

République Algérienne Démocratique et Populaire.
Ministère de l'Enseignement Supérieur et de la recherche scientifique.
Université des Sciences et de la Technologie Houari Boumediene.
USTHB –Alger-
Faculté de chimie.



THESE

Présentée pour l'obtention du diplôme de DOCTORAT

En : **Chimie**

Spécialité : **Chimie Physique et Théorique.**

Par

Yamina BELMILOUD.

Sur le sujet

Modélisation théorique de la séparation d'énantiomères dans un mélange racémique.

Soutenue publiquement le 12 Octobre 2009, devant le jury composé de :

M^r	C. CHELGHOUM	Professeur (U.S.T.H.B.)	Président.
M^r	M. BRAHIMI	Professeur (U.S.T.H.B.)	Directeur de Thèse.
M^r	D. CHERQAOU	Professeur (U. MAROC)	Examineur.
M^r	B. TANGOUR	Professeur (U. TUNIS)	Examineur.
M^r	A. RAHMOUNI	Professeur (U. SAÏDA)	Examineur.
M^{lle}	F. TAZEROUTI	Professeur (U.S.T.H.B.)	Examineur.

REMERCEMENTS

*Ce travail a été effectué au Laboratoire de Physico-chimie Théorique et Chimie Informatique de la faculté de chimie à l'Université des Sciences et de la Technologie HOUARI BOUMEDIENNE (U.S.T.H.B.) sous la direction scientifique de Monsieur **Meziane BRAHIMI**, Professeur de chimie à l'U.S.T.H.B. Qu'il me soit permis de le remercier très sincèrement et de lui exprimer ma reconnaissance pour l'aide précieuse, les conseils éclairés et les encouragements qu'il n'a cessé de me prodiguer, pour avoir cru en moi et pour m'avoir accordé sa confiance à de multiples reprises.*

*J'adresse mes plus chaleureux et respectueux remerciements à Monsieur, le Professeur **C. CHELGHOUM** pour l'honneur qu'il me fait d'avoir bien voulu accepter la présidence du jury de cette thèse.*

*Je suis très heureuse de témoigner ma gratitude à Monsieur **D. CHARQAoui**, Professeur à l'Université Cadi Ayyad, Marrakech (Maroc) pour sa disponibilité, ses encouragements et son aide précieuse dans le domaine des réseaux de neurones artificiels.*

*Je tiens à exprimer ma profonde gratitude à Monsieur **B. TANGOUR**, Directeur de recherche à l'université de Tunis, pour sa disponibilité sans limite, pour ses précieux conseils et pour ses nombreuses discussions très enrichissantes.*

*J'adresse également mes sincères remerciements à Monsieur le Professeur **A. RAHMOUNI** de l'Université de Saïda pour son aide et d'avoir bien voulu juger ce travail.*

*J'exprime mes vifs remerciements au professeur **F. TAZEROUTI** qui a bien voulu apporter sa contribution à ce travail par les critiques du spécialiste en chromatographie.*

J'exprime mes remerciements à tous mes collègues du laboratoire ainsi qu'aux collègues du C.R.A.P.C. et de l'Université de BOUMERDES pour l'atmosphère amicales que j'ai trouvé parmi eux.

Pour leur amitié sans faille, pour leur soutien moral, et pour leurs encouragements constants, je remercie tous mes amis, en particulier Amel, Nouara et Sihem.

Je dédie ce travail, très particulièrement à mes chers parents qui m'ont encouragé et soutenue tout au long de ces longues années.

Enfin, je tiens à remercier mon époux, ZIAD, pour avoir toujours été à mes côtés pendant ces années, pour avoir partagé mes doutes et mes espoirs, pour m'avoir encouragée et soutenue.

Pardon à mes petites filles, Melissa et Manal, pour avoir pris beaucoup de leur temps.

*« La réussite des enfants est un devoir sacré des parents.
La réussite des élèves est une fierté sans limite pour les formateurs. »
Meziane BRAHIMI.*

Le temps que j'ai mis à préparer cette thèse m'a permis de me plonger dans un travail multidisciplinaire et d'aborder plusieurs aspects et plusieurs domaines. Parfois indécise, ainsi j'ai pu satisfaire ma curiosité, même si le fait d'être confronté à plusieurs disciplines m'a parfois enlevé l'occasion de m'intéresser à chacune d'elles autant qu'elles le méritent ou que je l'aurais souhaité.

Y. BELMILOUD

Sommaire

INTRODUCTION GENERALE.....	9
Chapitre I Généralités.....	5
Partie 1: Relation Quantitative Structure Activité.....	6
1.1 Choix de l'échantillon à étudier.....	6
1.2 Description de l'activité.....	8
1.2.1 Activité Biologique.....	8
1.2.2 Réactivité chimique.....	8
1.2.3 Propriété Physique.....	8
1.2.4 Rétention chromatographique.....	9
1.3 Description des structures moléculaires.....	9
1.3.1 Les descripteurs moléculaires.....	9
1.4 Traitement mathématique des données.....	12
1.4.1 Méthode linéaire :.....	12
1.4.1.1 Régression multilinéaire (RML).	12
1.4.1.2 Analyse en Composantes Principales (APC).	13
1.4.1.3 Analyse discriminante (AD).	13
1.4.1.4 Méthode des moindres carrées (PLS).....	13
1.4.2 Méthode non linéaire :.....	13
1.4.2.1 Réseaux de Neurones Artificiels (RNA).	14
1.4.2.2 Algorithme Génétique (AG).	14
Conclusion.....	17
Partie 2: Les Réseaux de Neurones Artificiels.....	19
2.1 Historique.....	19
2.2 Les Réseaux de Neurones Artificiels.....	20
2.3 Le neurone.....	20
2.3.1 Le neurone biologique.....	20
2.3.1.1 Anatomie du neurone :.....	20
2.3.1.2 Le fonctionnement des neurones :.....	21
2.3.2 Le neurone formel.....	22
2.3.2.1 La fonction d'activation ou transfert.....	24
2.3.3 Le Réseau de neurones.....	24
2.3.3.1 Les Réseaux "FEED-FORWARD".....	24
Le perceptron monocouche :.....	25
Le perceptron multicouches :.....	25
2.3.3.2 Les réseaux à fonction radiale :.....	27
2.3.3.3 Les Réseaux "FEED-BACK".....	27
2.3.4 La phase d'apprentissage.....	28
2.3.5 La phase de prédiction.....	28
2.3.6 Applications.....	29
Partie 3: Le SIDA.....	33
3.1 Historique de la maladie.....	33
3.2 L'infection par le VIH.....	34
3.3 Les cellules cibles.....	35
3.4 Structure d'une particule virale.....	36
3.5 Le génome du VIH-1.....	37
3.5.1 Structure générale.....	37
- Les séquences LTR.....	37
- gag.....	38
- pol.....	38
- env.....	38

- Les gènes « annexes ».....	39
3.6 Le cycle de vie du VIH	40
3.6.1 Reconnaissance de la cellule hôte	41
3.6.2 Pénétration dans la cellule.....	41
3.6.3 Rétrotranscription du génome viral.....	42
3.6.4 Intégration dans le génome de la cellule hôte	42
3.6.5 Transcription du génome et expression des protéines virales	43
3.6.6 Formation de nouvelles particules virales.....	44
3.6.7 Maturation des particules virales	45
3.7 Les thérapies anti-VIH	45
3.7.1 Les vaccins	45
3.7.2 Les chimiothérapies	46
3.7.2.1 <i>Inhibition de la fusion :</i>	46
3.7.2.2 <i>Inhibiteur de la transcriptase inverse:</i>	46
- Les inhibiteurs nucléosidiques de la transcriptase inverse (INTI) :	47
- Les inhibiteurs non nucléosidiques de la transcriptase inverse (INNTI) :.....	47
3.7.2.3 <i>Inhibition de l'intégrase</i>	47
3.7.2.4 <i>Inhibition de la transcription du génome virale</i>	47
3.7.2.5 <i>Inhibition de la protéase</i>	48
Chapitre II Etude théorique et RQSA des TIBO (INNTI).....	52
Partie 1: Etude Théorique de l'effet de la contraction et de l'extension dans les TIBO.	54
Méthode de calcul	55
La configuration du Carbone C16 dans 6MRC15 :	56
1.1 Effet de la contraction du cycle à 7-sommets dans les TIBO.	57
1.1.1 Géométrie:.....	57
1.1.2 Charges:	63
1.1.3 Les propriétés électroniques et physico-chimiques.	65
1.1.4 La configuration Butterfly –like:	67
1.2 Effet de l'extension du cycle à 7-sommets dans les TIBO.	71
1.2.1 La configuration Butterfly –like :	71
1.2.2 Les propriétés électroniques et physico-chimiques	74
Conclusion	75
Partie 2: Etude des RQSA des dérivés TIBO par les RNA	78
Introduction:.....	78
Les Modèles de la littérature :	78
- Le premier Modèle de la littérature établie par la MLR.....	78
- Le premier Modèle de la littérature établie par les Réseaux de neurones artificiels.....	79
2.1 Modèle RQSA des dérivés TIBO établie dans ce présent travail	79
2.1.1 Calcul et sélection des descripteurs moléculaires	81
2.1.1.1 La sélection des descripteurs au moyen du programme Codessa	81
2.1.1.2 Addition de quelques descripteurs pris de la littérature.....	83
2.1.2 Application des RNA à l'étude des RQSA des dérivés TIBO.....	84
2.1.2.1 Caractéristiques des RNA	84
2.1.3 Résultats et discussion	86
2.1.3.1 <i>Phase d'apprentissage</i>	86
2.1.3.2 <i>Phase de prédiction</i>	88
2.1.3.3 Contribution des variables à la RQSA	89
Conclusion	92
Chapitre III Application des RNA à la Recherche de RQSA des Flavonoïdes.....	94
Introduction.....	95
Base de données.	98
Méthode de calcul.	98
La détermination du nombre Maximum de descripteurs	99

Résultat et discussions.	100
1- La construction de modèles linéaire.	100
1-1 Premier modèle construit en employant seulement les descripteurs du programme CODESSA.....	100
1-2 Deuxième modèle MLR construit en utilisant les descripteurs du programme CODESSA et les descripteurs donnés dans la littérature.	102
2- La construction du modèle non linéaire.....	104
2-1 Application de la méthode RNA au modèle 2	104
2-1-1 Phase d'apprentissage	104
2-1-2 Phase de prédiction	105
Conclusion	107
Chapitre IV Relations Quantitatives Structures-Rétention Chromatographique (RQSR) d'une série d'énantiomères.	110
Introduction.	111
Définitions des différentes grandeurs de rétention:	114
1- Rappel de la partie expérimentale.....	116
- Synthèse des différentes phases stationnaires chirales.	117
2- Recherche de Relations Quantitatives Structure – Rétention Chromatographique des solutés racémiques et des produits pharmaceutiques optiquement actifs séparés par les phases stationnaires chirales A, B, C, D, E et F.....	121
2.2 Base de données	121
2.3 Calculs des descripteurs	123
2.4 La détermination du nombre Maximum de descripteurs	123
2.5 Application de la Régression Multi linéaire.	124
2.5.1 Méthodologie Utilisée.....	124
2.5.2 Résultats et Discussions.	125
Conclusion.	129
Conclusion Générale	131

Abréviations :

RQSA : Relations Quantitatives Structure-Activité.

MLR : Méthodes de Régression Linéaire.

AD : Analyse Discriminante.

ACP : Analyse en Composantes Principales.

RNA : Réseaux de Neurones Artificiels.

AG : Algorithmes Génétiques.

MoRSE : Molecule Representation of Structures based on Electron diffraction.

RBF : Radial Basic Functions.

VIH : Virus de l'Immunodéficience Humaine.

SIDA : Syndrome de l'Immunodéficience Acquise.

IN : Intégrase.

PR : Protéase.

FDA: Food and Drug Administration.

HAART : Highly Active AntiRetroviral Therapy.

LTR : Long Terminal Repeat.

TI : Transcriptase Inverse.

INTI : Les Inhibiteurs Nucléosidiques de la Transcriptase Inverse.

INNTI : Les Inhibiteurs Non Nucléosidiques de la Transcriptase Inverse.

AZT : (3'-azido-3'deoxythymidine.

ddnTP : didéoxynucléoside triphosphate.

TIBO : 4,5,6,7-Tetrahydro-5-methylimidazo[4,5,1-jk][1,4] benzodiazepin-2(1H)-ones.

HF : Hartree Fock (HF).

DFT : Théorie de la Fonctionnelle Densité (DFT).

B3LYP : Becke Lee-Yang-Parr.

HOMO : Highest Occupied Molecular Orbital).

LUMO : Lowest Unoccupied Molecular Orbital.

CPG : La chromatographie en phase gazeuse.

CLHP : La chromatographie en phase liquide à haute performance.

INTRODUCTION

GENERALE

Un médicament est une substance ou une composition présentée comme possédant des propriétés curatives, préventives ou administré en vue d'établir un diagnostic. Il est plus souvent destiné à guérir et à soulager des maladies humaines ou animales. Lors de la synthèse de ce dernier deux énantiomères peuvent être obtenus s'il contient au moins un carbone asymétrique.

Deux énantiomères ont des réactivités chimiques identiques à d'autres molécules. Mais au niveau biologique, les deux énantiomères peuvent avoir des effets physiologiques différents, voir antagonistes. Cela s'explique par le fait que les systèmes biologiques sont eux-mêmes des énantiomères purs et interagissent différemment avec les deux énantiomères. De ce fait, l'analyse et la séparation des énantiomères sont capitales pour l'avancée des recherches dans le domaine pharmacologique.

Bien que plus d'une centaine de phases stationnaires chirales soient disponibles actuellement pour la séparation des énantiomères, de nombreux mélanges racémiques demeurent sans résolution. Par conséquent, le développement de nouvelles phases stationnaires reste un domaine d'actualité très coûteux en temps et en argent. Actuellement et avec l'avancé et le développement considérable de l'outil informatique et computationnel, plusieurs méthodes peuvent être utilisées pour améliorer et accélérer les processus de découverte dans ce domaine. Parmi ces méthodes nous trouvons les méthodes de recherche des **Relations Quantitatives Structure-Activité (RQSA)** - (Quantitative Structure Activity Relationships)-.

Les RQSA permettent de concevoir, classer, trier des molécules, ainsi que calculer ou prévoir leurs propriétés à partir de modèles mathématiques. Le terme activité est un terme général qui peut désigner une activité biologique, une propriété physique, une réactivité chimique ou même une rétention chromatographique. Pour effectuer une étude de relations structure-activité on doit disposer d'un ensemble de molécules dont on connaît à la fois la structure, donnée sous forme d'un ensemble de propriétés structurales et l'activité biologique de chacune d'elles.

Les relations RQSA entre la structure et l'activité sont le résultat d'une analyse statistique. Elles visent à mettre en évidence une corrélation entre les propriétés structurales et l'activité biologique. Si une telle relation est établie, on obtient une équation qui relie mathématiquement les influences réciproques de ces paramètres pris en considération par le modèle obtenu. En d'autres termes, les RQSA recherchent toutes les propriétés nécessaires

pour construire une structure géométrique de la clé (correspondant à celle de la molécule active) qui pourra s'ajuster dans la serrure que constitue le site actif de la cible biologique.

L'objectif est de comprendre, dans un premier temps, les forces gouvernant l'activité d'une série de dérivés afin d'orienter de manière efficace les synthèses des molécules non encore testées. En effet, un des buts principaux des RQSA concerne la compréhension et l'explication des mécanismes d'action des molécules étudiées. Au moyen de ces études, on essaie d'établir des relations entre la structure chimique et l'activité biologique d'une molécule. Les informations que l'on en retire sont fondamentales à la fois pour connaître le mode d'action des molécules et pour concevoir des analogues structuraux plus actifs.

Grâce au développement de l'informatique, plusieurs méthodes d'analyse de données peuvent être utilisées, comme les *Méthodes de régression : linéaire et non linéaire*, *l'analyse discriminante*, *l'analyse en Composantes Principales*, *les Réseaux de Neurones Artificiels (RNA)* et *les Algorithmes Génétiques (AG)*.

Les Réseaux de Neurones Artificiels (RNA) ont été utilisés et se sont montrés efficaces dans divers domaines, notamment celui de la chimie. Dans ce dernier, les applications sont de plus en plus nombreuses. Les RNA visent à reproduire, sous forme de programme informatique, les mécanismes d'apprentissage du cerveau humain.

Notre propos dans ce travail est de contribuer à la recherche de la modélisation de la séparation des énantiomères.

Le mode d'administration d'un médicament joue très souvent un rôle déterminant dans l'expression de l'activité pharmacologique. L'encapsulation du principe actif dans une molécule hôte constitue une des approches les plus récentes et le plus prometteuses du transport des médicaments.

De nombreuses molécules minérales (zéolithes, kaolinites, ...) ou organiques (éthers-couronnes, cyclophanes, cyclodextrines, ...) présentent une "cavité" leur permettant de donner naissance à des complexes d'inclusion. Dans ce processus, la molécule "hôte" admet à l'intérieur de sa cavité une ou plusieurs autres molécules "invitées" sans qu'aucune liaison covalente ne s'établisse. La stabilité du complexe ne repose donc que sur la qualité de l'adaptation entre les partenaires. Ces propriétés de "pièges moléculaires" sont mises à profit dans de nombreux domaines pour réaliser des échangeurs d'ions, des catalyseurs ou pour séparer les composants de mélanges complexes.

Les cyclodextrines sont des molécules "hôtes" naturelles obtenues par dégradation enzymatique de l'amidon. Elles se présentent sous forme d'oligomères cycliques du glucose et comportent de 6 à 12 unités. Les plus abondantes sont les hexamères (α -cyclodextrine), heptamères (β -cyclodextrine) et octamères (γ -cyclodextrine). La **Figure 1** montre les différentes structures des cyclodextrines. Tous les groupements polaires (hydroxyles -OH) sont localisés à l'extérieur, l'ensemble délimitant une cavité relativement hydrophobe. Ce caractère amphiphile permet aux cyclodextrines d'inclure dans leur cavité des molécules hydrophobes pour former des complexes d'inclusion solubles dans l'eau. Leur caractère biodégradable les prédispose à des applications importantes dans les domaines agro-alimentaires et pharmaceutiques. L'encapsulation dans les cyclodextrines permet en effet de protéger des molécules fragiles ou d'assurer leur libération lente et contrôlée.

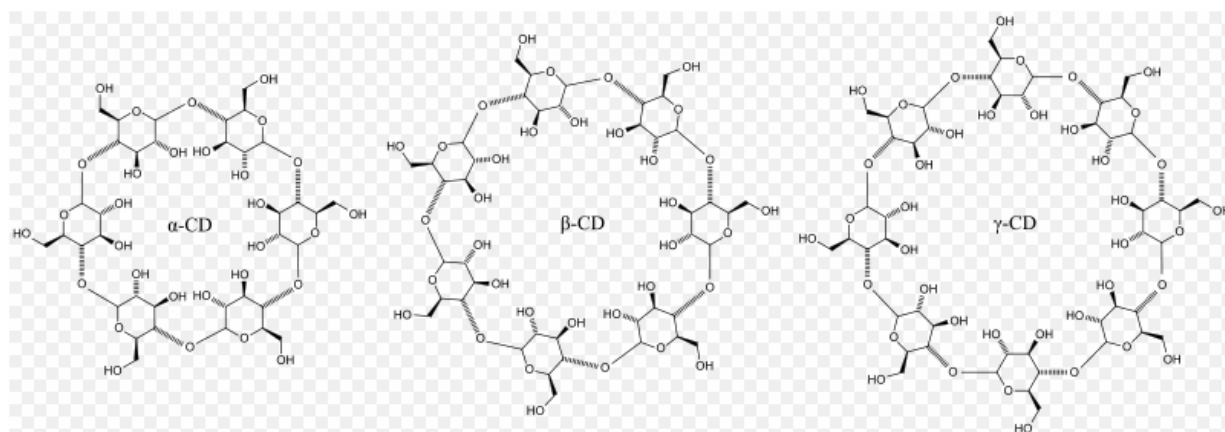


Figure 1: Structure de l' α -, β - et γ -cyclodextrines.

Plusieurs tentatives théoriques, de modéliser la séparation énantiosélectives en se basant uniquement sur la structure de la molécule isolée, sans tenir compte du complexe Droque - Récepteur et de son environnement, ont échoué. Dans ce cas, il est nécessaire de faire appel au docking moléculaire. Cette technique nécessite des bases de données assez larges et des moyens de calculs et de logiciels que nous n'avons pas pu acquérir. Nous nous sommes donc orientés, toujours dans le cadre de l'étude des molécules isolées, vers la recherche de paramètres physico chimiques aidant à expliquer l'activité biologique ou pharmacologique d'une molécule. La connaissance de ces descripteurs physico chimique corrélés à l'activité peuvent d'une part nous renseigner sur les éventuelles différences qui existeraient entre

deux énantiomères et d'autre part prévoir de nouvelles structures moléculaire susceptibles d'avoir des activités meilleurs.

C'est dans cette perspective que ce travail a été fait. Il comporte en plus de cette introduction générale, un premier chapitre dans lequel nous présentons, dans une première partie, les méthodes RQSA usuelles et dans la seconde partie, un aperçu détaillé sur l'intelligence artificielle, en particulier les réseaux de neurones artificiels pour lesquels nous donnerons tout d'abord quelques notions rudimentaires sur les neurones biologiques en précisant au passage les caractéristiques souvent conservées dans la conception des RNA. Nous expliciterons quelques exemples de réseaux, leurs architectures et leurs règles d'apprentissage.

Dans le second chapitre on présentera les résultats que nous avons obtenus dans la recherche de descripteurs susceptibles d'être corrélés à l'activité. Cette application porte sur une étude théorique de la contraction et de l'extension du cycle à sept sommets des TIBO (des anti-VIH) et leurs influences sur l'atome asymétrique. La seconde application sera de trouver, en appliquant les réseaux de neurones artificiels, des relations quantitatives Structure-Activité sur une série de molécules de la famille des flavonoïdes qui sont également des anti-VIH. Nous comparons les résultats statistiques à ceux donnés par d'autres méthodes de calculs, telles que les RML (Régression Multi Linéaire).

Les informations et les performances de ces méthodes recueillies lors de ces applications seront ensuite exploitées dans le but d'aider au développement de la recherche de la modélisation théorique de la séparation des énantiomères par la β -Cyclodextrine.

Nous terminerons ce travail par une conclusion générale et des perspectives qui sont la construction d'une base de donnée étoffée sur la β -cyclodextrine pour que l'utilisation des réseaux de neurones artificiels et des algorithmes génétiques soient plus crédible.

Chapitre I

Généralités

Partie 1: Relation Quantitative Structure Activité.

Introduction.

Les premiers essais de modélisation d'activités de molécules datent de la fin du 19^{ème} siècle, lorsque Crum-Brown et Frazer [1] postulèrent que l'activité biologique d'une molécule est une fonction de sa constitution chimique. Mais ce n'est qu'en 1964 que furent développés les modèles de "contribution de groupes", qui constitue les réels débuts de la modélisation RQSA. Depuis, l'essor de nouvelles techniques de modélisation par apprentissage, linéaires d'abord, puis non linéaires, ont permis la mise en place de nombreuses méthodes ; elles reposent pour la plupart sur la recherche d'une relation entre un ensemble de nombres réels, descripteurs de la molécule, et la propriété ou l'activité que l'on souhaite prédire.

La relation structure activité est depuis quelques années utilisée de façon systématique dans la conception de nouveaux médicaments et permet de prédire l'activité thérapeutique de molécules avant même de les avoir synthétisées. Cette prédiction, bien que non totale, a l'avantage d'économiser une partie de l'expérimentation initiale. En effet, cette méthode permet de concevoir de classer de trier des molécules, ainsi que de calculer ou de prévoir leurs propriétés à partir de modèles mathématiques. Le terme activité est un terme général qui peut désigner une activité biologique, une propriété physique, une réactivité chimique ou même une rétention chromatographique.

La recherche de relations entre les structures chimiques et leur activité passe par plusieurs étapes (Figure 1). Chaque étape doit être traitée avec beaucoup d'attention afin d'obtenir des RQSA fiables.

1.1 Choix de l'échantillon à étudier.

Pour l'obtention d'un échantillon représentatif d'une population bien déterminée, on s'appuie sur un certain nombre de règles. Celles-ci permettent d'obtenir l'image la plus exacte possible d'un grand ensemble en examinant qu'une petite partie de celui-ci [2]. Les règles à suivre dans l'étude des RQSA sont les suivantes :

a- Le nombre de molécules à étudier doit être suffisamment élevé pour permettre une analyse raisonnable de données. Plus le nombre est réduit, plus il risque de définir une population limitée.

b- Les valeurs des paramètres décrivant les structures moléculaires doivent être connues avec précision. Il est recommandé de prendre ces valeurs d'un même laboratoire, ou de laboratoires différents mais utilisant des procédés expérimentaux identiques.

c- Lors d'une classification de composés selon leur activité biologique, les molécules à l'intérieur d'une même classe doivent non seulement être structurellement proches, mais de surcroît avoir le même mécanisme d'action.

d- Les ensembles d'apprentissage et de test doivent contenir des composés de structures chimiques aussi voisines que possibles. En effet, la capacité de prédiction d'une RQSA dépend en grande partie de la composition de l'ensemble des données utilisées pour tester le modèle établi.

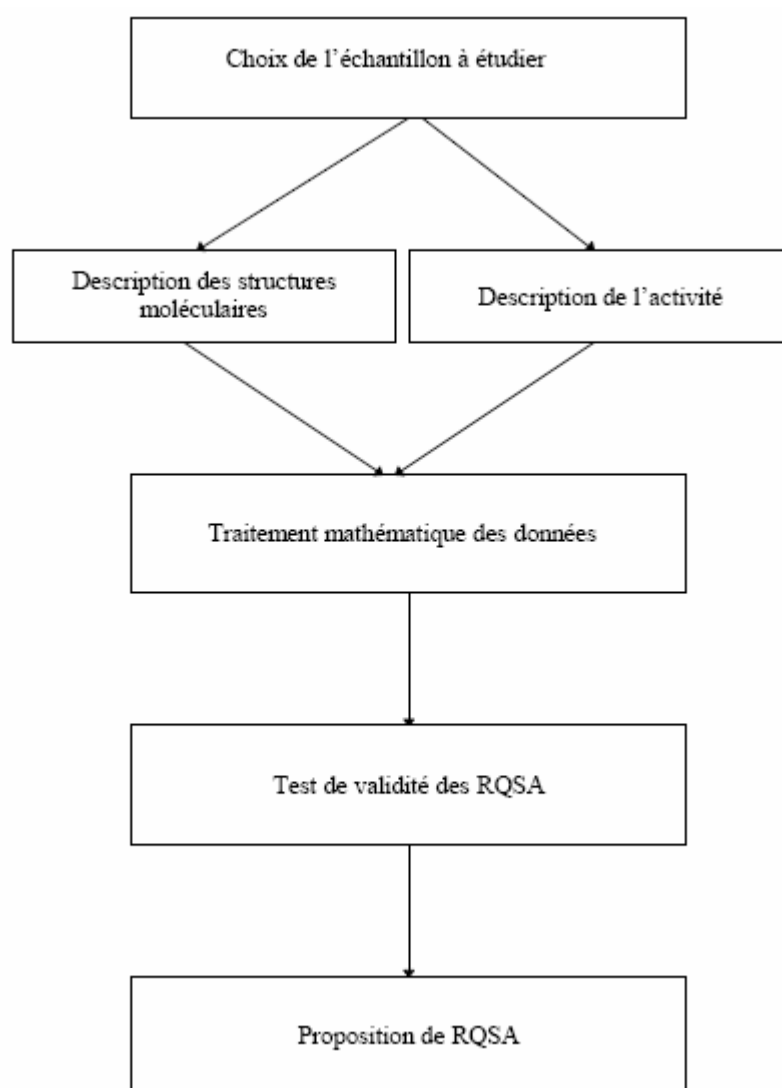


Figure 1 : Les étapes de recherche de RQSA [2].

1.2 Description de l'activité.

1.2.1 Activité Biologique

Il est possible, pour beaucoup de types de molécules, d'effectuer une étude pour chercher des liens entre la structure de la molécule et son activité biologique. Dans les études RQSA on prend en compte le moment du contact entre la molécule étudiée et son récepteur biologique correspondant. Autrement dit, dans l'interaction molécule-récepteur, tout ce qui concerne le récepteur biologique et le mécanisme d'action est une constante. Quand on veut préparer une nouvelle substance, il sera possible de connaître a priori son action biologique et les risques qu'elle présente. Avant de créer la substance, l'ordinateur peut proposer la structure des molécules qui accompliront la meilleure action biologique. Cela permet dès le début de limiter au maximum les substances à tester [2].

L'activité biologique peut être décrite de trois manières :

- Qualitative : l'activité biologique est codée (1 ou 0) de façon à distinguer les molécules biologiquement actives de celles qui sont inactives.
- Semi-Quantitative : dans ce cas, l'activité biologique est décrite selon plusieurs classes ou groupes.
- Quantitative : dans ce dernier cas, l'activité biologique peut prendre n'importe quelle valeur numérique. Par conséquent, les composés étudiés n'appartiennent pas à des classes prédéfinies.

1.2.2 Réactivité chimique

La synthèse organique est l'un des aspects fondamentaux de l'activité des chimistes organiciens. Cette synthèse fait appel à de nombreuses connaissances chimiques. Des connaissances générales relatives à la théorie chimique, à la structure et à la réactivité des molécules et des connaissances spécifiques de la synthèse sont en effet indispensables. Il est donc clair que la synthèse organique est un domaine très complexe à maîtriser et par conséquent des programmes informatiques ont été développés pour aider les chimistes à élaborer des plans de synthèses [2].

1.2.3 Propriété Physique

Les propriétés sont des données importantes utilisées pour caractériser des composés organiques. Elles sont généralement mesurées avec une précision et une reproductibilité satisfaisantes. Cependant, il arrive que les valeurs des propriétés concernant certains

produits ne soient pas disponibles ou qu'elles soient difficiles, voire impossibles à déterminer expérimentalement. Dans de tels cas, des méthodes théoriques ou empiriques sont utilisées. Les valeurs calculées ne correspondent pas aux valeurs réelles et par conséquent les RQSA établies à partir de ces valeurs peuvent être biaisées [2].

1.2.4 Rétention chromatographique

Les relations structure – rétention chromatographique sont étudiées pour trois raisons principales : explication des mécanismes des séparations chromatographiques, prédiction des grandeurs de rétention, et caractérisation des propriétés physiques des solutés.

1.3 Description des structures moléculaires.

De nombreuses recherches ont été menées, au cours des dernières décennies, pour trouver la meilleure façon de représenter l'information contenue dans la structure des molécules, et ces structures elles-mêmes, en un ensemble de nombres réels appelés descripteurs ; une fois que ces nombres sont disponibles, il est possible d'établir une relation entre ceux-ci et une propriété ou activité moléculaire, à l'aide d'outils de modélisation classiques. Ces descripteurs numériques réalisent de ce fait un codage de l'information chimique en un vecteur de réels. On en dénombre aujourd'hui plus de 3000 types, qui quantifient des caractéristiques physico-chimiques ou structurales de molécules. Ils peuvent être obtenus de manière empirique ou non-empirique, mais les descripteurs calculés, et non mesurés, sont à privilégier : ils permettent en effet d'effectuer des prédictions sans avoir à synthétiser les molécules, ce qui est un des objectifs de la modélisation. Il existe cependant quelques descripteurs mesurés : il s'agit généralement de données expérimentales plus faciles à mesurer que la propriété ou l'activité à prédire (coefficient de partage eau-octanol [3], polarisabilité, ou potentiel d'ionisation).

Avant toute modélisation, il est nécessaire de calculer ou de mesurer un grand nombre de descripteurs différents, car les mécanismes qui déterminent l'activité d'une molécule ou une de ses propriétés sont fréquemment mal connus. Il faut ensuite sélectionner parmi ces variables celles qui sont les plus pertinentes pour la modélisation.

1.3.1 Les descripteurs moléculaires

Nous allons présenter les descripteurs moléculaires les plus courants, en commençant par les descripteurs les plus simples, qui nécessitent peu de connaissances sur la structure moléculaire, mais véhiculent peu d'informations. Nous verrons ensuite comment les progrès de la modélisation moléculaire ont permis d'accéder à la structure 3D de la molécule, et de

calculer des descripteurs à partir de cette structure. Il est important de noter qu'on peut représenter une molécule par une structure 2D ou par une structure 3D. Par la structure 2D, qui est plane, on ne peut déterminer que les propriétés topologiques. Par la structure 3D, on peut déterminer un très grand nombre de propriétés physico-chimiques et géométriques à l'aide des méthodes de la chimie théorique.

a - Les descripteurs 1D : sont accessibles à partir de la formule brute de la molécule (par exemple C₆H₆O pour le phénol), et décrivent des propriétés globales du composé. Il s'agit par exemple de sa composition, c'est-à-dire les atomes qui le constituent, ou de sa masse molaire. On peut remarquer que ces descripteurs ne permettent pas de distinguer les isomères de constitution.

b - Les descripteurs 2D : sont calculés à partir de la formule développée de la molécule. Ils peuvent être de plusieurs types.

- **Les indices constitutionnels :** caractérisent les différents composants de la molécule. Il s'agit par exemple du nombre de liaisons simples ou multiples, du nombre de cycles...

- **Les indices topologiques :** Ils peuvent être obtenus à partir de la structure 2D de la molécule, et donnent des informations sur sa taille, sa forme globale et ses ramifications. Les plus fréquemment utilisés sont l'indice de Wiener [4], l'indice de Randić [5], l'indice de connectivité de valence de Kier-Hall [6] et l'indice de Balaban [7]. L'indice de Wiener permet de caractériser le volume moléculaire et la ramification d'une molécule : si l'on appelle distance topologique entre deux atomes le plus petit nombre de liaisons séparant ces deux atomes, l'indice de Wiener est égal à la somme de toutes les distances topologiques entre les différentes paires d'atomes de la molécule. L'indice de Randić est un des descripteurs les plus utilisés ; il peut être interprété comme une mesure de l'aire de la molécule accessible au solvant.

Ces descripteurs 2D reflètent bien les propriétés physiques dans la plupart des cas, mais sont insuffisants pour expliquer de façon satisfaisante certaines propriétés ou activités, telles que les activités biologiques. Des descripteurs, accessibles à partir de la structure 3D des molécules, ont pu être calculés grâce au développement des techniques instrumentales et de nouvelles méthodes théoriques.

c - Les descripteurs 3D : d'une molécule sont évalués à partir des positions relatives de ses atomes dans l'espace, et décrivent des caractéristiques plus complexes; leurs calculs nécessitent donc de connaître, le plus souvent par modélisation moléculaire empirique ou

ab-initio, la géométrie 3D de la molécule. Ces descripteurs s'avèrent donc relativement coûteux en temps de calcul, mais apportent davantage d'informations, et sont nécessaires à la modélisation de propriétés ou d'activités qui dépendent de la structure 3D. On distingue plusieurs familles importantes de descripteurs 3D :

- **Les descripteurs géométriques** les plus importants sont le volume moléculaire, la surface accessible au solvant, le moment principal d'inertie.
- **Les descripteurs électroniques** permettent de quantifier différents types d'interactions inter- et intramoléculaires, de grande influence sur l'activité biologique de molécules. Le calcul de la plupart de ces descripteurs nécessite la recherche de la géométrie pour laquelle l'énergie stérique est minimale, et fait souvent appel à la chimie quantique. Par exemple, les énergies de la plus haute orbitale moléculaire occupée et de la plus basse vacante sont des descripteurs fréquemment sélectionnés.

Le moment dipolaire, le potentiel d'ionisation, et différentes énergies relatives à la molécule sont d'autres paramètres importants.

- **Les descripteurs spectroscopiques** : les molécules peuvent être caractérisées par des mesures spectroscopiques, par exemples par leurs fonctions d'onde vibrationnelles. En effet, les vibrations d'une molécule dépendent de la masse des atomes et des forces d'interaction entre ceux-ci ; ces vibrations fournissent donc des informations sur la structure de la molécule et sur sa conformation. Les spectres infrarouges peuvent être obtenus soit de manière expérimentale, soit par calcul théorique, après recherche de la géométrie optimale de la molécule. Ces spectres sont alors codés en vecteurs de descripteurs de taille fixe. Le descripteur EVA [7] est ainsi obtenu à partir des fréquences de vibration de chaque molécule. Les descripteurs de type *MoRSE* [8] (*Molecule Representation of Structures based on Electron diffraction*) sont calculés à partir d'une simulation du spectre infrarouge ; ils font appel au calcul des intensités théoriques de diffraction d'électrons.

Réduction du nombre de variables

Comme nous l'avons rappelé, un grand nombre de descripteurs différents sont collectés pour la modélisation d'une grandeur donnée, car les facteurs déterminants du processus étudié ne sont a priori pas connus. Cependant, les descripteurs envisagés n'ont pas tous une influence significative sur la grandeur modélisée, et les variables ne sont pas toujours mutuellement indépendantes. De plus, le nombre de descripteurs, c'est-à-dire la dimension du vecteur d'entrée, détermine la dimension du vecteur des paramètres à ajuster.

Si cette dimension est trop importante par rapport au nombre d'exemples de la base d'apprentissage, le modèle risque d'être sur ajusté à ces exemples, et incapable de prédire la grandeur modélisée sur de nouvelles observations.

Il est donc nécessaire de réduire la dimensionnalité des variables d'entrée. Plusieurs approches sont possibles pour résoudre ce problème :

- réduire la dimension de l'espace des entrées.
- remplacer les variables corrélées par de nouvelles variables synthétiques, obtenues à partir de leurs combinaisons.
- sélectionner les variables les plus pertinentes.

1.4 Traitement mathématique des données.

Le traitement mathématique et statistique permet la sélection des variables et la détermination d'une équation reliant de façon optimale la structure et l'activité. En effet, le développement considérable des méthodes computationnelles rend à la disposition du praticien un grand nombre de techniques différentes qui permettent théoriquement, d'atteindre un objectif donné. Mais chaque technique a ses avantages et ses inconvénients, et le choix de la technique la plus appropriée est une des conditions essentielles du succès de la modélisation. Ces méthodes diffèrent par la stratégie de corréler la variable dépendante Y (ou l'activité) à un ensemble de variables indépendantes X_i (descripteurs moléculaires). Deux grands types de méthodes sont possibles : les méthodes linéaires et non-linéaires.

1.4.1 Méthode linéaire :

1.4.1.1 Régression multilinéaire (RML). [9]

La régression multilinéaire (RML) est la méthode la plus utilisée dans les études RQSA. Cette méthode suggère que l'augmentation de la valeur d'une variable indépendante entraîne une modification linéaire de la valeur de la variable dépendante Y. Dans cette approche, Y est exprimé sous forme d'une combinaison linéaire des X selon l'équation 1.

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n \dots\dots\dots(1)$$

a_i : coefficients de régression des variables X_i , ils sont déterminés par la méthode des moindres carrés.

1.4.1.2 Analyse en Composantes Principales (APC). [10]

L'APC est la plus simple et la plus connue des techniques d'Analyse de données multivariées. Etant donné un ensemble d'observations décrites par des variables exclusivement numériques $\{x_1, x_2, \dots, x_p\}$, l'ACP a pour objectif de décrire ce même ensemble de données par de nouvelles variables en nombre réduit. Ces nouvelles variables seront des combinaisons linéaires des variables originales, et porteront le nom de Composantes Principales (CP). En général, la réduction du nombre de variables utilisées pour décrire un ensemble de données provoque une perte d'information par détection de variables qui ont un comportement commun.

1.4.1.3 Analyse discriminante (AD). [11]

La technique d'analyse discriminante permet d'établir un modèle de discrimination entre les molécules considérées actives et celles considérées comme inactives. Ceci est particulièrement utile dans le cas de grandes bases de données issues de la chimie combinatoire, car l'on peut facilement effectuer un premier tri («actif» / «inactif») sur des milliers de molécules, sans tenir compte de leur mode d'action.

L'équation permettant la discrimination peut aussi être appliquée à des molécules qui n'ont pas été évaluées expérimentalement pour en prédire leur activité ou leur absence d'activité. Il ne permet cependant pas d'identifier les structures actives des molécules, mais représente un bon moyen pour identifier une « tête de série » ou une famille de molécules dont aucun élément ne permettait, a priori, de soupçonner une activité bien déterminée.

1.4.1.4 Méthode des moindres carrées (PLS)

La méthode des moindres carrés ou PLS (Partial Least squares error) a été présentée comme une alternative à la RML. Elle se révèle utile spécialement dans les études de RQSA où le nombre des descripteurs physico-chimiques est grand et où ces variables sont inter corrélées. Comme l'ACP, cette méthode essaie d'extraire un nombre réduit de variables orthogonales qui ne sont qu'une combinaison linéaire des variables initiales. Cependant, dans l'ACP, l'extraction des composantes principales se fait indépendamment de la variable à expliquer tandis que dans la PLS, les variables orthogonales doivent garder de bonnes corrélations avec la variable à expliquer.

1.4.2 Méthode non linéaire :

D'autres méthodes mathématiques d'analyse non linéaire ont été développées pour traiter les situations où les analyses linéaires se révèlent insuffisantes. En effet, les analyses

non linéaires ont l'avantage d'être des méthodes adaptatives, aptes à modifier le modèle construit en fonction de nouvelles données.

1.4.2.1 Réseaux de Neurones Artificiels (RNA). [12]

L'analyse par réseaux de neurones est une des méthodes les plus performantes. Ces réseaux sont des systèmes basés sur l'architecture simplifiée du cerveau humain, et essaient de reproduire son fonctionnement. Un réseau est constitué de «neurones », qui sont des unités de calcul disposées sur plusieurs niveaux reliés entre eux : un niveau «d'entrée» qui intègre les variables initiales, un ou plusieurs niveaux «intermédiaires» qui traitent l'information. Ces neurones ont la capacité de traiter l'information qui leur est fournie (les valeurs des descripteurs choisis), de façon à trouver la meilleure combinaison de ces descripteurs pour donner une valeur de sortie souhaitée (la valeur d'activité). Les niveaux de traitement sont interactifs et conduisent à un niveau de « sortie » qui donne le résultat final du réseau. Cette architecture permet donc d'effectuer une analyse complexe et évolutive, du fait des interactions permanentes entre les niveaux de traitement des informations.

1.4.2.2 Algorithme Génétique (AG). [13]

Un autre exemple d'analyse mathématique non linéaire est représenté par les algorithmes génétiques, fondés sur la théorie de l'évolution darwinienne, selon laquelle chaque génération d'une espèce est améliorée par combinaison des meilleurs de ses individus. Les algorithmes génétiques travaillent donc sur des «individus» (structures moléculaires), qui sont des solutions potentielles d'un problème à résoudre (une activité pharmacologique). Une évaluation préliminaire de chaque individu permet de juger de la pertinence des solutions possibles, et à l'élimination des moins performants. Les individus conservés sont combinés virtuellement entre eux et voire même «mutés», de façon à obtenir une nouvelle population (ensemble de molécules) mieux adaptée (plus efficace) que la précédente, et ainsi de suite. La population initiale donne lieu donc à des générations successives, et à la fin de ce processus l'on retrouve les «survivants», qui représentent la meilleure des solutions possibles au problème initialement posé.

1- Test de validité des RQSA.

La dernière étape des RQSA consiste à tester le pouvoir de prédiction des modèles établis. Ce test de validation nous permet d'accepter ou de rejeter le meilleur modèle proposé pour une étude donnée.

Deux types d'échantillons sont examinés pour tester la validité des RQSA :

- * échantillon d'apprentissage permettant la construction des règles de décision.
- * échantillon test permettant de juger la validité de ces règles.

- Concepts de base utilisés

Plusieurs paramètres statistiques permettent d'évaluer les modèles de corrélation. Il nous semble donc utile de définir quelques termes mathématiques que l'on ne manquera pas de rencontrer lors d'étude de RQSA. Dans le cas de la régression linéaire multiple, la qualité de cette corrélation peut se faire à l'aide des indicateurs suivants :

• Le coefficient de corrélation

Ce coefficient mesure la qualité de la corrélation. Plus R est proche de 1, meilleure est la corrélation. Cette condition est nécessaire mais elle n'est pas suffisante.

Le coefficient de corrélation multiple R est donné par la relation :

$$R = \sqrt{1 - \frac{\sum_i^n (Y_i^{cal} - Y_i^{obs})^2}{\sum_i^n \left(Y_i^{cal} - \frac{\sum_i^n Y_i^{obs}}{n} \right)^2}}$$

Y_i^{cal} : valeur de l'activité calculée ou prédite par le modèle.

Y_i^{obs} : valeur de l'activité mesurée ou expérimentale.

n : nombre d'observations qui correspond au nombre de molécules soumises à la recherche de régression.

En général, l'ajustement est validé dès que la valeur de ce coefficient est supérieure à $\frac{\sqrt{3}}{2}$

• Le test statistique

Un test sur R^2 permet d'apprécier la signification de l'ensemble de la régression, et cela en calculant le quotient :

$$F = \frac{R^2}{1 - R^2} \frac{n - K_2}{K_2 - K_1}$$

F : variable de *Fischer Snedecor*, elle représente la variance de la régression ou le taux de confiance. Plus F est grand, meilleure est l'équation de corrélation.

K_1 : nombre de paramètres dans l'équation de référence.

K_2 : nombre de paramètres ou de variables indépendantes dans l'équation de régression considérée.

n : le nombre de molécules soumises à la recherche de régressions.

• La déviation standard

L'écart type ou déviation standard d'une valeur correspond à l'incertitude basée sur la moyenne arithmétique des valeurs observées et calculées pour une même série de composés. Plus la valeur de s est petite, meilleure est la corrélation obtenue. Ce terme est donné par la relation suivante :

$$S = \sqrt{\frac{\sum_i^n (Y_i^{cal} - Y_i^{obs})^2}{n - K - 1}}$$

K : nombre de variables indépendantes (ou nombre de paramètres de la corrélation).

L'erreur standard permet d'apprécier si l'introduction d'un descripteur supplémentaire dans le modèle est bénéfique ou non.

• La validation croisée

L'application de la technique de RM requiert énormément de mémoire machine pour de gros systèmes, et nécessite de longs temps de calcul. En 1988, Cramer a essayé d'adapter cette méthode par l'introduction de la validation croisée [14]. Cette technique consiste à enlever, à chaque étape de la RM, des observations choisies aléatoirement. Le nouveau modèle ainsi obtenu donne une nouvelle équation de corrélation qui permet de prédire une nouvelle activité biologique. Par conséquent, la validation croisée ou cross-validation est une méthode qui donne une estimation du pouvoir prédictif du modèle, elle permet de contrôler la validité du modèle de la régression.

Le coefficient de cross-validation est donné par la relation :

$$R_{cv}^2 = 1 - \frac{\sum_i^n (Y_i^{obs} - Y_i^{cal})^2}{\sum_i^n (Y_i^{obs} - \bar{Y}^{obs})^2}$$

Y_i^{obs} : activité observée

Y_i^{cal} : activité calculée

$\bar{y}^{obs}$: moyenne de la somme des activités observées.

Plus cette valeur est proche de R^2 , meilleure est la prédiction du modèle obtenu. Un modèle ayant un R^2_{cv} nul ou très proche de zéro n'a pas de pouvoir prédictif.

Conclusion.

Les modèles des RQSA sont établis à partir d'un ensemble restreint de composés que l'on appelle ensemble d'apprentissage. Ces composés sont décrits à l'aide de descripteurs. Ces derniers sont sélectionnés par un modèle statistique approprié et constituent la base de RQSA. Les relations ainsi développées, grâce à un traitement mathématique, pourraient permettre l'estimation de données biologiques de composés non testés et le développement de nouveaux produits.

REFERENCE.

- [1]- Crum-Brown, A., Frazer, T. Transactions of the Royal Society of Edinburgh 1868-69, 25, p.151-203.
- [2]- D. Charqaoui. *Etude de Relations Structure-Activités Biologiques, Propriétés Physiques et Réactivité de composés organiques au moyen des réseaux de neurone* .
Thèse de Doctorat d'Etat. Université de Caen. France. 1994.
- [3]- C. Hansch, R. Muir, T. Fujita, P. Maloney, E. Geiger, M. Streich, Journal of the American Chemical Society, 1963, 85, 2817.
- [4]- Wiener, H. Structural determination of paraffin boiling points. Journal of Chemical Information and Computer Sciences, 1947, 69, p. 17-20.
- [5]- Randić, M. Journal of the American Chemical Society, 1975, 97, p. 6609-6614.
- [6]- Kier, L.B., et Hall, L.H. Molecular connectivity in chemistry and drug research. New-York : Academic Press, 1976.
- [7]- Balaban, A.T. Highly discriminating distance-based topological index. Chemical Physics Letters, 1982, 89, p. 399-404.
- [8]- Schuur, J.H., Selzer, P., et Gasteiger, J. Journal of Chemical Information and Computer Sciences, 1996, 36 (2), p. 334-344.
- [9]-R. Tomassone, S. Audrain, E. Lesquoy-de Turckheim, C. Millier, dans « *La régression. Nouveaux regards sur une ancienne méthode statistique* », édition « Masson ». Paris, 1992.
- [10]- L. Lebart, A. Morineau, J. P. Fenelon, dans « *Traitement des données statistiques* », éd. Dunod, Paris, 1979, p. 185.
- [11]- A. Mkhadri, Pattern Recognition Letter. 1995, 16, 267.
- [12]-Manallack, Livongstone DJ. Eur J Med Chem 1999 ;34 : 195-208.
- [13]- J. Devillers. *Genetic Algorithms in Computer-Aided Molecular Design. In «DEVILLERS J - Genetic Algorithms in Molecular Modeling*. Academic Press Limiteded, London, 1996, pp 1-34.
- [14]- R. D. Cramer,QSAR. 7, 18-25;1988.

Partie 2: Les Réseaux de Neurones Artificiels

2.1 Historique.

Traditionnellement, on fait remonter les travaux sur les réseaux de neurones à ceux de Mc Culloch et Pitts (1943) [1], qui ont montré qu'avec de tels réseaux, on pouvait, en principe, calculer n'importe quelle fonction arithmétique ou logique. Vers la fin des années 1940, Donald Hebb [2] a ensuite proposé une théorie fondamentale pour l'apprentissage.

La première application concrète des réseaux de neurones artificiels est survenue vers la fin des années 1950 avec l'invention du réseau dit «perceptron» par un dénommé Frank Rosenblatt. Rosenblatt [3] et ses collègues ont construit un réseau et démontré ses habilités à reconnaître des formes. Malheureusement, il a été démontré par la suite que ce perceptron simple ne pouvait résoudre qu'une classe limitée de problème. Environ au même moment, Bernard Widrow et Ted Hoff ont proposé un nouvel algorithme d'apprentissage pour entraîner un réseau adaptatif de neurones linéaires, dont la structure et les capacités sont similaires au perceptron.

Vers la fin des années 1960, un livre publié par Marvin Minsky et Seymour Papert [4] est venu jeter beaucoup d'ombre sur le domaine des réseaux de neurones. Ces deux auteurs ont démontré les limitations des réseaux développés par Rosenblatt et Widrow-Hoff. Beaucoup de gens ont été influencés par cette démonstration qu'ils ont généralement mal interprétée. Ils ont conclu à tort que le domaine des réseaux de neurones était un cul de sac et qu'il fallait cesser de s'y intéresser (et de financer la recherche dans ce domaine), d'autant plus qu'on ne disposait pas à l'époque d'ordinateurs suffisamment puissants pour effectuer des calculs complexes.

Heureusement, certains chercheurs ont persévéré en développant de nouvelles architectures et de nouveaux algorithmes plus puissants. En 1972, Teuvo Kohonen et James Anderson ont développé indépendamment et simultanément de nouveaux réseaux pouvant servir de mémoires associatives. Egalement, Stephen Grossberg a investigué ce qu'on appelle les réseaux auto-organisés.

Dans les années 1980, une pierre d'achoppement a été levée par l'invention de l'algorithme de rétro propagation des erreurs. Cet algorithme est la réponse aux critiques de Minsky et Papert formulées à la fin des années 1960. C'est ce nouveau développement, généralement attribué à David Rumelhart et James McClelland, mais aussi découvert plus ou moins en même temps par Paul Werbos et par Yann LeCun, qui a littéralement ressuscité le domaine

des réseaux de neurones. Depuis ce temps, c'est un domaine où bouillonnent constamment de nouvelles théories, de nouvelles structures et de nouveaux algorithmes.

2.2 Les Réseaux de Neurones Artificiels

Il y a des centaines d'exemples qui montrent à la fois combien on peut espérer de la modélisation du système nerveux mais aussi combien il sera difficile d'imaginer et de comprendre les divers aspects des problèmes de perception. Il paraît donc naturel d'essayer de comprendre comment les systèmes biologiques sont capables de telles performances, et si possible, de s'inspirer de leurs principes pour imaginer de nouveaux algorithmes ou de nouvelles machines plus efficaces que ceux dont nous disposons actuellement. Les techniques de réseaux de neurones relèvent d'une telle approche : comprendre les principes selon lesquels les systèmes biologiques traitent l'information et s'en inspirer pour élaborer de nouvelles techniques. C'est donc une double démarche, à la fois cognitive et synthétique où le monde biologique doit être considéré comme une source de référence et de connaissance.

Les modèles de réseaux de neurones artificiels sont, à l'origine, une imitation du fonctionnement du cerveau. Ces réseaux utilisent les connaissances que nous avons de notre système nerveux pour traiter une grande quantité d'informations au moyen de neurones artificiels reliés entre eux et activés de façon massivement parallèle. Le neurone constitue l'élément fonctionnel de base du système nerveux biologique ou artificiel. Il est donc intéressant de connaître son fonctionnement pour pouvoir comprendre le mécanisme des RNA.

2.3 Le neurone

2.3.1 Le neurone biologique

2.3.1.1 Anatomie du neurone :

Le neurone est une cellule nerveuse constituée d'un corps cellulaire d'où partent des prolongements plus ou moins ramifiés, les dendrites (voie d'entrée) et un axone (voie de sortie). La communication est assurée entre les neurones par des jonctions particulières, les synapses (Figure 1).

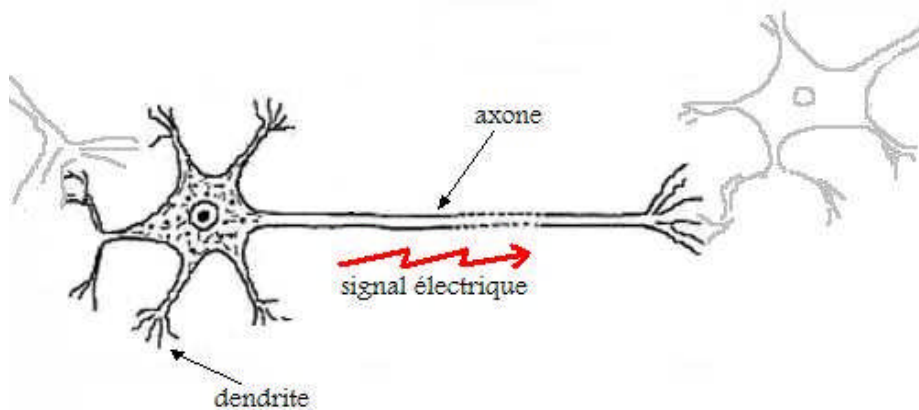


Figure 1 : Morphologie d'un neurone biologique.

2.3.1.2 Le fonctionnement des neurones :

Dans notre cerveau, les synapses sont innombrables; 100 milliards de neurones établissent chacun plusieurs milliers de contacts. En effet, le nombre de connections entre un neurone et ses voisins varie entre 10^3 et 10^4 , réalisant ainsi un véritable réseau de communication.

Les cellules nerveuses de notre organisme (les neurones) échangent de nombreuses informations entre elles au moyen de synapses. La synapse est la zone de jonction entre deux neurones : c'est le lieu de passage du message nerveux cheminant d'une cellule à l'autre.

Le neurone possède des sorties, les dendrites, par lesquelles il reçoit des impulsions de ses voisins. Si la somme des signaux dépasse un certain seuil, il renvoie un signal vers d'autres neurones, par l'intermédiaire de son "axone". Le message nerveux va donc se propager le long de l'axone sous la forme d'un courant électrique appelé influx nerveux. Au niveau de la synapse, le courant électrique est relayé par des substances chimiques, les neurotransmetteurs. Ces neurotransmetteurs traversent la fente synaptique et atteignent le second neurone, où elles recréent un influx nerveux électrique. Pour résumer, un neurone peut être schématisé ainsi (Figure 2): il fait la somme de toutes les informations qu'il reçoit et il émet un signal à condition que la somme soit suffisamment élevée.

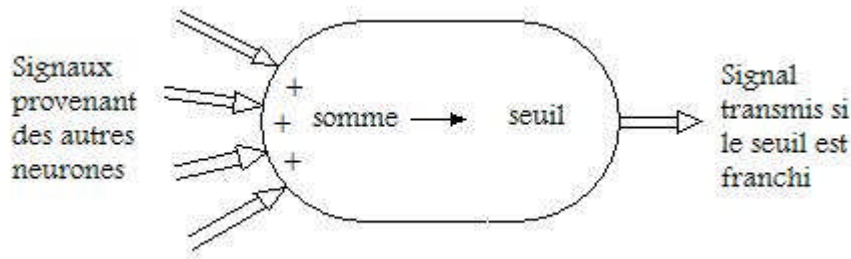


Figure 2 : Schématisation du passage de l'information dans un neurone.

Si l'on ramène la contribution d'un neurone au cerveau tout entier, on se rend compte que chacun effectue un travail très simple par rapport au résultat obtenu. En effet, les neurones réalisent des opérations basiques, et pourtant, lorsque l'on en met 10 milliards ensemble, on peut créer une entité pensante. Cependant, cela n'est pas suffisant: un cerveau ne peut rien faire s'il n'a pas de quoi apprendre. Il a besoin d'informations venant de l'extérieur. C'est pour cela qu'il est relié aux différents organes du corps. Par exemple, il reçoit les images provenant des yeux, les sons, les douleurs...etc. Grâce à ces informations il est capable de faire son apprentissage. Ce dernier est donc indispensable au développement du système nerveux.

Les neurones sont fortement connectés entre eux, mais ces connections ne sont pas aléatoires, elles correspondent à des réseaux dont les architectures sont assez bien connues. Les chercheurs ont essayé de modéliser les neurones biologiques et de mettre au point le neurone formel. C'est donc une modélisation mathématique qui reprend les grands principes du fonctionnement du neurone biologique.

2.3.2 Le neurone formel.

Un neurone formel est une représentation mathématique et informatique d'un neurone biologique, par ce fait, il constitue l'unité élémentaire des réseaux de neurones artificiels. Il possède généralement plusieurs entrées et une sortie qui correspondent respectivement aux dendrites et au cône d'émergence du neurone biologique (point de départ de l'axone). Les actions excitatrices et inhibitrices des synapses sont représentées, la plupart du temps, par des coefficients numériques (les poids synaptiques) associés aux entrées. Les valeurs numériques de ces coefficients sont ajustées dans une phase d'apprentissage. Dans sa version la plus simple, un neurone formel calcule la somme pondérée des entrées reçues,

puis applique à cette valeur une fonction d'activation ou fonction transfert. La valeur finale obtenue est la sortie du neurone.

La Figure 3 montre la correspondance entre le neurone biologique et le neurone formel.

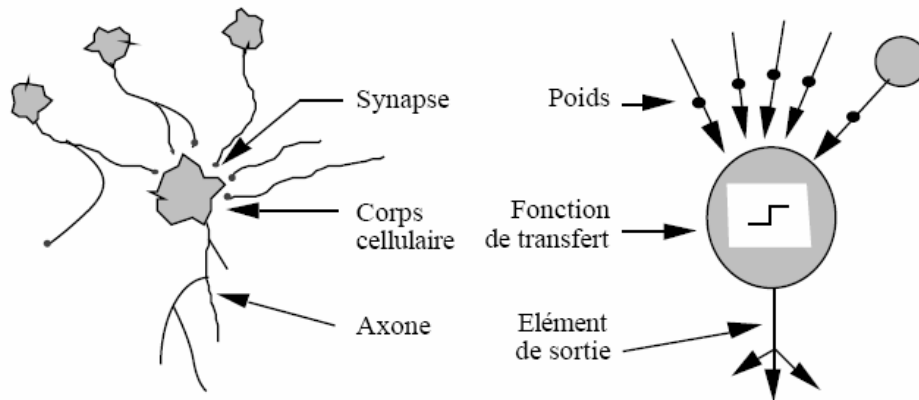


Figure 3 : Mise en correspondance neurone biologique / neurone artificiel.

Le premier modèle mathématique et informatique du neurone biologique est proposé par Warren McCulloch et Walter Pitts en 1943 [1]. En s'appuyant sur les propriétés des neurones biologiques, ces auteurs proposent un modèle simple de neurone formel. Il s'agit d'un neurone binaire, c'est-à-dire dont la sortie vaut 0 ou 1. Pour calculer cette sortie, le neurone effectue une somme pondérée de ses entrées (qui, en tant que sorties d'autres neurones formels, valent aussi 0 ou 1) puis applique une fonction d'activation à seuil : si la somme pondérée dépasse une certaine valeur, la sortie du neurone est 1 sinon elle vaut 0.

Nous pourrions schématiser cela dans la Figure 4.

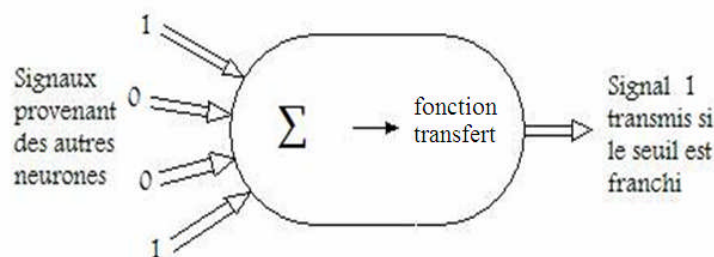


Figure 4 : modèle du neurone formel.

Malgré la simplicité de cette modélisation, le neurone formel dit McCulloch et Pitts reste aujourd'hui un élément de base des réseaux de neurones artificiels. De nombreuses variantes ont été proposées, plus ou moins biologiquement plausibles, mais s'appuyant

généralement sur les concepts inventés par les deux auteurs. On sait néanmoins aujourd'hui que ce modèle n'est qu'une approximation des fonctions remplies par le neurone réel.

2.3.2.1 La fonction d'activation ou transfert.

La fonction de transfert peut avoir plusieurs formes linéaires, binaires, probabilistes, à saturation ou continue. Les fonctions binaires à seuil telles que la fonction de Heaviside (ou signe) et les fonctions sigmoïdes, dont plusieurs formes existent, sont des exemples de ces fonctions. La Figure 5 regroupe quelques exemples de fonctions de transfert courantes.

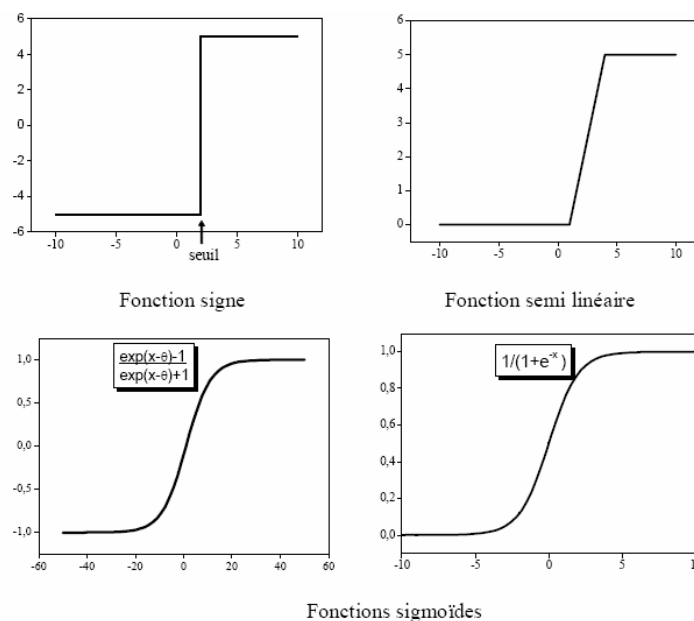


Figure 5 : Représentation de quelques exemples de fonctions de transfert.

2.3.3 Le Réseau de neurones

Les neurones formels que l'on vient de définir doivent être assemblés pour former un réseau.

On peut classer l'architecture des réseaux de neurones artificiels en deux grandes catégories:

2.3.3.1 Les Réseaux "FEED-FORWARD"

Appelés aussi réseaux de type Perceptron, ce sont des réseaux dans lesquels l'information se propage de couche en couche sans retour en arrière possible.

2.3.3.1.1 Les Perceptrons :

Le perceptron monocouche :

C'est historiquement le premier RNA, c'est le Perceptron de Rosenblatt. C'est un réseau simple, puisqu'il ne se compose que d'une couche d'entrée et d'une couche de sortie (Figure 6). Il est calqué, à la base, sur le système visuel et de ce fait a été conçu dans un but premier de reconnaissance des formes. Cependant, il peut aussi être utilisé pour faire de la classification et pour résoudre des opérations logiques simples (telle "ET" ou "OU"). Sa principale limite est qu'il ne peut résoudre que des problèmes linéairement séparables. Il suit généralement un apprentissage supervisé selon la règle de correction de l'erreur (ou selon la règle de Hebb [5]).

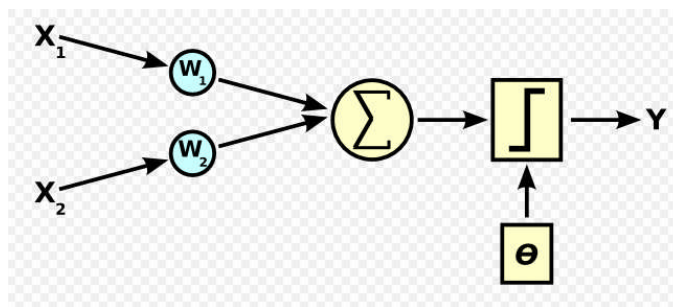


Figure 6 : Le perceptron monocouche.

Le perceptron multicouches [6] :

C'est une extension du précédent, avec une ou plusieurs couches cachées entre l'entrée et la sortie (Figure 7). Chaque neurone dans une couche est connecté à tous les neurones de la couche précédente et de la couche suivante (excepté pour les couches d'entrée et de sortie) et il n'y a pas de connexions entre les cellules d'une même couche. Les fonctions d'activation utilisées dans ce type de réseaux sont principalement les fonctions à seuil ou sigmoïdes. Il peut résoudre des problèmes non-linéairement séparables et des problèmes logiques plus compliqués [7].

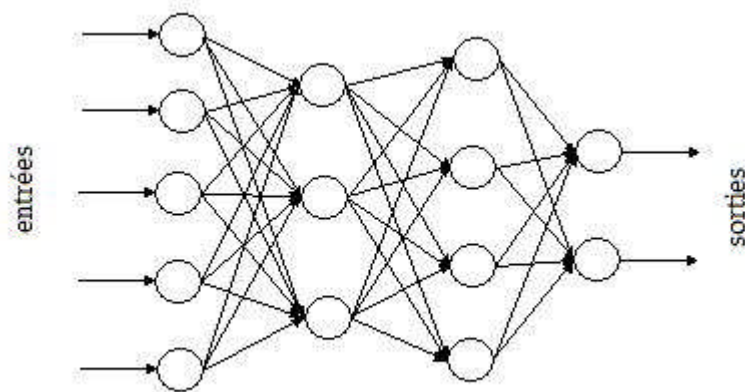


Figure 7 : Le Perceptron Multi-Couche (PMC).

Notons par ailleurs que les réseaux à couches sont les plus utilisés en chimie [8-9].

L'apprentissage de ces réseaux est supervisé c'est-à-dire, on présente au réseau des entrées, et au même temps les sorties que l'on désirerait pour cette entrée. Par exemple on lui présente en entrée une lettre " a " manuscrite et en sortie un code correspondant à la lettre " a ". Le réseau doit alors se reconfigurer, c'est-à-dire calculer ses poids afin que la sortie qu'il donne corresponde bien à la sortie désirée. L'algorithme utilisé pendant cet apprentissage est connu sous le nom de méthode de rétro propagation du gradient [10-12] ou encore « back-propagation ». Cette méthode d'apprentissage se divise en deux étapes:

- La propagation

La propagation consiste à présenter une configuration d'entrée au réseau, puis à propager cette entrée de proche en proche de la couche d'entrée à la couche de sortie en passant par les couches cachées. Autrement dit, à chaque exemple présenté au réseau, les entrées et les sorties des neurones cachés et des neurones de sortie sont calculées.

- La retro propagation

L'algorithme de rétro propagation consiste, après le processus de propagation, à minimiser l'erreur commise sur l'ensemble des exemples présentés, erreur considérée comme une fonction des poids synaptiques. Cette erreur représente la somme des différences au carré entre les réponses calculées et celles désirées pour tous les exemples contenus dans l'ensemble d'apprentissage.

2.3.3.2 Les réseaux à fonction radiale :

Ce sont les réseaux que l'on nomme aussi RBF ("Radial Basic Functions"). L'architecture est la même que pour les PMC cependant, les fonctions de base utilisées ici sont des fonctions Gaussiennes. Les RBF seront donc employés dans les mêmes types de problèmes que les PMC à savoir, en classification et en approximation de fonctions, particulièrement.

2.3.3.3 Les Réseaux "FEED-BACK"

Appelés aussi "réseaux récurrents", ce sont des réseaux dans lesquels il y a retour en arrière de l'information, autrement dit, ce type de réseau alimente leurs entrées avec leurs sorties contrairement au perceptron feed-forward.

Les réseaux de Kohonen:

Le réseau de Kohonen [13] repose sur des modèles biologiques comme celui de l'ouïe. Ces modèles présentent une organisation des cellules qui dépend souvent de leur spécialisation. Dans le cortex auditif, par exemple, on remarque un arrangement tonotopique des neurones, c'est-à-dire que des fréquences similaires sont détectées par des neurones voisins dans le cortex auditif. Le réseau de Kohonen est donc capable d'identifier des caractéristiques communes des formes d'entrée qui constituent l'ensemble d'apprentissage. Il utilise un apprentissage non supervisé pour modifier son état interne [14]. La topologie du réseau est formée de deux couches: une d'entrée classique et une de sortie où les neurones sont entièrement connectés dans un plan. Quand une donnée est présentée au réseau, l'un des neurones de la couche de sortie aura la valeur la plus importante. C'est le numéro de ce neurone, correspondant à une classe, qui constituera l'information extraite par le réseau de Kohonen. L'avantage de ce type de réseaux réside dans leur capacité d'auto-organisation. Il n'est pas nécessaire de faire d'hypothèses sur la nature ou la complexité des ensembles étudiés. Ce modèle de réseau a été appliqué à des problèmes de reconnaissance de parole [15].

Les réseaux de Hopfield [16] :

Celui-ci est typique des réseaux totalement connectés. Il propose un modèle à partir des neurones formels de McCulloch et Pitts qui permet de retrouver une forme mémorisée, en

stimulant le réseau par une partie caractéristique de cette forme et en le laissant se stabiliser. Cette propriété, qui permet de retrouver une donnée à partir de son contenu, fait du modèle de Hopfield une "mémoire associative". Le réseau de Hopfield est un RNA à une seule couche, formé de neurones totalement connectés ne pouvant envoyer et recevoir que des valeurs binaires. L'apprentissage du réseau de Hopfield est relativement différent de celui des modèles à couches. Les réseaux de Hopfield sont particulièrement bien adaptés à la résolution des problèmes caractérisés par un très grand nombre de solutions parmi lesquelles il faut trouver la meilleure, ou tout au moins l'une des meilleures [17-18]. Ce sont des problèmes dits d'optimisation combinatoire [19].

Les deux réseaux précédents utilisent un apprentissage non supervisé autrement dit on présente une entrée au réseau et on le laisse évoluer librement jusqu'à ce qu'il se stabilise.

2.3.4 La phase d'apprentissage

Parmi les propriétés désirables pour un réseau de neurones, la plus fondamentale est sûrement la capacité d'apprendre de son environnement, d'améliorer sa performance à travers un processus d'apprentissage. L'apprentissage des RNA consiste donc à trouver des relations entre les configurations d'entrée (exemple : molécules) et les réponses désirées (exemple activités biologiques).

On peut distinguer deux types d'apprentissage : supervisé et non supervisé.

- L'apprentissage supervisé consiste à présenter les exemples un par un ainsi que leurs réponses désirées, et à modifier les poids en fonction des réponses calculées par le réseau. Au cours de la phase d'apprentissage, l'erreur entre les réponses désirées et calculées est réduite progressivement. L'apprentissage est achevé une fois que l'erreur est inférieure à un seuil choisi. On dit alors que le RNA a appris.

- L'apprentissage non supervisé, apprentissage au cours duquel on présente des exemples en entrée sans leurs réponses désirées, on laisse le réseau s'auto-organiser uniquement au moyen de lois locales qui régissent l'évolution des poids synaptiques.

2.3.5 La phase de prédiction

Au cours de cette phase, on présente au réseau un ensemble d'exemples nouveaux (ensemble de test) mais proches des exemples appris et on mesure la qualité de ses réponses. On peut chercher à évaluer la capacité du réseau à reconnaître les exemples qu'on lui a soumis lors de l'apprentissage (mémorisation). Dans ce cas, l'ensemble de test sera extrait de l'ensemble d'apprentissage. Mais dans la plupart des applications, on s'intéresse à la capacité

du réseau à généraliser et à son aptitude à reconnaître des exemples nouveaux (test). Dans ce cas, l'ensemble de test doit être constitué d'exemples qui ne font pas partie de l'ensemble d'apprentissage. Il y a évidemment une dégradation des performances quand on passe de la mémorisation à la prévision.

2.3.6 Applications.

Les RNA ont fait objet de plusieurs applications dans plusieurs domaines tels que :

- Robotique: certains robots sont dotés de réseaux de neurones.
- La reconnaissance: un réseau peut servir à reconnaître des caractères. Cela est déjà utilisé à la Poste pour lire les codes postaux, ou même dans certaines banques pour lire les chèques. Il est aussi possible de retrouver la prononciation des mots à partir d'un texte.
- La prévision: on utilise de plus en plus les réseaux pour faire des prévisions en marketing (prédiction de comportement, de possibilité de vente d'un produit, ...) ou pour le trafic routier.
- Dans l'environnement : l'étude de la concentration de l'ozone dans l'atmosphère [20] et l'étude de la dégradation des composés organiques dans le troposphère [21].
- Les RNA connaissent un large succès dans le domaine de la chimie. En effet, grâce à leurs capacités d'apprentissage et de généralisation, les RNA apportent une solution satisfaisante en spectroscopie: ces applications incluent le dépouillement des spectres RMN du proton [22] la prédiction de la structure tridimensionnelle des molécules à partir des spectres IR [23, 24] la prédiction [25] des déplacements chimiques du ^{13}C et l'analyse des spectres de masse [26].
- Les capacités de généralisation et du traitement de fonctions non linéaires ont fait des RNA un outil performant pour la modélisation des RQSA. Les exemples sont de plus en plus nombreux et en évolution perpétuelle. Les RNA ont été utilisés avec succès pour la prédiction de la biodégradation des dérivés du benzène [27] pour la prédiction de l'effet corrosif des acides sur la peau à partir de la cytotoxicité de ces acides [28]. D'autres études ont été faites sur la relation structure-odeur du musc [29] sur le comportement mutagénique de composés organique et de ses produits de clivage [30] sur la capacité d'adsorption de composés aliphatiques et aromatiques sur le charbon actif [31]. Mghazli et coll. ont étudié l'activité antifongique des dérivés du 1-[2-(phényl) allyl] imidazoles [32] ainsi que des dérivés du 1-[1-(phényl)vinyl]imidazoles [33]. Les auteurs ont montré que les résultats obtenus sont meilleurs que ceux donnés dans la littérature par d'autres méthodes. Les

mêmes conclusions ont été établies lors de l'étude du pouvoir cancérogène des hydrocarbures aromatiques [34] et de l'activité inhibitrice des dérivés de benzyl-pyridines [35]. Douali et coll. Ont appliqué les RNA dans la modélisation des RQSA des anti-VIH agissant comme INNTIs [36], les résultats trouvés par ces auteurs sont nettement meilleurs que ceux trouvés par les régressions multi linéaires, trouvés dans la littérature.

- Un grand nombre d'applications relatives aux RNA concernent la prédiction des structures secondaires et tertiaires des protéines [37-39].

- Une autre application des RNA concerne l'étude de la réduction, par la levure du boulanger, des fonctions carbonyles [40, 41]. Aires-de-Soussa et Gasteiger [42] ont utilisé un RNA en vue d'étudier la chiralité des composés et son effet sur leurs propriétés. Une application qui peut servir aussi bien dans l'étude des réactions stéréosélectives que dans la chimie analytique.

- L'attention s'est aussi portée sur l'utilisation des RNA pour la prédiction des propriétés physiques. Ainsi la solubilité d'une large gamme de composés organiques [43] a été estimée par les RNA. Les points d'ébullition, les points de fusion, les volumes molaires et les températures critiques des alcanes ont été étudiés en utilisant des RNA multicouches [44]. Les résultats encourageants ont conduit les auteurs à étendre cette étude à la prédiction des points d'ébullition d'autres composés: éthers, peroxydes, acétals et leurs analogues sulfurés [45].

REFERENCE

- [1]- En particulier dans McCulloch, W. S., Pitts, W. Bulletin of Mathematical Biophysics, vol. 5, pp. 115-133, 1943.
- [2]- Donald Hebb, the organization of Behavior : A Neuropsychological Theory, Wiley, New York, 1949.
- [3]- F. Rosenblatt. Neurocomputing. Foundations of Research, MIT Press.
- [4]- Minsky, M.; Papert, S. Perceptrons: An Introduction to Computational Geometry; éd. MIT Press, Cambridge, USA, 1969.
- [5]-Hebb, D. O. (1949). The organization of behavior. New York: Wiley.
- [6]- Manallack, D. T.; Livingston, D. J. Eur. J. Med. Chem. 1999, 34, 195-208.
- [7]- Schneider, G. Neural Netw. 2000, 13, 15-16.
- [8]- Zupan, J.; Novic, M.; Gasteiger, J. Chemom. Intell. Lab. Syst. 1995, 27, 175-187.
- [9]- Vracko, M.; Mills, D.; Basak, S. C. Envir. Toxicol. Pharm 2004, 16, 25-36.
- [10]- Benigni, R.; Richard, A. M. Mut. Res./Genetic Toxicol 1996, 371, 29-46.
- [11]- Valkova, I.; Vracko, M.; Basak, S. C. Analytica Chim. Acta 2004, 509, 179-186.
- [12]- Zupan, J.; Novic, M. Analytica Chimica Acta 1999, 388, 243-250.
- [13]- Kohonen, T. Self-Organization and Associative Memory; Springer Verlag, Berlin, heidelberg, New York, 1984.
- [14]- Zupan, J.; Gasteiger, J. Neural Networks for Chemists; VCH-Wiley, Weinheim, 1993, P. 82.
- [15]- Masson, E.; Wang, Y.-J. European Journal of Operational Research 1990, 47, 1-28.
- [16]- Hopfield, J. Proc. Nat. Acad. Sc. USA 1982, 79, 2554-2558
- [17]- Campadelli, P.; Medici, D.; Schettini, R. Image and Vision Computing 1997, 15, 161-166.
- [18]- Agapkin, O. A.; Orlov, Y. V.; Persiantsev, I. G.; Dolenko, S. A. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 2003, 502, 520-522.
- [19]- Tatem, A. J.; Lewis, H. G.; Atkinson, P. M.; Nixon, M. S. Intern. J. App. Earth Obs. Geoinf. 2001, 3, 184-190.
- [20]- Hadjiiski, L.; Hopke, P. j. wast. Manag. Assoc. 2000, 50, 894-901.
- [21]- Gramatica, P.; Consonni, V.; Todeschini, R. Chemosphere 1999, 38, 1371-1378.
- [22]- Thomsen, J. U.; Meyer, B. J. Magn. Reson. 1989, 84, 212-217.
- [23]- Severcan, M.; Severcan, F.; Haris, P. I. Journal of Molecular Structure 2001, 565, 383-387.
- [24]- Hemmer, M.; J, G. Analytica Chimica Acta 2000, 420, 145-154.
- [25]- Kvasnicka, V. J. Math. Chem. 1991, 6, 63-76.
- [26]- Curry, B.; Rumelhart, D. E. Tetrahedron comput. Methodol. 1990, 3, 213-238.
- [27]- Zitko, V. Chemosphere 1991, 23, 305-315.
- [28]- Barratt, M. D.; Dixit, M. B.; Jones, P. A. Toxicology in Vitro 1996, 10, 283-290.
- [29]- Chastrette, M.; De saint Laumer, J. Y. Eur. J. Med. Chem. 1991, 26, 829-833.
- [30]- Sztandera, L.; Garg, A.; Hayik, S.; Bhat, K. L.; Bock, C. W. Dyes and Pigments 2003, 59, 117-133.
- [31]- Brasquet, C.; Le Cloirec, P. Water Research 1999, 33, 3603-3608.
- [32]- Mghazli, S.; Jaouad, A.; M., M.; Villemin, D.; Cherqaoui, D. Chemosphere 2001, 43, 385-390.
- [33]- Mghazli, S.; Villemin, D.; Mansour, M.; Jaouad, A.; Cherqaoui, D. Fresenius Envir.Bull. 2000, 9, 734-745.
- [34]- Villemin, D.; Cherqaoui, D.; Mesbah, A. J. Chem. Inf. Comput. Sci. 1994, 34, 1288-1293.
- [35]- So, S.; Richards, W. G. J. Med. Chem. 1992, 35, 3201-3207.
- [36]- Douali, L.; Villemin, D.; Cherqaoui, D. Curr. Pharm. Des. 2003, 9, 1817-1826.
- Douali, L.; Villemin, D.; Cherqaoui, D. Curr. Pharm. Des. 2003, 9, 1817-1826.

- Douali, L.; Villemin, D.; Cherqaoui, D. *Curr. Pharm. Des.* 2003, 9, 1817-1826.
- [37]- Severcan, M.; Severcan, F.; Haris, P. I. *Journal of Molecular Structure* 2001, 565, 383-387.
- [38]- Hirst, J. D.; Sternberg, M. J. *Biochemistry.* 1992, 31, 7211-7218.
- [39]- Stolorz, P.; Lapedes, A.; Xia, Y. J. *Mol. Biol.* 1992, 225, 363-377.
- [40]- Cherqaoui, D. *Etude De Relations Structure-Activités Biologiques, Propriétés Physiques Et Réactivité De Composés Organiques Au Moyen Des Réseaux De Neurones: Thèse de l'université, Université de Caen*, 1994.
- [41]- Zakarya, D.; Farhaoui, L.; Fkih-Tetouan, S. *Tetrahedron lett.* 1994, 35, 4985-4988.
- [42]- Aires-de-Sousa, J.; Gasteiger, J. J. *Mol. Graph. Mod.* 2002, 20, 373-388.
- Bodor, N.; Harget, A.; Huang, M. J. J. *Am. Chem. Soc.* 1991, 113, 9480-9483.
- [43]- Cherqaoui, D.; Villemin, D.; Kvasnicka, V. *Chemom. Intell. Labor. Syst.* 1994, 24, 117-128.
- [44]- Cherqaoui, D.; Villemin, D. J. *Chem. Soc. Faraday Trans.* 1994, 90, 97-102.
- [45]- Cherqaoui, D.; Villemin, D.; Mesbah, A.; Cense, J. M.; Kvasnicka, V. J. *Chem. Soc. Faraday Trans.* 1994, 90, 2015-2019.

Partie 3: Le SIDA

3.1 *Historique de la maladie.*

Le Virus de l'Immunodéficience Humaine (VIH) est l'agent pathogène responsable du Syndrome de l'Immunodéficience Acquise (SIDA).

Presque trente ans après le début de cette épidémie, le mystère plane toujours sur son origine, il semble certain que le sida est né du contact entre l'espèce humaine et les chimpanzés, porteurs d'un virus parent du VIH, le VIS (Virus de l'Immunodéficience Simien). Mais où, quand et comment ce virus est arrivé à l'homme ? Ces questions ne trouvent pas toujours de réponse. Aujourd'hui, seules deux hypothèses subsistent. La première suggère que le virus du sida serait le produit accidentel d'un vaccin oral contre la polio qui a été fabriqué à partir de reins de chimpanzés et a été administré à plus d'un million d'africain au Congo à la fin des années 50. La seconde hypothèse se repose sur la théorie du transfert naturel, par blessure ou par consommation du chimpanzé.

C'est en effet à la fin de l'année 1980 et au début de 1981 que sont rapportés les premiers cas suspects de déficiences immunitaires graves parmi les populations américaines [1]. Ceci mène en fin 1981 à une première définition d'un syndrome de l'immunodéficience acquise [2], qui deviendra le SIDA. A cette époque, la forte tendance du SIDA à sévir parmi les populations homosexuelles [3] lui vaut dans un premier temps le nom de « Gay Related Immune Deficiency » (GRID) [4,5]. Les premiers cas de déficiences immunitaires chez les consommateurs de drogues injectables sont cependant identifiés dès décembre 1981 [6]. Un peu plus d'un an plus tard, le virus de l'immunodéficience humaine (VIH), responsable du SIDA, est identifié presque simultanément par les professeurs Montagnier [7] et Gallo [8]. Les médecins constatent alors qu'il se transmet par les relations sexuelles et par le sang et qu'en l'absence de traitement, l'infection par le VIH mène lentement mais irrémédiablement à la mort du patient. Le SIDA, qui constitue la phase terminale de ce processus, résulte de la destruction progressive des défenses immunitaires. Il conduit le patient à contracter de multiples infections dites « opportunistes », normalement inoffensives pour un sujet sain. Très vite, on découvre que le SIDA n'est pas limité aux USA et qu'une épidémie, qui touche également les populations hétérosexuelles, sévit déjà en Afrique [9]. En 1985, au moins un cas de SIDA a été rapporté sur chaque continent. En l'absence de traitement efficace, l'heure est dans un premier temps à la prévention et la sensibilisation.

Le premier médicament antirétroviral agissant sur le VIH, l'AZT, est découvert en 1985 [10] et commercialisé aux USA dès 1987. Le premier traitement destiné à réduire les risques de transmission mère enfant est développé en 1994. A partir de 1996, l'introduction de nouvelles classes de médicaments mène à l'abandon des traitements classiques au profit des multithérapies (Highly Active AntiRetroviral Therapy, ou HAART) [11], qui combinent plusieurs types de molécules actives contre le VIH. Ces avancées ont un impact majeur sur l'espérance de vie et la qualité de vie des patients, sans pour autant permettre leur guérison. Les progrès réalisés dans les domaines du suivi médical et de la prévention ne parviennent cependant pas à mettre un terme à l'expansion rapide de l'épidémie. Actuellement le dénombrement des cas du sida dans le monde permet d'obtenir une certaine visibilité de l'épidémie. En effet, plus de quarante millions de personnes sont porteuses du VIH et 95% d'entre elles n'ont pas accès aux traitements antirétroviraux, ce qui conduit à la contamination de 8500 personnes par jour dont 50% de ces contaminations touchent en premier lieu les femmes.

La transmission de ce virus se fait :

- par voie sanguine :
Transfusion de produits sanguins non chauffés ou non traités.
Dons d'organe.
Toxicomanie intraveineuse (partage du matériel d'injection).
- par voie sexuelle.
- De la mère à l'enfant par voie transplacentaire, lors de l'accouchement, ou par l'allaitement.

3.2 L'infection par le VIH

Les cellules et les virus contiennent tous les deux une information génétique, mais seule la cellule est en mesure d'utiliser cette information pour se renouveler et se reproduire. Un virus nécessite d'infecter une ou plusieurs cellules d'un organisme hôte pour que celui-ci utilise l'information génétique du virus, et produire celui-ci. Le Virus de l'Immunodéficience Humaine, ou VIH, appartient à la famille des rétrovirus, c'est-à-dire des virus à ARN ayant la capacité de se transformer en ADN pro viral en faisant intervenir une enzyme qu'ils possèdent, et que l'on appelle la transcriptase inverse. Par voie de fait, le virus peut se propager dans les cellules.

Il existe deux types du VIH, le VIH de type 1 (VIH-1) et VIH de type 2 (VIH-2) [12]. Ils partagent de nombreuses propriétés biologiques ainsi qu'une structure génomique semblable. Le VIH-1 est la forme la plus courante dans les pays industrialisés, tandis que le VIH-2 est principalement localisé en Afrique Subsaharienne. L'homologie de leur séquence est d'environ 50% alors que les différents isolants du VIH-1 diffèrent en moyenne de 10%. Les souches de VIH-2 sont semble moins pathogène et moins transmissible que le VIH-1. En effet, chez les individus infectés par le VIH-2 la latence clinique est plus longue et la mortalité plus faible [12]. Il existe toutefois une certaine variété dans le patrimoine génétique de ces deux espèces. On a ainsi identifié onze sous-types du VIH-1. A cela s'ajoutent de multiples variantes et souches mutantes.

3.3 Les cellules cibles

Nous rappelons que le système immunitaire est la protection de l'organisme, il fonctionne à l'aide de différents types de cellules :

Les macrophages, qui détectent et « mangent » les intrus (bactéries, virus, débris de cellules mortes) et en présentent des éléments aux autres cellules du système immunitaire.

Les lymphocytes B, qui produisent des anticorps de surfaces, protéines spécifiques de reconnaissance. Ce type de cellule est sélectionné de façon à ce que les anticorps reconnaissent les fragments de protéines étrangères et activent les autres éléments du système immunitaire (lymphocytes B qui se transforment en monocytes, les T4 et T8 qui fabriquent des messages chimiques...)

Les lymphocytes T8 (ou T CD8), nommés ainsi par leur récepteur de surface. Ces cellules sont aussi appelées « natural killer » et détruisent les cellules infectées par les virus et reconnues par des anticorps. La production des **T CD8** est dédiée aux protéines perçant les membranes des cellules auxquelles se sont accrochées les anticorps.

Les lymphocytes T4 (ou TCD4), qui sont les cibles du VIH. Ces cellules sont les chefs d'orchestre du système immunitaire : elles produisent des messagers chimiques qui activent les autres cellules et leur donne des ordres (production d'anticorps, multiplication cellulaire...)

Le VIH s'attaque de manière presque exclusive aux cellules humaines porteuses des récepteurs CD4, qui sont principalement des lymphocytes T et des macrophages. Le taux habituel de CD4 dans le sang est compris entre 800 et 1000 lymphocytes par millimètre cube de sang. Quand le taux de CD4 descend en dessous de 200 par millimètre cube, le

risque de survenue de sida est possible. Comme le montre la **Figure 1**. En effet, après infection par le virus, le taux de cellules CD4 du patient diminue rapidement, tandis que sa charge virale augmente de manière exponentielle. Cette première période constitue la phase aiguë de l'infection. Après quelques semaines, un équilibre s'établit suite à la réaction du système immunitaire. Il marque le début d'une longue phase asymptomatique, caractérisé par un état presque stationnaire de la charge virale et du système immunitaire. Ce n'est généralement qu'au bout de plusieurs années que cet équilibre est rompu et que le patient évolue vers un SIDA clinique, caractérisé par une virémie en forte augmentation et par un taux de cellules CD4 extrêmement bas, qui permet l'apparition de maladies opportunistes [13].

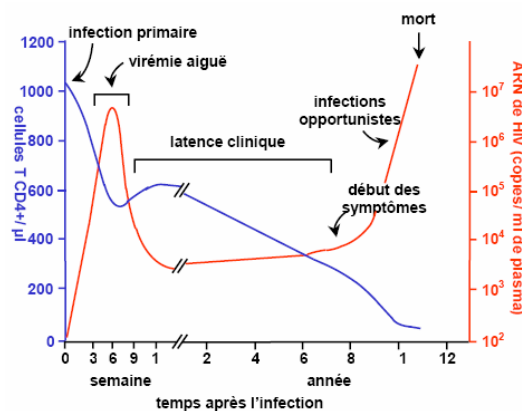


Figure 1 : Evolution, au cours du temps, de la charge virale d'un patient et de son taux de cellules CD4 [13].

3.4 Structure d'une particule virale [14]

Une particule virale du VIH-1 (**Figure 2**) mesure une centaine de nanomètres de diamètre et sa masse est d'environ $2.5 \cdot 10^8$ Dalton. Cette particule comporte :

Une enveloppe externe qui est constituée d'une double couche lipidique riche en cholestérol, associée d'une part à une glycoprotéine de surface (SU, ou gp120) trimérique, responsable de la reconnaissance des cellules ciblées par le VIH et d'autre part à une glycoprotéine transmembranaire (TM, ou gp41), elle aussi sous forme de trimère, qui joue un rôle important dans la pénétration du virus dans ces cellules. Une protéine de matrice (MA, ou p17) forme la partie externe du coeur viral. A l'intérieur de cette matrice se trouve la capside virale, construite par l'assemblage de protéines de capside (CA, ou p24). Elle contient les deux brins d'ARN qui constituent le génome viral, associés à une protéine de nucléocapside (NC, ou p7). Elle renferme également les trois enzymes virales

indispensables à la réplication que sont la transcriptase inverse (TI), l'intégrase (IN) et la protéase (PR), ainsi que plusieurs protéines virales annexes et de la cyclophiline A [15], appartenant à une précédente cellule hôte.

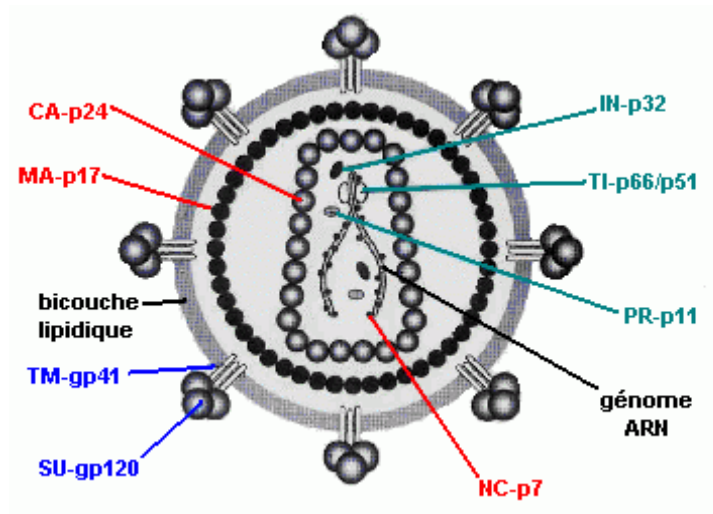


Figure 2 : Structure schématique d'une particule virale du VIH-1[14].

3.5 Le génome du VIH-1

3.5.1 Structure générale [14]

Le génome du virus de l'immunodéficience humaine de type 1 (Figure 3) est long d'environ 9800 bases. Il contient trois gènes principaux (gag, pol et env), communs à tous les rétrovirus, ainsi que plusieurs gènes annexes. Comme pour l'ensemble des rétrovirus (mais contrairement à d'autres catégories de virus), ce génome est diploïde. Etant constitué d'ARN, il doit être rétro transcrit par la transcriptase inverse pour que la cellule hôte puisse être effectivement infectée.

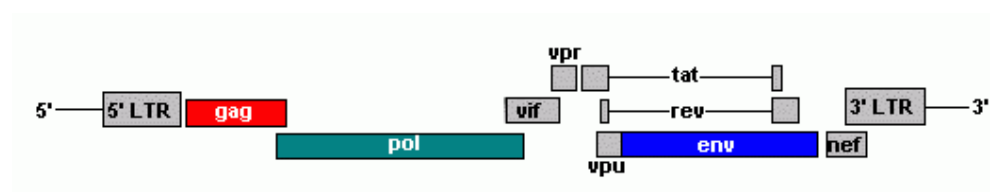


Figure 3 : Représentation schématique du génome du VIH-1

- Les séquences LTR

Une fois rétro transcrit, le génome du VIH-1 est bordé de part et d'autre d'une séquence LTR (Long Terminal Repeat), qui est présente chez tous les rétrovirus. Ces régions du génome sont essentielles à la réplication du virus. Elles permettent d'une part d'initier la

transcription inverse du génome grâce à un site d'amorçage désigné primer binding site, ou PBS. Elles facilitent également, au cours de ce processus, le transfert d'un brin d'ADN en cours de synthèse d'une extrémité à l'autre du génome viral. Enfin, elles contiennent de nombreuses séquences capables d'interagir avec des facteurs de transcription de la cellule hôte. La LTR 5' joue ainsi le rôle de promoteur de la transcription du génome viral par la cellule hôte.

- gag

Le gène gag (pour « group antigen ») code pour une polyprotéine de 55kDa (Gag, p55), qui est exprimée à partir d'un ARNm non épissé. Gag peut être clivé par la protéase en plusieurs autres protéines, qui permettent de structurer la particule virale. La protéine de matrice (MA, p17), qui forme la partie externe du coeur viral, est dérivée du domaine N-terminal myristoylé [16] et hydrophobe de Gag, ce qui l'amène à interagir avec la double couche lipidique du virus. Deux autres produits de ce clivage sont la protéine de capside (CA, p24) et la protéine de nucléocapside (NC, p7). Cette dernière est destinée à recouvrir le brin d'ARN viral. Elle le stabilise grâce à deux doigts de zinc [17], permet son incorporation dans de nouvelles particules virales et facilite sa transcription inverse [18]. Enfin, un polypeptide p6 joue le rôle de médiateur entre p55 et la protéine accessoire Vpr, ce qui permet l'intégration de cette dernière dans la particule virale [19]. p6 comporte également un domaine essentiel à la libération de nouveaux virions par la cellule hôte [20].

- pol

Les protéines codées par le gène pol comprennent la protéase virale, l'intégrase et la transcriptase inverse qu'on va définir ultérieurement. Le gène pol n'est jamais exprimé de manière indépendante, mais sous forme d'une polyprotéine de fusion Gag-Pol [21], générée suite à une lecture alternative des ARNm (« frame shifting ») viraux par les ribosomes. Cet événement, qui est déclenché par un motif d'ARN spécifique [22], ne se produit que pour environ 5% des ARNm viraux traduits. Les précurseurs Gag et Gag-Pol sont donc produits au rythme d'environ 20 Gag pour 1 Gag-Pol. Par la suite, la protéase virale sépare les domaines Gag et Pol du polypeptide et clive Pol en protéines fonctionnelles.

- env

Le précurseur Env de 160 kDa (gp160) est exprimé à partir d'un ARNm épissé issu du gène env. Après synthèse, il doit être glycosylé (addition de 25 à 30 chaînes d'hydrates de

carbones) pour conférer au virus son caractère infectieux [23]. Une protéase cellulaire clive ensuite env en protéines d'enveloppe. La glycoprotéine de surface (SU, gp120) qui résulte de cette opération permet la reconnaissance spécifique des récepteurs CD4 et des co-récepteurs portés par les cellules cibles du VIH-1 [24]. Elle est présente sous forme de trimère à la surface de la particule virale et comporte cinq domaines variables (désignés V1 à V5), dont certains sont essentiels à la réplication du virus. Sa structure a été déterminée par diffraction de rayons X [25]. La glycoprotéine transmembranaire (TM, gp41) est également rencontrée sous forme de trimère et interagit étroitement avec gp120. Elle permet la fusion de la membrane cellulaire et de la membrane virale, ce qui permet l'introduction du contenu viral dans le cytoplasme de la cellule infectée.

- Les gènes « annexes »

En plus des trois gènes principaux, communs à tous les rétrovirus, le génome du VIH-1 possède plusieurs autres gènes, plus spécifiques. Deux d'entre eux, **tat** et **rev**, codent pour des protéines de régulation. La protéine Tat (transactivateur de la transcription) permet principalement d'augmenter la transcription du génome viral par la cellule hôte [26]. La protéine Rev (Regulator of Viremia) est pour sa part destinée à promouvoir l'export d'ARNm viraux incomplètement épissés hors du noyau cellulaire [27]. Cette opération est essentielle à l'expression des précurseurs Gag, Gag-Pol et Env, produits à partir de ces ARNm. Quatre gènes « accessoires » sont également inclus dans le génome viral. Les protéines qui en sont issues, même si elles ne sont pas toujours indispensables au virus (tout au moins dans des expériences ex vivo), augmentent grandement son caractère infectieux. Vif (Viral Infectivity Factor) protège le brin d'ADN issu de la transcription inverse des effets d'une protéine endogène (CEM15/APOBEC3G) qui tend à y introduire de nombreuses mutations [28]. Vpu (Viral protein U) induit la dégradation des récepteurs CD4 de la cellule hôte et favorise la libération de particules virales [29]. Vpr (Viral Protein R) permet de bloquer le cycle cellulaire durant la phase (G2) où la transcription du génome viral est maximale [30]. Enfin, **Nef** (Negative Effector) est une protéine de 24 kDa aux multiples fonctions, qui tendent à modifier l'environnement cellulaire en vue d'optimiser la production de particules virales (blocage de l'apoptose, changement de l'état d'activation de la cellule,...).

3.6 Le cycle de vie du VIH

En tant que rétrovirus, le virus de l'immunodéficience humaine ne peut assurer seul sa subsistance et sa réplication. Il doit pour ce faire infecter une cellule hôte dont il détourne les fonctions en vue de produire de nouveaux virions. Comprendre ce processus nécessite la connaissance non seulement des gènes du VIH-1 et des fonctions qui y sont associées, mais aussi de la manière dont ces éléments interagissent avec ceux de la cellule hôte.

Le cycle de vie du VIH-1 (résumé à la figure 4) comprend sept étapes principales : la reconnaissance de la cellule hôte, la pénétration du virus au sein de la cellule par fusion des membranes, la rétrotranscription du génome viral en un ADN double brin, son intégration au sein du génome de la cellule hôte, la transcription du génome viral en vue de l'expression de polyprotéines, l'assemblage et la libération de virions immatures et enfin la maturation de ces nouvelles particules.

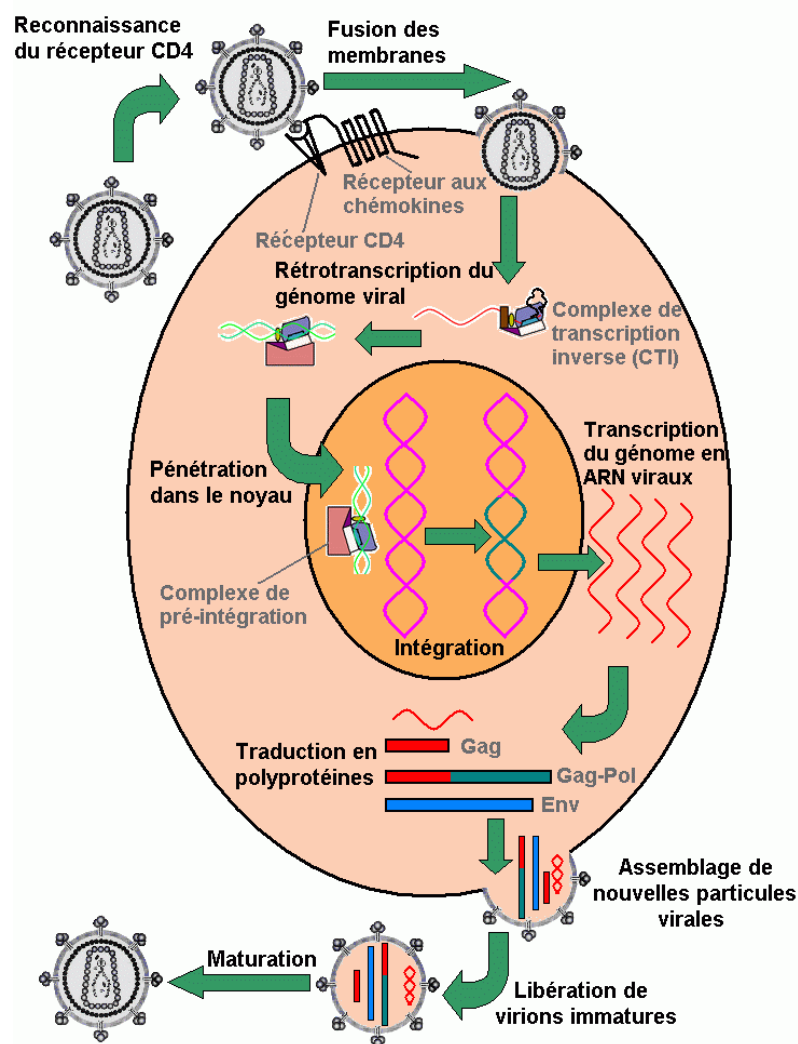


Figure 4: Cycle de réplication du VIH-1

3.6.1 Reconnaissance de la cellule hôte

La première étape du cycle de vie du virus consiste en une interaction peu spécifique entre une boucle (V3) de la glycoprotéine de surface (gp120, ou SU), qui est chargée positivement et le sulfate d'héparan, chargé négativement et présent à la surface de nombreuses lignées cellulaires [31, 32]. Ceci permet une première adsorption du virus. Le VIH-1 s'attaque cependant de manière quasiment exclusive à des lignées cellulaires porteuses du récepteur CD4 [33], c'est à dire principalement des lymphocytes T et certains macrophages comme nous avons indiqué précédemment.

3.6.2 Pénétration dans la cellule

Même si elle est indispensable à l'infection par le VIH, la liaison d'une glycoprotéine SU (gp120) à un récepteur CD4 n'est pas suffisante pour lui permettre d'entrer efficacement dans la cellule hôte. Elle provoque cependant chez SU une série de changements de conformation qui lui permettent d'interagir avec les récepteurs aux chémokines. Bien qu'une douzaine d'entre eux aient été identifiés en tant que co-récepteurs possibles du VIH-1, seuls deux jouent réellement un rôle *in vivo* [34]. Le premier, CXCR4 (aussi appelé fusine), permet l'entrée des souches virales responsables de l'infection des lymphocytes T (le virus est alors dit X4 ou T-tropique). Le deuxième est désigné CCR5 et est un récepteur naturel des chémokines MIP-1 et RANTES. Il est employé par les variétés s'attaquant aux macrophages (souches dites R5 ou macrophage-tropiques). Les deux sortes de virus (R5 ou X4), peuvent être rencontrées chez le même patient, mais sont généralement présentes à des périodes différentes de la maladie. Les virus de type R5 sont en effet souvent considérés comme responsables de la transmission du virus par voie intraveineuse ou mucoale, tandis que les souches X4 sont d'avantage rencontrées lorsque le patient évolue vers le SIDA [35]. Ceci explique pourquoi une délétion de 32 paires de bases dans le gène codant pour CCR5 confère aux individus homozygotes pour cette mutation (1 à 2% de la population caucasienne) une immunité virtuelle à l'infection par le VIH [36, 37].

Les interactions entre SU, le récepteur CD4 et les récepteurs aux chémokines induisent un autre changement de conformation, cette fois chez la glycoprotéine gp41 (TM), qui est le médiateur majeur de l'entrée du virus [38]. L'ouverture du trimère TM projette trois domaines de fusion peptidiques lipophiles, qui s'incrètent dans la membrane cellulaire. Ils provoquent la fusion des membranes du virus et de la cellule hôte et permettent ainsi au coeur viral de pénétrer dans la cellule.

3.6.3 Rétrotranscription du génome viral

Les informations génétiques du VIH sont sous forme d'ARN. Pour assurer l'intégration de ce matériel génétique à celui de la cellule il doit y avoir une étape permettant la *traduction* de l'ARN viral en ADN, c'est la transcriptase. La transcriptase inverse est une enzyme, contenue à l'origine dans la capside virale, qui permet à l'aide des nucléosides contenus dans la cellule de construire un brin d'ADN viral à partir de l'ARN [39, 40].

L'ADN ainsi produit peut être intégré à l'ADN cellulaire, ce qui est la première étape vers la synthèse de nouveaux virus [41, 42].

3.6.4 Intégration dans le génome de la cellule hôte

Afin d'être intégré dans le génome de la cellule hôte, l'ADN synthétisé par transcription inverse doit tout d'abord être conduit vers le noyau de la cellule hôte. Cette prise en charge est réalisée grâce à la formation d'un complexe de pré-intégration (CPI), composé des protéines virales IN, MA, Vpr et TI et de la protéine cellulaire HMGI(Y) [43]. Ce CPI est capable de migrer vers le noyau, en tirant parti des microtubules [44].

Contrairement aux autres rétrovirus, les lentivirus peuvent intégrer une séquence d'ADN dans le génome d'une cellule hôte sans que celle-ci soit en phase de division [45]. Cette capacité des lentivirus en fait d'ailleurs des vecteurs très étudiés dans le cadre de la thérapie génique [46, 47]. Pour ce faire, le complexe de pré-intégration, qui contient le provirus d'ADN, doit dans un premier temps franchir la membrane nucléaire, alors que sa taille (environ 28 nm) est le double de celle d'un pore nucléaire [48]. L'import du CPI dans le noyau réclame donc bon nombre d'interactions, et il n'est donc pas étonnant qu'au moins trois protéines virales impliquées dans ce complexe (MA, IN et Vpr) possèdent un signal de localisation nucléaire (SLN) [49-50]. Ces signaux interagissent avec la karyophérine α , qui prend en charge le CPI et le dirige vers le noyau. En outre, une zone du génome viral caractérisée par une triple hélice d'ADN (« DNA flap ») pourrait permettre la liaison à une protéine cellulaire contenant un SLN supplémentaire [51].

Une fois le CPI à l'intérieur du noyau, le provirus peut être intégré dans le génome de la cellule hôte. Pour ce faire, l'intégrase doit tout d'abord ôter deux nucléotides à chaque extrémité 3' du provirus d'ADN [43]. Cette première étape est en fait réalisée peu après la transcription inverse du génome viral. L'ADN génomique est ensuite clivé au niveau du site d'intégration. Le choix de ce site est probablement davantage conditionné par son accessibilité que par la reconnaissance d'une séquence spécifique [52, 53]. Enfin, une

activité ligase crée un seul lien covalent entre chaque extrémité du provirus et l'ADN génomique (réaction de transfert de brin). La réalisation de ces trois opérations nécessite la présence de cations métalliques divalents, tels que Mg^{2+} [54]. Une fois le transfert de brin effectué, les enzymes cellulaires réparent le site d'intégration.

3.6.5 Transcription du génome et expression des protéines virales

L'intégration du provirus dans le génome de la cellule hôte ne conduit pas systématiquement à la production immédiate de nouvelles particules virales [55]. Il est possible que l'infection de la cellule soit suivie d'une phase de latence, durant laquelle le génome du VIH-1 n'est pas exprimé. Un des facteurs qui conditionnent cette expression est l'activité de la cellule hôte. Il est par exemple bien établi que le VIH peut rester à l'état latent dans les lymphocytes T au repos, et ce pendant plusieurs années. Un autre facteur est le site d'intégration du génome viral. Un provirus incorporé dans une zone de la chromatine où la transcription est réprimée tend à se montrer moins productif [56]. Dans la plupart des cas cependant, plusieurs provirus ont été intégrés et il est probable qu'au moins un d'entre eux soit situé dans une zone active du point de vue de la transcription. D'autres causes doivent donc être trouvées pour expliquer le phénomène de latence.

L'une d'entre elles est le rôle des facteurs de transcription cellulaires capables de se lier à la séquence LTR du génome viral. On sait ainsi que la présence de facteurs comme NF- κ B ou NFAT permet de favoriser la transcription du génome viral [57]. Leur absence peut donc induire une phase de latence. Enfin, une expression efficace du génome viral nécessite la présence, en quantité suffisante, de la protéine virale Tat (transactivateur de la transcription). Tat se lie à l'ARN viral produit lors de la transcription, en interagissant avec une séquence désignée TAR (Transactivation Response element) [58] et recrute un complexe (P-TEFb) contenant la cycline T1 et la protéine kinase (Cdk9) [59]. Ce complexe phosphoryle alors l'ARN polymérase II, ce qui augmente d'un facteur 1000 le taux de transcription du génome viral [60].

Une fois initiée, la transcription du génome du VIH-1 peut conduire à plus d'une douzaine de transcrits différents [61]. Les ARNm complètement épissés, issus des gènes *nef*, *rev* et *tat*, sont libres de quitter immédiatement le noyau de la cellule hôte [62]. Ces gènes viraux sont les premiers à être traduits en protéines et constituent donc les gènes précoces du VIH. Les ARN non épissés ou incomplètement épissés ne peuvent par contre être traduits qu'en présence d'une quantité suffisante de la protéine Rev [62], qui joue le rôle de facteur

d'export nucléaire. Ils contiennent les gènes tardifs du VIH-1. Certains brins d'ARN ne seront pas traduits et constitueront le matériel génétique de futurs virions.

Si Tat et Rev agissent sur la transcription du génome viral et son expression, Nef a pour rôle de modifier l'environnement que constitue la cellule hôte, de manière à optimiser la réplication du virus [34]. Il peut par exemple diminuer l'expression des récepteurs CD4 à la surface de la cellule [63] ou causer des réarrangements dans le cytosquelette [64]. Son action n'est cependant pas limitée à l'intérieur de la cellule hôte. Nef joue en effet un rôle à la fois offensif et défensif vis à vis du milieu extérieur. Offensif, car en activant l'expression de FasL, Nef favorise l'apoptose des cellules voisines qui expriment le récepteur Fas [65]. Ceci inclut des lymphocytes T cytotoxiques, qui pourraient éliminer la cellule infectée. Défensif, car Nef diminue l'expression du complexe d'histocompatibilité de classe I (MHCI), ce qui évite la détection de la cellule infectée par les lymphocytes T cytotoxiques CD8+ [66]. En outre, Nef a la capacité de bloquer l'apoptose de la cellule hôte [67, 68]. Tous ces mécanismes contribuent à prolonger la vie de la cellule hôte et donc à augmenter la production de nouveaux virions. D'autres facteurs contribuent également à la modification de l'environnement intracellulaire. Ainsi, Vpr bloque le cycle cellulaire dans la phase G2 [30, 69], c'est à dire au moment où la transcription du virus est maximale [70].

3.6.6 Formation de nouvelles particules virales

L'assemblage et le bourgeonnement de nouvelles particules virales ont lieu au niveau de la membrane cellulaire. Ce processus est dirigé par les polyprotéines Gag, Gag-Pol et, dans une moindre mesure, Env. La formation d'une particule virale immature requiert l'assemblage d'environ 1500 précurseurs Gag [71]. Elle est facilitée par une protéine cellulaire (HP68), qui agit en tant que chaperonne et facilite les changements de conformation de Gag [72]. L'incorporation de la protéine Vif est également essentielle à la production de particules virales infectieuses. Les dernières étapes du bourgeonnement des particules virales dépendent d'un processus d'ubiquitination et de la protéine TSG 101, associée au tumor suppressor gene 101. TSG 101 se lie à une séquence d'acides aminés (PTAP) présente au sein de Gag et reconnaît les ubiquitines, ce qui permet la libération de virions [73, 74]. TSG 101 s'associe en temps normal avec d'autres protéines cellulaires, pour permettre la formation de corps vésiculaires dont le contenu est dégradé par les lysosomes [75]. Le VIH a donc détourné ce processus naturel de bourgeonnement pour permettre la libération de nouveaux virions.

3.6.7 Maturation des particules virales

Les particules virales libérées par la cellule infectée sont dans un premier temps non infectieuses et doivent encore subir une phase de maturation. La première étape de ce processus est la formation d'un dimère de précurseurs Gag-Pol, au niveau des domaines correspondant à la protéase virale. Le dimère peut alors s'autocliner, pour libérer une protéase mature [76]. Ce processus est le fruit d'un équilibre bien précis dans l'expression des polyprotéines. Ainsi, une expression trop élevée de Gag-Pol provoque l'activation intracellulaire de la protéase, ce qui interfère avec l'assemblage des particules virales [77]. Une fois la protéase virale libérée, les précurseurs Gag et Gag-Pol sont convertis en protéines structurales et fonctionnelles, ce qui donne au virion sa structure définitive. La nouvelle particule virale est alors capable de s'attaquer à une cellule porteuse du récepteur CD4.

3.7 Les thérapies anti-VIH

L'étude approfondie du VIH a permis de mieux comprendre son cycle de vie. Ces informations ont mené à la recherche et au développement de nombreux traitements destinés à combattre ce virus. Ceux-ci peuvent être classés en deux grandes catégories : les vaccins et la chimiothérapie. La première est une approche immunologique et vise soit à empêcher l'infection en préparant l'organisme à une exposition au VIH, soit à combattre l'infection en stimulant les mécanismes immunitaires du patient. La deuxième se base sur l'**emploi de substances chimiques** destinées à enrayer la réplication du virus.

3.7.1 Les vaccins

Dans le contexte d'une épidémie de SIDA qui prend des proportions alarmantes, la découverte d'un **vaccin** anti-VIH prophylactique (préventif) ou thérapeutique constituerait une véritable bouée de sauvetage pour un grand nombre de pays. Ceci est tout spécialement vrai pour les pays en voie de développement, qui ne peuvent se permettre les frais liés à la distribution de médicaments très coûteux aux millions de personnes touchées. Le développement de tels vaccins, même si il a été rapidement envisagé, s'est cependant heurté à plusieurs obstacles majeurs [78]. La persistance du virus intégré dans les cellules infectées en phase de latence (et plus particulièrement au sein des lymphocytes T CD4⁺ au repos) [79] rend très difficile l'éradication du VIH de l'organisme d'un patient. La grande variabilité du virus (il existe 11 sous-types du VIH-1 dans le monde) complique fortement la recherche d'un immunogène capable d'induire une réponse efficace face aux différentes

variantes du VIH. Enfin, le taux d'erreur élevé de la transcriptase inverse provoque l'apparition de nombreuses souches virales mutantes, qui peuvent échapper au contrôle immunologique induit par un vaccin [80].

3.7.2 Les chimiothérapies

Dans l'attente d'un vaccin efficace, la lutte contre le VIH est actuellement basée sur l'emploi de molécules capables de ralentir la prolifération du virus, ou idéalement de la bloquer. En effet, Une vingtaine de médicament antirétroviraux sont actuellement disponibles et agissent d'une part sur les enzymes nécessaire à la réplication du virus, et d'autre part sur les mécanismes d'entrée du virus dans la cellule (**Figure 5**).

Une brève description des différents inhibiteurs est schématisée sur la Figure 6.

3.7.2.1 Inhibition de la fusion :

Avant de pouvoir entrer dans une cellule, le VIH doit s'attacher à cette dernière par le biais d'un processus appelé *fusion*. Les composés qui empêchent le VIH de se fusionner aux cellules s'appellent des *inhibiteurs de fusion*. La recherche de ces derniers fut une nouvelle stratégie de lutte contre le SIDA.

Un certain nombre de composés tels que les lécithines, les polypeptides et les triterpène ont été élaborés pour perturber la fusion du virus dans les cellules.

Les chercheurs continus à proposer d'autres composés qui sont aux stades de développement les plus avancés.

3.7.2.2 Inhibiteur de la transcriptase inverse:

La rétrotranscription est une étape centrale du cycle répliatif de VIH et permet la synthèse de l'ADN proviral bicaténaire à partir de l'ARN génomique.

Grâce, à son rôle important dans le cycle répliatif du VIH, et à son absence au niveau des cellules hôtes, la transcriptase inverse (TI) représente la meilleure cible de la chimiothérapie antivirale. Elle acquis le plus grand intérêt des chercheurs dans ce domaine. La plupart des composés autorisés par la FDA (*Food and Drug Administration*) sont des inhibiteurs de la TI. En effet, elle est responsable du codage de l'information génétique virale à l'intérieur de la cellule cible. Elle transforme l'ARN génomique viral en un double brin d'ADN proviral qui s'intégrera ultérieurement dans les chromosomes de la cellule hôte.

Il existe deux catégories d'inhibiteurs ayant comme cible la transcriptase inverse.

- Les inhibiteurs nucléosidiques de la transcriptase inverse (INTI) :

Ils constituent historiquement la première classe de médicaments disponibles pour le traitement anti-SIDA. L'AZT (3'-azido-3'deoxythymidine), le premier analogue synthétisé des INTIs, est le médicament le plus administré jusqu'à aujourd'hui pour le traitement du SIDA. Tous les composés de cette classe sont des didéoxynucléosides (ddN). Pour leur activation, ils nécessitent une triple phosphorylation après leur entrée à la cellule. Le didéoxynucléoside triphosphate (ddNTP), ainsi formé, exerce une inhibition compétitive du substrat naturel, le déoxynucléoside triphosphate. Il agit donc comme terminateur de l'élongation de la chaîne d'ADN naissante³⁷. Les ddNTP inhibent, en outre, les ADN polymérases cellulaires. Ceci leur attribue une toxicité élevée vis-à-vis de la cellule hôte.

- Les inhibiteurs non nucléosidiques de la transcriptase inverse (INNTI) :

Ce sont des composés ayant une structure chimique très variable, d'une molécule à l'autre mais sont hautement actifs sur la transcriptase inverse. À la différence des INTIs, ils ont un effet inhibiteur direct sur la TI du VIH-1 en formant une liaison réversible et non compétitive avec l'enzyme. Ils ont fait preuve d'une activité remarquable en association avec les analogues nucléosidiques et/ou antiprotéases.

3.7.2.3 Inhibition de l'intégrase

L'intégrase est une enzyme qui joue un rôle important dans l'intégration du matériel génétique du virus à l'intérieur de la cellule hôte. Il est licite donc d'espérer trouver des molécules capables d'inhiber l'activité de cette enzyme [81].

3.7.2.4 Inhibition de la transcription du génome virale

En plus des inhibiteurs des gènes **tat** et **rev**, les oligonucléotides antisens se sont révélés capables d'inhiber la réplication du virus en agissant sur différentes étapes de la transcription. Leur fixation sur l'ADN proviral empêche sa transcription en ARNm et leur fixation sur l'ARNm bloque sa traduction en protéine. Cependant, leur emploi se heurte à des problèmes qu'il faudra résoudre. Le premier d'entre eux est en liaison avec leur structure polyanionique; leur couplage à des polycations ou à des molécules lipophiles est envisagé dans ce sens. Le second problème est leur dégradation par les nucléases.

3.7.2.5 Inhibition de la protéase

La protéase est une enzyme qui joue un rôle crucial dans le cycle de réplication du VIH. Son rôle est le clivage des protéines précurseurs Pr160gag-pol et Pr55gag en petits fragments qui correspondent aux protéines de la capsid virale, à la transcriptase et à l'intégrase, les principaux enzymes du VIH. L'action des inhibiteurs de la protéase aboutit donc à la formation de particules virales immatures et non infectieuses. Plusieurs composés agissant comme inhibiteurs de la protéase ont été approuvés par la FDA, d'autres sont en phase d'expérimentation.

Tous ces inhibiteurs inhibent non seulement la réplication du VIH lors d'un cycle lytique mais aussi la production virale dans les cellules chroniquement infectées.

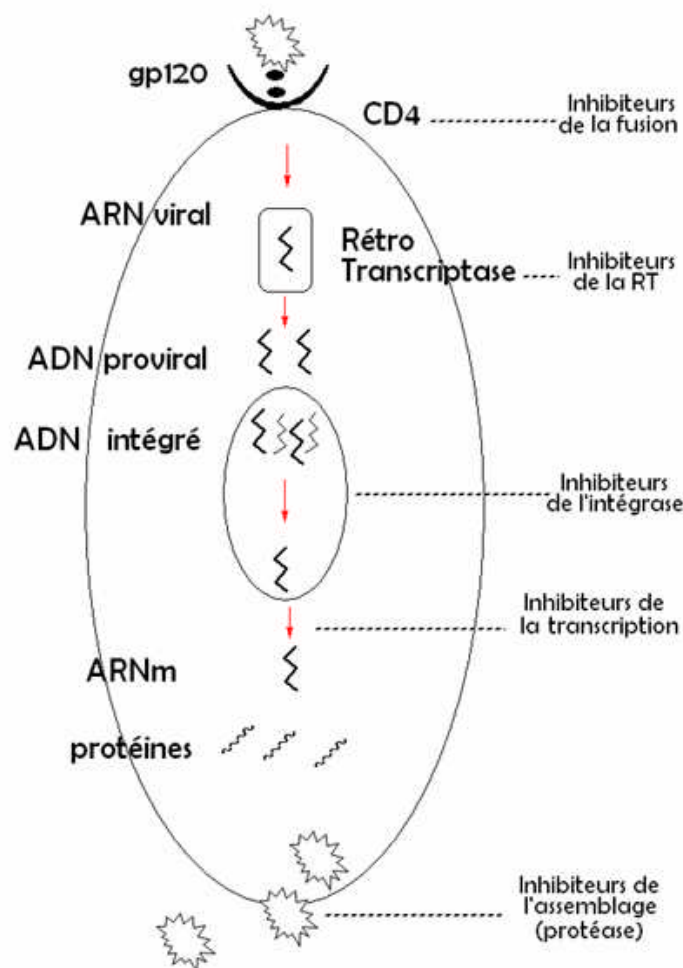


Figure 5: Stratégies Anti-VIH [81].

REFERENCE.

- [1]- Hymes, K.B.; J.B. Greene; A. Marcus.; al., Kaposi's sarcoma in homosexual men: A report of eight cases. *Lancet*, 1981. 2: p. 598-600.
- [2]- Gottlieb, M.S.; R. Schroff; H.M. Schanker. *The New England Journal of Medicine*, 1981. 305: p. 1425-1431.
- [3]- Altman, L.K., Rare cancer seen in 41 Homosexuals, in *The New York Times*. 1981
- [4]- Brennan, R.O.; D.T. Durack, Gay Compromise syndrome. *Lancet*, 1981. 2(1338-1339).
- [5]- Altman, L.K., New Homosexual Disorder Worries Officials, in *the New York Times*. 1982.
- [6]- Masur, H.; M.A. Michelis; J.B. Greene; I. Onorato; R.A. Stouwe; R.S. Holzman; G. Wormser; L. Brettman; M. Lange; H.W. Murray; S. Cunningham-Rundles. *The New England Journal Of Medicine*, 1981. 305(24): p. 1431-1438.
- [7]- Barré-Sinoussi, F.; J.C. Chermann; F. Rey; M.T. Nugeyre; S. Chamaret; J. Gruest; C. Danguet; C. Axler-Blin; F. Brun-Vezinet; C. Rozioux; W. Rozenbaum; L. Montagnier. *Science*, 1983. 220: p. 869-871.
- [8]- Gallo, R.C.; P.S. Sarin; E.P. Gelmann; M. Robert-Guroff; E. Richardson; V.S. Kalyanaraman; D. Mann; G.D. Sidhu; R.E. Stahl; S. Zolla-Pazner; J. Leibowitch; M. Popovic. *Science*, 1983. 220: p. 865-867.
- [9]- Van de Perre, P.; R. D.; P. Lepage. *Lancet*, 1984. 2: p. 62-65.
- [10]- Mitsuya, H.; K.J. Weinhold; P.A. Furman; M. St Clair; S. Nusinoff-Lehrman; R.C. Gallo; D. Bolognesi; D.W. Barry; S. Broker. *Proc Natl Acad Sci U S A.*, 1985. 82: p. 7096-7100.
- [11]- Holodniy, M.; V. Miller, Human immunodeficiency virus, in *Practical guidelines in antiviral therapy*, C.A.B.B.G.J. Galasso, Editor. 2002, Elsevier Science. p. 95-125.
- [12]- Clavel, F.D.; D. Guetard; F. Brun-Vezinet; S. Chamaret; F. Rey; M.O. Santos-Ferreira; A.G. Laurent; C. Danguet; C. Katlama; C. Rosioux. *Science*, 1986. 233: p. 343-346.
- [13]- Fauci, A.S.; R.C. Desrosiers, Pathogenesis of VIH and SIV. *Retroviruses*, 1997: p. 587-635.
- [14]- Ratner, R., VIH life cycle and genetic approaches. *Perspective in drug discovery and design*, 1993. 1: p. 3-22.
- [15]- Franke, E.K.; H.E. Yuan; J. Luban, Specific incorporation of cyclophilin A into VIH-1 virions. *Nature.*, 1994. 372(6504): p. 359-62.
- [16]- Bryant, M.; L. Ratner. *Proc Natl Acad Sci U S A.*, 1990. 87: p. 523-527.
- [17]- Roques, B.P.; N. Morellet; H. de Rocquigny; H. Démémé; W. Schueler; N. Jullian, *Biochimie*, 1997. 79(673-680).
- [18]- Lapadat-Tapolsky, M.; H. de Rocquigny; D. Van Gent. *Nucleic Acids Res.*, 1993. 21: p. 831-839.
- [19]- Paxton, W.; R.I. Connor; N.R. Landau. *J. Virol.*, 1993(67): p. 7229-7237.
- [20]- Garnier, L.; L.J. Parent; B. Rovinski; S.X. Cao; J.W. Wills. *J Virol.*, 1999. 73(3): p. 2309-20.
- [21]- Jacks, T.; M.D. Power; F.R. Masiarz. *Nature*, 1988. 331: p. 280-283.
- [22]- Parkin, N.T.; M. Chamorro; H.E. Varmus. *J. Virol.*, 1992. 66: p. 5147-5151.
- [23]- Capon, D.J.; R.H. Ward. *Annu Rev Immunol.*, 1991. 9: p. 649-678.
- [24]- Landau, N.R.; M. Warton; D.R. Littman. *Nature*, 1988. 334: p. 159-153.
- [25]- Kwong, P.D.; R. Wyatt; J. Robinson; R.W. Sweet; J. Sodroski; W.A. Hendrickson, *Nature.*, 1998. 393(6686): p. 648-59.
- [26]- Feinberg, M.B.; D. Baltimore; A.D. Frankel. *Proc Natl Acad Sci U S A.*, 1991. 88: p. 4045-4049.
- [27]- Malim, M.H.; B.R. Cullen. *Mol Cell Biol.*, 1993. 13(10): p. 6180-6185.
- [28]- Stopak, K.; C.M. de Noronha; W. Yonemoto; W.C. Greene. *Molecular Cell*, 2003. 12: p. 591-601.
- [29]- Schubert, U.; S.P. Bour; A.V. Ferrer-Montiel. *J. Virol.*, 1996. 70: p. 809-819.
- [30]- Emerman, M., Vpr and the cell cycle. *Curr. Biol.*, 1996. 6: p. 1096-1103.

- [31]- Ugolini, S.; I. Mondor; Q.J. Sattentau. another look. Trends in microbiology, 1999. 7(4): p. 144-9.
- [32]- Mondor, I.; S. Ugolini; Q.J. Sattentau. Journal of Virology, 1998. 72(5): p. 3623-34..
- [33]- Klatzmann, D.; E. Champagne; S. Chamaret; J. Gruet; D. Guetard; T. Hercend; J.C. Gluckman; L. Montagnier. Nature, 1984. 312: p. 767-768.
- [34]- Doms, R.W.; D. Trono, Genes Dev., 2000. 14(21): p. 2677-88.
- [35]- Scarlatti, G.; E. Tresoldi; A. Bjorndal; R. Fredriksson; C. Colognesi; H.K. Deng; M.S. Malnati; A. Plebani; A.G. Siccardi; D.R. Littman; E.M. Fenyo; P. Lusso. Nat Med., 1997. 3(11): p. 1259-65.
- [36]- Martinson, J.J.; N.H. Chapman; D.C. Rees; Y.T. Liu; J.B. Clegg. Nat Genet., 1997. 16(1): p. 100-3.
- [37]- Liu, R.; W.A. Paxton; S. Choe; D. Ceradini; S.R. Martin; R. Horuk; M.E. MacDonald; H. Stuhlmann; R.A. Koup; N.R. Landau. Cell. 1996. 86(3): p. 367-77.
- [38]- Chan, D.C.; P.S. Kim. Cell. 1998. 93(5): p. 681-4.
- [39]- Ventura, M.; L. Tarrago-Litvak; V. Dolle; C.H. Nguyen; M. Legraverend; H.J.A. Fleury; S. Litvak. ArcVHes of Virology, 1999. 144(3): p. 513-523.
- [40]- Karageorgos, L.; P. Li; C. Burrell, Characterization of VIH replication complexes early after cell-to-cell infection. AIDS Res Hum Retroviruses., 1993. 9(9): p. 817-23.
- [41]- Wu, X.; H. Liu; H. Xiao; J.A. Conway; E. Hehl; G.V. Klapana; V. Prasad; J.C. Kappes, Journal of Virology, 1999. 73(3): p. 2126-2135.
- [42]- Bukrinskaya, A.; B. Brichacek; A. Mann; M. Stevenson, J Exp Med., 1998. 188(11): p. 2113-25.
- [43]- Miller, M.D.; C.M. Farnet; F.D. Bushman. J Virol., 1997. 71(7): p. 5382-90.
- [44]- McDonald, D.; M.A. Vodicka; G. Lucero; T.M. Svitkina; G.G. Boris; M. Emerman; T.J. Hope. J Cell Biol., 2002. 159(3): p. 441-52.
- [45]- Weinberg, J.B.; T.J. Matthews; B.R. Cullen; M.H. Malim, J Exp Med., 1991. 174(6): p. 1477-82.
- [46]- Imren, S.; P. E.; K.A. Westerman; R. Pawliuk; M.E. Fabry; C.J. Eaves; B. Cavilla; L.D. Wadsworth; Y. Beuzard; E.E. Bouhassira; R. Russel; I.M. London; R.L. Nagel; P. Leblouch; R.K. Humphries. Proc Natl Acad Sci U S A., 2002. 99(22): p. 14380-14385.
- [47]- Poeschla, E.; P. Corbeau; F. Wong-Staal. Proc Natl Acad Sci U S A., 1996. 93: p. 11395-11399.
- [48]- Pemberton, L.F.; G. Blobel; J.S. Rosenblum. Curr Opin Cell Biol., 1998. 10(3): p. 392-9.
- [49]- Gallay, P.; T. Hope; D. Chin; D. Trono. Proc Natl Acad Sci USA., 1997. 94(18): p. 9825-30.
- [50]- Bukrinsky, M.I.; S. Haggerty; M.P. Dempsey; N. Sharova; A. Adzhubel; L. Spitz; P. Lewis; D. Goldfarb; M. Emerman; M. Stevenson. Nature., 1993. 365(6447): p. 666-9.
- Heinzinger, N.K.; M.I. Bukinsky; S.A. Haggerty; A.M. Ragland; V. Kewalramani; M.A. Lee; H.E. Gendelman; L. Ratner; M. Stevenson; M. Emerman, Proc Natl Acad Sci USA., 1994. 91(15): p. 7311-5.
- [51]- Zennou, V.; C. Petit; D. Guetard; U. Nerhbass; L. Montagnier; P. Charneau, Cell., 2000. 101(2): p. 173-85.
- [52]- Carreau, S.; C. Hoffman; F. Bushman. J. Virol., 1998. 72: p. 4005-4014.
- [53]- Stevens, S.W.; J.D. Griffith. J. Virol., 1996. 70: p. 6459-6462.
- [54]- Bushman, F.D.; R. Craigie. Proceedings of the National Academy of Sciences of the United States of America, 1991. 88(4): p. 1339-1343.
- [55]- Adams, M.; L. Sharmeen; J. Kimpton; J.M. Romeo; J.V. Garcia; B.M. Peterlin; M. Groudine; M. Emerman. Proc Natl Acad Sci U S A., 1994. 91(9): p. 3862-6.

- [56]- Jordan, A.; P. Defechereux; E. Verdin. *Embo J.*, 2001. 20(7): p. 1726-38.
- [57]- Jones, K.A.; B.M. Peterlin. *Annu Rev Biochem.*, 1994. 63: p. 717-43.
- [58]- Kao, S.Y.; A.F. Calman; P.A. Luciw; B.M. Peterlin. *Nature.*, 1987. 330(6147): p. 489-93.
- [59]- Wei, P.; M.E. Garber; S.M. Fang; W.H. Fischer; K.A. Jones. *Cell.*, 1998. 92(4): p. 451-62.
- [60]- Price, D.H., P-TEFb. *Mol Cell Biol.*, 2000. 20(8): p. 2629-34.
- [61]- Saltarelli, M.J.; E. Hadziyannis; C.E. Hart; J.V. Harrison; B.K. Felber; T.J. Spira; G.N. Pavlakis. *AIDS Res Hum Retroviruses.*, 1996. 12(15): p. 1443-56.
- [62]- Cullen, B.R. *Virology.*, 1998. 249(2): p. 203-10.
- [63]- Glushakova, S.; J. Munch; S. Carl; T.C. Greenough; J.L. Sullivan; L. Margolis; F. Kirchhoff. *J Virol.*, 2001. 75(21): p. 10113-7.
- [64]- Geyer, M.; O.T. Fackler; B.M. Peterlin. *EMBO Rep.*, 2001. 2(7): p. 580-5.
- [65]- Xu, X.N.; B. Laffert; G.R. Screaton; M. Kraft; D. Wolf; W. Kolanus; J. Mongkolsapay; A.J. McMichael; A.S. Baur. *J Exp Med.*, 1999. 189(9): p. 1489-96.
- [66]- Collins, K.L.; B.K. Chen; S.A. Kalams; B.D. Walker; D. Baltimore. *Nature.*, 1998. 391(6665): p. 397-401.
- [67]- Geleziunas, R.; W. Xu; K. Takeda; H. Ichijo; W.C. Greene. *Nature.*, 2001. 410(6830): p. 834-8.
- [68]- Greenway, A.L.; D.A. McPhee; K. Allen; R. Johnstone; G. Holloway; J. Mills; A. Azad; S. Sankovich; P. Lambert. *J Virol.*, 2002. 76(6): p. 2692-702.
- [69]- Jowett, J.B.; V. Planelles; B. Poon; N.P. Shah; M.L. Chen; I.S. Chen. *J Virol.*, 1995. 69(10): p. 6304-13.
- [70]- Goh, W.C.; M.E. Rogel; C.M. Kinsey; S.F. Michael; P.N. Fultz; M.A. Nowak; B.H. Hahn; M. Emerman. *Nat Med.*, 1998. 4(1): p. 65-71.
- [71]- Wilk, T.; I. Gross; B.E. Gowen; T. Rutten; F. de Haas; R. Welker; H.G. Krausslich; P. Boulanger; S.D. Fuller. *J Virol.*, 2001. 75(2): p. 759-71.
- [72]- Zimmerman, C.; K.C. Klein; P.K. Kiser; A.R. Singh; B.L. Firestein; S.C. Riba; J.R. Lingappa. *Nature.*, 2002. 415(6867): p. 88-92.
- [73]- Garrus, J.E.; U.K. von Schwedler; O.W. Pornillos; S.G. Morham; K.H. Zavitz; H.E. Wang; D.A. Wettstein; K.M. Stray; M. Cote; R.L. Rich; D.G. Myszka; W.I. Sundquist. *Cell.*, 2001. 107(1): p. 55-65.
- [74]- VerPlank, L.; F. Bouamr; T.J. LaGrassa; B. Agresta; A. Kikonyogo; J. Leis; C.A. Carter. *Proc Natl Acad Sci U S A.*, 2001. 98(14): p. 7724-9.
- [75]- Katzmann, D.J.; M. Babst; S.D. Emr. *Cell.*, 2001. 106(2): p. 145-55.
- [76]- Louis, J.M.; N.T. Nashed; K.D. Parris; A.R. Kimmel; D.M. Jerina. *Proc Natl Acad Sci U S A.*, 1994. 91: p. 7970-7974.
- [77]- Krausslich, H.G. *Proc Natl Acad Sci U S A.*, 1991. 88: p. 3213-3217.
- [78]- McMichael, A.; T. Hanke, VIH vaccines 1983-2003. *Nature Medicine*, 2003. 9(7): p. 874-880.
- [79]- Blankson, J.N.; D. Persaud; R.F. *Annu Rev Med.*, 2002. 53: p. 557-593.
- [80]- McMichael, A. *Cell.*, 1998. 93: p. 673-676.
- [81]- Mulligan, M.J.; J. Weber. *AIDS*, 1999. 13(Suppl A): p. S105-S112.

Chapitre II

Etude théorique et RQSA des TIBO (INNTI).

Introduction.

La Transcriptase Inverse du VIH est l'enzyme responsable de la conversion de l'ARN viral en ADN double brin qui est intégré ultérieurement dans la cellule hôte. Elle est une des cibles virales les plus étudiées dans le cadre du développement de nouveaux traitements. Quatre classes d'inhibiteurs de la transcriptase inverse ont pour l'instant été découvertes : Inhibiteurs nucléosidiques, non- nucléosidiques, Inhibiteurs de dimérisation et Inhibiteurs de l'activité RNase H.

La classe d'inhibiteurs non nucléosidiques de la transcriptase inverse (INNTI) regroupe une large gamme de composés dont le premier représentant, le HEPT, est identifié dès 1989 [1]. Cette découverte est rapidement suivie par celle de dérivés des benzodiazépines (TIBO [2], Névirapine [3]), des pyridinones [4], des bis-hétéroaryl-pipérazines (BHAP) [5], des isoindolinones [6,7] et des thiocarboxanilides [8]. A l'heure actuelle, plus d'une trentaine de familles chimiques sont reconnues pour leurs propriétés d'INNTI [9].

Les dérivés du 4,5,6,7-Tetrahydro-5-methylimidazo[4,5,1-jk][1,4] benzodiazepin-2(1H)-ones sont généralement connus par l'acronyme TIBO, constituent une famille très importante des INNTI. Nous proposons, d'étudier dans ce chapitre :

- Partie 1 : L'effet de la contraction et de l'extension du cycle à 7 – sommets sur la géométrie et les propriétés physico-chimiques de ce composé.
- Partie 2 : Les Relation Quantitative entre la Structure et l'Activité anti-VIH de ces composés, une comparaison entre les différents modèles obtenus par rapport à la littérature à été faite. La non linéarité de la relation entre l'activité anti-VIH et les différents descripteurs sélectionnés est discutée.

Partie 1: Etude Théorique de l'effet de la contraction et de l'extension dans les TIBO.

Introduction.

Les médicaments, luttant contre les infections, sont fonctionnels tant qu'ils peuvent interagir fortement avec les acides aminés des protéines sur lesquelles ils doivent se fixer. La résistance aux médicaments se développe lorsqu'un ou plusieurs, des acides aminés concernés, sont mutés. L'interaction médicament protéine-cible est alors perdue.

Les seules molécules commercialisées aujourd'hui comme anti-VIH appartiennent à la classe des nucléosides et inhibent la transcriptase inverse du virus. L'AZT (3'-azido-3'-didésoxythymidine, Wellcome) [10] a été la première molécule disponible dès 1987, suivie par la ddI (2', 3'-didésoxyinosine, Bristol-Myers Squibb) et du ddC (2', 3'-didésoxycytidine, Roche).

D'autres nucléosides sont sur le point d'être enregistrés il s'agit du d4T (2', 3'-didéhydro-2', 3'-didésoxythymidine, Bristol-Myers Squibb) et du 3TC (Glaxo). La deuxième famille de molécules à venir est celle des anti-protéases le saquinavir (Roche), l'indinavir (Merck) et le ritonavir (Abbott) sont les représentants les plus avancés dans leur développement. Enfin, les inhibiteurs non nucléosidiques de la transcriptase inverse, dont il existe de nombreux représentants comme le TIBO (Janssen) [11] et la névirapine (Boehringer Ingelheim) [12], bien que très puissants sur l'enzyme, se heurtent à un problème majeur dû au développement extrêmement rapide de résistance.

Dans le but d'améliorer la performance de ce type de molécule pour pouvoir établir des liens forts qui résisterai le plus longtemps possibles aux mutations du virus. Bahram Hemmateenejad et coll. ont publié deux articles où ils ont fait une étude par le Docking de nouveaux composés résultant de la contraction [13] et de l'extension [14] du cycle à 7 sommets du cycle diazèpine des TIBO conduisant à un cycle à six et huit sommets, respectivement (Figure 1 et 2) en employant pour l'optimisation des géométries la méthode semi-empirique (AM1) [15]. Les résultats du docking de ces travaux ont révélés que ces nouvelles molécules résultantes de la contraction et de

l'extension peuvent former des liens plus résistants avec le site récepteur par rapport à la molécule d'origine (TIBO) [13].

Etant donné l'importance de ces études basées seulement sur une approche théorique semi-empirique, nous avons jugé utile de reprendre et de compléter ces travaux avec des méthodes théoriques plus performantes. En effet, notre travail consiste à expliquer, par les méthodes de la chimie quantique, l'effet de la contraction et de l'extension des TIBO sur les différents paramètres géométriques et les propriétés physico-chimiques influençant l'activité anti-VIH des différentes molécules résultantes et de comparer entre les différents résultats trouvés par les différentes méthodes théoriques à savoir, la méthode semi empirique (PM3), Hartree Fock (HF) et la Théorie de la Fonctionnelle de la Densité (DFT).

Méthode de calcul

Les différents calculs effectués dans le cadre de ce travail ont été réalisés au moyen du programme Gaussian 98[16]. Pour chaque système nous avons procédé à une optimisation complète de la géométrie et nous avons obtenus des fréquences positives pour chaque structure. Le calcul des paramètres géométriques, des énergies d'ionisation, des moments dipolaires et des charges atomiques ont été effectués au moyen des méthodes, semi-empirique (PM3) [17] et ab-initio Hartree Fock (HF) [18]. Les résultats trouvés sont comparés avec ceux obtenus dans le cadre de la Théorie de la Fonctionnelle Densité (DFT) en utilisant la fonctionnelles d'échange exacte de Becke (B3) combinée avec la fonctionnelle du gradient de correction de Lee-Yang-Parr (LYP) [19]. La base Gaussienne utilisée dans ce travail est la 6-31G*. Les calculs du « Chemical Hardness » (η) [20] du « Chemical potential » (μ) [20] et de « l'électrophilicité globale » (ω) [21] ont été calculés à partir des énergies des Orbitales Moléculaires frontières de la HOMO (Highest Occupied Molecular Orbital) et de la LUMO (Lowest Unoccupied Molecular Orbital). En utilisant le théorème de Koopmans dans les expressions du η , μ et ω on obtient :

$$\eta = (I - A)/2 \quad \mu = (I + A)/2 \quad \omega = \mu^2/2\eta$$

Où : $I = -E_{\text{HOMO}}$ et $A = -E_{\text{LUMO}}$

I et A sont respectivement le potentiel d'ionisation et l'affinité électronique.

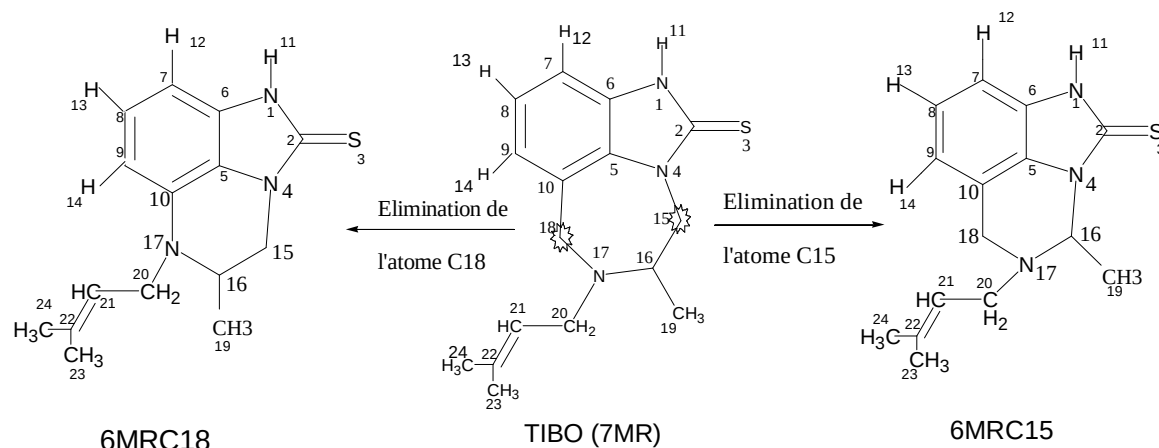


Figure 1 : Les structures et la numérotation atomique de TIBO et des deux molécules résultant de la contraction du cycle à 7 sommets [13].

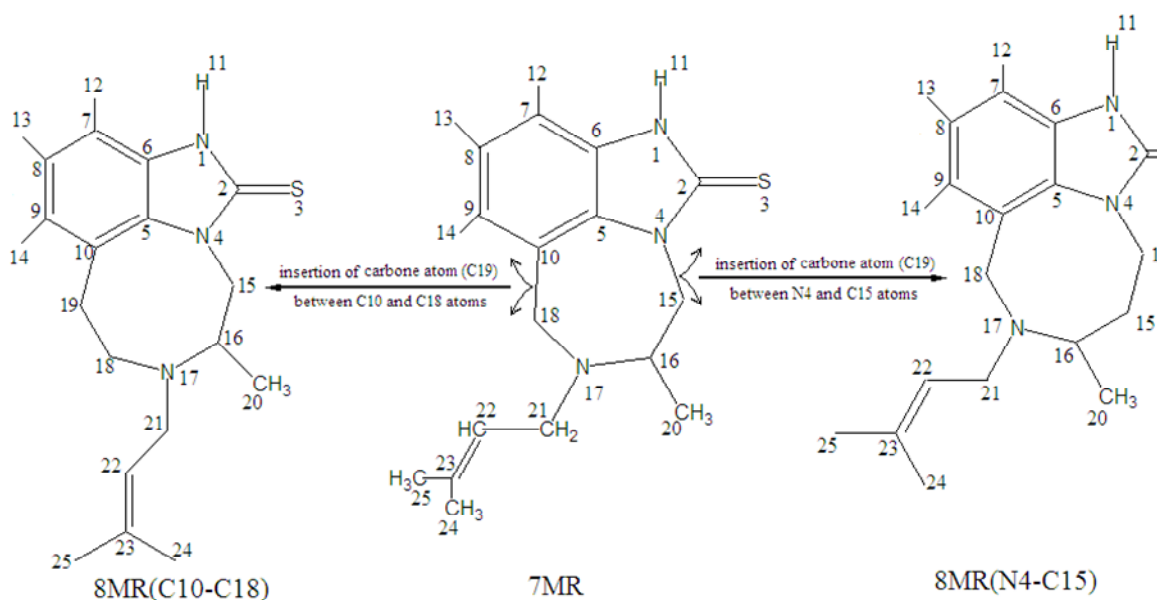


Figure 2 : Les structures et la numérotation atomique de TIBO et des deux molécules résultant de l'extension du cycle à 7 sommets [14].

La configuration du Carbone C16 dans 6MRC15 :

Le carbone asymétrique C₁₆ dans les TIBO adopte une configuration (S) [13]. La contraction du cycle à 7 sommets en position C₁₅ conduit à un changement de configuration de cet atome (C₁₆). En effet, il passe de la configuration (S) dans les TIBO à la configuration (R) dans les 6MRC₁₅.

Il est connu que les méthodes de la chimie Quantique ne permettent pas de séparer deux énantiomères [22]. En effet, les optimisations des géométries de la configuration (R) et (S) dans 6MRC₁₅ conduisent aux mêmes résultats (**Figure 3**).

Du fait que la disposition dans l'espace des groupements reliés au C₁₆ est la même pour la configuration (S) et (R) dans 6MRC₁₅, nous allons adopter dans la suite de notre travail la configuration (R) pour 6MRC₁₅.

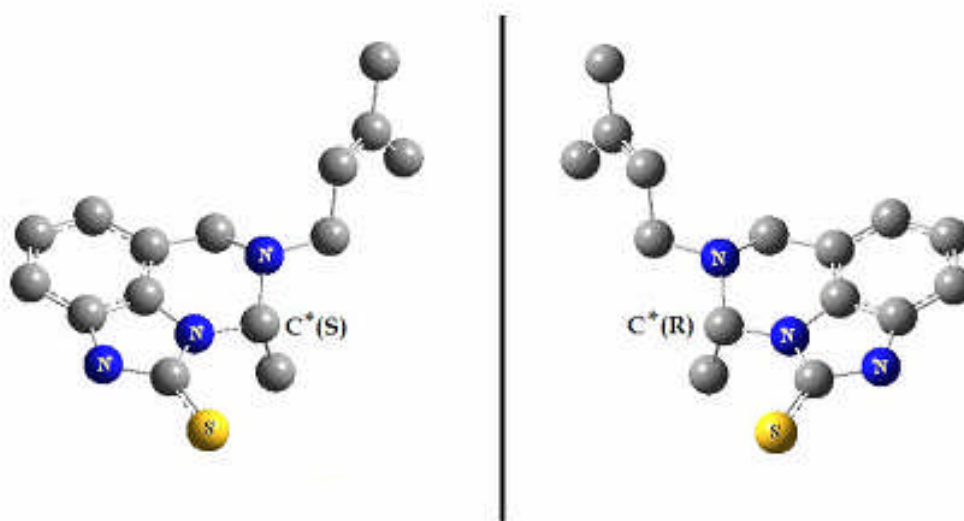


Figure 3 : Optimisation des géométries de configuration (R) et (S) dans 6MRC₁₅ sans les atomes d'hydrogène.

1.1 Effet de la contraction du cycle à 7-sommets dans les TIBO.

1.1.1 Géométrie:

La numérotation des différents atomes pour les composés TIBO, 6MRC₁₅ et 6MRC₁₈ sont montrés sur la Figure 4. La numérotation est identique pour le groupement benzodiazépinone et dans le cycle où on a effectué la contraction.

Les coordonnées internes des différents composés obtenus aux différents niveaux théoriques, après une optimisation complète des géométries en utilisant l'algorithme de Berny [23], sont rassemblées dans le tableau 1.

Dans la première partie de cette étude géométrique, on va examiner d'abord l'effet de la contraction du carbone C₁₅ et C₁₈ sur les distances, les angles de valences et les angles dièdres dans le cycle du benzène et le cycle imidazolidine-2-thione, après on va

voir son influence sur le cycle à 7 sommets (où la contraction à été faite) et sur le groupement diméthylallyl.

Les résultats obtenus pour le benzène dans les TIBO montrent que les distances C-C calculées varient entre [1.38-1.43] Å, [1.38-1.39] Å et [1.39-1.41] Å pour la méthode, PM3, HF/6-31G* et B3LYP/6-31G*, respectivement. Ces distances restent dans l'intervalle des distances expérimentales [24]. La contraction du carbone C₁₅ et C₁₈ conduit, dans tous les cas, à une petite variation dans les liaisons du benzène. En effet, nous remarquons un petit rétrécissement des liaisons C₅-C₆, C₅-C₁₀ et une petite elongation dans la liaison C₈-C₉, en moyenne de l'ordre de 0.02Å aux différents niveaux théoriques. La contraction des atomes C₁₅ et C₁₈ influe légèrement sur les angles de valence du benzène en moyenne de l'ordre de 3 degrés.

Dans le cycle imidazolidine -2- thione, la contraction du carbone C₁₅ et C₁₈ engendre une petite ouverture de l'angle de valence C₆C₅C₄ de 3 degrés, par conséquent nous remarquons le rétrécissement des distances C₂-C₄, C₄-C₅ et C₅-C₆ et l'elongation de l'ordre de 0.02Å des liaisons C₁-C₂ et N₁-C₆, aux différents niveaux théoriques.

La contraction du cycle à 7 sommets, dans les deux positions, n'influe pas sur la distance de la liaison C₂=S₃

Pour tous les composés étudiés, dans ce cas, les angles de torsions indiquent que les cycles, benzène et imidazolidine -2- thione sont portés par le même plan.

De ce fait, on conclut que les deux contractions conduisent aux mêmes résultats géométriques du côté des cycles benzène et imidazolidine -2- thione.

Dans les TIBO le cycle à 7-sommets peut avoir plusieurs conformations [25]. Ceci dépend des positions N₁₇ et C₁₆ par rapport au plan qui porte le benzène et l'imidazolidine -2- thione [26]. Nous reportons dans ce travail la structure la plus stable avec la plus basse énergie autrement dit c'est la structure qui comporte la conformation la plus stable.

Le cycle à 7-sommets adopte, dans cette étude, la conformation twist. Les différents paramètres géométriques obtenus pour cette structure sont en parfait accord avec la structure expérimentale [27].

Dans le TIBO, l'angle de torsion $C_5C_{10}C_{18}N_{17}$ est de l'ordre 40.94, 33.21 et 35.67 degrés et l'angle $C_5N_4C_{15}C_{16}$ est de l'ordre 50.60, 39.10 et 41.41 degrés au niveau PM3, HF et B3LYP, respectivement.

L'angle dièdre $C_{10}C_5N_4C_{15}$ est de l'ordre de -28.66, -17.31 et -18.46 degrés respectivement au niveau PM3, HF et B3LYP/6-31G*, cet angle devient presque plan quand on contracte le cycle à 7 sommets en position C_{18} . En effet, les calculs donnent un angle de l'ordre -5.49, -3.82 et -3.91 degrés respectivement au niveau PM3, HF et B3LYP/6-31G*. Dans le cas de la contraction en position C_{15} cet angle dièdre est remplacé par $C_{10}C_5N_4C_{16}$ qui est de l'ordre 2.98, 3.66 et 3.25 respectivement au niveau PM3, HF et B3LYP/6-31G*.

L'angle dièdre $N_4C_5C_{10}C_{18}$ dans le TIBO est de l'ordre de -5.11, -5.74 et -7.27 degrés respectivement au niveau PM3, HF et B3LYP/6-31G*. Les calculs donnent un angle plan de l'ordre de -0.08, 1.02, 0.93 pour le système MRC_{15} . Cet angle devient $N_{17}C_{10}C_5N_4$ dans le système MRC_{18} . Il est de l'ordre 4.77, 1.98 et 1.29 degrés respectivement au niveau PM3, HF et B3LYP/6-31G*.

La contraction de TIBO en position C_{18} , conduit au composé $6MRC_{18}$. L'optimisation de la géométrie de ce dernier conduit à une petite élongation de la liaison $N_{17}-C_{10}$ comparée à la littérature (1.352-1.368) Å [24]. En effet, nous trouvons des distances de l'ordre de 1.43, 1.39 et 1.40 Å au niveau PM3, HF/6-31G* et B3LYP/6-31G*, respectivement. Nous remarquons qu'il n'y a pas de délocalisation des électrons π entre les liaisons $C_{10}-C_5$, $C_{10}-C_9$ du benzène et $C_{10}-N_{17}$ dans le composé $6MRC_{18}$.

Dans le but de mieux vérifier cette localisation, nous avons examiné dans ce travail la forme des Orbitales Moléculaires Frontières LUMO, HOMO et HOMO-1 du $6MRC_{18}$. Celles-ci montre que le doublet non liant de N_{17} et les liaisons ($C_{10}-C_5$ et $C_{10}-C_9$) ne peuvent pas constituer des OM délocalisées à cause de leurs orientations dans l'espace. En effet, ces OM sont de symétrie π localisées respectivement sur les atomes N_{17} , C_{10} , C_5 et C_9 avec l'apparition d'un point nodal sur la liaison $N_{17}-C_{10}$ (**Figure 6**) [28].

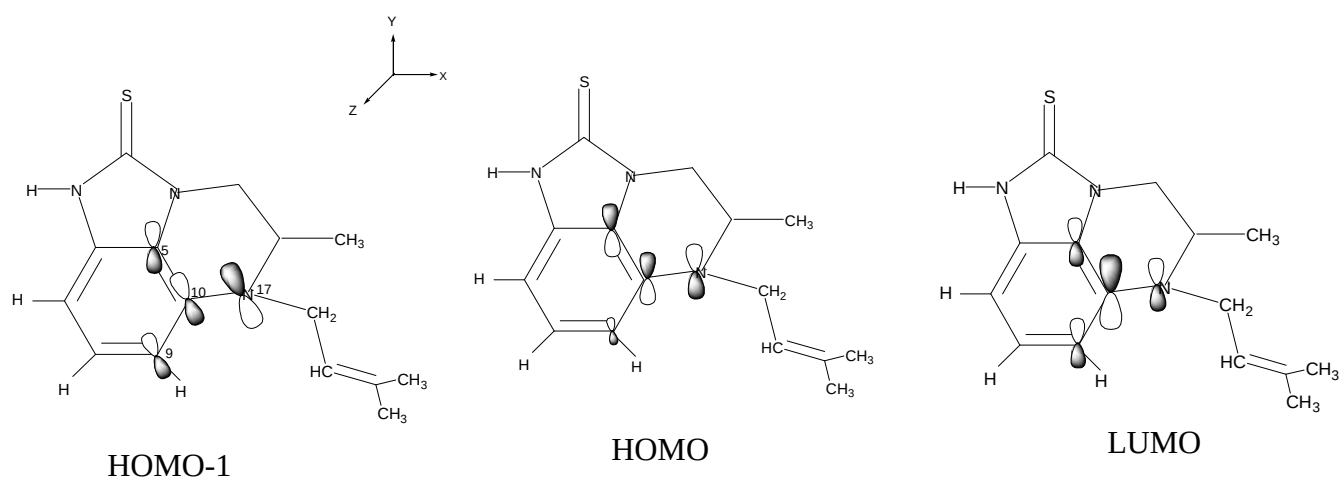


Figure 6: la plus grande contribution des OA sur les différentes OM au niveau HF/6-31G*

Les angles $N_4C_5C_{10}N_{17}$ et $C_{20}N_{17}C_{10}C_5$ sont portés dans le même plan dans la structure 6MRC₁₈. Pour ce composé l'azote N_{17} et le carbone C_{20} (le groupement diméthylallyl) sont dans le même plan portant les atomes $C_{10}C_5N_4C_{15}$. Pour cette planéité de ce système, nous remarquons qu'il y a une grande différence entre la contraction en opposition C_{18} et C_{15} . En effet, l'azote N_{17} dans le TIBO est en dehors du plan, l'angle dièdre $C_5C_{10}C_{18}N_{17}$ est de l'ordre de 40.94, 33.21 et 35.67 degrés respectivement au niveau PM3, HF et B3LYP/6-31G*. Cette géométrie est conservée dans la structure 6MRC₁₅, l'angle dièdre $C_5C_{10}C_{18}N_{17}$ est de l'ordre de 23.94, 22.46 et 23.30 degrés respectivement au niveau PM3, HF et B3LYP/6-31G*. On peut conclure que la position de l'azote N_{17} dans 6MRC₁₈ diffère de sa position dans TIBO et 6MRC₁₅.

Par conséquent la position de l'atome C_{20} , du groupement diméthylallyl, lié à N_{17} sera différente. En effet, l'angle dièdre $C_{20}N_{17}C_{10}C_5$ est de l'ordre de -161.07, -175.52 et -175.55 degrés pour la structure 6MRC₁₈. Dans le cas de TIBO et 6MRC₁₅, l'angle dièdre portant les atomes $C_{20}N_{17}C_{18}C_{10}$ est l'ordre 72.86 et 76.4 degrés respectivement pour ces deux structures au niveau B3LYP/6-31G*. Donc contrairement aux TIBO et 6MRC₁₅, le benzène l'imidazolidone et le groupement diméthylallyl sont presque dans le même plan dans la structure 6MRC₁₈.

Par ailleurs, nous remarquons que la contraction en C_{15} conduit à la même alternance des atomes C_{10} , C_{18} et N_{17} par rapport au TIBO contrairement à 6MRC₁₈.

En conclusion de cette partie, nous remarquons qu'il y a une grande similitude entre la géométrie 3D de 6MRC₁₅ et TIBO (**Figure 7**).

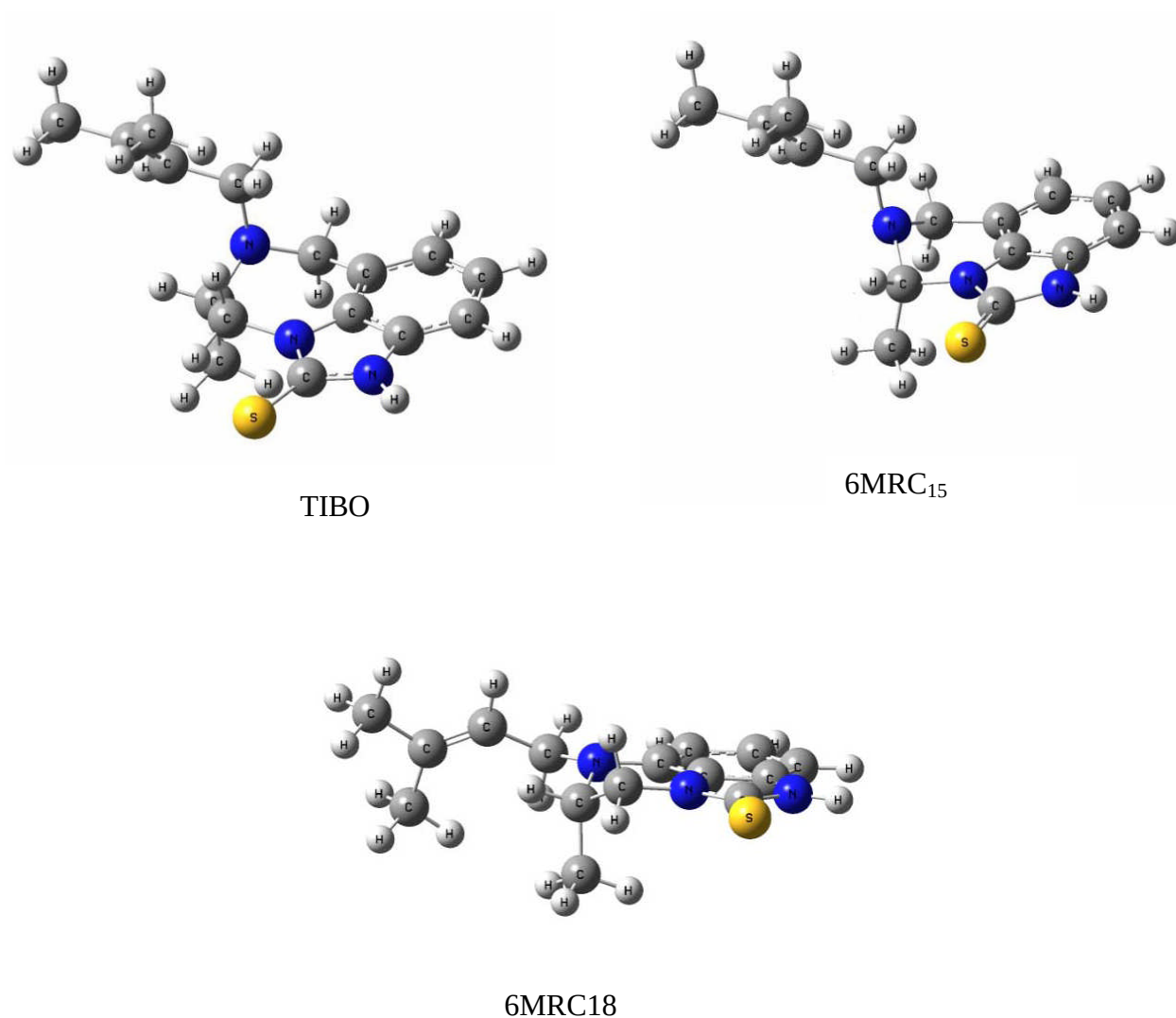


Figure 7 : La structure géométrique des TIBO, 6MRC₁₅ et 6MRC₁₈

Tableau 1 :

Comparaison entre les distances (Å), angles de valences (degrés) et angles dièdres (degrés) des TIBO, 6MRC₁₅ et 6MRC₁₈ aux différents niveaux théoriques PM3, HF/6-31G* et B3LYP/6-31G*.

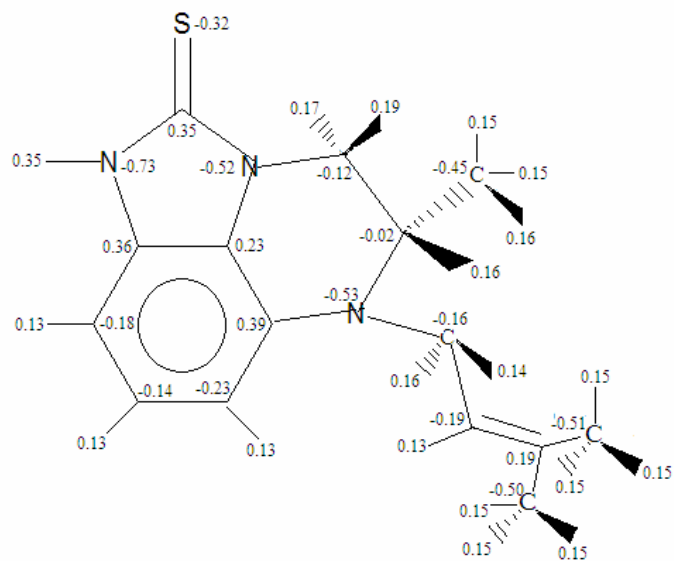
Method	PM3			HF/6-31G*			B3LYP/6-31G*		
Geometrical Parameters	MRC15	TIBO	MRC18	MRC15	TIBO	MRC18	MRC15	TIBO	MRC18
Distances (Å)									
N ₁ -C ₂	1.42	1.40	1.42	1.36	1.34	1.36	1.38	1.37	1.38
C ₂ -S ₃	1.64	1.64	1.64	1.67	1.68	1.67	1.67	1.67	1.67
C ₂ -N ₄	1.41	1.42	1.40	1.34	1.35	1.34	1.37	1.38	1.37
N ₄ -C ₅	1.41	1.43	1.41	1.37	1.40	1.37	1.38	1.40	1.38
C ₅ -C ₆	1.40	1.43	1.40	1.37	1.39	1.37	1.39	1.41	1.39
N ₁ -C ₆	1.42	1.42	1.42	1.38	1.38	1.38	1.39	1.38	1.39
C ₆ -C ₇	1.39	1.39	1.39	1.38	1.38	1.38	1.39	1.39	1.39
C ₇ -C ₈	1.39	1.38	1.39	1.39	1.38	1.38	1.40	1.39	1.40
C ₈ -C ₉	1.40	1.39	1.40	1.39	1.39	1.40	1.40	1.40	1.41
C ₉ -C ₁₀	1.38	1.39	1.39	1.38	1.39	1.39	1.39	1.40	1.40
C ₅ -C ₁₀	1.39	1.39	1.39	1.37	1.39	1.38	1.38	1.40	1.39
N ₁ -H ₁₁	0.99	0.99	0.99	0.99	0.99	0.99	1.01	1.01	1.01
C ₇ -H ₁₂	1.09	1.09	1.09	1.07	1.07	1.07	1.08	1.08	1.08
C ₈ -H ₁₃	1.09	1.09	1.09	1.07	1.07	1.07	1.08	1.08	1.08
C ₉ -H ₁₄	1.09	1.09	1.10	1.07	1.07	1.07	1.08	1.08	1.08
N ₄ -C ₁₅	-	1.47	1.47	-	1.46	1.44	-	1.46	1.45
N ₄ -C ₁₆	1.48	-	-	1.47	-	-	1.46	-	-
C ₁₅ -C ₁₆	-	1.53	1.54	-	1.53	1.53	-	1.54	1.55
C ₁₆ -N ₁₇	1.50	1.49	1.51	1.45	1.46	1.47	1.47	1.47	1.48
N ₁₇ -C ₁₈	1.50	1.48	-	1.46	1.44	-	1.47	1.52	-
C ₁₈ -C ₁₀	1.49	1.49	-	1.51	1.52	-	1.51	1.52	-
N ₁₇ -C ₁₀	-	-	1.43	-	-	1.39	-	-	1.40
C ₁₆ -C ₁₉	1.52	1.52	1.52	1.52	1.52	1.53	1.53	1.53	1.53
N ₁₇ -C ₂₀	1.49	1.49	1.49	1.45	1.46	1.45	1.47	1.47	1.47
C ₂₀ -C ₂₁	1.49	1.49	1.49	1.50	1.50	1.50	1.50	1.50	1.50
C ₂₁ -C ₂₂	1.34	1.34	1.34	1.32	1.32	1.32	1.34	1.34	1.34
C ₂₂ -C ₂₃	1.48	1.49	1.48	1.51	1.51	1.51	1.51	1.51	1.51
C ₂₂ -C ₂₄	1.49	1.49	1.48	1.51	1.51	1.51	1.51	1.51	1.51
Angles de Valence (deg)									
N ₁ C ₆ C ₅	105.57	107.19	105.49	104.51	106.03	104.45	104.66	111.66	-
C ₂ C ₄ C ₅	109.83	109.50	109.87	109.87	109.47	109.85	110.36	109.90	110.3
N ₁ C ₂ N ₄	104.7	105.50	104.60	106.01	106.86	105.75	104.80	105.64	104.5
N ₄ C ₅ C ₆	108.79	107.06	108.83	108.38	106.40	108.54	108.50	106.67	108.75
C ₇ C ₆ C ₅	119.16	122.11	119.44	119.58	122.88	120.15	119.30	122.52	119.88
C ₆ C ₅ C ₁₀	124.30	120.54	124.10	125.23	120.80	125.63	125.13	120.79	125.5
C ₁₀ C ₉ C ₈	120.40	122.20	120.35	120.52	122.81	120.45	120.61	122.85	120.51
C ₇ C ₈ C ₉	123.08	121.40	123.24	122.62	120.82	123.65	122.67	120.87	123.62
N ₄ C ₅ C ₁₀	126.9	132.3	127.07	126.37	132.74	125.81	126.36	132.48	125.75
C ₅ N ₄ C ₁₅	-	124.5	117.80	-	127.48	118.73	-	127.29	118.94
C ₅ N ₄ C ₁₆	118.72	101.11	85.45	119.39	-	-	119.51	-	-
N ₄ C ₁₅ C ₁₆	-	115.90	-162.71	-	115.70	108.53	-	116.02	108.75
C ₁₅ C ₁₆ N ₁₇	-	113.99	113.42	-	115.59	110.50	-	115.59	110.55
N ₄ C ₁₆ N ₁₇	106.65	-	-	106.32	-	-	106.36	-	-
C ₁₆ N ₁₇ C ₁₈	116.58	115.86	-114.04	116.77	116.75	-	116.05	116.27	-
N ₁₇ C ₁₈ C ₁₀	110.15	118.46	-	109.50	118.60	-	109.70	119.33	-
C ₁₆ N ₁₇ C ₁₀	-	-	115.16	-	-	114.99	-	-	115.46
N ₁₇ C ₂₀ C ₂₁	110.63	110.40	-149.25	111.05	111.62	111.68	111.12	111.68	111.88
C ₂₀ C ₂₁ C ₂₂	123.75	123.99	123.77	128.77	128.85	128.63	128.55	128.66	128.33
C ₂₁ C ₂₂ C ₂₃	123.72	123.711	121.05	120.54	125.71	120.52	125.22	125.21	120.24
C ₂₁ C ₂₂ C ₂₄	121.02	121.07	123.67	125.69	120.56	125.71	120.43	120.48	125.22
C ₂₄ C ₂₂ C ₂₃	115.25	115.21	115.27	113.76	113.71	113.76	114.33	114.29	114.34
Angles dièdres (deg)									
N ₁ C ₂ N ₄ C ₅	-1.02	-3.39	-2.08	-0.10	-1.26	-0.65	0.20	-1.12	-0.51
C ₂ C ₄ C ₅ C ₆	0.34	2.19	1.418	0.24	0.52	0.66	0.02	0.40	0.36
C ₅ C ₆ C ₇ C ₈	0.42	0.13	0.12	0.15	0.03	-0.33	0.15	0.15	-0.22
C ₇ C ₈ C ₉ C ₁₀	-0.61	0.72	0.02	-0.35	0.16	0.83	-0.36	0.17	0.89
S ₃ C ₂ N ₄ C ₅	179.65	179.57	179.88	179.77	179.98	179.60	179.70	179.78	179.88
C ₁₀ C ₅ N ₄ C ₁₅	179.47	177.98	178.95	179.95	177.79	179.36	179.83	177.98	179.50
C ₁₀ C ₅ N ₄ C ₁₆	-	-28.67	-5.49	-	-17.31	-3.82	-	-18.46	-3.91
C ₅ N ₄ C ₁₅ C ₁₆	2.98	-	-	3.66	-	-	3.25	-	-
C ₁₈ C ₁₀ C ₅ N ₄	-	50.6	25.95	-	39.10	29.21	-	41.41	29.43
N ₄ C ₁₅ C ₁₆ N ₁₇	-0.08	-5.11	-	1.02	-5.74	-	0.93	-7.27	-
C ₅ C ₁₀ C ₁₈ N ₁₇	-	-66.97	-46.51	-	-64.40	-53.32	-	-64.97	-52.72
C ₁₉ C ₁₆ N ₁₇ C ₂₀	23.94	40.94	-	22.46	33.21	-	23.30	35.67	-
N ₁₇ C ₂₀ C ₂₁ C ₂₂	64.5	174.05	56.49	69.55	172.00	76.06	67.93	171.59	76.34
	140.8	120.40	-118.77	131.25	135.71	-130.54	131.52	134.89	-129.43

1.1.2 Charges:

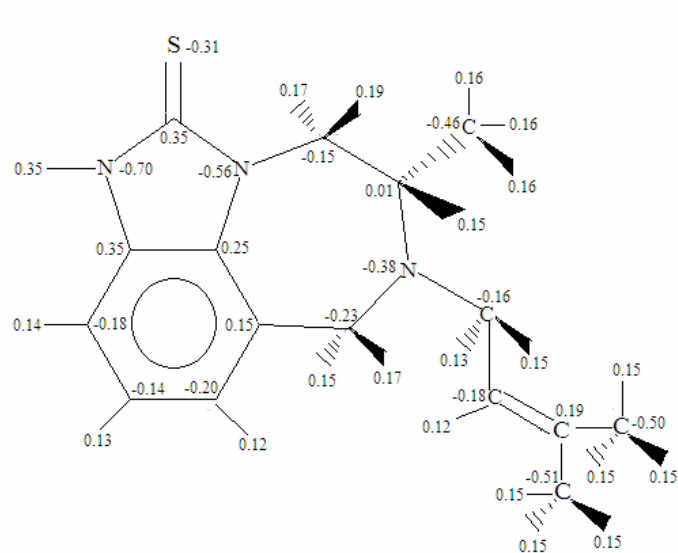
Les calculs des charges atomiques nets de Mulliken aux différents niveaux théoriques montrent que la contraction du cycle à 7 dans les TIBO en position C₁₅ n'influe pas sur la répartition des charges dans le benzène et dans cycle imidazolidone. Par contre dans le cas de la contraction en position C₁₈, nous constatons que le carbone C₁₀ du benzène devient plus électrophile, il passe d'une charge de 0.15 dans les TIBO à 0.39 dans 6MRC₁₈ au niveau B3LYP/6-31G* (**Figure 8**). En effet, ce carbone ne garde pas le même type de liaison, il passe d'une liaison de type C-C dans les TIBO à une liaison C-N dans 6MRC₁₈.

Quant au cycle à 7, nous remarquons que le carbone C₁₆ garde la même nature de charge dans le 6MRC₁₅ et il devient chargé légèrement négative dans le 6MRC₁₈. Dans les deux cas, l'azote N₁₇ devient plus nucléophile.

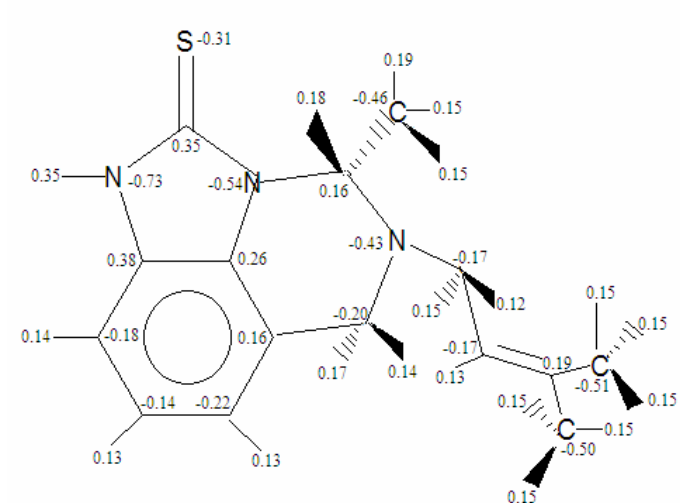
Nous remarquons aussi que les charges du groupement diméthylallyl n'ont pas été influencées par la contraction dans les deux positions.



MRC18



TIBO



MRC15

Figure 8: Les charges atomiques de Mulliken des TIBO, MRC15 et MRC18 au niveau B3LYP/6-31G* .

1.1.3 Les propriétés électroniques et physico-chimiques.

Dans cette partie quelques propriétés électroniques et physico-chimiques ont été calculées afin de montrer comment la contraction du cycle à 7-sommets dans le TIBO influence sur les propriétés moléculaires. L'analyse des différents résultats trouvés montre :

- Le Chemical Hardness (η) et le Chemical potential (μ) ont été calculés aux différents niveaux théoriques. L'analyse des résultats du **tableau 2** montre que le TIBO et 6MRC₁₅ ont les mêmes valeurs du chemical Hardness et du chemical potential comparés au 6MRC₁₈. Dans tous les cas, nous remarquons que MRC₁₈ possède un chemical hardness légèrement plus grand comparé au TIBO et 6MRC₁₅.
- Le calcul de l'indice global d'électrophilicity montre que 6MRC₁₈ possède un ω très petit comparé aux TIBO et 6MRC₁₅ qui ont des indices d'électrophilicity voisine. En effet, $\omega(\text{TIBO}) > \omega(6\text{MRC}_{15}) \gg \omega(6\text{MRC}_{18})$.
- Le calcul des énergies des Orbitales Moléculaires Frontières montre que le TIBO et 6MRC₁₅ ont les mêmes valeurs de la HOMO et de la LUMO (**Figure 9**).
- L'interaction des INNTI avec le site allostérique de la RT se fait selon le mode clé-serrur. Les molécules se sélectionnent en premier par leur forme pour rentrer à l'intérieur de la cavité du site récepteur, dans ce cas le calcul du volume de ces molécules dans le programme Gaussian en utilisant le mot clef (volum=tight) montre que 6MRC₁₅ possède un volume moyen par rapport aux deux autres molécules :

$$V(\text{TIBO}) > V(6\text{MRC}_{15}) > V(6\text{MRC}_{18}).$$

Nous remarquons aussi que 6MRC₁₅ possède une structure plus ovale comparé au TIBO et 6MRC₁₈.

Donc nous pouvons constater que quand on contracte le cycle à 7 en position C₁₅ le volume de la molécule diminue de 1.67% et sa structure devient plus ovale. Il est connu que les composés ayant de faibles tailles et qui peuvent établir les meilleures interactions ont des chances d'être plus active que les composés de grandes tailles [29]. En effet, les grosses molécules peuvent avoir des difficultés d'accès au site récepteur contrairement aux petites molécules qui pénètre plus aisément à l'intérieur de la cavité du site allostérique des INNTI qui est sous forme de Ω [30].

L'analyse de tous les résultats trouvés aux différents niveaux théorique montre qu'il y a une grande similitude de la géométrie en 3D et les propriétés moléculaires des TIBO et 6MRC₁₅.

Tableau 2 : Propriétés électronique et physicochimique des TIBO, 6MRC₁₅ et 6MRC₁₈.

DFT/6-31G*	X=H			X=Cl (in 9 position)		
	C ₁₅ atom	TIBO	C ₁₈ atom	C ₁₅ atom	TIBO	C ₁₈ atom
HOMO	-0.0919	-0.093	-0.082	-0.0991	-0.10	-0.091
LUMO	0.0797	0.079	0.093	0.0659	0.066	0.081
(HOMO-LUMO)	-0.1716	-0.172	-0.175	-0.165	-0.166	-0.172
Molar Refractivity	86.69	92.28	88.16	91.49	97.085	92.971
Ovality	1.491	1.465	1.471	1.503	1.474	1.470
V	588.67	598.92	552.10	592.08	674.00	568.68
LogP	3.62	3.53	4.00	4.178	4.24	4.566
η	0.0858	0.086	0.087	0.0825	0.083	0.086
μ	0.0061	0.007	0.005	0.0166	0.017	0.005
ω	2.168 e-04	2.841 e-04	1.728 e-04	1.67 e-03	1.74 e-03	0.14 e-03

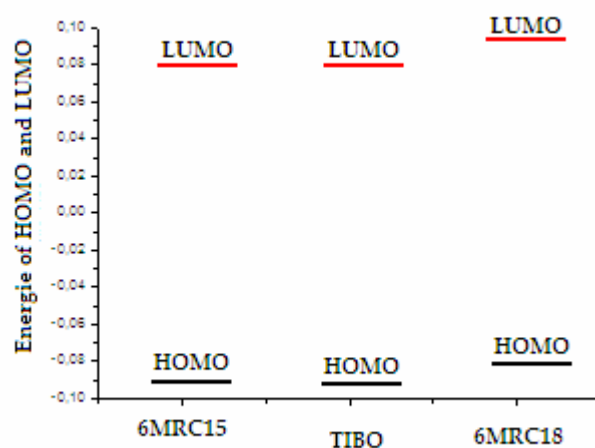


Figure 9 : Différence d'énergie entre la HOMO et la LUMO des TIBO, 6MRC₁₅ et 6MRC₁₈ au niveau HF/ 6-31G*.

1.1.4 La configuration Butterfly –like:

Bien que les INNTI aient des structures assez divergentes, ces molécules ont une conformation similaire et peuvent se superposer l'une à l'autre. Cette conformation peut être comparée à un papillon qu'on appelle aussi « configuration Butterfly-like ». Ce modèle de conformation a été élaboré, pour la première fois, par Schäfer et coll. en 1993 (**Figure 10**) [31]. Il a été construit grâce à la superposition des structures cristallographiques et des potentiels électrostatiques moléculaires de trois composés : la Névirapine, le TIBO et une thiazolidine. Comme le montre la **Figure 10**, ce modèle est principalement composé d'un cycle aromatique et d'un système π délocalisé, distants de 4.5 à 5 Å et formant entre eux un angle de 108 à 115°. Il comprend également un groupe de contrainte stérique, un carbonyle associé au système π délocalisé et d'un groupe méthyle placé sur le cycle aromatique. Ce modèle reste l'un des plus cités lors de la conception de nouveaux INNTI et a contribué à l'identification et au développement de plusieurs séries de composés. En utilisant cette méthode, Mager [32] montre que la plupart des inhibiteurs non nucléosidiques de la TI sont capables d'adopter une conformation compatible avec le Butterfly-like au sein du site allostérique de la TI.

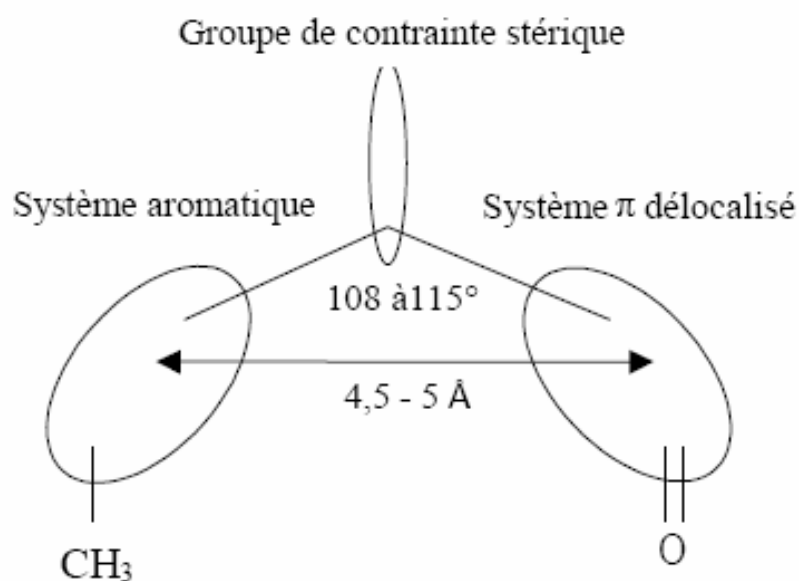


Figure 10 : La configuration butterfly-like tel que proposée par Schäfer et coll. [31].

La critique de la configuration Butterfly –like permet de dégager plusieurs informations importantes pour la sélection des composés résultant de la contraction du cycle à 7-sommets dans le TIBO.

Comme tout autre INNTI, les TIBO sont localisés dans le site allostérique de la RT en adoptant la conformation butterfly –like [33], comme le montre la **Figure 11**. Le plan portant le groupement diméthylallyl de TIBO c'est l'aile I [34], il interagit avec les amino acid Pro95, Tyr181, Tyr188, Gly190 et Trp229 du P66, et le plan portant le benzodiazepinone est l'aile II, il interagit avec Lys101, Lys103, Val106, Phe227, His235, Pro236 et Tyr318 du P66. Les amino-acides Leu100 et Leu234 du P66 interagissent avec les deux ailes I et II [33, 34]. Notons que la géométrie butterfly-like dépend de la conformation du cycle à 7-sommets dans TIBO [13, 35-36] et son degrés d'ouverture est en fonction de la taille des molécules, de sa polarisabilité et de sa lipophilicité [36].

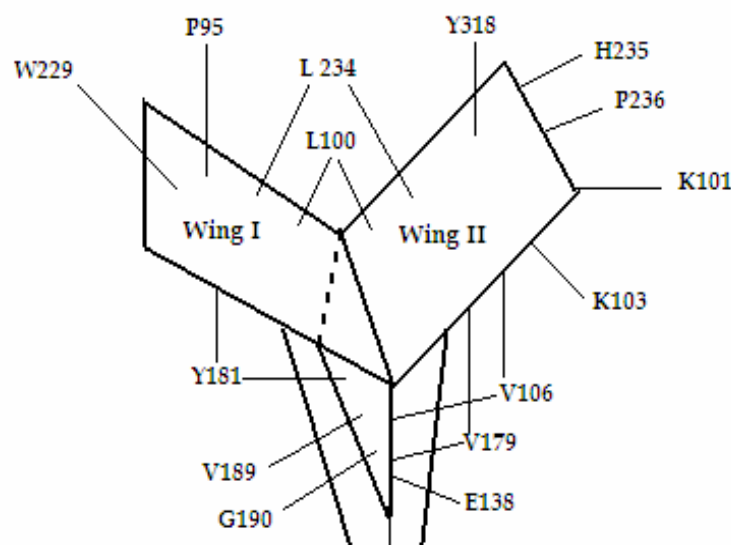


Figure 11 : La représentation Schématique de la configuration butterfly-like dans INNTI [33].

Une analyse conformationnelle des trois structures a été réalisée dans la première partie de ce travail, afin de prendre en compte leur flexibilité. Cette recherche vise à vérifier si ces composés sont en mesure d'adopter une conformation butterfly-like.

Nous pouvons conclure de l'étude géométrique précédente que les optimisations géométriques de TIBO aux différents niveaux théoriques conduisent dans tous les cas à une structure butterfly –like conformément à la littérature [13, 33-36]. En effet, l'angle séparant les deux plans portant l'aile I et II est de l'ordre de 111.43, 110.22 et 110.45 degrés respectivement au niveau PM3, HF/6-31G* et DFT/B3LYP/6-31G* (**Figure 12**).

La contraction du cycle à 7 sommets en position 15 conduit aussi à une géométrie butterfly –like. Dans ce cas, l'angle séparant les deux plans portant l'aile I et II est de l'ordre de 113.3, 114.5 et 113.0 degrés respectivement au niveau PM3, HF/6-31G* et DFT/B3LYP/6-31G* (**Figure 11**). Nous remarquons que l'angle trouvé pour cette molécule est proche de celui obtenue pour TIBO. Par ailleurs, nous remarquons que cet angle reste comparable par rapport à angle trouver dans les INNTI comme lanevirapine (108 – 115) degrés [37]

Par contre la contraction de TIBO en position 18 conduit à une légère planéité des deux ailes. En effet, nous remarquons que la géométrie butterfly –like n'a pas été conservée pour cette structure. L'angle séparant les deux plans portant l'aile I et II devient de l'ordre de 160.8, 169.22 et 157.04 degrés respectivement aux niveaux PM3, HF et B3LYP/6-31G* (**Figure 12**). Le **tableau 3** résume tous les angles α trouvés pour toutes les structures.

Les différentes méthodes théoriques conduisent à conclure la TIBO et 6MRC₁₅ possèdent des conformations communes, qui sont compatibles avec la configuration butterfly-like.

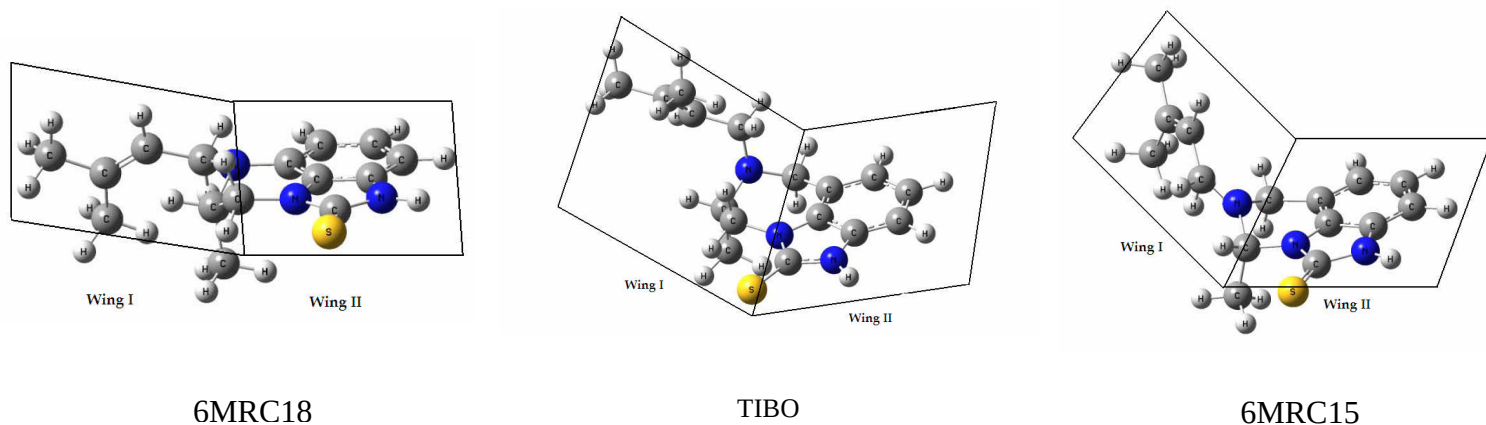


Figure 12: la structure butterfly –like adoptée par TIBO, 6MRC15 et 6MRC18 au niveau B3LYP/6-31G*.

Table 3: l'angle α séparant les deux plans portant l'aile I et II aux différents niveaux théoriques (en degrés).

Method	α					
	X=H			X=Cl (in 9 position)		
	C15 atom	TIBO	C18 atom	C15 atom	TIBO	C18 atom
PM3	113.3	111.43	160.8,	115.19	112.47	164.80
HF/6-31G*	114.5	110.22	169.22	115.50	114.26	169.87
B3LYP/6-31G*	113.0	110.45	157.04	111.9	113.41	163.09

1.2 Effet de l'extension du cycle à 7-sommets dans les TIBO.

Dans la second partie de ce travail, on va examine l'effet de l'extension du cycle à 7 sommets dans les TIBO à 8 sommets. Dans ce cas 2 types de molécules à 8MR peuvent se produire. L'insertion de l'atome de carbone (C_{19}) entre N_4 , C_{15} conduit au compose 8MR(N_4C_{15}) et son insertion entre C_{10} , C_{18} conduit au 8MR($C_{10}C_{18}$), comme il est montré sur la **figure 4**. Notons que l'insertion de l'atome C_{19} entre les atomes C_{15} et C_{16} ou entre C_{18} et N_{17} conduit au même composés que précédemment.

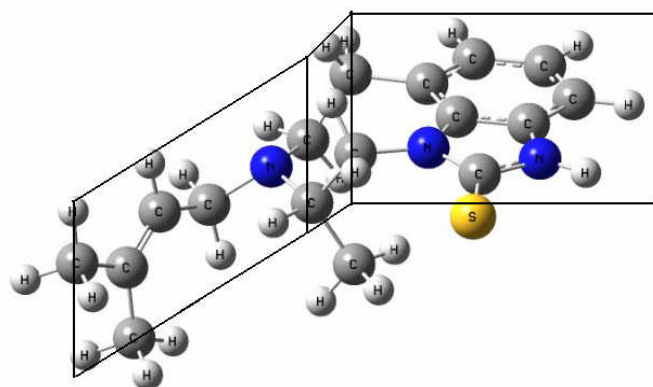
La numérotation des différents atomes du TIBO, 8MR (N_4C_{15}) et 8MR ($C_{10}C_{18}$) adoptée dans cette étude est représentée sur la **Figure 4**. Les principaux paramètres géométriques obtenus après une optimisation complète des différentes structures au niveau B3LYP/6-31G* sont portés sur le **tableau 4**.

1.2.1 La configuration Butterfly –like :

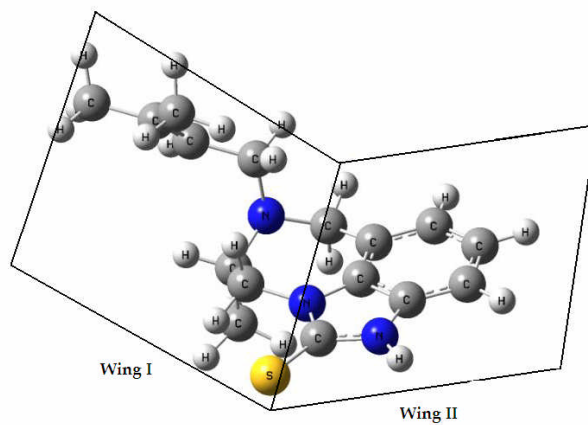
L'extension du cycle à 7 sommets entre les atomes N_4 , C_{15} conduit à une géométrie butterfly –like. L'angle séparant les deux plans portant l'aile I et II est de l'ordre de 108.55 degrés au niveau DFT/B3LYP/6-31G* (**Figure 12**). Nous remarquons que cet angle est proche de 110.45 degrés, angle obtenue pour TIBO au même niveau. Par ailleurs, cet angle reste comparable par rapport à angle trouver dans les INNTI comme la nevirapine (108 – 115) degrés [37]

Par contre l'extension de TIBO entre les atomes C_{10} , C_{18} ne conduit pas à la géométrie butterfly –like. En effet, nous remarquons que ce composé est constitué de trois plans en forme de chaise comme la montre **Figure 12**.

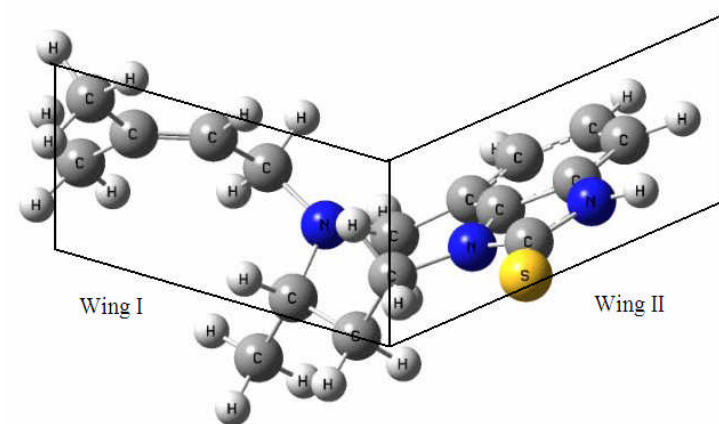
Ces différents résultats géométriques obtenus nous montrent qu'il y a une grande similarité entre la structure géométrique de 8MR (N_4C_{15}) et la TIBO.



8MR(C10-C18)



TIBO



8MR(N4-C15)

Figure 12: Les structures butterfly –like adoptées par les TIBO, 8MR(C10-C18) et 8MR(N4-C15) au niveau B3LYP/6-31G*.

Tableau 4:

Comparaison entre les distances (Å), angles de valences (degrés) et angles dièdres (degrés) des TIBO, 8MR(N₄C₁₅) et 8MR(C₁₀C₁₈) au niveau B3LYP/6-31G*.

Paramètres	B3LYP/6-31G*		
	8MR(C ₁₀ -C ₁₈)	TIBO	8MR(N ₄ -C ₁₅)
Distances (Å)			
N ₁ -C ₂	1.37	1.37	1.37
C ₂ -S ₃	1.67	1.67	1.67
C ₂ -N ₄	1.38	1.38	1.38
N ₄ -C ₅	1.40	1.40	1.41
C ₅ -C ₆	1.41	1.41	1.41
N ₁ -C ₆	1.38	1.38	1.38
C ₆ -C ₇	1.39	1.39	1.39
C ₇ -C ₈	1.39	1.39	1.39
C ₈ -C ₉	1.40	1.40	1.39
C ₉ -C ₁₀	1.40	1.40	1.40
C ₅ -C ₁₀	1.40	1.40	1.40
N ₁ -H ₁₁	1.01	1.01	1.01
C ₇ -H ₁₂	1.08	1.08	1.08
C ₈ -H ₁₃	1.08	1.08	1.08
C ₉ -H ₁₄	1.08	1.08	1.08
N ₄ -C ₁₅	1.46	1.46	-
N ₄ -C ₁₉	-	-	1.47
C ₁₅ -C ₁₉	-	-	1.53
C ₁₀ -C ₁₉	1.51	-	-
C ₁₈ -C ₁₉	1.55	-	-
C ₁₅ -C ₁₆	1.55	1.54	1.54
C ₁₆ -N ₁₇	1.47	1.47	1.47
N ₁₇ -C ₁₈	1.46	1.52	1.46
C ₁₈ -C ₁₀	-	1.52	1.51
C ₁₆ -C ₂₀	1.54	1.53	1.54
N ₁₇ -C ₂₁	1.47	1.47	1.47
C ₂₁ -C ₂₂	1.50	1.50	1.50
C ₂₂ -C ₂₃	1.34	1.34	1.34
C ₂₃ -C ₂₄	1.51	1.51	1.51
C ₂₃ -C ₂₅	1.51	1.51	1.51
Angles de valence (deg)			
N ₁ C ₆ C ₅	106.07	111.66	106.50
C ₂ C ₄ C ₅	110.10	109.90	109.78
N ₁ C ₂ N ₄	105.50	105.64	105.83
N ₄ C ₅ C ₆	106.64	106.67	106.17
C ₇ C ₆ C ₅	122.60	122.52	123.53
C ₆ C ₅ C ₁₀	121.21	120.79	119.40
C ₁₀ C ₉ C ₈	123.19	122.85	123.38
C ₇ C ₈ C ₉	120.92	120.87	120.33
N ₄ C ₅ C ₁₀	132.15	132.48	134.36
C ₅ N ₄ C ₁₅	127.56	127.29	-
N ₄ C ₁₅ C ₁₆	116.39	116.02	-
C ₁₅ C ₁₆ N ₁₇	112.48	115.59	108.87
N ₄ C ₁₉ C ₁₅	-	-	116.30
C ₁₆ N ₁₇ C ₁₈	117.36	116.27	114.87
N ₁₇ C ₁₈ C ₁₀	-	119.33	115.05
N ₁₇ C ₂₁ C ₂₂	111.86	111.68	112.14
C ₂₁ C ₂₂ C ₂₃	128.78	128.66	128.72
C ₂₂ C ₂₃ C ₂₄	120.50	125.21	120.51
C ₂₂ C ₂₃ C ₂₅	125.18	120.48	125.19
C ₂₅ C ₂₃ C ₂₄	114.31	114.29	114.29
Angle dièdre (deg)			
N ₁ C ₂ N ₄ C ₅	0.54	-1.12	-0.31
C ₂ C ₄ C ₅ C ₆	-0.45	0.40	0.00
C ₅ C ₆ C ₇ C ₈	0.00	0.15	0.00
C ₇ C ₈ C ₉ C ₁₀	0.00	0.17	0.31
C ₇ C ₆ C ₅ N ₄	180.00	179.78	179.55
S ₃ C ₂ N ₄ C ₅	-179.37	177.98	179.36
C ₁₀ C ₅ N ₄ C ₁₅	-0.45	-18.46	-
C ₅ N ₄ C ₁₅ C ₁₆	70.43	41.41	-
C ₁₈ C ₁₀ C ₅ N ₄	-	-7.27	-4.66
N ₄ C ₁₅ C ₁₆ N ₁₇	-100.50	-64.97	-
C ₅ C ₁₀ C ₁₈ N ₁₇	-	35.67	49.02
C ₂₀ C ₁₆ N ₁₇ C ₂₁	75.4	171.59	72.23
N ₁₇ C ₂₁ C ₂₂ C ₂₃	-132.58	134.89	-128.23
C ₅ N ₄ C ₁₉ C ₁₅	-	-	55.9
C ₁₀ C ₅ N ₄ C ₁₉	-	-	-13.7
C ₁₉ C ₁₀ C ₅ N ₄	-0.30	-	-
C ₁₈ C ₁₉ C ₁₀ C ₅	-67.5	-	-

1.2.2 Les propriétés électroniques et physico-chimiques

Dans le but de faire une comparaison entre les différentes molécules et de voir l'effet de l'extension du cycle à 7 sommets des TIBO en 8 sommets. Some electronic and physicochemical properties are calculated by application of DFT/B3LYP/6-31G* method.

Les différents résultats trouvés sont regroupés dans le **tableau 5**.

Le Chemical Hardness (η) et le Chemical potential (μ) ont été calculés pour les différents composés. Les trois structures ont des valeurs du chemical Hardness très proche. Par contre pour le Chemical Potentiel, les résultats trouvés montrent que les deux structures à 8 sommets ont des potentiel presque identique et elles sont supérieures à celle de TIBO.

Le calcul de l'indice global d'électrophilicity (ω) montre que les composés à 8 sommets ont des valeurs de (ω) très supérieures par rapport au TIBO. Donc comparés à ce dernier les deux composés 8MR(C₁₀-C₁₈), 8MR(N₄-C₁₅) peuvent être plus réactives [21].

$$\omega (8MR (C_{10}-C_{18})) > \omega (8MR (N_4-C_{15})) \gg \omega (TIBO).$$

Le résultat du calcul des énergies des Orbitales Moléculaires frontières montre la même chose que précédemment. En effet, nous remarquons que les deux structures à 8 sommets ont des énergies de la HOMO et de la LUMO très proches et différentes de celles de la TIBO. Mais la différence (HOMO-LUMO) reste presque constante pour les trois structures.

Le caractère Hydrophobe joue un rôle très important dans les processus de transport et de perméabilité [38]. Il convient de noter que les valeurs de LogP obtenus pour les deux structures 8MR(C₁₀-C₁₈), 8MR(N₄-C₁₅) sont dans l'intervalle [2.31 – 5.72] déterminé par Garg et coll. [38] pour la classe des INNTI. Pour les trois structures les valeurs de LogP varient comme suit:

$$\text{LogP} (8MR (C_{10}-C_{18})) > \text{LogP} (8MR (N_4-C_{15})) > \text{LogP} (TIBO).$$

Le volume des deux structures à 8 sommets a été aussi calculé dans cette partie et comparé à celui de TIBO. Les volumes trouvés pour les trois structures varient comme suit :

$$V (8MR (N_4-C_{15})) > V (8MR (C_{10}-C_{18})) > V (TIBO).$$

Nous remarquons que le composé 8MR (N₄-C₁₅) qui a une structure semblable au TIBO mais qui possède un plus grand volume avec une structure moins ovale, comparé au TIBO.

Son volume est de l'ordre de $711 \text{ \AA}^3$. Nous rappelons que le diamètre obtenu pour cette structure, qui nous a permis de calculer son volume, est de l'ordre de $a_0=5.49 \text{ \AA}$, ce diamètre reste inférieur à celui obtenu dans les TIBO quand on substitue l'hydrogène (H_{14}) porté par le carbone C_9 par le groupement ($-OC_2H_5$), cette dernière a un diamètre de l'ordre de $a_0=5.60 \text{ \AA}$ et son activité anti-VIH expérimentale est de 7.02 [38-42]. Ce résultat nous mène à conclure que 8MR (N_4-C_{15}) peut pénétrer dans la cavité du site allostérique des INNTI. A cet effet, les molécules résultantes de l'extension du cycle à 7 sommets entre les atomes N_4, C_{15} peuvent constituer de nouvelles classes de composés anti-VIH.

Table 5: Les propriétés électronique et physico-chimique des TIBO et des composés à 8MR.

B3LYP/6-31G*	8MR($C_{10}-C_{18}$)	TIBO	8MR(N_4-C_{15})
HOMO	-0.193	-0.093	-0.189
LUMO	-0.022	0.079	-0.021
(HOMO-LUMO)	-0.171	-0.172	-0.168
Molar Refractivity	97.04	92.28	97.15
Ovality	1.411	1.465	1.425
V	692	598.92	711
LogP	3.95	3.53	3.77
η	0.085	0.086	0.084
μ	0.107	0.007	0.105
ω	6.758 e-02	2.841 e-04	6.56 e-02

Nous notons que toutes les propriétés électroniques et physico-chimiques qui ont été calculés dans ce travail pourraient être des descripteurs susceptibles d'être corrélés à l'activité anti-VIH de TIBO. Afin de vérifier cette possibilité et éventuellement rechercher d'autre descripteurs, nous avons procédé à une étude RQSA en utilisant les méthodes statistiques MLR et RNA (décrites dans le chapitre II) à une série de dérivés TIBO.

Conclusion

Les recherches réalisées au cours de ce travail ont permis la proposition de nouvelles molécules qui pourraient être proposés comme de nouvelles séries d'inhibiteurs non nucléosidiques de la transcriptase inverse du VIH-1. Plusieurs des structures étudiées sont caractérisées par une excellente géométrie semblable à la molécule d'origine (TIBO) et devraient présenter de bonnes propriétés pharmacologiques.

REFERENCE.

- [1]- Miyasaka, T.; H. Tanaka; M. Baba; H. Hayakawa; R.T. Walker; J. Balzarini; E. De Clercq. *Journal of Medicinal Chemistry*, 1989. 32: p. 2507-2509.
- [2]- Pauwels, R.; K. Andries; J. Desmyter; D. Schols; M.J. Kukla; H.J. Breslin; A. Raeymaekers; J. Van Gelder; R. Woestenborghs; J. Heykants; K. Schellekens; M.A. Janssen; E. De Clercq; P.A.J. Janssen, *Nature*, 1990. 343: p. 470-474.
- [3]- Merluzzi, V.J.; K.D. Hargrave; M. Labadia; K. Grozinger; M.T. Skoog; J.C. Wu; C.K. Shih; K. Eckner; S. Hattox; J. Adams; A.S. Rosenthal; R. Faanes; R.J. Eckner; R.A. Koup; J.L. Sullivan. *Science*, 1990. 250: p. 1411-1413.
- [4]- Goldman, M.E.; J.H. Nunberg; J.A. O'Brien; J.C. Quintero; W.A. Schleif; K.F. Freund; S.L. Gaul; W.S. Saari; J.S. Wai. *Proceedings of the National Academy of Sciences of the United States of America*, 1991. 88(15): p. 6863-7.
- [5]- Romero, D.L.; M. Busso; C.K. Tan; F. Reusser; J.R. Palmer; S.M. Poppe; P.A. Aristoff; K.M. Downey; A.G. So; L. Resnick; W.G. Tarpley. *Proc Natl Acad Sci U S A.*, 1991. 88: p. 8806-8810.
- [6]- Schäfer, W.; W.-G. Friebe; H. Leinert; A. Mertens; T. Poll; W. von der Saal; H. Zilch; B. Nuber; M.L. Ziegler. *Journal of Medicinal Chemistry*, 1993. 36(726-732).
- [7]- Ren, J.; R.M. Esnouf; A.L. Hopkins; D.I. Stuart; D.K. Stammers. *J. Med. Chem.*, 1999. 42(19): p. 3845-51.
- [8]- Bader, J.P.; J.B. Mc Mahon; R.J. Schultz; V.L. Narayanan; J.B. Pierce; O.S. Weislow; M.C. F.; S.F. Stinson; M.R. Boyd. *Proc Natl Acad Sci U S A.*, 1991. 88: p. 6740-6744.
- [9]- De Clercq, E., *Antiviral Res.*, 1998. 38: p. 153-179.
- [10]- Mitsuya, H.; K.J. Weinhold; P.A. Furman; M. St Clair; S. Nusinoff-Lehrman; R.C. Gallo; D. Bolognesi; D.W. Barry; S. Broker. *Proc Natl Acad Sci U S A.*, 1985. 82: p. 7096-7100.
- [11]- Pauwels, R.; K. Andries; J. Desmyter; D. Schols; M.J. Kukla; H.J. Breslin; A. Raeymaekers; J. Van Gelder; R. Woestenborghs; J. Heykants; K. Schellekens; M.A. Janssen; E. De Clercq; P.A.J. Janssen. *Nature*, 1990. 343: p. 470-474.
- [12]- Merluzzi, V.J.; K.D. Hargrave; M. Labadia; K. Grozinger; M.T. Skoog; J.C. Wu; C.K. Shih; K. Eckner; S. Hattox; J. Adams; A.S. Rosenthal; R. Faanes; R.J. Eckner; R.A. Koup; J.L. Sullivan. *Science*, 1990. 250: p. 1411-1413.
- [13]- Bahram Hemmateenejad, S. Mohammad Hossein Tabaei and Fatemeh Namvaran. *J. of Molecular Structure*. 732 (2005) 39-45.
- [14]- B. Hemmateenejad, S. M. H. Tabaei, F. Namvaran. *Journal of the Iranian Chemical Society*, 4(4): 481-489 (2007).
- [15]- M. Dewar and W. Thiel, *J. Amer. Chem. Soc.* 99, 4499 (1977).
- [16]- Gaussian 98, Revision A.9, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, Cioslowski, Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Iashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1998 (2007).
- [17]- J. J. P. Stewart, *J. Comp. Chem.* 10, 209 (1989).
- [18]- C. C. J. Roothan, "New Developments in Molecular Orbital Theory," *Rev. Mod. Phys.* 23, 69 (1951).
- [19]- C. Lee, W. Yang and R. G. Parr. *Physical Review B* 37, 785 (1988).
- [20]- J. B. Collins, P. v. R. Schleyer, J. S. Binkley and J. A. Pople. *J. Chem. Phys.* 64: 5142

- (1976).
- [21]- R. G. Parr. J. Chem Phys 68: 3801 (1978).
- [22]- Chromatographic chiral separations. Edited by Morris Zief and Laura J. Crane. V(40).
- [23]- C. Peng, P. Y. Ayala, H. B. Schlegel and M. J. Frisch, "Using redundant internal coordinates to optimize geometries and transition states," J. Comp. Chem. 17, 49 (1996).
- [24]- Handbook of chemistry and physics. David R. Lide. Editor in Chief 73RD Edition 92-93.
- [25]- A. Peeters, C. Van Alsenoy, J. Dillen and H. J. Geise. J. Mo. Struct. 333 (1995) 99-110.
- [26]- I. K. Boessenkool and J. C. A. Boeyens, J. Cryst. Mol. Struct. 10 (1980) 11.
- [27]- Y. C. Liaw, Y. G. Gao, H. Robinson and A. H. J. Wang. J. Am. Chem. Soc. 113 (1991) 1857.
- [28]- Y. Belmiloud, D. Kheffache et M. Brahimi. J. of Molecular Structure. 730: 48-53 (2006).
- [29]- Latifa Douali, Didier Villemin and Driss Cherqaoui. Int. J. Mol. Sci. 5(2004) 548-55.
- [30]- Mager, P. P. Drug Des. Discovery. 14 (1996) 225-239.
- [31]- Schäfer, W.; W.-G. Friebe; H. Leinert; A. Mertens; T. Poll; W. von der Saal; H. Zilch; B. Nuber; M.L. Ziegler. Journal of Medicinal Chemistry, 1993. 36(726-732).
- [32]- Mager, P.P. Drug Design and Discovery, 1996. 14: p. 241-257.
- [33]- Ding, J., Das, K., Moereels, H., Koymans, L., Andries, K., Janssen, P. A. J., Hughes, S. H., Arnold, E., Nature Struct. Biol. 2 (1995a) 407-415.
- [34]- Kalyan Das, J. Mol. Biol. 265 (1996) 1085-1100.
- [35]- M. B. K. Smith, M. L. Lamb, J. Tirado-Rives, W. Jorgensen, C. J. Michejda, S. K. Ruby, R. Smith Jr., Protein Eng. 13 (2000) 413.
- [36]- M. LL. Barreca, A. Carotti, A. Carrieri, A. Chimirri, A. M. Monforte, M. P. Calace, A. Rao, Bioorg. Med. Chem. 7 (1999) 2283.
- [37]- Schafer, W.; Friebe, W.-G.; Leinert, H.; Mertens, A.; Poll, T.; von der Saal, W.; Zilch, H.; Nuber, B.; Zeigler, M. L. J. Med. Chem. 36(1993) 726.
- [38]- Rajni Garg, Satya P. Gupta, Hua Gao, Mekapati Suresh Babu, Asim Kumar Debnath, and Corwin Hansch. Chem. Rev. 99 (1999) 3525-3601.
- [39]- Kukla, M. J., Breslin, H. J., Pauwels, R., Fedde, C. L., Miranda, M., Scott, M. K., Sherrill, R. G., Raeymackers, A., Van Gelder, J., Andries, K., Janssen, M. A. C., De Clercq, E., Janssen, P. A. J. J. Med. Chem. 1991, 34, 746.
- [40]- Kukla, M. J., Breslin, H. J., Diamond, C. J., Grous, P. P., Ho, C. Y., Miranda, M., Rodgers, J. D., Sherril, R. G., De Clercq, E., Pauwels, R., Andries, K., Moens, L. J., Janssen, M. A. C., Janssen, P. A. J. J. Med. Chem. 1991, 34, 3187.
- [41]- Breslin, H. J., Kukla, M. J., Ludovici, D. W., Mohrbacher, R., Ho, W., Miranda, M., Rodgers, J. D., Hitchens, T. K., Leo, G., Gauthier, D. A., Ho, C. Y., Scott, M. K., De Clercq, E., Pauwels, R., Andries, K., Janssen, M. A. C., Janssen, P. A. J. J. Med. Chem. 1995, 38, 771.
- [42]- Ho, W., Kukla, M. J., Breslin, H. J., Ludovici, D. W., Grous, P. P., Diamond, C. J., Miranda, M., Rodgers, J. D., Ho, C. Y., De Clercq, E., Pauwels, R., Andries, K., Janssen, M. A. C., Janssen, P. A. J. J. Med. Chem. 1995, 38, 794. (b) Cramer, R. D., III, Bunce, J. D., Patterson, D. E., Frank, I. E. Quant. Struct.-Act. Relat. 1988, 7, 18.

Partie 2: Etude des RQSA des dérivés TIBO par les RNA

Introduction:

Le développement des dérivés TIBO continue de susciter encore l'intérêt de plusieurs chercheurs [1-3]. Les études RQSA ont un rôle important dans le développement de ces composés. Les études consacrées aux dérivés TIBO sont peu nombreuses comparées aux autres travaux RQSA sur des familles de composés inhibitrice de la transcriptase Inverse. Dans la littérature un certain nombre de techniques computationnelles a été jugé utile pour l'établissement des RQSA des dérivés TIBO, telles que la RML [1] et les RNA [2-3]. L'étude par les RNA qui a été faite dans la littérature sur ces dérivés s'est basée seulement sur les descripteurs proposés par MLR. L'objectif de ce travail est de rechercher d'autres descripteurs susceptibles de décrire l'activité de ces composés en appliquant la régression linéaire et non linéaire.

Les Modèles de la littérature :

- Le premier Modèle de la littérature établie par la MLR

Le premier modèle MLR a été proposé par Garg et coll. [1]. Ces auteurs ont essayé de corréler l'activité inhibitrice (IC50) des dérivés TIBO, synthétisés par le groupe de Kukla et coll. [4-6] à leur structure chimique. Ils ont réalisé l'étude sur une base de données composée de 89 dérivés TIBO. La structure générale de ces derniers est représentée dans la Figure 1.

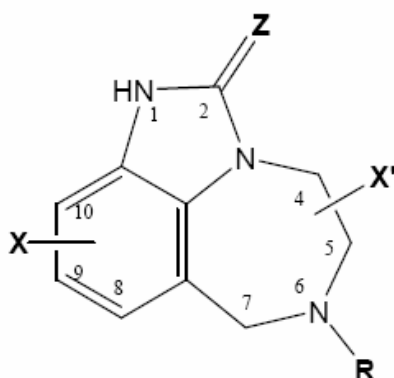


Figure 1 : Structure générale des dérivés TIBO étudiés.

Les auteurs se sont servis de 4 descripteurs pour décrire les composés étudiés.

Néanmoins ils n'ont pu introduire que 82 composés dans le modèle proposé en excluant 7 *outliers* (molécules exclues du modèle).

Le modèle proposé par ces auteurs est représenté par l'équation suivante:

$$\log(1/IC_{50}) = 0.39 \text{ LogP} + 1.16 \text{ B18} + 1.49 \text{ IZ} + 0.88 \text{ IR} + 1.80 \quad \text{Equation 1}$$

$$n = 82 \text{ (7 outliers)} \quad R^2 = 0.861 \quad S = 0.550 \quad R^2_{CV} = 0.840$$

Dans l'équation 1:

- LogP est le coefficient de partage déterminé pour la totalité des composés.
- B18 est le paramètre de Verloop déterminé pour le substituant en position 8 du noyau phénylique.
- IZ est un indicateur variable prenant la valeur 1 si Z est un oxygène et la valeur 0 si Z est un soufre.
- IR est un indicateur variable prenant la valeur 1 si R est un 3,3 diméthylallyle et 0 dans les autres cas.

- Le premier Modèle de la littérature établie par les Réseaux de neurones artificiels

Douali et coll. [2-3] ont essayé de corrélérer l'activité inhibitrice des dérivés TIBO par application des RNA à trois couches. Leur couche d'entrée est constituée de 4 neurones représentant les 4 descripteurs proposés par Garg et coll. [1], la couche de sortie, représenté par un seul neurone, fournit l'activité inhibitrice des dérivés TIBO. Ces auteurs ont constatés que les RNA sont mieux adaptés à l'étude des RQSA des dérivés TIBO et que cette méthode est capable de déterminer la non linéarité dans les relations entre l'activité et les 4 descripteurs moléculaires.

2.1 Modèle RQSA des dérivés TIBO établie dans ce présent travail

Dans ce présent travail nous allons utiliser la totalité de la base de donnée, 89 dérivés TIBO introduits par Garg et coll. [1] en calculant pour chacune des centaines de descripteurs. Nous allons ensuite utilisé des techniques statistiques qui permettent de réduire ce nombre de descripteurs. Nous allons donc proposer quelques nouveaux descripteurs jamais cités dans la littérature des RQSA des TIBO.

Les structures des différents substituant ainsi que les activités expérimentales sont données dans le tableau 1.

Tableau 1: Les structures chimiques des différents composés étudiés et leurs activités anti-VIH associées.

N°	Substituant				Activité anti-VIH Expérimentale
	X	Z	R	X'	log(1/IC50)
1	H	S	DMA	5-Me(S)	7.3600
2	9-Cl	S	DMA	5-Me(S)	7.4700
3	8-Cl	S	DMA	5-Me(S)	8.3700
4	8-F	S	DMA	5-Me(S)	8.2400
5	8-SMe	S	DMA	5-Me(S)	8.3000
6	8-OMe	S	DMA	5-Me(S)	7.4700
7	8-OC ₂ H ₅	S	DMA	5-Me(S)	7.0200
8	8-CN	O	DMA	5-Me(S)	5.9400
9	8-CN	S	DMA	5-Me(S)	7.2500
10	8-CHO	S	DMA	5-Me(S)	6.7300
11	8-CONH ₂	O	DMA	5-Me(S)	5.2000
12	8-Br	O	DMA	5-Me(S)	7.3300
13	8-Br	S	DMA	5-Me(S)	8.5200
14	8-I	O	DMA	5-Me(S)	7.0600
15	8-I	S	DMA	5-Me(S)	7.3200
16	8-C=CH	O	DMA	5-Me(S)	6.3600
17	8-C=CH	S	DMA	5-Me(S)	7.5300
18	8-Me	O	DMA	5-Me(S)	6.0000
19	8-Me	S	DMA	5-Me(S)	7.8700
20	9-NO ₂	O	CPM	5-Me(S)	4.4800
21	8-NH ₂	O	CPM	5-Me(S)	3.0700
22	8-NMe ₂	O	CPM	5-Me(S)	5.1800
23	9-NH ₂	O	CPM	5-Me(S)	4.2200
24	9-NMe ₂	O	CPM	5-Me(S)	5.1800
25	9-NHCOMe	O	CPM	5-Me(S)	3.8000
26	9-NO ₂	S	CPM	5-Me(S)	5.6100
27	9-F	S	DMA	5-Me(S)	7.6000
28	9-CF ₃	O	DMA	5-Me(S)	5.2300
29	9-CF ₃	S	DMA	5-Me(S)	6.3100
30	9-Me	O	DEA	5-Me(S)	6.5000
31	10-OMe	O	DMA	5-Me(S)	5.1800
32	10-OMe	S	DMA	5-Me(S)	5.3300
33	9,10-di-Cl	S	DMA	5-Me(S)	7.6000
34	10-Br	S	DMA	5-Me(S)	5.9700
35	H	O	CH ₂ CH=CH ₂	5-Me(S)	4.1500
36	H	O	2-MA	5-Me(S)	4.3300
37	H	O	CH ₂ CO ₂ Me	5-Me(S)	3.0700
38	H	O	CH ₂ C≡CH	5-Me(S)	3.2400
39	H	O	CH ₂ -2-furanyl	5-Me(S)	3.9700
40	H	O	CH ₂ CH=CH ₂ [S(+)]	5-Me(S)	4.1800
41	H	O	CH ₂ CH ₂ CH=CH ₂	5-Me(S)	4.3000
42	H	O	CH ₂ CH ₂ CH ₃	5-Me(S)	4.0500
43	H	O	2-MA[S(+)]	5-Me(S)	4.7200
44	H	O	CPM	5-Me(S)	4.3600
45	H	O	CH ₂ CH=CHMe(E)	5-Me(S)	4.2400
46	H	O	CH ₂ CH=CHMe(Z)	5-Me(S)	4.4600
47	H	O	CH ₂ CH ₂ CH ₂ Me	5-Me(S)	4.0000
48	H	O	DMA	5-Me(S)	4.9000
49	H	O	CH ₂ C(Br)=CH ₂	5-Me(S)	4.2100
50	H	O	CH ₂ C(Me)=CHMe(E)	5-Me(S)	4.5400
51	H	O	DMA[R(+)]	5-Me(S)	4.6600
52	H	O	DMA[S(+)]	5-Me(S)	5.4000
53	H	O	CH ₂ C(C ₂ H ₅)=CH ₂	5-Me(S)	4.4300
54	H	O	CH ₂ CH=CHC ₆ H ₅ (Z)	5-Me(S)	3.9100
55	H	O	CH ₂ C(CH=CH ₂)=CH ₂	5-Me(S)	4.1500
56	8-Cl	S	DMA	H	7.3400

57	9-Cl	S	DMA	H	6.8000
58	H	O	2-MA	5,5-di-Me	4.6400
59	H	O	2-MA	4-Me	4.5000
60	9-Cl	S	2-MA	4-Me(S)	6.1700
61	9-Cl	S	CPM	4-Me(R)	5.6600
62	H	O	C3H7	4-CHMe ₂	4.1300
63	H	O	2-MA	4-CHMe ₂	4.9000
64	H	O	2-MA	4-C ₃ H ₇	4.3200
65	H	O	DMA	7-Me	4.9200
66	8-Cl	O	DMA	7-Me	6.8400
67	9-Cl	O	DMA	7-Me	6.8000
68	H	S	C3H7	7-Me	5.6100
69	H	S	DMA	7-Me	7.1100
70	8-Cl	S	DMA	7-Me	7.9200
71	9-Cl	S	DMA	7-Me	7.6400
72	H	O	DMA	4,5-di-Me(<i>cis</i>)	4.2500
73	H	S	DMA	4,5-di-Me(<i>cis</i>)	5.6500
74	H	S	CPM	4,5-di-Me(<i>trans</i>)	4.8700
75	H	S	DMA	4,5-di-Me(<i>trans</i>)	4.8400
76	H	S	DMA	5,7-di-Me(<i>trans</i>)	7.3800
77	H	S	DMA	5,7-di-Me(<i>cis</i>)	5.9400
78	9-Cl	O	DMA	5,7-di-Me(R,R- <i>trans</i>)	6.6400
79	9-Cl	S	DMA	5,7-di-Me(R,R- <i>trans</i>)	6.3200
80	H	S	DMA	4,7-di-Me(<i>trans</i>)	4.5900
81	9-Cl	O	DMA	5-Me(S)	6.7400
82	9-Cl	S	CPM	5-Me(S)	7.4700
83	H	S	CPM	5-Me(S)	7.2200
84	H	O	C3H7	5-Me	4.2200
85	H	S	C3H7	5-Me	5.7800
86	H	O	2-MA	5-Me	4.4600
87	H	S	DMA	5-Me	7.0100
88	H	O	DMA	5-Me(S)	5.4800
89	H	S	2-MA	2-Me(S)	7.5900

DMA: 3,3-Dimethylallyl. **CPM:** Cyclopropylmethyl. **DEA :** 3,3-Diethylallyl. **MA:** 2-Methylallyl.

2.1.1 Calcul et sélection des descripteurs moléculaires

Dans une étude RQSA, Les informations concernant les données à étudier sont représentées par des descripteurs moléculaires. La détermination de ces derniers est une étape cruciale dans toute étude RQSA. Les données d'entrée doivent fournir le plus d'information possible sur les molécules étudiées. Dans ce travail les termes paramètres, variables, ou descripteurs moléculaires sont utilisés pour désigner les caractéristiques des molécules.

La sélection des différentes variables a été faite comme suit :

2.1.1.1 La sélection des descripteurs au moyen du programme Codessa

Dans ce présent travail, le calcul des différents descripteurs et leur sélection a été faite en employant le programme Codessa (*COMprehensive DEScriptors for Structural and Statistical Analysis* [7]). Ce programme calcul une grande variété de descripteurs moléculaires sur la base de la structure géométrique 3D. En effet, il calcul plus de 500

Chapitre II : Etude Théorique et RQSA des TIBO

descripteurs moléculaires appartenant à cinq classes différentes à savoir les descripteurs: Constitutionnels, Topologiques, Electrostatiques, Géométriques et Quantiques.

Les différentes étapes qui ont été suivies dans la sélection des meilleurs descripteurs appropriés à cette base de donnée, en utilisant au départ la Régression Multi Linéaire (MLR), sont résumées comme suit (cette procédure sera appliquée dans les prochains chapitres) :

1/- Structure et Calcul Quantique:

- Construction des structures 3-D au moyen du logiciel Agui [8].
- Optimisation des structures 3-D précédente au moyen de la méthode semi-empirique AM1 [9] qui se trouve dans le programme Ampac [8] avec deux séries de mots clefs permettant:
 - Le calcul SCF: (les mots clefs sont pris dans ce cas par défaut dans le programme.
 - Le calcul thermo: cette procédure permet le calcul des grandeurs thermodynamiques et nécessite les mots clefs suivant: thermo + force pour une température de 298.15 K. Pour calculer les grandeurs thermo à une température donnée il faut utiliser le mot clef: thermo.

2/-Calcul des Descripteurs:

- a/- Charger les fichiers de donnée (SCF.txt et Thermo.txt créés dans l'étape précédente pour chaque structure) dans le logiciel Codessa [7].
- b/- Lancer le calcul des descripteurs.
- c/- Lancer l'heuristique avec un choix du nombre de descripteurs désiré.
- d/- Procédures exécutées par Codessa pour le choix des meilleurs MLR:

1ère étape: tous les pairs de descripteurs quasi-orthogonaux sont sélectionnés dans une base de données initiale. Deux descripteurs sont orthogonaux si le coefficient d'inter-corrélation est inférieur à 0.1.

2ème étape: Codessa va prendre les paires de descripteurs orthogonaux pour établir des équations de régression bi-paramétrique.

Les dix meilleurs pairs de descripteurs (Tot=20 descripteurs) sont pris pour la troisième étape.

3ème étape: à un jeu de descripteur (n=20) défini dans l'étape 2 le programme ajoute automatiquement le (n+1)^{ème} descripteur si ce celui-ci est non significativement corrélé aux n descripteur procédant (le coefficient d'inter-corrélation est inférieur à 0.8). Cette itération est répétée jusqu'à ce que la meilleure corrélation soit obtenue avec le nombre de descripteurs (pris comme variables) choisi dans l'heuristique.

Pour notre base de donnée et dans la première étape de ce travail, sept descripteurs moléculaires ont été sélectionnés par le programme heuristique, en utilisant MLR. Les différents paramètres sélectionnés sont résumés dans le **tableau 2**. Ils font appel à des paramètres hydrophobiques, électroniques et topologiques.

Tableau 2: Descripteurs Moléculaires sélectionnés dans la première étape.

Notation	Descripteurs
MD1	<i>LogP: Logarithme du coefficient du partage dans le mélange n-octanol/eau.</i>
MD2	<i>Average nucleophyl reaction index for a N atom.</i>
MD3	<i>Minimum total interaction for a H-N bond.</i>
MD4	<i>Minimum (>0.1) bond order of a N atom.</i>
MD5	<i>ESP-HBSA H-bonding surface area.</i>
MD6	<i>Maximum atomic state energy for a N atom.</i>
MD7	³ χ : Indice de Connectivité du troisième ordre.

La sélection des descripteurs à l'aide du programme Codessa indique que les liaisons hydrogènes jouent un rôle important dans l'inhibition de la INNTI. On remarque aussi que l'azote dans les TIBO joue un rôle très important. En effet, le groupement (NH) de ce composé est capable de former des pont hydrogène dans le site de liaison des INNTI [10]. Les résultats du tableau 2 montrent que la connectivité intervient aussi comme un facteur dans l'inhibition.

2.1.1.2 Addition de quelques descripteurs pris de la littérature

Après avoir effectué la sélection des sept meilleurs descripteurs appropriés à cette base de donnée par le programme, nous allons ajoutés dans cette deuxième partie trois autres descripteurs. Ces descripteurs contribuent efficacement dans l'établissement des Relations

Quantitatifs Structure Activité anti-VIH des composés TIBO [1]. Ils caractérisent l'hydrophobe, les aspects stériques et électroniques. Ces paramètres ajoutés sont :

- I_X est représenté par le descripteur MD8. Le substituant X sur le cycle du benzène dans le TIBO peut avoir différentes positions (Figure 1), ce descripteur a un effet stérique. En effet, on va attribuer les valeurs : $I_X = 1$ quand il est en position 8, $I_X = 0,5$ pour la position 9 et $I_X = 0$ pour la position 10. Nous notons que ce descripteur diffère des paramètres de Verloop proposés par Garg et coll. [1].
- I_Z représenté par le descripteur MD9. Ce descripteur caractérise le type d'atome (S ou O) substitué en position Z sur le cycle imidazolidone (Figure 1). En effet, sa valeur sera $I_Z = 1$ si $Z = S$ et si $Z = O$ alors $I_Z = 0$.
- I_R représenté par MD10 va représenter le groupement 3, 3-diméthylallyl attaché à l'azote sur le cycle à 7-sommets en position 6 (voir Figure 2). Ce descripteur aura la valeur $I_R = 1$ si $R = 3, 3\text{-diméthylallyl}$ ($R = \text{DMA}$) et il sera nul ($I_R = 0$) dans le cas où on a un autre substituant ($R \neq \text{DMA}$).

Au total on aura 10 descripteurs. On a veillé dans ce travail à ce que le nombre de molécules soumises à la corrélation soit 5 fois supérieur au nombre de variables indépendantes.

2.1.2 Application des RNA à l'étude des RQSA des dérivés TIBO

2.1.2.1 Caractéristiques des RNA

2.1.2.1.1 - la détermination du nombre de neurones dans la couche cachée:

Nous allons utiliser dans toutes les applications qui vont être effectuées dans cette thèse un réseau à trois couches. La couche d'entrée contient autant de neurones que de descripteurs moléculaires. Des couches intermédiaires et une couche de sortie qui est constituée par un seul neurone fournissant l'activité calculée par le réseau.

Le problème posé dans l'application des RNA est le nombre de couches cachées à mettre dans un réseau pour obtenir une bonne paramétrisation. Certains auteurs [10, 11] ont proposé le rapport ρ qui joue un rôle important dans la détermination de l'architecture adéquate du RNA. Le paramètre ρ est défini de la manière suivante:

$$\rho = \frac{N}{H(I + O) + H + I}$$

Avec :

H : Nombre de neurones de la couche cachée.

I : Nombre de neurones de la couche d'entrée.

O : Nombre de neurones de la couche de sortie.

N : Nombre d'exemples.

Zupan et Gasteiger [12] proposent que le nombre de connections soit inférieur au nombre d'exemples. C'est-à-dire $\rho < 1$. Selon Andrea et Kalayeh [11], l'architecture optimale correspondrait à un ρ compris entre 1 et 2.2.

Charqaoui et coll. [13] ont adoptés dans leurs études un nombre de neurones dans la couche cachée qui permet de maintenir ρ dans l'intervalle $1 < \rho < 3$. Le critère de sélection sera basé sur l'examen des deux paramètres statistiques: le coefficient de corrélation r^2 et la déviation standard de l'erreur S définis précédemment dans le chapitre 1.

Dans ce programme utilisé la couche cachée et la couche de sortie sont munies d'un biais.

La fonction de transfert est une fonction sigmoïde ($f(x) = \frac{1}{1 + e^{-x}}$).

2.1.2.1.2- L'approche du Leave-one-out dans la phase de prédiction:

Un des attributs les plus importants des RNA est leur capacité à la généralisation, c'est-à-dire leur capacité de faire des prédictions fiables pour de nouvelles données avec une exactitude semblable à celle obtenue pour l'ensemble d'apprentissage. Le programme réalisé par Professeur Charqaoui [13-14] sur les RNA fonctionne avec la procédure du "Leave-one-out (LOO)" qui est une approche bien adaptée à l'estimation de cette capacité. Dans cette procédure, l'ensemble d'apprentissage est obtenu en retirant à chaque fois une molécule différente de l'ensemble des N molécules d'origine. Cette molécule retirée sert à tester le pouvoir de prédiction du RNA entraîné par les $N-1$ autres molécules. Selon ce programme les variables d'entrée sont réduites entre 0,1 et 0,9. Les poids de connections sont initialisés aléatoirement. Le coefficient d'apprentissage est initialisé avec une assez forte valeur (environ 1), puis décroît progressivement selon la fonction d'erreur. La fonction de transfert choisie est la fonction sigmoïde.

2.1.3 Résultats et discussion

2.1.3.1 Phase d'apprentissage

Dans ce présent travail, nous avons utilisé un réseau à trois couches. Dix neurones, décrivant les dix variables choisies auparavant, sept variables sélectionnées par Codessa et trois variables pris de la littérature, ces variables constituent la couche d'entrée. Un neurone dans la couche de sortie fournit l'activité anti-VIH calculée par le réseau. La couche cachée renferme un nombre variable de neurones (**Figure 2**).

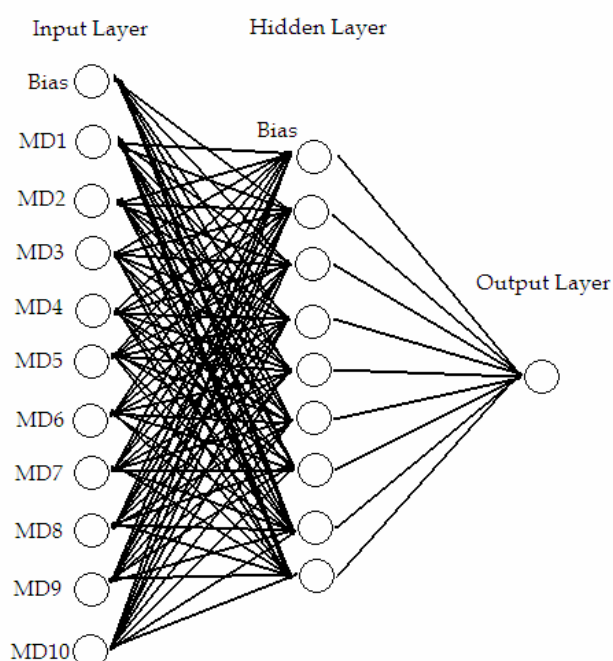


Figure 2 : Architecture du Réseaux de Neurones Artificiels adoptés dans ce travail.

Nous avons veillé à ce que ce nombre reste minimal afin d'éviter l'obtention de modèles complexes ou incapables de généraliser. Nous avons adopté des valeurs de p , comprises entre 1 et 3 [13]. Les architectures obtenues sont donc de la forme $10-n-1$. n varie de 3 à 8. Le pouvoir d'ajustement des différentes architectures a été estimé par le coefficient de corrélation r^2 et la déviation standard de l'erreur d'apprentissage S .

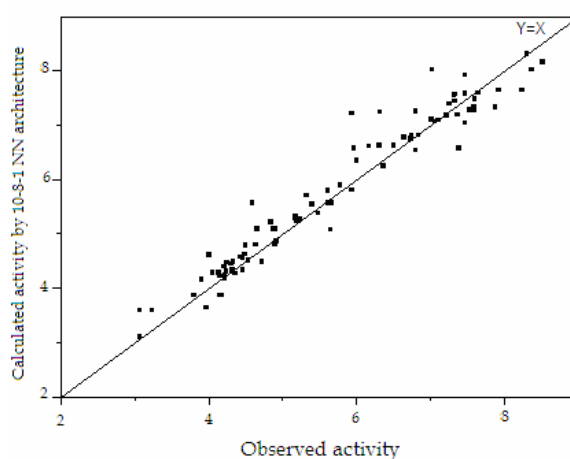
Les différents résultats de l'apprentissage des différentes architectures des RNA comparés aux résultats MLR sont résumés sur le tableau 3.

Tableau 3: coefficient de corrélation r^2 et la déviation standard de l'erreur S trouvés par RNA comparés à MLR.

Architecture	R^2	S
10-3-1	0.930	0.38
10-4-1	0.930	0.38
10-5-1	0.941	0.35
10-6-1	0.941	0.35
10-7-1	0.941	0.35
10-8-1	0.942	0.34
RML	0.857	0.54

L'analyse de ce tableau montre de bonnes performances de toutes les architectures des RNA entraînées comparés au RML. On note de faibles valeurs de S et des coefficients de corrélation élevés avec RNA comparés à l'analyse de régression, particulièrement pour l'architecture 10-8-1. Cette analyse nous conduit à conclure qu'il y a une forte corrélation entre l'activité anti-VIH et les dix descripteurs choisis. Les dix variables parviennent à bien décrire la variation de l'activité anti-VIH des dérivés TIBO.

La corrélation entre les valeurs des activités expérimentales et celles calculées par le réseau 10-8-1 est reportée sur la Figure 3.

**Figure 3 :** La corrélation entre les valeurs des activités expérimentales et celles calculées par RNA.

2.1.3.2 Phase de prédiction

Le pouvoir prédictif des RNA a été estimé par le coefficient de corrélation prédictif r^2_{pred} et la déviation standard de l'erreur de prédiction (S_{press}).

$$r^2_{pred} = \frac{SSY - PRESS}{SSY} \quad S_{press} = \sqrt{\frac{PRESS}{N}}$$

SSY: est la somme des carrées des déviations des valeurs expérimentales par rapport à leur moyenne

PRESS: est la somme des carrées des erreurs de prédiction.

N: est le nombre d'exemples dans la base de données

Les RNA ont la capacité de faire des prédictions sur de nouveaux exemples avec une performance semblable à celle de la phase d'apprentissage. En plus de l'évaluation du pouvoir prédictif des RNA, la procédure de LOO a été utilisée pour déterminer la meilleure architecture adaptée à l'étude des 89 dérivés TIBO. En d'autres termes, l'architecture qui donnerait la plus faible erreur de prédiction.

Les résultats obtenus après le test des différentes architectures sont regroupés dans le tableau 4.

Tableau 4: Les valeurs des coefficients de corrélation et des déviations standard de l'erreur de prédiction.

Architecture	R^2_{pred}	S_{press}
10-3-1	0.749	0.72
10-4-1	0.710	0.77
10-5-1	0.731	0.74
10-6-1	0.767	0.69
10-7-1	0.750	0.71
10-8-1	0.816	0.61

Les valeurs des coefficients de corrélation qui figurent dans le tableau 4 montrent un bon pouvoir prédictif des RNA. Le tableau 4 montre aussi que la valeur maximale de R^2_{pred} a été obtenue par l'architecture 10-8-1. Il s'en suit que sa capacité à la généralisation est importante. Nous rappelant que pour cette même architecture le coefficient de corrélation et la déviation standard de l'erreur étaient satisfaisants dans la phase d'apprentissage (Tableau

3). La capacité de généralisation étant le critère le plus important, l'architecture 10-8-1 serait donc la plus adéquate à l'étude des RQSA des dérivés TIBO. Nous avons donc adapté cette architecture pour la suite du travail.

La capacité de prédiction de l'architecture 10-8-1 peut être nettement identifiée par l'analyse de la figure 4. Cette figure présente les valeurs des activités anti-VIH prédites par l'architecture 10-8-1 en fonction des valeurs expérimentales.

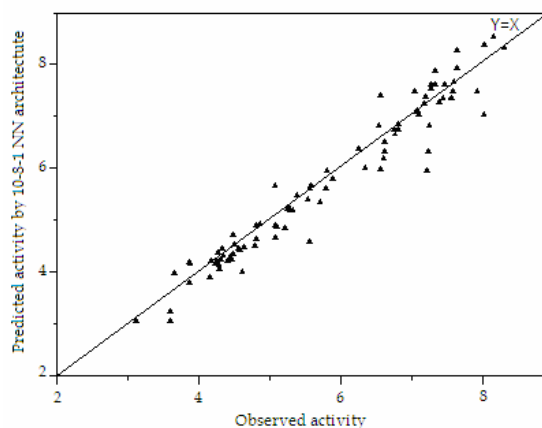


Figure 4 : valeurs des activités anti-VIH prédite par l'architecture 10-8-1 en fonction des valeurs expérimentales.

On note une corrélation importante entre les valeurs prédites et celles expérimentales. Plusieurs observations se trouvent très proches de, voire sur la droite d'équation $Y = x$.

2.1.3.3 Contribution des variables à la RQSA

Un des principaux objectifs des études RQSA est la détermination des facteurs qui gouvernent l'activité. Par conséquent, l'évaluation de la pertinence des variables est d'une assez grande importance pour les études RQSA. C'est pourquoi nous avons choisi d'estimer les contributions des différentes variables au modèle. La figure 5 représente les contributions (en %) des 10 variables figurant dans le modèle RQSA.

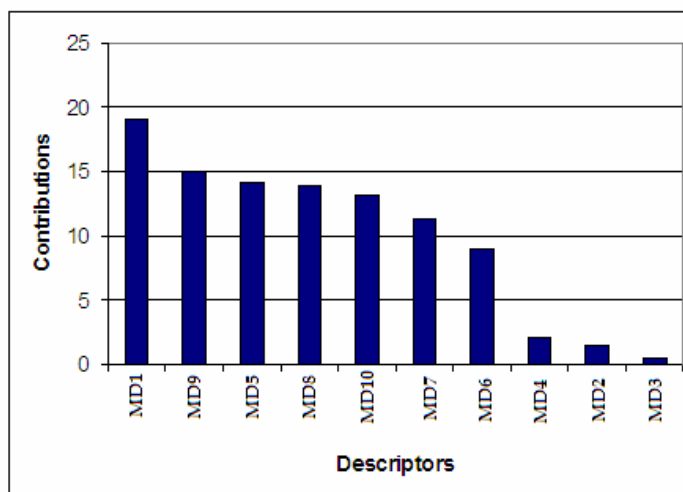


Figure 5 : Contribution en % des 10 variables au modèle RQSA.

La figure 5 indique l'importance relative des différents descripteurs, ils adoptent l'ordre suivant:

MD1 > MD9 > MD5 > MD8 > MD10 > MD7 > MD6 > MD4 > MD2 > MD3.

Cette ordre indique l'importance de la variable décrivant l'aspect hydrophobique MD1 (*LogP*) dans l'établissement de la relation structure-activité anti-VIH des dérivés **TIBO** de ce modèle. La principale raison qui pourrait expliquer la contribution la plus importante de ce descripteur par rapport aux autres: est le rôle protagoniste de l'hydrophobicité des dérivés TIBO dans le mécanisme de l'inhibition de la TI [1, 15-17]. Ce résultat est en parfait accord avec le modèle RNA proposé par Douali et Coll [2] construit sur les 82 molécules en utilisant les 4 descripteurs pris de la littérature [1].

Le transport, le passage à travers les membranes, la bioaccumulation ou encore l'activité pharmacologique d'une molécule peuvent être conditionnés par son partage entre une phase lipidique et une phase aqueuse, c'est-à-dire son caractère hydrophile.

Celui-ci peut être quantifié par le coefficient de partage eau-octanol, noté *LogP*, qui mesure la solubilité différentielle d'un soluté dans ces deux solvants non miscibles. Ce paramètre intervient fermement dans l'interaction entre les dérivés TIBO et leur site récepteur, nous avons jugé utile de suivre la variation de l'activité calculée par le réseau en fonction de la variation de *LogP*. Nous avons ainsi tracé la courbe de la Figure 6.

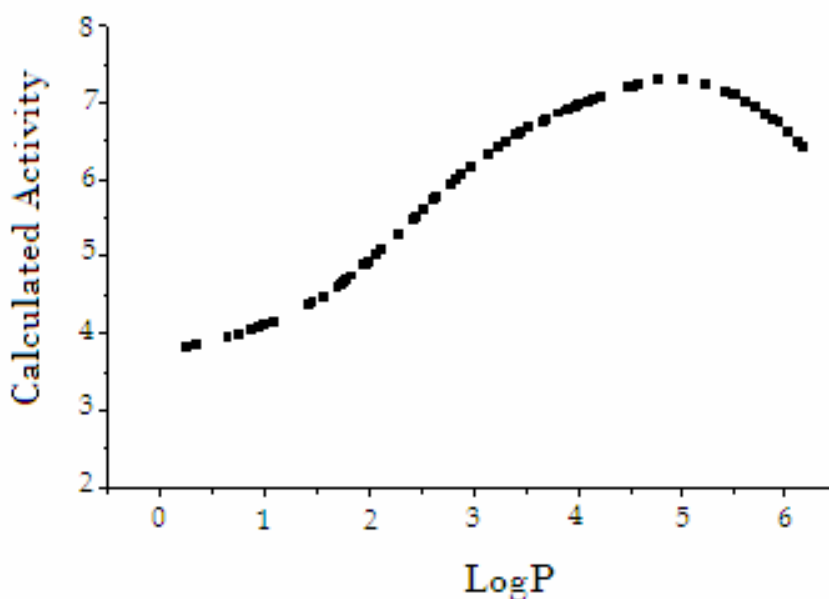


Figure 6 : variation de l'activité anti-VIH en fonction de la variation de LogP.

La figure 6 révèle une relation non linéaire entre l'activité anti-VIH calculée et la variable *LogP*. Une dépendance parabolique peut être constatée. L'activité anti-VIH augmente avec l'augmentation de *LogP* jusqu'à ce qu'il atteigne un maximum correspondant à la valeur 7.57, puis elle diminue. On peut donc déduire le rôle du caractère hydrophobe des composés non

seulement dans l'accroissement de l'activité anti-VIH des dérivés TIBO mais aussi dans les processus de transport et de perméabilité. Il convient de noter que la valeur optimum de *LogP*, valant 4.32, se trouve dans l'intervalle [2.31 - 5.72] déterminé par Garg et coll. [1] pour la classe des INNTIs.

Le rôle de la forme des molécules dans leur interaction avec la TI se traduit par les descripteurs I_z , I_R , I_x , $^3\chi$ et H-Binding Surface area (contribution > 10%) qui ont aussi une contribution considérable. En effet, ces descripteurs traduisent l'importance de la forme, de la connectivité et des liaisons hydrogène à l'établissement la relation structure-activité anti-VIH des dérivés **TIBO** de ce modèle.

Le modèle qu'on vient de construire décrit la variation de l'activité anti-VIH en fonction des propriétés des dérivés TIBO. Les paramètres statistiques de ce modèle suggèrent une relation non linéaire entre cette activité et les variables décrivant les propriétés moléculaires des dérivés TIBO.

Il est relativement difficile de décrire un modèle obtenu à partir des RNA par une équation mathématique, comme dans le cas d'une MLR.

Conclusion

Cette application est un travail complet de modélisation moléculaire, donnant toutes les étapes de l'application du RNA dans les domaines RQSA d'anti-VIH des dérivés de TIBO. En effet, l'application de RNA sur les 89 dérivés de TIBO conduit aux conclusions suivantes :

- Nous avons proposé 6 nouveaux descripteurs jamais cités dans la littérature des RQSA des TIBO. Il s'agit de descripteurs qui tiennent compte des effets électroniques et stériques.
- l'association de tous les descripteurs trouvés explique l'interaction entre INNTI et la cible de l'enzyme.
- Ce travail confirme les résultats trouvés dans la littérature au sujet de l'importance du paramètre hydrophobe, en effet, ce paramètre contribue avec le plus grand pourcentage (19.2%) comparé aux autres descripteurs.
- IX ajusté dans ce travail donne un bon résultat comparé à celui proposé par Garg et coll. [1] et Douali et coll. [2].

REFERENCE:

- [1]- Garg, R.; Gupta, S. P.; Gao, H.; Babu, M. S.; Debnath, A. K. Chem. Rev. 1999, 99, 3525-3601.
- [2]- Douali, L.; Villemin, D.; Cherqaoui, D. Curr. Pharm. Des. 2003, 9, 1817-1826.
- [3]- Douali, L.; Villemin, D.; Cherqaoui, D. J. Chem. Inf. Comput. Sci. 2003, 43, 1200-1207.
- [4]- Kukla, M. J.; Breslin, H. J.; Diamond, C. J.; Grous, P. G.; Ho, C. Y.; Miranda, M.; Rodgers, J.; Sherrill, R. G.; De Clercq, E.; Pauwels, R.; Andries, K.; Moens, L. J.; Janssen, M. A. C.; Janssen, P. A. J. J. Med. Chem. 1991, 34, 3187-3197.
- [5]- Kukla, M. J.; Breslin, H. J.; Pauwels, R.; Fedde, C. L.; Miranda, M.; Scott, M. K.; Sherrill, R. G.; Raeymaekers, A.; Van Gelder, J.; Andries, K.; Janssen, M. A. C.; De Clercq, E.; Janssen, P. A. J. J. Med. Chem. 1991, 34, 746-751.
- [6]- Breslin, H. J.; Kukla, M. J.; Ludovici, D. W.; Mohrbacher, R.; Ho, W.; Miranda, M.; Rodgers, J. D.; Hitchens, T. K.; Leo, G.; Gauthier, D. A.; Ho, C. Y.; Scott, M. K.; De Clercq, E.; Pauwels, R.; Andries, K.; Janssen, M. A. C.; Janssen, P. A. C. J. Med. Chem. 1995, 38, 771-793.
- [7]- A.R. Katritzky, V.S. Lobanov, M. Karelson, CODESSA: Reference Manual, University of Florida, Gainesville, FL, 1994.
- [8]- Michael J. S. Dewar, Andrew J. Holder, Roy D. Dennington, Daniel A. Liotard, Donald G. Truhlar, Todd A. Keith, John M. Millam, C. David Harris, Copyright © 1994, 1997, 2004 Semichem
- [9]-Dewar and W. Thiel, J. Amer. Chem. Soc. 99, 4499 (1977).
- [10]- So, S.; Richards, W. G. J. Med. Chem. 1992, 35, 3201-3207.
- [11]- Andrea, T. A.; Kalayeh, H. J. Med. Chem. 1991, 34, 2824-2836.
- [12]- Zupan, J., Gasteiger, J. Neural Networks in Chemistry and Drug Design, Wiley-VCH,, Weinheim, 1999.
- [13]- Cherqaoui, D., Villemin, D., J. Chem. Soc. Faraday Trans. 90 (1994) 97-102.
- [14]- Charqaoui. D. « Etude de Relations Structure-Activités Biologiques, Propriétés Physiques et Réactivité de composés organiques au moyen des réseaux de neurone ». Thèse de Doctorat d'Etat. Université de Caen. France. 1994.
- [15] Latifa Douali , Didier Villemin , Driss Cherqaoui., Int. J. Mol. Sci. 5 (2004) 48-55.
- [16] De Clercq, Il Farmaco., 54 (1999) 26-45.
- [17] Esnouf, R., Ren, J., Ross, C., Jones, Y., Stammers, D., Stuart, Struct. Biol., 2 (1995) 303-308.

Chapitre III

Application des RNA à la Recherche de RQSA des Flavonoïdes

Introduction

Les Flavonoïdes ou bioflavonoïdes sont des métabolites secondaires regroupés selon leur structure phénylbenzopyrone [1, 2]. Ils sont présents dans des organismes très divers, par exemple dans les végétaux. Les flavonoïdes sont connus principalement pour leur activité antioxydants [3], anti-microbiens [3, 4]. Ils se révèlent particulièrement efficaces pour réduire la perméabilité des vaisseaux sanguins (propriétés vasculoprotectrices et veinotoniques). Ils sont notamment utilisés pour traiter les crises hémorroïdaires, les jambes lourdes et les troubles de la fragilité capillaire. Ils sont aussi utilisés comme des anti-inflammatoires, antiviraux, anti-allergiques, antibiotiques et anticarcinogéniques [3, 5]. L'une des propriétés biologiques les plus intéressantes de ces composés est leur habilité à inhiber l'intégrase dans le cycle réplcatif du VIH [6,7].

Les flavonoïdes présentent un squelette de base à 15 atomes de carbone, fait de deux cycles en C₆ reliés par une chaîne en C₃. Le pont à 3 carbones entre les deux phényles forme généralement un troisième cycle pyrone (Figure 1).

Tous les flavonoïdes ont une origine biosynthétique commune, ce qui explique qu'ils partagent la même structure de base. Actuellement, près de 5000 flavonoïdes ont été décrits [6-8].

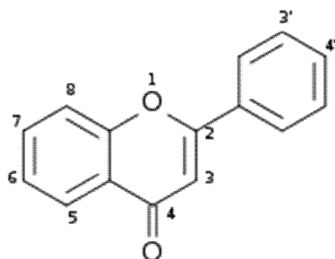


Figure 1 : Squelette de base des flavonoïdes.

La diversité des architectures moléculaires dans les flavonoïdes a conduit plusieurs études à s'intéresser à la recherche et à l'identification des paramètres moléculaires responsables des propriétés biologiques et physico-chimiques des flavonoïdes [5, 6-9].

La nécessité du développement des dérivés flavonoïdes imposée par le manque de remèdes efficaces contre le SIDA, a fait surgir un nombre important d'études consacrées à ces composés [3] pour la Recherche des Relations Quantitatives entre différents paramètres influençant l'activité anti-VIH-I des flavonoïdes. En effet, les RQSA contribuent efficacement dans l'étude des interactions entre les composés et leur site récepteur et génèrent ainsi des modèles qui orientent la recherche vers de nouveaux composés efficaces.

Dans la littérature, l'activité anti-VIH des 26 composés flavonoïdes a été étudiée par la régression Multi linéaire (MLR) [8]. Ces auteurs ont employé la méthode DFT pour l'optimisation des différentes géométries et le logiciel Matlab [10] pour l'analyse statistique. Pour une meilleure exactitude, ces auteurs utilisent la fonctionnelle hybride B3LYP avec la base 6-31G* pour calculer un ensemble de descripteurs [8]. Le modèle MLR a été construit avec seulement 24 molécules (1-24), deux molécules ont été considérées comme des molécules test (25 et 26). Les différentes structures sont rassemblées sur la Figure 2. Le meilleur modèle MLR obtenu par ces auteurs est basé sur les descripteurs suivants: l'électronegativité (χ), la charge atomique sur l'atome (C7) et le logarithme du coefficient de partage entre l'eau et l'octanol (CLogP), comme le montre l'équation 1 :

$$\text{Log} (1/\text{EC}_{50}) = 7.74 + 0.75 \text{ CLogP} - 0.49 \chi + 0.29 \text{ C7} \quad (\text{eq.1})$$

$$n = 23; R^2 = 0.86; F = 37.75; S_{\text{press}} = 0.45.$$

L'équation de la meilleure corrélation trouvée dans ces travaux [8] est donnée sous forme:

$$y = a_0 + \sum a_i x_i$$

Où:

"y": représente l'activité des différents composés pris des travaux expérimentaux de Hu et coll. [11].

" x_i ": représente les meilleurs descripteurs sélectionnés lors la recherche de la meilleur corrélation.

Dans cette équation nous trouvons aussi:

N: représente le nombre de molécules.

R²: coefficient de corrélation.

F: test de Fisher.

S_{press} : l'erreur standar de la prédiction qui est définit dans le chapitre précédent.

Dans ce présent travail, une recherche de Relation Quantitative entre la Structure et l'Activité anti-VIH de ces composés a été établit. En effet, plusieurs modèles MLR ont été construit et éventuellement comparés au modèle de la littérature [8]. La recherche de nouveaux descripteurs responsables de la variation de l'activité anti-VIH de ces composés a été effectuée. La non linéarité de la relation entre l'activité anti-VIH et les différents

descripteurs sélectionnés a été discutée en appliquant la méthode des réseaux de neurones artificiels.

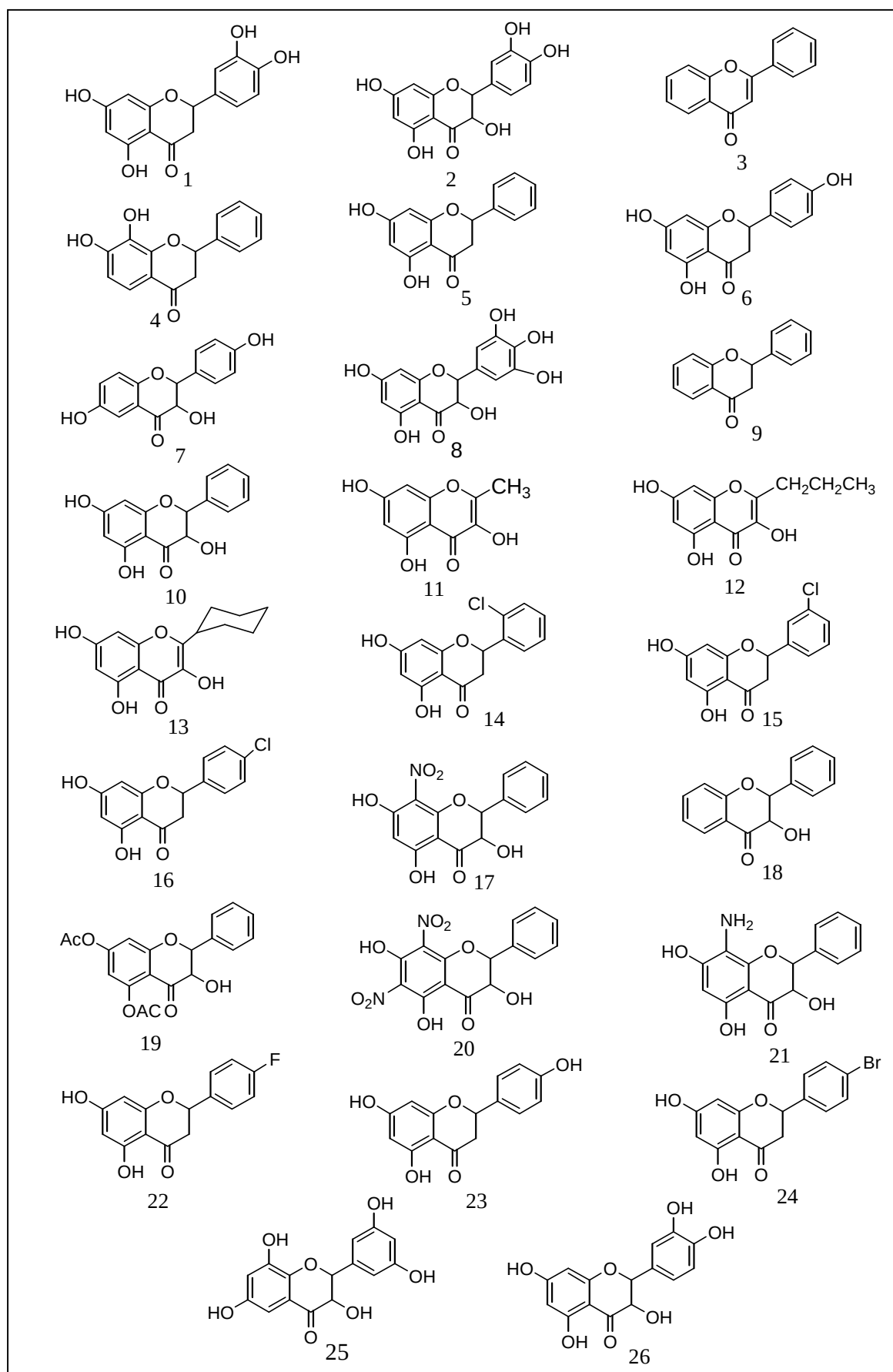


Figure 2: Structure des 26 flavonoïdes étudiée [8].

Base de données.

Notre travail concerne l'étude d'une série de 26 dérivés flavonoïdes indiquée sur la figure 2. Les valeurs des activités expérimentales anti-VIH, exprimées en $\text{Log}(1/\text{EC}_{50})$, sont données sur le tableau 1. EC_{50} est une concentration effective, elle indique la concentration requise pour protéger 50% des cellules contre la cytopathologie du virus. Nous rappelons que dans les études RQSA, c'est le logarithme de l'inverse de cette entité qui est utilisé comme variable dépendante.

Tableau 1: Les activités expérimentales anti-VIH des 26 flavonoïdes $\text{Log}(1/\text{EC}_{50})$ [8] avec les différents descripteurs associés [8].

composé	$\log(1/\text{EC}_{50})$	Composé	$\text{Log}(1/\text{EC}_{50})$
1	5.000	14	5.301
2	3.879	15	5.398
3	4.301	16	5.398
4	5.000	17	4.921
5	5.301	18	4.886
6	5.046	19	4.252
7	3.914	20	3.429
8	4.456	21	4.959
9	4.237	22	5.398
10	4.553	23	4.036
11	4.310	24	5.301
12	4.569	25	<3.480
13	4.721	26	<3.462

Méthode de calcul.

Les structures des composés ont été dessinées en utilisant le programme Chem-Office [11]. L'optimisation des différentes géométries a été faite au moyen de la méthode DFT/RB3LYP (B3LYP restreint) [12] avec la base 6-31G* [13] en utilisant le programme Gaussian 98W [14]. Les différents résultats géométriques trouvés ont été transférés dans le programme CODESSA développé par le groupe de Katritzky [15]. La procédure de calcul a été décrite dans le chapitre précédent.

Le succès d'une étude RQSA dépend énormément de la sélection des variables [16] et de la représentation de l'information. Les variables devraient donner le maximum d'informations sur la variation de l'activité et la colinéarité entre elles doit être maintenue au minimum possible.

Dans le but de rechercher de nouveaux descripteurs qui pourraient jouer un rôle dans l'activité anti-VIH de ces 26 composés flavonoïdes, nous avons construit plusieurs modèles

et nous avons commencé par une étude MLR en employant le programme CODESSA [15]. En effet, le programme heuristique de Codessa a permis le choix de quelques descripteurs avec la meilleure corrélation possible; nous avons comparé nos résultats avec ceux obtenus par J. Lameira et coll. [8]. Les meilleurs descripteurs, du meilleur modèle MLR, obtenus ont été employés comme base d'entrée pour l'étude avec les RNA.

La détermination du nombre Maximum de descripteurs

La détermination du nombre de variables qui décrivent les relations entre l'activité et la structure des composés chimiques est une étape cruciale dans les études RQSA. La méthode de 'breaking point' [17] est une des techniques utilisée. Cette dernière repose sur la variation du coefficient de corrélation R^2 en fonction du nombre de descripteurs. En effet, plusieurs études statistiques [17] ont révélés que le modèle de régression devient significatif quand la variation de R^2 devienne faible ($\Delta R^2 < 0.02 - 0.04$), ceci s'explique par une cassure de la courbe. Ainsi le nombre de descripteur se trouvant sur ce point est considéré comme le meilleur nombre. Dans notre cas on a reporté sur la figure 3 la variation du coefficient de corrélation R^2 en fonction de nombre de descripteur. L'analyse de cette courbe montre que la coupure (intersection de deux lignes) pour le coefficient de la corrélation est effectuée pour $n=3$. Par conséquent, le nombre maximum de descripteur est trois pour la suite de ce travail.

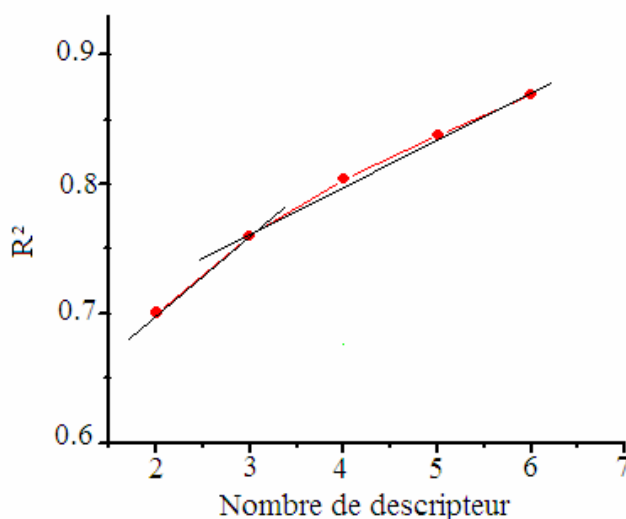


Figure 3: La variation de R^2 en fonction du nombre de descripteur.

Résultat et discussions.

1- La construction de modèles linéaire.

1-1 Premier modèle construit en employant seulement les descripteurs du programme CODESSA

Le premier modèle MLR construit pour cette base de donnée de 24 molécules (1-24), est basé uniquement sur les descripteurs sélectionnés par le programme Codessa. Les résultats montrent que le composé 23 pourrait être un outlier. Les outliers sont éliminées du modèle initial dans la tentative d'améliorer les résultats statistiques. En effet, un des problèmes majeurs des études RQSA par la RML est l'existence de quelques observations qui ne sont pas ajustables au modèle généré. Ces observations sont habituellement considérées comme des outliers et sont exclues du modèle [18]. Généralement, ce problème pourrait provenir de plusieurs sources:

- L'existence de dépendances non linéaires entre l'activité et les descripteurs. Ces relations peuvent être non apparentes dans le modèle linéaire.
- L'inadéquation des descripteurs moléculaires pour la description de la variation de l'activité étudiée.
- L'existence de quelques redondances dans la base de données.
- L'inhomogénéité dans la base de données.

Le modèle obtenu avec les descripteurs sélectionnés par Codessa est résumé par l'équation 2:

$$\text{Log} (1/\text{EC}_{50}) = 1.04 + 0.07 \text{ CLogP} + 3.23 \text{ Des1} + 0.002 \text{ Des 2} \quad (\text{eq.2})$$

$$R^2 = 0.8159 \quad F = 28.06 \quad S^2 = 0.0651 \quad R^2_{\text{CV}} = 0.7098$$

Les informations concernant l'équation de notre modèle sont données par :

- R^2 : Coefficient de corrélation.
- F: test de Fisher.
- S: L'erreur standard ou écart type résiduel.
- R^2_{CV} : Coefficient de cross-validated.

Dans ce premier modèle établi pour les dérivés flavonoïdes, trois descripteurs moléculaires ont été sélectionnés. Ils font appel à des paramètres hydrophobique et électronique. Le

paramètre hydrophobique CLogP est le coefficient de partage du composé entre l'eau et l'octanole. Ce descripteur représente la plus ou moins affinité des composés pour les lipides et pour les zones hydrophobes des protéines. Max partial charge for a Hydrogen atom (Des1) et PNSA-1 Partial negative surface area (Des2) [15], sont des descripteurs électroniques. Ces descripteurs expriment les propriétés électroniques des molécules. Ils tiennent en compte des possibilités d'interactions électroniques entre les flavonoïdes et son site récepteur (intégrase).

Nous avons obtenu la matrice de corrélation entre les descripteurs calculés (tableau 2). En effet, les corrélations entre les variables biaisent les résultats et, par conséquent, des modèles RQSA fortuits ou erronés pourraient être obtenus [19]. Pour éviter les difficultés d'interprétation des modèles résultants, nous avons procédé à une analyse des corrélations linéaires entre les variables. Les paires de variables ayant un coefficient de corrélation plus grand que 0.80 ont été considérées comme corrélées. La matrice obtenue, dans ce cas, prouve que ces descripteurs sont indépendants.

Tableau 2: Valeurs des corrélations linéaires entre les variables du modèle.

	LogP	Des1	Des2
<i>LogP</i>	1.0000		
<i>Des1</i>	-0.2151	1.0000	
<i>Des2</i>	-0.6981	0.3776	1.0000

La qualité d'ajustement des modèles établis par les RNA et les RML peut être jugée à partir des courbes de corrélation entre les valeurs des activités observées et celles calculées. Les valeurs des activités anti-VIH calculées dans ce modèle en fonction des valeurs expérimentales sont données sur la Figure 2.

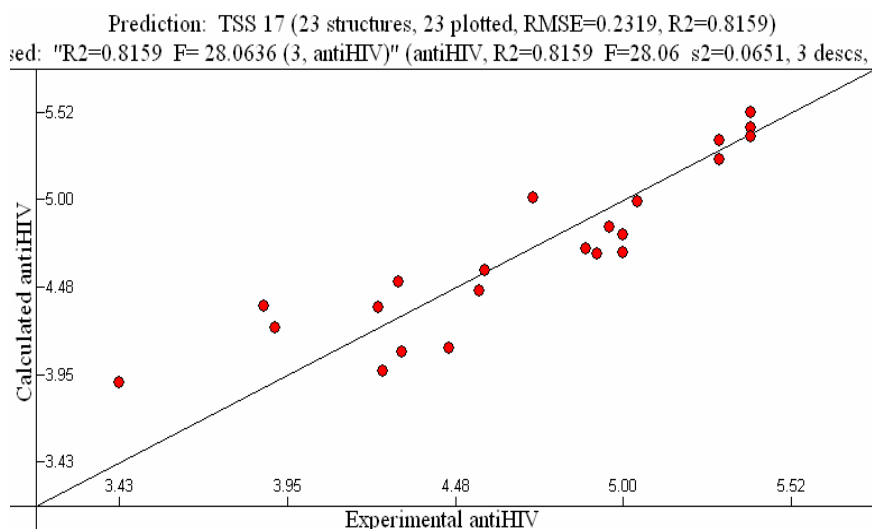


Figure 4: Les valeurs des activités anti-VIH calculées en fonction des valeurs expérimentales pour le premier modèle avec l'équation 2.

Sur la base de 26 composés flavonoïdes, on a gardé 2 molécules tests pour la validation du modèle, autrement dit pour la prédiction. L'utilisation de l'équation 2 nous a permis de calculer des valeurs théoriques prédites pour ces deux molécules, qui sont rassemblées sur le tableau 3. Nous remarquons que ces valeurs sont mieux corrélées que dans la littérature mais elles restent néanmoins légèrement surestimé par rapport aux valeurs expérimentales.

Tableau 3: l'activité anti-VIH prédite pour les molécules test en utilisant l'équation 2.

Structure	Experimental	Calc. Eq.2	Diff.	Calc. Eq.1 [3]	Diff
mol25	3.4800	3.6922	0.2122	3.860	0.380
mol26	3.4620	3.8106	0.3486	3.877	0.415

1-2 Deuxième modèle MLR construit en utilisant les descripteurs du programme CODESSA et les descripteurs donnés dans la littérature.

Afin de rechercher un nouvel ensemble de descripteurs pour établir la relation structure-activité de l'ensemble de notre base de donnée, nous proposons dans cette partie de travail d'ajouter les descripteurs trouvés dans la référence [8] à ceux de CODESSA. Les descripteurs additionnés sont : l'électronegativité (χ), la charge atomique sur quelques atomes principaux à savoir C6, C7, C8 et O11. L'équation trouvée pour ce modèle est:

$$\log(1/EC_{50}) = 0.77 + 0.042 \text{ CLogP} + 0.24 \text{ C7} - 0.21 \text{ X} \quad (\text{eq.3})$$

$$R^2 = 0.86 \quad F=37.75 \quad s^2=0.049 \quad R^2_{CV}=0.7780$$

Les trois descripteurs sélectionnés par l'heuristique correspondent au paramètre hydrophobique CLogP, et aux deux paramètres électroniques : l'électronégativité (χ) et la charge atomique sur atome C7. En effet, nous retrouvons le même modèle que la littérature (équation 1). En effet, le programme a sélectionné les mêmes descripteurs pris dans le modèle de la littérature avec le même poids. Ceci prouve que ces descripteurs jouent un rôle important dans l'activité anti-VIH des flavonoïdes. La matrice de corrélation entre ces 3 paramètres sélectionnés est donnée dans le tableau 4, nous remarquons qu'il y a une nette indépendance entre les trois variables.

Tableau 4 : Valeurs des corrélations linéaires entre les variables du modèle 2.

	LogP	C7	X
LogP	1.0000		
C7	0.1309	1.0000	
X	0.2144	0.0192	1.0000

L'activité anti-VIH prédite pour les molécules tests, 25 et 26, en utilisant l'équation 3 sont rassemblées dans le tableau 5. Nous remarquons qu'il n'y a pas une grande différence entre les résultats de la littérature [8] (l'équation 1) et les résultats trouvés dans l'équation 3 donnée par Codessa.

Table.5 l'activité anti-VIH prédite pour les molécules tests en utilisant l'équation 3.

Structure	Experimental	Calc. Eq.3	Diff.	Calc. Eq.1 [8]	Diff
mol25	3.4800	3.8602	0.3802	3.860	0.380
mol26	3.4620	3.8774	0.4154	3.877	0.415

Les valeurs des activités anti-VIH calculées pour ce modèle en fonction des valeurs expérimentales en utilisant l'équation 3 sont représentées sur la figure 5.

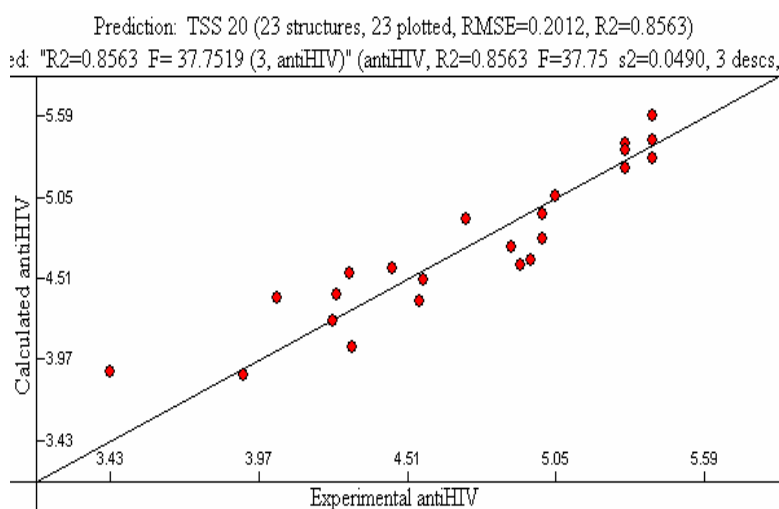


Figure 5: Les valeurs des activités anti-VIH calculées dans ce modèle en fonction des valeurs expérimentales pour le premier modèle avec l'équation 3.

La méthode MLR présente des limitations dans la plupart des cas, notamment quand la relation entre les variables dépendantes et indépendantes est si complexe qu'elle ne pourrait pas être simulée par une relation linéaire simple. Dans ce cas, les RNA ont montré un grand pouvoir de traitement de ce type de problèmes. D'ailleurs, il a été montré que la technique des RNA surpasse les techniques traditionnelles [20-23] et peut surmonter quelques limitations de ces approches RQSA [24]. Afin de vérifier la non linéarité entre les descripteurs du modèle 2, nous proposons pour cette deuxième partie de ce travail, d'établir un modèle RQSA par les RNA.

2- La construction du modèle non linéaire

2-1 Application de la méthode RNA au modèle 2 [25].

La force majeure des techniques RNA réside dans le fait qu'ils ne sont pas restreints à suivre un modèle spécifique. Ils sont capables d'extraire l'information à partir des données plus efficacement que les méthodes statistiques conventionnelles.

2-1-1 Phase d'apprentissage

Pour cette étude, nous avons utilisé un RNA à trois couches. La couche d'entrée est constituée de 3 neurones représentant les trois descripteurs proposés par la référence [8] et confirmés par notre étude (modèle 2), la couche de sortie, représenté par un seul neurone, fournit l'activité inhibitrice des dérivés flavonoïdes. Le nombre de neurones dans la couche

cachée peut varier selon l'application du RNA. En effet, Certains auteurs [26, 27] ont proposé le rapport ρ qui joue un rôle prédominant dans la détermination de l'architecture adéquate du RNA. Ce dernier a été décrit en détail dans le chapitre précédent. Dans ce présent travail, trois architectures (3-x-1, avec $x = 2, 3$ et 4) ont été entraînées en utilisant le programme des RNA élaboré par D. Cherqaoui [28] avec les caractéristiques des RNA décrites dans le chapitre précédent. Les résultats de l'apprentissage de ces architectures comparés à ceux des RML sont reportés dans le tableau 6.

Tableau 6: Résultats statistiques de l'apprentissage des RNA entraînées par 24 dérivés flavonoïde comparés aux résultats MLR.

Architecture	R^2	S
3-2-1	0.944	0.125
3-3-1	0.946	0.122
3-4-1	0.955	0.111
MLR (Codessa)	0.856	0.221
Literature MLR	0.860	/

L'analyse de ce tableau montre, de bonnes performances de toutes les architectures des RNA entraînées par rapport au modèle 2 et à la littérature [8]. On note des coefficients de corrélation élevés et de faibles valeurs de S, particulièrement pour l'architecture 3-4-1. Cette analyse nous conduit à conclure qu'il y a une forte corrélation entre l'activité anti-VIH et les trois descripteurs choisis. Les trois variables, corrélés par un modèle non linéaire décrivent bien la variation de l'activité anti-VIH des dérivés flavonoïdes.

2-1-2 Phase de prédiction

Nous avons testé le pouvoir de prédiction des réseaux sur l'ensemble des 24 dérivés flavonoïdes. Pour ce faire, nous avons accompli la procédure du LOO (leave-one – out) décrite dans le chapitre précédent. Toutes les architectures 3-x-1 (x variant de 2 à 4) ont été soumises à cette analyse. Les résultats obtenus sont reportés dans le tableau 7.

Tableau 7: Les coefficients de corrélation et les déviations standards de l'erreur de prédiction issus de l'étude des 24 dérivés flavonoïdes par les RNA et la RML.

Architecture	$R^2_{\text{préd}}$	S_{press}
3-2-1	0.837	0.214
3-3-1	0.699	0.291
3-4-1	0.663	0.308
MLR (Codessa)	0.778	0.201
Literature MLR [3]	0.80	0.45

L'examen, des valeurs statistiques représentées dans le tableau 7, montre que les RNA donnent des valeurs des coefficients de prédiction ($R^2_{\text{préd}}$) satisfaisants. L'architecture 3-2-1 donne la plus grande valeur du coefficient qui pourra prédire l'activité de 83.7% des molécules. Cette valeur reste élevée par rapport à la valeur donnée par MLR.

L'application des RNA au meilleur modèle MLR montre le succès du comportement non linéaire de quelques paramètres dans le modèle RQSA.

Le transport des molécules qui se traduit par le passage à travers les membranes peut être conditionnés par le partage entre une phase lipidique et une phase aqueuse, c'est-à-dire le caractère hydrophile de la molécule. Celui-ci peut être quantifié par le coefficient de partage eau-octanol, noté CLogP (décrit précédemment) qui mesure la solubilité différentielle d'un soluté dans ces deux solvants non miscibles. Le CLogP est ainsi figuré dans tous les modèles en tant que descripteur clé et cela pour la description et la prédiction de l'activité anti-VIH des flavonoïdes [8].

Tableau 8 : Les valeurs des différents descripteurs utilisés dans la littérature [8].

Compounds	CLogP	χ (eV)	EA	C6	C7	C8	O11	$\log(1/IC_{50})$	$\log(1/EC_{50})$
1	2.310	3.337	-0.528	-0.254	0.380	-0.247	-0.594	4.796	5.000
2	1.304	3.158	-0.496	-0.256	0.382	-0.246	-0.631	3.879	3.879
3	3.480	3.686	-0.449	-0.131	-0.130	-0.18	-0.521	4.167	4.301
4	2.733	3.560	-0.320	-0.190	0.318	0.249	-0.526	4.854	5.000
5	3.563	3.451	-0.624	-0.253	0.380	-0.246	-0.591	4.328	5.301
6	2.906	3.414	-0.478	-0.254	0.379	-0.247	-0.594	4.456	5.046
7	1.640	3.658	-0.450	0.360	-0.189	-0.185	-0.563	3.804	3.914
8	0.637	3.128	-0.563	-0.217	0.380	-0.282	-0.632	4.161	4.456
9	3.475	3.953	0.050	-0.157	-0.124	-0.187	-0.471	4.347	4.237
10	2.035	3.742	-0.290	-0.257	0.387	-0.251	-0.572	4.357	4.553
11	1.964	3.978	0.170	-0.215	0.378	-0.279	-0.584	3.670	4.310
12	3.022	3.924	0.122	-0.215	0.378	-0.28	-0.586	4.137	4.569
13	4.085	3.904	0.141	-0.216	0.378	-0.28	-0.588	4.377	4.721
14	4.026	3.456	-0.689	-0.252	0.380	-0.243	-0.587	4.770	5.301
15	4.276	3.377	-0.884	-0.252	0.380	-0.245	-0.588	4.854	5.398
16	4.276	3.511	-0.567	-0.252	0.380	-0.246	-0.589	4.796	5.398
17	1.803	3.469	-1.016	-0.263	0.377	0.183	-0.588	4.921	4.921
18	2.843	3.349	-0.544	-0.136	-0.128	-0.179	-0.561	4.770	4.886
19	2.198	3.769	-0.131	-0.268	0.400	-0.274	-0.504	< 3.428	4.252
20	0.043	3.510	-1.577	0.157	0.338	0.222	-0.586	3.429	3.429
21	2.370	3.606	-0.387	-0.231	0.368	0.097	-0.589	4.469	4.959
22	3.706	3.443	-0.671	-0.253	0.380	-0.246	-0.59	4.886	5.398
23	2.445	3.849	0.068	-0.220	0.384	-0.289	-0.546	3.532	4.036
24	4.195	3.541	-0.654	-0.214	0.379	-0.283	-0.587	4.678	5.301
25	0.334	3.230	-0.674	0.374	-0.246	0.299	-0.556	< 3.480	< 3.480
26	0.771	3.767	0.103	-0.221	0.387	-0.570	-0.558	< 3.462	< 3.462

D'autre part, les deux descripteurs électroniques : l'électronégativité (χ) et la charge atomique sur l'atome (C7) ressortaient comme étant des descripteurs importants dans le modèle. En effet, Selon la théorie de Pearson qui décrit l'électronégativité comme étant une mesure du caractère intrinsèque donneur - accepteur d'une molécule et détermine la direction du transfert d'électron entre deux molécules quand elles entrent en interaction

[29]. Fesen et coll [30] ont montré aussi que l'inhibition de l'intégrase dans le VIH par les flavonoïdes pourrait agir en chélateur avec le métal bivalent (Mg^{2+}) dans le site actif d'intégrase. Rappelons que le terme « Chélateur » se dit sur un corps qui a la propriété de se lier à un ion positif bivalent et trivalent avec lequel il forme un composé dans lequel l'atome central est lié aux atomes voisins par au moins deux liaisons en formant une structure sous forme d'anneau. Le Tableau 8 indique que les flavonoïdes les plus actifs possèdent la plus faible valeur d'électronégativité (χ). Le second descripteur électronique qui est la charge atomique sur l'atome C7 a été aussi sélectionnée comme étant le meilleur descripteur, ce descripteur peut être pris comme une description de la délocalisation électronique autour de l'anneau A. En observant la structure générale des flavonoïdes (Figure 2) on remarque que l'atome C6 et C8 peuvent aussi jouer le même rôle que C7. Rappelant que ces charges atomiques ont été aussi sélectionnés par d'autres modèles MLR dans la littérature [8] mais le modèle donné avec la charge atomique sur C7 reste le meilleur modèle de point de vue statistique par rapport aux autres modèles avec C6 et C8 [8].

On observe pour l'ensemble de composés étudiés dans ce travail que l'activité anti-VIH augmente quand un composé possède des valeurs d'électronégativité (χ) basse combinées avec une plus grande charge positives sur l'atome C7 et des valeurs élevées pour le CLogP.

Conclusion

L'examen des résultats des RNA comparativement à ceux donnés par l'analyse MLR dévoile des points intéressants:

- Les RNA ont donné de bons résultats statistiques dans les deux phases d'apprentissage et de prédiction. Les carrés des coefficients de corrélation étaient plus élevés et les déviations standards étaient inférieures que celles données par les RML.
- L'application des RNA au meilleur modèle MLR montre le succès du comportement non linéaire de quelques paramètres dans le modèle RQSA. Cette étude confirme une fois de plus que, l'électronégativité (χ), la charge atomique sur l'atome C7 et le CLogP jouent un rôle clés dans l'activité anti-VIH des flavonoïdes.

REFERENCE.

- [1]- E. de Rijke, P. Out, W.M.A. Niessen, F. Ariese, C. Gooijer, U.A.Th.Brinkman, J. Chromatogr. A 1112 (2006) 31.
- [2]- Ko, F.; Chu, C.; Lin, C.; Chang, C.; Teng, C. *Biochim. Biophys. Acta* **1998**, 1389, 81-90.
- [3]- W. Heller, G. Forkmann, in: J.B. Harborne (Ed.), *The Flavonoids*, Chapman & Hall, London, 1988, p. 399.
- [4]- E. Wollenweber, in: V. Cody, E. Middleton Jr., J.B. Harborne, A. Beretz (Eds.), *Plant Flavonoids in Biology and Medicine, II. Biochemical, Cellular and Medicinal Properties*, Liss, New York, 1988, p. 45.
- [5]- K. Janssen, R.P. Mensink, F.J. Cox, J.L. Harryvan, R. Hovenier, P.C. Hollman, M.B. Katan, *Am. J. Clin. Nutr.* 67 (1998) 255.
- [6]- Ng, T.; Huang, B.; Fong, W.; Yeung, H. *Life Sci.* 1997, 61, 933-949.
- [7]- Jesu's Olivero-Verbel and Leonardo Pacheco-Londono. *J. Chem. Inf. Comput. Sci.* 2002, 42, 1241-1246.
- [8]- J. Lameira, C.N. Alves, V. Moliner, E. Silla. *European Journal of Medicinal Chemistry* 41 (2006) 616-623.
- [9]- B.R. Kowalski, *Chemometrics "Mathematics and statistics in chemistry"*, D. Reidel Publ Comp, Dordrecht, 1984.
- [10]- Matlab, Mathworks Inc., Natick, MA, 1998.
- [11]- C. Hu, K. Chen, Q. Shi, R. E. Kilkuskie, Y. Cheng, K. H. Lee, K. H. J. *Nat. Prod.* 57 (1994) 42-51.
- [12]- C. Lee, W. Yang and R. G. Parr. *Physical Review B* 37, 785 (1988).
- [13] -W. J. Hehre, W. A. Lathan, R. Ditchfield, M. D. Newton and J. A. Pople, *Gaussian 70* (Quantum Chemistry Program Exchange, Program No. 237, 1970).
 - J. S. Binkley, R. A. Whiteside, P. C. Hariharan, R. Seeger, J. A. Pople, W. J. Hehre and M. D. Newton, *Gaussian 76* (Carnegie-Mellon University, Pittsburgh, PA, 1976).
 - J. S. Binkley, R. A. Whiteside, R. Krishnan, R. Seeger, D. J. Defrees, H. B. Schlegel, S. Topiol, L. R. Kahn and J. A. Pople, *Gaussian 80* (Carnegie-Mellon Quantum Chemistry Publishing Unit, Pittsburgh, PA, 1980).
- [14]- *Gaussian 98, Revision A.9*, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, *Gaussian, Inc., Pittsburgh PA*, 1998.
- [15]- A.R. Katritzky, V.S. Lobanov, M. Karelson, *CODESSA: Reference Manual*, University of Florida, Gainesville, FL, 1994.
- [16]- Cronin, M.T.D.; Schultz, T.W. *J. Mol. Sci.* 2003, 622, 39-51.
 - Zupan, J.; Novic, M. *Analytica Chimica Acta* 1999, 388, 243-250.
- [17]- Alan R. Katritzky, Dimilar A. Dobchev, Indrek Tulp, Mati Karelson and David A. Carlson. *Bioorganic and medicinal chemistry letters* 16 (2006) 2306-2311.
- [18]- del Rio, F. J.; Riu, J.; Rius, F. X. *Anal. Chim. Acta* 2001, 446, 49-58.

- [19]- Topliss, J. G.; Edwards, R. P. J. Med. Chem. 1979, 22, 1238-1244.
- [20]- Yan, A.; Jiao, G.; Hu, Z.; Fan, B. T. Comput. Chem. 2000, 24, 171-179.
- [21]- Aoyama, T.; Suzuki, Y.; Ichikawa, H. J. Med. Chem. 1990, 33, 2583-2590.
- [22]- Zhu, J.; Chen, D.; Wu, P. Comput. Chem. 1999, 23, 97-100.
- [23]- Mghazli, S.; Jaouad, A.; M., M.; Villemin, D.; Cherqaoui, D. Chemosphere 2001, 43, 385-390.
- [24]- Douali, L.; Villemin, D.; Cherqaoui, D. Curr. Pharm. Des. 2003, 9, 1817-1826.
- [25]- Y. Belmiloud, A. Kadari, L. Benahmed, D. Charqaoui, D. Villemin and M. Brahimi. Oriental Journal of Chemistry. Vol. (25), 49-56 (2009).
- [26]- So, S.; Richards, W. G. J. Med. Chem. 1992, 35, 3201-3207.
- [27]- Andrea, T. A.; Kalayeh, H. J. Med. Chem. 1991, 34, 2824-2836.
- [28]- Cherqaoui, D., Villemin, D., J. Chem. Soc. Faraday Trans. 90 (1994) 97-102.
- [29]- R. G. Parr, R. G. Pearson, J. Am. Chem. Soc. 105 (1983) 7512-7516.
- [30]- M. R. Fensen, F. Leteurtre, S. Hiroguchi, J. Yung, K. W. Kohn, Biochem. Pharmacol. 48 (1994) 595-608.

Chapitre IV

Relations Quantitatives Structures-
Rétention Chromatographique (RQSR)
d'une série d'énantiomères.

Introduction.

L'énantiométrie est une forme d'isométrie optique où une molécule possède deux structures images l'une de l'autre par rapport à un miroir, mais non superposables, dont les propriétés physiques peuvent différer (leur pouvoir rotatoire spécifique, caractéristique de leur activité optique). Une molécule ayant deux énantiomères est dite chirale. La configuration de ces isomères est indiquée par la lettre (R ou S), (d ou l) ou (+ et -).

Certains médicaments sont actifs sous une seule forme, l'autre pouvant même être parfois nocive, d'où la nécessité de synthétiser ou de procéder à la séparation énantiosélective. La chromatographie est une des techniques de séparation qui est le plus souvent utilisée.

La chromatographie repose sur l'entraînement d'un échantillon dissous par une phase mobile à travers une phase stationnaire. Celle-ci retient plus au moins fortement les substances contenues dans l'échantillon dilué selon l'intensité des forces d'interactions de faible énergie, comme les forces de Van de Waals, les liaisons hydrogènes et les forces électrostatiques, réalisées entre les différentes espèces moléculaires et la phase stationnaire [1]. Les différents composants de l'échantillon ont généralement une vitesse caractéristique qui permet de les séparer, voire de les identifier. Cette vitesse de séparation est fortement dépendante de la nature de la phase mobile et de la phase stationnaire.

Il existe de nombreux types de chromatographie ; on peut notamment les classer selon la nature de la phase mobile :

- la chromatographie sur couche mince.
- la chromatographie en phase gazeuse (CPG).
- la chromatographie en phase liquide.
- la chromatographie en phase liquide à haute performance (CLHP).
- la chromatographie en phase supercritique.

On peut aussi les nommer selon les interactions développées par la phase stationnaire :

- la chromatographie d'adsorption/d'affinité.
- la chromatographie de partage.
- la chromatographie à échange d'ions.
- la chromatographie chirale.
- la chromatographie d'exclusion stérique.

Ou selon le support de la phase stationnaire :

- la chromatographie sur colonne (regroupant notamment CLHP et CPG) : la phase stationnaire est dans un tube étroit et la phase mobile progresse par gravité ou différence de pression.
- la chromatographie planaire : la phase stationnaire est sur la surface d'un support plat ou dans une feuille de cellulose poreuse (chromatographie papier) et la phase mobile se déplace par capillarité ou par gravité.

Les cyclodextrines sont de plus en plus utilisées dans la séparation énantiomérique, notamment en CLHP. Rappelons que les cyclodextrines ou les cycloamyloses sont des molécules cages d'origine naturelle qui permettent d'encapsuler diverses molécules. Elles ont des constitutions dimensionnelles diverses selon le nombre de cycle d'amylose constituant la couronne de la cyclodextrine. Les plus abondantes sont les hexamères (α -cyclodextrine), heptamères (β -cyclodextrine) et octamères (γ -cyclodextrine) (Figure 1) [2]. Tous les groupes polaires (hydroxyles -OH) sont localisés vers l'extérieur de la couronne. L'ensemble délimite une cavité relativement hydrophobe. Ce caractère amphiphile permet aux cyclodextrines d'inclure dans leur cavité des molécules hydrophobes pour former des complexes d'inclusion solubles dans l'eau. Leur caractère biodégradable les prédispose à des applications importantes dans divers domaines.

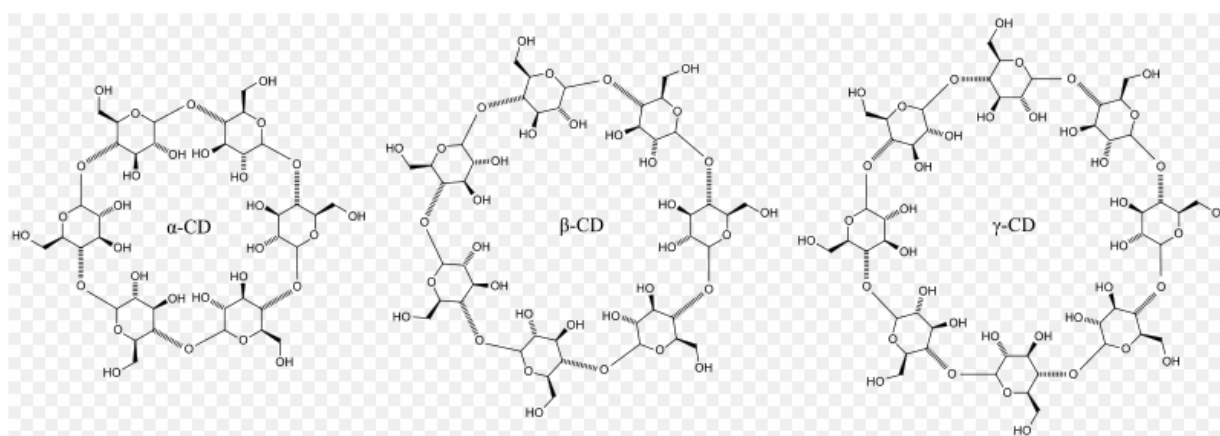


Figure 1: Structure de l' α -, β - et γ -cyclodextrines.

En chromatographie, les cyclodextrines sont utilisées comme greffes sur la phase stationnaire (silice en général) [1]. L'emploi de cyclodextrine permet de changer l'affinité des composés à analyser pour la phase stationnaire, donc d'améliorer leur temps de rétention. Actuellement des centaines de phases stationnaires chirales, à base de cyclodextrine, sont disponibles pour la séparation des énantiomères. Le développement de

nouvelles phases stationnaires reste un challenge de la recherche scientifique dans la séparation énantiosélective des médicaments.

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

Pour cette raison, l'industrie pharmaceutique s'oriente vers de nouvelles méthodes de recherche, qui consistent à prédire les propriétés, activités et rétentions de molécules avant même que celles-ci ne soient synthétisées. Plusieurs disciplines de la chimie computationnelle comme les relations, structure-activité ou RQSA (voir détails au chapitre II), qui se sont développées en réponse à ce besoin.

Dans les RQSA, le terme activité, peut désigner la rétention chromatographique, ce qui conduit à l'appellation ' Relations Quantitatives Structure – Rétention Chromatographique', (RQSR). L'utilisation de l'outil informatique dans la séparation chromatographique par les RQSR est devenue de plus en plus fréquente ces dernières années. Cette méthode a pour but d'établir des relations entre les indices de rétention d'une série d'énantiomères, d'une part et un ou plusieurs paramètres appelés descripteurs, d'autre part [3].

Généralement, l'équation de corrélation est développée à partir de données expérimentales (indices de rétention) et les descripteurs d'une série de composés qui peuvent être évalués expérimentalement ou déterminés par le calcul. Cette équation de régression sera utilisée pour prévoir les indices de rétention de nouveaux couples d'énantiomères. La recherche pharmacologique, au vu du nombre de travaux dévoués à la détermination des RQSR, a parfaitement réalisé l'intérêt scientifique et économique de cette démarche. En effet, les RQSR permettent [4]:

- de prédire l'indice de rétention des nouveaux couples non testés sur la même phase stationnaire.

- De mieux comprendre le mécanisme des séparations chromatographiques sur une phase stationnaire.
- prédire des grandeurs de rétention, et caractérisation des propriétés physiques des solutés.
- Et éventuellement de minimiser le coût de fabrication de nouvelles phases stationnaires.

Le but de ce travail consiste à modéliser la séparation, d'une série de molécules chirales, par la β - cyclodextrine. Dans une première étape, nous avons corrélié les indices de rétention, pris de la littérature [5] pour une série de couples d'énantiomères de masse molaire différente, avec des descripteurs moléculaires. On a construit plusieurs modèles linéaires que nous avons testé par la suite sur de nouvelles molécules chirales.

Définitions des différentes grandeurs de rétention:

La chromatographie permet aussi l'identification et le dosage de composés dans un mélange. Elle peut être couplée à un spectromètre de masse pour l'identification de composés inconnus. Pour l'exploiter pleinement il est important de connaître les différentes grandeurs de rétention.

a- Temps de rétention

C'est le temps que met le soluté à sortir de la colonne c'est à dire le temps écoulé entre l'injection et le maximum du pic du composé élué. C'est la principale **grandeur de rétention**. Il est noté t_R (en minutes)

Il varie en fonction du débit, de la température d'élution, de la composition de la phase mobile et du vieillissement de la colonne.

b- Temps mort

Il est noté t_M ou t_0 (en minutes).

C'est le temps que met un soluté non retenu (l'air par exemple dans le cas de la CPG) à sortir de la colonne.

En CLHP, c'est également le temps mis par la phase mobile (l'éluant) pour traverser la colonne.

c- Temps de rétention réduit

On ne prend plus comme origine l'injection mais le pic du temps mort.

Il est noté t'_R et est égal à $t'_R = t_R - t_M$

d- Volume de rétention

Au temps de rétention t_R d'un soluté correspond un volume de rétention V_R qui correspond au volume de phase mobile nécessaire à l'élution.

$$V_R = t_R \times D$$

Avec D le débit du gaz vecteur

e- Temps de rétention réduit

On ne prend plus comme origine l'injection mais le pic du temps mort.

Il est noté t'_R et est égal à $t'_R = t_R - t_M$

Il correspond au temps de séjour du composé en contact avec la phase stationnaire.

f- Volume de rétention réduit

Comme pour le temps mort réduit, on définit le volume mort

Il est noté V'_R et est égal à $t'_R = V_R - V_M$

g- Facteur de rétention k'

Le facteur de rétention caractérise la rétention d'un composé. C'est un paramètre important indépendant du débit.

Si pour 2 analyses identiques (même composé, même phase mobile):

- on a 2 valeurs de t_R identiques et 2 valeurs de k' différentes, cela signifie qu'il y a une fuite de phase mobile avant la colonne
- 2 valeurs de t_R et de k différentes indiquent des problèmes au niveau de la phase mobile ou de la phase stationnaire.

Le facteur de rétention est défini suivant la relation : $k' = (t_R - t_M)/t_M$

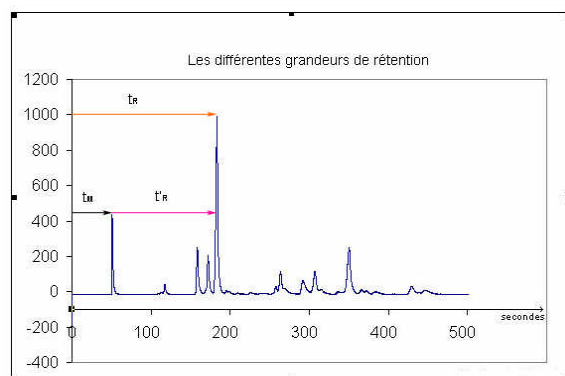


Figure 2 : Les différentes grandeurs de rétention

h- Coefficient de distribution K

$$K = C_s/C_m$$

Avec C_s la concentration du soluté dans la phase stationnaire

Avec C_m la concentration du soluté dans la phase mobile

i- Facteur de séparation, de rétention ou de sélectivité.

Il définit la position relative de 2 pics, le pic du soluté a sortant avant le pic du soluté b:

Il est égal à : $\alpha = k'_a/k'_b$

j- Facteur de résolution

Il définit la plus ou moins bonne séparation de 2 solutés :

$$R_s = 2(t_{R(a)} - t_{R(b)}) / (w_a + w_b)$$

Pour des valeurs de R_s supérieures à 1.5, les pics sont correctement séparés. On peut optimiser le facteur de résolution en modifiant les débits, les températures ou la composition de la phase mobile.

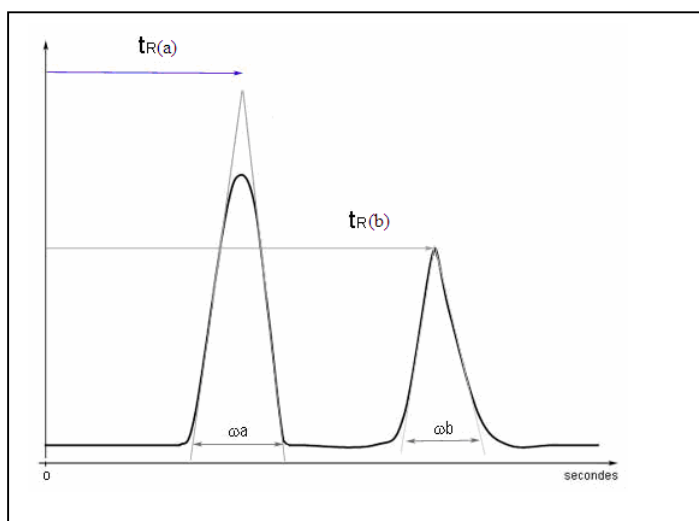


Figure 3 : Les grandeurs de rétention.

L'efficacité (N) :

Exprimé en nombre de plateaux théoriques, valeur calculée à partir du temps de rétention et de la largeur du pic (w) ou la largeur à mi-hauteur :

$$N = 16(t_R / w)^2 = 5.54 (t_R / w)^2$$

N : n'est pas une constante caractéristique de la colonne, mais n'est valable que pour un solvant et (ou un groupe) de composés donné.

En combinant les équations de N, α et R on aura : $R = f(N, \alpha, k')$.

Pour séparer deux composés il faut jouer :

1- utiliser des colonnes plus performantes.

2- des solvant plus éluants (k').

3- choisir avec discernement une colonne et un solvant appropriés c'est-à-dire jouer sur la sélectivité α (c'est le point le plus important et c'est là qu'intervient l'expérience de l'utilisateur).

1- Rappel de la partie expérimentale.

- Synthèse des différentes phases stationnaires chirales.

F. Tazerouti et coll. [5] se sont intéressés à la synthèse de six nouvelles phases stationnaires chirales à base de la β -cyclodextrine. La synthèse consiste à substituer les groupements hydroxyles primaires de la cyclodextrine par un spacer difonctionnel (isocyanatopropyltriéthoxysilane), à dériver les -OH libres restants en groupements phényle carbamate et enfin à greffer la cyclodextrine dérivée à la surface de la silice par une réaction de silanisation pour l'obtention de la première phase stationnaire notée A (PSC A). Concernant la deuxième phase stationnaire notée B (PSC B), ces auteurs ont suivi le même schéma de synthèse que précédemment. En utilisant cette fois ci l'hydroxypropyl- β -cyclodextrine (2-HP- β -CD) au lieu de la β -cyclodextrine.

La figure 4 résume le schéma de la synthèse des phases stationnaires A et B.

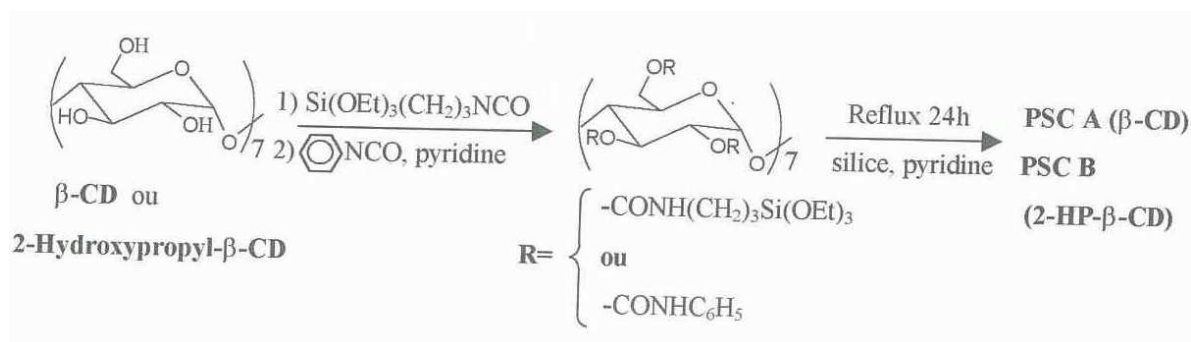


Figure 4 : Schéma de synthèse des phases stationnaires chirales A et B [5].

Concernant la phase stationnaire C (PSC C), ces auteurs [6] ont réalisé la même synthèse en améliorant la technique de séparation en utilisant la procédure « one-pot ».

Les phases stationnaires D, E et F (PSC D, E et F) ont été aussi synthétisées par ces auteurs [5] selon le schéma présenté sur la figure 5. Dans ce cas les dérivés de la cyclodextrine sont obtenus par la réaction successive de la cyclodextrine avec un chlorure d'acide (chlorure

d'undecenoyl) et un dérivé de phényle isocyanate (phényle isocyanate, p-méthyle ou p-chloro-phényle isocyanate). Ces dérivés ainsi préparés sont fixés à la silice allylique et ceci par une réaction de greffage radicalaire, en présence de l' α , α' -azobisisobutyronitrile (AIBN), mettant en jeu la double liaison C=C du groupement 10-undecenoyl préalablement lié à la cyclodextrine.

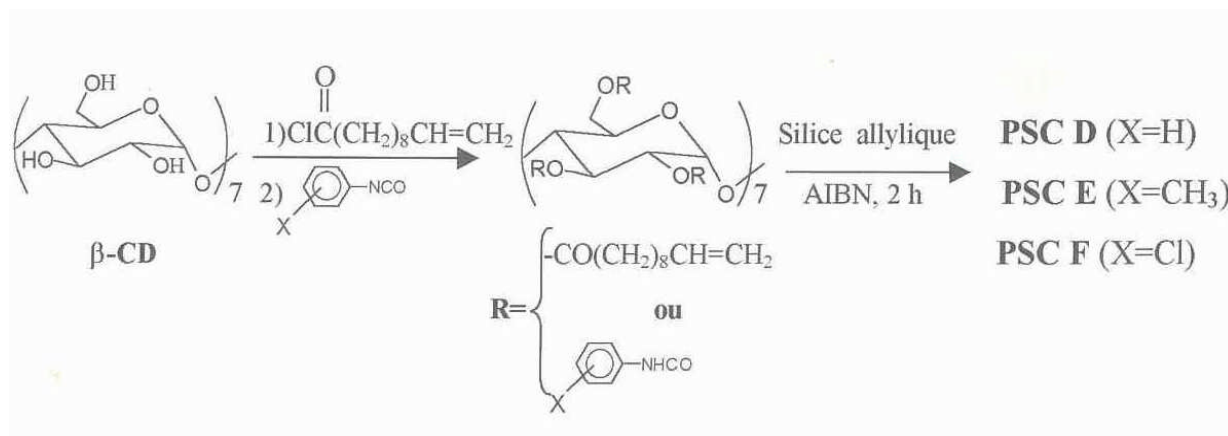


Figure 5 : Schéma de synthèse des phases stationnaires chirales D, E et F [5].

Les six phases stationnaires synthétisées ont été ensuite testées en mode de phase normal en utilisant aussi bien le mélange binaire heptane/2-propanol que le mélange heptane/chloroforme. Trente deux solutés ont été testés sur ces phases stationnaires chirales, nous trouvons parmi lesquels divers produits pharmaceutiques tels que les benzodiazépines, les agents β -bloquants adrénergiques et des anti-inflammatoires. La figure 6 rassemble les différentes structures des solutés racémiques testés. La grandeur de rétention donnée dans cette expérience est exprimée par le « facteur de rétention k' ». Le tableau 1 regroupe les différents résultats trouvés par cette équipe [5].

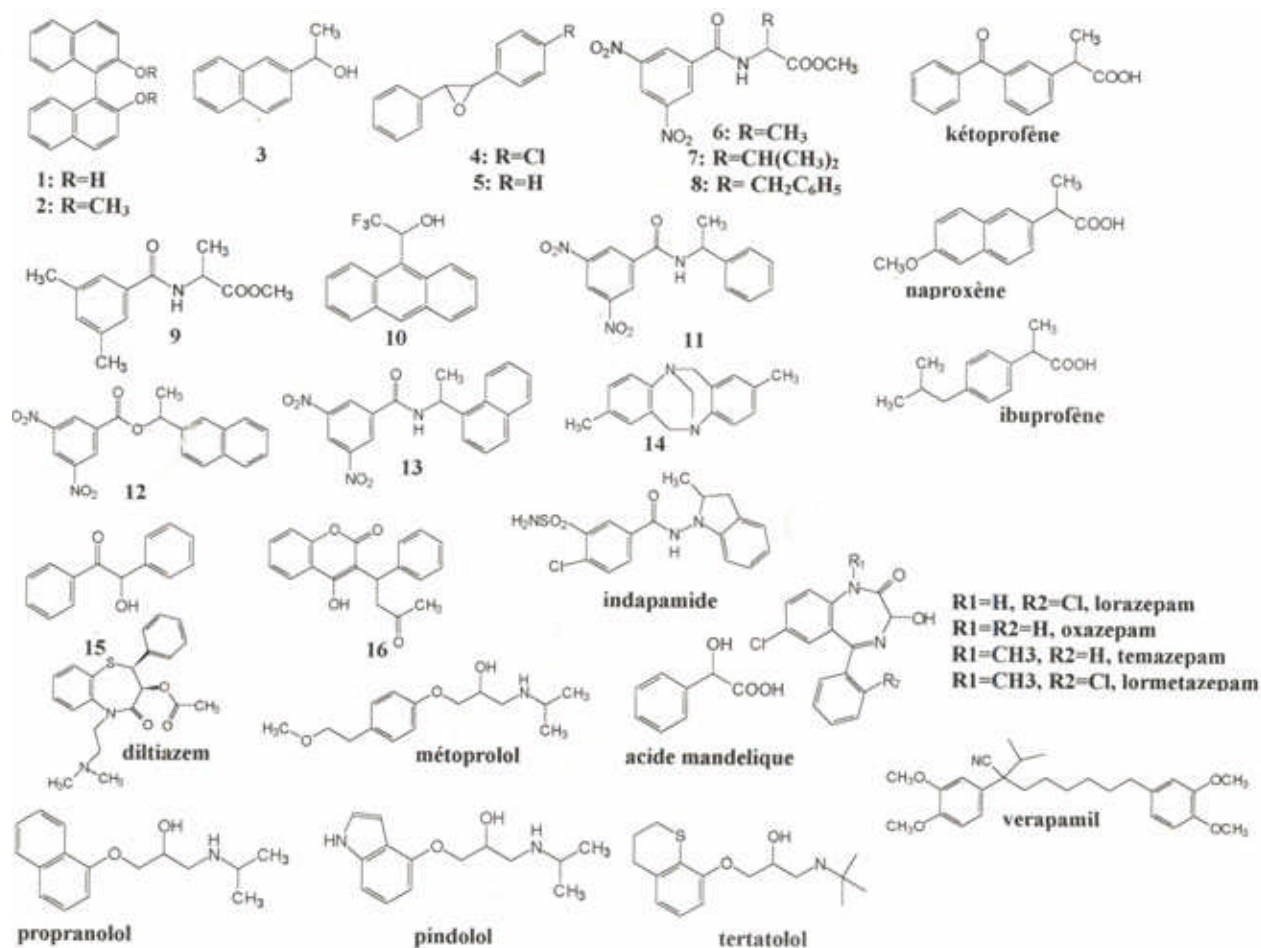


Tableau 1 : Séparation des solutés racémiques et des produits pharmaceutiques optiquement actifs sur les 4 phases stationnaires chirales C, D, E et F en utilisant le mélange binaire heptane/2-propanol pris de la référence [5].

Produit	Phase mobile	PSC C			PSC D			PSC E			PSC F		
	Hept/IPA	K'1	k'2	α	k'1	k'2	α	K'1	k'2	A	k'1	k'2	α
1 ^a	80/20	0.46	-	-	0.21	-	-	0.46	-	-	0.42	-	-
	90/10	0.81	0.88	1.08	0.36	-	-	0.39	-	-	0.32	-	-
	98/2	3.38	4.01	1.19	1.62	1.77	1.09	1.33	1.49	1.12	1.3	1.38	1.06
2 ^a	98/2	0.53	0.72	1.35	-	-	-	-	-	-	-	-	-
3 ^a	80/20	0.35	-	-	0.19	-	-	0.47	-	-	0.41	-	-
	90/10	0.58	-	-	0.36	-	-	0.56	-	-	0.33	-	-
	98/2	2.08	2.16	1.04	1.72	-	-	1.64	-	-	1.38	-	-
4 ^a	80/20	0.19	0.31	1.62	0.11	-	-	0.13	0.18	1.38	0.08	-	-
	90/10	0.2	0.3	1.49	0.17	-	-	0.14	0.2	1.43	0.08	-	-
	98/2	0.25	0.35	1.39	0.17	-	-	0.15	0.18	1.23	0.19	-	-
5 ^a	80/20	0.2	-	-	0.1	-	-	0.10	-	-	0.09	-	-
	90/10	1.42	-	-	0.1	-	-	0.10	-	-	0.15	-	-
	98/2	0.17	-	-	0.11	-	-	0.35	-	-	0.15	-	-
6 ^a	80/20	1.89(d)	2.07	1.09	0.66(d)	0.79	1.20	2.53(d)	4.04	1.59	1.21(d)	1.44	1.20
	90/10	4.05	3.24	1.05	1.41	1.75	1.24	2.66	4.69	1.76	2.76(d)	3.12	1.13
	98/2	-	-	-	9.8	12.4	1.26	-	-	-	8.76(d)	10.0	1.14
7 ^a	80/20	1.4	1.56	1.11	0.45	0.57	1.27	1.35	1.59	1.18	0.56	0.66	1.18
	90/10	2.73	-	-	0.82	1.1	1.34	1.28	2.00	1.57	0.82	0.98	1.20
	98/2	10.32	10.8	10.04	3.54	4.75	1.34	4.81	6.97	1.45	2.14	2.36	1.10
8 ^a	80/20	2.02(s)	2.74	1.35	0.83	-	-	3.28(s)	4.72	1.44	0.92	-	-
	90/10	4.01	6.49	1.62	1.77	-	-	3.87	5.46	1.41	4.18	-	-
	98/2	-	-	-	9.37(s)	12.3	1.31	12.94	22.2	1.71	4.59	-	-
9 ^a	80/20	0.84	-	-	-	-	-	-	-	-	-	-	-
	90/10	0.61	-	-	-	-	-	-	-	-	-	-	-
	98/2	5.37	5.73	1.07	-	-	-	-	-	-	-	-	-
10 ^a	90/10	1.54	1.75	1.13	-	-	-	-	-	-	-	-	-
11 ^a	80/20	1.28	1.5	1.17	0.56	-	-	1.88	32.7	1.45	1.10	1.43	1.30
	90/10	0.03	3.67	1.21	1.22	1.38	1.13	3.00	4.27	1.42	1.29	1.7	1.32
	98/2	19.74	24.9	1.26	8.48	11.3	1.33	12.15	19.9	1.63	4.10	5.45	1.33
12 ^a	80/20	0.68	0.75	1.09	-	-	-	0.51	-	-	0.24	-	-
	90/10	0.85	0.92	1.09	-	-	-	0.53	-	-	0.20	-	-
	98/2	1.43	1.56	1.09	0.49	-	-	0.74	0.78	1.06	0.33	-	-
13 ^a	80/20	1.24	1.5	1.21	0.46(s)	0.52	1.13	1.93	2.28	1.18	1.05	1.42	1.35
	90/10	2.82	3.34	1.18	1.03	1.18	1.14	2.33	2.75	1.18	0.65	0.88	1.36
	98/2	-	-	-	6.14	7.17	1.17	8.19	9.83	1.2	3.82	5.04	1.32
14 ^a	80/20	0.94	1.09	1.16	0.93	-	-	1.64	-	-	1.41	-	-
	90/10	1.15	1.3	1.14	1.42	-	-	1.96	-	-	1.34	-	-
	98/2	2.63	-	-	3.93	-	-	3.42	-	-	3.45	-	-
15 ^a	80/20	0.53	-	-	0.27	-	-	0.48	-	-	0.43	-	-
	90/10	0.77	-	-	0.41	-	-	0.54	-	-	0.65	-	-
	98/2	1.61	-	-	0.99	-	-	0.91	-	-	0.89	-	-
16 ^a	80/20	0.93	-	-	0.75	-	-	1.49	-	-	3.87	-	-
	90/10	2.29	-	-	1.75	-	-	1.86	-	-	4.04	-	-
	98/2	1.85	-	-	9.04	-	-	7.14	-	-	13.27	-	-
Lorazepam ^a	80/20	1.84	2.09	1.13	0.9	-	-	2.82	3.28	1.16	3.51	-	-
	90/10	4.45	5.22	1.17	2.2	-	-	4.47	6.24	1.39	4.26	-	-
	98/2	-	-	-	24.85	-	-	-	-	-	-	-	-
Lormeta-zepam ^a	80/20	2.88	-	-	1.39	-	-	3.74	-	-	3.16	-	-
	90/10	5.92	-	-	2.81	-	-	4.89	-	-	2.55	-	-
Oxazepam ^a	80/20	1.73	1.98	1.14	0.87	-	-	2.59	2.93	1.13	2.43	-	-
	90/10	4.13	4.97	1.2	2.14	-	-	4.15	4.88	1.17	4.00	-	-
Temazepam ^a	80/20	2.91	3.19	1.09	1.38	-	-	3.53	-	-	3.11	-	-
	90/10	5.92	6.35	1.07	2.99	-	-	5.07	-	-	4.36	-	-

Acide Mandelique ^b	98/2/0.5	6.44	6.95	1.08	-	-	-	-	-	-	-	-	-
Ibuprofène ^b	98/2/0.5	0.67	-	-	0.2	-	-	0.11	0.17	1.57	0.24	0.32	1.34
Ketoprofène ^b	98/2/0.5	2.76	-	-	0.77	-	-	1.62	-	-	0.29	0.36	1.26
Naproxène ^b	98/2/0.5	1.81	-	-	0.09	0.17	1.97	0.28	0.64	2.29	0.47	0.58	1.24
Flurbi- profène ^b	98/2/0.5	1.03	-	-	-	-	-	-	-	-	-	-	-
Indapamide ^c	80/10/0.1	2.68	3.03	1.12	1.55	-	-	2.08	-	-	3.83	-	-
	90/10/0.1	10.28	12.3	1.19	8.08	8.89	1.1	7.39	-	-	17.48	-	-
Diltiazem ^c	90/10/0.1	4.59	5.0	1.09	-	-	-	-	-	-	-	-	-
Metoprolol ^c	80/20/0.1	2.54	3.23	1.27	-	-	-	-	-	-	-	-	-
	90/10/0.1	5.98	7.43	1.24	-	-	-	-	-	-	-	-	-
Propranolol ^c	80/20/0.1	1.73	1.82	1.05	1.74	-	-	2.04	-	-	1.60	-	-
	90/10/0.1	3.83	4.03	1.05	5.43	-	-	4.01	-	-	7.90	-	-
Pindolol ^c	80/20/0.1	3.82	4.18	1.09	-	-	-	-	-	-	-	-	-
	90/10/0.1	12.94	14.1	1.09	-	-	-	-	-	-	-	-	-
Tertatolol ^c	90/10/0.1	4.34	-	-	3.4	-	-	4.25	-	-	7.67	-	-
Verapamil ^c	90/10/0.1	3.88	-	-	4.21	-	-	3.44	-	-	5.48	-	-
^a : hept/IPA ; 1ml/mn, l=254nm ; ^b :hept/IPA/TFA, 0.5ml/mn, 210nm; ^c :hept/IPA/DEA, 1ml/mn, 230nm hept : heptane ; IPA : 2-propanol ; TFA : acide trifluoroacétique ; DEA : diéthylamine													

2- Recherche de Relations Quantitatives Structure – Rétention Chromatographique des solutés racémiques et des produits pharmaceutiques optiquement actifs séparés par les phases stationnaires chirales A, B, C, D, E et F.

Les structures des différents composés ont été dessinées en utilisant le programme ChemOffice [7]. Ces structures ont été transférées ensuite vers le programme Semi-chem [8], l'utilisation en parallèle du programme Gaussian 98W [9] a servi pour l'optimisation des coordonnées internes des différents composés au moyen de la méthode AM1 [10]. Les différents résultats géométriques trouvés ont été transférés dans le programme CODESSA [11] pour le calcul des différents descripteurs.

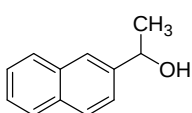
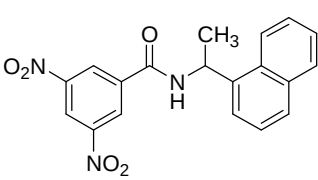
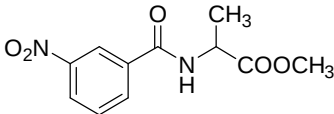
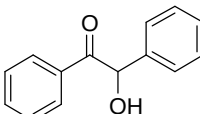
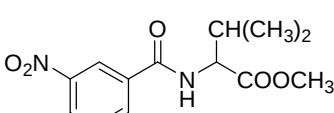
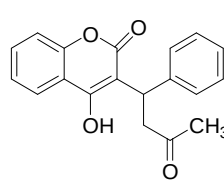
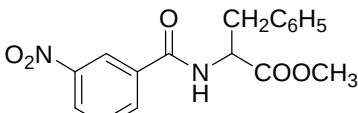
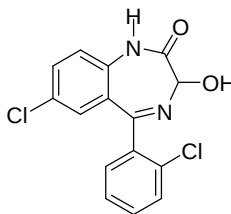
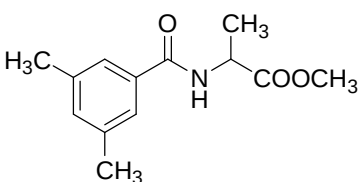
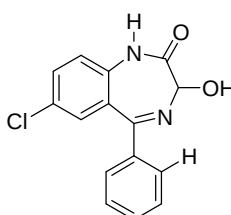
2.2 Base de données

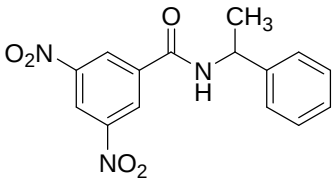
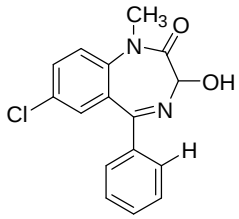
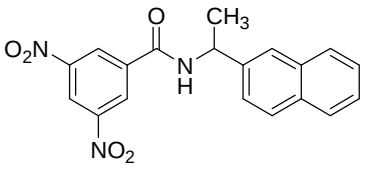
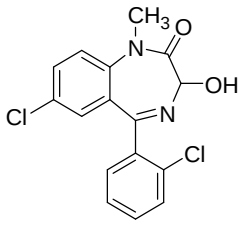
Expérimentalement la séparation des solutés racémiques et des produits pharmaceutiques optiquement actifs a été faite sur les 4 phases stationnaires chirales C, D, E et F en utilisant le mélange binaire heptane/2-propanol [5]. En effet, ces expérimentateurs ont constatés que sur les phases A et B, seuls quelques racémiques ont été séparés. Par contre sur les quatre phases C, D, E et F, testées en mode normal, les résultats obtenus montre que la phase stationnaire C présente une meilleure reconnaissance chirale par rapport aux autres. Ce qui nous incité à utiliser la phase stationnaire C dans notre modélisation. Les auteurs ont fait la séparation de 32 séries d'énantiomères avec des phases mobiles différentes dans ses proportions du mélange binaire heptane/2-propanol. Sur ces 32 solutés racémiques, seuls 14 ont fait objet de cette présente étude théorique. Leur sélection est basée sur le fait que la

molécule doit avoir un seul carbone asymétrique et transporté par le même type de phase mobile, à savoir les proportions heptane/2-propanole 80/20.

La structure des différents composés sélectionnés ainsi que leurs indices de rétentions expérimentaux sont regroupés sur le tableau 2.

Tableau 2: Les différents composés sélectionnés pour l'étude statistique.

Composé	K ₁	Composé	K ₁
1 	0.35	8 	1.24
2 	1.89	9 	0.53
3 	1.40	10 	0.93
4 	2.02	11 	1.84
5 	0.84	12 	1.73

6 	1.28	13 	2.91
7 	0.68	14 	2.88

2.3 Calculs des descripteurs

Par le biais de la fonction d'onde, la géométrie obtenue pour chaque molécule est utilisée pour le calcul des différents descripteurs en utilisant le programme CODESSA [11] qui calcule plus de 500 descripteurs de différentes classe à savoir, des descripteurs de constitution, topologiques, électrostatiques, thermodynamiques et géométriques. Rappelons que le logarithme du coefficient de partage dans le mélange n-octanol/eau (Log P) a été calculé en utilisant le programme Chem 3D [7].

2.4 La détermination du nombre Maximum de descripteurs

Pour la détermination du nombre maximum de descripteur, la méthode du 'breaking point' a été employée [12]. La variation du coefficient de corrélation R^2 en fonction du nombre de descripteurs a été reportée sur la Figure 7. L'analyse de cette courbe montre que la coupure (intersection de deux lignes) pour le coefficient de la corrélation est effectuée pour $n = 4$. Par conséquent, le nombre maximum de descripteurs sera de quatre pour toute la suite de ce travail.

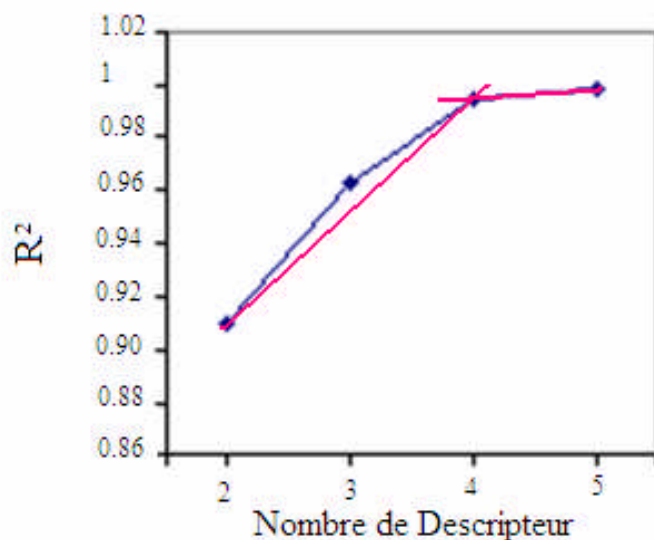


Figure 7: La variation de R^2 en fonction du nombre de descripteur.

2.5 Application de la Régression Multi linéaire.

2.5.1 Méthodologie Utilisée.

Notre base de donnée a été divisée en deux, on a pris 12 molécules pour la phase d'apprentissage et 2 molécules test. Les 12 molécules constituant le lot d'apprentissage ont été triés aléatoirement des 14 molécules de manière à construire plusieurs modèles. Cette méthode appelée « méthode de bloc » (Figure 8) [13] est utilisée de manière à faire passer toutes les molécules dans les deux phases et ceci dans le but de choisir le meilleur modèle de point de vue apprentissage et prédiction.

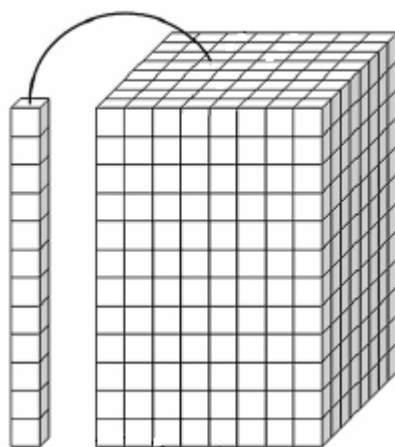


Figure 8 : Schématisation de la construction des modèles par bloc.

Les structures des molécules ont toutes été au préalable optimisées au moyen de la méthode semi empirique AM1 implémentée dans le programme Gaussian 98W en prenant une configuration (R) pour tous les carbones asymétriques. Ces calculs donnent toutes les grandeurs électroniques et thermodynamiques qui serviront à CODESSA pour faire ressortir les différents descripteurs susceptibles d'être corrélés aux indices de rétention.

Pour chaque modèle construit, l'heuristique dans CODESSA donne les dix meilleurs modèles de corrélations. On choisira le meilleur d'entre eux, d'un point de vue statistique (R^2 , F et S). Au total sept modèles de régression multi linéaire ont été construits. On choisira ensuite le meilleur modèle selon la bonne prédiction des molécules tests.

2.5.2 Résultats et Discussions.

Nous donnons ci-dessous les résultats des sept modèles obtenus lors de notre recherche de régressions. Pour chaque modèle nous donnons les différents descripteurs retenus ainsi que les paramètres statistiques : (R^2) le coefficient de corrélation. (F) le taux de confiance de la régression et (S) l'écart type de la régression. Ces différents résultats sont rassemblés dans le Tableau 3.

Table 3 : Les différents modèles obtenus ainsi que leurs descripteurs avec les paramètres statistiques.

MODELE	PARAMÈTRE STATISTIQUE			DESCRIPTEURS
	R^2	F	S	
Mod1	0.9969	559.92	0.0552	min(#HA, #HD) (MOPAC PC) HA dependent HDSA-1 (Zefirov) HOMO - LUMO energy gap Min partial charge for atom H (Zefirov)
Mod2	0.9971	611.93	0.0542	HA dependent HDSA-2 (MOPAC PC) min(#HA, #HD) (MOPAC PC) (all) Max electroph. react. index for atom O Moments of inertia B
Mod3	0.9971	611.93	0.0542	HA dependent HDSA-2 (MOPAC PC) min(#HA, #HD) (MOPAC PC) (all) Max electroph. react. index for atom O Moments of inertia B
Mod4	0.9989	1556.28	0.0353	Max SIGMA-SIGMA bond order HA dependent HDSA-2/TMSA Max atomic state energy for atom H Min total interaction for bond C-O

Chapitre IV : Etude RQSR d'une série d'énantiomère

Mod5	0.9929	245.08	0.0769	min(#HA, #HD) (MOPAC PC) HA dependent HDSA-1 (Zefirov) Max e-e repulsion for bond C-C. Min partial charge for atoms for atom H (Zefirov)
Mod6	0.9932	256.01	0.0891	min(#HA, #HD) (MOPAC PC) HA dependent HDSA-1 (Zefirov) Min partial charge for atoms for atom H (Zefirov) Max e-n attraction for bond C-C
Mod7	0.9945	316.63	0.0660	DPSA-1 Difference in CPSAs Min exchange energy for bond H-C HACA-1 (MOPAC PC) Min coulombic interaction for bond C-O.

Il ressort de ces résultats que le modèle (Mod4) semble être le meilleur d'un point de vue statistique pour la phase d'apprentissage. Notons par ailleurs que les valeurs des coefficients de corrélation sont importantes et les déviations standard de l'erreur d'apprentissage sont faibles pour tous les modèles.

Il est important de signaler que pratiquement dans tous les modèles de régression construite, ressort des descripteurs communs à savoir la formation de liaisons hydrogènes et les interactions électrostatiques ce qui pourrait signifier que ces descripteurs ont un effet prépondérant dans l'explication du mécanisme de rétention de ces énantiomères [14, 15].

Les différences (valeur calculée – valeur expérimentale) des indices calculés par les différents modèles de régressions (en noir dans le tableau), par rapport à l'expérience, sont regroupées dans le tableau 4 ci dessous.

Tableau 4: Les différences (valeur calculée – valeur expérimentale) des indices calculés pour les sept modèles construits. En noir c'est pour la phase d'apprentissage et en rouge pour les molécules test.

COMPOSE	EXP	MOD1	MOD2	MOD3	MOD4	MOD5	MOD6	MOD7
Mol 01	0.35	<u>0.205</u>	-0.002	0.005	0.017	0.067	0.024	0.021
Mol 02	1.89	0.017	<u>0.634</u>	-0.041	0.030	0.072	0.018	-0.118
Mol 03	1.4	-0.049	-0.045	<u>-0.754</u>	-0.036	-0.039	0.037	0.019
Mol 04	2.02	-0.039	-0.000	0.047	<u>-0.446</u>	-0.061	0.001	0.042
Mol 05	0.84	0.062	0.057	0.021	-0.019	<u>-0.255</u>	-0.077	0.011
Mol 06	1.28	-0.015	-0.020	-0.039	0.018	0.010	<u>0.155</u>	0.058
Mol 07	0.68	0.034	-0.018	0.011	0.043	-0.020	-0.033	<u>-1.295</u>
Mol 08	1.24	<u>-0.132</u>	0.037	0.001	-0.016	0.017	0.040	-0.049
Mol 09	0.53	-0.033	<u>2.624</u>	0.020	-0.033	-0.137	-0.110	-0.008
Mol 10	0.93	0.024	0.024	<u>0.478</u>	-0.024	0.076	0.146	-0.013
Mol 11	1.84	-0.021	-0.094	0.000	<u>0.127</u>	0.016	-0.055	0.039
Mol 12	2.88	0.082	0.058	0.016	-0.018	<u>0.157</u>	0.066	0.053
Mol 13	1.73	-0.049	-0.000	-0.072	0.033	-0.012	<u>-0.065</u>	-0.056
Mol 14	2.91	-0.012	0.004	0.027	0.004	0.011	-0.059	<u>1.742</u>

Nous constatons pour la phase d'apprentissage que les différences sont comprises entre [0.015- 0.082], [0 - 0.094], [0.001 - 0.072], [0.004 - 0.043], [0.011- 0.137], [0.001 – 0.146] et [0.011 - 0.118] respectivement pour les modèles 1, 2, 3, 4, 5, 6 et 7.

La figure 9 donne la corrélation entre les valeurs expérimentale et les valeurs théorique. Pratiquement tous les nombres d'observations se trouvent sur la droite d'équation $Y = X$.

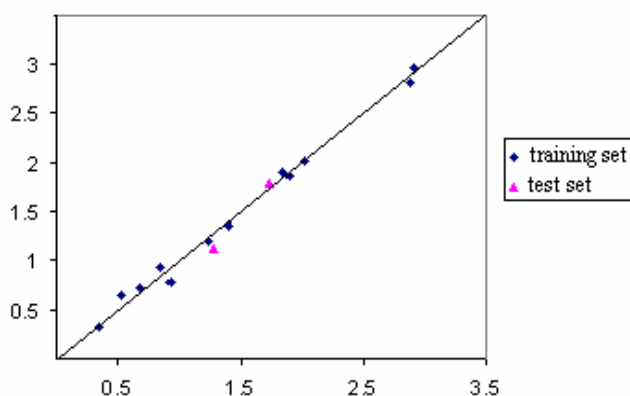


Figure 9 : Corrélation entre les valeurs des activités expérimentales et celles calculées par la RML.

La dernière étape des RQSR consiste à tester le pouvoir de prédiction des différents modèles établis. Ce test de validation nous permet d'accepter ou de rejeter le meilleur modèle proposé dans ce présent travail.

L'introduction des molécules tests conduit aux différences portées en rouge dans le tableau. Nous remarquons que, dans ce cas, c'est le modèle (Mod6) qui donne les plus petites différences. La corrélation entre les valeurs prédites et les valeurs expérimentales des molécules test sont reportées sur la Figure 8.

Nous donnons ci-dessous l'équation retenue de l'Indice de Rétention K (IRK) pour ce modèle:

$$\text{IRK} = 0.454 \text{ HAHD} + 0.0147 \text{ HDSA} + 116.457 \text{ MPC(H)} + 0.147 \text{ Matt(e-n) C-C.}$$

$$n = 12, R^2 = 0.9932 \text{ F} = 256.01 \text{ S} = 0.0891, R^2_{\text{CV}} = 0.9809$$

Cette équation se caractérise par un coefficient de corrélation et un taux de confiance élevés, la déviation standard de l'erreur est faible et la validation croisée donne une bonne estimation du pouvoir prédictif du modèle.

Les descripteurs retenus dans ce modèle sont:

HAHD : min(#HA, #HD) (MOPAC PC)

HDSA : HA dependent HDSA-1.

MPC(H) : Min partial charge for atoms for atom H.

Matt(e-n) C-C: Max e-n attraction for bond C-C.

Le choix restreint de la base de données est dû au fait que les résultats expérimentaux disponibles sur la phase stationnaire C synthétisé par F. Tazerouti et col. [5] soient limités, ce qui explique notre choix qui s'est limité à deux molécules testées uniquement. La perspective d'augmenter la taille de cette base de données pourrait être judicieuse pour confirmer les modèles de régression établis et prévoir la possibilité de trouver d'autres régressions multi variables et non linéaires à qui nous pourrions appliquer les réseaux de neurones artificiels, qui seraient à notre avis porteur.

Les modèles RQSR reportés ci-dessus renferment d'importantes informations sur les mécanismes de rétention des énantiomères lors de leurs séparations par la phase C. Les descripteurs sélectionnés par le modèle 6 indiquent que la rétention devrait dépendre essentiellement des forces électrostatiques et de la formation de liaisons hydrogènes [5, 6, 14, 15] et un descripteur quantique. En effet, le descripteur électrostatique HAHD : min(#HA, #HD) traduit les forces de liaison d'hydrogène de la molécule. Il décrit la force minimale de la liaison hydrogène entre le site Donneur et le site Accepteur. Plus ce minimum est important

(faible force) plus le facteur de rétention de l'énantiomère sera plus petit [16]. Ce résultat est en accord avec le fait que l'énantiomère doit sortir en dernier (facteur de rétention élevé) quand les liaisons hydrogène formées sont fortes. Le deuxième descripteur qui sort dans le modèle est : HA dependent HDSA-1. Ce dernier explique les forces des liaisons hydrogènes de la molécule, plus sa valeur est grande plus le composé est retenu [17]. Le troisième descripteur correspond au : 'Min partial charge for atoms H'. Ces résultats montrent que les liaisons hydrogènes jouent un rôle important dans la séparation énantiosélective.

Le dernier descripteur qui est un descripteur quantique correspond au 'Max e-n attraction for bond C-C'. Ce dernier décrit le processus d'attraction des molécules par les interactions électron-noyau qui joue un rôle important dans la stabilité du complexe [18].

Notons que le choix de ces descripteurs ne peut pas être unique si on étend la base de données d'une part et si l'on tient compte aussi du fait que la régression peut aussi être multi variable non linéaire.

Conclusion.

Le choix restreint de la base de données qui est dû au fait que les résultats expérimentaux disponibles sur les différentes phases stationnaires synthétisées par F. Tazerouti et col. soient limités, nous avons fait une tentative en utilisant les régressions multi variables et multi linéaire avec une base constituée de 12 molécules plus deux autres qui seront utilisées dans la validité du modèle théorique. Les résultats obtenus nous permettent de mieux comprendre le mécanisme des séparations chromatographiques sur la phase stationnaire C. La perspective d'augmenter la taille de cette base de données pourrait être judicieuse. En effet, La taille de la base de données influe sur la fiabilité des modèles établis. Un modèle contenant peu de prototypes donne de bons résultats statistiques. Il faut cependant noter que plus l'échantillon est réduit et plus il risque de définir une population limitée. C'est alors la portée de la conclusion qui en souffre. Il serait donc souhaitable de travailler avec une base de données de taille importante.

REFERENCE.

- [1]- J. Sherma, B. Fried, Handbook –Marcel Dekker- INC, New York. 1989.
- [2]- Pyka, A. J. Planar Chromatogr. 1990, 3, 425-428.
- [3]- N. Dimov, NIHFI, Quantitative Structure Retention Relationship by TLC. 1400-1411.
- [4]- Qing-You Zhang , Joao Aires de sousa. J. Chem. Inf. Model. 2006, 46, 2278-2287.
- [5]- F. Tazerouti. « Synthèse, caractérisation et application des phases stationnaires chirales en chromatographie liquide ».Thèse de Doctorat d'Etat, Faculté de chimie, USTHB. 2001.
- K. Si-Ahmed, F. Tazerouti, A. Y. Badjah-Hadj-Ahmed, Z. Aturki, G. D'Orazio, A. Rocco, S. Fanali, J. Chromtogr. 2009, 07, 053.
- [6]- F. Tazerouti, A. Y ; Badjah-Hadj-ahmed, B. Y. Meklati, P. Franco et C. Minguillon. Chirality 13, 2001
- [7]- CS Chem Office, Ultra. For Microsoft Windows, Cambridge Scientific Computing Inc., 875 Massachusetts Avenue, Suite 61, Cambridge, MA 02139, USA
- [8]- Michael J. S. Dewar, Andrew J. Holder, Roy D. Dennington, Daniel A. Liotard, Donald G. Truhlar, Todd A. Keith, John M. Millam, C. David Harris, Copyright © 1994, 1997, 2004 Semichem
- [9]- Gaussian 98, Revision A.9, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, Cioslowski, Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Iashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head- Gordon, E. S. Replogle, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1998 (2007).
- [10]- M. Dewar and W. Thiel, J. Amer. Chem. Soc. 99, 4499 (1977).
- [11]- A.R. Katritzky, V.S. Lobanov, M. Karelson, CODESSA: Reference Manual, University of Florida, Gainesville, FL, 1994.
- [12]- Alan R. Katritzky, Dimilar A. Dobchev, Indrek Tulp, Mati Karelson and David A. Carlson. Bioorganic and medicinal chemistry letters 16 (2006) 2306-2311.
- [13]- Varnek, A. ; Kireeva, N. ; Tetko, I. V. ; Baskin, I.I ; Solv'ev, V. P. J. Chem. Inf. Model. 2007, 47, 1111.
- [14]- Fairouz Tazerouti a; Ahmed Yacine Badjah-Hadj-Ahmed a; Toshihiko Hanai Journal of Liquid Chromatography and Related Technologies. 30: 3043–3057, 2007.
- [15]-Tazerouti, F.; Badjah-Hadj-Ahmed, A.Y.; Meklati, B.Y.; Pilar, F.; Minguillon, C. Chirality 2002, 14 (I), 59–66.
- [16]- M. Karelson, Molecular Descriptors in QSAR/QSPR Analysis, Wiley & Sons, New York, 2000.
- [17]- I. G. Csizmadia, Theory and Practice of MO Calculations on Organic Molecules, Elsevier, Amsterdam, 1976.
- [18]- E. Clementi, Computational Aspects of Large Chemical Systems, Springer Verlag, New York, 1980.

Conclusion Générale

- Le travail réalisé a permis de proposer de nouvelles molécules qui pourront donner lieu à de nouvelles familles d'inhibiteurs non nucléosidiques de la transcriptase inverse du VIH-1. Ces nouvelles structures sont caractérisées par une géométrie semblable à celle des molécules d'origine (TIBO) et devraient présenter de bonnes propriétés pharmacologiques. Elles possèdent la même conformation « butterfly-like » que les TIBO.

- Un travail complet de modélisation moléculaire qui a été fait, donne toutes les étapes de l'application du RNA dans la recherche de relation RQSA d'anti-VIH des dérivés de TIBO :

- Dans ce cadre là, nous avons proposé 6 nouveaux descripteurs. Il s'agit de descripteurs qui tiennent compte des effets électroniques et stériques, l'association de tous les descripteurs trouvés explique le mécanisme de l'inhibition de la TI par ces molécules.
- Ce travail confirme les résultats trouvés dans la littérature au sujet de l'importance du paramètre hydrophobe. En effet, ce paramètre contribue avec le plus grand pourcentage (19.2%) comparé aux autres descripteurs.
- I_X ajusté dans ce travail donne un bon résultat comparé à celui proposé dans la littérature.

- D'autre part, notre travail sur les flavonoïdes a conduit aux conclusions suivantes :

- Les RNA ont donné de bons résultats statistiques dans les deux phases d'apprentissage et de prédiction. Les carrés des coefficients de corrélation sont plus élevés et les déviations standards sont inférieures que celles données par les RML.
- Les RNA ont déterminé avec succès, le comportement non linéaire de quelques paramètres dans le modèle RQSA et la dépendance de l'activité anti-VIH des dérivés flavonoïdes aux aspects électroniques et hydrophobique. Cette étude confirme une fois de plus que, l'électronegativité (χ), la charge atomique sur l'atome C7 et le CLogP jouent un rôle clés dans l'activité anti-VIH des flavonoïdes.

- En ce qui concerne la séparation énantiosélective, au vu du choix restreint de la base de données et du fait que les résultats expérimentaux disponibles sur la phase stationnaire C synthétisé par F. Tazerouti et col. soient limités, nous avons fait une tentative en utilisant les régressions multi variables et multi linéaire avec une base constituée de 12 molécules plus deux autres qui ont été utilisées dans la validité du modèle théorique. La perspective d'augmenter la taille de cette base de données pourrait être judicieuse pour confirmer les modèles de régression établis et prévoir la possibilité de trouver d'autres régressions multi variables et non linéaires à qui nous pourrions appliquer les réseaux de neurones artificiels, qui seraient à notre avis porteur.