



MEMOIRE DE FIN D'ETUDE
Pour l'obtention du diplôme Master
En Génie des procédés
Option: Génie Pharmaceutique

Présenté par : **ABDELOUAHABI Samira**

ZOUIRI Manel

**Développement de modèles QSAR pour la
prédiction de l'activité anti-VIH des molécules
Pyranone**

Soutenu le : 27/06/2022

Devant le jury :

Mr. ARDJANI Taki eddine	Maître de Conférences A	Université de Saida	Président
Mme. TOUHAMI Moufida	Maître de Conférences B	Université de Saida	Encadrante
Mme. YAHIA CHERIF Fatima	Maître de Conférences A	Université de Saida	Examinatrice
Mme. BELARBI Ouassyla	Maître Assistante A	Université de Saida	Invitée

Promotion : 2021-2022

« تطوير نماذج QSAR للتنبؤ بالنشاط المضاد لفيروس نقص المناعة البشرية لجزيئات Pyranone »

الملخص

يتم استخدام العديد من تقنيات *in silico* في عملية تصميم الدواء وتطويره. إحدى هذه التقنيات هي العلاقة الكمية بين البنية والنشاط (QSAR).

بروتياز فيروس نقص المناعة البشرية هو واحد من الأهداف الرئيسية لتطوير علاجات كيميائية جديدة. لذا إن تثبيط الإنزيم البروتيني لفيروس نقص المناعة البشرية هو واحد من بين العلاجات المتوفرة لمحاربة المرض الذي يسببه هذا الفيروس الذي يتمثل دوره في منع تطور هذا الأخير في الجسم والمحافظة على التوازن بين وجوده في الجسم ونظام دفاعه. لذلك فإن تطوير مثبطات جديدة أقل سمية وأكثر احتمالا وملائمة ونشاطا ضد الفيروسات المقاومة للأدوية أمر مرغوب فيه للغاية.

في الواقع أجريت دراسة نوعية للعلاقة بين البنية والنشاط من أجل سلسلة نشطة بيولوجيا من مشتقات pyranone التي تظهر أنشطة قوية وانتقائية لتثبيط الأنزيم البروتيني لفيروس نقص المناعة البشرية.

قمنا بدراسة العلاقة الكمية بنية-فعالية البيولوجية (QSAR) لواحد أو أربعون مركب من مشتقات pyranone لهم فعالية بيولوجية (مثبط الأنزيم البروتيني لفيروس نقص المناعة البشرية) و قد استخدمنا الطريقة الإحصائية MLR و PLS و ذلك لتصميم نموذج رياضي QSAR لغرض التنبؤ بالقيم النظرية للفعالية البيولوجية عن طريق هذا النموذج.

ولتأكد من صحة وفعالية هذا النموذج استخدمنا طريقة LOO و LMO. و لقد لاحظنا شدة التقارب بين القيم النظرية والتجريبية للفعالية البيولوجية مما يؤكد فعالية و جودة النموذج.

كلمات مفتاحية: نشاط مكافحة فيروس نقص المناعة البشرية، الأنزيم البروتيني، pyranone، 2D/3D-QSAR، MLR و PLS.

« Développement de modèles QSAR pour la prédiction de l'activité anti-VIH des molécules Pyranone »

Résumé

Plusieurs techniques *in silico* sont utilisées dans le processus de la conception et du développement de médicaments. Une de ces techniques est la relation quantitative structure-activité (QSAR).

La protéase du VIH est l'une des principales cibles virales pour le développement de nouvelles chimiothérapies. Donc l'inhibition de protéase du VIH est l'un des différents traitements disponibles pour lutter contre la maladie causée par ce virus dont le rôle est de bloquer le développement de ce dernier dans le corps et de maintenir l'équilibre entre sa présence dans le corps et son système de défense. Par conséquent, le développement de nouveaux inhibiteurs moins toxiques, plus tolérables, pratiques et actifs contre les virus résistants aux médicaments est hautement souhaitable.

En effet, une étude quantitative de la relation structure-activité (QSAR) a été effectuée également pour une série bioactive de dérivés de pyranone présentant des activités d'inhibition de protéase du VIH puissantes et sélectives.

Une étude QSAR a été effectuée sur quarante un molécules de pyranone. La régression linéaire multiple (MLR) a été utilisée pour quantifier les relations entre les descripteurs moléculaires et la propriété de l'activité inhibitrice de dérivés de pyranone. La prédiction des modèles obtenus a été confirmée par la méthode de validation croisée LOO et LMO. Une forte corrélation a été observée entre les valeurs expérimentales et les valeurs prédites de l'activité inhibitrice de ces dérivés, ce qui indique la validité et la qualité des modèles QSARs obtenus.

Mots clés : Activité anti-VIH, protéase, pyranone, 2D/3D-QSAR, MLR et PLS.

« Development of QSAR models for the prediction of the anti-HIV activity of Pyranone molecules »

Abstract

Several in-silico techniques are utilized in the process of drug design and development. One such technique is quantitative structure-activity relationship (QSAR).

HIV protease is one of the major viral targets for the development of new chemotherapeutics. So HIV protease inhibition is one of the different treatments available to fight against the disease caused by this virus, the role of which is to block the development of the latter in the body and to maintain the balance between its presence in the body and its defense system.

Therefore, the development of new inhibitors that are less toxic, more tolerable, convenient and active against drug resistant viruses is highly desirable.

Indeed, a qualitative structure-activity relationship (QSAR) study was also performed for a bioactive series of pyranone derivatives exhibiting potent and selective HIV protease inhibition activities.

QSAR studies have been performed on forty-one molecules of pyranone. Multiple linear regression (MLR) analysis was performed to derive QSAR models which were further evaluated internally for the prediction of activity. The developed models were cross-validated by the leave one out 'LOO' and leave many out "LMO" technique as well as by the calculation of statistical parameters. High correlation between experimental and predicted activities values was observed, indicating the validation and the good quality of the derived QSARs models.

Keywords: Anti-HIV activity, pyranone, 2D-3D QSAR, MLR et PLS.

Table de Matières

Table des Matières

Liste des figures	3
Liste des tableaux	4
Abréviations	5
Introduction générale	7

Chapitre. I : Protéase et approches thérapeutiques du VIH

I. Introduction	10
II. Le virus de l'immunodéficience humaine (VIH)	10
II.1. Cycle de réplication	12
II.2. Thérapeutique	13
II.2.1. Les vaccins	13
II.2.2. La chimiothérapie	14
III. Les inhibiteurs de la protéase VIH	16
III.1. La protéase	16
III.2. Classes thérapeutiques de la protéase du VIH	17
III.2.1. Les inhibiteurs non-peptidique de la protéase	18
III.2.2. Les inhibiteurs peptidique de la protéase	20
IV. La trithérapie	21
V. Synthèse de molécules anti-VIH à base de 6-alkyl-5,6-dihydropyran-2-one	22
V.1. Synthèse des dérivés de pyranone	24
VI. Conclusion	27
Bibliographie	28

Chapitre. II : les méthodes in silico : les études QSARs

I. Introduction	31
II. Les relations quantitatives structures activités	31
II.1. Définition	31
III. Méthodologie QSAR	33
III.1. Qualité des données	33
III.2. Les descripteurs	34
III.2.1. Définition	34
III.2.2. Types de descripteurs	34
III.2.2.1. Les descripteur 1D	34
III.2.2.2. Les descripteur 2D	34
III.2.2.3. Les descripteur 3D	35
III.3. Méthodes statistiques	36
III.3.1. Régression linéaire	36
III.3.1.1. Régression linéaire simple (RLS)	36
III.3.1.2. Régression linéaire multiple (RLM)	36
III.3.2. Régression aux moindres carrés partiels (PLS)	37

Table de Matières

III.3.3. Réseaux neurones artificiels (RNA)	37
III.3.4. Les machines à vecteurs de support	37
III.4. Validation du modèle	38
III.4.1. Validation interne	38
III.4.1.1. La procédure Leave-One-Out (LOO)	38
III.4.1.2. La procédure Leave-Many-Out (LMO)	38
III.4.2. Validation externe	38
III.4.3. Coefficients et tests statistiques standards	39
III.5. Domaine d'applicabilité	40
IV. Conclusion	41
Bibliographie	42

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

I. Introduction	45
II. Méthodologie	45
II.1. Bases de données	45
II.2. Modèles QSARs	47
II.2.1. Les modèles 2D-QSAR	47
II.2.1.1. Calcul des descripteurs	47
II.2.1.2. Méthode statistique	48
II.2.1.3. Validation du modèle	55
II.2.2. Les modèles 3D-QSAR	55
II.2.2.1. Base de données et préparation des structures	56
II.2.2.2. L'alignement des molécules	56
II.2.2.3. Construction et validation du modèle	57
II.2.2.4. Développement du modèle	66
II.2.2.5. Validation	66
III. Conclusion	67
Bibliographie	68
Conclusion générale	70

Liste des figures

Liste des figures

Chapitre. I

Figure. 1 : Le virus de l'immunodéficience humaine	11
Figure. 2 : Le cycle de réplication du VIH	12
Figure. 3 : Quelques lieux d'actions de traitements contre le virus du SIDA	14
Figure. 4 : Structure de la protéase du VIH	16
Figure. 5 : Différentes classes d'inhibiteurs de la protéase du VIH	17
Figure. 6 : Inhibiteurs non-peptidiques de la protéase du VIH	18
Figure. 7 : Quelques inhibiteurs de la protéase du VIH approuvés par la FDA	19
Figure. 8 : Fixation du Tipranavir (5) sur la protéase du VIH	19
Figure. 9 : Inhibiteurs peptidiques de la protéase du VIH	20
Figure.10 : Structure générale de la Zidovudine (AZT)	21
Figure.11 : Structure du composé II occupant les 4 poches de l'enzyme	22
Figure.12 : Structure générale du composé II substituée	24

Chapitre. II

Figure. 1 : Les étapes de base de la construction du modèle de QSAR	32
Figure. 2 : Schéma représentatif d'un réseau de neurones artificiels	37

Chapitre. III

Figure. 1: Structure générale de Pyranone de notre travail	45
Figure. 2: La corrélation linéaire entre les valeurs observées et les valeurs calculées pour la 2D-QSAR	49
Figure.3 : L'ensemble de molécules alignées à l'aide de : (A) alignement de ligand flexible (B) alignement d'échafaudage commun	56
Figure. 4 : Les distributions gaussiennes : (A) électrostatique, (B)stérique, (C) hydrophobique, (D) accepteur de liaison H et (E) donneur de liaison H	58
Figure. 5 : La corrélation linéaire entre les valeurs observées et les valeurs calculées pour la 3D-QSAR	59

Liste des tableaux

Liste des tableaux

Chapitre. I

Tableau.1 : Liste des antirétroviraux actuellement disponibles	15
Tableau.2 : les structures chimiques ainsi que les concentrations efficaces des composés étudiés	24

Chapitre. III

Tableau.1 : Les structures chimiques ainsi que les données d'activité observées des composés.	46
Tableau. 2 : Les valeurs des activités observées, les valeurs calculées et le résiduel pour la 2D-QSAR.	49
Tableau. 3 : Les activités observées, les valeurs calculées et le résiduel pour la 3D-QSAR.	60

Abréviations

Abréviations

2D : Deux dimensions

3D : Trois dimensions.

ADN : L'acide DésoxyriboNucléique

ARN : Acide RiboNucléique

ARV : AIDS-associatedretrovirus.

AZT : Azidothymidine

CC : Cytotoxicité.

CCR5 : Le récepteur à C-C chimiokine de type 5

CD4: Cluster de Différenciation 4

CI₅₀ : Concentration inhibitrice.

CoMFA: Analyse Comparative de Champs Moléculaires

CoMSIA: Analyse Comparative d'indice de Similarité Moléculaire

CV : Validation croisée

DA : Domaine d'applicabilité

EC₅₀ : Concentration médiane effective

FDA : Food and Drug Administration

HBA : Accepteur de liaison hydrogène

HBD : Donneur de liaison hydrogène

HTLV-111 : Human T-ce/1/eukemia virus.

IC₅₀ : Concentration médiane inhibitrice

INNPs :Inhibiteurs non-peptidique de la Protéase

INNTI : Inhibiteurs non nucléosidiques de la transcriptase inverse

IP : Inhibiteurs de protéase

IT : Indice topologique.

LAV : Lymphadenopathy-associated virus

LMO : Leave- many-out

LOO : Leave-one-out

MST : Maladies sexuelles transmissibles

Abréviations

PLS : Moindres carrés partiels

PM : Poids moléculaires

QSAR : Relation Quantitative de Structure-Activité

QSPR : Relation Quantitative de Structure-Propriété

R&D : Recherche et Développement.

R^2_{adj} : Coefficient de corrélation ajusté

R^2 : Coefficient de corrélation

RLM : Régression linéaire multiple

RNA : Réseaux de neurones artificiels

SAR : Relation structure activité

SIDA : Syndrome d'immunodéficience acquise

SRM : Minimisation de risque structurel

SSPE : Somme des erreurs de prédiction des carrés

SVM : Machine à vecteurs de supports

T CD4⁺ : Taux de lymphocytes

TI : Transcriptase inverse

VIH : Virus de l'immunodéficience humaine

MLR : régression linéaire multiple

T : lymphocytes

DMF : Le N, N-Di-MéthylFormamide

TSET : Training Set

PSET : Prédiction Set

DS : déviation standard

F : Fisher

E state : état électrotopologique

PEOE : égalisation partielle de l'électronégativité orbitalaire

ALOGP : coefficient de partage octanol-eau de Ghose-Crippen-Viswanadhan

Q^2 : Coefficient de corrélation croisée

RMSE : Erreur quadrique moyenne

OPLS : Optimized Potentials for Liquid Simulations

Introduction générale

Introduction Générale

Le développement des médicaments est un processus lent et laborieux, exigeant un grand investissement. Une firme pharmaceutique aura besoin de huit à douze ans pour produire un médicament. Toute seule, la phase liée à la découverte « discovery phase » consomme une partie significative des ressources de la compagnie (pour effectuer la synthèse et les tests biochimiques) et celle-ci dure trois à cinq ans [1]. Les dépenses pour la recherche et développement (R&D) étaient estimées entre 500 et 600 millions de Dollars dans les années quatre-vingt pour dépasser 2.56 milliards de dollars US de nos jours [2], dont deux-tiers du coût sont dépensées pour la découverte des précurseurs potentielles (leads compound) qui ne surpassent malheureusement pas l'étape des tests précliniques [3].

L'industrie pharmaceutique est emmené à réduire le temps, le coût et de produire de nouveaux produits ayant les propriétés thérapeutiques optimums et sans, si possible, d'effets secondaires indésirables, de la manière la plus efficace et la moins coûteuse. Pour cette raison, l'industrie pharmaceutique continue sans cesse à chercher de nouveaux outils et des technologies modernes afin d'y parvenir [4].

Aujourd'hui, après plusieurs années de développement et d'amélioration, l'ordinateur est devenu un outil indispensable dans les différents modes de vie. Sa technologie de pointe et ses utilisations répandues ont accéléré considérablement son développement. De ce fait, plusieurs nouvelles disciplines sont apparues dans presque tous les domaines scientifiques.

La chimie informatique constitue le grand exemple issu de ce développement où elle est maintenant reconnue comme subdivision de la chimie et soutien principal à la chimie moderne[5]. Spécialement dans la chimie des médicaments, plusieurs méthodes de recherche utilisant l'ordinateur ont été créées. Ces méthodes ont été développées pour aider les chercheurs scientifiques pour identifier ou concevoir des ligands et des précurseurs qui pourraient agir plus favorablement avec un récepteur donné augmentant ainsi le nombre de composés présentant une affinité puissante [6].

Introduction générale

L'objectif principal de cette étude, est l'application d'une méthode de la modélisation moléculaire pour prédire l'activité biologique anti-VIH des molécules Pyranone reliant les activités expérimentales aux structures moléculaires.

Le présent travail comporte trois chapitres : Après une introduction générale, le premier chapitre est scindé en deux parties : dans la première partie, nous allons présenter des généralités sur les virus et plus spécialement le VIH. Dans la deuxième partie, nous allons s'étaler, sur le développement de la molécule 6-alkyl-5,6-dihydropyran-2-one (Pyranone) et ses dérivés.

Dans le deuxième chapitre, nous allons décrire le formalisme de la méthodologie choisie. Nous présenterons la méthodologie de l'étude QSAR, la description de la méthode statistique utilisée et les méthodes de calcul engagées dans notre travail.

Le troisième chapitre englobe les résultats des calculs effectués dans notre étude, il présente des relations structure-activité des 6-alkyl-5,6-dihydropyran-2-one. Ainsi que, le développement des modèles statistiques de l'activité biologique : l'activité anti-VIH.

Finalement, une conclusion générale, dans laquelle nous rappelons les principaux résultats obtenus.

Introduction générale

Bibliographie

- [1]. T. I. Oprea "Chemoinformatics in Drug Discovery" Ed. Wiley-vchVerlag. Allemagne, 2005.
- [2]. Gagnon, M. A. (2015). Pourquoi les médicaments sont-ils si chers? La dérive d'un modèle d'affaires'. Revue Prescrire, (35), 457-461.
- [3]. E. A. Reka ; P. N. Kourounakis "Chemistry and Molecular Aspects of Drug Design and Action" Ed. Taylor & Francis Group, LLC. Etats Unies, 2008.
- [4]. S. Ekins "Computer Applications in Pharmaceutical Research and Development" Edt. John Wiley& Sons, Inc. Etats Unies, 2006.
- [5]. F. Deanda "Development and Application of Software Tools for Computer-Assisted Drug Design" Thèse de doctorat; Août 1999. Université de Texas. Etats Unies.
- [6]. K. Gubernator; H.J. Bohm "Structure-Based Ligand Design" Ed. Wiley-vchVerlag. Allemagne, 1998.

Chapitre. I : Protéase et approches thérapeutiques du VIH

I. Introduction

Les virus sont des parasites intracellulaires obligatoires qui détournent à leurs profits les machineries cellulaires pour la réalisation de leurs cycles de réplication. Le développement de stratégies antivirales nécessite donc la compréhension des interactions du virus avec son hôte. Parmi les virus, on distingue les rétrovirus qui infectent spécifiquement les cellules eucaryotes [1].

Le syndrome d'immunodéficience acquise (SIDA) est une maladie grave qui rend inopérant le système immunitaire de l'individu, le rendant susceptible à toutes sortes d'infections opportunistes habituellement non mortelles mais laisse l'organisme vulnérable à toutes sortes d'infections par des microorganismes opportunistes comme le bacille de la tuberculose, le cytomégalovirus ou le parasite *Pneumocystis carinii* responsable de pneumonies graves. De plus, l'affaiblissement du système immunitaire s'accompagne souvent de sarcome de Kaposi ainsi que de divers lymphomes et d'autres formes de cancer [2].

Depuis les premiers signalements de cas d'immunodéficience au début des années 80, le nombre n'a cessé d'augmenter dangereusement et d'après certaines estimations, pourrait atteindre les 40 millions d'ici l'an 2000 au niveau mondial [3]. A ce jour, la maladie est toujours incurable et le patient atteint du SIDA survit rarement plus de deux ans après l'apparition des symptômes[4].

Cette maladie est causée par le VIH (virus de l'immunodéficience humaine), un virus de la famille Retroviridae. A ce jour, la maladie est toujours incurable mais la recherche pour combattre le virus va bon train[5].

II. Le virus de l'immunodéficience humaine (VIH)

Très rapidement après l'apparition des premiers cas, les milieux scientifiques et médicaux se sont mis à la recherche de l'agent étiologique du SIDA et dès le début, une origine virale fut soupçonnée. En 1983, un premier rétrovirus a été isolé chez un patient atteint du SIDA, il fut nommé LAV (lymphadenopathy-associated virus) [6]. Au même moment, Gallo et collaborateurs ont isolé un virus chez un patient atteint d'immunodéficience qu'ils ont dénommé HTLV-111(human T-cell/1/eukemia virus) d'une part pour son tropisme pour les lymphocytes T et d'autre part par analogie avec d'autres rétrovirus connus [7].

Une autre équipe a isolé un virus dans des conditions similaires et qu'ils ont dénommé ARV (AIDS-associated retrovirus) [8]. Le nom de HIV ou en français de VIH (virus d'immunodéficience humaine) a été adopté officiellement en 1986 une fois que l'on a démontré qu'il s'agissait dans tous les cas du même virus [9]. Au début de l'épidémie, le virus était d'abord localisé chez les populations homosexuelles, mais il s'est rapidement propagé par la suite dans tous les milieux urbains.

Ce mouvement fut accéléré par les transfusions de sang contaminé, les contacts sexuels par les personnes à risque, l'immigration et par la transmission verticale mère-enfant [10].

Chapitre. I : Protéase et approches thérapeutiques du VIH

Deux familles de VIH ont été identifiées à ce jour soit le VIH-1 qui est le plus répandu dans les pays industrialisés et le VIH-2 détecté en 1986 et qui se localise principalement en Afrique de l'ouest [11]. Ce dernier n'a que 50% d'homologie dans sa séquence nucléotidique avec le VIH-1 contre environ 80% avec le SIV, virus similaire affectant certains singes comme le Sooty mangabey et le singe vert [12].

Bien que ces deux familles du VIH causent une immunodéficience, les pathologies causées par le VIH-2 sont moins nombreuses et de moindre intensité en plus de s'accompagner d'une virémie plus faible [13].

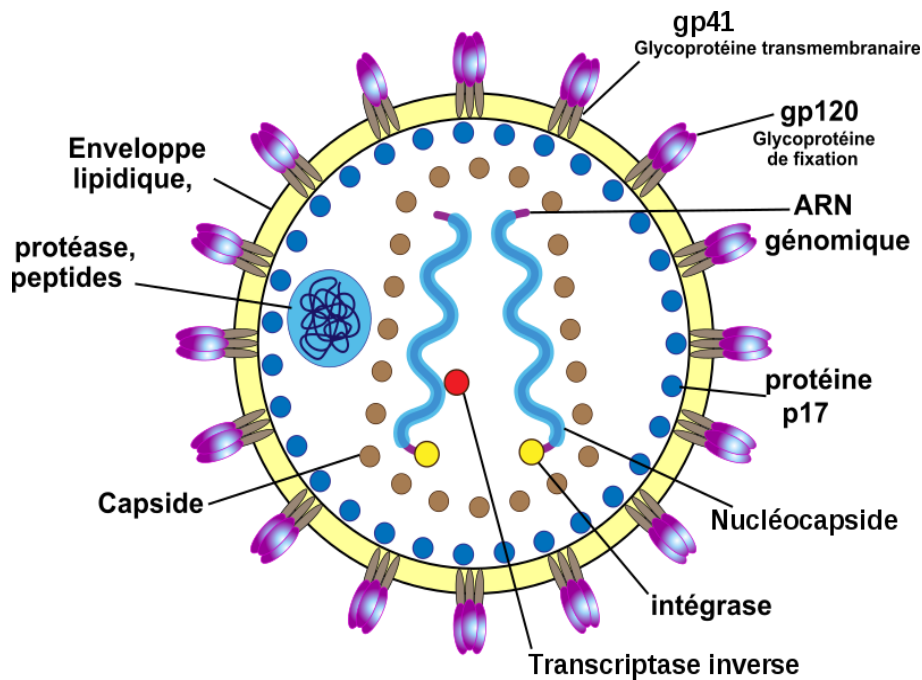


Figure. 1 : Le virus de l'immunodéficience humaine.

Chapitre. I : Protéase et approches thérapeutiques du VIH

II.1. Cycle de réplication

La réplication virale se caractérise par le fait que le virus détourne la machinerie cellulaire à son profit. En effet, les virus ne sont pas capables de se multiplier par eux-mêmes. La réplication virale consiste en l'introduction du génome viral dans une cellule hôte afin de fabriquer de nouveaux virus selon un procédé de biosynthèse.

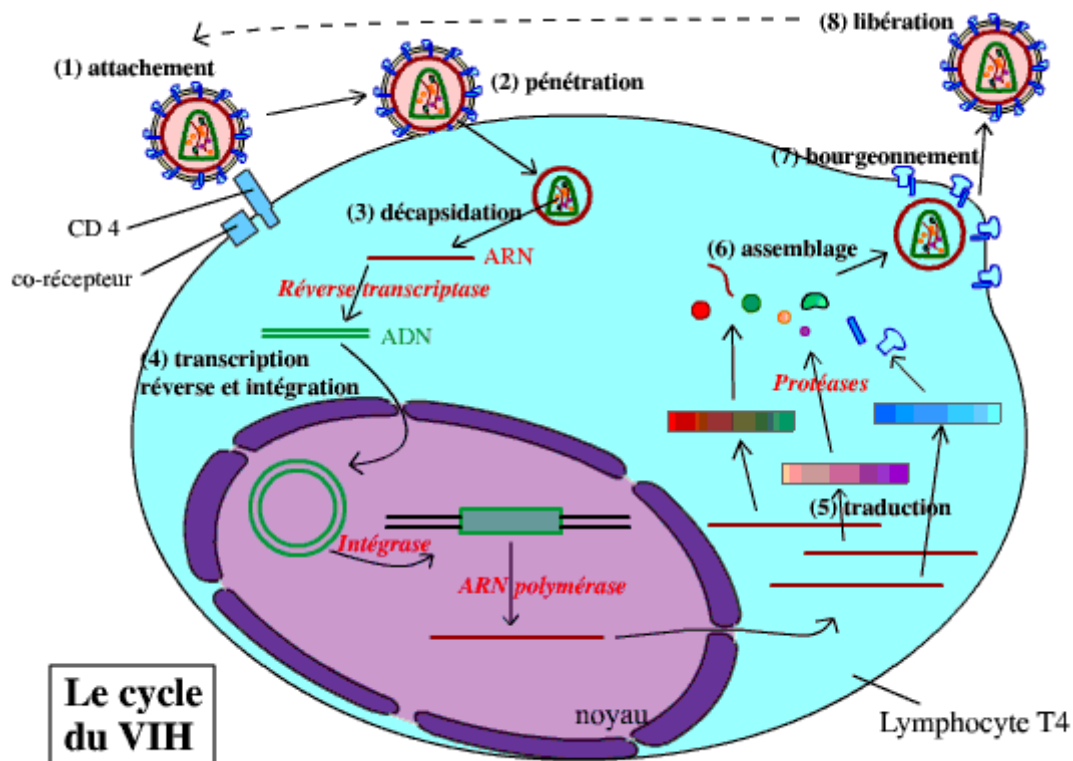


Figure. 2 : Le cycle de réplication du VIH

Le cycle de réplication du VIH comporte plusieurs phases successives :

- a. L'attachement** : La protéine gp120 du VIH se fixe tout d'abord sur le récepteur CD4 qui est présent sur la membrane plasmique de la cellule cible. La protéine gp120 va alors changer de conformation pour s'attacher à un second récepteur nommé corécepteur. Il en existe deux types : CCR-5 et CXCR4.
- b. La fusion** : Fusion entre les membranes du virus et celle de la cellule créant une discontinuité. La capsid du VIH pénètre dans la cellule et se désagrège en libérant les deux brins d'ARN ainsi que les enzymes.
- c. La transcription inverse** : C'est une étape très caractéristique des rétrovirus. La transcriptase inverse (TI), qui est une enzyme virale, transcrit le brin d'ARN viral en ADN et le duplique afin d'aboutir à un ADN proviral double-brin.

Chapitre. I : Protéase et approches thérapeutiques du VIH

- d. L'intégration : L'enzyme nommée intégrase va intégrer le double brin d'ARN proviral à l'ADN cellulaire qui va pouvoir ainsi porter une multitude de séquences codantes pour des protéines composant le VIH.
- e. L'assemblage : L'enzyme nommée protéase va découper les chaînes produites issues de la traduction des ARN messagers afin de former les différentes protéines qui composent le virus du VIH. L'assemblage d'une structure globulaire est permis grâce à des interactions.
- f. Le bourgeonnement : La structure globulaire sort de la cellule infectée et emporte avec elle un morceau de l'enveloppe virale.
- g. La maturation Les particules sorties suite au bourgeonnement sont encore immatures. L'étape de maturation donne lieu à la formation de la capsidie et du noyau. Cette étape rend les virions capables d'infecter d'autres cellules [14].

II.2. La thérapeutique

Au niveau individuel, on peut réduire le risque d'infection par le VIH en limitant l'exposition aux facteurs de risque. Parmi les principales démarches de prévention du VIH, souvent associées entre elles, figurent :

- L'utilisation du préservatif masculin ou féminin ;
- Le dépistage et conseil pour le VIH et les maladies sexuelles transmissibles MST ;
- La circoncision médicale volontaire de l'homme ;
- La réduction des effets nocifs pour les consommateurs de drogues par injection et l'élimination de la transmission mère-enfant du VIH [15].

II.2.1. Les vaccins

La vaccination est un moyen de prévention important pour éviter la propagation d'un virus dans la population. Dans le cas du VIH, la mise au point d'un vaccin efficace exige le développement de nouvelles stratégies d'approche.

En effet, contrairement à d'autres vaccins connus et mis au point contre certains virus comme l'influenza, le VIH ne peut être utilisé sous sa forme atténuée. On ne peut pas non plus garantir l'inactivation complète de tous les virions dans une préparation, ce qui complique l'élaboration d'un vaccin de type inactivé. En outre, le virus est capable d'infecter certaines cellules du système nerveux central difficilement accessibles par les mécanismes de la défense immunitaire, en plus de s'intégrer dans le génome cellulaire des cellules infectées.

Finalement, la très grande variabilité du virus rend très complexe la mise au point d'un vaccin efficace contre tous les variantes possibles [16].

Chapitre. I : Protéase et approches thérapeutiques du VIH

II.2.2. La chimiothérapie

Chacune de ces étapes du cycle de réplication rétroviral procure une cible de choix pour l'élaboration d'agents antirétroviraux. Cependant, la majorité de ces agents sont présentement dirigés surtout contre la transcriptase inverse et la protéase du virus[17].

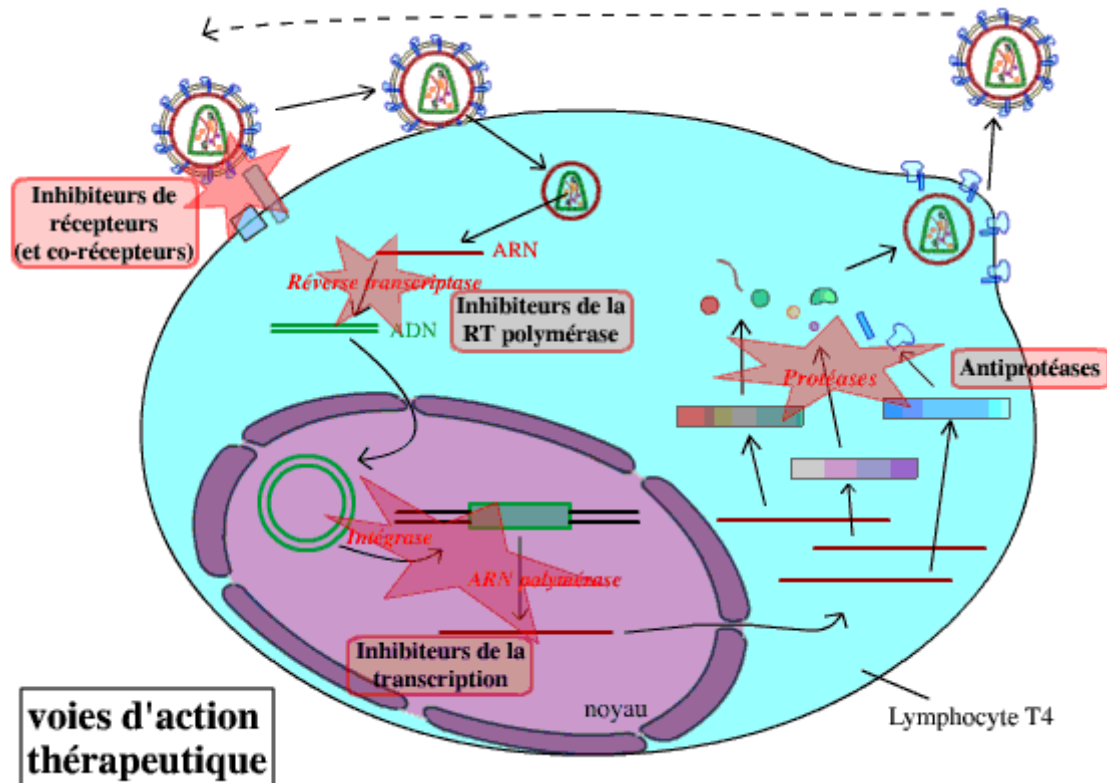


Figure. 3 : Quelques lieux d'actions de traitements contre le virus du SIDA

Il existe actuellement plusieurs classes de médicaments anti-VIH avec des mécanismes d'actions différentes[18] :

- Les inhibiteurs nucléosidiques et nucléotidiques de la transcriptase inverse;
- Les inhibiteurs non nucléosidiques de la transcriptase inverse (INNTI);
- Les inhibiteurs de protéase (IP);
- Les inhibiteurs de fusion ;
- Les inhibiteurs d'intégrase et les antagonistes du récepteur du CCR5.

Chapitre. I : Protéase et approches thérapeutiques du VIH

Tableau.1 : Liste des antirétroviraux actuellement disponibles [19]

<i>Dénomination Commune Internationale (DCI)</i>	<i>Abréviation</i>	<i>Nom commercial</i>
<i>Inhibiteurs nucléosidiques de la transcriptase inverse</i>		
Emtricitabine	FTC	EMTRIVA®
Lamivudine	3TC	EPIVIR®
Zidovudine	AZT	RETROVIR®
Didanosine	ddI	VIDEX®
Ténofovir	TDF	VUREAD®
Stavudine	d4T	ZERIT®
Abacavir	ABC	ZIAGEN®
<i>Inhibiteurs non nucléosidiques de la transcriptase inverse</i>		
Etravirine	ETR	INTELENCE®
Efavirenz	EFV	SUSTIVA®
Névirapine	NVP	VIRAMUNE®
Rilpivirine	RPV	EDURANT®
<i>Associations d'inhibiteurs de la transcriptase inverse</i>		
Lamivudine + Zidovudine	3TC+AZT	COMBIVIR®
Lamivudine+ Zidovudine + Abacavir	3TC+AZT+ABC	TRIZIVIR®
Emtricitabine+ Ténofovir	FTC+TDF	TRUVADA®
Emtricitabine + Ténofovir + Efavirenz	FTC+TDF+EFV	ATRIPLA®
Abacavir + Lamivudine	ABC+3TC	KIVEXA®
Emtricitabine + Ténofovir + Névirapine	FTC+RPV+TDF	EVIPLERA®
<i>Inhibiteurs de la Protéase</i>		
Tipranavir	TPV	APTIVUS®
Indinavir	IDV	CRIXIVAN®
Saquinavir	SQV	INVIRAS®
Lopinavir	LPV	KALETRA®
Darunavir	DRV	PREZISTA®
Atazanavir	ATV	REYATAZ®
Fosamprénavir	FPV	TELZIR®
Ritonavir(booster)	RTV	NORVIR®
<i>Inhibiteurs de l'intégrase</i>		
Raltégravir	RAL	ISENTRESS®
Elvitégravir	EVG	JTK-303®
Dolutegravir	DTG	TIVICAY®
<i>Inhibiteurs de récepteur CCR5</i>		
Maraviroc	MVC	CELESENTRI®
<i>Inhibiteurs de fusion</i>		
Enfuvirtide	T-20	FUSEON®

Chapitre. I : Protéase et approches thérapeutiques du VIH

III. Les inhibiteurs de la protéase du VIH

III.1. La protéase

La protéase des rétrovirus est une enzyme constituée de deux chaînes protidiques identiques qui se combinent pour former un seul site actif. Au cours de sa réplication, le VIH produit des chaînes polypeptidiques longues, dérivées essentiellement des gènes gag et pol.

La protéase virale permet de scinder ces protéines précurseurs pour former les différentes protéines de structure ou de régulation incluant la transcriptase inverse, l'ARNase et l'intégrase [20].

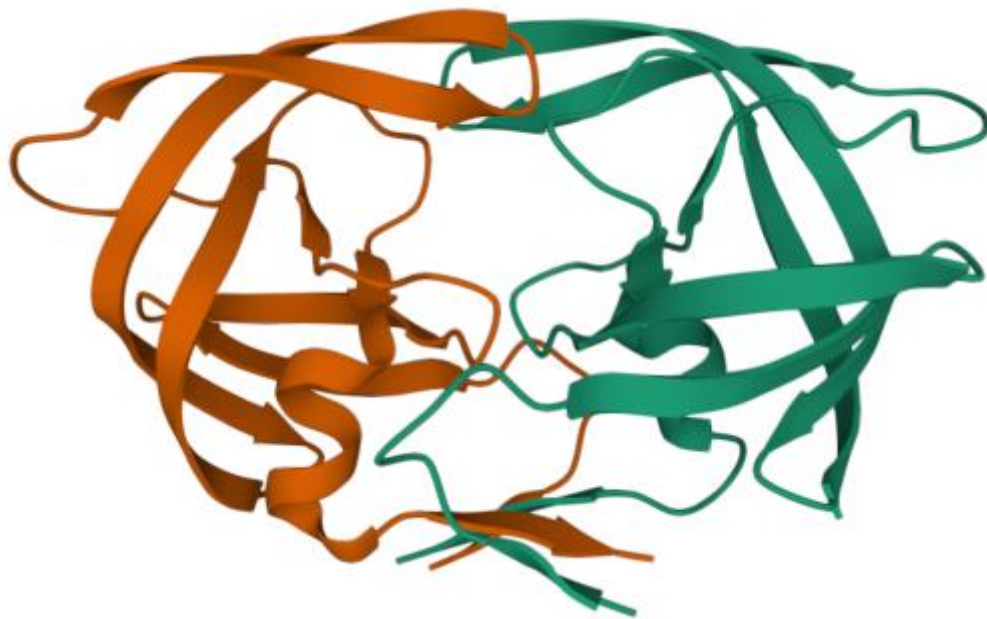


Figure.4 : Structure de la protéase du VIH [PDB : 2PC0]

En cas d'inhibition à ce stade, les particules virales qui sont formées sont désorganisées sur le plan structurel et fonctionnel et donc non infectieuses. La protéase du VIH est un membre de la famille des protéases de type aspartique. Elle fonctionne comme un homodimère, chacun des monomères comprenant un résidu de l'acide aspartique qui contribue au site enzymatique actif. La protéase du VIH a été considérée comme une cible potentielle pour un traitement antiviral à partir de 1986, lorsqu'il a été montré qu'une mutation au niveau du gène pol de la région des protéases prévenait la maturation du précurseur de la polyprotéine gag.

Plusieurs inhibiteurs ont été alors conçus avec une approche rationnelle de type "design moléculaire", en bénéficiant d'une recherche active et des connaissances acquises depuis une dizaine d'années sur d'autres inhibiteurs de protéase humaine comme par exemple les inhibiteurs de l'enzyme de conversion [21].

Chapitre. I : Protéase et approches thérapeutiques du VIH

III.2. Classes thérapeutiques de la protéase du VIH

Il existe deux grandes catégories d'inhibiteurs de protéase du VIH:

- Les inhibiteurs peptidiques qui sont basés sur l'inhibition compétitive du substrat de la protéase;
- Les inhibiteurs non peptidiques.

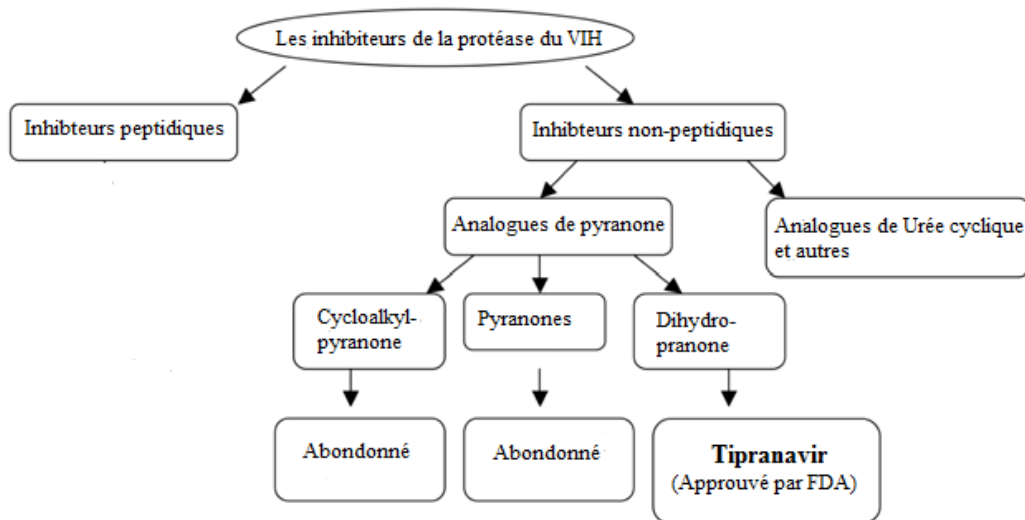


Figure. 5 : Différentes classes d'inhibiteurs de la protéase du VIH.

La première catégorie d'inhibiteurs nécessite des étapes complexes de fabrication (15 à 25 étapes successives). Il s'agit donc de médicaments complètement synthétiques avec des coûts de fabrication importants. Une des grandes caractéristiques de ce type d'inhibiteurs de protéase est leur grande spécificité vis-à-vis de la protéase du VIH. Ils n'interfèrent pratiquement pas avec les protéases cellulaires (qui sont des enzymes extrêmement répandues) et cette grande spécificité laissait augurer une toxicité plus faible qu'avec les analogues nucléosidiques qui interfèrent avec des enzymes cellulaires impliquées dans la synthèse des acides nucléiques [21].

In vitro, les inhibiteurs de protéase sont tous actifs sur le VIH-1 (avec de petites différences pour le VIH-2) à des concentrations nanomolaires alors que les inhibiteurs de la transcriptase inverse sont habituellement actifs à des concentrations micromolaires. A l'inverse des analogues nucléosidiques qui nécessitent plusieurs étapes de phosphorylation avant d'être efficaces, les inhibiteurs de protéase sont directement actifs.

Leur pénétration intracellulaire est bonne, y compris au niveau des macrophages et des cellules dendritiques. Ils agissent in vitro sur tous les sous-types du VIH-1. Ils agissent à un

Chapitre. I : Protéase et approches thérapeutiques du VIH

niveau post-transcriptionnel de la réplication du VIH et peuvent ainsi diminuer la réplication virale au niveau des cellules déjà infectées de façon aiguë et/ou chronique. Ils agissent également sur les macrophages alors que les inhibiteurs de la transcriptase inverse ont des effets variables en fonction des processus de métabolisation [21].

III.2.1. Les inhibiteurs non-peptidique de la protéase

Des recherches intenses ont été dirigées vers le développement des INNPs cycliques pour le traitement de l'infection par le VIH. Ils ont identifié plusieurs structures in silico (Figure.6) [22]. Toutes ces pistes ont été identifiées à l'aide d'informations structurales recueillies à partir de travaux approfondis sur les pistes peptidiques et d'une approche de conception d'analyse pharmacophore computationnelle. Beaucoup d'attention ont été accordées aux urées cycliques et aux dihydropyranones.

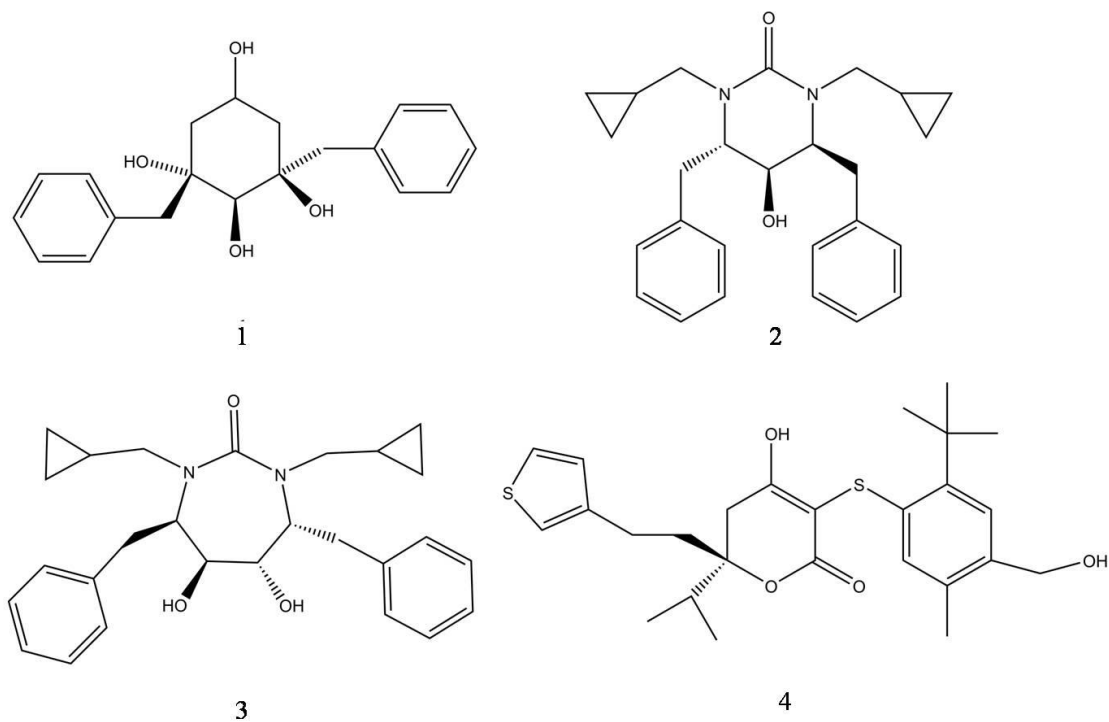


Figure.6: Inhibiteurs non-peptidiques de la protéase du VIH

Chapitre. I : Protéase et approches thérapeutiques du VIH

La recherche sur les dihydropyranones a abouti à l'approbation du Tipranavir (5) pour le traitement du VIH, qui est administré en association avec le Ritonavir (6).

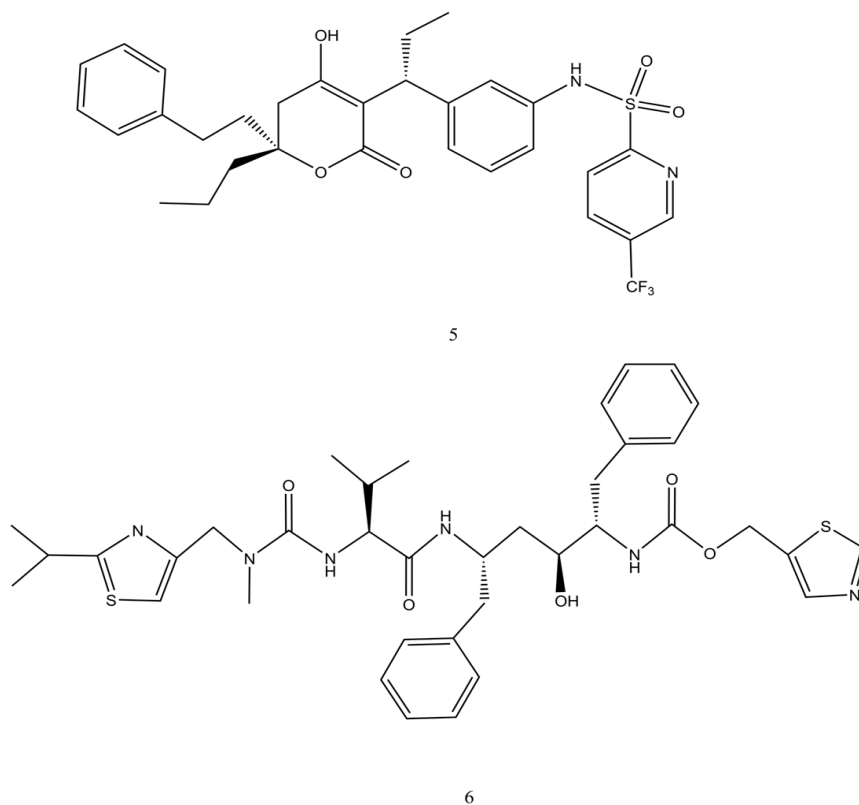


Figure.7 : Quelques inhibiteurs de la protéase du VIH approuvés par la FDA[23, 24].

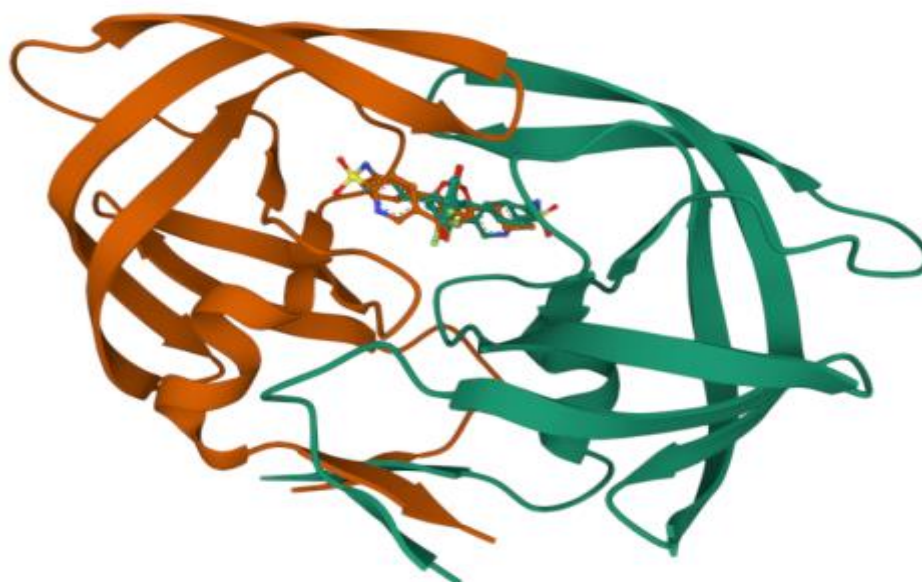


Figure.8 : Fixation du Tipranavir (5) sur la protéase du VIH [PDB : 3SPK]

Chapitre. I : Protéase et approches thérapeutiques du VIH

III.2.2. Les inhibiteurs peptidique de la protéase

Tous les IP approuvés par l'US-FDA, à l'exception du Tipranavir (5), sont des inhibiteurs peptidiques (Figure.9) à base de substrat. Ces inhibiteurs ont été conçus en utilisant le principe « peptidomimétique à l'état de transition », ce qui signifie que dans les inhibiteurs, la liaison peptidique hydrolysable est remplacée par un isostère à l'état de transition non hydrolysable [25]. Un certain nombre de ces isostères ont été étudiés, notamment la statine (7), l'hydroxyéthylène (8), l'hydroxyéthylamine (9), la norstatine (10), l'amide réduit (11), l'hydroxyéthylurée (12), le monoalcool (13), le diol (14) et les aminodiols (15) [25].

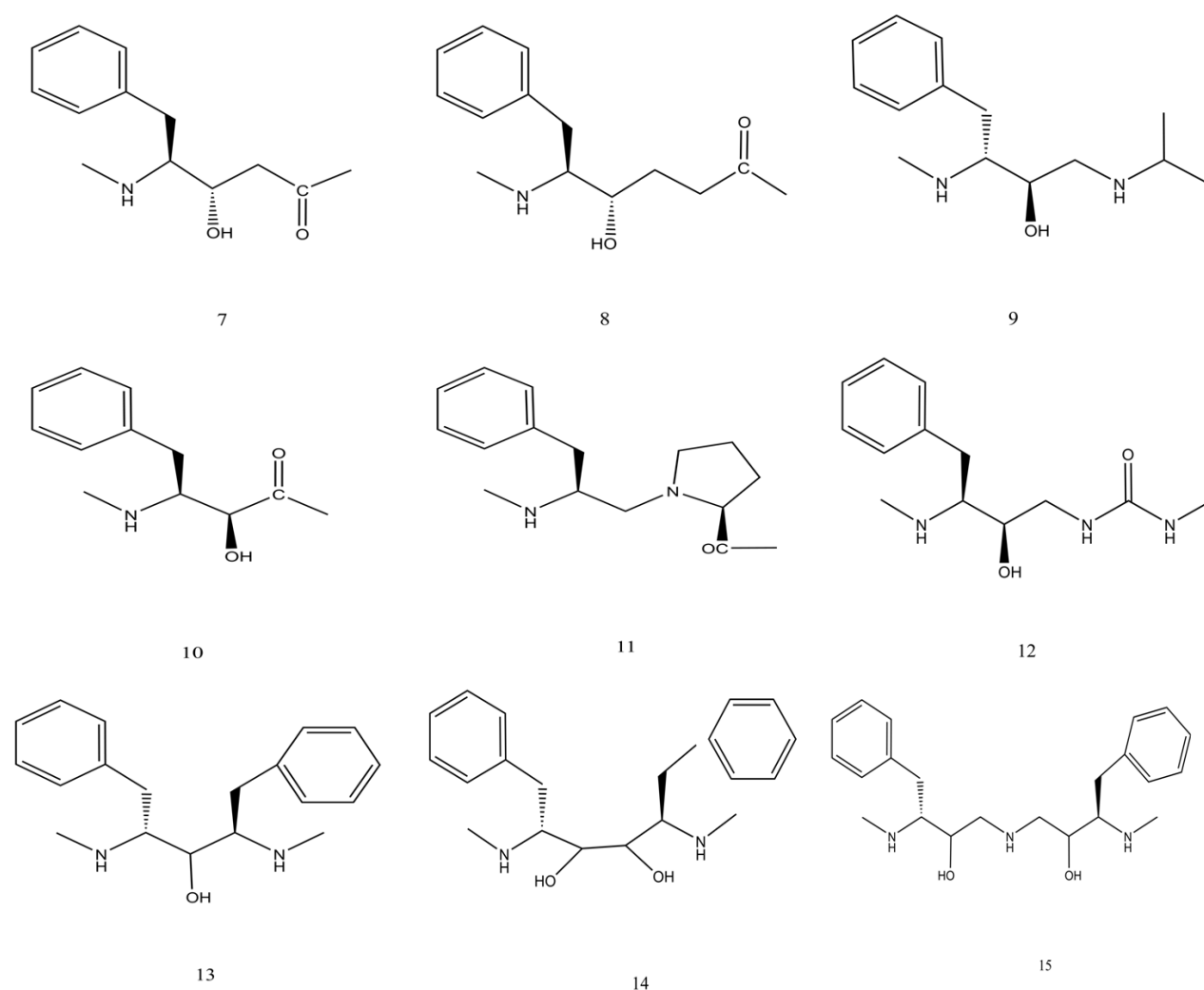


Figure.9 : Inhibiteurs peptidiques de la protéase du VIH

Chapitre. I : Protéase et approches thérapeutiques du VIH

IV. La trithérapie

Depuis le milieu des années 90, avec le développement par l'industrie pharmaceutique de molécules anti-VIH (antirétrovirales) autres que la zidovudine (AZT) [26-27], sont apparues les nouvelles stratégies de traitement de l'infection par le VIH connues sous le nom de trithérapies. Il s'agit d'associations de trois médicaments inhibant la réplication du VIH. Elles permettent de diminuer la quantité de virus circulant dans l'organisme (charge virale en nombre de copies du génome viral/ml de plasma) et surtout de faire remonter le nombre de lymphocytes T auxiliaires (taux de lymphocytes T CD4⁺ en nombre de cellules/ml de plasma) dont la raréfaction traduit l'affaiblissement du système immunitaire par le virus (immunodépression) [28].

Par rapport aux monothérapies utilisant l'AZT seul, ces traitements sont d'une efficacité majeure et surtout durable, responsable de l'amélioration considérable de la survie des personnes infectées par le VIH. En rétablissant indirectement le fonctionnement immunitaire, ils préviennent l'apparition des maladies opportunistes (qui ne se développent que chez les immunodéprimés) caractéristiques du stade avancé de l'infection par le VIH et responsables de sa mortalité (stade Sida) [29].

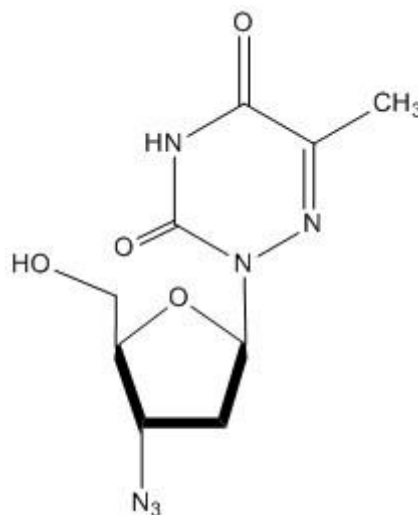


Figure.10 :Structure générale de la Zidovudine (AZT)

Cependant, les trithérapies ne guérissent pas de l'infection par le VIH. Tout arrêt du traitement est suivi à plus ou moins court terme d'une réaugmentation de la charge virale et d'une rediminution des taux de lymphocytes T auxiliaires. De plus, ces médicaments sont responsables d'effets secondaires importants, souvent transitoires, mais aussi à long terme [30].

Enfin, des résistances du virus aux différents médicaments peuvent apparaître au fil du traitement, d'autant plus que celui-ci est mal suivi.

Les trithérapies sont donc des traitements à vie, toujours contraignants et parfois pénibles, susceptibles de perdre leur efficacité.

Chapitre. I : Protéase et approches thérapeutiques du VIH

V. Synthèse de molécules anti-VIH à base de 6-alkyl-5,6-dihydropyran-2-one

Des recherches visant à identifier des inhibiteurs de la protéase du VIH simples, puissants et de faible poids moléculaire qui ne présentent pas de résistance croisée avec les agents existants sont toujours importantes. Des efforts vers la découverte de puissants inhibiteurs non peptidiques ont commencé avec des inhibiteurs, issu d'un criblage de masse d'une collection chimique [31]. D'autres SAR ont conduit à d'autres inhibiteurs, ne possédant pas de centre chiral mais n'occupant que trois poches internes de l'enzyme [32]. L'introduction d'un centre chiral en quaternisant la position 6 du cycle 4-hydroxy-pyran-2-one a conduit à la structure (II), dont les substituants occupent les quatre poches internes de l'enzyme [33].

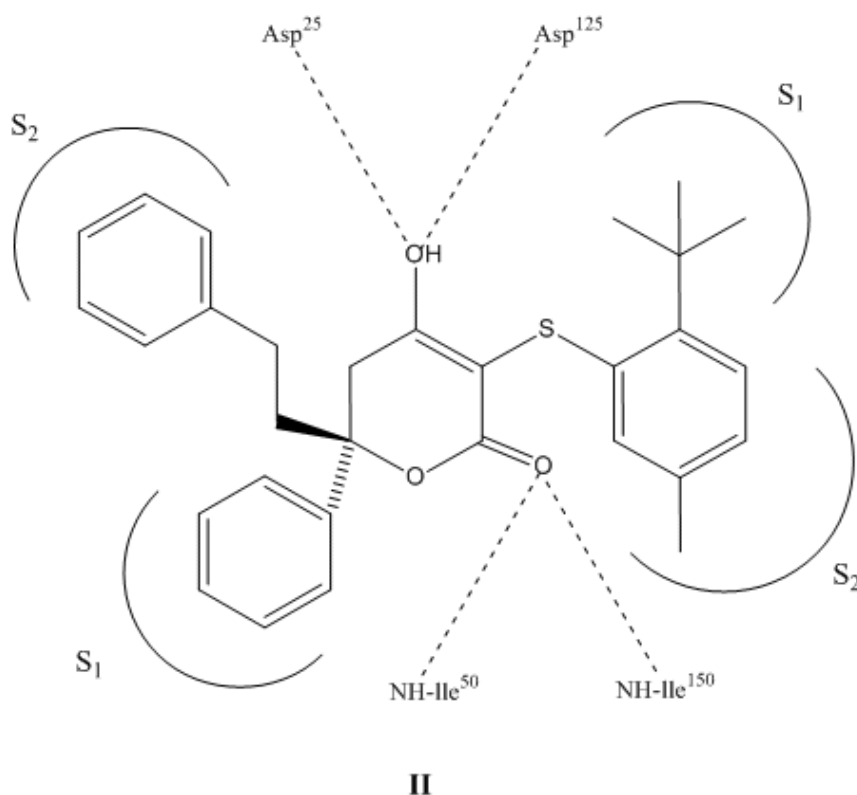


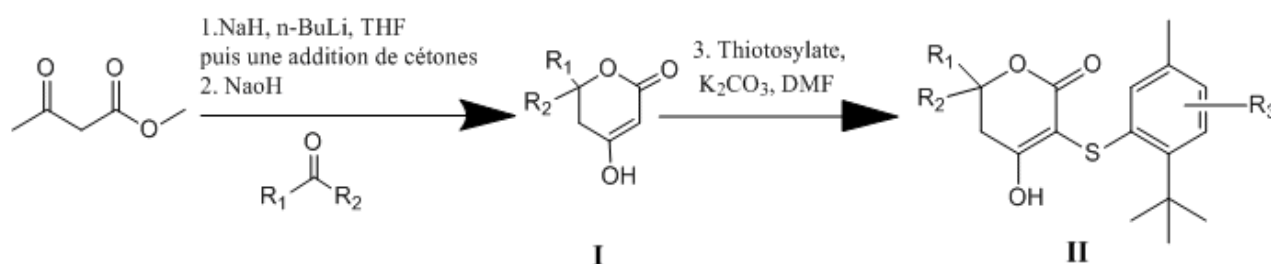
Figure.11 :Structure du composé II occupant les 4 poches de l'enzyme

Les structures cristallines aux rayons X [34-35, 33] des 5,6-dihydropyran-2-ones (II) liées à la protéase du VIH montrent que le groupe hydroxyle énolique forme des interactions de liaison hydrogène avec Asp²⁵ et/ou Asp¹²⁵ des résidus de l'enzyme.

Chapitre. I : Protéase et approches thérapeutiques du VIH

Le composé (II) décrit dans notre étude, a été synthétisé en associant le composé 5,6-dihydropyran-2-one 6,6-di substitué(I) avec le thiotosylate en présence de carbonate de potassium dans du DMF (Schéma.1) [36].

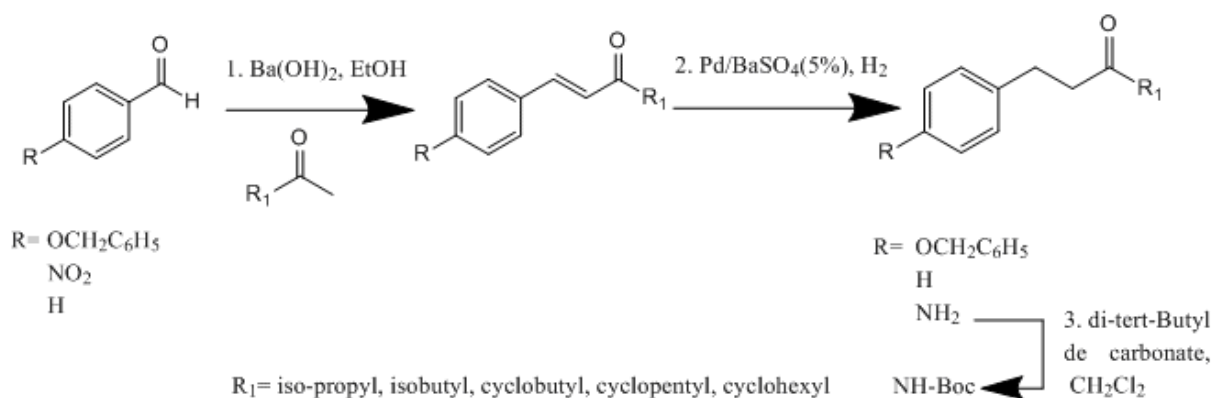
Schéma (1)



Les 5,6-dihydropyran-2-ones 6,6-di substituées (I) ont été synthétisées par condensation du dianion, un dérivé de l'acétoacétate de méthyle avec la cétone, suivie d'une cyclisation en présence d'une base généralement de la soude (Schéma.1).

Les cétones utilisées ci-dessus ont été synthétisées par la réaction de Claisen-Schmidt.

Schéma (2)



La réaction de Claisen-Schmidt (Schéma.2) a été réalisée par la condensation des aldéhydes correspondants avec les méthylcétones convenablement substituées en présence d'alcali à température ambiante. Les cétones insaturées ont été converties en cétones saturées par hydrogénation de la double liaison à l'aide de nickel de Raney ou de Pd/BaSO₄ (5%) (Schéma.2).

Chapitre. I : Protéase et approches thérapeutiques du VIH

V.1. Synthèse des dérivés de pyranone

Pour synthétiser les dérivés de pyranone, Prasad et ses collaborateurs [37] ont mis au point une stratégie de synthèse.

Une série de nouveaux dérivés de pyranone substitués à la fois aux positions (* et **) ont été synthétisés dans le but d'améliorer l'activité spécifique au VIH rapporté sur 6-alkyl-5,6-dihydropyran-2-one (II).

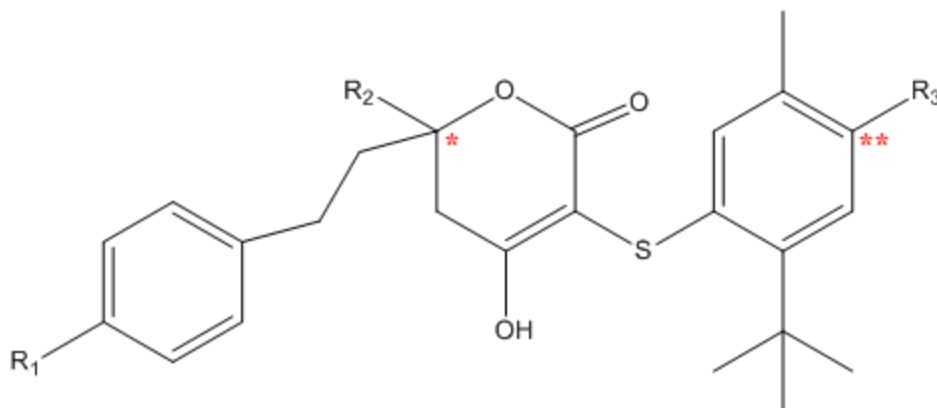


Figure.12 : Structure générale du composé II substitué

Les études se sont concentrées sur le sondage de la poche S₁ de l'enzyme en faisant varier les groupes alkyle de différentes exigences stériques en position * du cycle dihydropyran-2-one. Dans tous ces analogues de dihydropyran-2-one, le groupe 4-hydroxyphényl a été maintenu constant en position * de la matrice dihydropyran-2-one.

Ce substituant s'est avéré optimal pour occuper la poche S₂ de l'enzyme [38]. Le groupe 3-(2-isopropyl-5-méthylphénylthio), qui occupe les poches S₁ ainsi que S₂ de l'enzyme, a été utilisé à la position C-3 de la dihydropyran-2-one. La substitution du groupe 6-phényl (1) par un groupe méthyle a conduit à (2).

De manière surprenante, l'unité de liaison *in vitro* à la protéase du VIH est maintenue. De plus, il y a une augmentation de l'activité antivirale > sept fois (EC₅₀ : 9,4 μM contre > 69 μM pour (Pyranone 2) et (Pyranone 1), respectivement).

Tableau.2 : Les structures chimiques ainsi que les concentrations efficaces des composés étudiés.

Molécule	R ₁	R ₂	R ₃	EC ₅₀ (μM)
Pyranone 1	4-OH	Phenyl	H	69
Pyranone 2	4-OH	Methyl	H	9.4
Pyranone 3	4-OH	n-propyl	H	7.8
Pyranone 4	4-OH	n-Butyl	H	23
Pyranone 5	4-OH	n-Pentyl	H	27

Chapitre. I : Protéase et approches thérapeutiques du VIH

Pyranone 6	4-OH	Isopropyl	H	1.9
Pyranone 7	4-OH	cyclopropyl	H	5
Pyranone 8	4-OH	cyclopentyl	H	5.9
Pyranone 9	4-OH	cyclohexyl	H	4.1
Pyranone 10	4-OH	Phenyl	CH ₂ OH	4.7
Pyranone 11	4-OH	Methyl	CH ₂ OH	2.5
Pyranone 12	4-OH	n-propyl	CH ₂ OH	1
Pyranone 13	4-OH	n-Butyl	CH ₂ OH	0.9
Pyranone 14	4-OH	n-Pentyl	CH ₂ OH	0.6
Pyranone 15	4-OH	Isopropyl	CH ₂ OH	0.6
Pyranone 16	4-OH	Isobutyl	CH ₂ OH	1.5
Pyranone 17	4-OH	cyclopropyl	CH ₂ OH	1.6
Pyranone 18	4-OH	cyclobutyl	CH ₂ OH	0.8
Pyranone 19	4-OH	cyclopentyl	CH ₂ OH	0.6
Pyranone 20	4-OH	cyclohexyl	CH ₂ OH	0.5
Pyranone 21	4-OH	Methyl	OH	7
Pyranone 22	4-OH	Isopropyl	OH	1.6
Pyranone 23	4-OH	cyclohexyl	OH	3.1
Pyranone 24	4-OH	Methyl	OCH ₂ CH ₂ OH	3.4
Pyranone 25	4-OH	Methyl	OCH ₂ CH ₂ OCH ₂ OCH ₃	4.6
Pyranone 26	4-OH	n-propyl	OCH ₂ CH ₂ OH	4.3
Pyranone 27	4-OH	n-Butyl	OCH ₂ CH ₂ OH	1.7
Pyranone 28	4-OH	n-Pentyl	OCH ₂ CH ₂ OH	9
Pyranone 29	4-OH	Isopropyl	OCH ₂ CH ₂ OH	1.5
Pyranone 30	4-OH	cyclopropyl	OCH ₂ CH ₂ OH	1.6
Pyranone 31	4-OH	cyclopentyl	OCH ₂ CH ₂ OH	1.2
Pyranone 32	4-NH ₂	Phenyl	CH ₂ OH	3.7

Chapitre. I : Protéase et approches thérapeutiques du VIH

Pyranone 33	4-NH ₂	Isopropyl	CH ₂ OH	0.5
Pyranone 34	4-NH ₂	Isobutyl	CH ₂ OH	0.8
Pyranone 35	4-NH ₂	cyclohexyl	CH ₂ OH	1.4
Pyranone 36	H	Isopropyl	H	15
Pyranone 37	H	Phenyl	CH ₂ OH	2.5
Pyranone 38	H	Isopropyl	CH ₂ OH	0.6
Pyranone 39	H	Isopropyl	OH	1.8
Pyranone 40	H	Phenyl	OCH ₂ CH ₂ OH	20
Pyranone 41	H	Isopropyl	OCH ₂ CH ₂ OH	1.3

Les composés (Tableau.2) ont été testés pour leurs inhibitions de la protéase du VIH en utilisant le substrat peptidique à pH 6.2. L'enzyme clive la liaison *Leu-Phe* (*p-NO₂*) du substrat. L'EC₅₀ indique la concentration du médicament qui offre une protection de 50 % contre le VIH.

Chapitre. I : Protéase et approches thérapeutiques du VIH

VI. Conclusion

La maladie induite par le VIH peut être gérée au moyen de schémas thérapeutiques reposant sur une association de trois médicaments antirétroviraux ou plus. À l'heure actuelle, le traitement antirétroviral ne guérit pas de l'infection à VIH mais supprime activement la réplication du virus dans l'organisme et permet au système immunitaire de se renforcer et de reconstituer sa capacité à combattre les infections opportunistes et certains cancers.

Bien que plusieurs stratégies de conception de médicaments se soient concentrées sur le blocage des mécanismes structuraux et de réplication de la protéine du VIH, la protéase du VIH demeure une cible cruciale dans la découverte de médicaments contre le VIH. Malgré l'efficacité démontrée par les inhibiteurs de la protéase du VIH, seuls ou en association avec d'autres antirétroviraux, les effets secondaires indésirables, la toxicité et la résistance mutationnelle ont limité leur utilisation clinique [38].

Ceci explique la nécessité de nouveaux antagonistes de la protéine protéase du VIH présentant une affinité de liaison et des propriétés inhibitrices améliorées. A cet effet, les méthodes de simulation computationnelle ont été largement utilisées plus récemment, contrairement aux méthodes expérimentales conventionnelles qui sont laborieuses avec une énorme consommation de temps et de ressources.

Chapitre. I : Protéase et approches thérapeutiques du VIH

Bibliographie

- [1]. Gaston, F. (2008). Développement d'inhibiteurs d'entrée du virus VIH-1 (Doctoral dissertation, Université de Provence-Aix-Marseille I).
- [2]. Rane, D. F., Dasmahapatra, B., & Schwartz, J. (1994). Overview Anti-infectives: Recent advances in anti-HIV agents. *Expert Opinion on Therapeutic Patents*, 4(8), 941-955.
- [3]. Tai-Shun, L., Schinazi, R. F., Zhu, J., Birks, E., Carbone, R., Yikang, S., ... & Prusoff, W. H. (1993). Anti-HIV-1 activity and cellular pharmacology of various analogs of gossypol. *Biochemical pharmacology*, 46(2), 251-255.
- [4]. Nowak, M. A., May, R. M., Phillips, R. E., Rowland-Jones, S., Lalloo, D. G., McAdam, S., ... & McMichael, A. J. (1995). Antigenic oscillations and shifting immune dominance in HIV-1 infections. *Nature*, 375(6532), 606-611.
- [5]. ARNOLD, E., et G.F. ARNOLD. 1991. Human immunodeficiency virus structure: Implications for antiviral design: *Adv. Virus Res.* . . . : 1-87.
- [6]. Barré-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., ... & Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, 220(4599), 868-871.
- [7]. Gallo, R. C., Sarin, P. S., Gelmann, E. P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V. S., ... & Popovic, M. (1983). Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science*, 220(4599), 865-867.
- [8]. Levy, J. A., Hoffman, A. D., Kramer, S. M., Landis, J. A., Shimabukuro, J. M., & Oshiro, L. S. (1984). Isolation of lymphocytotropic retroviruses from San Francisco patients with AIDS. *Science*, 225(4664), 840-842.
- [9]. Coffin, J., Haase, A., Levy, J. A., Montagnier, L., Oroszlan, S., Teich, N., ... & Weiss, R. (1986). Human immune deficiency viruses. *Science*, 232(4751), 697-697.
- [10]. Hirsch, M. S., & Curran, J. (1990). Human immune deficiency viruses. *Biology and medical aspects*. Virology. Raven Press, New York, NY, 1545-1570.
- [11]. Clavel, F., Guyader, M., Guétard, D., Sallé, M., Montagnier, L., & Alizon, M. (1986). Molecular cloning and polymorphism of the human immune deficiency virus type 2. *Nature*, 324(6098), 691-695.
- [12]. Jaskolski, M., Tomasselli, A. G., Sawyer, T. K., Staples, D. G., Henrikson, R. L., Schneider, J., ... & Wlodawer, A. (1991). Structure at 2.5-Å resolution of chemically synthesized Human Immunodeficiency Virus Type 1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry*, 30(6), 1600-1609.
- [13]. Martinon, O., & Levy, D. (1993). Le virus de l'immunodéficience féline. *Veterinary Research*, 24, 151-158.
- [14]. CRIPS. La structure du VIH et son cycle de réplication. Disponible sur : <http://www.lecrisp-idf.net> (visité le : le 20/01/2022)
- [15]. Principaux repères sur le VIH/sida – WHO ; <https://www.who.int/fr/news-room/fact-sheets/detail/hiv-aids#> (visité : 30/11/2021)

Chapitre. I : Protéase et approches thérapeutiques du VIH

- [16]. Lévy-Bruhl, D. (1995). Decision regarding vaccination. *Revue D'épidemiologie et de Sante Publique*, 43(1), 5-6.
- [17]. Valcke, H. S. (2010). Étude du dysfonctionnement du compartiment des cellules B chez des patients à différents stades d'infection par le virus d'immunodéficience humaine (VIH).
- [18]. <https://www.vidal.fr/maladies/sexualite-contraception/ist-vih-sida/medicaments.html>
- [19]. Delessert, K., Schiffer, V., & Hirschel, B. (2005). Sida. *Rev Med Suisse*, 1, 73-8.
- [20]. [https://planet-vie.ens.fr/thematiques/microbiologie/virologie/les-trithérapies-antiretrovirales\(visité:12/03/2022\)](https://planet-vie.ens.fr/thematiques/microbiologie/virologie/les-trithérapies-antiretrovirales(visité:12/03/2022))
- [21]. https://www.pistes.fr/transcriptases/47_414.htm (visité: 04/04/2022)
- [22]. Chrusciel, R. A., & Strohbach, J. W. (2004). Non-peptidic HIV protease inhibitors. *Current Topics in Medicinal Chemistry*, 4(10), 1097-1114.
- [23]. Garg, R., & Bhatarai, B. (2006). QSAR and molecular modeling studies of HIV protease inhibitors. *QSAR and Molecular Modeling Studies in Heterocyclic Drugs I*, 181-271.
- [24]. De Clercq, E. (1995). Toward improved anti-HIV chemotherapy: therapeutic strategies for intervention with HIV infections. *Journal of medicinal chemistry*, 38(14), 2491-2517.
- [25]. Yarchoan, R., Weinhold, K., Lyerly, H. K., Gelmann, E., Blum, R., Shearer, G., ... & Markham, P. (1986). Administration of 3'-azido-3'-deoxythymidine, an inhibitor of HTLV-III/LAV replication, to patients with AIDS or AIDS-related complex. *The Lancet*, 327(8481), 575-580
- [26]. Balzarini, J., Baba, M., Pauwels, R., Herdewijn, P., & De Clercq, E. (1988). Anti-retrovirus activity of 3'-fluoro-and 3'-azido-substituted pyrimidine 2', 3'-dideoxynucleoside analogues. *Biochemical pharmacology*, 37(14), 2847-2856
- [27]. Boggetto, N., & Reboud-Ravaux, M. (2001). La protéase du VIH et le développement d'inhibiteurs. *Virologie*, 5(6), 9-18.
- [28]. Moyle, G. (2001). The emerging roles of non-nucleoside reverse transcriptase inhibitors in antiretroviral therapy. *Drugs*, 61(1), 19-26.
- [29]. Holodniy, m.; v. Miller, human immunodeficiency virus, in practical guidelines in antiviral therapy, c.a.b.b.g.j. Galasso, editor. 2002, elsevier science. P. 95-125.
- [30]. Vara Prasad, J. V. N.; Para, K. S.; Lunney, E. A.; Ortwine, D. F.; Dunbar, J. B.; Ferguson, D.; Tummino, P. J.; Hupe, D.; Tait, B. D.; Domagala, J. M.; Humblet, C.; Bhat, T. N.; Liu, B.; Guerin, D. M. A.; Baldwin, E. T.; Erickson, J. W.; Sawyer, T. K. J. *Am. Chem. Soc.* 1994, 116, 6989.
- [31]. Vara Prasad, J. V. N.; Lunney, E. A.; Ferguson, D.; Tummino, P. J.; Rubin, J. R.; Reyner, E. L.; Stewart, B. H.; Guttendorf, R. J.; Domagala, J. M.; Suvorov, L. I.; Gulnik, S. V.; Topol, I. A.; Bhat, T. N.; Erickson, J. W. *J. Am. Chem. Soc.* 1995, 117, 11070 and references therein.
- [32]. Tait, B. D.; Hagen, S.; Domagala, J. M.; Ellsworth, E.; Gajda, C.; Hamilton, H.; Vara Prasad, J. V. N.; Ferguson, D.; Graham, N.; Hupe, D.; Nouhan, C.; Tummino, P. J.; Humblet, C.; Lunney, E. A.; Pavlovsky, A.; Rubin, J.; Baldwin, E. T.; Bhat, T. N.; Erickson, J. W.; Gulnik, S. V.; Liu, B. *J. Med. Chem.* 1997, 40, 3781 and references therein.

Chapitre. I : Protéase et approches thérapeutiques du VIH

- [33]. Tait, B. D.; Hagen, S.; Domagala, J. M.; Ellsworth, E.; Gajda, C.; Hamilton, H.; Vara Prasad, J. V. N.; Ferguson, D.; Graham, N.; Hupe, D.; Nouhan, C.; Tummino, P. J.; Humb-let, C.; Lunney, E. A.; Pavlovsky, A.; Rubin, J.; Baldwin, E.T.; Bhat, T. N.; Erickson, J. W.; Gulnik, S. V.; Liu, B. J. *Med.Chem.* 1997, 40, 3781 and references therein.
- [34]. Romines, K. R.; Chrusciel, R. A. *Current Medicinal Chemistry* 1995, 2, 825 and references therein.
- [35]. Vara Prasad, J. V. N.; Lunney, E.; Para, K. S.; Tummino, P. J.; Ferguson, D.; Hupe, D.; Domagala, J. M.; Erickson, J. W. *Drug Design and Discovery* 1996, 13, 15 and references therein. nisterra, J. V.; Garcia-Raso, A.; Cabello, J. A.; Marinas, M. *Synthesis* 1984, 502.
- [36]. Prasad, J. V., Boyer, F. E., Domagala, J. M., Ellsworth, E. L., Gajda, C., Hamilton, H. W., ... & Erickson, J. W. (1999). Nonpeptidic HIV protease inhibitors possessing excellent antiviral activities and therapeutic indices. PD 178390: a lead HIV protease inhibitor. *Bioorganic & medicinal chemistry*, 7(12), 2775-2800.
- [37]. Hamilton, H.; Hagen, S.; Steinbaugh, B.; Lunney, E.; Pavlovsky, A.; Domagala, J.; Tummino, P.; Ferguson, D.; Gracheck, S. XIVth International Symposium on Medicinal Chemistry: Maastricht, The Netherlands, 8±12 September 1996.
- [38]. Olotu, F. A., Agoni, C., Soremekun, O., & Soliman, M. E. (2020). The recent application of 3D-QSAR and docking studies to novel HIV-protease inhibitor drug discovery. *Expert Opinion on Drug Discovery*, 15(9), 1095-1109.

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

I. Introduction

Le développement d'un médicament est un processus complexe, long et coûteux. On a estimé en 2004 que le coût de la mise sur le marché d'un seul médicament pouvait atteindre 800 millions de dollars [1]. Récemment, cette estimation est passée à 1,2 milliard de dollars [2]. Les étapes clés sont la recherche, l'optimisation et la conception d'un médicament-candidat (Drug design), l'évaluation préclinique, l'évaluation clinique, l'autorisation de mise sur le marché et la pharmacovigilance tout le long de la vie du médicament[3].

La Bio-informatique et la Chémo-informatique sont des disciplines devenues indispensables à la découverte de médicaments. De nos jours, les industries pharmaceutiques consacrent près de 10% de leur budget de recherche et développement, à la recherche de médicaments assistée par ordinateur[4]. Cette émergence peut s'expliquer à la fois par le développement des architectures et techniques de calculs mais aussi par le faible coût qu'engendrent des analyses sur ordinateur par rapport à des tests dans les laboratoires biologiques. Il est admis que, en moyenne, pour une molécule qui arrive sur le marché en tant que médicament innovant, 10000 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, coûteuses et parfois inutiles[5].

Les méthodes *in silico* telles que les relations quantitatives de structure-activité (QSAR) et le docking moléculaire peuvent réduire les échecs coûteux des candidats-médicaments en identifiant les principaux composés prometteurs et en réduisant ainsi le nombre d'expériences coûteuses. Ces méthodes font partie des principaux outils de découverte de médicaments [6].

II. Les relations quantitatives structures activités

Bien que l'étude des relations structure-activité ait commencé à la fin du 19^{ème} siècle, ce n'est qu'au début des années 60 que les travaux de Corwin et Hansch ont proposé un modèle mathématique pour corréler l'activité biologique et la structure chimique.

Pendant ces quarante dernières années, ce domaine a largement été étudié et les données bibliographiques disponibles sur cette approche sont maintenant importantes [7].

II.1. Définition

QSAR (Quantitative Structure Activity Relationship) est parmi les nouvelles techniques de modélisation, mettant en jeu des relations de structure avec l'activité. Elles prennent de plus en plus d'importance dans les études de la conception des précurseurs (lead compound). Ce sont des techniques alternatives et complémentaires aux techniques expérimentales *in vitro* et *in vivo* dans les phases très précoces du développement des médicaments. Un outil qui permet une prédiction rapide d'une ou plusieurs propriétés biologique. Elle a pu être mise en place dans les laboratoires et utilisée dans l'industrie pharmaceutique[8].

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

L'objectif d'une modélisation QSAR est de trouver des modèles précis, applicables et robustes afin de trouver une relation entre la structure et l'activité dans un but de prédiction mais également d'interprétation.

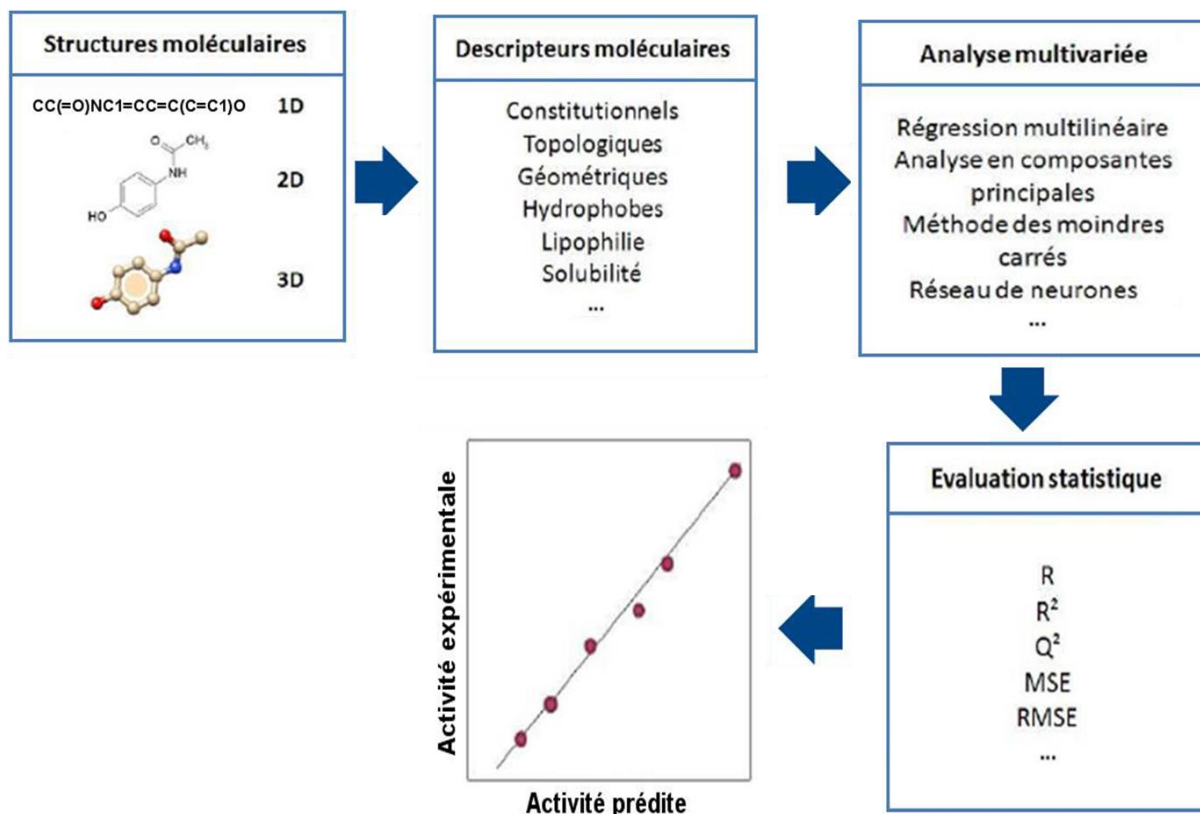
La forme mathématique la plus générale de QSAR est :

$$\text{Activité} = f(\text{propriétés physico-chimiques et / ou propriétés structurales})(1)$$

En général, pour prédire l'activité biologique par la méthode QSAR, la construction du modèle mathématique est effectuée en 5 étapes de base (Figure.1). L'ensemble des données (l'ensemble des molécules) utilisées dans l'étude de QSAR sont collectées soit directement à partir des mesures expérimentales, extraites de la littérature ou à partir de bases de données.

Les données collectées sont divisées en premier lieu aléatoirement en deux parties :

- Les données d'apprentissage « Training Set (TSET) » : utilisées pour la construction des modèles finaux.
- Les données de prédiction « Prédiction Set (PSET) » : utilisées pour la validation des modèles formés. Ces données ne subissent en revanche aucun traitement durant la construction des modèles et elles sont réservées uniquement pour tester la fiabilité des modèles[8].



Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Figure.1 : Les étapes de base de la construction du modèle de QSAR

La capacité de prédire une activité biologique est précieuse dans un grand nombre d'industries. Alors que certaines SAR semblent être un peu plus que des études universitaires, il existe un grand nombre d'applications de ces modèles au sein de l'industrie, du milieu universitaire et des organismes gouvernementaux (de réglementation). Un petit nombre d'utilisations potentielles sont énumérées ci-dessous :

- ✓ L'identification rationnelle et optimisation des nouvelles pistes avec une activité pharmacologique, biocide ou pesticide.
- ✓ La conception rationnelle de nombreux autres produits tels que les agents tensio-actifs, les parfums, les colorants et les produits chimiques fins.
- ✓ L'identification des composés dangereux aux premiers stades du développement du produit ou le criblage des inventaires des composés existants.
- ✓ La conception de la toxicité et des effets secondaires dans les nouveaux composés.
- ✓ La prédiction de la toxicité pour l'homme par une exposition délibérée, occasionnelle et professionnelle.
- ✓ La prédiction de la toxicité pour les espèces environnementales.
- ✓ La sélection de composés ayant des propriétés pharmacocinétiques optimales, que ce soit la stabilité ou la disponibilité dans les systèmes biologiques.
- ✓ La prédiction d'une variété de propriétés physico-chimiques des molécules (qu'il s'agisse de produits pharmaceutiques, de pesticides, de produits personnels, de produits de chimie fine, etc.).
- ✓ La prédiction du devenir des molécules libérées dans l'environnement. - La rationalisation et la prédiction des effets combinés des molécules, que ce soit dans des mélanges ou des formulations.

La caractéristique clé du rôle des technologies in silico dans tous ces domaines est que les prédictions peuvent être faites à partir de la seule structure moléculaire [9].

III. Méthodologie QSAR

III.1. Qualité des données

Les données devraient provenir du même protocole d'analyse, et il faudrait veiller à éviter la variabilité inter-laboratoire. Tout mauvais point de données aura tendance à corrompre la corrélation correcte de la structure et de l'activité [10].

Les règles empiriques pour un bon ensemble de données QSAR sont que la relation dose-réponse et l'activité (ou l'affinité) devrait être reproductible, la gamme d'activités devrait s'étendre de deux ordres de grandeur ou plus du produit chimique le moins actif au plus actif de

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

la série, le nombre de produits chimiques utilisés pour construire le modèle QSAR devrait être suffisamment important pour la stabilité, les activités des produits chimiques devraient être réparties uniformément dans toute la gamme d'activités, et les produits chimiques choisis pour l'ensemble d'entraînement devraient posséder suffisamment de diversité structurelle pour couvrir l'éventail de l'espace de chimie associé à l'activité biologique étudiée [11].

Des exemples de données, utilisés dans l'analyse de QSAR, sont :

Concentration Inhibitrice CI_{50} , Concentration Efficace CE_{50} , Cytotoxicité CC, fusion et ébullition [12].

II.2. Les descripteurs (Clusters)

II.2.1. Définition

Un descripteur moléculaire est un paramètre ayant une valeur numérique propre pour une structure chimique donnée [13]. Les descripteurs sont des représentations mathématiques formelles d'une molécule, obtenues par un algorithme bien défini et appliquées à une représentation moléculaire définie ou à une procédure expérimentale bien définie : le descripteur moléculaire est le résultat final d'une procédure logique et mathématique qui transforme l'information chimique codée dans une représentation symbolique d'une molécule en un nombre utile ou le résultat d'une expérience standardisée[14].

L'information codée par les descripteurs dépend généralement du type de représentation moléculaire et de l'algorithme défini pour son calcul.

La classification des descripteurs moléculaires est établie en fonction :

- ✓ Leurs origines (constitutionnel, géométrique, topologique, thermodynamique, quantique).
- ✓ Leurs dimensionnalités (1D, 2D ou 3D)[15].

La plupart des descripteurs, actuellement plus de 3000, peuvent être calculés théoriquement à partir de différents logiciels tels que : OASIS[16], ADAPT[17], CODESSA[18], DRAGON[19], GAUSSIAN [20], Canvas [21]....

III.2.2. Types de descripteurs

III.2.2.1. Les descripteurs 1D

L'activité est corrélée avec les propriétés moléculaires globales (constitutionnelles). Ce sont des descripteurs simples et couramment utilisés qui reflètent la composition moléculaire d'un composé sans aucune information sur sa topologie. Les descripteurs de constitution les plus courants sont le nombre d'atomes, le nombre de liaisons, le type d'atome, le nombre de cycles et le poids moléculaire (PM) [22].

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

III.2.2.2. Les descripteurs 2D

Les descripteurs 2D sont des descripteurs ou des indices topologiques (IT) qui prennent en compte l'arrangement atomique interne des composés. Ils sont dérivés de la représentation topologique des molécules et peuvent être considérés comme des descripteurs explicitement structurés.

Ces indices codent pour des informations sur la taille moléculaire, la forme, la ramification. La distance topologique d' entre deux atomes est alors définie comme le nombre de liaisons minimum qui relie ces atomes [23].

Les descripteurs 2D sont des propriétés numériques qui peuvent être calculées à partir de la table de connectivité d'une molécule ou d'une représentation planaire (2D) de la structure. Ils sont basés sur les éléments présents, les charges partielles, la nature des liaisons, etc., mais n'exploitent pas les coordonnées atomiques spatiales. On peut calculer ainsi :

- ✓ Des propriétés physiques : polarisabilité, charge totale, réfractivité moléculaire, masse, densité...
- ✓ Le dénombrement des atomes, liaisons et pivots.
- ✓ Des indices de formes et de connectivité : indices de Kier et Hall[24].
- ✓ Des descripteurs topologiques : matrices de distance et de connectivité.
- ✓ Les dénombrements de fonctions pharmacophoriques : donneurs, accepteurs de liaison hydrogène, polaire (donneur et accepteur), positif, négatif, hydrophobe et autres
- ✓ les descripteurs de charges partielles : charge partielle positive (négative), aire de la surface de Van der Waals positive (négative), aire de la surface de Van der Waals polaire (hydrophobe),[25].

Les descripteurs 2D sont employés pour obtenir des modèles QSAR plus simples, mais leur défaillance, comme pour les descripteurs 1D, est qu'ils ne permettent pas la bonne interprétation des mécanismes d'interaction mis en jeu pour l'activité/propriété étudiée.

III.2.2.3. Les descripteurs 3D

Il s'agit de descripteurs évalués à partir des positions relatives des atomes d'une molécule dans l'espace, ainsi que des rayons et masses atomiques. Ils nécessitent donc de connaître la structure 3D de la molécule[26].

Les descripteurs 3D décrivent des objets tridimensionnels et se répartissent en 2 groupes : ceux qui ne dépendent que des coordonnées internes de la molécule et ceux qui dépendent de son orientation absolue. Parmi tous ces descripteurs on peut citer :

- ✓ Les descripteurs de l'énergie potentielle : valeur de l'énergie potentielle et composante de cette énergie : Van der Waals, électrostatiques, atomes « hors-du-plan », torsion, solvation,
- ✓ Les descripteurs thermodynamiques : Le point d'ébullition, la constante de Henry, la

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

température, la pression, l'enthalpie et le point de fusion.

- ✓ Les descripteurs géométriques (Des descripteurs de formes et de volumes) : aire et volume de Van der Waals, surface accessible au solvant, moment d'inertie, globularité (molécule sphérique, plane ou linéaire).
- ✓ Les descripteurs du moment dipolaire : orientation et intensité.
- ✓ Les descripteurs physico-chimiques (les champs de potentiel d'interaction moléculaire) : stérique, électrostatique, lipophile, la densité, donneur ou accepteur d'hydrogène.
- ✓ Les descripteurs électroniques : densité électronique, distribution de charges.
- ✓ Les descripteurs de spectres énergétiques : IR, RAMAN, RMN [27].

III.3. Méthodes statistiques

Les techniques statistiques ou chimio-métriques constituent la base mathématique de la construction d'un modèle QSAR [28].

Dans la terminologie des statistiques, la propriété étudiée est appelée variable cible ou variable dépendante ou encore variable Y. C'est la caractéristique que l'on cherche à expliquer et à optimiser.

Les descripteurs, en l'occurrence chimiques, sont appelés variables explicatives, variables indépendantes ou encore variables X.

Le modèle statistique est une équation donnant la valeur de la variable cible en fonction de la somme des valeurs pondérées des descripteurs. Les principaux outils statistiques pour obtenir un modèle sont :

III.3.1. Régression linéaire

III.3.1.1. Régression linéaire simple (RLS)

Cette méthode effectue comme un calcul de régression linéaire standard dans la génération du modèle QSAR sous la forme d'équations qui incluent un seul descripteur indépendant x et y en tant que variable dépendante. Cette technique s'avère très prometteuse pour générer des relations de structure et d'activité en explorant certains des descripteurs les plus importants utilisés pour gouverner l'activité, alors que certaines des interactions de plusieurs descripteurs ont été négligées.

La régression linéaire simple peut être exprimée par l'équation 2 :

$$y = a + bx \quad (2)$$

Où y est la variable dépendante, x est la variable indépendante, a est la constante, b est le coefficient de régression.

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

III.3.1.2. Régression linéaire multiple (RLM)

La régression multilinéaire est la méthode la plus simple et la plus employée pour le développement de modèles prédictifs. Elle repose sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante Y et une série de n variables indépendantes Xi (ici, les descripteurs). L'objectif est d'obtenir une équation de la forme suivante [29] :

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n(3)$$

Où b_i sont les coefficients de la régression.

III.3.2. Régression aux moindres carrés partiels (PLS)

Régression aux moindres carrés partiels (*PLS, pour Partial Least Squares ou Projection on Latent Structures*)[30] donne une solution statistiquement robuste même lorsque les variables indépendantes sont fortement inter-reliées entre elles, ou lorsque les variables indépendantes dépassent le nombre d'observations. PLS est une méthode de régression itérative qui produit ses solutions basées sur la transformation linéaire d'un grand nombre de descripteurs originaux en un petit nombre de nouveaux termes orthogonaux appelés variables latentes. Ainsi, cette méthode est comptée comme statistique standard [31].

III.3.3. Réseaux de neurones artificiels (RNA)

Les RNA sont des méthodes utiles dans les études QSAR / QSPR, et en particulier dans les cas où il est difficile de spécifier un modèle mathématique exact pour décrire une relation de propriété de structure donnée.

La méthode des réseaux neuronaux artificiels provient des vrais neurones présents dans un cerveau animal. Les RNA sont des systèmes de calcul parallèles constitués de groupes d'éléments de traitement hautement interconnectés appelés neurones, qui sont disposés en une série de couches. Chaque couche peut faire ses calculs indépendamment et peut transmettre les résultats à une autre couche. En cela, le résultat de la fonction de transfert est communiqué aux neurones dans la couche de sortie. C'est le point où les résultats sont interprétés et présentés finalement[32].

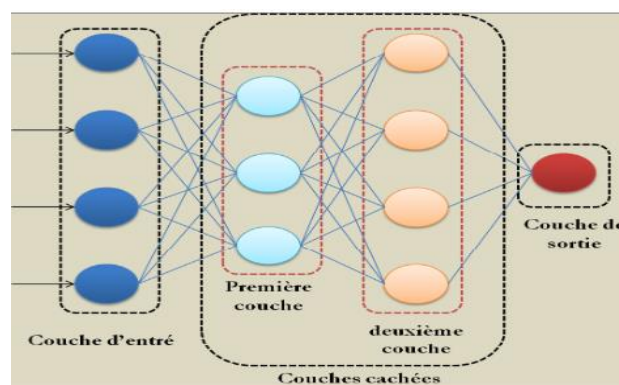


Figure. 2 : Schéma représentatif d'un réseau de neurones artificiels [33].

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

III.3.4. Les machines à vecteurs de support

Les machines à vecteurs de support ou séparateurs à vaste marge (en anglais Support Vector Machine, SVM) sont un ensemble de techniques d'apprentissage supervisé destinées à résoudre des problèmes de discrimination et de régression basés sur la théorie de l'apprentissage statistique (généralement considérés comme la 1^{ère} réalisation pratique de cette théorie [34] et qui respecte les principes du (SRM) « structural riskminimization ». L'émergence des SVMs a commencé autour des débuts des années 1990 par les mathématiciens russes Vladimir Vapnik et Alex Chervonenkis [35-36].

Les SVMs sont considéré comme l'un des modèles les plus importants parmi la famille des méthodes à Kernel (noeud).

III.4. Validation du modèle

Après l'obtention de l'équation du modèle, outre la stabilité et la qualité de l'ajustement du modèle, il est également important d'estimer la puissance et la validité du modèle avant de l'utiliser pour prédire l'activité biologique. La validité consiste à établir la fiabilité et la signification de la méthode pour un usage particulier. Par conséquent, la validation d'un modèle QSAR doit être effectuée. Il existe trois types de validation : validation interne / validation externe / Coefficients et tests statistiques standards[37].

III.4.1. Validation interne

La technique la plus employée pour déterminer la stabilité du modèle prédictif est de tester l'influence de chaque échantillon sur le modèle final. Pour ce faire, on emploie une technique de validation croisée (*cross validation* ou CV).

Ce processus consiste à extraire un certain nombre n de molécules du jeu initial à k molécules et à construire un nouveau modèle avec les $n-k$ molécules restantes à l'aide des descripteurs choisis (seules les constantes de la régression changent). Ce nouveau modèle est alors utilisé pour la phase de prédiction sur les n molécules retirées. Ce processus est ensuite réitéré pour retirer et prédire les valeurs de toutes les molécules du jeu d'entraînement[37].

III.4.1.1. La procédure Leave-One-Out(LOO)

On retire successivement une molécule de l'ensemble d'apprentissage (TSET). Un modèle QSAR est construit alors sur un ensemble de $k-1$ de composés et la molécule retirée est prédite par le modèle formé, « m » représente le nombre de molécules utilisées dans la construction du modèle (TSET). Cette procédure est réitérée m fois afin de prédire les propriétés de toutes les molécules [37].

III.4.1.2. La procédure Leave-Many-Out (LMO)

Elle correspond à un découpage en plusieurs parties de l'ensemble de données. A tour de rôle, une partie de l'ensemble de données est attribuée pour un ensemble de test interne. Les autres constituent l'ensemble d'entraînement. Les molécules de chaque groupe éliminé sont prédites

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

par le modèle formé. Cette procédure est répétée p fois pour prédire les propriétés de toutes les molécules, p est le nombre des groupes de molécules éliminés[37].

III.4.2. Validation externe

La validité parfaite du modèle est examinée par validation externe, qui évalue la généralisation du modèle. L'ensemble de données original est divisé en deux sous-groupes, l'ensemble d'apprentissage et l'ensemble d'essai. L'ensemble d'apprentissage ou d'étalonnage est utilisé pour dériver un modèle d'étalonnage qui sera utilisé plus tard pour prédire les activités des composés d'essai ou de validation. D'autre part, un ensemble de test externe qui n'a été inclus dans aucune étape de la construction du modèle peut être utilisé comme ensemble de test. Les paramètres quantifiant la supériorité de la prédiction de l'ensemble de test externe peuvent être les mêmes que ceux utilisés pour la validation interne. La somme des erreurs de prédiction des carrés (SSPE) est largement utilisée pour rendre compte de l'incohérence [37].

III.4.3. Coefficients et tests statistiques standards

Plusieurs tests et coefficient peuvent être utilisés pour juger et effectuer la validation de modèle QSAR [38], les tests standards pour validation de modèle sont :

- ✓ Le coefficient de détermination R^2 (coefficient de corrélation) est l'indicateur statistique le plus répandu qui évalue la part de la variance de l'activité cible expliquée par le modèle

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (4)$$

Où :

y: La valeur expérimentale de l'activité

\hat{y} : La valeur calculée de l'activité

\bar{y} : La valeur moyenne des valeurs calculées de l'activité.

Plus la valeur de R^2 sera proche de 1 plus les valeurs calculées et observées sont corrélées.

La valeur de R^2 dépend de la taille de l'échantillon et le nombre de variables prédictives dans le modèle. Il garde la même valeur ou augmente lors de l'ajout d'une nouvelle variable de prédiction, même si la variable ajoutée ne contribue pas à la réduction de la variance inexpliquée.

- ✓ Par conséquent, un autre paramètre statistique peut être utilisé, appelé R^2 ajusté (R^2_{adj})

$$R^2_{adj} = \sqrt{\frac{R^2 (n-1) - p}{n-p-1}} \quad (5)$$

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Avec : n est le nombre des variables dépendantes (les molécules) ; p est le nombre de variables indépendantes (les descripteurs) ; R² est le coefficient de détermination.

- ✓ La déviation standard DS est un autre indicateur statistique utilisé

$$DS = \sqrt{\frac{\sum(\hat{y}_i - y_i)^2}{n - p - 1}} \quad (6)$$

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

En particulier, plus la déviation standard est petite et plus la corrélation est meilleure.

- ✓ L'indice de Fisher *F* est également employé afin de mesurer le niveau de signification statistique du modèle, c'est-à-dire la qualité du choix des descripteurs constituant le modèle.

$$F = \frac{\sum((\hat{y}_i - \bar{y}_i)^2(n - p - 1))}{(y_i - \hat{y}_i)^2 \cdot p} \quad (7)$$

III.5. Domaine d'applicabilité

Un modèle QSAR/QSPR ne peut pas être considéré comme un modèle universel, parce qu'il est développé sur un nombre limité de composés qui ne couvrent pas tout l'espace chimique. Par conséquent, l'activité/propriété prédite d'un composé, chimiquement dissimilaire au jeu d'apprentissage, ne pourra pas être considérée fiable [39,40].

Le domaine d'applicabilité (DA) permet de définir la zone dans laquelle un composé pourra être prédit avec confiance. Le DA correspond donc à la région de l'espace chimique incluant les composés de la série d'apprentissage et les composés similaires, proches dans ce même espace [41].

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

IV. Conclusion

Les QSARs est un outil largement utilisé pour développer des relations entre les effets (par exemple, les activités et les propriétés d'intérêt) d'une série de molécules et leurs propriétés structurales. Il est utilisé dans de nombreux domaines scientifiques. C'est une zone dynamique qui intègre de nouvelles technologies à un rythme effréné[42].

Un modèle QSAR est dépendant des données expérimentales de référence, le choix de la base de données est un point critique dans le développement de ces modèles. Dans la plupart des cas, les données expérimentales sont issues de la littérature, et pour être de qualité, une base de données doit être composée de données expérimentales aussi fiables que possible, puisque les barres d'erreurs sur celles-ci se propageront dans le modèle final, étant donné que les paramètres de ce dernier sont ajustés par rapport à ces données.

Il est donc important de choisir des données présentant des incertitudes faibles, afin de limiter les barres d'erreur expérimentales. De plus, les données doivent être obtenues suivant un protocole expérimental unique. En effet, les conditions expérimentales ont une forte influence sur les valeurs obtenues. La définition de la propriété en termes de conditions expérimentales est d'ailleurs un point important de la démarche.

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Bibliographie

- [1]. Rawlins, M. D. (2004). Cutting the cost of drug development?. *Nature reviews Drug discovery*, 3(4), 360-364.
- [2]. Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., & Schacht, A. L. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature reviews Drug discovery*, 9(3), 203-214.
- [3]. Daminelli, S., Haupt, V. J., Reimann, M., & Schroeder, M. (2012). Drug repositioning through incomplete bi-cliques in an integrated drug–target–disease network. *Integrative Biology*, 4(7), 778-788.
- [4]. Meslamani, J. E. (2012). Développement de nouvelles méthodes de criblage in silico en chémogénomique (Doctoral dissertation, Strasbourg).
- [5]. MEDJAHED, S. (2018). Etude par des méthodes QSAR de plusieurs séries de molécules hétérocycliques à intérêt pharmaceutique (Doctoral dissertation, Université Mohamed Boudiaf des Sciences et de la Technologie-Mohamed Boudiaf d'Oran).
- [6]. Zekri, A. (2021). Une étude computationnelle des relations structureactivité d'un ensemble de composés à visées thérapeutiques (Doctoral dissertation, Université de mohamed kheider biskra).
- [7]. KHAIRELINE, K. (2009). études de QSAR sur des activités biologiques utilisant des produits d'origines naturels (Doctoral dissertation, Université 8 mai 1945 de Guelma).
- [8]. Ghose, A., & Viswanadhan, V. (2001). *Combinatorial Library Design and Evaluation: Principles, Software, Tools, and Applications in Drug Discovery*. CRC Press.
- [9]. Cronin, M. T. (2010). Quantitative structure–activity relationships (QSARs)–applications and methodology. In *Recent Advances in QSAR Studies* (pp. 3-11). Springer, Dordrecht.
- [10]. Blaney, J. M., & Martin, E. J. (1997). Computational approaches for combinatorial library design and molecular diversity analysis. *Current opinion in chemical biology*, 1(1), 54-59.
- [11]. Hopfinger, A. J., & Duca, J. S. (2000). Extraction of pharmacophore information from high-throughput screens. *Current Opinion in Biotechnology*, 11(1), 97-103.
- [12]. C.D. Selassie. History of quantitative structure-activity relationships", in Book "History of quantitative structure-activity relationships", D. J. Abraham, Editor., John Wiley& Sons, Inc., California (2003).
- [13]. Karelson, M. (2000). *Molecular descriptors in QSAR/QSPR*. Wiley-Interscience.
- [14]. Khan, A. U. (2016). Descriptors and their selection methods in QSAR analysis: paradigm for drug design. *Drug discovery today*, 21(8), 1291-1302.
- [15]. Shahlaei, M. (2013). Descriptor selection methods in quantitative structure–activity relationship studies: a review study. *Chemical reviews*, 113(10), 8093-8103.
- [16]. Mekenyan, O., & Bonchev, D. (1986). Oasis method for predicting biological-activity of chemical-compounds. *Acta Pharmaceutica Jugoslavica*, 36(2), 225-237.

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

- [17]. Guha, R., Serra, J. R., & Jurs, P. C. (2004). Generation of QSAR sets with a self-organizing map. *Journal of Molecular Graphics and Modelling*, 23(1), 1-14.
- [18]. AR. Katritzky, M. Karelson, R. Petrukhin, CODESSA PRO, University of Florida 2001-2005. <http://www.codessa-pro.com>.
- [19]. R. Todeschini, Milano Chemometrics and QSPR Group, (<http://www.disat.unimib.it/vhml>).
- [20]. Gaussian 16, Revision C.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc, Wallingford CT, 2016.
- [21]. Canvas Version 3.2.013, MMshare Version 3.8.013 Legal Notices Schrodinger Software Modules Copyright (c) Schrodinger, LLC. All rights reserved.
- [22]. Hicham, B., & Ayoub, K. (2018). Relation Structure Activité: Etude Qualitative et Quantitative et Développement de Recherche sur les Coumarines. Université Abou Bekr Belkaïd Tlemcen.
- [23]. Fayet, G. (2010). Développement de modèles QSPR pour la prédiction des propriétés d'explosibilité des composés nitroaromatiques (Doctoral dissertation, Chimie ParisTech).
- [24]. Hall, L. H., & Kier, L. B. (1991). The molecular connectivity chi indexes and kappa shape indexes in structure-property modeling. *Reviews in computational chemistry*, 367-422.
- [25]. N.Melkemi.(2013). Etude des corrélations entre les paramètres structuraux et les propriétés bioactives de quelques composés organiques (Doctoral dissertation, Université Mohamed Khider Biskra).
- [26]. Karelson, M., Lobanov, V. S., & Katritzky, A. R. (1996). Quantum-chemical descriptors in QSAR/QSPR studies. *Chemical reviews*, 96(3), 1027-1044.
- [27]. Fortuné, A. (2006). Techniques de Modélisation Moléculaire appliquées à l'Etude et à l'Optimisation de Molécules Immunogènes et de Modulateurs de la Chimiorésistance (Doctoral dissertation, Université Joseph-Fourier-Grenoble I).
- [28]. BEN SGHIER FATMA ZAHRA, A. A. Conception de nouvelles molécules à activité anti-oxydante par des études 3D QSAR.

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

- [29]. Montgomery D C., Elizabeth A P., G. Geoffrey V G., (2013). Introduction to Linear Regression Analysis, Fifth Edition by Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining. *International Statistical Review*, 81(2):318-319.
- [30]. Boulesteix, A. L., & Strimmer, K. (2007). Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefings in bioinformatics*, 8(1), 32-44.
- [31]. Patel, H. M., Noolvi, M. N., Sharma, P., Jaiswal, V., Bansal, S., Lohan, S., ...& Bhardwaj, V. (2014). Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery. *Medicinal chemistry research*, 23(12), 4991-5007.
- [32]. Hoffman, B. T., Kopajtic, T., Katz, J. L., & Newman, A. H. (2000). 2D QSAR modeling and preliminary database searching for dopamine transporter inhibitors using genetic algorithm variable selection of Molconn Z descriptors. *Journal of medicinal chemistry*, 43(22), 4151-4159.
- [33]. Hmamouchi, R., Bouachrine, M., & Lakhlifi, T. (2016). Tentative Pratique du Relation Quantitatives Structure-Activité/Propriété (QSAR/QSPR). *Revue Interdisciplinaire*, 1(1).
- [34]. Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.
- [35]. Vapnik, V. N. (1995). The nature of statistical learning. Theory.
- [36]. Conforti, D., & Guido, R. (2010). Kernel based support vector machine via semidefinite programming: Application to medical diagnosis. *Computers & Operations Research*, 37(8), 1389-1394.
- [37]. Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American statistical Association*, 88(422), 486-494.
- [38]. Wold, S., Eriksson, L., & Clementi, S. (1995). Statistical validation of QSAR results. *Chemometric methods in molecular design*, 309-338.
- [39]. Jaworska, J., Nikolova-Jeliazkova, N., & Aldenberg, T. (2005). QSAR applicability domain estimation by projection of the training set in descriptor space: a review. *Alternatives to laboratory animals*, 33(5), 445-459.
- [40]. Tetko, I. V., Sushko, I., Pandey, A. K., Zhu, H., Tropsha, A., Papa, E., ...& Varnek, A. (2008). Critical assessment of QSAR models of environmental toxicity against *Tetrahymena pyriformis*: focusing on applicability domain and overfitting by variable selection. *Journal of chemical information and modeling*, 48(9), 1733-1746.
- [41]. Netzeva, T. I., Worth, A. P., Aldenberg, T., Benigni, R., Cronin, M. T., Gramatica, P., ...& Yang, C. (2005). Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships: The report and recommendations of ecvam workshop 52. *Alternatives to Laboratory Animals*, 33(2), 155-173.
- [42]. Dguigui, K., Mbarki, M., Elhallaoui, M., Elasri, M., & Bouachrine, M. (2010). QSAR Study of conductors polymers based on thiophene. *J. Mater. Environ. Sci*, 1, 175-182.

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

I. Introduction

Les modèles quantitatifs de la relation structure-activité (QSARs) sont des méthodes de régression quantitative qui permettent de relier la structure chimique à l'activité biologique [1]. Elles ont été appliquées de manière extensive dans plusieurs disciplines scientifiques : la chimie, la biologie et la toxicologie. Les modèles QSARs sont considérés comme un outil scientifiquement crédible pour prédire et classer les activités biologiques des produits chimiques non testés. Il est donc devenu très essentiel dans l'industrie pharmaceutique comme méthode de découverte et d'optimisation de pointe pour diriger le développement de nouveaux produits [2].

Le rôle de la théorie QSAR est de proposer des modèles mathématiques capables d'estimer les propriétés d'intérêt pertinentes, qui ne peuvent être déterminées expérimentalement. Ces études reposent sur l'hypothèse que la structure d'un composé détermine ses propriétés. Ces propriétés doivent être traduites en descripteurs moléculaires.

L'analyse basée sur les relations quantitatives structure-activité suppose une relation entre les propriétés chimiques d'une molécule et son comportement physique, comme la liaison à une cible protéique [3]. Elle constitue donc une solution pour remédier au manque de données expérimentales dans les phénomènes chimiques complexes. A cet effet, il existe un intérêt pour le développement de telles techniques prédictives [4-5]. Les techniques utilisées dans QSAR reposent sur l'hypothèse que l'information requise pour décrire les propriétés biologiques d'une molécule donnée réside dans sa structure [6].

II. Méthodologie

II.1. Base de données

Dans nos études QSARs, nous avons utilisé une base de données comprenant Quarante-et-un composés de dérivés de pyranone ayant une activité biologique rapportées par les inhibiteurs anti-VIH de la protéase. Ils ont été tirés de la littérature [7]. L'activité biologique est exprimée en concentration efficace (EC_{50}) qui a été convertie en $\log(EC_{50})$ (Tableau.1).

Cette base de données a été divisée automatiquement en 75 % pour l'ensemble d'apprentissage et 25 % pour l'ensemble de test pour les études 2D et 3D QSARs.

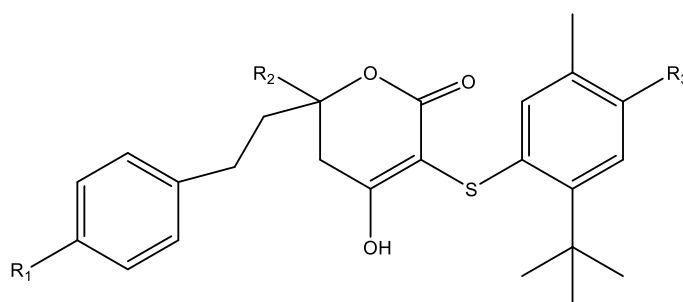


Figure.1. Structure générale de Pyranone de notre travail

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Tableau.1 : Les structures chimiques ainsi que les données d'activité observées des composés

Molécule	R ₁	R ₂	R ₃	Log EC ₅₀
Pyranone1	4-OH	Phenyl	H	4.1612
Pyranone 2	4-OH	Methyl	H	5.0269
Pyranone 3	4-OH	n-propyl	H	5.1079
Pyranone 4	4-OH	n-Butyl	H	4.6383
Pyranone 5	4-OH	n-Pentyl	H	4.5686
Pyranone 6	4-OH	Isopropyl	H	5.7212
Pyranone 7	4-OH	cyclopropyl	H	5.3010
Pyranone 8	4-OH	cyclopentyl	H	5.2291
Pyranone 9	4-OH	cyclohexyl	H	5.3872
Pyranone 10	4-OH	Phenyl	CH ₂ OH	5.3279
Pyranone 11	4-OH	Methyl	CH ₂ OH	5.6021
Pyranone 12	4-OH	n-propyl	CH ₂ OH	6.0000
Pyranone 13	4-OH	n-Butyl	CH ₂ OH	6.0458
Pyranone 14	4-OH	n-Pentyl	CH ₂ OH	6.2218
Pyranone 15	4-OH	Isopropyl	CH ₂ OH	6.2218
Pyranone 16	4-OH	Isobutyl	CH ₂ OH	5.8239
Pyranone 17	4-OH	cyclopropyl	CH ₂ OH	5.7959
Pyranone 18	4-OH	cyclobutyl	CH ₂ OH	6.0969
Pyranone 19	4-OH	cyclopentyl	CH ₂ OH	6.2218
Pyranone 20	4-OH	cyclohexyl	CH ₂ OH	6.3010
Pyranone 21	4-OH	Methyl	OH	5.1549
Pyranone 22	4-OH	Isopropyl	OH	5.7959
Pyranone 23	4-OH	cyclohexyl	OH	5.5086
Pyranone 24	4-OH	Methyl	OCH ₂ CH ₂ OH	5.4685

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Pyranone 25	4-OH	Methyl	OCH ₂ CH ₂ OCH ₂ OCH ₃	5.3372
Pyranone 26	4-OH	n-propyl	OCH ₂ CH ₂ OH	5.3665
Pyranone 27	4-OH	n-Butyl	OCH ₂ CH ₂ OH	5.7696
Pyranone 28	4-OH	n-Pentyl	OCH ₂ CH ₂ OH	5.0458
Pyranone 29	4-OH	Isopropyl	OCH ₂ CH ₂ OH	5.8239
Pyranone 30	4-OH	cyclopropyl	OCH ₂ CH ₂ OH	5.7959
Pyranone 31	4-OH	cyclopentyl	OCH ₂ CH ₂ OH	5.9208
Pyranone 32	4-NH ₂	Phenyl	CH ₂ OH	5.4318
Pyranone 33	4-NH ₂	Isopropyl	CH ₂ OH	6.3010
Pyranone 34	4-NH ₂	Isobutyl	CH ₂ OH	6.0969
Pyranone 35	4-NH ₂	cyclohexyl	CH ₂ OH	5.8539
Pyranone 36	H	Isopropyl	H	4.8239
Pyranone 37	H	Phenyl	CH ₂ OH	5.6021
Pyranone 38	H	Isopropyl	CH ₂ OH	6.2218
Pyranone 39	H	Isopropyl	OH	5.7447
Pyranone 40	H	Phenyl	OCH ₂ CH ₂ OH	4.6990
Pyranone 41	H	Isopropyl	OCH ₂ CH ₂ OH	5.8861

II.2. Modèles QSARs

II.2.1. Les modèles 2D-QSAR

II.2.1.1. Calcul des descripteurs

Le calcul de descripteurs des molécules nécessite une étape qui utilise des descripteurs 2D et plus de 750 descripteurs dans le logiciel Canvas (Schrödinger) [8]. Les descripteurs calculés ont été criblés pour leur nature invariante et sélectionnés par le module Feature Selection de Canvas. En outre, l'inter corrélation entre les paires de descripteurs a été effectuée pour l'optimisation de la sélection du nombre de descripteurs considérés pour l'étude.

Un dernier ensemble de 20Clusters identifiés après sélection a été utilisé dans l'étape de construction du modèle QSAR.

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Pendant l'étape de génération du modèle QSAR, nous avons utilisé le module Choose best subsets comme méthode pour choisir les meilleurs descripteurs (paramètres physico-chimiques et topologiques) tels que : l'indice de charge, nombre de liaison, l'indice d'état électrotopologique (E state), l'égalisation partielle de l'électronégativité orbitalaire (PEOE) et le coefficient de partage octanol-eau de Ghose-Crippen-Viswanadhan (ALOGP).

II.2.1.2.Méthode statistique

L'analyse de la régression est une méthode statistique pour la modélisation des relations entre différentes variables (dépendantes et indépendantes). Elle est utilisée pour décrire et analyser les relations entre les données. On peut faire des prédictions en utilisant des analyses de régression, par lesquelles les relations entre les données seront utilisées comme une base pour la prévision et la génération d'un modèle de prédiction.

Les modèles QSARs ont été générés en utilisant la méthode de régression multilinéaire(MLR) implémentée dans le logiciel Canvas (Schrödinger).

Le modèle a été évalué par la valeur R^2 , le SD et Fisher.

Mod.1:

La meilleure corrélation est exprimée par la relation mathématique suivante:

$$\text{LogEC}_{50} = -2.2264 + 0.3390D_1 + 0.0903D_2 + 0.2179D_3 + 170.915D_4 - 0.1537D_5 \text{ (Eq. 01)}$$

Le modèle exprimé par l'équation (Eq.01) reliant la variable dépendante $\log EC_{50}$ avec les cinq descripteurs est accompagné par les mesures statistiques citées ci-dessus :

$$R^2 = 0.803, SD = 0.261, F = 20.448$$

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

La figure.2 montre la corrélation linéaire entre les valeurs observées et les valeurs prédites de l'activité mentionnées dans le Tableau.2 obtenue en utilisant le modèle Mod.1.

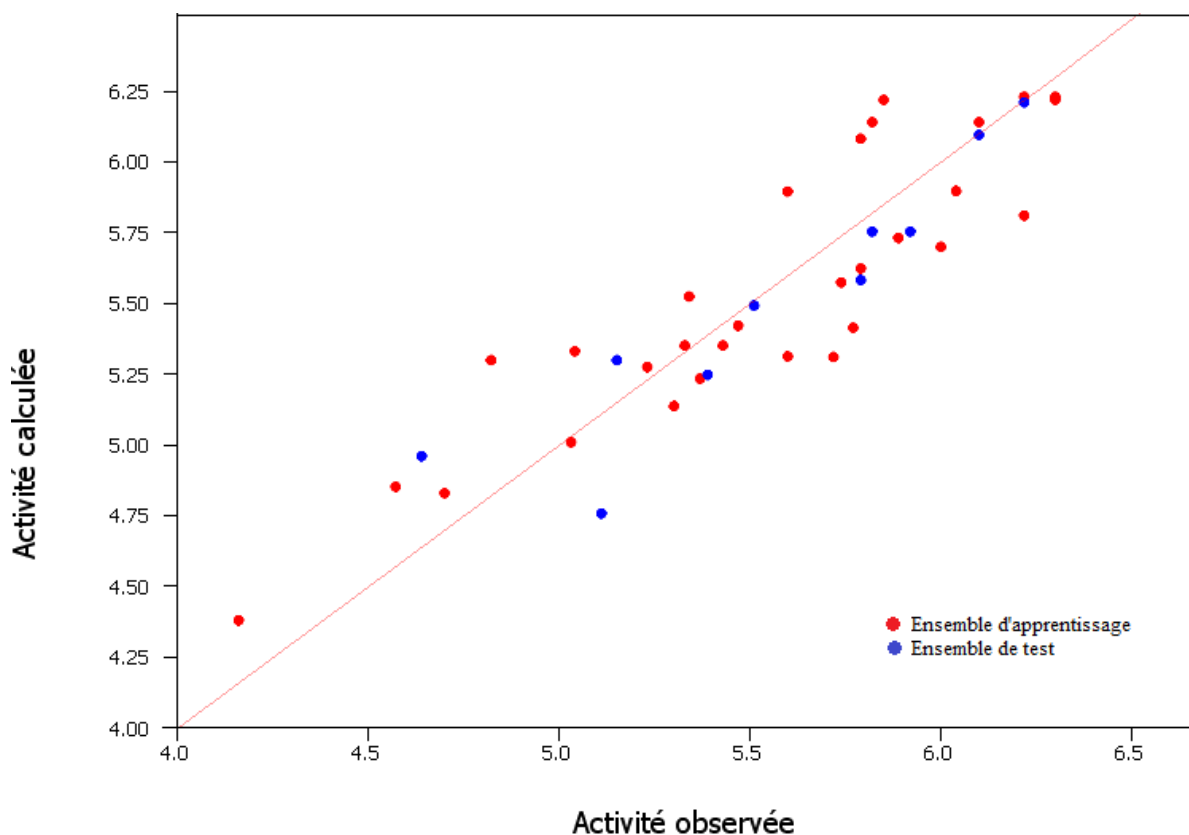
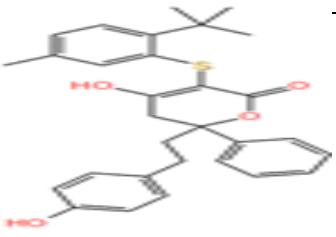
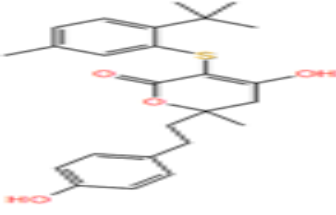
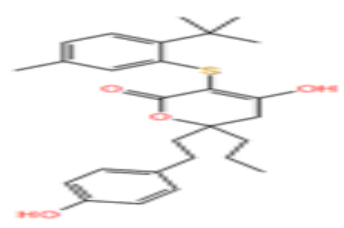
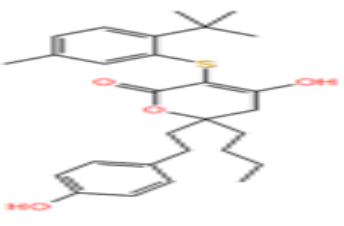
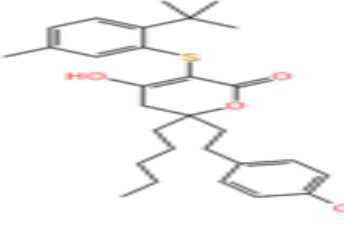
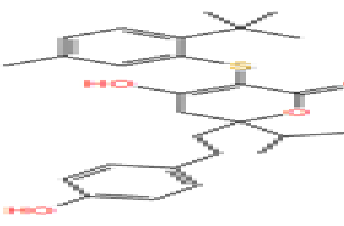
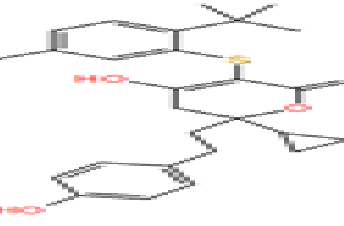
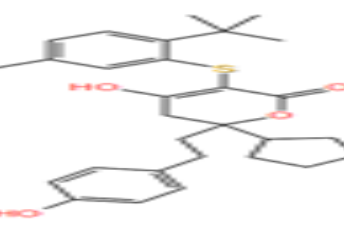
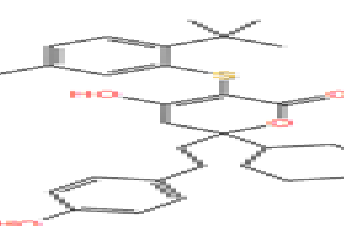


Figure.2: La corrélation linéaire entre les valeurs observées et les valeurs calculées pour la 2D-QSAR

Tableau.2 :Les valeurs des activités observées, les valeurs calculées et le résiduel du 2D-QSAR.

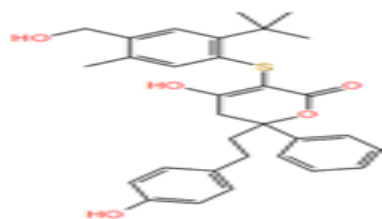
Molécule	Structure	Log EC ₅₀ Observée	Log EC ₅₀ Calculée	Résiduel
Pyranone 1		4.1612	4.37803	-0.21683
Pyranone 2		5.0269	5.00927	0.01763

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Pyranone 3		5.1079	4.75796	0.34994
Pyranone 4		4.6383	4.96065	-0.32235
Pyranone 5		4.5686	4.85102	-0.28242
Pyranone 6		5.7212	5.31017	0.41103
Pyranone 7		5.3010	5.13833	0.16267
Pyranone 8		5.2291	5.27506	-0.04596
Pyranone 9		5.3872	5.24678	0.14042

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Pyranone10

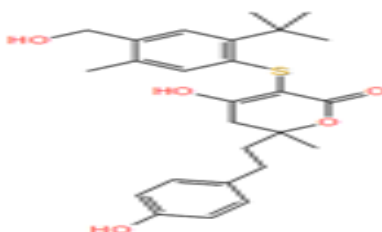


5.3279

5.35024

-0.02234

Pyranone11

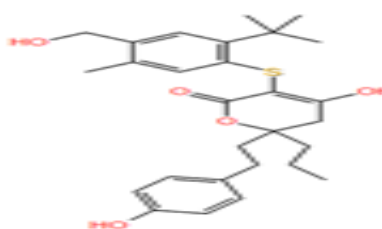


5.6021

5.89521

-0.29311

Pyranone12

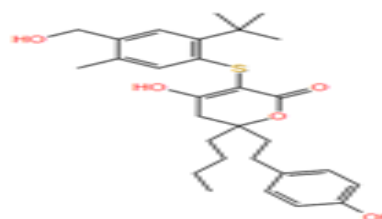


6.0000

5.70031

0.29969

Pyranone13

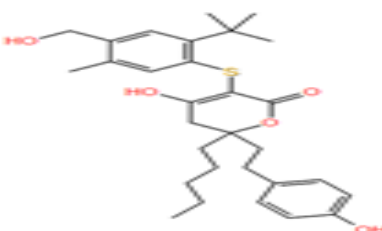


6.0458

5.89835

0.14745

Pyranone14

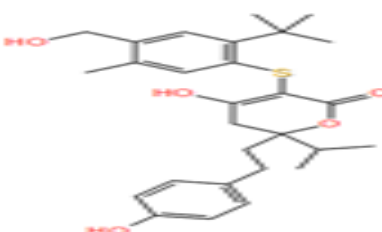


6.2218

5.8093

0.4125

Pyranone15

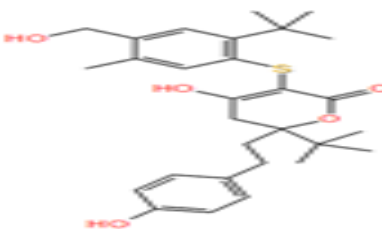


6.2218

6.2318

-0.01

Pyranone16

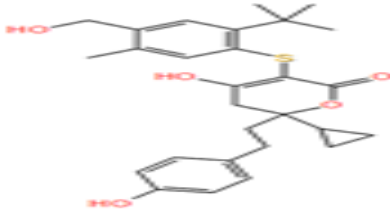
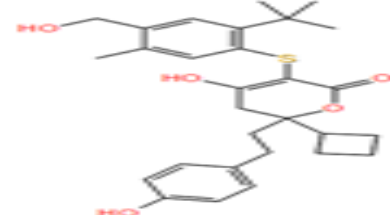
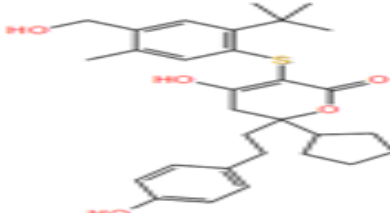
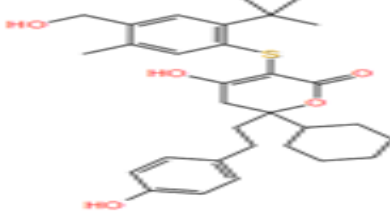
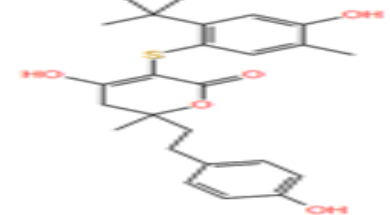
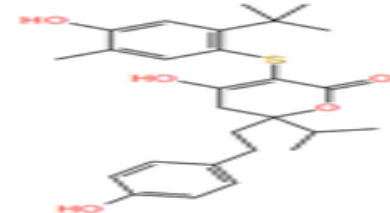
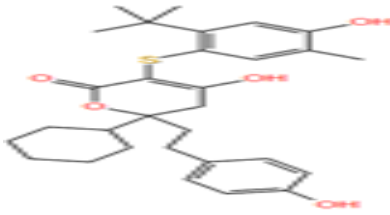


5.8239

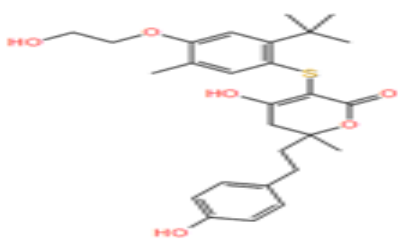
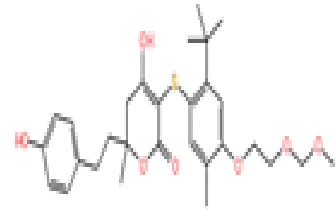
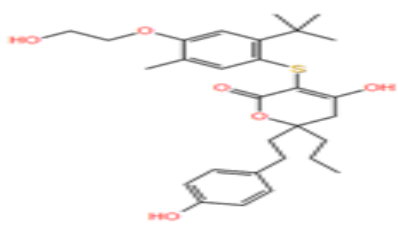
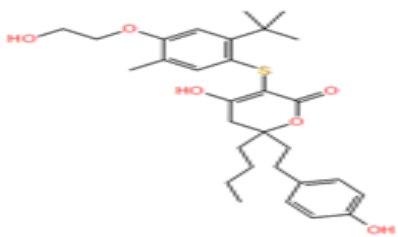
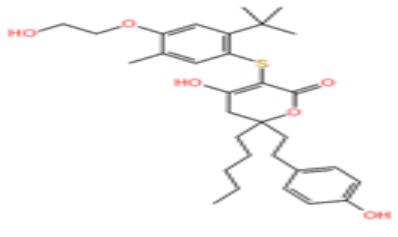
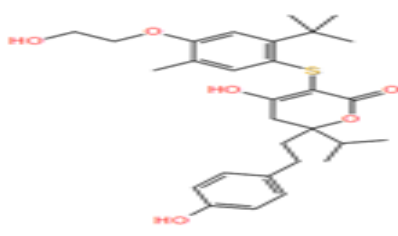
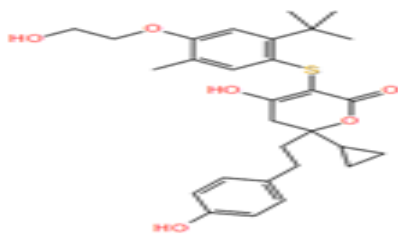
6.14192

-0.31802

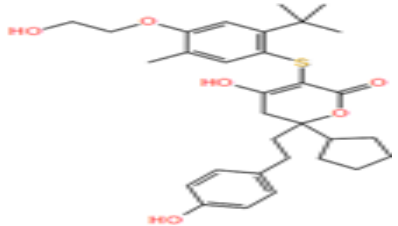
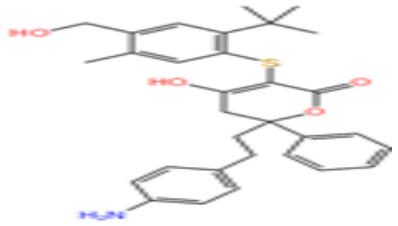
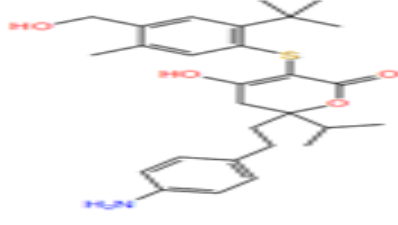
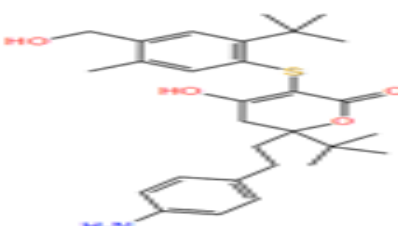
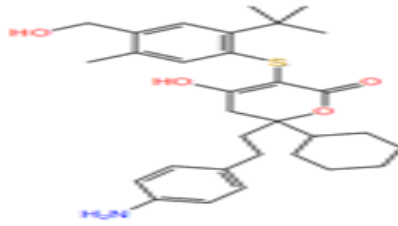
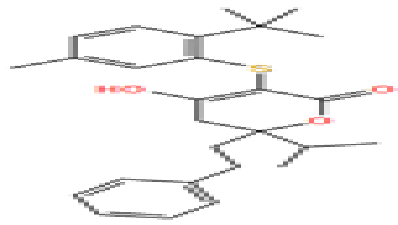
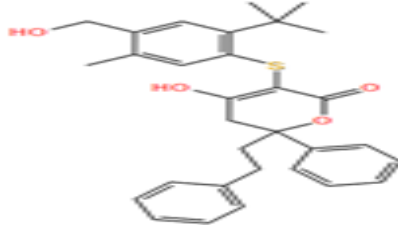
Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Pyranone17		5.7959	6.08337	-0.28747
Pyranone18		6.0969	6.09523	0.00167
Pyranone19		6.2218	6.23498	-0.01318
Pyranone20		6.3010	6.21898	0.08202
Pyranone21		5.1549	5.29979	-0.14489
Pyranone22		5.7959	5.5818	0.2141
Pyranone23		5.5086	5.4935	0.0151

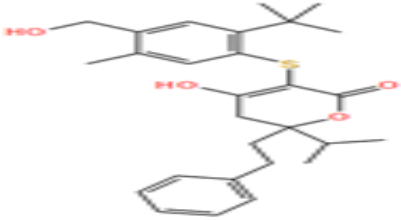
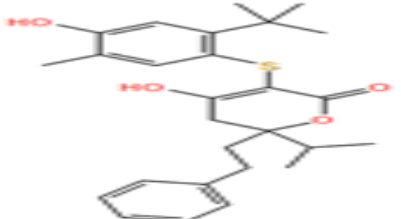
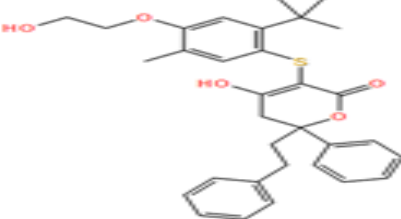
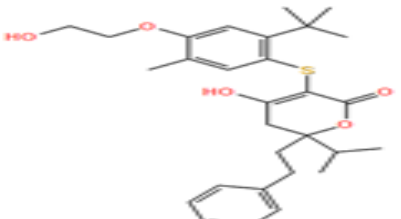
Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Pyranone24		5.4685	5.42136	0.04714
Pyranone25		5.3372	5.5234	-0.1862
Pyranone26		5.3665	5.234	0.1325
Pyranone27		5.7696	5.41411	0.35549
Pyranone28		5.0458	5.32983	-0.28403
Pyranone29		5.8239	5.75314	0.07076
Pyranone30		5.7959	5.62429	0.17161

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Pyranone31		5.9208	5.75294	0.16786
Pyranone32		5.4318	5.35024	0.08156
Pyranone33		6.3010	6.2318	0.0692
Pyranone34		6.0969	6.14192	-0.04502
Pyranone35		5.8539	6.21898	-0.36508
Pyranone36		4.8239	5.29916	-0.47526
Pyranone37		5.6021	5.31259	0.28951

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Pyranone38		6.2218	6.21149	0.01031
Pyranone39		5.7447	5.5728	0.1719
Pyranone40		4.6990	4.82889	-0.12989
Pyranone41		5.8861	5.73028	0.15582

II.2.1.3. Validation du modèle

Le modèle QSAR obtenu a été validé par une validation interne, obtenue par le calcul du coefficient de corrélation croisée (Q^2), en utilisant la LMO (Leave-Many-Out) avec un $Q^2=0.844$ et $RMSE=0.185$.

La valeur indique que notre modèle est stable et peut être efficacement utilisée pour estimer l'activité anti-VIH d'autres molécules pour lesquels les données expérimentales ne sont pas disponibles.

II.2.2. Les modèles 3D-QSAR

Le terme 3D-QSAR décrit le développement d'un modèle de régression mathématique entre les caractéristiques structurales 3D des molécules (la variable indépendante notée X) et leurs valeurs d'activité biologique correspondantes (variable dépendante notée Y), par analyse partielle des moindres carrés (PLS). Le modèle 3D-QSAR généré pourrait être utilisé pour extraire les caractéristiques structurales vitales des molécules nécessaires pour améliorer leurs activités biologiques et pour la prédiction de l'activité des molécules nouvellement conçues[9-11].

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Les outils tridimensionnels (3D) sont aujourd'hui des méthodes très employées pour les études QSARs. Elles permettent de comparer les molécules à travers leur aspect tridimensionnel. Les comparaisons se font sur des critères stériques, électrostatiques, phénomènes responsables des interactions dans les récepteurs [12-13].

II.2.2.1. Base de données et préparation des structures

Les structures chimiques 2D ont été converties en structures 3D de faible énergie en utilisant le module Ligprep (Ligprep, 2.5) du logiciels Schrödinger. Les structures ont été optimisées, les états d'ionisation (pH 7,0 \pm 2,0) ont été calculés.

II.2.2.2. L'alignement des molécules

Les activités biologiques Log EC₅₀des41 composés contre VIH ont constitué l'ensemble de données pour construire les modèles 3D-QSAR. En général, de bons alignements donnent de bons modèles QSAR. Nous avons utilisé comme référence la conformation énergétique la plus faible du composé le plus actif. Les autres composés ont été initialement alignés ; en utilisant le module Flexible Ligand Alignment (A)(Figure.3).

Dans l'alignement flexible des ligands (A), les coordonnées internes sont générées par une recherche conformationnelle et les noyaux sont alignés de manière rigide. Cet alignement n'a pas permis d'obtenir un bon modèle QSAR.

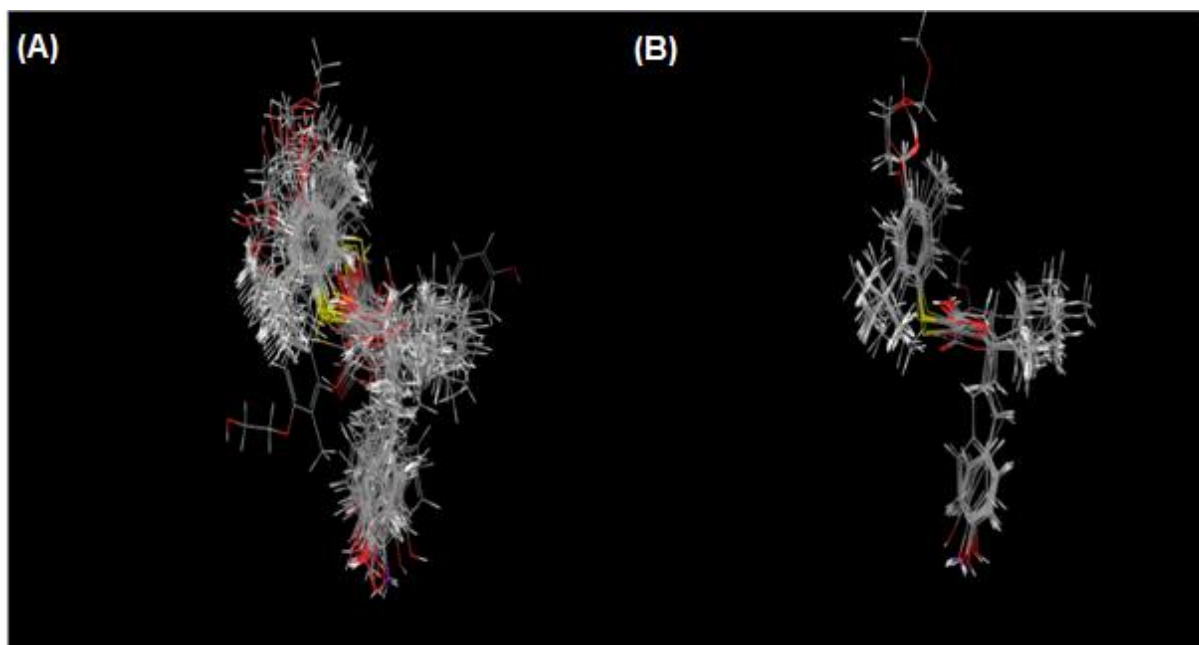


Figure.3 : L'ensemble de molécules alignées à l'aide de : (A) alignement de ligand flexible
(B) alignement d'échafaudage commun

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Donc, les molécules étaient à nouveau alignées en utilisant un alignement d'échafaudage commun (B) (Common scaffold alignment (Maximum Common Substructure)) pour s'assurer que toutes les molécules ont des coordonnées identiques du noyau commun. Dans l'alignement d'échafaudage commun, l'échafaudage commun maximal présent dans toutes les molécules est défini et les molécules sont alignées en conséquence.

Cela minimise les écarts provenant de parties analogues de molécules et fournit des coordonnées identiques pour des parties identiques de la molécule. Il ressort de l'analyse visuelle qu'un protocole d'alignement d'échafaudage commun génère un alignement propre.

II.2.2.3. Construction et validation du modèle

Les QSARs(3D field-based) est similaire à l'analyse comparative de champ moléculaire (CoMFA) et à l'analyse comparative des indices de similarité moléculaire (CoMSIA), avec de légères modifications [14-15]. Les 3D-QSAR utilise le champ de force : the Optimized Potentials for Liquid Simulations (OPLS) qui est un champ de force populaire pour la simulation des molécules organiques et des peptides en phase liquide [16], tandis que le CoMFA et le CoMSIA utilisent un champ de force tripos (qui est un champ de force de mécanique moléculaire implémenté dans le programme Sybyl)[17-18].

Un total de cinq champs gaussiens a été généré pour tous les molécules comprenant un accepteur de liaison H (HBA), stérique, électrostatique, hydrophobique et un donneur de liaison H (HBD). Ces champs gaussiens ont été utilisés pour générer des modèles QSARs.

Certaines molécules ont été choisies comme ensemble de test. Les autres, près de 75 % des molécules, ont été conservées pour l'ensemble d'apprentissage. Cela a assuré l'incorporation correcte de divers composés à la fois dans la formation et dans l'ensemble d'essai.

Les valeurs de l'activité ont été considérées comme des variables dépendantes, tandis que les intensités de champ gaussiennes ont été considérées comme des variables indépendantes.

Un réseau cubique 3D a été construit avec un espacement de grille de 1 Å pouvant s'étendre jusqu'à 3 Å au-delà des limites des molécules d'ensemble d'entraînement. Les seuils d'énergie pour la troncature des champs de force stérique et des champs de force électrostatique ont été réglés sur 30 kcal/mol chacun. Les variables dont l'écart type est inférieur à 0,01 ont été éliminées.

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

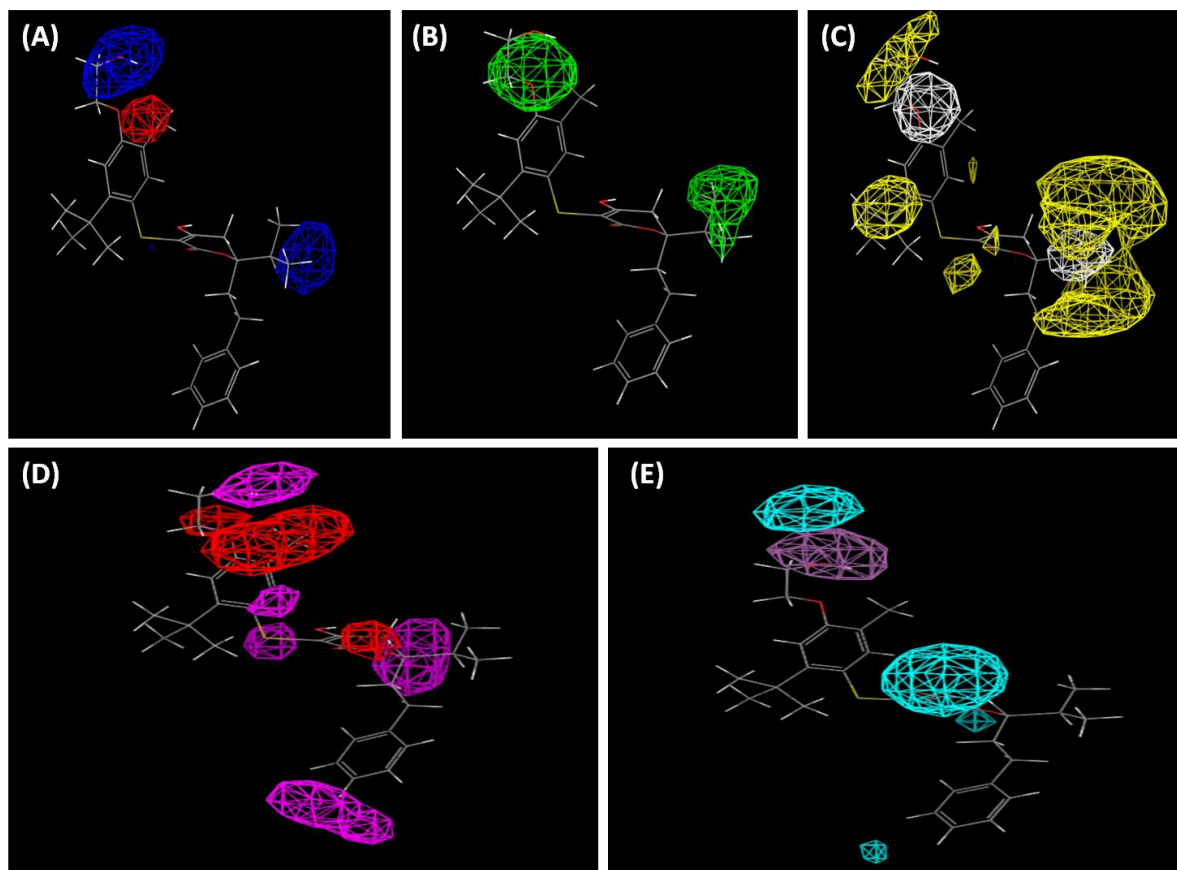


Figure.4 : Les distributions gaussiennes : (A) électrostatique, (B) stérique, (C) hydrophobique, (D) accepteur de liaison H et (E) donneur de liaison H.

Dans le champ électrostatique, les régions en bleu sont favorables au groupement électropositif et les régions en rouge sont favorables au groupement électronégatif.

La région électronégative au niveau des substituants indique que ces substituants renforcent l'activité s'ils sont riches en électrons. Le groupement phényle et les atomes électronégatifs comme l'oxygène, nitrogène sont bénéfiques dans cette région.

Dans le champ stérique, les régions en vert sont favorables à l'encombrement et doivent être occupées.

La Figure.5 ci-dessous montre la courbe de corrélation obtenue par PLS entre les valeurs calculées par rapport aux les valeurs observées de l'activité inhibitrice Log EC₅₀ des pyranones.

La courbe obtenue pour le modèle présente un coefficient de corrélation R² (0.806) de valeur significative. Il indique que ce modèle, peut être appliqué avec succès pour prédire l'activité anti-VIH.

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

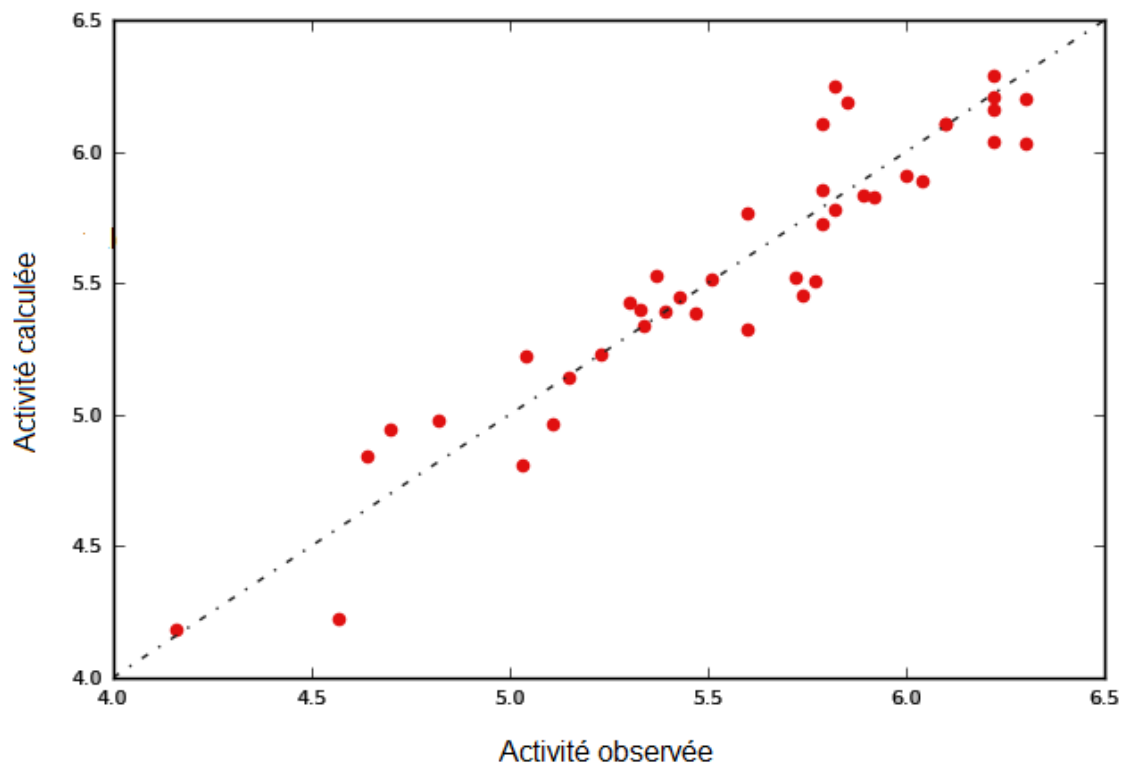
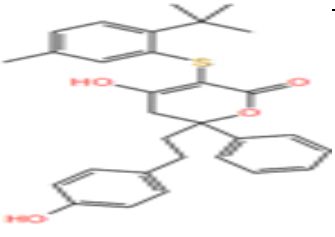
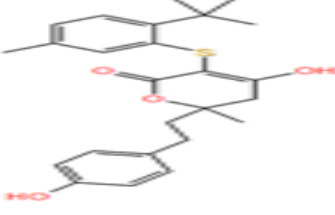
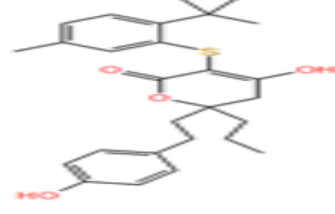
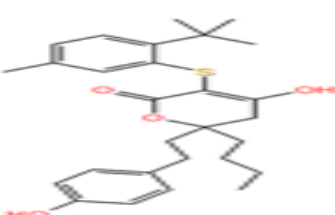
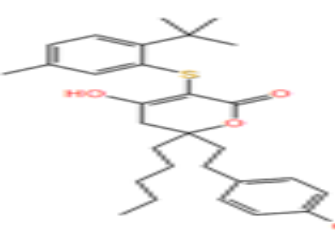
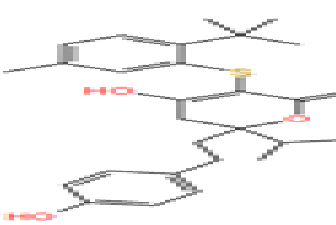


Figure.5: La corrélation linéaire entre les valeurs observées et les valeurs calculées pour la 3D-QSAR

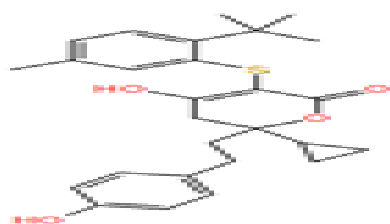
Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Tableau.3 : Les activités observées, les valeurs calculées et le résiduel pour la 3D-QSAR.

Molécule	Structure	Log EC ₅₀ Observée	Log EC ₅₀ Calculée	Résiduel
Pyranone 1		4.1612	5.66095	-1.49975
Pyranone 2		5.0269	4.94735	0.07955
Pyranone 3		5.1079	5.43045	-0.32255
Pyranone 4		4.6383	6.40225	-1.76395
Pyranone 5		4.5686	5.28666	-0.71806
Pyranone 6		5.7212	5.1431	0.5781

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Pyranone 7

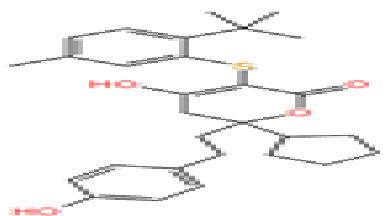


5.3010

6.01758

-0.71658

Pyranone 8

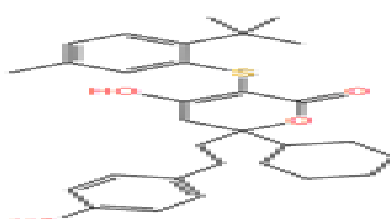


5.2291

5.90196

-0.67286

Pyranone 9

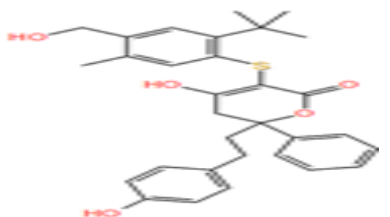


5.3872

6.06823

-0.68103

Pyranone10

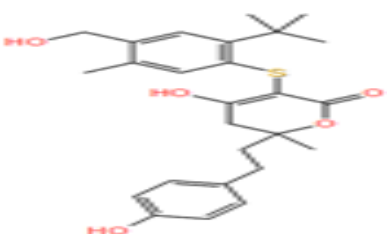


5.3279

5.39721

-0.06931

Pyranone11

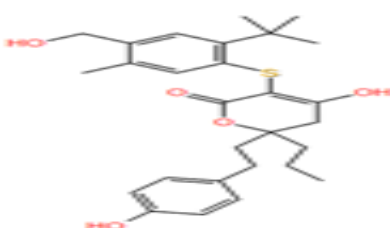


5.6021

5.82241

-0.22031

Pyranone12

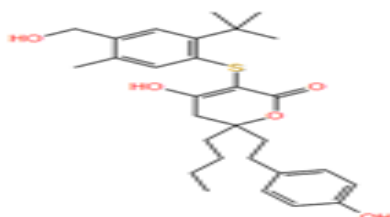


6.0000

5.76954

0.23046

Pyranone13

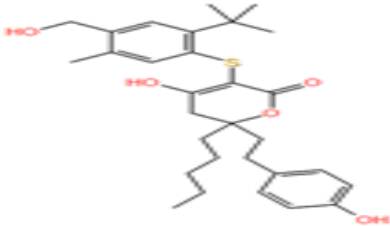
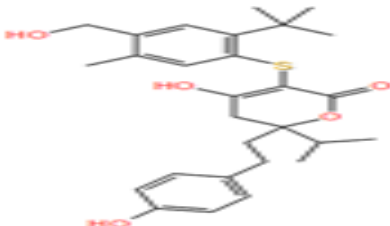
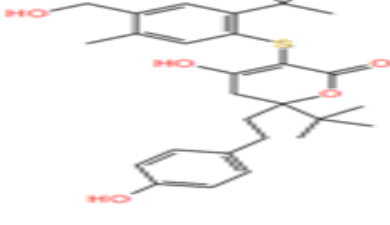
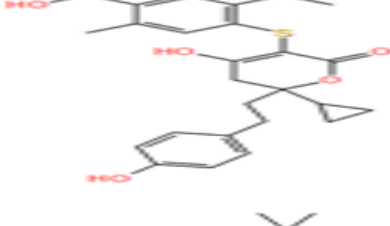
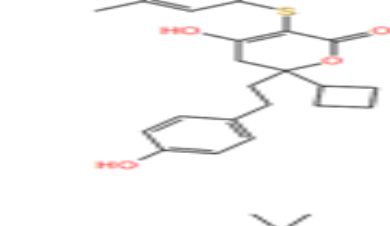
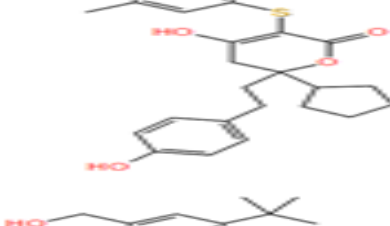
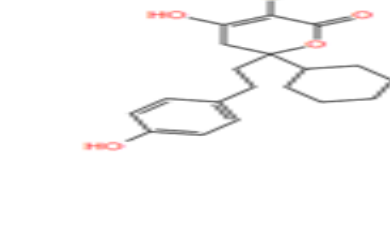


6.0458

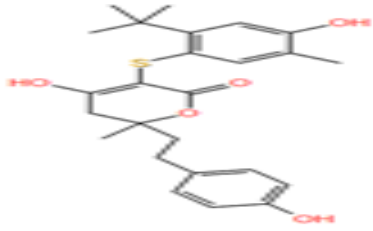
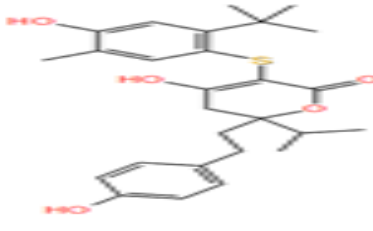
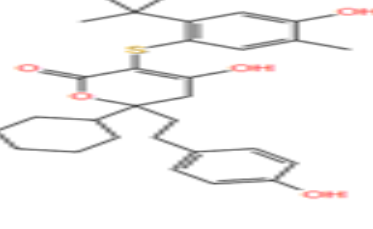
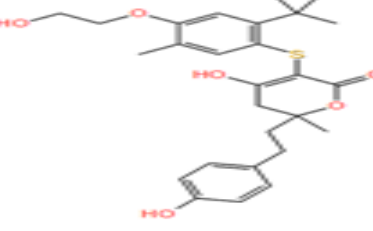
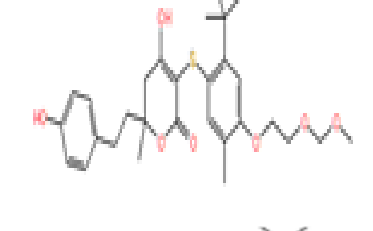
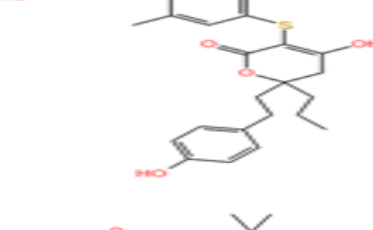
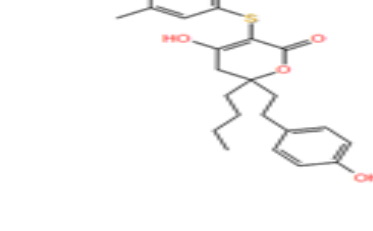
5.7981

0.2477

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

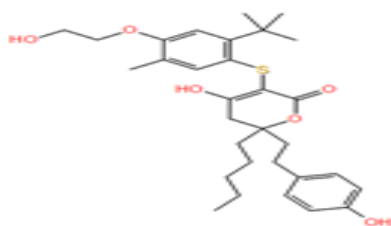
Pyranone14		6.2218	5.43653	0.78527
Pyranone15		6.2218	5.49721	0.72459
Pyranone16		5.8239	5.49645	0.32745
Pyranone17		5.7959	5.42613	0.36977
Pyranone18		6.0969	5.43386	0.66304
Pyranone19		6.2218	5.55953	0.66227
Pyranone20		6.3010	5.90311	0.39789

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Pyranone21		5.1549	5.24594	-0.09104
Pyranone22		5.7959	5.84827	-0.05237
Pyranone23		5.5086	6.16172	-0.65312
Pyranone24		5.4685	5.99325	-0.52475
Pyranone25		5.3372	6.10885	-0.77165
Pyranone26		5.3665	6.10499	-0.73849
Pyranone27		5.7696	6.13741	-0.36781

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Pyranone28

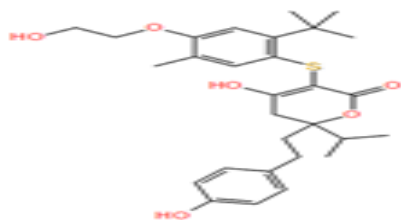


5.0458

5.9521

-0.9063

Pyranone29

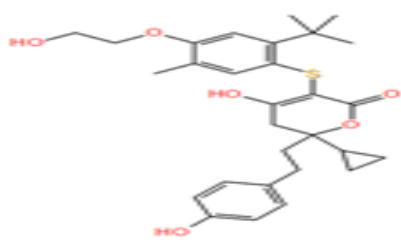


5.8239

5.83652

-0.01262

Pyranone30

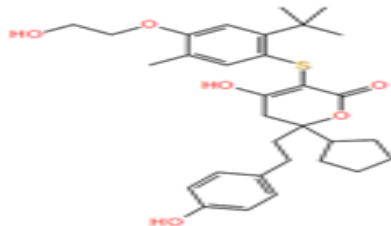


5.7959

5.83576

-0.03986

Pyranone31

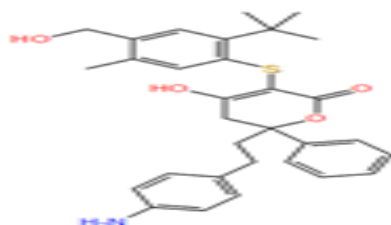


5.9208

5.77317

0.14763

Pyranone32

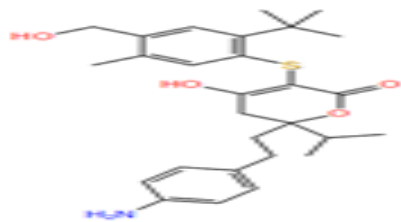


5.4318

5.46639

-0.03459

Pyranone33

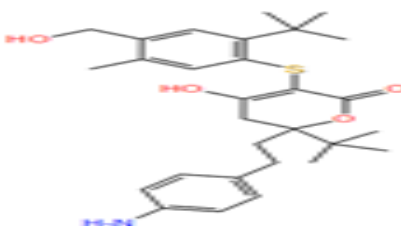


6.3010

5.69581

0.60519

Pyranone34



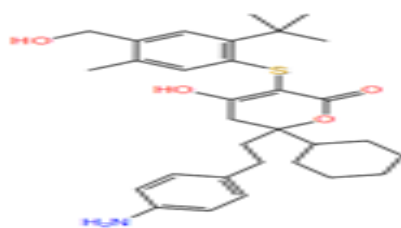
6.0969

5.21742

0.87948

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Pyranone35

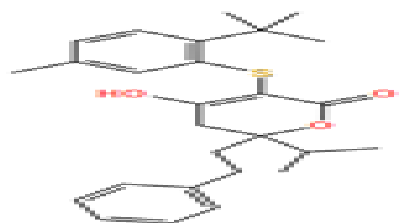


5.8539

5.46266

0.39124

Pyranone36

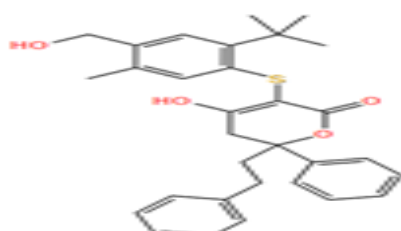


4.8239

5.47201

-0.64811

Pyranone37

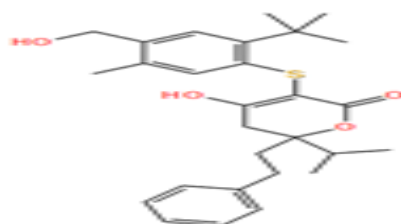


5.6021

4.55245

1.04965

Pyranone38

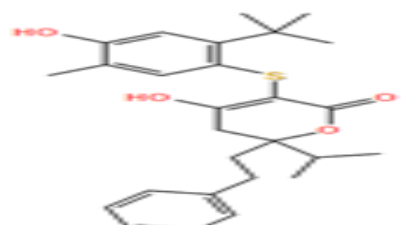


6.2218

4.3071

1.9147

Pyranone39

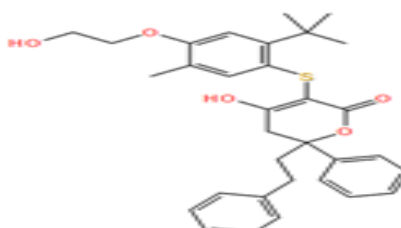


5.7447

5.12628

0.61842

Pyranone40

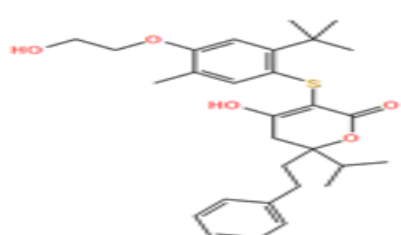


4.6990

4.91071

-0.21171

Pyranone41



5.8861

4.50571

1.38039

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

II.2.2.4. Développement du modèle

À l'aide de l'ensemble aligné de molécules, des modèles 3D-QSAR basés sur Gauss ont été développés à l'aide de l'outil (field-based QSAR) de Maestro (Schrödinger) [19].

Le meilleur modèle a montré un coefficient de régression R^2 de $0.806 > 0,6$ et $SD = 0.189$. La corrélation prévue entre les caractéristiques structurales gaussiennes (champs électrostatique, stérique, HBA, HBD et hydrophobique) est présentée sous forme de cartes de contour (Figure.4).

Les contributions des champs stériques, hydrophobiques, HBA, électrostatique et HBD étaient de 0.431, 0.266, 0.1187, 0.064 et 0.1182 respectivement.

Les champs stériques et hydrophobiques ont contribué le plus, alors que le HBD, HBA ont presque une contribution identique. Pour ce qui est électrostatiques, il a la contribution la plus petite. Cela indique que les caractéristiques stériques et hydrophobiques sont essentielles à l'activité étudiée.

II.2.2.5. Validation

Pour la validation du modèle, on a utilisé la méthode de validation croisée, qui est une méthode d'estimation de fiabilité d'un modèle fondé sur une technique d'échantillonnage. La Leave-One-Out (LOO) a été utilisé pour cette proposition dans laquelle un composé est retiré du jeu de données et reconstruit le modèle, le coefficient $Q^2 = 0.904 > 0.5$ et $RMSE = 0.15$ cela indique le meilleur modèle PLS prédictif.

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

III. Conclusion

En résumé, l'étude 2D-QSAR montre que les descripteurs physico-chimiques et topologiques, sont utiles pour la prédiction de l'activité inhibitrice des dérivés de Pyranone.

Le meilleur modèle 2D-QSAR est capable de décrire environ 81% de la variance de l'activité expérimentale et pourrait être utilisé efficacement pour estimer l'activité des dérivés de Pyranone pour lesquels les données expérimentales sont indisponibles.

En outre, pour les études 3D-QSAR réalisées sur 41 dérivés de Pyranone ont été initiées dans le but d'optimiser une série de ces dérivés, qui agissent comme agents perturbateurs de la protéase. 75% des molécules ont été utilisées dans l'ensemble d'apprentissage et 25% molécules ont été incluses dans l'ensemble de test. L'alignement des molécules a été réalisé en utilisant la méthode d'alignement d'échafaudage commun où l'échafaudage commun maximum présent dans toutes les molécules est défini et les molécules sont alignées en conséquence.

Le meilleur modèle QSAR généré a montré un coefficient de régression (R^2) de 0.806. Le coefficient de régression des molécules de l'ensemble de test (Q^2) s'est avéré être de 0.90 avec un RMSE de 0.15.

Les études ont indiqué que les facteurs stériques et hydrophobiques jouent un rôle majeur dans l'activité anti-VIH. Le modèle 3D-QSAR ainsi développé peut être utilisé pour prédire les activités anti-VIH des composés avant d'entreprendre une synthèse coûteuse, fastidieuse et chronophage.

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

Bibliographie

- [1]. Hansch C, Maloney PP, Fujita T, Muir RM. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature*. 1962;194(4824):178.
- [2]. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin, II, Cronin M, et al. QSAR modeling: where have you been? Where are you going to? *J Med Chem*. 2014;57(12):4977-5010.
- [3]. Nicholls A, MacCuish NE, MacCuish JD. Variable selection and model validation of 2D and 3D molecular descriptors. *Journal of computer-aided molecular design*. 2004;18(7-9):451-74.
- [4]. Hansch C, Fujita T. ρ - σ - π Analysis. A method for the correlation of biological activity and chemical structure. *Journal of the American Chemical Society*. 1964;86(8):1616-26.
- [5]. Hansch C. Quantitative approach to biochemical structure-activity relationships. *Accounts of Chemical Research*. 1969;2(8):232-9.
- [6]. Grassy G, Calas B, Yasri A, Lahana R, Woo J, Iyer S, et al. Computer-assisted rational design of immunosuppressive compounds. *Nature biotechnology*. 1998;16(8):748.
- [7]. Prasad JV, Boyer FE, Domagala JM, Ellsworth EL, Gajda C, Hamilton HW, Hagen SE, Markoski LJ, Steinbaugh BA, Tait BD, Humblet C, Lunney EA, Pavlovsky A, Rubin JR, Ferguson D, Graham N, Holler T, Hupe D, Nouhan C, Tummino PJ, Urumov A, Zeikus E, Zeikus G, Gracheck SJ, Erickson JW, et al. Nonpeptidic HIV protease inhibitors possessing excellent antiviral activities and therapeutic indices. PD 178390: a lead HIV protease inhibitor. *Bioorg Med Chem*. 1999 Dec;7(12):2775-800.
- [8]. Canvas Version 3.2.013, MMshare Version 3.8.013 Legal Notices Schrodinger Software Modules Copyright (c) Schrodinger, LLC. All rights reserved.
- [9]. PHASE, version 3.4, Schrödinger, LLC, New York, NY.
- [10]. S.L. Dixon, A.M. Smondyrev, E.H. Knoll, S.N. Rao, D.E. Shaw, R.A. Friesner, PHASE: a new engine for pharmacophore perception, 3D QSAR model development, And 3D database screening 1, Methodology and preliminary results, *J, Comput Aided Mol,Des*, 20 (2006) 647–671.
- [11]. O. Silakari, S. Chand, M.S. Bahia, Structural Basis of Amino Pyrimidine Derivatives For Inhibitory Activity of PKC- θ : 3D- QSAR and Molecular Docking Studies, *Mol. Inform*. 31 (2012) 659–668.
- [12]. Simon, Z.; Badileuscu, I.; Racovitan, T. *J. Theor. Biol.* 197, 66, 485.
- [13]. Simon, Z.; Dragomir, N.; Plauchithiu, M.G.; Holban, S.; Glatt, H.; Kerek, F. *Eur. J. Med. Chem.* 1980, 15, 521.
- [14]. Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc*. 1988 Aug 1;110(18):5959-67.
- [15]. Klebe G, Abraham U, Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem*. 1994 Nov 25;37(24):4130-46.

Chapitre. III : Etude de l'inhibition de la protéase du VIH par QSAR

- [16]. S.W.I. Siu, K. Pluhackova, R.A. Böckmann, Optimization of the OPLS-AA Force Field for Long Hydrocarbons, *J. Chem. Theory Comput.* 8 (2012) 1459–1470.
- [17]. Jorgensen WL, Tirado-Rives J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc.* 1988 Mar 1;110(6):1657-66.
- [18]. Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W. Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J Chem Theory Comput.* 2010 May 11;6(5):1509-19.
- [19]. Maestro Version 11.2.013, MMshare Version 3.8.013, Release 2017-2.

Conclusion générale

Conclusion générale

Le travail a été consacré au développement de modèles QSARs pour la prédiction de l'activité anti-VIH des molécules des dérivés de pyranone, pour construire de modèle du QSAR fiable, stable, capable de prédire efficacement cette activité.

Les QSARs se trouvent à l'intersection de la chimie, les statistiques et la biologie, pour cette raison les deux premiers chapitres de ce mémoire ont été consacrés à la présentation des dérivés de pyranone pour l'inhibition de la protéase du VIH, la méthodologie QSAR ainsi les outils nécessaires pour la mise en place d'un modèle QSAR.

Des études de la relation quantitative structure-activité (2D et 3D QSAR) sont appliquées à un ensemble de quarante-et-un composé des dérivés de pyranone, afin de prédire l'activité biologique à partir de descripteurs moléculaires appartenant à la série de dérivés du pyranone et de trouver une corrélation entre les différents paramètres moléculaires de ces composés et son activité biologique. La méthodologie basée sur la régression multilinéaire (MLR), a été utilisée pour la génération des 2D-QSAR et la PLS pour la 3D-QSAR.

Le pouvoir prédictif des modèles obtenus a été confirmé par les méthodes de validation Croisée LMO pour la 2D-QSAR et LOO pour la 3D-QSAR. Une forte corrélation a été observée entre les valeurs expérimentales et prédites des activités biologiques, ce qui indique la validité et la qualité des modèles QSARs obtenus.

Les valeurs des paramètres statistiques du modèle 2D-QSAR: $R^2 = 0.803$ et $Q^2 = 0.844$ et les valeurs des paramètres statistiques du modèle 3D-QSAR : $R^2 = 0.806$ et $Q^2 = 0.904$, montrent que ce dernier est statistiquement significatif possédant un bon pouvoir prédictif et une bonne stabilité.

Les modèles QSARs sont considérés comme des outils efficaces pour prédire l'activité inhibitrice de la protéase du VIH sur la base des descripteurs proposés.