



Université Mohamed Khider de Biskra
Faculté des Sciences Exactes et des Sciences de la Nature
et de la Vie

Département des Sciences de la Matière

MÉMOIRE DE MASTER

Domaine des Sciences de la Matière

Filière de Chimie

Spécialité Chimie Pharmaceutique

Présenté et soutenu par :

Benyaya Roumaïssa

Etude QSAR des nouveaux dérivés de flavonoïde comme HIV- Inhibiteurs

Jury

Dr. Khamouli Saida	Université Med Khider Biskra	Président
Dr. Mellaoui Malika	Université Med Khider Biskra	Rapporteur
Dr. Aggoun Siham	Université Med Khider Biskra	Examineur

Année universitaire : 2022-2023



الحمد لله الذي كتب لي الوصول لهذا المقام أحمد الله الكريم الذي منى علي
وأعانني على إتمام هذا البحث

أهدي عملي هذا الى الذي وهبني كل ما يملك للوصول لبر الأمان لأحقق أحلامه قبل أن
تكون آمالي الى الذي قدس العلم، سهر الليالي وقاسى الجراح لأجلنا، لمن أراد لنا العلى
أبي الغالي سندي وحببي الأول أطل الله في عمرك

الى نبع الحنان والعطاء الى صاحبة القلب الطيب من رعيتي من نعومة أظفاري الى
يومنا هذا حق رعاية ، الى من كانت دافعي الأول للعلم والسعي للعلي ملجئي الأول **نبع**
الحنان أمي أجمل من رأت عيوني وأحب قلبي

جزاكما الله عني خير الجزاء في الدارين

كما أهدى ثمرة جهدي الى من رافقتني في هذا العمل المتواضع التي لجأت إليها كلما
أضعت طريقي الأستاذة الفاضلة "**ملوي مليكة**" شكرا على كل دقيقة من وقتك
وهبتها لي بالرغم من تعدد مسؤولياتك

أهدي تخرجي الى من كانوا رفاقي طيلة هذه السنوات الى من تمنوا لي النجاح والتوفيق
أخواتي حبيباتي وإخواني الأعزاء على قلبي أهدى لهم ثمرة نجاحي

ولكل من تمنى لي النجاح والتوفيق في دراستي من عائلتي الكريمة وصديقات

دربي لكل من كانوا جزءا لمسح دموعي وتخفيف آلامي وفقكم الله الى ما تحبون



رئيسة لجنة إدارة



Remerciements

Tout d'abord, remercier le Tout-Puissant "Allah" pour son existence

Étant donné la force, la volonté et le courage d'entreprendre ce travail. Je remercie, tout d'abord, mon encadreur M^{me} Mellaoui Malika pour ses efforts, ses conseils successifs, ses indulgences ; les suivre intégralement pour ce travail. Je tiens également à remercier les membres du jury principal M^{me} Khamouli Saida et l'examinatrice M^{me} Aggoun Siham. D'accepter de juger mon humble travail

Je voudrais remercier tous mes professeurs du Département de Chimie sans exception qui ont donné notre enseignement, enfin, Dieu merci, je termine ce travail, C'est le résultat d'efforts et de fatigue.

A Tous, merci

SOMMAIRE

Liste de tableaux.....	(I)
Liste des figures.....	(II)
Liste des abréviations.....	(III)
Introduction Générale	(1)
Chapitre I: Généralité sur le VIH et FLAVONOIDE	
I.1 Introduction	(4)
I.2 Généralité sur le VIH /SIDA	(5)
I.2.1 Définition.....	(5)
I.2.2 Structure du VIH	(6)
I.2.3 Epidémiologie:.....	(6)
I.2.4 Mode de transmission du virus.....	(8)
I.2.4 A Transmission par voie sexuelle.....	(8)
I.2.4 B Transmission par le sang	(8)
I.2.4 C Transmission materno-fœtale.....	(8)
I.2.5 Le cycle viral du VIH	(9)
I.2.6 Diagnostic biologique du VIH	(10)
I.2.7 Traitement du VIH.....	(11).
I.3 LES FLAVONOIDE	
I.3.1 Les polyphénols	(12)
I.3.1.1 Définition	(12)
I.3.1.2 Classification.....	(12)
I.3.1.3 Biosynthèse des composés phénoliques.....	(12)
I.3.1.4 Effets biologiques des polyphénols.....	(12)
I.3.2 Les flavonoïdes.....	(14)
I.3.2.1 Définition	(14)
I.3.2.2 Etymologie	(15)
I.3.2.3 Structure chimique.....	(15)
I.3.2.4 Classification.....	(15)
I.3.2.5 Propriétés biologiques	(16)

a) Effets protecteurs vasculaires.....	(16)
b) Propriétés antiallergiques.....	(17)
c) Propriété anti-inflammatoire.....	(17)
d) Propriété anti-ulcérogène.....	(17)
e) Propriété antibactérienne.....	(17)
f) Activités antivirales.....	(17)
g) Activités anti-tumorales.....	(18)
I.4. TRAITEMENT ANTIRÉTROVIRAL.....	(18)
Référence.....	(19)

Chapitre II: QSAR : principe et méthodologie

II.1 Introduction.....	(22)
II.2 Historique.....	(23)
II.3 Définition.....	(24)
II.4 Principe.....	(24)
II.5 Outils et Méthodologie de QSAR.....	(25)
II.5.1 Les bases de données.....	(25)
II.5.2 Descripteurs moléculaires.....	(26)
II.5.2.1 Types de descripteurs.....	(26)
II.5.2.1.a Descripteurs constitutionnels.....	(26)
II.5.2.1.b Descripteurs topologiques.....	(27)
II.5.2.1.c Descripteurs géométriques.....	(28)
II.5.2.1.d Descripteurs quantiques.....	(28)
II.5.2.1.e Descripteurs physico-chimiques.....	(29)
II.6 Méthodes d'analyse de données.....	(30)
II.6.1 Régression linéaire simple (SLR).....	(30)
II.6.2 La régression linéaire multiple (MLR).....	(31)
II.7 Coefficients et tests statistiques standards.....	(32)
II.8 Validation du modèle.....	(34)
II.8.1 Validation interne.....	(34)

II.8.1.a La procédure leave-n-out.....	(34)
II.8.1.b La procédure leave-one-out.....	(35)
II.8.2 Validation externe.....	(36)
II.9 Application de QSAR.....	(37)
Référence.....	(39)

Chapitre III: Etude QSAR D'une Série De Dérivés De Flavonoïdes

III.1 Introduction.....	(43)
III.2 Matériels et méthodes.....	(44)
III.2.1 Base de donnée.....	(45)
III.2.2 Sélection des descripteurs et méthodes de calcul.....	(51)
III.2.3 Analyse statistique.....	(54)
III.3 Résultats et discussions.....	(55)
III.3.1 Interprétation du modèle QSAR.....	(55)
III.3.2 Validation des modèles.....	(57)
III.3.2.1 Validation interne.....	(57)
III.3.2.2 Validation externe.....	(60)
Références.....	(62)

Conclusion générale

Résumé

Abstrac

الملخص

Liste des tableaux

Chapitre I :

Tableau I.1: Médicaments Antirétroviraux disponibles en Suisse en 2006.....(18)

Chapitre II :

Tableau II. 1: Les descripteurs topologiques calculés dans étude QSAR.....(28)

Chapitre III :

Tableau III.1: les structures chimiques et l'activités biologiques de dérivés flavonoïdes en tant qu'inhibiteurs du VIH -1.....(46)

Tableau III.2: les descripteurs obtenu 1.....(52)

Tableau III.3: les descripteurs obtenu 2.....(53)

Tableau III.4: Modèles sélectionnés et paramètres statistiques des corrélations entre les propriétés moléculaires et l'activité biologique antivirale.....(55)

Tableau III.5: Paramètres de validation croisée.....(57)

Tableau III.6: Valeurs Expérimentales , Prédites Et Résiduelles e dep-IC50 Des dérivées des flavonoïdes ; état des molécules dans la validation(58)

Tableau III.7: Critères de Tropsha.....(60)

Liste de Figure

Chapitre I

Figure . I 1: Quelques symptômes de l'infection de VIH	(5)
Figure. I 2 : structure de virus de sida / VIH.....	(6)
Figure. I 3 : Mode de transmission du virus	(9)
Figure. I 4: Le Cycle du VIH	(10)
Figure. I 5: Effets biologiques des polyphénols.....	(13)
Figure. I 6 : Certaines plantes contiennent des flavonoïde	(14)
Figure. I 7: schéma de base des flavonoïdes.....	(15)
Figure . I 8: Structures des différentes classes de flavonoïdes	(16)

Chapitre II

Figure II .1: Modèle de l'étude de relation structure activité/propriété.....	(24)
Figure II.2: principe de la méthode QSPR	(25)
Figure II. 3: principe de la validation croisée leave-n-out.	(35)

Chapitre III

Figure III. 1 : les interfaces des programmes utilisées.	(44)
Figure III.2 : Dérivés de flavonoïdes en tant qu'inhibiteurs du VIH-1	(45)
Figure III .3: Diagramme de la contribution en pourcentage de chaque descripteur dans le modèle pIC50 développé expliquant la variation de l'activité.	(59)
Figure III .4: Corrélation Entre L'activité biologique expérimentale et prédite.....	(60)

Liste des Abréviations

- 3D, 2D : Trois dimensions, deux dimensions, une dimension
- Å : Angstrom
- ADN: Acide Désoxyribonucléique.
- ARN : Acide Ribonucléique
- ARV : Antirétroviraux
- CD4 : Cluster de différenciation 4
- CD8 : Cluster de différenciation 8
- CMH : Complexe Majeur d'Histocompatibilité
- HOMO: Orbitale Moléculaire plus Haut Occupée.
- Log P : Coefficient de partage.
- LOO : Leave One Out .
- LUMO: Orbitale Moléculaire plus bas Occupée .
- MD : moment dipolaire
- MLR : Régression linéaire multiple .
- P-IC50 : Inhibiteur concentration of 50% .
- Pol : polarisabilité
- PRESS: Somme des carrés des erreurs de prédiction .
- QSAR : Relations quantitatives structure- activités .
- QSPR : Relations quantitatives structure-propriétés .
- R : Coefficient de Corrélation.
- R² adj : Coefficient de Détermination Ajusté.
- SAG : surface moléculaire
- SIDA : Syndrome de l'Immunodéficience Acquise
- SPRESS: La capacité prédictive des modèles est évaluée par l'erreur de racine carrée moyenne.
- SSY : Somme des carrés des distances des valeurs observées .
- VIH : Virus de l'Immunodéficience Humaine

Introduction générale

Introduction générale

Le syndrome d'immunodéficience acquise (SIDA) est l'une des maladies infectieuses les plus dévastatrices causée principalement par l'infection par le virus de l'immunodéficience humaine (VIH). Il a été rapporté que plus de 13,9 millions de personnes sont décédées depuis l'identification de la maladie en 1984 [2]. Il a été découvert que le VIH cause de nombreuses maladies immunologiques telles que le lymphome de Burkitt, le sarcome de Kaposi, le cancer invasif du col de l'utérus, l'histoplasmosse, etc. [12]. Il se transmet principalement par les fluides corporels tels que le sperme sanguin, les sécrétions vaginales, le lait maternel, la sueur, etc. Bien que plusieurs mesures préventives aient été prises, il n'y a pas suffisamment de médicaments/vaccins pour prévenir la transmission et l'infection du VIH [13]. Malgré le taux de réussite des antirétroviraux tels que l'azidothymidine (AZT), le Nivavir et la Stavudine contre le VIH/SIDA, plusieurs études sur la conception et le développement de médicaments sont encore menées chaque jour. En raison de plusieurs complications dans les études cliniques, de nouveaux médicaments actifs sont actuellement nécessaires [10]. Ainsi, une stratégie définie est nécessaire pour étudier et analyser les composés bioactifs à utiliser comme agents thérapeutiques puissants contre l'infection par le VIH.

Avec l'avènement de la technologie scientifique moderne et de l'industrialisation, l'utilisation de produits chimiques a augmenté avec une augmentation du nombre de nouveaux produits chimiques produits quotidiennement [7]. Une nouvelle méthode qui permet à tout chimiste de mieux comprendre les problèmes liés au domaine moléculaire : la modélisation moléculaire [3][11]. Elle combine des techniques de visualisation, de traitement, d'analyse et de calcul de structure spatiale [1][8]. Il est aussi de plus en plus utilisé aujourd'hui pour étudier une réaction chimique.

La modélisation moléculaire est un terme général qui comprend diverses techniques de graphisme moléculaire et de chimie computationnelle pour afficher, simuler, analyser, calculer et stocker les propriétés physiques et chimiques des molécules. Il travaille à établir des relations entre la structure chimique des composés et leur activité biologique [9].

Les relations entre les structures des molécules et leurs propriétés ou activités sont généralement établies à l'aide de méthodes de modélisation par apprentissage statistique. Les techniques usuelles reposent sur la caractérisation des molécules par un ensemble de descripteurs, les nombres réels mesurés ou calculés de structures moléculaires. Il est alors possible d'établir une relation entre ces descripteurs et la grandeur modélisée [5]

Parmi les techniques de chimie computationnelle, nous pouvons citer les techniques de QSAR (Quantitative Structure-Activity Relationships) qui consiste à trouver une corrélation entre une activité biologique mesurée pour un panel de composés et certains descripteurs moléculaires. Elle permet

également de guider les processus de développement de nouvelles molécules, sans avoir à les synthétiser, ou à analyser des familles entières de composés. Les relations entre les structures des molécules et leurs propriétés ou activités sont généralement établies à l'aide de méthodes de modélisation par apprentissage statistique. Les techniques usuelles reposent sur la caractérisation des molécules par un ensemble de descripteurs ; nombres réels mesurés ou calculés traduisant des propriétés des structures moléculaires. Actuellement, il est possible de générer des milliers de descripteurs, cependant un seul petit nombre peut se révéler pertinent pour modéliser une activité biologique donnée[4][6] ; La prédiction de l'activité biologique est basée sur des calculs mathématiques et statistiques, c'est pour cette raison on met le choix d'une méthode statistique, appelée la régression multilinéaire (MLR) [9].

Étudier l'activité biologique de nouveaux dérivés de flavonoïdes qui inhibent le VIH. Notre objectif dans cette thèse est d'appliquer toutes les méthodes fournies par la chimie computationnelle pour déterminer les propriétés structurales de cette chaîne ; L'effet de la substitution sur ces propriétés ainsi que les associations entre la structure chimique et l'activité biologique sont discutés.

Pour mener à bien ce travail de manière appropriée, nous l'avons divisé en trois chapitres; Nous commençons par une introduction générale et terminons par une conclusion générale.

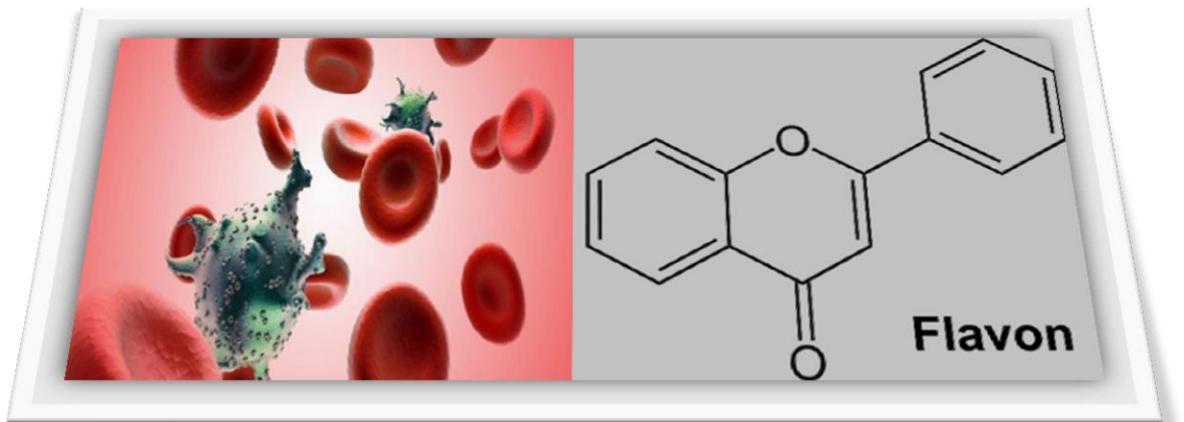
- ✚ Le premier chapitre : Nous introduisons des généralités sur la maladie à VIH plutôt que des généralités sur les flavonoïdes
- ✚ Le deuxième chapitre : contient une présentation des méthodes de calcul utilisées
- ✚ Le troisième chapitre : est consacré aux résultats obtenus et à leur discussion.

Référence :

- [1] A.R. Leach, *Molecular Modelling: Principals and Applications*, Edimbourg : Addison Wesley Longman Ltd,1996, 1-24
- [2] Bojak, A. , Deml, L. , Wagner, R. , 2002. The past, present and future of HIV-vaccine development: a critical view. *Drug Discov. Today* 7 (1), 36–46 .
- [3] E .Derty , j. *Mol. Structur .(theochem)*,1999,459,273
- [4] Götte, M., Rausch, J. W., Marchand , B., Sarafianos, S., & Le Grice, S. F. Reverse transcriptase in motion: conformational dynamics of enzyme–substrate interactions. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1804(5), (2010) ,1202-1212.
- [5] G-S-A. Aurélie, Thèse de doctorat, Université de Paris 6, 2008.
- [6] Hu, W. S., & Hughes, S. H. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* 2: a006882. (2012).
- [7] J.R. Seward, E. Hamblen, T.W. Schultz, Regression comparisons of *Tetrahymena pyriformis* and *Poecilia reticulata* toxicity, *Chemosphere*, 47 (2002) 93-101.
- [8] L. P. Graham, *Chimie pharmaceutique, de boeck*, 2003, 319-334.
- [9] Mehellou Mohammed Nadjib ; Etude des relations structure/activité quantitatives (QSAR/2D) d'une série de dérivés de Triazolothiadiazoles ; Université echahide hamma lakhder d'el-oued; 2018.
- [10] Prokofjeva, M.M. , Kochetkov, S.N. , Prassolov, V.S. , 2016. Therapy of HIV infection: current approaches and prospects. *Acta Naturae* 8 (4), 31 .
- [11] S.Belaidi ,M.Omari ,T.Lanez et A.Dibi ,*J.Soc .Alger.Chim*,2004,14,27.
- [12] Stephenson, K.E. , D' Couto, T.H. , Barouch, D.H. , 2016. New concepts in HIV-1 vaccine development. *Curr. Opin. Immunol.* 41, 39–46 .
- [13] Stolley, K.S; Glass, J.E. , 2009. HIV/AIDS. ABC-CLIO, Californi

Chapitre I

Généralité sur le VIH/ FLAVONOÏDE



I.1. Introduction

La maladie d'immunodéficience acquise est une maladie qui affecte les globules blancs responsables du système immunitaire, et provoque ainsi un défaut du système immunitaire, c'est-à-dire mettant en danger la vie humaine. Il n'existe aucun traitement efficace contre ce virus. Cependant, les médicaments agissent pour contrôler l'infection et prévenir les complications de la maladie. Les antiviraux qui traitent l'immunodéficience se sont avérés efficaces pour réduire les décès dus au sida partout dans le monde. Les organisations internationales s'emploient également à renforcer les mesures de prévention et de traitement et à fournir des médicaments plus efficaces contre ce virus [2][3][5].

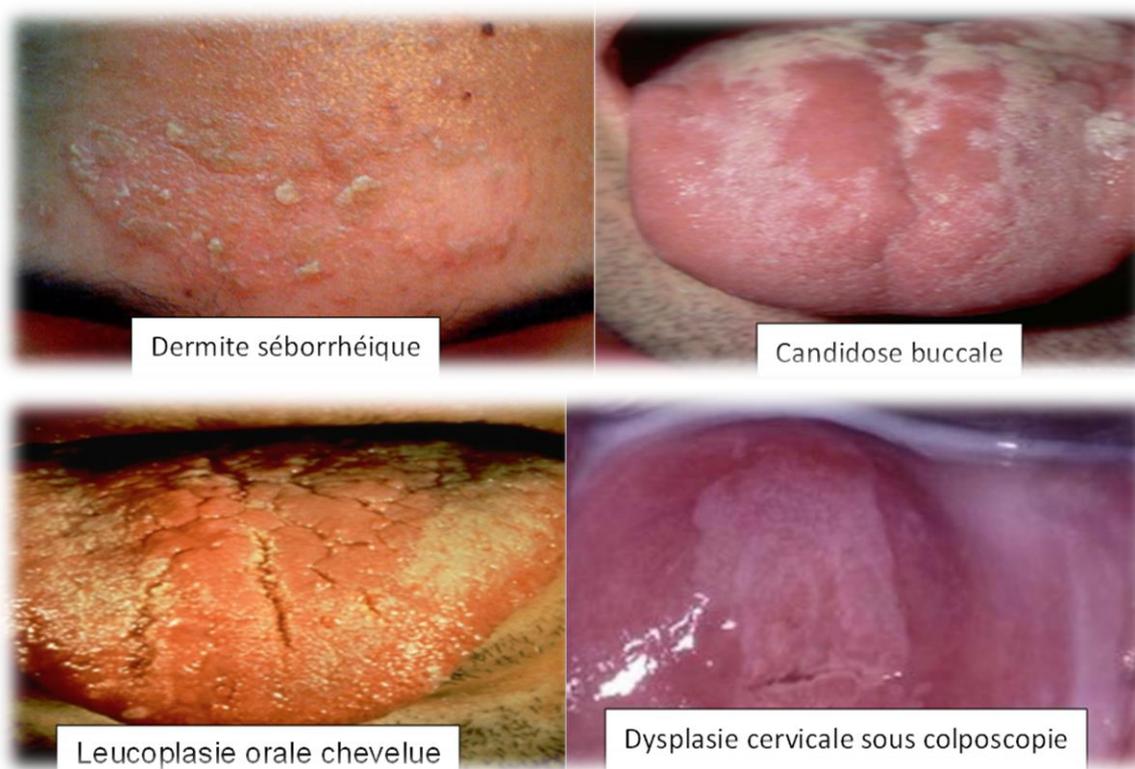
Les flavonoïdes à source unique sont des branches de la famille des polyphénols, un groupe de plus de 6 000 composés naturels qui sont presque universels dans les plantes vasculaires. Ils forment également des pigments responsables de la coloration jaune et rouge de divers organes végétaux.[1][4]

Les flavonoïdes sont les plus largement utilisés dans la production et la fabrication de médicaments pour leur efficacité et leur rôle biologique dans l'élimination des virus et des bactéries. Comme dans le cas du VIH / SIDA, il existe différents médicaments d'origine flavonoïde, tels que zidovudine(AZT), didanosine (ddI)...[7]

I.2. Généralité sur le VIH/ SIDA :

I.2.1. Définition :

C'est une infection sexuellement transmissible (IST) et autre. causée par un virus appelé de l'immunodéficience humaine (VIH) ,ce virus se caractérise par une diminution progressive du système de défense immunitaire et comme tout virus ,c'est un parasite intracellulaire obligatoire spécifique à un type des cellules qui contiennent des récepteurs membranaires CD4, des globules blancs (le cas des lymphocytes T CD4) des macrophages, des cellules dendritiques et même des cellules micro-gliales [2]. La multiplication et la libération des particules virales entraînent la destruction des cellules hôtes. Cette destruction aboutit à une immunodéficience qui favorise l'apparition de maladies opportunistes.



I.2.2. Structure du VIH :

Les virus de VIH mesurent environ 100 nm de diamètre. Il possède une enveloppe lipidique, dans laquelle est noyée la glycoprotéine transmembranaire trimérique gp41 à laquelle est fixée la glycoprotéine de surface gp120 (figure 2). Ces deux protéines virales sont responsables de l'attachement à la cellule hôte et sont codées par le gène env du génome de l'ARN viral, associées à deux enzymes appelées transcriptase inverse (p64)... Sous l'enveloppe, se trouve la protéine p17 et l'autre plus profonde par la protéine p24. Deux autres enzymes telles que la protéase p10 et l'intégrase p32 lui y sont associées également [6].

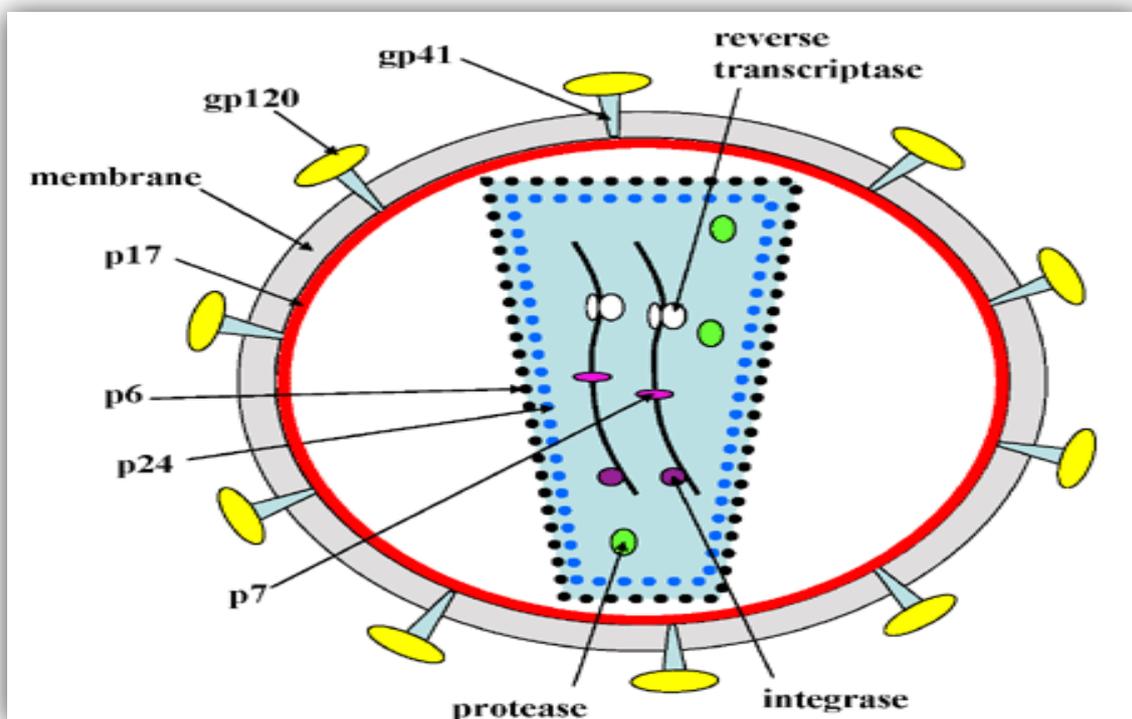


Figure I.2 : structure de virus de sida / VIH

I.2.3. Epidémiologie:

Les femmes infectées par le VIH ont, par rapport aux femmes non contaminées, une prévalence et une incidence plus grandes d'être infectées par le papillomavirus humain :[8]

*Elles présentent une fréquence 2 à 3 fois plus élevée d'infections à papillomavirus humain mais également une durée plus prolongée de l'infection, deux facteurs qui semblent corrélés au degré du déficit immunitaire.

*la prévalence et la persistance de l'infection à HPV augmentent avec la diminution du taux de lymphocytes CD4 et l'augmentation de l'ARN du VIH au niveau plasmatique .

*Deux études ont montré que les sous-types oncogéniques de HPV étaient plus fréquents lorsque les taux de lymphocytes CD4 étaient les plus bas et/ou lorsque la charge virale était plus élevée . Une plus grande charge virale de HPV est également associée avec un taux plus bas de lymphocytes CD4 [9] [3] .

*Une étude réalisée au Brésil sur 208 femmes séropositives pour le VIH a montré que 98 étaient positives pour le HPV par la technique d'amplification des séquences cibles de l'ADN du virus par la réaction en chaîne de la polymérase . L'étude prospective , multi-site , <<Human Immunodeficiency Virus Epidemiology Research Study >> (HERS) conduit pour évaluer l'impact biologique ,psychologique et social de l'infection à VIH sur 871 femmes américaines entre 1993et 1995 comparées à 439 femmes non infectées, âgées de 16 à 55 ans ,a montré que les femmes infectées par le VIH avaient une charge virale de HPV plus importante que les non infectées et une plus grande diversité dans les sous-types ,avec notamment une plus grande prévalence de sous-type oncogénique .le type de HPV est le même chez les patientes infectées ou non par le VIH (chez des patientes ayant les mêmes critères de risque d'infection à VIH) mais il existe un taux plus important de persistance de ces infections chez les patientes VIH positives ,surtout dans le cas d'un ce sous-type viral oncogène .

Une étude canadienne réalisée en 2002 a permis d'identifier les facteurs de risque d'infection au HPV chez les femmes infectées par le VIH :[10]

*un nombre de lymphocytes CD4 de moins de 200 par mm³.

*le stade plus avancé de l'infection à VIH .

*la charge virale de VIH plus élevée.

*la race autre que blanche .

*le nombre élevé de partenaires sexuels depuis la dernière consultation .

*la pratique de rapport sexuel anal réceptif non-protégé.

*l'utilisation inconstante du condom, l'utilisation régulière du condom étant associée à un risque diminué de 70%.

*l'utilisation de contraceptifs oraux .

* L'âge de moins de 30 ans .

I.2.4. Mode de transmission du virus :

Le virus du sida peut être transmis de diverses manières, qui impliquent différents fluides corporels : le sang ou les sécrétions génitales (voir [Les modes de transmission du virus du sida](#))[6].

I.2.4.A. Transmission par voie sexuelle :

Le virus est présent dans les sécrétions génitales, et peut donc être transmis lors d'un rapport sexuel, qu'il soit homosexuel ou hétérosexuel (la majorité des sidéens africains sont ainsi contaminés lors de rapports hétérosexuels). Certaines maladies sexuellement transmissibles, et surtout la multiplication des partenaires (sans protection lors des rapports) favorisent cette transmission. (70 à 80 % des cas d'infection).

I.2.4.B. Transmission par le sang :

Le virus étant présent dans le sang, il peut être transmis lors de tout « don » de sang d'un individu à un autre : lors de pratiques toxicomanes (échanges de seringues), de manière accidentelle, ou lors de transfusions. Un dépistage systématique des dons du sang a permis de réduire ce dernier mode de transmission (risque résiduel estimé à 1/500 000).[5]

I.2.4.C. Transmission materno-fœtale :

Le virus est capable de traverser la barrière hémato-placentaire, et ainsi de contaminer, *in utero*, un fœtus. Le cas le plus fréquent semble être toutefois lors de l'accouchement. De plus, le virus se retrouve dans le lait maternel, d'où une contamination lors de l'allaitement (cas fréquent surtout en Afrique). Sans traitement, le VIH-1 se transmet de la mère au fœtus dans 15 à 20 % des cas (30 % si allaitement). Le VIH-2 ne se transmet lui, que dans 2 % des cas.

Avec traitement préventif, le taux de transmission du VIH-1 baisse à moins de 8 % (moins de 2 % en Europe). Chaque jour, environ 1000 enfants porteurs du VIH naissent en Afrique...

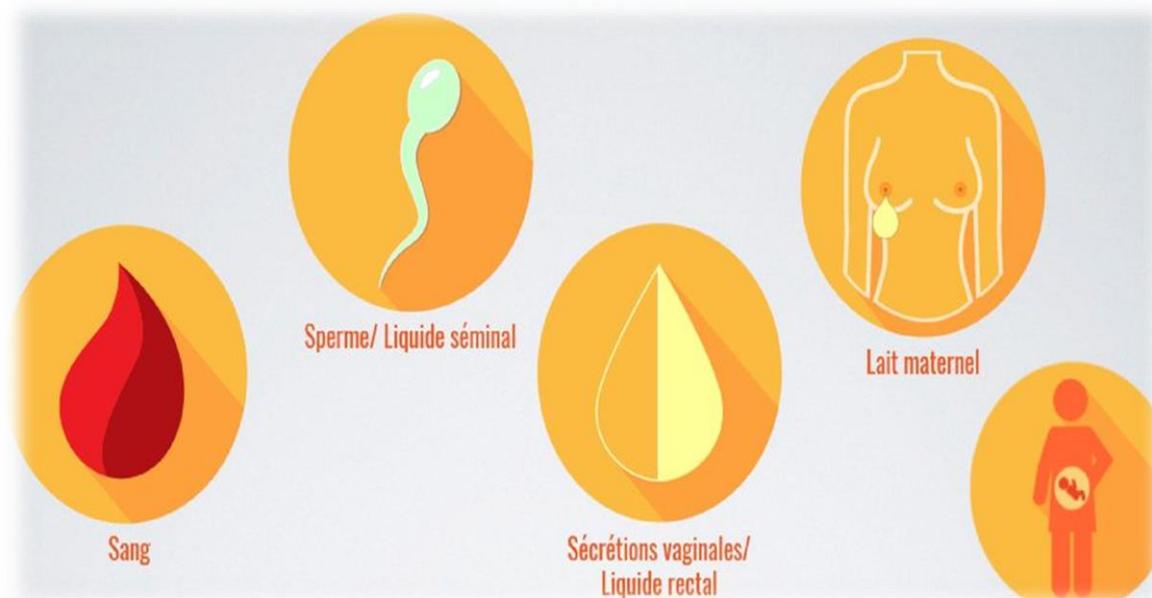


Figure I.3 : Mode de transmission du virus

I.2.5. Le cycle viral du VIH :

Le VIH est un virus qui se multiplie dans les cellules porteuses de la protéine membranaire de surface appelée CD4, essentiellement les lymphocytes T auxiliaires ou $CD4^+$ mais aussi les macrophages par exemple. Le virus interagit avec la cellule via la protéine CD4 grâce à la protéine virale appelée gp120, puis, par l'intermédiaire d'une autre protéine virale appelée gp41, la membrane virale fusionne avec la membrane cellulaire. L'ARN viral est ensuite rétro transcrit en ADN viral puis intégré au génome de la cellule infectée par deux protéines virales présentes dans la particule virale appelées transcriptase inverse et en donucléase. L'ARN génomique et les protéines des nouvelles particules virales sont synthétisés par la machinerie cellulaire. Ces protéines sont produites sous forme de poly-protéines qui sont ensuite clivées par une protéase virale pour libérer les protéines fonctionnelles [11].

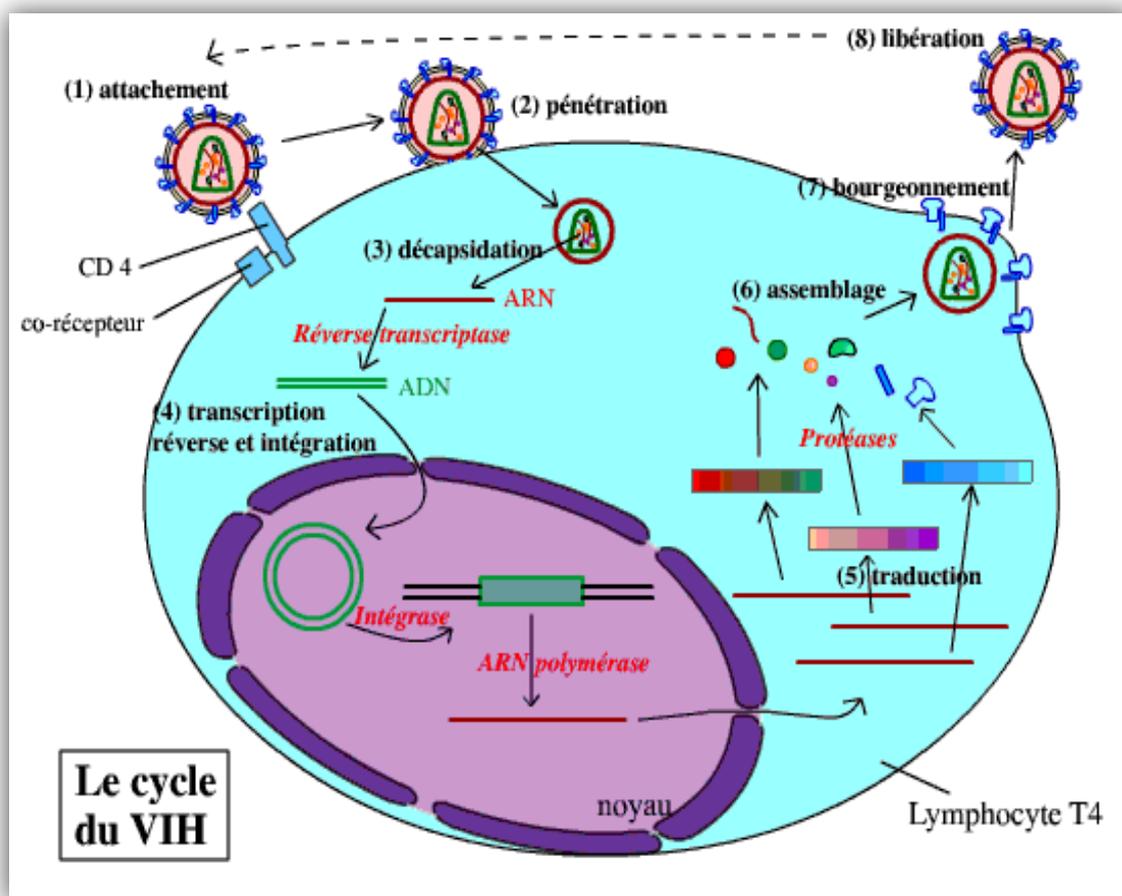


Figure I.4 : Le Cycle du VIH

I.2.6. Diagnostic biologique du VIH :

Divers tests permettent de rechercher le virus de l'immunodéficience humaine dans différents liquides biologiques. On distingue ainsi le diagnostic sérologique et le diagnostic moléculaire. Le diagnostic sérologique est basé sur la technique de détection directe permettant la mise en évidence chez le patient des anticorps dirigés contre des protéines spécifiques du virus de l'immunodéficience humaine (Anticorps anti VIH-1 et VIH-2). Si le résultat est positif, le diagnostic est alors confirmé par un test de référence tel que l'ELISA (Enzyme-Linked Immuno- Sorbent Assay) qui consiste à rechercher l'antigène p24 du virus. Le diagnostic moléculaire quant à lui, est basé sur la mise en évidence des acides nucléiques en utilisant une technique de polymérisation en chaîne, la RT-PCR (Réaction en Chaîne par Polymérase en Temps Réel), permettant de suivre en temps effectivement réel et cycle par cycle la formation des produits amplifiés grâce à des sondes fluorescentes . [9]

I.2.7. Traitement du VIH :

L'entrée du VIH dans l'organisme déclenche une réponse immunitaire innée impliquant les cellules dendritiques, les macrophages et les poly nucléaires . Ces cellules connues sous le nom de cellules présentatrices d'antigène (CPA) présentent les déterminants antigéniques du VIH aux cellules de l'immunité adaptative par le biais des molécules du complexe majeur d'histocompatibilité (CMH). Cette présentation permet aux lymphocytes T ainsi activés d'induire l'activation et la production d'anticorps anti-VIH par les lymphocytes B . L'infection causée par le VIH conduit à une diminution du nombre des cellules T CD4 due principalement à l'action des lymphocytes T cytotoxiques CD8, les cellules tueuses naturelles (NK-de l'anglais Natural killer) et à la libération de nouveaux virions . Ces cellules cytotoxiques reconnaissent et détruisent les lymphocytes CD4+ infectés et induisent leur mort par apoptose. Le système immunitaire se retrouve alors dans un état de délabrement qu'un simple rhume peut entraîner le décès du patient.En absence de traitement, l'organisme devient donc progressivement une cible pour les maladies opportunistes tels que : la tuberculose, le cancer, la méningite etc.. Pour corriger cet état immunitaire délabré, des moyens de restauration du système immunitaire ont été développés .Cependant, ces moyens ne permettent pas une guérison définitive . Il s'agit de combinaisons thérapeutiques appelées traitements antirétroviraux (ARV) .[3]

I.3. Les FLAVONOÏDES :

I.3.1. Les polyphénols :

I.3.1.1. Définition :

Les composés phénoliques ou les polyphénols sont des produits du métabolisme Secondaire des plantes, largement distribués possédant plusieurs groupements phénoliques, avec ou non d'autres fonctions et comportant au moins 8000 structures connues différentes [13], allant de molécules hénoliques simples de bas poids moléculaire tels que, les acides phénoliques à des composés hautement polymérisés comme les tannins. Ils font partie intégrante de l'alimentation humaine et animale [12].

I.3.1.2. Classification : Il existe de nombreuses classes de polyphénols qui sont [14] :

- ❖ Quinones
- ❖ Coumarines
- ❖ Phloroglucinols
- ❖ Acides-phénoliques
- ❖ Flavonoïdes
- ❖ Anthocyanes
- ❖ Tanins

I.3.1.3. Biosynthèse des composés phénoliques : [12]

Les polyphénols sont synthétisés par-deux voies biosynthétiques, celle :

- De l'acide shikimique
- Issue de l'acétate

De plus, la diversité structurale des composés poly-phénoliques due à cette double origine biosynthétique, est encore accrue par la possibilité d'une participation simultanée des voies dans l'élaboration de composés d'origine mixte.

I.3.1.4. Effets biologiques des polyphénols :

Les polyphénols sont associés à de nombreux processus physiologiques interviennent dans la qualité alimentaire, impliqués lorsque la plante est soumise à des blessures mécaniques. La capacité d'une espèce végétale à résister à l'attaque des insectes et des microorganismes est souvent corrélée avec la teneur en composés phénoliques [13]. Ces composés montrent des activités anti-carcinogènes, anti-inflammatoires, antiathérogènes, anti-thrombotiques, analgés-

siques, antibactériens, antiviraux, anticancéreux, anti-allergènes, vasodilatateurs et antioxydants [15][16].

Les composés poly-phénoliques sont d'ailleurs de plus en plus utilisés en thérapeutique. Ils sont regroupés dans la catégorie des veinotoniques et des vasculoprotecteurs. Parmi les veinotoniques, nous citerons le Relvenet ou le Cirkant renfermant du ruténoside, le Daflont ou le Diosmilt renfermant de la diosmine. Un certain nombre de molécules poly-phénoliques sont également en étude clinique comme des anti-agrégant plaquettaire, ou hypotenseur sans résultats probants [12].

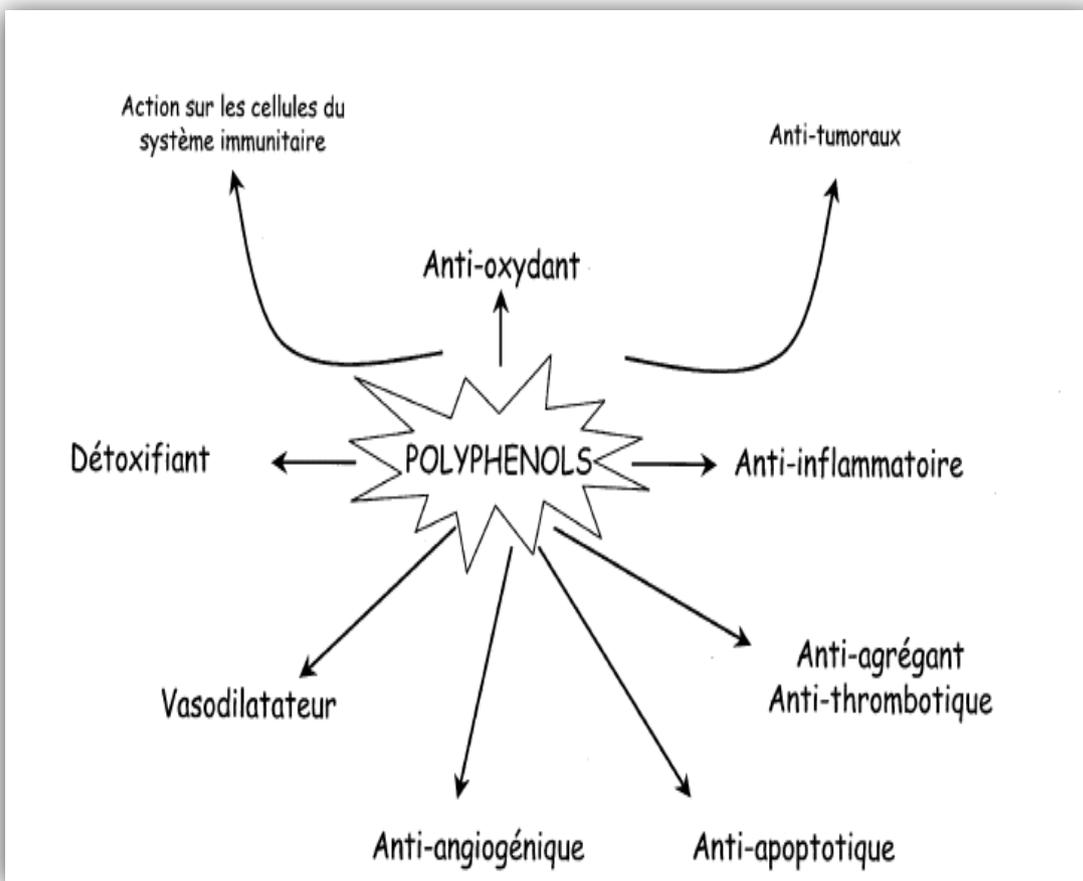


Figure I.5: Effets biologiques des polyphénols

I.3.2. Les flavonoïdes :

I.3.2.1. Définition :

Le terme « flavonoïde du grec flavus, (jaune) en latin » est le nom générique qui désigne une très large gamme de composés naturels appartenant à la famille des polyphénols. Ils sont considérés comme des pigments quasiment universels des végétaux, souvent responsables de la coloration des fleurs, des fruits et parfois des feuilles. À l'état naturel les flavonoïdes se trouvent le plus souvent sous forme d'hétérosides.[1]



Figure I.6 : Certaines plantes contiennent des flavonoïdes

I.3.2.2. Etymologie:

L'expression flavonoïde a été introduite en 1952 par Geissman et Hinreiner pour désigner les pigments ayant un squelette (C6-C3-C6), provenant du mot latin flavus qui signifie jaune [17].

I.3.2.3. Structure chimique :

Tous les flavonoïdes dérivent de l'enchaînement benzo- γ -pyrone et peuvent être classés selon la nature des différents substituants présents sur les cycles de la molécule et du degré de saturation du squelette benzo- γ -pyrone [17][1].

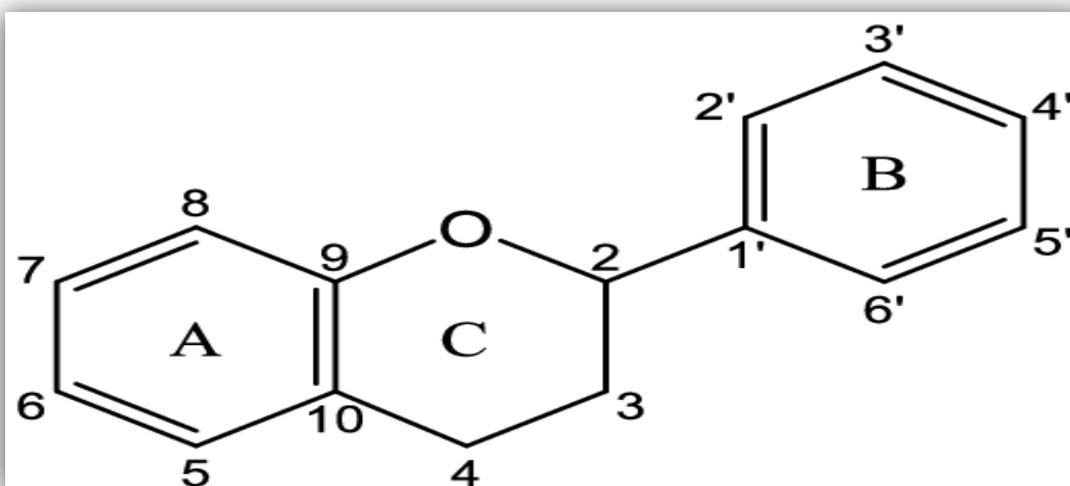


Figure I.7 : schéma de base des flavonoïdes

I.3.2.4. Classification :

Tous les flavonoïdes ont une origine biosynthétique commune et de ce fait possèdent le même élément structural de base. Ils peuvent être regroupés en différentes classes selon le degré d'oxydation du noyau pyranique central, le noyau Brelé à l'hétérocycle C dans les positions 2, 3. [4]

- Dans la position 2: le flavonoïde est appelé Flavane.
- Dans la position 4 de la flavane port un groupement carbonyle la flavane est appelé Flavanone.
- Si la liaison C2-C3 dans le squelette de la flavanone est insaturée le composé est nommé Flavone.
- Si le squelette est substitué en position 3 par un groupement hydroxyle il est désigné par le nom de Flavonol.
- Dans la position3 : le flavonoïde est désigné par le terme Isoflavane.

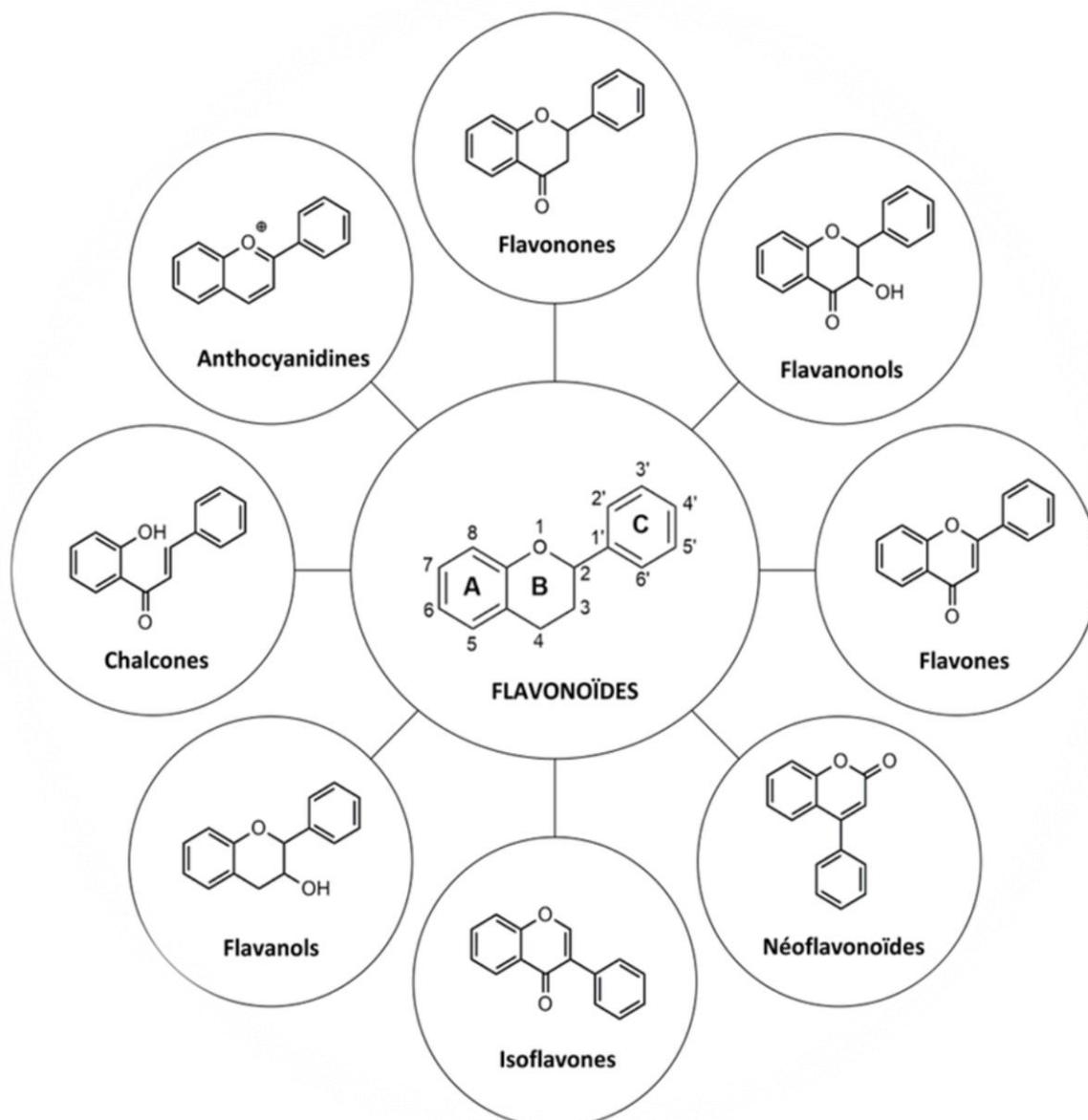


Figure I.8: Structures des différentes classes de flavonoïdes.

I.3.2.5. Propriétés biologiques :

La principale propriété initialement reconnue aux flavonoïdes est d'être "veino -actifs", c'est-à-dire capables de diminuer la perméabilité des capillaires sanguins et de renforcer leur résistance .

a) Effets protecteurs vasculaires:

Les flavonoïdes agissent sur les vaisseaux sanguins sous forme d'activité vitaminique P. Cette activité intervient dans le maintien d'une perméabilité vasculaire normale. Ils sont, de ce fait, utilisés dans certains états pathologiques caractérisés par un défaut affectant la perméabilité

vasculaire[18].

Les effets de l'O- β -hydroxyé-thyl rutoside (HR) ont été étudiés chez des patients présentant une insuffisance veineuse chronique : un traitement à base de HR a permis de restaurer les paramètres hémo-rhéologiques altérés. D'autres flavonoïdes sont responsables d'une augmentation de la résistance des capillaires. Cette activité serait en rapport avec les effets de certains flavonoïdes sur les plaquettes, les leucocytes et sur les enzymes intervenant dans la coagulation sanguine[19].

b) Propriétés antiallergiques:

Les flavonoïdes sont également connus pour leurs effets antiallergiques. Ils agissent par inhibition des enzymes qui favorisent la libération d'histamine à partir des mastocytes et des basophiles :

l'AMPc phospho- diestérase et la Ca²⁺ ATPase[1] .

En outre, la quercétine exerce un puissant effet inhibiteur de la libération d'histamine à partir des mastocytes[20] .

c) Propriété anti-inflammatoire:

Quelques études in vitro montrent que plusieurs flavonoïdes sont capables de modifier le métabolisme de l'acide arachidonique plaquettaire. C'est ainsi que la myricétine et la quercétine bloquent l'action des cyclo-oxygénase et lipo-oxygénase à des concentrations relativement élevées. À faibles concentrations, c'est la lipo-oxygénase qui est inhibée préférentiellement [23].

Les flavones et les flavonols glycosylée comme la quercétine, kaempférol, myrecétine ont une activité inhibitrice de COX (Cyclooxygénase) [2].

d) Propriété anti-ulcérogène:

Les flavonoïdes sont capables de protéger la muqueuse gastrique contre divers agents ulcérogènes. L'hypolaétine-8-glucos, une flavone présente dans diverses espèces du genre Sideritis, a une activité anti-ulcérogène significative. La naringine et la quercétine exercent également une activité anti-ulcérogène mise en évidence chez les rats dont l'ulcère gastrique a été induit par l'éthanol. Il a été suggéré que la quercétine exerce ses effets cytoprotecteurs grâce à un complexe impliquant la stimulation de la prostaglandine et l'inhibition de la production de leucotriènes via la production de mucus et ses propriétés antioxydantes [1] .

e) Propriété antibactérienne:

Les flavonoïdes sont reconnus par leur toxicité vis-à-vis des microorganismes. Le mécanisme de toxicité peut être lié à l'inhibition des enzymes hydrolytiques (les protéases et les carbohydrolases) ou d'autres interactions pour inactiver les adhésines microbiens, les protéines de transport et d'enveloppe cellulaire.

f) Activités antivirales :

L'activité antivirale des flavonoïdes contre VIH peut être liée directement par leurs effets sur les enzymes responsables de sa réplication (VIH-1 reverse transcriptase ou VIH-1 integrase) par ailleurs d'autres flavonoïdes montrait une activité antivirale contre le virus d'influenza, VIH-1, VIH-2. Quercétine , apigénine, catéchine et hespéridine sont parmi les flavonoïdes caractérisés par leurs propriétés antivirales contre onze types de virus .

Les flavonoïdes aglycones pourvus d'un groupement hydroxyle libre en C3 ont montré une bonne activité antivirale, les flavanes sont généralement plus efficaces que les flavones et les flavanones contre VIH-1et VIH-2 [21].

g) Activités anti-tumorales :

La plupart des flavonoïdes sont in vitro, antimutagènes ; quelques flavonols sont sur les mêmes modèles mutagènes et un petit nombre d'entre eux sont anti-cancérogènes et inhibiteur de la croissance des cellules tumorales in vitro .Les effets anti-carcinogènes de la quercétine et d'autres flavonoïdes deviennent de plus en plus évidents [10].

h) Autres propriétés :

Les flavonoïdes sont métabolites secondaires qui se sont avérés avoir des propriétés biologiques étendues telles que anti-inflammatoire , antidiabétique ,antibactérien , antioxydant , anticancéreux ,antivieillessement, Effets cardiovasculaires.....

I.4. TRAITEMENT ANTIRÉTROVIRAL :

Le choix initial du traitement est une décision essentielle pour l'avenir thérapeutique du patient et doit être réalisé par un médecin expérimenté. En 2006, plusieurs antirétroviraux sont disponibles en La Suisse (tableau 1) comme cible du premier traitement antirétroviral devrait rendre la charge virale indétectable en six mois. Pour ce faire, il faut utilisation d'une association de trois antirétroviraux (INRT ;IP ; Inhibiteur de fusion) [7]

Tableau I.1 : Médicaments Antirétroviraux disponibles en Suisse en 2006

Médicament (abréviation)	Nom de marque
zidovudin (AZT)	Retrovir®
AZT + 3TC	Combivir®
AZT + 3TC + abakavir	Trizivir®
didanozin (ddI)	Videx®
stavudin (d4T)	zeret®
Abacavir	Ziagen®

lamivudin (3TC) 3TC + Abacavir Emtricitabin (FTC)	3-TC® Kivexa® Emtriva®
Ritonavir Nelfinavir Tipranavir	Norvir® Viracept® Aptivus®

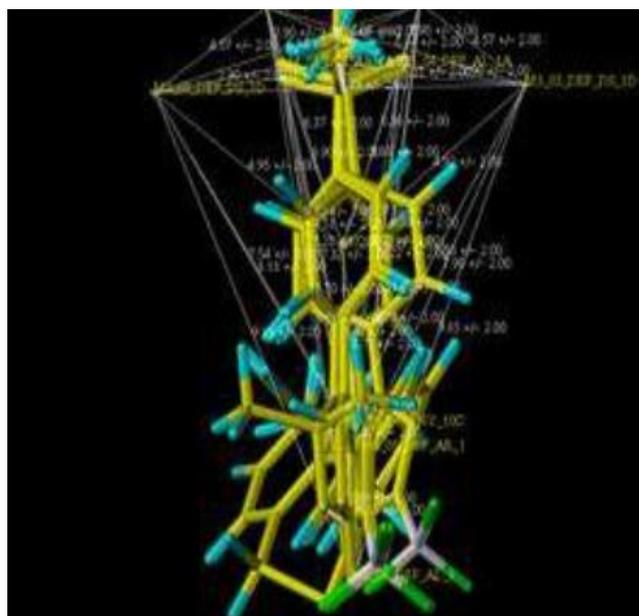
Références :

- [1] K ;Ghedira . "Les flavonoïdes: structure, propriétés biologiques, rôle prophylactique et emplois en thérapeutique. (2005). " *Phytothérapie*3(4): 162-169.
- [2] R; Landolfi et al. "Modification of platelet function and arachidonic acid metabolism by bioflavonoids: structure-activity relations. (1984). " *Biochemical pharmacology*33(9): 1525-1530
- [3]Pr Alitono Alan Jay et al., Université d'Abomey-Calavi ; Ecole Polytechniq d'Abomey-Calavi (EPAC) ; Prévalence de l'infection par le virus de l'hépatite B chez les personnes vivant avec le VIH/sida pendant le traitement Un médicament antirétroviral a été administré à l'hôpital Bethesda de Cotonou ; (18).
- [4]A. Meghsel et al ; Extraction caractérisation, et évaluation de L'activité biologique des flavonoïdes de *Rosmarinus officinalis* L;2012 (6-7).
- [5]V. Schiffer ;B. Hirschel ; VIH – sida Traitement de l'infection à VIH en 2006 :points forts .
- [6]Gilles Furelaud, Benjamin Pavie et al ; article : Le virus du sida ;2002.
- [7]Revue Médicale Suisse – www.revmed.ch – 10 janvier 2007.
- [8]A ;Castel, M ;Befus et al, Epidémiologie et utilisation social de la charge virale communautaire en tant que Biomarqueur de la charge du VIH/SIDA basé sur la population, 2012
- [9]I, Erlund Review of the flavonoids quercetin, hesperetin, and naringenin. Dietary sources, bioactivities, bioavailability and epidemiology. *Nutr Res* 24: 851-74. (2004)
- [10]Babar Ali, M , et al . Methyl Jasmonate and Salicylic Acid Induced Oxidative Stress and Accumulation of Phenolics in *Panax ginseng* Bioreactor Root Suspension Cultures. *Molecules*. 12: 607-621. (2007)
- [11]G. Furelaud et al, Planète-vie-ENS ;2000.
- [12]S. Martin, R. Andriantsitohaina. Mécanismes de la protection cardiaque et vasculaire des polyphénols au niveau de l'endothélium. *Annales de cardiologie et d'angéiologie*. 304-315. (2002)
- [13]T ,Bahorun. Substances naturelles actives : la flore mauricienne, une source d'approvisionnement potentielle. *Food and agricultural resarch council ., Réduit, Mauritus*.83-94. (1997).
- [14] O. Touafek ; Etude phyto-chimique de plantes médicinales du nord et du sud Algériens, Thèse de Doctorat Constantine, Algérie, 2010.
- [15] H, Falleh. C, Abdely. et al Phenolic composition of *Cynara cardunculus* L. organs, and their biological activities .*C. R. Biologies*. 331: 372. (2008)
- [16]Gomez-Caravaca , A.M. ,et al Advances in the analysis of phenolic compounds in products derived from bees. *J Pharmaceutical and Biomedical Analysis*. 41: 1220-1234. [17]Bouakaz, I. Etude phyto-chimique de la plante *Genista Microcephala*. Mémoire de magister. Batna.2006
- [18]Shih CM, Lin H, Liang YC, et al. Concentration-dependent differential effects of quercetin on rat aortic smooth muscle cells. *Eur J Pharmacol* 496 : 41-8. (2004)

- [19]Stoclet J C, Ndiaye M, et al. Vascular protection by dietary polyphenols. *Eur J Pharmacol* 500 : 299-313. (2004)
- [20]S, Yamamura , et al. Antihistaminic flavones and aliphatic glycosides from *Mentha spicata*. *Phytochemistry* 48 : 131-6. (1998).
- [21]David Germanaud, Gilles Furelaud ; *Les trithérapies antirétrovirales* ; 2003.
- [22]Folts J D Potential health benefits from the flavonoids in grape products on vascular disease. 505: 95-111. (2002)
- [23]Formica JV, Regelson W Review of the biology of quercetin and related bioflavonoids; *Food Chem Toxicol* 33: 1061-80. (1995)
- [24]Ghosh, M. N. et al Current trends in flavonoid research. *Ind. J. Pharmac.*12: 213-228. (1980)
- [25]H. Bauby et al ; Institut Cochin , Université Paris Descartes ,CNRS (UMR 8104),Paris France, TIP47, un cofacteur cellulaire essentiel à l'incorporation des glycoprotéines d'enveloppe du VIH-1 dans la particule virale,2008.
- [26] M, Amella. C, Bronner. et al.Inhibition of mast cell histamine release by flavonoids and biflavonoids . *Planta Med* 51 (1): 16-20. (1985)
- [27]B, Lee. M , Sharron. et al . Quantification of CD4, CCR5, and CXCR4 levels on lymphocyte subsets, dendritic cells, and differentially conditioned monocyte-derived macrophages. 96(9) : *Proc Natl Acad Sci USA*. 27avr 1999; 15-20

Chapitre II

QSAR : Principe et méthodologie



II.1 Introduction :

Les relations quantitatives structure-activitépropriété (QSAR/QSPR) sont de plus en plus utilisées, en raison de la croissance des moyens de calcul. Récemment, la mise en place d'un nouveau règlement européen REACH, qui préconise de limiter le recours à l'expertise, a donné un nouvel élan au développement de tels modèles prédictifs. Ces dernières années, l'utilisation des méthodes QSAR n'a cessé de croître. Il est même devenu indispensable en chimie pharmaceutique et en conception de médicaments. [2]

En fait, les premiers travaux QSPR/QSAR remontent au XIXe siècle. J'ai soumis la demande l'existence de relations entre les activités physiologiques et les structures chimiques par liaison les changements d'activité biologique ont des modifications structurelles. Cependant, dans ce a cette époque, non Les structures moléculaires n'étaient pas encore connues. Une avancée importante pour les mêmes modèles QSPR / QSAR grâce au développement équations de Hammett [1] dans lesquelles les constantes σ sont distinguées vitesses de réaction quantitatives des composés organiques.

$$\log K/K_0 = \sigma \rho \quad (\text{eq1})$$

Ou K et K₀ sont les constantes respectives de la réaction étudiée et de celle d'une référence et ρ une constante de réaction dépendant du type de réaction. [1] les premiers travaux utilisant la méthodologie QSPR/QSAR telle qu'employée actuellement sont dus à [3] et [5]. D'un côté, Hansch a proposé des modèles reliant directement l'activité biologique des composés avec les propriétés hydrophobes, électroniques et stériques à l'échelle moléculaire. D'un autre côté, Free et Wilson ont développé des modèles empiriques, dits de contributions de groupes, pour l'étude de l'activité biologique. Au cours de ces dernières décennies, l'utilisation des méthodes QSPR/QSAR n'a pas cessé de progresser. Elle est même devenue indispensable en chimie pharmaceutique, en toxicologie et pour la conception de médicaments.

II.2 Historique :

Les relations quantitatives structure-activité/propriété (QSAR/QSPR) sont de plus en plus utilisées, en raison de la croissance des ressources informatiques. Ils ont été largement utilisés dans les industries pharmaceutiques, chimiques et cosmétiques, notamment pour la conception rationnelle de nouveaux principes actifs et de nouvelles entités chimiques. Les premiers travaux utilisant la méthode (QSAR) ont débuté au début des années 1960 avec Hansch d'une part et Free et Wilson d'autre part, qui ont proposé un modèle mathématique pour corréler l'activité biologique et la structure chimique. Les développements de cette étude (QSAR) sont très anciens, de 1868, quand Alexander Crum-Brown et Thomas Fraser montrent l'existence d'une relation entre l'activité physiologique et la structure chimique, Richet établissant une relation entre la toxicité et les propriétés physico-chimiques. Indépendamment, Meyer et Overton ont décrit une corrélation linéaire entre la lipophilicité (coefficient de partage pétrole-eau) et biologiques (stupéfiants). Au cours des dernières décennies, ce domaine a été largement étudié et les données bibliographiques disponibles sur cette approche sont maintenant importantes [4].

II.3 Définition :

Une relation quantitative structure-activité (QSAR) est le processus par lequel une structure chimique est corrélée à un effet spécifique tel que l'activité biologique ou la réactivité chimique. Ainsi, l'activité biologique peut être exprimée quantitativement, comme la concentration de substance nécessaire pour obtenir une certaine réponse biologique. De plus, lorsque les propriétés ou structures physico-chimiques sont exprimées par des nombres, une relation mathématique, ou relation quantitative structure-activité, peut être proposée entre les deux. L'expression mathématique obtenue peut alors être utilisée comme moyen de prédiction de la réponse biologique pour des structures similaires [6].

Le QSAR le plus courant est de la forme [6] :

$$\text{Activité} = f(x) \quad (\text{eq2})$$

X = propriétés physico-chimiques et/ou structurelles.

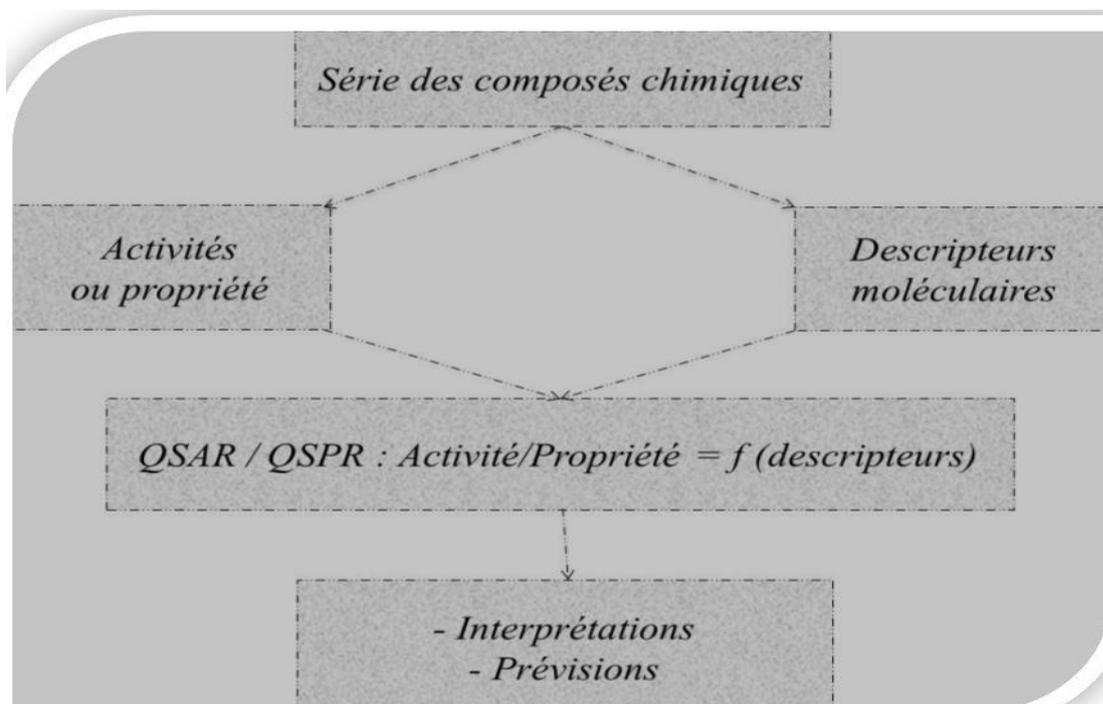


Figure II.1 :Modèle de l'étude de relation structure activité/propriété.

II.4 Principe :

Le principe d'une étude QSAR/QSPR consiste à trouver une relation mathématique liant quantitativement une activité biologique, ou une propriété, mesurée pour une série de composés similaires dans les mêmes conditions expérimentales, à des descripteurs moléculaires à l'aide de méthodes statistiques. L'objectif de ces études est d'analyser les données structurelles afin de détecter les déterminants de l'activité ou du bien étudié. Pour ce faire, différents types de méthodes statistiques peuvent être utilisées. L'expression mathématique obtenue peut alors être utilisée comme moyen prédictif de

l'activité/propriété étudiée pour de nouvelles molécules ou des molécules pour lesquelles les données expérimentales ne sont pas disponibles [7].

Ceci peut être traduit par l'équation suivante :

$$\text{Activité/ Propreté} = f(\text{Descripteurs moléculaires}) \quad (\text{eq3})$$

Le but de cette méthode est d'analyser les données structurales afin de découvrir les déterminants de la propriété/activité mesurée. Pour ce faire, différents types des outils statistiques peuvent être utilisés [8] :

- ❖ Régressions linéaires multiples (MLR).
- ❖ Régressions aux moindres carrées partielles (PLS).
- ❖ Analyse par composantes principales (ACP).
- ❖ Régressions non linéaire multiple (MNLR).

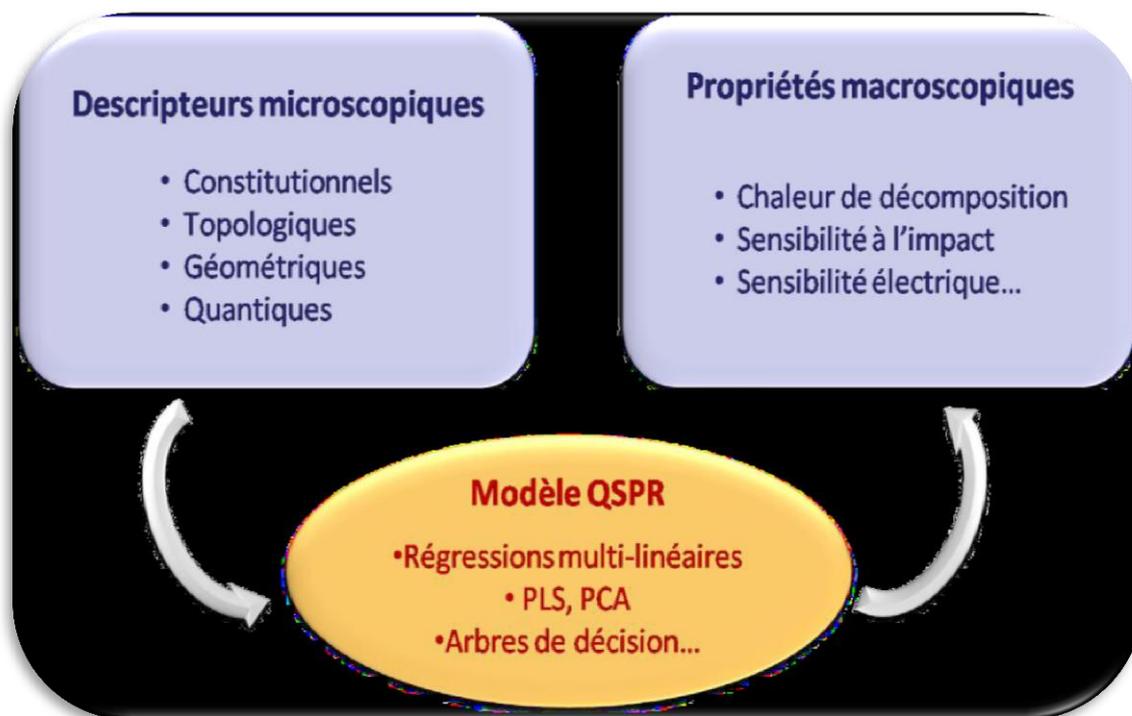


Figure II .2 : principe de la méthode QSPR

II.5 Outils et Méthodologie de QSAR :

II.5.1 Les bases de données :

Le modèle QSAR/QSPR s'appuie fortement sur des données de référence empiriques. sélection de fond de teint les données sont cruciales pour développer un tel modèle. Dans la plupart des cas, les données expérimentales sont extraites de la littérature. Pour être de qualité, une base de données doi-

vent consister en des données expérimentales aussi fiables que possible obtenues en suivant un protocole car les erreurs de ceux-ci vont se propager au modèle dernier. Il y a plusieurs choses à vérifier pour les étapes de nettoyage de la base de données. Assurez-vous d'abord que les structures sont chimiquement correctes, les structures défectueuses causent des dommages profils et donc mauvais modèles [7][8].

Les modèles QSPR sont généralement plus efficaces pour un jeu de molécules similaires. De même, il est important de considérer une base de données dont les molécules comporter de la même manière vis-à-vis de la propriété cible. De même, afin de réduire la probabilité de rencontrer des molécules très différentes de la plupart des autres, ce qui peut peser significativement sur le modèle et augmenter l'incertitude sur les prédictions (Outliers) dans le jeu de données, plus la diversité structurelle d'une base de données est grande important et de plus la taille de la base de données doit être importante [8].

II.5.2 Descripteur moléculaire :

Un descripteur moléculaire est un paramètre (une valeur numérique) spécifique à une structure chimique donnée. Ces valeurs peuvent être obtenues expérimentalement ou calculées de la structure de la molécule. Les descripteurs calculés permettent de réaliser des prédictions sans avoir à synthétiser les molécules, ce qui est un des objectifs de la modélisation moléculaire. Les descripteurs moléculaires jouent un rôle fondamental dans les études de la relation structure quantitative activité/propriété. Ils sont utilisés comme variables indépendantes pour prédire une variable dépendante (activité ou propriété). L'utilisation de descripteurs moléculaires dans le développement de modèles QSAR/QSPR n'est pas une tâche facile. Tout d'abord, un très grand nombre de descripteurs moléculaires, de différentes complexités et de diverses conceptions a été introduit au cours des dernières années. Puis, pendant ce temps, aucune règle stricte n'a été établie pour la sélection de descripteurs appropriés parmi le grand nombre de descripteurs disponibles [7].

II.5.2.1 Type de descripteur :

Les descripteurs moléculaires sont des représentations mathématiques formelles d'une molécule obtenue grâce à un algorithme bien défini et appliquée à la représentation structure moléculaire ou procédure expérimentale bien définie : un descripteur moléculaire c'est le résultat final d'une procédure logique et mathématique qui transforme l'information d'un produit chimique codé dans une représentation symbolique d'une molécule en un nombre utile ou le résultat d'une expérience standardisée [9].

II.5.2.1.a Descripteurs constitutionnels :

Les descripteurs constitutionnels sont plus simples en termes conceptuels et coût de calcul, qui est couramment utilisé. Ils reflètent la composition chimique d'un composé sans aucune information sur sa géométrie moléculaire ou sa structure. Parmi ces descripteurs, on cite [10]:

- ✓ Nombre total des atomes dans la molécule.
- ✓ Nombres absolu et relatif d'atomes de certaines identités chimiques (C, H, O, S, N, F, ... etc.) dans la molécule.
- ✓ Nombres absolu et relatif de certains groupes et fonctionnalités chimiques dans la molécule.
- ✓ Nombre total de liaisons dans la molécule.
- ✓ Nombres absolu et relatif de liaisons simples, doubles, triples, aromatiques ou autres existantes dans la molécule.
- ✓ Nombre total des noyaux, nombre de noyaux divisé par le nombre d'atomes.
- ✓ Nombre total et relatif de phényle ou d'autres noyaux aromatiques.

Cependant, ce type de descripteur est utilisé pour obtenir les modèles QSAR les plus simples dans certains cas, elle est imparfaite car elle ne permet pas de proposer des mécanismes les interactions impliquées dans l'activité étudiée. Par exemple, ils ne le permettent pas différencier les isomères. Cependant, dans la plupart des cas, cela change la position de la variable la valeur expérimentale de l'activité [6][11].

II.5.2.1.b Descripteurs topologiques :

Les descripteurs topologiques (ou indices topologiques) décrivent les liaisons atomiques dans une molécule. Ceux-ci sont considérés comme des descripteurs "plus complexes" qui ne contiennent pas nécessairement une signification chimique claire mais contiennent en leur sein des informations sur la taille globale du système, sa forme générale et ses ramifications . le principe est de trouver une valeur différente pour chaque squelette moléculaire .Ces descripteurs sont faciles à calculer, et leurs valeurs sont généralement précises ,Ils apparaissent souvent sur les modèles. Il vient de la théorie des graphes qu'il a développée Euler en 1736 ; Ce théorème est appliqué à la table de contact, qui est une représentation compacte de la liaison entre les atomes au sein d'une molécule [7][11].

Un graphe est un ensemble de points dont certains sont reliés par des droites ; Ça permet de représenter la topologie de la molécule sans se soucier de la géométrie spatiale exacte de celle –ci dernier [7]. Cette théorie est inventée à partir de quelques lois simples : deux points adjacents sont reliés par des lignes et deux lignes avec un point commun sont adjacentes. L'ordre ou le degré d'un point est le nombre de lignes reliées à ce point. Un pas est un enchaînement de points et de lignes adjacentes débutant par un point et se terminant par un point. Un chemin est un pas dans lequel aucun point n'est utilisé plus d'une fois. D'un point de vue moléculaire, un point représente un atome et une ligne une liaison covalente. Les chemins sont caractéristiques de l'architecture de l'ensemble des atomes constituant la molécule. Les atomes d'hydrogène sont exclus du graphe pour simplifier les calculs [7].

Tableau II.1 : Les descripteurs topologiques calculés dans étude QSAR [12].

Descripteur	Définition
	$\omega = \frac{1}{2} \sum_{ij} d_{ij} = \frac{1}{2} \sum_i s_i$
L'indice de Wiener (ω)	Est simplement la demi-somme de tous les éléments d'une telle matrice où S_i est la somme des distances
	$J = C \cdot \sum_k (s_i \cdot s_j)^{-0.5}$
L'index J de Balaban (J)	Où $C = m(1 + u)$; m est le nombre d'arêtes et u est le nombre cyclomatique.

II.5.2.1.c Descripteurs géométriques :

Ces descripteurs sont évalués à partir des positions relatives des atomes de la molécule dans l'espace, ainsi que des rayons et des masses atomiques. Il nécessite donc une connaissance de la structure 3D de la molécule. Cela peut être obtenu expérimentalement, bien sûr, mais le plus souvent par modélisation moléculaire, expérimentale ou par prototypage. Par conséquent, contrairement aux descripteurs précédents, ils nécessitent un certain temps de calcul. Parmi ces descripteurs, on retrouve la taille, la surface moléculaire, le moment d'inertie ou encore certaines distances, et les angles ou dodécaèdres entre les atomes de la molécule [13].

Le volume moléculaire : noté MV , en cm^3 , est défini par la formule suivante [13] :

$$MV = \frac{MW}{d} \quad (\text{eq4})$$

Avec : MW est le poids moléculaire et d la densité.

II.5.2.1.d Descripteurs quantiques :

Les approches chimiques quantitatives nous permettent d'accéder à des informations supplémentaires telles que les données énergétiques, vibrationnelles et orbitales du système. Il est alors possible, par exemple, de calculer des énergies de dissociation. Les propriétés électroniques telles que le moment dipolaire ou la polarisabilité sont également évaluées. Les descripteurs permettent également de définir différentes interactions entre et au sein des molécules, ces dernières ayant un impact significatif sur la plupart des propriétés des systèmes moléculaires. Il s'agit notamment de descripteurs d'interactions qui

peuvent être dérivés de DFT conceptuels ou encore de charges partielles qui peuvent être calculées selon différentes approches [14].

Énergie HOMO : notée E_{HOMO} , mesurée en eV, est le niveau d'énergie le plus élevé la molécule qui contient des électrons, elle est directement liée au potentiel d'ionisation. Lorsqu'une molécule agit comme une base de Lewis (une paire d'électrons donneurs) dans la formation d'une liaison, les électrons sont alimentés à partir de cette orbite. Il mesure la nucléophilie d'une molécule et caractérise la sensibilité de la molécule à l'attaque par électrophiles [15].

Énergie LUMO : notée E_{LUMO} , mesurée en eV, est le niveau d'énergie le plus bas molécule qui ne contient pas d'électrons, elle est directement liée à l'affinité électronique. Lorsqu'une molécule agit comme un acide de Lewis (une paire d'électrons accepteurs) dans la formation de liaisons, les paires d'électrons entrants sont reçues dans cette orbite. Elle mesure l'électrophile d'une molécule et caractérise la sensibilité de la molécule à l'attaque par nucléophiles [15].

II.5.2.1.e Descripteurs physico-chimiques :

Les descripteurs physicochimiques, (ou indices physicochimiques) certains d'entre eux reflètent la composition moléculaire du composé (le nombre et le type d'atomes et de liaisons présents dans la molécule, le nombre de cycle, les propriétés donneur/accepteur de liaison H, cation, anion, etc...). D'autres représentent le caractère hydrophile ou lipophile de la molécule généralement évalué à partir du coefficient de partage Octanol/eau représenté par le log P [16].

➤ **Le coefficient de partage P** [16]:

Le log P est une mesure de la solubilité différentielle de composés chimiques dans deux solvants, l'eau et l'octanol ; c'est le coefficient de partage octanol /eau :

$$\text{Log P} = \text{Log} (C_{\text{oct}} / C_{\text{eau}}) \quad (\text{eq5})$$

où C_{oct} et C_{eau} représentent les concentrations du polluant respectivement dans l'octanol et l'eau. Cette valeur permet d'appréhender le caractère hydrophile ou hydrophobe (lipophile) d'une molécule.

➤ **La réfractivité moléculaire** : notée (MR), en m³/mol, est le volume de la substance absorbée par mole de cette substance.

- Elle est définie par Lorentz-Lorenz (par la formule suivante [17] :

$$\text{MR} = \frac{n^2 - 1}{n^2 + 2} \frac{\text{MW}}{d} = \frac{n^2 - 1}{n^2 + 2} \text{MV} \quad (\text{eq6})$$

MV : C'est le volume molaire ;

MW : le poids moléculaire ;

d : la densité ;

n : est l'indice de réfraction .

-La réfractivité moléculaire est également proportionnelle à la polarisabilité α_e , par la relation suivante [18] : $MR = 4/3\pi NA \alpha_e$ (eq7)

NA : est le nombre d'Avogadro ; $NA = 6.022 * 10^{23}$.

➤ **L'indice de réfraction:** n, est défini par la formule de Lorentz suivante[16] :

$$n = \sqrt{\frac{2MR+MW}{MV-MR}} \quad (\text{eq8})$$

➤ **La densité :** (d), en (kg/m³), est liée à la masse et la taille de la molécule. C'est le rapport du poids moléculaire MW au volume moléculaire MV [7] :

$$d = \frac{MW}{MV} \quad (\text{eq9})$$

II.6 Méthodes d'analyse de données :

La préparation de modèles QSAR n'est pas une tâche facile. La première difficulté réside dans la différence d'échelles entre les données à corrélérer, où la structure est au niveau moléculaire alors que les propriétés à prédire sont à l'échelle macroscopique. De plus, il faut prendre en compte les problèmes d'incertitude au niveau des structures données moléculaires et expérimentales (protocoles de mesure). Traiter une grande quantité de données est également très difficile processus de développement du modèle QSAR. Les corrélations entre un grand nombre de descripteurs d'un grand nombre de molécules peuvent être analysés ainsi que leur activité. Cependant, il n'y a pas de règle stricte concernant la sélection des paramètres les plus structurés. important de tous disponibles [8].

En effet, de nombreux outils existent et il s'agit de trouver la méthode la plus appropriée pour obtenir un modèle fiable à partir des données disponibles. Selon les cas, plusieurs approches sont possibles, et il faut alors choisir celle qui permet la meilleure caractérisation du système. Les différentes méthodes présentées dans ce qui suit sont celles utilisées lors de l'étude, pour élaborer des modèles (plus ou moins de paramètres, linéaires ou non linéaires, interprétables ou non), pour sélectionner les paramètres les plus pertinents, pour valider ces modèles. (internes ou externes) et d'identifier leurs domaines d'application [13].

II.6.1 La régression linéaire simple :

La régression s'adresse à un type de problème où les 2 variables quantitatives continues X et Y ont un rôle asymétrique : la variable Y dépend de la variable X.

La liaison entre la variable Y dépendante et la variable X indépendante peut être modélisée par une fonction de type $Y = \alpha + \beta X$, représentée graphiquement par une droite [19].

$$Y = \alpha + \beta X \quad (\text{eq10})$$

II.6.2 La régression linéaire multiple :

La régression linéaire multiple est la méthode de modélisation statistique la plus simple et la plus largement appliquée dans les études de relation structure-activité. La méthode a été popularisée par Hansch en reliant l'activité biologique aux propriétés expérimentales lipophiles, électroniques et stériques d'une série de composés. La méthode MLR est basée sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante Y (dans notre cas l'activité biologique) et une série de n variables indépendantes X_i (ici, les descripteurs moléculaires). L'objectif est d'arriver à une équation mathématique de la forme [11] :

$$y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + \dots + a_n X_n \quad (\text{eq 11})$$

Où a_i ($i = 0, 1, \dots, n$) sont les coefficients de la régression.

La détermination de l'équation (eq12) à partir d'un ensemble de données de p échantillons revient à résoudre un système de p équation :

$$y_1 = a_0 + a_1 x_{1,1} + a_2 x_{2,1} + \dots + a_n x_{n,1} + b_1$$

$$y_2 = a_0 + a_1 x_{1,2} + a_2 x_{2,2} + \dots + a_n x_{n,2} + b_2$$

⋮

$$y_p = a_0 + a_1 x_{1,p} + a_2 x_{2,p} + \dots + a_n x_{n,p} + b_p$$

(eq12)

Où les résidus b_i représentent l'erreur du modèle, constituée par l'incertitude sur la variable dépendante Y_i d'une part, sur les variables indépendantes X_i d'autre part, mais aussi par les informations contenues dans les variables indépendantes mais non exprimées via les variables dépendantes. Ce système d'équation peut être donné sous la forme matricielle suivante :

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \cdots & \cdots & x_{n,1} \\ 1 & x_{1,2} & x_{2,2} & \cdots & \cdots & x_{n,2} \\ \vdots & \vdots & & & & \\ \vdots & \vdots & & & & \\ 1 & x_{1,p} & x_{2,p} & \cdots & \cdots & x_{n,p} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ \vdots \\ a_p \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ b_p \end{pmatrix} \quad (\text{eq13})$$

Soit de manière condensée :

$$Y = XA + B \quad (\text{eq14})$$

Où Y, X, A et B représentent respectivement le vecteur de propriété, la matrice des attributs (descripteurs), la matrice des coefficients et la matrice des erreurs de régression [6].

II.7. Coefficients et tests statistiques standards :

Afin de déterminer la qualité d'un modèle, différents paramètres statistiques sont employés, tels que les erreurs quadratiques moyennes (Mean Square Errors), les coefficients de corrélation qui sont régulièrement utilisés dans les études QSAR/QSPR, sont décrits en détail dans cette partie [7].

Coefficient de corrélation R (et coefficient de détermination R²) :

C'est l'indicateur statistique le plus courant qui évalue la fraction de variance de l'activité caractéristique cible expliquée par le modèle. La qualité du modèle est souvent visualisée sur un nuage de points, où les valeurs calculées de la caractéristique (activité biologique) sont tracées, en fonction des valeurs expérimentales. La qualité de la modélisation est meilleure lorsque les points de ce graphique sont proches de l'ajustement linéaire [5]. La modulation des points sur cette ligne peut être évaluée par le coefficient de détermination R² [5] :

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2} \quad (\text{eq15})$$

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.

R² qui varie de 0 à 1, mesure le rapport du contraste total de Y sur la moyenne s'explique par la régression. Plus sa valeur est proche de 1 (cas idéal), plus sa valeur est élevée les valeurs attendues et observées sont corrélées. Une valeur R² faible signifie que le modèle a une valeur faible le pouvoir explicatif et (certains) descripteurs n'ont aucun effet sur la réponse. Le coefficient est facile comprenant, il faut faire attention à ne pas y attacher trop d'importance car c'est loin de fournir un dossier un critère

suffisant pour juger de la qualité de la régression. R^2 non recommandé pour comparer des modèles avec un nombre différent de descripteurs, le coefficient R^2 il dira toujours choisir le modèle avec le plus de descriptions car R^2 le fera plus significatif (attendez-vous à plus d'espace), même si les variables n'ont aucun effet sur la réponse Y [6].

La valeur de R^2 dépend de la taille de l'échantillon et du nombre de variables prédictives dans l'équation. Elle garde la même valeur ou augmente lorsqu'une nouvelle variable est la prédiction est ajoutée à l'équation de régression, même si la variable ajoutée ne contribue pas réduire la variance inexpliquée. Par conséquent, un autre paramètre statistique peut être utilisé, appelé R^2 ajusté (R^2 ajusté) [20].

Coefficient de détermination ajusté R^2 ajusté : Ce coefficient est utilisé en régression multiple par ce qu'il tient compte du degré de liberté : Avec : est le nombre des observations (les molécules) ; est le nombre de variables indépendantes (les descripteurs) ; est le coefficient de détermination du modèle [7]

$$PRESS = \sum (Y_{obs} - Y_{cal})^2$$

$$TSS = \sum (Y_{obs} - \bar{Y})^2$$

$$R_{adj}^2 = (1 - r^2) \left(\frac{n - 1}{n - p - 1} \right)$$

$$R_{CV}^2 = 1 - \frac{PRESS}{TSS}$$

$$S_{PRESS} = \sqrt{\frac{PRESS}{n}}$$

➤ **PRESS :** Somme des carrés des écarts entre les valeurs observées et les valeurs prévues à partir du modèle [20]

➤ **SSY :** Somme des carrés des distances des valeurs observées pour une variable par rapport à la moyenne de cette variable, la somme des carrés permet de mesurer la variation totale dans une variable. [20] [26]

➤ **R^2_{cv} :** Est la variation de la statistique du R^2 obtenue en ajoutant ou en enlevant une variable indépendante .n [20][27]

➤ **SPRESS :** La capacité prédictive des modèles est évaluée par l'erreur de racine-carrée moyenne

➤ **Le test de Fisher F :** L'indice de Fisher F est également couramment employé afin de mesurer

le niveau de signifiante statistique du modèle, c'est-à-dire la qualité du choix du jeu de paramètres.

- **SE** : Il est l'erreur Standard.

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

$$SE = \frac{\sum(y_i - \hat{y}_i)^2}{n - p - 1}$$

II.8. 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 [21].

Cette qualité est vérifiée par ce que l'on appelle validation. Sa robustesse, c'est-à-dire l'influence des composés de la série d'apprentissage sur le modèle, est estimée par des méthodes de validation interne. Afin d'estimer son pouvoir prédictif, des données expérimentales supplémentaires sont nécessaires afin de déterminer la capacité du modèle à prédire ces valeurs c'est ce que l'on appelle validation externe. Enfin, il est important de savoir quel type de molécules utilisées avec quel modèle. On parle alors de domaine d'applicabilité [20].

II.8.1 Validation interne :

La validation interne est souvent la technique la plus utilisée dans les études QSAR déterminer la stabilité du modèle et tester l'influence de chaque échantillon de l'ensemble d'entraînement sur le modèle final. Pour ce faire, nous utilisons des techniques validation croisée leave-one-out et leave-n-out (validation croisée CV)[16].

II.8.1.a. La procédure leave-n-out :

Cette procédure consiste à extraire chaque fois un certain nombre n de molécules de l'ensemble initial d'apprentissage à k molécules et à construire un nouveau modèle avec les $k-n$ molécules restantes en utilisant les descripteurs choisis (les mêmes descripteurs utilisés dans la construction du modèle original) le nouveau modèle est ensuite utilisé pour la prédiction sur les n molécules retirées. Ce processus est ensuite réitéré p fois pour retirer et prédire les valeurs de toutes les molécules de l'ensemble d'entraînement (figure II.3)[16].

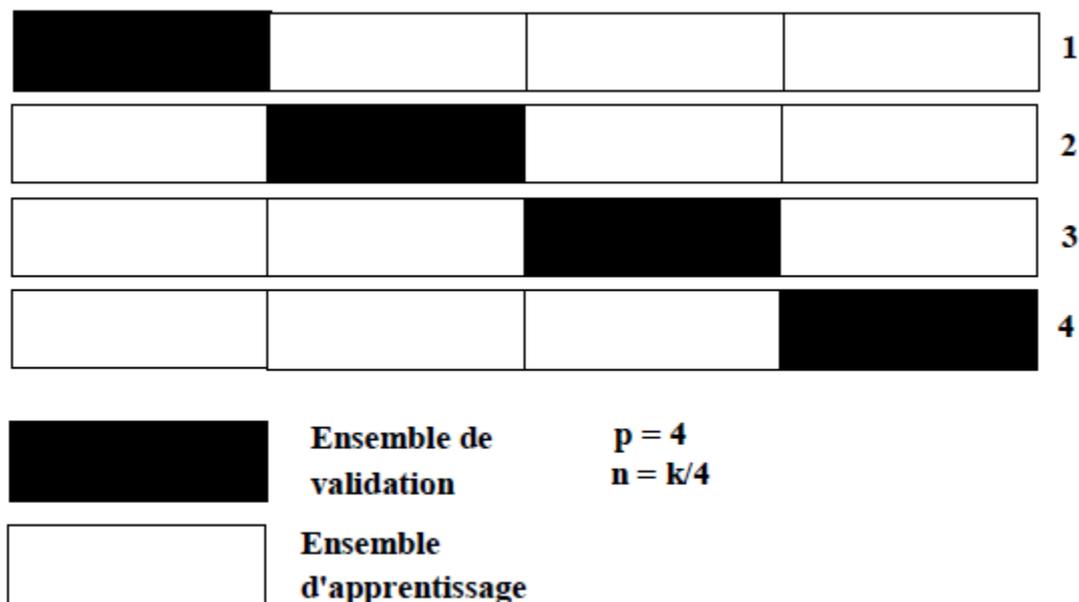


Figure II.3: principe de la validation croisée leave-n-out.

II.8.1.b. La procédure leave-one-out:

Pour un nombre N d'exemples d'apprentissage, on retire à chaque itération un exemple I de l'ensemble d'apprentissage initial. Une série d'apprentissage est réalisée pour les $N-1$ molécules restantes et la molécule retirée est prédite par le modèle formé. La performance des modèles de régression est estimée avec les paramètres statistiques de la validation croisée décrits par les équations ci-dessous [22] :

L'écart type de la validation croisée S_{CV-LOO} :

$$S_{CV-LOO} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_{icv})^2} \quad (\text{eq16})$$

Le coefficient de détermination de la validation croisée Q^2_{CV-LOO} :

$$Q^2_{CV-LOO} = \frac{\sum_{i=1}^N (y_i - \bar{y})^2 - \sum_{i=1}^N (y_i - y_{icv})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (\text{eq17})$$

y_{icv} : Représente la valeur de l'activité de la molécule i calculée par la méthode de la validation croisée,

\bar{y} : La valeur moyenne de l'activité,

y_i : Les valeurs observées de l'activité.

II.8.2 Validation externe :

Un modèle avec des valeurs élevées des indices internes Q^2 (ou R^2_{cv}) n'est pas encore dit valide, par conséquent, la validation interne est nécessaire mais insuffisante.

La puissance prédictive réelle d'un modèle QSAR/QSPR est de tester leur capacité à prédire parfaitement l'activité/propriété des composés à partir d'un ensemble de test externe (composés non utilisés pour le développement du modèle). Le but d'un bon modèle QSAR/QSPR est non seulement de prédire l'activité des composés d'ensemble d'apprentissage, mais aussi de prévoir les activités des molécules de test.

Le modèle QSAR/QSPR est bâti sur l'ensemble d'apprentissage et validé sur l'ensemble de test. La capacité prédictive du modèle est basée sur le coefficient de corrélation R^2 test entre les activités observées et les activités prédites pour l'ensemble de test, la valeur plus élevée de R^2_{test} ($> 0,5$) indique la bonne productivité du modèle [7].

La validation externe consiste à estimer l'activité d'une série de composés dite série de test qui ne sont pas inclus dans la série d'apprentissage. Le principe de cette technique est basée sur le calcul de coefficient de corrélation R^2 (test) et vérifier les critères de Tropsha (Tropsh acriteria) [22][23] :

les critères de Tropsha :

- ❖ **R^2_{pred}** : coefficient de corrélation pour les molécules de la série de test. $R^2_{pred} > 0.7$.
- ❖ **R^2_o** : coefficient de corrélation entre les valeurs prédites et expérimentales pour la série de test.
- ❖ **R'^2_o** : coefficient de corrélation entre les valeurs expérimentales et prédites pour la série de test.
- ❖ **K** : constante de la droite (à l'origine) de corrélation (valeurs prédites en fonction des valeurs expérimentales)
- ❖ **K'** : constante de la droite (à l'origine) de corrélation (valeurs expérimentales en fonction des valeurs prédites). pour Le choix du meilleur modèle on vérifié :

$$R_{pred}^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

$$R^{\circ 2} = 1 - \frac{\sum_{i=1}^{ntest} (\hat{y}_i - y_i^{r\circ})^2}{\sum_{i=1}^{ntest} (\hat{y}_i - \bar{y})^2}, y_i^{r\circ} = K \hat{y}_i$$

$$R'^{\circ 2} = 1 - \frac{\sum_{i=1}^{ntest} (y_i - \hat{y}_i^{r\circ})^2}{\sum_{i=1}^{ntest} (y_i - \bar{y})^2}, \hat{y}_i^{r\circ} = K' y_i$$

$$K = \frac{\sum_{i=1}^{ntest} y_i \hat{y}_i}{\sum_{i=1}^{ntest} \hat{y}_i^2}$$

$$K' = \frac{\sum_{i=1}^{ntest} y_i \hat{y}_i}{\sum_{i=1}^{ntest} y_i^2}$$

- ❖ $R^2 > 0.7$
- ❖ $|R^2 - R_0^2| / R^2 < 0.1$ et $0.85 \leq k \leq 1.15$
- ❖ $|R^2 - R_0'^2| / R^2 < 0.1$ et $0.85 \leq k' \leq 1.15$
- ❖ $|R^2 - R_0^2| \leq 0.3$

II.9. Application de QSAR :

Il existe un grand nombre d'applications des modèles QSAR en milieu industriel, dans le domaine de la recherche universitaire, en économie, en prévision météorologique... etc [12].

Un petit nombre d'utilisations potentielles sont énumérées ci-dessous [12] :

- L'identification rationnelle des nouvelles pistes avec une activité pharmacologique, biocide ou pesticide.
- L'optimisation de l'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.
- Activités biologiques : Anti VIH, Anti malaria, Anti Diabète, Anti Cancer, Anti -oxydante, Anti-inflammatoire
- 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.).

Références :

- [1]Hammett, L.P. The effect of structure upon the reactions of organic compounds Benzene derivatives. J. Am. Chem. Soc. 59, 96-103.1937
- [2]A.R. Katritzky , D.C. Fara, R.O. Petrukhin, Top. Med. Chem. 2002, 1333-1356.
- [3]Hansch, C., Leo, A., Hoekmann, D. Exploring QSAR: hydrophobic, electronic and steric constants. Washington, DC: Am. Chem. Soc.2015.
- [4]Rachid Hamamouchi ; Mohamed Bouachrine ; et al.; Une tentative pratique de la relation quantitative entre la structure, l'activité et la propriété(QSAR/QSPR); T 1 , N 1 (2016)
- [5] S.M, Free, J.W, Wilson, .A mathematical contribution to structure–activity studies. J. Med. Chem., 7, 395-399. 1964
- [6]BARI Zoulikha ; DJAHRA Khadidja ;Etude et développement des techniques QSAR des molécules Anti Alzheimer ; mémoire master. Université Mohamed Khider de Biskra ;19 ;32 ;22;35.2020.
- [7]Samir CHTITA. Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments . Thèse de Doctorat en Chimie. Université de MOULAY ISMAIL .Maroc. (2017).21.26.27.28.53.58
- [8]KEDIDI Naouel ; GUEBBABI Kamilia ; Etude 2D-QSAR DE Dérivés de benzophénone en tant qu'agents antipaludique ; mémoire de mastère ; univ Beskra 2021 ; .30.31.
- [9]Khan AU. Descriptors and their selection methods in QSAR analysis: paradigm for drug design. Drug discovery today , 21(8),(2016),1291-1302.
- [10]Khairidine KRAIM ; ETUDES DE QSAR SUR DES ACTIVITES BIOLOGIQUES UTILISANT DES PRODUITS D'ORIGINES NATURELS ;THES DOCTORAT UNIVER-SITE DE GUELMA ; 2009 ;54
- [11]M. SAIHI Youcef. Etude de la relation quantitative structure-activité inhibitrice des enzymes hydrolytiques : cas des alpha-glucosidases . THESE de DOCTORAT .UNIVERSITE BADJI MOKHTAR - ANNABA. 2015.62.71.75
- [12]Puzyn T.; Leszczynski J.; et al Recent advances in QSAR studies: methods and applications , Springer Science & Business Media. London New York, (2010).
- [13]Guillaume FAYET ;Developpement de modele QSPR pour la prédiction des propriétés d' explosibilité des composes nitro- aromatiques ; thèse de doctorat ; l'UNIVERSITÉ PIERRE ET MARIE CURIE;2010 ;63.69.70.71
- [14]KHAINNAR Mehdi et al . Etude des relations Quantitatives Structure-Activité (QSAR)de plusieurs familles de composés sur des données biologiques .2019 ;8/9)

- [15] Atkins, P.W.; and de Paula, J. Atkins' Physical Chemistry . Oxford University Press. Oxford. (2002).
- [16] Mme OUZLIFI FADIA ; ETUDE QSAR DE LA TOXICITE D'UNE SERIE DE COMPOSES PHENOLIQUES VIS-A-VIS DES ALGUES MARINES ;2019 ;20
- [17] Lorentz H.A. Ueber die Beziehung zwischen der Fortpflanzungsgeschwindigkeit des Lichtes und der Körperdichte. *Annalen der Physik*, 245(4), (1880), 641-665.
- [18] Hansen C.; Telzer B. R., and Zhang L.T. Comparative QSAR in toxicology: examples from teratology and cancer chemotherapy of aniline mustards. *Critical reviews in toxicology*, 25(1), (1995), 67-89.)
- [19] Castellani C. Régionalisation des précipitations annuelles par la méthode de la régression linéaire simple : l'exemple des Alpes du Nord . In: *Revue de géographie alpine*, tome 74, n°4, 1986. 393-403
- [20] BELLIFA, K . Etude des relations quantitatives structure–toxicité des composés chimiques à l'aide des descripteurs moléculaires« Modélisation QSAR » . Thèse de Doctorat. Université ABOU BEKR BELKAÏD DE TLEMCEN. (2015).
- [21] J, Shao . Linear model selection by cross-validation. *Journal of the American statistical Association*, 88(422), (1993), 486-494.
- [22] A. Tropsha, P. Gramatica, and V.K. Gombar, —the importance of Being Earnest: Validation is the Absolute Essential for Successful Application and interpretation of QSPR Models, *QSAR and Combinatorial Sciences*, 22(1), 2003, 69–77
- [23] S.K. Balasubramanian, J.A. Karrh, H. Patwardhan, Audience response to product placements: An integrative framework and future research agenda, *Journal of advertising*, 35 (2006) 115-141)
- [24] Esposito E. X.; Hopfinger A. J et al. Methods for applying the quantitative structure-activity relationship paradigm. In *Chemo informatics* , Humana Press, 275, (2004),131-213.
- [25] J. Jaworska, N.N. Jeliazkova, T. Aldenberg, —QSAR applicability domain estimation by projection of the training set descriptor space: a review, *Alternatives to Laboratory Animals*, 33, 2005, 445–459.
- [26] T.I. Netzeva , A.P. Worth, T. et al; Yang, —Current status of methods for defining the applicability domain of (quantitative) structure–activity relationships, *Alternatives to Laboratory Animals*, 33(2), 2005, 155–173
- [27] I, Tetko. Sushko, et al. Critical assessment of QSAR models of environmental toxicity against *Tetrahymena – pyriformis* : focusing on applicability domain and over fitting by variable selection, *Journal of Chemical Information and Modeling*, 48, 2008, 1733–1746.

[28]P. Gramatica, —Principles of QSAR models validation: internal & external, QSAR and Combinatorial Sciences, 26(5), 2007, 694–701.

Chapitre III

Etude QSAR d'une Série de dérivés de Flavonoïdes

III. 1. Introduction :

Le modèle QSAR (quantitative structure-activité Relationship) étude de corrélation basée sur le calcul statistique, des paramètres physiques et chimiques avec l'activité biologique [1]

La comparaison de l'activité biologique de molécules spécifiques et de leur structure a permis dans de nombreux cas d'établir des relations entre les paramètres structuraux et les propriétés de la molécule. La corrélation de la variance d'activité avec les paramètres a permis d'obtenir un système d'équations qui donne, pour une série chimique donnée et une activité déterminée, une équation de corrélation.

Dans ce travail, nous intéressons à l'étude des propriétés physiques et chimiques des flavonoïdes et de leurs dérivés qui permettent de prédire ces paramètres affectant l'activité biologique, afin de prédire une nouvelle activité biologique. Dans ce chapitre, le modèle final reliant l'efficacité des dérivés flavonoïdes en tant qu'inhibiteurs de l'infection par le VIH et la capacité prédictive de ce modèle est discuté.

III. 2. Matériel et méthodes :

Tout d'abord , les trente-deux molécules de dérivés de flavonoïdes ont été pré-optimisées au moyen de la mécanique moléculaire et du champ de force (MM+) puis les structures obtenues ont été optimisées à l'aide de la méthode semi-expérimentale PM3, le tout avec Hyper Chem 8.0 [2].

Après cela , Le module « Propriétés QSAR » d'Hyper Chem 8.0 a été utilisé pour calculer les paramètres suivants : polarisabilité (Pol), réfraction molaire (MR), coefficient de partage octanol /eau (log P), volume molaire (MV), surface moléculaire (SAG) et le poids moléculaire (MW) et l'énergie d'humidification (HE), ainsi que l'utilisation du logiciel suisse ADME pour calculer les paramètres suivants : Nombre de liaisons rotative (rot), Nombre de liaisons hydrogène accepteurs (N.H-accep) , Nombre de liaisons hydrogène donneurs (N.H-donn), Num. heavy atoms (HA) :Num. arom. heavy atoms (AHA) , Topological polar surface area(TPSA).

La méthode de régression multilinéaire (MLR) a été utilisée pour développer le modèle QSAR à l'aide du logiciel XLSTAT [3]

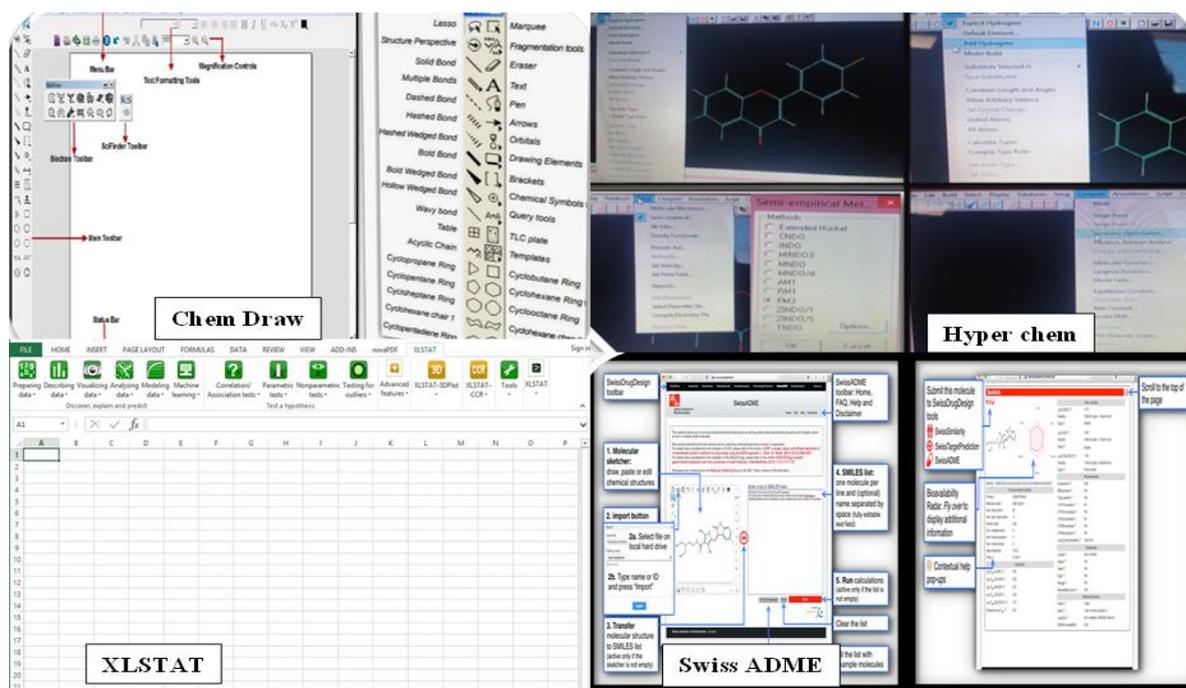


Figure. III 1: les interfaces des programmes utilisés.

III.2.1 Base de données :

De nombreux dérivés de flavonoïdes présentent des propriétés pharmacologiques importantes. Ils sont utilisés dans le traitement et la prévention d'un grand nombre de maladies. L'ensemble de données composite de 32 dérivés de flavonoïdes en tant qu'inhibiteurs du SIDA (VIH-1) a également été obtenu à partir de la littérature [4]. L'activité biologique de ces composés représentés en IC 50 (μ M) a été convertie en p IC 50.

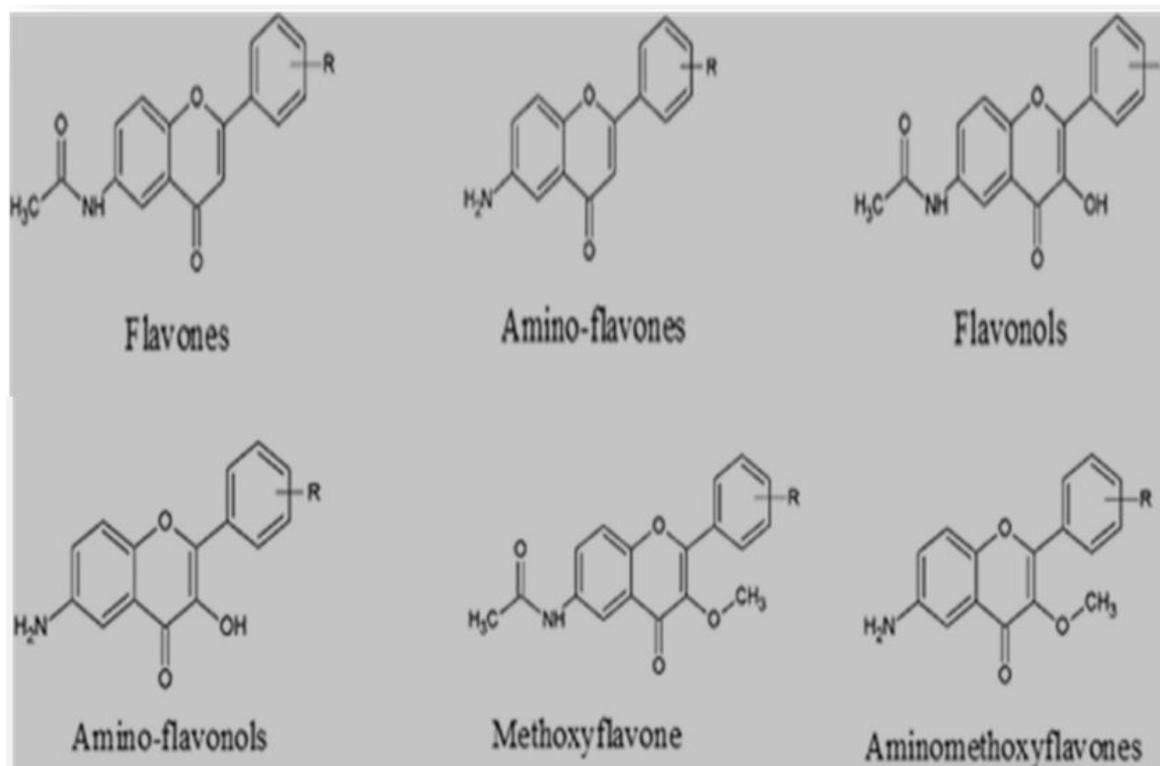
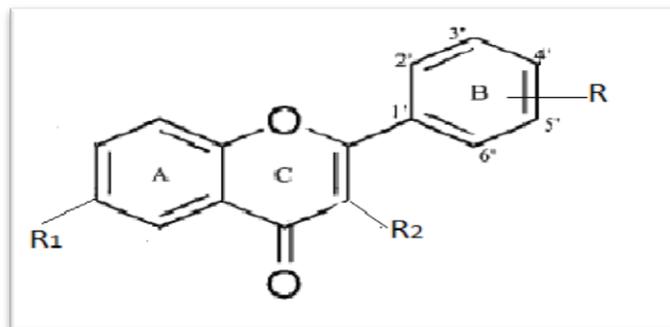
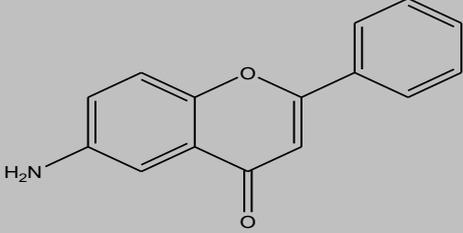
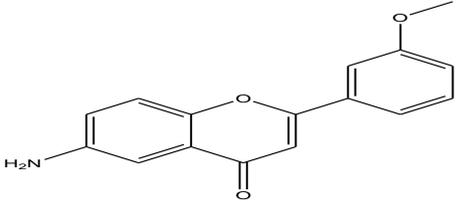
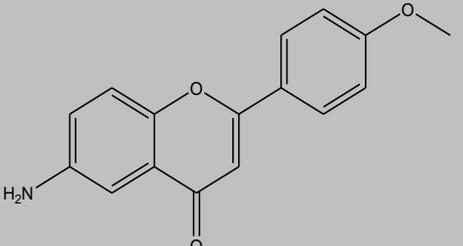
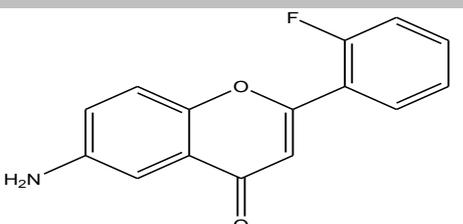
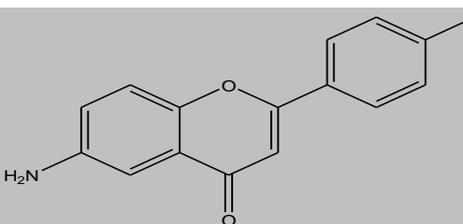
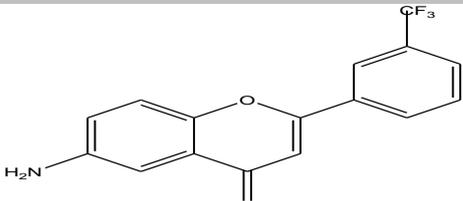


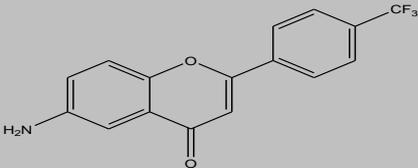
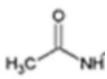
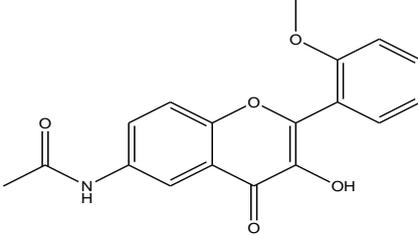
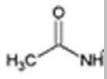
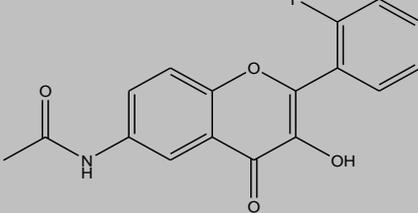
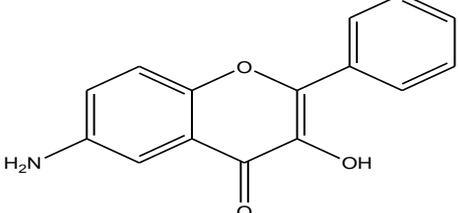
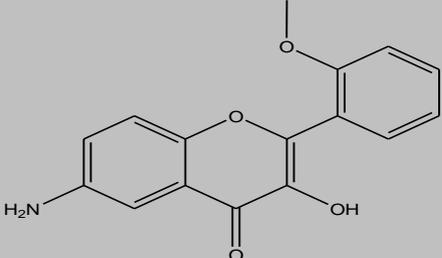
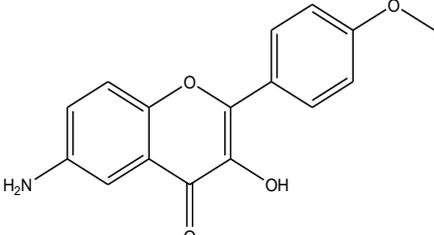
Figure. III 2 : Dérivés de flavonoïdes en tant qu'inhibiteurs du VIH-1.

Tableau. III 1 : les structures chimiques et l'activités biologiques de dérivés flavonoïdes en tant qu'inhibiteurs du VIH-1

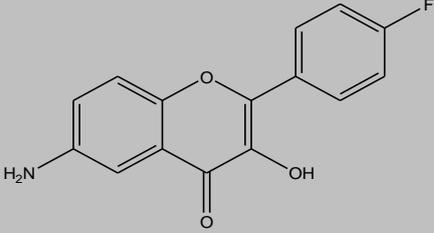
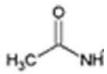
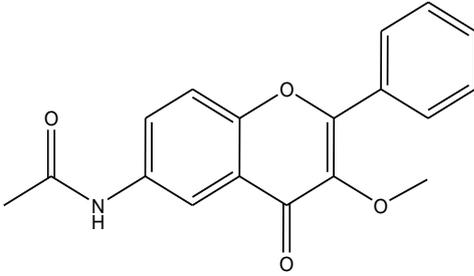
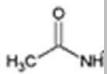
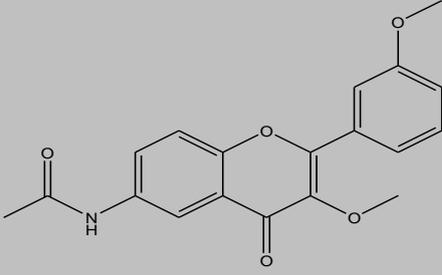
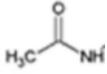
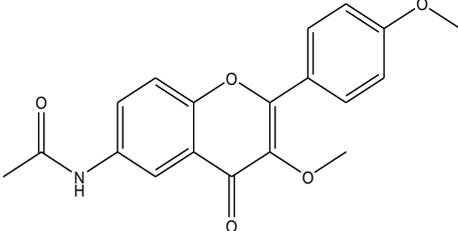
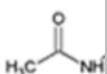
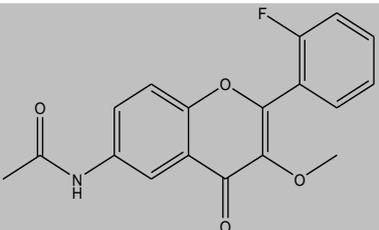
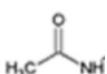
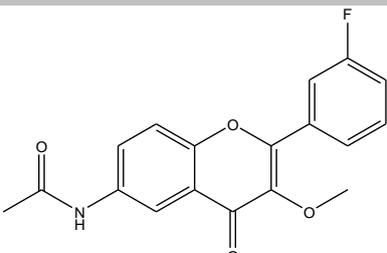
Nombres	R	R1	R2	Molécules	pIC50
1	2'-F		H		3.688
2	4'-F		H		3.692
3	3'-CF3		H		5.00
4	4'-CF3		H		3.482

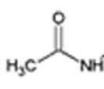
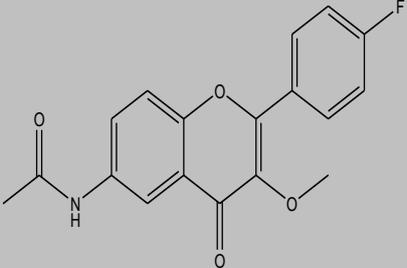
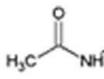
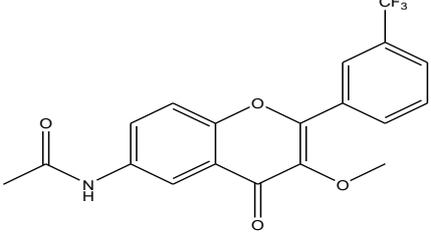
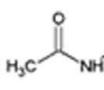
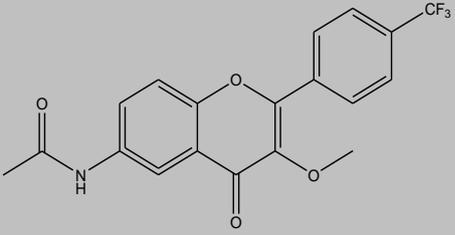
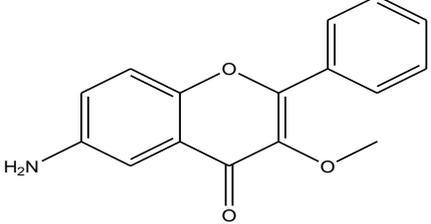
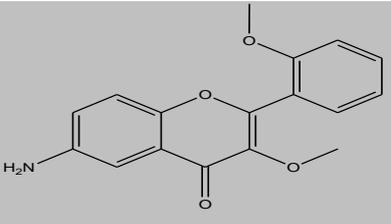
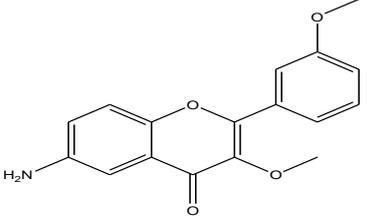
Chapitre III Etude QSAR D'une Série De Dérivés De Flavonoïdes

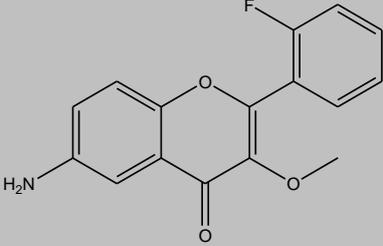
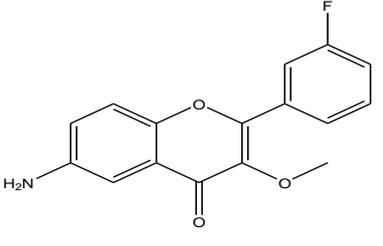
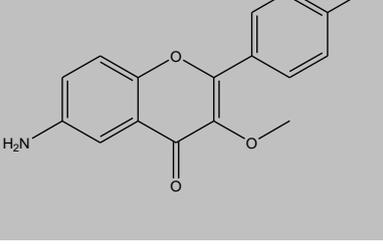
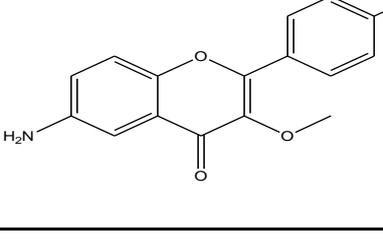
5	H	NH ₂	H		3.68
6	3'-OCH ₃	NH ₂	H		3.607
7	4'-OCH ₃	NH ₂	H		3.619
8	2'-F	NH ₂	H		3.546
9	4'-F	NH ₂	H		3.578
10	3'-CF ₃	NH ₂	H		3.861

11	4'-CF ₃	NH ₂	H	3.564
				
12	2'-OCH ₃		OH	3.415
				
13	2'-F		OH	3.766
				
14	H	NH ₂	OH	3.665
				
15	2'-OCH ₃	NH ₂	OH	3.665
				
16	4'-OCH ₃	NH ₂	OH	3.622
				

Chapitre III Etude QSAR D'une Série De Dérivés De Flavonoïdes

17	4'-F	NH ₂	OH		4.033
18	H		C-0Me		3.7
19	3'-OCH ₃		C-0Me		3.433
20	4'-OCH ₃		C-0Me		3.433
21	2'-F		C-0Me		3.744
22	3'-F		C-0Me		3.679

23	4'-F		C-0Me		3.499
24	3'-CF3		C-0Me		3.697
25	4'-CF3		C-0Me		3.48
26	H	NH ₂	C-0Me		3.899
27	2'-OCH3	NH ₂	C-0Me		3.64
28	3'-OCH3	NH ₂	C-0Me		3.601

29	2'-F	NH ₂	C-OMe		3.631
30	3'-F	NH ₂	C-OMe		4.217
31	4'-F	NH ₂	C-OMe		3.852
32	4'-CF ₃	NH ₂	C-OMe		4.321

III.2.2 Sélection des descripteurs et méthodes de calcul:

Le choix des descripteurs dépend des outils disponibles, de la nature des composés à décrire et des propriétés visées. L'information codée pour un descripteur moléculaire dépend du type de représentation moléculaire utilisée et de l'algorithme choisi pour son calcul. Dans cette partie :

- nous avons étudié huit propriétés physico-chimiques d'une série de dérivés de flavonoïdes à l'aide d'HyperChem 8.0, représentées dans le tableau suivant :
Volume moléculaire(MV), poids moléculaire (masse); coefficient de partage octanol/eau (logP); réfraction moléculaire (réf); polarisabilité (pol); surface moléculaire (SAG); Énergie d'humidification (HE) et moment dipolaire (MD).

- **Tableau. III 2** : les descripteurs obtenu 1

Molécules	M (a mu)	MV (Å ³)	Réf (Å ³)	Pol (Å ³)	SAG (Å ²)	HE (Kcal/mol)	log P	MD Debye
1	297.29	821.49	88.19	30.37	506.26	-6.14	-1.5	0.85
2	297.29	828.46	88.19	30.37	512.58	-6.13	-1.5	1.53
3	297.29	828.46	88.19	30.37	512.58	-6.13	-1.5	1.72
4	347.29	896.8	93.27	32.02	549.95	-5.8	-0.33	1.72
5	237.26	699.75	78.62	26.71	433.48	-9.83	-0.73	3.27
6	267.28	777.4	84.99	29.18	475.18	-11.28	-1.72	4.36
7	267.28	777.31	84.99	29.18	475.24	-11.47	-1.72	3.07
8	255.25	706.08	78.74	26.62	434.9	-9.2	-1.33	2.23
9	255.25	709.16	78.74	26.62	437.5	-9.53	-1.33	2.03
10	305.26	779.14	83.83	28.27	475.17	-9.3	-0.16	3.53
11	321.26	802.26	85.51	28.91	490.49	-14.33	-1.06	1.26
12	325.32	879.51	96.12	33.57	521.77	-9.69	-2.8	2.23
13	313.28	824.47	89.87	31.01	500.41	-10.26	-2.41	1.09
14	253.26	714.28	80.3	27.34	439.15	-14.57	-1.63	3.55
15	283.28	774.99	86.67	29.82	466.67	-13.53	-2.62	5.02
16	283.28	795.09	86.67	29.82	490.16	-16.57	-2.62	3.28
17	271.25	723.52	80.42	27.25	443.33	-14.27	-2.23	3.39
18	309.32	889.21	94.51	32.93	541.23	-7.16	-1.77	0.45
19	339.35	954.91	100.88	35.41	578.52	-8.64	-2.77	1.26
20	339.35	954.2	100.88	35.41	571.89	-8.81	-2.77	1.73
21	327.31	890.85	94.64	32.84	537.5	-6.74	-2.37	1.01
22	327.31	880.83	94.64	32.84	532.95	-6.64	-2.37	2.004
23	327.31	880.74	94.64	32.84	531.91	-6.68	-2.37	1.48
24	377.32	973.09	99.72	34.5	589.92	-6.42	-1.2	2.28
25	377.32	968.08	99.72	34.5	585.97	-6.57	-1.2	3.64
26	267.28	771.23	85.07	29.18	468.41	-10.88	-1.6	3.58
27	297.31	842.51	91.44	31.65	506.84	-10.58	-2.59	4.39
28	297.31	848.52	91.44	31.65	514.39	-12.39	-2.59	4.083
29	285.27	779.91	85.19	29.09	472.62	-10.41	-2.2	4.07
30	285.27	781.21	85.19	29.09	475.03	-10.58	-2.2	3.9

Chapitre III Etude QSAR D'une Série De Dérivés De Flavonoïdes

31	285.27	780.61	85.19	29.09	474.47	-10.59	-2.2	3.57
32	335.28	855.68	90.28	30.74	522.39	-10.23	-1.03	3.79

- Nous avons également étudié six autres propriétés physicochimiques d'une série de dérivés de flavonoïdes à l'aide du programme swiss ADME représenté dans le tableau suivant :

Nombre de liaisons rotative (rot), Nombre de liaisons hydrogène accepteurs (N.H-accept), Nombre de liaisons hydrogène donateurs (N.H-donn), Num. heavy atoms (HA) : Num. arom. heavy atoms (AHA), Topological polar surface area (TPSA).

- **Tableau. III 3** : les descripteurs obtenus 2.

Molécules	rot	N.H-accept	N.H-donn	HA	AHA	TPSA (Å ²)
1	3	4	1	22	16	59.31
2	3	4	1	22	16	59.31
3	4	6	1	25	16	59.31
4	4	6	1	25	16	59.31
5	1	2	1	18	16	56.23
6	2	3	1	20	16	65.46
7	2	3	1	20	16	65.46
8	1	3	1	19	16	56.23
9	1	3	1	19	16	56.23
10	2	5	1	22	16	56.23
11	2	6	2	23	16	76.46
12	4	5	2	24	16	88.77
13	3	5	2	23	16	79.54
14	1	3	2	19	16	76.46
15	2	4	2	21	16	85.69
16	2	4	2	21	16	85.69
17	1	4	2	20	16	76.46
18	4	4	1	23	16	68.54
19	5	5	1	25	16	77.77
20	5	5	1	25	16	77.77
21	4	5	1	24	16	68.54

22	4	5	1	24	16	68.54
23	4	5	1	24	16	68.54
24	5	7	1	27	16	68.54
25	5	7	1	27	16	68.54
26	2	3	1	20	16	65.46
27	3	4	1	22	16	74.69
28	3	4	1	22	16	74.69
29	2	4	1	21	16	65.46
30	2	4	1	21	16	65.46
31	2	4	1	21	16	65.46
32	3	6	1	24	16	65.46

III.2.3 Analyse statistique :

Des modèles QSAR ont été générés en utilisant des régressions MLR, Sur le programme XLSTAT [3].

XLSTAT est la principale solution d'analyse de données et de statistiques pour Microsoft Excel Nous avons choisi ce programme car il possède de nombreuses propriétés:

- ✓ XLSTAT c'est un logiciel très fiable et très stable.
- ✓ nous permet de réorganiser nos données et de les structurer afin de faciliter leur analyse.
- ✓ nous donne accès à un grand nombre de tests. Les résultats des tests comprennent une interprétation explicite.
- ✓ Il offre une large variété de tests statistiques pour que nous permette de tester des hypothèses.
- ✓ nous permet d'évaluer la qualité de nos données, et détermine les erreurs de calcul (écarts + résidus), ce que nous aide d'améliorer notre travail.

III. 3. Résultats et discussions :

III.3.1. Interprétation du modèle QSAR :

Dans cette étude, nous avons tenté de développer le meilleur modèle QSAR pour expliquer l'association entre les paramètres physiques, chimiques et les valeurs pIC50 de l'activité biologique des dérivés de flavonoïdes.

Une série de 32 dérivés de flavonoïdes a été utilisée pour produire le modèle de régression multilinéaire. Différents descripteurs physiques et chimiques ont été utilisés comme variables indépendantes et liés à l'activité biologique à l'aide du programme XLSTAT.

3 modèles de régression linéaire multiple (MLR) avec leur identification dans le tableau ci-dessous :

- **Tableau. III 4** : Modèles sélectionnés et paramètres statistiques des corrélations entre les propriétés moléculaires et l'activité biologique antivirale

Modèle	N	Equation	R	R ²	F	SE
1	32	$\text{pIC50} = -32,8893301804436 - 0,760076895905922 * M + 2,41918889682847E-02 * \text{SAG} - 3,45386707643426E-02 * \text{MV} - 8,34233876335496 * \text{POL} + 5,69512983019181 * \text{ref} + 3,63846278964491E-02 * \text{HE} - 0,106552101160206 * \text{MD} + 12,3321439737176 * \text{N-H accp} + 4,32979687675703 * \text{N-H den} - 0,402987849695985 * \text{TPSA}$	0.867	0.868	303.78	0.09
2	32	$\text{pIC50} = -12,3061573854792 - 0,16106150564563 * M + 7,22072418884878E-02 * \text{SAG} - 5,89876394618915E-02 * \text{MV} - 18,7369441568453 * \text{POL} + 7,01758341047153 * \text{ref} + 0,16373593179667 * \text{HE} - 0,751689252244767 * \text{Log P} - 4,02326665392097E-02 * \text{MD} + 5,5250219834442 * \text{ROT} + 0,156765114544179 * \text{TPSA}$	0.82	0.821	257.4	0.10

3	32	$pIC50 = 768,00743792032 - 6,16549231484062 * M + 1,26040861981453E-02 * SAG + 2,27595908698136E-02 * MV + 17,8001800608584 * POL + 7,13169926753012E-02 * HE + 5,49590186425887 * \text{Log} P + 0,265608528777205 * MD + 115,680309335496 * N - H_{accp} + 39,0148629401773 * ROT - 1,15357532790174 * TPSA$	0.751	0.752	1110.138	13.00
---	----	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------	-------	----------	-------

Sur les 3 équations QSAR, le meilleur modèle QSAR est choisi en fonction de différents critères statistiques tels que :

- ✚ On a une relation forte si $R \in [0.8, 1]$ ou $R \in [-1, -0.8]$;
- ✚ Corrélation modérée si $R \in [0.5, 0.8]$ ou $R \in [-0.8, -0.5]$;
- ✚ Relation faible si $R \in [-0.5, 0.5]$.

Le coefficient de corrélation au carré ($R^2 > 0,6$), qui est une mesure relative de la qualité de l'ajustement, et l'erreur type d'estimation ($SE < 0,3$) est une mesure absolue de la qualité de l'ajustement, et la valeur de Fisher (F), F est le ratio de Fisher et reflète la proportion de variance expliquée par le modèle et la variance due à l'erreur de régression. Des valeurs élevées du test F indiquent que le modèle est statistiquement significatif.[5]

Après analyse de régression linéaire multiple sur logiciel XLSTAT, les meilleures corrélations statistiquement significatives générées avec les paramètres statistiques pertinents (modèle 2) sont présentées ci-dessous.

$$pIC50 = -12,3061573854792 - 0,16106150564563 * M + 7,22072418884878E-02 * SAG - 5,89876394618915E-02 * MV - 18,7369441568453 * POL + 7,01758341047153 * ref + 0,16373593179667 * HE - 0,751689252244767 * \text{Log} P - 4,02326665392097E-02 * MD + 5,5250219834442 * ROT + 0,156765114544179 * TPSA$$

$$R^2 = 0.821 \quad , \quad R^2_{adj} = 0.731 \quad , \quad F = 257.4 \quad , \quad SE = 0.10, \quad N = 32 \quad , \quad P = 10$$

- ❖ Les valeurs $R = 0.82$, $R^2 = 0.821$ nous permis d'indiquer fermement la corrélation entre les différents descripteurs et l'activité biologique étudiée (P-IC50).
- ❖ La valeur de F calculée est importante 257.4 et indique donc que la relation de régression était statistiquement significative [5].
- ❖ SE: est une mesure de variation de la valeur d'un test statistique, d'un échantillon à l'autre il doit être très petits pour considérer le modèle significative.

Dans les équations de modèle 2 , le coefficient négatif de la masse molaire et de volume molaire des dérivés de flavonoïdes entraîne une diminution de l'activité biologique antivirale.

Le coefficient positif de l'énergie d'hydratation et le coefficient négative de Log P indique que les dérivés hydrophiles donnent une bonne activité biologique.

III.3.2. Validation De Modèle

III.3.2.1. validation interne :

Afin de tester la validité de pouvoir prédictif de modèle QSAR sélectionné on utilise la technique de validation croisée LOO (leave-one-out) , présenter les paramètres statistiques calculé pour la validation modèle :

- **Tableau. III 5** : Paramètres de validation croisée.

Model	PRESS	SPRESS	SSY	PRESS/SSY	R ² cv	R ² adj
MLR	0.555307	0.1317320908	3.0428753408	0.1824916684	0.8175	0.731

- ❖ PRESS (carrés de la somme résiduelle prédite) sa valeur étant inférieure à SSY souligne que le modèle prédit mieux que le hasard et peut être considéré comme statistiquement important, SSY (somme des carrés de la valeur de réponse)En outre, pour un modèle raisonnable de QSAR, le rapport PRESS/SSY doit être inférieur à 0,4[6]. Les résultats présentés dans le tableau (III 5) indiquent que ce rapport est 0.1824916684
- ❖ Le modèle QSAR développé présente une faible valeur de Spress (<0.200) ce qui indique que le modèle présente une petite valeur résiduelle entre l'activité biologique observée et prédite. Les résultats présentés dans le tableau (III 5) indiquent que ce rapport est 0.1317320908

❖ les valeurs élevées du R^2CV et R^2adj (dépassent 0,5) sont des critères essentiels pour une meilleure qualification des modèles QSAR [7]. Les résultats obtenus de R^2CV et R^2adj pour ce modèle QSAR est 0.817et 0.731.

✚ L'activité expérimentale, prédite et résiduelle des dérivés de flavonoïdes dans le tableau (III 6) a été déduite par le logiciel XLSTAT. La raison restante est en partie due à l'incertitude dans la mesure expérimentale et en partie à l'imperfection du modèle, donc si l'activité prédite est plus proche de l'activité expérimentale, le modèle peut être viable [8].

- **Tableau. III 6** : Valeurs Expérimentales , Prédites Et Résiduelles e dep-IC50 Des dérivées des flavonoïdes ; état des molécules dans la validation .

N	pIC50 exp	pIC50 pred	résidu	état
1	3.688	3.710	-0.022	Train
2	3.692	3.730	-0.038	Train
3	3.482	3.769	-0.287	Train
4	3.607	3.363	0.244	Train
5	3.619	3.393	0.226	Train
6	3.546	3.866	-0.320	Train
7	3.861	3.790	0.071	Train
8	3.564	3.869	-0.305	Train
9	3.665	3.931	-0.266	Train
10	3.665	3.564	0.101	Train
11	3.622	3.647	-0.025	Train
12	4.033	3.826	0.207	Train
13	3.697	3.616	0.081	Train
14	3.480	3.547	-0.067	Train
15	3.899	3.806	0.093	Train
16	3.640	3.694	-0.054	Train
17	3.601	3.600	0.001	Train
18	3.631	3.737	-0.106	Train
19	3.852	3.820	0.032	Train
20	4.321	4.296	0.025	Train
21	5.000	9.247	-4.247	Test
22	3.680	3.910	-0.230	Test
23	3.578	3.827	-0.249	Test

24	3.415	3.068	0.347	Test
25	3.766	3.506	0.260	Test
26	3.700	3.712	-0.012	Test
27	3.433	3.375	0.058	Test
28	3.433	2.892	0.541	Test
29	3.744	3.544	0.200	Test
30	3.679	3.783	-0.104	Test
31	3.499	3.728	-0.229	Test
32	4.217	3.813	0.404	Test

- D'après la figure III.3, les descripteurs ROT ; TPSA ; SAG ; MR et HE positives montrent que toute augmentation dans les valeurs de ces paramètres entraîne une augmentation de l'activité biologique. Nous notons également que MR a un impact significatif sur l'activité biologique alors que l'énergie d'humidification n'affecte pas significativement l'activité biologique ;

les descripteurs : log P ; MD ; Pol ; MV ; M négative mais cela fonctionne pour augmenter l'activité biologique.

Les descripteurs : N.H donneur et accepteur ; AHA ; HA n'affecte pas l'activité de l'activité biologique

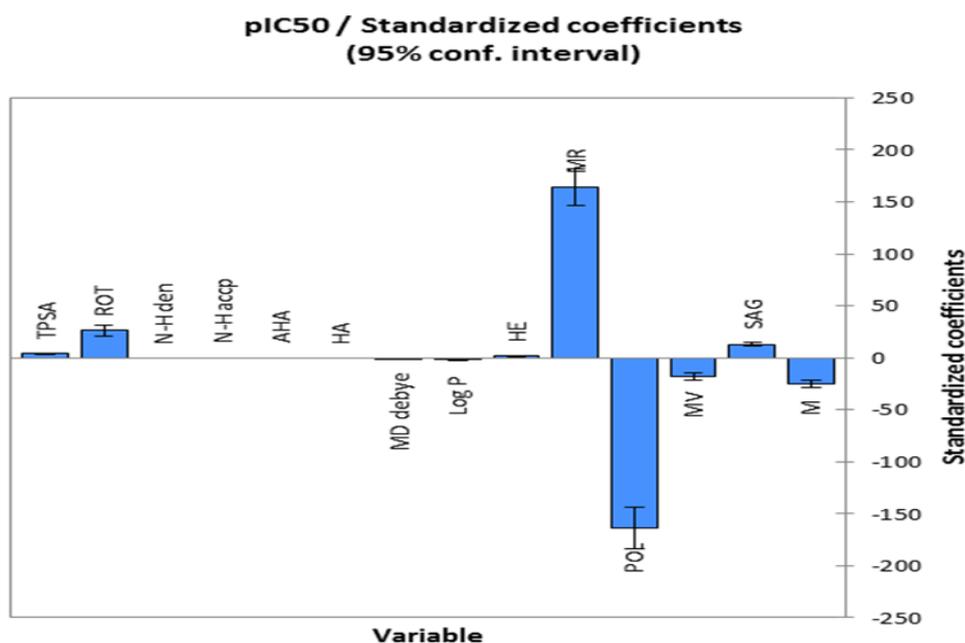


Figure. III3 : Diagramme de la contribution en pourcentage de chaque descripteur dans le modèle pic50 développé expliquant la variation de l'activité.

- La figure (III 4) présente la corrélation entre les valeurs prédites et expérimentales pour l'activité biologique étudiée des nouvelles dérivés des flavonoïdes . Du point de vue des performances statistiques, ce modèle présente de coefficient de corrélation R^2 significatif de valeur. La valeur de R^2 est égale à 0.821 , pour ce modèle. Il indique que ce modèle peut être appliqué avec succès pour prédire l'activité étudiée.

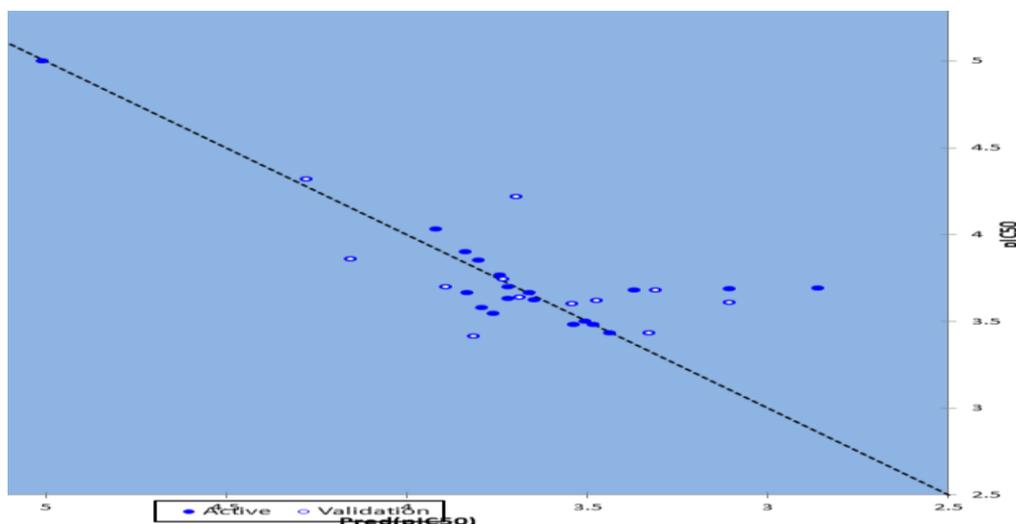


Figure. III 4 : Corrélation entre l'activité biologique expérimentale et prédite

III.3.2.2. Validation externe :

La bonne mesure de prévisibilité du modèle de QSAR obtenu par la validation interne (validation croisée) ne suffit pas. En effet, nous devons en plus généraliser ces prévisions en les appliquant sur un échantillon externe.

Selon Tropsha et ses collaborateurs [9] , un modèle de QSAR ne possède une puissance prédictive acceptable que si les critères de Tropsha sont vérifiés (Comme nous l'avons mentionné dans le deuxième chapitre).

- Tableau .III7 :** Critères de Tropsha

Model	R^2_{test}	R^2_{adj}	K	K'
MLR	0.9074	0.863	0.84	1.1018

R^2 : (de équations 2), est le coefficient de détermination entre les valeurs observées et celles prédits par le modèle (seulement pour l'ensemble d'apprentissage : 12 molécules).

Chapitre III Etude QSAR D'une Série De Dérivés De Flavonoïdes

K, K' : sont les pentes des lignes de régression qui passe par l'origine.

Selon ce tableau(III.7), toutes les normes MLR Tropsha sont vérifiées .Les résultats obtenus par MLR satisfaisant pour prédire l'activité des dérivés flavonoïdes en utilisant spécifications suggérées.

Références :

- [1] V. N. Viswanadhan, M.R.Reddy, R .J.Bacquet et M.D.Erion, *Journal of Computational Chemistry*, 14, 1993, 1019.
- [2] HyperChem v8. Molecular Modelling System, Hypercube Inc., 1115 NW 4th Street, Gainesville, FL 32601, USA, 2009.
- [3] www.xlstat.com
- [4] G. Casano. , A, Dumètre. , C ,Pannecouque , . et al . Anti-HIV and antiplasmodial activity of original flavonoid derivatives. *Bioorganic Med. Chem.* 18 (16), 6012–6023 . 2010.
- [5] K.J. Sanmati, J. Rahul, S. Lokesh, K. Y. Arvind, *Journal of Chemical and Pharmaceutical Research*, 4, 3215- 3223, 201
- [6] Reenu, & Vikas. Exploring the role of quantum chemical descriptors in modeling acute toxicity of diverse chemicals to *Daphnia magna*. *Journal of Molecular Graphics and Modelling*, 61, 89-101(2015)
- [7] S.O. Podunavac-Kuzmanović, D.D. Cvetković, D.J. Barna, QSAR analysis of 2-amino or 2-methyl-1-substituted benzimidazoles against *Pseudomonas aeruginosa*, *Int. J. Mol. Sci.*, 10, 1670-1682, 2009
- [8] H. Gonzalez-Diaz, E. Olazabal, L. Santana, et al, QSAR study of anticoccidial activity for diverse chemical compounds: Prediction and experimental assay of trans-2-(2-nitrovinyl) furan, *Bioorg. Med. Chem.*, 15, 962-968, 2007
- [9] A. Golbraikh .; and A Tropsha. Beware of q^2 !. *Journal of Molecular Graphics and Modelling* , 20(4), (2002), 269–276
- [10] Kadrilege et Datacross. CHEMDRAW 18.0, LE DESSIN MOLÉCULAIRE PAR EXCELLENCE [enligne] <https://ritme.com/fr/logiciels/chemdraw-chemoffice/> (consulté le 03/07/2019)

Conclusion générale

Conclusion générale

La recherche d'une relation entre la structure chimique et l'activité biologique ou d'autres propriétés est d'un grand intérêt pour l'industrie pharmaceutique. Connues sous le nom de relation structure-activité quantitative (QSAR), elles sont utilisées pour prédire les activités de molécules en se basant uniquement sur leurs structures chimiques. Afin d'obtenir de bonnes relations prédictives, il est nécessaire de découvrir et d'utiliser le jeu particulier de descripteurs moléculaires ayant de bonnes corrélations avec l'activité biologique cible.

Dans ce travail, nous avons appliqué des méthodes de chimie computationnelle à une série de 32 dérivés de flavonoïdes. Diverses méthodes de modélisation moléculaire ont été utilisées dans notre travail. Les molécules ont été utilisées pour étudier les propriétés physiques et chimiques des dérivés de flavonoïdes.

Les modèles QSAR ont été générés à l'aide de la méthode de régression multilinéaire (MLR). Les modèles obtenus peuvent être utilisés pour prédire l'activité spécifique des dérivés flavonoïdes et la production d'inhibiteurs du VIH tels que l'AZT.

La capacité prédictive des modèles obtenus a été confirmée par la méthode de validation LOO. Une forte corrélation a été observée entre les valeurs expérimentales et prédites des activités biologiques p-IC₅₀, indiquant la validité et la qualité des modèles QSAR obtenus. Dans le cadre des travaux de recherche, le présent travail comprend une étude fondamentale et originale sur les dérivés de flavonoïdes pour prédire l'activité biologique des composés étudiés.

Une étude quantitative de la relation structure-activité (QSAR) a été réalisée sur 32 molécules de dérivés de flavonoïdes, et une régression linéaire multiple (MLR) a été utilisée pour déterminer la relation entre les descripteurs moléculaires et l'activité antivirale des dérivés de flavonoïdes. La capacité prédictive du modèle a été confirmée par $R^2 = 0.821$, $R^2 \text{ test} = 0.9074$, et $F = 257.4$. La validation croisée de la méthode LOO nous montre une forte corrélation entre les valeurs expérimentales et prédites des activités biologiques, indiquant la validité et la qualité des modèles QSAR obtenus.

Mot Clés: antivirale, dérivées de flavonoïdes ,inhibiteur de VIH , QSAR, MLR.

RESUME

Dans ce travail, une relation quantitative structure-activité (QSAR) a été réalisée pour prédire les inhibiteurs du VIH d'une série de 32 dérivés de flavonoïdes. La régression linéaire MLR a été utilisée ; Où les modèles MLR ont montré des paramètres de validation internes et externes satisfaisants (R^2 ajusté = 0,731, R^2 test = 0,907) qui ont satisfait aux critères d'acceptation du modèle. Les modèles ont été rigoureusement validés par validation croisée. Les modèles montrent que les activités antivirales des dérivés flavonoïdes dépendent des descripteurs logP , ref , MD, HE, TPSA. Ces indices peuvent être utilisés pour concevoir de nouveaux inhibiteurs du VIH et prédire la valeur de leurs activités antivirales sur la base des équations MLR.

Mots clés : VIH /SIDA , inhibiteur de VIH , dérivés de flavonoïde , QSAR, MLR.

Abstract

In this work, a quantitative structure-activity relationship (QSAR) was performed to predict HIV inhibitors of a series of 32 flavonoid derivatives. MLR linear regression was used. Where the MLR models showed satisfactory internal and external validation parameters (adjusted $R^2 = 0.731$, test $R^2 = 0.907$) that met the model acceptance criteria. The models have been rigorously validated by cross-validation. The models show that the antiviral activities of the flavonoid derivatives depend on the descriptors logP , ref, MD, HE, TPSA. These indices can be used to design new HIV inhibitors and predict the value of their antiviral activities based on MLR equations.

Keywords: HIV / AIDS , HIV inhibitors , flavonoid derivatives , QSAR, MLR

المخلص

في هذا العمل تم اجراء علاقة كمية بين التركيب والنشاط (QSAR). للنتبؤ بمثبطات فيروس نقص المناعة البشرية لسلسلة من 32 جزئي لمشتقات الفلافونويد. تم استخدام الانحدار الخطي (MLR), حيث اظهرت نماذج MLR معلمات تحقق داخلية وخارجية مرضية (R^2 ajusté = 0,731, R^2 test = 0,907) التي تفي بمعايير قبول النموذج. تم التحقق من صحة النماذج بدقة من خلال التحقق المتقاطع. توضح النماذج ان الانشطة المضادة للفيروسات لمشتقات الفلافونويد تعتمد على الواصفات logP , ref , MD, HE, TPSA. يمكن استخدام هذه المؤشرات لتصميم مثبطات جديدة لفيروس نقص المناعة البشرية والتنبؤ بقيمة انشطتها المضادة للفيروسات بناء على معادلات MLR .

الكلمات المفتاحية: فيروس نقص المناعة البشرية : الإيدز ، مثبطات فيروس نقص المناعة البشرية ، مشتقات الفلافونويد، QSAR

MLR ،