



الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي



Ministère de l'enseignement Supérieure et de la Recherche Scientifique

Université Frères Mentouri Constantine 1

جامعة الإخوة منتوري قسنطينة 1

Faculté des Sciences de la Nature et de la Vie

كلية علوم الطبيعة والحياة

قسم الكيمياء الحيوية والبيولوجيا الخلوية والجزيئية

Département de Biochimie et Biologie Cellulaire et Moléculaire

Mémoire présenté en vue de l'obtention du Diplôme de Master

Filière : Sciences Biologiques

Spécialité : Biochimie appliquée

Intitulé

Analyse de la relation quantitative

Structure / Activité des inhibiteurs de la tyrosine kinase

Présenté et soutenu par :

Le : 19/09/2021

REDJAH ZAKARIA

SELLAMI MOHAMED AMINE

Devant le jury :

Président du jury : Mr BENSEGUENI. A. Pr. UFM Constantine-1

Encadreur : Mr. DEMS MOHAMED ABDESSELEM. MRA CRBt Constantine

Examineur : Mr. MOKRANI E. H Maitre- Conférences. B. UFM Constantine-1

Examineur : Mr. BOUKELIA ABDELBASSET. MRA CRBt Constantine

Année universitaire
2020- 2021



الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي



Ministère de l'enseignement Supérieure et de la Recherche Scientifique

Université Frères Mentouri Constantine 1

جامعة الإخوة منتوري قسنطينة 1

Faculté des Sciences de la Nature et de la Vie

كلية علوم الطبيعة والحياة

قسم الكيمياء الحيوية والبيولوجيا الخلوية والجزيئية

Département de Biochimie et Biologie Cellulaire et Moléculaire

Mémoire présenté en vue de l'obtention du Diplôme de Master

Filière : Sciences Biologiques

Spécialité : Biochimie appliquée

Intitulé

Analyse de la relation quantitative Structure / Activité des inhibiteurs de la tyrosine kinase

Présenté et soutenu par :

Le : 19/09/2021

REDJAH ZAKARIA

SELLAMI MOHAMED AMINE

Devant le jury :

Président du jury : Mr BENSEGUENI. A. Pr. UFM Constantine-1

Encadreur : Mr. DEMS MOHAMED ABDESSELEM. MRA CRBt Constantine

Examineur : Mr. MOKRANI E. H Maitre- Conférences. B. UFM Constantine-1

Examineur : Mr. BOUKELIA ABDELBASSET. MRA CRBt Constantine

Année universitaire
2020- 2021

Remerciements

Nous tenons à exprimer toute notre reconnaissance à notre encadreur MONSIEUR DEMS MOHAMED ABDESSELEM, maître de recherche classe A au centre de recherche en biotechnologie (CRBt) qui a dirigé notre travail. Ses conseils et ses commentaires précieux nous ont permis de surmonter notre difficulté et de progresser dans notre étude.

Nous tenons de remercier le chef du département de BIOCHIMIE de nous avoir guidé, orienté, et facilité notre travail.
Nous adressons notre plus grande et respectueuse reconnaissance au Pr. BENSEGUENI A. Professeur à l'université des frères Mentouri-Constantine 1 pour avoir accepté de présider le jury et pour le temps consacré à l'examen de ce mémoire.

Ainsi Mr. MOKRANI E.H. Docteur et Maître-assistant A, à l'université des frères Mentouri-Constantine 1, pour nous avoir fait l'honneur de juger ce travail et d'avoir participer au jury.

Nous adressons nos sincères remerciements à tous les enseignants qui ont assuré notre formation durant les cinq ans d'étude et toutes les personnes qui ont contribué de près ou de loin à toute notre cycle d'étude.

Ainsi nous remercions toute l'équipe du centre de recherche en Biotechnologie (C,R,Bt) labo 18, et tous ceux et celles qui ont contribué de près ou de loin à la réalisation de ce travail.

Nous remercions infiniment nos très chers parents, lesquels grâce à Dieu et à eux nous sommes ici ; ils ont toujours été là pour nous, merci pour tout ce que vous avez fait et entraîné de faire pour nous.

Nos remerciements à nos sœurs et frères pour leur soutien et encouragements.

Enfin nous remercions nos collègues et nos chers amis pour leur soutien inconditionnel.

A tous ces intervenants nous présentons notre remerciement, respect et gratitude.

Veillez agréer l'expression de nos sentiments les plus distingués.

Dédicace

Je dédie ce modeste travail

A mes chers parents Mohamed et Hanifa

Pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études.

A ma chère sœur Rania

A mon cher beau-frère Naoufel

A ma nièce Maysan

A mes amis Amine, aymen, Ouissal et Nadir

A toute ma famille pour leur soutien tout au long de mon parcours universitaire, Que ce travail soit l'accomplissement de vos vœux tant allégués, et le fruit de votre soutien infailible, Merci d'être toujours là pour moi.

Zakaria Redjah

Dédicace

Je dédie ce mémoire à :

Mes chers parents,

Que nulle dédicace ne puisse exprimer mes sincères sentiments, pour leur patience illimitée, leur encouragement contenu, leur aide, en témoignage de mon profond amour et respect pour leur grands sacrifices.

Ma grande mère Malika

Que dieu la procure bonne santé et longue vie.

A la mémoire de mes grands-parents : Achour, Lamri, ainsi ma Hamama

Que dieu les bénisses.

Mes chers frères : *Yahia, Djaber.*

Mes chères soeur : Nour, Djihène, ainsi que ma sœur aînée Sarah, son mari Salah, ses petits Ouais et Firras.

Mon cher oncle : *Nadir,*

Mon cher cousin : Seifeddine et sa femme Insaf.

Pour leur grand amour et leur soutien qu'ils trouvent ici l'expression de ma haute gratitude.

Mes amis et collègues,

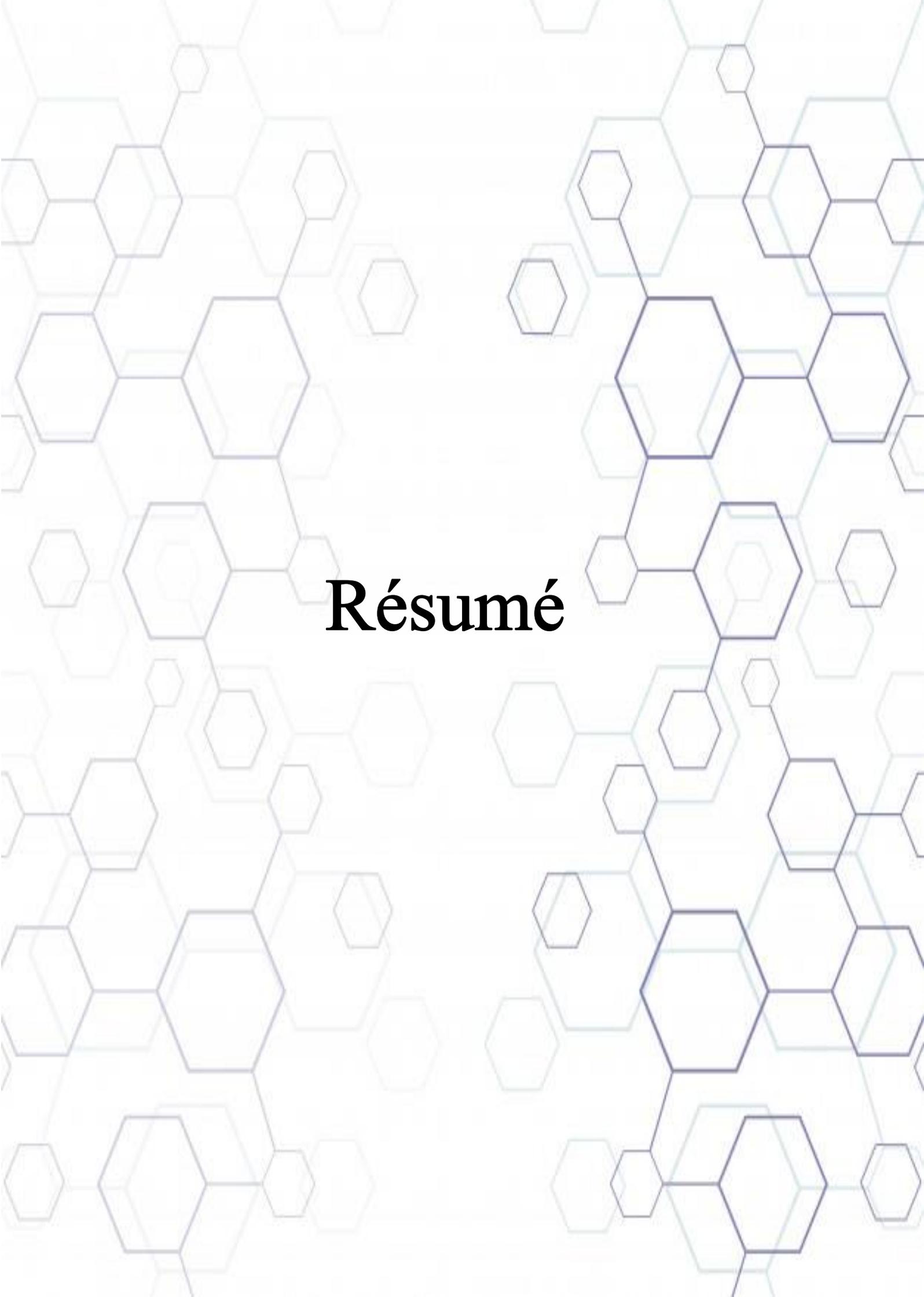
Pour leur encouragement et pour tous les bons moments qu'on a vécus ensemble.

J'espère que notre amitié durera éternellement.

Et à toute ma famille, à tous ceux qui m'aime et que j'aime.

A tout ceux que m'ont aidé de près ou de loin.

Mohamed Amine



Résumé

ملخص

تم استغلال العلاقة الكمية بين الهيكل والنشاط المثبط للتيروزين كيناز لـ 59 مشتقاً من الساليسيلانيليد. تحسين هندسة المركبات المستخدمة في هذا العمل مسبقاً باستخدام الميكانيكا الجزيئية MM + ثم تحسينها باستخدام طريقة PM3 شبه التجريبية، وبالتالي خضعت الأشكال الهندسية المحسنة لحساب واصفاتها. تم بناء نموذج QSAR على أساس الانحدار الخطي المتعدد (RLM) ، وتم ربط الخوارزمية الجينية (GA) لتحديد أهم الواصفات. تم تطوير نموذج QSAR للتنبؤ بالنشاط المثبط للتيروزين كيناز من 4 واصفات (MATS2v², ESpm12d², G1m², H5u²). تم استخدام عناصر تحكم Roy وشروط galbraikh و tropsha والتحقق منها. تم التحقق من صحة النموذج المعمول به وفقاً لعمليات التحقق الداخلية والخارجية للحكم على استقراره وقوته وقدرته على التنبؤ (R²=58.2% ، Q²LOO=50.66% ، Q²training=50.7% ، R²test=83.4 ، F=13.92). وقد درست مجال التطبيق لنموذجنا باستخدام مخطط وليامز للكشف عن المركبات الشاذة. يمكن استخدام هذا النموذج بفعالية للتنبؤ بالنشاط المثبط للجزيئات التي لم يتم اختبارها تجريبياً الكلمات الدالة: نموذج QSAR، النشاط المثبط للتيروزين كيناز، الخوارزمية الجينية (GA) ، الساليسيلانيليد

Abstract

The quantitative relationship between the structure and the Tyrosine Kinase inhibitory activity for 59 Salicylanilide derivatives has been exploited. The geometry of the compounds used in this work were pre-optimized with MM + molecular mechanics and then optimized using the semi-empirical PM3 method, the geometries thus optimized were subjected to a calculation of their descriptors. A QSAR model was built based on multiple linear regression (RLM), the genetic algorithm (GA) was associated to select the most important descriptors. The QSAR model for the prediction of Tyrosine Kinase inhibitory activity was developed from 4 descriptors (MATS2v², ESpm12d², G1m², H5u²), Roy's controls and Golbraikh et Tropsha conditions are used and verified. The established model has been validated according to internal and external validations to judge its stability, robustness and predictive capacity ($R^2=58.2\%$, $Q^2_{LOO}=50.66\%$, $Q^2_{training}=50.7\%$, $F=13.92$, $R^2_{test}=83.4$). The field of application of our model is studied using the Williams diagram to detect aberrant compounds. This model can be used effectively to predict inhibitory activity for molecules that have not been tested experimentally.

Keywords: QSAR model, Tyrosine Kinase inhibitory activity, genetic algorithm (GA), Salicylanilide.

Résumé

La relation quantitative entre la structure et l'activité inhibitrice de la Tyrosine Kinase pour 59 dérivés de Salicylanilide a été exploitée. La géométrie des composés utilisés dans ce travail ont été pré-optimisés avec le Mécanique moléculaire MM+ puis optimisé à l'aide de la méthode semi-empirique PM3, les géométries ainsi optimisées ont été soumises à un calcul de leurs descripteurs.

Un modèle QSAR a été construit en se basant sur la régression linéaire multiple (RLM), l'algorithme génétiques (GA) a été associé pour sélectionner les descripteurs les plus importants. Le modèle QSAR pour la prédiction l'activité inhibitrice de la Tyrosine Kinase a été développé à partir de 4 descripteurs ($MATS2v^2$, $ESpm12d^2$, $G1m^2$, $H5u^2$), les contrôles de Roy et les conditions de Golbraikh et Tropsha sont utilisés et vérifiés.

Le modèle établi, a été validé selon des validations internes et externes pour juger sa stabilité, sa robustesse et sa capacité prédictive ($R^2=58.2\%$, $Q^2_{LOO}=50.66$, $Q^2_{training}=50.7\%$, $F=13.92$, $R^2_{test}=83.4$). Le domaine d'application pour notre modèle est étudié à l'aide du diagramme de Williams pour détecter les composés aberrants.

Ce modèle peut être utilisé efficacement pour estimer des molécules qui n'ont pas été testées expérimentalement.

Mots clés : modèle QSAR, activité inhibitrice de la tyrosine kinase, algorithme génétique (AG), salicylanilide

Liste des tableaux

Tableau 1:L'activité inhibitrice de 32 dérivés de salicylanilide (A).....	18
Tableau 2:L'activité inhibitrice de 27 dérivés de salicylanilide (B).....	20
Tableau 3:repartition de l'ensemble de tests pour les trois méthodes de divisions Kennard- Stone, Euclidean Distance et Activity/Property.....	24
Tableau 4: les trois modèles pour les trois algorithmes par les 3 méthodes de division.....	25
Tableau 5: les critères de choix entre les trois méthodes de division	25
Tableau 6: Matrice de corrélation entre pIC50 et les descripteurs impliqués dans le modèle QSAR.....	27
Tableau 7: Les valeurs expérimentales prédit et calculé pour pIC50 pour les 59 dérivés de salicylanilide.	29
Tableau 8: Les paramètres de la validation interne.....	32
Tableau 9: Les paramètres de la validation externe.	32
Tableau 10:Signification des descripteurs.....	47
Tableau 11:Les valeurs d'IC50 et les quatre descripteurs.	47

Liste des figures

Figure 1:Progression histopathologique du foie vers le CHC[12].	4
Figure 2:Mécanismes de l'hépatocarcinogenèse[12].	5
Figure 3:Mode d'action des inhibiteurs de tyrosine kinase[23].	7
Figure 4 : Structure générale de sorafénib [27].	7
Figure 5:Structure générale des salicylanilides[32].	8
Figure 6:Processus de la construction d'un model QSAR [44].	10
Figure 7:Algorithme La méthode Basé sur la distance euclidienne.	14
Figure 8:Structure dérivés de salicylanilide.	18
Figure 9:La droite de régression des valeurs expérimentales et prédits de la notation pIC50 pour le modèle de QSAR.	28
Figure 10:principe du test de randomisation[86].	33
Figure 11:Test de randomisation associé au modèle QSAR 500 itérations.	34
Figure 12:Diagramme de williams des résidus standardisés.	35

Symboles et abréviation

A

A/P: Activity/Property.

AG : l'Algorithme Génétique.

ATP : Phosphate de l'Adénosine Triphosphate.

C

CHC : Le Carcinome Hépatocellulaire.

D

DS: Deviation Standard.

DE : Euclidean Distance.

E

EAM : Erreur absolue moyenne

F

F: L'indice de Fisher.

G

I

IC₅₀ : La concentration inhibitrice 50.

K

KS : Kennard-Stone.

M

MLR : Régression Linéaire Multiple.

N

NAFLD: Non-Alcoholic Fatty Liver Disease.

NASH : Non-Alcoholic SteatoHepatitis.

O

OMS : Organisation Mondiale de la Santé.

P

pIC₅₀ : $\log 1/(\text{IC}_{50})$.

PRESS : La somme résiduelle prédictive des carrés

Q

Q² : coefficient de validation.

QSAR : Quantitative Structure Activity Relationships (Relations Quantitatives Structure Activité).

Q²LMO : Coefficient de prédiction (Leave-Many-Out).

Q²LOO : Coefficient de validation (Leave-One-Out).

R

R²: Coefficient de détermination.

R²_{adj} : Coefficient de détermination ajusté.

RF: Radiofréquence.

RMSE: Root mean square error (l'écart type).

S

S : La déviation Standard.

SCE : La Somme Expliquée des Carrés.

SCT : La Somme Totale des Carrés.

T

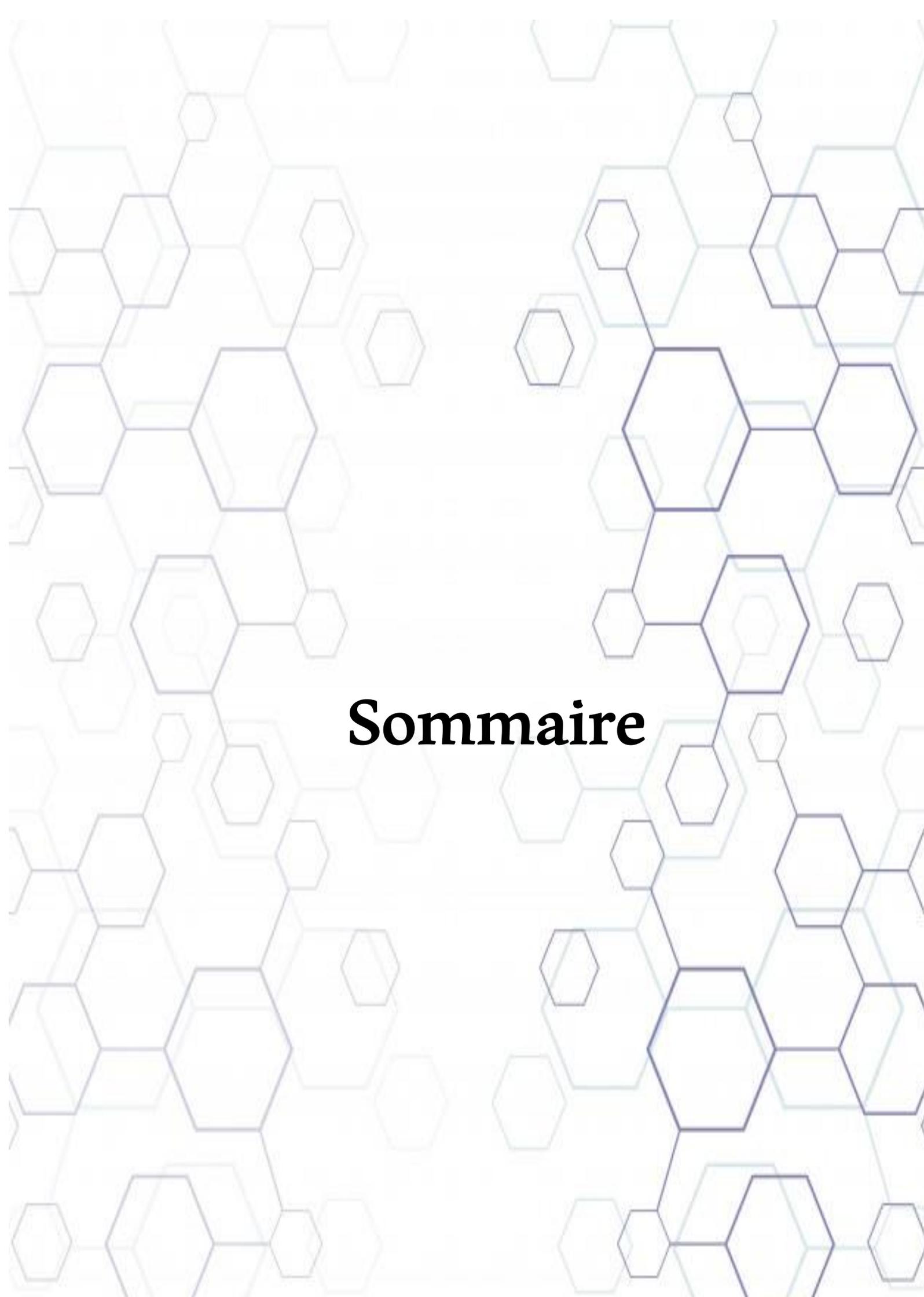
TCS : Systèmes de Régulation Bicomposants

V

VEGFR : Facteurs de Croissance Endothéliaux Vasculaires.

VHB : Virus de l'Hépatite B.

VHC : Virus de l'Hépatite C.



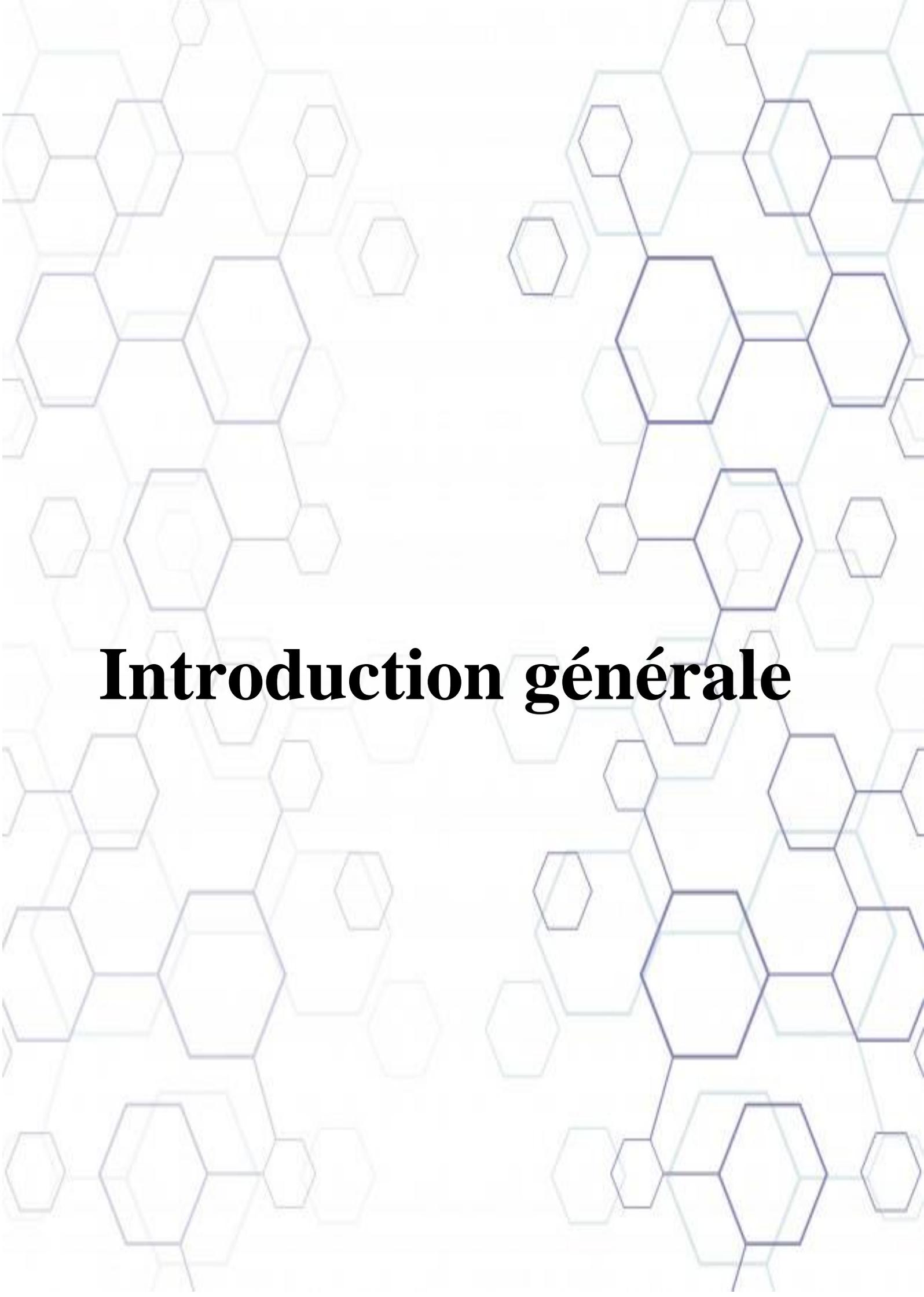
Sommaire

Sommaire

Remerciements	III
Dédicace	IV
Résumé	IX
Liste des tableaux	X
Liste des figures	XI
Symboles et abréviation	XII
Introduction générale.....	1
Partie I : Synthèse Bibliographique.....	3
Chapitre I : Carcinome Hépatocellulaire.....	4
1. Le Carcinome Hépatocellulaire :	3
2. Les facteurs de risque :.....	4
3. L'hépatocarcinogenèse :	5
3.1 Mécanismes de l'hépatocarcinogenèse :.....	5
4. Le traitement de carcinome hépatocellulaire :.....	5
4.1 Le traitement par chirurgie :	6
4.2 Le traitement par la chimiothérapie :.....	6
5. Tyrosine kinase :	6
5.1 Inhibiteur de tyrosine kinase :	6
5.1.1 Sorafénib :.....	7
5.1.1.1 Mode action de sorafénib :	8
5.1.2 Salicylanilide :	8
5.1.2.1 La relation entre le carcinome hépatocellulaire et les Salicylanilides :.....	8
Chapitre II : Relations Quantitatives Structure –Activité (QSAR)	9
1. Principe et processus du model QSAR :	9

2.	Les essentielles étapes de la construction du modèle QSAR :.....	11
3.	Les descripteurs :	11
3.1	Sélection du descripteur :	11
4.	Paramètres statistiques :.....	12
4.1	Les paramètres standards :.....	12
4.1.1	Coefficient de détermination :	12
4.1.2	Le Coefficient de détermination Ajuste noté Rajusté :	12
4.1.3	Le test de Fisher ou F :.....	12
4.1.4	La déviation standard S :.....	13
4.2	Les méthodes de sélection De l'ensemble d'apprentissage et de test :	13
4.2.1	La méthode Activité/propriété :	13
4.2.2	La méthode de Kennard Stone :	13
4.2.3	La méthode Basé sur la distance euclidienne :	13
5.	Validation :.....	15
5.1	Validation interne :	15
5.2	Validation externe :	16
5.2.1	Critères de Golbraikh et Tropsha :.....	16
5.3	Test de randomisation :.....	16
5.4	Domaine D'application :	17
Partie II : Etude Expérimentale		18
1.	Méthodologie	18
1.1	Collecte des données :	18
1.2	Traitements des données :.....	22
1.2.1	La Préparation de l'ensemble de données :.....	22
1.2.2	Le calcul des descripteurs :.....	22
1.2.3	La modélisation :.....	22
2.	Résultats et discussion :	23
2.1	Répartition des sous-ensembles (entraînement & test) :	23
2.1.1	Conclusion :	26

2.2	Analyse de modèle accepté :	27
2.2.1	Evaluation du modèle :	27
2.2.1.1	La colinéarité :	27
2.2.1.2	Analyse de la régression :	27
2.2.1.3	Qualité d'ajustement :	28
2.2.1.4	Validation :	29
2.2.1.4.1	Validation interne :	32
2.2.1.4.2	Validation externe :	32
2.2.1.5	Test de randomisation :	33
2.2.1.6	Domaine d'applications :	35
	Conclusion générale	36
	References	38
	Annexes :	47
	Résumé	50



Introduction générale

Introduction générale

Le carcinome hépatocellulaire (CHC) est un cancer primitif du foie. C'est une maladie silencieuse, dangereuse et mortelle, plus fréquente chez les personnes atteintes de cirrhose à cause d'un problème majeur de santé publique à la raison que le diagnostic est trop tardif [1].

Le traitement de cette maladie basé principalement sur certaines molécules, sont utilisées comme inhibiteurs des tyrosines kinases tel que le sorafénib [2].

Au cours de ces dernières années, le développement de nouvelles pistes thérapeutiques s'est amplifié. Il est important de noter que de nouvelles classes d'agents antiviraux comme la famille des salicylanilides de diverses structures ont été proposés et étudiés en tant qu'inhibiteurs de la tyrosine kinases[3,4].

Le développement de nouveaux médicaments plus efficace et moins chère constitue un enjeu majeur pour l'industrie pharmaceutique. Ce qui a obligé les chercheurs de s'orienter vers de nouvelles méthodes qui aide à modeler et prédire de façon rapide l'activité avant même que celles-ci ne soient synthétisées, parmi ces méthodes, les relations quantitatives structure-activité ou QSAR (en anglais Quantitative Structure-Activity Relationship), qui consiste à la recherche d'une relation entre les structures des composés et leurs activités en utilisant des méthodes d'analyses multidimensionnelles par apprentissage statistique afin de modéliser l'activité biologique[5,6].

L'objectif de ce travail vise à utiliser la méthodologie QSAR afin d'élaborer un modèle robuste et fiable qui explique et prédit l'activité anti-cancéreuse d'une série de 59 dérivés de salicylanilide inhibiteurs de tyrosine kinase.

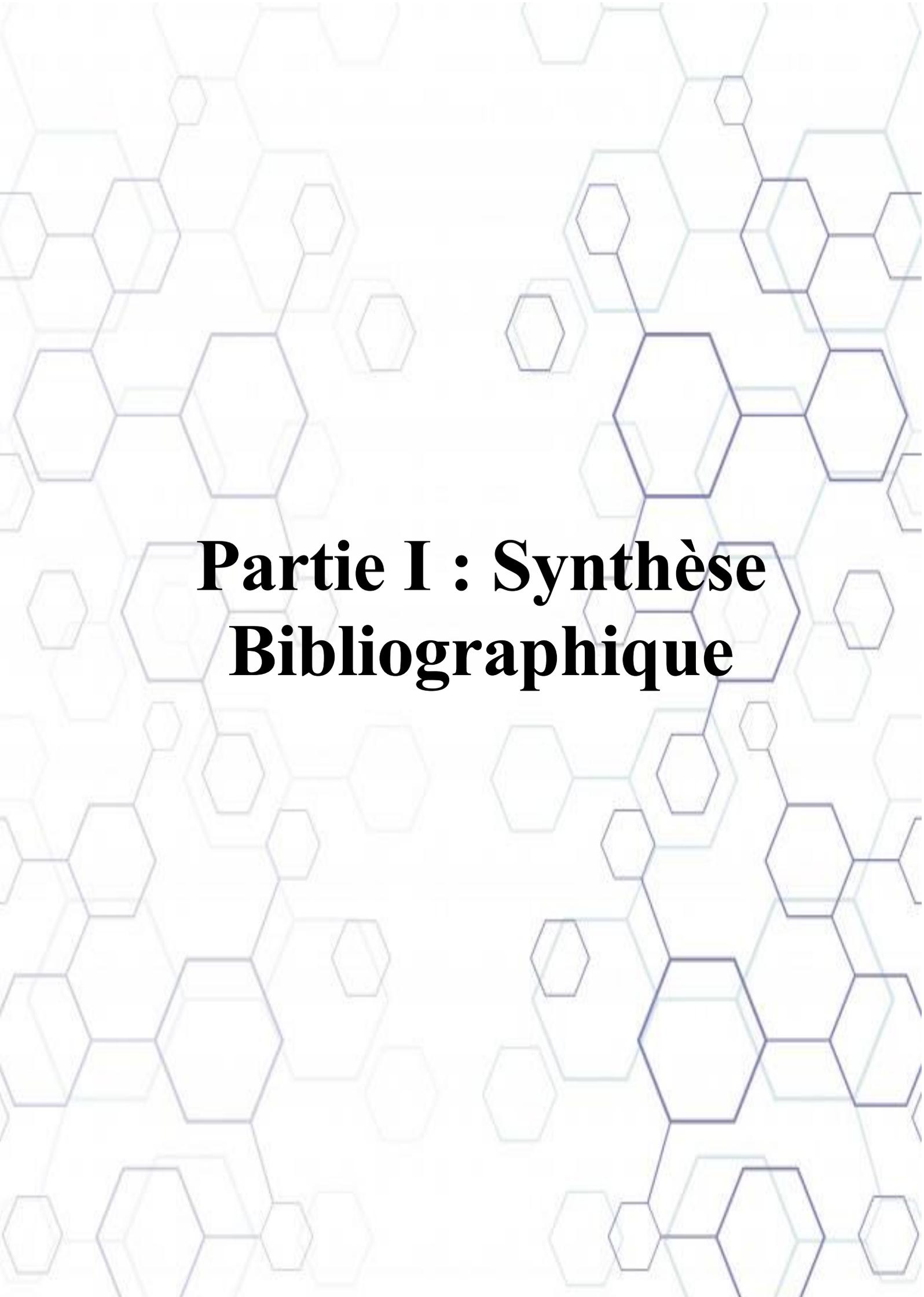
Ce manuscrit s'articulé sur deux grandes parties :

La première partie de ce mémoire présente une bibliographique contient

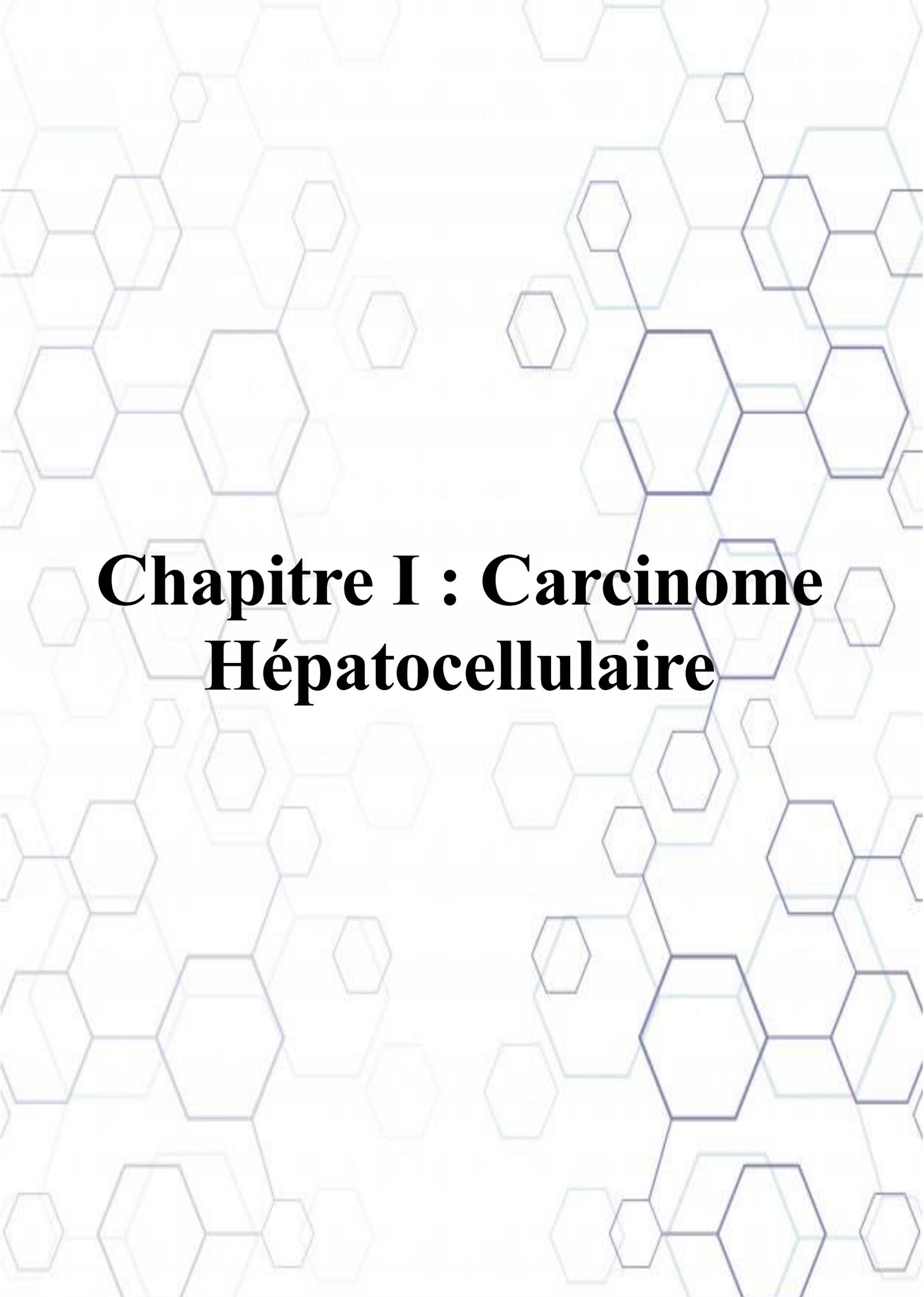
- ✓ Un aperçu général sur le carcinome hépatocellulaire.
- ✓ Rappelle sur la méthodologie QSAR.

Et une deuxième partie expérimentale ou nous présentons et nous discutons les résultats obtenus pour la méthodologie QSAR que nous avons effectué au long de notre séjour au centre de

recherche en biotechnologie de Constantine. Et nous achèverons le travail par une conclusion générale.



Partie I : Synthèse Bibliographique



Chapitre I : Carcinome Hépatocellulaire

1. Le Carcinome Hépatocellulaire :

Le carcinome hépatocellulaire (CHC) est une tumeur primaire du foie qui survient essentiellement sur un terrain de maladie hépatique chronique et est associée à une morbidité et une mortalité élevées ainsi qu'à une prise en charge complexe étant donné l'hépatopathie sous-jacente[7]. Le CHC est le principal cancer primitif du foie (85 à 90%) qui se développe après transformation maligne des hépatocytes[8, 9].

D'après le Centre International de Recherche sur le Cancer, chaque année environ 750 000 nouveaux cas de CHC sont diagnostiqués dans le monde. Le (CHC) est le sixième cancer le plus répandu et la troisième cause de décès dans le monde, c'est un problème majeur de santé publique parce que la raison du diagnostic est trop tardif, il est de très mauvais pronostic [1].

Le CHC a une distribution géographique qui se superpose à celle de l'hépatite B avec une fréquence augmentée en Afrique sub-saharienne et en Asie, particulièrement en Chine[10].

Le CHC peut survenir spontanément mais le plus souvent (70 à 90%) survient au terme d'une maladie chronique, qu'il s'agisse d'une hépatite virale B ou C devenue chronique [11]. Les atteintes du tissu hépatique (comme la cirrhose) induisent une inflammation chronique à l'origine de cycles de nécrose/régénération des hépatocytes, d'éclosion de cellules inflammatoires ainsi que de dépôts de tissu conjonctif et de collagène dans le foie. Ces modifications importantes qui changent la matrice et le microenvironnement du foie, précèdent l'apparition du CHC. Elles sont marquées par l'existence de foyers d'hépatocytes phénotypiquement altérés et plus tard, d'hépatocytes dysplasiques sous forme de nodules qui peuvent évoluer vers le CHC (Figure 1)[12].

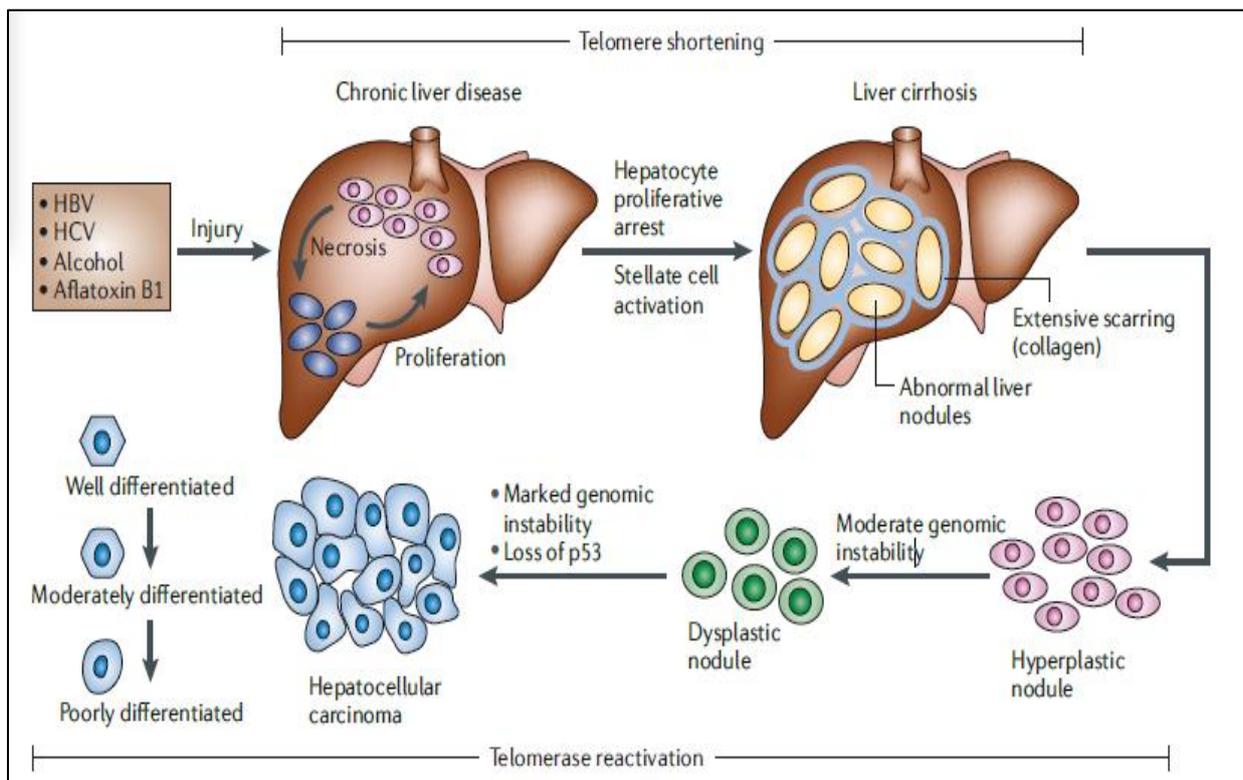


Figure 1: Progression histopathologique du foie vers le CHC[12].

2. Les facteurs de risque :

Les facteurs étiologiques du CHC sont très variables suivant les zones géographiques. Dans les zones à haut risque de CHC comme l'Asie du sud-est et l'Afrique sub-saharienne, La principale cause de survenue d'un CHC est l'infection par le virus de l'hépatite B.

Il a été constaté une forte augmentation d'incidence depuis 20 ans probablement liée à plusieurs facteurs [11, 13–16]:

- VHB, VHC.
- Facteurs alimentaires (aflatoxine B1).
- Obésité, diabète, stéatose.
- Age, sexe,
- Hormones sexuelles.
- Polymorphismes génétiques.
- Hémochromatose et surcharges en fer.
- Macronodules, dysplasie

3. L'hépatocarcinogénèse :

3.1 Mécanismes de l'hépatocarcinogénèse :

L'évolution vers le CHC répond à un processus multi-étages accompagnant l'exposition à différents facteurs de risque. Ces différents facteurs de risque peuvent affecter des voies de signalisation communes. C'est le cas par exemple du gène suppresseur de tumeur p53 qui peut être inactivé ou muté dans les CHC induits par le VHB, le VHC et l'aflatoxine B1. D'autre part, une inflammation accompagnée de cycles continus de nécrose/régénération et la présence d'un stress oxydatif sont caractéristiques de l'hépatocarcinogénèse induites par le VHB, le VHC et l'alcool (Figure 3). Actuellement, les mécanismes moléculaires de l'hépatocarcinogénèse ne sont pas très clairs, et ceci est dû principalement aux causes multifactorielles de ce cancer. Cependant, il est possible d'identifier un ensemble d'altérations spécifiques au CHC[12].

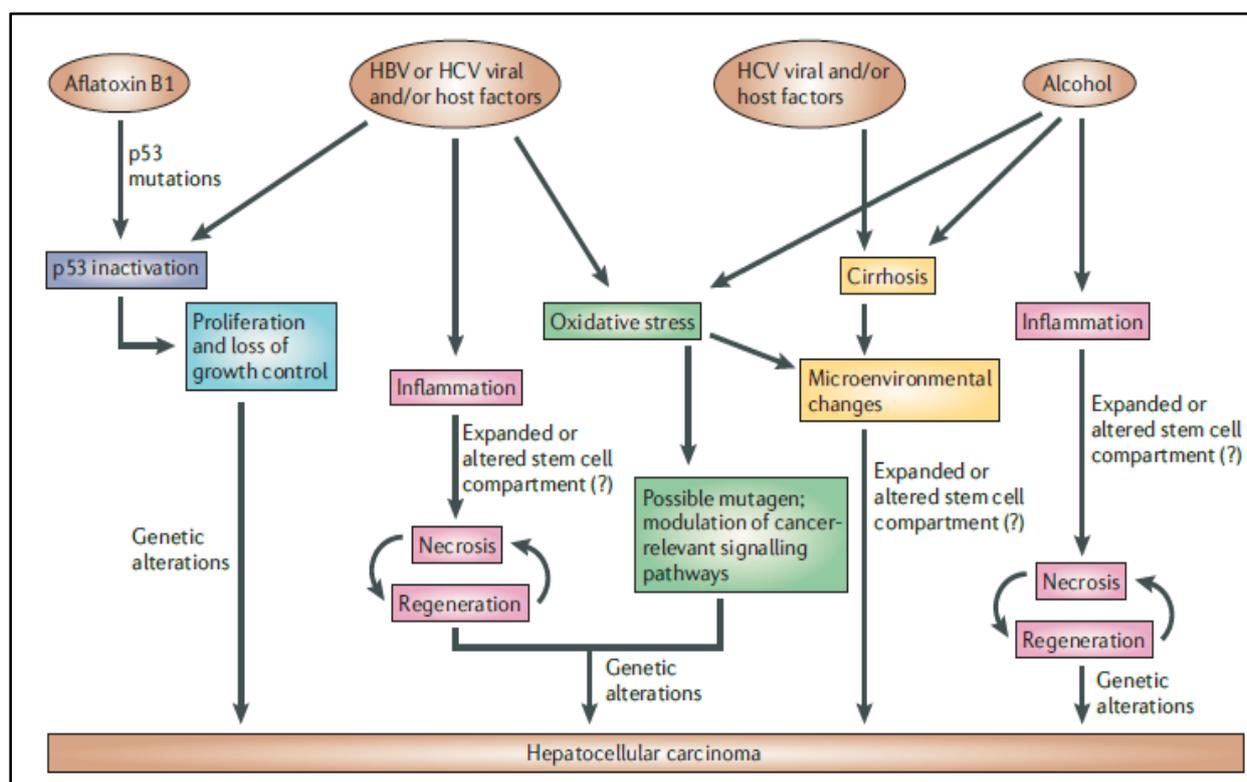


Figure 2: Mécanismes de l'hépatocarcinogénèse[12].

4. Le traitement de carcinome hépatocellulaire :

Tous les types de cancer du foie sont parmi les maladies les plus difficiles à traiter. Dans quelques cas, le cancer primitif du foie est détecté à un stade précoce : le stade où les chances de guérison sont bonnes[17]. De plus, il est difficile de traiter le cancer secondaire du foie ou les métastases d'un autre cancer qui s'est propagé au foie, car le cancer dans ce cas s'est métastasé[18].

4.1 Le traitement par chirurgie :

Les tumeurs détectées aux stades initiaux peuvent être retirées par chirurgie. Les patients dont la tumeur est découverte à un stade précoce ont les plus grandes chances de guérison[19].

Malheureusement, dans la plupart des cas de cancer du foie, la chirurgie ne peut pas être effectuée, soit parce que le cancer est à un stade avancé, soit parce que l'infection est si grave qu'elle ne peut survivre à la chirurgie pour la culture de greffe (transplantation)[20].

4.2 Le traitement par la chimiothérapie :

Il s'agit d'un traitement médicamenteux dans lequel des produits chimiques puissants sont utilisés pour détruire les cellules à croissance rapide dans le corps, Environ 20 thérapies moléculaires ciblées ont été approuvées dans différents types de cancer comme le *carcinome* hépatocellulaire par une inhibition multi-kinase, le sorafénib, qui inhibe la kinase *sérine* et les récepteurs des facteurs de croissance endothéliaux vasculaires (VEGFR)[21].

Abou-Alfa montre dans une Étude de phase II du sorafénib chez des patients atteints d'un stade avancé de Carcinome hépatocellulaire les effets indésirables ont été les suivants [22]:

- L'asthénie (9,5 %), La diarrhée (8,0 %), Le syndrome main-pied (5,1 %)[22].

5. Tyrosine kinase :

La tyrosine kinase est une transférase qui catalyse la réaction chimique d'un groupement phosphate de l'adénosine triphosphate (ATP) vers une protéine effectrice qui joue un rôle essentiel dans la signalisation et la régulation cellulaire. Cette enzyme permet d'induire la prolifération et la croissance cellulaire tumorale[23].

5.1 Inhibiteur de tyrosine kinase :

Un inhibiteur de tyrosine kinase est un médicament, antagoniste réversible, qui permet le blocage de l'activité des tyrosines kinases, enzymes impliquées dans le processus de la signalisation cellulaire (Figure 4). Les tyrosines kinases jouent un rôle dans la communication, le développement, la division et la croissance des cellules. Les inhibiteurs des tyrosines kinases sont donc un type de traitement par inhibiteur du facteur de croissance. On les utilise notamment dans le traitement des cancers[24].

Parmi ces inhibiteurs :

- Sorafénib, Lmatinib, Gefitinib
- salicylanilide, on peut classer les salicylanilide parmi les Inhibiteurs de la tyrosine kinase . Les relations structure-activité préliminaires ont été discuté et plusieurs composés se sont avérés être de puissants inhibiteurs de la protéine tyrosine kinase [24].

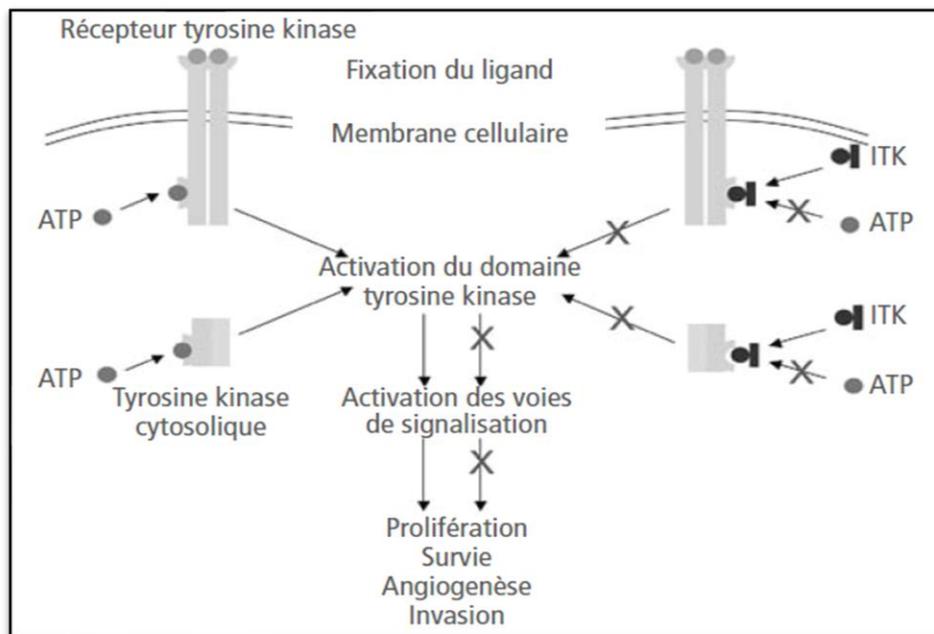


Figure 3: Mode d'action des inhibiteurs de tyrosine kinase[23].

5.1.1 Sorafénib :

Le sorafénib fait partie de la classe des phénylurées, c'est-à-dire l'urée dans laquelle l'un des azotes est substitué par un groupe 4-chloro-3-trifluorophényle tandis que l'autre est substitué par un groupe phényle qui, à son tour, est substitué en position para par un groupe [2-(méthylcarbamoyl)pyridin-4-yl]oxy (Figure 5)[25–27] .

Le sorafénib est un inhibiteur multi cible de tyrosines kinases actuellement indiqué dans le traitement du carcinome hépatocellulaire avancé et le premier médicament à obtenir une amélioration très significative de l'état général des patients atteints d'un cancer du foie avancé [2, 28].

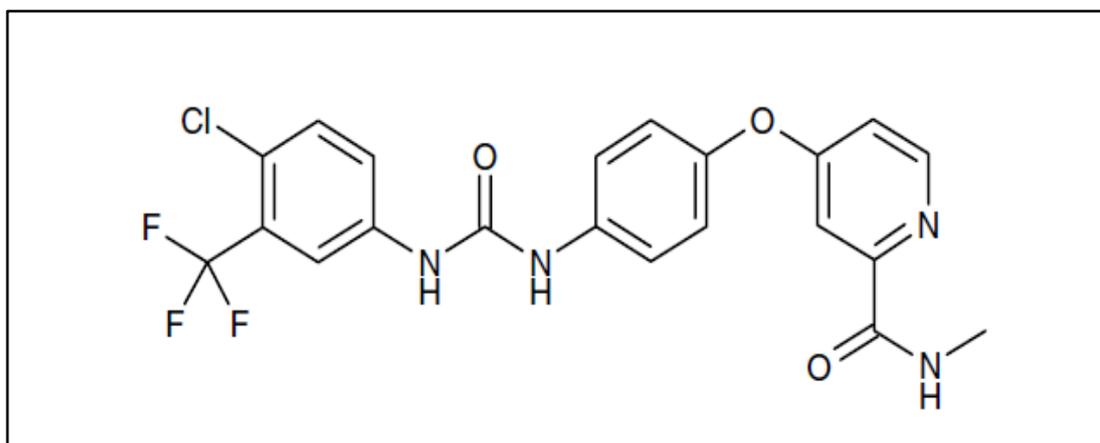


Figure 4 : Structure générale de sorafénib [27].

5.1.1.1 Mode action de sorafénib :

Le sorafenib est un inhibiteur de plusieurs protéines à activité sérine/thréonine kinase et de récepteurs à activité tyrosine-kinase[29]. Par une antiprolifératif sur les cellules tumorales du carcinome hépatocellulaire en inhibe la voie fibrosarcome à accélération rapide et par une antigénique en ciblant des récepteurs à activité tyrosine kinase tel que le Récepteur du facteur de croissance endothélial vasculaire (VEGFR)[30].

5.1.2 Salicylanilide :

Les salicylanilides ont été étudiés pendant de nombreuses années, montrant un large spectre de propriétés intéressantes. Activités biologiques avec une perspective d'utilisation potentielle. Les salicylamides (y compris les salicylanilides) ont montré une suppression de l'activité inhibitrice contre différentes espèces de virus[4].

Les salicylanilides étaient considérés comme des phénols en raison de la présence d'un groupe hydroxyle aromatique libre et de leur mécanisme d'action[31].

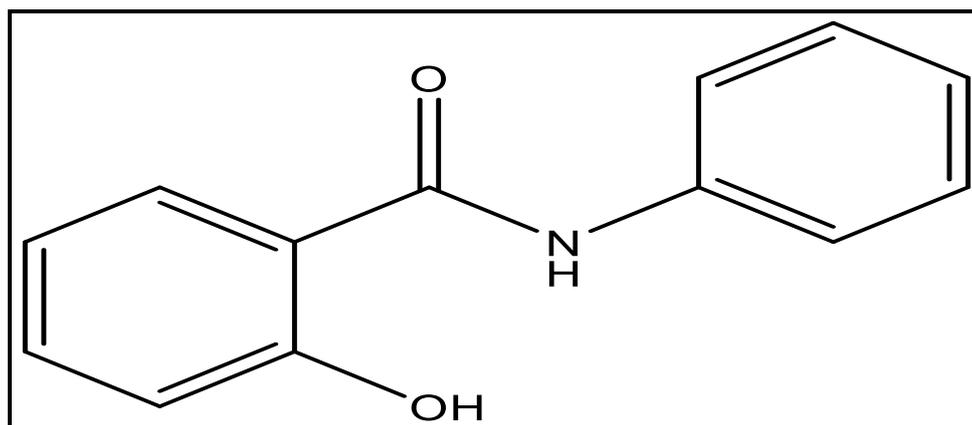
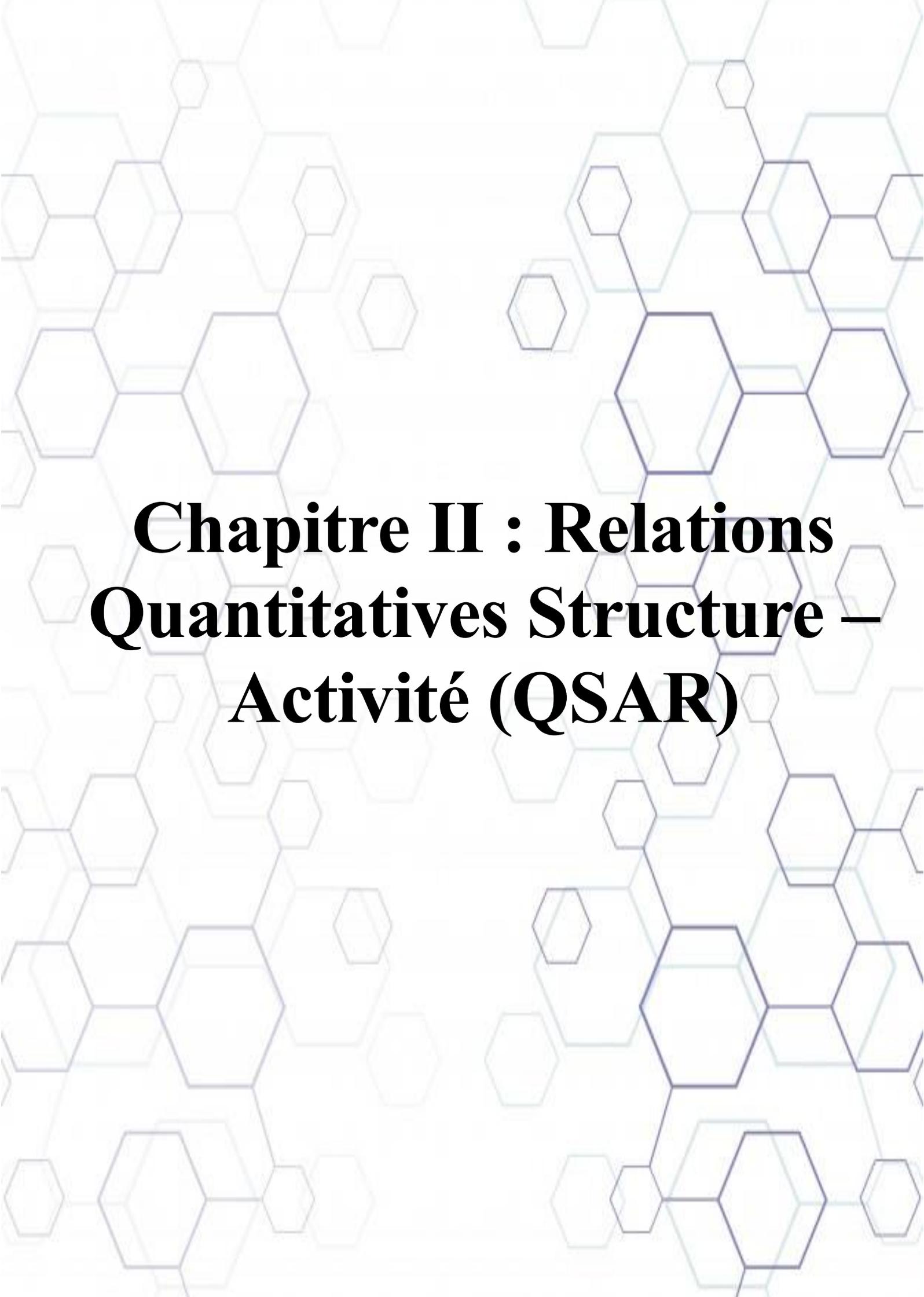


Figure 5: Structure générale des salicylanilides[32].

5.1.2.1 La relation entre le carcinome hépatocellulaire et les Salicylanilides :

Les salicylanilides sont très lipophiles et peuvent traverser les membranes, en particulier l'intérieur de membrane mitochondriale, éliminant ainsi le proton gradient et découpler la phosphorylation oxydative[33]. Ils sont étaient proposés et synthétisés comme agents cytotoxiques potentiels contre le carcinome hépatocellulaire et peuvent être considérés comme agents anticancéreux potentiels[34, 35].



**Chapitre II : Relations
Quantitatives Structure –
Activité (QSAR)**

Les premières études de modélisation des relations structure-activité ont commencé à la fin du 19^{ème} siècle, lorsque *Crum-Brown* et *Frazer* ont postulé que l'activité biologique d'une molécule est une fonction de sa constitution chimique[36, 37]. Mais ce n'est qu'au début des années 60 que les travaux de *Corwin Hansch* ont suggéré un modèle mathématique reliant l'activité biologique à la structure chimique[38].

Aujourd'hui, l'utilisation de QSAR n'a cessé de progresser[39,40]. Elle est devenue essentielle en chimie pharmaceutique et pour la composition de médicaments, notamment dans le cas où la disponibilité des échantillons est limitée ou les mesures expérimentales sont dangereuses, longues et chères[41,42]. Sans l'utilisation de grands outils analytiques, les résultats des études QSAR peuvent fournir des informations utiles pour obtenir une meilleure connaissance des structures moléculaires et probablement le mode d'action au niveau moléculaire. Ces informations peuvent être alors utilisées dans la prédiction des activités biologiques de nouveaux composés ainsi que dans la conception de nouvelles structures[43].

1. Principe et processus du modèle QSAR :

Le principe d'une étude QSAR consiste à trouver une relation mathématique reliant de manière quantitative les propriétés moléculaires, appelées descripteurs, avec une propriété ou une activité biologique pour une série de composés chimiques similaires à l'aide des méthodes d'analyses de données. Pour ce faire, différents types de méthodes statistiques peuvent être employées. La forme générale d'un tel modèle est la suivante [44]:

$$\text{Activité} = f(\text{descripteurs})$$

Pour vérifier que cette équation est significative, il faut suivre les conditions suivantes :

- (Nombre de molécules) / (Nombre de descripteurs) > 5
- Les descripteurs ne doivent pas être corrélés entre eux et en pratique, ils seront alors enlevés lors de l'analyse par la Régression linéaire multiple[45].

Pour prédire l'activité biologique par la méthode QSAR on doit passer par un processus qui est présenté par la Figure 7.

Nous devons Collecter et préparer les structures de chaque composé, pour que la valeur calculée des chaque descripteur soit optimal et ça se fait à l'aide des logiciels adaptés.

Les descripteurs sélectionnés seront utilisés pour la construction du modèle QSAR en utilisant l'algorithme génétique (AG), puis évaluer ce modèle par ces paramètres statistiques [46].

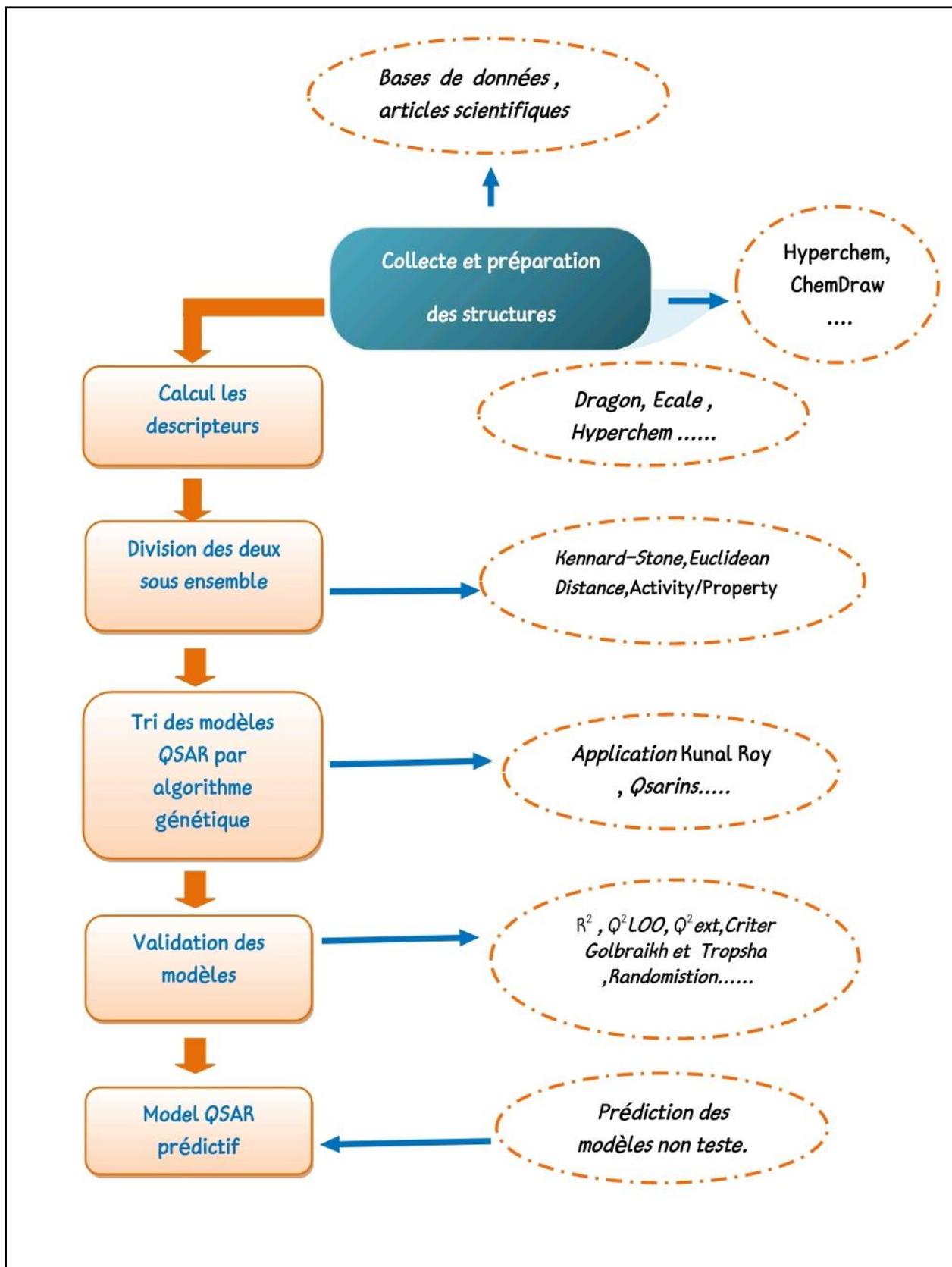


Figure 6:Processus de la construction d'un model QSAR [44].

2. Les essentielles étapes de la construction du modèle QSAR :

En général, pour prédire l'activité biologique par la méthode QSAR, la construction du modèle mathématique est effectuée en 6 étapes suivantes [47] :

- Etape 1 : collecte de données expérimentales.
- Etape 2 : préparation des structures.
- Etape 3 : calcul des descripteurs.
- Etape 4 : division de l'ensemble des données en deux ensembles, un ensemble d'apprentissage (training set) qui contient généralement les 2/3 de la base de données et un ensemble de test (test set) constitué par le 1/3 restant.
- Etape 5 : construction des modèles QSAR.
- Etape 6 : analyse du modèle accepté.

3. Les descripteurs :

Dans une étude QSAR, les descripteurs sont les caractéristiques chimiques d'une molécule sous forme numérique. Il existe différents types de descripteurs en fonction de leur origine (constitutionnel, topologique, géométrique, quantique, thermodynamique...)[48], et en fonction de leur dimensionnalité (1D, 2D, 3D ou 4D)[49].

les aspects 2D des composés (valeurs numériques uniques basées sur les propriétés physico-chimiques des composés)[48], sont calculés uniquement sur la base de la structure 2D, dans laquelle graphique la théorie est largement utilisée à cette fin.

Nous indiquons aussi les descripteurs 1D qui sont remontés directement à partir de la formule brute de la molécule, Exemples : le nombre d'atomes et la masse moléculaire, ...etc. Sans oublier les descripteurs 3D qui nécessitent la géométrie 3D des structures et sont sensibles à la conformation particulière adoptée par une molécule flexible. Des descripteurs physico-chimiques caractérisent les propriétés de la molécule entière en une seule valeur[50].

3.1 Sélection du descripteur :

De nombreux types de descripteurs de structure chimique sont disponibles à partir de logiciels. Obtenir une robustesse du modèle dépend beaucoup de la qualité des descripteurs sélectionnés. Certains de ces outils incluent l'algorithme génétique (GA) dans ses diverses incarnations, qui emploie les règles évolutives de la sélection naturelle pour sélectionner le plus approprié sous-ensemble de descripteurs pour un problème particulier[50].

4. Paramètres statistiques :

4.1 Les paramètres standards :

Afin de déterminer la qualité d'un modèle, différents paramètres statistiques sont employés, tels que les erreurs quadratiques moyennes (*Mean Square Errors*), les coefficients de corrélation qui sont régulièrement utilisés dans les études QSAR, sont décrits dans cette partie[51].

4.1.1 Coefficient de détermination :

Coefficient de détermination R^2 est une mesure de l'ajustement la droite de régression, le R^2 peut être utilisé pour déterminer la relation linéaire entre les valeurs mesurées et estimées[62, 63].

Le coefficient de détermination R^2 est calculé selon la formule suivante [53] :

$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (\text{eq 1})$$

\bar{y} : La valeur moyenne des valeurs observées pour l' ensemble d' apprentissage.

\hat{y} : La valeur prédite.

La racine de l'écart quadratique moyen de calcul (RMSE) (équation 2) :

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (\text{eq 2})$$

n : Le nombre total de composés.

4.1.2 Le Coefficient de détermination Ajusté noté $R_{\text{ajusté}}^2$:

Ce coefficient tient compte du nombre de paramètres du modèle. Le coefficient $R_{\text{ajusté}}^2$ est calculé selon la formule suivante (équation 3):

$$R_{\text{ajusté}}^2 = R^2 - \frac{k(1-R^2)}{n-k-1} = 1 - \frac{(1-R^2)(n-1)}{n-k-1} \quad (\text{eq 3})$$

K : nombre de variables utilisées.

Un bon ajustement correspondra à un R^2 proche de l' unité. (Proche de 1)

4.1.3 Le test de Fisher ou F :

L'indice de Fisher F est couramment employé pour mesurer le niveau de signification statistique du modèle, c'est-à-dire la qualité du choix des descripteurs constituant le modèle. La statistique F est lié au coefficient de détermination par la relation suivante :

$$F = \frac{R*(N-K-1)}{(1-R^2)*K} \quad (\text{eq 4})$$

Plus la valeur de F est grande, plus la probabilité pour que la présence de corrélation entre l'ensemble des descripteurs et l'activité biologique est grande.

4.1.4 La déviation standard S :

La déviation standard S est un autre indicateur statistique utilisé déterminé selon la formule suivante:

$$S = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n-p-1}} \quad (\text{eq 5})$$

Où **n** est le nombre de données d'apprentissage et **p** le nombre de descripteurs utilisés dans la construction du modèle.

En particulier, plus la déviation standard est petite et plus la corrélation est meilleure [44,54–56].

4.2 Les méthodes de sélection De l'ensemble d'apprentissage et de test :

Dans le développement de modèles QSAR dépend non seulement de la méthode statistique mais aussi de l'algorithme utilisé pour la sélection des ensembles d'apprentissage et de test, la sélection de ces deux sous-ensembles est très importante dans l'analyse QSAR.

L'ensemble d'apprentissage est utilisé pour construire et entraîner le modèle et l'ensemble de test est utilisé pour tester l'efficacité du modèle. En pratique l'ensemble d'entraînement est considérablement plus grand, environ 2/3 du nombre total de molécules de l'étude, que l'ensemble de test. et plusieurs différentes méthodes de validation de modèle sont utilisés pour tester sa capacité prédictive [48].

Il existe plusieurs approches possibles pour cela. Dans ce travail, nous décrivons la validation des modèles QSAR pour un ensemble de données basé sur une division par trois méthodes (Kennard Stone, Basé sur la distance euclidienne, Activité/Propriété)

L'activité biologique de l'ensemble de données a été triée de manière ascendante, puis un composé sur quatre a été attribué en tant que composé de l'ensemble de test

4.2.1 La méthode Activité/propriété :

L'activité biologique de l'ensemble de données a été triée de manière ascendante, puis un composé sur quatre a été attribué en tant que composé de l'ensemble de test. (référence On Selection of Training and Test Sets for the Development of Predictive QSAR models)

4.2.2 La méthode de Kennard Stone :

La méthode de Kennard Stone commence par trouver les deux échantillons les plus éloignés en termes de distance géométrique. Pour ajouter un autre échantillon à l'ensemble de sélection, l'algorithme sélectionne parmi les échantillons restants celui qui a la plus grande distance de séparation à partir des échantillons sélectionnés le processus est répété jusqu'à ce que le nombre requis d'échantillons [57, 58].

4.2.3 La méthode Basé sur la distance euclidienne :

La distance euclidienne est un moyen intuitif et habituel pour mesurer la distance entre deux vecteurs de nombres [71, 73]. Soient *x* et *y* deux vecteurs, pour *n* échantillons chaque

modèle correspond à une séquence d'attente $\mathbf{E} = \{E(x_1^k), E(x_2^k), \dots, E(x_n^k)\}$ qui représente les résultats prédictifs[60], le carré de la distance euclidienne est simplement :

$$D_{eucl}^2(x, y) = \sum_{i=1}^d (E(x_i^k) - y_i)^2 \quad (\text{eq 6})$$

i est le numéro de l'échantillon ; n est la quantité d'échantillon ; k est le nombre de modèle.

La distance euclidienne est intimement liée à la notion de produit scalaire. Le produit scalaire entre x et y , noté $\langle x, y \rangle$ (ou parfois $x \cdot y$ ou encore $x^T y$ en considérant les vecteurs comme des matrice à une colonne) est défini par :

$$\langle x, y \rangle = \sum_{i=1}^d x_i y_i \quad (\text{eq 7})$$

et il est facile de montrer que la distance euclidienne peut se réécrire :

$$D_{eucl}^2(x, y) = \langle x, x \rangle + \langle y, y \rangle - 2 \langle x, y \rangle \quad (\text{eq 8})$$

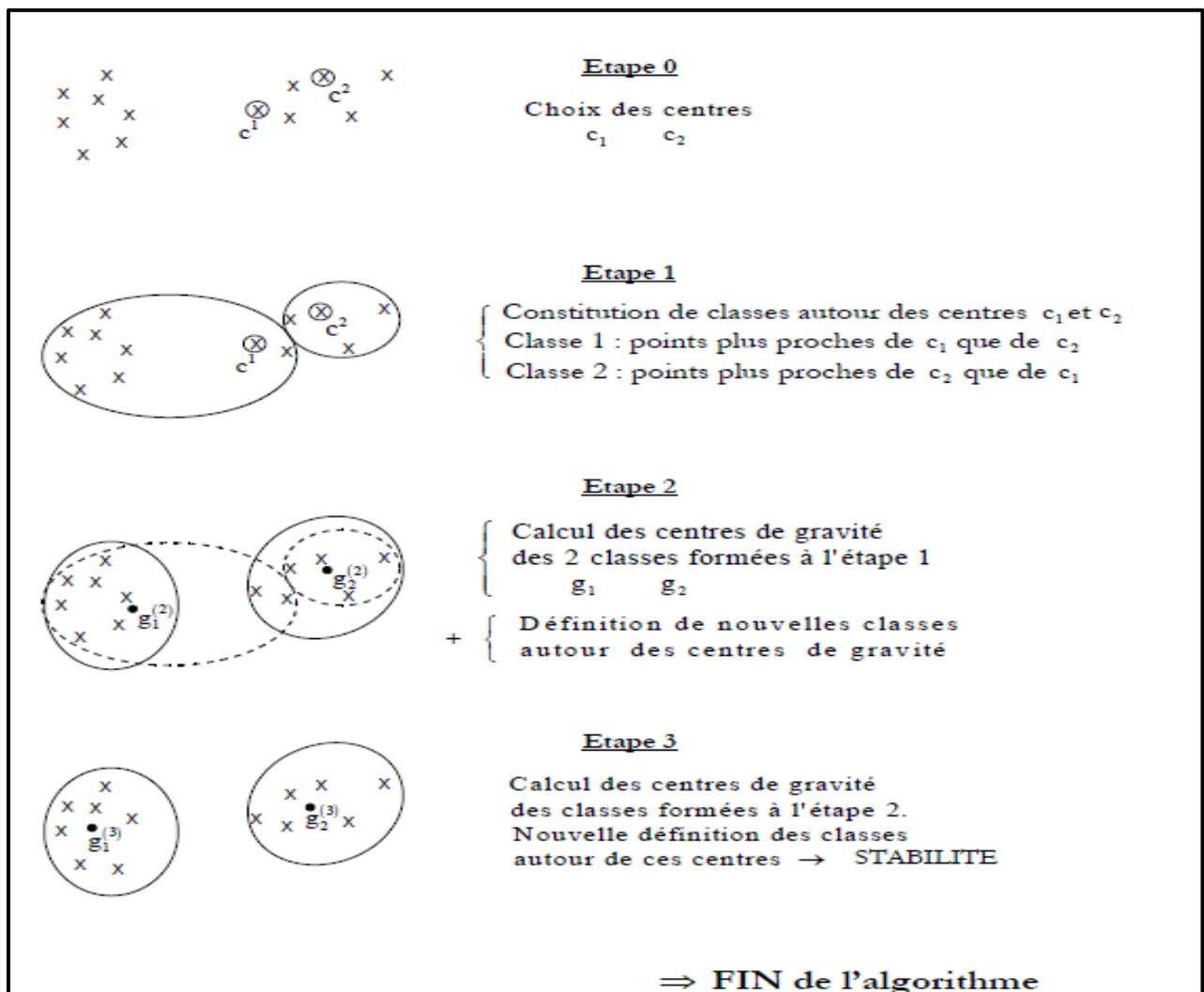


Figure 7:Algorithme La méthode Basé sur la distance euclidienne.

5. Validation :

Quelles que soient les méthodes utilisées pour créer les descripteurs et construire l'équation pour les modèles, il est nécessaire de valider le modèle en comparant les bios activités prédites avec les Bio activités expérimentales[61], en utilisant les données qui créé le modèle (une méthode interne) ou à l'aide d'un ensemble de données distinct (une méthode externe) peut aider à valider le modèle QSAR [62, 63].

5.1 Validation interne :

La validation interne dans les études QSAR est fréquemment utilisée pour étudier la stabilité du modèle et tester l'effet de chaque composant de l'ensemble d'apprentissage sur le modèle final. Il y a des différentes méthodes de validation croisée (CV) les plus couramment utilisés : Q_{LOO}^2 (Leave- One-Out (LOO)), Q_{LMO}^2 (Leave-Many-Out), (Bootstrap)[64].

Dans cette technique un composé est éliminé de l'ensemble de données au hasard à chaque cycle et le modèle est construit en utilisant le reste des composés, le modèle ainsi formé est utilisé pour prédire l'activité du composé éliminé, le processus est répété jusqu'à ce que tous les composés soient éliminés une fois [76,77] . La somme des carrés des erreurs de prédiction désignée par l'acronyme PRESS (La somme résiduelle prédictive des carrés) est utilisé pour déterminer le coefficient de prédiction (Q_{LOO}^2) par les équations suivants [48] :

$$Q_{LOO}^2 = 1 - \frac{PRESS}{\sum_{i=1}^n (Y_i - \bar{Y}_{obs})^2} \quad (\text{eq 9})$$

$$PRESS = \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (\text{eq 10})$$

\hat{Y}_i : La ou les valeurs des données non utilisées pour construire le modèle de CV.

Y_{obs} : La moyenne de l'activité observée de l'ensemble d'apprentissage.

- La métrique r_m^2 pour la validation interne :

Pour éviter le problème que les données d'activité/ propriété prévues se situent dans une proximité étroite avec celles observées et pour mieux indiquer la prévisibilité du modèle on peut exploiter les métriques r_m^2 [79, 80] elles sont définies par les équations suivantes :

$$\overline{r_m^2} = \frac{(r_m^2 + r_m'^2)}{2} \quad (\text{eq 11})$$

$$\Delta r_m^2 = |r_m^2 - r_m'^2| \quad (\text{eq 12})$$

5.2 Validation externe :

La validation externe est le seul moyen d'établir un modèle QSAR fiable. Une validation externe est utilisée pour déterminer la capacité prédictive du modèle développé pour la prédiction des valeurs d'activité de l'ensemble de test et le calcul de la valeur prédictive R_{pred}^2 (Q_{ext}^2) donnée par l'expression (eq 13) [48, 66]:

$$R_{pred}^2 = 1 - \frac{\sum(Y_{pred(TEST)} - Y_{(TEST)})^2}{\sum(Y_{(TEST)} - \bar{Y}_{(TRAINING)})^2} \quad (\text{eq 13})$$

$Y_{pred(TEST)}$ et $Y_{(TEST)}$: Prédit et observe les valeurs d'activité respectivement des composés d'essai.

5.2.1 Critères de Golbraikh et Tropsha :

Golbraikh et Tropsha (2002) ont proposé un ensemble de paramètres pour déterminer la prédiction externe du modèle QSAR. Selon ces auteurs, les modèles sont considérés comme satisfaisants, si l'ensemble des conditions suivantes sont simultanément réalisées[64, 69, 70] :

$$Q_{training}^2 > 0.5 \quad (\text{eq 14})$$

$$R_{test}^2 > 0.6 \quad (\text{eq 15})$$

$$\frac{r^2 - r_0^2}{r^2} < 0.1 \text{ et } 0.85 \leq K \leq 1.15 \quad \text{ou}$$

$$\frac{r^2 - r_0'^2}{r^2} < 0.1 \text{ et } 0.85 \leq K' \leq 0.15 \quad (\text{eq 16})$$

$$|r_0^2 - r_0'^2| < 0.3 \quad (\text{eq 17})$$

5.3 Test de randomisation :

Pour assurer que de bonnes corrélations entre les descripteurs et l'activité présentés dans le modèle QSAR ne sont pas dus au hasard, il existe un test de confirmation c'est le test de randomisation. Pour se faire, les observations sont aléatoirement désorganisées, en changeant itérativement et aléatoirement les valeurs des activités, mais les colonnes des descripteurs restent inchangées. Si la randomisation des observations conduit à des modèles de prévisions élevés cela signifie que les capacités prédictives du modèle QSAR construit sont dues aux corrélations de chance[71, 72].

5.4 Domaine D'application :

Le domaine d'application a été discuté à l'aide du diagramme de Williams traité en détail dans [69, 72], représentant les résidus de prédiction standardisés en fonction des valeurs des leviers h_i . La valeur critique du levier (h^*) est fixée à $(3p+1)/n$, où p est le nombre des descripteurs dans le modèle et n le nombre des molécules dans l'ensemble d'apprentissage[66].

Le domaine d'application est une région théorique de l'espace chimique, définie par le descripteur du modèle et de la réponse modélisée et donc par la nature de la molécule d'ensemble d'entraînement[73].



Partie II : Etude Expérimentale

1. Méthodologie

1.1 Collecte des données :

L'homogénéité des données est essentielle pour l'activité, elle doit être mesurée par un seul et même protocole et dans les mêmes conditions expérimentales. Un ensemble de données de 59 composés de dérivés de salicylanilide a été choisi selon la littérature avec leurs activités [74, 75]. Nous avons étudié l'activité inhibitrice de salicylanilide représenté par le cologarithme décimal d' IC_{50} (pIC_{50}) de 59 dérivés, qui inhibe la tyrosine kinase [74-77].

Les structures des composés sont présentées par la figure 9 et tableau 1.

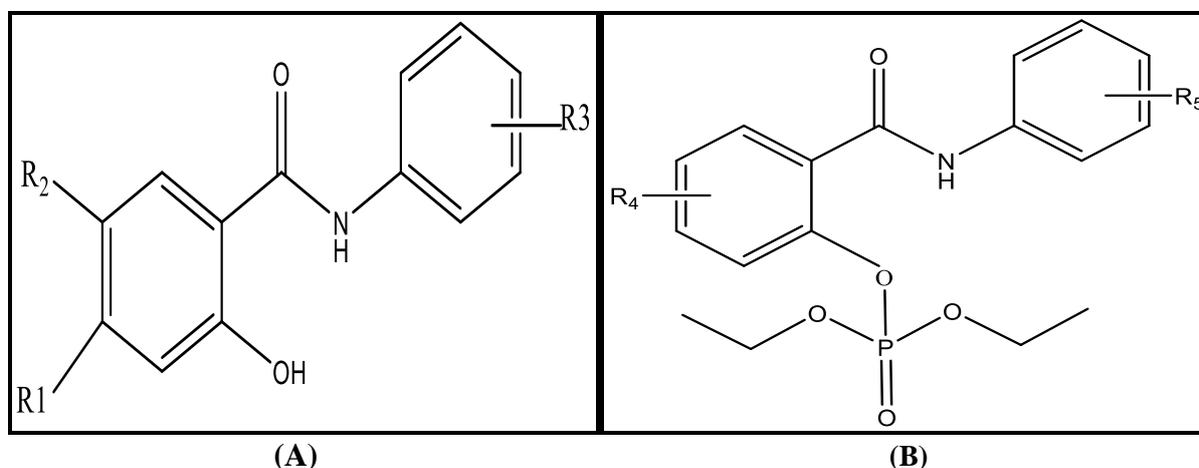


Figure 8: Structure dérivés de salicylanilide.

Tableau 1: L'activité inhibitrice de 32 dérivés de salicylanilide (A).

N°	CODE	R ₁	R ₂	R ₃	IC ₅₀ [μM]
1	1a'	Cl	H	4-Br	3.24
2	1b'	Cl	H	3,4-diCl	2.23
3	1c'	Cl	H	4-CF ₃	2.64
4	1d'	Cl	H	3,5-diCF ₃	9.55
5	2a'	Br	H	4-Br	7.59

6	2b'	Br	H	3,4-diCl	17.67
7	2c'	Br	H	4-CF ₃	15.04
8	2d'	Br	H	3,5-diCF ₃	8.08
9	3a'	H	Cl	4-Br	3.14
10	3b'	H	Cl	3,4-diCl	1.99
11	3c'	H	Cl	4-CF ₃	1.89
12	3d'	H	Cl	3,5-diCF ₃	10.19
13	4a'	H	Br	4-Br	23.6
14	4b'	H	Br	3,4-diCl	31.47
15	4c'	H	Br	4-CF ₃	1.81
16	4d'	H	Br	3,5-diCF ₃	7.8
17	5a'	Br	Cl	4-Br	9.78
18	5b'	Br	Cl	3,4-diCl	11.07
19	5c'	Br	Cl	4-CF ₃	6.52
20	5d'	Br	Cl	3,5-diCF ₃	4.8
21	6a'	Br	Br	4-Br	17.88
22	6b'	Br	Br	3,4-diCl	7.96
23	6c'	Br	Br	4-CF ₃	6.25
24	6d'	Br	Br	3,5-diCF ₃	3
25	7a'	Cl	Cl	4-Br	21.41

26	7b'	Cl	Cl	3,4-diCl	12.42
27	7c'	Cl	Cl	4-CF ₃	9.87
28	7d'	Cl	Cl	3,5-diCF ₃	4.86
29	8a'	Cl	Br	4-Br	10.83
30	8b'	Cl	Br	3,4-diCl	7.65
31	8c'	Cl	Br	4-CF ₃	11.03
32	8d'	Cl	Br	3,5-diCF ₃	5.07

Tableau 2:L'activité inhibitrice de 27 dérivés de salicylanilide (B).

N°	CODE	R ₄	R ₅	IC ₅₀ [μ M]
33	1a	4-Br	3-F	5.54
34	1b	4-Br	3-Cl	6.52
35	1c	4-Br	4-Cl	6.81
36	1d	4-Br	3,4-di-Cl	2.04
37	1e	4-Br	3-CF ₃	1.84
38	1f	4-Br	4-CF ₃	1.84
39	1g	4-Br	3-Br	4.41
40	1h	4-Br	4-Br	5.6
41	1i	4-Br	4-F	7.4
42	1j	4-Cl	3-Cl	2.68
43	1k	4-Cl	4-Cl	12.96

44	1l	4-Cl	3-F	17.17
45	1m	4-Cl	4-F	31.45
46	1n	4-Cl	4-Br	1.56
47	1o	4-Cl	3,4-di-Cl	2.84
48	1p	4-Cl	3-Br	1.81
49	1q	4-Cl	3-CF ₃	5.68
50	1r	4-Cl	4-CF ₃	2.33
51	1s	5-Cl	3-Cl	9.57
52	1t	5-Cl	3-Br	8.07
53	1u	5-Cl	3-F	11.17
54	1v	5-Cl	4-F	33.82
55	1w	5-Cl	4-Br	2.67
56	1x	5-Cl	4-Cl	5.6
57	1y	5-Cl	3,4-di-Cl	1.92
58	1z	5-Cl	3-CF ₃	3.85
59	1zz	5-Cl	4-CF ₃	2.51

1.2 Traitements des données :

Le traitement des données pour l'étude QSAR passe par trois étapes :

- Etape 1 Préparation de l'ensemble de données.
- Etape 2 : Calcule des descripteurs.
- Etape 3 : La modélisation

1.2.1 La Préparation de l'ensemble de données :

Nous avons utilisé deux logiciels :

- **ChemDraw 18.0** [78] : ce logiciel est utilisé pour le dessin des structures des molécules étudiés dans ce manuscrit.
- **HyperChem 08** [79]: ce logiciel est utilisé pour l'optimisation des molécules.

Tous les calculs d'optimisation des structures des molécules ont été effectués en utilisant une pré-optimisation avec le Mécanique moléculaire MM+ puis une optimisation avec les méthodes semi-empirique PM3 en appliquant un gradient limite standard de 0,001 Å kcal.mol⁻¹ comme critère d'arrêt.

1.2.2 Le calcul des descripteurs :

Nous avons utilisé le logiciel **DRAGON V5.5** [80] : Les géométries ainsi optimisées ont été transférées dans le logiciel Dragon version 5.5 qui calcul de plus de 1664 descripteurs.

1.2.3 La modélisation :

La modélisation consiste à construire des modèles mathématiques, afin de prévoir et de mieux en comprendre la relation entre la structure moléculaire et l'activité biologique. Pour formuler des prédictions exactes de l'activité pour de nouvelles structures qui n'étaient pas visibles précédemment et une évaluation précise des performances du modèle, une solution consiste à diviser l'ensemble de données en deux sous-ensembles :

- **Ensemble d'apprentissage** : sous-ensemble destiné à l'apprentissage d'un modèle.
- **Ensemble de Test** : sous-ensemble destiné à l'évaluation du modèle

a) Répartition des sous-ensembles (entraînement & test)

Pour la division de l'ensemble de données nous avons comparé entre trois méthodes de division. Nous avons utilisé l'application **Dataset division GUI 1.2** : C'est une application qui comprend trois méthodes différentes de la division (Kennard-Stone, basée sur la distance euclidienne et basée sur l'activité/la propriété).

b) Construction du model QSAR :

Pour le développement du model QSAR via l'application **MLRPlusValidation1.3**, (application de Roy) [81] Cet outil construit des modèles QSAR en effectuant une MLR et calcule les paramètres de validation internes et externes des modèles trouvés. Il juge en outre la qualité des prédictions de l'ensemble de test sur la base des erreurs de prédiction. Il vérifie également les critères d'acceptabilité des modèles de Golbarikh et Tropsha[81].a l'aide de l'algorithme génétique (AG) .

2. Résultats et discussion :

2.1 Répartition des sous-ensembles (entraînement & test) :

L'ensemble des données a été divisé selon les trois méthodes de division suivantes :

- ♣ Kennard-Stone ;
- ♣ Basée sur la distance euclidienne ;
- ♣ Basé Activité/Propriété,

Et chaque une de ces méthodes a réparties l'ensemble en deux séries, une série d'entraînement et une série de test. Le Tableau 3 montre les résultats de la répartition de l'ensemble de test pour les trois méthodes.

**Tableau 3:repartition de l'ensemble de tests pour les trois méthodes de divisions
Kennard-Stone, Euclidean Distance et Activity/Property.**

N	Kennard-Stone	la distance euclidienne	Activité/Propriété
1	4c'	1b'	1d'
2	4d'	2d'	3b'
3	6a'	3c'	3d'
4	6c'	4a'	4a'
5	7b	5a'	4d'
6	7c'	5b'	8d'
7	1b	6d'	1c
8	1c	8c'	1f
9	1g	1g	1l
10	1m	1j	1o
11	1v	1n	1q
12	1y	1o	1u
13		1p	1z
14		1u	1zz

D'après les résultats du Tableau 3, on remarque que les trois algorithmes des trois méthodes utilisées pour la division de l'ensemble des molécules ont données, différent nombre de molécules pour l'ensemble de Test :

Kennard-Stone a choisi 12 molécules pour le test, et les deux algorithmes (distance euclidienne et Activité/Propriété), ont choisi 14 molécules pour le test.

Et aussi que les molécules choisies pour le test ne sont pas les mêmes dans les trois méthodes. Pour le choix entre les trois algorithmes des trois méthodes étudiées dans ce manuscrit pour la construction du modèle QSAR.

L'algorithme génétique implanté dans l'application **MLRPlusValidation1.3** a donné un model établie en basant sur 4 descripteurs (MATS2v², ESpm12d², G1m², H5u².) pour les

trois divisions, le Tableau 4 montre les trois modèles pour trois algorithmes et le Tableau 5 montre les critères de choix.

Tableau 4: les trois modèles pour les trois algorithmes par les 3 méthodes de division.

<i>Les methods</i>	Equation
<i>Kennard-Stone</i>	$pIC_{50} = -3,26726(+/-0,45957) + 54,70919(+/-8,55849) MATS2v^2 + 0,00489(+/-0,00105) ESpm12d^2 + 14,79628(+/-9,39412) G1m^2 + 1,05908(+/-0,19758) H5u^2$
<i>la distance euclidienne</i>	$pIC_{50} = -3,81596(+/-0,43422) + 55,22916(+/-8,29596) MATS2v^2 + 0,00577(+/-0,00105) ESpm12d^2 + 23,52225(+/-9,18602) G1m^2 + 1,07853(+/-0,20274) H5u^2$
<i>Activité/Propriété</i>	$pIC_{50} = -3,71315(+/-0,54341) + 59,80287(+/-9,36056) MATS2v^2 + 0,00588(+/-0,00123) ESpm12d^2 + 19,18589(+/-11,52495) G1m^2 + 1,05578(+/-0,29918) H5u^2$

Tableau 5: les critères de choix entre les trois méthodes de division

	Les paramètres standards de Roy		Les paramètres de Golbreich et Tropsha		
	EAM	DS	$Q_{training}^2$	R_{test}^2	$ r_0^2 - r_0'^2 $
<i>Kennard-Stone (KS)</i>	0.146	0.140	0,546	0,710	0,229
<i>la distance euclidienne (DE)</i>	0.170	0.150	0,595	0,593	0,535
<i>Activité/ Propriété (A/P)</i>	0.068	0.058	0,507	0,834	0,010

Les modes de quatre dimensions pour les trois méthodes de division sont basés sur quatre descripteurs ((MATS2v², ESpm12d², G1m², H5u².) (Voir Tableau 4)

Pour le choix de la meilleure méthode de division les conditions de Roy et les Critères de Golbreich et Tropsha doivent être vérifiés. D'une notion générale, une Erreur absolue moyenne (EAM) de 10% devrait être acceptable alors qu'une valeur de EAM supérieure à 15% doit être une erreur très élevée et les prédictions qui ne relèvent pas de l'une des deux conditions peuvent être considérées comme de qualité moyenne, l'EAM et le DS (Déviation Standard) déterminent la qualité du modèle selon Roy [82, 83], d'après le tableau 5 les valeurs de l'EAM et la DS des 3 algorithmes:

♣ *Algorithme KS :*

$$EAM_{KS} = 14.6\% > 10\% \text{ et } SD_{KS} = 14\% > 10\% \quad \textbf{Inacceptable.}$$

♣ *Algorithme ED*

$$EAM_{ED} = 17\% > 10\% \text{ et } SD_{ED} = 15\% > 10\% \quad \textbf{Inacceptable.}$$

♣ *Algorithme A/P*

$$EAM_{A/P} = 6.8\% < 10\% \text{ et } SD_{A/P} = 5.8\% < 10\% \quad \textbf{Acceptable.}$$

Selon les paramètres de Roy La meilleure division pour cet ensemble de données c'est la division de la méthode Activité/Propriété.

Selon Les critères Golbreich et Tropsha le bon modèle devrait avoir un r^2 supérieur à 0.6 ($r^2 > 0.6$), l'intervalle de $|r_0^2 - r_0'^2|$ inférieur a 0.3 et Q^2 doit être supérieur à 0.5 ($Q^2 > 0.5$) pour accepte le modèle :

♣ *Algorithme KS :*

$$r_{KS}^2 0.71 > 0.6 \text{ et } |r_0^2 - r_0'^2| 0.229 < 0.3 \quad \textbf{acceptable}$$

♣ *Algorithme ED*

$$r_{DE}^2 0.593 < 0.6 \text{ et } |r_0^2 - r_0'^2| 0.535 > 0.3 \quad \textbf{inacceptable}$$

♣ *Algorithme A/P*

$$r_{A/P}^2 0.834 > 0.6 \text{ et } |r_0^2 - r_0'^2| 0.01 < 0.3 \quad \textbf{acceptable}$$

D'après ces critères deux méthodes de division sont accepté la méthode Activité/Propriété et La méthode de Kennard-Stone.

2.1.1 Conclusion :

Le choix de la meilleure méthode de division repose sur la vérification des conditions de Roy et aussi de Golbreich et Tropsha simultanément. Donc, la meilleure méthode pour la division de notre ensemble de données, c'est la méthode d'Activité/Propriété car elle vérifie toutes les conditions de Roy et les Critères de Golbreich et Tropsha ($MAE_{A/P} 0.068 < 0.1$ et $SD_{A/P} 0.058 < 0.1$, $r_{A/P}^2 0.834 > 0.6$ et $|r_0^2 - r_0'^2| 0.01 < 0.3$).

2.2 Analyse de modèle accepté :

Le model QSAR a été déterminée en prenant pour variable dépendante le logarithme de l'inverse de IC₅₀ (en mmol / litre), et pour variables explicatives quatre descripteurs (MATS2v², ESpm12d², G1m², H5u².)

Les données ont été modelées en utilisant la régression linéaire multiple à l'aide de l'algorithme génétique implanté dans l'application **MLRPlusValidation1.3**.

Les valeurs des données utilisées dans la suite de ce travail sont condensées dans le Tableau 1.

2.2.1 Evaluation du modèle :

2.2.1.1 La colinéarité :

La colinéarité est un problème majeur dans l'interprétation des résultats, Car Les descripteurs ne doivent pas être corrélés entre eux. Si le coefficient de corrélation r est supérieur à 0.97, les descripteurs seront considérés comme fortement corrélés[45] ,en convenable et ils seront enlevés. La matrice de corrélation entre les pIC₅₀ et tous les descripteurs est montrée par le Tableau 6.

Tableau 6: Matrice de corrélation entre pIC50 et les descripteurs impliqués dans le modèle QSAR.

	MATS2v ²	ESpm12d ²	G1m ²	H5u ²	PIC50
MATS2v ²	1	-0.575	-0.001	-0.247	0.346
ESpm12d ²	-0.575	1	-0.121	0.333	0.273
G1m ²	-0.001	-0.121	1	-0.476	-0.087
H5u ²	-0.247	0.333	-0.476	1	0.348
PIC50	0.346	0.273	-0.087	0.348	1

Cette matrice montre l'absence de la corrélation entre les descripteurs eux-mêmes.

2.2.1.2 Analyse de la régression :

L'activité inhibitrice dans ce travail est représentée par le cologarithme décimal d'IC₅₀ (pIC₅₀) de 59 dérivés de salicylanilide. La comparaison entre les méthodes de division nous a donné la meilleure méthode est la méthode Activité/Propriété, le modèle sélectionné à quatre variables explicatives (MATS2v², ESpm12d², G1m², H5u²) dont l'équation est :

$$pIC_{50} = -3,71315(+/-0,54341) + 59,80287(+/-9,36056) \text{ MATS2v}^2 + 0,00588(+/-0,00123) \text{ ESpm12d}^2 + 19,18589(+/-11,52495) \text{ G1m}^2 + 1,05578(+/-0,29918) \text{ H5u}^2 \quad (\text{eq1})$$

$$n = 59 \quad R^2 = 58.2 \% \quad R_{adj}^2 = 54.02\% \quad F = 13.92 \quad N_{ent} = 45 \quad N_{test} = 14$$

2.2.1.3 Qualité d'ajustement :

La valeur du test de Fisher $F = 13,92$ et les valeurs de $R^2 = 58.2\%$ et de $R_{adj}^2 = 54.02\%$ montrent la bonne qualité de l'ajustement, et la valeur du coefficient de détermination R^2 signifie que 58.2% de la variabilité de pIC_{50} est expliquée par ces quatre descripteurs.

Le Tableau 7 et la Figure 9 représente les données de l'activité expérimentale et leur prédiction donnée par l'équation (eq.1) pour l'entraînement et le test.

La dispersion des points des valeurs prédites de l'activité en fonction de celles mesurées autour de la droite (figure 9), montre que les valeurs prédites sont en adéquation avec les valeurs expérimentales, caractéristique d'un bon ajustement.

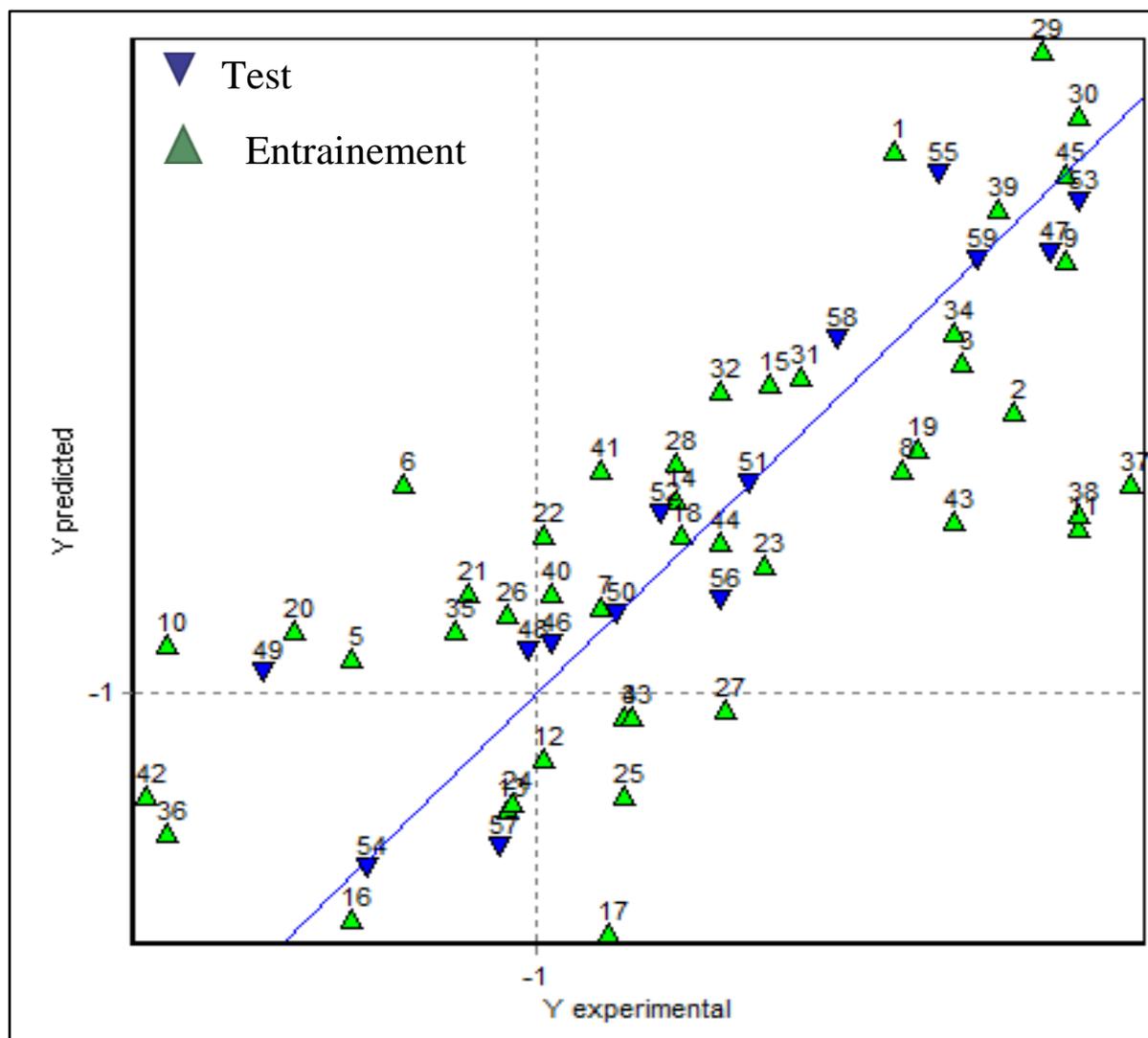


Figure 9: La droite de régression des valeurs expérimentales et prédites de la notation pIC_{50} pour le modèle de QSAR.

2.2.1.4 Validation :

L'ensemble des données a été divisé en deux sous-ensemble : un sous-ensemble d'entraînement (80% =45 molécules) pour la construction du model, et un sous-ensemble de Test (20% =14 molécules) pour la validation du model.

Tableau 7: Les valeurs expérimentales prédit et calculé pour pIC50 pour les 59 dérivés de salicylanilide.

N°	Code	Ensemble	pIC ₅₀ Exp	pIC ₅₀ Calc	pIC ₅₀ Pred	Err.Calc	Err.Pred
1	1a'	Entrainement	-0.51	-0.42	-0.25	0.09	0.26
2	1b'	Entrainement	-0.35	-0.58	-0.61	-0.23	-0.26
3	1c'	Entrainement	-0.42	-0.53	-0.54	-0.11	-0.12
4	1d'	Test	-0.98	-	-0.93	-	0.05
5	2a'	Entrainement	-0.88	-1.02	-1.03	-0.14	-0.15
6	2b'	Entrainement	-1.25	-0.97	-0.95	0.28	0.3
7	2c'	Entrainement	-1.18	-0.73	-0.71	0.44	0.47
8	2d'	Entrainement	-0.91	-0.89	-0.88	0.02	0.03
9	3a'	Entrainement	-0.5	-0.68	-0.69	-0.18	-0.19
10	3b'	Test	-0.3	-	-0.39	-	-0.1
11	3c'	Entrainement	-0.28	-0.39	-0.4	-0.11	-0.12
12	3d'	Test	-1.01	-	-0.94	-	0.07
13	4a'	Test	-1.37	-	-0.97	-	0.41
14	4b'	Entrainement	-1.5	-0.95	-0.93	0.54 *	0.57 *
15	4c'	Entrainement	-0.26	-0.74	-0.77	-0.48 *	-0.51
16	4d'	Test	-0.89	-	-0.89	-	0.01

17	5a'	Entrainement	-0.99	-1.08	-1.09	-0.09	-0.1
18	5b'	Entrainement	-1.04	-1.15	-1.16	-0.11	-0.11
19	5c'	Entrainement	-0.81	-0.74	-0.73	0.08	0.08
20	5d'	Entrainement	-0.68	-0.59	-0.57	0.09	0.11
21	6a'	Entrainement	-1.25	-1.3	-1.31	-0.05	-0.05
22	6b'	Entrainement	-0.9	-1.29	-1.33	-0.39	-0.43
23	6c'	Entrainement	-0.8	-0.78	-0.78	0.02	0.02
24	6d'	Entrainement	-0.48	-0.62	-0.66	-0.15	-0.18
25	7a'	Entrainement	-1.33	-0.92	-0.91	0.41	0.43
26	7b	Entrainement	-1.09	-0.87	-0.86	0.22	0.23
27	7c'	Entrainement	-0.99	-0.79	-0.78	0.2	0.21
28	7d'	Entrainement	-0.69	-0.8	-0.82	-0.11	-0.14
29	8a'	Entrainement	-1.03	-1.15	-1.15	-0.11	-0.12
30	8b'	Entrainement	-0.88	-1.12	-1.14	-0.24	-0.25
31	8c'	Entrainement	-1.04	-0.9	-0.89	0.14	0.15
32	8d'	Test	-0.71	-	-0.71	-	0
33	1a	Entrainement	-0.74	-1	-1.02	-0.26	-0.27
34	1b	Entrainement	-0.81	-0.68	-0.68	0.13	0.14
35	1c	Test	-0.83	-	-0.75	-	0.08
36	1d	Entrainement	-0.31	-0.15	-0.11	0.16	0.2
37	1e	Entrainement	-0.26	-0.21	-0.2	0.06	0.07

38	1f	Test	-0.26	-	-0.32	-	-0.05
39	1g	Entrainement	-0.64	-0.57	-0.56	0.08	0.08
40	1h	Entrainement	-0.75	-0.59	-0.58	0.16	0.17
41	1i	Entrainement	-0.87	-1.02	-1.03	-0.15	-0.16
42	1j	Entrainement	-0.43	-0.48	-0.5	-0.05	-0.07
43	1k	Entrainement	-1.11	-0.92	-0.91	0.19	0.21
44	1l	Test	-1.23	-	-1.24	-	0
45	1m	Entrainement	-1.5	-1.23	-1.19	0.27	0.31
46	1n	Entrainement	-0.19	-0.68	-0.71	-0.49 *	-0.51 *
47	1o	Test	-0.45	-	-0.28	-	0.17
48	1p	Entrainement	-0.26	-0.71	-0.75	-0.46	-0.49
49	1q	Test	-0.75	-	-0.87	-	-0.12
50	1r	Entrainement	-0.37	-0.35	-0.33	0.02	0.04
51	1s	Entrainement	-0.98	-0.87	-0.86	0.11	0.12
52	1t	Entrainement	-0.91	-0.7	-0.69	0.21	0.22
53	1u	Test	-1.05	-	-1.21	-	-0.16
54	1v	Entrainement	-1.53	-1.18	-1.14	0.35	0.39
55	1w	Entrainement	-0.43	-0.74	-0.76	-0.31	-0.34
56	1x	Entrainement	-0.75	-0.79	-0.79	-0.04	-0.05
57	1y	Entrainement	-0.28	-0.28	-0.28	0	0
58	1z	Test	-0.59	-	-0.51	-	0.08
59	1zz	Test	-0.4	-	-0.4	-	0

2.2.1.4.1 Validation interne :

Tableau 8: Les paramètres de la validation interne.

Validation interne	Les paramètres standards de Roy						
	R^2	R_{adj}^2	F	Q_{Loo}^2	PRESS	$\overline{r_m^2}$	Δr_m^2
	58.2	54.02	13.92	50.66	2,513	0,366	0,227

Le coefficient de détermination de la validation croisée du modèle

$Q_{Loo}^2=50.66\%$, témoigne d'une bonne corrélation entre l'activité prédite et l'activité réelle, et reflète une précision du modèle.

Pour mieux indiquer une meilleur prévisibilité du modèle on peut exploiter les métriques $\overline{r_m^2}$ et Δr_m^2 : ils doivent être proches de zéro [84]

$$\overline{r_m^2} = 0.366 \quad \text{Proche à 0.}$$

$$\Delta r_m^2 = 0.227 \quad \text{Proche à 0.}$$

2.2.1.4.2 Validation externe :

Tableau 9: Les paramètres de la validation externe.

Validation externe	Les paramètres standards de Roy			Les paramètres de Golbreich et Tropsha				
	R_{pred}^2	MAE	SD	$Q_{training}^2$	R_{test}^2	$ r_0^2 - r_0'^2 $	K	$[(r^2-r_0'^2)/r^2]$
	83.44	0.068	0.058	0,507	0,834	0,010	1,034	0,003

La validation interne, pour les nouveaux composés ne suffit pas pour juger la capacité prédictive du modèle. Une validation externe doit être réalisée pour une bonne prévision des composés n'ont pas été utilisé lors de l'élaboration du modèle.

Les paramètres standard de Roy ($R_{pred}^2 = 83.44$, MAE =0.068 et SD=0.058) et Les paramètres ($Q_{training}^2 = 0,507$, $R_{test}^2=0.834$ et $|r_0^2 - r_0'^2|=0,010$, $k=1.034$, $[(r^2-r_0'^2)/r^2] = 0.003$) vérifie les conditions de Golbreich et Tropsha :

1. $Q^2 > 0.5$)
2. $r^2 > 0.6$)
3. $|r^2 - r'^2| < 0.3$
4. $0.85 < k < 1.15$
5. $(r^2 - r'^2) / r^2 < 0.1$]

Confirment la bonne capacité prédictive du modèle pour les composés n'ayant pas participé aux calculs du model. Il est à noter que le modèle est à accréditer des meilleures performances.

2.2.1.5 Test de randomisation :

Le test de randomisation (Y- randomisé) permet de vérifier l'existence de corrélations imprévues [85]. Il consiste à mélanger les valeurs de la réponse Y entre elles sans changer la valeur des descripteurs X_i (figure10)[86].

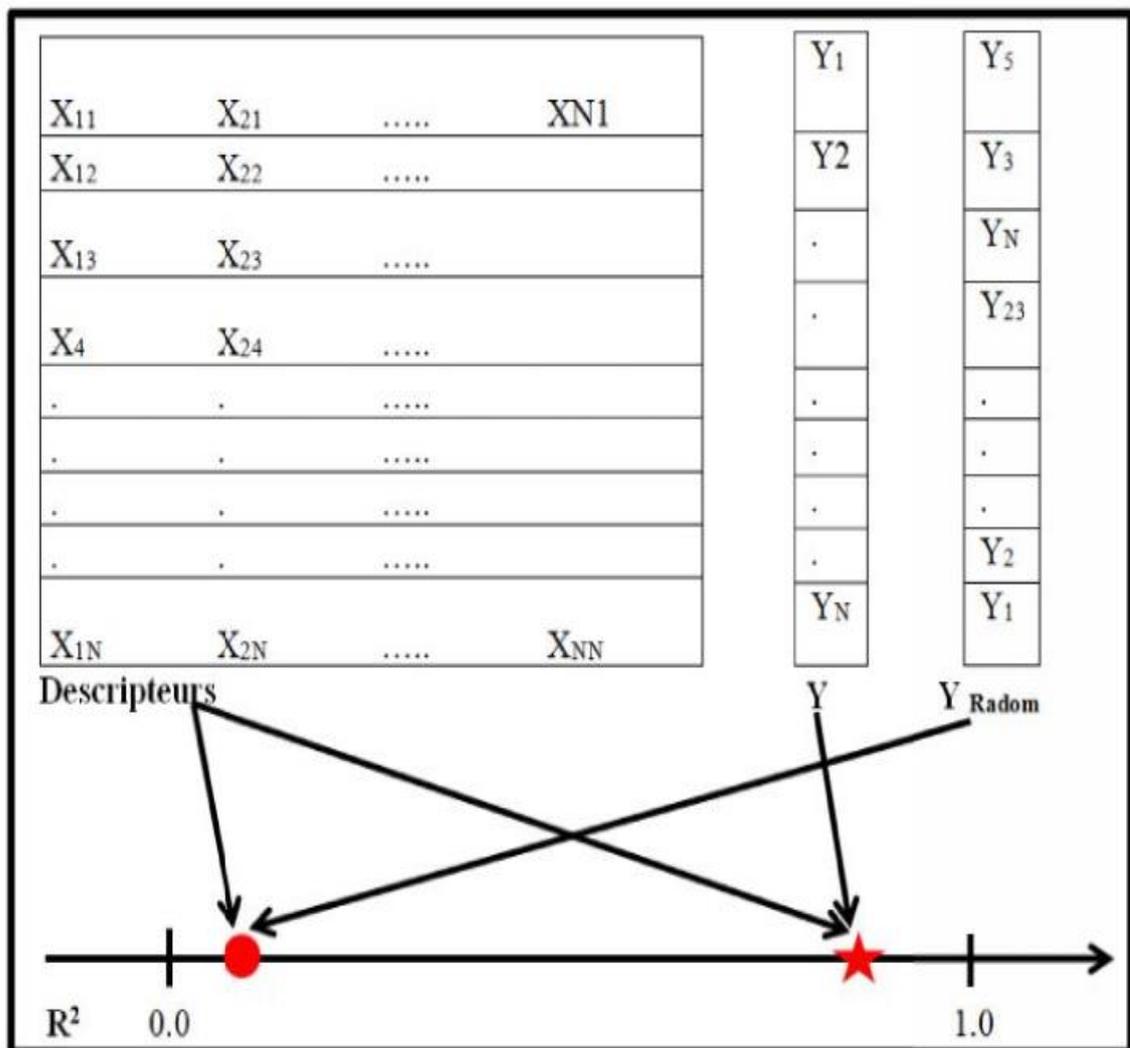


Figure 10: principe du test de randomisation[86].

Ce processus se fait entre 100 et 1000 itérations. Les modèles Randomisés doivent avoir des R^2_{Random} nettement inférieures à R^2 du modèle développé.

Les cercles bleus représentent les modèles mises au hasard, et le cercle rouge correspond au model réel. Le test de randomisation après 500 itérations, montre que notre modèle n'est pas dû au hasard ceci est justifié par la (figure12).

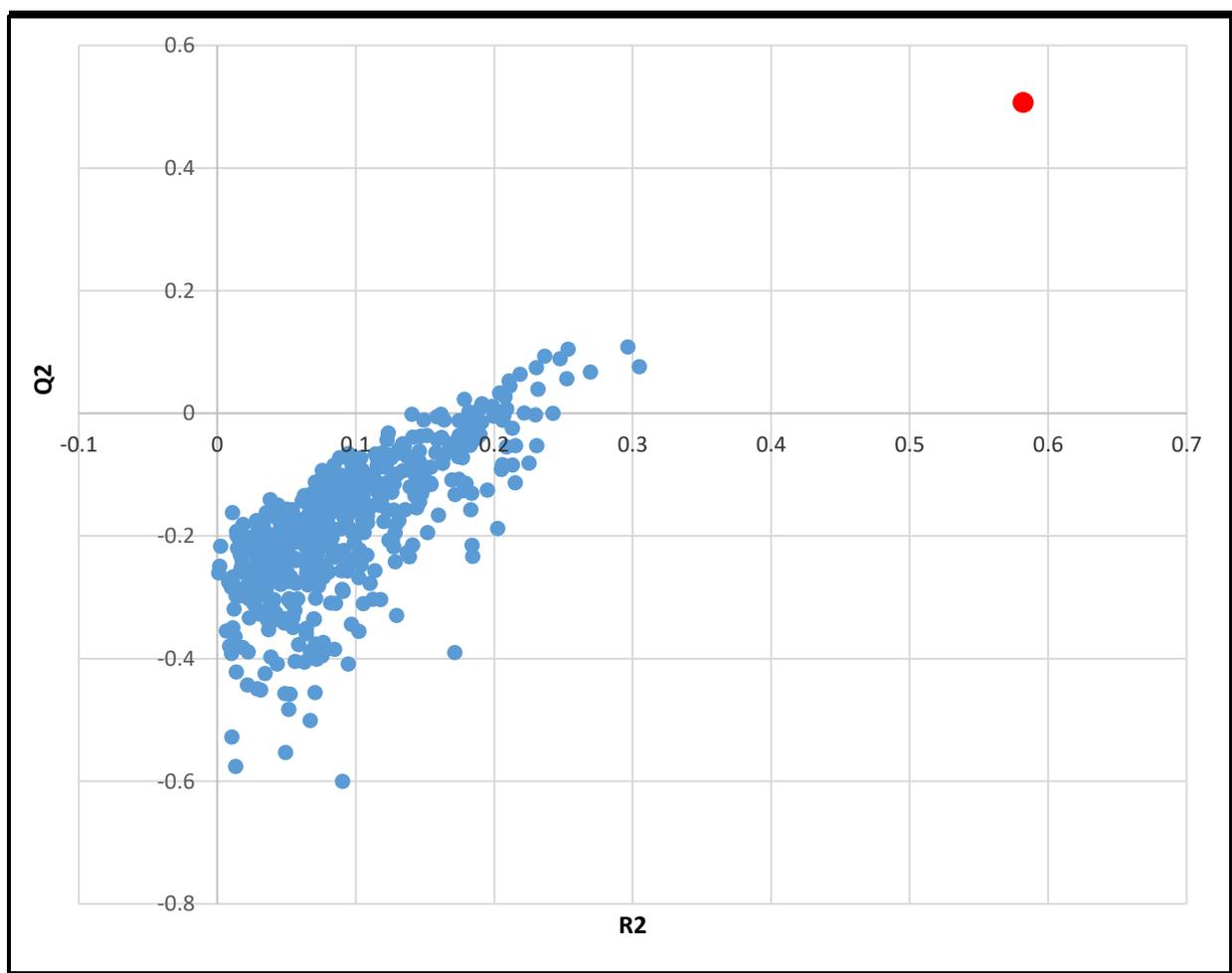


Figure 11: Test de randomisation associé au modèle QSAR 500 itérations.

2.2.1.6 Domaine d'applications :

Le diagramme de williams (voir figure 13) désigne les composés qui se trouvent dans le domaine d'application de notre model. Les valeurs des résidus standardisés sont toutes comprises dans l'intervalle ± 3 donc le modèle ne présente aucun point aberrant. Et les molécules étudiées se trouvent dans cet intervalle. $h^* = (3*(4+1))/45 = 0.333$

Ce model présente deux points levier pour l'ensemble de calibrage les molécules (1a' : 1) et (1r : 50).

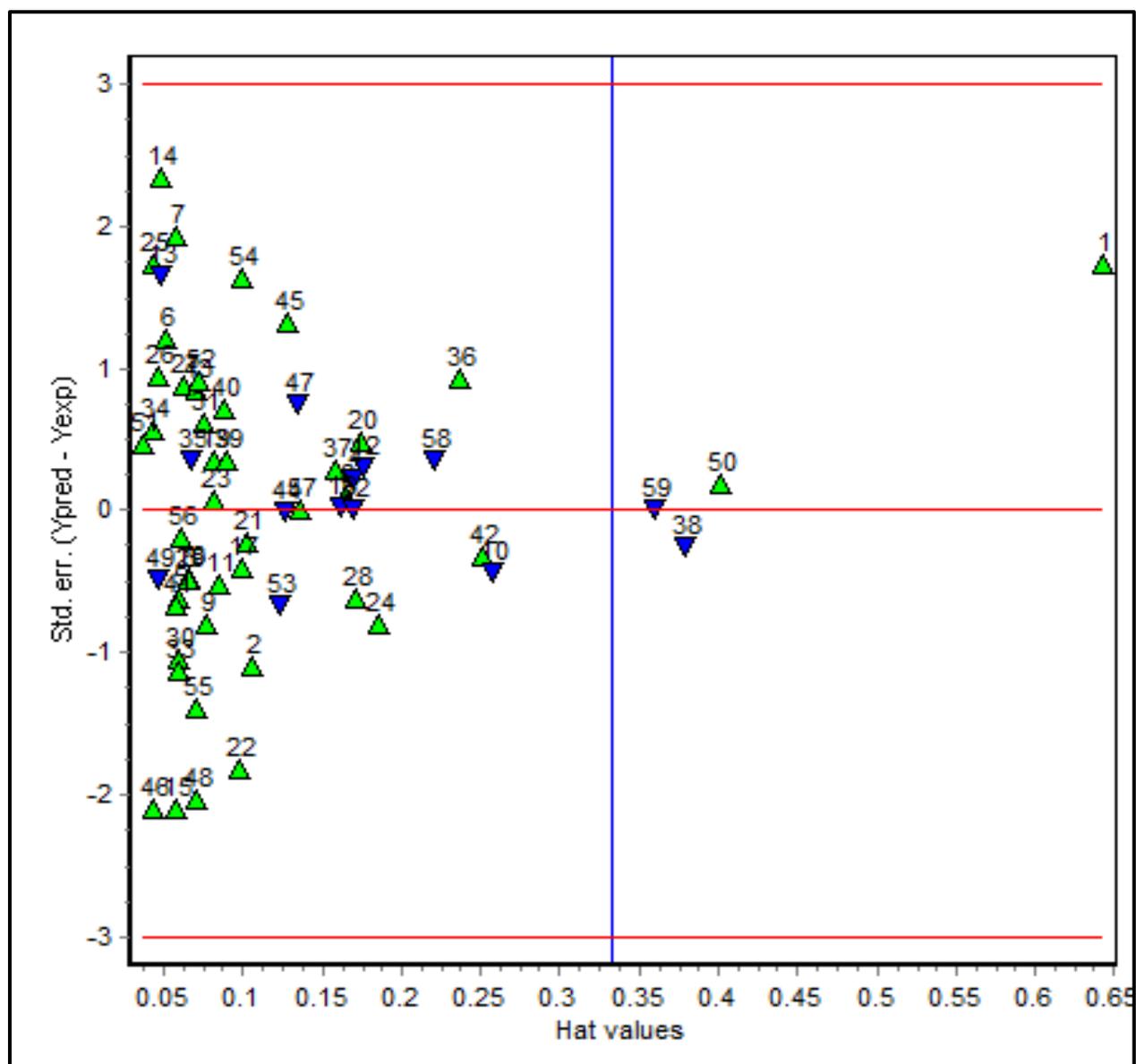
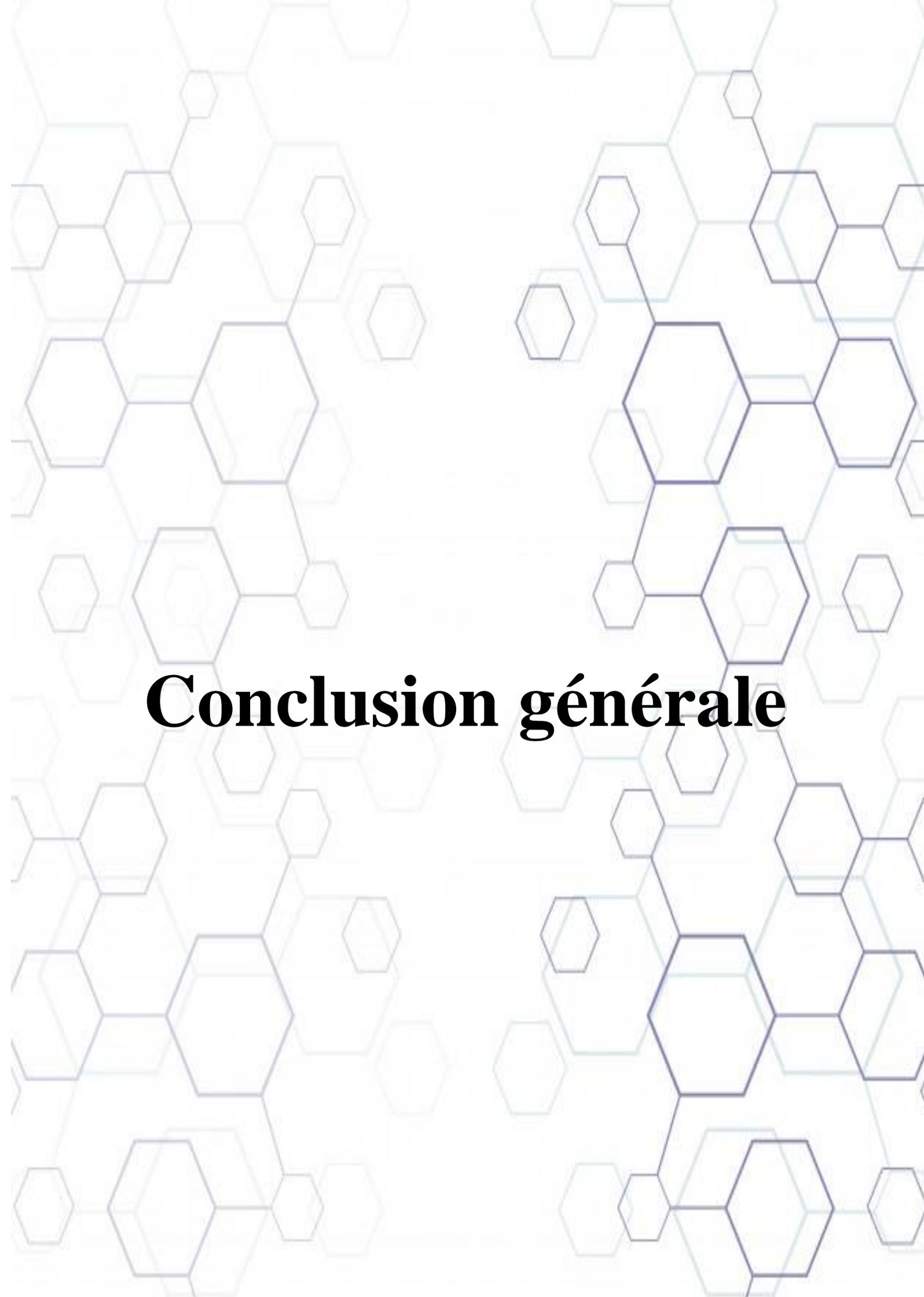


Figure 12: Diagramme de williams des résidus standardisés.



Conclusion générale

Conclusion générale

Le développement de nouveaux médicaments plus efficace et moins chère pour le traitement du carcinome hépatocellulaire (CHC) constitue un enjeu majeur pour l'industrie pharmaceutique.

La famille des salicylanilides est notée comme agents anti-CHC, un modèle QSAR a été construit afin d'expliquer et prédire l'activité anti-cancéreuse d'une série de 59 dérivés de salicylanilide inhibiteurs de tyrosine kinase.

Les données ont été modelées en utilisant l'analyse de la régression linéaire multiple (MLR = Multiple Linear Régression) à l'aide de l'algorithme génétique implanté dans l'application **MLRPlusValidation1.3** qui nous a permis a abouti à une corrélation entre pIC_{50} qui détermine l'activité inhibitrice de tyrosine kinase et quatre descripteurs ; $MATS2v^2$, $ESpm12d^2$, $G1m^2$, $H5u^2$.

L'ensemble des données a été divisé selon les trois méthodes de division suivantes :

- ♣ Kennard-Stone ;
- ♣ Basée sur la distance euclidienne ;
- ♣ Basé Activité/Propriété,

Selon les conditions de Roy et aussi de Golbreich et Tropsha simultanément ($MAE_{A/P} 0.068 < 0.1$ et $SD_{A/P} 0.058 < 0.1$, $r_{A/P}^2 0.834 > 0.6$ et $|r_0^2 - r_0'^2| 0.01 < 0.3$). Nous avons pu adopter la méthode d'Activité/Propriété comme meilleure méthode de division de cet ensemble de données. Les principales techniques de validation internes et externes des modèles ont été utilisés et nous avons vérifié également les critères d'acceptabilité du modèle de Golbarikh et Tropsha.

Le model QSAR a été déterminée en prenant pour variable dépendante le logarithme de l'inverse de IC_{50} (en mmol / litre), et pour variables explicatives quatre descripteurs ($MATS2v^2$, $ESpm12d^2$, $G1m^2$, $H5u^2$.)

Le modèle trouvé :

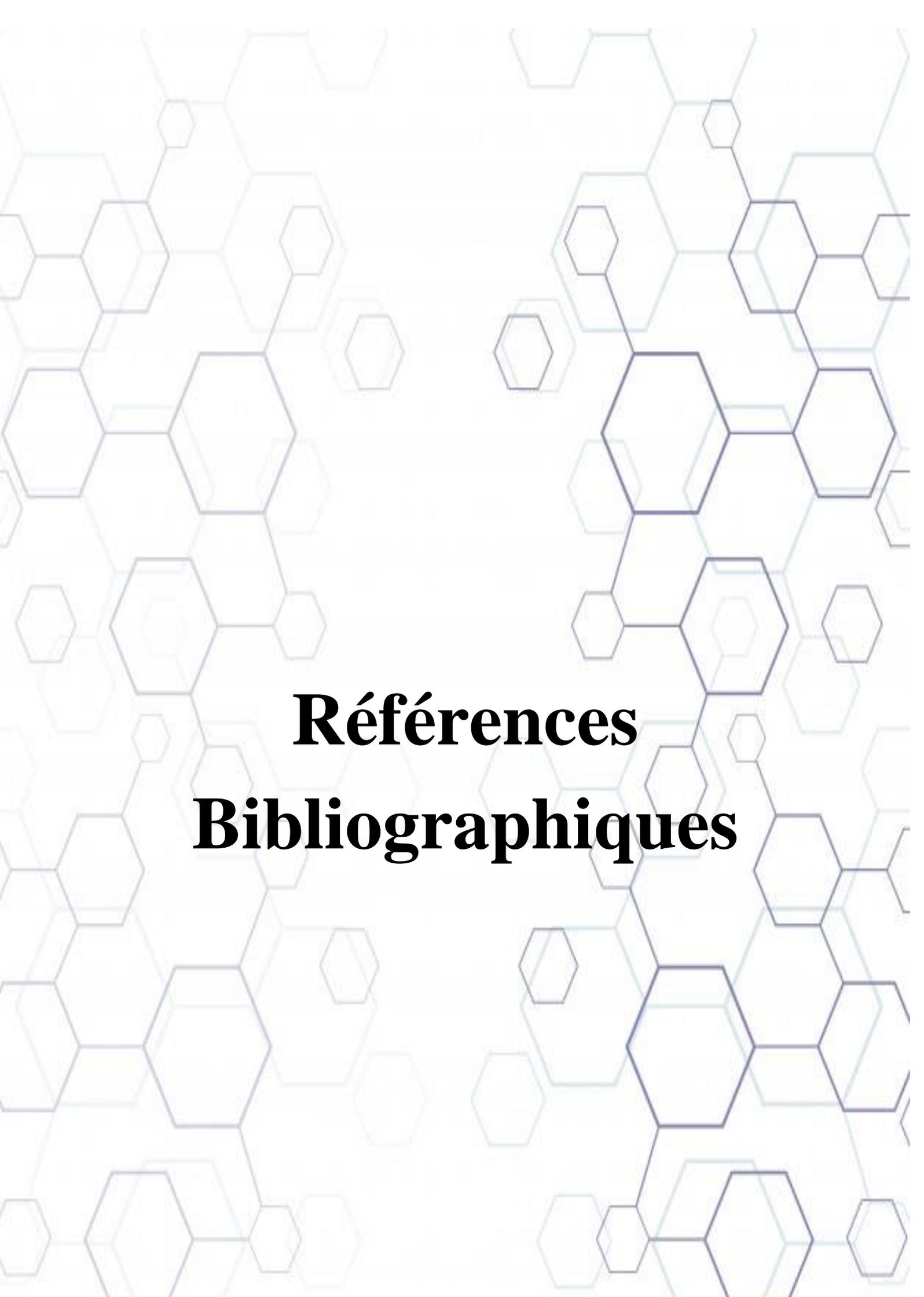
$$PIC_{50} = -3,71315(+/-0,54341) + 59,80287(+/-9,36056) MATS2v^2 + 0,00588(+/-0,00123) ESpm12d^2 + 19,18589(+/-11,52495) G1m^2 + 1,05578(+/-0,29918) H5u^2$$

$$n = 59 \quad R^2 = 58.2 \% \quad R_{adj}^2 = 54.02\% \quad F = 13.92 \quad N_{ent} = 45 \quad N_{test} = 14 \quad Q_{100}^2 = 50.66$$

$$Q_{training}^2 = 0,507 \quad R_{TEST}^2 = 0,834$$

Est très robuste, possède des excellents paramètres prédictifs internes et externes. Ce modèle est capable de prédire l'activité inhibitrice de tyrosine kinase de nouveaux composés.

Notre objectif principal de de ce mémoire a été atteint, et afin de poursuivre notre travail de recherche dans ce domaine nous prévoyons d'élaborer des modèles en utilisant d'autres méthodes telles que les réseaux de neurones artificiels (RNA), Support Vecteur Machine (SVM).



Références Bibliographiques

References

- [1] A. Forner, J. M. Llovet, et J. Bruix, « Hepatocellular carcinoma », *The Lancet*, vol. 379, n° 9822, p. 1245-1255, mars 2012, doi: 10.1016/S0140-6736(11)61347-0.
- [2] A.-L. Cheng *et al.*, « Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial », *Lancet Oncol.*, vol. 10, n° 1, p. 25-34, janv. 2009, doi: 10.1016/S1470-2045(08)70285-7.
- [3] N. Hamdane, T. F. Baumert, et M. B. Zeisel, « Carcinome hépatocellulaire après éradication du virus de l'hépatite C par les antiviraux à action directe: Prochains défis », *médecine/sciences*, vol. 34, n° 5, p. 391-394, mai 2018, doi: 10.1051/medsci/20183405007.
- [4] M. Kratky et J. Vinsova, « Antiviral Activity of Substituted Salicylanilides - A Review », *Mini-Rev. Med. Chem.*, vol. 11, n° 11, p. 956-967, oct. 2011, doi: 10.2174/138955711797068382.
- [5] C. Yap, Y. Xue, et Y. Chen, « Application of Support Vector Machines to In Silico Prediction of Cytochrome P450 Enzyme Substrates and Inhibitors », *Curr. Top. Med. Chem.*, vol. 6, n° 15, p. 1593-1607, août 2006, doi: 10.2174/156802606778108942.
- [6] W.-S. Hu et S. H. Hughes, « HIV-1 Reverse Transcription », *Cold Spring Harb. Perspect. Med.*, vol. 2, n° 10, p. a006882-a006882, oct. 2012, doi: 10.1101/cshperspect.a006882.
- [7] D. N. Goossens, T. Koessler, et P. L. Spahr, « Carcinome hépatocellulaire : nouvelles recommandations de prise en charge », *Rev. MÉDICALE SUISSE*, p. 4, 2018.
- [8] F. X. Bosch, J. Ribes, et J. Borràs, « Epidemiology of primary liver cancer », *Semin. Liver Dis.*, vol. 19, n° 3, p. 271-285, 1999, doi: 10.1055/s-2007-1007117.
- [9] J.-L. Raoul, J. Raimbourg, S. Hiret, X. Adhoute, et H. Senellart, « Épidémiologie du carcinome hépatocellulaire : simple augmentation d'incidence ou futur drame ? », *Bull. Cancer (Paris)*, vol. 105, n° 5, p. 502-507, mai 2018, doi: 10.1016/j.bulcan.2018.02.003.
- [10] J.-F. Dufour, « Le carcinome hépatocellulaire », *Schweiz. Gastroenterol.*, vol. 1, n° 1, p. 18-22, mars 2020, doi: 10.1007/s43472-020-00005-z.

- [11] H. B. El-Serag et K. L. Rudolph, « Hepatocellular Carcinoma: Epidemiology and Molecular Carcinogenesis », *Gastroenterology*, vol. 132, n° 7, p. 2557-2576, juin 2007, doi: 10.1053/j.gastro.2007.04.061.
- [12] P. A. Farazi et R. A. DePinho, « Hepatocellular carcinoma pathogenesis: from genes to environment », *Nat. Rev. Cancer*, vol. 6, n° 9, p. 674-687, sept. 2006, doi: 10.1038/nrc1934.
- [13] H.-O. Adami *et al.*, « Excess Risk of Primary Liver Cancer in Patients With Diabetes Mellitus », *JNCI J. Natl. Cancer Inst.*, vol. 88, n° 20, p. 1472-1477, oct. 1996, doi: 10.1093/jnci/88.20.1472.
- [14] D. L. Sudakin, « Dietary Aflatoxin Exposure and Chemoprevention of Cancer: A Clinical Review: REVIEW », *J. Toxicol. Clin. Toxicol.*, vol. 41, n° 2, p. 195-204, janv. 2003, doi: 10.1081/CLT-120019137.
- [15] I. Hsu, R. Metcalf, T. Sun, J. Welsh, N. Wang, et C. Harris, « Mutational hot spot in the p53 gene in human hepatocellular carcinomas | Nature ». <https://www.nature.com/articles/350427a0> (consulté le mai 27, 2021).
- [16] V. Moudgil, D. Redhu, S. Dhanda, et J. Singh, « A Review of Molecular Mechanisms in the Development of Hepatocellular Carcinoma by Aflatoxin and Hepatitis B and C Viruses », *J. Environ. Pathol. Toxicol. Oncol.*, vol. 32, n° 2, 2013, doi: 10.1615/JEnvironPatholToxicolOncol.2013007166.
- [17] M. Pompili, « Bridging and downstaging treatments for hepatocellular carcinoma in patients on the waiting list for liver transplantation », *World J. Gastroenterol.*, vol. 19, n° 43, p. 7515, 2013, doi: 10.3748/wjg.v19.i43.7515.
- [18] O. B. Gbolahan, M. A. Schacht, E. W. Beckley, T. P. LaRoche, B. H. O'Neil, et M. Pyko, « Locoregional and systemic therapy for hepatocellular carcinoma », *J. Gastrointest. Oncol.*, vol. 8, n° 2, p. 215-228, avr. 2017, doi: 10.21037/jgo.2017.03.13.
- [19] S. F. Ang *et al.*, « The Singapore Liver Cancer Recurrence (SLICER) Score for Relapse Prediction in Patients with Surgically Resected Hepatocellular Carcinoma », *PLOS ONE*, vol. 10, n° 4, p. e0118658, avr. 2015, doi: 10.1371/journal.pone.0118658.
- [20] J. Fan *et al.*, « Liver transplantation outcomes in 1,078 hepatocellular carcinoma patients: a multi-center experience in Shanghai, China », *J. Cancer Res. Clin. Oncol.*, vol. 135, n° 10, p. 1403-1412, oct. 2009, doi: 10.1007/s00432-009-0584-6.

- [21] Jessica ZUCMAN-ROSSI, « CARACTÉRISATION MOLÉCULAIRE DES CARCINOMES HÉPATOCELLULAIRES LIÉS AU VIRUS DE L'HÉPATITE B », Institut Universitaire d'Hématologie, 2014.
- [22] G. K. Abou-Alfa *et al.*, « Phase II Study of Sorafenib in Patients With Advanced Hepatocellular Carcinoma », *J. Clin. Oncol.*, vol. 24, n° 26, p. 4293-4300, sept. 2006, doi: 10.1200/JCO.2005.01.3441.
- [23] J.L. Merlin, « Les inhibiteurs de tyrosine kinase en oncologie 1 Tyrosine kinase inhibitors in oncology | Semantic Scholar ». <https://www.semanticscholar.org/paper/Les-inhibiteurs-de-tyrosine-kinase-en-oncologie-1-Merlin/68e6d1bbbfbf6fdfaa881f41e35fc9b56edb4480> (consulté le juin 04, 2021).
- [24] N. Ding, W. Zhang, H. L. Xiao, P. Wang, et Y. X. Li, « Synthesis and biological evaluation of a series of novel salicylanilides as inhibitors of EGFR protein tyrosine kinases », *Chin. Chem. Lett.*, vol. 23, n° 5, p. 529-532, mai 2012, doi: 10.1016/j.ccllet.2012.03.016.
- [25] PubChem, « Sorafenib ». <https://pubchem.ncbi.nlm.nih.gov/compound/216239> (consulté le juin 12, 2021).
- [26] A. A. Abdelgalil, H. M. Alkahtani, et F. I. Al-Jenoobi, « Sorafenib », in *Profiles of Drug Substances, Excipients and Related Methodology*, vol. 44, Elsevier, 2019, p. 239-266. doi: 10.1016/bs.podrm.2018.11.003.
- [27] B. I. Rini, « Sorafenib », *Expert Opin. Pharmacother.*, vol. 7, n° 4, p. 453-461, mars 2006, doi: 10.1517/14656566.7.4.453.
- [28] O. Rosmorduc *et al.*, « Bon usage du sorafénib chez les patients pris en charge pour un carcinome hépatocellulaire ou un cancer du rein », p. 7.
- [29] V. Ozenne *et al.*, « Tolerance and outcome of patients with unresectable hepatocellular carcinoma treated with sorafenib », *Eur. J. Gastroenterol. Hepatol.*, vol. 22, n° 9, p. 1106-1110, sept. 2010, doi: 10.1097/MEG.0b013e3283386053.
- [30] M. Nakano *et al.*, « Efficacy, Safety, and Survival Factors for Sorafenib Treatment in Japanese Patients with Advanced Hepatocellular Carcinoma », *Oncology*, vol. 84, n° 2, p. 108-114, 2013, doi: 10.1159/000342650.

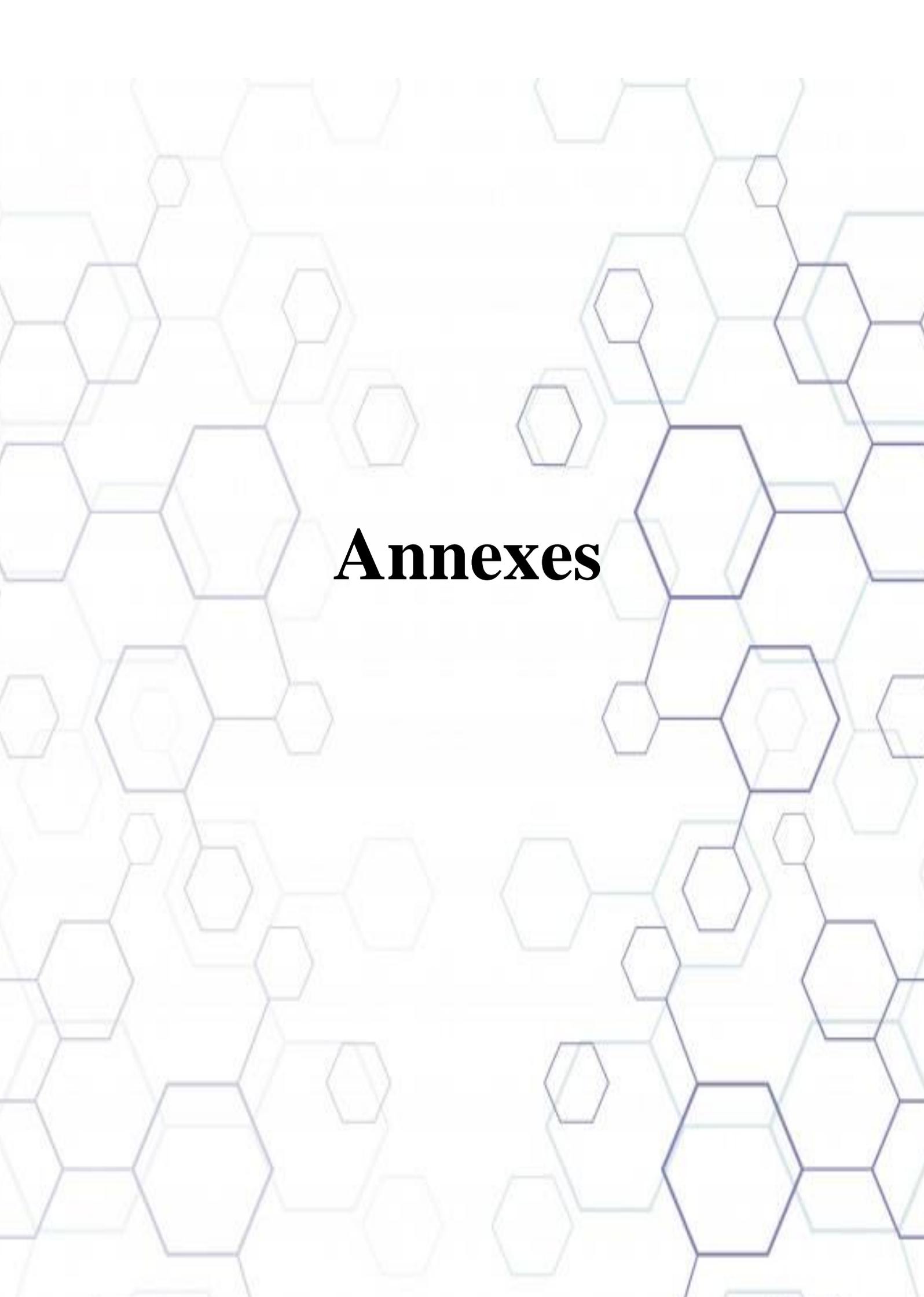
- [31] M. J. Macielag *et al.*, « Substituted Salicylanilides as Inhibitors of Two-Component Regulatory Systems in Bacteria », *J. Med. Chem.*, vol. 41, n° 16, p. 2939-2945, juill. 1998, doi: 10.1021/jm9803572.
- [32] Miyanou ROSALES HURTADO, « Approche pluridisciplinaire sur la problématique de la résistance bactérienne : conception, synthèse et évaluation de l'activité biologique de nouveaux agents antibactériens », UNIVERSITÉ DE NÎMES, 2019.
- [33] G. Hamilton et B. Rath, « Repurposing of Anthelmintics as Anticancer Drugs », *Oncomedicine*, vol. 3, p. 1-8, 2018, doi: 10.7150/oncm.20563.
- [34] F. Laudisi, M. Marônek, A. Di Grazia, G. Monteleone, et C. Stolfi, « Repositioning of Anthelmintic Drugs for the Treatment of Cancers of the Digestive System », *Int. J. Mol. Sci.*, vol. 21, n° 14, p. 4957, juill. 2020, doi: 10.3390/ijms21144957.
- [35] M. Krátký, J. Vinšová, E. Novotná, J. Mandíková, F. Trejtnar, et J. Stolaříková, « Antibacterial Activity of Salicylanilide 4-(Trifluoromethyl)-benzoates », *Molecules*, vol. 18, n° 4, p. 3674-3688, mars 2013, doi: 10.3390/molecules18043674.
- [36] S. Basak et R. Mishra, « Editorial: Descriptor Spaces for QSAR: Delving into the History of the Expanding Frontier », *Curr. Comput. Aided-Drug Des.*, vol. 13, n° 1, p. 4-7, janv. 2017, doi: 10.2174/157340991301170127200832.
- [37] C. Toshiofujita', « Analysis. A Method for the Correlation of Biological Activity and Chemical Structure », p. 11, 1964.
- [38] A. Debnath, « Quantitative Structure-Activity Relationship (QSAR) Paradigm - Hansch Era to New Millennium », *Mini-Rev. Med. Chem.*, vol. 1, n° 2, p. 187-195, juill. 2001, doi: 10.2174/1389557013407061.
- [39] M. Grover, B. Singh, M. Bakshi, et S. Singh, « Quantitative structure–property relationships in pharmaceutical research – Part 2 », vol. 3, n° 2, p. 8, 2000.
- [40] M. Grover, B. Singh, M. Bakshi, et S. Singh, « Quantitative structure–property relationships in pharmaceutical research – Part », vol. 3, n° 1, p. 8, 2000.
- [41] « Chemoinformatics in Drug Discovery | Oprea T.I. | download ». <https://3lib.net/book/2087582/a5978a> (consulté le juin 05, 2021).

- [42] E. A. Rekkà et P. N. Kourounakis, *Chemistry and Molecular Aspects of Drug Design and Action*. CRC Press, 2008.
- [43] A. K. Ghose et V. N. Viswanadhan, Éd., *Combinatorial library design and evaluation: principles, software tools, and applications in drug discovery*. New York: M. Dekker, 2001.
- [44] BOUFENNARA FERIAAL et BRIK MERIEM, « Etude chimio métrique des dérivés de quinazoline Pour le traitement de cancer du sein », Université des Frères Mentouri Constantine 1, 2020.
- [45] N. Trinajstić, S. Nikolić, S. C. Basak, et I. Lukovits, « Distance Indices and Their Hyper-Counterparts: Intercorrelation and Use in the Structure-Property Modeling », *SAR QSAR Environ. Res.*, vol. 12, n° 1-2, p. 31-54, avr. 2001, doi: 10.1080/10629360108035370.
- [46] A. S. Reddy, S. Kumar, et R. Garg, « Hybrid-genetic algorithm based descriptor optimization and QSAR models for predicting the biological activity of Tipranavir analogs for HIV protease inhibition », *J. Mol. Graph. Model.*, vol. 28, n° 8, p. 852-862, juin 2010, doi: 10.1016/j.jmgm.2010.03.005.
- [47] K. Kraim, « ETUDES DE QSAR SUR DES ACTIVITES BIOLOGIQUES UTILISANT DES PRODUITS D'ORIGINES NATURELS », p. 154, 2009.
- [48] E. X. Esposito, A. J. Hopfinger, et J. D. Madura, « Methods for Applying the Quantitative Structure-Activity Relationship Paradigm », in *Chemoinformatics*, vol. 275, J. Bajorath, Éd. Totowa, NJ: Humana Press, 2004, p. 131-213. doi: 10.1385/1-59259-802-1:131.
- [49] A. Dudek, T. Arodz, et J. Galvez, « Computational Methods in Developing Quantitative Structure-Activity Relationships (QSAR): A Review », *Comb. Chem. High Throughput Screen.*, vol. 9, n° 3, p. 213-228, mars 2006, doi: 10.2174/138620706776055539.
- [50] R. Perkins, H. Fang, W. Tong, et W. J. Welsh, « QUANTITATIVE STRUCTURE–ACTIVITY RELATIONSHIP METHODS: PERSPECTIVES ON DRUG DISCOVERY AND TOXICOLOGY », *Environ. Toxicol. Chem.*, vol. 22, n° 8, p. 1666, 2003, doi: 10.1897/01-171.
- [51] S. Chtita, « En vue de l'obtention de Doctorat en Chimie Spécialité : Chimie Physique et Modélisation », p. 186.
- [52] M. Şahin, Y. Kaya, et M. Uyar, « Comparison of ANN and MLR models for estimating solar radiation in Turkey using NOAA/AVHRR data », *Adv. Space Res.*, vol. 51, n° 5, p. 891-904, mars 2013, doi: 10.1016/j.asr.2012.10.010.

- [53] Almi Zineb, « Etude qualitative et quantitative des relations structuresactivités dans des hétérocycles à intérêt pharmaceutique. », Université Mohamed Khider-Biskra, 2015.
- [54] J. JACQUES, « Modélisation statistique », *Poly Tech Lille*.
- [55] D. Laffly, « Régression multiple : principes et exemples d'application », p. 33.
- [56] J. Confais et M. L. Guen, « Premiers pas en régression linéaire avec SAS® », p. 146.
- [57] *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Elsevier, 2015. doi: 10.1016/C2014-0-00286-9.
- [58] R. W. Kennard et L. A. Stone, « Computer Aided Design of Experiments », *Technometrics*, vol. 11, n° 1, p. 137-148, févr. 1969, doi: 10.1080/00401706.1969.10490666.
- [59] A. Mignon, « Apprentissage de métriques et méthodes à noyaux appliqués à la reconnaissance de personnes dans les images », p. 127.
- [60] P. Y. Chen et H. M. Yu, « Euclidean distance-based method for variable selection and its application in geological data », *J. Interdiscip. Math.*, vol. 19, n° 2, p. 215-232, mars 2016, doi: 10.1080/09720502.2014.932122.
- [61] F. Noé, S. M. Schwarzl, S. Fischer, et J. C. Smith, « Computational tools for analysing structural changes in proteins in solution », *Appl. Bioinformatics*, p. 8.
- [62] S. Wold, L. Eriksson, J. Trygg, et N. Kettaneh, « The PLS method -- partial least squares projections to latent structures -- and its applications in industrial RDP (research, development, and production). », p. 44.
- [63] D. Kovács, P. Király, et G. Tóth, « Sample-size dependence of validation parameters in linear regression models and in QSAR », *SAR QSAR Environ. Res.*, vol. 32, n° 4, p. 247-268, avr. 2021, doi: 10.1080/1062936X.2021.1890208.
- [64] A. Golbraikh et A. Tropsha, « Beware of q²! », *J. Mol. Graph. Model.*, vol. 20, n° 4, p. 269-276, janv. 2002, doi: 10.1016/s1093-3263(01)00123-1.
- [65] A. Yasri et D. Hartsough, « Toward an Optimal Procedure for Variable Selection and QSAR Model Building », *J. Chem. Inf. Comput. Sci.*, vol. 41, n° 5, p. 1218-1227, sept. 2001, doi: 10.1021/ci010291a.

- [66] U. Muhammad, A. Uzairu, et D. Ebuka Arthur, « Review on: quantitative structure activity relationship (QSAR) modeling », *J. Anal. Pharm. Res.*, vol. 7, n° 2, avr. 2018, doi: 10.15406/japlr.2018.07.00232.
- [67] N. BOUARRA, « Etudes QSPR des propriétés contrôlant l'évolution de quelques HAP dans l'environnement », UNIVERSITÉ BADJI MOKHTAR - ANNABA, 2018.
- [68] P. K. Ojha, I. Mitra, R. N. Das, et K. Roy, « Further exploring rm2 metrics for validation of QSPR models », *Chemom. Intell. Lab. Syst.*, vol. 107, n° 1, p. 194-205, mai 2011, doi: 10.1016/j.chemolab.2011.03.011.
- [69] A. Tropsha, P. Gramatica, et V. Gombar, « The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models », *QSAR Comb. Sci.*, vol. 22, n° 1, p. 69-77, avr. 2003, doi: 10.1002/qsar.200390007.
- [70] A. Tropsha, « Best Practices for QSAR Model Development, Validation, and Exploitation », *Mol. Inform.*, vol. 29, n° 6-7, p. 476-488, juill. 2010, doi: 10.1002/minf.201000061.
- [71] P. Gramatica, « Principles of QSAR models validation: internal and external », *QSAR Comb. Sci.*, vol. 26, n° 5, p. 694-701, mai 2007, doi: 10.1002/qsar.200610151.
- [72] L. Eriksson, J. Jaworska, A. P. Worth, M. T. D. Cronin, R. M. McDowell, et P. Gramatica, « Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs », *Environ. Health Perspect.*, vol. 111, n° 10, p. 1361-1375, 2003.
- [73] T. I. Netzeva *et al.*, « Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships: The Report and Recommendations of ECVAM Workshop 52 » *Altern. Lab. Anim.*, vol. 33, n° 2, p. 155-173, avr. 2005, doi: 10.1177/026119290503300209.
- [74] G. Paraskevopoulos *et al.*, « Novel salicylanilides from 4,5-dihalogenated salicylic acids: Synthesis, antimicrobial activity and cytotoxicity », *Bioorg. Med. Chem.*, vol. 25, n° 4, p. 1524-1532, févr. 2017, doi: 10.1016/j.bmc.2017.01.016.
- [75] J. Vinšová *et al.*, « Salicylanilide diethyl phosphates: Synthesis, antimicrobial activity and cytotoxicity », *Bioorg. Med. Chem.*, vol. 22, n° 2, p. 728-737, janv. 2014, doi: 10.1016/j.bmc.2013.12.016.

- [76] C. Liechti, U. Séquin, G. Bold, P. Furet, T. Meyer, et P. Traxler, « Salicylanilides as inhibitors of the protein tyrosine kinase epidermal growth factor receptor », *Eur. J. Med. Chem.*, vol. 39, n° 1, p. 11-26, janv. 2004, doi: 10.1016/j.ejmech.2003.09.010.
- [77] H. H. J. Gerets *et al.*, « Characterization of primary human hepatocytes, HepG2 cells, and HepaRG cells at the mRNA level and CYP activity in response to inducers and their predictivity for the detection of human hepatotoxins », *Cell Biol. Toxicol.*, vol. 28, n° 2, p. 69-87, avr. 2012, doi: 10.1007/s10565-011-9208-4.
- [78] « Chemdraw ultra: Molecular modelling, structure drawing, Semi-empirical calculations, structure display, MOPAC, solvation energy, MM2 ». http://www.chemistry-software.com/chemical_drawing.htm (consulté le juill. 07, 2021).
- [79] Hyper, « HyperChem 6.03 Package ». <https://hyper.com/> (consulté le juill. 07, 2021).
- [80] « Molecular descriptors, QSAR, chemometrics and chemoinformatics - Talete srl ». <http://www.taletemi.it/> (consulté le juill. 07, 2021).
- [81] « DTC lab », *DTC lab*. <https://dtclab.webs.com/> (consulté le juill. 07, 2021).
- [82] K. Roy, R. N. Das, P. Ambure, et R. B. Aher, « Be aware of error measures. Further studies on validation of predictive QSAR models », *Chemom. Intell. Lab. Syst.*, p. 16, 2016.
- [83] P. Ambure, « Understanding the structural requirements of cyclic sulfone hydroxyethylamines as hBACE1 inhibitors against A β plaques in Alzheimer's disease: a predictive QSAR approach », *RSC Adv.*, p. 16, 2016.
- [84] K. Roy, I. Mitra, S. Kar, P. K. Ojha, R. N. Das, et H. Kabir, « Comparative Studies on Some Metrics for External Validation of QSPR Models », *J. Chem. Inf. Model.*, vol. 52, n° 2, p. 396-408, févr. 2012, doi: 10.1021/ci200520g.
- [85] S. Wold, L. Eriksson, et S. Clementi, « Statistical Validation of QSAR Results », in *Methods and Principles in Medicinal Chemistry*, 1^{re} éd., H. van de Waterbeemd, Éd. Wiley, 1995, p. 309-338. doi: 10.1002/9783527615452.ch5.
- [86] M. A. DEMS *et al.*, « Etude théorique des réarrangements haptotropiques π - σ des fragments M(CO) $_3$ », Mohamed BOUDIAF University, M'sila. Algeria, 2015.



Annexes

Annexes :

Tableau 10: Signification des descripteurs.

Descripteurs	Descripteurs
MATS2v	Autocorrélation de Moran du décalage 2 pondéré par le volume de van der Waals (Moran autocorrelation of lag 2 weighted by van der Waals volume)
ESpm12d	Moment spectral 12 à partir du bord adj (Spectral moment 12 from edge adj)
G1m	Indice WHIM directionnel de symétrie 1ère composante / pondéré en masse (1st component symmetry directional WHIM index / weighted by mass)
H5u	H autocorrélation du décalage 5 / non pondéré (H autocorrelation of lag 5 / unweighted)

Tableau 11: Les valeurs d'IC50 et les quatre descripteurs.

Code	PIC50	MATS2v ²	ESpm12d ²	G1m ²	H5u ²
1a'	-0.51054501	0.014161	242.6429	0.046225	0.128164
1b'	-0.34830486	0.015625	249.3557	0.030276	0.1444
1c'	-0.42160393	0.006241	324.9727	0.028561	0.331776
1d'	-0.98000337	0.000081	348.9798	0.027225	0.1936
2a'	-0.88024178	0.008836	242.3315	0.030276	0.147456
2b'	-1.24723655	0.009025	249.1031	0.030276	0.152881
2c'	-1.17724784	0.002601	324.9367	0.028561	0.346921
2d'	-0.90741136	0.000729	348.9798	0.027225	0.198025
3a'	-0.49692965	0.014161	242.7052	0.030276	0.173056
3b'	-0.29885308	0.015625	249.3873	0.038025	0.178929
3c'	-0.2764618	0.006241	324.9727	0.028561	0.465124
3d'	-1.00817418	0.000081	348.9798	0.027225	0.183184

4a'	-1.372912	0.008836	242.3938	0.030276	0.201601
4b'	-1.49789674	0.009025	249.1347	0.030276	0.164836
4c'	-0.25767857	0.002601	324.9367	0.028561	0.341056
4d'	-0.8920946	0.000729	348.9798	0.027225	0.198916
5a'	-0.99033885	0.005329	248.9453	0.034225	0.187489
5b'	-1.04414762	0.005329	254.5939	0.030276	0.158404
5c'	-0.8142476	0.000361	325.6581	0.028561	0.465124
5d'	-0.68124124	0.003969	349.3535	0.033124	0.184041
6a'	-1.25236751	0.003025	248.4722	0.030276	0.178929
6b'	-0.90091307	0.002916	254.1792	0.030276	0.163216
6c'	-0.79588002	0.000004	325.5859	0.028561	0.446224
6d'	-0.47712125	0.005476	349.3161	0.027225	0.175561
7a'	-1.33061667	0.009025	249.4188	0.030276	0.191844
7b	-1.0941216	0.009801	255.009	0.030276	0.165649
7c'	-0.99431715	0.001764	325.6942	0.028561	0.332929
7d'	-0.68663627	0.002401	349.3909	0.027225	0.181476
8a'	-1.03462846	0.005329	248.9453	0.030276	0.192721
8b'	-0.88366144	0.005329	254.5939	0.030276	0.185761
8c'	-1.04257551	0.000361	325.6581	0.028561	0.308025
8d'	-0.70500796	0.003969	349.3535	0.027225	0.183184
1a	-0.74350976	0.005041	253.8924	0.027556	0.369664
1b	-0.8142476	0.010609	254.0198	0.027556	0.352836
1c	-0.83314711	0.010609	253.8286	0.024025	0.356409
1d	-0.30963017	0.020164	258.6307	0.024025	0.351649

1e	-0.26481782	0.011025	326.272	0.023409	0.451584
1f	-0.26481782	0.000729	326.0191	0.026569	0.877969
1g	-0.64443859	0.013225	253.7968	0.024025	0.379456
1h	-0.74818803	0.013225	253.6375	0.024025	0.358801
1i	-0.86923172	0.005041	253.7012	0.027556	0.349281
1j	-0.42813479	0.000064	326.0914	0.026569	0.760384
1k	-1.112605	0.007744	254.0517	0.024025	0.354025
1l	-1.2347703	0.002809	254.1155	0.024025	0.335241
1m	-1.49762065	0.002809	253.9242	0.024025	0.343396
1n	-0.1931246	0.010609	253.8605	0.027556	0.352836
1o	-0.45331834	0.017161	258.7916	0.027556	0.338724
1p	-0.25767857	0.010609	254.0198	0.024025	0.388129
1q	-0.75434834	0.007744	254.243	0.025921	0.363609
1r	-0.36735592	0.000064	326.0552	0.026569	0.887364
1s	-0.98091194	0.007744	254.243	0.027556	0.338724
1t	-0.90687353	0.010609	254.0198	0.024025	0.400689
1u	-1.04805317	0.002809	254.1155	0.024025	0.363609
1v	-1.5291736	0.002809	253.9242	0.025921	0.356409
1w	-0.42651126	0.010609	253.8605	0.024025	0.366025
1x	-0.74818803	0.007744	254.0517	0.031329	0.344569
1y	-0.28330123	0.017161	258.8237	0.027556	0.334084
1z	-0.58546073	0.000064	326.1275	0.026569	0.732736
1zz	-0.39967372	0.000064	326.0914	0.023409	0.893025

Année universitaire : 2020/2021

Présenté par :

REDJAH ZAKARIA

SELLAMI MOHAMED AMINE

Analyse de la relation quantitative Structure / Activité des inhibiteurs de la tyrosine kinase

Mémoire de fin de cycle en vue de l'obtention du diplôme de Master en Biochimie Appliquée.

Résumé

La relation quantitative entre la structure et l'activité inhibitrice de la Tyrosine Kinase pour 59 dérivés de Salicylanilide a été exploitée. La géométrie des composés utilisés dans ce travail ont été pré-optimisés avec le Mécanique moléculaire MM+ puis optimisé à l'aide de la méthode semi-empirique PM3, les géométries ainsi optimisées ont été soumises à un calcul de leurs descripteurs.

Un modèle QSAR a été construit en se basant sur la régression linéaire multiple (RLM), l'algorithme génétique (GA) a été associé pour sélectionner les descripteurs les plus importants.

Le modèle QSAR pour la prédiction l'activité inhibitrice de la Tyrosine Kinase a été développé à partir de 4 descripteurs ($MATS2v^2$, $ESpm12d^2$, $G1m^2$, $H5u^2$), les contrôles de Roy et les conditions de Golbraikh et Tropsha sont utilisés et vérifiés.

Le modèle établi, a été validé selon des validations internes et externes pour juger sa stabilité, sa robustesse et sa capacité prédictive ($R^2=58.2\%$, $Q^2_{LOO}=50.66$, $Q^2_{training}=50.7\%$, $F=13.92$, $R^2_{test}=83.4$).

Le domaine d'application pour notre modèle est étudié à l'aide du diagramme de Williams pour détecter les composés aberrants.

Ce modèle peut être utilisé efficacement pour estimer des molécules qui n'ont pas été testées expérimentalement.

Mots clés : modèle QSAR, activité inhibitrice de la tyrosine kinase, algorithme génétique (AG), salicylanilide

Laboratoire de recherche : laboratoire de l'informatique N°18 au niveau du centre de recherche de la biotechnologie (C.R.B.T), Constantine.

Jury de dévaluation:

Président du jury: Mr BENSEGUENI. A. Pr. UFM Constantine-1

Encadreur : Mr. DEMS MOHAMED ABDESSELEM. MRA CRBt Constantine

Examineur : Mr. MOKRANI E. H Maître- Conférences. B. UFM Constantine-1

Examineur: Mr BOUKELIA ABDELBASSET. MRA CRBt Constantine

Date de soutenance : 19/09/2021