REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEURE ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DE MEDEA

FACULTE DES SCIENCES



Mémoire

Pour obtenir le diplôme de **Master**

En chimie

Spécialité :
Chimie pharmaceutique

Présenté par :
BOUKHALFA Slimane
BOUALAMI Mohamed

Thème :

ETUDE QSAR D'INHIBITEUR DE NEURAMINIDASE DE LA VERUS DE LA GRIPPE

Évalué le : … /07 /2021 par le jury composé de :

MESELLEM Yamin	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

Dédicace

A mon trésor éternel et raison de ma vie, symbole de sacrifice et d'amour, qui m'a toujours soutenue et prêtée aide et assistance dans mes études : « Ma mère »

A l'homme le plus fort, le plus important dans ma vie, pour ses efforts et ses sacrifices : « Mon père »

A mes étoiles qui éclairent ma vie :

Mes frères : Ali, Omar, Fatah, Mes sœurs

Mes nièces et mes neveux : Mohamed, Rayen, Okba Mahdi Walid Chaima Achouak

Mon grand-pére et ma grand-mére

Mon oncle : khaled, rachid, hanachi et ma chére tante fadila et fatiha et ma niéce NADJWA.

Je tiens, enfin à remercier tous ceux qui se sont spontanément impliqués de près ou de loin dans la réalisation de ce travail.

Remerciements

Je remercie Dieu le tout puissant de m'avoir accordé

Le courage pour terminer ce travail

Je tiens à exprimer mes vifs remerciements à

Directeur de recherche Monsieur" BITAM. Said "pour son aide, ses

Encouragements et ses conseils précieux.

Mes sincères remerciements à tous les membres de jury pour avoir

Accepté de lire et d'évaluer ce travail

Nous remercions toute la famille enseignante de l'université de

Médéa

الملخص

الزكام هو مرض ينتج من العدوى التنفسية التي يسببها فيروس الأنفلونزا, وتشير التجارب إلى أن تثبيط بروتين النورامينيداز الغشائي يقلل من حدة المرض ,تم ذكر اوزيلتاميفير و الزانامیفیر ,على أساس أنهما المسؤولان على تثبيط النورامينيداز لذلك سيكون من المرغوب فيه تطوير طرق حاسوبية لتسهيل فحص هده النظائر وفي هذه الدراسة سيتم تطوير نماذج العلاقات الكمية الهيكلية والنشاط للتنبؤ بالقدرة العلاجية لهده النظائر

الكلمات المفتاحية : الزكام, الشبكات العصبية, الانحدار الخطي المتعدد, اوزيلتامفير واوزيلتامفير

Résumé

La grippe est une maladie respiratoire infectieuse causée par le virus influenza, les essais préliminaires indiquent l'inhibition de la protéine membranaire neuraminidase réduire la gravité de la maladie a été mentionnée oseltamivir et zanamivir, ce sont deux puisent inhibiteur de neuraminidase. Il serait donc souhaitable de développer des méthodes informatiques pour faciliter le criblage de ses analogues. Dans cette étude, il sera développé de relation quantitative structure-activité (QSAR) vont être développés pour prédire le potentiel thérapeutique des analogues de oseltamivir et zanamivir.

Mots-clés : la grippe, RNA, RLM, QSAR, inhibiteur de neuraminidase, oseltamivir, Zanamivir.

Abstract

Influenza is a respiratory illness infection caused by the influenza virus, preliminary tests indicate inhibition of membrane protein reduce disease severity, was mentioned oseltamivir and zanamivir these are two powerful inhibitors of neuraminidase, it would therefore be desirable to develop computer methods to facilitate the screening of its analogs. In this study it will be developed of relation quantitative structure-activity relationships (QSAR) will be

Developed to predict the therapeutic potential of analogues of oseltamivire and zanamivir.

Keywords : influenza, ANN, MLR, QSAR, inhibitor of neuraminidase, oseltamivir, Zanamivir.

Table de matières

Introduction Générale 1
Chapitre I : Rappels bibliographiques
I.1 La grippe 4
I.1.1 Définition 4
I.1.2 La structure général du virion 4
I.1.3 Le génome viral 5
I.1.4 Les protéine viral 5
I.1.5 La neuraminidase (NA) 6
I.1.6 Les inhibiteurs de neuraminidase 6
I.2 Les méthode de conception de médicaments 7
I.2.1 Première phase : Recherche exploration 7
I.2.2 Deuxième phase : Les essai préclinique 7
I.2.3 Les essais cliniques 7
Chapitre II : Développement de méthode QSAR
II.1Introduction 10
II.2 Le principe de méthode QSAR/QSPR 10
II.3 Les types de descripteurs 10
II.3.1 Descripteurs moléculaires théoriques 10
II.4 Les différents modèles QSAR 11
II.4.1 La régression linéaire multiple RLM 11
II.4.2 Les réseaux de neurones artificiels 11
II.5 La méthodologie général de modèle QSAR/QSPR 11
II.6 Les paramètres de validation de méthode QSAR 12
II.6.1 Le coefficient de détermination R2 : 12
II.6.2 Le coefficient de détermination ajusté R2adju 12
II.6.3 Le coefficient de prédiction Q2 13
II.6.4 Test de fischer F 13
II.6.5 Ecart type S 13
II.6.6 Le facteur d'inflation de la variance VIF 14
Chapitre III : Résultats et Discussion
III.1 Première partie 16
III.1.1 Introduction 16
III.1.2 La base de données 16
III.1.3 Les calcules 18
III.1.3.1 Génération des descripteurs 18
III.1.4 Les logiciels utilisée 18
III.1.5 Les modèles choisis 19
III.1.5.1 Le modèle MLR 19
III.1.5.2 Le modèle des réseaux de neurones artificiels RNA 19
III.2 Deuxième partie 20
III.2.1 Résultats et discussion 20
III.2.1.1 Sélection des descripteurs 20
III.2.1.2 Réfractivité (AMR) 20

III.2.1.3 Les descripteurs de structure topologique (ATS) pondérés 21
III.2.2 Le modèle RLM 21
III.2.3 Le modèle des réseaux de neurones artificiels RNA 25
Cocnlusion générale 31
Références 32
Annexe 01 34
Annexe 02 46

Liste des figures

Figure I.1. Représentation schématique du virus influenza de type A et du complexe ribonucléoprotéique

Figure I.2. Structure de la neuraminidase

Figure III.1. Structure moléculaire de zanamivir et Oseltamivir

Figure III.2. Structure moléculaire de composé 12

Figure III.3. Structure moléculaire général de Benzoïque acides

Figure III.4. Représentation graphique des concentrations inhibitrices prédites en fonction des concentrations inhibitrices expérimentales pour l'ensemble des analogue Train, Test

Figure III.5. Représentation graphique des concentrations inhibitrices prédites en fonction des concentrations inhibitrices expérimentales pour l'ensemble des analogue Train

Figure III.6. Représentation graphique des concentrations inhibitrices prédites en fonction des concentrations inhibitrices expérimentales pour l'ensemble des analogue Test

Liste des tableaux

Tableau III.1. Les analogues de zanamivir et Oseltamivir utilisés et valeurs des concentrations inhibitrices expérimentales.

Tableau III.2. Les descripteurs utilisés pour le développement des modèles QSAR :

Tableau III.3. Les valeurs de pIC_{50} expérimentales avec les descripteurs sélectionnés.

Tableau III.4. Les valeurs de pIC_{50} prédite obtenue.

Tableau III.5. Valeurs du coefficient VIF des descripteurs sélectionnés.

Tableau III.6. La matrice des coefficients de corrélation des descripteurs sélectionnés

Tableau III.7. Les différents paramètres de validation des deux ensembles (train, test)

Tableau III.8. Les concentrations inhibitrices prédites en fonction des concentrations inhibitrices expérimentales pour l'ensemble des analogue Train, Test

Tableau III.9. Les concentrations inhibitrices prédites en fonction des concentrations inhibitrices expérimentales pour l'ensemble des analogue Train.

Tableau III.10. Les concentrations inhibitrices prédites en fonction des concentrations inhibitrices expérimentales pour l'ensemble des analogue Test.

Liste des abréviations

AMM : Autorisation de mise sur le marché.

F : Test de Fisher

IC_{50} : la concentration inhibitrice médiane

MDCK : Madin-Darby Canine Kidney

MLP : Le perceptron multicouche

Q^2 : Coefficient de prédiction

QSAR : Relations Quantitatives Structure -Activité (Quantitative structure-activité Relationship).

QSPR : Relations Quantitatives Propriété-Activité (Quantitative structure propreté Relationship).

RLM : Régression linéaire multiple.

RNA : Réseaux neurone artificiel

R : Coefficient de détermination

R^2 : Coefficient de détermination.

R^2 adj. : Coefficient de détermination ajusté.

RNA : Réseaux de neurones artificiels

Introduction générale

Environ 500 000 ± 700 000 nouvelles structures chimiques sont synthétisées chaque année. Avec le développement de la chimie. Le développement de technologies rentables et efficaces pour tester et prédire les propriétés physiques, chimiques et biologiques des nouveaux composés, ce qui permettrait une estimation des dangers potentiels des composés organiques anciens et nouveaux et une évaluation efficace des risques. Les méthodes des relations quantitatives Structure - propriété (QSPR) et quantitatives structure-activité (QSAR) ont été Utilisé au cours des années pour développer des modèles à estimer / prédire l'activité thérapeutique en fonction des structures chimiques.

La grippe est une maladie respiratoire infectieuse causée par le virus influenza appartenant à la famille des (Orthomyxoviridae). C'est une maladie qui évolue sur un mode épidémique et qui a chaque hiver, un impact socio-économique important. Parmi les trois types de virus influenza (A, B et C), ceux de type A (sous-type H1N1 ou H3N2) et ceux de type B sont responsables des épidémies de grippe saisonnière chez l'homme. Cependant seuls les virus influenza de type A peuvent provoquer des pandémies grippales. Au cours du XXe siècle dernier, ils sont responsables de trois grandes pandémies grippales (1918 H1N1, 1957 H2N2, et 1968 H3N2), causant des millions de mort. Alors que l'humanité craint une pandémie grippale d'origine aviaire provoquée par le sous-type A (H5N1) depuis 2003, elle doit faire face à l'émergence d'un nouveau virus grippal de type A (H1N1) d'origine porcine**.** La propagation rapide de ce virus au niveau mondial a conduit l'Organisation Mondiale de la Santé (OMS) à élever son niveau d'alerte pandémique en phase 6, officialisant la venue de la première pandémie grippale du XXIe siècle. Cela remarque une apparition imprévisible de nouveaux virus de la grippe**.**

Le principal objectif de ce travail est d'obtenu un bon modèle qui garantir l'efficacité de médicament (l'effet thérapeutique) à partir de 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 des analogues de l'oseltamivir et zanamivir.

En plus d’une introduction et d’une conclusion générales, notre travail comporte trois chapitres, dans le premier chapitre nous présenterons une étude bibliographique consiste des généralités sur analogues de zanamivir et oseltamivir.

Dans le deuxième chapitre nous présenterons des généralités sur les méthodes QSAR et la régression non linéaire.

Le troisième chapitre de ce travail, présente les résultats obtenus par l’étude QSAR de l’activité thérapeutique analogues de zanamivir et oseltamivir et leurs discussions

Chapitre I :

Rappels Bibliographiques

I.1 La grippe

I.1.1 définition

La grippe est une maladie respiratoire infectieuse causée par le virus influenza appartenant à la famille des (Orthomyxoviridae). C'est une maladie qui évolue sur un mode épidémique et qui a, chaque hiver, un impact socio-économique important. Parmi les trois types de virus influenza (A, B et C), ceux de type A (sous-type H1N1 ou H3N2) et ceux de type B sont responsables des épidémies de grippe saisonnière chez l'homme. Cependant seuls les virus influenza de type A peuvent provoquer des pandémies grippales. Au cours du XXe siècle dernier, ils sont responsables de trois grandes pandémies grippales (1918 H1N1, 1957 H2N2, et 1968 H3N2), causant des millions de morts **[1]**. Alors que l'humanité craint une pandémie grippale d'origine aviaire provoquée par le sous-type A (H5N1) depuis 2003**[2]**, elle doit faire face à l'émergence d'un nouveau virus grippal de type A (H1N1) d'origine porcine **[3]**. La propagation rapide de ce virus au niveau mondial a conduit l'Organisation Mondiale de la Santé (OMS) à élever son niveau d'alerte pandémique en phase 6, officialisant la venue de la première pandémie grippale du XXIe siècle. Cela remarque une apparition imprévisible de nouveaux virus de la grippe **[4]**

I.1.2 La structure général du virion

Le virion du virus grippal comporte trois composants majeurs : l'enveloppe virale provenant de la cellule hôte, la couche protéique de matrice (M1), et la coré composée de segments d'ARN associés avec les protéines internes. La structure du virus influenza de type A est de beaucoup la plus étudiée. Son enveloppe est porteuse de trois protéines transmembranaires, l'hémagglutinine (HA) en trimère, la neuraminidase (NA) en tétramère, et le canal à proton (M2) en tétramère. L'hémagglutinine est la protéine majeure de surface (~80 %), la deuxième est la neuraminidase (~17 %), la protéine M2 est présente en faible quantité (~16 – 20 molécules pour chaque virion) **[5]**. En dessous de cette enveloppe, il existe une couche protéique de matrice (M1) qui forme un lien entre l'enveloppe et le génome interne. Une faible quantité de protéine d'export nucléaire (NEP, anciennement appelée NS2) est également associée à M1.

Les structures des virus de type B et de type A sont presque similaires. Cependant le type B possède quatre protéines transmembranaires de surface, l'hémagglutinine (HA), la neuraminidase (NA), le canal ionique BM2, et la protéine NB **[6]**.

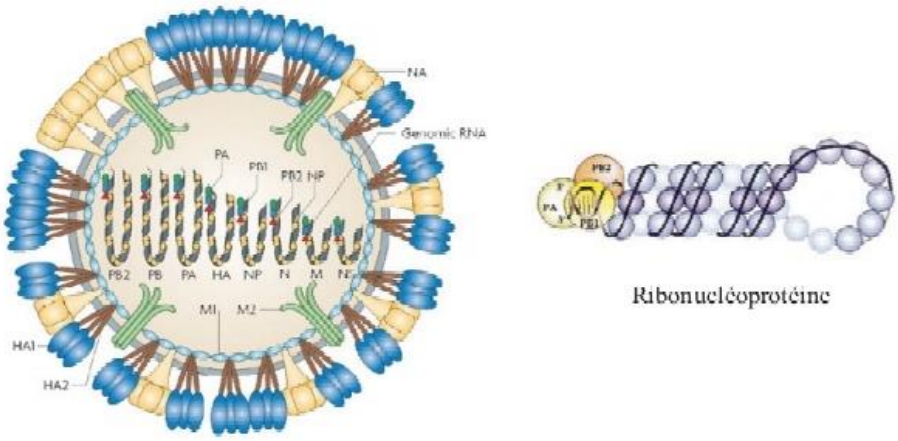


Figure 1.1. Représentation schématique du virus influenza de type A et du complexe ribonucléoprotéique

I.1.3 Le génome viral

Les virus grippaux possèdent un génome segmenté formé d'ARN monocaténaire de polarité négative. Les virus influenza de type A et de type B comportent huit segments de taille variable. Chaque ARN viral (ARNv) est trouvé associé avec plusieurs monomères de nucléoprotéine (NP), et avec un complexe polymérasique composé de trois sous-unités (PB1, PB2 et PA) aux extrémités 5' et 3'. Ce complexe à la structure hélicoïdale est appelé ribonucléoprotéide (RNP) ; il est responsable de la transcription et de la réplication de chaque segment d'ARN viral. La région codante de chaque segment est flanquée par deux régions non codantes aux extrémités 5' et 3'. Ces régions non codantes se composent d'une partie conservée pour tous les segments (11 à 13 et 9 à 12 nucléotides de l'extrémité 5' et3'respectivement) et d'une partie non conservée caractérisée par une taille variable et une séquence nucléotidique différente pour chaque segment **[7]**. Les régions non codantes sont impliquées dans la transcription, la réplication, et l'empaquetage des segments **[8].**

I.1.4 Les protéines virales

Il existe plusieurs protéines dans la structure de virus de type A :

L'hémagglutinine (HA), La neuraminidase (NA), La protéine de canal (A/M2 et BM2), La protéine NB, La protéine de matrice (M1), La nucléoprotéine NP, Le complexe polymérasique (PB1, PB2, PA) ……etc.

Dans cette étude nous traiton de la protéine neuraminidase (NA).

I.1.5 La neuraminidase (NA)

La neuraminidase (NA) du virus transmembranaire de type 2, est transportée directement vers la membrane plasmique apicale dans des cellules MDCK polarisées .une protéine transmembranaire bas latérale de type 2 nous avons étudié la localisation du signal de tri apical de virus de la grippe NA **[9]**, avec activité enzymatique de l'acétyl-neuraminosyl-glycohydrolase de type N(EC3.2.1.18).Le NA est une mesure homozygote fixée à la membrane viral par une séquence hydrophobe unique de environ 29 acide aminés se trouvent prés de l'extrémité N.

La glycoprotéine, obtenue par coloration négative, a la forme de champignon avec une partie supérieure sphérique reliée à une tige mince, une partie de celui-ci, le distal est incorporé dans la membrane viral **[10]**.

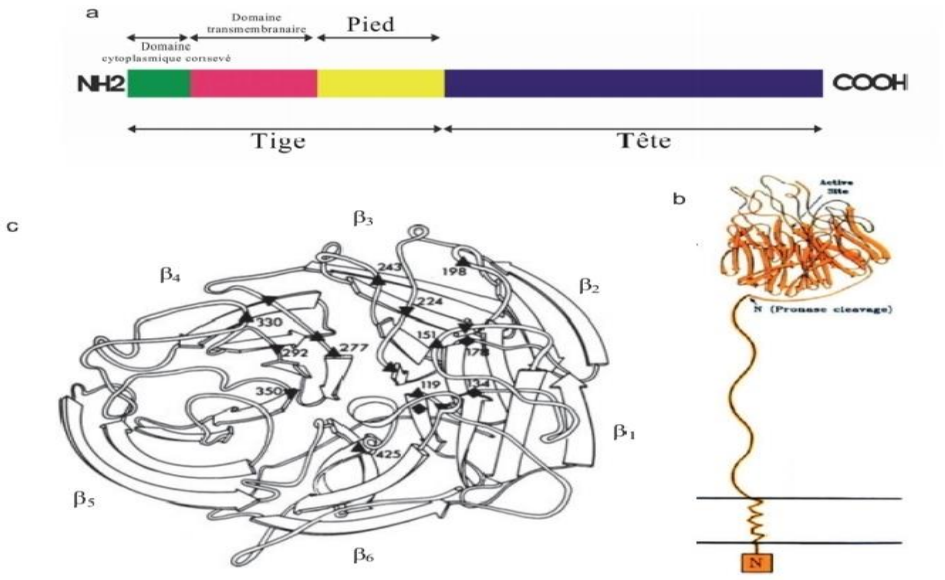


Figure I.2. Structure de la neuraminidase. [a)] Représentation schématique de la structure secondaire de la NA. [b)] La structure monomérique de la NA.

I.1.6 Inhibiteurs de la neuraminidase

Les inhibiteurs de la neuraminidase (INAs) sont des inhibitrice par compétition spécifique l'action de la neuraminidase en bloquant son site actif. Il s'ensuit un blocage du relargage des virions qui s'agrégeant à la surface de la cellule hôte et entre eux, par liaison de leur hémagglutinine aux résidus de l'acide sialique non clivé.

Les premiers inhibiteurs de la neuraminidase (INAs) ont été élaborés en fonction de l'état de transition du substrat durant l'hydrolyse de ce dernier. Le principal analogue obtenu fût l'acide 2-deoxy-2,3-di hydro-N-acetylneuraminique (DANA). Puis, à partir de la structure de la NA de virus influenza de type A et B déterminée par rayons X et de son complexe avec l'acide sialique ou le DANA, deux nouvelles molécules efficaces ont été synthétisées, le zanamivir et l'oseltamivir carboxylate. Le zanamivir (2,4-didéoxy-2,3- didéhydro-4- guanidine sialic acide) est un analogue synthétique de l'acide sialique, analogue à la structure du DANA, avec la substitution d'un groupement guanidinium. L'oseltamivir (ester éthylique (pro drogue), GS4071 (médicament actif) **[11]** ; éthyl-4-acétamido-5-amino-3-(1-ethylpropoxy) -1-cyclohexane-1-carboxylate) est, lui, analogue à un cyclohexane avec la substitution d'un groupement aminé **[12].**

I.2 Les méthodes de conception de médicaments

Le médicament, depuis sa conception jusqu'à commercialisation, en passant par s fabrication, il est soumis à une législation spécifique et rigoureuse ayant pour but de lui garantir la qualité, l'efficacité et la sécurité requis, il passe par 3étapes de base :

I.2.1 Premières phases : Recherche exploratoire

Lors de cette phase, il s'agit de comprendre les mécanismes de la maladie afin de déterminer la cible potentielle que le médicament devra atteindre. Une fois cette cible identifiée, c'est l'étape dite du criblage ou de très nombreuses molécules (plusieurs milliers) sont testées afin de ne retenir que celles éventuellement efficaces (généralement une centaine).

I.2.2 Les essais précliniques

La première phase de tests exigés est celle des essais précliniques. Elle consiste à effectuer des tests in vitro (sur des tissus organiques) et in vivo sur les animaux.

Les résultats doivent permettre de garantir la sécurité pour le passage chez l'homme (test de toxicologie). Un deuxième objectif de cette phase est d'effectuer une première évaluation de l'action que pourra avoir le produit sur l'organisme humain.

I.2.3 Les essais cliniques

Si le produit testé a garanti sa sécurité et si le laboratoire le juge suffisamment intéressent pour continuer à le développer, il passe par les phases sures :

- **La phase I** : Il s'agit de tests de tolérance (toxicité entre autres) sur l'homme sain. Pour que la sécurité soit assurée, les premiers tests sont effectués en administrant des doses très inférieures à la dose efficace et non toxique expérimentale du modèle animal. L'objectif est de déterminer le profil dose/effets secondaires, ainsi que le devenir du produit dans l'organisme : Le produit atteint-il la cible thérapeutique prévue par le modèle animal, et à quel taux ? Comment s'effectue l'élimination du produit, à quelle rapidité ? À la fin de cette phase, l'incertitude demeure élevée. Les tests sont caractérisés par un échantillon de petite taille de patients, et n'assure donc pas une validation fiable des résultats vivants.
- **La phase II** : Les essais cliniques sur des personnes malades. Elles peuvent être découpées en deux phases :
    a) **La phase A** : Est de nature exploratoire et vise à cerner l'efficacité du produit chez l'homme.
    b) **La phase B** : Permet d'affiner les indications notamment sur les doses.
- **Phase III** : Essais effectués à grande échelle, l'indication thérapeutique est choisie. Le médicament peut être mis sur le marché après la délivrance de l'AMM **[13].**

Chapitre II :

Développement de modèle QSAR

II.1. Introduction

Les modèles quantitatifs de QSAR présentent une solution statistique du problème de la difficulté du calcul direct des propriétés physiques et biologiques à partir de la structure. L'intérêt d'un modèle de QSAR est de tirer des informations à partir de l'ensemble des descripteurs numériques caractérisant la structure moléculaire et prédire ainsi les activités biologiques de nouvelles structures **[13].**

II.2. Le principe des méthodes QSPR / QSAR

La modélisation d'une propriété ou d'une activité moléculaire nécessite la disponibilité des informations qui caractérisent les molécules et des informations à partir desquelles cette quantité est dérivée Prédit. Ils peuvent être des descripteurs, mais il existe des méthodes alternatives de caractérisation Moléculaire. Le principe des méthodes QSPR / QSAR est d'utiliser des méthodes d'analyse de données pour trouver une relation mathématique reliant les propriétés moléculaires, appelées descripteurs, à une propriété ou à une activité biologique d'un certain nombre de composés chimiques similaires. L'aspect général d'un tel modèle est le suivant :

$$Propriété\ /\ Activité\ =\ f\ (D1, D2, \ldots Dn.)$$

II.3 Les types des descripteurs

II.3.1 Descripteurs moléculaires théoriques

- **Descripteur 1D :** Il est obtenu directement à partir de la formule brute de la molécule. Exemples : nombre d'atomes, masse moléculaire, etc.
- **Descripteurs bidimensionnels 2D :** la formule développée d'une molécule peut nous donner ce genre de Descripteurs. Nous différencions :
    - Indicateurs constitutionnels bidimensionnels.
    - Indicateurs topologiques bidimensionnels : fournissent des informations sur la taille et la forme en général, les effets sont structurels. Exemples : indice de Wiener, indice de Randić, indice de communication de parité Kier-Hall, Balaban Index, etc.
- **Descripteurs 3D :** estiment les positions relatives des atomes dans l'espace. Ils décrivent des propriétés plus complexes. Donc nos calculs nécessitent des connaissances géométrie tridimensionnelle de la molécule.

- **Descriptions techniques 3D :** la plus importante est la taille moléculaire. La surface accessible par le solvant est le moment d'inertie principal.
- **Descripteurs électroniques 3D :** permettent d'identifier différents types Interactions intermoléculaires et intramoléculaires qui ont une influence majeure sur l'activité biologique de Molécules. Le calcul de la plupart de ces descripteurs nécessite une étude d'ingénierie Dont l'énergie est minimale et comprend souvent la chimie quantique **[14]**.

II.4 Les différents modèles QSAR

II.4.1 La régression linéaire multiple RLM

La régression linéaire multiple est la méthode la plus simple et la plus utilisée pour la modélisation statistique dans les études de la relation entre la structure et l'activité **[15].** La méthode RLM est basée sur l'hypothèse qu'il existe une relation linéaire entre la variable dépendante Y (dans notre cas l'activité biologique) et un certain nombre de variables indépendantes n Xi (ici des descripteurs moléculaires**).** Le but est d'avoir une équation mathématique avec la formule :

$$Y = a0 + a1X1 + a2X2 + a3X3 + ..anXn$$

Où a_i (i = 0, 1, ..., n) sont les coefficients de régression. Cette méthode utilise généralement la variante de MLR qui crée une équation linéaire Plusieurs termes sont utilisés, mais pas toutes les variables indépendantes. Cette méthode est utile lorsque le nombre de descripteurs est important et lorsque les descripteurs Les principaux sont inconnus. De plus, des variables latentes orthogonales peuvent être utilisées dans MLR **[16]**.

II.4.2 Les réseaux de neurones artificiels

L'idée des réseaux de neurones artificiels (ARN) a été inspirée par la façon dont les neurones biologiques traitent l'information.

Cette technique s'avère très prometteuse pour la génération Relations de structure et d'activité en examinant certains des descripteurs les plus importants est utilisé pour contrôler l'activité lors de certaines interactions de plusieurs descripteurs ont été négligés [**17].**

II.5 La méthodologie générale de l'étude QSPR / QSAR

La méthodologie générale de l'étude **QSAR / QSPR** est la suivante :

1- collecte de la base de données

2- Recherche de descripteurs appropriés pour l'activité à l'étude

3- Choisissez la méthode d'analyse des données

4- Valider le modèle

Vérification interne :

Nous utilisons la chaîne d'apprentissage constituée des 2/3 de la base de données. Le but de ce type de vérification est de vérifier la stabilité et la validité du modèle QSAR spécifié.

Vérification externe :

La série de tests, qui comprend généralement 1/3 de la base de données, est utilisée. Le but de cette vérification est de vérifier la puissance prédictive du modèle développé.

D1, D2, … Dn sont des descripteurs des structures moléculaires **[18].**

II.6 Les paramètre de validation de méthode QSAR

II.6.1 Le coefficient de détermination R^2

Nous définissons le coefficient de détermination, qui mesure la proportion de la variance totale que le modèle interprète : R^2

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

$$\mathbf{R^2}= \frac{\sum(Yi\hat{}-\overline{Y})2}{\sum(Yi-\overline{Y})2} = \frac{b1\sum(xi-\overline{x})2}{\sum(Yi-\overline{Y})2}$$

II.6.2 Le coefficient de détermination ajusté R^2_{adj}

Ce coefficient est utilisé en régression multiple car il prend en compte les décomptes des paramètres (descripteur) du modèle **[18]** :

$$R^2 = \frac{R^2(n-1)-p}{n-p-1}$$

II.6.3 Le coefficient de prédiction (Q^2)

Si $\boldsymbol{R^2}$est une mesure de la validité de corrélation, alors $\mathbf{Q^2}$ est une mesure de la validité de la prédiction il mesure le pouvoir prédictif d'un modèle et est une donnée par **[19]** :

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

$$\mathbf{PRESS} = \sum(\boldsymbol{Yexp - Ypred})^2$$

$$\mathbf{SSY = TSS} = \sum(\boldsymbol{Yexp} - \mathbf{Y})^2$$

II.6.4 Test Fisher ($\boldsymbol{F}$)

Le test Fisher permet de justifier la liaison globale entre Yn et Xn, une version dérivée de cet indicateur peut juger du degré de pertinence des variables du modèle.

On peut définir quelques paramètres utilisés dans la régression :

Somme des carrés totaux : $\boldsymbol{TSS = \Sigma\,(Yexp - \bar{Y}exp)^2}$

Somme des carrés expliquée : $\boldsymbol{ESS = \Sigma\,(Ypred - \bar{Y}exp)^2}$

Somme des carrée résiduels **:** $\boldsymbol{RSS = \Sigma\,(Yexp - Ypred)^2}$

Ainsi, $\boldsymbol{TSS = ESS + RSS}$

$\boldsymbol{F}$ peut s'écrire comme suit:

$$\mathbf{F} = \frac{\mathbf{ESS}\,\mathbf{n-p-1}}{\mathbf{pRSS}}$$

Ou n représente le nombre des molécules, p représente le nombre des descripteurs.

II.6.5 Ecart type ($\boldsymbol{s}$)

L'écart type (s) est un autre paramètre habituellement rapporté ; il indique dans quelle mesure la fonction de régression prédite les données observées **[20]**, ce paramètre est donné par :

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

II.6.6 Les facteurs d'inflation de la variance (VIF)

Au sens strict, on parle de multi colinéarité parfaite lorsqu'une des variables explicatives d'un modèle est une combinaison linéaire d'une ou plusieurs autres variables explicatives introduites dans le même modèle **[21]**. L'absence de multi colinéarité parfaite est une des conditions requises pour pouvoir estimer un modèle linéaire.

L'approche la plus classique consiste à examiner les *facteurs d'inflation de la variance* (*FIV*) ou *variance inflation factor* (*VIF*) en anglais :

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

La valeur du paramètre VIF, calculée doit se situer dans la gamme de valeur de 1 à 5 **[22].**

Chapitre III :

Résultats et Discussion

III.1 PREMIRE PARTIE

III.1.1 Introduction

Dans ce chapitre nous allons discuter les études linéaire et non linéaire de la relation quantitative structure-activité inhibitrice de la neuraminidase de virus de la grippe de l'analogue de zanamivir et oseltamivir puissant inhibiteur de la neuraminidase.

III.1.2 La base de données

Le modèle QSPR / QSAR repose fortement sur des données empiriques. Concombre La base de données est essentielle pour développer un tel modèle. Une base de données doit être constitué de données empiriques, aussi fiables que possible et obtenues en suivant ces données.

Nous avons discuté des études QSAR sur les acides benzoïques dérivés carboxyliques, cyclopentanes, isoquinolines, pyrrolidines et composés divers pour leur influence sur les activités inhibitrices de la neuraminidase.

Les 24 analogues de Zanamivir et Oseltamivir ayant une activité inhibitrice sur la Neuraminidase. On a les valeurs de PIC_{50} de chaque analogue **[23]** :

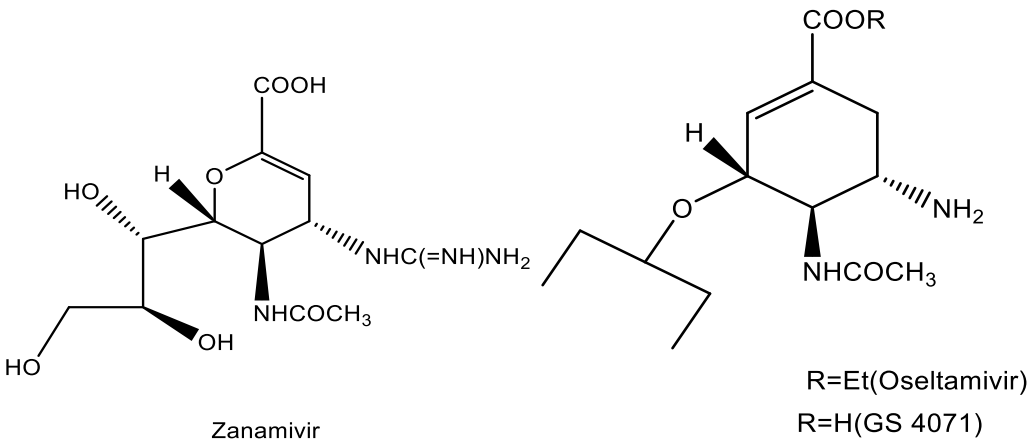


Figure III.1. Structure moléculaire de Zanamivir et Oseltamivir

Tableau III.1. Les analogues de Zanamivir et Oseltamivir utilisés et valeurs des concentrations inhibitrices expérimentales.

N°	Structure	PIC_{50}
1	$C_{10}H_{18}N_4O_3$	3.97
2	$C_{14}H_{26}N_4O_3$	6.44
3	$C_{14}H_{26}N_4O_3$	8.89
4	$C_{14}H_{23}F_3N_4O_4$	6.13
5	$C_{10}H_{19}N_3O_3$	4.36
6	$C_{12}H_{23}N_3O_4$	5.31
7	$C_{18}H_{26}N_2O_5$	7.38
8	$C_{14}H_{18}N_4O_5$	5.34
9	$C_{13}H_{15}NO_5$	3.14
10	$C_{12}H_{14}N_4O_3$	3.63
11	$C_{16}H_{22}N_2O_3$	3.68
12	$C_{10}H_{10}N_2O_4$	2.33
13	$C_{11}H_{12}N_2O_5$	2.68
14	$C_9H_{12}N_4O_4S$	4.12
15	$C_9H_{10}N_2O_5S$	2.78
16	$C_{12}H_{14}N_4O_3$	3.09
17	$C_{11}H_{14}N_2O_3S$	3.97
18	$C_{11}H_{17}N_3O_5$	5.89
19	$C_{13}H_{21}N_3O_5$	7.06
20	$C_{15}H_{25}N_3O_5$	8.3
21	$C_{13}H_{21}N_3O_5$	8.93
22	$C_{19}H_{25}N_3O_6$	8.7
23	$C_{15}H_{21}N_3O_5$	8.7
24	$C_{13}H_{21}N_5O_5$	6.6

III.1.3 Les calculs

III.1.3.1 Génération des descripteurs

Les descripteurs sont des valeurs numériques qui encodent les caractéristiques physicochimiques des molécules à partir de leurs représentations structurales.

La génération des descripteurs constitue l'étape cruciale lors de l'analyse de QSAR et la qualité d'un modèle dépendra essentiellement de la nature des descripteurs générés.

III.1.4 Les logiciels utilisés

Plusieurs logiciels sont disponibles pour faire les calculs des descripteurs, parmi ceux, que nous avons utilisé dans nos travaux ; **CHEMOFFICE 16, PaDEL**, **CSV viewer**,**IBM SPSS 21**, **STATISTICA10.**

Nous dessinons d'abord des molécules analogues d'ozeltamivir zanamivir dans CHEMDRAW Professional 16.0, puis nous vérifions la validité du dessin dans CHEM3D16.0

Ensuite on utilise le programme PADEL qui calcule environ 1445 descripteurs, et le fonctionnement de ce programme est le suivant :

- Sélectionnez un seul fichier structurel ou un répertoire contenant les fichiers structurels des molécules. Les formats de fichiers les plus courants (par exemple MDL mol, SMILES) sont pris en charge, mais le format de fichier recommandé est MDL mol.
- Sélectionnez un répertoire pour enregistrer les descripteurs calculés. Les descripteurs seront enregistrés au format de fichier CSV (valeurs séparées par des virgules).
- Sélectionnez l'option "1D & 2D" car nous allons calculer les descripteurs 1D et 2D.
- Sélectionnez l'option "Enlèvement de sel" pour supprimer les sels comme Na^+, Cl^- de la molécule avant le calcul des descripteurs.
- Cochez l'option "Détecter l'aromaticité" pour supprimer les informations d'aromaticité existantes et détecter automatiquement l'aromaticité dans la molécule avant le calcul des descripteurs.
- Cochez l'option "Normaliser les groupes nitro pour normaliser les groupes nitro à N (: O) :O. Ceci est nécessaire pour que les descripteurs étendus des atomes topochimiques (ETA) soient Calculés correctement.

- Cochez l'option "Journal" pour générer un fichier journal.
- Cochez l'option "Conserver l'ordre des molécules" pour conserver l'ordre des molécules dans les fichiers structurels pour le fichier descripteur.

Tous les descripteurs calculés par PDEL seront enregistrés dans un fichier au format CSV où nous ouvrons ce fichier par le programme de CSV viewer.

Nous avons utilisé la régression linéaire multiple avec le logiciel IBM SPSS pour réduire le nombre de descripteurs

Pour le développement des modèles QSAR, deux méthodes d'analyse de données ont été utilisées : la régression linéaire multiple (RLM ou MLR en anglais), les réseaux de neurones artificiels (RNA ou ANN en anglais). L'approche MLR a été réalisée à l'aide du logiciel IBM SPSS. Les approches ANN et, quant à elle, été réalisée à l'aide du logiciel STATISTICA.

III.1.5 Les modèles choisis

III.1.5.1 Le modèle MLR

Pour l'obtention du model MLR en a utilisé le du logiciel IBM SPSS suivant les étapes ci-dessous :

1- D'abord, nous ouvrons le fichier Excel qui contient la base de données dans le Programme IBM SPSS (nombre de descripteurs réduits + les analogues de molécule qu'on a traités).
2- Pour réaliser l'analyse, cliquez sur **Analyse**, **Régression**, puis **Linéaire**.
3- En cliquant sur « « ajouter », pour insérer la variable dépendante et les variable(s) Indépendante(s) dans les boîtes appropriées.
4- En a choisi la méthode **Ascendante pas à pas**.

III.1.5.2 Le modèle des réseaux de neurones artificiels RNA

L'utilisation des réseaux de neurones artificiels (RNA) est nécessaire et est devenue une technique de modélisation importante. Elle est largement utilisée pour les études de QSAR non linéaires et est un outil puissant pour construire des modèles prédictifs. Dans le cadre de ce travail, les réseaux de neurones artificiels (RAN) ont été utilisés pour construire un modèle non linéaire sur la base des mêmes descripteurs utilisés dans le cas du modèle MLR. Un réseau

typique à trois couches de type perceptron multicouche (MLP) avec une couche d'entrée, une couche cachée et une couche de sortie a été adoptée dans cette étude.

Des travaux théoriques ont montré qu'une seule couche cachée suffit pour que l'RNA se rapproche d'une fonction non linéaire complexe et que de nombreux résultats expérimentaux semblent Confirmer qu'une couche cachée peut suffire à la plupart des problèmes de prédiction**.**

Ainsi, l'utilisation d'une régression neuronale nécessite l'optimisation de l'architecture du réseau neuronal. Dans cette étude, la fonction sigmoïde a été utilisée comme fonction de transfert de la couche cachée et la fonction identité comme fonction de transfert de la couche de sortie.

Pour l'obtention du model RNA en a utilisé le du logiciel STATISTICA (STATISTICA 10.0, Stat Soft, Inc., Tulsa, OK, USA).

III.2 DUXIEME PARTIE

III.2.1 Résultats et discussion

III.2.1.1 Sélection des descripteurs

Tableau III.2. Les descripteurs utilisés pour le développement des modèles QSAR :

Type	Descripteurs	La description
2D	AMR	Réfractivité molaire-Ghose_Crippen
2D	ATSC4i	Descripteur de structure topologique (ATS) (Auto corrélation)

III.2.1.2 Réfractivité molaire (AMR)

La réfractivité est un cas particulier de volume moléculaire, c'est une propriété ramenée à une quantité de matière, ainsi que la réfractivité est la qualité de réfringence d'un corps, elle est utilisée dans la radioélectricité, en biologie… **[25].**

En chimie, la réfractivité moléculaire est un critère important pour mesurer lefacteur stérique, elle est importante dans le cas où le substituent possède des électrons π ou des doublets libres **[26].**

La réfractivité moléculaire est définie par la relation suivante :

$$\mathbf{Réf} = \boldsymbol{v}\frac{\boldsymbol{n^2_1}}{\boldsymbol{n^2+1}} = \frac{\mathbf{4}}{\mathbf{3}}\boldsymbol{\pi NaPol}$$

v: volume moléculaire

n : l'indice de réfraction

Na : nombre d'Avogadro

Pol : polarisation

III.2.1.3 Les descripteurs de structure topologique (ATS), pondérés

Selon les dessins (ATSc1, ATSc2, ATSc3, ATSc4, ATSc5), ATS

Descripteur, pondéré en fonction de la polarité (ATSp1, ATSp2, ATSp3, ATSp4, ATSp5) **[27].**

III.2.2 Le modèle RLM

Tableau III.3. Les valeurs de pIC_{50} expérimentales avec les descripteurs sélectionnés.

N°	pIC_{50}	AMR	ATSC4i	ETAT
1	3.97	55.1366	-6.5656	Train
2	6.44	68.4558	-2.5380	Train
3	8.89	68.4559	-2.5380	Train
4	6.13	68.4560	-2.5380	Train
5	4.36	68.4561	-2.5380	Train
6	5.31	68.4562	-2.5380	Train
7	7.38	68.4563	-2.5380	Train
8	5.34	68.4564	-2.5380	Train
9	3.14	40.5184	-10.3479	Train

10	3.63	38.0335	-17.5333	Train
11	3.68	51.1523	5.0824	Train
12	2.33	29.914	-4.6538	Test
13	2.68	32.9524	-9.9699	Train
14	4.12	37.1311	-19.1185	Test
15	2.78	35.7602	0.7399	Train
16	3.09	38.3857	-10.2219	Test
17	3.97	40.9163	-6.1027	Train
18	5.89	64.5262	-7.4211	Train
19	7.06	72.352	-2.8456	Train
20	8.3	80.1778	7.0467	Train
21	8.93	73.1138	-19.7878	Train
22	8.7	74.8519	-31.0929	Train
23	8.7	82.0446	-25.0094	Train
24	6.6	73.6329	-9.3230	Test

Le modèle MLR obtenu pour la prédiction de la concentration inhibitrice des dérivés de Zanamivir et Ozeltamivir, en utilisant 24 composés, est donné par la relation linéaire ci-dessous.

Les valeurs de **pIC$_{50}$** prédites par le modèle MLR pour l'ensemble des données (24 composés) Sont présentées dans le **Tableau III.4.**

La meilleure corrélation est exprimée par la relation mathématique suivante :

$$\boldsymbol{pIC}_{50} = -1.603 + 0.115(\boldsymbol{AMR}) - 0.052(\boldsymbol{ATSC4i}) \quad (1)$$

Après avoir extrait la relation, nous substituons les valeurs les descripteurs pour chaque composé et notons les valeurs de **pIC$_{50}$** obtenues ou plus les valeurs de **pIC$_{50}$** sont petites plus le composé sera efficace pour inhiber.

Les signes des coefficients de régression suggèrent la direction de l'influence des descripteurs dans un modèle donné. D'après les coefficients de régression de l'équation (**eq.1**), il convient de noter que le coefficient de régression du descripteur AMR est de signes positifs.

Ce descripteur a eu une influence positive et les valeurs les plus élevées vont contribuer à améliorer le pouvoir inhibiteur neuraminidase.

Tableau III.4. Les valeurs de pIC_{50} prédites obtenue.

N°	pIC_{50} observées	pIC_{50} prédite	ETAT
1	3.97	5. 0 7	Train
2	6.44	6. 40	Train
3	8.89	6.40	Train
4	6.13	6.40	Train
5	4.36	6.40	Train
6	5.31	6.40	Train
7	7.38	6.40	Train
8	5.34	6.40	Train
9	3.14	6.80	Train
10	3.63	3.68	Train
11	3.68	4.54	Train
12	2.33	2.07	Test
13	2.68	2.70	Train
14	4.12	3.66	Test
15	2.78	2.54	Train
16	3.09	3.34	Test
17	3.97	3.41	Train
18	5.89	6.20	Train
19	7.06	6.58	Train
20	8.3	7.98	Train
21	8.93	7.83	Train
22	8.7	8.62	Train
23	8.7	9.13	Train
24	6.6	7.34	Test

Après avoir effectué tous les calculs en substituant les valeurs de descripteur pour chaque composé, nous obtenons les valeurs **pIC$_{50}$ prédites**, ou plus les valeurs d'inhibition est petites, donc une meilleure efficacité du médicament.

A travers nos résultats nous notons que c'est **le composé 12** qui a la plus grande efficacité pour inhiber la neuraminidase de virus de la grippe.

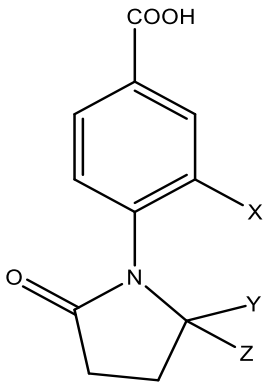


Figure III.2. La structure moléculaire de composé 12

Cet inhibiteur de neuraminidase de virus de la grippe de famille de **Benzoïque acides** de type **A by I** de racine **X (NHCH(C2H5)2** et **Y(H)** et **Z(H)**

Selon la structure générale de Benzoïque acides :

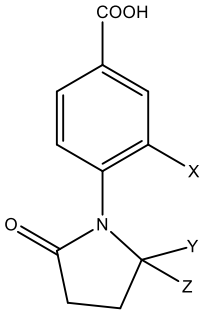


Figure III.3. La structure moléculaire général de Benzoïque acides

Pour évaluer la multi-colinéarité entre les deux descripteurs pertinents, leurs facteurs de variation d'inflation (VIF) ont été calculés. Si VIF est égal à 1, aucune inter corrélation n'existe pour chaque variable ; si VIF est situé dans l'intervalle de 1 à 5, la sélection des descripteurs est acceptable, et si VIF est supérieur à 10, le modèle associé aux descripteurs sélectionnés est instable et une nouvelle sélection est nécessaire **[28].** D'après le **Tableau III.5,** tous les descripteurs ont des valeurs VIF de 1 à 5.8, ce qui indique que la sélection effectuée est acceptable car les valeurs de VIF est inférieure à 10, et que les descripteurs obtenus ne sont pas corrélés entre eux.

Tableau III.5. Valeurs du coefficient VIF des descripteurs sélectionnés.

Descripteurs	**ATSC4i**	**AMR**
VIF	4.608	5.814

Afin d'étudier la corrélation entre les 2 descripteurs sélectionnés, la matrice de corrélation a été établie. La valeur du coefficient de corrélation de chaque paire de descripteurs sélectionnés a été examinée (**Tableau III.6**). La plus grande valeur du coefficient de corrélation appartenant à la paire de descripteurs **AMR**/ **pIC$_{50}$** étant égale à 0,885, cela signifie que tous les descripteurs sélectionnés sont indépendants les uns des autres.

Tableau III.6. La matrice des coefficients de corrélation des descripteurs sélectionnés

	pIC$_{50}$	**AMR**	**ATSC4i**
pIC50	1	0.885	0.225
AMR		1	-0.13
ATSC4i			1

III.2.3 Le modèle des réseaux de neurones artificiels RNA

L'objectif général d'un RNA est de trouver la configuration des poids de connexion entre neurones pour qu'il associe à chaque configuration d'entrée, une réponse adéquate. L'utilisation d'un RNA se fait en deux temps. Tout d'abord une phase d'apprentissage qui est chargée d'établir des valeurs pour chacune des connexions du réseau, puis une phase

d'utilisation proprement dite, où l'on présente au réseau une entrée et où il nous indique en retour « sa » sortie calculée **[29].**

Tableau III.7. Les différents paramètres de validation des deux ensembles (train, test)

Le paramètre de validation (RNA)		
Ntrain**= 20**	**N**test= 4	**N** train, test= 24
S = **0.57**	**S =** 0.54	**S =** 0.78
R = **0.966**	**R =** 0.961	**R =** 0.91
R^2 = **0.93**	**R^2 =** 0.92	**R^2 =** 0.82
R^2 ajusté = **0.92**	**R^2 ajusté =** 0.91	**R^2 ajusté =** 0.81
Q^2= **0.92**	**Q^2=** 0.66	**Q^2=** 0.88
Variation de F= **130.38**	**Variation de F=** 21.32	**Variation de F=** 79.01

Le coefficient de corrélation linéaire R, est une normalisation de la covariance par le produit des écart-type des variables.

L'objectif de coefficient de corrélation linéaire R est de quantifier la liaison entre deux variable X et Y de manière à mettre en évidence le sens de la liaison et son intensité.

R^2 qui appelle coefficient de détermination s'interprète comme la proportion de variance de Y (resp.X) linéairement expliquée par X (respY), on peut faire le rapprochement avec les résultats produit avec la régression linéaire.

Ainsi r = 0.9, on voit que la liaison est fort puisqu'elle rapproche de1. (Dans le cas test et train)

R^2 =0.81 on peut dire que 81% de la variance de Y est expliquée par X (et inversement).

On utilise le coefficient empirique du khi-deux noté Q^2

Les valeurs de Q^2 sont positives ou nulles, plus la valeur est grande et plus l'écart à l'indépendance observé sur l'échantillon est important

On a Q^2 = 0.88 (dans le cas de test, train).

Tableau III.8. Les concentrations inhibitrices prédites en fonction des concentrations inhibitrices expérimentales pour l'ensemble des analogue Train, Test

N°	pIC_{50} observes	pIC_{50} - Output	pIC_{50} - Résiduels	ETAT
1	3,970000	3,958839	0,01116	Train
2	6,440000	8,882807	-2,44281	Train
3	8,890000	8,547309	0,34269	Train
4	6,130000	6,248540	-0,11854	Train
5	4,360000	4,168707	0,19129	Train
6	5,310000	5,770036	-0,46004	Train
7	7,380000	6,736503	0,64350	Train
8	5,340000	5,571507	-0,23151	Train
9	3,140000	3,098428	0,04157	Train
10	3,630000	3,638276	-0,00828	Train
11	3,680000	3,645043	0,03496	Train
12	2,330000	2,492648	-0,16265	Test
13	2,680000	2,681583	-0,00158	Train
14	4,120000	3,447328	0,67267	Test
15	2,780000	2,790320	-0,01032	Train
16	3,090000	2,543871	0,54613	Test
17	3,970000	4,035437	-0,06544	Train
18	5,890000	5,850385	0,03962	Train
19	7,060000	7,054044	0,00596	Train

20	8,300000	8,332426	-0,03243	Train
21	8,930000	8,916315	0,01368	Train
22	8,700000	8,929999	-0,23000	Train
23	8,700000	8,930000	-0,23000	Train
24	6,600000	8,930000	-2,33000	Test

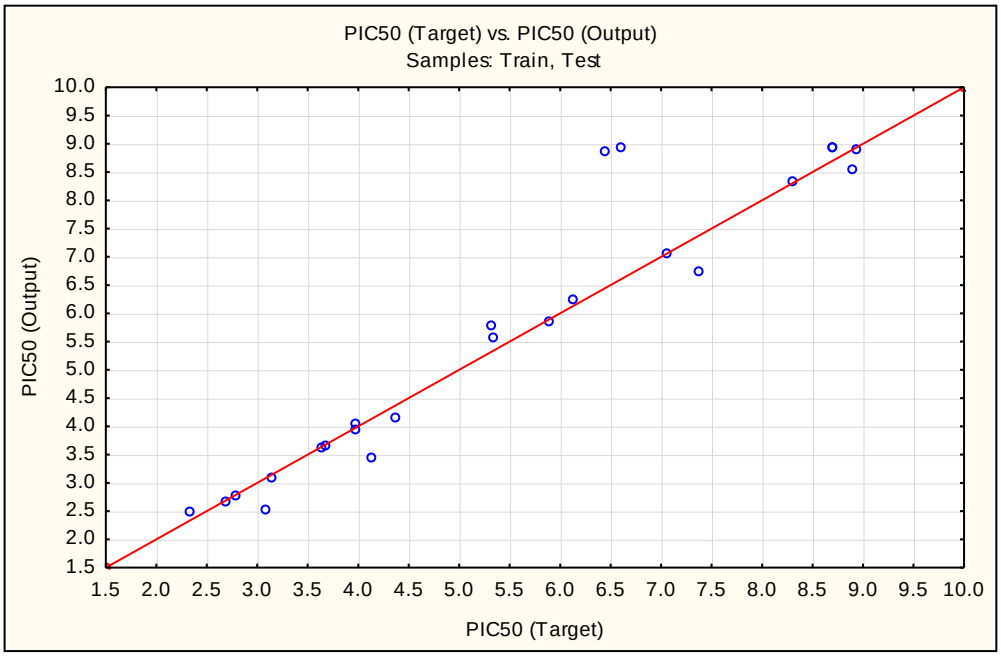


Figure III.4. Représentation graphique des concentrations inhibitrices prédites en fonction des concentrations inhibitrices expérimentales pour l'ensemble des analogue Train, Test

Tableau III.9. Les concentrations inhibitrices prédites en fonction des concentrations inhibitrices expérimentales pour l'ensemble des analogue Train.

N°	pIC_{50}observed	pIC_{50} - Output	Résiduels	ETAT
1	3,970000	3,958839	0,01116	Train
2	6,440000	8,882807	-2,44281	Train
3	8,890000	8,547309	0,34269	Train
4	6,130000	6,248540	-0,11854	Train
5	4,360000	4,168707	0,19129	Train

6	5,310000	5,770036	-0,46004	Train
7	7,380000	6,736503	0,64350	Train
8	5,340000	5,571507	-0,23151	Train
9	3,140000	3,098428	0,04157	Train
10	3,630000	3,638276	-0,00828	Train
11	3,680000	3,645043	0,03496	Train
13	2,680000	2,681583	-0,00158	Train
15	2,780000	2,790320	-0,01032	Train
17	3,970000	4,035437	-0,06544	Train
18	5,890000	5,850385	0,03962	Train
19	7,060000	7,054044	0,00596	Train
20	8,300000	8,332426	-0,03243	Train
21	8,930000	8,916315	0,01368	Train
22	8,700000	8,929999	-0,23000	Train
23	8,700000	8,930000	-0,23000	Train

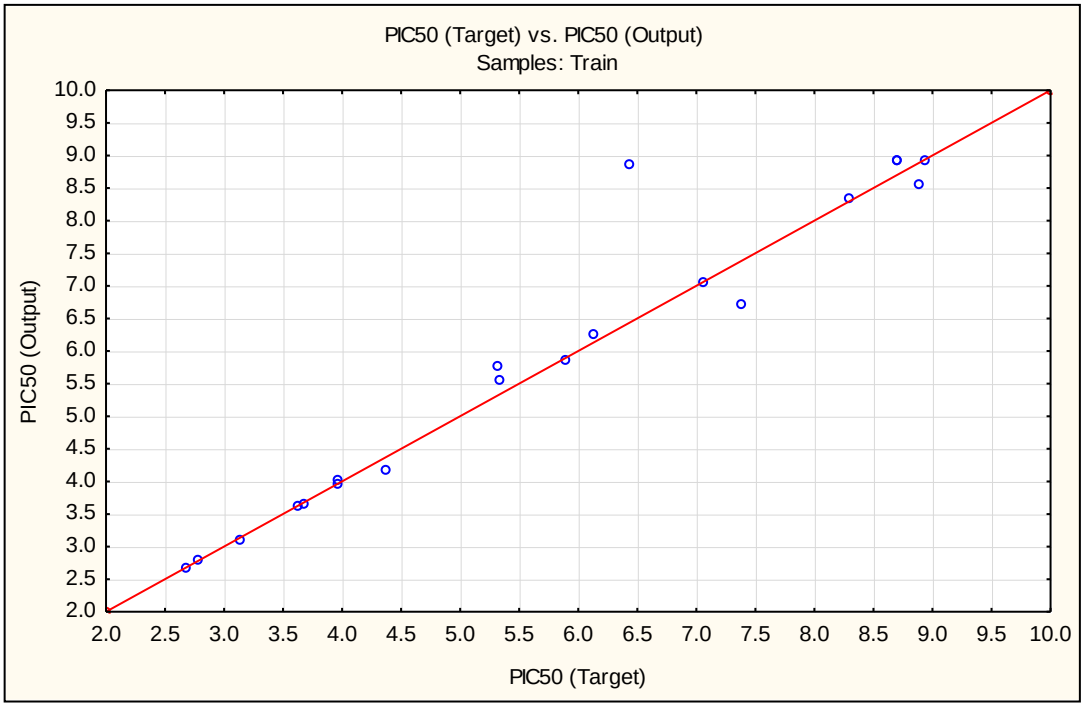


Figure III.5. Représentation graphique des concentrations inhibitrices prédites en fonction des concentrations inhibitrices expérimentales pour l'ensemble des analogue Train.

Tableau III.10. Les concentrations inhibitrices prédites en fonction des concentrations inhibitrices expérimentales pour l'ensemble des analogue Test.

	pIC_{50} observés	pIC_{50} - Output	pIC_{50} - Résidues	ETAT
12	2,330000	2,492648	-0,16265	TEST
14	4,120000	3,447328	0,67267	TEST
16	3,090000	2,543871	0,54613	TEST
24	6,600000	8,930000	-2,33000	TEST

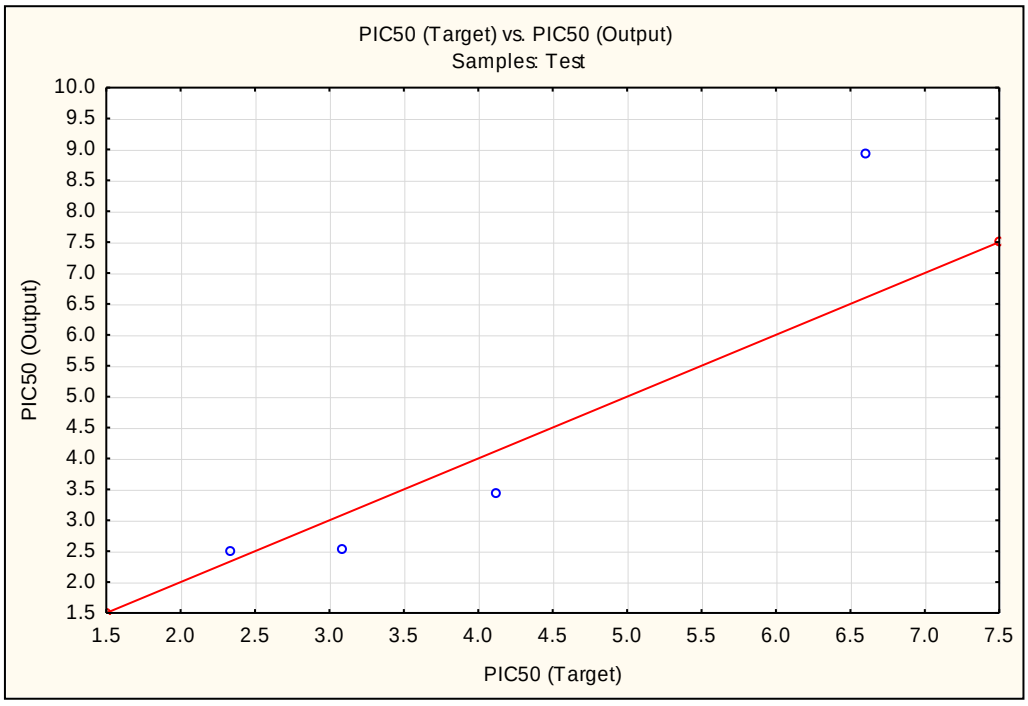


Figure III.6. Représentation graphique des concentrations inhibitrices prédites en fonction des concentrations inhibitrices expérimentales pour l'ensemble des analogue Test.

Conclusion Général

Dans ce travail nous avons fait une étude QSAR (Quantitative des relations structure activité) des analogues de zanamivir et oseltamivir afin de déterminer la relation entre l'activité thérapeutique et leur structure chimique, on a utilisé la régression non linéaire et les réseaux de neurones artificiels, nous avons établi des relations satisfaisantes entre de nombreuses descriptions et l'activité des récepteurs d'antigènes (neuraminidase), avec vérification mutuelle.

La modélisation d'une propriété ou d'une activité moléculaire nécessite la disponibilité des informations qui caractérisent les molécules et des informations à partir desquelles cette quantité est dérivée Prédit. Ils peuvent être des descripteurs, mais il existe des méthodes alternatives de caractérisation Moléculaire. Le principe des méthodes QSPR / QSAR est d'utiliser des méthodes d'analyse de données pour trouver une relation mathématique reliant les propriétés moléculaires, appelées descripteurs, à une propriété ou à une activité biologique d'un certain nombre de composés chimiques similaires.

À la fin du travail, on a obtenu deux descripteurs AMR et ATSC4i avec un coefficient de détermination R^2 = 0.78 et 0.83 et de coefficient empirique de khi Carré de test train Q^2= 0.88

Les résultats obtenus, montre que le modèle proposé dans ce travail peut prédire bonne activité thérapeutique,

Les résultats ont montré que le modèle proposé dans ce travail peut prédire l'activité avec précision et que les descriptions sélectionnées sont pertinentes. Les résultats montrent que le modèle proposé dans ce travail peut prédire l'activité avec précision et que les descriptions sélectionnées sont appropriées.

Les méthodes QSAR sont faciles à mettre en œuvre, ils fournissent un résultat immédiat et leurs coûts sont sans commune mesure avec ceux engendrés par la réalisation de tests expérimentaux. Ils peuvent être réalisés par des laboratoires équipés des logiciels adaptés ou par le demandeur lui-même s'il dispose des outils informatiques et des personnes compétentes pour interpréter et valider les résultats.

Références

1. **Kilbourne ED. Influenza pandemics of the 20th century. Emerg Infect Dis. 2006 Jan;12(1):9-14. doi: 10.3201/eid1201.051254. PMID: 16494710; PMCID: PMC3291411.Peiris, J. S. M., M. D. de Jong, and Y. Guan.** 2007. Avian influenza virus (H5N1): a threat to human health. Clinical Microbiology Reviews 20:243-267.
2. **Girard, M. P., J. S. Tam, O. M. Assossou, and M. P. Kieny.** 2010. The 2009 A (H1N1) influenza virus pandemic: A review. Vaccine 28:4895-4902.
3. **De Clercq, E.** 2006. Antiviral agents active against influenza A viruses. Nature Reviews Drug Discovery 5:1015-1025
4. **Nayak, D. P., R. A. Balogun, H. Yamada, Z. H. Zhou, and S. Barman.** 2009. Influenza virus morphogenesis and budding. Virus Research 143:147-161.
5. **Brassard, D. L., G. P. Leser, and R. A. Lamb.** 1996. Influenza B virus NB glycoprotein is a component of the virion. Virology 220:350-360.
6. **Robertson, J. S. 1979.** 5' and 3' terminal nucleotide sequences of the RNA genome segments of influenza virus. Nucleic Acids Research 6:3745-3757.
7. **Luytjes, W., M. Krystal, M. Enami, J. D. Parvin, and P. Palese.** 1989. Amplification, expression, and packaging of a foreign gene by influenza virus. Cell 59:1107-1113.
8. https://doi.org/10.1128/jvi.70.9.6508-6515.1996
9. **Air G.M., Laver W.G.,** 1989. The neuraminidase of influenza virus. Proteins, 6, 341–56.
10. **Taylor N.R., Von Itzstein M.,** 1994, Molecular modeling studies on ligand binding to sialidase from influenza virus and the mechanism of catalysis. J Med Chem, 37, 616–24.
11. 2008 de l'AFSAPPS, documents fournis par le Leem - Site http://www.med.univ-rennes1.fr/resped/s/pharmaco/BF1/dev.htm
12. **M.Ramdic.J.Am.chem.**Soc 1975,97,6609
13. OECD http ://WWW.Oecd.org/chemical
14. **H.Wiener.J.Am.**chem.Soc 1947,69,17
15. **L. B,LH.Hall,W.J.Murray.J.**Pharma.Sci 1975,64,1971
16. **Patel etal.**2010,2014 et 2017; Masandetal 2014,2018a,2018 b

17. **A.Z.Dudck,T.Arodz,and J.**Galvez.Caputational Méthods indeveloping quantitative structuro-activity relationships (qsar) : revieau.Cambinatorial Chemistry.High throughput Screening,9:213-228,2006,23,26

18. F. Bonachorcs.Les triplets pharmacophoriques flous : développement et application [thèse en ligne]. P4D thèses, Lille : Université Lille 1 science et technologie,2011, p :22-37

19. **J.J.P.Stewart**, Optimization of parameters for semiempirical methods I. Method. J.

20. **Worachartcheewan, Apilak, et al.** "QSAR study of H1N1 neuraminidase inhibitors from influenza a virus." Letters in Drug Design & Discovery 11.4 (2014): 420-427.

21. **Leach, A.R. and V.J.** Gillet, An Introduction to Chemoinformatics. 2007: Springer Netherlands

22. **Foucart, T.,** Colinéarité et régression linéaire. Mathématiques et sciences humaines. Mathematics and social sciences, 2006(173).

23. **Confais, J. and M.** Le Guen, Premiers pas en régression linéaire avec SAS®. 2007.

24. **P.J. Taylor.1990**, Hydrophobic Properties of Drugs, In Quantitative Drug Design‖, PergamonPress, Oxford (UK), 4, 241–294.

25. Site internet : www.formules-physique.com

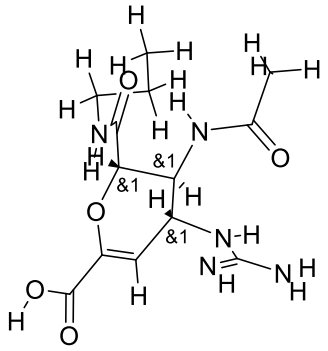
26. **N. Melkemi and S. Belaidi,** Structure-Property Relationships and QSAR Modeling of detoxication Properties of some 1,2-Dithiole-3-thione Derivatives, International journal of chemical research vol 4, pp. 134-139, 2012.

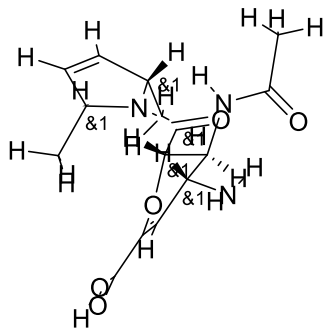
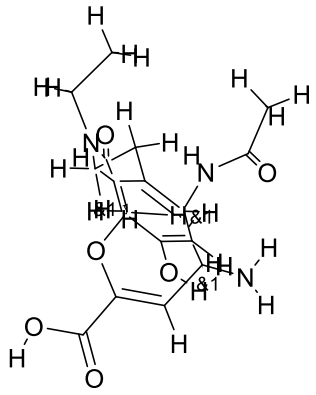
27. **https://doi.org/10.5487/TR.2019.35.4.343**

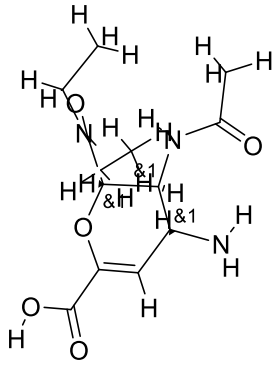
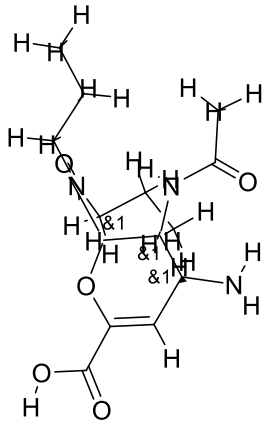
28. **Jaiswal, M., et al.,** Carbonic anhydrase inhibitors: the first QSAR study on inhibition of tumor-associated isoenzyme IX with aromatic and heterocyclic sulfonamides. Bioorganic & medicinal chemistry letters, 2004. **14**(12): p. 3283-3290.

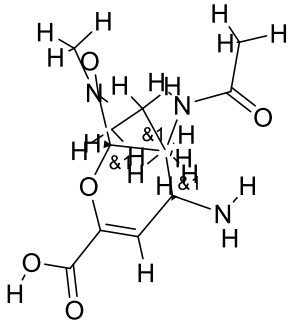
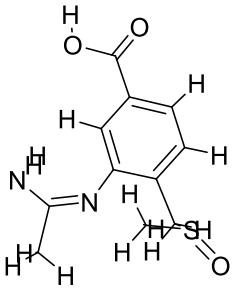
29. **Schmitt, A., et al.** Les réseaux de neurones artificiels. Bulletins et mémoires de la Société d'Anthropologie de Paris, 2001. **13(1): p. 2.**

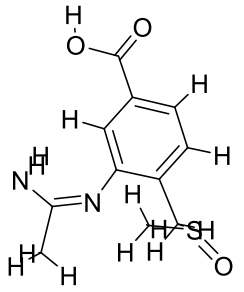
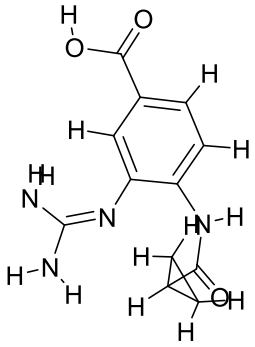
Annexe 01

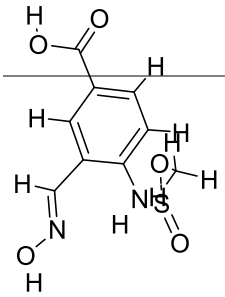
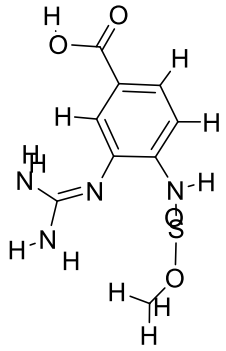
N°	STRUCTURE	pIC_{50}
1	 $C_{13}H_{21}N_5O_5$	5.0791

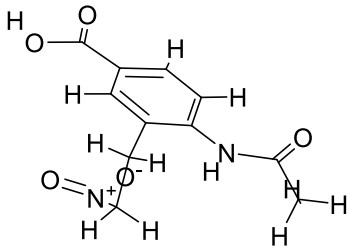
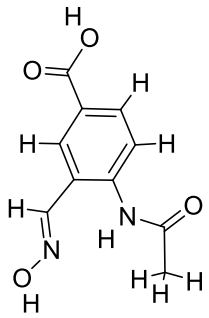
2	 $C_{15}H_{21}N_3O_5$	6.4014
3	 $C_{19}H_{25}N_3O_6$	6.4014

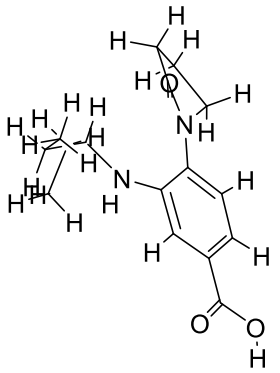
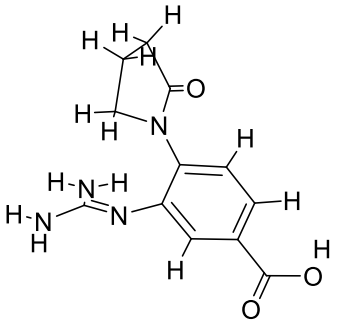
4	 $C_{13}H_{21}N_3O_5$	6.4014
5	 $C_{15}H_{25}N_3O_5$	6.4014

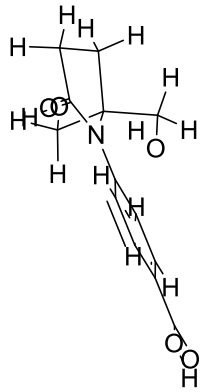
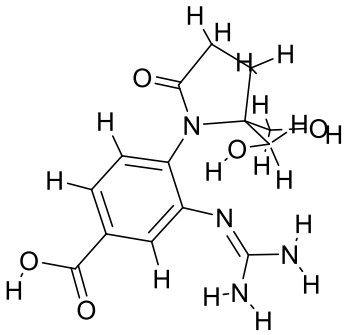
6	 $C_{13}H_{21}N_3O_5$	6.4014
7	 $C_{11}H_{14}N_2O_3S$	6.4014

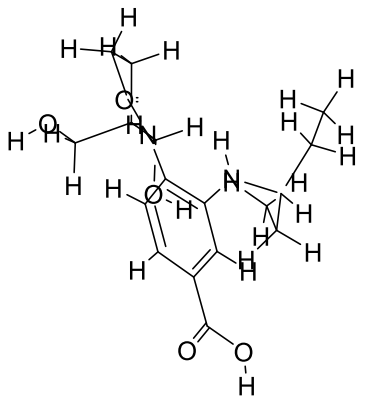
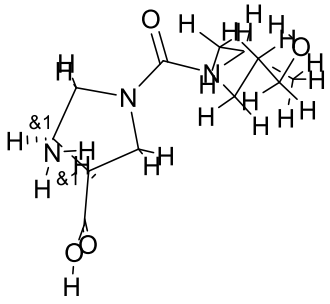
8	 $C_{11}H_{14}N_2O_3S$	6.4014
9	 $C_{12}H_{14}N_4O_3$	6.8007

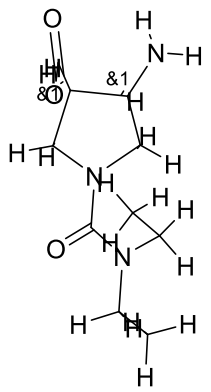
10	 $C_9H_{10}N_2O_5S$	3.6826
11	 $C_9H_{12}N_4O_4S$	4.5438

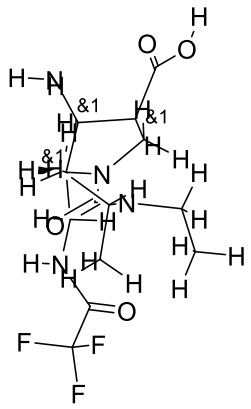
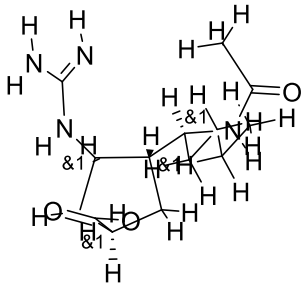
12	 $C_{11}H_{12}N_2O_5$	2.0791
13	 $C_{10}H_{10}N_2O_4$	2.7050

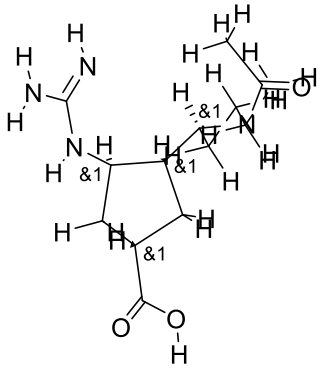
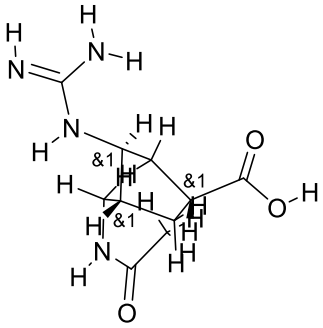
14	 $C_{16}H_{22}N_2O_3$	3.6612
15	 $C_{12}H_{14}N_4O_3$	2.5479

16	 $C_{13}H_{15}NO_5$	3.3429
17	 $C_{14}H_{18}N_4O_5$	3.4197

18	 $C_{18}H_{26}N_2O_5$	6.2034
19	 $C_{12}H_{23}N_3O_4$	6.5882

20	 $C_{10}H_{19}N_3O_3$	7.9839
21		7.8341

	 $C_{14}H_{23}F_3N_4O_4$	
22	 $C_{14}H_{26}N_4O_3$	8.6218

23	 $C_{14}H_{26}N_4O_3$	9.1326
24	 $C_{10}H_{18}N_4O_3$	7.3496

Annexe 2 :

N°	Structure	pIC_{50}	AMR	ATSC4i
1	$C_{10}H_{18}N_4O_3$	3.97	55.13	-6.56
2	$C_{14}H_{26}N_4O_3$	6.44	68.45	-2.53
3	$C_{14}H_{26}N_4O_3$	8.89	68.45	-2.58
4	$C_{14}H_{23}F_3N_4O_4$	6.13	68.45	-2.53
5	$C_{10}H_{19}N_3O_3$	4.36	68.45	-2.53
6	$C_{12}H_{23}N_3O_4$	5.31	68.45	-2.53
7	$C_{18}H_{26}N_2O_5$	7.38	68.45	-2.53
8	$C_{14}H_{18}N_4O_5$	5.34	68.45	-2.53
9	$C_{13}H_{15}NO_5$	3.14	40.51	-10.34
10	$C_{12}H_{14}N_4O_3$	3.63	38.03	-17.53
11	$C_{16}H_{22}N_2O_3$	3.68	51.15	5.08
12	$C_{10}H_{10}N_2O_4$	2.33	29.91	-4.65
13	$C_{11}H_{12}N_2O_5$	2.68	32.95	-9.96
14	$C_9H_{12}N_4O_4S$	4.12	37.13	-19.11
15	$C_9H_{10}N_2O_5S$	2.78	35.76	0.73
16	$C_{12}H_{14}N_4O_3$	3.09	38.38	-10.22
17	$C_{11}H_{14}N_2O_3S$	3.97	40.91	-6.10
18	$C_{11}H_{17}N_3O_5$	5.89	64.52	-7.42
19	$C_{13}H_{21}N_3O_5$	7.06	72.35	-2.84
20	$C_{15}H_{25}N_3O_5$	8.3	80.17	7.046
21	$C_{13}H_{21}N_3O_5$	8.93	73.11	-19.78
22	$C_{19}H_{25}N_3O_6$	8.7	74.85	-31.09
23	$C_{15}H_{21}N_3O_5$	8.7	82.04	-25.00
24	$C_{13}H_{21}N_5O_5$	6.6	73.63	-25.00