in silico* pour la série de dérivés de 4-phénoxypyridine conçus dans cette étude sur la base des modèles CoMFA/SE et CoMSIA/SEHD.

	pIC_{50}			pIC_{50}			pIC_{50}	
I	CoMFA /SE	CoMSIA/ SEHD	**II**	CoMFA /SE	CoMSIA/ SEHD	**III**	CoMFA /SE	CoMSIA/ SEHD
P27	7.577	7.530	***P42***	**8.758**	**8.847**	*P57*	7.460	7.408
P28	7.340	6.498	***P43***	**8.283**	**8.799**	*P58*	7.071	7.667
P29	7.477	7.398	***P44***	**8.371**	**8.704**	*P59*	6.929	6.767
P30	**8.114**	**8.130**	***P45***	**8.214**	**8.468**	*P60*	7.143	6.873
P31	7.716	6.907	***P46***	**8.659**	**8.201**	*P61*	7.279	7.817
P32	6.506	6.724	***P47***	**8.670**	**8.020**	*P62*	7.178	6.885
P33	7.396	7.311	***P48***	**8.453**	**8.637**	*P63*	7.161	6.912
P34	7.519	7.467	***P49***	**8.426**	**8.792**	*P64*	7.629	6.941
P35	7.490	7.417	***P50***	**8.397**	**8.743**	*P65*	7.637	7.615
P36	7.479	7.392	***P51***	**8.287**	**8.647**	*P66*	7.764	7.599
P37	7.435	7.439	***P52***	**8.360**	**8.769**	*P67*	7.490	7.731
P38	**8.028**	**7.967**	***P53***	**8.958**	**8.982**	*P68*	7.709	7.889
P39	7.516	7.153	***P54***	**8.483**	**8.494**	*P69*	7.843	7.791
P40	7.342	7.507	***P55***	**8.185**	**8.759**	***P70***	**7.829**	**7.802**
P41	7.214	7.654	***P56***	**8.200**	**7.976**	*P71*	6.864	6.789
-	-	-	-	-	-	***P72***	**7.798**	**7.893**

Référence : T14 (pIC_{50}=7.92)

À cette occasion, avant de considérer que les 19 molécules restantes peuvent être considérées comme des inhibiteurs spécifiques de c-Met, il faut d'abord savoir comment elles se comportent sur le plan pharmacocinétique et pharmacodynamique. Ceci a pour but de guider la stratégie de sélection des meilleurs candidats (H2L) pour la conception de nouveaux médicaments.

3.2 Exploration *in silico* fondée sur les prédictions Drug-like et ADME-Tox

3.2.1 Analyse des propriétés de biodisponibilité orale

Pour effectuer ce criblage, nous soumettons les 19 nouvelles molécules à des analyses *in silico* destinées à évaluer leur biodisponibilité, leur pharmacocinétique et leur pharmacodynamie. Les molécules candidates à la conception de médicaments sont tirées à partir des modèles Drug-like (**Tableau III.B6**) et ADME-Tox (**Tableau III.B7**) que nous avons élaborés dans le cadre de cette étude. À la suite du modèle prédictif de type Drug-like indiqué dans le **Tableau III.B6**, parmi les dix-neuf molécules considérées, seules neuf molécules (**P42**, **P43**, **P44**, **P45**, **P46**, **P54**, **P55**, **P56** et **P30**) présentent une excellente biodisponibilité, proche de celle des médicaments réels. Cependant, la molécule de référence T14 et le médicament standard Foretinib, ainsi que le reste des autres

composés, présentaient plusieurs problèmes de la biodisponibilité. Pour renforcer la validité de ces résultats, nous soumettons les mêmes données d'échantillons à un autre test *in silico* qui intègre les critères ADME-Tox (**Tableau III.B7**). L'objectif est de réduire la possibilité d'adopter des résultats faux/erronés dans d'autres futurs scénarios.

Tableau III.B6. Profil "Drug-like" des dérivés de la 4-phénoxypyridine proposés et du prototype (T14) et du médicament standard Foretinib.

	Propriété/règle	MW(Da) <500	LogP ≤5	n-RB <10	n-HBA <10	n-HBD ≤5	TPSA (A^2) <140
Références	**T14**	475.455	**4.9386**	**7**	**6**	**2**	**198.147**
	Foretinib	632.65	5.79	14	10	**2**	**111.25**
III	P70	592.008	5.6045	8	8	2	140.66
	P72	625.56	5.9699	8	8	2	140.66
II	**P42**	417.4	4.2407	6	5	3	109.00
	P43	431.427	4.2511	6	6	2	98.14
	P44	445.454	4.734	7	6	2	98.14
	P45	459.481	5.1241	8	6	2	98.14
	P46	473.508	5.5142	9	6	2	98.14
	P47	487.535	5.9043	10	6	2	98.14
	P48	507.525	5.7624	8	6	2	98.14
	P49	525.515	5.9015	8	6	2	98.14
	P50	541.97	6.4158	8	6	2	98.14
	P51	586.421	6.5249	8	6	2	98.14
	P52	525.515	5.9015	8	6	2	98.14
	P53	575.522	6.7812	8	6	2	98.14
	P54	457.465	4.9001	8	6	2	98.14
	P55	457.465	4.9001	8	6	2	98.14
	P56	460.469	3.6728	8	7	3	124.1
I	**P30**	397.41	3.8302	7	6	2	98.14
	P38	513.451	5.4873	7	6	2	98.14

MW : poids moléculaire, LogP : logarithme du coefficient de partage entre N-octanol et l'eau, n-RB : nombre de liaisons rotatives, n-HBA : Nombre des accepteurs de liaisons hydrogène, n-HBD : Nombre des donneurs de liaisons hydrogène, TPSA : Surface Polaire Topologique.

3.2.2 Analyse pharmacocinétique et pharmacodynamique par la modélisation ADMET-Tox.

D'après les prédictions les plus significatives recueillies grâce aux prédictions ADME-Tox présentées dans le **Tableau III.B7**, nous pouvons constater que tous les composés examinés ont montré des profils d'absorption intéressants et ont pu être jugés aptes à atteindre une bonne absorption dans le corps humain. De plus, les molécules examinées peuvent atteindre une perméabilité élevée à travers la peau, ce qui permet à cette propriété d'offrir la possibilité d'appliquer des médicaments à travers la peau. Par ailleurs, les faibles paramètres de distribution VDss et Fu indiquent une distribution optimale des substances examinées dans l'organisme et la possibilité qu'elles ne se lient pas ou n'interagissent pas avec d'autres médicaments et n'entraînent pas de problèmes thérapeutiques majeurs. De même, les valeurs des paramètres de perméabilité de la barrière hémato-encéphalique (BHE ou « BBB ») et du système nerveux central faible (SNC) indiquent l'incapacité de ces composés à pénétrer ces deux derniers, évitant ainsi d'attaquer le cerveau et de l'affecter dangereusement. Du point de vue du métabolisme, toutes les molécules, y compris celle de référence T14 et le Foretinib, étaient des inhibiteurs du CYP3A4 et des substrats du CYP2D6, à l'exception des molécules P42, P54 et P54, qui n'avaient pas de spécificité en termes de substrats.

Tableau III.B7. Propriétés ADME-Tox prédites *in silico* pour de nouvelles molécules à base de 4-phénoxypyridine conçues à partir de la structure de base T14 et du médicament standard Foretinib.

		Absorption							**Distribution**				**Métabolisme**				**Excrétion**		**Toxicité**					
		S	Pr	Ab	SP	P-S	P-I1	P-I2	VD	Fu	BBB	CNS	CYP-Sub		CYP-Inh		TC	ROS	AMT	MTD	LD_{50}	ORC	Hep	SS
													2D6	3A4	2D6	3A4								
Réf	**T14**	-2.9	0.7	100.0	-2.7	Oui	Oui	Oui	-0.8	0.2	-0.1	-3.0	Non	Oui	Non	Oui	0.60	Oui	Non	0.1	2.3	1.5	Oui	Non
	Fore.	-4.1	0.8	97.7	-2.7	Oui	Oui	Oui	-0.7	0.2	-1.7	-3.3	Non	Oui	Non	Oui	0.94	Non	Non	0.2	2.5	1.9	Oui	Non
III	P70	-2.9	0.7	92.3	-2.7	Oui	Oui	Oui	-0.3	0.2	-1.1	-3.5	Non	Oui	Non	Oui	0.01	Oui	Non	0.3	2.5	1.6	Oui	Non
	P72	-2.9	0.7	93.5	-2.7	Oui	Oui	Oui	-0.3	0.2	-1.2	-3.5	Non	Oui	Non	Oui	-0.14	Oui	Non	0.3	2.5	1.5	Oui	Non
II	**P42**	**-2.9**	**0.6**	**86.8**	**-2.7**	**Oui**	**Non**	**Oui**	**0.3**	**0.1**	**-1.4**	**-3.5**	**Non**	**Non**	**Oui**	**Oui**	**0.58**	**Oui**	**Non**	**0.5**	**2.4**	**2.5**	**Non**	**Non**
	P43	-3.1	1.0	97.1	-2.7	Oui	Oui	Oui	0.0	0.2	-0.7	-3.2	Non	Oui	Non	Oui	0.37	Oui	Non	0.4	2.5	1.3	Oui	Non
	P44	-3.1	1.0	98.3	-2.7	Oui	Oui	Oui	0.0	0.2	-0.7	-3.2	Non	Oui	Non	Oui	0.58	Oui	Non	0.4	2.5	1.2	Oui	Non
	P45	-3.1	1.0	99.0	-2.7	Oui	Oui	Oui	0.0	0.1	-0.6	-3.1	Non	Oui	Non	Oui	0.58	Oui	Non	0.4	2.5	1.2	Oui	Non
	P46	-3.1	1.0	99.8	-2.7	Oui	Oui	Oui	0.0	0.1	-0.5	-3.1	Non	Oui	Non	Oui	0.65	Oui	Non	0.3	2.6	1.2	Oui	Non
	P47	-3.1	1.0	100.0	-2.7	Oui	Oui	Oui	0.0	0.1	-0.5	-3.1	Non	Oui	Non	Oui	0.69	Oui	Non	0.3	2.6	1.3	Oui	Non
	P48	-2.9	0.9	100.0	-2.7	Oui	Oui	Oui	-0.3	0.2	-0.1	-3.1	Non	Oui	Oui	Oui	0.41	Oui	Non	0.2	2.5	0.3	Oui	Non
	P49	-2.9	0.8	100.0	-2.7	Oui	Oui	Oui	-0.6	0.1	-0.6	-3.1	Non	Oui	Non	Oui	0.42	Oui	Non	0.3	2.6	1.4	Oui	Non
	P50	-2.9	0.7	100.0	-2.7	Oui	Oui	Oui	-0.2	0.2	-0.2	-3.0	Non	Oui	Oui	Oui	0.40	Oui	Non	0.2	2.5	0.2	Oui	Non
	P51	-2.9	0.8	100.0	-2.7	Oui	Oui	Oui	-0.2	0.2	-0.2	-3.0	Non	Oui	Oui	Oui	0.38	Oui	Non	0.2	2.5	0.2	Oui	Non
	P52	-2.9	1.0	100.0	-2.7	Oui	Oui	Oui	-0.8	0.2	-0.4	-3.0	Non	Oui	Non	Oui	0.40	Oui	Non	0.2	2.5	1.5	Oui	Non
	P53	-2.9	0.7	100.0	-2.7	Oui	Oui	Oui	-0.1	0.2	-0.5	-3.0	Non	Oui	Non	Oui	0.25	Oui	Non	0.2	2.5	0.1	Oui	Non
	P54	**-3.0**	**0.9**	**95.4**	**-2.7**	**Oui**	**Oui**	**Oui**	**0.8**	**0.2**	**-0.4**	**-3.2**	**Non**	**Non**	**Non**	**Oui**	**0.63**	**Oui**	**Non**	**0.2**	**2.4**	**0.2**	**Non**	**Non**
	P55	**-3.0**	**0.9**	**95.4**	**-2.7**	**Oui**	**Oui**	**Oui**	**0.8**	**0.2**	**-0.4**	**-3.2**	**Non**	**Non**	**Non**	**Oui**	**0.63**	**Oui**	**Non**	**0.2**	**2.4**	**0.2**	**Non**	**Non**
	P56	-3.0	0.7	88.3	-2.7	Oui	Oui	Oui	0.2	0.2	-1.5	-3.4	Non	Oui	Non	Oui	0.51	Oui	Non	0.4	2.5	1.3	Oui	Non
I	P30	-2.9	0.4	82.9	-2.7	Oui	Oui	Non	0.6	0.2	-1.3	-2.9	Non	Non	Non	Oui	0.69	Oui	Oui	0.7	2.3	1.2	Oui	Non
	P38	-2.9	0.3	89.4	-2.7	Oui	Oui	Oui	0.04	0.1	-1.1	-2.9	Non	Non	Non	Oui	0.34	Oui	Non	0.2	2.4	1.4	Oui	Non

S (solubilité dans l'eau) ; **Pr** (perméabilité caco2) ; **Ab** (absorption intestinale) ; **SP** (perméabilité de la peau) ; **PS** (Substrat de P-glycoprotéine) ; **P-I1** (Inhibiteur de P-glycoprotéine I) ; **P-I2** (Inhibiteur de P-glycoprotéine II); **VD** (Volume de distribution chez l'homme à l'état stable.) ; **Fu** (Fraction non liée) ; **BBB** (perméabilité de la barrière hémato-encéphalique) ; **CNS** (perméabilité du système nerveux central) ; **CYP-sub** (Substrat d'enzyme Cytochrome P450) ; **CYP-Inh** (Inhibiteur d'enzyme Cytochrome P450) ; **TC** (Clairance totale) ; **ROS** (Substrat du transporteur de cations organiques rénaux 2 (OCT2)) ; **AMT** (Toxicité d'AMES) ; **MTD** (Dose maximale tolérée chez l'homme) ; **LD50** (Toxicité aiguë orale pour le rat (DL50)) ; **ORC** (Toxicité chronique orale pour le rat (LOAEL)); **Hep** (Hépatotoxicité) ; **SS** (Sensibilisation de la peau).

Par ailleurs, les valeurs prédictives de l'indice de clairance totale étaient faibles pour toutes les molécules par rapport au médicament de référence, à savoir le Foretinib. Cela indique la possibilité d'une bonne stabilité des doses de médicament dans l'organisme offerte par les molécules en question. À la lumière de ces données, il est clair que les paramètres ADME correspondent aux sorties obtenues par le modèle Drug-like. Toutefois, ces analyses ne sont pas suffisantes pour donner une indication prévisionnelle sur la pharmacodynamique de ces molécules dans le corps humain ainsi que sur les risques de toxicité potentiels. Les six critères de toxicité utilisés dans le cadre de cette étude ont montré que toutes les molécules examinées, y compris le T14 et le Foretinib, présentaient une hépatotoxicité potentielle, à l'exception des molécules P42, P54 et P54. En guise de synthèse générale des sorties obtenues par le modèle ADME-Tox, nous pouvons considérer les nouvelles molécules conçues **P42**, **P54** et **P54** comme parfaites (H2L) en vue de la conception de nouveaux inhibiteurs de c-Met (**Figure III.B7**).

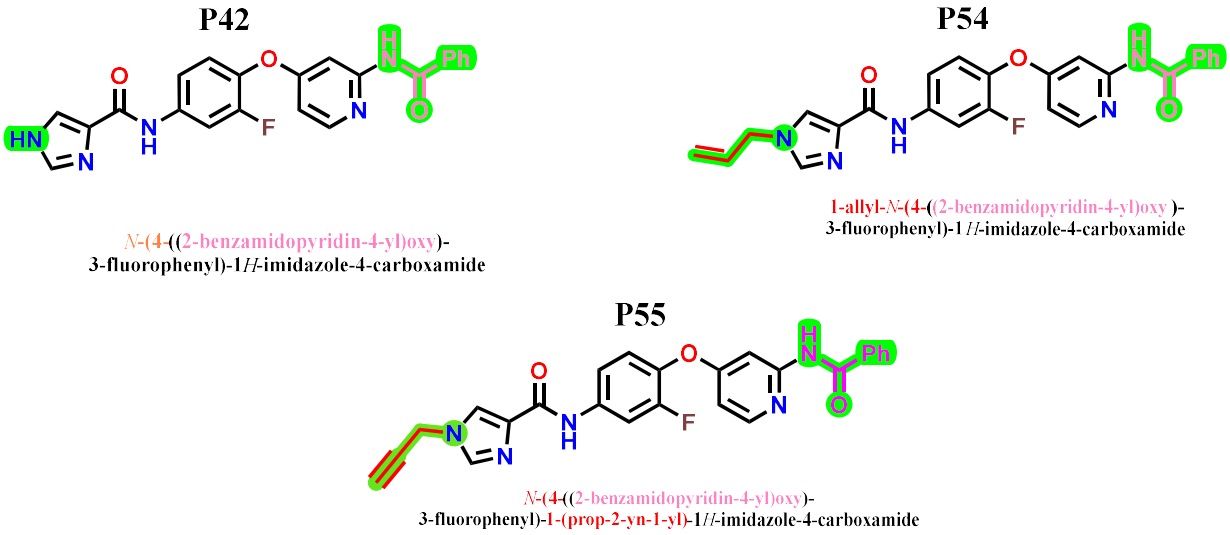


Figure III.B7. Dérivés de 4-phénoxypyridine (H2L) candidats en tant que molécules médicamenteuses potentielles *in silico*.

À cette occasion, afin de vérifier la compatibilité/l'affinité structurelle entre les trois molécules proposées **P42**, **P54** et **P55** et la protéine c-Met, nous avons soumis ces échantillons à un criblage virtuel par Docking moléculaire. Dans cette procédure de criblage supposé, nous n'avons pas exclu la molécule de référence **T14** et le médicament standard, le Foretinib, en dépit de leur toxicité potentielle. Ceci a pour but de rationaliser cette procédure et de fournir une vision plus complète sur le comportement structurel et énergétique des dérivés de la 4-phénoxypyridine vis-à-vis du récepteur c-Met.

3.3 Docking Moléculaire : Identification de candidats inhibiteurs de c-Met

3.3.1 Validation du protocole du Docking Moléculaire

Avant de soumettre les 3 molécules (H2L) retenues (**P42**, **P54**, et **P55**) accompagnées de la molécule de référence T14 à un criblage basé sur le Docking moléculaire appliqué à la protéine c-Met, il faut préalablement identifier la poche active, où les interactions c-Met-ligand sont susceptibles d'inhiber la prolifération des lignées cellulaires cancéreuses. La validité et la précision de la méthode d'ancrage moléculaire appliquée dans ce plan de test doivent également être confirmées. Pour

satisfaire ces exigences initiales, nous réalisons un protocole de re-docking moléculaire de l'inhibiteur Foretinib (GSK1363089) dans la poche active de c-Met et comparons les prédictions du modèle de docking moléculaire avec celles du modèle expérimental au sein du complexe cMet-Foretinib (code PDB : 3LQ8) [438]. Dans cette étape, nous avons effectué le re-docking moléculaire rigide du ligand Foretinib dans la poche active de c-Met dans le complexe 3LQ8. Cela afin d'accélérer la détection des résidus d'acides aminés actifs en volume de la poche active de c-Met avec lesquels le Foretinib interagit (**Figure III.B8**). Par conséquent, il est possible d'obtenir une valeur d'énergie de liaison très élevée du Foretinib avec c-Met par rapport aux 19 ligands soumis à évaluation et au ligand de référence T14 (dont les structures ont été adaptées et optimisées selon le modèle flexible). Ceci a pour but de rendre le processus de détection des aspects de la stabilité et de la compatibilité des conformations 3D des nouveaux ligands avec la poche active de c-Met plus rigoureux par rapport à la conformation de Foretinib déjà connue, qui sera simulée par la dynamique moléculaire dans la dernière étape de cette étude. La **Figure III.B8** présente les conformations superposées correspondantes de l'inhibiteur co-cristallisé (Foretinib, couleur jaune) et re-Docké (Foretinib, couleur noire) dans la poche de la protéine c-Met (RMSD=0.0128 Å, énergie de liaison de -18.6 kcal/mol).

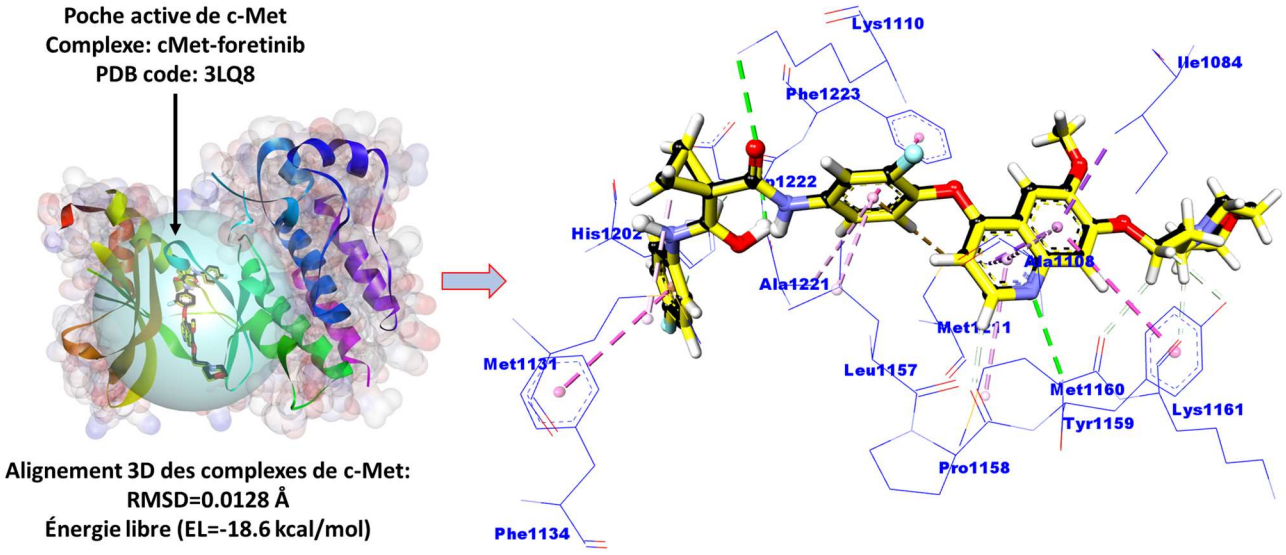


Figure III.B8. Visualisation 3D de l'alignement du complexe natif 3LQ8 et du complexe prototypique cMet-Foretinib obtenu par re-Docking moléculaire.

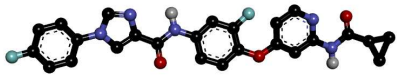
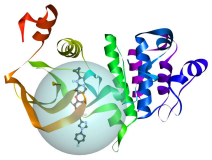
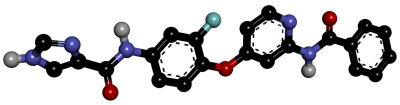
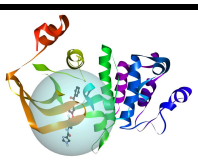
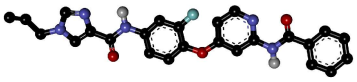
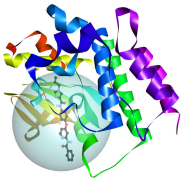
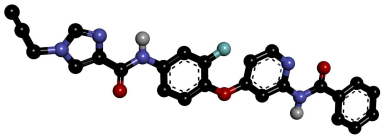
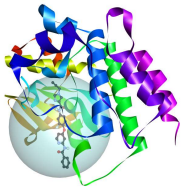
La valeur de RMSD calculée entre les conformations rigides du Foretinib original et du Foretinib re-Docké, proche de 0 Å, indique la performance élevée du modèle de Docking moléculaire que nous voulons suivre pour cribler les 19 ligands à l'intérieur de la poche du c-Met. La valeur élevée de liaison/affinité du Foretinib de -18.6 kcal/mol indique que le Foretinib re-Docké ne diverge pas de la poche active du c-Met. Par conséquent, les résidus d'acides aminés actifs avec lesquels le Foretinib interagit dans le complexe natif 3LQ8 et le complexe cMet-Foretinib modélisé peuvent être considérés comme des références pour valider la pertinence des ligands étudiés en tant que nouveaux inhibiteurs spécifiques de c-Met. Les résultats du protocole de re-Docking indiquent que le ligand Foretinib re-docké dans la poche de c-Met réalise les mêmes interactions que le Foretinib cristallisé avec c-Met dans le complexe original 3LQ8. Les résidus d'acides aminés **Lysine** (Lys1161 et Lys1110), **Isoleucine**

(Ile1084), **Methionine** (Met1211, Met1131, et Met1160), **Alanine** (Ala1211 et Ala1221), **Proline** (Pro1158), **Leucine** (Leu1157), **Phenylalanine** (Phe1223 et Phe1134), **Aspartic acid** (Asp1222), **Histidine** (His1202) sont identifiés dans cette étude comme références permettant d'évaluer et de suivre l'activité (affinité ou liaison) de nouveaux ligands de dérivées de 4-phenoxypyridine dans la poche active de la c-Met.

3.3.2 Caractérisation des conformations des dérivés de 4-phénoxypyridine adaptées à l'inhibition de c-Met

Dans cette étape, nous utilisons cette la simulation du Docking moléculaire afin de prédire l'orientation/conformation et l'affinité de liaison les plus probables des deux ligands examinés pour leur récepteur c-Met cibler. Le ligand ayant l'énergie de liaison la plus négative indique sa forte affinité et sa stabilité pour le récepteur. De fait, il a été démontré qu'une basse énergie de liaison est associée à une forte activité des ligands vis-à-vis des récepteurs. Par conséquent, les ligands ayant l'énergie de liaison la plus négative sont sélectionnés comme les molécules les plus susceptibles d'interagir et d'être stables vis-à-vis du récepteur visé au cours de la procédure de Docking moléculaire. Le **Tableau III.B8** présente les scores d'énergie/affinité de liaison et les constantes d'inhibition (Ki) prévues entre les ligands dockés et le récepteur Tyrosine Kinase c-Met visé.

Tableau III.B8. Sorties du Docking moléculaire en termes d'énergie de liaison et de constante d'inhibition estimée (Ki) des inhibiteurs de c-Met développés.

Ligand	Complexe	EL (Kcal/mol)	K_i (µM)
 **T14**	 **c-Met-T14**	-7.8	1.888
 **P42**	 **c-Met-P42**	-9.9	0.054
 **P54**	 **c-Met-P54**	-9.9	0.054
 **P55**	 **c-Met-P55**	-9.3	0.149

Grâce aux données de Docking moléculaire présentées dans le **Tableau III.B8**, nous pouvons noter que l'énergie libre de liaison de chacun des ligands conçus **P42** (-9.9 kcal/mol), **P54** (-9.9 kcal/mol) et **P55** (-9.3 kcal/mol) était plus négative que celle des ligands de référence **T14** (-7.8

kcal/mol), ce qui confirme davantage que les optimisations apportées à la structure de **T14** pour générer de nouveaux ligands ont conduit à une amélioration de l'activité biologique inhibitrice de ces ligands contre la c-Met. Ceci peut être confirmé par les valeurs K_i plus faibles que 100 µM (0.1 mM) pour **P42** (0.054 µM), **P54** (0.054 µM) et **P55** (0.140 µM) par rapport au **T14** (1.88 µM) de référence. Les molécules qui ont un K_i inférieur à 100 µM sont considérées comme des inhibiteurs très puissants du récepteur cible, tandis que celles qui ont un Ki supérieur à 100 µM ne sont pas des inhibiteurs puissants. Par conséquent, les nouveaux ligands **P42**, **P54** et **P55**, ainsi que la référence **T14**, peuvent être considérés comme de forts inhibiteurs de c-Met. Nous pouvons également remarquer que les conformations biactives des ligands P42 et P54 présentent les mêmes valeurs EL/K_i égales à (-9.9 kcal/mol-0.054 µM) vis-à-vis de la poche active du c-Met dans les complexes de têtes sélectionnés. Cela signifie que le groupe fonctionnel (but-1-ène) sur le site R2 du ligand P54 et son absence dans le ligand P42 n'ont pas d'effet sur l'énergie de liaison générée dans les complexes via les interactions non-covalentes (liaisons hydrogène, interactions de van der Waals, etc) dans les complexes cMet-inhibiteurs (cMet-P42 et cMet-P54). En revanche, la substitution du groupe fonctionnel (but-1-ène) par (but-1-yne) dans le site R2 du ligand P55 a entraîné un léger changement dans l'énergie de liaison de la conformation vers la poche active de c-Met (EL=-9.3 kcal/mol). Ces données, en général, conduisent à une relation étroite entre les conformations structurelles du récepteur cMet et les inhibiteurs proposés (**P42**, **P54** et **P55**) dans l'espace biomoléculaire 3D. En conséquence, la classe de petites molécules, les dérivés de la 4-phénoxypyridine portant le lien imidazole-4-carboxamide (pont), ont un fort potentiel pour devenir une clé structurelle prometteuse pour la conception de nouveaux inhibiteurs de c-Met de type II plus puissants et plus sûrs dans la thérapie ciblée contre le cancer. D'autres considérations concernant la compatibilité de ces structures avec la protéine c-Met peuvent être obtenues par des analyses visuelles des interactions prédites observées dans les complexes protéine-ligand obtenus par la simulation de Docking moléculaire. Le **Tableau III.B9** présente les résultats des analyses d'interaction protéine-ligand obtenues par la procédure de Docking moléculaire appliquée dans cette étude sur les ligands (P42, P54, P55, et T14) vis-à-vis de la poche active de c-Met. Grâce à l'analyse visuelle des interactions des ligands (T14, P42, P55 et P55) dans la poche active de la c-Met, nous avons remarqué que ceux-ci établissaient de nombreuses interactions non-covalentes avec la protéine c-Met, dont les plus importantes sont les interactions électrostatiques et hydrophobes, les interactions conventionnelles par liaisons hydrogène/carbone, etc. (**Tableau III.B9**). De plus, ces ligands ont établi le contact avec plusieurs résidus d'acides aminés de référence avec lesquels le Foretinib interagit dans la poche de c-Met (résidus indiqués en bleu) de même que la détection de nouveaux acides aminés (en vert). Il est également apparu que les résidus d'acides aminés Acide glutamique (Glu1127), Glycine (Gly1163, Gly1224, Gly1127, et Gly1128), Valine (Val1092, et Val1151) et Tyrosine (Tyr1159) contribuaient fortement aux interactions des nouveaux inhibiteurs de la c-Met que nous proposons dans cette étude. Par conséquent, ces acides aminés révélés dans cette étude (couleur verte) peuvent être considérés comme les sites actifs spécifiques contribuant à l'inhibition de l'activité enzymatique de la c-Met par les dérivés de la 4-phénoxypyridine que nous avons optimisés par voie computationnelle dans cette étude. Ensuite, nous avons soumis la base de données de petites molécules bioactives optimisées dans cette étude à un ensemble de rigoureux tests *in silico*, qui ont abouti à la sélection de

P42, P55 et P55 en tant que principales structures aptes à la conception de nouveaux médicaments pour lutter contre le cancer. Comme dernière étape post-Docking qui fait partie de la routine de validation et de confirmation des hypothèses et des conclusions, nous effectuons des simulations de la dynamique moléculaire sur un temps moyen (100 ns) pour simuler le comportement structurel et dynamique de la base de données optimisées et de la référence.

3.4 Validation rétrospective des résultats à l'aide des simulations de la Dynamique Moléculaire

Nous avons effectué des tests de dynamique moléculaire sur trois échantillons : T14 (modèle synthétique d'inhibiteur de c-Met), P55 (modèle de design de nouveaux inhibiteurs de c-Met) et Forotinib (inhibiteur de c-Met utiliser comme contrôle positif). La stabilité des systèmes est évaluée en suivant les valeurs de RMSD sur une période de 100 ns. La **Figure III.B9** montre que les valeurs RMSD pour tous les systèmes sont restées inférieures à 0.35 nm et ont augmenté régulièrement jusqu'à ce qu'elles atteignent un état stable après 10 ns. Les fluctuations des valeurs RMSD pour T14 enregistrées entre 40 ns et 60 ns étaient relativement légères. Cela indique que les complexes formés entre les ligands (P55 et T14) et la cible (c-Met) étaient stables tout au long de la simulation. La forte convergence entre les valeurs de RMSD obtenues pour le c-Met libre et complexé avec les ligands indique la grande stabilité des ligands étudiés dans la poche active du c-Met. Pour mémoire, dans cette étude, nous avons effectué des simulations de dynamique moléculaire via le logiciel open source Gromacs-2018, en suivant la stratégie générale de dynamique moléculaire décrite dans le deuxième chapitre de cette thèse.

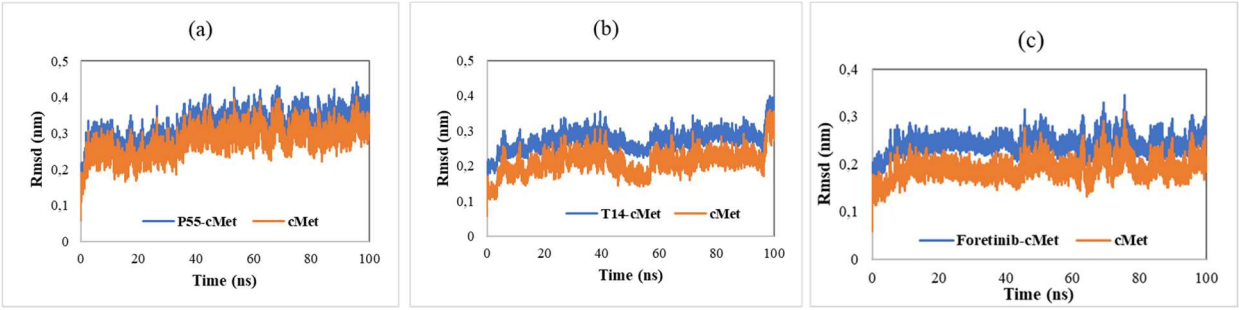


Figure III.B9. (a) Échelles de temps de l'écart quadratique moyen (RMSD) des atomes du squelette de la protéine cMet libre et complexée avec les ligands P55 (a), T14 (b) et Foretinib (c).

Dans cette étude, la fluctuation quadratique moyenne (RMSF) des atomes C-alpha de tous les acides aminés de la protéine c-Met est calculée afin d'évaluer l'effet des composés étudiés sur la flexibilité de la protéine (**Figure III.B10**). Les résultats obtenus ont montré que les profils de RMSF des complexes étudiés étaient similaires à ceux de la protéine libre. De plus, la plupart des résidus avaient des valeurs RMSF inférieures à 0.2 nm, ce qui indique la grande stabilité des systèmes étudiés. Le composé P55 a interagi avec neuf sites actifs connus pour inhiber l'activité de c-Met, et les fluctuations élevées observées dans certains résidus allant de 1233 à 1262 n'ont pas eu d'effet sur la stabilité globale.

Tableau III.B9. Interactions protéine-ligand identifiées par le modèle de Docking moléculaire.

Poche active : Lysine (Lys1161 et Lys1110), Isoleucine (Ile1084), Methionine (Met1211, Met1131, et Met1160), Alanine (Ala1211 et Ala1221), Proline (Pro1158), Leucine (Leu1157), Phenylalanine (Phe1223 et Phe1134), Aspartic acid (Asp1222), Histidine (His1202)

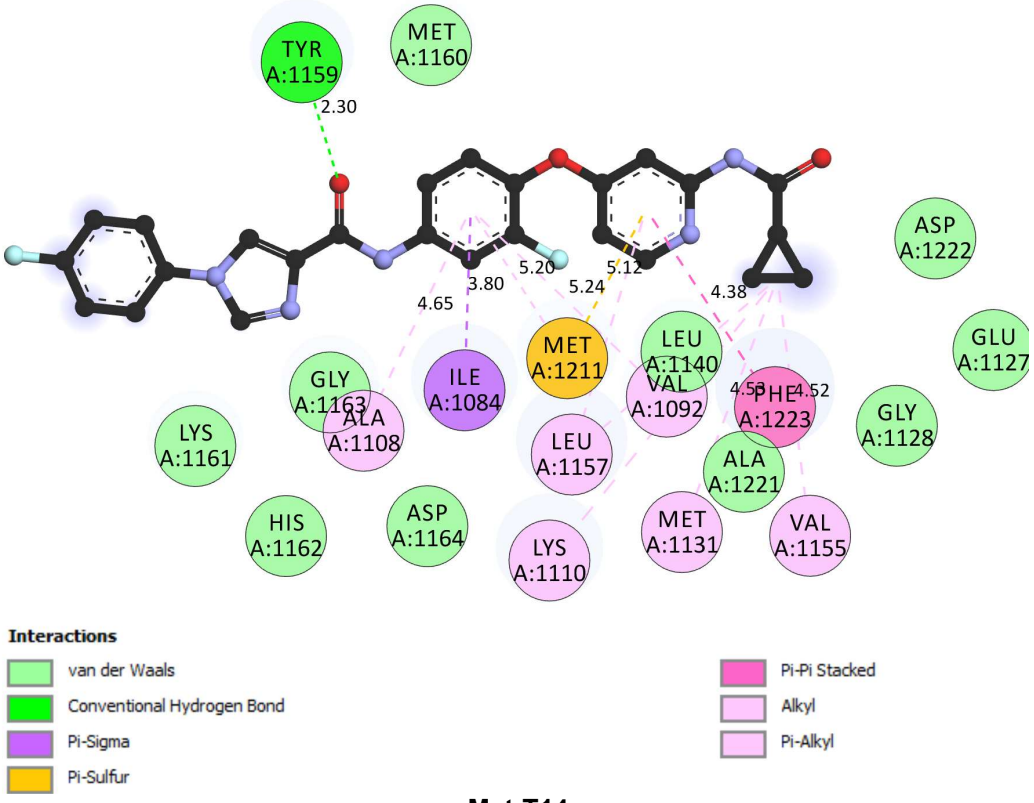
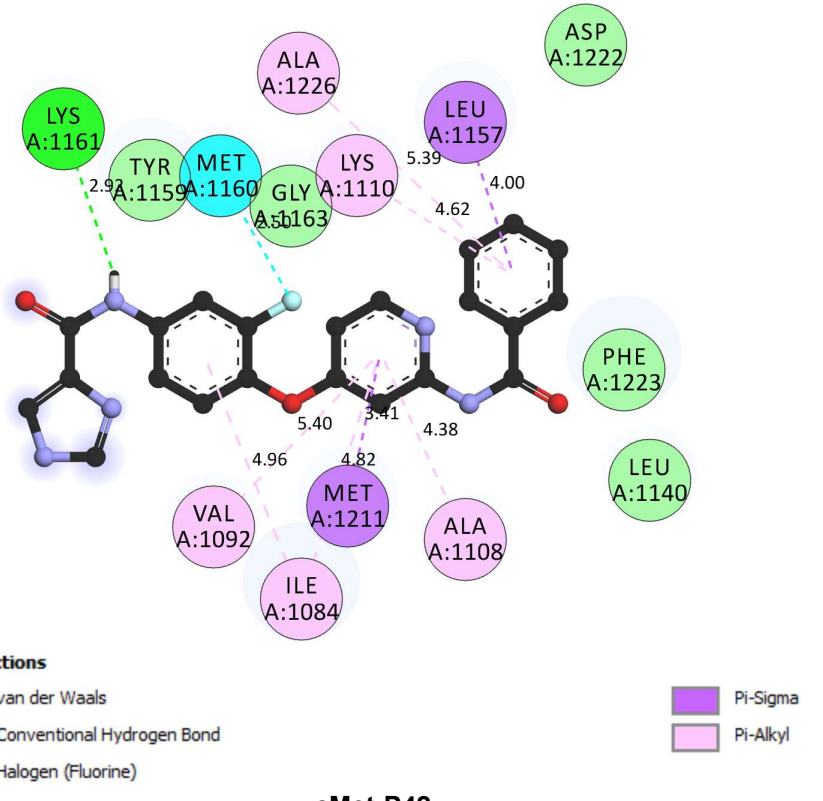
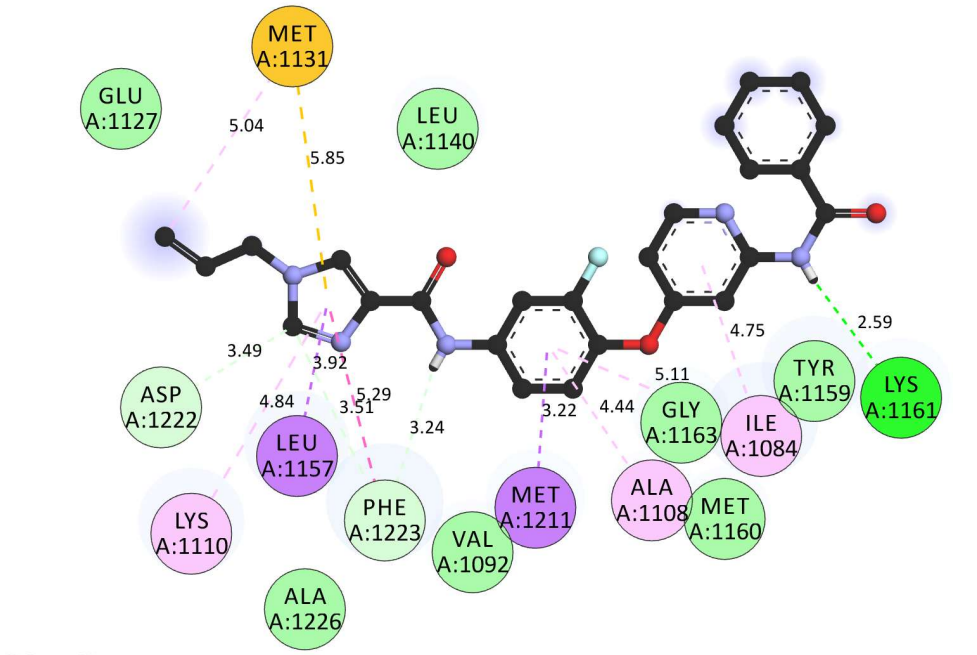
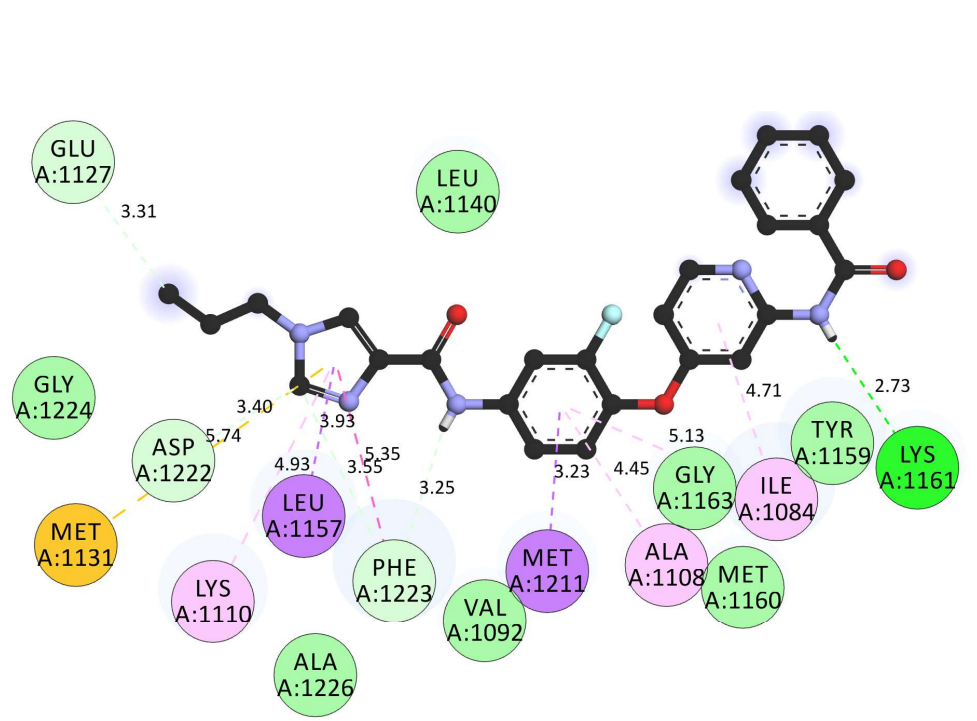
cMet-inhibiteur Interactions	Interactions générées
 **cMet-T14**	 **cMet-P42**
Tyr1159, Met1160, Lys1161, **Gly1163**, His1162, Ala1108, Asp1164, **Ile1084**, **Met1211**, Leu1157, **Lyss1110**, Leu1140, Val1092, **Met1131**, **Phe1223**, **Ala1221**, **Val1151**, **Gly1128**, **Asp1222**, **Gly1127**	**Lys1161**, **Tyr1159**, **Val1092**, **Met1160**, Ala1226, **Gly1163**, **Val1092**, **Ile1084**,, **Lys1110**, **Met1211**, , **Leu1157**, Ala1108, Asp1222, **Phe1223**, Leu1140

Tableau III.B10. Suite du tableau.

 **cMet-P54**	 **cMet-P55**
Glu1127, **Asp1222**, **Lys1110**, **Met1131**, **Leu1157**, Ala1226, **Phe1223**, Leu1140, **Val1092**, **Met1211**, Ala1108, **Gly1163**, **Met1160**, **Ile1084**, **Tyr1159**, **Lys1161**	**Glu1127**, **Gly1224**, **Met1131**, **Asp1222**, **Lys1110**, **Leu1157**, Ala1226, **Phe1223**, Leu1140, **Val1092**, **Met1211**, **Gly1163**, Ala1108, **Ile1084**, **Met1160**, **Tyr1159**, **Lys1161**
Poche active : Lysine (Lys1161 et Lys1110), Isoleucine (Ile1084), Methionine (Met1211, Met1131, et Met1160), Alanine (Ala1211 et Ala1221), Proline (Pro1158), Leucine (Leu1157), Phenylalanine (Phe1223 et Phe1134), Aspartic acid (Asp1222), Histidine (His1202)	

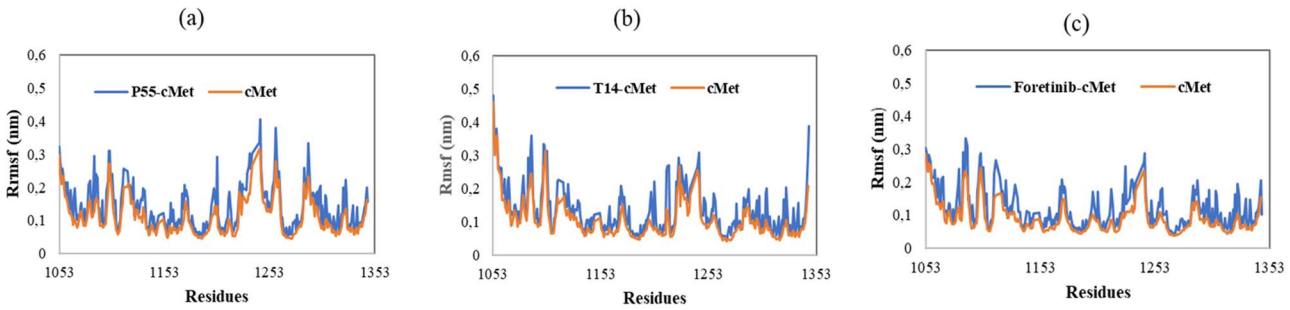


Figure III.B10. (a) Échelles de temps de la fluctuation quadratique moyenne (RMSF) des atomes Cα du squelette de la protéine cMet libre et complexée avec les ligands P55 (a), T14 (b) et Foretinib (c).

Cette étude porte également de façon générale sur les effets de certaines simulations MD sur le comportement de la molécule et sur l'évolution des paramètres. Le rayon de giration (Rg) n'a pas subi de changements conformationnels significatifs et ses valeurs moyennes étaient inférieures à 2 nm (**Figure III.B11a**). Le nombre de liaisons hydrogène entre les composés et le cMet était le plus élevé pour le P55 (**Figure III.B11b**) et la surface accessible (SASA) pour tous les systèmes était d'environ 150 nm^2 (**Figure III.B11c**). Ces résultats indiquent que tous les systèmes étudiés sont stables pendant la période de simulation, bien que le T14-cMet présente des variations proportionnellement plus marquées que pour P55-cMet. Au total, tous les aspects dynamiques abordés dans ce travail indiquent que le squelette de la C-Met n'est pas perturbé par son interaction avec les échantillons de ligands basés sur la Phénoxypyridine que nous avons examinés, ce qui indique la stabilité et la compatibilité structurelle possibles entre les deux parties : protéine/inhibiteurs. Par conséquent, les molécules optimisées dont on a indiqué dans cette étude qu'elles étaient des inhibiteurs de type II prometteurs pour c-Met doivent faire l'objet de recherches, de tests et de futures optimisations afin de contribuer à la conception d'un nouveau médicament contre le cancer basé sur les dérivés de la Phénoxypyridine portant un échafaudage d'imidazole-4-carboxamide.

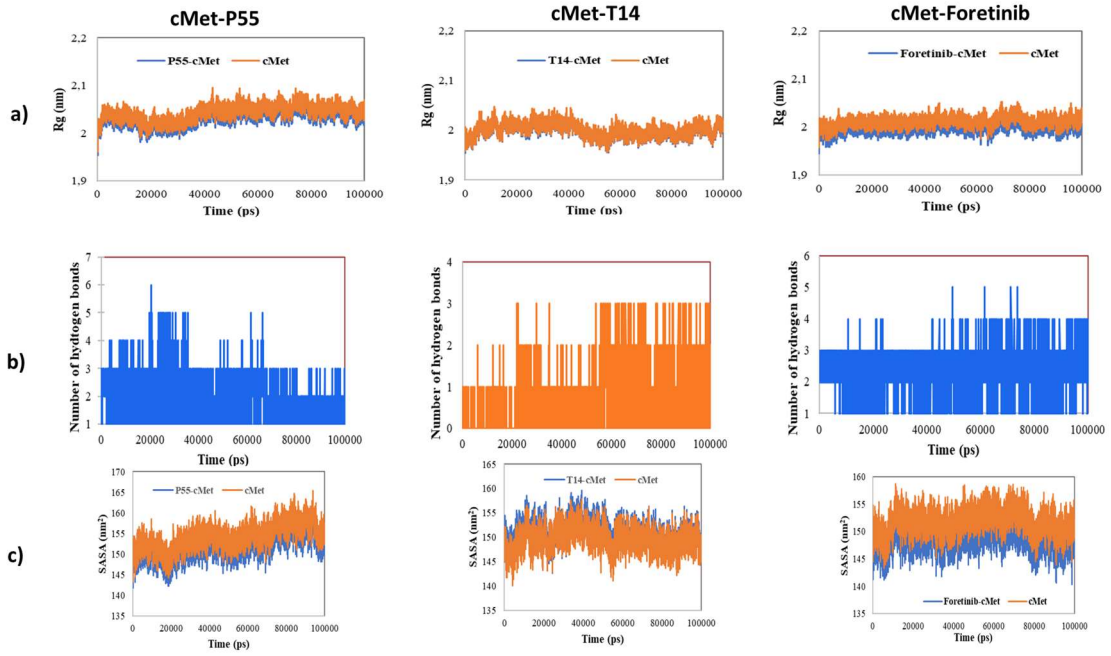


Figure III.B11. Rayon de giration (Rg) (a), Nombre de liaisons hydrogènes établies (b) et Surface accessible au solvant (c) des systèmes cMet non complexes et complexés (cMet-P55, cMet-T14, et cMet-Foretinib.

4 Conclusion

Le cancer reste une véritable menace mondiale, ce qui nécessite des recherches accrues pour développer les traitements chimio thérapeutiques les plus efficaces avec un minimum d'effets secondaires. Dans cette étude, nous avons mené une étude *in silico* sur la capacité des molécules dérivées de la 4-phénoxypyridine portant deux motifs imidazole-4-carboxamide et 1,2,4-triazole-3-carboxamide, pour inhiber l'activité enzymatique de la c-Met. Des modèles 3D-QSAR de type CoMFA et CoMSIA sont développés afin de comprendre les données pharmacologiques moléculaires associées à l'activité biologique inhibitrice pIC_{50} des 26 composés vis-à-vis de la protéine c-Met. Ensuite, après avoir appliqué de multiples scénarios de criblage *in silico*, notamment Drug-like, ADME-Tox, Docking moléculaire et simulations de dynamique moléculaire en vue d'une validation post-Docking, trois composés H2L, à savoir P42, P54 et P55, se sont révéler capables de présenter une activité biologique pIC_{50} estimée forte, une forte aptitude à se lier au site active de la protéine c-Met, la pharmacocinétique et dynamique favorables et une toxicité potentielle réduite par rapport à l'un des composés synthétisés (T14) et au médicament Foretinib. Cette étude met en évidence le potentiel de ces trois molécules dans le développement d'un nouveau médicament anticancer à faible toxicité. Il est recommandé de poursuivre les recherches sur leur capacité à cibler d'autres protéines associées à la croissance des cellules cancéreuses. Les résultats de cette étude complètent également les résultats de nos études précédentes et à venir relatives à la mise au point de nouvelles générations d'inhibiteurs spécifiques de la récepteur tyrosine kinase c-Met basés sur de petites molécules biologiquement actives contre la croissance des cellules cancéreuses malignes. Ce panorama de l'association de l'activité enzymatique anormale de la protéine c-Met avec le développement et la croissance de nombreuses cellules cancéreuses mortelles, en particulier le cancer du poumon non à petites cellules, mérite que l'on s'intéresse à cet aspect spécifique du cancer et à la recherche de stratégies thérapeutiques adaptées pour le combattre, dans l'espoir de réduire le nombre de victimes qu'il fait chaque année. Il convient de prêter attention à cet aspect spécifique du cancer et à la recherche de méthodes de traitement efficaces, dans l'espoir de réduire le nombre de victimes qu'il fait chaque année. À cet égard, nous abordons dans la section suivante une étude *in silico* à grande échelle sur une famille de petites molécules basées sur le cyclohexane-1,3-dione en tant que candidats prometteurs pour la conception de nouveaux médicaments dirigés spécifiquement contre le cancer du poumon non à petites cellules (NSCLC, en anglais Non-Small Cell Lung Cancer), en ciblant la protéine c-Met en tant que sujet de réflexion.

C. Criblage, Modélisation et Optimisation de Nouvelles Molécules Bioactives Dérivées de Cyclohexane-1,3-dione en tant qu'Agents Thérapeutiques Potentiels contre le Cancer du Poumon Non à Petites Cellules (CPNPC) en Ciblant le Récepteur Tyrosine Kinase c-Met

1 Introduction

Dans le cadre de la thérapie ciblée contre le cancer, un rapport de l'Organisation mondiale de la santé publié en 2020 révèle que 10.44 millions de personnes sont décédées des suites d'un cancer, avec un nombre élevé de décès, en particulier lié à un cancer du sein et du poumon [99]. Dans cette optique, l'étude présentée ici se concentre sur le cancer du poumon non à petites cellules, en tant que modèle pour son ciblage thérapeutique à l'aide de petites molécules dérivées de la cyclohexane-1,3-dione. Les lignées cellulaires H460 et A549 sont parmi les modèles *in vitro* les plus populaires pour les essais sur le cancer du poumon non à petites cellules (CPNPC) [439]. C'est dans ce contexte qu'une étude menée *in vitro* par Mohareb et al. est mise en évidence [409]. Pour cette étude, 40 composés hétérocycliques comportant des motifs pyraniques, Pyrazoliques, Thiophéniques et Thiazoliques à la base de cyclohexane-1,3-dione ont été synthétisés et testés vis-à-vis de six lignées cellulaires cancéreuses (H460 et A549 pour le cancer du poumon non à petites cellules, HT29 pour le cancer colorectal, MKN-45 pour le carcinome gastrique, U87MG pour le gliome malin, et SMMC-7721 pour le carcinome hépatocellulaire). Parmi les 40 composés (**Tableau III.C1**), 19 se sont révélés être les plus cytotoxiques contre les six lignées cellulaires ciblées. Ces 19 molécules ont ensuite été évaluées vis-à-vis de récepteurs tyrosine kinase, à savoir c-Kit, Flt-3, VEGFR-2, EGFR, PDGFR et Pim-1. La relation entre la structure des molécules à base de cyclohexane-1,3-dione et leur activité biologique suggère *in vitro* que ces petites molécules sont intéressantes en tant qu'agents anticancer dans le cadre de la thérapie ciblée contre le cancer. Donc, pour développer un modèle QSAR basé sur les valeurs IC_{50} pour les dérivés de cyclohexane-1,3-dione vis-à-vis du CPNPC, nous avons adopté la lignée cellulaire H460 comme référence. En choix des cibles thérapeutiques pour l'inhibition de la croissance des CPNPC, les récepteurs tyrosine kinases ont été privilégiés [440], [441].

La cyclohexane-1,3-dione a été signalée pour inhiber certains récepteurs à tyrosine kinase (par exemple, c-Kit, Flt-3, VEGFR-2, EGFR, PDGFR et Pim-1) [409], [442]. c-Met, ou tyrosine kinase de c-Met, a été décrit comme étant associé à l'expression du cancer du poumon non à petites cellules (CPNPC) [443], [444]. Dans ce cadre, le médicament anticancer Foretinib est considéré comme un agent de référence. En conséquence, pour développer un modèle *in silico* de représentation de l'activité biologique IC_{50} des dérivés de cyclohexane-1,3-dione, nous tenons compte des données d'entrée primaires (structure/activité) obtenues par les évaluations *in vitro* des structures moléculaires des 40 dérivés de cyclohexane-1,3-dione avec le Foretinib comme référence pour un calibrage et une étude *in silico* aboutissant à la découverte des nouveaux inhibiteurs spécifiques de c-Met pour applications thérapeutiques contre le cancer du poumon non à petites cellules. Toutefois, il est difficile de décider si une hypothèse est valable uniquement sur la base de données de criblage *in vitro*, notamment en l'absence d'informations telles que l'évaluation des propriétés de biodisponibilité, des paramètres ADME-Tox, l'affinité Protéine-inhibiteur, la stabilité et la compatibilité structurelle des nouveaux

inhibiteurs proposés vis-à-vis avec les protéines ciblées. Par conséquent, des nouvelles méthodes et fiables doivent être adoptées pour compléter le criblage classique et permettre à la modélisation moléculaire et à d'autres simulations de contribuer à la conception de futurs scénarios thérapeutiques. C'est ce que nous avons décidé de considérer dans cette étude comme un point de départ pour mettre en œuvre une stratégie hybride de CADD destinée à améliorer la pertinence des petites molécules bioactives portant l'échafaudage cyclohexane-1,3-dione dans le pipeline de développement et la conception rationnelle de nouveaux médicaments contre le cancer.

2 Stratégie de conception de médicaments assistée par ordinateur développée et appliquée

Dans nos deux études précédentes, nous avons modélisé avec succès de nouveaux inhibiteurs spécifiques de c-Met, candidats à une application dans la thérapie ciblée contre la majorité des cancers causés par l'activité enzymatique anormale du récepteur tyrosine kinase c-Met. En particulier, dans cette étude, nous considérons le cancer du poumon non à petites cellules comme un modèle de thérapie ciblée à l'aide de petites molécules bioactives. À cet égard, nous portons notre intérêt, du point de vue de la modélisation biomoléculaire basée sur les méthodes CADD, aux petites molécules bioactives portant l'échafaudage de cyclohexane-1,3-dione comme clés pour la conception rationnelle d'un candidat médicament par voie *in silico* contre le cancer du poumon non à petites cellules (CPNPC). La **Figure III.C1** illustre la méthodologie CADD hybride proposée et mise en œuvre dans cette étude pour le criblage, la modélisation et l'optimisation de nouvelles molécules candidates composées de dérivés de cyclohexane-1,3-dione en vue de la conception d'un nouveau médicament spécifique contre le CPNPC. Conformément au modèle de test biologique effectué *in vitro* qui utilisait le **Foretinib** comme standard pour le criblage, nous avons adopté le même médicament comme standard dans notre présente *in silico* étude. Les détails des méthodes et outils utilisés et tout ce qui est relatif à cette étude sont disponibles dans la version électronique de notre article publié à ce sujet. L'approche générale de ces méthodes est également discutée dans le deuxième chapitre de la thèse. Dans les sections suivantes, nous présentons les résultats et les conclusions les plus importants obtenus dans le cadre de cette étude.

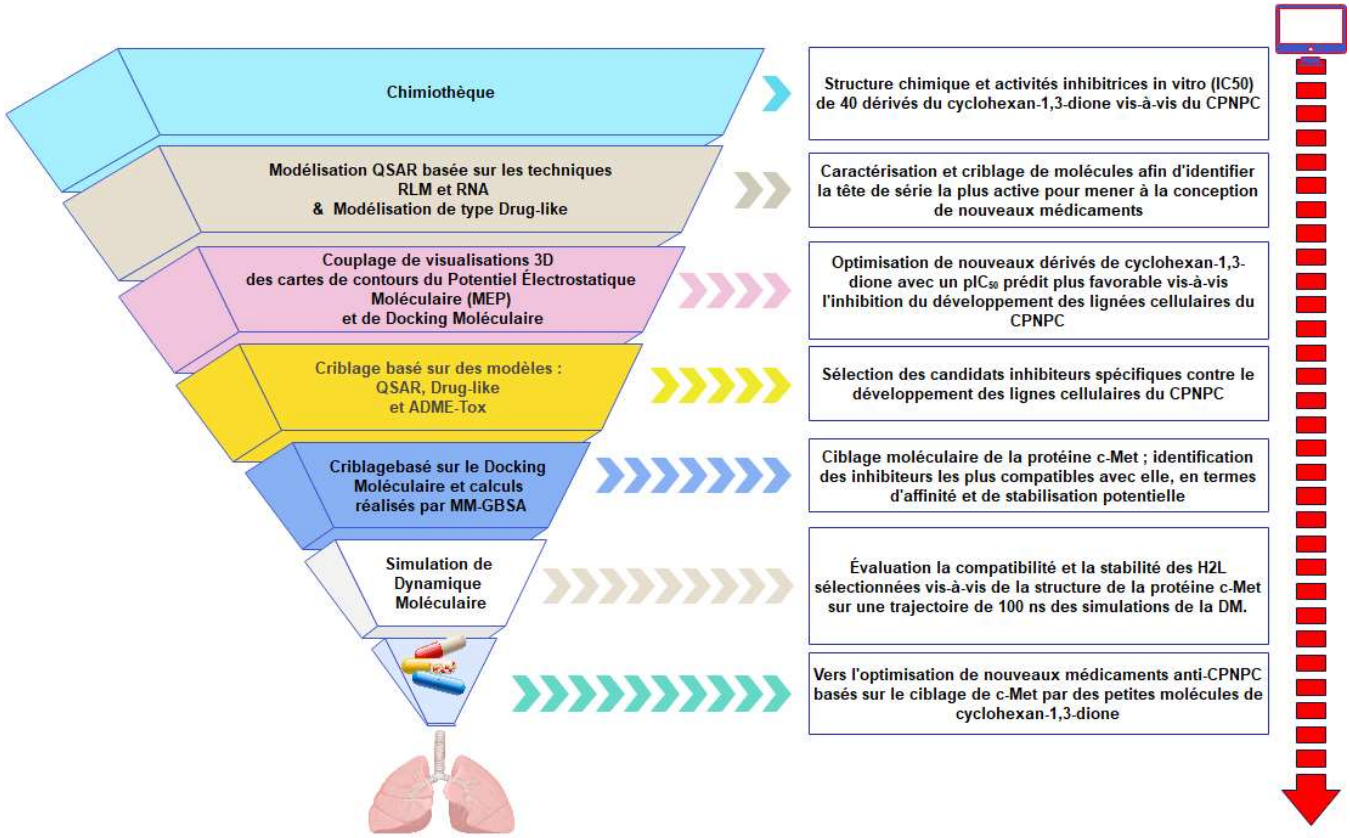
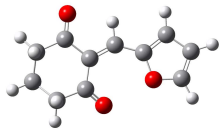
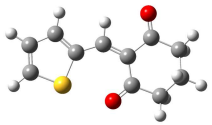
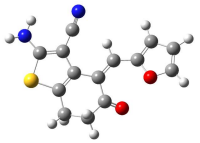
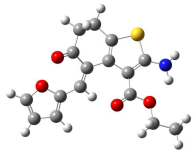
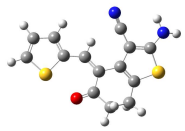
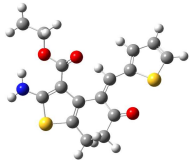
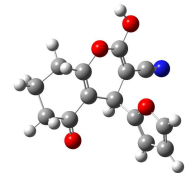
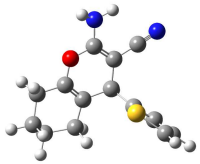
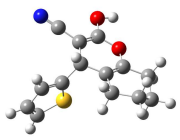
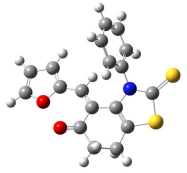
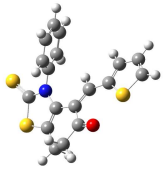
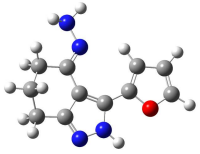
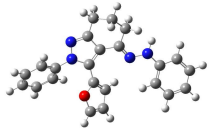
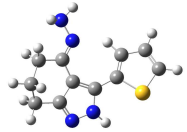
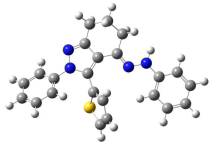
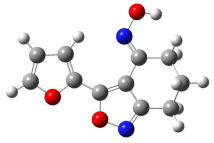
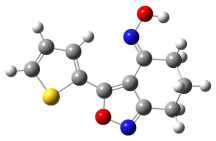
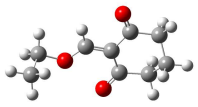
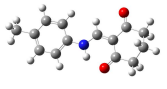
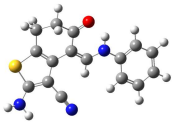
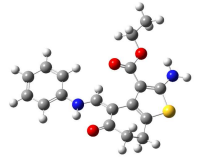
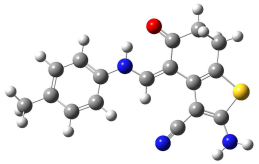
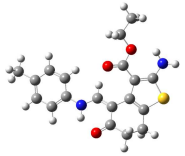
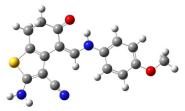
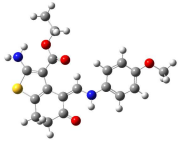
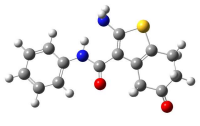
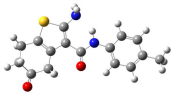
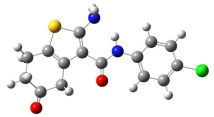
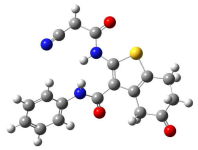
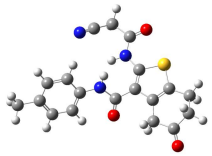
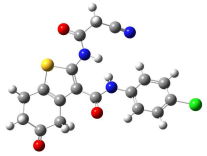
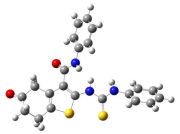
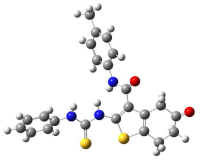
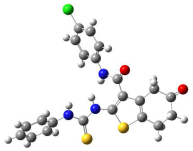
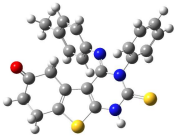
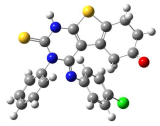


Figure III.C1. Stratégie de CADD hybride proposée et appliquée pour le criblage, la caractérisation et l'optimisation de nouvelles molécules de dérivés de cyclohexane-1,3-dione, candidates à la conception d'un nouveau médicament spécifique contre le CPNPC par le ciblage de c-Met.

Tableau III.C1. Structures moléculaires de dérivés de cyclohexane-1,3-dione optimisées par la méthode de la DFT conceptuelle avec informations *in vitro* sur leur activité inhibitrice « pIC_{50} » vis-à-vis de la lignée cellulaire H460 du cancer du poumon non à petites cellules (CPNPC).

	**3a**	**pIC_{50}=5.26**		**3b**	**pIC_{50}=6.41**		**5a**	**pIC_{50}=5.38**		**5b**	**pIC_{50}=5.93**		**5c**	**pIC_{50}=6.59**		**5d**	**pIC_{50}=6.44**		**6a**	**pIC_{50}=5.24**
	**6b**	**pIC_{50}=6.09**		**6c**	**pIC_{50}=6.43**		**6d**	**pIC_{50}=6.38**		**8a**	**pIC_{50}=6.55**		**8b**	**pIC_{50}=6.54**		**10a**	**pIC_{50}=5.12**		**10b**	**pIC_{50}=5.42**
	**10c**	**pIC_{50}=6.37**		**10d**	**pIC_{50}=6.24**		**12a**	**pIC_{50}=5.97**		**12b**	**pIC_{50}=6.44**		**14**	**pIC_{50}=5.17**		**16a**	**pIC_{50}=5.27**
	**16b**	**pIC_{50}=5.14**		**16c**	**pIC_{50}=6.17**		**17a**	**pIC_{50}=5.20**		**17b**	**pIC_{50}=5.34**		**17c**	**pIC_{50}=5.58**		**17d**	**pIC_{50}=5.78**
	**17e**	**pIC_{50}=6.29**		**17f**	**pIC_{50}=6.48**		**19a**	**pIC_{50}=5.08**		**19b**	**pIC_{50}=5.64**		**19c**	**pIC_{50}=6.49**		**20a**	**pIC_{50}=5.12**
	**20b**	**pIC_{50}=5.24**		**20c**	**pIC_{50}=6.46**		**21a**	**pIC_{50}=5.19**		**21b**	**pIC_{50}=5.79**		**21c**	**pIC_{50}=5.06**		**22a**	**pIC_{50}=5.08**
	**22b**	**pIC_{50}=5.63**		**22c**	**pIC_{50}=6.74**														

3 Résultats et discussion

3.1 Analyses structure-activité par une démarche hybride combine 2D-QSAR et cartographie structurelle en espace 3D

3.1.1 Modélisation Linéaire et Non Linéaire via 2D-QSAR

Nous avons procédé à une analyse en composantes principales (ACP) pour explorer la corrélation entre des descripteurs moléculaires hybrides (0D-3D) utilisés pour caractériser de petites molécules de dérivés de la cyclohexane-1,3-dione. Sur la base des résultats de l'ACP, les descripteurs Index de réfraction (**n**), Étirement-Flexion (**SB**), Acceptation de liaison hydrogène (**HBA**), Surface Moléculaire de Connolly (**CMA**), Surface Polaire (**PSA**), Connectivité totale (**TC**), Énergie totale ($\mathbf{E_T}$), Moment dipolaire (**DM**), Énergie HOMO ($\mathbf{E_{HOMO}}$), Énergie LUMO ($\mathbf{E_{LUMO}}$) et Indice d'électrophile (**w**) présentent une corrélation minimale entre eux (R<0.5). Par la suite, nous utilisons ces onze descripteurs moléculaires comme entrées pour développer un modèle QSAR applicable à l'activité biologique inhibitrice (pIC_{50}) des dérivés du cyclohexane-1,3-dione à l'aide des techniques d'apprentissage automatique de (RLM et RNA). Pour optimiser le modèle QSAR, nous avons divisé l'ensemble de données en un ensemble d'entraînement et un ensemble de test en utilisant la méthode du partitionnement univarié (UV). Sur la base de la division UV, huit observations (molécules) (6b, 6d, 10a, 10c, 12b, 19b, 20a et 22b) parmi les 40 de la base de données sont choisies comme éléments de l'ensemble de test pour valider les modèles QSAR. Les 32 molécules restantes sont considérées comme des éléments de l'ensemble d'entraînement. Le but de cette initiative est de développer un modèle 2D-QSAR (RLM-QSAR & RNA-QSAR) optimal permettant de définir la relation entre la structure et l'activité des dérivés de la cyclohexane-1,3-dione par le biais des descripteurs moléculaires hybrides (0D-3D). L'**Eq.** (**III.C1**) reflète le modèle RLM-QSAR optimal choisi pour décrire la relation structure-activité en 2D dans le cas de la chimiothèque des petites molécules de cyclohexane-1,3-dione visées par cette étude.

$$pIC_{50} = -0.93\times10^{-2}+1.58\times S\text{-}B+0.95\times HBA-5.73\times10^{-3}\times CMA-5.07\times10^{-2}\times PSA$$
$$-1.28\times TC-1.07\times10^{-4}\times E_T-0.74\times E_{HOMO}+0.44\times E_{LUMO} \qquad \text{(III.C1)}$$

$N=32,\ R=0.92,\ R^2=0.85,\ R^2_{Adj}=0.81,\ MSE=0.04$

$F=13.357,\ Pr<0.0001,\ Q^2=0.65,\ R^2_{test}=0.82$

La **Figure III.C2(a-e)** présente un bilan des informations les plus importantes que nous avons tirées et adoptées pour analyser et révéler la relation structure-activité des dérivés du cyclohexane-1,3-dione contre la CPNPC dans le cadre de la modélisation QSAR dans cette étude. Le modèle RLM-QSAR défini dans (**l'équation III.1**) indique précisément la corrélation entre les aspects structurels physicochimiques et électroniques des composés dérivés de la cyclohexane-1,3-dione et leur activité biologique inhibitrice contre les lignées cellulaires du CPNPC. Le coefficient de corrélation (R = 0.92) suggère une forte corrélation entre les valeurs observées et prédites du pIC_{50}, ce qui est confirmé par les niveaux élevés de R^2, R^2_{adj}, Q^2, R^2_{test} et F, ainsi que par les faibles valeurs de MSE et de P-value (<0.0001). La distribution linéaire régulière des valeurs observées et prédites du pIC_{50} obtenues par le modèle *in vitro* et le modèle RLM-QSAR, illustrées dans la **Figure III.C2(a)**, justifie encore davantage cette forte concordance. Les résultats du test LOOCV (Q^2=0.65>0.5), du test externe (R^2_{test}=0.82) et du

test de Y-randomisation (R^2_{Rand} et Q^2_{Rand}) démontrent que le modèle RLM-QSAR est fiable et a un fort pouvoir prédictif puisque les coefficients de corrélation sont supérieurs à 0.5. En outre, la différence significative entre les coefficients du modèle original et ceux du modèle randomisé ($^cR^2_p$ = 0.70>0.5) illustre la nature robuste du modèle proposé, à savoir que la corrélation entre la structure de chaque molécule et son activité biologique interprétée n'est pas due au hasard.

Par ailleurs, le diagramme de Williams du test du domaine d'applicabilité illustré à la **Figure III.C2(b)** indique que toutes les molécules des ensembles d'apprentissage et de test n'ont pas dépassé le seuil d'effet de levier (h^* = 0.7) et que les restes normalisés se sont situés dans la gamme acceptable de ±2.5. Cela indique que le modèle original de RLM-QSAR développé dans cette étude est fiable et précis pour prédire les activités des nouveaux composés basés sur les échafaudages du cyclohexan-1,3-dione. Cela confirme que le mélange de descripteurs multidimensionnels (0D-3D : Constitutionnels, Topologiques, Physico-chimiques et électronique issue de la DFT conceptuelle) que nous avons adopté dans la modélisation 2D-QSAR était approprié pour déchiffrer la relation structurelle des dérivés de cyclohexane-1,3-dione avec leur activité biologique anticancéreuse vis-à-vis de la CPNPC. Nous complétons cette analyse par une procédure d'investigation plus approfondie afin d'étudier plus avant les différents modèles de corrélation entre les huit descripteurs moléculaires représentatifs du modèle 2D-QSAR mis au point dans cette étude et l'activité biologique pIC_{50} vis-à-vis de la CPNPC. À cet égard, nous proposons de révéler la relation non linéaire entre les huit descripteurs (S-B, HBA, CMA, PSA, TC, E_T, E_{HOMO} et E_{LUMO}) et pIC_{50} à l'aide de la technique de RNA. Ceci est dû à l'efficacité du modèle RNA-QSAR à donner un aperçu plus approfondi et plus précis pour assurer le bon sélection des descripteurs moléculaires à travers le modèle linéaire, comme cela a été confirmé dans le cadre de notre étude précédente [294]. Un réseau neuronal de type Feed-forward (FFNN) avec des couches d'entrée, cachée et de sortie a été appliqué dans cette étude afin d'examiner le lien entre la variable pIC_{50} de sortie et les concurrents dans le modèle linéaire MLR (couche d'entrée). La **Figure III.C2(c)** présente l'architecture du modèle RNA-QSAR exploité dans le cadre de cette étude. Ce modèle RNA-QSAR non linéaire est extrêmement fidèle (R^2=0.98, Q^2=0.76, R^2_{test}=0.97, MSE=0.01 et ρ=1.03). Le coefficient de corrélation élevé (98%) peut être justifié par la distribution cohérente des valeurs pIC_{50} observées *in vitro* et prédites *in silico* par notre modèle non linéaire ANN-QSAR (**Figure III.C2d**). En dépit des performances apparemment élevées du RNA-QSAR, le modèle de RLM-QSAR offre une meilleure compréhension de la relation entre les descripteurs moléculaires et l'activité biologique, car il peut être représenté par une équation linéaire qui peut être évaluée et décrite sur la base des coefficients de chaque descripteur. Contrairement à ce modèle, le modèle non linéaire RNA-QSAR fournit de bonnes prédictions, mais n'est pas suffisamment précis du point de vue de la description des effets des descripteurs. Par conséquent, le modèle RLM-QSAR est plus pratique pour prédire l'activité biologique et la mettre en corrélation avec la structure des molécules. Dans cette optique, nous avons appliqué le modèle RLM-QSAR (équation **III.C1** et **Figure III.C2e**lors de cette étude afin d'analyser quantitativement la relation structure-activité des dérivés du cyclohexane-1,3-dione vis-à-vis du CPNPC.

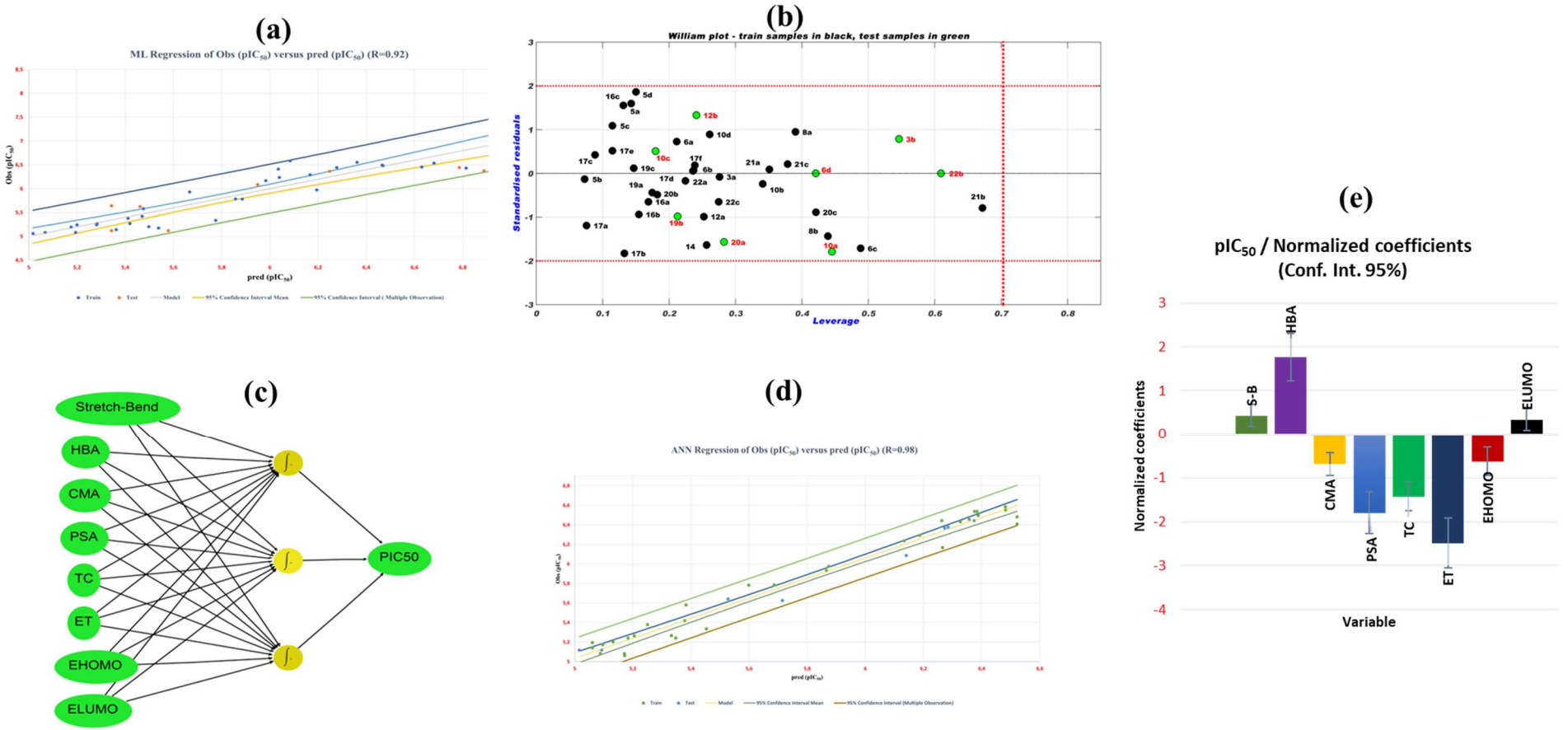


Figure III.C2. Valeurs des pIC_{50} des dérivés du cyclohexane-1,3-dione observées *in vitro* comparativement à celles prédites *in silco* à l'aide du modèle RLM-QSAR ; (b) Diagramme de William des résidus standard en fonction de l'effet de levier du modèle RLM-QSAR ; (c) Architecture du modèle non linéaire du RNA-QSAR ; (d) Valeurs des pIC_{50} des dérivés du cyclohexane-1,3-dione observées *in vitro* comparativement à celles prédites *in silco* à l'aide du modèle RNA-QSAR ; (e) Coefficients normalisés de chaque descripteur moléculaire choisi par le modèle RLM-QSAR.

Les coefficients normalisés issus du RLM-QSAR observé indiquent que l'activité biologique inhibitrice des dérivés du cyclohexane-1,3-dione hétérocycliques vis-à-vis de la CPNPC est largement caractérisée par les descripteurs moléculaires de type physico-chimique et électronique DFT tels que S-B, HBA, CMA, PSA, TC, E_T, E_{HOMO} et E_{LUMO}. Il est suggéré que les molécules présentant des valeurs S-B plus élevées, davantage de liaisons hydrogène, une surface moléculaire plus petite, une surface polaire plus faible, une faible énergie totale et une faible énergie LUMO soient incorporées dans les dérivés du cyclohexane-1,3-dione, afin de favoriser leurs activités biologiques inhibitrices optimales contre le CPNPC.

3.1.2 Sélection de la conformation bioactive prototype adaptée pour la conception de médicaments à partir du ligand

Dans cette phase, nous avons utilisé un modèle pharmacocinétique *in silico* pour évaluer les propriétés similaires à celles d'un médicament et les risques de toxicité de 19 dérivés de cyclohexane-1,3-dione qui avaient précédemment démontré une excellente activité inhibitrice contre des lignées de cellules cancéreuses. Ce modèle a évalué les paramètres de biodisponibilité de ces molécules sur la base des règles de Lipinski, Veber et Egan, et a prédit le niveau de flexibilité des molécules à l'aide des indices TPSA et n-RB (**Tableau III.C2**). L'indice d'accessibilité synthétique (SA) et les risques de toxicité ont également été évalués à l'aide des estimations d'AMES et d'hépatotoxicité. Les résultats issus de cette étape de criblage permettent d'identifier les ligands optimaux pour optimiser par la suite la conception de nouveaux inhibiteurs de c-Met en vue d'une thérapie ciblée contre le cancer.

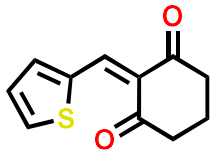
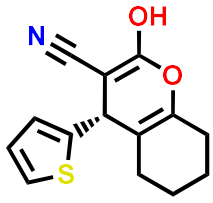
Tableau III.C2. Criblage *in silico* basé sur un modèle de type Drug-like.

Entrée	MW(Da)	n-RB	n-HBA	n-HBD	TPSA	Drug likeness			LogP	SA	Toxicité	
						Lipinski	Veber	Egan			AMES	Hépatotoxicité
Rule	<500	<10	<10	≤5	<140 A²	Yes/No			≤5	0<S.A<10	Oui/Non	
3b	**206.26**	**1**	**2**	**0**	**62.38**	**Oui**	**Oui**	**Oui**	**1.25**	**2.74**	**No**	**No**
5c	**286.37**	**1**	**2**	**1**	**123.36**	**Oui**	**Oui**	**Oui**	**1.25**	**3.27**	**Oui**	**Oui**
5d	**333.43**	**4**	**3**	**1**	**125.87**	**Oui**	**Oui**	**Oui**	**1.99**	**3.63**	**Oui**	**Oui**
6b	**257.24**	**1**	**5**	**1**	**83.46**	**Oui**	**Oui**	**Oui**	**-0.17**	**3.78**	**No**	**No**
6c	**258.34**	**1**	**2**	**1**	**87.28**	**Oui**	**Oui**	**Oui**	**1.59**	**3.97**	**Oui**	**Oui**
6d	**259.32**	**1.**	**3**	**1**	**81.49**	**Oui**	**Oui**	**Oui**	**1.59**	**3.61**	**No**	**No**
8a	**339.43**	**2**	**2**	**0**	**95.47**	**Oui**	**Oui**	**Oui**	**2.02**	**3.57**	**Oui**	**No**
8b	**355.50**	**2**	**1**	**0**	**110.57**	**Oui**	**Oui**	**Oui**	**2.90**	**3.60**	**Oui**	**No**
10c	**232.30**	**1**	**2**	**2**	**95.30**	**Oui**	**Oui**	**Oui**	**1.10**	**3.01**	**Oui**	**No**
10d	**384.50**	**4**	**2**	**1**	**70.45**	**Oui**	**Oui**	**Oui**	**4.10**	**3.71**	**Oui**	**Oui**
12a	**218.21**	**1**	**5**	**1**	**71.76**	**Oui**	**Oui**	**Oui**	**0.27**	**3.49**	**Oui**	**Oui**
12b	**234.27**	**1**	**4**	**1**	**86.86**	**Oui**	**Oui**	**Oui**	**1.10**	**3.51**	**Oui**	**No**
16c	**245.27**	**3**	**3**	**1**	**55.40**	**Oui**	**Oui**	**Oui**	**0.83**	**2.26**	**No**	**No**
17d	**356.44**	**5**	**3**	**2**	**109.66**	**Oui**	**Oui**	**Oui**	**2.10**	**3.72**	**No**	**No**
17[e]	**325.38**	**3**	**3**	**2**	**116.38**	**Oui**	**Oui**	**Oui**	**0.82**	**3.30**	**Oui**	**No**
17f	**372.44**	**6**	**4**	**2**	**118.89**	**Oui**	**Oui**	**Oui**	**1.55**	**3.71**	**No**	**Oui**
19c	**320.79**	**3**	**2**	**2**	**100.43**	**Oui**	**Oui**	**Oui**	**2.15**	**2.96**	**Oui**	**Oui**
20c	**387.84**	**6**	**4**	**2**	**127.30**	**Oui**	**Oui**	**Oui**	**1.48**	**3.37**	**Oui**	**Oui**
22c	**437.96**	**2**	**2**	**1**	**110.48**	**Oui**	**Oui**	**Oui**	**4.36**	**3.73**	**Oui**	**No**
For.	**632.65**	**14**	**10**	**2**	**111.25**	**Oui**	**No**	**Oui**	**2.62**	**4.05**	**No**	**Oui**

MW : poids moléculaire**, n-RB** : nombre de liaisons rotatives, **n-HBA** : Nombre d'accepteurs de liaisons hydrogène, **n-HBD** : Nombre de donneurs de liaisons hydrogène, **TPSA** : Surface Polaire Topologique, Loyauté avec le médicament (**Drug-likeness**) : Appropriée (Oui) ou Non Appropriée (Non), **Log P** : Le logarithme du coefficient de partage n-octanol-eau, **SA** : Accessibilité synthétique, **AMES** : Test de toxicité AMES, **Hépatotoxicité** : Risque d'hépatotoxicité.

Les résultats répertoires dans le **Tableau III.C2** indiquent que les 19 molécules satisfont aux règles de biodisponibilité de Lipinski, Veber et Egan. Cependant, seules 5 molécules (3b, 6b, 6d, 16c et 17d) se sont avérées non toxiques d'après les tests AMES et les tests d'hépatotoxicité. Par conséquent, les cinq molécules qui ont passé le criblage de type Drug-like réalisé *in silico* sont conservées pour une nouvelle étape de sélection, de même que la structure de médicament standard Foretinib (bien qu'elle soit considérée à risque de toxicité pour l'hépatotoxicité). Afin de sélectionner le meilleur ligand pour la modélisation de nouveaux composés cyclohexane-1,3-dione contre le CPNPC, nous avons analysé les valeurs pIC_{50} observées/prédites *in vitro*/ *in silico* des cinq candidats molécules. Les comparaisons entre les valeurs pIC_{50} observées ont révélé que les molécules **3b** (6.41) et 6d (6.38) sont les meilleurs en termes de pIC_{50} (observée/prédite), également les structures de ces dernières avaient une activité biologique inhibitrice très proches, ce qui rendait difficile de privilégier l'une par rapport à l'autre (**Tableau III.C3)**. La comparaison des valeurs prédites du pIC_{50} résultant des modèles MLR-QSAR/RNA-QSAR développés dans cette étude a révélé que la molécule **6d** (pIC_{50}=6.88) avait une activité prédite supérieure à celle de la molécule **3b** (pIC_{50}=6.035). Cela suggère que les descripteurs moléculaires proposés dans le modèle RLM-QSAR sont utiles pour améliorer l'activité pIC_{50} de la structure de la molécule **6d**. Par conséquent, la molécule **6d** de la famille des molécules est retenue comme base pour le développement de nouveaux dérivés de cyclohexane-1,3-dione en tant qu'inhibiteurs puissants de la protéine c-Me contre le cancer du poumon non à petites cellules (CPNPC).

Tableau III.C3. Structures-activités de cyclohexane-1,3-dione identifiées comme candidates à la conception de nouveaux médicaments.

	pIC_{50} observée *in vitro*	**pIC_{50} prédit *in silico***	
		RLM-QSAR	**RNA-QSAR**
 **3b**	6.41	6.035	6.598
 **6d**	6.38	**6.888**	6.354

3.2 Caractérisation 3D des propriétés optimales de la structure du ligand prototype

3.2.1 Cartographie 3D des contours du potentiel moléculaire électrostatique

Les cartes de contours du potentiel électrostatique moléculaire (MEP) en 3D peuvent être utilisées pour déterminer les sites appropriés pour les attaques électrophiles ou nucléophiles potentielles à l'intérieur de la poche active du c-Met. On peut s'en rendre compte en analysant la **Figure III.C3**, où

les contours rouges indiquent un potentiel électrostatique négatif (favorable aux attaques électrophiles, désignées par **R1**) et les contours bleus indiquent un potentiel électrostatique positif (favorable aux attaques nucléophiles, désignées par **R2** et **R3**). Pour favoriser la réactivité potentielle du composé 6d avec la poche active du c-Met, des substitutions devraient avoir lieu au niveau des sites **R1** et **R2**/**R3**, en remplaçant les groupements fonctionnels par ceux dont les propriétés pratiques favorisent les attaques électrophiles et nucléophiles, respectivement.

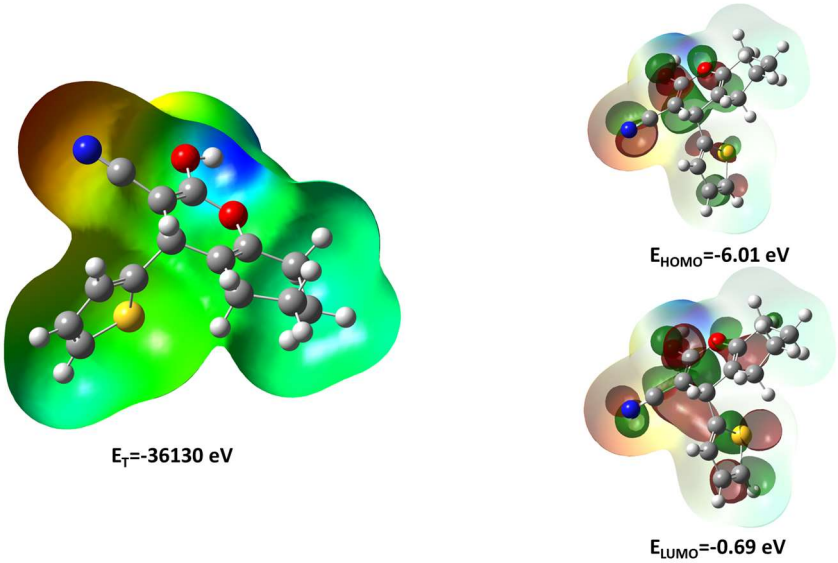


Figure III.C3. Cartographie du potentiel électrostatique moléculaire (MEP) et orbitales moléculaires HOMO/LUMO pour le composé optimal 6d.

3.2.2 Analyse de la compatibilité entre la surface c-Met et l'inhibiteur par Docking Moléculaire

Dans cette étape, nous employons l'approche de Docking Moléculaire semi-flexible pour docker le ligand **6d** à l'intérieur de la poche active de la protéine c-Met. Ceci afin d'obtenir une visualisation rapprochée dans l'espace 3D sur les propriétés structurelles les plus favorables à la compatibilité du ligand **6d** avec la structure 3D de la protéine c-Met. Pour ce faire, nous passons d'abord par deux étapes préliminaires : la validation du modèle de Docking moléculaire implémenté et l'identification de la poche active de la protéine c-Met. Les résultats de ces deux étapes peuvent être générés par la réalisation du protocole de réarticulation moléculaire de la structure de l'inhibiteur Foretinib avec le récepteur c-Met, qui repose sur la structure cristalline obtenue expérimentalement (code PDB : 3LQ8).

3.2.2.1 Validation du modèle du Docking moléculaire appliqué

La **Figure III.C4** présente la visualisation 3D observée pour l'alignement du Foretinib re-Docké (couleur jaune) avec le Foretinib original (couleur noire) à l'intérieur de la poche active de c-Met.

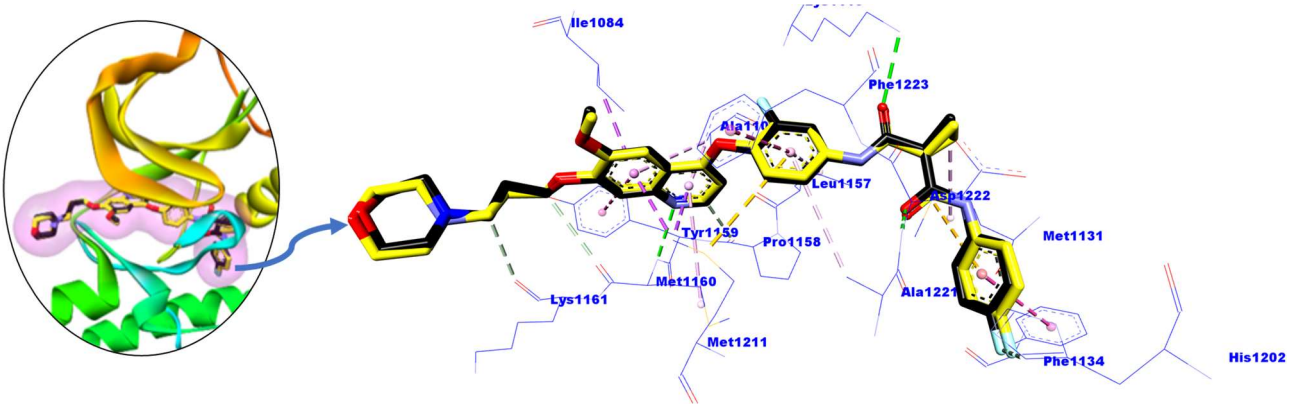


Figure III.C4. Modèle 3D des complexes cMet-Foretinib (3LQ8) et celui modélisé par le re-Docking (EL=-6.8 kcal/mol; RMSD=-0.2292 Å).

La superposition du ligand original du Foretinib par rapport à sa version re-Dockée est parfaite (RMSD=-0.2292 Å), ce qui indique la grande précision du protocole de Docking moléculaire mis en œuvre au cours de cette étape. Cela signifie que les prédictions de cette procédure sont considérées fiables et significatives. La superposition du ligand original du Foretinib par rapport à sa version re-dockée est parfaite (RMSD=-0.2292< 2 Å), ce qui indique la grande précision du protocole de Docking moléculaire mis en œuvre au cours de cette étape. Cela signifie que les prédictions de cette procédure sont considérées fiables et significatives. Ceci peut être confirmé par les interactions cMet-Foretinib dans la **Figure III.C4**, où nous constatons que le Foretinib re-Docké interagit avec le même résidu d'acide aminé que le Foretinib original présent dans 3LQ8. Sur cette base, nous pouvons considérer ces résidus actifs comme le site de liaison principal de c-Met. L'interaction entre un ligand biologiquement actif et ce site peut conduire à l'inhibition de l'activité enzymatique de c-Met, et donc à inhiber le développement de lignées cellulaires cancéreuse. Dans ce contexte, le site de liaison actif de c-Met se compose des résidus d'acides aminés suivants : **Isoleucine** (Ile1084), **Lysine** (Lys1110 et Lys1161), **Tyrosine** (Tyr1159), **Méthionine** (Met1160, Met1211 et Met1131), **Alanine** (Ala1108, Ala1221), **Proline** (Pro1158), **Leucine** (Leu1157), **Phénylalanine** (Phe1223, Phe1134), **Acide aspartique** (Asp1222), et **Histidine** (His1202). À titre de rappel, les résidus identifiés dans cette étude par le biais du protocole de re-Docking moléculaire semi-flexible sont les mêmes que ceux identifiés dans notre étude précédente par le biais d'un Docking rigide du Foretinib avec le c-Met. La seule différence concernait principalement l'énergie d'affinité, le Docking rigide produisant des énergies d'affinité plus négatives que le Docking semi-flexible.

3.2.2.2 Évaluation de la compatibilité entre le Ligand Prototype et la poche active de c-Met

À l'aide du modèle de Docking moléculaire développé, nous ancrons le ligand **6d** dans la maille de la poche active de c-Met (x = 0.20 Å, y = 4.04 Å, z = 28.96 Å avec une distance de 0.375 Å et une taille de 30 × 20 × 20 $Å^3$). La **Figure III.C5b** présente des visualisations 3D/2D du complexe cMe-6d après la procédure de Docking moléculaire. À partir de la **Figure III.C5**, nous pouvons constater que la structure du ligand 6d est adaptée à la poche de la protéine c-Met, les groupes Carbonitrile (**R1**), hydroxyde (**R2**) et atome d'oxygène (**R3**) interagissant respective ment avec les sites actifs **Lys**1110, Asp1222 et Glu1127. En outre, l'anneau Thiophénique et la structure du cercle interagissent avec

Glu1127, Asp1222, Phe1200, Phe1134, His1202 et Met1131. Bien que le Foretinib ait une énergie de liaison plus faible (-6.8 Kcal/mol) avec le récepteur c-Met, la structure du ligand 6d (-7.4 Kcal/mol) peut fournir de nouveaux dérivés de cyclohexane-1,3-dione susceptibles d'inhiber la croissance du CPNPC par l'inhibition de l'activité enzymatique de la c-Met. Au-delà des interactions non covalentes significatives établies entre le ligand 6d et la poche active de c-Met, nous pouvons observer que ce dernier est positionné à côté du Foretinib (**Figure III.C5a**). Ceci confirme la pertinence des structures des petites molécules portant l'échafaudage cyclohexan-1,3-dione à interagir avec c-Met dans le but de les inhiber.

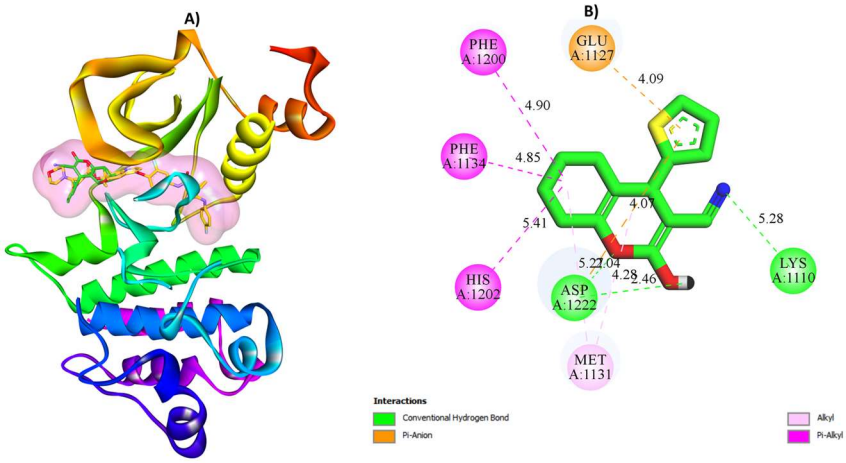


Figure III.C5. (a) Poses des ligands 6d (vert) et Foritinib (jaune) dans la poche active de c-Met, (b) Représentations 3D et 2D des interactions entre le ligand 6d et le site actif de liaison identifié (EL=-7.4 kcal/mol).

Pour développer davantage de connaissances sur les groupes fonctionnels du ligand 6d les plus favorables pour assurer la liaison avec le site actif de la protéine c-Met, nous avons reproduit la surface 3D du complexe cMet-6d obtenu par la méthode du Docking moléculaire en 3D (voir **Figure III.C7**). La **Figure III.C7(a-f)** illustre les surfaces de différentes formes favorables pour l'affinité en 3D concernant les interactions dans cMet-6d, y compris le bord/face aromatique, la liaison hydrogène donneur/accepteur, la charge, l'hydrophobicité, l'ionisation, et l'accessibilité au solvant. La **Figure III.C7(a)** (bleu (bord)/orange (face)) indique que les bords aromatiques du ligand 6d sont alignés favorablement, tandis que la **Figure III.C7(b)** (violet (donneur)/vert (accepteur)) montre que les sites **R1**, **R2** et **R3** sont appropriés pour les groupements donneurs de liaison hydrogène. La **Figure III.C7(c)** (bleu (charge positive) /rouge (charge négative)) suggère que le ligand 6d est favorable à une charge positive globale. La **Figure III.C7(d**) (bleu (hydrophile)/brun (hydrophobe)) indique que les sites **R1**, **R2** et **R3** sont favorables aux groupes hydrophobes. La **Figure III.C7(E)** (bleu (basique)/rouge (acide)) suggère que le ligand 6d est plus favorable aux conditions basiques, tandis que la **Figure III.C7(F)** (bleu (SAS élevé) /vert (SAS faible)) démontre la SASA élevée des résidus d'acides aminés de la protéine c-Met. Les résultats de la modélisation 2D-QSAR, des cartes de contours MEP/Docking Moléculaire et des prédictions Drug-like suggèrent que la structure du 6d peut être utilisée comme référence pour la conception de nouveaux inhibiteurs contre les CPNPC ciblant le récepteur de la tyrosine kinase c-Met. La **Figure III.C6** récapitule les dispositions structurelles favorables des dérivés de cyclohexan-1,3-dione qui peuvent optimiser à la fois leur pouvoir d'inhibition biologique contre le CPNPC et leur liaison et compatibilité avec le récepteur de la tyrosine kinase c-Met.

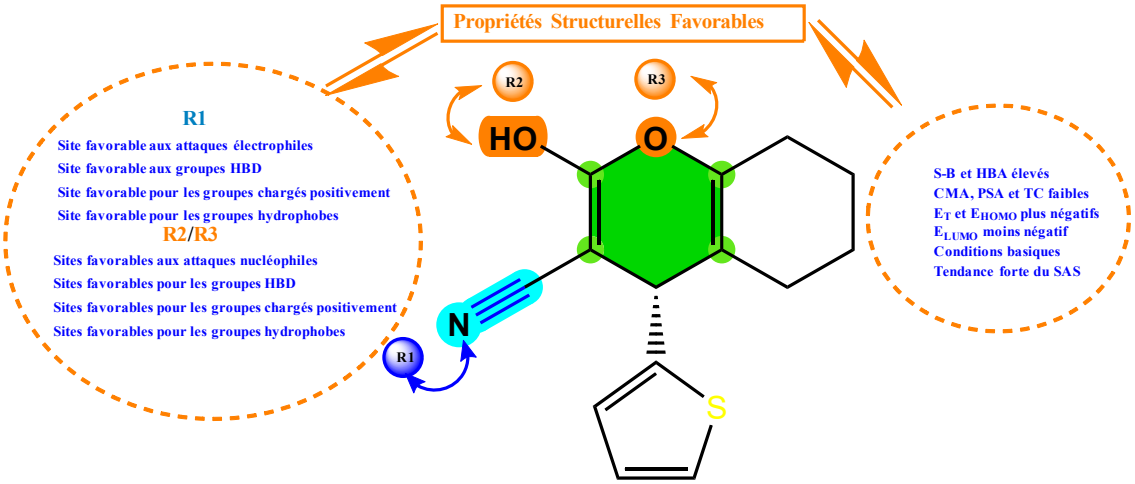


Figure III.C 6. Propriétés structurelles clés du ligand 6d favorables à l'optimisation de nouveaux inhibiteurs de c-Met à base de cyclohexan-1,3-dione en vue d'une thérapie future contre le CPNPC.

3.3 Concevoir, optimisation et criblage de nouveaux dérivés de Cyclohexane-1,3-dione

Nous avons conçu 36 nouveaux composés à base de cyclohexan-1,3-dione reposant sur la structure de la molécule de référence (6d) et des propriétés structurelles favorables pour renforcer leur activité biologique vis-à-vis du cancer du poumon non à petites cellules (CPNPC). La rétro-synthèse de ces composés peut être menée selon le même protocole que celui utilisé dans la littérature scientifique, à savoir des réactions chimiques du type "réaction de Michael", avec laquelle nous avons conçu les molécules proposées dans le cadre de cette étude (**Figure III.C8**). L'optimisation de ces nouveaux composés est réalisée selon la même stratégie de calcul et par la suite, les descripteurs (S-B, HBA, CMA, PSA, TC, E_T, E_{HOMO} et E_{LUMO}) sont calculés par la méthode DFT/B3LYP/6-31G(d,p). Les valeurs d'activité inhibitrice de ces 36 structures sont prédites à l'aide d'un modèle RLM-QSAR, dont les résultats sont indiqués dans le **Tableau III.C4.**

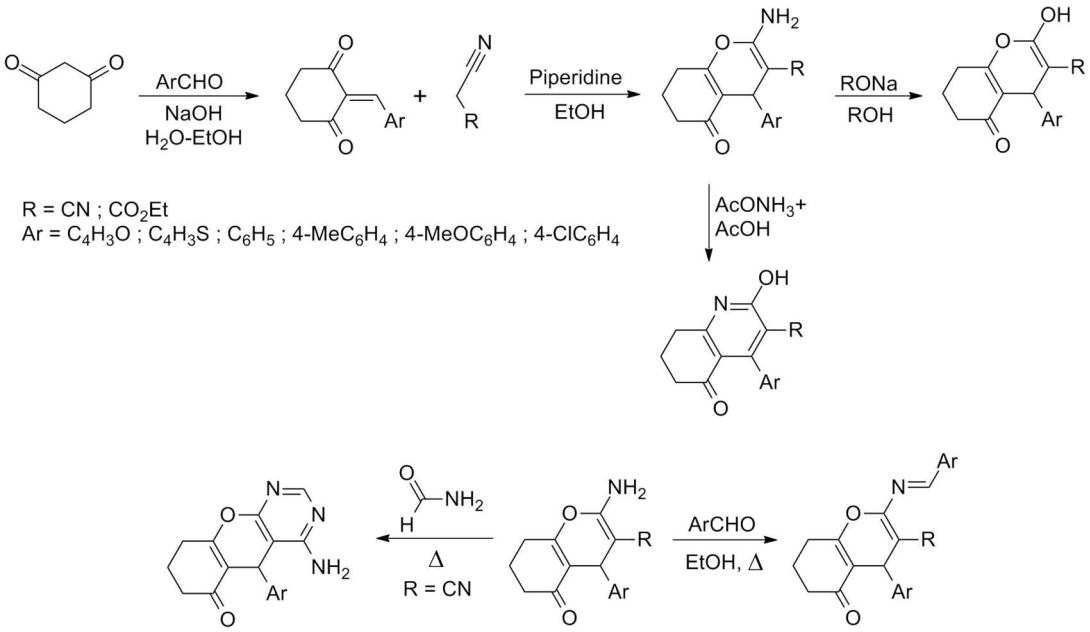


Figure III.C7. Voies de synthèse adoptées pour concevoir de nouveaux dérivés de cyclohexan-1,3-dione avec la base (6d).

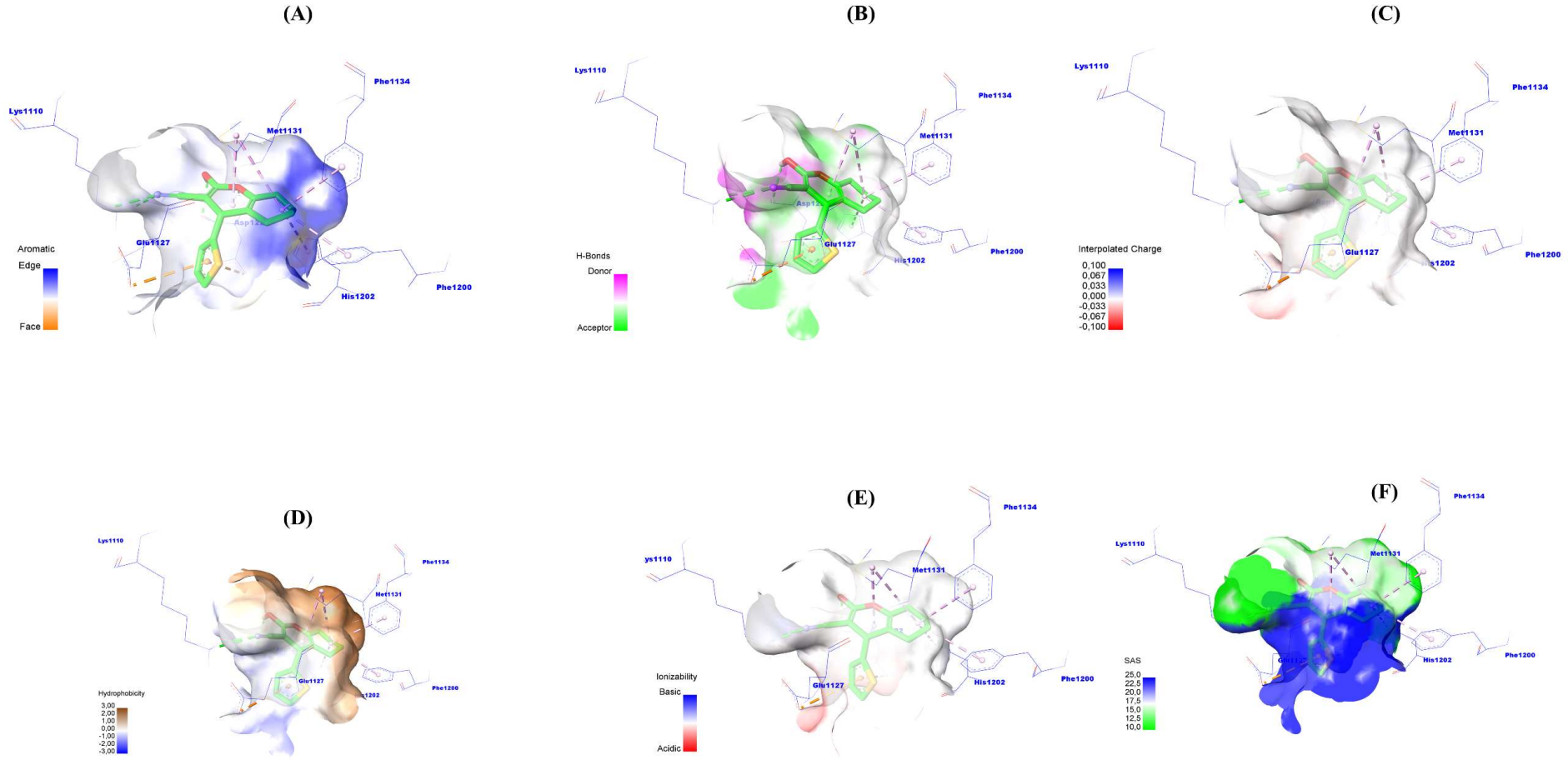
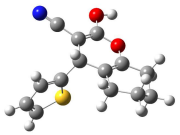
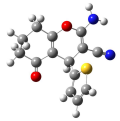
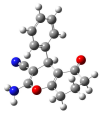
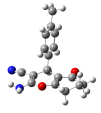
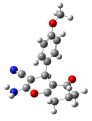
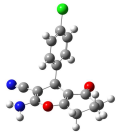
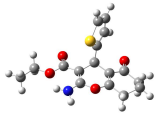
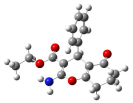
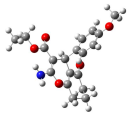
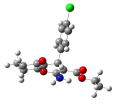
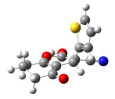
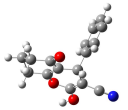
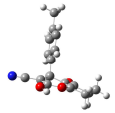
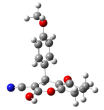
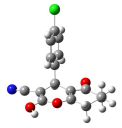
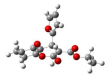
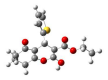
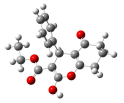
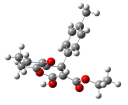
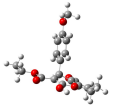
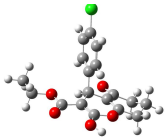
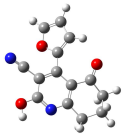
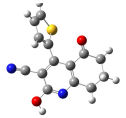
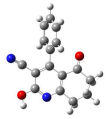
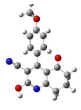
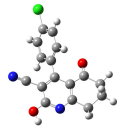
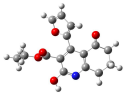
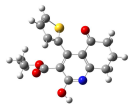
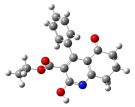
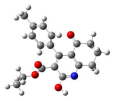
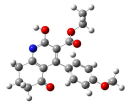
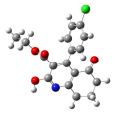


Figure III.C8. Représentation graphique de la surface (a) aromatique, (b) de liaison hydrogène, (c) de charge, (d) d'hydrophobie, (e) d'ionisation, (f) d'accessibilité au solvant du c-Met autour de 6d.

Tableau III.C4. Structure-activité optimisée *in silico* de nouveaux dérivés de cyclohexan-1,3-dione en tant qu'inhibiteurs prometteurs de la protéine c-Met anti-CPNPC.

															
6d	**pIC_{50}=6.88**	**H01**	**pIC_{50}=7.06**	**H02**	**pIC_{50}=6.50**	**H03**	**pIC_{50}=5.88**	**H04**	**pIC_{50}=8.81**	**H05**	**pIC_{50}=6.30**	**H06**	**pIC_{50}=7.22**	**H07**	**pIC_{50}=6.30**						
															
H08	**pIC_{50}=5.47**	**H09**	**pIC_{50}=5.69**	**H10**	**pIC_{50}=5.92**	**H11**	**pIC_{50}=5.28**	**H12**	**pIC_{50}=7.19**	**H13**	**pIC_{50}=5.50**	**H14**	**pIC_{50}=5.87**	**H15**	**pIC_{50}=5.34**						
															
H16	**pIC_{50}=5.36**	**H17**	**pIC_{50}=5.56**	**H18**	**pIC_{50}=7.01**	**H19**	**pIC_{50}=5.30**	**H20**	**pIC_{50}=7.49**	**H21**	**pIC_{50}=7.02**	**H22**	**pIC_{50}=5.84**	**H23**	**pIC_{50}=5.70**						
															
H24	**pIC_{50}=7.32**	**H25**	**pIC_{50}=6.76**	**H26**	**pIC_{50}=6.49**	**H27**	**pIC_{50}=6.41**	**H28**	**pIC_{50}=6.44**	**H29**	**pIC_{50}=6.70**	**H30**	**pIC_{50}=7.62**	**H31**	**pIC_{50}=7.05**						
																						
H32	**pIC_{50}=7.31**	**H33**	**pIC_{50}=6.47**	**H34**	**pIC_{50}=6.39**	**H35**	**pIC_{50}=6.81**	**H36**	**pIC_{50}=7.92**												

Les molécules optimisées qui présentent une activité biologique prédite supérieure à celle de la conformation **6d** de référence sont ombrées en gris

Le modèle RLM-QSAR appliqué dans cette étude a réussi à prédire la relation structure-activité pour les dérivés du cyclohexane-1,3-dione par la détermination des valeurs de pIC_{50}. Parmi les trente-six nouvelles molécules soumises à évaluation biologique de pIC_{50} *in silico*, onze ont enregistré une amélioration considérable dans leur activité biologique inhibitrice vis-à-vis du CPNPC par rapport à 6d (pIC_{50} *in vitro* = 6.38, pIC_{50} *in silico* = 6.88) et au médicament standard Foretinib (pIC_{50} = 6.67). Ces onze nouvelles molécules ont prédit des valeurs de pIC_{50} situées entre 7.01 et 7.92. Sur la base de ces résultats, les 11 nouveaux composés proposés (H01, H06, H12, H18, H20, H21, H24, H30, H31, H32, et H36) sont considérés comme ayant une forte activité inhibitrice vis-à vis de la prolifération des cellules du cancer du poumon non à petites cellules.

3.4 Sélection des candidats H2L à travers des prédictions de Drug-like et d'ADME-Tox

Pour filtrer les nouveaux dérivés de cyclohexan-1,3-dione conçus à des fins médicinales, ils doivent avoir des propriétés de type Drug-like et ADME-Tox favorables. À cette fin, on a soumis les 11 composés retenus à un criblage basé sur les propriétés pharmacocinétiques et pharmacodynamiques en utilisant les modèles *in silico* (SwissADME et pkCSM) disponibles sur Internet. Les profils de biodisponibilité des substances par voie orale et ADME-Tox définis dans le cadre de cette étude sont présentés dans le **Tableau III.C5(a)** et le **Tableau III.C5(b)**, respectivement.

Tableau III.C5. (a) Profil de type Drug-like appliqué à des nouveaux dérivés optimisés de la cyclohexane-1,3-dione.

Entrée	ABS	TPSA	n-RB	MW	Log P	n-HBA	n-HBD	Violations			Score de biodisponibilité	Pains	S.A
								Lipinski	Veber	Egan			
Règle	**-**	**<140 Å²**	**<10**	**<500 Da**	**<=5**	**<10**	**<5**	**<2**	**<2**	**<2**	**F>0.1**	**-**	**0<SA <10**
H01	Haut	89.25	1	256.26	2.09	4	1	0	0	0	0.56	0	3.92
H06	Haut	76.11	1	300.74	3.15	3	1	0	0	0	0.56	0	3.77
H12	Haut	78.62	4	347.79	3.19	4	1	0	0	0	0.56	0	4.09
H18	Haut	70.32	1	301.72	3.75	4	1	0	0	0	0.85	0	3.62
H20	Haut	101.07	4	320.36	3.20	5	1	0	0	0	0.56	0	4.01
H21	Haut	72.83	4	314.33	3.14	5	1	0	0	0	0.85	0	3.96
H24	Haut	72.83	4	348.78	3.79	5	1	0	0	0	0.85	0	3.92
H30	Haut	73.98	1	298.72	3.50	4	1	0	0	0	0.55	0	2.59
H31	Haut	89.63	4	301.29	2.74	6	1	0	0	0	0.55	0	3.08
H32	Haut	104.73	4	317.36	3.21	5	1	0	0	0	0.55	0	3.04
H36	Haut	76.49	4	345.78	3.80	5	1	0	0	0	0.55	0	2.95

Comme indiqué dans le **Tableau III.C5**, les 11 molécules proposées satisfont toutes les critères appropriés aux médicaments, et l'absence d'alerte à la douleur (0) suggère que les résultats obtenus sur le plan biologique sont valables. De plus, leur score (F>10%) correspond aux performances des règles de Lipinski, et leur indice d'accessibilité synthétique est compris entre 2.95 et 4.09. La forte adéquation entre la biodisponibilité et les structures proposées indique que ces dernières peuvent être considérées comme des candidats aptes aux applications thérapeutiques. Afin de valider cette hypothèse, nous avons procédé la modélisation ADME-Tox par une comparaison avec les ligands 6d et Foretinib. Les résultats obtenus sont présentés dans le **Tableau III.C6**. Sur la base des prédictions présentées par le modèle ADME-Tox illustré dans le **Tableau III.C6**, nous pouvons tirer les conclusions

suivantes : Les résultats de l'indice d'absorption indiquent un potentiel d'absorption élevé des molécules dans l'intestin humain, ce qui augmente leurs chances de succès sur le plan thérapeutique. Le taux de perméabilité cutanée, indiqué par la valeur K_p, était faible pour toutes les molécules, y compris le composé de référence Foretinib. Par conséquent, toutes les molécules proposées sont susceptibles de ne pas avoir d'effet dermatologique. Le volume de distribution (VDss) des molécules examinées dans le plasma humain était compris entre 0 et 0.3, ce qui indique la capacité des molécules proposées à se lier à la protéine c-Met à des niveaux importants. Les prévisions de pénétrabilité de la barrière hémato-encéphalique (BBB) et du système nerveux central (SNC) ont été évaluées ; à l'exception du composé **6d**, les résultats ont montré que toutes les molécules présentaient une gamme acceptable et étaient négatives dans le sang. L'enzyme cytochrome CYP3A4 s'est avérée n'avoir aucun effet inhibiteur sur les molécules, à l'exception du composé **H31** qui s'est avéré être un substrat de cette enzyme. L'indice d'excrétion de toutes les molécules était inférieur à 1, à l'exception du Foretinib, ce qui indique une meilleure stabilité des molécules étudiées dans l'organisme. Les résultats du test AMES, de la LD_{50} et du LOAEL indiquent une toxicité potentielle des composés **H31**, **H32** et Foretinib, mais il est peu probable que cette toxicité se produise en raison des faibles doses de l'IC_{50} micromolaire. Enfin, les prévisions relatives à la sensibilité cutanée indiquent que toutes les molécules ne présentent aucun danger pour le contact. Les prévisions ADME-Tox suggèrent que les 9 composés de tête/H2L (**H01**, **H06**, **H12**, **H18**, **H20**, **H21**, **H24**, **H30** and **H36**) sont appropriés pour la découverte de nouveaux médicaments et le développement thérapeutique ; cependant, en raison de problèmes de toxicité potentiels, les molécules **H31** et **H32** ont été exclues des résultats. Cela indique que la structure du ligand **6d** était appropriée pour la conception et l'optimisation des nouveaux médicaments potentiels contre le CPNPC.

3.5 Sélection des conformations optimales des dérivés de Cyclohexane-1,3-dione compatibles en tant qu'inhibiteurs de c-Met

En ancrant par la stratégie du Docking moléculaire les neuf ligands candidats (H2L) dans la poche active de c-Met, nous avons obtenu les résultats présentés dans le **Tableau III.C7**. Nous analysons l'affinité et la compatibilité des ligands avec la structure de c-Met par le biais des valeurs d'énergie de liaison (EL) et des constantes d'inhibition (K_i) estimées entre les conformations de cyclohexane-1,3-dione (ligands) et le récepteur de tyrosine kinase c-Met. Les résultats du Docking moléculaire montrent clairement que les ligands présentent des énergies de liaison négatives vis-à-vis de la poche active de c-Met. Cela signifie que les structures de ces ligands sont très compatibles avec la structure du c-Met. D'autre part, tous les ligands ont montré des valeurs Ki très inférieures à 100 µM, ce qui signifie que les nouveaux ligands conçus sont très appropriés comme inhibiteurs puissants de l'activité enzymatique de la protéine c-Met ciblée dans cette étude. Dans le même sens, les ligands **H06**, **H12**, **H30** et **H36** dans les complexes de c-Met ont montré des énergies de liaison plus négatives que la molécule de référence 6d et le Foretinib. Tous les résultats du Docking moléculaire obtenus confirment que le protocole d'optimisation moléculaire des dérivés du cyclohexane-1,3-dione utilisé dans cette étude a permis de générer de nouvelles molécules computationnelles dont les structures répondent parfaitement aux critères de conception pharmaceutique. Par conséquent, les neuf nouvelles petites

molécules optimisées peuvent être considérées comme une nouvelle génération d'inhibiteurs de c-Met susceptibles de contribuer à développer une nouvelle stratégie thérapeutique contre le cancer.

3.6 Analyse des Interactions entre cMet et l'Inhibiteur via le Docking Moléculaire et l'Approche Prime MM-GBSA

Sur la base des données EL/Ki présentées dans le **Tableau III.C7**, nous élaborons des visualisations des interactions des ligands H06, H12, H30 et H36 avec la poche de la c-Met. Ceci a pour but d'identifier les interactions les plus importantes impliquées dans l'affinité des dérivés du cyclohexane-1,3-dione avec la poche active de la c-Met (**Figure III.C9**). En parallèle, nous effectuons des simulations MM-GBSA selon le modèle de solubilité VSGB 2.0 pour recalculer et estimer l'énergie libre de liaison (ΔG_{bind}) résultant de la formation de liaisons non covalentes entre le c-Met et les échantillons H06, H12, H30 et H36. Cette procédure nous permet d'étudier la stabilité des interactions protéine-ligand lors de la présence de solvant comme de molécules d'eau. Le **Tableau III.C8** présente un aperçu général de contacts cMet-ligand les plus notables prévus par Docking moléculaire et confirmés par les calculs de l'énergie libre de liaison de Prime MM-GBSA. De l'ensemble des résultats des simulations de Docking moléculaire et MM-GBSA, nous pouvons constater que les principales interactions protéine-ligand générées étaient des liaisons hydrogène, des interactions électrostatiques, des interactions van der Waals et des interactions hydrophobes. Les liaisons faibles (interactions non covalentes) renforcent la compatibilité entre le récepteur c-Met et les ligands médicamenteux proposés. Ce résultat est principalement lié aux énergies libres de liaison nettement plus basses obtenues par Docking moléculaire (-8.9 à -6.4 kcal/mol) et par les simulations Prime MM-GBSA (ΔG_{bind}, ΔG_{bind} Vdw, and ΔG_{bind} Hbond allant de -43.65 à -20.05, -53.8 à -34.28 et -1.64 à -0.67 kcal/mol, respectivement). La **Figure III.C9** indique que les interactions non covalentes établies entre les 4 ligands de base et le site actif de c-Met restent préservées après les computations de Prime/MM-GBSA.

3.7 Vérification rétrospective des résultats à travers des Simulations de Dynamique Moléculaire

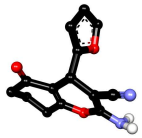
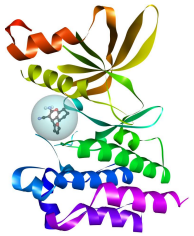
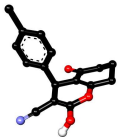
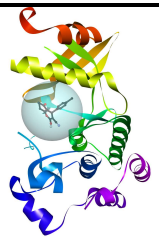
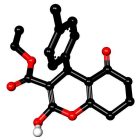
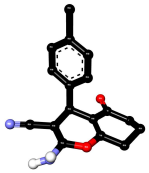
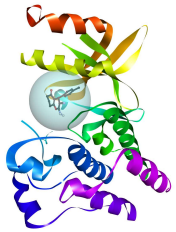
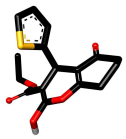
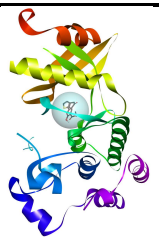
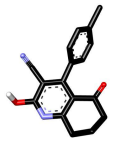
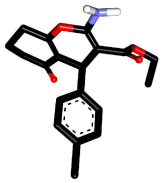
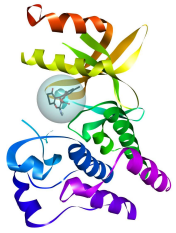
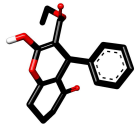
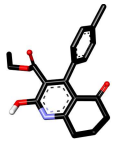
Pour approfondir davantage la stabilité temporelle des composés proposés dans cette étude comme inhibiteurs spécifiques de c-Met, nous avons soumis les 4 échantillons cMet-H06 (-7.9 kcal/mol), cMet-H12 (-8.1 kcal/mol), cMet-H30 (-8.3 kcal/mol), cMet-H36 (-8.9 kcal/mol), qui présentent les plus forts profils d'affinité avec la poche active de c-Met, à des simulations de la dynamique moléculaire. A cet égard, les complexes sont simulés dans un environnement aqueux et des conditions proches du modèle expérimental réel ; les profils de stabilité et la compatibilité des structures des échantillons soumis à cet essai avec la structure de la protéine c-Met sont étudiés. Dans cette étape, nous avons adopté la structure du c-Met libre (non complexé), et du complexe c-Met-Foretinib comme paramètres structurels de référence dans le suivi du comportement dynamique des systèmes d'inhibiteurs biomoléculaires de c-Met conçus (cMet-H06, cMet-H12, cMet-H30, et cMet-H36) pendant des cycles de simulation MD indépendants à l'échelle de temps de 0 à 100 nanosecondes.

Tableau III.C6. Propriétés ADME-Tox prédites *in silico*.

Modèle ADME-Tox		Absorption		Distribution				Métabolisme							Excrétion	Toxicité				
		Ab	SP	VDss	Fu	BBB	CNS	CYP-Substrat		CYP-Inhibiteur					TC	AMT	LD_{50}	ORC	Hep	SS
								2D6	3A4	1A2	2C19	2C9	2D6	3A4						
Réf.	**6d**	92.8	-2.7	-0.026	0.3	-2.02	-3.7	No	Yes	No	No	No	No	Yes	1.14	No	2.6	2.8	**Yes**	No
	Foretinib	90.7	-2.7	0.260	0.1	0.04	-1.7	No	Yes	Yes	Yes	No	No	No	0.01	No	2.7	1.9	No	No
Composé optimisé	**H01**	95.4	-3.3	-0.145	0.3	-0.20	-3.9	No	Yes	No	No	No	No	No	0.58	No	2.8	0.8	No	No
	H06	93.6	-3.0	0.102	0.09	-0.09	-2.9	No	Yes	No	Yes	No	No	No	-0.09	No	2.7	1.2	No	No
	H12	94.5	-3.9	-0.094	0.07	-0.25	-3.1	No	Yes	No	Yes	No	No	No	-0.02	No	2.9	1.1	No	No
	H18	92.3	-3.5	0.049	0.04	-0.15	-2.8	No	Yes	Yes	Yes	No	No	No	0.05	No	2.9	1.3	No	No
	H20	93.0	-3.6	-0.102	0.1	-0.38	-3.1	No	Yes	No	Yes	No	No	No	0.19	No	3.0	1.1	No	No
	H21	94.6	-3.5	-0.05	0.07	-0.16	-3.1	No	Yes	No	No	No	No	No	0.50	No	2.8	1.1	No	No
	H24	92.9	-3.4	-0.08	0.04	-0.31	-3.0	No	Yes	No	No	No	No	No	0.12	No	3.1	1.1	No	No
	H30	93.2	-2.9	0.08	0.04	-0.23	-2.8	No	Yes	Yes	Yes	Yes	No	No	0.02	No	2.6	0.8	No	No
	H31	95.9	-2.7	-0.3	0.20	-0.577	-3.0	No	No	Yes	No	No	No	No	0.822	No	3.0	1.3	**Yes**	No
	H32	93.3	-2.7	-0.3	0.02	-0.451	-3.3	No	Yes	No	Yes	No	No	No	0.194	**Yes**	2.6	0.8	**Yes**	No
	H36	93.4	-2.7	-0.2	0	-0.383	-3.2	No	Yes	No	Yes	Yes	No	No	0.122	No	2.6	0.8	No	No

Ab (absorption intestinale) ; **SP** (perméabilité de la peau) **VDss** (Volume de distribution chez l'homme à l'état stable.) ; **Fu** (Fraction non liée) ; **BBB** (perméabilité de la barrière hémato-encéphalique) ; **CNS** (perméabilité du système nerveux central) ; **CYP-Substrat** (Substrat d'enzyme Cytochrome P450) ; **CYP-Inhibiteur** (Inhibiteur d'enzyme Cytochrome P450) ; **TC** (Clairance totale); **AMT** (Toxicité d'AMES); **LD50** (Toxicité aiguë orale pour le rat (DL50)) ; **ORC** (Toxicité chronique orale pour le rat (LOAEL)); **Hep** (Hépatotoxicité) ; **SS** (Sensibilisation de la peau).

Tableau III.C7. Résultats du criblage virtuel basé sur le Docking moléculaire de candidats inhibiteurs de c-Met.

Ligand	Complexe		Ligand	Complexe		Ligand	Complexe	
H01	cMe-H01		H18	cMet-H18		H24	cMe-H24	
		EL=-7.7 kcal/mol Ki=2.23 µM			EL=-6.4 kcal/mol Ki= 20.1 µM			EL=-6.4 kcal/mol Ki= 20.1 µM
H06	c-Met-H06		H20	cMet-H20		H30	cMe-H30	
		EL=-7.9 kcal/mol Ki= 1.59 µM			EL=-7.8 kcal/mol Ki= 1.88 µM			EL=-8.3 kcal/mol Ki= 0.81 µM
H12	cMet-H12		H21	cMet-H21		H36	cMet-H36	
		EL=-8.1 kcal/mol Ki= 1.13 µM			EL=-6.9 kcal/mol Ki= 8.64 µM			EL=-8.9 kcal/mol Ki= 0.29 µM

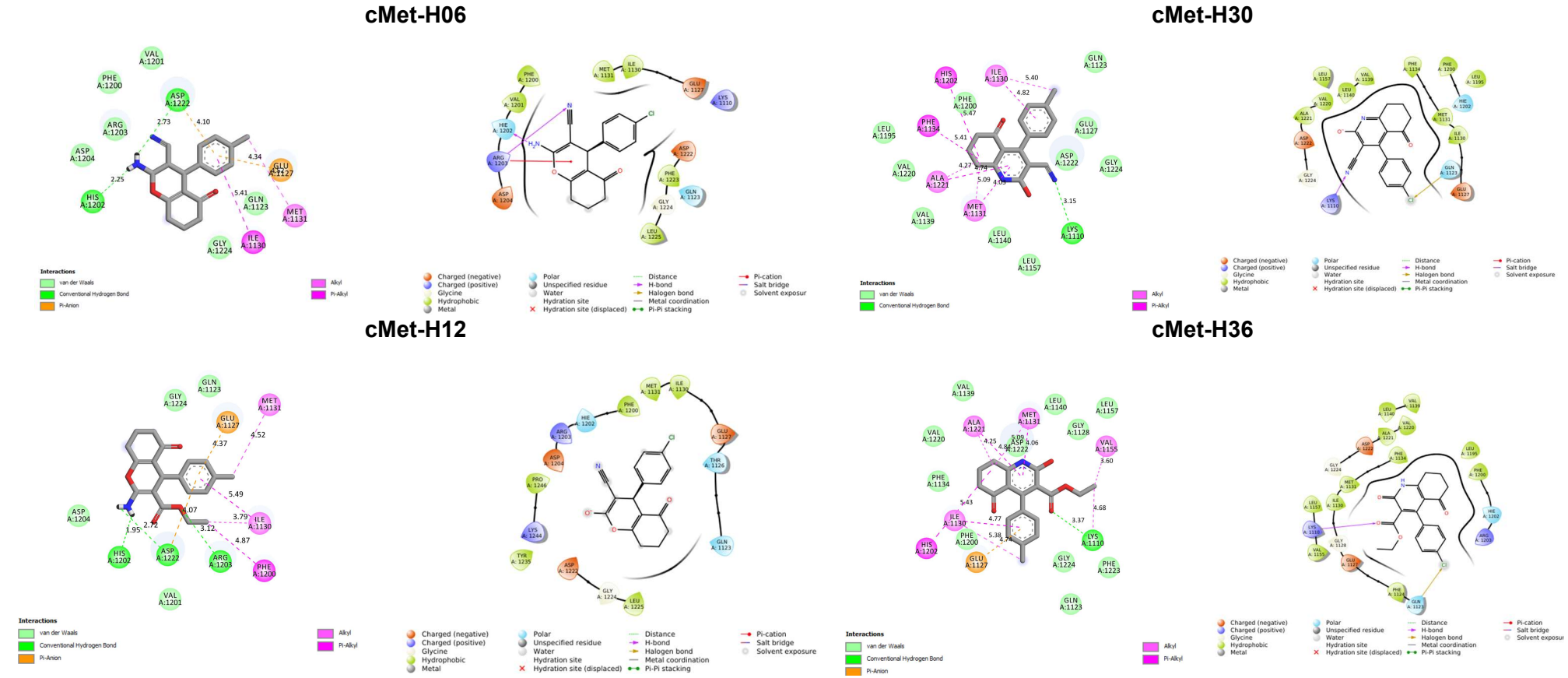


Figure III.C9. Visualisations 2D des interactions identifiées dans les complexes cMet-H06, cMet-H12, cMet-H30 et cMet-H36 par docking moléculaire (à gauche) et par calculs Prime MM-GBSA (à droite).

Tableau III.C8. Bilan des simulations de Docking moléculaire et du calcul de l'énergie libre de liaison par le modèle Prime/MM-GBSA.

Complexe	ΔG_{bind} (MM-GBSA) (kcal/mol)	ΔG_{bind} Vdw (MM-GBSA) (kcal/mol)	Énergie totale du complexe (kcal/mol)	ΔG_{bind} Hbond (kcal/mol)	Interaction cMet-ligand			
					Liaisons d'hydrogène	Électrostatique	van der Waals	Hydrophobique
cMet-H06	-20.05	-34.28	-10975.14	-1.64	Asp1222 (2.73 Å), His1202(2.25 Å)	Asp1222 (4.10 Å), Glu1127(4.34 Å)	Val1201, Phe1200, Arg1203, Asp1204, Glu1224, Gln1123	Met1131(4.52 Å), Ile1130 (5.41 Å)
cMet-H12	-20.32	-33.79	-10980.51	-1.84	His1202 (1.95 Å), Asp (1222), Arg1203 (3.12 Å)	Glu1127(4.37 Å), Asp1222(4.07 Å)	Glu1224, Gln1123, Val1201, Asp1204	Phe1200(4.87 Å), Ile1130(3.79 Å), Met1131(4.52 Å)
cMet-H30	-22.65	-47.49	-10976.08	-1.06	Lys1110 (3.15 Å)	-	Gln1123, Glu1127, Asp1222, Gly1224, Leu1157, Leu1140, Val1139, Val1220, Leu1195, Phe1200	Ile1130(5.40 Å), Ile1130(4.82 Å), His1202(5.47 Å), Phe1134(5.41 Å), Ala1221(4.27 Å), Ala1221(4.74 Å), Met1131(5.09 Å), Met1131(4.03 Å)
cMet-H36	-43.65	-53.80	-11019.61	-0.67	Lys1110(3.37 Å)	Glu1127(4.74 Å)	Val1139, Asp1222, Leu1140, Gly1128, Leu1157, Phe1223, Gly1224, Gln1123, Phe1200, Phe1134, Val1220	His1202(5.43 Å), Ile1130(4.77 Å), Ile1130(5.38 Å), Ala1221(4.84 Å), Ala1221(4.25 Å), Met1131(5.09 Å), Met1131(4.06 Å), Val1155(3.60 Å), Lys1110(4.68 Å)

Dans ce travail, nous avons réalisé cette exploration computationnelle critique sur la base du modèle de simulation de la dynamique moléculaire aidé par le package de Desmond du logiciel Schrödinger 2020-3. Dans cette démarche, nous nous sommes principalement appuyés sur l'analyse temporelle des systèmes cMet-inhibiteurs qui incluait les variations atomiques déterminées par le biais de la RMSD et de la RMSF, les profils des contacts cMet-inhibiteurs et les propriétés des inhibiteurs telles que la RMSD, le rayon de giration (rGyr), la surface moléculaire (MolSA), la liaison hydrogène intramoléculaire (intraHB), la surface accessible au solvant (SASA) et la surface polaire (PSA).

3.7.1 Analyses RMSD et RMSF

Pour analyser les fluctuations structurelles potentielles des nouveaux inhibiteurs de c-Met, les deux mesures de l'écart quadratique moyen (RMSD) et de la fluctuation quadratique moyenne (RMSF) ont été exploitées. La RMSD et la RMSF sont calculées par rapport au premier point de référence du squelette de la protéine (c-Met libre) après 100 ns de trajectoire temporelle. Les résultats enregistrés sont reproduits dans la **Figure III.C10(a-c)**.

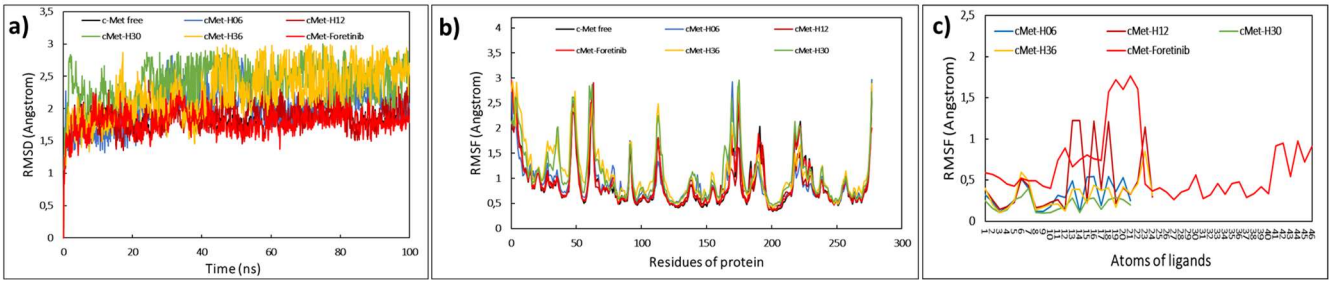


Figure III.C10. Trajectoires temporelles de RMSD (a) et de RMSF (b) des atomes du squelette des systèmes de c-Met à l'état libre et complexés avec H06, H12, H30, H36 et Foretinib. RMSF (c) trajectoires temporelles des inhibiteurs proposés H06, H12, H30, H36 et Foretinib dans la poche active de c-Met.

La **Figure III.C10a** montre que tous les systèmes étudiés - le c-Met libre et ses complexes - n'ont subi que des fluctuations mineures au début de la simulation, et qu'ils se sont stabilisés après 20 ns. Cette stabilité indique que les systèmes étaient alors à l'équilibre et le sont restés pendant toute la durée de la simulation. Les valeurs moyennes de RMSD pour le c-Met libre et ses complexes avec les molécules H06, H12, H30, H36 et le Foretinib étaient respectivement de 1.824, 2.011, 1.876, 2.420, 2.259 et 1.790 Å. Ces faibles niveaux de RMSD suggèrent que les complexes sont stables dans des conditions physiologiques. Les niveaux de RMSF des atomes de carbone Alpha des résidus c-Met, non complexés, et dans ces complexes avec les inhibiteurs H06, H12, H30, H36 et Foretinib ont montré des valeurs moyennes de 0.904, 0.940, 0.879, 1.086, 1.112 et 0.921 Å, respectivement (**Figure III.C10b**). Cela indique également la stabilité considérable de c-Met, qu'il soit non-complexé ou complexé avec les inhibiteurs proposés. D'autre part, de légères fluctuations ont pu être observées dans certains résidus, en particulier ceux situés en dehors des sites actifs de la c-Met, tels qu'Isoleucine (Ile1053, et Ile1118), Acide aspartique (Asp1054, et Asp1231), Acide glutamique (Glu1061), et Asparagine (Asn1100) (**Figure III.C10c**). Les fluctuations observées traduisent la flexibilité structurelle des ligands au sein de la poche active de c-Met.

3.7.2 Dynamique des contacts Protéine-ligand

La **Figure III.C11** montre les nombreuses interactions de contact entre H06, H12, H30, H36 et le Foretinib au cours de 100 ns dans des simulations de dynamique moléculaire. Il ressort des tracés générés que les liaisons hydrogène, hydrophobes, ioniques et les ponts d'eau sont parmi les principales interactions non covalentes qui préservent la stabilité des systèmes c-Met analysés. En outre, les diagrammes d'interaction 2D de cMet-ligand suggèrent qu'une majorité des interactions réalisées avec les résidus situés dans la poche spécifique de c-Met, telles que celles prévues par les simulations de Docking moléculaire et de MM-GBSA, sont conservées après 100 ns des simulations de MD. Par conséquent, les représentations visuelles des interactions protéine-ligand ont permis par la suite de comprendre davantage le mécanisme des interactions et de définir clairement l'environnement nécessaire aux ligands pour se positionner dans la poche spécifique de la protéine c-Met.

3.7.3 Profil des propriétés dynamiques des inhibiteurs de c-Met proposés

Les résultats obtenus par l'analyse des six propriétés -RMSD, rayon de giration (rGyr), surface moléculaire (MolSA), liaison hydrogène intramoléculaire (intraHB), surface accessible au solvant (SASA) et surface polaire (PSA) - des ligands étudiés (inhibiteurs) sur une période de 100 ns dans les simulations de dynamique moléculaire (DM) montrent une grande stabilité structurelle, comme les montres la **Figure III.C12**.

3.7.4 Analyse des propriétés thermodynamiques des systèmes cMet-inhibiteur

En soumettant les systèmes cMet-inhibiteurs à l'analyse de la qualité de la simulation, nous avons procédé à l'analyse du bilan de la qualité de la simulation MD en suivant la variation de l'énergie totale (E), de l'énergie potentielle (EP), de la température (T), de la pression (P) et du volume (V) au cours de la période de simulation de 100 ns. Les propriétés thermodynamiques du c-Met libre et complexé avec des inhibiteurs sont présentées dans la **Figure III.C13** et leurs valeurs moyennes sont présentées dans le **Tableau III.C9**. Les tracés thermodynamiques obtenus ont prouvé que les critères thermodynamiques examinés sont très voisins les uns par rapport aux autres et qu'ils sont très stables pendant toute la période de simulation de la DM de 100 ns.

Tableau III.C9. Score moyen des propriétés thermodynamiques des systèmes examinés.

	Propriétés Thermodynamiques				
Systèmes de c-Met	**Énergie totale (kcal/mol)**	**Énergie potentielle (kcal/mol)**	**Température (K)**	**Pression (bar)**	**Volume (Å^3)**
cMet-H06	-90109.333	-110165.406	298.674	0.707	327539.865
cMet-H12	-90087.429	-110144.368	298.683	0.653	327645.373
cMet-H30	-90340.770	-110444.722	298.685	1.157	328390.322
cMet-H36	-90340.788	-110448.703	298.691	1.204	328435.906
c-Met libre	-90302.277	-110373.609	298.685	1.207	327802.838
cMet-Foretinib	-89995.713	-110065.429	298.683	0.797	327581.987

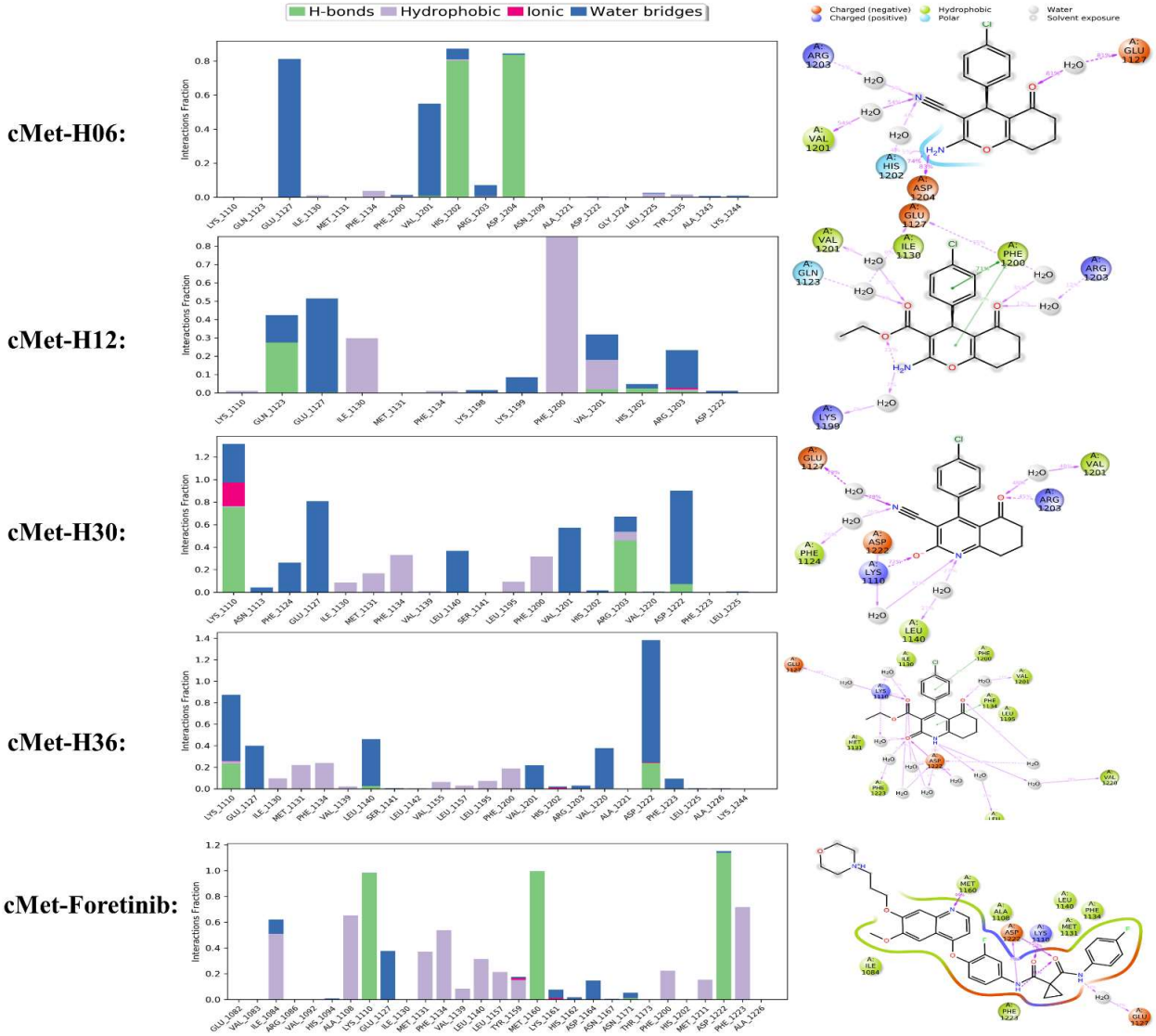


Figure III.C11. Diagrammes des interactions cMet-ligand générés par des simulations de dynamique moléculaire au cours de 100 ns.

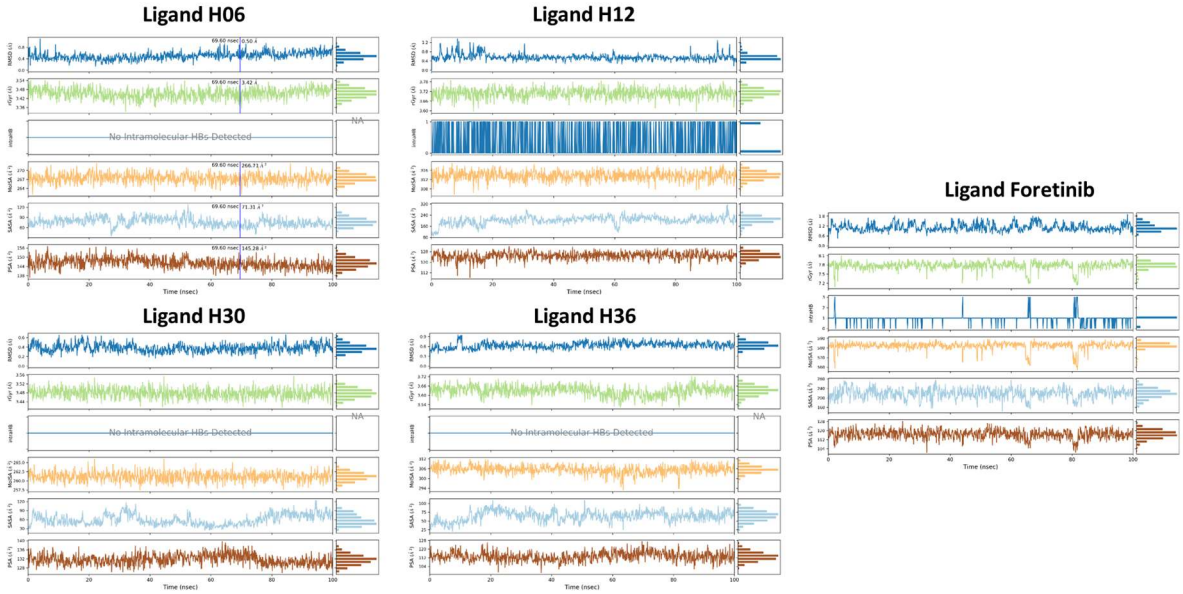


Figure III.C12. Profil des propriétés dynamiques des structures des ligands (H06, H12, H30, H36) avec la structure du c-Met pendent 100 ns de simulation de la DM.

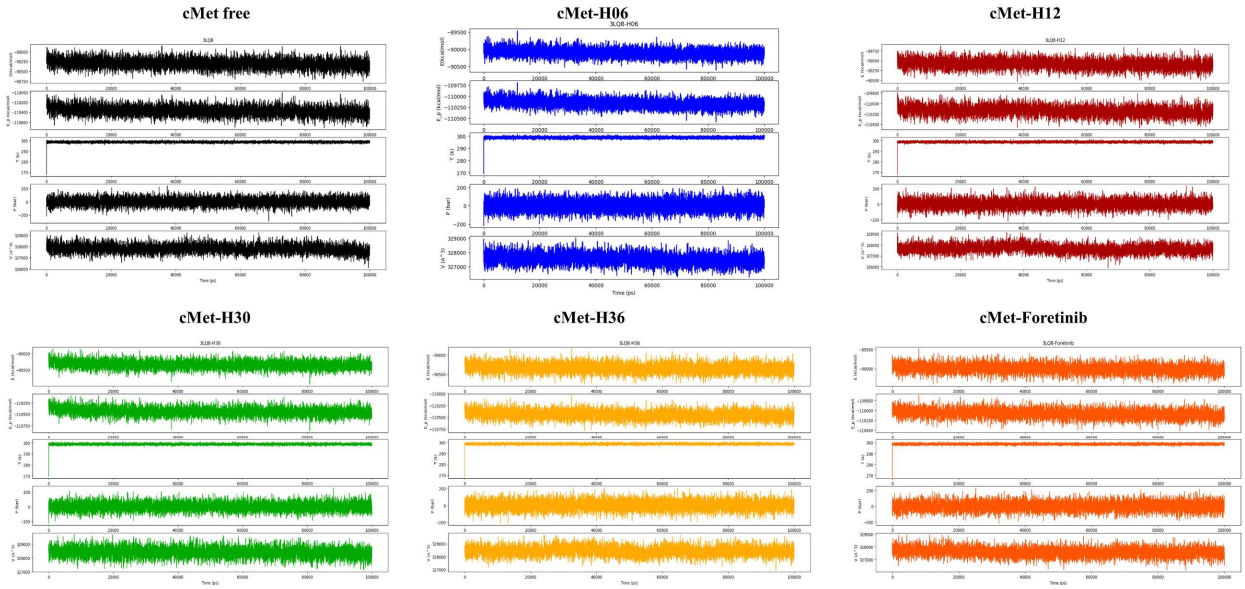


Figure III.C13. Qualité du protocole de dynamique moléculaire mis en œuvre à travers l'évaluation des propriétés thermodynamiques des systèmes cMet-inhibiteur.

Les résultats de l'analyse complète de la dynamique moléculaire effectuée dans le cadre de cette étude ont montré la forte stabilité entre la structure de la protéine c-Met et les structures des échantillons de petites molécules bioactives des dérivés du cyclohexane-1,3-dione H06, H12, H30 et H36 proposés comme médicaments. Cela confirme que les neuf dérivés de cyclohexane-1,3-dione optimisées dans cette étude sont considérées comme valide médicaments possibles (*in silico*) pour inhiber l'activité enzymatique de c-Met et atteindre ainsi la cible thérapeutique souhaitée contre le cancer du poumon non à petites cellules en tant que modèle.

4 Conclusion

Les dérivés du cyclohexane-1,3-dione montrent un grand intérêt pour la thérapie ciblée contre le cancer en raison de leurs propriétés bioactives contre une variété de lignées cellulaires cancéreuses, y compris le cancer du poumon non à petites cellules (CPNPC). Dans cet article, nous avons utilisé une approche hybride combinant des méthodes de conception de médicaments basées sur les ligands et sur la structure pour étudier quarante petites molécules hétérocycliques à base de cydohexane-1,3-dione pour leur activité inhibitrice contre la lignée cellulaire H460 du CPNPC. La modélisation statistique sur la base des modèles 2D-QSAR est utilisée pour modéliser et interpréter la relation structure-activité entre les composés étudiés. Les prédictions tirées du modèle MLR-QSAR, confirmées par le modèle ANN-QSAR, indiquent que l'activité inhibitrice des dérivés de la cyclohexane-1,3-dione est fortement corrélée aux descripteurs moléculaires physicochimiques et électroniques. La structure du composé 6d est identifiée comme un ligand optimal pour la conception et l'optimisation de nouveaux dérivés pour le développement et la conception de nouveaux agents médicamenteux destinés à lutter contre le CPNPC. De plus, l'activité enzymatique de c-Met est considérée comme un facteur important associé à la croissance cellulaire du cancer du poumon non à petites cellules et a fait l'objet d'une conception de médicaments basée sur la structure. Dans notre volonté de renforcer l'activité biologique, la biodisponibilité et les propriétés pharmacocinétiques/pharmacodynamiques de nouvelles molécules

basées sur la structure de référence 6d, nous avons extrait les propriétés structurales les plus importantes favorables à cet objectif. Pour ce faire, nous avons étayé les prédictions du modèle MLR-QSAR développé à partir des descripteurs moléculaires hybrides (0D-3D) par des visualisations 3D des cartes de contour MEP générées à partir de calculs DFT et des surfaces de contact protéine-ligand obtenues par ancrage de la structure 6d dans la poche active de la c-Met. La production de propriétés favorables du composé de référence 6d en vue de la modélisation/optimisation de la relation structure-activité menée dans le cadre de cette étude a conduit à la conception, à l'optimisation et au criblage de 36 nouvelles petites molécules considérées comme de nouveaux candidats inhibiteurs de la croissance des cellules cancéreuses par ciblage du récepteur de la tyrosine kinase c-Met. Après avoir défini le Foretinib comme standard, nous avons détecté par stratégie de Docking moléculaire les résidus d'acides aminés actifs suivants : **Phénylalanine** (Phe1200, et Phe1134), **Histidine** (His1202), **Acide Glutamique** (Glu1127), **Acide Aspartique** (Asp1222), **Méthionine** (Met1131) et **Lysine** (Lys1110), lesquels peuvent être adoptés comme des sites actifs spécifiques favorables pour inhiber l'activité enzymatique anormale de la c-Met par le biais de dérivés du cyclohexane-1, 3-dione, sujet de notre étude et de nos travaux de recherche et développement. Le Docking Moléculaire, les simulations Prime MM-GBSA, et la prédiction *in silico* Drug-like/ADME-Tox sont utilisés pour cribler et sélectionner les médicaments computationnels candidats (inhibiteurs spécifiques de c-Met) pour la thérapie du CPNPC. Les résultats fournis indiquent que parmi les 36 nouveaux dérivés de cyclohexane-1,3-dione conçus dans cette étude, 9 sont sélectionnés comme les meilleurs candidats médicaments contre le CPNPC par le biais du ciblage de l'activité enzymatique de c-Met. Finalement, des simulations de dynamique moléculaire sont appliquées pour valider et examiner le comportement structurel et la stabilité moléculaire des nouveaux candidats médicaments à l'intérieur de la poche active de la protéine c-Met, en fonction du temps (0 à 100 ns). Finalement, des simulations de dynamique moléculaire sont appliquées pour valider et examiner le comportement structurel et la stabilité moléculaire des nouveaux candidats médicaments à l'intérieur de la poche active de la protéine c-Met, en fonction du temps (0 à 100 ns). Globalement, le processus de conception de médicaments par voie *in silico* appliqué dans cette étude a permis de développer avec succès une nouvelle classe d'inhibiteurs de c-Met basés sur de petites molécules hétérocycliques de cyclohexane-1,3-dione comme médicaments potentiels pour le cancer du poumon non à petites cellules et d'autres formes éventuelles de cancer provoquées par l'activité enzymatique de la protéine c-Met.

CHAPITRE IV : EXPLORATION DES PHYTOCHIMIQUES POUR L'INHIBITION NON-COVALENTE DE LA PROLIFÉRATION CELLULAIRE CANCEREUSE : CIBLAGE MOLECULAIRE DU RÉCEPTEUR TYROSINE KINASE DE L'EGFR

Introduction

Le cancer du poumon et le cancer du sein restent parmi les types de cancer les plus fréquents et les plus dangereux, se classant au premier rang en termes de nombre de décès parmi les patients atteints de cancer chaque année [445]. Malgré les récents développements de l'industrie pharmaceutique et la diversité des méthodes thérapeutiques contre de nombreuses maladies réfractaires, une stratégie thérapeutique ou un médicament anticancéreux ayant une action définitive et le moins d'effets secondaires possibles n'a pas encore été mis sur le marché [446]. En raison des liens de dépendance entre le récepteur EGFR-TKD et le récepteur c-Met, les patients atteints de cancers causés par l'activité enzymatique de ces dernières sont traités avec des médicaments qui agissent en tant qu'inhibiteurs spécifiques de chaque récepteur. Plusieurs études ont montré que la combinaison des deux médicaments Foretinib et Erlotinib (inhibiteurs spécifiques de la tyrosine kinase) chez des patients atteints de cancer du poumon non à petites cellules donnait de bons résultats, mais non sans effets secondaires résultant de l'effet toxique de ces deux médicaments [41], [447]. Dans ce contexte, il est indispensable de renforcer les efforts et les ressources pour tenter de développer/concevoir de nouveaux médicaments efficaces et moins toxiques pour traiter le cancer en général et les cancers du sein et du poumon en particulier, causés par l'activité enzymatique explosive de la protéine EGFR-TKD. Pour contribuer à cet aspect, nous avons décidé, dans cette partie des études, de suggérer/développer des pistes thérapeutiques potentielles contre le cancer (poumon et sein) en ciblant l'EGFR-TKD. Cette initiative implique notamment la conception rationnelle *in silico* d'un nouveau médicament basé sur les structures de petites molécules biologiquement actives. En accord avec les orientations majeures liées aux objectifs de cette thèse, notamment la valorisation de la place des petites molécules biologiquement actives dans le pipeline de conception de médicaments, en particulier celles qui sont synthétisées (chimiquement) et/ou dérivées de substances naturelles à base de plantes (phytochimiques). Cela correspond également aux sous-objectifs liés à la conception et à l'identification d'un nouvel arsenal d'agents moléculaires potentiels en tant que clés pour le développement de médicaments contre le cancer au sens large. Dans ce chapitre de notre travail, nous avons concentré notre intérêt sur les molécules phytochimiques en tant que point de départ pour la mise en œuvre de stratégies hybrides de CADD pour concevoir et identifier de nouveaux inhibiteurs spécifiques de l'EGFR-TKD ayant un potentiel d'application dans la thérapie ciblée contre le cancer. Dans cette optique, le rôle des molécules phytochimiques dans la conception de médicaments a été reconnu comme étant extrêmement important [448], [449], [450], [451]. Les molécules phytochimiques sont des composés organiques naturels qui sont produits par des organismes vivants, généralement des plantes, et qui présentent des activités biologiques intéressantes. Ces composés peuvent être

utilisés pour développer des médicaments ou des produits qui peuvent être utilisés pour traiter ou prévenir diverses maladies. Les molécules phytochimiques sont particulièrement intéressantes car elles peuvent offrir une alternative aux composés organiques synthétisés, qui sont plus coûteux et moins efficaces [452]. Malgré leur importance, les molécules phytochimiques sont confrontées à plusieurs obstacles dans le processus de conception de médicaments [453]. Les principaux obstacles sont la variété des molécules et leur disponibilité limitée, les difficultés liées à l'extraction et à la purification des molécules, et les difficultés liées à la détermination de leurs mécanismes d'action, l'inaccessibilité, l'instabilité et l'insuffisance des données scientifiques et médicales [454]. Ces obstacles mettent en évidence l'importance de la recherche et des technologies d'extraction et de purification avancées pour faciliter l'utilisation des molécules phytochimiques dans la conception de médicaments.
Des progrès significatifs ont été réalisés ces dernières années pour surmonter ces obstacles. Les technologies modernes telles que l'extraction à l'aide de solvants, la chromatographie et l'analyse spectroscopique ont permis d'accéder aux molécules chimiques végétales et de mieux les comprendre. De plus, l'utilisation de modèles bio-informatiques et de modèles animaux a permis de mieux évaluer leur activité biologique et leur profil d'affabilité. Enfin, un autre défi important a été surmonté grâce à la mise en place de systèmes de gestion intégrés et à l'utilisation de technologies de l'information et de la communication [455]. Ces technologies ont contribué à améliorer la collecte, le traitement et l'analyse des données sur les molécules phytochimiques, ce qui a permis aux chercheurs et aux cliniciens d'accéder plus facilement à ces informations et d'accélérer le développement de médicaments à base de substances naturelle. Aujourd'hui, les molécules phytochimiques bioactives sont largement utilisées dans la conception de médicaments, et leur rôle dans le développement de médicaments sûrs et efficaces est plus important que jamais. En particulier, les molécules phytochimiques sont extrêmement importantes pour la découverte de médicaments anticancéreux [456], [457]. Ces composés peuvent offrir une solution non seulement pour traiter le cancer, mais aussi pour le prévenir et apporter de nouvelles solutions thérapeutiques contre diverses pathologies au-delà du cancer. Les molécules phytochimiques peuvent inhiber la prolifération des cellules cancéreuses et aider à réduire les effets indésirables des traitements anticancéreux [458]. Elles peuvent également être utilisées pour cibler les cellules cancéreuses et éviter les dommages aux cellules saines. Enfin, les molécules phytochimiques peuvent aider à améliorer l'efficacité des médicaments anticancéreux et à prévenir la résistance aux médicaments. De fait, les molécules phytochimiques offrent une alternative aux composés organiques synthétisés, et sont donc extrêmement importantes pour la découverte de médicaments anticancéreux [459]. Cette tendance est confirmée par les autorisations accordées par la FDA à plusieurs molécules à base de substances naturelles, comme les deux médicaments anticancéreux que sont l'Erlotinib (Tarceva) et le Gefitinib (Iressa), basés respectivement sur les plantes « Taxus brevifolia » et « Ancaria tomentosa » [460], [461]. Tous ces développements sont suffisants pour approfondir les études dans cette voie afin de découvrir d'autres classes de molécules phytochimiques candidates à une exploitation thérapeutique contre le cancer en tant que prototype. À cet égard, une récente étude *in vitro* dirigée par Min Zhoe et al [437]. a démontré la haute activité biologique de 51 composés isolés à partir de la plante *Magnolia officinalis* portant le radical 3-(4-aminobipyridin-1-yl) contre la croissance de lignées cellulaires de cancer du poumon non à petites cellules (CPNPC) induite par l'activité enzymatique

excessive de l'EGFR-TKD. De plus, ces nouveaux agents sont capables de renforcer l'action de l'autophagie, qui conduit à l'apoptose des cellules cancéreuses cibles. D'autre part, de nombreuses recherches récentes ont indiqué les multiples activités biologiques bénéfiques générées par les substances chimiques isolées à partir de la plante *Cannabis sativa*, notamment en ce qui concerne le traitement de certaines maladies immunitaires, d'anxiété, d'inflammation, de cancer et autres [462], [463], [464]. Cependant, étant donné que cette plante est principalement utilisée pour ses effets psychologiques plutôt que thérapeutiques, des controverses et des opinions divergentes entourent son utilisation. Il est donc nécessaire de mener plus d'études pour caractériser le *Cannabis sativa* et en comprendre davantage à des fins exclusivement thérapeutiques. Cela a pour but de comprendre ses potentialités liées à ses propriétés structurelles et à son activité biologique, ce qui est favorable à la conception de nouveaux médicaments capables de lutter contre les différentes pathologies, notamment le cancer. Dans ce contexte, une étude récente menée par Birenboim Matan et al. [463] rapporte que la combinaison des concentrations de Δ9-tétrahydrocannabinol (THC) et de cannabidiol (CBD) avec leurs profils terpéniques permet de classer les cultivars de cannabis médicinal de manière plus complète et plus précise. 279 échantillons de cannabis ont été extraits et caractérisés avec précision et classés en adoptant les modèles de classification du cultivar et du phénotype chimique [463]. L'étude a conclu qu'un mélange hybride de 50 molécules chimiques de cannabinoïdes isolés et de terpènes cannabis servait de meilleur modèle structurel décrivant la diversité structurelle des molécules isolées de *Cannabis sativa L*[463]. Dans ces scénarios, nous avons focalisé notre réflexion dans ce chapitre sur les molécules phytochimiques issues de *Magnolia officinalis* et de *Cannabis sativa L*. comme références chimiothèques pour rationaliser la conception et la modélisation d'un nouveau médicament contre le cancer. Pour ce faire, nous avons mené des études computationnelles *in silico* à grande échelle sur les données moléculaires dérivées de ces deux types de plantes. Ceci afin d'aider à la conception/développement de nouveaux agents médicamenteux potentiels qui inhibent l'activité enzymatique anormale de la protéine EGFR-TKD et suppriment la croissance des lignées cellulaires tumorales résultant de son activité. Par conséquent, les travaux menés dans cette partie de ma thèse peuvent apporter une contribution prometteuse à l'avancement de la recherche en oncologie en général, et à la découverte de nouvelles générations d'inhibiteurs de tyrosine kinase basés sur des sources de plantes médicinales traditionnelles qui sont disponibles dans notre écosystème. La **Figure IV.1** présente certaines structures moléculaires de substances phytochimiques isolées à partir de la plante de *Magnolia officinalis* et de *Cannabis sativa L.* que nous avons considérées dans cette partie de notre travail comme des inhibiteurs potentiels de l'EGFR-TKD pour la thérapie ciblée du cancer (CPNPC et du sein comme modèles). À cet égard, nous avons utilisé les deux médicaments Erlotinib et Tamoxifen comme références dans les évaluations et les comparaisons visant à établir des conclusions et des hypothèses concernant la compatibilité des structures des phytochimiques de *Magnolia officinalis* et *Cannabis sativa L.* pour devenir des agents médicamenteux candidats pour la thérapie ciblée contre le cancer.

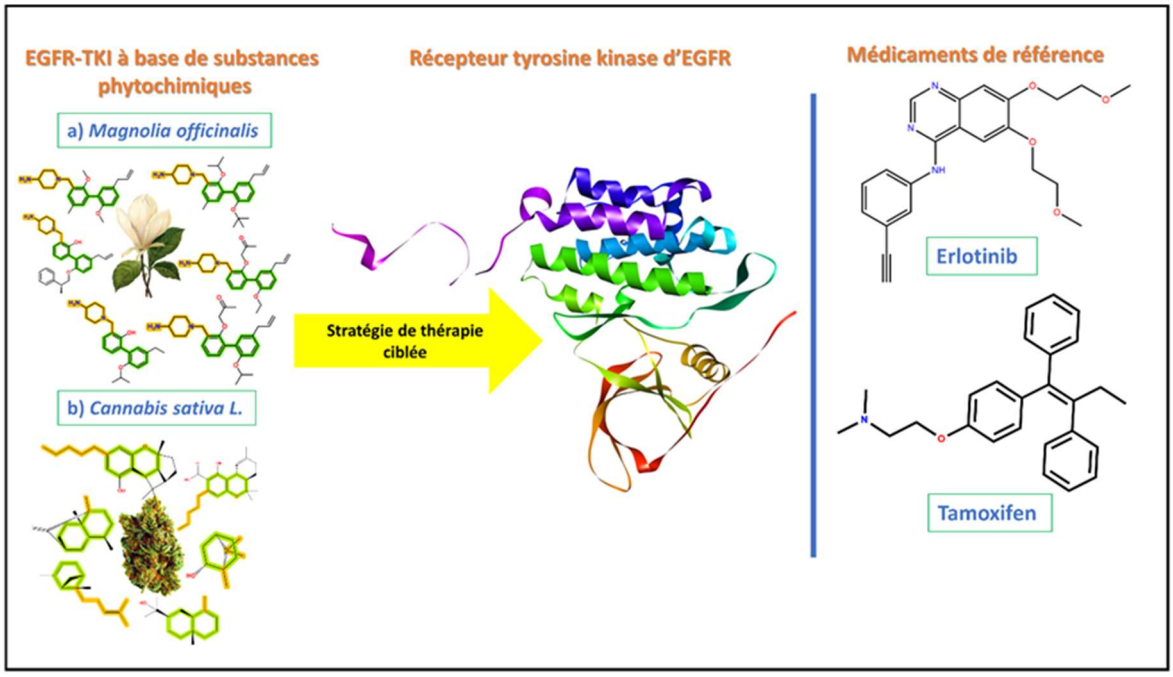


Figure IV.1. *Magnolia officinalis* (a), *Cannabis sativa L.* (b) : phytochimiques candidats pour le développement d'une nouvelle génération d'inhibiteurs EGFR-TKD comme stratégie de thérapie ciblée contre le cancer.

A. Modélisation assistée par ordinateur des composés dérivés de *Magnolia officinalis* pour le ciblage de l'EGFR en vue du traitement du cancer du poumon non à petites cellules

1 Introduction

Le cancer du poumon non à petites cellules (CPNPC) est le type de cancer du poumon le plus courant, avec 85 % du nombre total de personnes dans ce cas [445], [465]. Les traitements actuels de la maladie, tels que la chimiothérapie, sont efficaces, mais ils s'accompagnent d'effets secondaires indésirables dus à la destruction massive du système immunitaire [447]. De nombreuses tentatives sont actuellement menées pour trouver des traitements plus efficaces contre ce type de cancer, notamment des médicaments qui inhibent le récepteur du facteur de croissance épidermique (EGFR) . L'un de ces médicaments, appelé Erlotinib, a fait l'objet d'études approfondies, mais il n'est efficace que chez les patients présentant une mutation active de l'EGFR, ce qui le rend moins utile pour ceux qui n'en ont pas [439], [467], [468]. D'autres recherches ont été menées sur la substance naturelle isolée de *Magnolia officinalis.*, qui présente un potentiel en tant qu'agent anti-inflammatoires, antioxydants, antidépresseurs, et anticancéreux, et qui a donné des réponses encourageantes dans l'inhibition de la lignée cellulaire H460 chez les patients atteints de cancer du poumon non à petites cellules (CPNPC) [473], [474], [475], [476]. Selon l'étude de Min Zhoe et al[437], les 51 molécules des dérivés du Magnolol peuvent favoriser l'action de l'autophagie, qui entraîne la mort cellulaire programmée des cellules cancéreuses, ce qui donne à ce composé un fort pouvoir de développement de nouveaux agents thérapeutiques avec moins d'effets secondaires. Les résultats de l'évaluation biologique *in vitro* et *in vivo* de l'activité anticancéreuse de 51 dérivés de Magnolol portant la structure 3-(4-aminopiperidin-1-yl)méthyl ont indiqué leur activité biologique considérable (**Tableau IV.A1**). Malgré ces résultats intéressants qui soutiennent l'hypothèse de l'exploitation de la structure du Magnolol dans la conception de médicaments anticancéreux, cette étude reste presque brute et

nécessite d'autres investigations et évaluations à plus grande échelle et avec plus de précision. Sur cette base, et en raison de l'absence d'investigations et de contrôles relatifs à la détection des structures moléculaires les plus appropriées pour une conception efficace et saine de médicaments contre le cancer en tant que modèle, nous nous efforçons donc de compléter cette étude de façon qu'elle soit la plus complète possible et à ce qu'elle soit la plus exhaustive possible. Pour notre part, nous nous attachons à compléter et à optimiser les résultats de cette étude en impliquant des analyses plus approfondies et plus complètes *in silico* à l'aide d'approches CADD hybrides qui dépendent à la fois du ligand et de la structure. Ceci a pour but de décrypter la relation structure-activité que les dérivés du Magnolol développent contre les lignées cellulaires du cancer du poumon non à petites cellules en tant que modèle. La caractérisation quantitative de la relation structure-activité nous aidera à dégager et à définir les propriétés et les caractéristiques structurelles de base qui doivent être présentes dans les dérivés du Magnolol pour assurer une activité biologique élevée et une pertinence structurelle en vue de la conception de nouveaux médicaments. De même, grâce aux explorations *in silico*, nous pouvons identifier de manière approfondie la molécule de référence optimale pour les dérivés du Magnolol qui peut être adoptée pour modéliser de nouvelles molécules et prédire leurs activités biologiques inhibitrices vis-à-vis du CPNPC, ainsi que pour prédire leurs profils de propriétés similaires à celles des médicaments, leurs propriétés pharmacocinétiques et pharmacodynamiques, et leurs interactions avec le récepteur biologique cible, comme le récepteur EGFR-TKD, considéré comme un modèle dans cette étude.

2 Stratégie de conception de médicaments assistée par ordinateur développée et appliquée

Dans le cadre de notre tentative d'enrichir la base de données de molécules computationnelles candidates à la conception de nouveaux médicaments contre le cancer, nous nous intéressons dans cette étude au rôle des molécules phytochimiques bioactives. Ceci afin de compiler un arsenal diversifié de molécules dont les propriétés liées à la structure et à l'activité ont été validées et qui conviennent à la conception de médicaments et à leur application dans le cadre d'une thérapie ciblée contre le cancer en tant que modèle. Sur cette base, dans cette étude, nous avons procédé à une étude *in silico* à grande échelle sur une série de 51 structures de dérivés du Magnolol par le biais d'une stratégie de CADD hybride rationalisée. La **Figure IV. A1** présente un schéma indicatif de la méthodologie CADD hybride appliquée dans cette étude afin d'examiner, de concevoir, d'améliorer et d'identifier les structures optimales des dérivés du Magnolol susceptibles d'être exploitées dans la conception d'un nouveau type de thérapie contre le cancer du poumon non à petites cellules du point de vue qualitatif et contre d'autres types de cancers provoqués par l'expression anormale de l'activité enzymatique de la protéine EGFR-TKD. Concernant tous les détails relatifs aux méthodes et outils utilisés dans la mise en œuvre de chaque phase de cette étape, ils ont été présentés et détaillés dans le deuxième chapitre de cette thèse, ainsi que dans la version électronique de l'article que nous avons publié à ce sujet [324].

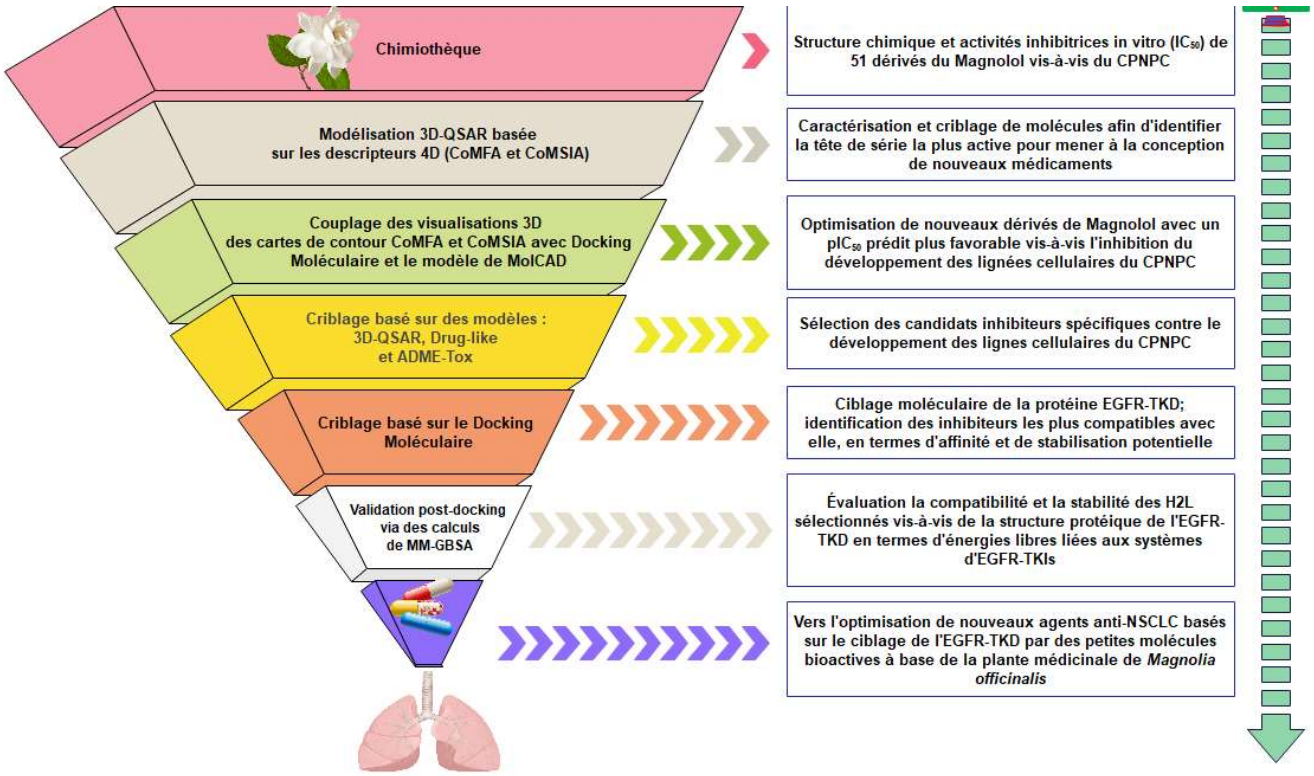
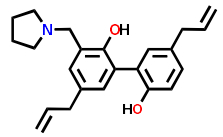
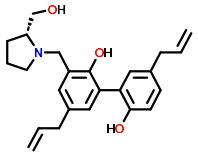
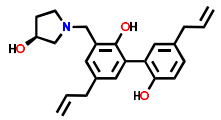
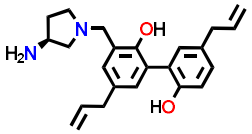
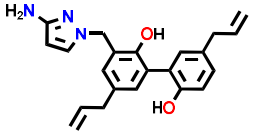
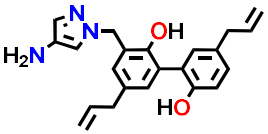
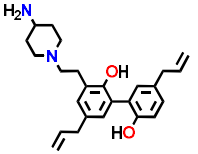
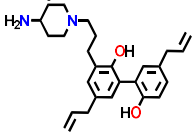
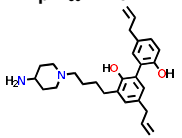
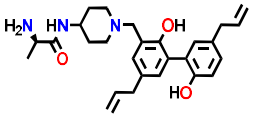
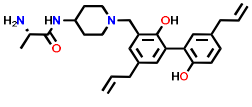
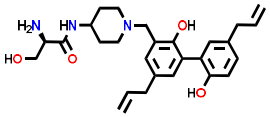
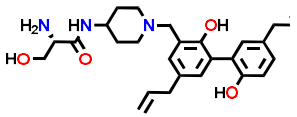
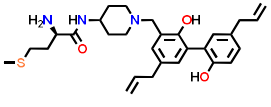
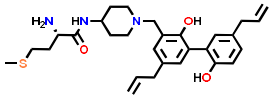
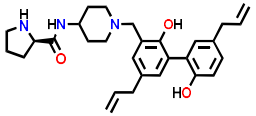
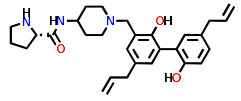
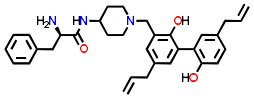
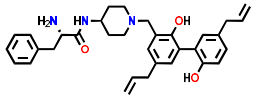
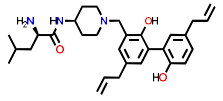
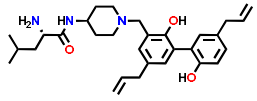
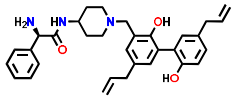
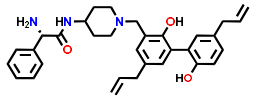
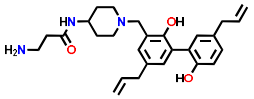
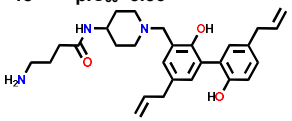
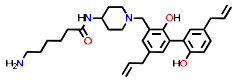
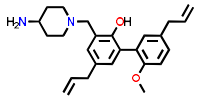
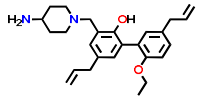
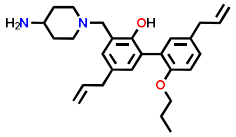
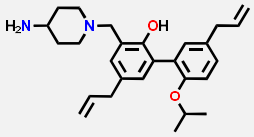
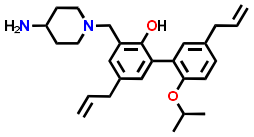
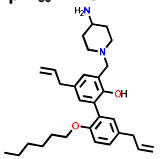
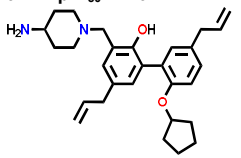
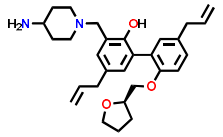
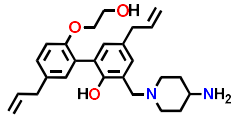
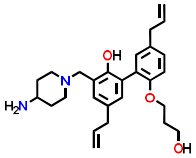
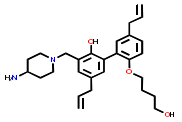
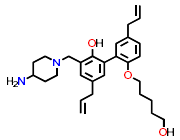
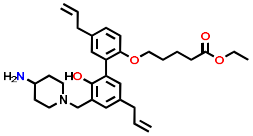
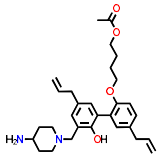


Figure IV. A1. Stratégie de CADD hybride proposée et appliquée pour le criblage, la caractérisation et l'optimisation de nouvelles molécules à base de Magnolol, candidates à la conception d'un nouveau médicament anti-CPNPC ciblant spécifiquement l'EGFR-TKD.

Tableau IV. A1. Données structure-activité des dérivés de Magnolol basés sur la plante médicinale *Magnolia officinalis* contre les lignées cellulaires du CPNPC obtenues par des essais *in vitro*.

						
01 pIC$_{50}$=4.70	02 pIC$_{50}$=4.70	03 pIC$_{50}$=4.70	04 pIC$_{50}$=4.70	05 pIC$_{50}$=4.70	06 pIC$_{50}$=4.70	07 pIC$_{50}$=4.70
						
08 pIC$_{50}$=4.70	09 pIC$_{50}$=4.70	10 pIC$_{50}$=4.70	11 pIC$_{50}$=4.76	12 pIC$_{50}$=5.08	13 pIC$_{50}$=5.18	14 pIC$_{50}$=4.70
						
15 pIC$_{50}$=4.70	16 pIC$_{50}$=4.70	17 pIC$_{50}$=4.70	18 pIC$_{50}$=5.36	19 pIC$_{50}$=4.90	20 pIC$_{50}$=4.70	21 pIC$_{50}$=4.70
						
22 pIC$_{50}$=5.35	23 pIC$_{50}$=4.80	24 pIC$_{50}$=4.70	25 pIC$_{50}$=4.70	26 pIC$_{50}$=4.70	27 pIC$_{50}$=5.37	28 pIC$_{50}$=5.90
						
29 pIC$_{50}$=5.31	30 pIC$_{50}$=6.07	31 pIC$_{50}$=5.06	32 pIC$_{50}$=5.02	33 pIC$_{50}$=5.23	34 pIC$_{50}$=5.56	35 pIC$_{50}$=5.21
						
36 pIC$_{50}$=5.18	37 pIC$_{50}$=5.06	38 pIC$_{50}$=4.99	39 pIC$_{50}$=4.70	40 pIC$_{50}$=4.70	41 pIC$_{50}$=4.70	42 pIC$_{50}$=5.21
43 pIC$_{50}$=5.16	44 pIC$_{50}$=5.44	45 pIC$_{50}$=5.54	46 pIC$_{50}$=5.95	47 pIC$_{50}$=5.21	48 pIC$_{50}$=5.84	49 pIC$_{50}$=5.88
50 pIC$_{50}$=5.54	51 pIC$_{50}$=5.76					

3 Résultats et discussion

3.1 Analyses structure-activité basée sur 3D-QSAR

3.1.1 Caractérisation biomoléculaire d'inhibiteurs de l'EGFR par CoMFA et CoMSIA

Après avoir dessiné et optimisé la structure géométrique et énergétique de la base de données des substances bioactives des dérivés de Magnolol, nous avons divisé les données générées en deux ensembles en utilisant la technique de division par diversité proposée par le logiciel SYBYL-X 2.1.1. La meilleure division obtenue comprenait les particules (01, 02, 03, 04, 05, 06, 07, 08, 09, 11, 12, 14, 15, 17, 19, 21, 22, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 et 51) en ensemble d'entraînement pour développer le modèle 3D-QSAR, tandis que les molécules (10, 13, 14, 16, 18, 20 et 23) sont désignées comme échantillons pour tester le modèle élaboré. Ensuite, en utilisant la technique de distillation à corps commun (alignement sous-structure), nous alignons les différentes conformations 3D bioactives des 51 molécules actives étudiées avec la molécule prototype (30, pIC_{50}=6.07) (**Figure IV. A2**).

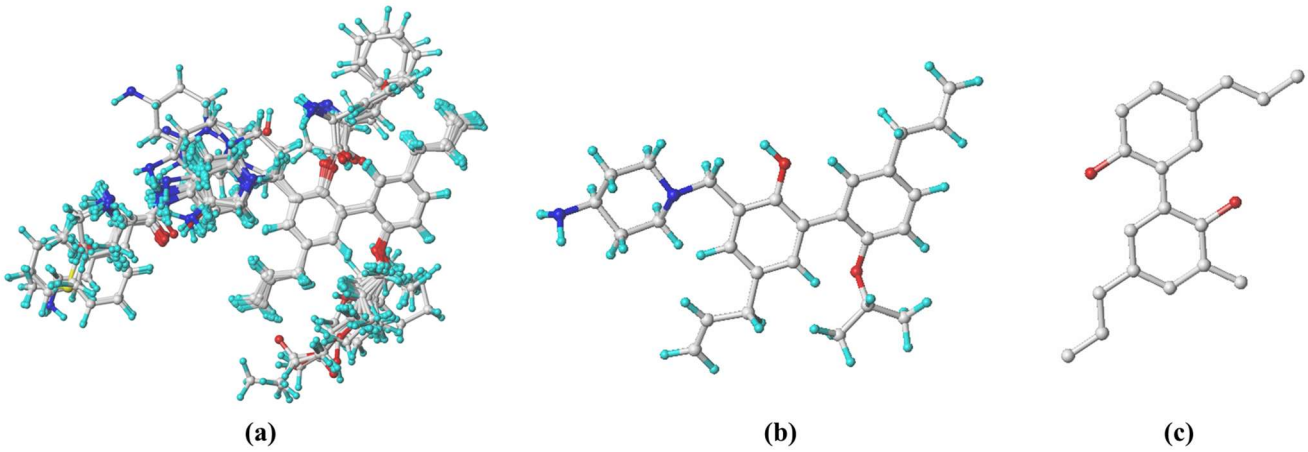


Figure IV. A2. (a) Base de données de dérivés de Magnolol alignés, (b) Structure de la molécule de référence "30", c) Unité de base commune.

La **Figure IV. A2** montre clairement que les données moléculaires des dérivés du Magnolol soumis au processus d'alignement sont correctement alignées avec la structure de la conformation de référence active n° 30. Les données sur les descripteurs des champs CoMFA et CoMSIA tirées de cette situation d'alignement parfait seront donc très utiles pour la construction d'un modèle 3D-QSAR fiable. Ce dernier nous permettra de formuler des hypothèses de pharmacopée dans l'espace 3D sur la relation structure-activité de ces molécules et de les attribuer à la conformation structurelle de référence qui présente l'activité biologique la plus élevée observée *in vitro*. Sur la base des descripteurs moléculaires stériques et électrostatiques (S et E) générés par les analyses du CoMFA, des analyses stériques (S), électrostatiques (E), hydrophobiques (H), des donneurs de liaisons d'hydrogène et des accepteurs de liaisons hydrogène (D et A) générés par les analyses du CoMSIA, nous avons élaboré le modèle 3D-QSAR et l'avons évalué à l'aide de la méthode PLS (**Tableau IV. A2**).

Tableau IV. A2. Analyse statistique selon la méthode PLS.

Models	Les seuils de performance des modèles développés selon les seuils d'acceptation								Contribution des descripteurs Fractions				
	Q^2	R^2	SEE	F	N	R^2_{pred}	R^2_{bs}	SEE_{bs}					
	>0.5	>0.6	Faible	Élevé	3<N<6	>0.6	>0.6	Faible	S	E	H	D	A
CoMFA/SE	0.67	0.90	0.158	68.8	4	0.863	0.92	0.031	0.523	0.477	-	-	-
CoMSIA/SEHDA	0.58	0.82	0.189	45.7	4	0.783	0.86	0.058	0.098	0.231	0.094	0.236	0.341

Q^2 est le coefficient de corrélation de validation croisée, **R^2** est le coefficient de corrélation non validé, **SEE** est l'erreur standard d'estimation, F est la valeur du F-test, **N** est le nombre optimal de composantes déterminé par PLS-loocv , **R^2_{pred}** est le coefficient de corrélation prédictif obtenu par validation externe, **R^2_{bs}** est le coefficient de corrélation après 100 cycles de bootstrap (ré-échantillonnage bootstrap), **SEE_{bs}** est l'erreur standard d'estimation pour 100 cycles de bootstrap.. Le terme "**Fraction**" désigne les contributions des champs S, E dans le modèle CoMFA et S, E, H, D, A dans le modèle CoMSIA.

À l'aide des résultats présentés dans le Tableau 1 nous pouvons constater que les modèles CoMFA et CoMSIA ont montré d'excellentes performances, avec des coefficients R^2 de 0.90 et 0.82, des valeurs Q^2 de 0.672 et 0.58, des valeurs F de 68.832 et 45.727, et des faibles SEE (0.158 et 0.189), respectivement, lors de l'utilisation du nombre optimal de 4 composants. De plus, les valeurs élevées du coefficient R^2_{pred} (0.86 et 0.78) indiquent que les modèles ont une excellente puissance prédictive externe. Les analyses Bootstrap ont également révélé une excellente performance grâce aux valeurs élevées de R^2_{bs} (0.926 et 0.865) et aux faibles valeurs de SEE_{bs} (0.031 et 0.058) des modèles CoMFA et CoMSIA respectivement. Les fractions de contribution des champs stérique (S) et électrostatique (E) pour le modèle CoMFA étaient respectivement de 0.523 et 0.477, indiquant que l'interaction entre ces champs ayant un effet significatif sur l'activité biologique, et que le champ stérique contribue plus que le champ électrostatique. Pour le modèle CoMSIA, les contributions du stérique (S), électrostatique (E), hydrophobe (H), donneur de liaison hydrogène (D) et accepteur de liaison hydrogène (A) étaient respectivement de 0.098, 0.231, 0.094, 0.236 et 0.341. Cette convergence montre que la contribution des cinq champs (S, E, H, D, A) a une forte corrélation avec l'activité biologique inhibitrice anticancéreuse (pIC_{50}) des composés dérivés de Magnolol vis-à-vis les lignes cellulaires du CPNPC. Dans l'ensemble, le modèle 3D-QSAR que nous avons généré dans cette étude à partir des descripteurs CoMFA et CoMSIA présente une bonne signification statistique et un bon pouvoir prédictif et est capable d'expliquer la relation structure-activité de la base de données biomoléculaires ciblée dans cette étude. Ces caractéristiques statistiquement significatives peuvent également être prouvées par la distribution homogène des valeurs de pIC50 observées *in vitro* par rapport à celles prédites par le modèle 3D-QSAR développé *in silico* sur la base des descripteurs de champ CoMFA et CoMSIA (**Figure IV. A3a**). Bien que tous les indicateurs de validation internes et externes indiquent que le modèle 3D-QSAR développé est statistiquement acceptable, cela n'exclut pas la possibilité d'une concordance aléatoire entre la structure et l'activité biologique des molécules étudiées. Par conséquent, il est nécessaire de procéder encore à des inspections plus rigoureuses pour écarter cette proposition. C'est pourquoi, dans cette étude, nous avons effectué deux autres tests, à savoir : Y-randomisation (**Tableau IV. A3**) et la définition du domaine d'applicabilité du modèle 3D-QSAR proposé (**Figure IV. A3b**).

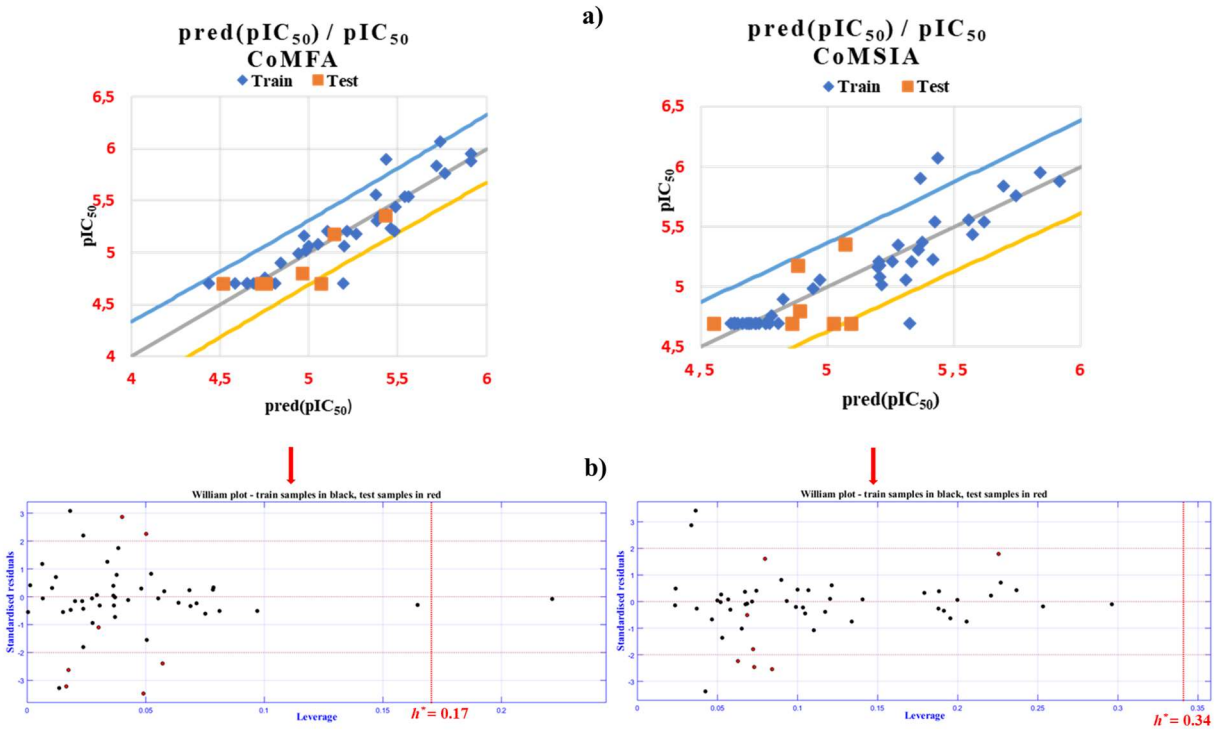


Figure IV. A3. (a) Distribution des valeurs pIC_{50} obtenues *in vitro* pour les dérivés du Magnolol par rapport à celles prédites par le modèle 3D-QSAR *in silico*. (b) Diagrammes de Williams des domaines d'applicabilité du modèle 3D-QSAR élaboré à l'aide des descripteurs CoMFA et CoMSIA.

Tableau IV. A3. Résultats du test de Y-randomisation.

		CoMFA			CoMSIA/SEHD		
Modèle		**Q^2_{yrand}**	**R^2_{yrand}**	**$^cR^2_p$**	**Q^2_{yrand}**	**R^2_{yrand}**	**$^cR^2_p$**
Original		**0.680**	**0.960**	-	**0.630**	**0.950**	-
Aléatoire	I1	0.320	0.584	0.528	0.289	0.487	0.522
	I2	0.126	0.534	0.569	0.476	0.513	0.502
	I3	-0.189	0.424	0.650	0.126	0.489	0.521
	I4	-0.114	0.413	0.657	0.237	0.478	0.529
	I5	-0.312	0.376	0.682	-0.478	0.382	0.598
Moyenne		-0.034	0.466	**0.617**	0.130	0.470	**0.534**

Grâce au test de Y-randomisation, nous pouvons constater que les valeurs Q^2 et R^2 des modèles randomisés après 5 mélanges aléatoires sont inférieures à ceux des modèles originaux CoMFA et CoMSIA; nous avons également obtenu des valeurs $^cR^2_p$ pour les modèles randomisés toutes supérieures à 0.5. Ceci signifie que les bons résultats obtenus par les modèles originaux pour CoMFA et CoMSIA ne sont pas dus au hasard. D'autre part, à partir des diagrammes Williams qui définissent les domaines d'applicabilité du modèle 3D-QSAR / CoMFA-CoMSIA, il apparaît que toutes les valeurs pIC_{50} correspondant aux ensembles d'entraînement et de test sont inférieures aux valeurs seuil de l'effet de levier (h^* = 0.17 pour CoMFA et h^* = 0.34 pour CoMSIA). Cela signifie que tous les composés appartiennent au domaine d'applicabilité, à l'exception du composé "05" de l'ensemble d'entraînement, qui se trouve en dehors du domaine d'application du modèle CoMFA. Nous pouvons supposer que la valeur d'activité pour ce composé n'a pas été correctement prédite ou qu'elle est due à la disponibilité de données expérimentales incorrectes pour cette molécule. De plus, toutes les valeurs pIC_{50} prédites par les modèles 3D-QSAR dans les ensembles de test s'inscrivent dans le domaine d'applicabilité

déterminé, aucun point aberrants ne faisant donc partie de cet ensemble. Tous les résultats obtenus à partir de la Y-randomisation et du domaine d'application confirment la haute performance et la validité du 3D-QSAR proposé pour rationaliser la conception et l'optimisation de nouveaux dérivés du Magnolol et la prédiction de leur activité inhibitrice biologique contre le cancer du poumon non à petites cellules (CPNPC) avec confiance et fiabilité. Après avoir réussi à décoder la relation structure-activité des dérivés du Magnolol par les descripteurs moléculaires CoMFA et CoMSIA, nous passons maintenant à l'étape du décodage des hypothèses de pharmacopée générées. Nous pouvons réaliser cette procédure en générant des cartes de contours des représentations des interactions des champs de force S, E, H, A, et D sur la structure #30 de conformation de plus forte pIC_{50} désignée comme référence (**Figure IV. A4**). L'analyse de ces cartes de contour nous permet d'extraire les propriétés et les sites dans l'espace 3D qui sont favorables à l'amélioration de l'activité biologique et de l'affinité des nouvelles molécules vis-à-vis de la cible moléculaire souhaitée afin d'atteindre l'objectif thérapeutique.

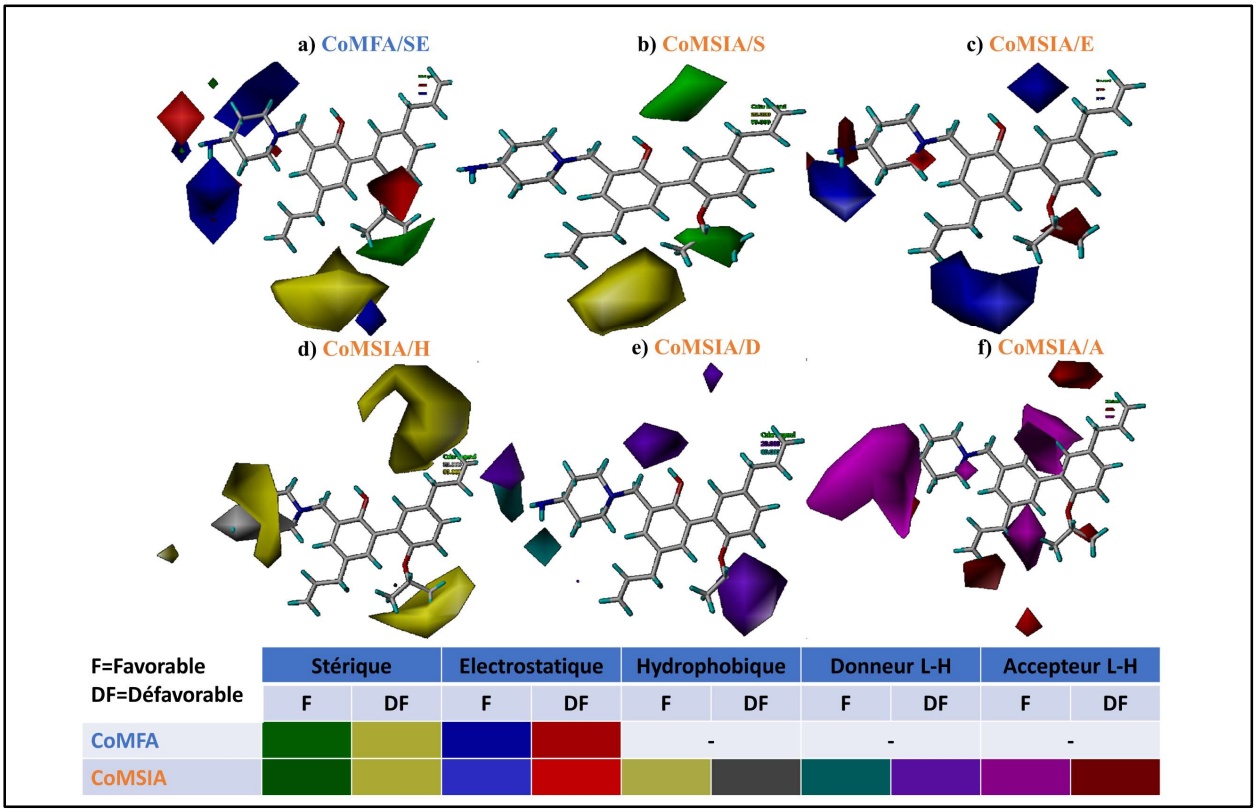


Figure IV. A4. Cartes de contours générées à partir des analyses CoMFA (a) et CoMSIA (b-f) des interactions stériques, électrostatiques, hydrophobiques, donneurs de liaisons H, accepteurs de liaisons H (favorables et défavorables) pour optimiser l'activité biologique inhibitrice de la structure de référence des dérivés de Magnolol examinés contre le cancer du poumon non à petites cellules (CPNPC).

L'analyse des cartes de CoMFA/Stérique (**Figure IV. A4a**) suggère que le contour jaune autour du groupe prop-1-ene indique que le remplacement par un petit groupe à cet endroit augmenterait l'activité inhibitrice biologique du dérivé de Magnolol 30. Le site **R1** est donc une région favorable avec le potentiel d'améliorer l'activité du composé. Le contour vert autour du groupe lié à l'oxygène en position para du cycle aromatique de la structure de référence Magnolol indique que la substitution par un groupe plus grand à cet endroit a le potentiel d'augmenter l'activité inhibitrice biologique. Ce site est étiqueté **R2** et peut être confirmé par l'évolution de la valeur d'activité biologique *in vitro* des composés 27, 28 et 30, les groupes plus grands à **R2** ayant des valeurs de pIC_{50} plus élevées. Pour l'analyse

CoMFA/Électrostatique, les grands contours bleus autour du groupe 3-(4-aminopiperidin-1-yl) methyl indiquent que les charges positives de ce groupe contribuent à l'activité biologique des dérivés de Magnolol. Cela est encore plus attesté par les valeurs de pIC50 associées au groupe 3-(4-aminopiperidin-1-yl) methyl (voir **Tableau IV. A1**), ce groupe étant donc considéré comme important dans la structure biomoléculaire des dérivés de Magnolol. Selon l'analyse des cartes de contours CoMSIA/Stérique dans la **Figure IV. A4b**, les contours verts et jaunes indiquent les sites favorables (étiquetés **R3**) pour des substitutions visant à augmenter la valeur du pIC_{50} des dérivés de Magnolol contre le CPNPC. Le **Table IV. A1** confirme que de telles substitutions conduisent à une meilleure activité inhibitrice biologique ; par exemple, 47 (pIC_{50} = 5.21) <48 (pIC_{50} = 5.84) <46 (pIC_{50} = 5.95). De la même manière, les cartes de contours CoMSIA / Électrostatique (**Figure IV. A4c**) montrent des contours bleus entourant les sites **R1**, **R2** et **R3**, indiquant que ces sites sont également adaptés pour des groupes chargés positivement. Les cartes de contours CoMSIA / Hydrophobiques (**Figure IV. A4d**) montrent que les contours jaunes dominent les contours blancs, indiquant une préférence pour les groupes hydrophobes, qui s'assemblent avec la protéine cible pour l'action inhibitrice. De plus, les cartes CoMSIA / DA (Donneur et Accepteur H-bond), (**Figures IV.4e-f**) suggèrent que le groupe 3-(4-aminopiperidin-1-yl) méthyle est favorable aux liaisons hydrogène, et que les sites **R2** et **R3** conviennent aux groupements accepteurs, tandis que R1 convient aux groupements donneurs. À la lumière de ces données sur la relation structure-activité, une expérience de Docking moléculaire sur la cible EGFR-TKD a été réalisée pour identifier les caractéristiques structurelles idéales des dérivés de Magnolol pour une activité inhibitrice améliorée contre le CPNPC. Ainsi, avant la conception de nouveaux inhibiteurs efficaces contre le CPNPC, nous avons obtenu des informations sur les caractéristiques structurelles des dérivés de Magnolol qui sont favorables à une plus grande activité biologique et à l'affinité avec la protéine cible.

3.1.2 Mécanisme d'affinité et de compatibilité des dérivés de Magnolol avec l'EGFR-TKD

À l'aide d'une stratégie de Docking moléculaire semi-flexible, nous ancrons le dérivé de référence #30 dans la poche active de l'EGFR-TKD (code PDB : 1M17). Nous réalisons cette procédure en tant qu'étape de soutien qui nous permet de mieux découvrir les profils d'affinité et de compatibilité des dérivés de Magnolol bioactifs avec la structure protéique de l'EGFR-TKD. En outre, elle sert de guide pour vérifier les propriétés des sites **R1**, **R2**, **R3** et 3-(4-aminopipéridin-1-yl)–méthyle modélisés par le modèle 3D-QSAR suite aux analyses de contour maps de CoMFA et de CoMSIA. Avant d'effectuer cette étape critique, il faut d'abord valider la pertinence du modèle de Docking moléculaire mis en œuvre ainsi que la qualité des outils bio-informatiques utilisés à cette fin. C'est pourquoi, dans ce travail, nous avons adopté le protocole de re-Docking moléculaire comme méthode de validation du modèle de Docking moléculaire implémenté via le logiciel AutoDock Vina. La **Figure IV. A5** présente des visualisations 3D et 2D des modes de Docking de l'Erlotinib dans la poche active EGFR-TKD (1M17), ainsi que les interactions des co-ligands de l'Erlotinib dans le complexe natif 1M17.

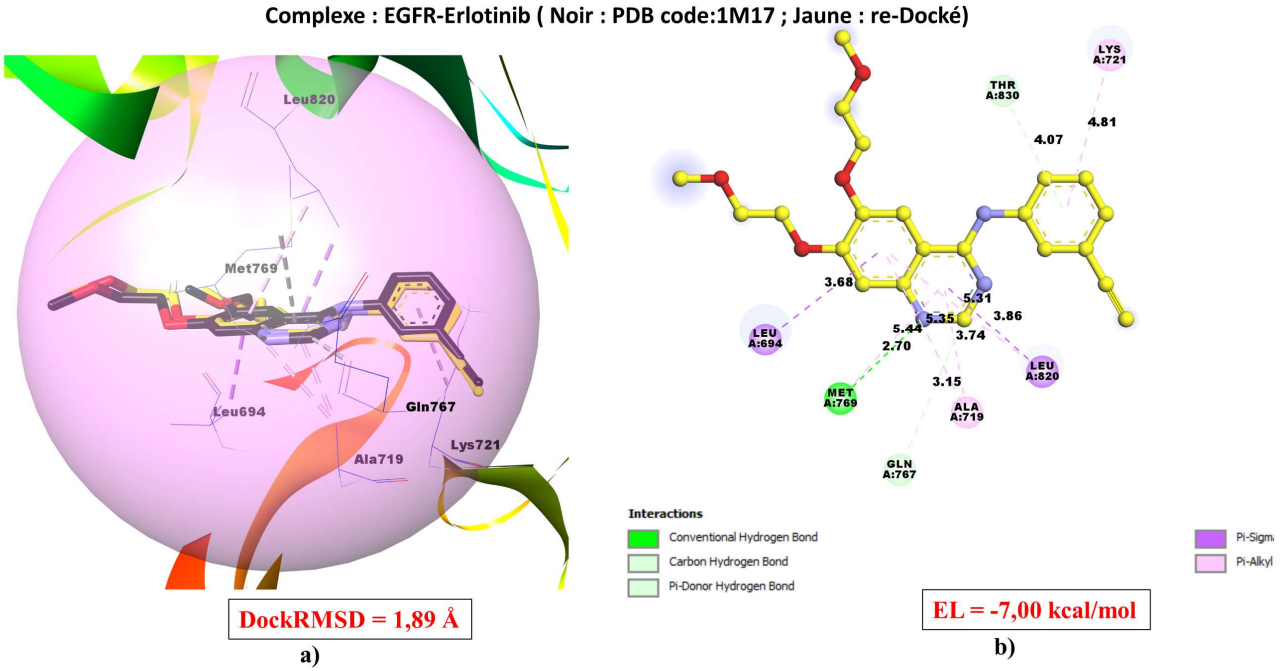


Figure IV. A5. Visualisation des modes de contacts EGFR-Erlotinib dans le complexe original 1M17 et le complexe remodelé par re-Docking moléculaire. (a) visualisation en 3D de la disposition de l'alignement du Foretinib dans le complexe original (couleur noire) et le complexe remodelé (couleur jaune) à l'intérieur de la poche active d'EGFR-TKD. (b) visualisation en 2D des interactions entre l'Erlotinib redocké dans la cavité active de l'EGFR-TKD.

Sans intervention des faibles interactions de van der Waals, les coordonnées de la poche active de l'EGFR-TKD dans l'espace 3D (dimension de la grille 60*60*60 $Å^3$, distance entre les points de la grille de 0.375 Å, coordonnées de la grille x = 21.014 Å, y = 0.253 Å, z = 0.253 Å) sont fixées. Les résidus d'acide aminé **Leucine** (Leu820 et Leu694), **Méthionine** (Met769), **Alanine** (Ala719), **Glutamine** (Gln767) et **Lysine** (Lys721) représentent les résidus d'acide aminé les plus importants de l'EGFR-TKD (**Figure IV. A5a**). La **Figure IV. A5b** montre que le ligand Erlotinib re-Docké dans le site de liaison d'EGFR-TKD que nous avons identifié dans cette étude présente les mêmes interactions que celles observées dans le complexe original, ce qui montre que le modèle de Docking utilisé dans cette étude est très performant. Cela peut s'expliquer davantage par la valeur de DockRMSD=1.89 Å inférieure à 2 Å qui a été calculée sur la base de critères de distance entres les atomes du ligand original (couleur noire) et du ligand re-Docké (couleur jaune). Après avoir confirmé la validité du modèle de Docking moléculaire implémenté dans cette étude, nous avons docké le dérivé de Magnolol de référence #30 dans la poche active d'EGFR-TKD (**Figure IV. A6**).

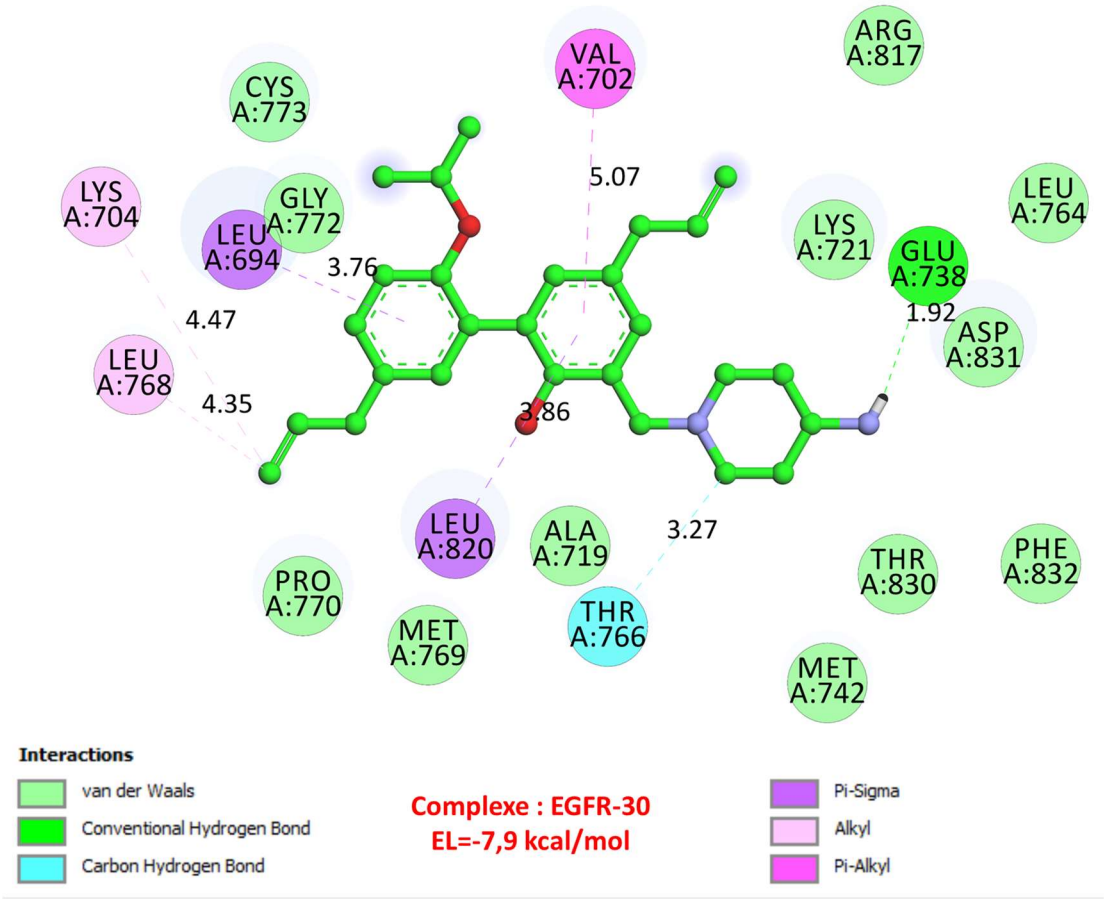


Figure IV. A6. Visualisation 2D des interactions entre le dérivé de Magnolol de référence #30 et la poche active d'EGFR-TKD (énergie de liaison estimée à -7.9 kcal/mol).

De la **Figure IV. A5**, il est clair que le composé Ligand #30 interagit avec les mêmes six résidus d'acides aminés que l'inhibiteur Erlotinib dans le complexe EGFR-Erlotinib (1M17). Une interaction hydrophobe de type p-sigma est observée entre Ligand #30 et les résidus actifs Leu694 (3.76 Å) et Leu820 (3.86 Å). D'autre part, une interaction van der Waals est observée entre Ligand #30 et Met769, Ala719, Lys721 et Thr830. En outre, une liaison hydrogène de type classique et de carbone est présente entre le ligand et les résidus Glu738 (1.92 Å) et Thr766 (3.27 Å). Cet effet est amplifié par la présence d'un groupe 3-(4-aminopiperidin-1-yl)-méthyle dans le ligand. Les parties aromatiques (**R1**, **R2**, et **R3**) du ligand interagissent avec Leu768, Lys704, Leu694, Leu820 et Val702 dans le site actif. Comparativement, l'inhibiteur Erlotinib présente des interactions avec Met (2.70 Å), Gln767 (3.15 Å) et Thr 830 (4.81 Å) respectivement avec des liaisons hydrogènes classiques, du carbone et du pi-Donneur. Notamment, toutes les distances entre les résidus d'acides aminés et Ligand #30 ont été observées comme étant plus courtes que celles d'Erlotinib. Par conséquent, on peut supposer que le Ligand #30 est plus stable dans le complexe EGFR-30 que Erlotinib dans le complexe EGFR-Erlotinib. Par conséquent, cela montre dans l'énergie d'affinité considérablement plus élevée du ligand #30 (-7.9 kcal / mol) que Erlotinib (-7.0 kcal / mol). Ainsi, avec l'énergie d'affinité élevée du Ligand #30 vers la poche active EGFR, il est susceptible de fournir une plus grande probabilité d'inhiber la croissance cellulaire CPNPC. De plus, Ligand #30 possède une qualité de stimulateur d'autophagie, qui peut encore contribuer à l'élimination du CPNPC. Ensemble, les propriétés structurelles du Ligand #30 évoquent la possibilité de l'utiliser comme modèle de référence ou comme clé biomoléculaire de

référence dans la conception de médicaments et le ciblage du traitement du CPNPC, afin de fournir un traitement plus efficace pour cette forme de cancer. Dans ce contexte, les résidus d'acides aminés avec lesquels le ligand #30 interagit dans le site de liaison actif de l'EGFR-TKD peuvent être considérés comme la clé de l'inhibition de l'activité enzymatique par des dérivés du Magnolol bioactifs. Voici les résidus principaux détectés dans le cadre de cette étude **Lysine** (Lys704 et Lys721), **Leucine** (Leu768, Leu694, Leu820 et Leu764), **Cystéine** (Cys773), **Proline** (Pro770), **Méthionine** (Met769, Met742 et Met742), **Valine** (Val702), **Alanine** (Ala719), **Thréonine** (Thr766, et Thr830), **Arginine** (Arg817), **Glycine** (Glu738), **Acide Aspartique** (Asp831), et **Phénylalanine** (Phe832).

3.1.3 Analyse biomoléculaire des dérivés du Magnolol via MolCAD

Pour approfondir la caractérisation structurale des profils d'affinité/compatibilité des dérivés du Magnolol en tant qu'inhibiteurs prometteurs de l'EGFR-TKD basés sur le ligand #30, nous avons soumis le complexe EGFR-30 à des analyses plus approfondies basées sur la technologie MolCAD (MOLecular Computer Aided Design). La conformation du complexe EGFR-30 ayant une énergie de liaison négative élevée est utilisée comme point d'entrée pour reproduire les cavités liées à des interactions protéine-ligand en utilisant la technique Surflex-Dock afin de les visualiser en 3D [477], [478]. Grâce au modèle MolCAD, la profondeur des cavités (CD), la lipophilie (LP), le potentiel électrostatique (EP) et la liaison hydrogène (HB) traduisent le niveau d'affinité entre les conformations 3D du ligand bioactif et la structure du récepteur ciblé. La **Figure IV. A7** présente des visualisations en 3D des contacts qui se produisent au sein du complexe EGFR-30, tirées du modèle MolCAD appliqué dans le cadre de cette étude. L'analyse de la **Figure IV. A7(a-d)** révèle que les sites **R2**, **R3** et 3-(4-aminopiperidin-1-yl)méthyle dans la structure de la molécule 30 ont des orientations différentes dans la poche réceptrice EGFR, indiquant des préférences distinctes pour interagir avec les sites actifs sur le récepteur. Plus précisément, **R2** et **R3** tendent à s'aligner sur la région bleue, indiquant une préférence pour les groupes plus grands et l'hydrophobicité pour interagir avec ces sites (**Figure IV. A7a**). Pendant ce temps, **R1** est proche de la région jaune, suggérant une préférence pour les groupes plus petits. En outre, le groupe 3-(4-aminopiperidin-1-yl)-méthyle est proche de la région rouge, indiquant une affinité pour former des liaisons hydrogène, et également proche de la région bleue, indiquant une préférence pour interagir avec des charges positives. Ces données suggèrent que la structure de la molécule 30 est optimisée pour engager plusieurs sites actifs sur le récepteur EGFR-TKD.

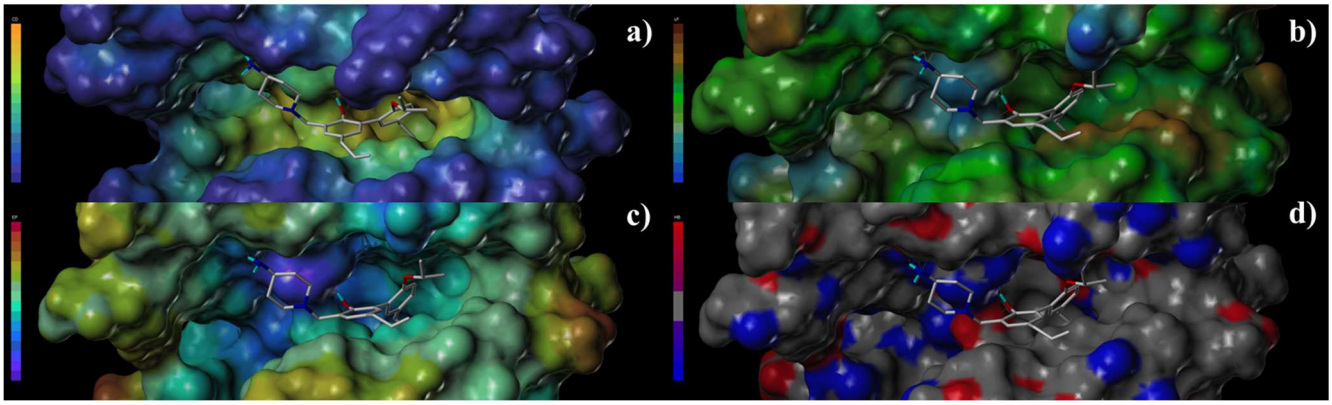


Figure IV. A7. Visualisation MolCAD de la surface moléculaire du ligand bioactif #30 à l'intérieur d'une poche active de l'EGFR-TKD. (a) (CD) surface de la profondeur de la cavité potentielle dans le complexe EGFR-30, la rampe de couleur de la profondeur de la cavité varie du bleu (valeurs de faible profondeur = à l'extérieur de la poche) à l'orange (valeurs de forte profondeur = cavités profondes à l'intérieur de la poche). (b) (LP) surface du lipophile potentiel dans le complexe EGFR-30, la rampe de couleur pour LP varie du brun (la région hydrophobe la plus élevée de la surface) au bleu (la région hydrophile la plus élevée). (c) (HB) liaison hydrogène au niveau du complexe EGFR-30, la rampe de couleur pour HB varie du rouge (donneurs d'hydrogène) au bleu (accepteurs d'hydrogène). (d) (EP) potentiel électrostatique de surface au niveau du complexe EGFR-30, la rampe de couleur pour EP varie du rouge (région la plus positive) au bleu (région la plus négative).

La **Figure IV. A7b** indique que les positions **R2** et **R3** sont alignées vers le côté bleu, ce qui leur permet de sortir de la cavité du récepteur. Cela signifie que les sites **R2** et **R3** dans la structure de la molécule 30 préfèrent les groupes les plus volumineux pour s'approcher de la poche active de l'EGFR. Alors que **R1** est proche de la région jaune, cela signifie que le site **R1** préfère les petits groupements pour atteindre la poche du récepteur de l'EGFR. La **Figure IV. A7b** indique que les sites **R2** et prop-1-ène sont proches de la cavité brune, ce qui signifie que ces deux sites conviennent aux groupements de type hydrophobe pour interagir avec les résidus actifs dans la poche de l'EGFR. La **Figure IV. A7c** montre que les parties **R2**, **R3** et 3-(4-aminopipéridine-1-yl) méthyle s'alignent sur la cavité rouge de la poche du récepteur de l'EGFR, ce qui signifie que ces derniers renforcent les propriétés qui conduisent à la formation de liaisons hydrogène avec les résidus actifs situés dans la poche du récepteur de l'EGFR-TKD. La **Figure IV. A7d** montre que le groupement 3-(4-aminopipéridine-1-yl)-méthyle est aligné sur la cavité bleue dans la poche du complexe (EGFR-30), ce qui signifie que ce groupement préfère les charges positives pour atteindre une haute affinité avec les sites de résidus actifs dans la poche du récepteur de l'EGFR-TKD.

3.1.4 Design in silico de nouveaux dérivés du Magnolol à activité anticancéreuse

Nous présentons dans la **Figure IV. A8** l'ensemble des aspects de la pharmacopée générés à partir des modèles 3D-QSAR, du Docking moléculaire et de MolCAD favorables à la conception et à l'optimisation de nouveaux dérivés du Magnolol qui présentent une affinité plus favorable vis-à-vis de l'EGFR-TKD et qui peuvent être considérés comme des candidats de choix pour concevoir des nouveaux médicaments anticancéreux.

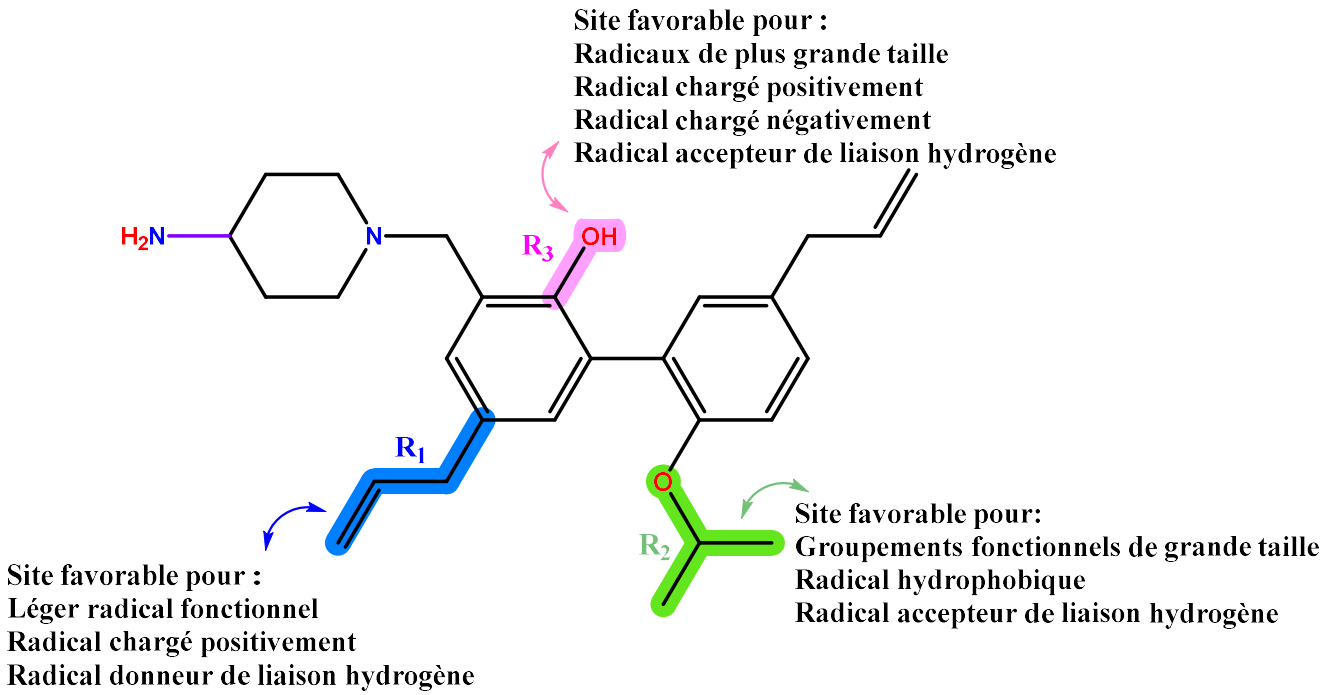
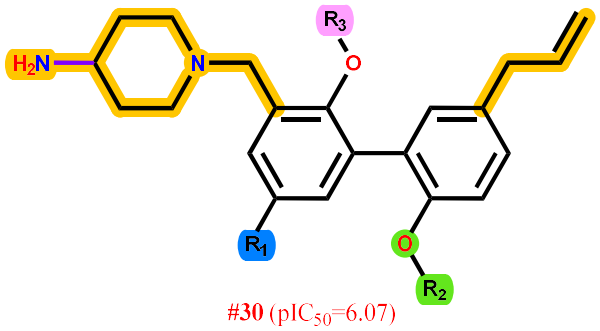


Figure IV. A8. Propriétés bio-structurales optimales favorables à l'amélioration de l'activité biologique inhibitrice et affinité de liaison au poche active de l'EGFT-TKD des dérivés du Magnolol pour lutter contre le CPNPC.

Sur la base des recommandations formulées lors de la caractérisation des propriétés structurelles favorables à l'amélioration des dérivés pIC_{50} du Magnolol présentées dans la **Figure IV. A8**, nous introduisons plusieurs adaptations structurelles au niveau de la molécule prototype #30 aux positions **R1**, **R2** et **R3** afin de concevoir une nouvelle classe de molécules candidates pour optimiser une nouvelle génération d'inhibiteurs spécifiques de l'EGFR-TKD. Ceci afin de se rapprocher d'une nouvelle stratégie thérapeutique contre les cancers liés notamment à l'activité enzymatique excessive de l'EGFR-TKD. Le **Tableau IV. A4** présente un ensemble de 11 dérivés de Magnolol développés dans le cadre de cette étude, avec les modifications structurelles apportées à la structure de référence #30. Les structures des autres molécules conçues qui n'ont pas montré sur le plan *in silico* d'amélioration significative de leur pIC_{50} prédit ou au moins proche du pIC_{50} de la référence #30, nous avons rejeté leur candidature en tant qu'EGFR-TKIs potentiels. Le **Tableau IV. A4** montre que les valeurs pIC_{50} des 11 nouveaux dérivés de Magnolol prédites *in silico* par le modèle 3D-QSAR sont optimisées. Concernant les dérivés X5, X7, X8, X9 et X10, ils ont montré des valeurs pIC_{50} *in silico* plus élevées que la valeur pIC_{50} observée *in vitro* pour le composé de base #30 (pIC_{50}=6.07). Par conséquent, les nouveaux dérivés de Magnolol conçus peuvent agir comme de nouveaux inhibiteurs puissants du développement du CPNPC. Par conséquent, ces données confirment la pertinence des nouveaux composés proposés X5, X7, X8, X9 et X10 pour la conception et la découverte rationnelles de nouveaux types de substances thérapeutiques. Nous pouvons également conclure que les connaissances cumulées engendrées à partir de l'intégration des modélisations 3D-QSAR, Docking, moléculaire et MolCAD sont suffisantes pour optimiser une nouvelle classe de dérivés de Magnolol avec une activité biologique et des propriétés de pharmacopée favorables à la lutte contre la croissance du CPNPC.

Tableau IV. A4. Activité biologique inhibitrice pIC_{50} prédite *in silico* contre le CPNPC de la série de dérivés du Magnolol optimisés dans cette étude.



#30 (pIC_{50}=6.07)

Composé	R_1	R_2	R_3	pIC_{50}	
				CoMFA	CoMSIA
X01	Méthane	Isobutane	Isobutane	5.950	5.797
X02	Méthane	Méthane	Méthane	5.974	5.854
X03	Hydrogène	Méthane	Méthane	5.890	5.867
X04	Hydrogène	Éthane	Propan-2-one	5.876	5.762
X05	**Hydrogène**	**Propane**	**Propan-2-one**	**6.162**	**5.948**
X06	Méthane	Propane	Propan-2-one	5.956	5.838
X07	**Méthane**	**Isobutane**	**Propane**	**6.392**	**6.112**
X08	**Hydrogène**	**Isobutane**	**Propane**	**6.145**	**5.974**
X09	**Méthane**	**Propane**	**Hydrogène**	**6.293**	**6.189**
X10	**Hydrogène**	**Propane**	**Hydrogène**	**6.170**	**6.261**
X11	Hydrogène	Cumène	Hydrogène	5.847	5.852

3.2 Explorer le potentiel thérapeutique des dérivés du Magnolol : Analyse *in silico* des propriétés Drug-like et ADME-Tox

Nous menons ces analyses pour identifier les structures moléculaires les plus pertinentes, sûres et appropriées en ce qui concerne les propriétés requises pour les substances médicamenteuses. À cet égard, nous avons évalué les cinq nouvelles molécules X5, X7, X8, X9 et X10 conçues aux côtés de la molécule de référence #30 et du médicament Foretinib. Cette évaluation comprenait notamment une analyse des critères de biodisponibilité orale (**Tableau IV. A5**) et des critères ADME-Tox (**Tableau IV. A6**).

Tableau IV. A5. Profil des propriétés de type drug-like des substances candidates.

Entrée	ABS	TPSA(Å²)	n-RB	MW (Da)	LogP	n-HA	n-HD	Violations			SA
								Lipinski	Veber	Egan	
Règle	**-**	**<140**	**<10**	**<500**	**<=5**	**<10**	**<5**	**<2**	**<2**	**<2**	**0<SA<10**
X05	Forte	64.79	10	436.59	4.27	5	1	0	0	0	4.13
X07	Forte	47.72	9	450.66	4.86	4	1	0	0	0	4.32
X08	Forte	47.72	9	436.63	4.50	4	1	0	0	0	4.20
X09	Forte	58.72	7	394.55	4.27	4	2	0	0	0	3.78
X10	Forte	58.72	7	380.52	3.96	4	2	0	0	0	3.67

ABS : Absorption, **TPSA** : Surface Polaire Topologique, **n-RB** : Nombre de liaisons rotatives, **MW** : Poids moléculaire, **Log P** : logarithme du coefficient de partage du composé entre n-octanol et l'eau, **n-HA** : Nombre d'accepteurs de liaisons hydrogène, **n-HD** : Nombre de donneurs de liaisons hydrogène, **SA** : Accessibilité synthétique.

Les résultats obtenus par les prédictions de type Drug-like montrent que les cinq composés proposés répondent aux règles de Lipinski, Veber et Egan pour la biodisponibilité et ont un indice d'accessibilité synthétique de 3-4, ce qui indique qu'ils peuvent être facilement synthétisés avec une faible complexité. Parmi les cinq, X5 avait un nombre relativement élevé de liaisons rotatives, ce qui peut réduire sa flexibilité et est donc exclu de toute analyse ultérieure. Les nouvelles molécules X7, X8, X9 et X10 sont potentiellement des agents viables pour l'optimisation et la conception de médicaments

contre le CPNPC, par conséquent, leurs paramètres pharmacocinétiques/pharmacodynamiques ADME-Tox ont été évalués pour sélectionner les candidats les plus appropriés (H2L) pour une utilisation médicinale. Les prédictions ADME-Tox ont montré que les quatre dérivés de Magnolol optimisés (X07, X08, X09 et X10) ont une forte capacité d'absorption par l'intestin dans le corps humain et sont des substrats de la p-glycoprotéine, les rendant appropriés pour l'utilisation comme médicament. Erlotinib se démarque seul par son absence de fonctionnement comme substrat de la p-glycoprotéine, ce qui lui permet d'être métabolisé dans le corps humain, réduisant ainsi son efficacité et son activité.

Tableau IV. A6. Profil ADME-Tox des dérivés de Magnolol proposés comme H2L computationnel.

	Propriétés														
	Absorption		**D**istribution			**M**étabolisme							**E**xcrétion	Toxicité	
						CYP450									
						Substrat		Inhibiteur							
models	Absorption intestinale (humain)	P-Gp substrat	VDss (humain)	Perméabilité de la BHE (BBB)	Perméabilité du SNC (CNS)	2D6	3A4	1A2	2C19	2C9	2D6	3A4	Clairance totale	AMES	
Unit	%	Oui/Non	Log L kg^{-1}	Log BB	(Log PS)	Oui/Non							log mL min^{-1} kg^{-1}	Oui/Non	
	Predicted values														
Erlo	95.226	Non	-0.273	-1.019	-3.468	Non	Oui	Oui	Oui	Oui	Non	Oui	0.646	Non	
#30	94.708	Oui	1.191	-0.423	-2.102	Non	Oui	Non	Oui	Non	Oui	Non	1.093	Non	
X7	91.604	Oui	1.046	-0.355	-1.165	Non	Oui	Non	Oui	Non	Oui	Oui	0.633	Non	
X8	91.794	Oui	1.009	-0.392	-1.203	Non	Oui	Non	Non	Non	Oui	Oui	0.644	Non	
X9	94.603	Oui	1.302	-0.336	-2.008	Non	Oui	Non	Oui	Non	Oui	Non	0.923	Non	
X10	91.222	Oui	1.636	-0.580	-2.089	Non	Oui	Non	Non	Non	Oui	Non	0.858	Non	

Les résultats de la distribution indiquent que #30, X07, X08, X09 et X10 ont un meilleur potentiel pour se distribuer dans le corps que Erlotinib. L'évaluation de leurs interactions avec les cytochromes P450 a montré que X07, X08, X09 et X10 sont des substrats du cytochrome 3A4 et du 2C9 et n'inhibent aucune enzyme CYP, tandis qu'Erlotinib inhibe les quatre co-enzymes CYP450. En ce qui concerne l'excrétion, toutes les molécules (y compris Erlotinib) présentent des valeurs faibles de clairance totale, ce qui signifie qu'elles peuvent rester dans le corps sans être rapidement éliminées. Enfin, l'évaluation de la toxicité suggère que tous ces composés, y compris Erlotinib, sont non toxiques. Selon les résultats de l'écran ADME-Tox, nous avons pu conclure que les dérivés de magnolol optimisés X07, X08, X09 et X10 sont prometteurs en tant qu'H2L pour la conception de médicaments anticancéreux. Nous avons également constaté que ces molécules étaient capables de discriminer avec succès une gamme de propriétés de médicaments lorsqu'elles sont comparées au médicament Erlotinib, utilisé dans le traitement du CPNPC. De plus, notre composé de référence #30, portant une 3-(4-aminopipéridine-1-yl)méthyle, a également montré un profil ADME-Tox acceptable. Par conséquent, la structure de cette molécule peut être considérée comme un idéal de référence pour la conception et l'optimisation de composés plus bioactifs contre le cancer et d'autres types de maladie.

3.3 Recherche d'inhibiteurs de l'EGFR-TKD par simulation de Docking Moléculaire

Après avoir complété les procédures de vérification de l'adaptation des conformations bioactives des ligands X07, X08, X09 et X10 à la conception médicamenteuse, nous passons à l'étape de la validation de ces derniers en tant que nouveaux inhibiteurs spécifiques de l'EGFR-TKD. À cette fin, nous effectuons un criblage virtuel en amarrant ces molécules au site de liaison actif de l'EGFR-TKD, puis nous dégageons des hypothèses et des estimations sur la compatibilité et l'affinité possibles des

ligands vis-à-vis de la protéine ciblée. Le **Tableau IV. A7** présente un résumé détaillé des résultats obtenus lors du criblage par Docking moléculaire en termes de scores d'affinité (énergie de liaison) et de profil d'interactions protéine-ligand.

Tableau IV. A7. Résumé des résultats d'un criblage virtuel basé sur le Docking moléculaire des inhibiteurs spécifiques de l'EGFR.

Récepteur ciblé	Affinité protéine-ligand			Interactions non covalentes spécifiques à l'interface protéine-ligand			
EGFR-TKD	Ligand	Énergie de liaison (kcal/mol)	Constante d'inhibition (Ki , µM)	Liaison hydrogène	Électrostatique	van der Waals	Hydrophobique
PDB: 1M17	**Erlotinib**	-7.0	7.29	Met769, Gln767, Thr830	-	Glu738, Leu764, Thr830, Thr766, Asp831, Gly695, Met769, Leu768, Phe771, Gly772, Cys773, et Gln768	Ala719, Lys721, Leu820 et Leu694
	#30	-7.9	1.59	Thr766, Glu738	-	Gly777, Cys773, Pro770, **Met769**, **Ala719**, Arg817, **Lys721**, Leu764, Asp831, Phe832, Asp831 et Leu764	Lys704, Leu768, **Leu694**, Val702, **Leu820**
	X07	-8.5	0.58	-	-	Gly695, Asp831, Thr830, Gly772, Pro770, Met742, **Lys721**, Leu753, Leu764, Thr766, Glu738	Val702, Phe699, Leu694, **Ala719**, Leu768, Leu820, et **Met769**
	X08	-8.3	0.81	Met742	-	Gly695, Pro770, Gly772, Asp831, Lys721, Thr830, Thr766, Leu753, GLu738, et Leu764	Phe699, Leu694, Val702, Ala719, Leu820, Met769, et Leu768
	X09	-8.6	0.49	Glu738, Met742, et Thr766	Met742	Pro770, Gly695, **Met769**, Gly772, **Ala719**, Thr830? Asp831, **Lys721**, et Leu764	Leu768, Lys704, **Leu694**, **Leu820**, et Val702
	X10	-9.2	0.18	Thr830, et Gly738		Pro770, Cys773, **Met769**, Gly772, **Ala719**, **Leu820**, Gly695, Thr766, **Lys721**, Met742, Leu764, Phe832, et Asp831	Leu768, **Leu694**, Lys704, et Val702
Site actif de référence	**Leucine** (Leu820 et Leu694), **Méthionine** (Met769), **Alanine** (Ala719), **Glutamine** (Gln767) et **Lysine** (Lys721)						

Grâce aux données présentées dans le **Tableau IV. A7** relatives au degré d'affinité protéine-ligand, nous pouvons noter que les quatre nouveaux ligands X07 (-8.5kcal/mol), X08 (-8.3 kcal/mol), X09 (-8.6 kcal/mol) et X10 (-9.2 kcal/mol) présentent une énergie de liaison plus stable et plus forte par rapport aux ligands #30 et au médicament de référence, l'Erlotinib. Cela indique la grande concordance entre les conformations actives des ligands inhibiteurs et la structure de la protéine EGFT-TKD. De même, les valeurs très faibles des constantes d'inhibition estimées (Ki<100 µM) présentées par tous les ligands analysés indiquent la grande aptitude de ces nouveaux agents à inhiber l'activité enzymatique de la protéine EGFR-TKD. Toutes ces indications confirment l'hypothèse que les quatre dérivés du Magnolol développés dans cette étude sont appropriés en tant qu'inhibiteurs spécifiques prometteurs de l'EGFR.

Le profil d'interactions non covalentes protéine-ligand observé montre également que de multiples interactions de type Van der Waals, hydrophobes et électrostatiques ainsi que des liaisons hydrogène sont responsables en partie de la stabilité des ligands #30, X07, X08, X09 et X10 à l'intérieur de la poche active de l'EGFR-TKD. Ceci peut être confirmé par les nombreuses interactions que l'inhibiteur de référence Erlotinib et les ligands examinés dans cette étude partagent avec les résidus d'acides aminés de référence dans l'inhibition de l'EGFR-TKD (ligne en **gras** au **Tableau IV. A7**). Sur la base de ces informations, on peut conclure que les conformations des quatre ligands X07, X08, X09 et X10 sont plus compatibles et plus en accord avec la structure de l'EGFR-TKD. Les structures de ces quatre ligands peuvent donc être considérées comme une extension idéale des candidats dérivés du Magnolol pour la conception de médicaments ayant une action sur l'EGFR-TKD. Parmi les nouvelles molécules proposées comme inhibiteurs spécifiques de l'EGFR-TKD, nous choisissons le ligand X10 comme inhibiteur computationnel idéal de cette classe conçue en raison de sa forte capacité de liaison avec l'EGFR et de sa concentration inhibitrice très favorable pour atteindre l'objectif thérapeutique avec une concentration minimale estimée à 0.18 µM. La **Figure IV. A9** présente une visualisation 2D des interactions non covalentes identifiées dans le complexe EGFR-X10 à l'aide des simulations de Docking moléculaire réalisées dans le cadre de cette étude.

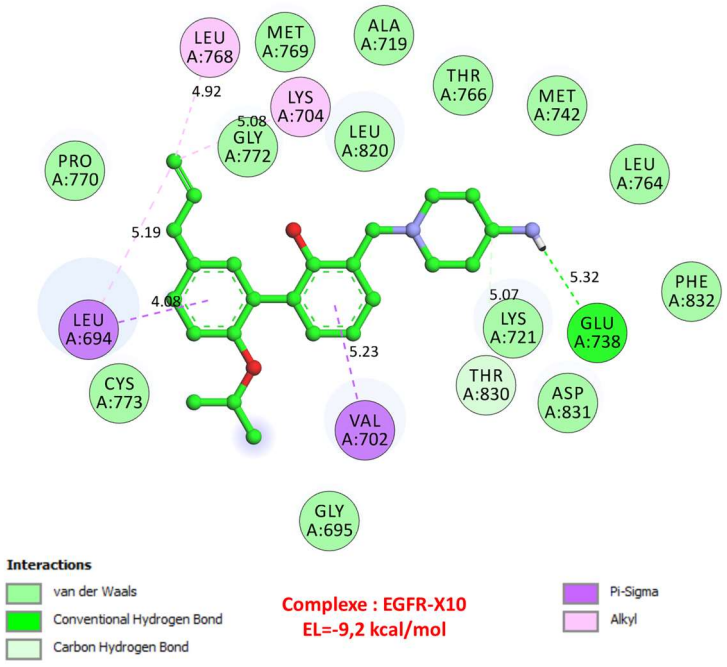


Figure IV. A9. Représentation en 2D des interactions non covalentes observées dans le complexe EGFR-X10.

La **Figure IV. A9** montre que le groupe fonctionnel 3-(4-aminopipéridine-1-yl)-méthyle interagit avec quatre sites de résidus dans la poche active de l'EGFR, par le biais d'interactions de van der Waals avec les acides aminés Leu764, Asp831, Phe832 et Lys721, ainsi que certaines interactions hydrogène conventionnelles et carbone avec Glu738 et Thr830, respectivement. Cela confirme que le groupe 3-(4-aminobipyridin-1-yl)-méthyle est un groupe d'échafaudage principal impliqué dans l'inhibition de l'activité enzymatique de l'EGFR-TKD et par conséquent dans l'inhibition de la croissance du cancer du poumon non à petites cellules (CPNPC). Ce pouvoir inhibiteur potentiel pourrait également être lié à la capacité des structures des dérivés du Magnolol portant un groupe 3-(4-aminobipyridin-1-yl)-méthyle à renforcer l'autophagie pour la suppression des cellules cancéreuses

programmées. Nous pouvons également constater que le ligand X10 interagit aux résidus d'acides aminés Leu768, Lys704 et Leu694 par des interactions alkyles, et aux résidus Leu694 et Val702 par des interactions pi-sigma. Ces diverses interactions entre le ligand X10 et la protéine cible EGFR-TKD justifient la grande compatibilité et la stabilité de la conformation biactive de ce ligand en complexe avec l'EGFR-TKD.

3.4 Affiner la sélection des inhibiteurs de l'EGFR-TKD par Prime/MM-GBSA

Dans cette étape de validation des résultats obtenus par le criblage due Docking moléculaire, nous utilisons la méthode Prime Molecular mechanics-generalized Born surface area (MM-GBSA) pour recorriger et recalculer l'énergie libre de liaison (ΔG_{bind}) des ligands examinés dans la poche active de l'EGFR-TKD. Ceci sert comme critère de reconnaissance du niveau de stabilité structurelle dans les complexes EGFR-TKIs modélisés dans cette étude. Dans ce contexte, nous avons appliqué les calculs MM-GBSA au complexe EGFR-X10 de la classe des dérivés de Magnolol conçus, au complexe EGFR-30 de la classe des dérivés de Magnolol synthétisés et au complexe EGFR-erlotinib de la classe des inhibiteurs de tyrosine kinase. Cette procédure a été réalisée après avoir minimisé les énergies des complexes par le champ de force OPLS3e à pH 7 ± 2 en utilisant le package Prime MM-GBSA disponible dans Maestro Schrödinger [479]. Au cours de cette procédure, l'énergie des complexes protéine-ligand ($E_{complex}$), ainsi que les ligands (E_{ligand}) sont minimisés, résultant ainsi dans l'énergie de liaison libre (ΔG_{bind}) pour chaque complexe protéine-ligand. Le **Tableau IV. A8** résume avec précision les paramètres des énergies libres de liaison que nous avons dégagées dans cette étude à l'aide de MM-GBSA concernant les complexes EGFR-X10, EGFR-30 et EGFR-Erlotinib.

Tableau IV. A8. Principaux résultats des calculs Prime/MM-GBSA.

Critères de mesure MM-GBSA	Complexes d'EGFR-TKIs		
	EGFR-X10	EGFR-30	EGFR-Erlotinib
ΔG_{bind} (MM-GBSA) (kcal/mol)	-65.59	-61.07	-42.56
ΔG_{bind} Vdw (MM-GBSA) (kcal/mol)	-67.98	-62.87	-43.87
ΔG_{bind} Hbond (MM-GBSA) (kcal/mol)	-1.11	-0.69	-0.32
Énergie du complexe (kcal/mol)	-12448.782	-12441.305	-10389.215
Énergie potentielle (Kcal/mol)	-1024.858	-1023.908	-873.247

Les résultats des computations effectuées par MM-GBSA présentés dans le **Tableau IV. A8** indiquent que l'énergie de liaison libre estimée pour les inhibiteurs X10 (ΔG_{bind}=-65.59 kcal/mol) et #30 (ΔG_{bind}=-61.07 kcal/mol) vis-à-vis de l'EGFR est une valeur plus élevée que celle de l'inhibiteur Erlotinib (ΔG_{bind}=-42.56 kcal/mol). Ce résultat peut également être confirmé par les scores énergétiques élevés de ΔG_{bind} Vdw, ΔG_{bind} Hbond, l'énergie complexe, l'énergie potentielle. De même, les visualisations 2D des interactions des complexes EGFR-inhibiteurs (EGFR-TKIs) obtenues à partir de la simulation MM-GBSA présentées dans la **Figure IV. A10** montrent que les inhibiteurs X10, #30 et Erlotinib ont conservé la plupart de leurs interactions avec EGFR-TKD prédites par le modèle de Docking moléculaire.

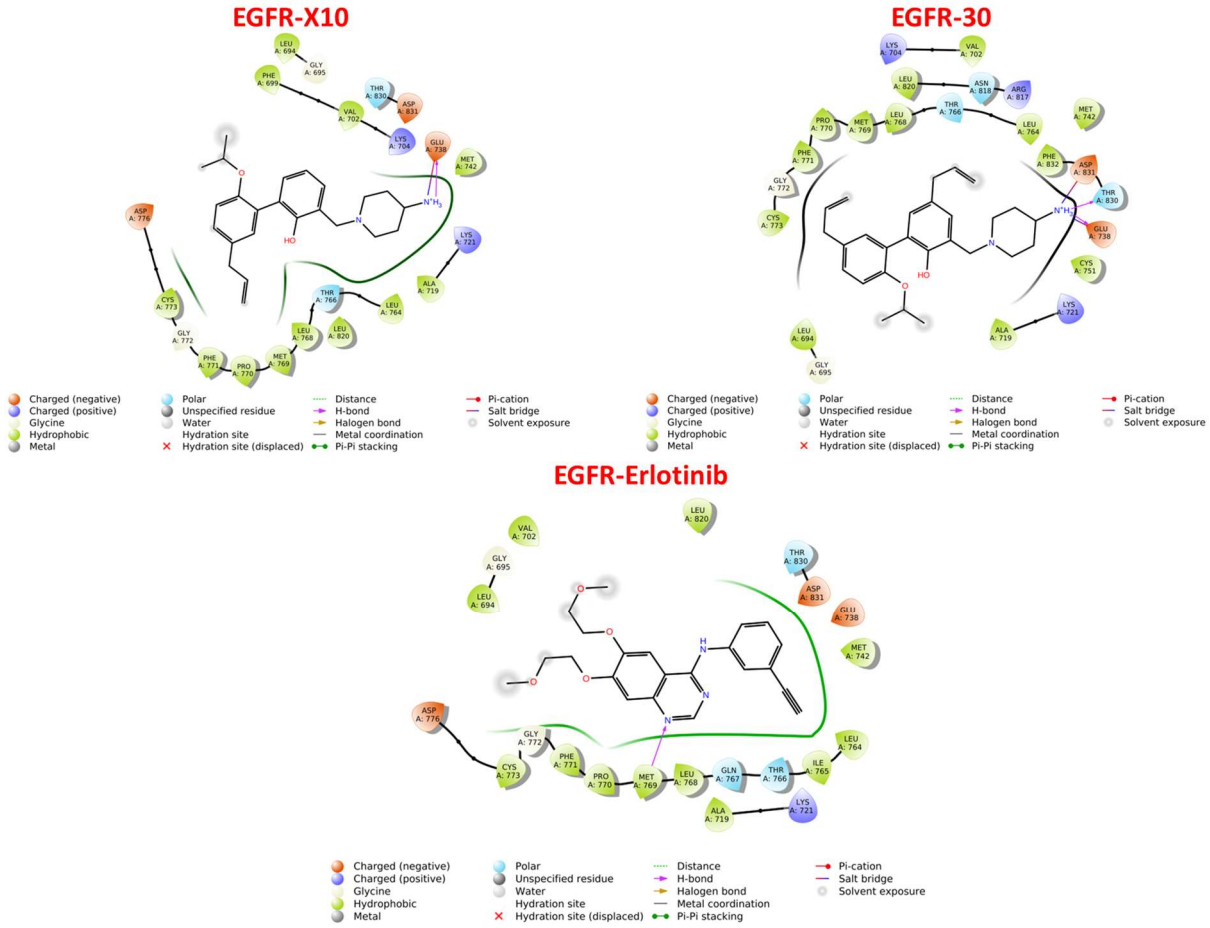


Figure IV. A10. Interactions non covalentes protéine-ligand régénérées après avoir soumis les complexes EGFR-X10, EGFR-30 et EGFR-Erlotinib aux simulations de MM-GBSA.

D'après les résultats globaux des calculs Prime/MM-GBSA, nous pouvons affirmer que les structures biomoléculaires X10 et #30 basées sur le Magnolol sont plus stables que celles de l'Erlotinib dans la poche active de l'EGFR-TKD (1M17). Cela confirme les prévisions obtenues à partir des modélisations 3D-QSAR, Docking Moléculaire, MolCAD et ADMET réalisées dans le cadre de cette étude. Par conséquent, l'incorporation de toutes les techniques de modélisation moléculaire basées sur le ligand et la structure utilisées dans ce travail en tant que stratégie CADD hybride est favorable à la conception et à l'optimisation de nouvelles molécules basées sur la structure 3-(4-aminobipyridin-1-yl)-méthyl du Magnolol, candidates EGFR-TKIs spécifiques pour la conception de médicaments contre le CPNPC. En général, dans cette étude, nous avons réalisé une évaluation *in silico* de 51 substances phytochimiques issues de *Magnolia officinalis*, considérées comme des candidats EGFR-TKI pour la conception de nouveaux médicaments destinés à lutter contre le cancer du poumon non à petites cellules (CPNPC). À cette fin, diverses méthodes de modélisation moléculaire ont été utilisées, telles que 3D-QSAR, Molecular Docking, MolCAD, ADME-Tox et la validation par énergie libre de liaison MM-GBSA. La modélisation 3D-QSAR est développée par le biais des descripteurs de champs moléculaires CoMFA et CoMSIA. Les modèles CoMFA/SE et CoMSIA/SEHDA élaborés ont démontré leur excellente prédictivité en termes de relations quantitatives entre la structure et l'activité des dérivés du Magnolol. Les cartes de contour générées par le modèle 3D-QSAR nous ont permis d'identifier les propriétés structurelles les plus importantes des dérivés du Magnolol, favorables pour améliorer leurs activités biologiques inhibitrices contre la croissance du CPNPC. Les hypothèses obtenues à partir de l'analyse

des cartes de contours sont confirmées par des simulations de Docking moléculaire et des visualisations de modèles MolCAD. Les résultats de ces procédures ont indiqué que les sites (**R1**, **R2**, **R3**, 3-(4-aminopipéridin-1-yl)) situés sur la structure moléculaire de référence #30 (activité biologique observée la plus élevée) représentent les bonnes positions pour modéliser l'activité biologique inhibitrice des dérivés du Magnolol contre le CPNPC. Sur la base des prédictions des modèles 3D-QSAR, Docking moléculaire, MolCAD, nous avons développé un ensemble de 11 (X01-X11) nouveaux dérivés de Magnolol basés sur le composé de référence #30. Le criblage *in silico* des nouveaux composés conçus par modélisation 3D-QSAR a indiqué que les pIC_{50} prédits pour les structures proposées étaient plus élevés que ceux synthétisés et évalués *in vitro*. Par conséquent, la possibilité d'inhiber la croissance des cellules du CPNPC peut être renforcée par les nouvelles structures de Magnolol développées. Les analyses *in silico* de type Drug-like et ADME-Tox nous ont permis de sélectionner les molécules (X07, X08, X09 et X10) susceptibles d'être utilisées comme médicaments, en raison de leurs propriétés structurelles adaptées à ce type du médicament. Pour confirmer davantage que ces molécules sont des H2L pour la thérapie ciblée contre le cancer, nous avons ciblé la structure de la protéine EGFR-TKD à l'aide par ces dernières. Cette étape a été réalisée en adoptant le modèle de Docking moléculaire que nous avons validé dans cette étude. Au cours de toutes les étapes du criblage et de la sélection, nous avons fixé la molécule de référence #30 et l'Erolotinib (inhibiteur spécifique de la tyrosine kinase) comme facteurs de comparaison lors de l'évaluation de nos résultats. Les résidus d'acides aminés suivants identifiés lors de la phase de Docking moléculaire ont été identifiés : Lysine (Lys704 et Lys721), Leucine (Leu768, Leu694, Leu820 et Leu764), Cystéine (Cys773), Proline (Pro770), Méthionine (Met769, Met742 et Met742), et Valine (Val702). Les sites actifs de référence, Alanine (Ala719), Thréonine (Thr766, et Thr830), Arginine (Arg817), Glycine (Glu738), Acide Aspartique (Asp831), et Phénylalanine (Phe832) servent de site actif de référence pour les interactions des dérivés du Magnolol dans la poche active de l'EGFR-TKD. Les énergies d'affinité et les constantes d'inhibition des complexes EGFR-TKI sont estimées par Docking moléculaire à à -7.0 kcal/mol (7.29 µM) pour le complexe EGFR-Erlotinib, -7.9 kcal/mol (1.59 µM) pour le complexe EGFR-30, et -8.5 kcal/mol (0.58 µM) pour le complexe EGFR-X07, -8.3 kcal/mol (0.81 µM) pour le complexe EGFR-X08, -8.6 kcal/mol (0.49 µM) pour le complexe EGFR-X09 et -9.2 kcal/mol (0.18 µM) pour le complexe EGFR-X10. Ces indications montrent que les ligands X07, X08, X09 et X10 optimisés dans cette étude peuvent générer un environnement inhibiteur très avantageux pour l'EGFR-TKD par rapport à la molécule de référence #30 et à l'inhibiteur spécifique Erlotinib. La procédure d'investigation post-Docking moléculaire via les énergies libres de liaison des MM-GBSA a confirmé la stabilité des complexes EGFR-TKIs examinés. Sur la base des résultats de cette étude computationnelle, il est évident que les structures des dérivés de Magnolol qui ont été optimisées dans cette étude ont des perspectives prometteuses dans la conception de médicaments contre le cancer par l'inhibition de l'activité enzymatique de l'EGFR-TKD. Ces molécules peuvent donc être considérées comme le début de la découverte d'une nouvelle génération d'EGFR-TKIs à base de *Magnolia officinalis*.

4 Conclusion

Ce travail a présenté une analyse *in silico* des dérivés phytochimiques obtenus à partir de *Magnolia officinalis* qui sont des candidats potentiels pour le développement de médicaments contre le cancer du poumon non à petites cellules (CPNPC). En particulier, les modèles 3D-QSAR CoMFA/SE et CoMSIA/SEHDA élaborés ont montré une excellente prédictivité en ce qui concerne les relations quantitatives entre la structure et l'activité inhibiteur des dérivés du Magnolol. Ces résultats ont permis d'identifier les propriétés structurelles optimales qui sont favorables à l'amélioration de l'activité biologique des produits chimiques issus de *Magnolia officinalis* contre le CPNPC. Ensuite, des simulations de Docking moléculaire, MolCAD, et des analyses ADME-Tox ont été réalisées afin de développer 11 (X01-X11) nouveaux dérivés de Magnolol basés sur le composé de référence #30. Les prédictions des modèles 3D-QSAR et des simulations de Docking ont suggéré que ces composés présentent un potentiel important en tant que nouvelles voies thérapeutiques, en raison de leurs propriétés adaptées à celles des médicaments. Par conséquent, les structures optimisées à partir des dérivés de Magnolol sont susceptibles d'inhiber efficacement la croissance des cellules du CPNPC. Ils peuvent donc servir de début dans la conception d'une nouvelle génération d'EGFR-TKIs à base de *Magnolia officinalis*.

B. Explorer le Potentiel Anticancéreux des Phytochimiques du Cannabis : Une étude *in silico*

1 Introduction

Le *Cannabis sativa L.* est cultivé et utilisé depuis longtemps dans le monde entier pour ses fibres, son alimentation, sa fonction médicinale et ses applications à caractère récréatif [463]. Dans des pays comme le Maroc et l'Afghanistan, un produit à base de cannabis connu sous le nom de haschisch est largement distribué et consommé pour ses propriétés psychoactives [480], [481], [482]. Cette plante est de plus en plus étudiée pour son potentiel à traiter des problèmes pathologiques tels que l'inflammation, la douleur, la neurodégénérescence, l'épilepsie, le cancer et les troubles du sommeil [360]. Le cannabis contient divers composés actifs, notamment des cannabinoïdes comme le THC (Tétrahydrocannabinol), le CBD (Cannabidiol) et le CBN (Cannabinol), ainsi que des terpènes, des flavonoïdes et des composés phénoliques [483]. Tous ces composés sont censés avoir des propriétés médicinales uniques, ce qui fait du cannabis une plante médicinale potentiellement puissante. Il existe différentes catégories de cannabis, basées sur les concentrations de THC, CBD et CBN, ainsi que sur le type de cultivar de la plante. Une meilleure compréhension des différents effets de chaque composé pourrait ouvrir un champ d'applications thérapeutiques pour cette plante [484]. La bioactivité de sept cannabinoïdes isolés (∆9-Tétrahydrocannabinol, Cannabidiol, Cannabigerol, Cannabichromene, Cannabidivarin, Tetrahydrocannabivirin) a été étudiée en relation avec ses propriétés thérapeutiques potentielles vis-à-vis de plusieurs formes de cancer [485]. Seuls le Δ9-Tétrahydrocannabinol et le Cannabidiol ont fait l'objet à ce stade d'essais cliniques, tandis que les autres cannabinoïdes n'ont progressé que jusqu'à des tests *in vivo* [485]. Le potentiel thérapeutique des cannabinoïdes, ainsi que d'autres produits naturels présents dans le cannabis tels que les terpènes, est encore largement inexploité. D'où la nécessité de développer la recherche dans ce sens afin de mieux comprendre les possibilités thérapeutiques de ces substances, en particulier pour lutter contre le cancer. En raison du manque de traitements efficaces accessibles à toutes les populations à un faible coût et avec moins de toxicité, les efforts des chercheurs se sont tournés vers les plantes médicinales telles que Cannabis sativa L et leurs avantages thérapeutiques contre le cancer du sein et du poumon. La WHO a signalé 10.44 millions de décès liés au cancer en 2020, dont 2.26 millions de cas de cancer du sein, 2.21 millions de cas de cancer du poumon, 1.93 million de cas de cancer colorectal et rectal, 1.41 million de cas de cancer de la prostate, 1.20 million de cas de cancer non mélanome de la peau et 1.09 million de cas de cancer d'estomac [486]. Malgré les progrès considérables réalisés dans les méthodes de traitement du cancer dans le monde, la perspective de transformer le cancer en une maladie qui puisse être vécue ou complètement traitée est encore loin. Cette réalité est toutefois considérée comme un puissant motivateur pour les chercheurs dans leurs différentes disciplines liées à ce contexte, avec l'horizon d'atteindre l'objectif souhaité. Cela se manifeste par la large gamme de thérapies contre le cancer disponibles, notamment la thérapie hormonale, la radiothérapie, la chirurgie, la thérapie ciblée par des médicaments / vaccins, la thérapie génique et d'autres traitements par chimiothérapie [487]. En ce sens, les thérapies moléculaires ciblées basées sur l'interférence / l'interaction avec des

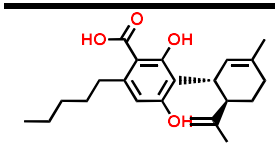
récepteurs moléculaires spécifiques, tels que les protéines / enzymes, ont révolutionné le traitement d'une variété de maladies cancéreuses, y compris le cancer du poumon, du sein, des voies digestives supérieures et de l'ovaire, en tant que moyen de prévenir la croissance du cancer [488]. Dans cette progression, les thérapies ciblées sur EGFR-TKD montrent une nouvelle ère en oncologie précise. Cela est dû au fait que l'EGFR (ou EGFR; ErbB-1; HER1) joue des rôles cruciaux dans le maintien des tissus épithéliaux ainsi que dans la régulation de la division et de la survie cellulaires [489]. Cependant, le récepteur régule de nombreuses voies de signalisation dans le corps, l'absence ou une expression faible d'EGFR et d'autres récepteurs kinases tyrosine chez l'homme entraînant de nombreuses maladies telles que la maladie d'Alzheimer, tandis que l'expression surexprimée d'EGFR-TKD est liée au développement de nombreuses maladies tumorales, notamment les cancers du sein et du poumon, qui sont actuellement les plus mortels [106]. La présente étude porte sur la recherche de nouveaux agents thérapeutiques pour les tumeurs cancéreuses résultant d'une surexpression du récepteur tyrosine kinase (EGFR). L'accent est mis sur les petites molécules inhibitrices de la tyrosine kinase en tant que stratégie thérapeutique ciblée, souvent combinée à la chimiothérapie. Les médicaments Erlotinib, Gefitinib, Afatinib, Dacomitinib, Neratinib, Osimertinib, Afatinib et Lazertinib sont approuvés par la FDA pour le traitement du cancer du sein et du poumon, ainsi que d'autres cancers [106], [490]. Toutefois, ces traitements présentent certains inconvénients, d'où la nécessité de poursuivre la recherche et le dépistage de nouveaux inhibiteurs de la tyrosine kinase (TKI) pour un traitement plus efficace du cancer [491]. C'est pourquoi dans cette étude, nous avons examiné les composés phytochimiques du *Cannabis sativa L*. en tant que nouvelle génération d'EGFR-TKIs qui pourraient avoir une valeur thérapeutique dans le traitement du cancer du sein et du poumon. Le récepteur de facteur de croissance épithélial (EGFR), qui appartient à la famille des kinases tyrosine récepteurs ErbB, médiate des voies de signalisation tyrosine pour transmettre des signaux à l'extérieur/intérieur de la cellule et modifier sa fonction [492]. Il est impliqué dans de nombreux processus physiologiques tels que la croissance cellulaire, la différenciation, la migration et la survie [493]. L'erlotinib, un inhibiteur de première et deuxième génération des kinases tyrosine (TKIs), est un médicament largement utilisé pour plusieurs types de cancer, y compris le cancer du poumon [494]. Cependant, une résistance du patient à cet inhibiteur a été observée [494]. Le tamoxifène, un autre médicament anti-œstrogène utilisé pour le traitement du cancer du sein depuis quatre décennies, a été trouvé pour inhiber l'œstrogène (ER) entraînant une réduction des niveaux de la protéine ERBB2 [495]. Le tamoxifène est toujours utilisé comme traitement du cancer du sein et sa résistance est l'un des principaux problèmes pour le traitement du cancer ; cependant, des études ont montré que ERBB2 et EGFR, étant tous les deux des kinases tyrosine membranaires avec des structures similaires, peuvent conduire à la résistance avec une sur-expression [495], [496]. Par conséquent, le développement précoce de nouveaux EGFR-TKI tels que le Tamoxifen semble être une approche prometteuse pour aborder la problématique de la résistance au Tamoxifen, et à cette fin, des études plus poussées doivent être menées pour étudier sa capacité à inhiber le développement des lignes cellulaires cancéreuses telles que le cancer du sein par le ciblage de l'EGFR-TKD. Pour notre part, dans cette étude, nous avons catégorisé les 50 composés phytochimiques en deux ensembles (Cannaboïdes et Terpènes). Le premier ensemble comprenait 12 Cannaboïdes (C1-C12) et le second ensemble comprenait 38 terpènes (T1-T38) (**Tableau IV. B1**). Sur

la base de techniques de modélisation moléculaire et de leur importance dans les stratégies de conception des médicaments assisté par l'ordinateur (CADD), nous avons effectué un criblage informatique complet de la base de données phytochimiques pour déterminer leurs propriétés d'activité et de biodisponibilité, leurs interactions et leur compatibilité structurelle avec le domaine du récepteur tyrosine kinase Epidermal Growth Factor (EGFR-TKD) impliqué dans la prolifération des lignées cellulaires du cancer du sein et du poumon, ainsi que d'autres types de cancer. Sur la base de nos prédictions de caractéristiques d'efficacité et de risques de toxicité, du criblage en silico ADME-Tox, des inspections moléculaires semi-flexibles et flexibles, des valeurs théoriques des constantes inhibitrices (Ki), des énergies libres MM-GBSA et de l'analyse dynamique des ligands, nous avons criblé les meilleurs ligands qui pourraient inhiber le mieux la croissance des cellules cancéreuses du sein et du poumon en ciblant le EGFR-TKD. Pour faciliter l'identification de nouveaux candidats substances phytochimiques à base de cannabis pour leur utilisation comme médicament computationnel du cancer du sein et du poumon par ciblage du domaine de la tyrosine kinase récepteur du facteur de croissance épidermique (EGFR-TKD). Dans notre présente étude, nous utilisons deux médicaments, le Tamoxifène (PubChem CID: 2733526 et l'Erlotinib (PubChem CID: 176870) comme référence pour construire le criblage hypothétique et faire les comparaisons rationnelles. Cela est dû à l'application clinique large de l'Erlotinib et du Tamoxifène chez les patients atteints de cancer du poumon et du sein.

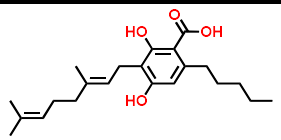
2 Stratégie de conception de médicaments assistée par ordinateur développée et appliquée

Dans l'objectif de concevoir de nouveaux médicaments contre le cancer, cette étude a porté sur les substances naturelles bioactives extraites du *Cannabis sativa L*. en tant que source structurelle pour mener un criblage par voie *in silico*. Cette recherche implique une stratégie de ciblage thérapeutique car l'activité enzymatique excessive de la protéine EGFR-TKD est liée au développement de nombreuses lignées cellulaires cancéreuses mortelles. En vue de découvrir de nouvelles structures moléculaires dotées d'un fort potentiel d'inhibition/modulation de l'EGFR-TKD, une stratégie CADD hybride a été mise en œuvre incluant le repositionnement innovant d'une cinquante de molécules phytochimiques extraites du *Cannabis sativa L*., ainsi que le criblage virtuel et les simulations biophysiques. Erlotinib et Tamoxifen fournissant des points de référence pour la sélection stricte des nouveaux inhibiteurs spécifiques de la protéine EGFR-TKD et pour la limitation des effets néfastes engendrés par le cancer sur la santé. Dans cette étude, nous avons utilisé Erlotinib comme agent standard pour le repositionnement et la sélection d'inhibiteurs spécifiques du EGFR-TKD et d'inhibiteurs de croissance des lignées cellulaires du cancer pulmonaire non à petites cellules. Nous avons également inclus le Tamoxifen pour sa potentialité comme inhibiteur du EGFR-TKD et pour comparer la croissance des lignées cellulaires du cancer du sein. La **Figure IV. B1** montre le processus suivi pour atteindre ces objectifs.

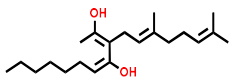
Tableau IV. B1. Chimiothèque des substances phytochimiques à base de Terpènes (T1-T38) et de Cannabinoïdes (C1-C12) extraits de *Cannabis sativa L.*



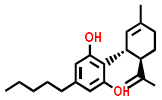
C1 : CBDA
Acide cannabidiolique



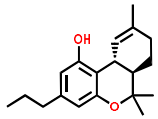
C2 : CBGA
Acide cannabigérolique



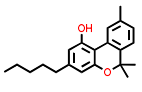
C3 : CBG
Cannabigérol



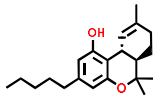
C4 : CBD
Cannabidiol



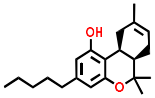
C5 : THCV
Tétrahydrocannabivarine



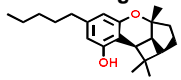
C6 : CNB
Cannabinol



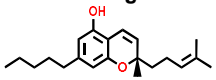
C7 : Δ-9-THC
Dronabinol



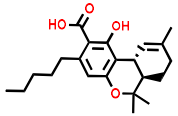
C8 : Δ-8-THC
Δ8-
Tétrahydrocannabinol



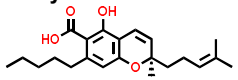
C9 : CBL
Cannabicyclol



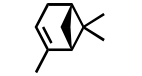
C10 : CBC
Cannabichromène



C11 : THCA
Acide Δ9-
tétrahydrocannabinolique



C12 : CBCA
Acide
cannabichroménique



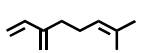
T1 : α-pinene
Alpha-Pinène



T2 : Camphène
Camphène



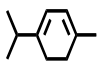
T3 : β-Pinène
Bêta-Pinène



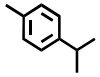
T4 : β-myrcène
Myrcène



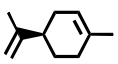
T5 : Δ-3-carène
Delta-3-Carène



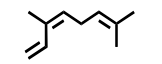
T6 : α-terpinène
Alpha-Terpinène



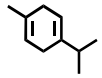
T7 : p-cymène
para-Cymène



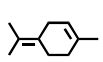
T8 : d-Limonène
D-Limonène



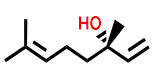
T9 : ocimène
cis-bêta-Ocimène



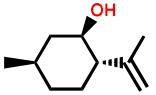
T10 : γ-terpinè
Gamma-Terpinène



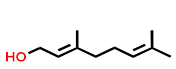
T11 : Terpinolène
Isoterpinène



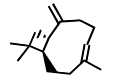
T12 : linalol
Linalol



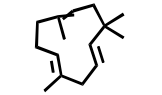
T13 : (-)-isopulégol
Isopulégol



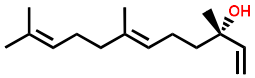
T14 : géraniol
Géraniol



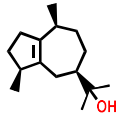
T15 : β-
caryophyllène
Bêta-caryophyllène



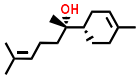
T16 : α-humulène
Alpha-Humulène



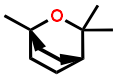
T17 : nérolidol
Nérolidol



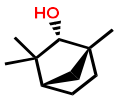
T18 : (-)-guaiol
Guaiol



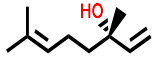
T19 : (-)-α-bisabolol
Alpha-Bisabolol



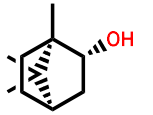
T20 : cinéole
Cinéole



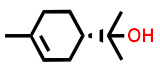
T21 : fenchol
Alcool fenchylique



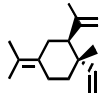
T22 : linalol
Linalol



T23 : bornéol
Bornéol



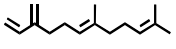
T24 : α-terpinéol
alpha-Terpinéol



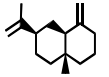
T25 : γ-élemène
Gamma-Elemène



T26 : α-bergomotène
Alpha-Bergamotène



T27 : Bêta-Farnésène
β-Farnésène



T28 : β-eudesmene.
Bêta-Eudesmène

T29 : valencène
Valencène

T30 : α-Bulnesène
Alpha-Bulnesène

T31 : α-farnésène
Alpha-Farnésène

T32 : β-gurjunène
Bêta-Gurjunène

T33 : eudesma-3,7(11)-diène
Eudesma-3,7(11)-diène

T34 : seychellène
seychellène

T35 : δ-sélinène
Delta-Sélinène

T36 : γ-eudesmol
Gamma-Eudesmol

T37 : α-eudesmol
Alpha-Eudesmol

T38 : bulnesol

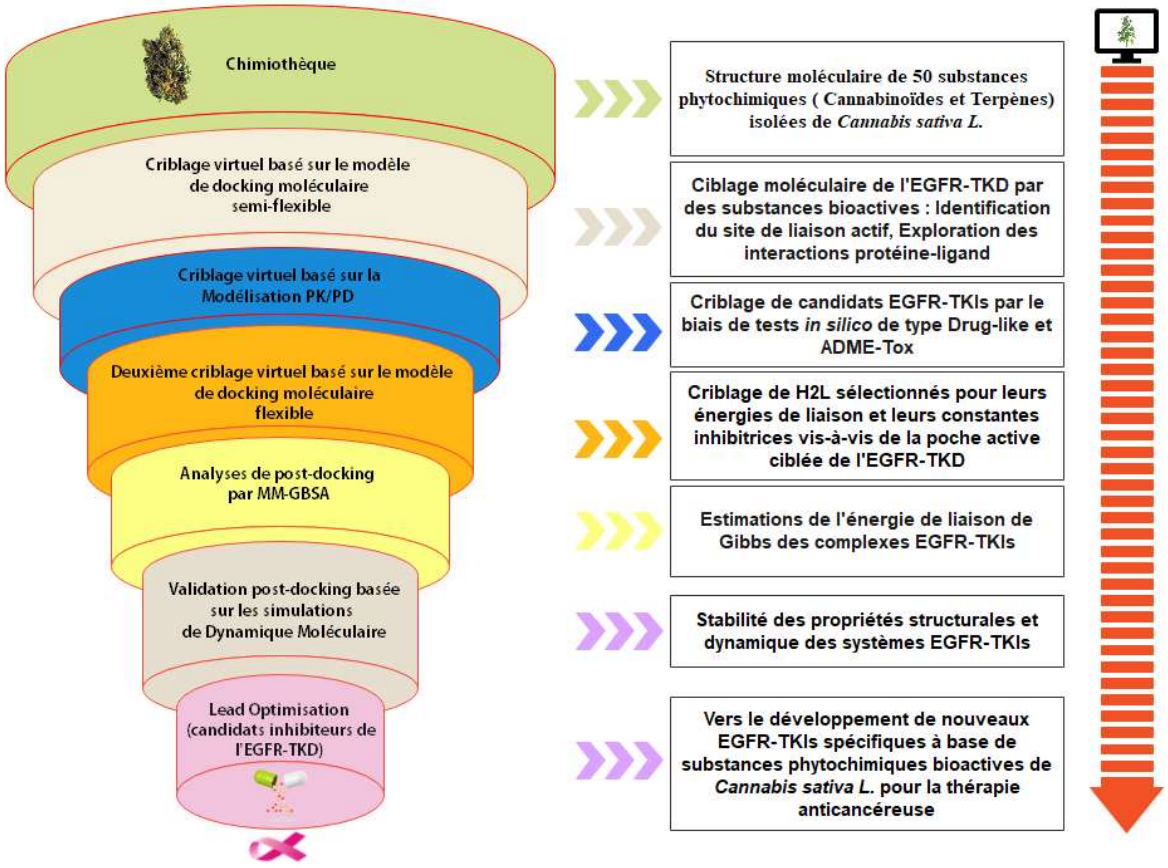


Figure IV. B1. La stratégie CADD implémentée dans cette étude pour identifier de nouveaux agents candidats comme EGFR-TKIs basés sur des substances phytochimiques bioactives de type cannabinoïdes et terpènes pour la thérapie ciblée contre le cancer.

3 Résultats et discussion

Dans cette section, nous présentons les résultats et les conclusions les plus importants auxquels nous sommes parvenus dans le cadre de cette étude. Le reste des détails, des matériels supplémentaires et des autres données relatives à cette étude peuvent être consultés dans la version électronique de notre article publié dans ce contexte [360].

3.1 Criblage virtuel des phytochimiques de Cannabinoïdes et de Terpènes de Cannabis

3.1.1 *Découverte de substances actives à l'aide du Docking Moléculaire*

Après avoir optimisé les structures des ligands et celle de la protéine EGFR-TKD (code PDB : 1M17), nous avons validé le procédé de Docking moléculaire semi-flexible qui sera appliqué dans le cadre de notre première étape de criblage. Nous avons considéré la cavité qui contient les structures des résidus d'acides aminés avec lesquels l'inhibiteur Erlotinib interagit dans le complexe 1M17, comme site de liaison active de la protéine EGFR-TKD. Ces résidus sont : **Lysine** (Lys721), **Valine** (Val702), **Méthionine** (Met769), **Alanine** (Ala719), **Leucine** (Leu694 et Leu820) et **Acide aspartique** (Asp776). Grâce aux packages AutoDock 4.2.6 et AutoDockVina1.1.2, une grille de taille 45×25×20 $Å^3$ et dont les points sont définis à x = 22.030 Å, y = 0.467 Å, z = 52.836 Å et un espacement de 0.375 Å, a été déterminée avant le re-Docking de la conformation flexible de l'inhibiteur Erlotinib dans la poche

active de l'EGFR-TKD. Et, ce dernier interagit comme dans le complexe original 1M17 (**Figure IV. B2**), et la valeur de DockRMSD (0.155<2 Å) confirme que la démarche de Docking moléculaire que nous avons appliquée est pertinente pour analyser l'affinité et la compatibilité des molécules phytochimiques avec la structure de l'EGFR-TKD.

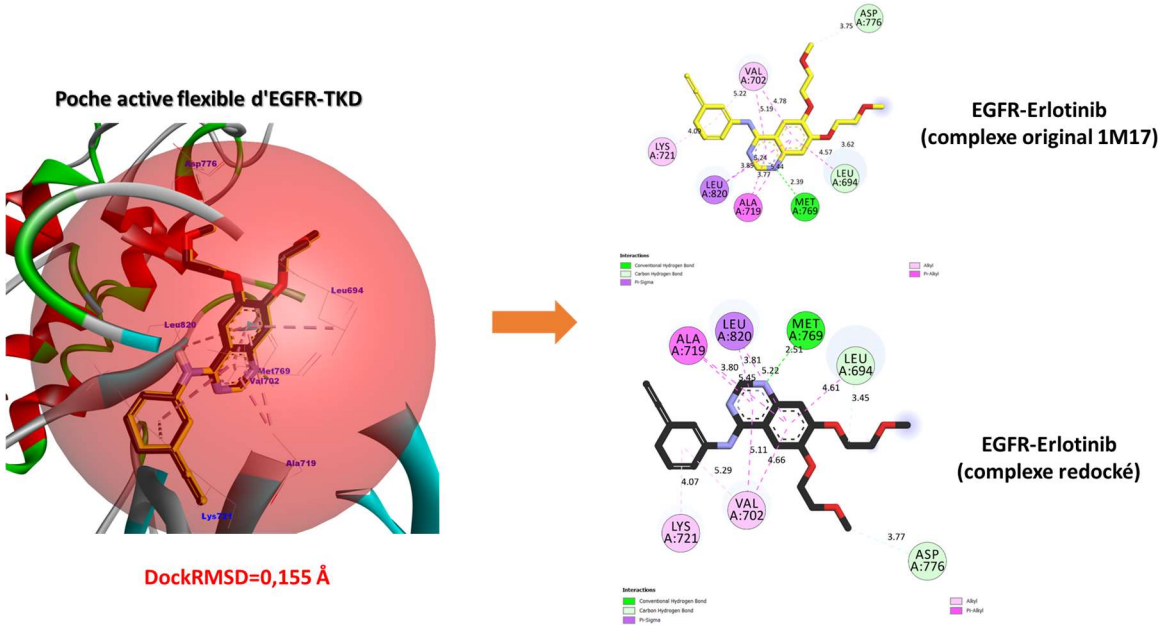


Figure IV. B2. Positions conformationnelles de l'inhibiteur Erlotinib original (jaune) et re-Docké (noir) au sein de la poche active flexible d'EGFR-TKD.

3.1.2 *Prédiction de l'affinité des substances phytochimiques dérivées du Cannabis vis-à-vis le site actif d'EGFR-TKD*

Les affinités de liaison des ligands phytochimiques de cannabinoïdes (C1-C12) et de terpènes de cannabis (T1-T38) et des médicaments de référence Tamoxifen et Erlotinib ont toutes été examinées vis-à-vis du récepteur EGFR-TKD (PDB ID : 1M17). À ce stade de la simulation virtuelle, nous utilisons la technique de Docking moléculaire rapide (semi-flexible) mise en œuvre par les algorithmes du logiciel AutoDockVina1.1.2. Ceci afin d'obtenir un aperçu préliminaire de la disposition et de la classification des substances phytochimiques en fonction de leurs conformations actives et de leur potentiel en tant que nouveaux inhibiteurs spécifiques de l'EGFR-TKD. Les résultats obtenus sont présentés dans le **Tableau IV. B2,** ces résultats indiquent que les cannabinoïdes isolés et les terpènes de cannabis présentent des affinités de liaison significativement plus basses (valeurs plus négatives) vis-à-vis de l'EGFR-TKD, ce qui indique une bonne compatibilité structurelle de ces substances avec la structure de l'EGFR-TKD. Bien que le Tamoxifen (-9.4 kcal/mol) et l'Erlotinib (-8. 2 kcal/mol) aient montré des valeurs plus élevées que la plupart des substances phytochimiques examinées, on a pu identifier certains cannabinoïdes isolés et terpènes de cannabis tels que **C1** (acide cannabidiolique, -9.9 kcal/mol), **C4** (cannabidiol, -9.8 kcal/mol), **C5** (tétrahydrocannabivarine, -9.9 kcal/mol), **C6** (cannabinol, -10.8 kcal/mol), **C7** (dronabinol, -10.7 kcal/mol), **C8** (delta-8-tétrahydrocannabinol, -10.1 kcal/mol), **C9** (cannabicyclol, -10.0 kcal/mol), et **C11** (acide delta9-tétrahydrocannabinolique, -11.3 kcal/mol) ont présenté des valeurs d'affinité de liaison plus basses et peuvent donc s'avérer de valables inhibiteurs non-covalents de l'EGFR (associé aux cancers du sein et du poumon), en comparaison avec les médicaments de référence que sont le Tamoxifen et l'Erlotinib. Les affinités de liaison de **C10**

(Cannabichromène, -8.4 kcal/mol) et **C12** (Acide cannabichroménique, -8.5 kcal/mol) se sont révélées supérieures à celles du Tamoxifen mais plus basses que celles de l'Erlotinib, ce qui indique que ces deux molécules pourraient être bénéfiques du point de vue thérapeutique pour le cancer du poumon. Il ressort du **Tableau IV. B2** que les ligands de la classe des terpènes du cannabis présentent des affinités de liaison significativement limitées vis-à-vis du Tamoxifen. Cependant, 4 substances de cette catégorie T15 (Bêta-caryophyllène, -8.4 kcal/mol), T17 (Nérolidol, -8.8 kcal/mol), T25 (Gamma-Elemène, -8.4 kcal/mol) et T28 (Bêta-Eudesmène, -8.5 kcal/mol) présentent des affinités de liaison plus négatives que l'Erlotinib, destiné à lutter contre le cancer du poumon, ce qui suggère qu'elles sont capables de fonctionner comme des nouveaux inhibiteurs réversibles pour l'EGFR-TKD dans le cadre de la lutte contre le cancer du poumon. De plus, les résultats de criblage du Docking moléculaire suggèrent que le Tamoxifen est davantage en contact avec l'EGFR que l'Erlotinib, ce qui lui permet de réguler la fonction de la tyrosine kinase. Cela signifie qu'il peut être considéré comme un traitement mixte approprié pour le cancer du poumon et le cancer du sein, bien que des recherches plus poussées soient encore nécessaires pour le valider.

Tableau IV. B2. Compilation des affinités de liaison des candidats inhibiteurs de l'EGFR-TKD.

Libellé du ligand	Affinités de liaison (Kcal/mol)	Libellé du ligand	Affinités de liaison (Kcal/mol)	Libellé du ligand	Affinités de liaison (Kcal/mol)	Libellé du ligand	Affinités de liaison (Kcal/mol)
C1	**-9.9**	T2	-5.5	**T15**	**-8.4**	**T28**	**-8.5**
C2	8.1	T3	-5.6	T16	-6.7	T29	-7.2
C3	-7.2	T4	-5.7	**T17**	**-8.8**	T30	-7.4
C4	**-9.8**	T5	-5.7	T18	-7.2	T31	-7.3
C5	**-9.9**	T6	-5.7	T19	-7.2	T32	-6.8
C6	**-10.8**	T7	-5.5	T20	-5.6	T33	-7.6
C7	**-10.7**	T8	-5.7	T21	-6.1	T34	-6.2
C8	**-10.1**	T9	-5.6	T22	-6.0	T35	-7.3
C9	**-10.0**	T10	-5.6	T23	-5.6	T36	-7.3
C10	**-8.4**	T11	-5.5	T24	-5.9	T37	-7.9
C11	**-11.3**	T12	-6.0	**T25**	**-8.4**	T38	-7.9
C12	**-8.5**	T13	-6.0	T26	-7.1	**Tamoxifen**	**-9.4**
T1	-5.6	T14	-6.1	T27	-7.3	**Erlotinib**	**-8.2**

3.2 Prédiction des propriétés pharmacologiques et des risques toxicologiques des inhibiteurs potentiels de l'EGFR-TKD

À ce stade, nous sélectionnons les molécules phytochimiques candidates comme H2L sur la base de leurs propriétés de biodisponibilité, adaptées ou non à un usage médical, ainsi qu'en fonction des risques de toxicité estimés par les calculs d'Osiris (**Tableau IV. B3**). Le **Tableau IV. B3** révèle que les phytochimiques de (*Cannabis sativa L.*) identifiés dans cette étude suivent généralement le protocole de bonne biodisponibilité lorsqu'ils suivent Lipinski RO5 et les règles de Veber. Les deux médicaments de référence Tamoxifen et Erlotinib n'ont pas suivi ces règles, indiquant une absorption insuffisante dans le texte lorsque le médicament est pris par voie orale. De plus, tous les composés avaient une prédiction médicamenteuse de 50% ou plus et un indice d'accessibilité synthétique inférieur à 10,

suggérant une synthèse et une évaluation simples des substances. Les calculs d'Osiris montrent des composés spécifiques (C10, C12 et T17) comme ayant des effets toxiques possibles en ce qui concerne la reproduction, la mutagénicité et l'irritation. Après ces composés sont exclus, les substances phytochimiques restants montrent des promesses en tant que H2L anticancer qui sont plus compatibles à l'utilisation que les médicaments de référence d'Erlotinib et Tamoxifen.

Tableau IV. B3. Modèle de type Drug-like et risques de toxicité associés aux substances criblées.

Substance	Critères de biodisponibilité						DruglikenessVio.		F	SA	Tox-R
	MW	LogP	n-HBA	n-HBD	n-ROB	MR	RO5	Viber	>10%	1-10	+/±/
	≤500 Da	<5	<10	<5	<10	40-130 Å	0≤Vio≤2				
C1 (CBDA)	358.47	3.79	4	3	7	106.81	0	0	0.56	4.25	-
C4 (CBD)	314.46	4.31	2	2	6	99.85	0	0	0.55	4.05	-
C5 (THCV)	286.41	3.94	2	1	2	88.30	0	0	0.55	4.05	-
C6 (CNB)	310.43	4.23	2	1	4	97.10	0	0	0.55	3.39	-
C7 (Δ-9-THC)	314.46	4.39	2	1	4	97.91	0	0	0.55	4.27	-
C8 (Δ-8-THC)	314.46	4.39	2	1	4	97.91	0	0	0.55	4.21	-
C9 (CBL)	314.46	4.48	2	1	4	96.01	0	0	0.55	4.37	-
C10 (CBC)	314.46	4.31	2	1	7	100.34	0	0	0.55	4.26	+RE, +MG
C11 (THCA)	358.47	3.88	4	2	5	104.87	0	0	0.85	4.43	-
C12 (CBCA)	358.47	3.79	4	2	8	107.30	0	0	0.85	4.40	+RE, +MG
T15 (BCP)	204.35	4.63	0	0	0	68.78	0	0	0.55	4.51	-
T17 (NRD)	222.37	3.86	1	1	7	74.00	0	0	0.55	3.53	+IRR
T25 (γ-Ele)	204.35	4.53	0	0	2	70.42	0	0	0.55	3.81	-
T28 (β-EUD)	204.35	4.63	0	0	1	68.78	0	0	0.55	3.42	-
Tamoxifen	371.51	5.10	2	0	8	119.72	1	0	0.55	3.01	+RE
Erlotinib	393.44	1.89	6	1	10	111.40	1	0	0.55	3.19	-

MW : poids moléculaire ; **MR** : réfractivité molaire ; **LogP** : lipophilicité ; **n-HBA** : Nombre d'accepteurs de liaisons hydrogène ; n-**HBD** : Nombre de donneurs de liaisons hydrogène ; **n-ROB** : Nombre de liaisons rotatives ; **nVio** : Nombre de violations ; **F** : score de biodisponibilité ; **SA** : accessibilité synthétique ; **Tox-R** : Risques de toxicité potentiels : **IRR** : Irritant, **RE**: Reproductive effective, **MG** : Mutagène, **TUMO** : Tumorigène, (**+** : hautement toxique ; **-** : non toxique ; **±** : faiblement toxique).

3.3 Caractérisation du comportement des médicaments et de leur toxicité potentielle par modélisation ADME-Tox

Le **Tableau IV. B4** présente les propriétés pharmacocinétiques et pharmacodynamiques obtenues *in silico* via les prédictions ADME-Tox appliquées aux substances phytochimiques C1 (CBDA), C4 (CBD), C5 (THCV), C6 (CNB), C7 (Δ-9-THC), C8 (Δ-8-THC), C9 (CBL), C10 (THCV) et C11 (THCV), et les terpènes de cannabis T15 (BCP), T25 (γ-Ele), T28 (β-EUD) retenus après le criblage de type Drug-like. Pour notre étude de phytocannabinoïdes C1 (CBDA), C4 (CBD), C5 (THCV), C6 (CNB), C7 (Δ-9-THC), C8 (Δ-8-THC), C9 (CBL), C11 (THCA) et des terpènes T15 (BCP), T25 (γ-ELE), T28 (β-EUD), la plupart des composés ont eu un taux d'absorption intestinales humaine élevé (>90%), ce qui était comparable aux deux médicaments, Tamoxifen et Erlotinib. Ces composés ont présenté une perméabilité cutanée optimale (log Kp) et un débit de clearance corporel total (CL), ainsi qu'une bonne distribution volume (VDss) et une perméabilité a la barrière hémato-encéphalique (BBB). Sauf pour C6 (CNB) et T28 (β-EUD), aucun des composés testés ne présentait de profil 'AMES toxique'. La plupart des composés avaient des valeurs LD_{50} plus élevées que le Tamoxifen, mais inférieures à l'Erlotinib. Il n'y avait pas de preuve d'hépatotoxicité ou d'interaction CYP450 pour tous les phytocannabinoïdes, à

l'exception de C6 (CBN), qui présentait une toxicité potentielle et pouvait potentiellement entrer en interaction avec le CYP-2C9. Cependant, CBD, CBG et CNB ont tous été suggérés pour agir comme EGFR-TKIs dans une autre étude [497], mais les propriétés PK et PD de ces molécules n'ont pas été estimées. Dans cette étude, nous concluons que C1 (CBDA), C4 (CBD), C5 (THCV), C7 (Δ-9-THC), C8 (Δ-8-THC), and C9 (CBL), C11 (THCA), T15 (BCP) pourraient s'avérer comme H2L sûrs pour une utilisation comme potentiels médicament, alors que C6 (CNB) et T28 (β-EUD) présentaient une toxicité potentielle.

3.4 Évaluation de la stabilité conformationnelle des complexes EGFR-TKIs

3.4.1 *Étude des modes de compatibilité, d'affinité et d'interaction Protéine-Ligand*

À ce stade, nous avons refait le Docking des neuf H2L qui ont satisfait les essais Drug-like et ADME-Tox dans la poche active de l'EGFR-TKD. Au cours de cette procédure, nous adaptons les structures des ligands et des résidus d'acides aminés actifs dans la poche EGFR-TKD (référence) en modèle totalement flexible, puis nous effectuons la procédure de Docking lente et rigoureuse à l'aide du logiciel AutoDock 4.2.6 avec l'algorithme génétique de type Lamarckien. Les **Figures II. B (3-4)** montrent les résultats obtenus par le Docking moléculaire flexible et les analyses Prime MM-GBSA, séparément. Le modèle de Docking moléculaire flexible a fourni des données sur l'énergie libre de liaison estimée (E.L), l'activité inhibitrice biologique estimée (K_i) et le mode des interactions entre les ligands et le site actif de l'EGFR-TKD. Les calculs Prime/MM-GBSA ont permis de générer à nouveau et de mesurer l'énergie libre de liaison (ΔG_{bind}) et les interactions protéine-ligand pour chaque complexe EGFR-inhibiteur obtenu par Docking moléculaire flexible. L'analyse des **Figures II. B3 et II. B4** nous a permis de tirer les conclusions suivantes : Les valeurs de l'énergie de liaison et de la concentration inhibitrice des cannabinoïdes isolés vont de -12.79 à -8.16 kcal/mol et de 0.4115 E^{-3} µM à 1.0281 µM, respectivement. Les terpènes de cannabis présentent des valeurs d'énergie de liaison de -7.61 à -7.41 kcal/mol, tandis que les valeurs de concentration inhibitrice sont comprises entre 2.603 µM et 3.650 µM. Les médicaments de référence ont des valeurs d'énergie de liaison de -8.79 à -8.29 kcal/mol et des valeurs de concentration inhibitrice de 0.4000 µM à 0.8253 µM. Le modèle Prime MM-GBSA a ajusté les énergies de liaison libres à -62.807 à -31.138 kcal/mol (Cannabinoïdes isolés) et de -39.782 à -33.980 kcal/mol (Terpènes de cannabis). Les interactions protéine-ligand des cannabinoïdes isolés présentent des liaisons hydrogène (conventionnelles/carbone), hydrophobes, électrostatiques (complexe EGFR-C11) et sulfureuses (complexes EGFR-C5 et C7). Les terpènes du cannabis n'interagissent que par liaison hydrophobique (alkyle). Le Tamoxifen et l'Erlotinib forment des liaisons de type hydrogène (conventionnel/carbone), hydrophobes et électrostatiques (π-Cation) avec la poche active de l'EGFR-TKD. Les simulations de conditions physiologiques obtenus par les prédictions d'Epik (prédiction empirique du pKa) ont révélé que l'acide carboxylique dans C1 (CBDA) et C11 (THCA) avait perdu un proton pour former l'anion carboxylate, tandis que les autres groupes fonctionnels dans C4 (CBD), C5 (THCV), C7 (Δ-9-THC), C8 (Δ-8-THC), C9 (CBL), T15 (Δ-8-THC), T25 (γ-Ele) et Erlotinib conservaient leur état de protonation. Pendant ce temps, Tamoxifène a acquis un proton à sa moitié ammoniaque, résultant en la formation d'un cation ammonium.

Tableau IV. B4. Analyse *in silico* des propriétés ADME-Tox des substances phytochimiques candidates comme H2L.

	Modèles																	
	Absorption		Distribution				Métabolisme							Excrétion	Toxicité			
							Cytochrome P450 (CYP450)											
							Substrat		Inhibiteur									
Propriétés	Absorption intestinale (humain)	Perméabilité de la peau	VDss (humain)	Fraction non liée (humain)	Perméabilité de la BHE (BBB)	Perméabilité du SNC (CNS)	2D6	3A4	1A2	2C19	2C9	2D6	3A4	Clairance totale	Toxicité d'AMES	ORAT (LD_{50})	Hépatotoxicité	Dose maximale tolérée (humain)
Unité	%	Log Kp	Log L kg-1	-	Log BB	Log PS	Oui/non							Log mL min^{-1} kg^{-1}	Oui/non	Mol/kg	Oui/non	Log mg/kg/jour
C1	98.61	-2.73	-1.412	0.09	-0.93	-1.97	Non	Non	Non	Non	Non	Non	Non	0.98	Non	2.47	Non	0.52
C4	91.02	-2.88	0.76	0.06	0.05	-1.40	Non	Oui	Oui	Non	Non	Non	Non	1.13	Non	2.29	Non	-0.37
C5	92.56	-2.75	0.78	0.06	0.30	-1.80	Non	Non	Oui	Oui	Non	Non	Non	0.93	Non	2.18	Non	-0.152
C6	93.18	-2.75	0.65	0.01	0.60	-1.06	Non	Oui	Oui	Oui	Oui	Non	Non	0.95	**Oui**	2.35	Non	0.26
C7	91.87	-2.75	0.85	0	0.42	-1.69	Non	Oui	Oui	Oui	Non	Non	Oui	0.98	Non	2.23	Non	-0.15
C8	91.87	-2.78	0.85	0	0.42	-1.69	Non	Oui	Oui	Oui	Non	Non	Oui	0.97	Non	2.23	Non	-0.15
C9	91.92	-2.80	0.86	0	0.47	-1.61	Non	Oui	Oui	Oui	Non	Non	Non	0.55	Non	2.33	Non	-0.52
C11	97.56	-2.73	-1.22	0.09	-0.13	-1.89	Non	Non	Non	Non	Non	Non	Non	0.74	Non	2.77	Non	0.56
T15	95.30	-1.59	0.65	0.26	0.74	-2.18	Non	Non	Non	Non	Non	Non	Non	1.08	Non	.1.67	Non	0.48
T25	95.43	-1.27	0.59	0.14	0.79	-1.68	Non	Non	Non	Non	Non	Non	Non	1.41	Non	1.58	Non	0.19
T28	94.76	-1.89	0.60	0.09	0.81	-1.40	Non	Oui	Oui	Non	Non	Non	Non	1.17	**Oui**	1.61	Non	-0.42
Tam	97.8	-2.73	0.40	0.16	1.40	-1.61	Non	Oui	Oui	Non	Non	Oui	Non	0.60	**Oui**	2.17	Non	0.52
Erl	95.26	-2.73	-0.27	0.14	-1.01	-3.46	Non	Oui	Oui	Oui	Oui	Non	Oui	0.64	Non	2.81	**Oui**	0.91

VDss : Volume de distribution par la méthode de l'état stable, BHE : Barrière hémato-encéphalique, SNC : Système nerveux central. ORAT : Toxicité aiguë par voie orale chez le rat.
Les molécules marquées en rouge présentent une toxicité potentielle.

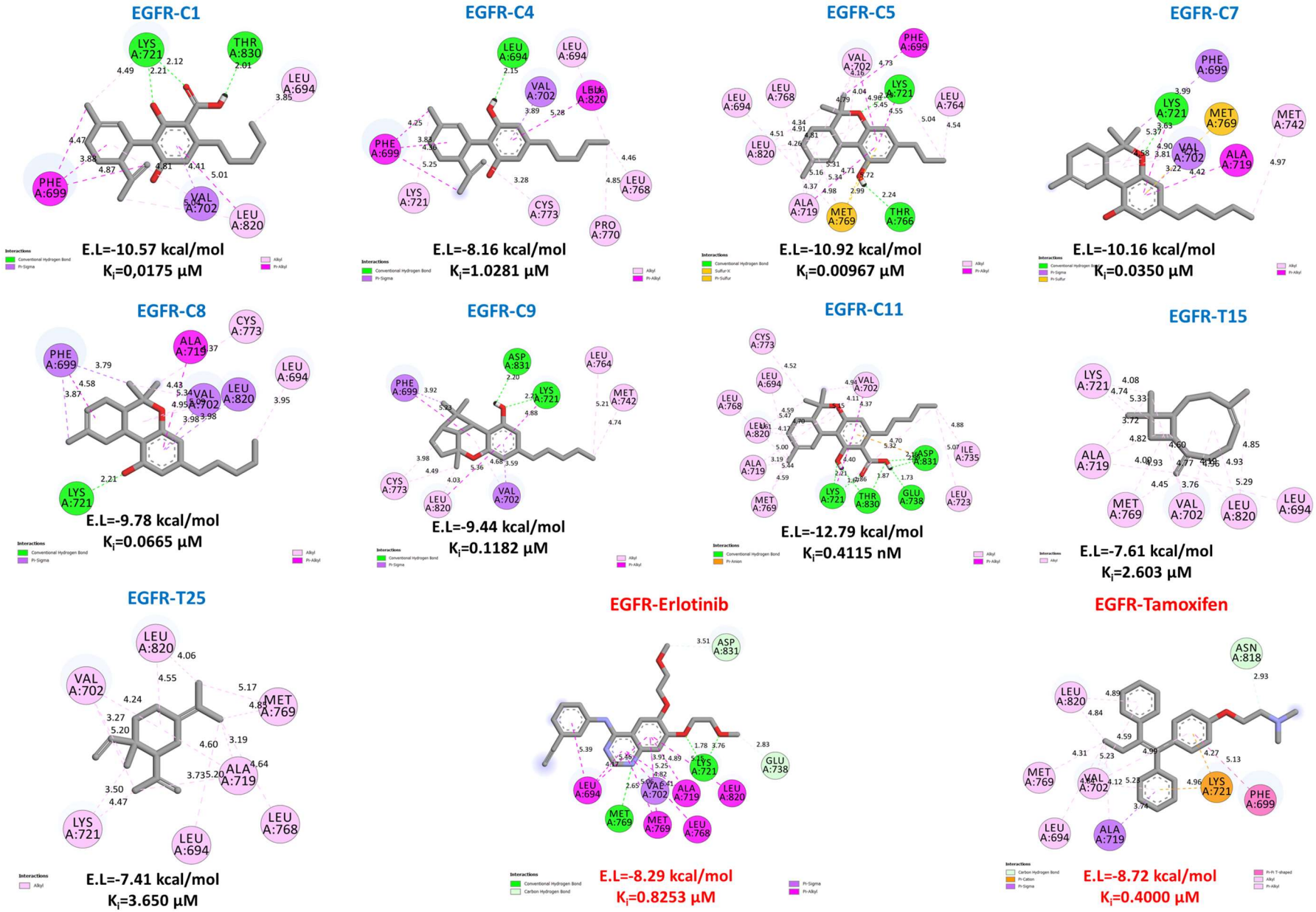


Figure IV. B3. Prédictions de modèle de Docking moléculaire flexible appliquées aux nouveaux EGFR-TKIs basés sur les phytochimiques de type Cannabinoïdes et Terpènes.

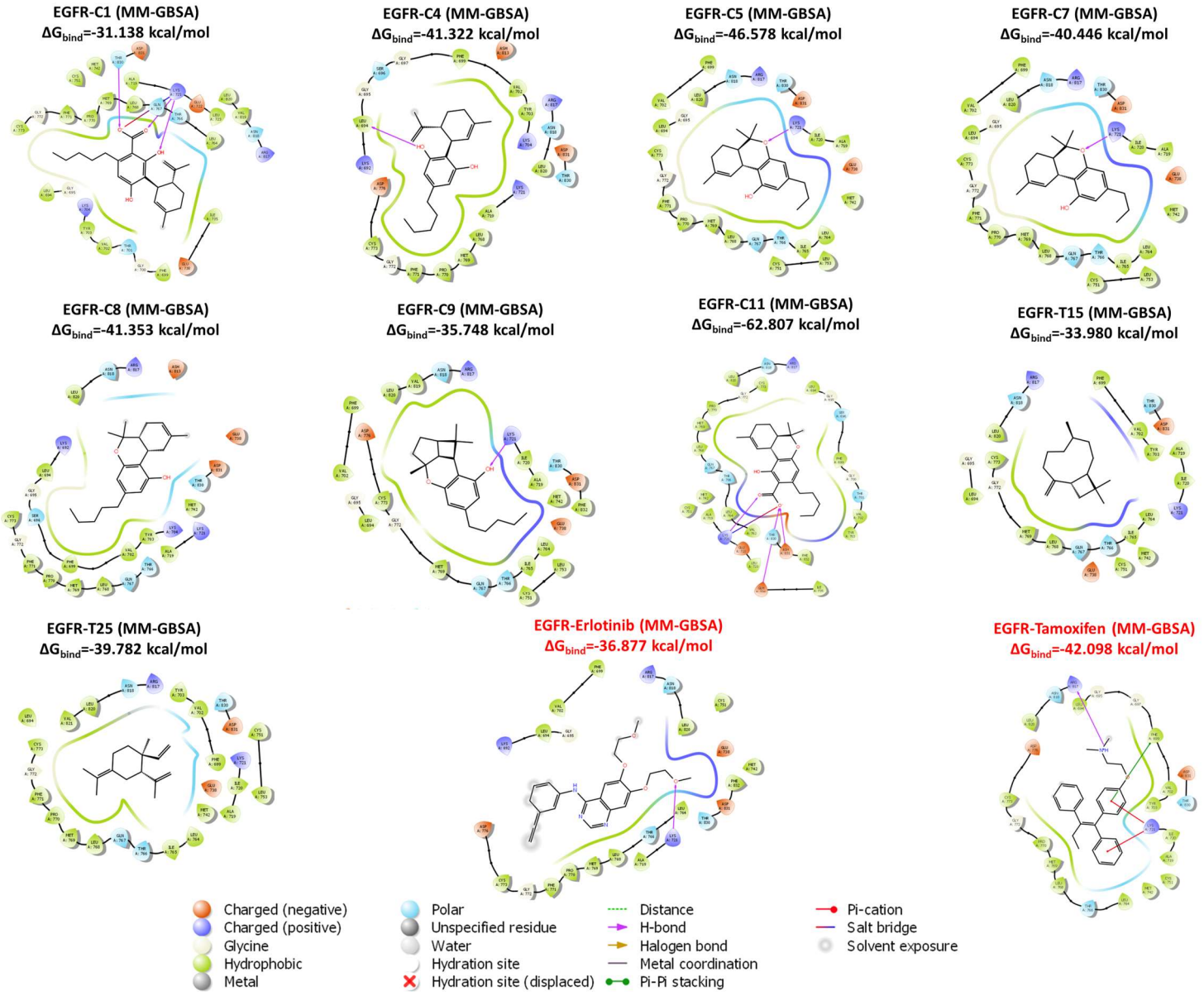


Figure IV. B4. Prédictions du modèle Prime/MM-GBSA appliquées aux complexes EGFR-TKIs générés.

3.4.2 *Découverte de candidats Lead Optimization (LO) à l'aide de simulations de Docking moléculaire et MM-GBSA*

Les composés phytochimiques du *Cannabis sativa L.* (CBDA, CBD, THCV, Δ-9-THC, Δ-8-THC, THCA et γ-Ele) ont des structures moléculaires de petite taille par rapport à deux médicaments standard, le Tamoxifen et l'Erlotinib, et présentent des propriétés ADME-Tox et de type "Drug-like" favorables, ce qui indique la possibilité de les utiliser dans le cadre d'une thérapie ciblée basée sur l'EGFR-TKD. Une analyse approfondie des interactions de ces composés avec une poche active de l'EGFR-TKD, prenant en compte des paramètres tels que la concentration inhibitrice estimée (Ki) et l'énergie libre de liaison (ΔG_{bind}), a montré que le C11 : THCA (K_i=0.4115 nM, ΔG_{bind}=-62.807 kcal/mol) était l'échantillon le plus prometteur de la catégorie des cannabinoïdes isolés et que le T15 : BCP (K_i=2.603 µM, ΔG_{bind}=-33. 980 kcal/mol) était le plus prometteur de la catégorie des terpènes de cannabis. Les deux échantillons C11 et T15 ont montré une concentration inhibitrice estimée très minime qui assure l'inhibition de 50 % de l'activité enzymatique de l'EGFR-TKD et une énergie de liaison libre plus basse que les autres échantillons retenus comme agents EGFR-TKI dans le cadre de cette étude. Par conséquent, nous pouvons adopter les structures C11 et T15 comme les deux molécules phytochimiques principales parmi les neuf molécules H2L en vue du développement optimal de ces substances. Par conséquent, les échantillons C11 (THCA) et T15 (BCP) ainsi que les deux références Tamoxifen et Erlotinib font l'objet davantage des investigations complémentaires. Les **Figures IV. B(5-8)** montre un profil des interactions non covalentes entre la poche active de l'EGFR-TKD et les échantillons de base C11, T15, Erlotinib et Tamoxifen, obtenues par le biais du Docking moléculaire flexible et des calculs Prime/MM-GBSA exécutés dans un modèle de solvatation VSGB 2.1 (en anglais, Variable dielectric Surface Generalized Born). Il est important de noter que les algorithmes de Schrödinger changent les noms des résidus standard tels que ARG, ASP, GLU, LYS et HIS en résidus non standard RNA, ASH, GLH, LYN, HIE et HIP. Il s'agit d'une convention informelle permettant d'attribuer différents états de proton à des résidus réguliers.

3.4.2.1 *Interactions EGFR-THCA*

L'analyse des résultats obtenus à partir du complexe EGFR-C11 (**Figure IV. B5**) a montré que la liaison du C11 (THCA) à la poche active de l'EGFR était obtenue par la formation d'interactions non covalentes de liaisons hydrogène et d'interactions alkyl à caractère hydrophobe avec des résidus spécifiques de la poche, tels que ALA719, MET769, LEU820, CYS773, VAL702, VAL768, LEU694, LEU723, ILE735, LYS721, LEU820, VAL702, LYS721, et ASP831. L'analyse de la Prime MM-GBSA (énergie) associée à la solution VSGB a confirmé la stabilité de ces interactions et la capacité de l'échantillon de THCA à inhiber l'activité enzymatique de l'EGFR. Cela suggère que ce type de cannabinoïdes peut être considéré comme un agent thérapeutique anticancéreux potentiel pour cibler/inhiber spécifiquement l'activité enzymatique de l'EGFR.

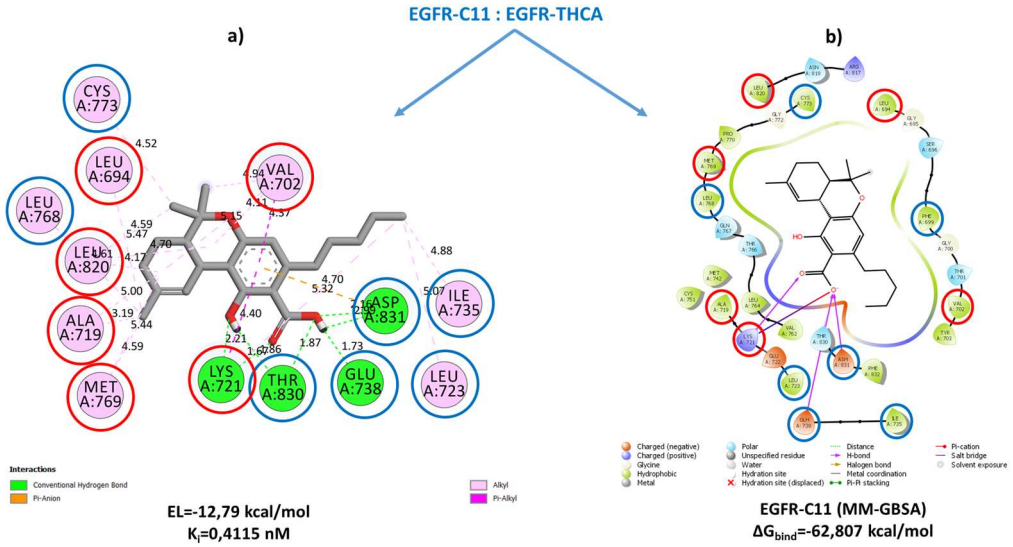


Figure IV. B5. Interactions EGFR-C11. (a) Prédictions de Docking moléculaire flexible. (b) Prédictions de Prime/MM-GBSA. Les acides aminés entourés d'un cercle rouge indiquant un site de liaison actif de référence, les acides aminés entourés d'un cercle bleu indiquant de nouveaux sites actifs de liaison.

3.4.2.2 *Interactions EGFR-BCP*

Sur la base du complexe EGFR-T15 (**Figure IV. B6**), les modèles basés sur le Docking moléculaire flexible et les simulations Prime/MM-GBSA indiquent que la substance phytochimique à base de terpène de cannabis (BCP) forme des interactions non covalentes de type alkyle hydrophobe avec les résidus ALA719, VAL702, LYS721, MET769, LEU820 et LEU694 situés dans la poche active de l'EGFR. Ces modélisations impliquent que le BCP et le γ-Ele peuvent être des inhibiteurs appropriés de l'activité enzymatique de l'EGFR-TKD, ce qui pourrait ouvrir la voie à de nouvelles thérapies anticancéreuses potentielles.

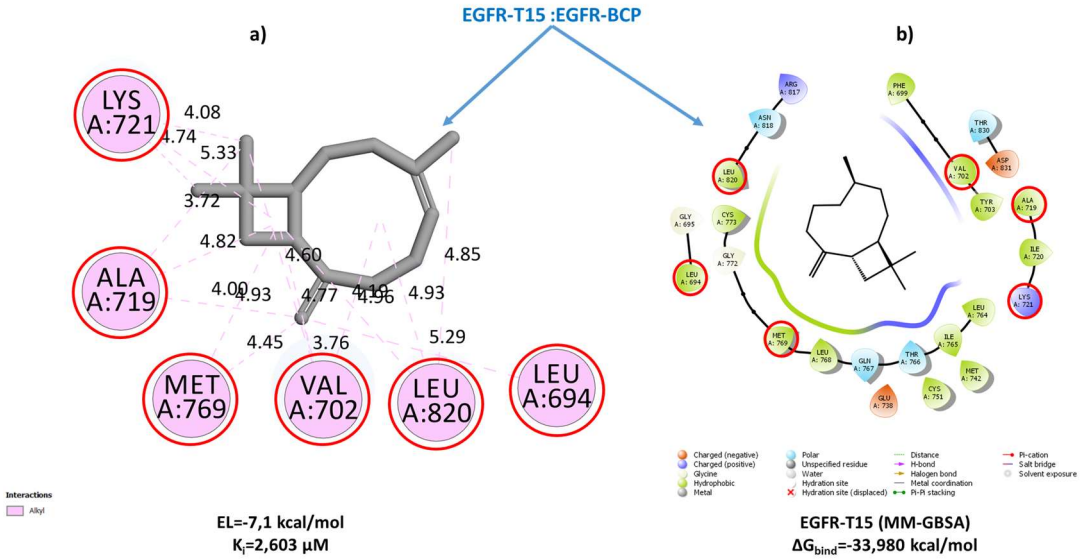


Figure IV. B6. Interactions EGFR-T15. (a) Prédictions de Docking moléculaire flexible. (b) Prédictions de Prime/MM-GBSA. Les acides aminés entourés d'un cercle rouge indiquant un site de liaison actif de référence, les acides aminés entourés d'un cercle bleu indiquant de nouveaux sites actifs de liaison.

3.4.2.3 *Interactions EGFR-Erlotinib*

La flexibilité conformationnelle du médicament de référence Erlotinib lui a permis d'établir plusieurs interactions non covalentes avec les résidus MET769, LYS721, ASP831, GLU738, VAL702, LEU694,

ALA719, LEU768, LEU769, LEU820 dans la poche active de l'EGFR (**Figure IV. B7**). Il s'agit notamment de liaisons hydrogène, de liaisons hydrogène carbone, d'interactions hydrophobes de type π-Sigma et π-alkyl. Les calculs Prime MM-GBSA ont confirmé que toutes les interactions entre l'Erlotinib et la poche active de l'EGFR étaient conservées. Ces résultats valident le modèle EGFR-Erlotinib (PDB ID : 1M17) en tant que référence utile pour simuler les mécanismes d'interactions protéine-médicament et pourraient être utilisés pour des comparaisons rationnelles et théoriques afin de soutenir la conception et la découverte de nouvelles générations d'EGFR-TKIs pour la thérapie ciblée anticancéreuse.

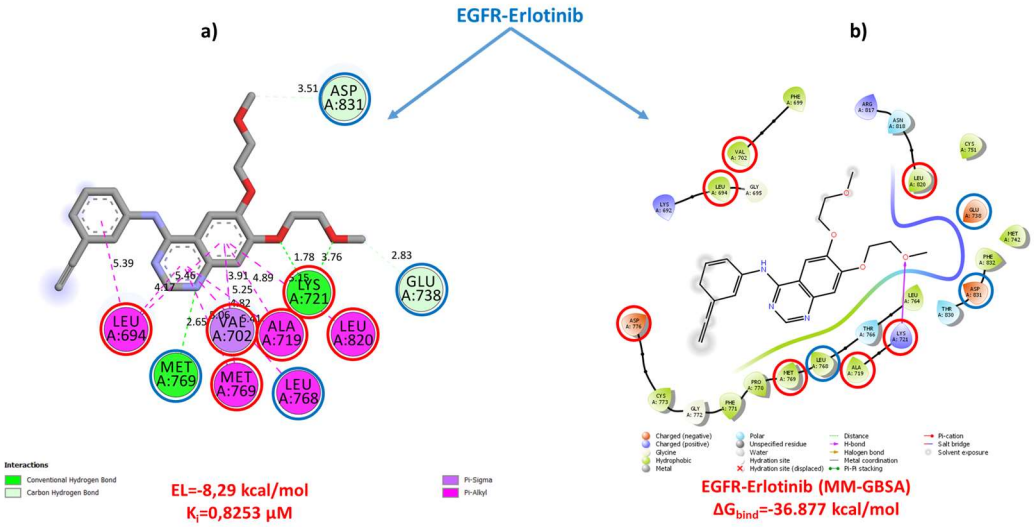


Figure IV. B7. Interactions EGFR-Erlotinib. (a) Prédictions de Docking moléculaire flexible. (b) Prédictions de Prime/MM-GBSA. Les acides aminés entourés d'un cercle rouge indiquent un site de liaison actif de référence, les acides aminés entourés d'un cercle bleu indiquent de nouveaux sites actifs de liaison.

3.4.2.4 *Interactions EGFR-Tamoxifen*

Les simulations de Docking moléculaire flexible du complexe EGFR-Tamoxifen ont révélé que le Tamoxifen se liait à la poche active de l'EGFR par des interactions non covalentes de type hydrogène, carbone, hydrophobique, de type alkyle, π-alkyle, π-π en forme de T et électrostatiques de type π-Cation avec les résidus d'acides aminés de référence ASN818 (2. 93 Å), ALA719 (3.73 Å), LEU694 (4.83 Å), VAL702 (5.22 Å, 4.98, 4.58, 5.23 Å), MET769 (4.30 Å), LEU820 (4.84 Å, 4.89 Å), PHE699 (5.13 Å) et LYS721 (4.26, 4.96 Å). Parallèlement, les simulations de Prime/MM-GBSA ont montré que le Tamoxifen conservait généralement ses interactions prédites lorsqu'il était en contact avec l'EGFR (**Figure IV. B8**). Notamment, les interactions du Tamoxifen avec ALA719, VAL702, MET769, LEU820 et LYS721 démontrent que ce médicament peut être considéré comme un agent inhibiteur de l'EGFR. Par conséquent, ce médicament pourrait être adapté à la recherche et à la classification des inhibiteurs enzymatiques de l'EGFR utilisés dans le traitement du cancer du sein.

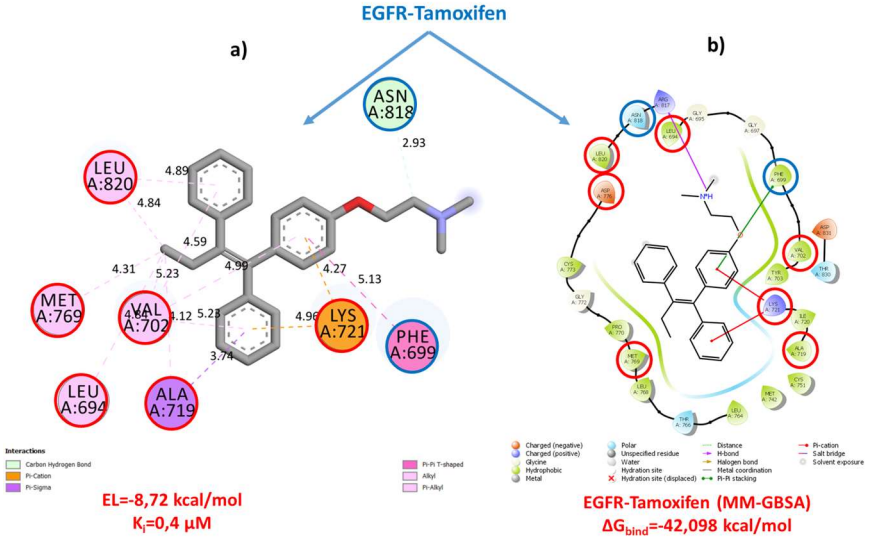


Figure IV. B8. Interactions EGFR-Tamoxifen. (a) Prédictions de Docking moléculaire flexible. (b) Prédictions de Prime/MM-GBSA. Les acides aminés entourés d'un cercle rouge indiquent un site de liaison actif de référence, les acides aminés entourés d'un cercle bleu indiquent de nouveaux sites actifs de liaison.

3.5 Analyse post-Docking des complexes EGFR-TKIs par simulations de Dynamique Moléculaire

Les résultats des criblages effectués dans le cadre de cette étude suggèrent que les substances phytochimiques extraites de *Cannabis sativa L* sont étroitement compatibles avec la structure de l'EGFR-TKD, ce qui indique qu'elles pourraient avoir le potentiel de développer une nouvelle génération d'inhibiteurs de l'EGFR-TKD. Pour confirmer et valider cette hypothèse, des simulations de dynamique moléculaire de quatre échantillons indépendants (EGFR-C11, EGFR-T15, EGFR-Erlotinib et EGFR-Tamoxifen) ont été réalisées en présence des médicaments de référence Erlotinib et Tamoxifen dans un environnement aqueux sur une période de 200 ns. Après avoir analysé les propriétés dynamiques et structurelles des systèmes échantillonnés, nous avons également déterminé avec précision l'énergie totale, l'énergie potentielle, la température, le volume et la pression. Tous ces paramètres peuvent être utilisés pour évaluer le potentiel thérapeutique des médicaments proposés comme base de référence pour combattre le cancer du poumon et du sein, ce qui encourage encore davantage la possibilité d'utiliser les substances phytochimiques indiquées comme inhibiteurs de l'EGFR-TKD.

3.5.1 Investigation dynamique des modifications de RMSD et RMSF

Les écarts quadratiques moyens (RMSD) et des fluctuations (RMSF) au cours des 200 ns de simulations de dynamique moléculaire (DM) des systèmes de backbone libre de l'EGFR et EGFR-Tamoxifen, EGFR-Erlotinib, EGFR-C11 et EGFR-T15 sont mesurés et reproduits dans la **Figure IV. B9(a-c).** D'après la **Figure IV. B9(a)**, nous pouvons constater que les valeurs RMSD se sont stabilisées après ~40 ns des trajectoires de simulations de la DM allant jusqu'à ~200 ns. Pour les cinq systèmes d'EGFG considérés, la moyenne des RMSD±SD des structures thermiques obtenues à partir du point de référence (t=0 ns) jusqu'au point final (t=200 ns) est de 2.10 ± 0.45 Å (EGFR libre), 3.48 ± 0,34 Å (EGFR-C11), 2.14 ± 0.37 Å (EGFR-T15), 3.44 ± 0.81 Å (EGFR-Tamoxifen) et 3.35 ± 0.30 Å (EGFR-Erlotinib), respectivement. Les faibles valeurs d'écart-type (SD), inférieures à 1 Å, suggèrent que la conformation structurelle de la protéine de l'EGFR libre et complexée n'a pas subi de changements

significatifs au cours des simulations de la DM. La **Figure IV. B9(b)** décrit la progression des valeurs de la RMSF dans la séquence de la protéine EGFR. Les systèmes sont mesurés régulièrement par la simulation DM pendant 200 ns afin d'évaluer les changements subis par la protéine EGFR. Les valeurs RMSF±SD moyennes pour l'EGFR libre et celui complexé avec les ligands THCA, BCP, Tamoxifen et Erlotinib étaient respectivement de 2.02±0.81 Å, 1.70±0.95 Å, 1.72±1.06 Å, 3.72±1.22 Å, 1.58±1.29 Å. Il en ressort que les chaînes latérales de la protéine EGFR non complexée (EGFR) et celles des complexes EGFR-C11, EGFR-T15 et EGFR-Erlotinib ont subi de légères variations, tandis que le système EGFR-Tamoxifen a subi des fluctuations plus importantes pendant le même laps de période. En outre, nous avons constaté que les chaînes latérales des acides aminés des queues N et C-terminales présentaient davantage de fluctuations que les autres parties de la protéine, en particulier pour les acides aminés numérotés entre 290 et 312. Pour mieux comprendre comment les ligands optimisés ("lead optimized") de THCA (C11) et BCP (T15) interagissent avec la protéine, les valeurs RMSF des ligands au cours de la simulation MD sont mesurées dans la **Figure IV. B9(c)**. Les grandes fluctuations observables des valeurs RMSF des ligands suggèrent la flexibilité de ces petites molécules qui peuvent fournir différents modèles d'interaction au sein de la poche active de la protéine d'EGFR-TKD.

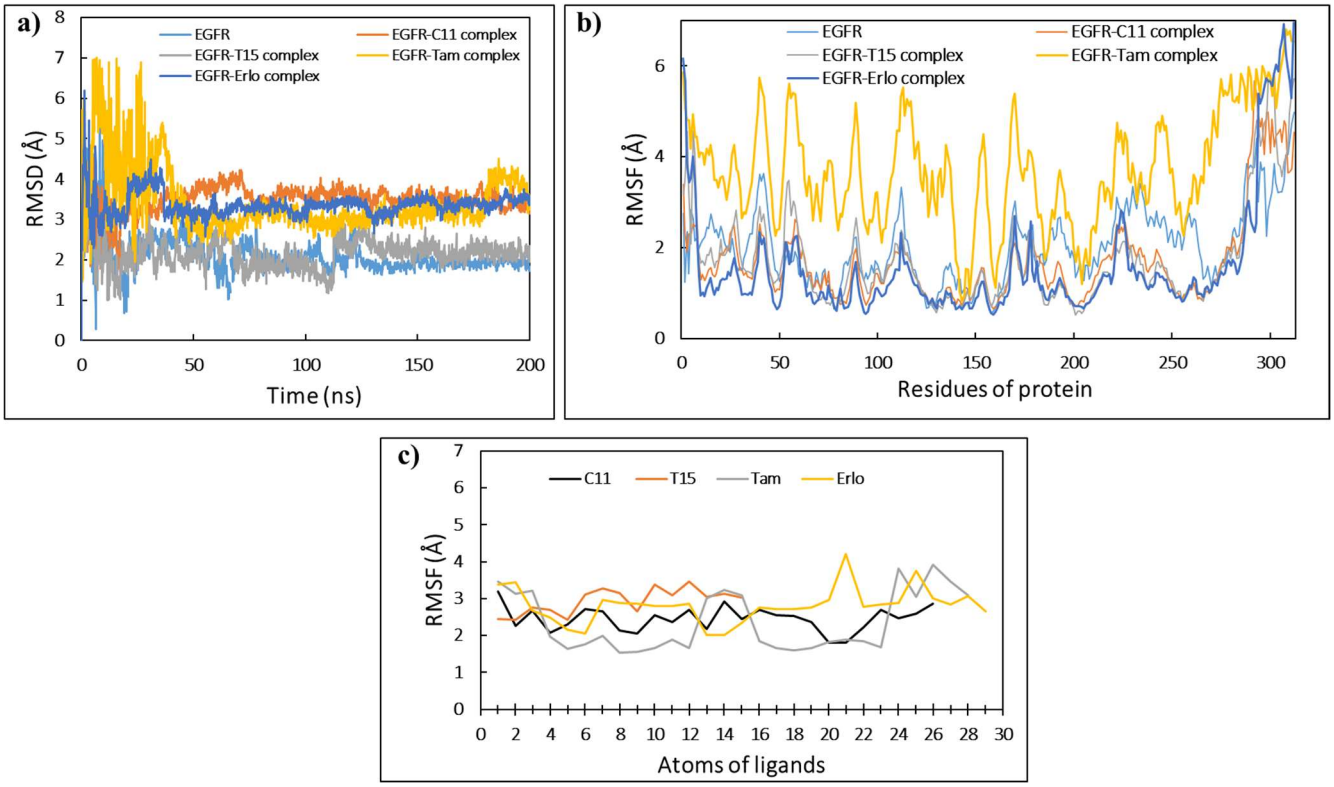


Figure IV. B9. Tracés temporels des simulations de la DM pour les systèmes d'EGFR libres et complexés avec les ligands C11, T15, Tamoxifen, et Erlotinib. (a) RMSD des squelettes C-alpha des systèmes EGFR, (b) RMSF des chaînes latérales des systèmes EGFR, (c) RMSF des ligands complexés avec l'EGFR.

3.5.2 *Caractérisation des interactions Protéine-Ligand au cours de la simulation de DM*

L'analyse de la trajectoire de la DM des interactions entre l'EGFR et les ligands C11, T15, Tamoxifen et Erlotinib présentée à la **Figure IV. B10** indique que le contact protéine-ligand a maintenu une grande partie du temps de simulation. Les contacts ligand-protéine ont été identifiés comme s'étant

produits pendant plus de 10 % du temps de simulation entre les atomes de la protéine et des ligands. De plus, des diagrammes d'interactions et de contacts ont été générés, qui présentent le nombre total de contacts spécifiques avec la protéine et détaillent les résidus qui ont interagi avec le ligand pendant chaque frame de la simulation. Les résidus ayant plus d'un contact avec le ligand sont plus foncés en orange sur le côté droit du diagramme. L'analyse des schémas de contact entre les substances phytochimiques de *Cannabis sativa L.* et les protéines EGFR-TKD a révélé que les résidus d'acides aminés Leucine, Phénylalanine, Valine, Lysine, Cystéine et Thréonine dans la poche protéique de l'EGFR étaient particulièrement intéressants, car ils interagissaient activement avec les ligands THCA et BCP par le biais de liaisons H, d'interactions de type hydrophobique, ionique et de formation de ponts d'eau. Ces acides aminés peuvent servir comme sites actifs de référence fondamentaux dans l'inhibition de l'activité enzymatique de l'EGFR-TKD et la modélisation de la conception de nouveaux médicaments, ainsi que pour atteindre une fonction biologique souhaitée, telle que l'inhibition de la croissance des lignées cellulaires du cancer du sein et du poumon.

3.5.3 Caractérisation temporelle des conformations des ligands

La **Figure IV. B11** présente l'évolution temporelle des propriétés majeures des ligands C11 (THCA), T15 (BCP), Tamoxifen et Erlotinib enregistrées au cours de la simulation de la DM. Le graphique de l'écart quadratique moyen (RMSD) indique les variations de la RMSD du ligand par rapport à la première conformation de référence au temps t=0, tandis que le graphique du rayon de giration (rGyr) reflète la dilatation du ligand. Le graphique des liaisons hydrogène intramoléculaires (intraHB) indique le nombre de liaisons hydrogène internes (HB) au sein du ligand, le graphique de la surface moléculaire (MolSA) mesure la surface moléculaire calculée avec un rayon de sonde de 1.4 Å, et les graphiques de la surface accessible au solvant (SASA) et de la surface polaire (PSA) mesurent la surface de la molécule qui peut être accessible aux molécules d'eau et la surface accessible au solvant, respectivement. L'analyse présentée à la **Figure IV. B11** révèle que les valeurs de la RMSD associée à la stabilité des ligands C11 (THCA), T15 (BCP), Tamoxifen et Erlotinib se sont situées dans la plage de 0.5-1Å, 0.3-0,9Å, 1-1.5Å, et 1-1,2Å, respectivement. Les valeurs du rayon de giration (rGyr) de ces ligands étaient de 3.9-4Å, 2.75-2.80Å, 4.00-4.25Å et 4.80-4.95Å. En outre, une liaison hydrogène intramoléculaire (IntraHB) a été observée dans le ligand THCA, alors qu'aucune n'a été observée dans les ligands BCP, Tamoxifen ou Erlotinib. En outre, le paramètre de stabilité de la surface moléculaire (MolSA) pour les ligands THCA, BCP, Tamoxifen et Erlotinib variait entre 350-360 $Å^2$, 232-236 $Å^2$, 375-390 $Å^2$ et 400-405 $Å^2$; les valeurs des paramètres de stabilité de la surface accessible au solvant (SASA) étaient de 50-55 $Å^2$, 90-100 $Å^2$, 140-160 $Å^2$, et 160-200 $Å^2$; et les valeurs des paramètres de stabilité de la surface polaire (PSA) étaient de 1120-120 $Å^2$, 0$Å^2$, 5-10 $Å^2$, 64-72 $Å^2$, et 64-120 $Å^2$, respectivement.

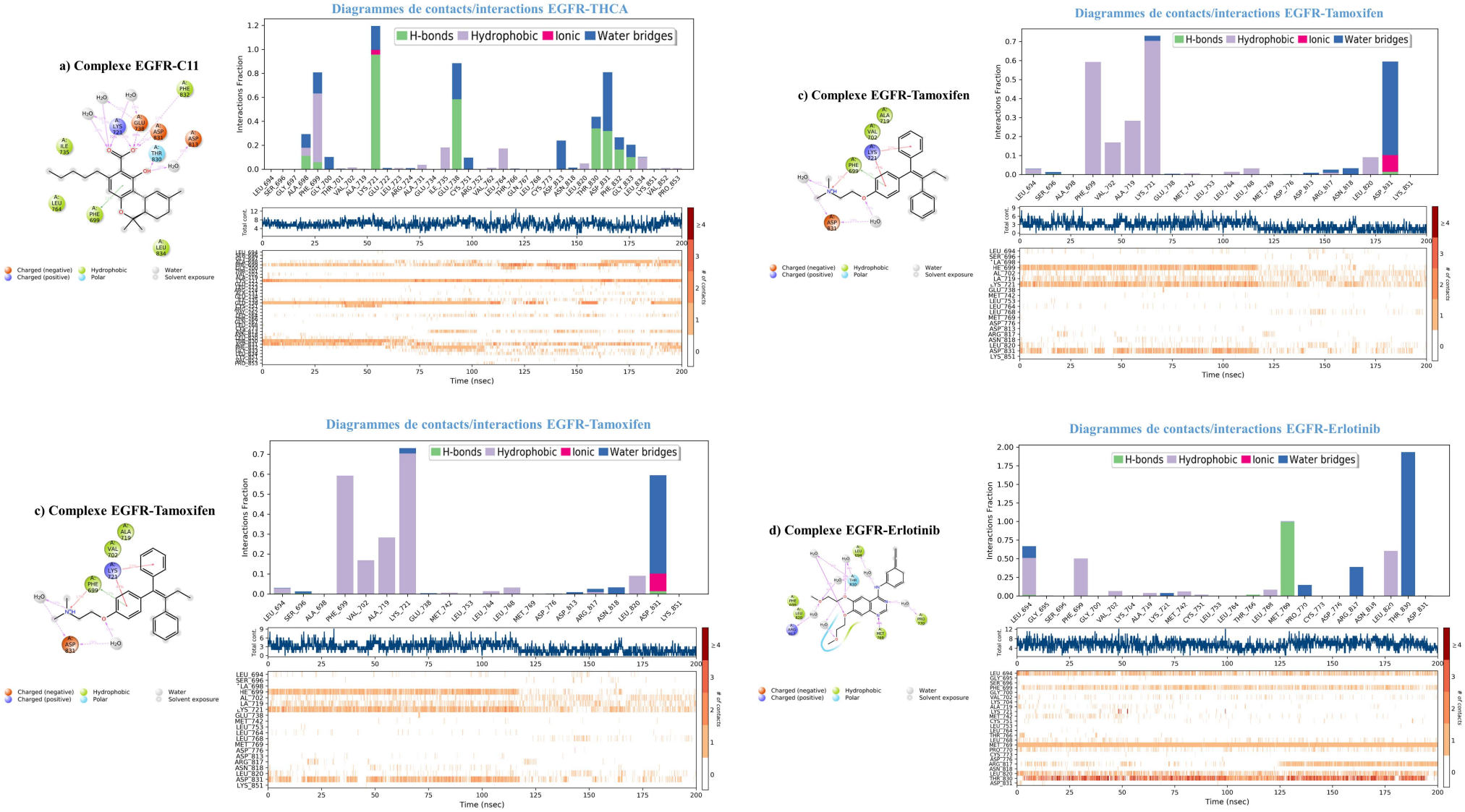


Figure IV. B10. Diagrammes détaillés de la distribution des interactions locales entre les ligands C11 (THCA), T15 (BCP), le Tamoxifen et l'Erlotinib dans la poche active de l'EGFR pendant 200ns de simulations de la DM.

3.5.4 *Investigation des influences des propriétés thermodynamiques*

Les valeurs moyennes des propriétés thermodynamiques de l'énergie totale (E), de l'énergie potentielle (EP), de la température (T), de la pression (P) et du volume (V) sur les trajectoires de dynamique moléculaire (MD) de 200 ns pour les systèmes d'EGFR-TKD, lesquels incluent l'EGFR libre, l'EGFR-C11, l'EGFR-T15, l'EGFR-Tamoxifen et l'EGFR-Erlotinib, sont reportées dans le **Tableau IV. B5**. Les résultats enregistrés indiquent que l'environnement ambiant dans lequel ils ont évolué pendant cette période était stable, caractérisé par des chiffres à peu près similaires entre tous les systèmes basés sur la structure de base de l'EGFR non complexé et ceux basés sur l'EGFR-TKIs complexé. Cela a également pu être confirmé par l'évolution approximativement uniforme des propriétés thermodynamiques de E, EP, T, P et V, illustrée à la **Figure IV. B12**.

Tableau IV. B5. Valeurs moyennes estimées pour les propriétés de l'énergie totale (E), de l'énergie potentielle (EP), de la température (T), de la pression (P) et du volume (V) pour les systèmes de protéines EGFR examinés.

	EGFR libre	EGFR-C11	EGFR-T15	EGFR-Tamoxifen	EGFR-Erlotnib
Énergie totale (kcal/mol)	-141829.047	-141737.555	-141819.536	-141574.711	-141737.842
Énergie potentielle (kcal/mol)	-172395.959	-172369.700	-172428.177	-172173.428	-172326.123
Température (K)	298.690	298.687	298.695	298.679	298.693
Pression (bar)	1.356	1.269	1.562	0.974	1.031
Volume ($Å^3$)	505443.948	506482.028	506029.727	505868.647	505560.214

Cette étude a identifié neuf composés présents dans *Cannabis sativa L* (CBDA, CBD, THCV, Δ-9-THC, Δ-8-THC, CBL, THCA, BCP et γ-Ele) comme des composants de base potentiels pour le développement d'inhibiteurs du récepteur du facteur de croissance épidermique-tyrosine kinase (EGFR-TKIs) destinés aux thérapies ciblées contre le cancer. En combinant des modèles de criblage virtuels, des approches de conception assistée par ordinateur de médicaments et des simulations biophysiques, cette recherche fournit des preuves que *Cannabis sativa L.* pourrait être utilisé pour développer des agents médicinaux pour traiter des maladies actuellement incurables, telles que le cancer. Pour faire avancer les composés identifiés le long du pipeline de conception de médicaments, des études et des évaluations préliminaires *in vivo* et *in vitro* devraient être menées. Les avancées en technologie analytique, en stratégies d'ingénierie et en analyse du génome de la plante n'ont fait qu'accroître le potentiel de *Cannabis sativa L* à des fins thérapeutiques, ce qui a conduit l'État marocain à récemment légaliser la production et l'exportation de (*Cannabis sativa L.)* à des fins médicales, industrielles et commerciales. Ce changement dans la région vers l'acceptation de (*Cannabis sativa L.)* à des fins médicinales pourrait avoir un impact majeur sur l'effort de développement et de mise à disposition de traitements innovants pour ceux qui en ont besoin.

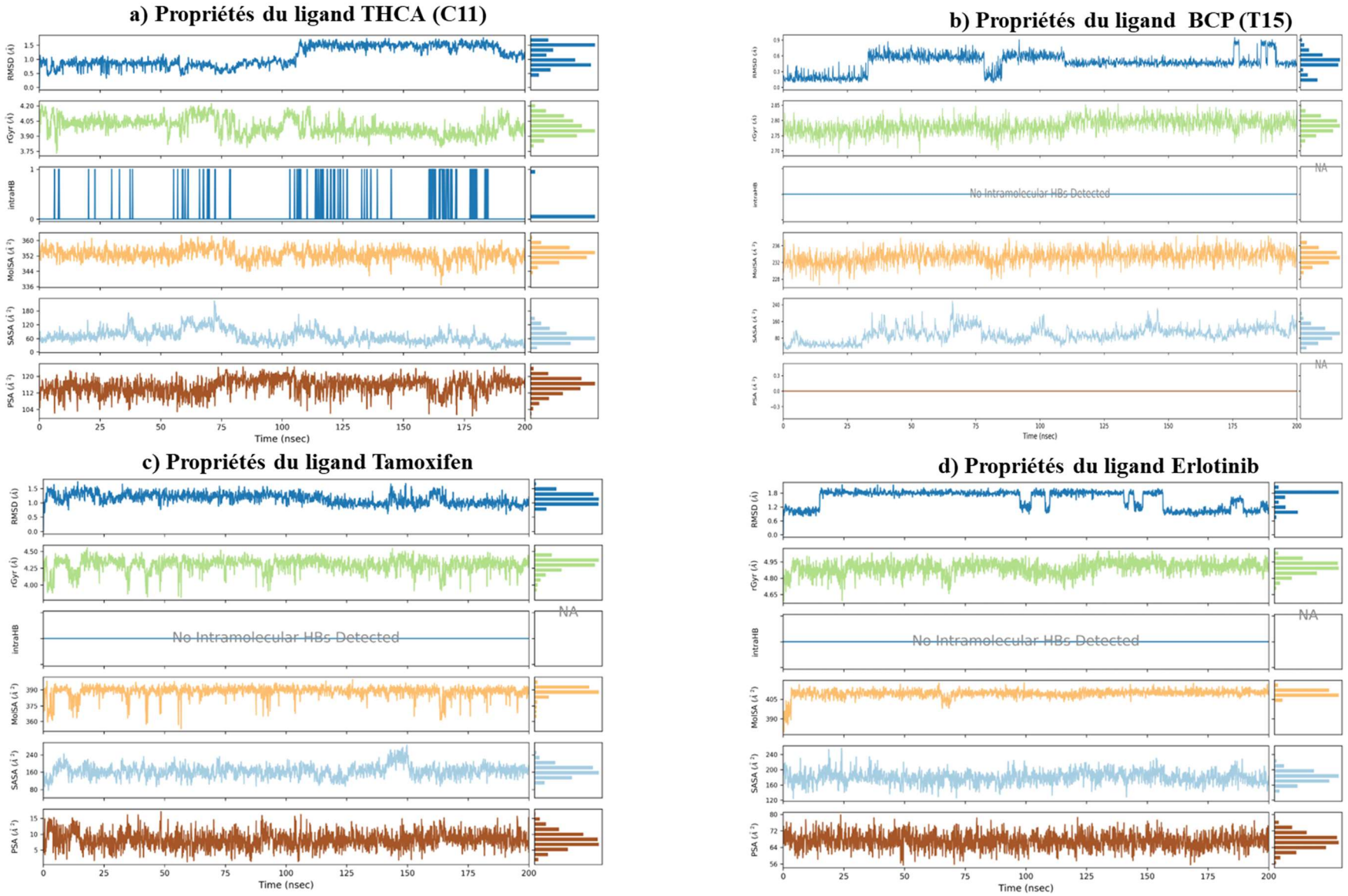


Figure IV. B11. Tracés des variations des propriétés des ligands tout au long de la trajectoire de la simulation de la dynamique moléculaire (0 à 200nsec).

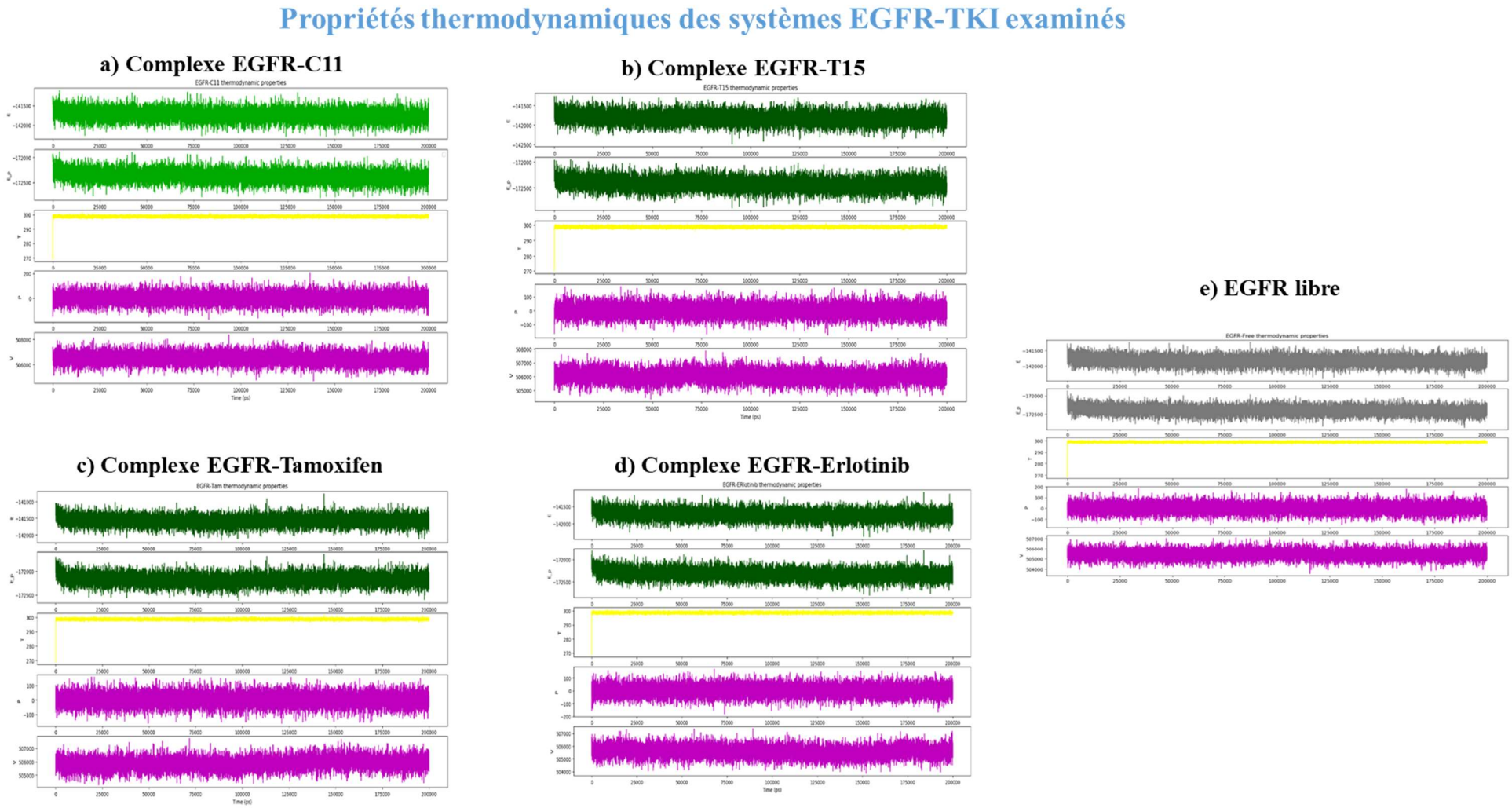


Figure IV. B12. Propriétés thermodynamiques en fonction du temps de l'énergie totale, de l'énergie potentielle, de la température, de la pression et du volume pour l'EGFR libre (a) et l'EGFR complexé avec les substances C11 (b), T15 (c), Tamoxifen (d) et Erlotinib (e).

4 Conclusion

Cette étude basée sur des approches hybrides de conception de médicaments assistée par ordinateur (CADD) est menée pour découvrir le potentiel des composés phytochimiques de *Cannabis sativa L.* comme agents thérapeutiques contre les cancers du sein et du poumon provoqués par l'activité enzymatique anormale de l'EGFR-TKD. Au total, cinquante substances phytochimiques à base de cannabinoïdes et de terpènes sont soumises à un criblage virtuel par Docking moléculaire semi-flexible et flexible, ainsi qu'à des simulations biophysiques de posk-Docking, telles que les calculs de Prime/MM-GBSA et la dynamique moléculaire. Les résultats de cette étude ont montré que certains échantillons spécifiques étaient mieux adaptés à l'inhibition de l'activité enzymatique de l'EGFR que les médicaments de référence tels que le Tamoxifen et l'Erlotinib. Les critères de similarité avec les médicaments et les paramètres ADME-Tox indiquent également que les substances candidates présentent des propriétés favorables, alors que les médicaments de référence présentent des risques de toxicité significatifs. La stabilité de la structure de la protéine EGFR cible et la parfaite adéquation entre la structure de cette dernière avec celle des substances phytochimiques examinées comme candidats nouveaux EGFR-TKI ont également confirmé par des simulations de la dynamique moléculaire. Par conséquent, cette étude explore que les substances CBDA (C1 : Acide cannabidiolique), CBD (C4 : Cannabidiol), THCV (C5 : Tétrahydrocannabivarine), Δ-9-THC (C7 : Dronabinol), Δ-8-THC (C8 : Delta-8-Tétrahydrocannabinol), CBL (C9 : Cannabicyclol), THCA (C11 : Acide delta9-tétrahydrocannabinolique), BCP (T15 : β-caryophyllène), et γ-Ele (T25 : γ-élemène) peuvent être adoptées comme facteurs structurels principaux pour la mise au point de nouveaux médicaments anticancéreux ayant une action spécifique en tant qu'EGFR-TKIs. En conclusion, ces petites molécules bioactives à base de cannabinoïdes isolés et de terpènes de cannabis peuvent être considérées comme une base qui mérite plus grande attention et plus de progrès de la part de l'industrie pharmaceutique en général et de la pharmacopée anti-cancéreuse en particulier.

CHAPITRE V : MODÉLISATION DE LIGANDS BIOACTIFS POUR LUTTER SPÉCIFIQUEMENT CONTRE LE LYMPHOME ET LES INFECTIONS VIRALES SARS-CoV-2 À TRAVERS LE CIBLAGE MOLÉCULAIRE RÉVERSIBLE DES PROTÉASES MALT1 ET 3CL^{PRO}

Introduction

Les petites molécules bioactives sont d'une grande importance dans le processus de conception rationnelle des médicaments car elles peuvent être utilisées pour comprendre le mécanisme de la maladie et développer des composés qui ciblent spécifiquement les récepteurs ou les enzymes liés à la pathologie. Ces molécules sont généralement de nature synthétique, ou produites par des plantes ou des organismes. Elles peuvent interagir avec des récepteurs biologiques, moduler ou réguler des processus physiologiques ou transporter des messages à l'intérieur des cellules. Ces molécules ont fait l'objet d'études approfondies pour leurs applications potentielles dans le développement de nouveaux médicaments en vue de lutter contre le cancer, les infections virales et d'autres types de pathologies. Par exemple, le développement de médicaments anticancéreux basés sur de petites molécules bioactives a connu d'énormes progrès au cours des dernières décennies. La majorité de ces molécules peuvent inhiber la croissance des cellules cancéreuses grâce à l'inhibition d'enzymes cibles, qui sont essentielles à la progression de la tumeur. De même, les petites molécules ciblant les protéines virales peuvent agir comme des médicaments antiviraux et contribuer à limiter l'infection par des virus ainsi que la progression de la maladie. En outre, les petites molécules peuvent être dirigées contre des organes et des tissus spécifiques, ce qui permet d'améliorer leur efficacité et de réduire les effets secondaires. Par exemple, le développement de médicaments anticancéreux basés sur de petites molécules bioactives a connu d'énormes progrès au cours des dernières décennies. La majorité de ces molécules peuvent inhiber la croissance des cellules cancéreuses grâce à l'inhibition d'enzymes cibles, qui sont essentielles à la progression de la tumeur. De même, les petites molécules ciblant les protéines virales peuvent agir comme des médicaments antiviraux et contribuer à limiter l'infection par des virus ainsi que la progression de la maladie. En outre, les petites molécules peuvent être dirigées contre des organes et des tissus spécifiques, ce qui permet d'améliorer leur efficacité et de réduire les effets secondaires. Dans ce contexte, les petites molécules bioactives jouent un rôle de plus en plus important dans le processus de conception rationnelle des nouveaux types de médicaments pour le combat contre le cancer et les infections virales. Le lymphome et le Covid-19 constituent un véritable fardeau pour la santé mondiale, avec environ 1.2 million de nouveaux cas de lymphome et 5 millions de cas de Covid-19 signalés en 2020 [498], [499]. Par conséquent, la recherche de nouvelles thérapies ciblées pour ces maladies est impérative. Ces évolutions permettront de réduire à la fois la morbidité et la mortalité associées à ces deux pathologies, ainsi que le bilan économique. En outre, les nouveaux traitements peuvent contribuer à la prévention des maladies et aux stratégies de contrôle, et

potentiellement conduire à d'autres découvertes dans le domaine des sciences médicales. La nécessité de mener des investigations et de mettre au point de nouveaux protocoles thérapeutiques pour lutter contre le lymphome et le Covid-19 est évidente, mais la complexité de ces maladies représente un véritable défi pour les scientifiques. Bien qu'il se trouve actuellement sur le marché plusieurs méthodes de traitement pour ces deux maladies, leur succès reste limité, d'où la nécessité de disposer toujours davantage de thérapies alternatives. La poursuite de la recherche de nouveaux traitements pour ces pathologies est essentielle pour augmenter les chances de succès et réduire la charge de morbidité associée à ces pathologies. Dans le cadre de notre contribution à cette proposition, nous visons à développer de nouvelles stratégies thérapeutiques pour ces pathologies incurables, basées sur de petites molécules synthétisées biologiquement actives.

A. Exploration, modélisation et optimisation de nouvelles petites molécules pyridiniques en tant que candidats inhibiteurs allostériques de la protéase MALT1 pour la thérapie ciblée du lymphome

1 Introduction

Parmi les inhibiteurs du MALT1 disponibles actuellement, on trouve le z-VRPR-fmk, le composé 3, le composé MI-2, β-lapachone, mepazine, MLT-747 (747, 748, 943, 985), la pyrazolopérimidine et la Thioridazine, qui ont tous des effets secondaires et des limites qui restreignent leur applicabilité clinique [145], [146]. Ceci est dû aux limitations structurelles et à la forte affinité de ces agents pour des récepteurs autres que MALT1, ainsi qu'au potentiel d'effets secondaires tels que des troubles de la motricité, des maladies inflammatoires de l'intestin, un rythme cardiaque irrégulier, etc [500], [501]. En raison des nombreuses limitations et des effets secondaires constatés au cours des essais cliniques pour de nombreux inhibiteurs sélectifs et non sélectifs du MALT1, il est urgent de poursuivre les recherches dans ce sens afin de surmonter les obstacles potentiels avant de commencer les essais cliniques chez l'homme. Plusieurs études ont indiqué que les inhibiteurs allostériques du MALT1 à base de petites molécules sont robustes et efficaces en phase préclinique [502]. Parmi les avantages de cette classe de inhibiteurs allostériques, citons les suivants :

- Amélioration de la sélectivité: Les inhibiteurs allostériques de MALT1 peuvent améliorer la sélectivité de la thérapie en s'attaquant spécifiquement aux cellules tumorales ou aux cellules auto-réactives impliquées dans les maladies auto-immunes, ce qui peut réduire les effets secondaires potentiels sur les cellules saines [502].
- Potentiel anticancéreux : MALT1 joue un rôle important dans la régulation de la voie de signalisation des récepteurs TNF et a été impliqué dans le développement de certains cancers, tels que le lymphome. Les inhibiteurs allostériques de MALT1 peuvent inhiber l'activité de MALT1 et bloquer la voie de signalisation TNF, ce qui pourrait potentiellement avoir un effet anticancéreux [161].
- Thérapie ciblée : Les inhibiteurs allostériques de MALT1 ciblent spécifiquement MALT1, ce qui signifie qu'ils ont un effet minimal sur d'autres protéines et fonctions cellulaires. Cela peut réduire les effets secondaires potentiels associés aux traitements actuels pour le cancer et les maladies auto-immunes [146].

À ce propos, plusieurs études ont montré que les petites molécules portant les échafaudages pyrazolo et pyrimidine étaient les plus abordées dans les essais précliniques comme inhibiteurs allostériques de MALT1 [502]. Cependant, en raison de la diminution rapide des doses de concentration approuvées, les agents à base de pyrazole et de pyrimidine peuvent entraîner de nombreux effets secondaires indésirables tels que des anomalies immunologiques et d'autres symptômes [146]. Par conséquent, il faut approfondir la recherche et les évaluations à cet égard afin de trouver une nouvelle classe de structures moléculaires capables de lutter contre les maladies résultant de la dérégulation de l'activité enzymatique de MALT1 et qui soient plus appropriées pour l'utilisation thérapeutique. À cet égard, une récente étude de criblage à haut débit menée par Tianbao Lu et al [503]. a conduit à la découverte d'une nouvelle série de 40 petites molécules (**Tableau V. A1**) portant un échafaudage pyridine puissant et hautement sélectif pour inhiber l'activité enzymatique de MALT1. Les résultats des essais cellulaires ont également montré la forte activité biologique de ces molécules contre la lignée cellulaire de lymphome à cellules B (TMD8). Les résultats de l'enzymologie mécaniste ont indiqué que ces composés se lient au site allostérique connu de la protéase MALT1. Ces données intéressantes nous ont incités à réfléchir à la mise en évidence et à l'amélioration des caractéristiques structurelles, des propriétés pharmacologiques et des mécanismes d'interaction de ces molécules avec la protéase MALT1, ce qui pourrait contribuer à favoriser la localisation de ces molécules dans le pipeline de conception et la découverte d'un nouveau médicament contre le lymphome. La **Figure V. A1** présente l'échafaudage de pyridine porté par la série de petites molécules proposées comme inhibiteurs allostériques de l'activité enzymatique excessive de la protéase MALT1. À cet égard, nous avons adopté le médicament Thioridazine (TH) comme référence pour conduire notre étude *in silico* sur ces nouvelles molécules.

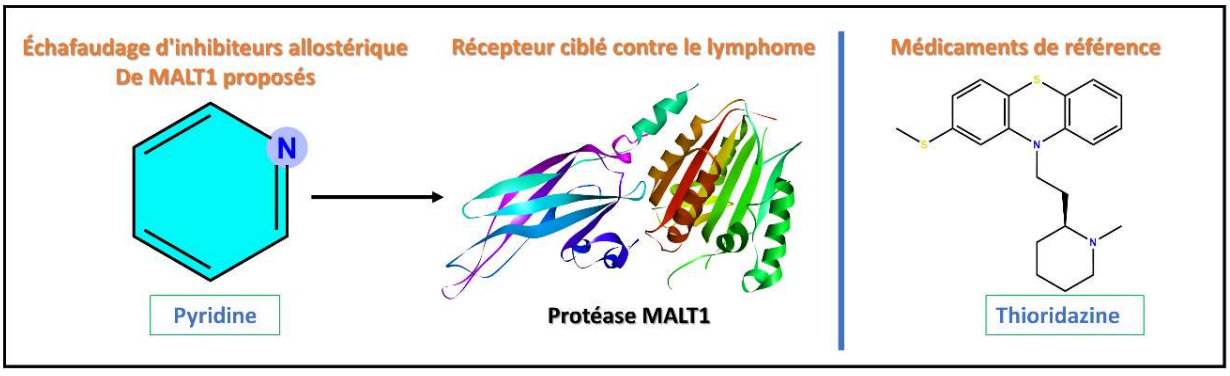


Figure V. A1. Conception de nouveaux inhibiteurs allostériques de MALT1 pour la thérapie ciblée contre le lymphome, utilisant un échafaudage moléculaire pyridinique.

Dans cette étude, la structure moléculaire et l'activité biologique inhibitrice d'une base de données de quarante petites molécules portant le motif de pyridine ont été examinées et optimisées *in vitro* pour leur effet inhibiteur (IC_{50}) observé vis-à-vis de l'activité enzymatique de MALT1. Ces données sont préparées par la conversion des valeurs observées de la IC_{50}(µM) au rang de pIC_{50} (pIC_{50} = -LogIC_{50}), suivie de la représentation et de l'optimisation géométrique et structurelle des 40 petites molécules, lesquelles sont regroupées dans le **Tableau V. A1**. Ces quarante petites molécules de dérivés de pyridine candidates comme inhibiteurs allostériques de la MALT1 ont été synthétisées par une série de réactions chimiques et de substitutions structurelles sur quatre références structurelles clés (**A**), (**B**),

(**C**) et (**D**), comme la montre la **Figure V. A2**. Grâce au **Tableau V. A1**, nous pouvons constater que les valeurs d'activité inhibitrice pIC_{50} contre MALT1 observées *in vitro* varient de 4.52 pour la conformation la moins active (composé B29) à 8.00 pour la conformation la plus active (composé D40). Cette variation significative dans la gamme structure/activité encourage la réalisation d'études approfondies de ces molécules en divers termes tels que la relation quantitative entre la structure et l'activité, la compatibilité structurelle et l'affinité de ce composé envers le biomarqueur macromoléculaire MALT1. Ceci dans le but principal de décoder et d'explorer les propriétés optimales liées aux structures de ces molécules favorables à atteindre un pIC_{50} optimal pour inhiber l'activité enzymatique anormale de MALT1 et de concevoir un nouveau médicament approprié pour lutter contre le lymphome.

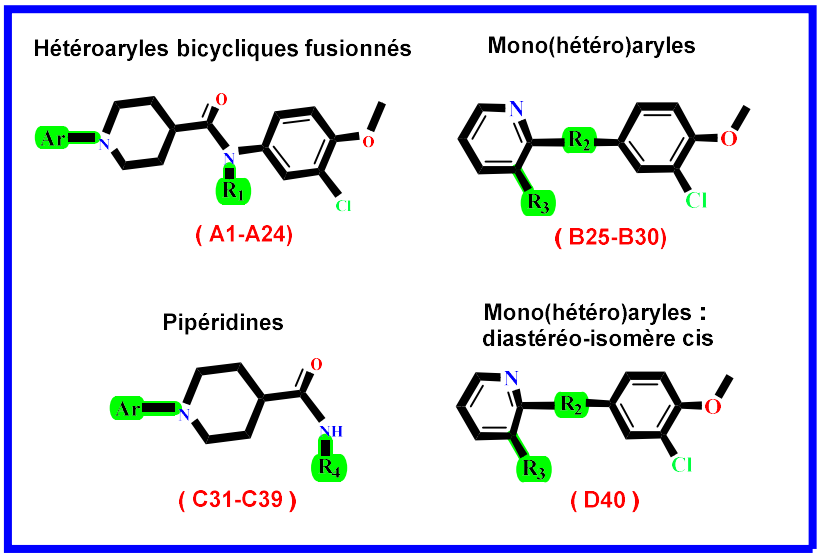


Figure V. A2. Les échafaudages à base de pyridine de 40 petites molécules étudiées.

2 Stratégie de conception de médicaments assistée par ordinateur développée et appliquée

Dans le cadre de notre intérêt porté à l'étude et à la valorisation de petites molécules bioactives à des fins pharmaceutiques, dans la présente étude, nous avons fait le point sur une base de données de 40 molécules optimisées et sélectionnées comme inhibiteurs allostériques de MALT1 par le biais d'une stratégie de criblage à haut débit ("high-throughput screening", HTS). En raison de l'association étroite entre l'activité enzymatique anormale de la protéase MALT1 et le développement de nombreuses pathologies et infections chroniques, dont le lymphome, etc. La découverte de nouveaux agents inhibiteurs de MALT1 dotés d'excellentes propriétés pharmacocinétiques et pharmacodynamiques pourrait contribuer à la naissance d'une stratégie thérapeutique prometteuse pour le lymphome. C'est ce que nous envisageons de commencer par mener la présente étude computationnelle *in silico* basée sur la stratégie de CADD hybride illustrée à la **Figure V. A3**. Concernant tous les détails relatifs aux méthodes et outils utilisés dans la mise en œuvre de chaque phase de cette étape, ils ont été présentés et détaillés dans le deuxième chapitre de cette thèse, ainsi que dans la version électronique de l'article que nous avons publié à ce sujet [217].

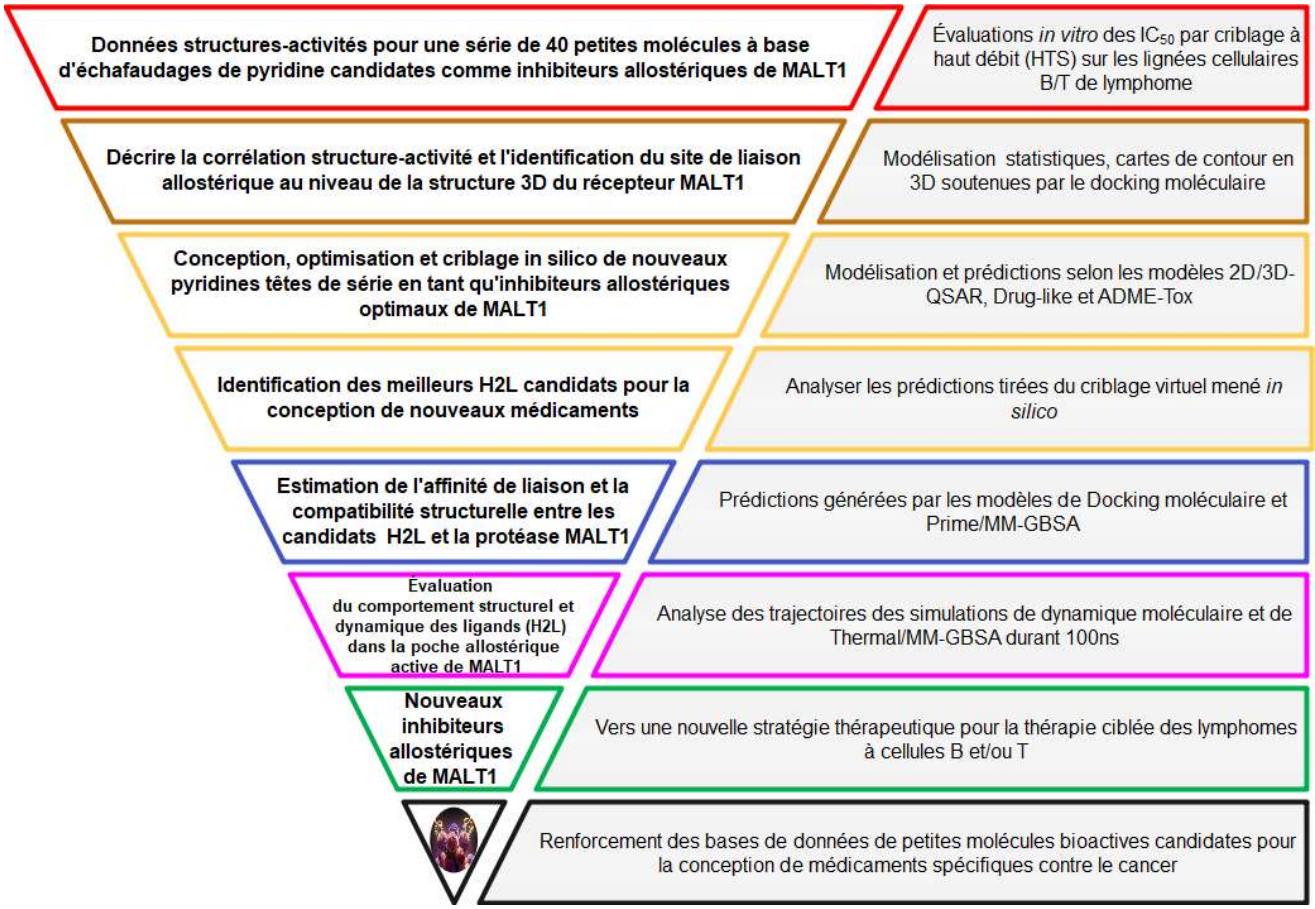
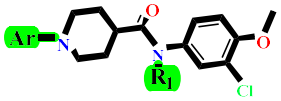


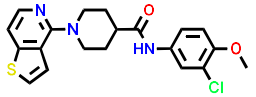
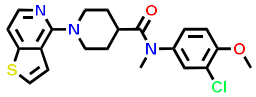
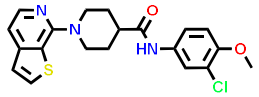
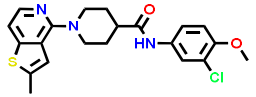
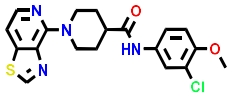
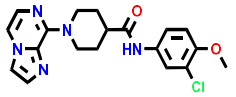
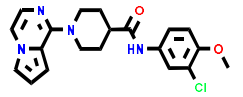
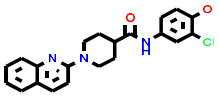
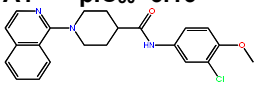
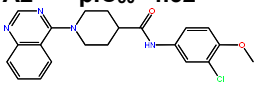
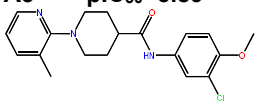
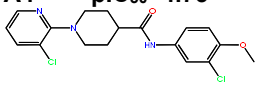
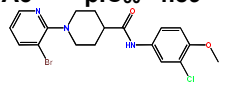
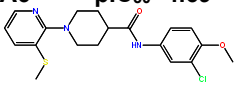
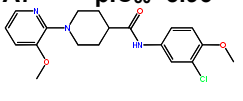
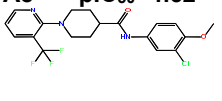
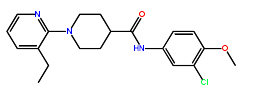
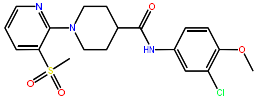
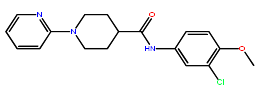
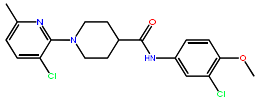
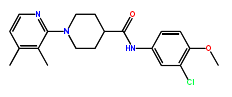
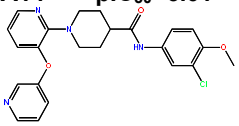
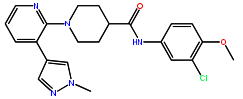
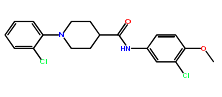
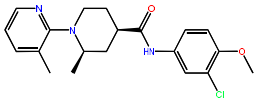
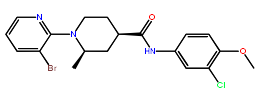
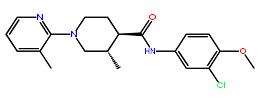
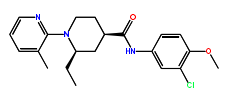
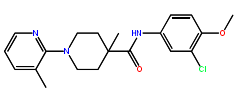
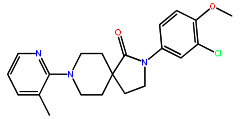
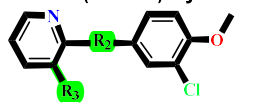
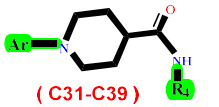
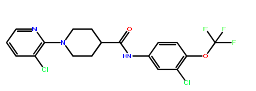
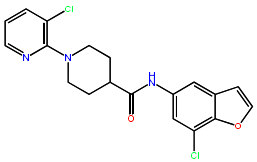
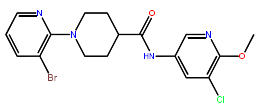
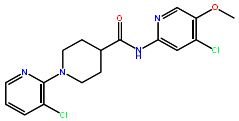
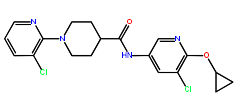
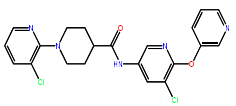
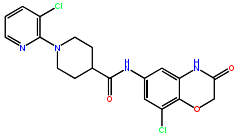
Figure V. A3. Stratégie de CADD hybride proposée et appliquée pour la conception, criblage, la caractérisation et l'optimisation de nouvelles inhibiteurs allostériques de la protéase MALT1 à base de Pyridine, candidates à la conception d'un nouveau médicament Anti-Lymphome.

Tableau V. A1. Informations sur la structure et l'activité obtenues *in vitro* pour les petites molécules à base de pyridine vis-à-vis de l'activité enzymatique de la MALT1.

Hétéroaryles bicycliques fusionnés



(A1-A24)

							
A1 pIC$_{50}$=6.16	**A2 pIC$_{50}$=4.52**	**A3 pIC$_{50}$=5.89**	**A4 pIC$_{50}$=4.75**	**A5 pIC$_{50}$=4.89**	**A6 pIC$_{50}$=4.69**	**A7 pIC$_{50}$=5.96**	**A8 pIC$_{50}$=4.52**
							
A9 pIC$_{50}$=5.64	**A10 pIC$_{50}$=5.40**	**A11 pIC$_{50}$=7.00**	**A12 pIC$_{50}$=7.00**	**A13 pIC$_{50}$=7.30**	**A14 pIC$_{50}$=6.01**	**A15 pIC$_{50}$=6.46**	**A16 pIC$_{50}$=6.39**
							
A17 pIC$_{50}$=6.08	**A18 pIC$_{50}$=5.22**	**A19 pIC$_{50}$=5.47**	**A20 pIC$_{50}$=6.26**	**A21 pIC$_{50}$=4.75**	**A22 pIC$_{50}$=5.37**	**A23 pIC$_{50}$=4.87**	**A24 pIC$_{50}$=6.64**
Mono(hétéro)aryles							Pipéridines
	**B25 pIC$_{50}$=7.30**	**B26 pIC$_{50}$=7.30**	**B27 pIC$_{50}$=6.64**	**B28 pIC$_{50}$=6.54**	**B29 pIC$_{50}$=4.52**	**B30 pIC$_{50}$=4.95**	
(B25-B30)							(C31-C39)
							
B31 pIC$_{50}$=5.76	**B32 pIC$_{50}$=7.00**	**B33 pIC$_{50}$=7.52**	**B34 pIC$_{50}$=6.89**	**B35 pIC$_{50}$=7.40**	**B36 pIC$_{50}$=6.89**	**B37 pIC$_{50}$=7.70**	**B38 pIC$_{50}$=6.00**
	Mono(hétéro)aryles : diastéréo-isomère cis					-	
B39 pIC$_{50}$=6.27	(D40)	**D40 pIC$_{50}$=8.00**	**Médicament de référence (Thioridazine) pIC$_{50}$=4.81**				

3 Résultats et discussion

Dans cette partie, nous présentons et discutons les résultats les plus importants que nous avons obtenus dans le cadre de cette étude. D'autres détails supplémentaires qui n'affectent pas le flux de données et l'analyse et la discussion de nos résultats peuvent être consultés dans la version originale de notre article publié dans ce contexte (Daoui et al. [217]).

3.1 Analyse structure-activité par modélisation 2D-QSAR

Dans cette phase, nous décodons la relation structure-activité de la base de données des molécules de pyridine examinées à l'aide d'un mélange diversifié de descripteurs multidimensionnels constitutionnels, topologiques, physico-chimiques et de descripteurs de réactivité basés sur la DFT (0D-3D). L'analyse en composantes principales (ACP) a révélé que l'énergie de l'orbitale moléculaire la plus occupée (E_{HOMO}), l'énergie de l'orbitale moléculaire la moins occupée (E_{LUMO}), le moment dipolaire (DM), l'énergie totale (E_T), la dureté chimique (η), la surface polaire (PSA), Le poids moléculaire (MW), le volume moléculaire (MV), l'indice de réfraction (n), la tension superficielle (ST), la densité (d), la polarisabilité (αe) et la lipophilie (LogP) représentent les descripteurs moléculaires les plus appropriés pour la modélisation 2D-QSAR de ce jeu de données. En appliquant l'algorithme des centroïdes k-means à la matrice résultante de 13 colonnes et 40 lignes (échantillons moléculaires), l'ensemble de données est partitionné en deux sous-ensembles, l'ensemble d'entraînement et l'ensemble de test. Les molécules A11, A17, A21, A22, B27, B30, C31, C36 sont classées dans l'ensemble de test et les molécules A1-A10, A12-A16, A18-A20, A23-B26, B28-B29, C32-C35, C37-D40 sont classées dans l'ensemble d'entraînement.

3.1.1 *Modèle 2D-QSAR linéaire*

En appliquant la technique de la régression linéaire multiple, nous avons pu découvrir l'équation mathématique linéaire (**L'Eq V.1**) qui quantifie la relation entre l'activité biologique inhibitrice des molécules (pIC_{50}) et sept descripteurs moléculaires (E_{HOMO}, DM, E_T, MV, n, αe, et Log P). En appliquant une technique de régression linéaire multiple, nous avons pu découvrir l'équation mathématique linéaire (équation 1) qui quantifie la relation entre l'activité biologique inhibitrice des molécules (pIC_{50}) et sept descripteurs moléculaires (E_{HOMO}, DM, E_T, MV, n, αe et Log P). Le modèle 2D-QSAR obtenu suggère que l'activité inhibitrice des molécules contenant l'échafaudage de pyridine envers l'activité enzymatique MALT1 peut être prédite/interprétée *in silico* par les sept descripteurs moléculaires hybrides.

$$pIC_{50} = -313.489 - 1.022 \times E_{HOMO} + 0.332 \times DM - 2.144 \times 10^{-5} \times E_T + 0.520 \times MV + 195.116 \times n - 3.857 \times \alpha e + 0.235 \times LogP \quad \text{(V.1)}$$

$$N = 32,\ R = 0.95,\ R^2 = 0.90,\ MSE = 0.02,\ F = 14.53$$
$$p\text{-}value < 0.0001,\ R^2_{cv} = 0.66,\ R^2_{test} = 0.85$$

Le modèle 2D-QSAR développé dans cette étude est très fiable, avec une valeur R^2 de 0.90, une statistique de test de Fisher élevée (F = 14.53), une faible erreur quadratique moyenne (MSE = 0.02)

et une valeur p <0.0001. Cela indique que le modèle sélectionné a une forte signification statistique et présente une corrélation quantitative forte et fiable entre les descripteurs moléculaires E_{HOMO}, MD, E_T, MV, n, αe et LogP et l'activité biologique inhibitrice (pIC_{50}) des molécules étudiées. La distribution uniforme des valeurs pIC_{50} observées *in vitro* par rapport aux valeurs prédites *in silico*, comme la montre la **Figure V. A4(a)**, confirme que ce modèle 2D-QSAR est capable de fournir une description fiable et précise de la relation entre la structure et l'activité des molécules bioactives dérivées de la pyridine étudiée, la **Figure V. A4(b)** illustrant les contributions respectives de chacun des sept descripteurs moléculaires.

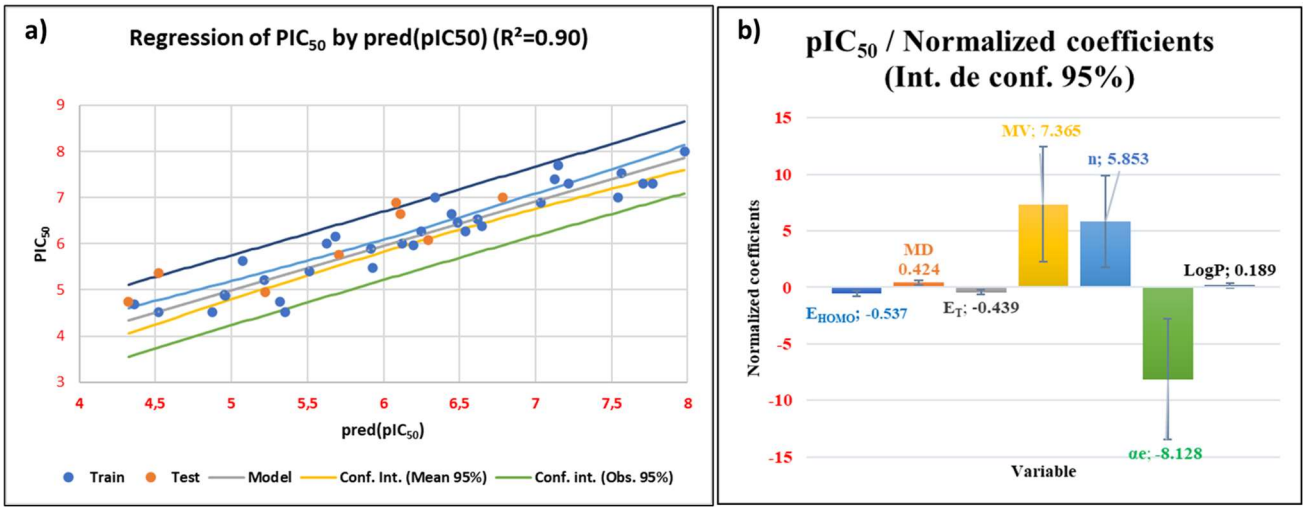


Figure V. A4. (a) Distribution des valeurs pIC_{50} observées *in viro* et prédites *in silico*. (b) Répartition des degrés des sept descripteurs moléculaires dans le modèle de RLM/2D-QSAR par les coefficients normalisés.

3.1.2 *Décryptage des corrélations entre les descripteurs moléculaires et la puissance d'inhibition*

Les résultats du diagramme de normalisation à la **Figure V. A4(b)** concernant les coefficients du modèle RLM /2D-QSAR proposé suggèrent que le coefficient du descripteur de polarisation (αe), l'indice de réfraction (n) et de volume moléculaire (MV) sont négatifs, ce qui indique que des valeurs plus élevées de ces descripteurs se traduisent par une activité biologique plus faible des molécules. L'inverse est vrai pour les descripteurs de lipophilie (LogP) et de moment dipolaire (MD), qui sont positifs et suggèrent que plus les valeurs de ces descripteurs sont élevées, plus l'activité des molécules vis-à-vis de la MALT1 est forte. L'énergie totale (E_T) et les coefficients d'énergie HOMO (E_{HOMO}) ont des valeurs négatives, ce qui indique que des valeurs d'énergie plus élevées provoquent une activité biologique plus faible des molécules. En utilisant les propriétés structurelles définies grâce au modèle RLM/2D-QSAR, de nouvelles molécules peuvent être conçues avec une meilleure activité biologique dans l'objectif d'inhiber l'activité enzymatique de la protéine MALT1.

3.1.3 *Évaluation de la validité interne et externe du modèle RLM/2D-QSAR*

La procédure de validation croisée "Leave-One-Out" (Loocv) appliquée dans cette étude a donné un coefficient de corrélation de $R^2_{cv} = 0.66$, ce qui indique que le modèle 2D-QSAR développé est robuste sous ce rapport. En outre, la validation externe basée sur l'ensemble de tests a donné $R^2_{pred} = 0.85$, ce qui confirme le fort pouvoir prédictif externe de ce modèle. Pour confirmer la stabilité du

modèle, le test de Y-randomisation est appliqué de manière répétée jusqu'à 50 fois (**Tableau V. A2**), ce qui révèle que les coefficients de corrélation (R^2 et R^2_{cv}) des modèles randomisés sont tous inférieurs à ceux du modèle 2D-QSAR original. Cela prouve que le modèle 2D-QSAR proposé est suffisamment stable et raisonnable pour prédire l'activité biologique inhibitrice (pIC_{50}) des petites molécules à base de pyridine, en raison de l'absence de valeurs aberrantes dans son domaine d'applicabilité (AD), comme le montre le diagramme de William (**Figure V. A5**). De ce fait, le modèle 2D-QSAR développé dans cette étude peut être appliqué en toute confiance comme support rationnel dans la conception et l'optimisation de nouveaux inhibiteurs allostériques de la MALT1.

Tableau V. A2. Test de Y-randomisation appliqué au modèle RLM/2D-QSAR.

Coefficient	Modèles 2D-QSAR randomisés	Modèle 2D-QSAR original
Nombre d'essais	50	-
Moyenne ***R***	0.47	0.95
Moyenne R^2	0.23	90
Moyenne R^2_{cv}	-0.47	0.66
$^{c}R_p^2$	0.78	

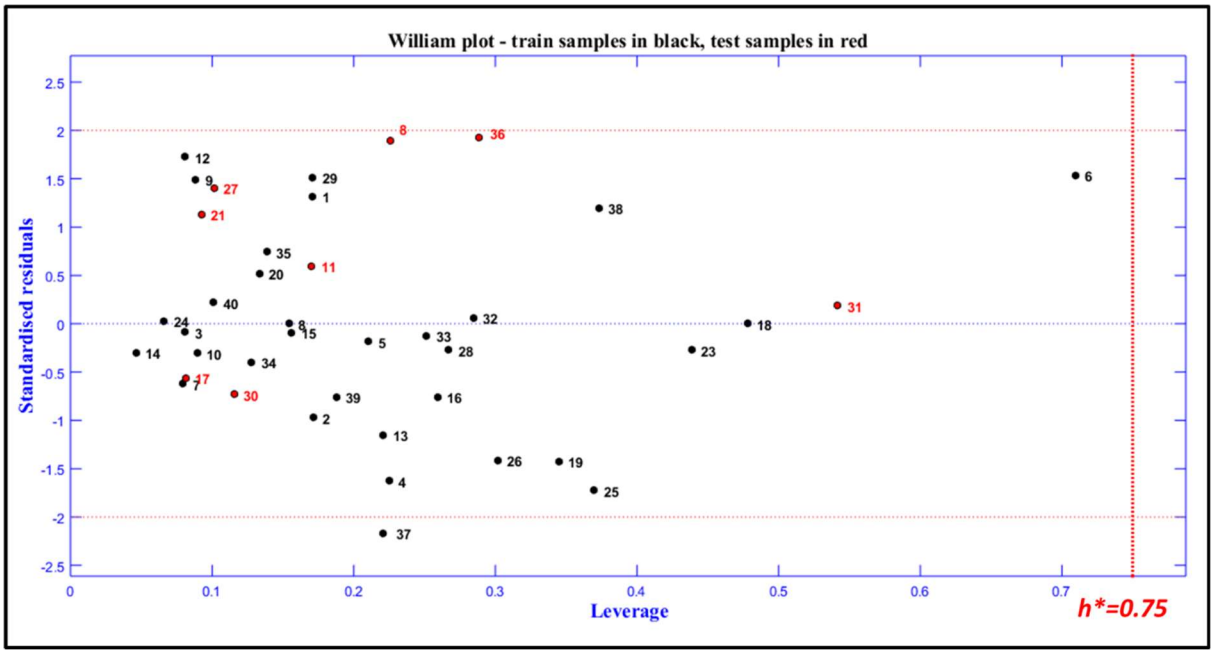


Figure V. A5. Domaine d'applicabilité du modèle RLM/2D-QSAR défini par le diagramme de Williams.

3.2 Étude des relations structure-activité par modélisation 3D-QSAR

Après avoir optimisé la structure géométrique 3D et le niveau énergétique des petites à base de pyridine, nous les alignons sur le ligand de référence bioactif (D40, pIC_{50} = 8,00) qui présente l'activité biologique inhibitrice la plus élevée observée *in vitro* vis-à-vis de l'activité enzymatique de MALT1, comme le reflète la **Figure V. A6**.

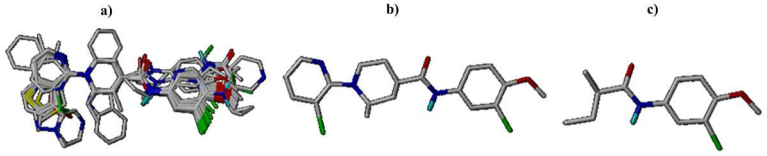


Figure V. A6. (a) Base de données alignée de petites molécules à base de pyridine, (b) conformation de référence du ligand bioactif D40, (c) sous- structure commune.

L'alignement des structures 3D des 40 petites molécules par rapport au ligand de référence D40 (**Figure V. A6a**) et la distillation du noyau commun de la sous-structure entre les 40 petites molécules (**Figure V. A6c**) ont été effectués avec succès. À partir de ces données moléculaires, il est possible de générer des descripteurs moléculaires de champ 4D (CoMFA et CoMSIA), qui peuvent ensuite être utilisés pour construire un modèle 3D-QSAR *in silico*. Ce modèle va ensuite apporter une meilleure connaissance sur les propriétés du champ qui peuvent influencer l'activité biologique et l'affinité des molécules vis-à-vis du récepteur MALT1. Grâce à la combinaison des résultats du modèle 3D-QSAR et de ceux du modèle 2D-QSAR, il est possible de générer des aperçus plus fidèles et des hypothèses pharmacophores plus concrètes.

3.2.1 *Évaluation statistique des modèles CoMFA et CoMSIA*

Le **Tableau V. A3** synthétise les données statistiques de l'analyse PLS obtenues pour le modèle 3D-QSAR optimal développé à partir des descripteurs moléculaires des champs de CoMFA/SE et de CoMSIA/SEHDA. Les résultats présentés dans le **Tableau V. A3** montrent que les modèles 3D-QSAR développés à l'aide des descripteurs de champ de CoMFA/SE et CoMSIA/SEHDA sont caractérisés par des valeurs élevées de R^2 (0.95 et 0.90), des valeurs F (55.016 et 37.834) et de faibles valeurs d'erreur standard d'estimation (SEE) (0.003 et 0.001), respectivement. En outre, les bons coefficients de validation interne $Q^2 > 0.5$ produits par les modèles CoMFA ($Q^2 = 0.672$) et CoMSIA ($Q^2 = 0.579$), même avec le nombre optimal de composants sélectionnés par la méthode de validation croisée leave-one-out, indiquent en outre que les deux modèles sont fiables. De plus, les valeurs de R^2_{pred} (CoMFA) = 0.83 et R^2_{pred} (CoMSIA) = 0.81 valident le pouvoir des deux modèles à prédire avec précision l'activité biologique des molécules de la base de données de test. Il est également noté que les valeurs élevées de R^2_{bs} (CoMFA) = 0.97 et R^2_{bs} (CoMSIA) = 0.92 combinées à des valeurs faibles de SEE_{bs} (CoMFA) = 0.031 et SEE_{bs} (CoMSIA) = 0.058 révèlent la nature stable et statistiquement acceptable du modèle CoMFA et du modèle 3D-QSAR (CoMFA et CoMSIA). La contribution des champs stériques (S) et électrostatiques (E) dans le modèle CoMFA est de 0.450 et 0.550, respectivement, ce qui implique une forte corrélation entre ses propriétés et le pIC_{50} des molécules étudiées vis-à-vis de l'activité inhibitrice envers MALT1. Pour le modèle CoMSIA, les contributions des champs stériques (S), électrostatiques (E), hydrophobiques (H), donneurs de liaisons hydrogène (D) et accepteurs de liaisons hydrogène (A) sont de 0.057, 0.212, 0.250, 0.155 et 0.326, respectivement. Cela indique à nouveau l'importance des contributions de ces différents champs moléculaires pour expliquer l'effet biologique spécifique de ces molécules. Ensuite, pour valider les performances du modèle 3D-QSAR, on a procédé à un test de Y-randomisation. Le **Tableau V. A4** présente les résultats de cinq essais randomisés indépendants réalisés dans le cadre de cette étude. Les paramètres statistiques mentionnés dans le **Tableau V. A4** indiquent que les valeurs de Q^2_r, R^2_r pour les modèles randomisés étaient inférieures à celles des modèles 3D-QSAR original et que la moyenne des $^cR^2_p$ pour les différents modèles dépassait 0.5. Cela montre la stabilité des relations structure-activité explorées par les modèles CoMFA et CoMSIA, ce qui justifie leur application dans le cadre de la découverte de médicaments à base de ligands. La production des cartes de contour des champs S et E, et H, D et A respectivement, pour les deux modèles, nous a permis de détecter avec succès les sites pharmacophores et les propriétés favorables et défavorables

qui peuvent être exploitées pour optimiser l'activité / l'affinité des conformations moléculaires envers le récepteur MALT1.

Tableau V. A3. Bilan statistique des analyses PLS de CoMFA et CoMSIA.

Modèle 3D-QSAR	Paramètres statistiques								Fractions de champs				
	Q^2	R^2	SEE	F-test	N	R^2_{pred}	R^2_{bs}	SEE_{bs}	S	E	H	D	A
CoMFA/SE	0.615	0.95	0.003	55.016	4	0.83	0.97	0.116	0.450	0.550	-	-	-
CoMSIA/SEHDA	0.52	0.90	0.001	37.834	4	0.81	0.92	0.168	0.057	0.212	0.250	0.155	0.326

Q^2 est le coefficient de corrélation de validation croisée (loocv), **R^2** est le coefficient de corrélation de validation non croisée, **SEE** est l'erreur standard d'estimation, **F** est la valeur du F-test, **N** est le nombre optimal de composantes défini par PLS-loocv, **R^2_{pred}** est le coefficient de corrélation de prédiction obtenu par validation externe, **R^2_{bs}** est le coefficient de corrélation après 100 cycles de bootstrap, **SEE_{bs}** est l'erreur standard d'estimation pour 100 cycles de bootstrap, **Fractions** est la contribution des différents champs stérique (S), électrostatique (E), hydrophobe (H), donneur de liaison hydrogène (D), et accepteur de liaison hydrogène (A).

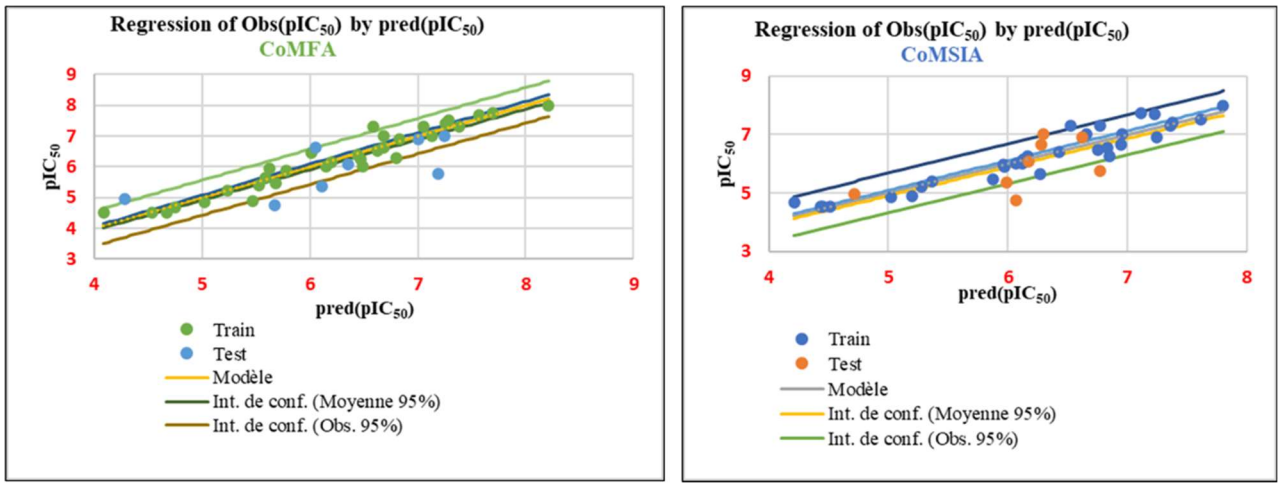


Figure V. A7. Distribution des valeurs pIC_{50} observées *in vitro* et prédites par le modèle 3D-QSAR *in silico* qui est basé sur les descripteurs des champs moléculaires de CoMFA et CoMSIA.

Tableau V. A4. Test de randomisation Y réalisé sur le modèle 3D-QSAR (CoMFA et CoMSIA).

Modèle 3D-QSAR	CoMFA/SE			CoMSIA/SEHDA		
Original	Q^2	R^2	-	Q^2	R^2	-
#1	**0.615**	**0.950**	**-**	**0.520**	**0.900**	**-**
Aléatoire	Q^2_r	R^2_r	$^cR^2_p$	Q^2_r	R^2_r	$^cR^2_p$
I1	0.427	0.792	0.573	0.389	0.812	0.603
I2	0.382	0.812	0.599	0.276	0.764	0.531
I3	-0.289	0.737	0.508	0.247	0.812	0.603
I4	-0.214	0.840	0.638	0.214	0.748	0.510
I5	-0.312	0.741	0.513	-0.512	0.747	0.509
Moyenne (I=5)	-0.0012	0.7844	0.5662	0.1228	0.7766	0.5512

3.2.2 *Corrélation des cartes de contour avec les données d'activité biologique*

L'analyse des cartes de contours générées par les modèles CoMFA (**Figure V. A8**) et CoMSIA (**Figure V. A9**) a facilité l'identification des caractéristiques structurelles dans l'espace 3D qui sont favorables (80%) ou défavorables (20%) à l'amélioration de l'activité biologique inhibitrice des dérivés pyridiniques étudiés contre la protéase MALT1. Dans cette partie de l'étude, la molécule D40 est utilisée comme référence pour présenter les cartes de contours analysées et déterminer les sites potentiels pour les différentes substitutions.

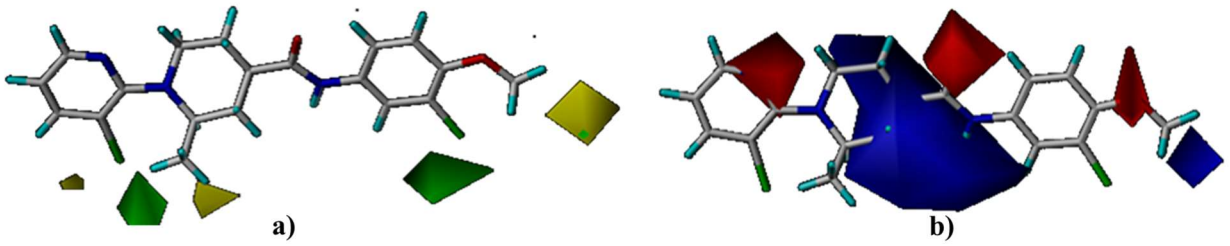


Figure V. A8. Cartes des contours stériques (a) et électrostatiques (b) générées par le modèle 3D-QSRA/CoMFA.

Les cartes de contours de CoMFA présentées dans la **Figures I.8(a,b)** indiquent que : (1) les contours jaunes situés près du méthyle (**R**) dans les groupes 1-chloro-2-méthoxybenzène et (2R, 4R)-2-méthylpipéridine-4-carboxamide (**R1**) suggèrent un site favorable pour l'ajout de petits groupes afin d'augmenter le pIC_{50} des composés étudiés ; (2) les contours verts près de l'atome de chlore (**R2**) du groupe 1-chloro-2-méthoxybenzène et de l'atome de brome du groupe 3-bromopyridine suggèrent un environnement favorable à l'ajout de groupements volumineux (**R3**), augmentant ainsi le pIC_{50} des molécules de dérivés pyridiniques contre MALT1 ; et, (3) l'abondance des contours électrostatiques bleus près des sites **R** et **R1** implique que l'ajout de groupements chargés positivement dans ces sites est bénéfique pour augmenter l'activité biologique pIC_{50} des molécules de dérivés pyridiniques. Les données expérimentales confirmant les conclusions tirées ont suivi, la substitution du méthyle par le trifluorométhoxy dans la structure de la molécule C31 entraînant une diminution de l'activité (pIC_{50} = 5.76), de même que la substitution du méthyle par l'éthyle dans la structure de la molécule D40 (pIC_{50} = 8.00) entraînant une diminution de l'activité biologique de la molécule B29 (pIC_{50} = 6.54). En outre, la substitution du méthyle par le brome dans la structure de la molécule B27 (pIC_{50} = 6.64) a été considérée comme augmentant l'activité biologique de la molécule, comme indiqué dans la molécule D40 (pIC_{50} = 8.0).

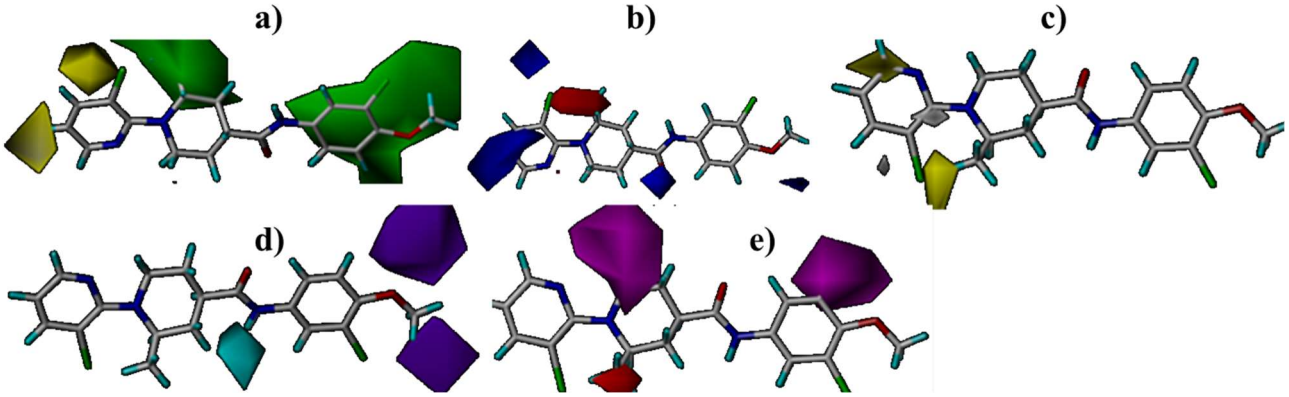


Figure V. A9. (a) stérique, (b) électrostatique, (c) hydrophobique, (d) donneur de liaison hydrogène, (e) accepteur de liaison hydrogène : cartes des contours fournies par le modèle 3D-QSAR/CoMSIA.

Les cartes de contours de la CoMSIA/SE dans la Figure 5(a-e) démontrent que les radicaux volumineux, les groupes donneurs d'électrons (+I), les groupes hydrophobes, les groupes donneurs de liaisons hydrogène et les groupes accepteurs de liaisons hydrogène peuvent améliorer la bio-activité inhibitrice des petites molécules de dérivés de pyridine étudiées contre l'activité enzymatique MALT1, la bio-activité la plus élevée étant observée lorsque ces groupements sont situés sur les sites R, R1, et R2. Le modèle 3D-QSAR/CoMFA s'est révélé plus précis que le modèle 3D-QSAR/CoMSIA pour

représenter ces champs, le modèle CoMFA/SE ayant une plus grande capacité prédictive que le modèle CoMSIA/SE. Les groupes hydrophobes sont considérés comme les plus favorables au site R2 pour augmenter la bioactivité des composés étudiés. En outre, nous avons constaté que le site R convient aux groupements accepteurs de liaisons hydrogène pour renforcer la liaison des composés étudiés à la poche allostérique active de la protéase MALT1, tandis que le site R1 convient mieux aux groupements donneurs de liaisons hydrogène.

3.3 Identification des propriétés 2D et 3D essentielles pour l'activité optimale des dérivés pyridiniques

La **Figure V. A10** présente un résumé illustratif des caractérisations structurelles favorables pour optimiser l'activité biologique inhibitrice des petites molécules de pyridine contre la poche allostérique active de la MALT1. Toutes ces descriptions sont tirées à partir des analyses des modèles 2D/3D-QSAR élaborés dans le cadre de la présente étude. Après avoir généré les propriétés structurelles et les caractéristiques de base des dérivés de la pyridine étudiés sur la base des conformations des ligands actifs, nous passons ensuite à la phase d'étude des aspects liés à l'affinité et à la compatibilité de ces molécules en tant qu'inhibiteurs allostériques spécifiques de la protéine MALT1. Ceci peut se faire par l'application de la stratégie de Docking moléculaire semi-flexible qui génère les profils de compatibilité/affinité associés aux interactions protéine-ligand.

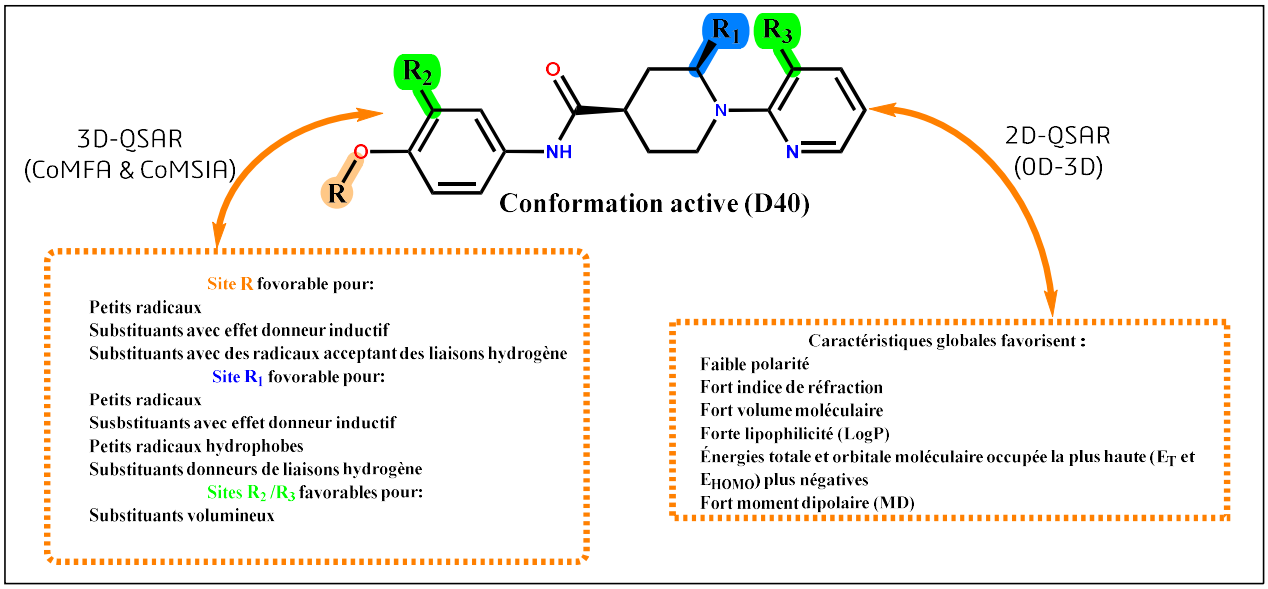


Figure V. A10. Propriétés structurelles optimales favorables à la génération de nouvelles petites molécules à base de pyridine ayant une activité inhibitrice et une affinité vis-à-vis de la poche allostérique active de la MALT1.

3.4 Identification des modes de liaison Protéine-ligand et des interactions clés par Docking Moléculaire

Avant de procéder à la modélisation des nouveaux inhibiteurs allostériques spécifiques de MALT1 susceptibles d'interagir avec la protéase MALT1 (code PDB : 4I1R), il est essentiel d'identifier le site actif allostérique de cette protéase. A cette fin, la complexité de la structure du composé de référence D40 et de l'inhibiteur Thioridazine (médicament de référence) dans la poche active de MALT1 est simulée à l'aide des modèles d'interaction ligand-protéine prédits. Avant de procéder à l'analyse et de

dégager des hypothèses, nous vérifions la validité et la performance du modèle de Docking moléculaire mis en œuvre dans ce contexte.

3.4.1 *Évaluation de la robustesse du modèle de Docking moléculaire par re-Docking*

Nous avons utilisé des simulations de re-Docking moléculaire pour confirmer les hypothèses générées par les modèles 2D/3D-QSAR et identifier les interactions actives entre les sites allostériques de la protéine MALT1 et les petites molécules à base de pyridine. Pour ce faire, nous utilisons le modèle 3D de la protéine MALT1 paracaspase complexée avec un inhibiteur Thioridazine. Après avoir préparé/optimisé soigneusement les modèles des structures de la protéine et des ligands, nous avons déterminé les coordonnées de la poche active (dimension XYZ de 20×20×20 $Å^3$; x = -28.84 Å, y = -14.97 Å, z = 20.81 Å ; 0.375 Å d'espacement entre les points). Ensuite, à l'aide du logiciel Autodock Vina, nous avons effectué des simulations de Docking semi-flexible multiples entre le récepteur MALT1 et les ligands bioactifs de la Thioridazine et du D40. Les résultats obtenus sont évalués sur la base du score d'énergie de liaison (EL)/constante inhibitrice estimée (K_i) et de l'analyse visuelle des interactions ligand-protéine. Enfin, la validation du processus repose sur la comparaison des résultats du modèle de Docking avec les informations disponibles sur le plan expérimental. La **Figure V. A11** montre l'alignement de l'inhibiteur allostérique Thioridazine complexé avec le MALT1 natif dans le complexe de référence original 4I1R avec son homologue re-Docké via le modèle de Docking moléculaire. L'analyse visuelle montre les positions conformationnelles parfaites du ligand Thioridazine natif (vert) et re-Docké (rouge) dans la poche active de MALT1. Cela explique la précision et la validité du modèle de Docking moléculaire mis en œuvre dans cette étude. Ceci est également confirmé par la faible valeur DockRMSD (0.103 < 2 Å) qui estime la distance atomique entre les deux échantillons de ligands (original/modélisé).

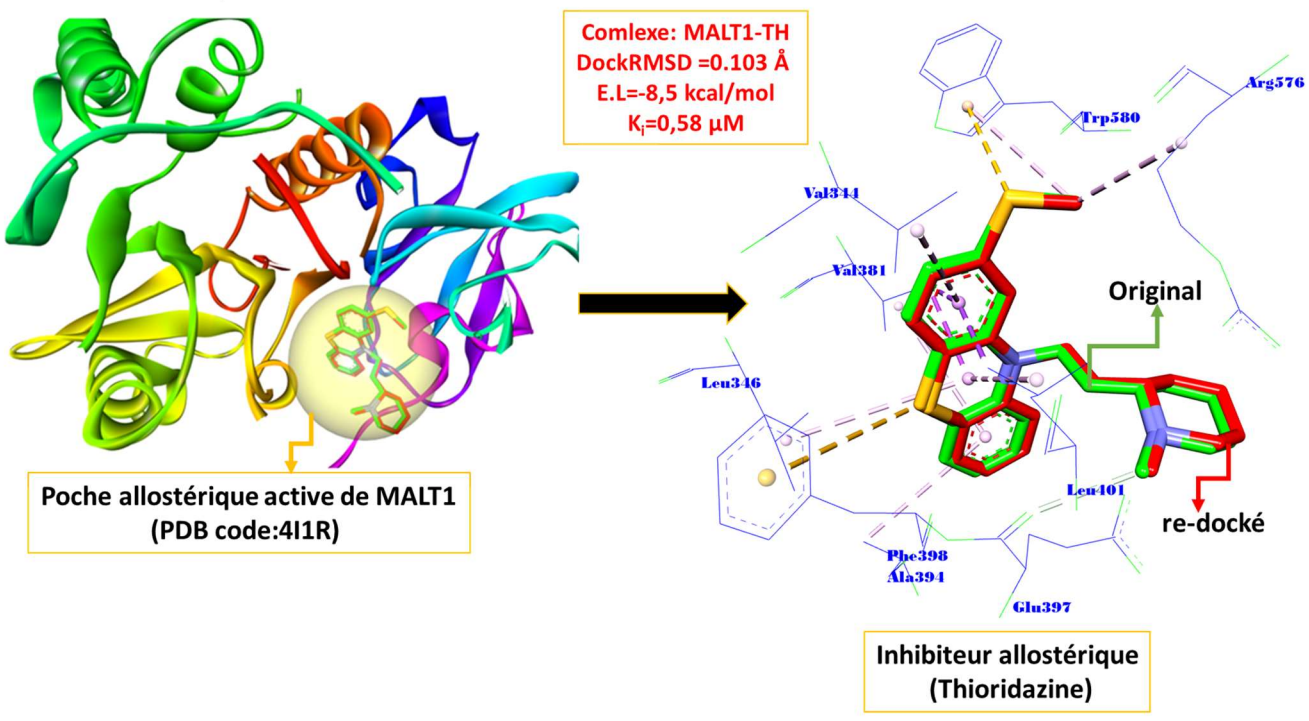


Figure V. A11. Positions des conformations de l'inhibiteur allostérique de référence Thioridazine (TH) (original/vert, re-docké/rouge) dans la poche allostérique active de MALT1.

Également, la visualisation 3D des interactions de la Thioridazine dans la poche active de MALT1 indique la présence des résidus d'acides aminés de référence suivants : **Leucine** (Leu346, Leu401), **Valine** (Val344, Val381), **Tryptophane** (Trp580), **Arginine** (Arg376), **Phénylalanine** (Phe398), **Alanine** (Ala394) et **l'acide glutamique** (Glu397). Sur cette base, ces résidus sont considérés comme un site actif allostérique susceptible d'être ciblé par de petites molécules (inhibiteurs spécifiques) et d'inhiber ainsi l'activité enzymatique anormale de la protéase MALT1. Les interactions non covalentes générées entre l'inhibiteur de Thioridazine et MALT1 étaient principalement de type carbone-hydrogène, électrostatiques et hydrophobiques.

3.4.2 Évaluer l'affinité des dérivés de pyridine avec la poche allostérique de MALT1

Après avoir validé la pertinence du protocole de Docking moléculaire pour identifier et analyser l'interaction entre la Thioridazine et le site allostérique actif de MALT1, nous avons procédé au Docking de la petite molécule D40 sur la poche de MALT1. Grâce à ce processus de Docking, nous avons pu identifier les interactions non covalentes potentielles entre les petites molécules étudiées et les sites allostériques actifs dans la poche du MALT1.

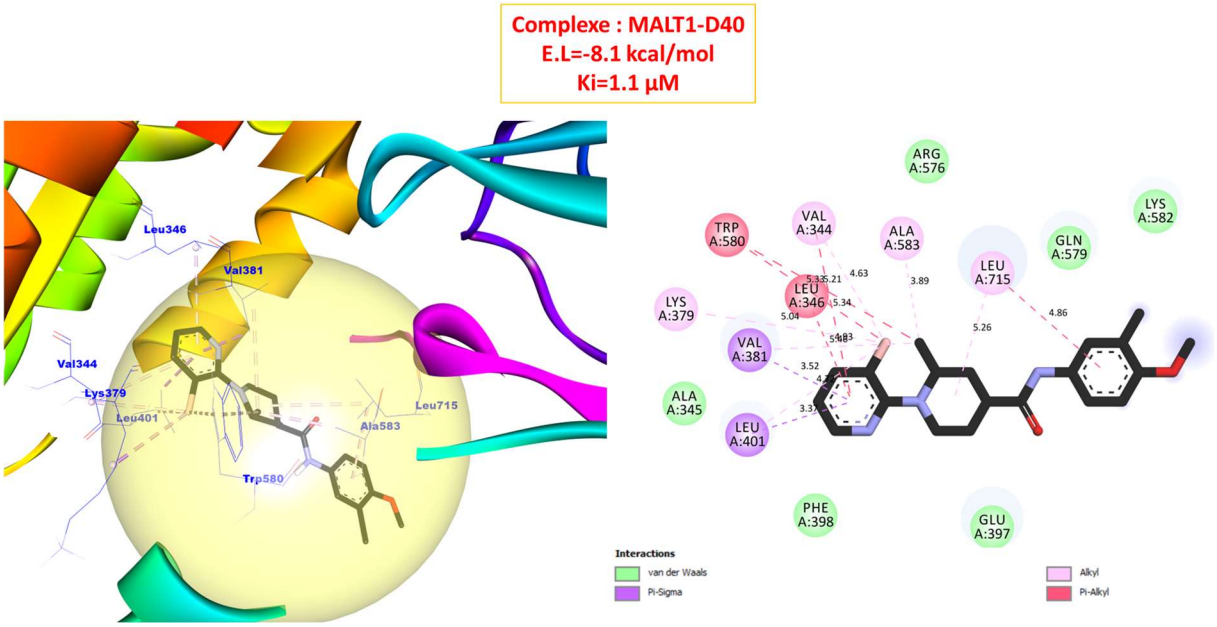


Figure V. A12. Pose optimale de la conformation de la petite molécule à base de pyridine (D40) au sein du complexe MALT1-D40.

La **Figure V. A12** montre la conformation optimale obtenue par Docking du ligand D40 dans la poche allostérique active de MALT1, avec une énergie de liaison de -8.1 kcal/mol et une constante d'inhibition (K_i) de 1.1 µM. Les énergies de liaison et les constantes d'inhibition estimées pour MALT1-D40 (-8.1 kcal/mol, Ki=1.1 µM) et MALT1-TH (-8.5 kcal/mol, K_i=0.58 µM) sont très proches, ce qui suggère qu'ils sont très efficaces pour inhiber l'activité enzymatique de MALT1. Cette découverte renforce l'idée que les petites molécules à base de pyridine peuvent être intéressantes pour le développement d'une nouvelle génération d'inhibiteurs allostériques spécifiques de MALT1, offrant ainsi des avantages thérapeutiques potentiels dans le traitement des lymphomes. Le **Tableau V. A5**

présente le bilan détaillé du profil des interactions protéine-ligand générées par le modèle de Docking moléculaire au sein des complexes MALT1-D40 et MALT1-TH.

Tableau V. A5. Profil de compatibilité/affinité protéine-ligand dans les complexes MALT1-D40 et MALT1-TH.

Complexe	Score d'affinité Protéine-ligand		Interactions non-covalentes Protéine-ligand			
	Énergie de liaison (E.L)	Constante d'inhibition prédite *in silico* (K_i)	Liaisons hydrogène	Électrostatiques	Hydrophobiques	Van der Waals
MALT1-TH	-8.5 kcal/mol	0.58 µM	Glu397 (3.32 Å)	Trp580(4.78 Å), Glu397 (4.46 Å)	Ala394(5.41 Å), Val381(5.14 Å), Val381 (3.45 Å), Val381 (3.66 Å) Leu401 (3.41 Å), Val344(5.23 Å), Arg576 (4.22 Å).	Phe398
MALT1-D40	-8.1 kcal/mol	1.1 µM	-	-	Leu401(3.37 Å), Val381 (3.52 Å), Lys379 (5.04 Å), Trp580 (5.33 Å), Leu346 (5.46), Val344 (4.63 Å), Ala583 (3.89 Å), Leu715 (4.86 Å), Leu715 (5.26 Å)	Ala345, Arg576, Gln579, Lys582, Phe398 and Glu397

D'après le **Tableau V. A5**, on peut voir que le petit ligand à base de pyridine (D40) se lie à sept sites de référence allostérique de Thioridazine, à savoir Leu401 (3.37 Å), Val381 (3.52 Å), Trp580 (5.33 Å), Val344 (4.63 Å), Ala345, Arg576, Gln579, Lys582, Phe398 et Glu397, par des interactions hydrophobes et des interactions de van der Waals. L'énergie de liaison de D40 dans le complexe 4I1R-D40 peut être comprise à travers l'activité inhibitrice élevée de cette molécule dans des essais *in vitro* (pIC_{50} = 8.00) et *in silico* tests RLM-QSAR, 3D-QSAR/CoMFA et 3D-QSAR/CoMSIA (pIC_{50} = 7.98, 8.21 et 7.80 respectivement). Afin d'optimiser la structure de D40 afin de le convertir en un médicament efficace pour l'inhibition de l'activité enzymatique de MALT1, il est important de créer des interactions non covalentes faibles entre D40 et les acides aminés au niveau de la poche active allostérique de MALT1, à savoir Alanine (Ala345, Ala583), Arginine (Arg576), Glutamine (Gln579), Lysine (Lys582, Lys379), Phénylalanine (Phe398), Acide glutamique (Glu397), Leucine (Leu401, Leu346, Leu715), Valine (Val381, Val344), et Tryptophane (Trp580). En agissant ainsi, des dérivés de D40 qui conviennent comme nouveaux médicaments pour l'inhibition de MALT1 peuvent être conçus.

3.5 Conception et optimisation d'inhibiteurs allostériques de MALT1

La caractérisation globale obtenue par la modélisation 2D/3D-QSAR et le Docking moléculaire confirme que la conformation bioactive du petit ligand D40 est tout à fait adaptée à la conception d'inhibiteurs allostériques puissants de MALT1 et qu'elle est également structurellement compatible avec la structure de la protéase MALT1. Pour renforcer cet aspect, nous proposons de visualiser en 3D les cavités de surface du complexe MALT1-D40 et de les comparer aux analyses des cartes de contours de CoMFA et CoMSIA. Il s'agit simplement d'établir une base rationnelle pour la conception de nouvelles molécules et leur criblage *in silico* sur la base de la méthodologie CADD hybride mise en œuvre dans cette étude. La **Figure V. A13(a-f)** illustre l'affinité du ligand D40 pour le récepteur de la

protéase MALT1 en termes de propriétés d'interactions favorables telles que l'affinité bord/face du radical aromatique, la surface d'accessibilité au solvant (SASA), la liaison hydrogène, la charge électrostatique, la susceptibilité à l'ionisation et les interactions hydrophobes.

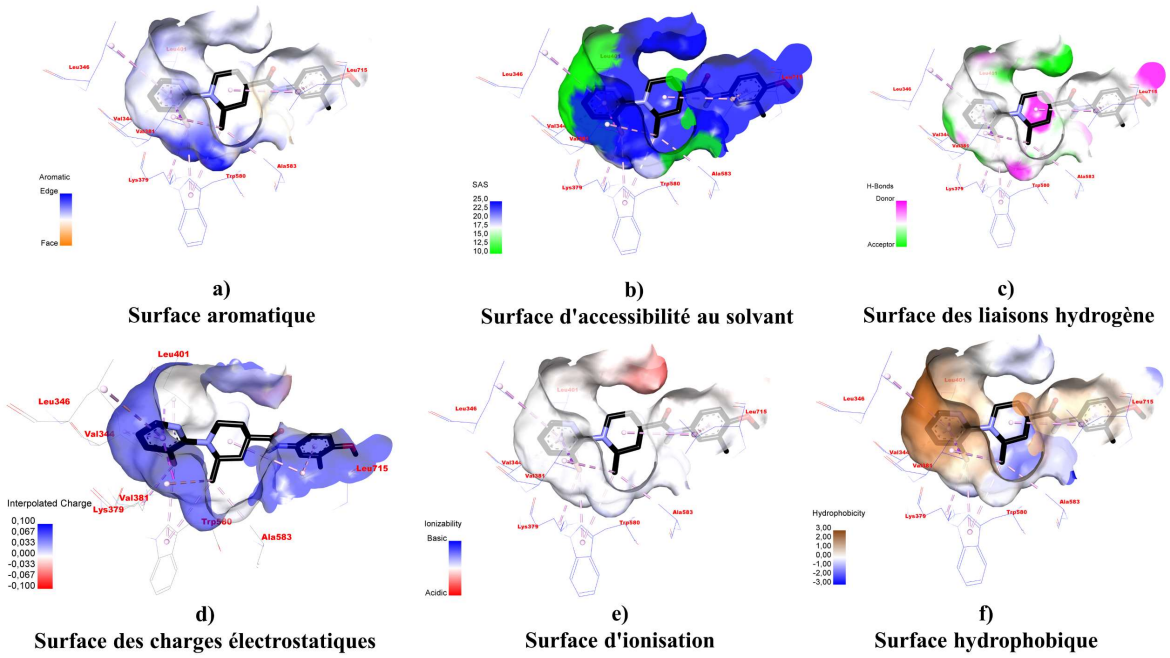


Figure V. A13. Représentation 3D des cavités de surface du système MALT1-D40. (a) Surface aromatique, (b) Surface d'accessibilité au solvant de MALT1 vis-à-vis de D40, (c) Surface des liaisons hydrogène, (d) Surface de charge, (e) Surface d'ionisation, (f) Surface hydrophobique.

La **Figure V. A13(a)** montre que la couleur bleue prédomine dans la structure du ligand D40 alors que la couleur orange est totalement absente, ce qui implique que les cycles aromatiques de la structure de la molécule D40 sont appropriés pour les contacts avec les résidus d'acides aminés alignés sur le bord de la poche active de MALT1. Cela est également validé par les cartes de contour CoMFA/S (**Figure V. A8a**) et CoMSIA/S (**Figure V. A9a**) qui se trouvent à proximité des sites R, R1, R2 et R3 situés sur les cycles aromatiques de la structure D40. La **Figure V. A13(b)** indique que le niveau bleu (SAS élevé) prédomine à celui du vert (SAS faible) ; les résidus d'acides aminés situés en dessous de la poche active de MALT1 ont donc une affinité significative pour la région d'accessibilité au solvant. Conformément à la **Figure V. A13(c)**, la présence de zones violettes (donneur) et vertes (accepteur) proches des sites R, R1, R2 et R3 dans la structure D40 indique que le renforcement de ces sites par des radicaux capables de donner ou d'accepter des liaisons hydrogène est susceptible de favoriser la formation d'interactions avec des résidus d'acides aminés déterminants dans la poche de la MALT1. Ceci est conforme aux cartes de contour de CoMSIA/DA visualisées dans la **Figure V. A9(d,e**). La **Figure V. A13(d)** indique que le bleu (charge positive) est relativement plus visible que le rouge (charge négative), ce qui implique que l'affinité du ligand D40 pour les résidus d'acides aminés actifs de MALT1 est favorisée par la présence en surface de charges positives. Cela corrobore les cartes de contour CoMFA/E (**Figure V. A8b**) et CoMSIA/E (**Figure V. A9b**). Parallèlement, la **Figure V. A13(e)** affiche une absence de bleu (basique), ce qui implique que l'environnement des interactions entre le D40 et les résidus MALT1 n'est pas adapté aux conditions basiques. Dans la dernière **Figure V. A13(f)**, on constate que la surface hydrophobe est relativement élevée par rapport à la surface hydrophile près de la structure du ligand D40, ce qui est également observé dans les cartes de contours de CoMSIA/H

(**Figure V. A9c**). Toutes ces connaissances tirées à la fois de l'analyse des cavités de surface et de la modélisation pharmacophore 3D-QSAR sont en accord les unes avec les autres, ce qui indique clairement qu'un repositionnement approprié de la D40 permet de renforcer la stabilité des interactions médicament-protéine et d'obtenir une activité et une sélectivité thérapeutiques optimales.

3.5.1 Conception computationnelle de nouveaux dérivés de pyridine

Sur la base des orientations et des caractérisations de la modélisation 2D/3D-QSAR et des visualisations de Docking moléculaire, nous réalisons de nouvelles adaptations structurelles au D40, puis nous évaluons leur niveau d'activité *in silico* vis-à-vis de la protéase MALT1. Au cours de cette procédure, il est possible de proposer plusieurs séries de petites molécules basées sur la molécule D40. Dans cette étude, nous considérons seulement 4 nouveaux composés qui ont montré des pIC50 prédits *in silico* plus forts que ceux observés *in vitro* dans le cas de la molécule de référence, le D40, et de l'inhibiteur standard, la Thioridazine. La **Figure V. A14** présente les substitutions clés sur lesquelles nous avons basé pour proposer les nouveaux dérivés pyridiniques (P1-P4) à partir de la structure de la molécule D40.

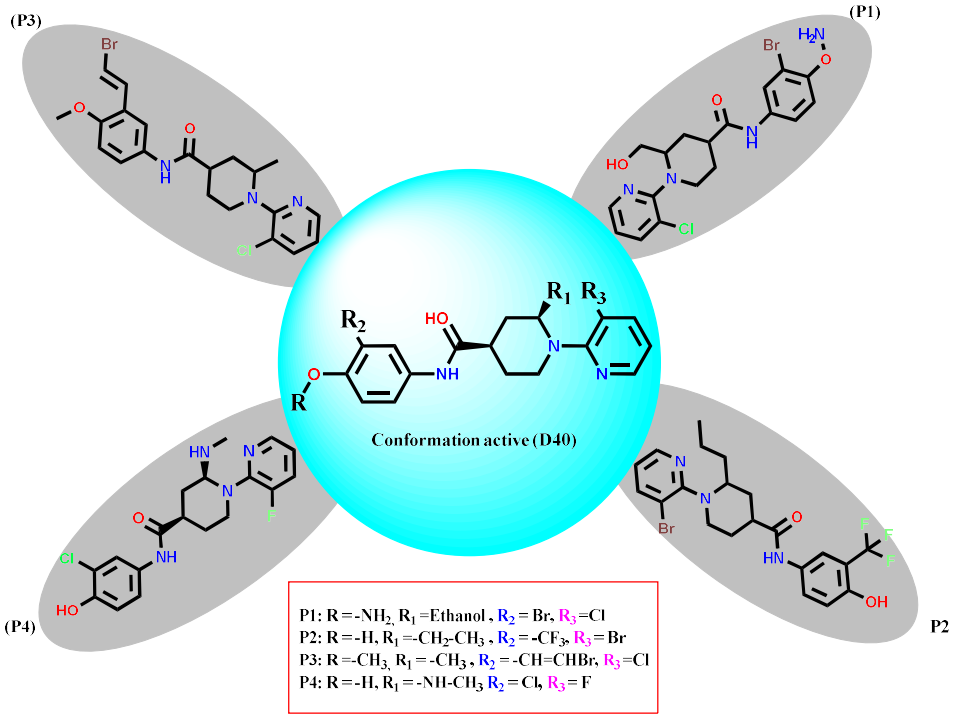


Figure V. A14. Schéma des substitutions proposées pour optimiser de nouveaux inhibiteurs allostériques de MALT1 à base de pyridine

3.5.2 Identification des composés de tête de la série "hits"

La structure géométrique et énergétique des 4 nouvelles structures est optimisée dans les mêmes conditions que celles dans lesquelles les données originales de la base de données ont été préparées. Ceci afin de préserver le critère d'homogénéité structurelle et d'obtenir des prédictions *in silico* fiables et homogènes. Le **Tableau V. A6** présente les informations sur les descripteurs moléculaires et les activités (*in vitro* et *in silico*) des petites molécules synthétisées (A1-D40) et des nouvelles molécules optimisées dans le cadre de cette étude (P1-P4). D'après les résultats présentés dans le **Tableau V. A6**, nous avons constaté que les modèles 2D/3D-QSAR révèlent que les nouveaux dérivés pyridiniques

(P1-P4) sont bien structurés, ce qui leur permet d'exercer une forte activité inhibitrice sur l'activité enzymatique de la protéase MALT1. Ces nouveaux dérivés à base de pyridine présentent des valeurs pIC_{50} prédites supérieures à celles de la molécule modèle D40 (pIC_{50} = 8.00) et de la Thioridazine (pIC_{50} = 4.81) [504]. Par conséquent, cette étude confirme que les modèles 2D/3D-QSAR que nous avons développé *in silico* peuvent être appliqués à la conception des autres dérivés ayant les propriétés inhibitrices appropriées et prévoir leurs effets pIC_{50}. Sur la base des résultats de la modélisation 2D/3D-QSAR, il est très clair que l'activité biologique inhibitrice des quatre molécules (P1-P4) est significativement renforcée par rapport à la molécule de référence D40 et à l'inhibiteur allostérique Thioridazine. Par conséquent, dans le reste des procédures de criblage *in silico* à caractère hypothétique, nous reposons sur l'étude et l'évaluation des molécules P1-P4 candidates en tant qu'inhibiteurs allostériques spécifiques de MALT1. Tout en excluant les molécules D40 et Thioridazine, dont l'activité biologique est la plus faible.

Tableau V. A6. Les valeurs pIC_{50} observées *in vitro* et prédites *in silico* pour les petites molécules à base de pyridine obtenues par la synthèse (A1 à D40) et celles optimisées dans le cadre de cette étude (P1 à P4).

	in vitro	**Prédiction 2D-QSAR *in silico***								**Prédiction 3D-QSAR *in silico***	
	pIC_{50}	E_{HOMO}	MD	E_T	MV	n	αe	LogP	pIC_{50} (RLM)	pIC_{50} CoMFA	pIC_{50} CoMSIA
A1	6.16	-5.31	2.20	-52954.60	293	1.679	43.89	4.86	5.684	6.196	6.131
A2	4.52	-5.34	2.51	-54023.35	311.5	1.662	45.75	4.65	4.874	4.669	4.448
A3	5.89	-5.12	4.43	-52954.67	293	1.679	43.89	3.90	5.919	5.778	5.970
A4	4.75	-5.14	3.92	-54023.92	309.3	1.669	45.8	5.32	5.319	7.697	7.118
A5	4.89	-3.79	4.44	-53405.66	286.2	1.684	43.13	3.57	4.955	5.467	5.194
A6	4.69	-2.54	5.78	-44080.28	272.2	1.686	41.08	2.55	4.361	4.750	4.213
A7	5.96	-4.96	4.06	-43629.45	283.6	1.664	41.7	3.16	6.195	5.623	5.962
A8	4.52	-5.30	1.93	-44230.66	303.5	1.661	44.53	4.28	4.522	4.084	4.509
A9	5.64	-5.12	4.27	-44230.40	303.5	1.661	44.53	4.11	5.073	5.583	6.271
A10	5.40	-5.25	4.58	-44666.36	296.8	1.666	43.77	3.27	5.514	5.527	5.363
A11*	7.00	-5.09	4.51	-41121.39	285.7	1.611	39.37	3.34	6.785	7.243	6.293
A12	7.00	-5.12	3.93	-52552.64	281.4	1.624	39.39	4.09	6.341	6.679	6.659
A13	7.30	-5.13	3.99	-109982.7	285.6	1.634	40.5	4.27	7.704	6.590	6.775
A14	6.01	-5.10	4.99	-51951.16	296.6	1.636	42.2	3.92	6.124	6.488	6.069
A15	6.46	-4.98	5.66	-43166.18	293.4	1.605	40.1	3.33	6.490	6.015	6.750
A16	6.39	-5.15	4.04	-49217.56	302.9	1.57	39.43	4.58	6.647	6.436	6.426
A17*	6.08	-5.08	4.60	-42190.39	302.2	1.604	41.24	3.87	6.295	6.347	6.170
A18	5.22	-5.23	3.02	-56037.02	311.2	1.605	42.52	1.57	5.219	5.236	5.275
A19	5.47	-5.15	3.41	-40052.46	269.4	1.618	37.45	2.88	5.929	5.678	5.871
A20	6.26	-5.12	4.20	-53621.92	297.6	1.617	41.31	4.55	6.249	6.469	6.166
A21*	4.75	-5.07	4.93	-42190.54	302	1.605	41.28	3.80	4.321	5.674	6.068
A22*	5.37	-5.12	4.35	-48816.05	333.2	1.632	47.18	4.25	4.522	6.110	5.985
A23	4.87	-5.28	4.11	-47212.18	319.5	1.655	46.49	2.97	4.961	5.019	5.025
A24	6.64	-5.14	3.67	-52116.64	288.1	1.62	40.15	4.76	6.446	6.685	6.944
B25	7.30	-5.08	4.20	-42190.47	305.7	1.595	41.23	3.84	7.217	7.056	6.529
B26	7.30	-5.09	5.03	-111051.77	305.6	1.616	42.37	4.76	7.770	7.383	7.363
B27*	6.64	-5.05	4.11	-42190.51	305.7	1.595	41.23	3.84	6.109	6.050	6.278
B28	6.54	-5.08	3.84	-43259.44	323.3	1.586	43.06	4.37	6.619	6.621	6.834
B29	4.52	-5.16	1.27	-42191.60	304.1	1.599	41.19	3.88	5.354	4.532	4.430
B30*	4.95	-5.21	1.99	-43227.78	295	1.637	41.99	3.35	5.224	4.277	4.714
C31*	5.76	-5.74	7.21	-60648.92	297.8	1.585	39.57	5.12	5.709	7.192	6.772
C32	7.00	-5.35	4.30	-53556.00	274.3	1.673	40.77	4.58	7.544	7.126	6.952
C33	7.52	-5.42	3.10	-110419.64	278.8	1.638	39.74	4.37	7.566	7.285	7.613
C34	6.89	-5.45	4.88	-53021.36	274.6	1.628	38.64	3.75	7.033	6.826	7.246
C35	7.40	-5.41	5.66	-55159.82	283.9	1.658	41.49	4.11	7.123	7.259	6.623
C36*	6.89	-5.58	5.53	-58638.38	314.4	1.654	45.72	4.00	6.082	7.008	7.375
C37	7.70	-5.43	6.25	-57107.05	287.8	1.652	41.76	3.41	7.150	7.568	7.225
C38	6.00	-4.98	0.74	-52448.85	270	1.665	39.8	3.40	5.624	6.147	6.121
C39	6.27	-5.18	5.42	-44606.73	275.8	1.648	39.82	2.32	6.539	6.798	6.849
D40	**8.00**	-5.51	1.77	-111052.85	305.6	1.616	42.37	4.76	**7.980**	**8.210**	**7.804**
P1	-	**-5.837**	**4.45**	**-113593.88**	**291.3**	**1.659**	**42.54**	**2.39**	**8.049**	**8.492**	**8.318**
P2	-	**-5.767**	**3.76**	**-118847.67**	**335.7**	**1.579**	**44.17**	**3.42**	**9.230**	**9.438**	**8.730**
P3	-	**-5.544**	**1.89**	**-113218.69**	**327.1**	**1.658**	**46.85**	**3.87**	**9.036**	**9.165**	**8.660**
P4	-	**-5.708**	**5.47**	**-111285.78**	**271**	**1.637**	**38.59**	**3.01**	**8.737**	**8.827**	**8.360**

* Éléments de l'ensemble de tests

3.6 Évaluation *in silico* des propriétés biodisponibilité, pharmacocinétique et pharmacocinétique de molécules candidates comme H2L

3.6.1 Prédiction in silico basée sur le modèle "Drug-like"

Dans cette étude, les propriétés pharmacocinétiques et de type Drug-like des composés conçus de manière computationnelle (P1-P4) ont été explorées afin de filtrer les candidats H2L. Plusieurs méthodes connues, notamment les règles de biodisponibilité de Lipinski, Veber et Egan, sont employées pour analyser la structure 2D des petites molécules afin de déterminer leur potentiel d'abondance biologique par voie orale. Le paramètre d'absorption (ABS), le coefficient de surface polaire totale (TPSA), le nombre de liaisons rotatives (n-RB), le poids moléculaire (MW), la lipophilie (LogP), le nombre d'accepteurs/donneurs de liaisons hydrogène (n-HBA, n-HBD) et l'accessibilité synthétique (SA) ont également été évalués afin de déterminer le potentiel de pénétration et de perméabilité des molécules dans les membranes cellulaires, ainsi que leur capacité à être synthétisées. Un modèle ADMET basé sur les descripteurs de surface polaire (2D-PSA) et l'AlogP98 (**Figure V. A15**) est utilisé dans ce contexte pour identifier le potentiel d'absorption et de passage de la barrière hémato-encéphalique (BHE). Le **Tableau V. A7** présente divers profils de propriétés de type drug-like associés aux molécules P1, P2, P3 et P4.

Tableau V. A7. Modèle de type Drug-like pour évaluer les molécules optimisées du point de vue de leurs biodisponibilités.

	Propriétés de type Drug-like							Violations			SA
	ABS	TPSA	n-RB	MW	LogP	n-HBA	n-HBD	Lipinski	Veber	Egan	
Règle	-	<140 Å²	<10	<500 Da	≤5	<10	<5	≤2	≤2	≤2	0-10
P1	Forte	100.71	6	455.73	2.39	5	3	0	0	0	3.71
P2	Forte	65.46	7	486.33	3.42	6	2	0	0	0	3.62
P3	Forte	54.46	6	464.78	3.87	3	1	0	0	0	3.63
P4	Forte	77.49	5	378.83	3.01	5	3	0	0	0	3.55

Les prédictions de type Drug-like obtenues dans le **Tableau V. A7** indiquent que les quatre nouveaux composés (P1-P4) respectent bien les règles de Lipinski, Veber et Egan, et ont une forte absorption par l'intestin et une bonne aptitude à la synthèse (tous ont des valeurs SA inférieures à 3.72). De plus, l'indice n-RB révèle que les structures de ces molécules contiennent toutes moins de dix liaisons rotatives, ce qui signifie qu'elles peuvent se lier efficacement aux sites actifs de la poche de la protéine MALT1. En outre, P2, P3 et P4 ont des valeurs d'indice TPSA inférieures à 90 Å², ce qui signifie que ces molécules peuvent potentiellement traverser la barrière hémato-encéphalique. En effet, la molécule P1 présente la valeur TPSA de 100 Å², cette molécule peut donc constituer un obstacle pour la pénétration de la BHE à partir de sa surface. Cette hypothèse peut être confirmée par la **Figure V. A15**, on peut noter que la molécule P1 existe en dehors de la zone d'éclipse d'absorption dans l'intestin.

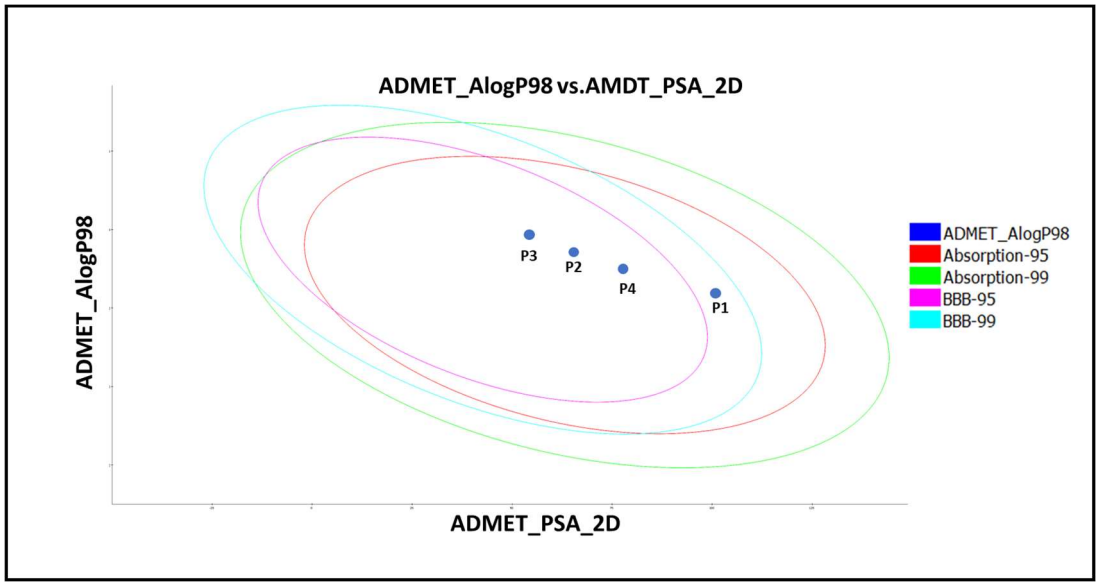


Figure V. A15. Graphique de AlogP98 vs 2D-PSA pour les quatre nouveaux composés optimisés à base de pyridine.

Les résultats présentés à la **Figure V. A15** indiquent que les composés P2, P3 et P4 possèdent à la fois des propriétés d'absorption intestinale et vis-à-vis de la barrière hémato-encéphalique, comme l'indique leur présence dans les éclipses verte et bleue, avec des intervalles de confiance de 99% et de 95%, respectivement. En outre, le composé P1 se trouve en dehors de l'éclipse pour l'absorption intestinale à la limite de confiance de 95 %. Par conséquent, les composés P2, P3 et P4 peuvent être sélectionnés comme les meilleurs H2L en vue de choix d'inhibiteurs allostériques spécifiques de MALT1. Toutefois, il convient de poursuivre l'évaluation du composé P1 afin de déterminer s'il doit rester dans la sélection ou être exclu.

3.6.2 Évaluations ADME-Tox in silico

Le **Tableau V. A8** présente les propriétés ADME-Tox prédites des H2L proposées (P1-C4). Ces propriétés comprennent l'absorption dans le sang, le transfert vers les sites cibles, les transformations métaboliques dans l'organisme, l'excrétion et la toxicité potentielle.

Tableau V. A8. Évaluation des propriétés pharmacocinétiques/pharmacodynamiques *in silico* des molécules candidates comme H2L via le modèle ADMT-Tox.

	Propriétés													
	Absorption		**Distribution**			**Métabolisme**							**Excrétion**	**Toxicité**
						CYP450								
						Substrat		Inhibiteur						
Modèles	Absorption intestinale (Humain)	P-Gp Substrat	VDss (humain)	Perméabilité à la BHE	Perméabilité du SNC	2D6	3A4	1A2	2C19	2C9	2D6	3A4	Clairance totale	Toxicité d'AMES
P1	76.07	Oui	-0.174	-1.177	-3.029	Non	Non	Non	Oui	Non	Oui	Non	-0.207	Non
P2	89.55	Oui	0.177	0.072	-2.791	Non	Oui	Non	Oui	Oui	Non	Oui	-0.209	Non
P3	92.30	Oui	0.204	0.038	-1.806	Non	Oui	Non	Oui	Oui	Non	Oui	-0.122	Non
P4	91.53	Oui	0.655	-1.258	-2.421	Non	Non	Non	Oui	Non	Oui	Non	0.600	Non

P-Gp : P-glycoprotéine ; VDss : volume de distribution à l'état stable ; BHE : barrière hémato-encéphalique ; SNC : système nerveux central ; CYP : Cytochrome

L'indice d'absorption d'un médicament reflète sa capacité à être absorbé dans l'intestin humain lorsqu'il atteint le sang, et les médicaments qui présentent une valeur d'absorption supérieure à 30 % sont considérés comme ayant de bonnes propriétés d'absorption. La glycoprotéine P peut contrôler

l'absorption de ses substrats en les recyclant activement et en permettant au CYP3A4 de les métaboliser plusieurs fois dans l'intestin. Pour assurer une bonne absorption du médicament, ses composés doivent être des substrats de la P-glycoprotéine. Parmi les 4 composés soumis à cet essai, P2, P3 et P4 ont une plus grande capacité d'absorption que P1 et agissent comme des substrats de la glycoprotéine P, ce qui les rend hautement absorbables dans l'intestin. Le volume de distribution (VDss) est un indicateur important pour prédire comment un médicament est distribué dans l'organisme, et les médicaments dont les valeurs de LogVDSS sont supérieures à -0.15 sont considérés comme ayant de bonnes capacités de distribution. D'après le **Tableau V. A8**, P2, P3 et P4 avaient des valeurs VDss élevées, tandis que P1 avait une valeur faible. La capacité des composés à traverser la barrière hémato-encéphalique (BHE) ou à accéder au système nerveux central (SNC) est très importante pour l'administration des médicaments, et les médicaments dont les valeurs LogBB sont supérieures à -1 et les valeurs LogPS supérieures à -2 ont de bonnes propriétés pour la BHE et le SNC. P1 et P4 ont montré une faible capacité à traverser la BHE et/ou à atteindre le SNC, tandis que P2 et P3 avaient une bien meilleure capacité. La bonne administration des médicaments dépend des enzymes cytochromes, et l'isoenzyme la plus importante, CYP450, est responsable de plus de 90 % de l'oxydation des médicaments dans l'organisme. Le **Tableau V. A8** indique que P2 et P3 agissent non seulement comme substrats du CYP3A4, mais sont également des inhibiteurs de l'enzyme, et que les composés P1 et P4 ne sont ni des substrats ni des inhibiteurs de cet enzyme. Les valeurs de la clairance totale des composés P1, P2 et P3 sont faibles, ce qui indique un temps de rétention plus élevé pour ces molécules dans l'organisme, ce qui en fait des candidats appropriés pour la thérapie. Par contre, le composé P4 a montré une valeur de clairance totale relativement élevée ; par conséquent, le composé P4 serait rapidement éliminé et le composé ne pourrait donc pas atteindre son objectif. En outre, la prédiction de la toxicité via le test AMES a révélé que les composés P1, P2, P3 et P4 ne présentent aucun comportement toxique, ce qui suggère qu'ils sont des bons candidats médicaments. Les prédictions des propriétés Drug-like et ADME-Tox de P2 et P3 indiquent qu'ils ont des profils prometteurs de biodisponibilité, de pharmacocinétique et de pharmacodynamique. Ces deux composés sont donc considérés comme des candidats médicaments positifs pour inhiber l'activité enzymatique de MALT1. En cas de succès, cela pourrait améliorer la régulation de la voie de signalisation NF-κB et jouer un rôle clé dans la protection des patients atteints de lymphome.

3.7 Analyse des interactions entre les systèmes biomoléculaires MALT et les inhibiteurs allostériques candidats

Pour vérifier les résultats du filtrage effectué jusqu'à ce stade, nous soumettons les molécules P2 et P3 à une analyse approfondie sur leur compatibilité structurelle en tant qu'inhibiteurs spécifiques de l'activité enzymatique anormale de la protéase MALT1. Pour réaliser cet examen, nous étudions les mécanismes de compatibilité/affinité des complexes MALT1-P2 et MALT1-P3 par des simulations de Docking moléculaire et des explorations post-Docking intégrant des calculs MM-GBSA et de la dynamique moléculaire.

3.7.1 Simulations de Docking Moléculaire et calculs Prime/MM-GBSA

La compatibilité des ligands P2 et P3 avec la poche active allostérique de MALT1 est évaluée à l'aide de modèles de Docking moléculaire et de simulations Prime/MM-GBSA. La **Figure V. A16** et le **Tableau V. A9** mettent en évidence les poses optimales de P2 et P3 avec MALT1 ainsi que les énergies de liaison et les interactions clés pour les deux complexes. Plus précisément, la meilleure énergie de liaison et l'énergie de liaison libre pour MALT1-P2 étaient respectivement de -9.4 kcal/mol et de -47.35 kcal/mol, tandis que la meilleure énergie de liaison et l'énergie de liaison libre pour MALT1-P2 étaient respectivement de -8.7 kcal/mol et de -32.18 kcal/mol.

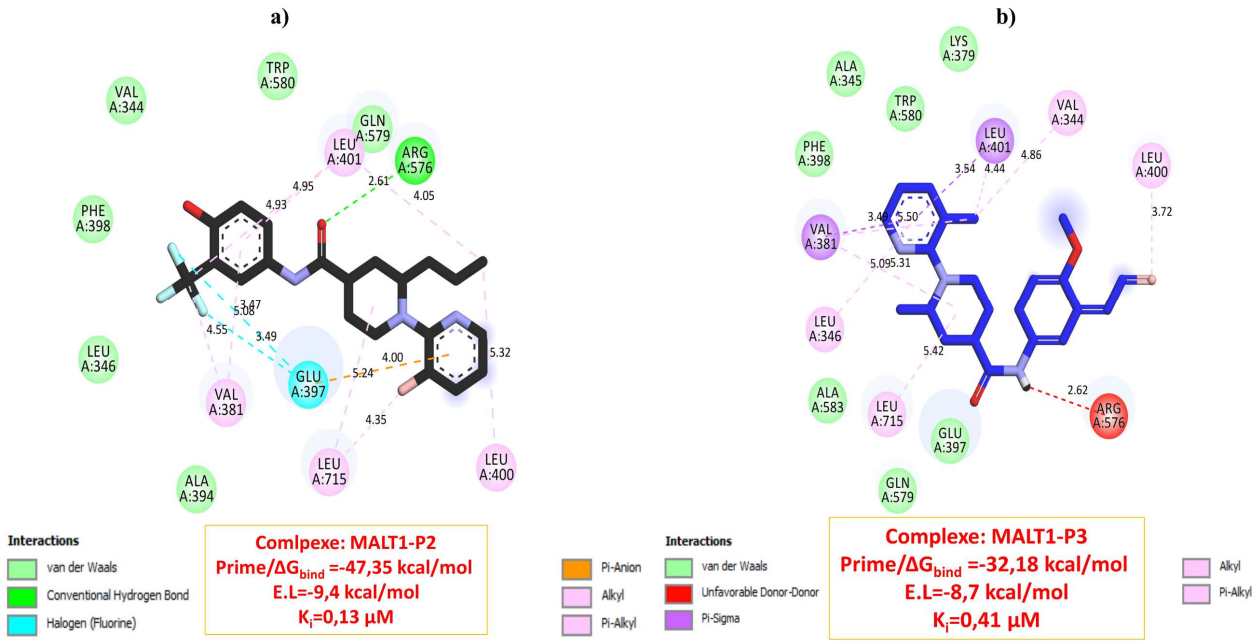


Figure V. A16. Caractérisations visuelles des profils de compatibilité/affinité des systèmes MALT1-P2 et MALT1-P3 modélisés par docking moléculaire et calculs de Prime/MM-GBSA.

Les énergies de liaison des ligands P2 et P3 optimisés *in silico* apparaissent plus stables que celles générées par l'optimisation *in vitro* (D40) et également par l'inhibiteur de référence (Thioridazine). Cela confirme la plus grande compatibilité de la structure des nouvelles molécules P2 et P3 avec l'environnement de la poche allostérique active de MALT1 par rapport aux références D40 et Thioridazine. Cela peut être confirmé par les valeurs très faibles des constantes d'inhibition estimées (Ki) pour P2 (0.13 µM) et P3 (0.41 µM) par rapport à D40 (1.1 µM) et Thioridazine (0.58 µM). La stabilité et la compatibilité de P2 et P3 avec la structure de MALT1 peuvent s'expliquer par la contribution significative des liaisons hydrophobes, électrostatiques, halogènes et hydrogène, ainsi que des interactions de van der Waals, qui conduisent à une liaison non covalente plus relaxante avec les sites actifs allostériques du récepteur MALT1. Nous pouvons également constater que les ligands modélisés P2 et P3 génèrent plusieurs interactions non covalentes avec les résidus d'acides aminés allostériques actifs de référence (marqués en **gras** dans le **Tableau V. A9**). Cela renforce encore l'hypothèse de la grande affinité et de la compatibilité des systèmes biomoléculaires MALT1-P2 et MALT1-P3. Cela confirme davantage la supposition que les nouvelles molécules P2 et P3 développées dans la présente étude par voie computationnelle sont de potentiels inhibiteurs allostériques de l'activité enzymatique anormale de la protéase MALT1.

Tableau V. A9. Mécanismes d'interactions non covalentes protéine-ligand des complexes MALT1-P1 et MALT1-P2 modélisés dans cette étude vis-à-vis des complexes de référence MALT1-TH et MALT1-D40.

Complexes optimisés	Nombre et type d'interactions non covalentes	Résidus d'acides aminés actifs
MALT-P2	Liaison H :1, Halogène :2, Alkyl : 6, π-Alkyl : 2, π-Anion : 1, van der Waals : 5	**Arg576** (2.61Å). **Glu397**(3.47, 3,49 Å), **Val381** (4.55 Å). **Leu715** (4.35 Å), Leu400 (5.32 Å), Leu401 (2.61, 4.05 Å). **Val381** (5.08 Å), **Leu715**(5.24 Å). Ala394, **Leu346**, **Phe398**, Val344, **Trp580**
MALT1-P3	Liaison H :1, Alkyl : 6, π-Alkyl : 1, π-σ: 2, van der Waals: 7	**Arg576** (2.62 Å). **Val381**(5.50, 5.31 Å), **Leu715**(5.42 Å), Leu400(3.72 Å), **Val344**(4.86 Å), Leu401(4.44 Å). **Leu346**(5.09 Å). **Val381**(3.49 Å), Leu401(3.54 Å). **Gln579**, **Glu397**, **Ala583**, **Phe398**, Ala345, **Trp580**, Lys379.
Complexes de référence		Poche active allostérique de MALT1
	MALT1-TH (PDB :4I1R)	Acide glutamique (**Glu397**), Tryptophane (**Trp580**), Alanine (**Ala394**), Valine (**Val381**, **Val344**), Leucine (**Leu401**), Arginine (**Arg576**), phénylalanine (**Phe398**)
	MALT1-D40	Acide glutamique (**Glu397**), Tryptophane (**Trp580**), Alanine (**Ala583**, **Ala345**), Valine (**Val381**, **Val344**), Leucine (**Leu346, Leu715**), Arginine (**Arg576**), Glutamine (**Gln579**), Lysine (**Lys582**), phénylalanine **(Phe398**)

3.7.2 Analyses post-Docking via simulations de dynamique moléculaire et calculs Thermal/MM-GBSA

Pour valider les résultats de l'identification précoce d'une tête de série optimisée (LO), les molécules impliquées doivent être soumises à diverses analyses expérimentales et/ou computationnelles à vaste échelle. À cet égard, dans la présente étude, nous avons mené une étude computationnelle exhaustive afin de confirmer la validité des résultats obtenus dans le cadre de cette étude. Pour ce faire, nous avons soumis les deux systèmes MALT1-P2 et MALT1-P3 à des simulations de dynamique moléculaire à une échelle de temps de 100ns. Au cours de cette étape, nous avons utilisé le complexe original MALT1-TH (code PDB : 4I1R) comme référence pour évaluer les caractérisations dynamiques au niveau biomoléculaire. Comme une dernière étape de vérification de la qualité du modèle de dynamique moléculaire mis en œuvre dans cette étude, nous avons soumis les systèmes d'inhibiteurs de MALT1 étudiés à deux tests de calcul très approfondis incluant l'évaluation de l'énergie libre de liaison totale et l'évaluation des paramètres thermodynamiques.

3.7.2.1 *Analyse des RMSD et RMSF*

Nous avons utilisé la RMSD et la RMSF pour évaluer le niveau de déviations et de fluctuations structurelles potentielles dans les systèmes MALT1 complexés (MALT1-P2, MALT1-P3 et MALT1-TH) par rapport au système de référence MALT1 non complexé (MALT1 libre). Le mouvement des atomes du squelette α-carbone pour chaque résidu d'acide aminé dans les systèmes MALT1 analysés est calculé à l'aide de ces deux mesures par rapport au premier point de référence du squelette de la protéine MALT pendant la période de 100ns. La **Figure V. A17** montre les délais des évolutions de la RMSD et de la RMSF relatives aux systèmes testés (MALT1 libre, MALT-P2, MALT1-P3 et MALT1-Thioridazine).

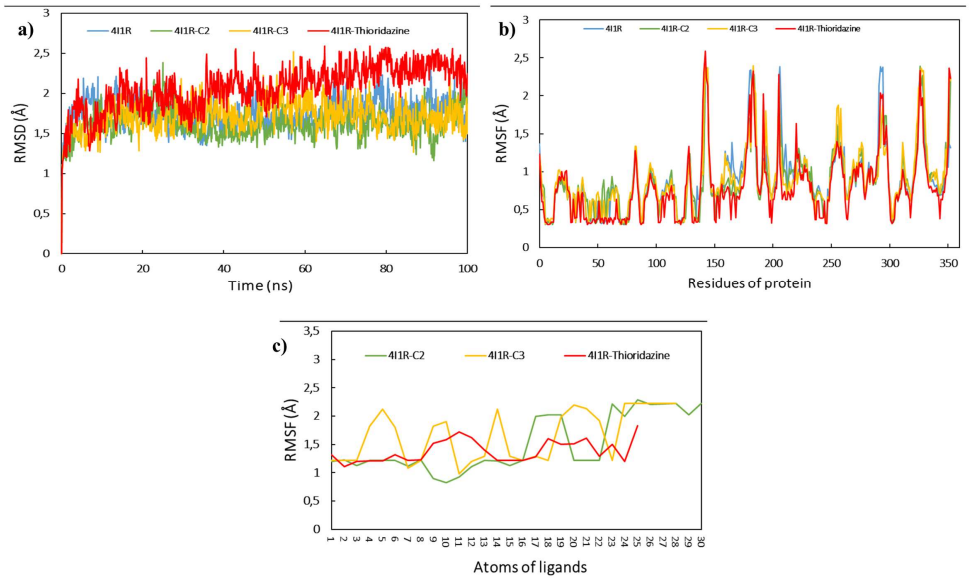


Figure V. A17. (a) RMSD (b) RMSF des systèmes : (MALT1 free, MAL1-P2, MALT1-P3 et MALT1- Thioridazine), (c) RMSF timeline des ligands (P2, P3 et Thioridazine) tout au long des 100ns de simulations de la dynamique moléculaire.

Les données RMSD présentées à la **Figure V. A17(a)** suggèrent que les complexes MALT1 optimisés avec les inhibiteurs P2 et P3 (MALT1-P2 et MALT1-P3) sont devenus rapidement stables du début de la simulation MD (0ns) jusqu'à sa fin (100ns). Malgré les légères fluctuations (ne dépassant pas 2.5 Å) observées dans le complexe MALT1-Thioridazine (code PDB : 4I1R), les écarts RMSD des complexes MALT1-P2 et MALT1-P3 sont restés inférieurs à ceux du complexe MALT1-Thioridazine tout au long des simulations. Par conséquent, les complexes MALT1-P2 et MALT1-P3 ont pu plus facilement conserver un équilibre dans leur environnement aqueux que le complexe de référence MALT1-Thioridazine (4I1R). Les résultats tirés de la **Figure V. A17(b)** indiquent un niveau élevé de stabilité pour les atomes (α-carbon) des résidus d'acides aminés allostériques de la protéase MALT1, avec des valeurs ne dépassant pas 0.85 Å. Cela suggère que ces résidus sont bien stables dans un environnement aquatique, comme en témoignent les valeurs RMSF proches entre les systèmes étudiés MALT1_free, MALT1-P2, MALT1-P3, et MALT1-Thioridazine. Certains acides aminés latéraux, notamment Lys543, Gly544, Ser573, Val659, Lys661 et Gly694, ont connu des fluctuations plus importantes allant jusqu'à 2 Å. La **Figure V. A17(c)** examine les calculs RMSF des nouveaux inhibiteurs de MALT1 (P2, P3 et Thioridazine) et suggère que tous ces ligands évoluent de manière dynamique dans la poche bioactive. Ce comportement indique que les interactions entre les nouveaux inhibiteurs optimisés P2 et P3 et l'inhibiteur standard Thioridazine et la poche allostérique de MALT1 sont stables.

3.7.2.2 *Propriétés dynamiques des ligands optimisés*

La **Figure V. A18** présente les propriétés dynamiques des inhibiteurs optimisés (P2, P3) et de l'inhibiteur standard (Thioridazine) dans la poche allostérique active de MALT1. La RMSD, le rayon de giration (rGyr), la liaison hydrogène intramoléculaire (intraHB), la surface moléculaire (MolSA), la surface accessible au solvant (SASA) et la surface polaire (PSA) de ces inhibiteurs ont été évalués au cours de la simulation MD de 100 ns dans un environnement aqueux, et les résultats ont montré que leurs propriétés étaient remarquablement stables au cours de cette période des simulations.

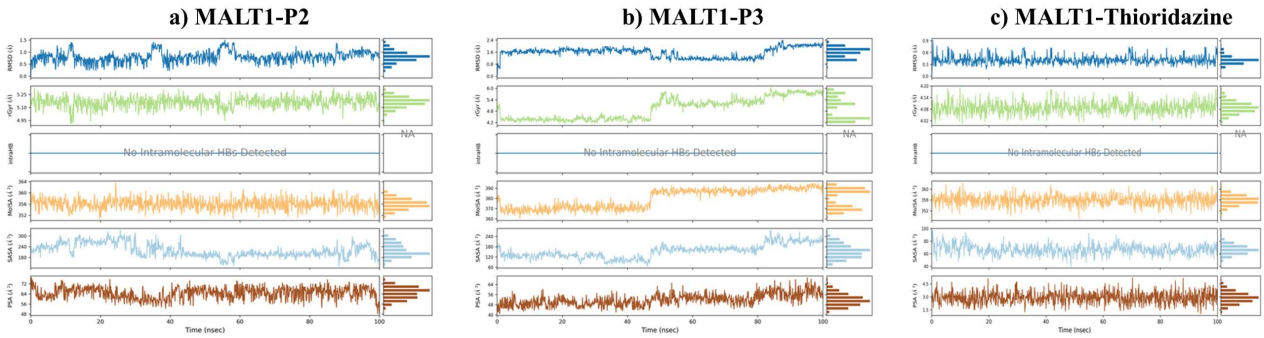


Figure V. A18. Propriétés dynamiques des nouveaux inhibiteurs allostériques optimisés de MALT1 P2 (a) et P3 (b), et de l'inhibiteur standard Thioridazine (c).

3.7.2.3 *Mécanisme des contacts protéine-ligand*

Les visualisations des contacts protéine-ligand donnent un aperçu dynamique plus clair sur les mécanismes d'interaction entre les ligands optimisés et le site actif de liaison au sein de la protéine ciblée. La **Figure V. A19** illustre les trajectoires des interactions non covalentes potentiellement stables entre les ligands (P2, P3 et Thioridazine) et le site allostérique actif de la protéase MALT1 pendant 100ns de la simulation par la dynamique moléculaire.

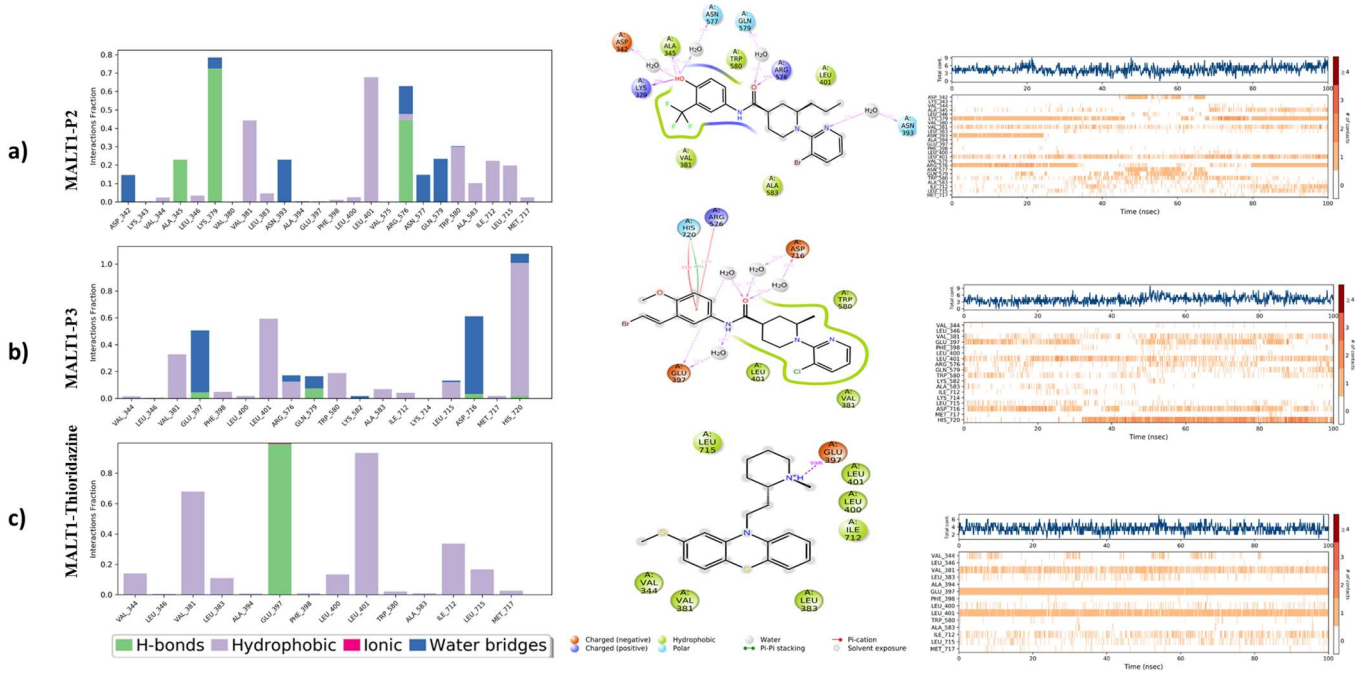


Figure V. A19. Rapports de diagrammes protéine-ligand obtenus après 100 ns de simulations de dynamique moléculaire pour les complexes MALT1-P2 (a), MALT1-P3 (b) et MALT1-Thioridazine (c).

La **Figure V. A19** représente une série de simulations qui suggèrent que P2, P3 et l'inhibiteur standard Thioridazine conservent leurs interactions avec la poche active allostérique de MALT1 aux cours de 100ns des trajectoires de dynamique moléculaire. Les interactions de type hydrophobique, les liaisons hydrogène et les ponts d'eau paraissent largement contribuer à la stabilité des ligands dans la poche active allostérique de MALT1. Les mêmes contacts protéine-ligand observés par Docking moléculaire et dynamique moléculaire dans les complexes MALT-P2, MALT1-P3 et MALT1-Thioridazine confirment la validité et la précision de la méthodologie de criblage virtuel hybride basée sur la CADD mise en œuvre dans le cadre de cette étude. En plus de cette attention, les simulations de dynamique moléculaire nous ont permis de détecter de nouveaux contacts que le protocole de

Docking moléculaire n'avait pas pu détecter (mis en évidence par la ligne bleue en **gras** dans le **Tableau V. A10**). Les nouveaux contacts Protéine-ligand concernent principalement les résidus d'acides aminés Lysine (Lys379), Méthionine (Met717), Isoleucine (Ile715), Acide Aspartique (Asp342), Asparagine (Asn393, Asn577), et Glutamine (Gln579) dans le complexe MALT1-P1 ; les résidus d'acide aminé Isoleucine (Ile712), Méthionine (Met717), Lysine (Lys582), Histidine (His720), Acide Aspartique (Asp716) dans le complexe MALT1-P3 ; une seule nouvelle interaction est détectée avec l'acide aminé Méthionine (Met717) dans le complexe de référence MALT1-Thioridazine. Grâce à cette analyse dynamique précise, l'ensemble des contacts protéine-ligand dans les complexes en question peut être considéré comme la poche allostérique active de MALT1 avec laquelle les petites molécules modélisées dans cette étude ou d'autres molécules candidates au développement de nouveaux inhibiteurs de MALT1 sont susceptibles d'interagir.

Tableau V. A10. Aperçu général des contacts protéine-ligand observés lors des simulations de dynamique moléculaire comparé aux prédictions du modèle de Docking moléculaire mis en œuvre sur les complexes MALT1-P2, MALT1-P3, et MALT1-Thioridazine.

Système	Contacts protéine-ligand
	Simulations de la dynamique moléculaire
MALT1-P2	Liaisons H : **Ala**345 (>20%), **Lys379** (>70%), **Arg576** (>50%). Hydrophobique : **Val344**, **Leu346**, **Leu**383, **Leu**400, **Met717** (>50%), **Val381** (>45%), **Leu401** (>70%), **Trp580** (>30%), **Ala**583, **Ile715**, **Leu**715 (>20%). Ponts d'eau : **Asp342** (>15%), **Asn393** (>20%), **Arg576** (>60%), **Asn577** (>15%), **Gln579** (>20%).
	Simulations de Docking moléculaire
	Arg576, Glu397, Val381, Leu715, Leu400, Leu401, Leu715, Ala394, Leu346, Phe398, Val344, Trp580
	Simulations de la dynamique moléculaire
MALT1-P3	Liaisons H : **Glu397** (>5%), **Gln579** (>10%), **Asp716** (>5%), **His720** (>3%). Hydrophobique : **Val344** (>3%), **Val381** (>25%), **Phe398** (>5%), **Leu401** (>50%), **Arg576** (>15%), **Trp580** (15 %), **Ala**583, **Ile712**, **Met717** (>3%), **His720** (>80%). Ponts d'eau : **Glu397** (>40%), **Arg576**, **Gln579**, **Lys582**, **His720** (>3%), **Asp716** (>60%).
	Simulations de Docking moléculaire
	Arg576, Val381, Leu715, Leu400, Val344, Leu401, Leu346, Val381, Leu401, Gln579, Glu397, Ala583, Phe398, Ala345, Trp580, Lys379.
	Simulations de la dynamique moléculaire
MALT1-Thioridazine	Liaisons H : **Glu397** (>80%). Hydrophobique : **Val344** (>15%), **Val381** (>60%), **Leu**383, **Leu**400 (>10%), **Trp580**, **Met717** (>3%), **Leu**715 (>15%). Ponts d'eau : Non observé.
	Simulations de Docking moléculaire
	Acide glutamique (Glu397), Tryptophane (Trp580), Alanine (Ala394), Valine (Val381, Val344), Leucine (Leu401), Arginine (Arg576), phénylalanine (Phe398)

Le trait **gras** indique le résidu d'acide aminé de référence (**noir**/original ; **bleu**/nouveau).

Également, à partir du résumé des interactions/contacts protéine-ligand présenté dans le **Tableau V. A10**, nous pouvons constater que le temps de proportion de certains contacts (<10%) entre les ligands P2, P3 et Thioridazine et la poche active allostérique de MALT1 a diminué sur 100ns de temps de simulation. Cela suggère que les faibles contacts non covalents produits entre les ligands et la poche active allostérique de MALT1 contribuent à la diminution du temps de contact dans les complexes protéines-ligands. Cette proposition peut être vérifiée par la détermination et l'évaluation de l'énergie de liaison libre totale (ΔG_{bind}) des systèmes protéine-ligand produits à partir de toutes les trajectoires de dynamique moléculaire. À cette fin, l'algorithme Thermal/MM-GBSA est utilisé dans cette étude comme protocole de vérification.

3.7.2.4 *Estimations de l'énergie libre de liaison par le modèle Thermal/MM-GBSA*

Le **Tableau V. A11** présente le bilan des paramètres moyens des énergies libres de liaison estimée par le modèle Thermal/MM-GBSA pour les systèmes d'inhibiteurs de MALT1 (MALT1-P2, MALT1-P3 et MALT1-Thioridazine) au cours de 100ns de simulation de la dynamique moléculaire. Nous avons représenté graphiquement les estimations recueillies dans la **Figure V. A20**. Les inhibiteurs de MALT1 modélisés (P2 et P3) présentent des énergies de liaison libres moyennes comprises entre -51.31 kcal/mol et -58.15 kcal/mol, plus élevées que le médicament de référence, la Thioridazine (-46.69 kcal/mol). Selon ces données, les faibles interactions non covalentes formées entre MALT1 et les molécules à base de pyridine permettent une inhibition favorable des enzymes MALT1 et sont donc plus appropriées que les interactions covalentes pour obtenir une réponse médicamenteuse. Cela peut être confirmé par les valeurs négatives d'autres paramètres d'énergie libre de liaison tels que (ΔG_{bind} Vdw), (G_{bind} Lipo), (G_{bind} Hbond), (ΔG_{bind} Packing), et (ΔG_{bind} Coulomb), qui contribuent à la production d'une énergie libre de liaison favorable (ΔG_{bind} totale) pour stabiliser les systèmes MALT1-inhibiteurs. Cependant, les contributions d'énergie de liaison positive de (ΔG_{bind} Solv_GB) et (G_{bind} Covalent) n'étaient pas favorables à la production de (ΔG_{bind} totale).

Tableau V. A11. Moyennes des énergies libres de liaison estimées par le modèle Thermal/MM-GBSA pendant 100ns de dynamique moléculaire pour les complexes MALT1-P2, MALT1-P3 et MALT1-Thioridazine.

Modèle Thermal/MM-GBSA	Complexe MALT1-inhibiteurs		
	Tête optimisée (LO)		Standard
ΔG_{bind} (kcal/mol)	**P2**	**P3**	**Thioridazine**
ΔG_{bind} totale	-58.15	-51.31	-46.69
ΔG_{bind} Vdw	-47.38	-44.81	-38.34
ΔGbind Lipo	-21.53	-19.68	-17.38
ΔG_{bind} Hbond	-2.35	-1.37	-2.03
ΔG_{bind} Packing	-3.89	-3.17	-2.89
ΔG_{bind} Coulomb	-18.48	-16.26	-8.34
ΔG_{bind} Solv_GB	20.11	23.78	29.54
ΔG_{bind} Covalent	6.34	8.44	7.13

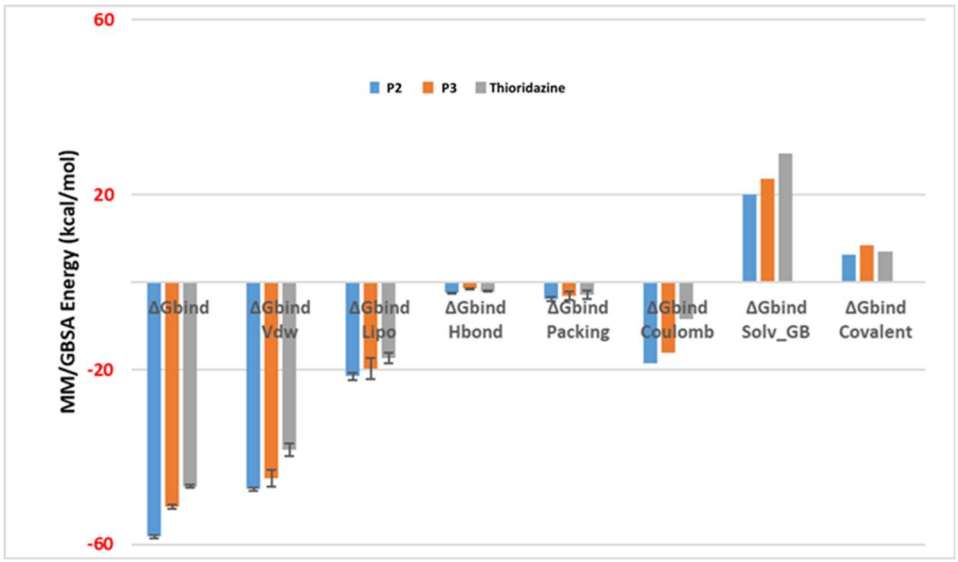


Figure V. A20. Représentation graphique du modèle Thermal/MM-GBSA concernant les scores d'énergie libre de liaison mesurés pendant 100ns de simulation de dynamique moléculaire.

3.7.3 Analyse des effets des propriétés thermodynamiques

Nous avons évalué la qualité des simulations MD effectuées sur les systèmes MALT1-AIs (inhibiteurs allostériques de MALT1) étudiés en analysant les effets thermodynamiques (énergie totale, énergie potentielle, température, pression et volume) au cours d'une période de 100ns pour chacun des systèmes. La **Figure V. A21(a-d)** présente les résultats de cette évaluation en termes de tracés thermodynamiques des systèmes MALT1 libre, MALT1-P2, MALT1-P3 et MALT1-Thioridazine. Les résultats présentés dans la **Figure V. A21** montrent que les paramètres thermodynamiques du MALT1_free et complexé aux inhibiteurs P2, P3 et Thioridazine sont très voisins et restent stables sur les trajectoires de 100ns des simulations de dynamique moléculaire. Le **Tableau V. A12** résume les valeurs moyennes des paramètres thermodynamiques (E, EP, T, P et V) obtenues à partir de ces simulations. Les valeurs des paramètres E, EP et T sont similaires dans tous les systèmes, mais le paramètre de pression du système MALT1-Thioridazine était plus élevé que les autres systèmes (4I1R_free, 4I1R-P2, 4I1R-P3), probablement en raison des interactions de la fraction 1-méthylpipéridine générant un ion (NH^+) et créant une réponse inhibitrice instable vis-à-vis de MALT1. Cela suggère que les inhibiteurs allostériques de MALT1 optimisés P2 et P3, à base de pyridine, peuvent être plus adaptés pour obtenir une bonne action biologique. Des analyses et validations post-Docking complètes par dynamique moléculaire ont confirmé la stabilité des nouveaux petits dérivés pyridiniques allostériques optimisés (LO) P2 et P3 en tant qu'inhibiteurs allostériques médicamenteux potentiels de la MALT1. La forte affinité de ces molécules vis-à-vis de la poche allostérique active de la MALT1 suggère qu'elles pourraient être des inhibiteurs spécifiques efficaces de la MALT1. Par conséquent, P2 et P3 pourraient être utilisées pour développer d'autres molécules susceptibles pour répondre aux pathologies provoquées par l'activité enzymatique anormale de MALT1.

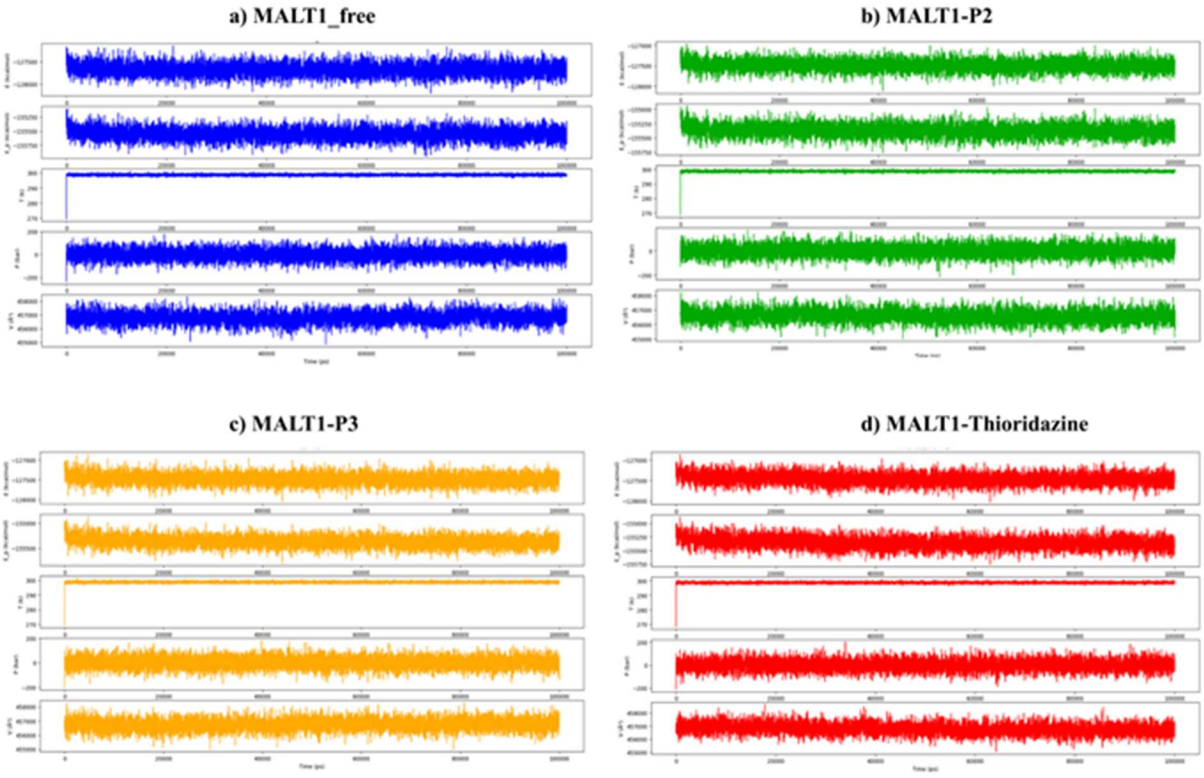


Figure V. A21. Caractéristiques thermodynamiques des inhibiteurs de MALT1 examinés, (a) MALT1_free, (b) MALT1-P2, (c) MALT1-P3, (d) MALT1-Thioridazine.

Tableau V. A12. Valeurs moyennes des propriétés thermodynamiques calculées à partir des trajectoires de 100ns simulées par la DM pour les systèmes biomoléculaires de MALT (MALT_free, MALT1-P2, MALT1-P3, et MALT1-Thioridazine).

Systèmes de MALT1 (PDB :4I1R)	Propriétés thermodynamiques				
	Énergie totale (kcal/mol)	Énergie Potentielle (kcal/mol)	Température (Kelvin)	Pression(bar)	Volume (Å3)
MALT1_free	-127676.9	-155553.5	298.677	0.115	456791.204
MALT1-P2	-127489.9	-155370.9	298.673	0.460	456584.454
MALT1-P3	-127489.9	-155366.9	298.682	0.821	456726.410
MALT1-Thioridazine	-127458.7	-155357.0	298.680	1.376	456798.343

4 Conclusion

Dans cette étude, nous avons rationalisé dans le monde *in silico* les méthodes LBDD et SBDD appliquées dans le cadre d'une stratégie de CADD hybride. Cette méthodologie a été appliquée au criblage, à la conception et à l'optimisation de nouvelles petites molécules biologiquement actives contre l'activité enzymatique de la protéase MALT1, qui est responsable du développement de nombreuses maladies incurables chez l'homme, y compris le lymphome, qui a servi de modèle pour l'étude. Pour mener à bien cette étude, une base de données de structure-activité identifiée par la méthode de criblage à haut débit *in vitro* appliqué sur une série de 40 petites molécules biologiquement actives en tant qu'inhibiteurs potentiels de la protéase MALT1 et de la lignée cellulaire de lymphome a été ciblée. À cet égard, un modèle hybride 2D/3D-QSAR basé sur divers descripteurs moléculaires 0D-4D (constitutionnels, topologiques, physico-chimiques, réactivité basée sur la DFT, et CoMFA&CoMSIA) est élaboré. Ceci a pour but d'explorer structurellement les propriétés et les caractéristiques favorables à l'optimisation de la relation structure-activité de cette classe de molécules. Grâce aux orientations et aux prédictions du modèle 2D/3D-QSAR assisté par des visualisations 3D des surfaces d'affinité/compatibilité du système protéine-ligand, la relation structure-activité de ces molécules est décodée et validée par le renforcement de l'activité biologique de quatre nouvelles molécules de pyridine (P1-P4) en comparaison de la conformation de référence active D40 et de l'inhibiteur allostérique standard, la Thioridazine. Pour examiner la pertinence des molécules proposées en tant que clés structurelles pour la conception et la découverte de médicaments, leur biodisponibilité (modèle de type Drug-like) et leurs propriétés pharmacocinétiques et pharmacodynamiques (modèle de type ADME-Tox) ont fait l'objet d'une évaluation rigoureuse *in silico*. Les résultats de ces évaluations ont abouti à la sélection des molécules P2 et P3, en tant que meilleures candidates 'Hit to lead' (H2L) pour la conception précoce de nouveaux médicaments. Ensuite, pour examiner la compatibilité de ces molécules en tant qu'inhibiteurs allostériques spécifiques de la protéase MALT1, nous avons généré des profils de compatibilité/affinité protéine-ligand à l'aide de modèles de Docking de molécules semi-flexibles et de Prime/MM-GBSA. Les résultats obtenus en termes de profils d'interaction protéine-ligand, d'énergies de liaison (EL), de constantes d'inhibition estimées (Ki) et d'énergies de liaison libres de Prime/MM-GBSA (ΔG_{bind}) ont abouti à une forte discrimination spécifique pour P2 et P3 par rapport à la molécule de référence D40 et à la Thioridazine allostérique standard inhibiteur de MALT1 (code

PDB : 4I1R). Pour vérifier la validité des résultats obtenus, nous avons soumis les systèmes biomoléculaires modélisés dans cette étude à des évaluations computationnelles post-Docking de grande précision. Le modèle de post-Docking a consisté à soumettre les systèmes MALT1-P2 et MALT1-P3 et les systèmes de référence MALT1-Thioridazine et MALT1_libre à des simulations de dynamique moléculaire associées à l'évaluation des énergies libres de liaison totales générées à partir des calculs Thermal/MM-GBSA et de l'analyse des propriétés thermodynamiques. Cette étude rigoureuse a permis de confirmer et valider la stabilité idéale de P2 et P3 dans la poche allostérique active de MALT1. À cet égard, la famille d'acides aminés contenue dans la poche allostérique active de MALT1 présentait la plus grande affinité et la plus grande stabilité avec les nouveaux inhibiteurs P2 et P3 et la thioridazine de référence. Cette poche active est constituée principalement : **Acide glutamique** (Glu397), **Tryptophane** (Trp580), **Alanine** (Ala394), **Valine** (Val381, Val344), **Leucine** (Leu401), **Arginine** (Arg576), **phénylalanine** (Phe398), **Méthionine** (Met717), **Isoleucine** (Ile715), **Acide Aspartique** (Asp342), **Asparagine** (Asn393, Asn577), et **Glutamine** (Gln579), **Isoleucine** (Ile712), **Lysine** (Lys582, Lys379), **Histidine** (His720), **Acide Aspartique** (Asp716). Les conclusions de cette étude confirment les propriétés stables des nouveaux petits dérivés allostériques à base de pyridine (LO) P2 et P3, considérés comme de bons candidats à l'utilisation en tant qu'inhibiteurs médicamenteux spécifiques de la MALT1. Les résultats de cette étude confirment les propriétés stables des nouveaux petits (LO) dérivées de pyridine (P2 et P3) modélisés ; ils peuvent être de bons candidats pour inhiber spécifiquement MALT1 dans la lutte contre les lymphomes. Ces molécules, qui ont une forte affinité avec la poche allostérique active de MALT1, peuvent être appliquées pour développer de nouvelles thérapies contre diverses pathologies causées par l'activité anormale de cette enzyme et ainsi contribuer à leur résolution. En résumé, la stratégie CADD hybride rationalisée et appliquée dans cette étude a conduit à l'introduction de concepts *in silico* très utiles pour renforcer les étapes de la découverte et de la conception d'un nouveau médicament. Par conséquent, l'applicabilité de cette approche, dans sa globalité ou dans sa spécificité, à l'étude des bases de données moléculaires et à leur évaluation et amélioration *in silico*, sera utile pour découvrir un mélange diversifié de molécules médicamenteuses computationnelles. C'est ce que nous allons aborder dans l'étude suivante en ciblant une base de données de petites molécules biologiquement actives vis-à-vis de la protéase $3CL^{pro}$ proposée pour la thérapie ciblée contre le SRAS-CoV-2 en tant que modèle.

B. Modélisation d'inhibiteurs non-covalents spécifiques de la protéase 3CL^{pro} basés sur le 9,10-Dihydrophénanthrène pour la thérapie ciblée contre le SARS-CoV-2 (COVID-19)

1 Introduction

La crise du coronavirus associé au syndrome respiratoire aigu sévère 2 (SARS-CoV-2) a provoqué d'immenses bouleversements dans le monde entier, tout comme la maladie à coronavirus 2019 (COVID-19) qui en a découlé. La prévention de la propagation de la maladie a fait l'objet de nombreuses interventions ; pas mal de vaccins ont été mis au point pour protéger la population contre ce virus [505], [506]. Cependant, le virus évolue progressivement vers de nouvelles variantes, telles que les variantes alpha (B.1.1.7), bêta (B.1.351), gamma (P.1), et delta (B.1.617.2), ce qui soulève des questions quant à l'impact sur les vaccins [171], [172]. Bien que quelques rapports de l'Organisation mondiale de la santé aient montré que les mutations ont eu un effet minime sur la vitesse de transmission [507], [508], la possibilité d'une réinfection après la vaccination reste possible [173]. C'est pourquoi pour tenter de comprendre le mécanisme de cette infection, les chercheurs explorent différentes voies possibles pour cibler certaines protéines comme la protéase de type 3C du coronavirus (3CL^{pro}) [509], la protéase de type papaïne (PL^{pro}) [510], l'ARN polymérase dépendante de l'ARN (RdRp) [511], la glycoprotéine de la pointe virale (protéine S) [512], la protéase transmembranaire sérine 2 (TMPRSS2) [513], l'enzyme de conversion de l'angiotensine 2 (ACE2) et le récepteur de l'angiotensine AT2 [514, p. 19], [515], afin de découvrir et de préparer des solutions thérapeutiques adaptées à cette infection. Dans ce contexte, notre contribution de recherche vise l'identification et l'optimisation de nouvelles petites molécules bioactives capables de détruire efficacement les processus du cycle de vie des coronavirus par l'inhibition de l'activité enzymatique anormale de la protéase à cystéine de type 3-chymotrypsine (3CL^{pro}). En effet, l'environnement enzymatique de la 3CL^{pro} convient parfaitement à la régénération des coronavirus, car elle est capable de cliver les protéines virales en fractions fonctionnelles distinctes, ce qui permet d'obtenir des informations précieuses sur l'inhibition pharmacologique de l'enzyme avec des effets optimaux contre les coronavirus [165, p. 19]. L'importance critique de l'enzyme 3CL^{pro} dans le cycle de vie du coronavirus en fait une cible thérapeutique idéale pour le développement de nouveaux agents antiviraux ayant une activité optimale contre le coronavirus [516]. L'inhibition réussie de l'activité enzymatique de la 3CL^{pro} peut réduire le potentiel de résistance aux médicaments associé aux variantes de COVID-19 [517]. Le mode d'inhibition de la 3CL^{pro} consiste à utiliser des agents spécifiques qui se lient à la zone active de l'enzyme, la bloquant et l'empêchant de couper les protéines. Cette inhibition fonctionne pour prévenir et retarder la progression du coronavirus, car elle empêche sa réplication. Pour ce faire, la FDA a approuvé différentes classes d'inhibiteurs qui inhibent spécifiquement l'activité de 3CL^{pro}, telle que la protéase, la transcriptase inverse et l'entrée du virus dans les cellules. Les inhibiteurs prennent diverses formes, notamment orale, intraveineuse et topique. Cependant, ils sont tous susceptibles de présenter des limitations telles que des interactions médicament-médicament possibles, une toxicité pour le foie et le développement de résistance. Par conséquent, le mode d'inhibition de la 3CL^{pro} peut être un outil précieux pour la prévention et le

traitement des infections à coronavirus, à prendre avec des précautions. À ce jour, il y a deux classes d'inhibiteurs de la 3CL^{pro} pour le SRAS-CoV-2, les inhibiteurs covalents et non-covalents [518]. Leur mode d'inhibition varie en fonction des mécanismes d'interaction entre les ligands du médicament et les sites actifs de résidus d'acides aminés dans la poche du 3CL^{pro} [519], [520]. À ce jour, parmi les nombreux inhibiteurs covalents de la 3CL^{pro}, seuls le Nirmatrelvir (PF-07321332) [521], PF-00835231 [522, p. 19], et l'inhibiteur de type peptide (N3) [523] ont atteint le stade des essais cliniques (**Figure V. B1a**) [524], [525], [526], [527]. D'autre part, plusieurs petites molécules ont été signalées comme étant des inhibiteurs non-covalents pour 3CL^{pro}, tels que la sciadopitysin [528], le 23R (Jun8-76-3A) [529], le CCF981, [530], Perampanel à base de pyridine [531], et le médicament Antabus (DSF) [532](**Figure V. B1b**).

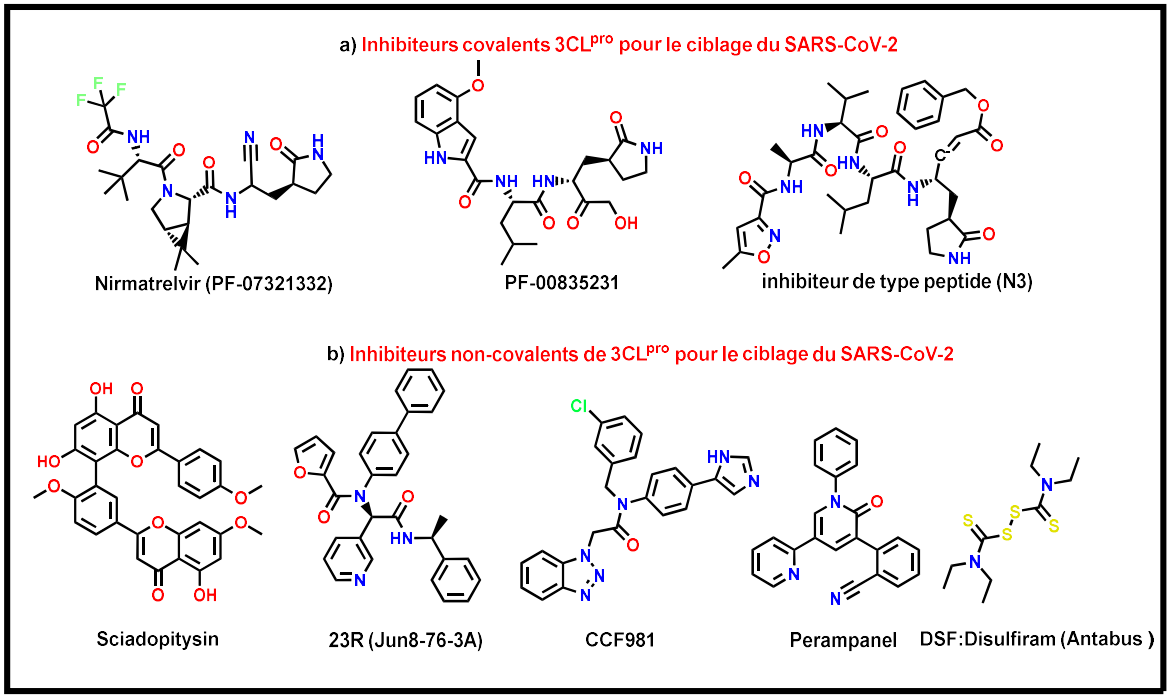


Figure V. B1. Récents inhibiteurs covalents (a) et non-covalents (b) de 3CL^{pro} destinés au SARS-CoV-2 dans le cadre de la recherche d'un médicament anti-COVID19.

Malgré tous ces efforts, la plupart des inhibiteurs covalents et non covalents de la 3CL^{pro} du SRAS-CoV-2 signalé présentent des propriétés similaires à celles des médicaments et une pharmacocinétique/pharmacodynamique ne sont pas convaincantes pour pouvoir être utilisées de manière appropriée (présentent une toxicité potentielle) ; l'utilisation d'inhibiteurs peptidiques covalents exacerbe les effets secondaires néfastes de l'utilisation clinique de ces médicaments [533], [534]. Il est donc devenu plus nécessaire encore d'explorer la piste de nouveaux inhibiteurs non-covalents de la 3CL^{pro} ayant des propriétés pharmaceutiques favorables, une bonne activité inhibitrice contre la COVID-19 et une bonne efficacité thérapeutique. Dans ce contexte, des études approfondies sont exigées pour développer des agents médicamenteux plus sûrs et plus efficaces pour la thérapie ciblée contre le SRAS-CoV-2 par inhibition de l'activité enzymatique de 3CL^{pro}. À cet égard, Zhang et al. [535, p. 19] ont récemment démontré par un modèle réalisé dans le cadre des évaluations d'activités biologique inhibitrice *in vitro* pou d'une chimiothèque de 46 petites molécules synthétiques portant l'échafaudage 9,10-Dihydrofenanthrène qui permet d'inhiber le SRAS-CoV-2-3CL^{pro} (**Tableau V. B1**). Cependant, plusieurs aspects clés dans le processus de conception et de découverte de nouveaux

médicaments ont été omis dans cette étude, notamment les évaluations des propriétés similaires à un médicament, la pharmacocinétique et la pharmacodynamique, les mécanismes d'interaction (Protéine-ligand), la compatibilité Protéine-ligand (les systèmes $3CL^{pro}$-Inhibiteurs) et d'autres paramètres d'équilibre et de stabilité dans des systèmes biomoléculaires géants protéine-ligand. Cela nous a incités à mener une étude de modélisation computationnelle à grande échelle *in silico* à l'aide de techniques bio-informatiques afin de renforcer la pertinence de l'échafaudage 9,10-Dihydrophénanthrène dans la conception et l'optimisation de nouveaux inhibiteurs non-covalents de la $3CL^{pro}$ ayant des propriétés pharmaceutiques/pharmacodynamiques favorables et une activité biologique inhibitrice dominante contre COVID-19. C'est ce que nous comptons accomplir dans ce volet des applications qu'ils ont exposé dans cette thèse. La **Figure V. B2** présente l'échafaudage du 9,10-Dihydrophénanthrène que nous avons adopté comme clé structurelle pour concevoir et modéliser de nouveaux inhibiteurs non-covalents de $3CL^{pro}$ qui pourraient être appliqués à la thérapie ciblée contre COVID-19. Dans ce contexte, nous avons pris l'inhibiteur covalent peptidyl (N3) comme point de référence covalent pour comparer la compatibilité structurelle des inhibiteurs non-covalents basés sur le 9,10-Dihydrophénanthrène avec celle de l'inhibiteur covalent peptidyl N3. Nous avons choisi cet inhibiteur comme référence car plusieurs études récentes ont confirmé que le peptidyl Michael N3 est capable d'inhiber l'enzyme $3CL^{pro}$ du Sars-CoV-2 et de réduire la réplication du virus [166], [536]. Parallèlement, de nombreuses études confirment que cette molécule peptidique est efficace pour inhiber les enzymes d'autres coronavirus, comme le SRAS-CoV-1. Par conséquent, N3 est en cours de développement pour son potentiel en tant que médicament pour traiter le COVID-19 via l'inhibition de l'enzyme $3CL^{pro}$ du SRAS-CoV-2. Dans le même cadre, nous avons considéré le **Disulfiram** (**DSF**), produit approuvé par la FDA (Food and Drug Administration) américaine pour le traitement de l'alcoolisme, au même titre qu'un inhibiteur non-covalent de $3CL^{pro}$ en tant que référence de contrôle. Il est apparu que le **DSF**, à base de petites molécules, possède des propriétés antivirales, inhibant la protéase M^{pro} et l'excrétion du zinc, et des effets anti-inflammatoires, inhibant la sécrétion de cytokines inflammatoires induite par le NF-kB et le NLRP, en relation avec le SRAS-CoV-2 [537]. Sur la base de ces orientations, nous avons sélectionné l'inhibiteur **N3** et le médicament DSF comme références non-covalent inhibiteur témoins dans le cadre de notre étude computationnelle afin de rationaliser le criblage et la modélisation de nouveaux inhibiteurs de la $3CL^{pro}$ plus efficaces et moins toxiques, basés sur de petites molécules à base de 9,10-dihydrophénanthrène.

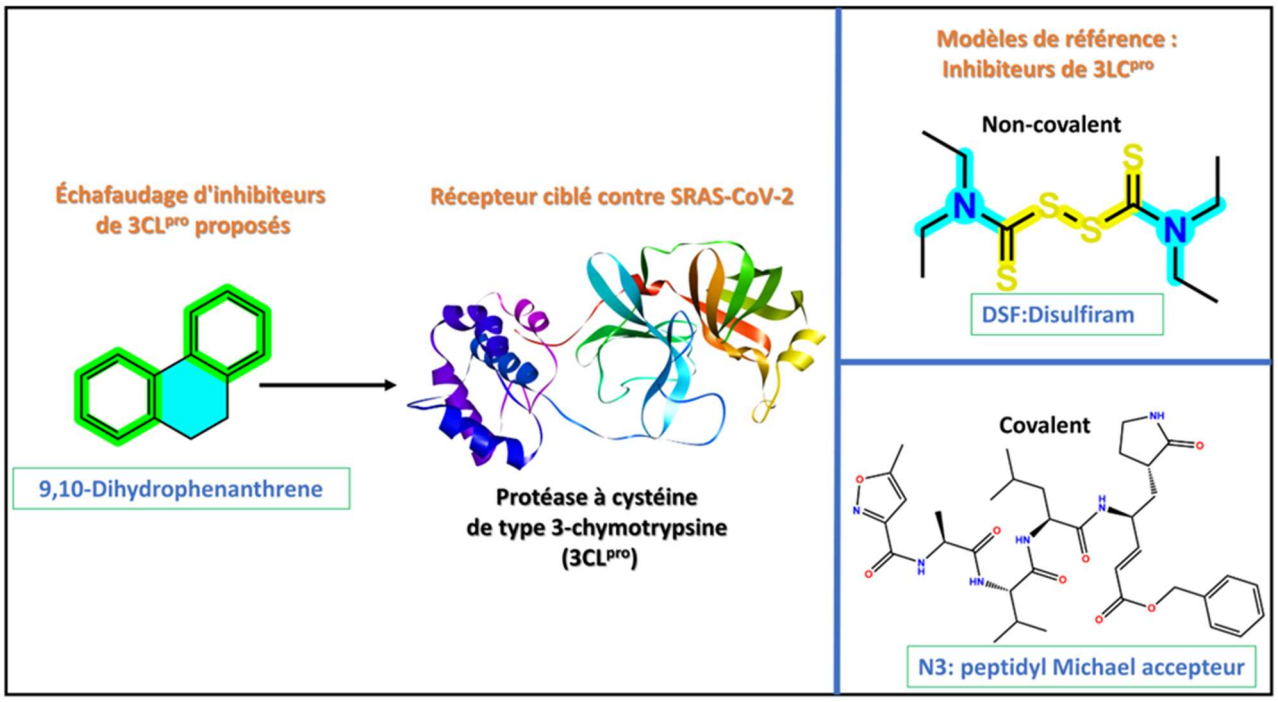
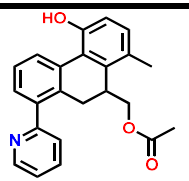
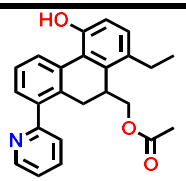
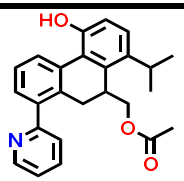
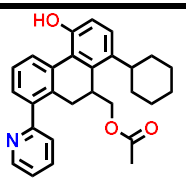
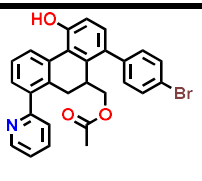
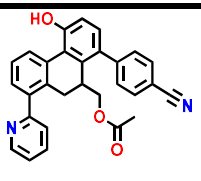
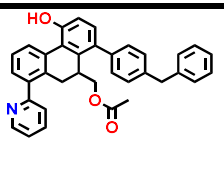
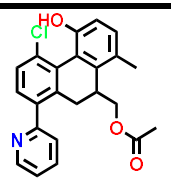
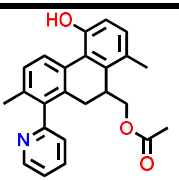
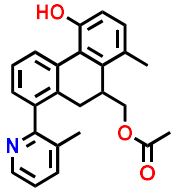
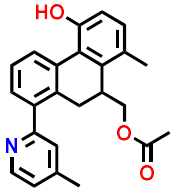
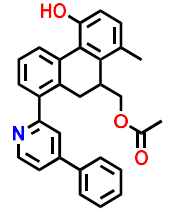
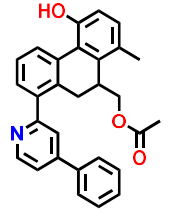
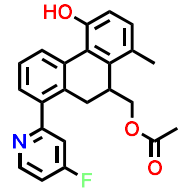
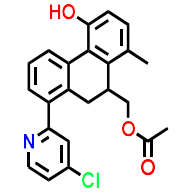
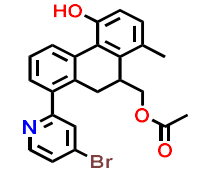
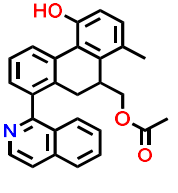
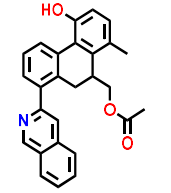
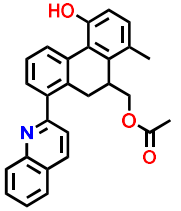
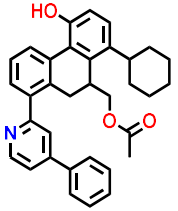
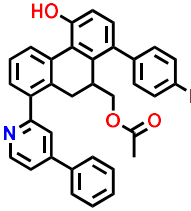
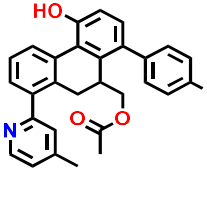
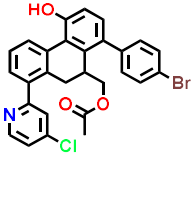
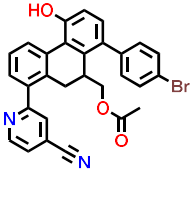
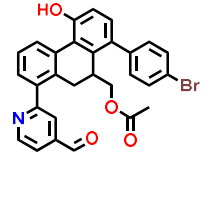
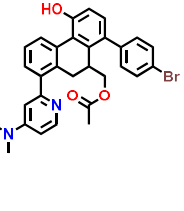
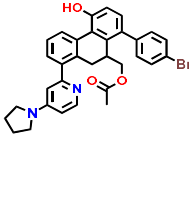
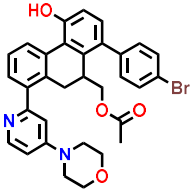
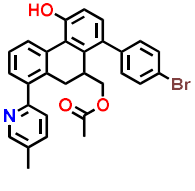
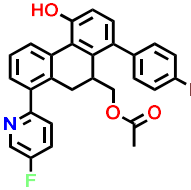
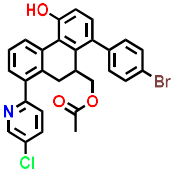
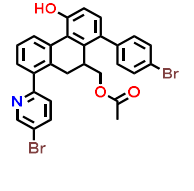
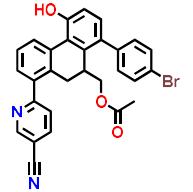
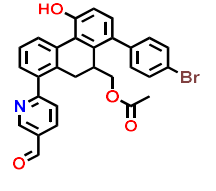
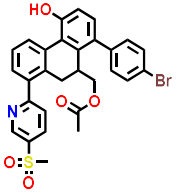
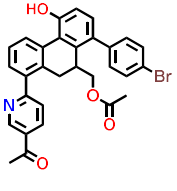
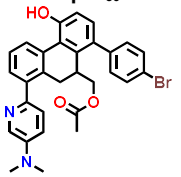
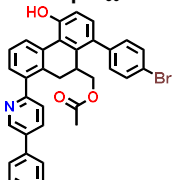
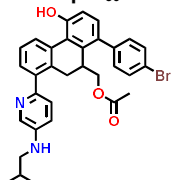
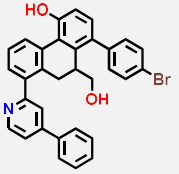


Figure V. B2. 9,10-Dihydrophénanthrène : échafaudage candidat pour le développement de nouveaux inhibiteurs de SRAS-CoV-2-3CL^{pro} pour la thérapie de COVID-19.

2 Stratégie de conception de médicaments assistée par ordinateur développée et appliquée

Conformément à l'importance critique des petites molécules biologiquement actives dans la conception et la découverte de nouveaux médicaments. Dans cette étude, nous appliquons la méthodologie CADD hybride décrite à la **Figure V. B3** pour le criblage virtuel et la modélisation biomoléculaire de nouveaux inhibiteurs non-covalents de SARS-CoV-2 3CL^{pro} sur la base d'un petit échafaudage moléculaire, à savoir 9,10-Dihydrophénanthrène. Cela afin de renforcer l'arsenal de petites molécules computationnelles candidates à des applications pharmaceutiques apparues au cours des études présentées dans cette thèse. Toutes les précisions relatives aux méthodes et outils utilisés pour mettre en œuvre chacune des parties de cette démarche sont décrites et détaillées dans le deuxième chapitre de cette thèse, ainsi que dans la version électronique de l'article que nous avons publié à ce sujet [309].

Tableau V. B1. Activité inhibitrice *in vitro* des dérivés du 9,10-dihydrophénanthrène (D1-D46) vis-à-vis de la $3CL^{pro}$ du SARS-CoV-2.

 D1 pIC_{50} =4.21	 D2 pIC_{50} =4.48	 D3 pIC_{50} =4.53	 D4 pIC_{50} =5.04	 D5 pIC_{50} =5.19	 D6 pIC_{50} =4.71	 D7 pIC_{50} =4.94	 D8 pIC_{50} =4.07	 D9 pIC_{50}=4.24
 D10 pIC_{50} =4.27	 D11 pIC_{50}=4.23	 D12 pIC_{50}=4.17	 D13 pIC_{50}= 5.25	 D14 pIC_{50}= 4.18	 D15 pIC_{50}= 4.45	 D16 pIC_{50}=4.73	 D17 pIC_{50}=4.85	 D18 pIC_{50}= 4.92
 D19 pIC_{50}= 4.86	 D20 pIC_{50}= 5.07	 D21 pIC_{50}=5.61	 D22 pIC_{50}=5.01	 D23 pIC_{50}= 5.32	 D24 pIC_{50}= 4.99	 D25 pIC_{50}= 5.48	 D26 pIC_{50}=5.00	 D27 pIC_{50}=4.95
 D28 pIC_{50}= 5.13	 D29 pIC_{50}= 5.28	 D30 pIC_{50}= 5.37	 D31 pIC_{50}=5.57	 D32 pIC_{50}=5.17	 D33 pIC_{50}= 5.56	 D34 pIC_{50}= 5.05	 D35 pIC_{50}= 5.18	 D36 pIC_{50}=5.23
 D37 pIC_{50}=5.44	 D38 pIC_{50}= 5.08	 D39 pIC_{50}= 5.48	 D40 pIC_{50}= 5.81	D41 pIC_{50}=5.74	D42 pIC_{50}=4.99	D43 pIC_{50}= 5.27	D44 pIC_{50}= 5.33	D45 pIC_{50}= 4.97
D46 pIC_{50}=5.26	DSF pIC_{50}=5.98	N3 pIC_{50}=5.26 [524]				-		

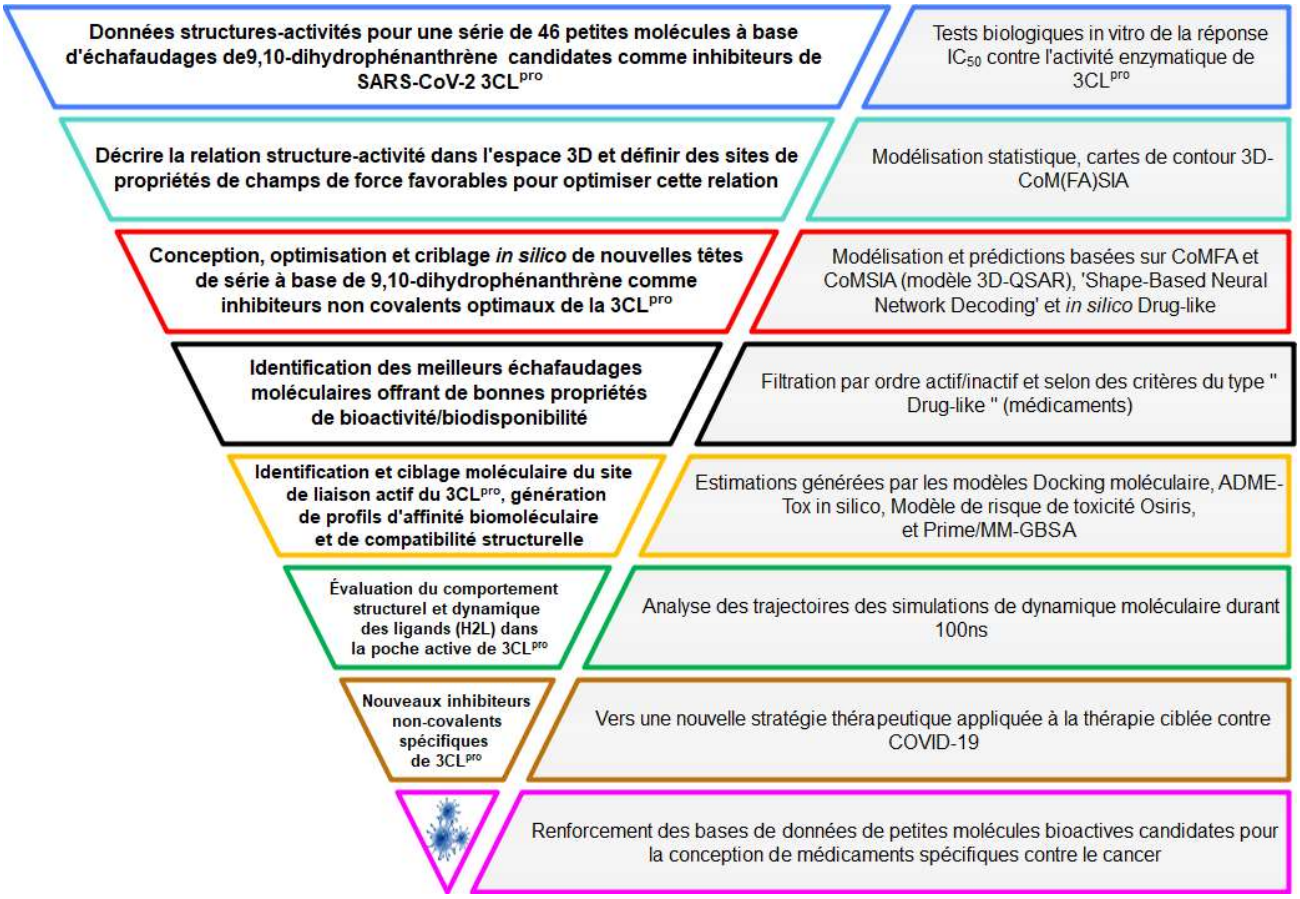


Figure V. B3. Stratégie de CADD hybride proposée et appliquée pour la conception, criblage, la caractérisation et l'optimisation de nouveaux inhibiteurs non-covalents de SARS-CoV-2 3CL^{pro} à base de 9,10-Dihydrophénanthrène, candidates à la conception d'un nouveau médicament Anti-COVID19.

3 Résultats et discussion

La présente section aborde et discute les principaux résultats obtenus dans le cadre de cette étude. D'autres détails complémentaires qui n'affectent pas le flux de données et l'analyse et la discussion de nos résultats peuvent être tirés de la version originale de notre article publié à cet égard (Daoui et al. [309]).

3.1 Exploration des relations structure-activité par 3D-QSAR

3.1.1 Modélisation PLS avec des descripteurs moléculaires CoMFA et CoMSIA

Après la préparation, l'optimisation et l'alignement de la base de données chimique, nous utilisons l'algorithme PLS pour définir la dépendance logique entre l'activité biologique observée (pIC_{50}) et les descripteurs moléculaires 4D calculés par les techniques CoMFA et CoMSIA pour les dérivés du 9,10-dihydrophénanthrène (**Figure V. B4**). L'alignement moléculaire idéal de la base de données étudiée, mis en évidence dans la **Figure V. B4**, permet éventuellement de calculer avec précision les descripteurs de champ CoMFA et CoMSIA, ce qui permet d'élaborer un modèle 3D-QSAR robuste et pertinent. Sur cette base et selon le principe de la diversité structurelle/activité, la base de données est divisée en deux ensembles. Le premier, destiné à l'entraînement et à la construction du modèle 3D-QSAR, comprend 37 composés, et le second, destiné à tester son pouvoir prédictif externe, comprend

9 composés. Le **Tableau V. B2** présente les scores des significations statistiques observées pour le meilleur modèle 3D-QSAR développé à partir des descripteurs moléculaires de CoMFA et de CoMSIA dans cette étude.

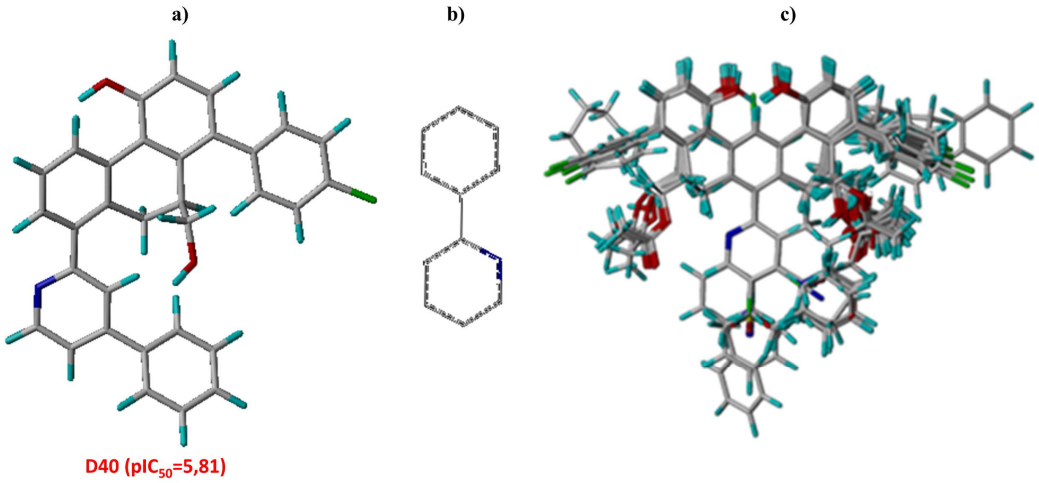


Figure V. B4. (a) conformation de référence D40, (b) base de sous-structure commune, (c) base de données alignée.

Tableau V. B2. Bilan des performances en matière de signification statistique du modèle 3D-QSAR de type CoMFA et CoMSIA.

Modèle 3D-QSAR	Paramètres statistiques							Fractions de champs				
	Q^2	R^2	SEE	F-test	N	R^2_{pred}	$^cR^2_p$	S	E	H	D	A
CoMFA/ SE	0.81	0.97	0.105	107.50	3	0.95	0.71	0.413	0.587	-	-	-
CoMSIA /SEHDA	0.76	0.94	0.131	67.45	4	0.91	0.65	0.135	0.299	0.275	0.079	0.212

Q^2 est le coefficient de corrélation de validation croisée (loocv), **R^2** est le coefficient de corrélation de validation non croisée, **SEE** est l'erreur standard d'estimation, **F** est la valeur du F-test, **N** est le nombre optimal de composantes défini par PLS-loocv, **R^2_{pred}** est le coefficient de corrélation de prédiction obtenu par validation externe, **$^cR^2_p$ est** le critère d'évaluation de la stabilité du modèle 3D-QSAR original après le test de Y-randomisation. **Fractions** est la contribution des différents champs stérique (S), électrostatique (E), hydrophobe (H), donneur de liaison hydrogène (D), et accepteur de liaison hydrogène (A).

Le modèle 3D-QSAR de CoMFA et CoMSIA décrit dans le **Tableau V. B2** démontre une forte capacité à expliquer la relation quantitative entre les descripteurs de champs (S, E, H, D, et A) et l'activité inhibitrice des dérivés du 9,10-dihydrophénanthrène contre le SARS-CoV-2-3CL^{pro}. Les modèles sont évalués à l'aide de diverses mesures, notamment les coefficients de détermination (R^2=0,97 et R^2=0.94), la validation interne par (Loocv) (Q^2=0.81 et Q^2=0.76), les tests de Fischer (F=107.50 et F=67.45) et le test de randomisation en Y ($^cR^2_p$=0.71 et $^cR^2_p$=0.65). De plus, le test externe a indiqué des coefficients de corrélation élevés entre les activités observées et prédites (R^2_{pred}=0.95 et R^2_{pred}=0.91) avec des valeurs SEE faibles (SEE=0.105 et SEE=0.131) pour les modèles CoMFA et CoMSIA, respectivement. Ces résultats suggèrent que les modèles CoMFA et CoMSIA proposés sont très précis dans la prédiction de l'activité inhibitrice des dérivés du 9,10-dihydrophénanthrène contre le SARS-CoV-2-3CL^{pro}. Les résultats de l'évaluation PLS suggèrent également que l'activité inhibitrice des dérivés du 9,10-dihydrophénanthrène contre le SARS-CoV-2-3CL^{pro} est influencée de manière significative par leurs propriétés électrostatiques (E), stériques (S), hydrophobes (H) et d'accepteur de liaison hydrogène (A). Cela est confirmé par l'effet important des descripteurs de champ moléculaire CoMFA/SE et CoMSIA/SEHDA indiqués en tant que fractions de contribution. Plus particulièrement, 58

% et 41.3 %, ainsi que 30 % et 27.5 % de la performance prédictive de l'activité inhibitrice sur la $3CL^{pro}$ sont attribués aux champs E et S respectivement selon les rapports des descripteurs des champs moléculaires CoMFA et CoMSIA. En outre, pour le modèle CoMSIA, les champs D et H ont contribué à 21.2 % et 13.5 %, tandis que les champs A ont contribué à 7.9 %. Par conséquent, le modèle 3D-QSAR développé sur la base des descripteurs CoMFA/SE et CoMSIA/SEHDA peut être utilisé pour prédire avec précision l'activité biologique inhibitrice de petites molécules de 9,10-dihydrophénanthrène vis-à-vis de l'activité enzymatique de la $3CL^{pro}$.

3.1.2 Interprétation des cartes de contours 3D issues des modèles CoMFA et CoMSIA

Le modèle 3D-QSAR généré peut être analysé sur la base des cartes de contour générées à partir des informations structure-activité décodées par les données de sortie de CoMFA et de CoMSIA. Les Figures **Figure V. B5** et **Figure V. B6** montrent les visualisations 3D de cartes de contours obtenues dans cette étude pour les modèles CoMFA et CoMSIA, respectivement.

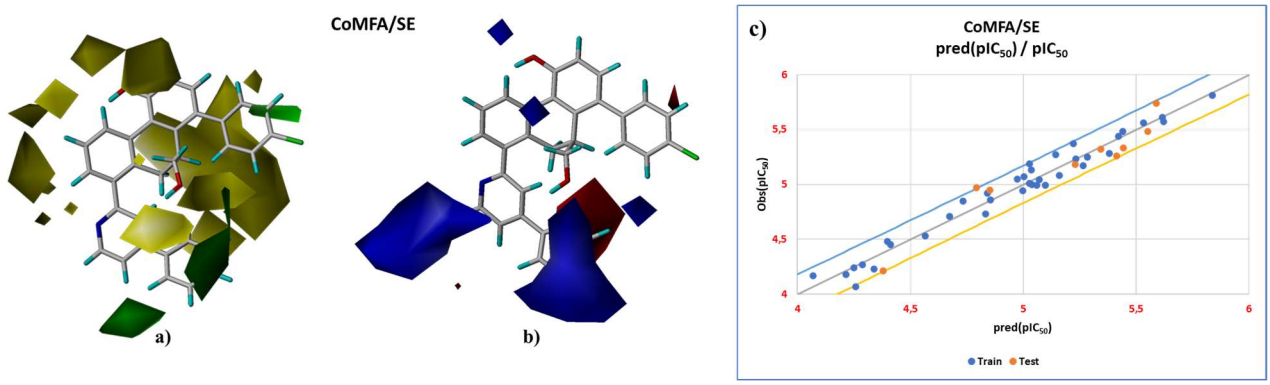


Figure V. B5. Cartes de contours 3D générées par le modèle CoMFA/SE, (a) interactions de champ stérique (vert = 80% favorable/jaune = 20% défavorable), (b) interactions de champ électrostatique (bleu = 80% favorable/rouge = 20% défavorable). (**c**) pIC_{50} observés vs pIC_{50} prédits.

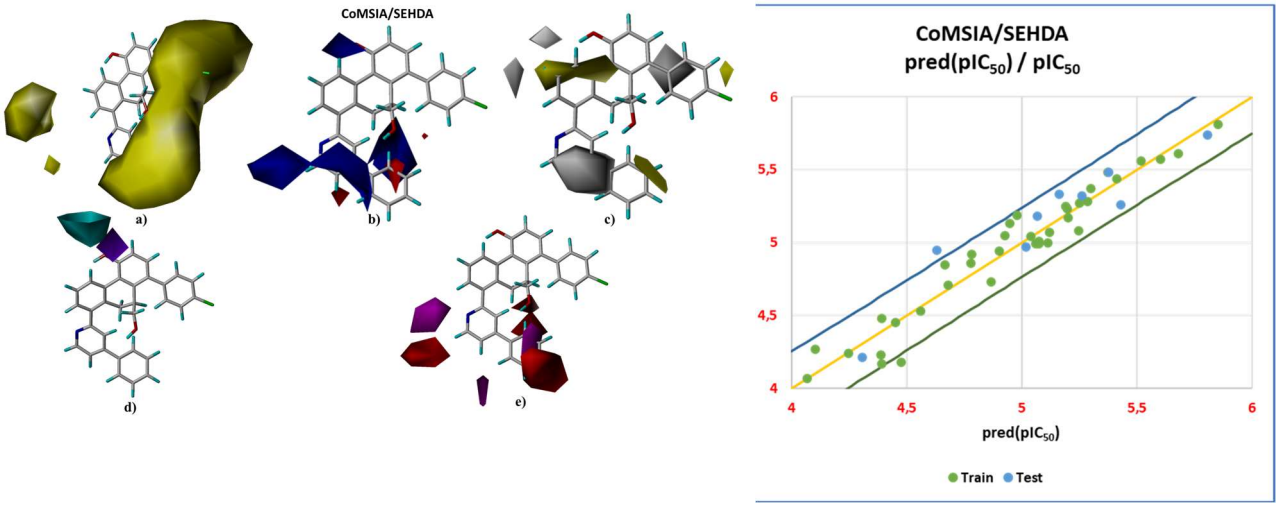


Figure V. B6. Cartes de contours 3D générées par le modèle CoMSIA, (a) interactions de champ stérique (vert = favorable/jaune = défavorable), (b) interactions de champ électrostatique (bleu = 80% favorable/rouge = 20% défavorable), (c) interactions de champ hydrophobe (jaune = 80% favorable / blanc = 20% défavorable), (d) interactions de champ de liaison hydrogène-donneur (cyan = 80% favorable / violet = 20% défavorable), (e) interactions de champ de liaison hydrogène-accepteur (magenta = 80% favorable / rouge= 20% défavorable). (f) pIC_{50} observés versus pIC_{50} prédits.

Sur la base des résultats de la **Figure V. B5(a,b)** et de la **Figure V. B6(a,b)**, nous pouvons observer une disposition spatiale similaire des contours stériques (S) et électrostatiques (E) sur la molécule modèle D40. Cela confirme que les descripteurs de type S et E exercent une influence significative sur l'activité inhibitrice vis-à-vis de l'enzyme, modélisée par la pIC_{50}. La prédominance des contours jaunes autour du D40 signifie que les grands radicaux cycliques ont eu un effet défavorable sur la bio activité des dérivés du 9,10-dihydrophénanthrène contre l'enzyme $3CL^{pro}$ du SARS-CoV-2. Pour renforcer le pIC_{50} associé aux molécules développées à partir du D40, il serait utile d'introduire des groupes fonctionnels de petite taille sur les sites couverts par les contours jaunes. En outre, la prédominance des contours bleus sur le groupement 4-phénylpyridine-2-yl indique que l'introduction de groupes donneurs d'électrons sur cet espace est susceptible de favoriser les activités biologiques des molécules dérivées du D40 contre l'enzyme SARS-CoV-2 $3CL^{pro}$. L'analyse des cartes de contours décrites dans la **Figure V. B6c** suggère que les sites C22 et C30 sont moins favorables pour la présence de groupements hydrophobes, ce qui signifie que l'ajout de groupements hydrophiles sur ces sites peut contribuer à l'amélioration de l'activité biologique inhibitrice (pIC_{50}) des composés étudiés. En outre, la **Figure V. B6e** montre que le groupe hydroxyle représente un site idéal pour l'insertion d'un radical donneur de liaison hydrogène, ce qui peut donc contribuer à renforcer l'activité biologique inhibitrice (pIC_{50}). Les corrélations linéaires observées dans les **Figures (V.5c, V.6f)** entre les valeurs observées et prédites du pIC_{50} indiquent que les propriétés de champ (S, E) et (H, D, A) des dérivés du 9,10-dihydrophénanthrène sont proportionnellement corrélées avec les valeurs de pIC_{50}. Par conséquent, les modèles pharmacophores 3D-QSAR (CoMSIA et CoMFA) peuvent être appliqués pour développer fidèlement de nouveaux dérivés de D40, plus probants, à utiliser comme inhibiteurs potentiels de la $3CL^{pro}$ du SARS-CoV-2 dans le cadre en particulier pour la thérapie proposée dans le cadre du projet COVID-19.

3.2 Modélisation générative basée sur la forme pour la conception de nouveaux médicaments

La **Figure V. B7** présente un bilan des hypothèses pharmacophores les plus favorables pour optimiser/modéliser l'activité biologique de nouvelles molécules en fonction de la structure de D40. Dans cette étude, la conformation moléculaire bioactive (D40) est utilisée comme modèle pour générer de 1 à 100 nouveaux composés présentant différents échafaudages et motifs chimiques. Des algorithmes de décodage de réseaux neuronaux basés sur la forme générative ont été employés [538], ce qui a permis de produire 96 nouveaux dérivés du 9,10-dihydrophénanthrène (DH1-DH96). Les molécules générées sont passées au tamis *in silico* qui évalue les propriétés de Lipinski (RO5), les violations de Ghose, la biodisponibilité et l'accessibilité de la synthèse. Par conséquent, un nouvel ensemble de composés aux propriétés potentiellement similaires à celles des médicaments (modèle de type Drug-like) est produit sur la base de la structure D40. Les résultats de la génération et du criblage de type Drug-like indiquent que 47 nouvelles molécules peuvent remplacer la conformation bioactive de référence (D40) et l'inhibiteur N3 en tant que médicament administré par voie orale contre l'activité enzymatique de la $3CL^{pro}$ SARS-CoV-2. Ces nouvelles molécules ont une bonne biodisponibilité (0.55) et une bonne accessibilité synthétique (2<SA<4) et ne violent ni les règles de

Lipinski ni celles de Ghose. D40 et N3, quant à elles, ont une faible biodisponibilité (0.17) et une faible accessibilité synthétique (4.07 et 6.43) et violent à la fois les règles de Lipinski (2 violations) et de Ghose (3 violations) [309].

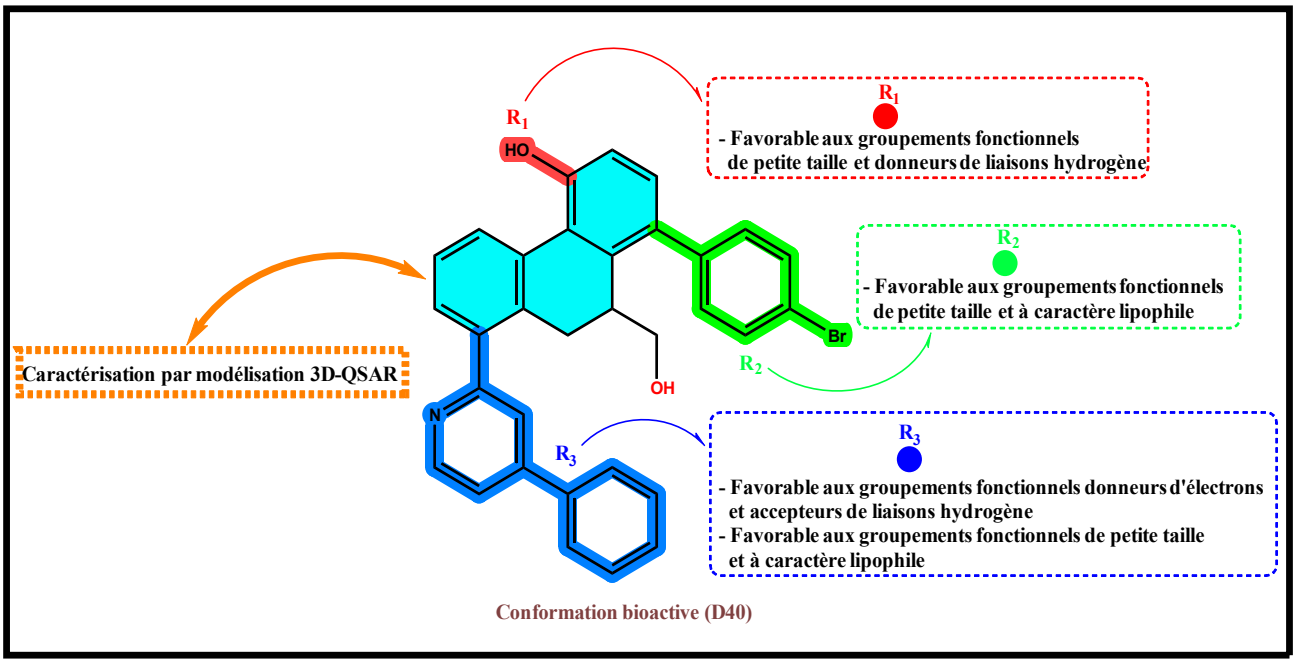


Figure V. B7. Caractérisation des propriétés structurelles favorables à la conception et à l'optimisation de nouveaux inhibiteurs du SARS-CoV-2 $3CL^{pro}$ par la structure (D40).

Pour valider davantage ces hypothèses, les modèles CoMFA et CoMSIA sont appliqués aux 47 nouvelles molécules pour prédire leur activité inhibitrice contre la $3CL^{pro}$. Après avoir préparé chaque structure 3D à l'aide de la même stratégie que celle utilisée pour la modélisation 3D-QSAR dans ce travail, les pIC_{50} des nouvelles molécules conçues sont prédits après leur alignement (**Figure V. B8**). Après l'alignement parfait entre les 47 molécules filtrées et la structure de référence présentée à la **Figure V. B8**, il est possible de déduire que les modifications structurelles du composé D40 étaient compatibles avec les propositions pharmacologiques issues de la caractérisation par CoMFA et CoMSIA. Le **Tableau V. B3** présente les activités inhibitrices prévues pour les 47 nouveaux inhibiteurs non-covalents de la $3CL^{pro}$ du SRAS-CoV-2, calculées par les modèles CoMFA et CoMSIA. Les résultats de ces modèles sont susceptibles d'être plus fiables et plus précis en termes de prédictions de l'activité des nouveaux dérivés du D40.

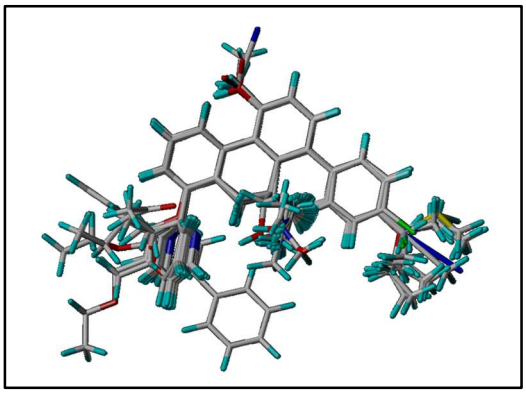


Figure V. B8. Visualisation en 3D des séquences d'alignement moléculaire entre la molécule de référence (D40) et ses nouvelles dérivées (DH1-DH96).

Tableau V. B3. Prédiction *in silico* des pIC_{50} de 47 nouveaux dérivés de 9,10-dihydrophénanthrène optimisés sur la base de la structure de référence D40 contre le SARS-CoV-2 $3CL^{pro}$.

Composé	pIC_{50} prédit *in silico*		Composé	pIC_{50} prédit *in silico*		Composé	pIC_{50} prédit in silico	
	CoMFA/ SE	CoMSIA/ SEHDA		CoMFA/ SE	CoMSIA/ SEHDA		CoMFA/ SE	CoMSIA/ SEHDA
D40	5.84	5.85	**DH25**	**6.144**	**6.878**	**DH76**	**6.300**	**6.986**
DH2	**6.302**	**6.938**	**DH25**	**6.144**	**6.878**	**DH76**	**6.300**	**6.986**
DH3	6.146	5.918	**DH26**	**6.290**	**6.913**	DH78	5.180	4.789
DH4	6.080	5.929	**DH27**	**6.149**	**6.988**	DH79	5.167	4.781
DH5	6.123	5.877	**DH28**	**6.238**	**6.858**	DH80	5.160	4.827
DH6	**6.261**	**6.967**	**DH30**	**6.105**	**6.847**	DH81	5.163	5.013
DH7	**6.187**	**6.893**	**DH38**	**6.165**	**6.736**	DH82	5.179	4.994
DH8	**6.259**	**6.852**	DH41	5.125	5.007	DH84	5.172	5.997
DH12	**6.240**	**6.826**	DH42	5.195	4.918	**DH87**	**6.221**	**6.005**
DH13	6.161	5.931	**DH43**	**6.246**	**6.860**	**DH88**	**6.268**	**6.015**
DH14	6.107	5.950	**DH44**	**6.159**	**6.724**	**DH92**	**6.239**	**6.969**
DH17	6.143	5.963	DH45	5.809	4.924	**DH94**	**6.226**	**6.850**
DH18	**6.218**	**6.857**	DH46	5.278	4.842	-	-	-
DH19	6.190	5.947	DH49	3.930	4.675	-	-	-
DH20	5.891	4.937	DH54	5.227	4.965	-	-	-
DH21	6.106	5.983	DH60	5.250	5.029	-	-	-
DH22	6.369	5.994	DH62	5.060	4.893	-	-	-
DH23	**6.104**	**6.231**	DH70	4.997	4.885	-	-	-
DH24	6.145	5.830	DH73	4.861	4.767	-	-	-

pIC_{50} observé *in vitro* [(compound **D40**: pIC_{50}=5.81), (**DSF**: pIC_{50}=5.98) et (**N3**:pIC_{50}=3.90 [524])].

Sur la base des données des valeurs prédites de pIC_{50} rapportées dans le **Tableau V. B3**, 20 nouveaux dérivés de **DH40** (D2, DH6, DH07, DH08, DH12, DH18, DH23, DH25, DH26, DH27, DH28, DH30, DH38, DH43, DH44, DH76, DH87, DH88, DH92, DH94) ont montré des activités inhibitrices *in silico* significativement plus puissantes (pIC_{50}>6) par rapport aux composés de référence (D40 avec pIC_{50}=5.8, DSF avec pIC_{50}=5.98 et N3 avec pIC_{50}=3.90). Les modifications structurelles et les substitutions effectuées sur la molécule modèle D40 se sont avérées favorables à l'amélioration de l'activité bio-inhibitrice de la nouvelle classe d'inhibiteurs non-covalents de la $3CL^{pro}$ du SARS-CoV-2. Ces 20 inhibiteurs non-covalents (**Figure V. B9**) ont le potentiel d'agir comme des agents thérapeutiques efficaces pour combattre le COVID-19. Pour étayer cette hypothèse, nous avons effectué des simulations par le Docking moléculaire afin de définir le niveau de compatibilité et les énergies d'affinité de liaison des 20 composés de tête identifiés comme inhibiteurs non-covalents de la $3CL^{pro}$ par modélisation 3D-QSAR et par criblage de type Drug-like.

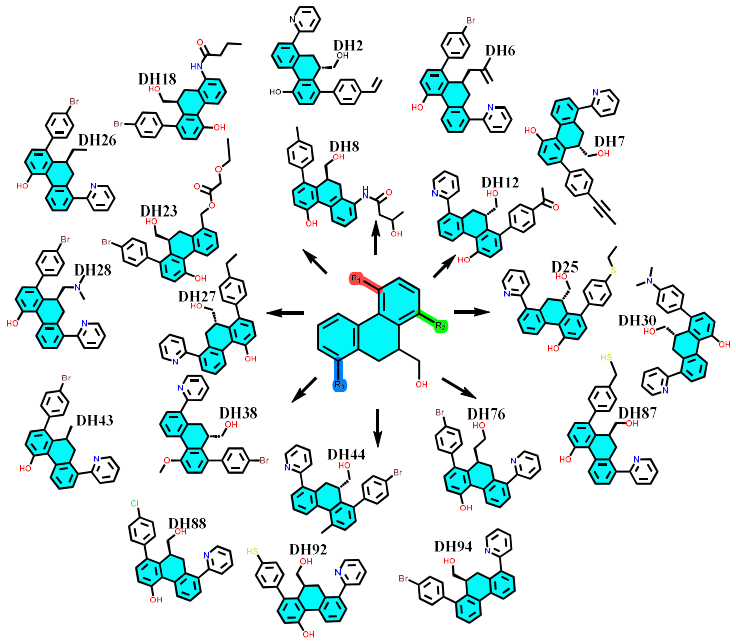


Figure V. B9. Nouveaux inhibiteurs (DH) proposés pour le SARS-CoV-2 $3CL^{pro}$ générés à partir du composé D40 et identifiés par le criblage in *silico* de 3D-QSAR et de Drug-like.

3.3 Analyse des profils de compatibilité et d'affinité Protéine-ligand

À ce stade du processus de conception virtuelle de médicaments via la méthodologie CADD hybride, nous adoptons le modèle du Docking moléculaire flexible comme procédure pour évaluer les profils de compatibilité/d'affinité structurelle des molécules étudiées (nouvelles et de références) vis-à-vis de la structure de la protéine $3CL^{pro}$ du SARS-CoV-2 (PDB : 6LU7). Dans le cadre de cette évaluation, nous avons réalisé des simulations du Docking moléculaire afin de prédire les interactions entre les sites d'acides aminés actifs dans la poche du récepteur de la $3CL^{pro}$ (code PDB : 6LU7) et les molécules médicamenteuses proposées. Nous avons sélectionné des résidus d'acides aminés qui interagissaient avec l'inhibiteur N3 de la $3CL^{pro}$ (code PDB : 6LU7) comme référence dans l'inhibition de l'activité enzymatique de la SARS-CoV-2-$3CL^{pro}$ par les petites molécules générées possédant une activité pIC_{50} élevée et des propriétés favorables semblables à celles des médicaments. Le site actif de liaison identifié dans la poche $3CL^{pro}$ contient les acides aminés suivants (**Figure V. B10**): Histidine (His41, His164, His172, His164 et His172), Méthionine (Met49, Met165), Glutamine (Gln189, Gln192), Cystéine (Cys145), Thréonine (Thr24, Thr25, Thr26, Thr190,), Alanine (Ala191), Proline (Pro168), Leucine (Leu167), Asparagine (Asn142), Acide glutamique (Glu166, Glu143), Phénylalanine (Phe140), Serine (Ser144), Glycine (Gly143), Asparagine (Asn142), Leucine (Leu141), et Acide aspartique (Asp187). Par la suite, nous avons préparé la protéine et les ligands en éliminant les molécules d'eau, en ajoutant des atomes d'hydrogène polaires et en mettant en place une grille flexible (x = -10.75 Å, y = 12.46 Å, z = 68.92 Å, espacement de grille 0.375 Å, dimension 40 × 60 × 40 $Å^3$). Nous avons effectué le Docking moléculaire en utilisant AutoDockTools-1.5.7, AutoDock vina et Discovery Studio 2016. Pour valider notre simulation, nous avons effectué une stratégie de re-Docking pour évaluer la RMSD entre l'inhibiteur N3 original et re-docké. Nous avons également évalué la stabilité du squelette du complexe $3CL^{pro}$-N3 (PDB :6LU7) dans un système aqueux par des simulations de la dynamique moléculaire à l'aide de package de logiciel Desmond (**Figure V. B11**).

3.3.1 Criblage virtuel basé sur le Docking moléculaire

Nous avons réalisé un modèle de re-Docking de l'inhibiteur covalent N3 sur la poche active du récepteur $3CL^{pro}$. Cela a permis de mieux comprendre et de valider le protocole de Docking moléculaire et de réviser les résidus d'acides aminés actifs de référence nécessaires au mécanisme d'interaction non-covalente à l'intérieur de la poche protéique de la $3CL^{pro}$ (code PDB : 6LU7). La **Figure V. B10** présente des visualisations 3D/2D à propos du processus de re-Docking de N3 dans la poche du récepteur $3CL^{pro}$.

3.3.1.1 Évaluation de la fiabilité des poses de Docking par re-Docking moléculaire

Notre modèle de Docking moléculaire s'est avéré d'une grande précision, avec DockRMSD = 0.131 Å indiquant une superposition parfaite du ligand N3 original et du ligand re-Docké. Les diagrammes d'interaction protéine-ligand 3D/2D ont montré que le site de liaison du N3 était composé des résidus His41, Met49, Gln189, Thr190, Ala191, Pro168, Leu167, Cys145, Glu166, His172, Met165, His164, Phe140 et Glu143, et que les interactions dominantes entre le ligand et la protéine étaient des

interactions hydrophobes, des liaisons hydrogène et des interactions de van der Waals. Pour valider davantage nos résultats, nous avons effectué une simulation de dynamique moléculaire pour le complexe cristallin 3CL^{pro}-N3 (PDB :6LU7) afin d'évaluer le degré de stabilité de ce complexe dans l'environnement aquatique par rapport à la poche active identifiée. Nos résultats peuvent être utilisés comme point de référence pour l'inhibition future de la 3CL^{pro} par des petites molécules à traverse des interactions non-covalentes.

3.3.1.2 Validation du modèle de Docking moléculaire par la simulation de la dynamique moléculaire

Les résultats présentés dans la **Figure V. B11a** indiquent que les variations de la RMSD sur la chaîne 3CL^{pro} étaient d'environ 1.8 Å, avec de faibles fluctuations allant jusqu'à 2.4 Å pendant 100 ns de trajectoires des simulations de la dynamique moléculaire. Les expériences de la dynamique moléculaire de l'inhibiteur N3 indiquent des déplacements considérables à partir du point de départ de la simulation avec des fluctuations entre 2.4 et 4.8 Å pour se stabiliser entre 1.2 et 3 Å après 35 ns jusqu'à la fin de la simulation. La **Figure V. B11 (b, c)** révèle que N3 a formé des interactions avec les mêmes résidus d'acides aminés que ceux prédits pour la poche de 6LU7 (3CL^{pro}-N3), avec des interactions principalement de type liaison hydrogène, hydrophobique et pont d'eau. Ces interactions étendues ont duré entre 25 et 100 % de la durée de la simulation. Ces résultats témoignent de la grande stabilité de la structure du squelette de la 3CL^{pro} complexée avec le N3 remodelé dans le cadre de la procédure de re-Docking moléculaire. Par conséquent, l'application du modèle de Docking moléculaire développé dans cette étude pour le criblage des molécules de tête candidates comme agents inhibiteurs de la 3CL^{pro} apparaît suffisamment précise et valide du point de vue computationnel.

3.3.2 Caractérisation de la sélectivité des inhibiteurs non-covalents de 3CL^{pro} obtenus par optimisation

Criblage virtuel basé sur les énergies de liaison (EL) et les constantes d'inhibition estimées (Ki)

En effectuant un criblage virtuel par Docking moléculaire, nous avons comparé l'énergie d'affinité de liaison (EL) et la constante d'inhibition (K_i) des vingt dérivés du 9,10-dihydrophénanthrène, avec celles de l'inhibiteur non-covalent de référence (DSF), puis de l'inhibiteur covalent (N3) et du composé synthétique (D40) vis-à-vis de la poche active du 3CL^{pro} du virus du SRAS-CoV-2 (**Figure V. B12**). Une fois les inhibiteurs non covalents optimaux identifiés pour cette poche active, nous avons choisis les inhibiteurs ayant les énergies de liaison les plus basses (**Tableau V. B4**).

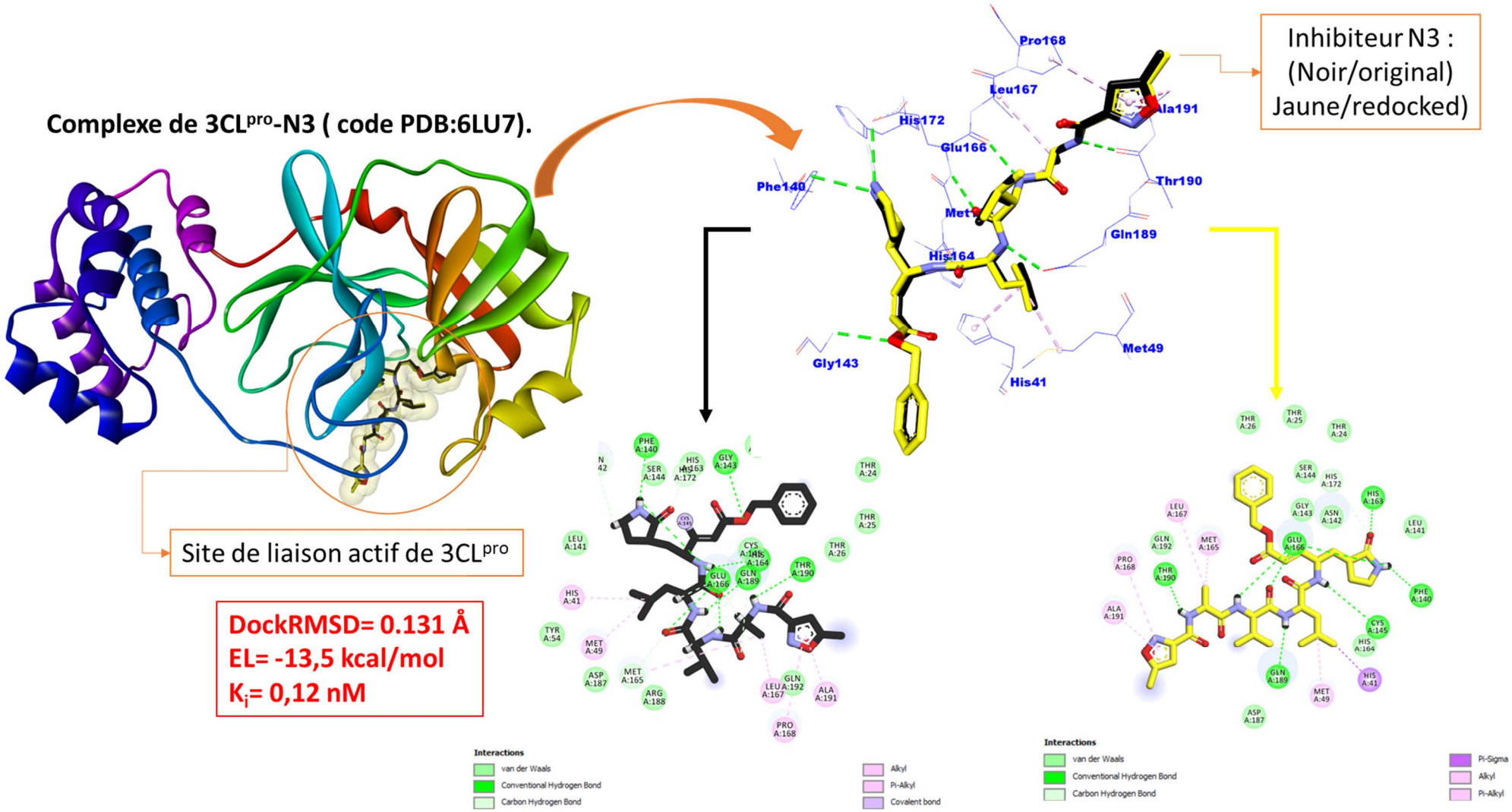


Figure V. B10. Poses de re-Docking par superposition 3D/2D de l'inhibiteur covalent N3 original (noir) et re-Docké (jaune) dans la poche active de la 3CL^{pro} (complexe 6LU7).

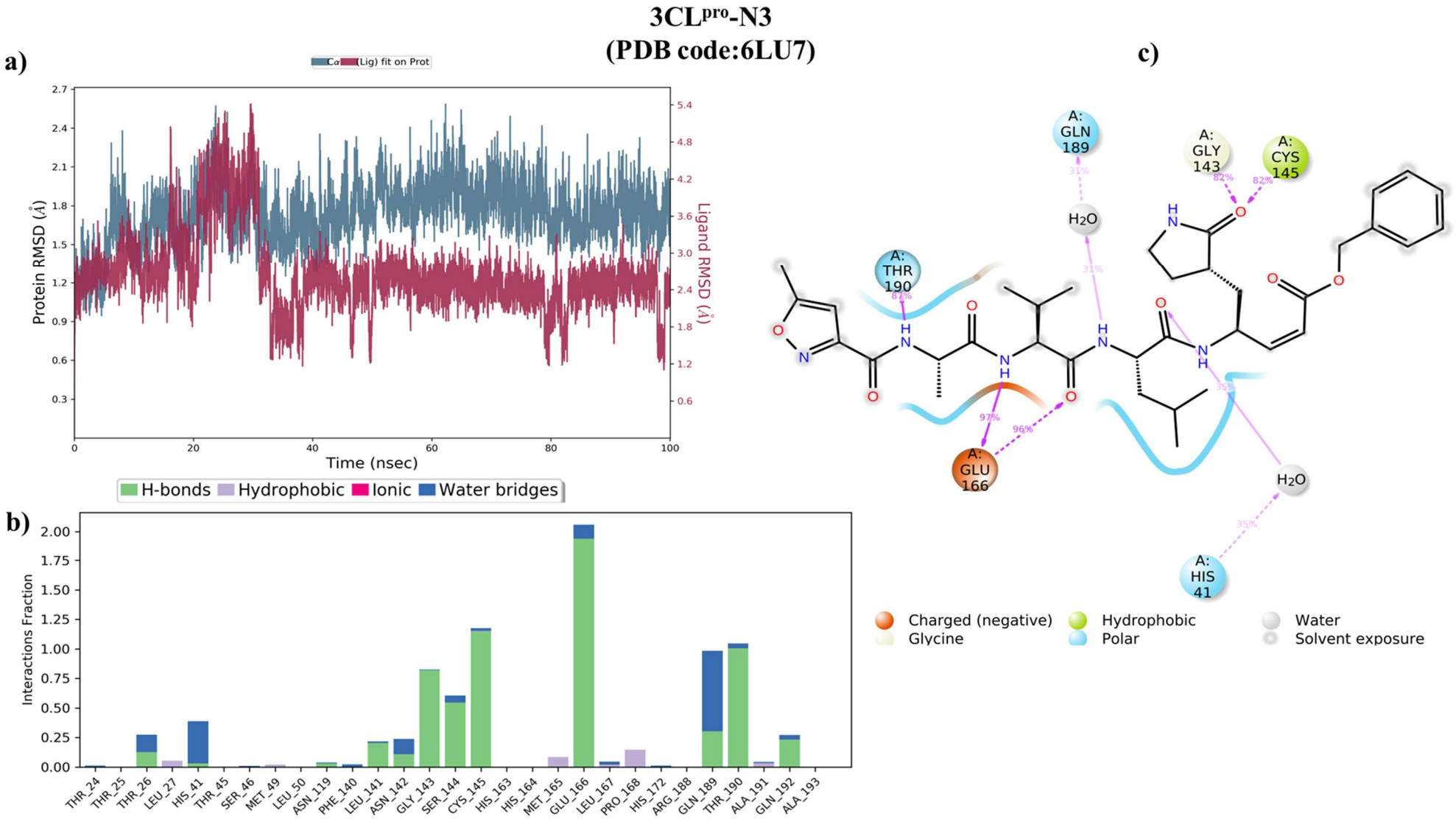


Figure V. B11. Écart quadratique moyen entre les positions atomiques de la protéine 3CL^{pro} et de l'inhibiteur N3 dans le complexe 6LU7. (b) Visualisation 2D des contacts établis entre 3CL^{pro} et N3 qui persistent au-delà de 30 % dans la simulation de la MD. (c) Histogramme des contacts entre 3CL^{pro} et N3 pendant 100 ns de la simulation de la MD.

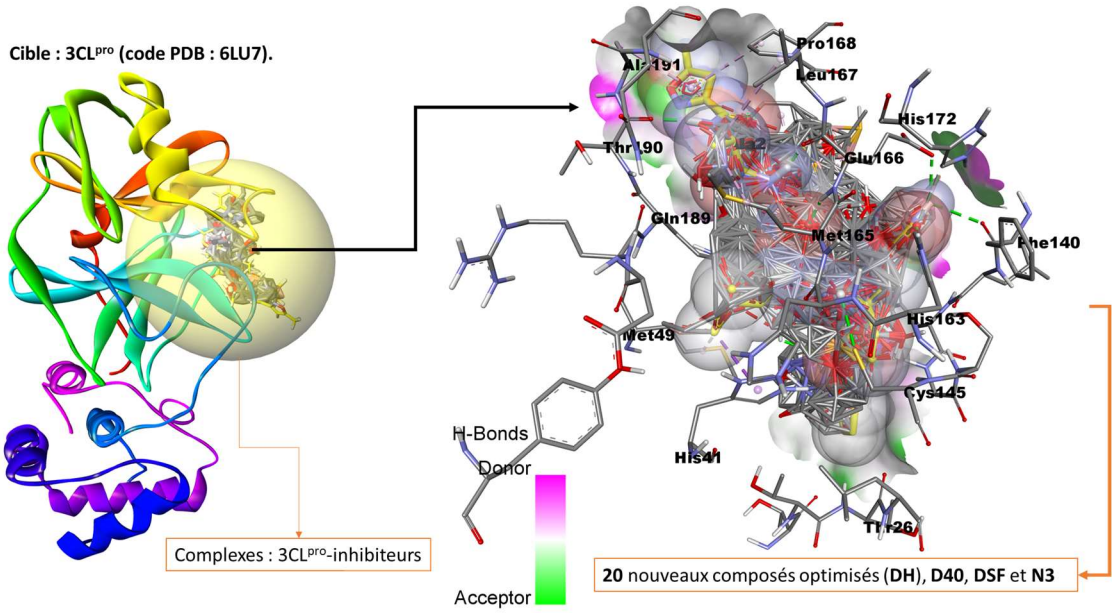


Figure V. B12. Visualisation 3D des poses du Docking moléculaire des meilleures conformations des vingt nouveaux composés DH optimisés, D40, N3 et DSF dans la poche active du SARS-CoV-2 3CL^{pro}.

La **Figure V. B12** montre la superposition des conformations les mieux classées pour les ligands étudiés par le biais du Docking moléculaire dans la poche active de la 3CL^{pro}. L'alignement parfait des inhibiteurs testés dans la poche active de la 3CL^{pro} signifie que les nouveaux petits dérivés optimisés du 9,10-dihydrophénanthrène (DH) peuvent cibler l'activité enzymatique de la 3CL^{pro} du SARS-CoV-2 et atteindre un fort potentiel d'inhibition de cette dernière par le biais des interactions non-covalentes.

Tableau V. B4. Score des énergies de liaison obtenues par le modèle de Docking moléculaire appliqué aux candidats inhibiteurs de 3CL^{pro} et aux références.

Libellé du ligand	Energie de liaison (Kcal/mol)	Constante d'inhibition estimée (nM)	Libellé du ligand	Energie de liaison (Kcal/mol)	Activité inhibitrice biologique (Ki) estimée (nM)
DH2	-9.6	90.00	**DH30**	**-10.5**	**20**
DH6	**-10.5**	**20.00**	DH38	-9.3	150
DH7	**-10.7**	**14.00**	DH43	-9.1	210
DH8	**-10.8**	**12**	DH44	-9.2	180
DH12	**-10.6**	**17**	**DH76**	**-11.1**	**7.2**
DH18	**-10.4**	**23**	DH87	-10.1	39
DH23	**-10.9**	**10**	DH88	-10.3	28
DH25	**-10.6**	**17**	DH92	-10.2	33
DH26	**-10.5**	**20**	DH94	-9	250
DH27	**-10.5**	**20**	**N3**	**-13.5**	**0.12**
DH28	-9.1	210	**DSF**	**-4.7**	**360000**
-	-	-	**D40**	**-10.2**	**33**

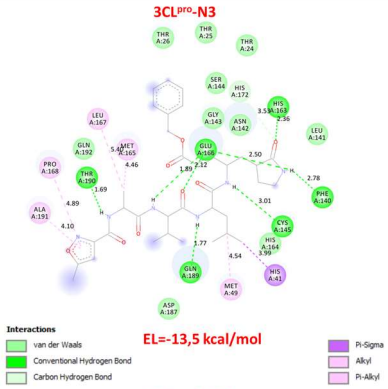
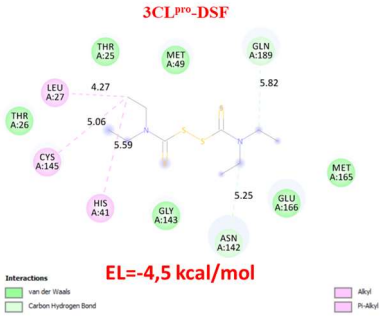
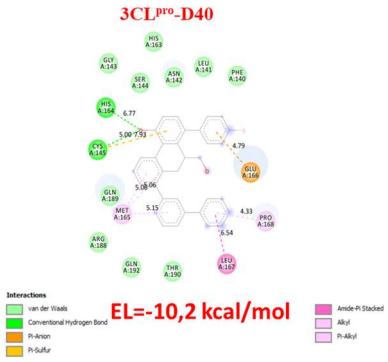
Les résultats du **Tableau V. B4** indiquent que parmi les 20 nouvelles molécules optimisées, onze petites molécules (**DH6-DH76**) présentent des énergies d'affinité de liaison plus stable que celles de la molécule de référence D40 (EL=-10.2 kcal/mol; Ki=33 nM) et de l'inhibiteur non-covalent DSF (EL=-4.7 kcal/mol; Ki=360000 nM). Ces petites molécules ont montré des caractéristiques structurelles compatibles, suggérant leur potentiel en tant que nouveaux inhibiteurs non-covalents de la 3CL^{pro}. Cependant, l'affinité de liaison estimée et les constantes d'inhibition de ces molécules DH étaient légèrement inférieures aux paramètres correspondants de l'inhibiteur covalent N3 (-13.5 kcal/mol; Ki=0.12 nM). De tels résultats confirment l'idée que les interactions non-covalentes protéine-ligand

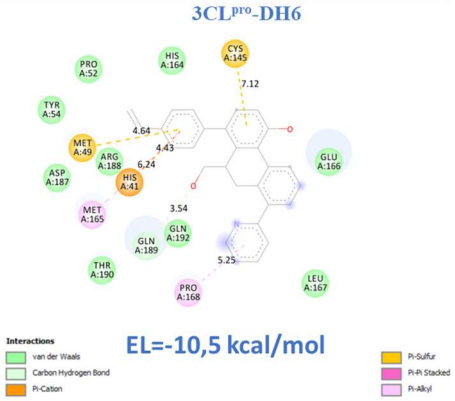
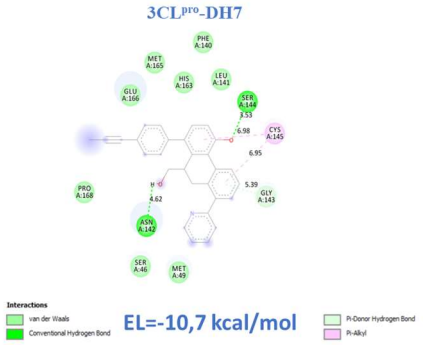
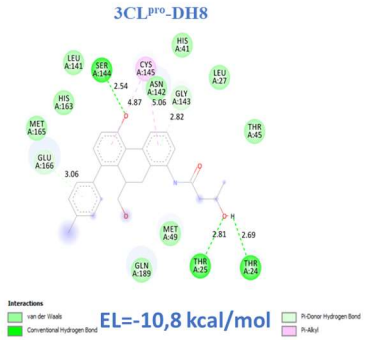
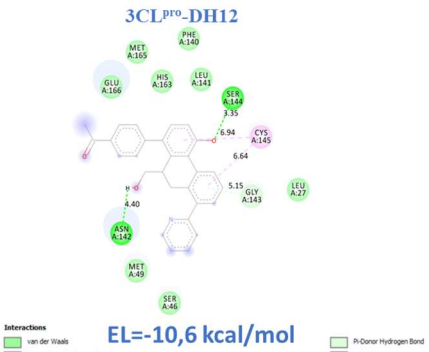
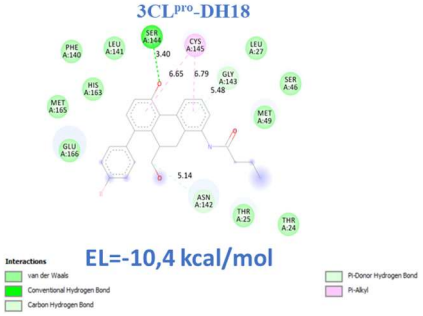
entre la 3CL^{pro} et les inhibiteurs DH pourraient être exploitées pour développer une nouvelle génération d'inhibiteurs non-covalents puissants contre la 3CL^{pro} du SARS-CoV-2, capables de concurrencer efficacement les inhibiteurs covalents tels que le modèle N3. En raison de la mauvaise action biologique prévue de l'inhibiteur DSF vis-à-vis de la 3CL^{pro}, indiquée par des indices d'énergie de liaison et de concentration biologique inhibitrice estimés par des analyses de Docking moléculaire flexible, nous excluons cette molécule en tant que référence dans les processus de criblage restent à effectuer.

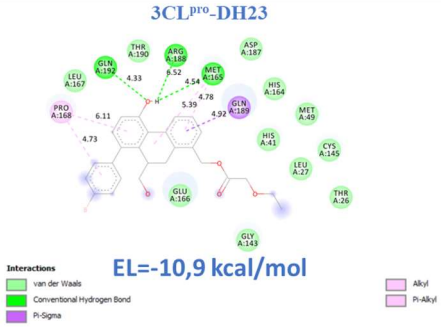
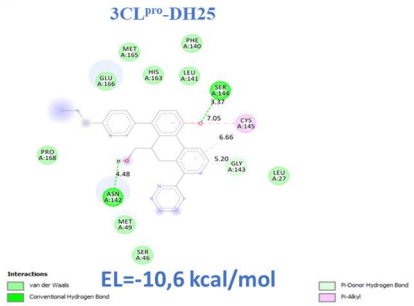
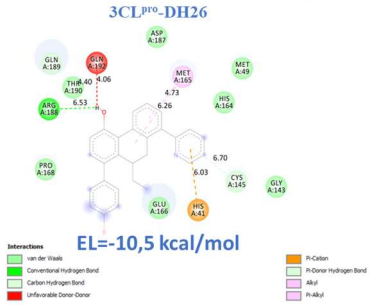
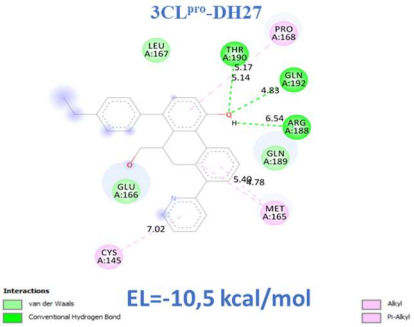
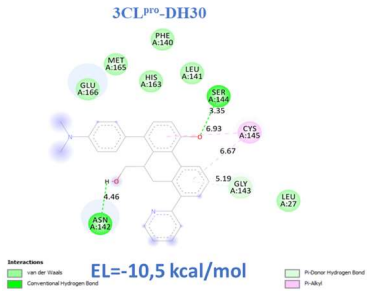
3.3.3 Sélection d'inhibiteurs prometteurs de la 3CL^{pro} basée sur l'analyse des interactions non-covalentes

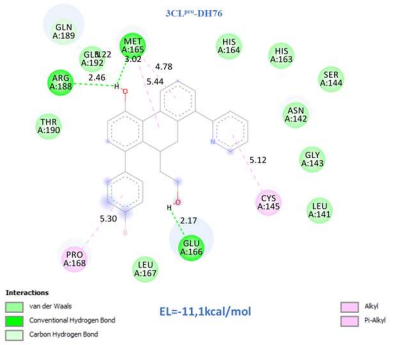
Le **Tableau V. B5** présente des diagrammes en 2D des interactions non-covalentes clés prédites dans cette étude qui se produisent entre la poche active de la 3CL^{pro} et les onze dérivés DH sélectionnés adjacents aux références molécules (N3, DSF, et D40).

Tableau V. B5. Les profils d'interactions non-covalentes protéine-ligand prédits par la simulation de docking moléculaire flexible.

Système protéine-ligand (3CL^{pro}-DH Inhibiteur)	Interactions non-covalentes			
	Liaisons Hydrogène	Électrostatique	van der Waals	Hydrophobique
	**Thr190**(1.69 Å), **Glu166**(1.89,2.12,2.50 Å), **His163**(2.36 Å), **Phe140**(2.78 Å), **Cys145**(3.01 Å), **Gln189**, **His172** (5.32 Å)	-	**Gln192**, **Thr26**, **Thr25**, **Thr24**, **Ser144**, **Gly143**, **Asn142**, **Leu141**, **His172**, **His164** et **Asp187**	**Leu167**(5.4 Å), **Pro168**(4.89 Å), **Ala191**(4.10 Å), **Met165**(4.46 Å), **Met49**(4.54 Å), **His41**(3.99 Å)
	**Gln189**(3.50 Å), **Asn142**(3.51 Å)	-	**Thr25**, **Thr26**, **Met49**, **Gly143**, **Glu166**, **Met165**	Leu27(4.06 Å), **Cys145**(4.63 Å), **His41**(4.99 Å)
	**Cys145**(3.02), **His164**(2.72)	**Cys145**(5.24 Å), **Glu166**(3.95 Å)	**His41**, **His163**, **Gly143**, **Ser144**, **Asn142**, **Phe140**, **Leu141**,	**Met165**(4.89,5.38, 5.41 Å), **Pro168**(4.04 Å), **Leu167**(4.50 Å)

Système protéine-ligand (3CL^{pro}-DH Inhibiteur)	Interactions non-covalentes			
	Liaisons Hydrogène	Électrostatique	van der Waals	Hydrophobique
3CL^{pro}-DH6  	**Gln189**(3.35 Å)	**Met49**(4.92 Å), **His41**(4.54 Å), **Cys145**(5.20 Å)	**Tyr54**, **Pro52**, **His164**, **Glu166**, **Gln192**, **Thr190**, **Asp187**, **Arg188**, **Leu167**	**Met165**(5.50 Å), **Pro168**(4.72 Å)
3CL^{pro}-DH7  	**Ser144**(3.53 Å), **Asn142**(2.81 Å), **Gly143**(2.83 Å),	-	**Glu166**, **Met165**, **His163**, **Phe140**, **Leu141**, **Met49**, **Ser46**, **Pro168**	**Cys145**(5.06 ,4.56 Å)
3CL^{pro}-DH8  	**Ser144**(2.54 Å), **Gly143**(2.82 Å), **Gly166**(3.06 Å), **Thr25**(2.81 Å), **Thr24**(2.69 Å)	-	**Met165**, **His163**, **Leu141**, **Asn142**, **His141**, **Leu27**, **Thr45**, **Met49**, **Gln189**	**Cyc145** (4.87,5.06 Å)
3CL^{pro}-DH12  	**Ser144**(2.42 Å), **Asn142**(2.64 Å), **Gly143**(2.64 Å)	-	**Glu166**, **Met165**, **His163**, **Phe140**, **Leu141**, **Leu27**, **Ser46**, **Met49**	**Cys145**(4.94,5.13 Å)
3CL^{pro}-DH18  	**Ser144**(2.33 Å), **Gly143**(2.80 Å), **Asn142**(3.48 Å)	-	**Glu166**, **Met165**, **His163**, **Phe140**, **Leu141**, **Leu27**, **Ser46**, **Met49**, **Thr24**, **Thr25**	**Cys145**(5.13,4.82 Å)

Système protéine-ligand (3CL^{pro}-DH Inhibiteur)	Interactions non-covalentes			
	Liaisons Hydrogène	Électrostatique	van der Waals	Hydrophobique
	**Gln192**(2.61 Å), **Arg188**(2.55 Å), **Met165**(2.84 Å)	-	**Leu167**, **Thr190**, **Asp187**, **His164**, **His41**, **Met49**, **Leu27**, **Cys145**, Thr26, Glu166, Gly143	**Met165**(5.32, 4.22 Å), **Gln189**(3.48 Å), **Pro168**(5.13,5.23 Å)
	**Asn142**(2.71 Å), **Ser144**(2.36 Å), **Gly143**(2.65 Å)	-	**Pro168**, **Glu166**, **Met165**, **His163**, **Leu141**, **Phe140**, **Leu27**, **Ser46**, **Met49**	**Cys145**(5.07,5.10 Å)
	**Arg188**(2.54 Å), **Cys145**(3.95 Å), **Gln189**(3.42 Å), **Gln192**(Défavorable Donneur-Donateur, 2.14 Å),	**His41**(3.49 Å)	**Pro168**, **Thr190**, **Asp187**, **His164**, **Met49**, **Gly143**, **Glu166**	**Met165**(5.16, 4.36 Å)
	**Thr190**(3.09 Å), **Gln192**(2.97 Å), **Arg188**(2.53 Å)	-	**Leu167**, **Gln189**, **His164**, **Gly143**, **Asn142**, **Glu166**	**Pro168**(5.42 Å), **Met165**(4.40, 4.92 Å), **Cys145**(5.07 Å)
	**Asn142**(2.68), **Ser144**(2.43), **Gly143**(2.67)	-	**Glu166**, **Met165**, **His163**, **Phe140**, **Leu141**, **Leu27**, **Ser46**, **Met49**	**Cys145**(4.96,5.17 Å)

Système protéine-ligand (3CL^{pro}-DH Inhibiteur)	Interactions non-covalentes			
	Liaisons Hydrogène	Électrostatique	van der Waals	Hydrophobique
	**Arg188** (2.46 Å), **Met165** (3.02 Å), **Glu166** (2.17 Å), **Gln189** (3.22 Å)	-	**Gln192**, **His164**, **Asn142**, **Ser144**, **Gly143**, **Leu141**, **Leu167**	**Met165** (4.78, 5.44 Å), **Pro168**(5.30 Å), **Cys145**(5.12 Å)

Chaîne d'acides aminés actifs (référence/noir **gras** ; nouveau/bleu **gras**),

Les résultats de l'analyse moléculaire du Docking des petits ligands optimisés des dérivés DH dans la poche active de la 3CL^{pro} ont démontré que, en plus de l'inhibiteur D40 et de l'inhibiteur DSF, les 11 inhibiteurs proposés avaient des interactions avec la majorité des résidus actifs. Ces interactions étaient principalement de nature hydrophobique, des liaisons hydrogène (pi-donneurs, carbone et liaisons hydrogène conventionnelles), électrostatique et de van der Waals. Cela montre que les dérivés du 9,10-dihydrophénanthrène proposés (DH) ont le potentiel d'inhiber l'activité enzymatique de la 3CL^{pro} de manière non covalente. Ainsi, les acides aminés actifs suivants : **Histidine** (His164, His41, et His163), **Cystéine** (Cys145), **Acide glutamique** (Glu166), **Glycine** (Gly143), **Serine** (Ser144), **Asparagine** (Asn142), **Phénylalanine** (Phe140), **Leucine** (Leu141 et Leu167), **Méthionine** (Met165), et **Proline** (Pro168) dans la poche de la 3CL^{pro}, avec lesquels la molécule modèle (D40) interagit, pourraient représenter des sites de liaison potentiels impliqués dans l'inhibition de l'activité enzymatique de la 3CL^{pro} du SARS-CoV-2 par le biais d'interactions non-covalentes. Les observations faites sur le complexe 3CL^{pro}-DSF indiquent que l'inhibiteur non covalent DSF interagit avec Gln189, Leu27, Cys145 et His41 par une liaison carbone-hydrogène, trois liaisons hydrophobes et d'autres interactions de la classe van der Waals. Dans le complexe 3CL^{pro}-D40, les groupements (OH, R1 ; Bromobenzyl, R2 ; et 4-phenylpyridine, R3) interagissent avec His164, Cys145, Glu166, et Leu167, respectivement, via deux liaisons hydrogène, une liaison électrostatique, et trois liaisons hydrophobes, ainsi que d'autres interactions de la classe de van der Waals. Cela conduit à une énergie d'affinité de liaison accrue du D40 (EL = -10,2 kcal/mol) par rapport à l'inhibiteur du DSF (EL = -4,7 kcal/mol). Les prédictions de la modélisation pharmacophore 3D-QSAR sont également validées par l'optimisation structurelle du D40 pour atteindre une forte activité inhibitrice contre la 3CL^{pro} du SARS-CoV-2. Les modifications apportées à D40 se sont traduites par la production de onze nouvelles petites molécules (DH) présentant un pouvoir inhibiteur et des énergies d'affinité de liaison prometteurs, ainsi que des conformations optimales pour le positionnement dans la poche active de la 3CL^{pro}, ce qui en fait des inhibiteurs non-covalents intéressants contre le COVID-19.

3.4 Identification de ligands potentiels H2L par analyse des propriétés ADME-Tox *in silico*

La conception et la découverte de nouveaux inhibiteurs non-covalents de la protéase $3CL^{pro}$ basés sur de petits composés bioactifs de DH peuvent constituer de nouvelles pistes thérapeutiques intéressantes pour lutter contre le COVID-19. En effet, la majorité des petites molécules ont des propriétés pharmacocinétiques et pharmacodynamiques plus favorables par rapport aux larges molécules telles que l'inhibiteur covalent N3 (connu sous le nom d'accepteur peptidique de Michael) [539], [540]. Pour s'assurer que nos onze inhibiteurs non-covalents optimisés pour $3CL^{pro}$ (DH6, DH7, DH8, DH12, DH18, DH23, DH25, DH26, DH27, DH30 et DH76) sont efficaces et sûrs, leurs propriétés d'absorption, de distribution, de métabolisme, d'excrétion et de toxicité (ADME-Tox) ont été évaluées. Le modèle *in silico* ADME-Tox développé à partir de cette évaluation est présenté sous forme de **Tableau V. B6**, qui fournit un support de sélection des meilleurs candidats H2L identifiés. D'après les prévisions ADME-Tox présentées dans le **Tableau V. B6**, toutes les molécules ont un taux d'absorption élevé dans l'intestin humain (>90%). Cependant, l'inhibiteur N3 a montré une absorption significativement plus faible de 57.88%. Cela peut indiquer que l'intestin est incapable d'absorber suffisamment de peptide N3, ce qui entraîne sa clairance rapide avant qu'il ne touche son objectif thérapeutique. Toutes les molécules examinées présentent des valeurs négatives pour le VDss, ce qui indique que leur distribution se fait principalement dans la circulation sanguine du corps. Les faibles valeurs de la fraction non liée (FU) suggèrent en outre que ces molécules sont plus susceptibles de se lier à la protéase qu'aux protéines plasmatiques. Toutes les molécules, à l'exception de la molécule DH12 (logBB = 0.29), n'ont pas pu pénétrer le système nerveux central (SNC) et la barrière hémato-encéphalique (BHE), ce qui indique qu'elles n'ont pas d'impact potentiel sur le cerveau. En ce qui concerne le métabolisme, toutes les molécules ont été caractérisées comme des substrats efficaces pour l'enzyme cytochrome 3A4, responsable du métabolisme des médicaments dans le corps humain. Cela suggère qu'elles ne seront pas rapidement métabolisées et éliminées par le corps. Par ailleurs, les faibles valeurs de clairance totale des molécules suggèrent qu'elles ont une demi-vie et une stabilité suffisantes dans l'organisme, ce qui leur permet d'atteindre leur objectif thérapeutique avant d'être excrétées, contrairement à l'inhibiteur N3 dont la valeur de clairance totale était beaucoup plus élevée (0.65). Selon les prédictions d'Osiris, pour la substance DH6, il peut y avoir une légère toxicité de type irritation. La molécule DH30 est répertoriée comme étant à la fois mutagène et cancérigène. Les autres composés analysés n'ont révélé aucun risque de toxicité. En raison des risques de toxicité des molécules DH6 et DH30, nous les éliminons et ne conservons que les molécules qui n'ont pas d'effet toxique prédit. Sur la base des prédictions ADME-Tox, les neuf molécules retenues (DH7, DH8, DH12, DH18, DH23, DH25, DH26, DH27 et DH76) sont considérées comme les candidats H2L susceptibles de faire l'objet par la suite de la conception de nouveaux médicaments.

3.5 Affinement des scores du criblage virtuel par le modèle Prime/MM-GBSA

À la suite de l'identification des candidats H2L présentant de bonnes propriétés pharmacocinétiques/pharmacodynamiques, nous les soumettons encore à des analyses plus poussées par voie de simulation. Cette étape a pour but d'évaluer le niveau de stabilité et de

compatibilité des structures des agents H2L vis-à-vis de la structure de la protéase $3CL^{pro}$ ciblée, dans un milieu aqueux virtuel similaire à des conditions réelles. Tout cela peut être reproduit à l'aide en partie des méthodes de validation du post-Docking. Le **Tableau V. B7** indique les énergies de liaison libre pour les complexes optimaux $3CL^{pro}$-DH7, $3CL^{pro}$-DH8, $3CL^{pro}$-DH12, $3CL^{pro}$-DH18, $3CL^{pro}$-DH23, $3CL^{pro}$-DH25, $3CL^{pro}$-DH26, $3CL^{pro}$-DH27 et $3CL^{pro}$-DH76, ainsi que pour les références $3CL^{pro}$-D40 et $3CL^{pro}$-N3 calculées à l'aide du modèle Prime/MM-GBSA. La gamme des énergies de liaison libre pour les ligands conçus (DH) est comprise entre -47.83 kcal/mol pour DH27 et -59.54 kcal/mol pour DH23, ce qui indique que les énergies de liaison libre des ligands médicamenteux (inhibiteurs) conçus sont relativement plus élevées que celles de la référence D40 (-47.77 kcal/mol) et moins élevées que celles du peptide inhibiteur covalent N3 (-83.84 kcal/mol). L'analyse des contributions des énergies des liaisons de van der Waals (ΔG_{bind} Vdw) et d'hydrogène (ΔG_{bind} H-bond), de Coulomb (ΔG_{bind} Coulomb), lipophiles (ΔG_{bind} Lipo) et de tassement (ΔG_{bind} Packing) à l'énergie de liaison moyenne (ΔG_{bind}) des neuf inhibiteurs non-covalents identifiés, ainsi que les agents de référence D40 et N3 ont montré que la contribution des interactions non-liantes est plus favorable à la stabilité des inhibiteurs conçus dans la poche active de la $3CL^{pro}$ que les interactions covalentes potentielles.

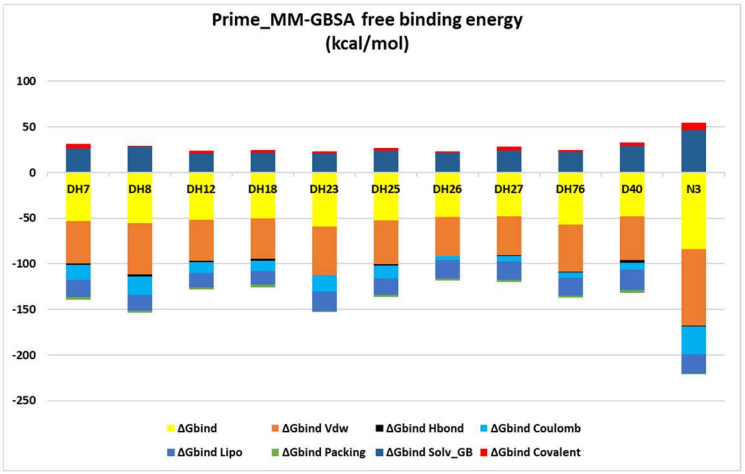


Figure V. B13. Histogramme des énergies libres de liaison recalculées par le modèle Prime/MM-GBSA pour les systèmes $3CL^{pro}$-inhibiteurs.

Les données rassemblées dans la **Figure V. B13** sont cohérentes avec les résultats de Docking moléculaire, à l'exception au ΔG_{bind} Solv_GB et au ΔG_{bind} Covalent, en ce qui concerne les énergies libres de liaison des systèmes protéine-ligand évalués par le modèle Prime/MM-GBSA. Cela indique que les petites molécules DH7, DH8, DH12, DH18, DH23, DH25, DH26, DH27 et DH76 proposées comme inhibiteurs non covalents sont très favorables à la liaison à la poche active de la $3CL^{pro}$. Par conséquent, ces neuf dérivés optimisés à base de 9,10-dihydrophénanthrène peuvent être considérés comme des inhibiteurs non-covalents médicamenteux prometteurs contre la $3CL^{pro}$ du SRAS-CoV-2, susceptibles d'être utilisés dans le cadre de la thérapie COVID-19. Cela confirme davantage que les liaisons non-covalentes entre les ligands H2L des dérivés DH et les résidus d'acides aminés actifs de la $3CL^{pro}$ sont compatibles. Par conséquent, les neuf inhibiteurs à base de 9,10-dihydrophénanthrène (DHIs) peuvent être considérés comme de nouveaux inhibiteurs antiviraux du type non covalent, susceptibles de se lier à la $3CL^{pro}$ et d'inhiber/moduler son activité enzymatique.

Tableau V. B6. Profil d'ADME-Tox des candidats H2L.

	Propriétés																	
	Absorption		Distribution				Métabolisme							Excrétion	Risques de toxicité			
							Cytochrome P450 (CYP)											
							Substrat		Inhibiteur									
Modèles	Absorption intestinale (humain)	Substrat de P-glycoprotéine	VDss (humain)	Fraction non liée (humain)	Perméabilité de BHE	Perméabilité du SNC	2D6	3A4	1A2	2C19	2C9	2D6	3A4	Clairance totale	Mutagène	Tumorigène	Irritant	Efficacité de la reproduction
Unité	Numérique (%)	Oui/Non	(Log L kg-1)	Numérique	(Log BB)	(Log PS)	Oui/Non							(log mL min^{-1} kg^{-1})	Critères catégoriels : Non toxique : , Toxique : , Faiblement toxique :			
Predicted values																		
N3	57.88	Yes	-0.76	0.06	-1.26	-3.56	Non	Oui	Non	Non	Non	Non	Oui	0.65				
D40	97.24	Yes	-0.96	0.33	-0.22	-1.15	Non	Oui	Oui	Non	Non	Non	Non	-0.07				
D06	100	Yes	-1.04	0.17	-0.02	-1.59	Non	Oui	Oui	Oui	Oui	Non	Oui	0.15				
D07	100	Yes	-1	0.26	-0.38	-1.55	Non	Oui	Oui	Non	Oui	Non	Non	0.13				
D08	92.53	Yes	-1.26	0.05	-0.74	-2.20	Non	Oui	Oui	Oui	Oui	Non	Oui	-0.16				
D12	100	Yes	-0.98	0.33	0.29	-1.81	Non	Oui	Oui	Oui	Oui	Non	Non	0.13				
D18	93.66	Yes	-1.01	0.06	-0.75	-1.77	Non	Oui	Non	Oui	Oui	Non	Oui	-0.18				
D23	97.17	Yes	-1.05	0.06	-0.22	-2.21	Non	Oui	Non	Oui	Oui	Non	Oui	-2.21				
D25	99.27	Yes	-0.87	0.26	-0.40	-1.62	Non	Oui	Oui	Oui	Oui	Non	Non	0.06				
D26	99.94	Yes	-0.75	0.26	0.14	-0.99	Non	Oui	Oui	Oui	Non	Non	Non	-0.21				
D27	100	Yes	-0.93	0.26	-0.37	-1.61	Non	Oui	Oui	Oui	Oui	Non	Non	-1.61				
D30	100	Yes	-0.87	0.27	-0.40	-1.74	Non	Oui	Oui	Oui	Oui	Non	Non	0.19				
D76	98.69	Yes	-0.87	0.26	-0.41	-1.44	Non	Oui	Oui	Oui	Oui	Non	Oui	-0.20				

VDss : volume de distribution ; BHE : barrière hémato-encéphalique ; SNC : système nerveux central

Cela renforce la possibilité que ces molécules parviennent à passer aux confins de l'inévitable pipeline de conception de nouveaux médicaments.

Tableau V. B7. Re-scores par Prime/MM-GBSA des énergies libres de liaison des nouveaux 3CL^{pro}-DHIs optimisés par rapport aux systèmes biomoléculaires de référence 3CL^{pro}-N3 et 3CL^{pro}-D40.

Prime/MM-GBSA (kcal/mol)	Nouveaux modèles de "3CL^{pro}-DHIs"									Système de référence 3CL^{pro}-ligand.	
	DH7	DH8	DH12	DH18	DH23	DH25	DH26	DH27	DH76	DH40	N3
ΔG_{bind}	-53.51	-55.15	-51.93	-50.43	-59.54	-52.24	-48.92	-47.83	-57.30	-47.77	-83.84
ΔG_{bind} Vdw	-45.89	-56.18	-44.44	-44.09	-52.87	-48.12	-42.18	-42.97	-51.58	-48.34	-84.31
ΔG_{bind} Hbond	-2.00	-2.21	-1.85	-1.99	-0.16	-1.86	-0.34	-0.44	-0.31	-2.72	-0.56
ΔG_{bind} Coulomb	-15.84	-20.62	-12.05	-11.07	-17.54	-13.56	-4.69	-6.17	-6.05	-7.24	-30.72
ΔG_{bind} Lipo	-19.33	-16.97	-15.47	-15.51	-21.99	-18.39	-20.00	-20.40	-19.92	-22.83	-20.61
ΔG_{bind} Packing	-2.54	-2.71	-2.28	-2.47	-1.01	-2.28	-2.21	-1.73	-2.26	-2.71	-0.03
ΔG_{bind} Solv_GB	26.11	28.50	21.15	21.87	21.02	23.87	21.34	24.24	22.76	28.82	46.07
ΔG_{bind} Covalent	5.00	1.04	3.02	2.85	2.02	3.11	2.16	3.95	2.07	4.10	8.50

3.6 Évaluation de la stabilité des interactions Protéine-Ligand par simulations de dynamique moléculaire

Dans cette étape de validation computationnelle, des simulations indépendantes de dynamique moléculaire sur une ligne de temps de 100 ns sur la 3CL^{pro} (6LU7) complexée et libre ont été réalisées. Le comportement structurel et dynamique et la stabilité des interactions non-covalentes potentielles entre la poche active de la 3CLpro et les nouveaux 3CL^{pro}-DHIs portant l'échafaudage 9,10-dihydrophénanthrène sont examinés en considérant les trois premiers systèmes avec le meilleur enregistrement de MM-GBSA/Prime_ΔG_{bind} (3CL^{pro}-DH8 : ΔG_{bind}=-55. 15 kcal/mol ; 3CL^{pro}-DH23 : ΔG_{bind}=-59.54 kcal/mol ; 3CL^{pro}-D76 : ΔG_{bind}=-57.30 kcal/mol) comme échantillons d'entrée pour l'analyse de la dynamique moléculaire. Le système 3CL^{pro}-D40 est considéré comme un échantillon de référence, tandis que le système 3CL^{pro}-N3 est exclu car il a déjà été validé lors de l'étape de validation du modèle du Docking moléculaire de cette étude.

3.6.1 Analyses de RMSD et de RMSF

Les paramètres d'écart quadratique moyen (RMSD) et de fluctuation quadratique moyenne (RMSF) sont employés pour évaluer les fluctuations potentielles du squelette de la macromolécule SARS-CoV-2 3CL^{pro} (code PDB : 6LU7) à l'état libre (non complexé) et lorsqu'elle est complexée aux échantillons DHIs (DH8, DH23 et DH76), optimisés *in silico* et au D40 identifié comme conformation bioactive de référence. Les indices RMSD et RMSF mesurent la moyenne du squelette α-carbone de tous les résidus d'acides aminés formant la macromolécule. Les variations RMSD et RMSF des systèmes 3CL^{pro}-DHIs sont calculées en utilisant comme référence le premier frame du squelette de 3CL^{pro} non complexé (6LU7_free). La **Figure V. B14** montre la progression des variations de RMSD et de RMSF obtenues à partir des trajectoires de simulation de dynamique moléculaire de 100ns.

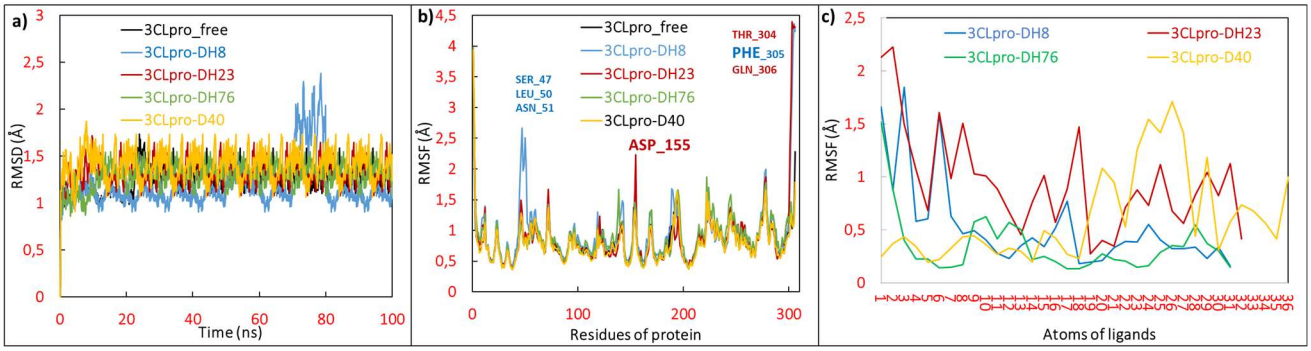


Figure V. B14. Diagrammes temporels des variations RMSD (a) et RMSF (b) correspondant aux atomes du squelette des systèmes 3CL^{pro}_free, 3CL^{pro}-DH8, 3CL^{pro}-DH23, 3CL^{pro}-DH76, et 3CL^{pro}-D40. (c) Diagrammes temporels RMSF des atomes ligands DH8, DH23, DH76, et D40 à l'intérieur de la poche active de 3CL^{pro}.

Les données de la **Figure V. B14(a)** indiquent que les atomes Cα du squelette de la 3CL^{pro} sont relativement stables, bien que les ligands DH8, DH23 et DH76 soient Dockés dans sa poche active lors d'une trajectoire de simulation de dynamique moléculaire de 100 ns. De légères fluctuations sont observées dans les chaînes latérales des acides aminés du complexe 3CL^{pro}-DH8 entre 70 et 80 ns, mais ces fluctuations ne dépassent pas 1.5 Å. Les valeurs RMSD moyennes enregistrées pour les systèmes 3CL^{pro} (3CL^{pro}_free, 3CL^{pro}-DH8, 3CL^{pro}-DH23, 3CL^{pro}-D76 et 3CL^{pro}-D40) sont les suivantes : 1. 239 Å, 1.179 Å, 1.876 Å, 1.305 Å, 1.284 Å et 1.400 Å, respectivement, ce qui montre que les inhibiteurs non-covalents de 3CL^{pro} proposés sont bien équilibrés et compatibles avec la poche active de 3CL^{pro}. Les valeurs RMSF moyennes des résidus α-carbone des systèmes 3CL^{pro} étudiés - 3CL^{pro}-free, 3CL^{pro}-DH8, 3CL^{pro}-DH23, 3CL^{pro}-DH76 et 3CL^{pro}-D40 - étaient respectivement de 0.821 Å, 0.961 Å, 0.886 Å, 0.873 Å et 0.788 Å, comme l'indique la **Figure V. B14(b)**. Bien que certaines séquences de résidus impliquant SER_47, LEU_50, ASN_51, ASP_155, THR_304, PHE_305 et GLN_306 aient présenté des fluctuations extrêmes, celles-ci n'ont pas dépassé 4.5 Å, ce qui suggère un degré élevé de stabilité des DHIs à l'intérieur de la poche active de la 3CL^{pro}. En outre, la **Figure V. B14(c)** montre que les petits ligands DHIs (DH8, DH23, DH76 et D40) présentent des fluctuations mineures, qui résultent probablement de certaines caractéristiques structurelles flexibles telles que les angles de rotation, la torsion et des interactions non covalentes libres et favorables dans les complexes 3CL^{pro}-ligands.

3.6.2 Caractérisation des interactions non-covalentes dynamiques dans les complexes 3CL^{pro}-DHIs

Les acides aminés actifs de la poche 3CL^{pro}-N3 (PDB ID : 6LU7) comprennent les résidus polaires Thr190, Thr24, Gln189, Gln192, Asn142, Asp187, Ser144 ; les résidus chargés négativement Glu166 ; les résidus chargés positivement His163, His172, His41, His164 ; les résidus non polaires Phe140, Met165, Met49, Leu167, Pro168, Ala191, et le résidu hydrophobe Cystéine 145. Les interactions entre les DHIs et ces résidus, et leur présence dans la poche active de la 3CL^{pro} contribuent probablement à l'inhibition de son activité enzymatique, comme le montrent les diagrammes de contacts protéine-ligand présentés dans les **Figures V. B(15-16)**, qui reflètent ces simulations sur 100ns de dynamique moléculaire.

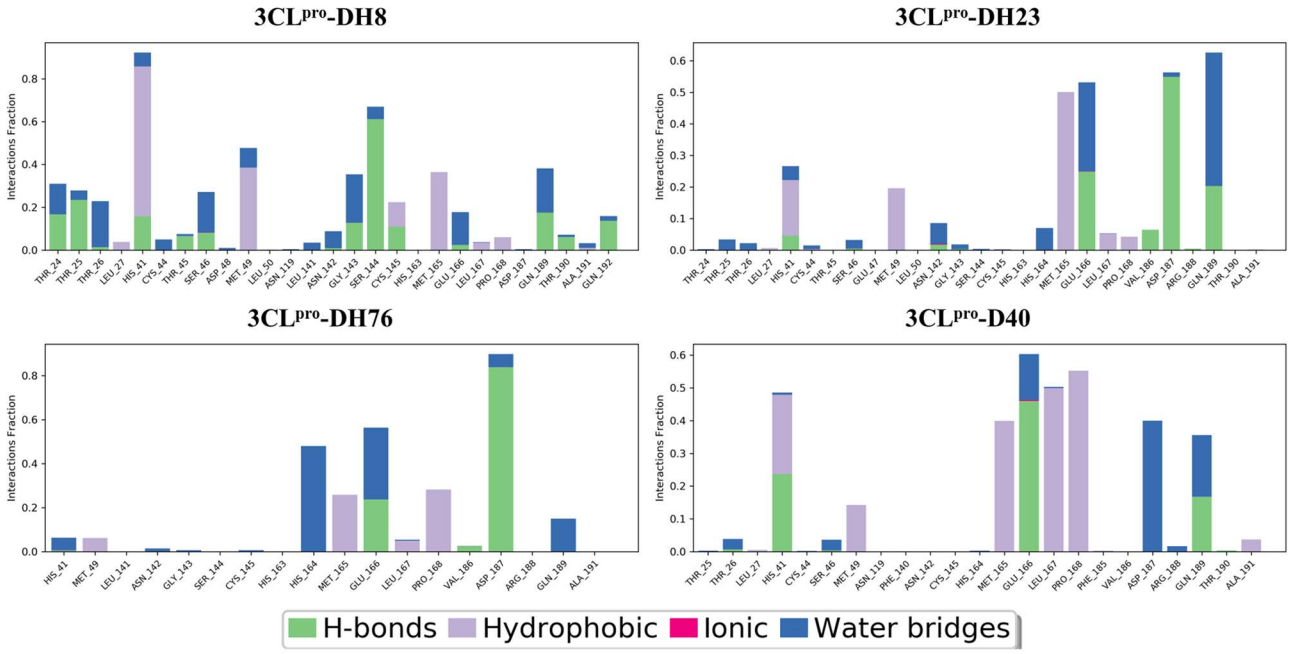


Figure V. B15. Tracés temporels en pourcentage des contacts protéine-ligand (PL) enregistrés au cours des 100ns de simulations de dynamique moléculaire effectuées sur les systèmes biomoléculaires 3CL^{pro}-DH8, 3CL^{pro}-DH23, 3CL^{pro}-DH76, et 3CL^{pro}-D40.

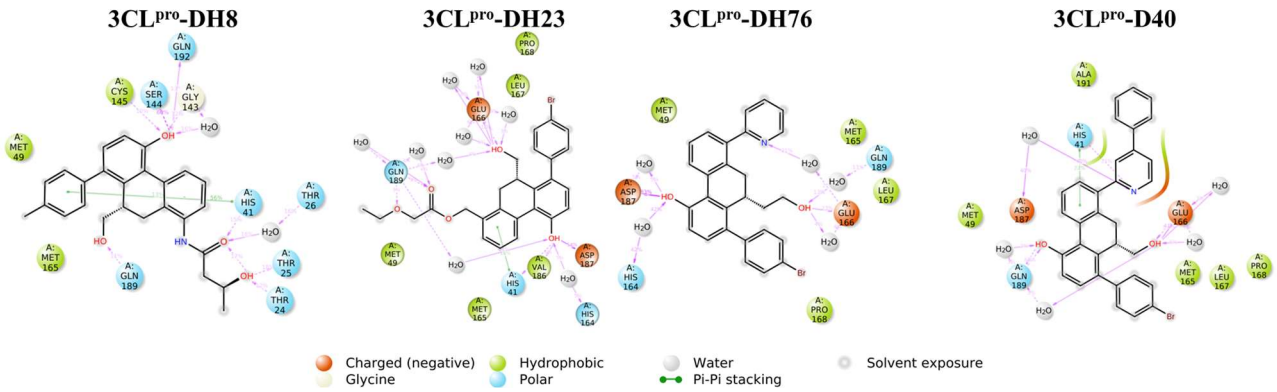


Figure V. B16. Visualisation 2D des contacts enregistrés au niveau de la poche active de la 3CL^{pro} et des DHIs optimisées (DH8, DH23, DH76), ainsi que de la référence D40 pendant 100ns de simulations de dynamique moléculaire.

Les diagrammes (**Figures V. B15 et V. B16**) montrent que les DHIs ont pu conserver la plupart de leurs interactions non-covalentes avec les acides aminés situés dans la poche de liaison du 3CL^{pro}, comme indiqué par le modèle de Docking moléculaire flexible. Le **Tableau V. B8** récapitule les principaux contacts protéine-ligand et les proportions de contact concernant 3CL^{pro}-DH8, 3CL^{pro}-DH23, 3CL^{pro}-DH76 et 3CL^{pro}-D40 pendant les 100ns de simulations de dynamique moléculaire. Ces données impliquent que les têtes de DHIs sont encore associées au site de liaison cible de la 3CL^{pro} pendant toute la durée des simulations. D'après les données sur les durées de contact protéine-ligand présentées dans le **Tableau V. B8**, obtenues après 100ns de simulation de dynamique moléculaire, nous pouvons clairement observer que les échantillons DH08, DH23, DH76 et D40 parviennent à conserver l'interaction dans la poche active de 3CL^{pro} avec la plupart des résidus d'acides aminés de référence qui ont été identifiés via le modèle flexible de Docking moléculaire. La contribution de certains nouveaux résidus d'acides aminés à la stabilité des ligands DH08, DH23, DH76 et D40 dans la poche active de la 3CL^{pro} a également été révélée.

Tableau V. B8. Contacts 3CL^{pro}-DHIs enregistrés sur les trajectoires de 100ns des simulations de dynamique moléculaire appliquées aux échantillons de 3CL^{pro}-DH8, 3CL^{pro}-DH23, 3CL^{pro}-DH76 et 3CL^{pro}-D40 de type macromoléculaire.

	Contacts non-covalents Protéine-ligand		
Complexe	**Liaisons hydrogène**	**Hydrophobique**	**Ponts d'eau**
3CL^{pro}-DH8	**Thr24** (12%), **Thr25** (18%), **His41** (15%), **Gln192** (13%), **Ser144** (60%), **Gln189** (20%) **Cy145** (10%), **Thr45** (8%), **Asn142** (3%), **Glu166** (4%), **Thr190** (5 %) and **Thr26** (16%)	**Leu27** (6%), **His41** (56%), **Met49** (40%), **Cys145** (10%), **Met165** (37%), **Leu167** (5%), **Pro168** (6%) and **Ala191** (3%)	**Thr24** (26%), **Cys44** (4%), **Ser46** (16%), **Leu141** (3%), **Asn142**(5%), **Cys143**(18%), **G1y166**(13%), **Gln189** (18%), **Thr190**, **Ala191**, and **Gln192** (<5%)
3CL^{pro}-DH23	**His41** (6%), **Asn142** (5%), **Val186** (6%), **Glu166** (25%), **Asp187** (54%) and **Gln189** (18%)	**His41** (16%), **Met49** (18%), **Cy145** (10%), **Met165** (38%), **Pro168** (6%) and **Leu166** (5%)	**Thr25** (4%), **Thr26** (3%), **His41** (4%), **Asn142** (7%), **His164** (8%), **Leu166** (24%) and **Gln189** (42%)
3CL^{pro}-DH76	**Asp187** (83%), **His164** (42%), **Glu166** (23%), **Gln189** (13%) and **Val186** (5 %).	**Met49** (5%), **Met165** (22%), **Leu167** (6%) and **Pro168** (26%).	**His41** (5%), **His164** (42%), **Glu166** (25%) and **Gln189** (13%),
3CL^{pro}-DH40	**Thr26** (<3%), **His41**(22%), **Glu166**(41%) and **Gln189**(18%).	**His41** (20%), **Met49** (12%), **Met165** (40%), **Leu167** (50%), **Pro168** (52%) and **Ala191** (<10%).	**Thr26** (5%), **Ser46**(6%), **Glu166** (15%), **Asp187** (41%), **Arg188**(3%) and **Gln189** (18%)

La ligne **noire** en gras représente les résidus initiaux (de référence) tandis que la ligne **bleue** en gras représente les résidus détectés après la simulation de la DyM

Parmi les nouvelles séquences d'acides aminés actifs les plus importantes détectées après la simulation de dynamique moléculaire, Cys44, Ser46, Cys143, Thr190 et Ala191 ont été identifiées pour le DH08, Val186, Leu166 et Thr25 pour le DH23, Val186 pour le DH76 et Ser46, Asp187 et Arg188 pour le D40. Ces résultats confirment la grande validité du modèle de Docking moléculaire développé dans cette étude sur la prédiction précise et fiable au niveau des mécanismes d'affinité/compatibilité protéine-ligand identifiés dans cette étude. Ce qui donne un caractère universel à ce modèle pour d'autres applications dans la détection précoce d'un nouveau médicament grâce au même plan et à la même méthodologie que ceux utilisés dans la présente étude. Les profils de contacts protéine-ligand générés par des simulations de dynamique moléculaire suggèrent que les énergies libres de liaison élevées des DHI (DH8, DH23, DH76 et D40) prédites par le modèle Prime/MM-GBSA sont confirmées par les faibles interactions non-covalentes entre les DHI et la poche active de la 3CL^{pro}. Les interactions les plus importantes qui contribuent à la stabilité de ces systèmes sont les liaisons hydrophobiques, les liaisons hydrogène et les ponts d'eau. Ces résultats signifient que les molécules médicamenteuses non-covalentes peuvent avoir un comportement thérapeutique et pharmacocinétique de qualité acceptable par rapport aux inhibiteurs covalents comme celle de Michael (N3). Par conséquent, les médicaments non-covalents peuvent être moins difficiles à séparer de la protéine cible afin d'éviter les effets indésirables.

3.6.3 Exploration des propriétés dynamiques des ligands DHIs au sein des complexes 3CL^{pro}

La **Figure V. B17** présente les propriétés des ligands DH8, DH23, DH76 et D40 optimisés (LO) évaluées pendant 100ns de trajectoires de simulations de dynamique moléculaire. Dans cette phase, nous avons analysé six propriétés différentes des ligands : la RMSD du ligand, le rayon de giration (rGyr), les liaisons hydrogène intramoléculaires (intra-HB), la surface moléculaire (MolSA), la surface accessible au solvant (SASA) et la surface polaire (PSA) pour chaque ligand. Pour le ligand D08 dans

le complexe 3CL^{pro}-DH8, la RMSD était comprise entre 0.5 et 1.5 Å et son équilibre se situait approximativement à 1 Å. rGyr était limité à la plage (4.25-4.75 Å) et son équilibre se situait autour de 4.5 Å. L'intra-HB était élevé tout au long de la simulation MD. MolSA se situait dans la plage (376-400 Å2) et son équilibre était d'environ 390 Å2. SASA a montré de fortes fluctuations allant d'environ (200-600 Å2) sur l'intervalle de temps (40-80 ns), puis s'est stabilisé à environ 400 Å. PSA était dans la gamme (120-180 Å2) et son équilibre était d'environ 160 Å2. Dans le complexe 3CL^{pro}-DH23, le ligand DH23 a montré quelques fluctuations dans la RMSD allant de (0.8-2.4 Å) avant de se stabiliser à environ 1.6 Å. rGyr a varié de (5.6-4.4 Å) et s'est stabilisé à environ 5 Å. Intra-HB a été faible tout au long de la simulation DyM. MolSA était compris entre (375-420 Å^2) et s'est parfaitement stabilisé tout au long de la simulation à environ 405 Å^2. La SASA a montré de fortes fluctuations autour de la plage (200-400 Å2) et s'est ensuite stabilisée à environ 200 Å^2. Le PSA a été limité dans la plage (80-140 Å^2) et s'est stabilisé à environ 120 Å^2. Dans le complexe 3CL^{pro}-DH76, le ligand D76 a montré une stabilité parfaite en termes de RMSD dans la gamme (0.5-1.2 Å) sur la période de simulation avec quelques légères fluctuations dans la période à peu près 80 ns. rGyr était dans la gamme (4.8-4.5 Å) et s'est stabilisé à environ 4.65 Å. Intra-HB n'a pas été détecté dans le ligand D76. La MolSA était comprise entre 376 et 392 Å^2 et s'est stabilisée à environ 384 Å^2. La SASA était comprise entre 180 et 300 Å^2 et s'est presque stabilisée à 240 Å^2. Le PSA se situait dans la plage (75-105 Å2) et s'est stabilisé à environ 90 Å^2. Dans le complexe 3CL^{pro}-D40, la RMSD du ligand D40 était comprise entre 0.5 et 1.8 Å et s'est stabilisée à environ 1.2 Å. rGyr était compris entre 4.8 et 5.7 Å et s'est stabilisé après 40 ns à environ 5.4 Å. L'intra-HB n'a pas été détecté dans le ligand D40. La MolSA était comprise entre 424 et 448 Å^2 et s'est stabilisée à environ 440 Å^2. La SASA était comprise entre (160-320 Å^2) ; elle se stabilise à environ 240 Å^2 après 40 ns de la trajectoire de la dynamique moléculaire. Le PSA était compris entre (70-100 Å^2) et s'est stabilisé à environ 80 Å^2.

3.6.4 Caractérisation thermodynamique des complexes Protéine-Ligand

Les propriétés thermodynamiques suivantes : énergie totale (E), énergie potentielle (E_P), température (T), pression (P) et volume (V) des cinq systèmes biomoléculaires - à savoir 3CL^{pro}_free, 3CL^{pro}-DH8, 3CL^{pro}-DH23, 3CL^{pro}-DH76 et 3CL^{pro}-D40 sont évaluées et analysées au cours d'une trajectoire de simulation de dynamique moléculaire d'une durée de 100ns. Comme les montres la **Figure V. B18**, les diagrammes thermodynamiques générés témoignent de la performance de la simulation. Les propriétés thermodynamiques des complexes 3CL^{pro}-DH8, 3CL^{pro}-DH23, 3CL^{pro}-DH76 et 3CL^{pro}-D40 sont restées stables et proches de celles du 3CL^{pro}_free (non complexé), comme le montrent les valeurs moyennes (E, EP, T, P et V) des complexes de 3CL^{pro} dans le **Tableau V. B9**, qui sont semblablement voisines de celles du 3CL^{pro}_free. Ces chiffres indiquent que les petites molécules optimisées à base de 9, 10-dihydrophénanthrène sont parfaitement stables et compatibles avec la poche active de la 3CL^{pro}. L'analyse de dynamique moléculaire post-Docking indique que les principaux échantillons de DHIs optimisés DH8, DH23, DH76 et D40 présentent une grande stabilité lorsqu'ils interagissent avec la poche active de la 3CL^{pro}. Par conséquent, les neuf nouveaux inhibiteurs DHI optimisés dans cette étude (DH7, DH8, DH12, DH18, DH23, DH25, DH26, DH27 et DH76) peuvent

être exploités pour la conception précoce de nouveaux médicaments et la découverte de nouveaux candidats traitements contre les infections virales de type COVID-19.

Tableau V. B9. Valeurs moyennes des propriétés thermodynamiques enregistrées pendant 100ns de simulations de dynamique moléculaire pour les systèmes biomoléculaires de $3CL^{pro}$.

Systèmes protéine-ligand	Valeur moyenne des propriétés				
	Énergie totale (kcal/mol)	L'énergie potentielle (kcal/mol)	Température (Kelvin)	Pression (bar)	Volume ($Å^3$)
$3CL^{pro}$_free	-99905.370	-122063.495	298.705	1.409	362814.329
$3CL^{pro}$ -DH8	-99756.824	-121927.605	298.714	1.565	362806.668
$3CL^{pro}$-DH23	-99661.259	-121823.752	298.702	1.193	362702.992
$3CL^{pro}$-DH76	-99649.019	-121807.582	298.709	1.247	362706.347
$3CL^{pro}$-D40	-99603.363	-121760.057	298.704	1.885	362615.184

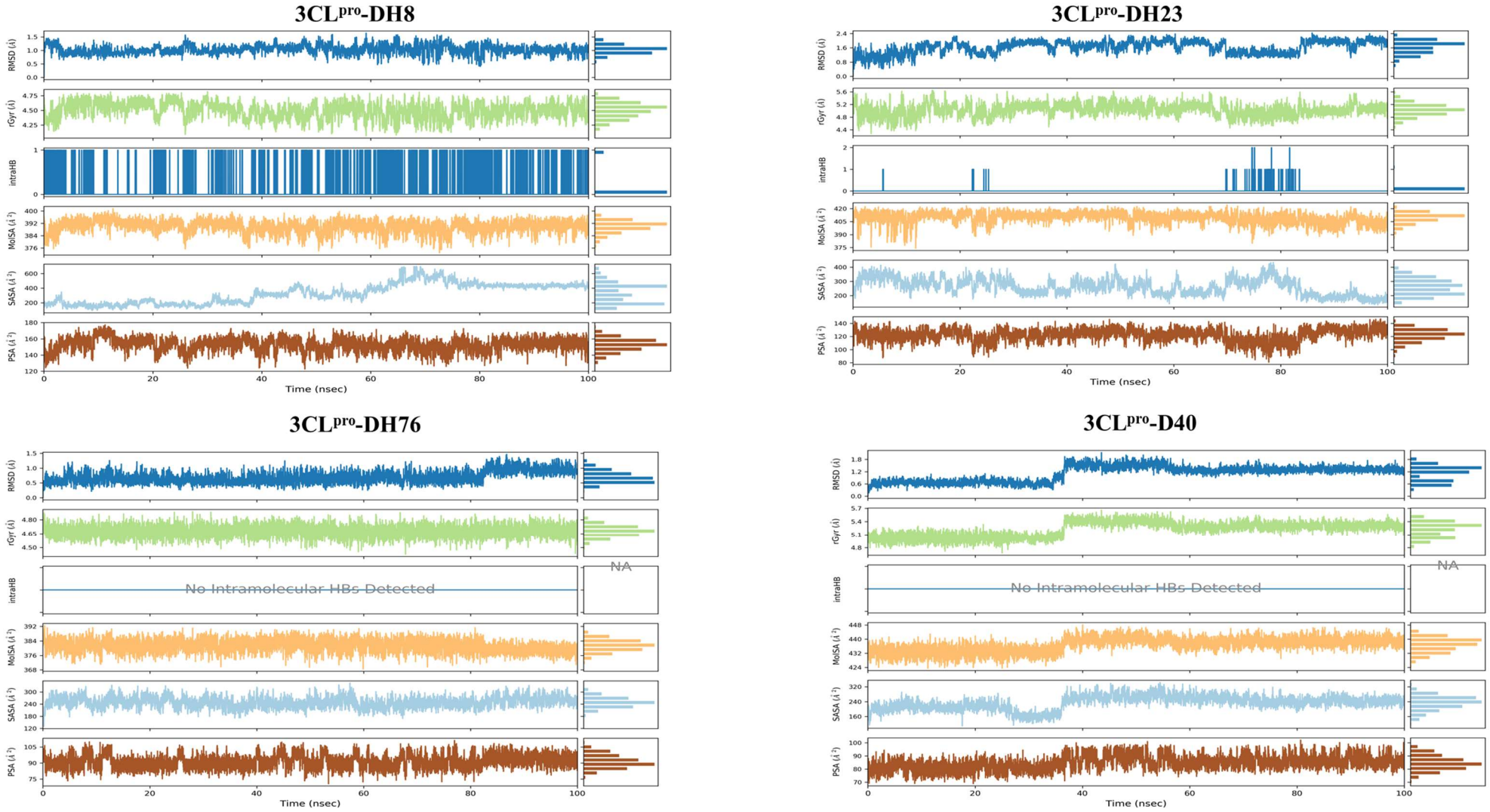


Figure V. B17. Diagrammes temporels de l'évolution des propriétés dynamiques des ligands DH8, DH23, DH76 et D40 examinés dans la poche active de la 3CL^{pro} au cours des 100ns de trajectoire des simulations de dynamique moléculaire.

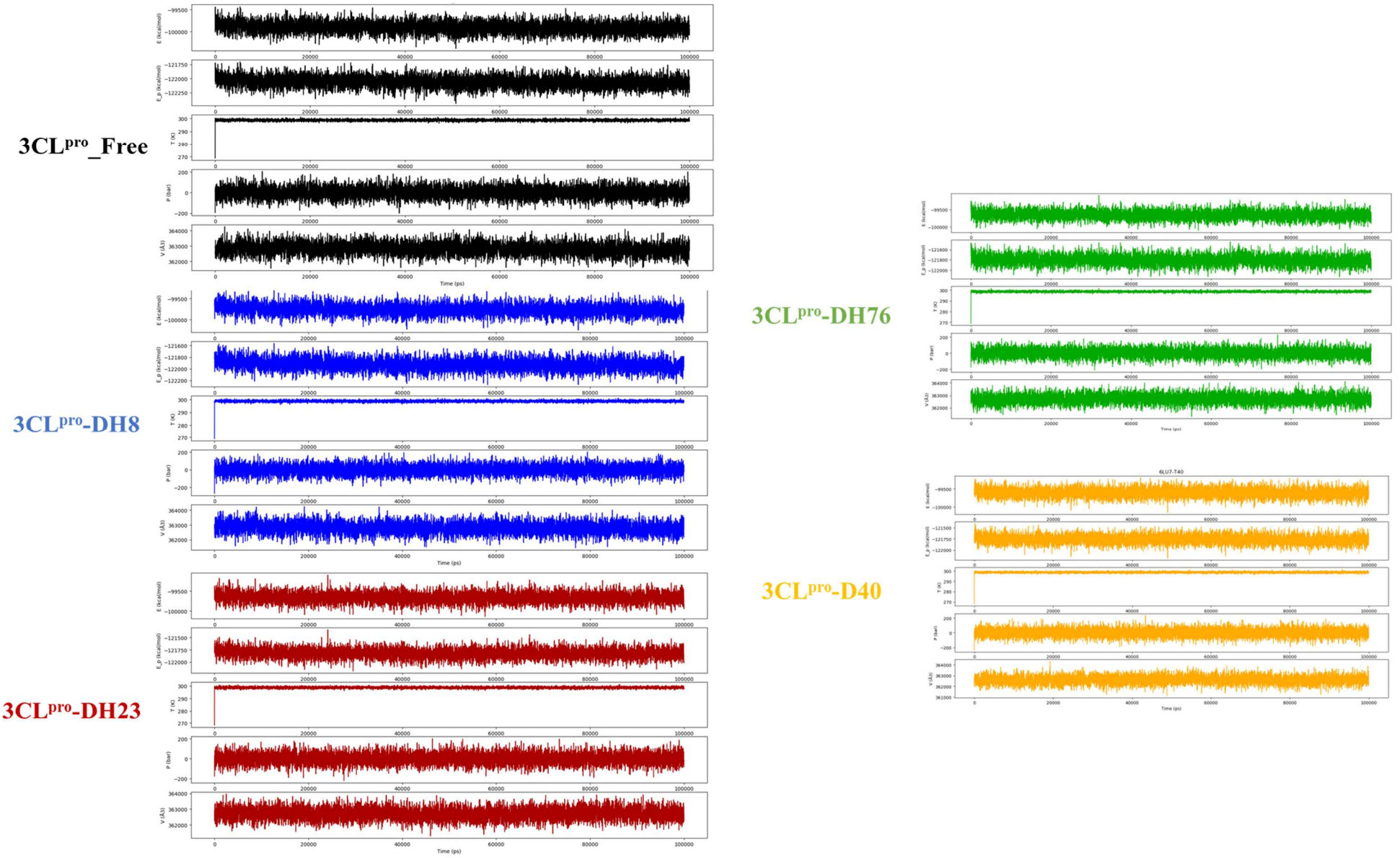


Figure V. B18. Diagrammes des propriétés thermodynamiques enregistrées au cours de 100ns des simulations de la dynamique moléculaire appliquées aux systèmes biomoléculaires de $3CL^{pro}$ ($3CL^{pro}$_free, $3CL^{pro}$-DH8, $3CL^{pro}$-DH23, $3CL^{pro}$-DH76, et $3CL^{pro}$-D40).

4 Conclusion

L'épidémie mondiale de COVID-19 causée par le coronavirus 2 du syndrome respiratoire aigu sévère (SARS-CoV-2) a eu un profond impact sur la vie quotidienne ; elle exige également au préalable de trouver les moyens de lutter efficacement contre cette épidémie. Dans la présente étude, la protéase à cystéine de type 3-chymotrypsine ($3CL^{pro}$) est identifiée comme un récepteur cible essentielle pour la réplication virale. L'inhibition de l'activité enzymatique de la $3CL^{pro}$ pourrait donc être une solution intéressante pour lutter contre la transmission et la croissance du syndrome COVID-19. Pour approfondir la réflexion sur le potentiel de cette stratégie, une petite chimiothèque de dérivés du 9,10-dihydrophénanthrène a été testée **in vitro** ; elle a fait l'objet à cette occasion d'une analyse computationnelle *in silico*. À la suite de cette analyse, les dérivés du 9,10-dihydrophénanthrène se sont vus optimisés dans le but de générer de potentiels inhibiteurs non-covalents de la $3CL^{pro}$, qui peuvent être bénéfiques pour développer de nouvelles stratégies thérapeutiques contre les infections virales associées à la COVID-19. Dans cette étude, la relation structure-activité quantitative 3D (3D-QSAR) pour les dérivés du 9,10-dihydrophénanthrène est analysée à l'aide des techniques CoMFA et CoMSIA, et deux modèles (CoMFA/SE et CoMSIA/SEHDA) sont mis au point. Les résultats ont montré que la structure des dérivés de 9,10-dihydrophénanthrènes pouvait prédire l'activité biologique du pIC_{50} contre le SARS-CoV-2 $3CL^{pro}$, et les sites pharmacologiques correspondants sont décrits. Les propriétés optimales requises pour optimiser l'activité biologique inhibitrice des molécules étudiées sont générées. On a également utilisé une technique de décodage de réseau neuronal génératif basée sur la forme comme base pour développer 96 nouveaux dérivés rationnels de 9,10-dihydrophénanthrènes (DH1-DH96) sur la base de la conformation biactive de référence D40. D'autres analyses sont menées pour évaluer l'activité inhibitrice (pIC_{50}) des dérivés DH par 3D-QSAR, et pour analyser la liaison non-covalente à la cible $3CL^{pro}$ par les modèles de Docking moléculaire flexible et Prime/MM-GBSA, l'adéquation pharmacocinétique et pharmacodynamique par le modèle ADMET-Tox, ainsi que les risques de toxicité à l'aide du modèle d'Osiris. Neuf des nouveaux dérivés de DH (DH7, DH8, DH12, DH18, DH23, DH25, DH26, DH27 et DH76) présentaient une activité inhibitrice biologique prédite élevée, une bonne liaison à la 3CLpro, des propriétés pharmacocinétiques acceptables et des risques de toxicité limités en comparaison avec le composé D40 synthétisé et évaluer à partir du modèle in vitro et avec l'inhibiteur peptidique covalent N3. Notre modèle CADD hybride est validé en termes computationnels par des simulations de dynamique moléculaire effectuées sur la $3CL^{pro}$ libre et complexée avec les dérivés optimisés du 9,10-dihydrophénanthrène (DHIs) dans un environnement aqueux sur un intervalle de temps de 100ns. Les résultats des simulations de dynamique moléculaire montrent que les inhibiteurs de $3CL^{pro}$ optimisés (DHI) pourront servir comme substances médicamenteuses non-covalentes destinées à la thérapie contre le COVID-19. Nous suggérons que ces petites molécules peuvent être exploitées comme base de la mise au point de nouveaux modes de thérapie contre le SARS-CoV-2. Enfin, ces DHIs peuvent être soumis à des analyses biologiques contre d'autres récepteurs liés à des maladies virales par le biais de stratégies hybrides de CADD similaires à celles utilisées dans la présente étude.

CONCLUSION GÉNÉRALE ET PERSPECTIVES

Cette thèse illustre l'exploitation optimale des techniques de modélisation moléculaire, telles que le criblage virtuel et les simulations biophysiques *in silico*, dans la recherche et le développement précoces de nouveaux ligands bioactifs à des fins thérapeutiques. Notre objectif principal était de concevoir une stratégie de CADD (Conception de Médicaments Assistée par Ordinateur) adaptable pour optimiser et identifier des ligands candidats à partir de bases de données de petites molécules bioactives de haute qualité. Cette stratégie hybride combine les approches LBDD (conception de médicaments basée sur les ligands) et SBDD (conception de médicaments basée sur la structure) avec des techniques de ML (apprentissage automatique) et d'IA (intelligence artificielle) pour le criblage virtuel et l'optimisation *in silico* des bibliothèques chimiques. Nous avons également cherché à développer une méthodologie CADD flexible et facilement applicable pour le criblage rationnel et l'optimisation de ligands candidats en vue de la conception de nouveaux médicaments. Cette approche flexible peut être adaptée à différentes situations, qu'il y ait ou non des données expérimentales sur l'activité biologique ou la structure du récepteur ciblé. Nous avons utilisé des approches biostatistiques avancées pour analyser les relations structure-activité et prédire les profils de compatibilité et d'affinité des systèmes récepteur-ligand. L'approche 2D/3D-QSAR nous a permis de développer une compréhension approfondie des relations structure-activité, en utilisant des descripteurs moléculaires multidimensionnels relatifs à la forme 2D et 3D des ligands. Pour prédire les profils d'affinité, de stabilité et de compatibilité des ligands dans les poches actives des récepteurs biologiques ciblés, nous avons utilisé des techniques de Docking moléculaire et de dynamique moléculaire. En ce qui concerne la prédiction des profils de biodisponibilité, de pharmacocinétique et de pharmacodynamie des ligands candidats, nous avons employé des modèles de criblage virtuel rigoureux de type Drug-like et ADME-Tox. Nous avons entériné avec succès la stratégie hybride de CADD à grande échelle, identifiant de nouveaux agents thérapeutiques potentiels pour le cancer et le SARS-CoV-2. Ces résultats ont contribué à l'optimisation et à l'identification d'un ensemble de ligands prometteurs, synthétisés chimiquement ou provenant de substances phytochimiques, capables d'agir comme des inhibiteurs spécifiques de récepteurs biologiques impliqués dans la propagation de lignées cellulaires cancéreuses et virales. Par exemple, dans le cas du cancer, nous avons proposé de nouvelles générations d'inhibiteurs potentiels de l'activité enzymatique de récepteurs tels que la Tyrosine kinase c-Met, la Protéase MALT1 et la Tyrosine kinase EGFR, ciblés par des dérivés synthétiques de Dimédone, 4-phénoxypyridine, Cyclohexane-1,3-dione, et Pyridine, ainsi que des molécules phytochimiques isolées à partir de *Magnolia officinalis* et *Cannabis sativa L*. Parallèlement avec la pandémie de SARS-CoV-2 (COVID-19), nous avons également proposé une nouvelle génération d'inhibiteurs de l'enzyme $3CL^{pro}$ à base de 9,10-Dihydrophénanthrène. Les résultats de l'application de la stratégie hybride de CADD ont montré des performances prometteuses en termes de fiabilité, de rapidité, de robustesse et de rationalité dans le criblage virtuel et l'optimisation biomoléculaire *in silico*. Pour l'avenir, nous envisageons d'étendre cette stratégie à d'autres protéines cibles et d'utiliser de nouvelles techniques de modélisation moléculaire pour améliorer la conception précoce de nouveaux médicaments. La validation *in vitro* et *in vivo* des ligands identifiés pourrait conduire à des avancées significatives dans

la recherche pharmaceutique. De plus, l'étude des propriétés physiques et chimiques des molécules ouvre de nouvelles perspectives dans des domaines tels que les insecticides et les énergies renouvelables. En conclusion, cette thèse représente une avancée majeure dans le domaine de la recherche pharmaceutique, soulignant le potentiel des techniques de modélisation moléculaire pour relever les défis de santé actuels et futur.

LISTE DES PUBLICATIONS

Articles dans des revues internationales avec comités de lecture

1. **O. Daoui**, S. Elkhattabi, S. Chtita, R. Elkhalabi, H. Zgou, A.T. Benjelloun, QSAR, molecular docking and ADMET properties *in silico* studies of novel 4,5,6,7-tetrahydrobenzo[D]-thiazol-2-Yl derivatives derived from dimedone as potent anti-tumor agents through inhibition of C-Met receptor tyrosine kinase, **Heliyon**. 7 (**2021**) e07463. https://doi.org/10.1016/j.heliyon.2021.e07463.

2. **O. Daoui**, S. Elkhattabi, S. Chtita, Rational identification of small molecules derived from 9,10-dihydrophenanthrene as potential inhibitors of $3CL^{pro}$ enzyme for COVID-19 therapy: a computer-aided drug design approach, **Structural Chemistry**. 33 (**2022**) 1667–1690. https://doi.org/10.1007/s11224-022-02004-z.

3. **O. Daoui**, S. Elkhattabi, S. Chtita, Rational design of novel pyridine-based drugs candidates for lymphoma therapy, **Journal of Molecular Structure**. 1270 (**2022**) 133964. https://doi.org/10.1016/j.molstruc.2022.133964.

4. **O. Daoui**, R. Haloui, K. Mkhayar, O. Abchir, H. Nour, S. Elkhattabi, S. Chtita, QSAR modeling, pharmacokinetics and molecular docking investigations of novel Neocryptolepine derivatives as Hits against HepG2 Cells, **RHAZES: Green and Applied Chemistry**. 16 (**2022**) 46–62. https://doi.org/10.48419/IMIST.PRSM/rhazes-v16.35254.

5. **O. Daoui**, K. Mkhayar, S. Elkhatabi, S. Chtita, H. Zgou, K. Elkhalabi, Design of novel carbocycle-fused quinoline derivatives as potential inhibitors of lymphoblastic leukemia cell line MOLT-3 using 2D-QSAR and ADME-Tox studies, **RHAZES: Green and Applied Chemistry**. 14 (**2022**) 36–61. https://doi.org/10.48419/IMIST.PRSM/rhazes-v14.31152.

6. **O. Daoui**, H. Nour, O. Abchir, S. Elkhattabi, M. Bakhouch, S. Chtita, A computer-aided drug design approach to explore novel type II inhibitors of c-Met receptor tyrosine kinase for cancer therapy: QSAR, molecular docking, ADMET and molecular dynamics simulations, **Journal of Biomolecular Structure and Dynamics**. 0 (**2022**) 1–18. https://doi.org/10.1080/07391102.2022.2124456.

7. **O. Daoui**, N. Mazoir, M. Bakhouch, M. Salah, A. Benharref, A. Gonzalez-Coloma, S. Elkhattabi, M.E. Yazidi, S. Chtita, 3D-QSAR, ADME-Tox, and molecular docking of semisynthetic triterpene derivatives as antibacterial and insecticide agents, **Structural Chemistry**. 33 (**2022**) 1063–1084. https://doi.org/10.1007/s11224-022-01912-4.

8. **O. Daoui**, R. El Mouhi, F. Barghady, K. Mkhayar, S. Elkhattabi, S. Chtita, R. Elkhalabi, A.T. Benjelloun, 3D-QSAR modeling, Molecular Docking and drug-like properties investigations of novel heterocyclic compounds derived from Magnolia Officinalis as Hit Compounds against NSCLC, **Moroccan Journal of Chemistry**. 10 (**2022**) 10–4. https://doi.org/10.48317/IMIST.PRSM/morjchem-v10i4.34498.

9. **O. Daoui**, S. Elkhattabi, S. Chtita, Design and Prediction of ADME/Tox Properties of Novel Magnolol Derivatives as Anticancer Agents for NSCLC Using 3D-QSAR, Molecular Docking, MOLCAD and MM-GBSA Studies, **Letters in Drug Design & Discovery**. 20 (**2023**) 545–569. http://dx.doi.org/10.2174/1570180819666220510141710.

10. **O. Daoui**, S. Elkhattabi, M. Bakhouch, S. Belaidi, R.R. Bhandare, A.B. Shaik, S.N. Mali, S. Chtita, Cyclohexane-1,3-dione Derivatives as Future Therapeutic Agents for NSCLC: QSAR Modeling, *In Silico* ADME-Tox Properties, and Structure-Based Drug Designing Approach, **ACS Omega**. (**2023**). https://doi.org/10.1021/acsomega.2c07585.

11. **O. Daoui**, S.N. Mali, K. Elkhattabi, S. Elkhattabi, S. Chtita, Repositioning Cannabinoids and Terpenes as Novel EGFR-TKIs Candidates for Targeted Therapy Against Cancer: A virtual screening model using CADD and biophysical simulations, **Heliyon**. 9 (**2023**) e15545. https://doi.org/10.1016/j.heliyon.2023.e15545.

12. R.T. Fouedjou, H.P.D. Fogang, M. Ouassaf, **O. Daoui**, F.A. Qais, S. Elkhattabi, M. Bakhouch, S. Belaidi, S. Chtita, R.T. Fouedjou, H.P.D. Fogang, M. Ouassaf, O. Daoui, F.A. Qais, S. Elkhattabi, M. Bakhouch, S. Belaidi, S. Chtita, Targeting the Main Protease and The Spike Protein of Sars-Cov-2 With Naturally Occurring Compounds from Some Cameroonian Medicinal Plants: An In-Silico Study for Drug

Designing, **Journal of the Chilean Chemical Society**. 67 (**2022**) 5602–5614. https://doi.org/10.4067/S0717-97072022000305602.

13. K. Mkhayar, **O. Daoui**, S. Khattabi, S. Chtita, R. Elkhalabi, QSAR study and Molecular Docking assisted design of novel cyclohexane-1, 3-dione Derivatives as anticancer agents for non-small cell lung cancer, **Biointerface Research in Applied Chemistry**. 30 (**2022**). https://doi.org/10.33263/BRIAC136.524.

14. M. Ouassaf, **O. Daoui**, S. Alam, S. Elkhattabi, S. Belaidi, S. Chtita, Pharmacophore-based virtual screening, molecular docking, and molecular dynamics studies for the discovery of novel FLT3 inhibitors, **Journal of Biomolecular Structure and Dynamics**. 0 (**2022**) 1–13. https://doi.org/10.1080/07391102.2022.2123403.

15. M. Chalkha, M. Akhazzane, F.Z. Moussaid, **O. Daoui**, A. Nakkabi, M. Bakhouch, S. Chtita, S. Elkhattabi, A.I. Housseini, M. El Yazidi, Design, synthesis, characterization, *in vitro* screening, molecular docking, 3D-QSAR, and ADME-Tox investigations of novel pyrazole derivatives as antimicrobial agents, **New Journal of Chemistry**. 46 (**2022**) 2747–2760. https://doi.org/10.1039/D1NJ05621B.

16. O. Abchir, **O. Daoui**, S. Belaidi, M. Ouassaf, F.A. Qais, S. ElKhattabi, S. Belaaouad, S. Chtita, Design of novel benzimidazole derivatives as potential α-amylase inhibitors using QSAR, pharmacokinetics, molecular docking, and molecular dynamics simulation studies, **Journal of Molecular Modeling**. 28 (**2022**) 1–17. https://doi.org/10.1007/s00894-022-05097-9.

17. H. Nour, **O. Daoui**, O. Abchir, S. ElKhattabi, S. Belaidi, S. Chtita, Combined computational approaches for developing new anti-Alzheimer drug candidates: 3D-QSAR, molecular docking and molecular dynamics studies of liquiritigenin derivatives, **Heliyon**. 8 (**2022**). https://doi.org/10.1016/j.heliyon.2022.e11991.

18. R. Haloui, **O. Daoui**, K. Mkhayar, M. El Yaqoubi, S. Elkhattabi, A. Haoudi, Y.K. Rodi, F.C. Ouazzani, S. Chtita, 3D-QSAR, drug-likeness, ADMET prediction, and molecular docking studies *in silico* of novel 5-oxo-1-thioxo-4,5-dihydro-1H-thiazolo[3,4-a]quinazoline derivatives as MALT1 protease inhibitors for the treatment of B cell lymphoma, **Chemical Papers**. (**2022**). https://doi.org/10.1007/s11696-022-02627-w.

19. M. Bouhadi, **O. Daoui**, H. El Hajjouji, S. Elkhattabi, S. Chtita, M. El Kouali, M. Talbi, H. Fougrach, Study of the competition between Pi and Cr (VI) for the use of Pi-transporter at Vicia faba L. using molecular modeling**, Plant Physiology and Biochemistry**. 196 (**2023**) 695–702. https://doi.org/10.1016/j.plaphy.2023.02.029.

20. R.E. Mouhi, **O. Daoui**, A. Fitri, A.T. Benjelloun, S.E. Khattabi, M. Benzakour, M. Mcharfi, M. Kurban, A strategy to enhance VOC of π-conjugated molecules based on thieno[2,3-b] indole for applications in bulk heterojunction organic solar cells using DFT, TD-DFT, and 3D-QSPR modeling studies, **New Journal of Chemistry**. 47 (**2023**) 812–827. https://doi.org/10.1039/D2NJ04281A.

21. O. Abchir, I. Yamari, H. Nour, **O. Daoui**, S. Elkhattabi, A. Errougui, S. Chtita, Structure-Based Virtual Screening, ADMET analysis, and Molecular Dynamics Simulation of Moroccan Natural Compounds as Candidates α-Amylase Inhibitors, **ChemistrySelect**. 8 (**2023**) e202301092. https://doi.org/10.1002/slct.202301092.

22. O. Abchir, H. Nour, **O. Daoui**, I. Yamari, S. ElKhattabi, M. El Kouali, M. Talbi, A. Errougui, S. Chtita, Structure-based virtual screening, ADMET analysis, and molecular dynamics simulation of Moroccan natural compounds as candidates for the SARS-CoV-2 inhibitors, **Natural Product Research**. 0 (**2023**) 1–8. https://doi.org/10.1080/14786419.2023.2281002.

23. I. Hammoudan, S. Chtita, **O. Daoui**, S. Elkhattabi, M. Bakhouch, M. El Yazidi, F. Siddique, D. Riffi-Temsamani, QSAR and DFT Studies of Some Tacrine-Hydroxamate Derivatives as Inhibitors of Cholinesterase (AChEs) in the Treatment of Alzheimer's Disease, **Letters in Drug Design & Discovery**. 20 (**2023**) 699–712. https://doi.org/10.2174/1570180819666220512174409.

24. M. Bouhadi, **O. Daoui**, H. El Hajjouji, S. Elkhattabi, S. Chtita, M. Talbi, H. Fougrach, Physiological and molecular modeling investigations of the relationship between sulfate and chromium VI uptake in Vicia faba L., **Biocatalysis and Agricultural Biotechnology**. 47 (**2023**) 102554. https://doi.org/10.1016/j.bcab.2022.102554.

25. K. Mkhayar, R. Haloui, **O. Daoui**, K. Elkhattabi, S. Chtita, S. Elkhattabi, *In silico* studies of 2-aryloxy-1,4- naphthoquinone derivatives as antibacterial agents against Escherichia coli using 3D-QSAR, ADMET properties, molecular docking, and molecular dynamics, **Chemical Data Collections**. 47 (**2023**) 101060. https://doi.org/10.1016/j.cdc.2023.101060.

26. R.T. Fouedjou, **O. Daoui**, H. Nour, M. Ayoub, H.P.D. Fogang, F. Siddique, S. Elkhattabi, M. Bakhouch, S. Belaidi, S. Chtita, *In Silico* Approach for Designing Novel SARS-CoV-2 Inhibitors from Medicinal Plants, **Physical Chemistry Research**. 11 (**2023**) 589–604. https://doi.org/10.22036/pcr.2022.349693.2138.

27. O. Abchir, **O. Daoui**, H. Nour, I. Yamari, S. Elkhattabi, A. Errougui, S. Chtita, Exploration of Cannabis constituents as potential candidates against diabetes mellitus disease using molecular docking, dynamics simulations and ADMET investigations, **Scientific African**. 21 (**2023**) e01745. https://doi.org/10.1016/j.sciaf.2023.e01745.

28. K. Mkhayar, K. Elkhattabi, R. Elkhalabi, R. Haloui, **O. Daoui**, E.I. Edache, S. Chtita, S. Elkhattabi, Evaluation of dimedone-derived compounds as inhibitors against human colon cancer: Insights from 2D-QSAR, ADMET prediction, Osiris, Molinspiration, and molecular modeling, **Chinese Journal of Analytical Chemistry**. 51 (**2023**) 100330. https://doi.org/10.1016/j.cjac.2023.100330.

29. R. Haloui, K. Mkhayar, **O. Daoui**, Y. Rhazi, K. ElKhattabi, M.E. Yazidi, S. Chtita, A. Haoudi, S. Elkhattabi, Computational Study on Benzofuran Derivatives as Potent Lysine-Specific Demethylase 1 Inhibitors using 3D QSAR, ADMET Prediction, and Molecular Docking, **Letters in Drug Design & Discovery**. 20 (**2023**) 1–1. http://dx.doi.org/10.2174/1570180820666230825122908.

30. R. Haloui, K. ElKhattabi, K. Mkhayar, **O. Daoui**, S. Chtita, A. Haoudi, S. Elkhattabi, Design of novel small molecules derived from styrylpyridine as potent HDAC1 inhibitors for the treatment of gastric cancer using 3D-QSAR, drug similarity, ADMET prediction, molecular docking, and molecular dynamics studies, **Scientific African**. 23 (**2024**) e01990. https://doi.org/10.1016/j.sciaf.2023.e01990.

31. R. Haloui, K. Mkhayar, **O. Daoui**, K. El Khattabi, A. El Abbouchi, S. Chtita, S. Elkhattabi, Design of new small molecules derived from indolin-2-one as potent TRKs inhibitors using a computer-aided drug design approach, **Journal of Biomolecular Structure and Dynamics** (**2024**) 1–18. https://doi.org/10.1080/07391102.2024.2302944.

32. K. Mkhayar, **O. Daoui**, R. Haloui, K. Elkhattabi, A. Elabbouchi, S. Chtita, A. Samadi, S. Elkhattabi, Ligand-Based Design of Novel Quinoline Derivatives as Potential Anticancer Agents: An In-Silico Virtual Screening Approach, **Molecules** 29 (**2024**) 426. https://doi.org/10.3390/molecules29020426.

33. M. Manachou, **O. Daoui**, O. Abchir, R. Dahmani, S. Elkhattabi, A. Samadi, S. Belaidi, S. Chtita, An antibacterial lead identification of novel 1,3,4-oxadiazole derivatives based on molecular computer aided design approaches, **Scientific African** 23 (**2024**) e02078. https://doi.org/10.1016/j.sciaf.2024.e02078.

34. N. Mounadi, H. Nour, **O. Daoui**, S. Elkhattabi, A. Errougui, M. Talbi, M. El Kouali, S. Chtita, Computational Studies of Cannabis Derivatives as Potential Inhibitors of SARS-CoV-2 Mpro, **Chemistry Africa** (**2024**). https://doi.org/10.1007/s42250-024-00914-5.

35. S. El Rhabori, A. El Aissouq, **O. Daoui**, S. Elkhattabi, S. Chtita, F. Khalil, Design of new molecules against cervical cancer using DFT, theoretical spectroscopy, 2D/3D-QSAR, molecular docking, pharmacophore and ADMET investigations, **Heliyon** 10 **(2024)** e24551. https://doi.org/10.1016/j.heliyon.2024.e24551.

Communications Internationales avec actes

1. **O. Daoui**, R. El Mouhi, F. Barghady, K. Mkhayar, S. Elkhattabi), S. Chtita, R. Elkhalabi , A. Touimi Benjelloun, 3D-QSAR modeling, Molecular Docking and drug-like properties investigations of novel heterocyclic compounds derived from Magnolia Officinalis as Hit Compounds against NSCLC, in: Environment and Natural Resources: Challenges and Solutions (**ENRSC2021**), **Mor.J.Chem**., **2022**: 881-890. https://revues.imist.ma/index.php/morjchem/article/view/34498

2. **O. Daoui**, R. Haloui, K. Mkhayar, O. Abchir, H. Nour, S. Elkhattabi, S. Chtita, QSAR modeling, pharmacokinetics and molecular docking investigations of novel Neocryptolepine derivatives as Hits against HepG2 Cells, in: Colloque Maghrébin sur la Chimie Hétérocyclique (**CMCH 5-2022**), **RHAZES**, **2022**: pp. 46-62. https://doi.org/10.48419/IMIST.PRSM/rhazes-v16.35254.

3. K. Mkhayar, **O. Daoui**, S. Elkhattabi, S. Chtita, R. Elkhalabi, *In silico* molecular investigations of derived cyclohexane-1,3-dione compounds as potential inhibitors of protein tyrosine kinase C-met: 2D QSAR, molecular docking and ADMET, in: 2022 International Conference on Intelligent Systems and Computer Vision (**ISCV**), **2022**: pp. 1–8. https://doi.org/10.1109/ISCV54655.2022.9806058.

4. O. Abchir, H. Nour, **O. Daoui**, S.E. Khattabi, S. Chtita, Identification of New Anti-SARS CoV-2 Agents through the Virtual Screening of Phytoconstituents Extracted from Moroccan Plants, **Medical Sciences Forum**. 14 (**2022**) 48. https://doi.org/10.3390/ECMC2022-13480.

5. K. Mkhayar, **O. Daoui**, R. Haloui, S. Elkhattabi, S. Chtita, Design of New Derivatives of Dimedone Molecules Using QSAR and Docking Molecular, **Medical Sciences Forum**. 14 (**2022**) 106. https://doi.org/10.3390/ECMC2022-13245.

6. O. Abchir, **O. Daoui**, S. Belaidi, M. Ouassaf, F. Abul qais, S. Elkhattabi, S. Chtita, Conception de nouveaux dérivés du benzimidazole en tant qu'inhibiteurs potentiels α l'amylase à l'aide d'études de simulation QSAR, pharmacocinétique, Docking moléculaire, in : Colloque Maghrébin sur la Chimie Hétérocyclique (**CMCH 5-2022**).

7. K. Mkhayar, **O. Daoui**, R. Haloui, S. Elkhattabi, S. Chtita, R. Elkhalabi, QSAR and Docking molecular studies of derived cyclohexane-1,3-dione compounds, in : Colloque Maghrébin sur la Chimie Hétérocyclique (**CMCH 5-2022**).

8. O. Abchir, **O. Daoui**, H. Nour, I. Yamari, S. Elkhattabi, A. Errougui, S. Chtita, Cannabis Constituents as Potential Candidates of Alpha Amylase Inhibitor, In: **8th INTERNATIONAL CONGRESS ON INNOVATIVE SCIENTIFIC APPROACHES** November 23-24, **2022** Samsun, Turkiye.

9. R. Haloui, K. Mkhayar, **O. Daoui**, K. ElKhattabi, S. Chtita, A. Haoudi, S. Elkhattabi, Identification of New Indolin-2-One Derivatives as Potent Trks Inhibitors Using 3d-Qsar, Admet Prediction, Molecular Docking, And Molecular Dynamics Studies, in: **5th International Harran Congress on Scientific Research** held on December 8-10, **2023** / Şanlıurfa, Türkiye.

RÉFÉRENCES

[1] M. Chan, S. Vijay, M. J. McElrath, E. C. Holland, et T. S. Gujral, « Machine Learning Identifies Ponatinib as a Potent Inhibitor of SARS-CoV2-induced Cytokine Storm ». bioRxiv, p. 2021.04.07.438871, 9 avril 2021. doi: 10.1101/2021.04.07.438871.

[2] D. D. Nguyen, K. Gao, J. Chen, R. Wang, et G.-W. Wei, « Potentially highly potent drugs for 2019-nCoV ». bioRxiv, p. 2020.02.05.936013, 13 février 2020. doi: 10.1101/2020.02.05.936013.

[3] C. Gao, *Computer-aided drug design approaches in developing anti-cancer inhibitors*. 2016.

[4] O. J. Garden, R. W. Parks, et S. J. Wigmore, *Principles and Practice of Surgery, E-Book*. Elsevier Health Sciences, 2022.

[5] C. M. Washington et D. T. Leaver, *Principles and practice of radiation therapy-e-book*. Elsevier Health Sciences, 2015.

[6] J. A. McKnight, « Principles of chemotherapy », *Clinical techniques in small animal practice*, vol. 18, nº 2, p. 67-72, 2003.

[7] P. Anand, S. Soni, et Y. S. Padwad, « Targeted Molecular Therapies in Cancer », *Mod. Cancer Ther. Tradit. Med. An Integr. Approach to Combat Cancers*, p. 59-101, 2021.

[8] R. J. Baber, N. Panay, et A. Fenton, « 2016 IMS Recommendations on women's midlife health and menopause hormone therapy », *Climacteric*, vol. 19, nº 2, p. 109-150, 2016.

[9] M. bin Umair *et al.*, « Viruses as tools in gene therapy, vaccine development, and cancer treatment », *Archives of Virology*, vol. 167, nº 6, p. 1387-1404, 2022.

[10] C. Thakur, « Chapter 2 - An Overview, Current Challenges of Drug Resistance, and Targeting Metastasis Associated With Lung Cancer », in *Nanotechnology-Based Targeted Drug Delivery Systems for Lung Cancer*, P. Kesharwani, Éd., Academic Press, 2019, p. 21-38. doi: 10.1016/B978-0-12-815720-6.00002-2.

[11] K. Stegmaier et W. R. Sellers, « 4 - Targeted Approaches to Drug Development », in *Oncology of Infancy and Childhood*, S. H. Orkin, D. E. Fisher, A. T. Look, S. E. Lux, D. Ginsburg, et D. G. Nathan, Éd., Philadelphia: W.B. Saunders, 2009, p. 57-98. doi: 10.1016/B978-1-4160-3431-5.00004-2.

[12] E. M. Michalak, M. L. Burr, A. J. Bannister, et M. A. Dawson, « The roles of DNA, RNA and histone methylation in ageing and cancer », *Nature reviews Molecular cell biology*, vol. 20, nº 10, p. 573-589, 2019.

[13] C. T. Chasapis, P.-S. A. Ntoupa, C. A. Spiliopoulou, et M. E. Stefanidou, « Recent aspects of the effects of zinc on human health », *Archives of toxicology*, vol. 94, p. 1443-1460, 2020.

[14] A. Dendebera, « Enzymes: how they work and what they do », in *Інноваційні тенденції підготовки фахівців в умовах полікультурного та мультилінгвального глобалізованого світу*, Київський національний університет технологій та дизайну, 2019.

[15] D. E. Koshland Jr, « Application of a theory of enzyme specificity to protein synthesis », *Proceedings of the National Academy of Sciences*, vol. 44, nº 2, p. 98-104, 1958.

[16] Y. Cheng *et al.*, « Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials », *Signal transduction and targeted therapy*, vol. 4, nº 1, p. 1-39, 2019.

[17] V. Gray-Schopfer, C. Wellbrock, et R. Marais, « Melanoma biology and new targeted therapy », *Nature*, vol. 445, nº 7130, Art. nº 7130, févr. 2007, doi: 10.1038/nature05661.

[18] G. Bjelaković, I. Stojanović, G. B. Bjelaković, D. Pavlović, G. Kocić, et A. Daković-Milić, « Competitive inhibitors of enzymes and their therapeutic application », *Med Biol*, vol. 9, nº 3, p. 201-206, 2002.

[19] Y. Blat, « Non-competitive inhibition by active site binders », *Chemical Biology & Drug Design*, vol. 75, nº 6, p. 535-540, 2010.

[20] Lakna, « Difference Between Reversible and Irreversible Enzyme Inhibition », Pediaa.Com. Consulté le: 30 décembre 2022. [En ligne]. Disponible sur: https://pediaa.com/difference-between-reversible-and-irreversible-enzyme-inhibition/

[21] F. D. Makurvet, « Biologics vs. small molecules: Drug costs and patient access », *Medicine in Drug Discovery*, vol. 9, p. 100075, mars 2021, doi: 10.1016/j.medidd.2020.100075.

[22] R. Ganellin, R. Jefferies, et S. M. Roberts, *Introduction to Biological and Small Molecule Drug Research and Development: Theory and Case Studies*. Academic Press, 2013.

[23] P. K. Patel, C. R. King, et S. R. Feldman, « Biologics and biosimilars », *Journal of Dermatological Treatment*, vol. 26, nº 4, p. 299-302, juill. 2015, doi: 10.3109/09546634.2015.1054782.

[24] P. J. Declerck *et al.*, « Biosimilars: controversies as illustrated by rhGH », *Current Medical Research and Opinion*, vol. 26, nº 5, p. 1219-1229, mai 2010, doi: 10.1185/03007991003719642.
[25] L. Ruddigkeit, R. van Deursen, L. C. Blum, et J.-L. Reymond, « Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17 », *J. Chem. Inf. Model.*, vol. 52, nº 11, p. 2864-2875, nov. 2012, doi: 10.1021/ci300415d.
[26] « Small vs Big: Understanding the Differences between Small Molecule Drugs and Biologic Drugs – IMMpress Magazine ». Consulté le: 28 octobre 2022. [En ligne]. Disponible sur: https://www.immpressmagazine.com/small-vs-big-understanding-the-differences-between-small-molecule-drugs-and-biologic-drugs/
[27] P. J. Declerck, « Biologicals and biosimilars: a review of the science and its implications », *GaBI J*, vol. 1, nº 1, p. 13-16, 2012.
[28] A. B. Deore, J. R. Dhumane, R. Wagh, et R. Sonawane, « The stages of drug discovery and development process », *Asian Journal of Pharmaceutical Research and Development*, vol. 7, nº 6, p. 62-67, 2019.
[29] T. Anwar, P. Kumar, et A. U. Khan, « Modern Tools and Techniques in Computer-Aided Drug Design », in *Molecular Docking for Computer-Aided Drug Design*, Elsevier, 2021, p. 1-30.
[30] Y. Zhang *et al.*, « A combined drug discovery strategy based on machine learning and molecular docking », *Chemical Biology & Drug Design*, vol. 93, nº 5, p. 685-699, 2019, doi: 10.1111/cbdd.13494.
[31] R. Lonsdale et R. A. Ward, « Structure-based design of targeted covalent inhibitors », *Chemical Society Reviews*, vol. 47, nº 11, p. 3816-3830, 2018.
[32] A. Aljoundi, I. Bjij, A. El Rashedy, et M. E. S. Soliman, « Covalent Versus Non-covalent Enzyme Inhibition: Which Route Should We Take? A Justification of the Good and Bad from Molecular Modelling Perspective », *Protein J*, vol. 39, nº 2, p. 97-105, avr. 2020, doi: 10.1007/s10930-020-09884-2.
[33] R. Sharma, *Enzyme Inhibition and Bioapplications*. BoD – Books on Demand, 2012.
[34] L. Boike, N. J. Henning, et D. K. Nomura, « Advances in covalent drug discovery », *Nat Rev Drug Discov*, vol. 21, nº 12, Art. nº 12, déc. 2022, doi: 10.1038/s41573-022-00542-z.
[35] A. A. Adeniyi, R. Muthusamy, et M. E. Soliman, « New drug design with covalent modifiers », *Expert opinion on drug discovery*, vol. 11, nº 1, p. 79-90, 2016.
[36] J. G. Robertson, « Mechanistic basis of enzyme-targeted drugs », *Biochemistry*, vol. 44, nº 15, p. 5561-5571, 2005.
[37] C. Lipinski et A. Hopkins, « Navigating chemical space for biology and medicine », *Nature*, vol. 432, nº 7019, p. 855-861, 2004.
[38] R. S. Herbst, M. Fukuoka, et J. Baselga, « Gefitinib—a novel targeted approach to treating cancer », *Nature Reviews Cancer*, vol. 4, nº 12, p. 956-965, 2004.
[39] F. A. Shepherd *et al.*, « Erlotinib in Previously Treated Non–Small-Cell Lung Cancer », *New England Journal of Medicine*, vol. 353, nº 2, p. 123-132, juill. 2005, doi: 10.1056/NEJMoa050753.
[40] S. J. Rodig et G. I. Shapiro, « Crizotinib, a small-molecule dual inhibitor of the c-Met and ALK receptor tyrosine kinases », *Curr Opin Investig Drugs*, vol. 11, nº 12, p. 1477-1490, déc. 2010.
[41] N. B. Leighl *et al.*, « A phase I study of foretinib plus erlotinib in patients with previously treated advanced non-small cell lung cancer: Canadian cancer trials group IND.196 », *Oncotarget*, vol. 8, nº 41, p. 69651-69662, sept. 2017, doi: 10.18632/oncotarget.18753.
[42] R. Gabizon *et al.*, « Efficient targeted degradation via reversible and irreversible covalent PROTACs », *Journal of the American Chemical Society*, vol. 142, nº 27, p. 11734-11742, 2020.
[43] A. Abdeldayem, Y. S. Raouf, S. N. Constantinescu, R. Moriggl, et P. T. Gunning, « Advances in covalent kinase inhibitors », *Chemical Society Reviews*, vol. 49, nº 9, p. 2617-2687, 2020.
[44] G. A. Holdgate, T. D. Meek, et R. L. Grimley, « Mechanistic enzymology in drug discovery: a fresh perspective », *Nat Rev Drug Discov*, vol. 17, nº 2, Art. nº 2, févr. 2018, doi: 10.1038/nrd.2017.219.
[45] S. De Cesco, J. Kurian, C. Dufresne, A. K. Mittermaier, et N. Moitessier, « Covalent inhibitors design and discovery », *European Journal of Medicinal Chemistry*, vol. 138, p. 96-114, sept. 2017, doi: 10.1016/j.ejmech.2017.06.019.
[46] U. H. Lodish *et al.*, *Molecular Cell Biology*. W. H. Freeman, 2008.
[47] H. C. Cheng, « The power issue: determination of KB or Ki from IC50 », *Journal of Pharmacological and Toxicological Methods*, vol. 46, nº 2, p. 61-71, sept. 2001, doi: 10.1016/S1056-8719(02)00166-1.

[48] B.-F. Krippendorff, R. Neuhaus, P. Lienau, A. Reichel, et W. Huisinga, « Mechanism-based inhibition: deriving KI and kinact directly from time-dependent IC50 values », *SLAS Discovery*, vol. 14, nº 8, p. 913-923, 2009.
[49] M. Vailati-Riboni, V. Palombo, et J. J. Loor, « What Are Omics Sciences? », *Periparturient Diseases of Dairy Cows*, p. 1-7, 2017, doi: 10.1007/978-3-319-43033-1_1.
[50] H. Matthews, J. Hanison, et N. Nirmalan, « "Omics"-Informed Drug and Biomarker Discovery: Opportunities, Challenges and Future Perspectives », *Proteomes*, vol. 4, nº 3, Art. nº 3, sept. 2016, doi: 10.3390/proteomes4030028.
[51] A. S. Kesselheim, M. S. Sinha, J. Avorn, et A. Sarpatwari, « Pharmaceutical policy in the United States in 2019: An overview of the landscape and avenues for improvement », *Stan. L. & Pol'y Rev.*, vol. 30, p. 421, 2019.
[52] « What Is ADME? », Drug Discovery from Technology Networks. Consulté le: 18 mai 2023. [En ligne]. Disponible sur: http://www.technologynetworks.com/drug-discovery/articles/what-is-adme-336683
[53] S. W. Baran *et al.*, « Emerging role of translational digital biomarkers within home cage monitoring technologies in preclinical drug discovery and development », *Frontiers in Behavioral Neuroscience*, vol. 15, p. 380, 2022.
[54] D. Sun, W. Gao, H. Hu, et S. Zhou, « Why 90% of clinical drug development fails and how to improve it? », *Acta Pharmaceutica Sinica B*, vol. 12, nº 7, p. 3049-3062, juill. 2022, doi: 10.1016/j.apsb.2022.02.002.
[55] H. Dowden et J. Munro, « Trends in clinical success rates and therapeutic focus », *Nat. Rev. Drug Discov*, vol. 18, nº 7, p. 495-6, 2019.
[56] R. K. Harrison, « Phase II and phase III failures: 2013–2015 », *Nat Rev Drug Discov*, vol. 15, nº 12, p. 817-818, 2016.
[57] K.-K. Mak et M. R. Pichika, « Artificial intelligence in drug development: present status and future prospects », *Drug discovery today*, vol. 24, nº 3, p. 773-780, 2019.
[58] W. Yu et A. D. MacKerell, « Computer-aided drug design methods », in *Antibiotics*, Springer, 2017, p. 85-106.
[59] S. J. Y. Macalino, V. Gosu, S. Hong, et S. Choi, « Role of computer-aided drug design in modern drug discovery », *Archives of pharmacal research*, vol. 38, nº 9, p. 1686-1701, 2015.
[60] T. Anwar, P. Kumar, et A. U. Khan, « Chapter 1 - Modern Tools and Techniques in Computer-Aided Drug Design », in *Molecular Docking for Computer-Aided Drug Design*, M. S. Coumar, Éd., Academic Press, 2021, p. 1-30. doi: 10.1016/B978-0-12-822312-3.00011-4.
[61] S. Surabhi et B. K. Singh, « COMPUTER AIDED DRUG DESIGN: AN OVERVIEW », *Journal of Drug Delivery and Therapeutics*, vol. 8, nº 5, Art. nº 5, sept. 2018, doi: 10.22270/jddt.v8i5.1894.
[62] K. L. Damm-Ganamet *et al.*, « Accelerating Lead Identification by High Throughput Virtual Screening: Prospective Case Studies from the Pharmaceutical Industry », *J. Chem. Inf. Model.*, vol. 59, nº 5, p. 2046-2062, mai 2019, doi: 10.1021/acs.jcim.8b00941.
[63] « Compound Libraries | TargetMol ». Consulté le: 4 novembre 2022. [En ligne]. Disponible sur: https://www.targetmol.com/all-compound-libraries?gclid=CjwKCAjw8JKbBhBYEiwAs3sxN9-hoFa2Bd9t6ustOfmW1xft4aLnECLvqrId-WnW2xvE-3W-ZDkIShoCmAAQAvD_BwE
[64] R. P. D. Bank, « RCSB PDB: Homepage ». Consulté le: 14 décembre 2022. [En ligne]. Disponible sur: https://www.rcsb.org/
[65] V. Rantos, K. Karius, et J. Kosinski, « Integrative structural modeling of macromolecular complexes using Assembline », *Nature Protocols*, vol. 17, nº 1, p. 152-176, 2022.
[66] S. Kalyaanamoorthy et Y.-P. P. Chen, « Structure-based drug design to augment hit discovery », *Drug Discovery Today*, vol. 16, nº 17, p. 831-839, sept. 2011, doi: 10.1016/j.drudis.2011.07.006.
[67] A. Najmi, S. A. Javed, M. Al Bratty, et H. A. Alhazmi, « Modern Approaches in the Discovery and Development of Plant-Based Natural Products and Their Analogues as Potential Therapeutic Agents », *Molecules*, vol. 27, nº 2, Art. nº 2, janv. 2022, doi: 10.3390/molecules27020349.
[68] A. N. Lima, E. A. Philot, G. H. G. Trossini, L. P. B. Scott, V. G. Maltarollo, et K. M. Honorio, « Use of machine learning approaches for novel drug discovery », *Expert Opinion on Drug Discovery*, vol. 11, nº 3, p. 225-239, mars 2016, doi: 10.1517/17460441.2016.1146250.
[69] A. Nayarisseri *et al.*, « Artificial Intelligence, Big Data and Machine Learning Approaches in Precision Medicine & Drug Discovery », *Current Drug Targets*, vol. 22, nº 6, p. 631-655, avr. 2021, doi: 10.2174/1389450122999210104205732.
[70] V. Sharma, S. Wakode, et H. Kumar, « Structure-and ligand-based drug design: Concepts, approaches, and challenges », *Chemoinformatics and bioinformatics in the pharmaceutical sciences*, p. 27-53, 2021.

[71] A. Raafat, S. Mowafy, S. M. Abouseri, M. A. Fouad, et N. A. Farag, « Lead generation of cysteine based mesenchymal epithelial transition (c-Met) kinase inhibitors: Using structure-based scaffold hopping, 3D-QSAR pharmacophore modeling, virtual screening, molecular docking, and molecular dynamics simulation », *Computers in Biology and Medicine*, vol. 146, p. 105526, juill. 2022, doi: 10.1016/j.compbiomed.2022.105526.
[72] D. Vemula, P. Jayasurya, V. Sushmitha, Y. N. Kumar, et V. Bhandari, « CADD, AI and ML in drug discovery: A comprehensive review », *European Journal of Pharmaceutical Sciences*, vol. 181, p. 106324, févr. 2023, doi: 10.1016/j.ejps.2022.106324.
[73] R. Saravanan et P. Sujatha, « A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification », in *2018 Second international conference on intelligent computing and control systems (ICICCS)*, IEEE, 2018, p. 945-949. Consulté le: 20 décembre 2023. [En ligne]. Disponible sur: https://ieeexplore.ieee.org/abstract/document/8663155/
[74] S. Dara, S. Dhamercherla, S. S. Jadav, C. M. Babu, et M. J. Ahsan, « Machine Learning in Drug Discovery: A Review », *Artif Intell Rev*, vol. 55, nº 3, p. 1947-1999, mars 2022, doi: 10.1007/s10462-021-10058-4.
[75] G. Klambauer, S. Hochreiter, et M. Rarey, « Machine Learning in Drug Discovery », *J. Chem. Inf. Model.*, vol. 59, nº 3, p. 945-946, mars 2019, doi: 10.1021/acs.jcim.9b00136.
[76] A. Becker, « Artificial intelligence in medicine: What is it doing for us today? », *Health Policy and Technology*, vol. 8, nº 2, p. 198-205, 2019.
[77] K.-K. Mak et M. R. Pichika, « Artificial intelligence in drug development: present status and future prospects », *Drug Discovery Today*, vol. 24, nº 3, p. 773-780, mars 2019, doi: 10.1016/j.drudis.2018.11.014.
[78] R. León, J. Soto-Delgado, E. Montero, et M. Vargas, « Development of Computational Approaches with a Fragment-Based Drug Design Strategy: *In Silico* Hsp90 Inhibitors Discovery », *International Journal of Molecular Sciences*, vol. 22, nº 24, Art. nº 24, janv. 2021, doi: 10.3390/ijms222413226.
[79] O. Aybars, « Using neural nets to recognize handwritten digits », PhD Thesis, Zak\lad Systemów Przetwarzania Informacji, 2021.
[80] « Givan: An introduction to Markov decision processes - Google Scholar ». Consulté le: 26 février 2023. [En ligne]. Disponible sur: https://scholar.google.com/scholar_lookup?title=An%20Introduction%20to%20Markov%20decision%20Processes&publication_year=2001&author=B.%20Givan&author=R.%20Parr
[81] K. Arafet *et al.*, « Mechanism of inhibition of SARS-CoV-2 Mpro by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity », *Chem. Sci.*, vol. 12, nº 4, p. 1433-1444, févr. 2021, doi: 10.1039/D0SC06195F.
[82] D. D'Amato *et al.*, « Real-world experience with obeticholic acid in patients with primary biliary cholangitis », *JHEP Rep*, vol. 3, nº 2, p. 100248, janv. 2021, doi: 10.1016/j.jhepr.2021.100248.
[83] S. Dhillon, « Tirabrutinib: first approval », *Drugs*, vol. 80, nº 8, p. 835-840, 2020.
[84] C. Selvaraj, I. Chandra, et S. K. Singh, « Artificial intelligence and machine learning approaches for drug design: challenges and opportunities for the pharmaceutical industries », *Mol Divers*, vol. 26, nº 3, p. 1893-1913, juin 2022, doi: 10.1007/s11030-021-10326-z.
[85] M. M. Salman, Z. Al-Obaidi, P. Kitchen, A. Loreto, R. M. Bill, et R. Wade-Martins, « Advances in Applying Computer-Aided Drug Design for Neurodegenerative Diseases », *International Journal of Molecular Sciences*, vol. 22, nº 9, Art. nº 9, janv. 2021, doi: 10.3390/ijms22094688.
[86] C.-Y. Jia, J.-Y. Li, G.-F. Hao, et G.-F. Yang, « A drug-likeness toolbox facilitates ADMET study in drug discovery », *Drug Discovery Today*, vol. 25, nº 1, p. 248-258, 2020.
[87] J. Shi et W. Zha, « Predicting Human Pharmacokinetics: Physiologically Based Pharmacokinetic Modeling and In Silico ADME Prediction in Early Drug Discovery », *Eur J Drug Metab Pharmacokinet*, vol. 44, nº 1, p. 135-137, févr. 2019, doi: 10.1007/s13318-018-0503-9.
[88] S. Winiwarter *et al.*, « Prediction of Fraction Unbound in Microsomal and Hepatocyte Incubations: A Comparison of Methods across Industry Datasets », *Mol. Pharmaceutics*, vol. 16, nº 9, p. 4077-4085, sept. 2019, doi: 10.1021/acs.molpharmaceut.9b00525.
[89] D. K. Saini *et al.*, « Comprehensive evaluation of mapping complex traits in wheat using genome-wide association studies », *Molecular Breeding*, vol. 42, nº 1, p. 1-52, 2022.
[90] M. M. Salman *et al.*, « Design and Validation of a Human Brain Endothelial Microvessel-on-a-Chip Open Microfluidic Model Enabling Advanced Optical Imaging », *Frontiers in Bioengineering and Biotechnology*, vol. 8, 2020, Consulté le: 15 novembre 2022. [En ligne]. Disponible sur: https://www.frontiersin.org/articles/10.3389/fbioe.2020.573775

[91] R. E. Gur et R. C. Gur, « Functional magnetic resonance imaging in schizophrenia », *Dialogues in clinical neuroscience*, 2022.
[92] R. Yuen, F. G. West, et F. Wuest, « Dual Probes for Positron Emission Tomography (PET) and Fluorescence Imaging (FI) of Cancer », *Pharmaceutics*, vol. 14, nº 3, p. 645, 2022.
[93] S. Zhang, J. Shen, D. Li, et Y. Cheng, « Strategies in the delivery of Cas9 ribonucleoprotein for CRISPR/Cas9 genome editing », *Theranostics*, vol. 11, nº 2, p. 614, 2021.
[94] S. Chtita, « Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments », PhD Thesis, Université Moulay Ismaïl, Meknès, 2017.
[95] « Molecular Docking for Computer-Aided Drug Design ». Consulté le: 30 décembre 2022. [En ligne]. Disponible sur: https://shop.elsevier.com/books/molecular-docking-for-computer-aided-drug-design/coumar/978-0-12-822312-3#full-description
[96] S. Q. Pantaleão, P. O. Fernandes, J. E. Gonçalves, V. G. Maltarollo, et K. M. Honorio, « Recent advances in the prediction of pharmacokinetics properties in drug design studies: a review », *ChemMedChem*, vol. 17, nº 1, p. e202100542, 2022.
[97] T. I. Adelusi *et al.*, « Molecular modeling in drug discovery », *Informatics in Medicine Unlocked*, p. 100880, 2022.
[98] H. Sung *et al.*, « Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries », *CA: A Cancer Journal for Clinicians*, vol. 71, nº 3, p. 209-249, 2021, doi: 10.3322/caac.21660.
[99] J. Ferlay *et al.*, « Cancer statistics for the year 2020: An overview », *Int J Cancer*, avr. 2021, doi: 10.1002/ijc.33588.
[100] O. Daoui *et al.*, « Cyclohexane-1,3-dione Derivatives as Future Therapeutic Agents for NSCLC: QSAR Modeling, In Silico ADME-Tox Properties, and Structure-Based Drug Designing Approach », *ACS Omega*, janv. 2023, doi: 10.1021/acsomega.2c07585.
[101] T. B. Tafesse, M. H. Bule, F. Khan, M. Abdollahi, et M. Amini, « Developing Novel Anticancer Drugs for Targeted Populations: An Update », *Current Pharmaceutical Design*, vol. 27, nº 2, p. 250-262, janv. 2021, doi: 10.2174/1381612826666201124111748.
[102] K.-W. Kim, J. K. Roh, H.-J. Wee, et C. Kim, *Cancer Drug Discovery*. Dordrecht: Springer Netherlands, 2016. doi: 10.1007/978-94-024-0844-7.
[103] I. Collins et P. Workman, « New approaches to molecular cancer therapeutics », *Nature chemical biology*, vol. 2, nº 12, p. 689-700, 2006.
[104] L. Zhong *et al.*, « Small molecules in targeted cancer therapy: advances, challenges, and future perspectives », *Sig Transduct Target Ther*, vol. 6, nº 1, Art. nº 1, mai 2021, doi: 10.1038/s41392-021-00572-w.
[105] D. G. Savage et K. H. Antman, « Imatinib Mesylate — A New Oral Targeted Therapy », *New England Journal of Medicine*, vol. 346, nº 9, p. 683-693, févr. 2002, doi: 10.1056/NEJMra013339.
[106] E. Levantini, G. Maroni, M. Del Re, et D. G. Tenen, « EGFR signaling pathway as therapeutic target in human cancers », *Seminars in Cancer Biology*, vol. 85, p. 253-275, oct. 2022, doi: 10.1016/j.semcancer.2022.04.002.
[107] S. M. Centuori et J. E. Bauman, « c-Met Signaling as a Therapeutic Target in Head and Neck Cancer », *The Cancer Journal*, vol. 28, nº 5, p. 346, oct. 2022, doi: 10.1097/PPO.0000000000000619.
[108] M. L. Uribe, I. Marrocco, et Y. Yarden, « EGFR in Cancer: Signaling Mechanisms, Drugs, and Acquired Resistance », *Cancers (Basel)*, vol. 13, nº 11, p. 2748, juin 2021, doi: 10.3390/cancers13112748.
[109] S. Ali *et al.*, « Molecular mechanisms and mode of tamoxifen resistance in breast cancer », *Bioinformation*, vol. 12, nº 3, p. 135-139, juin 2016, doi: 10.6026/97320630012135.
[110] M. R. Seshadri et A. M. Melnick, « Targeting MALT1 for the treatment of diffuse large B-cell lymphoma », *Leukemia & Lymphoma*, vol. 63, nº 4, p. 789-798, mars 2022, doi: 10.1080/10428194.2021.1999444.
[111] D. Pan, C. Jiang, Z. Ma, M. Blonska, M. J. You, et X. Lin, « MALT1 is required for EGFR-induced NF-κB activation and contributes to EGFR-driven lung cancer progression », *Oncogene*, vol. 35, nº 7, Art. nº 7, févr. 2016, doi: 10.1038/onc.2015.146.
[112] K. Ray, B. Ujvari, V. Ramana, et J. Donald, « Cross-talk between EGFR and IL-6 drives oncogenic signaling and offers therapeutic opportunities in cancer », *Cytokine & Growth Factor Reviews*, vol. 41, p. 18-27, juin 2018, doi: 10.1016/j.cytogfr.2018.04.002.

[113] Y. Zhang *et al.*, « Function of the c-Met receptor tyrosine kinase in carcinogenesis and associated therapeutic opportunities », *Molecular Cancer*, vol. 17, nº 1, p. 45, févr. 2018, doi: 10.1186/s12943-018-0796-y.
[114] W.-T. Liu *et al.*, « Hepatic stellate cell promoted hepatoma cell invasion via the HGF/c-Met signaling pathway regulated by p53 », *Cell Cycle*, vol. 15, nº 7, p. 886-894, 2016, doi: 10.1080/15384101.2016.1152428.
[115] Y. Imura *et al.*, « Functional and therapeutic relevance of hepatocyte growth factor/c-MET signaling in synovial sarcoma », *Cancer Science*, vol. 107, nº 12, p. 1867-1876, 2016, doi: 10.1111/cas.13092.
[116] J.-C. Wu *et al.*, « Heteronemin Is a Novel c-Met/STAT3 Inhibitor Against Advanced Prostate Cancer Cells », *The Prostate*, vol. 76, nº 16, p. 1469-1483, 2016, doi: 10.1002/pros.23230.
[117] H. Wang *et al.*, « The Function of the HGF/c-Met Axis in Hepatocellular Carcinoma », *Frontiers in Cell and Developmental Biology*, vol. 8, 2020, Consulté le: 26 janvier 2023. [En ligne]. Disponible sur: https://www.frontiersin.org/articles/10.3389/fcell.2020.00055
[118] H.-N. Mo et P. Liu, « Targeting MET in cancer therapy », *Chronic Diseases and Translational Medicine*, vol. 3, nº 3, p. 148-153, sept. 2017, doi: 10.1016/j.cdtm.2017.06.002.
[119] S. L. Organ et M.-S. Tsao, « An overview of the c-MET signaling pathway », *Ther Adv Med Oncol*, vol. 3, nº 1 Suppl, p. S7-S19, nov. 2011, doi: 10.1177/1758834011422556.
[120] D. Delitto, E. Vertes-George, S. J. Hughes, K. E. Behrns, et J. G. Trevino, « c-Met signaling in the development of tumorigenesis and chemoresistance: Potential applications in pancreatic cancer », *World J Gastroenterol*, vol. 20, nº 26, p. 8458-8470, juill. 2014, doi: 10.3748/wjg.v20.i26.8458.
[121] J. P. Eder, G. F. Vande Woude, S. A. Boerner, et P. M. LoRusso, « Novel therapeutic inhibitors of the c-Met signaling pathway in cancer », *Clin Cancer Res*, vol. 15, nº 7, p. 2207-2214, avr. 2009, doi: 10.1158/1078-0432.CCR-08-1306.
[122] Q.-W. Zhang, Z.-D. Ye, et L. Shi, « c-Met kinase inhibitors: an update patent review (2014-2017) », *Expert Opinion on Therapeutic Patents*, vol. 29, nº 1, p. 25-41, janv. 2019, doi: 10.1080/13543776.2019.1552261.
[123] L. Goyal, M. D. Muzumdar, et A. X. Zhu, « Targeting the HGF/c-MET Pathway in Hepatocellular Carcinoma », *Clinical Cancer Research*, vol. 19, nº 9, p. 2310-2318, avr. 2013, doi: 10.1158/1078-0432.CCR-12-2791.
[124] S. J. Klempner, A. Borghei, B. Hakimian, S. M. Ali, et S.-H. I. Ou, « Intracranial Activity of Cabozantinib in MET Exon 14–Positive NSCLC with Brain Metastases », *Journal of Thoracic Oncology*, vol. 12, nº 1, p. 152-156, janv. 2017, doi: 10.1016/j.jtho.2016.09.127.
[125] M. J. Wieduwilt et M. M. Moasser, « The epidermal growth factor receptor family: Biology driving targeted therapeutics », *Cell Mol Life Sci*, vol. 65, nº 10, p. 1566-1584, mai 2008, doi: 10.1007/s00018-008-7440-8.
[126] B. Linggi et G. Carpenter, « ErbB receptors: new insights on mechanisms and biology », *Trends in Cell Biology*, vol. 16, nº 12, p. 649-656, déc. 2006, doi: 10.1016/j.tcb.2006.10.008.
[127] J. Downward, P. Parker, et M. D. Waterfield, « Autophosphorylation sites on the epidermal growth factor receptor », *Nature*, vol. 311, nº 5985, Art. nº 5985, oct. 1984, doi: 10.1038/311483a0.
[128] M. Scaltriti et J. Baselga, « The Epidermal Growth Factor Receptor Pathway: A Model for Targeted Therapy », *Clinical Cancer Research*, vol. 12, nº 18, p. 5268-5272, sept. 2006, doi: 10.1158/1078-0432.CCR-05-1554.
[129] S. S. Ramalingam, T. K. Owonikoko, et F. R. Khuri, « Lung cancer: New biological insights and recent therapeutic advances », *CA: A Cancer Journal for Clinicians*, vol. 61, nº 2, p. 91-112, 2011, doi: 10.3322/caac.20102.
[130] U. McDermott, R. V. Pusapati, J. G. Christensen, N. S. Gray, et J. Settleman, « Acquired Resistance of Non–Small Cell Lung Cancer Cells to MET Kinase Inhibition Is Mediated by a Switch to Epidermal Growth Factor Receptor Dependency », *Cancer Research*, vol. 70, nº 4, p. 1625-1634, févr. 2010, doi: 10.1158/0008-5472.CAN-09-3620.
[131] W. Wang *et al.*, « Met kinase inhibitor E7050 reverses three different mechanisms of hepatocyte growth factor-induced tyrosine kinase inhibitor resistance in EGFR mutant lung cancer », *Clin Cancer Res*, vol. 18, nº 6, p. 1663-1671, mars 2012, doi: 10.1158/1078-0432.CCR-11-1171.
[132] E. Benedettini *et al.*, « Met activation in non-small cell lung cancer is associated with de novo resistance to EGFR inhibitors and the development of brain metastasis », *Am J Pathol*, vol. 177, nº 1, p. 415-423, juill. 2010, doi: 10.2353/ajpath.2010.090863.

[133] J.-C. Ko *et al.*, « Tamoxifen enhances erlotinib-induced cytotoxicity through down-regulating AKT-mediated thymidine phosphorylase expression in human non-small-cell lung cancer cells », *Biochem Pharmacol*, vol. 88, nº 1, p. 119-127, mars 2014, doi: 10.1016/j.bcp.2014.01.010.
[134] M. Yi, T. He, K. Wang, et Y. Wei, « Comparison of gefitinib plus chemotherapy versus gefitinib alone for advanced non-small-cell lung cancer: A meta analysis », *Clinics*, vol. 78, p. 100152, janv. 2023, doi: 10.1016/j.clinsp.2022.100152.
[135] M. Arnold *et al.*, « Current and future burden of breast cancer: Global statistics for 2020 and 2040 », *Breast*, vol. 66, p. 15-23, sept. 2022, doi: 10.1016/j.breast.2022.08.010.
[136] P. Seshacharyulu, M. P. Ponnusamy, D. Haridas, M. Jain, A. K. Ganti, et S. K. Batra, « Targeting the EGFR signaling pathway in cancer therapy », *Expert Opinion on Therapeutic Targets*, vol. 16, nº 1, p. 15-31, janv. 2012, doi: 10.1517/14728222.2011.648617.
[137] J. G. Paez *et al.*, « EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy », *Science*, vol. 304, nº 5676, p. 1497-1500, juin 2004, doi: 10.1126/science.1099314.
[138] T. Mitsudomi *et al.*, « Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial », *Lancet Oncol*, vol. 11, nº 2, p. 121-128, févr. 2010, doi: 10.1016/S1470-2045(09)70364-X.
[139] L. V. Sequist *et al.*, « Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations », *J Clin Oncol*, vol. 31, nº 27, p. 3327-3334, sept. 2013, doi: 10.1200/JCO.2012.44.2806.
[140] J. L. González-Larriba, M. Lázaro-Quintela, M. Cobo, M. Dómine, M. Majem, et R. García-Campelo, « Clinical management of epidermal growth factor receptor mutation-positive non-small cell lung cancer patients after progression on previous epidermal growth factor receptor tyrosine kinase inhibitors: the necessity of repeated molecular analysis », *Transl Lung Cancer Res*, vol. 6, nº Suppl 1, p. S21-S34, déc. 2017, doi: 10.21037/tlcr.2017.10.03.
[141] L. H. Sehn et G. Salles, « Diffuse Large B-Cell Lymphoma », *New England Journal of Medicine*, vol. 384, nº 9, p. 842-858, mars 2021, doi: 10.1056/NEJMra2027612.
[142] A. Rosenwald *et al.*, « Prognostic Significance of MYC Rearrangement and Translocation Partner in Diffuse Large B-Cell Lymphoma : A Study by the Lunenburg Lymphoma Biomarker Consortium », *Journal of Clinical Oncology*, p. JCO1900743, sept. 2019.
[143] S. Hailfinger *et al.*, « Malt1-dependent RelB cleavage promotes canonical NF-κB activation in lymphocytes and lymphoma cell lines », *Proc Natl Acad Sci U S A*, vol. 108, nº 35, p. 14596-14601, août 2011, doi: 10.1073/pnas.1105020108.
[144] Y.-Y. Zhang, J. Peng, et X.-J. Luo, « Post-translational modification of MALT1 and its role in B cell- and T cell-related diseases », *Biochemical Pharmacology*, vol. 198, p. 114977, avr. 2022, doi: 10.1016/j.bcp.2022.114977.
[145] A. Thys, T. Douanne, et N. Bidère, « Post-translational modifications of the CARMA1-BCL10-MALT1 complex in lymphocytes and activated B-cell like subtype of diffuse large B-cell lymphoma », *Frontiers in oncology*, vol. 8, p. 498, 2018.
[146] X. Liang, Y. Cao, C. Li, H. Yu, C. Yang, et H. Liu, « MALT1 as a promising target to treat lymphoma and other diseases related to MALT1 anomalies », *Medicinal Research Reviews*, 2021.
[147] B. Gomez Solsona, A. Schmitt, K. Schulze-Osthoff, et S. Hailfinger, « The Paracaspase MALT1 in Cancer. Biomedicines 2022, 10, 344 ». s Note: MDPI stays neutral with regard to jurisdictional claims in published …, 2022.
[148] I. Skordos *et al.*, « Normal lymphocyte homeostasis and function in MALT1 protease-resistant HOIL-1 knock-in mice », *The FEBS Journal*, 2022.
[149] F. Tokunaga et F. Ikeda, « Linear ubiquitination in immune and neurodegenerative diseases, and beyond », *Biochemical Society Transactions*, vol. 50, nº 2, p. 799-811, 2022.
[150] Q. Zhang, S. Qiao, C. Yang, et G. Jiang, « Nuclear factor-kappa B and effector molecules in photoaging », *Cutaneous and Ocular Toxicology*, vol. 41, nº 2, p. 187-193, 2022.
[151] N. Challagundla, B. Saha, et R. Agrawal-Rajput, « Insights into inflammasome regulation: cellular, molecular, and pathogenic control of inflammasome activation », *Immunologic Research*, vol. 70, nº 5, p. 578-606, 2022.
[152] X. Dong, X. Chen, et Y. Ren, « MALT1 reflects inflammatory cytokines, disease activity, and its chronological change could estimate treatment response to infliximab in Crohn's disease patients », *Journal of Clinical Laboratory Analysis*, vol. 36, nº 10, p. e24650, 2022.

[153] J. Quancard *et al.*, « An allosteric MALT1 inhibitor is a molecular corrector rescuing function in an immunodeficient patient », *Nat Chem Biol*, vol. 15, nº 3, Art. nº 3, mars 2019, doi: 10.1038/s41589-018-0222-1.
[154] B.-T. Xin *et al.*, « Development of new Malt1 inhibitors and probes », *Bioorganic & Medicinal Chemistry*, vol. 24, nº 15, p. 3312-3329, août 2016, doi: 10.1016/j.bmc.2016.03.035.
[155] D. Vercammen *et al.*, « Serpin1 of Arabidopsis thaliana is a Suicide Inhibitor for Metacaspase 9 », *Journal of Molecular Biology*, vol. 364, nº 4, p. 625-636, déc. 2006, doi: 10.1016/j.jmb.2006.09.010.
[156] G. Wu *et al.*, « Synthesis and structure–activity relationship studies of MI-2 analogues as MALT1 inhibitors », *Bioorganic & Medicinal Chemistry*, vol. 26, nº 12, p. 3321-3344, juill. 2018, doi: 10.1016/j.bmc.2018.04.059.
[157] J. K. Lue, O. A. O'Connor, et F. Bertoni, « Targeting pathogenic mechanisms in marginal zone lymphoma: from concepts and beyond », *Annals of lymphoma*, vol. 4, 2020.
[158] W. Magits et A. A. Sablina, « The regulation of the protein interaction network by monoubiquitination », *Current Opinion in Structural Biology*, vol. 73, p. 102333, 2022.
[159] J. Cheng, L. M. Maurer, H. Kang, P. C. Lucas, et L. M. McAllister-Lucas, « Critical protein-protein interactions within the CARMA1-BCL10-MALT1 complex: Take-home points for the cell biologist », *Cellular Immunology*, vol. 355, p. 104158, 2020.
[160] I. Gojo et J. E. Karp, « New strategies in acute myelogenous leukemia: leukemogenesis and personalized medicine », *Clinical Cancer Research*, vol. 20, nº 24, p. 6233-6241, 2014.
[161] I. Hamp, T. J. O'Neill, O. Plettenburg, et D. Krappmann, « A patent review of MALT1 inhibitors (2013-present) », *Expert Opinion on Therapeutic Patents*, vol. 31, nº 12, p. 1079-1096, 2021.
[162] « Coronavirus disease (COVID-19) – World Health Organization ». Consulté le: 7 février 2023. [En ligne]. Disponible sur: https://www.who.int/emergencies/diseases/novel-coronavirus-2019
[163] Z. Hamidi, S. Jabraeili-Siahroud, Y. Taati-Alamdari, P. S. Aghbash, A. Shamekh, et H. B. Baghi, « A comprehensive review of COVID-19 symptoms and treatments in the setting of autoimmune diseases », *Virology Journal*, vol. 20, nº 1, p. 1, janv. 2023, doi: 10.1186/s12985-023-01967-7.
[164] « Weekly epidemiological update on COVID-19 - 21 December 2022 ». Consulté le: 7 février 2023. [En ligne]. Disponible sur: https://www.who.int/publications/m/item/covid-19-weekly-epidemiological-update---21-december-2022
[165] W. Cui, K. Yang, et H. Yang, « Recent Progress in the Drug Development Targeting SARS-CoV-2 Main Protease as Treatment for COVID-19 », *Frontiers in Molecular Biosciences*, vol. 7, 2020, Consulté le: 7 février 2023. [En ligne]. Disponible sur: https://www.frontiersin.org/articles/10.3389/fmolb.2020.616341
[166] Z. Molavi *et al.*, « Identification of FDA approved drugs against SARS-CoV-2 RNA dependent RNA polymerase (RdRp) and 3-chymotrypsin-like protease (3CLpro), drug repurposing approach », *Biomedicine & Pharmacotherapy*, vol. 138, p. 111544, juin 2021, doi: 10.1016/j.biopha.2021.111544.
[167] S. Pushpakom *et al.*, « Drug repurposing: progress, challenges and recommendations », *Nat Rev Drug Discov*, vol. 18, nº 1, Art. nº 1, janv. 2019, doi: 10.1038/nrd.2018.168.
[168] S. Nhean *et al.*, « COVID-19: A Review of Potential Treatments (Corticosteroids, Remdesivir, Tocilizumab, Bamlanivimab/Etesevimab, and Casirivimab/Imdevimab) and Pharmacological Considerations », *Journal of Pharmacy Practice*, p. 08971900211048139, oct. 2021, doi: 10.1177/08971900211048139.
[169] D. M. Altmann, « The COVID-19 immunology masterclass enters its fourth year », *Nat Immunol*, vol. 24, nº 2, Art. nº 2, févr. 2023, doi: 10.1038/s41590-022-01393-x.
[170] M. T. Mascellino, F. Di Timoteo, M. De Angelis, et A. Oliva, « Overview of the Main Anti-SARS-CoV-2 Vaccines: Mechanism of Action, Efficacy and Safety », *Infection and Drug Resistance*, vol. 14, p. 3459-3476, mars 2021, doi: 10.2147/IDR.S315727.
[171] S. Nasreen *et al.*, « Effectiveness of COVID-19 vaccines against variants of concern, Canada », *Medrxiv*, 2021.
[172] X. He, C. He, W. Hong, K. Zhang, et X. Wei, « The challenges of COVID-19 Delta variant: prevention and vaccine development », *MedComm*, vol. 2, nº 4, p. 846-854, 2021.
[173] J. Singh, S. A. Rahman, N. Z. Ehtesham, S. Hira, et S. E. Hasnain, « SARS-CoV-2 variants of concern are emerging in India », *Nat Med*, vol. 27, nº 7, Art. nº 7, juill. 2021, doi: 10.1038/s41591-021-01397-4.
[174] J. Zhu *et al.*, « Progress on SARS-CoV-2 3CLpro Inhibitors: Inspiration from SARS-CoV 3CLpro Peptidomimetics and Small-Molecule Anti-Inflammatory Compounds », *Drug Design, Development and Therapy*, vol. 16, p. 1067-1082, déc. 2022, doi: 10.2147/DDDT.S359009.

[175] « Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern ». Consulté le: 8 février 2023. [En ligne]. Disponible sur: https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern
[176] H. M. Froggatt, B. E. Heaton, et N. S. Heaton, « Development of a fluorescence-based, high-throughput SARS-CoV-2 3CLpro reporter assay », *Journal of virology*, vol. 94, nº 22, p. e01265-20, 2020.
[177] S. Jo, S. Kim, D. Y. Kim, M.-S. Kim, et D. H. Shin, « Flavonoids with inhibitory activity against SARS-CoV-2 3CLpro », *Journal of Enzyme Inhibition and Medicinal Chemistry*, vol. 35, nº 1, p. 1539-1544, 2020.
[178] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, et Q. Zhou, « Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2 », *Science*, vol. 367, nº 6485, p. 1444-1448, mars 2020, doi: 10.1126/science.abb2762.
[179] Y. Huang, C. Yang, X. Xu, W. Xu, et S. Liu, « Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19 », *Acta Pharmacol Sin*, vol. 41, nº 9, Art. nº 9, sept. 2020, doi: 10.1038/s41401-020-0485-4.
[180] D. Shin *et al.*, « Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity », *Nature*, vol. 587, nº 7835, Art. nº 7835, nov. 2020, doi: 10.1038/s41586-020-2601-5.
[181] C. Empereur-Mot, « Développement d'outils statistiques d'évaluation de méthodes de criblage virtuel: courbes de prédictivité & Screening Explorer », PhD Thesis, Conservatoire national des arts et metiers-CNAM, 2017.
[182] J. Janes *et al.*, « The ReFRAME library as a comprehensive drug repurposing library and its application to the treatment of cryptosporidiosis », *Proceedings of the National Academy of Sciences*, vol. 115, nº 42, p. 10750-10755, 2018.
[183] A. Varela-Rial, M. Majewski, et G. De Fabritiis, « Structure based virtual screening: Fast and slow », *WIREs Computational Molecular Science*, vol. 12, nº 2, p. e1544, 2022, doi: 10.1002/wcms.1544.
[184] O. O. Grygorenko, D. S. Radchenko, I. Dziuba, A. Chuprina, K. E. Gubina, et Y. S. Moroz, « Generating Multibillion Chemical Space of Readily Accessible Screening Compounds », *iScience*, vol. 23, nº 11, p. 101681, nov. 2020, doi: 10.1016/j.isci.2020.101681.
[185] A. Gaulton *et al.*, « ChEMBL: a large-scale bioactivity database for drug discovery », *Nucleic Acids Research*, vol. 40, nº D1, p. D1100-D1107, janv. 2012, doi: 10.1093/nar/gkr777.
[186] A. D. Baxevanis, G. D. Bader, et D. S. Wishart, *Bioinformatics*. John Wiley & Sons, 2020.
[187] K. Malathi et S. Ramaiah, « Bioinformatics approaches for new drug discovery: a review », *Biotechnology and Genetic Engineering Reviews*, vol. 34, nº 2, p. 243-260, 2018.
[188] M. Pagano, K. Gauvreau, et H. Mattie, *Principles of biostatistics*. CRC Press, 2022.
[189] A. R. Feinstein, « Clinical biostatistics », *Clinical Pharmacology & Therapeutics*, vol. 22, nº 4, p. 485-498, 1977.
[190] W. W. Daniel et C. L. Cross, *Biostatistics: a foundation for analysis in the health sciences*. Wiley, 2018.
[191] V. S. Raj, A. Priyadarshini, M. K. Yadav, R. P. Pandey, A. Gupta, et A. Vibhuti, « Artificial Intelligence in Bioinformatics », in *Biomedical Data Mining for Information Retrieval*, John Wiley & Sons, Ltd, 2021, p. 21-51. doi: 10.1002/9781119711278.ch2.
[192] I. Inza, B. Calvo, R. Armananzas, E. Bengoetxea, P. Larranaga, et J. A. Lozano, « Machine learning: an indispensable tool in bioinformatics », in *Bioinformatics methods in clinical research*, Springer, 2009, p. 25-48.
[193] L. Herzog, E. Murina, O. Dürr, S. Wegener, et B. Sick, « Integrating uncertainty in deep neural networks for MRI based stroke analysis », *Medical image analysis*, vol. 65, p. 101790, 2020.
[194] N. Iqbal et P. Kumar, « From Data Science to Bioscience: Emerging era of bioinformatics applications, tools and challenges », *Procedia Computer Science*, vol. 218, p. 1516-1528, 2023.
[195] M. Torrisi, G. Pollastri, et Q. Le, « Deep learning methods in protein structure prediction », *Computational and Structural Biotechnology Journal*, vol. 18, p. 1301-1310, 2020.
[196] T. Nogrady et D. F. Weaver, *Medicinal chemistry: a molecular and biochemical approach*. Oxford University Press, 2005.
[197] M. A. Lill, « Multi-dimensional QSAR in drug discovery », *Drug Discovery Today*, vol. 12, nº 23, p. 1013-1017, déc. 2007, doi: 10.1016/j.drudis.2007.08.004.
[198] Y. C. Martin, « Exploring QSAR: Hydrophobic, Electronic, and Steric Constants C. Hansch, A. Leo, and D. Hoekman. American Chemical Society, Washington, DC. 1995. Xix + 348 pp. 22 × 28.5 cm. Exploring QSAR: Fundamentals and Applications in Chemistry and Biology. C. Hansch and A. Leo. American Chemical Society, Washington, DC. 1995. Xvii + 557 pp. 18.5 × 26 cm.

ISBN 0-8412-2993-7 (set). $99.95 (set). », *J. Med. Chem.*, vol. 39, nº 5, p. 1189-1190, janv. 1996, doi: 10.1021/jm950902o.
[199] H. Kubinyi, *QSAR: Hansch analysis and related approaches*, vol. 1. VcH Weinheim, 1993.
[200] C. Hansch, J. E. Quinlan, et G. L. Lawrence, « Linear free-energy relationship between partition coefficients and the aqueous solubility of organic liquids », *The Journal of Organic Chemistry*, vol. 33, nº 1, p. 347-350, 1968.
[201] C. Hansch, « Quantitative approach to biochemical structure-activity relationships », ACS Publications. Consulté le: 22 mars 2023. [En ligne]. Disponible sur: https://pubs.acs.org/doi/pdf/10.1021/ar50020a002
[202] I. E. Ebhohimen, L. Edemhanria, S. Awojide, O. H. Onyijen, et G. Anywar, « Chapter 3 - Advances in computer-aided drug discovery », in *Phytochemicals as Lead Compounds for New Drug Discovery*, C. Egbuna, S. Kumar, J. C. Ifemeje, S. M. Ezzat, et S. Kaliyaperumal, Éd., Elsevier, 2020, p. 25-37. doi: 10.1016/B978-0-12-817890-4.00003-2.
[203] N. Demise, Y. Miwa, M. Nakabayashi, Y. Nakoshi, et N. Demise, « OECD principles of corporate governance », *Corporate governance in Japan: From the viewpoints of management, accounting, and the market*, p. 109-117, 2006.
[204] G. Gómez-Jiménez *et al.*, « The OECD Principles for (Q)SAR Models in the Context of Knowledge Discovery in Databases (KDD) », *Adv Protein Chem Struct Biol*, vol. 113, p. 85-117, 2018, doi: 10.1016/bs.apcsb.2018.04.001.
[205] P. De, S. Kar, P. Ambure, et K. Roy, « Prediction reliability of QSAR models: an overview of various validation tools », *Archives of Toxicology*, vol. 96, nº 5, p. 1279-1295, 2022.
[206] I. Hoque, A. Chatterjee, S. Bhattacharya, et R. Biswas, « An approach of computer-aided drug design (CADD) tools for in silico pharmaceutical drug design and development », *Int. J. Adv. Res. Biol. Sci*, vol. 4, nº 2, p. 60-71, 2017.
[207] A. A. Toropov et A. P. Toropova, « QSAR as a random event: criteria of predictive potential for a chance model », *Structural Chemistry*, vol. 30, p. 1677-1683, 2019.
[208] M. A. Lill, « Multi-dimensional QSAR in drug discovery », *Drug Discovery Today*, vol. 12, nº 23, p. 1013-1017, déc. 2007, doi: 10.1016/j.drudis.2007.08.004.
[209] V. Kanakaveti, P. Anoosha, R. Sakthivel, S. k. Rayala, et M. M. Gromiha, « Quantitative structure?activity relationship in ligand-based drug design: Concepts and applications », in *Protein Interactions*, WORLD SCIENTIFIC, 2019, p. 333-349. doi: 10.1142/9789811211874_0013.
[210] T. Wang, X. Yuan, M.-B. Wu, J.-P. Lin, et L.-R. Yang, « The advancement of multidimensional QSAR for novel drug discovery-where are we headed? », *Expert opinion on drug discovery*, vol. 12, nº 8, p. 769-784, 2017.
[211] R. Muthukumaran, B. Sangeetha, R. Amutha, et P. P. Mathur, « Development of Anti-HIV Activity Models of Lysine Sulfonamide Analogs: A QSAR Perspective », *Current Computer - Aided Drug Design*, vol. 8, nº 1, p. 70-82, mars 2012, doi: 10.2174/157340912799218543.
[212] S. M. Free et J. W. Wilson, « A mathematical contribution to structure-activity studies », *Journal of medicinal chemistry*, vol. 7, nº 4, p. 395-399, 1964.
[213] R. D. Cramer, D. E. Patterson, et J. D. Bunce, « Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins », *J. Am. Chem. Soc.*, vol. 110, nº 18, p. 5959-5967, août 1988, doi: 10.1021/ja00226a005.
[214] V. Lukacova et S. Balaz, « Multimode Ligand Binding in Receptor Site Modeling: Implementation in CoMFA », *J. Chem. Inf. Comput. Sci.*, vol. 43, nº 6, p. 2093-2105, nov. 2003, doi: 10.1021/ci034100a.
[215] A. Vedani et M. Dobler, « 5D-QSAR: The Key for Simulating Induced Fit? », *J. Med. Chem.*, vol. 45, nº 11, p. 2139-2149, mai 2002, doi: 10.1021/jm011005p.
[216] A. Vedani, M. Dobler, et M. A. Lill, « Combining Protein Modeling and 6D-QSAR. Simulating the Binding of Structurally Diverse Ligands to the Estrogen Receptor », *J. Med. Chem.*, vol. 48, nº 11, p. 3700-3703, juin 2005, doi: 10.1021/jm050185q.
[217] O. Daoui, S. Elkhattabi, et S. Chtita, « Rational design of novel pyridine-based drugs candidates for lymphoma therapy », *Journal of Molecular Structure*, vol. 1270, p. 133964, déc. 2022, doi: 10.1016/j.molstruc.2022.133964.
[218] Z. Wu *et al.*, « MoleculeNet: a benchmark for molecular machine learning », *Chemical science*, vol. 9, nº 2, p. 513-530, 2018.
[219] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, et A. Walsh, « Machine learning for molecular and materials science », *Nature*, vol. 559, nº 7715, Art. nº 7715, juill. 2018, doi: 10.1038/s41586-018-0337-2.

[220] F. Grisoni, V. Consonni, et R. Todeschini, « Impact of Molecular Descriptors on Computational Models », in *Computational Chemogenomics*, J. B. Brown, Éd., in Methods in Molecular Biology. , New York, NY: Springer, 2018, p. 171-209. doi: 10.1007/978-1-4939-8639-2_5.
[221] P. Gramatica, « Principles of QSAR modeling: comments and suggestions from personal experience », *International Journal of Quantitative Structure-Property Relationships (IJQSPR)*, vol. 5, nº 3, p. 61-97, 2020.
[222] A. Mauri, V. Consonni, et R. Todeschini, « Molecular descriptors », in *Handbook of computational chemistry*, Springer, 2017, p. 2065-2093.
[223] R. Todeschini et V. Consonni, *Handbook of molecular descriptors*. John Wiley & Sons, 2008.
[224] A. Bird et E. Tobin, « Natural kinds », 2008.
[225] A. Cereto-Massagué, M. J. Ojeda, C. Valls, M. Mulero, S. Garcia-Vallvé, et G. Pujadas, « Molecular fingerprint similarity search in virtual screening », *Methods*, vol. 71, p. 58-63, 2015.
[226] G. Montavon *et al.*, « Learning invariant representations of molecules for atomization energy prediction », *Advances in neural information processing systems*, vol. 25, 2012.
[227] R. Todeschini et V. Consonni, « Molecular descriptors », *Recent Advances in QSAR Studies*, p. 29-102, 2010.
[228] R. Todeschini et V. Consonni, « Descriptors from Molecular Geometry », in *Handbook of Chemoinformatics*, John Wiley & Sons, Ltd, 2003, p. 1004-1033. doi: 10.1002/9783527618279.ch37.
[229] O. Devinyak, D. Havrylyuk, et R. Lesyk, « 3D-MoRSE descriptors explained », *Journal of Molecular Graphics and Modelling*, vol. 54, p. 194-203, nov. 2014, doi: 10.1016/j.jmgm.2014.10.006.
[230] J. U. Reveles et A. M. Köster, « Geometry optimization in density functional methods », *Journal of computational chemistry*, vol. 25, nº 9, p. 1109-1116, 2004.
[231] Y. J. Tseng, A. J. Hopfinger, et E. X. Esposito, « The great descriptor melting pot: mixing descriptors for the common good of QSAR models », *J Comput Aided Mol Des*, vol. 26, nº 1, p. 39-43, janv. 2012, doi: 10.1007/s10822-011-9511-4.
[232] E. S. Boek, D. S. Yakovlev, et T. F. Headen, « Quantitative molecular representation of asphaltenes and molecular dynamics simulation of their aggregation », *Energy & Fuels*, vol. 23, nº 3, p. 1209-1219, 2009.
[233] A. U. Khan, « Descriptors and their selection methods in QSAR analysis: paradigm for drug design », *Drug discovery today*, vol. 21, nº 8, p. 1291-1302, 2016.
[234] I. D. Kuntz, E. C. Meng, et B. K. Shoichet, « Structure-based molecular design », *Accounts of Chemical research*, vol. 27, nº 5, p. 117-123, 1994.
[235] K. Roy, S. Kar, et R. N. Das, « Chapter 8 - Introduction to 3D-QSAR », in *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*, K. Roy, S. Kar, et R. N. Das, Éd., Boston: Academic Press, 2015, p. 291-317. doi: 10.1016/B978-0-12-801505-6.00008-9.
[236] T. Gund, « 3 - Molecular Modeling of Small Molecules », in *Guidebook on Molecular Modeling in Drug Design*, N. C. Cohen, Éd., San Diego: Academic Press, 1996, p. 55-92. doi: 10.1016/B978-012178245-0/50004-4.
[237] X. Guo, D. He, L. Liu, R. Kuang, et L. Liu, « Use of QM/MM scheme to reproduce macromolecule–small molecule noncovalent binding energy », *Computational and Theoretical Chemistry*, vol. 991, p. 134-140, 2012.
[238] « ACDLABS 10. Advanced Chemistry Development. Inc. Toronto. ON. Canada 2015. Available:(www.acdlabs.com/ressources/freeware/chemsketch/). » Consulté le: 3 février 2022. [En ligne]. Disponible sur: https://chemdoodle.com/?gclid=CjwKCAiAl-6PBhBCEiwAc2GOVLnzGJRwd7uvt6NgRVXj-_Ho-4pmyJqE_B3ZxpYzKET2uK5zI_n_2BoCYVMQAvD_BwE
[239] N. L. Allinger, « Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms », *Journal of the American Chemical Society*, vol. 99, nº 25, p. 8127-8134, 1977.
[240] A. D. Becke, « Density-functional exchange-energy approximation with correct asymptotic behavior », *Phys. Rev. A*, vol. 38, nº 6, p. 3098-3100, sept. 1988, doi: 10.1103/PhysRevA.38.3098.
[241] M. Clark, R. D. Cramer, et N. Van Opdenbosch, « Validation of the general purpose tripos 5.2 force field », *J. Comput. Chem.*, vol. 10, nº 8, p. 982-1012, déc. 1989, doi: 10.1002/jcc.540100804.

[242] N. S. KL et C. Dasgupta, « Investigating the nature of students' reasoning in connecting molecular structures of stereoisomers with their physical properties using an ar app », in *2018 IEEE Tenth international conference on technology for education (t4e)*, IEEE, 2018, p. 53-60.
[243] J. R. De Laeter *et al.*, « Atomic weights of the elements. Review 2000 (IUPAC Technical Report) », *Pure and applied chemistry*, vol. 75, nº 6, p. 683-800, 2003.
[244] R. Siegfried, *From elements to atoms: A history of chemical composition*, vol. 92. American Philosophical Society, 2002.
[245] R. J. W. Le Fevre, « Molecular refractivity and polarizability », in *Advances in Physical Organic Chemistry*, vol. 3, Elsevier, 1965, p. 1-90.
[246] W. Y. Learn, « Gases », 2009.
[247] A. S. Danesh, A. Y. Dandekar, A. C. Todd, et R. Sarkar, « A modified scaling law and parachor method approach for improved prediction of interfacial tension of gas-condensate systems », in *SPE annual technical conference and exhibition*, OnePetro, 1991.
[248] L. Diéguez, N. Darwish, M. Mir, E. Martínez, M. Moreno, et J. Samitier, « Effect of the Refractive Index of Buffer Solutions in Evanescent Optical Biosensors », *Sensor Letters*, vol. 7, nº 5, p. 851-855, oct. 2009, doi: 10.1166/sl.2009.1161.
[249] J. W. Bush, D. L. Hu, et M. Prakash, « The integument of water-walking arthropods: form and function », *Advances in insect physiology*, vol. 34, p. 117-192, 2007.
[250] L. Senzamici et H. McMaster, « The Heaviness of Objects and Heaviness of a Material Kind: Some 11 to 12-Year-Old Children's Understandings of Mass and Density. », *Mathematics Education Research Group of Australasia*, 2019.
[251] G. S. Baker, T. E. Jordan, et J. Pardy, « An introduction to ground penetrating radar (GPR) », 2007.
[252] T. M. Miller et B. Bederson, « Atomic and molecular polarizabilities-a review of recent advances », *Advances in atomic and molecular physics*, vol. 13, p. 1-55, 1978.
[253] P. Li, Z.-M. Zong, Z.-K. Li, Y.-G. Wang, F.-J. Liu, et X.-Y. Wei, « Characterization of basic heteroatom-containing organic compounds in liquefaction residue from Shenmu–Fugu subbituminous coal by positive-ion electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry », *Fuel Processing Technology*, vol. 132, p. 91-98, 2015.
[254] A. L. Rockwood, J. R. Van Orman, et D. V. Dearden, « Isotopic compositions and accurate masses of single isotopic peaks », *Journal of the American Society for Mass Spectrometry*, vol. 15, p. 12-21, 2004.
[255] J. Yergey, D. Heller, G. Hansen, R. J. Cotter, et C. Fenselau, « Isotopic distributions in mass spectra of large molecules », *Analytical Chemistry*, vol. 55, nº 2, p. 353-356, 1983.
[256] M. W. Senko, S. C. Beu, et F. W. McLaffertycor, « Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions », *Journal of the American Society for Mass Spectrometry*, vol. 6, nº 4, p. 229-233, 1995.
[257] E. Gaubas *et al.*, « Pulsed photo-ionization spectroscopy of traps in as-grown and neutron irradiated ammonothermally synthesized GaN », *Scientific reports*, vol. 9, nº 1, p. 1-7, 2019.
[258] M. A. González, « Force fields and molecular dynamics simulations », *École thématique de la Société Française de la Neutronique*, vol. 12, p. 169-200, 2011.
[259] F.-Y. Lin et A. D. MacKerell, « Force fields for small molecules », *Methods Mol Biol*, vol. 2022, p. 21-54, 2019, doi: 10.1007/978-1-4939-9608-7_2.
[260] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. I. Goddard, et W. M. Skiff, « UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations », *J. Am. Chem. Soc.*, vol. 114, nº 25, p. 10024-10035, déc. 1992, doi: 10.1021/ja00051a040.
[261] J. Gräfenstein, E. Kraka, et D. Cremer, « Density functional theory for open-shell singlet biradicals », *Chemical physics letters*, vol. 288, nº 5-6, p. 593-602, 1998.
[262] P. Sinha, S. E. Boesch, C. Gu, R. A. Wheeler, et A. K. Wilson, « Harmonic vibrational frequencies: scaling factors for HF, B3LYP, and MP2 methods in combination with correlation consistent basis sets », *The Journal of Physical Chemistry A*, vol. 108, nº 42, p. 9213-9217, 2004.
[263] L. Zhang, J. Wan, et G. Yang, « A DFT-based QSARs study of protoporphyrinogen oxidase inhibitors: phenyl triazolinones », *Bioorganic & medicinal chemistry*, vol. 12, nº 23, p. 6183-6191, 2004.
[264] V. Kamath et A. Pai, « Application of molecular descriptors in modern computational drug design-an overview », *Research Journal of Pharmacy and Technology*, vol. 10, nº 9, p. 3237-3241, 2017.
[265] I. Zamora, T. Oprea, G. Cruciani, M. Pastor, et A.-L. Ungell, « Surface descriptors for protein-ligand affinity prediction », *Journal of medicinal chemistry*, vol. 46, nº 1, p. 25-33, 2003.

[266] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, et D. A. Case, « Development and testing of a general amber force field », *Journal of computational chemistry*, vol. 25, nº 9, p. 1157-1174, 2004.
[267] P. S. Kharkar *et al.*, « Three-dimensional quantitative structure- activity relationship of 1, 4-dihydropyridines as antitubercular agents », *Journal of medicinal chemistry*, vol. 45, nº 22, p. 4858-4867, 2002.
[268] P. E. Blöchl, « Electrostatic decoupling of periodic images of plane-wave-expanded densities and derived atomic point charges », *The Journal of chemical physics*, vol. 103, nº 17, p. 7422-7428, 1995.
[269] J. Gasteiger et M. Marsili, « Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges », *Tetrahedron*, vol. 36, nº 22, p. 3219-3228, 1980.
[270] M. Marsili et J. Gasteiger, « Pi charge distribution from molecular topology and pi orbital electronegativity », *Croatica Chemica Acta*, vol. 53, nº 4, p. 601-614, 1980.
[271] A. Streitwieser et H. W. Salzberg, « Molecular orbital theory for organic chemists », *Journal of The Electrochemical Society*, vol. 109, nº 4, p. 116Cb, 1962.
[272] G. Del Re, B. Pullman, et T. Yonezawa, « Electronic structure of the α-amino acids of proteins: I. Charge distributions and proton chemical shifts », *Biochimica et biophysica acta*, vol. 75, p. 153-182, 1963.
[273] J. Uthuppan et K. Soni, « Conformational analysis: a review », *International Journal of Pharmaceutical Sciences and Research*, vol. 4, nº 1, p. 34, 2013.
[274] G. Vistoli et A. Pedretti, « 5.24 - Molecular Fields to Assess Recognition Forces and Property Spaces », in *Comprehensive Medicinal Chemistry II*, J. B. Taylor et D. J. Triggle, Éd., Oxford: Elsevier, 2007, p. 577-602. doi: 10.1016/B0-08-045044-X/00142-5.
[275] U. Norinder, « Recent progress in CoMFA methodology and related techniques », *Perspectives in drug discovery and design*, vol. 12, nº 0, p. 25-39, 1998.
[276] M. M. Gromiha, « Chapter 7 - Protein Interactions », in *Protein Bioinformatics*, M. M. Gromiha, Éd., Singapore: Academic Press, 2010, p. 247-302. doi: 10.1016/B978-8-1312-2297-3.50007-2.
[277] D. Ballabio, V. Consonni, A. Mauri, M. Claeys-Bruno, M. Sergent, et R. Todeschini, « A novel variable reduction method adapted from space-filling designs », *Chemometrics and Intelligent Laboratory Systems*, vol. 136, p. 147-154, 2014.
[278] S. T. Roweis et L. K. Saul, « Nonlinear dimensionality reduction by locally linear embedding », *science*, vol. 290, nº 5500, p. 2323-2326, 2000.
[279] C. Yoo et M. Shahlaei, « The applications of PCA in QSAR studies: A case study on CCR5 antagonists », *Chemical biology & drug design*, vol. 91, nº 1, p. 137-152, 2018.
[280] J. T. Leonard et K. Roy, « On selection of training and test sets for the development of predictive QSAR models », *QSAR & Combinatorial Science*, vol. 25, nº 3, p. 235-251, 2006.
[281] M. Paesmans *et al.*, « Prognostic factors for survival in advanced non-small-cell lung cancer: univariate and multivariate analyses including recursive partitioning and amalgamation algorithms in 1,052 patients. The European Lung Cancer Working Party. », *Journal of Clinical Oncology*, vol. 13, nº 5, p. 1221-1230, 1995.
[282] R. D. Clark, « OptiSim: an extended dissimilarity selection method for finding diverse representative subsets », *Journal of Chemical Information and Computer Sciences*, vol. 37, nº 6, p. 1181-1188, 1997.
[283] D. Granato, J. S. Santos, G. B. Escher, B. L. Ferreira, et R. M. Maggio, « Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective », *Trends in Food Science & Technology*, vol. 72, p. 83-90, 2018.
[284] N. Bhardwaj, B. Parashar, A. Kumar, D. Deepankar, et P. K. Jaslam, « Principal component analysis for the assessment of genetic diversity », in *AIP Conference Proceedings*, AIP Publishing, 2022.
[285] B. M. S. Hasan et A. M. Abdulazeez, « A review of principal component analysis algorithm for dimensionality reduction », *Journal of Soft Computing and Data Mining*, vol. 2, nº 1, p. 20-30, 2021.
[286] R. Sathya et A. Abraham, « Comparison of supervised and unsupervised learning algorithms for pattern classification », *International Journal of Advanced Research in Artificial Intelligence*, vol. 2, nº 2, p. 34-38, 2013.
[287] Douglas. Steinley, « K-means clustering: A half-century synthesis », *British Journal of Mathematical and Statistical Psychology*, vol. 59, nº 1, p. 1-34, 2006, doi: 10.1348/000711005X48266.

[288] « k-means clustering », XLSTAT, Your data analysis solution. Consulté le: 4 août 2023. [En ligne]. Disponible sur: https://www.xlstat.com/en/solutions/features/k-means-clustering
[289] W. D. Fisher, « On grouping for maximum homogeneity », *Journal of the American statistical Association*, vol. 53, nº 284, p. 789-798, 1958.
[290] R. D. Clark et W. J. Langton, « Balancing representativeness against diversity using optimizable K-dissimilarity and hierarchical clustering », *Journal of chemical information and computer sciences*, vol. 38, nº 6, p. 1079-1086, 1998.
[291] T. M. Martin *et al.*, « Does rational selection of training and test sets improve the outcome of QSAR modeling? », *Journal of chemical information and modeling*, vol. 52, nº 10, p. 2570-2578, 2012.
[292] « Student | Statistical Software for Excel », XLSTAT, Your data analysis solution. Consulté le: 30 avril 2023. [En ligne]. Disponible sur: https://www.xlstat.com/en/pricing/student
[293] webdev, « Certara Enhances SYBYL-X Drug Design and Discovery Software Suite », Certara. Consulté le: 30 avril 2023. [En ligne]. Disponible sur: https://www.certara.com/pressrelease/certara-enhances-sybyl-x-drug-design-and-discovery-software-suite/
[294] O. Daoui, S. Elkhattabi, S. Chtita, R. Elkhalabi, H. Zgou, et A. T. Benjelloun, « QSAR, molecular docking and ADMET properties in silico studies of novel 4,5,6,7-tetrahydrobenzo[D]-thiazol-2-Yl derivatives derived from dimedone as potent anti-tumor agents through inhibition of C-Met receptor tyrosine kinase », *Heliyon*, vol. 7, nº 7, p. e07463, juill. 2021, doi: 10.1016/j.heliyon.2021.e07463.
[295] H. R. Seireg, Y. M. Omar, F. E. Abd El-Samie, A. S. El-Fishawy, et A. Elmahalawy, « Ensemble machine learning techniques using computer simulation data for wild blueberry yield prediction », *IEEE Access*, vol. 10, p. 64671-64687, 2022.
[296] K. J. Grimm et N. Ram, « Nonlinear growth models in M plus and SAS », *Structural Equation Modeling*, vol. 16, nº 4, p. 676-701, 2009.
[297] J. Kim, J. Hong, et H. Park, « Prospects of deep learning for medical imaging », *Precis Future Med*, vol. 2, nº 2, p. 37-52, juin 2018, doi: 10.23838/pfm.2018.00030.
[298] T. H. Abraham, « (Physio)logical circuits: The intellectual origins of the McCulloch–Pitts neural networks », *Journal of the History of the Behavioral Sciences*, vol. 38, nº 1, p. 3-25, 2002, doi: 10.1002/jhbs.1094.
[299] A. Riadi et M. H. Botutihe, « Visitor satisfaction prediction of the'Pantai Pohon Cinta'beach tourism using the backpropagation algorithm with particle swarm optimization feature selection », *no*, vol. 1, p. 85, 2021.
[300] M. B. Rao, « Feedforward neural network methodology », *Technometrics*, vol. 42, nº 4, p. 432, 2000.
[301] A. Ismail, D.-S. Jeng, L. L. Zhang, et J.-S. Zhang, « Predictions of bridge scour: application of a feed-forward neural network with an adaptive activation function », *Engineering Applications of Artificial Intelligence*, vol. 26, nº 5-6, p. 1540-1549, 2013.
[302] C. Lv *et al.*, « Levenberg–Marquardt Backpropagation Training of Multilayer Neural Networks for State Estimation of a Safety-Critical Cyber-Physical System », *IEEE Transactions on Industrial Informatics*, vol. 14, nº 8, p. 3436-3446, août 2018, doi: 10.1109/TII.2017.2777460.
[303] V. K\uurková, « Kolmogorov's theorem and multilayer neural networks », *Neural networks*, vol. 5, nº 3, p. 501-506, 1992.
[304] O. Daoui *et al.*, « 3D-QSAR, ADME-Tox, and molecular docking of semisynthetic triterpene derivatives as antibacterial and insecticide agents », *Struct Chem*, vol. 33, nº 4, p. 1063-1084, août 2022, doi: 10.1007/s11224-022-01912-4.
[305] R. E. Mouhi *et al.*, « A strategy to enhance VOC of π-conjugated molecules based on thieno[2,3-b] indole for applications in bulk heterojunction organic solar cells using DFT, TD-DFT, and 3D-QSPR modeling studies », *New J. Chem.*, vol. 47, nº 2, p. 812-827, janv. 2023, doi: 10.1039/D2NJ04281A.
[306] G. E. Kellogg et S. F. Semus, « 3D QSAR in modern drug design », in *Modern Methods of Drug Discovery*, A. Hillisch et R. Hilgenfeld, Éd., in EXS. , Basel: Birkhäuser, 2003, p. 223-241. doi: 10.1007/978-3-0348-7997-2_11.
[307] S. Pérez, C. Meyer, et A. Imberty, « Practical tools for molecular modeling of complex carbohydrates and their interactions with proteins », *Mol Eng*, vol. 5, nº 1, p. 271-300, mars 1995, doi: 10.1007/BF00999595.

[308] W. Zhu *et al.*, « QSAR analyses on ginkgolides and their analogues using CoMFA, CoMSIA, and HQSAR », *Bioorganic & Medicinal Chemistry*, vol. 13, nº 2, p. 313-322, janv. 2005, doi: 10.1016/j.bmc.2004.10.027.
[309] O. Daoui, S. Elkhattabi, et S. Chtita, « Rational identification of small molecules derived from 9,10-dihydrophenanthrene as potential inhibitors of 3CLpro enzyme for COVID-19 therapy: a computer-aided drug design approach », *Struct Chem*, vol. 33, nº 5, p. 1667-1690, oct. 2022, doi: 10.1007/s11224-022-02004-z.
[310] K. Jagiello *et al.*, « Advantages and limitations of classic and 3D QSAR approaches in nano-QSAR studies based on biological activity of fullerene derivatives », *Journal of Nanoparticle Research*, vol. 18, p. 1-16, 2016.
[311] R. D. Cramer III, J. D. Bunce, D. E. Patterson, et I. E. Frank, « Crossvalidation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies », *Quantitative Structure-Activity Relationships*, vol. 7, nº 1, p. 18-25, 1988, doi: 10.1002/qsar.19880070105.
[312] A. Golbraikh et A. Tropsha, « Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection », *Mol Divers*, vol. 5, nº 4, p. 231-243, déc. 2000, doi: 10.1023/A:1021372108686.
[313] F. Sahigara, K. Mansouri, D. Ballabio, A. Mauri, V. Consonni, et R. Todeschini, « Comparison of Different Approaches to Define the Applicability Domain of QSAR Models », *Molecules*, vol. 17, nº 5, Art. nº 5, mai 2012, doi: 10.3390/molecules17054791.
[314] C. Rücker, G. Rücker, et M. Meringer, « Y-randomization–a useful tool in QSAR validation, or folklore », *J. Chem. Inf. Model*, vol. 47, p. 2345-2357, 2007.
[315] A. Golbraikh et A. Tropsha, « Beware of q2! », *Journal of Molecular Graphics and Modelling*, vol. 20, nº 4, p. 269-276, janv. 2002, doi: 10.1016/S1093-3263(01)00123-1.
[316] R. Todeschini, V. Consonni, et P. Gramatica, « 4.05 - Chemometrics in QSAR », in *Comprehensive Chemometrics*, S. D. Brown, R. Tauler, et B. Walczak, Éd., Oxford: Elsevier, 2009, p. 129-172. doi: 10.1016/B978-044452701-1.00007-7.
[317] N. Nikolova-Jeliazkova et J. Jaworska, « An approach to determining applicability domains for QSAR group contribution models: an analysis of SRC KOWWIN », *Altern Lab Anim*, vol. 33, nº 5, p. 461-470, oct. 2005, doi: 10.1177/026119290503300510.
[318] P. Gramatica, « Principles of QSAR models validation: internal and external », *QSAR & combinatorial science*, vol. 26, nº 5, p. 694-701, 2007.
[319] G. Diana et C. Tommasi, « Cross-validation methods in principal component analysis: a comparison », *Statistical Methods and Applications*, vol. 11, nº 1, p. 71-82, 2002.
[320] T. I. Netzeva *et al.*, « Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships: The Report and Recommendations of ECVAM Workshop 521,2 », *Altern Lab Anim*, vol. 33, nº 2, p. 155-173, avr. 2005, doi: 10.1177/026119290503300209.
[321] « Microsoft Office 2019 | Previous Version | Microsoft 365 ». Consulté le: 13 mai 2023. [En ligne]. Disponible sur: https://www.microsoft.com/en-us/microsoft-365/previous-versions/microsoft-office-2019
[322] « DTC lab », DTC lab. Consulté le: 13 mai 2023. [En ligne]. Disponible sur: https://dtclab.webs.com/
[323] « Matlab toolboxes », Milano Chemometrics and QSAR Research Group. Consulté le: 13 mai 2023. [En ligne]. Disponible sur: https://michem.unimib.it/download/matlab-toolboxes/
[324] O. Daoui, S. Elkhattabi, et S. Chtita, « Design and Prediction of ADME/Tox Properties of Novel Magnolol Derivatives as Anticancer Agents for NSCLC Using 3D-QSAR, Molecular Docking, MOLCAD and MM-GBSA Studies », *Letters in Drug Design & Discovery*, vol. 20, nº 5, p. 545-569, 2022, doi: http://dx.doi.org/10.2174/1570180819666220510141710.
[325] M. Skalic, J. Jiménez, D. Sabbadin, et G. De Fabritiis, « Shape-Based Generative Modeling for de Novo Drug Design », *J. Chem. Inf. Model.*, vol. 59, nº 3, p. 1205-1214, mars 2019, doi: 10.1021/acs.jcim.8b00706.
[326] A. Heifetz, M. Southey, I. Morao, A. Townsend-Nicholson, et M. J. Bodkin, « Computational Methods Used in Hit-to-Lead and Lead Optimization Stages of Structure-Based Drug Discovery », in *Computational Methods for GPCR Drug Discovery*, A. Heifetz, Éd., in Methods in Molecular Biology. , New York, NY: Springer, 2018, p. 375-394. doi: 10.1007/978-1-4939-7465-8_19.

[327] H. Yu et A. Adedoyin, « ADME–Tox in drug discovery: integration of experimental and computational technologies », *Drug Discovery Today*, vol. 8, nº 18, p. 852-861, sept. 2003, doi: 10.1016/S1359-6446(03)02828-9.
[328] L. L. G. Ferreira et A. D. Andricopulo, « ADMET modeling approaches in drug discovery », *Drug Discov Today*, vol. 24, nº 5, p. 1157-1165, mai 2019, doi: 10.1016/j.drudis.2019.03.015.
[329] M. Danhof, E. C. de Lange, O. E. Della Pasqua, B. A. Ploeger, et R. A. Voskuyl, « Mechanism-based pharmacokinetic-pharmacodynamic (PK-PD) modeling in translational drug research », *Trends in pharmacological sciences*, vol. 29, nº 4, p. 186-191, 2008.
[330] H. Derendorf et B. Meibohm, « Modeling of pharmacokinetic/pharmacodynamic (PK/PD) relationships: concepts and perspectives », *Pharmaceutical research*, vol. 16, p. 176-185, 1999.
[331] S. Tian, J. Wang, Y. Li, D. Li, L. Xu, et T. Hou, « The application of in silico drug-likeness predictions in pharmaceutical research », *Advanced Drug Delivery Reviews*, vol. 86, p. 2-10, juin 2015, doi: 10.1016/j.addr.2015.01.009.
[332] A. Daina, O. Michielin, et V. Zoete, « SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules », *Scientific reports*, vol. 7, nº 1, p. 1-13, 2017.
[333] D. E. V. Pires, T. L. Blundell, et D. B. Ascher, « pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures », *J. Med. Chem.*, vol. 58, nº 9, p. 4066-4072, mai 2015, doi: 10.1021/acs.jmedchem.5b00104.
[334] « Lipinski: a, Lombardo, F., Dominy, BW & Feeney, PJ... - Google Scholar ». Consulté le: 17 mai 2023. [En ligne]. Disponible sur: https://scholar.google.com/scholar_lookup?title=Experimental%20and%20computational%20approaches%20to%20estimate%20solubility%20and%20permeability%20in%20drug%20discovery%20and%20development%20settings&journal=Adv.%20Drug.%20Deliv.%20Rev.&doi=10.1016%2FS0169-409X%2800%2900129-0&volume=46&pages=3-26&publication_year=2001&author=Lipinski%2CCA&author=Lombardo%2CF&author=Dominy%2CBW&author=Feeney%2CPJ
[335] A. K. Ghose, V. N. Viswanadhan, et J. J. Wendoloski, « A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases », *Journal of combinatorial chemistry*, vol. 1, nº 1, p. 55-68, 1999.
[336] D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward, et K. D. Kopple, « Molecular properties that influence the oral bioavailability of drug candidates », *Journal of medicinal chemistry*, vol. 45, nº 12, p. 2615-2623, 2002.
[337] « Muegge: Simple selection criteria for drug-like chemical... - Google Scholar ». Consulté le: 17 mai 2023. [En ligne]. Disponible sur: https://scholar.google.com/scholar_lookup?title=Simple%20selection%20criteria%20for%20drug-like%20chemical%20matter&journal=J.%20Med.%20Chem.&doi=10.1021%2Fjm015507e&volume=44&pages=1841-1846&publication_year=2001&author=Muegge%2CI&author=Heald%2CSL&author=Brittelli%2CD
[338] C. A. Lipinski, « Lead- and drug-like compounds: the rule-of-five revolution », *Drug Discovery Today: Technologies*, vol. 1, nº 4, p. 337-341, déc. 2004, doi: 10.1016/j.ddtec.2004.11.007.
[339] « Molecular Properties Prediction - Osiris Property Explorer ». Consulté le: 19 mai 2023. [En ligne]. Disponible sur: https://www.organic-chemistry.org/prog/peo/
[340] N. Fayyazi, A. Fassihi, S. Esmaeili, S. Taheri, J. B. Ghasemi, et L. Saghaie, « Molecular dynamics simulation and 3D-pharmacophore analysis of new quinoline-based analogues with dual potential against EGFR and VEGFR-2 », *International journal of biological macromolecules*, vol. 142, p. 94-113, 2020.
[341] S. Akkoc *et al.*, « Drug design of new therapeutic agents: molecular docking, molecular dynamics simulation, DFT and POM analyses of new Schiff base ligands and impact of substituents on bioactivity of their potential antifungal pharmacophore site », *Journal of Biomolecular Structure and Dynamics*, p. 1-14, 2022.
[342] Y. Wang, M. Liu, et J. Gao, « Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions », *Proceedings of the National Academy of Sciences*, vol. 117, nº 25, p. 13967-13974, 2020.
[343] V. D'Argenio, « The High-Throughput Analyses Era: Are We Ready for the Data Struggle? », *High-Throughput*, vol. 7, nº 1, Art. nº 1, mars 2018, doi: 10.3390/ht7010008.

[344] A. Cooper, « Thermodynamic analysis of biomolecular interactions », *Curr Opin Chem Biol*, vol. 3, nº 5, p. 557-563, oct. 1999, doi: 10.1016/s1367-5931(99)00008-3.
[345] T. Lengauer et M. Rarey, « Computational methods for biomolecular docking », *Current opinion in structural biology*, vol. 6, nº 3, p. 402-406, 1996.
[346] F.-F. Liu, T. Wang, X.-Y. Dong, et Y. Sun, « Rational design of affinity peptide ligand by flexible docking simulation », *Journal of Chromatography A*, vol. 1146, nº 1, p. 41-50, 2007.
[347] V. Salmaso et S. Moro, « Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview », *Frontiers in Pharmacology*, vol. 9, 2018, Consulté le: 19 mai 2023. [En ligne]. Disponible sur: https://www.frontiersin.org/articles/10.3389/fphar.2018.00923
[348] J. Liu et R. Wang, « Classification of current scoring functions », *J Chem Inf Model*, vol. 55, nº 3, p. 475-482, mars 2015, doi: 10.1021/ci500731a.
[349] A. D. Hill et P. J. Reilly, « Scoring functions for AutoDock », *Glycoinformatics*, p. 467-474, 2015.
[350] J. Li, A. Fu, et L. Zhang, « An Overview of Scoring Functions Used for Protein-Ligand Interactions in Molecular Docking », *Interdiscip Sci*, vol. 11, nº 2, p. 320-328, juin 2019, doi: 10.1007/s12539-019-00327-w.
[351] O. Trott et A. J. Olson, « AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading », *Journal of computational chemistry*, vol. 31, nº 2, p. 455-461, 2010.
[352] R. A. Friesner *et al.*, « Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy », *J. Med. Chem.*, vol. 47, nº 7, p. 1739-1749, mars 2004, doi: 10.1021/jm0306430.
[353] G. Wu, D. H. Robertson, C. L. Brooks, et M. Vieth, « Detailed analysis of grid-based molecular docking: A case study of CDOCKER-A CHARMm-based MD docking algorithm », *J Comput Chem*, vol. 24, nº 13, p. 1549-1562, oct. 2003, doi: 10.1002/jcc.10306.
[354] K. A, K. Pd, J. X, V. Cm, et W. M, « LigScore: a novel scoring function for predicting binding affinities », *Journal of molecular graphics & modelling*, vol. 23, nº 5, avr. 2005, doi: 10.1016/j.jmgm.2004.11.007.
[355] S. Dallakyan et A. J. Olson, « Small-Molecule Library Screening by Docking with PyRx », in *Chemical Biology: Methods and Protocols*, J. E. Hempel, C. H. Williams, et C. C. Hong, Éd., in Methods in Molecular Biology. , New York, NY: Springer, 2015, p. 243-250. doi: 10.1007/978-1-4939-2269-7_19.
[356] R. Wang, Y. Lu, X. Fang, et S. Wang, « An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein- ligand complexes », *Journal of chemical information and computer sciences*, vol. 44, nº 6, p. 2114-2125, 2004.
[357] X. Tao *et al.*, « Recent developments in molecular docking technology applied in food science: a review », *International Journal of Food Science & Technology*, vol. 55, nº 1, p. 33-45, 2020.
[358] « What is the difference between the blind docking and binding site based docking? », Bioinformatics Review. Consulté le: 19 mai 2023. [En ligne]. Disponible sur: https://bioinformaticsreview.com/question/what-is-the-difference-between-the-blind-docking-and-binding-site-based-docking/
[359] O. Daoui, H. Nour, O. Abchir, S. Elkhattabi, M. Bakhouch, et S. Chtita, « A computer-aided drug design approach to explore novel type II inhibitors of c-Met receptor tyrosine kinase for cancer therapy: QSAR, molecular docking, ADMET and molecular dynamics simulations », *Journal of Biomolecular Structure and Dynamics*, vol. 0, nº 0, p. 1-18, sept. 2022, doi: 10.1080/07391102.2022.2124456.
[360] O. Daoui, S. N. Mali, K. Elkhattabi, S. Elkhattabi, et S. Chtita, « Repositioning Cannabinoids and Terpenes as Novel EGFR-TKIs Candidates for Targeted Therapy Against Cancer: A virtual screening model using CADD and biophysical simulations », *Heliyon*, vol. 9, nº 4, p. e15545, avr. 2023, doi: 10.1016/j.heliyon.2023.e15545.
[361] J. B. Ghasemi, A. Abdolmaleki, et F. Shiri, « Molecular docking challenges and limitations », in *Pharmaceutical sciences: Breakthroughs in research and practice*, IGI Global, 2017, p. 770-794.
[362] R. Jakhar, M. Dangi, A. Khichi, et A. K. Chhillar, « Relevance of molecular docking studies in drug designing », *Current Bioinformatics*, vol. 15, nº 4, p. 270-278, 2020.
[363] O. Korb *et al.*, « Potential and limitations of ensemble docking », *Journal of chemical information and modeling*, vol. 52, nº 5, p. 1262-1274, 2012.
[364] R. Huey, G. M. Morris, A. J. Olson, et D. S. Goodsell, « A semiempirical free energy force field with charge-based desolvation », *Journal of Computational Chemistry*, vol. 28, nº 6, p. 1145-1152, 2007, doi: 10.1002/jcc.20634.

[365] S. Raschka, « Molecular docking, estimating free energies of binding, and AutoDock's semi-empirical force field », Sebastian Raschka, PhD. Consulté le: 24 mai 2023. [En ligne]. Disponible sur: https://sebastianraschka.com/Articles/2014_autodock_energycomps.html
[366] « Zentrum für Bioinformatik: Universität Hamburg - Proteins Plus Server ». Consulté le: 23 mai 2023. [En ligne]. Disponible sur: https://proteins.plus/
[367] T. Tuccinardi, « What is the current value of MM/PBSA and MM/GBSA methods in drug discovery? », *Expert opinion on drug discovery*, vol. 16, nº 11, p. 1233-1237, 2021.
[368] H. Alonso, A. A. Bliznyuk, et J. E. Gready, « Combining docking and molecular dynamic simulations in drug design », *Medicinal research reviews*, vol. 26, nº 5, p. 531-568, 2006.
[369] C. J. Morris et D. D. Corte, « Using molecular docking and molecular dynamics to investigate protein-ligand interactions », *Modern Physics Letters B*, vol. 35, nº 08, p. 2130002, 2021.
[370] M. Sgobba, F. Caporuscio, A. Anighoro, C. Portioli, et G. Rastelli, « Application of a post-docking procedure based on MM-PBSA and MM-GBSA on single and multiple protein conformations », *European Journal of Medicinal Chemistry*, vol. 58, p. 431-440, déc. 2012, doi: 10.1016/j.ejmech.2012.10.024.
[371] P. Dauber-Osguthorpe, V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, et A. T. Hagler, « Structure and energetics of ligand binding to proteins: Escherichia coli dihydrofolate reductase-trimethoprim, a drug-receptor system », *Proteins: Structure, Function, and Bioinformatics*, vol. 4, nº 1, p. 31-47, 1988.
[372] M. Zhao *et al.*, « MARTINI-compatible coarse-grained model for the mesoscale simulation of peptoids », *The Journal of Physical Chemistry B*, vol. 124, nº 36, p. 7745-7764, 2020.
[373] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, et A. Kolinski, « Coarse-Grained Protein Models and Their Applications », *Chem. Rev.*, vol. 116, nº 14, p. 7898-7936, juill. 2016, doi: 10.1021/acs.chemrev.6b00163.
[374] S. Riniker, « Fixed-charge atomistic force fields for molecular dynamics simulations in the condensed phase: an overview », *Journal of chemical information and modeling*, vol. 58, nº 3, p. 565-578, 2018.
[375] R. Dias, J. de Azevedo, et F. Walter, « Molecular docking algorithms », *Current drug targets*, vol. 9, nº 12, p. 1040-1047, 2008.
[376] S. Hug, « Classical Molecular Dynamics in a Nutshell », in *Biomolecular Simulations: Methods and Protocols*, L. Monticelli et E. Salonen, Éd., in Methods in Molecular Biology. , Totowa, NJ: Humana Press, 2013, p. 127-152. doi: 10.1007/978-1-62703-017-5_6.
[377] « Les méthodes in silico dans la recherche pharmaceutique, exemple d'application pour l'étude pharmacocinétique post-autorisation de mise sur le marché des β-bloquants utilisés pour le traitement du glaucome ». Consulté le: 28 mai 2023. [En ligne]. Disponible sur: https://www.idref.fr/256511551
[378] P. Mark et L. Nilsson, « Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K », *The Journal of Physical Chemistry A*, vol. 105, nº 43, p. 9954-9960, 2001.
[379] M. S. Valdés-Tresanco, M. E. Valdés-Tresanco, P. A. Valiente, et E. Moreno, « gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS », *Journal of chemical theory and computation*, vol. 17, nº 10, p. 6281-6291, 2021.
[380] E. Wang *et al.*, « VAD-MM/GBSA: A Variable Atomic Dielectric MM/GBSA Model for Improved Accuracy in Protein–Ligand Binding Free Energy Calculations », *Journal of Chemical Information and Modeling*, vol. 61, nº 6, p. 2844-2856, 2021.
[381] B. Anderson, « Prime, Schrodinger, LLC, New York, NY (2017) », 2008.
[382] L. J, A. R, Z. K, C. Y, Z. S, et F. Ra, « The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling », *Proteins*, vol. 79, nº 10, oct. 2011, doi: 10.1002/prot.23106.
[383] K. Roos *et al.*, « OPLS3e: Extending force field coverage for drug-like small molecules », *Journal of chemical theory and computation*, vol. 15, nº 3, p. 1863-1874, 2019.
[384] J. R. Greenwood, D. Calkins, A. P. Sullivan, et J. C. Shelley, « Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution », *J Comput Aided Mol Des*, vol. 24, nº 6-7, p. 591-604, juin 2010, doi: 10.1007/s10822-010-9349-1.
[385] W. F. van Gunsteren et H. J. C. Berendsen, « Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry », *Angewandte Chemie International Edition in English*, vol. 29, nº 9, p. 992-1023, 1990, doi: 10.1002/anie.199009921.
[386] C. Becquart et M. Perez, « Dynamique moléculaire appliquée aux matériaux », *Techniques de l'Ingénieur*, p. 169, 2010.

[387] T. Darden, D. York, et L. Pedersen, « Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems », *The Journal of chemical physics*, vol. 98, nº 12, p. 10089-10092, 1993.
[388] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, et L. G. Pedersen, « A smooth particle mesh Ewald method », *The Journal of chemical physics*, vol. 103, nº 19, p. 8577-8593, 1995.
[389] L. Verlet, « Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules », *Physical review*, vol. 159, nº 1, p. 98, 1967.
[390] L. Verlet, « Computer" experiments" on classical fluids. II. Equilibrium correlation functions », *Physical Review*, vol. 165, nº 1, p. 201, 1968.
[391] R. W. Hockney, « The Potenitial Calculation and Some Applications », *Methods Comput. Phys.*, vol. 20, p. 135, 1970.
[392] W. C. Swope, H. C. Andersen, P. H. Berens, et K. R. Wilson, « A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters », *The Journal of chemical physics*, vol. 76, nº 1, p. 637-649, 1982.
[393] R. T. Cygan, J. A. Greathouse, H. Heinz, et A. G. Kalinichev, « Molecular models and simulations of layered materials », *Journal of Materials Chemistry*, vol. 19, nº 17, p. 2470-2481, 2009.
[394] A. Kazakov, « Computer Simulation Methods of Soft Matter », 2023.
[395] X. Yong et L. T. Zhang, « Thermostats and thermostat strategies for molecular dynamics simulations of nanofluidics », *The Journal of chemical physics*, vol. 138, nº 8, p. 084503, 2013.
[396] Q. Ke, X. Gong, S. Liao, C. Duan, et L. Li, « Effects of thermostats/barostats on physical properties of liquids by molecular dynamics simulations », *Journal of Molecular Liquids*, vol. 365, p. 120116, 2022.
[397] O. M. H. Salo-Ahen *et al.*, « Molecular Dynamics Simulations in Drug Discovery and Pharmaceutical Development », *Processes*, vol. 9, nº 1, Art. nº 1, janv. 2021, doi: 10.3390/pr9010071.
[398] M. S. Friedrichs *et al.*, « Accelerating molecular dynamic simulation on graphics processing units », *Journal of computational chemistry*, vol. 30, nº 6, p. 864-872, 2009.
[399] J. D. Owens, M. Houston, D. Luebke, S. Green, J. E. Stone, et J. C. Phillips, « GPU computing », *Proceedings of the IEEE*, vol. 96, nº 5, p. 879-899, 2008.
[400] E. Chow *et al.*, « Desmond performance on a cluster of multicore processors », *DE Shaw Research Technical Report DESRES/TR–2008-01*, 2008.
[401] M. J. Abraham *et al.*, « GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers », *SoftwareX*, vol. 1-2, p. 19-25, sept. 2015, doi: 10.1016/j.softx.2015.06.001.
[402] J. C. Phillips *et al.*, « Scalable molecular dynamics on CPU and GPU architectures with NAMD », *The Journal of chemical physics*, vol. 153, nº 4, p. 044130, 2020.
[403] R. Salomon-Ferrer, D. A. Case, et R. C. Walker, « An overview of the Amber biomolecular simulation package », *WIREs Computational Molecular Science*, vol. 3, nº 2, p. 198-210, 2013, doi: 10.1002/wcms.1121.
[404] « Protein-Ligand Complex ». Consulté le: 5 juin 2023. [En ligne]. Disponible sur: http://www.mdtutorials.com/gmx/complex/index.html
[405] A. Hospital, J. R. Goñi, M. Orozco, et J. L. Gelpí, « Molecular dynamics simulations: advances and applications », *Advances and Applications in Bioinformatics and Chemistry*, vol. 8, p. 37-47, déc. 2015, doi: 10.2147/AABC.S70333.
[406] M. Batool, B. Ahmad, et S. Choi, « A Structure-Based Drug Discovery Paradigm », *International Journal of Molecular Sciences*, vol. 20, nº 11, Art. nº 11, janv. 2019, doi: 10.3390/ijms20112783.
[407] R. M. Mohareb, A. S. Abouzied, et N. S. Abbas, « Synthesis and Biological Evaluation of Novel 4, 5, 6, 7-Tetrahydrobenzo [D]-Thiazol-2-Yl Derivatives Derived from Dimedone with Anti-Tumor, c-Met, Tyrosine Kinase and Pim-1 Inhibitions », *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)*, vol. 19, nº 12, p. 1438-1453, 2019.
[408] J. Liu *et al.*, « Novel 4-phenoxypyridine derivatives bearing imidazole-4-carboxamide and 1, 2, 4-triazole-3-carboxamide moieties: design, synthesis and biological evaluation as potent antitumor agents », *Bioorganic Chemistry*, p. 105629, 2022.
[409] R. M. Mohareb, Y. R. Milad, et A. A. Masoud, « New Approaches for the Synthesis of Heterocyclic Compounds Derived from Cyclohexan-1, 3-dione with Anti-proliferative Activities », *Acta Chimica Slovenica*, vol. 68, nº 1, p. 72-87, 2021.
[410] G. Turan-Zitouni, M. D. Altıntop, A. Özdemir, Z. A. Kaplancıklı, G. A. Çiftçi, et H. E. Temel, « Synthesis and evaluation of bis-thiazole derivatives as new anticancer agents », *European*

Journal of Medicinal Chemistry, vol. 107, p. 288-294, janv. 2016, doi: 10.1016/j.ejmech.2015.11.002.
[411] R. Pignatello *et al.*, « Synthesis and biological evaluation of thiazolo-triazole derivatives », *European Journal of Medicinal Chemistry*, vol. 26, nº 9, p. 929-938, déc. 1991, doi: 10.1016/0223-5234(91)90135-A.
[412] G. Trapani *et al.*, « Synthesis and anticonvulsant activity of some 1,2,3,3a-tetrahydropyrrolo[2,1-b]benzothiazol-1-ones and pyrrolo[2,1-b]thiazole analogues », *European Journal of Medicinal Chemistry*, vol. 29, nº 3, p. 197-204, janv. 1994, doi: 10.1016/0223-5234(94)90038-8.
[413] H. Yu *et al.*, « Synthesis and Insecticidal Activity of N-Substituted (1,3-Thiazole)alkyl Sulfoximine Derivatives », *J. Agric. Food Chem.*, vol. 56, nº 23, p. 11356-11360, déc. 2008, doi: 10.1021/jf802802g.
[414] A. B. A. El-Gazzar, M. M. Youssef, A. M. S. Youssef, A. A. Abu-Hashem, et F. A. Badria, « Design and synthesis of azolopyrimidoquinolines, pyrimidoquinazolines as anti-oxidant, anti-inflammatory and analgesic activities », *European Journal of Medicinal Chemistry*, vol. 44, nº 2, p. 609-624, févr. 2009, doi: 10.1016/j.ejmech.2008.03.022.
[415] R. Dahmani, M. Manachou, S. Belaidi, S. Chtita, et S. Boughdiri, « Structural characterization and QSAR modeling of 1,2,4-triazole derivatives as α-glucosidase inhibitors », *New J. Chem.*, vol. 45, nº 3, p. 1253-1261, janv. 2021, doi: 10.1039/D0NJ05298A.
[416] X.-L. Qiu *et al.*, « Synthesis and Biological Evaluation of a Series of Novel Inhibitor of Nek2/Hec1 Analogues », *J. Med. Chem.*, vol. 52, nº 6, p. 1757-1767, mars 2009, doi: 10.1021/jm8015969.
[417] H.-R. Tsou *et al.*, « Discovery and optimization of 2-(4-substituted-pyrrolo [2, 3-b] pyridin-3-yl) methylene-4-hydroxybenzofuran-3 (2H)-ones as potent and selective ATP-competitive inhibitors of the mammalian target of rapamycin (mTOR) », *Bioorganic & medicinal chemistry letters*, vol. 20, nº 7, p. 2321-2325, 2010.
[418] K. M. Dawood, T. M. A. Eldebss, H. S. A. El-Zahabi, M. H. Yousef, et P. Metz, « Synthesis of some new pyrazole-based 1,3-thiazoles and 1,3,4-thiadiazoles as anticancer agents », *European Journal of Medicinal Chemistry*, vol. 70, p. 740-749, déc. 2013, doi: 10.1016/j.ejmech.2013.10.042.
[419] I. Cañadas, F. Rojo, M. Arumí-Uría, A. Rovira, J. Albanell, et E. Arriola, « C-MET as a new therapeutic target for the development of novel anticancer drugs », *Clin Transl Oncol*, vol. 12, nº 4, p. 253-260, avr. 2010, doi: 10.1007/s12094-010-0501-0.
[420] J. Tanizaki *et al.*, « MET Tyrosine Kinase Inhibitor Crizotinib (PF-02341066) Shows Differential Antitumor Effects in Non-small Cell Lung Cancer According to MET Alterations », *Journal of Thoracic Oncology*, vol. 6, nº 10, p. 1624-1631, oct. 2011, doi: 10.1097/JTO.0b013e31822591e9.
[421] J. J. Cui *et al.*, « Structure based drug design of crizotinib (PF-02341066), a potent and selective dual inhibitor of mesenchymal–epithelial transition factor (c-MET) kinase and anaplastic lymphoma kinase (ALK) », *Journal of medicinal chemistry*, vol. 54, nº 18, p. 6342-6363, 2011.
[422] Y. Xu et Y. Fan, « Responses to crizotinib can occur in c-MET overexpressing nonsmall cell lung cancer after developing EGFR-TKI resistance », *Cancer biology & therapy*, vol. 20, nº 2, p. 145-149, 2019.
[423] U. M. Zanger et M. Schwab, « Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation », *Pharmacology & therapeutics*, vol. 138, nº 1, p. 103-141, 2013.
[424] A. Puccini *et al.*, « Safety and tolerability of c-MET inhibitors in cancer », *Drug safety*, vol. 42, nº 2, p. 211-233, 2019.
[425] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, et P. Walter, « Protein Function », *Molecular Biology of the Cell. 4th edition*, 2002, Consulté le: 27 avril 2022. [En ligne]. Disponible sur: https://www.ncbi.nlm.nih.gov/books/NBK26911/
[426] T. Damghani, T. Sedghamiz, S. Sharifi, et S. Pirhadi, « Critical c-Met-inhibitor interactions resolved from molecular dynamics simulations of different c-Met complexes », *Journal of Molecular Structure*, vol. 1203, p. 127456, mars 2020, doi: 10.1016/j.molstruc.2019.127456.
[427] T. Damghani, M. Elyasi, S. Pirhadi, Z. Haghighijoo, et S. Ghazi, « Type II c-Met inhibitors: molecular insight into crucial interactions for effective inhibition », *Molecular Diversity*, p. 1-13, 2021.
[428] R. Shang *et al.*, « Cabozantinib-based combination therapy for the treatment of hepatocellular carcinoma », *Gut*, vol. 70, nº 9, p. 1746-1757, sept. 2021, doi: 10.1136/gutjnl-2020-320716.
[429] Z.-X. Wu *et al.*, « Tivantinib, A c-Met Inhibitor in Clinical Trials, Is Susceptible to ABCG2-Mediated Drug Resistance », *Cancers (Basel)*, vol. 12, nº 1, p. E186, janv. 2020, doi: 10.3390/cancers12010186.

[430] N. M. Ayoub, K. M. Al-Shami, M. A. Alqudah, et N. M. Mhaidat, « Crizotinib, a MET inhibitor, inhibits growth, migration, and invasion of breast cancer cells in vitro and synergizes with chemotherapeutic agents », *Onco Targets Ther*, vol. 10, p. 4869-4883, oct. 2017, doi: 10.2147/OTT.S148604.
[431] M. Zillhardt *et al.*, « Foretinib (GSK1363089), an orally available multi-kinase inhibitor of c-Met and VEGFR-2, blocks proliferation, induces anoikis, and impairs ovarian cancer metastasis », *Clin Cancer Res*, vol. 17, nº 12, p. 4042-4051, juin 2011, doi: 10.1158/1078-0432.CCR-10-3387.
[432] M. A. Shah *et al.*, « A Randomized Phase II Study of FOLFOX With or Without the MET Inhibitor Onartuzumab in Advanced Adenocarcinoma of the Stomach and Gastroesophageal Junction », *Oncologist*, vol. 21, nº 9, p. 1085-1090, sept. 2016, doi: 10.1634/theoncologist.2016-0038.
[433] Y. Dai et D. W. Siemann, « BMS-777607, a small-molecule met kinase inhibitor, suppresses hepatocyte growth factor-stimulated prostate cancer metastatic phenotype in vitro », *Mol Cancer Ther*, vol. 9, nº 6, p. 1554-1561, juin 2010, doi: 10.1158/1535-7163.MCT-10-0359.
[434] Y. Piao *et al.*, « Novel MET/TIE2/VEGFR2 inhibitor altiratinib inhibits tumor growth and invasiveness in bevacizumab-resistant glioblastoma mouse models », *Neuro Oncol*, vol. 18, nº 9, p. 1230-1241, sept. 2016, doi: 10.1093/neuonc/now030.
[435] J.-S. Shin *et al.*, « NPS-1034, a novel MET inhibitor, inhibits the activated MET receptor and its constitutively active mutants », *Invest New Drugs*, vol. 32, nº 3, p. 389-399, juin 2014, doi: 10.1007/s10637-013-0039-4.
[436] C. Liang *et al.*, « Self-controlled release of Oxaliplatin prodrug from d-α-tocopheryl polyethylene glycol 1000 succinate (TPGS) functionalized mesoporous silica nanoparticles for cancer therapy », *Journal of Colloid and Interface Science*, vol. 525, p. 1-10, sept. 2018, doi: 10.1016/j.jcis.2018.04.058.
[437] M. Zhao *et al.*, « Synthesis and evaluation of new compounds bearing 3-(4-aminopiperidin-1-yl)methyl magnolol scaffold as anticancer agents for the treatment of non-small cell lung cancer via targeting autophagy », *European Journal of Medicinal Chemistry*, vol. 209, p. 112922, janv. 2021, doi: 10.1016/j.ejmech.2020.112922.
[438] F. Qian *et al.*, « Inhibition of tumor cell growth, invasion, and metastasis by EXEL-2880 (XL880, GSK1363089), a novel inhibitor of HGF and VEGF receptor tyrosine kinases », *Cancer research*, vol. 69, nº 20, p. 8009-8016, 2009.
[439] M. H. Townsend *et al.*, « Non-small-cell lung cancer cell lines A549 and NCI-H460 express hypoxanthine guanine phosphoribosyltransferase on the plasma membrane », *Onco Targets Ther*, vol. 10, p. 1921-1932, mars 2017, doi: 10.2147/OTT.S128416.
[440] D.-D. Li *et al.*, « Molecular Dynamics Analysis of Binding Sites of Epidermal Growth Factor Receptor Kinase Inhibitors », *ACS Omega*, vol. 5, nº 26, p. 16307-16314, juill. 2020, doi: 10.1021/acsomega.0c02183.
[441] Q. Zhang *et al.*, « Discovery of Triazolo-pyridazine/-pyrimidine Derivatives Bearing Aromatic (Heterocycle)-Coupled Azole Units as Class II c-Met Inhibitors », *ACS Omega*, vol. 5, nº 27, p. 16482-16490, juill. 2020, doi: 10.1021/acsomega.0c00838.
[442] N. Y. Megally Abdo, R. Milad Mohareb, et P. A. Halim, « Uses of cyclohexane-1,3-dione for the synthesis of 1,2,4-triazine derivatives as anti-proliferative agents and tyrosine kinases inhibitors », *Bioorganic Chemistry*, vol. 97, p. 103667, avr. 2020, doi: 10.1016/j.bioorg.2020.103667.
[443] G. Pasquini et G. Giaccone, « C-MET inhibitors for advanced non-small cell lung cancer », *Expert Opin Investig Drugs*, vol. 27, nº 4, p. 363-375, avr. 2018, doi: 10.1080/13543784.2018.1462336.
[444] G. Tuğcu, F. E. Ö. Bayram, et H. Sipahi, « In silico Modeling and Toxicity Profiling of a Set of Quinoline Derivatives as c-MET Inhibitors in the treatment of Human Tumors », *Turkish Journal of Pharmaceutical Sciences*, vol. 18, nº 6, p. 738, 2021.
[445] Z. Benbrahim, T. Antonia, et N. Mellas, « EGFR mutation frequency in Middle East and African non-small cell lung cancer patients: a systematic review and meta-analysis », *BMC Cancer*, vol. 18, nº 1, Art. nº 1, déc. 2018, doi: 10.1186/s12885-018-4774-y.
[446] D. A. Fennell *et al.*, « Cisplatin in the modern era: The backbone of first-line chemotherapy for non-small cell lung cancer », *Cancer Treatment Reviews*, vol. 44, p. 42-50, mars 2016, doi: 10.1016/j.ctrv.2016.01.003.
[447] B. A. Chan et B. G. Hughes, « Targeted therapy for non-small cell lung cancer: current standards and the promise of the future », *Translational lung cancer research*, vol. 4, nº 1, p. 36, 2015.
[448] A. A. Koparde, R. C. Doijad, et C. S. Magdum, « Natural products in drug discovery », in *Pharmacognosy-medicinal plants*, IntechOpen, 2019.

[449] A. G. Atanasov, S. B. Zotchev, V. M. Dirsch, et C. T. Supuran, « Natural products in drug discovery: advances and opportunities », *Nature reviews Drug discovery*, vol. 20, nº 3, p. 200-216, 2021.
[450] F. Majolo, L. K. de O. B. Delwing, D. J. Marmitt, I. C. Bustamante-Filho, et M. I. Goettert, « Medicinal plants and bioactive natural compounds for cancer treatment: Important advances for drug discovery », *Phytochemistry Letters*, vol. 31, p. 196-207, 2019.
[451] M. Huang, J.-J. Lu, et J. Ding, « Natural products in cancer therapy: Past, present and future », *Natural products and bioprospecting*, vol. 11, p. 5-13, 2021.
[452] A. Altemimi, N. Lakhssassi, A. Baharlouei, D. G. Watson, et D. A. Lightfoot, « Phytochemicals: Extraction, Isolation, and Identification of Bioactive Compounds from Plant Extracts », *Plants*, vol. 6, nº 4, Art. nº 4, déc. 2017, doi: 10.3390/plants6040042.
[453] J. Zhu *et al.*, « Review of Natural Phytochemical-Based Self-Assembled Nanostructures for Applications in Medicine », *ACS Appl. Nano Mater.*, vol. 5, nº 3, p. 3146-3169, mars 2022, doi: 10.1021/acsanm.2c00056.
[454] S. M. Jachak et A. Saklani, « Challenges and opportunities in drug discovery from plants », *Current science*, p. 1251-1257, 2007.
[455] G. Aceto, V. Persico, et A. Pescapé, « The role of Information and Communication Technologies in healthcare: taxonomies, perspectives, and challenges », *Journal of Network and Computer Applications*, vol. 107, p. 125-154, 2018.
[456] P. Garcia-Oliveira *et al.*, « Status and challenges of plant-anticancer compounds in cancer treatment », *Pharmaceuticals*, vol. 14, nº 2, p. 157, 2021.
[457] T. Khan *et al.*, « Anticancer plants: A review of the active phytochemicals, applications in animal models, and regulatory aspects », *Biomolecules*, vol. 10, nº 1, p. 47, 2019.
[458] E. Khatoon *et al.*, « Phytochemicals in cancer cell chemosensitization: Current knowledge and future perspectives », in *Seminars in Cancer Biology*, Elsevier, 2022, p. 306-339.
[459] S. B. Chakraborty et C. Hancz, « Application of phytochemicals as immunostimulant, antipathogenic and antistress agents in finfish culture », *Reviews in Aquaculture*, vol. 3, nº 3, p. 103-119, 2011.
[460] A. Naeem *et al.*, « Natural Products as Anticancer Agents: Current Status and Future Perspectives », *Molecules*, vol. 27, nº 23, Art. nº 23, janv. 2022, doi: 10.3390/molecules27238367.
[461] Md. M. Rahman *et al.*, « Natural therapeutics and nutraceuticals for lung diseases: Traditional significance, phytochemistry, and pharmacology », *Biomedicine & Pharmacotherapy*, vol. 150, p. 113041, juin 2022, doi: 10.1016/j.biopha.2022.113041.
[462] G. Bar-Sela, M. Vorobeichik, S. Drawsheh, A. Omer, V. Goldberg, et E. Muller, « The medical necessity for medicinal cannabis: prospective, observational study evaluating the treatment in cancer patients on supportive or palliative care », *Evidence-Based Complementary and Alternative Medicine*, vol. 2013, 2013.
[463] M. Birenboim *et al.*, « Multivariate classification of cannabis chemovars based on their terpene and cannabinoid profiles », *Phytochemistry*, vol. 200, p. 113215, août 2022, doi: 10.1016/j.phytochem.2022.113215.
[464] S. Choi, B. C. Huang, et C. E. Gamaldo, « Therapeutic uses of cannabis on sleep disorders and related conditions », *Journal of Clinical Neurophysiology*, vol. 37, nº 1, p. 39-49, 2020.
[465] G. Zago, M. Muller, M. Van Den Heuvel, et P. Baas, « New targeted treatments for non-small-cell lung cancer–role of nivolumab », *Biologics: targets & therapy*, vol. 10, p. 103, 2016.
[466] J. Stamos, M. X. Sliwkowski, et C. Eigenbrot, « Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor », *Journal of Biological Chemistry*, vol. 277, nº 48, p. 46265-46272, 2002.
[467] A. M. Emam *et al.*, « Quinazoline-tethered hydrazone: A versatile scaffold toward dual anti-TB and EGFR inhibition activities in NSCLC », *Archiv der Pharmazie*, vol. 354, nº 12, p. 2100281, 2021, doi: 10.1002/ardp.202100281.
[468] G. M. Shaikh, M. Murahari, S. Thakur, M. S. Kumar, et M. Yc, « Studies on ligand-based pharmacophore modeling approach in identifying potent future EGFR inhibitors », *Journal of Molecular Graphics and Modelling*, vol. 112, p. 108114, mai 2022, doi: 10.1016/j.jmgm.2021.108114.
[469] D. Kiriwan *et al.*, « Identification of tripeptides against tyrosine kinase domain of EGFR for lung cancer cell inhibition by in silico and in vitro studies », *Chemical Biology & Drug Design*, vol. n/a, nº n/a, doi: 10.1111/cbdd.14010.

[470] R. Porta *et al.*, « Brain metastases from lung cancer responding to erlotinib: the importance of EGFR mutation », *European Respiratory Journal*, vol. 37, nº 3, p. 624-631, mars 2011, doi: 10.1183/09031936.00195609.
[471] Y. Wei *et al.*, « EGFR-mediated Beclin 1 phosphorylation in autophagy suppression, tumor progression, and tumor chemoresistance », *Cell*, vol. 154, nº 6, p. 1269-1284, 2013.
[472] A. Dang, S. Dang, et B. N. Vallish, « Efficacy and Safety of EGFR Inhibitors in the Treatment of EGFRPositive NSCLC Patients: A Meta-Analysis », *Reviews on Recent Clinical Trials*, vol. 16, nº 2, p. 193-201.
[473] H. Xu, W. Tang, G. Du, et N. Kokudo, « Targeting apoptosis pathways in cancer with magnolol and honokiol, bioactive constituents of the bark of *Magnolia officinalis* », *Drug Discoveries & Therapeutics*, vol. 5, nº 5, p. 202-210, 2011, doi: 10.5582/ddt.2011.v5.5.202.
[474] Y. Fu *et al.*, « Magnolol inhibits lipopolysaccharide-induced inflammatory response by interfering with TLR4 mediated NF-κB and MAPKs signaling pathways », *Journal of Ethnopharmacology*, vol. 145, nº 1, p. 193-199, janv. 2013, doi: 10.1016/j.jep.2012.10.051.
[475] J.-L. Shen *et al.*, « Honokiol and Magnolol as Multifunctional Antioxidative Molecules for Dermatologic Disorders », *Molecules*, vol. 15, nº 9, Art. nº 9, sept. 2010, doi: 10.3390/molecules15096452.
[476] H. Hu, Z. Wang, W. Hua, Y. You, et L. Zou, « Effect of Chemical Profiling Change of Processed Magnolia officinalis on the Pharmacokinetic Profiling of Honokiol and Magnolol in Rats », *Journal of Chromatographic Science*, vol. 54, nº 7, p. 1201-1212, août 2016, doi: 10.1093/chromsci/bmw052.
[477] A. N. Jain, « Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine », *Journal of medicinal chemistry*, vol. 46, nº 4, p. 499-511, 2003.
[478] R. Spitzer et A. N. Jain, « Surflex-Dock: Docking benchmarks and real-world application », *Journal of computer-aided molecular design*, vol. 26, nº 6, p. 687-699, 2012.
[479] « Schrödinger Release 2020-3. (2020). Prime. Schrödinger, LLC. » Consulté le: 9 janvier 2022. [En ligne]. Disponible sur: https://scholar.google.com/scholar_lookup?hl=en&publication_year=2020&author=Schr%C3%B6dinger+Release+2020-3&title=Prime.
[480] M. Tayyab et D. Shahwar, « GCMS analysis of Cannabis sativa L. from four different areas of Pakistan », *Egyptian Journal of Forensic Sciences*, vol. 5, nº 3, p. 114-125, 2015.
[481] F. Bachir, M. Eddouks, M. Arahou, et M. Fekhaoui, « Origin, Early History, Cultivation, and Characteristics of the Traditional Varieties of Moroccan Cannabis sativa L. », *Cannabis and cannabinoid research*, 2021.
[482] Z. K. Punja, « Emerging diseases of Cannabis sativa and sustainable management », *Pest Management Science*, vol. 77, nº 9, p. 3857-3870, 2021, doi: 10.1002/ps.6307.
[483] L. Ramos-Guerrero, G. Montalvo, M. Cosmi, C. García-Ruiz, et F. E. Ortega-Ojeda, « Classification of Various Marijuana Varieties by Raman Microscopy and Chemometrics », *Toxics*, vol. 10, nº 3, p. 115, 2022.
[484] A. Aviz-Amador, N. Contreras-Puentes, et J. Mercado-Camargo, « Virtual screening using docking and molecular dynamics of cannabinoid analogs against CB1 and CB2 receptors », *Computational Biology and Chemistry*, vol. 95, p. 107590, déc. 2021, doi: 10.1016/j.compbiolchem.2021.107590.
[485] A. M. Tomko, E. G. Whynot, L. D. Ellis, et D. J. Dupré, « Anti-Cancer Potential of Cannabinoids, Terpenes, and Flavonoids Present in Cannabis », *Cancers*, vol. 12, nº 7, Art. nº 7, juill. 2020, doi: 10.3390/cancers12071985.
[486] T. Dyba *et al.*, « The European cancer burden in 2020: Incidence and mortality estimates for 40 countries and 25 major cancers », *Eur J Cancer*, vol. 157, p. 308-347, nov. 2021, doi: 10.1016/j.ejca.2021.07.039.
[487] M. Saha, « Medical Oncology in Cancer Treatment », in *Cancer Diagnostics and Therapeutics : Current Trends, Challenges, and Future Perspectives*, S. K. Basu, C. K. Panda, et S. Goswami, Éd., Singapore: Springer, 2022, p. 271-285. doi: 10.1007/978-981-16-4752-9_12.
[488] Y. T. Lee, Y. J. Tan, et C. E. Oon, « Molecular targeted therapy: Treating cancer with specificity », *European Journal of Pharmacology*, vol. 834, p. 188-196, sept. 2018, doi: 10.1016/j.ejphar.2018.07.034.
[489] L. Bains et T. Chawla, « Anti-EGFR Therapy in Gallbladder Cancer », in *Gallbladder Cancer: Current Treatment Options*, Springer, 2023, p. 331-349.
[490] G. Iancu *et al.*, « Tyrosine kinase inhibitors in breast cancer », *Experimental and Therapeutic Medicine*, vol. 23, nº 2, p. 1-10, 2022.

[491] Q. Pan *et al.*, « Recent Advances in Boosting EGFR Tyrosine Kinase Inhibitors-Based Cancer Therapy », *Mol. Pharmaceutics*, vol. 20, nº 2, p. 829-852, févr. 2023, doi: 10.1021/acs.molpharmaceut.2c00792.
[492] R. Mishra, H. Patel, S. Alanazi, L. Yuan, et J. T. Garrett, « HER3 signaling and targeted therapy in cancer », *Oncol Rev*, vol. 12, nº 1, p. 355, mai 2018, doi: 10.4081/oncol.2018.355.
[493] V. Subramaniyan *et al.*, « A review on epidermal growth factor receptor's role in breast and non-small cell lung cancer », *Chemico-Biological Interactions*, vol. 351, p. 109735, janv. 2022, doi: 10.1016/j.cbi.2021.109735.
[494] J.-K. Dong *et al.*, « Overcoming erlotinib resistance in EGFR mutation-positive lung adenocarcinomas through repression of phosphoglycerate dehydrogenase », *Theranostics*, vol. 8, nº 7, p. 1808-1823, 2018, doi: 10.7150/thno.23177.
[495] W. S. Ramadan *et al.*, « Interplay between Epigenetics, Expression of Estrogen Receptor- α, HER2/ERBB2 and Sensitivity of Triple Negative Breast Cancer Cells to Hormonal Therapy », *Cancers*, vol. 11, nº 1, Art. nº 1, janv. 2019, doi: 10.3390/cancers11010013.
[496] H. G. Russnes, O. C. Lingjærde, A.-L. Børresen-Dale, et C. Caldas, « Breast Cancer Molecular Stratification: From Intrinsic Subtypes to Integrative Clusters », *Am J Pathol*, vol. 187, nº 10, p. 2152-2162, oct. 2017, doi: 10.1016/j.ajpath.2017.04.022.
[497] T. Lamtha, L. Tabtimmai, N. Songtawee, N. Tansakul, et K. Choowongkomon, « Structural analysis of cannabinoids against EGFR-TK leads a novel target against EGFR-driven cell lines », *Curr Res Pharmacol Drug Discov*, vol. 3, p. 100132, sept. 2022, doi: 10.1016/j.crphar.2022.100132.
[498] P. Auberger, J. Tamburini-Bonnefoy, et A. Puissant, « Drug Resistance in Hematological Malignancies », *International Journal of Molecular Sciences*, vol. 21, nº 17, Art. nº 17, janv. 2020, doi: 10.3390/ijms21176091.
[499] R. C. Khanna, M. V. Cicinelli, S. S. Gilbert, S. G. Honavar, et G. V. Murthy, « COVID-19 pandemic: Lessons learned and future directions », *Indian journal of ophthalmology*, vol. 68, nº 5, p. 703, 2020.
[500] S. Pernas et S. M. Tolaney, « Targeting HER2 heterogeneity in early-stage breast cancer », *Current Opinion in Oncology*, vol. 32, nº 6, p. 545, nov. 2020, doi: 10.1097/CCO.0000000000000685.
[501] U. Mehraj *et al.*, « The tumor microenvironment as driver of stemness and therapeutic resistance in breast cancer: New challenges and therapeutic opportunities », *Cell Oncol.*, vol. 44, nº 6, p. 1209-1229, déc. 2021, doi: 10.1007/s13402-021-00634-9.
[502] C. Pissot Soldermann *et al.*, « Discovery of potent, highly selective, and in vivo efficacious, allosteric MALT1 inhibitors by iterative scaffold morphing », *Journal of medicinal chemistry*, vol. 63, nº 23, p. 14576-14593, 2020.
[503] T. Lu *et al.*, « Discovery and optimization of a series of small-molecule allosteric inhibitors of MALT1 protease », *Bioorganic & Medicinal Chemistry Letters*, vol. 29, nº 23, p. 126743, déc. 2019, doi: 10.1016/j.bmcl.2019.126743.
[504] F. Schlauderer *et al.*, « Structural Analysis of Phenothiazine Derivatives as Allosteric Inhibitors of the MALT1 Paracaspase », *Angewandte Chemie International Edition*, vol. 52, nº 39, p. 10384-10387, 2013, doi: 10.1002/anie.201304290.
[505] T. Wu *et al.*, « Original hosts, clinical features, transmission routes, and vaccine development for coronavirus disease (COVID-19) », *Frontiers in Medicine*, vol. 8, 2021.
[506] P. Cihan, « Forecasting fully vaccinated people against COVID-19 and examining future vaccination rate for herd immunity in the US, Asia, Europe, Africa, South America, and the World », *Applied Soft Computing*, vol. 111, p. 107708, nov. 2021, doi: 10.1016/j.asoc.2021.107708.
[507] « SARS-CoV-2 Evolution ». Consulté le: 8 avril 2022. [En ligne]. Disponible sur: https://www.who.int/news-room/questions-and-answers/item/sars-cov-2-evolution
[508] C. Liu *et al.*, « Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum », *Cell*, vol. 184, nº 16, p. 4220-4236.e13, août 2021, doi: 10.1016/j.cell.2021.06.020.
[509] V. Kumar, S. Kar, P. De, K. Roy, et J. Leszczynski, « Identification of potential antivirals against 3CLpro enzyme for the treatment of SARS-CoV-2: A multi-step virtual screening study », *SAR and QSAR in Environmental Research*, vol. 0, nº 0, p. 1-30, avr. 2022, doi: 10.1080/1062936X.2022.2055140.
[510] S. A. Amin, S. Banerjee, K. Ghosh, S. Gayen, et T. Jha, « Protease targeted COVID-19 drug discovery and its challenges: Insight into viral main protease (Mpro) and papain-like protease (PLpro) inhibitors », *Bioorganic & medicinal chemistry*, vol. 29, p. 115860, 2021.

[511] J. F.-W. Chan *et al.*, « Improved molecular diagnosis of COVID-19 by the novel, highly sensitive and specific COVID-19-RdRp/Hel real-time reverse transcription-PCR assay validated in vitro and with clinical specimens », *Journal of clinical microbiology*, vol. 58, nº 5, p. e00310-20, 2020.
[512] L. Wang et Y. Xiang, « Spike glycoprotein-mediated entry of SARS coronaviruses », *Viruses*, vol. 12, nº 11, p. 1289, 2020.
[513] J. Jankun, « COVID-19 pandemic; transmembrane protease serine 2 (TMPRSS2) inhibitors as potential drugs. », *Translation: The University of Toledo Journal of Medical Sciences*, vol. 7, p. 1-5, 2020.
[514] W. Ni *et al.*, « Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19 », *Critical Care*, vol. 24, nº 1, p. 1-10, 2020.
[515] G. P. Rossi, V. Sanga, et M. Barton, « Potential harmful effects of discontinuing ACE-inhibitors and ARBs in COVID-19 patients », *Elife*, vol. 9, p. e57278, 2020.
[516] S. A. Khan, K. Zia, S. Ashraf, R. Uddin, et Z. Ul-Haq, « Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach », *Journal of Biomolecular Structure and Dynamics*, vol. 39, nº 7, p. 2607-2616, mai 2021, doi: 10.1080/07391102.2020.1751298.
[517] M. G. Deshmukh *et al.*, « Structure-guided design of a perampanel-derived pharmacophore targeting the SARS-CoV-2 main protease », *Structure*, vol. 29, nº 8, p. 823-833.e5, août 2021, doi: 10.1016/j.str.2021.06.002.
[518] S. Liu, Q. Zheng, et Z. Wang, « Potential covalent drugs targeting the main protease of the SARS-CoV-2 coronavirus », *Bioinformatics*, vol. 36, nº 11, p. 3295-3298, 2020.
[519] P. Delre, F. Caporuscio, M. Saviano, et G. F. Mangiatordi, « Repurposing known drugs as covalent and non-covalent inhibitors of the SARS-CoV-2 papain-like protease », *Frontiers in Chemistry*, vol. 8, p. 1032, 2020.
[520] J. Osipiuk *et al.*, « Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors », *Nature communications*, vol. 12, nº 1, p. 1-9, 2021.
[521] B. Halford, « Pfizer unveils its oral SARS-CoV-2 inhibitor », *Chemical & Engineering News*, p. 7-7, 2021.
[522] B. Boras *et al.*, « Discovery of a Novel Inhibitor of Coronavirus 3CL Protease for the Potential Treatment of COVID-19 ». bioRxiv, p. 2020.09.12.293498, 12 février 2021. doi: 10.1101/2020.09.12.293498.
[523] K. Sanachai *et al.*, « Identification of repurposing therapeutics toward SARS-CoV-2 main protease by virtual screening », *PLOS ONE*, vol. 17, nº 6, p. e0269563, juin 2022, doi: 10.1371/journal.pone.0269563.
[524] Z. Jin *et al.*, « Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors », *Nature*, vol. 582, nº 7811, Art. nº 7811, juin 2020, doi: 10.1038/s41586-020-2223-y.
[525] L. Zhang *et al.*, « Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors », *Science*, vol. 368, nº 6489, p. 409-412, avr. 2020, doi: 10.1126/science.abb3405.
[526] J. Qiao *et al.*, « SARS-CoV-2 Mpro inhibitors with antiviral activity in a transgenic mouse model », *Science*, vol. 371, nº 6536, p. 1374-1378, mars 2021, doi: 10.1126/science.abf1611.
[527] C. Ma *et al.*, « Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease », *Cell Res*, vol. 30, nº 8, Art. nº 8, août 2020, doi: 10.1038/s41422-020-0356-z.
[528] Y. Xiong *et al.*, « Discovery of naturally occurring inhibitors against SARS-CoV-2 3CLpro from Ginkgo biloba leaves via large-scale screening », *Fitoterapia*, vol. 152, p. 104909, juill. 2021, doi: 10.1016/j.fitote.2021.104909.
[529] N. Kitamura *et al.*, « Expedited Approach toward the Rational Design of Noncovalent SARS-CoV-2 Main Protease Inhibitors », *J. Med. Chem.*, vol. 65, nº 4, p. 2848-2865, févr. 2022, doi: 10.1021/acs.jmedchem.1c00509.
[530] S. H. Han *et al.*, « Structure-Based Optimization of ML300-Derived, Noncovalent Inhibitors Targeting the Severe Acute Respiratory Syndrome Coronavirus 3CL Protease (SARS-CoV-2 3CLpro) », *J. Med. Chem.*, vol. 65, nº 4, p. 2880-2904, févr. 2022, doi: 10.1021/acs.jmedchem.1c00598.
[531] C.-H. Zhang *et al.*, « Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations », *ACS Cent. Sci.*, vol. 7, nº 3, p. 467-475, mars 2021, doi: 10.1021/acscentsci.1c00039.
[532] P. A. Nogara, F. B. Omage, G. R. Bolzan, C. P. Delgado, L. Orian, et J. B. T. Rocha, « Reactivity and binding mode of disulfiram, its metabolites, and derivatives in SARS-CoV-2 PLpro: insights

from computational chemistry studies », *J Mol Model*, vol. 28, nº 11, p. 354, oct. 2022, doi: 10.1007/s00894-022-05341-2.
[533] E. De Vita, « 10 years into the resurgence of covalent drugs », *Future Medicinal Chemistry*, vol. 13, nº 2, p. 193-210, janv. 2021, doi: 10.4155/fmc-2020-0236.
[534] M. Erak, K. Bellmann-Sickert, S. Els-Heindl, et A. G. Beck-Sickinger, « Peptide chemistry toolbox – Transforming natural peptides into peptide therapeutics », *Bioorganic & Medicinal Chemistry*, vol. 26, nº 10, p. 2759-2765, juin 2018, doi: 10.1016/j.bmc.2018.01.012.
[535] J.-W. Zhang *et al.*, « Discovery of 9,10-dihydrophenanthrene derivatives as SARS-CoV-2 3CLpro inhibitors for treating COVID-19 », *European Journal of Medicinal Chemistry*, vol. 228, p. 114030, janv. 2022, doi: 10.1016/j.ejmech.2021.114030.
[536] M. A. Alamri *et al.*, « Pharmacoinformatics and molecular dynamics simulation studies reveal potential covalent and FDA-approved inhibitors of SARS-CoV-2 main protease 3CLpro », *Journal of Biomolecular Structure and Dynamics*, vol. 39, nº 13, p. 4936-4948, sept. 2021, doi: 10.1080/07391102.2020.1782768.
[537] S. A. L. PhD MD, « DISulfiram for COvid-19 (DISCO) Trial: A Phase 2 Double-Blind, Randomized Placebo-Controlled Trial of Disulfiram Compared to Standard Care in Patients With Symptomatic COVID-19 », clinicaltrials.gov, Clinical trial registration NCT04485130, mai 2021. Consulté le: 7 avril 2022. [En ligne]. Disponible sur: https://clinicaltrials.gov/ct2/show/NCT04485130
[538] H. I. Umar *et al.*, « Novel Molecules derived from 3-O-(6-galloylglucoside) inhibit Main Protease of SARS-CoV 2 In Silico », *Chem. Pap.*, vol. 76, nº 2, p. 785-796, févr. 2022, doi: 10.1007/s11696-021-01899-y.
[539] A. Belhassan *et al.*, « In silico detection of potential inhibitors from vitamins and their derivatives compounds against SARS-CoV-2 main protease by using molecular docking, molecular dynamic simulation and ADMET profiling », *Journal of Molecular Structure*, vol. 1258, p. 132652, juin 2022, doi: 10.1016/j.molstruc.2022.132652.
[540] S. Chtita *et al.*, « In silico investigation of phytoconstituents from Cameroonian medicinal plants towards COVID-19 treatment », *Struct Chem*, avr. 2022, doi: 10.1007/s11224-022-01939-7.