

Année 2022

N°:MM502022

MEMOIRE DE MASTER

MASTER DE « **BIOTECHNOLOGIE MEDICALE** »

OPTION : « **Biomédicale** »

Intitulé

Exploration des inhibiteurs de STS par approche QSAR pour le traitement de cancer du sein

Soutenue par :

Zineb SOUBAI

Le 06/07/2022

Devant le jury composé de :

Pr. OUADGHIRI Mouna, Faculté de Médecine et de Pharmacie de Rabat, **Présidente**

Pr. MABROUKI Jamal, Faculté des Sciences de Rabat, **Encadrant**

Dr. ABOU-ELAAZ Fatima-zahra, Institut Scientifique de Rabat, **Co-encadrante**

Pr. KANDOUSSI Ilham, Faculté de Médecine et de Pharmacie de Rabat, **Examinatrice**

Remerciement

Je tiens tout d'abord à remercier tous les membres du laboratoire de biotechnologie de la Faculté de Médecine et de Pharmacie de Rabat, et surtout au **Pr. IBRAHIMI Azzedine**, pour cette formation qui nous a permis de réaliser nos propres ambitions et de perfectionner nos compétences.

Je veux bien remercier **Pr. OUADGHIRI Mouna**, la coordinatrice de ce Master pour son orientation et pour ses conseils pertinents et pour sa gentillesse durant toute cette formation.

Je suis reconnaissante à mon encadrant **Pr. MABROUKI Jamal**, professeur à la faculté de science, Rabat pour son encadrement et soutien et pour son encouragement.

Je remercie chaleureusement ma co-encadrante **ABOU-ELAAZ Fatima-zahra** Doctorante à l'institut scientifique, Rabat pour m'avoir aidé pendant tout ce travail, et pour son soutien et encouragement et pour sa gentillesse

Je remercie infiniment **Pr. KANDOUSSE Ilham**, Veuillez accepter ma profonde reconnaissance et admiration et mes remerciements les plus sincères d'avoir accepté de juger ce travail.

Je tiens à remercier très sincèrement tous **les membres de ma famille** pour leur soutien et leurs encouragements constants durant tous ces années d'études.

Dédicace

Avec l'expression de ma reconnaissance, je dédie ce modeste travail à ceux qui, quels que soient les termes embrassés, je n'arriverais jamais à leur exprimer mon amour sincère.

A la mémoire de mon père, mon cher père **SOUBAI Ahmed**, Que Dieu le tout puissant l'accueille en son vaste Paradis.

A la femme qui a souffert sans me laisser souffrir, qui n'a jamais dit non à mes exigences et qui n'a épargné aucun effort pour me rendre heureuse: mon adorable mère **SENNABI Fatima**.

A mes frères **SOUBAI Youssef** et **SOUBAI Mohamed** qui n'ont pas cessé de me conseiller, encourager et soutenir tout au long de mes études. Que Dieu les protège et leurs offre la chance et le bonheur.

A mon adorable sœur **SOUBAI Meryam** qui sait toujours comment procurer la joie et le bonheur pour toute la famille.

A mes petits frères **MOUAD**, **HAMZA** et **ISRAE**. Que Dieu leur donne une joyeuse vie.

A tous les amis que j'ai connu jusqu'à maintenant.

Résumé

Le stéroïde sulfatase (STS) est une enzyme de stéroïdogénèse qui augmente la concentration des œstrogènes dans la circulation sanguine, et considérée comme un des principaux facteurs conduisant au cancer du sein. Au niveau moléculaire, 87 % des STS sont exprimées. Et un inhibiteur ciblant les STS pourrait diminuer le niveau d'œstrogènes.

Une analyse 2D QSAR utilisant la méthode de régression linéaire multiple (MLR) a été menée sur un ensemble de données de 29 dérivés de 4-(thio) chroménone 6-O-sulfamate 2- substitué agissant comme inhibiteurs de STS afin de trouver une relation quantitative entre l'activité biologique et les descripteurs moléculaires et de savoir les descripteurs potentiels qui affecte l'inhibition de STS. Un modèle robuste et prédictif a été développé avec un bon R^2 (0.89) et une valeur de coefficient de corrélation obtenu par la validation croisée LOO Q^2 (0.83) explique bien la relation de manière quantitative. En plus de modèle QSAR les profils pharmacocinétiques de 29 molécules ont été validés selon les règles de Lipinski et Veber pour tester la probabilité que ces molécules soient des candidats médicaments.

Mots clés: Sulfamates, QSAR, MLR, Validation croisée LOO.

Abstract

Steroid sulfatase (STS) is a steroidogenic enzyme that increases the concentration of estrogen in the circulation, one of the main factors leading to breast cancer. At the molecular level, 87% of STS is expressed. And an inhibitor targeting STS could decrease estrogen levels.

A 2D QSAR analysis using the multiple linear regression method was conducted on a dataset of 29 2- substituted 4-(thio)chromenone 6-O-sulfamate derivatives acting as STS inhibitors to find a quantitative relationship between biological activity and molecular descriptors and to identify potential descriptors that affect STS inhibition. A robust and predictive model was developed with a good R^2 (0.89) and a correlation coefficient value obtained by LOO Q2 cross-validation (0.83) explains the relationship well in a quantitative way. In addition to the QSAR model, the pharmacokinetic profiles of 29 molecules were validated according to the Lipinski and Veber rules to test the probability that these molecules are drug candidates.

Key words: Sulfamates, QSAR, MLR, LOO cross-validation.

ملخص

ستيرويد سيلفاتاز (STS) هو انزيم ستيرويدوجين يزيد من تركيز هرمون الاستروجين في الدورة وهو احد العوامل الرئيسية المؤدية الى سرطان الثدي. على المستوى الجزيئي, يتم التعبير عن 87% من هذا الانزيم ومكن للكابح الذي يستهدف STS ان يقلل من نسبة هرمون الاستروجين.

تم اجراء دراسة QSAR ثنائية الابعاد باستخدام طريقة الانحدار الخطي المتعدد على مجموعة بيانات مكونة من 29 من مشتقات الكرومينون 4- (ثيو) 6- O- Sulfamate تعمل ككابحات STS, لايجاد علاقة كمية بين النشاط البيولوجي والاصاف الجزيئية وتحديد الالوصاف المحتملة التي تؤثر على عملية كبح STS تم تطوير نموذج تنبؤي مع R^2 جيد (0.89) وقيمة معامل الارتباط Q^2 التي تم الحصول عليها بواسطة التحقق من الصحة LOO (0.83) يشرح العلاقة جيدا بطريقة كمية. بالاضافة الى نموذج QSAR تم التحقق من صحة خصائص الحرائك الدوائية ل 29 جزيئة وفقا لقواعد Lipinski و Veber لاختبار احتمال كون هذه الجزيئات ادوية مرشحة.

الكلمات المفاتيح : QSAR, Sulfamates , MLR , التحقق المزدوج

Table des Matières

Introduction	1
REVUE BIBLIOGRAPHIE	2
I. LE CANCER DE SEIN	3
I.1 Epidémiologie du cancer du sein	3
I.1.1 Incidence et mortalité	3
I.1.2 Facteurs de risques	3
I.1.2.1 Facteurs hormonaux endogènes.....	3
I.1.2.1.1 La nulliparité	3
I.1.2.1.2 L'âge à la première maternité	4
I.1.2.1.3 L'allaitement	4
I.1.2.1.4 Puberté et ménopause.....	4
I.1.2.2 Facteurs hormonaux exogènes.....	4
I.1.2.3 Exposition environnementale	5
I.1.2.4 Facteurs génétiques.....	5
I.1.3 Développement du cancer du sein.....	5
I.1.3.1 Anatomie et physiologie du tissu mammaire	6
I.1.3.2 Diagnostique et physiopathologie	7
I.1.4 Exposition aux œstrogènes et conséquences sur le cancer du sein	7
I.1.4.1 Généralités	7
I.1.4.2 La signalisation œstrogénique	9
I.2 Les types de cancer de seins.....	10
I.2.1 Classification anatomopathologique	10

I.2.2	Classification moléculaire de cancer de sein.....	12
I.3	Les traitements	13
I.3.1	La privation en estrogènes des cellules tumorales	13
I.3.1.1	La suppression ovarienne	13
I.3.1.2	Les inhibiteurs de l'aromatase et de sulfatase	13
I.3.2	Le blocage des RE.....	14
II.	LES SULFAMATES	14
II.1	Les dérivés de sulfamates.....	14
II.2	Les sulfamates naturels	15
II.3	Les applications pharmacologiques	15
II.3.1	Les sulfamates antiviraux	16
II.3.2	Les sulfamates anti-cancéreux.....	16
II.3.3	Les sulfamates Antiépileptique	16
II.3.4	Le traitement de l'obésité.....	16
II.3.5	Traitement de l'hyperlipidémie et de l'athérosclérose	17
II.3.6	Les sulfamates anti-arthrite et anti-ostéoporose	17
III.	GÉNÉRALITÉ SUR QSAR.....	17
III.1	Définition	17
III.2	Principe.....	18
III.3	Présentation de la méthode QSAR	18
III.3.1	Le choix de base de données	18
III.3.2	La division de donnée.....	19
III.3.3	Génération de Descripteurs	19
III.3.4	Le développement du modèle.....	19

III.3.5	Validation de modèle.....	19
III.3.5.1	Validation interne.....	20
III.3.5.2	Validation externe.....	20
III.3.6	Le domaine d'applicabilité.....	21
MATÉRIEL ET MÉTHODES.....		22
I.	La construction de modèle QSAR.....	23
I.1	Préparation des données.....	23
I.1.1	Calcul de descripteurs.....	23
I.1.2	L'équation du modèle QSAR.....	23
I.1.3	Validation de modèle.....	24
I.2	Le domaine d'applicabilité.....	24
I.3	Etude des propriétés physico-chimiques des paramètres de Lipinski et de Veber.....	24
RÉSULTATS ET DISCUSSION.....		25
I.	Validation de Model choisis.....	33
I.1	Validation interne.....	33
I.2	La validation externe.....	34
II.	Le domaine d'application.....	37
III.	Contribution du descripteur.....	39
III.1	Interprétation des descripteurs sélectionnés.....	40
IV.	Les règles de 5 de Lipinski et de veber.....	40
Conclusion et Perspective.....		44

Liste des figures

Figure 1: Structure du sein chez un sujet sein	6
Figure 2 : synthèse d'estrogène à partir de cholestérol	8
Figure 3 : Différents mécanismes d'action des œstrogènes	10
Figure 4 : les dérivés de sulfamates	15
Figure 5 : Principe de l'approche QSAR	18
Figure 6 : Technique de validation LOO	20
Figure 7: Relation linéaire entre la pIC50 expérimentale et la pIC50 observée	36
Figure 8: Représentations graphique des valeurs résiduelles en fonction des valeurs	37
Figure 9: La contribution des descripteurs choisis dans le modèle	39
Figure 10: les valeurs de paramètres de lipinski (NHA, NHD, LogP) des 29 composés	42
Figure 11: Les valeurs de poids moléculaire des 29 composés.....	42
Figure 12: Les paramètres de Veber des 29 composés	42

Liste des tableaux

Tableau I: Structure chimique et activité biologique expérimentale des dérivés de 4-(thio)chroménone 6-O-sulfamate 2- substitué.....	26
Tableau II: les modèles QSAR obtenus à partir de 15 combinaisons	30
Tableau III: Matrice de corrélation entre les descripteurs sélectionnés	32
Tableau IV: Les facteurs d'inflation de la variance (VIF).....	32
Tableau V: Résultats de validation interne	33
Tableau VI: Résultats de la validation externe.....	34
Tableau VII: Statut du domaine d'applicabilité du de l'ensemble d'apprentissage.....	38
Tableau VIII: Statut du domaine d'applicabilité du de l'ensemble d'essai	39
Tableau IX: Singnification des descripteurs choisis	40

Liste des abréviations

ACAT: cholesterol acyltransferase
AD: Applicability Domain
ANA: Anastrozole
AND: acide désoxyribonucléique
AP-1: Activator protein 1
BRCA: Breast cancer
CO: Contraceptifs oraux
CTNB: Breast Cancer Triple Negative
DDT: Dichlorodiphényltrichloroéthane
DNP: Natural Products data base
ERE: estrogène response element
ESR1: Estrogen receptor 1
EXE: Exémestane
FAD: Food and Drug Administration
FSH: follicle-stimulating hormone
GH: Growth Hormon
GPER: G protein-coupled estrogen receptor 1
HER2: human epidermal growth factor receptor 2
HR+: Hormone Receptors +
IGF1: insulin-like growth factor 1
LET: Létrozole
LH-RH: Luteinizing hormone-releasing hormone
LOO: Leave One Out
MLR: Multiple Linear Regression
PCB: Polychlorobiphényle
PLS: partial least squares
PR: Progesterone Receptor
QSAR: Quantitative structure-activity relationship
RE: Récepteur d'estrogène
RNA: Artificial Neural Networks
SERD: selective estrogen receptor degraders
SERM: Selective estrogen receptor modulators
SO: Supression ovarienne
STS: Stéroïd sulfatase
THS: Traitement Hormonal Substitutif
TP53: Tumor protein 53
VIH: virus de l'immunodéficience humaine

Introduction

Le cancer du sein est le cancer le plus fréquemment diagnostiqué chez les femmes. Malgré les progrès récents en matière de détection précoce et de traitement, le cancer du sein reste la deuxième cause de mortalité par cancer chez les femmes dans le monde.

Au moment de diagnostic, environ 70% à 75% des patients ont un cancer du sein a récepteurs d'oestrogènes (RE) positifs, parmi les deux types de récepteurs d'oestrogènes α et β le RE $_{\alpha}$ est le sous type prédominant exprimé dans les tissus des tumeurs mammaires [1]. Ce récepteur est un médiateur majeur de la tumorigénicité des œstrogènes dans les cellules.

En se liant à ses récepteurs, l'œstrogène active des cascades de signalisation qui régulent l'expression de nombreux gènes qui contrôlent le métabolisme, la progression du cycle cellulaire ou l'apoptose, et finalement favorisent la croissance tumorale.

La dépendance aux œstrogènes a été utilisée avec succès dans les cancers ER-positifs afin de développer des thérapies antihormonales. Cette dernière est basée principalement sur la privation en estrogènes des cellules tumorales par suppression ovarienne (SO) ou inhibiteurs de l'aromatase ou sur le blocage des RE [2] [1].

Nombreuses études ont été menées pour identifier une nouvelle cible pour inhiber la synthèse des oestrogènes dans la glande mammaire, elles ont révélé l'existence de deux enzymes : L'aromatase et stéroïde sulfatase (STS) qui augmentent la concentration des oestrogènes dans la circulation, Néanmoins l'activité de STS dans les tissus cancéreux est cent fois plus que la voie d'aromatase, la chose qui le rend une cible plus prometteuse pour découvrir de nouveaux candidats pour le traitement du cancer du sein [3].

Plusieurs études ont montré l'importance de sulfamates substitués dans l'inhibition de STS. Le but de notre étude est de développer un modèle 2D QSAR à partir de 29 dérivés de 4-(thio) chroménone 6-O-sulfamate 2- substitué, et de vérifier le profil pharmacocinétique des 29 molécules en se basant sur les règles de Lipinski et de Veber.

REVUE BIBLIOGRAPHIE

I. LE CANCER DE SEIN

I.1 Epidémiologie du cancer du sein

I.1.1 Incidence et mortalité

Le cancer est un problème majeur de la santé publique, selon l'organisation mondiale de la santé il constitue la deuxième cause de décès dans le monde, à l'origine de 8,8 millions de décès en 2015, et près de 10 millions de décès en 2020.

En 2020, le cancer le plus courant en termes de nombre de cas recensés était le cancer du sein avec 2,26 millions de cas et 685 000 décès.

Au Maroc, la localisation la plus fréquente, en considérant les deux sexes, était le cancer du sein qui occupait le premier rang et il représentait 19,2%, suivi par le cancer du poumon 12,3% et le cancer colorectal 7,8%, La quasi-totalité des patients atteints du cancer du sein était du sexe féminin avec une proportion de 99,1%. Le sexe masculin représentait moins de 1% des cas enregistrés [4] [5].

I.1.2 Facteurs de risques

Le cancer du sein est une maladie multifactorielle, il n'y a donc jamais de cause unique au développement d'un cancer. Malgré les progrès qui ont permis d'identifier un certain nombre de facteurs (facteurs hormonaux, facteurs génétiques...), favorisant le cancer du sein, il est toujours difficile d'évaluer l'implication et le poids de nombreux de ces facteurs. Néanmoins, la connaissance de ces facteurs de risque permet la mise en œuvre de stratégies efficaces de prévention des cancers et de dépistages ciblés.

I.1.2.1 Facteurs hormonaux endogènes

I.1.2.1.1 La nulliparité

L'un des principaux facteurs recensés est la nulliparité. Les femmes n'ayant jamais vécues d'accouchement sont plus risqué par rapport aux femmes ayant d'enfants, et ce risque diminue avec l'augmentation du nombre de naissances. Cela peut être expliqué par le fait que la grossesse entraîne l'interruption de l'exposition des cellules mammaires à l'œstrogène circulant [6].

I.1.2.1.2 L'âge à la première maternité

L'âge des mères à la naissance de leur premier enfant est fortement corrélé au cancer du sein. Plus les femmes mènent leur première grossesse à terme avant l'âge de 30 ans, plus le risque est réduit. Cet effet protecteur est lié à la grossesse qui va permettre la différenciation de la glande mammaire pour transformer le sein dans le sens de la lactation et les cellules différenciées sont moins sensibles aux carcinogènes en comparaison avec les cellules immatures [6].

I.1.2.1.3 L'allaitement

Les études récentes montrent que l'allaitement favorise la réduction du risque de cancer du sein mais uniquement dans les cas où il est prolongé. Ce risque décroît d'environ 4 % pour chaque année où l'enfant allaite. Le rôle protecteur de l'allaitement pourrait être dû à la sécrétion de prolactine et la réduction de la sécrétion d'oestrogènes [6].

I.1.2.1.4 Puberté et ménopause

La puberté précoce généralement avant 13 ans et la ménopause tardive, après 55 ans augmente le risque de cancer du sein. Ceci pourrait être justifié par l'exposition prolongée des cellules mammaires aux hormones ovariennes et aux oestrogènes en particulier [6].

I.1.2.2 Facteurs hormonaux exogènes

Contraceptifs oraux

Les contraceptifs oraux sont des composés d'hormones qui imitent l'action des hormones naturelles. Ils visent à inhiber l'ovulation pour empêcher la survenue d'une grossesse. Son utilisation accroît le risque de cancer de sein d'environ 25%. Ce risque ne varie pas de manière significative avec la durée d'utilisation ou bien en fonction du type d'oestrogène ou de combinaison des préparations utilisées. Néanmoins, il chute dès l'arrêt de la consommation de ces médicaments[7].

Traitement hormonal substitutif (THS)

Le THS est utilisé pour compenser la baisse des taux d'hormones ovariennes circulantes lors de la ménopause. Comparativement, aux femmes qui n'ont jamais utilisé de THS, les femmes qui l'ont utilisé ont un risque plus élevé de cancer du sein. Ce risque

augmente avec la durée d'utilisation, la prise pendant cinq ans ou plus, et le risque d'atteindre le cancer de 26 % à 35 %. Cependant, une fois le traitement arrêté, le risque diminue. De plus, les effets du THS varient en fonction de composition de traitement, il est plus élevé lors de l'utilisation de l'association œstrogène-progestatif que lors de l'utilisation de l'œstrogène seul[7].

I.1.2.3 Exposition environnementale

Les facteurs environnementaux ont aussi une influence sur le risque de cancer du sein. Les plus décrits sont les produits chimiques organochlorés comme : le polychlorobiphényle (PCB) dichlorodiphényltrichloroéthane (DDT) ou le qui ont des structures similaires aux œstrogènes [6].

I.1.2.4 Facteurs génétiques

Depuis une longue période, l'impact génétique sur le cancer du sein est suggéré et il est estimé que presque de 5 à 10% des cas de cancer du sein sont liés à l'hérédité, aux mutations dans des gènes autosomiques dominantes. Ces mutations affectent généralement deux types de gènes : des gènes codants pour des protéines activatrices de tumeur, qui provoquent la croissance et la division incontrôlée comme HER2 codé par le gène ERBB2, ou des gènes suppresseurs de tumeur comme (BRCA1/2) (TP53) qui réparent les dommages d'ADN et favorisent l'apoptose [8].

I.1.3 Développement du cancer du sein

Normalement les cellules sont programmées pour se multiplier un nombre défini de fois puis mourir, selon le processus de l'apoptose. En revanche, les cellules cancéreuses se prolifèrent d'une façon incontrôlée en engendrant la formation d'une masse appelée tumeurs maligne. Un cancer du sein est donc une tumeur maligne qui se développe au niveau du sein, et pour mieux comprendre comment il se développe, il est nécessaire de comprendre d'abord l'anatomie du sein.

I.1.3.1 Anatomie et physiologie du tissu mammaire

Le sein est composé de tissu glandulaire (sécrétoire) et adipeux (gras) soutenu par un cadre lâche de tissu conjonctif fibreux appelé ligaments de Cooper.

Chez la femme, le tissu glandulaire est composé de lobes comprenant des lobules contenant de 10 à 100 alvéoles, d'environ 0,12 mm diamètre.

Les alvéoles se drainent par un système canalaire (canaux galactophores) qui décrit comme étant de nombreux petits conduits, qui fusionnent pour aboutir à un conduit principal qui se dilate légèrement pour former un sinus lactifère qui transportent le lait jusqu'au mamelon.

Concernant le drainage lymphatique du sein la lymphe est drainée par deux voies principales : les ganglions axillaires et les ganglions mammaires internes [9].

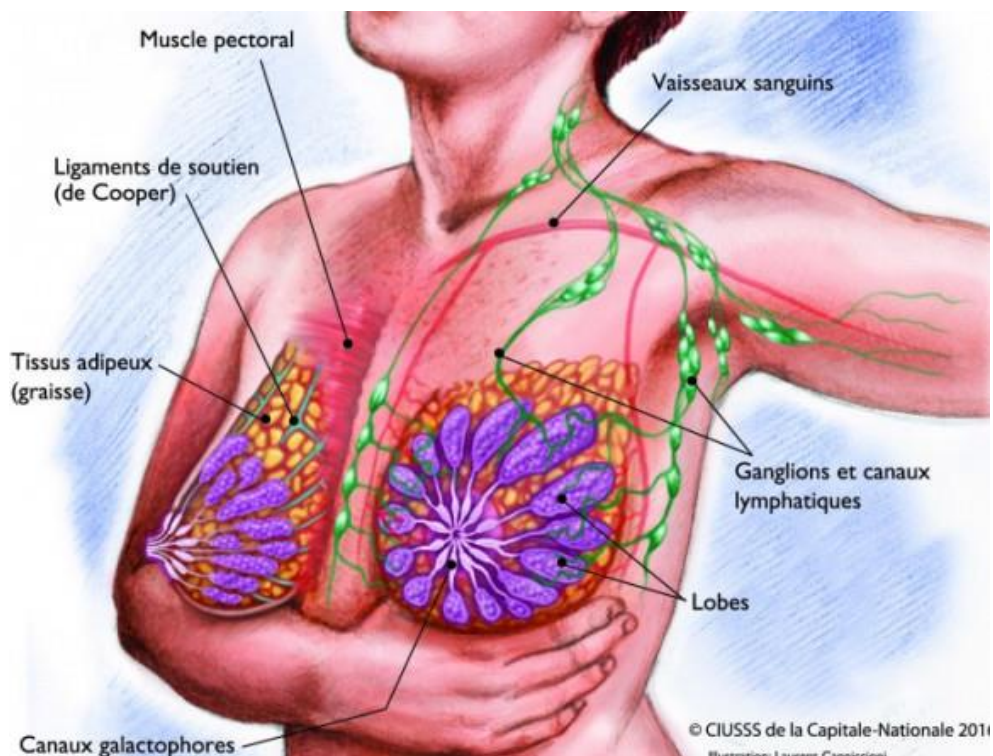


Figure 1: Structure du sein chez un sujet sein [10]

La glande mammaire se développe en plusieurs étapes distinctes : croissance pubertaire, grossesse, lactation et involution, grâce à l'intervention d'hormones qui régulent son développement, comme l'hormone de croissance (GH) et les œstrogènes, et le facteur de croissance essentiel (IGF1) pour le développement des seins à la puberté. Pendant la grossesse, l'action combinée de la progestérone et de la prolactine crée les alvéoles qui sécrètent le lait pendant la lactation [11].

I.1.3.2 Diagnostique et physiopathologie

Le cancer du sein est un diagnostic histologique établi selon des critères pathologiques normalisés. L'histologie la plus courante du cancer du sein est le carcinome canalaire invasif, suivi par le carcinome lobulaire invasif, avec des carcinomes mixtes canaux/lobulaires et d'autres carcinomes plus rares.

Deux cibles moléculaires principales dans la pathogenèse du cancer du sein ont été identifiées. L'une d'entre elles est le récepteur d'œstrogène alpha ($ER\alpha$), qui est exprimé dans environ 70 % des cancers du sein invasifs. $ER\alpha$ est un récepteur des hormones stéroïdes et un facteur de transcription qui, lorsqu'il est activé par les œstrogènes, active les voies de croissance oncogènes dans les cellules cancéreuses du sein.

L'expression du récepteur de la progestérone (PR), une hormone stéroïde étroitement liée, est également un marqueur de la signalisation de l' $ER\alpha$.

L'utilisation d'agents endocriniens pour réduire la signalisation du RE est le principal traitement systémique des cancers du sein ER-positifs ou PR-positifs.

La deuxième cible moléculaire principale est le HER2, un récepteur transmembranaire à tyrosine kinase situé dans la région de l'épiderme.

Le cancer du sein est classé dans les stades I à IV, le stade IV correspondant à une maladie métastatique à distance. Le stade I définis anatomiquement comme une tumeur mammaire de moins de 2 cm et sans atteinte ganglionnaire [12].

I.1.4 Exposition aux œstrogènes et conséquences sur le cancer du sein

I.1.4.1 Généralités

Les œstrogènes sont des hormones sexuelles femelles constituant un groupe de stéroïdes, synthétisées de façon endogène à partir d'un précurseur lipidique, le cholestérol

(figure 2), lors d'un processus complexe appelé stéroïdogénèse. Leur structure est basée sur un noyau œstrane, constitué de quatre cycles hydrocarbonés, les modifications de substitutions telles que des groupes méthyle ou hydroxyle, permettent de différencier les molécules de cette famille.

On distingue:

- Les œstrogènes naturels : il existe trois formes: l'œstrone (E1), le 17βœstradiol (E2) et l'œstriol (E3). Le 17β-œstradiol est l'œstrogène endogène dont l'effet biologique est le plus puissant de l'organisme (c'est un régulateur clé de la croissance, de la différenciation et de la fonction des voies reproductives masculines et féminines, la glande mammaire, les systèmes squelettique et cardiovasculaire.), suivi par l'œstrone, l'œstriol quant à lui n'est observé qu'à de forte concentration pendant la grossesse.
- Les œstrogènes de synthèse : sont le plus souvent dérivés du noyau du 17βœstradiol, parmi les molécules les plus administrée, on trouve l'éthinylestradiol, qui se distingue de l'E2 par son radical éthinyl en C17. Cet œstrogène rentre dans la composition des contraceptifs hormonaux et a pour principal mode d'action l'inhibition de (FSH), provoquant l'atrophie de l'ovaire et inhibition de l'ovulation [13] [14].

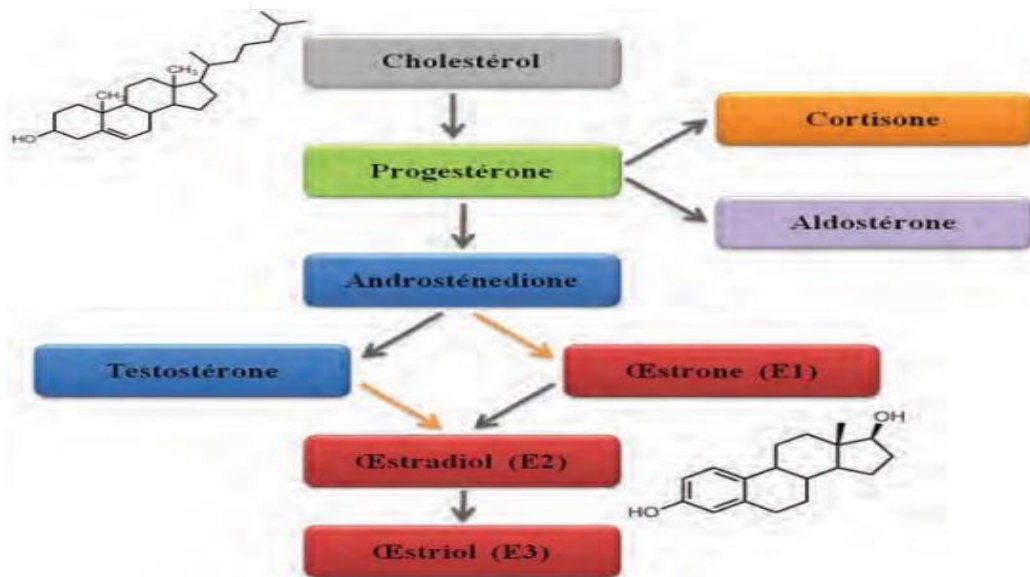


Figure 2 : synthèse d'estrogène à partir de cholestérol [14]

I.1.4.2 La signalisation œstrogénique

Les œstrogènes amplifient les réponses des cellules sensibles aux hormones stéroïdes par leur liaison à des récepteurs nucléaires spécifiques, les récepteurs aux œstrogènes α et β codés respectivement par les gènes ESR1 et ESR2. Ces récepteurs agissent comme un facteur de transcription après fixation d'un ligand.

Les quatre voies suivantes résument le mécanisme moléculaire de signalisation des œstrogènes (fig.3) :

- ❖ **Mécanismes classique de l'action des œstrogènes (dépend de ligand)** : Après liaison d'E2 avec le récepteur un changement de conformation activateur au sein du RE est induit ce qui favorise l'homodimérisation et la liaison avec des éléments de réponse de l'ADN (ERE= œstrogen response element). Après la liaison à l'ADN les récepteurs contactent l'appareil général de transcription soit directement soit indirectement par l'intermédiaire de protéine cofacteurs, et exercent un effet positif ou négatif sur l'expression des gènes cibles en aval [15].
- ❖ **La signalisation indépendante de ligand** : la fonction du RE peut être modulée par des signaux extracellulaire en absence de E2, ceci grâce à la capacité des facteurs EGE, IGF1, ainsi que la capacité de 8-bromo-adénosine monophosphate à imiter le comportement de E2 et activer le RE par une cascade de phosphorylation et à augmenter l'expression de gènes cible du RE[15].
- ❖ **Signalisation indépendante de l'ERE** : le complexe E2-ER modifie la transcription des gènes contenant des éléments de réponse alternatifs tels que l'AP-1 par le biais d'une association avec d'autres facteurs de transcription liés à l'ADN, entraînant une régulation positive de l'expression génique[15].
- ❖ **Signalisation de surface cellulaire (non génomique)** : Généralement à des effets biologiques rapides, ces derniers sont médiés par le biais de forme de RE. Cette fois, le récepteur est membranaire, et il est couplé à la protéine G (GPER) connu sous le nom de GPER30 [16], qui sont liés à des protéines de transduction de signal cellulaire qui génèrent des réponses tissulaires rapides.

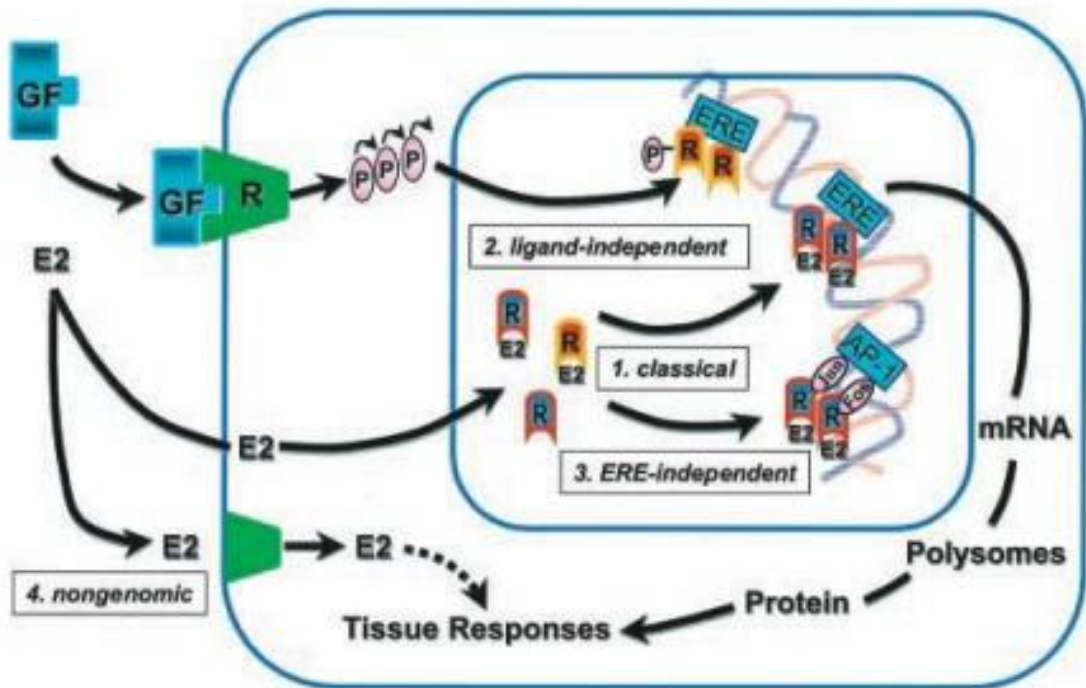


Figure 3 : Différents mécanismes d'action des œstrogènes [13].

I.2 Les types de cancer de seins

I.2.1 Classification anatomopathologique

Le type des cellules à partir desquelles le cancer de sein se développent génère différents types de tumeurs [17].

- **Les adénocarcinomes**

Les adénocarcinomes sont les plus fréquents et représentent 95 % des cancers du sein. Ils se développent à partir des cellules épithéliales de la glande mammaire.

Le plus souvent, les adénocarcinomes naissent à partir des cellules des canaux (cancer canaux) et plus rarement à partir des cellules des lobules (cancer lobulaire).

On distingue les cancers in situ et les cancers infiltrant.

- **Les adénocarcinomes in situ**

Dans ce type de cancer, les cellules cancéreuses se trouvent uniquement à l'intérieur des canaux ou des lobules, et la tumeur ne franchit pas la membrane basale qui les entoure et donc n'infiltré pas le tissu voisin, dans ce cas on l'appelle cancer ou carcinome in situ [17].

- **Les adénocarcinomes infiltrant**

Lorsque les cellules cancéreuses infiltrent le tissu qui entoure les canaux et les lobules. On parle du cancer ou carcinome infiltrant. Presque tous les cancers infiltrants sont des cancers canaux et les cancers lobulaires infiltrants sont très rares. Ils peuvent se propager vers les ganglions ou bien vers d'autres parties du corps. Souvent les ganglions les plus atteints par les cellules du cancer du sein existent sous les bras, au niveau des aisselles : appelés les ganglions axillaires [17].

- **Les formes rares de carcinomes**

- ✚ Le carcinome médullaire: Ce type représente 1 % des cancers du sein infiltrants. Il apparaît davantage chez les femmes de moins de 50 ans. Caractérisé par une bordure le plus souvent ronde et bien définie, il est similaire au fibroadénome à la mammographie ou à la palpation lors de l'examen clinique des seins. Le carcinome médullaire a un pronostic plus favorable que le carcinome canalaire infiltrant. Lorsqu'un carcinome médullaire est découvert chez une femme jeune, les médecins soupçonnent vivement une forme génétique. [17].

- ✚ Le carcinome mucineux : il est rare, représente environ 2 % de tous les cancers du sein infiltrants. Il est aussi appelé carcinome colloïde. La tranche d'âge la plus vulnérable est entre 60 et 70 ans. Il est caractérisé par des cellules cancéreuses qui sécrètent du mucus. Le carcinome mucineux a un meilleur pronostic par rapport aux autres types de cancers du sein infiltrants [17].

- ✚ Le carcinome tubuleux : ce type représente environ 1 à 2 % de tous les cancers du sein infiltrant. Il se manifeste davantage chez les femmes \geq de 55 ans. En raison de l'aspect des cellules au microscope, il est nommé tubulaire. Les cancers tubulaires ont tendance à être petits et ne se propagent généralement pas aux

ganglions lymphatiques de l'aisselle. Il a également un meilleur pronostic que le carcinome canalaire invasif ou le carcinome lobulaire [17].

- ✦ Le carcinome papillaire : ce type de cancer peut être infiltrant ou bien stable. Lorsqu'il est infiltrant, il ne dépasse pas plus de 1 à 2 % de tous les cancers du sein infiltrant. Les femmes âgées sont les plus concernées par ce type de cancer, et généralement a un pronostic plus favorable que le carcinome canalaire infiltrant[17].

I.2.2 Classification moléculaire de cancer de sein

La classification moléculaire est basée généralement sur la présence ou l'absence de récepteurs hormonaux et de récepteurs Her2.

Les informations sur ces récepteurs facilitent la classification de chaque tumeur afin que les traitements les plus efficaces puissent être sélectionnés pour vaincre la maladie.

Cette Classification donne naissance à 4 groupes de cancer du sein [18] [19] :

- Groupe HR+ : qui est subdivisé en deux groupes pronostiques distincts :
 - Luminal A = HER2-, Protéine Ki-67 faible, PR élevé.
 - Luminal B = HER2-, Protéine Ki-67 élevé, PR faible.
- Groupe HER2+ : caractérisé par l'expression des récepteurs HER2 à la surface cellulaire (Avec ou sans expression RH). Ce type de cancer est particulièrement agressif.
- Groupe TNCB: tumeur ER-, PR, HR2-, maladie très hétérogène, généralement ne répond pas à l'hormonothérapie, d'autres cibles thérapeutiques doivent être identifiées.
- Groupe Claudin low : décrite plus récemment, souvent triple-négative, mais qui se distingue par une faible expression des protéines de la jonction cellule-cellule, dont la E-cadhérine.

I.3 Les traitements

Plusieurs alternatives sont disponibles pour le traitement de cancer du sein, Néanmoins le type et le stade de cancer qui permettent d'orienter le traitement.

Pour le cancer du sein ER_α positif, il répond généralement à l'hormonothérapie. Donc la connaissance de la sensibilité du cancer du sein aux estrogènes est indispensable. La sensibilité aux estrogènes des cellules tumorales est médiée par les récepteurs d'estradiol (RE) et de progestérone (RP). La présence dans le tissu tumoral d'au moins un des deux récepteurs est nécessaire pour définir l'hormonosensibilité de la tumeur. L'hormonothérapie du cancer du sein est conçue pour supprimer les effets stimulants des estrogènes sur les cellules tumorales. Elle est basée sur la privation en estrogènes des cellules tumorales par suppression ovarienne (SO) ou inhibiteurs de l'aromatase et de sulfatase ou sur le blocage des RE [2] [1].

I.3.1 La privation en estrogènes des cellules tumorales

La privation en estrogènes des cellules tumorales comme déjà mentionné peut être faite soit par suppression ovarienne ou bien par inhibition de de l'aromatase et de sulfatase.

I.3.1.1 La suppression ovarienne

La suppression ovarienne peut être réalisée par différentes manières:

- chirurgie
- radiothérapie
- analogues de LH—RH

I.3.1.2 Les inhibiteurs de l'aromatase et de sulfatase

Le mécanisme d'action des inhibiteurs de l'aromatase se figure dans le blocage de la synthèse périphérique des estrogènes vu que l'aromatase assure la conversion de testostérone en estrone et en estradiol respectivement, néanmoins les inhibiteurs de l'aromatase ne peuvent être utilisés que chez les femmes ménopausées, Trois molécules sont actuellement disponibles : (ANA), Létrozole (LET) et Exémestane (EXE) Mais qui présentent des effets secondaires comme les douleurs ostéoarticulaires et le risque accru de fracture ostéoporotique [20] [2].

Le stéroïde sulfatase est aussi un enzyme responsable de l'augmentation du taux d'estrogène dans la circulation. Nombreuses études ont suggéré l'importance de composés à base de sulfamates dans l'inhibition de STS [1].

I.3.2 Le blocage des RE

Le blocage des récepteurs des estrogènes se fait généralement soit par des modulateurs sélectifs des récepteurs d'estrogènes (SERM) qui augmentent la liaison des estradiol au RE_{β} et non pas à ER_{α} ou bien par (SERD) régulateurs réducteurs sélectifs de récepteurs des œstrogènes qui dégradent les RE_{α} [1].

II. LES SULFAMATES

Les chimistes médicaux utilisent un nombre assez restreint de blocs de construction privilégiés utilisés de manière récurrente dans des molécules médicamenteuses puissantes.

L'une de ces composantes chimiques préférées est le groupe sulfonamide (substitué) et sa variante étroitement apparentée, le groupe sulfamate.

Les groupes sulfamates primaires (-OSO₂NH₂) sont des groupes de liaisons au zinc bien connus, ils figurent dans la structure de presque 26 médicaments américains approuvés par la FDA et d'un nombre important de médicaments expérimentaux, qui sont connus par leur action sur une série de cibles moléculaires dont la métalloenzyme zinc anhydrase carbonique, les récepteurs de la dopamine, les canaux ioniques [21].

II.1 Les dérivés de sulfamates

L'acide sulfamique (1) est la molécule mère à partir duquel quatre types principaux de dérivés peuvent être générés.

Les dérivés O-substitués(2), les composés N-substitués(3) et les dérivés O, N-di/tri substitués(4,5), et qui ont tous des applications pharmacologique [22].

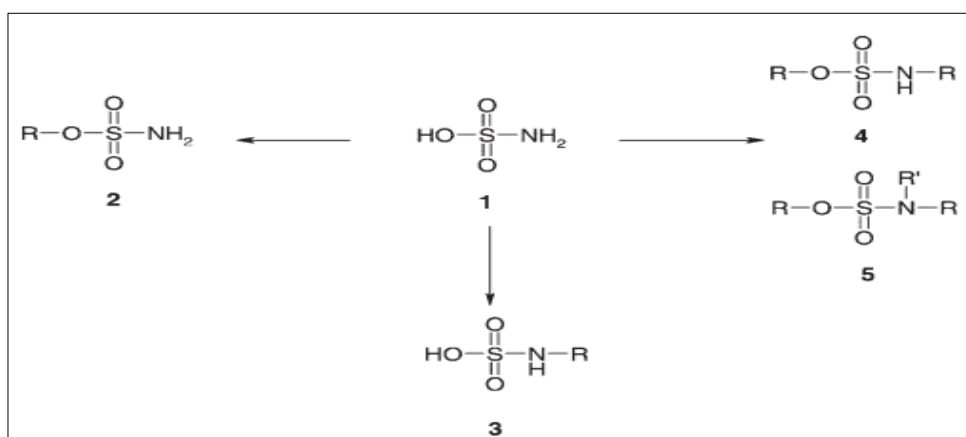


Figure 4 : les dérivés de sulfamates [22]

II.2 Les sulfamates naturels

Les produits naturels qui contiennent un groupe sulfamate primaire sont rares. D'après une recherche documentaire dans la base de données DNP, cinq produits naturels contenant le groupe sulfamates ont été trouvés :

- Les dérivés nucléosidiques de l'adénosine (la nucléocidine) : Découvert dans le microbe du sol streptomyce calvus.
- La 5' O sulfamoyl adénosine, 5'-O- sulfamoyl-2- chloroadénosine, 5'-O-sulfamoyl-2- bromoadénosine, et le 7-déaza-adénosine 5'-O-sulfamoyltubercidine. Ces quatre dérivés ont été isolés de divers actinomycètes terrestre [21].

II.3 Les applications pharmacologiques

Les sulfamides ont été les premiers médicaments largement et systématiquement utilisés comme agents prophylactiques et chimiothérapeutiques pour diverses maladies. Plus de 30 médicaments dotés de cette propriété sont en utilisation clinique, y compris les médicaments antifongique, anti-inflammatoire, antiprotozoaire, antibactérien, agent antihypertenseur, et les antagonistes des récepteurs de la vasopressine.. Ils sont également efficaces pour le traitement des infections urinaires, intestins et ophtalmiques, brûlures, rectocolite hémorragique, polyarthrite rhumatoïde, la dysfonction érectile comme le sildénafil inhibiteur de la phosphodiesterase. De plus récemment, les sulfamides sont

utilisés comme un agent anticancéreux, comme l'amprénavir inhibiteur de protéase antiviral VIH et dans la maladie d'Alzheimer [23].

II.3.1 Les sulfamates antiviraux

Des agents antiviraux incorporant des motifs sulfamate ont également été récemment signalés, avec au moins deux types de ces dérivés très étudiés : les dérivés nucléosidiques/non nucléosidiques de la transcriptase inverse du VIH et les IP du VIH.

La plupart des composés de ce type présentent une activité antivirale significative mais n'ont pas été développés pour un usage clinique [23].

II.3.2 Les sulfamates anti-cancéreux

Dans l'arsenal croissant des médicaments anticancéreux, les sulfamates occupent une position particulière, avec au moins deux cibles importantes mises en évidence ces dernières années : les STS et les CA (les isozymes associés aux tumeurs, CA IX et XII)[23].

Ainsi, un nombre impressionnant de STSI de type sulfamate ont été signalés au cours des dix dernières années par plusieurs groupes de recherche, tant dans le milieu universitaire que dans l'industrie. Plusieurs composés, tels que 667COUMATE (un sulfamate à base de coumarine tricyclique qui inhibe de manière irréversible l'activité du STS) entre autres, sont passés aux essais cliniques pour le traitement des tumeurs hormono-dépendantes (principalement les cancers du sein et de la prostate).

II.3.3 Les sulfamates Antiépileptique

Le topiramate est un antiépileptique de type monosaccharide sulfamate-substitué. Outre son indication dans l'épilepsie, il est prescrit dans le traitement prophylactique de la crise migraineuse. Son action anti-épileptique repose sur un blocage des canaux sodium voltage-dépendants, l'inhibition de l'activité excitatrice du glutamate et sur des effets de type GABAergique [24]

II.3.4 Le traitement de l'obésité

Un grand nombre de sulfamates anticonvulsivants ont été signalés, dont l'un de ces composés, le topiramate, qui est largement utilisé en clinique comme médicament antiépileptique, et comme traitement des formes réfractaires de la maladie, qui ne

répondent pas aux autres médicaments antiépileptiques. Compte tenu d'un effet secondaire de ce médicament, il a été récemment proposé d'utiliser le topiramate et les sulfamates apparentés pour le traitement de l'obésité. La justification de cette utilisation est basée sur l'inhibition de l'isozyme mitochondriale CA V, impliquée dans la lipogenèse [23].

II.3.5 Traitement de l'hyperlipidémie et de l'athérosclérose

Certains sulfamates possèdent une puissante activité inhibitrice contre l'ACAT, une enzyme impliquée dans le métabolisme du cholestérol. L'un de ces agents, l'avasimibe, fait l'objet d'essais cliniques avancés pour le traitement de l'hyperlipidémie et de l'athérosclérose [23].

II.3.6 Les sulfamates anti-arthrite et anti-ostéoporose

Les maladies arthritiques sont des maladies inflammatoires systémiques caractérisées par une atteinte considérable des structures du tissu conjonctif de l'articulation et des structures adjacentes, entraînant une destruction progressive du cartilage et des os. Les bisulfamates tels que l'AHR-15010 ont été signalés comme possédant une activité anti-arthritique chez les rats atteints d'arthrite adjuvante, agissant par un mécanisme d'action unique, représenté par l'inhibition de CA, impliqué dans la résorption osseuse [23].

III. GÉNÉRALITÉ SUR QSAR

III.1 Définition

Une approche QSAR, relation quantitative structure activité est un procédé qui relie de manière quantitative un ensemble de descripteurs calculés à partir de la structure moléculaire avec l'activité biologique, suivant un modèle mathématique : [Activité biologique] = f [Descripteurs]. Cette approche est décrite généralement comme une application des méthodes d'analyses des données et des statistiques pour développer des modèles qui pourraient prédire avec précision les activités biologiques en fonction de l'information de structure moléculaire d'une série de composés [25].

Cette méthode est l'un des principaux outils informatiques utilisés en chimie médicinale [26].

III.2 Principe

Le principe d'une approche QSAR est comme leur nom l'indique, repose sur l'établissement d'une relation mathématique qui relie l'activité biologique déterminée expérimentalement d'un ensemble de composés dans les mêmes conditions expérimentales, à des descripteurs moléculaires calculés à partir de la structure des molécules en se basant sur des méthodes statistiques. QSAR cherche à analyser des données structurales afin de détecter des facteurs influençant l'activité biologique.

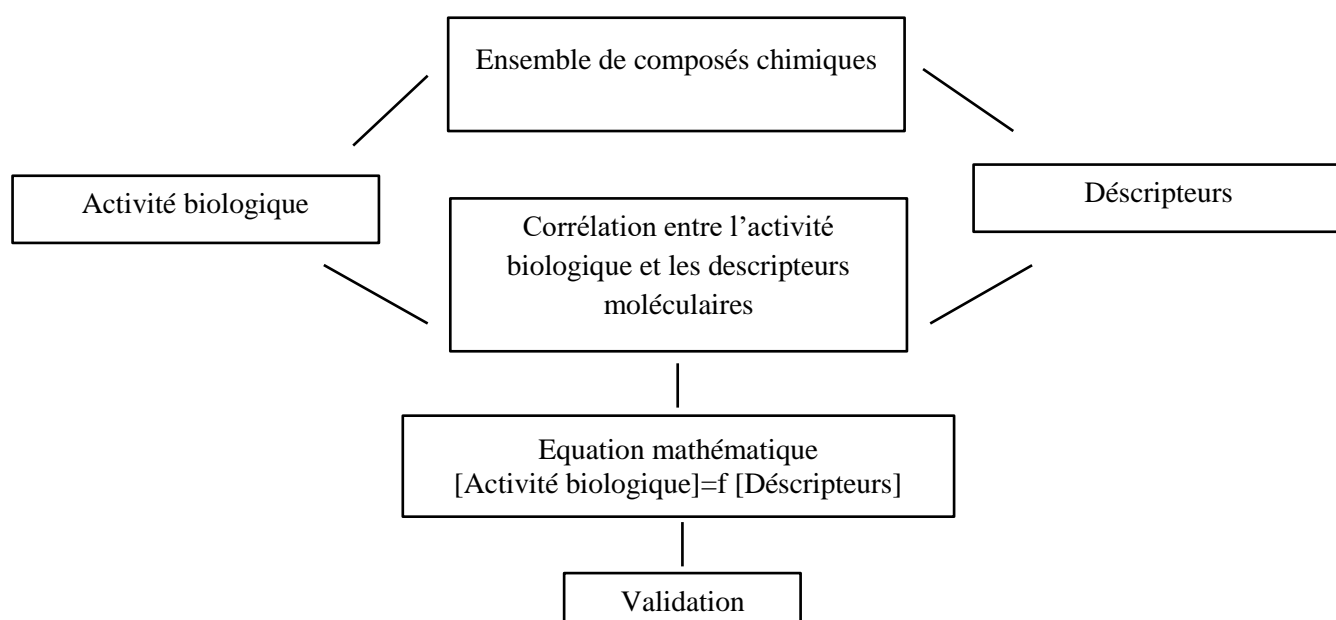


Figure 5 : Principe de l'approche QSAR

III.3 Présentation de la méthode QSAR

Généralement pour prédire l'activité biologique par une méthode QSAR il faut tout d'abord construire un modèle mathématique, ce qui exige le respect d'un certain nombre d'étape :

III.3.1 Le choix de base de données

Le choix de la base de données est une étape cruciale pour développer un modèle QSAR parce que ce dernier est très dépendant des données expérimentales de référence. Il faut que ces données soit fiables que possible et issues du même protocole expérimentale.

Plusieurs bases de données sont disponibles pour collecter les données à savoir : ChEMBL, Ligand Data Base, PubChem... [27]

III.3.2 La division de donnée

Les données collectées doit faire l'objet d'une division en deux types de données : Une donnée d'apprentissage « Training test » qui vont être utilisées pour la construction de modèle QSAR, et une donnée de prédiction ou bien de test « test set » qui va servir pour la validation de modèle formé. Il existe plusieurs approches possibles pour faire cette sélection. L'approche la plus utilisée est la division aléatoire qui consiste à diviser la base de données par simple sélection aléatoire [28].

III.3.3 Génération de Descripteurs

Les descripteurs sont définis comme des valeurs numériques caractérisant la structure moléculaire des molécules. Ils peuvent être calculés à partir de la formule brute des molécules (Descripteur 1D), ou bien à partir de la formule développée d'une molécule (Descripteur 2D), comme ils peuvent être estimés à partir de la position relatives des atomes dans l'espace. Plusieurs logiciels peuvent effectuées cette tâche : Codessa, Dragon, PaDEL...

III.3.4 Le développement du modèle

Afin d'élaborer un modèle il faut une méthode d'analyse des données qui va permettre de relier de manière quantitative l'activité étudié avec l'ensemble des descripteurs calculés. En effet, il existe plusieurs méthode statistique pour construire un modèle et analyser les données statistiques de ce dernier certaines sont linéaires telles que la régression linéaire multiple (MLR), la régression aux moindres carrés partiels (PLS), d'autres sont non linéaires comme, les réseaux de neurones artificiels (RNA)... [27]

III.3.5 Validation de modèle

Après l'élaboration du modèle tous les paramètres statistiques doivent être analysés et interprétés, ainsi il faut vérifier sa robustesse et sa qualité par le processus de validation. Donc cette validation va être divisée en deux: pour évaluer la robustesse de modèle on fait recours à la validation interne, et pour évaluer la capacité prédictive il faut faire une validation externe.

III.3.5.1 Validation interne

Pour ce type de validation, la méthode de validation croisée est la plus utilisée. Elle peut être réalisée suivant différents processus, celle qu'on va discuter c'est la technique (LOO) leave one out. Cette procédure retire à chaque fois une molécule de l'ensemble d'apprentissage et construit le modèle à partir du reste de molécules, et le modèle construit prédit l'activité de la molécule retirée. Ce processus est répété afin de prédire l'activité de toutes les molécules de l'ensemble d'apprentissage.

Pourtant, cette validation est insuffisante pour étudier la capacité prédictive d'un modèle. C'est pour cela la validation externe des modèles est devenue un élément standard et obligatoire de la modélisation (QSAR) [29].

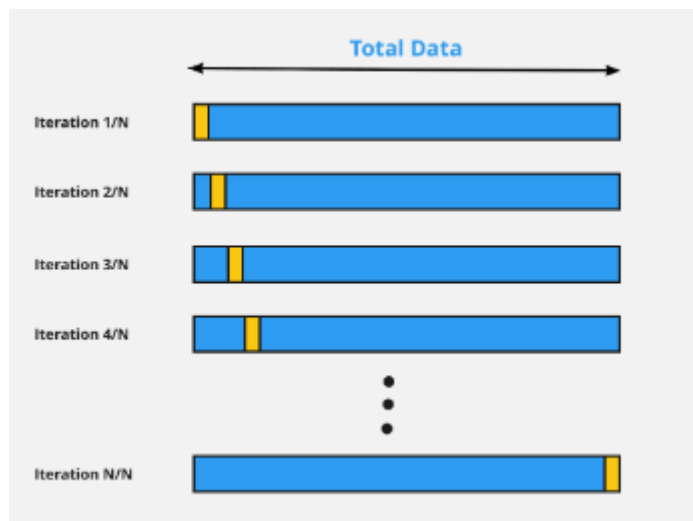


Figure 6 : Technique de validation LOO [30]

III.3.5.2 Validation externe

S'assurer de la robustesse de modèle n'est généralement pas suffisant pour dire que ce modèle est parfait et peut-être utilisé pour prédire l'activité biologique de nouvelles molécules. Alors la validation externe sert à généraliser la prévision en appliquons le modèle sur l'échantillon externe (test set) qui n'est pas utilisé pour développer le modèle [31].

III.3.6 Le domaine d'applicabilité

Le domaine d'applicabilité d'un modèle QSAR est l'espace physico-chimique, structurel ou biologique, les connaissances ou les informations sur lesquelles l'ensemble d'apprentissage du modèle a été développé, et pour lequel il est applicable pour faire des prédictions pour de nouveaux composés [32]. En effet L'absence de ce domaine admet que chaque modèle QSAR peut formellement prévoir l'activité d'un produit chimique même si sa structure est complètement différente de ceux inclus dans le sous ensemble.

MATÉRIEL ET MÉTHODES

I. La construction de modèle QSAR

I.1 Préparation des données

Les structures chimiques ont été sélectionnées à partir de PubChem ainsi que les valeurs de l'inhibition en nm.

Après la sélection des molécules utilisées pour cette étude, une optimisation de structures avec minimisation d'énergie a été effectuée à l'aide du logiciel Avogadro version 1.2.0. Les structures optimisées ont été utilisées comme entrées pour calculées les descripteurs.

L'inhibition de STS en nm est convertis à l'échelle logarithmique en prenant le logarithme négatif ($pIC_{50} = -\log_{10}IC_{50}$) pour qu'elle sera utilisé comme variable dépendante pour le développement de model QSAR.

L'ensemble de données est divisé en deux ensembles : 70% training set, qui va servir pour l'élaboration de modèle et 30% Test set, qui va être utilisé pour faire la validation externe de modèle. Cette sélection est faite en utilisant « Random Selection » par le logiciel DTC QSAR version 1.0.6.

I.1.1 Calcul de descripteurs

Les structures optimisées sont utilisées pour le calcul des 2D descripteurs par le logiciel PaDEL version 2.2.1. Un nombre important de descripteurs sont calculés, ce qui appelle la nécessité de faire un filtrage, l'option Data prétraitement disponible sur DTC QSAR va permettre de supprimer les valeurs constantes, et exclue les valeurs zéro pour éviter la corrélation entre les descripteurs, > 0.70 .

I.1.2 L'équation du modèle QSAR

A l'aide du logiciel DTC QSAR Tools, l'option « best set selection » nous a permis de choisir la combinaison de descripteur qui va donner des bons résultats en ce qui concerne la qualité de prédiction, Donc nous avons analysé des combinaisons de deux descripteurs et de trois descripteurs.

Les combinaisons choisies sont utilisées pour élaborer l'équation de modèle QSAR par : régression linéaire multiple (MLR) qui est utile pour la génération de modèles QSAR hautement prédictifs et interprétables.

I.1.3 Validation de modèle

La robustesse du modèle QSAR est validée par la technique de validation croisée pour évaluer la robustesse de modèle et pour s'assurer de son capacité prédictive.

Dans cette étude toutes les mesures de validation interne et externe requises sont calculées en utilisant respectivement les ensembles de formation (training set) et de test (test set).

I.2 Le domaine d'applicabilité

Nous avons étudié l'AD en utilisant l'approche de standardisation afin de déterminer les composés de l'ensemble de test qui ne relèvent pas du domaine d'applicabilité (outside) et de détecter les composés de l'ensemble d'apprentissage qui sont aberrants (outlier). Le logiciel adopté pour cette analyse s'appelle « AD using standardization approach ».

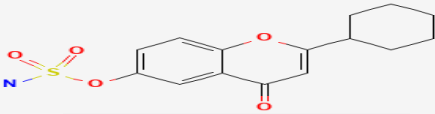
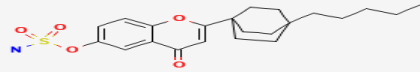
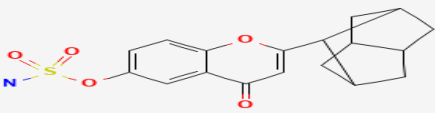
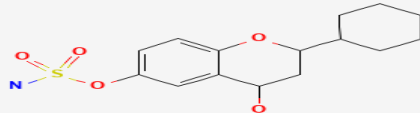
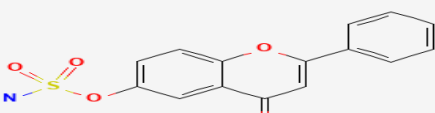
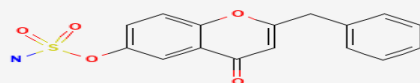
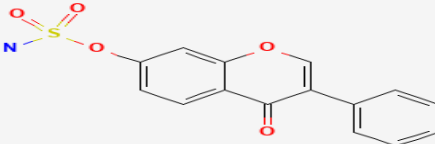
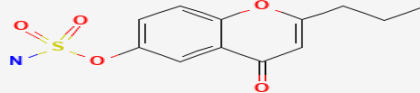
I.3 Etude des propriétés physico-chimiques des paramètres de Lipinski et de Veber

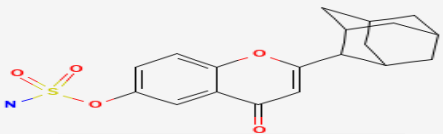
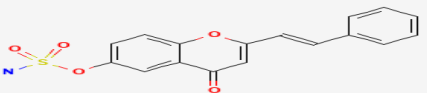
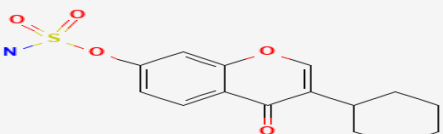
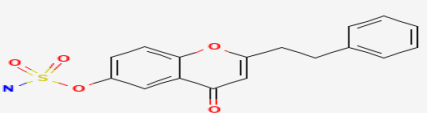
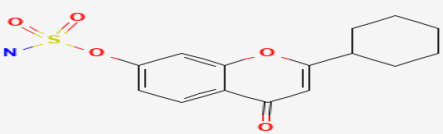
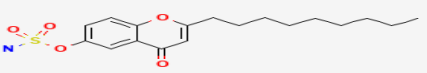
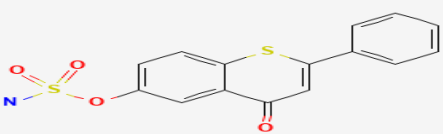
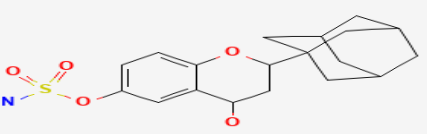
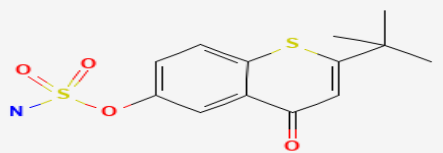
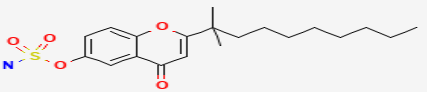
Les règles de Lipinski et de Veber ont été appliquées afin d'identifier « Drug like » parmi les 29 inhibiteurs, pour calculer les descripteurs suivantes : NHA – NHD – MW – LogP -PAS-RBC nous avons utilisé le logiciel : Marvin Sketch.

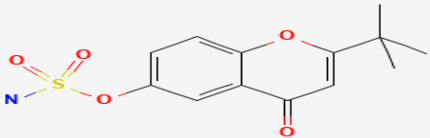
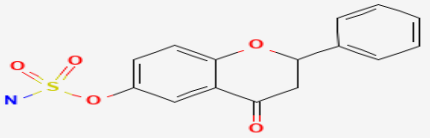
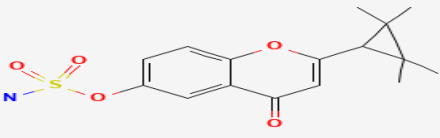
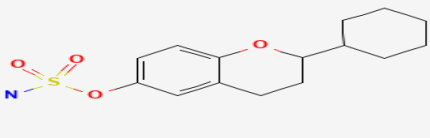
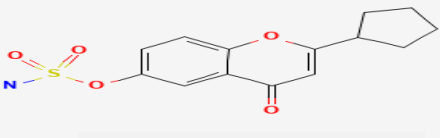
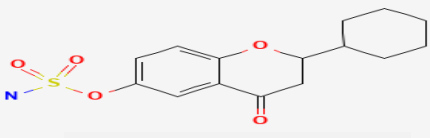
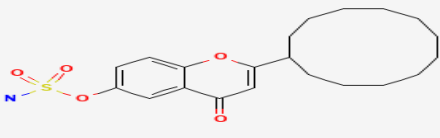
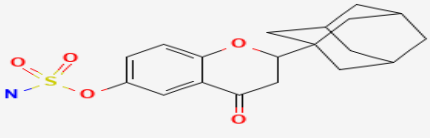
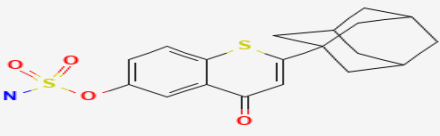
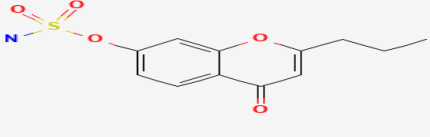
RÉSULTATS ET DISCUSSION

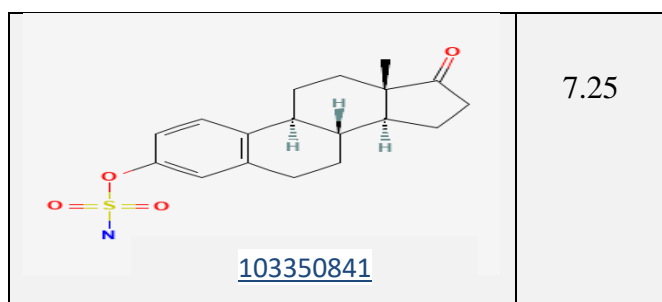
Une ensemble de données de dérivées de 4thio chroménone est collecté et analysé pour déterminer la relation entre les structures chimiques représentées par les descripteurs et l'activité biologique, les structures chimiques de 29 dérivées et leur pIC50 déterminer expérimentalement sont illustrés dans le tableau 1.

Tableau 1: Structure chimique et activité biologique expérimentale des dérivés de 4-(thio)chroménone 6-O-sulfamate 2- substitué

Structure	pIC50	structure	pIC50
 103373446	7.21	 103373441	7.96
 103373977	7.96	 103373390	4.61
 103373976	6.17	 103373388	5.79
 103373550	6.36	 103373175	6.14

 <p>103373513</p>	6.29	 <p>103373134</p>	5.79
 <p>103373512</p>	7.3	 <p>103373108</p>	6.72
 <p>103373510</p>	5.20	 <p>103373100</p>	6.39
 <p>103373509</p>	7.21	 <p>103373058</p>	6.58
 <p>103373504</p>	8.25	 <p>103373057</p>	7.11

 <p>103373485</p>	7.66	 <p>103373034</p>	5.59
 <p>103373484</p>	7.35	 <p>103372957</p>	4.95
 <p>103373483</p>	6.87	 <p>103372906</p>	6.55
 <p>103373480</p>	7.35	 <p>103372905</p>	6.85
 <p>103373465</p>	9.47	 <p>103373511</p>	4



Parmi les 29 composés, huit (103350841-103372905-103372957-103373388-103373057-103373504-103373510-103373550) qui sont choisis aléatoirement pour tester le modèle (test set) (30%), et le reste (21composés) sont gardé pour le test d'entraînement (training set) (70%).

Le logiciel PaDEL a calculé un groupe de 1343 de descripteurs 2D (physicochimique, stérique, électronique, topologique, etc.).

Le prétraitement de données a permis d'éliminer 426 descripteurs avec valeurs constantes ou égale à 0, et 254 descripteurs intercorrélés.

Dans l'étape de sélection de meilleure combinaison de descripteurs, une analyse de 219453 combinaisons de 2 descripteurs a été réalisée, et toutes les combinaisons générées ont donné une qualité de prédiction soit modérée soit mauvaise. Alors que l'analyse de 2147483647 combinaisons de 3 descripteurs a élaboré 15 combinaisons avec bonne qualité de prédiction.

La régression multilinéaire a permet d'élaborer les modèles QSAR à partir des 15 combinaisons, le tableau ci-dessous montre les résultats obtenues.

Tableau 2: les modèles QSAR obtenus à partir de 15 combinaisons

Nb	Equation MLR	Internal Validation metrics								External Validation metrics					
		R ²	R ² a	SEE	Q ² _{LOO}	SPED (LOO)	Avg. rm ² (LOO)	Δrm ² (LOO)	MAE	Q ² (F1)Test	Q ² (F2)Test	Avg. rm ² (test)	Δrm ² (test)	CCC (test)	MAE, (Test)
1	PIC50 = -26,1865 -1,2716 maxssO -0,1934 AATSC6v +8,2311 MWC2														
2	PIC50 = -7,0318 -0,1997 AATSC6v +0,1283 TWC -0,2533 SssO	0.82	0.79	0.54	0.73	0.60	0.65	0.077	0.48	0.76	0.74	0.71	0.11	0.88	0.35
3	PIC50 = 6,9325 -0,1933 AATSC6v -1,2817 maxssO +0,8526 SP-3	0.79	0.75	0.59	0.71	0.62	0.61	0.14	0.50	0.65	0.63	0.63	0.04	0.83	0.53
4	PIC50 = 22,2899 +0,0765 ATSC2i -3,5523 GATS8p -15,9647 GATS1s	0.80	0.77	0.57	0.73	0.60	0.64	0.12	0.45	0.77	0.76	0.71	0.15	0.88	0.34
5	PIC50 = -23,9697 -0,0023 ATSC6v +4,9383 MWC4 -0,2173 SssO	0.77	0.73	0.62	0.70	0.63	0.59	0.15	0.43	0.63	0.61	0.53	0.14	0.79	0.53
6	PIC50 = 44,584 -0,257 AATS3i +8,1492 GATS6v +27,0155 MATS1s	0.84	0.81	0.52	0.73	0.60	0.64	0.10	0.41	0.69	0.67	0.61	0.14	0.83	0.43
7	PIC50 = 3,023 -0,2413 SssO -0,1952AATSC6v +0,7208 SP-3	0.78	0.74	0.60	0.71	0.62	0.61	0.15	0.46	0.68	0.67	0.64	0.13	0.84	0.45
8	PIC50 = -0,796 -3,3957 AATSC3i +1,3062 MAXDP +1,3734 SdssC	0.76	0.72	0.63	0.63	0.70	0.51	0.20	0.51	0.67	0.66	0.57	0.21	0.82	0.49
9	PIC50 = 26,3519 -3,7003 GATS8p -10,7364 GATS2p +23,4981 MATS1s	0.89	0.87	0.42	0.83	0.48	0.77	0.009	0.32	0.82	0.81	0.83	0.008	0.92	0.37
10	PIC50 = 25,2715 -9,042GATS2i -3,7251 GATS8p +27,4367MATS1s	0.87	0.85	0.45	0.79	0.52	0.73	0.006	0.35	0.83	0.82	0.82	0.04	0.92	0.36
11	PIC50 = 8,7327+1,4332MAXDP -6,8511SIC5 -6,6343GATS2c	0.74	0.70	0.65	0.64	0.69	0.52	0.19	0.51	0.69	0.67	0.67	0.17	0.84	0.47

12	PIC50 = 8,0914 +4,3825 GATS3v -8,0743GATS1s +1,3179maxdssC	0.7 3	0.68	0.67	0.57	0.76	0.44	0.16	0.54	0.62	0.60	0.45	0.28	0.77	0.51
13	PIC50 = -4,7795+3,911 GATS3i +1,3695MAXDP +1,3013 SdssC	0.7 6	0.71	0.63	0.63	0.70	0.50	0.20	0.52	0.70	0.68	0.63	0.18	0.84	0.45
14	PIC50 = 12,1843 +1,0117 maxdssC -8,7358GATS1s +0,9308 VC-3	0.7 0	0.65	0.70	0.56	0.76	0.43	0.18	0.54	0.64	0.62	0.51	0.24	0.79	0.50
15	PIC50 = 2,4715 +0,6777 VP-2 +0,9467 ndssC - 0,0071SpAbs_Dzp	0.6 7	0.61	0.74	0.54	0.78	0.39	0.24	0.53	0.66	0.64	0.50	0.25	0.79	0.42

D'après les modèles obtenus nous avons remarqué que le modèle numéro 9 a la meilleure valeur de R^2 (0.89), avec une valeur de SEE (standard error of the estimate) (0.42) la plus faible. Cependant parmi ces 15 modèles choisissant le modèle numéro 9.

PIC50 = 26,3519 (+/-1,7493) -3,7003(+/-0,5982) GATS8p -10,7364(+/-1,2478) GATS2p +23,4981(+/-2,6116) MATS1s

N=29 $R^2 = 0.89$ $R^2_{adj} = 0.87$ S = 0.42 F=46.14

Le R^2 c'est le coefficient de corrélation au carré ou bien le coefficient de détermination. Il représente la proportion de la variance de y (pIC50) qui a été expliquée par les variables indépendants (les descripteurs) dans le modèle. Plus sa valeur est proche de 1 plus le modèle est adéquat, la valeur élevé de R^2 indique que la ligne de régression correspond parfaitement aux données.

Dans notre cas $R^2=0.89$ ce qui indique fermement la corrélation entre les descripteurs utilisés et l'activité biologique étudiée (l'inhibition de STS).

La limitation de R^2 c'est que sa valeur augmente avec l'addition de variables indépendants au modèle chose qui est trompeuse, par ce que certains variables ajoutées peuvent être inutile, c'est-à-dire qu'ils ont une signification minimales. Le R^2_{adj} surmonte ce problème par ce que si le variable ajouté au modèle est unitile le R^2_{adj} va diminuer et

dans le cas contraire le R^2_{adj} va augmenter. La valeur de R^2_{adj} dans notre équation égale à 0.87 ce qui indique une très bonne concordance entre la corrélation et la variation des données.

La valeur de $F = 46.14$ indique fortement que la relation de régression étaient statistiquement significative et ne s'est produite par hasard.

La matrice de corrélation pour l'activité biologique pIC50 et les descripteurs sélectionnés pour construire le modèle QSAR est présentée dans le tableau 3, Pour montrer l'absence de corrélation entre les descripteurs sélectionnés.

Tableau 3: Matrice de corrélation entre les descripteurs sélectionnés

	MATS1s	GATS2p	GATS8p
MATS1s	1		
GATS2p	,298	1	
GATS8p	-,013	-,189	1

Le tableau 4 montre les facteurs d'inflation de la variance (VIF) des descripteurs dans le modèle QSAR. Les valeurs de VIF des trois descripteurs sont toutes inférieures à 5 ce qui suggère l'absence de colinéarité entre les descripteurs sélectionnés et que le modèle qui en résulte a une bonne stabilité.

Tableau 4: Les facteurs d'inflation de la variance (VIF)

Term	Coef	VIF
Constant	26,35	
MATS1s	23,5	1,1
GATS2p	-10,74	1,14
GATS8p	-3,69	1,04

I. Validation de Model choisis

I.1 Validation interne

Les pouvoirs prédictifs de l'équation sont validés par la méthode de validation croisée (cross validation). Il s'agit d'une méthode pratique et fiable pour tester la signification d'un modèle.

Les paramètres : Q^2_{LOO} , SEE, $SPED_{(LOO)}$, $Avg. rm^2_{(LOO)}$, $\Delta rm^2_{(LOO)}$, MAE ont été calculé par cette méthode de validation.

Tableau 5: Résultats de validation interne

Q^2_{LOO}	SEE	$SPED_{(LOO)}$	$Avg. rm^2_{(LOO)}$	$\Delta rm^2_{(LOO)}$	MAE
0.83	0.42	0.48	0.77	0.009	0.32

Q^2_{LOO} est le coefficient de corrélation de la validation croisée (leave one out), décrit la stabilité de modèle de régression obtenu en se concentrant sur la sensibilité du modèle à l'élimination d'un point de données. Notre $Q^2_{LOO} = 0.83$, cette valeur illustre bien la stabilité de modèle.

SEE (l'erreur standard de l'estimation) est une mesure de la précision de prédiction. Il représente la distance moyenne entre les valeurs prédites et la ligne de régression, d'une façon pratique, il nous indique à quel point le mode de régression est faux. Ça valeur est faible = 0.42 ce qui montre que les valeurs prédites sont plus proches de la ligne de régression.

MAE (Mean Absolute Error), l'Erreur moyenne absolue est obtenu par le calcul de la différence absolue moyenne entre la prédiction de modèle et les valeurs expérimentales. Plus sa valeur est proche de 0 plus le modèle est parfait. Pour notre modèle $MAE = 0.32$, ceci confirme l'efficacité de modèle.

$SPED_{(LOO)}$ (Standard Deviation of the Prediction Errors) représente l'écart-type des erreurs de prédiction L'objectif principal de ce paramètre est d'obtenir un modèle

avec la plus grande capacité de prédiction, c'est-à-dire avec la valeur numérique la plus basse de SDEP. Sa valeur dans notre modèle été la plus basse par rapport aux autres modèles (0.48).

Les deux paramètres **Avg. $\text{rm}^2_{(\text{LOO})}$** et **$\Delta\text{rm}^2_{(\text{LOO})}$** sont aussi important pour évaluer la qualité de la prédiction de modèles QSAR. Ses valeurs doivent être > 0.5 et < 0.2 respectivement. Pour notre cas **Avg. $\text{rm}^2_{(\text{LOO})} = 0.77$** et **$\Delta\text{rm}^2_{(\text{LOO})} = 0.009$** ce qui prouve la capacité prédictive de modèle.

➤ D'une façon générale tous les paramètres calculés pour la validation interne ont démontré la robustesse de modèle. Mais cette validation seule n'est pas suffisante, il faut absolument une validation externe pour évaluer la capacité prédictive de nouveaux composées qui n'ont été pas utilisé pour générer le modèle (le test set).

I.2 La validation externe

La validation externe est comme déjà mentionnée s'effectue par l'application de modèle obtenu par le jeu d'entraînement (training test) sur l'ensemble de données de jeu d'essai (test set), et la mesure des performances à l'aide de différents critères tels que : $Q^2(\text{F1})_{\text{Test}}$, $Q^2(\text{F2})_{\text{Test}}$, **Avg. $\text{Rm}^2_{(\text{test})}$** , **$\Delta\text{rm}^2_{(\text{test})}$** , **$\text{CCC}_{(\text{test})}$** , **$\text{MAE}_{(\text{Test})}$** les résultats obtenus sont illustrés dans le tableau 6.

Tableau 6: Résultats de la validation externe

$Q^2(\text{F1})_{\text{Test}}$	$Q^2(\text{F2})_{\text{Test}}$	Avg. $\text{Rm}^2_{(\text{test})}$	$\Delta\text{rm}^2_{(\text{test})}$	$\text{CCC}_{(\text{test})}$	$\text{MAE}_{(\text{Test})}$
0.82	0.81	0.83	0.008	0.92	0.37

Du premier aperçu, nous observons, que les valeurs obtenues (**Avg. $\text{Rm}^2_{(\text{test})} = 0.83$** , **$\Delta\text{rm}^2_{(\text{test})} = 0.008$** , **$\text{MAE}_{(\text{Test})} = 0.37$**) dans la validation externe sont très proches de ceux de la validation interne.

Pour le $Q^2(\text{F1})_{\text{Test}}$ externe pour l'ensemble d'essai est déterminé par l'équation suivante :

$$Q^2(F1) = 1 - \frac{PRESS}{SS_{ext} \bar{y}_{tr}}$$

Où \bar{y}_{tr} représente la moyenne des réponses de l'ensemble d'entraînement, $PRESS$ est la somme des carrés prédictifs, $SS_{ext} \bar{y}_{tr}$ est la somme totale des carrés de l'ensemble externe calculée par la moyenne de l'ensemble d'apprentissage.

Alors que la valeur de $Q^2(F2)_{Test}$ est calculée par l'équation suivante :

$$Q^2(F1) = 1 - \frac{PRESS}{SS_{ext} \bar{y}_{test}}$$

Le $Q^2(F2)_{Test}$ diffère de $Q^2(F1)_{Test}$ par l'utilisation du moyenne de l'ensemble d'essai au lieu de l'ensemble l'apprentissage

La moyenne de l'ensemble d'apprentissage est utilisée pour se disposer d'une valeur de référence

Donc on va comparer le $Q^2(F2)_{Test}$ avec $Q^2(F1)_{Test}$, et si leurs valeurs sont proches et élevées ça va montrer la capacité prédictive du modèle.

Le CCC est le coefficient de corrélation de concordance. Il est utilisé pour mesurer la concordance entre les données expérimentales et prédites. En effet le CCC mesure à la fois la précision (à quel point les observations sont éloignées de la ligne de régression) et l'exactitude (combien la ligne de régression s'écarte de la ligne de concordance), et par conséquent toute divergence de la ligne de régression par rapport à la ligne de concordance donne comme conséquence une valeur de CCC inférieure à 1.

D'après les résultats de la validation externe nous remarquons que les valeurs (**Avg. $Rm^2_{(test)} = 0.83$, $\Delta rm^2_{(test)} = 0.008$, $MAE_{(Test)} = 0.37$) dans la validation externe sont très proches de celles de la validation interne, et la valeur de $Q^2(F1)_{Test} (0.82) \approx Q^2(F2)_{Test} (0.83)$ et la $CCC = 0.92$ s'approche beaucoup de 1 donc la ligne de régression ne s'écarte pas beaucoup de la ligne de concordance. Ceci montre la capacité prédictive du model sur l'ensemble d'essai.**

→ Les résultats de validation interne et de validation externe montre que ce modèle peut être appliqué avec succès pour prédire l'activité étudié, et la figure 7 confirme ceci vu que les valeurs prédites sont très proches des valeurs expérimentales soit

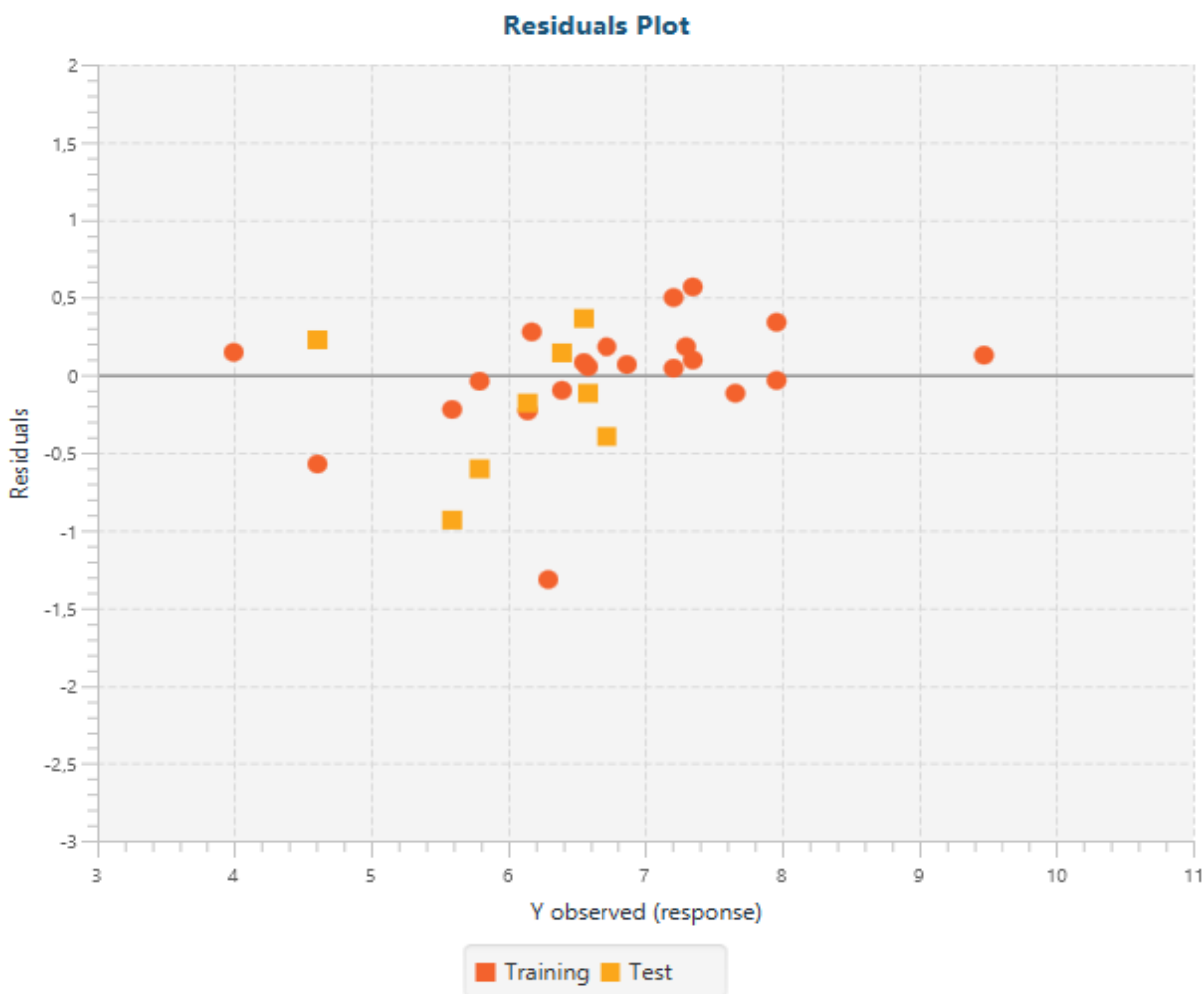


Figure 8: Représentations graphique des valeurs résiduelles en fonction des valeurs

II. Le domaine d'application

L'AD ou bien le domaine d'application c'est la zone théoriques de l'espace chimique, définies par les descripteurs de modèles et les réponses modélisées, et donc par les propriétés chimiques de l'ensemble d'apprentissage, représentées dans chaque modèle par des descripteurs moléculaires spécifiques. Étant donné que même un QSAR robuste, significatif et validé ne peut pas prédire de manière fiable les propriétés modélisées pour l'univers des produits chimiques, son domaine d'application doit être défini et les prévisions ne sont considérées comme fiables que pour les produits chimiques appartenant à ce domaine.

D'après l'analyse du tableau 7, nous observons que le composé (103373511) dans l'ensemble d'apprentissage est considéré comme aberrant, donc il est dissimilaire à la majorité des composés de l'ensemble d'apprentissage.

Le tableau 8 indique que le composé(103373504) est hors domaine d'applicabilité du modèle.

Tableau 7: Statut du domaine d'applicabilité du de l'ensemble d'apprentissage

Name	GATS8p	GATS2p	MATS1s	PIC50	Predicted Response	Status(AD)
103372906	0.94	1.09	-0.2	6.55	6.471242100950132	In
103373034	0.94	1.02	-0.26	5.59	5.812903953923634	In
103373058	0.85	1.05	-0.23	6.58	6.528785843582778	In
103373100	0.79	1.14	-0.2	6.39	6.489473792960702	In
103373108	0.72	1.05	-0.25	6.72	6.539868480567609	In
103373134	0.75	1.04	-0.28	5.79	5.8312792677126914	In
103373175	0.76	1.03	-0.26	6.14	6.371601699964248	In
103373390	0.82	1.12	-0.26	4.61	5.183305648778564	In
103373441	0.8	1.04	-0.18	7.96	7.996071713342541	In
103373446	0.84	1.08	-0.21	7.21	6.713659357683788	In
103373465	0.9	0.88	-0.18	9.47	9.343860531592934	In
103373480	0.75	1.17	-0.18	7.35	6.7853575885894095	In
103373483	0.81	1.06	-0.22	6.87	6.804416567311726	In
103373484	1.02	0.88	-0.25	7.35	7.254952698425587	In
103373485	0.63	0.9	-0.28	7.66	7.778415597581713	In
103373509	0.98	0.88	-0.26	7.21	7.1679854489782535	In
103373511	1.44	1.03	-0.26	4.0	3.855368652003037	Outlier
103373512	0.91	1.04	-0.2	7.3	7.119072098105768	In
103373513	0.94	1.05	-0.17	6.29	7.605640747905287	In
103373976	0.82	1.01	-0.28	6.17	5.894347106858152	In
103373977	0.93	1.03	-0.18	7.96	7.622391103181425	In

Tableau 8: Statut du domaine d'applicabilité du de l'ensemble d'essai

Name	GATS8p	GATS2p	MATS1s	PIC50	Predicted Response	Status(AD)
103350841	1.21	0.98	-0.19	7.25	6.888133899733221	Inside-AD
103372905	0.95	1.03	-0.17	6.85	7.783365206980561	Inside-AD
103372957	0.88	1.11	-0.26	4.95	5.068649028848751	Inside-AD
103373057	0.8	1.07	-0.21	7.11	6.969037009512974	Inside-AD
103373388	0.81	1.03	-0.26	5.79	6.186584564084748	Inside-AD
103373504	0.78	0.77	-0.27	8.25	8.85407640872678	Outside-AD
103373510	1.2	1.08	-0.21	5.2	5.38153597935138	Inside-AD
103373550	0.9	0.96	-0.28	6.36	6.135139405078888	Inside-AD

III. Contribution du descripteurs

La figure 9 montre que les 3 descripteurs (tableau 9) sont indispensables pour la construction de notre modèle. Chacun des trois contribue soit positivement soit négativement sur la pIC50.

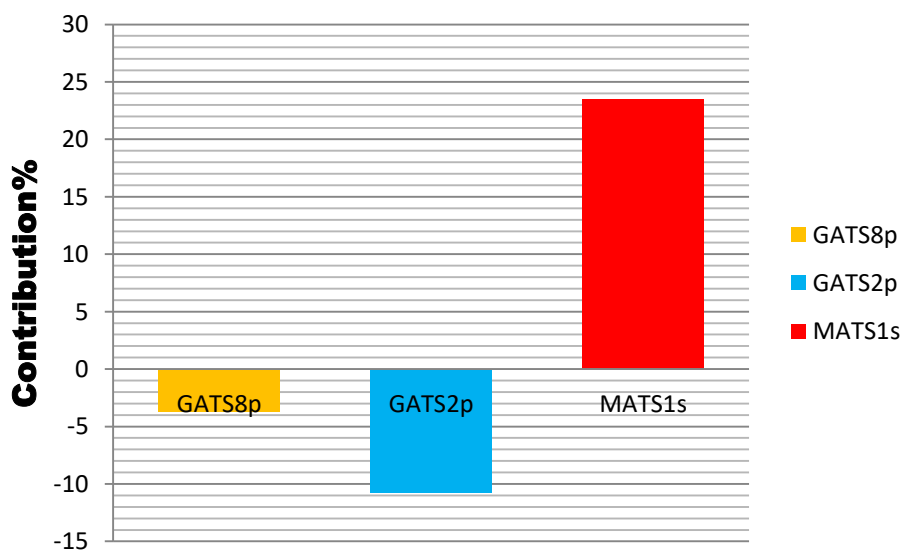


Figure 9: La contribution des descripteurs choisis dans le modèle

Tableau 9: Signification des descripteurs choisis

Le descripteur	Signification
GATS8p	Geary autocorrelation-lag8/weighted by atomic polarizabilites
GATS2p	Geary autocorrelation-lag2/weighted by polarizabilities
MATS1s	Moran autocorrelation of lag 1 weighted by I-state

III.1 Interprétation des descripteurs sélectionnés

- GATS2p est un descripteur d'autocorrélation. En effet les descripteurs d'autocorrélation des composés chimique peuvent être représentés au niveau atomique ou au niveau de surface moléculaire ou autre, et qui reflète les caractéristiques structurelles de la topologie moléculaire et les propriétés atomiques. Le signe négatif de GATS2p a indiqué qu'une augmentation de ce descripteur entraîne une diminution de l'activité inhibitrice.
- GATS8p indique que la présence d'atomes polarisables à une distance topologique égale à 8 a une influence sur l'activité inhibitrice. Il est important de noter que cette contribution est toujours négative car la relation mathématique n'implique que des valeurs négatives.
- MATS1s est aussi un descripteur qui appartient au bloc d'autocorrélations 2D.

IV. Les règles de 5 de Lipinski et de Veber

La règle de 5 de Lipinski est une méthode utilisé afin d'identifier les composés ayant une forte absorption ou perméabilité de ceux qu'ayant une faible absorption. Donc pour qu'un composé soit comme un médicament sa masse moléculaire ne doit pas dépasser 500 DA.

Dans le cas contraire sa diffusion à travers les membranes biologiques sera difficile est va être mal absorbé. Ainsi il doit y avoir 5 liaisons hydrogènes ou moins par ce que trop de ponts d'hydrogène rendent la molécule trop polaire, ce qui empêche sa diffusion dans les parties lipophiles et aussi les accepteurs de liaisons hydrogènes doivent être 10 ou moins, et la valeur de log P doit être inférieur à 5.

En plus de règle de Lipinski, Veber a introduit deux autres critères pour identifier les composés qui n'ont pas de problèmes avec la biodisponibilité orale, il s'agit du nombre de liaisons rotatives < 10 et de la surface polaire $< 140 \text{ \AA}$.

Les figures 10, 11 et 12 montrent les résultats de paramètres calculés de Lipinski et Veber

La figure 10 indique que Les valeurs de toutes les molécules sont en concordance avec la règle de Lipinski, les nombres de liaisons accepteurs d'hydrogène sont inférieur à 10, les nombre de liaison donneurs d'hydrogène sont inférieur à 5 et le logarithme p et inférieur à 5.

La figure 11 montre que le poids moléculaire de tous les composés est inférieur à 500 DA.

Dans la figure 12, nous observons que le nombre des liaisons rotatives est inférieur à 10 pour tous les composés et la surface polaire est inférieure à 140 Å.

D'après tous ces résultats, aucun composé n'a violé la règle des 5 de Lipinski donc ces inhibiteurs ont un profile pharmaceutique et peuvent être considérés comme des candidats médicaments potentiels pour le cancer de sein.

La même chose pour veber tous les composés de notre série n'ont pas transgressé la règle, ce qui témoigne une bonne concordance avec le transport moléculaire passif à travers les membranes.

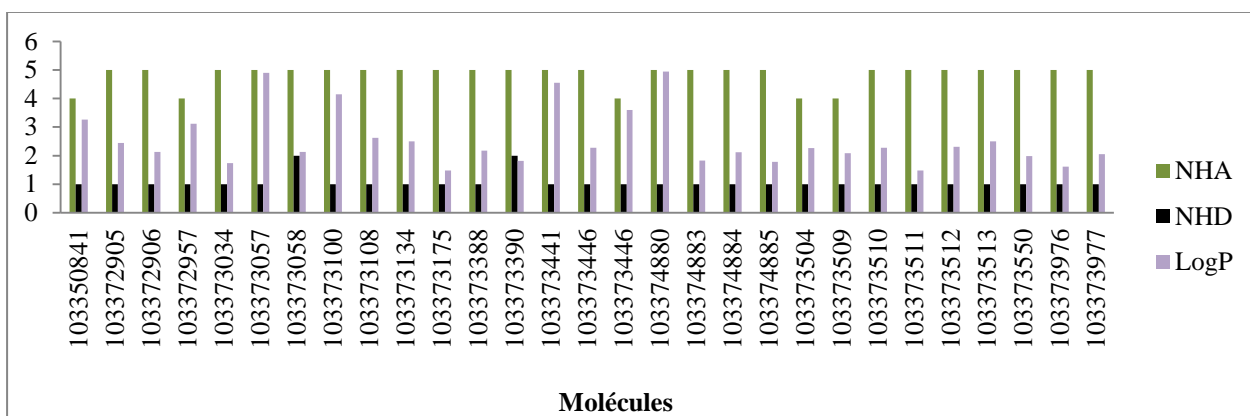


Figure 10: les valeurs de paramètres de Lipinski (NHA, NHD, LogP) des 29 composés

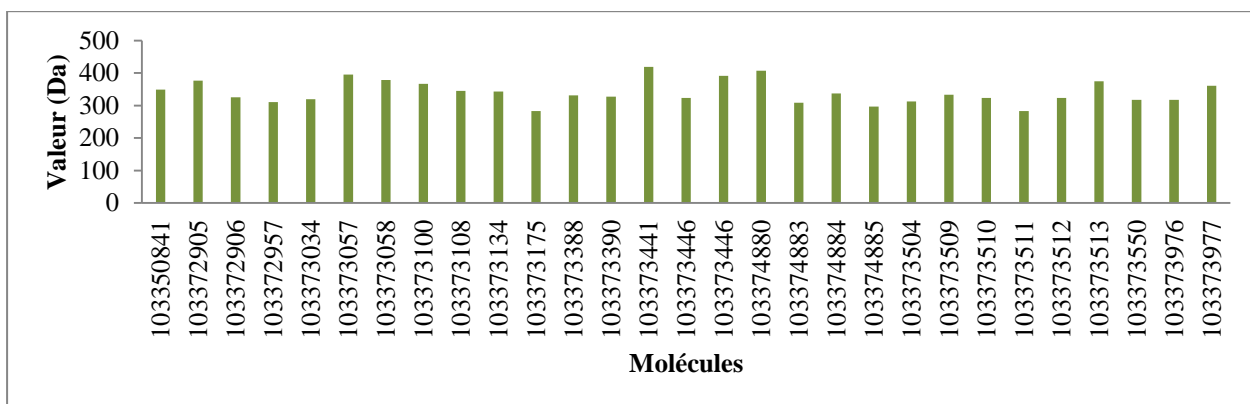


Figure 11: Les valeurs de poids moléculaire des 29 composés

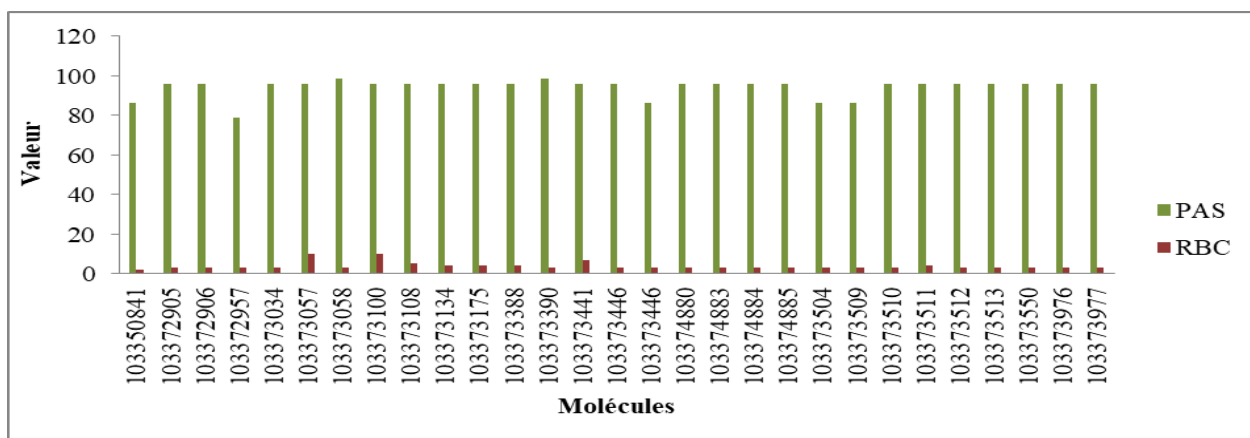


Figure 12: Les paramètres de Veber des 29 composés

CONCLUSION ET PERSPECTIVE

Conclusion et Perspective

Durant ce travail nous avons effectué une étude QSAR sur une série de 29 molécules de 4-(thio) chroménone 6-O-sulfamate 2- substitué par l'utilisation de la méthode statistique d'analyse RLM. Le but de cette étude c'est l'établissement d'une relation linéaire quantitative structure-activité.

Cette étude a montré que l'activité inhibitrice de STS est influencée par 3 descripteurs d'autocorrélation 2D (GATS2p, GATS8p, MATS1s), GATS2p, GATS8p contribuent négativement sur l'activité biologique alors que le descripteur MATS1s contribuent positivement, donc ce modèle donne un aperçu de la conception de nouveaux composés par modification structurelle.

La robustesse de modèle obtenu a été évaluée par la méthode de validation croisée (Leave one out), une forte corrélation a été observée ($Q^2 = 0.83$) entre les valeurs expérimentales et les valeurs prédites de activité biologique, Ainsi le pouvoir prédictif est confirmé par la méthode de validation externe par l'utilisation de données d'essai (Data set) et a montré une meilleure capacité prédictive.

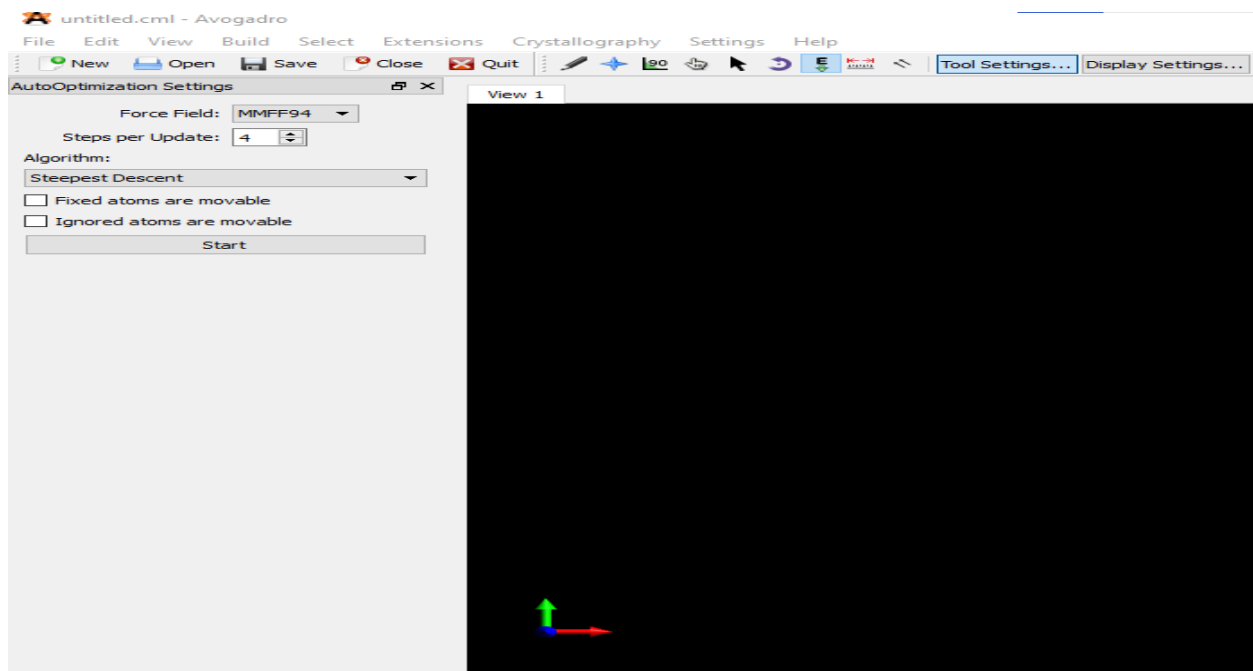
Les résultats obtenus de validation nous a permis de dire que ce modèle peut être utilisé pour prédire l'activité biologique de molécules non testés mais à condition qu'ils soient dans le domaine d'applicabilité de modèle.

La comparaison des résultats de calcul et des critères de la règle de Lipinski et de Veber a induit que les 29 molécules peuvent être considérées comme candidats médicaments vu qu'ils sont en concordance avec la règle de 5 de Lipinski et aussi la règle de Veber.

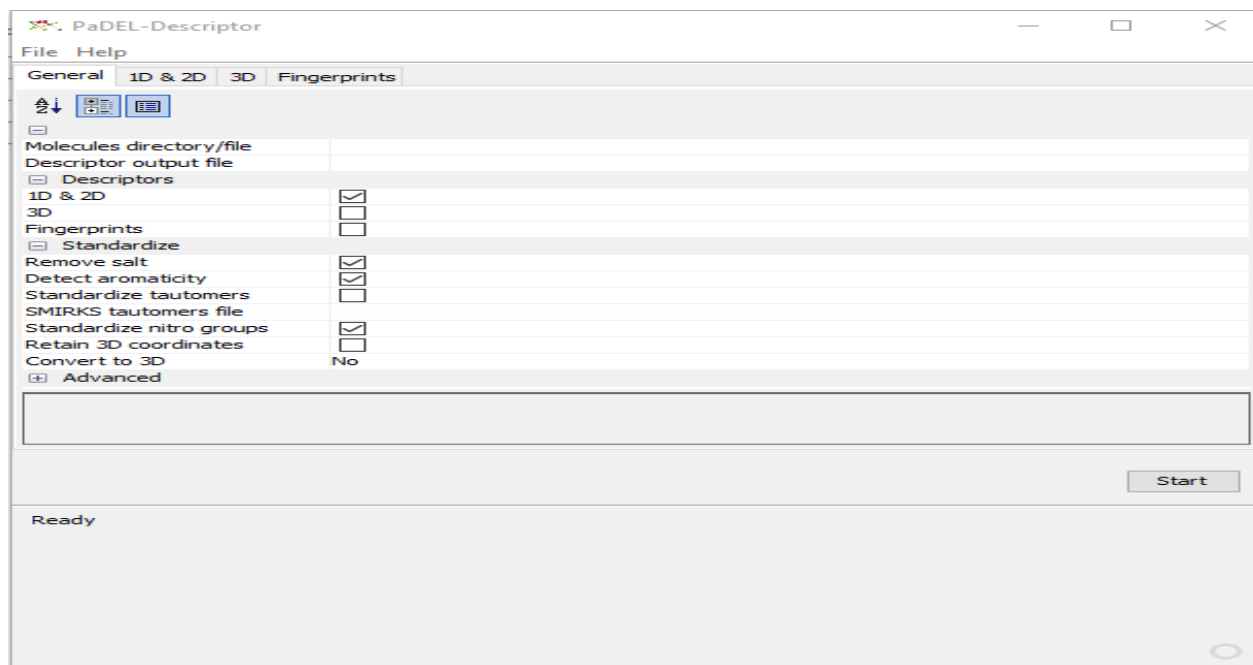
En plus de modèle QSAR qui nous a donné un aperçu sur les descripteurs qui affecte l'activité inhibitrice en peu faire aussi un docking moléculaire et une étude de DFT qui peut nous aider à déterminer la structure clé nécessaire pour inhiber le STS.

ANNEXES

Annexe I :



Annexe II :



Annexe III :

The screenshot displays the 'Model Development Module' interface. At the top, there is a navigation bar with 'Model Development Module' and 'Predict Module' tabs. The status bar indicates 'Status: Setting Parameters' and '*DTC-QSAR version 1.0.5'. The main settings area includes:

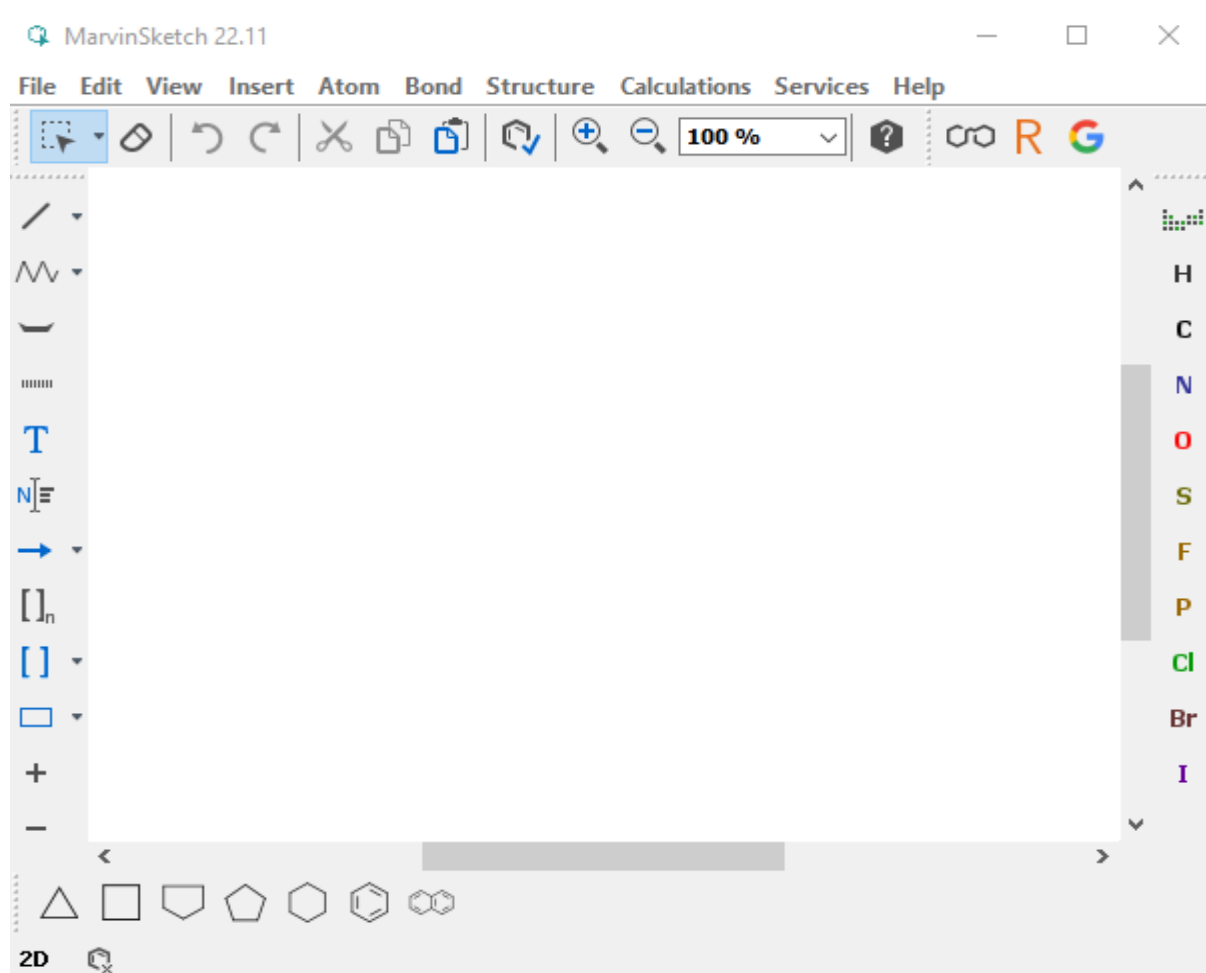
- Select the QSAR approach:** Regression (dropdown)
- Positive Group Notation:** P (text input)
- Negative Group Notation:** N (text input)
- Perform Data-PreTreatment:** (checkbox)
- Variance cut-off:** 0.001 (text input)
- Inter-R² cut-off:** 0.70 (text input)
- Remove less discriminating features:** (checkbox)
- Threshold:** 0.05 (text input)
- Choose Division Method:** Random(*Faster) (dropdown)
- Percent of compounds in Training set:** 70 (text input)
- Seed value (Random division):** 1 (text input)
- Choose Model Development Method:** BSS-MLR (dropdown)
- Settings:** (button)

Annexe IV :

The screenshot shows a window titled 'DTC Lab.' with the subtitle 'Applicability Domain Using Std. Approach'. The interface includes the following elements:

- Select Training set file:** (text input)
- Select Test set file :** (text input)
- Select Output file :** (text input)
- Submit:** (button)

Annexe V :



RÉFÉRENCES

Références

- [1] “Steroidal N-Sulfonylimidates: Synthesis and biological evaluation in breast cancer cells - PubMed.” <https://pubmed.ncbi.nlm.nih.gov/31284080/> (accessed Jun. 22, 2022).
- [2] “Hormonothérapie du cancer du sein - ScienceDirect.” <https://www.sciencedirect.com/science/article/abs/pii/S0368231510002693> (accessed Jun. 23, 2022).
- [3] S. R and K. Mk, “Lead optimization of 4-(thio)-chromenone 6-O-sulfamate analogs using QSAR, molecular docking and DFT - a combined approach as steroidal sulfatase inhibitors,” *J. Recept. Signal Transduct. Res.*, vol. 41, no. 2, pp. 123–137, Apr. 2021, doi: 10.1080/10799893.2020.1794004.
- [4] “Profil épidémiologique et anatomopathologique du cancer de sein au CHU Ibn Rochd, Casablanca - PMC.” <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7648477/#ref1> (accessed Jun. 22, 2022).
- [5] “Cancer.” <https://www.who.int/fr/news-room/fact-sheets/detail/cancer> (accessed Jun. 22, 2022).
- [6] H. Sancho-Garnier and M. Colonna, “Épidémiologie des cancers du sein,” *Presse Médicale*, vol. 48, no. 10, pp. 1076–1084, Oct. 2019, doi: 10.1016/j.lpm.2019.09.022.
- [7] A. Nkondjock and P. Ghadirian, “Facteurs de risque du cancer du sein,” *médecine/sciences*, vol. 21, no. 2, Art. no. 2, Feb. 2005, doi: 10.1051/medsci/2005212175.
- [8] “JAKO201528551641823.pdf.” Accessed: Jun. 27, 2022. [Online]. Available: <https://www.koreascience.or.kr/article/JAKO201528551641823.pdf>
- [9] “Anatomy of the human mammary gland: Current status of knowledge - PubMed.” <https://pubmed.ncbi.nlm.nih.gov/22997014/> (accessed Jun. 22, 2022).
- [10] “Anatomie du sein.” <http://www.depistagesein.ca/anatomie-du-sein/#.YrOZG3bMLgz> (accessed Jun. 22, 2022).
- [11] “Mammary gland development - Macias - 2012 - WIREs Developmental Biology - Wiley Online Library.” <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wdev.35> (accessed Jun. 22, 2022).
- [12] A. G. Waks and E. P. Winer, “Breast Cancer Treatment: A Review,” *JAMA*, vol. 321, no. 3, pp. 288–300, Jan. 2019, doi: 10.1001/jama.2018.19323.

- [13] "M2812019.pdf." Accessed: Jun. 22, 2022. [Online]. Available: <http://ao.um5.ac.ma/xmlui/bitstream/handle/123456789/17710/M2812019.pdf?sequence=1&isAllowed=y>
- [14] E. Fleurot, "Mise en évidence d'un dialogue entre la signalisation œstrogénique et le syndécane-1 dans les cellules de carcinome mammaire humain MCF7," p. 273.
- [15] J. M. Hall, J. F. Couse, and K. S. Korach, "The multifaceted mechanisms of estradiol and estrogen receptor signaling," *J. Biol. Chem.*, vol. 276, no. 40, pp. 36869–36872, Oct. 2001, doi: 10.1074/jbc.R100029200.
- [16] J. P. Tiano and F. Mauvais-Jarvis, "Importance of oestrogen receptors to preserve functional β -cell mass in diabetes," *Nat. Rev. Endocrinol.*, vol. 8, no. 6, pp. 342–351, Feb. 2012, doi: 10.1038/nrendo.2011.242.
- [17] "Cancers du sein - Les maladies du sein." <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Les-maladies-du-sein/Cancers-du-sein> (accessed Jun. 23, 2022).
- [18] A. Fouad, A. Yousra, Z. Kaoutar, E. M. Omar, A. Afaf, and B. Sanae, "Classification moléculaire du cancer du sein au Maroc," *Pan Afr. Med. J.*, vol. 13, p. 91, Dec. 2012.
- [19] "Precision medicine in breast cancer: reality or utopia? | Journal of Translational Medicine | Full Text." <https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-017-1239-z> (accessed Jun. 23, 2022).
- [20] Z. Desta *et al.*, "Antiestrogen pathway (aromatase inhibitor)," *Pharmacogenet. Genomics*, vol. 19, no. 7, pp. 554–555, Jul. 2009, doi: 10.1097/FPC.0b013e32832e0ec1.
- [21] P. Mujumdar and S.-A. Poulsen, "Natural Product Primary Sulfonamides and Primary Sulfamates," *J. Nat. Prod.*, vol. 78, no. 6, pp. 1470–1477, Jun. 2015, doi: 10.1021/np501015m.
- [22] "Sulfamates and their therapeutic potential - Winum - 2005 - Medicinal Research Reviews - Wiley Online Library." <https://onlinelibrary.wiley.com/doi/abs/10.1002/med.20021> (accessed Jun. 23, 2022).
- [23] "(PDF) Biological activity and synthesis of sulfonamide derivatives: A brief review."

https://www.researchgate.net/publication/286361669_Biological_activity_and_synthesis_of_sulfonamide_derivatives_A_brief_review (accessed Jun. 23, 2022).

[24] “Epitomax 15 mg, gélule, boîte de 1 flacon de 28.” <https://sante.lefigaro.fr/medicaments/3483051-epitomax-15mg-gelule-28> (accessed Jun. 23, 2022).

[25] A. Tropsha, “Best Practices for QSAR Model Development, Validation, and Exploitation,” *Mol. Inform.*, vol. 29, no. 6–7, pp. 476–488, 2010, doi: 10.1002/minf.201000061.

[26] A. Cherkasov *et al.*, “QSAR Modeling: Where have you been? Where are you going to?,” *J. Med. Chem.*, vol. 57, no. 12, pp. 4977–5010, Jun. 2014, doi: 10.1021/jm4004285.

[27] M. Ghamali, S. Chtita, M. Bouachrine, and T. Lakhlifi, “Méthodologie générale d’une étude RQSA/RQSP,” *Rev. Interdiscip.*, vol. 1, no. 1, Art. no. 1, 2016, Accessed: Jun. 23, 2022. [Online]. Available: <https://revues.imist.ma/index.php/Revue-Interdisciplinaire/article/view/6222>

[28] “Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment - 1st Edition.” <https://www.elsevier.com/books/understanding-the-basics-of-qsar-for-applications-in-pharmaceutical-sciences-and-risk-assessment/roy/978-0-12-801505-6> (accessed Jun. 23, 2022).

[29] L. Zhang, H. Zhu, T. I. Oprea, A. Golbraikh, and A. Tropsha, “QSAR modeling of the blood-brain barrier permeability for diverse organic compounds,” *Pharm. Res.*, vol. 25, no. 8, pp. 1902–1914, Aug. 2008, doi: 10.1007/s11095-008-9609-0.

[30] “LOOCV Leave One Out Cross Validation.” <https://dataaspirant.com/7-loocv-leave-one-out-cross-validation/> (accessed Jun. 23, 2022).

[31] P. Pratim Roy, S. Paul, I. Mitra, and K. Roy, “On two novel parameters for validation of predictive QSAR models,” *Mol. Basel Switz.*, vol. 14, no. 5, pp. 1660–1701, Apr. 2009, doi: 10.3390/molecules14051660.

[32] S. Kar, K. Roy, and J. Leszczynski, “Applicability Domain: A Step Toward Confident Predictions and Decidability for QSAR Modeling,” in *Computational Toxicology: Methods and Protocols*, O. Nicolotti, Ed. New York, NY: Springer, 2018, pp. 141–169. doi: 10.1007/978-1-4939-7899-1_6.

