Université Sidi Mohamed Ben Abdellah
Faculté des Sciences Dhar El Mahraz-Fès
Centre d'Etudes Doctorales
"Sciences et Technologies"

Formation Doctorale : Ressources Naturelles, Environnement et Développement Durable

Discipline : Chimie

Spécialité : Chimie-physique Appliquée

Laboratoire : Ingénierie des Matériaux, de Modélisation et d'Environnement

THESE DE DOCTORAT

Présentée par

HADNI HANINE

Modélisation moléculaire de séries d'hybrides hétérocycliques comme agents antipaludiques, par les méthodes QSAR et docking moléculaire avec le récepteur *PF-DHFR* dans ses deux formes sauvage et mutante

Soutenue le 04 /05 / 2019 devant le jury composé de :

Pr. Hamid Toufik	**Faculté Polydisciplinaire - Taza**	**Président**
Pr. Mohamed Zahouily	**FSTM-Université HassanII - Mohammadia**	**Rapporteur**
Pr. Mohammed Fahim	**Faculté des Sciences - Meknès**	**Rapporteur**
Pr. Adil Touimi Benjelloun	**Faculté des Sciences Dhar El Mahraz - Fès**	**Rapporteur**
Pr. Fatima Lamchouri	**Faculté Polydisciplinaire - Taza**	**Examinateur**
Pr. Mouhcine Sfaira	**Faculté des Sciences Dhar El Mahraz – Fès**	**Examinateur**
Pr. Mohammed Benzakour	**Faculté des Sciences Dhar El Mahraz – Fès**	**Examinateur**
Pr. Menana Elhallaoui	**Faculté des Sciences Dhar El Mahraz – Fès**	**Directeur de thèse**

Année universitaire : 2018-2019

Résumé

Les travaux de recherche présentés dans cette thèse portent sur l'exploitation de différentes techniques de modélisation moléculaire de molécules hybrides agissant comme agents antipaludiques, via les méthodes QSAR et le Docking moléculaire. L'objectif principal est de contribuer à mettre au point de nouveaux médicaments antipaludiques, capables d'agir comme inhibiteurs du *dihydrofolate réductase* (DHFR) récepteur responsable du paludisme. Ainsi, trois séries de molécules hybrides à base de quinoline-triazine testées pour leurs activités antipaludiques vis-à-vis du récepteur DHFR sous ses deux formes, sauvage et mutante, ont été étudiées. Enfaite, cette étude est une contribution à la découverte de nouvelles molécules capables de surmonter la résistance du paludisme constituant un problème majeur face à la santé humaine, via la détermination des caractéristiques structurales de molécules hybrides cibles. Pour cela, un certain nombre de descripteurs moléculaires multidimensionnels (1D, 2D, 3D) ont été choisis pour identifier les régions de l'espace qui sont corrélées avec l'activité biologique. Pour le calcul de ces descripteurs nous avons utilisé la méthode DFT avec la fonctionnelle hybride à trois paramètres de Becke (Lee-Yang-Paar) (B3LYP) et la base 6-31G dans le cas des descripteurs électroniques, et la mécanique moléculaire MM2 pour les autres descripteurs. Les modèles QSAR ont été générés à l'aide des méthodes statistiques, à savoir : L'analyse en Régression Linéaire Multiple (MLR) et les Réseaux de Neurones Artificiels (ANN). La fiabilité et la capacité des modèles établis ont été testés par différentes techniques de validation, telles que: La validation interne, la validation externe et la méthode de Y-randomisation. La méthode du Docking moléculaire a permis d'une part de mettre en évidence les caractéristiques structurales des molécules étudiées, d'autre part, de visualiser et de comprendre les modes des interactions des pharmacophores que comportent les dérivés de la quinoline-triazine avec la protéine DHFR. Les résultats obtenus dans cette étude, pourraient être efficacement utilisées dans des projets de recherche d'avenir dédiés à la conception de nouvelles molécules ayant une activité antipaludique.

Mots clés: Quinoline-triazine, activité antipaludique, *Dihydrofolate Réductase* (*DHFR*), QSAR, Docking moléculaire, DFT.

ملخص

يتناول البحث المقدم في هذه الرسالة استغلال تقنيات النمذجة الجزيئية المختلفة للجزيئات المهجنة التي تعمل كعوامل مضادة للملاريا ، عبر طرق QSAR و الالتحام الجزيئي. الهدف الرئيسي هو المساهمة في تطوير عقاقير جديدة مضادة للملاريا ، قادرة على العمل كمثبطات لمستقبلات إنزيم ثنائي هيدروفولات (DHFR) المسؤول عن الملاريا. وهكذا ، تمت دراسة ثلاث سلاسل من جزيئات الكينولين-تريازين الهجينة التي تم اختبارها من خلال نشاطها المضاد للملاريا ضد مستقبلات DHFR في كل من الأشكال البرية والمتحولة. في الواقع ، هذا البحث هو اتجاه نحو اكتشاف جزيئات جديدة قادرة على التغلب على مقاومة الملاريا، والتي تعد مشكلة كبيرة في صحة الإنسان ، من خلال تحديد الخصائص الهيكلية للجزيئات الهجينة المستهدفة. لهذا الغرض ، لقد تم اختيار عدد من الواصفات الجزيئية متعددة الأبعاد (3D, 2D,1D) لتحديد مناطق الفضاء المرتبطة بالنشاط البيولوجي. من أجل حساب هذه الواصفات ، استخدمنا طريقة DFT مع الوظيفة الهجينةBecke ذات الثلاثة معلمات B3LYP (Lee-Yang-Paar) مع قاعدة 6-31G في حالة الواصفات الإلكترونية، والميكانيكا الجزيئية MM2 من أجل الواصفات الأخرى. لقد تم إنشاء نماذج QSAR باستخدام الأساليب الإحصائية،c وهي: تحليل الانحدار الخطي المتعدد (MLR) والشبكات العصبية الاصطناعية (ANN). تم اختبار موثوقية وقدرة النماذج المنشأة من خلال أساليب التحقق المختلفة ، مثل: التحقق الداخلي ، التحقق الخارجي وطريقة التوزيع العشوائي. تتضمن طريقة الإرساء الجزيئي من ناحية تسليط الضوء على الخصائص الهيكلية للجزيئات التي تمت دراستها ، من ناحية أخرى ، لتصور وفهم أنماط تفاعلات المستحضرات الصيدلانية التي تحتويها مشتقات الكينولين- تريازين مع البروتين DHFR. النتائج التي تم الحصول عليها في هذه الدراسة، يمكن استخدامها بفعالية في مشاريع البحوث المستقبلية المكرسة لتصميم جزيئات جديدة ذات نشاط مضاد للملاريا.

الكلمات المفتاحية: الكينولين- تريازين ، نشاط مضاد للملاريا ، اختزال ثنائي هيدروفولات (DHFR)، QSAR ، الالتحام الجزيئي،DFT.

Abstract

The research presented in this thesis deals with the exploitation of different molecular modeling techniques of hybrid molecules acting as antimalarial agents, via QSAR methods and molecular Docking. The main objective is to contribute to the development of new antimalarial drugs, able of acting as inhibitors of the dihydrofolate reductase (DHFR) receptor responsible of malaria. Thus, three series of quinoline-triazine hybrid molecules tested by their antimalarial activity against the DHFR receptor in its both wild and mutant forms were studied. In fact, this study is an contribution towards the discovery of new molecules able to overcome the resistance of malaria, which is a major problem in human health, by determining the structural characteristics of the target hybrid molecules. For this, a number of multidimensional molecular descriptors (1D, 2D, 3D) have been chosen to identify the regions of space that are correlated with biological activity. For the calculation of these descriptors we used the Density Functional Theory (DFT) method with the Becke three-parameter hybrid functional (Lee-Yang-Paar) (B3LYP) with the 6-31G basis in the case of the electronic descriptors, and the molecular mechanics MM2 for the other descriptors. The QSAR models were generated using statistical methods: Multiple Linear Regression (MLR) analysis and Artificial Neural Networks (ANN). The reliability and the capacity of the established models were tested by different validation techniques, such as: Internal validation, external validation and Y-randomization methods. The Molecular Docking method has first, highlight the structural characteristics of the molecules studied, secondly, has helped us to visualize and understand the modes of pharmacophore interactions that quinoline-triazine derivatives established with the protein DHFR. The results obtained in this study, could be effectively used in future research projects dedicated to the design of new molecules with antimalarial activity.

Key Words: Quinoline-triazine, antimalarial activity, *Dihydrofolate Reductase* (DHFR), QSAR, Molecular docking, DFT.

Dédicace

A ma chère mère

J'aimerai bien exprimer mon immense amour et ma profonde gratitude à ma mère qui n'a guère cessé de m'encourager et de m'apporter toute l'assurance et l'épaulement tout au long de mon parcours, que Dieu la protège et l'honore. Fort précieuse qu'elle est, je ne saurais jamais exprimer la gratitude que je lui dois.

A la mémoire de mon père et de ma grand-mère, auxquels je ne cesse d'y penser.

A mon frère Mohamed et mes sœurs Soumia , houria et sa famille, votre affection et vos encouragements ont toujours été pour moi des plus précieux. Que ce travail soit pour vous le gage de mon profond amour.

J'espère avoir répondu aux espoirs que vous avez fondés en moi.

A mes ami (e)s et à mes collègues qui n'ont pas manqué de m'encourager, qu'ils trouvent ici le témoignage de ma profonde gratitude

.

REMERCIEMENTS

« La reconnaissance est la mémoire du cœur »

(Jean-Baptiste Massieu, 1742-1822)

Ce travail de thèse a été réalisé au sein du Laboratoire d'Ingénierie des Matériaux, de Modélisation et d'Environnement (LIMME) de la Faculté des Sciences Dhar El Mahraz-Fès, Université Sidi Mohamed Ben Abdallah USMBA, sous la direction du Professeur **Menana ELHALLAOUI.**

En premier lieu, je tiens à exprimer ma profonde reconnaissance et gratitude au Professeur **Menana ELHALLAOUI** pour m'avoir accueilli au sein de son équipe de recherche et pour m'avoir permis de mener ce travail dans d'excellentes conditions. J'adresse à elle mes sincères remerciements pour son soutien sans faille, ses directives avisées ainsi que sa patience pour avoir suivi avec attention les travaux de cette thèse jusqu'à son aboutissement.

Je tiens à exprimer ma respectueuse gratitude à Monsieur **Hamid TOUFIk** professeur à la Faculté Polydisciplinaire - Taza, qui m'a fait l'honneur d'accepter de présider le jury de cette thèse.

Mes sincère remerciements vont également à Monsieur **Mohamed ZAHOUILY** professeur à la Faculté des Sciences et techniques-Mohammadia, Université Hassan II, d'avoir accepté d'examiner mon travail en tant que rapporteur.

J'adresse mes remerciements à Monsieur **Mohammed Fahim** Professeur à la Faculté des Sciences-Meknès, qui s'est manifesté favorablement pour examiner mon travail de thèse en tant que rapporteur.

Je tiens à remercier Monsieur **Adil TOUIMI BENJELLOUN** Professeur à la Faculté des Sciences Dhar El Mahraz-Fès, Université Sidi Mohamed Ben Abdellah et directeur du Laboratoire d'Ingénierie des Matériaux, Modélisation et Environnement (LIMME), pour l'honneur qu'il me fait en acceptant d'être rapporteur de ma thèse.

Je tiens également à adresser mes remerciements à Madame **Fatima LAMCHOURI** Professeur à la Faculté Polydisciplinaire de Taza, Université Sidi Mohamed Ben Abdellah, pour avoir accepté d'examiner mon travail.

Mes remerciements vont également à Monsieur **Mohammed BENZAKOUR** Professeur à la Faculté des Sciences Dhar El Mahraz de Fès, Université Sidi Mohamed Ben Abdellah, pour l'honneur qu'il me fait en acceptant d'être membre du jury de ma thèse.

Mes plus vifs remerciements s'adressent à Monsieur **Mouhcine SFAIRA** Professeur à la Faculté des sciences Dhar El Mahraz de Fès, Université Sidi Mohamed Ben Abdellah, pour avoir voulu faire partie du jury de ma thèse, qu'il trouve ici l'expression de ma profonde gratitude.

Mes cordiaux remerciements s'adressent également à tout le corps professoral et administratif de la Faculté des sciences Dhar El Mahraz-Fès et particulièrement les Professeurs du département de chimie. Qu'ils trouvent ici l'expression de ma reconnaissance.

Je ne saurais oublier de remercier mes collègues du Laboratoire LIMME, je nomme en particulier : Khalid DGUIGUI, Nidal NACEIRI MRABTI, Fatima zohra TASSINE, Mohamed Mazigh, El'mbarki Charif. Pour toutes les discussions scientifiques ainsi que pour tous les moments agréables qu'on a passé ensemble.

Finalement, je remercie du fond de cœur mes parents, mes frères, sœurs et toute ma famille pour le constant et précieux soutien moral et matériel qu'ils m'ont apporté pendant mes études. Un grand merci à tous ceux qui m'ont aidé, renseigné ou conseillé.

Table de matières

Introduction générale **1**

Chapitre I: Généralités sur le paludisme **5**

I Historique du paludisme 5

II Introduction 5

III Aspect biologique du paludisme 7

III.1 Cycle chez l'Homme 8

III.1.1 Phase hépatique ou exo-érythrocytaire 8

III.1.2 Cycle intra-érythrocytaire 8

III.2 Cycle chez l'anophèle 9

IV Composés antipaludiques 9

IV.1 Dérivés des aurones 10

IV.2 Dérivés des quinolines 10

IV.2.1 Quinine 10

IV.2.2 Dérivés de la 4-Aminoquinoline 10

IV.2.3 Dérivés de la 8-aminoquinoline 11

IV.2.5 Mode d'action des antipaludiques quinolines 12

IV.2.6 Résistance à la chloroquine 12

IV.3 Triazines et composés similaires 13

IV.3.1 Mode d'action des composés antifolates 14

IV.3.2 Résistance aux antifolates 15

V Intégration des molécules hybrides dans la chimiothérapie antipaludique à base de Chloroquine (dérivées de la quinoline) et cycloguanil (dérivées de la triazine). 15

VI Présentation de la PfDHFR-TS 17

Chapitre II: Techniques de modélisation moléculaire **29**

I. Introduction 29

II. Méthodes théoriques de calcul 29

II.1 Mécanique moléculaire 29

II.1.1 Champ de force 29

II.1.1.1 Exemples de champs de force 30

II.1.2 Énergie d'interaction entre atomes liés 32

II.1.2.1 Energie d'élongation 32

II.1.2.2 Energie de flexion 32

II.1.2.3 Energie de torsion 32

II.1.3 Énergie d'interaction entre atomes non liés 33
II.1.3.1 Energie de Van der Waals 33
II.1.3.2 Liaisons hydrogènes 34
II.1.3.3 Interactions électrostatiques 34
II.2 Théorie de la Fonctionnelle de la Densité (DFT) 34
II.2.1 introduction 34
II.2.2 Approximation de Kohn-Sham 37
II.2.3 Fonctionnelle de densité de première génération (LDA) 38
II.2.3.1 Approximation de la densité de spin locale (LSDA) 38
II.2.3.2 Approximation du gradient généralisée (GGA) 39
II.3 Descripteurs moléculaires 41
II.3.1 Descripteurs constitutionnels 41
II.3.2 Descripteurs géométriques 42
II.3.3 Descripteurs topologiques 43
II.3.3.1 Indice de Vienne, indice de Balaban 43
II.3.4 Descripteurs physico-chimiques 44
II.3.5 Descripteurs quantiques 44
III. Relations quantitatives structures activités/propriétés 46
III.1 Introduction 46
III.2 Principe du QSAR 46
IV. Méthodes d'analyse statistiques 47
VI.1 Régression linéaire multiple 47
VI.1.2 Coefficients et tests statistiques de la régression 48
VI.1.2.1 Coefficient de détermination (R^2): 48
VI.1.2.2 Coefficient de corrélation (R) 49
VI.1.2.3 Erreur quadratique moyenne « RMSE » et Erreur de l'écart-type « SD » 49
VI.1.2.4 Test d'hypothèses 49
VI.1.2.4.1 Hypothèse nulle et hypothèse alternative 50
VI.1.2.5 Test de Student 50
VI.1.2.6 Test de Fisher-Snedecor (F) 51
VI.2 Réseaux de Neurones Artificiels 52
VI.2.1 Structure des réseaux de neurones 53
a) Perceptron simple 53
b) Perceptron multicouche 53
VI.2.2 Types de réseaux de neurones 54
VI.2.3 Fonction d'activation 55

VI.2.4 Fonction de perte 55
VI.3 Validation des modèles 56
VI.3.1 Coefficients et tests statistiques standards 56
VI.3.2 Validation interne 56
VI.3.2.1 Procédure leave-one-out 57
V.3.3 Y-Randomization 58
VI.3.4 Validation externe 58
VI.4 Domaine d'applicabilité 59
V. Docking moléculaire 60
V.1 Généralités sur le Docking Moléculaire 60
V.1.1 Principe du docking moléculaire 61
V.1.2 Algorithme du docking 61
V.1.3 Fonction de score 62
V.1.4 Choix du logiciel AUTODOCK 63
V.2 AUTODOCK 64
V.3 Avantages et limites du docking moléculaire 65
V.3.1 Avantages du docking moléculaire 65
V.3.2 Limites du docking moléculaire 66
Références 67
Chapitre III: Molecular Modeling of Antimalarial Agents by 3D-QSAR Study and Molecular Docking of Two Hybrids 4-Aminoquinoline-1,3,5-triazine and 4-Aminoquinoline-oxalamide Derivatives with the Receptor Protein in Its Both Wild and Mutant Types 78
I Introduction 78
II. Materials and Methods 80
II.1 Experimental Data 80
II.2 Molecular Descriptors Calculation 86
II.3 Analysis Methods 86
III. Results and Discussion 88
III.1 Multiple Linear Regression 89
III.2 Artificial Neural Networks 90
III.3 Cross validation 91
III.4 Y-randomization 92
III.5 External Validation 93
III.6 Docking studies 94
IV. CONCLUSION 97
References 99

CHapitre IV : QSAR and Molecular docking studies of 4-anilinoquinoline-triazine hybrids as *pf-DHFR* inhibitors 102

I. Introduction 102

II. Materials and Methods 103

II.1 Experimental Data 103

II.2 Molecular descriptors calculation 105

II.3 Analysis Methods 106

II.3.1 Multiple Linear Regression 106

II.3.2 Artificial Neural Networks 106

II.3.3 Cross-validation 107

II.3.4 Y-randomization test 107

II.3.5 External validation 107

II.3.6 Molecular Docking 108

III. Results and Discussion 108

III.1 Multiple Linear Regression (MLR) 109

III.2 Artificial Neural networks (ANN) 111

III.3 Cross-validation (CV) 111

III.4 Y-randomization 111

III.5 External Validation 112

III.6 Docking studies 113

IV. Conclusion 115

References 117

Chapitre V : Molecular docking and QSAR studies for modeling the antimalarial activity of 4-anilinoquinoline-triazines hybrids Derivatives with the wild-type and mutant Receptor pf-DHFR 120

I. Introduction 120

II. Materials and methods 121

II.1 Studied molecules 121

II.2 Molecular descriptors calculation 127

II.3 Statistical methods 128

II.4 Molecular Docking modeling 128

III. Results and Discussion 129

III.1 Multiple Linear Regression 130

III.2 Artificial Neural networks 132

III.3 Cross Validation 132

III.4 Y-Randomization 133

III.5 External Validation 134
III.6 Docking Studies 135
IV. Conclusion 137
References 138
Conclusion générale 142

Liste de figures

Figure 1 : Structure chimiques des 4-aminoquinoline-oxalamides **(a)**, 4-aminoquinoline–triazines **(b)** et 4-anilinoquinoline-triazines **(c)**. 2
Figure 2 : Répartition géographique du paludisme dans le monde. 6
Figure 3 : Cycle de vie des parasites plasmodiums responsables du paludisme [18]. 7
Figure 4 : Structures des aurones et des aza-aurones. 10
Figure 5 : Structures chimiques de la chloroquine, l'amodiaquine, la pamaquine, la mepacrine, la quinine, et de la primaquine. 12
Figure 6 : Structures chimiques de la cycloguanil, de la pyriméthamine, de la proguanil et de la sulfadoxine. 14
Figure 7 : Structures de l'acide folique et de l'acide folinique 15
Figure 8 : Structures de (a) chloroquine, (b) cycloguanil et (c) hybrides dérivés du 4-aminoquinoline-1,3,5-triazine. 17
Figure 9 : Structure du PfDHFR-TS type sauvage (en vert et jaune) couplé avec WR99210 (en rouge), NADPH (en bleu) et dUMP (en orange). Les régions de jonction et les domaines TS sont en violet et bleu ciel. 18
Figure 10 : Structure chimique du WR99210 19
Figure 11 : Schéma de principe d'un perceptron simple 53
Figure 12: Architecture d'un réseau de neurones artificiel. 54
Figure 13: Différents types de fonction de transfert pour le neurone artificiel. 55
Figure 14: Mesure pour l'évaluation du docking moléculaire. 60
Figure 15: Procédures du docking avec AutoDock 64

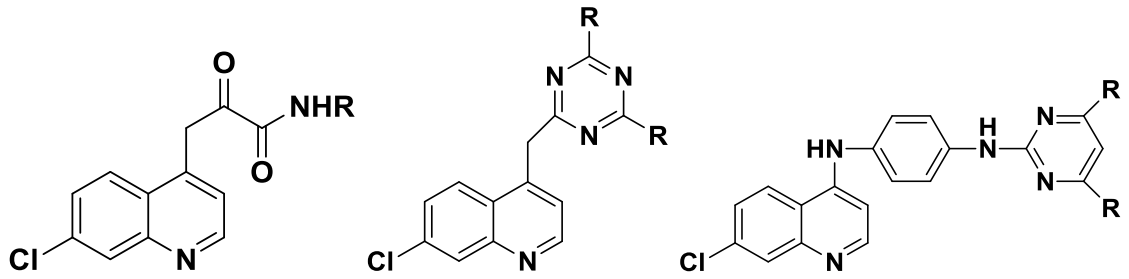
Liste des abréviations

ANN	Artificial neural networks
B3LYP	Becke's exchange with three-parameters of the Lee-Yang-Parr's correlation functional
CV	Cross validation
DFT	Density functional theory
DHFR	*Dihydrofolate reductase*
DHPS	*Dihydropteroate synthase*
G6PD	*Glucose-6-phosphate dehydrogenase*
GGA	Generalized gradient approximation
Hb	Hemoglobin
HOMO	Highest occupied molecular orbital energy
IC_{50}	Median inhibitory concentration
LCO	The “leave-many-out” cross validation method
LDA	Local density approximation
LOO	The “ leave-one-out ” cross validation method
LSDA	Local spin density approximation
LUMO	Lowest unoccupied molecular orbital energy
MD	Molecular dynamics
MLR	Multiple linear regression
MM	Molecular mechanics
Pfcrt	*Plasmodium falciparum chloroquine resistance transporter*
PF-DHFR- TS	*Plasmodium falciparum dihydrofolate reductase–thymidylate synthase*
Pfmdr1	*Plasmodium falciparum multidrug resistance 1*
QSAR	Quantitative structure–activity relationship
VDW	Van Der Waals
WHO	The world health organization

Introduction générale

De nos jours, le paludisme reste l'une des principales causes de morbidité et de mortalité dans le monde [1,2]. Il est causé par des parasites protozoaires du genre Plasmodium qui s'exprime sous plusieurs types, dont le *Plasmodium falciparum (P. falciparum)* est le plus mortel [2]. Les succès dans la lutte pour l'éradication du paludisme ont été de courte durée à cause de l'augmentation de la résistance des parasites aux traitements antipaludiques. En plus, l'émergence de la résistance aux médicaments antipaludiques dans le parasite du paludisme augmente considérablement le taux de mortalité [3,4], d'où la nécessité de mettre au point de nouveaux agents thérapeutiques capables de lutter contre la maladie et de surmonter la résistance.

Le développement de la résistance aux médicaments antipaludiques constitue un problème majeur face à la santé humaine. Cela nécessite une concentration des efforts pour la recherche de nouveaux agents antipaludiques. La conception d'une nouvelle génération de molécules hybrides pourrait remédier à ce problème [5]. Dans ce contexte, plusieurs équipes de recherche ont montré beaucoup d'intérêts aux composés hétérocycliques hybrides qui manifestent une activité biologique antipaludique. Dans cette thèse, nous nous somme intéressées à trois types d'hybrides: 4-anilinoquinoline-triazines, 4-aminoquinoline-triazines, et 4-aminoquinoline-oxalamides (**Figure 1**) qui constituent une classe de molécules hybrides approuvant une efficacité quant à la lutte contre la résistance des parasites au sein de la protéine *Plasmodium falciparum dihydrofolate réductase* (pf-DHFR) avec ses deux formes mutante et sauvage [6–8]. Nous avons procédé à la modélisation de ces hybrides et exploré différentes pistes dans le but de comprendre leur comportement structural, en essayant de trouver une explication de l'activité et de la non-activité de certaines molécules.



(a) **(b)** **(c)**

Figure 1 : Structure chimiques des 4-aminoquinoline-oxalamides **(a)**, 4-aminoquinoline–triazines **(b)** et 4-anilinoquinoline-triazines **(c)**.

Les travaux de cette thèse portent ainsi sur l'exploitation des techniques de la modélisation moléculaire afin de contribuer à la recherche de nouveaux médicaments pour cette maladie. Ainsi deux objectifs ont été visés dans le cadre de ce travail :

Le premier objectif est d'étudier l'influence des substituants portés par la structure principale, pour les différentes séries, sur l'activité antipaludique. Ainsi les techniques QSAR exploitées nous ont permis d'analyser la relation entre l'activité antipaludique et la structure des molécules ce qui a abouti à l'établissement d'un 'modèle QSAR' prédictif.

La première étape de cet objectif consiste à calculer les descripteurs électroniques, constitutionnels, géométriques, lipophiles, topologiques, physico-chimiques et stériques. Pour le calcul des descripteurs électroniques, nous avons utilisé la méthode DFT avec la fonction hybride à trois paramètres de Becke (Lee-Yang-Paar) (B3LYP) [9] et la base 6-31G la mécanique moléculaire MM2 pour les autres descripteurs. La deuxième étape est consacrée à l'application des méthodes statistiques linéaire et non linéaire qui permettent d'élaborer des modèles capables de prédire l'activité antipaludique. La performance de chaque modèle établi a été testée à travers différentes méthodes de validation. Ceci nous a donné une vision claire sur la nature et la pertinence des paramètres physico-chimiques sélectionnés.

Le deuxième objectif est de comprendre le comportement des molécules en interaction avec le recepteur pf-DHFR dans ses deux formes naturelle et mutante puisque la résistance aux médicaments antipaludiques, et qui reste le plus grand problème dans ces dernières années, a été expliquée par la mutation des gènes de la protéine [10,11]. Dans ce sens, nous avons fait appel à la méthode du docking moléculaire pour déterminer dans un premier temps les acides aminés les plus importants pour l'activité antipaludique. Ensuite, chercher comment la mutation de quelques acides aminés peut développer la résistance aux médicaments.

Enfin, ces mêmes étapes avec les mêmes techniques et méthodes ont été appliquées dans les chapitres III, IV et V. De plus, les chapitres I et II, sont réservés à une recherche bibliographique sur la maladie du paludisme, la protéine réceptrice, et les différentes techniques utilisées pour la modélisation des séries des molécules étudiées.

Références

[1] WHO Global Malaria Mapper, WHO. (2016). http://www.who.int/malaria/publications/world_malaria_report/global_malaria_mapper/en/.

[2] N.J. White, S. Pukrittayakamee, T.T. Hien, M.A. Faiz, O.A. Mokuolu, A.M. Dondorp, Malaria, Lancet. 383 (2014) 723–735.

[3] A.R. Brock, J. V. Ross, S. Parikh, A. Esterman, The role of antimalarial quality in the emergence and transmission of resistance, Med. Hypotheses. 111 (2018) 49–54.

[4] C. Prosser, W. Meyer, J. Ellis, R. Lee, Evolutionary ARMS Race: Antimalarial Resistance Molecular Surveillance, Trends Parasitol. 34 (2018) 322–334.

[5] F.W. Muregi, A. Ishih, Next-generation antimalarial drugs: hybrid molecules as a new strategy in drug design, Drug Dev. Res. 71 (2009) n/a-n/a.

[6] A. Kumar, K. Srivastava, S. Raja Kumar, M.I. Siddiqi, S.K. Puri, J.K. Sexana, P.M.S. Chauhan, 4-Anilinoquinoline triazines: A novel class of hybrid antimalarial agents, Eur. J. Med. Chem. 46 (2011) 676–690.

[7] N. Sunduru, M. Sharma, K. Srivastava, S. Rajakumar, S.K. Puri, J.K. Saxena, P.M.S. Chauhan, Synthesis of oxalamide and triazine derivatives as a novel class of hybrid 4-aminoquinoline with potent antiplasmodial activity, Bioorganic Med. Chem. 17 (2009) 6451–6462.

[8] M. Sharma, K. Chauhan, S.S. Chauhan, A. Kumar, S.V. Singh, J.K. Saxena, P. Agarwal, K. Srivastava, S. Raja Kumar, S.K. Puri, P. Shah, M.I. Siddiqi, P.M.S. Chauhan, Synthesis of hybrid 4-anilinoquinoline triazines as potent antimalarial agents, their in silico modeling and bioevaluation as Plasmodium falciparum transketolase and ??-hematin inhibitors, Medchemcomm. 3 (2012) 71–79.

[9] A.D. Becke, Density-functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648–5652.

[10] A.R. Gomes, M. Ravenhall, E.D. Benavente, A. Talman, C. Sutherland, C. Roper, T.G. Clark, S. Campino, Genetic diversity of next generation antimalarial targets: A baseline for drug resistance surveillance programmes, Int. J. Parasitol. Drugs Drug Resist. 7 (2017) 174–180.

[11] J. Yuvaniyama, P. Chitnumsub, S. Kamchonwongpaisan, J. Vanichtanankul, W. Sirawaraporn, P. Taylor, M.D. Walkinshaw, Y. Yuthavong, Insights into antifolate resistance from malarial DHFR-TS structures, Nat. Struct. Biol. 10 (2003) 357–365.

Chapitre I: Généralités sur le paludisme

I Historique du paludisme

Le paludisme est une maladie très ancienne, 3000 av. J.-C, les Égyptiens en souffraient et en mourraient déjà. Cette certitude est issue de la découverte de Plasmodies dans des momies par Miller en 1994 [1]. À peu près à la même époque, soit à partir de 2700 av. J.-C., le paludisme est aussi présent dans la mythologie chinoise [2], le paludisme à l'époque était symbolisée par un dragon à trois têtes : la première ; un marteau, représente les maux de tête, la deuxième ; un seau d'eau glacée, symbolise les frissons et la troisième ; un diadème de fer porté au rouge, représente la fièvre [3]. Depuis, la maladie a également été évoquée dans d'autres civilisations. En raison de son association avec les marais et les zones humides, le paludisme a été appelé "mala-aria" - ce qui se traduit par un "mauvais air" en italien médiéval. La lutte contre cette maladie a connu un très grand intérêt par la communauté scientifique impliquée dans la recherche thérapeutique du paludisme. De ce faite les vertus de l'écorce du quinquina ont été utilisées contre les « fièvres des marais » par Don Francisco Lopez en 1630 au Pérou [4]. Mais c'est en 1880 que Laveran, médecin militaire français en Algérie, mis en évidence le parasite du paludisme [5]. Et c'est en 1898 que Grassi démontra que l'anophèle est le vecteur du paludisme humain [6]. Dans les années 1957, L'OMS entreprit à l'éradication du paludisme mais il a dû rapidement faire place aux stratégies de contrôle (dans le sens de lutte ou maîtrise) au début des années 1970.

II Introduction

Le paludisme reste l'une des principales causes de mortalité dans le monde, malgré des douzaines d'années d'interventions stratégiques visant à réduire la mortalité liée au paludisme. Selon le rapport de l'organisation mondiale de la santé (OMS). En 2016, 216 millions de cas de paludisme dans 91 pays, soit une augmentation de 5 millions par rapport à l'année précédente. Le nombre de décès associé a atteint 445 000, comme l'année 2015. Même si l'incidence du paludisme a diminué au niveau mondial depuis 2010, cette tendance ralentit, voire s'inverse dans certaines régions depuis 2014, et l'évolution de la mortalité liée au paludisme a augmenté [7]. Le paludisme sévit dans 91 pays, essentiellement situés dans les zones intertropicales d'Amérique, d'Asie et d'Afrique. La **figure 2**, illustre cette répartition géographique [8] . La région Afrique concentre toujours 90 % des cas de paludisme et des

décès associés dans le monde [7]. La grande majorité des victimes sont les enfants âgés de moins de cinq ans et les femmes enceintes.

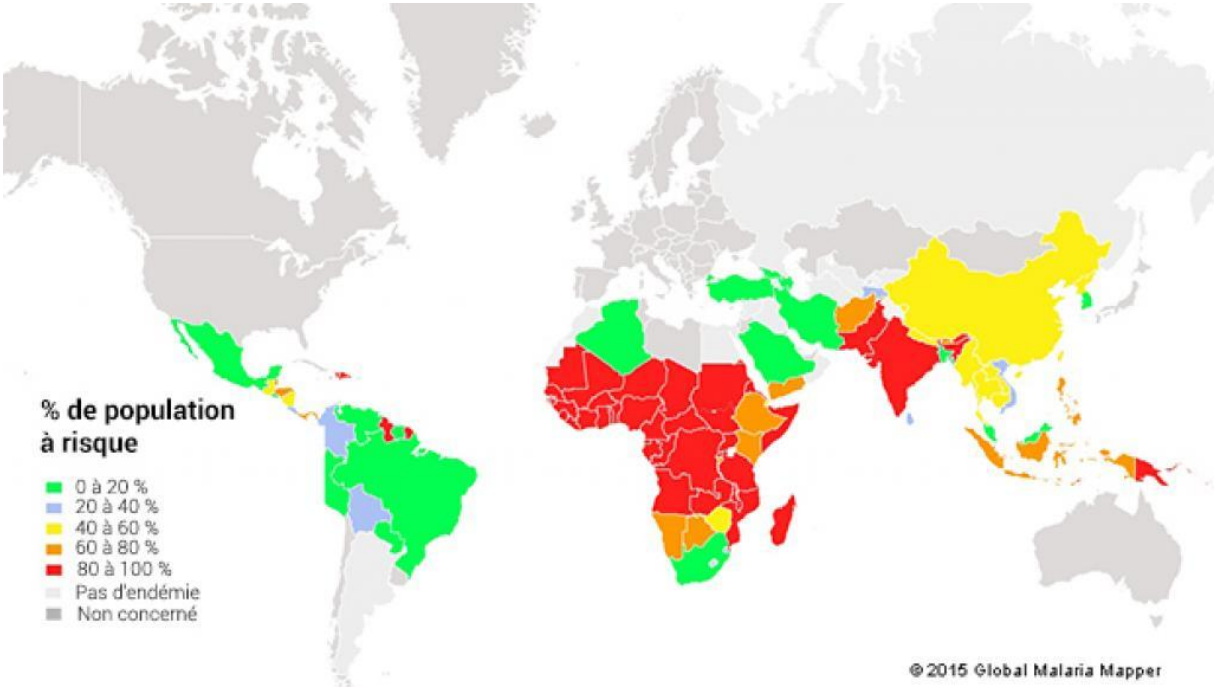


Figure 2 : Répartition géographique du paludisme dans le monde.

En général, le paludisme est causé par des parasites protozoaires du genre Plasmodium [9] et est transmis par les moustique Anophèles femelles [10]. Les parasites infectent et détruisent les globules rouges, entraînant la fièvre, une anémie sévère et, dans certains cas, le paludisme cérébral et, si elles ne sont pas traitées, la mort. Cinq espèces de plasmodium qui infectent les humains et qui sont responsables du paludisme sont identifiées [11–13] :

- *Plasmodium falciparum* est présent dans les zones tropicales d'Amérique Latine, d'Asie et dominante en Afrique. C'est l'espèce la plus répandue et la plus virulente car elle se multiplie rapidement dans le sang et peut entraîner la mort par obstruction des vaisseaux sanguins dans le cerveau.
- *Plasmodium vivax* que l'on retrouve principalement en Asie, en Amérique latine et dans quelques régions d'Afrique, a la particularité de créer des formes hépatiques «dormantes» (hypnozoïtes), pouvant s'activer et envahir le sang dans des semaines voire des mois après la première infection (rechutes cliniques).
- *Plasmodium ovale* est plutôt observé en Afrique de l'ouest et dans les îles de l'ouest du Pacifique. Cette espèce a une biologie et une morphologie proche de *plasmodium vivax* et existe également sous forme d'hypnozoïtes.

- *Plasmodium malariae* a une distribution mondiale. S'il n'est pas traité, il entraîne une infection chronique, durable, pouvant dans certains cas persister toute la vie. Chez certains patients, il peut entraîner de sévères complications.
- *Plasmodium knowlesi* est principalement rencontré en Asie du Sud-Est. Il était d'abord connu chez deux espèces de macaque en 1930, mais il a été découvert chez l'Homme en 1965. Cette espèce peut rapidement progresser vers une infection sévère, pouvant être mortelle.

III Aspect biologique du paludisme

Le cycle de développement parasitaire est pratiquement identique à toutes les espèces de Plasmodium qui infectent l'Homme [14]. Le cycle de vie des parasites du paludisme implique deux hôtes : les êtres humains et les moustiques anophèles femelles, qui agissent respectivement comme hôtes principaux et porteurs du paludisme. La phase sexuelle, sporogonie, a lieu dans le moustique tandis que la phase asexuée, schizogonie, se produit dans l'hôte humain [15]. Chez l'Homme le cycle est divisé en deux phases : une phase hépatique, *exo-érythrocytaire*, correspondant à une période d'incubation cliniquement asymptomatique et une phase sanguine, intra-érythrocytaire, la phase symptomatique [16,17]. Le cycle de vie des parasites plasmodiums responsables du paludisme est décrit dans la **figure 3**.

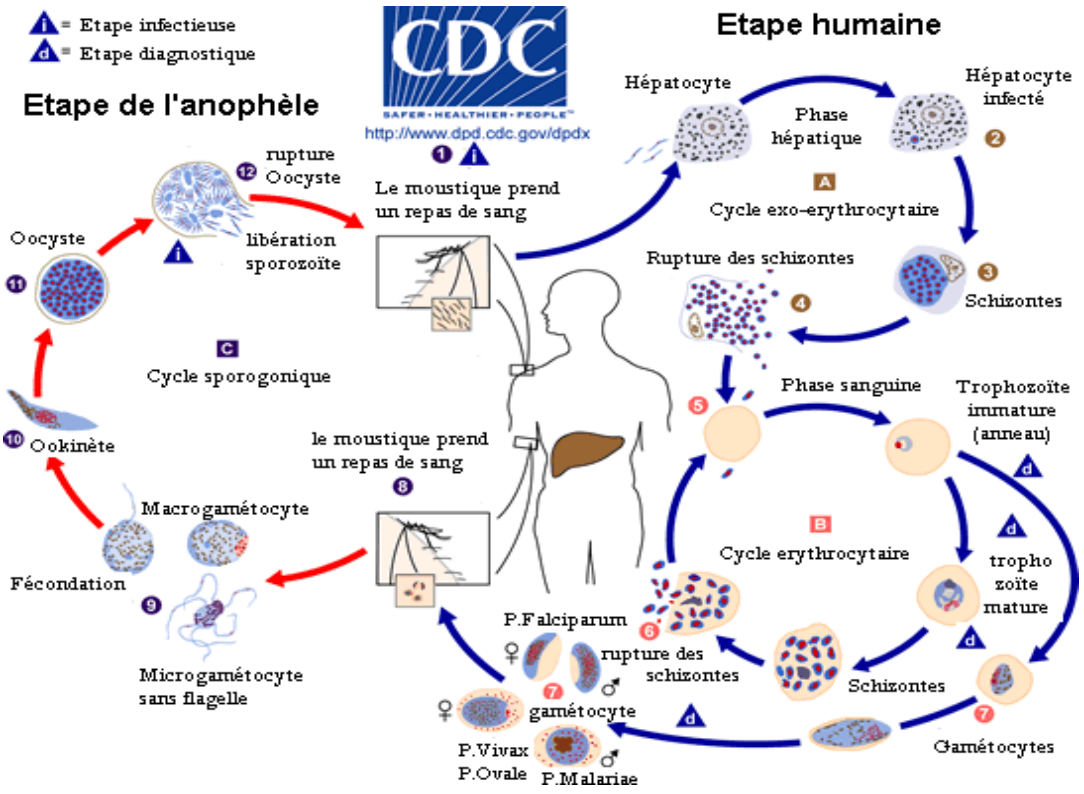


Figure 3 *:* Cycle de vie des parasites plasmodiums responsables du paludisme [18].

III.1 Cycle chez l'Homme

III.1.1 Phase hépatique ou exo-érythrocytaire

Lorsqu'une personne est mordue par un moustique Anophèle femelle infecté, le moustique suce le sang pour nourrir ses œufs et en même temps injecte de la salive qui contient des sporozoïtes, la forme infectieuse du parasite, dans la circulation sanguine de la personne [19,20]. Les sporozoïtes sont transportés dans le sang vers le foie où ils envahissent les hépatocytes se dissimulant ainsi du système immunitaire de l'hôte. Les sporozoïtes qui n'atteindront pas le foie, sont éliminés par les cellules phagocytaires. A ce stade, le plasmodium est incapable de poursuivre son cycle dans les cellules d'autres organes. Après multiplication, cela donne lieu à une phase *exo-érythrocytaire* ou hépatique qui va durer environ de 7 à 15 jours pour le *P. falciparum*, de 15 jours à 9 mois pour le *P. vivax*, de 15 jours à 10 mois pour le *P. ovale* et de 3 semaines pour le *P. malariae*. Les cellules hépatocytes se transforment en schizonte [21].

Le schizonte est une structure qui contient des milliers de minuscules mérozoïtes arrondis. Lors de la maturation des schizontes, les cellules hépatiques vont éclater et libérer des milliers de mérozoïtes dans le sang (10 000 à 30 000 mérozoïtes) [22,23]. Pour le *P. vivax* et le *P. ovale*, certains parasites restent en phase dormante dans le foie (hypnozoïtes) et sont responsables d'une schizogonie hépatique retardée qui va entraîner une libération des mérozoïtes plusieurs semaines, mois ou années plus tard après l'infection. Les hypnozoïtes sont à l'origine des reviviscences tardives observées [24,25]

III.1.2 Cycle intra-érythrocytaire.

Dans la schizogonie érythrocytaire, chaque mérozoïte envahit une globule rouge annonçant le début du paludisme symptomatique. L'invasion sélective des globules rouges est médiée par l'interaction entre les ligands de mérozoïtes et les récepteurs des globules rouges. L'invasion est à la fois rapide et efficace ce qui est nécessaire pour la survie des mérozoïtes, compte tenu du fait que les antigènes sur la surface du mérozoïte sont très vulnérables à l'attaque immunitaire [26]. À ce stade, le parasite consomme l'hémoglobine de l'hôte et modifie la membrane cellulaire du globule rouge pour créer des passages d'importation des nutriments. La digestion de l'hémoglobine (Hb) libère l'hème et fournit au parasite des acides aminés nécessaires pour sa croissance. L'hème est un produit toxique pour le parasite et pour se protéger des effets toxiques de ce produit, le parasite le cristallise en «hémozoïne» .

Après la maturation du schizonte, les globules rouges éclatent et les mérozoïtes s'écoulent dans la circulation sanguine où ils envahissent de nouveaux érythrocytes pour réinitialiser le cycle sanguin et augmenter les niveaux de parasitémie (pourcentage de globules rouges infectés) [27,28]. Le parasite n'est pathogène que dans le stade sanguin; les érythrocytes infectés par le paludisme au stade sanguin adhèrent aux cellules endothéliales et séquestrées dans la microvasculature des organes vitaux, y compris le cerveau, le cœur et les poumons. La séquestration dans le cerveau conduit souvent à une malaria cérébrale [29,30].

III.2 Cycle chez l'anophèle

Certains des mérozoïtes qui pénètrent dans les globules rouges ne se transforment pas en schizontes, mais dans des formes sexuelles connues sous le nom de gamétocytes qui circulent dans la circulation sanguine. Quand un moustique anophèle femelle mord une personne infectée, elle suce ces gamétocytes avec du sang. Dès l'arrivée dans l'estomac de l'anophèle, les gamétocytes se transforment en gamètes mâles semblables à des spermatozoïdes ou à de grands gamètes femelles semblables à des œufs. Lors de la fécondation des gamètes masculins et féminins, on produit un oocyste rempli de sporozoïtes infectieux. Lorsque l'oocystème mûrit, il se rompt, et des milliers de sporozoïtes migrent vers la glande salivaire du moustique, prêt à recommencer le cycle lorsque le moustique se nourrit à nouveau [31].

IV Composés antipaludiques

La chimiothérapie du paludisme peut spécifiquement cibler les parasites. Comme le stade hépatique est l'étape unique la plus longue du cycle de vie du parasite, des études ont montré que le foie est la cible idéale pour les médicaments et les vaccins destinés à traiter et à lutter contre le paludisme [32]. Durant cette période, le patient ne présente aucun symptôme de la maladie. Cette période est donc une formidable opportunité pour éliminer les parasites, mais il est extrêmement difficile d'étudier les stades hépatiques des parasites, et la plupart des médicaments thérapeutiques agissent contre les stades sanguins des parasites.

Tous les médicaments antipaludiques sont principalement des schizonticides sanguins. Les principales classes de médicaments antipaludiques qui ont été ou sont actuellement utilisés en clinique sont les médicaments contenant de la quinoline, les antagonistes du folate et les dérivés de l'artémisinine [33]. Les composés qui ont fait l'objet des études QSAR dans cette thèse dont certains dérivés de la 4-aminoquinoline-triazine, les aurones et aza-aurones n'ont encore pas trouvé accès au marché médical.

IV.1 Dérivés des aurones

Les aurones, 2-Benzylidene-1-benzofuran-3(2H)-one (**figure 4**) sont des métabolites secondaires structurellement apparentés aux flavones. Les aurones se sont montrées actives contre le *P. falciparum* dans le stade intra-erythrocytaire. En revanche, les aza-aurones, (z)-2-benzylidene-4,6-dimethoxyindolin-3-one (**figure 4**), inspirées des aurones naturelles, restent largement insuffisamment exploitées comme agents antipaludiques [34]. Cependant ils agissent comme antipaludiques à deux stades, c'est-à-dire capables d'inhiber les parasites érythrocytaires et hépatiques, comme Il a été étonnamment constaté que la souche résistante à la chloroquine (phénomène détaillé plus bas) était plus sensible aux aurones [35].

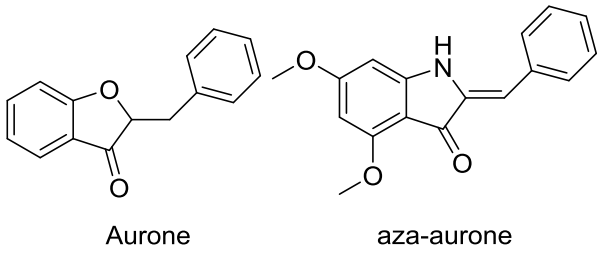


Figure 4 : Structures des aurones et des aza-aurones.

IV.2 Dérivés des quinolines

IV.2.1 Quinine

Les antipaludiques de la quinoline proviennent de la quinine (**figure 5**), qui est un produit naturel antipaludique actif trouvé dans l'écorce péruvien de l'arbre quinquina (Cinchona) [36]. En raison des effets toxiques associés, on a développé des antipaludéens 4-aminoquinolines synthétiques [37]. La plus importante parmi celles-ci est la chloroquine qui a été le pilier de la chimiothérapie du paludisme par *P. falciparum* pendant des décennies [38].

IV.2.2 Dérivés de la 4-Aminoquinoline

Les 4-aminoquinolines se sont prouvés être la classe de composés la plus efficace pour le traitement et la prophylaxie du paludisme parce qu'ils sont faciles à synthétiser et moins couteux, ils ont des profils de toxicité acceptables, et sont bien tolérés et efficaces contre un certain nombre de souches résistantes à la chloroquine [39]. Les modifications du système cyclique de la pamaquine (**figure 5**) ont conduit au développement de la mépacrine et à la synthèse éventuelle de la chloroquine en 1934 par des scientifiques des laboratoires Bayer [40,41]. La chloroquine est devenue le médicament antipaludique le plus utilisé dans les années 1940. La popularité de la chloroquine est attribuable à son efficacité et au faible risque

d'effets secondaires associés à son utilisation aux doses prescrites [40], la chloroquine est efficace contre les infections causées par *P. vivax, P. ovale* ou *P. malariae* [42]. Néanmoins, en raison de la résistance à la chloroquine, d'autres aminoquinolines, l'amodiaquine et la primaquine (**figure 5**) sont plus facilement utilisées. L'amodiaquine fonctionne de manière similaire à la chloroquine et a une certaine efficacité contre les souches du *P. falciparum* résistant à la chloroquine [41]. Cependant, en raison de la résistance croisée (à savoir que la résistance à un médicament rend le parasite résistant à d'autres médicaments de la même famille chimique ou ayant un mode d'action comparable), il est totalement inefficace dans les endroits où les niveaux de résistance à la chloroquine sont plus élevés [43,44].

IV.2.3 Dérivés de la 8-aminoquinoline

L'observation que le bleu de méthylène présentait une certaine activité antipaludique a encouragé la synthèse des 8-aminoquinolines [45]. Le premier antipaludique synthétique dérivé de la 8-aminoquinoline, la pamaquine (**figure 5**), est synthétisée dans les années 1920 [46] . Toutefois, il a été constaté que la pamaquine provoquait une méthémoglobinémie (taux de méthémoglobine dans le sang et par extension le nom de la maladie correspondant à un taux trop important de méthémoglobine dans le sang) grave chez les patients souffrant d'un déficit en glucose-6-phosphate déshydrogénase (G6PD) (favisme, ou déficit en G6PD, est le déficit enzymatique le plus répandu dans le monde. Il est d'origine génétique, il se caractérise par une destruction des globules rouges) [47,48], ce qui a conduit à la recherche d'analogues plus sûrs. La primaquine, l'analogue d'amine primaire moins toxique du pamaquine, est actuellement considérée comme le médicament de premier choix pour la guérison radicale du paludisme [49]. En plus de leurs activités contre les hypnozoïtes *plasmodium vivax* et *plasmodium ovale,* les 8-aminoquinolines ont une activité gamétocytocide forte (diminution de la densité gamétocytaire dans le stade de transmission sexuelle.) [50].

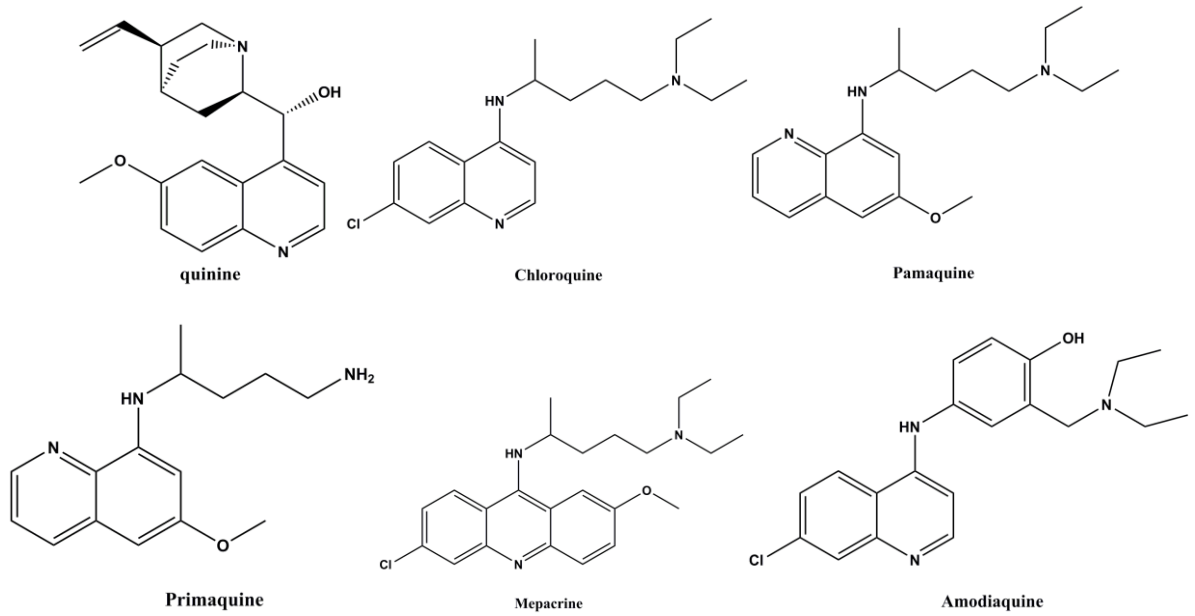


Figure 5 : Structures chimiques de la chloroquine, l'amodiaquine, la pamaquine, la mepacrine, la quinine, et de la primaquine.

IV.2.5 Mode d'action des antipaludiques quinolines

On croit que les antipaludiques de la quinoline en général et les 4-aminoquinolines en particulier, exercent leur activités antipaludiques principalement en diminuant le taux de formation des hemozoïnes [51]. Il en résulte des concentrations élevées d'hème dans la vacuole alimentaire qui sont fatales pour le parasite. Dans les globules rouges, les parasites ont besoin de dégrader l'hémoglobine pour acquérir les acides aminés essentiels de l'hème. Par contre, l'hème est toxique pour les parasites et est donc biocristallisée pour former des molécules d'hémozoïnes non toxiques. La chloroquine se lie ensuite à l'héme, ce qui donne lieu à un complexe chloroquine-héme extrêmement toxique qui provoque la lyse des cellules parasitaires et l'autodigestion subséquente [52,53]. Les quinines ont tendance à ne s'associer que faiblement avec l'hème et peuvent exercer leurs effets en interférant avec d'autres étapes du métabolisme de l'hémoglobine [40].

IV.2.6 Résistance à la chloroquine

Le développement de la résistance à la chloroquine est apparu au début des années 1960. Il a été localisée en Amérique du Sud et en Asie du Sud-Est, mais à ce jour, la résistance s'est propagée à pratiquement toutes les zones où le paludisme est endémique [54].

Pour comprendre l'apparition de résistance aux médicaments à la quinoline chez *P. falciparum*, trois points fondamentaux doivent être pris en considération:

- les mécanismes cellulaires par lesquels les parasites échappent à l'action toxique des médicaments;
- les changements possibles dans l'expression des gènes ou des mutations génétiques qui peuvent conduire au développement de la résistance;
- l'inversion de résistance par des modifications aux médicaments existants de quinoline.

Des études récentes ont montré que la résistance à la chloroquine peut être associée à deux protéines qui permettent au parasite d'échapper à la toxicité du médicament en expulsant le médicament hors de la vacuole digestive du parasite [55]. La première protéine est codée par *Pfmdr1* (*Plasmodium falciparum multidrug-resistance 1*), le gène impliqué dans la multi-chimiorésistance aux antipaludiques. Et de plus, cette protéine a été considérée comme étant impliquée dans le transport d'importation vers la vacuole digestive du parasite. La mutation du codon *N86Y* du gène *Pfmdr1* est la plus étudiée. Des échecs thérapeutiques à la chloroquine et à l'amodiaquine liée à cette mutation ont été rapportés. Cependant cette mutation n'est pas hautement corrélée à la résistance à la chloroquine [56]. La seconde protéine est codée par Pfcrt (*Plasmodium falciparum chloroquine resistance transporter*) est située sur la membrane de la vacuole alimentaire dans le parasite et a été directement liée à la résistance à la chloroquine [57].

Des études ont révélé que seul un changement d'acide aminé K76T (responsable de la résistance à la chloroquine, par la mutation qui consiste à un remplacement d'un acide aminé lysine par une thréonine au niveau de la position 76 de la protéine *Pfcrt)* [58,59]. Il est situé dans le premier domaine transmembranaire, où se trouve constamment dans tous les parasites de la résistance à la chloroquine. Les mutations ont montré une réduction marquée de l'accumulation de la chloroquine par le parasite. Cela conduit à une réduction de la concentration globale de chloroquine sur son site d'action, et donc une diminution de la sensibilité du parasite.

IV.3 Triazines et composés similaires

Une recherche intensive d'antipaludiques au Royaume-Uni pendant la Seconde Guerre mondiale a identifié le proguanil (**figure 6**), ce dérivé de pyrimidine ouvert (biguanide) montre une meilleure activité que la drogue quinine, bien que son action ait été lente contre le paludisme aigu et il semblait entraîner une résistance à certains parasites [60]. Le proguanil est un médicament converti au niveau hépatique en un métabolite triaziné actif [41], le

cycloguanil (**figure 6**), qui possède une activité antipaludique prometteuse [61,62]. Le cycloguanil est efficace contre les sporozoïtes qui agissent comme inhibiteurs de la *dihydrofolate réductase* pour stopper la voie biosynthétique du folate est donc la prolifération du parasite, dont la plus élevée était la pyriméthamine (**figure 6**), développée en 1951 par Falco et Hitchings [63]. Cependant, ce composé a facilement induit une résistance et, par conséquent, la pyriméthamine a été utilisée plus tard en association avec la sulfadoxine (**figure 6**), un antibiotique sulfonamide (figure 4), qui a un effet synergique. Ces composés ciblent la voie essentielle de biosynthèse du folate par les parasites antipaludiques en inhibant les enzymes *dihydrofolate réductase* et la *dihydroptéroate synthétase* [64].

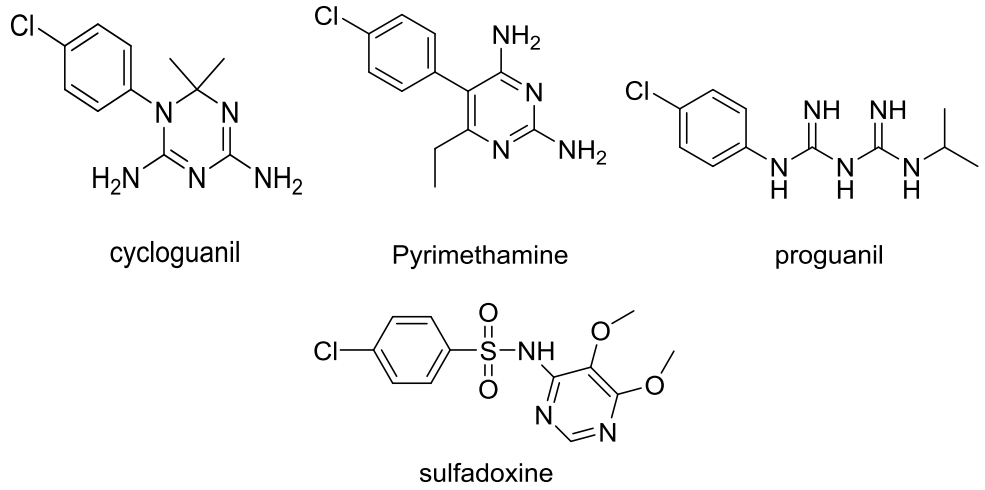


Figure 6 : Structures chimiques de la cycloguanil, de la pyriméthamine, de la proguanil et de la sulfadoxine.

IV.3.1 Mode d'action des composés antifolates

La pyriméthamine et le cycloguanil interfèrent avec la synthèse d'acide tetrahydrofolate à partir d'acide folique en inhibant la *dihydrofolate réductase* (DHFR) qui produit l'acide folinique (**figure 7**) [65,66]. Ils privent le parasite des facteurs généraux de l'acide folinique réduits, qui n'est pas seulement un moyen efficace de tuer les parasites, mais il bloque également la capacité du parasite à se multiplier [67].

Néanmoins, la pyriméthamine et le cycloguanil peuvent entraîner une carence en acide folique chez les humains en raison des effets sur la DHFR. Il est efficace contre toutes les formes de paludisme humain, cependant, la résistance à la pyrimethamine est apparue rapidement après son introduction [68–70], la raison pour laquelle, elle a été par la suite utilisée en combinaison avec les sulfamides [71]. L'inhibition de la DHPS par les sulfamides diminue la quantité de la dihydroptéroate et donc de la dihydrofolate disponible pour l'activité de la DHFR. La

diminution concomitante du substrat et de l'activité de la DHFR permet à la combinaison de ces deux drogues d'être redoutablement efficace. Malheureusement, la résistance à la combinaison de sulfadoxine-pyriméthamine augmente chez le *P. vivax*. Dans les zones où ce dernier est encore susceptible, cette combinaison est encore un traitement largement utilisé [72].

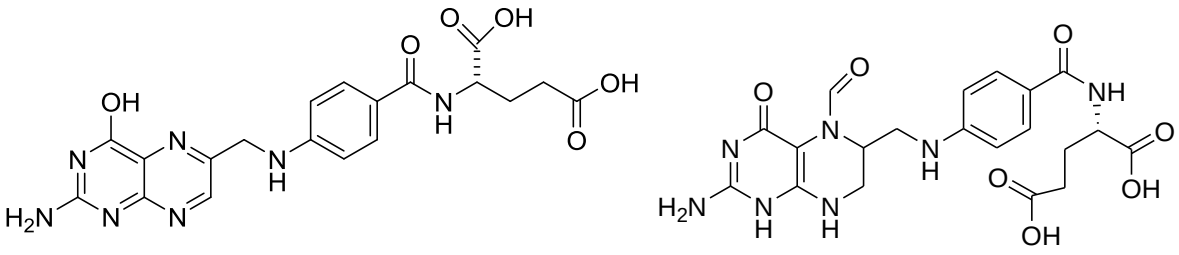


Figure 7 : Structures de l'acide folique et de l'acide folinique

IV.3.2 Résistance aux antifolates

La résistance aux antagonistes du folate s'est développée rapidement après l'introduction de ces médicaments et elle est associée à des mutations ponctuelles sur la *dihydrofolate réductase* (DHFR) et *dihydroptéroate synthétase* (DHPS).

La résistance est causée par des mutations ponctuelles dans les gènes de la DHFR et de la DHPS. Ces mutations atteignent les acides aminés en positions 51, 59, 108 et 164 de la DHFR et sont codés respectivement N51I, C59R, S108N et I164L. Ceux-ci sont fortement associés à l'échec thérapeutique à la pyriméthamine et cycloguanil chez le *P. falciparum*, tandis que les mutations dans la DHPS génèrent les codons S436A, A437G, K540E, A581G et A613S, qui semblent jouer un rôle majeur dans le développement de la résistance au sulfadoxine chez le *P. falciparum* [73–76]. Il a également été suggéré que les parasites présentent une limitation des possibilités de mutation. De telles observations ont servi d'impulsion au développement des nouvelles molécules hybrides capables de résoudre le problème de la résistance aux médicaments antifolates.

V Intégration des molécules hybrides dans la chimiothérapie antipaludique à base de Chloroquine (dérivées de la quinoline) et cycloguanil (dérivées de la triazine).

L'un des défis de la future chimiothérapie antipaludique consiste à mettre au point des composés novateurs en ce qui concerne la structure chimique et la cible moléculaire [77].

Récemment, dans le cadre d'une approche rationnelle de conception de médicaments, des molécules hybrides uniques à double fonctionnalité et/ou des cibles ont été mises au point en tant que nouveaux médicaments antipaludiques. Il a été démontré que certains de ces médicaments hybrides sont des agents antipaludiques puissants, avec une toxicité acceptable [78].

Toutefois, compte tenu du grand intérêt pour les molécules hybrides et des nombreux rapports encourageants sur l'efficacité et la non-toxicité, la prochaine génération d'antipaludiques pourrait aussi bien être des médicaments hybrides que des médicaments à composants multiples. L'utilisation de molécules hybrides présente de nombreux avantages par rapport aux médicaments à plusieurs composants dans la thérapie antipaludique. Comparativement à ces derniers, les médicaments hybrides peuvent être moins coûteux puisque, en principe, les risques et les coûts ne peuvent être différents d'aucune autre entité. Un autre avantage est le risque moins élevé d'interactions médicamenteuses indésirables que les médicaments à composantes multiples. L'inconvénient, cependant, est qu'il est difficile d'ajuster le coefficient des activités aux différentes cibles thérapeutiques [79]. Les molécules hybrides peuvent être classées dans la catégorie suivante [13] :

- Molécules hybrides conjuguées, dans lesquelles les structures moléculaires qui contiennent les pharmacophores pour chaque cible sont séparées par un groupe de liaison différent que l'on ne retrouve dans aucun des médicaments individuels. La plupart des hybrides conjugués contiennent un agent de liaison métaboliquement stable [79].
- Les hybrides conjugués de clivage ont un agent de liaison conçu pour être métabolisé afin de libérer les deux médicaments qui interagissent indépendamment avec chaque cible.
- les hybrides fusionnés (fused hybrid) ont la taille du lien réduite de sorte que la structure des pharmacophores entre en contact avec elles.
- Les hybrides combinés (Merged hybrids) ont leurs cadres de travail fusionnés en tirant parti des points communs dans les structures des composés de départ, qui donnent naissance à des molécules plus petites et plus simples [79].

De nombreux efforts sont déployés pour tenter de découvrir des inhibiteurs efficaces pour la protéine dihydrofolate réductase–thymidylate synthèse (PfDHFR-TS) à cause de la résistance aux médicaments. La fraction quinoline a attiré l'attention des chimistes médicinaux, car c'est l'une des pharmacophores cruciales responsables de l'action antipaludique [80,81]. D'autre

part, les dérivés du 1,3,5-triazine à savoir le cycloguanil, le chlorcycloguanil, le clociguanil et le WR99210 sont déjà approuvés comme inhibiteurs efficaces de la DHFR, inhibiteurs spécifiques du domaine du *P. falciparum* et inhibiteurs sélectifs des processus biochimiques vitaux pour la croissance des parasites [82]. Les propriétés pharmacocinétiques respectives de la chloroquine et de la cycloguanil ont déjà été très largement étudiées [83,84], La chloroquine et de la cycloguanil ont des profils pharmacocinétiques raisonnablement semblables avec un mode d'action varié sur des sites biochimiques parasitaires du gène de la *dihydrofolate réductase* (DHFR) [84]. De ce faite les molécules hybrides à base de la chloroquine et de la cycloguanil se sont montrées les meilleurs remèdes thérapeutiques contre la résistance aux antifolates. Dans ce travail nous nous sommes intéressés à la modélisation des molécules hybrides à base de chloroquine et de 1,3,5-triazine (**figure 8**) dans une tentative de palier à ce problème de résistance au médicament antipaludique en essayant d'abord de comprendre le mode de liaison et les principales interactions moléculaires clés entre les hybrides et la protéine DHFR sauvage et mutante (**figure 9**), ce qui ensuite nous a permis d'expliquer la différence d'activité de ces molécules [84,85].

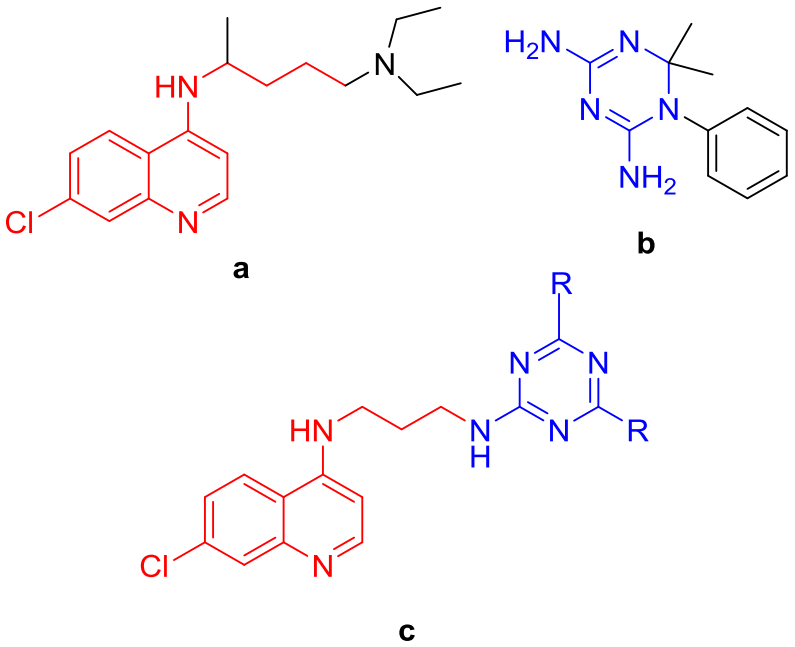


Figure 8 : Structures de (a) chloroquine, (b) cycloguanil et (c) hybrides dérivés du 4-aminoquinoline-1,3,5-triazine.

VI Présentation de la PfDHFR-TS

Plasmodium falciparum dihydrofolate réductase-thymidylate synthase (PfDHFR-TS) est une cible importante pour les médicaments antipaludiques. Les Pf-DHFR de type sauvage et le

quadruple mutant ont été cristallisé la première fois en 2003 par Jirundon Yuvaniyama et al [86]. Le DHFR et le TS coexistent en tant qu'enzyme bifonctionnelle à chaîne unique. Le polypeptide Pf-DHFR-TS contient 608 acides aminés, les 231 premiers résidus constituant le domaine DHFR. On trouve ensuite 89 résidus désignés comme région de jonction entre le domaine (DHFR ,TS) et 288 résidus du domaine TS **(figure 9)** [87–89]

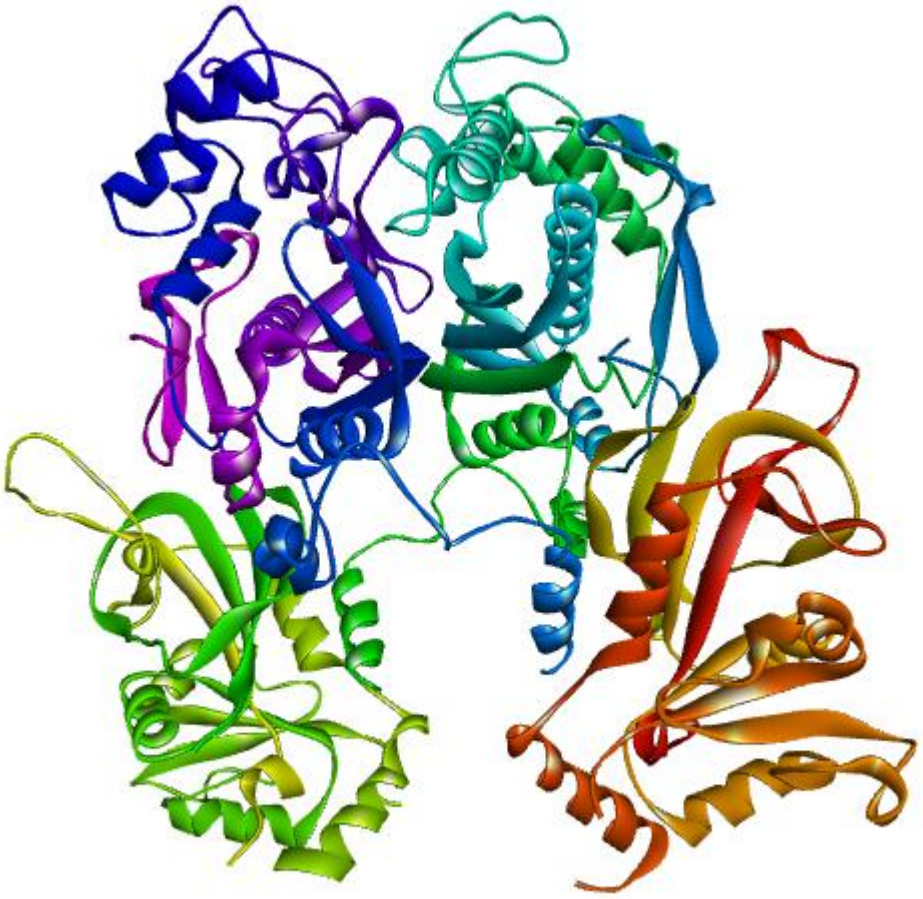


Figure 9 : Structure du PfDHFR-TS type sauvage (en vert et jaune) couplé avec WR99210 (en rouge), NADPH (en bleu) et dUMP (en orange). Les régions de jonction et les domaines TS sont en violet et bleu ciel.

Le domaine DHFR du PfDHFR-TS présente une cible d'inhibition préférentielle pour la chimiothérapie. L'efficacité des médicaments inhibiteurs du DHFR est maintenant diminuée en raison de mutations qui empêchent la liaison médicamenteuse tout en conservant l'activité enzymatique. La résistance aux antifoliés est due à des mutations du PfDHFR-TS, la première mutation est observée au niveau du résidu 108, ensuite trois autres résidus (51,59,164) ont subi la mutation, ceci entraîne une affinité de liaison de plus en plus faible de l'enzyme avec les inhibiteurs [76,90]. Les quatre mutations situées dans le domaine DHFR sont codées : N51I/C59R/S108N/I164L, consistant à un remplacement des acides aminés : l’asparagine par

l’isoleucine, la cystéine par l’arginine, la sérine par l’asparagine et l’isoleucine par la leucine au niveau des positions 51, 59, 108 et 164 respectivement.

Les structures cristallines de PfDHFR-TS du type sauvage et le quadruple mutant, en complexe avec un puissant inhibiteur WR99210 **(figure 10)**, ainsi que le double mutant résistant avec la pyriméthamine antipaludique**,** révèlent des caractéristiques pour vaincre la résistance [86].

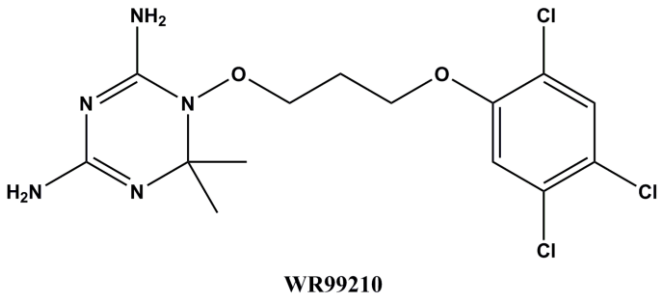


Figure 10 *:* Structure chimique du WR99210

Le quadruple mutation de DHFR qui présente une résistance élevée à la pyriméthamine et au cycloguanil reste encore sensible au WR99210**.** Contrairement à la pyriméthamine, la chaîne latérale flexible du WR99210 qui peut adopter une conformation s'adaptant bien dans le site actif fournissant des explications pour la mauvaise liaison observée de la pyriméthamine par rapport à la bonne liaison du WR99210 (**tableau 1**), de plus, ils fournissent de nouvelles stratégies pour la conception de nouveaux médicaments pour vaincre la résistance antifoliée [86].

Tableau 1 : Concentration d'inhibition des composés pyriméthamine, cycloguanil, et WR99210 de la PfDHFR-TS

Types d'enzymes DHFR	Activité antiplasmodiale IC50 (μM)		
	pyriméthamine	cycloguanil	WR99210
type sauvage	0,08 ± 0,01	0,037 ± 0,012	0,00057 ± 0,0001
double mutant (C59R/S108N)	30,9 ± 8,4	2,4 ± 1,1	0,0023 ± 0,001
Quadruple mutant N51I/C59R/S108N/I164L	>100	>100	0,018 ± 0,01

La région du site actif contient des résidus qui ont été identifiés comme importants dans l'activité du DHFR: Ile14, Ala16, Trp48, Asp54, Phe58, Ser108, Ile164 et Thr185. Ces résidus interagissent avec le dihydrofolate et les inhibiteurs antifoliés. Cependant, les résidus qui ont été impliqués dans le quadruple mutation conduisant à la résistance (Ala16, Cys50, Asn51, Cys59, Ser108 et Ile164) sont tous localisés dans la région du site actif [86].

Références

[1] R.L. Miller, S. Ikram, G.J. Armelagoss, R. Walker, B. Harers, C.J. Shiffi, D. Baggett, M. Carrigan, S.M. Maret, Diagnosis of Plasmodium falciparum infections in mummies using the rapid manual ParaSightTM-F test, Trans. R. Soc. Trop. Med. Hyg. 88 (1994) 31–32.

[2] F.E.G. Cox, History of human parasitology., Clin. Microbiol. Rev. 15 (2002) 595–612. http://www.ncbi.nlm.nih.gov/pubmed/12364371.

[3] M. Ni, The Yellow Emperor's Classic of medicine : a new translation of the Neijing Suwen with commentary, Shambhala, 1995. https://books.google.co.ma/books?hl=ar&lr=&id=CpfOtTRofwgC&oi=fnd&pg=PP1&dq=The+Yellow+Emperor's+Classic+of+Medicine+:+A+New+Translation+of+the+Neijing&ots=PAOhQBu8a7&sig=A3u4yGv_Jlr-WcOomD3MzJ-hRXE&redir_esc=y#v=onepage&q=The Yellow Emperor's Classic of.

[4] K.S. Zimmerer, HUMBOLDT'S NODES AND MODES OF INTERDISCIPLINARY ENVIRONMENTAL SCIENCE IN THE ANDEAN WORLD*, Geogr. Rev. 96 (2010) 335–360.

[5] L.J. Bruce-Chwatt, Alphonse Laveran's discovery 100 years ago and today's global fight against malaria., J. R. Soc. Med. 74 (1981) 531–6. http://www.ncbi.nlm.nih.gov/pubmed/7021827.

[6] S.R. Meshnick, M.J. Dobson, The History of Antimalarial Drugs, (2001) 15–25.

[7] OMS | Rapport sur le paludisme dans le monde 2017, WHO. (2017). http://www.who.int/malaria/publications/world-malaria-report-2017/report/fr/.

[8] WHO | Global Malaria Mapper, WHO. (2016). http://www.who.int/malaria/publications/world_malaria_report/global_malaria_mapper/en/.

[9] J.B. Vaishnani, Cutaneous findings in five cases of malaria., Indian J. Dermatol. Venereol. Leprol. 77 (2011) 110.

[10] I. Weissbuch, L. Leiserowitz, Interplay Between Malaria, Crystalline Hemozoin Formation, and Antimalarial Drug Action and Design, Chem. Rev. 108 (2008) 4899–4914.

[11] CDC - Malaria, (n.d.). https://www.cdc.gov/malaria/.

[12] P. Newton, BM, DPhil, MRCP, N. White, DSc, MD, FRCP, MALARIA: New Developments in Treatment and Prevention, Annu. Rev. Med. 50 (1999) 179–192.

[13] M.C. Murray, M.E. Perkins, Chapter 15. Chemotherapy of Malaria, Annu. Rep. Med. Chem. 31 (1996) 141–150.

[14] C. Zhang, Z. Li, H. Cui, Y. Jiang, Z. Yang, X. Wang, H. Gao, C. Liu, S. Zhang, X. Su, J. Yuan, Systematic CRISPR-Cas9-Mediated Modifications of *Plasmodium yoelii* ApiAP2 Genes Reveal Functional Insights into Parasite Development, MBio. 8 (2017) e01986-17.

[15] S. Antinori, L. Galimberti, L. Milazzo, M. Corbellino, Biology of human malaria plasmodia including Plasmodium knowlesi., Mediterr. J. Hematol. Infect. Dis. 4 (2012) e2012013.

[16] A.F. Cowman, B.S. Crabb, A parasite genome sheds light on an old enemy, Nat. Biotechnol. 20 (2002) 1098–1099.

[17] N.J. White, S. Pukrittayakamee, T.T. Hien, M.A. Faiz, O.A. Mokuolu, A.M. Dondorp, Malaria, Lancet. 383 (2014) 723–735.

[18] C.-C. for D.C. and Prevention, CDC - Malaria - About Malaria - Biology, (n.d.). https://www.cdc.gov/malaria/about/biology/ (accessed December 16, 2017).

[19] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of State Calculations by Fast Computing Machines, J. Chem. Phys. 21 (1953) 1087–1092.

[20] B.J. Alder, T.E. Wainwright, Phase Transition for a Hard Sphere System, J. Chem. Phys. 27 (1957) 1208–1209.

[21] G. Valkiūnas, T.A. Iezhova, Exo-erythrocytic development of avian malaria and related haemosporidian parasites., Malar. J. 16 (2017) 101.

[22] R. Ménard, The journey of the malaria sporozoite through its hosts: two parasite proteins lead the way., Microbes Infect. 2 (2000) 633–42. http://www.ncbi.nlm.nih.gov/pubmed/10884614.

[23] M. Schlitzer, Malaria Chemotherapeutics Part I: History of Antimalarial Drug Development, Currently Used Therapeutics, and Drugs in Clinical Development, ChemMedChem. 2 (2007) 944–986.

[24] B.M. Greenwood, D.A. Fidock, D.E. Kyle, S.H.I. Kappe, P.L. Alonso, F.H. Collins, P.E. Duffy, Malaria: progress, perils, and prospects for eradication, J. Clin. Invest. 118 (2008) 1266–1276.

[25] S. Meister, D.M. Plouffe, K.L. Kuhen, G.M.C. Bonamy, T. Wu, S.W. Barnes, S.E. Bopp, R. Borboa, A.T. Bright, J. Che, S. Cohen, N. V. Dharia, K. Gagaring, M. Gettayacamin, P. Gordon, T. Groessl, N. Kato, M.C.S. Lee, C.W. McNamara, D.A. Fidock, A. Nagle, T. -g. Nam, W. Richmond, J. Roland, M. Rottmann, B. Zhou, P.

Froissard, R.J. Glynne, D. Mazier, J. Sattabongkot, P.G. Schultz, T. Tuntland, J.R. Walker, Y. Zhou, A. Chatterjee, T.T. Diagana, E.A. Winzeler, Imaging of Plasmodium Liver Stages to Drive Next-Generation Antimalarial Drug Discovery, Science (80-.). 334 (2011) 1372–1377.

[26] A.F. Cowman, B.S. Crabb, Invasion of Red Blood Cells by Malaria Parasites, Cell. 124 (2006) 755–766.

[27] V.L. Lew, Malaria: Endless Fascination with Merozoite Release, Curr. Biol. 15 (2005) R760–R761.

[28] L. Bannister, G. Mitchell, The ins, outs and roundabouts of malaria., Trends Parasitol. 19 (2003) 209–13. http://www.ncbi.nlm.nih.gov/pubmed/12763426.

[29] L. Zhang, B. Leifate, A. Kelala, The dynamic alteration in Gq/11 protein expression in multiple organs during acute respiratory distress syndrome in rat, Chinese Crit. Care Med. 22 (2010) 477–481.

[30] H. Cai, R. Kuang, J. Gu, Y. Wang, Proteases in Malaria Parasites - A Phylogenomic Perspective, Curr. Genomics. 12 (2011) 417–427.

[31] L.A. Baton, L.C. Ranford-Cartwright, Spreading the seeds of million-murdering death: metamorphoses of malaria in the mosquito, Trends Parasitol. 21 (2005) 573–580.

[32] C.W. McNamara, M.C. Lee, C.S. Lim, S.H. Lim, J. Roland, O. Simon, B.K. Yeung, A.K. Chatterjee, S.L. McCormack, M.J. Manary, A.-M. Zeeman, K.J. Dechering, T.S. Kumar, P.P. Henrich, K. Gagaring, M. Ibanez, N. Kato, K.L. Kuhen, C. Fischli, A. Nagle, M. Rottmann, D.M. Plouffe, B. Bursulaya, S. Meister, L. Rameh, J. Trappe, D. Haasen, M. Timmerman, R.W. Sauerwein, R. Suwanarusk, B. Russell, L. Renia, F. Nosten, D.C. Tully, C.H. Kocken, R.J. Glynne, C. Bodenreider, D.A. Fidock, T.T. Diagana, E.A. Winzeler, Targeting Plasmodium PI(4)K to eliminate malaria., Nature. 504 (2013) 248–253.

[33] P. Olliaro, Mode of action and mechanisms of resistance for antimalarial drugs., Pharmacol. Ther. 89 (2001) 207–19. http://www.ncbi.nlm.nih.gov/pubmed/11316521.

[34] O. Kayser, A.F. Kiderlen, U. Folkens, H. Kolodziej, In Vitro Leishmanicidal Activity of Aurones, Planta Med. 65 (1999) 316–319.

[35] O. Kayser, A. Kiderlen, R. Brun, *In Vitro* Activity of Aurones against *Plasmodium falciparum* Strains K1 and NF54, Planta Med. 67 (2001) 718–721.

[36] C.E. Song, ed., Cinchona Alkaloids in Synthesis and Catalysis, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2009.

[37] R.G. Ridley, Medical need, scientific opportunity and the drive for antimalarial drugs,

Nature. 415 (2002) 686–693.

[38] P.A. Winstanley, Chemotherapy for Falciparum Malaria: The Armoury, the Problems and the Prospects, Parasitol. Today. 16 (2000) 146–153.

[39] P.M. O'Neill, P.G. Bray, S.R. Hawley, S.A. Ward, B.K. Park, 4-Aminoquinolines--past, present, and future: a chemical perspective., Pharmacol. Ther. 77 (1998) 29–58. http://www.ncbi.nlm.nih.gov/pubmed/9500158 .

[40] M. Foley, L. Tilley, Quinoline antimalarials: mechanisms of action and resistance and prospects for new agents., Pharmacol. Ther. 79 (1998) 55–87. http://www.ncbi.nlm.nih.gov/pubmed/9719345 (accessed December 16, 2017).

[41] P.M. O'Neill, V.E. Barton, S.A. Ward, J. Chadwick, 4-Aminoquinolines: Chloroquine, Amodiaquine and Next-Generation Analogues, in: Treat. Prev. Malar., Springer Basel, Basel, 2011: pp. 19–44.

[42] J.L. Deen, L. von Seidlein, A. Dondorp, Therapy of uncomplicated malaria in children: a review of treatment principles, essential drugs and current recommendations, Trop. Med. Int. Heal. 13 (2008) 1111–1130.

[43] M. Lemnge, M. Alifrangis, M.Y. Kafuye, M.D. Segeja, S. Gesase, D. Minja, J.J. Massaga, A.M. Rønn, I.C. Bygbjerg, High reinfection rate and treatment failures in children treated with amodiaquine for falciparum malaria in Muheza villages, Northeastern Tanzania., Am. J. Trop. Med. Hyg. 75 (2006) 188–93. http://www.ncbi.nlm.nih.gov/pubmed/16896117.

[44] B. Nsimba, V. Guiyedi, M. Mabika-Mamfoumbi, J. Mourou-Mbina, E. Ngoungou, M. Bouyou-Akotet, R. Loembet, R. Durand, J. Le Bras, M. Kombila, Sulphadoxine/pyrimethamine versus amodiaquine for treating uncomplicated childhood malaria in Gabon: A randomized trial to guide national policy, Malar. J. 7 (2008) 31.

[45] J.L. Vennerstrom, M.T. Makler, C.K. Angerhofer, J.A. Williams, Antimalarial dyes revisited: xanthenes, azines, oxazines, and thiazines., Antimicrob. Agents Chemother. 39 (1995) 2671–7.

[46] V.R. Solomon, H. Lee, Chloroquine and its analogs: A new promise of an old drug for effective and safe cancer therapies, Eur. J. Pharmacol. 625 (2009) 220–233.

[47] P.E. Carson, R. Hohl, M. V Nora, G.W. Parkhurst, T. Ahmad, S. Scanlan, H. Frischer, Toxicology of the 8-aminoquinolines and genetic factors associated with their toxicity in man., Bull. World Health Organ. 59 (1981) 427–37. http://www.ncbi.nlm.nih.gov/pubmed/6976850.

[48] R.E. Howes, K.E. Battle, A.W. Satyagraha, J.K. Baird, S.I. Hay, G6PD Deficiency, in: Adv. Parasitol., 2013: pp. 133–201.

[49] J.K. Baird, S.L. Hoffman, Primaquine Therapy for Malaria, Clin. Infect. Dis. 39 (2004) 1336–1345.

[50] D.F. Clyde, Clinical problems associated with the use of primaquine as a tissue schizontocidal and gametocytocidal drug., Bull. World Health Organ. 59 (1981) 391–5. http://www.ncbi.nlm.nih.gov/pubmed/6976846.

[51] T.J. Egan, K.K. Ncokazi, Quinoline antimalarials decrease the rate of β-hematin formation, J. Inorg. Biochem. 99 (2005) 1532–1539.

[52] E. Hempelmann, Hemozoin Biocrystallization in Plasmodium falciparum and the antimalarial activity of crystallization inhibitors, Parasitol. Res. 100 (2007) 671–676.

[53] A. Saifi, T. Beg, A.H. Harrath, F.S.H. Altayalan, S. Al Quraishy, Antimalarial drugs: Mode of action and status of resistance, African J. Pharm. Pharmacol. 7 (2013) 148–156.

[54] M. Foley, L. Tilley, Quinoline antimalarials: mechanisms of action and resistance., Int. J. Parasitol. 27 (1997) 231–40. http://www.ncbi.nlm.nih.gov/pubmed/9088993.

[55] P.E. Ferreira, G. Holmgren, M.I. Veiga, P. Uhlén, A. Kaneko, J.P. Gil, PfMDR1: Mechanisms of Transport Modulation by Functional Polymorphisms, PLoS One. 6 (2011) e23875.

[56] S. Picot, P. Olliaro, F. de Monbrison, A.-L. Bienvenu, R.N. Price, P. Ringwald, A systematic review and meta-analysis of evidence for correlation between molecular markers of parasite resistance and treatment outcome in falciparum malaria, Malar. J. 8 (2009) 89.

[57] S. Nessler, O. Friedrich, N. Bakouh, R.H.A. Fink, C.P. Sanchez, G. Planelles, M. Lanzer, Evidence for Activation of Endogenous Transporters in *Xenopus laevis* Oocytes Expressing the *Plasmodium falciparum* Chloroquine Resistance Transporter, PfCRT, J. Biol. Chem. 279 (2004) 39438–39446.

[58] P.G. Bray, R.E. Martin, L. Tilley, S.A. Ward, K. Kirk, D.A. Fidock, Defining the role of PfCRT in Plasmodium falciparum chloroquine resistance, Mol. Microbiol. 56 (2005) 323–333.

[59] M. Imwong, A.M. Dondorp, F. Nosten, P. Yi, M. Mungthin, S. Hanchana, D. Das, A.P. Phyo, K.M. Lwin, S. Pukrittayakamee, S.J. Lee, S. Saisung, K. Koecharoen, C. Nguon, N.P.J. Day, D. Socheat, N.J. White, Exploring the Contribution of Candidate Genes to Artemisinin Resistance in Plasmodium falciparum, Antimicrob. Agents Chemother. 54

(2010) 2886–2892.

[60] Bruce-Chwatt LJ, Chemotherapy of Malaria. Second edition. Geneva:, World Heal. Organ. Monogr. Ser. No. 27. (1986) 9–19.

[61] S. Melato, D. Prosperi, P. Coghi, N. Basilico, D. Monti, A Combinatorial Approach to 2,4,6-Trisubstituted Triazines with Potent Antimalarial Activity: Combining Conventional Synthesis and Microwave-Assistance, ChemMedChem. 3 (2008) 873–876.

[62] B. Klenke, M.P. Barrett, R. Brun, I.H. Gilbert, Antiplasmodial activity of a series of 1,3,5-triazine-substituted polyamines, J. Antimicrob. Chemother. 52 (2003) 290–293.

[63] A. V. MYATT, T. HERNANDEZ, G.R. COATNEY, Studies in human malaria. XXXIII. The toxicity of pyrimethamine (daraprim) in man., Am. J. Trop. Med. Hyg. 2 (1953) 788–794.

[64] N. Nirmalan, P. Wang, P.F.G. Sims, J.E. Hyde, Transcriptional analysis of genes encoding enzymes of the folate pathway in the human malaria parasite Plasmodium falciparum., Mol. Microbiol. 46 (2002) 179–90. http://www.ncbi.nlm.nih.gov/pubmed/12366841.

[65] I.B. Müller, J.E. Hyde, Folate metabolism in human malaria parasites—75 years on, Mol. Biochem. Parasitol. 188 (2013) 63–77.

[66] MPMP - Home, (n.d.). http://mpmp.huji.ac.il/.

[67] Y. Yuthavong, S. Kamchonwongpaisan, U. Leartsakulpanich, P. Chitnumsub, Folate metabolism as a source of molecular targets for antimalarials, Future Microbiol. 1 (2006) 113–125.

[68] J.H. MEUWISSEN, Resistance of Plasmodium falciparum to pyrimethamine and proguanil in Netherlands New Guinea., Am. J. Trop. Med. Hyg. 10 (1961) 135–9. http://www.ncbi.nlm.nih.gov/pubmed/13769941.

[69] R.W. BURGESS, M.D. YOUNG, The development of pyrimethamine resistance by Plasmodium falciparum., Bull. World Health Organ. 20 (1959) 37–46. http://www.ncbi.nlm.nih.gov/pubmed/13638788.

[70] D.F. CLYDE, G.T. SHUTE, Resistance of East African varieties of Plasmodium falciparum to pyrimethamine., Trans. R. Soc. Trop. Med. Hyg. 48 (1954) 495–500. http://www.ncbi.nlm.nih.gov/pubmed/13216953.

[71] J.E. Hyde, Drug-resistant malaria - an insight., FEBS J. 274 (2007) 4688–98.

[72] V.N. Hawkins, H. Joshi, K. Rungsihirunrat, K. Na-Bangchang, C.H. Sibley, Antifolates can have a role in the treatment of Plasmodium vivax, Trends Parasitol. 23 (2007) 213–

222.

[73] T. Jelinek, A.M. Rønn, M.M. Lemnge, J. Curtis, J. Mhina, M.T. Duraisingh, I.C. Bygbjerg, D.C. Warhurst, Polymorphisms in the dihydrofolate reductase (DHFR) and dihydropteroate synthetase (DHPS) genes of Plasmodium falciparum and in vivo resistance to sulphadoxine/pyrimethamine in isolates from Tanzania., Trop. Med. Int. Health. 3 (1998) 605–9. http://www.ncbi.nlm.nih.gov/pubmed/9735930.

[74] T. Jelinek, P. Schelbert, T. Löscher, D. Eichenlaub, Quinine resistant falciparum malaria acquired in east Africa., Trop. Med. Parasitol. 46 (1995) 38–40. http://www.ncbi.nlm.nih.gov/pubmed/7631126.

[75] N. Berens, B. Schwoebel, S. Jordan, V. Vanisaveth, R. Phetsouvanh, E.-M. Christophel, S. Phompida, T. Jelinek, Plasmodium falciparum: correlation of in vivo resistance to chloroquine and antifolates with genetic polymorphisms in isolates from the south of Lao PDR., Trop. Med. Int. Health. 8 (2003) 775–82. http://www.ncbi.nlm.nih.gov/pubmed/12950663.

[76] W. Sirawaraporn, T. Sathitkul, R. Sirawaraporn, Y. Yuthavong, D. V Santi, Antifolate-resistant mutants of Plasmodium falciparum dihydrofolate reductase., Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 1124–9. http://www.ncbi.nlm.nih.gov/pubmed/9037017.

[77] P. Olliaro, T. Wells, The Global Portfolio of New Antimalarial Medicines Under Development, Clin. Pharmacol. Ther. 85 (2009) 584–595.

[78] F. Coslédan, L. Fraisse, A. Pellet, F. Guillou, B. Mordmüller, P.G. Kremsner, A. Moreno, D. Mazier, J.-P. Maffrand, B. Meunier, Selection of a trioxaquine as an antimalarial drug candidate., Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 17579–84.

[79] R. Morphy, Z. Rankovic, Designed Multiple Ligands. An Emerging Drug Discovery Paradigm, J. Med. Chem. 48 (2005) 6523–6543.

[80] A. Mishra, H. Batchu, K. Srivastava, P. Singh, P.K. Shukla, S. Batra, Synthesis and evaluation of new diaryl ether and quinoline hybrids as potential antiplasmodial and antimicrobial agents, Bioorg. Med. Chem. Lett. 24 (2014) 1719–1723.

[81] N. Boechat, M. de L.G. Ferreira, L.C.S. Pinheiro, A.M.L. Jesus, M.M.M. Leite, C.C.S. Júnior, A.C.C. Aguiar, I.M. de Andrade, A.U. Krettli, New Compounds Hybrids 1 *H* -1,2,3-Triazole-Quinoline Against *Plasmodium falciparum*, Chem. Biol. Drug Des. 84 (2014) 325–332.

[82] M.D. Hastings, C.H. Sibley, Pyrimethamine and WR99210 exert opposing selection on dihydrofolate reductase from Plasmodium vivax, Proc. Natl. Acad. Sci. 99 (2002) 13137–13141.

[83] A. Kumar, K. Srivastava, S. Raja Kumar, S.K. Puri, P.M.S. Chauhan, Synthesis and bioevaluation of hybrid 4-aminoquinoline triazines as a new class of antimalarial agents, Bioorganic Med. Chem. Lett. 18 (2008) 6530–6533.

[84] A. Kumar, K. Srivastava, S. Raja Kumar, M.I. Siddiqi, S.K. Puri, J.K. Sexana, P.M.S. Chauhan, 4-Anilinoquinoline triazines: A novel class of hybrid antimalarial agents, Eur. J. Med. Chem. 46 (2011) 676–690.

[85] N. Sunduru, K. Srivastava, S. Rajakumar, S.K. Puri, J.K. Saxena, P.M.S. Chauhan, Synthesis of novel thiourea, thiazolidinedione and thioparabanic acid derivatives of 4-aminoquinoline as potent antimalarials, Bioorganic Med. Chem. Lett. 19 (2009) 2570–2573.

[86] J. Yuvaniyama, P. Chitnumsub, S. Kamchonwongpaisan, J. Vanichtanankul, W. Sirawaraporn, P. Taylor, M.D. Walkinshaw, Y. Yuthavong, Insights into antifolate resistance from malarial DHFR-TS structures, Nat. Struct. Biol. 10 (2003) 357–365.

[87] D.J. Bzik, W.B. Li, T. Horii, J. Inselburg, Molecular cloning and sequence analysis of the Plasmodium falciparum dihydrofolate reductase-thymidylate synthase gene., Proc. Natl. Acad. Sci. U. S. A. 84 (1987) 8360–4.

[88] J.E. Hyde, The dihydrofolate reductase-thymidylate synthetase gene in the drug resistance of malaria parasites, Pharmacol Ther. 48 (1990) 45–59. http://www.ncbi.nlm.nih.gov/pubmed/2274577.

[89] J. Yuvaniyama, P. Chitnumsub, S. Kamchonwongpaisan, J. Vanichtanankul, W. Sirawaraporn, P. Taylor, M.D. Walkinshaw, Y. Yuthavong, Insights into antifolate resistance from malarial DHFR-TS structures, Nat. Struct. Mol. Biol. 10 (2003) 357–365.

[90] A.F. Cowman, M.J. Morry, B.A. Biggs, G.A. Cross, S.J. Foote, Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of Plasmodium falciparum., Proc. Natl. Acad. Sci. U. S. A. 85 (1988) 9109–13. http://www.ncbi.nlm.nih.gov/pubmed/3057499.

Chapitre II: Techniques de modélisation moléculaire

I. Introduction

Au cours des dernières décennies, la simulation numérique est devenue un outil important dans divers domaines comme la chimie, le génie mécanique, la physique et les matériaux. En particulier, les simulations informatiques qui tiennent compte des structures moléculaires [1,2]. A l'heure actuelle, la simulation par ordinateur est l'un des principaux outils permettant d'identifier les régularités de comportement des systèmes moléculaires sous diverses influences externes [3]. Ce comportement moléculaire est analysé par les méthodes de modélisation moléculaire qui servent non seulement à le prédire, mais aussi à étudier la structure, la dynamique, les propriétés de surface et la thermodynamique des systèmes inorganiques, biologiques et polymères. La modélisation moléculaire a été ainsi introduite en tant que méthodologie précieuse pour la recherche scientifique, qui fournit des outils utiles pour l'analyse et l'estimation des paramètres physico-chimiques et l'activité biologique [4].

II. Méthodes théoriques de calcul

II.1 Mécanique moléculaire

La mécanique moléculaire calcule l'énergie d'une molécule, comme l'énergie stérique, l'énergie due à la géométrie ou à la conformation d'une molécule.

II.1.1 Champ de force

Les méthodes de la mécanique moléculaire (MM) modélisent le système moléculaire en mécanique classique et peuvent généralement être utilisées dans l'optimisation de la géométrie moléculaire et la simulation de la dynamique moléculaire. Les atomes sont traités comme des points de masse classiques liés par des interactions interatomiques, qui comprennent des termes liés relatifs aux atomes liés par des liaisons covalentes, et des termes non liés décrivant les interactions électrostatiques à longue distance et van der Waals. L'énergie potentielle totale du système en mécanique moléculaire est calculée comme la somme des termes énergétiques individuels définis pour un champ de forces. Les paramètres dans les champs de force sont optimisés en fonction des résultats de calculs de la mécanique quantique (QM) expérimentaux ou de haut niveau [5].

Dans certains champs de force populaires (par ex. AMBER [6], CHARMM [7], GROMOS [8], CFF [9], MMFF [10], MM3 [11], MM4 [12], UFF [13]), l'énergie potentielle est donnée dans la formule analytique.

L'énergie potentielle en champ de force MM peut être écrite:

$$E_{MM} = E_{bond} + E_{ang} + E_{tors} + E_{vdw} + E_{ele} + E_{cross}$$

Où E_{bond} est l'énergie de liaison entre les atomes ou l'énergie d'élongation. E_{ang} est l'énergie de déformation angulaire entre deux atomes liés à un troisième atome. E_{tors} est l'énergie associée au mouvement de rotation des liaisons. E_{vdw} décrit la répulsion et la dispersion d'énergie combinées des interactions de Van der Waals entre deux atomes. E_{ele} représente l'énergie dans le système due aux interactions électrostatiques. E_{cross} représente un effet hybride de couplage des termes énergétiques.

II.1.1.1 Exemples de champs de force

De nombreux champs de force ont déjà été développés. Pettersson et Liljefors (1996) ont publié une comparaison complète de plusieurs champs de force ciblant le calcul des énergies conformationnelles des molécules organiques [14].

Le tableau 1 présente une liste de certains champs de force MM couramment utilisés qui ont été conçus pour les molécules organiques et qui sont utilisés dans plus d'un logiciel. Il y a généralement des différences négligeables dans la mise en œuvre, ce qui entraîne de légères différences dans les résultats d'un champ de force MM à l'autre.

Tableau 1 : Les champs de force MM les plus utilisés.

Champ de force	La description
MM2/MM3/MM4	Il a été développé par Allinger et ses collaborateurs [15], MM2 est principalement conçu pour les hydrocarbures. Il a été étendu par de nombreux groupes fonctionnels (alcanes, alcènes et alcynes non conjugués, les composés carbonylés, les sulfures, les amines…), pour couvrir presque tous les types de petites molécules organiques. La méthode MM3 qui apporte des améliorations significatives de la forme fonctionnelle est probablement l'une des méthodes les plus précises de modélisation des hydrocarbures. Il a également été étendu aux amides, polypeptides et protéines.
MM+	Le champ de force MM+ est une extension du champ de force MM2, avec l'ajout de quelques paramètres géométriques[16]**.** MM+ est un champ de force robuste, il a l'aptitude de prendre en considération les paramètres négligés dans d'autres champs de force et peut donc s'appliquer pour des molécules plus complexes tels que les composés inorganiques[17]**.**
AMBER	AMBER «Assisted Model Building with Energy Refinement» est mis au point par Kollman [18]. Un des champs de force les plus populaires pour la modélisation des protéines et des acides nucléiques. Il a été utilisé pour les polymères comme pour d'autres petites molécules.
CHARMM	CHARMM « Chemistry at HARvard Macromolecular Mechanics» est le nom d'un ensemble de champs de force largement utilisé en dynamique moléculaire ainsi que le nom du programme de simulation et d'analyses de mécanique moléculaire associé , Il a été développé par Brooks et Karplus[19], utilise une fonction d'énergie empirique. Il a été appliqué pour les systèmes macromoléculaires et molécules biologiques (protéines, acides nucléiques,…).
UFF	UFF «Universal Force Field » Il s'agit d'un ensemble de formes fonctionnelles simples et de paramètres utilisés pour modéliser la structure, le mouvement et l'interaction des molécules contenant n'importe quelle combinaison d'éléments dans le tableau périodique.

MMFF	MMFF «Merck Molecular Force Field» Il est basé principalement sur des calculs quantiques mécaniques de la surface énergétique et est capable de traiter tous les groupes fonctionnels d'intérêt dans la conception pharmaceutique.

II.1.2 Énergie d'interaction entre atomes liés

II.1.2.1 Energie d'élongation

L'élongation est le changement de la longueur d'une liaison covalente entre deux atomes, comme C-H ou C-C. Cette déformation est régie en première approximation par la loi de "Hooke" d'élongation des ressorts. On peut ainsi lui associer une énergie d'élongation de la forme:

$$E(L) = \sum_{L} (L - L_0)^2$$

Où $\mathbf{K_r}$: est la constante d'élongation ou constante de Hooke

$\mathbf{L_o}$: la longueur de la liaison de référence.

L : la longueur de la liaison dans le modèle.

II.1.2.2 Energie de flexion

La déformation est le changement de l'angle entre deux liaisons covalentes, ce phénomène est régi par une énergie de flexion qui peut s'exprimer sous les mêmes formes que précédemment à savoir, pour la plus simple :

$$E(\theta) = \sum_{\theta} (\theta - \theta_0)^2$$

K_θ : constante de flexion.

θ : angle de valence de référence.

θ_0 : angle de valence dans la molécule.

II.1.2.3 Energie de torsion

En géométrie, un angle dièdre ou angle de torsion est l'angle entre deux plans, où chaque plan est défini par trois atomes liés, l'énergie potentielle est exprimée en fonction des angles

dièdres ou angles de torsion. L'évaluation de cette énergie $E(\varphi)$ se fait par une fonction développée en série de Fourier.

$$E(\varphi) = \sum_{\varphi} \frac{V_n}{2}(1 + \cos(n\varphi))$$

L'expression de la contribution de chaque angle dièdre est un (ou plusieurs) terme(s) d'une fonction développée en série de Fourier. L'entier n est l'ordre de la série de Fourier pris en compte pour l'angle dièdre considéré, V_n est la constante du potentiel de l'énergie de torsion.

II.1.3 Énergie d'interaction entre atomes non liés

Les interactions non liées agissent entre des atomes qui ne sont pas liés par des liaisons covalentes.

II.1.3.1 Energie de Van der Waals

L'interaction non liée la plus simple est l'interaction de van der Waals (VDW). Les forces de VDW sont des forces attractives ou répulsives entre entités moléculaires (ou entre groupes au sein de la même entité moléculaire) autres que celles dues à la formation de liaisons ou à l'interaction électrostatique d'ions ou de groupes ioniques entre eux ou avec des molécules neutres [20,21]. Elles incluent les interactions attractives ou répulsives entre dipôles permanents (les atomes, les molécules et les surfaces), ce qui comprend: la force entre deux molécules polaires (force de Keesom) et la force entre une molécule polaire et un dipôle induit (force de Debye), également la force entre deux dipôles induits (force de dispersion de Londres).

Le potentiel le plus connu est celui de Lennard-Jones appelé potentiel 6-12 :

$$E_{ij} = \sum_i \sum_j -\frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}}$$

Le terme en puissance 12 est répulsif à très courte distance (terme de Pauli), alors que le terme en puissance de 6 est attractif (forces dispersives) à grande distance.

r_{ij} : distance entre les deux atomes non liés.

A_{ij} et B_{ij} sont des constantes de VDW.

II.1.3.2 Liaisons hydrogènes

Les liaisons hydrogènes sont un cas particulier des liaisons de van der Waas. Elles sont en partie de nature dipôle-dipôle mais ont une énergie dix fois plus importante que celle observée pour une interaction de VDW classique. Elles renferment donc un type de liaison chimique non covalente de type dipôle-dipôle entre 2 molécules ou entre 2 groupements d'une molécule, qui implique un arrangement du type : A-H…B où A et B sont généralement fluor, oxygène ou azote. Elle consiste essentiellement dans l'interaction entre deux molécules :

- 1 molécule possédant un atome donneur d'électrons (O, N, F)
- 1 molécule possédant un atome H accepteur d'électrons (OH, NH_2)

Les liaisons hydrogènes sont très importantes en chimie et biochimie, car, elles assurent la stabilité de nombreux biopolymères, protéines et acides nucléiques.

II.1.3.3 Interactions électrostatiques

La manipulation de l'électrostatique est un peu plus compliquée. Les interactions électrostatiques sont entre les cations et anions, qui peuvent être attractifs ou répulsifs selon les signes des charges.

$$E_{elect} = \sum \frac{q_i q_j}{D r_{ij}}$$

D : constante diélectrique locale du milieu

$q_i q_j$: charges partielles atomiques des atomes i,j

r_{ij} : distance interatomique.

II.2 Théorie de la Fonctionnelle de la Densité (DFT)

II.2.1 introduction

La théorie de la fonctionnelle de la densité (DFT) est une méthode qui repose sur la densité électronique pour calculer la structure électronique des systèmes à plusieurs électrons au lieu de la fonction d'onde [22,23].

L'équation de Schrödinger est l'équation clé de la mécanique quantique. Cette équation différentielle partielle de second ordre détermine la forme spatiale et l'évolution temporelle d'une fonction ondulatoire dans un potentiel donné et pour des conditions limites données.

Le but de la plupart des approches chimiques quantiques qui sont utilisées pour des problèmes de structure électronique a une solution approximative de l'équation de Schrödinger non relativiste et indépendante du temps pour les électrons avec des noyaux fixes(Eq. 1.1):

$$H\psi(\vec{r}_1,\vec{r}_2,....,\sigma_1,\sigma_2,....) = E\psi(\vec{r}_1,\vec{r}_2,....,\sigma_1,\sigma_2,....)\,(1.1)$$

Avec H est l'opérateur Hamiltonien et E est l'énergie totale d'un système spécifique ($\vec{r}_1,\vec{r}_2$,… représentent les coordonnées des électrons et σ_1,σ_2 ,….indiquent leurs coordonnées de spin), L'équation de Schrödinger est une équation aux valeurs propres En avec des fonctions propres $\psi_n(\vec{r}_1,\vec{r}_2,....,\sigma_1,\sigma_2,....)$. Les valeurs propres de l'équation de Schrödinger En sont discrètes pour les états liés des électrons et sont des formes continues pour les états non liés. La plus basse valeur propre représente l'énergie de l'état fondamental alors que les autres valeurs propres plus grandes sont les énergies des états excités.

La connaissance de $\psi_n(\vec{r})$ et E_n implique la détermination de tous les paramètres physiques pertinents du système. Cependant, la résolution de l'équation de Schrödinger est une tache compliquée. Par exemple, l'équation de Schrödinger pour le benzène implique les coordonnées de 42 électrons, c'est-à-dire 126 degrés de liberté spatiale et 42 coordonnées de spin d'électrons. En raison du principe de Pauli, $\psi_{\text{benzène}}$ est une fonction asymétrique dans les coordonnées des électrons, et il a 126 degrés de liberté spatiale et 42 coordonnées de spin; par conséquent, il est essentiellement impossible de la résoudre complètement avec précision. Néanmoins, la chimie quantique cherche à faire des prédictions avec une précision chimique.

Après la proposition de l'équation de Schrödinger, Thomas et Fermi ont indépendamment développé un modèle pour décrire la densité approximative d'électrons $\rho(\vec{r})$ et l'énergie d'état fondamental E d'un grand système ou d'une molécule avec un grand nombre d'électrons [24–26]. Puisque la théorie est formulée à base de la densité électronique, elle est considérée comme un précurseur de la DFT. Dans cette théorie, on écrit l'énergie de l'état fondamental d'un atome comme:

$$E = T + U_{eN} + U_{ee} \qquad (1.2)$$

$$= C_F \int \rho(\vec{r})^{5/3} dr + \int \rho(\vec{r}) V_N(\vec{r}) dr + \frac{1}{2} e^2 \int \frac{\rho(\vec{r})\rho(\vec{r}')}{\left|\vec{r} - \vec{r}'\right|} dr dr'$$

Dans laquelle, l'énergie totale d'un système ayant N électrons est la somme des trois termes T, U_{eN} et $U_{ee.}$ Le premier terme représente l'énergie cinétique des électrons où C_F est une constante et $\rho(\vec{r})$ est la densité d'électrons à un point donné $\vec{r}$ dans un espace tridimensionnel. Le deuxième terme représente l'énergie potentielle des électrons due au noyau, où $V_N(\vec{r})$ est le potentiel d'attraction noyau-électron à la position $\vec{r}$. Le troisième terme est l'énergie potentielle due à la répulsion des électrons. Même si ce modèle pousse les chercheurs à prendre l'attention à l'approche de la densité des électrons, il reste incapable de prédire l'optimisation exacte des molécules stables [27].

La théorie de la fonctionnelle de la densité que nous connaissons aujourd'hui est née avec Hohenberg et Kohn comme il est rapporté dans leur travail publié en 1964 [28]. Ils ont montré que dans l'approximation de Born-Oppenheimer que tous systèmes atomiques ou moléculaires sont complètement déterminés par la densité des électrons $\rho(\vec{r})$. Dans le paragraphe suivant, nous allons présenter brièvement les deux théorèmes de Hohenberg-Kohn:

- Premier théorème de Hohenberg-Kohn:

Hohenberg et Kohn ont prouvé que l'énergie électronique de l'état fondamental E0, et toutes les autres propriétés électroniques de l'état fondamental sont uniquement déterminées par la connaissance de la densité électronique $\rho(\vec{r})$ en chaque point rdr du volume moléculaire. E0 est une fonctionnelle de $\rho(\vec{r})$ et est représentée par $E_0[\rho]$ avec $\rho = \rho(\vec{r})$. En pratique, on travaille avec la fonctionnelle E [ρ] écrite sous une forme qui fait intervenir la contribution cinétique T0 d'un système d'électrons sans interaction entre eux et la contribution coulombienne d'un système classique :

$$E[\rho] = T[\rho] + \int V_{ext}(\bar{r})\rho(\bar{r})dr + \frac{1}{2}\int \frac{\rho(\bar{r})\rho(\bar{r}')}{\left|\bar{r}-\bar{r}'\right|} d^3r d^3r' + E_{xc}[\rho] \qquad (1.3)$$

Pour tout système de particules en interaction avec eux dans un potentiel externe $V_{ext}(\vec{r})$, la densité est déterminée de façon unique (en d'autres termes, le potentiel externe est une fonction unique de la densité).

- Second théorème de Hohenberg-Kohn:

Il existe une fonctionnelle universelle E[ρ] exprimant l'énergie totale en fonction de la densité électronique $\rho(\vec{r})$, valide pour tout potentiel externe $V_{ext}(\vec{r})$. L'énergie de l'état fondamental

du système est la valeur qui minimise cette fonctionnelle et la densité $\rho_0(\vec{r})$ correspondante est la densité exacte de l'état fondamental.

Donc on peut appliquer le principe variationnel.

$$\frac{\delta E[\rho]}{\delta \rho} = 0 \qquad \text{avec} \qquad \int \rho d^3\vec{r} = N$$

L'énergie de l'état fondamental correspond au minimum de la fonctionnelle de la densité.

$$E_0 = minE[\rho(\vec{r})]$$

Trouver $\rho(\vec{r})$ de façon à ce que E soit minimal revient à un problème d'optimisation avec contrainte. Le calcul de la dérivée, de la fonctionnelle E [ρ], donne l'équation fondamentale de la DFT, qui est une équation de type Euler-Lagrange :

$$\frac{\delta E[\rho]}{\delta \rho} = \frac{\delta T[\rho]}{\delta P} + \int \frac{\rho(\vec{r}')}{\left|\vec{r} - \vec{r}'\right|} dr' + v(\vec{r}) + \frac{\delta E_{xc}[\delta]}{\delta \rho} = \mu$$

Les théorèmes de Hohenberg-Kohn n'ont pas une grande valeur pratique. Ils déclarent fondamentalement qu'une fonctionnelle exacte de la densité existe cependant ils ne disent rien au sujet de la façon de l'obtenir. Dans cette expression (1.3) les formules de l'énergie cinétique et d'échange - corrélation restent toujours inconnues. Puisque la contribution de premier terme à l'énergie totale est importante il ne peut pas être exprimé par une expression approximative contrairement au terme

II.2.2 Approximation de Kohn-Sham

Pour pallier à ce problème concernant le terme exprimant l'énergie cinétique et l'échange-corrélation, l'approche proposée par Kohn et Sham en 1965 [29] exprime la densité d'un système fictif sans interaction par un seul déterminant de Slater (Eq. 1.4) :

$$\phi_{SD} = \frac{1}{\sqrt{N!}} \det\left|\chi_1(\vec{x_1})\chi_2(\vec{x_2})....\chi_N(\vec{x_N})\right| \qquad (1.4)$$

Le déterminant de Slater ϕ_{SD} est une approximation de la fonction d'onde N-électronique.

χ_N est le i^{ème} spin-orbitale de type α, et $\bar{x}_n$ indique spin-orbitale de type β. L'énergie totale du système peut s'écrire comme suit :

$$E[\rho] = E_{Ne}[\rho] + T[\rho] + J[\rho] + E_{xc}[\rho] \qquad (1.5)$$

Dans laquelle E_{Ne}, T et J sont respectivement l'attraction électron-noyau, l'énergie cinétique et répulsion électronique. E_{xc} est l'énergie d'échange corrélation, qui représente toutes les parties restantes de l'énergie. En se basant sur le minimum variationnel de l'énergie totale par rapport à la densité des électrons r, on obtient l'équation de Kohn-Sham (Eq. 1.6):

$$f^{KS}\chi_i = \varepsilon_i \chi_i \qquad (1.6)$$

où f^{KS} est appelé l'opérateur Fock, qui peut être écrit comme suit (Eq. 1.7):

$$f^{KS}(r_1) = -\sum_{A}^{M} \frac{Z_A}{r_{iA}} - \frac{1}{2}\nabla_i^2 + \int \frac{\rho(r_2)}{r_{12}} dr_2 + V_{xc}(r_i) \qquad (1.7)$$

où les quatre termes de l'équation 1.7 sont les équivalents de ceux de l'équation 1.4. Le potentiel d'échange-corrélation est défini comme la dérivée de l'énergie d'échange-corrélation avec respectivement la densité d'électrons (Eq. 1.8):

$$V_{xc} = \frac{\delta E_{xc}}{\delta \rho} \qquad (1.8)$$

Nous pouvons résoudre l'équation 1.5 et 1.7 par la méthode du champ auto consistant (SCF), où nous introduisons une supposition initiale pour le terme de densité d'électrons en actualisant tous les autres termes itérativement jusqu' à ce que la convergence soit atteinte. On peut alors obtenir l'énergie totale en intégrant l'équation 1.5

Les équations de Kohn-Sham sont exactes en principe. Si nous connaissons la forme exacte d'E_{xc} et de V_{xc}, nous pouvons obtenir l'énergie exacte avec une approximation dans les équations de Kohn-Sham. L'approximation entre seulement lorsque nous voulons la forme explicite de E_{xc} et V_{xc}. La fonctionnelle de densité a été développé principalement en se concentrant sur l'obtention de bonnes formes fonctionnelles d'énergie de corrélation d'échange, et à partir de celles-ci, on peut obtenir le potentiel de corrélation d'échange.

II.2.3 Fonctionnelle de densité de première génération (LDA)

II.2.3.1 Approximation de la densité de spin locale (LSDA)

En 1965, Kohn et Sham ont introduit l'approximation de la densité locale (LDA) pour l'énergie d'échange corrélation. Dans cette approximation, on suppose qu'à chaque point de l'espace, l'énergie d'échange-corrélation par particule est donnée par sa valeur pour un milieu gazeux homogène (également connu sous le nom gaz d'électron uniforme ou UEG)

$E_{xc}\left[\rho(\vec{r})\right]$. L'UEG est défini comme un grand nombre N d'électrons dans un cube de volume V, à travers lequel il y a une répartition uniforme de la charge positive suffisante pour rendre le système neutre. Ensuite elle est définie comme la limite $N \to \infty, V \to \infty$, avec la densité $\rho = N/V$ reste fini. Les approximations locales de la densité de spin (LSDA) sont celles de classe des énergies de corrélation et d'échange qui ne dépendent que des densités de spin ρ_α et ρ_β [30–32].

$$E_{XC}^{LSDA}\left[\rho_\alpha, \rho_\beta\right] = \int d\vec{r}\rho(\vec{r})\varepsilon_{XC}(\rho_\alpha, \rho_\beta) \qquad (1.9)$$

Dans l'équation (1.9), $\varepsilon_{XC}(\rho_\alpha, \rho_\beta)$ représente l'énergie de corrélation d'échange de l'électron gazeux uniforme par électron à un point donné de l'espace [32] (10). La LSDA est très simple et a été largement utilisée en physique des solides, avec de bons résultats pour les géométries et les fréquences de vibration solide [33].

Cependant la densité moléculaire n'est pas uniforme, et apporte des changements rapides. Par conséquent, les énergies moléculaires LSDA ont des erreurs systémiques, y compris la surestimation des énergies de liaison et de grandes erreurs absolues pour les énergies d'atomisation.

II.2.3.2 Approximation du gradient généralisée (GGA)

Dans les années 1980, la GGA a été développée pour rapprocher l'énergie de corrélation d'échange; une GGA ne dépend pas seulement de la densité mais aussi du gradient de la densité.

$$E_{XC}^{GGA}\left[\rho(\vec{r})\right] = \int d\vec{r} f\left[\rho(\vec{r}), \nabla\rho(\vec{r})\right] \qquad (1.10)$$

Cette approximation est basée sur l'expansion du gradient du trou d'échange qui est basé sur des valeurs limites de l'espace réel choisies pour garantir que le trou est négatif partout et représente un déficit d'un électron [34]. En comparaison à la LSDA,

Les GGA ont une tendance à améliorer les énergies de liaison et les énergies d'atomisation [35]. La révolution dans la vulgarisation du KS-DFT a été initiée par l'échange GGA proposé par Becke en 1988 [36] et par les fonctions de corrélation de Perdew (P86) [37], Perdew et Wang (PW91) [38], et Lee, Yang, et Parr (LYP) [39]. La fonction de corrélation LYP contient des paramètres adaptés à l'atome d'hélium et permet d'annuler exactement l'erreur d'auto-interaction dans les systèmes à un électron. Lorsque l'échange est séparé de la corrélation, la

mise à l'échelle correcte de l'échange sous une expansion uniforme de toutes les coordonnées n'est réalisée que si la densité d'énergie d'échange est séparable [40], c'est-à-dire, si c'est une somme sur spins σ d'une fonction de ρ_α multiplie à une fonction de s_σ^2, où s_σ est le gradient sans dimension proportionnel à $|\rho_\sigma| / \rho_\sigma^{4/3}$; les GGA standards pour l'échange satisfont tous à cette condition. La fonction d'échange B88 a le comportement asymptotique correct à longue portée pour la densité d'énergie d'échange, et elle incorpore un paramètre optimisé en s'adaptant aux énergies d'échange Hartree-Fock de six atomes de gaz noble (de He à Rn). Les formes fonctionnelles populaires pour les fonctions d'échange GGA sont des extensions rationnelles de fonctions dans le gradient réduit [41,42] et les polynômes de ces fonctions [43].

Il faut encore citer les fonctionnelles dites «hybrides», basées sur le formalisme de la connexion adiabatique [44,45]. Le principe émerge de la question demandant s'il est possible d'utiliser l'échange de Hartree-Fock dans le formalisme de Kohn-Sham. La formule de la connexion adiabatique justifie théoriquement la détermination de l'énergie d'échange HF à partir de l'énergie des orbitales Kohn-Sham. L'utilisation de la partie d'échange HF associée aux fonctionnelles GGA fournit des résultats comparables à ceux de l'approximation des gradients généralisés. La première fonctionnelle de ce type a été proposée par Becke, et contient 50% d'échange HF, c'est la fonctionnelle «half and half» [46]. Elle présentait l'inconvénient de contenir une trop forte proportion d'échange HF. Pour y réduire, une nouvelle fonctionnelle a été proposée (contenant 20%d'échange HF), actuellement la plus utilisée et connue sous l'acronyme B3LYP [47]. Celle-ci est une fonctionnelle à trois paramètres combinant les fonctionnelles d'échange local, d'échange de Becke et d'échange HF, avec les fonctionnelles de corrélation locale (VWN) et corrigée du gradient de Lee, Yang et Parr comme le montre l'expression suivante :

$$E_{xc}^{B3LYP} = E_{xc}^{LSDA} + a_0(E_x^{HF} - \varepsilon_x^D) + a_x(E_x^{B88} - \varepsilon_x^D) + a_c(E_c^{LYP} - E_c^{VWN})$$

où E_X^{B88} est la fonctionnelle d'échange, aussi de type gradient corrigé de Becke. Les valeurs suivantes pour les trois paramètres a0=0.20, ax=0.72 et ac=0.81 ont été déterminées par optimisation des énergies d'atomisation, énergies d'ionisation, affinités électroniques ou protoniques d'un ensemble de molécules tests. Ces paramètres peuvent également être considérés comme des paramètres variationnels.

Récemment plusieurs fonctionnelles ont été développées à savoir : CAM-B3LYP [48], M05-2X [49], M06-2X [50] où l'échange de HF est compris entre 20% et 80%.

Enfin, de nouveaux travaux ont récemment été entrepris afin de développer de nouvelles fonctionnelles ab-initio sans paramètres. A l'heure actuelle, il n'existe qu'une seule fonctionnelle de ce type, élaborée par Perdew, Burke et Ernzerhof (PBE) [42], qui s'est montrée très efficace pour les calculs de géométries, de fréquences et d'énergies d'excitations électroniques.

Ces deux méthodes théoriques la DFT et la mécanique moléculaire MM ont été exploitées pour optimiser les molécules étudiées et calculer les descripteurs moléculaires, les descripteurs électroniques ont été optimisées en utilisant la base 6-31G avec la fonctionnelle B3LYP qui est une sorte de la méthode de théorie de la fonctionnelle de la densité DFT. Leur fiabilité a d'ailleurs été démontrée via différentes analyses théoriques, dédiées principalement à la réactivité chimique [51,52], Le calcul des descripteurs commence par le dessin des molécules dans le logiciel GaussView 3.09 puis l'optimisation de ces structures dans le programme Gaussian 03 [53], et les autres sont calculés en utilisant la mécanique moléculaire MM à l'aide de deux les logiciels ACD/ChemSketch program[54] et Chembiooffice [55].

II.3 Descripteurs moléculaires

Afin d'exploiter au maximum les informations contenues dans les structures moléculaires, celles-ci sont traduites en une série de grandeurs (en général scalaires) qui quantifient leurs caractéristiques physico-chimiques et structurelles. Ces grandeurs sont appelées descripteurs.

Depuis des décennies, de nombreux travaux ont été menés afin de développer des descripteurs capables décrire les structures moléculaires. aujourd'hui, il y a environ des milliers des descripteurs [56], ils peuvent être calculés ou empiriques. Ces derniers nécessitant d'être mesurés, les descripteurs calculés seront privilégiés [57].

Parmi les types de descripteurs que nous avons employés dans notre travail : les descripteurs constitutionnels, les descripteurs physico-chimiques, les descripteurs géométriques, les descripteurs topologiques, les descripteurs lipophiles, les descripteurs stériques et les descripteurs quantiques.

II.3.1 Descripteurs constitutionnels

Les descripteurs constitutionnels sont les descripteurs les plus simples et les plus couramment utilisés pour décrire la composition moléculaire d'un composé sans aucune information sur sa géométrie moléculaire.

Ce groupe de descripteurs comprend le poids moléculaire, le nombre de différents types d'atomes présents dans une molécule, le nombre d'anneaux, le nombre de différents types de liaisons, le nombre de différents groupes fonctionnels. Parmi les descripteurs constitutionnels que nous avons pris en compte dans ce travail :

Le poids moléculaire (MW): Il s'agit d'une variable importante dans les études QSAR.

L'optimisation des molécules de haute activité pour une cible biologique, presque souvent, va de pair avec l'augmentation des poids moléculaires. Ainsi, essayer de garder des poids moléculaires aussi bas que possible devrait être l'objectif pour établir un médicament [58].

II.3.2 Descripteurs géométriques

Les descripteurs géométriques d'une molécule sont évalués à partir des positions relatives de ses atomes dans l'espace, Ils nécessitent donc de connaître la relation entre la configuration géométrique moléculaire (3D) et les propriétés bioactives ou physico-chimiques. Pour l'interaction récepteur-ligand, la disposition géométrique des ligands détermine si les ligands peuvent entrer dans la poche réceptrice. Pour saisir ces caractéristiques moléculaires, les descripteurs géométriques sont conçus pour représenter un ensemble de fonctions mathématiques de la taille, de la forme, de la position et des propriétés moléculaires dans l'espace. Le calcul des descripteurs géométriques n'est pas aussi simple que celui des descripteurs constitutionnels, puisqu'il implique une intégration mathématique et une approximation numérique. L'interprétation de certains descripteurs exige une compréhension approfondie des théories fondamentales.

Nous présentons ci-dessous quelques descripteurs géométriques utilisés dans ce travail :

- Le volume moléculaire (MV) le volume occupé par une mole de la molécule considérée. il est défini par la formule suivante :

$$MV = \frac{MW}{D}$$

  Où : MW est le poids moléculaire et D la densité.

- La réfractivité moléculaire (MR) est le volume de la substance absorbée par mole de cette substance. Elle est définie par l'équation de Lorentz-Lorenz[59] :

$$\mathrm{MR} = \frac{\mathrm{MW}}{\mathrm{D}}\frac{\mathrm{n}^2 - 1}{\mathrm{n}^2 + 1}$$

  où : MW est le poids moléculaire, D la densité et n est l'indice de réfraction

- La densité (D), est une relation entre le poids moléculaire et le volume moléculaire.

II.3.3 Descripteurs topologiques

Les descripteurs topologiques, comme son nom l'indique, considèrent la topologie d'une molécule. Ces descripteurs sont des quantificateurs numériques de la topologie moléculaire [60]. Les caractéristiques structurales de la molécule telles que la taille, la forme, la symétrie, la dérivation et la cyclicité et peuvent également coder des informations chimiques concernant le type d'atome et la multiplicité des liaisons. Étant donné que les descripteurs topologiques considèrent la molécule comme un graphique mathématique, un certain nombre de ces descripteurs sont simplement des invariants graphiques divers ou d'autres fonctions du graphique moléculaire.

De nombreux indices topologiques ont été créés et utilisés dans les études QSAR. Leur calcul est simple, basé uniquement sur la structure moléculaire 2D et ne nécessitent pas d'analyse conformationnelle ou d'optimisation 3D de la structure. Le principal inconvénient des descripteurs topologiques est leur interprétation complexe et difficile.

II.3.3.1 Indice de Vienne, indice de Balaban

Ces deux indices sont calculés à partir de la matrice de distance d'un graphique moléculaire. L'élément ij de la matrice de distance est égal au nombre de liaisons entre les atomes i et j dans le chemin le plus court entre eux. Les éléments diagonaux de la matrice des distances sont égaux à zéro.

L'indice de Vienne (W) a été proposé par Wiener en 1947 [61]. Il est calculé comme la demi-somme des éléments de la matrice des distances:

$$W = \frac{1}{2}\sum\sum D_{ij}$$

Où Dij est l'élément ij de la matrice de distance de la molécule. La somme est faite sur tous les atomes i et j de la molécule.

L'indice Balaban (J) [62] est défini comme suit:

$$J = M / (\mu + 1)\sum (D_i D_j)^{-0.5}$$

où M est le nombre de liaisons dans la molécule

Di est la somme des éléments de la rangée i de la matrice de distance, correspondant au i^{ème} atome; $D_i = \sum D_{ij}$

Dij est l'élément ij de la matrice de distance de la molécule.

μ est le nombre cyclomatique du graphe moléculaire, égal au nombre minimum de bords qui doivent être enlevés avant qu'un graphe polycyclique ne devienne acyclique:

$$\mu = n - (N - 1)$$

où n est le nombre de bords de graphe (liaisons), N est le nombre de sommets de graphe (atomes).

L'indice de Balaban augmente avec la taille de la molécule, le degré de ramification et le degré d'insaturation. Sa valeur diminue fortement avec le nombre d'anneaux dans une molécule. [63].

II.3.4 Descripteurs physico-chimiques

le caractère hydrophile ou lipophile de la molécule généralement évalué à partir du coefficient de partage Octanol/eau représenté par le log P [64]. Enfaite, Le transport, le passage à travers les membranes et l'activité pharmacologique d'une molécule peuvent être conditionnés par son partage entre une phase lipidique et une phase aqueuse, c'est-à-dire son caractère hydrophile. Celui-ci peut être quantifié par le coefficient de partage Octanol-Eau, noté (log P), qui mesure la solubilité différentielle d'un soluté dans ces deux solvants non miscibles [65]. Le coefficient de partage octanol-eau est défini ainsi.

$$\log P = \log \frac{[Octanol]}{[H_2O]}$$

[Octanol] et [H_2O] sont les concentrations du soluté dans l'Octanol et l'eau.

Les composés qui ont les valeurs de log P > 0 sont dites lipophiles, et les composés qui ont les valeurs de log P <0 sont dites hydrophiles. Si le Log P est positif et très élevé, cela exprime le fait que la molécule est plus soluble dans l'Octanol que dans l'eau, ce qui reflète son caractère lipophile, et inversement, si le Log P est négatif cela signifie que la molécule est hydrophile. Un Log P nul signifie que la molécule est aussi soluble dans un solvant que dans l'autre.

II.3.5 Descripteurs quantiques

Pour décrire les structures moléculaires, des caractéristiques supplémentaires de la structure moléculaire peuvent encore être quantifiées. Ces descripteurs qui pour être calculés de manière fiable, nécessitent des calculs plus précisé sont regroupés ici sous l'étiquette de « descripteurs quantiques ».

Les approches de chimie quantique nous donnent accès à des informations supplémentaires telles que des données énergétiques, structurales, vibrationnelles et orbitalaires du système étudié [66].

Les principaux descripteurs calculés dans ce travail sont :

- **L'énergie HOMO** (E_{HOMO}), est le niveau d'énergie le plus élevé dans la molécule qui contient des électrons, il est directement lié au potentiel d'ionisation IP d'après l'approximation de Koopmans [67] [128] selon laquelle : $IP=-E_{HOMO.}$
- **L'énergie LUMO** (E_{LUMO}), est le niveau d'énergie inoccupé le plus bas dans la molécule, il est directement lié à son affinité électronique EA, (pour après l'approximation de Koopmans [67] [128]) $EA=-E_{LUMO.}$
- **L'électronégativité** (χ), est l'opposé du potentiel chimique qui mesure la tendance du nuage électronique à s'échapper de la molécule, c'est un paramètre global du système moléculaire égal à la pente de l'énergie en fonction du nombre d'électrons N à potentiel externe v(r) constant telle que définie par Parr et Mulliken [68,69]:

$$\chi = -\frac{(E_{LUMO} + E_{HOMO})}{2}$$

- **La dureté (**η**) et la mollesse (**S), peuvent être obtenues à partir de la première dérivée du potentiel chimique [70,71]:

$$\eta = \frac{1}{S} = \frac{E_{LUMO} - E_{HOMO}}{2}$$

- **L'indice d'électrophilicité**(ω), utilisée pour caractériser la capacité d'une molécule à engendrer un transfert d'électron, elle est calculée selon la formule suivante [72]:

$$\omega = \frac{\chi^2}{2\eta}$$

- **Le moment dipolaire** (μ) mesure la polarité de la molécule et décrit la séparation de charge dans une molécule où la densité d'électrons est partagée inégalement entre les atomes. L'existence d'un moment dipolaire dans une molécule a son origine dans la différence d'électronégativité entre les atomes [73]. La densité électronique est plus élevée au voisinage de l'atome le plus électronégatif. Ceci entraîne une dissymétrie dans la répartition des électrons de liaison. Ainsi, plus le moment dipolaire d'une molécule est élevé, plus la dissymétrie dans la molécule est importante.

- **L’énergie totale (E)**, pour une molécule isolée à l’état fondamental, l’énergie totale calculée peut être utilisée comme descripteur moléculaire quantique. Cette énergie approximative a été calculée pour une conformation optimisée de la géométrie la plus stable dont la structure d'énergie est minimale.

III. Relations quantitatives structures activités/propriétés

III.1 Introduction

Les relations quantitatives structure-activité/propriété (QSAR/QSPR) sont de plus en plus utilisées, de par le développement des moyens de calculs. La capacité de prédire a priori les propriétés des molécules, en particulier des médicaments et d'autres molécules bioactives, a suscité beaucoup d'intérêt pour la recherche, en particulier dans les industries pharmaceutiques.

En fait, les premiers développements dans le sens de telles méthodologies sont plutôt anciens. Dès 1868, Crum-Brown et Fraser ont postulé l’existence de relations entre les activités physiologiques et les structures chimiques [74] en reliant les changements d’activité biologique à des modifications structurales simples, ne disposant alors pas de moyens pour caractériser les structures chimiques en termes quantitatifs. les premiers travaux utilisant la méthodologie QSAR telle qu’employée à l’heure actuelle sont à attribuer aux contributions de Hansch [75] d’une part et de Free et Wilson [76] d’autre part.

Il s’agit de présenter ici le principe des modèles QSPR ainsi que ceux des différents outils employés pour leur mise en place et leur évaluation : bases de données expérimentales, descripteurs, outils d’analyse de données.

III.2 Principe du QSAR

Grâce à l'analyse de la forme des équations QSAR/QSPR, il est possible d'interpréter quelles caractéristiques structurales sont bénéfiques pour l'activité, ces informations peuvent être utilisées pour guider le processus de conception des produits pharmaceutiques et aider à fabriquer des composés ayant des activités plus appropriées ou d'autres propriétés.

L’établissement d’une relation quantitative Structure-Activité (QSAR) est une méthode par laquelle les propriétés numériques dérivées de la structure moléculaire sont mathématiquement liées à son activité. Dans sa forme la plus générale, une équation QSAR prend la forme :

$$\text{Activité} = f(X1, X2, \ldots .. Xn) \quad (1.1)$$

L'activité est une fonction des propriétés moléculaires ou des descripteurs Xi,…Xn qui codent des informations chimiques importantes sur la molécule. Il est donc important qu'une équation QSAR contienne des descripteurs qui englobent les propriétés qui ont une influence sur l'activité.

La prévision de l'activité biologique est l'un des objectifs majeurs de ces études. Toutefois, la capacité d'un médicament proposé d'agir contre une cible particulière n'est pas pertinente si la molécule ne peut atteindre la cible. L'indépendance d'une méthode prédictive par rapport aux informations observées expérimentalement permet également de prédire les propriétés chimiques sans avoir besoin de synthèse. Cependant, l'utilisation de données expérimentales observées dans la prédiction des propriétés est toujours avantageuse lorsqu'une propriété difficile à mesurer peut être reliée à une propriété facilement mesurable.

IV. Méthodes d'analyse statistiques

Au cours des 50 dernières années, les relations quantitatives Structure-Activité (QSAR) ont été considérées comme un outil polyvalent pour la prédiction des activités des molécules médicamenteuses et de médicaments, le but du développement d'un modèle QSAR étant de construire une relation (en utilisant des méthodes statistiques) entre les structures et les activités [77].Dans la plupart des cas, il est plus commode de considérer une relation linéaire entre l'activité/la propriété et les descripteurs [78], dont la méthode la plus largement utilisée est l'approche de régression linéaire multiple (MLR) proposée par Bartlett et Youd (1995) [79].

Dans ce travail, les méthodes de modélisation utilisées pour établir un modèle QSAR sont : la régression linéaire multiple (MLR), les réseaux de neurones artificiels (ANN), la validation croisée (CV), la technique de Y-Randomization, la validation externe avec les critères de Golbraikh et Tropsha.

VI.1 Régression linéaire multiple

La Régression linéaire multiple (MLR, pour Multiple Linear Regression) [80] est la méthode la plus simple et la plus communément employée pour le développement de modèles prédictifs qui permet d'étudier la relation entre une variable dépendante et deux ou plusieurs variables indépendantes. Lorsque la régression linéaire correspond à la valeur d'une ligne dans un espace bidimensionnel, la régression linéaire multiple correspond à une surface multidimensionnelle. Pour une série p de variables X indépendantes, la forme générale de la régression linéaire multiple est la suivante

$$Y = \varepsilon + a_1 X_1 + a_2 X_2 + \ldots\ldots + a_n X_n$$

où Y est la variable dépendante (à expliquer ou à prédire), ε est le modèle constant, X_1, ..., X_n sont les variables indépendantes avec les coefficients correspondants a_1, ..., a_n (pour les descripteurs 1 à n). Les valeurs des coefficients peuvent montrer le degré d'influence des descripteurs moléculaires individuels sur la propriété du composé. En outre, le signe du coefficient indique une contribution positive ou négative du descripteur donné à la propriété étudiée. Cependant, ces interprétations ne sont pas précises avec des descripteurs fortement corrélés, car ils peuvent avoir des valeurs incorrectes des coefficients. Par conséquent, les descripteurs moléculaires du modèle devraient être indépendants les uns des autres et le nombre de descripteurs entrant dans le modèle devrait être au moins cinq fois plus petit que le nombre de composés [81].

VI.1.2 Coefficients et tests statistiques de la régression.

VI.1.2.1 Coefficient de détermination (R^2):

Le coefficient de détermination (R^2) est une mesure de la façon dont la ligne de régression représente les données sur le diagramme de dispersion. R^2 est une mesure de l'ajustement du modèle de régression [82]. Le R^2 peut être utilisé pour déterminer la relation linéaire entre les valeurs mesurées et les valeurs estimées [83]. R^2 est un écart de 0 à 1.

Le coefficient de détermination R^2 est le rapport entre la somme des carrés expliquée et la somme totale des carrés.

$$R^2 = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

TSS est la somme des carrés totale ou variabilité totale:

$$TSS = \sum (Y_{obs} - \overline{Y})^2$$

ESS est la somme des carrés expliquée ou variabilité expliquée:

$$ESS = \sum (Y_{cal} - \overline{Y})^2$$

RSS est la somme des carrés des résidus ou variabilité des résidus:

$$RSS = \sum (Y_{obs} - Y_{cal})^2$$

VI.1.2.2 Coefficient de corrélation (R)

Coefficient de corrélation (R), est un coefficient de corrélation entre valeurs observées et prédites de la variable dépendante Y,

$$R = (ESS / TSS)^{0.5} = (1 - RSS / TSS)^{0.5}$$

VI.1.2.3 Erreur quadratique moyenne « RMSE » et Erreur de l'écart-type « SD »

L'erreur quadratique moyenne (RMSE) représente la racine carrée de la moyenne d'écart par rapport à la moyenne. Ceci est utilisé pour évaluer l'exactitude des valeurs calculées par le modèle.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - Y_i)^2}{n}}$$

Ou encore, l'erreur de l'écart-type (Sd).

$$Sd = \sqrt{\frac{\sum(\hat{y}_i - y_i)^2}{n - p - 1}}$$

Avec : $\hat{y}_i \; et \; y_i$ sont, respectivement, les valeurs observées et calculées de la variable dépendante ; n est le nombre des observations ; p est le nombre de variables indépendantes.

Une mesure similaire est l'écart-type (Sd). C'est presque identique à RMSE sauf que la moyenne des carrés est divisée par n-1. Ceci est fait pour éviter la sous-estimation de la variance de la population totale. Bien que pour les ensembles de données où n > 20, la différence entre les deux méthodes statistiques devient si faible que l'une ou l'autre peut être utilisée pour calculer la variation des données.

VI.1.2.4 Test d'hypothèses

Avant de tester des hypothèses dans le modèle de régression multiple, nous allons donner un aperçu général de la vérification des hypothèses.

Les tests d'hypothèse nous permettent de tirer des conclusions sur les paramètres de la population à partir des données d'un échantillon. Afin de tester une hypothèse en statistique, nous devons effectuer les étapes suivantes :

- Formuler une hypothèse nulle et une hypothèse alternative sur les paramètres de population,

- Construire une statistique pour tester l'hypothèse posée,
- Définissez une règle de décision pour rejeter ou non l'hypothèse nulle.

VI.1.2.4.1 Hypothèse nulle et hypothèse alternative

Il est très important de noter que la vérification des hypothèses porte toujours sur les paramètres de la population. Les tests d'hypothèses impliquent de prendre une décision, sur la base des données de l'échantillon, sur le rejet ou non de certaines restrictions satisfaites par le modèle de base supposé. Les restrictions que nous allons tester sont connues sous le nom d'hypothèse nulle, est notée par H_0. N'importe quelle autre hypothèse qui diffère de l'hypothèse H_0 s'appelle l'hypothèse alternative, est notée H_1. Dans la démarche de l'établissement de H_0 : la variable y est linéairement indépendante des variables x_j. Par contre pour H_1 : la variable y est expliquée linéairement par au moins une des variables x_j. Toutefois, cette décision est fondée sur une information partielle ou les résultats d'un échantillon. Il est donc statistiquement impossible de prendre la bonne décision à coup sûr. En pratique, on met en œuvre une démarche qui nous permettrait, à long terme de rejeter à tort une hypothèse nulle vraie dans une faible proportion de cas. La conclusion qui sera déduite des résultats de l'échantillon aura un caractère probabiliste : on ne pourra prendre une décision qu'en ayant conscience qu'il y a un certain risque qu'elle soit erronée. Ce risque nous est donné par le seuil de signification du test.

VI.1.2.5 Test de Student

Pour savoir si une variable joue un rôle explicatif dans un modèle, on effectue un test de Student ou test de significativité du coefficient de la variable explicative. Pour une régression multilinéaire, cela revient à supposer le coefficienta_i qui lui est associé comme nul.

La statistique du test est :

$$|t_i| = \left|\frac{(a_i)}{s(a_i)}\right| > t_{1-\frac{\alpha}{2}}^{n-p-2}$$

Avec : t_i est le *t-test* pour le descripteur « i », a_i est le coefficient associé au descripteur « i »

Dans le modèle, $s(a_i)$ est l'erreur type associée au descripteur « i », α est l'intervalle de confiance, n est le nombre des observations (les molécules), p est le nombre de variables indépendantes (les descripteurs).

Cette hypothèse est rejetée (avec un intervalle de confiance α) si le ratio t_i entre a_i et son erreur type $s(a_i)$ atteint la valeur du fractile d'ordre (1- α/2) de la loi de Student à (n-p-2) degrés de liberté.

L'indice de Student (le t-test de Student) est employé aussi pur évaluer la significativité du modèle complet. Le test s'écrit : H_0 : r = 0 et H_1 : r≠0

Si le coefficient de corrélation est différent de zéro, on rejette l'hypothèse H_0 et on accepte H_1 donc le modèle est significatif.

Sous H_0, la loi de Student à (n-p-1) degré de liberté t_{cal} s'écrit :

$$t_{cal} = \left[\frac{r}{\sqrt{\frac{1-r^2}{(n-p-1)}}} \right]$$

La règle de décision est la suivante :

Si | t_{cal} | > t* où t* est la valeur critique de la table de Student pour un risque fixé et un nombre de degré de liberté égal à (n-p-1) à une Probabilité (1- $\frac{\alpha}{2}$) , on rejette H_0 et on accepte H_1 : le coefficient est significativement différent de zéro et la variable joue un rôle explicatif dans le modèle.

VI.1.2.6 Test de Fisher-Snedecor (F)

F le test de Fisher, reflète le rapport de la variance expliquée par le modèle et la variance due à l'erreur dans le modèle, les valeurs élevées du test F indiquent la signification de l'équation [82]. Les hypothèses du test de Fisher sont les suivantes :

H_0: l'ensemble des coefficients du modèle est non significatif.

H_1: il existe au moins un coefficient non nul.

On calcule le F (observé) à partir de la formule :

$$F_{obs} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \hat{y}_i)^2} \frac{n-p-1}{p}$$

Avec : F est l'indice de Fisher, $y_i$et $\hat{y}_i$ sont, respectivement, les valeurs observées et calculées de la variable dépendante, $\bar{y}$est la valeur moyenne des valeurs prédites, n est le nombre des observations (les molécules), p est le nombre de variables indépendantes (les descripteurs).

La règle de décision est donc :

Si F >F*(p,q) où F*(p,q) est la valeur donnée par la table de Fisher pour p et q donnés et pour un risque fixé

⇒ On accepte H_1 : il existe au moins un coefficient non nul. Cela signifie que les deux variances ont des valeurs suffisamment proches pour qu'on accepte l'idée qu'elles soient homogènes.

VI.2 Réseaux de Neurones Artificiels

Les réseaux de neurones sont une sorte d'architecture informatique inspirée du réseau de neurones biologiques du cerveau [84,85]. Les réseaux de neurones artificiels (ANN) «Artificial Neural Networks» constituent l'une des méthodes bien connues d'intelligence artificielle pour résoudre des problèmes complexes. Les applications les plus courantes sont la reconnaissance de formes et plus généralement l'analyse des données pour trier, prendre une décision ou contrôler un procédé. L'ANN fera alors intervenir des paramètres tels que la vitesse d'apprentissage et la mémoire tout comme le cerveau humain. Une fois l'apprentissage effectué, il sera à même de résoudre de nouveaux problèmes.

Les avantages d'un réseau de neurones sont [86]:

- La possibilité de fournir une réponse soit linéaire soit non linéaire selon le contexte. Ceci est très utile car dans le cadre de cas réels, les données évoluent souvent de façon non linéaire et les modes de calcul linéaires peuvent s'avérer insuffisants.
- La possibilité de construire un modèle uniquement basé sur la connaissance des données d'entrée et de sortie, sans nécessiter de connaître le type de relation mathématique qui les relie.
- La possibilité d'utiliser l'ANN dans un environnement différent de celui qui a servi à la phase d'apprentissage grâce à sa capacité d'adaptation et sa faible sensibilité aux données aberrantes. En effet, chaque neurone est sensible à l'activité globale générée par tous les autres neurones collectivement.
- La possibilité de pouvoir interpréter un résultat tel qu'un tri avec un taux de confiance.
- La possibilité de continuer à fonctionner malgré une panne ou un dysfonctionnement de l'un des neurones. Dans ce cas, la réponse de l'ANN est légèrement dégradée, mais reste fiable car l'information a été transmise de façon distribuée sur tous les neurones. L'ANN est donc particulièrement robuste.

VI.2.1 Structure des réseaux de neurones

Un réseau de neurones artificiels (ANN) est une association de plusieurs perceptrons interconnectés. Plusieurs architectures existent selon le nombre de couches de neurones dans la chaîne de transmission, le nombre de neurones dans chaque couche et aussi selon les règles de propagation du signal dans le réseau.

a) Perceptron simple

Ce réseau est dit simple car il ne se compose que de deux couches : une couche d'entrée et une couche de sortie ce qui implique une seule matrice de poids. L'ensemble des unités de la couche d'entrée sont connectés à celles de la couche de sortie. Les données d'entrée ne sont pas toutes connectées au perceptron avec le même poids, mais au contraire chacune d'entre elles est affectée d'un poids différent, indiquant son importance (**Figure 11**). Ces structures lui permettent d'être considéré comme un classifieur linéaire [87]

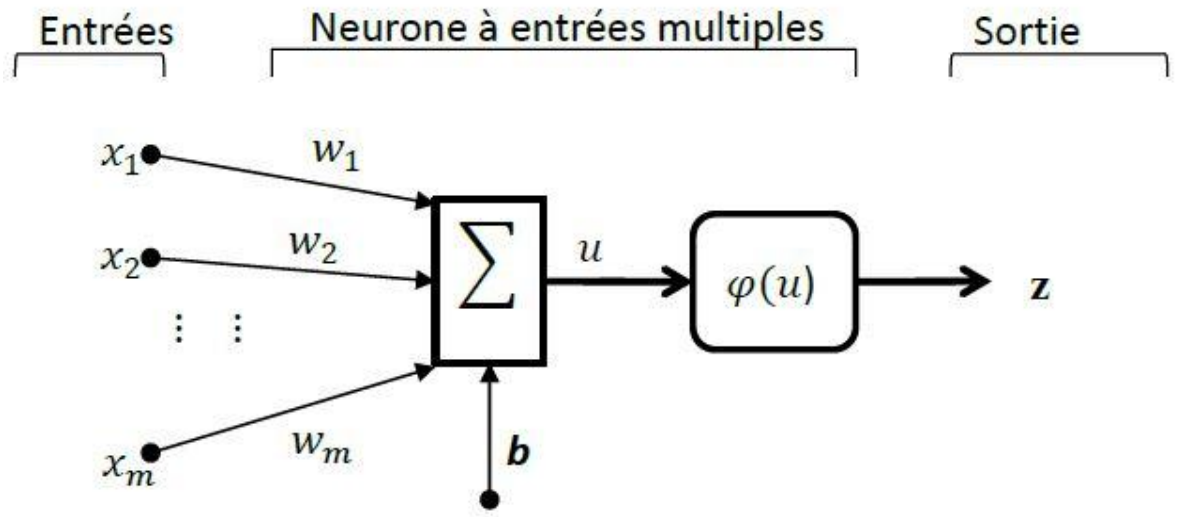


Figure 11 : Schéma de principe d'un perceptron simple

b) Perceptron multicouche

Le perceptron multicouche se compose de trois couches: une couche d'entrée, une couche de sortie et une ou plusieurs couches cachées. Si le réseau possède n couches, alors il possède n-1 matrice de poids. Le perceptron multicouche doté d'une couche cachée est théoriquement un approximateur universel de fonctions. En théorie, il suffit donc d'ajouter un nombre de neurones suffisant au niveau de la couche cachée pour approximer n'importe quelle fonction non linéaire [88].

VI.2.2 Types de réseaux de neurones

Il existe deux types de réseaux de neurones : des réseaux bouclés et des réseaux non bouclés.

Les réseaux de neurones bouclés peuvent avoir une topologie de connexions quelconque, ils ont été largement étudiés en tant que mémoires associatives, dans ce cadre, le réseau est considéré comme un système dynamique non linéaire, dont les attracteurs sont mis à profit pour retrouver des informations mémorisées pendant l'apprentissage.

Un réseau de neurones non bouclé est représenté graphiquement par un ensemble de neurones connectés entre eux, l'information circulant des entrées vers les sorties dans un seul sens "sans retour" [88] (**Figure 12**). Les réseaux non bouclés à couche sont structurés de manière telle que les neurones appartenant à une même couche ne sont pas connectés entre eux, chacune des couches recevant des signaux de la couche précédente, et transmettant le résultat de ses traitements à la couche suivante [89]. Les deux couches extrêmes correspondent à la couche d'entrée la couche d'entrée représentée par les descripteurs pertinents obtenus en technique MLR, et à la couche de sortie qui représente les valeurs d'activités expérimentales logIC5. Les couches intermédiaires sont appelées couches cachées. Un problème majeur dans le réseau de neurones est de savoir comment déterminer le nombre de nœuds dans la couche cachée. Bien qu'il n' y ait pas de règle rigoureuse sur laquelle s'appuyer [90], un moyen pratique est d'utiliser un paramètre ρ pour déterminer le nombre d'unités cachées [91]. ρ est défini comme le nombre de composés représentés dans le réseau / nombre de connexions dans le réseau. La valeur raisonnable de ρ devrait se situer entre 1 et 3. Si $\rho < 1$, le réseau mémorise simplement les données, alors que si $\rho > 3$, le réseau n'est pas capable de généraliser [90].

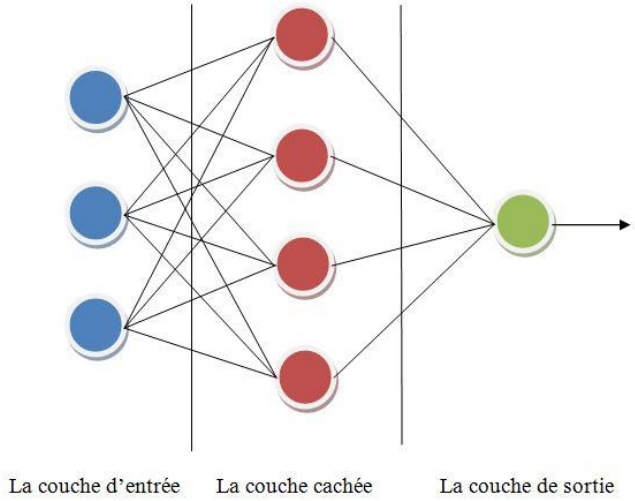


Figure 12: Architecture d'un réseau de neurones artificiel.

VI.2.3 Fonction d'activation

La fonction d'activation du neurone est choisie pour avoir un certain nombre de propriétés. Les fonctions d'activation sont la fonction continue et différentielle par partie, ce qui permet de former le réseau lors de la rétro-propagation. Le problème avec la fonction de transfert linéaire est que tout réseau neuronal multicouche peut se convertir efficacement en un réseau monocouche équivalent et ne supporte pas la fonction non linéaire. Par conséquent, une fonction d'activation non linéaire est nécessaire pour bénéficier des avantages d'un réseau multicouches [92]. Différentes fonctions d'activation ont été suggérées et expérimentées, mais les plus courantes et largement utilisées sont les fonctions à seuil, les fonctions sigmoïdes et les fonctions linéaires (**Figure 13**). La fonction sigmoïde est la plus utilisée car elle représente un bon compromis entre les fonctions seuils et linéaires [93] .

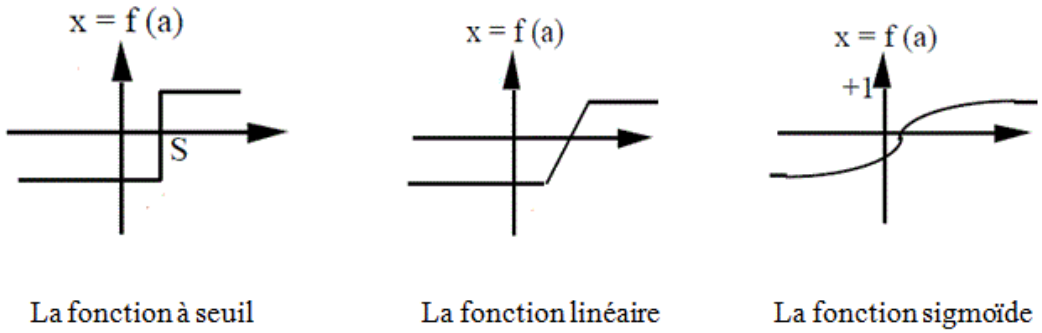


Figure 13: Différents types de fonction de transfert pour le neurone artificiel.

VI.2.4 Fonction de perte

La fonction de perte est utilisée pour l'estimation des paramètres et estime la précision du réseau neuronal. Il tente de saisir la différence entre la valeur du prédicteur et la valeur réelle. Les paramètres du réseau de neurones sont optimisés afin de minimiser la fonction de perte. Ceci permet de traiter la mise à jour des paramètres comme un problème d'optimisation. Le type de la fonction de perte dépend fortement du type de problème [94]. La fonction de perte la plus courante pour les problèmes de régression est:

- Erreur quadratique moyenne (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(t_i - p_i)^2}$$

- Erreur absolue moyenne (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|t_i - p_i|$$

où n est le nombre total de points de données, p_i et t_i sont respectivement les valeurs prédite et la réelle pour le $i^{ème}$ point de données.

VI.3 Validation des modèles

La validation d'une relation quantitative structure-activité est probablement l'étape la plus importante de toutes. La validation permet d'estimer la fiabilité et l'exactitude des prévisions avant la mise en pratique du modèle. De mauvaises prédictions désorientent le développement de médicaments et rendent inutiles les efforts de recherche. La technique la plus employée pour déterminer la stabilité du modèle prédictif est de tester l'influence de chaque échantillon ou plusieurs échantillons sur le modèle final. Pour ce faire, on emploie une technique de validation externe et interne [95].

VI.3.1 Coefficients et tests statistiques standards

Afin de déterminer la qualité d'un modèle, différents indicateurs statistiques sont employés. Le plus répandu d'entre eux est le coefficient de corrélation R^2 qui évalue la part de la variance expliquée par le modèle [90].

$$R^2 = 1 - \frac{\sum(y_i - y_i)^2}{\sum(y_i - \bar{y}_i)^2}$$

Où y_i les valeurs observées de l'activité, y_i les valeurs de l'activité prédites par l'activité et $\bar{y}_i$ est la valeur moyenne des valeurs prédites.

VI.3.2 Validation interne

La technique la plus employée pour déterminer la stabilité du modèle prédictif est de tester l'influence de chaque échantillon sur le modèle final. Pour ce faire, on emploie une technique de validation croisée (cross validation ou CV) avec le processus «Leave- One-Out» ou «Leave-Many-Out».

Ce processus consiste à extraire un certain nombre *n* de molécules du jeu initial à *k* molécules et à construire un nouveau modèle avec les *k-n* molécules restantes à l'aide des descripteurs choisis (seules les constantes de la régression changent). Ce nouveau modèle est alors utilisé

pour la phase de prédiction sur les *n* molécules retirées. Ce processus est ensuite réitéré pour retirer et prédire les valeurs de toutes les molécules du jeu d'entraînement.

En fonction du nombre de molécules retirées à chaque itération, si en retire une seule molécule, on parlera de Leave-One-Out(LOO), mais si plusieurs molécules sont retirées, on parlera de Leave-Many-Out(LMO).

D'une manière générale, ces techniques de validation interne permettent l'évaluation de la robustesse du modèle, autrement dit la stabilité des paramètres du modèle QSAR vis-àvis des molécules du jeu d'entraînement. Cela dit, elles ne permettent en aucun cas de démontrer le pouvoir prédictif des modèles [90,96] .

VI.3.2.1 Procédure leave-one-out

Pour un nombre N d'exemples d'apprentissage, on retire à chaque itération un exemple I de l'ensemble d'apprentissage initial. Une série d'apprentissage est réalisée pour les N-1 molécules restantes et la molécule retirée est prédite par le modèle formé [96].

La performance des modèles de régression est estimée avec les paramètres statistiques de la validation croisée décrits par les équations ci-dessous :

Le coefficient de détermination de la validation croisée $R^2_{cv(loo)}$:

$$R^2_{cv(loo)} = \frac{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2 - \sum_{i=1}^{N}(y_i - y_{icv})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

Avec y_{icv} : Représente la valeur de l'activité de la molécule i calculée par la méthode de la validation croisée,

$\bar{y}$: La valeur moyenne de l'activité,

y_i : Les valeurs observées de l'activité.

L'écart type de la validation croisée $S_{cv(LOO)}$:

$$S_{cv(LOO)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - y_{icv})^2}$$

V.3.3 Y-Randomization

Y-randomisation est une technique répandue pour valider la robustesse des modèles QSAR. En effet, il n'est pas rare d'obtenir des corrélations fortuites (ou « chance corrélation »), c'est-à- dire un modèle affichant de bons résultats statistiques (R^2, SD) pour l'apprentissage, , mais impliquant des descripteurs qui ne sont pas reliés à la propriété/activité modélisée. Ces modelés aléatoires peuvent être détectes par la procédure Y-randomization [97]. Elle consiste à mélanger aléatoirement les propriétés/activités expérimentales pour la série d'apprentissage en utilisant les mêmes descripteurs, de nouveaux modèles sont obtenus. Pour un modèle QSAR acceptable, le coefficient de corrélation moyen (Rr) des modèles randomisés doit être inférieur au coefficient de corrélation (R) du modèle non randomisé, cela signifie que les capacités prédictives du modèle QSAR construit ne sont pas dues aux corrélations de chance [95,98].

Cependant, La validation externe du modèle est indispensable pour vérifier le pouvoir prédictif d'un modèle QSAR. D'après certains chercheurs, la validation externe est une norme et une partie obligatoire dans la modélisation QSAR [99,100].

VI.3.4 Validation externe

Plusieurs stratégies différentes sont employées pour la Validation externe. Une procédure très courante consiste à faire une répartition aléatoire des données entre jeux d'entraînement et de validation. Afin de valider l'efficacité de la subdivision, il est recommandé, soit un ensemble de molécules divisé en deux : un ensemble d'apprentissage sur lequel le modèle est développé et un ensemble de validation utilisé pour déterminer sa validité externe. Une fois l'ensemble de validation mis en place, il suffit alors d'appliquer le modèle QSAR aux molécules qui le composent et de déterminer la corrélation existante entre les activités calculées et celles expérimentales. En effet, un tel partitionnement risque de donner lieu à des jeux de données dont les distributions sont incompatibles. Il faut donc vérifier que la partition choisie ne pose pas ce genre de problème. Plus cette corrélation est importante, plus le modèle est capable de prédire les activités pour des molécules hors l'ensemble d'apprentissage [101,102].

$$R^2_{ext} = 1 - \frac{\sum_{i=1}^{N} (y_i - y_{iext})^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$$

Avec y_{iext} : est les valeurs calculées des molécules composants l'ensemble de validation externe.

Récemment plusieurs études [90,103] ont montré l'insuffisance des paramètres R^2(test), R^2_{cv} (test) pour vérifier la capacité prédictive des modèles QSAR. Par conséquent, d'autres paramètres doivent être respectés pour cet objectif. Ces paramètres sont connus sous le nom de ≪critères de validation externe≫ ou souvent appelés ≪critères de Trophsa et Golbraikh≫[104].

Cinq critères de Trophsa et Golbraikh sont définis :

Critère 1 : $R^2_{ext} > 0.5$

Critère 2 : $R^2 > 0.6$

Critère 3 : $\frac{R^2 - R_0^2}{R^2} < 0.1$ et $0.85 < K < 1.15$

Critère 4 : $\frac{R^2 - R_0'^2}{R^2} < 0.1$ et $0.85 < K' < 1.15$

Critère 5 : $|R^2 - R_0^2| < 0.3$

Avec

R^2:Coefficient de corrélation pour les molécules de la série de test.

R_0^2: Coefficient de corrélation entre les valeurs prédites et expérimentales pour la série de test.

$R_0'^2$: Coefficient de corrélation entre les valeurs expérimentales et prédites pour la série de test.

k : Constante de la droite (à l'origine) de corrélation (valeurs prédites en fonction des valeurs expérimentales)

k` : Constante de la droite (à l'origine) de corrélation (valeurs expérimentales en fonction des valeurs prédites).

VI.4 Domaine d'applicabilité

Un modèle QSAR ne peut pas être considéré comme un modèle universel, parce qu'il est développé sur un nombre limité de composés qui ne couvrent pas tout l'espace chimique. Le domaine d'applicabilité est la région de l'espace chimique définie par les molécules de l'ensemble d'apprentissage du modèle [105]. Les modèles QSAR ne peuvent pas prédire des propriétés de manière fiable pour l'intégralité des composés chimiques existants. En effet, un

modèle QSAR n'est pas destiné à être employé en dehors de son domaine d'applicabilité, c'est-à-dire en dehors de l'espace chimique couvert par son ensemble d'apprentissage. La détermination des domaines d'applicabilités est donc d'une grande importance. Cette partie de l'analyse est d'ailleurs explicitement demandée dans les démarches de validation mises en place au niveau de l'OCDE (Organisation de Coopération et de Développement Economique)[95]. Les méthodes les plus utilisées pour décrire le domaine d'applicabilité consistent à prendre en compte l'intervalle des descripteurs individuels et la présence de fragments structuraux dans l'ensemble d'apprentissage.

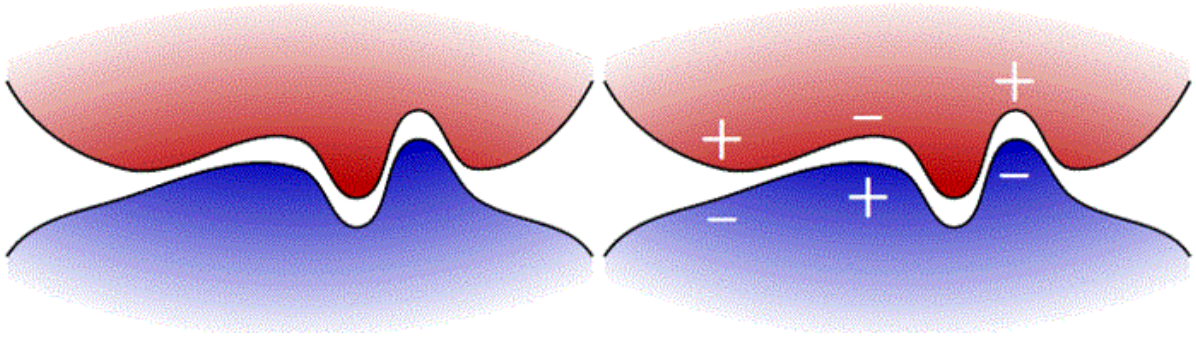
V. Docking moléculaire

V.1 Généralités sur le Docking Moléculaire

Le docking moléculaire est un outil informatique couramment utilisé dans les projets de découverte de nouveaux médicaments et les études biologiques fondamentales sur les interactions protéines-ligand (**figure 14**). En général, Le docking moléculaire est utilisé pour répondre à l'une des trois questions suivantes:

- Prédire le mode de liaison entre un ligand et un récepteur,
- Tester une série de petites molécules sur un récepteur et identifier les ligands d'un potentiel actif,
- prédire l'affinité de liaison entre le ligand et un récepteur cible.

À l'heure actuelle, ces trois questions ne sont pas de la même difficulté, où la prédiction de position est considérée comme la plus facile et la prédiction de l'affinité de liaison est la plus difficile [106–108].



Complémentarité géométrique complémentarité électrostatique

Figure 14: Mesure pour l'évaluation du docking moléculaire.

Le docking fait référence au processus de détermination de l'orientation et de la conformation d'un ligand dans un site de liaison macromoléculaire (récepteur). Les algorithmes du docking ont deux objectifs: trouver la bonne conformation liée et prédire correctement l'activité biologique, ce qui implique d'estimer l'énergie libre de liaison du complexe ligand-récepteur par une fonction de score [109].

Les algorithmes du docking actuellement utilisés diffèrent par leur représentation du ligand et du récepteur et leur flexibilité, les méthodes utilisées pour la recherche conformationnelle et leur description de la fonction de score. Chacun de ces paramètres est traité en détail dans les sections suivantes.

V.1.1 Principe du docking moléculaire

Le docking moléculaire vise à prédire la structure d'un complexe moléculaire à partir des molécules isolées, dans lesquelles différentes approches sont combinées pour étudier les modes d'interaction entre deux molécules. Les logiciels du docking moléculaire sont donc des outils très utiles en biologie, pharmacie et médecine, car la plupart des principes actifs sont de petites molécules (ligands) qui interagissent avec une cible biologique d'intérêt thérapeutique. Le récepteur macromoléculaire étant le plus souvent une protéine, le terme docking seul est couramment employé pour désigner un « docking protéine-ligand » [108]. Le docking moléculaire a ainsi pour but de déterminer le mode d'interaction d'un complexe avec une ou plusieurs molécules, en cherchant l'orientation et la conformation favorables pour la fixation d'un ligand à un récepteur [110].

Une simulation du docking comprend essentiellement deux étapes :

- L'algorithme du docking : c'est une étape de sélection, consiste à placer le ligand dans le site actif de la protéine et à échantillonner les conformations, positions et orientations (poses) possibles, en ne retenant que celles qui représentent les modes d'interactions les plus favorables.
- La fonction score : elle consiste à évaluer l'affinité entre le ligand et la protéine et de donner un score aux poses obtenues lors de la phase du docking. Ce score permettra de retenir la meilleure pose parmi toutes celles proposées [85,109].

V.1.2 Algorithme du docking

En principe, un docking peut être fait de façon « manuelle », directement par le modélisateur, en plaçant le ligand dans le site actif de la protéine à l'aide d'une interface graphique. Ensuite,

la géométrie de l'ensemble est optimisée de manière à corriger les problèmes stériques et obtenir un complexe énergétiquement stable. Cette approche est appliquée quand on a une idée précise du mode d'interaction réel du ligand [111].

Néanmoins, le plus souvent, le mode d'interaction réel n'est pas connu. Dans ce cas, tester manuellement toutes les conformations et orientations des ligands s'avère impossible d'un point de vue pratique, même en considérant la protéine comme un corps rigide. Pour contourner cette difficulté, les algorithmes du docking ont été conçus pour rechercher de façon objective, rapide et efficace les modes d'association protéine-ligand les plus favorables [112].

Les algorithmes du docking peuvent être séparés en deux grandes classes : ceux qui ne tiennent pas compte de la flexibilité de la protéine, en traitant celle-ci comme un corps rigide, et ceux qui sont capables de prendre en compte, partiellement la flexibilité du récepteur.

- Dans les procédures du docking considérant la protéine comme un corps rigide, la flexibilité du ligand seule est prise en compte pour l'obtention de complexes. Selon la méthode utilisée pour générer les conformations du ligand et les placer dans la cavité catalytique du récepteur, les algorithmes peuvent être sous-divisés en algorithmes de simulation de mécanique moléculaire (MM) et dynamique moléculaire (DM) [110,113], de forme systématique et stochastique. Plusieurs articles de revues décrivant le principe, les points forts et les limitations de ces méthodes sont disponibles [114,115].
- Certaines protéines présentent naturellement des régions de grande flexibilité, subissant des réarrangements considérables en présence d'un ligand. Dans ce cas, négliger la flexibilité de la protéine peut mettre en péril la fiabilité des résultats du docking, et rend nécessaire l'utilisation d'approches capables de tenir compte de la flexibilité du système entier [116].

V.1.3 Fonction de score

Une bonne fonction de score doit être capable de différencier les ligands actifs et non actifs, d'identifier les poses correctes et d'estimer l'énergie de liaison. Les calculs d'énergie libre utilisant des simulations de dynamique moléculaire peuvent fournir une meilleure précision, mais ils sont extrêmement lents et donc inadéquats pour évaluer les grandes bibliothèques moléculaires. Les fonctions de score mises en œuvre dans les algorithmes du docking tendent à simplifier ou à ignorer les termes physiques et chimiques complexes comme l'entropie ou

l’hydrophobie. Néanmoins, il est possible d'utiliser à la place des descripteurs physiques qui sont corrélés avec ces termes complexes (p. ex. nombre de liaisons en rotation au lieu de pertes entropiques). Les fonctions de score peuvent être divisées en trois catégories:

- La fonction du champ de forces basée sur l’approximation de la force de l'interaction en tenant compte des interactions électrostatiques et van-der-waals entre les couples d'atomes [117,118],
- Les fonctions empiriques utilisent un ensemble de descripteurs physiques et chimiques pour rendre compte de l'interaction entre un ligand et un récepteur [119]. Des termes comme le nombre de liaisons hydrogène, le contact hydrophobe et le nombre de liaisons rotatives peuvent être utilisés pour estimer approximativement les contributions enthalpiques et entropiques,
- Les méthodes basées sur la connaissance utilisent des modèles 3D de complexes protéine-ligands déterminés expérimentalement pour analyser statistiquement les préférences des interactions entre différents types d'atomes ou groupes chimiques. Ces méthodes partent de l'hypothèse que les types d'interactions hautement observées sont énergétiquement favorables [120].

V.1.4 Choix du logiciel AUTODOCK

Un grand nombre de programmes (commerciaux ou non) du docking moléculaire sont disponibles. Parmi ceux-ci, nous citerons par exemple AUTODOCK [121], FLEXX [118], GOLD [122], SURFLEX- DOCK [123], MOLEGRO VIRTUAL DOCKER [124] , UCSF CHIMERA [125] et Schrödinger (Glide) [126] .

Dans ce travail nous avons choisi l’AUTODOCK en raison des caractéristiques suivantes : à partir d’une position initiale aléatoire, à l’extérieur du site actif, le ligand explore le site étudié par la répétition successive de mouvements et d’évaluations de l’interaction ligand-récepteur. Les mouvements sont effectués par des opérations de translation, de rotation et de changement de conformation. L’énergie d’interaction est calculée par une fonction énergétique. Les mouvements du cycle à venir sont guidés par les variations d’énergie induites par les mouvements des cycles précédents. L’algorithme s’arrête lorsqu’il a trouvé la position idéale du ligand dans le récepteur. Ces techniques sont plus lentes, mais ils prennent mieux en compte la flexibilité du ligand et permettent l’exploration de régions plus vastes.

V.2 AUTODOCK

L'AUTODOCK est constitué d'un ensemble de trois programmes et d'utilitaires exécutés en ligne de commandes : 1) AutoDockTools, qui définit les rotations des liaisons avec les coordonnées du ligand en entrée et permet de préparer l'expérience du docking moléculaire, 2) AutoGrid, qui calcule la grille tridimensionnelle de l'énergie d'interaction basée sur les coordonnées macromoléculaires ; et 3) AutoDock4, qui effectue la simulation du docking moléculaire.

La procédure AUTODOCK 4.2 schématisée dans la **figure15** se décompose en cinq étapes:

- Préparation des fichiers de coordonnées atomiques du ligand et de la macromolécule ;
- Définition des pivots flexibles du ligand,
- Calcul des grilles de potentiel d'interaction du récepteur,
- Recherche des solutions du docking,
- Analyse des résultats.

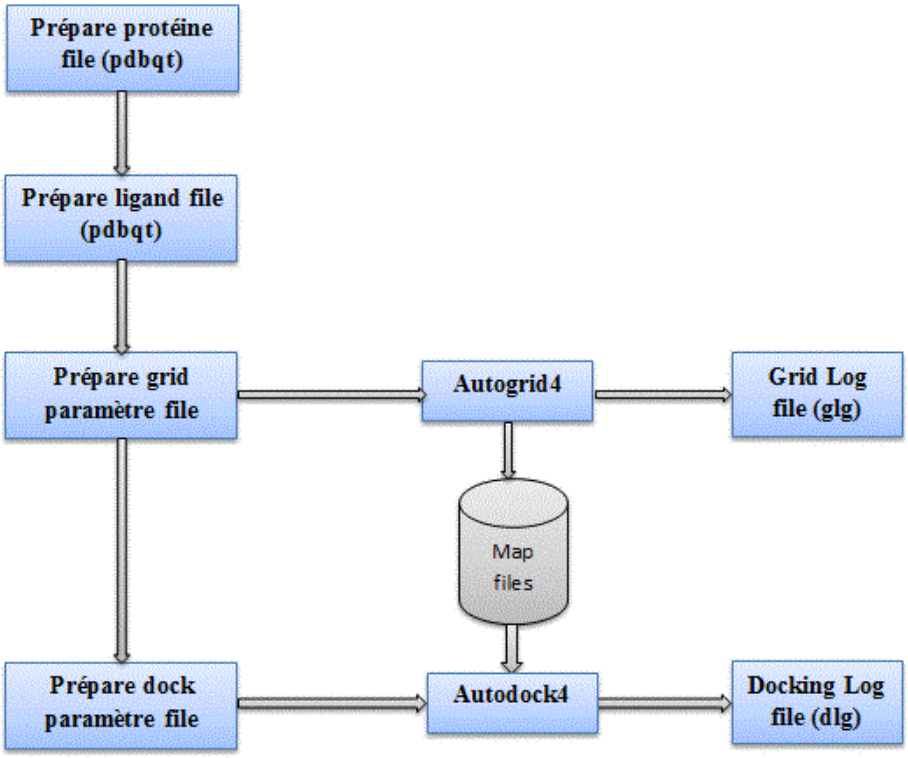


Figure 15: Procédures du docking avec AutoDock

Algorithme AutoDock permet d'utiliser trois méthodes de recherche différentes [117] :

- une méthode de recherche utilisant un algorithme génétique classique, où les chromosomes sont réellement des chaînes de caractères. Ils sont composés dans l'ordre des trois cordonnées cartésiennes représentant la translation du ligand, suivies de quatre variables définissant un quaternion, qui est un ensemble mathématique qui définit la rotation du ligand dans l'espace, et enfin une valeur réelle pour chaque angle de torsion de chaque liaison du ligand.
- une méthode de recherche locale, basée sur un simulé de type Monte Carlo, peut être ajoutée pour minimiser l'énergie [127].
- une méthode mixte de recherche globale et locale, l'algorithme génétique lamarckien, basé comme son nom l'indique sur les principes génétiques énoncés par Jean Baptiste Lamarck au début du XXe siècle.

AutoDock 4 utilise une fonction de score pour l'évaluation de l'énergie libre basée sur un champ de force semi-empirique [128]. Ce champ de force a été calibré en utilisant un large groupe hétérogène de données expérimentales de complexes protéine-ligand. Cette fonction de score évalue les énergies intramoléculaires du ligand et celles de la protéine sont calculées indépendamment ; elles correspondent à des partenaires non liés et séparés dans l'espace. Dans un second temps, l'énergie intermoléculaire est calculée ; elle correspond à l'énergie du complexe protéine/ligand, une fois que ce dernier est positionné dans le site actif. Ces énergies intra et inter-moléculaires prennent en compte les énergies de van der Waals de dispersion et répulsion entre atomes, l'énergie des liaisons hydrogène, l'énergie électrostatique et l'énergie de VDW aussi bien positive (des atomes de carbones) que négative (des atomes polaires et des atomes chargés). L'énergie libre d'interaction est ainsi la différence entre les valeurs des énergies liées et non liées (du ligand, de la protéine et du complexe) auquel est ajouté un terme correspondant à la perte d'entropie du ligand lors de sa fixation à la protéine.

V.3 Avantages et limites du docking moléculaire

V.3.1 Avantages du docking moléculaire

- L'application du docking dans un système ciblé de livraison de médicaments est un énorme avantage. On peut étudier la taille, la forme, la distribution des charges, la polarité, la liaison de l'hydrogène et les interactions hydrophobes du ligand-récepteur.

- Le docking moléculaire aide à identifier les sites cibles du ligand et de la molécule réceptrice.
- Le docking aide également à comprendre les différentes enzymes et leur mécanisme d'action.
- La fonction "score" du docking aide à choisir le meilleur ajustement ou le meilleur médicament parmi une série d'options.
- Tout ne peut pas être prouvé expérimentalement, car les méthodes expérimentales pour la découverte de médicaments prennent beaucoup de temps. Le docking moléculaire aide à accélérer le processus de conception des médicaments assistés par ordinateur et fournit également toutes les conformations possibles en fonction du récepteur et de la molécule ligand.
- Le docking a un avantage énorme quand il s'agit de l'étude des interactions ligand-protéines.
- Il existe des millions de composés, ligands, médicaments et récepteurs dont la structure 3D a été cristallisée. Le criblage virtuel de ces composés peut être réalisé.

V.3.2 Limites du docking moléculaire

- Dans le docking ligand-récepteur, il peut y avoir des problèmes dans la structure du récepteur. Une valeur de résolution fiable pour le docking de petites molécules est inférieure à 1,2 Å [129], tandis que la plupart des structures cristallographiques ont une résolution comprise entre 1,5 et 2,5 Å. L'utilisation accrue des modèles d'homologie dans Le docking devrait être examinée avec soin, car ils ont une résolution encore plus faible [130]. La plupart des applications acceptent et donnent de bons résultats pour les structures inférieures à 2,2 Å. Tout de même, il faut être prudent lors du choix d'une structure.
- Les fonctions de score utilisées dans Le docking, presque toutes, ne tiennent pas compte du rôle joué par les inhibiteurs ou ions liés de manière covalente [131].
- La méthodologie et la recherche dans le domaine du docking protéine-protéine doivent être considérablement augmentées car le succès dans ce domaine est grandement entravé par de nombreux faux positifs et faux négatifs [132].

Références

[1] T. Igarashi, T. Nakazawa, C. Suzuki, T. Tsuru, Y. Kaji, Fast calculation approach to semi-empirical molecular orbital method using real space division method, Comput. Mater. Sci. 50 (2011) 3346–3349.

[2] S. Ogata, F. Shimojo, R.K. Kalia, A. Nakano, P. Vashishta, Hybrid quantum mechanical/molecular dynamics simulation on parallel computers: Density functional theory on real-space multigrids, Comput. Phys. Commun. 149 (2002) 30–38.

[3] O.E. Glukhova, G.V. Savostyanov, M.M. Slepchenkov, A New Approach to Dynamical Determination of the Active Zone in the Framework of the Hybrid Model (Quantum Mechanics/ Molecular Mechanics), Procedia Mater. Sci. 6 (2014) 256–264.

[4] G.L. Patrick, An Introduction to Medicinal Chemistry, UCL Lect. Notes. (2009).

[5] W.D. Cornell, A.E. Howard, P. Kollman, Molecular mechanical potential functions and their application to study molecular systems: Current Opinion in Structural Biology 1991, 1:201–212, Curr. Opin. Struct. Biol. 1 (1991) 201–212.

[6] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, J. Comput. Chem. 26 (2005) 1668–1688.

[7] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.

[8] M. Christen, P.H. Hünenberger, D. Bakowies, R. Baron, R. Bürgi, D.P. Geerke, T.N. Heinz, M.A. Kastenholz, V. Kräutler, C. Oostenbrink, C. Peter, D. Trzesniak, W.F. van Gunsteren, The GROMOS software for biomolecular simulation: GROMOS05, J. Comput. Chem. 26 (2005) 1719–1751.

[9] S. Lifson, A. Warshel, Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and *n* -Alkane Molecules, J. Chem. Phys. 49 (1968) 5116–5129.

[10] T.A. Halgren, Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules, J. Comput. Chem. 17 (1996) 616–641.

[11] N.L. Allinger, Y.H. Yuh, J.H. Lii, Molecular mechanics. The MM3 force field for hydrocarbons. 1, J. Am. Chem. Soc. 111 (1989) 8551–8566.

[12] N.L. Allinger, K. Chen, J.-H. Lii, An improved force field (MM4) for saturated hydrocarbons, J. Comput. Chem. 17 (1996) 642–668.

[13] A.K. Rappe, C.J. Casewit, K.S. Colwell, W.A. Goddard, W.M. Skiff, UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, J. Am. Chem. Soc. 114 (1992) 10024–10035.

[14] I. Pettersson, T. Liljefors, Molecular Mechanics Calculated Conformational Energies of Organic Molecules: A Comparison of Force Fields, in: 2007: pp. 167–189.

[15] N.L. Allinger, H.L. Flanagan, Isotope effects in molecular mechanics (MM2). Calculations deuterium compounds, J. Comput. Chem. 4 (1983) 399–403.

[16] G.B. Jones, B.J. Chapman, π Stacking Effects in Asymmetric Synthesis, Synthesis (Stuttg). (1995) 475–497.

[17] A. Hocquet, M. Langgård, An Evaluation of the MM+ Force Field, J. Mol. Model. 4 (1998) 94–112.

[18] S.J. Weiner, P.A. Kollman, D.T. Nguyen, D.A. Case, An all atom force field for simulations of proteins and nucleic acids, J. Comput. Chem. 7 (1986) 230–252.

[19] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.

[20] A.D. McNaught, A. Wilkinson, International Union of Pure and Applied Chemistry., Compendium of chemical terminology : IUPAC recommendations, Blackwell Science, 1997. http://www.old.iupac.org/publications/books/author/mcnaught.html (accessed January 10, 2019).

[21] D. Chandler, J. Weeks, H.A.- Science, undefined 1983, Van der Waals picture of liquids, solids, and phase transformations, Science.Sciencemag.Org. (n.d.). http://science.sciencemag.org/content/220/4599/787.short.

[22] R.G. Parr, W. Yang, Density-Functional Theory of Atoms and Molecules, 1989.

[23] E.K.U.G. R. M. Dreizler, Density Functional Theory, Springer -Ver Lag Berlin Heidelb. New York London. (1990) 1–16.

[24] L.H. Thomas, The calculation of atomic fields, Math. Proc. Cambridge Philos. Soc. 23 (1927) 542–548.

[25] E. Fermi, Statistical method to determine some properties of atoms, Rend. Accad. Naz. Lincei. 6 (1927) 602–607.

[26] P.A.M. Dirac, Note on Exchange Phenomena in the Thomas Atom, Math. Proc. Cambridge Philos. Soc. 26 (1930) 376.

[27] E. Teller, On the Stability of Molecules in the Thomas-Fermi Theory, Rev. Mod. Phys. 34 (1962) 627–631.

[28] P. Hohenberg, W. Kohn, Inhomogeneous Electron Gas, Phys. Rev. 136 (1964) B864–B871.

[29] W. Kohn, L.J. Sham, Self-Consistent Equations Including Exchange and Correlation Effects, Phys. Rev. 140 (1965) A1133–A1138.

[30] J.P. Perdew, A. Zunger, Self-interaction correction to density-functional approximations for many-electron systems, Phys. Rev. B. 23 (1981) 5048–5079.

[31] S.H. Vosko, L. Wilk, M. Nusair, Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis, Can. J. Phys. 58 (1980) 1200–1211.

[32] J.P. Perdew, Y. Wang, Pair-distribution function and its coupling-constant average for the spin-polarized electron gas, Phys. Rev. B. 46 (1992) 12947–12954.

[33] S. Lundqvist, N.H. March, eds., Theory of the Inhomogeneous Electron Gas, Springer US, Boston, MA, 1983.

[34] J.P. Perdew, Density-functional approximation for the correlation energy of the inhomogeneous electron gas, Phys. Rev. B. 33 (1986) 8822–8824.

[35] J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, C. Fiolhais, Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation, Phys. Rev. B. 46 (1992) 6671–6687.

[36] A.D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, Phys. Rev. A. 38 (1988) 3098–3100.

[37] J.P. Perdew, Density-functional approximation for the correlation energy of the inhomogeneous electron gas, Phys. Rev. B. 33 (1986) 8822–8824.

[38] PERDEW, J. P., Unified Theory of Exchange and Correlation beyond the Local Density Approximation, Electron. Struct. Solids ’91. 11 (1991). https://ci.nii.ac.jp/naid/10004115120/.

[39] C. Lee, W. Yang, R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, Phys. Rev. B. 37 (1988) 785–789.

[40] M. Levy, J.P. Perdew, Density functionals for exchange and correlation energies: Exact conditions and comparison of approximations, Int. J. Quantum Chem. 49 (1994) 539–548.

[41] A.D. Becke, Density functional calculations of molecular bond energies, J. Chem.

Phys. 84 (1986) 4524–4529.

[42] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett. 77 (1996) 3865–3868.

[43] A.D. Becke, Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals, Http://Oasc12039.247realmedia.Com/RealMedia/Ads/Click_lx.Ads/Www.Aip.Org/Pt/Adcenter/Pdfcover_test/L-37/56140772/X01/AIP-PT/JCP_ArticleDL_110117/AIP-3075_JCP_Perspective_Generic_1640x440.Jpg/434f71374e315a556e61414141774c75?X. (1998).

[44] K. Sharkas, J. Toulouse, A. Savin, Double-hybrid density-functional theory made rigorous, J. Chem. Phys. 134 (2011) 064113.

[45] J. Harris, Adiabatic-connection approach to Kohn-Sham theory, Phys. Rev. A. 29 (1984) 1648–1659.

[46] A.D. Becke, A new mixing of Hartree–Fock and local density-functional theories, J. Chem. Phys. 98 (1993) 1372–1377.

[47] A.D. Becke, Density-functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648–5652.

[48] T. Yanai, D.P. Tew, N.C. Handy, A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP), Chem. Phys. Lett. 393 (2004) 51–57.

[49] Yan Zhao, and Nathan E. Schultz, D.G. Truhlar*, Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions, (2006).

[50] Y. Zhao, D.G. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals, Theor. Chem. Acc. 120 (2008) 215–241.

[51] † Claudia A. Caro, † José H. Zagal, § Fethi Bedioui, § and Carlo Adamo, ‡ Gloria I. Cárdenas-Jirón*, Solvent Effect on Density Functional Reactivity Indexes Applied to Substituted Nickel Phthalocyanines, (2004).

[52] † Pratim K. Chattaraj, ‡ Patricia Pérez, ‡ and Jenny Zevallos, ‡ Alejandro Toro-Labbé*, Ab Initio SCF and DFT Studies on Solvent Effects on Intramolecular Rearrangement Reactions, (2001)..

[53] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman,

J. Montgomery J. A., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. V Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision C.02, 2004.

[54] ACD/ChemSketch for Academic and Personal Use :: ACD/Labs.com, (n.d.). http://www.acdlabs.com/resources/freeware/chemsketch/.

[55] ChemOffice Professional – PerkinElmer Informatics Logiciels, (n.d.). http://www.cambridgesoft.com/Ensemble_for_Chemistry/ChemOffice/ChemOfficeProfessional/.

[56] A.G. Mercader, P.R. Duchowicz, F.M. Fernández, E.A. Castro, Modified and enhanced replacement method for the selection of molecular descriptors in QSAR and QSPR theories, Chemom. Intell. Lab. Syst. 92 (2008) 138–144.

[57] M.K. and, V.S. Lobanov, A.R. Katritzky, Quantum-Chemical Descriptors in QSAR/QSPR Studies, (1996).

[58] S. Rekkab, " Drug Design " et synthèse de nouveaux calix [8] arènes sulfoniques flexibles à activités anticorrosive et anticoagulante, (2014).

[59] H.A. Lorentz, Ueber die Beziehung zwischen der Fortpflanzungsgeschwindigkeit des Lichtes und der Körperdichte, Ann. Phys. (1880).

[60] R. Todeschini, V. Consonni, Wiley InterScience (Online service), Handbook of molecular descriptors, Wiley-VCH, 2000. https://books.google.co.ma/books?hl=ar&lr=&id=TCuHqbvgMbEC&oi=fnd&pg=PP2&dq=Handbook+of+Molecular+Descriptors,&ots=jvzGwcxQl9&sig=N8sZ4tKQEcOzD1Fr2_p-6bt-6_w&redir_esc=y#v=onepage&q=Handbook of Molecular Descriptors%2C&f=false.

[61] H. Wiener, Structural Determination of Paraffin Boiling Points, J. Am. Chem. Soc. 69

(1947) 17–20.

[62] A.T. Balaban, Highly discriminating distance-based topological index, Chem. Phys. Lett. 89 (1982) 399–404.

[63] S.C. Basak, V.R. Magnuson, G.J. Niemi, R.R. Regal, G.D. Veith, Topological indices: their nature, mutual relatedness, and applications, Math. Model. 8 (1987) 300–305.

[64] V.N. Viswanadhan, A.K. Ghose, G.R. Revankar, R.K. Robins, Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics, J. Chem. Inf. Model. 29 (1989) 163–172.

[65] C. Hansch, Structure-activity relationships of chemical mutagens and carcinogens, Sci. Total Environ. 109–110 (1991) 17–29.

[66] F. Neese, A critical evaluation of DFT, including time-dependent DFT, applied to bioinorganic chemistry, JBIC J. Biol. Inorg. Chem. 11 (2006) 702–711..

[67] T. Koopmans, Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms, Physica. 1 (1934) 104–113.

[68] R.G. Parr, R.A. Donnelly, M. Levy, W.E. Palke, Electronegativity: The density functional viewpoint, J. Chem. Phys. 68 (1978) 3801–3807.

[69] R.S. Mulliken, A new electroaffinity scale; Together with data on valence states and on valence ionization potentials and electron affinities, J. Chem. Phys. 2 (1934) 782–793.

[70] R.G. Parr, R.G. Pearson, Absolute Hardness: Companion Parameter to Absolute Electronegativity, J. Am. Chem. Soc. 105 (1983) 7512–7516.

[71] W. Yang, R.G. Parr, Hardness, softness, and the fukui function in the electronic theory of metals and catalysis., Proc. Natl. Acad. Sci. 82 (1985) 6723–6726.

[72] R.G. Parr, L. V. Szentpály, S. Liu, Electrophilicity index, J. Am. Chem. Soc. 121 (1999) 1922–1924.

[73] J.K. Gregory, D.C. Clary, K. Liu, M.G. Brown, R.J. Saykally, The water dipole moment in water clusters, Science (80-.). 275 (1997) 814–817.

[74] A.C. Brown, T.R. Fraser, V.—On the Connection between Chemical Constitution and Physiological Action. Part. I.—On the Physiological Action of the Salts of the Ammonium Bases, derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia, Trans. R. Soc. Edinburgh. 25 (1868) 151–203.

[75] C. Hansch, T. Fujita, *p* -σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure, J. Am. Chem. Soc. 86 (1964) 1616–1626.

[76] S.M. Free, J.W. Wilson, A Mathematical Contribution to Structure-Activity Studies, J. Med. Chem. 7 (1964) 395–399.

[77] A. Shayanfar, S. Shayanfar, Is regression through origin useful in external validation of QSAR models?, Eur. J. Pharm. Sci. 59 (2014) 31–35.

[78] J. Ghasemi, S. Saaidpour, S.D. Brown, QSPR study for estimation of acidity constants of some aromatic acids derivatives using multiple linear regression (MLR) analysis, J. Mol. Struct. THEOCHEM. 805 (2007) 27–32.

[79] A.T.C. Goh, W.G. Zhang, An improvement to MLR model for predicting liquefaction-induced lateral spread using multivariate adaptive regression splines, Eng. Geol. 170 (2014) 1–10.

[80] É.M.-L. Pierre-André Cornillon, No TitleRégression. Théorie et applications, Springer-V, n.d.

[81] J.G. Topliss, R.P. Edwards, Chance factors in studies of quantitative structure-activity relationships., J. Med. Chem. 22 (1979) 1238–1244.

[82] Y. Belmiloud, A. Kadari, L. Benahmed, D. Cherquaoui, D. Villemin, M. Brahimi, Application of neural network to quantitative structure anti-HIV activity relationships of flavonoid compounds, Orient. J. Chem. 25 (2009) 49–56.

[83] M. Şahin, Y. Kaya, M. Uyar, Comparison of ANN and MLR models for estimating solar radiation in Turkey using NOAA/AVHRR data, Adv. Sp. Res. 51 (2013) 891–904.

[84] G. Poinar, Plasmodium dominicana n. sp. (Plasmodiidae: Haemospororida) from Tertiary Dominican amber, Syst. Parasitol. 61 (2005) 47–52.

[85] W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biophys. 5 (1943) 115–133.

[86] A. Plaquet, L'université bordeaux 1, (2011) 1–182.

[87] Ikram Chraibi Kaadoud, Architecture des réseaux de neurones : Réseaux de neurones artificiels classiques (2/3) ! - Intelligence mécanique, Intell. Mécanique. (2018). http://www.scilogs.fr/intelligence-mecanique/architecture-des-reseaux-de-neurones-reseaux-de-neurones-artificiels-classiques-2-3/.

[88] A. Chamekh, Optimisation des procédés de mise en forme par les réseaux de neurones artificiels, (2008). https://tel.archives-ouvertes.fr/tel-00445341.

[89] D. González-Arjona, G. López-Pérez, A. Gustavo González, Non-linear QSAR modeling by using multilayer perceptron feedforward neural networks trained by back-propagation, Talanta. 56 (2002) 79–90.

[90] A. Golbraikh, A. Tropsha, Beware of q2!, J. Mol. Graph. Model. 20 (2002) 269–276.

[91] T.A. Andrea, H. Kalayeh, Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors, J. Med. Chem. 34 (1991) 2824–2836.

[92] L. Douali, A. R. Schmitzer, D. Villemin, A. Jarid, D. Cherqaoui, NEURAL NETWORKS AND THEIR APPLICATIONS IN CHEMISTRY AND BIOLOGY, Phys. Chem. News. 34 (2007) 131–144.

[93] J.D. Cowan, A mathematical theory of central nervous activity, (1967). http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.622820.

[94] M.K. Gupta, S. Gupta, R.K. Rawal, Impact of Artificial Neural Networks in QSAR and Computational Modeling, in: Artif. Neural Netw. Drug Des. Deliv. Dispos., Elsevier, 2016: pp. 153–179..

[95] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci. 26 (2007) 694–701.

[96] A. Tropsha, P. Gramatica, V. Gombar, The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models, QSAR Comb. Sci. 22 (2003) 69–77.

[97] L. He, P.C. Jurs, Assessing the reliability of a QSAR model's predictions, J. Mol. Graph. Model. 23 (2005) 503–523.

[98] L. Eriksson, J. Jaworska, A.P. Worth, M.T.D. Cronin, R.M. McDowell, P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs., Environ. Health Perspect. 111 (2003) 1361–75.

[99] A. Tropsha, P. Gramatica, V. Gombar, The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models, QSAR Comb. Sci. 22 (2003) 69–77.

[100] P. Pratim Roy, S. Paul, I. Mitra, K. Roy, P. Pratim Roy, S. Paul, I. Mitra, K. Roy, On Two Novel Parameters for Validation of Predictive QSAR Models, Molecules. 14 (2009) 1660–1701.

[101] L.M. Shi, H. Fang, W. Tong, J. Wu, R. Perkins, R.M. Blair, W.S. Branham, S.L. Dial, C.L. Moland, D.M. Sheehan, QSAR models using a large diverse set of estrogens., J. Chem. Inf. Comput. Sci. 41 (2001) 186–95. http://www.ncbi.nlm.nih.gov/pubmed/11206373.

[102] N. Chirico, P. Gramatica, Real External Predictivity of QSAR Models: How To

Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient, J. Chem. Inf. Model. 51 (2011) 2320–2335.

[103] T.M. Martin, P. Harten, D.M. Young, E.N. Muratov, A. Golbraikh, H. Zhu, A. Tropsha, Does rational selection of training and test sets improve the outcome of QSAR modeling?, J. Chem. Inf. Model. (2012).

[104] H. Hadni, M. Mazigh, E. Charif, A. Bouayad, M. Elhallaoui, Molecular Modeling of Antimalarial Agents by 3D-QSAR Study and Molecular Docking of Two Hybrids 4-Aminoquinoline-1,3,5-triazine and 4-Aminoquinoline-oxalamide Derivatives with the Receptor Protein in Its Both Wild and Mutant Types, Biochem. Res. Int. 2018 (2018) 1–15.

[105] T.I. Netzeva, A. Worth, T. Aldenberg, R. Benigni, M.T.D. Cronin, P. Gramatica, J.S. Jaworska, S. Kahn, G. Klopman, C.A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G.Y. Patlewicz, R. Perkins, D. Roberts, T. Schultz, D.W. Stanton, J.J.M. van de Sandt, W. Tong, G. Veith, C. Yang, Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52., Altern. Lab. Anim. 33 (2005) 155–73. http://www.ncbi.nlm.nih.gov/pubmed/16180989.

[106] D. Plewczynski, M. Łaźniewski, R. Augustyniak, K. Ginalski, Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database, J. Comput. Chem. 32 (2011) 742–755.

[107] P. Englebienne, N. Moitessier, Docking Ligands into Flexible and Solvated Macromolecules. 4. Are Popular Scoring Functions Accurate for this Class of Proteins?, J. Chem. Inf. Model. 49 (2009) 1568–1580.

[108] *,† Gregory L. Warren, ‡ C. Webster Andrews, # Anna-Maria Capelli, ∥ Brian Clarke, †,§ Judith LaLonde, ‡ Millard H. Lambert, ⊥,· Mika Lindvall, † Neysa Nevins, † Simon F. Semus, ⊥ Stefan Senger, # Giovanna Tedesco, ∥ Ian D. Wall, ⊥ James M. Woolven, † and Catherine E. Peishoff, M.S. Head†, A Critical Assessment of Docking Programs and Scoring Functions, (2005).

[109] D.B. Kitchen, H. Decornez, J.R. Furr, J. Bajorath, Docking and scoring in virtual screening for drug discovery: methods and applications, Nat. Rev. Drug Discov. 3 (2004) 935–949.

[110] D. Duhovny, R. Nussinov, H.J. Wolfson, Efficient Unbound Docking of Rigid Molecules, in: Springer, Berlin, Heidelberg, 2002: pp. 185–200.

[111] R.A. Coupez, B.; Lewis, Docking and Scoring - Theoretically Easy, Practically

Impossible?, Curr. Med. Chem. 13 (2006) 2995–3003.

[112] † Andrew R. Leach, ‡ and Brian K. Shoichet, § Catherine E. Peishoff*, Prediction of Protein−Ligand Interactions. Docking and Scoring: Successes and Gaps, (2006).

[113] I. Halperin, B. Ma, H. Wolfson, R. Nussinov, Principles of docking: An overview of search algorithms and a guide to scoring functions, Proteins Struct. Funct. Genet. 47 (2002) 409–443.

[114] B.D. Bursulaya, M. Totrov, R. Abagyan, C.L. Brooks III, Comparative study of several algorithms for flexible ligand docking, J. Comput. Aided. Mol. Des. 17 (2003) 755–763.

[115] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, C.R. Corbeil, Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go, Br. J. Pharmacol. 153 (2009) S7–S26.

[116] J.A. Erickson, M. Jalaie, D.H. Robertson, R.A. Lewis, M. Vieth, Lessons in Molecular Recognition: The Effects of Ligand and Protein Flexibility on Molecular Docking Accuracy, J. Med. Chem. 47 (2004) 45–55.

[117] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, J. Comput. Chem. 19 (1998) 1639–1662.

[118] B. Kramer, M. Rarey, T. Lengauer, Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking, Proteins Struct. Funct. Genet. 37 (1999) 228–241.

[119] H.J. Böhm, Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs., J. Comput. Aided. Mol. Des. 12 (1998) 309–323.

[120] I. Muegge, PMF scoring revisited, J. Med. Chem. 49 (2006) 5895–5902.

[121] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785–2791.

[122] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997) 727–748.

[123] A.N. Jain, Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine, J. Med. Chem. 46 (2003) 499–511.

[124] R. Thomsen, M.H. Christensen, MolDock: A new technique for high-accuracy molecular docking, J. Med. Chem. 49 (2006) 3315–3321.

[125] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—A Visualization System for Exploratory Research and Analysis, J Comput Chem. 25 (2004) 1605–1612.

[126] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy, J. Med. Chem. 47 (2004) 1739–1749.

[127] N. Metropolis, S. Ulam, The Monte Carlo Method, J. Am. Stat. Assoc. 44 (1949) 335–341.

[128] R. Huey, G.M. Morris, A.J. Olson, D.S. Goodsell, Software news and update a semiempirical free energy force field with charge-based desolvation, J. Comput. Chem. 28 (2007) 1145–1152.

[129] H. Gohlke, G. Klebe, Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors, Angew. Chemie Int. Ed. 41 (2002) 2644–2676.

[130] M. Mihăşan, Basic protein structure prediction for the biologist: A review, Arch. Biol. Sci. 62 (2010) 857–871.

[131] M. Mihăşan, What in silico molecular docking can do for the ‘bench-working biologists,’ J. Biosci. 37 (2012) 1089–1095.

[132] I.S. Moreira, P.A. Fernandes, M.J. Ramos, Protein-protein docking dealing with the unknown, J. Comput. Chem. 31 (2009) NA-NA.

Chapitre III: Molecular Modeling of Antimalarial Agents by 3D-QSAR Study and Molecular Docking of Two Hybrids 4-Aminoquinoline-1,3,5-triazine and 4-Aminoquinoline-oxalamide Derivatives with the Receptor Protein in Its Both Wild and Mutant Types

Hanine Hadni, Mohamed Mazigh, El'mbarki Charif, Asmae Bouayad, and Menana Elhallaoui

Engineering Materials, Modeling and Environmental Laboratory, Faculty of Sciences of DharElmehraz,

Sidi Mohammed Ben Abdellah University, B.P. 1796, Atlas, Fes, Morocco

Abstract:

Modeling studies by 3D-QSAR and molecular docking methods were performed on a set of 34 4-aminoquinoline derivatives previously studied as effective antimalarial agents of wild-type as well as quadruple mutant *Plasmodium falciparum dihydrofolaten reductase* (DHFR). So, the famous mathematical methods multiple linear regression (MLR) and artificial neural network (ANN) were explored to build the QSAR model. The DFT-B3LYP method with the basis set 6-31G was used to calculate quantum chemical descriptors, chosen to represent the electronic descriptors of molecular structures. On the other hand MM2 method was used to calculate lipophilic, geometrical, physicochemical and steric descriptors. The QSAR model tested first with ANN method shows high performance towards its predictability by using leave-one-out (LOO) cross-validation, which was confirmed secondly by external validation, and Y-randomization in the case of the MLR model. The molecular docking study of three compounds 9, 11 and 26 on both wild and quadruple mutant type of pf-DHFR-TS as the protein target helps to more understand and then to predict the binding modes with the binding sites.

I Introduction

Malaria is one of the world's greatest global public health challenges. It is most prevalent in sub-African, Asian, and South American countries, and it mostly affects children under the age of five and pregnant women [1, 2]. According to a world health organization (WHO) report 2015, estimated 3.2 billion people were at risk of malaria, approximately 212 million cases of malaria worldwide, and 429.000 deaths occurred worldwide in 2015. Of these

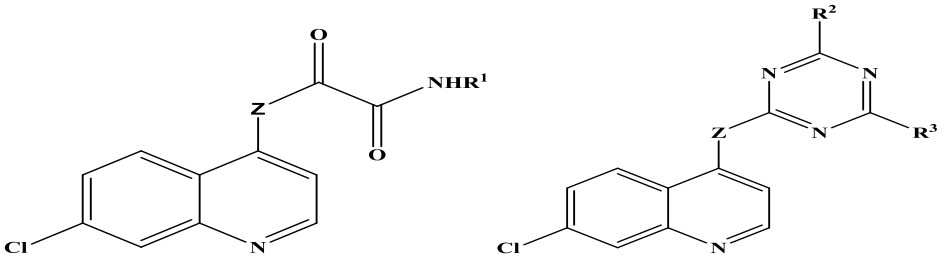
estimated deaths, 90% occurred in sub-Saharan Africa [1]. Malaria is an infectious and contagious disease caused by the protozoa of the genus *Plasmodium* [3]. There are five species that infect humans (*P. falciparum, P. vivax, P. malariae, P. ovale,* and *P. knowlesi*) [4]. However, amongst these five species, *Plasmodium falciparum* is the most severe and lethal species [5]. Many efforts are made in attempts to find out efficient inhibitors for this protein by testing several molecular structures. The quinoline moiety has attracted a great consideration of the medicinal chemists, as it is one of the crucial pharmacophores accountable for imparting antimalarial action [6, 7]. In the other hand, the 1,3,5-triazine derivatives cycloguanil, chlorcycloguanil, clociguanil, and WR99210 are already approved as effective *dihydrofolate reductase* (DHFR), specific inhibitor of *P. falciparum* domain, and they selectively inhibit biochemical processes that are vital for parasite growth [8]. Nowadays, to overcome drug resistance problems the concept of hybrid molecules has been introduced, in which two or more pharmacophores are linked together (as quinoline-triazine and qunoline-oxalamide), and it is believed that these compounds act by inhibiting simultaneously two conventional targets [9]. In this study, we worked on these two pharmacophores, as two types of hybrids: 4-aminoquinoline-triazine and 4-aminoquinoline-oxalamide [10].

The discovery of new antimalarial drugs is very challenging; the aim of developing a QSAR model is to construct a relationship (using statistical methods) between structural properties and activities using a training set which is capable of predicting the activity of compounds which are not used to build the model by multiple linear regression (MLR) and artificial neural network (ANN) calculations. The QSAR models have been validated by using internal and external validation as well as Y-randomization. To develop the binding modes of this set of hybrids in the active sites, we have performed molecular docking of three compounds; the highest active compound 26 belonging to the triasines series, the highest active compound 9 from the oxalamide series and the lowest active, of the entire series, compound 11, with *Plasmodium falciparum dihydrofolate reductase–thymidylate* synthase (PfDHFR-TS) in its two forms; the wild type and the quadruple mutant [11]. This study allows the developing of models that not only provide details of the binding modes and key molecular interactions, but also allow the prediction of relative inhibition and binding affinities that could be reproduced in silico.

II. Materials and Methods

II.1 Experimental Data

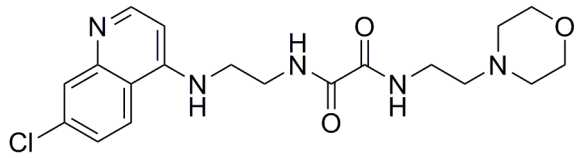
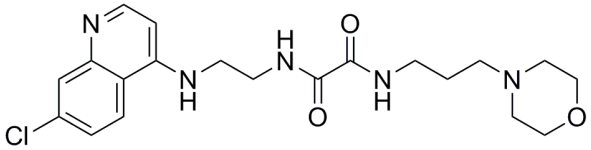
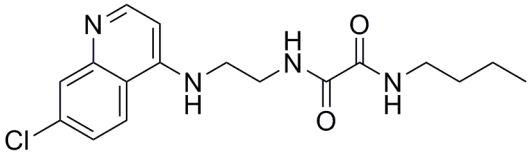
In this work, a data set of 34 hybrids of 4-aminoquinoline [10] constituting two groups is explored. The first group (4-aminoquinoline-oxalamides) accounts for 16 molecules numerated from 1 to 16, and the second group (4-aminoquinoline-triazines) contains 18 compounds numerated from 17 to 34 (**Figure 1**). The chemical structures of these hybrid derivatives with their antimalarial activities (IC50 (ng/ml)) are presented in Tables 1 and 2. The observations are converted into logarithm scale log IC50.

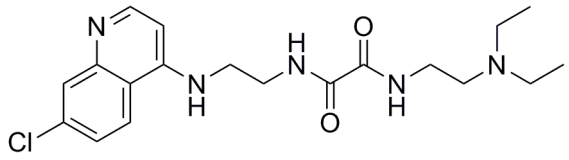
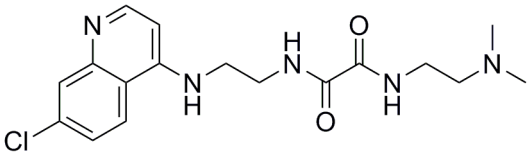
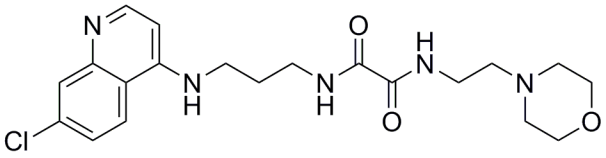
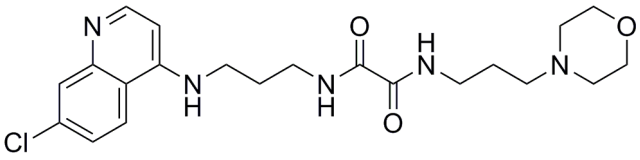
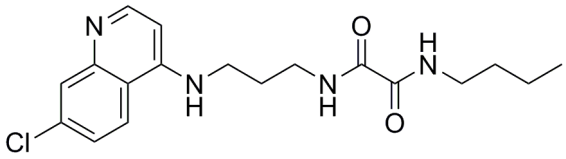
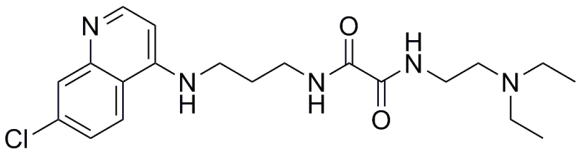
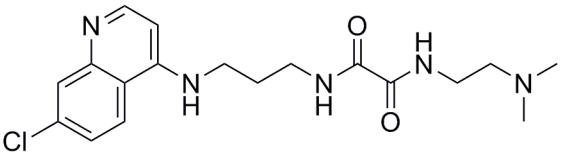
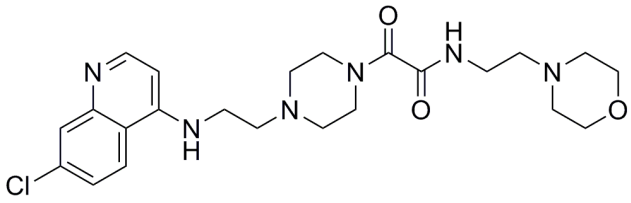
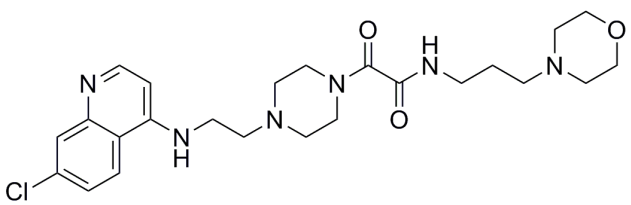


(a) ***(b)***

Figure 1: The general structure of 4-aminoquinoline-oxalamides (a) and 4-aminoquinoline-triazines (b).

Table 1: Chemical structures and activities of 4-aminoquinoline oxalamide and derivatives..

Compounds	Structure of compound	$LogIC_{50}$ (obs)
1		1.958
2		1.496
3		2.367

4		1.701
5		1.500
6		1.915
7		1.405
8		2.368
9		1.193
10		1.422
11		2.418
12		1.611

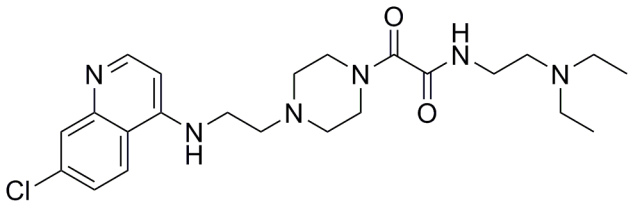
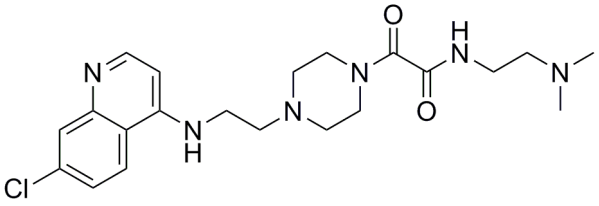
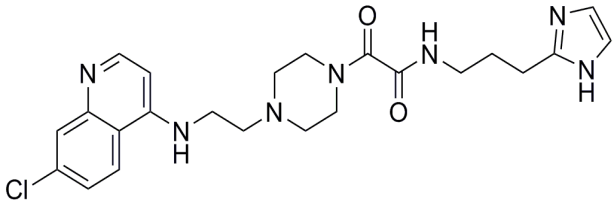
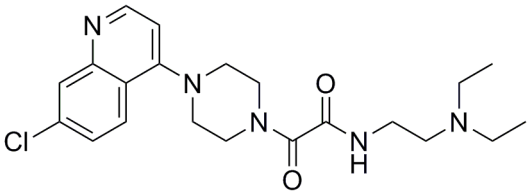
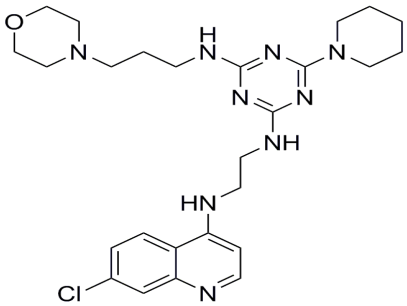
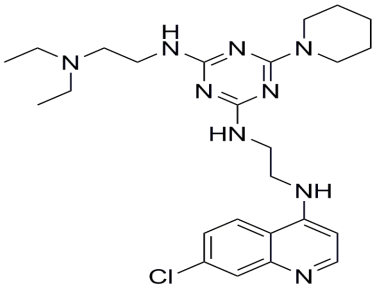
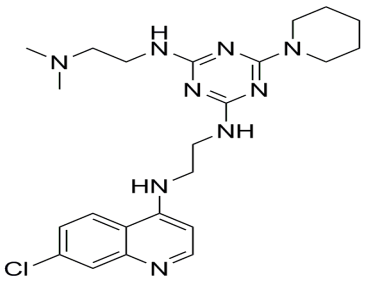
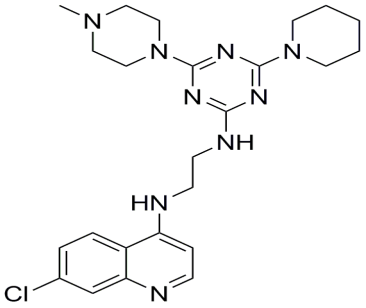
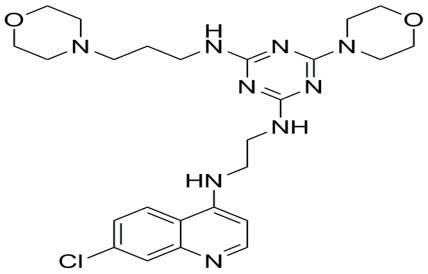
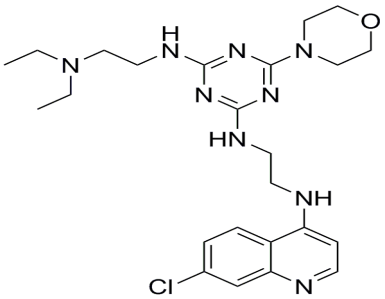
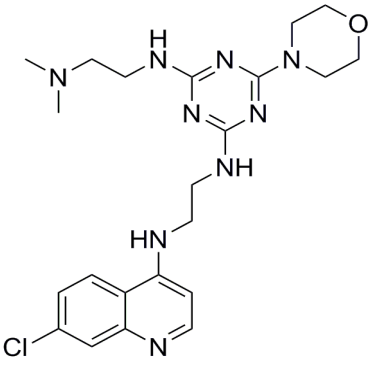
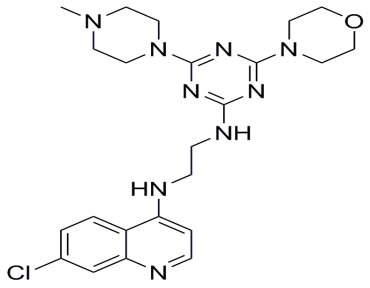
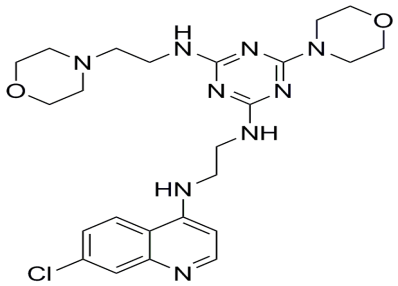
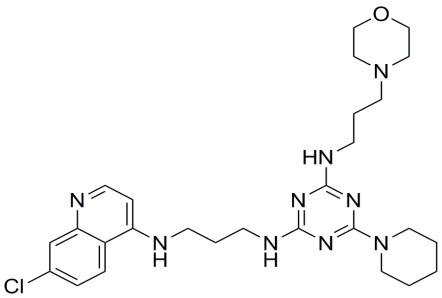
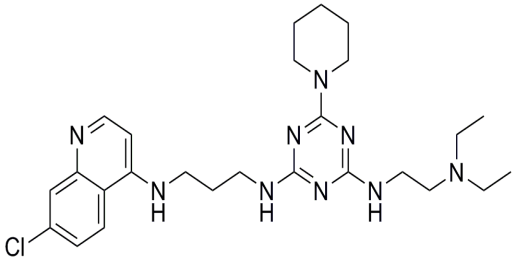
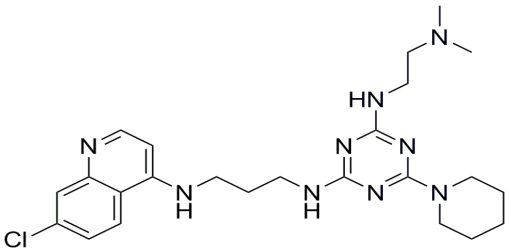
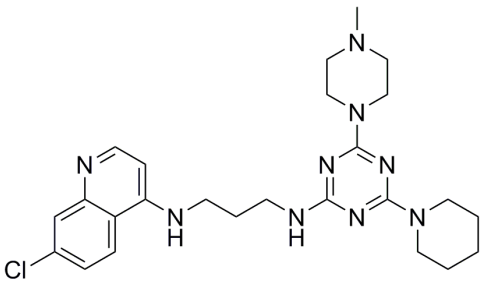
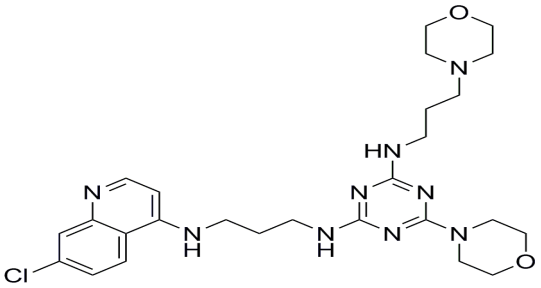
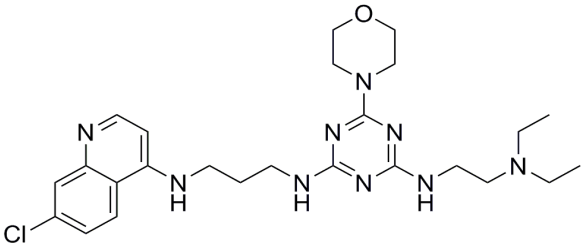
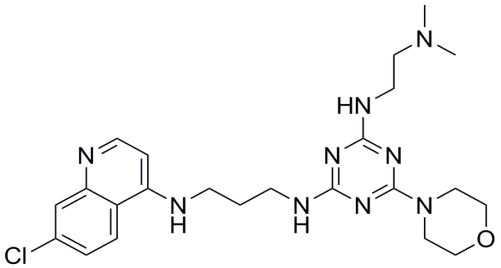
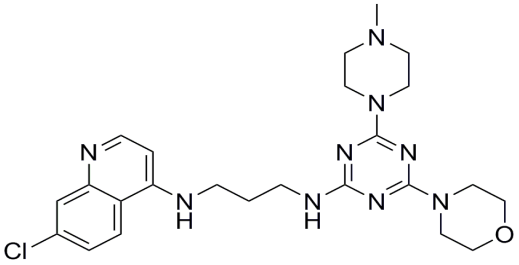
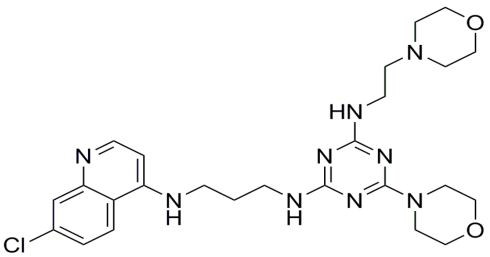
13		1.902
14		1.927
15		2.130
16		1.623

Table 2: Chemical structures and activities of the 4-aminoquinoline-triazine and derivatives.

Compounds	**Structure of compound**	**LogIC_{50} (obs)**
17		1.817
18		1.167

19		1.201
20		0.897
21		1.346
22		1.473
23		1.595

24		1.001
25		1.683
26		0.719
27		0.953
28		1.194

29		0.960
30		1.717
31		1.605
32		2.215
33		1.603
34		1.252

We note that in this study the training set included 28 compounds and the test set is constituted of 6 compounds.

II.2 Molecular Descriptors Calculation

In order to accurately model and predict inhibitors activities, 16 descriptors listed in Table 3 were introduced. 11 descriptors which are lipophilic, geometrical, physicochemical, and steric descriptors were calculated with the MM2 method with the aid of the ACD/ChemSketch program [12] and the ChemBioOffice software [13]. On the other hand, 5 electronic descriptors were calculated with the DFT method [14], using the Gaussian03 quantum chemistry package [15]. The optimization of compounds was performed with the DFT method using Becke's three-parameter hybrid function (B3LYP) [16], with a 6-31G basis set in the case of electronic descriptors calculation and with the MM2 method for the remaining descriptors. The totality of descriptors used in this work is represented in Table 3.

Table 3 : List of descriptors constituting the database.

Category of descriptors	**Electronic**	**Lipophilic**	**Geometrical**	**Physicochemical**	**Steric**
Name of the descriptors	HOMO energy (E_{HOMO}) LUMO energy (E_{LUMO}) Dipole moment (Dp) total energy (E) Repulsion energy (Er)	LogP (Octanol-water partition coefficient) VDW energy (E_{VDW})	Torsion (T) stretch-bend (SB)	Critical pressure (CP) Critical temperature (CT) Critical volume (CV)	Density (D) D = $\frac{MW}{MV}$ Wiener index (W) Molar volume (MV) Number of H-bonding substituents (Hs)

II.3 Analysis Methods

Multiple linear regression (MLR) [17] analysis with the descendent selection method was used to select the most appropriate descriptors. It is a mathematical technique to study the relation between one dependent variable and several independent variables. The regression method is based on three criteria: correlation of determination (*R*2), the Fisher ratio value (*F*), and the root mean square error (RMSE). The MLR model was generated using the software XLSTAT version 2013 [18]. We note that the MLR model has been served to select the used descriptors as the input parameters in the artificial neural network (ANN).

The ANN analysis is performed using the SAS JMP package (v8.0, SAS Institute Inc., Cary, NC, USA). The neurons networks are arranged in three layers: The input layer contains six neurons representing the relevant descriptors obtained with the MLR technique, the output

layer contains one neuron representing the calculated activities values log IC50, and the hidden layer is composed of 3 neurons determined by ρ = (number of weight)/(number of connection). In this work, we used the ρ value interval $1 < \rho < 3$ [19, 20].

Since a high-correlation coefficient only indicates how well the equations fit the data, in order to explore the stability of this equation; the cross-validation method with "leave-one-out" was carried out using the ANN method. Based on this technique, a number of modified data sets are created by deleting in each case one individual [21], and thereafter, the corresponding models serve to predict the activity of the removed compound.

The LOO cross-validation coefficient R^2 was calculated as follows [22]:

$$Rcv^2 = 1 - \frac{\sum_{i=1}^{n} (\mathrm{Yexp} - \mathrm{Ypred})^2}{\sum_{i=1}^{n} (Y\,\mathrm{exp} - Y)^2}$$

where Yexp and Ypred are the observed and predicted values for the dependent variables, respectively, and Y is the average observed value.

In order to ensure the reliability of the QSAR model, the Y-randomization test has been used. This approach consists to randomly mix many properties/experimental activities for the learning series using the same descriptors; the new QSAR model is constructed to exclude the possibility of random correlation in the obtained model [23].

Furthermore, external validation is necessary as the validation method is used to ensure the ability of the QSAR model. However, the data set in this work has been randomly divided into a training set with 28 compounds for the model developed through MLR, and a predicted set with 6 compounds has been reserved to external validation.

The ability of the built model based on the external prediction set was evaluated by R^2_{ext}, which could be calculated as follows [22]:

$$R^2_{ext} = 1 - \frac{\sum (\mathrm{Y}_{pred\,(test)} - Y_{(test)})^2}{\sum (Y_{test} - \overline{Y_{tr}})^2}$$

where $\mathrm{Y}_{pred\,(test)}$ and $Y_{(test)}$ are the predicted and experimental values of the samples for the prediction set, respectively. $\overline{Y_{tr}}$ is the average value for the dependent variable for the training set.

The value of $R^2_{ext} \geq 0.5$ is considered as an indicator of the reliability of the model. However, Golbraikh and Tropsha showed that R^2_{ext} is not a good parameter to estimate the reliability of the QSAR model. Indeed, an external validation based on the Golbraikh and Tropsha criteria is necessary [22].

In order to gain insight on the key structural requirements of the antimalarial activity, molecular docking studies are carried out using the AutoDock4.2 program [24]. X-ray crystallography structures of *Plasmodium falciparum* of wild type (coded as 1J3I.pdb) and quadruple mutant (coded as 1J3K.pdb) pf-DHFR-TS was obtained from the Protein Data Bank [25]. The minimized protein structures were defined as receptors, the first step in the preparation of the receptor was the removal of the ligands and the water molecules. In order to simplify the docking analysis, In this docking, the 3D grid was created by the AUTOGRID algorithm [26] to evaluate the interacting energy between protein-ligands. The grid maps were constructed using 60, 60, 60, pointing in x, y and z directions, with grid point spacing of 0.375 Å. The center grid box is of 29.39 Å, 5.56 Å, 52.49 Å, By the ligand location in the complex. Discovery Studio software was used for the 2D and 3D visualizations of the established interactions [27].

III. Results and Discussion

The selected descriptors values, and predicted activities values of 28 compounds composing the training set obtained by MLR, ANN, and CV methods, are summarized in Table 4.

Table 4: Values of the selected descriptors and the observed/predicted log IC50 values.

N	EHOMO	E	Er	T	SB	CT	logIC50	MIR	ANN	CV(Loo)
1	5.496	-46195.5	66323.4	-1.509	0.463	1170.7	1.958	1.730	1.993	1.901
2	5.447	-47264.9	68848	-1.354	0.485	1179.5	1.496	1.763	1.547	1.638
3	5.848	-40538.9	52297.9	-5.165	0.098	1107.9	2.367	2.393	2.343	2.363
4	5.531	-44182.8	63712.2	-0.391	0.397	1134.3	1.701	1.777	1.575	1.815
6	5.459	-47265	69290.9	-1.482	0.484	1179.5	1.915	1.744	1.849	1.730
7	5.41	-48334.4	71918.9	-1.405	0.505	1189.3	1.405	1.754	1.408	1.434
8	5.764	-41608.5	55210.6	-4.728	0.12	1115.3	2.368	2.358	2.381	2.328
10	5.709	-43113.3	58873.4	-1.405	0.505	1189.3	1.422	1.268	1.421	1.452
11	5.385	-51945.2	86354.4	1.319	0.73	1176.1	2.418	1.842	2.308	1.915
12	5.347	-53014.6	88771.4	1.425	0.753	1188.6	1.611	1.877	1.704	1.824
13	5.426	-49932.6	83218.3	2.558	0.664	1143.3	1.902	1.940	1.961	1.947
14	5.574	-47793.5	74926.6	-1.955	0.53	1119.3	1.927	1.985	1.942	2.093
16	5.465	-46288.2	74260.7	9.347	0.874	1098.8	1.623	1.708	1.591	1.632
17	5.391	-55513.4	103363.9	-4.042	0.744	1153.3	1.817	1.399	1.743	1.554
18	5.394	-52431.4	97285.7	-4.294	0.685	1107.7	1.167	1.003	1.148	1.088
19	5.483	-50292.4	88485.6	-7.332	0.519	1079.5	1.201	1.291	1.260	1.246
21	5.351	-56489.4	105348.6	-2.732	0.826	1157.2	1.346	1.448	1.253	1.517
22	5.475	-53407.4	97851.6	-2.68	0.753	1111.5	1.473	1.563	1.516	1.606
23	5.547	-51268.4	88785.8	-6.024	0.603	1083.3	1.595	1.727	1.723	1.780
24	5.675	-52305	94806.5	-6.472	0.736	1103	1.001	1.386	0.984	0.970
25	5.363	-55420	102004.1	-2.701	0.801	1142.4	1.683	1.460	1.661	1.506
26	5.28	-56582.9	108704.5	-4.016	0.768	1169.1	0.719	0.989	0.737	0.847
28	5.461	-51361.9	92126.6	-7.315	0.544	1093.1	1.194	1.246	1.194	1.199
29	5.574	-52398.5	97419.4	-7.874	0.675	1111.5	0.96	0.921	1.027	1.146
30	5.425	-57558.9	107184.7	-2.663	0.849	1173	1.717	1.786	1.669	1.630
31	5.438	-54476.9	101111	-2.977	0.787	1127.1	1.605	1.419	1.593	1.525
32	5.538	-52337.9	92059.3	-6.003	0.626	1096.9	2.215	1.751	2.098	1.777
34	5.371	-56489.5	105092	-2.767	0.822	1157.2	1.252	1.528	1.429	1.600

III.1 Multiple Linear Regression

The QSAR model corresponding to the training set built using the MLR method is represented by the following equation:

$$LogIC_{50} = -17.79 + 2.54*E_{HOMO} - 310^{-4}*E - 0.810^{-4}*Er + 0.146*T - 3.47*SB - 6.3410^{-3}*CT$$

$$N = 28 \qquad R = 0.83 \qquad R^2 = 0.69 \qquad F = 57.86 \qquad RMSE = 0.27$$

Where N is the number of compounds, *R* is the correlation coefficient, R^2 is the determination coefficient, RMSE is the root mean square error, and *F* is the Fisher test. The relevant descriptors involved in the MLR model are HOMO energy, total energy, repulsion energy, torsion, critical temperature, and stretch-bend. The corresponding normalized coefficients are

presented in **Figure 2**, and the correlation of the observed activities with the MLR calculated ones is illustrated in **Figure 3**.

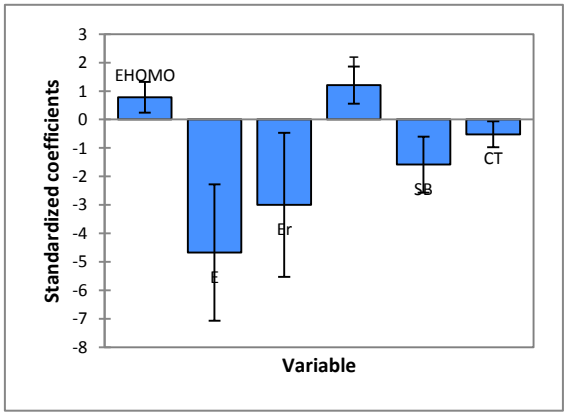


Figure 2: Modeling characterization by the normalized coefficients.

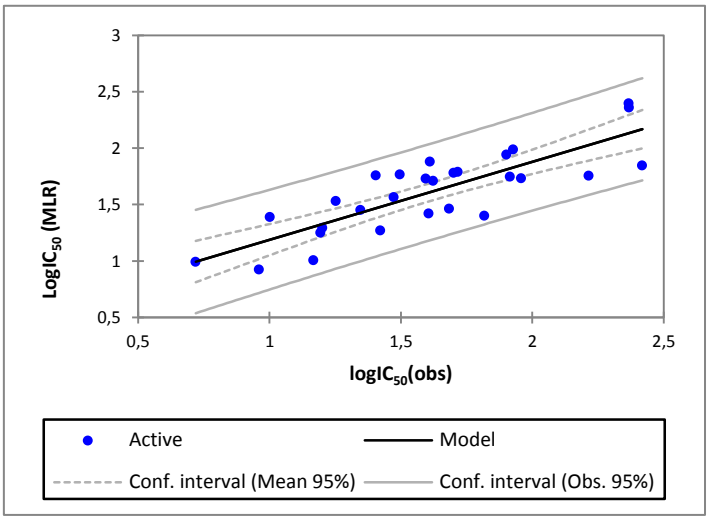


Figure 3: Correlation of observed and predicted activities calculated using the MLR method.

As indicated by the statistical coefficient values of the correlation between the observed and calculated activities based on this model using the training set are quite significant, and the low RMSE indicates that the model is reliable to a better prediction precision.

III.2 Artificial Neural Networks

In order to increase the probability of good characterization of studied compounds, artificial neural networks (ANN) are used as the nonlinear method to generate predictive nonlinear model between observed antimalarial activities values and the set of molecular descriptors obtained by MLR with that of the architecture network (6-3-1).The correlation of the observed activities with the ANN predicted ones is illustrated graphically in **Figure 4**.

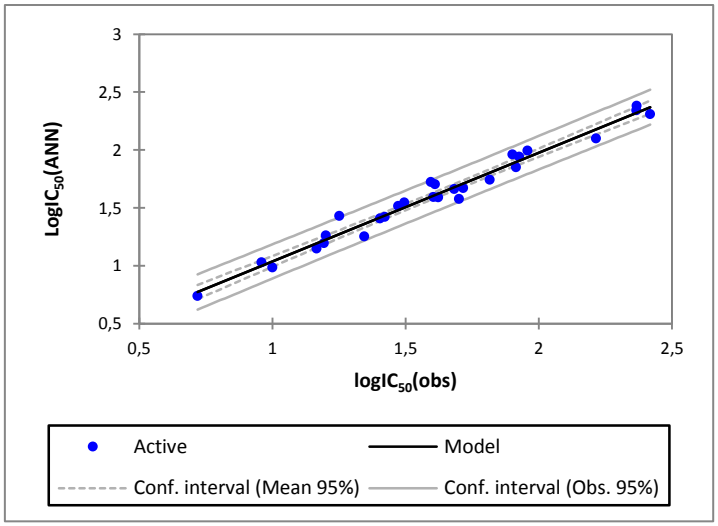


Figure 4: Correlation of observed and predicted activities calculated using ANN.

As it is shown in **figure 4**, a good correlation between observed antimalarial activities values and predicted activities by ANN is obtained, in fact the correlation coefficient R = 0.98, the determination coefficient R^2 = 0.97, and the Standard Error of Estimate RMSE = 0.09. Such results show that the selected descriptors by MLR are pertinent, the ANN model possesses a significantly statistical quality and the model proposed to predict antimalarial activity is relevant.

III.3 Cross validation

The QSAR model proposed to predict the activity of new compounds should be tested. To validate our results, we used the LOO procedure, which involves removing a single molecule from the set containing 28 molecules and making a prediction for antimalarial activity. This procedure is repeated 28 times in order to estimate the predictive ability of such models. The correlation of the observed activities with the calculated Cross-validation ones is shown graphically in **figure 5**.

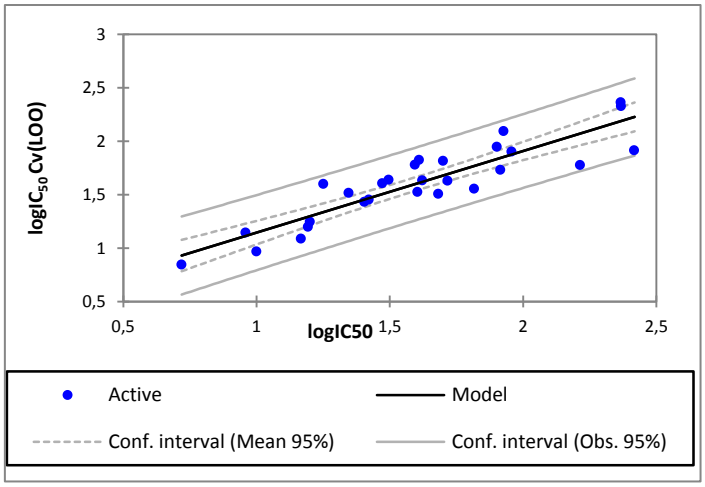


Figure 5: Correlation of observed and predicted activities calculated using CV (LOO).

The obtained correlation (R=0.90, R^2=0.81 and RMSE=0.16) shows a high predictive power of the MLR model. This result shows that our QSAR model is not sensitive to this operation of putting a molecule aside and putting it back into the learning series. This is a first indication of the stability of the selected QSAR model.

III.4 Y-randomization

The Y-randomization test was performed to make sure that there is no random correlation. In this way, we test the validity of established QSAR model and check that the selected descriptors are not random, and consequently; the result model should have correct statistical coefficients. So, this method is applied on a set of 28 compounds randomly chosen from the entire series. The results of y-randomization method are presented in the table 5 and illustrated in the **figure 6.**

Table 5: Comparison between observed and predicted activities obtained using Y-randomization method.

N	3	11	19	26	29	31	33	2	7	9	18	20	22	24
$logIC_{50}$	2.367	2.418	1.201	0.719	0.96	1.605	1.603	1.496	1.405	1.193	1.167	0.897	1.473	1.001
$predLogIC_{50}$	2.124	1.813	1.294	0.834	0.882	1.439	1.458	1.637	1.604	1.706	1.009	0.843	1.612	1.442
N	30	27	34	8	14	15	4	1	12	5	13	23	25	17
$logIC_{50}$	1.717	0.953	1.252	2.368	1.927	2.13	1.701	1.958	1.611	1.5	1.902	1.595	1.683	1.817
$predLogIC_{50}$	1.706	1.125	1.489	2.084	1.982	2.259	1.686	1.614	1.828	1.646	1.932	1.800	1.470	1.300

The new QSAR model built using the y-randomization method is represented by the following Eq.4:

$$logIC_{50} = -13.6+2,32*E_{HOMO}-4.5E^{-4}*E-0.9E^{-4}*Er+0,141*T-2,43*SB-9.33E^{-3}*CT$$

N = 28 R = 0.825 R^2 = 0.682 F = 7.51 RMSE = 0.29

The correlation coefficient value of the mixture samples is close to that obtained by applying the model by the training set. This result provides the absence of dependence between descriptors included in QSAR model.

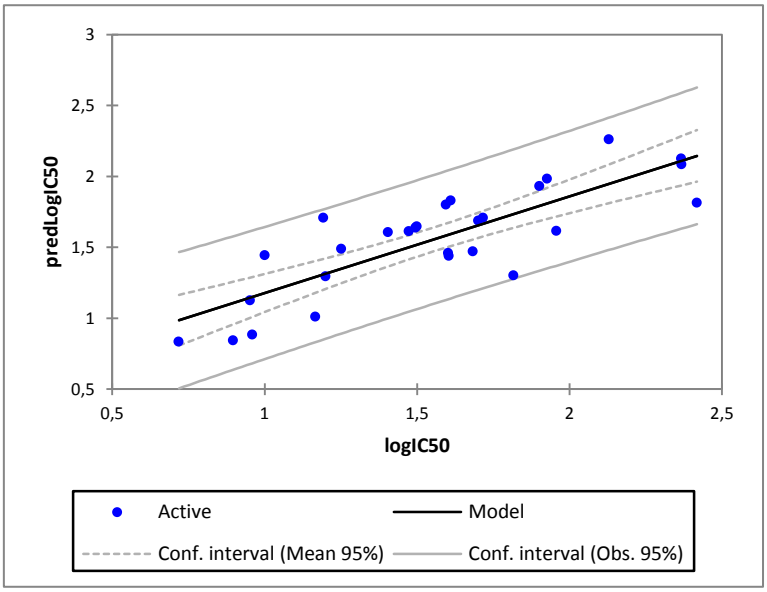


Figure 6 : Correlation between observed/predicted by the Y-randomization test.

III.5 External Validation

In a study on efficient methods of validation for QSAR models Golbraikh and Tropsha showed that LOO methods are necessary but not sufficient [22], claiming that external validation is inevitable and proposed some criteria which help to validate a QSAR model. This validation is done in two steps: 1) The prediction of the six compounds activities constituting the test set using the obtained training set models, MLR and ANN models (Table 6). 2) Verification of the Tropsha criteria's (Table 7).

Table 6: Comparison between experimental and predicted log IC50 values of the external test set from the MLR and ANN models.

N	$logIC_{50}$	Pred($logIC_{50}$) MLR	Residuals	Pred($logIC_{50}$) ANN	Residuals
5	1.5	1.565	-0.04	1.704	0.0478
9	1.193	1.589	0.467	1.799	-0.337
15	2.13	2.888	0.291	2.447	0.067
20	0.897	0.517	-0.138	1.138	-0.09
27	0.953	0.842	0.098	1.038	0.048
33	1.603	1.089	-0.678	1.566	0.264

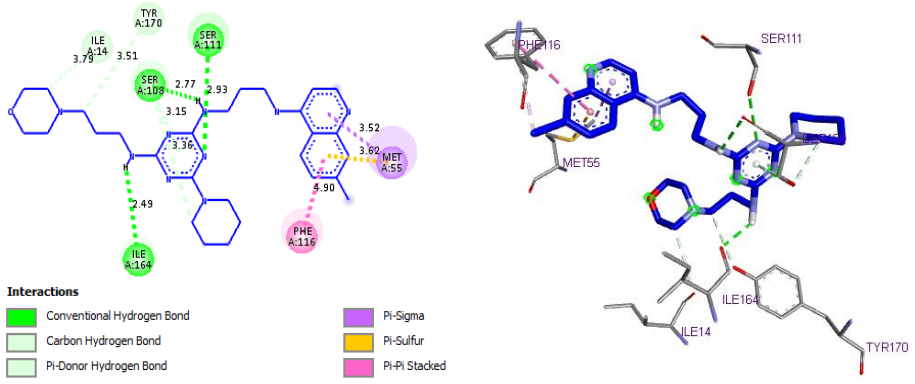
Table 7: Calculation of Golbraikh and Tropsha criteria.

Parameter	*Formula*	*Threshold*	Modelscore
R^2_{ext}	$R^2_{ext} = 1 - \frac{\sum(Y_{pred(test)} - Y_{(test)})^2}{\sum(Y_{(test)} - \bar{Y}_{tr})^2}$	R^2_{ext}>0.6	0.75
r^2	Coefficient of determination for the plot of predicted versus observed for test set	r^2 >0.6	0.77
r_0^2	r2 at zero intercept		0.71
$r_0'^2$	r2 for the plot of observed versus predicted activity for the test set at zero intercept		0.59
$\lvert r_0^2 - r_0'^2 \rvert$		$\lvert r_0^2 - r_0'^2 \rvert < 0.3$	0.12
k	Slope of the plot of predicted versus observed activity for test set at zero intercept	0.85<k<1.15	1.07
$\frac{r^2 - r_0^2}{r^2}$		$\frac{r^2 - r_0^2}{r^2} < 0.1$	0.07
k'	Slope of the plot of observed versus predicted activity at zero intercept	0.85<k'<1.15	0.92
$\frac{r^2 - r_0'^2}{r^2}$		$\frac{r^2 - r_0'^2}{r^2} < 0.1$	0.06

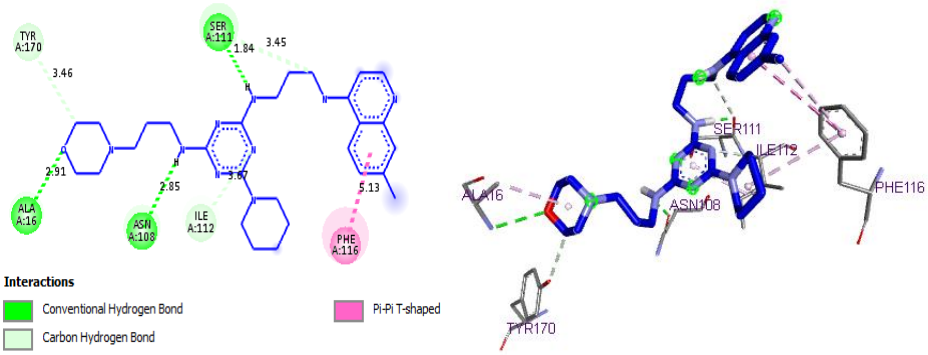
The results show that Golbraikh and Tropsha criteria's are successfully validated. All validations indicate that the built QSAR model is robust and satisfactory. The model established in this study meets all of the principles for QSAR validation and can be used to predict the antimalarial activity.

III.6 Docking studies

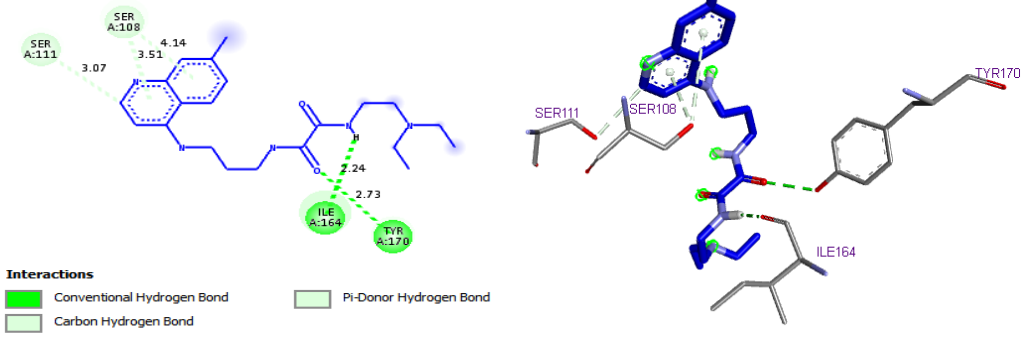
In a pioneering study on the binding modes and the localization of the principal active sites in wild and mutant protein performed with a potent inhibitor 1,3,5-triazine derivative which is a pre-clinical molecule called WR99210, it is found that the important sites are located in : Ile14, Ala16, Met55, Asp54, Ser108, Ile164 and Tyr170 in the case of wild type, and Ala16, Cys50, Asn51, Cys59, Asn108, Leu164 and Tyr170 in the case of mutant protein [28]. In a tentative to give insight on the interaction modes and to find out the interaction types established with this protein (Pf-DHFR-TS) in its two forms, wild and mutant, the molecular docking study performed in this work is applied on three compounds 9 (IC_{50}=15.58) 11(IC_{50}=261.84) and 26 (IC_{50}=5.23) with the binding sites of both wild type and quadruple mutant. The docking results and docked conformations of ligands in the active sites are represented in **Figure 7.**


TYR A:170
ILE A:14
SER A:111
SER A:108
MET A:55
PHE A:116
ILE A:164
3.79
3.51
2.77
2.93
3.15
3.36
3.52
3.62
4.90
2.49
Interactions
Conventional Hydrogen Bond
Carbon Hydrogen Bond
Pi-Donor Hydrogen Bond
Pi-Sigma
Pi-Sulfur
Pi-Pi Stacked
PHE116
SER111
MET55
ILE164
ILE14
TYR170

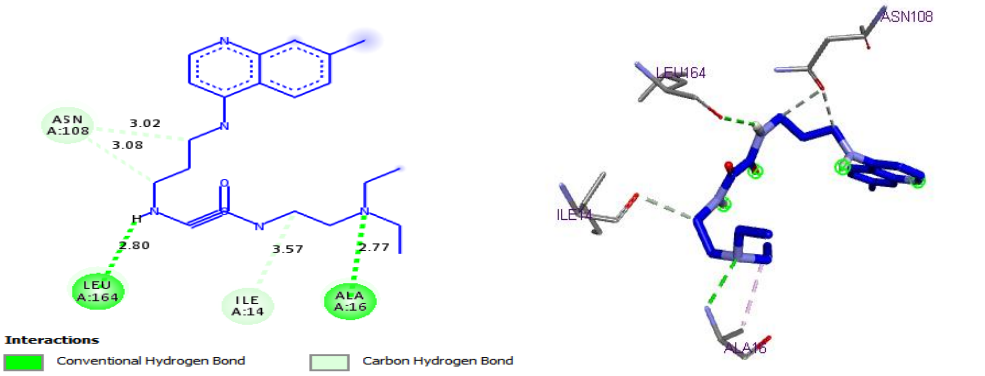
(a)


TYR A:170
SER A:111
ALA A:16
ASN A:108
ILE A:112
PHE A:116
3.46
1.84
3.45
2.91
2.85
3.67
5.13
Interactions
Conventional Hydrogen Bond
Carbon Hydrogen Bond
Pi-Pi T-shaped
SER111
ILE112
PHE116
ALA16
ASN108
TYR170

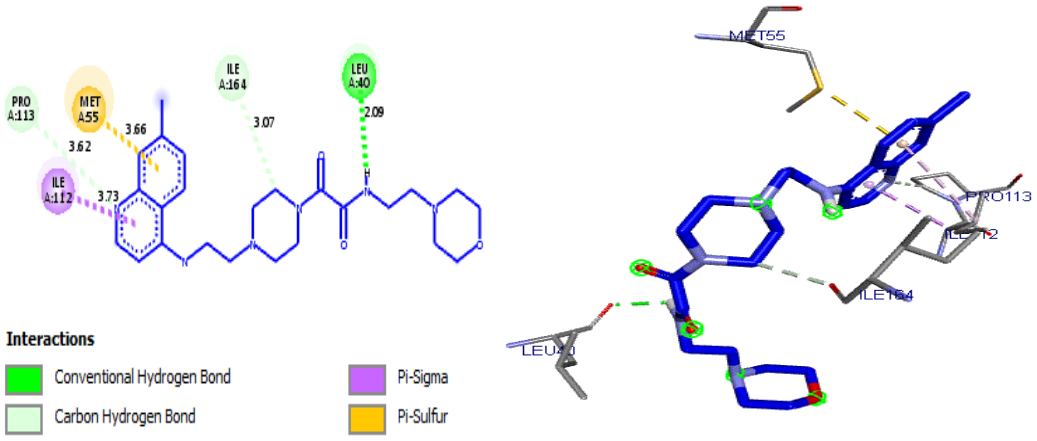
(b)


SER A:108
SER A:111
ILE A:164
TYR A:170
4.14
3.51
3.07
2.24
2.73
Interactions
Conventional Hydrogen Bond
Carbon Hydrogen Bond
Pi-Donor Hydrogen Bond
SER111
SER108
TYR170
ILE164

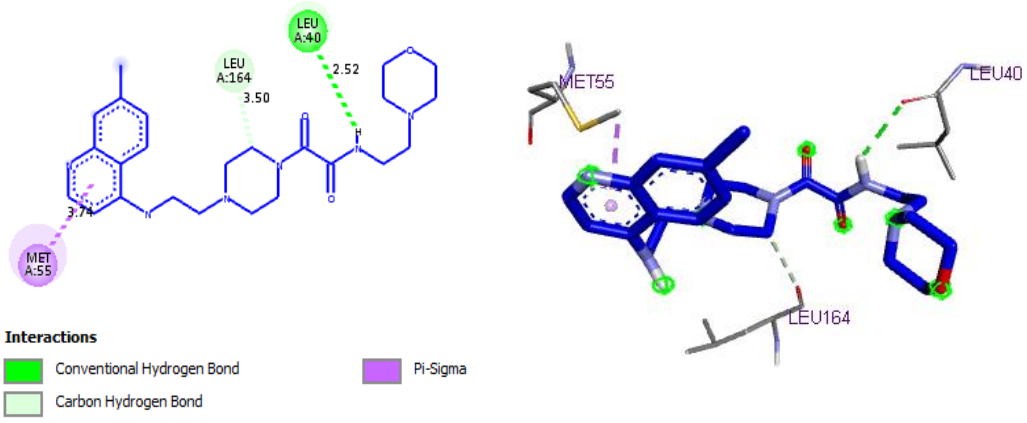
(c)



(d)



(e)



(f)

FIGURE 7: 2D and 3D docking poses showing interactions of compounds 9, 26, and 11 in the binding sites of wild type and quadruple mutant of pf-DHFR-TS. (a) Compound 26: wild type of pf-DHFR (binding energy −8.9 kcal/mol). (b) Compound 26: quadruple mutant of pf- DHFR-TS (binding energy −8.3 kcal/mol). (c) Compound 9: wild type of pf-DHFR (binding energy −7.42 kcal/mol). (d) Compound 9: quadruple mutant of pf-DHFR-TS (binding energy −7.96 kcal/mol). (e) Compound 11: wild type of pf-DHFR (binding energy −8.58 kcal/mol). (f) Compound 11: quadruple mutant of pf-DHFR-TS (binding energy −9.2 kcal/mol)

In the case of the wild type, compound 26 performs hydrogen bonding with the carboxylate oxygen atoms of ILE164, SER108 and SER111 by the involvement of the two NH groups bounded to the triazine group and one of triazine nitrogens, with respectively the distances 2.49 Å, 2.77 Å and 2.93 Å, and a non-bonded p-sigma interaction between phenyl of quinoline with MET55 at a distance of 3.52 Å. Whereas in the case of the quadruple mutant, three-hydrogen bond with ASN108, SER111 and ALA16 were observed by the involvement of two NH groups linked to 1,3,5-triazine and the oxygen of the morpholino group with a distance of 2.85 Å, 1.84 Å and 2.91 Å, respectively. For the compound 9, two hydrogen bonds are formed between an oxygen and an nitrogen of the oxalamide group with ILE164, TYR170 with respectively the distances 2.24 Å and 2.73 Å in the case of wild type. But in the case of the quadruple mutant, it forms two hydrogen bonds with ALA16 and LEU164 through the involvement of two nitrogens (the first is linked to the oxalamide group and the second belongs to the diethylamine group), with distances of 2.77 Å and 2.80 Å respectively. However, compound 11 showed only one hydrogen bonding interaction with LEU40 in both cases.

In the analysis of these results we have at first observed that the residues with that the compounds 26 and 9 have formed their interactions are mentioned as the most important binding sites for antimalarial activity [28], which is not the case for the compound 11. Secondly we observed that the number of hydrogen bonds differs from the most active compound which belongs to the triasine family, to the less active compound which belongs to the oxalamide family. So, this could explain the potent antimalarial activity for the compound 26 and the importance of triazine group to enhance the antimalarial activity compared to the oxalamide group

IV. CONCLUSION

The present study on a series of 4-aminoquinoline-triazines and 4-aminoquinoline-oxalamides was carried out using 3D-QSAR and docking techniques in the aim to predict the antimalarial activity. Group contribution method (for both training and test set) was used to develop a reliable QSAR model for predicting antimalarial activity. The result of MLR and ANN methods using the training set clearly show a strong relationship between the structural properties and the activity. Thus, the correlation coefficient for both methods shows good predictive ability of model. The model is validated by internal and external validation methods including (leave-one-out) cross-validation, y-randomization. The obtained model shows good quality of the robustness to predicting the antimalarial activity. The observed

activity was further corroborated via molecular docking study which gave explanation to the differences observed of activities of compounds especially between triazine family and oxalamide one. Results of these studies provided details of the predicted binding modes and the key molecular interactions. These will provide opportunities for medicinal chemists to develop new antimalarial drugs, by using new hybrid molecules.

References

[1] World Health Organization., World Malaria Report 2015., World Health Organization, 2016.

[2] S. Manohar, M. Tripathi, D.S. Rawat, 4-aminoquinoline based molecular hybrids as antimalarials: an overview., Curr. Top. Med. Chem. 14 (2014) 1706–33. http://www.ncbi.nlm.nih.gov/pubmed/25116580.

[3] J.B. Vaishnani, Cutaneous findings in five cases of malaria., Indian J. Dermatol. Venereol. Leprol. 77 (2011) 110.

[4] M.C. Murray, M.E. Perkins, Chapter 15. Chemotherapy of Malaria, Annu. Rep. Med. Chem. 31 (1996) 141–150.

[5] P. Newton, BM, DPhil, MRCP, N. White, DSc, MD, FRCP, MALARIA: New Developments in Treatment and Prevention, Annu. Rev. Med. 50 (1999) 179–192.

[6] N. Boechat, M. de L.G. Ferreira, L.C.S. Pinheiro, A.M.L. Jesus, M.M.M. Leite, C.C.S. Júnior, A.C.C. Aguiar, I.M. de Andrade, A.U. Krettli, New Compounds Hybrids 1 *H* -1,2,3-Triazole-Quinoline Against *Plasmodium falciparum*, Chem. Biol. Drug Des. 84 (2014) 325–332.

[7] A. Mishra, H. Batchu, K. Srivastava, P. Singh, P.K. Shukla, S. Batra, Synthesis and evaluation of new diaryl ether and quinoline hybrids as potential antiplasmodial and antimicrobial agents, Bioorg. Med. Chem. Lett. 24 (2014) 1719–1723.

[8] M.D. Hastings, C.H. Sibley, Pyrimethamine and WR99210 exert opposing selection on dihydrofolate reductase from Plasmodium vivax, Proc. Natl. Acad. Sci. 99 (2002) 13137–13141.

[9] F.W. Muregi, A. Ishih, Next-generation antimalarial drugs: hybrid molecules as a new strategy in drug design, Drug Dev. Res. 71 (2009) n/a-n/a.

[10] N. Sunduru, M. Sharma, K. Srivastava, S. Rajakumar, S.K. Puri, J.K. Saxena, P.M.S. Chauhan, Synthesis of oxalamide and triazine derivatives as a novel class of hybrid 4-aminoquinoline with potent antiplasmodial activity, Bioorganic Med. Chem. 17 (2009) 6451–6462.

[11] W. Sirawaraporn, T. Sathitkul, R. Sirawaraporn, Y. Yuthavong, D. V Santi, Antifolate-resistant mutants of Plasmodium falciparum dihydrofolate reductase., Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 1124–9. http://www.ncbi.nlm.nih.gov/pubmed/9037017.

[12] ACD/ChemSketch for Academic and Personal Use :: ACD/Labs.com, (n.d.). http://www.acdlabs.com/resources/freeware/chemsketch/.

[13] ChemOffice Professional – PerkinElmer Informatics Logiciels, (n.d.). http://www.cambridgesoft.com/Ensemble_for_Chemistry/ChemOffice/ChemOfficeProfessional/.

[14] R.G. Parr, W. Yang, Density-Functional Theory of Atoms and Molecules, 1989.

[15] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J. Montgomery J. A., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. V Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision C.02, 2004.

[16] A.D. Becke, Density-functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648–5652.

[17] R.B. Darlington, Regression and linear models, McGraw-Hill, 1990.

[18] XLSTAT Logiciel de statistique pour MS Excel | Logiciel statistique pour Excel, (n.d.). https://www.xlstat.com/fr/.

[19] S.S. So, W.G. Richards, Application of neural networks: quantitative structure-activity relationships of the derivatives of 2,4-diamino-5-(substituted-benzyl)pyrimidines as DHFR inhibitors, J. Med. Chem. 35 (1992) 3201–3207.

[20] T.A. Andrea, H. Kalayeh, Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors, J. Med. Chem. 34 (1991) 2824–2836.

[21] B. Efron, Estimating the error rate of a prediction rule: Improvement on cross-validation, J. Am. Stat. Assoc. 78 (1983) 316–331.

[22] A. Golbraikh, A. Tropsha, Beware of q2!, J. Mol. Graph. Model. 20 (2002) 269–276.

[23] A. Tropsha, Best practices for QSAR model development, validation, and exploitation, Mol. Inform. 29 (2010) 476–488.

[24] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785–2791.

[25] RCSB PDB: Homepage, (n.d.). http://www.rcsb.org/ .

[26] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, J. Comput. Chem. 19 (1998) 1639–1662.

[27] D.S. BIOvIA, Discovery studio modeling environment., San Diego, Dassault Syst. Release, 4. (2015) 98–104.

[28] J. Yuvaniyama, P. Chitnumsub, S. Kamchonwongpaisan, J. Vanichtanankul, W. Sirawaraporn, P. Taylor, M.D. Walkinshaw, Y. Yuthavong, Insights into antifolate resistance from malarial DHFR-TS structures, Nat. Struct. Biol. 10 (2003) 357–365.

CHapitre IV : QSAR and Molecular docking studies of 4-anilinoquinoline-triazine hybrids as *pf-DHFR* inhibitors

Hanine HADNI, Mohamed Mazigh, and Menana ELHALLAOUI

Engineering Materials, Modeling and Environmental Laboratory, Faculty of Sciences of DharElmehraz, Sidi Mohammed Ben Abdellah University, B.P. 1796, Atlas, Fes, Morocco

Abstract: A quantitative structure-activity relationship (QSAR) investigation was performed towards 41 hybrids of 4-anilinoquinoline-triazines as potential antimalarial agents. The study was carried out using descendant multiple linear regression analyses (MLR), and artificial neural networks (ANN). Quantum chemical descriptors were calculated using DFT-B3LYP method, with the basis set 6-31G. The values obtained for the correlation coefficient of 0.87 and 0.92 by MLR and ANN, respectively, show a good predictive quality of the established model. In addition, the predicted model has been confirmed by several validation methods such as leave-one-out (LOO) cross-validation, Y-randomization, and external validation. The observed activity and the structural features of the studied molecules were further highlighted by molecular docking study on both wild and quadruple mutant type of *pf-DHFR* protein. Furthermore, the present work deals to study the binding modes and the key protein-ligand interactions. This methodology will be used to design new antimalarial drugs.

Keywords: antimalarial activity; QSAR; DFT; MLR; ANN; *pf-DHFR* protein; Molecular docking.

I. Introduction

Despite the scientific advances made in medicinal chemistry. Nowadays, the malaria disease continues to cause the death of huge numbers of human beings. According to the World malaria report, 2018 draws on data from 90 countries and areas, there were an estimated 219 million cases and 435 000 related deaths in 2017 [1,2]. Malaria is caused by protozoa of the genus *Plasmodium* [3]. In fact, we can distinguish five species that infect humans, *P. falciparum*, *P. vivax*, P. *malariae*, *P. ovale*, and *P. Knowlesi* [4]. Among all five species, *Plasmodium falciparum* is the most severe and deadly type [5]. For this reason, the development of new drugs able to fight malaria is still in great interest. Indeed, 1,3,5-triazine derivatives as cycloguanil, chlorcycloguanil, clociguanil, WR99210 are already approved as effective *dihydrofolate reductase* (DHFR) inhibitors. They inhibit selectively the biochemical processes that are vital for parasite growth [6]. However, the development of the resistance to antimalarial drugs such as chloroquine, amodioquine, artemisinin, and antifolates becomes a serious health concern [7]. To overcome this problem, the concept of hybrid molecules has been introduced as one of the most used solutions, in which two or more pharmacophores are linked together and act by inhibiting simultaneously two conventional targets [8]. Herein, we

report the molecular modeling of 4-anilinoquinoline-triazine derivatives [9]. In order to pursue, our ongoing research on molecular modeling of antimalarial activity [10]. We aim in this study to develop a predictive QSAR model, which will be used to analyze the antimalarial activity of a series of 41 hybrid 4-anilinoquinoline-triazine derivatives (table 1). The proposed quantitative models are relying on the multivariate statistical analysis of experimental results published previously [11]. The MLR and ANN methods have served to establish a predictive model of antimalarial activity. Moreover, internal and external validations, as well as Y-randomization methods, were used to test the reliability of the built model. In addition, to develop geometrical model and to analyze the interactions of the hybrid system with the active sites of the protein, we performed the docking of two compounds, 11 and 38 which are two position isomers (meta and para), with *Plasmodium Falciparum Dihydrofolate Reductase–Thymidylate Synthase* (PfDHFR-TS) in its two forms; the wild type and the quadruple mutant[12].

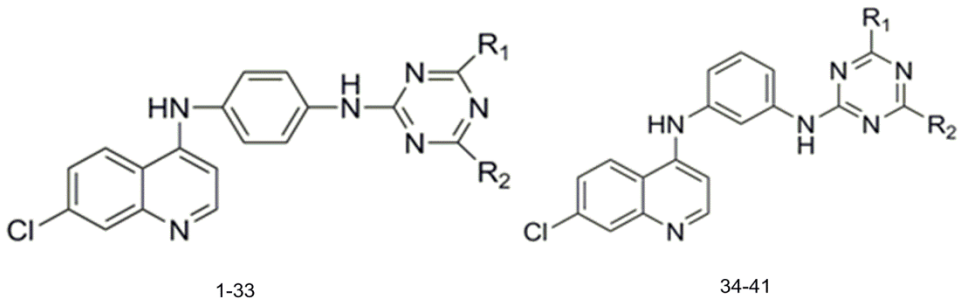


Figure 1: The structures of para and meta 4-anilinoquinoline-triazine derivatives.

II. Materials and Methods

II.1 Experimental Data

The experimental data of 4-anilinoquinoline-triazine derivatives are collected from recent publication [9], The observed activities (IC_{50} (nM)) are converted into logarithm scale $logIC_{50}$ and listed in table 1. The studied set is subdivided in two groups numerated from 1 to 33 and from 34 to 41, where the triazine frameworks are linked to the anilinoquinoline pattern via nitrogen atom in para and meta position respectively. The general structure of the 4-anilinoquinoline triazine hybrids is represented in **figure1**.

Table 1: The observed activities logIC50 of studied compounds.

Compounds	R_1	R_2	$LogIC_{50}$ (obs)
1	piperidino	N-methylpiperazino	1.541
2	piperidino	N-ethylpiperazino	1.473
3	piperidino	N.N-diethylethylenediamino	1.588
4	piperidino	4-(3-aminopropyl)morpholino	1.364
5	piperidino	Cyclohexylamino	2.495
6	piperidino	Benzylamino	2.370
7	anilino	N-methylpiperazino	0.767
8	anilino	N-ethylpiperazino	0.772
9	anilino	n-butylamino	1.267
10	anilino	N.N-diethylethylenediamino	1.781
11	anilino	4-(3-aminopropyl)morpholino	0.478
12	anilino	Cyclohexylamino	1.811
13	anilino	Benzylamino	1.526
14	anilino	Amino	1.618
15	anilino	Methylamino	1.279
16	o-toluidino	N-methylpiperazino	1.275
17	o-toluidino	N-ethylpiperazino	2.010
18	o-toluidino	n-butylamino	1.498
19	o-toluidino	N.N-diethylethylenediamino	1.415
20	o-toluidino	4-(3-aminopropyl)morpholino	1.127
21	o-toluidino	Cyclohexylamino	2.212
22	o-toluidino	Benzylamino	1.443
23	o-toluidino	Amino	1.472
24	o-toluidino	Methylamino	1.575
25	p-toluidino	N-methylpiperazino	1.094

26	p-toluidino	N-ethylpiperazino	1.075
27	p-toluidino	N-butylamino	1.648
28	p-toluidino	N.N-diethylethylenediamino	1.918
29	p-toluidino	4-(3-aminopropyl)morpholino	0.846
30	p-toluidino	Cyclohexylamino	1.868
31	p-toluidino	Benzylamino	1.802
32	p-toluidino	Amino	1.071
33	p-toluidino	Methylamino	0.806
34	anilino	N-methylpiperazino	1.673
35	anilino	N-ethylpiperazino	1.778
36	anilino	n-butylamino	1.4197
37	anilino	N.N-diethylethylenediamino	2.464
38	anilino	4-(3-aminopropyl)morpholino	2.044
39	p-toluidino	N-methylpiperazino	1.635
40	p-toluidino	N-ethylpiperazino	1.612
41	p-toluidino	4-(3-aminopropyl)morpholino	1.538

II.2 Molecular descriptors calculation

In order to save computational time, initial geometries optimization was carried out with molecular mechanics MM2 using the MM+ force fields. The lowest energy conformations obtained by the MM method were further optimized using the DFT method [13] available in the Gaussian 09 quantum chemistry package [14], by employing Becke's three-parameter hybrid functional (B3LYP) [15], with a 6-31G basis set, Thereafter, the related electronic parameters: HOMO energy, LUMO energy and Dipole moment, were computed with the same method.

In the other hand, Advanced Chemistry Development's ACD/ChemSketchprogram [16] and Chembiooffice [17] software are used to calculate constitutional, lipophilic, topological and Steric descriptors. The totality of descriptors used in this work is listed in table 2.

Table 2 : Descriptors forming the database.

Category of descriptors	Electronic	Lipophilic	Topological	Constitutional	Steric
Name of the descriptor	HOMO energy (E_{HOMO}) LUMO energy (E_{LUMO}) Dipole moment (Dp)	LogP (Octanol-water partition coefficient),	Balaban Index (blndx) Molecular Topological Index (Mtdx) Sum Of Degrees (Sde) Sum Of Valence Degrees (Svde) Wiener Index (Windx)	Molecular weight (MW)	Ovality (O) Density (D) Surface of tension (ST) Reactivity index (Rdx) Molar refractivity (MR)

II.3 Analysis Methods

II.3.1 Multiple Linear Regression

Multiple linear regression (MLR) analysis with descendent selection and elimination of variables is used to study the relation between one dependent variable (biological activity) and several independent variables (molecular descriptors). The independent variables were individually added to or deleted from the model at each step of the regression based on three criteria: Determination coefficient R^2. Fisher ratio value (F), and the Root Mean Squared Error (RMSE). This procedure is a mathematic technique that minimizes differences between actual and predicted values. The MLR model related to antimalarial activity, is generated using the software xlstat [18]. It has served also to select the descriptors used as the input parameters for the artificial neural network (ANN).

II.3.2 Artificial Neural Networks

Artificial neural networks are artificial systems that simulate the learning process of neurons in the human brain. The ANN analysis is performed using Matlab software [19]. In this, the neurons are arranged in layers: the input layer, hidden layer, and output layer. Neurons in the same layer are not connected together. In this work, the input layer contains six neurons representing the relevant descriptors obtained with MLR techniques. The output layer represents the calculated activities values $\log IC_{50}$. The hidden layer has been determined by ρ = (number of compounds) / (number of connections). In our case, it is recommended to take into account the ρ interval $1 < \rho < 3$ [20,21] . Hence, with 32 compounds and 6 descriptors, the ρ value is 1.88 when the hidden layer is composed of two neurons [22]. As a result, the final ANN architecture is (6-2-1).

II.3.3 Cross-validation

The cross-validation method has been performed by the Leave-One-Out (LOO) procedure [23,24], which removes successively one molecule from a training set containing 32 molecules. Then it was repeated 32 times, in order to predict the properties of all molecules. The outcome from such a test is the LOO cross-validation correlation coefficient Q^2_{cv}, which is calculated according to equation 1 (eq. 1). The high average of $Q^2_{cv}$was used by several authors as an indicator of robustness and predictive ability of a model [25].

$$Q^2_{cv} = 1\text{-} \frac{\sum_{i=1}^{n}(Y_{exp}-Y_{pred})^2}{\sum_{i=1}^{n}(Y_{exp}-Y)^2} \quad \text{Eq. 1}$$

Where Y_{exp}, Y_{pred} and Y are, respectively, the measured, predicted and the averaged values for the dependent variables.

II.3.4 Y-randomization test

This test has been used to ensure the robustness of the built QSAR model. In this test, validation is performed by permuting the response parameters (Y) with respect to the (X) matrix which has been kept unchanged [26]. The Y-randomization test ensures that the correlation coefficient of the obtained model is not found by chance. For an acceptable QSAR model, the average correlation coefficient (R_r) of randomized models should be less than the correlation coefficient (R) of nonrandomized model [27] . The basis for this method is to test the validity of the original QSAR model and to ensure that the selected descriptors are appropriate.

II.3.5 External validation

Validation strategies are recognized as one the most methods, which implies a quantitative assessment of QSAR model robustness, predictive power and application domain based on new data. According to Golbraikh and Tropsha study on validation methods, internal validation is not a good parameter to estimate the capability of QSAR models [25] . However, external validation has used to estimate the performance accuracy of the QSAR on the test set that determines the true predictive power of a QSAR model. The predictive power of the built QSAR was estimated by an external Q^2_{ext}defined as follows (eq.2) [25]:

$$Q^2_{ext} = 1\text{-} \frac{\sum_{i=1}^{n}(Y_{\exp(test)}-Y_{(test)})^2}{\sum_{i=1}^{n}(Y_{test}-\bar{Y}_{tr})^2} \qquad \text{Eq. 2}$$

Where $Y_{exp(test)}$ and $Y_{(test)}$ are, respectively, the measured and the predicted values of the dependent variable test set) and $\bar{Y}_{tr}$is the average value for the dependent variable for the training set.

Q^2_{ext} is an important indicator of the reliability of the proposed model. For this purpose, it must be greater than 0.5. Golbraikh and Tropsha also proposed several other parameters for analyzing the external predictive ability of the developed QSAR model which must be respected [25].

II.3.6 Molecular Docking

A molecular modeling study was conducted in order to gain insight into the key structural requirements for the antimalarial activity. The x-ray crystal structures of both wild type (coded as 1J3I.pdb) and quadruple mutant (coded as 1J3K.pdb) of pf-DHFR-TS were obtained from the Protein Data Bank RCSB [28]. Both of these structures contain the third-generation Pf-DHFR inhibitor WR99210 bounded to the active site in the presence of NADPH [11]. The minimized protein structures was defined as receptor, consequently, the first step in the preparation of the receptor uses the Discovery Studio software for removal all waters and ligands as well as other non-protein, binding site was defined as volume occupied by the ligand in the receptor, and an input site sphere was defined over the binding site with a radius of 5 Å [29,30]. The analysis of the conformational changes of compound 11 and 38 for their ability to antimalarial activity and the analysis of the formed interactions between the ligands and the receptor were performed with program AutoDock 4.2 [31]. In this docking, the 3D grid was created by the AUTOGRID algorithm [32] to evaluate the interacting energy between protein-ligands. The grid maps were constructed using 60, 60, 60, pointing in x, y and z directions, with grid point spacing of 0.375 Å. The center grid box is of 28.09 Å, 5.76 Å, 52.59 Å, by the ligand location in the complex.

III. Results and Discussion

The dataset was randomly divided into two sets: The training set which contains 32 compounds and the test set contain with 9 compounds. The selected descriptors and the predicted activity values using the training set obtained by MLR and ANN methods are represented in table 3.

Table 3: Values of the selected descriptors and the observed/predicted logIC50 values.

N	E_{HOMO}	E_{LUMO}	Dp	MR	blndx	Svde	$LogIC_{50}$ (obs)	RLM	ANN
1	-5.241	-1.455	1.338	155.107	2438365	125	1.541	1.395	1.610
3	-5.215	-1.439	1.294	162.337	3252644	125	1.588	1.863	1.587
4	-5.237	-1.46	1.335	166.781	3665731	135	1.364	1.224	1.406
5	-5.218	-1.447	1.34	155.835	2466903	123	2.495	2.421	2.430
6	-5.277	-1.487	1.366	159.434	2820083	131	2.37	1.837	2.247
7	-5.341	-1.526	1.38	157.023	2778844	133	0.767	1.003	0.843
9	-5.351	-1.552	1.36	150.611	2559675	125	1.267	1.282	1.284
10	-5.326	-1.537	1.332	164.253	3687134	133	1.781	1.610	1.599
11	-5.346	-1.538	1.374	168.697	4119305	143	0.478	0.701	0.419
13	-5.379	-1.567	1.409	161.351	3196796	139	1.526	1.575	1.591
14	-5.376	-1.553	1.473	131.245	1528082	117	1.618	1.375	1.652
15	-5.338	-1.532	1.438	136.738	1735996	119	1.279	1.244	1.275
16	-5.328	-1.534	1.359	162.064	3097437	135	1.275	1.405	1.253
18	-5.316	-1.531	1.34	155.653	2873094	127	1.498	1.389	1.578
19	-5.324	-1.528	1.314	169.294	4094456	135	1.415	1.739	1.705
20	-5.332	-1.541	1.355	173.738	4543619	145	1.127	1.014	1.158
21	-5.309	-1.523	1.361	162.793	3131051	133	2.212	2.365	2.129
23	-5.37	-1.559	1.442	136.286	1746777	119	1.472	1.458	1.327
25	-5.323	-1.511	1.359	162.064	3129875	135	1.094	1.059	1.017
26	-5.318	-1.51	1.336	166.812	3513906	137	1.075	1.058	1.121
27	-5.333	-1.54	1.34	155.653	2904599	127	1.648	1.422	1.527
28	-5.306	-1.516	1.314	169.294	4134633	135	1.918	1.608	1.765
29	-5.314	-1.529	1.355	173.738	4584594	145	0.846	0.882	0.883
30	-5.293	-1.51	1.361	162.793	3163523	133	1.868	2.216	2.011
31	-5.355	-1.559	1.387	166.382	3584488	141	1.802	1.693	1.743
33	-5.304	-1.524	1.411	141.779	1998964	121	0.806	1.343	0.939
34	-5.588	-1.637	1.38	157.023	2657230	133	1.673	1.655	1.721
35	-5.584	-1.634	1.355	161.771	2997968	135	1.778	1.588	1.717
36	-5.61	-1.638	1.36	150.611	2445964	125	1.419	1.503	1.439
38	-5.443	-1.643	1.374	168.697	3959243	143	2.044	1.963	2.063
39	-5.57	-1.622	1.359	162.064	2996271	135	1.635	1.721	1.631
40	-5.566	-1.618	1.336	166.812	3367587	137	1.612	1.681	1.622

III.1 Multiple Linear Regression (MLR)

The training set was used to build a QSAR model using MLR method. The obtained QSAR model is represented by the following equation 3:

$$LogIC_{50} = -54.5 + 4.4 * E_{HOMO} - 14.9 * E_{LUMO} + 34 * D + 0.38 * MR - 8.210^{-7} * Blndx - 0.36 * Svde \quad Eq3$$

$$N = 32 \qquad R = 0.88 \qquad R^2 = 0.78 \qquad F = 14.69 \qquad RMSE = 0.23$$

Where N is the number of compounds, R is the correlation coefficient, R^2 is the determination coefficient, RMSE is the root mean square error, F is the Fisher F-statistic. The relevant descriptors involved in the MLR model of the training set are HOMO energy (E_{HOMO}), LUMO energy (E_{LUMO}), Density (D), Molar refractivity (MR), Sum of Valence Degrees (Svde) and Balaban Index (blndx). The corresponding normalized descriptors coefficients and the correlation of the observed activity obtained by the MLR method are presented, respectively, in **figure 2** and **figure 3**.

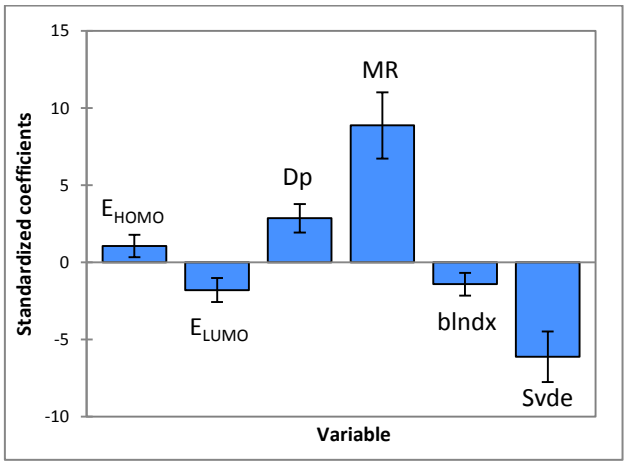


Figure 2: Modeling characterization by the normalized coefficients.

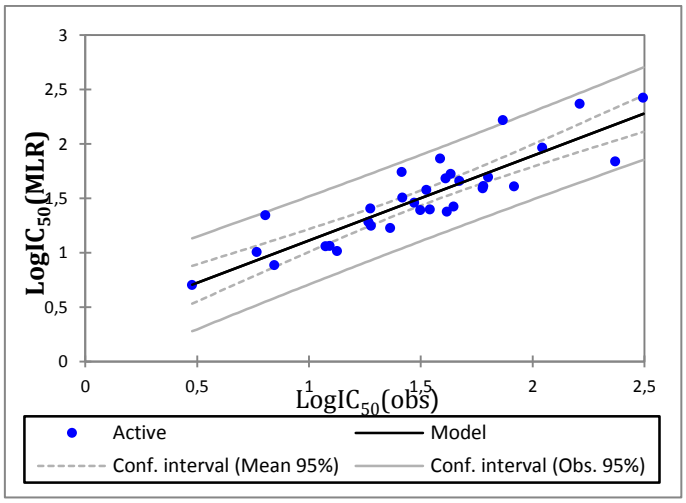


Figure 3: Correlation between observed and predicted activities calculated using MLR model

The correlation between the experimental and calculated activities based on MLR model for a series of 32 compounds was quite significant, as indicated by the statistical values (R=0.88), and the RMSE value (0.23). The value obtained for RMSE showed that the model has a good prediction precision.

III.2 Artificial Neural networks (ANN)

In order to increase the probability of good characterization of studied compounds, artificial neural networks (ANN) has used as a non-linear method to generate a predictive non-linear model between the set of molecular descriptors selected by MLR method and the observed activity. The correlation between the observed and the predicted activities using the ANN method is illustrated in **figure 4**.

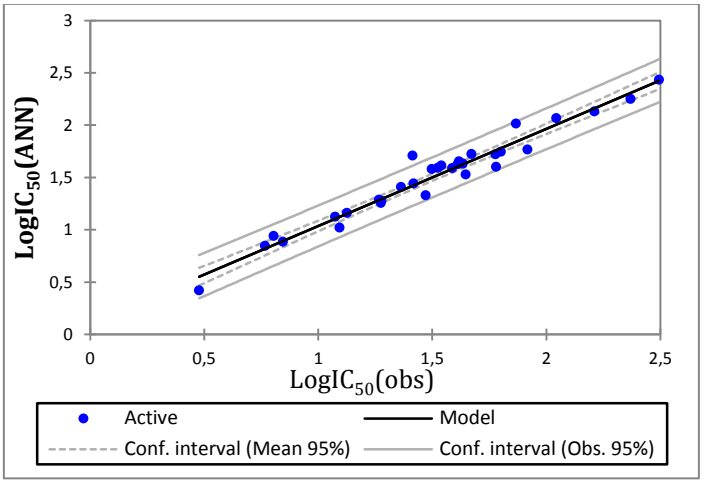


Figure 4: The correlation between observed and predicted activities using the ANN method.

Figure 4 shows a good correlation between observed and predicted (ANN) activities. Thus, this model has a significant statistical quality and an excellent prediction ability (R=0.97, R^2=0.95 and RMSE=0.09).

III.3 Cross-validation (CV)

In this study, we have used to validate our model through "leave-one-out" (LOO) cross-validation. The outcomes achieved (Q_{cv}=0.95, Q^2_{cv}=0.91 and RMSE= 0.12) reveals the robustness and the predictive ability of the built QSAR model. Further tests have used to ensure the applicability of this obtained model.

III.4 Y-randomization

This technique was widely used to ensure the robustness of a QSAR model. The results of the QSAR model obtained in the Y-randomization test (table 4) showed relatively a less average correlation coefficient (R_r^2 = 0.69) than obtained by the nonrandomized model (R^2 = 0.78). Furthermore, this result implies that the obtained QSAR model is robust and has a good predictive ability.

Table 4: Comparison between observed and predicted activities obtained using Y-randomization method.

N	5	9	13	17	22	25	29	35	41	15	18	24	26	30	32	37
logIC_{50}	2.495	1.267	1.526	2.010	1.443	1.094	0.846	1.778	1.538	1.279	1.498	1.575	1.075	1.868	1.071	2.464
pred logIC_{50}	2.500	1.078	1.443	1.275	1.498	1.279	1.007	1.616	1.313	1.337	1.211	1.325	1.247	2.082	1.335	2.449
N	34	39	11	14	16	20	28	31	36	1	4	6	19	2	12	8
logIC_{50}	1.673	1.635	0.478	1.618	1.275	1.127	1.918	1.802	1.419	1.541	1.364	2.370	1.415	1.473	1.811	0.772
pred logIC_{50}	1.684	1.804	0.922	1.465	1.301	1.061	1.585	1.491	1.621	1.680	1.572	1.944	1.639	1.675	1.992	1.087

III.5 External Validation

The external validation has used in order to estimate the true predictive power of the proposed QSAR model. For this reason, nine compounds have been randomly removed from the original set. Thus, the model has been built with 32 compounds through MLR and ANN methods. Then, we have tested the applicability of the built model on the nine compounds (table 5). In addition, Golbraikh and Tropsha have proposed a set of parameters for determining the external predictability of QSAR model (table 6) [25].

Table 5: Comparison between experimental and predicted LogIC50 values of an external test set for the MLR and ANN models based on descriptors of equations 3.

N	**LogIC_{50}**	**Pred logIC_{50} (MLR)**	**Residuals**	**Pred logIC_{50} (ANN)**	**Residuals**
2	1.473	1.542	0.003	1.526	-0.02
8	0.772	0.963	-0.017	1.029	-0.292
12	1.811	2.038	0.229	1.986	-0.078
17	2.010	1.352	-0.615	1.342	0.675
22	1.443	1.519	0.004	1.518	-0.043
24	1.575	1.499	-0.121	1.516	0.09
32	1.071	1.170	-0.049	1.175	-0.119
37	2.464	2.542	0.213	2.533	0.101
41	1.538	1.944	0.353	1.939	-0.312

Table 6: Golbraikh and Tropsha's criteria.

Parameter	Formula	Threshold	Modelscore
Q^2_{ext}	$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{n}(Y_{\exp(test)} - Y_{(test)})^2}{\sum_{i=1}^{n}(Y_{test} - \bar{Y}_{tr})^2}$	Q^2_{ext}>0.5	0.65
r^2	The coefficient of determination for the plot of predicted versus observed for the test set	r^2>0.6	0.67
r_0^2	r2 at zero intercept		0.65
$r_0'^2$	r2 for the plot of observed versus predicted activity for the test set at zero intercept		0.61
$\lvert r_0^2 - r_0'^2 \rvert$		$\lvert r_0^2 - r_0'^2 \rvert < 0.3$	0.04
k	The slope of the plot of predicted versus observed activity for the test set at zero intercept	0.85<k<1.15	0.96
$\frac{r^2 - r_0^2}{r^2}$		$\frac{r^2 - r_0^2}{r^2} < 0.1$	0.03
k'	The slope of the plot of observed versus predicted activity at zero intercept	0.85<k'<1.15	1.01
$\frac{r^2 - r_0'^2}{r^2}$		$\frac{r^2 - r_0'^2}{r^2} < 0.1$	0.09

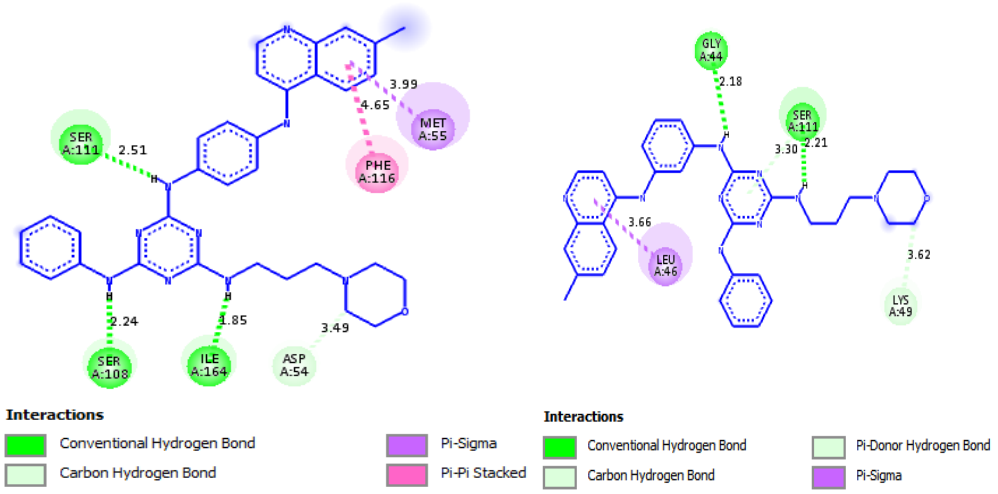
The results show that Golbraikh and Tropsha criteria's are successfully validated which indicated the validity and quality of the QSAR model developed in this work who meets all of the principles for QSAR validation. The most important result of this investigation is that in vitro antimalarial activity could be predicted using QSAR methods.

III.6 Docking studies

In an attempt to better understand the high antimalarial activity potency manifested with certain compounds and the lack activity observed with others, we have decided to perform molecular docking with the binding sites of both wild type and quadruple mutant of Pf-DHFR-TS for the highest active compound 11 and the lowest active compound 38, having the same radicals R1 and R2 but belong to different structural categories as it is detailed in paragraph II.1. The reported study by Yuvaniyama et al [10,33] has found the binding modes and has localized the active sites on both wild and mutant type of protein (*Pf*-DHFR). The study performed with a potent inhibitor 1,3,5-triazine derivative which is a preclinical molecule called WR99210. It is found that the important sites in the case of the wild type are located in ILE14, ALA16, MET55, ASP54, SER108, ILE164 and TYR170. It was also found that important sites are located in ALA16, CYS50, ASN51, CYS59, ASN108, LEU164 and TYR170 in the case of the wild type. The interactions obtained for the two compounds are illustrated in **figure 5**.

Wild type of *pf-DHFR*

Compound 11 Compound 38



Quadruple mutant *pf-DHFR-TS*

Compound 11 Compound 38

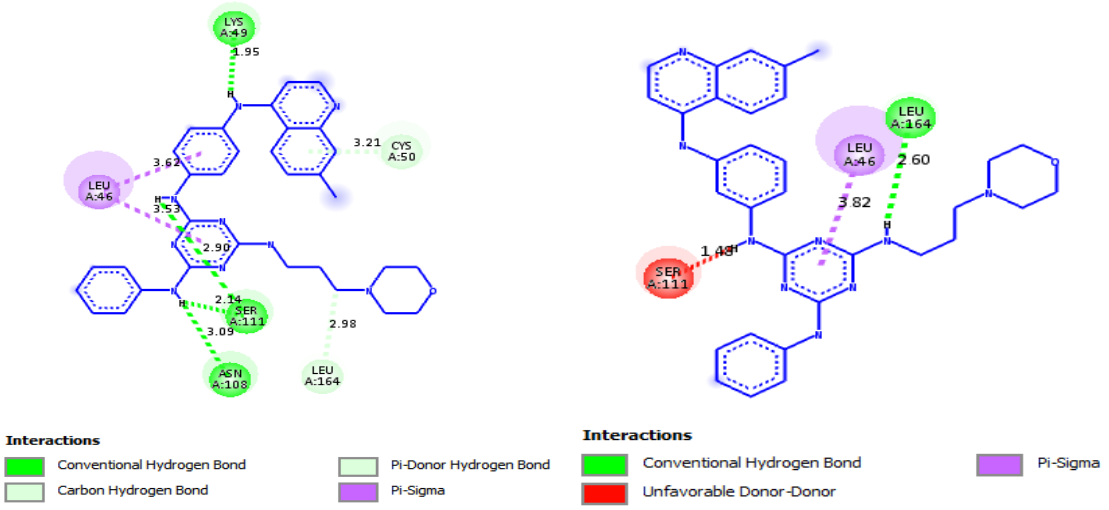


Figure 5: Binding interaction from docking simulation of 11 and 38 compounds into the active site of wild type and a quadruple mutant of *pf-DHFR*.

The molecular docking of compound 11 with wild type shows three hydrogen bond between the three nitrogens linked to the triazine group, and SER111, SER108 and ILE164 at respectively 2.51 Å, 2.24 Å and 1.84 Å, in the other hand a P-P and a p-sigma interactions are observed between a phenyl group of the quinoline moiety and respectively PHE116 and MET55 at a distance of 4.65 Å and 3.99 Å. However compound 38 forms only two hydrogen bonds with two less important binding sites, SER111 and GLY44, as they are not cited as active sites for antimalarial activity [28]. In the case of quadruple mutant, compound 11 forms three hydrogen bond with ASN108, SER111 and lys49, and two interaction P-sigma throughout two phenyl groups, one belongs to the aniline attached to the anilinoquinoline

moiety, and a second with the triazine moiety with LEU46, however compound 38 forms only one hydrogen bond and one P-sigma interaction with respectively LEU146 and LEU46..

In the analysis of these results we have observed that the residues with that the compound 11 has undertaken its interactions present the most important binding sites for antimalarial activity referring with an antecedent study on the antimalarial protein characterization[33], which might be the plausible reason for observing differences in the activity of wild and mutant strains. So, as compounds 11 and 38 are two position isomers (meta and para) we can note how the position of radicals could make changes in activities potency through the changes of interactions with receptor. Further We noticed that the 3D visualization of compound 11–protein complex (wild type) (**figure 6**) showed that the plans of phenyl rings that form the P-P interaction (at a distance of 4.65 Å) are almost parallel, which is considered in a previous study [34] as a strong interaction.

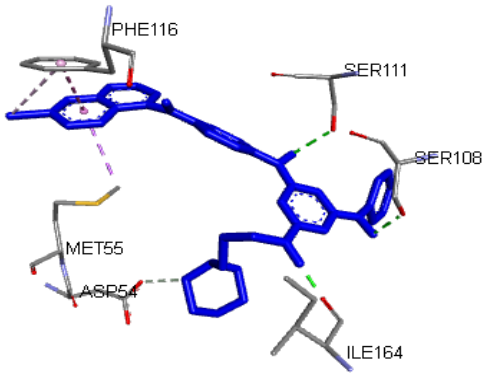


Figure 6: 3D visualization of compound 11-protein complex (wild type)

IV. Conclusion

The present work shows how antimalarial activities of 41 hybrids of 4-Anilinoquinoline triazines may be treated statistically to uncover the molecular characteristics which are essential for high activity. The generated models were analyzed and validated for their statistical significance and external prediction power. The QSAR analysis revealed a set of important descriptors which influence the activity: HOMO energy, LUMO energy, Density, Sum of Valence Degrees and Balaban Index. The molecular docking studies performed with compounds 11 (para-isomer) and 38 (meta-isomer) shows that the most active compound 11 forms important interactions with the active sites as it was found with known inhibitors (WR99210) of the wild-type and quadruple mutant forms of *Pf-DHFR*. Therefore, the docked

poses provide details of the predicted binding modes and the key molecular interactions, which might provide opportunities for medicinal chemists to develop new antimalarial drugs.

References

[1] N. Kumar, S.I. Khan, H. Atheaya, R. Mamgain, D.S. Rawat, Synthesis and in vitro antimalarial activity of tetraoxane-amine/amide conjugates, Eur. J. Med. Chem. 46 (2011) 2816–2827.

[2] T.A. Ghebreyesus, Avant-propos, (2018). https://www.who.int/malaria/media/world-malaria-report-2018/fr/.

[3] J.B. Vaishnani, Cutaneous findings in five cases of malaria., Indian J. Dermatol. Venereol. Leprol. 77 (2011) 110.

[4] M.C. Murray, M.E. Perkins, Chapter 15. Chemotherapy of Malaria, Annu. Rep. Med. Chem. 31 (1996) 141–150.

[5] P. Newton, BM, DPhil, MRCP, N. White, DSc, MD, FRCP, MALARIA: New Developments in Treatment and Prevention, Annu. Rev. Med. 50 (1999) 179–192.

[6] M.D. Hastings, C.H. Sibley, Pyrimethamine and WR99210 exert opposing selection on dihydrofolate reductase from Plasmodium vivax., Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 13137–41.

[7] I.B. Müller, J.E. Hyde, Antimalarial drugs: modes of action and mechanisms of parasite resistance, Future Microbiol. 5 (2010) 1857–1873.

[8] F.W. Muregi, A. Ishih, Next-generation antimalarial drugs: hybrid molecules as a new strategy in drug design, Drug Dev. Res. 71 (2009) n/a-n/a.

[9] A. Kumar, K. Srivastava, S. Raja Kumar, M.I. Siddiqi, S.K. Puri, J.K. Sexana, P.M.S. Chauhan, 4-Anilinoquinoline triazines: A novel class of hybrid antimalarial agents, Eur. J. Med. Chem. 46 (2011) 676–690.

[10] H. Hadni, M. Mazigh, E. Charif, A. Bouayad, M. Elhallaoui, Molecular Modeling of Antimalarial Agents by 3D-QSAR Study and Molecular Docking of Two Hybrids 4-Aminoquinoline-1,3,5-triazine and 4-Aminoquinoline-oxalamide Derivatives with the Receptor Protein in Its Both Wild and Mutant Types, Biochem. Res. Int. 2018 (2018) 1–15.

[11] D. Hoekman, Exploring QSAR Fundamentals and Applications in Chemistry and Biology, Volume 1. Hydrophobic, Electronic and Steric Constants, Volume 2 J. Am. Chem. Soc. 1995, 117, 9782, (1996).

[12] W. Sirawaraporn, T. Sathitkul, R. Sirawaraporn, Y. Yuthavong, D. V Santi, Antifolate-resistant mutants of Plasmodium falciparum dihydrofolate reductase., Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 1124–9. http://www.ncbi.nlm.nih.gov/pubmed/9037017.

[13] R.G. Parr, W. Yang, Density-Functional Theory of Atoms and Molecules, 1989.

[14] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J. Montgomery J. A., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. V Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision C.02, 2004.

[15] A.D. Becke, Density-functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648–5652.

[16] ACD/ChemSketch for Academic and Personal Use :: ACD/Labs.com, (n.d.). http://www.acdlabs.com/resources/freeware/chemsketch/.

[17] G.W.A. Milne, Software Review of ChemBioDraw 12.0 ChemBioDraw 12.0 CambridgeSoft 100 CambridgePark Drive, Cambridge, MA 02140, United States . http://www.cambridgesoft.com/software/ChemBioDraw/, J. Chem. Inf. Model. 50 (2010) 2053–2053.

[18] XLSTAT Logiciel de statistique pour MS Excel | Logiciel statistique pour Excel, (n.d.). https://www.xlstat.com/fr/.

[19] The MathWorks Inc., MATLAB and statistics toolbox, Release 2010b [computer software], 2010.

[20] S.S. So, W.G. Richards, Application of neural networks: quantitative structure-activity relationships of the derivatives of 2,4-diamino-5-(substituted-benzyl)pyrimidines as DHFR inhibitors, J. Med. Chem. 35 (1992) 3201–3207.

[21] T.A. Andrea, H. Kalayeh, Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors, J. Med. Chem. 34 (1991) 2824–2836.

[22] V. Kůrková, Kolmogorov's theorem and multilayer neural networks, Neural Networks.

(1992).

[23] B. Efron, Estimating the error rate of a prediction rule: Improvement on cross-validation, J. Am. Stat. Assoc. 78 (1983) 316–331.

[24] M. Stone, Cross-Validatory Choice and Assessment of Statistical Predictions, J. R. Stat. Soc. (1974).

[25] A. Golbraikh, A. Tropsha, Beware of q2!, J. Mol. Graph. Model. 20 (2002) 269–276.

[26] * Christoph Rücker, § and Gerta Rücker, M. Meringer‖, y-Randomization and Its Variants in QSPR/QSAR, (2007).

[27] K. Roy, I. Mitra, On Various Metrics Used for Validation of Predictive QSAR Models with Applications in Virtual Screening and Focused Library Design, Comb. Chem. High Throughput Screen. 14 (2011) 450–474.

[28] RCSB PDB: Homepage, (n.d.). http://www.rcsb.org/.

[29] D.S. BIOvIA, Discovery studio modeling environment., San Diego, Dassault Syst. Release, 4. (2015).

[30] H.R. Bhat, U.P. Singh, A. Thakur, S. Kumar Ghosh, K. Gogoi, A. Prakash, R.K. Singh, Synthesis, antimalarial activity and molecular docking of hybrid 4-aminoquinoline-1,3,5-triazine derivatives, Exp. Parasitol. 157 (2015) 59–67.

[31] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785–2791.

[32] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, J. Comput. Chem. 19 (1998) 1639–1662.

[33] J. Yuvaniyama, P. Chitnumsub, S. Kamchonwongpaisan, J. Vanichtanankul, W. Sirawaraporn, P. Taylor, M.D. Walkinshaw, Y. Yuthavong, Insights into antifolate resistance from malarial DHFR-TS structures, Nat. Struct. Biol. 10 (2003) 357–365.

[34] M. Elhallaoui, M. Laguerre, A. Carpy, F.C. Ouazzani, Molecular modeling of noncompetitive antagonists of the NMDA receptor: proposal of a pharmacophore and a description of the interaction mode, J. Mol. Model. 8 (2002) 65–72.

Chapitre V : Molecular docking and QSAR studies for modeling the antimalarial activity of 4-anilinoquinoline-triazines hybrids Derivatives with the wild-type and mutant Receptor pf-DHFR

Hanine HADNI, Menana ELHALLAOUI

Engineering Materials, Modeling and Environmental Laboratory, Faculty of Sciences of Dhar Elmehraz, Sidi Mohammed Ben Abdellah University, B.P. 1796, Atlas, Fes, Morocco

Abstract:

Plasmodium falciparum dihydrofolate reductase (pf-DHFR) is one of many targets for malaria treatment. Double and quadruple mutations at residues 51, 59, 108 and 164 of pf-DHFR were related to antifolate resistance. Several efforts are under way to overcome this drug resistance and to produce potential inhibitors. In continuity of our previous studies in this way, QSAR and docking studies were carried out for molecular hybrids based on 4-aminoquinoline. The generated model showed good correlation coefficients (R = 0.9 and R^2 = 0.81) and test set prediction coefficients (R = 0.88 and R^2 = 0.78). These results showed the good predictive potency of the established QSAR model. the most active 4 and the least active compounds 5 are docked at the active sites of the pf-DHFR protein. So The docking results revealed that these molecules interact specifically with SER108 and ILE164 binding pocket known as active sites.

Keywords: 4-Aminoquinoline-triazine, antimalarial activity, pf-DHFR, QSAR, molecular docking.

I. Introduction

Malaria is an infectious disease caused by parasites of the genus *Plasmodium* [1]. Five species of malaria parasites are known, namely, *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale* and *P. knowlesi* [2]. Approximately 90 % of deaths (generally of children in Africa) related to malaria infections are caused by *Plasmodium falciparum* [3,4]. In 2016, 216 million of human beings are counted by the World Health Organization (WHO) over 91 countries of the world, which are affected by malaria, with an increase of 5 million cases compared to previous years [5]. Despite of the great efforts devoted to discover an effective antimalarial drugs, these efforts suffer from many obstacles, including drug resistance issues [6,7]. To overcome this problem, the concept of hybrid molecules has been introduced as one of the most used

solutions, in which two or more pharmacophores are linked together and act by inhibiting simultaneously two conventional targets [8]. In this regard, the 1,3,5-triazine derivatives such as cycloguanil, chlorcycloguanil and WR99210 are already approved as effective *dihydrofolate reductase* (DHFR) inhibitors, which selectively inhibit biochemical processes that are vital for parasite growth [9]. In addition, quinoline nucleus has attracted much interest of medicinal chemists, as an imperative pharmacophore accountable for imparting antimalarial action[10,11]. In the purpose to pursue our previous works [12], in this paper we have performed the molecular modeling of 4-anilinoquinoline-triazines (**figure1**) as a potential antimalarial compounds by using QSAR and docking studies [13].

Quantitative Structure Activity Relationships (QSAR) is very important tool in drug discovery [14]. The multidimensional molecular descriptors (1D, 2D, 3D) have been calculated to identify regions in space which are correlated to the biological activities (table 1) [15]. QSAR models could be generated by using statistical methods. In this study, we have used a training set of 37 4-anilinoquinoline-triazines derivatives to build the QSAR model [16]. For this purpose, we used multiple linear regression analysis (MLR)) and artificial neural networks (ANN). The predictive ability of the established model has been tested by several validation techniques such as: internal and external validations as well as Y-randomization methods. Further, in order to gain structural insight for improved antimalarial activity, we investigated the interaction of two hybrids 4 and 5 with the Pf-DHFR protein binding sites in its two forms; the wild type and the quadruple mutant [17].

II. Materials and methods

II.1 Studied molecules

To perform the molecular modeling we have used the experimental antimalarial activities data of 37 hybrids molecules described previously [13]. Thus, the observed activities (IC50 (nM)) are converted into logarithm scale logIC50 and they are presented in table 1.

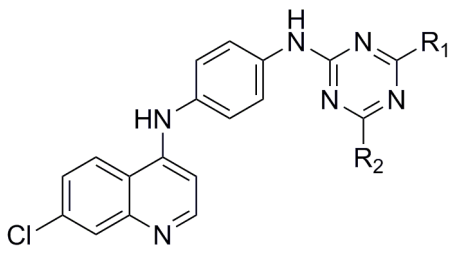
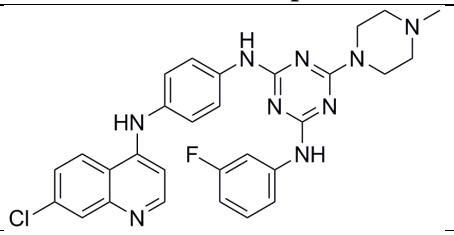
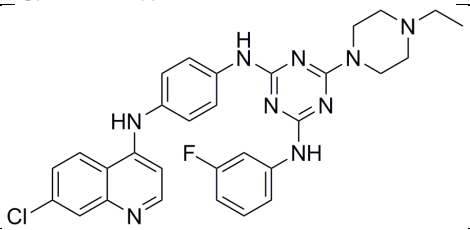
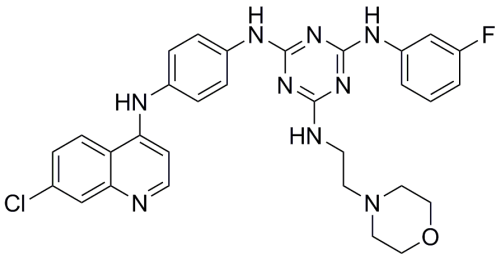
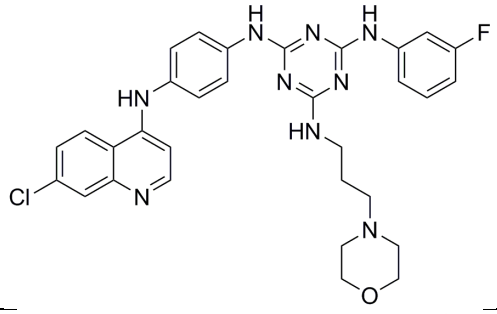
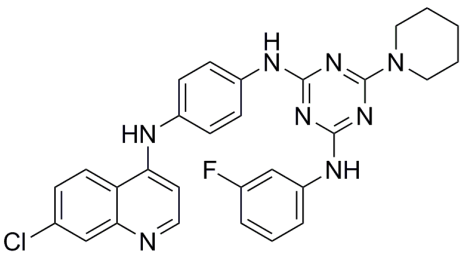
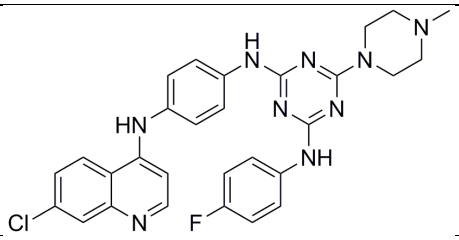
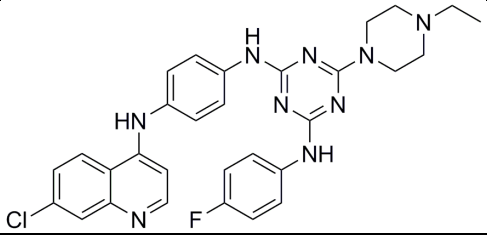
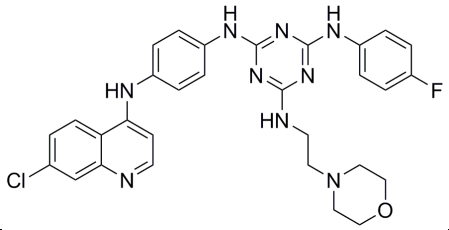
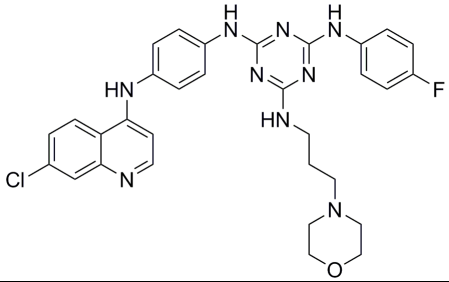
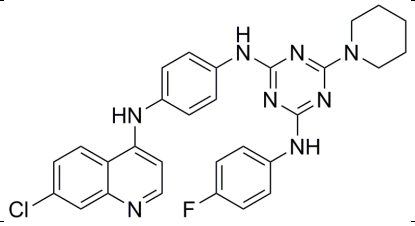
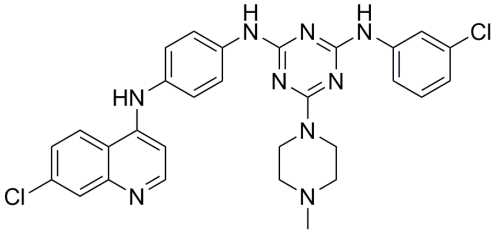
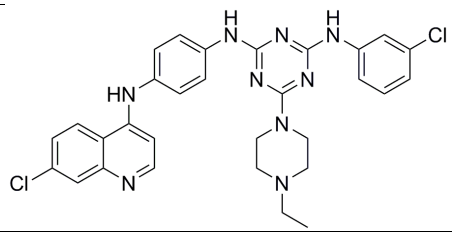
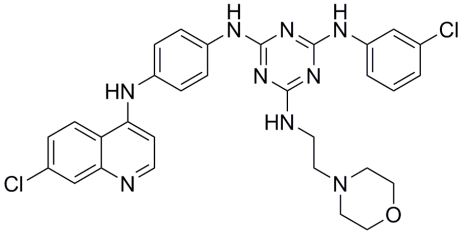
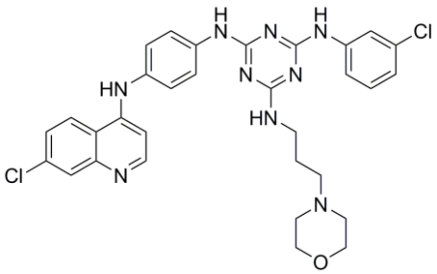
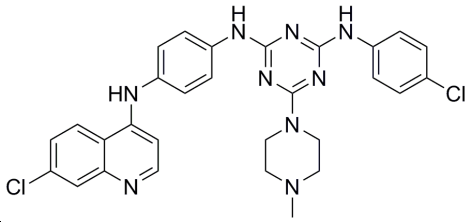
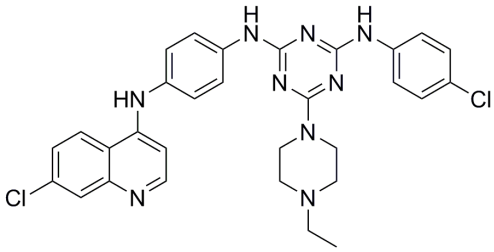
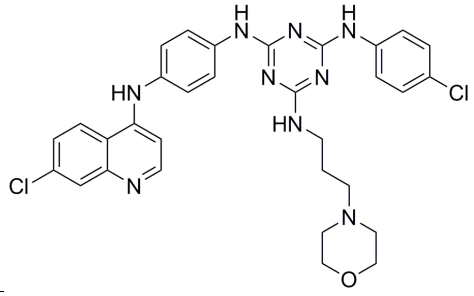
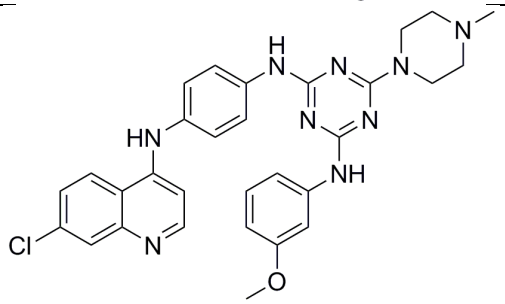
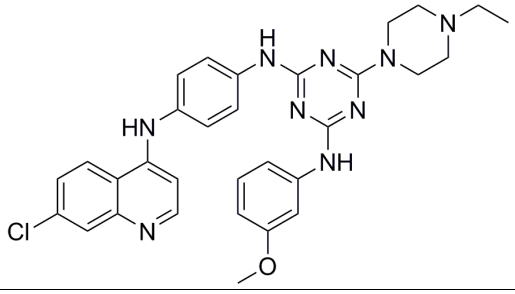
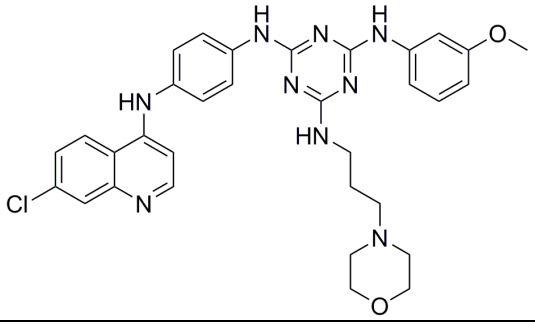


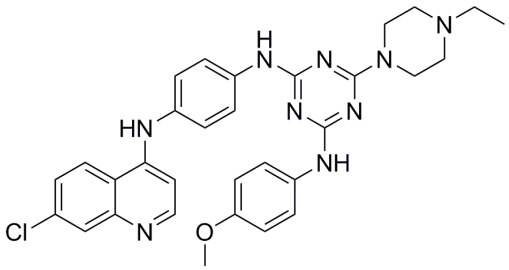
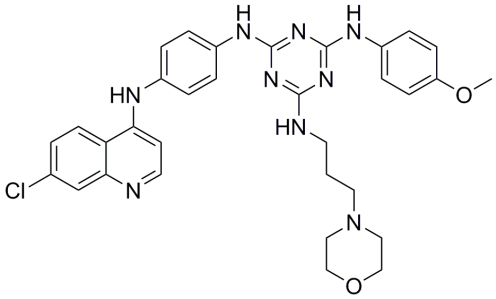
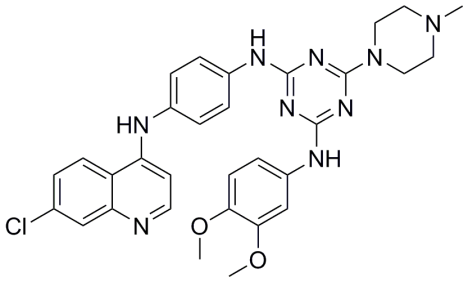
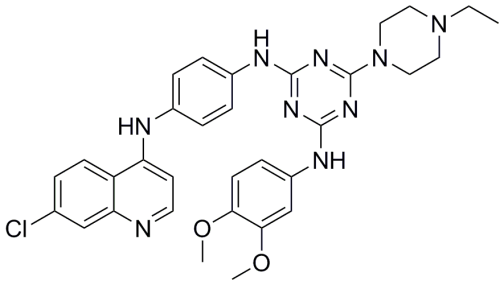
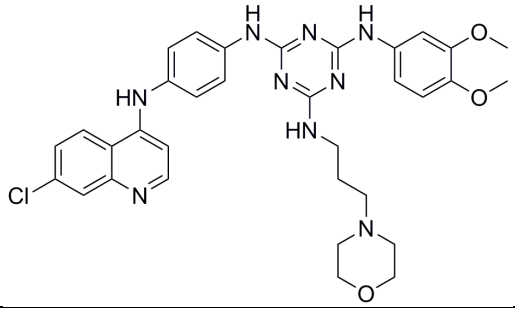
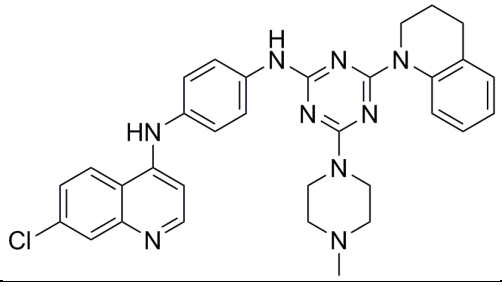
Figure1: The general structure of the 4-anilinoquinoline-triazine.

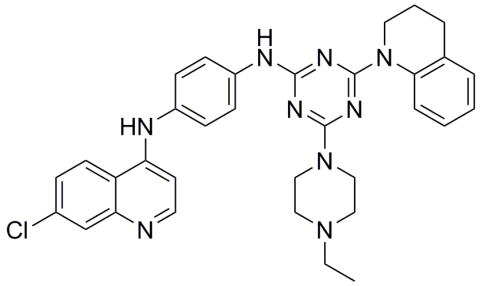
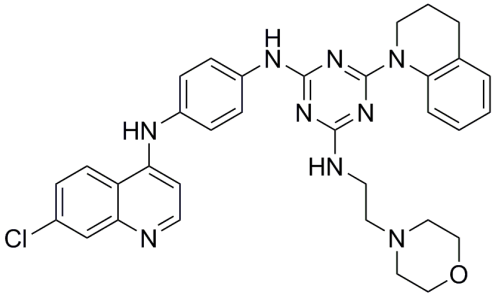
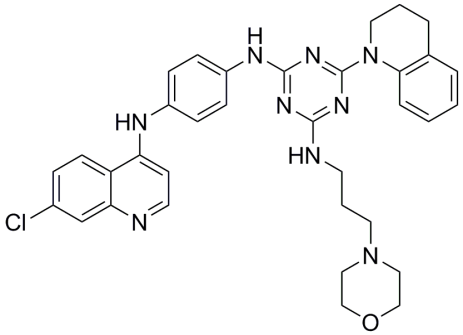
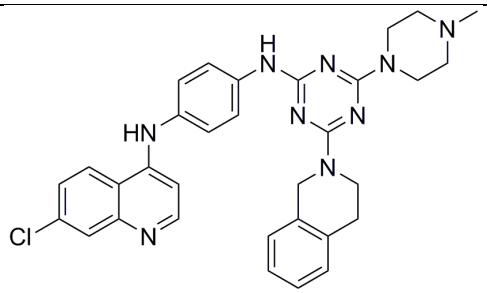
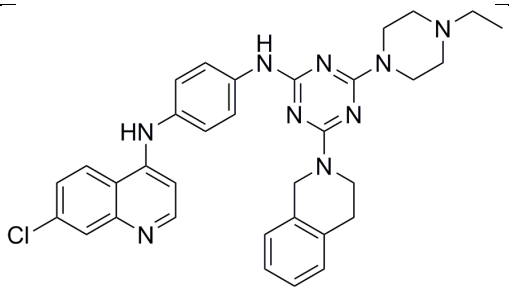
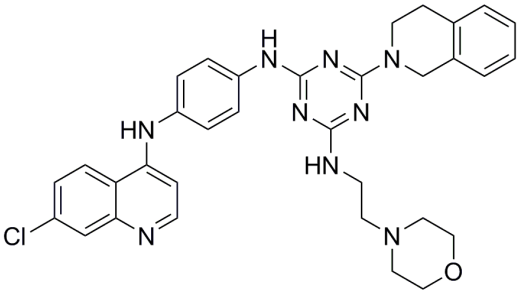
Table 1: Studied compounds and their observed activities $logIC_{50}$.

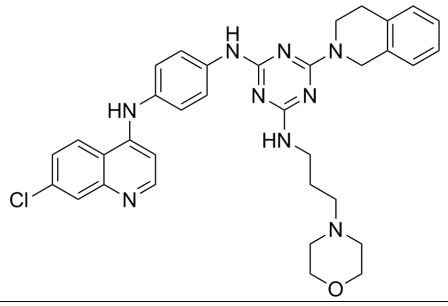
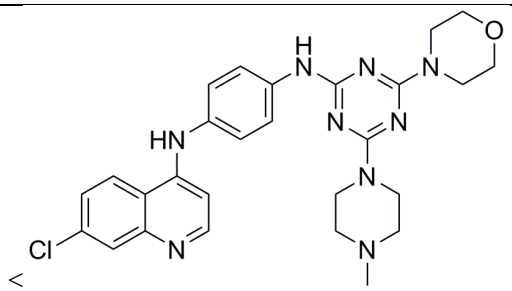
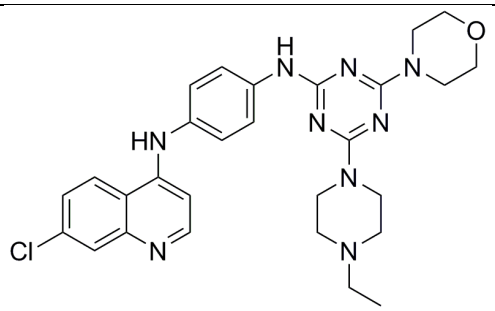
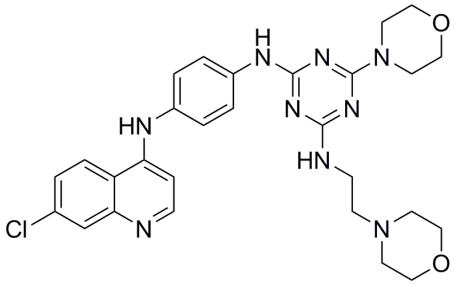
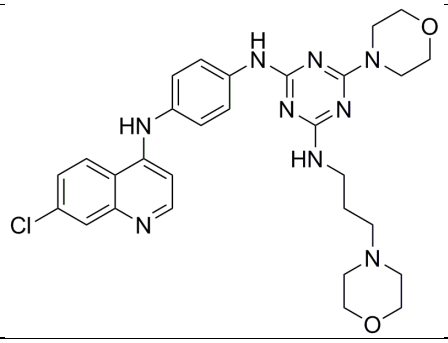
Compounds	Structure of compound	$LogIC_{50}$(obs)
1		1.358
2		0.393
3		1.602
4		0.134
5		2.110
6		0.953
7		1.046

8		0.992
9		1.199
10		1.621
11		0.382
12		0.324
13		1.486
14		0.582

15		0.892
16		0.617
17		0.666
18		0.629
19		0.949
20		0.474

21		0.629
22		0.581
23		0.702
24		0.412
25		0.661
26		1.102

27		1.319
28		1.239
29		1.167
30		1.449
31		1.084
32		1.655

33		1.308
34		0.508
35		0.334
36		0.769
37		0.441

II.2 Molecular descriptors calculation

In order to build a reliable QSAR model, a total of 14 descriptors including electronic, lipophilic, geometrical, physicochemical and steric. These later are calculated using DFT/B3LYP(6-31G) method [20,21] available in Gaussian 03 quantum chemistry software[22]. However the MM2 method using ACD/ChemSketch [18] and ChemBioOffice

softwares [19] is used to clalculate the remained descriptors. The descriptors used in this work are presented in **table 2**.

Table 2: List of the calculated descriptors.

Type of descriptors	Electronic	lipophilic	Geometrical	physicochemical	Steric
Name of the descriptors	HOMO energy (E_{HOMO}) LUMO energy (E_{LUMO}) Dipole moment (Dp) total energy (E)	LogP lipophilic(Octanol -water partition coefficient) VDW energy (E_{VDW})	Torsion(T) stretch-bend(SB)	Critical pressure (CP) Critical volume (CV)	Density (D) $D = \frac{MW}{MV}$ Refractive Index (R) Surface Tension (ST) parchor(P)

II.3 Statistical methods

The complete set of 37 compounds was randomly divided into two subsets, a training set of 29 compounds to build the model and a test set of 8 compounds to evaluate the reliability of the established model. Various statistical methods were used to build the QSAR model, *viz*.: Multiple Linear Regression (MLR), and Artificial Neural Networks (ANN). Indeed, the MLR with descendent selection of variables was used to study the relation between one dependent variable (antimalarial activity) and independent variables (calculated molecular descriptors). Furthermore, the MLR was used to select the descriptors that will serve as the input parameters for ANN. Hence, ANN could be considered as suitable tools that have powerful mechanism to capture patterns in data, which has been widely used to model nonlinear system [23]. In addition, the high values of the correlation coefficient indicate how the equations fit the data. However, a validation strategy is required to inquire about the applicability of the QSAR models on a new data. So, in this study we have used Cross-Validation method with "leave-one-out" procedure [24,25] and External Validation using a test set of 8 compounds. The external validation should be in perfect agreement with the criteria of Golbraikh and Tropsha [26]. Finally, Y-randomization test has been used to exclude the possibility of random correlation between molecular descriptors and the corresponding activities in the obtained model [27].

II.4 Molecular Docking modeling

A molecular modeling study was conducted in order to gain insight into the key structural requirements of a geometrical model and to analyze the interactions of the hybrid systems with the active sites of the protein *pf-DHFR* [12,17] of both wild type (coded as 1J3I.pdb) and

quadruple mutant type (coded as 1J3K.pdb), which are obtained from the Protein Data Bank RCSB [28]. In this study, two compounds were chosen to be docked with pf-DHFR protein; the highest and the lowest active compounds of the studied series, **4** and **5**. At first, we removed all waters and ligands as well as other non-protein by using the Discovery Studio software [29]. The AutoDock 4.2, has been used to analyze the interactions between the ligand and the protein [30].The 3D grid was created by the AUTOGRID algorithm [31]. The grid maps were constructed using 60, 60, 60, pointing in x, y and z directions, with grid point spacing of 0.375 Å. The center grid box is of 28.09 Å, 5.76 Å, 52.59 Å, by the ligand location in the complex. However, the center grid box of the quadruple mutant type protein is about (29.987 Å, 5.56 Å and 57.424 Å). Discovery Studio software was used for the 2D and 3D visualizations of the established interactions[29].

III. Results and Discussion

The values of the selected descriptors and the predicted values of antimalarial activity of the training set which obtained by using MLR, ANN and CV methods are presented in Table 3.

Table 7 : Values of selected descriptors and observed/predicted activities ($logIC_{50}$).

N	E	S-B	T	E_{VDW}	LogP	ST	$LogIC_{50}$ (obs)	MLR	ANN	Cv
2	-2213.007	0.520	-15.547	42.004	7.347	72.1	0.393	0.741	0.724	0.739
3	-2288.181	0.532	-12.073	43.096	6.065	77.9	1.602	1.195	1.311	1.299
4	-2327.486	0.556	-12.039	43.729	6.170	75.7	0.134	0.858	0.667	0.772
5	-2118.390	0.371	-15.879	39.069	7.986	76.6	2.110	2.111	2.099	2.100
7	-2213.006	0.522	-15.540	42.003	7.347	72.1	1.046	0.738	0.725	0.737
8	-2288.181	0.534	-12.048	43.091	6.065	77.9	0.992	1.195	1.310	1.298
9	-2327.485	0.558	-12.034	43.726	6.170	75.7	1.199	0.856	0.665	0.769
11	-2534.064	0.483	-15.753	41.898	7.409	76.7	0.382	0.779	0.649	0.628
12	-2573.369	0.533	-15.553	42.645	7.747	73.8	0.324	0.534	0.473	0.472
13	-2648.544	0.545	-12.055	43.739	6.465	79.6	1.486	0.992	1.416	1.288
14	-2687.848	0.570	-12.044	44.369	6.570	77.4	0.582	0.650	0.629	0.740
15	-2534.064	0.485	-15.692	41.838	7.409	76.7	0.892	0.777	0.641	0.625
16	-2573.369	0.533	-15.557	42.558	7.747	73.8	0.617	0.520	0.461	0.464
18	-2188.972	0.518	-15.798	44.081	6.724	73.2	0.629	0.711	0.675	0.678
19	-2228.277	0.571	-15.560	44.843	7.062	70.6	0.949	0.525	0.762	0.761
20	-2342.756	0.624	-12.094	46.444	5.886	74.1	0.474	0.569	0.495	0.515
21	-2228.275	0.578	-15.558	44.824	7.062	70.6	0.629	0.511	0.764	0.756
22	-2342.754	0.615	-12.040	46.569	5.886	74.1	0.581	0.613	0.527	0.550
23	-2303.447	0.477	-13.677	47.151	6.598	70.6	0.702	0.763	0.718	0.768
25	-2457.232	0.593	-9.705	49.569	5.759	71.6	0.661	0.677	0.659	0.611
26	-2191.186	0.989	-15.240	49.085	7.482	71.7	1.102	1.220	1.110	1.177
27	-2230.491	1.046	-14.930	49.892	7.820	69.2	1.319	1.064	1.320	1.174
28	-2305.667	1.058	-11.375	51.018	6.538	74.4	1.239	1.432	1.243	1.280
30	-2191.189	0.605	-15.188	46.528	7.381	71.7	1.449	1.384	1.403	1.300
31	-2230.494	0.660	-14.869	47.331	7.719	69.2	1.084	1.231	1.108	1.204
32	-2305.670	0.669	-11.361	48.401	6.437	74.4	1.655	1.587	1.661	1.600
34	-2074.665	0.796	-6.763	43.187	5.271	72.0	0.508	0.525	0.508	0.392
35	-2113.970	0.851	-6.497	43.975	5.609	69.3	0.334	0.327	0.356	0.390
37	-2228.449	0.885	-3.101	45.637	4.433	72.9	0.441	0.417	0.436	0.427

III.1 Multiple Linear Regression

The MLR method is based on three criteria: Coefficient of determination (R^2), the root mean square error (RMSE) and the Fisher ratio value (F) [12]. The QSAR model of the training set is represented by the following equation:

$$LogIC_{50} = -15.8 + 0.0026 * E - 1.67 * S\text{-}B + 0.18 * T + 0.15 * E_{VDW} + 0.91\ LogP + 0.18 * ST$$

$$N = 29 \qquad R = 0.84 \qquad R^2 = 0.70 \qquad F = 8.84 \qquad RMSE = 0.29$$

The MLR results which contain the corresponding normalized descriptors coefficients and the correlation between the observed and predicted activities are presented in **figures 2** and **3** respectively.

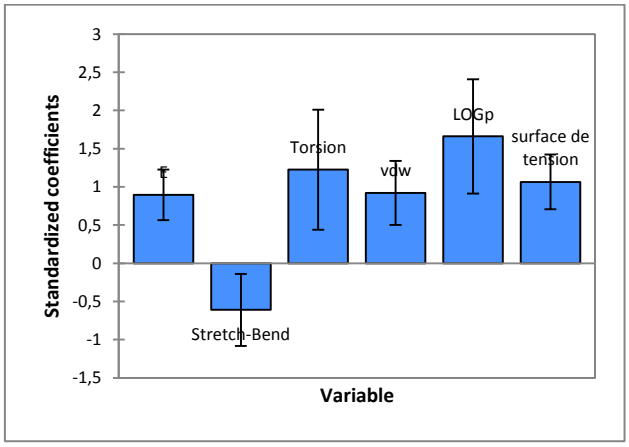


Figure 2: Modeling characterization by the normalized coefficients.

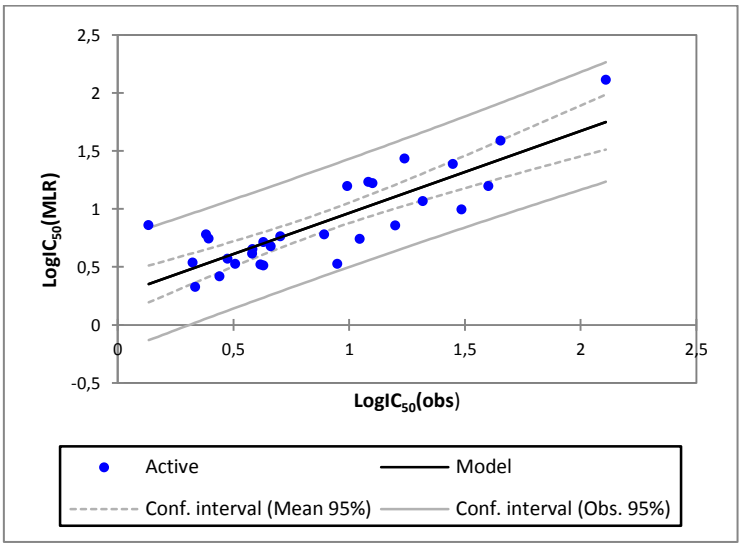


Figure 3: The correlation between the observed and the predicted activities.

The selected descriptors involved in the MLR model of the training set are: Total energy (E), stretch-bend (SB), Torsion (T), VDW energy (E_{VDW}), lipophilic (LogP) and Surface Tension (ST). High correlation coefficient (R = 0.84) of the built QSAR model based on the training set indicates good variance explanation of the model, which is supported by low standard deviation (RMSE = 0.23). Furthermore, the evaluation of the degree of visual significance of the Fischer Test (F) confidence ($p < 0.0001$) reflects the good predictive potency of the generated model.

III.2 Artificial Neural networks

In this study, we have used 3 layers Neural Networks: The input layer contains six neurons representing the selected descriptors, the output layer which represent the observed activity values ($logIC_{50}$) and the hidden layer constituted of 2 neurons. It should be noted that there are no theoretical or empirical rules to determinate the number of hidden layers. While, some authors[32,33] recommended to take into consideration the ρ parameter (ρ= number of weights/number of connections) which must be in the range of $1 < \rho < 3$ [34,35]. Thus, our final ANN architecture is (6-2-1). The **figure 4** shows the correlation between the observed and the predicted activities established by the ANN.

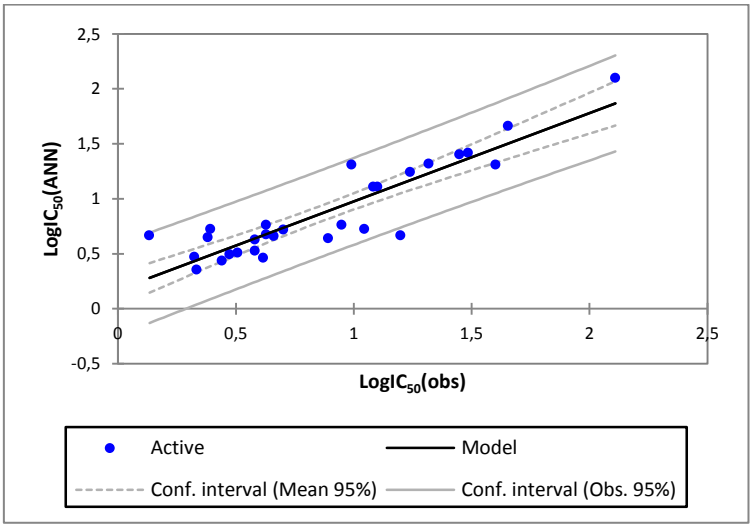


Figure 4: The correlation between the observed and the predicted activities established by ANN.

The plot of observed versus predicted activity (Figure 4) shows a good fit with R^2 value of 0.81 and the mean squared error RMSE = 0.18 indicate that the model possesses a significant statistical quality and the selected descriptors by MLR are pertinent.

III.3 Cross Validation

The results obtained by Cross Validation (CV) with "leave-one-out" are represented in **figure 5**. Therefore, the obtained parameters R^2=0.78 and RMSE= 0.19 show that the built QSAR model is stable and robust.

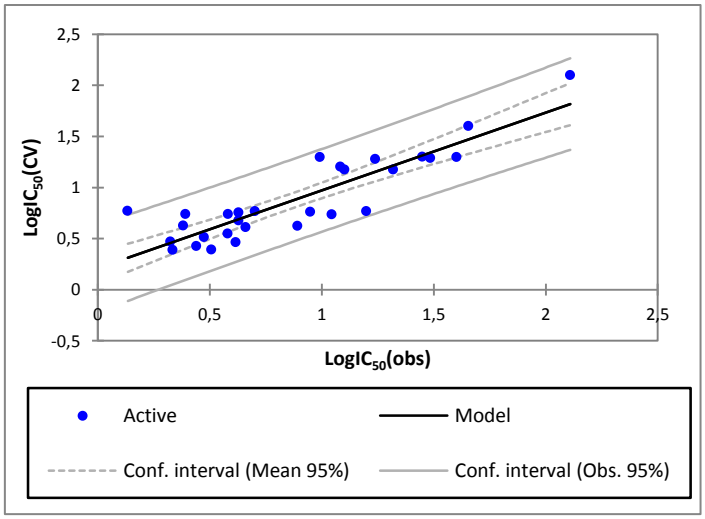


Figure 5: Cross-Validation Correlation (LOO).

However, according to a study conducted by Golbraikh and Tropsha [26], Cross Validation method is not sufficient to estimate the ability of QSAR models. For this reason we have preceded to others validation techniques such as External Validation and Y-randomization.

III.4 Y-Randomization

In order to ensure the reliability of the proposed QSAR model, in a second test, we again removed 8 compounds randomly from the original set of 37 compounds, so that a new model is constructed through the MLR method. The obtained results are presented in table 5.

Table 5 : Calculated results using Y-randomization.

N	4	23	5	26	30	37	2	32	24	21	28	9	7	20	11
$logIC_{50}$	0.134	0.702	2.110	1.102	1.449	0.441	0.393	1.655	0.412	0.629	1.239	1.199	1.046	0.474	0.382
pred $logIC_{50}$	0.869	0.732	2.102	1.254	1.368	0.401	0.774	1.534	0.588	0.543	1.428	0.867	0.772	0.585	0.802
N	36	1	27	25	16	19	13	31	8	12	34	17	33	15	-
$logIC_{50}$	0.769	1.358	1.319	0.661	0.617	0.949	1.486	1.084	0.992	0.324	0.508	0.666	1.308	0.892	-
pred $logIC_{50}$	0.716	1.032	1.080	0.610	0.528	0.556	0.980	1.196	1.211	0.541	0.545	0.626	1.261	0.800	-

The newly QSAR model using the Y-randomization method is represented by the following equation:

$$LogIC50= -15.08 + 0.0026*E - 1.44*S-B +0.16*T + 00.14*VDW + 0.85*LogP + 0.17*ST$$

$$N = 29 \qquad R = 0.83 \qquad R^2 = 0.69 \qquad F = 8.14 \quad RMSE = 0.29$$

The obtained correlation coefficient value R^2=0.69 of the second training set is compared to that obtained with the first training set R^2=0.70, which confirmed that the correlation coefficient of the QSAR model is not found incidentally and attested the good predictive ability

III.5 External Validation

In **table 6**, are presented the 8 compounds that have been reserved for external validation as a test set, maintaining their original numbers as presented in **table 1**.

Table 6 : The results of external validation by the MLR and ANN methods.

N	LogIC$_{50}$	Pred (LogIC$_{50}$) MLR	Residual	Pred (LogIC50)	Residual
1	1.358	0.99	0.388	0.776	0.443
6	0.953	0.984	-0.013	0.769	0.042
10	1.621	2.119	-0.158	2.099	-0.109
17	0.666	0.65	-0.061	0.611	-0.146
24	0.412	0.646	-0.312	0.525	-0.347
29	1.167	1.155	0.079	1.014	0.106
33	1.308	1.326	0.097	1.393	0.013
36	0.769	0.736	-0.02	0.544	-0.002

The results of Golbraikh and Tropsha criteria's **[26]** validation are presented in table 7.

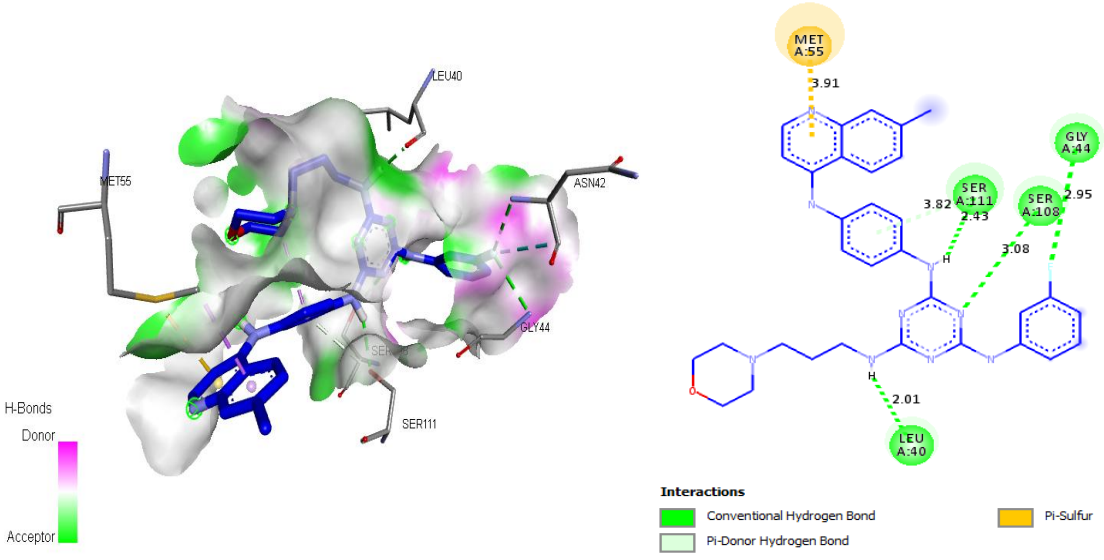
Table 7 : Golbraikh and Tropsha criteria's.

Parameter	Formula	Threshold	Modelscore
$R^2_{ANN\ ext}$	$R^2_{ext} = 1 - \frac{\sum(Y_{pred(test)} - Y_{(test)})^2}{\sum(Y_{(test)} - \bar{Y}_{tr})^2}$	R^2_{ext}>0.6	0.68
r^2	Coefficient of determination for the plot of predicted versus observed for test set by MLR	r^2>0.6	0.74
r_0^2	r^2 at zero intercept		0.67
$r_0'^2$	r2 for the plot of observed versus predicted activity for the test set at zero intercept		0.68
$\lvert r_0^2 - r_0'^2 \rvert$		$\lvert r_0^2 - r_0'^2 \rvert < 0.3$	0.01
k	Slope of the plot of predicted versus observed activity for test set at zero intercept	0.85<k<1.15	0.92
$\frac{r^2 - r_0^2}{r^2}$		$\frac{r^2 - r_0^2}{r^2} < 0.1$	0.07
k'	Slope of the plot of observed versus predicted activity at zero intercept	0.85<k'<1.15	0.93
$\frac{r^2 - r_0'^2}{r^2}$		$\frac{r^2 - r_0'^2}{r^2} < 0.1$	0.08

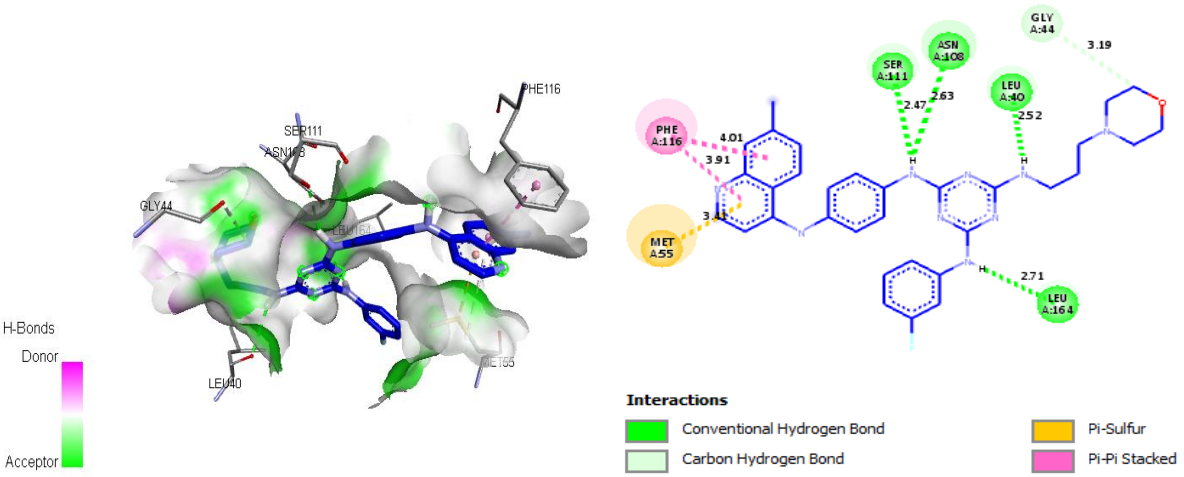
Overall, we can conclude that Golbraikh, Tropsha criteria's and external validation techniques are successfully validated, which indicate that the established QSAR model is in perfect agreement with the different validation methods. In the other hand, the antimalarial activity could be accurately predicted using this QSAR model.

III.6 Docking Studies

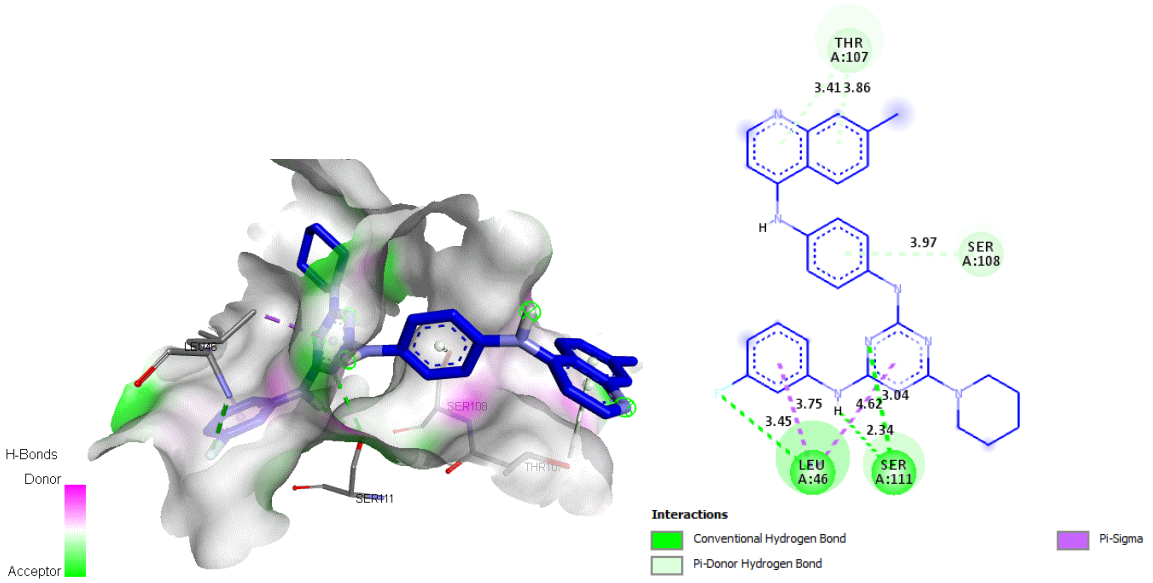
The molecular docking study was performed for two reasons. The first is to understand the good antimalarial activity potency manifested with some compounds. The second is to find out the key interaction types established with the protein (pf-DHFR) in its two types (wild and mutant) [12]. The reported study of Yuvanyama et al [36] has found the binding modes, and has localized the active sites in wild and mutant of protein (pf-DHFR). This study was performed with a potent inhibitor 1,3,5-triazine derivative which is a preclinical molecule called WR99210. It was found that the important sites in the case of the wild type are located in Ile14, Ala16, Met55, Asp54, Ser108, Ile164 and Tyr170, but in the case of the wild type it was found that important sites are located in Ala16, Cys50, Asn51, Cys59, Asn108, Leu164 and Tyr170. The interactions obtained by the docking of compounds 4 and 5 are visualized in **figure 6**.



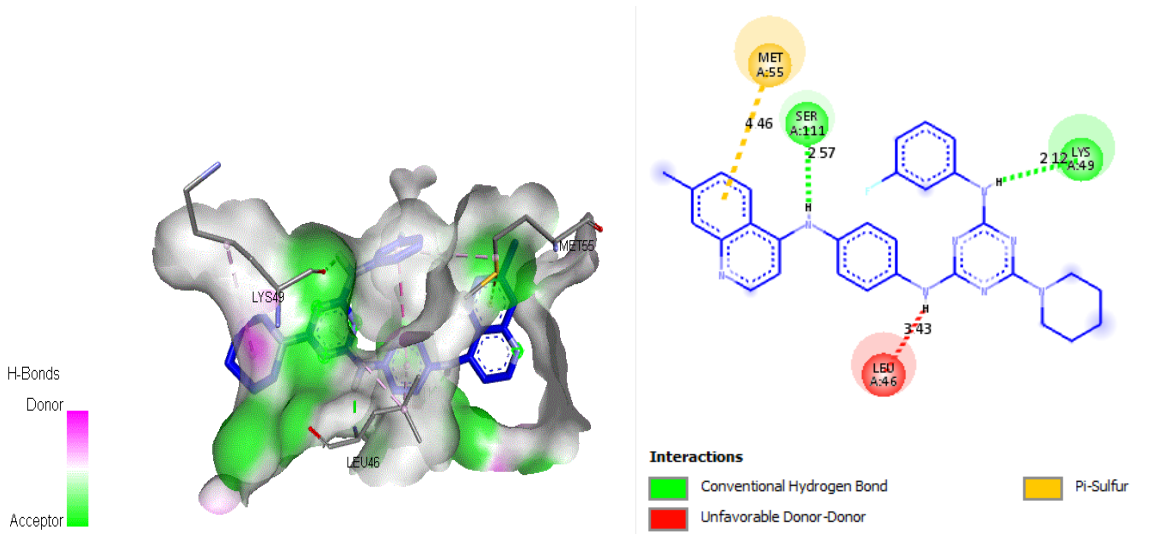
(a)


PHE116
SER111
ASN108
GLY44
LEU164
LEU40
MET55
H-Bonds
Donor
Acceptor
GLY A:44
3.19
ASN A:108
SER A:111
LEU A:40
PHE A:116
4.01
3.91
2.47
2.63
2.52
3.41
MET A:55
2.71
LEU A:164
Interactions
Conventional Hydrogen Bond
Carbon Hydrogen Bond
Pi-Sulfur
Pi-Pi Stacked

(b)


LEU46
SER108
THR107
SER111
H-Bonds
Donor
Acceptor
THR A:107
3.41
3.86
3.97
SER A:108
3.75
4.62
3.04
3.45
2.34
LEU A:46
SER A:111
Interactions
Conventional Hydrogen Bond
Pi-Donor Hydrogen Bond
Pi-Sigma

(c)


MET55
LYS49
LEU46
H-Bonds
Donor
Acceptor
MET A:55
4.46
SER A:111
2.57
2.12
LYS A:49
3.43
LEU A:46
Interactions
Conventional Hydrogen Bond
Unfavorable Donor-Donor
Pi-Sulfur

(d)

FIGURE 6: 2D and 3D docking poses showing interactions of compounds 4 and5 in the binding sites of wild type and quadruple mutant of pf-DHFR. (a) Compound 4 with wild type of pf-DHFR (binding energy −10.6 kcal/mol). (b) Compound 4: quadruple mutant of pf- DHFR (binding energy −10.9 kcal/mol). (c) Compound 5: wild type of pf-DHFR (binding energy −11.4 kcal/mol). (d) Compound 5: quadruple mutant of pf-DHFR (binding energy −10.3 kcal/mol).

In the case of the wild type, compound 4 performs four interactions through hydrogen bonding with SER111, LEU40, SER108 and GLY44 amino acids. These interactions involve three nitrogen atoms linked to the triazine group; the fourth bonding implicates the fluorine atom linked to the benzene group, with the distances of 2.43 Å, 2.01 Å, 3.08 Å and 2.95 Å, respectively. In the case of quadruple mutant, compound 4 performs four hydrogen bonds between four nitrogen atoms linked to the triazine group and SER111, ASN108, LEU164 and LEU40 amino acids with the distance of 2.47 Å, 2.63 Å, 2.71 Å and 2.52 Å, respectively. However, compound 5 performs only three hydrogen bonds with Ser111 and Gly46 in the case of wild type, and two interactions with SER 111 and LYS49 in the case of the quadruple mutant, we note that these binding sites are not cited as active sites for antimalarial activity.

In summary, the interactions formed by the compounds 4 with the binding sites of pf-DHFR are in good agreement with a previous study for antimalarial activity [37]. Moreover, no significant interactions with critical amino acids in pf-DHFR protein on both wild and mutant types, showed for compound 5. In addition we noted that with the 1-substituent N-methy-3-morpholinopropan-1-amine, in the case of compound 4, the number of rotations increased (11 rotations versus 7 rotations in the case of compound 5), which has led to the molecule to be more flexible inside the active site of the protein, furthermore the number of nitrogen atoms increased which gave more chance to the molecule to be more active, in fact the number of hydrogen bonding has been increased compared to compound 5.

IV. Conclusion

In the purpose of producing new effective antimalarial drugs, a QSAR model was developed using the *in vitro* antimalarial data. The built model is statistically significant. Molecular docking study highlights the interaction mode of the ligands with the active sites residues i.e. ILE164, SER108 and LEU40, helps to explain the differences of IC50 values and showed the importance of the structure of R_2 substituent.

References

[1] J.B. Vaishnani, Cutaneous findings in five cases of malaria., Indian J. Dermatol. Venereol. Leprol. 77 (2011) 110.

[2] M.C. Murray, M.E. Perkins, Chapter 15. Chemotherapy of Malaria, Annu. Rep. Med. Chem. 31 (1996) 141–150.

[3] N.J. White, S. Pukrittayakamee, T.T. Hien, M.A. Faiz, O.A. Mokuolu, A.M. Dondorp, Malaria, Lancet. 383 (2014) 723–735.

[4] P. Newton, BM, DPhil, MRCP, N. White, DSc, MD, FRCP, MALARIA: New Developments in Treatment and Prevention, Annu. Rev. Med. 50 (1999) 179–192.

[5] WHO | Global Malaria Mapper, WHO. (2016). http://www.who.int/malaria/publications/world_malaria_report/global_malaria_mapper/en/.

[6] A.R. Brock, J. V. Ross, S. Parikh, A. Esterman, The role of antimalarial quality in the emergence and transmission of resistance, Med. Hypotheses. 111 (2018) 49–54.

[7] C. Prosser, W. Meyer, J. Ellis, R. Lee, Evolutionary ARMS Race: Antimalarial Resistance Molecular Surveillance, Trends Parasitol. 34 (2018) 322–334.

[8] F.W. Muregi, A. Ishih, Next-generation antimalarial drugs: hybrid molecules as a new strategy in drug design, Drug Dev. Res. 71 (2009) n/a-n/a.

[9] M.D. Hastings, C.H. Sibley, Pyrimethamine and WR99210 exert opposing selection on dihydrofolate reductase from Plasmodium vivax., Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 13137–41.

[10] N. Boechat, M. de L.G. Ferreira, L.C.S. Pinheiro, A.M.L. Jesus, M.M.M. Leite, C.C.S. Júnior, A.C.C. Aguiar, I.M. de Andrade, A.U. Krettli, New Compounds Hybrids 1 *H* -1,2,3-Triazole-Quinoline Against *Plasmodium falciparum*, Chem. Biol. Drug Des. 84 (2014) 325–332.

[11] A. Mishra, H. Batchu, K. Srivastava, P. Singh, P.K. Shukla, S. Batra, Synthesis and evaluation of new diaryl ether and quinoline hybrids as potential antiplasmodial and antimicrobial agents, Bioorg. Med. Chem. Lett. 24 (2014) 1719–1723.

[12] H. Hadni, M. Mazigh, E. Charif, A. Bouayad, M. Elhallaoui, Molecular Modeling of Antimalarial Agents by 3D-QSAR Study and Molecular Docking of Two Hybrids 4-Aminoquinoline-1,3,5-triazine and 4-Aminoquinoline-oxalamide Derivatives with the Receptor Protein in Its Both Wild and Mutant Types, Biochem. Res. Int. 2018 (2018) 1–15.

[13] M. Sharma, K. Chauhan, S.S. Chauhan, A. Kumar, S.V. Singh, J.K. Saxena, P. Agarwal, K. Srivastava, S. Raja Kumar, S.K. Puri, P. Shah, M.I. Siddiqi, P.M.S. Chauhan, Synthesis of hybrid 4-anilinoquinoline triazines as potent antimalarial agents, their in silico modeling and bioevaluation as Plasmodium falciparum transketolase and ??-hematin inhibitors, Medchemcomm. 3 (2012) 71–79.

[14] M.A. Lill, Multi-dimensional QSAR in drug discovery, Drug Discov. Today. 12 (2007) 1013–1017.

[15] Q.-S. Du, J. Gao, Y.-T. Wei, L.-Q. Du, S.-Q. Wang, R.-B. Huang, Structure-Based and Multiple Potential Three-Dimensional Quantitative Structure–Activity Relationship (SB-MP-3D-QSAR) for Inhibitor Design, J. Chem. Inf. Model. 52 (2012) 996–1004.

[16] D. Hoekman, Exploring QSAR Fundamentals and Applications in Chemistry and Biology, Volume 1. Hydrophobic, Electronic and Steric Constants, Volume 2 J. Am. Chem. Soc. 1995, 117, 9782, (1996).

[17] W. Sirawaraporn, T. Sathitkul, R. Sirawaraporn, Y. Yuthavong, D. V Santi, Antifolate-resistant mutants of Plasmodium falciparum dihydrofolate reductase., Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 1124–9. http://www.ncbi.nlm.nih.gov/pubmed/9037017.

[18] T. Österberg, U. Norinder, Prediction of drug transport processes using simple parameters and PLS statistics The use of ACD/logP and ACD/ChemSketch descriptors, Eur. J. Pharm. Sci. 12 (2001) 327–337.

[19] G.W.A. Milne, Software Review of ChemBioDraw 12.0 ChemBioDraw 12.0 CambridgeSoft 100 CambridgePark Drive, Cambridge, MA 02140, United States . http://www.cambridgesoft.com/software/ChemBioDraw/, J. Chem. Inf. Model. 50 (2010) 2053–2053.

[20] R.G. Parr, W. Yang, Density-Functional Theory of Atoms and Molecules, 1989..

[21] A.D. Becke, Density-functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648–5652.

[22] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J. Montgomery J. A., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich,

A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. V Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision C.02, 2004.

[23] I. Flood, N. Kartam, Neural Networks in Civil Engineering. II: Systems and Application, J. Comput. Civ. Eng. 8 (1994) 149–162.

[24] B. Efron, Estimating the error rate of a prediction rule: Improvement on cross-validation, J. Am. Stat. Assoc. 78 (1983) 316–331.

[25] M. Stone, Cross-Validatory Choice and Assessment of Statistical Predictions, J. R. Stat. Soc. (1974)..

[26] A. Golbraikh, A. Tropsha, Beware of q2!, J. Mol. Graph. Model. 20 (2002) 269–276.

[27] A. Tropsha, Best practices for QSAR model development, validation, and exploitation, Mol. Inform. 29 (2010) 476–488.

[28] A. Kouranov, L. Xie, J. de la Cruz, L. Chen, J. Westbrook, P.E. Bourne, H.M. Berman, The RCSB PDB information portal for structural genomics, Nucleic Acids Res. 34 (2006) D302–D305.

[29] D.S. BIOvIA, Discovery studio modeling environment., San Diego, Dassault Syst. Release, 4. (2015).

[30] G. Morris, R. Huey, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J. …. 30 (2009) 2785–2791.

[31] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, J. Comput. Chem. 19 (1998) 1639–1662..

[32] S.S. So, W.G. Richards, Application of neural networks: quantitative structure-activity relationships of the derivatives of 2,4-diamino-5-(substituted-benzyl)pyrimidines as DHFR inhibitors, J. Med. Chem. 35 (1992) 3201–3207.

[33] T.A. Andrea, H. Kalayeh, Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors, J. Med. Chem. 34 (1991) 2824–2836..

[34] M.E. N. Naceiri Mrabti, K .Dguigui, H. Hadni, QSAR study and molecular docking of benzimidazole derivatives as potent activators of AMP-activated protein kinase, J. Taibah Univ. Sci. 11 (2017) 18–39.

[35] V. Kůrková, Kolmogorov's theorem and multilayer neural networks, Neural Networks. (1992).

[36] J. Yuvaniyama, P. Chitnumsub, S. Kamchonwongpaisan, J. Vanichtanankul, W. Sirawaraporn, P. Taylor, M.D. Walkinshaw, Y. Yuthavong, Insights into antifolate resistance from malarial DHFR-TS structures, Nat. Struct. Biol. 10 (2003) 357–365.

[37] J. Yuvaniyama, P. Chitnumsub, S. Kamchonwongpaisan, J. Vanichtanankul, W. Sirawaraporn, P. Taylor, M.D. Walkinshaw, Y. Yuthavong, Insights into antifolate resistance from malarial DHFR-TS structures, Nat. Struct. Mol. Biol. 10 (2003) 357–365.

Conclusion générale

Le travail mené dans le cadre de cette thèse, a été consacré à la modélisation de l'activité antipaludique, manifesté par des molécules hétérocycliques hybrides à base de quinoline, triazine et oxalamide. Pour mener à bien cette étude, nous avons fait appel à plusieurs méthodes de modélisation moléculaire à savoir :

- La méthode QSAR pour établir des modèles statistiques sur la base d'une variété de séries de molécules ayant une activité antipaludique. Ces modèles ont montré une robustesse, fiabilité et applicabilité remarquables pour prédire efficacement l'activité étudiée sur des molécules de la même famille.
- La méthode du docking moléculaire a été utilisée pour prévoir le mode d'interaction du pharmacophore responsable de l'activité biologique avec la protéine *Plasmodium falciparum dihydrofolate réductase* (pf-DHFR), sous ses deux formes sauvage et mutante.

Les études 3D-QSAR menées sur trois séries de molécules, une série de 34 molécules hybrides dérivées du 4-aminoquinoline-triazine et de 4-aminoquinoline-oxalamide, une série de 41 hybrides dérivés du 4-anilinoquinoline-triazine, et une série de 37 hybrides dérivés du 4-anilinoquinoline-triazine, ont montré que les descripteurs calculés tel que l'énergie HOMO, l'énergie totale, l'énergie de répulsion, torsion, stretch-bend, et la lipophilie s'avèrent être les descripteurs les plus pertinents pour la construction des modèles MLR et ANN. Ces modèles prédictifs ont été validés par différentes méthodes de validation interne et externe montrant la fiabilité de prédiction de l'activité antipaludique.

L'analyse effectué par le docking moléculaire sur la série composée des hybrides dérivés du 4-aminoquinoline-triazine et du 4-aminoquinoline-oxalamide, a montré que l'hybride le plus actif de la série des drivés du 4-aminoquinoline-triazine a formé trois interactions de liaisons hydrogènes, et le plus actif de la série des 4-aminoquinoline-oxalamides n'a formé que deux, cependant l'hybride le moins actif de la série et faisant partie des dérivés du 4-aminoquinoline-oxalamide, n'a formé qu'une seule liaison hydrogène. Ces observations nous ont permis d'expliquer la différence d'activité observée entre les composés, en particulier entre la famille des triazines, considérée comme la famille la plus active, et celle de l'oxalamide. En ce qui concerne le docking moléculaire effectué sur la série des 4-

anilinoquinoline-triazines a révélé que la géométrie des molécules est d'une grande importance pour l'activité antipaludique. En effet, le composé 11 (isomère para) ayant une activité très élevée par rapport au composé 38 (isomère méta), ceci a été traduit par le nombre d'interactions établies entre le composé 11 et les sites actifs de la protéine. Finalement le docking moléculaire réalisé avec les composés le plus actif et le moins actif 4 et 5 de la série des 4-anilinoquinoline-triazine a montré que le composé 4 forme plus d'interactions avec les sites les plus actifs du récepteur pf-DHFR. Enfaite ceci peut être expliqué par la flexibilité de la structure du compose 4 par rapport au composé 5 ce qui lui permettrait de faire des rotations simples de façon à s'adapter aux sites du récepteur.

Enfin, la forte fiabilité des modèles QSAR établis dans ce travail, ainsi que les observations faites avec le docking moléculaire concernant les modes de liaison prévus et les interactions moléculaires clés, pourraient être exploités pour l'orientation du design et ensuite de la synthèse dans le domaine de la recherche pharmaceutique de l'activité antipaludique.