



Université Mohamed Khider de Biskra
Faculté des sciences exactes et des sciences de la nature et de la vie
Département des sciences de la matière

MÉMOIRE DE MASTER

Science de la matière
Chimie
Chimie pharmaceutique

Réf. : N°1

Présenté et soutenu par :
KHELIL Ida HIRANE Fouzia

Le : mercredi 26 juin 2019

Etude QSAR des dérivés de la coumarine comme inhibiteur de la monoamine oxydase B

Jury :

Mme. HOUHOU Salima	MAA	Université Mohamed Khider de Biskra	Présidente
Mme. LEHRAKI Faiza	MAA	Université Mohamed Khider de Biskra	Rapporteur
Mme. ALMI Zineb	MCB	Université Mohamed Khider de Biskra	Examinatrice

Remerciement

Tout d'abord, nous tiens à remercier le bon Dieu le tout Puissant de notre avoir donné la force et le courage de mener à bien ce modeste travail, également je remercie infiniment notre parents, qui notre encouragé et aidé à arriver à ce stade de formation.

*Nous tenons tout à adresser nos profonds remerciements et gratitude à notre encadreuse **lehraki faiza**, pour sa disponibilité tout au long de la réalisation de ce mémoire, ainsi pour l'orientation, la confiance, la patience et le temps qui ont constitué un apport considérable sans lequel ce travail n'aurait pas pu être mené au bon port.*

*Nous tiens aussi à remercier les membres de jury **HOUHOU SALIMA** et **Almi Zineb** pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner notre travail et de l'enrichir par leurs propositions.*

A la fin nous tiens à remercier tous ceux et celle qui ont contribué à finaliser ce modeste travail.

Dédicace

À mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études.

À mes chères sœurs Mounira et Hadjer leurs encouragements permanents, et leur soutien moral.

À mes chers frères, Abed moumen et Abed ghani pour leur appui et leur encouragement.

À toute ma famille pour leur soutien tout au long de mon parcours universitaire.

Ma chère amie, mon binôme Fouzia H qui j'ai trouvé l'entente dont j'avais besoin.

À mes fidèles amis Housseem Dj ; Nadjela K et Fatima D ; Amina H ,qui ont m'accompagne tout au long de mes 5 ans.

Mes chers parents, ma mère Zohra et mon père Ahmed, qui m'ont encouragé à réussir mes études pour leurs sacrifices et leur patience

Aux membres de ma famille, à mes frères et à mes sœurs pour les encourager auprès de mes amis de l'université, j'ai vécu avec eux de beaux moments au cours de mes études à l'université, notamment à Ida, Fatima et Nedjla.

"Pour ma meilleure amie, je suis heureuse et merveilleuse qu'elle me comprenne, me conseille et m'encourage à communiquer dans mes études et je l'appelle pardon et miséricorde

"Bouthaina Merzoug "Que la bénisse.

À ma chère mère ZOÛRA pour son soutien infatigable, sa patience merveilleuse. Mon cher père Habib a poursuivi sa passion, ce qui m'a beaucoup aidé dans ma vie et pendant mes études.

Aux frères Kamal, Walid et Issam qui n'ont pas cessé de m'encourager et de prier pour moi.

À toutes mes sœurs et à tous les membres de leur famille. Pour ma deuxième mère, ma tante est Saida de me soutenir. Mes amis au travail:

Mohamed, Nabiha, Yazid, Hayatte, Saida, Sara, Sofia, Soumia, Manel, Chaima, Rim, Taheni, abir, Mabrouk, Sabrina, Radhia, Nessrin, Samira, Saber, Hadjer, Abdrazek, Adel, Soufian, Yahia, Haroun, Karim, Malek, Moussa, Hicham, Nabil et Bilal.

En fin, à toute personne qui m'a aidé de près ou de loin à la réalisation de ce travail.

HIRANE FOUZIA

SOMMAIRE

<i>Liste des figures</i>	I
<i>Liste des tableaux</i>	II
<i>Introduction générale</i>	1
<i>Références bibliographique</i>	3
CHAPITRE I: Bases théoriques	
Partie A : Quelques notions théoriques de base	4
I.A.1. Introduction	4
I.A.2. Objectifs	4
I.A.3. Approximation de Born-Oppenheimer	5
I.A.4. méthodes de calculs	5
I.A.4.1. Méthodes quantiques	5
I.A.4.1.1. Méthodes semi empiriques	5
I.A.4.1.2. Méthodes ab initio	5
I.A.4.1.3. La théorie de la fonctionnelle de la densité DFT	6
I.A.4.2. Mécanique moléculaire	10
I.A.4.2.1. Différents champs de force en mécanique moléculaire	11
I.A.4.2.2. Principe de la minimisation	12
I.A.4.3. Dynamique moléculaire	13
Partie B : QSAR	13
I.B.1. Introduction	13
I.B.2. Principe de QSAR	14
I.B.3. Les applications de l'étude QSAR	15
I.B.4. Outils et méthodologie de QSAR	15
I.B.4.1. Paramètres biologiques	15
I.B.4.2. Utilisation des descripteurs moléculaires	16
I.B.4.2.1. Sélection et réduction du nombre de descripteurs	21
Partie C : Méthodes statistiques	22
I.C.1. Définition	22
I.C.2. Méthodes statistiques	23
I.C.2.1. La statistique descriptive	23
I.C.2.2. La statistique décisionnelle ou prédictive	23
I.C.3. La régression linéaire multiple	24
I.C.4. Techniques de validation	25
I.C.4.1. Coefficients et tests statistiques standards	25
I.C.4.1.1. Coefficients de corrélation r	25
I.C.4.1.2. Le coefficient de détermination ajusté r^2_{adj}	26
I.C.4.1.3. L'erreur type résiduel «S»	26
I.C.4.1.4. Le test de Fisher F	27
I.C.4.2. Pouvoir de prévision	28
I.C.4.2.1. Interne	28
I.C.4.2.2. Externe	29
<i>Références bibliographiques</i>	<i>30</i>
CHAPITRE II: Généralité sur la coumarine, L'IMAO-B et la maladie de parkinson	
II.1. Coumarine	34
II.1.1. Structure chimique et nomenclature	34
II.2. Monoamine oxydases (MAO)	35
II.2.1. Isoformes	36
II.2.2. Structure de la MAO-B	36
II.2.3. Inhibiteurs de la MAO B	37

<i>II.3. Maladie de parkinson (MP)</i>	<i>37</i>
<i>II.3.1. Symptômes</i>	<i>38</i>
<i>II.3.2. Epidémiologie</i>	<i>38</i>
<i>II.3.3. Traitement actuel de la maladie de parkinson</i>	<i>39</i>
<i>Références bibliographiques</i>	<i>40</i>
CHAPITRE III: Etude QSAR et MPO des dérivées de Coumarine	
<i>III.1. Présentation de la série de molécule et des descripteurs</i>	<i>43</i>
<i>III.1.1. Série de molécule</i>	<i>43</i>
<i>III.1.2. Descripteurs</i>	<i>44</i>
<i>III.2. Méthodes et matériels</i>	<i>46</i>
<i>III.3. Corrélation entre différent descripteurs</i>	<i>47</i>
<i>III.4. Equation du modèle QSAR</i>	<i>48</i>
<i>III.4.1. Analyse de contribution des descripteurs</i>	<i>50</i>
<i>III.5. Validation de modèle</i>	<i>50</i>
<i>III.6. Méthodes d'optimisation multiparamétriques (MPO)</i>	<i>54</i>
<i>III.6.1. Règle de lipinski</i>	<i>54</i>
<i>III.6.2. Les règles de Veber</i>	<i>57</i>
<i>III.6.3. Efficacité de ligand (LE)</i>	<i>58</i>
<i>III.6.4. Efficacité lipophile de ligand (LLE)</i>	<i>59</i>
<i>III.6.5. Indice de lipophilicité dépendant de l'efficacité des ligands (LELP)</i>	<i>61</i>
<i>Références bibliographiques</i>	<i>62</i>
<i>Conclusion générale</i>	<i>65</i>

<i>Liste des figures</i>		<i>Pages</i>
<i>Chapitre I</i>		
<i>Figure .1:</i>	Organigramme des méthodes utilisées en chimie quantique	<i>10</i>
<i>Figure .2:</i>	Schéma de la méthodologie utilisée dans nos travaux.	<i>29</i>
<i>Chapitre II</i>		
<i>Figure .1:</i>	Combinaison d'un benzène avec un pyrane en une coumarine	<i>34</i>
<i>Figure .2:</i>	Parenté chimique de la coumarine et de son isomère structural	<i>34</i>
<i>Figure .3:</i>	Désamination oxydative des amines par les MAO.	<i>36</i>
<i>Figure .4:</i>	Structure cristallographique de la MAO-B humaine.	<i>37</i>
<i>Figure .5:</i>	Anatomie des ganglions de la base chez l'humain	<i>38</i>
<i>Chapitre III</i>		
<i>Figure .1:</i>	Structures chimiques de coumarine et ses dérivés	<i>43</i>
<i>Figure .2:</i>	La contribution des descripteurs choisis dans le modèle	<i>50</i>
<i>Figure .3:</i>	Corrélation entre l'activité biologique expérimentale et prédite.	<i>53</i>
<i>Figure .4:</i>	Corrélation entre l'activité biologique expérimentale et résiduelle.	<i>54</i>
<i>Figure .5:</i>	Règle de Lipinski.	<i>55</i>

<i>Liste des tableaux</i>		<i>Pages</i>
<i>Chapitre III</i>		
<i>Tableau.1</i>	Type de descripteurs moléculaire utilisés pour l'étude QSAR	<i>44</i>
<i>Tableau.2</i>	les descripteurs utilisés dans le développement de modèle QSAR	<i>45</i>
<i>Tableau.3</i>	La matrice de corrélation entre différent descripteurs	<i>47</i>
<i>Tableau.4</i>	Les descripteurs sélectionnés par MLR	<i>48</i>
<i>Tableau.5</i>	Table de Fisher.	<i>49</i>
<i>Tableau.6</i>	La matrice de corrélation du modèle.	<i>49</i>
<i>Tableau.7</i>	Paramètres de validation croisée.	<i>51</i>
<i>Tableau.8</i>	Les valeurs activité expérimentales, prédites et résiduelles des dérivées de Coumarine.	<i>52</i>
<i>Tableau.9</i>	Les règles de Lipinski des dérivés de coumarine	<i>56</i>
<i>Tableau.10</i>	Les règles de Veber des dérivés de coumarine.	<i>57</i>
<i>Tableau.11</i>	L'efficacité de ligand des dérivés de Coumarine	<i>59</i>
<i>Tableau.12</i>	Efficacité lipophile de ligand pour les dérivés de coumarine	<i>60</i>
<i>Tableau.13</i>	Indice de lipophilie des dérivés de coumarine	<i>61</i>



Introduction Générale

La maladie de Parkinson est une maladie neurodégénérative liée à une destruction des neurones dopaminergiques du ganglion basal du cerveau, impliquant une baisse de dopamine la plus connue et étudiée [1].

La communication neuronale se fait grâce à des centaines de neurotransmetteurs qui sont synthétisés au niveau des neurones, grâce à une machinerie d'enzymes. La Monoamine oxydase (MAO) est une enzyme qui joue un rôle important dans la dégradation des neurotransmetteurs mono aminergique (comme Dopamine).

Face à ce pathologie (maladie de parkinson), plusieurs stratégies ont été développés parmi lesquelles l'inhibition de la monoamine oxydase était historiquement la première voie qui a donné espoir à guérir des maladies longtemps considérées comme inguérissables. Ces composés appelés IMAO (Inhibiteurs de la Monoamine Oxydases) qui inhibent l'action de la MAO ont pour effet d'augmenter le niveau de neurotransmetteurs monoaminergique [2].

Au cours des dernière années, les dérivés de la coumarine ont été identifiés comme des inhibiteurs d'enzymes thérapeutiquement importants tels que l'aromatase, l'acétylcholinestérase (AChE) et de la MAO [3, 4].

La quantification de la relation existante entre la structure moléculaire et l'activité biologique (QSAR : Quantitative Structure Activity Relationship) est devenue, actuellement, un outil indispensable dans le domaine de la conception de médicaments, notamment dans le cas d'absence d'information sur le site actif de l'enzyme.

Techniquement, cette approche se base principalement sur: le choix de la structure de base, le choix des descripteurs moléculaires, la méthode de sélection des descripteurs et le choix la méthode statistique. La découverte d'une telle relation permet de prédire les propriétés physiques et chimiques ainsi que l'activité biologique des composés, de développer de nouvelles théories ou de comprendre les phénomènes observés. Elle permet également de modéliser d'analyser des familles de composés et de prévoir et de sélectionner ceux possédant une activité biologique potentielle [5-7].

Dans cette thèse, notre objectif est développer un modèle QSAR, par régression linéaire multiple, capable de prédire l'activité inhibitrice de la monoamine oxydase B d'une série dérivés de coumarine.

Le présent travail comporte trois chapitres contenant :

🚩 Le premier chapitre sera consacré à une étude bibliographique subdivisée en trois parties:

- *Partie A: La base théorique des différentes méthodes de la modélisation moléculaire appliquées à fin de réaliser ce travail.*
 - *Partie B: Intitulé «QSAR : principe et méthode » présente la méthode QSAR, son principe, son application et les différentes étapes à respecter pour la réaliser.*
 - *Partie C: la méthode statistique, nous avons utilisé la régression linéaire multiple MLR.*
- ✚ *Dans le deuxième chapitre comporte une recherche bibliographique sur coumarine et son activité biologique (inhibiteur de monoamine oxydase MAO) et la maladie de parkinson.*
- ✚ *Dans le troisième chapitre une étude quantitative qui visait à décrire les relations entre les propriétés physicochimiques, électronique,...etc et l'activité biologique d'une série de dérivés de coumarine par l'utilisation d'un Modèle QSAR, aussi une application des méthodes de sélection (drug likeness et indice de lipophilicité) .*
- Enfin, nous terminerons ce manuscrit par une conclusion générale.*

- [1] C. Christelle, *Monoamine oxydase et athérosclérose: signalisation mitogène et études in vivo. Thèse de doctorat d'université: innovation pharmacologique Toulouse: l'Université Toulouse III - Paul Sabatier. France. 2007, p 147.*
- [2] R. Benarab et M. Khenoussi, *Les inhibiteurs de la Monoamine oxydase en psychopharmacologie, l'Université Abderrahmane Mira de Bejaia, 2012, p1.*
- [3] C. Bruhlmann, F. Ooms, P.A. Carrupt, B. Testa, M. Catto, F. Leonetti, C. Altomare, A. Carotti. *Coumarins derivatives as dual inhibitors of acetylcholinesterase and monoamine oxidase. J.Med. Chem, 2001, 44, 3195-3198.*
- [4] C. Gnerre, M. Catto, F. Leonetti, P. Weber, P.A. Carrupt, C. Altomare, A. Carotti, B. Testa, *Inhibition of Monoamine Oxidases by Functionalized Coumarin Derivatives: Biological Activities, QSARs, and 3D-QSARs. J. Med. Chem, 2000, 43, 4747-4758.*
- [5] T. Puzyn, J. Leszczynski, M. T. D. Cronin, « *Recent advances in QSAR studies -Methods and applications* » Springer Dordrecht Heidelberg London New York, 2010.
- [6] M. Tichý, *Science.Tot. Env. 109, 1991, 407-410.*
- [7] B. Peric, J. Sierra, E. Martí, R. Cruañas, M. A. Garau, *Ecotoxicol. Env. Saf. 115, 2015, 257-262.*



Chapitre I :

Bases théoriques

Partie A: Quelques notions théoriques de base

I.A.1. Introduction

Le Développement important des moyens informatiques (mémoire plus importante, calcul scientifique, infographie, ...) au cours de ces dernières années a permis à la chimie de s'enrichir d'outils informatiques spécialement dédiés à la représentation des molécules. Ainsi un nouvel outil est apparu permettant à tout chimiste de mieux appréhender les problèmes liés au domaine moléculaire "*la Modélisation Moléculaire*".

La modélisation moléculaire implique l'utilisation de méthodes de calculs théoriques (mécanique moléculaire, dynamique moléculaire, mécanique quantique ab-initio ou semi-empirique, ...) permettant de déterminer la représentation graphique de la géométrie et d'évaluer les propriétés physico-chimiques de la molécule étudiée. La modélisation moléculaire permet d'interpréter des phénomènes physico-chimiques, de suggérer des nouvelles expériences et d'analyser ainsi des résultats d'une façon plus critique que les expériences classiquement utilisées, elle est donc complémentaire à l'expérience [1, 2].

I.A.2. Objectifs de la modélisation moléculaire

Nous citons dans ces lignes les principaux objectifs de la chimie théorique [3]:

- Visualisation et dessin des molécules à partir de données structurales.
- Utilisation de banques de données pour identifier les systèmes moléculaire; Obtention d'informations sur les mouvements dynamiques des molécules et sur leurs énergies.
- Déterminer les propriétés moléculaires et les distances interatomiques.
- Reproduction et prédiction qualitative des propriétés des molécules et des macromolécules.
- Corrélation entre des propriétés moléculaires et une structure moléculaire donnée
- Évaluation de la validité d'une structure moléculaire.

La plupart des méthodes de calculs quantiques utilisent l'approximation de Born-Oppenheimer (BO).

I.A.3. Approximation de Born-Oppenheimer

L'approximation de Born-Oppenheimer [4] est utilisée pour simplifier le problème moléculaire en séparant les mouvements électroniques et nucléaires. Cette approximation est basée sur le fait que les électrons, beaucoup moins lourds que les noyaux, peuvent s'adapter quasi instantanément aux déplacements nucléaires. Ainsi, la distribution électronique au sein d'une molécule dépend surtout de la position des noyaux et non pas de leurs vitesses.

I.A.4. Méthodes de calculs

I.A.4.1. Méthodes quantiques

I.A.4.1.1. Méthodes semi empiriques

Les méthodes semi empiriques sont des techniques de résolution de l'équation de Schrödinger de systèmes à plusieurs électrons. Contrairement aux *ab initio*, les méthodes semi empirique utilisent de données ajustées sur des résultats expérimentaux afin de simplifier les calculs. La longueur et la difficulté des calculs est en grande partie due aux intégrales biélectroniques qui apparaissent aux cours du processus de résolution.

Plusieurs méthodes semi-empiriques ont été développées telles que CNDO, MNDO, NDDO, [5] ...etc. Les méthodes les plus connues sont la méthode AM1 "*Austin Model 1*" [6] et la méthode PM3 "*Parametrization Model 3*" [7] qui offrent des avantages réellement significatifs concernant la rapidité du calcul et une plus grande flexibilité dans la dimension du système à étudier.

I.A.4.1.2. Méthodes *ab initio*

Le but ultime en chimie quantique est de pouvoir résoudre l'équation de Schrödinger, ce qui revient à chercher les énergies. Les méthodes *ab-initio* sont des méthodes non empiriques, il n'y'a pas d'approximation à faire sauf celle de Born Oppenheimer et l'approximation OM-CLOA [8]. Dans les méthodes *ab-initio*, toutes les particules (noyaux et électrons) sont traitées explicitement.

Cette méthode permet d'obtenir des résultats fiables et précis pour des molécules de petite taille. Cependant il est également possible d'obtenir de bons résultats à partir de méthodes quantiques moins exigeantes en moyens informatiques : les méthodes basées sur la théorie de la fonctionnelle de la densité (DFT) ont atteint un niveau de fiabilité compétitif avec des méthodes *ab initio* incluant la corrélation électronique, tout en permettant de réduire considérablement la charge de calculs.

1.A.4.1.3. La théorie de la fonctionnelle de la densité DFT

Au cours des trente dernières années, la Théorie de la Fonctionnelle de la Densité DFT (Density Functional Theory), venant compléter les méthodes HF et Post HF, s'est révélée comme particulièrement performante. Cette approche présente plusieurs avantages qui facilitent son utilisation dans différents domaines, surtout en chimie et sciences des matériaux.

- D'abord, elle inclut dans son formalisme une grande part de la corrélation électronique.
- L'aspect mono-électronique permet de donner une interprétation chimique de la fonction d'onde.

Dans le cas des méthodes Hartree-Fock, il s'agissait d'exprimer l'énergie du système en fonction de la fonction d'onde Ψ , alors que dans le cas de la DFT, l'énergie sera une fonctionnelle de la densité électronique ρ du système [9].

a. Objectif

L'objectif principal de la théorie de la fonctionnelle de la densité est de remplacer la fonction d'onde multiélectronique par la densité électronique en tant que quantité de base pour les calculs. Alors que la fonction d'onde multiélectronique dépend de $3N$ variables (où N est le nombre total de particules du système), la densité est seulement fonction de trois variables ; il s'agit donc d'une quantité plus facile à traiter tant mathématiquement que conceptuellement.

Le principe de la DFT consiste en une reformulation du problème quantique à N corps en un problème monocorps (ou, à la rigueur, bi-corps si l'on considère les problèmes de spin) avec pour paramètre la densité électronique. L'idée centrale de la DFT est que la seule densité électronique de l'état fondamental du système détermine entièrement les valeurs moyennes des observables, comme l'énergie [10].

La DFT trouve ses origines dans le modèle développé par Llewellyn Thomas et Enrico Fermi à la fin des années 1920. Néanmoins il faudra attendre le milieu des années 1960 et les contributions de Pierre Hohenberg, Walter Kohn et Lu Sham pour que soit établi le formalisme théorique sur lequel repose la méthode actuelle [11].

b. Approximations en DFT

b.1. Approximation locale de la densité (LDA)

La fonctionnelle d'échange-corrélation est une grandeur physique dépendante de plusieurs paramètres comme la densité électronique, l'énergie cinétique, etc.

Afin d'obtenir une bonne approximation de cette fonction, elle doit reproduire au mieux les caractéristiques physiques du trou d'échange-corrélation. Elle se base sur les propriétés d'un gaz homogène d'électrons ; la densité électronique et la fonction d'onde sont considérées localement comme constantes.

Par le concept de LDA il est possible d'estimer l'énergie d'échange-corrélation d'un système inhomogène en utilisant les résultats d'un gaz homogène d'électrons de densité égale à la densité locale d'un système inhomogène. L'énergie de corrélation d'un gaz uniforme d'électrons n'est pas déterminée exactement. Cependant, certaines expressions approchées permettent de décrire les densités électroniques faibles ou élevées. Cette approximation considérée localement permet de décrire correctement les systèmes ayant une densité électronique variant faiblement dans l'espace [12].

b.2. L'approximation du gradient généralisé (GGA)

Ainsi une amélioration peut donc être apportée sur ce point en utilisant des corrections qui prennent en compte les variations locales de la densité (gradient). Ces approximations sont définies comme semi-locales et sont regroupées sous le nom de GGA (Generalized Gradient Approximation). Les variations semi-locales impliquent à la fois l'échange et la corrélation. Ces fonctionnelles améliorent dans de nombreux cas les résultats structuraux et énergétiques. Les plus utilisées sont celles proposées par Perdew (P86), [13] Becke (B88), [14] Perdew et Wang (PW86 et PW91) [15] et par Perdew, Burke et Ernzerhof (PBE) sous sa forme initiale [16] ainsi que sous sa forme modifiée. La fonctionnelle d'échange corrélation reste la seule approximation liée à la DFT, même si sa qualité s'améliore constamment avec par exemple l'apparition de fonctionnelles non locales [17]. Mais d'autres approximations sont nécessaires pour l'étude de nos systèmes.

b.3. Fonctionnelles hybrides pour le calcul moléculaire

Les fonctionnelles hybrides, quant à elles, reprennent l'énergie d'échange calculée par méthode HF qui peut être calculée de façon exacte contrairement aux fonctionnelles approximatives d'échange disponible en DFT. Elles reprennent également une partie des énergies d'échange et de corrélation de fonctionnelles LDA et GGA, le tout pondéré par des coefficients bien déterminés.

c. La fonctionnelle B3LYP

La fonctionnelle B3LYP est comptée parmi les fonctionnels hybrides les plus utilisées pour l'étude de systèmes moléculaires.

d. Bases d'orbitales atomiques

Les orbitales moléculaires obtenues par les méthodes *ab-initio* ou DFT sont exprimés comme des combinaisons linéaires d'orbitales atomiques (LCAO) [18]. Une orbitale atomique de type 1s, 2s, 2px, 2py peut être représentée mathématiquement par une ou plusieurs fonctions (proches des solutions de l'équation de Schrödinger pour l'atome H). L'ensemble de ces fonctions pour tous les atomes d'une molécule est appelé «base d'orbitales atomiques». Cette dernière est nommée « base simple zêta » si chaque orbitale atomique d'un atome est représentée par une seule fonction, « base double zêta » si elle est représentée par deux fonctions.

Plusieurs catégories de bases peuvent être distinguées, les plus développées sont :

1- Les bases minimales : dans ces bases, les orbitales de coeur et de valence comprennent le même nombre de primitives gaussiennes. Les résultats obtenus sont moins coûteux en temps de calcul mais restent insuffisants pour la recherche et les comparaisons avec les analyses expérimentales.

2- Les bases de Pople : [19] ce sont des bases à valence séparée qui ont été introduites par l'équipe de John Pople et sont typiquement en forme de « X-YZG ». Dans ce cas, X représente le nombre de gaussiennes utilisées pour décrire chaque orbitale atomique de coeur. Le Y et le Z indiquent que les orbitales de valence sont composées chacune de deux fonctions, la première étant composée d'une combinaison linéaire de Y fonctions gaussiennes primitives, l'autre d'une combinaison linéaire de Z fonctions gaussiennes. Dans ce cas, la présence de deux nombres après le trait d'union indique que la base est une base double zêta à valence séparée. La différence entre deux fonctions Y et Z réside essentiellement dans les exposant ξ des fonctions gaussiennes, plus grands pour Y (la fonction Y est plus contractée que Z). Les bases les plus utilisées dans cette catégorie sont la 6-31G et la 3-21G.

La base 6-31G (d, p) est celle utilisée pour tous les calculs de cette mémoire.

Base 6-31G :

La base 6-31G est constituée de deux familles de fonctions :

- (i) une fonction de base approchée par 6 gaussiennes pour décrire les électrons de coeur.
- (ii) Deux fonctions de base approchées respectivement par 3 et 1 gaussiennes pour décrire les électrons de valence.

La qualité des bases peut encore être améliorée en ajoutant :

⇒ des fonctions de polarisation.

⇒ des fonctions diffuses.

Pour une plus grande flexibilité on peut encore rajouter des fonctions de polarisation (déformation du nuage électronique, et des distorsions dues à l'environnement). La dénomination la plus ancienne est l'ajout d'un astérisque sur la base en question (par exemple 6-31G*), et dans une désignation plus récente, le caractère de la fonction ajoutée est explicitement donné : 6-31G(d). La base 6-31G(d) signifie ainsi qu'un jeu de fonctions « d » a été ajouté à tous les atomes (sauf H) dans la molécule, alors que 6-31G** ou 6-31G (p, d) signifie qu'un jeu de fonctions « p » a été ajouté aux hydrogènes et que des fonctions « d » ont été ajoutées aux autres atomes.

Un autre type de fonctions est indispensable à inclure dans la base d'orbitale atomique chaque fois que le phénomène physique décrivant la propriété étudiée nécessite une bonne description de l'espace situé au-delà des orbitales de valence (espace diffus). Ce sont les fonctions diffuses, qui augmentent la taille du nuage électronique. Pour les espèces ayant des doublets libres et les espèces chargées (anions), la présence d'orbitales diffuses est indispensable. On note par le signe (+) la présence d'orbitales diffuses, et par un astérisque(*) les orbitales de polarisation.

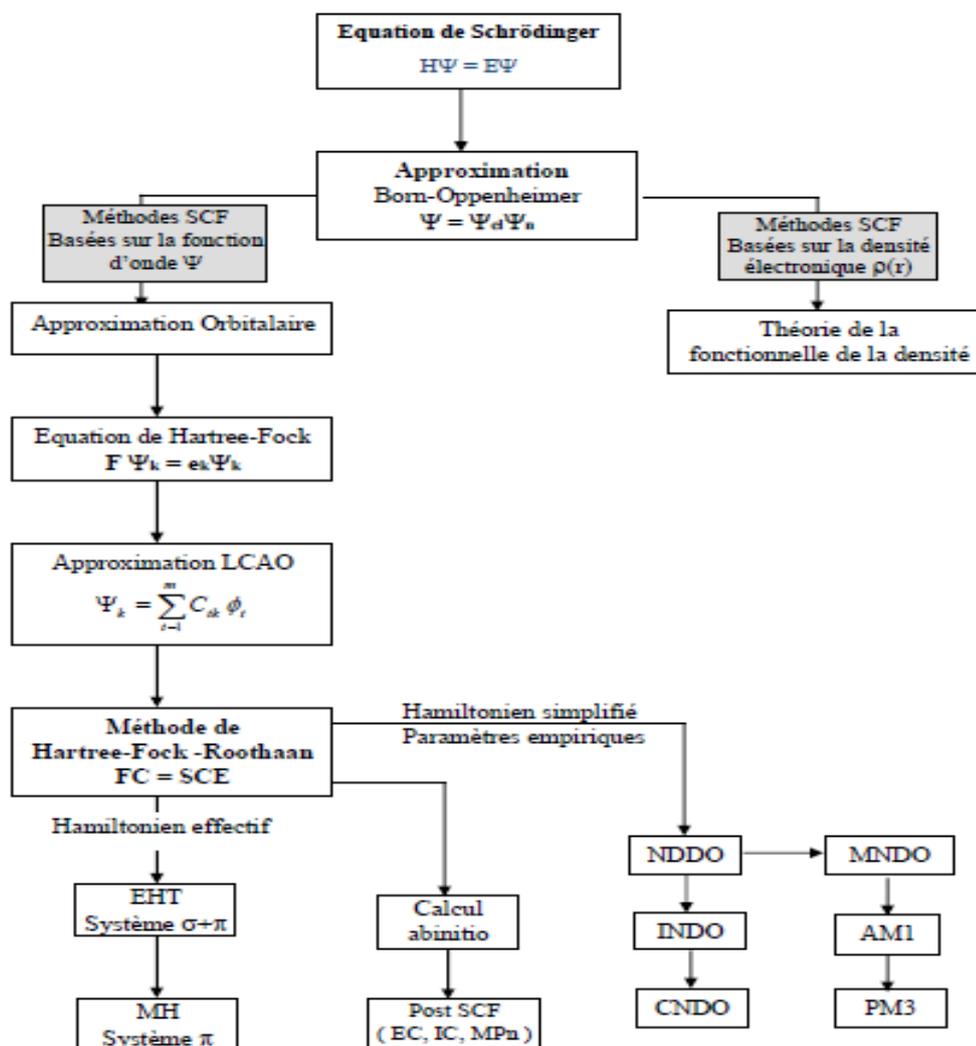


Figure. 1: Organigramme des méthodes utilisées en chimie quantique.

I.A.4.2. Mécanique moléculaire

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

La mécanique moléculaire est apparue en 1930 [20], mais s'est développée à partir des années soixante, quand les ordinateurs furent plus accessibles et plus performant.

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

Les électrons n'y sont pas traités explicitement comme dans un calcul de mécanique quantique, mais les atomes y sont représentés par des masses ponctuelles chargées reliées les unes aux autres par des ressorts auxquelles sont associés des paramètres dérivant de l'expérience ou de calculs précis quanta-chimiques. Le « champ de force » établi par cette méthode représente aussi bien que possible les variations de l'énergie potentielle avec la géométrie moléculaire. L'énergie de la molécule est exprimée sous la forme d'une somme de contributions associées aux écarts de la structure par rapport à des paramètres structuraux de référence :

$$E = E_{\text{Liaison}} + E_{\text{Angle}} + E_{\text{Dièdre}} + E_{\text{Van der Waals}} + E_{\text{Electrostatique}}$$

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

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

1.A.4.2.1. Différents champs de force en mécanique moléculaire

Différents champs de force sont proposés dans la littérature, ils se distinguent les uns des autres par les termes dans le développement de l'expression de l'énergie de la molécule. Chacun a un domaine d'application spécifique de sorte que le choix d'un champ de force dépend des propriétés et de l'application du système que l'on veut étudier. Les champs de forces en MM peuvent être groupés en trois classes principales [21].

- Champs de force contenant uniquement les termes harmoniques.
- Champs de force utilisant les termes d'ordre supérieur (cubique, quadratique,...).
- Champs de force suggérés par *Allinger et col.* [22] ne considérant pas que les termes de la mécanique moléculaire classique mais aussi les effets chimiques comme l'électronégativité.

MM2 : Il a été développé par Allinger en 1976 et c'est le champ de force le plus utilisé par la communauté des chimistes organiciens [23]. Il a été conçu au début pour les molécules simples (alcanes, alcènes et alcynes non conjugués, les composés carbonylés, les sulfures, les amines...), mais ses version améliorées MM3 (1989) [24] et MM4 (1996) [25] lui permet de traiter des molécules organiques de plus en plus complexes.

MM+: Est une extension du champ de force MM2, avec l'ajout de quelques paramètres additionnels [26]. MM+ est un champ de force robuste, il a l'aptitude de prendre en considération les paramètres négligés dans d'autres champs de force et peut donc s'appliquer pour des molécules plus complexes tels que les composés inorganiques [27].

AMBER: (*Assisted Model Building with Energy Refinement*), ce champ de force a été paramétrisé pour les protéines et les acides nucléiques. Il a été utilisé pour les polymères et pour d'autres petites molécules [28].

OPLS: (*Optimised Potentials for Liquid Simulation*), comme l'indique son nom, est conçu pour optimiser le potentiel qui permet la description des propriétés de solvation [29,30].

GROMOS : (*Groningen Molecular Simulation Program Package*), est conçu spécialement pour les biomolécules en milieu aqueux en vue de l'étude des interactions entre les molécules d'eau et les groupements polaires des protéines [31].

CHARMM (Bio+): (*Chemistry Harvard Macromolecular Mechanic*), Son concept est semblable à celui d'AMBER [32].

MMFF94: C'est l'un des champs de force de mécanique moléculaire les plus récents, il a été développé par Tom Halgren [33,34]. Les termes d'énergie d'élongation, de valence et de torsion sont les même que ceux de MM3 (expansion de MM2). MMFF94 comprend des termes croisés de liaison-valence [35].

1.A.4.2.2. Principes de la minimisation

La mécanique moléculaire nous permet de minimiser l'énergie calculée [36]. Le rôle de la minimisation est essentiellement de relaxer la structure initiale et d'éliminer les mauvais contacts interatomiques.

Le principe de base de la minimisation est qu'à partir d'une géométrie très approximative, on recherche le jeu de coordonnées cartésiennes qui réduit à son minimum la somme de toute les

contributions énergétiques dues aux déformations des coordonnées internes et aux interactions entre atomes non liés. En principe, il suffit de prendre la dérivée première de l'énergie stérique par rapport à chacun des degrés de liberté de la molécule et de trouver l'endroit sur l'hypersurface énergétique où pour chaque coordonnée r_i : $\frac{\partial E}{\partial r_i} = 0$

Les différentes méthodes de minimisation les plus utilisées sont :

- La méthode de la plus grande pente appelée «Steepest descent».
- La méthode du gradient conjuguée (Polak-Ribiere). (Cette méthode que nous avons utilisée)
- La méthode de Newton Raphson.

1.A.4.3. Dynamique moléculaire

Les premiers pas de la dynamique moléculaire n'ont pu se faire que grâce à l'arrivée des premiers ordinateurs (1957). Mais les premières réelles simulations ont été faites par Rahman, grâce à ses travaux sur la simulation de l'argon liquide en 1964 avec un temps de simulation de 10-11 s, puis de l'eau liquide [37] en 1971.

La mécanique moléculaire permet de calculer la position des atomes et des molécules dans leurs états énergétiques minima, à 0 °K c'est à dire sans la moindre vibration ni le moindre mouvement.

La dynamique moléculaire tente à simuler le mouvement des atomes et des molécules au cours du temps pour une température supérieure à 0 °K. Elle consiste à étudier la trajectoire d'une molécule en appliquant les lois de la mécanique classique Newtonienne et elle permet de simuler les mouvements intramoléculaires que l'on peut visualiser ensuite en temps réel. Ces mouvements correspondent à des vibrations autour d'un minimum, ou au passage d'un minimum à un autre minimum d'énergie [37].

Partie B: QSAR

1.B.1. Introduction

Les premiers essais de modélisation des relations structure-activité ont commencé à la fin du 19ème siècle, lorsque *Crum-Brown* et *Frazer* ont postulé que l'activité biologique d'une molécule est une fonction de sa constitution chimique [38-40]. 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 [41].

Aujourd'hui, l'utilisation de QSAR n'a cessé de progresser [42, 43]. 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 [44, 45].

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 [46].

1.B.2. Principe de QSAR

Le principe des méthodes QSAR est, comme leur nom l'indique, de mettre en place une relation mathématique reliant de manière quantitative la structure moléculaire, codée par des propriétés moléculaires appelées descripteurs, avec une activité en utilisant des méthodes d'analyse de données [47]. Par ces relations nous pouvons développer des modèles prédictifs de la forme générale suivante:

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

L'objectif de ces méthodes est alors d'analyser les données structurales afin de déterminer les facteurs influençant l'activité mesurée [48].

En pratique, le développement d'un modèle commence par la collecte de données expérimentales fiables et en nombre important, ensuite il est nécessaire de générer un nombre de descripteurs, caractérisant les structures moléculaires, utilisés dans le développement du modèle QSAR [49]. Une fois le modèle construit, il doit être évalué par des méthodes de validation (validations interne et externe) afin d'estimer sa robustesse et son pouvoir prédictif [50-52]. En fin, pour tout modèle, il est important de connaître quel type de molécules ce modèle est utilisable (connaître le domaine d'applicabilité) [53].

Il y a plusieurs raisons pratiques qui justifient l'utilisation des études QSAR [44].

- Pour prédire les propriétés de l'activité biologique et physico-chimiques par des moyens rationnels.
- Economie sur le coût de développement de produits (exemple de l'industrie pharmaceutique, des pesticides, des détergents,...etc.).
- Les prévisions pourraient réduire l'exigence de tests longs et coûteux pratiqués sur les animaux.

1.B.3. 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 [54]:

- 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 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 des effets conjugués de molécules, que ce soit dans des mélanges ou des formulations.

1.B.4. Outils et Méthodologie de QSAR

1.B.4.1. Paramètres biologiques

Les modèles QSAR sont dépendants des données expérimentales utilisées pour leur construction. Le modélisateur doit tenir compte des données à modéliser. Le choix de la base de données est donc une étape très importante dans le développement des modèles QSAR.

Ces données devraient, idéalement, être de grande qualité, ce qui signifie qu'elles devraient être fiables et cohérentes. Il est donc important de les choisir parmi celles présentant des incertitudes faibles afin de limiter les barres d'erreurs expérimentales. De plus, le modélisateur doit s'assurer que les données expérimentales utilisées ont été obtenues selon le même protocole. En effet les conditions expérimentales ont, généralement, une forte influence sur les valeurs obtenues [55].

Il faut également que la distribution des données soit la plus homogène et normale que possible, car la plupart des méthodes statistiques sont basées sur ce type de distribution [56].

L'efficacité d'un modèle QSAR dépend également du type de molécules qui y sont incluses, plus le modèle présentera des composés de structures proches et similaires, plus il aura de chance d'être performant.

Les données biologiques sont généralement exprimées en logarithmes inverses ($\log 1/C$) afin d'obtenir des valeurs mathématiques plus élevées lorsque les structures sont biologiquement très efficaces [57, 58].

1.B.4.2. Utilisation des descripteurs moléculaires

L'activité biologique ne peut pas être reliée directement à la structure moléculaire ; cette dernière est codée par des grandeurs qui représentent d'une manière quantitative les informations contenues dans la structure moléculaire telles que les caractéristiques physico-chimiques et structurelles. Ces grandeurs sont appelées descripteurs [59]. Une fois ces derniers disponibles, il est possible d'établir des relations entre ceux-ci et une activité à l'aide des outils de la modélisation [60]. On en dénombre aujourd'hui des milliers, qui peuvent être calculés ou obtenus de manière empirique. De nombreux logiciels ont été développés pour calculer les différents descripteurs moléculaires tels que :

Gaussian, ChemOffice, ChemSketch, Marvin Sketch, Dragon, QSARIS.

Dans ce qui suit, nous allons présenter que les descripteurs moléculaires qui ont été utilisés dans l'ensemble de nos travaux.

⇒ Les descripteurs 1-D

Ces descripteurs sont calculés à partir de la formule brute de la molécule à l'aide de la composition moléculaire, c'est-à-dire les atomes qui la constituent, et ils représentent des propriétés générales telles que : les pourcentages massiques des atomes, la masse molaire, le poids moléculaire...

Dans nos travaux nous avons utilisé :

Le poids moléculaire, noté MW (appelé aussi le poids de formule), mesuré en daltons (Da). C'est la somme des poids atomiques des différents atomes constituant la molécule. Il est utilisé dans l'étude de transport dont la diffusion et le mode de fonctionnement. Les composés avec des poids plus élevés sont moins susceptibles d'être absorbés et donc ne peuvent pas atteindre le site d'action. Ainsi, essayer de garder des poids moléculaires aussi bas que possible devrait être l'objectif pour établir un médicament. Pour les médicaments délivrés par voie orale le poids moléculaire doit être inférieur ou égal à 500 daltons (optimum autour de 300 daltons) [61].

⇒ *Les descripteurs 2-D*

Les descripteurs 2D sont obtenus à partir de la structure plane de la molécule. Dans cette catégorie on trouve principalement les descripteurs topologiques.

Les descripteurs topologiques (ou indices topologiques) décrivent les connectivités atomiques dans la molécule. Ce sont des descripteurs plus "sophistiqués" qui n'ont pas forcément un sens chimique évident mais ils contiennent en leur sein des informations sur la taille globale du système, sa forme globale et ses ramifications. Le principe est de trouver une valeur différente pour chaque squelette moléculaire.

La superficie de la surface polaire, notée (PSA), en (Å²), est un paramètre très utile pour la prédiction des propriétés du transport des médicaments. Elle est définie comme la somme des surfaces des atomes polaires (habituellement, l'oxygène, l'azote, le soufre, le chlore et l'hydrogène ci-jointes) dans une molécule.

⇒ *Les descripteurs 3-D*

Ce type de descripteurs nécessitent une conformation 3D de la molécule ; Ils sont évalués à partir des positions relatives de leurs atomes dans l'espace et décrivent des caractéristiques plus complexes ; leurs calculs nécessitent donc de connaître, le plus souvent par « modélisation moléculaire empirique » ou « ab-initio », la géométrie 3D de la molécule. La plupart de ces descripteurs s'avèrent relativement coûteux en temps de calcul, mais apportent davantage d'informations et sont nécessaires à la modélisation de propriétés ou d'activités qui dépendent de la structure 3D. On distingue plusieurs familles de descripteurs 3D:

Les descripteurs géométriques

Ces descripteurs peuvent être obtenus expérimentalement ou par modélisation moléculaire, empirique ou ab-initio. Ils sont basés sur l'arrangement spatial des atomes constituant la molécule et sont définis par les coordonnées des noyaux atomiques et la grosseur de la molécule représentée. Ces descripteurs incluent des informations sur la surface moléculaire obtenue par les aires de Van Der Waals et leur superposition. Les volumes moléculaires peuvent être obtenus par les volumes de Van Der Waals. Parmi ceux qui sont les plus importants, que nous avons utilisé dans nos travaux, on trouve:

Le volume moléculaire, noté MV, en cm³, est défini par la formule suivante :

$$MV = \frac{MW}{d}$$

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

Le nombre de liaisons rotatives, noté NROT, La liaison rotative est définie comme une liaison d'un composé non cyclique, associée à un atome non lourd (qui n'est pas l'hydrogène). Les liaisons CN (amide) ne sont pas considérées en raison de leur barrière d'énergie de rotation élevée. Le nombre de liaisons rotatives est utilisé pour identifier la flexibilité de la molécule, il a été montré pour être un descripteur de très bonne biodisponibilité orale de médicaments, et pour qu'une structure chimique puisse présenter de bons effets inhibiteurs et être similaire aux médicaments, selon la règle de Lipinski, il faut que le nombre de liaisons rotatives soit inférieur ou égal à 5 [62].

Les descripteurs physico-chimiques

Les descripteurs physicochimiques, (ou indices physicochimiques) certains d'entre eux reflètent la composition moléculaire du composé (le nombre et le type d'atomes et de liaisons présents dans la molécule, le nombre de cycle, les propriétés donneur/accepteur de liaison H, cation, anion, etc...). D'autres représentent le caractère hydrophile ou lipophile de la molécule généralement évalué à partir du coefficient de partage Octanol/eau représenté par le log P [62]. Parmi ceux que nous avons utilisés dans nos travaux, on trouve:

Le coefficient de partage Octanol/Eau: Le transport, le passage à travers les membranes et l'activité pharmacologique d'une molécule peuvent être conditionnés par son partage entre une phase lipidique et une phase aqueuse, c'est-à-dire son caractère hydrophile. Celui-ci peut être quantifié par le coefficient de partage Octanol-Eau, noté (log P), qui mesure la solubilité différentielle d'un soluté dans ces deux solvants non miscibles. C'est une mesure importante pour l'identification de la similarité médicamenteuse, selon la règle de Lipinski, les médicaments délivrés par voie orale doivent avoir des valeurs de log P supérieures ou égales à -2 et inférieures ou égales à 5) [62]. Il est défini par la formule suivante:

$$\log P = \frac{[\text{Octanol}]}{[\text{H}_2\text{O}]}$$

[Octanol] et [H₂O] sont les concentrations du soluté dans l'Octanol et l'eau.

Les composés qui ont les valeurs de log P > 0 sont dites lipophiles, et les composés qui ont les valeurs de log P < 0 sont dites hydrophiles. Si le Log P est positif et très élevé, cela

exprime le fait que la molécule est plus soluble dans l'Octanol que dans l'eau, ce qui reflète son caractère lipophile, et inversement, si le Log P est négatif cela signifie que la molécule est hydrophile.

Un Log P nul signifie que la molécule est aussi soluble dans un solvant que dans l'autre.

La réfractivité moléculaire, notée (MR), en m³/mol, est le volume de la substance absorbée par mole de cette substance. Elle est définie par Lorentz-Lorenz par la formule suivante:

$$MR = \frac{n^2 - 1}{n^2 + 2} \frac{MW}{d} = \frac{n^2 - 1}{n^2 + 2} MV$$

Où : MW est le poids moléculaire ; d est la densité ; n est l'indice de réfraction ; MV est le volume molaire.

La réfractivité moléculaire est également proportionnelle à la polarisabilité α_e , par la relation suivante [41] : $MR = \frac{4}{3} \pi N A \alpha_e$

Où : NA est le nombre d'Avogadro qui est, le nombre de molécules dans une mole de substance, NA = 6.022 10²³

La polarisabilité, notée (α_e), en (m³), est l'aptitude à la déformation du nuage électronique de la molécule sous l'influence d'un champ électrique uniforme. C'est l'un des paramètres qui traduisent les propriétés moléculaires liées à l'hydrophobie et par conséquent aux activités biologiques. Elle est calculée à partir de la réfractivité molaire ou du volume molaire comme suit :

$$\alpha_e = 0,3964308 \times MR = 0,3964308 \times \frac{n^2 - 1}{n^2 + 2} MV$$

Le nombre de donneurs de liaisons hydrogène, noté (NHD), calcule le nombre de donneurs de liaison hydrogène dans la molécule. Il s'agit du nombre d'atomes possédant une case quantique vide et contenant un hydrogène acide, c'est-à-dire un atome d'hydrogène lié à un hétéroatome (comme dans les amines, alcools, thiols).

Le nombre d'accepteurs de liaisons hydrogène, noté (NHA), calcule le nombre d'accepteurs de liaison hydrogène dans la molécule. Il s'agit du nombre d'atomes possédant des doublets non liants (azote, oxygène ou fluor) et capable de se lier par liaisons hydrogène à d'autres molécules.

Selon la « règle des cinq de Lipinski », lors de l'identification de la similarité médicamenteuse, les médicaments délivrés par voie orale doivent avoir un nombre d'accepteurs de liaisons hydrogène (NHA) inférieur ou égal à 10 (optimum d'environ 5) et un nombre de donneurs de liaisons hydrogène (NHD) inférieur ou égal à 5 (optimum de 2).

Les descripteurs quantiques/électroniques

Ces descripteurs caractérisent la distribution de charge des molécules (polarité des molécules) mais aussi les paramètres de la chimie quantique qui, pour être calculés de manière fiable, nécessitent des calculs plus sophistiqués.

Les approches de la chimie quantique nous donnent accès à des informations supplémentaires telles que des données structurales, énergétiques, électroniques et spectroscopiques des systèmes étudiés.

Parmi ces propriétés, que nous avons utilisées dans nos travaux, on trouve:

L'énergie totale : Pour une molécule isolée à l'état fondamental, l'énergie totale calculée, notée E_t , mesurée en eV, peut être utilisée comme descripteur moléculaire quantique. Cette énergie approximative a été calculée pour une conformation optimisée de la géométrie la plus stable dont la structure d'énergie est minimale.

Le moment dipolaire, noté μ , mesuré en debye (D), mesure la polarité nette moléculaire, et décrit la séparation de charge dans une molécule où la densité d'électrons est partagée inégalement entre les atomes. L'existence d'un moment dipolaire dans une molécule a son origine dans la différence d'électronégativité entre les atomes. La densité électronique est plus élevée au voisinage de l'atome le plus électronégatif. Ceci entraîne une dissymétrie dans la répartition des électrons de liaison. Ainsi, plus le moment dipolaire d'une molécule est élevé, plus la dissymétrie dans la molécule est importante.

Les énergies des orbitales frontières, jouent un rôle majeur dans de nombreuses réactions chimiques et dans les mécanismes réactionnels. Les énergies de ces orbitales sont des paramètres très populaires dans la chimie quantique et dans les études RQSA:

L'énergie HOMO, notée E_{HOMO} , mesurée en eV, est le niveau d'énergie le plus élevé dans la molécule qui contient des électrons, il est directement lié au potentiel d'ionisation.

Lorsqu'une molécule agit comme une base de Lewis (un doublet d'électrons donneur) dans la formation d'une liaison, les électrons sont alimentés à partir de cette orbite. Il mesure la nucléophilie d'une molécule et caractérise la susceptibilité de la molécule à l'attaque par des électrophiles.

L'énergie LUMO, notée E_{LUMO} , mesurée en eV, est le niveau d'énergie le plus bas dans la molécule qui ne contient pas d'électrons, il est directement lié à l'affinité d'électron. Lorsqu'une molécule agit comme un acide de Lewis (un doublet d'électrons accepteur) dans la formation de liaisons, des doublets d'électrons entrants sont reçus dans cette orbite. Il mesure l'électrophilicité d'une molécule et caractérise la susceptibilité de la molécule à l'attaque par les nucléophiles.

Le Gap énergétique, ou l'écart HOMO-LUMO, noté E_{gap} , mesuré en eV, traduit l'énergie comprise entre l'orbitale moléculaire la plus haute occupée et la plus basse vacante, C'est un indice de stabilité important. Cette différence d'énergie sert de mesure de l'excitabilité d'une molécule. Ainsi, plus l'intervalle d'énergie sera faible et plus la molécule pourra interagir avec l'environnement.

Un grand écart HOMO-LUMO implique une grande stabilité pour la molécule dans le sens de sa faible réactivité dans les réactions chimiques, et de même, un faible écart implique une grande réactivité de la molécule. L'écart HOMO-LUMO a également été utilisé comme une approximation de la plus faible énergie d'excitation de la molécule.

Descripteurs de charge, noté q_i , Ce sont des descripteurs électroniques définis en termes de charges atomiques utilisées pour décrire les aspects électroniques aussi bien de la molécule entière que de régions particulières, comme les atomes, les liaisons, les fragments moléculaires et les orbitales. Les charges renseignent sur l'importance de la localisation de la densité électronique dans une molécule: des valeurs q_i négatives signifient la localisation d'un excès de charge électronique au centre i , alors que des valeurs positives indiquent que le centre i est électroniquement déficitaire.

1.B.4.2.1. Sélection et réduction du nombre de descripteurs

Un grand nombre de descripteurs différents sont collectés pour la modélisation d'une grandeur donnée (activité), car les facteurs déterminants du processus étudié ne sont a priori

pas connus. Cependant, les descripteurs envisagés n'ont pas tous une influence significative sur la grandeur modélisée, et les variables ne sont pas toujours indépendantes. De plus, le nombre de descripteurs, c'est-à-dire la dimension de la base de données d'entrée, détermine la dimension du vecteur des paramètres à ajuster. Si cette dimension est trop importante par rapport au nombre des observations (molécules) de la base d'apprentissage, le modèle risque d'être sur-ajusté à ces exemples, incapable de prédire la grandeur modélisée sur de nouvelles molécules et peut contenir des informations redondantes. Les descripteurs moléculaires employés doivent être porteurs de sens et interprétables d'un point de vue chimique. Et par conséquent, lorsque les descripteurs sélectionnés sont pertinents, ils offrent des idées sur les mécanismes, et les modèles RQSA seront simples, transparents et compréhensibles [63].

Avant d'entamer le développement effectif des équations de régression RQSA, il est hautement recommandé d'examiner la qualité statistique des données de départ, à la fois les données à corrélérer (variable dépendante) et les descripteurs utilisés dans la corrélation (variables indépendantes).

Pour un ensemble de descripteurs différents, il est nécessaire d'effectuer une analyse des données bivariée, c'est-à-dire de calculer le coefficient de corrélation entre chacune des paires de l'ensemble des descripteurs. Si ce coefficient est statistiquement significatif ($R > 0,95$), ces deux descripteurs sont considérés comme fortement corrélés et ne peuvent être utilisés simultanément lors de l'analyse RQSA [64] et en pratique, ils seront alors enlevés dans le procédé de sélection. Ce type d'analyse est appelé l'analyse objective qui permet de réduire le nombre de descripteurs sans faire participer la variable dépendante (l'activité).

Finalement, pour que les relations RQSA ne soient pas statistiquement non significatives ou en cas d'erreur ponctuelles, il faut que le rapport composés/descripteurs doive être supérieur à 5 [65, 66].

Partie C: Méthodes statistiques

I.C.1. Définition

Par définition, la statistique est « la science dont l'objet est de recueillir, de traiter et d'analyser des données issues de l'observation de phénomènes dans lesquels le hasard intervient (phénomène aléatoire) ». Par conséquent, l'objectif principal de la statistique est de *Approches de* maîtriser au mieux l'incertitude pour extraire des informations utiles des données, par l'intermédiaire de l'analyse des variations dans les observations. En outre, l'analyse des données est utilisée pour décrire, comprendre et gérer les phénomènes étudiés, faire des prévisions et prendre des décisions [61].

I.C.2. Méthodes statistiques

Faire de la statistique suppose que l'on étudie un ensemble d'objets équivalents sur lesquels on observe des caractéristiques appelées « variables ». Dans notre cas, les objets (ou individus) sont les molécules et les variables sont les descripteurs moléculaires précédemment décrits dans ce chapitre.

Après le recueil des descripteurs, la démarche statistique consiste à traiter et interpréter les informations recueillies sur ces molécules. Cette démarche comporte deux grandes classes : la statistique descriptive et la statistique décisionnelle ou prédictive.

I.C.2.1. La statistique descriptive

La statistique descriptive (appelée aussi l'analyse des données), a pour but d'extraire le maximum de l'information contenue dans les données d'une façon efficace, simple et compréhensible. Elle permet de résumer les caractéristiques essentielles du phénomène étudié et de suggérer des hypothèses pour des études plus sophistiquées. Elle utilise pour cela des représentations de données sous forme de graphiques, de tableaux et d'indicateurs statistiques. Elle est utilisée aussi pour diviser et classer les données dans des classes homogènes.

- L'analyse en composantes principales (ACP) comme technique pour l'analyse des données
- La méthode du partitionnement en k-moyennes (ou k-means en anglais)
- La classification ascendante hiérarchique (CAH) pour la classification des données.

I.C.2.2. La statistique décisionnelle ou prédictive

Contrairement à la statistique descriptive, dans ce type de statistiques les probabilités jouent un rôle fondamental. Cette statistique a pour but de prendre des décisions et de faire des prévisions au vu des observations. En général, il faut pour cela proposer des modèles probabilistes du phénomène aléatoire étudié et savoir gérer les risques d'erreurs. Dans notre cas, il faut rechercher une relation approximative entre une activité ou propriété et plusieurs variables quantitatives (descripteurs moléculaires), la forme de cette relation peut être linéaire ou non linéaire.

Dans l'ensemble de nos travaux, nous avons utilisé la régression linéaire multiple MLR pour la construction des modèles RQSA/RQSP.

I.C.3. La régression linéaire multiple

La régression linéaire multiple MLR est l'une des méthodes de modélisation les plus populaires grâce à sa simplicité d'utilisation et facilité d'interprétation. L'avantage important de la régression linéaire multiple est qu'elle est très transparente, puisque l'algorithme est disponible, et que les prédictions peuvent être réalisées facilement. Cette méthode a été utilisée aussi pour la sélection des descripteurs moléculaires utilisés dans les autres méthodes statistiques [67]. La méthode MLR se base sur l'hypothèse que la propriété y dépend linéairement des différentes variables (les descripteurs) $x_1, x_2, x_3, \dots, x_i$, selon la relation:

$$Y = a_0 + \sum_{i=0}^n a_i X_i$$

Avec : y est la variable dépendante (à expliquer ou à prédire).

X_i : sont les variables indépendantes (explicatives).

n : est le nombre de variables explicatives.

a_0 : est la constante de l'équation du modèle.

a_i : sont les coefficients de descripteurs dans l'équation du modèle.

La taille de ces coefficients indique le degré d'influence des descripteurs moléculaires correspondants sur l'activité cible. Un coefficient positif indique que le descripteur moléculaire correspondant contribue positivement à la cible, tandis qu'un coefficient négatif suggère la contribution négative.

La méthode de progression que nous avons utilisée c'est la progression ascendante.

La MLR progressive ascendante, qui consiste à incorporer les variables au modèle une à une, en sélectionnant, à chaque étape, la variable dont la corrélation partielle avec la grandeur modélisée est la plus élevée.

I.C.4. Techniques de validation**I.C.4.1. Coefficients et tests statistiques standards**

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

I.C.4.1.1. Coefficient de corrélation r (et coefficient de détermination r^2)

C'est l'indicateur statistique le plus répandu est le coefficient de corrélation qui évalue la part de la variance de l'activité / la propriété cible expliquée par le modèle.

$$r = \sqrt{1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2}}$$

Avec : r : est le coefficient de corrélation.

y_i , \hat{y}_i : sont, respectivement, les valeurs observées et calculées de la variable dépendante.

\bar{y}_i : est la valeur moyenne des valeurs observées.

Ces coefficients ne sont pas affectés par l'unité de mesure choisie et traduisent une bonne corrélation entre l'activité cible et l'activité initiale si r^2 est proche de 1 (cas idéal).

Le jugement sur la valeur de r ou r^2 est très subjectif. Bien que ce coefficient soit très facile à comprendre, il faut se garder d'y attacher trop d'importance car il est loin de fournir un critère suffisant pour juger la qualité d'une régression. Il n'est pas recommandé d'utiliser r^2 pour comparer des modèles avec un nombre différent de descripteurs, le coefficient r^2 nous dira toujours de choisir le modèle avec le plus grand nombre de descripteurs car son r^2 sera plus important (on projette sur un espace plus grand), même si les variables sont sans effets sur la réponse (l'activité étudiée).

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 ajusté (r^2_{adj}). Bien entendu, un autre indicateur est l'erreur quadratique moyenne (MSE, pour Mean Square Error), à laquelle est parfois préférée la déviation standard s .

1.C.4.1.2. Le coefficient de détermination ajusté r^2_{adj}

Ce coefficient est tient compte du degré de liberté :

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

Avec : n : est le nombre des observations (les molécules).

p : est le nombre de variables indépendantes (les descripteurs).

r^2 : est le coefficient de détermination du modèle.

1.C.4.1.3. L'erreur type résiduel « s » (La déviation standard)

$$s = \sqrt{\frac{\sum(\hat{Y}_i - Y_i)^2}{n - p - 1}}$$

Avec:

y_i, \hat{y}_i : sont, respectivement, les valeurs observées et calculées de la variable dépendante.

n : est le nombre des observations (les molécules).

p : est le nombre de variables indépendantes (les descripteurs).

Ces paramètres mesurent la variation de l'activité cible non expliquée par le modèle RQSA. En particulier, plus la déviation standard est petite et plus la corrélation est meilleure. Sa valeur est toujours fonction de l'unité de mesure de l'activité cible et tient également compte des erreurs expérimentales ce qui explique qu'une valeur trop petite n'ait aucune signification.

I.C.4.1.4. Le test de Fisher F

L'indice de Fisher F-test est employé afin de mesurer le niveau de signification statistique du modèle à « x% » (le niveau usuel est 95%), c'est-à-dire la qualité du choix du jeu de paramètres [61].

Hypothèses :

H₀: les variances des échantillons sont homogènes.

H₁: les variances des échantillons ne sont pas homogènes.

La valeur à calculer est:

On calcule le F (observé) à partir de la formule :

$$F(\text{observée}) = \frac{\sum(\hat{Y}_i - \bar{Y}_i)^2}{\sum(Y_i - \hat{Y}_i)^2} \frac{n - p - 1}{p}$$

Avec: F est l'indice de Fisher.

y_i, \hat{y}_i : sont, respectivement, les valeurs observées et calculées de la variable dépendante.

n: est le nombre des observations (les molécules).

p: est le nombre de variables indépendantes (les descripteurs).

\bar{y}_i : est la valeur moyenne des valeurs prédites.

Après le calcul de F (observé) on le compare avec le F théorique obtenu à partir des tables statistiques usuelles (la table de Fisher).

- Si F observé est plus grand que le F théorique : refus de l'hypothèse nulle H₀ et cela signifie que les variances des échantillons sont trop différentes pour être considérées comme homogènes.
- Si F observé est plus petit que le F théorique : acceptation de l'hypothèse nulle H₁ et cela signifie que les deux variances ont des valeurs suffisamment proches pour qu'on accepte l'idée qu'elles soient homogènes.

I.C.4.2. Pouvoir de prévision**I.C.4.2.1. Interne**

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) 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 : $PRESS$, SSY , S_{press} , R^2_{CV} et R^2_{adj} ; ces paramètres sont calculés à partir des équations suivantes [68] :

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

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

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

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

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

Avec: n : nombre de molécules de la série étudiée.

$PRESS$: Somme des carrés des écarts entre les valeurs observées et les valeurs prévues à partir du modèle.

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

S_{PRESS} : La capacité prédictive des modèles est évaluée par l'erreur de racine-carrée-moyenne.

R^2_{adj} : Estimation de la qualité d'ajustement du modèle à la population.

R^2_{CV} : Est la variation de la statistique du R^2 obtenue en ajoutant ou en enlevant une variable indépendante.

I.C.4.2.2. Externe

Cette méthode consiste à prédire l'activité d'une série de molécules appelée généralement série de test qui ne sont pas dans la série de développement du modèle, cette validation est caractérisée par le paramètre R^2_{test} . Récemment plusieurs études [69,70] ont montré l'insuffisance des paramètres R^2 , R^2_{CV} pour vérifier le pouvoir prédictif des modèles RQSA. Par conséquent, d'autres paramètres doivent être vérifiés pour cet objectif. Ces paramètres sont connus sous le nom «critères de validation externe» ou souvent appelés «critères de Trophsa » [69].

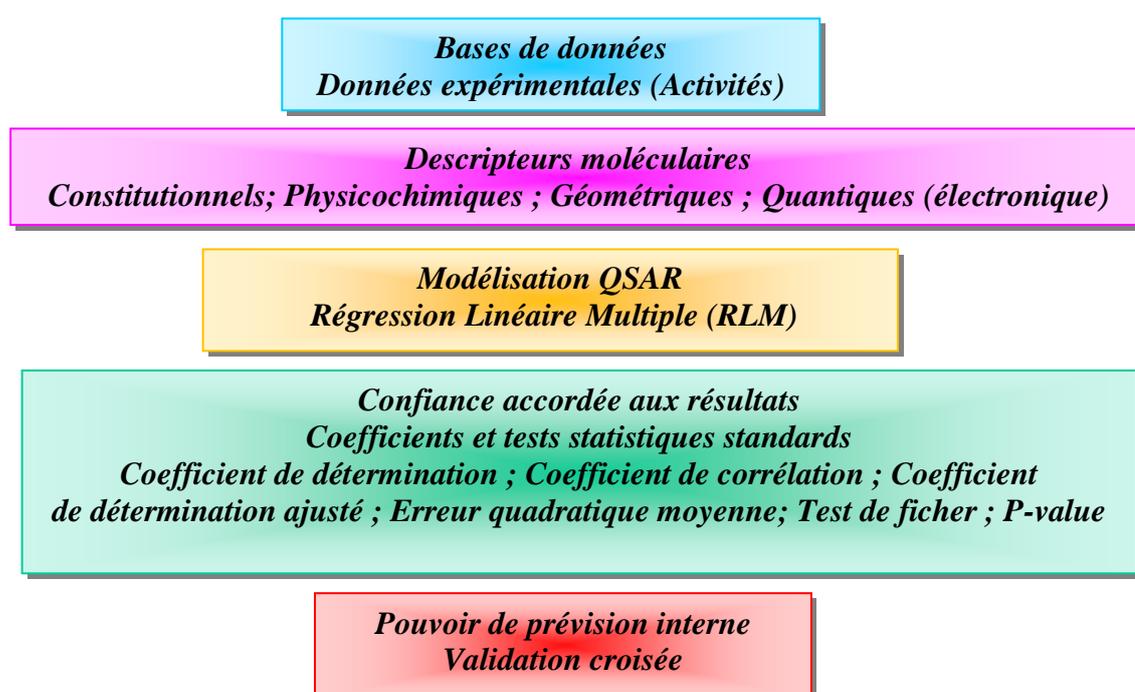


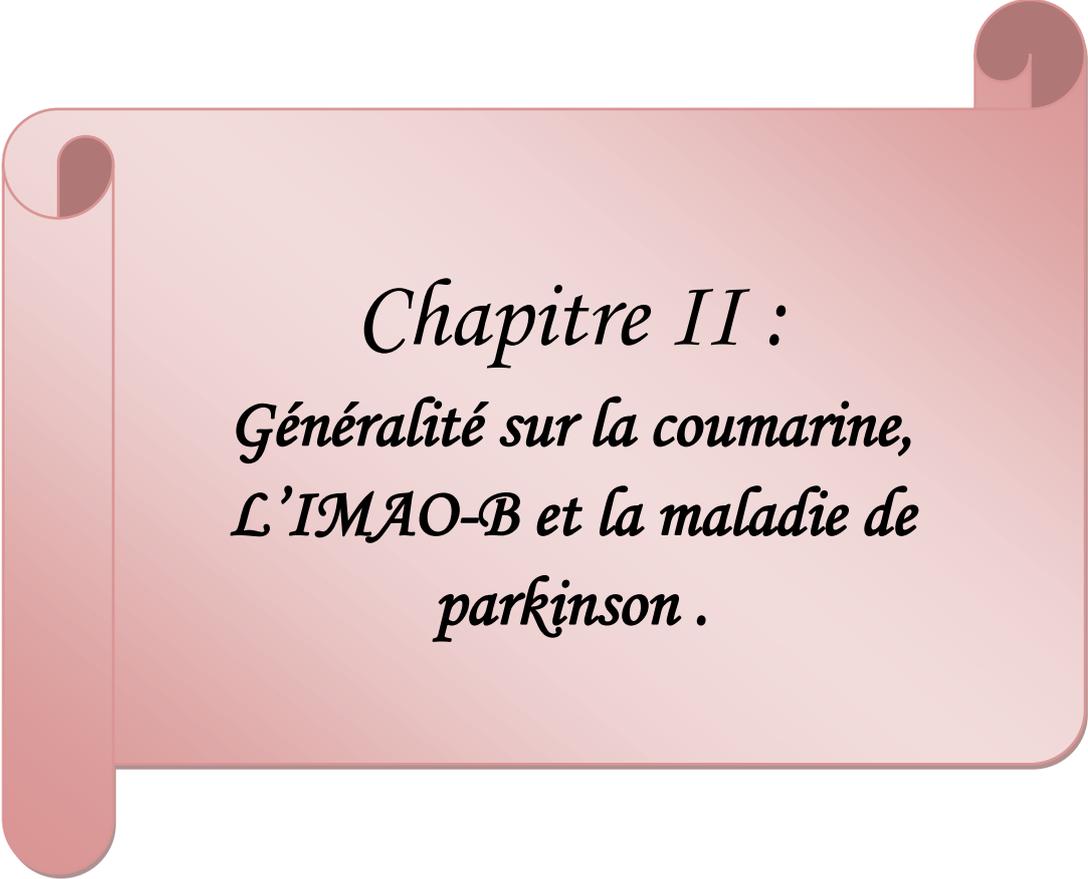
Figure .2 : Schéma de la méthodologie utilisée dans nos travaux.

- [01] E. Derety, *J. Mol. Structr. (Theochem)*, 1999, 459, 273.
- [02] S. Belaidi, M. Omari, T. Lanez et A. Dibi, *J. Soc. Alger. Chim.*, 2004, 14, 27.
- [03] S. K. Niazi, *Preformulation Handbook of Chemical, Pharmaceutical Scientist Inc., USA*, 2007.
- [04] M. Born, R. Oppenheimer, *Ann. Phys.*, 1927, 84, 457.
- [05] J. L. Rivail, *Elements de chimie quantique à l'usage des chimistes*, CNRS Editions, Paris, 1999.
- [06] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, 107, 3902-3909.
- [07] J. J. P. Stewart, *J. Comput. Chem.*, 1989, 10, 209-220.
- [08] G. Boucekkine, *Méthodes de la chimie quantique, Techn. Ing.*, 2007.
- [09] N. Merabet, *Thèse de Doctorat: Université 8 mai 1945 Guelma*, 2017.
- [10] E. Fermi, *Springer Handbook of Atomic, Molecular and Optical Physics, Rend. Accad* 6 (1927) 602.
- [11] P. Hohenberg, W. Kohn, *Phys. Rev A* 136 (1964) 864.
- [12] M. Rahim, *Thèse de Doctorat soutenue 2015 à l'université de Guelma*.
- [13] A.D. Becke, *Phys. Rev. A*, 38(1988)3098.
- [14] J.P. Perdew, Y. Wang, *Phys. Rev. B* 1986, 33, 8800, 1989, 403399
- [15] J.P. Perdew, Y. Wang, *Phys. Rev. B*.45 (1992) 13244.
- [16] Y. Zhang, W. Yang, *Phys. Rev. Lett.*80 (1998) 890.
- [17] J.P. Perdew, S. Kurth, A. Zupan, P. Blaha, *Phys. Rev. Lett.*82 (1999)2544.
- [18] R.S. Mulliken, *J. Chem. Phys.*23 (1955) 1833-1846.
- [19] W. J. Hehre ; Lathan, W. A.; Ditchfield, R.; Newton, M. D.; Pople, J. A. *Gaussian*.
- [20] L.P. Hammett, *J. Am. Chem. Soc.*, 1937, 59, 96.
- [21] J. C. Smith, M. Karplus, *J. Am. Chem. Soc*, 1992, 114, 805.
- [22] T. Shimanouchi, *Pure Appl. Chem*, 1963, 7, 131.
- [23] N.L. Allinger, *Q. C. P. E.*, 1983, 3, 32.
- [24] N.L. Allinger, Y.H. Yuh, J-H. Lii, *J. Am. Chem. Soc.*, 1989, 111, 8522.
- [25] N.L. Allinger, K. Chen, J-H. Lii, *J. Comp. Chem.*, 1996, 17, 642.
- [26] G.B. Jones, B.J. Chapman, *Synthesis*, 1995, 475.
- [27] A. Hocquet, M. Langgard, *J. Mol. Model.*, 1998, 4, 94.
- [28] S.J. Weiner, P.A. Kollman, T. Nguyent, D.A. Case, *J. Comput. Chem.*, 1986, 7, 230.
- [29] W.L. Jorgensen, J.T. Rives, *J. Am. Chem. Soc.*, 1988, 110, 1657.
- [30] J.C. Smith & M. Karplus, *J. Am. Chem. Soc.*, 1992, 114, 805.

- [31] W.F. Van Gusten, M. Karplus, *Macromolecules*, 1982, 15, 1528.
- [32] B.R. Brooks, R.E. Bruccoleri, B.D. Ofolson, D.J. States, S. Swaminathan, M. Karplus, *J. Comp. Chem.*, 1983, 4, 187.
- [33] F. Tristram, V. Durier, G. Vergoton, *J. Mol. Struct.*, 1996, 378, 249.
- [34] T.A. Halgren, *J. Comp. Chem.*, 1996, 17, 490.
- [35] T.A. Halgren, R.B. Nachbar, *J. Comp. Chem.*, 1996, 17, 587.
- [36] M. Baaden, *Outils pour étudier la structure et la dynamique des peptides et des protéines*, (Institut de Biologie Physico-chimique, PARIS, CNRS UPR9080, 2003, 36.
- [37] S.J. Weiner, P.A. Kollman, T. Nguyen, D.A. Cas. *An all atom force field for simulations of proteins and nucleic acids. J. Comput. Chem.*, 1986, 7, 230-252.
- [38] A. Crum Brown, T.R. Fraser, *Trans. Roy. Soc.* 25, 1868, 151-203.
- [39] Hansch, *Acc. Chem. Res.* 2, 1969, 232-239.
- [40] Hansch, T. Fujita, *J. Am. Chem. Soc.* 86, 1964, 1616-1626.
- [41] A. K. Debnath, *Mini Reviews in Medicinal Chemistry. I*, 2001, 187-195.
- [42] M. Grover, B. Singh, M. Bakshi, S. Singh, *Pharm. Sci. Tech. Today.* 3, 2000, 50-57.
- [43] M. Grover, B. Singh, M. Bakshi, S. Singh, *Pharm. Sci. Tech. Today.* 3, 2000, 28-35.
- [44] T. I. Oprea, "Chemoinformatics in Drug Discovery" Ed. WILEY-VCH Verlag. Allemagne, 2005.
- [45] E. A. Rezza ; P. N. Kourounakis "Chemistry and Molecular Aspects of Drug Design and Action" Ed. Taylor & Francis Group, LLC. Etats Unies, 2008.
- [46] A. K. Ghose, V. N. Viswanadhan, "Combinatorial Library Design and Evaluation: Principles, Software Tools, and Applications in Drug Discovery" Ed. Marcel Dekker, Inc. Etats Unies, 2001.
- [47] F. Li, H. Wu, L. Li, X. Li, J. Zhao, W. J. G. M. Peijnenburg, *Ecotoxicol. Environm. Saf.* 80, 2012, 273-279.
- [48] A. Srinivas Reddy, S. Kumar, R. Garg, *J Mol Graph Model.* 28(8), 2010, 852-862.
- [49] M. Zhao, Z. Li, L. Peng, Y. R. Tang, C. Wang, Z. Zhang, S. Peng, *Bioorg. Med. Chem.* 15, 2007, 2815-2826.
- [50] A. A. Toropov, A. P. Toropova, E. Benfenati, G. Gini, D. Leszczynska, J. Leszczynski, G. De Nucci, *Biochem. Biophys. Res. Comm.* 432(2), 2013, 214-225.
- [51] A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* 20, 2002, 269-276.
- [52] A. Tropsha, P. Gramatica, K.V. Gombar, *QSAR Comb. Sci.* 22, 2003, 69-77.
- [53] P. Gramatica, E. Giani, E. Papa, *J. Mol. Graph. Mod.* 25, 2007, 755-766.

- [54] T. Ziani, *Prédiction de l'effet inhibiteur des dérivés d'imidazole sur l'enzyme Cox-2*, Université Med Khider Biskra, 2012-2013.
- [55] D. Fourches, E. Muratov, A. Tropsha, *J. Chem. Inf. Model.* 50, 2010, 1189–1204
- [56] A. Tropsha, *Mol Inform.* 29, 2010, 476–488.
- [57] S. Chatterjee, A. S. Hadi, B. Price, « *Regression analysis by example* » Wiley-Interscience, New York. 2000.
- [58] H. T. N. Phuong, « *Synthèse et étude des relations structure/activité quantitatives (QSAR/2D) d'analogues Benzo[c]phénanthridiniques* » Thèse de doctorat, Université d'Angers. France, 2007.
- [59] N. L. Mai, Y-M Koo, *Biochem. Engin. J.* 87, 2014, 33-40.
- [60] J. Wu, Y. Wang, Y. Shen, *Chemomet. Intel. Lab. Syst.* 131, 2014, 22-30.
- [61] S. CHTITA, *Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments. Thèse de doctorat. Université Moulay Ismaïl, Meknès, 2017.*
- [62] C.A. Lipinski, F. Lombardo, B.W. Dominy, and P.J. Feeney, —*Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*‡, *Advanced Drug Delivery Reviews*, 6(1–3), 1997, 3–25.
- [63] S. Mannan, *Lee's Loss Prevention in Process Industries: Hazard Identification, Assessment and Control*‡, Elsevier Butterworth-Heinemann, Burlington, 2005.
- [64] N. Trinajstić, S. Nikolić, S.C. Basak, and I. Lukovits, —*Distances indices and their hypercounterparts: Intercorrelation and use in the structure-property modeling*‡, *SAR and QSAR in Environmental Research*, 12, 2001, 31–54.
- [65] P.P. Roy, S. Paul, I. Mitra, and K. Roy, —*Two novel parameters for validation of predictive QSAR models*‡, *Molecules*, 14, 2009, 1660–701.
- [66] J.G. Topliss and R.P. Edwards, —*Chance factors in studies of quantitative structure-activity relationships*‡, *Journal of Medicinal Chemistry*, 22(10), 1979, 1238–1244.
- [67] K. Roy, S. Kar and R. Narayan Das, —*Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment, Chapter 6 – Selected Statistical Methods in QSAR*, Academic Press, Boston, 2015, 191–229.
- [68] A. Tropsha, P. Gramatica, and V.K. Gombar, —*the importance of Being Earnest: Validation is the Absolute Essential for Successful Application and interpretation of QSPR Models*, *QSAR and Combinatorial Sciences*, 22(1), 2003, 69–77.

- [69] R. PRATIM, P. Partha, Somnath, MITRA, Indrani, et al. On two novel parameters for validation of predictive QSAR models. *Molecules*, 2009, vol. 14, no 5, p. 1660-1701.
- [70] A. GOLBRAIKH, A. TROPSHA, Beware of q^2 !. *Journal of molecular graphics and modelling*, 2002, vol. 20, no 4, p. 269-276.



Chapitre II :
Généralité sur la coumarine,
L'IMAO-B et la maladie de
parkinson .

II.1 .Coumarine

II.1.1. Structure chimique et nomenclature

➤ Structure chimique

La coumarine est un composé chimique hétérocyclique, produit par la combinaison d'un noyau benzène avec un pyrane, présentant une fonction cétone en position α par rapport à l'oxygène (Figure .1) [1].

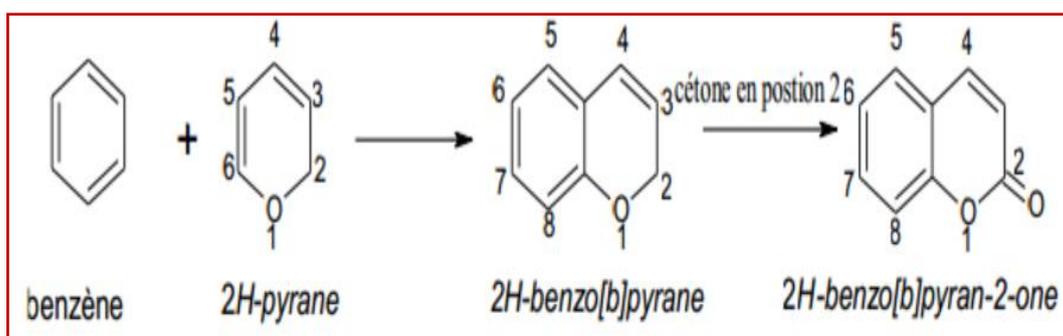


Figure 1: Combinaison d'un benzène avec un pyrane en une coumarine

➤ Formule brute: $C_9H_6O_2$

➤ Noms chimiques:

Nom IUPAC: 2H-chromen-2-one ou 2H-1-benzopyran-2-one.

Autres: 1,2-benzopyrone (α -benzopyrone) [2].

L'IUPAC considère la coumarine comme, étant, un dérivé de 2H-chromène (ou chrom-3-ène), un des deux isomères structuraux du produit d'insaturation de la chromane. (Figure. 2)

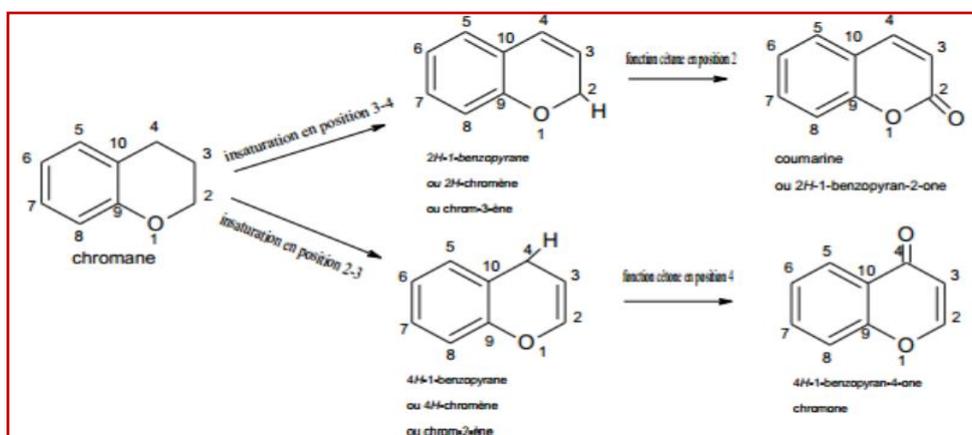


Figure. 2: Parenté chimique de la coumarine et de son isomère structural.

➤ **Utilisation**

- ⇒ Les coumarines sont utilisées comme intermédiaires pour la synthèse des molécules importantes dans le domaine pharmaceutique [3,4].
- ⇒ Elles sont utilisées en parfumerie. Leur odeur se rapproche de la vanilline et du foin fraîchement coupé. On la retrouve dans plusieurs plantes comme l'aspérule, la cannelle de Chine, le céleri, la fève tonka, la lavande,.... [5,6].
- ⇒ Du fait de l'accessibilité et la variabilité de ces composés hétérocycliques synthétiques différemment substitués ces derniers jouent un rôle important non seulement dans la chimie organique, mais également dans le domaine de la chimie médicinale [7-10]. En fait, les coumarines ont été précédemment décrits comme:
 - anticancéreux,
 - antiviraux,
 - anti-inflammatoires,
 - antimicrobiens,
 - agents antioxydants et *inhibiteurs enzymatiques* [11-13]
- ⇒ Au cours des dernière années, les dérivés de la coumarine ont été identifiés comme des inhibiteurs d'enzymes thérapeutiquement importants tels que l'aromatase [14,15], l'acétylcholinestérase (AChE) [16] et de la MAO [17,18].

II.2. Monoamines oxydases (MAO)

MAO sont des enzymes ubiquitaires situées au niveau de la membrane mitochondriale externe et appartenant à la famille des flavines amines réductase (nommés flavoenzymes). Elles catalysent la désamination oxydative des amines primaires, secondaires et tertiaires [19] tels que : la sérotonine, la norépinephrine, la β -phényléthylamine et la dopamine (voir figure 3). La MAO est impliqué dans plusieurs processus physiologiques notamment dans le maintien de l'homéostasie de ces molécules en contrôlant leur concentration au niveau du SNC (le système nerveux central) [20].

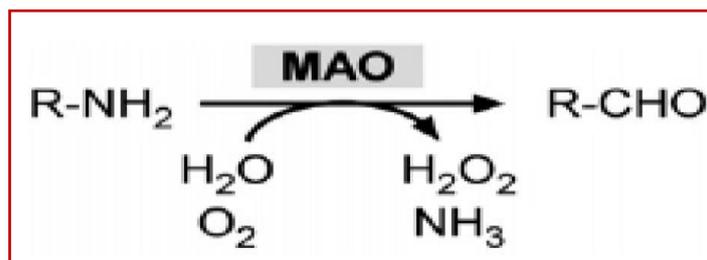


Figure.3 : Désamination oxydative des amines par les MAO.

II.2.1. Isoformes

Dans cette famille d'enzyme, on dénombre deux isoformes :

a. MAO-A

Présente dans la muqueuse gastro-intestinale, elle joue un rôle dans le métabolisme de la sérotonine dans le cerveau [21].

b. MAO-B

Enzyme majeur des plaquettes et des cellules gliales [22], joue un rôle important dans la dégradation de la dopamine [23].

II.2.2. Structure de la MAO-B

MAO-B est une enzyme de poids moléculaire d'environ 63 KDa qui constitue un dimère de 520 acides aminés [24, 25]. La structure cristalline de la MAO-B humaine montre que l'enzyme est composée de trois domaines (figure 4):

- Le premier est le domaine de liaison au FAD (cofacteur se liant par une liaison covalente à l'enzyme au niveau du résidu Cys 397) [24].
- Le deuxième est un domaine hydrophobe de liaison au substrat entouré par des résidus principalement aromatiques et aliphatiques (triade catalytique) [26].
- Le dernier est un domaine transmembranaire attaché à la membrane externe des mitochondries. Il est constitué de la séquence des résidus acides aminés de 489 jusqu'à 520 [27].

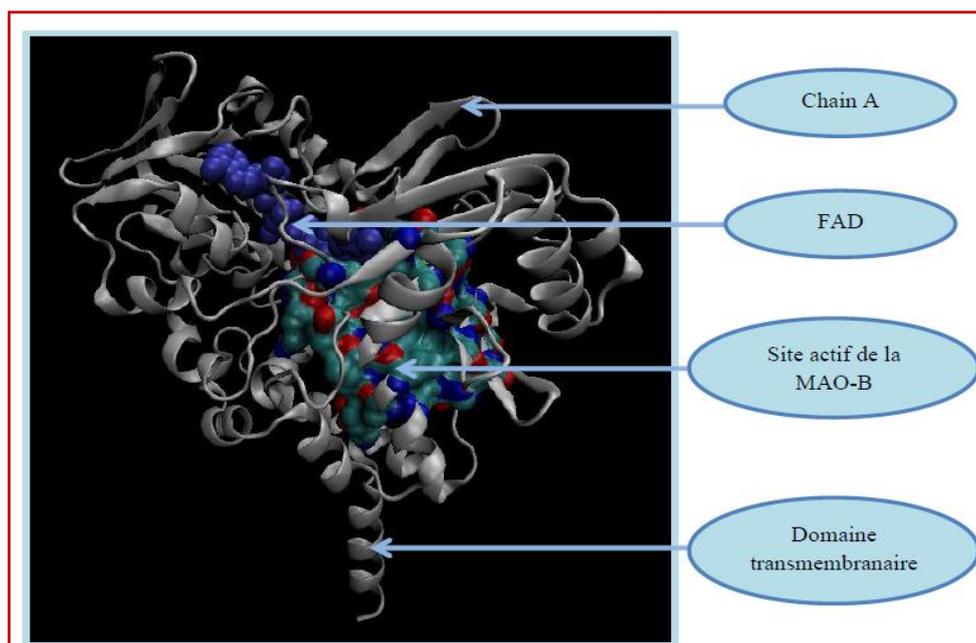


Figure.4: Structure cristallographique de la MAO-B humaine.

II.2.3. Inhibiteurs de la MAOB

La MAOB constitue une cible thérapeutique importante dans le traitement de la maladie de Parkinson (MP) car elle catalyse la décomposition de la dopamine (neurotransmetteur) en générant plusieurs produits chimiques neuro-toxiques, telles que le peroxyde d'hydrogène, l'ammoniac et les aldéhydes [28].

Les inhibiteurs de la MAOB sont de petites molécules très utilisées dans le traitement de la MP ayant pour effet d'augmenter la disponibilité de la dopamine dans le cerveau par la réduction de sa dégradation [28] et donc augmenter ses effets tout en diminuant la formation des produits toxiques.

II.3. Maladie de Parkinson (MP)

La maladie de Parkinson (MP) est une maladie neuro-dégénérative du système nerveux centrale qui fut découverte en 1817 par James Parkinson. De nos jours, c'est une pathologie assez fréquente qui constitue la seconde affection dégénérative après la maladie d'Alzheimer (MA) [29].

La MP est caractérisée par une perte des neurones dopaminergiques situés dans une structure des ganglions de la base: la pars compacta du locus niger ou SNpc (substantia nigra pars compacta). Les ganglions de la base, également appelés noyaux gris centraux, forment un ensemble de structures sous-corticales anatomiquement interconnectées (figure 5) [30]. Au

moment où les symptômes apparaissent, 60 à 80% des cellules ont déjà disparu. Ces cellules dopaminergiques sont responsables de la production d'un élément chimique appelé dopamine qui est un neurotransmetteur, c'est-à-dire un messager chimique responsable de la transmission de signaux, entre la substance noire et le striatum. Ces signaux interviennent dans le contrôle des mouvements. La perte de ces neurones entraîne des altérations dans l'activité des circuits neuronaux au sein des noyaux de la base, lesquels régulent les mouvements [31].

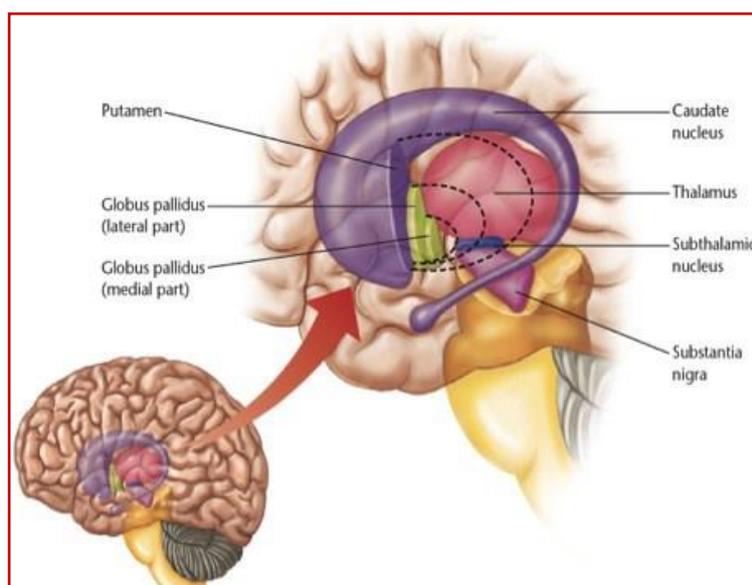


Figure.5 : Anatomie des ganglions de la base chez l'humain

II.3.1. Symptômes [32]

- ⇒ Tremblement
- ⇒ Mouvement ralenti (bradykinésie)
- ⇒ Les muscles rigides
- ⇒ Perte de mouvements automatiques
- ⇒ Modifications de la parole
- ⇒ Difficultés d'écriture.

II.3.2. Epidémiologie

La MP est désormais en passe de devenir un véritable problème public qui touche de plus en plus la santé des adultes dans les quatre coins du monde. En effet, l'Organisation Mondiale de

la Santé (OMS) estime qu'environ 4 millions de personnes à travers le monde sont atteintes de la MP. Selon cette même organisation, ce nombre dédoublera d'ici 2040 si rien n'est fait pour lutter contre cette maladie. La MP touche légèrement plus les hommes que les femmes, et survient le plus souvent chez le sujet âgé, avec une apparition entre 50 et 70 ans [33]. Les dernières études épidémiologiques montrent que plus de quatre milles algériens souffrent de la MP [34].

II.3.3. Traitement actuel de la maladie de parkinson

Dans l'état actuel des connaissances de la médecine, aucun traitement curatif ne permet de guérir de la MP, encore moins pour la prévenir. Cependant, certains médicaments ont la capacité de ralentir parfois la progression des symptômes de cette maladie. Ces traitements peuvent améliorer la qualité de vie durant plusieurs années. Leur efficacité est très variable d'une personne à l'autre et les effets secondaires peuvent également être importants. Trois types de médicaments sont principalement utilisés :

⇒ **La lévodopa (L-dopa)** [31].

⇒ **Les agonistes dopaminergiques** [33].

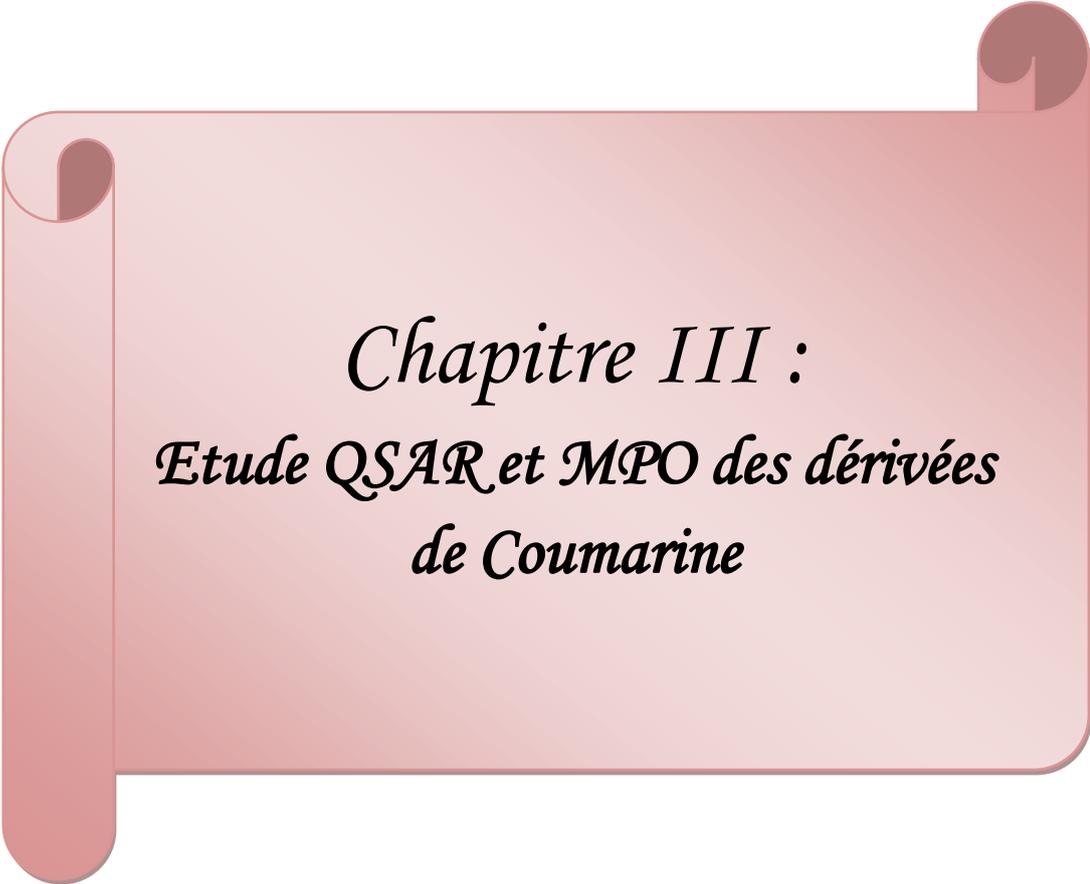
Ces deux traitements ne ralentissent pas la progression de la maladie et peuvent être liés à des effets secondaires. En effet, la lévodopa peut provoquer des crampes douloureuses dans les jambes (dystonie) et leur efficacité diminue avec le temps [35]. De même, les agonistes dopaminergiques provoquent rapidement des effets gastro-intestinaux indésirables comme les vomissements [31]. Cela conduit les chercheurs à développer des traitements alternatifs retardant réellement la progression de la maladie avec moins d'effets secondaires.

⇒ **Les inhibiteurs des enzymes du catabolisme**: les inhibiteurs de la monoamine oxydase B (IMAOB) et de la catéchol-O-méthyl-transférase (ICOMT), constituent la troisième classe de traitement de la MP. Il s'agit de petites molécules capables de diminuer la dégradation de la dopamine en soulageant des symptômes liés à la MP tels que : la bradykinésie (lenteur lors de l'exécution d'un mouvement), l'instabilité posturale (déséquilibre vers l'avant ou vers l'arrière) qui favorisera le risque de chutes, les tremblements, ou encore, la rigidité musculaire et ensuite le handicap du patient [31]. De même, ces composés peuvent réduire les risques d'apparition des complications observés en cas d'utilisation d'un des deux types de médicaments de la MP cités plus haut et ce avec des effets secondaires plus tolérables [36]. C'est raisons expliquent notre intérêt pour cette classe de traitement et en particulier les inhibiteurs de la MAOB.

- [01] R. Fethi, M. S. Fedouche, *thèse de docteur en pharmacie: Coumarines à intérêt thérapeutique: Synthèse et contrôle analytique*, 2017, p: 21-22.
- [02] J. Roberts D, M. C. Caserio. *Chimie Organique Moderne*. Paris: Ediscience ; 1968 .
- [03] M. A. Khalil, S. M. Sayed, and M. A. Raslan. *Synthesis of New Pyrazolo[5,1-c]triazine, Triazolo[5,1-c]triazine, Triazino[4,3-b]indazole and Benzimidazo [2,1-c]triazine Derivatives Incorporating Chromen-2-one Moiety*. *Journal of the Korean Chemical Society*, 2013, 57, 5-612.
- [04] M. E. Jung, D. A. Allen. *Organic Lett*, 2009; 11:757. Chen G, Tokunaga N, Hayashi T. *Organic Lett*, 2005; 7: 2285. (c) Stoffman E. J. L, Clive D. L. K. *Org. Biomol. Chem*, 2009, 7, 4862.
- [05] F. Borges, et al. *Simple coumarins and analogues in medicinal chemistry: occurrence, synthesis and biological activity*. *Curr. Med. Chem*, 2005, 12, 887- 916.
- [06] F. Borges, et al. *Simple coumarins: privileged scaffolds in medicinal chemistry*. *Front. Med. Chem*, 2009, 4, 23-85.
- [07] L. Kabeya, et al. *Inhibition of horseradish peroxidase catalytic activity by new 3-phenylcoumarin derivatives: synthesis and structure-activity relationships*. *Bioorg. Med.Chem*, 2007, 15, 1516-1524.
- [08] F. Belluti, et al. *Design, synthesis and anticancer activities of stilbene-coumarin hybrid compounds: identification of novel proapoptotic agents*. *Bioorg. Med. Chem*, 2010, 18, 3543-3550.
- [09] M. Roussaki, et al. *A novel synthesis of 3-aryl coumarins and evaluation of their antioxidant and lipoxygenase inhibitory activity*. *Bioorg. Med. Chem. Lett*, 2010, 20, 3889-3892.
- [10] J. Neyts, et al. *Structure-activity relationship of new anti-hepatitis C virus agents: heterobicycle-coumarin conjugates*. *J. Med. Chem*, 2009, 52, 1486-1490.
- [11] M.J. Matos, et al. *Synthesis and study of a series of 3-arylcoumarins as potent and selective monoamine oxidase B inhibitors*. *J. Med. Chem*, 2011, 54, 7127-7137.
- [12] D. Viña, et al. *3-substituted coumarins as dual inhibitors of AChE and MAO for the treatment of Alzheimer's disease*. *MedChemComm*, 2011, 3, 213-218.
- [13] S. Serra, et al. *Hydroxycoumarins as selective MAO-B inhibitors*. *Bioorg. Med. Chem. Lett*, 2012, 22, 258-261.

- [14] F. Chimenti, et al. *Synthesis, molecular modeling, and selective inhibitory activity against human monoamine oxidases of 3-carboxamido-7-substituted coumarins*. *J. Med. Chem*, 2009, 52, 1935-1942.
- [15] S. Chen, M. Cho, K. Karlsberg, D. Zhou, Y.C. Yuan. *Biochemical and biological characterization of a novel antiaromatase coumarin derivative*. *J. Biol. Chem*, 2004, 279, 48071-48078.
- [16] F. Leonetti, A. Favia, A. Rao, R. Aliano, A. Paluszczak, R.W. Hartmann, A. Carotti. *Design, synthesis, and 3D QSAR of novel potent and selective aromatase inhibitors*. *J. Med. Chem*, 2004, 47, 6792-6803.
- [17] C. Bruhlmann, F. Ooms, P.A. Carrupt, B. Testa, M. Catto, F. Leonetti, C. Altomare, A. Carotti. *Coumarins derivatives as dual inhibitors of acetylcholinesterase and monoamine oxidase*. *J. Med. Chem*, 2001, 44, 3195-3198.
- [18] C. Gnerre, M. Catto, F. Leonetti, P. Weber, P.A. Carrupt, C. Altomare, A. Carotti, B. Testa, *Inhibition of Monoamine Oxidases by Functionalized Coumarin Derivatives: Biological Activities, QSARs, and 3D-QSARs*. *J. Med. Chem*, 2000, 43, 4747-4758.
- [19] K. Markošová., I. Dolejš, R. Stloukal., L. Rios-Solis, M. Rosenberg, M. Micheletti et al. *Immobilization and kinetics of monoamine oxidase (MAO-N-D5) enzyme in polyvinyl alcohol gels*. *Journal of Molecular Catalysis B: Enzymatic*. 2016, 129: 69-74.
- [20] C. Christelle, *Monoamine oxydase et athérosclérose : signalisation mitogène et études in vivo*. Thèse de doctorat d'université : innovation pharmacologique Toulouse: l'Université Toulouse III - Paul Sabatier . France. 2007, p 147.
- [21] M.S. Nel, A. Petzer, J.P. Petzer, L.J. Legoabe, *2-Heteroarylidene-1-indanone derivatives as inhibitors of monoamine oxidase*. *Bioorganic Chemistry*. 2016 ; 69: 20-28.
- [22] P. Jenner, *Mitochondria, monoamine oxidase B and Parkinson's disease*. *Basal ganglia*. 2012 ; 2: S3-S7.
- [23] P. Riederer, T. Muller, *Use of monoamine oxidase inhibitors in chronic neurodegeneration*. *Expert Opinion on Drug Metabolism & Toxicology*. 2017.
- [24] H. Gaweska, P. F. Fitzpatrick, *Structures and mechanism of the monoamine oxidase family*. *Biomolecular concepts*. 2011 ; 2 : 365-377.
- [25] R. R. Ramsay, *Molecular aspects of monoamine oxidase B*. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. 2016; 69: 81-89.

- [26] A. Bolasco, R. Fioravanti, & S. Carradori, *Recent development of monoamine oxidase inhibitors. Expert Opinion on Therapeutic Patents*. 2005 ; 15 : 1763-1782.
- [27] http://proteopedia.org/wiki/index.php/Monoamine_oxidase.
- [28] <http://pontt.net/wp-content/uploads/2016/03/TFE-Maladiede-Parkinson.pdf>.
- [29] <https://www.uclouvain.be/cps/ucl/doc/commission-gbio/documents/PARK.pdf>.
- [30] Cl. Gibrat, *Propriétés thérapeutiques de la cystamine pour la maladie de parkinson: étude chez divers modèles animaux de la maladie*, Faculté de médecine: université laval Québec, 2013.
- [31] M. Ghardaoui, C. Merrad, *Mémoire de Master: Conception in silico de nouveaux inhibiteurs de la monoamine oxydase B pour le traitement de la maladie de Parkinson*, 2017, P:2-11.
- [32] http://www.medecine.unige.ch/enseignement/apprentissage/module4/immersion/arches/2008_2009/travaux/09_r_parkinson.pdf.
- [33] B. Cyrine, *thèse de doctorat: Maladie de parkinson et thérapies innovantes: perspectives d'avenir?*, 2015.
- [34] www.cochrane.org/.../inhibiteurs-de-lamonoamine.
- [35] R. Benarab et M. Khenoussi, *Les inhibiteurs de la Monoamine oxydase en psychopharmacologie*, 2012.
- [36] L. Manon, *thèse présentée à l'université du Québec à Trois-Rivières, rôle des protéines responsables du maintien du cytosquelette neuronal dans la maladie de parkinson et les dyskinésies induites par la L-dopa*, 2010.



Chapitre III :
Etude QSAR et MPO des dérivées
de Coumarine

Dans cette partie de ce travail nous allons établir une relation linéaire structure-activité quantitative entre l'activité inhibitrice de la mono amine oxydase B (MAO-B) d'une série dérivées de coumarine et leurs descripteurs structuraux, par conséquent nous proposons un modèle quantitatif, et nous essayons d'interpréter l'activité de ces molécules en se basant sur la méthode statistique suivante : Régression linéaire multiple RLM.

Puis nous allons utiliser les méthodes d'optimisation multiparamétriques (MPO) pour prédire le meilleur équilibre des propriétés, parmi ces méthodes: Les règles de Lipinski, Veber, l'efficacité du ligand (LE) et en fin l'efficacité lipophile de ligand (LipE).

III.1. Présentation de la série de molécule et des descripteurs

III.1.1. Série de molécule

Nous avons étudié et analysé la série de la molécule de coumarine constituée de 20 dérivés Figure 1. Ceci dans le but de trouver une relation quantitative entre l'activité inhibitrice MAO et la structure de ces molécule qui sont décrites par leurs substituant R_3 , R_4 , R_6 et R_7 .

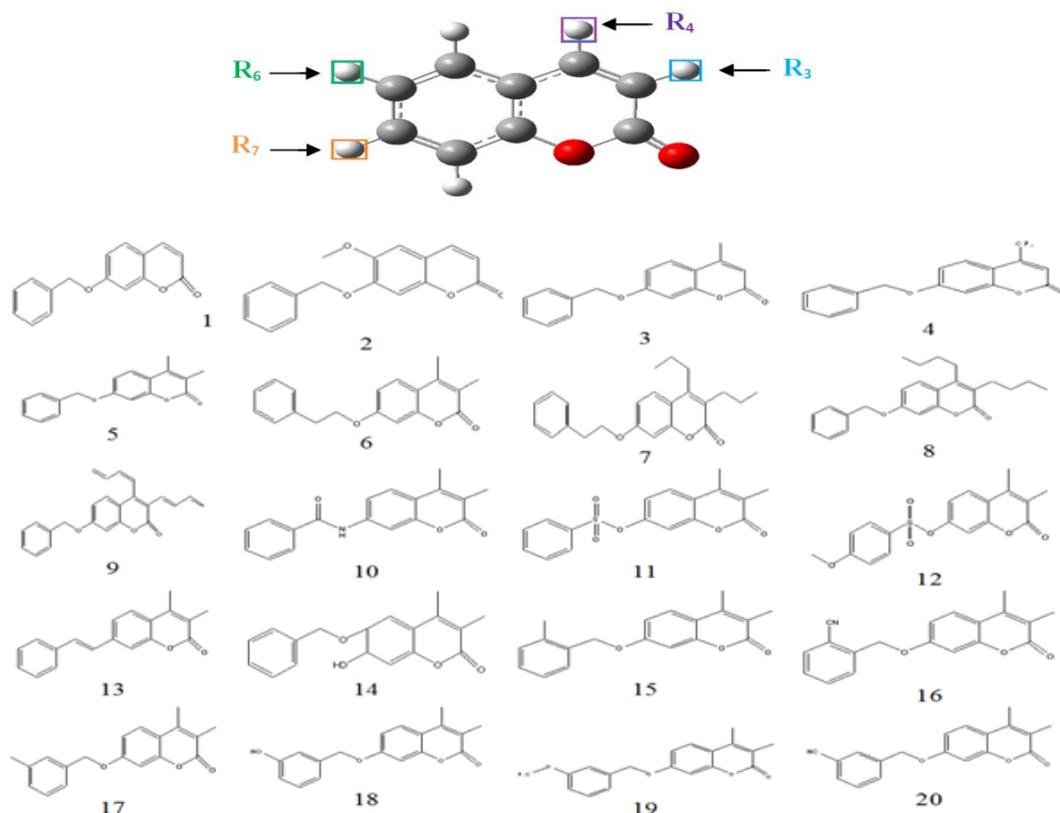


Figure.1: Structures chimiques de coumarine et ses dérivés

III.1.2. Descripteurs

Dans ce travail les descripteurs ont été choisis pour décrire la structure des molécules constituant la série à étudier sont données dans le tableau 1 suivant:

Tableau.1: Type de descripteurs moléculaire utilisés pour l'étude QSAR

<i>Descripteurs moléculaire</i>	<i>Type</i>
Réfractivité molaire (RM)	<i>Propriétés physicochimique</i>
Polarisabilité (Pol)	
Lipophile (logP)	
Donneur de liaison hydrogène (HBD)	
Accepteur de liaison hydrogène (HBA)	
Surface moléculaire (SAG)	
Densité (D)	
Energie totale (E_T)	<i>Electronique</i>
Moment dipolaire (MD)	
Charge (qC3, qC4, qC6, qC7)	
Energie de l'orbitale la plus haute occupée (E_{HOMO})	
Energie de l'orbitale la plus basse vacante (E_{LUMO})	
Gap énergétique (E_{gap})	
Volume molaire (MV)	<i>Géométrique</i>
Nombre de liaisons rotatives (NROT)	<i>Constitutionnel</i>
Masse moléculaire (WM)	<i>Topologique</i>
Superficie de la surface polaire (PSA)	

Le tableau 2 regroupe tout les Descripteurs utilisés dans le développement de modèle QSAR

Tableau. 2: Descripteurs utilisés dans le développement des modèles QSAR

composés	Log p	D	HE (Kcal/mol)	ET (a.u)	Réf (Å ³)	Poi (Å ³)	MW (amu)	V (Å ³)	S (Å ³)	HOMO (a.u)	LUMO (a.u)	E _{gap} (a.u)	PSA(Å ³)	NROT	qC3	qC4	qC6	qC7	MD (D)	HBD	HBA
01	0,34	0,339	-5,98	-842,618	80,78	27,83	252,27	743,40	458,98	-0,2208	-0,0586	-0,162	35,53	3	-0,383	0,029	-0,271	0,398	4,303	0	2
02	-0,66	0,359	-6,09	-957,135	87,10	30,30	282,30	784,75	477,89	-0,2160	-0,0588	-0,157	44,76	4	-0,399	0,052	0,222	0,185	4,608	0	3
03	0,49	0,336	-5,09	-881,941	85,01	29,66	266,30	791,78	483,00	-0,2185	-0,0542	-0,164	35,53	3	-0,480	0,270	-0,302	0,413	4,446	0	2
04	0,91	0,389	-5,80	-1179,650	85,94	29,39	320,27	823,09	500,09	-0,2321	-0,0773	-0,155	35,53	4	-0,329	0,055	-0,280	0,423	4,381	0	5
05	0,77	0,335	-4,42	-921,260	89,36	31,50	280,32	835,74	504,45	-0,2125	-0,0513	-0,161	35,53	3	-0,212	0,104	-0,302	0,406	3,896	0	2
06	1,02	0,333	-4,24	-960,577	94,12	33,33	294,35	882,79	530,47	-0,2132	-0,0531	-0,160	35,53	4	-0,235	0,126	-0,316	0,491	4,030	0	2
07	2,35	0,327	-3,39	-1078,525	107,77	38,84	336,43	1027,74	610,09	-0,2115	-0,0500	-0,162	35,53	7	-0,277	0,062	-0,308	0,406	3,789	0	2
08	3,15	0,321	-2,74	-1157,409	116,97	42,51	364,48	1134,17	667,47	-0,2112	-0,0497	-0,162	35,53	9	-0,250	0,093	-0,322	0,421	3,739	0	2
09	2,19	0,327	-8,35	-1152,231	119,29	41,74	356,42	1088,51	645,55	-0,1968	-0,0763	-0,1205	35,53	7	-0,163	0,182	-0,327	0,430	4,084	0	2
10	0,01	0,346	-6,02	-975,431	91,54	32,30	293,32	845,55	510,10	-0,2219	-0,0665	-0,1554	55,40	2	-0,214	0,097	-0,244	0,376	5,522	1	2
11	-0,09	0,378	-7,79	-1430,518	93,84	30,80	330,36	872,82	526,26	-0,2341	-0,0700	-0,1641	69,67	3	-0,189	0,085	-0,287	0,391	3,089	0	3
12	-1,08	0,379	-9,43	-1545,046	100,22	33,27	360,38	949,79	568,14	-0,2311	-0,0669	-0,1642	78,90	4	-0,197	0,101	-0,310	0,426	2,979	0	4
13	1,89	0,325	-3,83	-884,146	93,07	32,50	276,33	848,99	512,95	-0,2071	-0,0744	-0,1327	26,30	2	-0,197	0,070	-0,296	0,284	4,784	0	1
14	-0,26	0,348	-5,95	-996,472	90,97	32,13	296,32	849,76	507,52	-0,2107	-0,0529	-0,1578	55,76	3	-0,201	0,101	0,178	0,281	4,013	1	3
15	0,92	0,336	-7,21	-960,578	93,65	33,33	294,35	875,40	520,09	-0,2161	-0,0546	-0,1615	35,53	3	-0,244	0,105	-0,333	0,450	4,256	0	2
16	0,49	0,345	-3,24	-1013,498	94,34	33,35	305,33	884,45	529,05	-0,2114	-0,0507	-0,1607	59,32	3	-0,223	0,110	-0,311	0,406	0,899	0	3
17	0,92	0,331	-10,85	-960,580	93,65	33,33	294,35	888,75	533,90	-0,2160	-0,0616	-0,1544	35,53	3	-0,218	0,086	-0,322	0,420	3,973	0	2
18	-0,26	0,345	-5,81	-996,480	90,97	32,13	296,32	858,60	517,09	-0,2119	-0,0508	-0,1611	55,76	3	-0,208	0,089	-0,317	0,428	3,243	1	3
19	1,71	0,383	-9,51	-1333,513	96,85	33,70	364,32	951,49	571,63	-0,2109	-0,0498	-0,1611	44,76	5	-0,211	0,104	-0,316	0,423	3,558	0	6
20	0,49	0,363	-9,15	-1013,501	94,34	33,35	305,33	893,31	538,40	-0,2197	-0,0583	-0,1614	59,32	3	-0,223	0,108	-0,308	0,408	4,266	0	3

III.2. Méthodes et matériels

Pour réaliser cette étude nous disposons des logiciels suivants :

- Tout d'abord, les vingt molécules de la coumarine ont été pré-optimisées au moyen de la mécanique moléculaire, champ de force (MM+) en utilisant le logiciel HyperChem 8.0.6 [1]. Ensuite, ces dérivées ont été réoptimisées en utilisant le logiciel gaussien 09 [2], au niveau de la théorie de la densité fonctionnelle DFT en utilisant le Lee-Yang-Parr à trois paramètres de Becke (B3LYP), avec de base 6 -31G (d, p).
- Le module « propriétés QSAR » de l'HyperChem 8.0.6 a été utilisé pour calculer les paramètres suivants: La polarisabilité (Pol), la réfractivité molaire (MR), le coefficient de partage octanol/eau (log P), le volume molaire (MV), la surface Moléculaire (SAG) et le poids moléculaire (MW).
- Les descripteurs électroniques: le moment dipolaire (MD), l'énergie des orbitales frontières (E_{HOMO} , E_{LUMO}), les charges (qC3, qC4, qC6 et qC7), l'énergie totale (E_{T}) ont été calculés par la méthode DFT (B3LYP /6 -31G d, p et les charges partielles atomiques peuvent être calculées par le modèle CHELPG [3]) en utilisant logiciel Gaussian 09 et GaussView (5.0).
- Logiciel MarvinSketch [4] a été utilisé pour calculer les descripteurs suivantes: Le nombre des Liaisons hydrogène donneur et accepteur (HBD, HBA), Superficie de la surface polaire (PSA) et aussi le nombre de liaison rotatif (nrotb)
- La méthode de régression multilinéaire (MLR) a été utilisée pour développer de modèle QSAR en utilisant le logiciel SPSS 20 [5].

III.3. Corrélation entre différent descripteurs

d'entamer le développement effectif des équations de régression QASR, il est hautement recommandé d'examiner la qualité statistique des données de départ, à la fois les données à corrélérer (variable dépendante) et les descripteurs utilisés dans la corrélation (variables indépendantes).

Pour un ensemble de descripteurs différents, il est nécessaire d'effectuer une analyse des données bivariée, c'est-à-dire de calculer le coefficient de corrélation entre chacune des paires de l'ensemble des descripteurs. Si ce coefficient est statistiquement significatif ($R > 0,95$), ces deux descripteurs sont considérés comme fortement corrélés et ne peuvent être utilisés simultanément lors de l'analyse QSAR [6] et en pratique, ils seront alors enlevés dans le procédé de sélection.

Tableau.3: La matrice de corrélation entre différent descripteurs

Desc	qC3	qC4	qC6	qC7	MD	ET	LogP	HE	MR	Pol	W	V	S	D	HOMO	LUMO	Egap	HBD	HBA	NRot	PSA	
qC3	1,000																					
qC4	-0,164	1,000																				
qC6	-0,244	-0,211	1,000																			
qC7	0,212	0,300	-0,837	1,000																		
MD	-0,278	-0,009	0,210	-0,265	1,000																	
ET	-0,380	0,056	0,170	-0,223	0,376	1,000																
LogP	0,289	0,037	-0,374	0,293	-0,030	-0,168	1,000															
HE	-0,246	-0,045	0,052	-0,127	-0,078	0,414	0,184	1,000														
MR	0,489	0,133	0,189	-0,029	-0,125	-0,259	0,534	0,078	1,000													
Pol	0,454	0,134	0,161	-0,007	-0,107	-0,143	0,609	0,160	0,984	1,000												
W	0,492	0,010	0,156	0,055	-0,198	-0,570	0,350	-0,135	0,872	0,805	1,000											
V	0,500	0,121	0,025	0,119	-0,202	-0,326	0,638	0,107	0,974	0,971	0,881	1,000										
S	0,497	0,128	-0,053	0,171	-0,214	-0,353	0,681	0,105	0,948	0,948	0,859	0,993	1,000									
D	-0,014	-0,220	0,167	-0,066	-0,183	-0,676	-0,385	-0,412	-0,272	-0,378	0,112	-0,278	-0,290	1,000								
HOMO	0,239	0,250	0,032	-0,080	0,002	0,423	0,515	0,230	0,476	0,573	0,044	0,438	0,439	-0,680	1,000							
LUMO	-0,137	0,080	0,064	0,140	-0,305	0,268	0,155	0,282	0,074	0,182	-0,078	0,122	0,100	-0,286	0,283	1,000						
Egap	0,311	0,132	-0,030	-0,186	0,265	0,110	0,281	-0,058	0,320	0,305	0,103	0,248	0,268	-0,302	0,562	-0,634	1,000					
HBD	0,236	-0,049	0,346	-0,199	0,172	0,167	-0,288	0,059	0,269	0,275	0,226	0,193	0,134	0,008	0,059	0,109	-0,047	1,000				
HBA	0,036	-0,144	0,086	0,080	-0,285	-0,625	-0,088	-0,406	-0,099	-0,149	0,197	-0,071	-0,077	0,871	-0,363	-0,005	-0,289	-0,012	1,000			
NRot	0,042	0,048	-0,127	0,214	-0,106	-0,342	0,799	0,179	0,598	0,634	0,518	0,699	0,748	-0,192	0,341	0,142	0,152	-0,295	0,060	1,000		
PSA	0,330	-0,053	0,134	-0,031	-0,459	-0,599	-0,461	-0,355	0,100	-0,008	0,394	0,096	0,079	0,554	-0,484	-0,013	-0,379	0,318	0,408	-0,238	1,000	

Les descripteurs hautement corrélés ont été exclus:

- La polarisabilité est parfaitement corrélée avec la réfractivité molaire pour $r = 0,984$.
- La surface est fortement corrélée avec le volume pour $r = 0,993$.
- Le volume est parfaitement corrélé avec la polarisabilité et réfractivité molaire ($r = 0,971, 0,974$).

III.4. Equation du modèle QSAR

Le tableau 4 rassemble les descripteurs sélectionnés, leurs coefficients et p value (pour un seuil de signification inférieur à 0,05).

Tableau.4: Les descripteurs sélectionnés par MLR

Modèle	coefficient	p-value
constant	27,898	0,000
D	-61,339	0,000
qC6	-3,656	0,000
HBA	0,583	0,016
LUMO	32,348	0,032

Donc l'équation de la régression linéaire obtenue est la suivante (équation 1):

$$PIC_{50} = 27,898 - 61,339 D - 3,656 qC6 + 0,583 HBA + 32,348 LUMO$$

$$N=20; R=0.945; R^2= 0,893 S= 0.451; F=31,446$$

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 [7]. Les valeurs $R = 0,945$ ($R^2 = 0,893$) nous permis d'indiquer fermement la corrélation entre les différents descripteurs utilisés tels que: LUMO, qC6, D et HBA et l'activité biologique étudiée (inhibiteur de MAO-B).

La valeur de F calculée est supérieure à la valeur lue dans la table du Fischer ($31,446 > 3,06$) (Tableau 5), et indique donc que la relation de régression était statistiquement significative et ne s'est pas produites par hasard [7].

Tableau.5: Table de Fisher

F limite à p 0,05

TABLE DU F de FISCHER (comparaison de 2 variances)

degré de liberté du numérateur

degré de liberté du dénominateur

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

Dans ce modèle, nous remarquons que toutes augmentations de nombre de l'accepteur de liaison hydrogène et l'énergie LUMO et que toutes diminutions de qC6, et la densité provoquent à une augmentation de l'activité biologique.

La matrice de corrélation pour l'activité biologique pIC₅₀ et les descripteurs sélectionnés pour construire le modèle QSAR est présentée dans le tableau 6. Les paramètres utilisés dans ce modèle sont presque indépendants, ce qui peut être vu à partir de la matrice de corrélation.

Tableau.6: La matrice de corrélation du modèle.

	PIC50	qC6	D	LUMO	HBA
PIC50	1				
qC6	-0,574	1			
D	-0,691	0,167	1		
LUMO	0,510	0,064	-0,286	1	
HBA	-0,378	0,086	0,871	-0,005	1

III.4.1. Analyse de contribution des descripteurs

En fonction des valeurs absolues des coefficients de corrélation, l'importance du poids des descripteurs impliqués dans le modèle (équation 1) diminue dans l'ordre suivant: $D > qC6 > LUMO > HBA$. La densité a la contribution la plus importante avec (-0,691) et le plus petit la HBA (n^{bre} d'accepteur de liaison hydrogène) avec (-0,378) (Figure 2). L'activité IMAO-B de ces dérivés est donc favorisée par les molécules ayant une densité, charge atomique et de l'énergie LUMO élevée et de valeur faible de HBA. Toute proposition d'une molécule visant à améliorer l'activité IMAO de ces dérivés devrait prendre en compte ces critères.

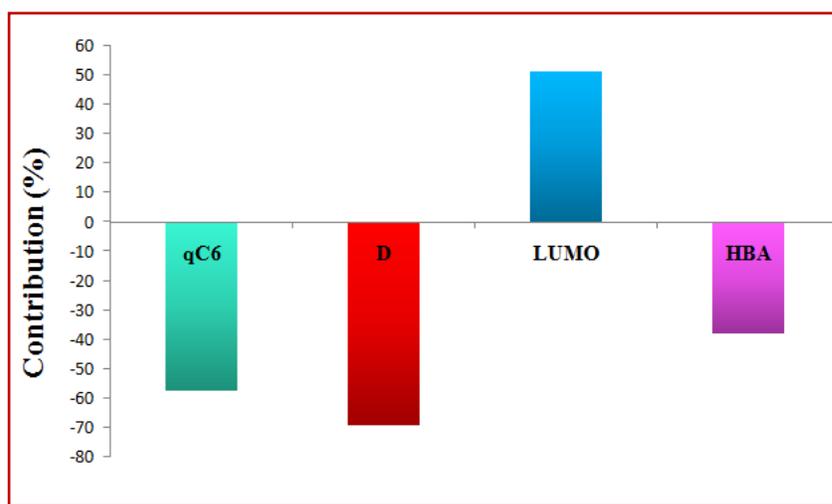


Figure.2 : La contribution des descripteurs choisis dans le modèle

D'après cet histogramme on constate que les quatre descripteurs sont nécessaires pour la construction du modèle QSAR.

Nous avons remarqué aussi les descripteurs qui sont former le modèle en deux types: électronique et physico-chimique, ces sont responsable d'améliorer l'activité inhibitrice de mono amine oxydase

III.5. Validation de modèle

Les puissances prédictives des équations ont été validées par la méthode de validation croisée "leave-one-out" (LOO) [8-10]. La validation croisée est une méthode pratique et fiable pour tester la signification d'un modèle.

Afin de tester la validité de pouvoir prédictif de modèle QSAR sélectionné, la technique de validation croisée LOO (leave-one-out) a été utilisée. Le modèle développé a été validé par le calcul des paramètres statistiques suivants: la somme des carrés résiduels (PRESS), la somme des carrés totaux (TSS ou SSY), qualité d'ajustement R^2_{adj} , coefficient de corrélation de validation croisée R^2_{CV} et la validation standard des erreurs de prédiction (SPRESS) (Tableau .7).

Tableau.7: Paramètres de validation croisée.

PRESS	SSY	PRESS/SSY	SPRESS	R^2_{CV}	R^2_{adj}
3,058	28,698	0,106	0,073	0,894	0,865

PRESS est un important paramètre de validation croisée car il s'agit d'une bonne approximation de l'erreur prédictive réelle du modèle. Sa valeur étant inférieure à SSY indique que ce modèle prédit mieux que le hasard et peut être considéré statistiquement significatif. La petite valeur de PRESS signifie la meilleure prévisibilité du modèle [9]. D'après les résultats présentés dans le tableau 7, cette valeur est égale à 3,058, le modèle est statistiquement significatif.

De plus, pour un modèle QSAR fiable, le rapport PRESS / SSY devrait être inférieur à 0,4 [7]. D'après les données présentées dans le tableau 7 indiquent que pour le modèle développé, ce rapport est de 0.106. Les valeurs élevées du R^2_{CV} et R^2_{adj} sont des critères essentiels pour une meilleure qualification des modèles QSAR [11]. Les résultats obtenus de R^2_{CV} et R^2_{adj} pour ce modèle QSAR est 0.894 et 0.865.

SPRESS (incertitude de prédiction) est un bon paramètre à utiliser pour décider de l'incertitude dans la prédiction. Plus la valeur de ce paramètre est basse, la capacité prédictive du modèle sera mieux. Le modèle QSAR développé présent une faible valeur de Spress (<0.200) ce qui indique que ce modèle présente des petites valeurs résiduelles entre l'activité biologique observée et prédite. Dans notre cas, ce paramètre porte une petite valeur 0.073 ce qui explique que l'habilité de prédiction est la meilleure pour ce modèle.

L'activité inhibitrice MAO-B expérimentale, prédite et résiduelle de coumarine et ses dérivés sont représentés dans le tableau 8. 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 [12].

**Tableau.8: Les valeurs activité expérimentales, prédites et résiduelles
des dérivées de Coumarine**

N°	PIC ₅₀ EXP	PIC ₅₀ PRED	PIC ₅₀ RESID	N°	PIC ₅₀ EXP	PIC ₅₀ PRED	PIC ₅₀ RESID
01	7,26	7,347	-0,087	11	5,28	5,222	0,058
02	5,17	4,871	0,299	12	4,77	5,928	-1,158
03	7,74	7,787	-0,047	13	7,55	7,197	0,353
04	5,86	5,471	0,389	14	5,51	5,897	-0,387
05	8,36	7,936	0,424	15	8,06	7,894	0,166
06	8,25	8,052	0,198	16	7,64	7,971	-0,331
07	8,46	8,497	-0,037	17	8,36	7,940	0,420
08	8,46	8,926	-0,466	18	8,01	7,995	0,015
09	7,30	7,710	-0,410	19	7,94	7,462	0,478
10	6,72	6,533	0,187	20	7,97	8,035	-0,065

La figure ci-dessous (figure 3) présente la corrélation entre les valeurs prédites et expérimentales pour l'activité biologique étudiée des dérivés de coumarine. Du point de vue des performances statistiques, ce modèle présente de coefficient de corrélation R^2 significatif de valeur. La valeur de R^2 est égale à 0.893, pour ce modèle. Il indique que ce modèle peut être appliqué avec succès pour prédire l'activité étudiée.

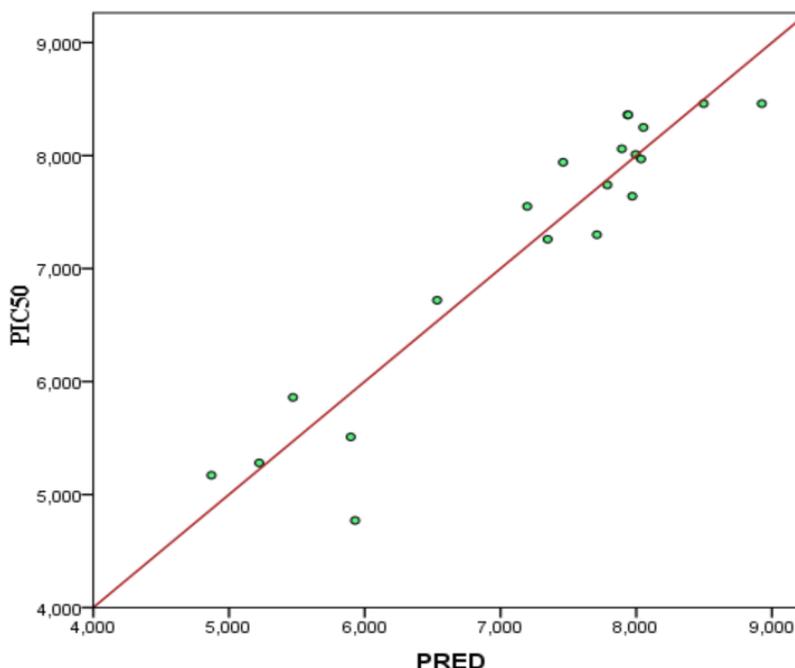


Figure.3: Corrélation entre l'activité biologique expérimentale et prédite.

Pour rechercher la présence d'une erreur systématique dans l'élaboration des modèles QSAR, les valeurs résiduelles de l'activité biologique inhibitrice de MAO-B ont été tracées en fonction des valeurs expérimentales, comme le montre la figure 4.

La répartition des résidus dans les deux côtés de l'axe des abscisses indique l'absence de l'erreur systémique [13]. Il indique que ce modèle peut être appliqué avec succès pour prédire l'activité biologique inhibitrice de MAO-B pour l'ensemble des coumarines utilisées dans le développement de modèle QSAR.

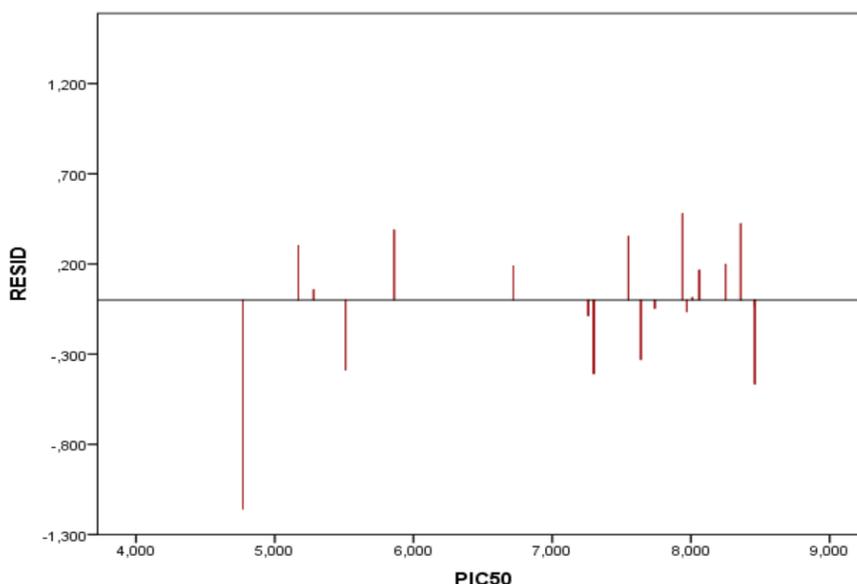


Figure.4: Corrélation entre l'activité biologique expérimentale et résiduelle.

III.6. Méthodes d'optimisation multiparamétriques (MPO)

Nous pouvons utiliser les méthodes d'optimisation multiparamétriques (MPO) pour prédire le meilleur équilibre des propriétés, parmi ces méthodes, nous appliquons des règles de base. Les plus couramment utilisés sont les règles de Lipinski et Veber [14, 15]. D'autre part les méthodes métriques les plus anciennes et les plus couramment utilisées sont l'efficacité du ligand (LE) , l'efficacité lipophile de ligand (LipE) et indice de lipophilicité dépendant de l'efficacité des ligands (LELP) [16].

III.6.1. Règle de lipinski

Le médicament apparaît comme un paradigme prometteur pour coder l'équilibre entre les propriétés moléculaires d'un composé qui influence sa pharmacodynamie et sa pharmacocinétique et optimise finalement leur absorption, leur distribution, leur métabolisme et leur excrétion (ADME) dans le corps humain comme un médicament. Les conditions empiriques pour satisfaire la règle de Lipinski et manifester une bonne biodisponibilité orale impliquent un équilibre entre la solubilité aqueuse d'un composé et sa capacité à diffuser passivement à travers les différentes barrières biologiques [14, 17].

Ces règles sont la plus utilisée pour l'identification des composés « drug-like ». D'après ces règles, les composés ne validant pas au moins deux des critères suivants ont de très fortes chances d'avoir des problèmes d'absorption ou de perméabilité par voie orale [18].

Ces règles, communément appelées « règles de Lipinski » ou « la règle de 5 », comportent cinq critères physico-chimiques qui décrivent la molécule, sont résumé sur le schéma suivant:

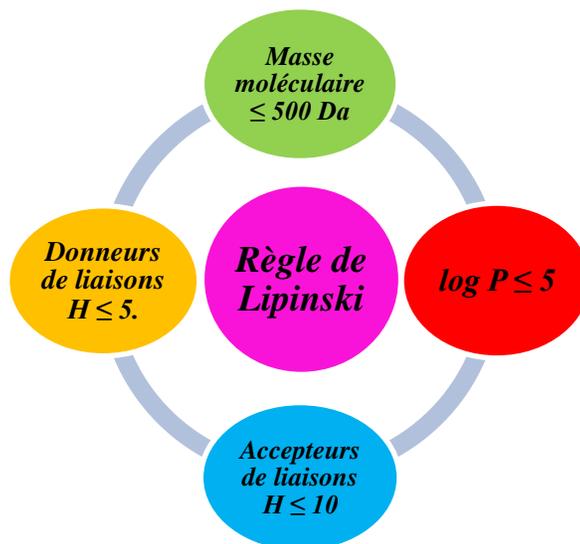


Figure.5: Règle de Lipinski

Les propriétés de divers paramètres des composés (1-20) se trouvent dans le tableau 9.

Tableau.9: Les règles de Lipinski des dérivés de coumarine

Composés	MW	Log P	HBA	HBD
01	252,27	0,34	2	0
02	282,30	-0,66	3	0
03	266,30	0,49	2	0
04	320,27	0,91	5	0
05	280,32	0,77	2	0
06	294,35	1,02	2	0
07	336,43	2,35	2	0
08	364,48	3,15	2	0
09	356,42	2,19	2	0
10	293,32	0,01	2	1
11	330,36	-0,09	3	0
12	360,38	-1,08	4	0
13	276,33	1,89	1	0
14	296,32	-0,26	3	1
15	294,35	0,92	2	0
16	305,33	0,49	3	0
17	294,35	0,92	2	0
18	296,32	-0,26	3	1
19	364,32	1,71	6	0
20	305,33	0,49	3	0

HBD : liaisons hydrogène donneurs, HBA : liaisons hydrogène accepteurs.

- A travers le tableau 9, tous les composés ont des valeurs moins de 5 pour la lipophilie, allant de -1.08 à 3.15. A cet égard, ces composés sont mieux solubilisés dans des solutions aqueuses et lipidiques.
- Nous pouvons également voir dans le tableau 9 que tous ces composés ont le nombre accepteurs d'hydrogène inférieurs à 10
- Le nombre donneurs d'hydrogène inférieurs à 5.
- Aussi, pour les poids moléculaires, tous les composés ont des valeurs inférieures à 500 Da, donc ils traversent facilement les membranes cellulaires, la valeur maximale de MW est de 364.48 pour le composé 19.

Donc, ces composés répondent aux règles de Lipinski (règles des cinq), en suggérant que ces composés théoriquement n'auraient pas de problèmes avec la biodisponibilité orale.

III.6.2. Les règles de Veber

Les règles Veber suggèrent que la flexibilité moléculaire et la surface polaire (PSA) sont des déterminants importants de la biodisponibilité orale [15]. La réduction de la flexibilité moléculaire (mesurée par le nombre de liaisons rotatives) et la faible surface polaire sont des prédicateurs importants de la bonne biodisponibilité orale [19].

Les propriétés Veber des composés (1-20) se trouvent dans le tableau 10.

Tableau.10 : Les règles de Veber des dérivés de coumarine.

Composés	PSA	Nrotb	%ABS
01	35,53	3	96,742
02	44,76	4	93,558
03	35,53	3	96,742
04	35,53	4	96,742
05	35,53	3	96,742
06	35,53	4	96,742
07	35,53	7	96,742
08	35,53	9	96,742
09	35,53	7	96,742
10	55,40	2	89,887
11	69,67	3	84,964
12	78,90	4	81,780
13	26,30	2	99,927
14	55,76	3	89,763
15	35,53	3	96,742
16	59,32	3	88,535
17	35,53	3	96,742
18	55,76	3	89,763
19	44,76	5	93,558
20	59,32	3	88,535

NRB : Le nombre des liaisons rotatives, PSA : La surface polaire , %ABS = $109 \pm 0,345x$ PSA. [16]

Pour la règle de Veber, nous remarquons que :

- Les valeurs de nombre de liaisons rotatives sont toutes inférieures à 10. Le faible nombre de liaisons rotatives (flexibilité réduite) dans les composés étudiés indique que ces ligands lors de la liaison avec une protéine ne changent que légèrement leur conformation.

➤ Pour les résultats de PSA, nous remarquons que tous les composés de la série étudiée ont des valeurs inférieures à 140 \AA^2 ce qui montre la bonne prédiction de la biodisponibilité orale et le transport à travers les membranes.

Nous pouvons conclure que ces composés étudiés sont en accord avec la règle de Veber.

On a aussi les valeurs de pourcentage d'absorption (%ABS), on peut observer que tous les composés présentent un grand pourcentage d'absorption allant 81.78% à 96.742%. Cela signifie que ces composés ont une bonne perméabilité membranaire cellulaire.

III.6.3. Efficacité de ligand « LE »

La lipophilie est une propriété physico-chimique qui joue un rôle fondamental dans la détermination des propriétés de l'ADME (absorption, distribution, métabolisme et excrétion). La lipophilie est en corrélation avec un trop grand nombre d'autres propriétés, telles que le stockage dans les tissus, la biodisponibilité, la perméabilité, la toxicité, le volume de distribution, la liaison aux protéines plasmatiques et la liaison aux récepteurs des enzymes [20, 21]. Le plus petit composé tend à avoir les meilleures propriétés physico-chimiques et un bon ADME en ce qui concerne l'efficacité du ligand [22, 23].

Dans cette partie, nous avons étudié l'efficacité du ligand (Tableau 11) pour pénaliser les grands composés sur des petits composés ayant une puissance similaire, car les composés en plus grande taille ont des propriétés physicochimiques et ADME plus faibles [23]. L'efficacité du ligand (LE) est le plus souvent définie comme le rapport de l'affinité d'un ligand divisé par le nombre d'atomes lourds (non hydrogène) dans la molécule.

Ce rapport a été décrit pour la première fois en 1999 [24] et a gagné une popularité répandue dans les cercles de découverte de médicaments peu de temps après [23].

$$LE = \frac{1,4 \text{ pIC}_{50}}{NH}$$

Où: NH est le nombre d'atomes lourds.

$$\text{pIC}_{50} = -\log(\text{IC}_{50})$$

Tableau.11: L'efficacité de ligand des dérivés de Coumarine

Composés	PIC50	NH	LE
01	7,26	19	0,535
02	5,17	21	0,345
03	7,74	20	0,542
04	5,86	23	0,357
05	8,36	21	0,557
06	8,25	22	0,525
07	8,46	25	0,474
08	8,46	27	0,439
09	7,30	27	0,379
10	6,72	22	0,428
11	5,28	23	0,321
12	4,77	25	0,267
13	7,55	21	0,503
14	5,51	22	0,351
15	8,06	22	0,513
16	7,64	23	0,465
17	8,36	22	0,532
18	8,01	22	0,510
19	7,94	26	0,428
20	7,97	23	0,485

D'après les résultats du tableau 11 on remarque que :

- Les valeurs de l'efficacité de ligand (LE) diminue avec le nombre croissant d'atomes lourds autrement dit l'obtention d'une efficacité de ligand élevé nécessite des composés avec des atomes lourds faibles et ça peut être expliqué par la corrélation qui existe entre la taille du composé et leur propriétés physicochimique.

III.6.4. Efficacité lipophile de ligand (LLE)

D'autre part nous avons étudié l'efficacité lipophile (LipE) pour maximiser la puissance tout en maintenant la lipophilie aussi faible que possible, en raison de l'association entre une lipophilie élevée et plusieurs problèmes, y compris une faible solubilité, perméabilité membranaire, stabilité métabolique, etc [25].

Cependant, pour obtenir des propriétés ADMET optimales, la taille moléculaire et la lipophilie sont des facteurs importants. Pour faciliter l'optimisation de l'affinité vis-à-vis de la lipophilie,

Leeson et Springthorpe [22] ont défini l'efficacité du ligand-lipophilie (LLE), également connu sous le nom efficacité lipophylique (Lipophilic Efficiency LipE):

$$LLE = pIC_{50} - \log P$$

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

Tableau.12: Efficacité lipophilique de ligand pour les dérivés de coumarine

Composés	PIC50	Logp	LipE (LLE)
01	7,26	0,34	6,92
02	5,17	-0,66	5,83
03	7,74	0,49	7,25
04	5,86	0,91	4,95
05	8,36	0,77	7,59
06	8,25	1,02	7,23
07	8,46	2,35	6,11
08	8,46	3,15	5,31
09	7,30	2,19	5,11
10	6,72	0,01	6,71
11	5,28	-0,09	5,37
12	4,77	-1,08	5,85
13	7,55	1,89	5,66
14	5,51	-0,26	5,77
15	8,06	0,92	7,14
16	7,64	0,49	7,15
17	8,36	0,92	7,44
18	8,01	-0,26	8,27
19	7,94	1,71	6,23
20	7,97	0,49	7,48

D'après nos résultats, nous avons constaté que :

- Les composés 1, 2, 7, 8, 9, 10, 11, 12, 13, 14 et 19 se situent dans la plage de 5 à 7. Pour les composés restants, il a été constaté que LipE prenait des valeurs supérieures à la plage décrite. Parmi ces molécules, la plus hydrophile est le composé 18, qui présente la plus haute valeur de LipE (8,27).

III.6.5. Indice de lipophilicité dépendant de l'efficacité des ligands (LELP)

Pour obtenir des propriétés ADMET optimales, la taille moléculaire et la lipophilie sont des facteurs importants à prendre en compte. Keseru et Makara [26] proposent un indice de lipophilicité dépendant de l'efficacité des ligands (LELP) pour combiner la taille moléculaire et la lipophilie en une seule mesure d'efficacité. L'intervalle LELP optimal est $-10 < \text{LELP} < 10$.

$$\text{LELP} = \frac{\log p}{LE}$$

Tableau. 13: Indice de lipophilie des dérivés de coumarine

Composés	Logp	LE	LELP
01	0,34	0,535	0,64
02	-0,66	0,345	-1,91
03	0,49	0,542	0,90
04	0,91	0,357	2,55
05	0,77	0,557	1,38
06	1,02	0,525	1,94
07	2,35	0,474	4,96
08	3,15	0,439	7,18
09	2,19	0,379	5,78
10	0,01	0,428	0,02
11	-0,09	0,321	-0,28
12	-1,08	0,267	-4,04
13	1,89	0,503	3,76
14	-0,26	0,351	-0,74
15	0,92	0,513	1,79
16	0,49	0,465	1,05
17	0,92	0,532	1,73
18	-0,26	0,510	-0,51
19	1,71	0,428	4,00
20	0,49	0,485	1,01

D'après nos résultats (tableau 13), nous constaté que:

- Tous les composés se situent dans la plage suggérée $-10 < \text{LELP} < 10$.
- De l'autre, tous les composés ont un LELP inférieur à 10, ce qui signifie que ces composés sont situés dans la zone de Lipinski.

- [01] *HyperChem (Molecular Modeling System) Hypercube, Inc., 1115NW, 4th Street, Gainesville, FL 32601, USA, 2008.*
- [02] *Gaussian 09, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian Inc., Wallingford, CT (2010).*
- [03] *C. M. Breneman and K. B. Wiberg, Journal of Computational Chemistry, 1990, 11, 361–373.*
- [04] *MarvinSketch. ChemAxon (<http://www.chemaxon.com>).*
- [05] *Spss 20 For Windows, SPSS software packages, SPSS Inc., 444 North Michigan Avenue, Suite 3000, Chicago, Illinois, 60611, USA. <http://www.spss.com>.*
- [06] *W.A. Wakeham, G.S. Cholakov, and R.P. Stateva, —Liquid density and critical properties of hydrocarbons estimated from molecular structure^{el}, Journal of Chemical & Engineering Data, 47, 2002, 559–570.*
- [07] *B.P. Ajeet, K. Bijander. Quantitative structure activity relationship (QSAR) modeling of 2-x-5,9-dimethoxy-1,4-naphthoquinones against 11210 cells. Int. J. Pharm. & Pharm. Sci., 2012, 4, 445-448.*
- [08] *M. Clark, R.D. Cramer. The probability of chance correlation using partial least squares (PLS). Quant. Struct.-Act. Relat., 1993, 12, 137-45.*
- [09] *S. Wold, L. Eriksson. Statistical validation of QSAR results. Validation tools. Methods and Principles in Medicinal Chemistry, 1995, 2, 309-18.*
- [10] *H. Van der Voet. Comparing the predictive accuracy of models using a simple randomization test. Chemometr. Intell. Lab., 1994, 25, 313-23.*

- [11] S.O. Podunavac-Kuzmanović, D.D. Cvetković, D.J. Barna. *QSAR analysis of 2-amino or 2-methyl-1-substituted benzimidazoles against pseudomonas aeruginosa*. *Int. J. Mol. Sci.*, 2009, 10, 1670-1682.
- [12] H. Gonzalez-Diaz, E. Olazabal, L. Santana, E. Uriarte, Y. Gonzalez-Diaz, N. Castanedo. *QSAR study of anticoccidial activity for diverse chemical compounds: Prediction and experimental assay of trans-2-(2-nitrovinyl) furan*. *Bioorg. Med. Chem.*, 2007, 15, 962-968.
- [13] M. Jalali-Heravi, A. Kyani. *Use of computer-assisted methods for the modeling of the retention time of a variety of volatile organic compounds*. *J. Chem. Inf. Comput. Sci.*, 2004, 44, 1328-1335.
- [14] C. Lipinski, F. Lombardo, B. Dominy, P. Feeney. *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*. *Adv. Drug Deliv. Rev.*, 2001, 46, 3-26.
- [15] D. F. Veber, S. R. Johnson, H.Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple. *Molecular properties that influence the oral bioavailability of drug candidates*. *J. Med. Chem.* 2002, 45, 2615-23.
- [16] D. M. Segall. *Multi-parameter optimization: identifying high quality compounds with a balance of properties*. *J. Curr. Pharm. Des.*, 2012, 18, 1292-1310.
- [17] G. Vistoli, A. Pedretti, B. Testa. *Assessing drug-likeness – what are we missing?*. *Drug Discov. Today*, 2008, 13, 285-294.
- [18] M. Aurélien. *Création et utilisation de chimiothèques optimisées pour la recherche « in silico » de nouveaux composés bioactifs*. Thèse de Doctorat. Université d'ORLEANS, 2006.
- [19] M. Remko, M. Swart, F. M. Bickelhaupt. *Theoretical study of structure, pKa, lipophilicity, solubility, absorption, and polar surface area of some centrally acting antihypertensives*. *J. Bioorg. Med. Chem.*, 2006, 14, 1715-1728.
- [20] H. Patrick, W. Langenacker, J. Tollenaere. *Computational Medicinal Chemistry for Drug Discovery*, marceldekker, new York, 1995.
- [21] V. Pliska, B. Testa, H. van de Waterbeemd, R. Mannhold, H. Kubinyi, H. Timmerman. *Lipophilicity in drug action and toxicology*, Wiley-VCH, Federal Republic of Germany, 1996.
- [22] P.D. Leeson, B. Springthorpe. *The influence of drug-like concepts on decision making in medicinal chemistry*, *Nat. Rev. Drug. Discov.*, 2007, 6, 881-890.

- [23] A.L. Hopkins, C.R.Groom, A.Alexander. *Ligand efficiency: a useful metric for lead selection. J. Drug. Discov.*, 2004, 9,430-1.
- [24] I.D. Kuntz, K. Chen, K.A. Sharp, P.A. Kollman. *The maximal affinity of ligands. Proc. Natl Acad. Sci., USA* 1999, 96(18), 9997–10002.
- [25] J. D. Hughes, J. Blagg, D. A. Price, S. Bailey, G. A. Decrescenzo, R. V. Devraj, E. Ellsworth, Y. M. Fobian, M. E. Gibbs, R. W. Gilles. *Physicochemical drug properties associated with in vivo toxicological outcomes. J. Bioorg. Med. Chem. Lett.*, 2008, 18, 4872-4875.
- [26] G.M. Keserü, G.M. Makara, *The influence of lead discovery strategies on the properties of drug candidates. Nat. Rev. Drug Discov.* 2009, 8, 203–212.



Conclusion Générale

Dans ce travail nous avons effectué une étude QSAR sur une série de molécule de coumarine par l'utilisation de méthode statistique d'analyse. Le but de cette étude est d'établir une relation linéaire quantitative structure-activité. Cette étude à montre que l'activité de coumarine est liée à l'effet d'accepteur de liaison hydrogène l'effet de densité, l'effet de E_{LUMO} et la charge atomique de carbone 6 qui ont grande influence sur l'activité inhibitrice de monoamine oxydase (MAO-B). Ce modèle QSAR peut être utilisé pour le développement ultérieur de nouvelles molécules afin de présenter une bonne activité inhibitrice des enzymes.

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.

Par comparaison, des résultats de calcul et les critères de la règle de Lipinski on a remarqué que tous les composés de la séries confirment aux ces critères, ils n'ont pas des problèmes d'absorption ou de perméabilité et ils sont acceptable d'être administrer par voie orale. Notre travail est ensuite porté l'étude la règle de Veber explique que tous les composés de la série ayant le nombre des liaisons inférieure à 10, et aussi que tous les composés de notre série avec des valeurs de la surface polaire sont en accord au la règle 2^{ème} de Veber, c'est-à dire tous les valeurs de la surface polaire inférieur à 140 \AA^2 , et cette résultat qui montre une bonne corrélation avec le transport moléculaire passive à travers les membranes. Donc, il permet l'estimation des propriétés de transport des médicaments.

Pour les résultats reliés de l'efficacité de ligand, on a trouvé une grande valeur de LE est de (0.557) pour le composé 5 avec NH égale 21, donc cette valeur de LE explique que ce composé a tendance à avoir les meilleures propriétés physico-chimiques et ADME.

L'étude de l'efficacité lipophile dans la série des de dérivés de coumarine, à partir de nos résultats on a trouvés les composés (1, 2, 7, 8, 9, 10, 11, 12, 13, 14 et 19) qui se situé dans la gamme suggérée de 5-7, ce qui indique que ce composé été optimisé avec succès. En fin en termine par l'étude de l'indice de lipophilicité dépendant de l'efficacité des ligands (LELP) nous constaté que tous les composés se situent dans la plage suggérée - $10 < LELP < 10$.

Résumé

L'objectif principal de cette étude était de développer un modèle QSAR pour vingt nouvelles séries de dérivés de la coumarine en tant qu'inhibiteur de la monoamine oxydase-B qui peuvent être des agents potentiels pour traiter la maladie de parkinson. Le modèle QSAR implique l'utilisation de la méthode d'analyse MLR par régression linéaire multiple à l'aide du logiciel SPSS. Le pouvoir prédictif du modèle obtenu a été confirmé par $R = 0,945$, $R^2 = 0,893$, $S = 0,451$ et $F = 31\ 446$. La méthode de validation croisée LOO nous montre 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.

L'application des règles de Lipinski, des règles de Veber et des indices de lipophilie sur les dérivés de coumarine étudiés montre que la plupart de ces composés n'auront théoriquement pas de problèmes de biodisponibilité orale.

Mot-clés: Coumarine, QSAR, MLR, Inhibition MAO-B, indices de lipophilie.

Abstract

The main objective of this study was to develop a QSAR model for twenty new series of coumarin derivatives as a monoamine oxidase-B inhibitor that may be potential agents for treating Parkinson's disease. The QSAR model involves the use of MLR multiple linear regression analysis using the SPSS software. The predictive power of the obtained model was confirmed by $R = 0.945$, $R^2 = 0.893$, $S = 0.451$ and $F = 31\ 446$. The LOO cross validation method shows a strong correlation was observed between experimental and predicted values of biological activities. , which indicates the validity and quality of the QSAR models obtained.

The application of Lipinski rules, Veber rules and lipophilicity indices on the coumarin derivatives studied shows that most of these compounds will not have problems with oral bioavailability.

Keywords: Coumarin, QSAR, MLR, MAO inhibitor, Lipophilicity indices.