REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEURE ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DE MEDEA

FACULTE DES SCIENCES



Mémoire

Présenté pour obtenir le diplôme de **Master**

En chimie

Spécialité :
Chimie pharmaceutique

Présenté par :

KAIROUANI Malika
BACHOUCHE Nadjet

Thème :

Etude QSAR sur les dérivés de l'imidazopyridazine comme inhibiteurs de la PfPK7

Soutenue à distance le : … /07 /2021 et évalué par le jury composé de :

MESELLEM Yamine	Grade : MCB	Université de Médéa	Président
ABIZA Amirouche	Grade : MAA	Université de Médéa	Examinateur
BITAM Said	Grade : MCA	Université de Médéa	Encadreur

Année universitaire : 2020/2021

Remerciement

Nous remercions dieu abondamment de nous avoir accordé la santé et la volonté de commencer et de terminer cette thèse, puis j'exprime mes remerciements et ma gratitude au Prof. **Dr. BITAM Said** qui a supervisé la thèse, qui nous a donné beaucoup de son temps, et sa générosité, la grandeur de son caractère et son style distingué dans le suivi de cette thèse ont eu le plus grand impact sur moi. Aidez-nous á mener á bien ce travail, et je demande au tout-puissant de le récompense et d'écrire son acte dans le l'équilibre de ses bonnes actions.

J'exprime également mes remerciements et ma reconnaissance á tous nos professeurs pour leur générosité et leur grande patience dont ils ont fait preuve tout au long des années scolaires, ainsi qu'aux membres de l'administration universitaire pour leur bonne prestation aux étudiants et leur aide dans tous les domaines cela leur donnerait un espace confortable pour étudier.

Dédicace

Je dédie ce modeste travail à mes parents qui on' ont Soutenus depuis mon plus jeune âge ont toujours fait leurs maximum, en consacrant temps et argent pour m'éveiller et m'encourager dans mes passions, c'est grâce à vous et pour vous que j'ai fois mon mémoire.

- À mes chères sœurs : Nacira, Naima et Fatiha.
- À mes chères Frères : Nouradine, Hamid et Ayoub.
- À tous mes vrais amis qui ont répondues présentes à chaque fois que j'ai besoin d'eux : Amira, Sara …
- À tout la promotion de chimie pharmaceutique 2021-2022
- À ceux qui me connaissant, et surtout qui m'aiment.

Avec mes profonds sentiments.

Nadjet

Dédicace

❀ A ma chère, chère âme de mère.

❀ Aux personnes les plus chères et les plus proches de mon coeur, mon père, qui a été mon soutien et mon aide.

❀ A mes soeurs bien-aimées: Nabila et Salima.

❀ A tous ceux qui m'ont soutenu et m'ont conseillé de lutter et m'on apporté un soutien moral.

A tous, je dédie cet humble travail, demandant à Dieu Tout puissant de nous en faire bénéficier et de nous assurer le succès et d'éclairer notre chemin sur le chemain de la connaissance et de la bonté.

Malika

SOMMAIRE

Introduction Générale

Chapitre I : Rappel bibliographique

I.1Introduction 2
I.2 Définition 3
I.3 L'histoire du paludisme en Algérie 3
I.4 Les causes du paludisme 4
I.5 Les méthodes de traitement 4
I.6 Les médicaments efficaces actuellement en circulation 5
I.7 Les méthodes de conception des médicaments 6
I.8 La nécessité d'utiliser l'outil informatique dans la conception du médicament 9
Conclusion 10

Chapitre II : Etude quantitative de la relation structure-activité QSAR

II.1 Introduction 12
II.2 Définition 12
II.3 Principe 12
II.4 Objectifs 14
II.5 Application de QSAR 14
II.6 Activité thérapeutique 14
II.7 Descripteurs moléculaires 15
II.8 Les différentes approximations utilisées 17
II.8.1 La régression multiple linéaire RLM 17
II.8.2 Réseaux de neurones artificiels 17
II.9 Le principe des approximations utilisées 19
II.9.1 Le principe de RLM 19
II.9.2 Le principe de RNA 19
II.10 Les facteurs 20
II.10.1 Coefficient de corrélation (R) 20
II.10.2 Coefficient de détermination R^2 21
II.9.3 Le coefficient de détermination ajusté R^2_{adj} 21
II.10.4 Coefficient de corrélation de validation croisée R^2CV 22
II.9.5 Ecart type (s) 22
II.10.6 Fischer-Snedecor 22
II.10.7 Coefficient de prédiction 23
II.10.8 Validation des modèles 23
II.10.9 Colinéarité des variables explicatives 24

Chapitre III : Résultats et Discussion

III. Introduction 26
III.1 Partie 1 26

III.1.1 La synthèse des dérivés imidazopyridazine 26
III.1.2 Définition des logiciels utilisés 28
III.2 Partie 2 30
III.2.1 Sélection des descripteurs 30
III.2.2 Modèle RLM 30
III.2.3 Modèle réseau des neurones RNA 32
III.3 Conclusion 38
Conclusion Générale 40
Références 42
Annex 48

LISTE DES FIGURES

Chapitre II

Figure II.1. Modèle de l'étude QSAR 13

Figure II.2. Représentation d'un réseau de neurone 18

Figure II.3. Schéma représentatif de la structure du neurone 20

Chapitre III

Figure III.1. Structure d'imidazopyridazine 27

Figure III.2. Représentation graphique des concentrations inhibitrices (PIC50) expérimentales et celle prédite par le modèle RNA pour la série d'ensemble de test des dérives des imidazopyridazine 33

Figure III.3. Représentation graphique des concentrations inhibitrice (PIC50) expérimentales et celle prédite par le modèle RNA pour la série d'ensemble d'apprentissage des dérives de imidazopyridazine 34

Figure III.4. Représentation graphique des concentrations inhibitrice (PIC50) expérimentales et celle prédite par le modèle RNA pour la série d'ensemble d'apprentissage et test des dérives de imidazopyridazine. 35

LISTE DES TABLEAUX

Chapitre I

Tableau I.1. Evolution des pourcentages d'utilisation des traitements curatifs 6

Chapitre III

Tableau III.1. Dérives d'imidazopyridazine utilisés et valeurs de concertations inhibitrices expérimentales 27

Tableau II.2. Descripteur utilisée pour le développement des modèle QSAR 30

Tableau III.3. Valeurs du coefficient VIF des descripteurs sélectionnés.

Tableau III.4. Matrice de corrélation du modèle 31

Tableau III.5. Valeurs PIC50 expérimentales avec les descripteurs sélectionnés 31

Tableau III.6. Les différents paramètres MLP des deux ensembles d'apprentissage et test)

36

Tableau III.7. Les valeurs expérimentales, prédites et résiduelles 36

LISTE DES ABREVIATIONS

DUD : La Houque d'évolution de référence

IC_{50} : La concentration inhibitrice médiane

MLP : Perceptron multicouche

OMS : Organisation mondiale de la santé

PfPK7: Plasmodium falciparum protéine kinase 7

PRESS: predicted residual error some square

QSAR : Relation quantitative structure-activité

QSPR : Relation quantitative structure-propriété

RLM : Régression linéaire multiple

RNA : Réseaux de neurones artificiels

CV: cross validation

LOO: Leave-One-Out

VS : Le criblage virtuel

R : coefficient de corrélation

R^2 : coefficient de détermination

F : Statique de Fisher

S : Ecran type

Q^2 : coefficient de prédiction

S_{PRESS} : La validation standard des erreurs de prédiction

R^2CV : Coefficient de corrélation de validation croisée

TSS : Somme de carrés totaux

R^2_{adj} : Qualité d'ajustement

الملخص

تمثل الملاريا التي تسببها طفيليات البلازموديوم المنجلي تهديدا رئيسا للصحة العالمية. وهي واحدة من أكثر الأمراض الخطيرة التي تؤدي إلى الموت. يعد تحديد أهداف الطفيليات لتطوير الأدوية أولوية وتقترح مشتقات imidazopyridazine كأهداف مناسبة لعلاج تثبيط بروتين كيناز البلازموديوم فالسيباروم (pfpk7) وهي مجموعة جديدة نسبيا من مضادات الملاريا ذات التقارب والنوعية العالمية للموقع المستهدف حسب دراسات تقدير العلاقة الكمية بين البنية والنشاط (QSAR) الذي يتضمن نموذج غير خطي من خلال استخدام طريقة الشكات العصبية (RNA) كطريقة تعلم وإختيار متغير على التوالي: R=0.97، R^2=0.94، بالمقارنة مع تلك التي تم الحصول عليها من خلال الإنحدار الخطي المتعدد (RLM) باستخدام برنامج SPSS.

تثبت هذه النتائج الجيدة القدرة الكبيرة للشبكات العصبية الاصطناعية على نمذجة الفعالية المثبطة لمشتقات imidazopyridazine وقوة قدرتها التنبؤية، وبالتالي صحة وجود نماذج QSAR التي تم الحصول عليها.

الكلمات المفتاحية:

الملاريا، البلازموديوم المنجلي، تثبيط بروتين كيناز البلازموديوم فالسيباروم، الانحدار الخطي المتعدد،الشبكات العصبية الإصطناعية.

Résumé

Le paludisme causé par le parasite plasmadium falciparum représente une menace majeure pour la santé mondiale et est l'une des maladies les plus graves et mortelles. L'identification de cible parasitaires pour le développement de médicament de priorité et les dérivés imidazopyridazin sont suggérés comme cibles appropriées pour le traitement de l'inhibition de la protéine kinase Plasmodium falciparum pfpk7,un groupe relativement nouveau d'antipaludiques avec une affinité et une spécificité élevées pour le site cible selon des études estimant la relation quantitative entre la structure et l'activité QSAR, qui comprend un modèle non linéaire grâce à l'utilisation de la méthode du réseau neuronal RNA comme méthode d'apprentissage et de sélection de variables respectivement : R=0.97, R^2=0.94, comparé à celui obtenu par régression linéaire multiple RLM à l'aide de logiciel SPSS.

Ces bons résultats indiquent la grande capacité des RNA à modéliser l'activité inhibitrice des dérivés imidazopyridazine et la force de leur pouvoir prédictif, et donc la validité des modèles QSAR obtenus.

Mots clés :

Paludisme, plasmodium falciparum, QSAR, RLM, RNA, inhibiteurs de la pfpk7, imidozapyridazine.

Abstract

Malaria is caused by the parasite plasmodium falciparum represents a major threat to the global health and is one of the most serious and fatal diseases.

Identification of parasite targets for drug development is a priority and imidazopyridazine derivatives are suggested as appropriate targets for the treatment of Pfpk7, a relatively new group of antimalarials with high affinity and specificity to the target site according to studies estimating the quantitative relationship between structure and QSAR activity, Which includes a nonlinear model through the use of ANN neural network method as learning and variable selection method respectively : R=0.97, R^2=0.94, compared with that obtained by MLR multiple linear regression using SPSS software.

These good results indicate the large ability of ANN to model the inhibitory activity of imidazopyridazin derivatives and the strength of their predictive ability, and thus the validity of the obtained QSAR models

Keywords:

Malaria, QSAR, MLR, ANN, pfpk7 inhibitors.

Introduction Générale

Découvrir de nouveaux médicaments de la manière la plus efficace et la moins couteuse possible constitue un enjeu majeur pour les années à venir. Il est admis que, en moyenne, pour une molécule qui arrive sur le marché en tant que médicament innovant, 10 000 molécules sont synthétisées et testées. De plus, le développement d'un médicament demande généralement entre 10 et 15 ans de recherches. Il s'agit en effet de trouver une molécule qui doit à la fois présenter des propriétés thérapeutiques particulières, et posséder le minimum d'effets secondaires indésirables. Le prix de revient d'un médicament est essentiellement dû à ces synthèses longues et couteuses.

Pour cette raison, l'industrie pharmaceutique s'oriente vers de nouvelles méthodes de recherche, qui consistent à prédire les propriétés et activités de molécules avant même que celles-ci ne soient synthétisées. Deux disciplines de la « chimie computationnelle » se sont développées en réponse à ce besoin : les relations structure-activité ou QSAR (Quantitative Structure-Activity Relationship) et les relations structure-propriété ou QSPR (Quantitative Structure-Property Relationship).

La découverte d'une telle relation permet de prédire les propriétés physiques et chimiques et l'activité biologique de composés, de développer de nouvelles théories ou de comprendre les phénomènes observés. Elle permet également de guider les synthèses de nouvelles molécules, sans avoir à les réaliser, ou à analyser des familles entières de composés.

Le paludisme reste l'une des maladies les plus répandues et les plus mortelles dans le monde, et c'est une source de morbidité qui constitue un sérieux obstacle aux progrès social et économique de nombreux pays du monde en développement. Les parasites plasmodiums falciparum sont les souches les plus mortelles de plasmodium, étant responsables de la grande majorité es décès. En 2016, selon l'Organisation mondiale de la santé, le nombre estimé de décès dus au paludisme dans le monde était de 445000. La plupart de ces décès sont survenus en Afrique (92%), suivie de l'Asie de l'Est (6%) et de la région de la méditerranée orientale (2%). Ce fléau touche particulièrement les femmes enceintes et les enfants de moins de cinq ans dans les pays africains.

Le principal objectif de ce travail est l'application de la modélisation moléculaire dans le développement des méthodes QSAR des propriétés chimiques pour prédire activité thérapeutique attendue dans des nouvelles molécules bioactives pour de dérivés d'imidazopyridazine.

Notre travail est divisé en trois chapitres.

Le premier chapitre présente un rappel bibliographe sur le paludisme au Sahara algérien comment se produit le paludisme comment il peut traiter. Les médicaments efficaces actuellement comme nous avons abordé les méthodes de conception pour améliorés afin de rendre plus acceptable et désirable et aussi la nécessité d'utiliser des outils informatiques

Le deuxième chapitre intitulé « QSAR » : « principe et méthode » présente la méthode QSAR, son principe, son application et les différentes étapes à réaliser.

Le troisième chapitre est consacré à la méthodologie du travail réalisé dans cette thèse, aux résultats obtenus et à leurs discussions.

Enfin, nous terminerons ce manuscrit par une conclusion générale et des perspectives.

Chapitre I

Rappel bibliographique

I.1 Introduction

Le paludisme est une des maladies les plus transmises sur le globe dans toutes les régions tropicales, subtropicales et même dans certaines aires tempérées.

La disponibilité des moyens de traitement et de prévention chimique, les moustiques produites, la pulvérisation intra-domiciliaire d'insecticide à effet rémanent peuvent constituer une association importante dans la lutte contre le paludisme. Cette étude est fondée afin d'avoir des informations précises dans une zone couverte par ces différents moyens de lutte antipaludique. Elle a pour but de mesurer les paramètres entomologique et parasitologiques de la transmission du paludisme, une étude descriptive transversale à passage répété.

L'étude a été lancée dans 2 sites : Sirakorola et Kenenkou. Les enquêtes entomologique et cliniques ont été faites entre juillet 2010 et octobre 2013. Les données entomologiques ont été définies par la méthode d'aspersion intradomiciliaire, capture sur appât humain et par test à cône. La collecte des données a été faites parasito-cliniques par goutte épaisse et par examen clinique. La densité moyenne par chambre variait par site et par état de pulvérisation des chambres. La densité variait de façon décroissante de 2010 en 2011 dans les deux sites d'étude. Mais relativement élevés en 2012 surtout au mois de novembre dans les deux sites. L'anthropophile était élevée dans les deux sites, une grande proportion de l'été observée au cours de notre étude.

Les moustiques avaient plus tendance à piquer à l'intérieur des habitations. Le taux de l'inoculation entomologique était relativement élevé en saison des pluies et variable par passage dans les deux sites. Il ressort de cette étude que les insecticides utilisés en pulvérisation intra-domiciliaire avaient une rémanence assez bonne pour lutter contre les moustiques. Le paludisme était méso-endémiques dans les deux sites durant notre étude **[1].**

I.2 Définition

Le terme « malaria » vient du l'italien *mal'aria* et signifie mauvais air. Le terme « paludisme » est, quant à lui, d'origine latine *paludis* signifiant marais et est apparu en 1857. Le mot paludisme sera associé à la maladie en 1867 par le professeur A. Verneuil et sera ensuite repris par Charles Louis Alphones Laveran (1845-1922) (Vaucel et Féron 1968).

Le paludisme ou malaria est une maladie qui remonte au « temps des momies » (Millet et al.1994). En effet dans le papyrus d'Ebers (1570 av. JC), des symptômes tels que la fièvre, la splénomégalie, ainsi que des traitements contre la fièvre ont été décrits (Ebers GM and Stern L, 1875). De plus, une étude récente, réalisée par Hawass et son équipe, a montré que Toutankhamon était atteint de cette maladie (Hawass et al 2010). Au fil des siècles, la malaria a tué de nombreuses personnalités telles qu'Alexandre le grand en 323 av. JC, Dante en 1321. Elle a également touché le Roi Soleil qui fut guéri par l'absorption d'infusions à base de plantes et de quinquina. Par la suite, Laveran a démontré que la prise de quinine, extrait de l'écorce de quinquina, permettait l'élimination du parasite (Laveran 1982) **[2].**

- **Les symptômes**

Le paludisme est une maladie caractérisée par des épisodes fébriles aigues. Les symptômes apparaissent au bout de sept (7) jours ou plus (généralement 10 à 15 jours) après la piqûre infectante de moustique. Les premiers symptômes (fièvre, maux de tête, frissons et vomissements) peuvent été modérés et difficiles à attribuer au paludisme. S'il né est pas traité dans les 24 heures. Le paludisme à P. falciparum peut évoluer vers une affection sévère mortelle (OMS, 2014) **[3].**

I.3 L'histoire du paludisme en Algérie

Grace à la compagne d'éradication du paludisme lancée en Algérie en 1968, le nombre de cas a chuté de manière considérable, passant de 95 424 en 1960 à 30 en 1978. A cette date, le nord était déclaré indemne de paludisme à plasmodium falciparum. Très peu de cas dus à P. vivax persistaient dans des foyers résiduels au centre du pays.

A contrario, le sud de l'Algérie, non concerné par cette compagne, commençait à présenter à partir des années 1980 un profil épidémiologique inquiétant. L'augmentation régulière du nombre des cas importés et la résurgence des foyers épidémiques dans les oasis sahariennes coïncidaient avec l'ouverture de la route transsaharienne et avec l'augmentation du volume

des échanges commerciaux avec les pays du sahel. Plusieurs auteurs ont insisté sur le risque d'introduction du paludisme ou de ses vecteurs exotiques en Algérie via cette nouvelle voie de trafic. Actuellement, tous les cas autochtones de paludisme doivent être notifiés par les wilayas (départements) du sud du pays, ce qui représentait 300 cas déclarés entre 1980 et 2007. Les récentes épidémies enregistrées à la frontière Algéro-malienne et l'introduction d'Anophèles gambiae dans le territoire algérien démontrent la vulnérabilité du sud du pays au paludisme, accentuée vraisemblablement par des changements environnementaux locaux.

Les auteurs dressent un tableau de l'évolution du paludisme au Sahara algérien et de la distribution des anophèles dans cette région **[4].**

I.4 Les cause du paludisme

Le paludisme est causé par des parasites du genre *plasmodium* et transmis par des moustiques femelles appartenant au genre *anophèles*. Quatre espèces différentes causent le paludisme chez l'homme (P.falaparum, P. vivax, P. malanae et P.avale). Parmi elles, P. falaparum et P. Vivax ont la plus forte prévalence tandis que P. falaparum est la plus dangereuse. P. knowlesi est une espèce zoonotique qui peut également infester l'homme.

Quoique parfaitement évitable, le paludisme continue d'avoir des conséquences dévastatrices sur la santé et les moyens de subsistance des populations à travers le monde. Selon les dernières données disponibles, environ 3.2 milliards de personnes dans 97 pays, territoires et zones étaient exposées à la maladieen 2013, et l'on estime à 198 millions de nombre de cas survenus (fourchette : 124 millions-283 millions). Toujours en 2013, la maladie a tué 584 000 personnes (fourchette : 367 000-755 000), principalement des enfants de moins de 5 ans en Afrique subsaharienne. Dans la plupart des pays d'endémie palustre, la maladie touche de manière disproportionnée les pauvres et les groupes défavorisés, qui ont un accès limité aux établissements de santé et peuvent à peine se payer le traitement recommandé.

I.5 Les méthodes de traitement

Tous les cas suspects du paludisme doivent être confirmés au moyen de méthodes de détection des parasites, telles que la microscopie de qualité garantie ou les tests de diagnostic rapide. Les services de santé du secteur public et du secteur privé devraient confirmer le diagnostic avant administration du traitement antipaludique. Chaque cas confirmé devrait faire l'objet d'un suivi et être enregistré dans le système de surveillance, le but géant

d'orienter la planification des programmes. L'universalité des tests de diagnostic permettra de réduire la consommation excessive des combinaisons thérapeutiques à base de d'artérnisinine (le traitement de première intention contre le paludisme simple) et la pression médicamenteuse sur les parasites.

La généralisation des tests de diagnostic mettra à disposition des données de surveillance précises, communiquées en temps opportun, fondées sur les cas confirmés plutôt que sur les cas suspects. En outre, elle améliorera l'identification et la prise en charge des nombreuses maladies fébriles non liées au paludisme qui y sont à tort assimilées simplement en raison de la présence de fièvre. Les efforts visant à étendre l'accès aux tests de diagnostic rapide ont pris du retard par rapport à ceux de prévention et de lutte anti vectorielle, mais quoiqu'il en soit le renforcement du diagnostic et du traitement contribuera à réduire la mortalité imputable au paludisme **[5].**

I.6 Les médicaments efficaces actuellement en circulation

Les médicaments utilisés dans le traitement de première intention sont la quinine (Quinimax, Quinine Lafran), l'halofantnine (Halfan), la méfloquine et la chloroquine, comme indiqué dans le **tableau I.1,** La quinine reste le médicament de référence dans les accès graves et compliqués ou pour les patients dans l'impossibilité de recevoir un traitement per os : elle est parfois associée à la doxycycline pour le traitement d'accès provenant de pays connus pour présenter des risques élevés de résistance. On observe également que, faisant suite aux recommandations de la conférence de consensus de 1999, la quinine, et plus discrètement la méfloquine, sont de plus en plus employées dans les prises en charge thérapeutiques, au détriment de l'halofantrine, qui reste dominante en usage pédiatrique (75.0%) (2.8). L'atovaquon-proguanil, à diffusion uniquement hospitalière jusqu'en octobre 2001 est encore peu utilisée mais représentera une alternative pour le traitement des accès simples à *plasmodium falciparun.*

La durée de l'hospitalisation varie suivant le type et la gravité de l'accès de 1 à 50 jours avec une valeur médiane de 3 jours pour les accès simples et de 7 jours pour les accès graves. Le cout médian global *a minima* d'un accès palustre simple a pu être chiffré (en 1997) à environ 5000€ en hospitalisation et à 1 500 € en ambulatoire **[6,7].**

Tableau I.1. Evolution des pourcentages d'utilisation des traitements curatifs.

Années	1998	1999	2000
Halofantrine	54.1	43.4	24.6
Quinine	25.9	32.9	45.7
Méfloquine	8.5	13.2	19.5
Chloroquine	8.2	7.3	8.1

I.7 Les méthodes de conception des médicaments

Conception de médicaments

Le médicament, de sa conception à sa commercialisation, en passant par sa fabrication, est soumis à une législation spécifique et stricte visant à garantir la qualité, l'efficacité et la sécurité requises.

La phrase de conception actuelle d'un médicament

Les démarches de conception pour tous sont principalement pointées sur les tâches à réaliser, l'utilisateur est souvent réduit dans le meilleur des cas, à des caractéristique cognitives, sensorielles et motrice. Les facteurs esthétiques, hédoniques et émotionnels dans sa relation avec les produits sont peu, voire pas pris en compte. Pour nous, les méthodologies de conception pour tous peuvent être améliorées afin de rendre plus acceptables et désirables, d'un point de vue fonctionnel, social et émotionnel les produits conçus pour tous **[8].**

Le processus de découverte de médicament a un succès limité malgré tous les progrès réalisés. En effet, on estime actuellement que le développement d'un médicament nécessite environ 1.8milliard de dollars américains sur environ 13ans **[9].**

Ce processus est scindé en trois phases : une première phase dédie à la recherche exploratoire, une seconde phase englobe les études précliniques et une troisième phase consacrée à la recherche clinique **[10].**

La première phase : la recherche exploratoire

Les recherches, qu'elles soient de nature quantitative ou qualitative, visent la description, l'exploration, l'évaluation, l'explication ou la prédiction u monde-vie (Schultz) tel que nous le percevons et le construisons socialement. Camper la recherche qualitative dans le domaine exploratoire représente une amputation de son pouvoir d'intelligibilité de la réalité. Il est plutôt souhaitable de qualifie l'ambition descriptive, exploratoire, évaluative, explicative ou prédictive d'une recherche.

Identification de la cible

Cs cibles sont nombreuse, On y dénombre plusieurs familles :

- Les enzymes visées par environ 25% des molécules active actuelles. Ces molécules sont essentiellement des inhibiteurs de l'activité enzymatique.
- Les pompes, transporteurs et canaux ionique, protéines membranaires qui régissent les équilibres transmembranaires des principaux ions, visés qui environ 15% des molécules actives, des inhibiteurs des mouvement ioniques.
- Les récepteurs membranaires visés par environ 40% des molécules médicamenteuses, agonistes ou antagonistes.
- Les récepteurs nucléaires visés par environ 10%des molécules actives, agonistes ou antagonistes.

Ces quatre familles de cibles correspondent à près de 90% des médicament. Il faut y ajouter 5% des molécules se liant à des cibles diverses

Microtubules qui lient la colchicine, et les alcaloïdes de la pervenche. Enfin, les cibles d'environ 5% des molécules actives reconnues par déterminées, par exemple la cible du paracétamol **[11].**

Le criblage

Lors de la conception de médicament, l criblage in silico est de plus en plus utilisé et les méthodes disponibles nécessitent d'être évaluées. L'évaluation de 8 méthodes a mis en évidence l'efficacité des méthodes de criblage in silico et des problèmes de construction de la banque d''évaluation de référence (DUD), la conformation choisie pour les sites de liaison n'étant pas toujours adaptée à tous les actifs. La puissance informatique actuelle le permettant, plusieurs structures expérimentales ont été choisisses pou tente de mimer la flexibilité désistes de liaison. Un autre problème a été mis en évidence : les métriques d'évaluation des méthodes souffrent e biais. De nouvelles métriques ont once été proposées, telles que BEROC et RIE. Une autre alternative est proposée ici, mesurant la capacité prédictive d'une méthode en actifs. Enfin, une petite molécule active sur le TNFa*in vitro* et *in vivo* sur souris a été identifiée par un protocole de criblage in silico. Ainsi, malgré le besoin d'amélioration des méthodes, le criblage in silico peut être d'un important soutien à l'identification de nouvelles molécules a visée thérapeutique **[12].**

Leurs applications sont polyvalentes : elles permettent d'identifier des candidats médicaments pour des cibles thérapeutiques connues, d'anticiper des effets secondaires

potentiels, ou de proposer de nouvelle indication thérapeutique pour des médicaments connus. Cette thèse est conçue selon deux cadres d'approches dans lesquelles les données sont décrites numériquement sur la base des connaissances des experts, et les approches basées sur l'apprentissage automatique de la représentation numérique à partir du graphe moléculaire et de la séquence protéique.

Nous discutons ces approches et les appliquons pour guider la découverte de médicaments. **[13].**

Seconde phase : études précliniques

Un vaccin contre plasmodium falciparum est une grande priorité, il est difficile de tester les candidats vaccins en essai clinique, et sans un bon modèle animal, des tests fonctionnels d'anticorps in vitro sont nécessaires pour choisir les meilleurs candidats. Ce travail décrit un test mesurant l'activation des neutrophiles (poussée respiratoire) et la phagocytose en présence d'anticorps antimérozoïte, par chimiluminescence. Ce test est reproductible et relié à la protection clinique en zone d'endémie.

Les anticorps spécifiques de PfMSP1p19 et PfMSP4p20 exprimé de baculovirus sont fonctionnels dans ce test. Avec une reproductibilité confirmée ailleurs, il pourrait devenir un outil de référence. Une autre approche immuno-épidémiologique démontre une corrélation entre le taux d'anticorps anti-PfMSP1p19 ou anti-PfMSP5 et la protection clinique. **. [14]**

Troisième phase : Recherche clinique

Nous présentons une méthodologie de co-analyse du travail que nous appelons « autoconfrontationcroisée ». Elle est fondée sur la distinction entre activité réalisée et réel de l'activité et vise à identifier les développements possibles ou empêchés de l'activité pour éventuellement en transformer le cours.

Une première phase de co-conception du milieu de travail précède les enregistrements vidéo des tâches accomplies par les opérateurs, puis de ce qu'ils font de ce qu'ils ont dit. A partir des variantes de genre professionnel, chaque operateur analyse le style de ses actions. Confrontés par binôme à leur activité, ils s'engagent alors dans des controverses professionnelles portant sur les genres à partir des styles de leurs actions. L'autoconfrontation croisée, indissociable d'une clinique de l'activité, vise aussi à renouveler l'analyse du travail. Elle nous permet d'approcher la santé comme pouvoir d'action du sujet sur son milieu et sur lui-même **[15].**

I.8 La nécessité d'utiliser l'outil informatique dans la conception du médicament

Le criblage virtuel (VS) a émergé dans la découverte de médicaments comme un outil informatique puissant pour cribler de grandes bibliothèques de petites molécules pour de nouveaux hits avec les propriétés souhaitées qui peuvent ensuite être testées expérimentalement.

Semblable a d'autres approches informatiques, l'intention de VS n'est pas de remplacer les tests in vitro ou in vivo, mais d'accélérer le processus de découverte, de réduire le nombre de candidats à tester expérimentalement et de rationaliser leur choix. De plus, VS est devenu très populaire dans les sociétés pharmaceutiques et les organisations universitaires en raison de son économie de temps, de couts, de ressources et de main-d'œuvre. Lorsque la cible biologique n'est pas connue ou lorsque sa structure 3D ou le mode de liaison des ligands ne sont pas disponibles, l'analyse quantitative de la relation structure-activité (QSAR) peut être utilisée dans les VS en raison de leur grande vitesse de criblage et de leur bon taux de réussite. Cette approche est connue sous le nom VS basé sur QSAR. Dans cette méthode, les composés ayant des propriétés biologiques connues sont collectés à partie de bases de données et de la littérature. Ensuite, après la conservation des données, les descripteurs chimiques sont calculés à différents niveaux de représentation de la structure moléculaire, allant de 1D à nD, ou n= 0-7, puis corrélés avec la propriété biologique à l'aide de méthodes d'apprentissage automatique. Une fois développés et validés, les modèles QSAR sont appliqués pour prédire la propriété biologique de composés non testés. Bien que la validation expérimentale des résultats de calcul ne représente pas une partie de la méthodologie QSAR, cela devait être effectué comme une étape finale de cette approche. Dans cette mini-revue, nous résumons et analysons de manière critique les tendances récentes des approches VS basées sur QSAR dans la découverte de médicaments, et démontrer des applications réussies dans l'identification de composés en perspective avec les propriétés souhaitées. De plus, nous fournissons quelques recommandations sur les meilleures pratiques pour le dépistage virtuel basé sur QSAR ainsi que les perspectives futures de cette approche **[16].**

Conclusion

Le paludisme est une des causes principales de la mortalité infantile en Afrique, où son agent, le Plasmodium falciparum, représente une des infections majeures de transmissibilité transfusionnelle. Des études menées sur les continents africains ont apporté des données sur sa prévalence et ses facteurs de risque, sur ses manifestations cliniques chez le donneur de sang et sur les possibilités de prévention de la contamination transfusionnelle. La symptomatologie peut être discrète ou inexistante chez e nombreux donneurs de sang porteur du parasite, ce qui limite considérablement l'efficacité de la sélection des candidats au don lors de l'entretien précédant le don. De surcroît, la limitation des ressources biologiques sécuritaires et l'approvisionnement insuffisant en sang sur le continent rendent également problématique l'identification et l'exclusion des donneurs parasitérniques. La destruction du paludisme dans la poche de sang et le traitement prophylactique de receveur sont des solutions possibles, à côté de la prévention générale du paludisme auprès des populations en contact avec ses vecteurs.

Chapitre II

Etude quantitative de la relation structure-activité

QSAR

II.1 Introduction

La relation quantitative structure-activité (QSAR) est l'un des outils les plus pratiques en physico-chimie computationnelles. L'idée fondamentale de QSAR consiste en la possibilité d'une relation entre un ensemble de descripteurs, et une réponse moléculaire. QSAR peut être considérée comme une règle calculée par ordinateur qui décrit quantitativement l'activité biologique en termes de descripteurs chimiques, il a été fréquemment utilisé pour prédire les activités biologiques des nouveaux composés **[17].**

La prédiction de l'activité biologique est basée sur des calculs mathématiques et statistiques, c'est pourquoi nous choisissons deux méthodes statistiques :

- ✓ La régression linéaire multiple RLM.
- ✓ Les réseaux de neurones artificiels RNA.

II.2 Définition

Le modèle QSAR (Quantitative Structure-Activity Relationship) est une corrélation basée sur un calcul statistique, de paramètres physico-chimiques avec l'activité biologique **[18].**

La relation Quantitative Structure et Activité (QSAR) est un outil permet d'expliquer l'interaction des composés chimiques avec les sujets actuels, le principe des méthodes QSAR est comme leur nom l'indique, de mettre en place une relation mathématique reliant de manière quantitative des propriétés moléculaires, appelées descripteurs, avec une observable macroscopique (activité biologique, toxicité physico-chimique, etc.), pour une série des molécules biologiquement actives à l'aide des méthodes d'analyse de données "une équation de corrélation" **[19].**

II.3 Principe

Le principe d'une étude QSAR **(Figure II.1),** consiste à trouver une relation mathématique reliant de manière quantitative une activité biologique, mesurée pour une série de composés similaires dans les mêmes conditions expérimentales, avec des descripteurs moléculaires à l'aide des méthodes statistiques. Ces études permettent à analyser les données structurales afin de détecter les facteurs déterminants pour l'activité étudiée. Pour ce faire, différents types de méthodes statistiques peuvent être employées, les plus répondues sont : la méthode de moindre carrées partielles, la régression linéaire simple et multiple et les réseaux de neurones artificiels.

La modélisation QSAR comporte généralement trois étapes **[20]** :

- Recueillir ou, si possible, concevoir un ensemble d'information sur la structure et les propriétés des produits chimiques.
- Choisir des descripteurs capables de relier correctement la structure chimique à l'activité biologique.
- Appliquer des méthodes statistiques qui corrèlent les changements de structure avec les changements dans l'activité biologique.

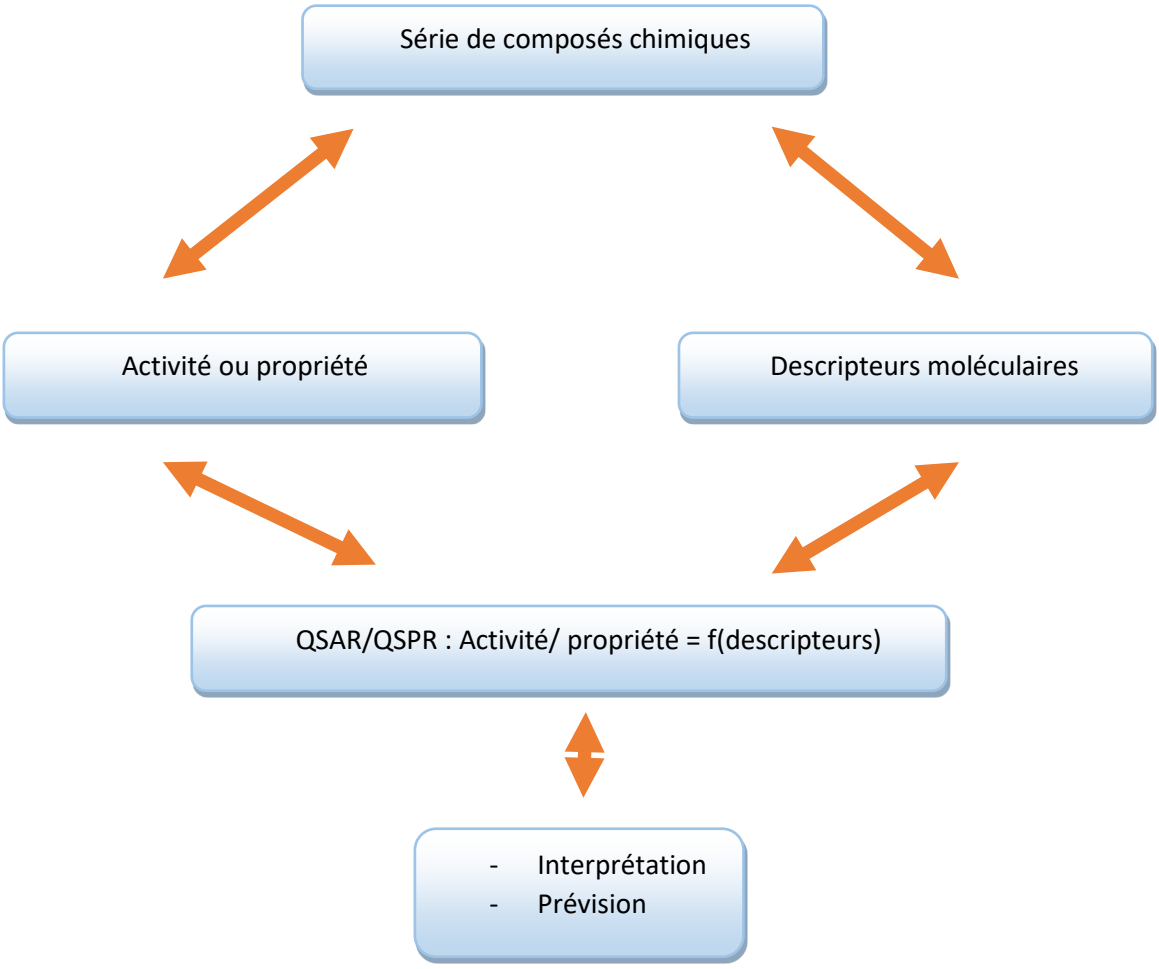


Figure II.1. Modèle de l'étude QSAR **[21]**

L'expression mathématique obtenue peut alors être utilisée comme moyen prédictif de l'activité étudiée pour de nouvelles molécules ou des molécules pour lesquelles les données expérimentales ne sont pas disponibles.

II.4 Objectifs

Les objectifs pratiques d'une QSAR sont nombreux et ces techniques sont largement utilisées dans de nombreuses situations. Le but des études in silico comprend donc ce qui suit [**22**] :

- Prédire l'activité biologique et les propriétés physico-chimiques par des moyens rationnels.
- Comprendre les mécanismes d'action au sein d'une série de produits chimiques.
- Economies sur les couts de développement de produits.
- Les prévisions pourraient réduire la nécessité de tests sur les animaux longs et couteux.

II.5 Applications de QSAR

Il existe un grand nombre d'application des modèles QSAR en milieu industriel, dans le domaine de la recherche universitaire, en économie, en prévision météorologique…etc. **[23].**

Nous rapportons ci-dessous quelques applications possibles des modèles QSAR **[24,25] :**

- L'identification rationnelle de nouvelles pistes avec :
- L'optimisation de l'activité pharmacologique, biocide ou pesticide.
- La conception rationnelle de nombreux produits tels que des agents tensioactifs, des parfums, des colorants et des produits chimiques fins.
- L'identification des composés dangereux aux premiers stades de développement.
- La prédiction de la toxicité et les effets secondaires de nouveaux composés.
- La sélection des composés ayant des propriétés pharmacocinétique optimales, que ce soit la stabilité ou la disponibilité dans les systèmes biologiques.

II.6 Activité thérapeutique

Il a été rapporté que les composés d'imidazopyridazine possèdent de bonnes activités inhibitrices de protéine kinase 7 de p. falciparum pfpk7 **[26,27]**. Pfpk7 est exprimé à plusieurs stades du cycle de vie du parasite y compris les stades asexuel et sexuel chez l'homme et aussi

chez le moustique, mais n'a pas d'homologue humain. Il représente pour l'intervention de petites molécules dans le cadre d'un traitement antipaludique **[28]**. Un certain nombre d'études quantitatives de relation structure-activité QSAR liées à la conception de médicaments antipaludiques a également été rapportées jusqu'à présent **[29,30]**, mais une étude systématique QSAR doit encore être réalisée pour les imidazopyridazines en tant qu'inhibiteurs de pfpk7.

La chimie computationnelle est devenue un contributeur important à la conception rationnelle des médicaments. L'approche QSAR permet de corréler les activités biologiques spécifiques ou les propriétés physiques d'une série de composés avec les propriétés moléculaires mesurées ou calculées des composés, en termes de descripteurs **[31,32]**. Les méthodologies QSAR économisent des ressources et accélèrent le processus de développement de nouvelles molécules et médicaments.

Cette étude vise à élucider les caractéristiques structurelles des dérivés d'imidazopyridazine nécessaires à l'inhibition de pfpk7 et à obtenir des QSAR 2D et 3D prédictifs.

II.7 Descripteurs moléculaires

Les descripteurs moléculaires jouent un rôle fondamental dans les études de la relation quantitative structure activité /propriété. Ils sont utilisés en tant que variables indépendantes pour prédire une variable dépendante (activité).

L'utilisation des descripteurs moléculaires dans le développement de modèles QSAR n'est pas une tâche aisée. Tout d'abord, un très grand nombre de descripteurs moléculaires, de différentes complexités et de conceptions diverses a été introduit au cours des dernières années. Ensuite, pendant ce temps, aucune règle stricte n'a été établie pour la sélection de descripteurs adaptés parmi le grand nombre de descripteurs disponibles. Ce choix a souvent été basé sur l'intuition chimique des chercheurs, ou en se pliant à la tradition **[33].**

Dans ce qui suit, nous allons présenter que la classification des descripteurs moléculaires les plus utilisés

- **Les descripteurs 1D**

Sont accessibles à partir de formule brute de la molécule et décrivent des propriétés globales du composé comme le nombre d'atomes et la masse moléculaire, …etc.

Ces descripteurs sont couramment utilisés du fait de leur extrême simplicité. Cependant, ils peuvent poser problème pour une bonne interprétation des mécanismes d'interaction du fait qu'ils ne permettent pas de tenir en compte des effets stériques et d'isomérie **[34]**.

- **Les descripteurs 2D**

Ce sont les descripteurs ou des indices topologiques qui prennent en compte l'arrangement atomique interne des composés. Ils sont calculés à partir de la formule développée de la molécule. On distingue :

a) **Les indices 2D constitutionnels** : qui caractérisent les différents composants de la molécule. Il s'agit par exemple du nombre de liaisons simples ou multiples, du nombre de cycles…etc.
b) **Les indices 2D topologiques** : prouvent être obtenus à partir de la structure 2D de la molécule, et donnent des informations sur sa taille, sa forme globale et ses ramifications.

Exemples : indice de Wiener **[35]**, indice de Randié **[36]**, indice de connectivité de valence de Kier-Hall **[37]**, indice de Balaban **[38]** …etc.

Ces descripteurs 2D permettent de prédire les propriétés physiques mais sont insuffisantes pour expliquer certaines propriétés et activités biologiques comme la toxicité.

- **Les descripteurs 3D**

Sont évalués à partir des positions relatives des atomes dans l'espace, et décrivent des caractéristiques plus complexes. Leurs calculs nécessitent donc de connaitre la géométrie 3D de la molécule **[34]**.

a) **Les descripteurs 3D géométriques** : les plus importants sont le volume moléculaire, la surface accessible au solvant, le moment principal d'inertie.
b) **Les descripteurs 3D électroniques** : permettent de quantifier les différents types d'interactions inter- et intramoléculaires, de grande influence sur l'activité biologique des molécules. Le calcul de la plupart de ces descripteurs nécessite la recherche de la géométrie pour laquelle l'énergie est minimale, et fait souvent appel à la chimie quantique.

c) Les descripteurs 4D

Ils correspondent à la mesure des propriétés 3D (potentiel électrostatique, d'hydrophobicité, de liaison hydrogène…) d'une molécule en tout point de l'espace. Ils permettent d'avoir l'information sur la structure de la cible (protéine). On pourra ainsi distinguer les descripteurs 4D qui nécessitent un alignement de la molécule guidé par l'étude des complexes ligand-cible (ou, au moins, par des contraintes visant d'optimiser le recouvrement spatial des champs électriques et stériques des ligands, faute d'information sur le vrai mode de fixation dans la cible) avant d'être calculés. Ces descripteurs sont obtenus par le calcul des chams d'interactions moléculaires (CoMFA, CoMSIA) entre une molécule et une sonde représentée par une molécule (eau, amide, …). **[39,40]**

II.8 Les différentes approximations utilisées

II.8.1 La régression linéaire multiple RLM

La régression linéaire multiple (RLM) est une méthode statistique utilisée pour quantifier la relation entre les variables indépendantes obtenues par calcul et une variable dépendante données par l'expérience. Cette méthode repose sur un ensemble de paramètres statistiques qui définissent la combinaison linéaire entre les variables indépendantes (descripteurs quantiques) et l'activité biologique des molécules étudiées **[41].**

Le problème de la régression linéaire est de déterminer les paramètres d'un modèle reliant Y à X, à partir de mesures {(Y1, X1), (Y2, X2) ……, (Yn, Xn)}. Le cas simple du modèle du premier ordre avec une seule variable s'écrit **[42]** :

$$Y = \boldsymbol{\beta} 0 + \boldsymbol{\beta} 1 X$$

Le calcul des paramètres $\boldsymbol{\beta}_0$ et $\boldsymbol{\beta}_1$ est réalisé en minimisant l'erreur commise sur leur estimation.

II.8.2 Réseaux de neurones artificiels

Les réseaux de neurones artificiels sont devenus en quelques années des outils précieux dans des domaines très divers de l'industrie et divers services. Ils sont particulièrement utilisés pour

résoudre des problèmes de classification, de prédiction, de reconnaissance des formes, de catégorisation, de mémoire associative et d'optimisation **[43].**

Réseau de neurones artificiels (ANN) est une technique de reconnaissance qui ressemble de près le fonctionnement interne du cerveau qui est essentiellement composé de neurones interconnectés. Ce phénomène est affecté par la conception architecturale de ANN ou les unités neuronales sont interconnectés les uns aux autres.

Une architecture couramment utilisé comme représenté dans la figure contient un réseau d'anticipation à trois couches qui est composé d'une couche d'entrée, une couche cachée, et une couche de sortie. La couche d'entrée passe essentiellement des informations des variables indépendantes dans le système ANN : par conséquent, le nombre d'unités de neurones présents dans la couche d'entrée est égal au nombre de variables indépendantes dans l'ensemble de données. Les connexions entre les neurones sont affectées des valeurs numériques connues comme poids.

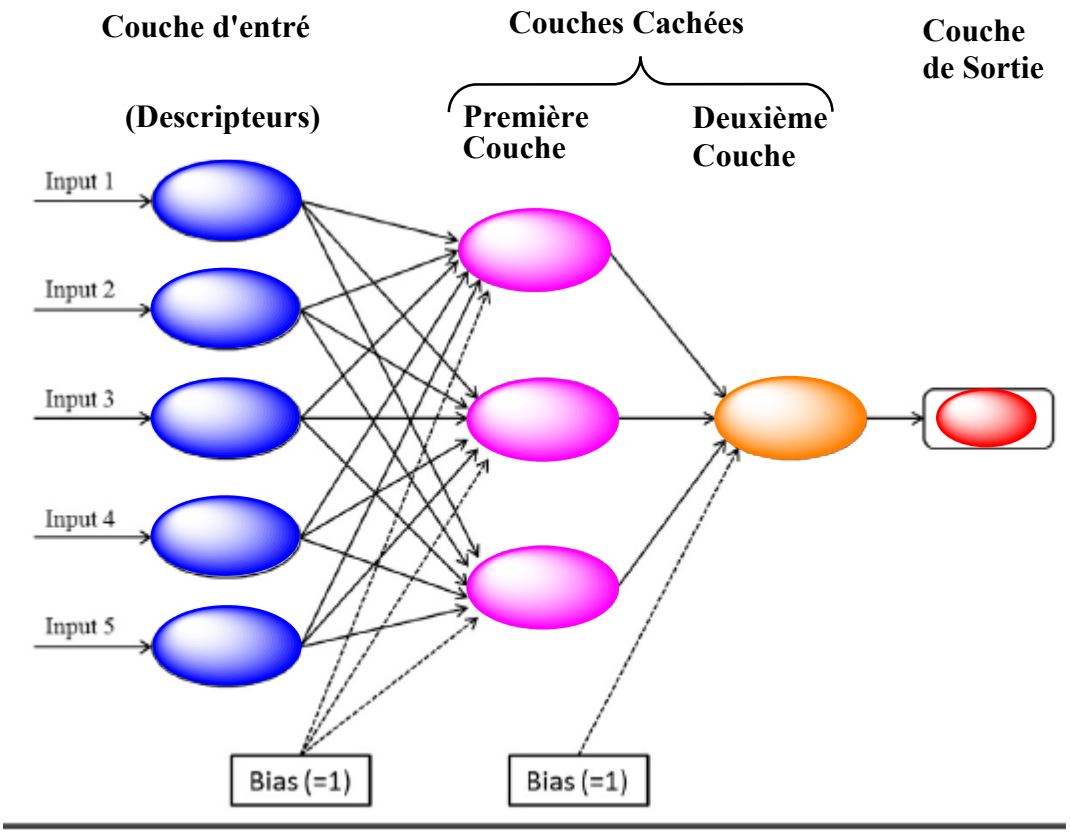


Figure II.2. Représentation d'un réseau de neurones

Les informations provenant de la couche d'entrée sont transmises à la couche cachée pour la configuration de traitement seront ensuite transmises de la couche cachée à la couche de sortie. Dans un algorithme de rétropropagation, l'erreur calculée est dérivée de la différence entre la valeur prédite et la valeur réelle, et si elle est acceptable, le processus d'apprentissage ne cesse autrement signaux seront envoyés vers l'arrière de la couche cachée à un traitement ultérieur et réajustements de poids. Ceci est réalisé de manière itérative jusqu'à ce qu'une solution soit trouvée et l'apprentissage est terminé.

Afin de répondre à la problématique de cette recherche, telle que la prédiction des paramètres électroniques (E_{HOMO}), (gap), (μ), (E_T), (E_a), (λ_{max}) d'un ensemble de structures chimiques de différentes familles qui contiennent une activité biologique recherchés, nous avons choisi l'architecture de perceptrons multicouches PMC en utilisant l'algorithme de Levenberg-Marquardt avec un couple de fonction comme fonction comme fonction de transfert.

II.9 Le principe des approximations utilisées

II.9.1 Le principe de MLR

Le modèle linéaire a été développé en appliquant MLR. La technique MLR s'est avérée être une technique multidisciplinaire d'une applicabilité précieuse pour établir des modèles

Prédictifs QSAR par différents groupes de recherche **[44].**

Un modèle MLR est construit en supposant qu'une relation linéaire existe entre un entre

Un ensemble de descripteurs moléculaires d'un composé (qui est représenté par un

Vecteur caractéristique (X) avec chaque descripteur comme composant) et une activité

Thérapeutique (Y) **[45].**

II.9.2 Le principe de RNA

Un réseau de neurones est constitué d'un très grand nombre de petites unités de traitement identiques appelées neurones artificiels.

Chacun de ces neurones est par ailleurs fort complexe. Essentiellement, il s'agit de tissu vivant et de chimie. Les spécialistes des neurones biologiques commencent à peine à comprendre quelques-uns de leurs mécanismes internes. On croit en général que leurs différentes fonctions neuronales, y compris celles de la mémoire, sont stockées au niveau des connexions (synapses)

entre les neurones. C'est ce genre de théories qui a inspiré la plupart des architectures de réseaux de neurones. L'apprentissage consiste alors soit à établir de nouvelles connexions, soit à en modifier des existences **[46]**.

L'origine des réseaux de neurones vient de l'essai de modélisation du neurone biologique par Warren McCulloch et Waiter pitts **[47]**. Ils supposent que l'impulsion nerveuse est le résultat d'un calcul simple effectué par chaque neurone et que la pensée née grâce à l'effet collectif d'un réseau de neurone.

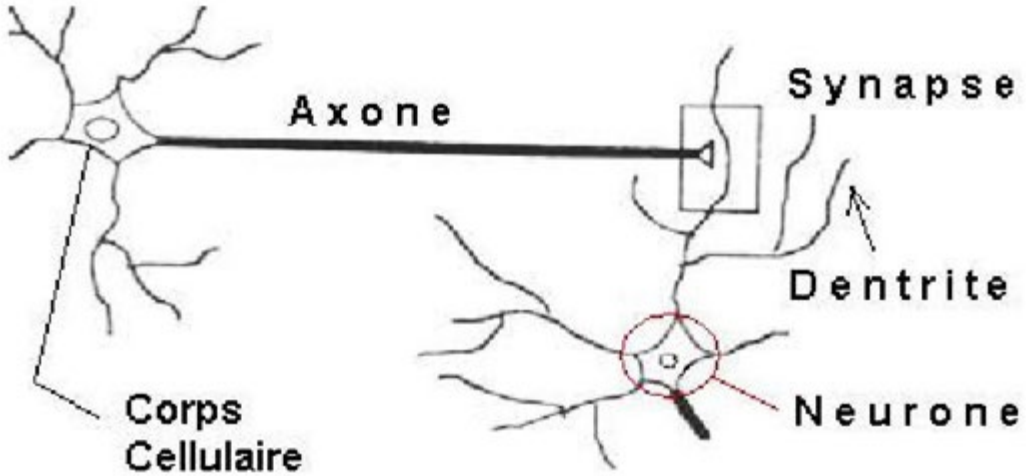


Figure II.3. Schéma représentatif de la structure du neurone.

II.10 Les facteurs

II.10.1 Coefficient de corrélation (R)

Le coefficient de corrélation est un coefficient statistique permettant de mettre en évidence, une liaison entre deux types de séries de données statistique. Alors pour savoir s'il existe ou non une relation linéaire entre les variables X_i et Y_i, on calcule le coefficient de corrélation de Pearson **r**, donné par la relation mathématique **[48]** suivante :

$$r_p = \frac{\sum_{i=1}^{N}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

Ou **n** est le nombre de d'observations, et $\sum i = 1$ à n.

La valeur de r doit être comprise entre +1 et -1. Plus elle est proche de +1 ou de -1 plus grande la probabilité qu'il existe une relation linéaire entre les variables x et y, valeur proche de +1 indique que la pente de la corrélation est positive, une valeur proche de -1, indique qu'elle est négative, pour une valeur tendent vers zéro indique qu'il n'existe pas de relation linéaire entre x et y (mais elles peuvent reliées par une autre fonction) **[49].**

II.10.2 Coefficient de détermination R^2

Pour évaluer la précision avec laquelle la dépendance trouvée la variance de la variable dépendante (c'est-à-dire la qualité de l'ajustement statistique), le carré du coefficient de détermination (R^2) est utilisé. On définit alors le coefficient de détermination, qui mesure la part de variance expliquée par le modèle dans la variance totale **[50,51].**

$$\mathbf{R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}}$$

II.10.3 Le coefficient de détermination ajusté R^2 adj

Ce coefficient est utilisé en régression multiple par ce qu'il tient compte du degré de liberté **[52]** :

$$\mathbf{R^2_{adj} = \sqrt{\frac{R^2(n-1)-p}{n-p-1}}}$$

Avec :

n est le nombre des observations (les molécules)

p est le nombre de variables indépendantes (les descripteurs)

R^2 est le coefficient de détermination du modèle **[53]**

II.10.4 Coefficient de corrélation de validation croisée R^2_{CV}

La procédure statistique cross-validation peut être utilisée pour évaluer prédictif des modèles QSAR, le coefficient qui décrit la validation est donné par l'équation ci-dessous **[54-56].**

$$R^2_{CV} = 1 - \frac{\sum_i (y_i^{pred} - y_i^{abs})^2}{\sum (y_i^{abs} - y^{mean})^2}$$

Ce coefficient peut être calculé à partir de PRESS comme suit

$$R^2_{CV} = 1 - \frac{PRESS}{SStotal}$$

II.10.5 Ecart type (s)

L'écart type (s) **[57]** est un autre paramètre habituellement rapporté, il indique dans quelle mesure la fonction de régression prédite les données observées, ce paramètre est donné par :

$$S = \sqrt{\frac{RSS}{n-p-1}}$$

Dont p est le nombre de variables indépendants.

II.10.6 Fischer-Snedecor

Le test Fischer permet de justifier la liaison globale entre y_n et x_j, une version dérivée de cet indicateur peut juger du degré de pertinence des variables du modèle. Il s'agit de vérifier pour chaque variable X_j si, alors lorsqu'on passe du modèle complet à p prédicateurs au modèles simplifié obtenu en t mesure en fait l'apport marginal de variable X_j à l'explication de Y_n.

On peut définir quelques paramètres **[58-60]** utilisés dans la régression multilinéaire.

Somme des carrés totaux ; $TSS = \sum (y_{abs} - \ddot{y})^2$

Somme des carrés expliqués : $ESS = \sum (y_{cal} - \ddot{y})^2$

Somme des carrés résiduels : $RSS = \sum (y - y_{cal})^2$

Ainsi, TSS = ESS +RSS

F peut s'écrire comme suit **[61]** :

$$\mathbf{F = \frac{ESS}{P} = \frac{n-p-1}{RSS}}$$

II.10.7 Coefficient de prédiction

Si R^2 est une mesure de la justesse de la corrélation, Q^2 est une mesure de la justesse de prédiction. Il mesure la capacité prédictive d'un modèle et donnée par **[62]** :

$$\mathbf{Q^2 = 1 - \frac{PRESS}{SSY}}$$

II.10.8 Validation des modèles

Quel que soit le but ultime du modèle QSAR, il doit être obligatoirement validé avant d'être interprété ou utilisé à fins prédictives. Il existe différentes solutions pour s'assurer de validité d'un modèle.

La validation croisée LOO (leave-one-out) est un processus qui permet de tester la précision prédictive d'un modèle, cette méthode est basée sur le calcul de quelques paramètres statistiques tels que :

- ✓ La somme des carrés résiduels RSS (PRESS)
- ✓ Somme des carrés totaux TSS
- ✓ Qualité d'ajustement R^2_{adj},
- ✓ Coefficient de corrélation de validation R^2_{CV}
- ✓ La validation standard des erreurs de prédiction (SPRESS)
- ✓ L'erreur de prédiction (PE)

Ces paramètres statistiques **[63-68],** sont calculés à partir des relations suivantes :

- **PRESS** = $\sum(Y_{abs} - Y_{cal})^2$
- **TSS** = $\sum(y_{abs} - \ddot{y})^2$
- $\mathbf{R^2 = (1 - r^2)\left(\frac{n-1}{n-p-1}\right)}$
- $\mathbf{R^2_{CV} = 1 - \frac{PRESS}{TSS}}$

- **SPRESS** = $\sqrt{\frac{PRESS}{n}}$
- **PE = 0.6745 (1-r²) / $\sqrt{n}$**

II.10.9 Colinéarité des variables explicatives

Quand les variables sont très corrélées et donc quasi-colinéaires, ce qui rend le déterminent de XX proche de 0 : on dit que le système est mal conditionné. L'inversion de la matrice conduit alors à des estimations ayant une variance très importante, voir même parfois à des problèmes numériques. Il est donc important de diagnostiquer de tels problèmes.

Nous essayons de donner des outils de diagnostics. Les solutions sont d'autres méthodes telles que la régression « Ridge », la méthode PLS « partial least-square » et la régression sur composante principale.

- Matrice de corrélation : elle permet de détecter des fortes corrélations entre deux variables.
- Facteur d'inflation de la variance VIF **[69]** : on définit le facteur de la variance (VIF) par :

$$\mathbf{VIF}=\frac{1}{1-R_j^2}$$

R_j^2 est le coefficient de détermination de la régression de la variable X_i sur les autres variables.

Ou Plus X_j est linéairement proche des autres variables, plus R_j^2 est proche de 1 et le VIF grand. L'avantage du VIF par rapport à la matrice de corrélation est qu'il prend en compte des corrélations multiples **[70,71].**

Chapitre III

Résultats et Discussion

III. Introduction

La régression linéaire multiple est parmi les méthodes statistiques la plus utilisée. du fait de la simplicité de ses modèles linéaire et de la facilité de l'interprétation de ses résultats **[72].**

Cependant cette méthode présente des limitations, car elle dépend des lois et des hypothèses statistiques., notamment dans le cas où les données ne sont pas distribuées normalement **[73]** ; Les réseaux de neurones ont été présentés comme une méthode alternative aux méthodes statistiques potentiellement intéressants lorsque les effets des variables sont non linéaires **[74]**. Les réseaux de neurones présentent plusieurs avantages tels que la meilleure capacité de prédiction., d'adaptation et de généralisation au-delà de l'échantillon étudié. Une meilleure stabilité des coefficients par une moindre sensibilité aux points aberrants et absence d'hypothèses sur la distribution des variables comme la normalité en régression linéaire par exemple.

Dans ce chapitre nous présentons une étude non linéaire de la relation quantitative structure-activité inhibitrice de protéine Kinase plasmodium falciparum par les dérivés d'imidazopyridazine. Nous avons fait appel aux réseaux de neurones et aux algorithmes génétique en tant que méthodes d'apprentissage et de sélection de variables respectivement. L'architecture du réseau de neurones utilisée dans cette étude est constituée de trois couches : Couches d'entrée. Couche cachée. Couche de sortie. Les résultats obtenus sont analysés. Validés avec ceux obtenus par la régression linéaire multiple.

III.1 Partie 1

III.1.1 La synthèse des dérivés imidazopyridazine

Les produits appartenant à la famille des pyridopyrimidines sont caractéristisés par leur intense utilisation dans le domaine pharmacologique, ce qui a poussé différentes équipes de recherche à les étudier chimiquement et biologiquement. Dans ce travail, nous sommes intéressés au groupe des pyridopyrimidines et plus particulièrement à savoir les pyrido [3.2-d] pyrimidines. Les composés ciblés sont synthétisés à partir de la 2. 7-dichloropyrido [3. 2-4] pyrimidine, via des substitutions nucléophiles aromatiques et des couplages pallado-catalysés et ce dans le but d'obtenir de puissants inhibiteurs de kinases. Ce but a été atteint puisqu'en effet plusieurs des molécules

élaborées inhibent les kinases testées avec des concentrations de l'ordre du nanomolaire. Des résultats pharmacologiques aussi concluants nous ont amenés à étendre nos études à d'autres pyridopyrimidines. A savoir les pyrido [2. 3-d] pyrimidines ainsi qu'à d'autres types de bicycles polyazotés, en l'occurrence les imadazo [1. 2-a] pyridines et les imidazo [1. 2-b] pyridazines.

- **La structure**

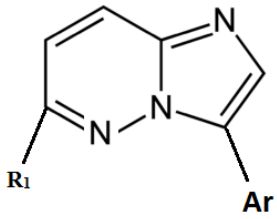


Figure III.1. Structure d'imidazopyridazine

Tableau III.1. Dérivés d'imidazopyridazine utilisés et valeurs des concertations inhibitrices expérimentales.

N	Structure	PIC_{50}	N	Structure	PIC_{50}
1	C27H28N6O	-1.06446	**17**	C20H18FN5O	-1.62325
2	C24H22N6O	-0.89763	**18**	C20H17FN6O	-1.53148
3	C18H18N6O	0.09151	**19**	C19H22N4O	-1.74819
4	C19H19N5O	-0.10037	**20**	C18H23N5O	-2.01284
5	C19H19N5O	-0.02938	**21**	C17H21N5O	0.35655
6	C19H19N5O	0.12494	**22**	C19H21N5O	0.00000
7	C19H19N5O	0.23657	**23**	C18H22N4O	0.00000
8	C17H17N5O	0.55284	**24**	C18H21N5O	-0.14613
9	C16H15N5O	0.11351	**25**	C18H19N5O	0.08619
10	C15H13N5O	-0.86923	**26**	C17H15N5	0.88606
11	C16H15N5O	-1.39794	**27**	C18H17N5O	0.58503
12	C16H15N5O	-1.53148	**28**	C20H15FN6	0.65758
13	C19H23FN6O	-2.01703	**29**	C20H16N6	0.46852
14	C17H16FN7O	-2.04532	**30**	C19H14N6	-0.32222
15	C21H21FN6O	-0.04139	**31**	C17H14N6	-0.90309
16	C19H17FN6O	-1.32222			

III.1.2 Définition des logiciels utilisés

- ChemOffice

ChemOffice est une suite intégrée et scientifiquement intelligente d'outilles de productivité personnelle qui aide les scientifiques à suivre efficacement leur travail. à mieux comprendre leurs données et à produire des rapports scientifiques de manière professionnelle et efficace **[75].**

- ChemDraw

ChemDraw est un outil permettant aux scientifiques professionnelles. aux étudiants en sciences et aux auteurs scientifiques de communiques. Il est conçu pour fonctionner selon les conventions que nous trouvées les plus intuitives pour ces utilisateurs. Notre objectif a été de faire chemDraw aussi facile à utiliser que possible tout en offrant une qualité de dessin supérieure **[76].**

- Chem3D

Chem3D est une interface de modélisation moléculaire avec des fonctions pou la construction et l'analyse de modèles de structure chimique 3D en combinaison avec le calcul de la chimie computationnelle. L'application consiste en un moteur de rendu professionnel de la structure 3D Open GL et de la surface moléculaire avec des options d'affichage communes pour les structures (boule et bâton). Liaison cylindrique, remplissage d'espace. etc.) et surfaces (solides, mailles métalliques, translucides. etc.) **[77].**

- PaDEL

PaDEl-Descriptors a été programmé en utilisant le langage Java et se composait de deux composants : un composant de bibliothèque et un composant d'interface. Le composant de bibliothèque est le véritable bourreau de travail du système et est autonome. Cela signifie qu'il peut fonctionner sans le composant d'interface et peut être facilement intégré à d'autres logiciels QSAR pour fournir la fonction de calcul des descripteurs. Le composant de bibliothèque fournit les classes wrapper autour des 43 algorithmes de descripteurs moléculaires et sept algorithmes d'empreintes digitales qui ont été implémentés dans CDK **[78].**

- CSVieweres

C'est un programme qui vous permet de visualise les fichiers DCM acquis avec les scanners et de les exporter en tant que fichiers DCM simples (fichiers ne contenant aucune information sur le patient). Fichiers STL ou fichiers PLY. Vous pouvez également afficher des fichiers STL ou PLY. Si vous avez dessiné des lignes de marge 3D. Vous pouvez afficher les lignes de marge avec la visionneuse de maillage. Le visualiser de maillage contient également un convertisseur d'orientation. qui vous permet de réorienter le maillage afin qu'il soit compatible avec le logiciel CAD/CAM spécifique d'un laboratoire **[79].**

- SPSS

SPSS (statical package for the social sciences) est un logiciel utilisé pour l'analyse statique. C'est aussi le nom de la société qui le revend (SPSS inc). En 2009, la compagnie décide de changer le nom de ses produits en **PASW**. pour *predictives analytics* Software et est rachetée par IBM pour 1.24 milliard de dollars **[80].**

- Statistica

Est le nom générique de la gamme de logiciels édités par la société StatSoft. inc.

Cette gamme de produits comprend une soixantaine de modules logiciels différents qui couvrent tout ou partie de l'analyse statistique des données, de l'exploration de données et de l'informatique décisionnelle **[81].**

III.2. Partie 2

III.2.1 Sélection des descripteurs

Après élimination des descripteurs dont la valeur ne varie pas ou peu sur l'ensemble des molécules. Une analyse multi variée suivie d'une évaluation statistique sont menées pour développer le meilleur modèle QSAR **[82].**

L'analyse de régression RLM a été utilisée pour développer le modèle et prédire de l'activité biologique à partir de descripteurs moléculaires apprenant à la série dérivés de imidazopyridazine. Notre modèle de QSAR développé est basé sur les descripteurs suivants : STN-AATSC7e. AATSC8e-C1SP3-MDEO11.

La description de cette sélectionnées pour le modèle QSAR est données dans le Tableau II.2.

Tableau II.2. Descripteur utilisée pour le développement des modèle QSAR **[83-85]**

Type	Descripteur	La description	Classe de description
2D	AATSC7e	Autocorrélation Brato. Moreau centrée moyenne. décalage 7/pondère par les électronégativités des sanderson	Descripteur d'autocorrélation
2D	AATSC8e	Autocorrélation Brato. Moreau centrée moyenne. décalage 8/pondère par les électronégativités des sanderson	Descripteur d'autocorrélation
2D	C1SP3	Autocorrélation Brato. Moreau	Descripteur d'autocorrélation
2D	STN	Autocorrélation Brato. Moreau	Descripteur d'autocorrélation
2D	MDEO11	Autocorrélation Brato. Moreau	Descripteur d'autocorrélation

III.2.2 Modèle MLR

Premièrement le MLR a été utilisé pour modéliser les relations quantitatives structure-activité avec un ensemble différent de descripteurs. Afin de construire et de tester le modèle. Un ensemble de données composé de 31 composées a été divisé au hasard en un ensemble d'apprentissage de 25 composés (80%), qui ont été appliqués pour construire le modèle et un ensemble de test de 6 composés (20%), qui ont servi à tester le modèle QSAR.

PIC_{50}= -0.611 + 0.092 **STN** + 1.020 **ATSC7e** + 1.371 **ATSC8e** – 0.290 **C1SP3** – 2.985 **MDEO11**

R=0.962. R^2=0.926. R^2_{adj} = 0.912. F=251. N=31. Q^2=0.999. S=0.418

Les descripteurs d'autocorrélation mentionnée ont été définis sur des structures moléculaires en utilisant l'électrophorèse de Sanderson e. Cette observation suggérait que les masses atomiques et la distribution d'électronégativité des atomes qui composent la molécule avaient en outre une influence significative sur l'activité anticonvulsivant de l'ensemble de données. Les scores des coefficients de régression pour chaque descripteur ont indiqué la direction de leur influence dans le modèle. De sorte que le coefficient de régression positif plus élevé associé au descripteur AATSC8e améliore le pouvoir inhibiteur des dérivés de l'imidazopyridazine.

Les facteurs d'inflation de la variance VIF entre les cinq descripteurs pertinents ont été calculés pour évaluer la multi-colinéarité. Les valeurs VIF correspondantes des cinq descripteurs sont présentées dans le Tableau III.3. Comme le montre ce tableau, tous les descripteurs ont des valeurs VIF inférieurs à cinq indiquant que le modèle obtenu a une signification statique et les descripteurs se sont avérés être raisonnablement orthogonal.

Tableau III.3. Valeur du coefficient VIF des descripteurs sélectionnés.

Descripteur	**STN**	**ATSC72**	**ATSC82**	**C1SP3**	**MDEO11**
VIF	1.471	1.452	1.817	1.876	1.648

Tableau III.4. Matrice de corrélation du modèle

	STN	**ATSC7e**	**ATSC8e**	**C1SP3**	**MDEO11**
STN	**1**				
ATSC7e	0.09	**1**			
ATSC8e	0.27	0.26	**1**		
C1SP3	0	0.22	0.52	**1**	
MDEO11	0.56	0.16	0.26	0.79	**1**

Tableau III.5. Valeur PIC_{50} expérimentale avec les descripteurs sélectionnés

N	PIC50 (exp)	PIC50 (pred)	ATSC7e	ATSC8e	C1SP3	StN	MDEO-11	Etat
1	-1.064	-1.086	0.416	-0.285	1.000	0.000	0.000	Train
2	-0.898	-1.094	-0.215	0.078	1.000	0.000	0.000	Train
3	0.092	0.297	-0.030	0.261	1.000	8.928	0.000	Train
4	-0.100	0.322	-0.198	0.357	1.000	8.960	0.000	Train
5	-0.029	0.127	-0.245	0.249	1.000	8.954	0.000	Train
6	0.125	-0.029	-0.523	0.595	2.000	8.918	0.000	Train
7	0.237	0.292	-0.162	0.405	1.000	8.888	0.000	Train
8	0.553	0.055	-0.323	0.733	3.000	8.911	0.000	Train
9	0.114	0.071	0.064	1.221	3.000	0.000	0.077	Train
10	-0.869	-0.782	-0.259	0.766	4.000	0.000	0.077	Train
11	-1.398	-1.285	-0.658	0.364	3.000	8.777	0.000	Train
12	-1.531	-1.626	-0.324	0.161	3.000	0.000	0.000	Train
13	-2.017	-1.740	-0.447	0.363	4.000	0.000	0.000	Test
14	-2.045	-2.070	-0.591	0.981	3.000	0.000	0.410	Train
15	-0.041	-0.001	0.103	0.592	1.000	0.000	0.000	Train
16	-1.322	-1.214	-0.212	0.223	2.000	0.000	0.000	Train
17	-1.623	-1.531	-0.280	-0.269	1.000	0.000	0.000	Train
18	-1.531	-1.646	-0.511	-0.007	2.000	0.000	0.000	Test
19	-1.748	-1.706	-0.456	-0.101	2.000	0.000	0.000	Train
20	-2.013	-1.960	-0.462	0.121	4.000	0.000	0.000	Test
21	0.357	0.175	0.011	0.432	2.000	8.866	0.000	Test
22	0.000	0.175	0.011	0.432	2.000	8.866	0.000	Train
23	0.000	0.299	-0.157	0.681	2.000	8.841	0.000	Test
24	-0.146	-0.052	-0.344	0.514	2.000	8.872	0.000	Train
25	0.086	-0.179	-0.174	0.273	2.000	8.896	0.000	Test
26	0.886	0.912	0.341	0.331	0.000	8.935	0.000	Train
27	0.585	0.190	0.076	-0.120	0.000	8.936	0.000	Train
28	0.658	0.656	-0.375	0.591	0.000	8.936	0.000	Train
29	0.469	0.209	-0.161	0.065	0.000	8.935	0.000	Train
30	-0.322	-0.133	-0.182	0.265	2.000	8.930	0.000	Train
31	-0.903	-0.946	-0.396	0.462	5.000	9.006	0.000	Train

III.2.3 Modèle de réseau des neurones artificiels RNA

Les modèles QSAR intrinsèquement non linéaires peuvent être développés en utilisant les réseaux de neurones artificiels.

Les Réseaux de Neurones Artificiels ont émergé pour des outils d'apprentissage pour modélisation. la classification et l'approximation de fonction **[86,87].** Dans les dernières années. Ils ont été utilisés dans la prédiction de la charge électrique pour

pallier aux problèmes trouvés dans les anciennes approches **[88]**. L'avantage des Réseaux de Neurones Artificiels est leur capacité d'apprendre du passé et de faire une généralisation **[89]** qui a rendu la prédiction de la charge électrique plus précise **[90].** Leur principe se base sur l'extraction du lien qui existe entre les entrées et sorties du système **[91].** Ils interprètent la dépendance qui existe entre les valeurs de la charge électrique (son passé) et les facteurs qui influent **[92,93].**

Sur les figures III.2 et III.3 et III.4 sont représentées les concentrations inhibitrices prédites en fonction des concentrations inhibitrices expérimentales pour l'ensemble d'apprentissage et test.

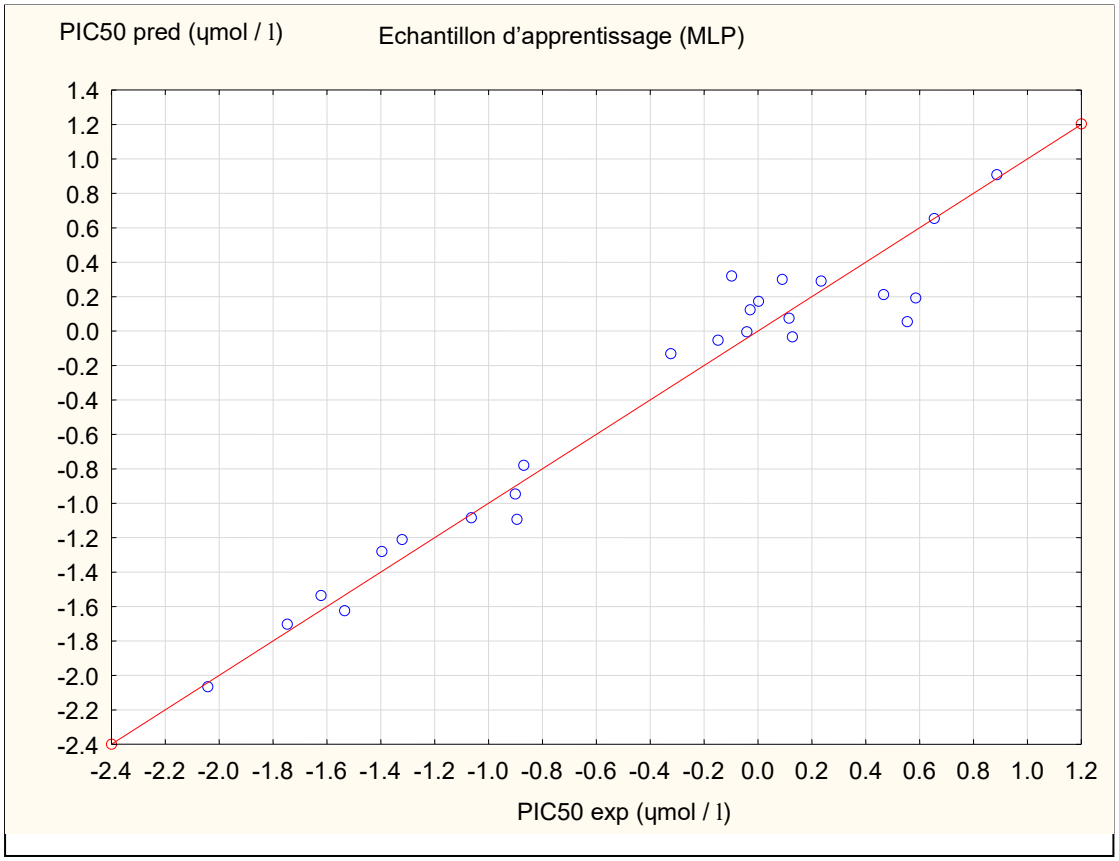


Figure III.2. Représentation graphique des concentrations inhibitrices (pIC_{50}) expérimentales et celle prédite par le modèle RNA pour la série d'ensemble de test des dérives des imidazopyridazine

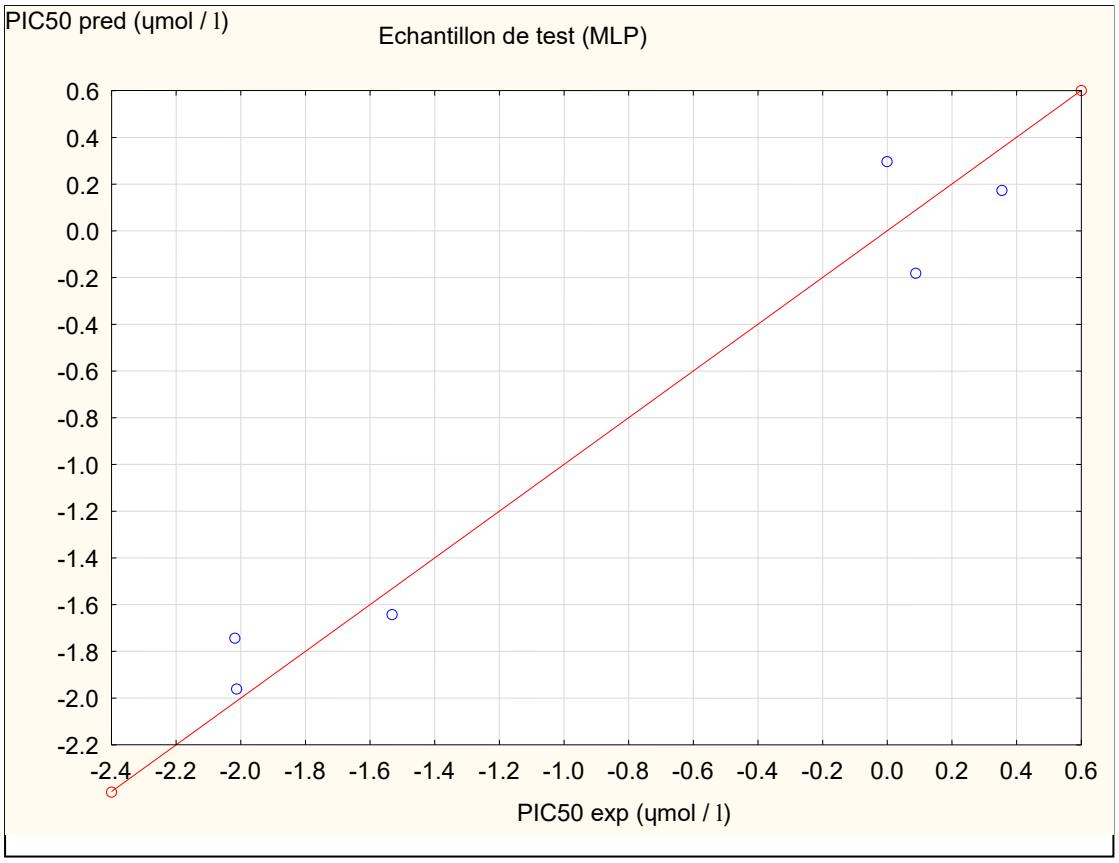


Figure III.3. Représentation graphique des concentrations inhibitrice (pIC_{50}) expérimentales et celle prédite par le modèle RNA pour la série d'ensemble d'apprentissage des dérives de imidazopyridazine.

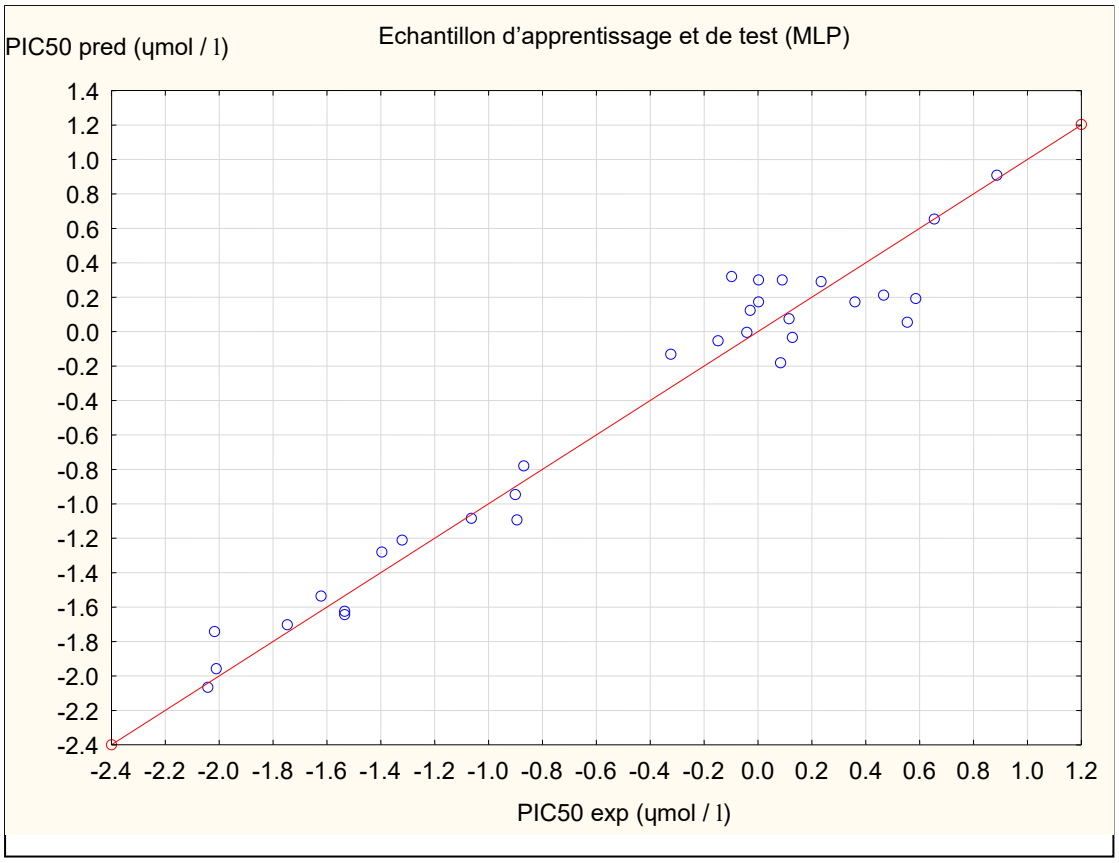


Figure III.4. Représentation graphique des concentrations inhibitrice (pIC_{50}) expérimentales et celle prédite par le modèle RNA pour la série d'ensemble d'apprentissage et test des dérives de imidazopyridazine.

D'après les Figures III.2 et III.3 et III.4, le modèle de réseaux de neurones artificiels : perceptron multicouche (MLP) a donné de bons résultats.

Le modèle ANN a suscité un grand intérêt car ses approximateurs de fonctions universelles sont capables de cartographier toutes les fonctions linéaires ou non linéaires. Le modèle de réseau neuronal perceptron multicouche (MLP) est un réseau neuronal supervisé basé sur le modèle perceptron simple original avec rétropropagation pour l'entrainement du réseau. Il se compose généralement d'une couche d'entrée de nœuds sources, d'une couche de sortie et d'une ou plusieurs couches cachées de nœuds de calcul (neurones) qui augmentent la puissance d'apprentissage du modèle MLP. Le nombre de neurones cachés détermine la capacité d'apprentissage du réseau MLP. Il est plus recommandé de sélectionner le réseau qui fonctionne le mieux avec le moins de neurones cachés possible.

Tableau III.6. Les différents paramètres MLP des deux ensemble (apprentissage et test).

Les paramètres de validation (RNA)	
N appr = 25	N test = 6
R = 0.973	R = 0.977
R^2 = 0.947	R^2 = 0.955
R^2 ajusté = 0.933	R^2 ajusté = 1.000
Q^2 = 0.999	Q^2 = 0.999
S (écart type) = 0.286	S = 0
F =237	F = 0

Les résultats obtenus ont montré que le réseau contenant cinq neurones est le meilleur modèle. Comme le montre le Tableau III.6. Le modèle MLP non linéaire a donné de bons résultats avec des coefficients de corrélation plus élevés R^2 (0.947. 0.955). En plus d'une meilleure robustesse Q^2 (0.999. 0.999) dans les ensembles d'apprentissage et de test. Par conséquent, un MLP avec l'architecture 5-2-1 a été choisi comme modèle final dans ce travail.

Tableau III.7. Les valeurs expérimentales, prédites et résiduelles.

N	PIC50-exp	PIC50-pred	PIC50 - Résiduelles
Ensemble d'apprentissage			
1	-1.06446	-1.08559	0.021132
2	-0.89763	-1.09440	0.196768
3	0.09151	0.29749	-0.205977
4	-0.10037	0.32168	-0.422053
5	-0.02938	0.12683	-0.156214
6	0.12494	-0.02922	0.154157
7	0.23657	0.29232	-0.055744
8	0.55284	0.05455	0.498289
9	0.11351	0.07120	0.042306
10	-0.86923	-0.78158	-0.087649
11	-1.39794	-1.28461	-0.113334

12	-1.53148	-1.62578	0.094305
14	-2.04532	-2.06950	0.024172
15	-0.04139	-0.00057	-0.040822
16	-1.32222	-1.21408	-0.108137
17	-1.62325	-1.53071	-0.092539
19	-1.74819	-1.70639	-0.041796
22	0.00000	0.17496	-0.174957
24	-0.14613	-0.05247	-0.093663
26	0.88606	0.91195	-0.025897
27	0.58503	0.19041	0.394613
28	0.65758	0.65642	0.001160
29	0.46852	0.20871	0.259816
30	-0.32222	-0.13325	-0.188968
31	-0.90309	-0.94557	0.042483
Ensemble de test			
13	-2.01703	-1.74018	-0.276854
18	-1.53148	-1.64605	0.114568
20	-2.01284	-1.96021	-0.052630
21	0.35655	0.17496	0.181590
23	0.00000	0.29885	-0.298850
25	0.08619	-0.17858	0.264769

III.3 Conclusion

La régression linéaire multiple donne des résultats assez acceptables pour prédire l'activité inhibitrice mais elle n'est pas supérieure au RNA.

Des études cumulatives ont montré que les réseaux de neurones artificiels sont meilleurs pour la prédiction que la régression linéaire multiple. Dans ce chapitre. Nous avons pu confirmer cette théorie. Le RNA est l'outil le plus puissant pour prédire l'activité inhibitrice en comparant les résultats des deux modèles. Les valeurs R et R^2 les plus élevées ont été trouvées pour le modèle de RNA. Cette force est due à la capacité du réseau à reconnaître les modèles entre l'entrée et la sortie ou à la généralisation qui a été mise en œuvre.

Conclusion Générale

Conclusion Générale

Les méthodes QSAR représentent un domaine de recherche très important de la chimie-physique. L'objectif est d'établir une relation ou une équation mathématique reliant des descripteurs structurels d'une série de composes étudies aux activités biologiques. Ces méthodes ont permis de nouvelles avancées vers la compréhension des données biologiques en utilisant des descripteurs simples et cibles.

Comme nous l'avons mentionné précédemment, l'objectif de notre travail est de modéliser les activités inhibitrices de protéine kinase plasmodium falciparum pfpk7 pour former des modèles de QSAR robustes, stables, et précis capables de prédire efficacement ces activités.

QSAR se trouve à l'intersection de la chimie, les statistiques et la biologie, pour cette raison les deux premiers chapitres de cette thèse ont été consacrés à la cause du paludisme, à son mode de transmission et à l'effet de la protéine kinase pf sur la maladie. La méthodologie QSAR ainsi tous les outils nécessaires pour la mise en place d'un modèle QSAR.

Une étude de la relation quantitatives structure activité QSAR est appliquée à un ensemble de 32 molécules dérivés de l'imidazopyridazine, afin de prédire l'activité biologique inhibitrice de protéine kinase plasmodium falciparum (pfpk7) et de trouver une corrélation entre les différents paramètres moléculaires (descripteurs) de ces composés et son activité biologique, en utilisant l'analyse en composantes principales ACP pour la sélection de variables, la régression linéaire multiple RLM, les réseaux des neurones RNA.

Les résultats statiques de RLM, du RNA indiquent que les coefficients de détermination R^2 étaient respectivement de 0.92, 0.97 et coefficients de détermination R^2ajusté 0.91, 0.90 respectivement. Les résultats montrent que les méthodes de modélisation permettent de bien prédire l'activité et peuvent être utiles pour prédire l'activité biologique de nouveaux composés. La validation des modèles a été utilisée pour déterminer la qualité statistique et le pouvoir prédictif du QSAR des deux modèles RLM et RNA.

Enfin, nous avons soutiré des informations utiles à partir des équations des modèles. Ainsi, nous avons pu interpréter la contribution du descripteur sur la variation de l'activité biologique.

A travers les différents résultats obtenus au cours de ce travail, nous pouvons dire que les traitements statistiques et les techniques informatiques utilisés, lors du développement et l'analyse des modèles de QSAR, ont donné de bons résultats, ce qui nous permet d'entrevoir des perspectives assez prometteuses dans ce domaine par l'amélioration des traitements statiques et par l'utilisation d'autres méthodes de sélection.

Référence

Bibliographie :

[1] Bibliosante.ml – A nianglay – 2014-page1436
https://www.bibliosante.ml/handle/123456789.

[2] OMS, (2003). Entomologie du paludisme et contrôle es vecteurs 1, Rapport du département du contrôle, de la prévention et de l'éradication. Groupe des maladies transmissibles. Edition provisoire, WHO/CPE/SMT/2002,18 Rev.1.

[3] World malaria report 2014, Geneva : Wold Health Oganisation; 2014 (résumé disponible enfrançais)
(https:/www.who.int/malaria/publications/wrold_malaria_report_2014/en/,consulté le 10/05/2021).

[4] OMS. Accès universel aux tests diagnostiques de paludisme : manuel pratique, novembre 2011 (rév. Février 2013). Genève. Oganisaion mondiale de la Santé, 2012 (https://apps.who.int/iris/bitstream/1066/78877/1/9789242502091_fre.pdf ?ua=1, consulté le 10/05/2021).

[5] DANIS M. LEGROS F, GAY F et Coll – Paludisme d'importation en France. Med Mal Infect 1999 ; 29 suppl 3 : 357-373.

[6] COLLECTIF-Conférence de Consensus – Prise en charge et prévention du Paludisme d'importation à *Plasmodium falciparum.*

[7] Majoie CHARRIER, Florence BAZZARO, Jean-Claude SAGOT, ACTES## DU# CONGRES, 120, 2014.

[8] Paris siences letters, 2019.

[9] Bitam said, Thèse de doctorat.

[10] Louis Trudel, Claudine Simard, Nicolas Vonarx, Recherches qualitatives 5, 38-55, 2006

[11] Bitam said la thèse de doctorat.

[12] Hélène Guillemain, Conservatoirenational des arts et metiers-CNAM, 2012.

[13] Benoit Playe, Paris Sciences et Lettres, 2019.

[14] Charlotte Josepha Joos, Paris 6, 2007.

[15] Yves Clot, Daniel Faïta, Gabriel Fernandez, Livia Scheller, Prespectives interdisciplinaires sur le travail et la santé, 2000.

[16] M Houde, L Desbiens, P D »Orléas-Juste-Elsevier-Advances in Pharmacology, 2016.

[17] G. Wermuth, "Designing prodrugs and bioprecursors", The practice of Medicinal Chemistry (Third Edition), 2008

[18] V. N. Viswanadhan, M. R. Reddy, R. J. Bacquet et M. D. Erion, Journal of computational chemistry. 14, 1019 (1993).

[19] G. Fayet, Thèse de doctorat, Développement de modèle QSAR pour la prédiction des propriétés d'explosibilité des composés nitro-aromatiques. Université pierre et marie Curie, paris VI, (2010).

[20] Perkins R, Fang H, Tong W, Welsh WJ. Quantitative structure-activity relationship methods: Perspectives on drug discovery and toxicology. Environmental Toxicology and Chemistry. 2003, 22 (8): 1666-79.

[21] Free SM, Wilson JW. A mathematical Contribution to Structure-activity Studies. Journal of Medicinal Chemistry. 1964, 7 (4): 395-9.

[22] Antoine DAINA, Olivier MICHIELIN et Vincent ZOETE, «iLOGP: a simple, robust, and efficient description of n-octanol /water partition coefficient for drug design using the GB/SA approach», in: Journal of chemical information and modelling 54.12 (2014), P. 3284-3301.

[23] T. Puzyn, J. Leszcznski, M. T. D. Cronin, «Recent advances in QSAR Studies. Methods and application» Springer Dor drecht Heidelberg London New york, 2010.

[24] M. Tichy, Science. Tot. Env. 109, 1991, 407-410.

[25] O. M. Bautista-Aguilera, G. Esteban, I. Bolea, K. Nikolic, D. Agbaba, I. Moraleda, I. Iriepa, A. Samadi, E. Soriano, M. Unzeta, J. Marco-Contelles, Europ. J. Med. Chem. 75, 2014, 82-95.

[26] N. Bouloc, M. Jonathan ،E. S. Large, W. David, H. Keith, C. D. Ansell, and S. B. Justin, Synthesis and in vitro evaluation of imidazopyridazine as novel inhibitors of the malarial kinse PfPK7, Bioorg, Med. Chem. Lett. 18 (2008), PP. 5294-5298.

[27] A. Merckx, A. Echalier, K. Langford, A. Sicard, G. Langsley, J. Joore, C. Doerig, M. Noble, and J. Endicott, Structures of P. falciparum protein kinase 7 identify an activation motif and leads for inhibitor design, Structure 16 (2008), PP. 228-238.

[28] D. Dorin, J. Semblat, P. Poullet, P. Alano, J.P.D. Goldring, C. Whittle, S. Patterson, D. Chakrabarti, and C. Doerig, PfPK7. an atypical MEK-related protein kinase, reflects the absence of classical three-component MAPK pathways in the human malaria parasite plasmodium falciparum, Mol. Microbiol. 55 (2005), PP. 184-196.

[29] V. K. Agrawal, R. Srivastava, and P. V. Khadikar, QSAR studies on some antimalarial sulfonamides, Bioorg. Med. Chem. 9 (2001), PP. 287-293.

[30] A. Xie, S. R. Clark, S. Prasanna, and R. J. Doerksen, Imidazole-containing farnesyltransferase inhibitors: 3D quantitative structure-activity relationships and molecular docking, J. Comput. Aided Mol. Des. 23 (2009), PP. 431_448.

[31] C. Hansch, A quantitative approch to biochemical structure- activity relationships, Acc. Chem. Res. 2 (1969), P. 232.

[32] T. Scior, J. L. Medina-Franco, Q. T. Do, K. Martínez-Mayorga, J. A. Yunes Rojas, and P. Bernard, How to recognize and workaround pitfalls in QSAR studies: A critical review, Curr. Med. Chem. 16 (2009), PP. 4297_4313.

[33] M. Karelson, Molecular descriptors in QSAR/QSPR, Wiley, New York, 2000.

[34] bellifa Khadidja, doctorat en chimie physique: étude des relations quantitatives structue-toxicité des composés chimiques à l'aide des descripteurs moléculaires. « Modélisation QSAR », universite abou bekr belkaïd de tlemcen, 2015.

[35] H. Wiener, Structural determination of paraffin boiling points. J. Am. Chem. Soc. 1947, 69, 17_20.

[36] M. Randic, Characterization of molecular branching. J. Am. Chem. Soc. 97, 6609_6615.

[37] L. B. Kier, L. H. Hall, W. J. Murray, Molecular connectivity I: Relationship to nonspecific local anesthesia. J. Pharm. Sci. 1975, 64, 1971_1974.

[38] A. T. Balaban, Highly Discriminating Distance-Based Topological Index, Chem. Phys. Lett. 1982, 89, 399_404.

[39] R. D. Cramer, D. E. Patterson, and J. D. Bunce, "Comparative molecular field analysis (CoMFA): Effect of shape on binding of steroids to carrier proteins", Journal of the American Chemical Society, 110(18), 1988, 5959_5967.

[40] F. Bonachera, " Les triplets pharmacophoriques flous: développement et applications", Thèse de Doctorat, Université Lille 1 Sciences et Téchnologies, France, 2011.

[41] N. Goudarzi, M. Goodarzi, and T. Chen. " QSAR prediction of HIV inhibition activity of styrylquinoline derivatives by genetic algorithm Compled with multiple linear regressions". Medicinal Chemistry Research Vol. 21, no. 4 PP. 437_443, 2012.

[42] N. L. Allinger, "Calculation of molecular Structure and energy by force- field methods". In Advances in physical organic Chemistry, Vol. 13, PP. 1_82. Academic Press, 1976.

[43] Hoffman BT, Kopajtic T, Katz JL, Newman AH. 2D QSAR modeling and preliminary database searching for dopamine transporter inhibitors using genetic algorithm variable selection of Molconn Z descriptors. Journal of medicinal chemistry. 2000; 43(22):4151-9.

[44] K. Y. Wong, A. G. Mercader, L.M. Saavedra, B. Honarparvar, G. P. Romanelli, and P.R. Duchowicz, QSAR analysis on tacrine-related acetylcholinesterase inhibitors, J. Biomed. Sci. 21 (2014), p. 84.

[45] C. W. Yap, H. Li. Z.L. Ji, and Y.Z. Chen, Regression methods for developing QSAR and QSPR models to predict compounds of specific pharmacodynamic, pharmacokinetic and toxicological properties, Mini Rev. Med. Chem. 7 (2007), 1097_1107.

[46] Marc Parzeau, "RÉSEAUX DE NEURONES, GIF-21140 et GIF-64326", UNIVERSITE DE LAVAL, automne 2004.

[47] N. L. Allen and P. N. Mikropoulos, "STREAMER PROPAGATION ALONG INSULATING SURFACES", IEEE Transaction on Dielectrics Electrical InsuLating Vol. 6 No. 3, June 1999.

[48] R. Mazri, Nouvelle approche des relations Structures-activités dans des molécules antibiotiques, Thèse de doctorat, université Biskra, 2015.

[49] G. Philippe, utilisation du modèle linéaire rappel de base méthodes de validation novembre, 2000.

[50] L. Zekri, étude QSAR par doking moléculaire des propriétés cytotoxiques de certaines dérives de nitrobenzoxadiazole, mémoire master, université Biskra, 2015.

[51] MA. Leach, Quantum Mechanical Models, in Molecular modelling: Principles and applications. Addison Wesley Longman Ltd: Harlow. pp. 25-130, 1996.

[52] H. Dugas, Princeps de basent modélisation moléculaire, Aspects théorique et pratiques, Chapitre 3 introduction aux méthodes de minimisation d'énergie, quatrième éditions, Libraire d L'Université de Montréal, 1996.

[53] N.Melkemi, Etude des corrélations entre les paramètres structuraux et les propriétés bioactives de quelques composés organiques, thèse doctorant, Université Biskra, 2015.

[54] Y. Yousfi, étude QSAR de l'activité anti-oxydante d'une série de composes phénoliques, mémoire master université Tlemcen, 2017.

[55] J. Tirado-Rives, and William L. Jorgensen. "Molecular dynamics of proteins with the OPLS potential functions. Simulation of the third domain of silver pheasant ovomucoid in water." Journal of the American Chemical Society, Vol. 112, no. 7, pp. 2773-2781, 1990.

[56] O. Afoufou, étude quantitative des relations structure-activité dans les 4-aniloquinozalines, agents anticancéreux, mémoire master, université Biskra, 2013.

[57] C. Hansch and T. Fujita," P-∂-∏ Analysis. A Method for the Correlation of Biological Activity and Chemical Structure", Journal of the American Chemical Society, 86(8), 1964, 1616-1626.

[58] E. Overton, Studien über die Narkose zugleich ein Beitrag zur allgemeinen Pharmakologiel. Ed. G. Fischer, Jena, 1901.

[59] R. L. Lipnick, "Hans Horst Meyer and the lipoid theory of narcosis", Trends in Pharmacological Sciences, 10(7), 1989, 265_269.

[60] H. Fühner and E. Neubauer, " ămolyse durch Substanzen homologen Reihen", Archiv für experimentelle Pathologie und Pharmakologie, 56, 1907, 333_345.

[61] O. R. Hansen, " Hammett Series with Biological Activity", Acta Chemica Scandinavica, 16, 1962, 1593_1600.

[62] S. M. Free and J. W. Wilson, "A Mathematical Contribution to Structure-Activity Studies", Journal of Medicinal Chemistry, 7(4), 1964, 395_399.

[63] C. Hansch and E. J. Lien, Structure-activity relationships in antifungial agents. A survey, Journal of Medicinal Chemistry, 14(8), 1971, 653_670.

[64] S. Y. Tham and S. Agatonovic-Kustrin, Application of the artificial neural network in quantitative structure-gradient elution retention relationship of phenylthiocarbamyl amino acids derivatives, Journal of Pharmaceutical and Biomedical Analysis, 28(3), 2002, 581_590.

[65] M. Karelson, Molecular descriptors in QSAR/QSPR, Wiley, New York, 2000.

[66] R. Todeschini, V. Consonni, and R. Mannhold_Molecular Descriptors for Chemoinformatics, Drug Discovery & Development, 41(2), 2009.

[67] bellifa Khadidja, doctorat en chimie physique: étude des relations quantitatives structure_ toxicité des composés chimiques à l'aide des descripteurs moléculaires. «Modélisation QSAR», université abou bekr belkaid de Tlemcen, 2015.

[68] H. Wiener, Structural determination of paraffin boiling points. J. Am. Chem. Soc. 1947, 69, 17_20.

[69] T. M. Martin, P. Harten, D. M. Young, E.N. Muratov, A. Golbraikh, H. Zhu and A. Tropsha, Dose Rational Selection of Training and Test Sets Improve the Outcome of QSAR Modeling, J. Chem. Inf. Model. 2012, 52, 2570-2578.

[70] J. JACQUES, ‘‘ Modélisation Statistique ’’ http://labomath.univ-lille 1.fr/ ~jacques.

[71] D. Laffly, ‘‘Régression multiple : principes et exemples d’application’’, 2006.

[72] H. Nocairi, M. Hanafi, E. M. Qannari. Rev. Statistique Appliquée, LIII (2), 2005, 29-41.

[73] P. Desmet. Recherche et application en Marketing, 11 (2), 1996, 17-28.

[74] E. Tenorio-Borroto, C. G. Rivas, J. C. Vásquez chagoyán, N. Castañedo, F. J. Prado-Prado, X. Garcia-Mera, H. Gonwáler-Diaz. Bioorg. Med. Chem. 20, 2012, 6181-6194.

[75] https://www.combridgesoft.com. cosulté le 19/05/2021.

[76] https://www.findwords.info .cosulté le 19/05/2021.

[77] J.Chem. Inf Comput. Sci, 2004, 44, 2225-2226.

[78] Laggner, C. SMARTS Patterns for Functional Group Classification; 2009. Available at http //code.google.com /P/semanticchemistry/source/browse/wiki/InteLigand.wiki,spec=snv4 1&r=41. (last accessed: 15 May 2010).

[79] https://www.fr.freedownloodmanager.org, consulté le 20/05/2021.

[80] https://www.fr.m.wikipedia.org, consulté le 21/05/2021.

[81] https://www.statistica.com, consulté le 20/05/2021.

[82] Damale, M. G. HarKe, S. N. Kalam khan, F. A, et al. Recent advances in multi dimensional QSAR (4D-6D): a critical review. Mini Rev. Med. Chem , 2014, 14(1) : 35-55.

[83] Todeschini, R. and Consonni, V. (2009). Molecular descriptors for chemoinformatics, (Weinheim: Wiley VCH) pg 27-37.

[84] Hall, L. H, and kier, L. B. (1995). Electrotopological state indices for atom types: A novel combination of electronic, topological, and valence state information. J Chem Inf Comput Sci 35, 1039-1045.

[85] Liu, S. and Cao, C. And Li, Z. Approach to Estimation and Prediction for Normal Boiling Point (NBP) of Alkanes Based on Novel Molecular Distance Edge (MDE) Vector, lambda, Journal of Chemical In formation and Computer Sciences, 1998, 38:387-394.

[86] O. A. S. Carpinteiro et A. PA. Dasilva, A Hierarchical Neural Model in Short-Term Load Forecasting, the VI Brazilian Symposium on Neural networks, pp 120-124, 2000.

[87] H. Mori, State of the art overview on artificial neural networks in power systems El-Sharkawi et Niebur edition, pp 51-70, 1996.

[88] D. Srtnivasan, A. C. Liew et J.S.P. Chen, A Novel Approach To Electrical Load Forecasting Based On A Neural Network, IEEE Neural Networks Vol 2, pp 1172-1177, 1991.

[89] P. K. Dash, A.C. Liew, S. Rahman et S. Dash, Fuzzy and Neuro-fuzzy Computing Models for Electric Load Forecasting, Engineering Applications of Artificial Intelligence Vol. 8, N° 4, pp 423-433, Août 1995.

[90] T. Senjyu, H. Takara et K. Uezato, Two-hour-ahead Load Forecasting Using A Neural Network, Bulletin of the Faculty of Engineering, University of the RyuKyus Vol 60, pp 61-68, 2000.

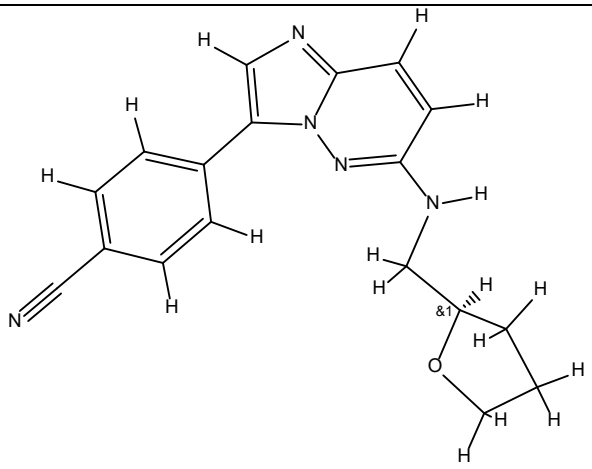
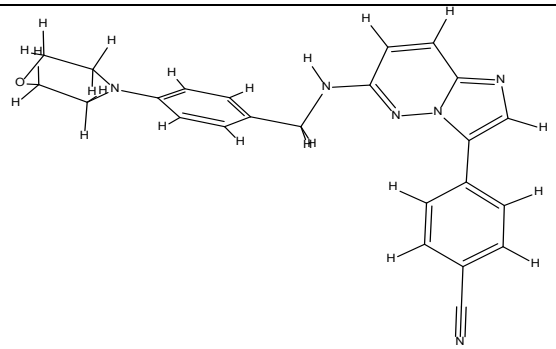
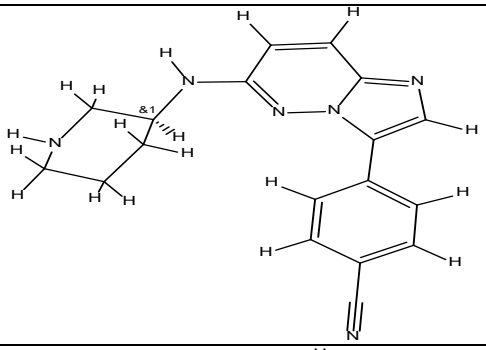
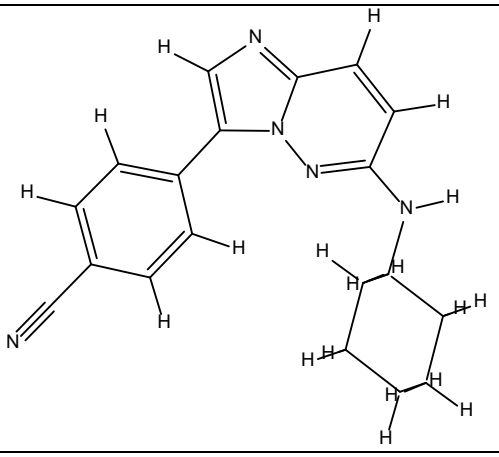
[91] D.K. Ranaweera, N.F. Hubele et G.G. Karady, Fuzzy Logic Term Load Forecasting, Electrical Power & Energy Systems Vol. 18, No. 4, pp 215-222, 1996.

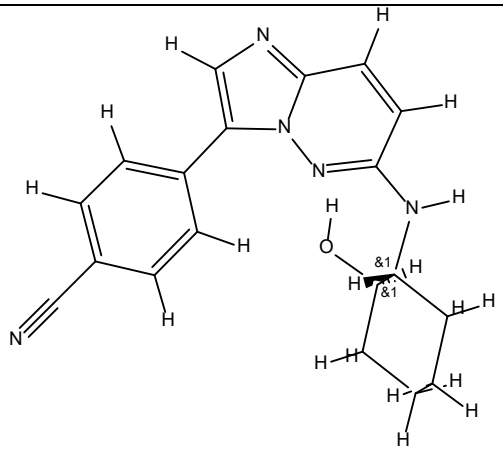
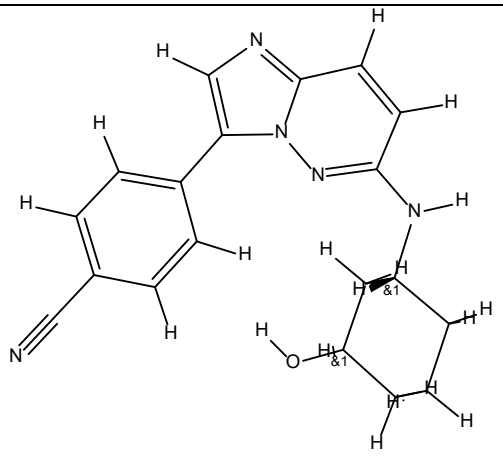
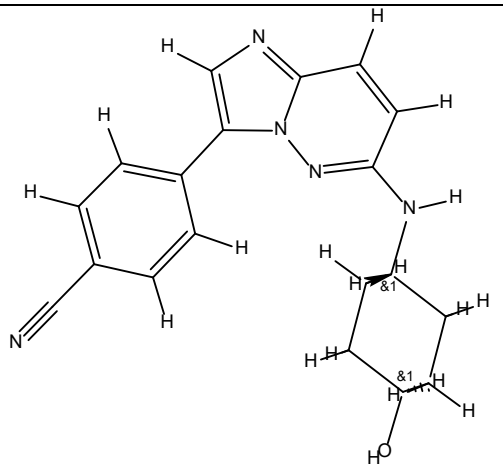
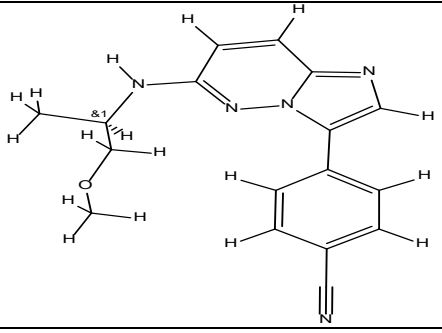
[92] Y. Zhu et Y. He, Short-Term Load Forecasting Model Using Fuzzy C Means Based Radial Basis Function Network, IEEE 6^{th} International Conference on intelligent Systems Design and Applications Vol 1, pp 579-582, 2006.

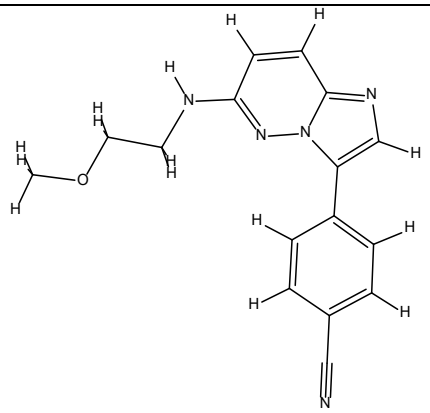
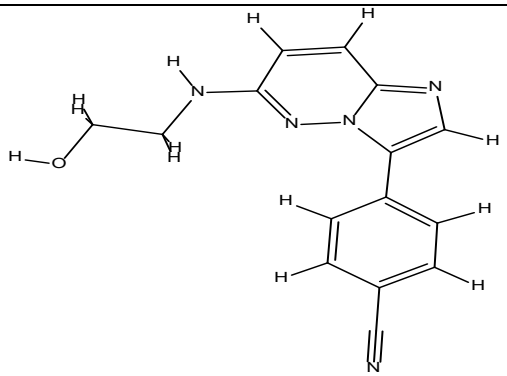
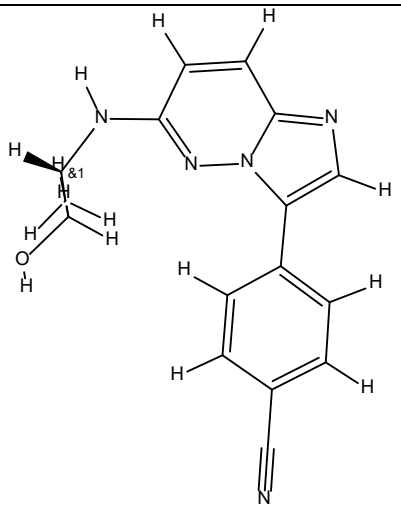
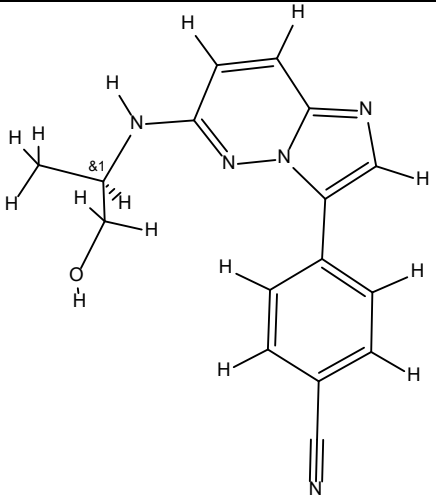
[93] W. Charytoniuk et M.S. Chen, Neural Network design for Short-Term Load Forecasting, Proceeding of International Conference on Electric Utility Deregulation and restructuring and Power Technologies, London, pp 554-561, 2000.

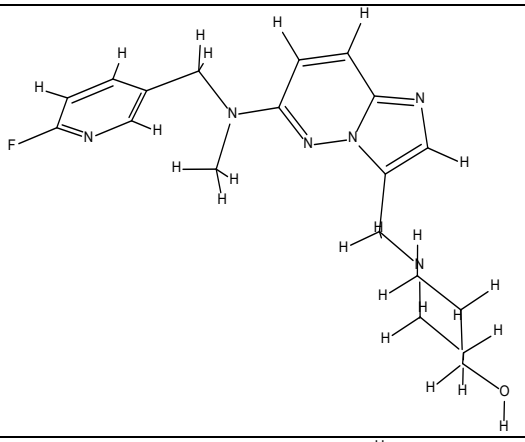
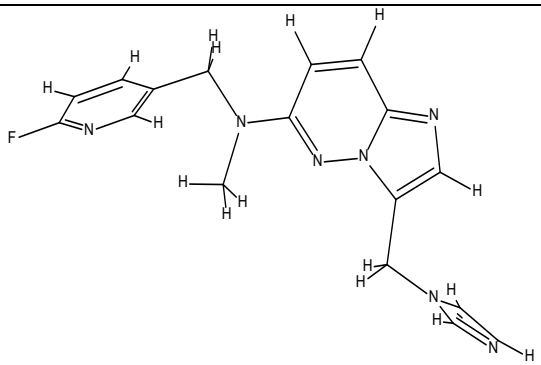
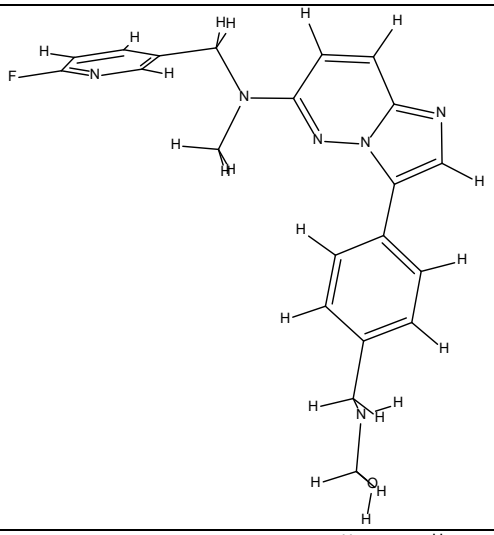
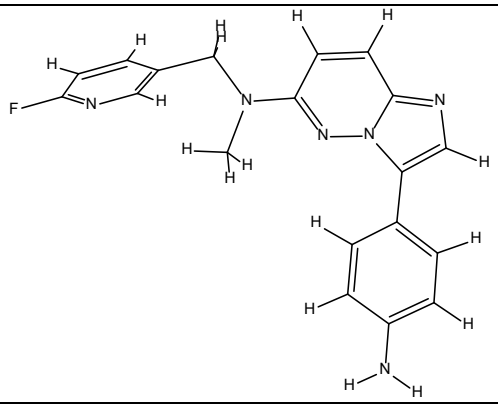
ANNEXE

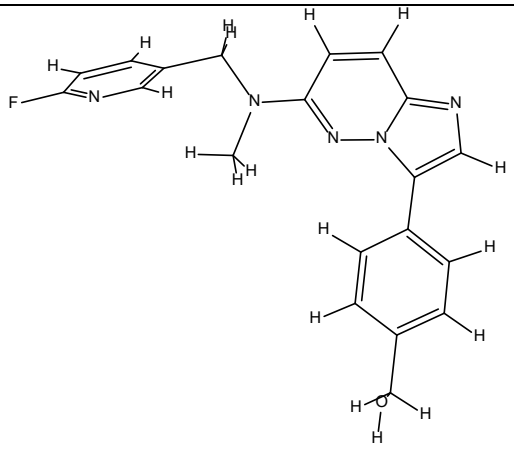
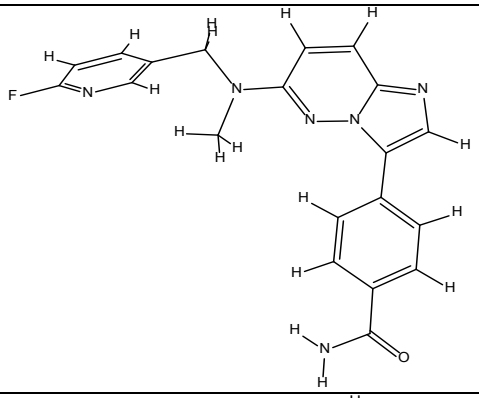
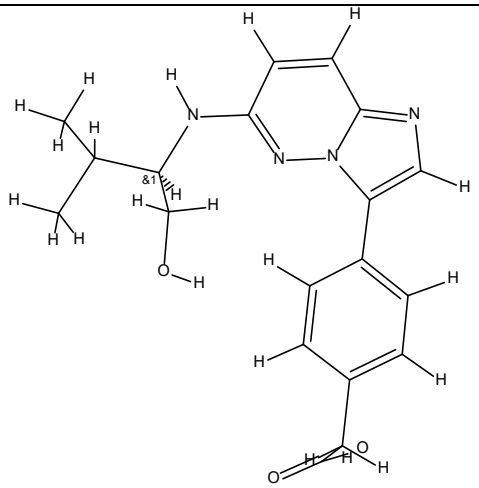
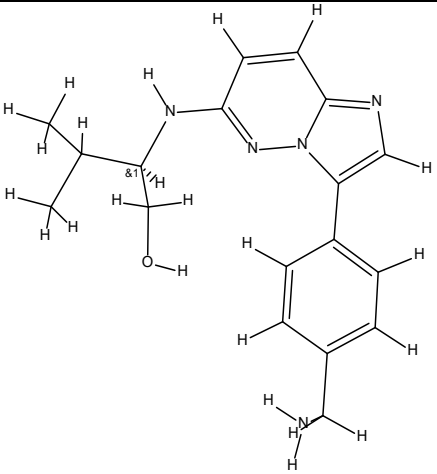
ANNEXE

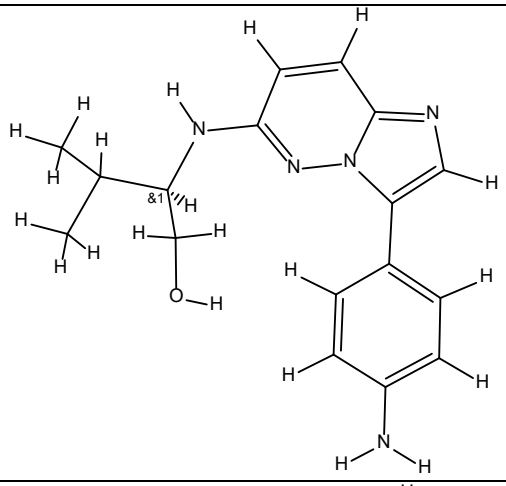
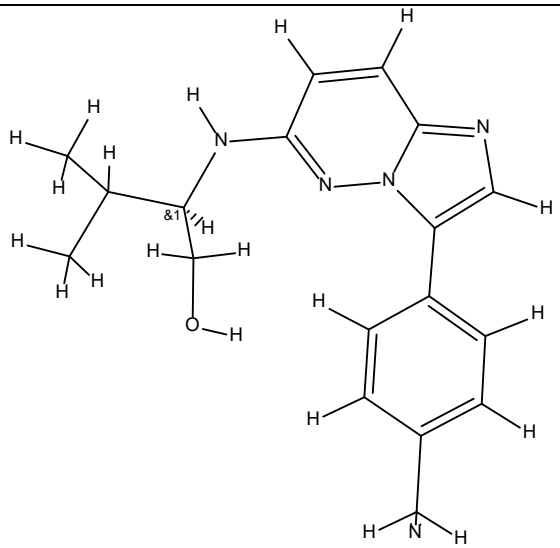
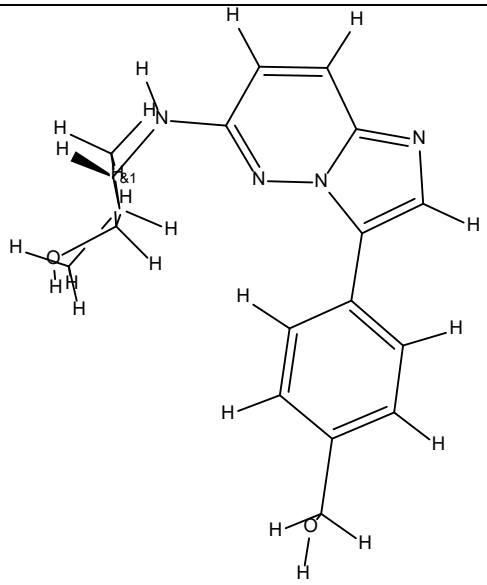
N°	Structure		PIC_{50}
1	$C_{27}H_{28}N_6O$		-1.06
2	$C_{24}H_{22}N_6O$		-0.9
3	$C_{18}H_{18}N_6O$		0.99
4	$C_{19}H_{19}N_5$		-0.1

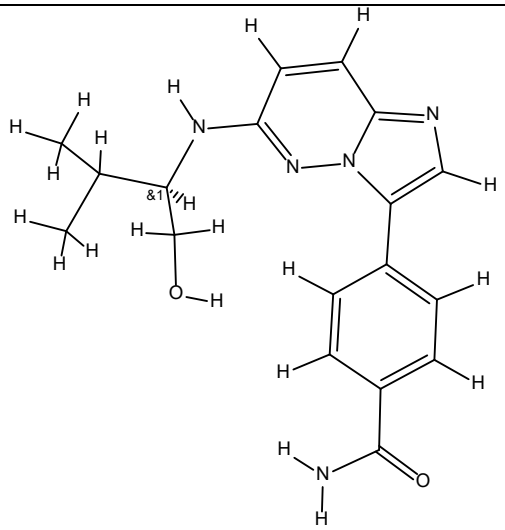
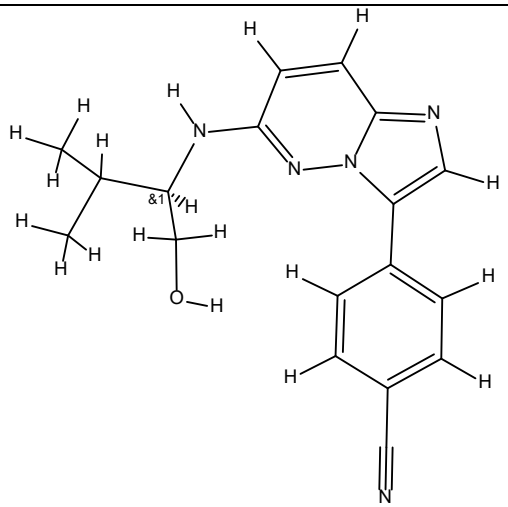
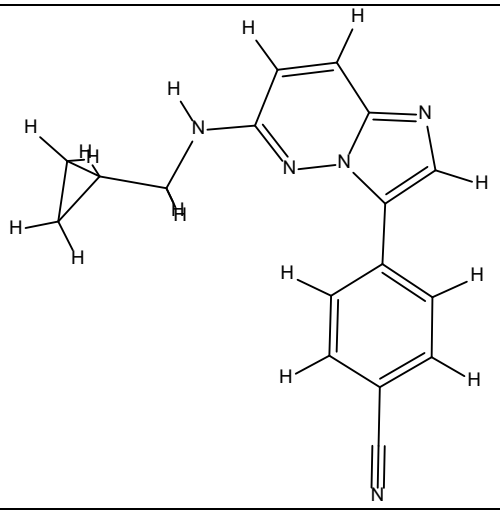
5	**$C_{19}H_{19}N_5O$**		-0.03
6	**$C_{19}H_{19}N_5O$**		0.12
7	**$C_{17}H_{17}N_5O$**		0.24
8	**$C_{16}H_{15}N_5O$**		0.55

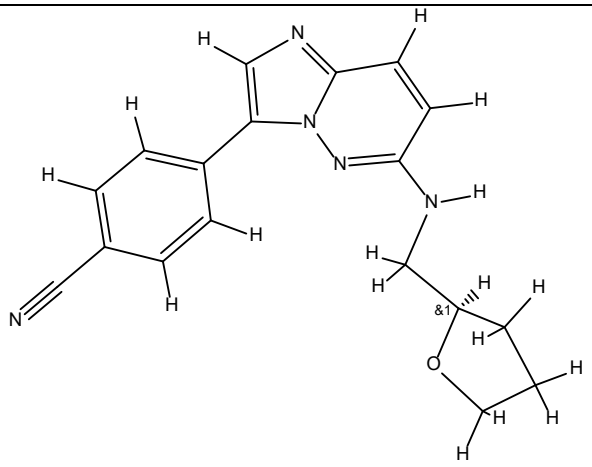
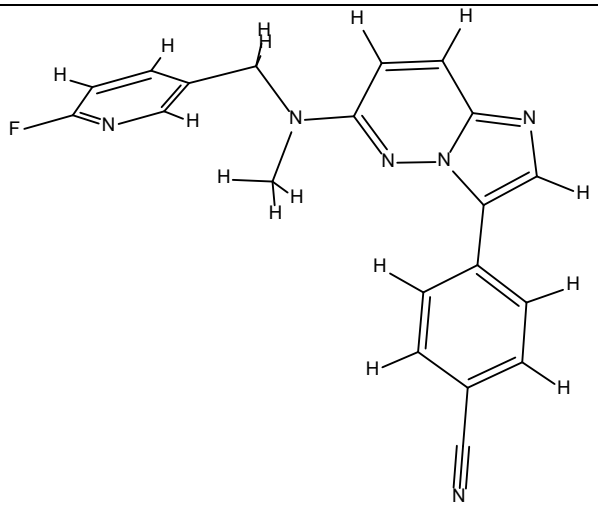
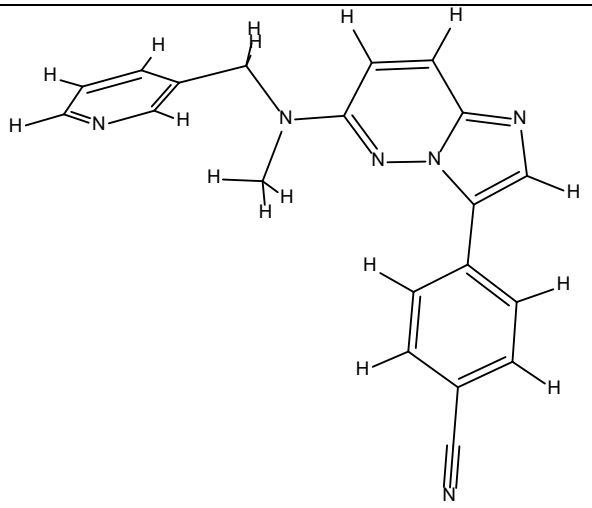
9	**$C_{15}H_{13}N_5O$**		0.11
10	**$C_{16}H_{15}N_5O$**		-0.87
11	**$C_{16}H_{15}N_5O$**		-1.4
12	**$C_{19}H_{23}FN_6O$**		-0.53

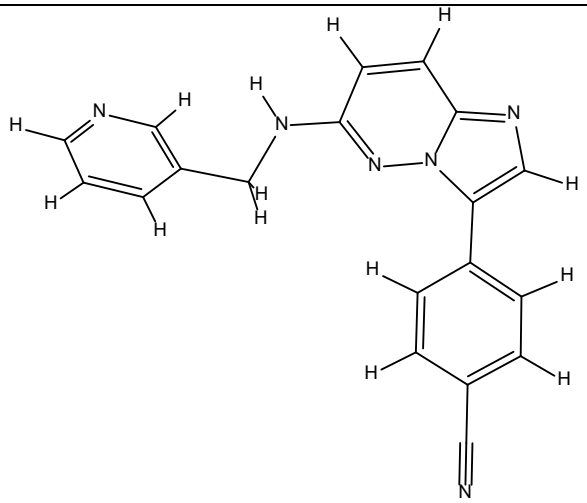
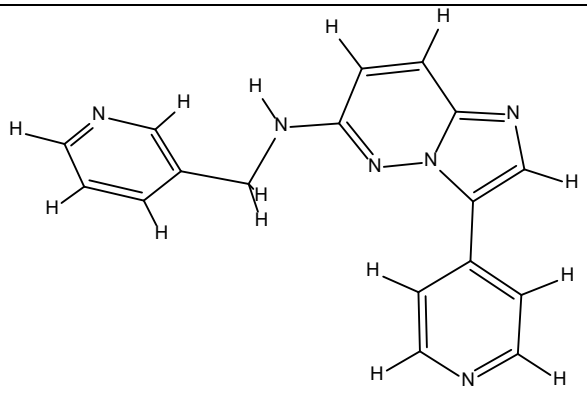
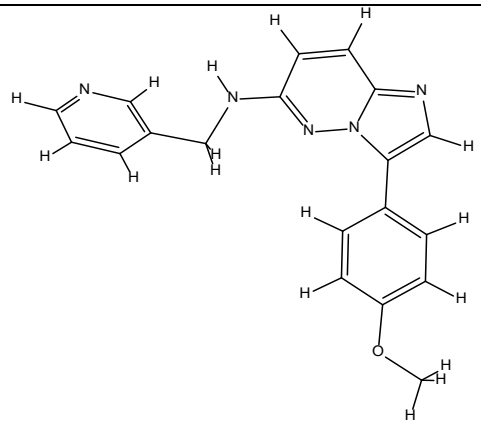
13	**$C_{17}H_{16}FN_7O$**		-2.02
14	**$C_{21}H_{21}FN_6O$**		-2.05
15	**$C_{19}H_{17}FN_6O$**		-0.04
16	**$C_{20}H_{18}FN_5O$**		-1.32

17	**$C_{20}H_{17}FN_6O$**		-1.62
18	**$C_{19}H_{22}N_4O$**		-1.53
19	**$C_{18}H_{23}N_5O$**		-1.75
20	**$C_{17}H_{21}N_5O$**		-2.01

21	**$C_{19}H_{21}N_5O$**		0.36
22	**$C_{18}H_{22}N_4O$**		0
23	**$C_{18}H_{21}N_5O$**		0

24	**$C_{18}H_{19}N_5O$**		-0.15
25	**$C_{17}H_{15}N_5$**		0.09
26	**$C_{18}H_{17}N_5O$**		0.89

27	**$C_{20}H_{15}FN_6$**		0.59
28	**$C_{20}H_{16}N_6$**		0.66
29	**$C_{19}H_{14}N_6$**		0.47

30	**$C_{19}H_{14}N_6$**		-0.32
31	**$C_{17}H_{14}N_6$**		-0.9
32	**$C_{19}H_{17}N_5O$**		No inhibition