



République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Med Khider Biskra

Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie

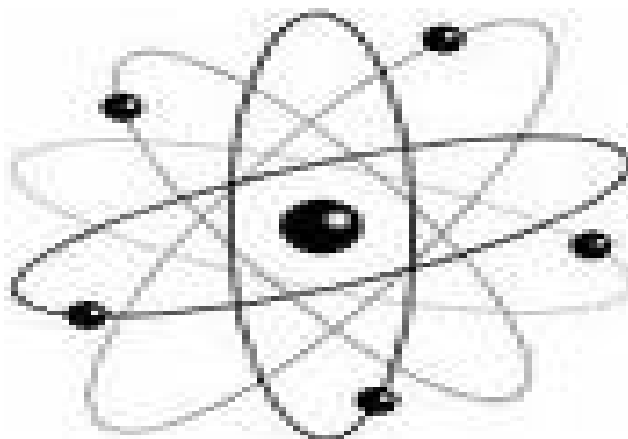


Département des Sciences de la Matière

Domaine : Sciences de la Matière

Filière : Chimie

Spécialité : Chimie des matériaux



*Mémoire de fin d'étude en Master*

*Intitulé :*

**Etude QSAR basée sur la DFT conceptuelle des propriétés antiprolifératives  
de certains dérivés de 2-iminothiazolidin-4-one**

*Présenté par :*

**Refrafi Salah eddine**

*Devant le jury :*

**Mr MELKEMI Nadjib,**

**MCA, Université de Biskra**

**Président**

**Mr BOUMEDJANE Youcef,**

**MCA, Université de Biskra**

**Rapporteur**

**M<sup>me</sup> LEHRAKI Faiza,**

**MAA, Université de Biskra**

**Examineur**

## DEDICACE

*A mes parents, je les remercie avec affection pour leurs sacrifices, leurs encouragements et pour le soutien moral qu'ils m'ont toujours apporté ; sans eux je ne serai jamais arrivée jusque là.*

*A mes frères et mes sœurs*

*A toute ma famille*

*A mes amies*

*Je dédie ce modeste travail.*

# *Remerciement*

*Tout d'abord, je tiens à remercier Allah, le Tout Puissant et le Miséricordieux, de m'avoir donné la santé la volonté et la patience pour mener à terme ma formation de master.*

*Je tiens à remercier monsieur «Boumedjane Yousef», le rapporteur de ce mémoire, pour avoir encadré ce travail. Je tiens à le remercier pour son aide précieuse, son objectivité, sa disponibilité et ses précieux conseils qui ont fait progresser ce travail. Il m'est aussi d'un agréable devoir de vous adresser un grand merci pour la sympathie, la confiance et la liberté d'action dont j'ai bénéficié tout au long de ce mémoire.*

*Je remercie également les membres de jury Monsieur Melekmi Nadjib et Madame Lehraki Faiza pour l'honneur qu'ils m'ont fait en acceptant de juger notre travail.*

*J'adresse un très grand et sincère remerciement à tous les enseignants du département de chimie de l'université de Biskra...*

*Je remercie en fin toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce travail.*

*Merci à tous*

# SOMMAIRE

Liste de figures .....	iii
Liste de tableaux .....	iv
Liste des abréviations .....	viii
Introduction général .....	1

## *Chapitre I : les méthodes de la chimie quantique*

I.1. Généralités sur les méthodes de la chimie quantique.....	4
I.1.1. Equation de Schrödinger et historique de la chimie quantique.....	4
I.1.2. L'approximation Born-Oppenheimer.....	5
I.1.3. L'Approximation de hartree-fock.....	5
I.2. La relation entre la chimie et l'informatique.....	6
I.3. Théorie de la fonctionnelle de la densité(DFT).....	7
I.3.1. Fondement de la théorie DFT.....	7
I.3.1.1. 1 <sup>er</sup> théorème de Hohenberg et Kohn.....	7
I.3.1.2. 2 <sup>ème</sup> théorème de Hohenberg et Kohn.....	8
I.3.2. Méthode de Kohn et Sham.....	8
I.3.3. Les différentes approximations.....	9
I.3.3.1. Approximation de la densité locale LDA.....	9
I.3.3.2. Approximation de la densité de spin locale LSDA.....	10
I.3.3.3. Approximation du Gradient Généralisé (GGA).....	10
I.4. Fonctionnelle hybride B3LYP.....	11
I.5. Références Bibliographiques.....	12

## *Chapitre II : Approches théoriques des méthodes de calculs utilisées.*

II.1. Approches théoriques de la réactivité chimique.....	13
II.1.1. Modèles de la réactivité chimique.....	13
II.1.1.1. Modèles empiriques.....	13
II.1.1.1.1. Principe HSAB (Hard and Soft Acids and Bases).....	13
II.1.1.2. Modèles théoriques.....	13

II.1.1.2.1. Modèle théorique basé sur la théorie des orbitales frontières.....	13
II.1.1.2.2. Modèle théorique basé sur la théorie de la fonctionnelle de la densité.....	14
II.1.1.2.2.1. Descripteurs globaux.....	14
a- Le potentiel chimique ( $\mu$ ) et l'électronégativité ( $\chi$ ).....	14
b- Dureté et mollesse chimiques.....	14
c- Indice d'électrophilicité globale .....	15
II.1.1.2.2.2. Descripteurs locaux.....	15
a- Fonction de Fukui.....	16
b- Molleses locales.....	16
c- Electrophilie locale.....	17
d- Nucléophilie locale.....	17
II.2. Etude quantitative des relations structure-activité (QSAR).....	17
II.2.1. Introduction.....	17
II.2.2. Principe des méthodes QSAR.....	18
II.2.3. Méthodologie générale d'une étude QSPR/QSAR.....	19
II.2.4. Outils et techniques de QSAR.....	19
II.2.4.1. Paramètres biologiques.....	19
II.2.4.2. Descripteurs moléculaires.....	20
II.2.4.2.1. Les descripteurs physico-chimiques.....	20
II.2.4.2.2. Les descripteurs électroniques.....	20
II.2.4.2.3. Les descripteurs topologiques.....	20
II.2.5. Importance des descripteurs.....	21
II.2.6. Les techniques de QSAR.....	21
II.3. Statistiques et analyse des données.....	22
II.3.1. Introduction.....	22
II.3.2. Régression linéaire multiple (MLR).....	22
II.3.3. Estimation des paramètres statistiques du modèle.....	23
II.3.4. Confiance accordée aux résultats.....	24
II.3.5. Tests sur le modèle linéaire.....	24
II.3.6. Test de la signification globale de la régression (F-Fisher).....	24
II.3.7. Sélection des variables et choix du modèle.....	25

II.3.7.1. Critères de comparaison de modèle.....	25
II.3.7.1.1. Limitation du coefficient de détermination $R^2$ .....	25
II.3.7.1.2. Coefficient de détermination ajusté $R^2_{\text{ajusté}}$ .....	25
II.4.7.1.3. Critère de validation croisée : PRESS.....	26
II.3.8. Validation d'un modèle.....	26
II.4. Applications de l'étude QSAR.....	26
II.5. Références Bibliographiques.....	28

### *Chapitre III : Résultats et discussion*

III.1. Introduction.....	31
III.2. Définition de 2-iminothiazolidine-4-one.....	31
III.3. Cellules de cancer du poumon A549 et H460.....	32
III.4. Méthodologie de calcul.....	32
III.5. Effet de solvant sur la structure de noyau de base de 2-iminothiazolidin-4-one.....	32
III.5.1. Effet de solvant sur les paramètres géométriques.....	33
III.5.2. Effet de solvant sur les paramètres électroniques.....	35
III.6. Etude de réactivité chimique de 2-iminothiazolidin-4-one.....	36
III.6.1. Indices et descripteurs de réactivité.....	36
III.6.1.1. Indices globaux de réactivité.....	36
III.6.1.2. Indices locaux de réactivité.....	37
III.7. Etude quantitative des relations structure-activité (QSAR) des propriétés antiprolifératives D'une série des dérivés de 2-iminothiazolidin-4-one.....	41
III.7.1. Méthodologie.....	41
III.7.1.1. Données expérimentales.....	41
III.7.1.2. Sélection des descripteurs et méthodes de calcul.....	42
III.7.2. Développement des modèles QSAR.....	44
III.7.3. Validation des modèles QSAR.....	48
III.8. Références bibliographiques.....	52
Conclusion générale .....	54
Annexe.....	55

# Liste des figures

## Chapitre II:

**Figure II. 1:** Modèle d'étude quantitative des relations structure/activité (QSAR)..... 18

**Figure II.2 :** L'analyse de données multivariées..... 22

## Chapitre III:

**Figure III.1 :** Structure 3D de noyau de base de 2-iminothiazolidin-4-one..... 31

**Figure III.2 :** Structure optimisée de 2-iminothiazolidin-4-one dans la phase gazeuse (a)  
et dans la phase aqueuse (b)..... 34

**Figure III.3 :** Visualisation du 3D-MESP des fonctions de Fukui  $f^-$ ,  $f^+$  et  $f^0$  dans  
la phase gazeuse..... 39

**Figure III.4 :** Visualisation du 3D-MESP des fonctions de Fukui  $f^-$ ,  $f^+$  et  $f^0$  dans  
la phase aqueuse..... 40

**Figure III.5 :** Corrélation entre l'activité biologique expérimentale et prédite pour  
les modèles (1) et (2)..... 50

**Figure III.6 :** Corrélation entre les valeurs résiduelles et l'expérimentale des modèles (1) et (2).. 51

# Liste des tableaux

## Chapitre II:

**Tableau II.1** : Types de données biologiques utilisées dans l'analyse QSAR..... 19

## Chapitre III:

**Tableau III.1** : Classification des types et sous-types du cancer du poumon..... 32

**Tableau III.2** : Distances interatomique en (Å)..... 33

**Tableau III.3** : Angles de valence en (°)..... 33

**Tableau III.4** : Angles dièdres en (°)..... 34

**Tableau III.5** : Charges atomiques de 2-iminothiazolidin-4-one..... 35

**Tableau III.6** : Moment dipolaire..... 36

**Tableau III.7** : Descripteurs de réactivité utilisés..... 36

**Tableau III.8** : Valeurs des indices de Fukui de 2-iminothiazolidin-4-one..... 38

**Tableau III.9** : Structure et activité antiprolifératives expérimentale des dérivés  
de 2-iminothiazolidin-4-one..... 41

**Tableau III.10** : Valeurs des descripteurs électroniques et physico-chimiques utilisés dans l'étude  
QSAR.....43

**Tableau III.11** : Matrice de corrélation du modèle QSAR pour l'activité H460..... 46

**Tableau III.12** : Matrice de corrélation du modèle QSAR pour l'activité A549..... 47

**Tableau III.13** : Paramètres de validation croisée..... 48

**Tableau III.14** : Activité antiproliférative expérimentale, prédite et résiduelle des dérivés de  
2-iminothiazolidin-4-one..... 49



Dans l'industrie pharmaceutique, pour rechercher de nouveaux composés chimiques ayant un intérêt thérapeutique potentiel, la méthode classiquement utilisée consiste à synthétiser de nombreuses séries de molécules et à les soumettre à des tests pharmacologiques pour déceler une éventuelle activité biologique. Cette méthode de « screening » présente un défaut évident : son caractère aléatoire lui confère une faible rentabilité, une importante mobilisation humaine et un coût financier important. Pourtant, la quasi-totalité des médicaments actuellement commercialisés est encore issue de ce type d'approche [1].

C'est dans ce sens que l'industrie pharmaceutique s'oriente vers des nouvelles méthodes de recherche, qui consistent à prédire les propriétés et les activités des molécules avant même que celles-ci ne soient synthétisées, pour cette raison ils utilisent les ordinateurs qui sont devenus des outils indispensables en chimie pharmaceutique moderne.

Les diverses opérations qui sont effectuées dans le cadre de la modélisation moléculaire impliquent la mise en œuvre de programmes et d'algorithmes qui permettent de calculer les données relatives à la structure et aux propriétés de la molécule prise en considération [2].

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 ou l'activité biologique [3]. Elle 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, comme la mécanique moléculaire, dynamique moléculaire, mécanique quantique. Ainsi, elle permet de déterminer la représentation graphique de la géométrie ou de la conformation des atomes d'une molécule et d'évaluer les propriétés physico-chimiques de la molécule étudiée [4].

C'est très intéressant de noter que la réactivité chimique des molécules a été souvent interprétée à l'aide d'indices et descripteurs basés sur l'intuition et l'expérience accumulée des chimistes [5]. Récemment, de nouveaux concepts chimiques et indices de réactivité dérivant la théorie de la fonctionnelle de densité (DFT) ont émergé comme des outils puissants pour la prédiction des sites réactifs des systèmes moléculaires [6,7]. Certains de ces concepts sont bien connus par les chimistes comme le potentiel chimique électronique ( $\mu$ ) et l'électronégativité ( $\chi$ ).

De nouveaux concepts comme l'électrophilie ( $\omega$ ) et la dureté ( $\eta$ ) ont été dérivés à partir des équations fondamentales de la DFT.

Les relations entre les structures des molécules et leurs propriétés ou activités sont généralement établies à l'aide des méthodes de modélisation par apprentissage statistique.

Parmi les techniques de chemoinformatique nous pouvons citer les techniques de QSAR qui consiste à trouver une corrélation entre une activité biologique mesurée pour un panel de composés et certains descripteurs moléculaires.

La régression linéaire multiple (MLR) est une méthode statistique utilisée pour quantifier la relation entre les variables indépendantes obtenues par les calculs et des variables dépendantes données par l'expérience. Cette méthode repose sur un ensemble des paramètres statistiques qui définissent la combinaison linéaire entre les variables indépendantes (descripteurs moléculaires) et l'activité biologique des molécules étudiées [8].

Nous avons fixé comme des objectifs principaux, Etude d'effet de solvant (eau) sur les paramètres structuraux, électroniques et la réactivité chimique de 2-iminothiazolidin-4-one, En fin Etude quantitative des relations structure/activité (QSAR) des propriétés antiprolifératives des dérivés de 2-iminothiazolidin-4-one.

Notre travail comprend deux parties :

- ❖ La première partie servant de rappel bibliographique, comporte deux chapitres contenant respectivement :

**Chapitre I** : Les méthodes de la chimie quantique.

**Chapitre II** : Approches théoriques des méthodes de calculs utilisées.

- ❖ La deuxième partie comporte le **Chapitre III** : Résultats et discussion.

## Références bibliographiques

- [1] Johann le M. (1989). Implication des récepteurs sérotoninergiques 5-HT<sub>4</sub> et 5-HT<sub>6</sub> dans le Traitement de la maladie d'Alzheimer, thèse docteur en pharmacie, université de rennes 1, faculté de pharmacie.
- [2] Graham L. Patrick. (2003). Chimie pharmaceutique . 2e Edition ; De Boeck Paris ,319.
- [3] Mighell, A. D., Himes, V. L. & Rodgers. (1983). J. R. Space-group frequencies for organic compounds. Acta Cryst. A39, 737.
- [4] Chermette, H. Spectra 2000., (1993). 291, 255.
- [5] L. Salem, J. Wiley, (1982). Electrons in Chemical Reaction: First Principles, NewYork.
- [6] P. Geerlings, F. De Proft, W. Langenaeker, (2003). Chem. Rev. 103, 1793.
- [7] R. G. Pearson. (1999). J. Chem. Ed. 76, 267.
- [8] N. Goudarzi, M. Goodarzi, T. Chen. (2012). Med. Chem. Rev, 21, 437.

## I.1. Généralités sur les méthodes de la chimie quantique

### I.1.1. Equation de Schrödinger et historique de la chimie quantique

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, qui a permis d'expliquer de nombreuses propriétés physiques, telles que les propriétés chimiques des éléments et la formation des liaisons chimiques [1].

Comme il va l'être vu, les méthodes de chimie quantique permettent le calcul de la structure électronique de systèmes tels que les atomes, les molécules neutres, les espèces radicalaires, les ions, les clusters d'atomes, les surfaces de solides, etc. Des algorithmes de calculs très précis sont utilisés pour minimiser l'énergie totale en fonction des paramètres structuraux et pour prédire la structure la plus stable des composés étudiés. Les fonctions d'onde ainsi obtenues permettent de calculer des propriétés électriques et magnétiques, de même qu'elles conduisent à l'obtention d'indices de réactivité et d'autres caractéristiques encore. Ces méthodes permettent donc aussi bien l'interprétation de résultats expérimentaux, que la prédiction de propriétés pour les quelles aucune expérience n'a pu encore fournir d'informations.

La chimie quantique consiste en l'utilisation de méthodes basées sur la résolution de l'équation de Schrödinger indépendante du temps. En résolvant l'équation aux valeurs

propres et vecteurs propres 
$$\mathbf{H} \psi = \psi \mathbf{E} \quad (1)$$

Où :

- H est l'opérateur hamiltonien du système.
- $\Psi$  est la fonction d'onde du système.
- E est l'énergie totale du système [2,3].

Cette équation ne peut être rigoureusement résolue que pour les systèmes mono-électroniques.

La description des systèmes plus complexes nécessite la mise en œuvre d'un certain nombre d'approximations [4].

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

$$\hat{H} = -\frac{1}{2} \sum_{K=1}^{n_e} \frac{\hbar^2}{m_e} \cdot \nabla_K^2 - \sum_{K=1}^{n_e} \sum_{A=1}^N \frac{e^2}{4\pi\epsilon_0} \cdot \frac{Z_A}{r_{kA}} \sum_{K=1}^{n_e} \sum_{l>k}^{n_e} \frac{e^2}{4\pi\epsilon_0 r_{kl}} \frac{1}{2} \sum_{A=1}^N \frac{\hbar^2}{M_A} \nabla_{R_A}^2 + \sum_{A=1}^N \sum_{B>A}^N \frac{e^2}{4\pi\epsilon_0} \cdot \frac{Z_A Z_B}{R_{AB}} \quad (2)$$

### I.1.2. L'approximation Born-Oppenheimer

En 1927, Born et Oppenheimer ont proposé de simplifier la résolution de l'équation (1) en séparant la partie électronique de la partie nucléaire dans la fonction d'onde  $\Psi$ . Cette approximation est basée sur le fait que les électrons se déplacent beaucoup plus rapidement que les noyaux, ceci étant dû à la masse beaucoup plus faible des électrons (environ 1836 fois moindre de celle du proton). Par conséquent, les électrons réagissent quasi instantanément à une modification de la position des noyaux [5]. Dans le cadre de cette approximation l'hamiltonien  $H$  peut se réduire à la forme suivante :

$$H = -\frac{\hbar^2}{2m_e} \sum_{i=1}^{n_e} \Delta_i - \sum_{K=1}^N \sum_{i=1}^{n_e} \frac{Z_K e^2}{R_{Ki}} + \sum_{i>j}^{n_e} \frac{e^2}{r_{ij}} \quad (3)$$

La résolution exacte de l'équation (1) n'est possible que pour l'atome d'hydrogène et les systèmes hydrogénéoïdes. Pour les systèmes poly-électroniques, il est nécessaire de faire appel aux méthodes d'approximation (méthode des variations et méthode des perturbations) pour résoudre l'équation de Schrödinger d'une manière approchée.

Les propriétés moléculaires qui peuvent être calculées en se basant sur la résolution de l'équation de Schrödinger sont la géométrie moléculaire, et donc les stabilités relatives, les spectres de vibrations, les moments dipolaires et quadripolaires, les spectres électroniques et aussi des fonctions descriptives de la réactivité, telles que les charges atomiques et les fonctions de Fukui. Toutefois, la précision avec laquelle on peut espérer calculer ces quantités est très variable en fonction de la nature de ces propriétés. Cette équation ne peut en effet pas être résolue de manière exacte pour les systèmes moléculaires, et l'on doit donc effectuer un certain nombre d'approximations.

### I.1.3. L'Approximation de Hartree-Fock

L'approximation orbitale, introduite par Hartree en 1928 [6], consiste à écrire la fonction d'onde pour un système poly électronique sous forme de produit de spin-orbitales Mono-électroniques supposées normalisées

$$\Psi_e (1,2,3,\dots,n_e) = \varphi(1) \cdot \varphi(2) \cdot \varphi(3) \dots \varphi_{n_e}(n_e) \quad (4)$$

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

$$\phi_i(\mathbf{r}_i) = \phi_i(\mathbf{r}_i) \cdot \eta(\mathbf{s}_i) \quad (5)$$

Ce concept a été ensuite généralisé par Hartree et Fock [7] 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 [8]. 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.

$$\Psi_e(1,2,3,\dots,n_e) = \frac{1}{\sqrt{n_e!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \dots & \phi_n(1) \\ \phi_1(2) & \phi_2(2) & \dots & \phi_n(2) \\ \dots & \dots & \dots & \dots \\ \phi_1(n_e) & \phi_2(n_e) & \dots & \phi_{n_e}(n_e) \end{vmatrix} \quad (6)$$

$\frac{1}{\sqrt{n_e!}}$  : Est le facteur de normalisation.

## I.2. La relation entre la chimie et l'informatique

La chimie assistée par ordinateur (« Computational Chemistry » en anglais) est le domaine de la chimie qui fait intervenir l'ordinateur ; ses applications peuvent être de différente nature, telles que l'élucidation et l'analyse de structures chimiques, le traitement d'informations chimiques ou encore la chimie théorique [9]. Les domaines de la chimie théorique sont, de même, très nombreux : chimie quantique, mécanique moléculaire, dynamique moléculaire ou encore représentation moléculaire. La chimie a cette particularité, avec l'informatique, de réunir sous un même vocable à la fois une activité scientifique et une activité industrielle de grande ampleur, visible et clairement identifiée dans l'opinion publique. Ce n'est le cas ni des mathématiques, ni de la physique, de la biologie ou de l'archéologie égyptienne, pour ne prendre que quelques exemples. Même si certaines des branches de ces disciplines comme l'électronique ou le génie génétique ont aussi, mais de façon clairement moins marquée dans l'esprit du public, cette double identité scientifique et industrielle [10].

La chimie et l'informatique sont très largement rependues dans notre quotidien malgré que certains les considèrent trop complexe d'autres personnes trouvent que la chimie et l'informatique des moyens plus au moins facile à aborder.

On juge l'informatique comme envahissante à cause du progrès qu'elle a subi ces derniers temps qui a touché plusieurs domaines (comme la médecine et la pharmacie, l'environnement, la

biologie). La chimie a aussi connue un grand développement au cours des années pour céder la place à la chimie numérique.

### I.3. Théorie de la fonctionnelle de la densité (DFT)

#### I.3.1. Fondement de la théorie DFT

Historiquement, les premiers à avoir exprimé l'énergie en fonction de la densité furent Thomas (1927), Fermi (1927, 1928) et Dirac (1930) sur le modèle du gaz uniforme d'électrons non interagissant. Le but des méthodes DFT est de déterminer des fonctionnelles qui permettent de relier la densité électronique à l'énergie [11]. Cependant, la DFT a véritablement débuté avec les théorèmes fondamentaux de Hohenberg et Kohn en 1964 [12] qui établissent une relation fonctionnelle entre l'énergie de l'état fondamental et sa densité électronique.

##### I.3.1.1. 1<sup>er</sup> théorème de Hohenberg et Kohn

Ce théorème de Hohenberg et Kohn montre que la densité électronique est la seule fonction nécessaire pour obtenir toutes les propriétés électroniques d'un système. Si on considère un gaz d'électrons, le potentiel externe agissant sur ces particules détermine l'état fondamental de ce système et la densité de charge correspondante [12]. Ainsi, toutes les propriétés concernant cet état sont des fonctionnelles du potentiel externe. Comme cela a été démontré initialement par Hohenberg et Kohn, en raison de la correspondance biunivoque existant entre le potentiel externe  $v_{\text{ext}}$  et la densité électronique  $\rho(\vec{r})$ . L'expression de l'Hamiltonien électronique d'un système poly-électronique est :

$$\hat{H} = -\frac{1}{2} \sum_i^n \Delta_i + \sum_{i>j}^n \frac{1}{r_{ij}} + v_{\text{ext}}(\mathbf{r}_i) \quad (7)$$

$$\text{Avec :} \quad v_{\text{ext}}(\mathbf{r}_i) = -\sum_{K=1}^N \frac{Z_K}{R_{Ki}} \quad (8)$$

$v_{\text{ext}}(\mathbf{r}_i)$  : Potentiel externe de l'électron  $i$ .

$\rho(\mathbf{r})$  : La densité électronique. Elle est égale à  $n$ , le nombre total des électrons, lorsqu'elle est intégrée sur tout l'espace.

$$\int_0^\infty \rho(\mathbf{r}) d\mathbf{r} = n \quad (9)$$

L'énergie totale du système peut s'écrire comme la somme de trois fonctionnelles : celle de l'énergie potentielle noyau-électron, de l'énergie cinétique et de l'énergie potentielle électron-électron.

$$E[\rho] = V_{ne}[\rho] + T[\rho] + V_{ee}[\rho] \quad (10)$$

En posant:  $V_{ne}[\rho] = \int(\mathbf{r}) v_{ext}(\mathbf{r}) d\mathbf{r} \quad (11)$

$$F[\rho] = T[\rho] + V_{ee}[\rho] \quad (12)$$

La fonctionnelle d'énergie s'écrit alors :

$$E[\rho] = \int(\mathbf{r}) v_{ext}(\mathbf{r}) d\mathbf{r} + F[\rho] \quad (13)$$

### I.3.1.2. 2<sup>ème</sup> théorème de Hohenberg et Kohn

Le second théorème applique le principe variationnel à la DFT. Ainsi, l'énergie fonctionnelle d'une densité électronique approchée est supérieure ou égale à l'énergie exacte du système dans son état fondamental [13].

$$E[\rho] \geq E[\rho_0] = E_0 \quad (14)$$

### I.3.2. Méthode de Kohn et Sham

W.Kohn et L.J.Sham (KS) [14] ont apporté des applications pratiques de la théorie de la densité. En 1965, ils ont proposé un ensemble d'équations mono électroniques à partir des quelles, il est possible d'obtenir la densité électronique et ensuite l'énergie totale.

Les orbitales de Kohn-Sham diffèrent des orbitales Hartree-fock :

Ce sont celles d'un système fictif d'électrons sans interaction, mais elles possèdent la même densité que le système réel. L'énergie totale d'un système à n électrons peut être écrite sans approximation [12], en exprimant les densités en fonction des orbitales moléculaires de Kohn-Sham [13].

Le système de référence est choisi de telle façon à avoir :

$$\rho_s(\mathbf{r}) = \rho_0 \quad (15)$$

Etant donné que les électrons n'interagissent pas entre eux dans le système de référence, l'hamiltonien de système de référence s'écrit :

$$\hat{H}_s = \sum_{i=1}^n [ -1/2 \nabla_i^2 + v_s(\mathbf{r}_i) ] = \sum_{i=1}^n h_i^{KS} \quad (16)$$



Avec :

$$\mathbf{h}_i^{\text{KS}} = -1/2 \nabla_i^2 + V_s(\mathbf{r}_i) \quad (17)$$

Par conséquent, les équations de Kohn et Sham, pour l'électron  $i$ , peuvent s'écrire comme suit :

$$\mathbf{h}_i^{\text{KS}} \theta_i^{\text{KS}} = \epsilon_i^{\text{KS}} \theta_i^{\text{KS}} \quad (18)$$

$\theta_i^{\text{KS}}$  : Orbitale de Kohn et Sham de l'électron  $i$ .

### I.3.3. Les différentes approximations

Comme c'est le problème de la fonction de corrélation et d'échange reste toujours délicat l'utilisation des approximations est toujours demandable et pour cela des nouvelles approximations ont été trouvées avec la DFT. D'après Ziegler les fonctionnels d'énergie d'échange-corrélation ont été classifiés en trois générations.

#### I.3.3.1. Approximation de la densité locale LDA

Hohenberg et Kohn ont montré que si  $r$  varie extrêmement lentement avec la position, l'énergie d'échange-corrélation  $E_{xc}[\rho]$  peut s'écrire comme suit :

$$E_{xc}^{\text{LDA}}[\rho] = \int \rho(\mathbf{r}) \epsilon_{xc}(\rho) d\mathbf{r} \quad (19)$$

$\epsilon_{xc}$  : étant l'énergie d'échange-corrélation par électron. Cette quantité est exprimée comme la somme des deux contributions (énergie d'échange  $\epsilon_x$  et énergie de corrélation  $\epsilon_c$ ) telle que:

$$\epsilon_{xc}(\rho) = \epsilon_x(\rho) + \epsilon_c(\rho) \quad (20)$$

avec :

$$\epsilon_x(\rho) = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{1/3} (\rho(\mathbf{r}))^{1/3} \quad (21)$$

Donc :

$$E_X^{\text{LDA}} = \int \rho \epsilon_x \, dr = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{1/3} \int [\rho(\mathbf{r})]^{4/3} \, dr \quad (22)$$

Le terme de corrélation  $\epsilon_c(\rho)$  est exprimé par la formule de Vosko, Wilk, et Nusair (VWN) [15]. Cette formule assez compliquée est donnée dans la référence [16].

### I.3.3.2. Approximation de la densité de spin locale LSDA

Pour les molécules à couches ouvertes et les géométries des molécules près de leur état de dissociation, l'approximation LSDA donne des résultats meilleurs que l'approximation LDA. Dans LDA, les électrons ayant des spins opposés ont les mêmes orbitales KS spatiales. En revanche, LSDA distingue entre les orbitales des électrons de spins opposés ( $\theta_{i\alpha}^{\text{KS}}$  pour les e- de spin  $\alpha$  et  $\theta_{i\beta}^{\text{KS}}$  pour les e- de spin  $\beta$ ). Par conséquent, on aura :

$$E_{\text{XC}} = E_{\text{XC}}[\rho^\alpha, \rho^\beta] \quad (23)$$

C'est l'équivalent de la méthode UHF (Unrestricted Hartree-Fock) pour les chaînes ouvertes.

### I.3.3.3 Approximation du Gradient Généralisé (GGA)

Cette méthode consiste à introduire le gradient de la densité dans la description des effets d'échange et de corrélation. Alors que la méthode LDA est locale de par la considération d'une densité équivalente à celle d'un gaz homogène, on incorpore ici des éléments non-locaux. En effet, le gradient de la densité permet de prendre en compte la variation de la densité au voisinage de chaque point.

L'énergie d'échange-corrélation, dans le cadre de l'approximation du gradient généralisé GGA (Generalized –gradient approximation), s'écrit alors:

$$E_{\text{XC}}^{\text{GGA}}[\rho^\alpha, \rho^\beta] = \int f(\rho^\alpha(\mathbf{r}), \rho^\beta(\mathbf{r}), \nabla\rho^\alpha(\mathbf{r}), \nabla\rho^\beta(\mathbf{r})) \, d\mathbf{r} \quad (24)$$

Où  $f$  est une fonction des densités de spin et de leurs gradients.

$E_{\text{XC}}^{\text{GGA}}$  : est divisé en deux contributions échange et corrélation.

$$E_{\text{XC}}^{\text{GGA}} = E_X^{\text{GGA}} + E_C^{\text{GGA}} \quad (25)$$

## I.4. Fonctionnelle hybride B3LYP

La fonctionnelle hybride B3LYP (Becke 3-paramètres Lee-Yang-Parr) est une fonctionnelle à trois paramètres combinant les fonctionnelles d'échange local, d'échange de Becke et d'échange HF, avec les fonctionnelles de corrélation locale (VWN) et corrigée du gradient de Lee, Yang et Parr :

$$E_{XC}^{B3LYP} = (1-a_0-a_x) E_X^{LDA} + a_0 E_X^{HF} + a_x E_X^{B88} + (1-a_c) E_C^{VWN} + a_c E_C^{LYP} \quad (26)$$

Où les paramètres  $a_0$ ,  $a_x$  et  $a_c$  ont été ajustés respectivement à 0,20, 0,72 et 0,81 [17].

## I.5. Références Bibliographiques

- [1] Landau.L et Lifchiz.E. (1967). Mécanique quantique, Editions Mir, Moscou.
- [2] Doucet, J.P.; Weber. J. (1996). Computer-aided molecular design: Theory and applications, Academic Press, London, 266.
- [3] Levine, I.N. (1999). Quantum Chemistry, 5th edition, 455.
- [4] T. Epstein. (1974). The Variation Method in Quantum Chemistry, Academic Press: New-York.
- [5] M. Born et J. R. Oppenheimer. (1927). Ann. Phys. 84, 457.
- [6] a) D. R. Hartree, Proc. Cambridge Philos. Soc, (1928), 24, 89.  
b) D. R. Hartree, Proc. Cambridge Phil. Soc, (1928), 24, 426.
- [7] a) V. Z. Fock, Z. Phys. (1930), 61, 126.  
b) V. Fock, Z. Phys. (1930), 62, 795.  
c) J. C. Slater, Phys. Rev. (1930), 48, 35.  
d) J. C. Slater. (1972), Adv. Quantum Chem. 6, 1.
- [8] J. L. Rivail. (1989). Eléments de chimie quantique à l'usage des chimistes, Inter-Editions/ Editions du CNRS: Paris.
- [9] Doucet, J.P. Weber. (1996). J. Computer-aided molecular design: Theory and applications, Academic Press, London, 266.
- [10] Bernard Bigot. (2007). dialogue confiant entre la chimie et la société. Discours prononcé à la Maison de la Chimie, à l'occasion des journées SFC 07.
- [11] R. G. Parr and W. Yang . (1989). Density Functional Theory, Oxford University Press.
- [12] P. Hohenberg and W. Kohn. (1964). Phys. Rev. 136, 864.
- [13] P. Chaquin. (2000). Manuel de Chimie théorique ; application à la structure et à la réactivité en chimie moléculaire, Ed, Ellipses.
- [14] W. Kohn and L. J. Sham. (1965). Phys. Rev. 140, A1133.
- [15] S. J. Vosko, L. Wilk and M. Nusair. (1980). Can. J. Phys., 58, 120.
- [16] F. Jensen. (1999). Introduction to Computational Chemistry. John Wiley & Sons, Chichester, England.
- [17] A.D.Becke, (1993). J. Chem. Phys, 98, 5648.

## II.1. Approches théoriques de la réactivité chimique

L'une des principales propriétés d'une molécule, c'est sa réactivité chimique, l'étude de cette réactivité et surtout sa prédiction permet de faciliter et d'orienter les expérimentateurs pour leurs synthèses, beaucoup de méthodes théoriques basées sur les principes de la mécanique quantique sont employées actuellement pour étudier la réactivité chimique [1,2]. A présent, deux théories sont largement utilisées par les chimistes : la théorie des orbitales frontières [3] initiée par Fukui en 1952 et récompensée par le Prix Nobel de Chimie en 1981 et la théorie conceptuelle de la fonctionnelle de la densité. Ces modèles ont conduit à la définition de divers indices de réactivité.

### II.1.1. Modèles de la réactivité chimique

#### II.1.1.1. Modèles empiriques

##### II.1.1.1.1. Principe HSAB (Hard and Soft Acids and Bases)

Pearson [4] a proposé, en 1963, de classer les acides et les bases en quatre catégories :

- les acides durs qui regroupent les électrophiles de petite taille, très positivement chargés, et faiblement polarisables,
- les acides mous qui regroupent les électrophiles de grosse taille, faiblement positivement chargés, et fortement polarisables.
- les bases dures qui regroupent les nucléophiles peu polarisables,
- les bases molles qui regroupent les nucléophiles très polarisables, Il en ressort alors le principe de réactivité suivant : Le principe HSAB stipule que les acides durs se lient fortement aux bases dures et les acides mous se lient fortement aux bases molles.

#### II.1.1.2. Modèles théoriques

##### II.1.1.2.1. Modèle théorique basé sur la théorie des orbitales frontières

Durant les années 50, La théorie des orbitales moléculaires frontières FMO a été développée par Fukui pour expliquer la régiosélectivité observée lors de réactions mettant en jeu des composés aromatiques. Notamment aux systèmes  $\pi$ , puis aux systèmes  $\sigma$ , où l'idée originale de Fukui consiste à postuler qu'au cours d'une réaction entre un nucléophile et un électrophile, le transfert de charge qui a lieu au voisinage de l'état de transition met en jeu principalement les électrons de l'orbitale moléculaire la plus haute occupée (HOMO) du nucléophile.

La théorie FMO joue un rôle primordial dans la rationalisation de la réactivité chimique. Cependant, elle a échoué dans plusieurs cas. La principale raison en est son champ d'application, qui

n'a été correctement établi qu'à partir de l'équation de Klopman- Salem proposé en 1968 et qui a été critiquée également dans les années 89 par Dewar [5].

### II.1.1.2.2. Modèle théorique basé sur la théorie de la fonctionnelle de la densité

Parallèlement à la DFT fondamentale, présentée précédemment, une seconde branche s'est développée à partir des années 1980, dont le but consiste à donner un cadre théorique précis à des concepts souvent mal définis caractérisant la réactivité chimique. Il s'agit du modèle de « DFT conceptuelle », selon l'appellation de R. G. Parr [6]. Ce modèle a été jugé le plus fructueux et semble toujours le plus prometteur [7]. Il constitue actuellement une vraie source de concepts chimiques comme le potentiel chimique électronique, l'électronégativité, la dureté, la mollesse, l'électrophilicité, ...etc.

#### II.1.1.2.2.1. Descripteurs globaux

##### a- Le potentiel chimique ( $\mu$ ) et l'électronégativité ( $\chi$ )

La théorie de la fonctionnelle de densité (DFT) constitue actuellement une vraie source de concepts chimiques comme le potentiel chimique électronique, l'électronégativité, la dureté, la mollesse, l'électrophilicité, ...etc. La DFT est fondée sur le principe variationnel. En effet, l'énergie d'un système est une fonctionnelle de la densité électronique. Pour obtenir la densité optimale, on minimise l'énergie  $E$ . En se basant sur la méthode de variations, cette contrainte est introduite via la méthode de multiplicateur de Lagrange conduisant à la condition variationnel suivante :

$$\delta[E - \mu(\int \rho(\mathbf{r}) - N)] = 0 \quad (1)$$

Où  $\mu$  est le multiplicateur de Lagrange :

$$v(\mathbf{r}) + \frac{\delta F_{\text{HK}}}{\delta \rho} = \mu \quad (2)$$

$v(\mathbf{r})$  : est le potentiel externe (i.e. du au noyaux) et  $F_{\text{HK}}$  est la fonctionnelle de Hohenberg et Kohn contenant les opérateurs de l'énergie cinétique des électrons et des répulsions inter électroniques [8].

##### b- Dureté et mollesse chimiques

Vue la discontinuité de l'énergie en fonction de  $N$ , on utilise généralement l'approximation de la différence finie pour obtenir  $\eta$  et  $S$ . Dans le cadre de cette approximation,  $\eta$  et  $S$  peuvent être écrits comme suit [9]:

$$\eta = \frac{I - A}{2} \quad (3)$$

$$S = \frac{1}{I - A} \quad (4)$$

I et A sont le potentiel de la 1<sup>ère</sup> ionisation verticale et l'affinité électronique de la molécule respectivement.

### Remarque

Le potentiel chimique électronique  $\mu$  et la dureté globale  $\eta$  peuvent être calculés à partir des énergies des orbitales moléculaires frontières  $\epsilon_{\text{HOMO}}$  et  $\epsilon_{\text{LUMO}}$  comme suit [9,10]:

$$\mu = (\epsilon_{\text{HOMO}} + \epsilon_{\text{LUMO}}) / 2 \quad (5)$$

$$\eta = (\epsilon_{\text{LUMO}} - \epsilon_{\text{HOMO}}) \quad (6)$$

La mollesse [11] est définie comme l'inverse de la dureté, c'est la capacité d'un atome ou d'une molécule de conserver une charge acquise, cette propriété est donnée par la relation suivante :

$$S = \frac{1}{2\eta} = \frac{1}{2} (I - A) \quad (7)$$

### c- Indice d'électrophilicité globale

Parr et al [12] ont défini un indice pour la puissance globale d'électrophilicité d'un système en termes de sa potentiel chimique  $\mu$  et la dureté chimique  $\eta$ :

$$\omega = \frac{\mu^2}{2\eta} \quad (8)$$

Cet indice est une mesure de la capacité d'espèce d'accepter un nombre arbitraire d'électrons.

#### II.1.1.2.2.2. Descripteurs locaux

Pour déterminer les sites réactifs d'une molécule lors de l'approche d'un agent électrophile, nucléophile ou radicalaire, les chimistes utilisent les charges nettes pour favoriser une interaction par rapport à une autre. Cependant, il est bien connu que les charges nettes calculées sur les différents sites d'une molécule ne sont pas un bon descripteur pour décrire les interactions entre molécules, particulièrement, pour les réactions contrôlées par les frontières c'est à dire les interactions Soft-Soft. En effet, la prédiction de la réactivité à l'aide des charges nettes peut conduire à des prédictions contradictoires avec l'expérience [13,14]. Dans ce qui suit, nous présentons succinctement le fondement théorique des principaux indices locaux utilisés actuellement pour la prédiction des sites réactifs d'une molécule, en l'occurrence : les indices de Fukui et les nucléophilies locales.

**a- Fonction de Fukui**

La fonction de Fukui, définie par Parr et Yang, est la quantité locale :

$$f(r) = \left( \frac{\delta \rho(r)}{\delta N} \right)_{v(r)} \quad (9)$$

L'évaluation de ces valeurs de  $f(r)$  est complètement compliquée. Yang et Mortier [15] ont donné un simple procédé pour calculer  $f_K$  (fonction condensée de Fukui), basée sur l'analyse de la population de Mulliken (MPA) et l'approximation de différence finie de l'équation (9) dans une approximation de différence finie, pour un système de N électrons, les valeurs de  $f_K$  sont données par:

$$f_K^+ = [q_K(N+1) - q_K(N)] \quad \text{pour une attaque nucléophile} \quad (10a)$$

$$f_K^- = [q_K(N) - q_K(N-1)] \quad \text{pour une attaque électrophile} \quad (10b)$$

$$f_K^0 = ([q_K(N+1) - q_K(N-1)]) / 2 \quad \text{pour une attaque radicalaire} \quad (10c)$$

Où  $q_K(N)$ ,  $q_K(N+1)$  et  $q_K(N-1)$  représentent la population électronique dans l'atome k pour N, N-1 et N+1 systèmes électroniques respectivement.  $f_K^+$  et  $f_K^-$  représentent l'habilité de l'atome k à réagir avec un nucléophile et électrophile, respectivement. Une valeur élevée de  $f_K^+$  indique que l'atome k présente un caractère électrophile indiquant de ce fait une haute probabilité pour une attaque nucléophile sur ce site, d'autre part, une valeur élevée de  $f_K^-$  signifie que le site k est plus nucléophile et disponible pour des attaques électrophiliques.

Une valeur élevée de  $f_K^0$  sur le site k indique que ce dernier présente une forte probabilité pour une attaque radicalaire.

**b- Molleses locales**

Alors que la dureté et la mollesse sont des paramètres globaux, les expérimentateurs ont parfois utilisé les termes durs et mous à propos d'atomes ou des groupements fonctionnels d'une molécule pour rationaliser leur réactivité. Ainsi pour dériver un analogue local à la mollesse, Parr et Yang [16] ont proposé :

$$S(r) = \left( \frac{\partial \rho(r)}{\partial \mu} \right)_{v(r)} \quad (11)$$

La relation avec la fonction de Fukui est immédiate, lorsqu'on applique la règle des dérivées en chaîne :



$$S(\mathbf{r}) = \left( \frac{\partial \rho(\mathbf{r})}{\partial N} \right)_{\mathbf{v}(\mathbf{r})} \left( \frac{\partial N}{\partial \mu} \right)_{\mathbf{v}(\mathbf{r})} = f(\mathbf{r}) S \quad (12)$$

Les molleses locales condensées  $S_K^\pm$  peuvent être facilement calculées à partir des fonctions de Fukui condensées  $f_K^\pm$  et de la mollesse globale  $S$  :

$$S_K^- = S [q_K(N) - q_K(N - 1)] = S f_K^- \quad (13)$$

$$S_K^+ = S [q_K(N + 1) - q_K(N)] = S f_K^+ \quad (14)$$

### c- Electrophilie locale

Le site le plus électrophile peut être facilement identifié par l'indice de l'électrophilie locale,  $\omega_K$  [17]; défini comme le produit de l'indice d'électrophilie globale  $\omega$  et l'indice de Fukui électrophilique  $f_K^+$ :

$$\omega_K = \omega f_K^+ \quad (15)$$

### d- Nucléophilie locale

Le site le plus nucléophile peut être facilement identifié par l'indice de la nucléophilie locale,  $N_K$  [18]; défini comme le produit de l'indice de nucléophile globale  $N$  et l'indice de Fukui nucléophilique  $f_K^-$ .

$$N_K = N f_K^- \quad (16)$$

## II.2. Etude quantitative des relations structure-activité (QSAR)

### II.2.1. Introduction

Les premiers essais de modélisation des relations structure-activité ont commencé à la fin du 19<sup>ème</sup> siècle, lorsque Crum-Brown et Frazer ont postulé que l'activité biologique d'une molécule est une fonction de sa constitution chimique [19,20]. Mais ce n'est qu'au début des années 60 que les travaux de Corwin Hanch ont proposé un modèle mathématique reliant l'activité biologique à la structure chimique [21]. Aujourd'hui, l'utilisation de QSAR n'a cessé de progresser [22,23]. Elle est devenue indispensable en chimie pharmaceutique et pour la conception de médicaments, notamment dans le cas où la disponibilité des échantillons est limitée ou les mesures expérimentales sont dangereuses, longues et chères [24,25]. Sans l'utilisation de grands instruments analytiques, les résultats des études QSAR peuvent fournir des informations utiles pour obtenir une meilleure

connaissance des structures moléculaires et probablement le mode d'action au niveau moléculaire. Ces informations peuvent être alors utilisées dans la prédiction des activités biologiques de nouveaux composés ainsi que dans la conception de nouvelles structures [26].

### II.2.2. Principe des méthodes QSAR

Le principe des méthodes QSPR/QSAR est de trouver une relation mathématique reliant de manière quantitative des propriétés moléculaires, appelées descripteurs, avec une propriété ou une activité biologique pour une série de composés chimiques similaires à l'aide des méthodes d'analyses de données. La forme générale d'un tel modèle est la suivante :

$$\text{Propriété/ Activité} = f(D1, D2, \dots Dn\dots) \quad (17)$$

D1, D2, ...Dn sont des descripteurs des structures moléculaires.

Les techniques QSAR s'appuient sur le concept postulant que des structures similaires ont des propriétés similaires [27], plus les molécules sont différentes, plus il est difficile à corrélérer les propriétés physico-chimiques et l'activité biologique, alors que le contraire est plus aisé.

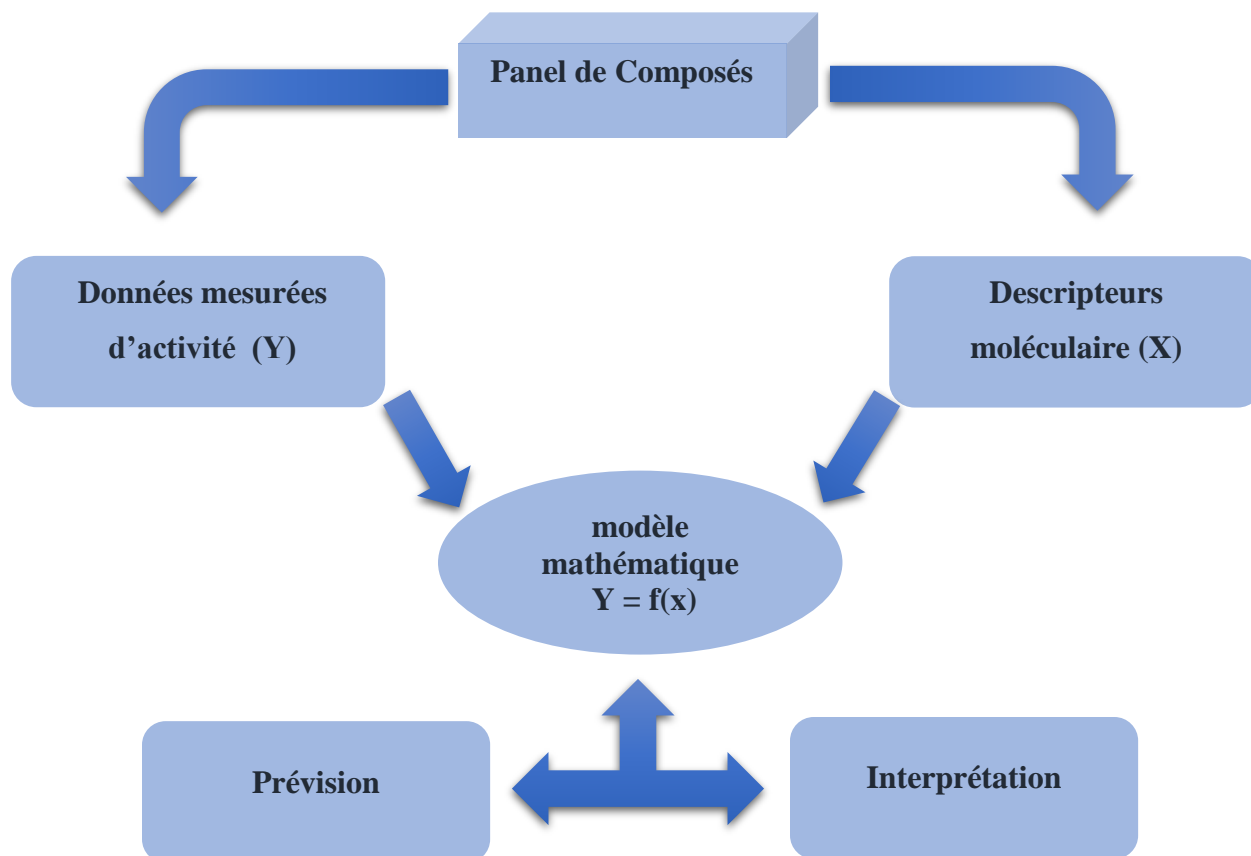


Figure II. 1 : Modèle d'étude quantitative des relations structure/activité (QSAR).

### II.2.3. Méthodologie générale d'une étude QSPR/QSAR

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

- 1- Collecte d'une base de données
- 2- Recherche de descripteurs adéquats pour l'activité étudiée
- 3- Le choix d'une méthode d'analyse des données
- 4- Validation du modèle
  - Validation interne : on utilise la série d'apprentissage constituée de 2/3 de la base de données. Ce type de validation a pour but de vérifier la stabilité et le pouvoir explicatif du modèle QSAR retenu.
  - Validation externe : on utilise la série de test constituée généralement de 1/3 de la base de données. Le but de cette validation est de vérifier le pouvoir prédictif du modèle élaboré.

### II.2.4. Outils et techniques de QSAR

#### II.2.4.1. 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 II.1 [28].

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

Source d'activité	Paramètres biologiques
<b>1. Récepteurs isolés</b> Constante de vitesse Constante de Michaelis-Menten Constante d'inhibition	Log k Log 1/K <sub>m</sub> Log 1/K <sub>i</sub>
<b>2. Systèmes cellulaires</b> Constante d'inhibition Résistance croisée Données biologiques in vitro Mutation de gène	Log 1/IC <sub>50</sub> Log CR Log 1/C Log TA <sub>98</sub>
<b>3. Systèmes in vivo</b> Facteur de bioconcentration Vitesses de la réaction in vivo Vitesses pharmacodynamiques	Log BCF Log I (induction) Log T (clairance totale)

### II.2.4.2. Descripteurs moléculaires

Les descripteurs moléculaires sont généralement classés en trois catégories ; les descripteurs physicochimiques, topologiques et électroniques. Ces descripteurs sont caractéristiques de la structure bidimensionnelle ou tridimensionnelle de la molécule.

L'étude QSAR-2D permet de rationaliser une propriété avec des descripteurs calculés à partir de la formule développée avec une équation de régression linéaire. Les études QSAR3D, décrites pour la première fois par Cramer et ses collaborateurs [29], permettent d'établir une relation entre une propriété et des descripteurs calculés à partir d'une conformation spatiale de la molécule définie.

#### II.2.4.2.1. Les descripteurs physico-chimiques

Les descripteurs physico-chimiques caractérisent généralement la structure bidimensionnelle de la molécule.

Parmi ces descripteurs, certains reflètent la composition moléculaire du composé, soit le nombre et le type d'atomes et de liaisons présents dans la molécule, son nombre de cycle, son poids moléculaire : ce sont les descripteurs constitutionnels [30]. Certains représentent la surface accessible au solvant (nommée « Connolly Accessible Surface »), le volume de solvant couvert par cette surface (« Connolly Solvent- Excluded Volume ») [31], le caractère hydrophile ou lipophile de la molécule généralement évalué à partir du coefficient de partage octanol/eau représenté par le log P [32].

Ces descripteurs ne fournissent pas assez d'information sur la structure des molécules pour l'élaboration de modèles prédictifs plus complexes ; il est nécessaire d'ajouter d'autres types de descripteurs.

#### II.2.4.2.2. Les descripteurs électroniques

Ces descripteurs caractérisent la distribution de charge des molécules (polarité des molécules) mais aussi les paramètres de la chimie quantique : le moment dipolaire, les énergies HOMO (Orbitale moléculaire occupée de plus haute énergie) et LUMO (Orbitale moléculaire non-occupée de plus basse énergie) [30].

#### II.2.4.2.3. Les descripteurs topologiques

Ces descripteurs décrivent les connectivités atomiques dans la molécule. Ils sont pour la plupart représentés par des indices comme ceux de Balaban, Wiener... et sont issus de la théorie des graphes. Cette théorie a été développée par Euler en 1736 avec les sept ponts de Königsberg. Un graphe est un

ensemble de point, certains reliés par des lignes ; il permet de représenter la topologie des molécules sans se soucier de la géométrie spatiale exacte de cette dernière [33].

### II.2.5. Importance des descripteurs

Les descripteurs moléculaires jouent un rôle fondamental en chimie, sciences pharmaceutiques, la protection de l'environnement, recherche de la santé et contrôle de qualité, être obtenu quand les molécules sont transformées dans une représentation moléculaire qui autorise quelque traitement mathématique. Les descripteurs moléculaires sont très importants pour :

- Indiquons la description de la configuration de la molécule à étudier.
- Décrivons tous les paramètres descriptifs de la molécule.

Les descripteurs moléculaires sont utilisés pour, une connaissance de statistiques, chimio métriques, et les principes des approches QSAR/QSPR sont nécessaires en plus de la connaissance spécifique du problème [34].

### II.2.6. Les techniques de QSAR

Certaines recherches mettent en jeu des données biologiques et physico-chimiques appropriées. Ces données peuvent être représentées et analysées de diverses manières. Le groupement et la classification des composés, basés sur leurs propriétés, sont les éléments principaux lors d'études de similarité moléculaire. Les études de régression ou de corrélation entre les données biologiques et chimiques sont généralement utiles pour rationaliser les relations structure-activité. Deux types principaux d'études peuvent être définis :

(1) par corrélation entre les données biologiques et physicochimiques en utilisant les techniques de régression.

(2) par classification des composés ou méthodes d'identification de modèle. Ces deux types d'études sont nommés analyse de données statistiques multi-variables, ou étude QSAR [35].

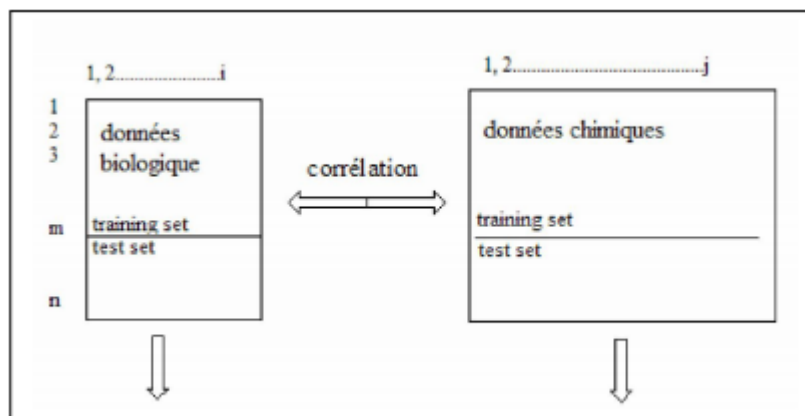


Figure II.2 : L'analyse de données multivariées.

## II.3. Statistiques et analyse des données

### II.3.1. Introduction

Pour élaborer un modèle QSPR/QSAR nous avons besoin d'une méthode d'analyse de données, cette méthode permet de quantifier la relation qui existe entre la propriété/Activité et la Structure (descripteurs).

Il existe plusieurs méthodes pour construire un modèle et analyser les données statistiques de ce dernier, certaines sont linéaires telles que la régression linéaire multiple (MLR), la régression aux moindres carrés partiels (PLS), d'autres sont non linéaires comme les arbres de décisions, les réseaux de neurones... ces méthodes sont disponibles dans des logiciels tels que, Excel, Origin Microcal, Minitab, Statistica, SPSS, R, ...

La méthode utilisée dans notre étude est la méthode de Régression Linéaire Multiple (MLR).

### II.3.2. Régression linéaire multiple (MLR)

Une analyse de la régression multiple est la méthode la plus simple et la plus employée pour le développement de modèles prédictifs [36]. Elle cherche à approximer une relation trop complexe en général, par une fonction mathématique simple. Elle repose sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante (à expliquer)  $Y$  (ici, 'activité biologique) et une série de  $p$  variables indépendantes (explicatives)  $X_i$  (ici, les descripteurs). L'objectif est d'obtenir une équation de la forme suivante :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p + \varepsilon \quad (18)$$

Cette équation est linéaire par rapport aux paramètres (coefficients de régression)  $\beta_0, \beta_1, \dots, \beta_p$ .

La détermination de l'équation (18) se fait alors à partir d'une base de données de n échantillons pour laquelle à la fois les variables indépendantes et la variable dépendante sont connues. Il s'agit donc de considérer un système d'équations :

$$Y_1 = \beta_0 + \beta_1 X_{1,1} + \beta_2 X_{2,1} + \dots + \beta_p X_{n,1} + \varepsilon_1 \tag{19}$$

$$Y_2 = \beta_0 + \beta_1 X_{1,2} + \beta_2 X_{2,2} + \dots + \beta_p X_{n,2} + \varepsilon_2 \tag{20}$$

.....

$$Y_n = \beta_0 + \beta_1 X_{1,p} + \beta_2 X_{2,p} + \dots + \beta_p X_{n,p} + \varepsilon_n \tag{21}$$

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'équation peut être donné sous la forme matricielle suivante :

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \dots \\ Y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & & & & \\ 1 & & & & \\ \dots & & & & \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_n \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \dots \\ \varepsilon_n \end{pmatrix}$$

### II.3.3. Estimation des paramètres statistiques du modèle

Dans le cas d'un modèle à p variables régresseurs, le critère des moindres carrés s'écrit :

$$S(\beta_0, \dots, \beta_p) = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{1i} - \dots - \beta_p X_{pi})^2 \tag{22}$$

Les valeurs des  $\beta$  qui minimisent ce critère seront les solutions  $\beta_0, \beta_1, \dots, \beta_p$  du système linéaire de (p+1) équations à (p+1) inconnues.

La méthode consiste alors à choisir les coefficients du vecteur b 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 Y (estimé par le modèle de régression) s'écrit :

$$\hat{Y} = X \hat{\beta} = XB \tag{23}$$

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

$$Y_i - \hat{Y}_i = \varepsilon_i \quad (24)$$

#### II.3.4. 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  $\beta_1 = 0$  et à calculer l'intervalle de confiance de  $\beta_1$ ,
- 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.

#### II.3.5. Tests sur le modèle linéaire

Comme pour le modèle linéaire simple, les hypothèses de régression linéaire doivent être vérifiées pour un modèle de régression multiple.

#### II.3.6. Test de la signification globale de la régression (F-Fisher)

Ce test permet de connaître l'apport global de l'ensemble des variables  $X_1, \dots, X_p$  à la détermination de Y.

On veut tester l'hypothèse nulle :

$H_0 : \beta_1 = \dots = \beta_p = 0$  contre  $H_a$  : il existe au moins un  $\beta_j$  parmi  $\beta_1, \dots, \beta_p$  non égal à 0.

On calcule la statistique de test  $F = \frac{MS \text{ model}}{MS \text{ error}} \quad (25)$

Avec :  $MS \text{ model} = \frac{SS \text{ model}}{P} \quad (26)$

Et  $MS \text{ error} = \frac{SS \text{ error}}{n - p - 1} \quad (27)$

Les expressions (26) et (27) représentent respectivement la somme des carrés des écarts moyens pour le modèle et pour l'erreur, avec p et n-p-1 sont les 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 de Fisher permet de tester l'hypothèse nulle selon laquelle chaque  $\beta$  est significativement différent de zéro, ce qui est signe d'une relation



évidente entre la variable expliquée et les variables explicatives.

### II.3.7. Sélection des variables et choix du modèle

Parmi l'ensemble des  $p$  variables disponibles, toutes n'ont pas nécessairement un intérêt dans la modélisation de  $Y$ , et il peut alors être néfaste de les utiliser. Nous sommes alors en présence de différents modèles possibles parmi lesquels il faut faire un choix.

#### II.3.7.1. Critères de comparaison de modèle

##### II.3.7.1.1. Limitation du coefficient de détermination $R^2$

$R^2$  qui varie entre 0 et 1, mesure la proportion de variation totale de  $Y$  autour de la moyenne expliquée par la régression. Plus la valeur de  $R^2$  sera proche de 1 (cas idéal) et plus les valeurs prédites et observées sont corrélées. Un  $R^2$  faible signifie que le modèle a un faible pouvoir explicatif et les descripteurs (certains d'eux) sont sans effet sur la réponse.

La valeur de  $R^2$  dépend de la taille de l'échantillon et le nombre de variables prédictives dans l'équation. Il garde la même valeur ou augmente lors d'une nouvelle variable de prédiction est ajoutée à l'équation de régression, même si la variable ajoutée ne contribue pas à la réduction de la variance inexpliquée. Par conséquent, un autre paramètre statistique peut être utilisé, appelé  $R^2_{\text{ajusté}}$  ( $R^2_{\text{ajusté}}$ ).

##### II.3.7.1.2. Coefficient de détermination ajusté $R^2_{\text{ajusté}}$

Il est obtenu par une expression semblable à celle de  $R^2$ , mais la SSR et SST sont divisés par leurs degrés de liberté correspondants. Ce coefficient est utilisé en régression multiple par ce qu'il tient compte du nombre de paramètres du modèle.

$$R^2_{\text{adj}} = 1 - \frac{(n - \text{int except})(1 - R^2)}{n - p} \quad (28)$$

Ou bien

$$R^2_{\text{adj}} = \frac{R^2(n-1) - p}{n - p - 1} \quad (29)$$

Cette formule indique notamment que  $R^2_{\text{adj}}$  est toujours inférieur à  $R^2$ , et ceci d'autant plus que le modèle contient un grand nombre de prédicteurs (descripteurs).

### II.4.7.1.3. Critère de validation croisée : PRESS

C'est la somme des carrées des écarts entre les valeurs observées et les valeurs prévues à partir du modèle.

$$\text{PRESS} = \sum (Y_{\text{obs}} - Y_{\text{calc}})^2 \quad (30)$$

Ce critère permet de sélectionner les modèles ayant un bon pouvoir prédictif. (On cherche toujours le PRESS le plus petit).

### II.3.8. Validation d'un modèle

Quel que soit le but ultime du modèle QSAR, il doit être obligatoirement validé avant d'être interprété ou utilisé à fins prédictives. Il existe différentes solutions pour s'assurer de validité d'un modèle. La validation croisée LOO (leave-one-out) [37] est un processus qui permet de tester la précision prédictive d'un modèle, cette méthode basée sur le calcul de quelques paramètres statistiques tels que : la somme des carrés résiduels RSS (PRESS), Somme des carrés totaux TSS, qualité d'ajustement  $R^2_{\text{adj}}$ , la validation standard des erreurs de prédiction (SPRESS) et le coefficient de corrélation de validation croisée  $R^2_{\text{CV}}$  c'est le coefficient qui décrit la validation est donné par l'équation ci-dessous

$$R^2_{\text{CV}} = 1 - \frac{\text{PRESS}}{\text{TSS}} \quad (31)$$

## II.4. Les applications de l'étude QSAR

Certaines études QSAR semblent être un peu plus d'études universitaires, il existe un grand nombre d'applications de ces modèles tels que [38]:

- L'optimisation de l'activité pharmacologique.
- La conception rationnelle de nombreux autres produits tels que des agents tensioactifs, des parfums, des colorants et des produits chimique fins.
- L'identification des composés dangereux dans les premiers stades de développement de produits ou la projection des stocks de composés existants.
- La prédiction de la toxicité et les effets secondaires de nouveaux composés.
- La prédiction de la toxicité pour les espèces.
- La prédiction de la toxicité pour les espèces environnementales.
- La sélection des composés ayant des propriétés pharmacocinétiques optimales, que ce soit la stabilité ou la disponibilité dans les systèmes biologiques.

- La prédiction d'une variété de propriétés physico-chimique des molécules.
- La prédiction du devenir des molécules qui sont libérées dans l'environnement.
- La prédiction des effets conjugués de molécules, que ce soit dans des mélanges ou des formulations.

## II.5. Références Bibliographiques

- [1] H. Chermette. (1999). Chemical reactivity indexes in density functional theory, *J. Comp. Chem.*, 20 :129–154.
- [2] F. De Proft, P. Geerlings. (2001). Conceptual and computational DFT in the study of aromaticity, *Chem.Rev.* 101 : 1451–1464.
- [3] K. Fukui, T. Yonezawa et H. Shingu. (1952). A Molecular Orbital Theory of Reactivity in Aromatic Hydrocarbons, *J. Chem. Phys.*, 20 : 722-725.
- [4] R. G. Pearson. (1963). Hard and Soft Acids and Bases, *J. Am. Chem. Soc.*, 85 : 3533-3539.
- [5] M.J.S. Dewar. (1989). A Critique of Frontier Orbital Theory, *J. Mol. Struct (Theochem)*, 259: 301,323.
- [6] R. G. Parr et W. Yang. (1995). Density-functional theory of the electronic- structure of molecules, *Annu. Rev. Phys. Chem.* 46 : 701-728.
- [7] P.Geerlings, F.Proft, W.Langenaeker. (2003). *Chem. Rev.*, 103 : 1793.
- [8] R. G. Parr, R. A. Donnelly, M. Levy, W. E. Palk. (1978). *J. Chem. Phys.*, 68, 3801.
- [9] R. G. Parr, W. Wang. (1989). *Density Theory for atoms and Molecules*, Oxford University Press: Oxford.
- [10] T. M. Lowry. (1923). *Chem. Ind.*, 42, 43.
- [11] R. G. Pearson. (1983). *J. Am. Chem. Soc.*, 105 :7512.
- [12] R. G.Parr, L. Szentpály, S. J. Liu. (1999). *Am. Chem. Soc.* 121 : 1922.
- [13] P. Geerlings, F. De Proft, W. Langenaeker. (2003). Conceptual density functional theory, *Chem. Rev.*, 103(5) : 1793–873.
- [14] W. Langenaeker, K. Demel, P. Geerlings. (1992). Quantum-chemical study of the Fukui function as a reactivity index Part 3. Nucleophilic addition to  $\alpha,\beta$ -unsaturated compounds I, *J. Mol. Struct. (THEOCHEM)*. 259 : 317-330.
- [15] W. Yang, W. Mortier, (1986). *J. Am. Chem. Soc.* 108, 5708.
- [16] W. Yang, R.G. Parr, *Proc.(1985). Natl. Acad. Sci. U.S.A.* 82, 6723.
- [17] a) P. Pérez, A. Toro-Labbe', A. Aizman, R. Contreras. (2002). Comparison between Experimental and Theoretical Scales of Electrophilicity in Benzhydryl Cations, *J. Org. Chem.* 67, 4747-4752.
- b) E. Chamorro, P.K. Chattaraj, P. Fuentealba. (2003). Variation of the electrophilicity index along the reaction path, *J. Phys. Chem. A.* 107 : 7068.
- [18] P. Pérez, L.R. Domingo, M. Duque-Noreña, E. Chamorro. (2009). A condensed  $\pi$ -atom

- nucleophilicity index. An application to the director effects on the electrophilic aromatic substitutions", *J. Mol. Struct. (Theochem)*. 895 : 86-91.
- [19] A. Crum Brown, T.R. Fraser. (1868). *Trans. Roy.Soc.* 25, 151-203.
- [20] Hansch, T. Fujita. (1964). *J. Am. Chem. Soc.* 86, 1616-1626.
- [21] A. K. Debnath. (2001). *Mini Reviews in Medicinal Chemistry*. I, 187-195.
- [22] M. Grover, B. Singh, M. Bakshi, S. Singh. (2000). *Pharm. Sci. Tech. Today*. 3, 50-57.
- [23] M. Grover, B. Singh, M. Bakshi, S. Singh. (2000). *Pharm. Sci. Tech. Today*. 3, 28-35.
- [24] T. I. Oprea. (2005). *Chemoinformatics in Drug Discovery*" Ed. WILEY-VCH Verlag. Allemagne.
- [25] E. A. Rekká, P. N. Kourounakis (2008). *Chemistry and Molecular Aspects of Drug Design and Action*. Ed. Taylor & Francis Group, LLC. Etats Unies,
- [26] A. K. Ghose, V. N. Viswanadhan. (2001). *Combinatorial Library Design and Evaluation: Principles, Software Tools, and Applications in Drug Discovery*. Ed. Marcel Dekker, Inc. Etats Unies.
- [27] A. G. Maldonado, J. P. Doucet, M. Petitjean, B. T. Fan. (2006). *Molecular similarity and diversity in Chemoinformatics: from theory to applications*. *Mol Divers*, 10 (1) : 39-79.
- [28] C. D. Selassie. (2003). *History of quantitative Structure-Activity relationships*, in Book "History of quantitative Structure-Activity relationships", D. J. Abraham, Editor. John Wiley & Sons, Inc.: California.
- [29] R.D. Cramer. D.E. Patterson, and J.D. Bunce. (1988). *Comparative molecular field analysis (CoMFA). Effect of shape on binding of steroids to carrier proteins*. *J. Am.Chem. Soc.* 110 : 5959-5967.
- [30] R. Bosque, J.Sales, E.Bosch, M. Rosès, Garcia-Alvarez-Coque M.C., and Torres- Lapasio J.R. (2003). *A QSPR study of the p solute polarity parameter to estimate retention in HPLC*. *J. Chem. Inf. Comput. Sci.* 43 : 1240-1247.
- [31] M.L. Connolly . (1985). *Computation of molecular volume*. *J. Am. Chem. Soc.* 107 : 118- 1124.
- [32] V.N. Viswanadhan, A.K. Ghose, G.R. Revankar, and R.K. Robins. (1989). *Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics*. *J. Chem. Inf. Comput.* 29 : 163- 172.
- [33] H.P. Schultz. (1989). *Topological organic chemistry. 1. Graph theory and topological indices of alkanes*. *J. Chem. Inf. Comput. Sci.* 29 : 227-228.
- [34] AI ACCESS, 91940, Les Ulis, France.

- [35] H. Waterbeemd, S. Rose. (2003). Quantitative approaches to structure-activity relationships, in Book Quantitative approaches to structure-activity relationships , Elsevier. 351-367.
- [36] I. Charif. (2012). Elaboration des corrélations quantitatives structure-activité des acides carbonés. Etude théorique des effets de solvants sur les équilibres libres cétoénoliques des composés b-dicarbonylés cycliques, Thèse de doctorat , Université de TLEMCEM.
- [37] R. Wehrens, H. Putter, L. M. C. Buydens. (2000). Chemom. Int. Lab. Syst., 54 : 35- 52.
- [38] Jerzy leszczynski. (2010). Challenges and advances in computational chemistry and physics, department of chemistry, jackson state university, U.S.A.

### III.1. Introduction

Ces dernières années, la grande avancée dans le domaine de l'informatique et les progrès réalisés dans les méthodes de calcul de chimie quantique permettent de donner une bonne prédiction et une bonne description des propriétés d'une entité donnée, ces propriétés sont reliées à la nature et la forme de la molécule.

L'objectif de notre travail, qu'été orientés vers l'obtention de l'effet de solvant sur les propriétés structurales et électroniques de la molécule de 2-iminothiazolidin-4-one. Ainsi nous avons utilisé les théories quantiques de la réactivité chimique [1,2] qui permettent actuellement, de justifier et prédire les propriétés de réactivité de la molécule en état gazeux et en état solvato dans l'eau.

Enfin, nous avons appliqué une méthode quantitative des relations structure/activité (QSAR) sur une série des dérivés de 2-iminothiazolidin-4-one. Cette méthode permette d'étudier les corrélations entre les paramètres structuraux et l'activité biologique d'une série bioactive des dérivés de 2-iminothiazolidin-4-one.

### III.2. Définition de 2-iminothiazolidine-4-one

2-iminothiazolidine-4-one est un hétérocycle caractérisé par un méthylène actif en position-5. De plus, ce sont tous des hétérocycles contiens une atome d'oxygène, d'azote ou de soufre successivement. Les dérivés du nos hétérocycles possède de propriétés biologique antiproliférative peut être utilisé comme inhibiteur de cancer.

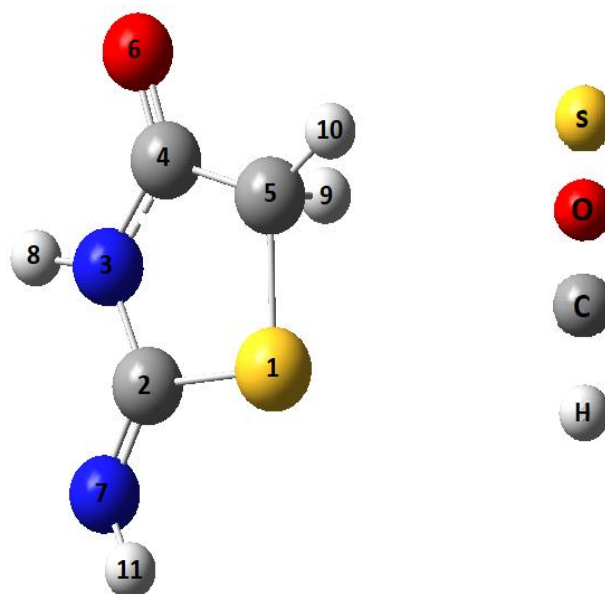


Figure III.1 : Structure 3D de noyau de base de 2-iminothiazolidin-4-one.

### III.3. Cellules de cancer du poumon A549 et H460

Il existe deux principaux types de cancer du poumon : le cancer du poumon non à petites cellules (le plus répandu) et le cancer du poumon à petites cellules (Tableau III.1) [3].

**Tableau III.1** : Classification des types et sous-types du cancer du poumon.

Cancer du poumon non à petites cellules	Cancer du poumon à petites cellules
Adénocarcinome (A549)	Carcinome à petites cellules
Carcinome épidermoïde (H460)	Carcinome mixte à petites et grandes cellules
Carcinome à grandes cellules indifférenciées	Carcinome mixte à petites cellules
Localisation : aucune partie du poumon n'est Privilégiée	Localisation : alvéoles, bronches, bronchioles, tissu de soutien des poumons

- Les cellules de la lignée A549 : dérivées d'un carcinome bronchiole-alvéolaire humain dont les caractéristiques ont été décrites par Lieber et al. (1976) [4].

- Les cellules de la lignée H460 : dérivées d'un carcinome pulmonaire épidermoïde humain. Les caractéristiques de ces cellules ont été décrites par Gruenert et al. (1995) [5].

### III.4. Méthodologie de calcul

Cette étude a été réalisée par l'utilisation de programme HyperChem 8.07 [6] pour optimiser la structure de 2-iminothiazolidin-4-one au moyen de la mécanique moléculaire, champ de force (MM+). Après cela, les structures minimisées obtenus précédemment ont été optimisées par utilisation de la méthode DFT par une fonctionnelle hybride B3LYP [7] et la base 6-311++G (d,p). Les calculs sont effectués par le logiciel gaussien 09 [8].

### III.5. Effet de solvant sur la structure de noyau de base de 2-iminothiazolidin-4-one

Pour obtenir l'effet de solvant, on a calculé les paramètres géométriques et électroniques de structure de noyau de base de 2-iminothiazolidin-4-one dans la phase gazeuse et dans la phase aqueuse.



## III.5.1. Effet de solvant sur les paramètres géométriques

Tableau III.2 : Distances interatomique en (Å).

Liaisons (Å)	phase gazeuse	phase aqueuse	L'écart
<b>S1_C2</b>	<b>1.7986</b>	<b>1.7915</b>	<b>-0.0071</b>
<b>S1_C5</b>	1.8342	1.8338	-0.0004
<b>C2_N3</b>	1.3919	1.3918	-0.0001
<b>N3_C4</b>	1.3752	1.3696	-0.0056
<b>N3_H8</b>	1.0122	1.0132	0.0010
<b>C4_C5</b>	1.5270	1.5211	-0.0059
<b>C4_O6</b>	<b>1.2084</b>	<b>1.2156</b>	<b>0.0072</b>
<b>C5_H9</b>	1.090	1.0897	-0.0004
<b>C5_H10</b>	1.0900	1.0896	-0.0004
<b>N7_C2</b>	1.2639	1.2679	0.0040
<b>N7_H11</b>	1.0166	1.0174	0.0008

Tableau III.3 : Angles de valence en (°).

Angles (°)	phase gazeuse	phase aqueuse	L'écart
<b>S1 C2 N3</b>	109.0243	109.2872	0.2629
<b>S1 C5 C4</b>	107.6132	107.3842	-0.229
<b>S1 C5 H9</b>	110.921	110.6117	-0.3093
<b>S1 C5 H10</b>	110.9215	110.6107	-0.3108
<b>C2 S1 C5</b>	92.4903	92.4322	-0.0581
<b>C2 N3 C4</b>	120.2033	119.7462	-0.4571
<b>C2 N3 H8</b>	119.0656	119.4431	0.3775
<b>C2 N7 H11</b>	<b>112.9984</b>	<b>112.3057</b>	<b>-0.6927</b>
<b>N3 C4 C5</b>	110.6688	111.1502	0.4814
<b>N3 C4 O6</b>	125.1491	124.7182	-0.4309
<b>C4 N3 H8</b>	120.7311	120.8107	0.0796
<b>C4 C5 H9</b>	109.1333	109.4911	0.3578
<b>C4 C5 H10</b>	109.1303	109.4889	0.3586
<b>C5 C4 O6</b>	124.1821	124.1317	-0.0504
<b>N7 C2 S1</b>	129.1465	128.7714	-0.3751
<b>N7 C2 N3</b>	121.8292	121.9414	0.1122
<b>H9 C5 H10</b>	109.0788	109.2253	0.1465

Tableau III.4 : Angles dièdres en (°).

Angles (°)	Phase gazeuse	phase aqueuse	L'écart
S1 C2 N3 C4	0.0250	0.0226	-0.0024
S1 C2 N3 H8	-179.9852	-179.9783	0.0069
C2 S1 C5 C4	-0.0049	-0.0042	0.0007
C2 S1 C5 H9	119.3044	119.4214	0.1170
C2 S1 C5 H10	-119.3108	-119.4266	-0.1158
C2 N3 C4 C5	-0.0293	-0.0263	0.0030
C2 N3 C4 O6	179.9784	179.9847	0.0063
N3 C4 C5 S1	0.0186	0.0166	-0.0020
N3 C4 C5 H9	-120.4250	-120.1236	0.3014
N3 C4 C5 H10	120.4612	120.1545	-0.3067
C5 S1 C2 N3	-0.0096	-0.0089	0.0007
C5 S1 C2 N7	179.9892	179.9933	0.0041
O6 C4 C5 S1	180.0111	180.0057	-0.0054
O6 C4 C5 H9	59.5674	59.8655	0.2981
O6 C4 C5 H10	-59.5464	-59.8564	-0.3100
N7 C2 N3 C4	-179.9739	-179.9794	-0.0055
N7 C2 N3 H8	0.0159	0.0197	0.0038
H8 N3 C4 C5	<b>179.9811</b>	<b>-180.0254</b>	<b>- 360.0065</b>
H8 N3 C4 O6	-0.0113	-0.0144	-0.0031
H11 N7 C2 S1	-0.0004	-0.0030	-0.0026
H11 N7 C2 N3	179.9983	179.9993	0.0010

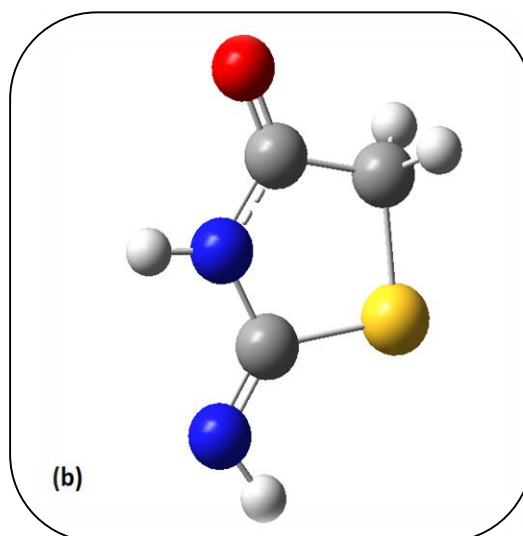
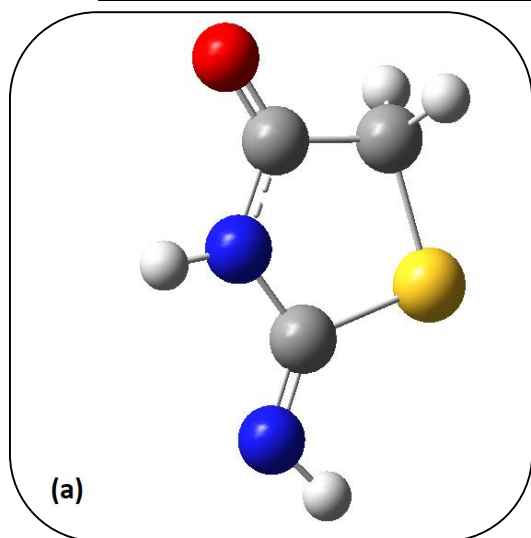


Figure III.2 : Structure optimisée de 2-iminothiazolidin-4-one dans la phase gazeuse (a) et dans la phase aqueuse (b).

Les paramètres géométriques de noyau de base de 2-iminothiazolidin-4-one sont rapportés dans les tableaux (III.2. III.3 et III.4).

D'après l'analyse des résultats obtenus dans les tableaux III.2. III.3 et III.4 on peut noter les remarques suivantes :

Les valeurs des distances interatomiques obtenus dans la phase gazeuse et dans la phase aqueuse sont similaires à l'exception des liaisons S1-C2 et C4-O6 où l'écart est : 0.0071 Å et 0.0072 Å respectivement.

L'augmentation de la longueur de la liaison polarisée C4-O6 dans la phase aqueuse est expliqué par l'effet de solvant polaire (l'eau) qui provoque l'augmentation de la polarisation de cette liaison et par la suite l'augmentation de sa longueur. La diminution de la longueur de liaison S1-C2 dans la phase aqueuse est expliqué par l'augmentation de la polarisation de liaison C4-O6.

Les valeurs des angles de valence obtenus dans la phase gazeuse et la phase aqueuse sont similaires et se diffèrent que dans l'angle C2-N7-H11.

Les valeurs des angles de torsion obtenus dans la phase gazeuse et la phase aqueuse sont totalement similaires. Les valeurs des angles de torsion confirment que la structure de 2-iminothiazolidin-4-one est plane.

Le plus grande variation dans les valeurs des angles de torsion ont été remarquées pour l'angle (H8-N3-C4-C5) où l'écart est 360.0065°.

Ces variations sont expliquées par l'effet des interactions entre le solvant et la molécule étudiée.

### III.5.2. Effet de solvant sur les paramètres électroniques

**Tableau III.5 :** Charges atomiques de 2-iminothiazolidin-4-one.

Atomes	phase gazeuse	phase aqueuse	$\Delta Q$
<b>S1</b>	0.2241	0.2406	0.0165
<b>C2</b>	0.3054	0.3086	0.0032
<b>N3</b>	-0.6371	-0.6290	0.0081
<b>C4</b>	<b>0.6833</b>	<b>0.6995</b>	<b>0.0162</b>
<b>C5</b>	-0.5794	-0.5770	0.0024
<b>O6</b>	<b>-0.5744</b>	<b>-0.6269</b>	<b>-0.0525</b>
<b>N7</b>	-0.6809	-0.7211	-0.0401
<b>H8</b>	0.4255	0.4350	0.0095
<b>H9</b>	0.2411	0.2535	0.0123
<b>H10</b>	0.2411	0.2535	0.0123
<b>H11</b>	0.3512	0.3632	0.0120

Tableau III.6 : Moment dipolaire.

	Moment dipolaire
Phase gazeuse	2.1850
phase aqueuse	2.9595

Les charges atomiques obtenues dans la phase gazeuse et dans la phase aqueuse sont différentes et les plus importantes qui sont trouvées dans les atomes C4 et O6 de la fonction carbonyle dont les quelles on remarque une augmentation de la charge positive du carbone fonctionnel et de la charge négative d'oxygène dans la phase aqueuse. Cette augmentation des charges des atomes C4 et O6 est expliquée par l'augmentation de la polarisation de la liaison C=O dans la phase aqueuse.

Les résultats du moment dipolaire représentés dans le tableau III.6 présentent une augmentation de ce paramètre dans la phase aqueuse ce qui indique l'effet induisant du solvant polaire sur la molécule étudiée.

## III.6. Etude de réactivité chimique de 2-iminothiazolidin-4-one

### III.6.1. Indices et descripteurs de réactivité

#### III.6.1.1. Indices globaux de réactivité

L'étude de la réactivité globale des molécules est basée sur le calcul des indices globaux déduits des propriétés électroniques. Le tableau ci-après récapitule les valeurs de l'énergie totale, énergies des orbitales moléculaires frontières, le gap énergétique, les valeurs de la dureté et du potentiel chimique électronique et l'indice d'électrophilicité.

Tableau III.7 : Descripteurs de réactivité utilisés.

	Etot (u,a)	HOMO (u,a)	LUMO (u,a)	GAP (u,a)	$\eta$ (u,a)	$\mu$ (u,a)	$\omega$ (eV)
Phase gazeuse	-699.7743	-0.2686	-0.0395	0.2291	0.1145	-0.1540	2.8197
Phase aqueuse	-699.7866	-0.2628	-0.0386	0.2242	0.1121	-0.1507	2.7578

Les orbitales moléculaires frontières la plus haute occupée (HOMO) et la plus basse vacante (LUMO) jouent un rôle fondamental dans l'interprétation qualitative de la réactivité chimique [9].

Les valeurs et les signes des coefficients des O.F constituent un indice important pour favoriser le mode d'attaque de la contribution des orbitales atomiques à l'orbitale moléculaire. Ainsi, l'attaque du réactif électrophile se produira préférentiellement sur le site ayant le plus gros coefficient dans la HOMO. Alternativement, une attaque nucléophile aura lieu de préférence sur les sites moléculaires ayant les plus grandes valeurs de la densité LUMO.

La valeur de gap énergétique de 2-iminothiazolidin-4-one dans la phase aqueuse ( $GAP = 0.2242$  u.a) est inférieur à celui dans l'état gazeux ( $GAP = 0.2291$  u.a) ce qui implique que l'écoulement des électrons sont facile ; donc la réactivité chimique est importante et la molécule est molle, par contre dans l'état gazeux, l'écoulement des électrons sont difficile ; donc la réactivité de la molécule est faible ce qui rend la molécule dur ( $\eta = 0.1145$  u.a).

Le potentiel chimique  $\mu$  du 2-iminothiazolidin-4-one dans la phase aqueuse ( $\mu = -0.15074$  a.u) est supérieur à celui dans la phase gazeuse ( $\mu = -0.1548$  a.u) ; ce qui implique que le transfert des électrons du 2-iminothiazolidin-4-one dans la phase aqueuse est mieux que dans la phase gazeuse.

L'indice d'électrophilicité du 2-iminothiazolidin-4-one dans la phase aqueuse ( $\omega = 2.7578$  eV) est légèrement inférieur à celui dans la phase gazeuse ( $\omega = 2.8197$  eV), Par conséquent, 2-iminothiazolidin-4-one dans l'état aqueuse est moins électrophile que dans la phase gazeuse. En d'autres termes, dans l'état aqueuse est plus nucléophile que dans la phase gazeuse.

Les valeurs de l'énergie totale calculées montre que 2-iminothiazolidin-4-one est stable dans la phase aqueuse que dans la phase gazeuse. La différence entre l'énergie totale dans les deux phase est ( $-1.23 \cdot 10^{-2}$  u.a) correspond à l'énergie de solvatation de 2-iminothiazolidin-4-one.

### III.6.1.2. Indices locaux de réactivité

Afin de différencier les comportements réactifs des atomes formant une molécule, différents indices ont été utilisés. Plus précisément, il s'agit des indices locaux et des descripteurs duals de la réactivité. Les descripteurs locaux de réactivité ont été employés pour caractériser les sites préférentiels pour les attaques nucléophiles, électrophiles et radicalaires.

Tableau III.8 : Valeurs des indices de Fukui de 2-iminothiazolidin-4-one.

Atomes	Phase gazeuse			Phase aqueuse		
	$f^-$	$f^+$	$f^0$	$f^-$	$f^+$	$f^0$
<b>S1</b>	0.0439	0.2292	<b>0.2816</b>	0.0831	0.2633	<b>0.3432</b>
<b>C2</b>	0.0050	-0.0612	-0.0281	0.1141	-0.0596	0.0272
<b>N3</b>	0.0486	0.0315	0.0401	-0.0001	0.0288	0.0144
<b>C4</b>	-0.0144	-0.0064	-0.0104	<b>0.2249</b>	0.0030	0.1139
<b>C5</b>	0.0617	-0.0482	0.0067	-0.0174	-0.0459	-0.0317
<b>O6</b>	0.0290	0.1013	0.0651	0.2095	0.0484	0.1290
<b>N7</b>	0.0176	<b>0.3003</b>	0.1590	0.1799	<b>0.2960</b>	0.2380
<b>H8 (N3)</b>	<b>0.3831</b>	0.0262	0.2046	0.0324	0.0156	0.0240
<b>H9 (C5)</b>	0.1478	0.0541	0.1009	0.0716	0.0461	0.0589
<b>H10 (C5)</b>	0.1480	0.0541	0.1010	0.0716	0.0461	0.0589
<b>H11 (N7)</b>	0.1298	0.0291	0.0794	0.0303	0.0179	0.0241

Le tableau III.8 représente les valeurs des indices de Fukui dans la phase gazeuse et dans la phase aqueuse, les figures III.3 et III.4 ci-dessous représentent la visualisation du potentiel électrostatique moléculaire (3D-MESP) des fonctions de Fukui dans les deux phases gazeuse et aqueuse.

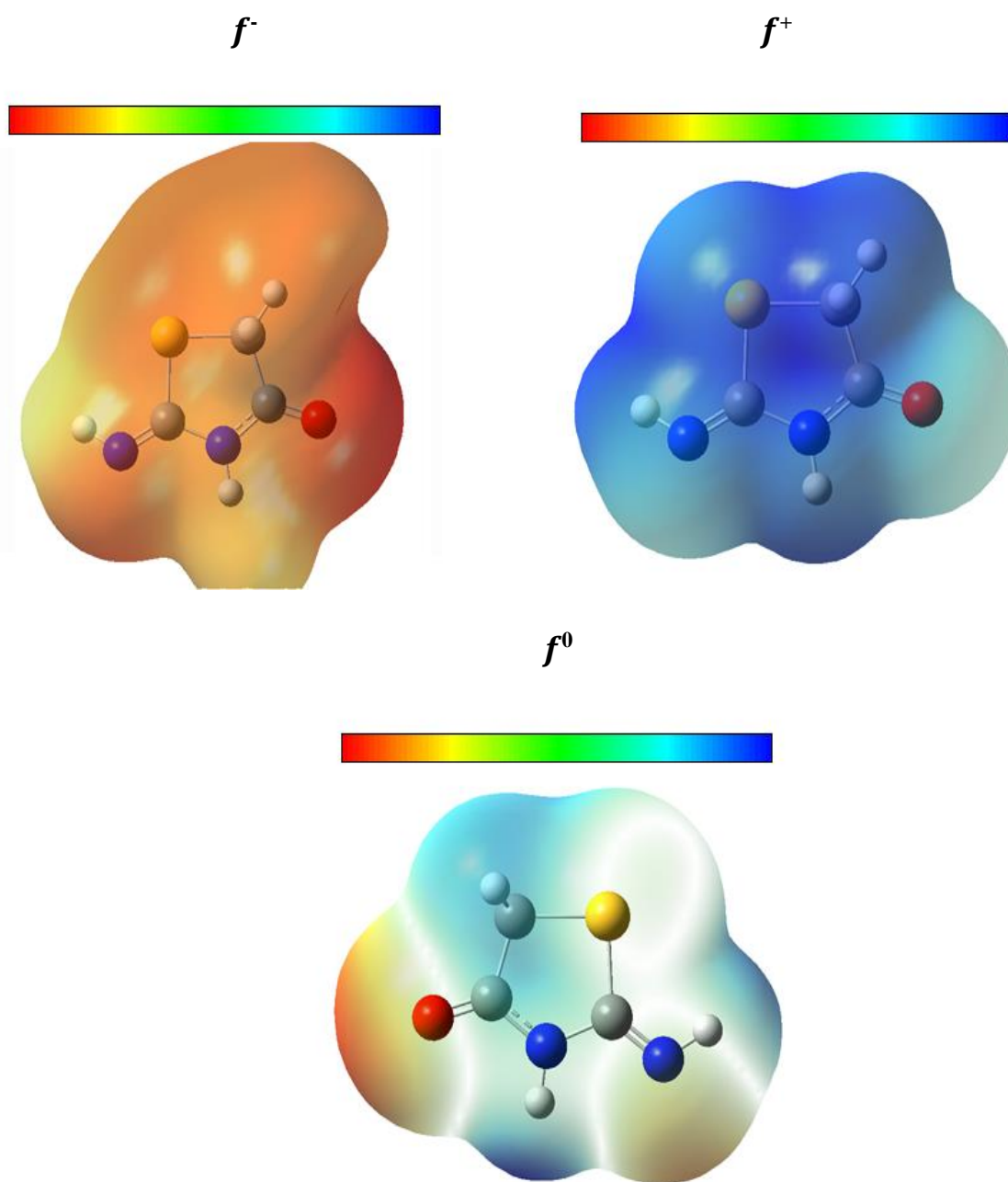
Les fonctions de Fukui décrivent la réactivité de molécule par rapport à un autre réactif électrophile, nucléophile ou radicalaire.

D'après les résultats du tableau III.8 on peut constater que l'atome d'azote N7 possède une grande valeur de  $f^+$  pour les deux phases ( gazeuse et aqueuse ).Ce résultat montre que la valeur élevée de  $f^+$  est indiquée le site de l'atome d'azote présente une haute disponibilité pour une attaque nucléophile.

L'atome d'soufre S1 possède une grande valeur de  $f^0$  Ce qui indique que cet atome est un site le plus important aux attaques radicalaires.

L'atome d'hydrogène H8 porté par N3 caractérisé par une grande valeur de  $f^-$  ce qui indiqué le site trop nucléophile et plus sensible aux attaques électrophiles ( $f^- = 0.3831$ ) à le cas gazeuse.

L'atome de carbone C4 représente une valeur plus élevés de  $f^-$  dans le cas aqueuse. Ce résultat montre que le site trop nucléophile et plus sensible aux attaques électrophiles ( $f^- = 0.2249$ )

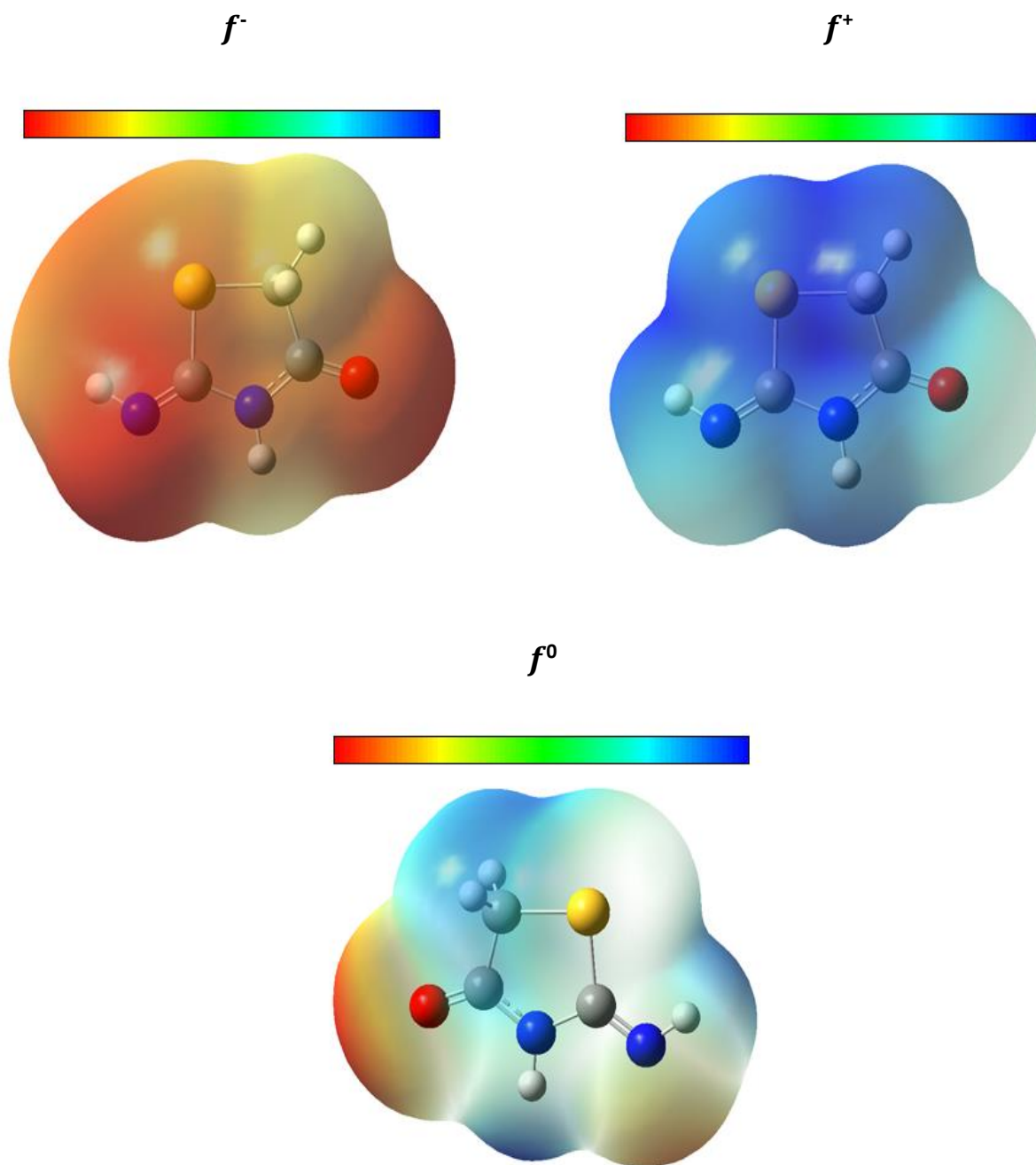


La constante de densité de charge d'électron (0.0004 e/au<sup>3</sup>).

La région de potentiel électro négatif de couleur rouge.

La région de potentiel électro positif de couleur rouge bleu.

**Figure III.3 :** Visualisation du 3D-MESP des fonctions de Fukui  $f^-$ ,  $f^+$  et  $f^0$  dans la phase gazeuse.



La constante de densité de charge d'électron (0.0004 e/au<sup>3</sup>).

La région de potentiel électro négatif de couleur rouge.

La région de potentiel électro positif de couleur rouge bleu.

**Figure III.4 :** Visualisation du 3D-MESP des fonctions de Fukui  $f^-$ ,  $f^+$  et  $f^0$  dans la phase aqueuse.



### III.7. Etude quantitative des relations structure-activité (QSAR) des propriétés antiprolifératives d'une série des dérivés de 2-iminothiazolidin-4-one

#### III.7.1. Méthodologie

##### III.7.1.1. Données expérimentales

Dans notre travail nous avons utilisé une série de 18 dérivés de 2-iminothiazolidin-4-one pour développer des modèles QSAR. Le travail de Xin zhai et al [10] nous permet de rassembler les données des structures et l'activité biologique expérimentale des dérivés de 2-iminothiazolidin-4-one sont rapportées dans le tableau III.9. L'activité antiprolifératives a été exprimée par PIC<sub>50</sub>.

**Tableau III.9 :** Structure et activité antiprolifératives expérimentale des dérivés de 2-iminothiazolidin-4-one.

composes	R	PIC <sub>50</sub> A549	PIC <sub>50</sub> H460
8c	2.5-dimethoxy	5.119	4.431
8f	2.3,4-trihydroxy	4.586	4.401
8L	2.4-dimethoxy	5.251	5.376
8N	3,5-dibromo-4-hydroxy	4.141	4.774
8T	2.4-difluoro	5.050	6.522
7	/	4.413	4.079
8Y	pyridin-4-yl	4.679	4.332
8E	/	5.420	5.522
8I	3,4-dihydroxy	4.767	4.707
8M	2.4.6-trimethoxy	4.943	5.744
8P	2-hydroxy	5.769	6.522
8Q	4-hydroxy	5.494	5.327
8R	2-fluoro	6,045	5.920
8S	4-fluoro	5.958	8.000
8X	thiophen-2-yl	4.393	4.642
8O	2-hydroxynaphthalene	5.040	5.091
8A	3,4.5-trimethoxy	4.567	5.387
8B	2.3,4-trimethoxy	4.013	4.326

### III.7.1.2. 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 [11].

Le choix des descripteurs dépend des outils dont on dispose, de la nature des composés décrits et de la propriété ciblée.

les dix-huit dérivés de 2-iminothiazolidin-4-one ont été pré-optimisées au moyen de la mécanique moléculaire, champ de force (MM+) en utilisant le logiciel HyperChem 8.03 [6]. Le module "propriétés QSAR" de l'HyperChem 8,03 a été utilisé pour calculer les paramètres suivants : la polarisabilité (Pol), le coefficient de partage octanol/eau (logP), l'énergie d'hydratation (HE), le volume moléculaire (V), la surface moléculaire (SAG) et le poids moléculaire (W). Les structures minimisées obtenus précédemment ont été optimisées en utilisant la méthode DFT avec une fonctionnelle hybride B3LYP [7] et la base 6-311++G(d,p) dans modèle CPCM ( eau ) de logiciel Gaussien 09[8].

Les molécules optimisées ont été utilisées pour calculer un certain nombre des descripteurs électroniques tel que : le moment dipolaire (MD), le gap énergétique, la charge atomique nette (qN7), la densité électronique des orbitales frontières électrophiles  $f_{N7}^E$  de l'atome N7.

La densité électronique d'orbitale moléculaire d'un atome quelconque est définie comme suit :

- ❖ Densité électronique des orbitales frontières électrophiles :

$$f_{i}^E = \sum (C_{HOMO})^2 \times 100 \quad (1)$$

- ❖ Densité électronique des orbitales frontières nucléophiles :

$$f_{i}^N = \sum (C_{LUMO})^2 \times 100 \quad (2)$$

**Tableau III.10** : Valeurs des descripteurs électroniques et physico-chimiques utilisés dans l'étude QSAR.

composé	qN7	$f_{N7}^E$	GAP	MD	HE	Log P	V	SAG	W	Pol
8c	-0.542	0.2542	0.1199	12.3804	-16.76	2.91	1343.13	794.86	490.53	51.6
8f	-0.5469	0.7903	0.1266	12.3123	-30.66	2.56	1251.17	740.71	478.48	48.57
8L	-0.5501	0.6335	0.1232	14.1115	-16.72	2.98	1346.55	801.18	490.53	52.15
8N	-0.5369	0.7875	0.1284	5.7883	-19.87	3.41	1233.3	733.86	482.46	47.12
8T	-0.5353	0.7738	0.1303	6.2931	-14.09	3.62	1210.57	722.17	466.46	46.48
7	-0.5426	3.625	0.1778	6.3283	-13.24	1.19	952.5	548.97	342.37	35.48
8Y	-0.5289	0.5596	0.1302	4.4114	-16.22	3.03	1181.26	708.36	431.47	45.95
8E	-0.5392	0.7252	0.1323	8.4606	-14.54	3.41	1198.85	717.49	430.48	46.66
8I	-0.5444	0.7101	0.1269	10.1688	-27.01	2.84	1238.88	736.21	462.48	47.93
8M	-0.5473	1.8169	0.1325	12.7679	-16.63	2.66	1426.82	839.51	520.56	54.08
8P	-0.5419	0.5968	0.1284	10.7657	-19.28	3.13	1216.57	723.65	446.48	47.3
8Q	-0.5455	0.8181	0.1287	10.4594	-21.36	3.13	1219.11	726.65	446.48	47.3
8R	-0.5357	0.6508	0.1304	8.245	-14.39	3.55	1200.95	718.04	448.47	46.57
8S	-0.5389	0.8032	0.132	6.6361	-14.24	3.55	1207.62	723.77	448.47	46.57
8X	-0.5419	0.6398	0.127	9.2126	-15.35	2.3	1172.46	699.76	436.5	46.18
8O	-0.5396	2.1611	0.1295	10.3031	-17.21	4.13	1333.04	778.99	496.54	54.57
8A	-0.5415	2.3858	0.1262	8.9965	-17.32	2.66	1411.62	824.19	520.56	54.08
8B	-0.5468	0.7196	0.1257	12.6297	-16.15	2.66	1415.1	829.36	520.56	54.08

### III.7.2. Développement des modèles QSAR

Les modèles Structure-activité ont été générés en utilisant la méthode de régression multilinéaire (MLR) par le logiciel SPSS 19 [12]. Les valeurs  $PIC_{50}$  sont indiquées en ( $\mu M$ ), ce sont des variables indépendantes. Les modèles QSAR sont évalués par la valeur de  $R^2$  (coefficient de détermination), la valeur de S (erreur standard de l'estimation) et la valeur F (statistique Fischer).

N : indique le nombre d'observations.

Les corrélations entre l'activité biologique et les descripteurs utilisées sont exprimées par les relations mathématiques suivantes :

#### 1- Dans l'activité A549 :

$$PIC_{50} = 77.750 + 1.066 f_{N7}^E - 69.387 GAP + 0.414 MD + 1.012 \text{ Log P} - 0.545 POL \\ + 0.033 SAG + 110.780 qN7 - 0.021 W \quad (1)$$

$$N = 18 ; \quad R^2 = 0.952; \quad F = 17.292; \quad S = 0.171$$

#### 2- Dans l'activité H460 :

$$PIC_{50} = 21.958 + 1.325 f_{N7}^E - 96.411 GAP + 0.193 MD + 0.061 HE + 1.291 \text{ Log P} \\ - 0.722 POL + 0.033 SAG \quad (2)$$

$$N = 18 ; \quad R^2 = 0.797; \quad F = 4.478; \quad S = 0.424$$

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.

Dans notre modèles QSAR obtenus  $R^2 = 0.952$  et  $R^2 = 0.797$ . Nous permis d'indiquer fermement la corrélation entre les différents descripteurs utilisés et les activité antiproliférative A549 et H460 respectivement.

La valeur F est jugée la signification statistique au niveau de 95%, pour toutes les valeurs de F calculées sont supérieures par rapport aux valeurs lus dans la table du Fischer (voire l'annexe).

Dans l'équation du modèle QSAR pour l'activité A549. Les coefficients négatifs de GAP, POL et W montrent que toute augmentation dans les valeurs de ces paramètres entraîne une diminution de

l'activité biologique, par contre pour l'activité H460 l'augmentation des paramètres GAP et POL provoque la diminution de l'activité antiproliférative des molécules étudiées.

Dans l'équation (1), les coefficients positives de  $f^E_{N7}$ . Log P, MD, SAG et qN7 montrent que toute augmentation dans les valeurs de ces paramètres provoque l'augmentation de l'activité antiproliférative. Même chose, dans l'équation (2), l'augmentation des paramètres  $f^E_{N7}$ . MD, HE, Log P, et SAG provoque l'augmentation de l'activité biologique.

Les matrices de corrélation entre les descripteurs obtenus par l'analyse MLR et l'activité biologique sont reportée dans les tableaux III.11 et III.12.

Tableau III.11 : Matrice de corrélation du modèle QSAR pour l'activité H460.

	PIC <sub>50</sub>	$f_{N7}^E$	GAP	MD	HE	Log P	POL	SAG
PIC <sub>50</sub>	1							
$f_{N7}^E$	-0.103	1						
GAP	-0.243	0.767	1					
MD	-0.021	-0.172	-0.411	1				
HE	0.285	0.233	0.305	-0.311	1			
Log P	<b>0.563</b>	-0.456	-0.612	-0.051	0.0143	1		
POL	0.229	-0.248	-0.760	0.642	-0.140	0.466	1	
SAG	0.262	-0.324	-0.780	0.640	-0.137	0.411	0.977	1

Tableau III.12 : Matrice de corrélation du modèle QSAR pour l'activité A549.

	PIC <sub>50</sub>	$f_{N7}^E$	GAP	MD	Log P	POL	SAG	qN7	W
PIC <sub>50</sub>	1								
$f_{N7}^E$	-0.231	1							
GAP	-0.167	0.767	1						
MD	0.118	-0.172	-0.411	1					
Log P	<b>0.489</b>	-0.456	-0.612	-0.051	1				
POL	0.034	-0.248	-0.760	0.642	0.466	1			
SAG	0.029	-0.324	-0.780	0.641	0.411	0.977	1		
qN7	0.103	-0.108	0.056	-0.863	0.382	-0.325	-0.341	1	
W	-0.08	-0.283	-0.763	0.574	0.433	0.955	0.970	-0.307	1

Les résultats que nous avons obtenus dans les tableaux (Tableau III.11 et Tableau III.12) montrent que le descripteur Log P est de paramètre le plus important dans la corrélation entre les descripteurs sélectionnés et l'activité antiproliférative des dérivés de 2-iminothiazolidin-4-one dans les deux modèles (H460 et A549).

### III.7.3. Validation des modèles QSAR

Afin de tester la validité de la puissance prédictive des modèles MLR sélectionnés (Eq. 1 et 2), la technique de 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 (TSS), qualité d'ajustement  $R^2_{adj}$ , coefficient de corrélation de validation croisée  $R^2_{cv}$ , la validation standard et les erreurs de prédiction ( $S_{PRESS}$ ) (Tableau III.13).

**Tableau III.13** : Paramètres de validation croisée.

Modèle	PRESS	TSS	PRESS/TSS	Spress	$r^2_{cv}$	$r^2_{adj}$
(1) A549	0.206	4.268	0.0482	0.106	0.952	0.897
(2) H460	1.438	7.071	0.203	0.282	0.797	0.619

PRESS est un paramètre important dans la validation croisée ainsi c'est une bonne approximation de la véritable erreur de prédiction des modèles. La petite valeur de PRESS signifie la meilleure prévisibilité du modèle [14]. En outre, pour un modèle raisonnable de QSAR, le rapport PRESS/TSS doit être inférieur à 0.4 [13].

Le tableau III.13 indiquent que les modèles développés ont des rapports de 0.0482 pour l'activité A549 et 0.203 pour l'activité H460 et les deux sont inférieurs à 0.4.

Les valeurs élevées du  $R^2_{cv}$  et  $R^2_{adj}$  sont des critères essentiels pour une meilleure qualification des modèles QSAR [14].

Dans notre étude on constate que  $R^2_{cv} = 0.952$  et  $R^2_{adj} = 0.897$  (dans le modèle 1), et  $R^2_{cv} = 0.797$  et  $R^2_{adj} = 0.619$  (dans le modèle 2).

Les activités antiprolifératives expérimentales, prédites et résiduelles de 2-iminothiazolidin-4-one sont représentés dans le tableau III.14.



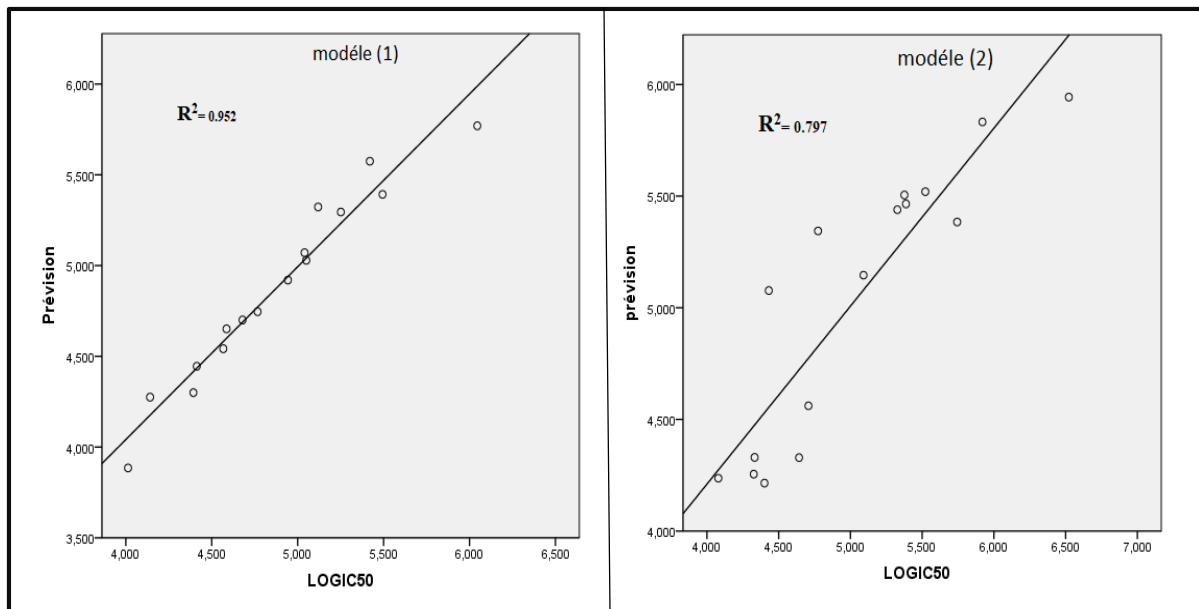
Le résidu est dû d'une part à l'incertitude de la mesure expérimentale et d'autre part à l'imperfection du modèle, donc si l'activité prédite est plus proche de l'activité expérimentale, le modèle peut être applicable [15].

**Tableau III.14 :** Activité antiprolifératives expérimentale, prédite et résiduelle des dérivés de 2-iminothiazolidin-4-one.

composés	Dans l'activité A549			Dans l'activité H460		
	Activité expérimentale	Activité prédite	Activité Résiduelle	Activité expérimentale	Activité prédite	Activité Résiduelle
1	5.119	5.323	-0.204	4.431	5.077	-0.646
2	4.586	4.651	-0.065	4.401	4.215	0.186
3	5.251	5.295	-0.044	5.376	5.505	-0.129
4	4.141	4.275	-0.134	4.774	5.344	-0.570
5	5.05	5.029	0.021	6.522	5.943	0.579
6	4.413	4.445	-0.032	4.079	4.237	-0.158
7	4.679	4.700	-0.021	4.332	4.330	0.002
8	5.42	5.575	-0.155	5.522	5.520	0.002
9	4.767	4.746	0.021	4.707	4.561	0.146
10	4.943	4.920	0.023	5.744	5.384	0.360
11	5.769	5.608	0.161	6.522	5.943	0.579
12	5.494	5.392	0.102	5.327	5.439	-0.112
13	6.045	5.770	0.275	5.92	5.832	0.088
14	5.958	5.364	0.594	8.000	7.201	0.799
15	4.393	4.300	0.093	4.642	4.329	0.313
16	5.04	5.071	-0.032	5.091	5.146	-0.055
17	4.567	4.542	0.025	5.387	5.465	-0.078
18	4.013	3.885	0.128	4.326	4.255	0.071

Figure III.5 ci-dessous montre les courbes de régression linéaire des valeurs prédites par rapport aux valeurs expérimentales de l'activité biologique de 2-iminothiazolidin-4-one. Les courbes des modèles 1 et 2 représentent pour être plus pratique avec  $r^2 = 0.952$  et  $0.797$  respectivement. Il indique

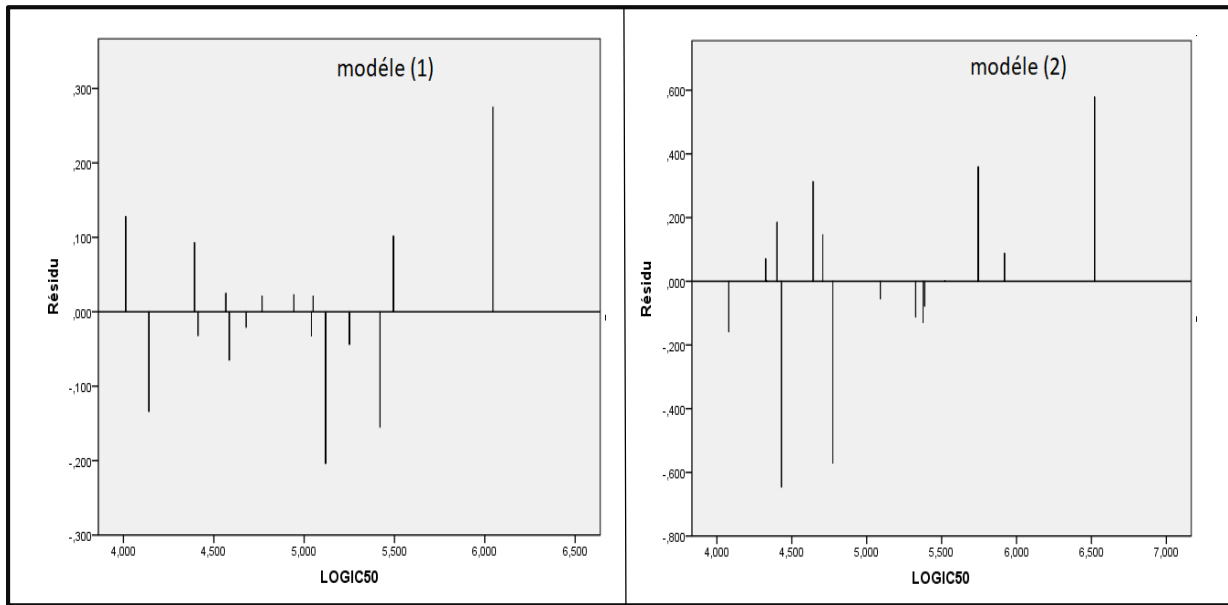
que les modèles 1 et 2 peuvent être appliqués avec succès pour prédire l'activité antiproliférative contre A549 et H460 dans cette série de molécules.



**Figure III.5 :** Corrélation entre l'activité biologique expérimentale et prédite pour les modèles (1) et (2).

Pour étudier la présence de l'erreur systématique dans l'élaboration des modèles QSAR, les valeurs résiduelles de l'activité biologique ont été tracées en fonction des valeurs expérimentales, comme le montre la Figure III.6.

La propagation des résidus des deux côtés de zéro indique qu'il n'y a aucune erreur systématique. Il indique que ces deux modèles peuvent être appliqués avec succès pour prédire l'activité antiproliférative pour la série de molécules utilisée dans le développement des modèles QSAR.



**Figure III.6 :** Corrélation entre les valeurs résiduelles et l'expérimentale des modèles (1) et (2).

### III.8. Références bibliographiques

- [1] H. Chermette. (1999). Chemical reactivity indexes in density functional theory, *J. Comp.Chem.* 20, 129– 154.
- [2] F. De Proft, P. Geerlings. (2001). Conceptual and computational DFT in the study of aromaticity, *Chem. Rev.* 101. 1451–1464.
- [3] Delphine Feron. (2010). Significations physiopathologiques des hémorphines de type 7 dans le diabète et les cancers broncho-pulmonaires, Thèse de doctorat, Université de La Rochelle.
- [4] Smith B. (1977). Cellline A549: A model system for the study of alveolar type II cell function. *Am. Rev. Respir. Dis.* 115 : 285-293.
- [5] William Raoul. (2005). Effets des facteurs angiogéniques et des cellules progénitrices dans la réparation de la barrière alvéolo-capillaire au cours des agressions pulmonaires aiguës, Thèse de doctorat, Université paris XII.
- [6] HyperChem (Molecular Modeling System) Hypercube, Inc., 1115 NW, 4th Street, Gainesville, FL 32601. USA, (2008).
- [7] F. Jensen; J. Wiley. (1999). *Introduction to Computational Chemistry*, Chichester, Angleterre.
- [8] Gaussian 09, Revision A.02, 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, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, 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. J. 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, (2009).
- [9] R. B. Woodward, R. Hoffmann. (1970). *The Conservation of Orbital Symmetry*; Chemie: Weinheim, Germany.
- [10] X. Zhai, W. Li, D. Chen, R. Lai, J. Liu, P. Gong, (2011). Design and Synthesis of 2-Iminothiazolidin-4-one Moiety Containing Compounds as Potent Antiproliferative Agents. *Arch Pharm*, 1-8.
- [11] H. Gonzalez-Diaz, E. Olazabal, L. Santana, E. Uriarte, Y. Gonzalez-Diaz, N. Castanedo, Bioorg.

Med. (2007). Chem., 15, 962.

[12] SPSS 19 for Windows, SPSS Incorporation, <http://www.spss.com>.

[13] K.Ajeet, Bijander. (2012). Int. J. Pharm. & Pharm. Sci., 4, 445.

[14] S.O. Podunavac-Kuzmanović, D.D. Cvetković, D.J. Barna. (2009). Int. J. Mol. Sci., 10, 1670.

[15] M.Clark, and R. D. Cramer, III. (1993). The probability of chance correlation using partial Least squares (PLS). Quantitative Structure-Activity Relationships, 12 (2): 45,137,

## *Conclusion générale*

Ce présent travail s'inscrit dans le cadre de l'étude sur des différents types de descripteurs de réactivité moléculaire selon une approche quantique basée sur la théorie de fonctionnelle de la densité (DFT). Des concepts chimiques et des indices de réactivité dérivant de la théorie de la fonctionnelle de densité ont été utilisé pour prédire les sites réactifs des systèmes moléculaires étudiés.

Pour cela nous avons étudié l'effet de solvant (eau) sur les paramètres géométriques, électroniques et énergétiques de la structure de noyau de base de 2-iminothiazolidin-4-one, en utilisant la base (B3LYP/6-311++ G(d,p)) pour faire les calculs. Les résultats montrent que la comparaison structurale et électronique pour le noyau de base de 2-iminothiazolidin-4-one dans la phase gazeuse et dans la phase aqueuse présente une différence de ces paramètres dans les deux phases, ce qui indique la déformation de la structure de 2-iminothiazolidin-4-one dans le solvant.

Après, nous avons étudiés l'effet de solvant sur la réactivité chimique de noyau de base de 2-iminothiazolidin-4-one. Les résultats obtenus montrent que la molécule de 2-iminothiazolidin-4-one possède une grande affinité pour donner des réactions ioniques (électrophile ou nucléophile).

Dans le dernier volet de ce travail nous avons réalisé une étude quantitative de la relation structure-activité (QSAR) sur une série des dérivés de 2-iminothiazolidin-4-one. Deux modèles QSAR ont été établis en utilisant la méthode de régression multilinéaire (MLR). Les modèles obtenus peuvent être utilisé pour prédire l'activité antiproliférative de cellules de cancer du poumon A549 et H460.

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 antiprolifératives de cellules de cancer du poumon A549 et H460, ce qui indique la validité et la qualité des modèles QSAR obtenus.



## Résumé

Le présent travail comporte une recherche fondamentale sur la molécule de 2-iminothiazolidine-4-one dont le but est de prédire l'effet de solvant sur la réactivité chimique et l'activité biologique des dérivés de 2-iminothiazolidine-4-one. Des concepts chimiques et des indices de réactivité dérivant de la théorie de la fonctionnelle de densité (DFT/B3LYP/6-311++G(d,p)) ont été utilisés pour prédire les sites réactifs des systèmes moléculaires étudiés dans les deux phases, gazeuse et aqueuse.

Une étude QSAR a été effectuée sur 18 dérivés de 2-iminothiazolidine-4-one contre les cellules de cancer du poumon A549 et H460 dans la phase aqueuse. La régression linéaire multiple (MLR) a été utilisée pour quantifier les relations entre les descripteurs moléculaires et l'activité antiproliférative des dérivés de 2-iminothiazolidine-4-one. 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, ce qui indique la validité et la qualité des modèles QSAR obtenus.

**Mots Clés :** 2-iminothiazolidine-4-one, Effet de solvant, DFT, QSAR, MLR.

## Abstract

This work involves a fundamental on the 2-iminothiazolidine-4-one molecules that aim to predict the solvent effect on the chemical reactivity and biological activity of derivatives of 2-iminothiazolidine-4-one. Chemical concepts and indications of reactivity derived from the density functional theory (DFT / B3LYP / 6-311 ++ G (d, p)) were used to predict the reactive sites of the molecular systems studied in gas phase and in water phase.

QSAR studies have been performed on 18 molecules of 2-iminothiazolidine-4-one derivatives against lung cancer cells A549 and H460 in the aqueous phase. A multiple linear regression (MLR) procedure was used to design the relationships between molecular descriptor and ant-proliferative activity of derivatives of 2-iminothiazolidine-4-one. The predictive power of the models obtained was confirmed by the method of cross-validation LOO. High correlation between experimental and predicted activity values of biological activities, indicating the validity and quality of QSAR models obtained.

**Keywords:** 2-iminothiazolidine-4-one, Effect of solvent, DFT, QSAR, MLR

## ملخص

العمل الحالي يدخل في إطار بحث أساسي حول جزيء 2-امينوثيازوليدين-4-أون والذي يهدف إلى التنبؤ بتأثير المذيب على الفاعلية الكيميائية والنشاط البيولوجي للمشتقات 2-امينوثيازوليدين-4-أون. المفاهيم الكيميائية ومؤشرات التفاعلية المستمدة من نظرية الكثافة الوظيفية (DFT/B3LYP/6-311++G(d,p)) تم استخدامها للتنبؤ مواقع التفاعلية لأنظمة الجزيئية في مرحلتين غازية ومائية.

لقد أجريت الدراسة QSAR على 18 جزيء من مماثلات 2-امينوثيازوليدين-4-أون ضد خلايا سرطان رئة A549 و H460 في الطور المائي. استعملنا طريقة الانحدار الخطي المتعدد لتكميم الارتباطات بين الواصفات الجزيئية والنشاط المضاد للتكاثر لمشتقات 2-امينوثيازوليدين-4-أون. تم تقدير التنبؤ للنماذج التي تم الحصول عليها من خلال طريقة اترك من اصل واحد (LOO). وقد لوحظ وجود ارتباط عال بين القيم التجريبية والمتوقعة للأنشطة البيولوجية، مما يشير إلى صحة وجودة نماذج QSAR التي تم الحصول عليها.

**كلمات مفتاحية:** 2-امينوثيازوليدين-4-أون، تأثير المذيب، الانحدار الخطي المتعدد، DFT، QSAR