

# THÈSE

## En vue de l'obtention du Diplôme de Doctorat LMD

*Présenté par : SOUALMIA Fatima*

### *Intitulé*

*Analyse multivariée des relations structures-activités dans des nouveaux  
composés organiques à utilisation médicinale*

*Faculté : Chimie*

*Département : Chimie Physique*

*Domaine : Sciences de la matière*

*Filière : Chimie*

*Intitulé de la Formation : Chimie physique et théorique appliquée*

*Devant le Jury Composé de :*

<i>Membres de Jury</i>	<i>Grade</i>	<i>Qualité</i>	<i>Domiciliation</i>
<i>Mr. DAAOU Mortada</i>	<i>Professeur</i>	<i>Président</i>	<i>USTO MB</i>
<i>Mr. BELAIDI Salah</i>	<i>Professeur</i>	<i>Encadrant</i>	<i>Université de Biskra</i>
<i>Mr. TCHOUAR Noureddine</i>	<i>Professeur</i>	<i>Co-Encadrant</i>	<i>USTO MB</i>
<i>Mme. SEBBA Fatima Zohra</i>	<i>Professeur</i>	<i>Examinatrice</i>	<i>Université Oran 1 ABB</i>
<i>Mr. ADIL Ali-Othman</i>	<i>Professeur</i>	<i>Examineur</i>	<i>USTO MB</i>
<i>Mr. MEHDID Mohammed Amine</i>	<i>MCA</i>	<i>Examineur</i>	<i>USTO MB</i>

# *Remerciements*

*En tout premier lieu, je remercie le bon dieu, tout puissant, de m'avoir donné la force pour survivre, ainsi que l'audace pour dépasser toutes les difficultés.*

*Cette thèse a été réalisée essentiellement au Laboratoire de Modélisation et Optimisation des Systèmes Industriels (LAMOSI) à l'USTO-MB en collaboration avec le laboratoire de la chimie moléculaire et environnement (LMCE) à l'université de Biskra.*

*Je tiens tout d'abord à exprimer ma profonde reconnaissance et ma gratitude à Monsieur le **Pr. BELAIDI Salah** directeur de thèse, de m'avoir accueilli au sein de son équipe, de m'avoir permis d'effectuer ce travail, pour son encadrement, ses conseils constructifs, ses encouragements continus et pour la confiance qu'il m'a accordé.*

*J'exprime mes sincères remerciements et ma profonde reconnaissance à Monsieur le co-encadreur de thèse **Pr. TCHOUAR Noureddine** de m'avoir accueilli au sein du Laboratoire pour la réalisation de ce travail, pour m'avoir confié la réalisation de ce projet, pour son encadrement ainsi que ses conseils et remarques pour l'amélioration de ce manuscrit.*

*J'exprime toute ma gratitude au **Pr. DAAOU Mortada** pour m'avoir fait l'honneur de présider le jury malgré tous ses engagements, ainsi que pour ses conseils et ses recommandations constructives.*

*Je tiens à adresser aussi mes vifs remerciements et à exprimer ma profonde gratitude au **Pr. SEBBA Fatima Zohra**, pour avoir accepté d'examiner et de juger ce travail.*

*Je remercie vivement le **Pr. ADIL Ali-Othman** d'avoir aimablement accepté d'examiner mon travail de thèse et aussi de faire partie du jury malgré tout ses engagements.*

*Mes remerciements vont au **Dr. MEHDID Mohammed Amine**, d'avoir bien voulu examiner mon travail et de faire partie de ce jury ainsi que pour ses remarques constructives.*

*Merci **Mr. BENTAYEB Kamel** d'avoir tout fait pour assurer que ce travail se déroule dans les meilleures conditions.*

*Mes plus chaleureux remerciements s'adressent également, à toutes les personnes du laboratoire de LAMOSI. Je les remercie tous d'avoir assuré une ambiance particulièrement favorable pour mener à bien mon travail.*

*Sans oublier de remercier également mes amies, pour leur soutien, leur gentillesse et leur sympathie en leur souhaitant bon courage et bonne continuation.*

*Je tiens aussi à remercier vivement les membres de ma famille qui m'ont toujours soutenue et encouragée et pour leurs aides précieuses durant toutes ces années.*

# SOMMAIRE

REMERCIEMENTS

SOMMAIRE

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES PRINCIPALES ABREVIATIONS

INTRODUCTION GENERALE	1
Références bibliographiques	4

## \*\*\*Rappel bibliographique\*\*\*

### Chapitre I : Généralités sur les pyrazines et l'enzyme de télomérase

I.1. Les pyrazines	7
I.1.1. Réactivité	7
I.1.2. Réactions	7
I.1.3. Utilisations et propriétés pharmacologiques	9
I.2. Enzyme de télomérase	
I.2.1. Les télomères	10
I.2.1.1. Structure et description	11
I.2.1.2. Les protéines associées aux télomères humains : complexe Shelterin	12
I.2.2. La télomérase	13
I.2.2.1. Structure de la télomérase	13
I.2.2.2. Techniques de mesure de l'activité télomérase	15
I.2.2.3. Stratégies thérapeutiques ciblant la télomérase	15
Références bibliographiques	18

## \*\*\* Matériels et méthodes\*\*\*

### Chapitre II : Méthodes de la Modélisation moléculaire

II.1. Introduction	24
II.2. Méthodes de la mécanique quantique	25
II.2.1. Base de la mécanique quantique	25

II.2.2. Approximation de Born-Oppenheimer	26
II.2.3. Fonction d'onde électronique	27
II.2.4. Méthode quantique : <i>Ab initio</i>	27
II.2.4.1. Approximation de Hartree-Fock	29
II.2.4.2. Méthode post-Hartree-Fock	30
II.2.5. Théorie de la fonctionnelle de la densité (DFT)	32
II.2.6. Méthodes semi-empiriques	33
<b>II.3. Mécanique moléculaire</b>	<b>35</b>
II.3.1. Énergie d'interaction entre atomes liés	36
II.3.2. Énergie d'interaction entre atomes non liés	38
II.3.3. Quelques exemples de champs de force utilisés dans la mécanique moléculaire	39
<b>II.4. Dynamique moléculaire</b>	<b>41</b>
II.4.1. Principe de la dynamique moléculaire(DM)	41
<b>Références bibliographiques</b>	<b>43</b>

### **Chapitre III : Etude de la relation structure-propriété/activité**

<b>III.1. Introduction</b>	<b>47</b>
<b>III.2. Historique</b>	<b>47</b>
<b>III.3. Outils et techniques de QSAR</b>	<b>49</b>
III.3.1. Paramètres biologiques	49
III.3.2. Descripteurs moléculaires	49
III.3.2.1. Volume et surface moléculaire	50
III.3.2.2. Polarisabilité	51
III.3.2.3. réfractivité molaire	52
III.3.2.4. Energie d'hydratation	52
III.3.2.5. Coefficient de partition (Log P)	54
III.3.3. Modèles statistiques	56
<b>III.4. Régression linéaire multiple</b>	<b>57</b>
III.4.1. Description de la méthode	57
III.4.2. Confiance accordée aux résultats	59
III.4.3. Test de la signification globale de la régression	59

III.4.4. Validations des modèles	61
<b>III.5. Optimisation multi-paramètres (MPO)</b>	<b>62</b>
III.5.1. Les règles de lipinski	63
III.5.2. Les règles de Veber	64
III.5.3. Efficacité de ligand « LE »	64
III.5.4. Efficacité lipophylique de ligand « LipE »	65
<b>Références bibliographiques</b>	<b>66</b>

**\*\*\* Résultats et discussions\*\*\***

**Chapitre IV : Etude des propriétés structurales, électroniques et énergétiques de la pyrazine et ses dérivés**

<b>IV.1. Introduction</b>	<b>71</b>
<b>IV.2. Méthodes de calcul utilisées</b>	<b>71</b>
<b>IV.3. Etude des propriétés structurales et électroniques de la pyrazine</b>	<b>72</b>
<b>IV.4. Etude de l'effet de la substitution sur le noyau de base de la pyrazine</b>	<b>75</b>
IV.4.1. Effet d'un substituant électrodonneur (méthyle et éthyle)	76
IV.4.2. Effet d'un substituant électro-attracteurs(bromure et fluore)	77
<b>IV.5. Conclusion</b>	<b>79</b>
<b>Références bibliographiques</b>	<b>81</b>

**Chapitre V : Etude qualitative et quantitative des relations structure-activité des dérivés de la pyrazine**

<b>V.1. Introduction</b>	<b>84</b>
<b>V.2. Etude qualitative des relations structure/activité (SAR) dans les dérivés de la Pyrazine</b>	<b>85</b>
V.2.1. Sélection d'une série bioactives des pyrazines	85
V.2.1.1. Résultats et discussions	86
<b>V.3. Paramètres MPO</b>	<b>89</b>
V.3.1. Résultats et discussions	89
<b>V.4. Etude quantitatives des relations structure/activité (QSAR) dans les dérivés de la pyrazine</b>	<b>90</b>
V.4.1. Données expérimentales	90
V.4.2. Sélection des descripteurs et méthodes de calcul	92
V.4.3. Résultats et discussions	94

<b>V.5. Conclusion</b>	<b>98</b>
<b>Références bibliographiques</b>	<b>99</b>
<b>CONCLUSION GENERALE</b>	<b>100</b>
<b>ANNEXES</b>	<b>103</b>

## *Liste des tableaux*

---

### **Chapitre III**

**Tableau III.1 :** Types de données biologiques utilisées dans l'analyse QSAR.

**Tableau III.2 :** Table d'analyse de variance.

### **Chapitre IV**

**Tableau IV.1 :** Résultats de calcul des longueurs des liaisons, les angles de valence de la pyrazine.

**Tableau IV.2 :** Résultats de calcul des charges Mulliken de la pyrazine.

**Tableau IV.3 :** Résultats de calcul des charges NBO de la pyrazine.

**Tableau IV.4:** Résultats de calcul des paramètres énergétique et le moment dipolaire des dérivés de la pyrazine.

**Tableau IV.5:** Valeurs des charges de NBO de la pyrazine et ses dérivés (série 1).

**Tableau IV.6:** Valeurs des charges de NBO de la pyrazine et ses dérivés (série 2).

### **Chapitre V**

**Tableau V.1:** Propriétés physicochimiques des dérivés de la pyrazine.

**Tableau V.2:** Activités pharmacologiques et propriétés des méthodes MPO pour « drug-likeness » des dérivés de la pyrazine.

**Tableau V.3 :** Structures chimiques et les activités biologiques expérimentales des molécules utilisées dans l'étude QSAR.

**Tableau V.4 :** Descripteurs moléculaires utilisés dans l'étude QSAR.

**Tableau V.5 :** Paramètres de validation croisée.

**Tableau V.6 :** Valeurs des activités prévues et résiduelles de l'ensemble de dérivés de la pyrazine.

## *Liste des figures*

---

### **Chapitre I**

**Figure I.1:** Formule générale des pyrazines.

**Figure I.2:** Effet de résonance du pyrazine.

**Figure I.3:** Structure chimique de l'AEZS-131.

**Figure I.4:** Structure chimique de ethyl 2-(8-([5-(diethylamino)pentan-2-yl]amino)-2-methyl-3-phenylpyrido[2,3-b]pyrazin-6-yl)acetate.

**Figure I.5:** Structure chimique de 2,3-dihydro-6,7-dichloro-pyrido[2,3-b]pyrazine-8-oxyde.

**Figure I.6 :** Représentation schématique de la séquence d'ADN télomérique humaine.

**Figure I.7:** Représentation schématique de la télomérase humaine.

**Figure I.8:** Représentation schématique des stratégies d'inhibition de la télomérase.

**Figure I.9 :** structure chimique de pyrazinamide.

### **Chapitre II**

**Figure II.1 :** Énergie d'élongation entre deux atomes liés.

**Figure II.2:** Énergie de déformation des angles de valence.

**Figure II.3 :** Énergies de torsion.

### **Chapitre III**

**Figure III.1 :** Atome entouré par une grille.

**Figure III.2:** Atomes A, B et la zone de recouvrement X.

**Figure III.3 :** La liaison hydrogène.

**Figure III.4 :** Polarité des différents milieux cellulaires.

**Figure III.5 :** Etapes du processus de découverte des médicaments.

## Chapitre IV

**Figure IV.1 :** Conformation 3D de la pyrazine (GaussView 5.08) et 2D (MarvinSketch 15.8.31).

**Figure IV.2 :** Représentation graphique des charges NBO et Mulliken des atomes de la pyrazine.

**Figure IV.3 :** Représentation des surfaces potentielles électrostatiques moléculaires de la pyrazine en 2D et 3D.

**Figure IV.4:** Les deux séries des pyrazines substitués.

**Figure IV. 5:** Orbitales moléculaires HOMO et LUMO du composé B3 (2,3-dibromo-pyrazine).

## Chapitre V

**Figure V.1:** Modèle d'étude quantitative de relations structure activité (QSAR).

**Figure V.2 :** Structure chimique des dérivés de la pyrazine.

**Figure V.3 :** Liaisons hydrogènes accepteurs(HBB) du composé 18.

**Figure V.4 :** Corrélation entre les activités biologiques expérimentales et prédites pour les modèles 1 et 2.

**Figure V.5 :** Corrélation entre les valeurs résiduelles et valeurs expérimentales de  $\log(1/IC_{HEPG2})$  et  $\log(1/IC_{SW1116})$ .

## *Liste des abréviations*

---

**ADME** : Absorption, Distribution, Métabolisme et Excrétion  
**AM1**: Austrin Model 1.  
**AMBER** :Assisted Model Building with Energy Refinement  
**B3LYP** : Trois paramètres hybrides de becke  
**DFT** :Density Functional Theory  
**HBA**: Hydrogen Bond Acidity  
**HBB**: Hydrogen Bond Basicity  
**HE**: Hydration Energy  
**HEPG2** : Lignée cellulaire d'hépatoblastome  
**HF**: Hartree-Fock  
**HOMO** :Highest Occupied Molecular Orbital  
**HSAB** :Hard and Soft, Acids and Bases theory  
**hTERT** : Human Telomerase Reverse Transcriptase  
**hTR** : Human Telomerase RNA  
**IC 50** : Concentration Inhibitrice de 50% de l'activité  
**LOO** :Leave-One-Out  
**LUMO** :Lowest Unoccupied Molecular Orbital  
**MLR** :Multiple Linear Regression  
**MM** : Mécanique Moléculaire  
**MM2**:Molecular Mechanic 2  
**MP2**:Møller-Plesset level 2  
**MR**: Molar refractivity  
**MV**: Molar Volume  
**MW** : Molar Weight  
**NBO** : Analyse de liaison naturelle  
**nrotb** :Number of Rotatable Bonds  
**PE** : Prediction Error  
**Pol** : Polarisabilité  
**PRESS**: Predictive Residual Sum of Squares  
**PSA** :Polar Surface Area  
**QSAR**: Quantitative Structure-Activity Relationships  
**SAG**: Surface Area Grid  
**SAR** : Relation Structure Propriété  
**SCF** :Self-Consistent-Field procedure  
**SW1116**:Lignée cellulaire d'adénocarcinomes coliques  
**ΔE** : gap énergétique  
**μ** : Moment dipolaire

# *Introduction Générale*

La chimie informatique, appelée aussi chimie computationnelle (traduction du terme anglais : Computational Chemistry), est une branche de la chimie qui utilise les lois de la mécanique quantique et la mécanique classique ; exploitées dans des programmes informatiques spécifiques afin de calculer à l'aide des ordinateurs, les structures et les propriétés de substances chimiques tels que les molécules libres, les solides, les agrégats atomiques (ou clusters), les surfaces, etc., en appliquant autant que possible ces programmes à des problèmes chimiques réels.

En chimie informatique, les chimistes, les physiciens et les mathématiciens développent des algorithmes et des codes afin de prédire des propriétés atomiques, moléculaires ou autres, et éventuellement des chemins de réactions chimiques. Il existe deux aspects distincts de la chimie numérique :

- les études menées pour trouver un point de départ pour une synthèse de laboratoire, c.à.d. de faire un tri avant de passer au stade expérimentale, ou pour expliciter des résultats expérimentaux, comme la position et la source des pics spectroscopiques.
- les études menées pour prédire la possibilité d'existence pour des systèmes inconnus ou d'explorer des mécanismes réactionnels qui ne peuvent être étudiés par des moyens expérimentaux.
- Les études de corrélation entre la structure chimique et l'activité biologique (méthodes QSAR).

La modélisation moléculaire est un terme général qui englobe différentes techniques de graphisme moléculaire et de chimie computationnelle permettant de dessiner, afficher, simuler, analyser, calculer et stocker les propriétés des molécules.

Le développement des outils informatiques a permis de mettre au point des techniques de calculs de plus en plus poussées, rendant ainsi possible l'étude des systèmes de plus en plus complexes [1]. Actuellement, les applications de la chimie computationnelle couvrent en plus des branches de la chimie d'autres disciplines de sciences et de l'environnement. Son but n'est pas seulement l'obtention, par le calcul, de résultats appropriés aux problèmes rencontrés en

chimie telles les propriétés thermodynamiques, les grandeurs géométriques,...[2], mais elle s'intéresse à établir des relations mathématiques entre les propriétés physicochimiques des composés et leurs activités biologiques, pharmacologiques et environnementales [3-5].

Les hétérocycles actifs jouent un rôle thérapeutique très important. En raison de leur intérêt pharmaceutique, la chimie des hétérocycles est en développement croissant [6,7].

- Dans cette thèse, nous rapportons une étude computationnelle de la pyrazine et une série de ses dérivés ; donc notre objectif est d'appliquer l'ensemble des méthodes fournies par la chimie informatique pour déterminer les propriétés structurales et électroniques du noyau de la pyrazine ; l'effet de substitution sur ces propriétés est également discuté et enfin des corrélations entre la structure chimique et l'activité biologique d'une série de molécules bioactives.

Le premier chapitre est scindé en deux parties : dans la première partie, nous allons étudier les propriétés physicochimiques et la réactivité chimique des pyrazines. Dans la deuxième partie, nous allons présenter des généralités sur l'enzyme de télomerase.

Dans le deuxième chapitre, nous présentons la base théorique des différentes méthodes de la modélisation moléculaire appliquées à fin de réaliser ce travail (la mécanique quantique (MQ) et la mécanique moléculaire (MM)).

Le troisième chapitre, nous présenterons des généralités sur la méthode de corrélation quantitative (QSAR).

Le quatrième chapitre comporte une étude des propriétés structurales, électroniques et énergétique sur la pyrazine et ses dérivés. Aussi dans ce chapitre, nous présentons les résultats d'une étude comparative sur les méthodes utilisées dans le calcul de la pyrazine et les données expérimentales.

Le cinquième chapitre est une étude qualitative et quantitative des relations structure-activité des pyrazines. Nous présenterons la méthodologie de l'étude

QSAR et la description de la méthode statistique utilisée (MLR), ainsi que le développement des modèles statistiques de deux activités biologiques: l'activité antiproliférative contre HEPG2 et SW1116 (travail publié dans: *Journal of bionanoscience*, volume 11(6), 2017 ;584-591).

**Références bibliographiques**

- [1] Rivail J. L., "Eléments de chimie quantique à l'usage des chimistes", *CNRS Edition*, Paris(1999).
- [2] Mc Quarrie D. A., Simon J. D., "Chimie Physique Approche Moléculaire", Dunod, Paris(2000).
- [3] Smith J., Karplus M., Empirical force field study of geometries and conformational transitions of some organic molecules, *Journal of the American Chemical Society*, 114, 801(1992).
- [4] Takigawa T., Ashida T., Sasada Y., Kakudo M., The crystal structures of L-tryptophan hydrochloride and hydrobromide, *Bull. Chem. Soc. Jpn.* 39, 2369(1966).
- [5] Buyukuslu H., M. Akdogan, Yildirim G., Parlak C. , *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. 75 ,1362(2010).
- [6] Belaidi S., Dibi A., Omari M., A Conformational Exploration of Dissymmetric. Macrolides Antibiotics, *Turk. J. Chem*, 26, 491(2002).
- [7] Van Bambeke F. , Verhaegen J. , Tyteca D, Auckenthaler R., Tulken P. M., Erythromycine et néomacrolides actuels, usages cliniques et perspectives. , *Louv. Med.* 119 ,259(2000).

# *Rappel bibliographique*

*Chapitre I :*  
*Généralités sur les*  
*pyrazines et l'enzyme*  
*de télomérase*

## I.1. Les pyrazines

Les pyrazines 1,4 (diazines) sont des hétérocycles azotés et aromatiques dont leur structure générale est présentée sur la figure I.1

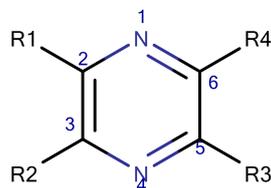


Figure I.1: Formule générale des pyrazines.

### I.1.1. Réactivité

Puisque la pyrazine est un hétérocycle aromatique, elle doit donc subir une substitution aromatique électrophile. L'effet de résonance de la pyrazine est présenté dans la figure I.2.

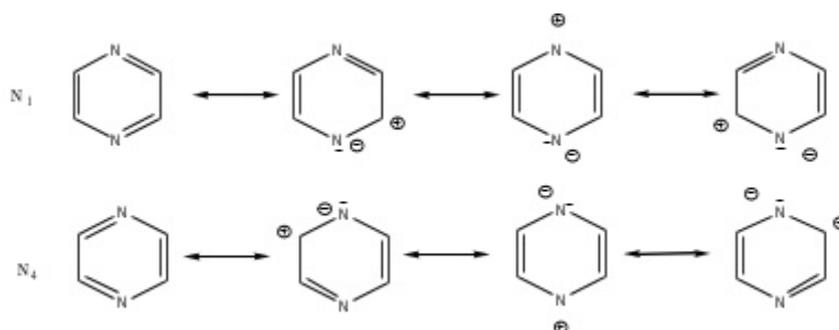


Figure I.2: Effet de résonance de la pyrazine.

La présence des deux atomes d'azote dans le cycle influence à la fois sur la basicité et les propriétés aromatiques des pyrazines.

### I.1.2. propriétés chimique

#### a. Formation des sels

La pyrazine forme des sels monoquaternaires qui sont instables (préparés à une température inférieure à 40°C) avec des halogénures d'alkyle. La pyrazine réagit également avec des acides halogénés pour former un sel de pyrazinium comme indiqué dans le schéma I.1, par lequel la 2,5-diméthylpyrazine (2) réagit avec l'acide iodo ou bromo-acétique pour donner le sel de 1,2,5-triméthylpyrazinium (3) [1,2].

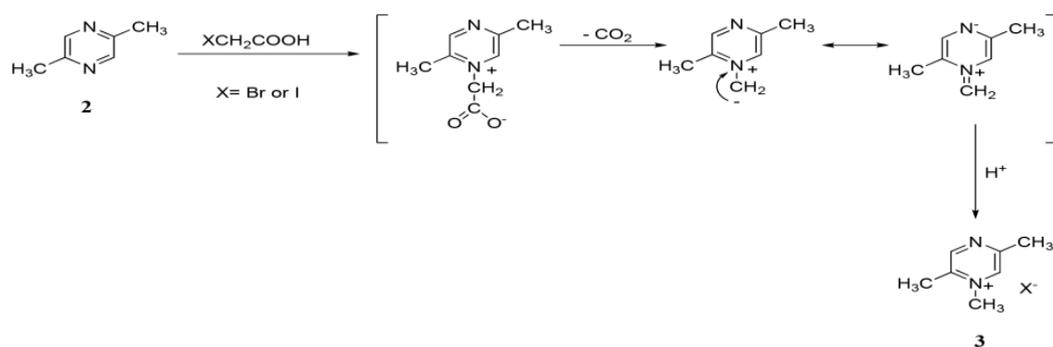


Schéma I.1

### b. Oxydation

La pyrazine subit un processus d'oxydation, par lequel elle décolore la solution alcaline du permanganate [3,4] mais la pyrazine est moins stable que la pyridine à l'action des agents oxydants. Mono (4) et di-N-oxydes (5) [5,6] ont été obtenus lorsque l'alkyl pyrazine réagit avec le peroxyde d'hydrogène comme indiqué dans le schéma I.2.

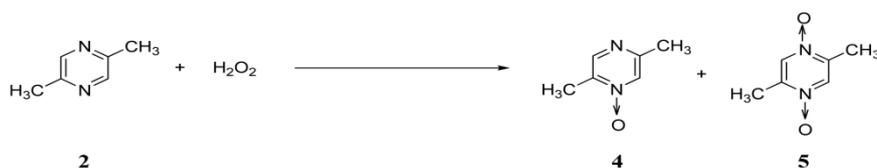


Schéma I.2

### c. Réduction

Les pyrazines sont facilement réduites en pipérazines saturées par le sodium et l'alcool [7], l'amalgame d'aluminium, l'amalgame de sodium, l'étain et l'acide chlorhydrique [8]. Par exemple, la 2,5-diphényl-3,6-dihydropyrazine (7) a été obtenue lorsque la 2,5-diphénylpyrazine (6) a été traitée avec de l'acide iodhydrique et du phosphore rouge, comme le montre le schéma I.3.

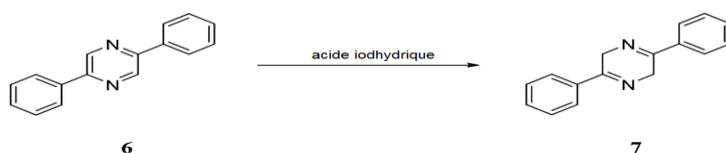


Schéma I.3

En général, les pyrazines sont stables aux alcalins et aux acides, mais des expériences avec certains dérivés aryliques montrent que l'acide hydriodique peut cliver le noyau du cycle pyrazine. On pense que ce réactif réduit d'abord la pyrazine en un dérivé dihydro qui est facilement hydrolysé comme le montre dans le schéma I.4. Le mécanisme de la réaction n'a pas été étudié, mais la transformation de la 2,5-diphénylpyrazine (6) en aminoacétophénone (8) serait possible [9].

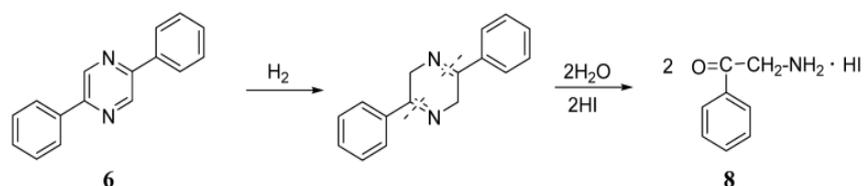
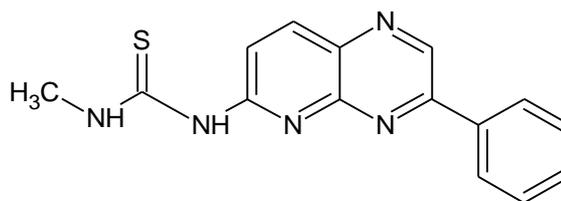


Schéma I.4

### I.1.3. Utilisations et propriétés pharmacologiques

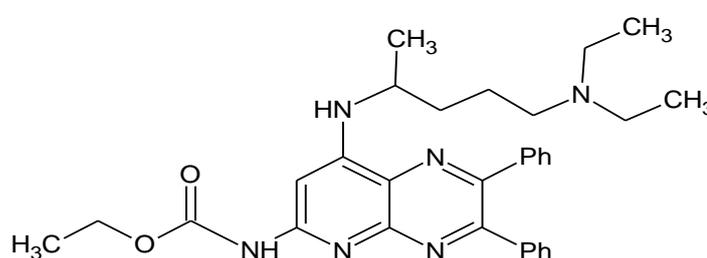
Les molécules possédant le motif pyrazine, sont utilisées pour leurs valeurs thérapeutiques; elles ont des applications très variées puisqu'elles ont démontré des activités anticancéreuses, antiprolifératives, antibactériennes et autres. L'inhibition de l'activité découlant de la kinase Erk 1/2 comme une cible thérapeutique est intéressante, car l'inhibition pharmacologique de la Erk 1/2 renverse l'activation de la voie de transduction des signaux Ras-Raf-Mek-Erk par de nombreuses lésions moléculaires communes dérégées en cancer. Cette voie contrôle un certain nombre de processus cellulaires fondamentaux, incluant la survie, la prolifération, la motilité et la différenciation cellulaire. Elle est constitutivement active en cancer du poumon, du côlon, du pancréas, du rein et de l'ovaire.

AEZS-131(3-méthyl-1-{3-phénylpyrido[2,3-b]pyrazin-6-yl}thiourea) est un petit composé moléculaire oralement actif qui inhibe sélectivement Erk 1/2 avec un IC<sub>50</sub> de 4nM. L'efficacité antiproliférative *in vitro* s'est avérée excellente pour des lignées de cellules tumorales humaines diversifiées [10].



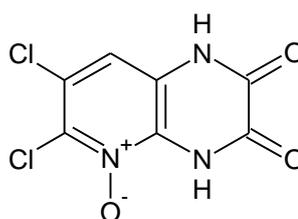
**Figure I.3:** Structure chimique de AEZS-131.

Une série de dérivés de pyridopyrazine a été synthétisée par Bini Mathew et al. [11] qui montre une activité contre FtsZ de *Mycobacterium tuberculosis* (Mtb) in vitro et une activité antibactérienne contre Mtb H37Ra et le Mtb H37Rv.



**Figure I.4:** Structure chimique d'éthyl 2-(8-{{5-(diethylamino)pentan-2-yl}amino}-2-méthyl-3-phénylpyrido[2,3-b]pyrazin-6-yl)acétate.

A.Cugola et al.[12] ont synthétisé le 2,3-dihydro-6,7-dichloro-pyrido [2,3-b] pyrazine-8-oxyde qui a été révélé être un outil très utile pour tester l'efficacité d'un antagoniste de la glycine comme étant un potentiel médicament neuroprotecteur pour les lésions ischémiques.



**Figure I.5:** Structure chimique de 2,3-dihydro-6,7-dichloro-pyrido [2,3-b] pyrazine-8-oxyde.

## I.2. Enzyme de télomérase

### I.2.1. Les télomères

Le télomère est un terme qui vient du grec télós (fin) et mère (segment). Le concept de télomères est apparu suite aux travaux de H.J. Muller sur les

chromosomes de la Drosophile en 1938 et qui a montré que les extrémités des chromosomes interagissent rarement avec les extrémités issues de cassures de l'ADN radio-induites. Il a été proposé ainsi que les chromosomes possèdent une structure particulière.

### **I.2.1.1. Structure et description**

Les télomères sont des structures nucléoprotéiques spécialisées qui coiffent et protègent l'extrémité des chromosomes. L'ADN télomérique consiste en une région double-brin dont la taille moyenne varie de 12 à 15 kb dans les cellules somatiques humaines, contenant de nombreuses répétitions en tandem de séquences hexa-nucléotidiques de séquence 5'-TTAGGG-3'[13]. L'extrémité terminale du télomère est composée d'une extension 3' simple-brin d'environ 250 à 300 bases (brin G) (Figure I.6). Les télomères ont au moins deux fonctions : protéger les extrémités chromosomiques d'une dégradation de l'ADN par les exonucléases et prévenir les fusions de deux extrémités de chromosome capables d'entraîner des cassures et des recombinaisons chromosomiques [14]. En plus de leur rôle dans le maintien de la stabilité du chromosome, les télomères sont impliqués dans l'organisation spatiale du noyau et dans la séparation des chromosomes pendant la division cellulaire [15]. Les télomères joueraient aussi un rôle sur la régulation transcriptionnelle de gènes localisés près des extrémités chromosomiques [16].

Du fait de l'incapacité de la ADN polymérase à répliquer complètement les extrémités terminales d'une molécule d'ADN linéaire, le télomère se raccourcit à chaque division cellulaire, ce qui aboutit à un arrêt de la prolifération des cellules somatiques en phase G0 (sénescence répllicative) lorsque le télomère atteint une taille critique [17]. Les télomères représentent un mécanisme de contrôle de la durée de vie des cellules normales et leur altération permet l'immortalisation indispensable à la prolifération des cellules cancéreuses [18].



**Figure I.6:** Représentation schématique de la séquence d'ADN télomérique humaine.

L'extrémité télomérique contient des répétitions d'ADN double brin de 5 à 15 kb et d'une extension simple-brin en 3' de 200 à 300 b.

### I.2.1.2. Les protéines associées aux télomères humains : complexe Shelterin

Les télomères sont des complexes nucléoprotéiques spécialisés constitués de protéines spécifiques associées aux séquences TTAGGG des extrémités chromosomiques. Ce complexe appelé « télosome » ou « capuchon télomérique » ou « shelterin » permet à la cellule de distinguer entre des cassures double brin de l'ADN et les extrémités naturelles des chromosomes [19]. Ces complexes protéiques permettent ainsi la protection des extrémités chromosomiques vis-à-vis de l'activation des voies de reconnaissance de dommages à l'ADN et de l'action d'exonucléases impliquées dans les voies de réparation. Il est également impliqué dans la régulation de la longueur des télomères en interagissant avec la télomérase ou en remodelant la chromatine.

Le shelterin est composé de six protéines, il reconnaît spécifiquement les répétitions télomériques TTAGGG grâce à trois protéines de ses composants : "Telomeric Repeat binding Factor" 1 et 2 (TRF1 et TRF2) qui se fixent les répétitions télomériques par un domaine C-terminal Myb et diffèrent par leur partie N-terminale (acide pour TRF1 et basique pour TRF2).

Tandis que la protéine des Télomères 1 (POT1) peut lier les répétitions TTAGGG simple-brin de l'extrémité 3'. Trois autres protéines ; TRF2- and TRF1-Interacting Nuclear protein 2 (TIN2), Rap1 (Repressor Activator Protein 1), et

TPP1 (appelé aussi TINT1, PTOP, ou PIP1) servent d'adaptateurs entre TRF1, TRF2 et POT1.

En plus des protéines du complexe shelterin, les télomères humains interagissent avec de nombreuses protéines qui contribuent au maintien et à la protection des extrémités chromosomiques. Ces facteurs « accessoires » sont associés aux télomères de façon transitoire au cours du cycle cellulaire contrairement aux protéines du complexe shelterin. La plupart des protéines accessoires ont une fonction connue autre que leur fonction télomérique, qui pour la majorité est impliquée dans les mécanismes de réparation de l'ADN [20].

Ainsi, les protéines impliquées dans les mécanismes de réparation de l'ADN telles que KU70/KU80 [21], Apollo [22], PARP1 et PARP2 [23].

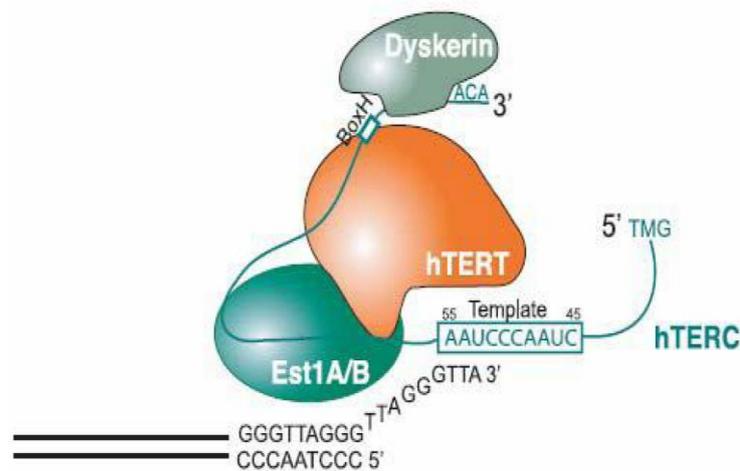
### **I.2.2. La télomérase**

L'identification de la télomérase et la réplication des télomères a été faite en 1985 par Elizabeth Blackburn, Jack Szostak et Carol Greider, ce qui leurs a valu le Prix Nobel de Médecine et physiologie en 2009. Ainsi, Greider et Blackburn avaient découvert la télomérase, qu'ils avaient préalablement décidé d'appeler «telomere terminal transférase», permettant d'ajouter des séquences TTGGGG in vitro sur une amorce simple brin synthétique est identifiée chez *Tetrahymena thermophila* [24]. L'activité télomérase a ensuite été mise en évidence dans de nombreux autres organismes à savoir, chez *Oxytrichia* [25], les *Euplotes* [26] et chez la levure [27]. Chez l'Homme, l'activité enzymatique a été mise en évidence d'abord dans les cellules tumorales HeLa [28]. Ensuite, il a été montré que la télomérase est réactivée dans plus de 90% des cancers, toutes origines confondues [29].

#### **I.2.2.1. Structure de la télomérase**

La télomérase est un complexe ribonucléoprotéique (127 Kda) dont l'holoenzyme comporte deux sous-unités (Figure I.7), une sous-unité catalytique à activité transcriptase inverse, hTERT (human TELomerase Reverse Transcriptase), et d'une sous-unité ribonucléique composée d'une matrice d'ARN, hTR (ou hTER pour human Telomerase RNA) ou hTERC (human telomerase RNA component),

permettant l'addition de séquences TTAGGG à l'extrémité 3' du télomère [30]. La télomérase est active dans les cellules germinales et embryonnaires indifférenciées, mais très faiblement exprimée dans les cellules somatiques. Une très forte activité télomérase a été détectée dans les cellules à prolifération rapide, hématopoïétiques et interstitielles. Les cellules souches et les cellules germinales possèdent également une activité télomérase qui permet d'assurer leurs fonctions respectives dans le renouvellement de tissus et dans la reproduction.



**Figure I.7** : Représentation schématique de la télomérase humaine.

La télomérase Enzyme constituée par une partie ARN codée par hTERC et une sous-unité protéique codée par hTERT. Le complexe télomérase regroupe d'autres protéines telles que Est1 et la dyskérine [31].

La sous-unité hTERT est une ADN polymérase, assurant l'activité catalytique de l'enzyme. Sur son extrémité N-terminale, TERT possède une large extension de 400 acides aminés nommée NTE (acid N-terminal extension) et une courte extension en C-terminal d'environ 150-200 acides aminés nommée CTE (C-terminal extension). Les domaines N-terminaux et NLS sont impliqués dans la localisation nucléolaire de la télomérase [32], alors que le domaine C-terminal est essentiel pour le fonctionnement de hTERT, et semble jouer un rôle dans l'activité processive de la télomérase [33]. La région N-terminale de TERT contribue aux propriétés uniques de la télomérase telles que son association à la matrice ARN, la fixation d'autres composants protéiques et la modulation de sa processivité [34].

La sous-unité ribonucléique hTR (human Telomerase RNA) est constitué d'un ARN intégral de 451 nucléotides dont 11 constituent la matrice pour la synthèse des répétitions télomériques [35]. La sous unité d'ARN hTR est exprimé avec un taux 5 fois supérieur dans les cellules cancéreuses humaines que dans les cellules normales [36]. En plus de ces deux sous-unités, d'autres composants du complexe de la télomérase interviennent dans l'assemblage, la conformation et la localisation nucléaire de la télomérase.

Parmi, ces protéines : TEP1, P23/ p90 et 14-3-3 se lient à la sous-unité hTERT. Cependant, les protéines REP1, hGAR1, Dyskérine, hNOP10, hNHP2, C1/C1, A1/UP1, L22 et hStau se lient à la sous-unité hTR [37].

### **I.2.2.2. Techniques de mesure de l'activité télomérase**

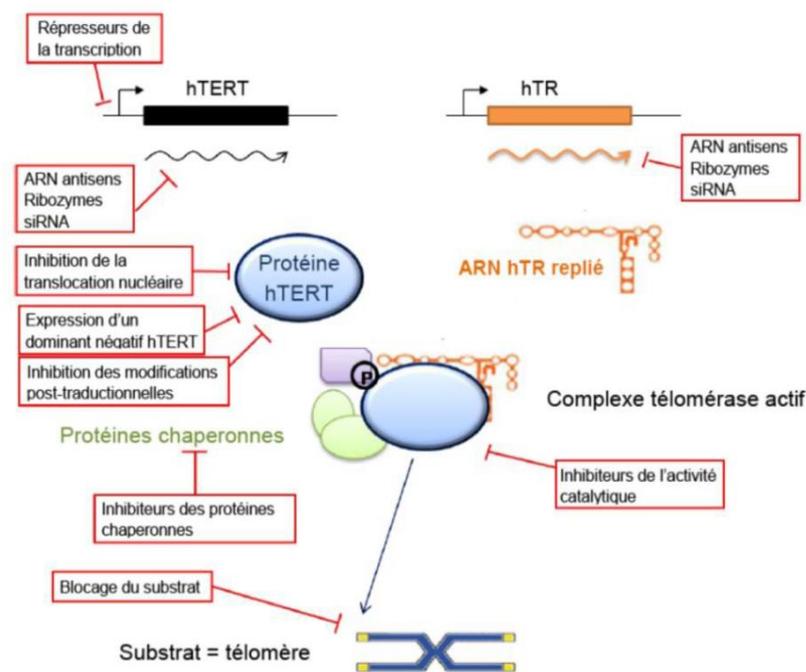
La détection in vitro de l'activité télomérasique dans des échantillons biologiques peut être effectuée grâce à la méthode TRAP (Telomere Repeat Amplification Protocol). Cette technique est composée de deux phases, la première permet l'élongation in vitro d'un fragment télomérique synthétique par la télomérase contenue dans l'extrait cellulaire, la deuxième permet la détection de ces produits d'élongation par amplification par PCR [38].

Plusieurs techniques permettent par la suite de détecter les produits d'amplification de la PCR, qui vont permettre une mesure semi-quantitative ou quantitative de l'activité télomérase. Une mesure semi-quantitative de l'activité télomérase peut être obtenue soit grâce à la méthode de transfert d'énergie en fluorescence (FRET) en utilisant une amorce-sonde d'hybridation marquée par un fluorochrome (TRAPEZEXL, Chemicon International) [39], soit à l'aide d'un test ELISA, en marquant l'amorce TS avec la biotine (TeloTAGGG Telomerase PCR-ELISA kit, Roche) [40]. Une quantification précise de l'activité télomérase peut être réalisée grâce à l'utilisation de la PCR quantitative en temps réel.

### **I.2.2.3. Stratégies thérapeutiques ciblant la télomérase**

L'activité télomérase procure à la cellule un potentiel réplicatif illimité et est considérée comme une des six altérations essentielles nécessaires à la croissance tumorale [41]. La télomérase étant surexprimée dans la plus de 85% des cancers

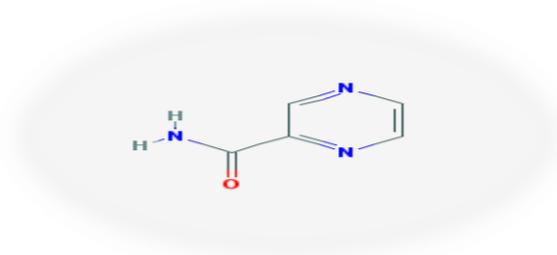
humains. Cette spécificité met en évidence une cible attrayante pour la conception des inhibiteurs en vue du traitement de cancers, tout en minimisant les effets secondaires toxiques dans les cellules normales qui n'expriment pas la télomérase. Ces stratégies ont pour but d'induire un raccourcissement prématuré des télomères et un arrêt de la prolifération anarchique des cellules. De plus les inhibiteurs de la télomérase peuvent aussi posséder une grande spécificité, faible toxicité et réduction des autres effets indésirables [42]. Le ciblage des télomères et la modulation de l'activité télomérasique dans les cellules cancéreuses peut être assurée par différents types d'inhibiteurs (Figure I.8).



**Figure I.8:** Représentation schématique des stratégies d'inhibition de la télomérase.

L'inhibition de la télomérase a pour but d'inhiber l'élongation des télomères, aboutissant au raccourcissement de la taille de l'extrémité chromosomique, à l'arrêt de la prolifération cellulaire et l'entrée des cellules cancéreuses en sénescence et/ou en apoptose. Il existe plusieurs stratégies d'inhibition de la télomérase, soit par inhibition de la sous-unité ribosomale hTR, l'inhibition de la sous-unité catalytique hTERT ou inhibition de l'association du complexe hTERT / hTR.

La pyrazine et ses dérivés ont attiré une attention considérable depuis plusieurs années en raison de leurs activités biologiques importantes. Parmi ces dérivés, on trouve la pyrazinamide (pyrazine-2-carboxamide) (figure I.9) [43]. Cette molécule possède une activité remarquable dans le traitement de la maladie de la tuberculose. D'autres dérivés ont des propriétés anti-cancérogènes et antidiabétiques [44].



**Figure I.9** : structure chimique de pyrazinamide.

**Références bibliographiques**

- [1] Barlin G.B., , The Chemistry of Heterocyclic Compounds, 41, *John Wiley and Sons*(1982).
- [2] Beak P. , Brown R. A. , Yamamoto J., Chiang C. C., and Paul I. C. , Reaction of 2,3-diphenyl-5,6-dihydropyrazine and malononitrile to give 2,6-diamino-3,5-dicyano-4,10-diphenyl-1,7-diazatricyclo[5.2.1.0 4,10]deca-2,5-diene. A novel synthesis of a substituted diazadihydrotriquinacene, *J. Org. Chem.*, 41,3389 (1976).
- [3] Meyer V. and Treadwell F. P. Action of Ammonium Cyanate on Aldehydes, *Ber.*, 14, 1150 (1881).
- [4] Jones J. H. , Bicking J. B. and Cragoe E. Pyrazine diuretics. II. N-amidino-3-amino-5-substituted 6-halopyrazinecarboxamides, *J., Medicin. J. Chem.*, 10,66(1967).
- [5] Mislow K., Binary systems of some carboxylic acid; systems containing heterocyclic acids as the components, *J. Phys. and Colloid Chem.*,52(4), 740(1948).
- [6] Fischer H. and Fink E., *physiol. Chem.*, 283,152(1948).
- [7] Neber P. W. and Friedolsheim A.V, Justus Liebig's Annals of Chemistry *Annalen*, 449,109(1926).
- [8] Kurihara N. H. and Bublitz D. E. , *US. Pat. 3,853,901 through Chem Abs.*, 82, 2084(1975).
- [9] Barlin G.B., the Chemistry of Heterocyclic Compounds,*John Wiley and Sons*, 41,194-196(1982).
- [10] Gerlach M., Seipelt I., Mueller G., Schuster T., Blumenstein L., Aicher B., Guenther E. et Teifel M., « Novel pyrido[2,3-b]pyrazines as orally active ERK-inhibitors», Aeterna Zentaris (2011).
- [11] Mathew B., Srivastava S., Novel pyridopyrazine and pyrimidothiazine derivatives as FtsZ inhibitors, *Bioorg. & Medicin.Chem.*,19 ,7120-7128(2011).
- [12] Cugola A., Synthesis and biological evaluation of pyrido[2,3-b]pyrazine and pyrido[2,3-b]pyrazine-n-oxide as selective glycine antagonists, *Bioorg. & Medicin.Chem. Lett.*, 6 (22) 2749-2754(1996).

- [13] Blackburn E.H., Switching and signaling at the telomere. *Cell*. 106, 661–673(2001).
- [14] Hahn W.C., Role of telomeres and telomerase in the pathogenesis of human cancer. *J. Clin. Oncol.* 21, 2034–2043(2003).
- [15] Kirk K.E., Harmon B.P., Reichardt I.K., Sedat J.W., E.H., B., Block in anaphase chromosome separation caused by a telomerase template mutation. *Science* 275, 1478-1481, (1997).
- [16] Shore D., Aging; Silence is golden. *Curr. Biol.*, 5 , 822-825(1995).
- [17] McEachern M.J., Krauskopf A., Blackburn E.H.,Telomeres and their control *Annu. Rev. Genet.* ,34, 331-358(2000).
- [18] Harley C.B., Futcher A.B., Telomeres shorten during ageing of human fibroblasts. *Nature* ,345, 458 (1990).
- [19] De Lange T., Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes. Dev.*, 19, 2100-2110 (2005).
- [20] Palm W., and De Lange T.,How Shelterin Protects Mammalian Telomeres. *Annu. Rev. Genet.*, 42 ,21-34(2008).
- [21] Hsu H.L., Gilley D., Galande S.A., Hande M.P., Allen B., Kim S H., Li G.C., Campisi J., Kohwi-Shigematsu T., Chen D.J., Ku acts in a unique way at the mammalian telomere to prevent end joining. *Genes. Dev.*,14(22),2807-2812( 2000).
- [22] Beneke S., Cohausz O., Malanga M., Boukamp P., Althaus F., Burkle A, Rapid regulation of telomere length is mediated by poly(ADP-ribose) polymerase-1. *Nucleic. Acids. Res.* 36 ,6309-6317(2008). .
- [23] Gomez M., Wu J., Schreiber V., Dunlap J., Dantzer F., Wang Y., Liu Y., PARP1 Is a TRF2-associated poly(ADP-ribose)polymerase and protects eroded telomeres , *Mol. Biol. Cell.* , 17, 1686-1696 (2006).
- [24] Greider C.W., Blackburn E.H., Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell.*, 43, 405-413 (1985).
- [25] Zahler A.M., Prescott D.M., Telomere terminal transferase activity in the hypotrichous ciliate *Oxytricha nova* and a model for replication of the ends of linear DNA molecules. *Nucleic. Acids. Res.* 16, 6953-6972(1988).
- [26] Shippen-Lentz D., Blackburn E.H.,Telomere terminal transferase activity from *Euplotes crassus* adds large numbers of TTTTGGGG repeats onto telomeric primers. *Mol. Cell. Biol.* ,9 2761-2764(1989).

- [27] Kramer K. M., Haber J. E., New telomeres in yeast are initiated with a highly selected subset of TG1-3 repeats. *Genes. Dev.*, 7, 2345-2356(1993).
- [28] Morin G.B., The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell.*, 59, 521-529 (1989).
- [29] Kim N.W., Piatyszek M. A., Prowse K. R., Harley C. B., West M. D., L., H.P., Coviello G. M., Wright W. E., Weinrich S. L., Shay J. W., Specific association of human telomerase activity with immortal cells and cancer. *Science.*, 266, 2011-2015(1994).
- [30] Greider C.W., Blackburn E.H., A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature.* 337, 331-337(1989).
- [31] Smogorzewska A., De Lange T., Regulation of telomerase by telomeric proteins. *Annu. Rev. Biochem.*, 73, 177-208 (2004).
- [32] Yang Y., Chen Y., Zhang C., Huang H., Weissman S.M., Nucleolar Localization of hTERT Protein Is Associated with Telomerase Function. *Exp. Cell. Res.* 2, 201-209(2002).
- [33] Huard S., Moriarty T.J., Autexier C., The C terminus of the human telomerase reverse transcriptase is a determinant of enzyme processivity. *Nucleic Acids. Res.*, 31, 4059-4070(2003).
- [34] Autexier C., Lue N.F., The Structure and Function of Telomerase Reverse Transcriptase. *Annu. Rev. Biochem.*, 75, 493-517(2006).
- [35] Chen Q.M., Liu J., Merrett J.B., Apoptosis or senescence-like growth arrest: influence of cell-cycle position, p53, p21 and bax in H<sub>2</sub>O<sub>2</sub> response of normal human fibroblasts. *Biochem. J.*, 347, 543-551(2000).
- [36] Yi X., Tesmer V.M., Savre-Train I., Shay J.W., Wright W.E., Both transcriptional and posttranscriptional mechanisms regulate human telomerase template RNA levels. *Mol. Cell. Biol.* 19, 3989-3997(1999).
- [37] Cong Y.S., Wen J., Bacchetti S., The human telomerase catalytic subunit hTERT: organization of the gene and characterization of the promoter. *Hum. Mol. Genet.* 8, 137-142(1999).
- [38] Kim N.W., Piatyszek M. A., Prowse K. R., Harley C. B., West M. D., L., H.P., Coviello G. M., Wright W. E., Weinrich S. L., Shay J. W., Specific association of human telomerase activity with immortal cells and cancer. *Science.* 266, 2011-2015(1994).
- [39] Uehara H., Nardone G., Nazarenko I., Hohman R.J., Detection of telomerase activity utilizing energy transfer primers: comparison with gel- and ELISA-based detection. *Biotechniques.* 26, 552-558(1999).

[40] Wu Y.Y., Hruszkewycz A.M., Delgado R.M., A., Y., Vortmeyer A.O., Moon Y.W., Weil R.J., Zhuang Z., Remaley A.T., Limitations on the quantitative determination of telomerase activity by the electrophoretic and ELISA based TRAP assays. *Clin. Chim. Acta.* 293, 199-212 (2000).

[41] Hanahan D., Weinberg R.A., The hallmarks of cancer. *Cell.* 100, 57-70(2000).

[42] Gellert G.C., Jackson S.R., Dikmen Z.G., Wright W.E., Shay J.W., Telomerase as therapeutic target in cancer. *Drug. Discovery. Today. Dis. Mech.* 2, 159-164(2005).

[43] Spaia S, Magoula I, Tsapas G, Vayonas G "Effect of pyrazinamide and probenecid on peritoneal urate transport kinetics during continuous ambulatory peritoneal dialysis". *Perit. Dial. Int.* 20, 47-52(2000).

[44] pharmacology, k.d.tripathy, 6<sup>th</sup> edition(2008).

# *Matériels et Méthodes*

*Chapitre II :*  
*Méthodes de la*  
*Modélisation*  
*moléculaire*

## II.1. Introduction

Avec l'arrivée de méthodes théoriques de calcul de plus en plus sophistiquées et les ressources de computation plus accessibles, la chimie informatique est maintenant devenue un outil de plus en plus utile à la fois pour l'industrie et le milieu académique [1].

La modélisation par ordinateur d'une molécule implique généralement une présentation graphique de la géométrie ou de la configuration des atomes de la molécule, suivie de l'application d'une méthode théorique [2].

La modélisation moléculaire constitue une approche scientifique novatrice, originale, qui permet de travailler *in silico* dans des domaines extrêmement variés, que ce soit dans un but très fondamental, ou beaucoup plus appliqué.

La modélisation moléculaire est un terme général qui englobe les différentes techniques de graphisme moléculaire et de chimie computationnelle permettant d'afficher, simuler, analyser, calculer et stocker les propriétés des molécules [3].

La modélisation moléculaire (Mécanique et Dynamique Moléculaire) est une méthode qui permet la détermination de la structure et de l'énergie d'entités moléculaires.

La modélisation moléculaire a pour but de prévoir la structure et la réactivité des molécules ou des systèmes de molécules. La recherche et la synthèse de nouveaux composés chimiques et biochimiques sont aujourd'hui souvent associées à une étude par modélisation moléculaire.

Grace au développement informatique de ces dernières années et à l'essor du calcul parallèle intensif en particulier, la modélisation moléculaire est devenue un véritable enjeu.

En effet les systèmes moléculaires qui sont amenés à être étudiés tendent à devenir de plus en plus complexes. Cette complexité est bien sur liée à la taille des molécules envisagées (plusieurs centaines de milliers d'atomes pour les molécules biologiques par exemple) ainsi qu'à la structure intrinsèque des atomes eux-mêmes, mais aussi au degré de précision exigé pour le calcul de certaines grandeurs physiques [4].

La modélisation moléculaire consiste en la construction de modèles tridimensionnels à partir des données. Elle trouve sa raison d'être d'une part dans les limitations expérimentales

des méthodes de détermination de structure de protéines et d'autre part dans l'incapacité actuelle de prédire la structure 3D à partir de la seule information de séquence. En outre, elle permet d'investiguer les changements de conformations liées à des mutations à partir de structures 3D expérimentales.

La modélisation moléculaire est une application des méthodes théoriques et des méthodes de calcul pour résoudre des problèmes impliquant la structure moléculaire et la réactivité chimique [5]. Ces méthodes peuvent être relativement simples et utilisables rapidement ou au contraire elles peuvent être extrêmement complexes et demander des centaines d'heures de temps d'ordinateur, même sur un superordinateur. En plus, ces méthodes utilisent souvent des moyens infographiques très sophistiqués qui facilitent grandement la transformation de quantités impressionnantes de nombres en quelques représentations graphiques facilement interprétables [6].

Les trois méthodes de la modélisation moléculaire peuvent être rangées en trois catégories:

- Les méthodes quantiques
- La mécanique moléculaire
- La dynamique moléculaire.

## II.2. Méthodes de la mécanique quantique

### II.2.1. Base de la mécanique quantique

La chimie quantique repose essentiellement sur les méthodes basées sur la résolution de l'équation de Schrödinger [7], qui décrit le mouvement des électrons et des noyaux d'un système moléculaire. Cette équation dans le cas stationnaire indépendante du temps s'écrit :

$$\hat{H} \Psi = E \Psi \quad (1)$$

où

-  $\hat{H}$  est l'opérateur hamiltonien non relativiste du système,

-  $\Psi$  est la fonction d'onde du système, -  $E$  est l'énergie totale du système.

Cette équation ne peut être rigoureusement résolue que pour les systèmes monoélectroniques. La description des systèmes plus complexes nécessite la mise en œuvre d'un certain nombre d'approximations [8].

L'hamiltonien exact d'un système comportant N noyaux et n e électrons, où les noyaux sont désignés par A et B et les électrons par k et l, s'écrit :

$$\hat{H} = \underbrace{-\frac{1}{2} \sum_{k=1}^{n_e} \frac{\hbar}{m_e} \cdot \nabla_k^2}_{\text{Opérateur cinétique des } \bar{e}} - \underbrace{\sum_{k=1}^{n_e} \sum_{A=1}^N \frac{e^2}{4\pi\epsilon_0} \cdot \frac{Z_A}{r_{kA}}}_{\text{Opérateur d'attraction des } \bar{e} \text{ par les noyaux}} + \underbrace{\sum_{k=1}^{n_e} \sum_{l>k}^{n_e} \frac{e^2}{4\pi\epsilon_0 \cdot r_{kl}}}_{\text{Opérateur de répulsion entre les } \bar{e}} - \underbrace{\frac{1}{2} \sum_{A=1}^N \frac{\hbar}{M_A} \cdot \nabla_{R_A}^2}_{\text{Opérateur cinétique des noyaux}} + \underbrace{\sum_{A=1}^N \sum_{B>A}^N \frac{e^2}{4\pi\epsilon_0} \cdot \frac{Z_A \cdot Z_B}{R_{AB}}}_{\text{Opérateur cde répulsion entre les noyaux}} \quad (2)$$

### II.2.2. Approximation de Born-Oppenheimer

L'approximation de Born-Oppenheimer (1927) [9] consiste à simplifier la résolution de l'équation de Schrödinger en séparant la partie électronique de la partie nucléaire dans la fonction d'onde. Cette approximation est basée sur le fait que les électrons étant plus légers que les noyaux (la masse d'un proton est égale à 1836 fois celle de l'électron), ils se déplacent plus rapidement et par conséquent réagissent instantanément à toute modification de la position des noyaux.

L'équation de Schrödinger peut ainsi être séparée en une partie nucléaire et une partie électronique. La fonction d'onde approchée du système, solution de cette équation, s'écrit alors sous la forme d'un produit de deux fonctions :

$$\Psi(R, r) = \Psi_e(r, R) \cdot \Psi_N(R) \quad (3)$$

Où  $\Psi_N(R)$  est la fonction d'onde nucléaire et  $\Psi_e(r, R)$  la fonction d'onde électronique pour des positions données des noyaux figés. Les positions des électrons et des noyaux sont désignées respectivement par r et R.

L'hamiltonien dans ce cas s'écrit:

$$\hat{H} = \hat{H}_N + \hat{H}_e(r, R) \quad (4)$$

où  $\hat{H}_N$  est l'opérateur nucléaire et  $\hat{H}_e$  l'opérateur électronique.

Par suite, l'équation de Schrödinger électronique s'écrit :

$$\hat{H}_e(r, R) \Psi_e(r, R) = \varepsilon(R) \Psi_e(r, R) \quad (5)$$

L'énergie de Born-Oppenheimer, pour un ensemble de positions R des noyaux fixés est :

$$E(\mathbf{R}) = \varepsilon(\mathbf{R}) + \sum_{A=1}^N \sum_{B>A}^N \frac{e^2 Z_A Z_B}{4\pi\epsilon_0 R_{AB}} \quad (6)$$

Cette énergie permet, pour la totalité des positions des noyaux, de définir la surface d'énergie potentielle dite de Born-Oppenheimer dont le minimum de plus basse énergie correspond à la géométrie d'équilibre du système. Connaissant cette énergie, on peut résoudre l'équation nucléaire (7) et avoir ainsi accès aux constantes de force du système et à ses fréquences de vibration.

$$\left[ -\frac{1}{2} \sum_{A=1}^N \frac{\hbar^2}{M_A} \nabla_{\mathbf{R}_A}^2 + E(\mathbf{R}) \right] \cdot \psi_N(\mathbf{R}) = E_N \cdot \psi_N(\mathbf{R}) \quad (7)$$

Cependant, l'équation ainsi formulée ne peut être résolue de manière exacte que pour les systèmes hydrogénoïdes. Pour un système polyélectronique, il est indispensable d'introduire de nouvelles approximations afin de décrire les électrons en interaction. Deux approches sont connues, la première fait appel à la fonction d'onde, l'autre à la densité électronique.

### II.2.3. Fonction d'onde électronique

La résolution de l'équation aux valeurs propres, impose impérativement une fonction d'onde qui doit satisfaire le principe de Pauli ; pour cela, elle doit inclure le spin de l'électron, sous la forme de spin-orbitale. Sa détermination exigera en outre des approximations qui lui imposent une forme a priori : l'approximation orbitale, et le développement en une combinaison linéaire d'un nombre limité d'orbitales atomiques (LCAO) [10].

### II.2.4. Méthode quantique : *ab initio*

Les travaux effectués au début du vingtième siècle par Planck, Einstein, Bohr, De Broglie, Schrödinger et Heisenberg ont abouti à l'élaboration de la mécanique des microsystemes. En 1925, grâce aux efforts de W. Heisenberg et E. Schrödinger et de P. Dirac, J. Von Neumann, N. Bohr, M. Born et d'autres, une nouvelle mécanique a été créée : la Mécanique quantique [11-13], qui a permis d'expliquer de nombreuses propriétés physiques, telles que les propriétés chimiques des éléments et la formation des liaisons chimiques [14].

Les méthodes de modélisation basées sur la mécanique quantique [15] visent à décrire

le système étudié par une fonction d'onde qui peut théoriquement être déterminée par résolution de l'équation de Schrödinger [16]. Cette équation relie les états stationnaires d'un système moléculaire et les énergies qui y sont associées à un opérateur Hamiltonien et à leur fonction d'onde.

L'un des principaux avantages de la mécanique quantique, par rapport aux autres techniques de modélisation moléculaire, est qu'elle ne fait appel qu'à un nombre minimal d'approximations et considère de manière explicite les électrons de la molécule étudiée. Ceci permet le calcul de propriétés importantes telles que les charges atomiques, le potentiel électrostatique moléculaire ou la topologie et l'énergie des orbitales frontières (HOMO, LUMO...). Son principal inconvénient est d'être particulièrement coûteux du point de vue du temps de calcul. Malgré ce problème, le traitement d'un système par mécanique quantique reste la méthode de référence pour le calcul des propriétés moléculaires, comme le rappelle cette citation de Burkert et Allinger [17]:

"For calculating molecular properties, quantum chemistry seems to be the obvious tool to use. Calculations that do not use the Schrödinger equation are acceptable only to the extent that they reproduce the results of high level quantum mechanical calculations."

En termes simples....

Le point de départ de toute discussion en mécanique quantique est donc l'équation de Schrödinger.

Où  $h$  est la constante de Planck et  $m$  est la masse de la particule traitée.

Dans sa forme indépendante du temps, cette équation peut être écrite :  $H \Psi (r) = E \Psi (r)$

Ces expressions font intervenir la fonction d'onde du système,  $\psi$ , qui en tant que telle n'a aucune signification physique. Son carré représente par contre une probabilité de présence d'un électron au sein d'un volume donné. La détermination de la fonction d'onde permet donc en théorie de calculer l'énergie associée au système auquel elle s'applique.

L'approximation de Born-Oppenheimer [18] permet de traiter séparément le mouvement des noyaux et des électrons. Le terme d'énergie cinétique contenu dans l'opérateur Hamiltonien ( $H$ ) fait en effet intervenir, au sein de son dénominateur, la masse de la particule traitée. Les protons et les neutrons étant 1836 fois plus lourds qu'un électron, la contribution des noyaux atomiques dans cet opérateur peut être donc considérée comme négligeable. En pratique, ceci revient à postuler que les électrons, de masse beaucoup plus faible, peuvent s'adapter de manière quasiment instantanée aux changements de position des noyaux

et que l'énergie d'une molécule peut donc être écrite comme une fonction des seules coordonnées nucléaires.

$$\left[ \left( -\frac{\hbar^2 \nabla^2}{8\pi^2 m} \right) + v \right] \Psi(r, t) = i\hbar / 2\pi \left( \frac{\partial \Psi(r, t)}{\partial t} \right) \quad (8)$$

La résolution exacte de l'équation Schrödinger n'est possible que pour l'atome d'hydrogène et les systèmes hydrogénoïdes. Pour les systèmes polyélectroniques, on fait appel aux méthodes d'approximation [19-20].

Les principales variantes sont la méthode de Huckel et les méthodes de champ auto cohérent (Self consistent Field, SCF). Les méthodes *ab initio* sont caractérisées par l'introduction d'une base arbitraire pour étendre les orbitales moléculaires et alors le calcul explicite toutes les intégrales exigées qui impliquent cette base.

Les calculs *ab initio* peuvent être exécutés au niveau d'approximation de Hartree-Fock, qui est équivalent à un calcul du champ auto – cohérent SCF (Self Consistent Field).

L'option et les niveaux d'Hartree-Fock incluent les effets de corrélation qui n'est pas incluse au niveau d'approximation d'Hartree-Fock d'une solution non-relativiste pour l'équation de Schrödinger.

#### II.2.4.1. Approximation de Hartree-Fock

L'approximation orbitalaire, introduite par Hartree en 1928 [21], consiste à écrire la fonction d'onde pour un système polyélectronique sous forme de produit de spin-orbitales monoélectroniques supposées normalisées.

$$\psi_e(\mathbf{1}, \mathbf{2}, \mathbf{3}, \dots, \mathbf{n}_e) = \varphi_1(\mathbf{1}) \cdot \varphi_2(\mathbf{2}) \cdot \varphi_3(\mathbf{3}) \cdot \dots \cdot \varphi_{n_e}(\mathbf{n}_e) \quad (9)$$

Sachant que chaque spin-orbitale est le produit d'une fonction de position de l'électron  $\varphi$  et d'une fonction de spin  $\eta(s_i)$ .

$$\varphi_i(\mathbf{n}_i) = \phi_i(\mathbf{r}_i) \cdot \eta(s_i) \quad (10)$$

On associe à la fonction de spin  $\eta(s_i)$  deux formes :  $\alpha$  pour le spin  $(\frac{1}{2})$  et  $\beta$  pour le spin  $(-\frac{1}{2})$ .

Ce concept a été ensuite généralisé par Hartree et Fock [22] en écrivant la fonction d'onde sous forme d'un déterminant de Slater construit sur la base des fonctions d'onde de chaque électron pour satisfaire le principe de Pauli [23]. Il est possible, dans ce cas,

d'évaluer la meilleure fonction d'onde, pour un tel système en minimisant, l'énergie à l'aide du principe variationnel.

$$\psi(1, 2, 3, \dots, n_e) = \frac{1}{\sqrt{n_e!}} \begin{vmatrix} \varphi_1(1) & \varphi_2(1) & \varphi_3(1) \dots & \varphi_{n_e}(1) \\ \varphi_1(2) & \varphi_2(2) & \varphi_3(2) \dots & \varphi_{n_e}(2) \\ \dots & \dots & \dots & \dots \\ \varphi_1(n_e) & \varphi_2(n_e) & \varphi_3(n_e) \dots & \varphi_{n_e}(n_e) \end{vmatrix} \quad (11)$$

$\frac{1}{\sqrt{n_e!}}$  est le facteur de normalisation

Malgré les résultats très satisfaisants obtenus, cette approche négligeant le terme de corrélation, présente une limitation majeure : elle surestime la distance moyenne entre les électrons puisque l'on suppose que chaque électron se trouve dans un champ moyen créé par tous les autres sans tenir compte des interactions interélectroniques et elle surestime l'énergie de répulsion électronique. Pour décrire correctement le système, Löwdin [24] a défini l'énergie de corrélation ' $E_{corr}$ ' comme étant la différence entre l'énergie définie par la méthode Hartree-Fock (EHF) et l'énergie exacte non relativiste du système.

$$E_{corr} = E_{exacte} - E_{HF} \quad (12)$$

Ces interactions électroniques, qui ne sont pas négligeables, ont été prises en compte pour améliorer les résultats et ont donné naissance aux nouvelles méthodes dites post-Hartree-Fock dont la méthode d'interaction de configuration (CI) [25], la méthode multi-configuration SCF (MC-SCF) [26], la perturbation Moller-Plesset (MP) [27] et la méthode Coupled-Cluster (CC) [28]. La fonction d'onde s'exprime sous forme d'une combinaison linéaire de plusieurs déterminants de Slater.

Cependant, ces méthodes sont très faramineuses en termes de temps et en puissance de calculs et ne sont, par conséquent, applicables que pour de petits systèmes.

#### II.2.4.2. Méthode post-Hartree-Fock

Les méthodes les plus importantes, dans l'utilisation courante, pour introduire la corrélation électronique sont d'une part les méthodes appelées post-HF avec l'interaction de configuration (CI), les méthodes de "coupled cluster" (CC) et les "many-body perturbation

theory" (MP2, MP4, ...) et, d'autre part, les méthodes qui dérivent de la théorie de la fonctionnelle de la densité (DFT).

La résolution des équations d'HF donne une fonction d'onde de référence sous la forme d'un déterminant de Slater.

La fonction d'onde de la méthode CI est une combinaison linéaire de déterminants de Slater représentant l'état fondamental et des configurations excitées. Ces dernières correspondent à l'excitation d'un, de deux ou plus électrons d'une orbitale occupée à une virtuelle. Les coefficients de cette combinaison linéaire sont déterminés selon le principe variationnel; l'énergie qui en découle est donc une limite supérieure à l'énergie exacte. L'IC totale ("Full CI") est la limite que l'on peut atteindre dans une base donnée, c'est-à-dire qu'elle comprend toutes les excitations possibles des  $n$  électrons. Cependant l'IC totale accroît énormément les calculs avec le nombre d'électrons et la dimension des bases utilisées. Pour ces raisons de tels calculs servent habituellement de référence pour des systèmes comprenant un petit nombre d'électrons ( $n \leq 20$ ). La méthode CI limitée à un nombre d'excitation inférieure à la totalité des possibilités ("truncated CI") n'est pas "size consistent" (c'est à dire, par exemple, que l'énergie d'un dimère placé à distance infinie n'est pas égale à deux fois l'énergie du monomère). Cette propriété importante est garantie par des méthodes non variationnelles comme "many-body perturbation theory" ou "coupled cluster methods".

Dans le formalisme de Moller-Plesset, l'hamiltonien est représenté comme la somme de l'hamiltonien HF de l'état fondamental et d'une différence, entre l'hamiltonien exact et HF, traitée comme une perturbation. Dans la théorie de perturbation, la fonction d'onde et l'énergie, pour un état donné, donnent l'état appelé "zéro" ou fondamental auquel est ajouté des corrections successives résultant des différents ordres de perturbation pour le traitement du système. L'approximation MP2 comprend des substitutions simples et doubles, la théorie des perturbations d'ordre 4 (MP4) ajoute des substitutions triples et quadruples.

Les méthodes "coupled cluster" (CC) sont actuellement les plus puissantes des méthodes *ab initio*. La théorie CC commence par un postulat sur la fonction d'onde à  $n$  électrons à savoir que cette fonction d'onde est multipliée par une exponentielle naturelle d'une somme d'opérateurs d'excitation d'électrons (T); T2 implique une double substitution.

CCSDT comprend jusqu'à un opérateur de triple excitation.

### II.2.5. Théorie de la fonctionnelle de la densité (DFT)

En mathématique, une fonction  $f$  fait correspondre à la variable  $x$  le nombre  $f(x)$ . Une fonctionnelle  $F$  est une fonction de fonctions qui fait correspondre à une fonction  $f$  un nombre noté  $F[f]$ .

Avant d'aborder la théorie de la fonctionnelle de la densité, il paraît évident de définir son élément central : la densité électronique.

La densité électronique notée  $\rho$ , est une fonction positive qui dépend de trois variables seulement, les coordonnées d'espace. Cette quantité, intégrée sur tout l'espace est égale au nombre total des électrons et s'annule à l'infini.

$$\int \rho(\mathbf{r}) d\mathbf{r} = n_e \quad (13)$$

$\rho(\mathbf{r})$  représente, par définition, la probabilité de présence d'un électron dans un élément de volume défini par  $\mathbf{r}$ . Elle permet la description d'un système chimique et la détermination complète de ses propriétés contrairement à la fonction d'onde ; c'est pour cette raison que plusieurs recherches ont été entreprises pour la mise en place d'un formalisme se basant sur cette quantité et qui ont abouti à la théorie de la fonctionnelle de la densité.

#### II.2.5.1. Fondements de la théorie de la fonctionnelle de la densité

Les méthodes de la théorie de la fonctionnelle de la densité (DFT de l'anglais Density Functional Theory) ont acquis une popularité grandissante pendant ces dernières années. Elles constituent une alternative de choix aux méthodes ab-initio citées précédemment. L'énergie est donc une fonctionnelle de la densité électronique et s'écrit  $E[\rho]$ .

Par suite, l'énergie électronique totale s'écrit sous forme d'une somme de fonctionnelles : énergies cinétique et potentielle d'attraction électron-noyaux et de répulsion électron-électron.

$$E[\rho] = T[\rho] + E_{ne}[\rho] + E_{ee}[\rho] \quad (14)$$

La DFT a été introduite à l'origine par Thomas [29] et Fermi [30] en 1927 ; quand ils proposèrent de décrire les propriétés électroniques d'un système en terme de fonctionnelles de la densité. En considérant un gaz homogène à électrons libres non interagissant, ils ont pu donner l'expression analytique de l'énergie dans laquelle l'interaction électronique est

limitée seulement à la répulsion coulombienne classique. En 1930 Dirac [31], compléta cette théorie, en introduisant une fonctionnelle d'échange, mais ces deux modèles étaient insuffisants pour traiter la liaison chimique. Ils furent complétés par les travaux de Wigner (1932) [32] et de Von Weizsäcker (1935) [33] qui eurent l'idée de prendre en compte l'hétérogénéité de la distribution électronique dans la molécule et de rajouter un terme non local dépendant du gradient de la densité électronique à la fonctionnelle énergie cinétique, celle-ci étant très grande devant les autres composantes de l'énergie électronique du système. Il est en effet, très important de calculer le plus précisément possible l'énergie cinétique et toute approximation conduit à des erreurs sur l'énergie totale du système. Un peu plus tard, Hohenberg et Kohn en 1964 [34], ensuite Kohn et Sham en 1965 [35] ont pu définir les bases modernes de la DFT.

### II.2.6. Méthodes semi-empiriques

La mécanique quantique est une technique mathématique rigoureuse basée sur l'équation de Schrödinger. La solution de cette équation permet d'obtenir des informations précises sur les propriétés géométriques et électroniques de la molécule. Les calculs peuvent être de type *ab initio* ou semi-empirique (ex : CNDO, PM3). En *ab initio*, on tient compte de tous les électrons de la molécule et on vise une solution rigoureuse de l'Hamiltonien.

Les calculs semi-empiriques traitent seulement les électrons de valence et utilisent un Hamiltonien plus simple ayant des facteurs de correction basés sur des données expérimentales. L'équation de Schrödinger d'un système moléculaire peut être résolue sans approximation (*ab initio*) ou en introduisant des approximations (semi-empirique).

En mécanique quantique, on se préoccupe de la distribution des électrons (orbitales) dans l'espace. Les meilleurs programmes comportent des processus d'optimisation de la géométrie [36].

L'objectif de la mécanique quantique est principalement de déterminer l'énergie et la distribution électronique. Ainsi les énergies moléculaires sont calculées en utilisant l'équation de Schrödinger avec le formalisme des orbitales moléculaires (MO). Par les méthodes MO, les énergies peuvent être extraites via le formalisme SCF ("Self Consistent Field") ou par une méthode de perturbation comme PCILO.

En mécanique quantique, on étudie la distribution électronique (orbitale) dans l'espace. Les propriétés moléculaires qui peuvent être calculées sont la géométrie, les spectres de

vibration, les spectres d'absorption UV, les moments dipolaires, les spectres électroniques et les fonctions responsables de la réactivité chimique (Orbital HOMO, LUMO).

Les termes énergétiques les plus difficile à calculer sont estimés à partir des données expérimentales. Les temps de calculs sont considérablement raccourcis, mais la méthode est tributaire des composés qui ont servi à l'étalonner. Selon la nature des approximations utilisées [5], on distingue plusieurs variantes :

- CNDO : (Complete Neglect of Differential Overlap) 1<sup>ère</sup> méthode semi empirique, elle a été proposée par Pople, Segal et Santry en 1965. Méthode présentant certains défauts entre autres : elle ne tient pas compte de la règle de Hund.
- INDO: (Intermediate Neglect of Differential Overlap) Proposée par Pople, Beveridge et Dobosh en 1967. Elle permet de distinguer entre les états singulets et les états triplets d'un système en conservant les intégrales d'échange.
- MINDO/3: Proposée par Bingham, Dewar et Lo en 1975. Paramétrisation effectuée en se référant aux résultats expérimentaux et non pas aux résultats ab-initio, de plus l'algorithme d'optimisation utilisé est très efficace (Davidon-Fletcher-Powell). Cependant, elle surestime la chaleur de la formation des systèmes insaturés et sous-estime celle des molécules contenant des atomes voisins ayant des paires libres.
- MNDO: (Modified Neglect of Diatomic Overlap) Proposée par Dewar et Thiel en 1977. Méthode basée sur l'approximation NDDO (Neglect of Diatomic Differential Overlap) qui consiste à négliger le recouvrement différentiel entre orbitales atomiques sur des atomes différents. Cette méthode ne traite pas les métaux de transition et présente des difficultés pour les systèmes conjugués.
- AM 1 : (Austin Model 1) Proposée par Dewar en 1985. Il a tenté de corriger les défauts de MNDO.
- PM 3 : (Parametric Method 3) Proposée par Stewart en 1989. Présente beaucoup de points en commun avec AM1. D'ailleurs, il existe toujours un débat concernant les mérites relatifs de paramétrisation de chacune d'elles. PM3 et AM1 généralement sont les méthodes les plus rigoureuses dans le logiciel HyperChem, PM3 a été paramétré pour beaucoup des éléments principaux des groupes et quelques métaux de transition. PM3 est différent d'AM1 seulement dans les valeurs des paramètres.
- SAM 1 : (Semi-ab-initio Model 1) La méthode la plus récente proposée par Dewar en 1993. Elle inclut la corrélation électronique.

### II.3. Mécanique moléculaire

L'expression "Mécanique moléculaire" désigne actuellement une méthode de calcul qui permet, d'obtenir des résultats de géométrie d'énergie moléculaires en se basant sur la mécanique classique.

La mécanique moléculaire est apparue en 1930 [37], mais s'est développée à partir des années soixante, quand les ordinateurs furent plus accessibles et plus performant.

Les méthodes de la mécanique moléculaire permettent le calcul de propriétés structurales et thermodynamiques de systèmes moléculaires comportant jusqu'à plusieurs milliers d'atomes.

Les électrons n'y sont pas traités explicitement comme dans un calcul de mécanique quantique, mais les atomes y sont représentés par des masses ponctuelles chargées reliées les unes aux autres par des ressorts.

Contrairement à la mécanique quantique, l'énergie des systèmes moléculaires ne provient pas de la résolution de l'équation de Schrödinger, mais est décrite par les fonctions empiriques auxquelles sont associés des paramètres dérivant de l'expérience ou de calculs précis quanto-chimiques. Le « champ de force » établi par cette méthode représente aussi bien que possible les variations de l'énergie potentielle avec la géométrie moléculaire.

L'énergie de la molécule est exprimée sous la forme d'une somme de contributions associées aux écarts de la structure par rapport à des paramètres structuraux de référence :

$$E = E_{\text{liaison}} + E_{\text{angle}} + E_{\text{dièdre}} + E_{\text{van der Waals}} + E_{\text{électrostatique}} \quad (15)$$

Les variables du calcul sont les coordonnées internes du système : longueurs de liaisons, angles de valence, angles dièdres, ainsi que les distances entre atomes non liés dont les interactions sont représentées par un potentiel de Van der Waals et un potentiel électrostatique le plus souvent de type Coulombien.

Habituellement, on distingue dans l'équation de l'énergie du champ de force les termes intramoléculaires concernant les atomes liés chimiquement (liaisons, angle, dièdres, etc.) des termes intermoléculaires concernant les interactions entre les atomes non-liés chimiquement (électrostatiques, Van der Waals, etc.).

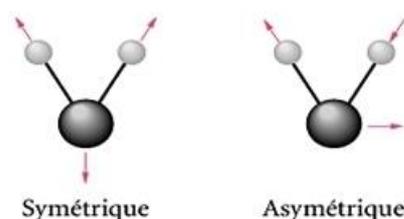
### II.3.1. Énergie d'interaction entre atomes lié

Les potentiels utilisés sont du même type que ceux utilisés en analyse vibrationnelle. La déformation du squelette est décrite par l'élongation des liaisons et distorsion des angles de valence.

Ces déformations des liaisons et des angles sont représentées par un oscillateur harmonique avec une sommation sur toutes les liaisons covalentes ou sur tous les angles entre atomes liés par liaisons "covalentes" [38].

#### II .3.1.1. Énergie d'élongation

Les liaisons entre les atomes dans un édifice moléculaire ont souvent tendance à s'allonger ou à se contracter (figure II.1).



**Figure II.1.** Énergie d'élongation entre deux atomes liés.

Cette déformation est régie en première approximation par loi de "Hooke" d'élongation des ressorts. On peut ainsi lui associer une énergie d'élongation de la forme:

$$E(L)=1/2 K_r(L-L_0)^2 \quad (16)$$

où

$K_r$  : est la constante d'élongation ou constante de Hooke.

$L_0$  : la longueur de la liaison de référence.

$L$  : la longueur de la liaison dans le modèle.

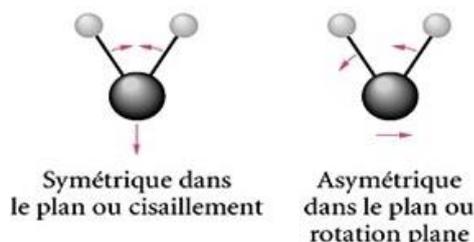
Tous ces termes d'élongation sont sommés sur l'ensemble des liaisons de la molécule.

Un terme cubique  $(L-L_0)^3$  est généralement ajouté pour des déformations importantes. Le calcul de cette énergie impose donc de connaître au minimum les deux paramètres indissociables ( $K_r$  et  $L_0$ ) qui représentent un sous-ensemble du champ de force. En effet, il ressort du développement en série de l'expression mathématique de la courbe

de Morse traduisant l'interaction existante, entre deux atomes en fonction de leur distance respective.

### II.3.1.2. Énergie de flexion

La fluctuation des atomes autour de leur position d'équilibre engendre une déformation des angles de valence (figure II.2).



**Figure II.2:** Énergie de déformation des angles de valence.

Ce phénomène est régi par une énergie de flexion qui peut s'exprimer sous les mêmes formes que précédemment à savoir, pour la plus simple :

$$E(\theta) = \frac{1}{2} K_{\theta} (\theta - \theta_0)^2 \quad (17)$$

ou,

$K_{\theta}$  : constante de flexion.

$\theta_0$  : angle de valence de référence.

$\theta$  : angle de valence dans la molécule.

Le couple ( $K_{\theta}$ ,  $\theta_0$ ) représente ici encore un sous-ensemble du champ de force.

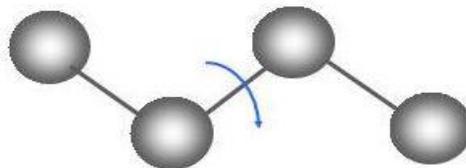
### II.3.1.3. Énergie de torsion

La variation des angles dièdres (angles de torsion), E torsion, impose l'utilisation d'un terme périodique. La torsion correspond à la rotation autour d'une liaison simple. L'énergie potentielle s'exprime en fonction de l'angle de rotation  $\Phi$  (angle dièdre) (figure II.3).

Le terme énergétique représentant la déformation des angles dièdres est représenté par une fonction développée en série de Fourier et il est calculé par la formule [39] :

$$E(\phi) = \frac{1}{2} [V_1(1 + \cos \phi) + V_2(1 - \cos 2\phi) + V_3(1 + \cos 3\phi)] \quad (18)$$

$V_1, V_2, V_3$  sont les constantes du potentiel de l'énergie de torsion.



**Figure II.3 :** Énergies de torsion.

### II.3.2. Énergie d'interaction entre atomes non liés

Les deux derniers termes traduisent les interactions entre atomes non liés de manière covalente, seuls les atomes séparés de plus de trois liaisons sont considérés comme pouvant avoir des interactions non liées. Ce choix se justifie par le fait que les interactions, entre atomes séparés par une, deux ou trois liaisons, sont suffisamment bien décrites par les termes de déformation des longueurs de liaisons et d'angles formés par deux liaisons covalentes.

#### II.3.2.1. Énergie de Van der Waals

Le quatrième terme de la fonction d'énergie est le potentiel de Lennard-Jones qui traduit les interactions de Van der Waals. Il est constitué d'un terme répulsif en  $r_{ij}^{-12}$  représentant le principe d'exclusion de Pauli et la répulsion internucléaire à courte distance. Il contient également un terme attractif en  $r_{ij}^{-6}$  représentant les interactions d'origine dipolaire dues aux fluctuations des nuages électroniques de chacun des atomes (forces de London) [40]. La distribution asymétrique des électrons dans les orbitaux induits de dipôles instantanés. Ces dipôles oscillent et produisent une force attractive : la force de dispersion de London. À très courte distance, la répulsion entre les deux nuages électroniques est très forte.

$$E_{ij} = \sum_i \sum_j -\frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} \quad (19)$$

Il s'agit d'une double somme sur tous les atomes chargés mais ne considérant que les paires d'atomes non-liés par des liens covalents (on exclut les paires prises en considération pour les liens covalents, les angles entre les liens et les angles de torsions).

$r_{ij}$  : La distance entre les deux atomes.

$A_{ij}$  et  $B_{ij}$  constantes de Van Der Waals.

Chaque atome possède un rayon de Van der Waals caractéristique.

La distance de contact ou distance optimale entre deux atomes, c'est-à-dire la distance pour laquelle l'énergie de Van der Waals est minimum, correspond à la somme de leurs rayons de Van der Waals.

### II.3.2.2. Interactions électrostatiques

Enfin le cinquième terme est le potentiel coulombien qui traduit les interactions électrostatiques entre les différentes paires d'atomes [41]. Il s'agit des interactions entre deux atomes portant chacun une charge. L'énergie de cette interaction est décrite par la loi de Coulomb :

$$E_{elect} = \sum \frac{q_i \cdot q_j}{D'_{ij}} \quad (20)$$

### II.3.2.3. Énergie de liaison hydrogène

Les liaisons hydrogène sont le résultat des interactions électrostatiques (70%) et de Van der Waals (30%) entre un atome électronégatif (généralement un atome d'oxygène ou d'azote) portant un doublet d'électron libre et un atome d'hydrogène porté par un atome électronégatif. Les deux atomes sont distants d'environ 3 Å. L'énergie des liaisons hydrogène est de l'ordre de 3 kcal.mol<sup>-1</sup>. Les glucides polaires peuvent ainsi former des liaisons hydrogène entre eux ou avec des molécules d'eau. Ils se dissolvent donc facilement dans l'eau : ils sont hydrophiles.

Les phénomènes de répulsion et de délocalisation électronique interviennent. Plusieurs types de fonctions d'énergie potentielle ont été développés pour tenir compte de la directivité de la liaison hydrogène. Actuellement, les fonctions les plus utilisées permettant d'exprimer ces interactions dans des systèmes moléculaires importants sont souvent simplifiées:

$$\text{La fonction } E_H = A/r_{ij}^{12} - B/r_{ij}^{10}$$

$$\text{La fonction } E_H = A'/r_{ij}^{12} - B'/r_{ij}^6$$

Les coefficients A, B, A', B' sont spécifiques des liaisons hydrogènes [42].

### II.3.3. Quelques exemples de champs de force utilisés dans la mécanique moléculaire

Différents champs de force sont proposés dans la littérature, ils se distinguent les uns des autres par les termes dans le développement de l'expression de l'énergie de la molécule.

Chacun a un domaine d'application spécifique de sorte que le choix d'un champ de force dépend des propriétés et de l'application du système que l'on veut étudier.

Type de composé : carbohydrate, complexe métallique.

Environnement : gaz, solution.

Type d'interaction à étudier : liaison hydrogène, ....

De tels champs de force sont apparus en début des années 1970 et continuent à évoluer aujourd'hui.

- MM2 : il a été développé par Allinger en 1976 et c'est le champ de force le plus utilisé par la communauté des chimistes organiciens [43]. Il a été conçu au début pour les molécules simples (alcanes, alcènes et alcynes non conjugués, les composés carbonylés, les sulfures, les amines...), mais ses versions améliorées MM3 (1989) [44] et MM4 (1996) [45] lui permet de traiter des molécules organiques de plus en plus complexes.
- MM+ : est une extension du champ de force MM2, avec l'ajout de quelques paramètres additionnels [46]. MM+ est un champ de force robuste, il a l'aptitude de prendre en considération les paramètres négligés dans d'autres champs de force et peut donc s'appliquer pour des molécules plus complexes tels que les composés inorganiques [47].
- AMBER (Assisted Model Building with Energy Refinement) : est un champ de force de mécanique moléculaire mis en point par Kollman [48]. Ce champ de force a été paramétrisé pour les protéines et les acides nucléiques. Il a été utilisé pour les polymères et pour d'autres petites molécules.
- OPLS (Optimised Potentials for Liquid Simulation) : comme l'indique son nom, est conçu pour optimiser le potentiel qui permet la description des propriétés de solvation. Il est écrit par W. L. Jorgensen et J. Tirado Rives [49]. Il a été développé par Karplus et al. [50].
- GROMOS (Groningen Molecular Simulation Program Package) : est écrit par Van Gunsteren [51] et conçu spécialement pour les biomolécules en milieu aqueux en vue de l'étude des interactions entre les molécules d'eau et les groupements polaires des protéines.
- CHARMM (Bio+) (Chemistry Harvard Macromolecular Mechanic) : a été développé par R. Brooks et M. Karplus (1983) [52]. Il utilise une fonction d'énergie empirique

pour les systèmes macromoléculaires et molécules biologiques (protéines, acides nucléiques,...). Son concept est semblable à celui d'AMBER.

- SPASIBA (Spectroscopic Potentiel Algorithm for Simulating Biomolecular conformational Adaptability) : a été élaboré par Gérard Vergoten et col. (1995). Il combine les avantages du champ spectroscopique modifié de Urey-Bradley-Shimanouchi [53] et du champ de force de mécanique moléculaire AMBER. Il permet de trouver en même temps les structures, les énergies de conformation et les fréquences vibrationnelles au minimum énergétique d'une molécule [54].
- MMFF94 : est l'un des champs de force de mécanique moléculaire les plus récents, il a été développé par Tom Halgren [55,56]. Les termes d'énergie d'élongation, de valence et de torsion sont les mêmes que ceux de MM3 (expansion de MM2). MMFF94 comprend des termes croisés de liaison-valence.
- UFF : (Universal Force Field). A. K. Rappé, C. J. Casewit et collaborateurs ont essayé de concevoir un champ de force universel [57] capable de simuler des molécules contenant n'importe quelle combinaison d'atomes dans le tableau périodique.

## II.4. Dynamique Moléculaire (DM)

La dynamique moléculaire (DM) est une méthode de simulation du mouvement des atomes et des molécules en calculant leurs déplacements. Cette technique est largement utilisée pour simuler les propriétés des solides, des liquides, et des gaz. Elle permet de calculer les propriétés physico-chimiques d'un système sachant que le potentiel d'interaction est connu.

### II.4.1. Le principe de la Dynamique Moléculaire (DM)

La dynamique moléculaire utilise la seconde loi de Newton pour décrire le mouvement d'une molécule en fonction du temps.

$$\vec{F}_i = m_i \vec{a}_i \quad (21)$$

$\vec{F}_i$  : est le vecteur force agissant sur l'atome i.

$m_i$  : est la masse de l'atome i.

$\vec{a}_i$  : est le vecteur d'accélération de l'atome i.

Cette équation montre que la vitesse et le sens du mouvement atomique dépendent des forces qui s'exercent entre les atomes.

La force  $F_i$  qui s'exerce sur un atome  $i$  se trouvant en position  $r_i(t)$  est déterminée par :

$$\vec{F}_i = - \frac{dE(\vec{r}_i)}{dr_i} \quad (22)$$

L'énergie potentielle totale du système, se calcule en utilisant les techniques de la mécanique moléculaire.

Le protocole typique de simulation de dynamique moléculaire contient :

- La création de la structure moléculaire de départ
- La minimisation de l'énergie
- Le chauffage du système
- L'équilibration du système à la température d'étude
- La production de la trajectoire
- L'analyse de la trajectoire

Typiquement, avant d'effectuer la simulation de dynamique moléculaire, les systèmes ont été chauffés de 0 K à 300 K pendant 20 ps et équilibrés à la même température pendant encore 20 ps. Les simulations ont été effectuées à une température et une pression constantes (300K et 1 atm) dans des boîtes de solvant rectangulaires [58].

## Références bibliographiques

- [1] Leach A. R., *Molecular Modelling. Principles and Applications*, Longman , New York(1996).
- [2] Clark T., ‘‘Handbook of Computational Chemistry’’, Wiley, London (1985).
- [3] Kollmann P., Advances and Continuing Challenges in Achieving Realistic and predictive Simulations of the Properties of Organic and Biological molecules, *Acc. Chem. Res.*, 29 ,462(1996).
- [4] Audouze C., Vers une parallélisation par bandes en chimie quantique, Laboratoire de Mathématique, UMR CNRS 8628, Université Paris-Sud (2003).
- [5] Liotta D., *Advances in Molecular Modeling*, JAI Press Inc, London(1988).
- [6] Dugas H., *Principes de base en modélisation moléculaire, Aspects théoriques et pratiques*, Librairie de l’Université de Montréal, Montréal(1996).
- [7] E. Schrödinger, *Ann. Phys*, 79 ,361-489(1926).
- [8] Epstein v, *The Variation Method in Quantum Chemistry*, *Academic Press*, New-York(1974).
- [9] Born M., Oppenheimer, À la théorie quantique des molécules,*J., Ann. Phys*, 84 ,361(1927).
- [10] Meskaldji S., Thèse de Magistère, Université de Constantine (2008).
- [11] Dirac P. A. M. , *Lectures on Quantum Mechanics*, Yeshiva University, New-York(1964).
- [12] Heisenberg W. , *the Physical Principles of the Quantum Theory*, Dover, New York (1949).
- [13] Landau L. et Lifchitz E., *Mécanique quantique*, Editions Mir, Moscou(1967).
- [14] Marle C. M., *De la mécanique classique a la mécanique quantique : pourquoi et comment quantifier?*, Université Pierre et Marie Curie Paris, France(2003).
- [15] Leach v, *Quantum Mechanical Models*, in *Molecular modelling: Principles and applications*, Addison Wesley Longman Ltd., Harlow(1996).
- [16] E. Schrodinger, *Ann. phys. Leipzig*, 76 ,361(1926).
- [17] Burkert U., Allinger N.L., *Molecular Mechanics*, Editors, ACS Monograph, Washington(1982).
- [18] Born M. et Oppenheimer J. R., On the Quantum Theory of Molecules, *Ann. Phys*, 84 ,457(1927).

- [19] Doucet J. P., Weber J., Computer-aided molecular design: Theory and applications, Academic Press, London(1996).
- [20] Levine I. N., Quantum Chemistry, 5th ed. ,Prentice Hall, New Jersey(1999).
- [21] Hartree D. R., The Wave Mechanics of an Atom with a non-Coulomb Central Field,*Proc. Cambridge Philos. Soc.*, 24 ,89-426(1928).
- [22] Fock V. Z., Méthode d'approximation pour résoudre le problème multi-corps en mécanique quantique, *Z. Phys.* 61 ,126-795(1930).
- [23] Rivail J. L. , Eléments de chimie quantique à l'usage des chimistes, CNRS: Paris(1989).
- [24] Löwdin P.-O., *Advan. Chem. Phys.*, 2 ,207(1959).
- [25] Pople J. A., Binkley J. S. and Seeger R., Theoretical Models Incorporating Electron Correlation, *Int. J. Quant. Chem. Symp.* 10,1(1976).
- [26] Sinanoglu A., Brueckner K. A., Three Approches to Electron Correlation in Atoms ; New Haven, Connecticut : Yale-University Press (1970).
- [27] Møller C., Plesset M. S., Note on an Approximation Treatment for Many-Electron Systems, *Phys. Rev.*, 46 ,61(1934).
- [28] Ripka W.C. & Blaney J.M., Computer graphics and molecular modeling in the analysis of synthetic targets, *Eds .Eliel. & Wilen.*, 20 ,85(1991).
- [29] Thomas L. H., The effect of the orbital velocity of the electrons in heavy atoms on their stopping of  $\alpha$ -particles,*Proc. Cambridge Phil. Soc.*, 23 , 713(1927).
- [30] Fermi E., Statistical method to determine some properties of atoms ,*Rendiconti*, 6 ,602(1927).
- [31] Dirac P. A. M., Note on Exchange Phenomena in the Thomas Atom, *Proc. Cambridge Phil. Soc.*, 26 ,376(1930).
- [32] Wigner E., On the Quantum Correction For Thermodynamic Equilibrium, *Phys. Rev.*, 40 ,749( 1932).
- [33] von Weisacker C. F., *Z. Phys*, 96,431(1935).

- [34] Hohenberg P., Kohn W., Inhomogeneous Electron Gas, *Phys. Rev. B*, 136, 864(1964).
- [35] Kohn W., Sham L. J., Self-Consistent Equations Including Exchange and Correlation Effects, *Phys. Rev. A*, 140, 1133(1965).
- [36] Sordo J. A., Probst M., Corongiu G., Chin S. and Clementi E.,  
Ab initio pair potentials for the interactions between aliphatic amino acids, *J. Am. Chem. Soc.*, 109, 1708(1987).
- [37] Andrews D.H., the Relation Between the Raman Spectra and the Structure of Organic Molecules, *Phys. Rev.*, 36, 544(1930).
- [38] Allinger N.L., Molecular Mechanics, *Adv. Phys. Org. Chem.*, 13, 1(1976).
- [39] Young D.C., Computational chemistry- A Practical Guide for Applying Techniques to Real-World Problems- A John Wiley & Sons, Inc., New York (2001).
- [40] Burkert U., Allinger N.L., Molecular Mechanics, ACS Monograph 177. American Chemical Society, Washington DC(1982).
- [41] Dugas H., Principes de base en modélisation moléculaire, Aspects théorique et pratiques, Chapitre 3 introduction aux méthodes de minimisation d'énergie, quatrième édition, Librairie de L'Université de Montréal(1996).
- [42] Dugas H., Principes de base en modélisation moléculaire, Aspects théoriques et pratiques, Hermann Dugas, quatrième édition, librairie de l'université de Montréal(1996).
- [43] Allinger N.L., Q. C. P. E., 3, 32(1983).
- [44] Allinger N.L., Yuh Y.H., Lii J-H., Molecular mechanics. The MM3 force field for hydrocarbons. 1, *J. Am. Chem. Soc.*, 111, 8522(1989).
- [45] Allinger N.L., Chen K., Lii J-H., An improved force field (MM4) for saturated hydrocarbons, *J. Comp. Chem.*, 17, 642 (1996).
- [46] Jones G.B., Chapman B.J.,  $\pi$  stacking effects in asymmetric synthesis, *Synthesis*. 475(1995).
- [47] Hocquet A., Langgard M., An Evaluation of the MM+ Force Field, *J. Mol. Model.*, 4, 94(1998).
- [48] Weiner S.J., Kollman P.A., Nguyent T., Case D.A., An all atom force field

- for simulations of proteins and nucleic acids, *J. Comput. Chem.*, 7, 230(1986).
- [49] Jorgensen W.L., Rives J.T., The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin, *J. Am. Chem. Soc.*, 110, 1657(1988).
- [50] Smith J.C. & Karplus M., Empirical force field study of geometries and conformational transitions of some organic molecules, *J. Am. Chem. Soc.*, 114, 805(1992).
- [51] Van Gusten W.F., Karplus M., Effect of constraints on the dynamics of macromolecules, *Macromolecules*, 15, 1528(1982).
- [52] Brooks B.R., Bruccoleri R.E, Ofalson B.D., States D.J., Swaminathan S., Karplus M., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations\*, *J. Comp. Chem.*, 4, 187(1983).
- [53] Shimanouchi T., Force constants of small molecules, *Pure Appl. Chem.*, 7, 131(1963).
- [54] Tristram F., Durier V., Vergoton G., *J. Mol. Struct.*, 378, 249(1996).
- [55] Halgren T.A., Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94, *J. Comp. Chem.*, 17, 490(1996).
- [56] Halgren T.A., Nachbar R.B., Merck molecular force field II, *J. Comp. Chem.*, 17, 587(1996).
- [57] Rappé A.K., Casewit C.J., Kolwell K.S., Skiff W.M., UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, *J. Am. Chem. Soc.*, 114, 10028(1992).
- [58] E. Schrodinger, Quantization as an eigenvalue problem, *Ann. phys. Leipzig*, 76, 361 (1926).

*Chapitre III :*  
*Etude de la relation*  
*structure-*  
*propriété/activité*

### III.1. Introduction

Plusieurs techniques *in silico* sont utilisés dans le processus dans la découverte et la conception des médicaments. La principale technique est les méthodes QSAR, qui étudient les relations qualitatives et quantitatives des structure-activité (SAR/QSAR).

QSAR sont des tentatives pour corrélérer entre les paramètres physico-chimiques de la structure chimique et leur activité biologique. La reconnaissance que les propriétés ont un effet majeur sur la réalisation d'expériences biologiques de recherche de médicaments.

La comparaison de l'activité biologique de certaines molécules et de leurs structures a permis d'établir dans de nombreux cas des corrélations entre les paramètres structuraux et les propriétés d'une molécule [1,2].

De nombreuses corrélations ont été établies entre l'activité biologique d'une molécule et la surface accessible de certains de ses substituant non polaires. La surface accessible et la lipophilie de groupements non polaires sont très étroitement liées [3-5].

### III.2. Historique

L'exemple le plus ancien de RSA remonte à 1868, lorsque deux scientifiques britanniques, Alexander Crum-Brown et Thomas Fraser, publièrent une étude sur la relation entre la structure chimique et l'activité physiologique. Dans les décennies qui suivirent, Richet établit une relation entre la toxicité et les propriétés physicochimiques d'une série de composés organiques polaires (éthers, alcools, cétones), indiquant que « plus ils sont solubles, moins ils sont toxiques » [6,7]. De manière indépendante, Meyer et Overton ont décrit une corrélation linéaire entre la lipophilie (coefficient de partage huile-eau) et les effets biologiques (narcotiques), « la théorie lipoïde de la narcose » [8,9].

En 1937, Hammett introduisit une méthode pour séparer les effets des substituants (propriétés donneuses ou capteuses d'électrons) sur les mécanismes des réactions de chimie organique [10]. Quelques années plus tard, Taft améliora le modèle de Hammett en proposant le premier paramètre stérique,  $E_s$  [11].

En 1962, Hansch et al. ont publié leur étude sur la corrélation de l'activité biologique de substances régulatrices de la croissance végétale (acides phénoxyacétiques monosubstitués sur le cycle benzénique) avec les constantes de substitution de Hammett et l'hydrophobicité [12].

Deux ans plus tard, le même groupe de recherche introduisit un nouveau paramètre,  $\pi$  (hydrophobicité relative d'un substituant), qui peut être défini comme suit :

$$\pi_X = \log P_X - \log P_H \quad (1)$$

Où  $\pi_X$  représente l'hydrophobicité relative d'un substituant, et  $P_X$  et  $P_H$  les coefficients de partage respectifs de la molécule dérivée et de la molécule parente [13]. Cette méthode est à présent connue sous le terme de QSAR bidimensionnelle de Hansch-Fujita ou 2D-QSAR [14,15].

En 1964, Free et Wilson ont suggéré l'approche dite "*de novo*" qui postule que chaque substituant apporte un effet additif et constant à l'activité biologique, quels que soient les autres substituants présents dans le reste de la molécule [16,17]:

$$BA = \sum a_i x_i + u \quad (2)$$

où "BA" représente l'activité biologique, "u" la contribution moyenne de la molécule parente,  $a_i$  la contribution de chaque motif structural et  $x_i$  la présence ( $x_i = 1$ ) ou l'absence ( $x_i = 0$ ) d'un motif structural particulier.

Quelques années plus tard, les indices topologiques décrivant l'intégralité de la structure moléculaire ont été introduits. L'origine de cette terminologie remonte au début des années 1970 (Hosoya, 1971) [18], bien que les premiers travaux sur les descripteurs constitutionnels numériques aient démarré pratiquement 25 ans en arrière avec Wiener (Wiener, 1947) [19]. Ces indices topologiques de première génération étaient des nombres entiers basés sur des invariants vertex locaux entiers, et intéressaient essentiellement les propriétés physiques des hydrocarbures saturés (alcanes). Parmi d'autres travaux importants, les indices topologiques de Randić (*molecular branching* ; Randić, 1975) [20], de Balaban (Balaban, 1982) [21], la différence topologique minimale de Simon (*Minimum Topological Difference, MTD* ; Simon, 1974) [22] et l'approche de connectivité moléculaire selon Kier et Hall (Kier, 1976 ; Hall, 1977 ; Kier, 1986) [23] constituèrent le socle nécessaire à l'avènement de l'ère moderne des RSAQ / 3D-RSAQ.

### III.3. Outils et techniques de QSAR

#### III.3.1. Paramètres biologiques

Les données biologiques sont habituellement exprimées 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. Des exemples de données biochimiques ou biologiques, utilisés dans l'analyse de QSAR, sont décrits dans le tableau III.1 [24].

**Tableau III.1:** Types de données biologiques utilisées dans l'analyse QSAR.

Source d'activité	Paramètres biologiques
<b>1. Récepteurs isolés</b>	
Constante de vitesse	$\text{Log } k$
Constante de Michaelis-Menten	$\text{Log } I/K_m$
Constante d'inhibition	$\text{Log } I/K_i$
<b>2. Systèmes cellulaires</b>	
Constante d'inhibition	$\text{Log } I/IC_{50}$
Résistance croisée	$\text{Log } CR$
Données biologiques <i>in vitro</i>	$\text{Log } I/C$
Mutation de gène	$\text{Log } TA_{98}$
<b>3. Systèmes <i>in vivo</i></b>	
Facteur de bioconcentration	$\text{Log } BCF$
Vitesses de la réaction <i>in vivo</i>	$\text{Log } I$ (induction)
Vitesses pharmacodynamiques	$\text{Log } T$ (clairance totale)

#### III.3.2. Descripteurs moléculaires

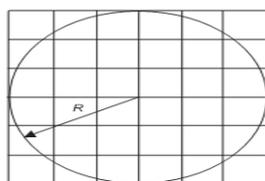
Les descripteurs moléculaires sont des propriétés numériques qui peuvent être calculées à partir de la table de connectivité d'une molécule ou d'une représentation planaire (2D) de la

structure. Ils sont basés sur les éléments présents, les charges partielles, la nature des liaisons, mais n'exploitent pas les coordonnées atomiques spatiales [25]. On peut calculer ainsi les paramètres caractérisant la structure moléculaire:

### III.3.2.1. Volume et surface moléculaire

Le volume moléculaire est souvent calculé par la méthode des grilles [26]. Le volume d'un atome est déterminé à partir de rayon de Van der Waals  $R$  donné par  $\frac{4}{3}\pi R^3$ .

La figure III.1 donne une représentation d'un atome dont le rayon de Van der Waals est  $R$ , entouré d'une grille tridimensionnelle avec même distance entre les points. Le centre de cette grille est le centre de l'atome et les bords de la grille correspondent au rayon de Van der Waals.

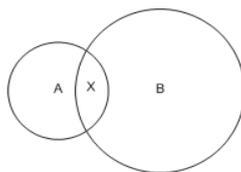


**Figure III.1:** Atome entouré par une grille.

Pour chaque point de la grille, on calcule sa distance par rapport au centre et on détermine si le point de la grille se trouve à l'intérieur ou à l'extérieur de l'atome. Si  $n$  est le nombre total de points de grille et  $n_a$  le nombre de points qui se trouvent à l'intérieur de l'atome, le volume est donnée par :

$$\frac{V}{8R^3} = \frac{n_a}{n} \quad (3)$$

Pour un système polyatomique, nous devons donner une attention particulière aux points de la grille qui se trouvent dans la zone de recouvrement. Figure III.2 montre deux atomes A et B, avec des rayons  $R_A$  et  $R_B$ . La zone de recouvrement est marquée par X.



**Figure III.2:** Atomes A, B et la zone de recouvrement X.

Nous savons que le volume de l'atome A est donné par  $\frac{4}{3}\pi R^3$ . Nous entourons maintenant l'atome B par une grille comme décrit ci-dessus, et nous testons chaque point de la grille à son tour. Si le point de la grille se trouve dans la sphère B, nous testons pour voir si elle se trouve dans la région X et donc déjà a été considérée comme faisant partie du volume d'atome A. L'algorithme se produit jusqu'à ce que tous les atomes aient été testés. Le volume moléculaire est obtenu en additionnant toutes les contributions atomiques. Il existe des méthodes similaires pour l'estimation de la surface moléculaire.

### III.3.2.2. Polarisabilité

La polarisabilité d'une molécule est définie par la capacité de son système électronique de se déformer par un champ électrique externe, il joue un rôle important dans la modélisation de nombreuses propriétés moléculaires et des activités biologiques [27].

Une molécule très polarisée peut donner des attractions fortes avec d'autres molécules. La polarisabilité d'une molécule peut également améliorer sa solubilité aqueuse [28].

### III.3.2.3. Réfractivité molaire

La réfractivité molaire (MR) est un critère important dans la mesure des facteurs stériques. Elle est généralement désignée comme une simple mesure du volume occupé soit par un atome individuel ou par un groupe d'atomes [29]. La réfractivité molaire est particulièrement importante dans une situation où le substituant possède les électrons  $\pi$  ou les doublets électroniques libres. La réfractivité molaire est donnée par la relation suivante :

$$MR = \frac{(n^2-1)}{(n^2+2)} \cdot \frac{MW}{d} \quad (4)$$

Où

n : indice de réfraction.

MW : masse moléculaire.

d: densité.

MW/d : volume.

$(n^2 - 1) / (n^2 + 2)$  : facteur correctif.

La détermination expérimentale de la polarisabilité est basée sur l'équation de Lorentz-Lorentz [30]. Cette équation établit une relation entre l'indice de réfraction d'une molécule et sa polarisabilité ( $\alpha$ ). La forme générale de cette équation est donnée par :

$$MR = \frac{(n^2-1)}{(n^2+2)} \cdot \frac{MW}{d} = \frac{4\pi}{3} N\alpha \quad (5)$$

### III.3.2.4. Energie d'hydratation

La liaison hydrogène (ou liaison H) est de type électrostatique (charge partielle, dipôle) et stérique entre deux groupements d'une même molécule ou de deux molécules voisines. La liaison hydrogène joue un rôle primordial dans la solubilité des molécules médicamenteuses et leurs interactions avec les récepteurs biologiques [3].

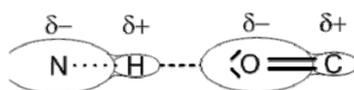


Figure III.3 : La liaison hydrogène.

Dans le figure III.3, l'oxygène (O) est appelé l'accepteur (accepteur de proton H<sup>+</sup>) et l'azote (NH) est appelé le donneur (donneur de proton H<sup>+</sup>).

Les liaisons hydrogènes sont à la fois plus longues et plus faibles que les liaisons covalentes établies entre les mêmes atomes. La solubilité des substances dans un milieu aqueux dépend fortement de leur capacité de former des liaisons hydrogènes avec l'eau.

Hine et Mookerjee ont proposé un modèle qui considère que l'énergie libre des interactions entre les petites molécules et l'eau peut être estimée par une fonction additive de leurs groupes constitutifs [31]. Pour les petites molécules, dont tous les fragments accessibles au solvant, l'énergie libre d'hydratation est:

$$\Delta G_{Hyd} = \sum_{i=1}^N \Delta G_{fr_i} \quad (6)$$

Où  $\Delta G_{fr_i}$  est l'énergie libre d'hydratation du fragment  $i$  et  $N$  est le nombre total de fragments dans la molécule.

Dans le modèle proposé, les systèmes aromatiques sont traités dans une représentation de Kekulé pour l'analyse fragmentaire, mais l'énergie obtenue avec les groupes C=C doit être corrigée comme suit:

$$\Delta G_{Hyd_{aro}} = n \times \Delta G_{C=C} - m \times 1.8 \quad (7)$$

Où  $n$  est le nombre de doubles liaisons dans la représentation de Kekulé, et  $m$  le nombre des centres aromatiques du système, par exemple:

Dans le benzène  $n = 3$  et  $m = 1$ .

Dans le naphthalène  $n = 5$  et  $m = 2$ .

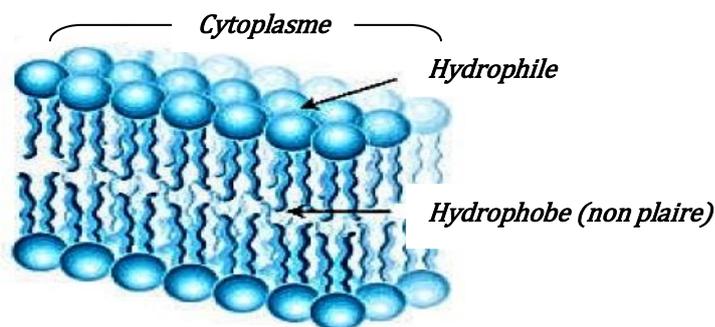
Ce qui donne les valeurs  $-1.05$  et  $-2.1$  kcal/mole de l'énergie d'hydratation pour ces molécules respectivement.

### III.3.2.5. Coefficient de partition (Log P)

Le partage d'une molécule entre une phase aqueuse et une phase lipidique conditionne en partie ses propriétés biologiques telles que le transport, le passage à travers les membranes, la biodisponibilité (distribution et accumulation), l'affinité pour un récepteur et la fixation par une protéine, l'activité pharmacologique ou encore la toxicité. Depuis les travaux de Collander à la fin des années 1950, puis ceux du groupe de Hansch quelques années plus tard, le coefficient de partage  $P$  d'une molécule dans un système bi-phasique constitué de deux solvants non-miscibles (le plus souvent le système  $n$ -octanol/eau), est reconnu pour sa faculté à mimer le passage de cette molécule à travers les membranes biologiques. Pour des solutions diluées, ce coefficient de partage  $n$ -octanol/eau est le rapport de la concentration d'une molécule de soluté dans le  $n$ -octanol sur sa concentration dans l'eau lorsque le système biphasique est en équilibre [32].

Log  $P$ , est l'un des paramètres largement utilisé dans des études de relations structure-activité quantitatives (QSARs) dans les sciences pharmaceutiques, biochimiques, toxicologiques et dans les sciences de l'environnement. La lipophilie intéresse donc tout autant la communauté qui étudie les problèmes de santé humaine que celle qui est impliquée dans les problèmes de l'environnement.

Un médicament doit être caractérisé par une balance correcte entre ses propriétés hydrophiles et hydrophobes. Pendant leur parcours à travers du corps, les molécules médicamenteuses doivent se dissoudre dans les milieux aqueux de cytoplasme et des liquides extracellulaires ainsi que dans le milieu apolaires de la membrane cellulaire (Figure III.4). Ainsi, la solubilité dans des milieux à la fois polaires et apolaires joue un rôle important dans la disponibilité et de la distribution des médicaments.



**Figure III.4** : Polarité des différents milieux cellulaires.

Le coefficient de partage d'une substance chimique entre deux phases non miscibles est le rapport des concentrations à l'équilibre du composé dans la phase organique apolaire et la phase aqueuse.

Le solvant organique couramment utilisés pour mesurer expérimentalement le coefficient de partage des molécules médicamenteuses est le 1-octanol. Le coefficient de partage  $P$  se calcule de la manière suivante:

$$\text{Log } P = \frac{\text{Concentration du composé en solution dans l'octanol}}{\text{Concentration du composé en solution dans l'eau}} \quad (8)$$

La phase aqueuse est habituellement tamponnée à pH 7,4.  $\text{Log } P$ , est souvent appelée la constante de lipophilie d'une molécule. Une valeur positive pour  $\text{log } P$ , ( $P > 1$ ), indique que le composé est lipophile. Une valeur négative pour  $\text{log } P$ , ( $P < 1$ ) indique que le composé est hydrophile.

La plupart des méthodes expérimentales de détermination de  $\text{log } P$  souffrent du même inconvénient, à savoir que leur domaine d'application est relativement étroit. D'autre part, du fait de la nature intrinsèque de certaines molécules, leurs  $\text{log } P$  sont inaccessibles à l'expérience. C'est le cas en particulier des surfactants qui ont tendance à s'accumuler à l'interface du système bi-phasique au lieu de se disperser dans les deux phases [32]. Enfin, dans le domaine de la conception assistée par ordinateur ou dans le domaine de la chimie combinatoire, les chercheurs travaillent sur des modèles moléculaires avant même que les molécules aient été synthétisées. Ceci explique le succès des nombreuses méthodes d'estimation de  $\text{log } P$  qui ont été décrites dans la littérature depuis plus de trente ans. Les plus anciennes sont des méthodes fragmentables dans lesquelles une molécule est divisée en

fragments prédéfinis et les contributions correspondantes sont sommées pour conduire à une valeur estimée du log P [33].

Parmi les méthodes utilisées dans l'estimation de log P, on trouve la méthode Hansch. Cette méthode repose sur l'utilisation de paramètre d'hydrophobicité d'un substituant qui mesure l'intensité hydrophobe du substituant en question par rapport à celle de l'hydrogène.

On peut mesurer log P expérimentalement pour un composé donné sans et avec le substituant X en utilisant l'équation suivante:

$$\Pi_x = \log P_x - \log P_H \quad (9)$$

Une valeur positive de  $\pi$  indique que le substituant est plus hydrophobe que l'hydrogène. Une valeur négative signifie que moins hydrophobe que l'hydrogène.

La méthode de Rekker permet de calculer le coefficient de partage octanol/ eau. Elle consiste à décomposer la molécule en fragments à chacun desquels on attribue une lipophilie  $f$  appelée constante fragmentale hydrophobe. La formule générale de Rekker est donnée par :

$$\log P = \sum f + k.CM \quad (10)$$

Avec  $k$  entier et  $CM = 0,219$  (l'entier  $k$  dépend du type d'interaction).

### III.3.3. Modèles statistiques

L'objectif de l'analyse statistique est justement de « démêler » ces descripteurs et d'identifier ceux qui sont corrélés à la variable cible, qui produisent du signal, de ceux qui ne le sont pas, qui produisent du bruit. L'analyse statistique permet également d'identifier les descripteurs qui sont corrélés entre eux pour ne garder que les principaux et réduire ainsi la redondance d'informations.

Dans la terminologie des statistiques, la propriété étudiée est appelée variable cible ou variable dépendante ou encore variable Y. C'est la caractéristique que l'on cherche à expliquer et à optimiser. Les descripteurs, en l'occurrence chimiques, sont appelés variables explicatives, variables indépendantes ou encore variables X.

L'analyse statistique détermine et quantifie les corrélations entre les descripteurs et la variable cible. Elle indique également la contribution relative de chaque descripteur dans l'explication globale de l'activité. Le modèle statistique est une équation donnant la valeur de la variable

cible en fonction de la somme des valeurs pondérées des descripteurs. Les principaux outils statistiques pour obtenir un modèle sont :

- la régression linéaire multivariée (*Multivariate Linear Regression* - MLR) ;
- la régression en composante principale (*Principale Component Regression* - PCR) ;
- la régression des moindres carrés partiels (*Partial Least Squares* - PLS) [34] ;
- les réseaux de neurones artificiels (*Artificial Neural Network* - ANN).

### III.4. Régression linéaire multiple

Une analyse de régression où la variable dépendante  $Y$  dépend linéairement de plusieurs variables indépendantes  $X_1, X_2, \dots, X_j$  est appelée régression linéaire multiple.

L'équation de régression linéaire multiple est de la forme :

$Y = f(X_1, X_2, \dots, X_j)$  où  $f(X_1, X_2, \dots, X_j)$  est une fonction linéaire de  $X_1, X_2, \dots, X_j$  [35].

Le modèle mathématique peut se présenter sous la forme :

$$Y^i = a + \sum_{j=1}^{j=p} b_j x_j^i \quad (11)$$

où  $b_j$  sont les coefficients de régression et  $a$  la constante de régression.

Considérons un modèle à  $N$  individus,  $p$  variables  $X_j$ , où  $Y_{obs}$  et  $Y_{calc}$  sont respectivement les valeurs observées et calculées de la propriété et  $\bar{Y}$  la valeur moyenne de la variable à expliquer.

#### III.4.1. Description de la méthode

La régression multilinéaire est la méthode la plus simple et la plus employée pour le développement de modèles prédictifs [36]. Elle repose sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante  $Y$  (ici, la propriété) et une série de  $p$  variables indépendantes  $X_i$  (ici, les descripteurs). L'objectif est d'obtenir une équation de la forme suivante :

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p + \varepsilon \quad (12)$$

Où  $\beta_i$  sont les coefficients de la régression.

La détermination de l'équation (2) se fait alors à partir d'une base de données de  $n$  échantillons pour laquelle à la fois les variables dépendantes et la variable indépendante sont connues. Il s'agit donc de considérer un système de  $p$  équations.

$$\begin{aligned}
 Y_1 &= \beta_0 + \beta_1 X_{1,1} + \beta_2 X_{2,1} + \dots + \beta_p X_{n,1} + \varepsilon_1 \\
 &\vdots \\
 Y_2 &= \beta_0 + \beta_1 X_{1,2} + \beta_2 X_{2,2} + \dots + \beta_p X_{n,2} + \varepsilon_2 \\
 &\vdots \\
 Y_n &= \beta_0 + \beta_1 X_{1,p} + \beta_2 X_{2,p} + \dots + \beta_p X_{n,p} + \varepsilon_n
 \end{aligned}
 \tag{13}$$

Où les résidus  $\varepsilon_i$  représentent l'erreur du modèle, constituée par l'incertitude sur la variable dépendante  $Y_i$  d'une part, sur les variables indépendantes  $x_i$  d'autre part, mais aussi par les informations contenues dans les variables indépendantes mais non exprimées via les variables dépendantes.

Ce système d'équations peut être écrit sous la forme matricielle suivante :

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \mathbf{1} & X_{1,1} & X_{2,1} & \dots & X_{n,1} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \mathbf{1} & X_{n,1} & X_{n,2} & \dots & X_{n,p} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}
 \tag{14}$$

la matrice  $X(n,p)$  comporte dans sa première colonne un vecteur constitué uniquement de 1. Ce vecteur correspond à la constante  $X_0$ . La matrice  $X$  est alors de dimension  $(n, p+1)$ . Le modèle s'écrit alors sous forme  $Y = X b + \varepsilon$ .

La méthode consiste alors à choisir les coefficients du vecteur  $\beta$  en faisant en sorte de minimiser la somme des carrés des écarts entre les valeurs prédites et les valeurs réelles sur l'intégralité de la base de données et ceci sous couvert de certaines hypothèses de départ.

En premier lieu, les variables indépendantes  $X_i$ , comme leur nom l'indique, sont supposées indépendantes entre elles et leur incertitude est négligeable. Ensuite, les différents échantillons  $Y_i$  sont supposés indépendants entre eux. Enfin, par nature, la dépendance de  $y$  vis-à-vis des  $x_i$  est supposée linéaire.

La valeur prédite de la variable dépendante est alors :

$$\hat{Y} = \beta_0 + \hat{\beta}_1 X_{1,i} + \dots + \hat{\beta}_p X_{n,i}
 \tag{15}$$

Les résidus peuvent donc être définis comme la différence entre les valeurs prédites et observées de  $Y$ .

$$\varepsilon = Y_i + \hat{Y}_i \quad (16)$$

Il s'agit alors de trouver les coefficients  $\hat{B}_i$  afin de minimiser la somme des carrés de ces résidus (*RSS*) pour l'intégralité de la base de données.

#### III.4.2. Confiance accordée aux résultats

La logique statistique nous conduit à effectuer des tests d'hypothèses, et à déterminer des intervalles de confiance autour des valeurs estimées. Successivement, on va chercher à:

- tester la signification globale de la régression,
- tester l'hypothèse nulle  $\beta_l=0$  et à calculer l'intervalle de confiance de  $\beta_l$ ,
- tester l'hypothèse nulle  $\beta_0=0$  et à calculer l'intervalle de confiance de  $\beta_0$ ,
- calculer la précision de l'estimation de  $y$  pour la moyenne et pour une observation individuelle.

#### III.4.3. Test de la signification globale de la régression

##### a. Coefficient de détermination ( $R^2$ )

Le coefficient de détermination  $R^2$  est la mesure de degré de liaison entre  $Y_N$  et  $X_j$ .

$R^2$  est déterminé par les différentes relations suivantes :

$$R^2 = \frac{ESS}{TSS} = \frac{TSS-RSS}{TSS} = 1 - \frac{RSS}{TSS} \quad (17)$$

Un bon ajustement correspondra à un  $R^2$  proche de l'unité.

##### b. Coefficient de corrélation ( $R$ )

C'est le coefficient de corrélation de Bravais Pearson entre  $Y$  et  $Y'$ , c'est à dire entre valeurs observées et prédites par le modèle de régression, il est noté  $R$ , sa valeur variée entre 0 et 1.

##### c. Test Fisher-Snedecor ( $F$ )

Le test Fisher permet de justifier la liaison globale entre  $Y_n$  et  $X_n$ , une version dérivée de cet indicateur peut juger du degré de pertinence des variables du modèle. Il s'agit de vérifier pour chaque variable  $X_j$  si, lorsqu'on passe du modèle complet à  $p$  prédicteurs au modèle simplifié obtenu en  $t$  mesure en fait l'apport marginal de variable  $X_j$  à l'explication de  $Y_n$ .

On peut définir quelques paramètres utilisés dans la régression.

$$\text{Somme des carrés totaux :} \quad TSS = \sum (Y_{obs} - \bar{Y})^2 \quad (18)$$

$$\text{Somme des carrés expliqués :} \quad ESS = \sum (Y_{cal} - \bar{Y})^2 \quad (19)$$

$$\text{Somme des carrés résiduels :} \quad RSS = \sum (Y_{obs} - Y_{cal})^2 \quad (20)$$

$$\text{Ainsi,} \quad TSS = ESS + RSS \quad (21)$$

F peut s'écrire comme suit :

$$F = \frac{ESS}{P} \frac{n-p-1}{RSS} \quad (22)$$

La forme de l'équation (8) représente le nombre de degrés de liberté associé avec chaque paramètre. Le *ESS* associé avec *p* degrés de liberté et le *RSS* associé n-p-1 degrés de liberté.

Le test de Fisher mesure le rapport entre la variance de la variable dépendante expliquée et non expliquée par le modèle de régression. En d'autres termes, le test Fisher permet de tester l'hypothèse nulle selon laquelle chaque *b* est significativement différent de zéro, ce qui est signe d'une relation évidente entre la variable expliquée et les variables explicatives.

Intuitivement, nous rejetterons l'hypothèse nulle lorsque la somme des carrés expliquée par la régression est grande. En d'autres termes, la région critique de ce test est de la forme ( $F > \text{seuil}$ ). Si la quantité *F* observée dépasse le seuil, on rejette l'hypothèse  $H_0$  dans le cas contraire, on conserve  $H_0$ .

Pour éviter de raisonner sur *F*, le programme fournit la *p*-value associée au *F* observé.

La *p*-value est le niveau de significativité du test de Fisher-Snedecor, c'est-à-dire la probabilité de dépasser le *F* observé si l'hypothèse nulle est vraie. On compare la *p*-value au risque  $\alpha$  choisi (par exemple  $\alpha=0.05$ ). Si  $p\text{-value} \leq \alpha$ , alors on rejette l'hypothèse nulle  $\beta_1 = \dots = \beta_p = 0$ .

Ces résultats permettent d'interpréter les tables d'analyse de variance complètes fournies par tout logiciel mettant en œuvre la régression linéaire. La table complète est du type suivant :

Tableau III.2 : Table d'analyse de variance.

Source de variation	Somme des carrés	Degrés de liberté	Carrés moyens
Expliquée	$ESS = \sum(Y_{cal} - \bar{Y})^2$	$P$	$EMS = \frac{ESS}{P}$
Résiduelle	$RSS = \sum(Y_{obs} - Y_{cal})^2$	$n - p - 1$	$RMS = \frac{RSS}{n - P - 1}$
Totale	$TSS = \sum(Y_{obs} - \bar{Y})^2$	$n - 1$	

### c. Ecart type (s)

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

$$s = \sqrt{\frac{RSS}{n-p-1}} \quad (23)$$

Dont  $p$  est le nombre de variables indépendants.

### d- Coefficient de prédiction ( $Q^2$ )

Si  $R^2$  est une mesure de la justesse de la corrélation,  $Q^2$  est une mesure de la justesse de la prédiction. Il mesure la capacité prédictive d'un modèle et donnée par :

$$Q^2 = 1 - \frac{PRESS}{SSY} \quad (24)$$

#### III.4.4. Validations des modèles

Quel que soit le but ultime du modèle QSAR, il doit être obligatoirement validé avant d'être interprété ou utilisé à des fins prédictives. Il existe différentes solutions pour s'assurer de validité d'un modèle. La validation croisée *LOO* (leave-one-out) est un processus qui permet de tester la précision prédictive d'un modèle. Cette méthode est basée sur le calcul de quelques paramètres statistiques tels que : la somme des carrés résiduels  $RSS$  ( $PRESS$ ), la somme des carrés totaux  $TSS$ , la qualité d'ajustement  $R^2_{adj}$ , le coefficient de corrélation de validation croisée  $R^2_{CV}$ , la validation standard des erreurs de prédiction ( $S_{PRESS}$ ) et l'erreur de prédiction ( $PE$ ). Ces paramètres statistiques sont calculés à partir des relations suivantes :

$$PRESS = \sum (Y_{obs} - Y_{cal})^2 \quad (25)$$

$$TSS = \sum (Y_{obs} - \bar{Y})^2 \quad (26)$$

$$R^2_{adj} = (1 - r^2) \left( \frac{n-1}{n-p-1} \right) \quad (27)$$

$$R^2_{CV} = 1 - \frac{PRESS}{TSS} \quad (28)$$

$$S_{PRESS} = \sqrt{\frac{PRESS}{n}} \quad (29)$$

$$PE = 0.6745 (1 - r^2) / \sqrt{n} \quad (30)$$

### III.5. Optimisation multi-paramètres (MPO)

La découverte de nouveaux médicaments ( drug discovery ) est un processus extrêmement long et fastidieux. Les différentes étapes sont illustrées de façon schématique dans la figure III.5 Au cours des phases successives, des milliers de molécules doivent être triées et sélectionnées, pour aboutir à un nombre très restreint de candidats. Pour chaque nouveau médicament, le temps passe depuis la première étape, au cours de laquelle une cible biologique pertinente dans un processus pathologique donné est identifiée, jusqu'à la mise sur le marché d'un composé tourne autour d'une moyenne d'environ 12 à 14 ans.



**Figure III.5 :** Etapes du processus de découverte des médicaments.

Quelle que soit l'origine des composés criblés, la construction d'une base de molécules implique généralement l'application d'un certain nombre de filtres employés pour s'assurer que les molécules possèdent des propriétés compatibles avec une application biologique.

Les activités de découverte de médicaments produisent des volumes toujours plus importants de données complexes qui comportent des niveaux d'incertitude importants. Les méthodes d'optimisation multi paramètres permettent de mieux utiliser ces données pour cibler

rapidement des composés ayant un bon équilibre des propriétés, mais elles ont toutes leurs forces et leurs faiblesses [37]. Par conséquent, nous pouvons utiliser les méthodes d'optimisation multiparamétriques (MPO) pour prédire le meilleur équilibre des propriétés, parmi ces méthodes, nous appliquons des règles de base. Les plus couramment utilisés sont les règles de Lipinski et Veber [38]. D'autre part les méthodes métriques les plus anciennes et les plus couramment utilisées sont l'efficacité du ligand (LE) et l'efficacité lipophile (LipE) .

### III.5.1. Les règles de lipinski

Le médicament apparaît comme un paradigme prometteur pour coder l'équilibre entre les propriétés moléculaires d'un composé qui influence sa pharmacodynamie et sa pharmacocinétique et optimise finalement leur absorption, leur distribution, leur métabolisme et leur excrétion (ADME) dans le corps. Les conditions empiriques pour satisfaire la règle de Lipinski et manifester une bonne biodisponibilité orale impliquent un équilibre entre la solubilité aqueuse d'un composé et sa capacité à diffuser passivement à travers les différentes barrières biologiques [39]. Cette règle est la plus utilisée pour l'identification des composés « drug-like ».

Lipinski a défini un ensemble de règles permettant d'estimer la biodisponibilité d'un composé par voie orale à partir de sa structure bidimensionnelle (2D). Ces règles concernant les propriétés physico-chimiques ont été définies après l'analyse de 2245 médicaments commercialisés ou en phase finale de développement:

- Le poids moléculaire du composé ne doit pas être supérieur à 500 u.m.a.
- Le logarithme décimal du coefficient de partage eau / 1-octanol, noté Log P, doit être inférieur à 5.
- Le nombre de donneurs de liaisons hydrogène doit être inférieur à 5.
- Le nombre d'accepteurs de liaisons hydrogène doit être inférieur à 10.

Les composés dont les propriétés physico-chimiques ne satisfont pas au moins 2 des règles sont fortement susceptibles de présenter des problèmes d'absorption ou de perméabilité.

La « règle des 5 » a été mise au point à partir de composés administrables par voie orale ayant passé avec succès la phase II des tests cliniques. Ce n'est donc pas une méthode pour distinguer les composés étant potentiellement des médicaments de ceux n'en étant pas, mais plutôt une méthode pour identifier les composés ayant une faible absorption ou une faible perméabilité.

### III.5.2. Les règles de Veber

Les règles Veber suggèrent que la flexibilité moléculaire et la surface polaire (PSA) sont des déterminants importants de la biodisponibilité orale. Il est bien connu qu'une biodisponibilité orale élevée est un facteur important pour le développement de molécules bioactives en tant qu'agents thérapeutiques. La réduction de la flexibilité moléculaire (mesurée par le nombre de liaisons rotatives) et la faible surface polaire sont des prédicateurs importants de la bonne biodisponibilité orale [40].

Pour une biodisponibilité orale idéale, il existe deux autres descripteurs identifiés par Veber et al :

- Les liaisons rotatives sont inférieures à 10.
- La surface polaire est inférieure à  $140 \text{ \AA}^2$

Le nombre de liaisons rotatives (NRB) a été défini comme une simple liaison, non pas dans un cycle, liée à un atome non terminal (c'est-à-dire non hydrogène). Sont exclues les liaisons amide C-N en raison de leur barrière à l'énergie de rotation élevée. D'autre part, la surface polaire (PSA) qui est formée par des atomes polaires d'une molécule. C'est un descripteur qui montre une bonne corrélation avec le transport moléculaire passif à travers les membranes, et permet donc d'estimer les propriétés de transport des médicaments.

*Note* : Ces paramètres sont calculés en utilisant le logiciel molinspiration [41].

### III.5.3. Efficacité de ligand « LE »

La lipophilie est une propriété physico-chimique qui joue un rôle fondamental dans la détermination des propriétés de l'ADME (absorption, distribution, métabolisme et excrétion). La lipophilie est en corrélation avec un trop grand nombre d'autres propriétés, telles que le stockage dans les tissus, la biodisponibilité, la perméabilité, la toxicité, le volume de distribution, la liaison aux protéines plasmatiques et la liaison aux récepteurs des enzymes [42,43]. Le plus petit composé tend à avoir les meilleures propriétés physico-chimiques et un bon ADME en ce qui concerne l'efficacité du ligand [44]. Dans cette partie, nous avons étudié l'efficacité de ligand pour pénaliser les grands composés sur de petits composés ayant une puissance similaire, car les composés de plus grande taille ont des propriétés physicochimiques et ADME plus faibles [45]. L'efficacité du ligand (LE) est le plus souvent

définie comme le rapport de l'affinité d'un ligand divisé par le nombre d'atomes lourds (non hydrogène) dans la molécule.

$$LE = \frac{\Delta G}{HAC} = \frac{RT \ln K}{NH} \quad (31)$$

Ce rapport a été décrit pour la première fois en 1999 [46] et a gagné une popularité répandue dans les cercles de découverte de médicaments peu de temps après. Dans les unités plus communes, cela peut être exprimé comme suit:

$$LE = 1.4pIC50/NH \quad (32)$$

Où: NH est le nombre d'atomes lourds.

$$pIC50 = -\log(IC50)$$

#### III.5.4. Efficacité lipophile de ligand (LipE)

D'autre part, nous avons étudié l'efficacité lipophile (LipE) pour maximiser la puissance tout en maintenant la lipophilie aussi faible que possible, en raison de l'association entre une lipophilie élevée et plusieurs problèmes, y compris une faible solubilité, perméabilité membranaire, stabilité métabolique, etc [47,48]. Cependant, pour obtenir des propriétés ADMET optimales, la taille moléculaire et la lipophilie sont des facteurs importants à considérer. Si la lipophilie est trop élevée, la probabilité qu'un composé se lie à de multiples cibles augmente. Pour faciliter l'optimisation de l'affinité vis-à-vis de la lipophilie, Leeson et Springthorpe [18] ont défini l'efficacité du ligand-lipophilie (ELL), également connu sous le nom d'efficacité lipophile (Lipophilic Efficiency) (LipE):

$$LipE = pIC50 - \log P \quad (33)$$

LipE élevé favorise les composés qui gagnent beaucoup de leur affinité grâce à des interactions dirigées rendant ainsi l'interaction avec le récepteur plus spécifique.

**Références bibliographiques**

- [1] Grant G. , Richards W., "Computational Chemistry", Oxford Chemistry Primers, Oxford(1995).
- [2] de Leeuw N. H. , Mkhonto D. and Catlow C. R. A., A computer modelling study of the adhesion of apatite thin films on silicate surfaces,*J. Phys. Chem. B.*, 107, 1(2003).
- [3] Deleu M., thèse de doctorat es sciences, FUSAGx, Belgique(2000).
- [4] Lipkowitz K. B., Abuses of Molecular Mechanics: Pitfalls to Avoid, *J. Chem. Educ.*, 72,1070(1995).
- [5] Melkemi N , Etude des corrélations entre les paramètres structuraux et les propriétés bioactives de quelques composés organiques, thèse de doctorat .université mohamed khider,biskra,149(2015).
- [6] Crum-Brown A., Fraser T.R., On the connection between chemical constitution and physiological action-continued.On the physiological action of the salts of trimethylsulphin, *Trans. R. Soc. Edinburgh*, 25, 151 (1868).
- [7] Richet C., On the relationship between the toxicity and the physical properties of substances,*C. R. Séances Soc. Biol. Ses. Fil.*, 9, 775 (1893).
- [8] Meyer Zur H. Theorie der Alkoholnarkose, *Arch. Exp. Pathol. Pharmacol.*, 42, 109 (1899).
- [9] Overton E., Studien über die Narkose. Zugleich ein Beitrag zur allgemeinen Pharmakologie. Jena: Gustav Fischer, Germany(1901).
- [10] Hammett L.P., The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives, *J. Am. Chem. Soc.*, 59, 96 (1937).
- [11] Taft R.W., Polar and Steric Substituent Constants for Aliphatic and o-Benzoate Groups from Rates of Esterification and Hydrolysis of Esters, *J. Am. Chem. Soc.*, 74, 3120 (1952).
- [12] Hansch C., Maloney P.P., Fujita T., Muir R.M., Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients, *Nature*, 194, 178 (1962).
- [13] Fujita T., Iwasa J., Hansch C., A New Substituent Constant,  $\pi$ , Derived from Partition Coefficients, *J. Am. Chem. Soc.*, 86, 5175 (1964).
- [14] Hansch C., Fujita T., A New Substituent Constant,  $\pi$ , Derived from Partition Coefficients, *J. Am. Chem. Soc.*, 86, 1616 (1964).

- [15] Martin Y.C., Quantitative drug design, Marcel Dekker, New York (1978).
- [16] Free S.M., Wilson J., A Mathematical Contribution to Structure-Activity Studies, *J. Med. Chem.*, 7, 395 (1964).
- [17] Selassie C.D.. In: Abraham DJ, Ed. Burger's medicinal chemistry and drug discovery. 6th ed, John Wiley & Sons Inc., New-York(2003).
- [18] Hosoya H., The Bright Side of Mathematical Chemistry, *Bull. Chem. Soc. Jpn.*, 44, 2332 (1971).
- [19] Wiener H., Structural Determination of Paraffin Boiling Points, *J. Am. Chem. Soc.*, 69, 17 (1947).
- [20] Randić M., Characterization of molecular branching, *J. Am. Chem. Soc.*, 97, 6609 (1975).
- [21] Balaban A.T., Highly discriminating distance-based topological index , *Chem. Phys. Lett.*, 89, 399 (1982).
- [22] Simon Z., Specific interactions. Intermolecular forces, steric requirements and molecular size, *Angew. Chem. Int. Ed. Eng.*, 13, 719 (1974).
- [23] Kier L.B., Hall L.H., Molecular connectivity in chemistry and drug research, Academic Press, New-York(1976).
- [24] Selassie C.D., History of quantitative Structure-Activity relationships", in Book "History of quantitative Structure-Activity relationships", D. J. Abraham, Editor., John Wiley & Sons, Inc., California( 2003).
- [25] Fortuné A., Thèse de doctorat, Techniques de Modélisation Moléculaire Appliquées à l'Etude et à l'Optimisation de Molécules Immunogènes et de Modulateurs de la Chimiorésistance, Université Joseph Fourier – Grenoble I(2006).
- [26] Hinchliffe A., Molecular Modeling for Beginners, JohnWiley & Sons Ltd, United Kingdom (2008).
- [27] Wang J., Xie X.Q., Hou T., Xu X., *J. Phys. Chem.*, 111, 4443 (2007).
- [28] Qaneinasab S., Bayat Z., *J. Chem. Pharm. Res.*, 3, 561 (2011).
- [29] Patrick G.L., Winter H.D., Langenaeker W., Tollenaere J.P., Computational Medicinal Chemistry for Drug Discovery, Marcel Dekker, New York (1995).
- [30] Autin Ludovic M., Thèse de Doctorat, Analyse des systèmes tenase et prothrombinase par bioinformatique structurale: prédiction de complexes macromoléculaires et proposition d'agents anti-coagulants, Université Paris 5(2005).
- [31] Pépe G., Guiliani G., Loustalet S., Halfon P., *Eur. J. Med. Chem.*, 37, 865 (2002).

- [32] Carpy A., Importance de la lipophilie en modélisation moléculaire, *Analisis*, 27 (1999).
- [33] Mannhold R., Kubinyi H., Timmerman H., *Methods and Principles in Medicinal Chemistry*, VCH Publishers, New York (1996).
- [34] Khadikar P., Jaiswal M., Gupta M., Mandloi D. and Sisodia R. S., *Bioorg. Med. Chem. Lett.*, 15, 1249 (2005).
- [35] Dodge Y., *Statistique Dictionnaire encyclopédique*, Springer-Verlag France, Paris(2007).
- [36] Charif I., *Elaboration des corrélations quantitatives structure-activité des acides carbonés. Etude théorique des effets de solvants sur les équilibres libres céto-énoliques des composés b-dicarbonyles cycliques*, Thèse de doctorat, Université de Tlemcen(2012).
- [37] Segall D. M., *J. Curr. Pharm. Des.* 18, 1292 (2012).
- [38] Veber D. F., Johnson S. R., Cheng H. Y., Smith B. R., Ward K. W. and Kopple K. D, *J. Med. Chem.* 45, 2615 (2002).
- [39] Vistoli G., Pedretti A. and Testa B., *Drug Discov Today*, (7-8), 285-294 (2008).
- [40] Remko M., Swart M. and Bickelhaupt F. M., *J. Bioorg. Med. Chem.* 14, 1715 (2006).
- [41] Database,[<http://www.molinspiration.com>].
- [42] Kerns E, Elsevier Science,1–552(2008).
- [43] Pliska V, Testa B, van de Waterbeemd H, Mannhold R, Kubinyi H, and Timmerman H, *Lipophilicity in Drug Action and Toxicology*, Wiley-VCH, Federal Republic of Germany (1996).
- [44] Leeson, P.D. and Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.* 6, 881–890 (2007).
- [45] Hopkins A L., Groom C R., and Alexander A., *J. Drug. Discov. T*, 9,430(2004).
- [46] Kuntz ID, Chen K, Sharp KA, Kollman PA.The maximal affinity of ligands. *Proc. Natl Acad. Sci. USA* 96(18), 9997–10002 (1999).
- [47] Hughes J. D., Blagg J., Price D. A., Bailey S., Decrescenzo G. A., Devraj R. V., Ellsworth E., Fobian Y. M., Gibbs M. E. and Gilles R. W., *J. Bioorg. Med. Chem. Lett.* 18, 4872 (2008).
- [48] Edwards R. W. and Price D., *Annu. Rep. Med. Chem.* 45, 381 (2010).

# *Résultats et discussions*

*Chapitre IV:*  
*Etude des propriétés*  
*structurales,*  
*électroniques et*  
*énergétiques de la*  
*pyrazine et ses dérivés*

## IV.1. Introduction

La modélisation moléculaire s'avère indispensable pour l'interprétation et la compréhension des observations expérimentales. Dans le domaine moléculaire, toutes les propriétés sont reliées à la nature et la forme de la molécule. Pouvoir optimiser la géométrie d'une molécule par un modèle théorique (méthodes de chimie quantique) c'est approcher sa conformation moléculaire observée expérimentalement [1,2].

L'objectif de cette partie de notre travail a été orienté vers la recherche d'une conformation moléculaire de la pyrazine calculée théoriquement et similaire à celle donnée par l'expérience. Aussi, l'autre objectif est d'apporter une contribution à la compréhension des propriétés structurales, énergétiques et électroniques pour présenter la conformation moléculaire la plus stable et également l'arrangement et l'environnement atomique ainsi que la réactivité chimique de la molécule étudiée.

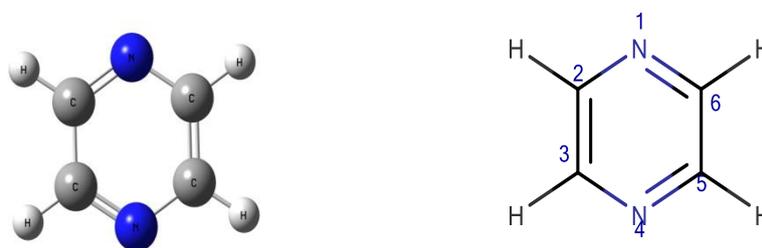
Dans cette étude, nous définirons la meilleure méthode ainsi que la base nécessaire appliquée à cette dernière.

## IV.2. Méthodes de calcul utilisées

Les molécules ont été optimisées initialement par logiciel HyperChem 8.0.6 [3]. D'abord, les structures de la pyrazine et ses dérivés ont été optimisées par la mécanique moléculaire, avec un champ de force MM+ à l'aide de l'algorithme steepest descent avec pour critère une racine du carré moyen du gradient égale à 0,01 kcal.mol<sup>-1</sup>[4]. Les structures obtenues ont été ré-optimisées en utilisant la méthode AM1 [5]. Dans l'étape suivante, une étude parallèle a été faite en utilisant le logiciel Gaussien 09 [6], nous avons effectué des calculs d'optimisation de géométrie de la pyrazine. Nous nous sommes basés sur la méthode ab initio de type Hartree-Fock (HF)[7] et de type Moller Plesset (MP2) et la théorie de la fonctionnelle de la densité (en anglais DFT: Density Functional Theory) avec la fonctionnelle B3LYP [8] en utilisant les bases 6-311++G (d,p) et cc-PVDZ. Cette investigation conduira à définir la meilleure méthode ainsi que la base nécessaire pour ces systèmes.

### IV.3. Etude des propriétés structurales et électroniques de la pyrazine

Les paramètres géométriques de la pyrazine ont été optimisés par les méthodes *ab initio*/HF, *ab initio*/MP2 et la méthode DFT, les résultats de calcul sont donnés dans le tableau IV.1. La structure et la numérotation des atomes correspondant à la molécule sont présentées dans la figure IV.1. L'efficacité des méthodes théoriques utilisées dans les calculs peuvent être évalués par la comparaison des résultats obtenus avec les résultats expérimentaux [9]. Les calculs visent les longueurs des liaisons, les angles de valence pour la pyrazine (Tableau IV.1) et les charges de Mulliken et NBO respectivement (Tableau IV.2) et (Tableau IV.3).



**Figure IV.1** : Conformation 3D de la pyrazine (GaussView 5.08) et 2D (MarvinSketch 15.8.31).

**Tableau IV.1** : Résultats de calcul des longueurs des liaisons, les angles de valence de la pyrazine.

		EXP[9]	DFT/B3LYP		<i>Ab initio</i> /HF		<i>Ab initio</i> /MP2	
			6-311++ G(d,p)	cc- PVDZ	6-311++ G(d,p)	cc-PVDZ	6-311++ G(d,p)	cc-PVDZ
Longueur de liaison (Angstrom)	C-N	1.338	1.335	1.339	1.317	1.320	1.343	1.349
	C-C	1.397	1.394	1.398	1.386	1.388	1.399	1.405
	C-H	1.083	1.086	1.095	1.075	1.082	1.087	1.096
Angle de valence(degré)	CCH	120.0	120.0	120.8	120.8	120.8	120.7	120.6
	CNC	115.7	116.1	115.6	116.6	116.3	115.2	114.6

D'après les résultats obtenus, nous pouvons noter une bonne corrélation entre les valeurs calculées et expérimentales pour les paramètres géométriques. L'efficacité de la méthode DFT / B3LYP peut être affirmée par la comparaison avec les résultats obtenus par des calculs plus élaborés tels que *Ab initio* HF et MP2.

A cet égard, nous trouvons que tous les résultats obtenus par les calculs sont proches aux résultats expérimentaux. Les résultats obtenus par la méthode DFT avec la base cc-PVDZ sont les plus proches donc nous pouvons déduire que la méthode DFT/cc-PVDZ est la meilleur méthode pour approfondir notre étude sur les propriétés structuraux de la molécule de la pyrazine et ses dérivés.

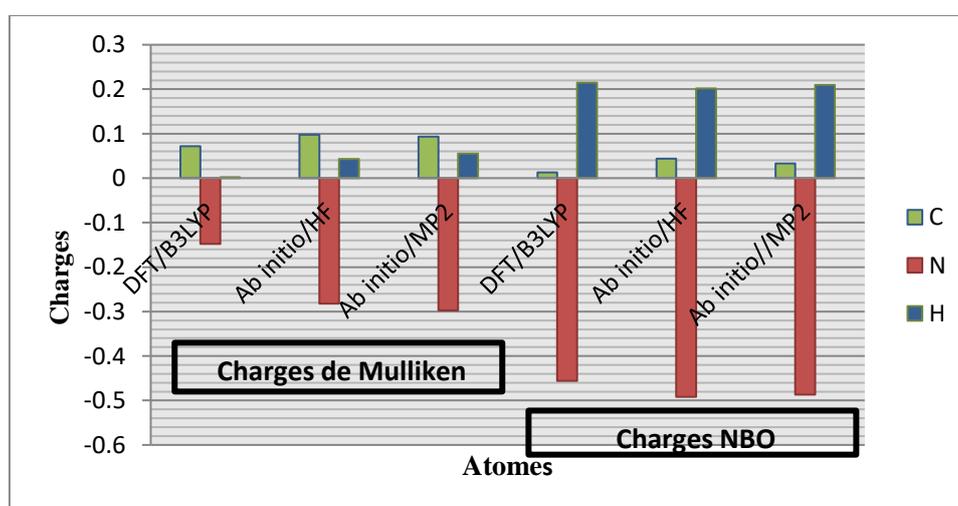
La géométrie de la pyrazine est plane dont les angles dièdres du cycle dans cette molécule sont autour de 0 degré.

**Tableau IV.2 :** Résultats de calcul des charges Mulliken de la pyrazine.

Atomes	DFT/B3LYP cc-PVDZ	<i>Ab initio</i> /HF cc-PVDZ	<i>Ab initio</i> /MP2 cc-PVDZ
C	0,072	0,098	0,093
N	-0,148	-0,282	-0,297
H	0,002	0,043	0,055

**Tableau IV.3 :** Résultats de calcul des charges NBO de la pyrazine.

Atomes	DFT/B3LYP cc-PVDZ	<i>Ab initio</i> /HF cc-PVDZ	<i>Ab initio</i> /MP2 cc-PVDZ
C	0.013	0.044	0.033
N	-0.456	-0.492	-0.487
H	0.215	0.202	0.210



**Figure IV.2 :** Représentation graphique des charges NBO et Mulliken des atomes de la pyrazine.

A partir de la figure IV.2, nous constatons que les valeurs des charges NBO sont plus proches et plus efficaces pour les différentes méthodes de calcul par rapport aux charges de Mulliken. Donc les valeurs des charges NBO calculées par les méthodes ab initio (HF et MP2) sont similaires à celles calculées par la méthode DFT.

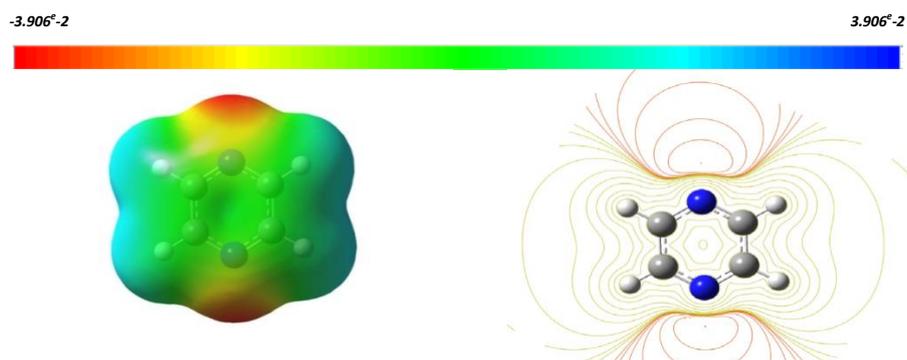
Le potentiel électrostatique moléculaire MESP donne des informations sur les régions moléculaires qui sont préférées où évitées par un électrophile ou un nucléophile. Tout système chimique crée un potentiel électrostatique autour de lui-même [10]. MESP est un facteur important par lequel nous pouvons confirmer la répartition de la région du potentiel électrostatique de la taille et la forme des molécules ainsi que la physiologie totale des molécules [11]. Une partie de la molécule qui présente un potentiel électrostatique négatif est susceptible d'une attaque électrophile. Les régions rouge et bleue de la carte MESP se réfèrent aux régions des potentiels négatifs et positifs et correspondent respectivement aux régions riches en électrons et déficientes en électrons tandis que la couleur verte indique le potentiel électrostatique neutre [12].

Les augmentations potentielles dans l'ordre rouge <orange <jaune <vert <bleu.

La surface de MESP de la pyrazine (Figure IV.3) montre trois régions :

- La couleur rouge (potentiel électrostatique négative) autour des deux atomes d'azote qui expliquent la capacité d'une attaque électrophile sur ces positions.
- La couleur bleue (potentiel électrostatique positif) autour des quatre atomes d'hydrogène qui expliquent que ces régions sont sensibles à une attaque nucléophile.
- La couleur verte située au milieu entre les régions rouges et bleues explique la surface du potentiel électrostatique neutre.

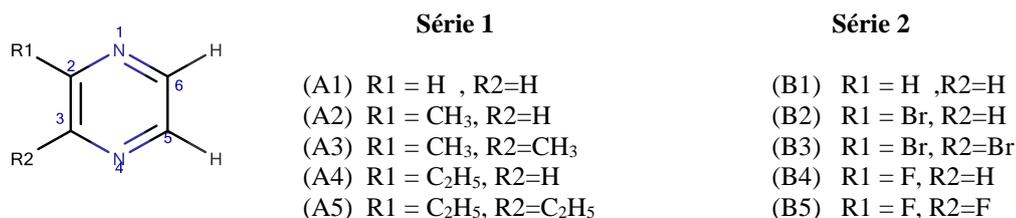
Le code de la couleur de ces cartes est dans l'intervalle compris entre  $-0,0391$  u.a à  $0,0391$  u.a



**Figure IV.3 :** Représentation des surfaces potentielles électrostatiques moléculaires de la pyrazine en 3D et 2D.

#### IV.4. Etude de l'effet de la substitution sur le noyau de base de pyrazine

Pour avoir l'effet de la substitution sur les paramètres électroniques et énergétiques et leur incidence sur la stabilité et la réactivité chimique de la molécule étudiée, deux séries de molécules substituées ont été étudiées (Figure IV.4). La première série comporte des molécules substituées par des substituants électrodonneurs (méthyle et éthyle) et la deuxième série comporte des molécules substituées par des substituants électroattracteurs (fluore et brome).



**Figure IV.4:** Les deux séries des pyrazines substitués.

Le tableau IV.4 présente les valeurs calculées de différents paramètres énergétiques tels que la chaleur de formation  $\Delta H_f$ , le moment dipolaire  $\mu$ , HOMO (orbitale moléculaire plus haute occupée), LUMO (orbitale moléculaire plus basse occupée) et la différence  $\Delta E$  entre les énergies des orbitales HOMO et LUMO. Le tableau IV.4 présente également les valeurs calculées du moment dipolaire  $\mu$ .

Les paramètres HOMO, LUMO,  $\Delta E$  et  $\mu$  ont été calculés par la méthode DFT (B3LYP / cc-PVDZ) en utilisant logiciel Gaussian 09, par contre la chaleur de formation a été calculée par la méthode semi-empirique PM3 avec l'utilisation du logiciel HyperChem8.0.6.

**Tableau IV.4:** Résultats du calcul des paramètres énergétiques et le moment dipolaire des dérivés de la pyrazine.

		$\Delta H_f$ [kcal/mol]	HOMO [au]	LUMO [au]	$\Delta E$ [au]	$\mu$ [Debye]
A1/B1	Pyrazine	44.09	-0.252	-0.055	0.197	0.00
A2	2-methyl pyrazine	37.05	-0.247	-0.051	0.196	0.59
A3	2,3-di-methyl pyrazine	31.78	-0.243	-0.044	0.199	0.80
A4	2-ethyl pyrazine	30.97	-0.247	-0.051	0.195	0.59
A5	2,3-di-ethyl pyrazine	20.48	-0.242	-0.045	0.196	0.69
B2	2-bromopyrazine	49.73	-0.269	-0.068	0.201	1.50
B3	2,3-di-bromopyrazine	55.88	-0.268	-0.075	0.192	2.05
B4	2-fluoro pyrazine	04.15	-0.272	-0.065	0.207	1.33
B5	2,3-di-fluoropyrazine	-33.52	-0.280	-0.069	0.211	2.24

#### IV.4.1. Effet d'un substituant électrodonneur (méthyle et éthyle)

Pour la substitution par un électrodonneur, la chaleur de formation est diminuée approximativement de 6 kcal/mol pour chaque addition d'un groupement méthyle et 12 kcal/mol pour chaque addition d'un groupement éthyle.

La charge atomique négative sur les atomes d'azote est augmentée considérablement pour les dérivés méthyléniques et éthyléniques, sauf l'atome N1 dans les dérivés A3 et A5 (Tableau IV.5).

Dans les dérivés alkyles monosubstitué, le 2-éthyl-pyrazine (le composé A4) présente une charge positive maximale sur le carbone C2 (0.206), cet atome est privilégié pour subir une attaque nucléophile (Tableau IV.5).

Ce résultat est conforté par le gap énergétique HOMO-LUMO le plus bas (0,195) (Tableau IV.4), ce qui montre la réactivité chimique du composé A4.

**Tableau IV.5:** Valeurs des charges de NBO de la pyrazine et ses dérivés (série 1).

	A1	A2	A3	A4	A5
N1	-0.456	-0.472	-0.471	-0.476	-0.476
N4	-0.456	-0.452	-0.473	-0.452	-0.472
C2	0.013	0.204	0.215	0.206	0.216
C3	0.013	0.016	0.208	0.020	0.213
C5	0.013	0.003	0.010	0.004	0.013
C6	0.013	0.022	0.012	0.023	0.015
C-methyl- 2	-	-0.665	-0.669	-	-
C-methyl -3	-	-	-0.673	-	-
C <sup>1</sup> -ethyl- 2	-	-	-	-0.458	-0.459
C <sup>2</sup> -ethyl -2	-	-	-	-0.628	-0.627
C <sup>1</sup> -ethyl -3	-	-	-	-	-0.461
C <sup>2</sup> -ethyl-3	-	-	-	-	-0.627

HOMO représente la capacité d'une molécule à donner un électron et LUMO représente l'aptitude à accepter un électron. La différence entre les énergies HOMO et LUMO (HOMO-LUMO), appelée la dureté chimique, représente la résistance d'une molécule au transfert d'électrons [13,14].

Quand l'écart énergétique HOMO-LUMO est élevé, l'écoulement des électrons à l'état d'énergie plus élevée est difficile, ce qui rend la molécule dure et moins réactive. D'autre part, le gap énergétique HOMO-LUMO plus faible, a pour conséquence un écoulement facile des électrons, ce qui rend la molécule molle, ceci est gouverné par le principe de HSAB (Hard Soft Acid and Base) [15].

Dans le cas des pyrazines disubstitués par deux méthyles ou deux éthyles, le composé A5 présente une charge maximale (0.216) sur le carbone C2 et un petit écart énergétique HOMO-LUMO (0,196) (Tableaux IV.3 et IV.4). Le carbone C2 est un site préférentiel pour l'attaque nucléophile.

Nous notons également que les substituants alkyles (méthyle et éthyle) ont un effet sur l'énergie des orbitales frontières. La présence de ces groupes dans la structure des pyrazines provoque l'augmentation des énergies des orbitales HOMO et LUMO (Tableau IV.4).

#### IV.4.2. Effet d'un substituant électroattracteurs (bromure et fluore)

Dans ce travail, nous avons également étudié l'effet de la substitution des substituants électroattracteurs (bromure et fluore) sur la stabilité et la réactivité chimique des pyrazines.

La chaleur de formation est diminuée approximativement de 39 kcal/mol pour chaque addition d'un groupement fluore par contre le groupement bromure est augmenté de 6 kcal/mol.

**Tableau IV.6:** Valeurs des charges de NBO de la pyrazine et ses dérivés (série 2).

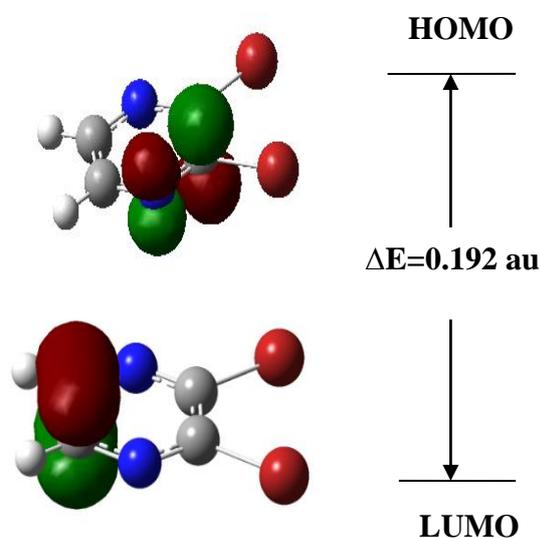
	<b>B1</b>	<b>B2</b>	<b>B3</b>	<b>B4</b>	<b>B5</b>
<b>N1</b>	-0.456	- 0.458	-0.446	-0.497	-0.485
<b>N4</b>	-0.456	-0.441	-0.446	-0.441	-0.485
<b>C2</b>	0.013	0.112	0.102	0.634	0.586
<b>C3</b>	0.013	0.018	0.102	-0.040	0.586
<b>C5</b>	0.013	0.006	0.018	-0.008	0.002
<b>C6</b>	0.013	0.024	0.018	0.024	0.002
<b>Brome-2</b>	-	0.064	0.100	-	-
<b>Brome-3</b>	-	-	0.100	-	-
<b>Fluor- 2</b>	-	-	-	-0.338	-0.327
<b>Fluor -3</b>	-	-	-	-	-0.327

Dans les dérivés monosubstitués du bromure et du fluore, le 2-bromo-pyrazine (composé B2) est prévu pour être plus réactif chimiquement, ce composé présente un faible gap énergétique HOMO-LUMO (0,201) (Tableau IV.4).

Le carbone C2 dans le 2-bromo-pyrazine présente une charge positive maximale (0.112), ce qui est privilégié pour subir une attaque nucléophile (Tableau IV.6). Dans les dérivés de bromure et de fluore disubstitués, le 2,3-dibromo-pyrazine (composé B3) est plus réactif que le difluoro-pyrazine (composé B5), cela est expliqué par le plus faible gap énergétique HOMO-LUMO (0.192) (Tableau IV.4).

La présence des groupes électroattracteurs dans la structure des pyrazines provoque la diminution des énergies des orbitales HOMO et LUMO (Tableau IV.4).

Les orbitales frontières HOMO et LUMO de l'état fondamental du composé B3 (2,3-dibromo-pyrazine) sont représentés dans la figure IV.5. Nous pouvons constater que l'orbitale moléculaire HOMO se concentre principalement sur les atomes C5 et C6 avec une certaine délocalisation le long de liaison C5-N1, tandis que l'orbitale moléculaire LUMO se concentre principalement sur les atomes N1 et C2. Cette distribution des orbitales frontières a montré l'existence de la délocalisation des électrons  $\pi$  dans le système conjugué de la molécule.



**Figure IV. 5:** Orbitales moléculaires HOMO et LUMO du composé B3 (2,3-dibromopyrazine).

Le moment dipolaire égale à 0 ce qui confirme la symétrie du groupe  $D_{2h}$  de la pyrazine par contre le 2,3 difluoro-pyrazine (composé B5) présente une valeur maximale du moment dipolaire, il provient de la présence d'un effet de résonance.

## IV.5. Conclusion

Les résultats obtenus de calcul par les méthodes *ab initio* et la théorie de la densité fonctionnelle nous permet de déterminer plusieurs propriétés moléculaires de la pyrazine. La comparaison structurale et électronique de noyau de base de la pyrazine présente des résultats similaires entre les résultats expérimentaux et les différentes méthodes de calcul.

Les méthodes *ab initio* /HF, MP2 et DFT/B3LYP peuvent être utilisés pour prédire la réactivité chimique des molécules étudiées et pour déterminer l'effet de la substitution sur les propriétés énergétiques et électroniques des molécules étudiées. L'étude de la substitution sur le noyau de base de la pyrazine montre une influence de la nature des substituants, donneurs (méthyl, éthyl) et accepteurs (fluore, bromure) sur les propriétés énergétiques et électroniques du noyau de base. Le composé B3 est prévu pour être le plus réactif chimiquement avec un gap énergétique HOMO-LUMO le plus faible (0.192). Parmi tous les systèmes des pyrazines étudiés, nous notons aussi que la présence d'un substituant attracteur : bromure et fluore

baisse les énergies des orbitales frontières HOMO et LUMO, la diminution est plus importante pour l'énergie de l'orbitale LUMO.

## Références bibliographiques

- [1] Maseras F., Morokuma K., *Journal of Computational Chemistry*. 16, 1170(1995).
- [2] Humbel S. , Sieber S. , Morokuma K. , A new integrated *ab initio* + molecular mechanics geometry optimization scheme of equilibrium structures and transition states, *The Journal of chemical physics*.105, 1959(1996).
- [3] HyperChem (Molecular Modeling System) Hypercube, Inc., 1115 NW, 4th Street, Gainesville, FL 32601, USA (2008).
- [4] Smith J., Karplus M., *Journal of the American Chemical Society*.114, 801 (1992).
- [5] Stewart J.J.P., Optimization of parameters for semiempirical methods I.,*J. Comp. Chem.*, 10, 221 (1989).
- [6] Gaussian 09, Frisch M. J., Trucks G. W., Schlegel H. B., Scuseria G. E., Robb M. A., Cheeseman J. R., Scalmani G., Barone V., Mennucci B., Petersson G. A., Nakatsuji H., Caricato M., Li X., Hratchian H. P., Izmaylov A. F., Bloino J., Zheng G., Sonnenberg J. L., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Vreven T., Montgomer J. A, Peralta J. E., Ogliaro F., Bearpark M., Heyd J. J., Brothers E., Kudin K. N., Staroverov V. N., Keith T., Kobayashi R., Normand J., Raghavachari K., Rendell A., Burant J. C., Iyengar S. S., Tomasi J., Cossi M., Rega N., Millam J. M., Klene M., Knox J. E., Cross J. B., Bakken V., Adamo C., Jaramillo J., Gomperts R., Stratmann R. E., Yazyev O., Austin A., Cammi R., Pomelli C., Ochterski J. W., Martin R. L., Morokuma K., Zakrzewski V. G., Voth G. A., Salvador P., Dannenberg J. J., Dapprich S., Daniels A.D., Farkas O., Foresman J. B., Ortiz J. V., Cioslowski J., and Fox D. J., Gaussian Inc., Wallingford, CT (2010).
- [7] Boucekkine G., Méthodes de la chimie quantique, Techn. Ing., AF6050(2007).
- [8] Jensen F. and Wiley J., Introduction to Computational Chemistry, Chichester, Angleterre, (1999).
- [9] Kanno M., Ito Y., Shimakura N., Koseki S., Kono H., Fujimura Y. , *Ab initio* quantum dynamical analysis of ultrafast nonradiative transitions *via* conical intersections in pyrazine, *J.of Phys. Chem. - Chemical Physics* ,17 , 2012 (2015).
- [10] Murray J. S. and Sen K., *Molecular Electrostatic Potentials*, Concepts and Applications. Elsevier, Amsterdam (1996).
- [11] Abhisek B., Apoorva D., Subodh P., Anoopande and V. Narayan, DFT Study of Picric Acid and its Derivative by First Principles, *I.J.E.T.T* . 6, 3(2013).

[12] Srivastava R. K., Narayan V., Kumar A., Prasad O. and Sinha L., Theoretical Studies on the Isomers of Quinazolinone by first Principles, *Res. J. Recent Sci.* 1(3), 11-18 (2012).

[13] Parr R.G., Von Szentpály and Liu S., Electrophilicity Index, *J. Am. Chem. Soc.*, 121, 1922-1924(1999)

[14] Parr R.G. and Pearson R.G., Absolute hardness: companion parameter to absolute electronegativity, *Am. Chem. Soc.*, 105, 7512-7516(1983).

[15] Miessler G.L., Tarr, D.A., Inorganic chemistry, 2<sup>nd</sup> edition, Prentice-hall Upper Saddle River, NJ, USA(1999).

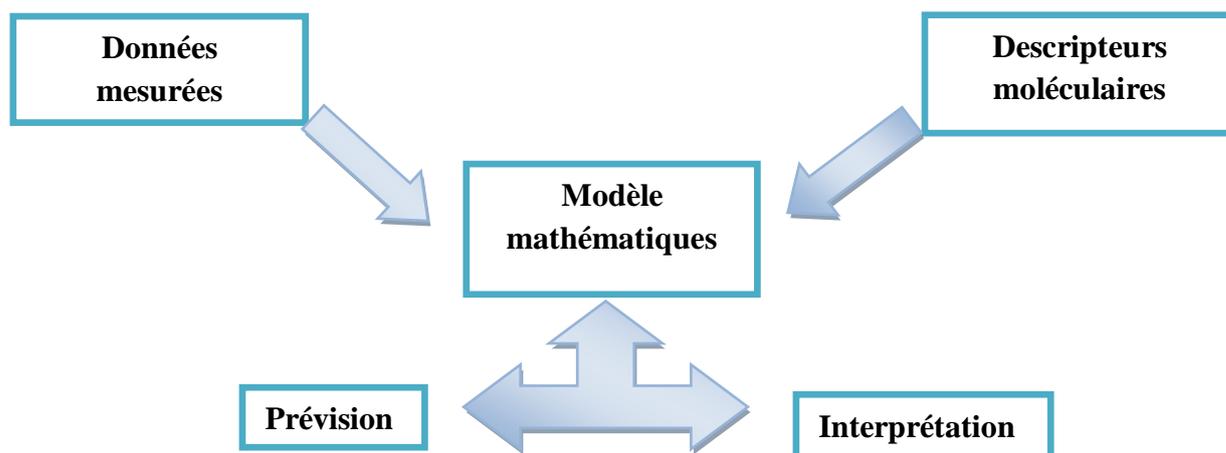
*Chapitre V:*  
*Etude*  
*Qualitative et*  
*quantitative des*  
*relations structure-*  
*activité des dérivés de*  
*pyrazine*

## V.1. Introduction

Plusieurs techniques *in silico* sont utilisés dans le processus dans la découverte et la conception des médicaments. La principale technique est les méthodes QSAR, qui étudient les relations qualitatives et quantitatives des structure-activité (SAR/QSAR).

Le principe des méthodes QSAR (Quantitative Structure Activity Relationship) consiste comme leurs noms l'indiquent à mettre en place une relation mathématique à l'aide de méthodes d'analyse de données, reliant des propriétés moléculaires microscopiques appelées descripteurs, à un effet expérimental (activité biologique, toxicité), pour une série de composés chimiques similaires. Le point de départ de telles méthodes se construit sur la définition des descripteurs moléculaires empiriques ou théoriques. Ces dernières prennent en compte des informations sur la structure et les caractéristiques physico-chimiques des molécules. Le choix de la base des données expérimentales de référence est décisif dans une étude QSAR. Elle doit être composée de données expérimentales fiables obtenues en suivant un protocole expérimental unique. En effet, la robustesse du modèle dépend fortement de la base sur lequel il se fonde. Enfin, le lien entre les descripteurs et la base de données sont déterminés grâce à des outils d'analyse comme les régressions multi-linéaires (MLR), les régressions aux moindres carrés partiels (PLS), les arbres de décisions, les réseaux de neurones, et les algorithmes génétiques.

Les informations extraites à partir des résultats des études QSAR peuvent être utilisées pour obtenir une meilleure connaissance des structures moléculaires et probablement le mode d'action au niveau moléculaire. Ces informations peuvent alors être utilisées pour prévoir les propriétés physicochimiques et les activités biologiques de nouveaux composés ainsi que pour concevoir de nouvelles structures (Figure V.1) [1].



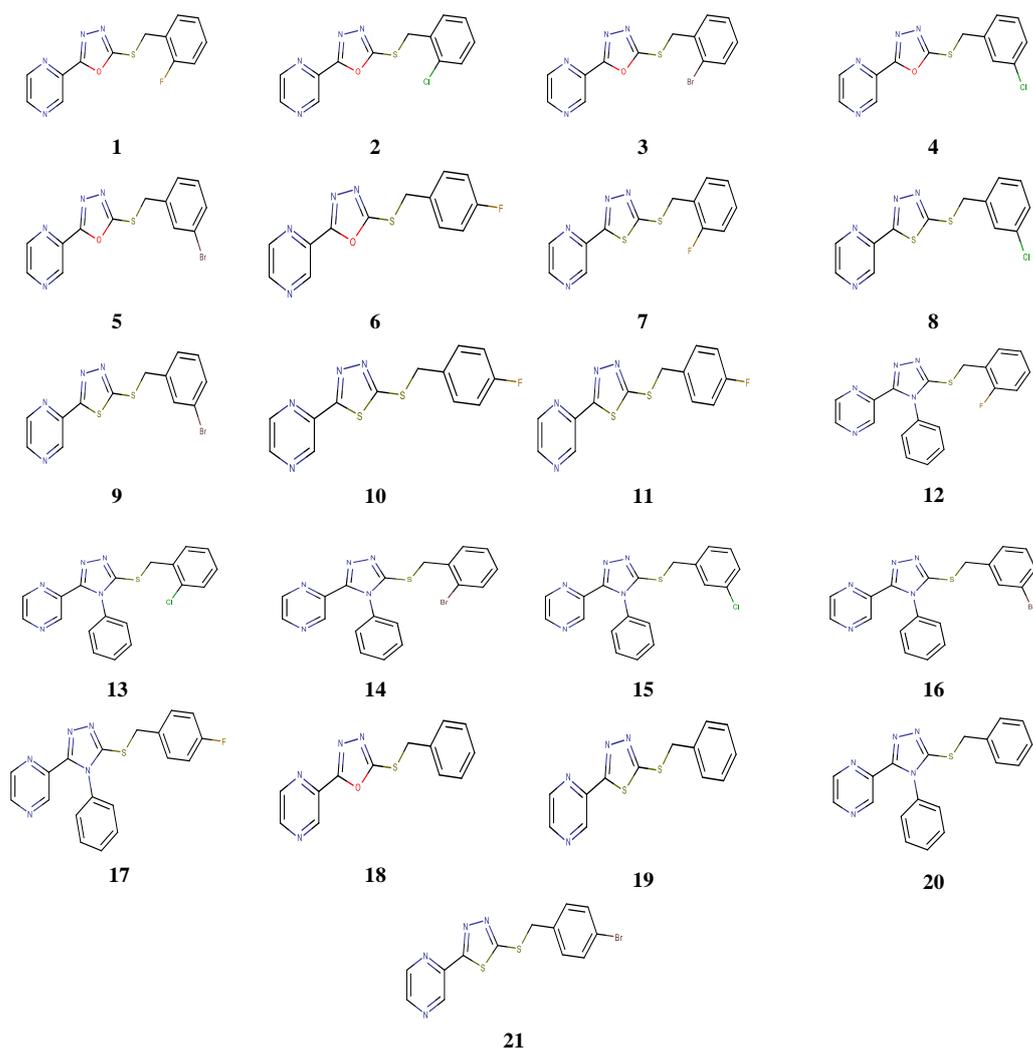
**Figure V.1:** Modèle d'étude quantitative de relations structure activité (QSAR).

Notre présente étude vise à développer des modèles QSAR pour un groupe des vingt un dérivés de la pyrazine qui ont une activité inhibitrice de télomérase contre la prolifération des lignées cellulaires HEPG2(human colorectal carcinoma cell) et la lignées cellulaires SW1116 (human colorectal carcinoma cell).

## V.2. Etude qualitative des relations structure/activité (SAR) dans les dérivés de la pyrazine

### V.2.1. Sélection d'une série bioactives des pyrazines

Plusieurs dérivés de la pyrazine présentent des propriétés pharmacologiques importantes. Ils sont utilisés dans le traitement et la prévention d'un grand nombre de maladies. Dans le but est d'étudier l'effet de la substitution sur les propriétés physicochimiques des pyrazines, nous avons sélectionné une série bioactive de dérivés de la pyrazine [2]. Les structures correspond de ces molécules sont représentées dans la figure V.2.



**Figure V.2 :** Structures chimiques des dérivés de la pyrazine.

Tout d'abord, les vingt un dérivés sélectionnés de la pyrazine ont été optimisés par la méthode de mécanique moléculaire (champ de force MM+) en utilisant un logiciel HyperChem 8.0.6.

Après cela, le module propriétés QSAR de l'HyperChem 8.0.6 a été utilisé pour calculer les paramètres suivants: la polarisabilité (Pol), la réfractivité molaire (MR), le coefficient de partage octanol/eau (log P), l'énergie d'hydratation (HE), le volume molaire (MV), la surface moléculaire (SAG) et le poids moléculaire (MW).

### V.2.1.1. Résultats et discussions

Les résultats du calcul des paramètres physicochimiques de la série sélectionnée sont représentés dans le tableau V.1.

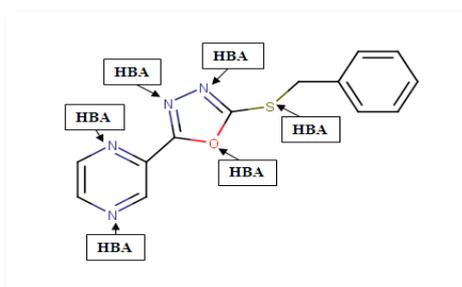
Tableau V.1: Propriétés physicochimiques des dérivés de la pyrazine.

N°	MW [amu]	SAG [A <sup>o2</sup> ]	MV [A <sup>o3</sup> ]	Pol [A <sup>o3</sup> ]	MR [A <sup>o3</sup> ]	LogP	HE [kcal·mol <sup>-1</sup> ]
1	288.30	466.47	770.17	28.82	79.14	1.94	-12.54
2	304.75	474.61	791.62	30.84	83.73	2.32	-12.63
3	349.20	485.20	810.26	31.54	86.54	2.60	-12.58
4	304.75	498.29	809.75	30.84	83.73	2.32	-13.29
5	349.20	505.96	828.55	31.54	86.54	2.60	-13.24
6	288.30	479.96	777.83	28.82	79.14	1.94	-13.38
7	304.36	486.18	800.84	31.18	85.58	2.29	-11.39
8	320.81	498.70	822.05	33.20	90.17	2.67	-12.25
9	320.81	511.10	830.07	33.20	90.17	2.67	-12.54
10	365.27	520.26	848.39	33.90	92.99	2.94	-12.50
11	304.36	496.49	797.40	31.18	85.58	2.29	-12.61
12	363.41	543.24	961.11	39.20	110.97	2.48	-10.86
13	379.87	550.54	984.28	41.21	115.56	2.86	-10.69
14	424.32	554.53	997.06	41.91	118.37	3.13	-10.63
15	379.87	562.49	980.74	41.21	115.56	2.86	-11.45
16	424.32	570.87	999.40	41.91	118.37	3.13	-11.41
17	363.41	543.20	948.38	39.20	110.97	2.48	-11.54
18	270.31	475.71	769.23	28.91	079.01	2.55	-13.67
19	286.37	490.32	789.01	31.27	085.45	2.89	-12.89
20	345.42	539.34	939.49	39.29	110.84	3.08	-11.79
21	365.27	526.34	854.34	33.90	092.99	2.94	-11.49

La polarisabilité et la réfractivité molaire augmentent relativement avec la taille et le poids moléculaire des pyrazines étudiés (Tableau V.1). Ce résultat est en accord avec la formule de Lorentz-Lorenz qui donne une relation entre la polarisabilité, la réfractivité molaire et le volume [3]. Cette relation montre que la réfraction molaire et la polarisabilité sont proportionnelles avec le volume et la masse moléculaire. Par exemple, le composé 14 et le composé 16 sont des pyrazines substitués par des radicaux volumineux, ces deux composés possèdent les mêmes grandes valeurs de polarisabilité (41.91 A<sup>o3</sup>) et de réfractivité molaire (118.37 A<sup>o3</sup>) et une légère différence par rapport aux valeurs des surfaces et des volumes (Tableau V.1).

Le 2-(5-((2-fluorophenyl)methyl)sulfanyl)-1,3,4-oxadiazol-2-yl)pyrazine (composé 1) est le dérivé dans la série étudiée des pyrazines, il possède une petite valeur de la polarisabilité (28.82 A<sup>o3</sup>) et de réfractivité molaire (79.14 A<sup>o3</sup>). La présence des groupes hydrophobes dans la structure des pyrazines provoquent la diminution de l'énergie d'hydratation; ainsi que, la présence des groupes hydrophiles augmentent

l'énergie d'hydratation (Tableau V.1). L'énergie d'hydratation en valeur absolue, la plus importante (13.67 kcal/mol) a été observée pour le composé 2-[5-(benzylsulfanyl)-1,3,4-thiadiazol-2-yl]pyrazine (composé 18). La plus faible valeur d'énergie d'hydratation (10.63 kcal/mole) a été observée pour le 2-(5-[(2-bromophenyl) methyl]sulfanyl)-4-phenyl-4H-1,2,4-triazol-3-yl)pyrazine (composé 14). En effet, dans les milieux biologiques, les molécules polaires sont entourées par les molécules d'eau où les liaisons hydrogènes peuvent être établies entre les molécules d'eau et ces molécules. Les sites donneurs de protons interagissent avec l'atome d'oxygène de l'eau et les sites accepteurs de protons interagissent avec l'atome d'hydrogène. Ces molécules hydratées sont partiellement déshydratées avant et lors de leurs interactions. Ces interactions de faible énergie et généralement réversibles, que l'on observe généralement avec les différents récepteurs biologiques [4]. Le composé 18 possède six sites accepteurs de proton (O en groupement hydroxyle et S du groupe S=O) et ne possède aucun site donneur (Figure V.3). Pour une bonne biodisponibilité orale, le log P doit être supérieur à zéro et inférieur à 3 ( $0 < \log P < 3$ ). En effet, pour un log P trop élevé, le médicament présente une faible solubilité aqueuse. Dans le cas d'un log P très faible, le médicament présente une difficulté de se pénétrer dans les bicouches lipidiques des membranes cellulaires [5].



**Figure V.3:** Liaisons hydrogènes accepteurs (HBA) du composé 18.

La plupart des composés possèdent un log P dans la plage des normes (intervalle 0-3). Ces composés possèdent généralement une bonne absorption intestinale, en raison d'un bon équilibre entre la solubilité et la perméabilité par diffusion passive; l'élimination métabolique est faible, en raison de faibles liaisons avec les enzymes métaboliques. En outre, ces molécules ont tendance à être libérées par le rein, en raison de leur polarité.

### V.3. Paramètres MPO

Nous avons utilisé les règles de Lipinski pour identifier les composés qui posent des problèmes d'absorption et de perméabilité si ces composés ne valident pas au moins deux de ces règles. La règle de Lipinski est la plus utilisée pour caractériser les composés « drug-like ». Nous rappelons que cette règle a pour but d'identifier les composés posant des problèmes d'absorption et de perméabilité. Les composés ne validant pas au moins deux de leurs critères ont de très fortes chances d'avoir des problèmes d'absorption ou de perméabilité [6].

**Tableau V.2:** Activités pharmacologiques et propriétés des méthodes MPO pour « drug-likeness » des dérivés de la pyrazine.

Compounds	pIC50	logP	MW [uma]	HBA	HBD	Règle de Lipinski's	NRB	PSA	LE	LIPE
1	4.90	1.94	288.30	5	0	0	4	64.71	0.34	2.96
2	4.91	2.32	304.75	5	0	0	4	64.71	0.34	2.59
3	5.37	2.60	349.20	5	0	0	4	64.71	0.38	2.77
4	5.12	2.32	304.75	5	0	0	4	64.71	0.36	2.80
5	5.24	2.60	349.20	5	0	0	4	64.71	0.37	2.64
6	4.78	1.94	288.30	5	0	0	4	64.71	0.33	2.84
7	5.02	2.29	304.36	4	0	0	4	51.57	0.35	2.73
8	4.97	2.67	320.81	4	0	0	4	51.57	0.35	2.30
9	5.15	2.67	320.81	4	0	0	4	51.57	0.36	2.48
10	5.75	2.94	365.27	4	0	0	4	51.57	0.40	2.81
11	4.99	2.29	304.36	4	0	0	4	51.57	0.35	2.70
12	4.72	2.48	363.41	5	0	0	5	56.50	0.25	2.24
13	4.73	2.86	379.87	5	0	0	5	56.50	0.25	1.87
14	4.93	3.13	424.32	5	0	0	5	56.50	0.27	1.80
15	4.71	2.86	379.87	5	0	0	5	56.50	0.25	1.85
16	4.97	3.13	424.32	5	0	0	5	56.50	0.27	1.84
17	5.00	2.48	363.41	5	0	0	5	56.50	0.27	2.52
18	4.95	2.55	270.31	5	0	0	4	64.71	0.36	2.40
19	4.95	2.89	286.37	4	0	0	4	51.57	0.36	2.06
20	4.77	3.08	345.42	5	0	0	5	56.50	0.27	1.69
21	6.11	2.94	365.27	4	0	0	4	51.57	0.43	3.17

#### V.3.1. Résultats et discussions

A travers le tableau V.2, tous les composés ont des valeurs moins de 5 pour la lipophilie, allant de 1.94 à 3.13. A cet égard, ces composés sont mieux solubilisés dans des solutions aqueuses et lipidiques. Nous pouvons également voir dans le

tableau V.2 que tous ces composés ont le nombre accepteurs d'hydrogène inférieurs à 10 (O) et le nombre donneurs d'hydrogène inférieurs à 5 (OH, NH). Aussi, pour les poids moléculaires, tous les composés ont des valeurs inférieures à 500 Da, donc ils traversent facilement les membranes cellulaires.

Pour le nombre de violation, nous remarquons que tous les composés ont une violation nulle. Donc, ces composés répondent aux règles de Lipinski (règles des cinq), en suggérant que ces composés théoriquement n'auraient pas de problèmes avec la biodisponibilité orale. Pour la règle de Veber, nous remarquons que les valeurs de nombre de liaisons rotatives sont toutes inférieures à 10. Le faible nombre de liaisons rotatives (flexibilité réduite) dans les composés étudiés indique que ces ligands lors de la liaison avec une protéine ne changent que légèrement leur conformation.

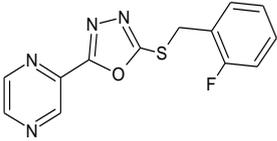
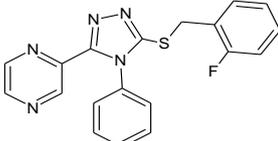
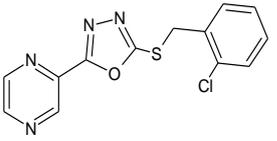
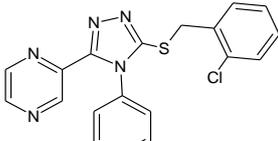
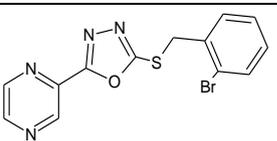
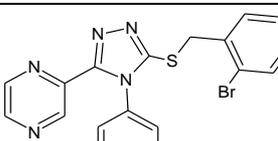
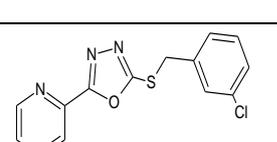
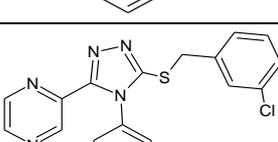
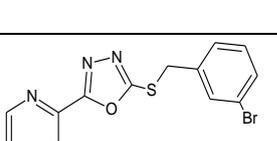
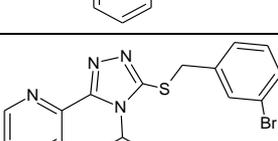
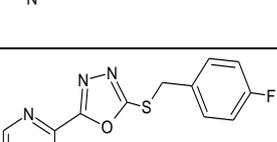
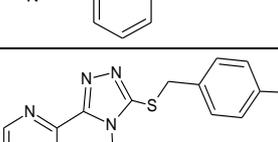
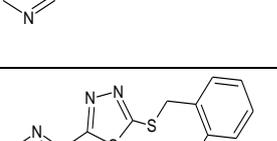
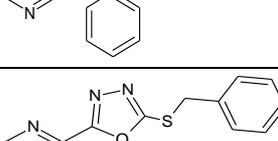
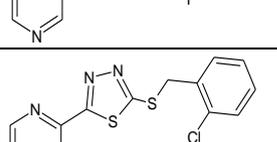
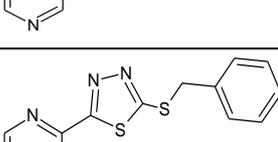
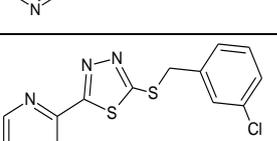
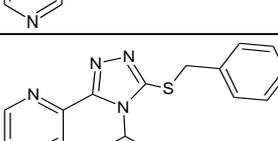
Pour les résultats de PSA, nous remarquons que tous les composés de la série étudiée ont des valeurs inférieures à  $140 \text{ \AA}^2$  ce qui montre la bonne prédiction de la biodisponibilité orale et le transport à travers les membranes. Nous pouvons conclure que ces composés étudiés sont en accord avec la règle de Veber. Le composé 21 possède la valeur maximale de pIC50, ainsi que pour l'étude de l'efficacité de ligand nous trouvons que le composé 21 a la plus grande valeur de LE qui est 0.43 cela lui permet d'avoir une bonne propriété physico-chimique et ADME. À partir de ces résultats obtenus ; nous pouvons conclure que il y a une relation proportionnelle entre pIC50 et (LE et LipE).

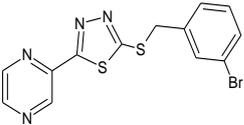
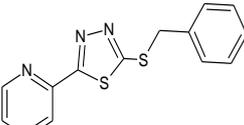
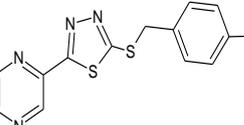
#### **V.4. Etude quantitatives des relations structure/activité (QSAR) dans les dérivés de la pyrazine**

##### **V.4.1 Données expérimentales**

Dans notre travail, nous avons utilisé deux paramètres biologiques: l'activité inhibitrice de télomérase contre la prolifération des cellules HEPG2 ( $1/IC_{50_{HEPG2}}$ ) et contre la prolifération des cellules SW1116 ( $1/IC_{50_{SW1116}}$ ). Cela, nous a permis de rassembler les données des deux activités pour 21 dérivés de la pyrazine (Tableau V.3) [2]. Les résultats biologiques de ces 21 dérivés de la pyrazine sont exploités par la méthode de régression linéaire multiple (MLR = multiple linear regression) dans le but d'établir les modèles QSAR.

**Tableau V.3 :** Structures chimiques et les activités biologiques expérimentales des molécules utilisées dans l'étude QSAR.

No	Structure	log(1/IC <sub>H</sub> EPG2)	log(1/IC <sub>SW</sub> 1116)	N o	Structure	log(1/IC <sub>HE</sub> PG2)	log(1/IC <sub>SW</sub> 1116)
1		4.90	4.78	12		4.72	4.73
2		4.91	4.82	13		4.73	4.74
3		5.37	5.61	14		4.93	4.71
4		5.12	5.10	15		4.71	5.18
5		5.24	5.30	16		4.97	5.30
6		4.78	4.65	17		5	5.39
7		5.02	5.17	18		4.95	4.47
8		4.97	5.12	19		4.95	5.10
9		5.15	4.92	20		4.77	4.81

10		5.75	5.28	21		6.11	5.83
11		4.99	4.77				

#### V.4.2. Sélection des descripteurs et méthodes de calcul

Un descripteur moléculaire peut être considéré comme la conséquence d'un processus logique et mathématique, appliqué à l'information chimique codifiée à travers la représentation d'une molécule [7]. Le choix des descripteurs dépend des outils dont on dispose, de la nature des composés décrits et de la propriété ciblée. L'information codée d'un descripteur moléculaire dépend du type de représentation moléculaire employée et de l'algorithme défini pour son calcul.

Tout d'abord, les vingt-un molécules de la pyrazine ont été pré-optimisées au moyen de la mécanique moléculaire, champ de force (MM+) en utilisant le logiciel HyperChem 8.0.6. Après cela, les structures minimisées obtenues précédemment ont été optimisées en utilisant la méthode semi-empirique PM3 par le même logiciel. Nous avons choisi une limite de norme du gradient de 0.01kcal/mol pour l'optimisation de la géométrie. Le module "propriétés QSAR" de l'HyperChem 8.0.6 a été utilisé pour calculer les paramètres suivants: la polarisabilité (Pol), la réfractivité molaire (MR), le coefficient de partage octanol/eau (log P), l'énergie d'hydratation (HE), le volume molaire (MV), la surface Moléculaire (SAG) et le poids moléculaire (MW) (Tableau V.4). Par contre, les descripteurs électroniques: le moment dipolaire ( $\mu$ ), l'énergie des orbitales frontières ( $E_{\text{HOMO}}$ ,  $E_{\text{LUMO}}$ ), les charges NBO ( $q_{\text{N1}}$ ,  $q_{\text{N4}}$ ,  $q_{\text{C2}}$ ,  $q_{\text{C3}}$ ,  $q_{\text{C5}}$  et  $q_{\text{C6}}$ ) ont été calculés par la méthode DFT (B3LYP / cc-PVDZ) en utilisant logiciel Gaussian 09. L'analyse de régression linéaire multiple des descripteurs moléculaires a été réalisée au moyen de logiciel SPSS 19 pour Windows [8].

Le tableau V.4 montre les 16 descripteurs retenus pour construire les modèles de QSAR.

Tableau V.4 : Descripteurs moléculaires utilisés dans l'étude QSAR.

N°	MV (A <sup>o3</sup> )	HE (kcal/mol)	log P	MR (A <sup>o3</sup> )	SAG (A <sup>o2</sup> )	MW (uma)	Pol (A <sup>o3</sup> )	E <sub>HOMO</sub> (au)	E <sub>LUMO</sub> (au)	μ (debye)	qN1	qN4	qC2	qC3	qC5	qC6
1	770.170	-12.540	1.940	79.140	466.470	288.300	28.820	-0.2392	-0.0794	0.886	-0.427	-0.450	0.026	0.020	0.046	0.107
2	791.620	-12.630	2.320	83.730	474.610	304.750	30.840	-0.2494	-0.0807	5.144	-0.424	-0.453	0.025	0.021	0.050	0.108
3	810.260	-12.580	2.600	86.540	485.200	349.200	31.540	-0.2397	-0.0796	0.887	-0.427	-0.450	0.026	0.020	0.046	0.107
4	809.750	-13.290	2.320	83.730	498.290	304.750	30.840	-0.2427	-0.0814	1.269	-0.426	-0.450	0.026	0.021	0.046	0.107
5	828.550	-13.240	2.600	86.540	505.960	349.200	31.540	-0.2468	-0.0817	1.498	-0.425	-0.450	0.026	0.021	0.046	0.107
6	777.830	-13.380	1.940	79.140	479.960	288.300	28.820	-0.2388	-0.0797	2.243	-0.445	-0.451	0.022	0.018	0.064	0.108
7	800.840	-11.390	2.290	85.580	486.180	304.360	31.180	-0.2339	-0.0842	5.024	-0.467	-0.449	0.024	0.014	0.066	0.137
8	822.050	-12.250	2.670	90.170	498.700	320.810	33.200	-0.2351	-0.0856	5.023	-0.467	-0.449	0.024	0.015	0.066	0.137
9	830.070	-12.540	2.670	90.170	511.100	320.810	33.200	-0.2378	-0.0860	1.986	-0.465	-0.449	0.026	0.016	0.065	0.136
10	848.390	-12.500	2.940	92.990	520.260	365.270	33.900	-0.2378	-0.0860	1.924	-0.465	-0.449	0.026	0.016	0.065	0.136
11	797.400	-12.610	2.290	85.580	496.490	304.360	31.180	-0.2349	-0.0846	2.137	-0.465	-0.449	0.025	0.015	0.065	0.136
12	961.110	-10.860	2.480	110.970	543.240	363.410	39.200	-0.2212	-0.0652	3.636	-0.470	-0.453	0.021	0.007	0.063	0.137
13	984.280	-10.690	2.860	115.560	550.540	379.870	41.210	-0.2200	-0.0634	4.963	-0.471	-0.453	0.021	0.006	0.063	0.137
14	997.060	-10.630	3.130	118.370	554.530	424.320	41.910	-0.2199	-0.0634	4.949	-0.471	-0.453	0.021	0.006	0.063	0.137
15	980.740	-11.450	2.860	115.560	562.490	379.870	41.210	-0.2244	-0.0671	4.190	-0.470	-0.453	0.022	0.008	0.063	0.136
16	999.400	-11.410	3.130	118.370	570.870	424.320	41.910	-0.2230	-0.0653	4.523	-0.471	-0.453	0.021	0.007	0.063	0.137
17	948.380	-11.540	2.480	110.970	543.200	363.410	39.200	-0.2233	-0.0640	4.275	-0.431	-0.454	0.022	0.013	0.040	0.135
18	769.230	-13.670	2.550	79.010	475.710	270.310	28.910	-0.2397	-0.0810	4.278	-0.424	-0.451	0.027	0.019	0.043	0.109
19	789.010	-12.890	2.890	85.450	490.320	286.370	31.270	-0.2327	-0.0829	1.449	-0.465	-0.449	0.025	0.015	0.065	0.137
20	939.490	-11.790	3.080	110.840	539.340	345.420	39.290	-0.2198	-0.0646	4.399	-0.470	-0.453	0.021	0.007	0.063	0.137
21	854.340	-11.490	2.940	92.990	526.340	365.270	33.900	-0.2363	-0.0857	2.507	-0.465	-0.449	0.026	0.016	0.065	0.136

### V.4.3. Résultats et discussions

Dans la première étape de notre étude, les différentes pyrazines substitués (Tableau V.2) ont été évaluées pour leur activité antiproliférative. Deux paramètres biologiques ( $pIC_{HEPG2}$  et  $pIC_{SW1116}$ ) ont été introduits dans cette étude et les résultats sont illustrés dans le tableau V.2. Afin de déterminer le rôle des caractéristiques structurelles, une étude QSAR a été réalisée. Une série de 21 dérivés de la pyrazine a été utilisée pour la production du modèle de régression multilinéaire. Différents descripteurs physico-chimiques ont été utilisés comme variables indépendantes et ont été corrélées avec l'activité biologique.

Le développement d'un modèle QSAR exige un ensemble de diverses données, pour cet effet, un nombre considérable de descripteurs doivent être considérés dans l'étude QSAR. Les descripteurs sont des valeurs numériques qui codent les différentes caractéristiques structurelles des molécules. La sélection d'un ensemble de descripteurs appropriés à partir d'un grand nombre de descripteurs utilisés, nécessite un procédé qui est capable de discriminer entre les paramètres.

La corrélation entre les activités biologiques ( $pIC50_{HEPG2}$  et  $pIC50_{SW1116}$ ) et les descripteurs exprimés par les relations suivantes:

$$\text{Modèle 1: } \log (1/IC50_{HEPG2}) = 2.784 + 0.416HE + 0.692\log P + 0.021SAG - 0.181Pol + 73.301qC3 \quad (1)$$

$$n=21; R=0.943; S= 0.131; F=24.090; Q=7.198$$

$$\text{Modèle 2: } \log (1/IC50_{SW1116}) = 18.243 - 0.041MV + 0.700HE + 0.583\log P + 0.140MR + 0.045SAG - 289.948qC2 - 49.834qC5 - 30.846qC6 - 0.104\mu - 86.539ELUMO \quad (2)$$

$$n=21; R=0.955; S= 0.143; F=10.427; Q=6.678$$

Les valeurs de la fraction de la variance peuvent varier entre 0 et 1. Le modèle QSAR doit considérer un  $R^2 > 0.6$  pour qu'il soit valide. Les valeurs  $R = 0.943$  ( $R^2 = 0,889$ ) (modèle 1) et  $R = 0.955$  ( $R^2 = 0.912$ ) (modèle 2); nous permet d'indiquer clairement la corrélation entre les différents descripteurs utilisés et les activités biologiques  $pIC50_{HEPG2}$  et  $pIC50_{SW1116}$  respectivement.

La valeur F est jugée la signification statistique au niveau de 95%, pour toutes les valeurs de F calculées est supérieures par rapport aux valeurs lues dans la table du Fischer (voire l'annexe). La valeur positive de facteur de qualité (Q) de ces deux modèles de QSAR **1** et **2** suggère son haut pouvoir prédictif [9].

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

Dans les deux modèles, nous remarquons que toute augmentation de la surface moléculaire provoque à une augmentation de l'activité biologique, qui se traduit par l'augmentation de la surface du contact entre le ligand et le récepteur.

La figure IV.2 présente la corrélation entre les valeurs prédites et expérimentales pour les deux activités biologiques ( $pIC_{50_{HEPG2}}$  et  $pIC_{50_{SW1116}}$ ) des dérivés de la pyrazine. Du point de vue des performances statistiques, ces modèles présentent des coefficients de corrélation  $R^2$  significatifs de valeur. Les valeurs de  $R^2$  est égale à 0,889 et 0,912, pour les modèles **1** et **2**, respectivement. Ils indiquent que ces modèles peuvent être appliquées avec succès pour prédire l'activité antiproliférative contre HEPG2 et SW1116 pour les pyrazines étudiés.

Afin de tester la validité de pouvoir prédictif des modèles MLR sélectionnés (**1** et **2**), la technique de validation croisée LOO (leave-one-out) a été utilisée. Les modèles développés ont été validés par le calcul des paramètres statistiques suivants: la somme des carrés résiduels (PRESS), la somme des carrés totaux (SSY), qualité d'ajustement  $R^2_{adj}$ , coefficient de corrélation de validation croisée  $R^2_{CV}$ , la validation standard des erreurs de prédiction ( $S_{PRESS}$ ) et l'erreur de prédiction (PE) (Tableau V.5).

**Tableau V.5** : Paramètres de validation croisée.

Model	PRESS	SSY	PRESS/SSY	$S_{PRESS}$	$R_{cv}^2$	$R_{adj}^2$	6PE
<b>1</b>	0.259	2.341	0.111	0.111	0.889	0.852	0.098
<b>2</b>	0.205	2.342	0.087	0.099	0.912	0.825	0.077

La somme résiduelle prédictive des carrés (PRESS, *Predictive Residual Sum of Squares*) est une mesure des capacités prédictives d'un modèle. La petite valeur de PRESS signifie la meilleure prévisibilité du modèle [9]. En outre, pour un modèle

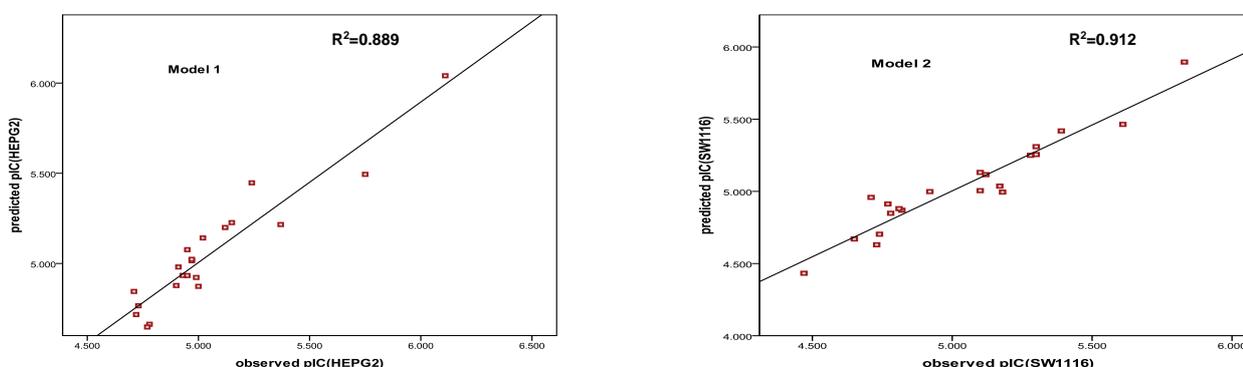
raisonnable de QSAR, le rapport PRESS/SSY doit être inférieur à 0,4 . Les résultats présentés dans le tableau V.5 indiquent que ce rapport est de 0,111 et 0.087 pour les modèles **1** et **2** respectivement.

Les valeurs élevées du  $R^2_{cv}$  et  $R^2_{adj}$  sont des critères essentiels pour une meilleure qualification des modèles QSAR [9]. Les résultats obtenus de  $R^2_{cv}$  ( $R^2_{adj}$ ) pour les modèles QSAR **1** et **2** sont 0.889 (0.852) et 0.912 (0.825) respectivement.

Nous pouvons également utiliser le paramètre  $S_{PRESS}$  qui rend compte des variations d'erreur des prédictions. Les modèles QSAR développés présente des faibles valeurs de  $S_{press}$  (<0.200) ce qui indique que les deux modèles présentent des petites valeurs résiduelles entre les activités biologiques observées et prédites.

Le calcul des valeurs prédites ( $\log(1/IC_{HEPG2})_{pred.}$ ,  $\log(1/IC_{SW1116})_{pred.}$ ) et résiduelles ( $\log(1/IC_{HEPG2})_{resd.}$ ,  $\log(1/IC_{SW1116})_{resd.}$ ) de l'activité biologique des pyrazines est un autre moyen pour estimer le pouvoir prédictif des modèles développés (Tableau V.6).

La figure V.4 montre les courbes de corrélation obtenues par la régression linéaire entre les valeurs prédites par rapport aux valeurs expérimentales pour les deux activités biologiques ( $pIC_{50_{HEPG2}}$  et  $pIC_{50_{SW1116}}$ ). Les courbes obtenues pour les deux modèles présentent des valeurs élevées de coefficient de détermination, avec  $R^2 = 0,889$  et 0,912. Ils indiquent que les modèles **1** et **2**, peuvent être appliquées avec succès pour prédire l'activité antiproliférative contre HEPG2 et SW1116 pour des dérivés de la pyrazine.



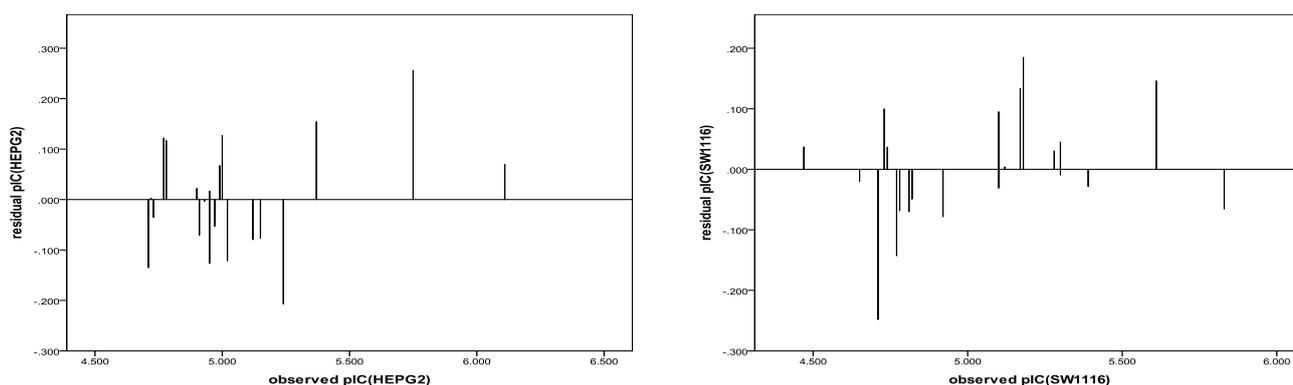
**Figure V.4** : Corrélation entre les activités biologiques expérimentales et prédites pour les modèles **1** et **2**.

**Tableau V.6** : Valeurs des activités prédites, observables et résiduelles de l'ensemble de dérivés de la pyrazine.

N°	Log(1/IC <sub>HEPG2</sub> )			Log(1/IC <sub>SW1116</sub> )		
	Obs.	Pred.	Res.	Obs.	Pred.	Res.
1	4.900	4.878	0.022	4.780	4.848	-0.068
2	4.910	4.981	-0.071	4.820	4.869	-0.049
3	5.370	5.216	0.154	5.610	5.464	0.146
4	5.120	5.199	-0.079	5.100	5.131	-0.031
5	5.240	5.447	-0.207	5.300	5.309	-0.009
6	4.780	4.663	0.117	4.650	4.670	-0.020
7	5.020	5.142	-0.122	5.170	5.036	0.134
8	4.970	5.016	-0.046	5.120	5.116	0.004
9	5.150	5.227	-0.077	4.920	4.998	-0.078
10	5.750	5.494	0.256	5.280	5.250	0.030
11	4.990	4.923	0.067	4.770	4.913	-0.143
12	4.720	4.717	0.003	4.730	4.630	0.100
13	4.730	4.766	-0.036	4.740	4.703	0.037
14	4.930	4.934	-0.004	4.710	4.958	-0.248
15	4.710	4.845	-0.135	5.180	4.995	0.185
16	4.970	5.023	-0.053	5.300	5.255	0.045
17	5.000	4.873	0.127	5.390	5.419	-0.029
18	4.950	4.933	0.017	4.470	4.433	0.037
19	4.950	5.077	-0.127	5.100	5.005	0.095
20	4.770	4.648	0.122	4.810	4.880	-0.070
21	6.110	6.041	0.069	5.830	5.896	-0.066

Pour rechercher la présence d'une erreur systématique dans l'élaboration des modèles QSAR, les résidus des valeurs prédites des deux activités biologiques  $pIC_{HEPG2}$  et  $pIC_{SW1116}$  ont été tracés en fonction des valeurs expérimentales, comme le montre la figure V.5.

La répartition des résidus dans les deux côtés de l'axe des abscisses indique l'absence de l'erreur systématique [10]. Il indique que ces deux modèles peuvent être appliqués avec succès pour prédire les activités biologiques  $pIC_{HEPG2}$  et  $pIC_{SW1116}$  pour l'ensemble des pyrazines utilisées dans le développement des modèles QSAR.



**Figure V.5 :** Corrélation entre les valeurs résiduelles et valeurs expérimentales de  $\log(1/IC_{HEPG2})$  et  $\log(1/IC_{SW1116})$ .

## V.5. Conclusion

La présente étude offre une comparaison structurale entre les vingt un composés des pyrazines. Elle fournit également une discussion de plusieurs approximations qualitatives de la structure relation activité/propriété. Les valeurs de la polarisabilité sont généralement proportionnelles aux valeurs des volumes et des surfaces et la réfractivité. L'énergie d'hydratation la plus importante, en valeur absolue, est celle du composé 18 (13.67 kcal/mol).

Par conséquent, ils ont la meilleure distribution dans les tissus. La plupart des composés étudiés, ont un Log P dans la plage des normes (intervalle 0-3). Ces composés ont pour conséquence une meilleure distribution dans les tissus c'est à-dire une bonne solubilité et une perméabilité efficace.

Les dérivés de la pyrazine utilisées dans cette étude sont caractérisés par son activité antiproliférative. La modélisation moléculaire et l'analyse QSAR ont été effectuées pour déterminer quantitativement l'effet de la structure moléculaire de composés étudiés sur leur activité biologique. Des modèles mathématiques ont été développés pour prédire l'activité inhibitrice de télomérase contre la lignée cellulaire HEPG2 et la ligne cellulaire SW1116 de certains dérivés de la pyrazine. La validité des modèles ont été établie par la détermination des paramètres statistiques appropriés. Le modèle mis en place a été utilisé pour prédire l'activité biologique des pyrazines étudiés. Les valeurs prédites sont plus proches que les valeurs expérimentales pour les deux activités biologiques  $pIC_{HEPG2}$  et  $pIC_{SW1116}$ .

## Références bibliographiques

- [1] Phuong H.T., Thèse de doctorat, Synthèse et étude des relations structure/activité quantitatives (QSAR/2D) d'analogues Benzo[c]phénanthridiniques, Université d'Angers (2007).
- [2] Zhang Y. B., Wang X. L., Liu W., Yang Y. S., Tang J. F. and Zhu H. L., Design, synthesis and biological evaluation of heterocyclic azoles derivatives containing pyrazine moiety as potential telomerase inhibitors, *Bioorg. Med. Chem.*, 20, 6356 (2012).
- [3] G.L. Patrick, H.D. Winter, W. Langenaeker, J.P. Tollenaere, Computational Medicinal Chemistry for Drug Discovery, Marcel Dekker, New York(1995).
- [4] Kier L.B. Molecular Orbital Theory in Drug Research, Academic Press, New York, NY, USA(1981).
- [5] Kerns E.H., Di L., Drug-like Properties: Concepts, Structure Design and Methods: from ADME to Toxicity Optimization, Academic Press, USA (2008).
- [6] MONGE Aurélien, "Création et utilisation de chimio-thèques optimisées pour la recherche « in silico » de nouveaux composés bioactifs. » ; Thèse de Doctorat de l'université D'Orléans; 29, 30(2006).
- [7] Gonzalez-Diaz H., Olazabal E., Santana L., Uriarte E., Gonzalez-Diaz Y., Castanedo N., QSAR study of anticoccidial activity for diverse chemical compounds: prediction and experimental assay of trans-2-(2-nitrovinyl)furan, *Bioorg. Med. Chem.*, 15, 962 (2007).
- [8] SPSS 19 software packages, SPSS Inc., 444 north Michigan avenue, suite 3000, Chicago, illinois, 60611, USA.
- [9] S.O. Podunavac-Kuzmanović, D.D. Cvetković, D.J. Barna, QSAR Analysis of 2-Amino or 2-Methyl-1-Substituted Benzimidazoles Against *Pseudomonas aeruginosa*, *Int. J. Mol. Sci.*, 10, 1670- (2009).
- [10] Jalali-Heravi M. and Kyani A., Use of Computer-Assisted Methods for the Modeling of the Retention Time of a Variety of Volatile Organic Compounds: A PCA-MLR-ANN Approach, *J. Chem. Inf. Comput. Sci.*, 44, 1328 (2004).

# *Conclusion Générale*

Dans ce travail, nous avons présenté une étude computationnelle dans le champ d'application est à priori très étendu puisque les méthodes de la chimie théorique et informatique sont désormais capables de reproduire la plupart des grandeurs physico-chimiques ou de la réactivité de manière satisfaisante. Différentes méthodes de modélisation moléculaire ont été utilisées dans notre travail.

Dans cette étude, nous avons décrit les propriétés structurales et électroniques sur le noyau de la pyrazine. Les résultats obtenus ont été comparés à des données expérimentales. Les propriétés géométriques, électroniques et énergétiques ont été déterminées en utilisant les différentes méthodes de calcul : *ab initio*/HF, *ab initio*/MP2 et DFT/B3LYP et les différentes bases 6-311++G (d,p) et cc-PVDZ.

La comparaison structurale, électronique du noyau de base de la pyrazine présente des résultats similaires entre les résultats expérimentaux et les différentes méthodes de calcul. Cette étude démontre que les calculs établis par la méthode DFT/B3LYP avec la base cc-PVDZ sont les plus proches aux valeurs expérimentales, donc la méthode DFT/B3LYP présente une approche puissante pour comprendre la structure moléculaire des composés organiques de taille moyenne. L'efficacité de ces méthodes utilisées a été confirmée par la comparaison des paramètres structuraux entre les données expérimentales et les résultats obtenus par les méthodes théoriques *ab initio*/HF, *ab initio*/MP2 et DFT/B3LYP.

En suite, nous avons étudié l'effet de la substitution sur les paramètres électroniques et énergétiques du noyau de base ; les résultats montrent que la nature de type de substituant (donneur, accepteur) influe sur les paramètres électroniques et énergétiques du noyau de base de la pyrazine. En effet, cette étude nous a permis de prédire la réactivité chimique des dérivés de la pyrazine.

Pour avoir l'effet de la substitution sur les paramètres électroniques et énergétiques et leur incidence sur la stabilité et la réactivité chimique de la pyrazine, deux séries de molécules substituées ont été étudiées.

Une étude qualitative de la relation structure-propriétés / activités a été réalisée sur une série des dérivés de la pyrazine. Les molécules étudiées présentent un intérêt pharmacologique ou bien biologique. La nature des groupements qui se lient sur le noyau de base des molécules étudiées affecte sur leurs propriétés physicochimiques et par conséquence sur leurs propriétés pharmacologiques. Tous les critères de la règle de la Lipinski et Veber sont vérifiés.

L'étude quantitative de la relation structure-propriétés (QSAR) a été effectuée sur une série de dérivés de la pyrazine. Deux modèles QSAR ont été établis en utilisant la méthode de régression multilinéaire (MLR). Les modèles obtenus peuvent être utilisés pour prédire l'activité inhibitrice de télomérase contre la ligne cellulaire HEPG2 ( $pIC_{50_{HEPG2}}$ ) et la l'activité inhibitrice de télomérase contre la ligne cellulaire SW1116 ( $pIC_{50_{SW1116}}$ ).

Le pouvoir prédictif des modèles obtenus a été confirmé par la méthode de validation croisée LOO. Une forte corrélation a été observée entre les valeurs expérimentales et prédites des activités biologiques ( $pIC_{50_{HEPG2}}$ ) et ( $pIC_{50_{SW1116}}$ ), ce qui indique la validité et la qualité des modèles QSAR obtenu.

En perspectives, une étude sur la structure moléculaire des dérivés de la pyrazine avec la dynamique moléculaire pour avoir une approche sur le milieu biologique, ainsi qu'une étude de la relation structure-activité par le docking.

# *Annexes*

# *Annexe A*

# Table du F (de Fischer)

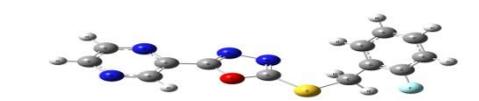
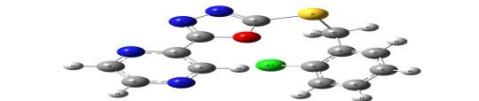
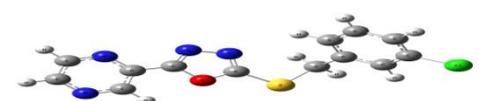
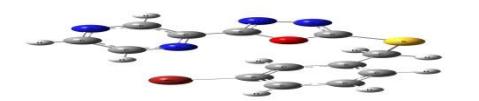
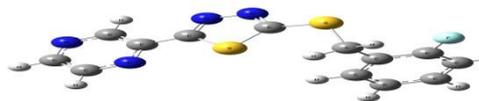
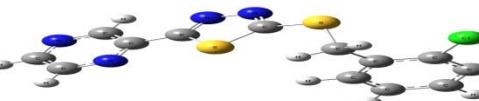
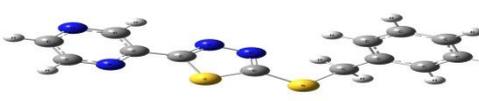
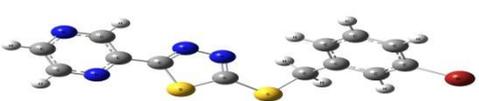
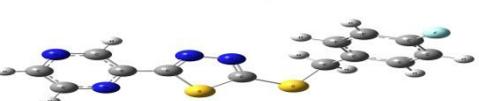
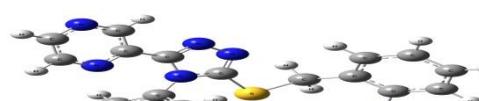
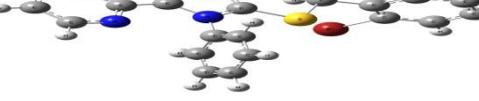
F limite à p 0,05

(comparaison de 2 variances)

		degré de liberté du numérateur																		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	30	40	50
degré de liberté du dénominateur	1	161	199	216	225	230	234	237	239	241	242	243	244	245	245	246	248	250	251	252
	2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5
	3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73	8,71	8,70	8,66	8,62	8,59	8,58
	4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89	5,87	5,86	5,80	5,75	5,72	5,70
	5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62	4,56	4,50	4,46	4,44
	6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94	3,87	3,81	3,77	3,75
	7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,51	3,44	3,38	3,34	3,32
	8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,22	3,15	3,08	3,04	3,02
	9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	3,01	2,94	2,86	2,83	2,80
	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85	2,77	2,70	2,66	2,64
	11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72	2,65	2,57	2,53	2,50
	12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,62	2,54	2,47	2,43	2,40
	13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58	2,55	2,53	2,46	2,38	2,34	2,31
	14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,46	2,39	2,31	2,27	2,24
	15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45	2,42	2,40	2,33	2,25	2,20	2,18
	16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,35	2,28	2,19	2,15	2,12
	17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35	2,33	2,31	2,23	2,15	2,10	2,08
	18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,27	2,19	2,11	2,06	2,04
	19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28	2,26	2,23	2,16	2,07	2,03	2,00
	20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25	2,22	2,20	2,12	2,04	1,99	1,97
	21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,22	2,20	2,18	2,10	2,01	1,96	1,94
	22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,20	2,17	2,15	2,07	1,98	1,94	1,91
	23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24	2,20	2,18	2,15	2,13	2,05	1,96	1,91	1,88
	24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,22	2,18	2,15	2,13	2,11	2,03	1,94	1,89	1,86
	25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,14	2,11	2,09	2,01	1,92	1,87	1,84
	26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,12	2,09	2,07	1,99	1,90	1,85	1,82
	27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,17	2,13	2,10	2,08	2,06	1,97	1,88	1,84	1,81
	28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	2,09	2,06	2,04	1,96	1,87	1,82	1,79
	29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,08	2,05	2,03	1,94	1,85	1,81	1,77
	30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,06	2,04	2,01	1,93	1,84	1,79	1,76
	31	4,16	3,30	2,91	2,68	2,52	2,41	2,32	2,25	2,20	2,15	2,11	2,08	2,05	2,03	2,00	1,92	1,83	1,78	1,75
	32	4,15	3,29	2,90	2,67	2,51	2,40	2,31	2,24	2,19	2,14	2,10	2,07	2,04	2,01	1,99	1,91	1,82	1,77	1,74
	33	4,14	3,28	2,89	2,66	2,50	2,39	2,30	2,23	2,18	2,13	2,09	2,06	2,03	2,00	1,98	1,90	1,81	1,76	1,72
	34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17	2,12	2,08	2,05	2,02	1,99	1,97	1,89	1,80	1,75	1,71
	35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11	2,07	2,04	2,01	1,99	1,96	1,88	1,79	1,74	1,70
	36	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15	2,11	2,07	2,03	2,00	1,98	1,95	1,87	1,78	1,73	1,69
	37	4,11	3,25	2,86	2,63	2,47	2,36	2,27	2,20	2,14	2,10	2,06	2,02	2,00	1,97	1,95	1,86	1,77	1,72	1,68
	38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,05	2,02	1,99	1,96	1,94	1,85	1,76	1,71	1,67
	39	4,09	3,24	2,85	2,61	2,46	2,34	2,26	2,19	2,13	2,08	2,04	2,01	1,98	1,95	1,93	1,85	1,75	1,70	1,66
	40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97	1,95	1,92	1,84	1,74	1,69	1,66
45	4,06	3,20	2,81	2,58	2,42	2,31	2,22	2,15	2,10	2,05	2,01	1,97	1,94	1,92	1,89	1,81	1,71	1,66	1,63	
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95	1,92	1,89	1,87	1,78	1,69	1,63	1,60	
55	4,02	3,16	2,77	2,54	2,38	2,27	2,18	2,11	2,06	2,01	1,97	1,93	1,90	1,88	1,85	1,76	1,67	1,61	1,58	
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,89	1,86	1,84	1,75	1,65	1,59	1,56	
65	3,99	3,14	2,75	2,51	2,36	2,24	2,15	2,08	2,03	1,98	1,94	1,90	1,87	1,85	1,82	1,73	1,63	1,58	1,54	
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89	1,86	1,84	1,81	1,72	1,62	1,57	1,53	
75	3,97	3,12	2,73	2,49	2,34	2,22	2,13	2,06	2,01	1,96	1,92	1,88	1,85	1,83	1,80	1,71	1,61	1,55	1,52	
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88	1,84	1,82	1,79	1,70	1,60	1,54	1,51	
85	3,95	3,10	2,71	2,48	2,32	2,21	2,12	2,05	1,99	1,94	1,90	1,87	1,84	1,81	1,79	1,70	1,59	1,54	1,50	
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,90	1,86	1,83	1,80	1,78	1,69	1,59	1,53	1,49	
95	3,94	3,09	2,70	2,47	2,31	2,20	2,11	2,04	1,98	1,93	1,89	1,86	1,82	1,80	1,77	1,68	1,58	1,52	1,48	
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85	1,82	1,79	1,77	1,68	1,57	1,52	1,48	
110	3,93	3,08	2,69	2,45	2,30	2,18	2,09	2,02	1,97	1,92	1,88	1,84	1,81	1,78	1,76	1,67	1,56	1,50	1,47	
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,87	1,83	1,80	1,78	1,75	1,66	1,55	1,50	1,46	
130	3,91	3,07	2,67	2,44	2,28	2,17	2,08	2,01	1,95	1,90	1,86	1,83	1,80	1,77	1,74	1,65	1,55	1,49	1,45	
140	3,91	3,06	2,67	2,44	2,28	2,16	2,08	2,01	1,95	1,90	1,86	1,82	1,79	1,76	1,74	1,65	1,54	1,48	1,44	
150	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,85	1,82	1,79	1,76	1,73	1,64	1,54	1,48	1,44	
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80	1,77	1,74	1,72	1,62	1,52	1,46	1,41	
250	3,88	3,03	2,64	2,41	2,25	2,13	2,05	1,98	1,92	1,87	1,83	1,79	1,76	1,73	1,71	1,61	1,50	1,44	1,40	
300	3,87	3,03	2,63	2,40	2,24	2,13	2,04	1,97	1,91	1,86	1,82	1,78	1,75	1,72	1,70	1,61	1,50	1,43	1,39	
400	3,86	3,02	2,63	2,39	2,24	2,12	2,03	1,96	1,90	1,85	1,81	1,78	1,74	1,72	1,69	1,60	1,49	1,42	1,38	
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81	1,77	1,74	1,71	1,69	1,59	1,48	1,42	1,38	
600	3,86	3,01	2,62	2,39	2,23	2,11	2,02	1,95	1,90	1,85	1,80	1,77	1,74	1,71	1,68	1,59	1,48	1,41	1,37	
700	3,85	3,01	2,62	2,38	2,23	2,11	2,02	1,95	1,89	1,84	1,80	1,77	1,73	1,71	1,68	1,59	1,48	1,41	1,37	
1000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,80	1,76	1,73	1,70	1,68	1,58	1,47	1,41	1,36	

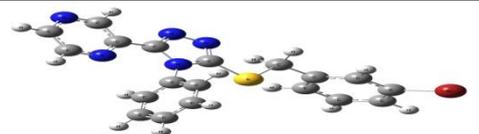
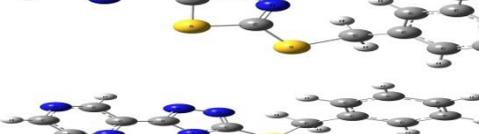
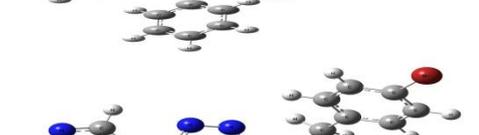
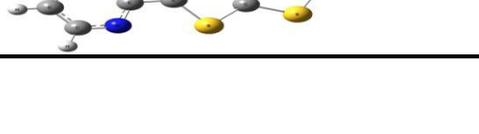
# *Annexe B*

# *Nomenclature des molécules*

<i>N°</i>	<i>Structure chimique</i>	<i>Nomenclature</i>
1		2-(5-{{[(2-fluorophenyl)methyl]sulfanyl}}-1,3,4-oxadiazol-2-yl)pyrazine
2		2-(5-{{[(2-chlorophenyl)methyl]sulfanyl}}-1,3,4-oxadiazol-2-yl)pyrazine
3		2-(5-{{[(2-bromophenyl)methyl]sulfanyl}}-1,3,4-oxadiazol-2-yl)pyrazine
4		2-(5-{{[(3-chlorophenyl)methyl]sulfanyl}}-1,3,4-oxadiazol-2-yl)pyrazine
5		2-(5-{{[(3-bromophenyl)methyl]sulfanyl}}-1,3,4-oxadiazol-2-yl)pyrazine
6		2-(5-{{[(4-fluorophenyl)methyl]sulfanyl}}-1,3,4-oxadiazol-2-yl)pyrazine
7		2-(5-{{[(2-fluorophenyl)methyl]sulfanyl}}-1,3,4-thiadiazol-2-yl)pyrazine
8		2-(5-{{[(2-chlorophenyl)methyl]sulfanyl}}-1,3,4-thiadiazol-2-yl)pyrazine
9		2-(5-{{[(3-chlorophenyl)methyl]sulfanyl}}-1,3,4-thiadiazol-2-yl)pyrazine
10		2-(5-{{[(3-bromophenyl)methyl]sulfanyl}}-1,3,4-thiadiazol-2-yl)pyrazine
11		2-(5-{{[(4-fluorophenyl)methyl]sulfanyl}}-1,3,4-thiadiazol-2-yl)pyrazine
12		2-(5-{{[(2-fluorophenyl)methyl]sulfanyl}}-4-phenyl-4H-1,2,4-triazol-3-yl)pyrazine
13		2-(5-{{[(2-chlorophenyl)methyl]sulfanyl}}-4-phenyl-4H-1,2,4-triazol-3-yl)pyrazine
14		2-(5-{{[(2-bromophenyl)methyl]sulfanyl}}-4-phenyl-4H-1,2,4-triazol-3-yl)pyrazine
15		2-(5-{{[(3-chlorophenyl)methyl]sulfanyl}}-4-phenyl-4H-1,2,4-triazol-3-yl)pyrazine

## *Nomenclature des molécules*

---

16		2-(5-[[3-bromophenyl)methyl]sulfanyl)- 4-phenyl-4H-1,2,4-triazol-3-yl)pyrazine
17		2-(5-[[4-fluorophenyl)methyl]sulfanyl)- 4-phenyl-4H-1,2,4-triazol-3-yl)pyrazine
18		2-[5-(benzylsulfanyl)-1,3,4-oxadiazol-2- yl]pyrazine
19		2-[5-(benzylsulfanyl)-1,3,4-thiadiazol-2- yl]pyrazine
20		2-[5-(benzylsulfanyl)-4-phenyl-4H-1,2,4- triazol-3-yl]pyrazine
21		2-(5-[[4-bromophenyl)methyl]sulfanyl)- 1,3,4-thiadiazol-2-yl)pyrazine

---

# *Annexe C*

# Quantitative Structure Anti-Proliferative Activity Against HEPG2 and SW1116 Relationships in a Series of Pyrazine Derivatives

Fatima Soualmia<sup>1,2</sup>, Salah Belaidi<sup>2,\*</sup>, Houmam Belaidi<sup>2,3</sup>, Nouredine Tchouar<sup>1</sup>, and Zineb Almi<sup>2</sup>

<sup>1</sup>Laboratory of Modeling and Optimization of Industrial Systems, University of Science and Technology of Oran-Mohamed Boudiaf (USTO-MB) 31036, Algeria

<sup>2</sup>Group of Computational and Medicinal Chemistry, LMCE Laboratory, University of Biskra, BP 145 Biskra 07000, Algeria

<sup>3</sup>Institut des Sciences Chimiques de Rennes, UMR 6226 CNRS-Université de Rennes 1, Campus de Beaulieu, 35042, Rennes Cedex, France

Quantitative structure activity relationship (QSAR) methods have been applied in several scientific studies including chemistry, biology, and toxicology and drug discovery to predict and classify biological activities of virtual or newly-synthesized compounds. QSAR models can also be used in designing new chemical entities and are now regarded as essential tools in pharmaceutical industries to identify promising hits and generate high quality leads in the early stages of drug discovery. QSAR studies were performed on a series of pyrazine as anti-proliferative agents. A multiple linear regression (MLR) procedure was used to envisage the relationships between molecular descriptors and the activity of pyrazine derivatives. The predictivity of the model was estimated by cross-validation with the leave-one-out method. Our results suggest a QSAR model based of the following descriptors:  $\log P$ , HE, SAG, Pol,  $qC_3$  for the anti-proliferative activity against the HEPG2 (human liver cancer cell) and  $\log P$ , HE, MR, SAG, Vol,  $qC_2$ ,  $qC_5$ ,  $qC_6$ , LUMO for the anti-proliferative activity against the SW1116 (human colorectal carcinoma cell). To confirm the predictive power of the models, an external set of molecules was used. High correlation between experimental and predicted activity values was observed, indicating the validation and the good quality of the derived QSAR models.

**Keywords:** Anti-Proliferative, MLR, Pyrazine Derivatives, SAR, QSAR.

## 1. INTRODUCTION

Quantitative Structure–Activity Relationships (QSARs)<sup>1–8</sup> are attempts to correlate between physicochemical parameters of chemical structures and their biological activity. QSAR is based on the general principle of medicinal chemistry that the biological activity of a ligand or compound is related to its molecular structure or properties, and structurally similar molecules may have similar biological activities.<sup>9</sup>

Such molecular structural information is encoded in molecular descriptors and a QSAR model defines mathematical relationships between descriptors and biological activities of known ligands to predict unknown ligand activities. QSAR methods have been applied in several scientific studies including chemistry, biology,

toxicology and drug discovery to predict and classify biological activities<sup>10,11</sup> of virtual or newly synthesized compounds.<sup>12–16</sup>

QSAR models can also be used in designing new chemical entities (NCEs) and are now regarded as essential tools in pharmaceutical industries to identify promising hits and generate high quality leads in the early stages of drug discovery.<sup>15–18</sup>

The molecular modeling and QSAR calculations are used in many fields specially, physics, chemistry, ecology, material science as well as tissue engineering<sup>19,20</sup> and drug design<sup>21–24</sup> molecular modeling and quantum calculations<sup>25</sup> helps in studying the biological activity of the new inhibitors and the physical aspect of the interaction with the virus and selecting the possibly best compound to synthesize,<sup>26,27</sup> thus, saving time, effort and money.<sup>28</sup>

\*Author to whom correspondence should be addressed.

The process of drug development is time-consuming and cost-intensive. Several years are required for lead identification, optimization,<sup>29</sup> *in vitro*<sup>30–39</sup> and *in vivo*<sup>40–46</sup> testing before starting the first clinical trials.<sup>47</sup>

Density functional theory (DFT) has become the most popular quantum mechanical method for calculating molecular properties. It has been used to calculate the excited state properties of physical, chemical, and biological systems.<sup>48</sup>

The heterocyclic compounds have attracted numerous attentions due to their wide applications in medicinal chemistry research. Heterocyclic compounds are cyclic compounds with at least two different elements as ring members' atoms, the commonest atoms include nitrogen, oxygen and sulphur.<sup>49</sup>

Pyrazine is the heterocyclic compound having two nitrogen atoms in the para-position of the six-membered ring. Ligands containing a pyrazine ring are widely studied and their *p*-donor properties are interesting.<sup>50</sup> Pyrazine has been paid great attention, because the diazine rings form an important class of compounds presented in several natural and synthetic compounds.<sup>51</sup> Pyrazine derivatives have been widely used in the fields of medicinal chemistry for the skeleton of biologically active sites,<sup>52,53</sup> some of the pyrazine derivative contains various pharmacological effects: anti-cancer, antidepressant, anti-diabetic, antitumor, antibiotic and diuretic activities.<sup>54</sup>

Following our interest in this field, our present research is aimed to describe the quantitative structure-activity relationship study on pyrazines and to develop QSAR models for these compounds with respect to their anti-proliferative activity against the HEPG2 (human liver cancer cell) and the SW1116 (human colorectal carcinoma cell).

The paper deals with a specific organization form of molecular matter. Other forms are given for example in the Refs. [55, 56].

## 2. EXPERIMENTAL DETAILS

### 2.1. Data Set

In the present study a data set of pyrazine derivatives as anti-proliferative activity against the HEPG2: human liver cancer cell  $\log(1/IC_{HEPG2})$  and SW1116: human colorectal carcinoma cell  $\log(1/IC_{SW1116})$ . Twenty one molecules presented in (Table I), were adopted as reported by Zhang et al.<sup>57</sup> The reported IC 50 values ( $\mu M$ ) have been converted to the logarithmic scale [pIC 50], for QSAR study.

### 2.2. Descriptors Generation

Firstly, the twenty-one investigated molecules were pre-optimized using the Molecular Mechanics Force Field (MM+) method included in HyperChem version 8.0.6 package.<sup>58</sup> The resultant minimized structures were further refined using the semi empirical PM3 Hamiltonian as implemented also in HyperChem. The gradient norm

limit of 0.01 kcal/Å was chosen for the geometry optimization. PM3 optimized geometry was used to calculate a number of physicochemical descriptors: surface area grid (SAG), molar volume (Vol), hydration energy (HE), partition coefficient octanol/water ( $\log P$ ), the molar refractivity (MR), molar polarizability (Pol) and molar weight (MW). The Quantum Chemical descriptors: dipole moment (DM), Energy of frontier orbital's  $E_{HOMO}$  and  $E_{LUMO}$  and atomic net charges (qN1, qN4, qC2, qC3, qC5 and qC6) were computed using Gaussian 09W software<sup>59</sup> by using DFT/B3LYP with cc-PVDZ basis set.

Calculation of  $\log P$  is carried out using atomic parameters derived by Viswanadhan and coworkers.<sup>60</sup> Computation of molar refractivity was made via the same method as  $\log P$ . Atomic contributions to the refractivity presented by Ghose and Crippen have been used in our study.<sup>61</sup> Solvent-accessible surface bounded molecular volume and Van der Waals surface-bounded molecular volume calculations are based on a grid method derived by Bodor et al.,<sup>62</sup> using the atomic radii of Gavezzotti.<sup>63</sup> Polarizability was estimated from additively scheme given by Miller with a 3% in precision for the calculation,<sup>64</sup> where different increments are associated with different atom types.

### 2.3. Regression Analysis

A relationship between independent and dependent variables (physicochemical descriptors and biological activities, respectively) were determined statistically using regression analysis. In the present work, Multiple Linear Regression MLR analysis of molecular descriptors was carried out using the stepwise strategy in SPSS version 19 for Windows.<sup>65</sup>

## 3. RESULTS AND DISCUSSION

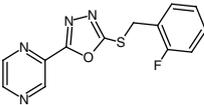
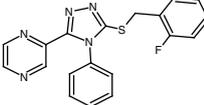
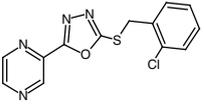
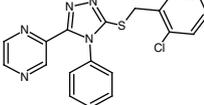
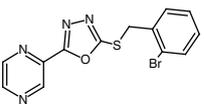
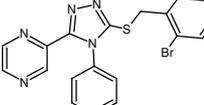
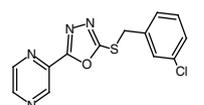
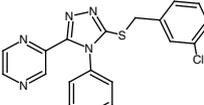
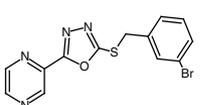
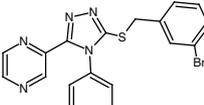
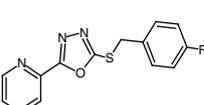
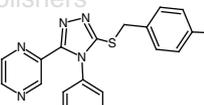
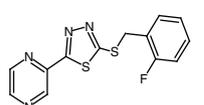
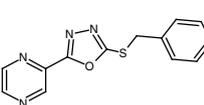
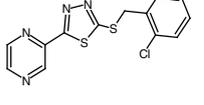
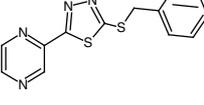
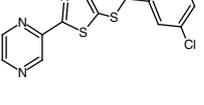
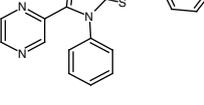
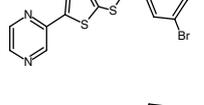
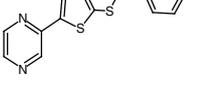
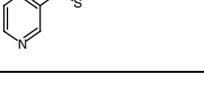
In the present work, twenty-one derivatives of pyrazine were evaluated for their inhibitory activities against HEPG2 and SW1116 (Table I). In order to determine the role of structural features, we used QSAR studies.

### 3.1. Structure Property/Activity Relationships (SPR/SAR)

We have studied sixteen physical and chemical proprieties of a series of twenty-one pyrazine derivatives; these proprieties are calculated and described in (Table II).

The attractive part of the Van der Waals interaction is a good measure of the polarizability. Molecular polarizability of a molecule characterizes the capability of its electronic system, and it plays an important role in modeling many molecular properties and biological activities,<sup>66</sup> also, the molar refractivity (MR) is important criterion to measure the steric factor and designated as a simple measure of the volume occupied either by individual atom or cluster (group) of atoms.<sup>67</sup> Molar refractivity and polarizability relatively increase with the size and the molecular weight

**Table I.** Chemical structures and experimental activity of the molecules under study.

No	Structure	log(1/ IC <sub>HEPG2</sub> )	log(1/ IC <sub>SW1116</sub> )	No	Structure	log(1/ IC <sub>HEPG2</sub> )	log(1/ IC <sub>SW1116</sub> )
1		4.90	4.78	12		4.72	4.73
2		4.91	4.82	13		4.73	4.74
3		5.37	5.61	14		4.93	4.71
4		5.12	5.10	15		4.71	5.18
5		5.24	5.30	16		4.97	5.30
6		4.78	4.65	17		5	5.39
7		5.02	5.17	18		4.95	4.47
8		4.97	5.12	19		4.95	5.10
9		5.15	4.92	20		4.77	4.81
10		5.75	5.28	21		6.11	5.83
11		4.99	4.77				

IP: 154.121.4.150 On: Wed, 07 Mar 2018 15:34:52  
 Copyright: American Scientific Publishers  
 Delivered by Ingenta

**Table II.** Values of molecular descriptors used in the regression analysis.

No.	Vol ( $\text{\AA}^3$ )	HE (kcal/mol)	logP	MR ( $\text{\AA}^3$ )	SAG ( $\text{\AA}^2$ )	MW (uma)	POL ( $\text{\AA}^3$ )	$E_{\text{HOMO}}$ (au)	$E_{\text{LUMO}}$ (au)	$\mu$ (debye)	qN1	qN4	qC2	qC3	qC5	qC6
1	770.170	-12.540	1.940	79.140	466.470	288.300	28.820	-0.2392	-0.0794	0.886	-0.427	-0.450	0.026	0.020	0.046	0.107
2	791.620	-12.630	2.320	83.730	474.610	304.750	30.840	-0.2494	-0.0807	5.144	-0.424	-0.453	0.025	0.021	0.050	0.108
3	810.260	-12.580	2.600	86.540	485.200	349.200	31.540	-0.2397	-0.0796	0.887	-0.427	-0.450	0.026	0.020	0.046	0.107
4	809.750	-13.290	2.320	83.730	498.290	304.750	30.840	-0.2427	-0.0814	1.269	-0.426	-0.450	0.026	0.021	0.046	0.107
5	828.550	-13.240	2.600	86.540	505.960	349.200	31.540	-0.2468	-0.0817	1.498	-0.425	-0.450	0.026	0.021	0.046	0.107
6	777.830	-13.380	1.940	79.140	479.960	288.300	28.820	-0.2388	-0.0797	2.243	-0.445	-0.451	0.022	0.018	0.064	0.108
7	800.840	-11.390	2.290	85.580	486.180	304.360	31.180	-0.2339	-0.0842	5.024	-0.467	-0.449	0.024	0.014	0.066	0.137
8	822.050	-12.250	2.670	90.170	498.700	320.810	33.200	-0.2351	-0.0856	5.023	-0.467	-0.449	0.024	0.015	0.066	0.137
9	830.070	-12.540	2.670	90.170	511.100	320.810	33.200	-0.2378	-0.0860	1.986	-0.465	-0.449	0.026	0.016	0.065	0.136
10	848.390	-12.500	2.940	92.990	520.260	365.270	33.900	-0.2378	-0.0860	1.924	-0.465	-0.449	0.026	0.016	0.065	0.136
11	797.400	-12.610	2.290	85.580	496.490	304.360	31.180	-0.2349	-0.0846	2.137	-0.465	-0.449	0.025	0.015	0.065	0.136
12	961.110	-10.860	2.480	110.970	543.240	363.410	39.200	-0.2212	-0.0652	3.636	-0.470	-0.453	0.021	0.007	0.063	0.137
13	984.280	-10.690	2.860	115.560	550.540	379.870	41.210	-0.2200	-0.0634	4.963	-0.471	-0.453	0.021	0.006	0.063	0.137
14	997.060	-10.630	3.130	118.370	554.530	424.320	41.910	-0.2199	-0.0634	4.949	-0.471	-0.453	0.021	0.006	0.063	0.137
15	980.740	-11.450	2.860	115.560	562.490	379.870	41.210	-0.2244	-0.0671	4.190	-0.470	-0.453	0.022	0.008	0.063	0.136
16	999.400	-11.410	3.130	118.370	570.870	424.320	41.910	-0.2230	-0.0653	4.523	-0.471	-0.453	0.021	0.007	0.063	0.137
17	948.380	-11.540	2.480	110.970	543.200	363.410	39.200	-0.2233	-0.0640	4.275	-0.431	-0.454	0.022	0.013	0.040	0.135
18	769.230	-13.670	2.550	79.010	475.710	270.310	28.910	-0.2397	-0.0810	4.278	-0.424	-0.451	0.027	0.019	0.043	0.109
19	789.010	-12.890	2.890	85.450	490.320	286.370	31.270	-0.2327	-0.0829	1.449	-0.465	-0.449	0.025	0.015	0.065	0.137
20	939.490	-11.790	3.080	110.840	539.340	345.420	39.290	-0.2198	-0.0646	4.399	-0.470	-0.453	0.021	0.007	0.063	0.137
21	854.340	-11.490	2.940	92.990	526.340	365.270	33.900	-0.2363	-0.0857	2.507	-0.465	-0.449	0.026	0.016	0.065	0.136

of the studied pyrazine derivatives (Table II). This result is in agreement with the formula of Lorentz-Lorenz,<sup>68</sup> which gives a relationship between polarizability, molar refractivity and the molecular size.<sup>69</sup>

From the results obtained in (Table II), we observed that polarizability data and molecular refractivity are generally proportional to the size and the molecular weight of pyrazine derivatives. This explains the congruity of our results with Lorentz-Lorenz expression, for instance compound number 14 and compound 16 show the same maximum values of polarizability (41.91 (Å<sup>3</sup>)) and refractivity (118.37 (Å<sup>3</sup>)). These compounds have also high values of Molecular weight (424.32 uma), and a slight difference in surfaces and volumes. As for Hydration energy (in absolute value), the most important is that of the compound 18 (13.67 kcal · mol<sup>-1</sup>) and the smallest value is that of the compound 14 (10.63 kcal · mol<sup>-1</sup>). Indeed, in the biological environments water molecules surround the polar molecules. Thus they establish hydrogen bonds between them.

We noticed that compounds 18 possess six (HBA) hydrogen bond acceptors and no (HBD) hydrogen bond donors, the presence of hydrophilic groups in this compound results in an increase of the hydration energy. This property explains the ability of these compounds, not only fixing to the receptor, but also activating it. Hydration energy measures the degree of agonist character of a potential drug molecule.

All studied molecules have almost optimal (log *P*) values; for good oral bioavailability, the log *P* must be greater than zero and less than 3 (0 < log *P* < 3). If log *P* is too high (>3), the drug has low solubility. Whereas for too low log *P* (<0), the drug has difficulty penetrating the lipid membranes.<sup>70,71</sup> Thus the compound 18 has important hydration energy and optimal log *P* value.

Compound 2 shows the maximum dipole moment value (5.14 D). It would originate from a resonance effect, involving a donor effect from nucleus toward the electro-attractive. The least *E*<sub>HOMO</sub> – *E*<sub>LUMO</sub> energy gap is (0.149 a.u), which depicts the chemical reactivity of the compound. The higher the HOMO-LUMO energy gap, the lesser is the flow of electrons to the higher energy state making the molecule hard and less reactive.

On the other hand, in lesser *E*<sub>HOMO</sub> – *E*<sub>LUMO</sub> gap, there is easy flow of electrons to the higher energy state making it softer and more reactive. Hard bases have highest occupied molecular orbitals (HOMO) of low energy and hard acids have the lowest unoccupied molecular orbitals (LUMO) of high energy.<sup>72</sup>

Compounds 8 present the least *E*<sub>LUMO</sub> – *E*<sub>HOMO</sub> energy gap (0.149 a.u). The *E*<sub>LUMO</sub> – *E*<sub>HOMO</sub> energy gap is an important stability index that should be considered. As the *E*<sub>LUMO</sub> – *E*<sub>HOMO</sub> energy gap decreases, the interactions between the reacting species become stronger such as ligand-receptor interaction.<sup>73</sup>

Charges based descriptors have been widely employed as chemical reactivity indices or as measures of weak intermolecular interactions. Many quantum-chemical descriptors are derived from the partial charge distribution in a molecule or from the electron densities on particular atoms.<sup>74</sup>

The nitrogen atoms have negative charges, which lead to electrophilic attack, whereas the carbon atoms have positive charges, which lead to preferential site to nucleophilic attack.

### 3.2. Quantitative Structure-Activity Relationships Studies (QSAR) of Pyrazine Derivatives

In the second step, the several physical and chemical properties<sup>75</sup> known as physicochemical descriptors were used as independent variables and were correlated with biological activities of pyrazine derivatives for the generation of QSAR models by multiple linear regressions (MLR) analyzes. Developing a QSAR model requires a diverse set of data, and thereby, a large number of descriptors have to be considered. Descriptors are numerical values that encode different structural features of the molecules. The selection of an appropriate set of descriptors from a large number of them requires a method, which is able to discriminate between the parameters. Pearson's correlation matrix has been performed on all descriptors by using SPSS statistics 19 Software. The analysis of the matrix revealed five descriptors for the development of MLR models. The values of descriptors used in MLR analysis are presented in Table II.

The correlation between the biological activities and descriptors expressed by the following relations:

$$\begin{aligned} \text{Log}(1/\text{IC}_{50_{\text{HEPG2}}}) &= 2.784 + 0.416\text{HE} + 0.692\log P \\ &+ 0.021\text{SAG} - 0.181\text{Pol} + 73.301\text{qC3} \end{aligned} \quad (1)$$

$$n = 21; r = 0.943; S = 0.131; F = 24.090; Q = 7.198$$

$$\begin{aligned} \text{Log}(1/\text{IC}_{50_{\text{SW1116}}}) &= 18.243 - 0.041\text{Vol} + 0.700\text{HE} + 0.583\log P \\ &+ 0.140\text{MR} + 0.045\text{SAG} - 289.948\text{qC2} - 49.834\text{qC5} \\ &- 30.846\text{qC6} - 0.104\mu - 86.539E_{\text{LUMO}} \end{aligned} \quad (2)$$

$$n = 21; r = 0.955; S = 0.143; F = 10.427; Q = 6.678$$

The values of fraction variance may lie between 0 and 1. QSAR model having *r*<sup>2</sup> > 0.6 will only be considered

**Table III.** Cross-validation parameters.

Model	PRESS	SSY	PRESS/SSY	<i>S</i> <sub>PRESS</sub>	<i>r</i> <sub>cv</sub> <sup>2</sup>	<i>r</i> <sub>adj</sub> <sup>2</sup>	6PE
1	0.259	2.341	0.111	0.111	0.889	0.852	0.098
2	0.205	2.342	0.087	0.099	0.912	0.825	0.077

**Table IV.** Experimental and predicted activities IC 50 of the molecules under study.

No.	Log(1/IC <sub>HEPG2</sub> )			Log(1/IC <sub>SW1116</sub> )		
	Obs.	Pred.	Res.	Obs.	Pred.	Res.
1	4.900	4.878	0.022	4.780	4.848	-0.068
2	4.910	4.981	-0.071	4.820	4.869	-0.049
3	5.370	5.216	0.154	5.610	5.464	0.146
4	5.120	5.199	-0.079	5.100	5.131	-0.031
5	5.240	5.447	-0.207	5.300	5.309	-0.009
6	4.780	4.663	0.117	4.650	4.670	-0.020
7	5.020	5.142	-0.122	5.170	5.036	0.134
8	4.970	5.016	-0.046	5.120	5.116	0.004
9	5.150	5.227	-0.077	4.920	4.998	-0.078
10	5.750	5.494	0.256	5.280	5.250	0.030
11	4.990	4.923	0.067	4.770	4.913	-0.143
12	4.720	4.717	0.003	4.730	4.630	0.100
13	4.730	4.766	-0.036	4.740	4.703	0.037
14	4.930	4.934	-0.004	4.710	4.958	-0.248
15	4.710	4.845	-0.135	5.180	4.995	0.185
16	4.970	5.023	-0.053	5.300	5.255	0.045
17	5.000	4.873	0.127	5.390	5.419	-0.029
18	4.950	4.933	0.017	4.470	4.433	0.037
19	4.950	5.077	-0.127	5.100	5.005	0.095
20	4.770	4.648	0.122	4.810	4.880	-0.070
21	6.110	6.041	0.069	5.830	5.896	-0.066

for validation. For example, the value  $r = 0.955$  and  $r^2 = 0.912$  (model 2) allowed us to indicate firmly the correlation between different parameters (independent variables) with anti-proliferative activity against HEPG2; human colorectal carcinoma cell.

The  $F$  value has found to be statistically significant at 95% level, since all the calculated  $F$  value is higher as compared to tabulated values. The positive value of quality factor ( $Q$ ) for these two QSAR's models 1 and 2 suggests its high predictive power.

The positive coefficient of HE indicates that any increase in hydration energy causes an increase in biological activity, it may be concluded that hydrophilic molecules are more important for anti-proliferative activity against the HEPG2 and SW1116.

In models 1 and 2, it is observed that any increase in the molecular surface causes an increase in activity, which results in an increase in the surface area of contact between the ligand and the receptor.

It can be observed that low positive charge for C3 (qC3) in model 01. But in model 02, high coefficients of atomic charges on C2, C5 and C6 atoms (qC2, qC5 and qC6 respectively), thus, high positive charges for C2 and low negative for C5 and C6 respectively lead to increasing anti-proliferative activity.

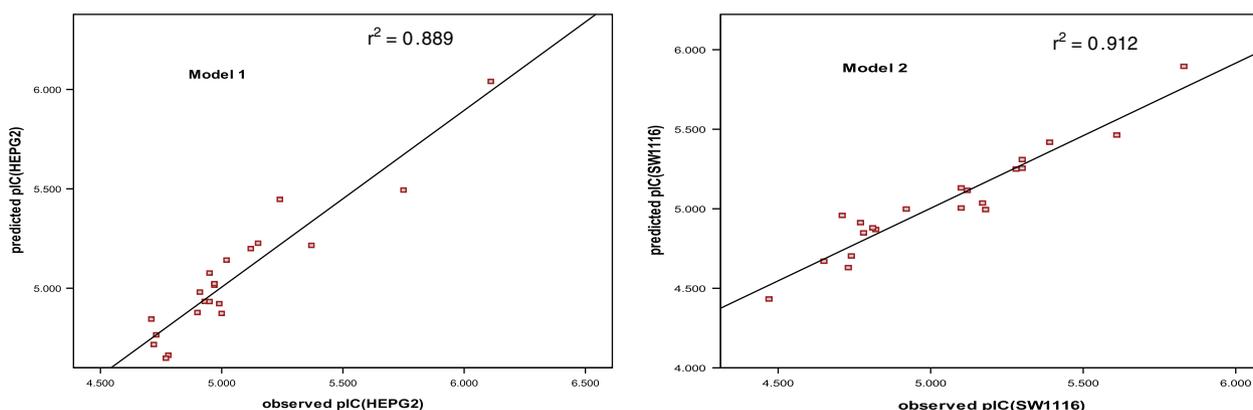
The charges allowed a physical explanation and electronic molecular properties contributing to antiproliferative potency as the electronic character related directly to the electron distribution of interacting molecule at the active site.

In order to test the validity of the predictive power of the selected MLR models (Eqs. (1) and (2)), the leave-one out technique (LOO technique) was used. The developed models were validated by calculating the following statistical parameters: predicted residual sum of squares (PRESS), total sum of squares deviation (SSY), the predictive error of the coefficient of correlation (PE) and cross-validated correlation coefficients ( $r_{adj}^2$  and  $r_{cv}^2$ ) (Table III).

PRESS is an important cross-validation parameter as it is a good approximation of the real predictive error of the models. Its value being less than SSY points out that model predicts better than chance and can be considered statistically significant. The smaller PRESS value of the better model predictability is from the results shown in Table III. Models 1 and 2 are statistically significant.

Further, for reasonable QSAR model, the PRESS/SSY ratio should be lower than 0.4.<sup>76</sup> The data presented in Table III indicate that for the developed models this ratio is 0.041 for the first model and 0.102 for the second one. The result  $r_{cv}^2$  for these QSAR models is 0.889 for the first model and 0.912 for the second one. The high value of  $r_{cv}^2$  and  $r_{adj}^2$  are essential criteria for the best qualification of the QSAR models 1 and 2.

The predictive error for the coefficient of correlation (PE) is yet another parameter used to evaluate the

**Fig. 1.** Plots of predicted versus experimentally observed anti-proliferative activity against the HEPG2 and SW1116.

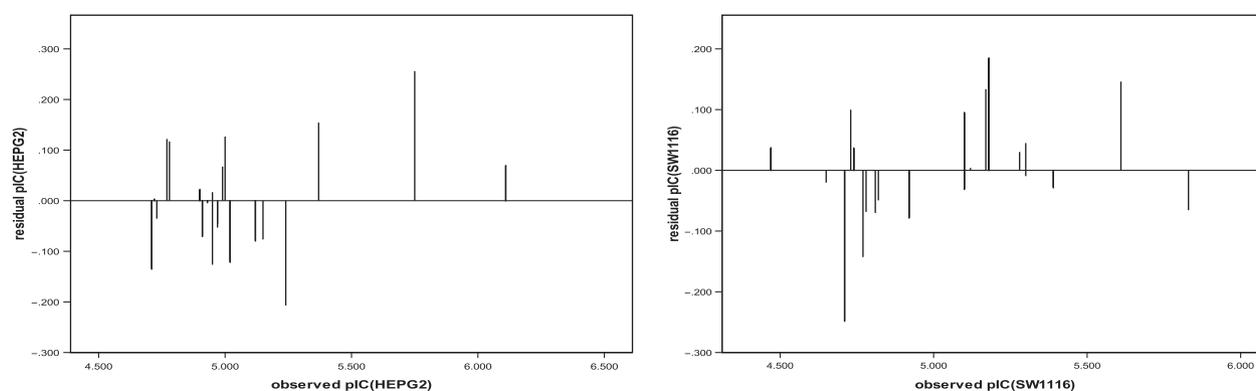


Fig. 2. Plots of residual against experimental values  $\log(1/IC_{\text{HEPG2}})$  and  $\log(1/IC_{\text{SW1116}})$ .

predictive power of the proposed models.<sup>77</sup> We have calculated the PE value of the proposed models and they are reported in Table III. For the two models developed the condition  $r > 6$  PE is satisfied and hence they can be said to have a good predictive power.

However, the only way to estimate the true predictive power of the developed model is by calculating the predicted values of  $\log(1/IC_{\text{HEPG2}})$  and  $\log(1/IC_{\text{SW1116}})$  for the investigated pyrazines using model 1 and 2 respectively (Table IV).

Figure 1 shows the plots of the predicted linear regression versus the experimental values of the biological activity of pyrazines outlined above. The plots for models 1 and 2 show a good deal of correspondence with experimentally reported data with  $r^2 = 0.889$  and  $0.912$ , respectively. The present QSAR study shows that models 1 and 2 can be successfully applied to predict anti-proliferative activities against the HEPG2 and SW1116 in these molecules generations. To investigate the presence of a systematic error in developing the QSAR models, the residuals of the predicted values for the biological activity  $\log(1/IC_{\text{HEPG2}})$  and  $\log(1/IC_{\text{SW1116}})$  were plotted against the experimental values, as it's shown in Figure 2.

The propagation of the residuals on both sides of zero indicates that no systemic error exists, as suggested by Jalali-Heravi and Kyani.<sup>78</sup> It indicates that these models can be successfully applied to predict the anti-proliferative activity against the HEPG2 and SW1116 of this class of molecules.

#### 4. CONCLUSION

The present study provides a discussion of several qualitative approximations of the structure activity relationship for pyrazine derivatives. From the results discussed above, accurate mathematical models were developed for predicting the anti-proliferative activity against HEPG2 and SW1116. The validity of the models has been established by the determination of suitable statistical parameters. Henceforth the developed QSAR models were used to predict the anti-proliferative activity of the investigated

pyrazines, and close agreement between experimental and predicted values has been achieved. The low residual activity and high cross-validated values obtained suggests a good predictive ability of the developed QSAR models. It indicates that the studied activities ( $IC_{\text{HEPG2}}$  and  $IC_{\text{SW1116}}$ ) of pyrazine derivatives series can be successfully modeled using various molecular descriptors. The developed QSAR models 1 and 2 show that hydrophilic derivatives of pyrazines give a good anti-proliferative activity against HEPG2 and SW1116.

#### References and Notes

- H. Languer, K. Kassali, and N. Lebga, *J. Comput. Theor. Nanosci.* 10, 86 (2013).
- N. Melkemi and S. Belaidi, *J. Comput. Theor. Nanosci.* 11, 801 (2014).
- M. Mellaoui, S. Belaidi, D. Bouzidi, and N. Gherraf, *Quantum Matter* 3, 435 (2014).
- S. Belaidi, H. Belaidi, and D. Bouzidi, *J. Comput. Theor. Nanosci.* 12, 1737 (2015).
- K. Dermeche, N. Tchouar, S. Belaidi, and T. Salah, *J. Bionanosci.* 9, 395 (2015).
- S. Belaidi, R. Mazri, H. Belaidi, T. Lanez, and D. Bouzidi, *Asian J. Chem.* 25, 9241 (2013).
- S. Medjahed, S. Belaidi, S. Djekhaba, N. Tchouar, and A. Kerassa, *J. Bionanosci.* 10, 118 (2016).
- S. Belaidi, H. Belaidi, A. Kerassa, M. Saoula, and D. Bouzidi, *Quantum Matter* 6, 798 (2016).
- E. X. Esposito, A. J. Hopfinger, and J. D. Madura, *Methods Mol. Biol.* 275, 131 (2004).
- G. Poovi, *J. Bionanosci.* 11, 87 (2017).
- S. Roy, *J. Bionanosci.* 11, 131 (2017).
- S. P. Bradbury, *Toxicol. Lett.* 79, 229 (1995).
- C. Hansch and A. Leo, *Exploring QSAR, Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, Washington, DC, USA (1995).
- C. Hansch, B. R. Telzer, and L. Zhang, *Toxicol.* 25, 67 (1995).
- R. Perkins, H. Fang, W. Tong, and W. Welsh, *Environ. Toxicol. Chem.* 22, 1666 (2003).
- J. Z. Chen, X. W. Han, Q. Liu, A. Makriyannis, J. Wang, and X. Q. Xie, *J. Med. Chem.* 49, 625 (2006).
- L. Salum and A. Andricopulo, *Mol. Divers.* 13, 277 (2009).
- J. Campos, C. Jiménez, C. Trigo, P. Ibarra, D. Rana, R. Thiruganesh, M. Ramalingam, and Z. S. Haida, *J. Bionanosci.* 9, 249 (2015).

19. C. Romagnoli, R. Zonefrati, D. Puppi, C. Rosati, A. Aldinucci, G. Palmi, G. Galli, F. Chiellini, F. S. Martelli, A. Tanini, and M. L. Brandi, *J. Biomater. Tissue Eng.* 7, 622 (2017).
20. F. Straka, D. Schornik, J. Masin, E. Filova, T. Mirejovsky, Z. Burdikova, Z. Svindrych, H. Chlup, L. Horny, J. Vesely, J. Pirk, and L. Bacakova, *J. Biomater. Tissue Eng.* 7, 527 (2017).
21. M. Ibrahim, N. A. Saleh, W. M. Elshemey, and A. A. Elsayed, *J. Comput. Theor. Nanosci.* 7, 224 (2010).
22. Y. Belmiloud, M. Ouraghi, M. Brahimi, A. Benaboura, D. Charqaoui, and B. Tangour, *J. Comput. Theor. Nanosci.* 9, 1101 (2012).
23. M. Ibrahim, N. A. Saleh, J. H. Ali, W. M. Elshemey, and A. A. Elsayed, *Spectrochim. Acta, Part A* 75, 702 (2010).
24. M. Ibrahim, *J. Comput. Theor. Nanosci.* 6, 682 (2009).
25. A. Tamulis and M. Grigalavicius, *Quantum Matter* 3, 460 (2014).
26. O. A. Ojo, B. E. Oyinloye, A. B. Ojo, O. B. Afolabi, O. A. Peters, O. Olaiya, A. Fadaka, J. Jonathan, and O. Osunlana, *J. Bionanosci.* 11, 292 (2017).
27. H. I. Salaheldin, *J. Bionanosci.* 11, 105 (2017).
28. A. S. Noha, A. E. Abdo, A. E. Ahmed, M. E. Wael, and I. Medhat, *J. Comput. Theor. Nanosci.* 2, 11 (2014).
29. H. Zheng, J. Jie, and Y. Zheng, *J. Bionanosci.* 7, 643 (2013).
30. D. Painuly, A. Bhatt, and V. K. Krishnan, *J. Biomater. Tissue Eng.* 4, 823 (2014).
31. D. Li, L. Yang, J. Shi, E. Hou, and C. Zuo, *J. Bionanosci.* 7, 465 (2013).
32. Y. Ge and L. Liu, *J. Bionanosci.* 7, 606 (2013).
33. A. Ayala, J. D. Alexander, and A. U. Kargol, *J. Bionanosci.* 8, 309 (2014).
34. N. Kaur, K. Yadav, R. Garg, K. Saroha, and D. Yadav, *J. Bionanosci.* 10, 191 (2016).
35. G. Sekar, A. Sivakumar, A. Mukherjee, and N. Chandrasekaran, *J. Bionanosci.* 10, 94 (2016).
36. B. Özden, S. Yilmaz, B. Baş, F. O. Özden, and H. Duran, *J. Biomater. Tissue Eng.* 7, 413 (2017).
37. H. Fengyan, K. An, X. Wen, P. Zhao, D. Shi, C. Lin, and H. Liu, *J. Biomater. Tissue Eng.* 7, 35 (2017).
38. N. Jha, J. S. Choi, J. H. Kim, R. Jung, E. H. Choi, J. J. Ryu, and I. Han, *J. Biomater. Tissue Eng.* 7, 662 (2017).
39. K. A. Almoammar, E. Alkofide, A. Alkhatlan, Y. Alateeq, A. Alqahtani, and M. M. AlShaaqim, *J. Biomater. Tissue Eng.* 7, 671 (2017).
40. B. Wei, Z. Shi, J. Xiao, Y. Xu, and L. Lv, *J. Biomater. Tissue Eng.* 7, 418 (2017).
41. H. Xu, G. Fang, J. Gou, S. Wang, Q. Yao, Y. Zhang, X. Tang, and Y. Zhao, *J. Biomater. Tissue Eng.* 5, 919 (2015).
42. X. Li, Q. Rong, and S. L. Chen, *J. Biomater. Tissue Eng.* 5, 445 (2015).
43. H. G. Guo, H. S. Guo, F. L. Yao, S. G. Yang, Z. Chen, and T. Wang, *J. Biomater. Tissue Eng.* 6, 114 (2016).
44. Y. Liu and I. M. Hsing, *J. Biomater. Tissue Eng.* 5, 687 (2015).
45. K. Kim, D. Y. Kim, J. H. Baek, J. H. Kim, Y. H. Park, Y. J. Kim, B. H. Min, and M. S. Kim, *J. Biomater. Tissue Eng.* 5, 372 (2015).
46. R. Jeyapragas, G. Poovi, M. S. Rao, R. Gopal, and M. Sivabalan, *J. Bionanosci.* 9, 465 (2015).
47. A. Zakharov, A. Lagunin, D. Filimonov, and V. Poroikov, *J. Chem. Res. Toxicol.* 25, 2378 (2012).
48. Y. Chi-Yung, Z. Xiao, and Ch. GuanHua, *J. Comput. Theor. Nanosci.* 3, 5 (2006).
49. S. Khamouli, S. Belaidi, Z. Almi, S. Medjahed, and H. Belaidi, *J. Bionanosci.* 11, 301 (2017).
50. A. K. Singh, P. Kumar, M. Yadav, and D. S. Pandey, *J. Organomet. Chem.* 695, 567 (2010).
51. X. H. Zhao, S. S. Liu, Y. Z. Li, and M. D. Chen, *Spectrochim. Acta, Part A* 75, 794 (2010).
52. T. Asaki, T. Hamamoto, Y. Sugiyama, K. Kuwano, and K. Kuwabara, *Bioorg. Med. Chem.* 15, 6692 (2007).
53. J. W. Corbett, M. R. Rauckhorst, F. Qian, R. L. Hoffman, C. S. Knauer, and L. W. Fitzgerald, *Bioorg. Med. Chem. Lett.* 17, 6250 (2007).
54. C. P. Meher, A. M. Rao, and M. D. Omar, *Asian J. Pharm. Sci. Res.* 3, 52 (2013).
55. A. Herman, *Rev. Theor. Sci.* 1, 3 (2013).
56. E. L. Pankratov and E. A. Bulaeva, *Rev. Theor. Sci.* 1, 58 (2013).
57. Y. B. Zhang, X. L. Wang, W. Liu, Y. S. Yang, J. F. Tang, and H. L. Zhu, *Bioorg. Med. Chem.* 20, 6356 (2012).
58. HyperChem (Molecular Modeling System) Hypercube, Inc., 1115 NW, 4th Street, Gainesville, FL 32601, USA (2008).
59. Gaussian 09, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian Inc., Wallingford, CT (2010).
60. V. N. Viswanadhan, A. K. Ghose, G. R. Revankar, and R. K. Robins, *J. Chem. Inf. Comp. Sci.* 29, 163 (1989).
61. A. K. Ghose and G. M. Crippen, *J. Chem. Inf. Comp. Sci.* 27, 21 (1987).
62. N. Bodor, Z. Gabanyi, and C. K. Wong, *J. Am. Chem. Soc.* 111, 3783 (1989).
63. A. Gavezzotti, *J. Am. Chem. Soc.* 105, 5220 (1983).
64. K. J. Miller, *J. Am. Chem. Soc.* 112, 8533 (1990).
65. SPSS software packages, SPSS Inc., 444 North Michigan Avenue, Suite 3000, Chicago, Illinois, 60611, USA.
66. A. Kerassa, S. Belaidi, and T. Lanez, *Quantum Matter* 4, 45 (2015).
67. J. Wang, X. Q. Xie, T. Hou, and X. Xu, *Fast. J. Phys. Chem. A* 24, 4443 (2007).
68. A. Kerassa, S. Belaidi, D. Harkati, T. Lanez, O. Prasad, and L. Sinha, *Rev. Theor. Sci.* 3, 85 (2015).
69. Z. Almi, S. Belaidi, and L. Segueni, *Rev. Theor. Sci.* 3, 264 (2015).
70. N. I. Zhokhova, I. I. Baskin, V. A. Palyulin, A. N. Zefirov, and N. S. Zefirov, *Russ. Chem. B* 52, 1061 (2003).
71. S. Belaidi, A. Kerassa, T. Lanez, and M. Cinar, *J. Comput. Theor. Nanosci.* 12, 2127 (2015).
72. E. H. Kerns and L. Di, *Drug-Like Properties*, Academic Press, Cambridge, USA (2008), p. 43.
73. S. Belaidi, A. Landreau, S. Djebbar, O. BenaliBaitich, G. Bouet, and J. P. Bouchara, *J. Inorg. Biochem.* 55, 102 (2008).
74. M. Karelson, *Molecular Descriptors in QSAR/QSPR*, John Wiley & Sons, New York (2000).
75. S. Nithyanantham and L. Palaniappan, *J. Comput. Theor. Nanosci.* 11, 1721 (2014).
76. S. O. P. Kuzmanovic, D. D. Cvetkovic, and D. J. Barna, *Int. J. Mol. Sci.* 10, 1670 (2009).
77. A. K. Srivastava and N. Shukla, *Microsoft Academic Search* 15, 215 (2011).
78. M. J. Heravi and A. Kyani, *J. Chem. Inf. Comput. Sci.* 44, 1328 (2004).

Received: 22 June 2017. Accepted: 4 September 2017.

## ملخص

العمل الحالي يحتوي على بحث أساسي وأصلي على البيرازينات بهدف التنبؤ بالفاعلية الكيميائية والفاعلية البيولوجية ووضع نموذج فارماكوفوري لجزئيات جديدة نشطة بيولوجيا. أساليب النمذجة الجزيئية المستخدمة في عملنا: الطرق الكمية، طرق النمذجة الجزيئية ' وقد استخدمت هذه الأساليب لتحديد العوامل الهيكلية والإلكترونية والعوامل الطاقية المرتبطة بالجزئيات المدروسة. هذه المعايير المحسوبة هي متوافقة مع القيم التجريبية، وطبيعة نوع مستبدل (المانح، المستقبل) الذي يؤثر على العوامل الإلكترونية والطاقية للنواة الأساسية البيرازين. الدراسة النوعية للعلاقة هيكل-خاصيات/فاعلية قدمت أيضا لسلسلة من مشتقات البيرازينات النشطة بيولوجيا. طبيعة المجموعات على الهيكل الحلقي للجزئيات المدروسة تؤثر على خصائصها الفيزيائية، وبالتالي على خصائصها الصيدلانية. وأخيرا قمنا بدراسة الكمية بنية فاعلية لواحد وعشرين مركب من مشتقات البيرازين (عوامل مضادة لتكاثر الخلايا السرطانية) ولقد لاحظنا شدة التقارب بين القيم النظرية والتجريبية للفاعلية البيولوجية مما يؤكد فاعلية وجودة النموذج QSAR المحصل عليها.

## Résumé

Le présent travail comporte une recherche fondamentale et originale sur les pyrazines, dans le but est de prédire la réactivité chimique et l'activité biologique et d'établir des modèles pharmacophore pour des nouvelles molécules bioactives. Nous avons utilisé différentes méthodes de modélisation moléculaire telles que les méthodes quantiques, les méthodes empiriques. Le choix de ces méthodes est pour déterminer les différents paramètres structuraux, électroniques et énergétiques associés aux molécules étudiées. Ces paramètres optimisés ont été trouvé en bon accord avec les valeurs expérimentales. La nature de type de substituant (donneur, accepteur) influe sur les paramètres électroniques et énergétiques du noyau de base des pyrazines. Une étude qualitative de la relation structure–propriétés/activité a été effectuée également pour une série bioactive de dérivés de la pyrazine. La nature des groupements sur le noyau hétérocyclique des molécules étudiées affecte leurs propriétés physicochimiques et par conséquence sur leurs propriétés pharmacologiques. Une étude QSAR a été effectuée sur vingt un molécules analogues de pyrazine. Une forte corrélation a été observée entre les valeurs expérimentales et les valeurs prédites de l'activité antiproliférative, ce qui indique la validité et la qualité des modèles QSAR obtenus.

## Abstract

This work involves a fundamental and original research on pyrazines, the aim is to predict the chemical reactivity and biological activity and to establish a pharmacophore models for new bioactive molecules. The molecular modeling methods used in our work are: quantum methods, empirical methods. These methods were used to determine the structural parameters, electronics and energy associated with molecules studied. These optimized parameters are in good agreement with experimental values. The nature of such substituent (donor, acceptor) affects the electronic and energy parameters of basic structure of pyrazines. A study of the structure-properties/activity has been carried out for a series of bioactive derivatives of pyrazine. The nature of the groups on heterocyclic ring of the studied molecules affects on their physico-chemical properties and by consequence on their pharmacological properties. QSAR studies have been performed on twenty one molecules of pyrazine as anti-proliferative agents. High correlation between experimental and predicted activity values was observed, indicating the validation and the good quality of the derived QSAR models.