

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE



UNIVERSITE Dr. TAHAR MOULAY – SAIDA–

FACULTE DES SCIENCES  
DEPARTEMENT DE CHIMIE

N° D'ORDRE : .....

# THÈSE

Présentée par

**TOUHAMI Moufida**

Pour l'obtention du diplôme de Doctorat LMD en

\*\*\*\*\* **CHIMIE COMPUTATIONNELLE** \*\*\*\*\*

## THEME

**Modélisation de l'activité biologique de composés  
hétérocycles**

Soutenu publiquement, le 02/03/2019 devant le jury composé de :

N°	Nom et prénom	Qualité	Grade	Établissement
1	MOSTEFAI Asma	Présidente	Prof	Université Dr. Tahar Moulay -Saida-
2	RAHMOUNI Ali	Directeur de thèse	Prof	Université Dr. Tahar Moulay -Saida-
3	BENAISSA Tahar	Co-directeur	MCA	Université Dr. Tahar Moulay -Saida-
4	ALI-OTHMANE Adil	Examineur	Prof	Université d'USTOMB -Oran-
5	BELMILOUD Yamina	Examinatrice	MCA	Université d'USTHB -Alger-
6	Guendouzi Abdelkrim	Examineur	MCA	Université Dr. Tahar Moulay -Saida -

Année Universitaire 2018-2019

**Thèse préparée au Laboratoire de Modélisation et Méthodes de calcul – Université de Saida**

## "نمذجة النشاط البيولوجي للمركبات الحلقية غير المتجانسة "

### المخلص:

إن تثبيط إنزيم النسخ العكسي لفيروس نقص المناعة البشرية هو واحد من بين العلاجات المتوفرة لمحاربة المرض الذي يسببه هذا الفيروس الذي يتمثل دوره في منع تطور هذا الأخير في الجسم والمحافظة على التوازن بين وجوده في الجسم ونظام دفاعه.

في الواقع، حددت المراجع عدد 107 من مشتقات 1- [(2-hydroxyethoxy) methyl] -6 (phenylthio) thymine (HEPT) الذي يوفر أنشطة مثبطة لفيروس نقص المناعة البشرية.

يظهر تحليل البيانات التجريبية التي تم جمعها على أن النشاط البيولوجي يعتمد اعتمادا قويا على التباين في طبيعة وعدد و موضع الجذور التي تحملها حلقة الثيوفينيل و حلقة الثيمين. لفهم وتفسير هذه الاعتماد من ناحية وتطوير نماذج QSAR من ناحية أخرى، تم إجراء حسابات كيمياء الكم على مستوى DFT/B3LYP 6-311 G(d, p) على 83 مركب.

بالإضافة إلى ذلك، نماذج QSAR الموثوقة والمستقرة للتنبؤ بالنشاط المضاد لفيروس نقص المناعة البشرية للمركبات (HEPT) باستخدام الانحدار الخطي المتعدد (MLR) ومزيج من الواصفات الجزيئية ذات الطبيعة الكمية التي تحتوي من بينها على معاملات Fukui تم تطوير أفضل نموذج QSAR و الذي تم الحصول عليه باستعمال 60 مركب. وهو قادر على وصف حوالي 81% من التباين في النشاط التجريبي ويمكن استخدامه بفعالية لتقدير نشاط المشتقات التي لا تتوفر لها بيانات تجريبية.

**كلمات مفتاحية:** نشاط مكافحة فيروس نقص المناعة البشرية، DFT، مؤشرات Fukui، HEPT، QSAR، واصفات الكم

## « Modélisation de l'activité biologique de composés hétérocycles »

### Résumé

L'inhibition de la transcriptase inverse du VIH est l'un des différents traitements disponibles pour lutter contre la maladie causée par ce virus dont le rôle est de bloquer le développement de ce dernier dans le corps et de maintenir l'équilibre entre sa présence dans le corps et son système de défense. En effet, l'étude bibliographique a permis de recenser un nombre de 107 dérivés de la 1 - [(2-hydroxyéthoxy) méthyl] -6 (phénylthio) thymine (HEPT) présentant des activités d'inhibition de la transcriptase inverse du VIH puissantes et sélectives. L'analyse des données expérimentales ainsi collectées montre une forte dépendance de l'activité biologique par rapport à la variation de la nature, le nombre et la position des groupements portés par le cycle thiophényle et par la thymine. Pour comprendre et interpréter ces dépendances d'une part et de développer des modèles QSAR d'autre part, des calculs de chimie quantique au niveau DFT / B3LYP / 6-311 + G (d, p) ont été réalisés sur 83 dérivés de HEPT. Des caractéristiques structurales et des propriétés électroniques ont été évaluées et analysé à savoir, le potentiel chimique, le potentiel d'ionisation, l'affinité électronique, la mollesse, la dureté, l'électronégativité, l'indice électrophile.

Par ailleurs, des modèles QSAR fiables, stables et prédictifs pour la prédiction de l'activité anti-VIH des HEPT en utilisant la régression linéaire multiple et d'une combinaison de descripteurs moléculaires de nature quantiques parmi les quels des indices de Fukui. Le meilleur modèle QSAR obtenu a été développé sur une série de 60 molécules. Il est capable de décrire environ 81% de la variance de l'activité expérimentale et pourrait être utilisé efficacement pour estimer l'activité des dérivés pour lesquels les données expérimentales sont indisponibles.

**Mots clés** Activité anti-VIH, DFT, indices de Fukui, HEPT, QSAR, descripteurs quantique

## « Modeling the biological activity of heterocyclic compounds »

### **Abstract:**

Inhibiting HIV reverse transcriptase is one of the different treatments available to fight the disease caused by this virus, whose role is to block its development in the body and maintain the balance between its presence in the body and its defense system. Indeed, the literature review identified a number of 107 derivatives of 1 - [(2-hydroxyethoxy) methyl] -6 (phenylthio) thymine (HEPT) with potent and selective HIV reverse transcriptase inhibition activities. The analysis of the experimental data thus collected shows a strong dependence of biological activity on the variation in the nature, number and position of the groups carried by the thiophényl cycle and by the thymine. To understand and interpret these dependencies on the one hand and to develop QSAR models on the other hand, quantum chemistry calculations at the DFT / B3LYP / 6-311 + G (d, p) level were performed on 83 HEPT derivatives. Structural characteristics and electronic properties were evaluated and analyzed, namely, chemical potential, ionization potential, electronic affinity, softness, hardness, electronegativity, electrophilicity index.

In addition, reliable, stable and predictive QSAR models for predicting the HIV activity of HEPTs using multiple linear regression and a combination of molecular descriptors of a quantum nature among which of the Fukui indices. The best QSAR model obtained was developed on a series of 60 molecules. It is able to describe about 81% of the variance of experimental activity and could be used effectively to estimate the activity of derivatives for which experimental data are unavailable.

**Key words:** Anti-HIV activity, DFT, Fukui indexes, HEPT, QSAR, quantum descriptors

*« Tout obstacle renforce la détermination. Celui qui s'est fixé un but n'en change pas..... »*

*Léonard De Vinci*

*A Muma*

*A mes parents*

*A mes sœurs et mes frères*

*A mes amis et mes collègues*

*A tous les gens qui m'ont soutenu*

## **Remerciements**

*Le présent travail a été réalisé au sein du laboratoire de Modélisation et Méthodes de Calcul de l'université Dr Moulay Tahar de Saida.*

*Il me sera très difficile de remercier tout le monde car c'est grâce à l'aide de nombreuses personnes que j'ai pu mener cette thèse à son terme.*

*Je voudrais tout d'abord remercier grandement mon directeur de thèse Monsieur RAHMOUNI Ali, Professeur à l'université Dr Moulay Tahar de Saida, qui m'a encadré tout au long de cette thèse et qui m'a fait partager ses brillantes intuitions. Je suis ravie d'avoir travaillé en sa compagnie car outre son appui scientifique. Qu'il soit aussi remercié pour sa gentillesse, sa disponibilité permanente et pour les nombreux encouragements qu'il m'a prodigués.*

*Je voudrais aussi remercier mon co-encadreur de thèse, Monsieur BENAÏSSA Tahar, Maître de conférence A à l'université de Saida, il a toujours été là pour me soutenir et me conseiller au cours de l'élaboration de cette thèse.*

*J'exprime également ma profonde gratitude à Madame MOSTEFAI Asma Professeur à l'université de Saida, pour nous avoir fait l'honneur de présider le jury de cette thèse.*

*Je suis très honorée par la présence de Madame BELMILOUD Yamina, Maître de conférence A à l'université de USTHB d'Alger, à qui j'adresse mes remerciements les plus sincères et l'expression de mon profond respect pour avoir accepté d'examiner ce travail.*

*Je tiens à adresser mes vifs remerciements et l'expression de mon profond respect à Monsieur ALI-OTHMANE Adil, professeur à l'université USTOMB d'Oran, pour l'intérêt qu'il a porté à ce travail en acceptant d'en être l'un des rapporteurs de cette thèse.*

*Mes vifs et sincères remerciements vont également à Monsieur Guendouzi Abdelkrim, Maître de conférences A à l'université de Saida pour l'honneur qu'il nous a fait d'accepter de juger notre travail.*

*Enfin, j'adresse mes remerciements à tous mes collègues de l'équipe «chimie computationnelle» du laboratoire de recherche «Modélisation et Méthodes de Calcul de l'université de Saida» et à tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail.*

## Table des Matières

# Table des Matières

Liste des figures	4
Liste des Tableaux	5
Abréviations	6
Introduction générale	7
<b>Chapitre I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH</b>	
I. Introduction	12
II. Le virus de l'immunodéficience humaine (VIH)	13
II.1 Cycle de l'intégration et de la reproduction du rétrovirus VIH	14
II.2 Thérapeutique	15
II.2.1 Les Vaccins	15
II.2.2 La chimiothérapie	16
III. Inhibiteurs de la transcriptase inverse du VIH	19
III.1 La transcriptase inverse	19
III.2. Classes thérapeutiques de la transcriptase inverse	20
III.2.1 Les analogues nucléosidiques de la TI	20
III.2.2 Les inhibiteurs non nucléosidiques de la TI	23
IV. Les trithérapies contre le VIH	25
V. Développement de molécules anti-VIH à base de 1-2-[(Hydroxy-éthoxy) Methyl]-6-(Phenylthio) Thymine	
V.1 Synthèse de la 1-2-[(Hydroxyéthoxy) Methyl]-6-(Phenylthio) Thymine (HEPT)	26
V.1.1 Préparation de 1-[(2-hydroxyéthoxy) methyl] uracil	26
V.1.1.a A partir de bis (triméthylsilyl)	26
V.1.1.b A partir du 5-fluoro-4- (méthylthio) pyrimidin-2- (IH) -one	27
V.2 Activité anti-VIH de la 1-2-[(Hydroxyéthoxy) Methyl]-6-(Phenylthio) Thymine	28
V.3 Synthèse des dérivés de HEPT	30
V.3.1 Modification en C-6	31
V.3.2 Modification en C-5	32
V.3.3 Modification en C-2	33
V.3.4 Modification en N-1	36
V.3.5 Modification en C-4	37
V.3.6 Modification en N-3	38
V.3.7 Modification en position Y	39
V.4 Activité anti-VIH des dérivés de HEPT	39
V. Conclusion	45
Bibliographie	51
	52

## Table des Matières

### Chapitre II : Méthodologie QSAR

I. Introduction	57
II. Les relations quantitatives structures activités	60
II.1. Définition	60
II.2. Les premiers développements de QSAR	62
III. Méthodologie générale d'une étude QSAR	63
III.1. La base de données	63
III.2. Les descripteurs moléculaires	63
III.2.1. Définition	63
III.2.2. Types de descripteurs	64
III.2.2.1 Les descripteurs 1D	64
III.2.2.2 Les descripteurs 2D	64
III.2.2.3 Les descripteurs 3D	65
III.3 Méthodes d'évaluation de descripteurs quantiques	67
III.4. Sélection des descripteurs moléculaires	74
IV. Méthodes d'analyse de données	75
IV.1 Méthode basée sur les descripteurs	75
IV.1.1 Approches linéaires	75
IV.1.1.1 Analyse par composantes principales	75
IV.1.1.2 La régression linéaire multiple (MLR)	77
IV.1.1.3 La méthode de régression des moindres carrés partiels	79
IV.1.1.4 Analyse discriminante linéaire (LDA).	79
IV.1.2 Approches non linéaires	81
IV.1.2.1 Réseaux de neurones artificiels	81
IV.1.2.2 Arbre de décision	84
IV.1.2.3 Machine à vecteurs de support	85
V. Validation du modèle	87
VI. Conclusion	92
Bibliographie	93

### Chapitre III : Structures et propriétés électroniques de dérivés HEPT

I. Introduction	103
II. Détail de calculs	103
III. Résultats	103
III.1 Aspect structural des dérivés de HEPT	103
a. Position du radical R <sub>1</sub> sur le thiophényle	104
b. Nature du radical R <sub>2</sub>	106
c. Longueur de la chaîne carbonée du R <sub>3</sub>	107
III.2. Étude des descripteurs en fonction de la structure des dérivés	109
a. Le potentiel d'ionisation, l'affinité électronique en fonction des radicaux	109
b. La mollesse, l'indice électrophile, la dureté, l'électronégativité et le potentiel	

## *Table des Matières*

chimique en fonction de $R_1$	112
c. Les valeurs des charges nettes en fonction du radical $R_1$	113
d. Les valeurs des indices de Fukui en fonction du radical $R_1$	114
IV. Conclusion	115
Bibliographie	116

### **Chapitre IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT**

I. Introduction	118
II. Modèles QSAR de l'activité anti-VIH de dérivé HEPT publiés dans la littérature	119
II.1 Méthodes de régression linéaire multiple	120
II.1.1 Modèle de Hansch et Zhang	120
II.1.2 Modèle de Hannongbua	122
II.1.3 Modèle de Kireev	124
II.1.4 Modèle de Luco et Ferretti	125
II.1.5 Modèle de Jalali-Heravi et Parastar	126
II.1.6 Modèle de Bazoui et ses collaborateurs	126
II.2 Méthodes des moindres carrés partiels	127
II.3 Méthodes des réseaux neurones	127
III. Modèles QSAR de l'inhibition de la transcriptase inverse par les dérivés HEPT à l'aide des indices de Fukui	129
III.1 Méthodologie	129
III.1.1 Base de données	129
III.1.2 Calcul des descripteurs quantiques	132
III.1.3 Méthode statistique	136
III.2 Modèles QSAR	136
III.3 Validation du meilleur modèle	141
IV. Conclusion	144
Bibliographie	145
Conclusion générale	147
Annexe	150

## Table des Matières

# Liste des figures

### Chapitre I

Figure.1 : Représentation de la structure d'un rétrovirus	21
Figure.2 : Cycle de réplication du VIH	23
Figure.3 : Quelques lieux d'actions de traitements contre le virus du SIDA	25
Figure.4 : Structure de la transcriptase inverse du VIH	28
Figure.5 : Inhibiteurs nucléosidiques de transcriptase inverse (INTIs)	30
Figure.6 : Quelques-unes des principales classes d'INNRTs	31
Figure.7 : Fixation des INTIs et INNRTs sur la transcriptase inverse du VIH	32
Figure.8 : Inhibiteurs non nucléosidiques de transcriptase inverse (INNTIs)	33
Figure.9 : Structure de 5 Fluoro -1 - [(2-hydroxyéthoxy) méthyl] uracil	36

### Chapitre II

Figure.1 : Quelques grandes étapes de la découverte de médicaments	66
Figure.2 : Représentation simplifiée des principales étapes de la mise au point d'un candidat médicament	68
Figure.3 : Présentation de la méthodologie QSAR	70
Figure.4 : Une représentation graphique simple des bases de l'ACP	85
Figure.5 : Régression linéaire multiple	86
Figure.6 : Régression linéaire simple	87
Figure.7 : Distribution d'un échantillon à deux classes sur un axe transformé	89
Figure.8 : Le schéma d'un neurone biologique	90
Figure.9 : Le schéma d'un neurone formel	91
Figure.10 : Architecture des réseaux neurones	92
Figure.11 : Exemple d'un arbre de décision	94

### Chapitre III

Figure.1 : Numérotation des différents atomes	112
Figure.2 : Les géométries optimisées au niveau DFT B3LYP/6-311g + (d, p)	113
Figure.3 : Les géométries optimisées au niveau DFT	115
Figure.4 : Les géométries optimisées	116

### Chapitre IV

Figure.1 : Structure générale de la molécule HEPT	129
Figure.2 : Structure générale de HEPT	131
Figure.3 : Structure de la molécule HEPT dans cette étude	133
Figure.4 : Structure de la molécule HEPT dans ce modèle	134
Figure.5 : La structure de HEPT étudiée	135
Figure.6 : Structure générale de HEPT de notre travail	138
Figure.7 : La corrélation $R^2$ en fonction du nombre de descripteurs	147

## Table des Matières

# Liste des Tableaux

### Chapitre I

Tableau.1 : Liste des antirétroviraux actuellement disponibles	27
Tableau.2 : Inhibition de la réplication du VIH par les composés 11 et 12	39
Tableau.3 : L'activité biologique des dérivés de HEPT-O modifiés en C-5	54
Tableau.4 : L'activité biologique des dérivés de HEPT-O modifiés en N-1	55
Tableau.5 : L'activité biologique des dérivés de HEPT-O modifiés en C-6	56
Tableau.6 : Les dérivés de HEPT-O modifiés en N-1	57
Tableau.7 : L'activité anti-VIH des dérivés HEPT-S comparée aux dérivés HEPT-O	58
Tableau.8 : L'activité biologique des dérivés de HEPT-S modifiés en C-5 et C-6	58

### Chapitre III

Tableau.1 : Effets de la position du radical $R_1$ sur certaines caractéristiques structurales de quelques dérivés de HEPT	114
Tableau.2 : Effets de la nature du radical $R_2$ sur certaines caractéristiques structurales de quelques dérivés de HEPT	116
Tableau.3 : Effets de la longueur de la chaîne carbonée du $R_3$ sur certaines caractéristiques structurales de quelques dérivés de HEPT	117
Tableau.4 : Le potentiel d'ionisation et l'affinité électronique des dérivés de HEPT en fonction de la position et le type du radical $R_1$	118
Tableau.5 : Le potentiel d'ionisation et l'affinité électronique des dérivés de HEPT en fonction du radical $R_2$	119
Tableau.6 : Le potentiel d'ionisation et l'affinité électronique des dérivés de HEPT en fonction du radical $R_3$	120
Tableau.7 : la mollesse(S), l'indice électrophile ( $\omega$ ), la dureté( $\eta$ ), l'électronégativité ( $\chi$ ) et le potentiel chimique ( $\mu$ ) de quelques dérivés de HEPT-O en fonction de $R_1$	121
Tableau.8 : Les charges nettes de quelques dérivés de HEPT-O en fonction du groupement $R_1$	122
Tableau.9 : Les indices de Fukui de quelques dérivés de HEPT-O en fonction du $R_1$	123

### Chapitre IV

Tableau.1 : Les inhibiteurs HEPT utilisés dans ce modèle	130
Tableau.2 : Les dérivés HEPT utilisés dans cette étude	132
Tableau.3 : Les descripteurs impliqués dans cette étude	136
Tableau.4 : Les résultats statistiques de différentes architectures de réseaux de neurones	137
Tableau.5 : Les structures chimiques ainsi que les activités observées des composés	139
Tableau.6 : Les descripteurs utilisés dans les modèles	141
Tableau.7 : Les valeurs des descripteurs	142
Tableau.8 : Les valeurs des activités observées/calculées en utilisant Mod10	148
Tableau.9 : La validation croisée du meilleur modèle Mod10	151
Tableau.10 : Comparaison de différentes méthodes d'analyse de la relation structure – activité	152

## Abréviations

### Abréviations

AA: Acide Aminé  
ADMET: Absorption, Distribution, Métabolisme, Excrétion et Toxicité  
ADN: L'acide DésoxyriboNucléique  
ARN: Acide RiboNucléique  
ARV: Un AntiRétroViral  
B3LYP: Becke, Three-Parameter, Lee-Yang-Parr  
CCR5 : Le récepteur à C-C chimiokine de type 5  
CD4: Cluster de Différenciation 4  
CE: Concentration efficace  
CI: Concentration Inhibitrice  
CoMFA: Analyse Comparative de Champs Moléculaires  
CoMSIA: Analyse Comparative d'indice de Similarité Moléculaire  
DMF: Le N, N-DiMéthylFormamide  
HAART: Highly Active Antiretroviral Therapy  
HEPT: 1 - [(2-hydroxyéthoxy) méthyl] -6- (phénylthio) thymine  
HTLV-III: Human T-cell lymphotropic virus-III  
In vitro: dans le verre  
In vivo: au sein du vivant  
k – NN: Méthode des k plus proches voisins  
kDa : kiloDalton  
LAS : lymphadénopathie  
LDA: Lithium DiisopropylAmide  
LMO: Leave Many out  
LOO: Leave One Out  
LTMP: le Lithium 2,2,6,6-TétraMéthylPipéridide  
OCDE: Organisation de Coopération et de Développement Economiques  
ONUSIDA: Le Programme Commun des Nations Unies sur le SIDA  
PBMC: Les cellules mononuclées du sang périphérique  
PSO: Particle Swarm Optimization  
QSAR: Quantitative Structure-Activity Relationship  
QSPR: Quantitative Structure- Property Relationship  
REACH: Enregistrement, Evaluation, Autorisation et Restriction des Substances Chimiques  
RF: Random Forest  
SIDA: Le Syndrome d'ImmunoDéficiency Acquis  
SRM: Structural Risk Minimization  
TBDMS: Tert-ButylDiMéthylSilyle  
THF: TétraHydroFurane  
TI: La Transcriptase Inverse  
VC: Validation Croisée  
VIH: Le Virus de l'Immunodéficiency Humaine

# *Introduction générale*

La découverte de nouveaux médicaments de la manière la plus efficace et la moins coûteuse possible constitue un enjeu majeur [1]. 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 [2]. Il a été estimé que seulement 11% des médicaments qui ont terminé le stade de développement ont été approuvés par les organismes de contrôle américains ou européens. En 2000, 10% de l'attrition durant le développement du médicament était dû à une pharmacocinétique et à une biodisponibilité médiocre, alors qu'au stade clinique, 30% de l'attrition était attribuable au manque d'efficacité et 30% à la toxicité ou à la sécurité clinique [3-4]. Ainsi, il sera utile de prévoir ces échecs avant le stade clinique afin de réduire les coûts de développement des médicaments. Il a été affirmé que des économies de 100 millions USD en coûts de développement par médicament pourraient être réalisées avec un gain de prévision de 10% [5]. Par conséquent, diverses méthodes, telles que les méthodes *in vitro*, *in vivo* ou *in silico*, sont utilisées dès le début de développement du médicament pour filtrer les échecs potentiels.

Un exemple d'une méthode *in silico* est la relation quantitative structure-activité (QSAR), elles consistent essentiellement en la recherche de similitudes entre molécules dans de grandes bases de données de molécules existantes dont les propriétés sont connues. La découverte d'une telle relation permet de prédire les propriétés physiques et chimiques et l'activité biologique de composés [6-7]. Elle permet également de guider les processus de développement de nouvelles molécules, sans avoir à les synthétiser, ou à analyser des familles entières de composés. Les relations entre les structures des molécules et leurs propriétés ou activités sont généralement établies à l'aide de méthodes de modélisation par apprentissage statistique. Les techniques usuelles reposent sur la caractérisation des molécules par un ensemble de descripteurs ; nombres réels mesurés ou calculés traduisant des propriétés des structures moléculaires. Actuellement, il est possible de générer des milliers de descripteurs, cependant un seul petit nombre peut se révéler pertinent pour modéliser une activité biologique donnée. Pour sélectionner les descripteurs les plus efficaces pour les composés, les méthodes de sélection de descripteur appropriées sont nécessaires [8-9].

## *Introduction générale*

Le virus de l'immunodéficience humaine de type 1 (VIH) est l'un des virus les plus meurtriers qui affectent la santé publique dans le monde. Le Programme Commun des Nations Unies sur le VIH / SIDA (ONUSIDA) et l'Organisation mondiale de la santé estiment à plus de 15 millions le nombre de personnes infectées par ce VIH dans le monde. Pendant des décennies, la recherche sur VIH a été un défi constant dans la recherche du médicament le plus efficace [10]. Des résultats des nombreuses recherches effectués, il a été noté que l'approche de l'inhibition de l'enzyme transcriptase inverse (TI) du VIH s'est révélée efficace [11-12]. Cette enzyme traduit l'ARN viral en un ADN au sein de la cellule hôte [13–15]. Les recherches ont montré qu'un inhibiteur non nucléosidique de la transcriptase inverse appelé 1 - [(2-hydroxyéthoxy) méthyl] -6- (phénylthio) thymine (HEPT) offre une grande puissance dans l'inhibition de la TI du VIH.

De la littérature, plusieurs modèles QSAR ont été développés pour différents ensembles de dérivés HEPT pour la prédiction de l'inhibition de la transcriptase inverse du VIH. Le développement de ces modèles QSAR est basé sur les descripteurs classiques de type: constitutionnels, topologiques et géométriques. L'utilisation des descripteurs quantiques est moins fréquente alors qu'ils présentent l'avantage de pouvoir exprimer les propriétés de la densité électronique des systèmes moléculaires.

L'objectif principal des travaux de cette thèse est l'application d'une méthode de la modélisation moléculaire pour prédire l'activité biologique anti-VIH des molécules HEPT reliant les activités expérimentales aux structures moléculaires par des descripteurs quantiques et plus particulièrement les indices de Fukui.

Le présent manuscrit comporte quatre chapitres : Après une introduction générale, le premier chapitre est scindé en deux parties : dans la première partie, nous allons présenter des généralités sur les virus et plus spécialement le VIH. Dans la deuxième partie, nous allons s'étaler, sur le développement de la molécule 1-[2-hydroxyéthoxy-méthyl]-6-(phénylthio) thymine (HEPT) et ses dérivés.

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

## **Introduction générale**

Le troisième chapitre englobe les résultats des calculs effectués dans notre étude et à la présentation des résultats de calcul de chimie quantique concernant la structure et quelques propriétés électroniques de la molécule HEPT.

Le quatrième chapitre présente des relations structure-activité des 1-[2-hydroxyéthoxy-méthyl]-6-(phenylthio) thymine. Ainsi que, le développement des modèles statistiques de l'activité biologique : l'activité anti-VIH.

Finalement, une conclusion générale, dans laquelle nous rappelons les principaux résultats obtenus, et quelques perspectives seront proposées.

## *Introduction générale*

### **Bibliographie**

- [1].Snodin, D.J. (2002) *Toxicology Letters*, 127, 161–168.
- [2].Kraljevic, S., Stambrook, P.J., and Pavelic, K. (2004) *EMBO Reports*, 5, 837–842.
- [3].Adams, C.P. and Brantner, V.V. (2006) *Health Aff (Millwood)*, 25, 420–428.
- [4].Kola, I. and Landis, J. (2004) *Nature Reviews. Drug Discovery*, 3, 711–716.
- [5].Critical Path Opportunities Reports > Challenges and Opportunities Report – March (2004) [http://www.fda.gov/ ScienceResearch/SpecialTopics/ CriticalPathInitiative/ CriticalPathOpportunitiesReports/ ucm077262.htm#f15](http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/ucm077262.htm#f15).
- [6].Yap, C.W., Xue, Y., Li, Z.R., and Chen, Y.Z. (2006) *Current Topics in Medicinal Chemistry*, 6, 1593–1607.
- [7].Guido, R.V., Oliva, G., and Andricopulo, A.D. (2008) *Current Medicinal Chemistry*, 15, 37–46.
- [8].Go'tte M, Rausch JW, Marchand B, Sarafianos S, Le Grice SF (2010) *Biochim Biophys Acta* 1804:1202
- [9].Hu WS, Hughes SH (2012) *Cold Spring Harbor Perspect Med* 2:a006882
- [10].Global AIDS update 2016. (2016) UNAIDS. [http://www.unaids.org/sites/default/files/media\\_asset/global-AIDS-update-2016\\_en.pdf](http://www.unaids.org/sites/default/files/media_asset/global-AIDS-update-2016_en.pdf). Accessed 20 Feb 2017
- [11].De Clercq E (1998) *Antivir Res* 38:153
- [12].De Clercq E (2004) *Chem Biodivers* 1:44
- [13].Telesnitsky A, Goff SP (1997) *Reverse transcriptase and the generation of retroviral DNA*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, *Retroviruses*
- [14].Shafer RW, Dupnik K, Winters MA, Eshleman SH (2001) *A guide to HIV-1 reverse transcriptase and protease sequencing for drug resistance studies*. *Compendium, HIV Sequence*, p 1
- [15].Sarafianos S, Marchand B, Das K, Himmel D, Parniak MA, Hughes SH, Arnold E (2009) *J Mol Biol* 385:693

*Chapitre I*  
*Dérivés HEPT pour*  
*l'inhibition de la*  
*transcriptase inverse*  
*du VIH*

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

### I. Introduction

La compréhension des virus et de leurs interactions avec le reste du vivant a beaucoup évolué ces dernières années. Tout d'abord il faut bien réaliser que la mauvaise réputation des virus s'est construite autour des effets – certes parfois dévastateurs – causés chez l'homme par une infime proportion de la totalité des virus présents sur notre planète. Moins d'une dizaine sont responsables des maladies virales les plus fréquentes. S'il est bien normal que ce petit nombre de virus nous préoccupe particulièrement, il faut savoir que l'écrasante majorité des virus non seulement ne peut pas infecter l'homme mais joue un rôle crucial dans son « écosystème » interne. [1]

L'impact de ces virus n'est pas encore complètement compris, mais ils jouent un rôle important dans la régulation de la composition des communautés bactériennes vivant en symbiose avec l'homme [2].

Les Virus ont été relatés depuis des temps immémoriaux. Toutefois, l'origine exacte de ces minuscules organismes qui portent uniquement sur l'information génétique dans une couche de protéines est encore inconnue. Le principal problème est qu'aucun fossile de virus n'a été détecté.

Un virus est un parasite intracellulaire ne pouvant se multiplier qu'à l'intérieur d'une cellule hôte et utilisant sa machinerie cellulaire. Il contient : une information génétique (sous forme d'ADN ou d'ARN) protégée par une structure compacte, souvent protéique, la capsid. Il existe de nombreux types différents de virus, parmi lesquels on retrouve en particulier les rétrovirus.

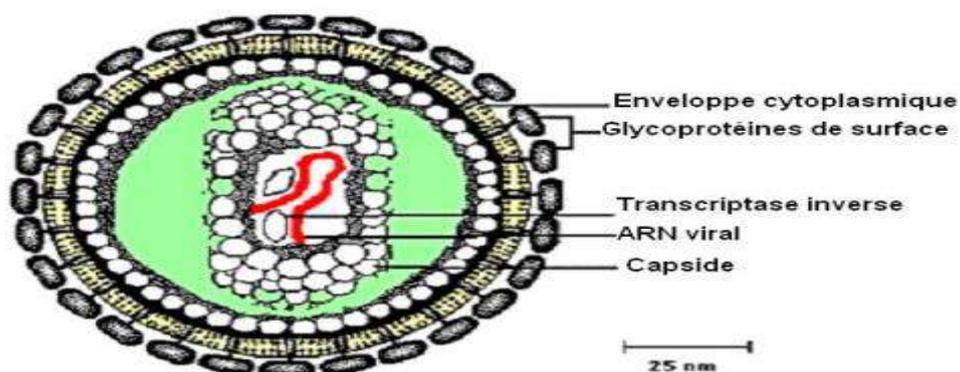


Figure.1 Représentation de la structure d'un rétrovirus

## **Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH**

Les rétrovirus sont des virus d'un diamètre de 110 à 125 nanomètres, très répandus dans le monde animal. Ils sont la cause de différentes formes de cancers, d'immunodéficiences, dont le Sida, et de dégénérescences du système nerveux central. Leur génome s'intègre sous forme d'ADN dans celui de la cellule hôte, pour ensuite s'exprimer pendant toute la vie active de la cellule. Les lentivirus (ou Lentivirinae) font partie de cette famille : ces virus sont responsables de pathologies à évolution lente. L'exemple le plus connu de tels virus est le virus du Sida.

### **II. Le virus de l'immunodéficiency humaine (VIH)**

Le syndrome d'immunodéficiency acquise (SIDA) causé par le virus de l'immunodéficiency humaine (VIH), constitue un problème majeur dans le monde entier. Le premier rapport publié en 1981 décrit cette épidémie comme un syndrome clinique de l'immunodéficiency bien avant l'identification du VIH [3].

Le SIDA est la manifestation d'une infection prolongée au VIH au stade terminal. Les personnes infectées par le VIH meurent généralement du SIDA après environ 10 ans d'infection.

Depuis le début de l'épidémie, plus de 60 millions de personnes ont été infectées dans le monde [4]. Face à cette pandémie mondiale, un effort important a été déployé dans les années 1980 pour trouver les moyens de combattre la maladie. Bien qu'un remède ou un traitement global reste à trouver, des progrès ont été réalisés au cours des dernières décennies.

L'agent mis en cause dans le SIDA a été isolé en 1983 et reconnu comme un nouveau rétrovirus du ganglion lymphatique d'un homme atteint du syndrome de lymphadénopathie persistante (LAS) [5-6]. En 1984, un autre rétrovirus humain nommé HTLV-III (virus de la leucémie à lymphocytes T humains III) a été isolé à partir de cellules mononuclées du sang périphérique (PBMC) d'adultes et de patients pédiatriques atteints du SIDA [7]. La même année Levy et ses collaborateurs ont isolé un rétrovirus de patients atteints du SIDA et l'ont appelé le rétrovirus associé au SIDA (ARV) [8]. Les trois virus LAV, HTLV-III et ARV ont été rapidement reconnus comme des membres antigéniquement indiscernables du même groupe de rétrovirus appartenant à la famille des lentiviridae. En 1986, un comité international sur la taxonomie des virus recommandait de donner au virus du sida un nom distinct : Virus de l'Immunodéficiency Humaine (VIH) [9]. En 1986, un deuxième virus du sida a été découvert chez des malades du SIDA en Afrique de l'Ouest [10].

Pour distinguer ces virus, le rétrovirus initialement découvert associé à la plupart des maladies à VIH dans le monde est désigné comme VIH-1 et le second virus détecté en Afrique de l'Ouest est désigné par VIH-2. Les deux souches sont antigéniquement distinctes. Le VIH-2 diffère de plus de 55% de la souche VIH-1. Le VIH-2 est rare dans la plupart des régions du monde en raison d'une transmission moins efficace.

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

Le taux de mortalité par infection par le VIH-2 est également deux fois moins élevé que celui du VIH-1. Dans l'usage courant, « VIH » implique habituellement le VIH-1. VIH est un virus à ARN de masse moléculaire élevée, dotée d'une enzyme spécifique dénommée La Transcriptase Inverse (TI) ou Réverse Transcriptase permet de synthétiser, à partir de l'ARN viral, un ADN bicaténaire dit « proviral » qui va s'intégrer dans le génome de la cellule infectée [11].

Les études menées sur le VIH ont permis d'élucider des événements moléculaires critiques nécessaires à la réplication virale et ont identifié plusieurs molécules cibles possible pour de le traitement thérapeutique. Les détails des composants structurales et le cycle de vie du VIH ont été discutés [12, 13]. La plupart des travaux de recherche pour le développement d'agents chimiothérapeutiques ont effectué sur deux enzymes majeures, la transcriptase inverse (TI) et la protéase (PR), du code génétique du VIH [14, 15].

### II.1 Cycle de l'intégration et de la reproduction du rétrovirus VIH

Le schéma ci-dessous résume le cycle d'intégration. Les 8 étapes de ce cycle sont :

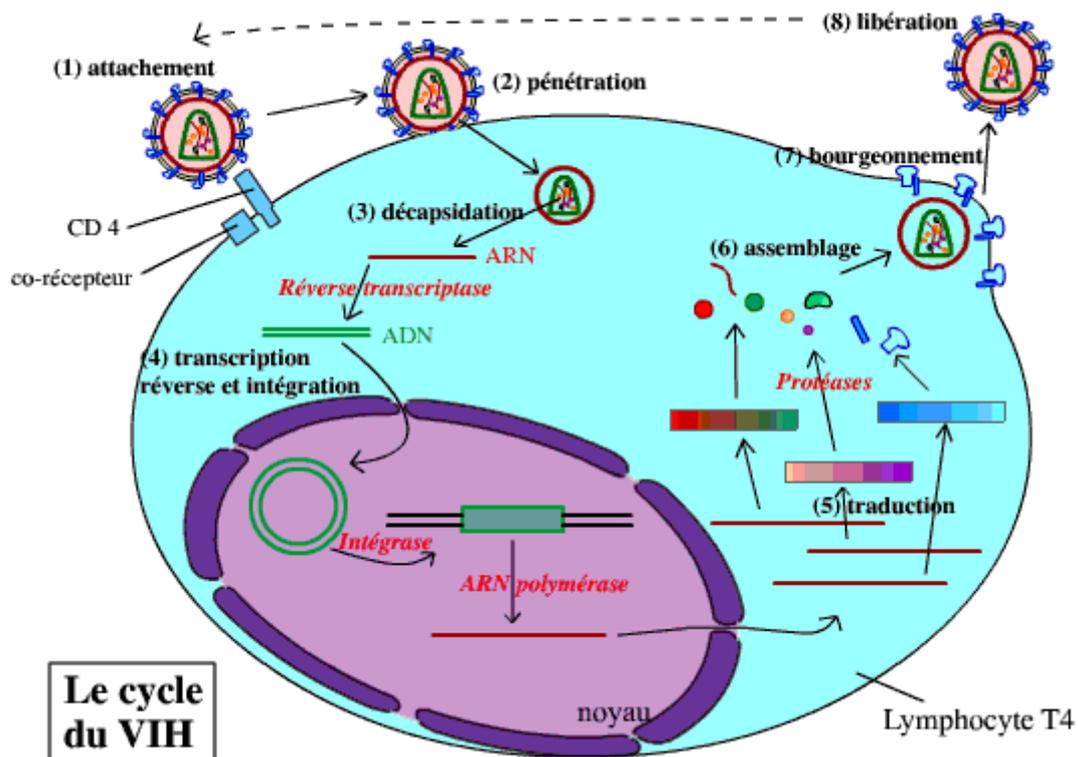


Figure.2 Cycle de réplication du VIH-1

## **Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH**

- (1) Attachement : Le virus se fixe sur le lymphocyte T CD4, par reconnaissance entre la protéine virale gp120 et la protéine CD4 du lymphocyte (ainsi qu'un co-récepteur).
- (2) Pénétration : Les deux membranes (du virus et du lymphocyte) fusionnent, ce qui permet la pénétration de la nucléocapside (les deux capsides + le matériel génétique) du virus dans le cytoplasme.
- (3) Décapsidation : Les deux capsides se dissocient, libérant l'ARN viral dans le cytoplasme.
- (4) Transcription inverse et intégration : Grâce à la transcriptase inverse virale, l'ARN viral est rétrotranscrit en ADN double brin. Cet ADN pénètre dans le noyau, où il s'intègre au génome du lymphocyte. Il est ensuite transcrit en ARN.
- (5) Traduction : Après avoir été transcrits par l'ARN polymérase de la cellule, les ARN messagers viraux sont traduits en trois précurseurs protéiques. Ces précurseurs sont clivés par des protéases, pour donner les différentes protéines du virus.
- (6) Assemblage : Les protéines virales et l'ARN viral (transcrit par ailleurs) sont associés pour reformer des virus (sans la membrane). Les protéines virales membranaires sont intégrées à la membrane du lymphocyte.
- (7) Bourgeonnement : Le virus bourgeonne, emportant un fragment de la membrane plasmique du lymphocyte (qui contient uniquement les protéines membranaires virales).
- (8) Libération : Les nouveaux virus sont libérés dans le milieu intérieur. Ils peuvent infecter de nouveaux lymphocytes T CD4. [16]

### **II.2 Thérapeutique**

Les recherches de traitement contre le virus du Sida sont multiples, la compréhension du cycle de réplication du VIH a fait apparaître de nombreuses approches d'inhibitions du virus. Les composés antiviraux ont pour vocation d'agir à de différents stades de la réplication du virus, allant de sa fixation sur la cellule cible jusqu'à sa libération dans l'organisme.

L'objectif idéal étant de pouvoir d'une part prévenir l'infection et d'autre part de guérir les cellules infectées, 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 [17-18].

#### **II.2.1 Les Vaccins**

Au même titre que pour les autres maladies infectieuses, la meilleure approche pour contrôler la propagation du SIDA consiste à développer un vaccin empêchant l'infection par le VIH. Malgré d'intenses recherches biomédicales et des études cliniques encourageantes, aucun vaccin anti-VIH n'a encore été mis au point. En effet, le développement d'un vaccin fait face à de nombreux obstacles. La diversité génétique du VIH-1 et sa capacité de mutation sont des éléments qui rendent la conception d'un vaccin problématique [19-20].

# Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

## II.2.2 La chimiothérapie

Les autorités sanitaires ont approuvé plusieurs médicaments pour contrôler l'évolution des infections à VIH. La lutte contre le VIH est actuellement basée sur l'emploi de molécules (pour la plupart synthétiques) capables de diminuer et de maintenir une présence virale indétectable, de restaurer une défense immunitaire en augmentant le nombre de LT CD4+, d'allonger l'espérance et d'améliorer la qualité de vie et de réduire le risque de transmission [21]. En théorie, chaque étape du cycle de vie du virus constitue une cible potentielle dans le cadre du traitement de l'infection par le VIH.

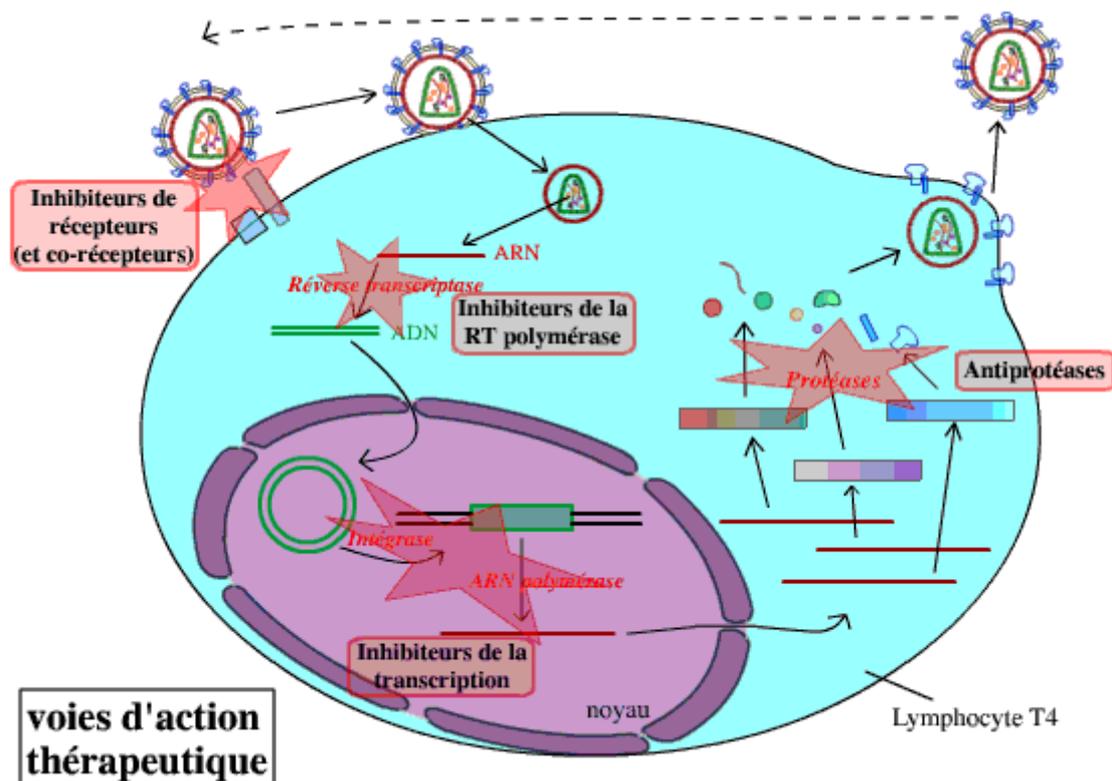


Figure. 3 : Quelques lieux d'actions de traitements contre le virus du SIDA

## **Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH**

Les différentes thérapies antirétrovirales [16] possibles sont :

- ✓ L'inhibition de la fusion VIH-CD4.
- ✓ L'inhibition de la transcriptase inverse : Inhibiteurs nucléosidiques / nucléotidiques / non nucléosidiques.
- ✓ L'inhibition de l'intégrase.
- ✓ L'inhibition du co-recepteur CCR5.
- ✓ L'inhibition de la protéase.

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

Tableau.1 : Liste des antirétroviraux actuellement disponibles [22]

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

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

### III. Inhibiteurs de la transcriptase inverse du VIH

#### III.1 La transcriptase inverse

Elle a été la première cible des thérapies antirétrovirales en raison de son rôle primordial dans la réplication du virus. Structurellement, la TI du VIH est un hétérodimère composé de deux sous-unités (figure 4). La plus grande d'entre elles, d'un poids moléculaire de 66 kDa (p66) et d'une longueur totale de 560 acides aminés (AA), est l'unité fonctionnelle de l'enzyme.

La seconde sous-unité, d'un poids moléculaire de 51 kDa (p51), est le résultat du clivage protéolytique de l'extrémité C-terminale de p66 au niveau du codon 440 [13] et joue un rôle de stabilisation structurale en favorisant le contact entre p66 et le complexe amorce/matrice (complexe A/M = complexe ARN/ADN). Cette unité p51 n'a pas de fonction RNase H à elle seule, car elle débute au niveau de l'AA 441, alors que l'activité ribonucléasique de la TI du VIH-1 dépend des AA 427 à 560 [23]. En revanche, de nombreuses études ont montré que l'interaction de ces deux sous unités est indispensable à l'activité de l'enzyme *in vivo* [24-27]. L'analyse de la structure cristallographique de ce domaine évoque une main droite constituée de sous-domaines dénommés. (Figure.4)

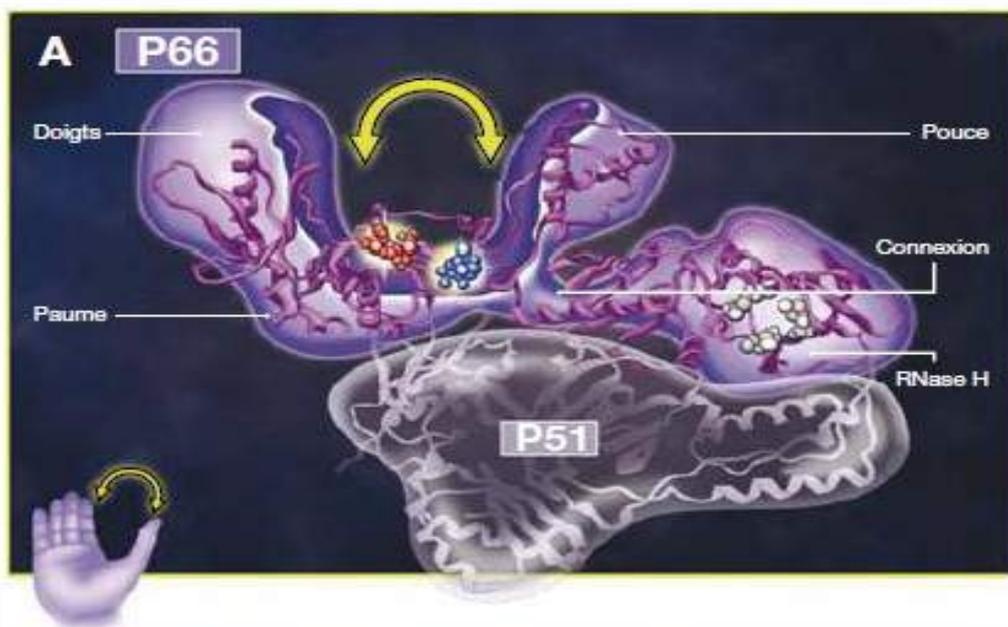


Figure.4 Structure de la transcriptase inverse du VIH-1

## **Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH**

### **III.2. Classes thérapeutiques de la transcriptase inverse**

#### **III.2.1 Les analogues nucléosidiques de la TI**

Les INTIs sont la première classe [28-29] mise sur le marché avec la zidovudine (AZT) en 1987[30-32]. Cette molécule a été développée initialement dans le cadre de la recherche de nouveaux agents antimitotiques anticancéreux. D'efficacité très modeste dans cette indication, l'AZT s'est révélé plus tard un inhibiteur efficace de la réplication du VIH [33]. Et elles restent largement utilisés dans la thérapie antirétrovirale en combinaison avec les autres classes (INNTI, IP, IE et INI).

Les molécules de cette classe agissent durant la phase précoce du cycle de réplication virale par inhibition de la TI. Les INTI sont des analogues des nucléosides naturels, privés de groupement hydroxyle en 3' (didésoxynucléosides ou ddNTP), les empêchant d'assurer la liaison 5'-3' avec le nucléoside suivant. Ils sont ainsi appelés « terminateur de chaîne » [34].

Ces dérivés entrent en compétition avec les nucléosides naturels et sont incorporés dans le premier brin d'ADN proviral lors de sa synthèse par la TI. L'arrêt de la synthèse d'ADN viral correspond à l'effet antiviral. Les INTIs constituent des prodrogues qui, une fois dans la cellule, subissent une triphosphorylation par des kinases cellulaires pour être actifs. Le ténofovir (TDF), commercialisé en 2002[35], est le seul traitement dont la particularité est d'être déjà monophosphorylé ce qui présente l'avantage de ne nécessiter que deux étapes de phosphorylation intracellulaire.

Les INTI actuellement commercialisés sont au nombre de sept : l'AZT, le didanosine (ddI), la stavudine (d4T), l'abacavir (ABC), la lamivudine (3TC), l'emtricitabine (FTC) pour les analogues nucléosidiques et le ténofovir (TDF) pour les analogues nucléotidiques.

**Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH**

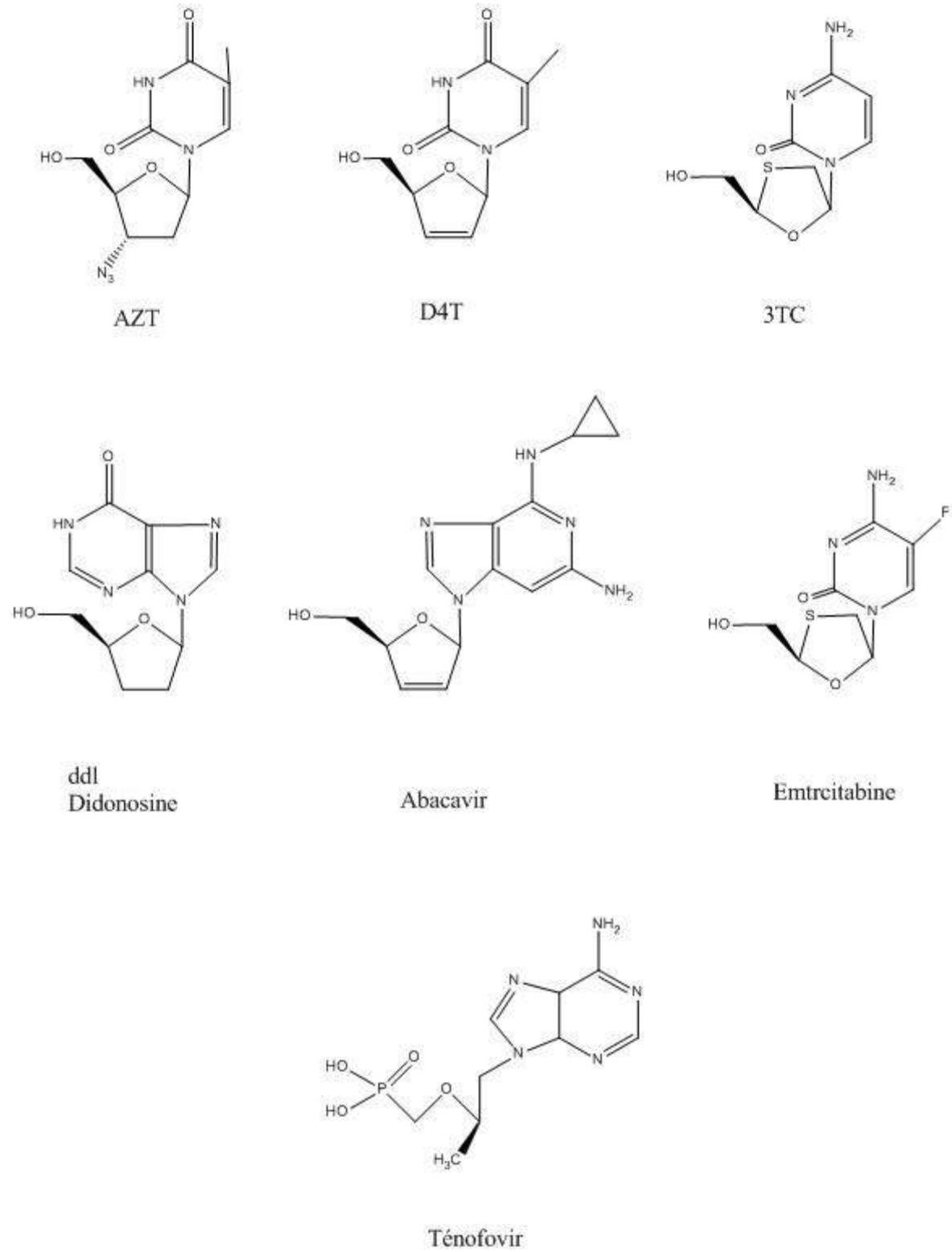


Figure.5 : Inhibiteurs nucléosidiques de transcriptase inverse (INTIs)

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

Cependant, l'utilisation à long terme de l'AZT est souvent limitée par des effets secondaires graves tels que la suppression du rôle la moelle osseuse aussi que la molécule se révèle extrêmement toxique et on doit recourir à des transfusions sanguines très fréquentes, en raison de l'anémie qu'elle provoque chez les patients [36]. Il semblerait donc impératif de rechercher de nouveaux composés ayant une activité anti-VIH puissante, une moindre toxicité et, de préférence, agissant par un mécanisme différent.

Il existe un grand nombre de substances dont l'efficacité in vitro a été démontrée [37]. Ces molécules montrent une grande diversité de structure, comme l' $\alpha$ -APA, le PETT ou le TIBO [38]. Mais la classe homogène la plus étudiée et qui a montré les meilleurs résultats d'inhibition in vitro est celle des HEPT qui est identifié dès 1989[39].

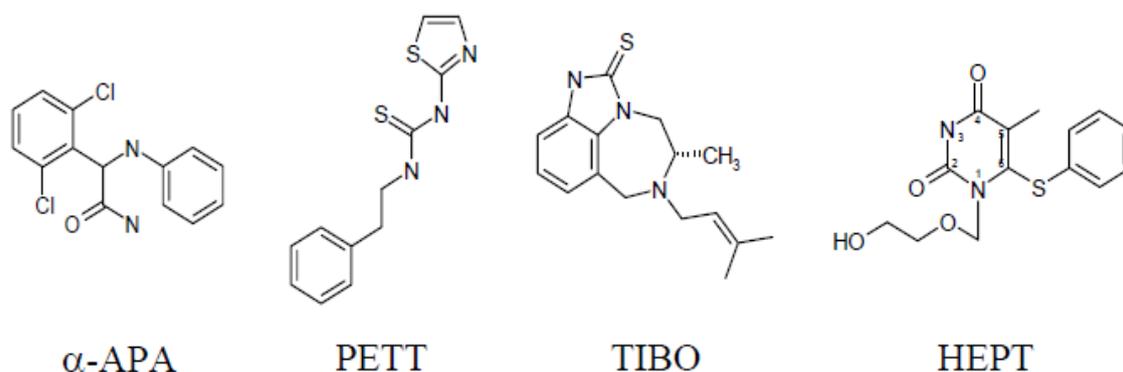


Figure. 6 : Quelques-unes des principales classes d'INNRTs

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

### III.2.2. Les analogues non nucléosidiques de la TI

De structure chimique différente des INTIs, ils ont une activité inhibitrice importante sur la TI du VIH-1, alors qu'ils sont inactifs sur le VIH-2. À la différence des analogues nucléosidiques, ils inhibent la TI de façon non compétitive, en se fixant au niveau d'une poche hydrophobe située à proximité du site catalytique de l'enzyme. Les INNTIs ont la particularité d'être métabolisés de façon prédominante par le foie avec une demi-vie longue permettant une prise unique quotidienne.

Les INNTIs actuellement commercialisés sont au nombre de quatre. Les molécules de première génération, dont la barrière génétique est faible et présentant des résistances croisées, sont la névirapine (NVP) et l'éfavirenz (EFV). Les molécules de deuxième génération dont le chef de file est l'étravirine (ETR) capable d'inhiber des virus porteurs de mutations de résistance aux inhibiteurs de première génération lorsqu'il est associé à des antirétroviraux pleinement actifs. La rilpivirine (RPV), nouvelle molécule de deuxième génération a récemment obtenu l'autorisation de mise sur le marché chez les patients naïfs d'antirétroviraux. Cette nouvelle molécule semble avoir une barrière génétique plus faible [40-41]

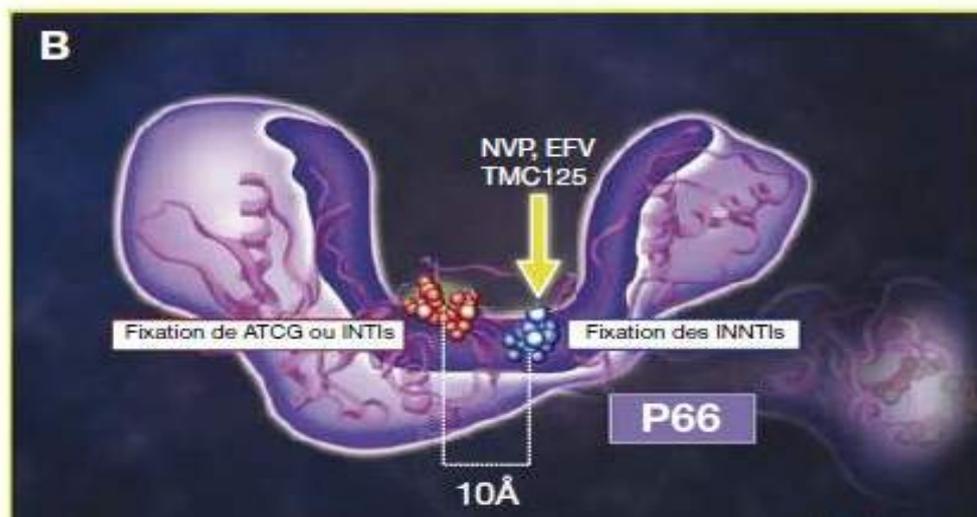
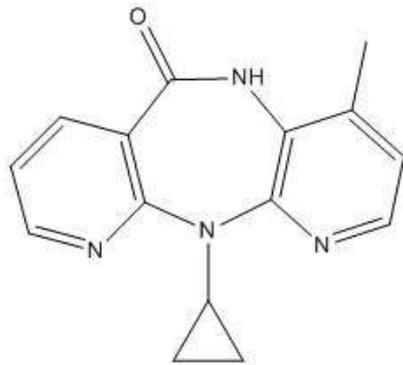
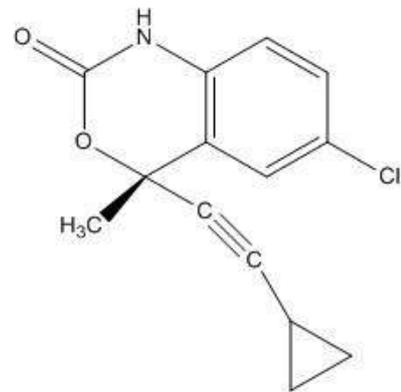


Figure.7 Fixation des INTIs et INNTs sur la transcriptase inverse du VIH-1

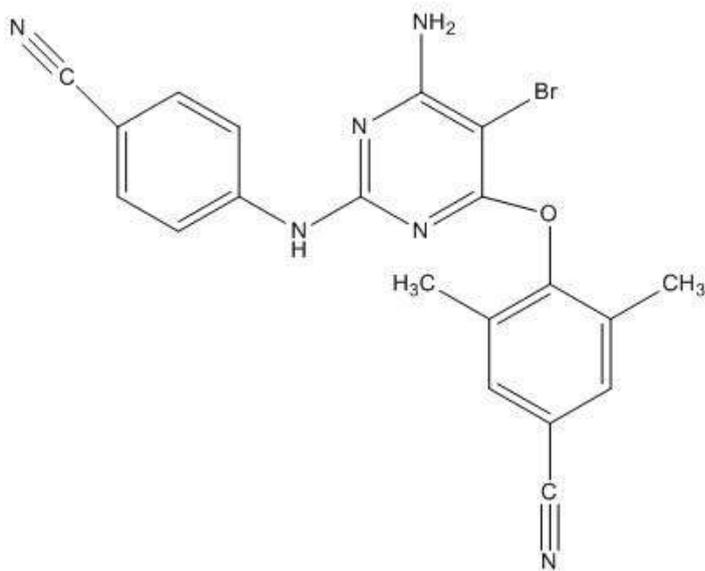
**Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH**



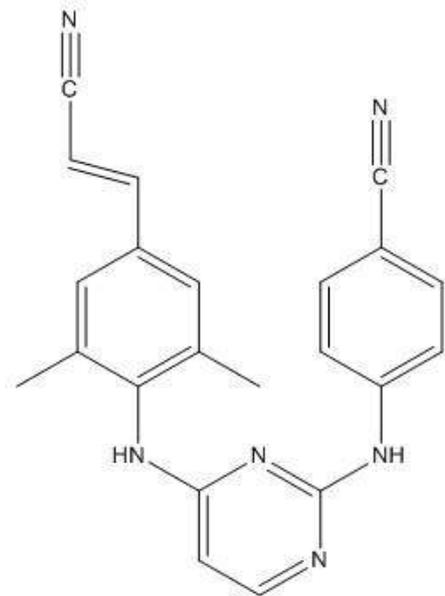
Névirapine



Efavirenz



Etravirine



Rilpivirine

Figure.8 : Inhibiteurs non nucléosidiques de transcriptase inverse (INNTIs)

## **Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH**

### **IV. Les trithérapies contre le VIH**

Lorsque le premier traitement du VIH est arrivé, personne ne savait vraiment s'il serait efficace ou pendant combien de temps. De nos jours, grâce aux progrès réalisés, il semble que le traitement réussisse mieux qu'ils avaient osé l'espérer au début.

Les médicaments contre le VIH sont prescrits sous forme de « trithérapies », des associations de trois substances antivirales qui appartiennent à des classes différentes, également appelées multithérapies ou highly active antiretroviral therapy (HAART). Ces derniers assurent également un meilleur suivi des effets secondaires rencontrés ou de la résistance du virus, en permettant de remplacer un des composés entrant dans leur composition, en fonction de l'état général du patient [42].

Trois types d'associations peuvent être envisagés. Les combinaisons de 1er niveau utilisent deux composés inhibant des étapes différentes du cycle de réplication. Les combinaisons de 2eme niveau font appel à des molécules agissant sur des sites différents d'une même protéine (INTI et INNTI). Les combinaisons de 3eme niveau ont recours à des inhibiteurs possédant un même mécanisme d'action, mais un profil de résistance différent [43].

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

### V. Développement de molécules anti-VIH à base de 1-2-[(Hydroxy-éthoxy) Methyl]-6-(Phenylthio) Thymine

Un dérivé d'acyclouridine, le 1 - [(2-hydroxy-éthoxy) méthyl] 6-phénylthiothymine (HEPT) s'est avéré être un inhibiteur puissant et sélectif du virus de l'immunodéficience humaine de type 1 in vitro. HEPT inhibe la réplication du VIH dans diverses cultures de cellules T4 ainsi que dans les lymphocytes du sang et les macrophages. La concentration efficace antivirale de 50% pour le VIH-1 dans les cellules MT-4 est de 7,0  $\mu\text{M}$ , tandis que la concentration cytotoxique de 50% pour les cellules MT-4 infectées est de 740  $\mu\text{M}$ .

Bien que HEPT inhibe diverses souches de VIH-1, il n'a aucun effet sur la réplication d'autres rétrovirus, y compris le VIH-2, contrairement à l'AZT [44].

#### V.1 Synthèse de la 1-2-[(Hydroxyéthoxy) Méthyl]-6-(Phenylthio) Thymine (HEPT)

La molécule HEPT est synthétisée à partir du 1 - [(2-hydroxyéthoxy) méthyl] uracil [39].

##### V.1.1 Préparation de 1-[(2-hydroxyéthoxy) méthyl] uracil

La synthèse du 5 Fluoro -1 - [(2-hydroxyéthoxy) méthyl] uracil peut se faire par 2 méthodes :

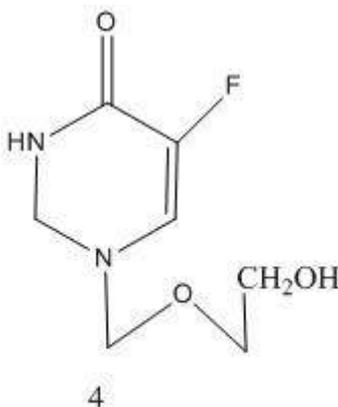


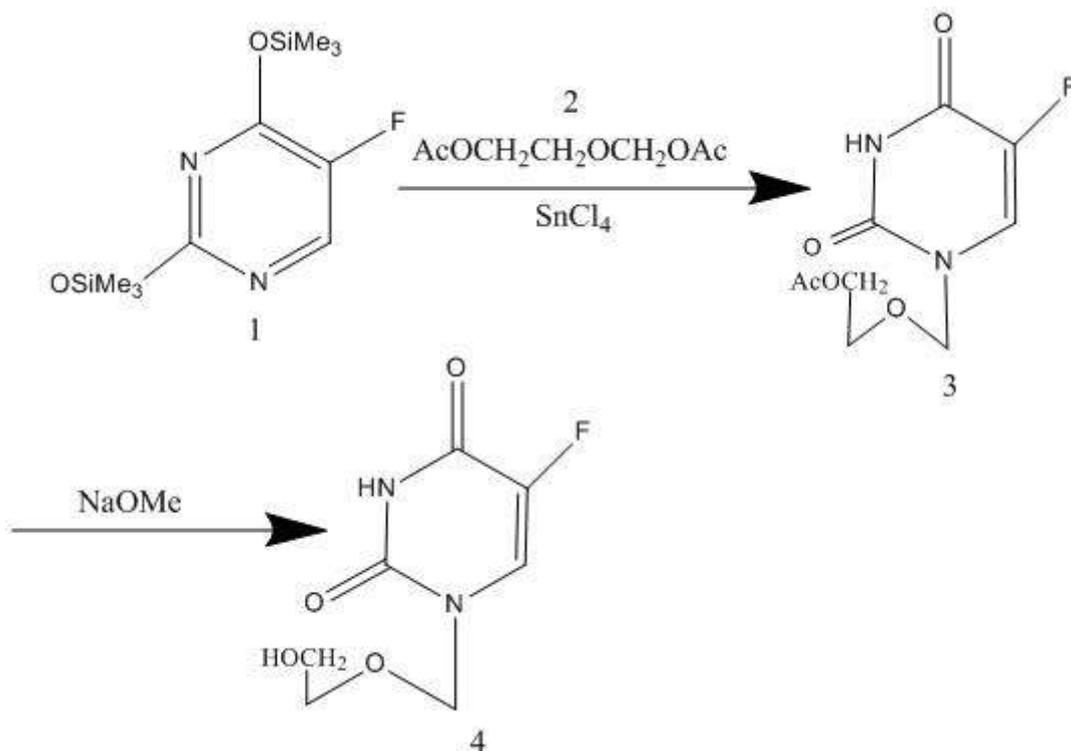
Figure.9 Structure de 5 Fluoro -1 - [(2-hydroxyéthoxy) méthyl] uracile

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

### V.1.1.a A partir de bis (triméthylsilyl)

L'ester acétate (3) provient de la condensation d'un dérivé bis (triméthylsilyl) (1) et du 2-acétoxyéthyl acétoxyéthyl éther (2) en présence de chlorure stannique ( $\text{SnCl}_4$ ). La molécule 4 est obtenue par la déprotection de la molécule (3) avec du méthoxyde de sodium avec un rendement de 83% [45].

Schéma (1)

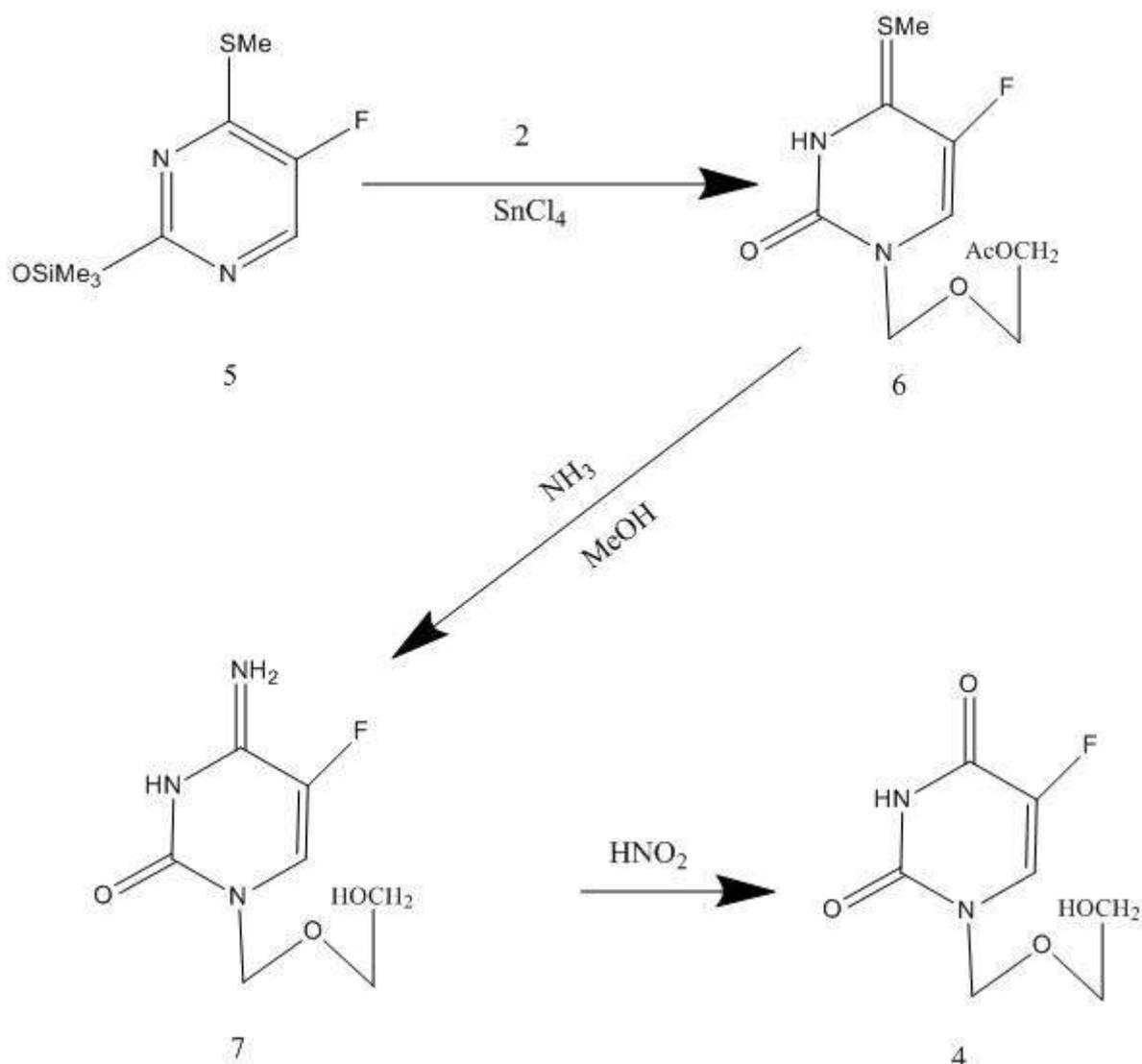


## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

### V.1.1.b A partir du 5-fluoro-4- (méthylthio) pyrimidin-2- (1H) -one

L'amine (7) a été préparée à partir de 5-fluoro-4- (méthylthio) pyrimidin-2- (1H) -one (5) par triméthylsilylation suivie d'une condensation avec le diacétate (2) en présence de chlorure stannique pour donner le produit (6). La désacylation simultanée et le remplacement du groupe méthylthio par un amino ont été effectués par chauffage dans l'ammoniac méthanolique.

Schéma (2)

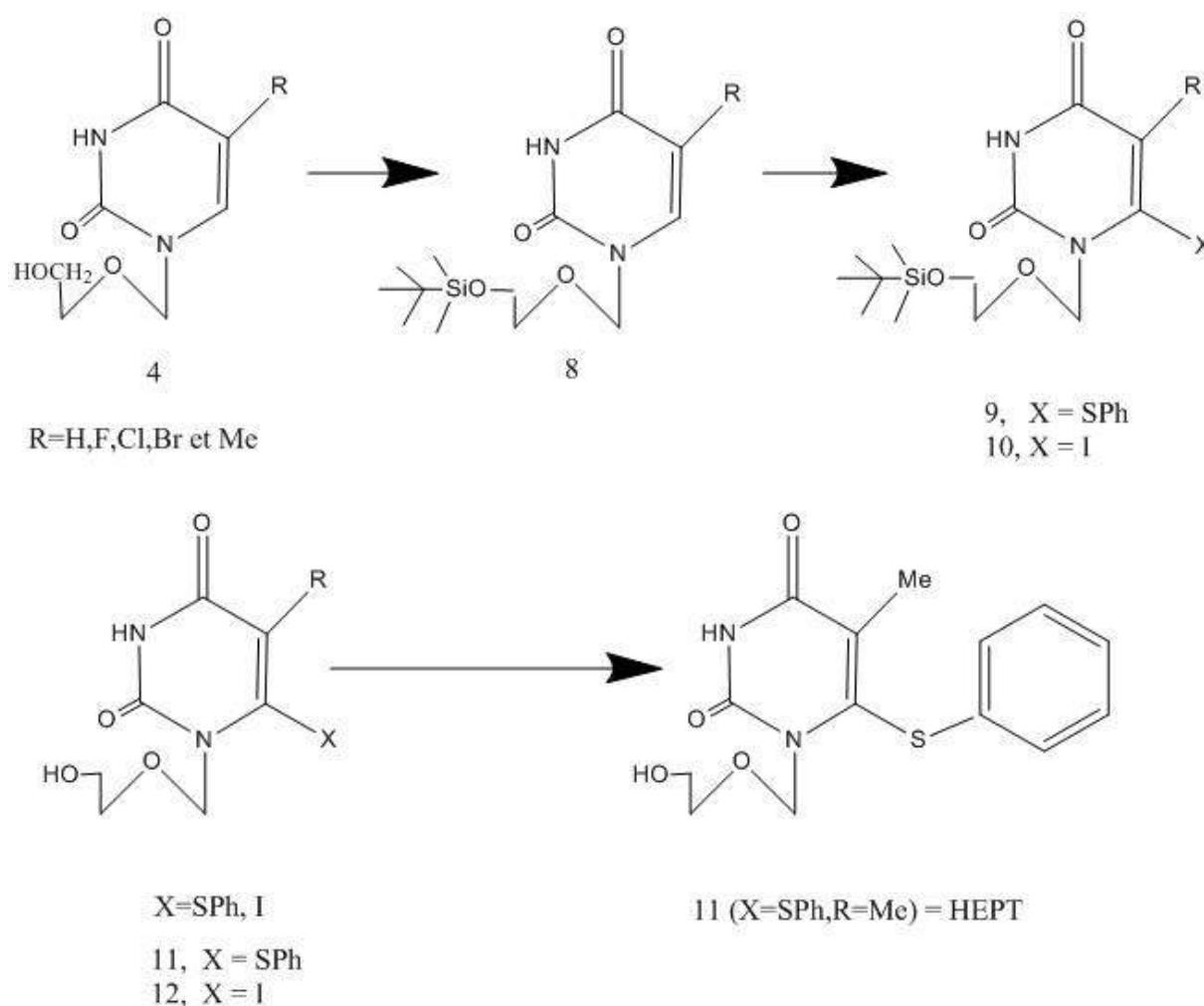


## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

Le groupe hydroxyle dans (4) a été silylé avec du chlorure de TBDMS dans le DMF en présence d'imidazole pour donner (8) (Schéma (3)). La lithiation de ces acylonucléosides protégés (8) a été effectuée dans du THF à une température  $-70^{\circ}\text{C}$  en utilisant du LDA. La solution obtenue de l'espèce lithiée en C-6 a ensuite été traitée avec le diphenyle électrophile correspondant disulfure ou iode, à  $-70^{\circ}\text{C}$  pendant 1 heure.

La purification par Chromatographie sur colonne du mélange réactionnel a donné le 6-phénylthio respectivement (9a, 81% ; 9b, 92% ; 9c, 26% ; 9d, 42% ; 9e, 73 %) et 6-iodo (10a, 76% ; 10b, 94% ; 10c, 68% ; 10d, 61% ; 10e, 79%), les composés 11 et 12 ont tous été obtenus sous forme cristalline.

Schéma (3)



## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

### V.2 Activité anti-VIH de la 1-2-[(Hydroxyethoxy) Methyl]-6-(Phenylthio) Thymine

L'activité anti-VIH et la cytotoxicité des composés (11a-e et 12a-e) sont montrées dans le tableau ci-dessous avec celles de l'AZT, la 2',3'-didésoxycytidine (ddCyd) et la 2', 3'-didésoxyadénosine (ddAdo). Parmi les composés présents dans le tableau.2, la 1 - [(2-hydroxyéthoxy) méthyl] -6- (phénylthio) thymine 11e (HEPT) s'est révélée comme l'agent le plus sélectif avec un indice de sélectivité de 106. Lorsque l'activité de HEPT a été comparée à celle des nucléosides connus, l'AZT et le ddCyd étaient plus actifs mais plus toxiques que le 11e dans les cellules MT-4 [46].

En termes d'activité et de cytotoxicité, HEPT était presque comparable à ddAdo. Lorsque 11e a été examiné pour son inhibition du VIH-2 dans les mêmes conditions d'essai que celles utilisées pour le VIH-1, il n'avait aucune activité à des concentrations allant jusqu'à 250  $\mu\text{M}$ .

Tableau.2 : Inhibition de la réplication du VIH-1 dans les cellules MT-4 par les composés 11 et 12

Composé	CE <sub>50</sub> <sup>a</sup> ( $\mu\text{M}$ )	CC <sub>50</sub> <sup>b</sup> ( $\mu\text{M}$ )	Si <sup>c</sup>
11a(R=H)	>500	1850	<3.7
11b(R=F)	>100	288	<2.9
11c(R=Cl)	>30	100	<3.4
11d(R=Br)	>20	44	<2.2
11e(R=Me)	7.0	740	106
12a(R=H)	>30	75	<2.5
12b(R=F)	>15	32	<2.1
12c(R=Cl)	>6	13	<2.2
12d(R=Br)	>5	11.5	<2.3
12e(R=Me)	>80	400	<5
AZT	0.016	20	1250
ddCyd	0.3	40	133
ddAdo	6.3	890	141

a : Concentration efficace du composé nécessaire pour atteindre la protection de 50% des cellules MT-4 contre l'effet cytopathique du VIH-1, b : Concentration cytotoxique du composé nécessaire pour réduire la viabilité de 50% des cellules MT-4 infectées et c : Indice de sélectivité : rapport CC<sub>50</sub> / CE<sub>50</sub>.

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

En conclusion, les résultats obtenus indiquaient clairement que le composé 11e peut être considéré comme un analogue nucléosidique qui manifeste son activité anti-VIH-1 par un mécanisme différent de celui connu pour d'autres analogues nucléosidiques. Étant donné que HEPT n'était pas très toxique contre les cellules de la moelle osseuse humaine in vitro à une concentration de 100  $\mu\text{M}$ , soit 20 fois à laquelle l'AZT est significativement toxique (5  $\mu\text{M}$ ). Les résultats obtenus pouvaient conduire à la découverte d'une nouvelle zone dans laquelle les chercheurs recherchaient des agents anti-VIH qui pourraient être considérablement moins toxiques.

### V.3 Synthèse des dérivés de HEPT

Pour synthétiser les dérivés de HEPT, Tanaka et ses collaborateurs ont mis au point une stratégie de synthèse qui a pris plusieurs années.

En 1990, une série de nouveaux dérivés d'acyclouridine variés à la fois aux positions C-5, C-6, N-1, C-2, N-3, et C-4 ont été synthétisés dans le but d'améliorer l'activité spécifique au VIH rapporté sur 1- [(2-hydroxyéthoxy) méthyl-6- (phénylthio) thymine.

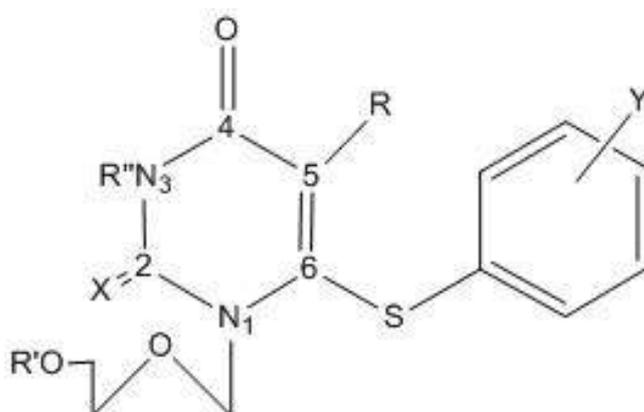


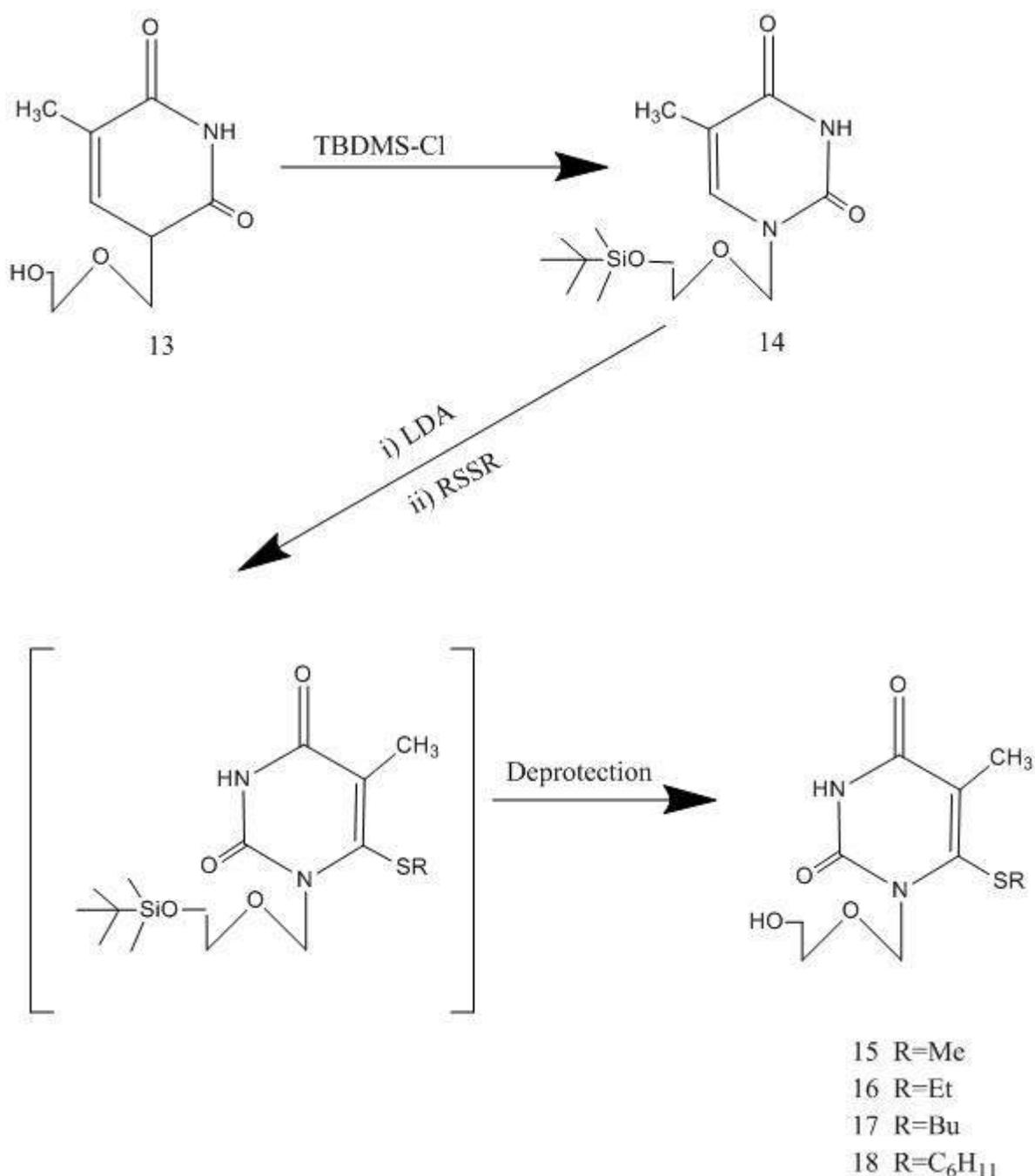
Figure.10 Structure de 1- [(2-hydroxyéthoxy) méthyl-6- Y (phénylthio) thymine.

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

### V.3.1. Modification en C-6

La préparation de dérivés substitués en C-6 a été réalisée sur la base des trois méthodes suivantes : (1) LDA (lithium diisopropylamide) lithiation d'un dérivé de thymine (14) suivie d'une réaction avec des électrophiles, (2) une réaction d'addition-élimination de HEPT, ou (3) le couplage croisé catalysé par le palladium entre un dérivé iodo et des alcynes terminaux [46].

Schéma (4)



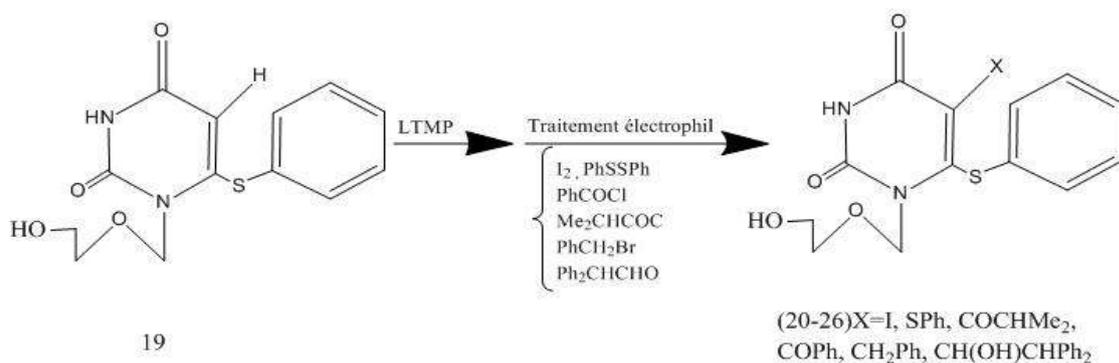
## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

La 1 - [[2- [(tert-butyldiméthylsilyl) oxy] éthoxy] méthyl] thymine (14) a été préparée à partir du composé connu (13) [47] avec du chlorure de tert-butyldiméthylsilyle (TBDMS-C1) et de l'imidazole. Le composé (14) a été lithié avec du LDA au-dessous d'une température de  $-70^{\circ}\text{C}$ . Le produit de lithiation est traité par une série de dialkyl disulfures inférieurs à une température  $-70^{\circ}\text{C}$ . Après un traitement acide du mélange réactionnel, les dérivés 6-alkylthio correspondants (15-18) ont été obtenus (Schéma (4)).

### V.3.2. Modification en C-5

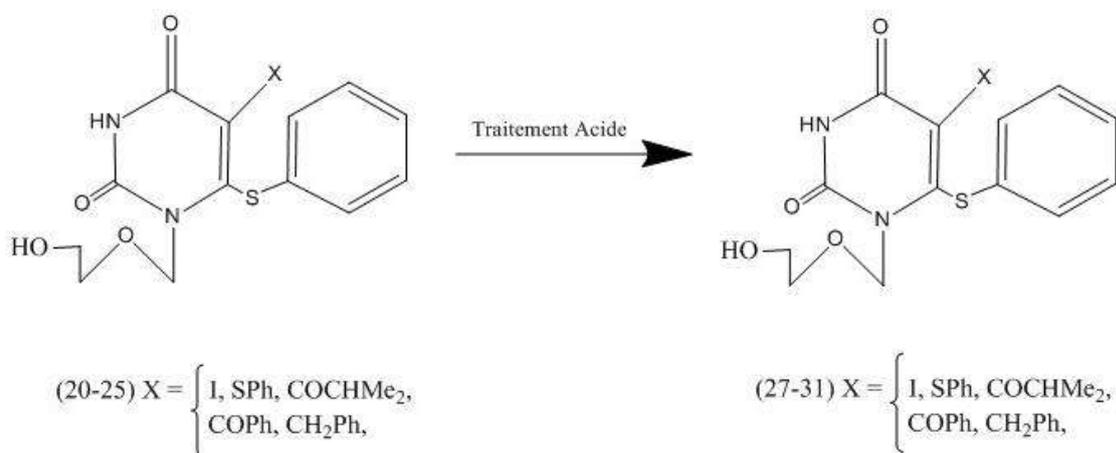
L'approche de lithiation a semblé être efficace également pour la modification à la position C-5 de HEPT. Cependant, un agent de lithiation plus basique, le lithium 2,2,6,6-tétraméthylpipéridide (LTMP) s'est avéré nécessaire. La lithiation du composé (19) avec du LTMP suivie d'un traitement avec un électrophile ( $\text{I}_2$ , PhSSPh,  $\text{Me}_2\text{CHCOCl}$ , PhCOCl, PhCH<sub>2</sub>Br et Ph<sub>2</sub>CHCHO) a donné les composés 20-26 [46].

Schéma (5)



Le traitement acide des composés 20-25 a fourni les composés 27-31.

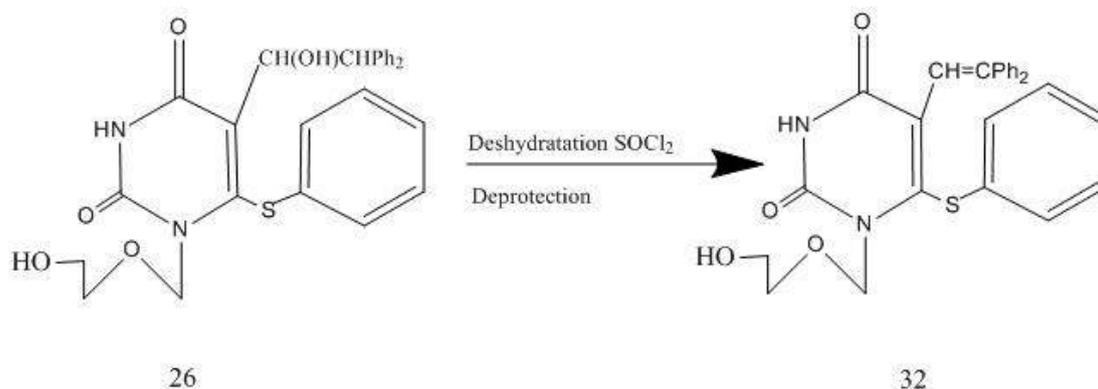
Schéma (6)



## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

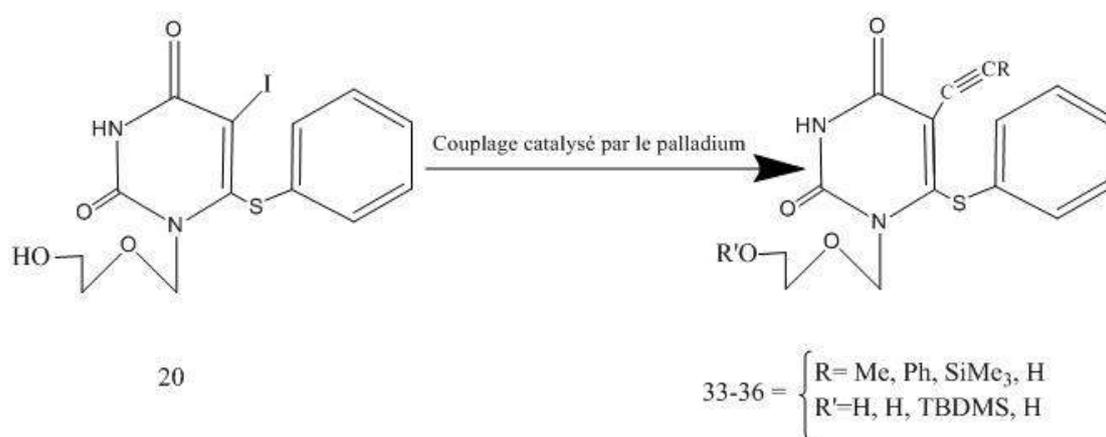
La déshydratation du composé (26) avec  $\text{SOCl}_2$  dans la pyridine suivie de la déprotection a donné le dérivé 5-(2,2-diphénylvinyle) (32).

Schéma (7)



L'application du couplage croisé catalysé au palladium mentionné ci-dessous a été effectuée en utilisant le dérivé 5-iodo (20) suivi de déprotection pour donner les composés (33-36).

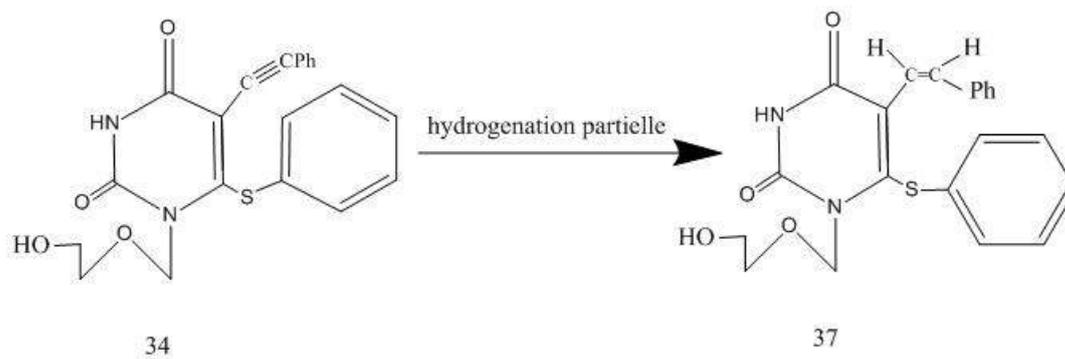
Schéma (8)



## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

Une hydrogénation partielle du composé (34) a donné le composé (37)

Schéma (9)



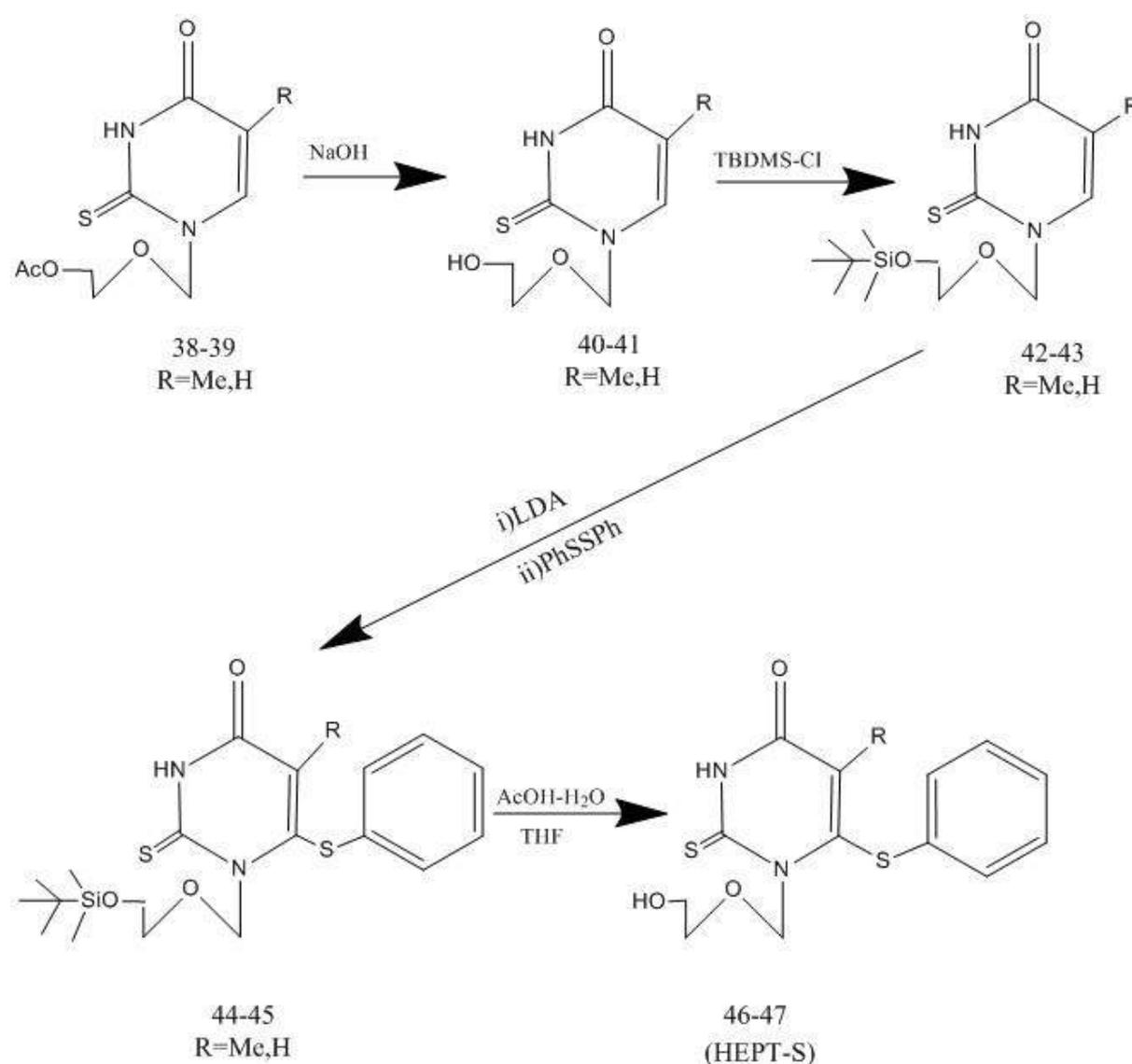
## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

### V.3.3 Modification en C-2

La préparation des dérivés de 6- (phénylthio) -2-thioacycloïdine a été effectuée comme indiqué sur le schéma (10). Les composés (38) et (39) [48] ont été désacétylés pour donner les 2-thioacycloïdines libres respectives (40) et (41). Pour la lithiation par LDA, le groupe hydroxyle des composés (40) et (41) a été protégé avec un groupe tert-butyldiméthylsilyle (TBDMS) pour donner les composés (42) et (43).

Les composés (42) et (43) ont été traités avec LDA (phényle disulfure à une température de -70 ° C) pour donner les produits (44) et (45) comme indiqué dans le schéma ci-dessous. [49]. La désilylation des composés (44) et (45) dans AcOH-THF-H<sub>2</sub>O a donné les résultats correspondants 2-thioacycloïdines libres (46) et (47) [50].

Schéma (10)



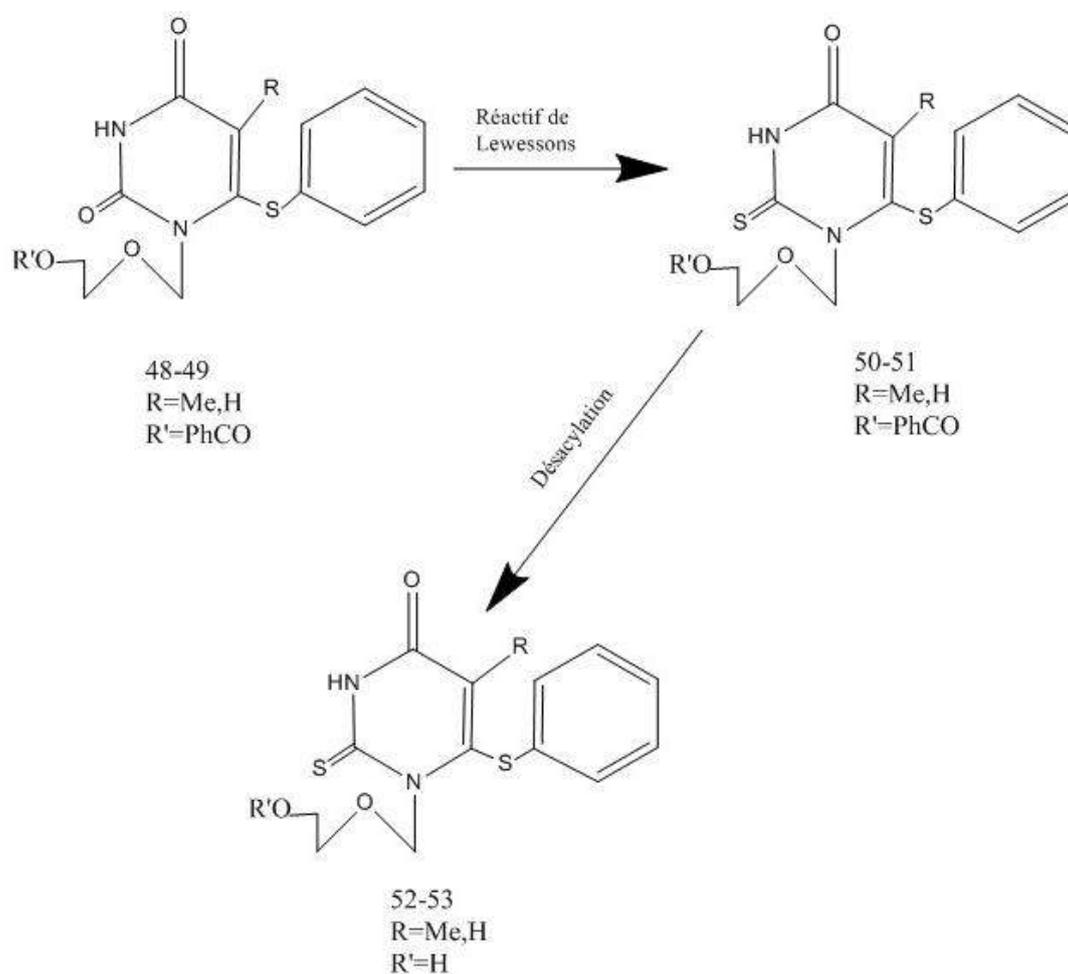
## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

### V.3.4 Modification en N-1

Pour la préparation des équivalents 4-thio, les dérivés 0-benzoyle (48) et (49) des 6-(phénylthio) - acyclonucléosides préformés ont été traités avec du 2,4-bis (4-méthoxyphényl) -1,3-dithia-2, 4-diphosphétane 2,4-disulfure appelé réactif de Lawesson [51].

Suite à la désacylation des produits (50) et (51), les composés cibles (52) et (53) ont été obtenus [50].

Schéma (11)



## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

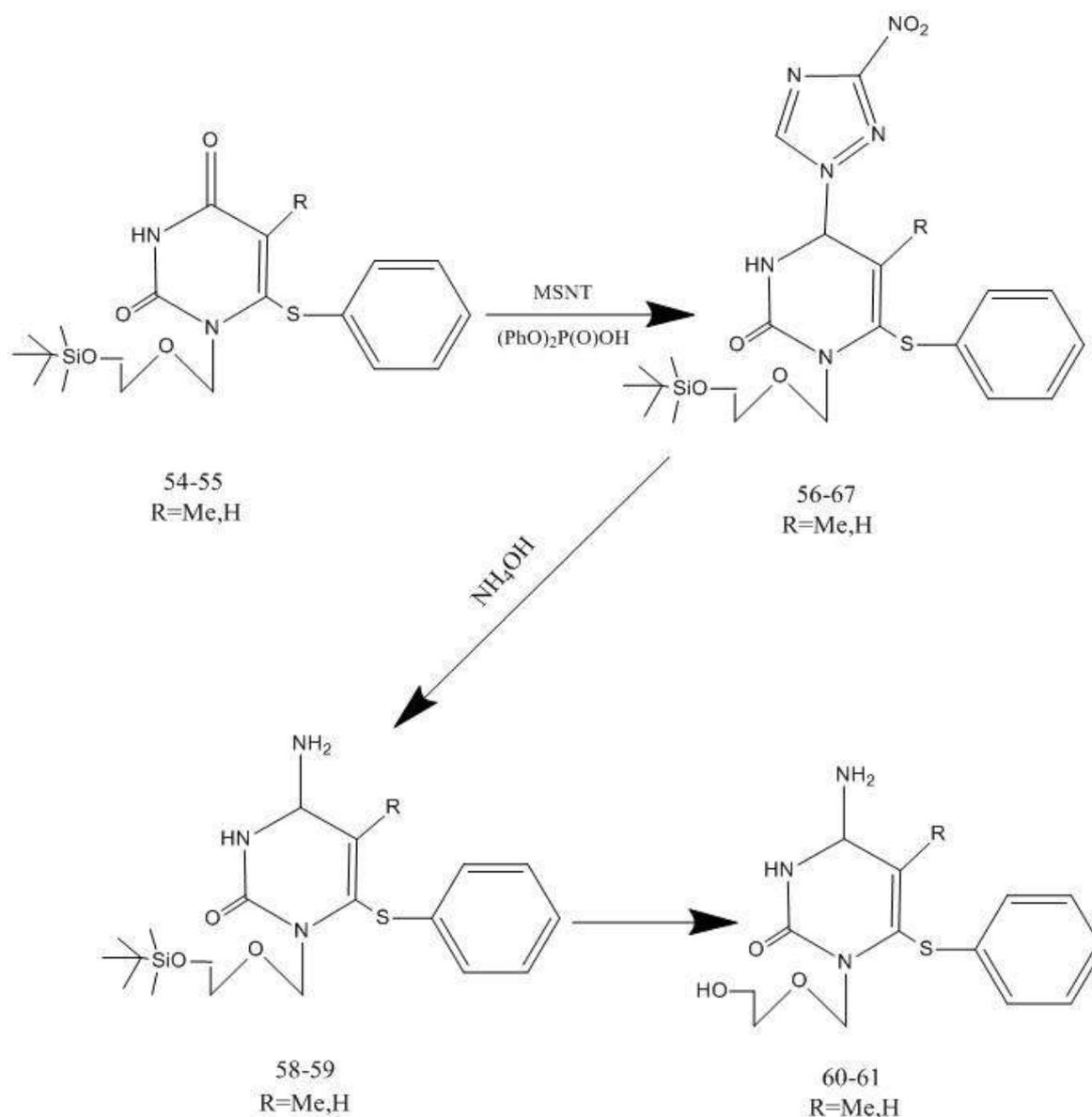
### V.3.5 Modification en C-4

Pour l'introduction d'un groupe amino à la position C-4 des composés (54) et (55) par une méthode déjà publiée[52-53], on a fait réagir le composé (54) avec du 1- (mésitylène-2-sulfonyl) -3- nitro-1,2,4-triazole (MSNT) dans de la pyridine en présence de phosphate de diphenyle (schéma 12). [54]

Le mélange a été traité avec de l'ammoniaque dans du dioxane, cela a permis l'isolement du composé (57) à partir du composé (54). Contrairement au cas du composé (55), le dérivé 4-(3-nitro-1,2,4-triazol-1-yl) (58) a été obtenu à partir du composé (55).

Cependant, lors de l'ammonolyse a donné trois produits. Sur la base de leur masse et des spectres RMN <sup>1</sup>H et des analyses, les structures ont été confirmées à (59) (57%), (60) (21%) et (61) (20%).

Schéma (12)

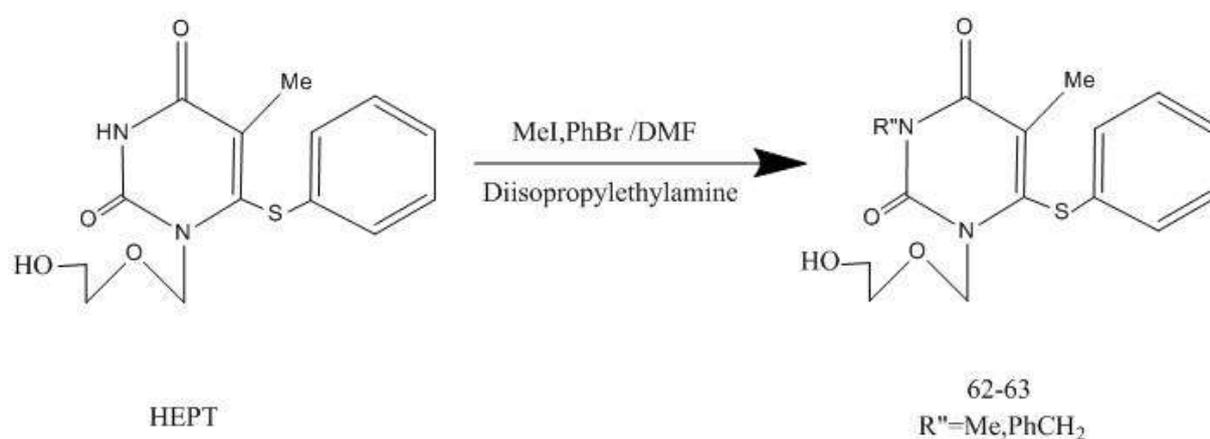


## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

### V.3.6 Modification en N-3

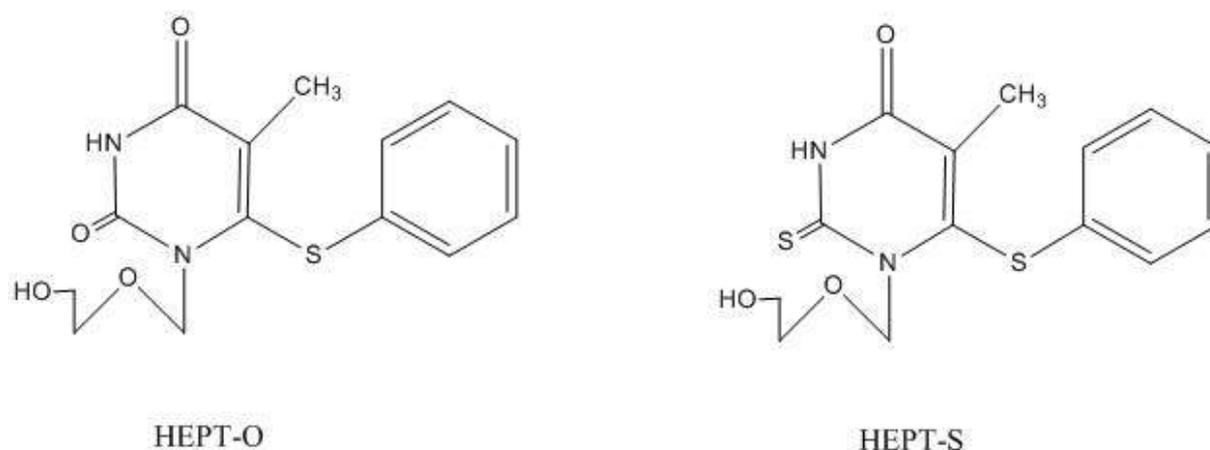
Enfin, la préparation des dérivés N-3 méthyle (62) et N-3 benzyle (63) a été réalisée par alkylation de HEPT par MeI ou bromure de benzyle dans le DMF en présence de diisopropyléthylamine [50].

Schéma (13)



### V.3.7 Modification sur la position Y

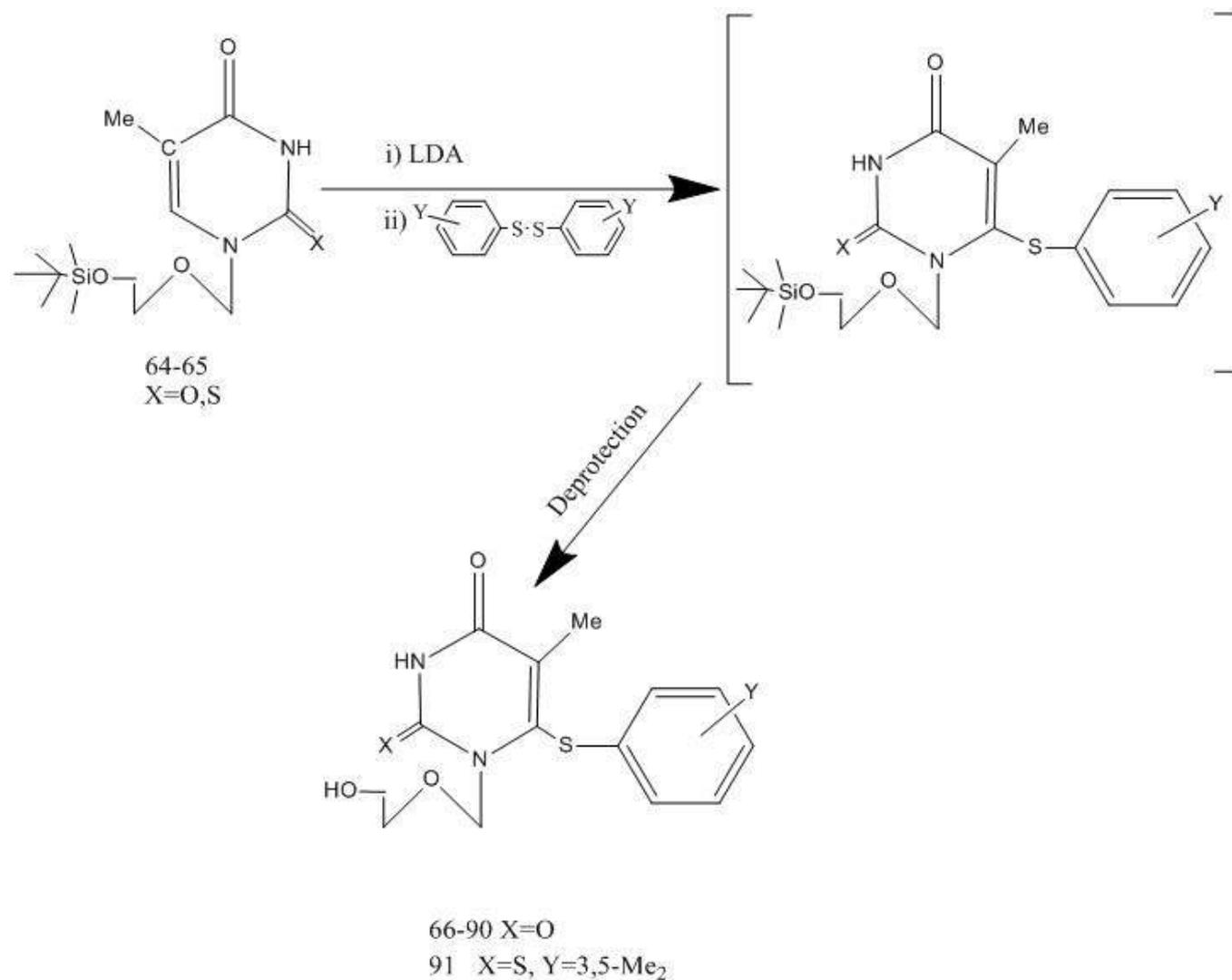
En 1991, une grande majorité des analogues HEPT-O et HEPT-S modifiés dans le cycle phényle C-6 (Y) ont été synthétisés à partir de dérivés de thymine et de 2-thiothymine (64) et (65) protégés par le tert-butyldiméthylsilyle (TBDMS), par une lithiation LDA (lithium diisopropylamide) [54].



## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

Le traitement du (64) ou (65) avec LDA dans le THF en dessous d'une température de -70 °C a généré les espèces lithiées C-6 respectives (66)-(91) [46].

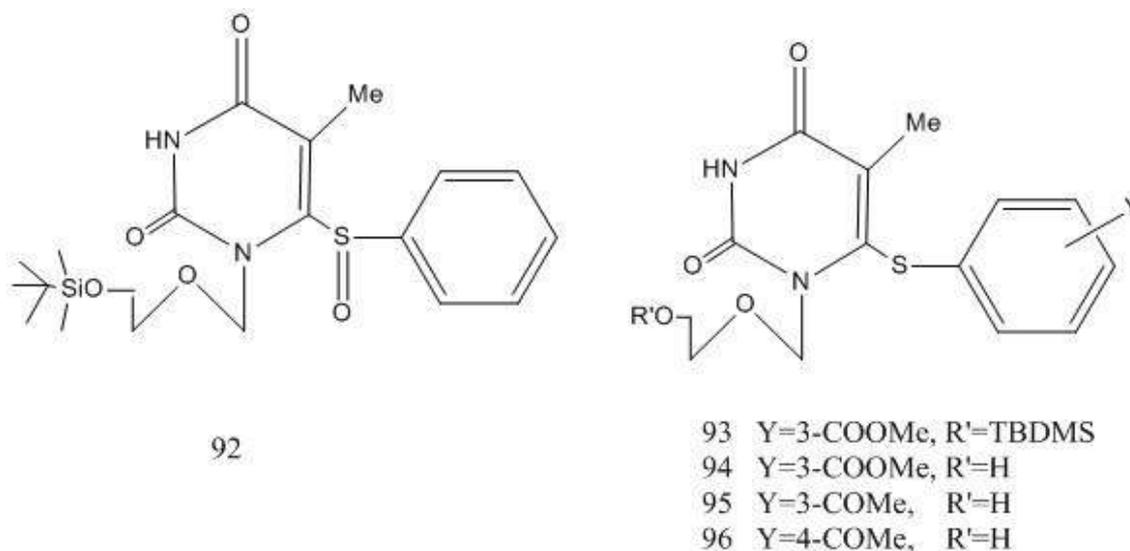
Schéma (14)



Y=2-Me, 2-Cl, 2-NO<sub>2</sub>, 2-OMe, 3-Me, 3-Et, 3-t-Bu, 3-CH<sub>2</sub>OH, 3-CF<sub>3</sub>, 3-F, 3-Cl, 3-Br, 3-I, 3-NO<sub>2</sub>, 3-OH, 3-OMe, 4-Me, 4-F, 4-Cl, 4-NO<sub>2</sub>, 4-CN, 4-OH, 4-OMe, 3,5-Me<sub>2</sub>, 3,5-Cl<sub>2</sub>

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

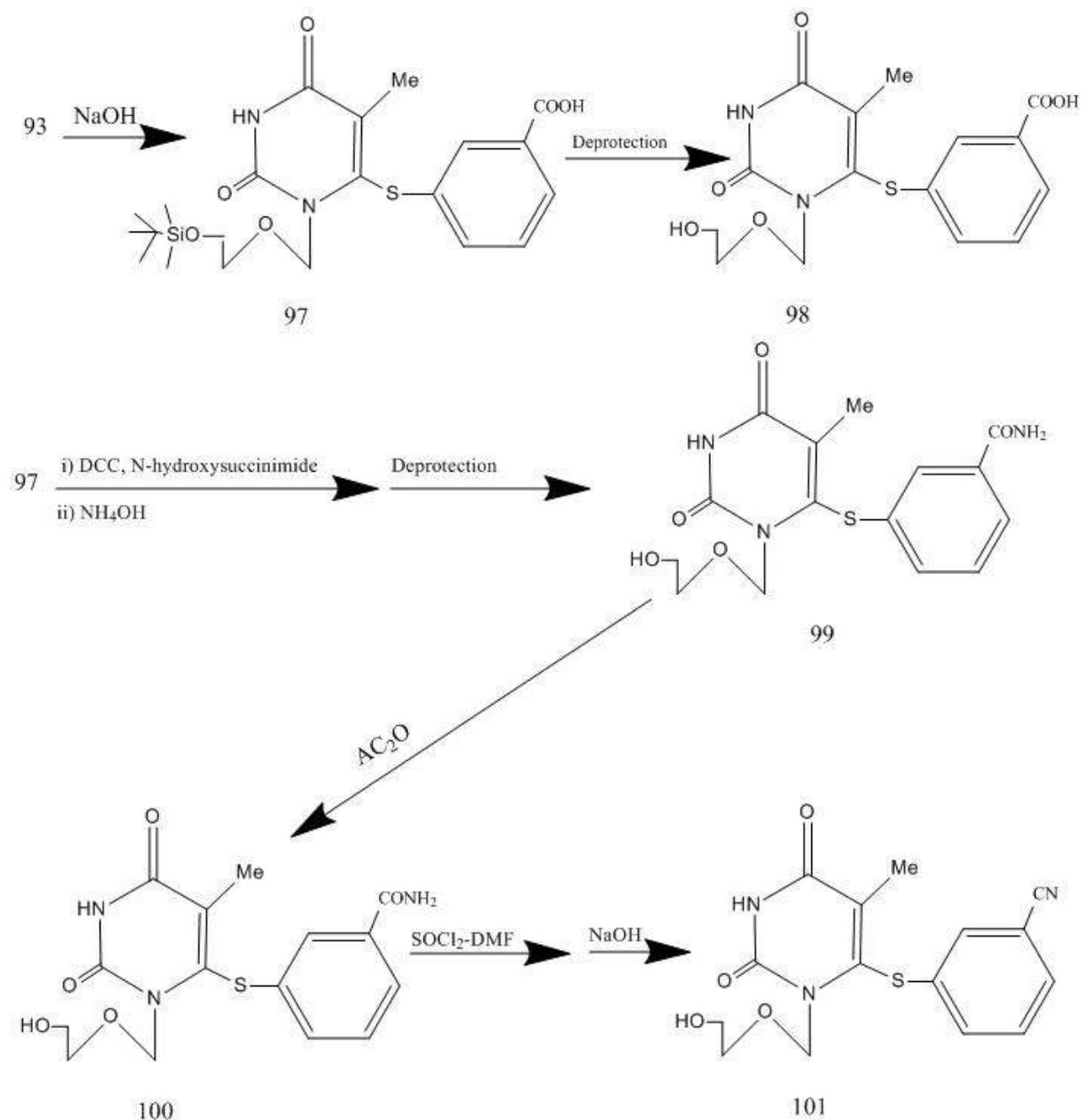
Schéma (15)



Lorsque le composé (92) a été traité avec du 3- (méthoxycarbonyl) benzènethiolate de sodium dans du DMF à température ambiante, le (93) s'est formé avec un rendement de 48%. Après élimination du groupe TBDMS, l'acyclothymidine (94) a été obtenue. Les composés (95) et (96) ont été préparés à partir de (92) d'une manière similaire.

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

Schéma (16)

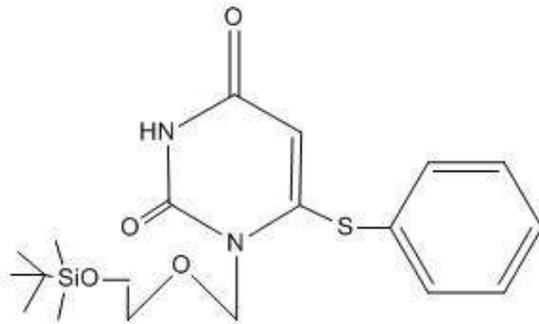


La transformation conventionnelle du groupe méthoxycarbonyle du (93) a également été réalisée comme indiqué sur le schéma (16). L'hydrolyse alcaline a donné (97), qui a ensuite été désilylée pour donner la 6 - [(3-carboxyphényl) thio] acyclocthamine (98). L'amidation de (97) a été réalisée via un intermédiaire d'ester actif. Après élimination du groupe TBDMS, on a obtenu le dérivé 6- [(3-carbamoylphényl) thio] (99).

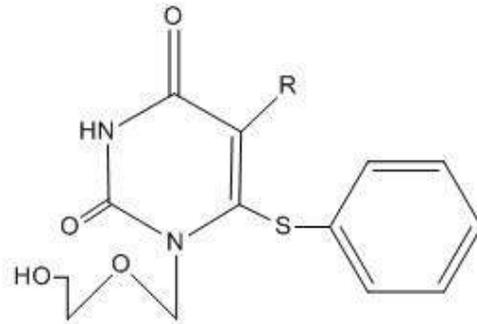
La déshydratation du groupe carbamoyle a été effectuée en utilisant un dérivé O-acétylé (100) pour donner la 6- [(3-cyanophényl) thio]acyclocthamine (101).

**Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH**

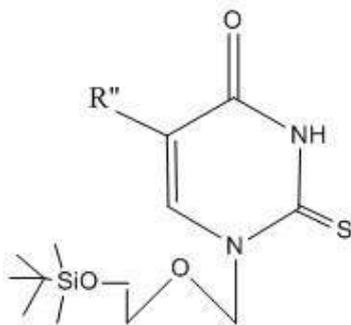
Schéma (17)



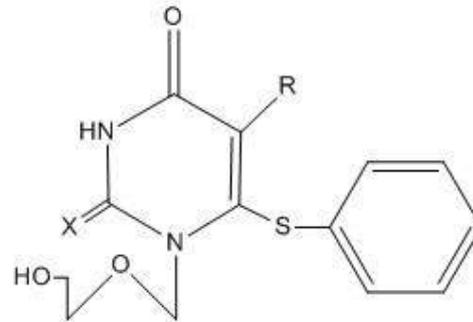
102



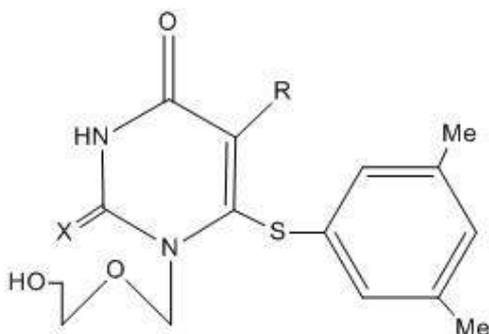
103 R=   
 104 R=COOMe  
 105 R=NHCOPh



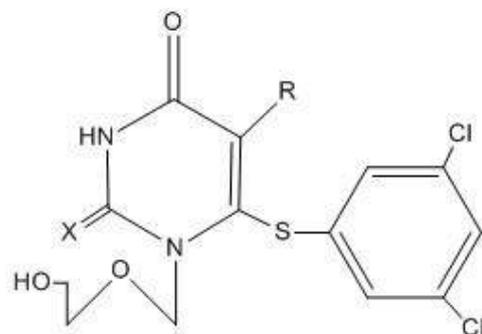
106 R''=Et  
 107 R''=Pr  
 108 R''=i-Pr



109 R=Et, X=S  
 110 R=Pr, X=S  
 111 R=i-Pr, X=S  
 115 R=Et, X=O  
 116 R=Pr, X=O  
 117 R=i-Pr, X=O



112 R=Et, X=S  
 113 R=i-Pr, X=S  
 118 R=Et, X=O  
 119 R=i-Pr, X=O



114 R=Et, X=S  
 120 R=Et, X=O

## **Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH**

LTMP (lithium 2,2,6,6-tétraméthylpipéridide) été utilisé pour la lithiation C-5 de la 6-(phénylthio) uridine. Lorsque le (102) été lithié avec LTMP, puis a réagi avec des électrophiles ( $\text{BrCH}_2\text{CH} = \text{CH}_2$ ,  $\text{ClCO}_2\text{Me}$  et  $\text{PhNCO}$ ), les produits (103)-(105) ont été obtenus après déprotection. Les dérivés (106)-(108) protégés par 0-TBDMS ont été préparé à partir de 5-alkyl-2-thiouracil, puis ils ont été soumis à une lithiation par LDA.

L'utilisation de diaryl disulfures appropriés a permis l'obtention des analogues 5-alkyl-2-thio (109)-(114) après déprotection. Les analogues d'uracile correspondants (115)-(120) ont été préparés par hydrolyse oxydative ( $\text{H}_2\text{O}_2$  en  $\text{NaOH}$  aqueux) de la fonction thione de (109)-(114) en solution.

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

### V.4 Activité anti-VIH des dérivés de HEPT

#### Activité biologique des dérivés HEPT-O

Généralement, une variation de l'inhibition de la réplication du VIH dans les cellules MT-4 par des dérivés de HEPT-O qui subissent des modifications du substituant en position 5.

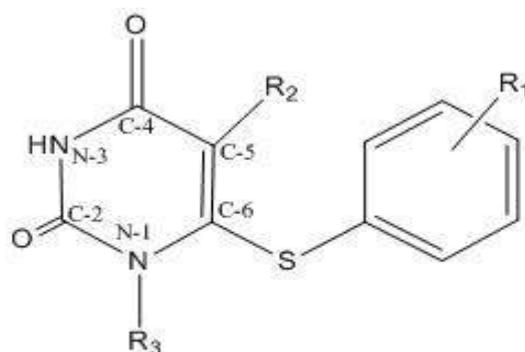


Tableau.3 : L'activité anti-VIH des dérivés de HEPT-O modifiés en C-5

N°	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	CE <sub>50</sub> (μM)
30	H	i-Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	0,063
28	H	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	0,120
35	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,079
37	H	I	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	3,630
104	H	Br	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	19,952
103	H	Cl	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	30,199
102	H	F	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	100,000
32	3,5-Me <sub>2</sub>	i-Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	0,003
31	3,5-Me <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	0,013
15	3,5-Me <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	0,257

Le tableau.3 permet de comparer la variation de l'activité anti-VIH du HEPT-O avec le changement des radicaux R<sub>1</sub> et R<sub>2</sub>, R<sub>3</sub> étant maintenu constant et qui est CH<sub>2</sub>OCH<sub>2</sub>CH<sub>2</sub>OH. Des trois premières lignes de ce tableau, on note que quand le phényle porté par l'atome de soufre n'est pas substitué, l'activité anti-VIH des molécules augmente avec la variation le radical R<sub>2</sub> de méthyle, éthyle puis isopropyl [55-56]

Le sens de cette variation est opposé à celle de l'effet inductif dans ces radicaux. On note que la variation de l'activité anti-VIH lorsque R<sub>2</sub> passe de méthyle à éthyle est largement plus importante que celle correspondante au passage de R<sub>2</sub> de éthyle à isopropyl.

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

Ceci peut être relié à l'augmentation de la chaîne carbonée de R<sub>2</sub> qui induit un effet stérique sur le comportement conformationnel des molécules. On note que l'activité anti-VIH est multipliée par 10 lorsque thiophényl est substitué dans les positions méta par un méthyle.

En variant le radical R<sub>3</sub> de la position N-1, on remarque une amélioration spectaculaire au niveau de l'activité biologique de certaines dérivés.

Tableau.4 : L'activité biologique des dérivés de HEPT-O modifiés en N-1

N°	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	CE <sub>50</sub> (µM)
66	H	Me	Bu	1,202
65	H	Me	Et	2,188
107	H	Me	Me	151,356
106	H	Me	H	251,189
35	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,079

Le tableau.4 permet d'étudier l'effet et le changement du radical en N-1 sur l'activité anti-VIH des dérivés HEPT-O ayant des thiophényl en C-6 et un méthyle en C-5.

On note alors que la concentration CE<sub>50</sub> augmente avec la longueur de la chaîne et l'effet inductif du radical en N-1. Dans cette série de molécules, la variation de cette concentration diminue fortement quand on passe de zéro à 1 puis 2 atomes de carbone. Au de-là de 2 atomes de carbone, l'augmentation de la chaîne ne présente qu'un faible effet. L'introduction des atomes d'oxygène sous forme de fonction éther et alcool terminal permet d'augmenter faiblement la concentration même avec une chaîne ayant plus de deux atomes de carbone.

Pour cette série de molécules, la modification au niveau du radical R<sub>3</sub> influe l'activité anti-VIH. La substitution du radical CH<sub>2</sub>OCH<sub>2</sub>CH<sub>2</sub>OH par un hydrogène conduit à l'utilisation d'une grande quantité de la molécule et donc une baisse de l'activité.

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

Tableau.5 : L'activité biologique des dérivés de HEPT-O modifiés en C-6

N°	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	CE <sub>50</sub> (µM)
15	3,5-Me <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	0,257
16	3,5-Cl <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	1,288
4	3-Me	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	2,570
5	3-Et	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	2,691
8	3-F	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	3,311
10	3-Br	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	5,754
19	3-COMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,244
18	3-COOMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,943
11	3-I	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	10,000
20	3-CN	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	10,000
6	3-t-Bu	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	12,022
9	3-Cl	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	12,882
2	2-OMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	19,054
14	3-OMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	21,877
12	3-NO <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	33,884
7	3-CF <sub>3</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	44,668
1	2-Me	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	70,794
90	4-COMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	109,647
2	2-NO <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	141,253
34	4-Me	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	218,776
87	4-CN	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	251,188
89	4-OMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	251,188
88	4-OH	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	275,422

Le tableau.5 permet de voir la variation de l'activité biologique pour une série de HEPT-O ayant un méthyle en R<sub>2</sub>, un CH<sub>2</sub>OCH<sub>2</sub>CH<sub>2</sub>OH en R<sub>3</sub> et le changement du R<sub>1</sub>. Il ressort dès résultats dressé dans ce tableau que pour une monosubstitution du thiophényl, la position méta permet d'avoir les molécules les plus efficaces. Parmi tous les radicaux utilisés, la molécule ayant un méthyle en position méta du thiophényl est la meilleure candidate pour l'inhibition de la TI du VIH.

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

L'augmentation de la chaîne et la variation des effets électroniques du groupement en méta du thiophényl est en défaveur de l'activité biologique de ce type des dérivés HEPT-O. Quand le thiophényl est substitué deux fois en position méta l'efficacité est maximale [57].

L'ajout d'un deuxième méthyle en position méta ne présente pas la même influence. Ce qui suggère que l'effet du méthyle en position méta dépend aussi de ces interactions avec les autres groupements de la molécule.

Ce pendant, l'ajout d'un chlore permet d'augmenter radicalement l'efficacité de la molécule, puisque la molécule avec un seul chlore en position méta du thiophényl a une mauvaise activité anti-VIH. L'interprétation de tels résultats nécessitent des calculs de structures et de propriétés électroniques des molécules. Un autre résultat qui attire l'attention est que la monosubstitution du thiophényl en position ortho et méta offrent des efficacités presque équivalente alors que la substitution en position para donne une assez haute CE<sub>50</sub>.

Dès résultats collectés dans le tableau.6, on peut suivre la variation de l'activité des molécules HEPT-O avec un thiophényl non substitué et portant un méthyle sur le radical R<sub>2</sub> avec la variation du groupement d'atome (Z) final du radical R<sub>3</sub> du type CH<sub>2</sub>OCH<sub>2</sub>CH<sub>2</sub>-Z [58]. Le meilleur résultat est celui de la molécule avec Z = H. Il est à noter que le simple allongement de la chaîne Z=Me et le changement des propriétés électroniques de la queue du groupement R<sub>3</sub> (Z = F, Cl, N<sub>3</sub>, OH) abaisse l'activité anti-VIH de ces dérivés HEPT-O.

Tableau.6 : Les dérivés de HEPT-O modifiés en N-1

N°	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	CE <sub>50</sub> (μM)
45	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>3</sub>	0,331
48	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> F	1,096
46	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> Cl	1,513
49	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> Me	3,311
47	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> N <sub>3</sub>	5,754
35	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,079

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

### Activité biologique des dérivés HEPT-S

Tableau.7 : L'activité anti-VIH des dérivés HEPT-S comparé aux dérivés HEPT-O

N°	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	X	CE <sub>50</sub> (µM)
30	H	i-Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	0,063
24	H	i-Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	0,059
28	H	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	0,120
22	H	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	0,110
35	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	7,079
36	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	0,977

Le tableau.7 regroupe les activités anti-VIH (CE<sub>50</sub>) des HEPT-S ayant un thiophényl non substitué et où N-1 porte un alkyle ayant des longueurs de chaîne différente. Comme les HEPT-O du même type, on note une augmentation de l'activité anti-VIH avec la longueur de la chaîne. On note aussi que le volume de R<sub>2</sub> n'a pas une influence importante sur cette activité. Par ailleurs, on note que le remplacement de l'oxygène du groupement oxo par un atome de soufre permet une faible amélioration quand le R<sub>2</sub> est éthyle ou isopropyl alors que cette amélioration est importante pour R<sub>2</sub> = méthyle. Ainsi, il serait intéressant d'étudier les éventuelles interactions entre l'oxygène ou le soufre avec le R<sub>2</sub>.

Finalement, la présence de deux groupements méthyle en position méta sur le cycle phényle améliore l'activité de dérivés soufrés.

Tableau.8 : L'activité biologique des dérivés de HEPT-S modifiés en C-5 et C-6

N°	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	CE <sub>50</sub> (µM)
24	H	i-Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	0,059
22	H	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	0,110
36	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	0,977
26	3,5-Me <sub>2</sub>	i-Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	0,005
25	3,5-Me <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	0,008
17	3,5-Me <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	0,219

## **Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH**

Dès résultats du tableau.8, on note que la substitution par deux méthyles dans les deux positions méta du thiophényl conduit a cette amélioration de l'activité anti-VIH des HEPT-S. Cette activité augmente avec les longueurs de la chaîne alors que le volume du groupe  $R_2$  n'a qu'un faible effet. Dans ce cas ou  $R_2$  est un méthyle, on note une très faible amélioration de l'activité anti-VIH lorsqu'on remplace l'oxygène du groupement oxo par un atome de soufre.

## *Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH*

### **V. Conclusion**

La mise au point de nouveaux médicaments antiviraux contre le VIH est toujours d'actualité mais les nouveaux traitements se heurtent à une barrière fondamentale qui est l'apparition de résistance envers des molécules qui ont un mode d'action similaire.

L'influence de la nature, le nombre et la position des groupements sur les dérivés de HEPT a été étudié. En particulier, nous avons étudié la relation entre ce changement et l'activité biologique des molécules (quantité de médicaments).

Dans un premier temps, on a étudié la variation des substituants sur les molécules ayant un oxygène en position X (HEPT-O). Ces changements ont été élaborés pour différentes séries de HEPT-O, la première variation était pour le radical R<sub>2</sub> (nombre de la chaîne carbonée et le type) vis-à-vis la quantité de médicament. Pour la longueur de la chaîne, le changement est significatif et il a apparue sur la CE<sub>50</sub>, aussi pour l'introduction d'un halogène, la variation est spectaculaire qu'on passe d'un Iode à un Fluor. Pour les HEPT-O substitués en N-1, on a remarqué une forte diminution lors du passage de zéro à 1 puis à 2 atomes de carbones.

Cependant, une influence significative a été observée pour les HEPT-O ayant une modification au niveau du radical R<sub>1</sub>, aussi, quand le thiophényl est substitué deux fois, l'efficacité est maximale.

Dans le cas des molécules ayant un soufre en position X (HEPT-S), l'étude a été établie sur plusieurs ensembles. On note que la substitution par deux méthyles dans les deux positions méta du thiophényl améliore l'activité anti-VIH de ces composés.

La plupart des études exploraient les modifications d'une manière conjointe. D'autres modifications structurelles plus profondes, notamment par la création de cycles supplémentaires ont été étudiées mais sans apporter de résultats dans les études in vitro [59-60].

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

### Bibliographie

- [1]. Feschotte, C., & Gilbert, C. (2012). Endogenous viruses: insights into viral evolution and impact on host biology. *Nature Reviews Genetics*, 13(4), 283.
- [2]. [https://www.news-medical.net/health/virus-origins-\(french\).aspx](https://www.news-medical.net/health/virus-origins-(french).aspx)
- [3]. Gottliebms, schroff r, schanker hm, weisman jd, fan pt, wolf ra, saxon a (1981) n engl j med 305:1425
- [4]. Unaid, aids epidemic update, 2005 <http://www.unaids.org> (accessed 2/9/2006)
- [5]. Gallo rc, sarin ps, gelmann ep, robert-guroff m, richardson e, kalyanaraman vs, mann d, sidhu gd, stahl re, zolla-pazner s, leibowitch j, popovicm (1983) science 220:865
- [6]. Barre-sinoussi f, chermann jc, rey f, nugeyremt, chamaret s, gruest j, dauguet c, axler-blin c, vezinet-brun f, rouzioux c, rozenbaum w, montagnier l (1983) science 220:868
- [7]. Sarngadharan mg, popovic m, bruch l, schupbach j, gallo rc (1984) science 224:506
- [8]. Levy ja, hoffman ad, kramer sm, landis ja, shimabukuro jm, oshiro ls (1984) science 225:840
- [9]. Coffin j, haase a, levy ja, montagnier l, orozlan s, teich n, temin h, toyoshima k, varmus h, vogt p (1986) science 232:697
- [10]. Clavel f, guetard d, brun-vezinet f, chamaret s, rey ma, santos-ferreira mo, laurent ag, dauguet c, katlama c, rouzioux c (1986) science 233:343
- [11]. Fauci as (1988) science 239:617
- [12]. De clerq e (1995) clin microbiol rev 8:200
- [13]. De clerq e (2002) biochim biophys acta 258:1587
- [14]. Gullick rm (2003) clin microbiol infect 9:186
- [15]. Deshpande, A., Jauvin, V., Pinson, P., Jeannot, A. C., & Fleury, H. J. (2009). Phylogenetic analysis of HIV-1 reverse transcriptase sequences from 382 patients recruited in JJ Hospital of Mumbai, India, between 2002 and 2008. *AIDS research and human retroviruses*, 25(6), 633-635.
- [16]. <http://planet-vie.ens.fr/article/1463/virus-sida>.
- [17]. Mcmichael, a.; t. Hanke, hiv vaccines 1983-2003. *Nature medicine*, 2003. 9(7): p. 874- 880.
- [18]. Blankson, j.n.; d. Persaud; r.f. Silciano, the challenge of viral reservoirs in hiv-1 infection. *Annu rev med.*, 2002. 53: p. 557-593
- [19]. Mcmichael, a., t-cell responses and viral escape. *Cell.*, 1998. 93: p. 673-676.
- [20]. Barouch, d.h.; e. Al, control of viremia and prevention of aids in rhesus monkeys by cytokine-augmented dna vaccination. *Science*, 2000. 290: p. 486-492
- [21]. Jacobo-molina, a.; j. Ding; r.g. Nanni; a.d. Clark; x. Lu; c. Tantillo; r.l. Williams; g. Kamer; a.l. Ferris; p. Clark; a. Hizi; s.h. Hughes; e. Arnold, crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded dna at 3.0Å resolution shows bent dna. *Proc natl acad sci u s a.*, 1993. 90: p. 6320-6324.
- [22]. Hirschel, K. D. V. S. B. (2005). Sida Molécules récentes, stratégies thérapeutiques d'avenir et nouvelles cibles du traitement. *Rev Med Suisse*, 1, 30010.

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

- [23]. Smith JS, Roth MJ. Purification and characterization of an active human immunodeficiency virus type 1 RNase H domain. *J Virol* 1993; 67: 4037-49.
- [24]. Smith JS, Gritsman K, Roth MJ. Contributions of DNA polymerase subdomains to the RNase H activity of human immunodeficiency virus type 1 reverse transcriptase. *J Virol* 1994; 68: 5721-9.
- [25]. Becerra SP, Clore GM, Gronenborn AM, et al. Purification and characterization of the RNase H domain of HIV-1 reverse transcriptase expressed in recombinant *Escherichia coli*. *FEBS Lett* 1990; 270: 76-80.
- [26]. Davies 2nd JF, Hostomska Z, Hostomsky Z, Jordan SR, Matthews DA. Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science* 1991; 252: 88-95.
- [27]. Nanni, r.g., ding, j., jacobo-molina, a., hughes, s.h. Et arnold, e., review of hiv-1 reverse transcriptase three-dimensional structure: implications for drug design (1993), perspectives in drug discovery and design, 1:129-150
- [28]. Varmus, h., reverse transcription (1987), scientific american, 257:48-54
- [29]. Kohlstaedt, I.a., wang, j., friedman, j.m., rice, p.a. Et steitz, t.a., crystal structure at 3.5 .ang. Resolution of hiv-1 reverse transcriptase complexed with an inhibitor (1992), science (washington, d.c., 1883-), 256:1783-1790
- [30]. Yarchoan, R., Weinhold, K., Lyerly, H. K., Gelmann, E., Blum, R., Shearer, G., ... & Markham, P. (1986). Administration of 3'-azido-3'-deoxythymidine, an inhibitor of HTLV-III/LAV replication, to patients with AIDS or AIDS-related complex. *The Lancet*, 327(8481), 575-580
- [31]. Balzarini, J., Baba, M., Pauwels, R., Herdewijn, P., & De Clercq, E. (1988). Anti-retrovirus activity of 3'-fluoro-and 3'-azido-substituted pyrimidine 2', 3'-dideoxynucleoside analogues. *Biochemical pharmacology*, 37(14), 2847-2856
- [32]. Fischl, M. A., Richman, D. D., Grieco, M. H., Gottlieb, M. S., Volberding, P. A., Laskin, O. L., ... & Jackson, G. G. (1987). The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. *New England Journal of Medicine*, 317(4), 185-191.
- [33]. Volberding, P. A., Lagakos, S. W., Koch, M. A., Pettinelli, C., Myers, M. W., Booth, D. K., ... & Murphy, R. L. (1990). Zidovudine in asymptomatic human immunodeficiency virus infection: a controlled trial in persons with fewer than 500 CD4-positive cells per cubic millimeter. *New England Journal of Medicine*, 322(14), 941-949.
- [34]. Furman, P. A., Fyfe, J. A., St Clair, M. H., Weinhold, K., Rideout, J. L., Freeman, G. A., ... & Mitsuya, H. (1986). Phosphorylation of 3'-azido-3'-deoxythymidine and selective interaction of the 5'-triphosphate with human immunodeficiency virus reverse transcriptase. *Proceedings of the National Academy of Sciences*, 83(21), 8333-8337.
- [35]. Raffi, F. (2002). Nouvelles molécules antirétrovirales. *Médecine thérapeutique*, 8(1), 47-51.
- [36]. Richman, D. D., Fischl, M. A., Grieco, M. H., Gottlieb, M. S., Volberding, P. A., Laskin, O. L., ... & Jackson, G. G. (1987). The toxicity of azidothymidine (AZT) in

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

- the treatment of patients with AIDS and AIDS-related complex. *New England Journal of Medicine*, 317(4), 192-197
- [37]. De Clercq, E., Perspectives of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection
- [38]. Pauwels, R., Andries, K., Desmyter, J., Schols, D., Kukla, M.J., Breslin, H.J., Raeymaeckers, A., Van Gelder, J., Woestenborghs, R. et Heykants, J., Potent and selective inhibition of HIV-1 replication in vitro by a novel series of TIBO derivatives (1990), *Nature*, 343:470-474
- [39]. Miyasaka, T., Tanaka, H., Baba, M., Hayakawa, H., Walker, R. T., Balzarini, J., & De Clercq, E. (1989). A novel lead for specific anti-HIV-1 agents: 1-[(2-hydroxyethoxy) methyl]-6-(phenylthio) thymine. *Journal of Medicinal Chemistry*, 32(12), 2507-2509
- [40]. Madruga, J. V., Cahn, P., Grinsztejn, B., Haubrich, R., Lalezari, J., Mills, A., ... & Smedt, G. (2007). Efficacy and safety of TMC125 (etravirine) in treatment-experienced HIV-1-infected patients in DUET-1: 24-week results from a randomised, double-blind, placebo-controlled trial. *The Lancet*, 370(9581), 29-38.
- [41]. Cohen, C. J., Molina, J. M., Cahn, P., Clotet, B., Fourie, J., Grinsztejn, B., ... & Crauwels, H. (2012). Efficacy and safety of rilpivirine (TMC278) versus efavirenz at 48 weeks in treatment-naïve HIV-1-infected patients: pooled results from the phase 3 double-blind randomized ECHO and THRIVE trials. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 60(1), 33-42.
- [42]. Moyle, g., the emerging roles of non-nucleoside reverse transcriptase inhibitors in antiretroviral therapy. *Drugs*, 2001. 61(1): p. 20-26
- [43]. 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.
- [44]. Baba, M., Tanaka, H., De Clercq, E., Pauwels, R., Balzarini, J., Schols, D., ... & Miyasaka, T. (1989). Highly specific inhibition of human immunodeficiency virus type 1 by a novel 6-substituted acyclouridine derivative. *Biochemical and biophysical research communications*, 165(3), 1375-1381.
- [45]. Rosowsky, A., Kim, S. H., & Wick, M. (1981). Synthesis and antitumor activity of an acyclonucleoside derivative of 5-fluorouracil. *Journal of medicinal chemistry*, 24(10), 1177-1181.
- [46]. Tanaka, H., Baba, M., Hayakawa, H., Sakamaki, T., Miyasaka, T., Ubasawa, M., ... & Walker, R. T. (1991). A new class of HIV-1 specific 6-substituted acyclouridine derivatives: synthesis and anti-HIV-1 activity of 5-or 6-substituted analogs of 1-[(2-hydroxyethoxy) methyl]-6-(phenylthio) thymine (HEPT). *Journal of medicinal chemistry*, 34(1), 349-357.
- [47]. Rosowsky, A.; Kim, Sa-H.; Wick, M. *J. Med. Chem.* 1981,24, 1177-1181.
- [48]. Kim, Y. H., Kim, J. Y., & Lee, C. H. (1988). Efficient synthesis of acyclic nucleosides by N-alkylation of pyrimidine and purine-bases using a new coupling agent of cesium iodide. *Chemistry Letters*, 17(6), 1045-1048
- [49]. Nawrot, B., & Malkiewicz, A. (1989). The tRNA "WOBBLE POSITION" Uridines. III. 1 the Synthesis of 5-[S-Methoxycarbonyl (Hydroxy) Methyl] Uridine and its 2-Thio Analogue. *Nucleosides & nucleotides*, 8(8), 1499-1512

## Chapitre.I : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH

- [50]. Tanaka, H., Baba, M., Ubasawa, M., Takashima, H., Sekiya, K., Nitta, I., ... & Miyasaka, T. (1991). Synthesis and anti-HIV activity of 2-, 3-, and 4-substituted analogs of 1-[(2-hydroxyethoxy) methyl]-6-(phenylthio) thymine (HEPT). *Journal of medicinal chemistry*, 34(4), 1394-1399
- [51]. Niu, C.-H. *Anal. Biochem.* 1984, 139, 404-407.
- [52]. Reese, C. B.; Ubasawa, A. *Tetrahedron Lett.* 1980, 21, 2265-2268
- [53]. Divakar, K. J.; Reese, C. B. *J. Chem. SOC.P, erkin Trans. 1* 1982, 1171-1176
- [54]. Tanaka, H., Takashima, H., Ubasawa, M., Sekiya, K., Nitta, I., Baba, M., ... & Miyasaka, T. (1992). Structure-activity relationships of 1-[(2-hydroxyethoxy) methyl]-6-(phenylthio) thymine analogs: effect of substitutions at the C-6 phenyl ring and at the C-5 position on anti-HIV-1 activity. *Journal of medicinal chemistry*, 35(2), 337-345.
- [55]. Tanaka, H., Baba, M., Saito, S., Miyasaka, T., Takashima, H., Sekiya, K., ... & Walker, R. T. (1991). Specific anti-HIV-1 acyclonucleosides which cannot be phosphorylated: synthesis of some deoxy analogs of 1-[(2-hydroxyethoxy) methyl]-6-(phenylthio) thymine. *Journal of medicinal chemistry*, 34(4), 1508-1511.
- [56]. Tanaka, H., Baba, M., Hayakawa, H., Haraguchi, K., Miyasaka, T., Ubasawa, M., ... & De Clercq, E. (1991). Lithiation of Uracilnucleosides and its Application to the Synthesis of a New Class of Anti-HIV-1 Acyclonucleosides. *Nucleosides & Nucleotides*, 10(1-3), 397-400
- [57]. Tanaka, H., Takashima, H., Ubasawa, M., Sekiya, K., Nitta, I., Baba, M., ... & Miyasaka, T. (1992). Synthesis and antiviral activity of deoxy analogs of 1-[(2-hydroxyethoxy) methyl]-6-(phenylthio) thymine (HEPT) as potent and selective anti-HIV-1 agents. *Journal of medicinal chemistry*, 35(25), 4713-4719.
- [58]. Tanaka, H., Takashima, H., Ubasawa, M., Sekiya, K., Inouye, N., Baba, M., ... & Miyasaka, T. (1995). Synthesis and antiviral activity of 6-benzyl analogs of 1-[(2-hydroxyethoxy) methyl]-5-(phenylthio) thymine (HEPT) as potent and selective anti-HIV-1 agents. *Journal of medicinal chemistry*, 38(15), 2860-2865.
- [59]. Fossey, C., Laduree, D. et Robba, M., Synthesis of acyclic thieno[3,2-d]pyrimidine nucleosides and azido derivatives as potential anti-HIV agents (1994), *Nucleosides & Nucleotides*, 13:883-889.
- [60]. Maruenda, H. et Johnson, F., Design and synthesis of novel inhibitors of HIV-1 reverse transcriptase (1995), *J.Med.Chem.*, 38:2145-2151.

***Chapitre II***  
***Méthodologie QSAR***

## Chapitre.II: Méthodologie QSAR

### I. Introduction

La transmission des connaissances médicaux, reposant sur la capacité de partager les savoirs acquis, est depuis toujours un facteur déterminant de la découverte de médicaments. L'exploration de ces derniers est toujours d'actualité et constitue une discipline à part entière, l'ethnopharmacologie. La publication au niveau mondial des résultats scientifiques a largement contribué aux nombreux succès de la recherche de médicaments [1]. Pendant longtemps, des substances naturelles, principalement d'origine végétale mais aussi minérale et animale, étaient utilisées. Ces substances étaient sélectionnées par l'observation empirique de leurs effets sur le cours des maladies. Ainsi, certaines substances naturelles traditionnelles sont toujours utilisées pour leurs vertus thérapeutiques [2]. Cependant, la nécessité de rechercher d'autres sources de médicaments est rapidement apparue et a permis de nombreuses avancées tout au long du XXe siècle et jusqu'à aujourd'hui (Figure.1).

Le début du XIXe siècle marque un tournant historique dans la recherche de nouveaux médicaments grâce à l'isolement de principes actifs de substances naturelles précédemment utilisées dans la médecine traditionnelle. Ainsi, la morphine est isolée du pavot en 1803 et ses effets sont décrits par Friedrich Wilhelm Adam Sattürner quelques temps plus tard [2] (Figure.1).

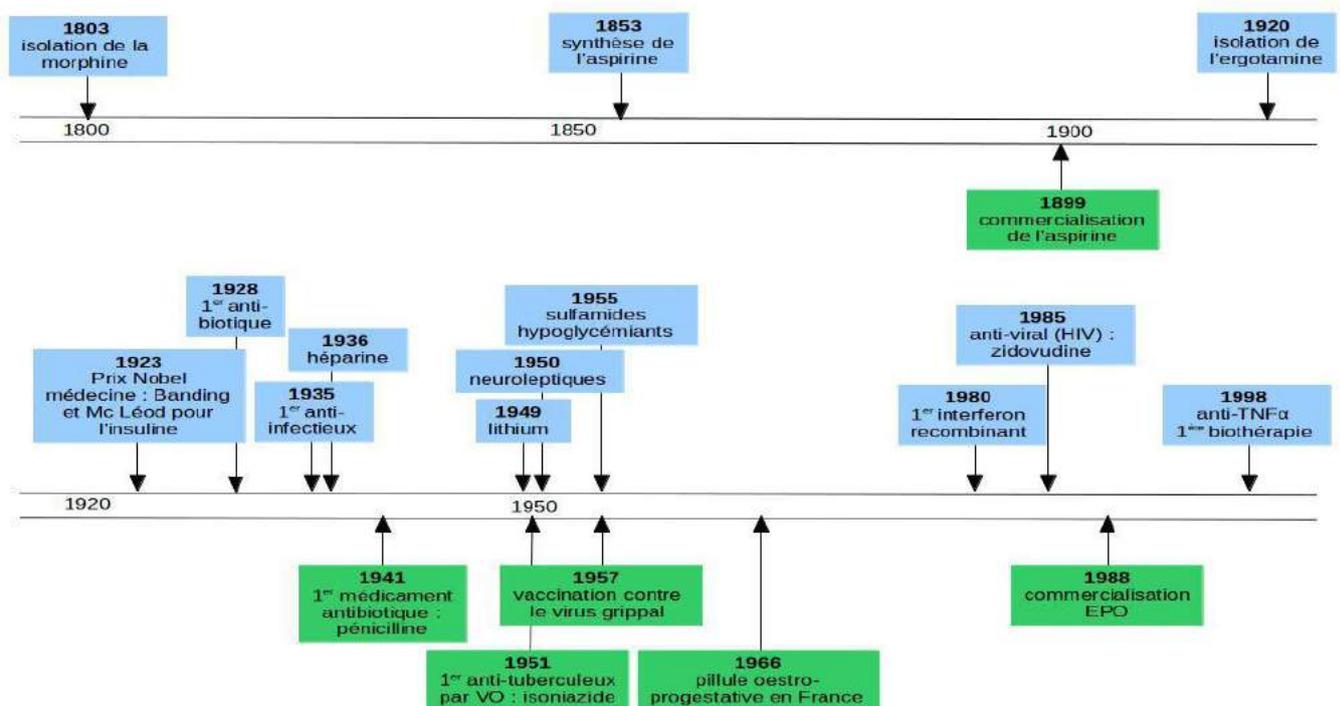


Figure 1. Quelques grandes étapes de la découverte de médicaments au cours du XIXe et XXe Siècle

## Chapitre.II: Méthodologie QSAR

L'essor des médicaments de synthèse débute dans le milieu du XIXe siècle avec la synthèse en 1853 de l'acide acétylsalicylique par Charles Gerhardt [3-4] puis sa commercialisation vingt ans plus tard sous le nom de marque « Aspirine » par les Laboratoires Bayer [5].

L'identification de molécules à intérêt thérapeutique est un long processus, faisant intervenir de très nombreuses connaissances scientifiques dans les domaines de la chimie (bio) organique, de la pharmacognosie, de la pharmacologie, de la pharmacocinétique, de la pharmacie galénique, de la pharmacie clinique. La découverte d'un nouveau médicament nécessite avant tout la mise au point d'une nouvelle substance active. Le travail à entreprendre est difficile, long, fastidieux mais bénéfique.

Le lancement d'un programme de recherche ayant pour objectif la mise au point d'une nouvelle molécule à intérêt thérapeutique dépend d'un grand nombre de paramètres et de connaissances scientifiques. Pour les pharmaco-chimistes, il sera primordial de prendre en compte certains d'entre eux pour élaborer leur projet de recherche. Par exemple, ils seront amenés à :

- ✓ Choisir une maladie (SIDA, diabète de type 2, hépatite virale B,) ;
- ✓ Choisir une cible (Transcriptase inverse, ADN-gyrase bactérienne, protéasome 26S, tyrosine kinase Bcr-Abl) ;
- ✓ L'identification d'un ligand naturel d'un récepteur (Sérotonine, adrénaline, acétylcholine);
- ✓ Bénéficier de la mise au point d'un test biologique (L'inhibition de la caséine kinase 2, cytotoxicité sur lignées tumorales) ;
- ✓ L'identification d'un « hit » suite à un criblage à haut débit d'une librairie de molécules (chimiothèque) ;
- ✓ L'isolement et la purification d'une substance active d'un extrait végétal ou marin.

Cette liste n'est pas limitative et peut intégrer d'autres données importantes comme(i) la connaissance de la structure tridimensionnelle de la cible et (iii) l'identification des relations structure-activité (RSA). De toute manière, le point de départ d'un projet de recherche n'est pas standardisé et dans tous les cas une extrême flexibilité est requise pour analyser l'ensemble des données préliminaires obtenues et ainsi définir la meilleure stratégie pour rechercher la molécule « hit », la molécule chef de file, le candidat médicament.

## Chapitre.II: Méthodologie OSAR

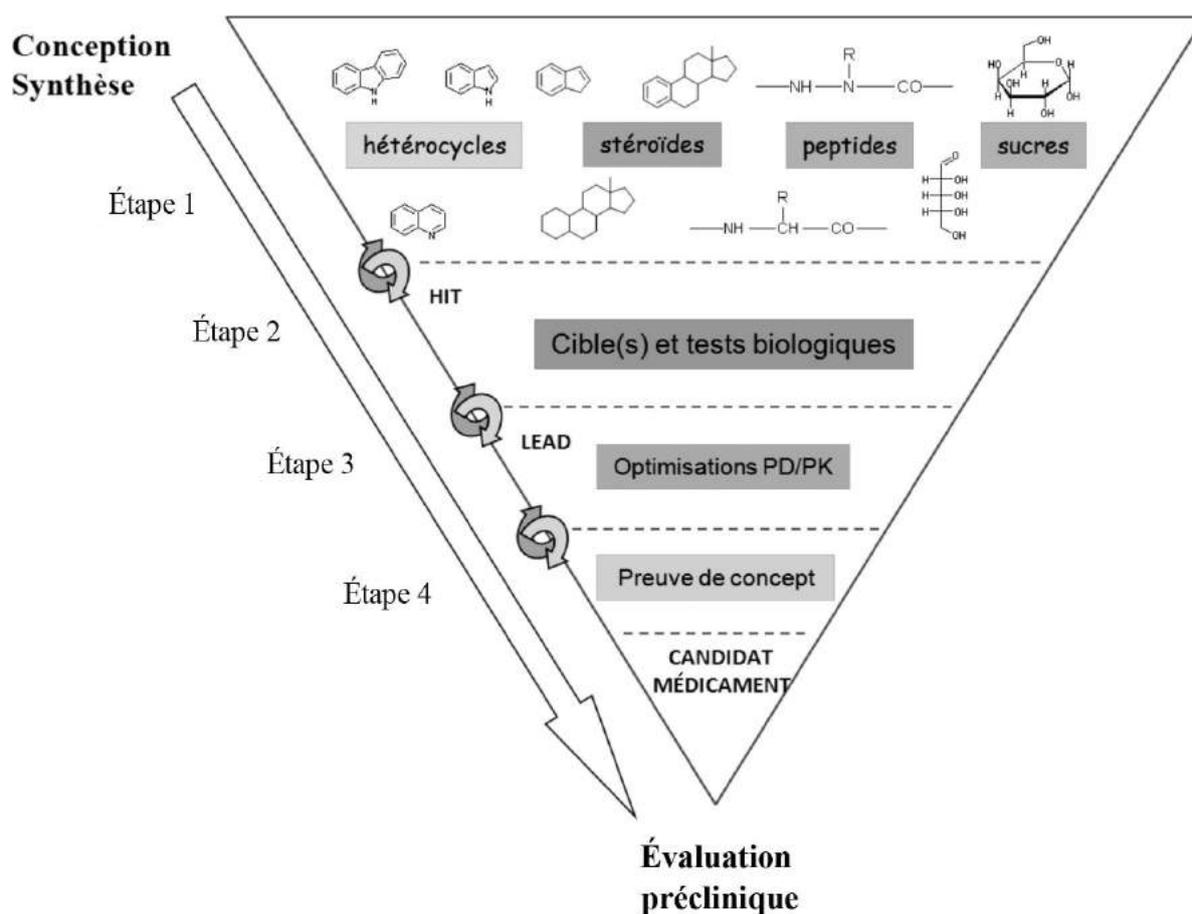


Figure 2 : Représentation simplifiée des principales étapes de la mise au point d'un candidat médicament

Les principales étapes de la mise au point d'un candidat médicament sont listées sur la figure 2. Dans un premier temps (étape 1), il s'agit de la découverte d'un « hit ». Par exemple, il est possible d'identifier un « hit » soit après la mise en place d'une évaluation à haut débit soit à partir d'une molécule isolée d'une plante. La modulation structurale du « hit » peut être réalisée pas à pas (exp : l'introduction de groupements électro-attracteurs puis électro-donneurs) ou en s'appuyant sur des données 3D de la cible (groupements donneurs ou accepteurs de liaison H).

Ensuite, l'étape 2 consiste à l'évaluation de la librairie de composés synthétisés (la chimiothèque) et ainsi établir les premières RSA. La diversité structurale générée par ces recherches de pharmaco-modulation est essentielle pour découvrir une molécule chef de file. L'étape 3 se focalisera sur l'optimisation structurale du « lead », sur le plan pharmacodynamique (PD – optimisation du potentiel d'activité) mais également sur le plan pharmacocinétique (PK – optimisation des paramètres ADMET) [6].

L'étape 4, dite « preuve de concept » est une étape clé qui doit permettre de démontrer l'efficacité du « lead optimisé » sur différents modèles biologiques (*in vitro*, mais également *in vivo*). Les frontières entre ces différentes étapes ne sont pas étanches et de nombreux allers et retours sont nécessaires pour mettre au point le « lead » à partir du « hit », puis de disposer d'un candidat médicament issu de l'optimisation progressive du « lead ».

## Chapitre.II: Méthodologie QSAR

### II. Les relations quantitatives structures activités

Les relations quantitatives structure-activité (QSAR) sont de plus en plus utilisées, du fait de la croissance des moyens informatiques. Le nouveau règlement européen REACH [7-10] recommande leur utilisation pour limiter le recours à l'expérience. Dans les dernières années, l'utilisation des méthodes QSAR n'a cessé de progresser. Elle est même devenue indispensable en chimie pharmaceutique et pour la conception de médicaments [11].

#### II.1. Définition

Les méthodes QSAR sont basées sur l'hypothèse que l'activité d'un composé chimique est liée à sa structure. Plus précisément cette approche affirme que l'activité biologique et la structure d'un composé chimique sont liées par une formule mathématique.

Le principe des méthodes QSAR est, comme leur nom l'indique, de mettre en place une relation mathématique reliant de manière quantitative la structure moléculaire, codée par des propriétés moléculaires appelées descripteurs, avec une activité biologique en utilisant des méthodes d'analyse de données [12].

On développe des modèles prédictifs de la forme suivante :

$$\text{Activité} = f(\text{descripteurs moléculaires}) \quad (1)$$

La relation mathématique obtenue peut alors être utilisée comme moyen prédictif de l'activité biologique étudiée de nouvelles molécules ou des molécules pour lesquelles les données expérimentales ne sont pas encore disponibles. Ils peuvent être également utilisés pour mieux comprendre les mécanismes et les modes d'action.

La méthodologie générale d'une étude QSAR est la suivante :

1. Constituer la base de données structure – activité à partir de mesures quantitatives, fiables pour chaque composé ;
2. Sélectionner des descripteurs moléculaires en relation avec l'activité cible afin de traduire de manière numérique la structure des molécules ;
3. Diviser ce jeu de données aléatoirement en un d'apprentissage et un autre de test ;
4. Établir le modèle mathématique en utilisant la série d'apprentissage à l'aide des méthodes statistiques ;
5. Caractériser ce modèle par leurs indices statistiques et par une validation interne ;
6. Valider les modèles avec la série de test et calculer leur indice de corrélation externe ;
7. Élaborer le domaine d'applicabilité du modèle proposé ;
8. Explorer et exploiter les modèles validés pour comprendre les mécanismes possibles et faire des prévisions d'activité biologique de nouvelles molécules [13].

## Chapitre.II: Méthodologie QSAR

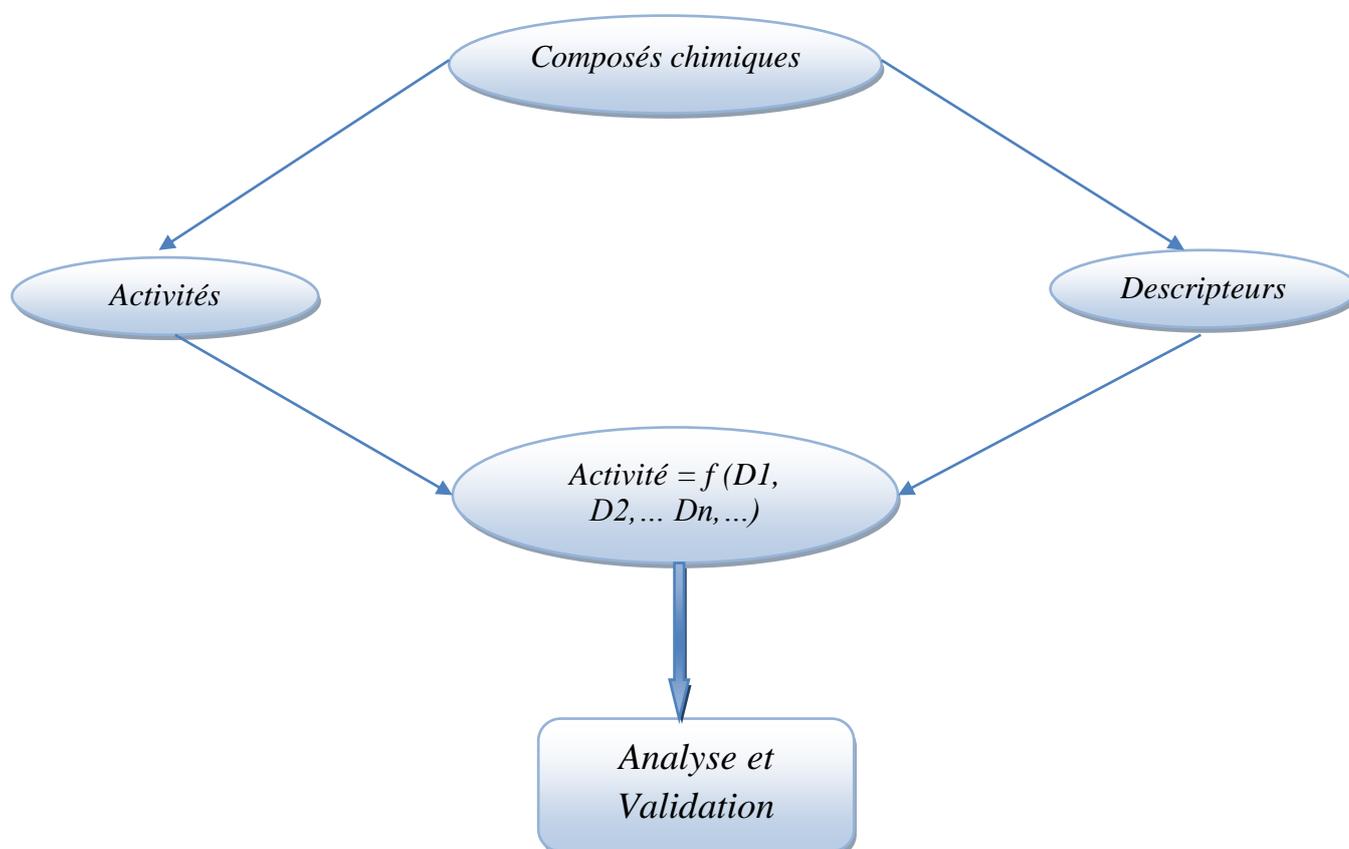


Figure.3 : Présentation de la méthodologie QSAR.

## Chapitre.II: Méthodologie QSAR

Il y a plusieurs raisons pratiques qui justifient l'utilisation des études QSAR

- ✓ Prédire les propriétés de l'activité biologique par des moyens rationnels.
- ✓ Économiser le coût de développement de médicament.
- ✓ Les prévisions pourraient réduire l'exigence de tests longs et coûteux pratiqués sur les animaux.

### **II.2. Les premiers développements de QSAR**

Il y a plus d'un siècle, Crum-Brown et Fraser ont exprimé l'idée que l'action physiologique d'une substance était fonction de sa composition chimique et de sa constitution [14]. Quelques décennies plus tard, en 1893, Richet a montré que les cytotoxicités d'un ensemble divers de molécules organiques simples étaient inversement liées à leurs solubilités correspondantes dans l'eau [15]. Au début du 20<sup>ème</sup> siècle, Meyer et Overton suggèrent indépendamment que l'action narcotique (dépressive) d'un groupe de composés organiques est parallèle à leurs coefficients de partition huile d'olive / eau [16-17].

En 1939, Ferguson a introduit une généralisation thermodynamique à la corrélation de l'action dépressive avec la saturation relative des composés volatils dans le véhicule dans lequel ils ont été administrés [18]. Les travaux approfondis d'Albert, Bell et Roblin ont établi l'importance de l'ionisation des bases et des acides faibles dans l'activité bactériostatique [19-21]. Pendant ce temps, sur le front organique physique, de grands progrès ont été réalisés dans la délimitation des effets des substituants sur les réactions organiques, menées par le travail séminal de Hammett, qui a donné naissance à la culture « sigma rho » [22-23]. Taft a conçu un moyen de séparer les effets polaires, stériques et de résonance et d'introduire le premier paramètre stérique, ES [24].

Les contributions de Hammett et Taft ont jeté les bases mécanistiques du développement du paradigme QSAR par Hansch et Fujita. En 1962, Hansch et Muir ont publié leur brillante étude sur les relations structure-activité des régulateurs de croissance des plantes et leur dépendance aux constantes de Hammett et à l'hydrophobie [25]. En utilisant le système octanol / eau, toute une série de coefficients de partition ont été mesurés, et ainsi une nouvelle échelle hydrophobe a été introduite [26]. Le paramètre  $\pi$ , qui est l'hydrophobicité relative d'un substituant, a été défini d'une manière analogue à la définition de sigma [27].

## **Chapitre.II: Méthodologie QSAR**

### **III. Méthodologie générale d'une étude QSAR**

#### **III.1. La base de données**

Le choix de la base de données expérimentale initiale est une étape critique pour le développement des modèles QSAR. Généralement, les composés testés ont deux origines possibles, soit des produits synthétisés ou des produits d'extraction à partir de plantes. Les conditions expérimentales ont une forte influence sur les valeurs obtenues.

La base de données doit être composée de données expérimentales fiables, puisque les marges d'erreurs sur celles-ci se propageront dans le modèle final. Il est donc important de choisir des données présentant de faibles incertitudes afin de limiter les marges d'erreur expérimentales.

Les données biologiques sont exprimées habituellement sur une échelle logarithmique en raison de la relation linéaire entre la réponse et le logarithme de dose dans la région centrale de la courbe de log dose-réponse. Les logarithmes inverses de l'activité ( $\log 1/C$ ) sont également utilisés pour obtenir des valeurs mathématiques plus élevées lorsque les structures sont biologiquement très efficaces [28].

L'obtention d'un modèle efficace sera également dépendante du type de molécules qui y sont incluses. La série devrait contenir des composés pour lesquels le mécanisme mis en jeu est unique. Finalement, la présence de molécules au comportement déviant au sein de la base de données peut influencer de manière importante le modèle final. De même, la mise en évidence de sous-groupes de molécules dans un modèle a tendance à le dégrader.

Avant toute modélisation, il est nécessaire d'exploiter au maximum les informations contenues dans les structures moléculaires, celles-ci sont traduites en une série de grandeurs (en général scalaires) qui quantifient leurs caractéristiques physico-chimiques et structurelles, ces grandeurs sont appelées descripteurs [29].

#### **III.2. Les descripteurs moléculaires**

##### **III.2.1. Définition**

Un descripteur moléculaire est un paramètre ayant une valeur numérique propre pour une structure chimique donnée. Ils peuvent être obtenus de manière empirique ou non-empirique. 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 moléculaire.

Depuis des années, plusieurs travaux ont été menés afin de développer des descripteurs capables de décrire de la manière la plus exhaustive possible les structures moléculaires.

## **Chapitre.II: Méthodologie QSAR**

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 [30].

### **III.2.2. Types de descripteurs**

Historiquement, deux grands schémas pour la classification des descripteurs moléculaires ont été établis : l'un en fonction de leurs origines (constitutionnels, topologiques, géométriques, quantiques, thermodynamiques...), et un autre sur leurs dimensionnalité (1D, 2D ou 3D) [31].

#### **III.2.2.1 Les descripteurs 1D**

Les descripteurs 1D sont appelés «descripteurs constitutionnels» et sont faciles et rapides à calculer, ils sont calculés à partir de la formule brute de la molécule ou à partir de la composition moléculaire, c'est-à-dire les atomes qui la constituent en décrivant des propriétés globales du composé [32-33].

Des exemples de descripteurs constitutionnels sont :

- le nombre d'atome ;
- le nombre de liaison ;
- les nombres de cycles.

Ce type de descripteurs est employé pour obtenir les modèles QSAR les plus simples.

#### **III.2.2.2 Les descripteurs 2D**

Les descripteurs moléculaires qui utilisent la représentation des molécules comme des graphes sont dits «descripteurs 2D». Ils contiennent des informations à propos de la connectivité ou à propos de certains fragments moléculaires, mais aussi des estimations des propriétés physico-chimiques [34-35]

##### ***Descripteurs topologiques***

La structure du composé est représentée par un graphe : les atomes étant les sommets et les liaisons sont les arêtes. De nombreux indices quantifiant la connectivité moléculaire ont été développés en se basant sur cette approche. Les plus fréquemment utilisés sont l'indice de Wiener [36], l'indice de Randić [37], l'indice de connectivité de valence de Kier-Hall et l'indice de Balaban [38]. L'indice de Wiener permet de caractériser le volume moléculaire et la ramification d'une 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.

Les indices topologiques sont souvent considérés comme des descripteurs convenables d'un point de vue numériques. Cela dit, dans la plupart des cas, l'interprétation des équations QSAR qui en résultent n'est pas aisée, puisqu'il est difficile de les relier aux mécanismes sous-jacents.

## **Chapitre.II: Méthodologie QSAR**

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.

### **III.2.2.3 Les descripteurs 3D**

Ils décrivent des caractéristiques complexes ; leurs calculs nécessitent donc de connaître la géométrie 3D de la molécule. Ces descripteurs s'avèrent donc relativement coûteux en temps de calcul, mais apportent davantage d'informations, et sont souvent nécessaires à la modélisation d'activités biologiques [39].

On distingue plusieurs familles importantes de descripteurs :

#### ***Descripteurs géométriques***

Ils sont évalués à partir des positions relatives des atomes d'une molécule dans l'espace, ainsi que des rayons et masses atomiques. Celles-ci peuvent être obtenues expérimentalement bien entendu mais le plus souvent par modélisation moléculaire, empirique ou *ab initio*. Ils sont basés sur l'arrangement spatial des atomes constituant la molécule et sont définis par les coordonnées des noyaux atomiques de la molécule représentée. Ces descripteurs incluent des informations sur la surface moléculaire obtenue par les aires de Van Der Waals et leur superposition. Les volumes moléculaires peuvent être obtenus par les volumes de Van Der Waals [40].

Parmi ces descripteurs, on retrouve le volume, la surface moléculaire, le moment d'inertie ou encore des distances, angles ou angles dièdres particuliers entre atomes dans la molécule, le nombre de liaisons, la surface de Van Der Waals et le volume de Van Der Waals,...ect.

#### ***Descripteurs physico-chimiques***

Les descripteurs physico-chimiques, certains d'entre eux reflètent la composition moléculaire du composé, d'autres représentent le caractère hydrophile ou lipophile de la molécule généralement évalué à partir du coefficient de partage Octanol/eau représenté par le log P. Parmi ces descripteurs, le log P, la réfractivité moléculaire, l'indice de réfraction, la polarisabilité, la densité, le parachor, la surface de tension, le nombre de donneurs de liaisons hydrogène, le nombre d'accepteurs de liaisons hydrogène,...ect [41].

#### ***Descripteurs spectroscopiques***

Les molécules peuvent être caractérisées par des mesures spectroscopiques, par exemples par leurs fonctions d'onde vibrationnelles. Les vibrations d'une molécule dépendent des forces d'interaction et de la masse des atomes. Ces vibrations donnent donc des informations sur la structure et sur la conformation de la molécule. Après la recherche de la géométrie optimale on peut obtenir les spectres infrarouges. Ces spectres sont alors codés en vecteurs de descripteurs [42].

Ces descripteurs sont souvent spécifiques à certaines méthodes (CoMFA, CoMSIA...) [43-44].

## **Chapitre.II: Méthodologie QSAR**

### ***Descripteurs thermodynamiques***

Ce sont des descripteurs peu utilisés dans les études QSAR. Ils peuvent être exprimés par la fonction de partition Q de la molécule utilisée en thermodynamique statistique ainsi que de ses dérivées. Cette fonction décrit la façon avec laquelle l'énergie d'un système de molécules est répartie parmi les individus moléculaires. Sa valeur dépend du poids moléculaire, de la température, du volume moléculaire, des distances inter nucléaires, des mouvements moléculaires et des forces intermoléculaires [45].

Par exemples : Le point d'ébullition, la constante de Henry, la température, la pression, l'enthalpie et le point de fusion.

### ***Descripteurs quantiques/électroniques***

Afin d'aller plus loin dans la description des structures moléculaires, des caractéristiques supplémentaires de la structure moléculaire peuvent encore être calculées. Ces *descripteurs électroniques* [46-47] permettent de quantifier différents types d'interactions inter et intramoléculaires de grande influence sur l'activité biologique. Ces descripteurs nécessitent des calculs plus sophistiqués pour la recherche de la géométrie pour laquelle l'énergie est minimale, et fait souvent appel à la chimie quantique qui nous donne accès à des informations supplémentaires telles que des données énergétiques, vibrationnelles et orbitales du système [48-49].

Il est important de noter qu'il n'existe pas de « bon » ou de « mauvais » descripteur : l'utilité et l'efficacité sont étroitement liées aux types de molécules à traiter ainsi qu'aux calculs à effectuer. Par conséquent, la plupart des descripteurs connus aujourd'hui sont employés de préférence dans le contexte pour lesquels ils ont été créés [50-51].

De nombreux logiciels ont été développés pour calculer les différents descripteurs moléculaires tels que : Gaussian[52], ChemOffice[53], ChemSketch[54], Marvin Sketch [55], QSARIS [56], Cerius2 [57], Vol Surf [58], Dragon [59], Codessa [60], ISIDA [62],MOPAC[62].

## **Chapitre.II: Méthodologie QSAR**

### **III.3 Méthodes d'évaluation de descripteurs quantiques**

Les méthodes de modélisation basées sur la mécanique quantique présentent des avantages évidents car elles dépendent beaucoup moins des paramètres empiriques du système étudié. En théorie quantique un système moléculaire est complètement décrit par la donnée de sa fonction d'onde. De ce fait toute propriété du système moléculaire peut être évaluée à partir de la fonction d'onde. Ainsi l'évaluation des descripteurs quantiques d'une molécule nécessite la connaissance de sa fonction d'onde ou au moins sa densité électronique. D'une manière générale, pour les systèmes moléculaires conservateurs de l'énergie, il suffit la détermination de sa fonction d'onde spatiale. Théoriquement cette fonction peut être obtenue par la résolution de l'équation de Schrödinger indépendante du temps [63].

En pratique, on ne peut obtenir que des fonctions approchées de systèmes modèles en utilisant les méthodes d'approximations, en l'occurrence la méthode des variations et la méthode de perturbation [64]. Le modèle le plus simple d'une molécule est obtenu lorsqu'on considère les noyaux immobiles et les mouvements des électrons indépendants. Ces simplifications sont justifiées par l'approximation de Born-Oppenheimer [65], qui s'appuie sur la différence importante entre la masse de l'électron et celles des noyaux pour permettre la séparation des mouvements nucléaires et électroniques, et l'approximation orbitale qui néglige la corrélation entre les mouvements de deux électrons. Ainsi et pour respecter le principe de Pauli [66], on recherche sous la forme d'un déterminant de Slater d'orbitales moléculaires, la fonction d'onde spatiale de l'état fondamental des électrons de la molécule évoluant dans le champ des noyaux immobiles. Cette approche est dite le cadre Hartree-Fock [67-68]. La mise en œuvre de cet approche en application informatique nécessite l'introduction de la troisième approximation appelée la Combinaison Linéaire des Orbitales Atomiques (CLOA) [69] : une orbitale moléculaire est développée sur des orbitales centrées sur tous les noyaux de la molécule. Comme la stipule le théorème des variations, on est certain que l'énergie obtenue à partir de la fonction d'onde ainsi déterminée est supérieure à l'énergie exacte de l'état fondamental. Cependant on ne peut pas se prononcer sur la position des valeurs de propriétés physiques calculées à partir de cette fonction d'onde par rapport à leurs valeurs exactes. Il existe plusieurs types de méthodes dites post-Hartree-Fock [70] qui tentent d'améliorer la qualité de la fonction d'onde en intégrant la corrélation électronique de diverses manières et en appliquant les principes des méthodes des variations et des perturbations.

Les méthodes Post-Hartree-Fock améliorent la qualité de la fonction d'onde mais exigent des ressources informatiques plus importantes. La recherche directe de la densité électronique sans passer par la fonction d'onde permet l'obtention de meilleurs résultats que le cadre Hartree-Fock en utilisant de modestes ressources informatiques. L'idée de base de cette approche est connue depuis les premières décennies du vingtième siècle mais elle n'a été acceptée que vers les années soixante du même siècle après la publication des théorèmes de Hohenberget Kohn [71] qui affirme que la densité électronique et l'Hamiltonien électronique peuvent avoir une relation fonctionnelle. Depuis, cette approche est connue sous le nom de la Théorie de la Fonctionnelle de la Densité (DFT) [72]. La meilleure description de la densité électronique correspondant au potentiel créé par les noyaux immobiles est obtenue par la méthode des variations.

## Chapitre.II: Méthodologie QSAR

La qualité de cette densité électronique dépend des bases d'orbitales atomiques utilisées pour exprimer les orbitales moléculaires et la fonctionnelle utilisée pour décrire le potentiel d'échange et de corrélation électronique.

Dans ce qui suit je ne présente pas les méthodes d'obtention de la fonction d'onde et la densité électronique de l'état fondamental d'une molécule mais je présente les méthodes de calcul de descripteurs quantiques en particulier des indices globaux et locaux dérivant de la DFT conceptuelle.

### III.3.1 Concepts dérivant de la DFT conceptuelle

Les outils théoriques de la fonction de densité peuvent également être utilisés pour une description qualitative de la réactivité chimique, l'accent étant mis sur la description des forces motrices des réactions chimiques, formant ainsi une base solide pour plusieurs descripteurs de réactivité. Dans la DFT, des définitions précises sont fournies pour certains concepts et principes chimiques clés. Le calcul de ces quantités pour diverses espèces chimiques s'est avéré d'une grande utilité pour la rationalisation de données expérimentales et théoriques dans différents domaines de la chimie [73]. L'énergie électronique est exprimée en fonction du nombre d'électrons  $N$  et du potentiel externe  $v(r)$ . En supposant que les réactions chimiques à l'étude sont motivées par les changements de l'énergie électronique, il faut considérer le changement de l'énergie électronique d'une molécule en changeant le nombre d'électrons  $N$  ou le potentiel externe. L'étude du transfert d'électrons sera traduite dans l'étude de la façon dont l'énergie de la molécule change en tant que nombre de changements d'électrons. La présence d'un agent attaquant sera étudiée via le changement d'énergie d'une molécule par rapport à un changement de potentiel externe. Afin d'étudier ces changements [74], l'équation différentielle totale est la suivante :

$$dE[v; N] = \left( \frac{\partial E[v; N]}{\partial N} \right)_{v(r)} dN + \int \left( \frac{\delta E[v; N]}{\delta v(r)} \right)_N \delta v(r) dr \quad (2)$$

Le premier terme de l'équation (2) est le potentiel chimique électronique  $\mu$ . Cela représente à quel point il est énergétiquement favorable pour un système électronique d'accepter des électrons.

## Chapitre.II: Méthodologie QSAR

La question est maintenant de trouver des expressions pour  $\mu$  pour obtenir des valeurs numériques et la méthode de la différence finie donne une définition opérationnelle et approximative de  $\mu$  [75].

$$\mu \approx \frac{E[v; N + 1] - E[v; N] + E[v; N] - E[v; N - 1]}{2} \quad (3)$$

### *Le potentiel d'ionisation PI*

L'énergie nécessaire pour arracher un électron d'un système. C'est à dire l'énergie nécessaire pour passer de la molécule neutre (N électrons) au cation (N-1 électrons) :  $(M + I \rightarrow M^+)$

$$PI = E(N - 1) - E(N) \quad (4)$$

### *L'affinité électronique AE*

L'énergie gagnée par un système lorsqu'il capte un électron. C'est à dire le gain d'énergie qu'accompagne le passage d'un système neutre à un anion (N+1 électrons) :  $(M \rightarrow M^- + A)$

$$AE = E(N) - E(N + 1) \quad (5)$$

Et donc on peut écrire le potentiel chimique suivant la relation :

$$\mu \approx \frac{-AE - PI}{2} \quad (6)$$

D'autre part la quantité  $\chi$  a été définie comme étant l'électronégativité [76].

$$\mu \approx -\chi \quad (6)$$

Dans la validité du théorème de Koopmans [77], le potentiel d'ionisation et l'affinité électronique sont donnés par les relations suivantes :

$$PI = -\varepsilon_{HOMO} \quad (7)$$

$$AE = -\varepsilon_{LUMO} \quad (8)$$

## Chapitre.II: Méthodologie QSAR

La première dérivée partielle de  $\mu$  par rapport à  $N$  (le nombre total d'électrons) est définie comme la dureté globale  $\eta$  du système [78].

$$\eta = \left( \frac{\partial^2 E[v; N]}{\partial N^2} \right)_{v(r)} = \left( \frac{\partial \mu}{\partial N} \right)_{v(r)} \quad (9)$$

$$\eta = (E[v; N + 1] - E[v; N]) - (E[v; N] - E[v; N - 1]) \quad (10)$$

$$\eta = \frac{PI - AE}{2} \quad (11)$$

L'inverse de la dureté, la mollesse ( $S$ ), est défini par l'équation suivante [79].

$$S = \frac{1}{\eta} = \left( \frac{\partial N}{\partial \mu} \right)_{v(r)} \quad (12)$$

Le concept de mollesse est associé à la polarisabilité. Plus le système chimique est gros, plus il est mou. Un descripteur global final dérivé de la dureté moléculaire est défini par Parr [80] sous le nom d'indice d'électrophilicité, mesurant la propension à l'attaque électrophile

$$\omega = \frac{\mu^2}{2\eta} \quad (13)$$

Les indices globaux cités plus haut donnent des caractérisations du comportement du système moléculaire vis-à-vis de son environnement. Ils ne permettent pas de différencier entre les comportements de deux sites différents du système vis-à-vis de son environnement. Des indices évalués localement ont été alors proposés. Il s'agit d'indice caractérisant localement le comportement de la densité électronique du système en réponse à des actions d'agent externes. L'utilisation première de ces indices est l'étude de la réactivité.

Pour déterminer les sites réactifs d'une molécule lors de l'approche d'un agent électrophile, nucléophile ou radicalaire, les chimistes utilisent les charges nettes pour favoriser une interaction par rapport à une autre. Cependant, il est bien connu que le calcul des charges nettes sur les différents sites d'une molécule ne permet pas de bien décrire les interactions entre molécules.

A partir des densités électroniques du système à  $N$ ,  $N + 1$  et  $N - 1$  électrons, il est possible d'avoir accès aux fonctions de Fukui.

## Chapitre.II: Méthodologie QSAR

### *Les indices de Fukui*

La fonction de Fukui mesure la réponse d'un réactif vis-à-vis des modifications de  $N$  et de  $v(r)$ . C'est un descripteur local en 3 dimensions qui permet de comprendre et de prédire les réactivités relatives de différents sites dans une molécule. Les fonctions de Fukui,  $f^+(r)$  et  $f^-(r)$ , mesurent le changement d'énergie lorsque le nombre de particules varie de  $N \rightarrow N \pm dN$ . Dans la notation des dérivées fonctionnelles, les fonctions de Fukui sont dérivées en tant que dérivées fonctionnelles du potentiel chimique par rapport au potentiel externe, et peuvent être écrites par une relation de Maxwell comme dérivées de la densité électronique  $\rho(r)$  en ce qui concerne  $N$ .

$$f^+(r) = \left( \frac{\partial \left( \frac{\delta E}{\delta v(r)} \right)_N}{\partial N} \right)_{v(r)}^+ = \left( \frac{\delta \left( \frac{\partial E}{\partial N} \right)_{v(r)}^+}{\delta v(r)} \right)_N \quad (14)$$

$$f^-(r) = \left( \frac{\partial \left( \frac{\delta E}{\delta v(r)} \right)_N}{\partial N} \right)_{v(r)}^- = \left( \frac{\delta \left( \frac{\partial E}{\partial N} \right)_{v(r)}^-}{\delta v(r)} \right)_N \quad (15)$$

Les indices supérieurs indiquent si la dérivée par rapport au nombre de particules  $N$  augmente (+) ou diminue (-). Les équations + et - montrent que les fonctions de Fukui sont utiles pour décrire des réactions dans lesquelles le nombre d'électrons d'un substrat et le potentiel externe du réactif changent. La fonction de Fukui est normalisée à un.

$$\int f(r) dr = 1 \quad (16)$$

Une valeur élevée de  $f(r)$  sur un site donné indique une réactivité élevée. L'insertion de l'équation  $\rho(r) = \left( \frac{\partial E}{\partial v(r)} \right)_N$  dans les équations (14) et (15) donne les définitions suivantes pour les fonctions de Fukui (équation 17) :

$$f^\pm(r) = \left( \frac{\partial \rho(r)}{\partial N} \right)_{v(r)}^\pm \quad (17)$$

## Chapitre.II: Méthodologie OSAR

L'équation (17) indique que la fonction de Fukui vue d'en haut,  $f^+(r)$ , représente la façon dont la densité d'électrons  $\rho(r)$ , change lorsque le nombre d'électrons augmente de  $N$  à  $N + dN$  à potentiel externe constant  $v(r)$ .

De même, la fonction de Fukui d'en bas,  $f^-(r)$ , représente l'évolution de la densité électronique lorsque le nombre d'électrons décroît de  $N$  à  $N - dN$  à potentiel externe constant. Parce qu'à zéro température et  $N$  entier, les formules interpolant entre l'électron  $N$  et les densités d'électrons  $N + 1$  et  $N - 1$  sont linéaires [81-83], les équations (20) et (21) peuvent être déduites à l'aide des équations (18) et (19) [84-85].

$$\rho_{N+dN,v(r)}(r) = (1 - dN)\rho_{N,v(r)}(r) + dN\rho_{N+1,v(r)}(r) \quad (18)$$

$$\rho_{N-dN,v(r)}(r) = (1 - dN)\rho_{N,v(r)}(r) + dN\rho_{N-1,v(r)}(r) \quad (19)$$

$$f^+(r) = \rho_{N+1,v(r)}(r) - \rho_{N,v(r)}(r) \quad (20)$$

$$f^-(r) = \rho_{N,v(r)}(r) - \rho_{N-1,v(r)}(r) \quad (21)$$

Si la relaxation orbitale est négligée, alors  $f^\pm(r)$  peuvent être approchée en tant que densité de l'orbitale moléculaire frontière, reliant ainsi la théorie orbitale frontière de Fukui à la DFT. Au point  $r$ ,  $f^+(r)$  et  $f^-(r)$  sont des mesures directes de la réactivité face aux attaques nucléophiles et électrophiles. C'est-à-dire que les régions où  $f^+(r)$  est grand stabilisent de manière satisfaisante la densité électronique supplémentaire et sont donc particulièrement réactives vis-à-vis des réactifs riches en électrons. Les régions où  $f^-$  est grand abandonnent facilement leurs électrons et sont donc réactives vis-à-vis de réactifs pauvres en électrons. Dans la mesure où on ne sait jamais vraiment si le réactif du problème à l'étude agira comme un acide ou une base, il semble approprié d'étudier le comportement de la fonction dite de double de Fukui. Ce descripteur peut capturer les deux effets. En 2005, Grand, Morell et Toro-Labbe [86] ont étudié le descripteur dual  $f^{(2)}(r)$ :

$$f^{(2)}(r) = \left( \frac{\partial f(r)}{\partial N} \right)_{v(r)} = \left( \frac{\partial^2 \rho(r)}{\partial N^2} \right)_{v(r)} \quad (22)$$

Dans l'approche des différences fini, on peut écrire par analogie avec les équations (20) et (21), le descripteur dual comme suit :

$$f^{(2)}(r) = \rho_{N+1}(r) - 2\rho_N(r) + \rho_{N-1}(r) \quad (23)$$

## Chapitre.II: Méthodologie QSAR

La fonction de Fukui condensée en atomes peut être définie comme la fonction de Fukui intégrée dans la région attribuée à un atome, comme proposé dans les équations (24), (25) et (26) [87-88].

$$f_k^+ = [q_k(N+1) - q_k(N)] \quad \text{Pour une attaque nucléophile} \quad (24)$$

$$f_k^- = [q_k(N) - q_k(N-1)] \quad \text{Pour une attaque électrophile} \quad (25)$$

$$f_k^\circ = [q_k(N+1) - q_k(N-1)]/2 \quad \text{Pour une attaque radicalaire} \quad (26)$$

$q_k(N)$  : population électronique de l'atome k dans la molécule neutre.

$q_k(N+1)$  : population électronique de l'atome k dans la molécule anionique.

$q_k(N-1)$  : population électronique de l'atome k dans la molécule cationique.

## Chapitre.II: Méthodologie QSAR

### III.4. Sélection des descripteurs moléculaires

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.

Lorsqu'une grande quantité de descripteurs est introduite, certains d'entre eux peuvent contenir des informations redondantes, entraînant un problème de colinéarité. De plus, les descripteurs calculés n'ont pas nécessairement une influence sur l'activité à modéliser.

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.

Différentes techniques permettent de réduire le nombre de descripteurs finalement inclus dans les modèles. L'une d'entre elles est la technique *Best Multi-Linear Régression*(BMLR). La première étape de cette analyse consiste à exclure tous les descripteurs présentant une variance non significative. Cela permet de s'assurer que de tels descripteurs ne sont pas inclus par chance dans le modèle final alors qu'ils n'ont aucun lien avec l'activité étudiée.

Cette étape préliminaire évite l'introduction dans le modèle de descripteurs inappropriés et rend la suite de l'analyse moins coûteuse en termes de temps de calcul, puisqu'elle réduit le nombre de variables restant à traiter.

### IV. Méthodes d'analyse de données

L'objectif principal des statistiques est de maîtriser au mieux l'incertitude pour extraire des informations utiles des données, par l'intermédiaire de l'analyse des variations dans les observations. L'analyse des données est utilisée pour décrire, comprendre et gérer les phénomènes étudiés. Les données peuvent être de toute nature, ce qui rend la statistique utile dans la plupart des disciplines : économie, sociologie, psychologie, agronomie, biologie, médecine, chimie, physique, géologie, sciences de l'ingénieur, sciences de l'information et de la communication, etc...[89].

La mise en place des modèles QSAR n'est pas une tâche aisée. La première difficulté réside dans la différence d'échelles existant entre les données à corrélérer, la structure étant à une échelle moléculaire alors que les propriétés à prédire sont à une échelle macroscopique. De

## **Chapitre.II: Méthodologie QSAR**

plus, on doit tenir compte des problèmes d'incertitudes à la fois au niveau des structures moléculaires (liées niveau de calcul) et des données expérimentales (protocoles de mesures).

Le traitement d'une grande quantité de données représente aussi une difficulté majeure dans le processus de mise au point de modèles QSAR. Les corrélations entre un grand nombre de descripteurs d'un grand nombre de molécules avec leur activité peuvent être à analyser. Cependant aucune règle stricte n'existe quant au choix des paramètres structuraux les plus importants parmi le jeu complet de ceux disponibles [90-91].

En fait, de nombreux outils existent pour trouver le moyen le plus adapté pour obtenir un modèle fiable à partir des données disponibles. Selon les cas, plusieurs approches sont envisageables, il faut alors choisir celle permettant de caractériser au mieux le système.

Les méthodes de modélisation moléculaire sont des outils informatiques qui permettent d'explorer les relations entre les descripteurs de molécules calculés à partir des structures chimiques, et des propriétés ou des activités déterminées expérimentalement.

### **IV.1 Méthode basée sur les descripteurs**

Les méthodes basées sur les descripteurs utilisent une représentation numérique de la structure chimique pour en déduire un modèle. Une autre application importante de cette méthodologie est de pouvoir calculer la capacité d'une molécule d'être un candidat médicament ou une tête de série afin de réduire les risques d'échecs aux étapes expérimentales [92].

#### **IV.1.1 Approches linéaires**

En général, les fonctions linéaires sont facilement interprétables et suffisamment précises pour de petites séries de composés identiques, spécialement lorsque les descripteurs sont sélectionnés avec soin pour une grandeur donnée.

Ces méthodes sont par exemple la régression linéaire multiple (MLR : Multiple Linear Régression), La méthode des moindres carrés (PLS : Partial Least Squares) ou encore l'analyse discriminante linéaire (LDA : Linear Discriminant Analysis).

Cependant, dans plusieurs cas, elles ne peuvent pas être utilisées pour corrélérer avec succès les molécules qui ont une grande diversité, les données avec du bruit et les dépendances non linéaires. Les approches non linéaires peuvent permettre de surmonter ces limitations.

##### **IV.1.1.1 Analyse par composantes principales**

L'analyse en composantes principales (ACP), ou *principal component analysis (PCA)* en anglais, qui est sans doute la technique exploratoire la plus répandue pour décrire les données d'entrée, rechercher d'éventuels aberrations et à connaître les corrélations entre les variables d'entrée. Elle permet d'analyser et de visualiser un jeu de données concernant des individus décrits par plusieurs variables quantitatives. Elle est non supervisée, c'est-à-dire sans phase d'apprentissage [93-95].

C'est une méthode statistique qui permet d'explorer des données dites multivariées (données avec plusieurs variables). Chaque variable pourrait être considérée comme une dimension différente.

## Chapitre.II: Méthodologie OSAR

Si vous avez plus de 3 variables dans votre jeu de données, il pourrait être très difficile de visualiser les données dans un “hyper-espace” multidimensionnel.

L’analyse en composantes principales est utilisée pour extraire et de visualiser les informations importantes contenues dans une table de données multivariées. L’ACP synthétise cette information en seulement quelques nouvelles variables appelées *composantes principales*. Ces nouvelles variables correspondent à une combinaison linéaire des variables originales. Le nombre de composantes principales est inférieur ou égal au nombre de variables d’origine. L’information contenue dans un jeu de données correspond à la variance ou l’*inertie totale* qu’il contient. L’objectif de l’ACP est d’identifier les directions (*axes principaux* ou composantes principales) le long desquelles la variation des données est maximale.

En d’autres termes, l’ACP réduit les dimensions de données multivariées à deux ou trois composantes principales, qui peuvent être visualisées graphiquement, en perdant le moins possible d’information. La compréhension des détails de l’ACP nécessite des connaissances de l’algèbre linéaire.

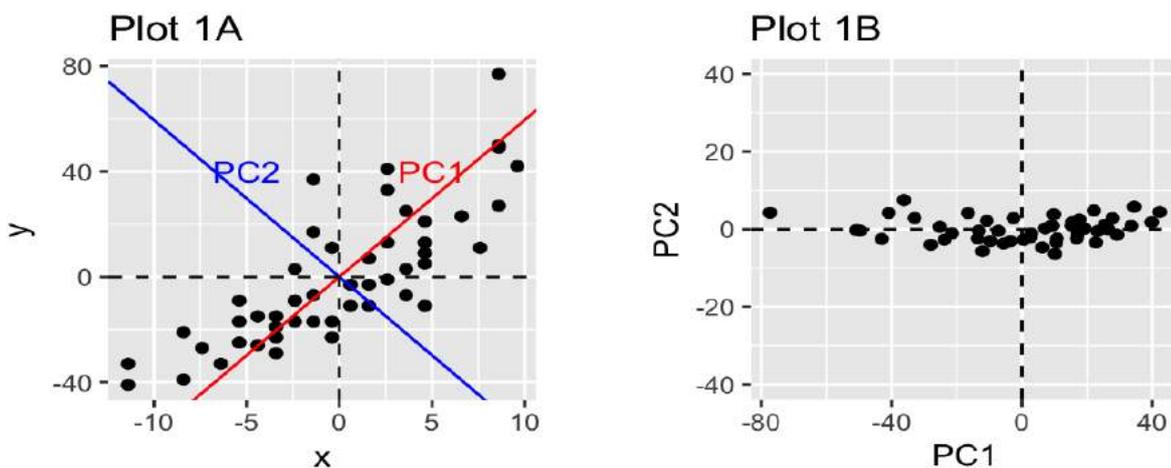


Figure.4 : Une représentation graphique simple des bases de l’ACP

Dans le Plot 1A ci-dessus, les données sont représentées dans le système de coordonnées X-Y. La réduction de la dimension est obtenue en identifiant les directions principales, appelées composantes principales, suivant les quelles les données varient. L’ACP suppose que les directions avec les plus grandes variances sont les plus “importantes” (principales).

Dans la figure ci-dessus, l’axe PC1 est le premier axe principal le long duquel les échantillons présentent la plus grande variation. L’axe PC2 est la seconde direction la plus importante et orthogonal à l’axe PC1.

Les dimensions de notre jeu de données peuvent être réduites à une seule dimension en projetant chaque l’ensemble des données sur le premier axe principal (Plot 1B) [96-97].

## Chapitre.II: Méthodologie QSAR

L'ACP est particulièrement utile lorsque les variables, dans le jeu de données, sont fortement corrélées. La corrélation indique qu'il existe une redondance dans les données.

En raison de cette redondance, l'ACP peut être utilisée pour réduire les variables d'origine en un nombre plus petit de nouvelles variables (= **composantes principales**), ces dernières expliquant la plus grande partie de la variance contenue dans les variables d'origine.

Les limites de l'ACP viennent du fait que c'est une méthode de projection, et que la perte d'information induite par la projection peut entraîner des interprétations erronées.

### IV.1.1.2 La régression linéaire multiple (MLR)

La régression linéaire multiple MLR est l'une des méthodes de modélisation les plus populaires grâce à sa simplicité d'utilisation et facilité d'interprétation. L'avantage important de la régression linéaire multiple est qu'elle est très transparente, puisque l'algorithme est disponible, et que les prédictions peuvent être réalisées facilement [98].

L'analyse de régression linéaire multiple repose sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante  $Y$  et une série de  $n$  variables indépendantes  $X_i$ .

Pour les études de régression multiple, le nombre de variables doit être inférieur ou égal au nombre d'individus (molécules). L'objectif est d'obtenir une équation de la forme suivante :

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_nX_n \quad (27)$$

$X_1, \dots, X_n$  sont des descripteurs moléculaires affectés de leurs coefficients  $a_1, \dots, a_n$ .

Le résultat de l'analyse permet de fixer les valeurs des différents coefficients. La taille des coefficients peut révéler le degré d'influence des descripteurs moléculaires correspondants sur la propriété cible. De plus, un coefficient positif suggère que le descripteur moléculaire correspondant contribue positivement à la propriété cible, tandis qu'un coefficient négatif suggère une contribution négative. Cependant, ces interprétations peuvent ne pas être précises car les descripteurs colinéaires ont le potentiel d'influencer les coefficients de sorte que des valeurs erronées peuvent être attribuées.

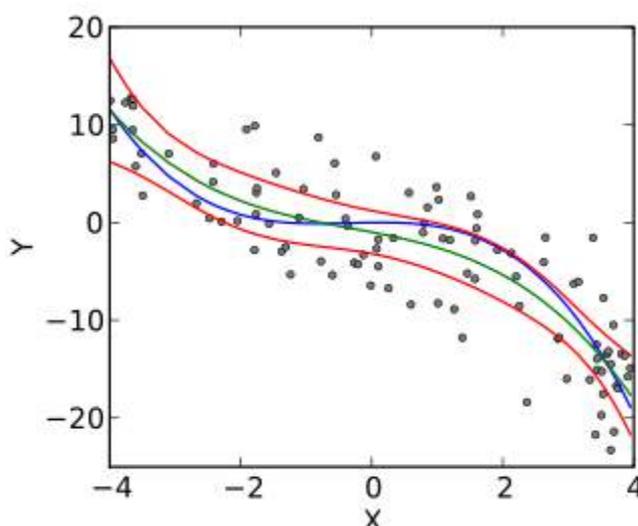


Figure 5: Régression linéaire multiple.

## Chapitre.II: Méthodologie QSAR

Le cas d'une seule variable explicative s'appelle une régression linéaire simple.

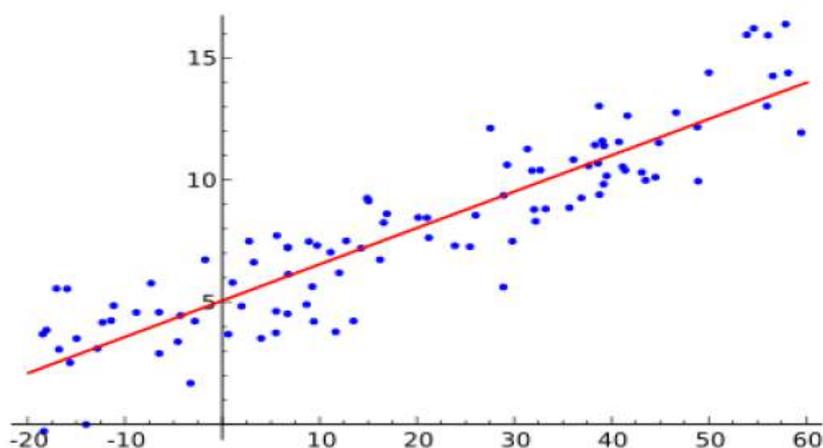


Figure 6: Régression linéaire simple

Les descripteurs moléculaires du modèle devraient être indépendants les uns des autres et le nombre d'instances pour la construction de modèles devrait être au moins cinq fois plus élevé que le nombre de descripteurs utilisés [99]. En outre, l'hypothèse d'une relation linéaire rend la MLR moins approprié pour modéliser des problèmes complexes comme la toxicité, où de multiples mécanismes peuvent interagir pour provoquer une réponse toxique.

À ce jour, la MLR reste utilisée avec des améliorations ou en combinaison avec la sélection de fonctionnalités pour améliorer ses performances. Des exemples d'améliorations sont: l'utilisation de l'analyse indépendante des composants en QSRP pour la solubilité aqueuse [100], régression paresseuse locale [101], rétro-régression appliquée sur les points d'ébullition des nonanes [102], sélection de caractéristiques d'ensemble [103] des méthodes de sélection comme l'algorithme génétique, la régression de crête.

Si le nombre de variables est élevé, il est plus judicieux d'utiliser la méthode du Partial Least Squares (PLS).

## Chapitre.II: Méthodologie QSAR

### IV.1.1.3 La méthode de régression des moindres carrés partiels

La régression par les moindres carrés partiels (PLS) est une technique qui diminue le nombre de descripteurs à un plus petit ensemble de composantes non corrélées et qui effectue la régression par les moindres carrés sur ces composantes, plutôt que sur les données initiales. La fonctionnalité PLS est particulièrement utile lorsque les descripteurs sont fortement colinéaires, ou lorsqu'il y a plus de descripteurs que d'observations et que la régression sur les moindres carrés échoue complètement ou produit des coefficients avec des erreurs élevées. Les moindres carrés partiels (PLS) ne supposent pas que les descripteurs sont fixes, à la différence de la régression multiple. Ainsi, les mesures des descripteurs tolèrent des erreurs, ce qui signifie que l'analyse PLS gère mieux l'incertitude des mesures [104-105].

La régression par les moindres carrés partiels est essentiellement utilisée dans les secteurs chimique, pharmaceutique, alimentaire et plastique. Une application courante consiste à modéliser la relation entre des mesures spectrales (spectre du proche infrarouge, IR UV), qui comprennent de nombreuses variables souvent corrélées les unes avec les autres, et une composition chimique ou d'autres propriétés physiques et chimiques. La régression PLS met l'accent sur le développement de modèles de prévision. Par conséquent, elle n'est généralement pas utilisée pour éliminer les variables qui ne permettent pas d'expliquer la réponse.

### IV.1.1.4 Analyse discriminante linéaire (LDA)

L'Analyse discriminante de Fisher, qui a été élaborée en 1936 par Ronald Aylmer Fisher, est couramment utilisée pour les problèmes de classification et également pour la réduction de la dimensionnalité. Elle est adaptée aux propriétés caractéristiques et des descripteurs moléculaires qui sont des variables continues [106].

LDA essaie de trouver un hyperplan qui peut le mieux séparer différentes classes d'une propriété cible. L'hyperplan est défini par une fonction discriminante linéaire qui est une combinaison linéaire de descripteurs moléculaires.

$$L = \sum_{i=1}^k w_i x_i \quad (28)$$

où L est le discriminant,  $x_1, \dots, x_k$  les descripteurs moléculaires et  $w_1, \dots, w_k$  leurs poids correspondants.

## Chapitre.II: Méthodologie QSAR

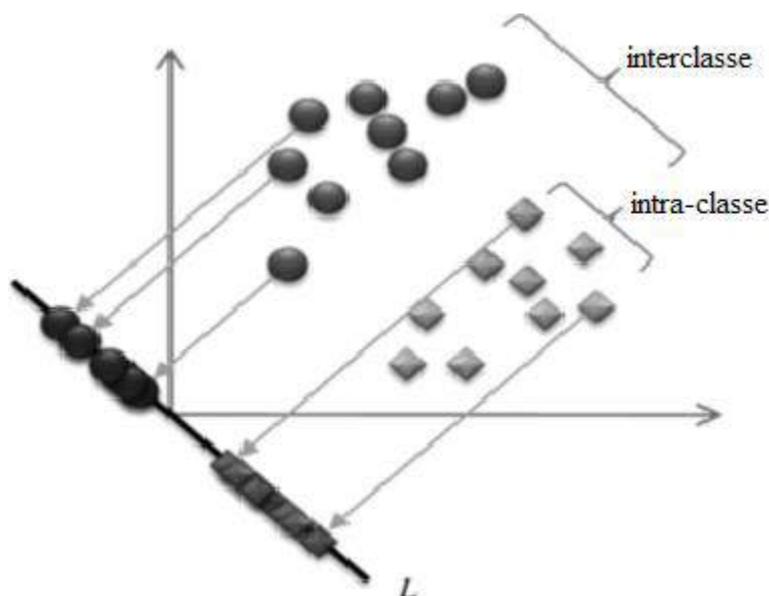


Figure.7 Distribution d'un échantillon à deux classes sur un axe transformé.

Comme l'illustre la Figure.6, la fonction discriminante  $L$  représente la fonction sur laquelle un ensemble de points de données est projeté. Cette fonction est obtenue en optimisant un ensemble de poids  $w_i$  qui maximise le rapport de la variance interclasse à la variance intra-classe pour obtenir la plus grande séparation de classe.

La classe d'un composé inconnu est attribuée en fonction de sa valeur du discriminant  $L$  qui peut être inférieur ou supérieur à un seuil  $C$ . Lorsque deux classes sont de cardinalité égale et ont des distributions similaires, une valeur possible pour  $C$  est la moyenne des discriminants moyens des deux classes :

$$C_{\text{seuil}} = \frac{\mu_{L1} + \mu_{L2}}{2} \quad (29)$$

LDA a été utilisé dans des études QSAR telles que: la prédiction de la répartition passive sang-cerveau [107], la mutagénicité [108], la prédiction des antiparasitaires contre différentes espèces de parasites [109], la prédiction des composés antitrypanosomiens [110], la toxicité aquatique du pesticide [111] l'activité inhibitrice de la télomérase [112], la prédiction des trichomonacides [113] et l'activité antiproliférative [114].

## Chapitre.II: Méthodologie QSAR

### IV.1.2 Approches non linéaires

D'autres méthodes, « non-linéaires », étendent l'approche QSAR à des relations plus complexes. Ces méthodes génèrent des modèles qui se révèlent être plus précis, spécialement pour des ensembles de données plus larges et plus diversifié. Cependant, ces modèles se heurtent parfois à des difficultés de compréhension et sont parfois sur-ajustés (ils se borneront dans ce cas à décrire du bruit au lieu de la relation sous-jacente entre descripteurs et activité). La recherche pharmaceutique tire un grand bénéfice de l'application des méthodes non linéaires telles que la classification de Bayes, la méthode des k plus proches voisins (k – NN), les réseaux de neurones, mais aussi des arbres de décision, ou des méthodes des machines à vecteurs de support (SVM ), le partitionnement récursif (RP113) et les méthodes de clustering.

Ces dernières années, la méthode des réseaux de neurones a gagné une popularité et est devenue l'une des principales méthodes d'analyse statistique en industrie pharmaceutique.

#### IV.1.2.1 Réseaux de neurones artificiels

Conçus à l'origine comme des modèles mathématiques très simplifiés de notre cerveau, les réseaux de neurones sont des systèmes de traitement de l'information. Concrètement, ces techniques neuronales sont des outils mathématiques et algorithmiques qui s'avèrent être puissants et commodes pour résoudre des problèmes complexes.

Le cerveau humain se compose de cellules distinctes appelées *neurones* formant un ensemble dense d'environ 10 à 100 milliards d'unités interconnectées. La principale caractéristique de ces neurones est qu'ils permettent de véhiculer et de traiter des informations en faisant circuler des messages électriques dans le réseau ténu et massivement parallèle formé par leur *axone*. L'extrémité des axones se divise en une multitude de ramifications [115].

A l'inverse, les arborescences qui amènent l'information vers le corps cellulaire sont appelés *dendrites*. Les informations sont transmises d'un neurone à l'autre, de manière unidirectionnelle, par l'intermédiaire de points de jonction appelés *synapses*. En général, le nombre de connexions entre neurones est de 1000 à 10000 synapses par neurone. La figure suivante présente le schéma d'un neurone biologique.

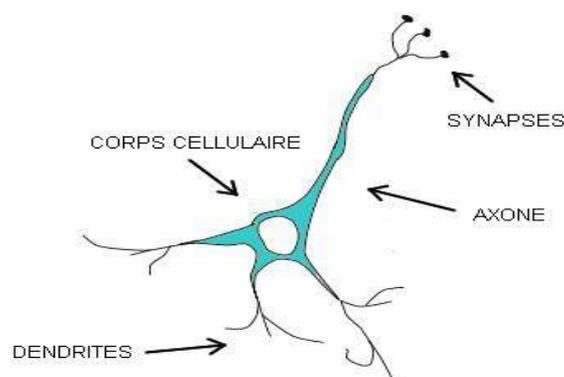


Figure.8 : le schéma d'un neurone biologique

## Chapitre.II: Méthodologie OSAR

C'est en 1943 que le premier modèle de neurone formel est proposé. Il s'inspire du neurone biologique : il possède un certain nombre d'entrées, similaires aux dendrites du neurone biologique, un corps servant d'unité de traitement, et un axone permettant la transmission d'un potentiel d'action à d'autres neurones [116].

### *Neurone formel :*

Le neurone formel est un modèle mathématique simplifié du neurone biologique, il présente un certain nombre d'entrées, les dendrites, un corps traitant les entrées suivant la méthode de tout ou rien, et un axone véhiculant la réponse du neurone.

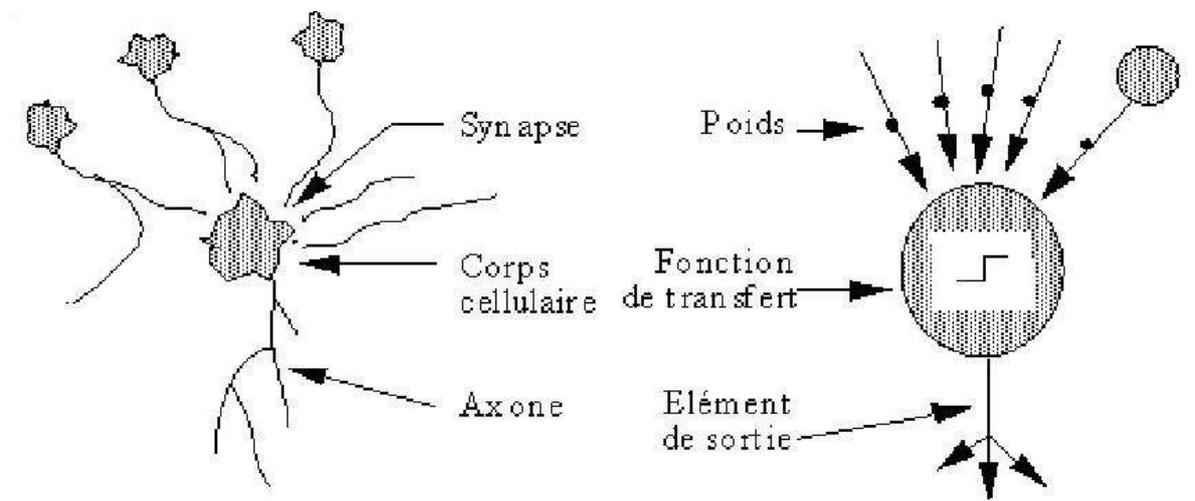


Figure.9 : le schéma d'un neurone formel.

Un réseau de neurones est un processeur massivement distribué en parallèle qui a une propension naturelle pour stocker de la connaissance empirique et la rendre disponible à l'usage. Il ressemble au cerveau sur deux aspects :

1. La connaissance est acquise par le réseau au travers d'un processus d'apprentissage.
2. Les connexions entre les neurones, connues sous le nom de poids synaptiques servent à stocker la connaissance

Depuis une dizaine d'années, l'utilisation des réseaux de neurones artificiels (RNA) s'est développée dans de nombreuses disciplines (sciences économiques, écologie et environnement, chimie, biologie et médecine...). Ils sont notamment appliqués pour résoudre des problèmes de classification, de prédiction, de catégorisation, d'optimisation, de reconnaissance des formes et de mémoire associative [117].

Dans le cadre du traitement des données, les RNA constituent une méthode d'approximation de systèmes complexes, particulièrement utile lorsque ces systèmes sont difficiles à modéliser à l'aide des méthodes statistiques classiques. Les RNA sont également applicables dans toutes les situations où il existe une relation non linéaire entre une variable prédictive et une variable prédite.

## Chapitre.II: Méthodologie OSAR

Par leur nature et leur fonctionnement, les RNA peuvent détecter les interactions multiples non linéaires parmi une série de variables d'entrée, ils peuvent donc gérer des relations complexes entre les variables indépendantes et les variables dépendantes.

Avec les méthodes statistiques traditionnelles, il faut d'abord penser à un modèle, le tester, penser à un autre modèle jusqu'à ce qu'on obtienne un modèle suffisamment précis. Un réseau est entraîné sur des données grâce à un mécanisme d'apprentissage qui agit sur les constituants du réseau pour réaliser au mieux la tâche souhaitée.

Les réseaux de neurones fonctionnent en répartissant les valeurs des variables dans des automates (les neurones). Ces unités sont chargées de combiner entre elles leurs informations pour déterminer la valeur du paramètre de discrimination. C'est de la connexion de ces unités entre elles qu'émerge la capacité de discrimination du RNA. Chaque neurone reçoit des informations numériques en provenance de neurones voisins ; à chacune de ces valeurs est associé un poids représentatif de la force de la connexion. Chaque neurone effectue localement un calcul dont le résultat est transmis ensuite aux neurones avals [118].

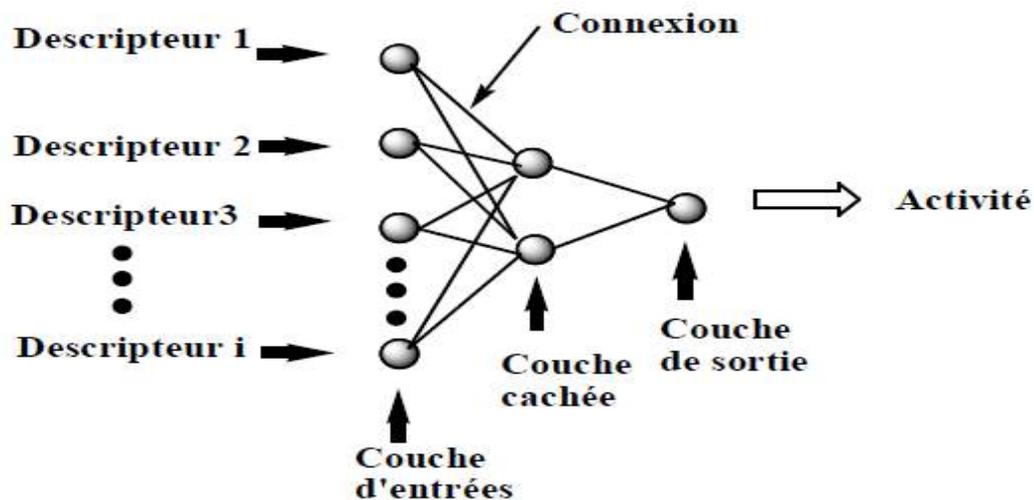


Figure.10 : Architecture des réseaux neurones

Les avantages d'un réseau de neurones sont :

- ✓ La possibilité de fournir une réponse soit linéaire soit non linéaire selon le contexte. Ceci est très utile car dans le cas de prédiction de l'activité biologique de molécules les données structures moléculaires évoluent souvent de façon non linéaire et les modes de calcul linéaires peuvent s'avérer insuffisants.
- ✓ La possibilité de construire un modèle uniquement basé sur la connaissance des données d'entrée et de sortie, sans nécessité de connaître le type de relation mathématique qui les relie.
- ✓ La possibilité d'utiliser l'RNA dans un environnement différent de celui qui a servi à la phase d'apprentissage grâce à sa capacité d'adaptation et sa faible sensibilité aux

## Chapitre.II: Méthodologie OSAR

données aberrantes. En effet, chaque neurone est sensible à l'activité globale générée par tous les autres neurones collectivement.

- ✓ La possibilité de pouvoir interpréter un résultat tel qu'un tri avec un taux de confiance.
- ✓ La possibilité de continuer à fonctionner malgré une panne ou un dysfonctionnement de l'un des neurones. Dans ce cas, la réponse de l'RNA est légèrement dégradée mais reste viable car l'information a été transmise de façon distribuée sur tous les neurones. L'RNA est donc particulièrement robuste [119-120].

### IV.1.2.2 Arbre de décision

En théorie des graphes, un *arbre* est un graphe non orienté, acyclique et connexe. L'ensemble des nœuds se divise en trois catégories :

- Nœud *racine* (l'accès à l'arbre se fait par ce nœud) ;
- Nœuds *internes* : les nœuds qui ont des descendants (ou *enfants*), qui sont à leur tour des nœuds ;
- Nœuds *terminaux* (ou *feuilles*) : nœuds qui n'ont pas de descendant.

Un arbre de décision est un schéma qui représente les résultats possibles d'une série de choix interconnectés. Il permet à une personne ou une organisation d'évaluer différentes actions possibles en fonction de leur coût, leur probabilité et leurs bénéfices. Il peut être utilisé pour alimenter une discussion informelle ou pour générer un algorithme qui détermine le meilleur choix de façon mathématique.

Les arbres de décision (AD) sont une catégorie d'arbres utilisée dans l'exploration de données et en informatique décisionnelle. Ils emploient une représentation hiérarchique de la structure des données sous forme des séquences de décisions (tests) en vue de la prédiction d'un résultat ou d'une classe. Chaque individu (ou observation), qui doit être attribué(e) à une classe, est décrit(e) par un ensemble de variables qui sont testées dans les nœuds de l'arbre. Les tests s'effectuent dans les nœuds internes et les décisions sont prise dans les nœuds feuille [121-124].

Par exemple, l'arbre de la figure.11 ci-dessus décide une réponse booléenne (classification dans l'ensemble {oui, non}) pour la question « Cette présentation est-elle intéressante ? » en fonction des valeurs discrètes des attributs {difficile, durée, motivation, surprenant}.

## Chapitre.II: Méthodologie OSAR

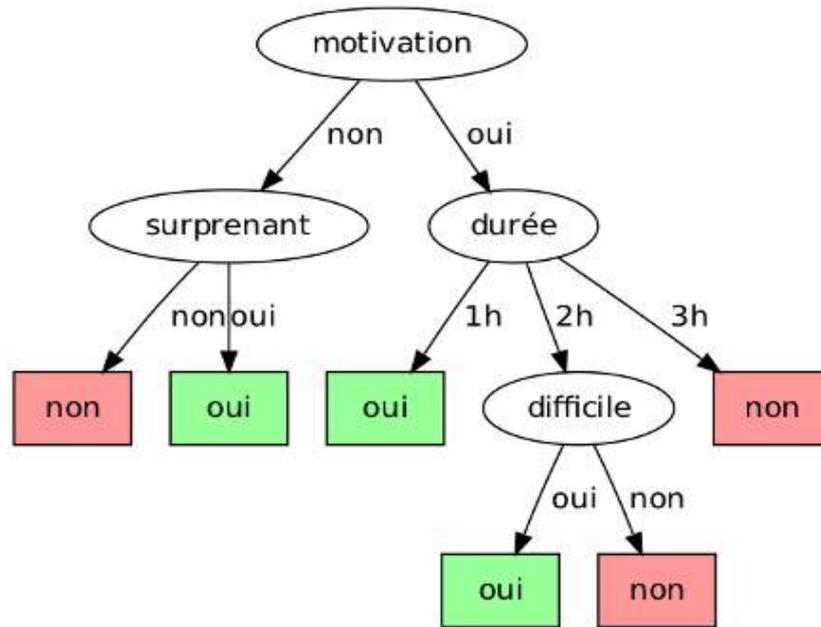


Figure 11: Exemple d'un arbre de décision

Dans l'exploration de données, les arbres de décision peuvent également être décrits comme la combinaison de techniques mathématiques et de calcul pour faciliter la description, la catégorisation et la généralisation d'un ensemble de données.

### IV.1.2.3 Machine à vecteurs de support

Les machines à vecteurs de support ou séparateurs à vaste marge (en anglais Support Vector Machine, SVM) sont un ensemble de techniques d'apprentissage supervisé destinées à résoudre des problèmes de discrimination et de régression basés sur la théorie de l'apprentissage statistique (généralement considérés comme la 1<sup>ère</sup> réalisation pratique de cette théorie [125]) et qui respecte les principes du (SRM) « structural risk minimization ». L'émergence des SVMs a commencé autour des débuts des années 1990 par les mathématiciens russes Vladimir Vapnik et Alex Chervonenkis [126-131].

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

Vu leur composition comme des techniques d'apprentissage, les SVMs sont utilisés dans les domaines de :

- Reconnaissance de formes/Classification :
  - Vision Machine : Identification de visage, reconnaissance d'expression faciale.
  - Catégorisation de texte
- Bioinformatique : prédiction de la structure des protéines, prédiction du progrès d'une maladie.
- Régression : estimation et prédiction des valeurs des fonctions.

## **Chapitre.II: Méthodologie QSAR**

L'application des SVMs s'est étendue aux domaines d'apprentissage non supervisé comme :

- Réduction de dimension : ACP non linéaire
- Clustering

L'algorithme SVM est implémenté en pratique en utilisant un noyau, l'apprentissage de l'hyperplan en SVM linéaire est fait en transformant le problème en utilisant une algèbre linéaire, ce qui sort du cadre de cette introduction à SVM.

Une idée puissante est que le SVM linéaire peut être reformulé en utilisant le produit interne de deux observations quelconques, plutôt que les observations elles-mêmes. Le produit interne entre deux vecteurs est la somme de la multiplication de chaque paire de valeurs d'entrée.

## Chapitre.II: Méthodologie QSAR

### V. Validation du modèle

La nouvelle politique de la commission européenne sur les substances chimiques (REACH: Enregistrement, Evaluation et Autorisation des Produits Chimiques) [132] stipule explicitement qu'au niveau de l'enregistrement des produits chimiques devrait inclure des informations provenant de sources alternatives (comme des modèles QSAR) qui peuvent aider à identifier la présence ou l'absence de propriétés dangereuses de la substance et qui peuvent dans certains cas remplacer les résultats d'essais sur les animaux. Évidemment, aux fins de la législation REACH, il est essentiel d'utiliser des modèles QSAR qui produisent des estimations fiables, c'est-à-dire des modèles QSAR validés [133].

La validation des modèles a fait l'objet de beaucoup de débats dans les milieux scientifiques et réglementaires. Il a été jugé important d'élaborer un ensemble de principes internationalement reconnus pour la validation QSAR, de fournir aux organismes de réglementation une base scientifique pour prendre des décisions sur l'acceptabilité des estimations QSAR des paramètres réglementaires et de promouvoir l'acceptation des modèles QSAR. Plusieurs principes d'évaluation de la validité des QSAR ont été proposés en 2002 sous le nom de «Principes de Setubal», lors d'un atelier international tenu à Setubal (Portugal) [134]; ceux-ci ont ensuite été modifiés en 2004 par le programme de travail de l'OCDE sur les QSAR, en tant que principes de l'OCDE [135] pour la validation QSAR.

Pour faciliter la prise en compte d'un modèle QSAR à des fins de réglementation, il convient d'associer ce dernier aux informations suivantes :

- ✓ Un point final défini;
- ✓ Un algorithme non ambigu;
- ✓ Un domaine d'applicabilité défini;
- ✓ Des mesures appropriées de la qualité de l'ajustement, de la robustesse et de la prédictivité;
- ✓ Une interprétation mécanistique, si possible. La révision la plus substantielle des principes Setubal dans les nouveaux principes de l'OCDE a été l'unification des principes précédents (validation interne) et (validation externe) en un seul contrat.

Tout modèle QSAR doit être correctement validé avant son utilisation pour interpréter et prédire les réponses biologiques des composés non étudiés. Mais la question se pose: comment s'assurer de la validité d'un QSAR spécifique et qu'entend-on par qualité du modèle? Il existe plusieurs façons d'exprimer la performance d'un modèle. L'approche classique adoptée dans l'analyse QSAR, basée sur la régression linéaire multiple (MLR), consiste à prendre en compte  $R^2$ , le «coefficient de corrélation» et S, l'écart type résiduel (RSD). La première quantité varie entre 0 et 1, 1 signifie un modèle parfait, expliquant 100% des données de réponse (Y), et 0 un modèle dépourvu de tout pouvoir explicatif. Ainsi, un  $R^2$  élevé (proche de 1) et un RSD bas sont des conditions nécessaires à la validité du modèle. Cependant, les excellentes valeurs de  $R^2$  et RSD ne constituent pas des indicateurs suffisants de la validité du modèle. Cela dépend de la propriété des modèles de régression pour un meilleur ajustement: plus la capacité de réplification du modèle est bonne à partir des données, plus de paramètres et de termes sont incorporés au modèle [136-138].

## Chapitre.II: Méthodologie QSAR

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

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (30)$$

Où :

$y$ : La valeur expérimentale de l'activité

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

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

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

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

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

$$R^2_{adj} = \sqrt{\frac{R^2(n-1) - p}{n - p - 1}} \quad (31)$$

Avec :  $n$  est le nombre des variables dépendantes (les molécules) ;  $p$  est le nombre de variables indépendantes (les descripteurs) ;  $R^2$  est le coefficient de détermination.

- ✓ La déviation standard  $S$  est un autre indicateur statistique utilisé

$$S = \sqrt{\frac{\sum(\hat{y}_i - y_i)^2}{n - p - 1}} \quad (32)$$

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

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

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

$$F = \frac{\sum((\hat{y}_i - \bar{y}_i)^2(n - p - 1))}{(y_i - \hat{y}_i)^2 \cdot p} \quad (33)$$

## **Chapitre.II: Méthodologie QSAR**

Cependant, quatre outils d'évaluation de la validité des modèles QSAR peuvent être différenciés. Ce sont: (i) la randomisation des données de réponse dans des vecteurs de réponse réordonnés, (ii) la validation croisée, (iii) La fragmentation des composés chimiques en un ensemble d'apprentissage et de validation, et (iv) une confirmation à l'aide d'un ensemble de validation externe indépendant. Sans le luxe d'un ensemble de validation indépendant, considéré comme le plus fiable de ces outils, la validité du modèle peut donc être vérifiée par l'une ou l'autre des trois procédures.

### *Outil 1: Randomisation des réponses dans un tableau des variables réordonnées*

Le premier des quatre outils est basé sur des randomisations répétitives des données de réponse (Y) de N composés dans l'ensemble d'apprentissage. Ainsi, un générateur de nombres aléatoires est utilisé pour affecter les entiers compris entre 1 et N à des séquences de N nombres. Dans chaque cycle, la disposition résultante de nombres entiers aléatoires est utilisée afin de réordonner les données Y - en laissant les données X intactes - puis l'analyse complète des données est effectuée sur ces données brouillées. Chaque exécution donnera des estimations de  $R^2$  et  $Q^2$ , qui sont enregistrées. Si, dans chaque cas, les données brouillées donnent des valeurs  $R^2$  et  $Q^2$  bien inférieures aux données d'origine, on peut alors avoir confiance en la pertinence du modèle «réel» QSAR. La randomisation des données Y un certain nombre de fois (au moins dix) donne une idée assez précise de l'importance de la QSAR réelle, mais pour améliorer la précision du niveau de probabilité, il est généralement nécessaire de disposer de centaines d'analyses de données rerandomisées. Lorsque des centaines d'essais ont été réalisés, les histogrammes de  $R^2$  et  $Q^2$  peuvent fournir une estimation précise du niveau de signification du modèle QSAR réel [139-140].

### *Outil 2: Validation croisée*

Contrairement à la méthode précédente, la validation croisée (CV) est basée sur des prédictions. CV fonctionne en effectuant un nombre (G) de modifications légèrement réduites sur le jeu de données parent, en estimant les paramètres de chacun de ces jeux de données modifiés, puis en calculant la précision des prédictions par chacun des modèles résultants. Ainsi, CV crée G ensembles de données modifiées en enlevant un ou plusieurs groupes de composés de manière à ce que chaque observation (ici: composé) soit supprimée une fois, et une seule fois, sur le nombre total de cycles CV.

G, le modèle est ensuite ajusté aux données, dépourvu de la partie omise, et est ensuite utilisé pour calculer des prévisions sur la base des données de réponse des composés laissés de côté. Ceci est répété pour chaque ensemble de données modifié, après quoi les différences au carré entre les valeurs de réponse prédites et réelles sont résumées pour former PRESS (Somme prédictive récurrente des carrés ou, alternativement, Somme d'erreur de prédiction des carrés). En fin de compte, PRESS contiendra une contribution de chaque observation et constitue donc un bon indicateur de la capacité prédictive réelle du modèle. Ensuite, PRESS est comparé à la somme des carrés des valeurs de réponse (SS,) et si le premier est plus petit que le second, la QSAR prédit mieux que le hasard et peut être considérée comme «statistiquement significative». Ceci parce que la meilleure "estimation" pour l'activité de chaque composé, ce qui donne "l'erreur d'estimation" égale à SS.

## Chapitre.II: Méthodologie QSAR

Les performances prédictives du modèle QSAR peuvent être exprimées sous la forme  $Q^2$ , la variance prédite ou validée de manière croisée, qui est (1-PRESS / SS) et accompagne le paramètre  $R^2$ . Dans certaines situations, le CV peut ne pas fonctionner comme on le souhaite. Le premier est lorsque les composés sont considérablement regroupés et ne sont donc pas indépendants. Cela peut par exemple se produire lorsque deux types de composés différents ou plus sont incorporés dans le même modèle et que l'activité de ces composés ne diffère que selon ce groupe. Tout modèle tiendra alors principalement compte de cette différence d'activité entre les groupes et le CV, ainsi que tout autre test de signification (à l'exception de la randomisation), identifiera cette trivialité. Une autre situation où le CV est trompeur se produit lorsque le CV est appliqué après sélection de variables dans une MLR par étapes. Ici, le problème est que le modèle final est basé sur des descripteurs sélectionnés en fonction de leur corrélation avec les valeurs de réponse, et que les corrélations apparentes résultantes sont également, rétrospectivement, stables en CV. Enfin, le CV peut donner des résultats trop conservateurs si la matrice X est générée à partir d'un plan statistique expérimental orthogonal, mais ceci est exceptionnellement rare dans les QSAR [141-143].

La procédure de CV est alors souvent appelée "laisser-sortir-un" (LOO). Toutefois, pour des raisons de calcul, les CV avec des modèles multivariés (PLS, PCA, etc.) ont généralement été exécutés avec beaucoup moins de groupes, généralement entre cinq et dix. Il est à noter que d'une manière théorique et pratique, l'approche «laisser-plusieurs-sortir» (LMO) est préférable à LOO. Ce résultat peut être compris lorsque l'on considère ce qui se passe lorsque le nombre de composés N augmente. La technique LMO laisse toujours de côté une certaine partie des données, créant ainsi une perturbation constante de la structure de données. L'approche LOO perturbe la structure des données en supprimant 1 de N à chaque tour du CV, ce qui entraîne une perturbation de plus en plus petite avec l'augmentation de N. Par conséquent, à la limite, le  $Q^2$  de LOO approche  $R^2$ , ce qui est très peu satisfaisant.

### *Outil 3: Division du jeu de données parent en jeux de formation et de validation*

La validation croisée fournit une approximation raisonnable de la capacité avec laquelle le QSAR prédit les valeurs d'activité de nouveaux composés. On parle généralement de validation interne, car tous les produits chimiques considérés appartiennent au même ensemble de données. Toutefois, si le nombre de composés disponibles est suffisamment important, ils peuvent être divisés pour former un ensemble d'entraînement séparé et un ensemble de validation séparé, permettant ainsi la validation externe. Cette subdivision de l'ensemble de données peut être réalisée de nombreuses manières, mais il est souhaitable que les deux séries de composés couvrent environ des plages approximativement similaires des réponses biologiques et des propriétés structurelles [144].

## **Chapitre.II: Méthodologie QSAR**

### *Outil 4: Validation externe à l'aide d'un jeu de validation conçu*

Une étape souvent négligée dans le QSAR est la sélection des ensembles appropriés de formation et de validation, c'est-à-dire comment sélectionner les ensembles pour répondre au critère statistique fondamental de la représentativité. L'ensemble de formation et les composés de validation doivent être représentatifs de la classe de composés dont ils sont issus, ce qui signifie qu'ils doivent être choisis de manière à couvrir de manière adéquate les propriétés chimiques et structurelles des composés considérés. Un moyen pratique d'obtenir de tels ensembles de produits chimiques consiste à utiliser un modèle expérimental statistique, qui a déjà été abordé par Sjöström et ses collaborateurs [145]. Le recours à un modèle expérimental statistique pour générer des ensembles bien équilibrés d'entraînement et de validation de composés représentatifs est peu fréquent dans les QSAR [145]. Cependant, dans le cas où cela aurait été fait, il existerait un ensemble de validation couvrant l'ensemble de l'espace X de manière égale et indépendant de l'ensemble d'apprentissage. À condition que des analogies et des relations prévalent entre les propriétés chimiques et structurelles et les réponses biologiques de ces produits chimiques, ce type de jeu de validation de haute qualité permettrait de valider expérimentalement la QSAR sur l'ensemble de la plage d'activités biologiques.

## Chapitre.II: Méthodologie QSAR

### VI. Conclusion

Le développement de l'outil informatique aussi bien au niveau théorique qu'au niveau matériel a permis la mise en place de techniques de simulation numérique centrées sur la chimie et la biologie [146-149]. Par analogie avec les expressions *in vivo* et *in vitro*, le terme « *in silico* » a été introduit pour qualifier la mise en œuvre de ces méthodes numériques. De par son nom, ce terme fait référence au silicium, matériau principal retrouvé dans le matériel informatique. Le champ *in silico* regroupe un très large ensemble de méthodes numériques fondées sur les lois de la physique et de la chimie qui, utilisant les approches des mathématiques, permettent de simuler ou de modéliser un phénomène [150-154]

Les méthodes *in silico* sont complémentaires des études *in vivo* et *in vitro* et ne peuvent remplacer l'expérience. Elles apportent des informations sur la nature des interactions entre les principes actifs et les récepteurs et sont nécessaires à la bonne compréhension de ces mécanismes. Par la suite, les résultats de telles études peuvent être utilisés dans le développement de nouvelles expériences ou lors de la sélection de molécules à tester. En d'autres termes, l'utilisation combinée de la théorie et de l'expérience engendre une meilleure compréhension des phénomènes traités dans un domaine de recherche [155-158]

## Chapitre.II: Méthodologie OSAR

### Bibliographie

- [1]. Sneader, W. Drug Discovery, A History, ed. J.W.S. Ltd. 2005. 472.
- [2]. Schmitz, R. Friedrich Wilhelm Serturmer and the discovery of morphine. *PharmHist*, 1985. 27(2): p. 61-74.
- [3]. Gerhardt, C. Recherches sur les acides organiques anhydres. *Ann. de Chim. et de Phys.*, 1853. 37(3): p. 285-342.
- [4]. Gerhardt, C. Untersuchungenüber die WassefreienorganischenSäuren. *Ann. d. Chem. u. Pharm.*, 1853. 87: p. 149-179.
- [5]. Lafont, O. Du saule à l'aspirine. *Revue d'histoire de la pharmacie*, 2007. 94(354): p. 209-216.
- [6]. Nikiéma, J.B.;Djierro, K.; Simporé, J.; et al. Stratégie d'utilisation des substances naturelles dans la prise en charge des personnes vivant avec le VIH: expérience du Burkina Faso. *Ethnopharmacologia*, 2009. 43: p. 47-51.
- [7]. Règlement (CE) n° 1907/2006 du Parlement Européen et du Conseil du 18 décembre 2006 concernant l'enregistrement, l'évaluation et l'autorisation des substances chimiques, ainsi que les restrictions applicables à ces substances (REACH), instituant une agence européenne des produits chimiques, modifiant la directive 1999/45/CE et abrogeant le règlement (CEE) n°793/93 du Conseil et le règlement (CE) n° 1488/94 de la Commission ainsi que la directive 76/769/CEE du Conseil et les directives 91/155/CEE, 93/67/CEE, 93/105/CE et 2000/21/CE de la Commission.
- [8]. N. Margossian, Le règlement REACH - La réglementation européenne sur les produits chimiques, Dunod / L'Usine Nouvelle, Paris, 2007
- [9]. Institut National de Recherche et de Sécurité (INRS). [www.inrs.fr](http://www.inrs.fr) [20 janvier 2010]
- [10]. European Chemicals Agency (ECHA). <http://echa.europa.eu/> [28 août 2009]
- [11]. T. I. Oprea, "Chemoinformatics in Drug Discovery" Ed. WILEY-VCH Verlag.Allemagne, 2005.
- [12]. F. Li, H. Wu, L. Li, X. Li, J. Zhao, W. J. G. M. Peijnenburg, *Ecotoxicol. Environm.Saf.* 80, 2012, 273–279.
- [13]. A. Srinivas Reddy, S. Kumar, R. Garg, *J Mol Graph Model.* 28(8), 2010, 852–862.
- [14]. A. Crum-Brown and T. R. Fraser, *Trans. R. Soc. Edinburgh*, 25, 151 (1868).
- [15]. C. Richet and C. R. Seancs, *Soc. Biol. Ses. Fil.*, 9, 775 (1893).
- [16]. H. Meyer, *Arch. Exp. Pathol. Pharmakol.*, 42, 109 (1899).
- [17]. E. Overton, *StudienÜber die Narkose*, Fischer, Jena, Germany, 1901
- [18]. J. Ferguson, *Proc. R. Soc. London Ser. B*, 127, 387 (1939).
- [19]. A. Albert, S. Rubbo, R. Goldacre, M. Darcy, and J. Stove, *Br. J. Exp. Pathol.*, 26, 160 (1945).
- [20]. A. Albert, *Selective Toxicity: The Physicochemical Bases of Therapy*, 7th ed., Chapman and Hall, London, 1985, p. 33.
- [21]. P. H. Bell and R. O. Roblin, Jr.*J. Am. Chem. Soc.*, 64, 2905 (1942).
- [22]. L. P. Hammett, *Chem. Rev.*, 17, 125 (1935).
- [23]. L. P. Hammett, *Physical Organic Chemistry*, 2nd ed., McGraw-Hill, New York, 1970.
- [24]. R. W. Taft, *J. Am. Chem. Soc.*, 74, 3120 (1952).

## **Chapitre.II: Méthodologie QSAR**

- [25]. C. Hansch, P. P. Maloney, T. Fujita, and R. M. Muir, *Nature*, 194, 178 (1962).
- [26]. R. Nelson Smith, C. Hansch, and M. M. Ames, *J. Pharm. Sci.*, 64, 599 (1975).
- [27]. T. Fujita, J. Iwasa, and C. Hansch, *J. Am. Chem. Soc.*, 86, 5175 (1964).
- [28]. C. D. Selassie, "History of quantitative Structure-Activity relationships", in Book "History of quantitative Structure-Activity relationships", D. J. Abraham, Editor. 2003, John Wiley & Sons, Inc.: California
- [29]. A.Z. Dudek, T. Arodz, and J. Gàlvez, —Computational methods in developing quantitative structure-activity relationships (QSARs): A review, *Combinatorial Chemistry & High Throughput Screening*, 9, 2006, 213–228
- [30]. M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley, New York, 2000
- [31]. R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley, Weinheim, 2000
- [32]. S.W. Benson, J.H. Buss, *J. Chem. Phys.*, 1958, 29, 546-572.
- [33]. H. Wiener, *J. Am. Chem. Soc.*, 2002, 69, 17-20.
- [34]. P. Geerlings, F. De Proft, W. Langenaeker, *Chem. Rev.*, 2003, 103, 1793-1874
- [35]. H. Chermette, *J. Comput. Chem.*, 1999, 20, 129-154.
- [36]. Wiener, H. Structural determination of paraffin boiling points. *Journal of Chemical Information and Computer Sciences*, 1947, 69, p. 17-20.
- [37]. Randić, M. On characterization of molecular branching. *Journal of the American Chemical Society*, 1975, 97, p. 6609-6614.
- [38]. Balaban, A.T. Highly discriminating distance-based topological index. *Chemical Physics Letters*, 1982, 89, p. 399-404
- [39]. P. Labute, A widely applicable set of descriptors, *Journal of Molecular Graphics and Modelling*, 18, 2000, 464–477
- [40]. R. Bosque, J. Sales, E. Bosch, M. Rosès, M.C. Garcia-Alvarez-Coque, and J.R. Torres-Lapasio, A QSPR study of the p-solute polarity parameter to estimate retention HPLC, *Journal of Chemical Information and Modeling*, 43, 2003, 1240–1247
- [41]. V.N. Viswanadhan, A.K. Ghose, G.R. Revankar and R.K. Robins, —Atomic physicochemical parameters for 3D structure directed quantitative structure-activity relationships: Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics, *Journal of Chemical Information and Modeling*, 29, 1989, 163–172
- [42]. J. Schuur, P. Selzer, J. Gasteiger, The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity. *J. Chem. Inf. Comput. Sci.* 1996, 36, 334–344.
- [43]. R.D. Cramer, D.E. Patterson, and J.D. Bunce, —Comparative molecular field analysis (CoMFA): Effect of shape on binding of steroids to carrier proteins, *Journal of the American Chemical Society*, 110(18), 1988, 5959-5967.
- [44]. C. Navajas, A. Poso, K. Tuppurainen, and J. Gynther, —Comparative Molecular Field Analysis (CoMFA) of MX Compounds using different Semi-Empirical Methods: LUMO Field and its Correlation with Mutagenic Activity, *Molecular Informatics*, 15(3), 1996, 189–193.

## Chapitre.II: Méthodologie QSAR

- [45]. D.A. McQuarrie, *Statistical Thermodynamics*, harper row publishers, New York, 1973.
- [46]. K. Fukui, —Theory of Orientation and Stereoselection, *Reactivity and Structure Concepts in Organic Chemistry*, 2, 1975, 34–39.
- [47]. R. Franke, —Theoretical Drug Design Methods, *Elsevier Amsterdam*, 1984, 115–123.
- [48]. D.F.V. Lewis, C. Ioannides, and D.V. Parke, —Interaction of a series of nitriles with the alcoholinducible isoform of P450: Computer analysis of structure—activity relationships, *Xenobiotica*, 24(5), 1994, 401–408.
- [49]. Z. Zhou and R.G. Parr, —Activation hardness: new index for describing the orientation of electrophilic aromatic substitution, *Journal of the American Chemical Society*, 112(15), 1990, 5720–5724.
- [50]. A.I. Akhiezer and S.V., —Peltminskii, *Methods of Statistical Physics*, PergamonPress,Oxford, 1981.
- [51]. P.W. Atkins, —Physical Chemistry, 2nd edition, W.H. Freeman and Company, San Francisco,1982.
- [52]. M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. Vreven, K. N. Jr., Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J.Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W.Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G.Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck,K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski,B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, 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, C. Gonzalez and J.A. Pople, —Gaussian 03, Revision B.04, Gaussian, Inc. Pittsburgh PA, 2003.
- [53]. ChemOffice, PerkinElmerInformatics, 2010.
- [54]. ACDLABS 10, Advanced Chemistry Development Inc., Toronto, ON, Canada, 2015
- [55]. Marvin Sketch 5.11.4, ChemAxon, 2012.
- [56]. QSARIS. [www.scivision.com/qsaris.html](http://www.scivision.com/qsaris.html).
- [57]. Cerius2. [www.accelrys.com/products/cerius2](http://www.accelrys.com/products/cerius2).
- [58]. Vol Surf. [www.moldiscovery.com/softvolsurf.php](http://www.moldiscovery.com/softvolsurf.php).
- [59]. DRAGON. <http://www.taletе.mi.it/products/dragondescription.htm>.
- [60]. AR. Katritzky, M. Karelson, R. Petrukhin, CODESSA PRO, University of Florida 2001-2005.<http://www.codessa-pro.com>.
- [61]. Solov'ev, V. P.; Varnek, A. A. *ISIDA (In Silico Design and Data Analysis) program*; version 5.79; 2008 - 2014. <http://infochim.ustrasbg.fr/spip.php?rubrique53><http://vpsolovev.ru/programs/>
- [62]. <http://openmopac.net/>

## Chapitre.II: Méthodologie OSAR

- [63]. Schrödinger E., Quantization as an Eigenvalue Problem, *Ann. Phys. Leipzig.*, 79, (1926),361.
- [64]. Szabo, A.; Ostlund, N. S. in: *Modern Quantum chemistry*, Mc Graw-Hill, 1982.
- [65]. Born M. R., Oppenheimer, *Zur Quantentheorie der Molekeln*, *Ann. Phys.*, (Leipzig), 84, (1927), 457.
- [66]. Pauli W., Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren, *Z. Phys.*, 31, (1925), 765.
- [67]. Minkine V., B. Simkine, R. Minaev, *Théorie de la structure moléculaire*, (1982), Edition Mir, Moscou.
- [68]. Fock V., Z. Näherungsmethode zur Lösung des quanten-mechanischen Mehrkörperprobleme, *Physik.*, 61, (1930), 126.
- [69]. Roothaan C. C. J., *New Developments in Molecular Orbital Theory*, *Rev. Mod. Phys.*, 23,(1951), 69.
- [70]. C. Møller et M.S. Plesset, « Note on an Approximation Treatment for Many-Electron Systems », *Phys. Rev.*, vol. 46, 1934, p. 618–622
- [71]. Hohenberg P., Kohn, W. Inhomogeneous Electron Gas, *Phys. Rev.*, 136, (1964), 864.
- [72]. Kohn W., Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects, *Phys. Rev. A.*, 140, (1965), 1133.
- [73]. P. Geerlings, F. De Proft, and W. Langenaeker. Conceptual density functional theory. *Chemical Reviews*, 103(5):1793{1873, 2003.
- [74]. R. G. Parr, R. A. Donnelly, M. Levy, and W. E. Palke. Electronegativity - Density Functional Viewpoint. *Journal of Chemical Physics*, 68(8):3801{ 3807, 1978.
- [75]. A. Kourounakis and N. Bodor. Quantitative structure activity relationships of catechol derivatives on nerve growth factor secretion in L-M cells. *Pharmaceutical research*, 12(8):1199{1204, 1995.
- [76]. R. S. Mulliken. A new Electronegativity Scale; Together with Data on Valence States and on Valence Ionization Potentials and Electron Affinities. *Journal of Chemical Physics*, 2(11):782{793, 1934.
- [77]. T. A. Koopmans. Ordering of Wave Functions and Eigenenergies to the Individual Electrons of an Atom. *Physica*, 1(1):104{113, 1933.
- [78]. R. G. Parr and R. G. Pearson. Absolute Hardness - Companion Parameter to Absolute Electronegativity. *Journal of the American Chemical Society*, 105(26):7512{7516, 1983.
- [79]. W. T. Yang and R. G. Parr. Hardness, Softness, and the Fukui Function in the Electronic Theory of Metals and Catalysis. *Proceedings of the National Academy of Sciences of the United States of America*, 82(20):6723{6726, 1985.
- [80]. R. G. Parr, L. Von Szentpaly, and S. B. Liu. Electrophilicity index. *Journal of the American Chemical Society*, 121(9):1922{1924, 1999.
- [81]. J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz. Density-Functional Theory for Fractional Particle Number - Derivative Discontinuities of the Energy. *Physical Review Letters*, 49(23):1691{1694, 1982.

## Chapitre.II: Méthodologie QSAR

- [82]. Y. K. Zhang and W. T. Yang. Perspective on "Density-functional theory for fractional particle number: derivative discontinuities of the energy" Perdew JP, Parr RG, Levy M, Balduz JL Jr. *Theoretical Chemistry Accounts*, 103(3-4):346-348, 2000
- [83]. W. T. Yang, Y. K. Zhang, and P. W. Ayers. Degenerate ground states and a fractional number of electrons in density and reduced density matrix functional theory. *Physical Review Letters*, 84(22):5172-5175, 2000.
- [84]. P. W. Ayers and M. Levy. Perspective on "Density functional approach to the frontier-electron theory of chemical reactivity" - Parr RG, Yang W (1984) *J Am Chem Soc* 106: 4049-4050. *Theoretical Chemistry Accounts*, 103(3-4):353-360, 2000
- [85]. P. W. Ayers and R. G. Parr. Variational principles for describing chemical reactions: The Fukui function and chemical hardness revisited. *Journal of the American Chemical Society*, 122(9):2010-2018, 2000
- [86]. C. Morell, A. Grand, and A. Toro-Labbe. New dual descriptor for chemical reactivity. *Journal of Physical Chemistry A*, 109(1):205-212, 2005.
- [87]. W. Yang and W. J. Mortier. The Use of Global and Local Molecular- Parameters for the Analysis of the Gas-Phase Basicity of Amines. *Journal of the American Chemical Society*, 108(19):5708-5711, 1986
- [88]. C. T. Lee, W. T. Yang, and R. G. Parr. Local Softness and Chemical- Reactivity in the Molecules CO, SCN- and H<sub>2</sub>CO. *Theochem-Journal of Molecular Structure*, 40:305-313, 1988.
- [89]. Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br J Pharmacol*. 2007; 152:9-20.
- [90]. Van de Waterbeemd H, Gifford E. ADMET in silicomodelling: towards prediction paradise? *Nat Rev Drug Discov*. 2003;2:192-204
- [91]. Kasnanen H, Myllymaki MJ, Minkkilä A et al. 3-Heterocycle-phenyl Nalkylcarbamates as FAAH inhibitors: design, synthesis and 3D-QSAR studies. *ChemMedChem*. 5:213-31.
- [92]. Yamashita F, Hashida M. In silico approaches for predicting ADME properties of drugs. *Drug Metab Pharmacokinet*. 2004;19:327-38
- [93]. Abdi, Hervé, and Lynne J. Williams. 2010. "Principal Component Analysis." *John Wiley and Sons, Inc. WIREsComp Stat* 2: 433-59.
- [94]. Husson, Francois, Sebastien Le, and Jérôme Pagès. 2017. *Exploratory Multivariate Analysis by Example Using R*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <http://factominer.free.fr/bookV2/index.html>.
- [95]. Jolliffe, I.T. 2002. *Principal Component Analysis*. 2nd ed. New York: Springer-Verlag. <https://goo.gl/SB86SR>.
- [96]. Kaiser, Henry F. 1961. "A Note on Guttman's Lower Bound for the Number of Common Factors." *British Journal of Statistical Psychology* 14: 1-2.
- [97]. Peres-Neto, Pedro R., Donald A. Jackson, and Keith M. Somers. 2005. "How Many Principal Components? Stopping Rules for Determining the Number of Non-Trivial Axes Revisited." *British Journal of Statistical Psychology* 49: 974-97
- [98]. Douglas Montgomery, Elizabeth Peck, and Geoffrey Vinig, "Introduction to linear Regression Analysis". Book, ISBN: 978-0-470-54281-1, 672 pages, April 2012.

## **Chapitre.II: Méthodologie QSAR**

- [99]. Topliss, J.G. and Edwards, R.P. (1979) *Journal of Medicinal Chemistry*, 22, 1238–1244.
- [100]. Hiromasa, K., Masamoto, A., and Kimito, F. (2008) *Journal of Chemical Information and Modeling*, 48, 534–541.
- [101]. Guha, R., Dutta, D., Jurs, P.C., and Chen, T. (2006) *Journal of Chemical Information and Modeling*, 46, 1836–1847
- [102]. Randi\_c, M. (2001) *Journal of Chemical Information and Computer Sciences*, 41, 602–606.
- [103]. Dutta, D., Guha, R., Wild, D., and Chen, T. (2007) *Journal of Chemical Information and Modeling*, 47, 989–997.
- [104]. An Introduction to Partial Least Squares Regression. <http://www.ats.ucla.edu/stat/sas/library/pls.pdf>
- [105]. Wold, S., Sjostrom, M., and Eriksson, L. (2001) *Chemometrics and Intelligent Laboratory Systems*, 58, 109–130.
- [106]. Yee, L. C., & Wei, Y. C. (2012). Current modeling methods used in QSAR/QSPR. *Statistical Modelling of Molecular Descriptors in QSAR/QSPR, Volume 2*, 1-31.
- [107]. Vilar, S., Chakrabarti, M., and Costanzi, S. (2010) *Journal of Molecular Graphics and Modelling*, 28, 899–903.
- [108]. P\_erez-Garrido, A., Helguera, A.M., Rodr\_iguez, F.G., and Cordeiro, M.N.D.S. (2010) *Dental Materials*, 26, 397–415.
- [109]. Prado-Prado, F.J., Garc\_ia-Mera, X., and Gonz\_alez-D\_iaz, H. (2010) *Bioorganic and Medicinal Chemistry*, 18, 2225–2231.
- [110]. Castillo-Garit, J.A., Vega, M.C., Rolon, M., Marrero-Ponce, Y., Kouznetsov, V.V., Torres, D.F.A., G\_omez-Barrio, A., Bello, A.A., Montero, A., Torrens, F., and P\_erez-Gim\_enez, F. (2010) *European Journal of Pharmaceutical Sciences*, 39, 30–36.
- [111]. Wang, G., Li, Y., Liu, X., and Wang, Y. (2009) *the QSAR and Combinatorial Science*, 28, 1418–1431.
- [112]. Castillo-Gonz\_alez, D., Cabrera-P\_erez, M.A., P\_erez-Gonz\_alez, M., Helguera, A.M., and Dur\_an-Mart\_inez, A. (2009) *European Journal of Medicinal Chemistry*, 44, 4826–4840.
- [113]. Rivera-Barroto, O.M., Marrero-Ponce, Y. Meneses-Marcel, A., Escario, J.A., Barrio, A.G., Ar\_an, V.J., Alho, M.A.M., Pereira, D.M., Nogal, J.J., Torrens, F., Ibarra-Velarde, F., Montenegro, Y.V., Huesca-Guill\_en, A., Rivera, N., and Vogel, C. (2009) *The QSAR and Combinatorial Science*, 28, 9–26.
- [114]. Sa\_iz-Urra, L., P\_erez-Castillo, Y., Gonz\_alez, M.P., Ruiz, R.M., Cordeiro, M.N.D.S., Rodr\_iguez-Borges, J.E., and Garc\_ia-Mera X. (2009) *The QSAR and Combinatorial Science*, 28, 98–110
- [115]. Drew, P. J., & Monson, J. R. (2000). Artificial neural networks. *Surgery*, 127(1), 3–11.
- [116]. <http://www.moreere.eu/index.php/spip.php?article19>
- [117]. Ritchie, M.D., White, B.C., Parker, J.S., Hahn, L.W., and Moore, J.H. (2003) *BMC Bioinformatics*, 4, 28.
- [118]. Hornik, K., Stinchcombe, M., and White, H. (1989) *Neural Networks*, 2, 359–366.
- [119]. Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986) *Nature*, 323, 533–536.

## Chapitre.II: Méthodologie QSAR

- [120]. Mundie, D.B. and Massengill, L.W. (1991) IEEE Transactions on Neural Networks, 2, 168–170.
- [121]. [https://fr.wikipedia.org/wiki/Apprentissage\\_supervis%C3%A9](https://fr.wikipedia.org/wiki/Apprentissage_supervis%C3%A9)
- [122]. PayamEmamiKhoonsari and AhmadRezaMotie, "A Comparison of Efficiency and Robustness of ID3 and C4.5 Algorithms Using Dynamic Test and Training Data Sets" International Journal of Machine Learning and Computing, Vol. No.5, October 2012.
- [123]. Caruana R, Niculescu-Mizil A. "An empirical comparison of supervised learning algorithms". Proceedings of the Twenty-Third International Conference; New York ACM; pp. 161–168. 2003. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, "An Introduction to Statistical Learning". New York: Springer.p. 315. ISBN 978-1-4614-7137-0, 2015.
- [124]. Letham, Ben; Rudin, Cynthia; McCormick, Tyler; Madigan, David "Interpretable Classifiers Using Rules And Bayesian Analysis: Building A Better Stroke Prediction Model". Annals of Applied Statistics.9: pp1350–1371. 2015
- [125]. Trevor Hastie, Robert Tibshirani, Jerome Friedman (2009) *The Elements of Statistical Learning*, Springer-Verlag
- [126]. Vapnik, V. (1995) *the Nature of Statistical Learning Theory*, Springer, New York, London.
- [127]. Bennett, K.P. and Campbell, C. (2000) SIGKDD Explorations Newsletter, 2, 1–13.
- [128]. Kim, H.S. and Sohn, S.Y. (2010) European Journal of Operational Research, 201, 838–846.
- [129]. Zuluaga, M.A., Magnin, I.E., HernandezHoyos, M., Delgado Leyton, E.J.F., Lozano, F., and Orkisz, M. (2010) International Journal of Computer Assisted Radiology and Surgery, vol 6, 163–174.
- [130]. Fernandez, M., Ahmad, S., and Sarai, A. (2010) Journal of Chemical Information and Modeling, 50, 1179–1188. Shen, J., Cheng, F., Xu, Y., Li, W., and Tang, Y. (2010) Journal of Chemical Information and Modeling, 50, 1034–1041.
- [131]. Conforti, D. and Guido, R. (2010) Computers and Operations Research, 37, 1389–1394.
- [132]. Regulation (EC) No 1907/2006 of the European Parliament and of the Council (18/12/2006) concerning REACH; [http://eur-lex.europa.eu/LexUriServ/site/en/oj/2006/l\\_396/l\\_39620061230en00010849.pdf](http://eur-lex.europa.eu/LexUriServ/site/en/oj/2006/l_396/l_39620061230en00010849.pdf) (accessed January 22, 2007).
- [133]. <http://ecb.jrc.it/qsar/> (accessed January 22, 2007).
- [134]. J. Jaworska, M. Comber, C. Auer, C. J. van Leeuwen, Environ. Health Perspect.2003, 111, 1358 – 1360.
- [135]. <http://www.oecd.org/dataoecd/33/37/37849783.pdf> (accessed January 22, 2007).
- [136]. Turner, L., Choplin, F., Dugard, P., Hermens, J., Jaekh, R., Marsmann, M., & Roberts, D. (1987). Structure-activity relationships in toxicology and ecotoxicology: an assessment. *Toxicology in vitro*, 1(3), 143-171.
- [137]. Hermens, J. L. (1989). Quantitative structure-activity relationships of environmental pollutants. In *Reactions and processes* (pp. 111-162). Springer, Berlin, Heidelberg.

## **Chapitre.II: Méthodologie QSAR**

- [138]. Blum, D. J., & Speece, R. E. (1990). Determining chemical toxicity to aquatic species. *Environmental Science & Technology*, 24(3), 284-293.
- [139]. Topliss, J. G., & Edwards, R. P. (1979). Chance factors in studies of quantitative structure-activity relationships. *Journal of Medicinal Chemistry*, 22(10), 1238-1244.
- [140]. Wold, S., & Dunn III, W. J. (1983). Multivariate quantitative structure-activity relationships (QSAR): conditions for their applicability. *Journal of Chemical Information and Computer Sciences*, 23(1), 6-13.
- [141]. Wold, S. (1994). PLS for Multivariate Linear Modelling, QSAR: Chemometric methods in Molecular Design. *Methods and Principles in Medicinal Chemistry, Van de Waterbeemd, H.(Ed.). Verlag-Chemie, Weinheim, Germany.*
- [142]. Cramer III, R. D., Bunce, J. D., Patterson, D. E., & Frank, I. E. (1988). Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quantitative Structure-Activity Relationships*, 7(1), 18-25.
- [143]. Clark, M., & Cramer III, R. D. (1993). The probability of chance correlation using partial least squares (PLS). *Quantitative Structure-Activity Relationships*, 12(2), 137-145.
- [144]. Eriksson, L., Verhaar, H. J., & Hermens, J. L. (1994). Multivariate characterization and modeling of the chemical reactivity of epoxides. *Environmental toxicology and chemistry*, 13(5), 683-691.
- [145]. Sjöström, M., & Eriksson, L. (1995). Applications of statistical experimental design and PLS modeling in QSAR. *Chemometric methods in molecular design*, 2, 63-90.
- [146]. CRAMER CJ. Essentials of computational chemistry: theories and models. Wiley, 2e edition, 2004
- [147]. DEVILLERS J, MARCHAND-GENESTE N, CARPY A, PORCHER JM. SAR and QSAR modeling of endocrine disruptors. SAR and QSAR Environmental research 2006, 17 : 393-412
- [148]. GAO H, KATZENELLENBOGEN JA, GARG R, HANSCH C. Comparative QSAR analysis of estrogen receptor ligands. Chem Rev 1999, 99 : 723-744
- [149]. JACOBS MN. In silico tools to aid risk assessment of endocrine disrupting chemicals. Toxicology 2004, 205 : 43-53
- [150]. KARELSON M. Molecular Descriptors in Qsar/Qspr. John Wiley & Sons Inc, New York 2000
- [151]. LEACH AR. Molecular modelling: principles and applications. Prentice Hall, 2e edition, 2001
- [152]. LI J, GRAMATICA P. The importance of molecular structures, endpoints' values and predictivity parameters in QSAR research: QSAR analysis of a series of estrogen receptor binders. Mol Divers 2010 14 : 687-696
- [153]. LIU H, PAPA E, GRAMATICA P. QSAR Prediction of estrogen activity for a large set of diverse chemicals under the guidance of OECD principles. Chem Res Toxicol 2006, 19 : 1540-1548

## **Chapitre.II: Méthodologie QSAR**

- [154]. LIU H, PAPA E, WALKER JD, GRAMATICA P. In silico screening of estrogen-like chemicals based on different nonlinear classification models. *Journal of Molecular Graphics and Modelling* 2007, 26 : 135-144
- [155]. RONCAGLIONI A, BENFENATI E. In silico-aided prediction of biological properties of chemicals: oestrogen receptor-mediated effects. *ChemSoc Rev* 2008, 37 : 441-450
- [156]. SHI LM, FANG H, TONG W, WU J, PERKINS R, et coll. QSAR Models using a large diverse set of estrogens *J ChemInfComputSci* 2001, 41 : 186-195
- [157]. TODESCHINI R, CONSONNI V. Handbook of molecular descriptors. In : MANNHOLD R, KUBINYI H, TIMMERMAN H (eds). WILEY VCH, Weinheim, 2000
- [158]. ZHU BT, HAN GZ, SHIM JY, WEN Y, JIANG XR. Quantitative structure-activity relationship of various endogenous estrogen metabolites for human estrogen receptor  $\alpha$  and  $\beta$  subtypes: insights into the structural determinants favoring a differential subtype binding. *Endocrinology* 147 : 4132-4150

***Chapitre III  
Structures et propriétés  
électroniques de dérivés  
HEPT***

## Chapitre.III : Structures et propriétés électroniques de dérivés HEPT

### I. Introduction

Dans ce chapitre, nous avons englobé quelques résultats de calcul effectués dans notre travail mais nous nous sommes intéressés, dans un premier temps à la présentation de tous les résultats de calcul de chimie quantique concernant la structure et quelques propriétés électroniques des dérivés de HEPT.

### II. Détail de calculs

Afin de réaliser notre travail, il a été nécessaire de définir une méthode de calcul ainsi que des bases atomiques de manière à développer la fonction d'onde,

Nous avons utilisé :

- ✓ La méthode DFT B3LYP/6-311g + (d, p) [1-5] pour les calculs quantiques,
- ✓ Le programme Gaussview 5,0 pour la construction et la visualisation des molécules,
- ✓ Le programme GAUSSIAN 09W [6] pour l'optimisation des géométries,
- ✓ Les géométries des molécules neutres ont été maintenues constantes pour les systèmes cationiques et anioniques pour le calcul des indices globaux,
- ✓ L'analyse de population naturelle (NPA) [7] pour le calcul des indices locaux de réactivité.

### III. Résultats

#### III.1 Aspect structural des dérivés de HEPT

Pour la molécule HEPT, nous avons adopté la numérotation donnée dans la figure suivante, pour caractériser ses différents atomes.

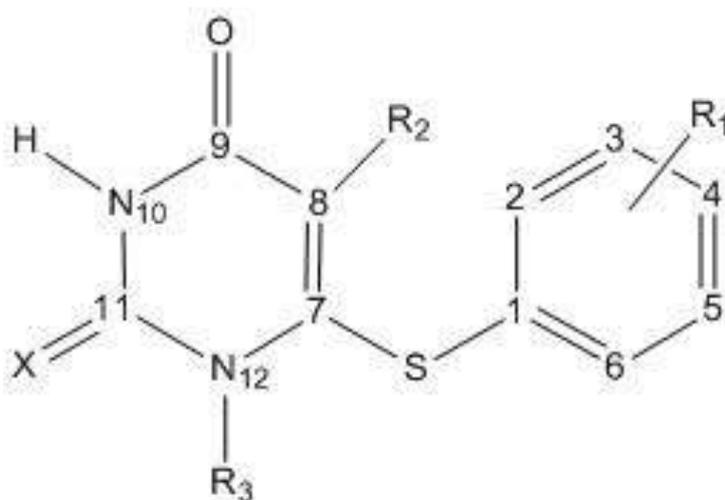
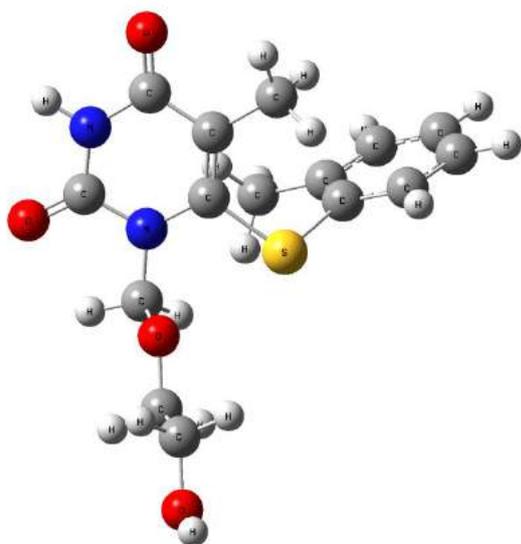


Figure.1 : Numérotation des différents atomes

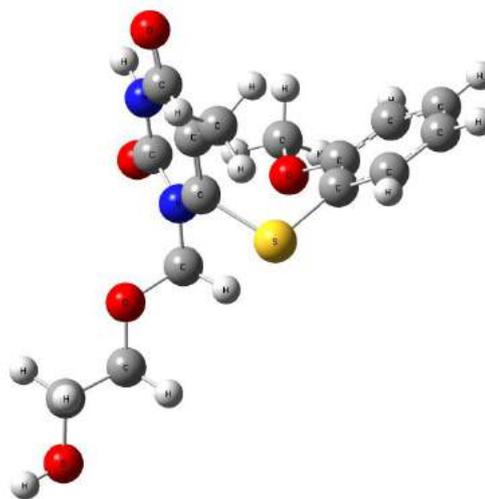
## Chapitre.III : Structures et propriétés électroniques de dérivés HEPT

### *a. Position du radical $R_1$ sur le thiophényl*

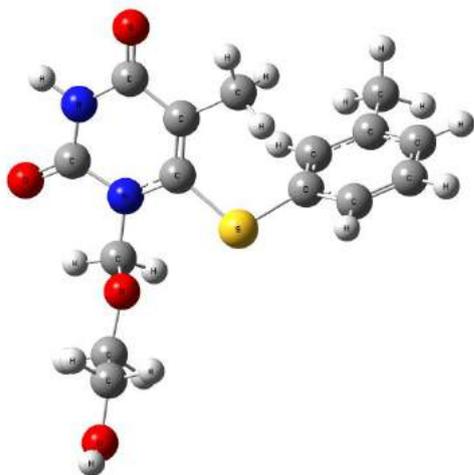
L'optimisation de géométries des 4 dérivés de HEPT-O ayant un méthyle en  $R_2$  et  $\text{CH}_2\text{OCH}_2\text{CH}_2\text{OH}$  en  $R_3$  et un  $R_1$  (2-Me, 3-Me, 2-OMe, 3-OMe) a donné les structures décrites ci-dessous.



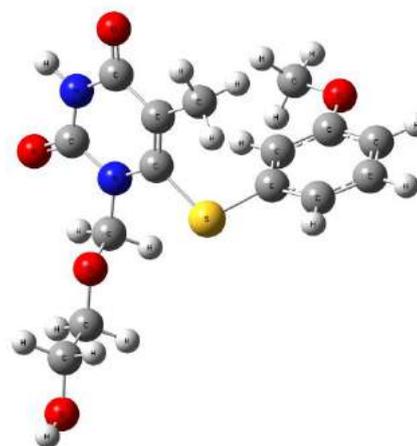
HEPT-O (2-Me)



HEPT-O (2-OMe)



HEPT-O (3-Me)



HEPT-O (3-OMe)

Figure.2 : Les géométries optimisées au niveau DFT B3LYP/6-311g + (d, p)

### Chapitre.III : Structures et propriétés électroniques de dérivés HEPT

Afin de déterminer dans quelle mesure la position du radical R<sub>1</sub> sur le thiophényl peut apporter des changements au niveau de la structure de certaines dérivés de HEPT-O, nous proposons d'analyser les résultats du tableau suivant, qui contient quelque uns de leurs paramètres géométriques optimisés au niveau DFT B3LYP/6-311g + (d, p). Ces paramètres permettent de comparer les orientations relatives des deux cycles à six chaînons.

Tableau.1 : Effets de la position du radical R<sub>1</sub> sur certaines caractéristiques structurales de quelques dérivés de HEPT-O.

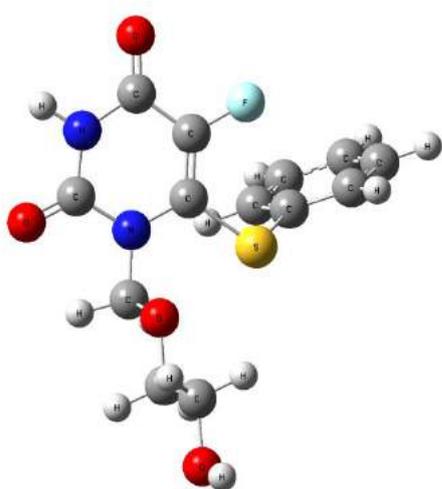
N°	R <sub>1</sub>	CE <sub>50</sub> (μM)	C <sub>8</sub> -C <sub>2</sub> (Å)	C <sub>7</sub> -S-C <sub>1</sub> -C <sub>2</sub> (°)	C <sub>8</sub> -C <sub>7</sub> -S-C <sub>1</sub> (°)
4	3-Me	2,570	3,653	-50,509	-43,500
3	2-OMe	19,054	4,313	-54,888	-60,521
14	3-OMe	21,877	3,533	-42,976	-58,896
1	2-Me	70,794	3,873	-62,140	-52,965

De ces résultats, on conclue que les positions relatives des deux cycles sont gouvernées par les interactions et les encombrements stériques entre le méthyle en C<sub>8</sub> et le groupement R<sub>1</sub> du thiophényl. Des valeurs des paramètres géométriques, il est clair que, en position ortho, les groupements méthyle ou méthoxy ont des interactions répulsives avec le méthyle en C<sub>8</sub> plus importantes qu'en position méta. Il est à noter qu'en position méta le méthoxy est moins répulsive avec le méthyle en C<sub>8</sub> que le méthyle. Dans le tableau.1, les CE<sub>50</sub> sont dressées à côté de paramètres géométriques calculés des quatre molécules. La comparaison des évolutions de ces paramètres géométriques avec celle des CE<sub>50</sub> ne permet pas d'identifier de corrélations entre les orientations relatives des deux cycles et l'activité biologique des molécules étudiées.

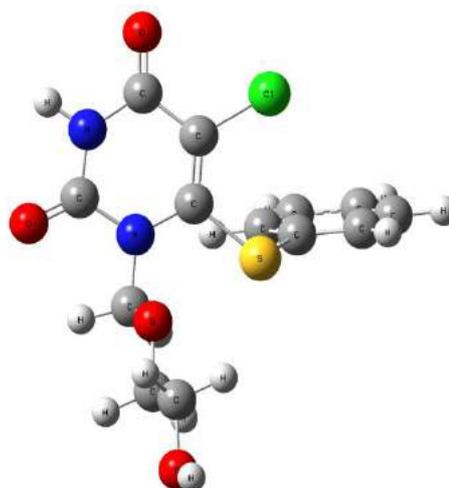
## Chapitre.III : Structures et propriétés électroniques de dérivés HEPT

### *b. Nature du radical R<sub>2</sub>*

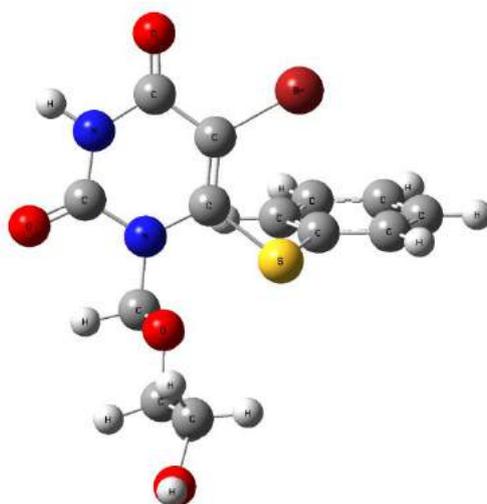
Dans le but de voir l'effet de la nature du groupement R<sub>2</sub> sur la géométrie (longueur de liaisons, paramètres angulaires) et afin de déterminer dans quelle mesure la bromation, la chloration et la fluoration peuvent apporter des changements au niveau de la structure, nous allons analyser les résultats d'une série de HEPT-O ayant un H en R<sub>1</sub> et un CH<sub>2</sub>OCH<sub>2</sub>CH<sub>2</sub>OH en R<sub>3</sub>.



HEPT-O (R<sub>2</sub>=F)



HEPT-O (R<sub>2</sub>=Cl)



HEPT-O (R<sub>2</sub>=Br)

Figure.3 : Les géométries optimisées au niveau DFT

## Chapitre.III : Structures et propriétés électroniques de dérivés

### HEPT

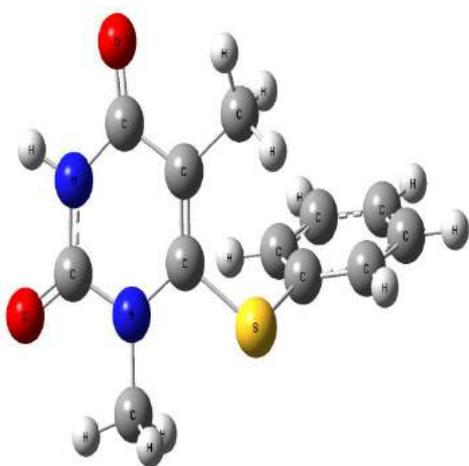
Tableau.2 : Effets de la nature du radical R<sub>2</sub> sur certaines caractéristiques structurales de quelques dérivés de HEPT-O

N°	R <sub>2</sub>	CE <sub>50</sub> (µM)	R <sub>2</sub> -C <sub>2</sub> (Å)	C <sub>8</sub> -C <sub>2</sub> (Å)	R <sub>2</sub> -C <sub>8</sub> -C <sub>7</sub> -S(°)	H-C <sub>2</sub> -C <sub>1</sub> -S(°)
102	F	100,000	3,584	3,641	-0,997	3,863
103	Cl	30,200	3,826	3,617	-1,728	4,686
104	Br	19,953	3,924	3,602	-2,131	4,961

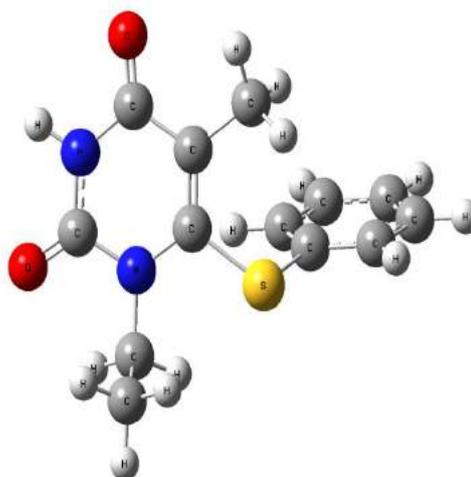
Les résultats du tableau.2 montrent que l'introduction d'un atome de Fluore, Chlore ou un Brome agit sur les caractéristiques structurales et plus particulièrement sur la position des deux phényles. Les liaisons R<sub>2</sub>-C<sub>2</sub> ont des valeurs allant de 3,584 Å à 3,924 Å, ces valeurs varient dans l'ordre F < Cl < Br, le sens de cette variation est le même que l'effet inductif de ces atomes. Quant aux angles dièdres, les angles R<sub>2</sub>-C<sub>8</sub>-C<sub>7</sub>-S diminuent lorsqu'on passe de Fluor, Chlore et Brome, ce qui est de même pour le deuxième angle.

#### *c. Longueur de la chaîne carbonée du R<sub>3</sub> :*

La modification de la queue hydrophobe de certains HEPT-O ayant un méthyle en R<sub>2</sub> et un H en R<sub>1</sub> tout en variant le radical R<sub>3</sub>, en remplaçant le méthyle par un éthyle ou un butyle donne les structures suivantes :

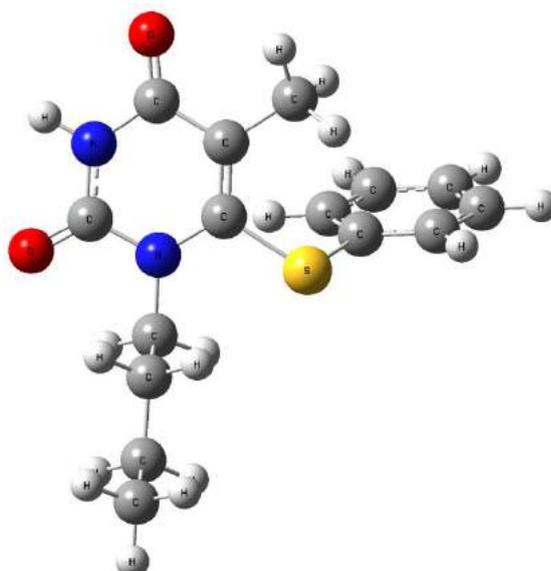


HEPT-O (R<sub>3</sub>=Me)



HEPT-O (R<sub>3</sub>=Et)

### Chapitre.III : Structures et propriétés électroniques de dérivés HEPT



HEPT-O (R<sub>3</sub>=Bu)

Figure.4 : Les géométries optimisées

Tableau.3 : Effets de la longueur de la chaîne carbonée du R<sub>3</sub> sur certaines caractéristiques structurales de quelques dérivés de HEPT-O

N°	R <sub>3</sub>	CE <sub>50</sub> (μM)	C <sub>7</sub> -S-C <sub>1</sub> (°)	R <sub>3</sub> -N <sub>12</sub> -C <sub>7</sub> -S(°)
107	Me	151,356	103,238	0,770
65	Et	2,188	102,840	0,683
66	Bu	1,202	102,900	1,473

En analysant les résultats reportés dans ce tableau, nous pouvons remarquer clairement que la variation des valeurs des angles de liaisons C<sub>7</sub>-S-C<sub>1</sub> n'a aucune relation avec la variation du nombre de carbone du groupement R<sub>3</sub> (la longueur de la chaîne carbonée), il est de même pour l'angle de dièdre R<sub>3</sub>-N<sub>12</sub>-C<sub>8</sub>-S.

Cependant, la longueur de la chaîne carbonée paraît agir sur l'activité biologique des molécules qui est clairement visible avec la valeur de la concentration efficace (CE<sub>50</sub>) alors qu'elle n'apporte pas d'effet sur les caractéristiques structurales de ces derniers.

## Chapitre.III : Structures et propriétés électroniques de dérivés HEPT

Après avoir déterminé et analysé les caractéristiques structurales de quelques dérivés de HEPT, nous nous intéressons, dans un second temps, à la détermination de leurs propriétés électroniques.

### III.2. Etude des descripteurs en fonction de la structure des dérivés

a. Le potentiel d'ionisation et l'affinité électronique des dérivés de HEPT-O en fonction de la position et le type du radical  $R_1$

Tableau.4 : Le potentiel d'ionisation et l'affinité électronique des dérivés de HEPT-O en fonction de la position et le type du radical  $R_1$

N°	$R_1$	$R_2$	$R_3$	PI (e.V)	AE (e.V)
1	2-Me	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,646	0,707
2	2-NO <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	8,653	3,945
3	2-OMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	8,027	2,857
4	3-Me	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,836	0,19
5	3-Et	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,962	0,299
7	3-CF <sub>3</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	8,054	0,490
8	3-F	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	8,201	0,457
12	3-NO <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	8,163	1,034
13	3-OH	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,994	0,305
14	3-OMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,836	0,816
15	3,5-Me <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,885	0,267
16	3,5-Cl <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	8,109	1,088
18	3-COOMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	8,168	0,702
19	3-COMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	8,212	0,849
34	4-Me	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,910	0,269
84	4-F	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	8,141	0,395
85	4-Cl	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	8,095	0,493
90	4-COMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,828	1,371

Le tableau.4 présente le potentiel d'ionisation et l'affinité électronique des molécules HEPT-O ayant un méthyle en  $R_2$  et CH<sub>2</sub>OCH<sub>2</sub>CH<sub>2</sub>OH en  $R_3$ . On note que le changement de position du méthyle sur le thiophényl donne une faible variation du PI suivant l'ordre ortho < méta < para. Ce pendant, cette position influe considérablement sur AE puisque la position ortho lui attribue une assez forte valeur alors qu'elle attribue deux faibles et équivalentes valeurs pour les positions méta et para.

### Chapitre.III : Structures et propriétés électroniques de dérivés HEPT

On note qu'il n'y a pas d'évolution des PI et AE de ces molécules, elles ne présentent pas des corrélations linéaires avec les propriétés électroniques des radicaux liés au thiophényl. Il n'y a pas de corrélation évidente avec leurs positions sur les thiophényl.

Dans le tableau ci-après, nous reportons quelques valeurs de potentiel d'ionisation et affinité électronique d'une série de dérivés de HEPT-O en fonction du groupement R<sub>2</sub>.

Tableau.5 : Le potentiel d'ionisation et l'affinité électronique des dérivés de HEPT-O en fonction du radical R<sub>2</sub>

N°	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	PI (e.V)	AE (e.V)
28	H	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	8,041	0,395
93	H	COOMe	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	8,166	0,498
40	H	CH <sub>2</sub> Ph	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	8,008	0,528
104	H	Br	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	8,046	0,544
103	H	Cl	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	8,198	0,585
94	H	CONHPh	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,793	0,601
100	H	COPh	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,684	0,669
41	H	CH=CPh <sub>2</sub>	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,148	0,770
39	H	CH=CHPh	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,202	1,292
101	H	C≡CMe	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	6,498	1,431
97	H	C≡C-Ph	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	6,163	1,663
21	H	CH <sub>2</sub> CH=CH <sub>2</sub>	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	5,524	1,687

Le tableau.5 présente les PI et AE des molécules HEPT-O ayant CH<sub>2</sub>OCH<sub>2</sub>CH<sub>2</sub>OH en R<sub>3</sub> et non substitué en thiophényl. D'après les résultats de ce tableau, on remarque que les valeurs du potentiel d'ionisation varient dans l'intervalle [5,52 - 8,20] e.V. D'autre part, on peut voir clairement que les dérivés avec un potentiel d'ionisation élevé ont les faibles valeurs de l'affinité électronique. On remarque que les radicaux avec une double ou triple liaison affectent à leurs molécules des valeurs d'affinité les plus élevées.

### Chapitre.III : Structures et propriétés électroniques de dérivés

#### HEPT

Tableau.6 : Le potentiel d'ionisation et l'affinité électronique des dérivés de HEPT-S en fonction du radical R<sub>3</sub>.

N°	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	PI (e.V)	AE (e.V)
22	H	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	7,836	0,607
52	H	Et	CH <sub>2</sub> OCH <sub>2</sub> Me	7,790	0,569
57	H	Et	CH <sub>2</sub> OCH <sub>2</sub> Ph	7,739	0,612
68	H	Et	CH <sub>2</sub> O-i-Pr	7,755	0,612
69	H	Et	CH <sub>2</sub> O-c-Hex	7,722	0,544
70	H	Et	CH <sub>2</sub> OCH <sub>2</sub> -c-Hex	7,768	0,596
71	H	Et	CH <sub>2</sub> OCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-Me)	7,630	0,623
72	H	Et	CH <sub>2</sub> OCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-Cl)	7,747	0,732
73	H	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> Ph	7,657	0,571

Le tableau.6 présente le PI et AE de dérivés HEPT-S ayant le thiophényl non substitué et un radical éthyle lié à l'atome de carbone. Ces dérivés varient par leurs radical R<sub>3</sub>. On note que la longueur de la chaîne et les propriétés électroniques de ce radical ont un faible effet sur le PI et AE de HEPT-S. On remarque aussi que les valeurs du PI sont voisine du 7,700 e.V, la variation du groupement R<sub>3</sub> influe légèrement sur ces valeurs. Aussi, pour les affinités électroniques, elles varient entre 0,571 e.V et 0,732 e.V. On peut conclure que le groupement R<sub>3</sub> participe très faiblement aux orbitales HOMO et LUMO.

### Chapitre.III : Structures et propriétés électroniques de dérivés HEPT

*b. la mollesse, l'indice électrophile, la dureté, l'électronégativité et le potentiel chimique en fonction de R<sub>1</sub>*

Dans le tableau qui suit, on donne la mollesse(S), l'indice électrophile ( $\omega$ ), la dureté( $\eta$ ), l'électronégativité ( $\chi$ ) et le potentiel chimique ( $\mu$ ) de quelques dérivés de HEPT-O, avec R<sub>2</sub> un méthyle et R<sub>3</sub>=CH<sub>2</sub>OCH<sub>2</sub>CH<sub>2</sub>OH. Ainsi, on peut analyser la variation de ces indices avec la variation du nombre, la nature et la position sur le thiophényl du radical R<sub>1</sub>.

Tableau.7 : la mollesse(S), l'indice électrophile ( $\omega$ ), la dureté( $\eta$ ), l'électronégativité ( $\chi$ ) et le potentiel chimique ( $\mu$ ) de quelques dérivés de HEPT-O en fonction de R<sub>1</sub>.

N°	R <sub>1</sub>	S (e.V)	$\omega$ (e.V)	$\eta$ (e.V)	$\chi$ (e.V)	$\mu$ (e.V)
4	3-Me	0,131	2,107	3,823	4,013	-4,013
15	3,5-Me <sub>2</sub>	0,131	2,181	3,809	4,076	-4,076
5	3-Et	0,131	2,227	3,831	4,130	-4,130
13	3-OH	0,130	2,239	3,845	4,150	-4,150
7	3-CF <sub>3</sub>	0,132	2,413	3,782	4,272	-4,272
8	3-F	0,129	2,420	3,872	4,329	-4,329
18	3-COOMe	0,134	2,635	3,733	4,435	-4,435
19	3-COMe	0,136	2,788	3,682	4,530	-4,530
12	3-NO <sub>2</sub>	0,140	2,966	3,565	4,598	-4,598
16	3,5-Cl <sub>2</sub>	0,142	3,012	3,510	4,598	-4,598
3	2-OMe	0,193	5,728	2,585	5,442	-5,442
2	2-NO <sub>2</sub>	<u>0,212</u>	<u>8,429</u>	<u>2,354</u>	<u>6,299</u>	<u>-6,299</u>

On peut conclure facilement que la position du groupement -NO<sub>2</sub> en 2 ou en 3 influe sur la valeur de certains indices. On note que les meilleurs pouvoirs électrophiles ( $\omega$ ) sont pour les molécules ayant un radical en position ortho, ce qui est le même pour d'autres indices comme la mollesse globale. Un bon électrophile est toujours associé à une électronégativité élevée et une dureté basse ce qui est en bon accord avec nos résultats.

### Chapitre.III : Structures et propriétés électroniques de dérivés HEPT

c. Les valeurs des charges nettes en fonction du radical  $R_1$

Dans le but de calculer les indices de Fukui des dérivés de HEPT-O, il a été nécessaire de présenter la distribution de charge sur les différents sites  $C_7$ ,  $C_8$ ,  $S$ ,  $O$  et le  $N_{12}$ . Le choix de ces atomes était sur la base que ces différents sites sont en relation avec la position des différents groupements ( $R_1$ ,  $R_2$  et  $R_3$ ) Les calculs ont été effectués avec l'analyse de population naturelle (NPA).

Dans le tableau ci-dessous, nous avons donné les valeurs charges nettes des sites  $C_7$ ,  $C_8$ ,  $S$ ,  $O$  et le  $N_{12}$  pour une série de dérivés de HEPT-O en fonction du groupement  $R_1$ .

Tableau. 8 : Les charges nettes de quelques dérivés de HEPT-O en fonction du groupement  $R_1$ .

N°	$R_1$	$C_7$	$C_8$	$S$	$O$	$N_{12}$
1	2-Me	0,828	0,652	0,267	-0,612	-0,625
2	2-NO <sub>2</sub>	0,828	0,645	0,286	-0,523	-0,606
3	2-OMe	0,820	0,648	0,154	-0,600	-0,624
4	3-Me	0,833	0,649	0,302	-0,611	-0,631
5	3-Et	0,827	0,653	0,306	-0,601	-0,629
7	3-CF <sub>3</sub>	0,834	0,649	0,310	-0,607	-0,628
84	4-F	0,828	0,653	0,307	-0,599	-0,629
85	4-Cl	0,828	0,654	0,311	-0,598	-0,629
90	4-COMe	0,828	0,653	0,338	-0,597	-0,628

L'analyse des valeurs prévoient que les sites  $O$  et  $N_{12}$  sont riches en électrons (charge nette négative). Cependant, les sites  $C_7$ ,  $C_8$  et  $S$  sont pauvres en électrons (Charge positive).

On remarque aussi que la nature et la position du radical  $R_1$  n'influence pas beaucoup sur les valeurs des charges comme celle du  $C_7$  ou  $C_8$ . Aussi, pour les charges négatives, celles du site  $O$  ont des valeurs allant de -0,523 à -0,612. Cependant, le site du  $S$  est le plus déficitaire en électrons.

### Chapitre.III : Structures et propriétés électroniques de dérivés HEPT

d. Les valeurs des indices de Fukui en fonction du radical  $R_1$

La fonction de Fukui  $f_k^+$  représente la réaction de la densité électronique d'une molécule au niveau d'un atome k à l'extraction d'un électron. Elle est essentiellement utilisée comme une mesure de réactivité lorsque la molécule est attaquée par un réactif nucléophile. Et la  $f_k^-$  est utilisée comme une mesure de réactivité lorsque la molécule est attaquée par un réactif électrophile.

La fonction  $\Delta f_k$  permet d'obtenir de la réaction de la densité électronique: à l'augmentation et à la diminution d'un électron.

Tableau. 9 : Les indices de Fukui de quelques dérivés de HEPT-O en fonction du  $R_1$ .

N°	$R_1$	$\Delta f_{C7}$	$\Delta f_{C8}$	$\Delta f_S$	$\Delta f_O$	$\Delta f_{N12}$
82	2-Cl	0,839	0,682	0,140	-0,727	-0,639
1	2-Me	0,817	0,643	0,304	-0,581	-0,623
2	2-NO <sub>2</sub>	0,838	0,654	0,107	-0,535	-0,598
3	2-OMe	0,830	0,676	-0,133	-0,663	-0,621
19	3-COMe	0,839	0,666	0,089	-0,689	-0,633
8	3-F	0,841	0,657	0,097	-0,722	-0,637
4	3-Me	0,844	0,658	0,206	-0,718	-0,641
12	3-NO <sub>2</sub>	0,835	0,690	0,120	-0,691	-0,624
14	3-OMe	0,831	0,646	0,041	-0,650	-0,520
85	4-Cl	0,840	0,655	0,104	-0,688	-0,639
90	4-COMe	0,835	0,674	0,118	-0,683	-0,632
84	4-F	0,841	0,654	0,095	-0,697	-0,640

La variation de la position et la nature du radical  $R_1$  sur le thiophényl pour une série de HEPT-O ayant le radical  $R_2 = \text{Me}$  et  $R_3 = \text{CH}_2\text{OCH}_2\text{CH}_2\text{OH}$  agissent sur les valeurs des indices de Fukui calculés pour les sites C<sub>7</sub>, C<sub>8</sub>, S, O et le N<sub>12</sub>. La variation de l'indice  $\Delta f_{C7}$  est légèrement remarquable et qui est de l'ordre de 0,03 ce qui est le même pour l'indice  $\Delta f_{C8}$ . Cependant, la différence pour les indices  $\Delta f_O$  et  $\Delta f_{N12}$  est de l'ordre de 0,2. Les valeurs de l'indice  $\Delta f_S$  varient entre -0,133 et 0,304.

Cependant, ces propriétés de la densité électronique d'une molécule peuvent être utilisées pour exprimer les interactions avec d'autres molécules. C'est le cas de l'interaction d'une molécule biologiquement active avec le site actif.

## Chapitre.III : Structures et propriétés électroniques de dérivés HEPT

### IV. Conclusion

Dans ce chapitre, on s'est attaché à étudier la structure et quelques propriétés électroniques pour différentes séries de molécules.

En se basant sur différentes approches théoriques, à savoir :

- ✓ Les énergies des systèmes neutres, anioniques et cationiques.

On a pu identifier les différents indices globaux tels que le potentiel chimique, le potentiel d'ionisation, l'affinité électronique, la dureté, la mollesse globale, l'électronégativité et l'indice électrophile.

D'autres part et en utilisant :

- ✓ Les charges atomiques sur les différents sites des dérivés de HEPT.

On a calculé les indices de Fukui ainsi que les descripteurs duals de la réactivité.

L'importance relative de ces résultats est affectée par la présence des radicaux en différentes positions sur les deux cycles. Une étude détaillée sur les aspects structuraux et sur les propriétés électroniques pour voir l'effet de ces substituants sur ces derniers.

Dans un premier temps, une analyse sur la position du radical  $R_1$  sur le thiophényl a montré que les positions des deux cycles sont guidées par les interactions et les encombrements stériques entre le radical  $R_2$  de la thymine et le  $R_1$  du thiophényl.

Dans un second temps, on a étudié les propriétés électroniques, à savoir, le potentiel d'ionisation, l'affinité électronique et certains indices de la réactivité en fonctions de la position et la nature des groupements portés par les deux cycles.

D'un point de vue général, l'étude de ces différentes analyses peut être un point clés dans le développement de modèles prédictifs des activités pour de telles molécules, que ce soit pour la recherche de descripteurs ou pour la mise en place des bases de données. Cette étude est donc tout à fait complémentaire de l'analyse QSAR puisqu'elle permet de mieux saisir certains points clés du développement des modèles.

## Chapitre.III : Structures et propriétés électroniques de dérivés

### HEPT

#### Bibliographie

- [1]. Parr, R. G., & Yang, W. (1989). *Density-Functional Theory of Atoms and Molecules* New York: *Oxford Univ. Press*.
- [2]. McLean, A. D., & Chandler, G. S. (1980). Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z= 11–18. *The Journal of Chemical Physics*, 72(10), 5639–5648.
- [3]. Krishnan, R. B. J. S., Binkley, J. S., Seeger, R., & Pople, J. A. (1980). Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *The Journal of Chemical Physics*, 72(1), 650–654.
- [4]. Clark, T., Chandrasekhar, J., Spitznagel, G. W., & Schleyer, P. V. R. (1983). Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+ G basis set for first-row elements, Li–F. *Journal of Computational Chemistry*, 4(3), 294–301.
- [5]. Frisch, M. J., Pople, J. A., & Binkley, J. S. (1984). Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *The Journal of Chemical Physics*, 80(7), 326.
- [6]. Frisch, M. J. E. A., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., ... & Nakatsuji, H. (2009). Gaussian 09, revision D. 01.
- [7]. Hirshfeld, F. L. (1977). Bonded-atom fragments for describing molecular charge densities. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling. Theoretica Chimica Acta*, 44(2), 129–138.

*Chapitre IV*  
*Modèles QSAR de l'inhibition*  
*de la transcriptase inverse de*  
*dérivés HEPT*

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

### I. Introduction

Comme toute activité biologique, la compréhension de l'inhibition de la TI de VIH peut être obtenue en utilisant une approche QSAR. De telles approches sont basées sur la déduction de l'activité biologique à partir de descripteurs traduisant des propriétés structurales de molécules. Cependant, il existe un nombre important de descripteurs et seulement un petit nombre s'avèrent pertinents dont l'utilisation permet d'obtenir des modèles QSAR efficaces.

Pour sélectionner les descripteurs les plus pertinents pour les composés HEPT, les méthodes de sélection des descripteurs appropriées sont nécessaires. De nombreuses recherches ont mené des sélections de descripteurs sur les composés HEPT et une variété de techniques statistiques pour produire des modèles QSAR.

Afin de mener bien la conception de nouveaux modèles QSAR l'activité anti-VIH de dérivés du HEPT, j'ai dans un premier temps choisi d'examiner les différents modèles de la littérature que la recherche bibliographique approfondie m'a révélé. Tenant compte des résultats de ces analyses, il m'a semblé intéressant de mettre au point des modèles QSAR de l'activité anti-VIH de dérivés HEPT en utilisant les indices de Fukui comme descripteurs.

Dans les QSAR décrits ci-dessous, divers paramètres biologiques ont été utilisés. Cependant, dans la plupart d'entre eux, l'activité anti-VIH des composés est exprimée par  $IC_{50}$  ou  $EC_{50}$ . La  $IC_{50}$  se réfère à la concentration minimale du composé conduisant à une inhibition de 50% de l'enzyme TI. La  $EC_{50}$  (ou  $ED_{50}$ ) se réfère à la concentration ( $\text{mol L}^{-1}$  ou  $\text{mol g}^{-1}$ ) du composé nécessaire pour atteindre une protection de 50% des cellules MT-4 contre l'effet cytopathique du virus. Les logarithmes de l'inverse de ces paramètres se réfèrent aux points finaux biologiques ( $\log 1 / C$ ) dans les études QSAR

Les corrélats de ces paramètres biologiques sont divers paramètres physico-chimiques, de nature hydrophobe, électronique, stérique ou topologique. Ceux qui sont importants dans plusieurs modèles QSAR sont brièvement expliqués ici [1].

- ✓ Les descripteurs hydrophobes : sont principalement obtenus expérimentalement :  $\log P$ , où  $P$  est le coefficient de partage octanol-eau et  $\pi$  est la constante hydrophobe des substituants.
- ✓ Les descripteurs électroniques : c'est les constantes de Hammett  $\sigma$ ,  $\sigma^-$  et  $\sigma^+$  s'appliquent aux effets de substitution sur les systèmes aromatiques et  $\sigma^*$  constante de Taft s'applique aux systèmes aliphatiques.
- ✓ Les descripteurs stériques : sont le paramètre stérique Tafts  $E_s$ , le volume de McGowan  $MgVol$ , le volume  $V_w$  de van der Waals et le poids moléculaire  $MW$ .
- ✓ Les descripteurs Stérimol de Verloop  $B_1$ ,  $B_5$  et  $L$ : ils servent à mesurer la largeur et la longueur des substituants.  $B_1$  est une mesure de la largeur du premier atome d'un substituant,  $B_5$  est une tentative pour définir le volume global et  $L$  est la longueur du substituant.
- ✓ La réfractivité molaire  $MR$  est en grande partie une mesure du volume avec une petite correction pour la polarisabilité.

## **Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT**

### **II. Modèles QSAR de l'activité anti-VIH de dérivé HEPT publiés dans la littérature**

Parmi les différentes méthodes de chimiométrie, la régression linéaire multiple (MLR) et les moindres carrés partiels (PLS) sont deux méthodes qui ont été utilisées dans plusieurs études QSAR des inhibiteurs de la transcriptase inverse. Kireev et ses collègues ont utilisé MLR pour relier l'activité inhibitrice de la TI des analogues de HEPT et Luco et Ferretti ont développé un modèle QSAR basé sur les méthodes MLR et PLS pour l'activité anti-VIH d'un large groupe de dérivés HEPT. Ces derniers ont conclu que le PLS est une meilleure approche que MLR pour améliorer l'interopérabilité des données et aussi pour présenter des modèles ayant un meilleur pouvoir prédictif.

Cependant, un problème dans les méthodes MLR et PLS est le fait qu'en utilisant des méthodes de calcul pour générer, à partir des descripteurs, un ensemble de données. Dans le cas où le nombre de données à générer est inférieur au nombre de descripteurs, les corrélations observées peuvent être des corrélations du hasard. Un deuxième problème rencontré dans l'utilisation de l'analyse de régression est l'hypothèse d'une relation linéaire entre l'activité biologique et un ou plusieurs descripteurs. Cependant, les phénomènes biologiques sont considérés comme non-linéaires par nature, et par conséquent la contribution de certains des paramètres aux propriétés d'inhibition de TI peut être non linéaire. La clé de la résolution de ces problèmes est l'utilisation des méthodes non linéaires comme les réseaux neuronaux artificiels (RNA) ou les machines à vecteurs de support (SVM).

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

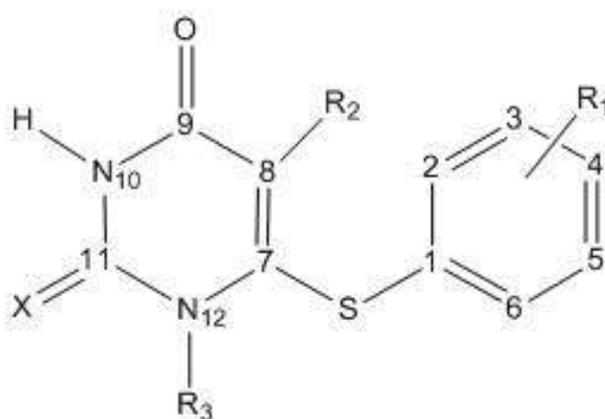
### II.1 Méthodes de régression linéaire multiple MLR

A notre connaissance, la première étude de type QSAR concernant l'activité anti-HIV des composés HEPT a été réalisée par Hansch et Zhang [2].

Toutes les données expérimentales ont été tirées de diverses études de H. Tanaka et de collaborateurs [3-8].

#### II.1.1 Modèle de Hansch et Zhang

En 1992, les travaux de Hansch et Zhang permettent l'élaboration du premier modèle de QSAR sur les dérivés de HEPT (Fig.1) en étudiant la concentration molaire du médicament qui offre une protection de 50% aux cellules MT-4 des dérivés HEPT, le modèle comprend un ensemble de 33 molécules [10].



Avec  $R_3 = \text{CH}_2\text{OCH}_2\text{CH}_2\text{OH}$

Figure.1 Structure générale de la molécule HEPT.

Le modèle suivant a été proposé pour prédire l'activité avec un coefficient de corrélation de 0.941.

$$\begin{aligned} \text{Log}1/C = & 0.88(\pm 0.19)\Sigma\pi + 12.1(\pm 3.8)L_{R_2} - 1.59(\pm 0.49)(L_{R_2})^2 \\ & + 1.17(\pm 0.86)(B_1 - 3R_1) + 1.53(\pm 0.82)(E_s - 2R_1) - 15.3(\pm 7.4) \end{aligned} \quad (1)$$

$n = 33, R^2 = 0.941, S = 0.500$

- ✓ C : est la concentration molaire du médicament offrant une protection de 50% aux cellules MT-4 ;
- ✓  $\Sigma\pi$  : Les constantes hydrophobes pour les substituants  $R_1$  et  $R_2$  ;
- ✓ L est la longueur du substituant ;
- ✓  $B_1-3R_1$  est la largeur de  $R_1$  en position 3 (lorsque les  $R_1$  sont présents dans les deux positions 3 et 5, seulement un seul  $B_1-3R_1$  est utilisé, c'est-à-dire qu'un seul substituant en position méta semble montrer l'effet stérique positif) ;
- ✓  $E_s-2R_1$  est la constante stérique Taft pour les substituants en position ortho.

**Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT**

Tableau.1 : Les inhibiteurs HEPT utilisés dans ce modèle

N°	X	R <sub>1</sub>	R <sub>2</sub>	Obs <sup>a</sup>	Calc <sup>b</sup>
1	O	2-Me	Me	4.15	3.90
2	O	2-NO <sub>2</sub>	Me	3.85	4.08
3	O	2-OMe	Me	4.12	4.85
4	O	3-Me	Me	5.59	5.44
5	O	3-Et	Me	5.51	5.84
6	O	3-t-C <sub>4</sub> H <sub>9</sub>	Me	4.92	5.52
7	O	3-CF <sub>3</sub>	Me	4.35	5.22
8	O	3-F	Me	5.48	5.26
9	O	3-Cl	Me	4.89	5.27
10	O	3-Br	Me	5.24	5.24
11	O	3-I	Me	5.00	5.26
12	O	3-NO <sub>2</sub>	Me	4.41	4.52
13	O	3-OH	Me	4.09	4.55
14	O	3-OMe	Me	4.66	5.12
15	O	4-Me	Me <sub>2</sub>	3.66	6.00
16	O	3,5-di-Me	Me	6.59	5.93
17	O	3,5-di-Me	Me	5.89	5.88
18	S	3,5-di-Me	Me	6.66	5.93
19	O	3-COOMe	Me	5.10	4.82
20	O	3-COMe	Me	5.14	4.39
21	O	3-CN	Me	5.00	4.31
22	O	H	Allyl	5.60	4.71
23	O	H	Et	6.96	7.18
24	S	H	C <sub>3</sub> H <sub>7</sub>	5.00	5.83
25	S	H	CHMe <sub>2</sub>	7.23	1.63
26	S	3,5-di-Me	Et	8.11	1.60
27	S	3,5-di-Me	CHMe <sub>2</sub>	8.30	8.04
28	S	3,5-di-Cl	Et	7.31	7.56
29	O	H	Et	6.92	7.18
30	O	H	C <sub>3</sub> H <sub>7</sub>	5.41	5.83
31	O	H	CHMe <sub>2</sub>	7.20	1.62
32	O	3,5-di-Me	Et	1.89	7.60
33	O	3,5-di-Me	CHMe <sub>2</sub>	8.57	8.04
34	O	3,5-di-Cl	Et	1.85	7.56

<sup>a</sup> : Activité observée, <sup>b</sup>: Activité calculée.

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

### II.1.2 Modèle de Hannongbua

Le caractère électronique et stérique des molécules joue un rôle important dans les interactions médicament-récepteur. Le caractère électronique se rapporte directement aux propriétés de densité électronique des molécules interagissant sur le site actif, alors que les effets stériques dépendent de leur forme géométrique

Dans le but de bien présenter cette corrélation, une étude basée sur les propriétés électroniques et moléculaires était faite sur 40 inhibiteurs du VIH-1[9]. Les géométries optimisées par la méthode semi-empirique AM1, ont été utilisées pour calculer les propriétés électroniques et moléculaires de tous les composés.

Le modèle de Hannongbua, publié en 1995, se base sur la structure décrite ci-dessous.

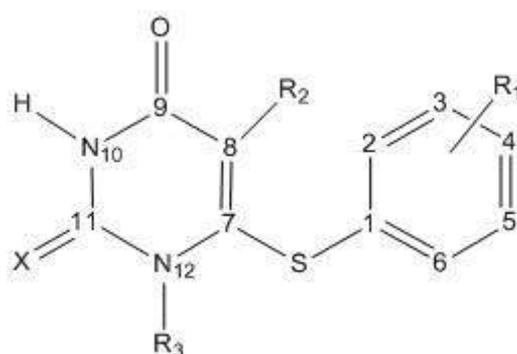


Figure.2 Structure générale de HEPT.

Afin d'examiner la relation entre les activités biologiques et les propriétés structurales, plusieurs modèles de régression linéaire ont été développés. Le meilleur modèle QSAR obtenu, montrant une signification statistique et aussi une capacité prédictive est basé sur l'utilisation de descripteurs moléculaires qui sont les charges atomiques, les énergies d'hydratation, les réfractivités molaires et le coefficient de partage log P.

$$\text{Log}1/C = 0.228(\pm 0.121)HE + 0.118(\pm 0.049)MR + 31.072(\pm 33.111)C8 + 28.979(\pm 24.851)C7 + 10.195(\pm 8.339) \quad (2)$$

$$n = 35, R^2 = 0.869, S = 0.743$$

- ✓ C est la concentration molaire du médicament offrant une protection de 50% aux cellules MT-4
- ✓ HE : L'énergie d'hydratation ;
- ✓ MR : La réfractivité molaire ;
- ✓ C8 : La charge de l'atome C<sub>8</sub> ;
- ✓ C7 : La charge de l'atome C<sub>7</sub>.

**Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT**

Tableau.2 : Les dérivés de HEPT utilisés dans cette étude

N°	X	R <sub>1</sub>	R <sub>2</sub>	Obs <sup>a</sup>	Calc <sup>b</sup>
1	O	2-Me	Me	4.146	5.124
2	O	2-Cl	Me	3.885	5.104
3	O	2-NO <sub>2</sub>	Me	3.853	3.857
4	O	2-OMe	Me	4.720	5.186
5	O	3-Et	Me	5.568	5.484
6	O	3-tBu	Me	4.920	6.717
7	O	3-CH <sub>2</sub> OH	Me	3.534	4.023
8	O	3-CF <sub>3</sub>	Me	4.346	5.664
9	O	3-F	Me	5.480	4.758
10	O	3-Cl	Me	4.885	4.978
11	O	3-NO <sub>2</sub>	Me	4.468	3.645
12	O	3-OMe	Me	4.657	4.842
13	O	3,5-Me <sub>2</sub>	Me	6.584	6.456
14	O	3,5-Cl <sub>2</sub>	Me	5.885	5.459
15	S	3,5-Me <sub>2</sub>	Me	6.656	6.442
16	O	3-COOMe	Me	5.102	4.769
17	O	3-COMe	Me	5.136	5.052
18	O	4-COMe	Me	3.958	4.788
19	O	3-COOH	Me	3.453	3.566
20	O	3-CN	Me	4.999	4.458
21	O	H	CH <sub>2</sub> CH=CH <sub>2</sub>	5.601	6.402
22	O	H	COOMe	5.180	5.439
23	O	H	CONHPh	4.744	4.360
24	S	H	Et	6.957	6.202
25	S	H	i-Pr	7.228	7.422
26	S	3,5-Me <sub>2</sub>	i-Pr	8.299	8.468
27	S	3,5-Cl <sub>2</sub>	Et	7.365	7.386
28	O	H	Et	6.919	6.175
29	O	H	Pr	5.467	6.071
30	O	H	i-Pr	7.199	7.007
31	O	3,5-Me <sub>2</sub>	Et	7.885	7.360
32	O	3,5-Me <sub>2</sub>	i-Pr	8.567	7.720
33	O	3,5-Cl <sub>2</sub>	Et	7.852	6.603
34	O	H	Me	5.154	4.527
35	S	H	Me	6.008	5.183

<sup>a</sup> : Activité observée, <sup>b</sup> : Activité calculée.

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

### II.1.3 Modèle de Kireev

Le modèle proposé par Kireev et ses collaborateurs en 1997[10] était dans le cadre d'une étude pour une famille de 87 dérivés de HEPT qui est un puissant inhibiteur du VIH-1 de la transcriptase inverse (TI). Parmi ces 87 inhibiteurs non nucléosidiques de la TI, 9 nouveaux analogues de HEPT ont été synthétisés par Pontikis et ses collaborateurs [11] et les autres ont été tirés de la littérature [3-8].

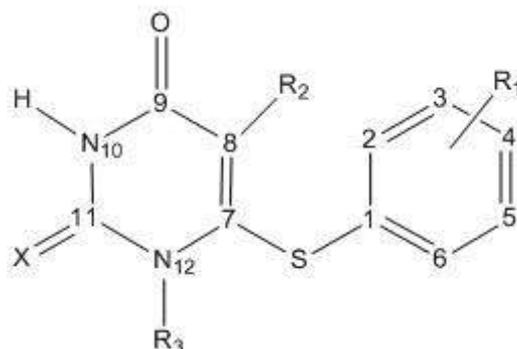


Figure.3 Structure de la molécule HEPT dans cette étude

En dérivant les modèles QSAR partiels, cinq descripteurs ont été sélectionnés comme les plus pertinents. Enfin, ces cinq descripteurs ont été utilisés pour obtenir la meilleure équation reproduisant l'activité anti-VIH d'une série de 33 composés. C'est une équation à quatre termes.

$$\begin{aligned} \text{Log}1/EC_{50} = & -1.55(\pm 0.302) \left( \sum q^- \right)_7 + 2.10(\pm 0.190) \left( \frac{1}{2} W_8 \right) - 0.728(\pm 0.122) (W_{cis}) \\ & + 0.362(\pm 0.0889) (RB1_{cis}) + 2.86(\pm 0.0995) \end{aligned} \quad (3)$$

n=33, R<sup>2</sup>=0.95, S=0.43

- ✓  $(\sum q^-)_7$  : La somme des charges négatives sur les atomes du thiophényl ;
- ✓  $\frac{1}{2}W_8$  : La largeur du substituant R<sub>2</sub> ;
- ✓ W<sub>cis</sub> : La largeur de la conformation Cis ;
- ✓ RB1<sub>cis</sub> : La barrière de rotation pour la liaison rotative S-1 dans les conformères 1,6-cis.

La capacité prédictive de ces relations a été évaluée en utilisant un ensemble de 54 composés. Des descripteurs liés aux changements conformationnels ont été trouvés comme un facteur important qui sous-tend l'activité inhibitrice de la TI.

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

Ce modèle QSAR fournit des preuves concernant les transformations conformationnelles que peuvent subir les molécules pendant le processus d'inhibition. Les relations établies étaient complémentaires à une étude expérimentale sur la liaison des inhibiteurs de type HEPT à la RT par Hopkins et ses collaborateurs [12].

### II.1.4 Modèle de Luco et Ferretti

Dans la même année Luco et Ferretti [13] ont publié une étude QSAR sur un ensemble de 107 dérivés HEPT en utilisant plusieurs descripteurs structuraux et physico-chimiques.

L'activité de ces composés a été étudiée en utilisant les techniques de régression linéaire multiple.

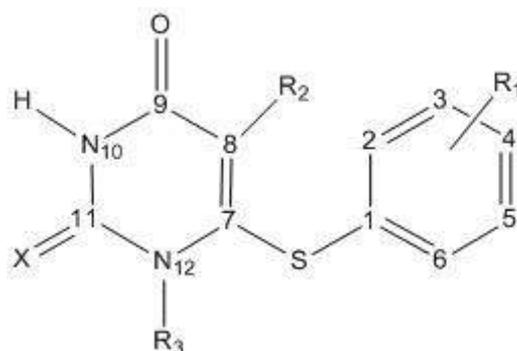


Figure.4 Structure de la molécule HEPT dans cette étude

$$\begin{aligned} \log 1/C = & 0.65(\pm 0.09) \sum \pi(R_1 + R_2) + 130(\pm 17.05)^1 \chi^N(R_2) \\ & - 197.65(\pm 23.2) (\chi^N(R_2))^2 - 1.14(\pm 0.19) B_1 - 3R_1 - 1.60(\pm 0.28) E_s \\ & - 2R_1 + 63.31(\pm 7.23) \chi^N_P + 12.04(\pm 2.3) Vx/100 \\ & - 2.13(\pm 0.43) (Vx/100)^2 - 2.27(\pm 0.30) Ich - R_3 - 1.25(\pm 0.20) I - 4R_1 \\ & - 0.72(\pm 0.21) \Delta X(R_3) - 47.7(\pm 2.96) \end{aligned} \quad (4)$$

n=107, R<sup>2</sup>= 0.951, R<sup>2</sup><sub>cv</sub>=0.820, S=0.49.

- ✓  $\sum \pi(R_1 + R_2)$  : les constantes de Hansch pour les substituants R<sub>1</sub> et R<sub>2</sub> ;
- ✓  $\chi^N(R_2)$  : l'indice de connectivité de la variation structurale à la position C-8 (R<sub>2</sub>) ;
- ✓ B<sub>1</sub>-3R<sub>1</sub> : la largeur de R<sub>1</sub> en position 3 ;
- ✓ E<sub>s</sub>-2R<sub>1</sub> : la constante stérique de Taft pour les substituants en ortho ;
- ✓  $\chi^N_P$  : l'indice de connectivité moléculaire au quatrième ordre ;
- ✓ Vx : le volume moléculaire caractéristique de McGowan [14-15] ;
- ✓ Ich-R<sub>3</sub> : une variable qui prend la valeur 1 ou 0 pour la présence ou l'absence d'un cycle saturé à six chaînons dans R<sub>3</sub> ;
- ✓ I-4R<sub>1</sub> : une variable indicatrice prenant la valeur 1 ou 0 pour la présence ou l'absence d'un substituant en position 4 du cycle aromatique C-7 ;
- ✓  $\Delta X(R_3)$  : l'indice de connectivité moléculaire [16].

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

### II.1.5 Modèle de Jalali-Heravi et Parastar

Un autre modèle de QSAR pour les dérivés HEPT était établi en 2000 [17], en utilisant toujours la méthode MLR pour un ensemble de 107 molécules.

$$\begin{aligned} \text{Log}1/C = & 2.977(\pm 1.335) + 3.307(\pm 0.679) 1/S + 2.013(\pm 0.741) \text{POS/NEG} + 6.689 \\ & * 10^{-3}(\pm 1.631 * 10^{-3}) \Delta H_f + 0.141(\pm 0.028)(\text{NCSP3} - R_2)^2 \\ & - 0.804(\pm 0.171) \text{NOH} - R_3 + 5.079 \\ & * 10^{-3}(\pm 9.938 * 10^{-4})(\text{NS} - R_3)^3 \end{aligned} \quad (5)$$

n=80, R2=0.901, S=0.607

- ✓ 1/S : l'inverse de la zone d'ombre standard sur le plan YZ ;
- ✓ POS/NEG : Le rapport des charges partielles sur les atomes les plus positifs et les plus négatifs ;
- ✓  $\Delta H_f$  : La chaleur de formation (kcal / mol) ;
- ✓  $(\text{NCSP}^3 - R_2)^2$  : Le carré du nombre d'atomes de carbone  $\text{SP}^3$  du substituant  $R_2$  ;
- ✓  $\text{NOH} - R_3$  : Le nombre de groupes hydroxyle sur le substituant  $R_3$  ;
- ✓  $(\text{NS} - R_1)^3$  : Le cube de la sommation des positions de  $R_1$  sur le cycle aromatique.

### II.1.6 Modèle de Bazoui et ses collaborateurs :

En utilisant la méthode MLR avec seulement 4 descripteurs moléculaires, Bazoui [18] ont établi un modèle QSAR pour un groupe de 95 dérivés HEPT. Malheureusement, ces paramètres ne tiennent compte que de  $R_1$  à la position 5,  $R_2$  et  $R_3$  de la structure décrite ci-dessous (Fig.5).

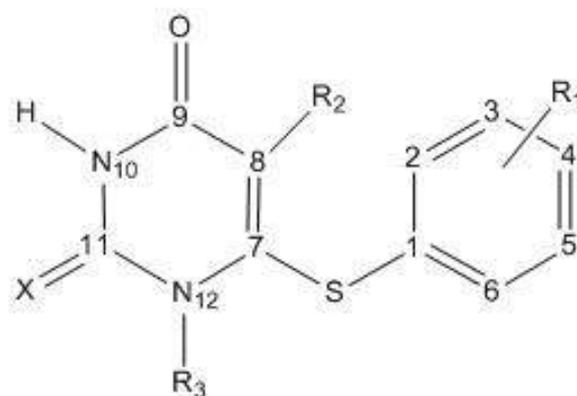


Figure.5 La structure de HEPT étudiée

Les descripteurs impliqués dans ce modèle sont :

- ✓  $\text{Log } P(R_3)$  : Le coefficient de partition du  $R_3$  ;
- ✓  $\text{Log } P(R_1)$  : Le coefficient de partition du  $R_1$  en position 5 ;
- ✓  $\text{MW}(R_2)$  : Le poids moléculaire du  $R_2$  ;
- ✓  $\text{MR}(R_2)$  : La réfractivité moléculaire du  $R_2$ .

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

### II.2 Méthodes des moindres carrés partielles PLS

Pour tenir compte de la variation chimique et structurale des substituants  $R_1$  et  $R_2$ , Luco et Ferritti [13] ont considéré les descripteurs physico-chimiques suivants : la constante de Hansch, la constante de Hammett, la constante stérique de Taft pour les substituants ortho et la réfractivité molaire MR.

De plus, pour tenir compte de la capacité des substituants  $R_1$  à former des liaisons hydrogène, le paramètre  $\Lambda_{alk}$  tel que défini par Van de Waterbeemd et Kansy [19] a été utilisé.

Il est important de souligner aussi que le modèles PLS présentés dans cette étude pour un nombre de 107 molécules (avec les mêmes descripteurs de l'eq.4) a une capacité prédictive nettement meilleures avec les statistiques suivantes :  $R^2 = 0,944$ ,  $R^2_{cv} = 0,931$ ,  $S=0.503$ .

### II.3 Méthodes des réseaux neurones

Il y a eu beaucoup d'intérêt pour les réseaux neurones artificiels (RNA) dans de nombreux domaines de la science et de la technologie tels que l'ingénierie, la biologie et dans le domaine des sciences cognitives. L'application du RNA apparaît également dans plusieurs domaines de la chimie, y compris l'étude de QSAR, l'étude des relations structure-odeur, l'estimation des propriétés physiques, la prédiction de la réactivité chimique et la détermination de la structure protéique.

Jalali et ses collaborateurs sont publié la première application de RNA en 2000[17] pour un ensemble de 107 dérivés de HEPT.

Le meilleur modèle de RNA était une méthode feed-forward avec une architecture 6-6-1. L'erreur quadratique moyenne (MSE) pour l'ensemble de prédiction utilisant ce réseau était de 0,372 par rapport à 0,780 obtenue en utilisant la technique MLR (l'eq.5).

Tableau.3 : Les descripteurs impliqués dans cette étude.

<i>Descripteur</i>	<i>Notation</i>
L'inverse de la zone d'ombre standard sur le plan YZ	1/S
Le rapport des charges partielles sur les atomes les plus positifs et les plus négatifs	POS/NEG
La chaleur de formation (kcal / mol)	$\Delta H_f$
Le carré du nombre d'atomes de carbone $SP^3$ du substituant $R_2$	$(NCSP^3-R_2)^2$
Le nombre de groupes hydroxyle sur le substituant $R_3$	NOH- $R_3$
Le cube de la somation des positions de $R_1$ sur le cycle aromatique	$(NS-R_1)^3$

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

La comparaison de la qualité du RNA avec le modèle MLR (II.1.5) montre que le réseau neurone a un meilleur pouvoir prédictif donc on peut conclure que la contribution de certains des paramètres à la propriété d'inhibition de la TI peut être non linéaire, cette conclusion découle du fait que les mêmes descripteurs ont été utilisés pour le développement du modèle MLR et du réseau neurone.

Cette étude permet de conclure que la méthode RNA est capable d'établir une relation satisfaisante entre l'activité anti-VIH et les descripteurs moléculaires choisis. Cependant, ces descripteurs n'étaient pas satisfaisants car ils ne tenaient pas compte des effets hydrophobes moléculaires. En fait, de nombreux QSAR, utilisant la méthode MLR, montrent que l'hydrophobie joue le rôle le plus important dans la description de l'activité moléculaire.

Douali et ses collaborateurs [20] ont remarqué évidemment que le substituant R<sub>1</sub> en position 3 a un impact sur l'activité anti-VIH donc il était nécessaire de prendre en considération la contribution de R<sub>1</sub> à cette position.

Le modèle QSAR était pour un ensemble de 80 composés en utilisant les mêmes descripteurs de Luco et Ferretti [13] avec un coefficient de corrélation de R<sup>2</sup>= 0,984.

Tableau.4 : Les résultats statistiques de différentes architectures de réseaux de neurones.

Architecture	R <sup>2</sup>	S
8-3-1	0.958	0.272
8-4-1	0.971	0.227
8-5-1	0.974	0.215
8-6-1	0.980	0.187
8-7-1	0.984	0.165

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

### III. Modèles QSAR de l'inhibition de la transcriptase inverse par les dérivés HEPT à l'aide des indices de Fukui

Une part du succès de tout modèle QSAR réside dans le choix des descripteurs employés. En général, les descripteurs classiques utilisés pour de telles études sont des descripteurs constitutionnels, topologiques et des fois géométriques. L'emploi des descripteurs quantiques est moins fréquent en QSAR alors qu'ils présentent l'avantage de pouvoir exprimer les propriétés de la densité électronique des systèmes moléculaires.

Dans cette étude, la pertinence de tels descripteurs pour prédire l'activité anti-HIV des 60 dérivés de HEPT représentés sur le tableau.6 a été analysée.

#### III.1 Méthodologie

##### III.1.1 Base de données

Les travaux de Tanaka et collaborateurs [3-8] ont permis de produire des données d'activité cytotoxique *in vitro* (EC<sub>50</sub>) de plusieurs analogues de HEPT.

L'évaluation biologique de la cytotoxicité de ces structures sur la lignée cellulaire humaine de cellules MT-4, en utilisant la méthode d'exclusion au bleu trypan, la microscopie indirecte par immunofluorescence et la cytofluorographie par flux laser a permis d'obtenir des valeurs de EC<sub>50</sub>.

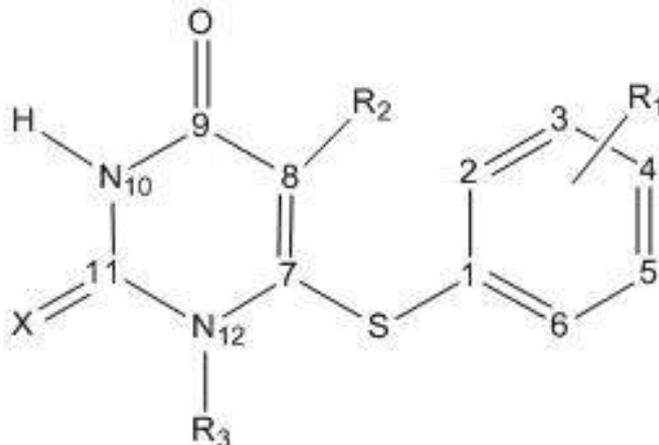


Figure.6. Structure générale de HEPT de notre travail.

Les activités biologiques ( $\log 1 / EC_{50}$ ) sont indiquées dans le tableau ci-dessous

**Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT**

Tableau.5 : Les structures chimiques ainsi que les activités observées des composés.

N°	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	X	Log1/EC <sub>50</sub>
1	2-Me	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	4,15
2	2-NO <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	3,85
3	2-OMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	4,72
4	3-Me	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	5,59
5	3-Et	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	5,57
7	3-CF <sub>3</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	4,35
8	3-F	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	5,48
12	3-NO <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	4,47
13	3-OH	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	4,09
14	3-OMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	4,66
15	3,5-Me <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	6,59
16	3,5-Cl <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	5,89
18	3-COOMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	5,10
19	3-COMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	5,14
20	3-CN	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	5,00
21	H	CH <sub>2</sub> CH=CH <sub>2</sub>	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	5,60
22	H	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	6,96
24	H	i-Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	7,23
25	3,5-Me <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	8,11
26	3,5-Me <sub>2</sub>	i-Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	8,30
29	H	Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	5,47
32	3,5-Me <sub>2</sub>	i-Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	8,57
35	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	5,15
36	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	6,01
38	H	CH=CH <sub>2</sub>	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	5,69
40	H	CH <sub>2</sub> Ph	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	4,37
41	H	CH=CPh <sub>2</sub>	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	6,07

**Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT**

44	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OCOPh	O	5,12
46	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> Cl	O	5,82
48	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> F	O	5,96
49	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> Me	O	5,48
50	H	Me	CH <sub>2</sub> OCH <sub>2</sub> Ph	O	7,06
51	H	Et	CH <sub>2</sub> OCH <sub>2</sub> Me	O	7,72
52	H	Et	CH <sub>2</sub> OCH <sub>2</sub> Me	S	7,58
53	3,5-Me <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> Me	O	8,24
54	3,5-Me <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> Me	S	8,30
57	H	Et	CH <sub>2</sub> OCH <sub>2</sub> Ph	S	8,09
58	3,5-Me <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> Ph	S	8,14
61	H	i-Pr	CH <sub>2</sub> OCH <sub>2</sub> Me	S	7,89
62	H	i-Pr	CH <sub>2</sub> OCH <sub>2</sub> Ph	S	8,14
63	H	Me	CH <sub>2</sub> OMe	O	5,68
64	H	Me	CH <sub>2</sub> OBu	O	5,33
65	H	Me	Et	O	5,66
66	H	Me	Bu	O	5,92
67	3,5-Cl <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> Me	S	7,89
68	H	Et	CH <sub>2</sub> O-i-Pr	S	6,66
70	H	Et	CH <sub>2</sub> OCH <sub>2</sub> -c-Hex	S	6,45
71	H	Et	CH <sub>2</sub> OCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-Me)	S	7,11
72	H	Et	CH <sub>2</sub> OCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-Cl)	S	7,92
73	H	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> Ph	S	7,04
76	H	Et	CH <sub>2</sub> O-c-Hex	O	5,40
77	H	Et	CH <sub>2</sub> OCH <sub>2</sub> -c-Hex	O	6,35
90	4-COMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	3,96
93	H	COOMe	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	5,18
95	H	SPh	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	4,68
96	H	C≡CH	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	4,74
97	H	C≡CPh	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	5,47

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

101	H	C≡CMe	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	4,72
103	H	Cl	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	4,52
104	H	Br	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	4,70

### III.1.2 Calcul des descripteurs quantiques

Quatre descripteurs globaux de réactivité, issus de la DFT conceptuelle, ont été calculés : potentiel d'ionisation (PI), affinité électronique (AE), Mollesse globale (S) et l'indice d'électrophilicité ( $\omega$ ) et cinq indices de Fukui. Plus de détails sur ces descripteurs sont disponibles dans le chapitre II.

Ces descripteurs représentent un moyen simple de rationaliser les différences de comportement chimique existant entre les molécules.

Tableau.6 : Les descripteurs utilisés dans les modèles.

Descripteurs	Significations
D1	Le potentiel d'ionisation
D2	L'affinité électronique
D3	La Mollesse globale
D4	L'indice d'électrophilicité
D5	Descripteur dual du site C <sub>7</sub>
D6	Descripteur dual du site C <sub>8</sub>
D7	Descripteur dual du site S
D8	Descripteur dual du site O
D9	Descripteur dual du site N <sub>12</sub>

**Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT**

Tableau.7 : Les valeurs des descripteurs.

<b>N°</b>	<b>D1</b>	<b>D2</b>	<b>D3</b>	<b>D4</b>	<b>D5</b>	<b>D6</b>	<b>D7</b>	<b>D8</b>	<b>D9</b>
<b>1</b>	7,646	0,707	0,144	2,514	0,817	0,643	0,304	-0,581	-0,623
<b>2</b>	8,653	3,945	0,212	8,429	0,838	0,654	0,107	-0,535	-0,598
<b>3</b>	8,027	2,857	0,193	5,728	0,830	0,676	-0,133	-0,663	-0,621
<b>4</b>	7,836	0,190	0,131	2,107	0,844	0,658	0,206	-0,718	-0,641
<b>5</b>	7,962	0,299	0,131	2,227	0,838	0,652	0,077	-0,694	-0,639
<b>7</b>	8,054	0,490	0,132	2,413	0,849	0,667	0,125	-0,727	-0,631
<b>8</b>	8,201	0,457	0,129	2,420	0,841	0,657	0,097	-0,722	-0,637
<b>12</b>	8,163	1,034	0,140	2,966	0,835	0,690	0,120	-0,691	-0,624
<b>13</b>	7,994	0,305	0,130	2,239	0,839	0,652	0,099	-0,686	-0,638
<b>14</b>	7,836	0,816	0,142	2,666	0,831	0,646	0,041	-0,650	-0,520
<b>15</b>	7,885	0,267	0,131	2,181	0,829	0,651	0,080	-0,691	-0,638
<b>16</b>	8,109	1,088	0,142	3,012	0,839	0,676	-0,056	-0,699	-0,637
<b>18</b>	8,168	0,702	0,134	2,635	0,843	0,669	0,093	-0,698	-0,632
<b>19</b>	8,212	0,849	0,136	2,788	0,839	0,666	0,089	-0,689	-0,633
<b>20</b>	6,748	3,428	0,301	7,799	0,824	0,653	0,328	-0,572	-0,631
<b>21</b>	5,524	1,687	0,261	3,388	0,407	0,587	0,453	-0,703	-0,640
<b>22</b>	7,836	0,607	0,138	2,465	0,308	0,659	0,139	-0,631	-0,654
<b>24</b>	7,798	0,637	0,140	2,484	0,309	0,651	0,137	-0,630	-0,649
<b>25</b>	5,173	-0,239	0,185	1,124	2,816	3,569	0,444	-0,823	-0,675
<b>26</b>	7,681	0,577	0,141	2,400	0,310	0,649	0,135	-0,631	-0,649
<b>29</b>	8,041	0,395	0,131	2,326	0,839	0,649	0,055	-0,675	-0,643
<b>32</b>	5,578	0,680	0,204	1,999	0,304	0,592	0,464	-0,733	-0,750
<b>35</b>	8,049	0,313	0,129	2,260	0,841	0,656	0,083	-0,698	-0,640
<b>36</b>	7,918	1,007	0,145	2,881	0,307	0,632	0,144	-0,662	0,646
<b>38</b>	20,388	21,428	-0,962	-420,572	0,870	0,690	0,219	-0,567	-0,690
<b>40</b>	8,008	0,528	0,134	2,435	0,839	0,646	0,104	-0,657	-0,641
<b>41</b>	7,148	0,770	0,157	2,458	0,830	0,661	0,249	-0,589	-0,634

**Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT**

44	7,986	0,694	0,137	2,583	0,837	0,659	0,111	-0,675	-0,635
46	8,027	0,544	0,134	2,454	0,847	0,657	0,149	-0,709	-0,632
48	7,953	0,906	0,142	2,784	0,823	0,637	0,327	-0,597	-0,641
49	8,008	0,280	0,129	2,222	0,841	0,655	0,086	-0,699	-0,639
50	5,524	0,735	0,209	2,045	0,296	0,589	0,460	-0,724	-0,751
51	8,019	0,335	0,130	2,270	0,296	0,589	0,460	-0,724	-0,751
52	7,790	0,569	0,138	2,419	0,307	0,647	0,143	-0,633	-0,652
53	5,524	0,571	0,202	1,875	0,299	0,594	0,461	-0,725	-0,751
54	7,700	0,463	0,138	2,302	0,249	0,649	0,100	-0,669	-0,645
57	7,739	0,612	0,140	2,446	0,307	0,644	0,177	-0,626	-0,646
58	7,619	0,541	0,141	2,352	0,301	0,645	0,158	-0,631	-0,647
61	7,752	0,590	0,140	2,430	0,309	0,651	0,143	-0,632	-0,650
62	7,709	0,631	0,141	2,457	0,308	0,633	0,175	-0,626	-0,647
63	8,049	0,770	0,137	2,671	0,844	0,669	0,074	-0,703	-0,639
64	8,000	0,280	0,130	2,220	0,840	0,654	0,083	-0,697	-0,640
65	7,828	0,770	0,142	2,619	0,828	0,673	0,110	-0,714	-0,634
66	7,973	0,294	0,130	2,225	0,835	0,656	0,104	-0,710	-0,636
67	8,049	6,941	0,903	50,725	0,256	0,662	0,195	-0,645	-0,680
68	7,755	0,612	0,140	2,450	0,303	0,644	0,111	-0,626	-0,653
70	7,768	0,596	0,139	2,439	0,305	0,646	0,138	-0,630	-0,652
71	7,630	0,623	0,143	2,430	0,344	0,824	0,189	-0,620	-0,645
72	7,747	0,732	0,143	2,562	0,303	0,644	0,174	-0,618	-0,644
73	7,657	0,571	0,141	2,389	0,296	0,642	0,184	-0,626	-0,649
76	7,959	0,318	0,131	2,242	0,842	0,652	0,070	-0,687	-0,637
77	8,002	0,359	0,131	2,287	0,840	0,651	63,000	-0,682	-0,642
90	7,828	1,371	0,155	3,277	0,835	0,674	0,118	-0,683	-0,632
93	8,166	0,498	0,130	2,447	0,839	0,673	0,084	-0,669	-0,672
95	7,540	0,629	0,145	2,414	0,832	0,632	0,281	-0,603	-0,644
96	13,034	0,599	0,080	3,736	0,826	0,671	0,374	-0,579	-0,650
97	6,163	1,663	0,222	3,402	0,824	0,653	0,352	-0,554	-0,642

### Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

<b>101</b>	6,498	1,431	0,197	3,102	0,827	0,651	0,377	-0,580	-0,642
<b>103</b>	8,198	0,585	0,131	2,533	0,839	0,635	0,105	-0,654	-0,642
<b>104</b>	8,046	0,544	0,133	2,459	0,837	0,633	0,111	-0,648	-0,641

Les fonctions Dual de Fukui sont des indicateurs locaux de la distribution électronique. Ils ont été utilisés comme descripteurs dans la conception du modèle QSAR. Les indices duals de Fukui ont été évalués pour des atomes C<sub>7</sub>, C<sub>8</sub>, N<sub>12</sub>, O et S. Le tableau 6 montre que l'indice de Fukui de l'atome O (D8) est légèrement le même pour tous les dérivés HEPT étudiés. En général, les valeurs des autres indices locaux de Fukui ont montré qu'elles dépendent étroitement de groupements liés à l'atome. Excepté le numéro de molécule 38, toutes les molécules présentent un potentiel d'ionisation similaire (PI). Cette molécule présente une grande valeur d'IP, d'affinité électronique (AE) et d'indice d'électrophile global par rapport aux autres molécules. Cette différence peut s'expliquer par un effet mésomère. Il est à noter qu'aucune corrélation directe entre un descripteur et l'activité biologique anti-VIH n'a été détectée.

Les deux indices de Fukui,  $\Delta f_{C7}$  et  $\Delta f_{C8}$  semblent jouer un rôle important dans l'équation de corrélation.

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

### III.1.3 Méthode statistique

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

Les modèles QSAR ont été générés en utilisant la méthode de régression multilinéaire (MLR) implémentée dans le logiciel Minitab [21]. Les modèles ont été évalués par les valeurs des coefficients  $R^2$ ,  $R^2_{\text{ajusté}}$  et  $R^2_{\text{cv}}$ .

### III.2 Modèles QSAR

Les valeurs calculées des descripteurs sont reportées dans le Tableau 7.

Plusieurs modèles QSAR linéaires impliquant un, deux et plusieurs descripteurs sont élaborés.

#### *Mod.1*

$$\text{Log1/EC}_{50} = 6,85 - 1,25 \text{ D5} \quad (6)$$

$$N=60, S = 1, 26636, R^2 = 12, 9\%, R^2_{\text{adj}} = 11,4\%$$

#### *Mod.2*

$$\text{Log1/EC}_{50} = 7,18 - 0,146 \text{ D1} \quad (7)$$

$$N=60, S = 1,32672, R^2 = 4,4\%, R^2_{\text{adj}} = 2,8\%$$

On remarque que le meilleur modèle QSAR à un seul descripteur est obtenue avec l'indice de Fukui  $\Delta f_{C7}$  de  $C_7$  (Mod.1,  $R^2 = 0,129$ ).

Afin d'améliorer le pouvoir prédictif des modèles QSAR, il est nécessaire d'élaborer des modèles QSAR impliquant plusieurs descripteurs.

#### *Mod.3*

$$\text{Log1/EC}_{50} = 6,12 - 4,35 \text{ D5} + 4,07 \text{ D6} \quad (8)$$

$$N=60, S = 0,754355, R^2 = 69,6\%, R_{\text{adj}} = 68,6\%$$

#### *Mod.4*

$$\text{Log1/EC}_{50} = 5,20 - 4,49 \text{ D5} + 4,07 \text{ D6} - 1,64 \text{ D9} \quad (9)$$

$$N=60, S = 0,693379, R^2 = 74,8\%, R^2_{\text{adj}} = 73,4\%$$

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

### **Mod.5**

$$\text{Log1/EC}_{50} = 4,03 - 4,56 D5 + 3,92 D6 - 1,23 D9 - 2,42 D8 \quad (10)$$

$$N=60, S = 0,667610, R^2 = 77,0\%, R^2_{\text{adj}} = 75,4\%$$

### **Mod.6**

$$\text{Log1/EC}_{50} = 4,06 - 4,62 D5 + 3,98 D6 - 1,23 D9 - 2,35 D8 + 0,0189 D7 \quad (11)$$

$$N=60, S = 0,654639, R^2 = 78,3\%, R^2_{\text{adj}} = 76,3\%$$

### **Mod.7**

$$\text{Log1/EC}_{50} = 3,98 - 4,67 D5 + 4,00 D6 - 1,18 D9 - 2,53 D8 + 0,0192 D7 - 0,00183 D4 \quad (12)$$

$$N=60, S = 0,652539, R^2 = 78,9\%, R^2_{\text{adj}} = 76,5\%$$

### **Mod.8**

$$\text{Log1/EC}_{50} = 4,01 - 4,33 D5 + 3,60 D6 - 1,10 D9 - 2,08 D8 - 0,0329 D4 - 0,371 D2 + 5,09 D3 \quad (13)$$

$$N=60, S = 0,649105, R^2 = 79,5\%, R^2_{\text{adj}} = 76,7\%$$

### **Mod.9**

$$\text{Log1/EC}_{50} = 4,03 - 4,40 D5 + 3,67 D6 - 1,10 D9 - 2,05 D8 - 0,0311 D4 - 0,348 D2 + 4,79 D3 + 0,0173 D7 \quad (14)$$

$$N=60, S = 0,638080, R^2 = 80,6\%, R^2_{\text{adj}} = 77,1\%$$

### **Mod.10**

$$\text{Log1/EC}_{50} = 3,74 - 4,42 D5 + 3,69 D6 - 1,11 D9 - 2,05 D8 - 0,0324 D4 - 0,373 D2 + 5,15 D3 + 0,0172 D7 + 0,0313 D1 \quad (15)$$

$$N=60, S = 0,643725, R^2 = 80,6\%, R^2_{\text{adj}} = 77,5\%$$

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

La figure.6 montre l'évolution du coefficient de corrélation  $R^2$  des modèles en fonction du nombre de descripteurs. On note une nette amélioration de la qualité des modèles QSAR qui utilise plus de 3 descripteurs.

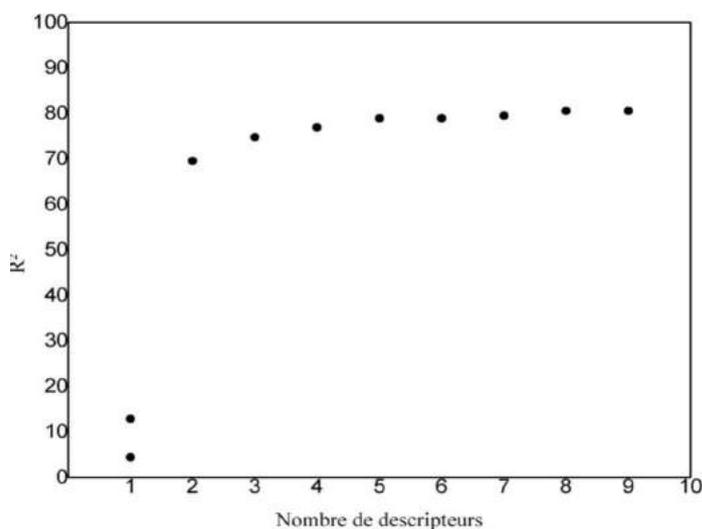


Figure.6 : La corrélation  $R^2$  en fonction du nombre de descripteurs.

La figure.7 montre la corrélation linéaire entre les valeurs observées et les valeurs prédites de l'activité mentionnées dans le tableau.7 obtenue en utilisant le modèle 10.

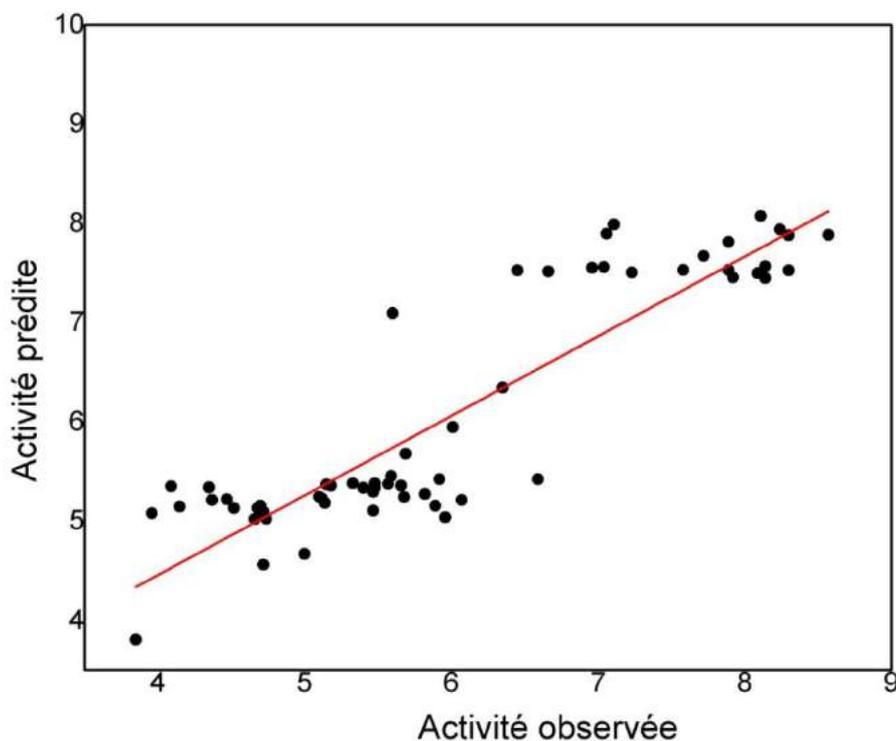


Figure.7 : La corrélation linéaire entre les valeurs observées et les valeurs prédites.

### Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

Comme le montre la figure.7, de bons accords entre les valeurs observées et calculées de (Log1/EC<sub>50</sub>). Il est noté que la fonction de Fukui sur l'atome C<sub>7</sub> et sur l'atome C<sub>8</sub> a la meilleure contribution dans ce modèle.

Tableau. 8 : Les valeurs des activités observées/calculées en utilisant le *Mod.10*

N°	Observée	Calculée	Résiduel
1	4,15	5,15	-1,00
2	3,85	3,81	0,04
3	4,72	4,56	0,16
4	5,59	5,46	0,13
5	5,57	5,37	0,20
7	4,35	5,34	-0,99
8	5,48	5,34	0,14
12	4,47	5,22	-0,75
13	4,09	5,35	-1,26
14	4,66	5,02	-0,36
15	6,59	5,42	1,17
16	5,89	5,15	0,74
18	5,10	5,25	-0,15
19	5,14	5,19	-0,05
20	5,00	4,67	0,33
21	5,60	7,10	-1,50
22	6,96	7,55	-0,59
24	7,23	7,51	-0,28
25	8,11	8,08	0,03
26	8,30	7,53	0,77
29	5,47	5,30	0,17
32	8,57	7,88	0,69
35	5,15	5,37	-0,22
36	6,01	5,95	0,06

**Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT**

---

38	5,69	5,68	0,01
40	4,37	5,21	-0,84
41	6,07	5,21	0,86
44	5,12	5,23	-0,11
46	5,82	5,27	0,55
48	5,96	5,04	0,919
49	5,48	5,38	0,10
50	7,06	7,90	-0,84
51	7,72	7,67	0,05
52	7,58	7,53	0,05
53	8,24	7,94	0,30
54	8,30	7,88	0,42
57	8,09	7,50	0,59
58	8,14	7,57	0,57
61	7,89	7,53	0,36
62	8,14	7,45	0,69
63	5,68	5,24	0,44
64	5,33	5,38	-0,05
65	5,66	5,36	0,30
66	5,92	5,42	0,50
67	7,89	7,82	0,08
68	6,66	7,52	-0,86
70	6,45	7,53	-1,08
71	7,11	7,99	-0,88
72	7,92	7,46	0,46
73	7,04	7,56	-0,52
76	5,4	5,34	0,06
77	6,35	6,35	0,00
90	3,96	5,08	-1,12
93	5,18	5,36	-0,18

---

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

95	4,68	5,13	-0,45
96	4,74	5,02	-0,28
97	5,47	5,11	0,36
101	4,72	5,10	-0,38
103	4,52	5,13	-0,61
104	4,70	5,15	-0,45

### III.3 Validation du meilleur modèle

Pour valider notre modèle (Mod.10), nous avons utilisé la validation interne avec les deux méthodes, la première : Leave-One-Out avec un coefficient  $R^2_{cv} = 79,47\%$  et la deuxième la Leave-Many-Out en suivant les étapes suivantes :

- ✓ Les valeurs de données expérimentales ont été divisées en trois sous-ensembles (A, B et C). Les molécules portant les numéros  $(n*3+1)$  forment le premier sous-ensemble A, les molécules portant les numéros  $(n*3+2)$  forment le sous-ensemble B et les molécules portant les numéros  $(n*3+3)$  forment le troisième sous-ensemble C avec n variant de 0 à 19.
- ✓ En combinant deux des sous-ensembles A, B et C on obtient trois combinaisons et l'équation de corrélation est dérivée avec les mêmes descripteurs.
- ✓ L'équation obtenue a été utilisée pour prédire les données pour le sous-ensemble restant.
- ✓ Il s'avère que les valeurs prédites en utilisant  $R^2$  pour les sous-ensembles (A + B), (A + C) et (B + C) sont très proches de celle correspondant à l'ensemble complet de la série d'apprentissage (A + B + C) et les valeurs moyennes de  $R^2$  pour l'ensemble d'apprentissage et  $R^2$  pour l'ensemble de test (Tableau 8) sont également proches. Notons que la valeur  $R^2$  des modèles correspondant à des sous-ensembles (A + B), (A + C) et (B + C) sont plus proches de 0,80 ce qui indique que notre modèle est stable et peut être efficacement utilisée pour estimer l'activité anti-VIH d'autres molécules pour lesquels les données expérimentales ne sont pas disponibles.

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

Tableau.9 : La validation croisée du meilleur modèle *Mod.10*

Ensemble d'apprentissage	N	R <sup>2</sup>	R <sup>2</sup> <sub>adj</sub>	Ensemble de test	N	R <sup>2</sup>	R <sup>2</sup> <sub>adj</sub>
A + B	40	0,839	0,797	C	20	0,883	0,773
A + C	40	0,830	0,773	B	20	0,882	0,778
B+ C	40	0,791	0,730	A	20	0,945	0,900
<b>Avérage</b>		0,820	0,764			0,910	0,815

De nombreux modèles de relation structure-activité des composés HEPT ont été réalisés au cours des années avec des résultats louables. Pour vérifier la fiabilité du modèle de notre travail et d'autres références, une comparaison entre le R<sup>2</sup> et la sélection de descripteurs des autres travaux est présentée dans le tableau.10.

**Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT**

Tableau.10. Comparaison de différentes méthodes d'analyse QSAR de composés HEPT.

Réf	Méthode	N <sup>br</sup> .mol	N <sup>br</sup> .Descip	R <sup>2</sup>	Type de descrip
[2]	MLR	33	5	0.941	Hydrophobiques
[9]	MLR	40	11	0.869	Electroniques
[13]	MLR	107	10	0.951	Hydrophobiques
	PLS	107	10	0.944	et géométriques
[10]	MLR	87	8	0.940	Flexibilité conformationnelle des substituant, physico-chimiques, géométriques et électroniques
[17]	MLR	107	6	0.901	Topologiques, géométriques, Electroniques,
	RNA	107	6	/	et physico-chimiques
[18]	MLR	76	4	0.960	Physico-chimiques, électroniques et géométriques
[20]	RNA	80	8	0.984	Hydrophobiques et géométriques
[22]	PSO	132	40	0.900	Hydrophobiques, longueur de substituant et indices de connectivité
[23]	MLR	79	6	0.840	Stériques,
	RNA	79	6	0.862	hydrophobiques,
	SVM	79	6	0.880	électroniques et structurales
[24]	RF	132	5	0.830	Hydrophobiques en relation avec la surface de van der Waals et le nombre d'atomes d'oxygène
[25]	MLR	34	9	0.919	Physico-chimiques et quantiques
	MLR	60	9	0.806	Indices de Fukui

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

### IV. Conclusion

En résumé, notre étude montre que les descripteurs de la chimie quantique, à savoir, le potentiel d'ionisation, l'affinité électronique, la mollesse, l'indice électrophile en combinaison avec les indices de Fukui « descripteur dual », sont utiles pour la prédiction de l'activité inhibitrice des dérivés de HEPT.

Le meilleur modèle QSAR (*Mod.10*) est capable de décrire environ 81% de la variance de l'activité expérimentale et pourrait être utilisé efficacement pour estimer l'activité des dérivés de HEPT pour lesquels les données expérimentales sont indisponibles, l'étude montre aussi que l'indice dual de Fukui pour l'atome C<sub>7</sub> constitue le descripteur principal.

Cependant, un autre descripteur qui a été choisi fréquemment dans des études antérieures est la connectivité moléculaire des substituants de HEPT. Luco [13] a montré que les facteurs hydrophobes avaient une importance primordiale pour l'activité anti-VIH. Kireev [10] a suggéré que la flexibilité conformationnelle des composés HEPT joue un rôle important car ils sont extrêmement flexibles. Ceci nous amène à se concentrer sur les descripteurs électroniques, physico-chimiques et géométriques liés aux changements de conformation de HEPT. Plusieurs recherches ont révélé que le coefficient de partage octanol / eau, la connectivité moléculaire et la caractéristique stérique des composés HEPT sont importants dans l'interaction [20-23].

Les résultats peuvent varier en fonction de la nature aléatoire de l'algorithme lors de la sélection de descripteurs ayant une signification mineure et un sur-ajustement des données et la méthode statistique utilisé mais cela n'exclut pas le fait que les indices de Fukui constituent de très bons descripteurs pour les modèles QSAR.

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

### Bibliographie

- [1].Hansch C, Leo A (1995) Exploring QSAR, vol 1: Fundamentals and applications in chemistry and biology. Am ChemSoc, Washington, DC
- [2].Hansch, C.; Zhang, L. QSAR of HIV Inhibitors.*Bioorg. Med. Chem. Lett.* **1992**, *2*, 1165-1169.
- [3].Miyasaka, T., Tanaka, H., Baba, M., Hayakawa, H., Walker, R. T., Balzarini, J., & De Clercq, E. (1989). A novel lead for specific anti-HIV-1 agents: 1-[(2-hydroxyethoxy) methyl]-6-(phenylthio) thymine. *Journal of Medicinal Chemistry*, *32*(12), 2507-2509.
- [4].Tanaka, H., Baba, M., Hayakawa, H., Sakamaki, T., Miyasaka, T., Ubasawa, M., ... & Walker, R. T. (1991). A new class of HIV-1 specific 6-substituted acyclouridine derivatives: synthesis and anti-HIV-1 activity of 5-or 6-substituted analogs of 1-[(2-hydroxyethoxy) methyl]-6-(phenylthio) thymine (HEPT). *Journal of medicinal chemistry*, *34*(1), 349-357.
- [5].Tanaka, H., Baba, M., Ubasawa, M., Takashima, H., Sekiya, K., Nitta, I., ...&Miyasaka, T. (1991). Synthesis and anti-HIV activity of 2-, 3-, and 4-substituted analogs of 1-[(2-hydroxyethoxy) methyl]-6-(phenylthio) thymine (HEPT). *Journal of medicinal chemistry*, *34*(4), 1394-1399.
- [6].Tanaka, H., Takashima, H., Ubasawa, M., Sekiya, K., Nitta, I., Baba, M., ...&Miyasaka, T. (1992). Structure-activity relationships of 1-[(2-hydroxyethoxy) methyl]-6-(phenylthio) thymine analogs: effect of substitutions at the C-6 phenyl ring and at the C-5 position on anti-HIV-1 activity. *Journal of medicinal chemistry*, *35*(2), 337-345.
- [7].Tanaka, H., Takashima, H., Ubasawa, M., Sekiya, K., Nitta, I., Baba, M., ...&Miyasaka, T. (1992). Synthesis and antiviral activity of deoxy analogs of 1-[(2-hydroxyethoxy) methyl]-6-(phenylthio) thymine (HEPT) as potent and selective anti-HIV-1 agents. *Journal of medicinal chemistry*, *35*(25), 4713-4719
- [8].Tanaka, H., Takashima, H., Ubasawa, M., Sekiya, K., Inouye, N., Baba, M., ...&Miyasaka, T. (1995). Synthesis and antiviral activity of 6-benzyl analogs of 1-[(2-hydroxyethoxy) methyl]-5-(phenylthio) thymine (HEPT) as potent and selective anti-HIV-1 agents. *Journal of medicinal chemistry*, *38*(15), 2860-2865.
- [9].Hannongbua, S., Lawtrakul, L., &Limtrakul, J. (1996).Structure-activity correlation study of HIV-1 inhibitors: Electronic and molecular parameters.*Journal of computer-aided molecular design*, *10*(2), 145-152.
- [10].Kireev, D. B., Chrétien, J. R., Grierson, D. S., &Monneret, C. (1997). A 3D QSAR study of a series of HEPT analogues: the influence of conformational mobility on HIV-1 reverse transcriptase inhibition.*Journal of medicinal chemistry*, *40*(26), 4257-4264.
- [11].Pontikis, H.; Benhida, R.; Aubertin, A.-M.; Grierson, D. S.; Monneret, C. Synthesis and Anti-HIV Activity of Novel N-1 Side Chain Modified Analogs of 1-[(2-Hydroxyethoxy)methyl]-6-phenylthio) thymine (HEPT). *J. Med. Chem.* **1997**, *40*, 1845-1854

## Chapitre.IV : Modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT

- [12].Hopkins, A. L., Ren, J., Esnouf, R. M., Willcox, B. E., Jones, E. Y., Ross, C., ... & Stuart, D. I. (1996). Complexes of HIV-1 reverse transcriptase with inhibitors of the HEPT series reveal conformational changes relevant to the design of potent non-nucleoside inhibitors. *Journal of medicinal chemistry*, 39(8), 1589-1600.
- [13].Luco, J. M., & Ferretti, F. H. (1997). QSAR based on multiple linear regression and PLS methods for the anti-HIV activity of a large group of HEPT derivatives. *Journal of chemical information and computer sciences*, 37(2), 392-401.
- [14].Abraham, M. H.; McGowan, J. C. The Use of Characteristic Volumes to Measure Cavity Terms in Reversed Phase Liquid Chromatography. *Chromatographia* **1987**, 23, 243-246.
- [15].Abraham, M. H. Scales of Solute Hydrogen-Bonding: Their Construction and Application to Physicochemical and Biochemical Processes. *Chem. Soc. Rev.* **1993**, 22, 73-83
- [16].Kier, L. B.; Hall, L. H. A Differential Molecular Connectivity Index. *Quant. Struct.-Act. Relat.* **1991**, 10, 134-140.
- [17].Jalali-Heravi, M., & Parastar, F. (2000). Use of artificial neural networks in a QSAR study of anti-HIV activity for a large group of HEPT derivatives. *Journal of chemical information and computer sciences*, 40(1), 147-154.
- [18].Bazoui, H., Zahouily, M., Boulajaaj, S., Sebti, S., & Zakarya, D. (2002). QSAR for anti-HIV activity of HEPT derivatives. *SAR and QSAR in Environmental Research*, 13(6), 567-577.
- [19].Van de Waterbeemd, H.; Kansy, M. Hydrogen-Bonding Capacity and Brain Penetration. *Chimia* **1992**, 46, 299-303.
- [20].Douali, L., Villemin, D., Zyad, A., & Cherqaoui, D. (2004). Artificial neural networks: Non-linear QSAR studies of HEPT derivatives as HIV-1 reverse transcriptase inhibitors. *Molecular diversity*, 8(1), 1.
- [21].Minitab, I. N. C. (2010). MINITAB statistical software. *Minitab Release*, 13.
- [22].Prakasvudhisarn, C., & Lawtrakul, L. (2008). Feature set selection in QSAR of 1-[(2-Hydroxyethoxy) methyl]-6-(phenylthio) thymine (HEPT) analogues by using swarm intelligence. *Monatshefte für Chemie-Chemical Monthly*, 139(3), 197-211.
- [23].Darnag, R., Schmitzer, A., Belmiloud, Y., Villemin, D., Jarid, A., Chait, A., ... & Cherqaoui, D. (2009). QSAR studies of HEPT derivatives using support vector machines. *Molecular Informatics*, 28(6-7), 709-718.
- [24].Inthajak, K., Khamsemanan, N., Nattee, C., Toochinda, P., & Lawtrakul, L. (2017). A prediction approach for anti-HIV activity of HEPT compounds using random forest technique. *Monatshefte für Chemie-Chemical Monthly*, 148(10), 1697-1709.
- [25].Tomar, S., Nadreen, B., Kaleem, A. (2015). QSAR study of anti-HIV drugs of 1-2-[(hydroxyethoxy) methyl]-6-(phenylthio) thymine (HEPT) derivatives by using quantum descriptors. *Journal of Chemical and Pharmaceutical Research*, 7(3), 146-151.

## Conclusion générale

# Conclusion générale

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

L'objectif central de cette thèse était d'étudier l'applicabilité des résultats de calculs de la chimie quantique dans la mise au point de modèles QSAR. De nouveaux descripteurs moléculaires informatifs sont recherchés. Comme les postulats de la chimie quantique impliquent que la fonction d'onde et sa densité électronique associée sont porteurs de toutes les informations d'une molécule, on peut affirmer que la densité électronique est porteuse de toutes informations relative à sa réactivité et par extension de son activité pharmacologique. L'utilisation de la densité électronique et de ses caractéristiques dérivées comme descripteurs moléculaires dans la prévision des activités biologiques n'est donc qu'une conséquence logique, en supposant que les descripteurs moléculaires de la chimie quantique peuvent présenter des corrélations importantes avec l'activité biologique.

Tout d'abord, et en se basant sur la recherche bibliographique, on a pu recenser toutes les données expérimentales concernant la synthèse et le développement de la molécule HEPT et ses dérivés. Ensuite, on a analysé certains dérivés et leurs activités biologiques en fonction de la nature, la position et le nombre des différents groupements pour trouver la relation entre ces derniers et la quantité de médicament utilisée pour l'inhibition de la transcriptase inverse du VIH.

Dans un premier temps, on a étudié la variation des groupements sur les HEPT-O, ces changements ont été élaborés pour différentes séries, la première variation était pour le nombre de la chaîne carbonée et le type du radical  $R_2$  vis-à-vis la concentration efficace. Pour le nombre de carbone, le changement est significatif et il a apparu sur la  $CE_{50}$ , aussi pour l'introduction d'un halogène, la variation est spectaculaire quand on passe d'un Iode à un Fluor.

## Conclusion générale

Pour les HEPT-O substitués en N-1, on a remarqué une forte diminution lors du passage de zéro à 1 puis à 2 atomes de carbones. Cependant, une influence significative a été observée pour les HEPT-O ayant une modification au niveau du radical R<sub>1</sub>, aussi, quand le thiophényl est substitué deux fois, l'efficacité est maximale.

Dans le cas des molécules ayant un soufre en position X (HEPT-S), on note que la substitution par deux méthyles dans les deux positions méta du thiophényl améliore l'activité anti-VIH des HEPT-S.

Des calculs de chimie quantique ont été appliqués sur un ensemble de 83 molécules dérivés de HEPT. Nous nous sommes intéressés à étudier leurs caractéristiques structurales et les propriétés électroniques. En s'appuyant sur les énergies des systèmes, on a pu identifier les différents indices globaux tels que le potentiel chimique, le potentiel d'ionisation, l'affinité électronique, la dureté, la mollesse globale, l'électronégativité et l'indice électrophile. D'autre part, et en utilisant la théorie de l'analyse de population, on a calculé les indices de Fukui. Une étude détaillée sur l'aspect structural et sur les propriétés électroniques pour voir l'effet de groupements attachés sur les cycles sur ces derniers. En premier temps, une analyse sur la position du radical R<sub>1</sub> sur le thiophényl a montré que les positions des deux cycles sont guidées par les interactions et les encombrements stériques entre le radical R<sub>2</sub> de la thymine et le radical R<sub>1</sub> du thiophényl.

Dans une seconde étude, et à fin de bien mener la conception des modèles QSAR, une étude comparative pour les différents modèles antérieurs a été faite pour voir le type des descripteurs et méthodes statistiques déjà utilisés. Le succès de tout modèle QSAR réside dans le choix des descripteurs employés. En général, les descripteurs classiques utilisés pour de telles études sont des descripteurs constitutionnels, topologiques et des fois géométriques. L'emploi des descripteurs quantiques était moins fréquent en QSAR alors qu'ils présentent l'avantage de pouvoir exprimer les propriétés de la densité électronique des systèmes moléculaires.

Nous avons utilisé la méthode MLR (Multiple Linear Régression) pour élaborer des modèles QSAR pour l'inhibition de la transcriptase inverse du VIH en utilisant des informations issues de la chimie quantique. Le modèle le plus fiable, pour prédire l'activité anti-VIH a été développé sur une série de 60 dérivés de HEPT dont les valeurs expérimentales sont déjà connues. Un ensemble de descripteurs moléculaires de nature quantique a été testé avec succès, dont on cite les indices de Fukui. Le meilleur modèle QSAR est capable de décrire environ 81% de la variance de l'activité expérimentale et pourrait être utilisé efficacement pour estimer l'activité de nouvelles molécules pour lesquels les données expérimentales sont indisponibles.

Sur la base de ces études, on peut conclure que la chimie quantique peut apporter une valeur ajoutée au paradigme QSAR. D'une part, avec les théories de la chimie quantique appropriées, on peut clarifier l'action des descripteurs classiques et des définitions appropriées pour des concepts souvent mal définis. D'autre part, on peut importer avec succès de nouvelles informations ou améliorer le comportement prédictif des modèles QSAR classiques. Cependant, la chimie quantique ne génère pas toujours un modèle plus prédictif. De plus, l'intégration d'informations chimiques implique des exigences plus élevées en puissance informatique par rapport aux modèles classiques.

## Conclusion générale

Ainsi, l'utilisation des informations sur les produits chimiques dans les QSAR devrait toujours être considérée par rapport au coût de calcul élevé. Il est conseillé d'utiliser les informations de la chimie quantique lorsque les modèles classiques échouent.

Les perspectives de ce travail sont diverses. D'une part, nous avons l'intention de reprendre les mêmes bases de données pour élaborer d'autres modèles en utilisant les résultats de calculs de chimie quantique du type DFT basés sur des fonctionnelles d'échange-corrélation et des bases d'orbitales atomiques non utilisées dans les travaux de cette thèse. Aussi, d'autres méthodes d'analyse telle que PLS, RNA, et SVM, seront expérimentées. D'autre part, nous projetons d'élaborer des modèles QSAR pour d'autres bases de données et d'autres activités biologiques en utilisant toujours les indices de Fukui.

Ce travail a donné lieu à une publication internationale ainsi qu'à trois présentations publiques, dont les références sont indiquées ci-dessous :

1. Rahmouni, A., Touhami, M., & Benaïssa, T. (2017). Fukui Indices as QSAR Model Descriptors: The Case of the Anti-HIV Activity of 1-2-[(Hydroxyethoxy) Methyl]-6-(Phenylthio) Thymine Derivatives. *International Journal of Chemoinformatics and Chemical Engineering (IJCCE)*, 6(2), 31-44.
2. Touhami M, Rahmouni A. Are the quantum descriptors able to predict anti HIV activity of HEPT derivatives? ” .Workshop: Theory and Computation in Sciences and Bio-Sciences, Tlemcen, Algérie, 13-17 Novembre 2014.
3. Touhami M, Rahmouni A. Modèles QSAR utilisant des descripteurs quantiques pour la prediction de l'activité anti-VI-H des dérives HEPT. Mini workshop de chimie computationnelle, Guelma Algérie, 8-9 Décembre 2014.
4. Touhami M, Rahmouni A, Benaïssa, T. “QSAR for anti-HIV activity of HEPT derivatives using Fukui indexes”. Workshop on Computational Chemistry for Junior Scientists, Sidi belabess, Algérie, 09-11 Decembre 2018.

# *Annexe*

Tableau.1 : Les structures des dérivés HEPT avec les valeurs de concentrations efficaces issues de la synthèse.

N°	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	X	CE <sub>50</sub> (μM)
1	2-Me	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	70,795
2	2-NO <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	141,254
3	2-OMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	19,055
4	3-Me	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	2,570
5	3-Et	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	2,692
6	3-t-Bu	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	12,023
7	3-CF <sub>3</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	44,668
8	3-F	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	3,311
9	3-Cl	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	12,882
10	3-Br	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	5,754
11	3-I	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	10,000
12	3-NO <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	33,884
13	3-OH	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	81,283
14	3-OMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	21,878
15	3,5-Me <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	0,257
16	3,5-Cl <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	1,288
17	3,5-Me <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	0,219
18	3-COOMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	7,943
19	3-COMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	7,244
20	3-CN	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	10,000
21	H	CH <sub>2</sub> CH=CH <sub>2</sub>	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	2,512
22	H	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	0,110
23	H	Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	10,000
24	H	<i>i</i> -Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	0,059
25	3,5-Me <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	0,008
26	3,5-Me <sub>2</sub>	<i>i</i> -Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	0,005

## Annexe

---

27	3,5-Cl <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	0,043
28	H	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	0,120
29	H	Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	3,388
30	H	<i>i</i> -Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	0,063
31	3,5-Me <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	0,013
32	3,5-Me <sub>2</sub>	<i>i</i> -Pr	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	0,003
33	3,5-Cl <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	0,014
34	4-Me	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	218,776
35	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	7,079
36	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	S	0,977
37	H	I	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	3,631
38	H	CH=CH <sub>2</sub>	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	2,042
39	H	CH=CHPh	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	6,026
40	H	CH <sub>2</sub> Ph	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	42,658
41	H	CH=CPh <sub>2</sub>	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	0,851
42	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OMe	O	8,710
43	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OAc	O	6,761
44	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OCOPh	O	7,586
45	H	Me	CH <sub>2</sub> OCH <sub>2</sub> Me	O	0,331
46	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> Cl	O	1,514
47	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> N <sub>3</sub>	O	5,754
48	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> F	O	1,096
49	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> Me	O	3,311
50	H	Me	CH <sub>2</sub> OCH <sub>2</sub> Ph	O	0,087
51	H	Et	CH <sub>2</sub> OCH <sub>2</sub> Me	O	0,019
52	H	Et	CH <sub>2</sub> OCH <sub>2</sub> Me	S	0,026
53	3,5-Me <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> Me	O	0,006
54	3,5-Me <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> Me	S	0,005
55	H	Et	CH <sub>2</sub> OCH <sub>2</sub> Ph	O	0,006
56	3,5-Me <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> Ph	O	0,003

---

## Annexe

57	H	Et	CH <sub>2</sub> OCH <sub>2</sub> Ph	S	0,008
58	3,5-Me <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> Ph	S	0,007
59	H	<i>i</i> -Pr	CH <sub>2</sub> OCH <sub>2</sub> Me	O	0,010
60	H	<i>i</i> -Pr	CH <sub>2</sub> OCH <sub>2</sub> Ph	O	0,003
61	H	<i>i</i> -Pr	CH <sub>2</sub> OCH <sub>2</sub> Me	S	0,013
62	H	<i>i</i> -Pr	CH <sub>2</sub> OCH <sub>2</sub> Ph	S	0,007
63	H	Me	CH <sub>2</sub> OMe	O	2,089
64	H	Me	CH <sub>2</sub> OBu	O	4,677
65	H	Me	Et	O	2,188
66	H	Me	Bu	O	1,202
67	3,5-Cl <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> Me	S	0,013
68	H	Et	CH <sub>2</sub> O- <i>i</i> -Pr	S	0,219
69	H	Et	CH <sub>2</sub> O-c-Hex	S	1,622
70	H	Et	CH <sub>2</sub> OCH <sub>2</sub> -c-Hex	S	0,355
71	H	Et	CH <sub>2</sub> OCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-Me)	S	0,078
72	H	Et	CH <sub>2</sub> OCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> (4-Cl)	S	0,012
73	H	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> Ph	S	0,091
74	3,5-Cl <sub>2</sub>	Et	CH <sub>2</sub> OCH <sub>2</sub> Me	O	0,007
75	H	Et	CH <sub>2</sub> O- <i>i</i> -Pr	O	0,339
76	H	Et	CH <sub>2</sub> O-c-Hex	O	3,981
77	H	Et	CH <sub>2</sub> OCH <sub>2</sub> -c-Hex	O	0,447
78	H	Et	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> Ph	O	0,095
79	H	<i>c</i> -Pr	CH <sub>2</sub> OCH <sub>2</sub> Me	S	0,095
80	H	<i>c</i> -Pr	CH <sub>2</sub> OCH <sub>2</sub> Me	O	0,100
81	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OC <sub>5</sub> H <sub>11-n</sub>	O	34,674
82	2-Cl	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	128,825
83	3-CH <sub>2</sub> OH	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	295,121
84	4-F	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	251,189
85	4-Cl	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	251,189
86	4-NO <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	190,546

## Annexe

---

87	4-CN	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	251,189
88	4-OH	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	275,423
89	4-OMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	251,189
90	4-COMe	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	109,648
91	3-COOH	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	354,813
92	3-CONH <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	309,030
93	H	COOMe	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	660,693
94	H	CONHPh	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	181,970
95	H	SPh	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	208,930
96	H	C≡CH	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	181,970
97	H	C≡C-Ph	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	3,388
98	3-NH <sub>2</sub>	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	251,189
99	H	COCHMe <sub>2</sub>	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	12,023
100	H	COPh	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	12,882
101	H	C≡CMe	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	19,055
102	H	F	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	100,000
103	H	Cl	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	30,200
104	H	Br	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OH	O	19,953
105	H	Me	CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OCH <sub>2</sub> Ph	O	19,953
106	H	Me	H	O	251,189
107	H	Me	Me	O	151,356

---

## Annexe

Tableau.2 : Les énergies des trois systèmes exprimées en e.V des dérivés de HEPT

N°	Système Neutre (e.V)	Système Anionique (e.V)	Système Cationique (e.V)
1	-37858,3342	-37859,0417	-37850,6882
2	-42351,5215	-42355,4669	-42342,8687
3	-39903,0296	-39905,8867	-39895,0027
4	-37858,0893	-37858,2798	-37850,2528
5	-38928,6069	-38928,9062	-38920,6452
7	-45961,7171	-45962,2069	-45953,6629
8	-39489,5546	-39490,0118	-39481,3535
12	-42354,5690	-42355,6030	-42346,4060
13	-38835,9677	-38836,2725	-38827,9734
14	-39905,6418	-39906,4581	-39897,8053
15	-38928,6640	-38928,9307	-38920,7786
16	-61801,0662	-61802,1546	-61792,9577
18	-41921,0294	-41921,7314	-41912,8609
19	-40943,1509	-40943,9999	-40934,9390
20	-38258,4028	-38261,8313	-38251,6547
21	-38859,0636	-38860,7506	-38853,5399
22	-46645,8554	-46646,4622	-46638,0189
23	-47715,8778	-47716,5090	-47708,0494
24	-47715,7988	-47716,4356	-47708,0005
25	-48754,3719	-48754,1324	-48749,1993
26	-49856,0259	-49856,6027	-49848,3445
27	-71618,3798	-71619,4138	-71612,7691
28	-38928,4137	-38928,8082	-38920,3731
30	-38927,9865	-38928,4491	-38919,7963
32	-41040,1628	-41040,8430	-41034,5847
33	-62871,0179	-62867,0207	-62862,9365
34	-37858,5491	-37858,8185	-37850,6392
35	-36788,4343	-36788,7472	-36780,3856
36	-45575,5256	-45576,5323	-45567,6074
38	-37824,8795	-37846,3074	-37804,4910
39	-44113,5622	-44114,8547	-44106,3597
40	-43076,8095	-43077,3374	-43068,8016

## Annexe

---

41	-50401,7605	-50402,5306	-50394,6125
42	-37829,7637	-37830,4439	-37824,1040
44	-46162,0806	-46162,7745	-46154,0945
45	-34712,9942	-34713,6473	-34707,389
46	-46177,0298	-46177,5740	-46169,0029
48	-37442,2389	-37443,1450	-37434,2854
49	-35811,3286	-35811,6089	-35803,3207
50	-39931,3553	-39932,0899	-39925,8316
51	-35811,2742	-35811,6089	-35803,2554
52	-44598,7328	-44599,3015	-44590,9426
53	-37922,8491	-37923,4205	-37917,3255
54	-46738,3449	-46738,8075	-46730,6445
55	-41029,5753	-41029,9699	-41021,6763
56	-43169,8078	-43170,1289	-43162,0394
57	-49817,0394	-49817,6516	-49809,3009
58	-51957,2637	-51957,8052	-51949,6449
59	-36881,2149	-36881,5714	-36873,2451
60	-42098,4141	-42098,9039	-42090,6048
61	-45668,6790	-45669,2694	-45660,9269
62	-50886,9829	-50887,6141	-50879,2743
63	-33671,1533	-33671,9233	-33663,1046
64	-36881,3401	-36881,6204	-36873,3404
65	-31624,2511	-31625,0211	-31616,4228
66	-33764,2931	-33764,5870	-33756,3206
67	-69603,6072	-69610,5485	-69595,5585
68	-45668,8395	-45669,4517	-45661,0847
69	-48845,7131	-48846,2573	-48837,9909
70	-49915,6484	-49916,2443	-49907,8800
71	-50887,1570	-50887,7801	-50879,5273
72	-62323,3867	-62324,1186	-62315,6400
73	-50887,0400	-50887,6114	-50879,3831
74	-60823,7973	-60824,8667	-60815,4302
76	-40058,2491	-40058,5674	-40050,2902
77	-41128,1871	-41128,5463	-41120,1846

---

## Annexe

---

<b>78</b>	-42099,6440	-42100,0412	-42091,8429
<b>82</b>	-49294,6972	-49295,8618	-49286,9124
<b>84</b>	-39489,5492	-39489,9437	-39481,4080
<b>85</b>	-49294,7870	-49295,2795	-49286,6920
<b>90</b>	-40943,1591	-40944,5305	-40935,3308
<b>93</b>	-41974,8099	-41975,3079	-41966,6442
<b>94</b>	-46598,3250	-46598,9263	-46590,5320
<b>95</b>	-52841,9343	-52842,5628	-52834,3944
<b>96</b>	-37790,6874	-37791,2860	-37777,6538
<b>97</b>	-43040,7998	-43042,4623	-43034,6367
<b>100</b>	-45091,1440	-45091,8134	-45083,4599
<b>101</b>	-37822,1993	-37823,6306	-37815,7016
<b>102</b>	-38419,1214	-38419,0126	-38410,8822
<b>103</b>	-48224,4054	-48224,9905	-48216,2071
<b>104</b>	-105744,1360	-105744,680	-105736,0900
<b>107</b>	-30554,1661	-30554,4763	-30546,0875

---

## Annexe

Tableau.3 : Les valeurs des indices globaux des dérives de HEPT exprimés en e.V

<b>N°</b>	<b>PI (e.V)</b>	<b>AE (e.V)</b>	<b><math>\eta</math> (e.V)</b>	<b>S (e.V)</b>	<b><math>\chi</math> (e.V)</b>	<b><math>\mu</math> (e.V)</b>	<b><math>\omega</math> (e.V)</b>
<b>1</b>	7,646	0,707	3,469	0,144	4,177	-4,177	2,514
<b>2</b>	8,653	3,945	2,354	0,212	6,299	-6,299	8,429
<b>3</b>	8,027	2,857	2,585	0,193	5,442	-5,442	5,728
<b>4</b>	7,836	0,190	3,823	0,131	4,013	-4,013	2,107
<b>5</b>	7,962	0,299	3,831	0,131	4,130	-4,130	2,227
<b>7</b>	8,054	0,490	3,782	0,132	4,272	-4,272	2,413
<b>8</b>	8,201	0,457	3,872	0,129	4,329	-4,329	2,420
<b>12</b>	8,163	1,034	3,565	0,140	4,598	-4,598	2,966
<b>13</b>	7,994	0,305	3,845	0,130	4,150	-4,150	2,239
<b>14</b>	7,836	0,816	3,510	0,142	4,326	-4,326	2,666
<b>15</b>	7,885	0,267	3,809	0,131	4,076	-4,076	2,181
<b>16</b>	8,109	1,088	3,510	0,142	4,598	-4,598	3,012
<b>18</b>	8,168	0,702	3,733	0,134	4,435	-4,435	2,635
<b>19</b>	8,212	0,849	3,682	0,136	4,530	-4,530	2,788
<b>20</b>	6,748	3,428	1,660	0,301	5,088	-5,088	7,799
<b>21</b>	5,524	1,687	1,918	0,261	3,605	-3,605	3,388
<b>22</b>	7,836	0,607	3,615	0,138	4,222	-4,222	2,465
<b>23</b>	7,828	0,631	3,599	0,139	4,230	-4,230	2,486
<b>24</b>	7,798	0,637	3,581	0,140	4,218	-4,218	2,484
<b>25</b>	5,173	-0,239	2,706	0,185	2,467	-2,467	1,124
<b>26</b>	7,681	0,577	3,552	0,141	4,129	-4,129	2,400
<b>27</b>	5,611	1,034	2,288	0,218	3,322	-3,322	2,412
<b>28</b>	8,041	0,395	3,823	0,131	4,218	-4,218	2,326
<b>30</b>	8,190	0,463	3,864	0,129	4,326	-4,326	2,422
<b>32</b>	5,578	0,680	2,449	0,204	3,129	-3,129	1,999
<b>33</b>	8,081	-3,997	6,039	0,083	2,042	-2,042	0,345
<b>34</b>	7,910	0,269	3,820	0,131	4,090	-4,090	2,189
<b>35</b>	8,049	0,313	3,868	0,129	4,181	-4,181	2,260
<b>36</b>	7,918	1,007	3,456	0,145	4,462	-4,462	2,881
<b>38</b>	20,388	21,428	-0,520	-0,962	20,908	-20,908	-420,572
<b>39</b>	7,202	1,292	2,955	0,169	4,247	-4,247	3,053
<b>40</b>	8,008	0,528	3,740	0,134	4,268	-4,268	2,435

## Annexe

---

<b>41</b>	7,148	0,770	3,189	0,157	3,959	-3,959	2,458
<b>42</b>	5,660	0,680	2,490	0,201	3,170	-3,170	2,018
<b>44</b>	7,986	0,694	3,646	0,137	4,340	-4,340	2,583
<b>45</b>	5,605	0,653	2,476	0,202	3,129	-3,129	1,977
<b>46</b>	8,027	0,544	3,741	0,134	4,286	-4,286	2,454
<b>48</b>	7,953	0,906	3,524	0,142	4,430	-4,430	2,784
<b>49</b>	8,008	0,280	3,864	0,129	4,144	-4,144	2,222
<b>50</b>	5,524	0,735	2,394	0,209	3,129	-3,129	2,045
<b>51</b>	8,019	0,335	3,842	0,130	4,177	-4,177	2,270
<b>52</b>	7,790	0,569	3,611	0,138	4,179	-4,179	2,419
<b>53</b>	5,524	0,571	2,476	0,202	3,048	-3,048	1,875
<b>54</b>	7,700	0,463	3,619	0,138	4,082	-4,082	2,302
<b>55</b>	7,899	0,395	3,752	0,133	4,147	-4,147	2,291
<b>56</b>	7,768	0,321	3,724	0,134	4,045	-4,045	2,197
<b>57</b>	7,739	0,612	3,563	0,140	4,175	-4,175	2,446
<b>58</b>	7,619	0,541	3,539	0,141	4,080	-4,080	2,352
<b>59</b>	7,970	0,356	3,807	0,131	4,163	-4,163	2,276
<b>60</b>	7,809	0,490	3,660	0,137	4,150	-4,150	2,352
<b>61</b>	7,752	0,590	3,581	0,140	4,171	-4,171	2,430
<b>62</b>	7,709	0,631	3,539	0,141	4,170	-4,170	2,457
<b>63</b>	8,049	0,770	3,639	0,137	4,409	-4,409	2,671
<b>64</b>	8,000	0,280	3,860	0,130	4,140	-4,140	2,220
<b>65</b>	7,828	0,770	3,529	0,142	4,299	-4,299	2,619
<b>66</b>	7,973	0,294	3,839	0,130	4,133	-4,133	2,225
<b>67</b>	8,049	6,941	0,554	0,903	7,495	-7,495	50,725
<b>68</b>	7,755	0,612	3,571	0,140	4,184	-4,184	2,450
<b>69</b>	7,722	0,544	3,589	0,139	4,133	-4,133	2,380
<b>70</b>	7,768	0,596	3,586	0,139	4,182	-4,182	2,439
<b>71</b>	7,630	0,623	3,503	0,143	4,126	-4,126	2,430
<b>72</b>	7,747	0,732	3,507	0,143	4,239	-4,239	2,562
<b>73</b>	7,657	0,571	3,543	0,141	4,114	-4,114	2,389
<b>74</b>	8,367	1,069	3,649	0,137	4,718	-4,718	3,050
<b>76</b>	7,959	0,318	3,820	0,131	4,139	-4,139	2,242
<b>77</b>	8,002	0,359	3,822	0,131	4,181	-4,181	2,287

---

## Annexe

---

<b>78</b>	7,801	0,397	3,702	0,135	4,099	-4,099	2,270
<b>82</b>	7,785	1,165	3,310	0,151	4,475	-4,475	3,025
<b>84</b>	8,141	0,395	3,873	0,129	4,268	-4,268	2,351
<b>85</b>	8,095	0,493	3,801	0,132	4,294	-4,294	2,425
<b>90</b>	7,828	1,371	3,228	0,155	4,600	-4,600	3,277
<b>93</b>	8,166	0,498	3,834	0,130	4,332	-4,332	2,447
<b>94</b>	7,793	0,601	3,596	0,139	4,197	-4,197	2,450
<b>95</b>	7,540	0,629	3,456	0,145	4,084	-4,084	2,414
<b>96</b>	13,034	0,599	6,217	0,080	6,816	-6,816	3,736
<b>97</b>	6,163	1,663	2,250	0,222	3,913	-3,913	3,402
<b>100</b>	7,684	0,669	3,507	0,143	4,177	-4,177	2,487
<b>101</b>	6,498	1,431	2,533	0,197	3,964	-3,964	3,102
<b>102</b>	8,239	-0,109	4,174	0,120	4,065	-4,065	1,980
<b>103</b>	8,198	0,585	3,807	0,131	4,392	-4,392	2,533
<b>104</b>	8,046	0,544	3,751	0,133	4,295	-4,295	2,459
<b>107</b>	8,079	0,310	3,884	0,129	4,194	-4,194	2,265

---

## Annexe

Tableau.4 : Les valeurs des charges nettes sur les atomes choisis pour les dérivés de HEPT.

N°	C <sub>7</sub>	C <sub>8</sub>	S	O	N <sub>12</sub>
1	0,828	0,652	0,267	-0,612	-0,625
2	0,828	0,645	0,286	-0,523	-0,606
3	0,819	0,648	0,154	-0,600	-0,624
4	0,833	0,649	0,302	-0,611	-0,631
5	0,827	0,653	0,308	-0,601	-0,513
7	0,834	0,592	0,457	-0,674	-0,612
8	0,828	0,584	0,436	-0,702	-0,611
12	0,829	0,577	0,448	-0,696	-0,698
13	0,828	0,660	0,267	-0,614	-0,622
14	0,827	0,601	0,458	-0,696	-0,615
15	0,823	0,641	0,273	-0,604	-0,606
16	0,828	0,669	0,310	-0,595	-0,629
18	0,829	0,591	0,454	-0,681	-0,619
19	0,827	0,591	0,453	-0,682	-0,619
20	0,829	0,654	0,385	-0,606	-0,618
21	0,445	0,591	0,453	-0,680	-0,617
22	0,295	0,594	0,455	-0,685	-0,619
23	0,295	0,657	0,266	-0,612	-0,607
24	0,295	0,649	0,310	-0,607	-0,628
25	-0,029	0,660	0,295	-0,617	-0,616
26	0,297	0,654	0,320	-0,593	-0,627
27	-0,232	0,654	0,232	-0,594	-0,628
28	0,828	0,654	0,316	-0,591	-0,628
29	0,828	0,656	0,325	-0,591	-0,629
30	0,826	0,654	0,303	-0,598	-0,629
32	0,444	0,654	0,307	-0,600	-0,629
33	0,830	0,210	0,259	-0,607	-0,655
34	0,828	0,655	0,332	-0,604	-0,630

## Annexe

---

<b>35</b>	0,828	0,653	0,338	-0,597	-0,628
<b>36</b>	0,284	0,674	0,344	-0,576	-0,635
<b>38</b>	0,826	0,633	0,340	-0,576	-0,629
<b>39</b>	0,821	0,650	0,294	-0,604	-0,629
<b>40</b>	0,831	0,656	0,310	-0,630	-0,512
<b>41</b>	0,828	0,656	0,328	-0,592	-0,629
<b>42</b>	0,444	0,661	0,357	-0,590	-0,631
<b>44</b>	0,828	0,653	0,307	-0,602	-0,628
<b>45</b>	0,441	0,656	0,292	-0,603	-0,627
<b>46</b>	0,836	0,655	0,302	-0,597	-0,627
<b>48</b>	0,828	0,660	0,309	-0,595	-0,605
<b>49</b>	0,828	0,654	0,309	-0,600	-0,629
<b>50</b>	0,446	0,669	0,302	-0,579	-0,629
<b>51</b>	0,828	0,656	0,297	-0,595	-0,607
<b>52</b>	0,293	0,659	0,305	-0,597	-0,605
<b>53</b>	0,441	0,650	0,290	-0,605	-0,628
<b>54</b>	0,234	0,653	0,309	-0,594	-0,606
<b>55</b>	0,828	0,656	0,296	-0,600	-0,629
<b>56</b>	0,827	0,655	0,309	-0,590	-0,606
<b>57</b>	0,294	0,656	0,299	-0,599	-0,629
<b>58</b>	0,292	0,655	0,310	-0,592	-0,606
<b>59</b>	0,827	0,657	0,300	-0,598	-0,629
<b>60</b>	0,816	0,654	0,306	-0,602	-0,629
<b>61</b>	0,293	0,656	0,306	-0,593	-0,607
<b>62</b>	0,294	0,656	0,306	-0,592	-0,607
<b>63</b>	0,828	0,656	0,299	-0,599	-0,629
<b>64</b>	0,828	0,654	0,317	-0,596	-0,628
<b>65</b>	0,819	0,653	0,307	-0,599	-0,629
<b>66</b>	0,820	0,654	0,311	-0,598	-0,629
<b>67</b>	0,260	0,654	0,311	-0,599	-0,628

---

## Annexe

---

<b>68</b>	0,294	0,654	0,306	-0,598	-0,629
<b>69</b>	0,294	0,653	0,307	-0,601	-0,629
<b>70</b>	0,293	0,660	0,308	-0,595	-0,606
<b>71</b>	0,293	0,659	0,297	-0,604	-0,627
<b>72</b>	0,293	0,653	0,306	-0,601	-0,629
<b>73</b>	0,293	0,653	0,310	-0,597	-0,629
<b>74</b>	0,830	0,656	0,303	-0,594	-0,607
<b>76</b>	0,828	0,656	0,306	-0,592	-0,606
<b>77</b>	0,827	0,656	0,298	-0,600	-0,629
<b>78</b>	0,827	0,660	0,298	-0,603	-0,628
<b>82</b>	0,828	0,660	0,306	-0,596	-0,606
<b>84</b>	0,828	0,653	0,307	-0,601	-0,629
<b>85</b>	0,828	0,650	0,290	-0,606	-0,627
<b>90</b>	0,828	0,656	0,303	-0,595	-0,607
<b>93</b>	0,831	0,656	0,302	-0,593	-0,607
<b>94</b>	0,816	0,656	0,297	-0,594	-0,607
<b>95</b>	0,828	0,656	0,296	-0,602	-0,627
<b>96</b>	0,829	0,656	0,295	-0,600	-0,629
<b>97</b>	0,833	0,649	0,327	-0,579	-0,631
<b>100</b>	0,830	0,664	0,329	-0,582	-0,630
<b>101</b>	0,833	0,604	0,316	-0,572	-0,633
<b>102</b>	0,827	0,350	0,212	-0,263	-0,310
<b>103</b>	0,415	0,670	0,369	-0,563	-0,640
<b>104</b>	0,828	0,667	0,369	-0,587	-0,636
<b>107</b>	0,820	0,279	0,293	-0,589	-0,671

---

## Annexe

Tableau.5 : Les valeurs des indices de Fukui des dérivées de HEPT.

N°	$\Delta f_{C7}$	$\Delta f_{C8}$	$\Delta f_S$	$\Delta f_O$	$\Delta f_{N12}$
1	0,817	0,643	0,304	-0,581	-0,623
2	0,838	0,654	0,107	-0,535	-0,598
3	0,830	0,676	-0,133	-0,663	-0,621
4	0,844	0,658	0,206	-0,718	-0,641
5	0,838	0,652	0,077	-0,694	-0,639
7	0,849	0,667	0,125	-0,727	-0,631
8	0,841	0,657	0,097	-0,722	-0,637
12	0,835	0,690	0,120	-0,691	-0,624
13	0,839	0,652	0,099	-0,686	-0,638
14	0,831	0,646	0,041	-0,650	-0,520
15	0,829	0,651	0,080	-0,691	-0,638
16	0,839	0,676	-0,056	-0,699	-0,637
18	0,843	0,669	0,093	-0,698	-0,632
19	0,839	0,666	0,089	-0,689	-0,633
20	0,824	0,653	0,328	-0,572	-0,631
21	0,407	0,587	0,453	-0,703	-0,64
22	0,308	0,659	0,139	-0,631	-0,654
23	0,307	0,644	0,133	-0,628	-0,653
24	0,309	0,651	0,137	-0,63	-0,649
25	2,816	3,569	0,444	-0,823	-0,675
26	0,310	0,649	0,135	-0,631	-0,649
27	-0,263	0,573	0,446	-0,779	-0,856
28	0,840	0,649	0,053	-0,679	-0,643
29	0,839	0,649	0,055	-0,675	-0,643
30	0,841	0,656	0,04	-0,715	-0,631
32	0,304	0,592	0,464	-0,733	-0,75
33	0,837	0,672	0,159	-0,672	-0,635
34	0,839	0,654	0,086	-0,694	-0,640
35	0,841	0,656	0,083	-0,698	-0,64
36	0,307	0,632	0,144	-0,662	-0,646
38	0,870	0,690	0,219	-0,567	-0,690
39	0,817	0,212	0,298	-0,621	-0,673

## Annexe

---

<b>40</b>	0,839	0,646	0,104	-0,657	-0,641
<b>41</b>	0,830	0,661	0,249	-0,589	-0,634
<b>42</b>	0,300	0,589	0,460	-0,723	-0,752
<b>44</b>	0,837	0,659	0,111	-0,675	-0,635
<b>45</b>	0,293	0,833	0,460	-0,724	-0,753
<b>46</b>	0,847	0,657	0,149	-0,709	-0,632
<b>48</b>	0,823	0,637	0,327	-0,597	-0,641
<b>49</b>	0,841	0,655	0,086	-0,699	-0,639
<b>50</b>	0,296	0,589	0,460	-0,724	-0,751
<b>51</b>	0,840	0,651	0,062	-0,682	-0,651
<b>52</b>	0,307	0,647	0,143	-0,633	-0,652
<b>53</b>	0,299	0,594	0,461	-0,725	-0,751
<b>54</b>	0,249	0,649	0,100	-0,669	-0,645
<b>55</b>	0,839	0,648	0,125	-0,670	-0,638
<b>56</b>	0,835	0,646	0,115	-0,670	-0,636
<b>57</b>	0,307	0,644	0,177	-0,626	-0,646
<b>58</b>	0,301	0,645	0,158	-0,631	-0,647
<b>59</b>	0,840	0,655	0,060	-0,681	-0,640
<b>60</b>	0,827	0,656	0,118	-0,706	-0,627
<b>61</b>	0,309	0,651	0,143	-0,632	-0,65
<b>62</b>	0,308	0,633	0,175	-0,626	-0,647
<b>63</b>	0,844	0,669	0,074	-0,703	-0,639
<b>64</b>	0,840	0,654	0,083	-0,697	-0,640
<b>65</b>	0,828	0,673	0,110	-0,714	-0,634
<b>66</b>	0,835	0,656	0,104	-0,71	-0,636
<b>67</b>	0,256	0,662	0,195	-0,645	-0,686
<b>68</b>	0,303	0,644	0,111	-0,626	-0,653
<b>69</b>	0,310	0,647	0,159	-0,636	-0,650
<b>70</b>	0,305	0,646	0,138	-0,630	-0,652
<b>71</b>	0,344	0,824	0,189	-0,620	-0,645
<b>72</b>	0,303	0,644	0,174	-0,618	-0,644
<b>73</b>	0,296	0,642	0,184	-0,626	-0,649
<b>74</b>	0,846	0,666	0,083	-0,67	-0,642
<b>76</b>	0,842	0,652	0,070	-0,687	-0,637

---

## Annexe

---

<b>77</b>	0,840	0,651	0,063	-0,682	-0,642
<b>78</b>	0,834	0,646	0,124	-0,654	-0,64
<b>82</b>	0,839	0,682	0,140	-0,727	-0,639
<b>84</b>	0,841	0,654	0,095	-0,697	-0,64
<b>85</b>	0,840	0,655	0,104	-0,688	-0,639
<b>90</b>	0,835	0,674	0,118	-0,683	-0,632
<b>93</b>	0,839	0,673	0,084	-0,669	-0,642
<b>94</b>	0,804	-0,115	0,148	-0,626	-0,723
<b>95</b>	0,832	0,632	0,281	-0,603	-0,644
<b>96</b>	0,826	0,671	0,374	-0,579	-0,65
<b>97</b>	0,824	0,653	0,352	-0,554	-0,642
<b>100</b>	0,898	0,575	-0,051	-0,721	-0,402
<b>101</b>	0,827	0,651	0,377	-0,58	-0,642
<b>102</b>	0,841	0,604	0,103	-0,671	-0,643
<b>103</b>	0,013	0,067	-0,119	-0,032	-0,002
<b>104</b>	0,837	0,633	0,111	-0,648	-0,641
<b>107</b>	0,831	0,656	0,096	-0,708	-0,638

---