



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

UNIVERSITE ABOU-BEKR BELKAID - TLEMCCEN

MEMOIRE

Présenté à

FACULTE DES SCIENCES – DEPARTEMENT DE CHIMIE

Pour l'obtention du diplôme de :

MASTER

Filière: **Chimie**

Option: Chimie Théorique et Modélisation

Par :

M^r YOUSFI YUCEF

Sur le thème

ETUDE QSAR DE L'ACTIVITE ANTI-OXYDANTE D'UNE SERIE DE COMPOSES PHENOLIQUES

Soutenu publiquement le 19 Juin 2017 à Tlemcen devant le jury composé de :

Melle NEGADI Latifa	Professeur	Université de Tlemcen	Président
Mr CHARIF Imad Eddine	Maître de Conférences B	Université de Tlemcen	Examineur
Mme ERRAHOUÏ- BELLIFA Khadidja	Maître de Conférences B	Université de Tlemcen	Encadreur

*Laboratoire de Thermodynamique Appliquée et Modélisation Moléculaire (LATA2M), N° 53
BP 119, 13000 Tlemcen - Algérie*

A la mémoire de ma grand-mère

A mes chers parents

A mes sœurs et frères

Remerciements

Le travail présenté dans ce mémoire a été réalisé dans le Laboratoire de Thermodynamique Appliquée et Modélisation Moléculaire (LATA2M) de l'université A. Belkaid de Tlemcen dirigé par Mademoiselle Pr Latifa NEGADI.

J'adresse mes chaleureux remerciements à Mme Dr K. ERRAHOUÏ-BELLIFA pour m'avoir confié le sujet de ce travail. Je tiens vivement à la remercier pour son investissement et sa célérité pour son attention à se soumettre à mes questions pressantes, ses réponses sont désormais partie intégrante de ce mémoire. Je la remercie pour la grande autonomie qu'elle m'a accordé. J'espère m'être montré digne de sa confiance.

J'adresse ma plus grande et respectueuse reconnaissance au Pr Latifa NEGADI pour avoir accepté de présider le jury et pour le temps consacré à l'examen de ce mémoire.

Je tiens également à remercier Monsieur Dr Imad Eddine CHARIF pour m'avoir fait l'honneur de juger ce travail et d'avoir participer au jury.

Grand merci revient de droit à Monsieur Pr Sidi Mohamed MEKELLECHE qui a inculqué en nous l'amour de la science. Grand respect pour votre personne.

Un autre grand merci à Mme Dr W. BENCHOUK et Mme Dr H. CHEMAÛRI pour leur disponibilité à répondre à tout questionnement.

Je ne remercierai jamais assez Dr Mohammed BENABDALLAH pour son soutien inestimable. Grace à vous on en est là.

Merci aux responsables de la faculté des sciences de l'université A. Belkaid de Tlemcen pour m'avoir offert cette chance.

Liste des abréviations

AM1	Austin Model 1
B3LYP	Becke 3-Parameter Lee-Yang-Parr
CLOA	Combinaison Linéaire d'Orbitales atomique
DA	Domaine d'Applicabilité
DPPH	2,2-diphényle-1-picrylhydrazyl
DFT	Density Function Theory
DM	Dipôle Moment
ERO	Espèce Réactive Oxygénée
GGA	Generalized Gradient Approximation
GTO	Gaussian Type Orbital
HF	Hartree-Fock
HOMO	Highest Occupied Molecular Orbital
KS	Kohn et Sham
LDA	Local Density Approximation
LMO	Leave Many Out
LOO	Leave One Out
MLR	Multiple Linear Regression
MSD	Mulliken Spin Density
NHFO	Numerical Hartree fock Orbital
OA	Orbitale Atomique
OCDE	Organisation de Coopération et de Développement Economiques
OM	Orbitale Moléculaire
ORAC	Oxygene Radical Absorbance Capacity
PCA	Principal composante Analysis
PLS	Partial Least Squares
PM3	Parametric Method 3
PM6	Parametric Method 6
QSPR	Quantitative Structure-Property Relationships
QSAR	Quantitative Structure-Activity Relationships
RL	Radical Libre
SCF	Self Consistant Field
STO	Slater Type Orbital
SLR	Simple Linear Regression
SSR	Sum of Squares Regression
SSE	Sum of Squares Error
TEAC	Trolox Equivalent Antioxydant Capacity

SOMMAIRE

INTRODUCTION GENERALE

CHAPITRE I : GENERALITES-ACTIVITE ANTIOXYDANTE

Introduction	5
I.1. Oxydants et problèmes des radicaux libres.....	5
I.2. Classes des radicaux libres.....	5
I.3. Productions des espèces réactives à l'oxygène ERO.....	8
I.3.1 Production endogène.....	8
I.3.2. Productions exogène.....	8
I.4. Les antioxydants.....	8
I.4.1. Antioxydants enzymatiques.....	9
I.4.2. Antioxydants non enzymatiques.....	9
I.5. Les composés phénoliques et leur activité anti-oxydante.....	12
I.5.1. Classification des composés phénoliques.....	12
I.5.2. Modes d'action anti-oxydante des phénols.....	15
I.6. Méthodes de quantification de l'activité anti-oxydante.....	16
I.6.1. Test DPPH.....	16
I.6.2. Test TEAC.....	17
I.6.3. Test ORAC.....	17
I.6.4. Potentiel Redox.....	17
I.7. QSAR et l'activité anti-oxydante des composés phénoliques.....	18
Références Bibliographiques	20

CHAPITRE II : BASES THEORIQUES

I. METHODES QSPR/QSAR

1. Introduction à la modélisation QSPR/QSAR	21
2. Principe des méthodes QSAR/QSPR	21
3. Méthodologie générale d'une étude QSPR/QSAR	21
3.1. Bases de données	23
3.2. Descripteurs moléculaires.....	23
3.2.1. Les descripteurs moléculaires théoriques.....	23
3.2.2. Les descripteurs moléculaires empiriques	24
3.3. Méthodes d'analyse des données :	25
3.4. Interprétation et validation d'un modèle QSPR/QSAR	26
3. 4. 1.Validation interne	26
3. 4. 2. Validation externe	27
3.4. 3. Domaine d'applicabilité	28
Références Bibliographiques	29

II. METHODES QUANTIQUES

Introduction	30
1. Méthode de Hartree-Fock-Roothaan.....	31

1.1. Approximation du champ moyen de Hartree	32
1.2. Méthode de Hartree-Fock	33
1.3. Méthode de Hartree-Fock-Roothaan	33
2. Méthodes Post-SCF	35
2.1. Méthode d'interaction de configuration (CI)	35
2.2. Méthode de Möller-Plesset d'ordre 2 (MP2)	36
3. Théorie de la fonctionnelle de densité (DFT)	38
3.1. Fondement de la théorie DFT :	38
3.2 Méthode de Kohn et Sham :	39
3.3 Approximation de la densité locale LDA :	41
3.4 Méthode $X\alpha$:	42
3.5 Approximation de la densité de spin locale LSDA :	42
3.6. Approximation du Gradient Généralisé (GGA) :	43
3.7 Fonctionnelle hybride B3LYP : ..	43
3.8. Processus SCF d'un calcul DFT :	43
4. Bases d'orbitales atomiques :	44
5. Les méthodes semi-empiriques	46
5.1. MNDO	47
5.2. AM1.....	47
5.3. PM3.....	47
5.4. PM6.....	48
Références Bibliographiques :	49
III. METHODES D'ANALYSE DES DONNES	
Introduction.....	50
1. Régression linéaire.....	50
1.1- Méthode d'estimation des paramètres β_0, β_1	52
1.2- Décomposition de la variance et qualité de la régression.....	54
1.3. Représentation de l'analyse de la variance :	56
1.4. Qualité de la régression linéaire.....	56
1.4.1 Coefficient de détermination R^2	56
1.4.2 Déviation standard (SD).....	57
1.5. Confiance à accorder aux résultats.....	57
1.5.1 Test de Fisher-Snedecor.....	57
1.5.2 Test de Student.....	58
1.6. Sélection de variables et choix du modèle.....	58
1.6.1. Coefficient de détermination ajusté $R^2_{\text{ajusté}}$	58
1.6.2. Critère de validation croisée : PRESS.....	59
1.6.3. Cross validated R^2	59
Références Bibliographiques.....	60
CHAPITRE III : RESULTATS ET DISCUSSION	
Introduction.....	61
1. Méthodologie.....	61
1.1. Base de données.....	61

1.2. Calcul des descripteurs.....	63
1.3. Analyse statistique.....	64
2. Résultats et Discussion.....	64
2.1. Modèles QSAR avec un seul descripteur.....	65
2.2. Modèles QSAR avec deux descripteurs.....	65
2.3. Modèles QSAR avec trois descripteurs.....	67
2.4. Etude du meilleur modèle.....	68
2.4.1. Validation du meilleur modèle QSAR.....	70
2.4.2. Domaine d'applicabilité du meilleur modèle QSAR.....	74
2.4.3. Interprétation mécanistique du meilleur modèle.....	75
Conclusion.....	75
Références Bibliographiques.....	76
CONCLUSION GENERALE.....	77

Introduction générale

Introduction générale

L'activité anti-oxydante demeure un axe très dynamique de la recherche scientifique multidisciplinaire. En 1956, Harman [1] a émis une hypothèse faisant le lien entre les radicaux libres générés dans l'organisme responsables de l'endommagement de nombreux composants cellulaires tels que les lipides, les protéines et les acides nucléiques ... causant des signes de vieillissement précoce. Depuis, le lien entre la propagation des radicaux libres communément connus sous le nom des Espèces Réactives Oxygénées (ERO) [2] et certaines pathologies plus moins graves ne cesse de se confirmer, dont les maladies cardiovasculaires [3], les maladies inflammatoires [4], et certains cancers [5], et d'autres.

Le corps humain se dispose de son propre dispositif de défense contre la génération et la propagation des ERO. Ce sont des systèmes enzymatiques endogènes comme le glutathion peroxydase, superoxyde dismutase, les catalases et autres [6]. Néanmoins, une génération excessive de ces espèces radicalaires prédatrices peut causer un déséquilibre entraînant un stress oxydatif. Pour la détoxification, le corps a besoin de certains composés connus par leur activité antioxydante apportée essentiellement par l'alimentation.

Les antioxydants sont indispensables pour la lutte contre les ravages du stress oxydatif. Ce sont des composés exogènes, essentiellement d'origine végétale. Les plus connus sont le β -carotène Vitamine A, l'acide ascorbique vitamine C, les tocophérols vitamine E, ainsi que les dérivés phénoliques [7].

Les composés phénoliques sont des substances organiques caractérisés par la présence d'au moins un motif phénolique dans leurs structures chimiques. Il existe de nombreuses structures, ils sont divisés en plusieurs classes, les flavonoïdes, les acides phénoliques, ... ils sont distribués dans les fruits (ex : pomme, Kiwi, ...), les légumes (ex : brocoli, artichaut...), les huiles (ex : huile d'olive, huile d'argan). Ils sont des antioxydants puissants, leur activité anti-oxydante est attribuée i) à la capacité de piéger les ERO, ii) inhibition de la peroxydation des lipides iii) chélation les métaux et enfin ils peuvent

inhiber les enzymes impliqués dans le stress oxydant. Ils sont aussi utilisés dans l'industrie agroalimentaire comme additifs et conservateurs contre l'oxydation [8].

En s'appuyant sur cette vision, il devient indispensable de renforcer l'effort de recherche pour l'étude de l'activité anti-oxydante de ces composés dans le but de d'identifier les paramètres responsables de cet effet antioxydant et de comprendre le mécanisme avec lequel un antioxydant peut inhiber un radical libre et pour chercher d'améliorer le pouvoir antioxydant de ces composés.

La modélisation QSAR (Quantitative Structure-Activity Relationship) constitue actuellement un outil fiable et performant pour combler les lacunes des études *in vivo* et *in vitro* et limiter le recours aux tests biologiques. C'est une méthode qui tente à identifier une relation quantitative entre l'activité mesurée expérimentalement et des descripteurs théoriques ou empiriques liés à la structure moléculaire.

L'activité anti-oxydante a été discutée dans plusieurs ouvrages [9, 10-13], et de nombreux auteurs ont essayé de construire des modèles QSAR afin de corréliser l'activité anti-oxydante des composés phénoliques avec des descripteurs physico-chimiques calculées. Cependant, plusieurs travaux présentent certains inconvénients, tels que l'utilisation arbitraire de méthodes théoriques pour calculer les propriétés moléculaires, les critères d'OCDE (Organisation de Coopération et de Développement Economiques) ne sont pas respectés (validation et domaine d'applicabilité du modèle QSAR).

Dans ce contexte, l'objectif de ce travail est d'élaborer des modèles QSAR robustes et fiables, en respectant toutes les étapes d'une étude QSAR pour expliquer et prédire l'activité anti-oxydante d'une série constituée de 41 composés phénoliques en utilisant des descripteurs ciblés et pertinents.

Ce mémoire se divise en trois chapitres :

- Dans le premier chapitre nous exposons le problème des radicaux libres, les différents antioxydants, les composés phénoliques et les études QSAR en activité anti-oxydante.

- Le deuxième chapitre englobe une synthèse bibliographique nécessaire pour aboutir à une étude QSAR, ce chapitre est divisé en trois parties :
 - ✓ La modélisation QSPR/QSAR
 - ✓ Les méthodes quantiques.
 - ✓ Les méthodes statistiques et analyse des données
- Dans le troisième chapitre nous présentons et nous discutons les résultats de notre étude QSAR de l'activité anti-oxydante de 41 composés phénoliques.
Nous terminerons par une conclusion générale et les perspectives envisagés pour ce travail.

Références Bibliographiques

- [1] D. Harman, J. Gerontol, Aging : a theory based on free radical and radiationchemistry, 1956, 11, 298-300.
- [2] J. R. Roede, D.P. Jones, Environmental and Molecular Mutagenesis, 2010, 51, 380.
- [3] K. Sachidanandam, S. C. Fagan, A. Ergul, Cardiovascular Drug Reviews, 2005, 23, 115.
- [4] P. G. Winyard, D. R. Blake, C. H. Evans, Ed. Springer Basel AG, 2000. 11.
- [5] S. Kawanishi, Y. Hiraku, M. Murata, S. Oikawa, Am. cancer Soc., 2002, 32, 822.
- [6] F. ZERARGUI, Thèse-Doctorat, 2015, Université F. A. Sétif
- [7] R. Ksouri, W. Megdiche, A. Debez, H. Falleh, C. Grignon, C.Abdelly, Physio. Biochem., 2007, 45, 244.
- [8] A. Sarkar, T. R. Middy, A. Dipnagar, J. Mol. Model., 2012, 18, 2621.
- [9] V. Rastija, M. Medic, S. Eur, J. Med. Chem., 2009, 44, 400.
- [10] E. Sergediene, K. Jonsson, FEBS Lett., 1999, 462, 393.
- [11] D. Amic, D. Davidovic, D. Amic, N. Trinajstic, Croat. Chem. Acta., 2003, 76, 55.
- [12] D. Amic, D. Davidovic, V. Rastija, N. Trinajsti, C. Med. Chem, 2007, 14, 827.
- [13] M. Filipovic, Z. Markovic, C. R. Chimie, 2015, 18,492-498.

CHAPITRE I

Activité Anti-oxydante/ Généralités

Introduction

L'humain comme tout organisme aérobie, utilise l'oxygène pour fabriquer de l'énergie. Les tests biologiques sur une cellule de rat [1] lors de son métabolisme normal, métabolise 10^{12} molécules d'oxygènes (O_2) dont 98% subissent une réduction normale en eau. Les 2% restant subissent une réduction incomplète et engendrent une quantité d'anions superoxydes et peroxydes d'hydrogène connu comme espèces réactives de l'oxygène ERO. Sous un équilibre fragile ces espèces avec des enzymes appropriés jouent un rôle fondamental dans la défense contre les bactéries mais quand les moyens de défenses sont dépassés, les ERO participent à la progression de diverses pathologies plus ou moins graves.

I. 1. Oxydants et problèmes des radicaux libres

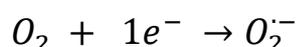
Les ERO espèces réactives oxygénées incluant les radicaux libres de l'oxygène tel que ; anion superoxyde $O_2^{\cdot-}$, radical hydroxyle $\cdot OH$, monoxyde d'azote $NO\cdot$ Et certain dérivé oxygéné non radicalaire comme H_2O_2 peroxyde d'hydrogène et peroxyde d'azote $ONOO\cdot$ [2]. Le radical libre qui est une entité chimique neutre possédant au moins un électron célibataire non apparié sur sa couche externe. Un radical libre est instable et donc très réactif avec les molécules environnantes qui tente de se stabiliser soit en acceptant un électron soit par le transfert de cet électron libre vers une autre molécule qui devient à son tour instable déclenchant ainsi une réaction en chaîne.

I.2. Classes des radicaux libres

➤ Radicaux libres primaires

Les radicaux libres primaires sont formés directement à partir de l'oxygène [3] on peut citer

- Anion superoxyde $O_2^{\cdot-}$: Un RL chargé négativement résultant de la réduction monovalente de l'oxygène moléculaire O_2 .



Sa durée de vie est assez longue lui permettant d'être propagé dans l'organisme loin de son lieu de production. Il peut être à la fois oxydant et réducteur. L'anion superoxyde est neutralisé par l'enzyme superoxyde dismutase.

- Radical hydroxyle $\cdot\text{OH}$: Produit essentiellement à partir des anions superoxydes et des peroxydes d'hydrogènes, supporté par des ions ferriques selon la réaction de Haber Weiss



Le radical hydroxyle est un oxydant très puissant capable de réagir avec presque tous les composants cellulaires, par un échange d'électron, addition sur des double liaisons ou arrachement électron ou d'un atome d'hydrogène. En biologie il est considéré comme L'ERO le plus toxique [4].

➤ Radicaux libres secondaires

Ce sont des radicaux libres générés indirectement sous l'action oxydante des RL primaires anion superoxyde ou radical hydroxyle, sur une composante cellulaire (acide aminé, acide gras polyinsaturé, protéine, ou acide nucléique). Ces RL sont connus pour être moins réactifs que les primaires mais plus sélectifs.

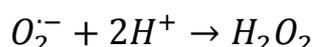
Radicaux alkyles $R\cdot$ les radicaux Alkoxydes $RO\cdot$ et Alkoperoxydes $ROO\cdot$. Le radical peroxyde $ROO\cdot$ est formé suite à la réaction de peroxydation lipidique d'un RL primaire $O_2^{\cdot-}$ ou $\cdot\text{OH}$ sur un acide gras insaturé.

➤ Espèces réactives de l'oxygène ERO non radicalaires

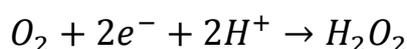
Sont considérés comme des précurseurs.

- Peroxyde d'hydrogène H_2O_2 : Ce n'est pas un radical libre au sens propre puisqu'il ne contient pas d'électrons libres mais il entre dans la catégorie des dérivés actifs de l'oxygène.

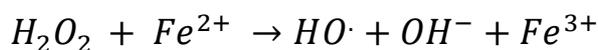
Il est généré par la réaction spontanée, ou induite de l'anion superoxyde



Ou par la réduction bivalente de l'oxygène moléculaire



Les peroxydes d'hydrogène ont la capacité de générer des radicaux peroxyyles en présence des cations métalliques Cu^{2+} et Fe^{2+} (réaction de Fenton) [5].



- **L'oxygène singulet $^1\text{O}_2$** : la molécule d'oxygène singulet est une forme excitée de O_2 . Il est généré selon plusieurs processus dont l'action du rayonnement UV, la peroxydation lipidique ou encore des réactions enzymatiques. C'est un agent oxydant puissant capable d'attaquer les acides nucléiques les protéines et les lipides.

- **Autres** : d'autres ERO existent sous forme non radicalaire qui sont peu réactifs tel que l'ozone, le peroxy-nitrite ONOO^- , l'acide hypochloreux HOCl [6]. Ils sont plutôt précurseurs d'autres radicaux libres.

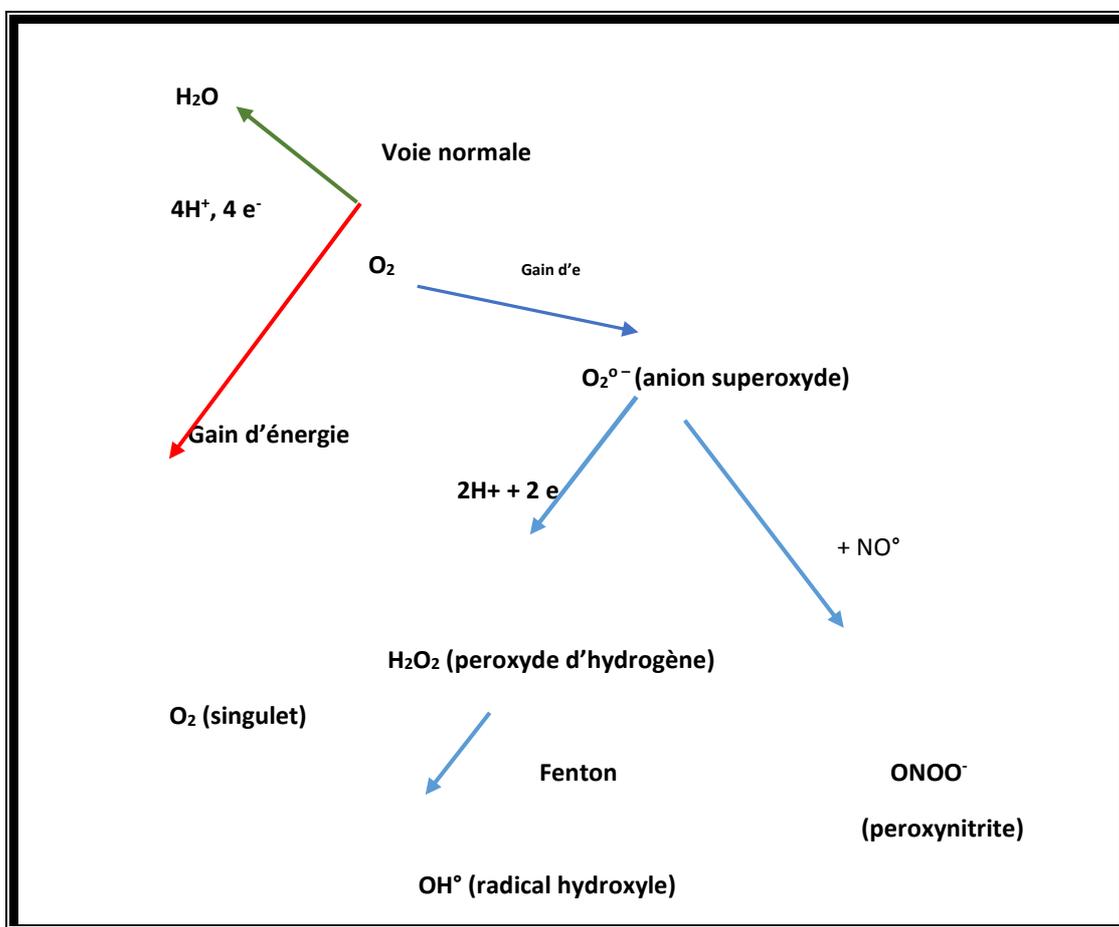


Figure.1 : Formation des Espèces Oxygénées Réactives (EOR)

I.3. Productions des espèces réactives à l'oxygène ERO

Il existe deux grandes catégories de production des radicaux libres.

I.3.1 Production endogène

Les ERO sont produits de manière naturelle au sein de l'organisme au cours de la respiration cellulaire et au cours de processus biologique comme la réponse immunitaire de défense contre les bactéries et corps étrangers.

Au cours du processus respiratoire cellulaire, des anions peroxydes sont produits résultants de la réduction incomplète de l'oxygène moléculaire en eau au niveau des mitochondries [6].

I.3.2. Production exogène

Des éléments externes peuvent contribuer à la génération des radicaux libres et entraînent ainsi plus de dommages à l'organisme ; Une mauvaise hygiène de vie, l'alcool, le tabagisme actif ou passif, le stress émotionnel. Il existe aussi des agents physiques comme la chaleur les radiations, les rayons du soleil causant des brûlures de peau, le rayonnement UV les ultrasons et les champs électromagnétiques et d'autres polluants.

I.4. Les antioxydants

Selon Halliwell et Gutteridge 1989 [7] un antioxydant est défini comme étant une substance qui, présente à faible concentration par rapport à un substrat oxydable, retarde ou prévient significativement l'oxydation de ce substrat. Les antioxydants par leurs pouvoirs d'offrir leurs propres électrons sont capables de **ralentir ou d'inhiber l'oxydation des ERO** et de détruire les radicaux libres propagés dans l'organisme. Ils constituent une ligne de défense pour les cellules pour maintenir des concentrations non cytotoxiques de radicaux libres.

Une forme endogène d'antioxydants sous forme d'enzymes permette au corps de réguler constamment la détoxification des cellules. Une autre forme exogène apportée par l'alimentation est nécessaire pour maintenir l'équilibre.

I.4.1. Antioxydants enzymatiques

Différents systèmes enzymatiques sont connus pour leur rôle antioxydant :

- **Superoxyde dismutase SOD** qui catalyse la dismutation de l'anion superoxydes en oxygène moléculaire et peroxyde d'hydrogène.
- **Catalase** une enzyme qui intervient dans la décomposition de la molécule peroxyde d'hydrogène en eau et oxygène moléculaire.
- **Glutathion peroxydase**, avec l'apport du Sélénium la glutathion peroxydase peut réduire le peroxyde d'hydrogène et les hydroperoxydes lipidiques.

I.4.2. Antioxydants non enzymatiques

Constituent un groupe de molécules exogènes apportées essentiellement par l'alimentation. Certains de ces composés ont le pouvoir de complexer des métaux de transition et empêcher la production des radicaux libres ce sont des chélateurs métalliques. D'autres antioxydants agissent en captant les électrons célibataires des radicaux libres et les transformer en molécules neutres ou ions stables ce sont des piègeurs. Différents vitamine, composés phénoliques et

➤ **Vitamine C** : Un puissant antioxydant connu sous le nom d'acide ascorbique. La vitamine C intervient aussi dans la protection des vitamines liposolubles A et E et les acides gras polyinsaturés de l'oxydation.

➤ **Vitamine E** : Connue sous le nom des α -tocophérols. Elle possède un pouvoir antioxydant très intéressant. La nature lipidique de l' α -tocophérol lui donne la capacité d'agir sur le site de la peroxydation lipidique membranaire. Le radical tocophéroxyde généré est peu réactif stabilisé de par sa structure cyclique. La vitamine E agit conjointement avec la vitamine C pour donner lieu à une chaîne de détoxications des radicaux libres.

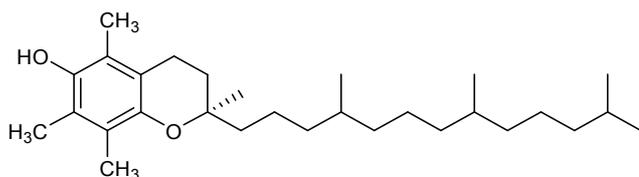


Figure.2 : Structure d'une des formes de la vitamine E, l' α -tocophérol.

➤ **Composés phénoliques** : Ce sont des composés organiques qui font partie des métabolites secondaires des plantes. Ils sont caractérisés par la présence d'un ou plusieurs groupes phénoliques. Leur pouvoir antioxydant varie en fonction de leurs structures et leurs caractéristiques physico-chimiques.

➤ **Métaux** : des métaux oligo-éléments comme le Cuivre le Manganèse et le Zinc, sont indispensables pour nombreuses réactions enzymatiques inhibitrices des ERO. Leur présence joue un rôle important dans la protection des cellules contre les stress oxydatifs.

➤ **Antioxydants synthétiques** : il existe d'autres composés synthétiques qui possèdent un pouvoir antioxydants. Ce sont des dérivés phénoliques synthétiques hautement actifs. Leur utilisation en excès peut être toxique.

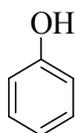
I.5. Les composés phénoliques et leur activité anti-oxydante

Plusieurs études ont montré que les composés phénoliques possèdent une activité biologique intéressante liée à leurs propriétés anti-oxydantes et anti-radicalaires [8], grâce à la mobilité de l'hydrogène phénolique. Ils sont capables de piéger les différents radicaux générés en permanence par l'organisme et ou causés par des agressions externes (polluants, alcool, tabac, infection, ...). Cette activité anti-oxydante permet à l'organisme de se protéger contre les agressions de l'oxygène qui sont à l'origine de plusieurs maladies. Ce qui suscite un intérêt croissant à étudier l'activité de ces composés dans la prévention et le traitement de nombreuses maladies comme, le cancer, les maladies cardiovasculaires, les inflammations, l'Alzheimer, parkinson et autres.

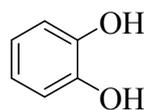
I.5.1. Classification des composés phénoliques

➤ Les phénols simples

Les composés phénoliques simples comprennent un noyau aromatique avec un ou plusieurs substituants hydroxylés.



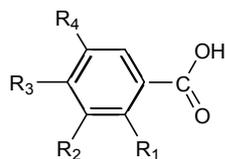
Phénol



Catéchol

➤ Les acides phénoliques dérivés de l'acide benzoïque

Tableau 2 : Structure des dérivés de l'acide benzoïque

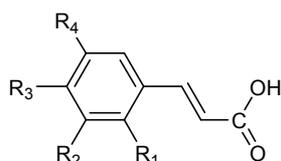


<i>Acides Benzoliques</i>	R1	R2	R3	R4
<i>Ac. Salicylique</i>	OH	H	H	H
<i>Ac. P-hydroxy</i>	H	H	OH	H
<i>-benzoïque</i>	H	OH	OH	OH
<i>Ac. Gallique</i>	H	OH	OH	H
<i>Ac. Protocatechique Ac. Vanilique</i>	H	OCH ₃	OH	H

➤ Les acides phénoliques dérivés de l'acide cinnamique (C₆-C₃)

Ces composés phénoliques comprennent une structure en commun constituée d'un noyau aromatique, un ou plusieurs substituants hydroxyles et une chaîne de 3 carbones avec un groupe carboxyle.

Tableau.3 : Structure des dérivés de l'acide cinnamique

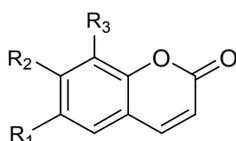


<i>Acides cinnamiques</i>	R1	R2	R3	R4
<i>Ac. cinnamique</i>	H	H	H	H
<i>Ac. o-coumarique</i>	OH	H	H	H
<i>Ac. m-coumarique</i>	H	OH	H	H
<i>Ac. P-coumarique</i>	H	H	OH	H
<i>Ac. caféique</i>	H	OH	OH	H

➤ Les coumarines

Dérivés des acides cinnamiques ortho-hydroxylés. Les coumarines et leurs dérivés ont la capacité de capter les ERO. Ils préviennent aussi la peroxydation lipidique de membranes [9].

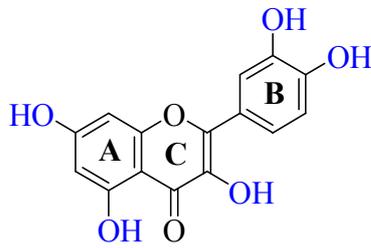
Tableau 4 : Structure des dérivés de coumarines



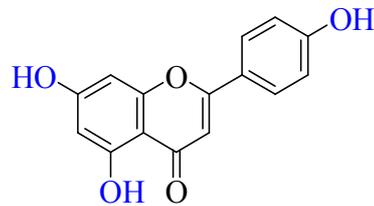
<i>Coumarines</i>	R1	R2	R3
<i>Umbelliférol</i>	H	OH	H
<i>Aescultol</i>	OH	OH	H
<i>Scopolétol</i>	OCH ₃	OH	H
<i>Fraxétol</i>	OCH ₃	OH	OH
<i>Daphnétol</i>	H	OH	OH

➤ Les flavonoïdes

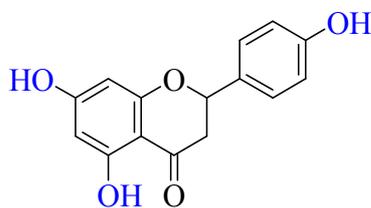
Leurs structures constituées d'une squelette de base de 15 carbones formée de deux noyaux aromatiques à 6 carbones A et B reliés par une chaîne de 3 carbone souvent sous forme d'hétérocycle C [10]. Les substituants sur l'hétérocycle C donnent les différents sous-groupes des flavonoïdes.

Flavonols

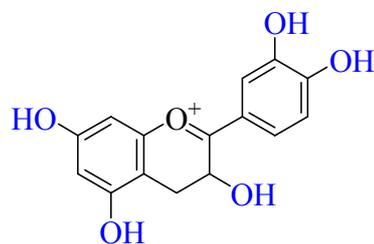
Ex. Quercétine

Flavones

Ex. Apigénine

Flavanones

Ex. Naringinine

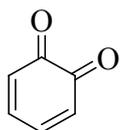
Anthocyanidine

Ex. Cyanidine

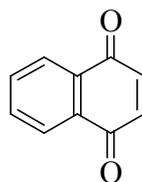
Les flavonoïdes sont connus pour leurs capacités de protéger l'organisme contre les effets délétères des agressions oxydantes environnementales. Par leurs potentiels redox faibles ils sont capables de réduire les RL. Ce sont aussi de bons chélateurs d'ions métalliques.

➤ **Les quinones**

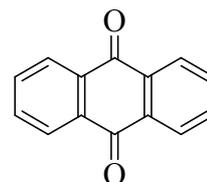
Les quinones sont des composés aromatiques contenant un noyau de benzène avec deux substitutions cétones. Elles sont attachées aux dérivés phénoliques simples. Elles ont des caractéristiques intéressantes dont le transport d'électrons dans les membranes mitochondriales.



1,2 benzoquinone



naphthoquinone



anthraquinone

➤ Les tanins

Les tanins sont des substances polyphénoliques avec une structure complexe constituée d'unités répétitives monomérique. Les tanins ont la capacité de fixer les protéines et par conséquent ont le pouvoir d'inhiber la peroxydation lipidique. Il existe deux classes de tanins :

- Les tanins condensés formés de polymères flavonoïques.
- Les tanins hydrolysables ce sont des polyesters d'acides phénoliques et de glucides.

I.5.2. Modes d'action anti-oxydante des phénols

Plusieurs modes d'actions sont connus pour l'activité anti-oxydante des composés phénoliques :

- Le piégeage direct des radicaux libres.
- Chélation des ions métalliques.
- L'inhibition des enzymes responsables de la génération des ERO.

a. Piégeage direct des radicaux libres

Par leur faible potentiel redox les composés phénoliques peuvent réduire rapidement des radicaux libres comme l'anion superoxyde $O_2^{\cdot-}$, les alkoxyles $RO\cdot$, les peroxydes $ROO\cdot$ et les hydroxyles $\cdot OH$ par transfert d'hydrogène.



b. Chélation des ions métalliques

Les acides phénoliques et les flavonoïdes sont connus pour leur capacité à chélater les ions des métaux de transition comme le cuivre et le fer, en formant des complexes ou des composés de coordination avec ces ions. Les travaux menés par Van Acker et al. [11] ont mis en évidence 3 sites essentiels pour les chélation comme points d'attache pour chez les flavonoïdes à savoir :

- La structure catéchol (3,4 hydroxy) sur le cycle B
- La structure 4-céto, 3-hydroxy sur le cycle C
- La structure 4-céto et 5-hydroxy entre les cycles A et C [12].

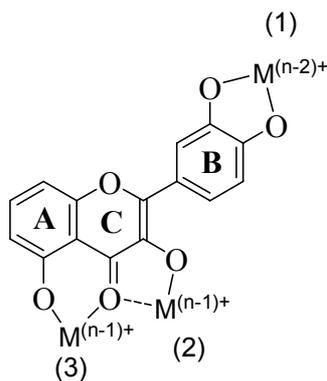


Figure: Complexation d'ions métalliques par les flavonoides.

c. L'inhibition des enzymes responsables de la génération des ERO

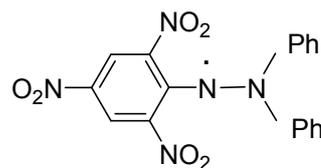
Les composés phénoliques révèlent aussi un pouvoir inhibiteur envers une grande variété d'enzyme. Cos et al. en 1998 [13] ont établi une relation entre la structure de quelques flavonoids et leurs capacité d'inhiber la generation des radicaux libres superoxydes par la formation de complexes enzyme-inibiteur avec la xanthine oxydase.

I.6. Méthodes de quantification de l'activité anti-oxydante

Compte tenu de la nature diversifiée des antioxydants contenant des composants à la fois hydrophiles et hydrophobes, et la complexité des processus d'oxydations, il n'existe pas une méthode universelle pour laquelle l'activité anti-oxydante peut être mesurée quantitativement d'une façon précise. Plusieurs méthodes expérimentales sont élaborées pour évaluer l'activité anti-oxydante [14, 15].

I.6.1. Test DPPH

C'est une méthode largement utilisée pour l'évaluation expérimentale de l'activité anti-oxydante. Elle est basée sur la réduction du radical libre DPPH. C'est un radical stable centré sur l'Azote. Sa réduction par l'antioxydant donneur d'atome H, notamment les composés phénoliques, conduit à la formation du DPPH-H (composé incolore) et au radical AH·.

Figure : Structure DPPH[•]

La quantification du pouvoir antioxydant en EC₅₀ correspond à la détermination de la concentration en antioxydant nécessaire pour réduire 50% de la concentration initiale en DPPH[•].

I.6.2. Test TEAC

La méthode TEAC (Trolox Equivalent Antioxydant Capacity) est basée quant à elle sur la capacité de l'antioxydant à inhiber le cation radical ABTS^{•+} comparativement à un antioxydant de référence, le trolox de structure similaire à la vitamine E. Le ABTS^{•+} de coloration bleu-vert en contact avec un donneur de H[•] conduit au cation incolore ABTS⁺. Un EC₅₀ est déterminé.

I.6.3. Test ORAC

La méthode ORAC (Oxygene Radical Absorbance Capacity) est assez récente, appliquée pour la mesure de la capacité antioxydante aussi bien pour les composés hydrophiles que lipophiles. Elle est basée sur le principe de la décroissance de la fluorescence de la fluorescéine en présence du générateur de radicaux peroxydes AAPH. La capacité antioxydante est évaluée en fonction de la cinétique de dégradation de la fluorescence de la fluorescéine en fonction du temps. Le processus de dégradation stoppé ou ralenti indique un pouvoir croissant du composé testé.

I.6.4. Potentiel Redox

Le potentiel redox est une mesure électrochimique du pouvoir antioxydant exprimée en volt. C'est une technique utilisée dans l'évaluation du pouvoir antioxydant du phénol et ses analogues substitués. En présence de radicaux libres (agents oxydants) les composés phénoliques sont convertis en radicaux phenoxyles. Expérimentalement [16] la méthode de radiolyse à impulsion est utilisée pour produire un équilibre entre les phénols et les

phénoxyles à un pH donné. Cet équilibre est suivi cinétiquement en utilisant la méthode de détection optique pour déterminer le potentiel redox.

Tableau 5 : Potentiel redox [17]

Composé	$E_0(v)$	Famille
myricétine	0,712	flavonols
quercétine	0,872	
kaempférol	0,929	
épicatéchine	0,937	flavanols
catéchine	0,957	
naringénine	0,961	flavanone



Le pouvoir antioxydant décroît avec l'augmentation du potentiel redox.

I.7. QSAR et l'activité anti-oxydante des composés phénoliques

Les antioxydants phénoliques sont d'un grand intérêt en raison de leur implication dans des processus biologiques et industriels importants. En général, les composés ayant une activité anti-oxydante ont été trouvés qu'ils possèdent d'autres activités, anticancéreuse, anti-cardio-vasculaire, anti-inflammation et de nombreuses autres activités [18]. Les calculs théoriques utilisant des méthodes de chimie quantique telle que la DFT ou les méthodes semi-empiriques ont été utilisées pour élucider et confirmer la relation structure-activité anti-oxydante des composés phénoliques [19-32].

Plusieurs chercheurs ont essayé d'élaborer des relations quantitatives structure-activité (QSAR) des antioxydants en utilisant différents paramètres physico-chimiques comme la chaleur de formation (ΔH_f), le nombre de groupe hydroxyle (OH), la constante de Hammet (σ), le potentiel d'ionisation (IP), l'enthalpie de dissociation de la liaison O-H (BDE), etc. [18]. Lien et al. [29] ont rapporté un modèle QSAR pour l'étude de l'activité anti-

oxydante d'une série de phénols en utilisant des descripteurs calculés par la méthode semi-empirique AM1. Dans cette étude, quatre paramètres, à savoir $\Delta H_f(r)$ (chaleur de formation du radical phénoxy), $\Delta H_f(p)$ (chaleur de formation du phénol parent), OH (nombre de groupes hydroxyle) et LUMO(r) (LUMO du radical), étaient responsables de l'activité anti-oxydante. Néanmoins, les méthodes semi-empiriques approchent certaines propriétés jusqu'à un certain niveau, et il est évident que la variation des valeurs d'énergie LUMO de divers congénères est très faible, donc une erreur importante peut être introduite si un procédé quantique de bas niveau est utilisé dans les calculs [18]. D'autre part, Wright et al. [19] ont fait une étude QSAR de l'activité anti-oxydante d'une série de flavonoïdes. Ils ont utilisé la DFT pour le calcul des descripteurs. Ils ont constaté que ces antioxydants agissent soit par transfert d'atomes d'hydrogène, pour lesquels le calcul de BDE (enthalpie de dissociation de la liaison) est pertinent, ou par transfert d'un seul électron, pour lequel le calcul de l'IP (potentiel d'ionisation) est adéquat. Très récemment, Jeremic et al. [32] ont élaboré un modèle QSAR pour 21 composés phénoliques, les descripteurs utilisés sont de trois types : thermodynamiques, NPA, indices d'aromaticité. Un coefficient de corrélation $R= 0,97$ est obtenu, la stabilité et le pouvoir prédictif n'ont pas été discutés.

Tous les travaux QSAR qui ont été effectués ont tenté de stabiliser un modèle fiable capable d'expliquer le mécanisme par lequel un antioxydant peut désactiver un radical libre. Cependant, plusieurs modèles QSAR ne répondent pas entièrement aux critères de l'OECD (Organisation de coopération et de développement économique) concernant les principes de validation. En plus il a été constaté que c'est difficile de stabiliser un mécanisme vu la différence des bases de données, multiplicités des méthodes de quantification expérimentale de l'activité anti-oxydante ainsi la nature complexe de ce processus antioxydant [20, 31].

Références Bibliographiques

- [1] B. N. Ames, M. K. Shigenaga, T. M. Hagen, Proc. Natl. Acad. Sci. **1993**, 90, 7915.
- [2] G. P. Novelli. J. Physiol. Pharmacol. **1997**, 48, 517.
- [3] Oldham, Review. **1998**, 98.
- [4] L. Castro, B. A. Freeman. Nutrition. **2001**, 17, 295.
- [5] G. K. Lopes, H. M. Schulham, M. Hermes-Lima. J. Agr. Food Chem. **1999**, 74, 1749.
- [6] K. M. Oldham, P. E. Bowen, J. Am. Diet. Assoc. **1998**, 98, 1001.
- [7] B. Halliwell, J. M. C. Gutteridge. Anal. Biochem. **1987**, 1, 215.
- [8] J. Kanner, E. Frankel, R. Granit, B. German, J. E. Kinselva. **1994**, 42, 64.
- [9] J. Bruneton, Pharmacognesie, Phytochimie, Plantes médicinales. Lavoisier Technique & Documentation, 4^{ème} édition. Paris
- [10] J. B. Harborne, Phytochemical methods. A guide to modern techniques of plants analysis. Third Edition. 1998.
- [11] S. A. Van Acker, D. G. Van Den Berg, Tromp MN, D. H. Griffioen, W. P. Van Bennekom, W. J. Van Der Vijgh, A. Bast. Free Radic. Biol. Med. **1996**, 20, 331.
- [12] C. A. Rice-Evans, N. J. Miller, G. Paganga. Free Radic. Biol. Med, **1996**, 20, 933.
- [13] P. Cos, L. Ying, M. Calomme, J. H. Hu, K. Cimanga, B. Van Poel, L. Pieters, A. J. Vlietinck, D. Vanden Berghe. J. Nat. Prod, **1998**, 61, 71.
- [14] P. Cristina, S. Ilonka, T. Bartek, Rev. Génie Ind. **2009**, 4, 25.
- [15] R. Apak, S. Gorinstein, Pure Appl. Chem. **2013**, 85, 957.
- [16] E. Rozoy, M. Araya-Farias, S. Simard, D. Kitts, J. Lessard,. Food Chem. **2013**, 138, 1982.
- [17] A. Fougerousse, Bull.Un.Phy. **1996**, 781, 319.
- [18] M. Reis, B. Lobato, J. Lameira, A.S. Santos, C.N. Alves. Eu. J. Med. Chem. **2007**, 42, 440.
- [19] a- J. S. Wright, D. J. Carpenter, D. J. McKay, K. U. Ingold, J. Am. Chem. Soc. **1997**, 119, 4245.
b- J. S. Wright, E. R. Johnson, G. A. DiLabio, J. Am. Chem. Soc. **2001**, 123, 1173.
- [20] D. Amic, D. Davidovic, V. Rastija1, N. Trinajsti, C. Med. Chem, **2007**, 14, 827.
- [21] M. Leopoldini, I. P. Pitarch,; N. Russo,; M. Toscano, , J. Phys. Chem. A. **2004**, 108, 92.
- [22] M. Guerra, R. Amorati, G. F. Pedulli, J. Org. Chem. **2004**, 69, 5460
- [23] K. I. Priyadarsini, D. K. Maity, G. H. Naik, M. S. Kumar, M. K. Unnikrishnan, J. G. Satav, H. Mohan. Free Rad Biol & Med. **2003**, 35 , 475.
- [24] N. Russo, M. Toscano, N. Uccella. J. Agr. and Food Chemistry. **2000**, 48 , 3232.
- [25] P. Trouillas, C. Fagnere, R. Lazzaroni, C. Calliste, A. Marfak, J. L. Duroux. Food Chem. **2004**, 88, 571.
- [26] J. S. Wright, E. R. Johnson, G. A. DiLabio. J. Am. Chem. Soc. **2001**, 123 , 1173.
- [27] Zhang, H. Y.; Sun, Y. M.; Chen, D. Z. Quant. Struct-Act. Rel. **2001**, 20, 148.
- [28] H. Y. Zhang, Y. M. Sun, X. L. Wang. Chem. Eur. J. **2003**, 9, 502.
- [29] E. J. Lien, S. Ren, H. H. BUI, R. Wang. Free Radic. Biol. Med. **1999**, 26, 285.
- [30] M. Leopoldini, T. Marino, N. Russo, M. Toscano, , T. Chem. Ac. **2004**, 111 , 210.
- [31] M. Filipovic, Z. Markovic, J. Dorovic, J. Dimitric, C. Markovic, C. R. Chem. **2015**, 18, 492
- [32] S. Jeremic, S. Radenkovic, M. Filipovic, M. Antic, A. Amic, Z. Markovic, J. Mol. Graph. Model., **2017**, 01, 11.

CHAPITRE II

Bases Théoriques

CHAPITRE II

Bases Théoriques: Méthodes QSPR/QSAR

1. Introduction à la modélisation QSPR/QSAR

Un modèle QSPR/QSAR est une formule mathématique qui permet de relier, d'une manière quantitative, la structure d'une molécule à une propriété ou à une activité donnée.

Les méthodes QSPR/QSAR sont anciennes, elles remontent au 19^{ème} siècle [1,2].

En effet, les premiers travaux utilisant la méthodologie QSPR/QSAR telle qu'employée actuellement sont dus à Hansch [3] et Free et Wilson [4]. D'un côté, des modèles reliant directement l'activité biologique des composés avec les propriétés hydrophobes, électroniques et stériques à l'échelle moléculaire ont été proposés par Hansch. D'un autre côté, des modèles empiriques, dits de contributions de groupes, pour l'étude de l'activité biologique ont été développés par Free et Wilson.

A nos jours, l'utilisation des méthodes QSPR/QSAR n'a pas cessé de progresser [5-7].

2. Principe des méthodes QSPR/QSAR

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

$$\text{Propriété/ Activité} = f(D_1, D_2, \dots, D_n, \dots) \quad (1)$$

D_1, D_2, \dots, D_n sont des descripteurs des structures moléculaires.

3. Méthodologie générale d'une étude QSPR/QSAR

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

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

- Validation externe : on utilise la série de test constituée généralement de 1/3 de la base de données. Le but de cette validation est de vérifier le pouvoir prédictif du modèle élaboré.

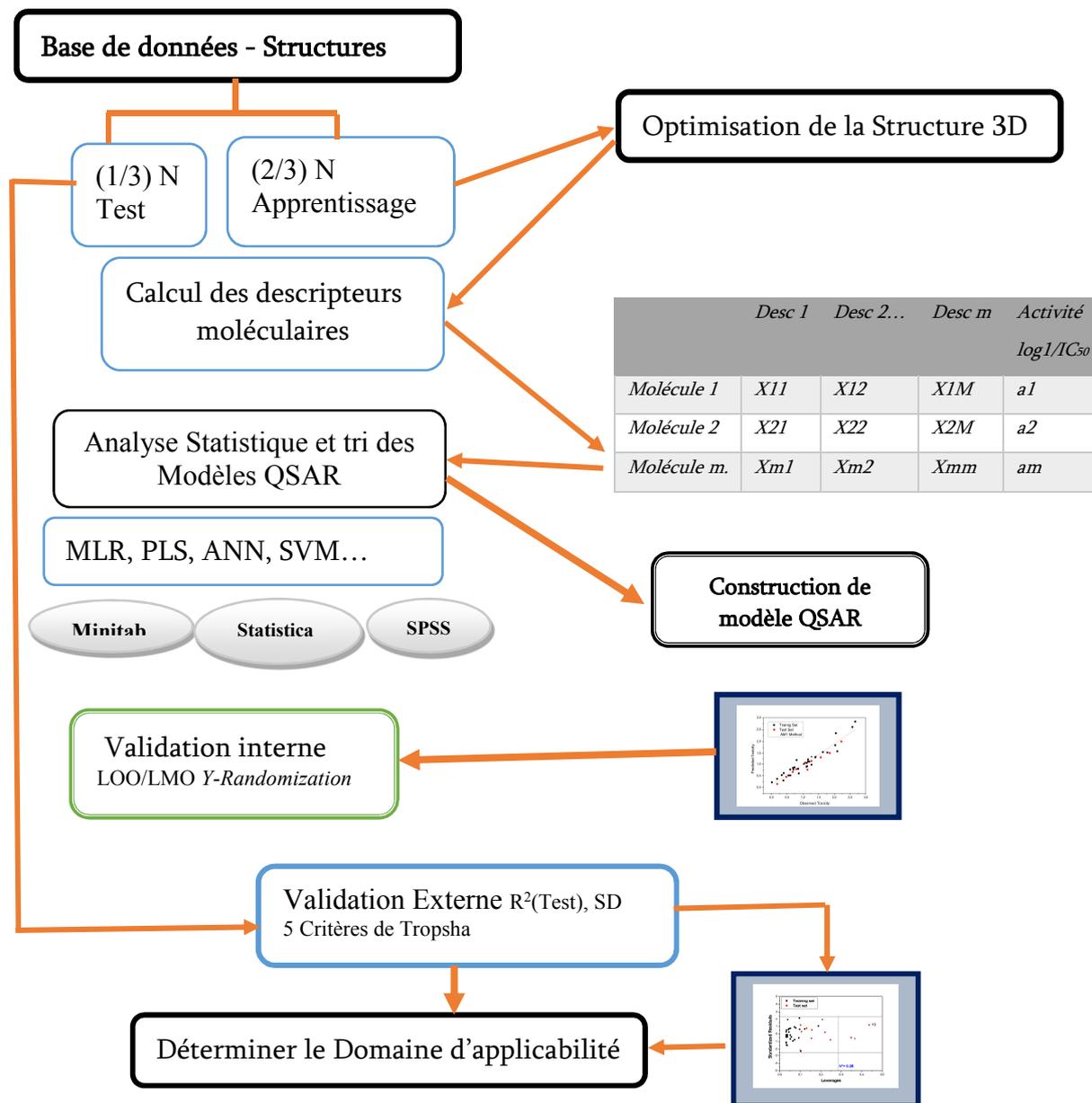


Figure.1 : Procédure d'obtention et de validation d'un modèle QSPR/QSAR

3.1. Bases de données

Un modèle QSPR/QSAR est très dépendant des données expérimentales. Le choix de la base de données est décisif dans le développement de tel modèle. Une base de données doit être composée de données expérimentales aussi fiables que possible obtenues en suivant un protocole unique puisque les erreurs sur celles-ci se propageront dans le modèle final. Plusieurs bases de données sont disponibles sur le net, on peut citer :

- ECHA : <http://echa.europa.eu/fr>
- Oracdatabase : <http://oracdatabase.com/list-of-foods/search/>
- TOXNET...

3.2. Descripteurs moléculaires

Définition d'un descripteur : c'est l'information contenue dans la structure d'une molécule.

3.2.1. Les descripteurs moléculaires théoriques

Les descripteurs 1D : sont remontés directement à partir de la formule brute de la molécule. Exemples : le nombre d'atomes et la masse moléculaire, ...etc.

Les descripteurs 2D : la formule développée d'une molécule peut nous offrir ce type de descripteurs. On distingue :

- **Les indices 2D constitutionnels** :
- **Les indices 2D topologiques** : Ils donnent des informations sur la taille, la forme globale et les ramifications d'une structure.

Exemples: indice de Wiener [8], indice de Randić [9], indice de connectivité de valence de Kier-Hall [10], indice de Balaban [11], ...etc

Les descripteurs 3D : sont estimés à partir des positions relatives des atomes dans l'espace, et décrivent des caractéristiques plus complexes. Leurs calculs nécessitent donc de connaître la géométrie 3D de la molécule.

- **Les descripteurs 3D géométriques** : les plus importants sont le volume moléculaire, la surface accessible au solvant, le moment principal d'inertie.

- **Les descripteurs 3D électroniques** : permettent de quantifier les différents types d'interactions inter- et intramoléculaires, de grande influence sur l'activité biologique des molécules. Le calcul de la plupart de ces descripteurs nécessite la recherche de la géométrie pour laquelle l'énergie est minimale, et fait souvent appel à **la chimie quantique**.

Exemples :

<i>Descripteur</i>	<i>Définition</i>
E_{LUMO}	Energie de l'orbitale la plus basse vacante.
E_{HOMO}	Energie de l'orbitale la plus haute occupée
μ	Potentiel chimique
η	Dureté chimique
σ	Mollesse chimique
ω	Indice d'électrophilie globale
AEI	Indice de l'énergie d'activation
MSD	Densité de Spin de Mulliken
E_{SOMO}	Energie de l'orbitale occupée par un électron

Tableau.1. Exemples de descripteurs électroniques de la chimie quantique

Descripteurs thermodynamiques :

Les descripteurs thermodynamiques sont calculés sur la base de la fonction de partition totale Q de la molécule. Exemples : l'enthalpie de formation, l'enthalpie libre...

3.2.2. Les descripteurs moléculaires empiriques

Il existe plusieurs descripteurs qui peuvent être mesurés expérimentalement tel que le coefficient de partage (octanol-eau) ($\log P$).

Coefficient de partage (octanol-eau) K_{ow} , $\log P$

Le coefficient de partition est défini comme le rapport de la concentration du soluté dans la phase huileuse C' à la concentration du soluté non ionisé dans la phase aqueuse C , à l'équilibre.

Il est largement utilisé dans des études de relations structure-activité quantitatives (QSARs) dans les sciences pharmaceutiques, biochimiques, toxicologiques et dans les sciences de l'environnement.

Détermination expérimentale de log P

La méthode expérimentale utilisée pour mesurer le coefficient de partage d'une molécule est la méthode des flacons agités ou «*shake-flask*». Cette méthode est préconisée comme procédure standard de caractérisation par l'OCDE [12] (Organisation de Coopération et de Développement Economique).

Méthodes de calcul et d'estimation de log P

Plusieurs méthodes d'estimation de log P ont été décrites dans la littérature depuis plusieurs décennies [13]. Les plus anciennes sont des méthodes fragmentales dans lesquelles une molécule est divisée en fragments prédéfinis et les contributions correspondantes sont sommées pour conduire à une valeur estimée du log P . Parmi ces méthodes on peut citer

- Méthode de Hansch [14], Méthode de Rekker [15], La méthode de Ghose et Viswanadhan [16], La méthode de Klopman et Iroff [17], La méthode de Bodor [18]

Plusieurs logiciels sont utilisés pour l'estimation de log P , on peut citer : ACD/LABS, Clog P, Hyperchem, KOWWIN, ChemSpider (online)

3.3. Méthodes d'analyse des données

Une méthode d'analyse des données est nécessaire pour élaborer un modèle QSPR/QSAR.

Il existe plusieurs méthodes pour construire un modèle et analyser les données statistiques de ce dernier, ces méthodes sont disponibles dans des logiciels tels que, Excel, Origin Microcal, **Minitab**, Statistica, SPSS, R,...

La méthode utilisée dans ce travail est la méthode de Régression Linéaire (SLR/MLR) implémentée sur Minitab et qui est présentée en détail dans la partie 3 de ce chapitre.

3.4. Interprétation et validation d'un modèle QSAR/QSPR

Une fois développé, le modèle doit être validé afin de vérifier sa stabilité, son pouvoir explicatif et son pouvoir prédictif.

3.4.1. Validation interne

Validation croisée, LOO (*leave-one out*) ou LMO (*leave-many out*) : Ce processus consiste à extraire un certain nombre k de molécules du jeu initial à N molécules et à construire un nouveau modèle avec les $(N-k)$ molécules restantes à l'aide des descripteurs choisis (seules les constantes de la régression changent). Ce processus est ensuite réitéré pour retirer et prédire les valeurs de toutes les molécules de la série d'apprentissage. En fonction du nombre de molécules retirées à chaque itération, on parlera de *Leave-One-Out* (LOO) ou de *Leave-Many-Out* (LMO) selon qu'une ou plusieurs molécules est (sont) retirée(s) [19]. Elle est quantifiée par le coefficient R^2_{cv} .

Dans ces dernières années, d'autres méthodes sont utilisées pour faire la validation interne, tel que la hasardisation de la réponse (*Y-Randomization*).

- **Y-Randomization**

Il n'est pas rare d'obtenir des corrélations fortuites (ou « *chance correlation* »), c'est-à-dire un modèle affichant de bons résultats statistiques (R^2 , SD) pour l'apprentissage, mais impliquant des descripteurs qui ne sont pas reliés à la propriété/activité modélisée. Ces modèles aléatoires peuvent être détectés par la procédure *Y-randomization* [20]. Elle consiste à mélanger aléatoirement les propriétés/activités expérimentales pour la série d'apprentissage en utilisant les mêmes descripteurs, de nouveaux modèles sont obtenus. Ces derniers doivent avoir des performances très faibles.

La validation externe du modèle est indispensable pour vérifier le pouvoir prédictif d'un modèle QSPR/QSAR. D'après certains chercheurs, la validation externe est une norme et une partie obligatoire dans la modélisation QSPR/QSAR [21, 22].

3. 4. 2. Validation externe

Cette méthode consiste à prédire la propriété/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 les paramètres $R^2(\text{test})$ et $R_{cv}^2(\text{test})$.

Récemment plusieurs études [23, 24] ont montré l'insuffisance des paramètres $R^2(\text{test})$, $R_{cv}^2(\text{test})$ pour vérifier la capacité prédictive des modèles QSAR. 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 Tropsha » (*Tropsha criteria*) [24].

Critères de validation Externe (série de test)

- $R^2 > 0.7$ (critère 1)
- $R_{cv}^2 > 0.6$ (critère 2)
- $\frac{R^2 - R_0^2}{R^2} < 0.1$ et $0.85 \leq k \leq 1.15$ (critère 3)
- $\frac{R^2 - R_0'^2}{R^2} < 0.1$ et $0.85 \leq k' \leq 1.15$ (critère 4)
- $|R^2 - R_0^2| \leq 0.3$ (critère 5)

Avec

R^2 : coefficient de corrélation pour les molécules de la série de test.

R_0^2 : coefficient de corrélation entre les valeurs prédites et expérimentales pour la série de test.

$R_0'^2$: coefficient de corrélation entre les valeurs expérimentales et prédites pour la série de test.

k : constante de la droite (à l'origine) de corrélation (valeurs prédites en fonction des valeurs expérimentales)

k' : constante de la droite (à l'origine) de corrélation (valeurs expérimentales en fonction des valeurs prédites). La figure 2 montre que $R^2(\text{test})$ élevé ne signifie pas un bon pouvoir prédictif, il faut que le k soit proche de 1 pour juger que le modèle est prédictif.

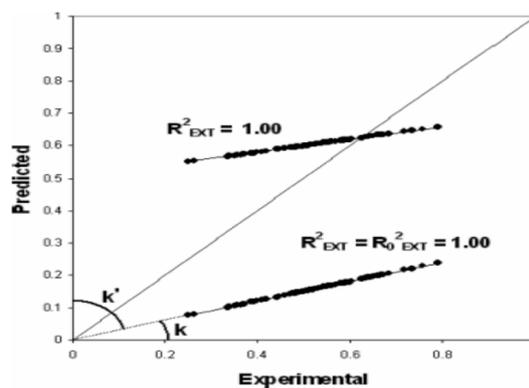


Figure.2 : Importance de la valeur de k dans la validation externe

3.4. 3. Domaine d'applicabilité

Un modèle QSAR/QSPR ne peut pas être considéré comme un modèle universel, parce qu'il est développé sur un nombre limité de composés qui ne couvrent pas tout l'espace chimique. Par conséquent l'activité/propriété prédite d'un composé, chimiquement dissimilaire au jeu d'apprentissage, ne pourra pas être considérée fiable [25]. Le domaine d'applicabilité (DA) permet de définir la zone dans laquelle un composé pourra être prédit avec confiance. Le DA correspond donc à la région de l'espace chimique incluant les composés de la série d'apprentissage et les composés similaires, proches dans ce même espace [26].

Références Bibliographiques

- [1] A. Crum Brown, T.R. Fraser. Trans. Roy.Soc. Edinburgh. **1868**, 25, 151.
- [2] CrumBrown,A.and Fraser,T.R.J.Anat.Physiol.**1868**, 2, 224.
- [3] C. Hansch, A. Leo, D. Hoekmann. American. Chemical. Society. **1995**.
- [4] S.M. Free, J.W. Wilson. J. Med. Chem., **1964**, 7, 395.
- [5] M. Grover, B. Singh, M. Bakshi, S. Singh. Pharm. Sci. Tech. Today. **2000**, 3,50.
- [6] M. Grover, B. Singh, M. Bakshi, S. Singh, Pharm. Sci. Tech. Today, **2000**, 3, 28.
- [7] A.R.D. Katritzky, C. Fara, R.O. Petrukhin, D.B. Tatham, U. Maran, A. Lomaka, M.Karelson, Curr.Top.Med.Chem. **2002**, 2,1333.
- [8] H. Wiener. J. Am. Chem. Soc. **1947**, 69, 17.
- [9] M. Randic. J. Am. Chem. Soc. **1975**,97, 6609.
- [10] L.B. Kier, L.H. Hall, W.J. Murray. J. Pharm. Sci. **1975**, 64,1971.
- [11] A.T. Balaban. Chem. Phys. Lett. **1982**, 89, 399.
- [12] OECD Guidelines for Testing of Chemicals No. 107, OECD,Paris, **1992**.
- [13] B. Lee, F.M. Richards. J. Mol. Biol. **1971**, 55, 379.
- [14] R.S. Pearlman, Eds., Pergamon, New York, **1986**, 3.
- [15] R. F. Rekker, H. M. de Kort. (1979). Eur. J. Med. Chem. – Chim. Ther. **1979**, 14, 479.
- [16] Ghose AK, Crippen GM. J. Chem. Inf. Comput. Sci. **1987**, 27,21.
- [17] G.Klopman, L.D. Iroff, J.Comp.chem. 1981, 2, 157.
- [18] Bodor N, Buchwald P. J. Phys. Chem B. **1997**, 101, 3404
- [19] L. Zhang, H. Zhu, T.I. Oprea, A. Golbraikh, A. Tropsha, Pharm. Res. **2008**, 25, 1902.
- [20] H. Linnan, P.C. Jurs, J. Mol. Graph. Model. **2005**, 23, 503.
- [21] A. Tropsha, P. Gramatica, V.K. Gombar, QSAR Comb.Sci. **2003**, 22, 69
- [22] P.P. Roy, S. Paul, I. Mitra, K. Roy, Molecules, **2009**, 14, 1660..
- [23] A. Golbraikh, A. Tropsha, Beware of q²!, J. Mol. Graph. Model. **2002**, 20, 269.
- [24] T. M. Martin, P. Harten, D. M. Young, E.N. Muratov, A. Golbraikh, H. Zhu and A.Tropsha, J. Chem. Inf. Model. **2012**, 52, 2570.
- [25] I.V. Tetko, I. Sushko, A.K. Pandey, H. Zhu, A. Tropsha, E. Papa, T. Oberg, R. Todeschini, D. Fourches, Varnek A. J. Chem.Inf. Model. **2008**, 48, 1733.
- [26] J. Jaworska, N.N. Jeliaskova, T. Aldenberg Altern. Lab. Anim **2005**, 33, 445.

CHAPITRE II

Bases Théoriques: Méthodes Quantiques

Introduction

La chimie quantique a fait l'objet de nombreux développements de logiciels permettant de réaliser des calculs plus ou moins compliqués sur des systèmes moléculaires, en se basant sur le formalisme de la mécanique quantique et certains niveaux d'approximations. Un calcul de chimie quantique permet d'obtenir des données importantes sur les propriétés du système étudié. Des propriétés d'ordre structural, énergétique, magnétique, électronique... Ce qui mène à bien définir, comprendre et même anticiper la réactivité du système considéré.

Le comportement du système peut être décrit complètement par l'équation de Schrödinger [1].

$$H\Psi = E\Psi \quad (1)$$

$$\text{Avec : } \hat{H} = \hat{T} + \hat{V} \quad (2)$$

Ψ : sont les fonctions propres de H

E : sont les valeurs propres de H

L'hamiltonien complet non relativiste et indépendant du temps décrivant un système isolé constitué de N noyaux et n électrons est constitué de cinq termes :

$$H = -\frac{\hbar^2}{2m_e} \sum_i^n \Delta_i - \frac{\hbar^2}{2M_K} \sum_K^N \Delta_K + \sum_{i>j}^n \frac{e^2}{r_{ij}} + \sum_{K>L}^N \frac{Z_K Z_L e^2}{r_{KL}} - \sum_{K=1}^N \sum_{i=1}^n \frac{Z_K e^2}{R_{Ki}} \quad (3)$$

<i>energie cinétique des electrons</i>	<i>energie cinétique des noyaux</i>	<i>energie de repulsion electrostatique entre electrons</i>	<i>energie de repulsion electrostatistique entre electrons</i>	<i>energie d'attraction electron - noyau</i>
--	---	---	--	--

❖ Approximation de Born-Oppenheimer

L'approximation de Born-Oppenheimer [2] considère que l'on traite uniquement le cas des électrons et supposer les noyaux fixes dans l'espace. Cette approximation est justifiée du fait qu'elle s'appuie sur la grande différence de masse de l'électron 1880 fois plus faible que celle du proton (un électron voit le noyau immobile). Les noyaux sont considérés comme fixe dans l'espace ($T_N=0$) et le traitement se fera uniquement sur les électrons.

L'hamiltonien s'écrit donc :

$$H = T_e + V_{eN} + V_{ee} + V_{NN} \quad (4)$$

Du fait que les noyaux sont considérés fixes, le potentiel électrostatique répulsif V_{NN} est constant. Il peut être traité à part et ajouté à la solution énergétique globale, une fois les solutions électroniques définies.

L'hamiltonien électroniques sera :

$$H_e = T_e + V_{eN} + V_{ee} \quad (5)$$

❖ Approximation des électrons indépendants

Le terme V_{ee} est pratiquement très difficile à calculer. Visant à négliger ce terme dans le traitement du système, on fait appel à l'approximation des électrons indépendants qui considère que les électrons sont indépendants les uns des autres.

Un hamiltonien électronique approché est défini comme suit :

$$H_e = H_e(1) + H_e(2) + \dots + H_e(n) \quad (6)$$

Par conséquent la fonction d'onde totale peut être considérée comme le produit des solutions mono-électroniques de cette équation.

$$\Psi_{\text{Totale}} = \Psi_{(1)} \cdot \Psi_{(2)} \dots \Psi_{(n)} \quad (7)$$

1. La méthode de Hartree-Fock-Roothan/ab-initio

La résolution exacte de l'équation de Schrödinger [1] n'est possible que pour les systèmes mono-électroniques comme l'atome d'hydrogène ou les systèmes hydrogénoïdes. Pour passer aux systèmes poly-électroniques il est indispensable d'utiliser un certain niveau d'approximations pour résoudre l'équation de Schrödinger d'une manière approchée.

1.1. L'approximation du champ moyen de Hartree

Hartree a proposé en 1927 l'approximation du champ moyen [3] dans le but de remplacer le terme d'interaction biélectronique par une interaction de chaque électron avec un champ moyen $U_{(i)}$, constitué de la totalité des autres électrons, dans la résolution de l'équation de Schrödinger. Donc pour un système de n électrons il y aura n fonctions d'ondes monoélectronique $\Psi_{(i)}$, et selon le théorème des électrons indépendants la fonction d'onde totale sera le produit de ces fonctions d'ondes.

$$\Psi_{Tot(1,2,\dots,n)} = \Psi_{(1)} \Psi_{(2)} \dots \Psi_{(n)} \quad (8)$$

L'approximation de Hartree permet d'exprimer l'hamiltonien total du système comme étant la somme des hamiltonien monoélectronique

$$H_{Tot} = -\frac{\hbar^2}{2m} \Delta_i - \sum_{k=1}^n \frac{z_k e^2}{r_{ik}} + U_{(i)} \quad (9)$$

Energie de Hartree :

$$H_{Tot} \Psi_{Tot} = E_{E_{Tot}} \Psi_{Tot} \quad (10)$$

En utilisant la méthode quantique variationnel.

On multiplie à gauche par ψ^* et on intègre :

$$E_{Tot} = \frac{\int \psi^* H \psi d\tau}{\int \psi^* \psi d\tau} \quad (11)$$

Après intégration :

$$E_{Total} = \sum_{i=1}^n H_{ii}^c + \sum_{i=1}^n \sum_{k=1}^n J_{ij} \quad (12)$$

Avec : H_{ii}^c : Intégrale monoélectronique de cœur.

$$H_{ii}^c = \int \psi_{(i)} \left[-\frac{\hbar^2}{2m} \Delta_i - \sum_{k=1}^n \frac{z_k e^2}{r_{ik}} \right] \psi_{(i)} d\tau \quad (13)$$

J_{ij} : Intégrale biélectronique coulombienne.

1.2. Méthode de Hartree-Fock

Selon la méthode de Hartree la fonction d'onde totale (eq:8) est exprimée en fonction des fonctions d'onde de chaque électron occupant chacun sa propre orbitale. Chose qui ne vérifie pas les principes de l'indiscernabilité des électrons, et l'exclusion de Pauli [4].

Fock en 1928 [5] a eu l'idée d'exprimer la fonction d'onde totale sous la forme d'un déterminant de Slater [6] . Chaque orbitale moléculaire peut contenir au maximum 2 électrons (spin α et spin β).

$$\Psi_{(1,2,\dots,n)} = \frac{1}{\sqrt{2^n n!}} \begin{vmatrix} \phi_1 x_1 & \phi_2 x_1 \dots & \phi_{2n} x_1 \\ \phi_1 x_2 & \phi_1 x_2 \dots & \phi_1 x_2 \\ \dots & \dots & \dots \\ \phi_1 x_{2n} & \phi_1 x_{2n} \dots & \phi_1 x_{2n} \end{vmatrix} \quad (14)$$

En exprimant la fonction d'onde avec le déterminant de Slater (Eq :14) et après intégration l'énergie de Hartree-Fock sera :

$$E_{\text{Tot}} = 2 \sum_{i=1}^n H_{ii}^c + \sum_{i=1}^n \sum_{j=1}^n (2J_{ij} - K_{ij}) \quad (15)$$

K_{ij} : Intégrale biélectronique d'échange.

1.3. Méthode de Hartree-Fock-Roothan

La méthode de Hartree-Fock se base sur des orbitales moléculaires réelles c'est une méthode très puissante mais les calculs sont très couteux voir même impossible. Pour remédier à ce point Roothan et Hall [7] ont eu l'idée d'exprimer ces orbitales en une combinaison linéaire d'orbitales atomiques OM-CLOA plus simple à utiliser conduisant à des expressions matricielles plus facile à résoudre.

$$\psi_k = \sum_{r=1}^N C_{kr} \varphi_r \quad (16)$$

En utilisant la méthode vibrationnelle on minimise l'énergie par rapport aux coefficients.

Le système séculaire obtenu est le suivant :

$$\sum_{s=1}^N C_{ks} (F_{rs} - \varepsilon_k S_{rs}) = 0 \quad r = 1, 2, \dots, N \quad (17)$$

N : Nombre d'orbitales atomiques.

S_{rs} : intégrale de recouvrement.

$$S_{rs} = \int \varphi_{r(i)} \varphi_{s(i)} d\tau \quad (18)$$

F_{rs} : élément matriciel de Fock

$$F_{rs} = h_{rs}^c + \sum_{p=1}^n \sum_{q=1}^n P_{pq} \{ 2\langle rs | tu \rangle - \langle rt | su \rangle \} \quad (19)$$

h_{rs} : hamiltonien monoélectronique de cœur

$$h_{rs}^c = \int \phi_{r(i)}^* h^c \phi_{s(i)} d\tau \quad (20)$$

$\langle rs | tu \rangle$: Intégrale biélectronique coulombienne

$$\langle rs | tu \rangle = \iint \varphi_{r(i)} \varphi_{s(i)} \frac{e^2}{r_{ij}} \varphi_{t(j)} \varphi_{u(j)} d\tau \quad (21)$$

$\langle rt | su \rangle$: Intégrale biélectronique d'échange

$$\langle rt | su \rangle = \iint \varphi_{r(i)} \varphi_{t(j)} \frac{e^2}{r_{ij}} \varphi_{s(i)} \varphi_{u(j)} d\tau_i d\tau_j \quad (22)$$

- **Processus SCF (Self Consistant Field) de résolution des équations de Roothan :**

Le problème donc revient à résoudre les équations de Roothan au lieu de résoudre l'équation de Schrödinger.

L'élément matriciel de Fock F_{rs} (Eq :19) dépend des coefficients ce qui rend le système séculaire (Eq :17) non linéaire. Pour résoudre les équations un processus itératif est utilisé :

- ♦ **Initiation** : négliger les coefficients dans F_{rs} .

$$C_{ks}^0 = 0 \Rightarrow F_{rs}^0 = h_{rs}^c \quad (23)$$

$$\sum_{s=1}^N C_{ks}^1 (F_{rs}^0 - \varepsilon_k^1 S_{rs}) = 0 \quad (24)$$

c'est un système linéaire, la résolution donne ε_k^1 et C_{ks}^1 .

- ♦ **Etape 1** : injecter les coefficients C_{ks}^1 dans F_{rs} et obtenir F_{rs}^1
- ♦ **Etape 2** : le processus est répété d'une manière itérative jusqu'à convergence

$$\left| E_{tot}^k - E_{tot}^{k-1} \right| \leq \varepsilon \quad \varepsilon : \text{test de convergence}$$

Donnant lieu à une distribution électronique cohérente correspondant à une énergie minimale invariable.

2. Méthodes post SCF / traitement de la corrélation électronique

L'énergie obtenue après un processus SCF est une énergie approximative car la corrélation électronique n'est pas prise en compte. Les électrons de spin opposés (en particulier ceux situés dans des orbitales ayant des parties spatiales similaires) exercent les uns sur les autres des forces répulsives instantanées dépendant de leurs positions. Cet effet est en partie négligé, puisque l'on suppose que chaque électron se trouve dans le champ moyen créé par tous les autres. C'est une faible énergie mais sa contribution dans l'énergie totale devient importante lorsque de petites énergies doivent être calculées.

D'après Lowdin [8] l'énergie de corrélation électronique est la différence entre l'énergie SCF et l'énergie exacte non relativiste du système. Un traitement de la corrélation électronique plus poussé se révèle donc essentiel pour l'obtention des certaines propriétés du système tel que : l'énergie de dissociation ou encore les propriétés thermo-chimiques. Plusieurs méthodes dites post-SCF ont été développées dont : CI (Configuration Interaction), CC (Coupled Cluster), MPn ou bien la DFT.

2.1. Méthode d'interaction de configuration (CI)

La méthode CI [9] est une méthode à référence unique qui décrit l'état fondamental du système (état fondamental proche de l'équilibre) par une combinaison linéaire de

déterminants de Slater. C'est une combinaison qui représente les différents états d'excitation d'un ou plusieurs électrons des orbitales moléculaires occupées vers les orbitales vacantes.

$$\Psi = \sum_k^A C_k \Phi_k \quad (25)$$

Les déterminants $\Phi_k, k = 1, 2, 3, \dots, A$, déterminent respectivement l'état fondamental et les états excités (mono, bi, tri excités...). Au nombre de configurations considérées. Une combinaison très étendue de déterminants est nécessaire pour obtenir une bonne représentation. Par conséquent une valeur exacte de l'énergie nécessitera une infinité de déterminants.

L'énergie SCF de la méthode HFR correspond à l'état fondamental ou $k=0$.

2.2. Méthode de Möller-Plesset MPn

La méthode de perturbation Möller-Plesset MPn [10] est une adaptation de la théorie plus générale, connue sous le nom théorie de perturbation à plusieurs corps (MBPT), développée par Rayleigh et Schrödinger.

L'hamiltonien total de la méthode Möller-Plesset est séparé en deux. Une partie constituée des fonctions propres et valeurs propres du déterminant de Hartree-Fock, et une partie perturbée V .

$$H = H^0 + \lambda V \quad (26)$$

H^0 : Hamiltonien non perturbé (ordre zéro) issu de la somme des opérateurs de Fock.

$$H^0 = \sum_i F(i) = \sum_i \left\{ h^c(i) + \sum_j [J_j(i) - K_j(i)] \right\} \quad (27)$$

V : la perturbation

λ : paramètre compris entre {0 et 1}; 0 pour un système non perturbé et 1 pour une perturbation maximale.

$$\lambda V = \sum_i \sum_{j>i} \frac{1}{r_{ij}} - \sum_i \sum_j [J_j(i) - K_j(i)] \quad (28)$$

La théorie de perturbation considère que si la perturbation V est petit par rapport à F on peut développer l'opérateur $H = H^0 + \lambda V$ en une série de Taylor selon λ , d'où :

$$E_n = E_n^0 + \lambda E_n^1 + \lambda^2 E_n^2 + \dots \quad (29)$$

$$\Psi_n = \Psi_n^0 + \lambda \Psi_n^1 + \lambda^2 \Psi_n^2 + \dots \quad (30)$$

E_0^i : Sont les corrections énergétiques d'ordre i dans la méthode Möller-Plesset.

La correction énergétique d'ordre n pour E_n^i s'obtient quand a-t-elle en appliquant la méthode de perturbation de Rayleigh et Schrödinger.

La perturbation du second ordre MP2 est la plus couramment utilisée. Elle permet de récupérer une grande partie de l'énergie de corrélation électronique.

- Système non perturbé / ordre 0 :

$$E^0 = \int \Psi^0 H^0 \Psi^0 d\tau \quad (31)$$

- Correction d'ordre 1/MP1 :

$$\begin{aligned} E_{MP1} = E^0 + E^1 &= \int \Psi^0 H^0 \Psi^0 d\tau + \int \Psi^1 H \Psi^1 d\tau \\ &= \int \Psi^0 (H^0 + H^1) \Psi^0 d\tau \\ &= E_{HFR} \end{aligned} \quad (32)$$

La correction d'ordre 1 conduit à l'énergie HRF donc l'énergie de corrélation nécessitée un niveau supérieur.

- Correction d'ordre 2/ MP2 :

$$E_{MP2} = E_n^{HF} + E_n^2 \quad (33)$$

$$\Rightarrow E_{MP2} = \sum_k e_k - \frac{1}{2} \sum_i \sum_j (J_{ij} - K_{ij}) + \frac{1}{4} \sum_r \sum_s \sum_t \sum_u \frac{\langle rs / tu \rangle^2}{e_r + e_s + e_t + e_u}$$

Avec :

$$\langle rs / tu \rangle = \iint \phi_r(1) \phi_s(1) \frac{1}{r_{12}} \phi_t(1) \phi_u(2) d\tau_1 d\tau_2 - \iint \phi_r(1) \phi_s(2) \frac{1}{r_{12}} \phi_u(1) \phi_t(2) d\tau_1 d\tau_2 \quad (34)$$

La méthode MP2 permet des récupérer une grande partie de l'énergie de corrélation électronique. Elle est très efficace et requière dans la pratique des temps de calculs appréciable, proportionnels à N^5 , N étant le nombre d'électrons du système. Les ordres de perturbations supérieurs sont obtenus de la même manière amis sont plus couteux.

3. Théorie de la Fonctionnelle de Densité DFT

3.1. Fondement de la théorie DFT

La théorie de la fonctionnelle de la densité est fondée sur le postulat proposé par Thomas et Fermi. Le postulat stipule que les propriétés électroniques peuvent être décrites en termes de fonctionnelles de la densité électronique en appliquant localement des relations appropriées [11]. Thomas et Fermi ont utilisé leur théorie pour la description de systèmes gaz uniforme d'électrons non interagissant. Le manque de précision et l'impossibilité de traiter des systèmes moléculaires ont en fait un modèle trop simpliste.

Hohenberg et Kohn [12] ont repris en 1964 ont repris la théorie de Thomas-Fermi et ont démontré qu'il existe une fonctionnelle de l'énergie $E[\rho(r)]$ associée à un principe variationnel.

Mathématiquement on désigne par fonctionnelle une entité qui fait correspondre un nombre à chaque fonction provenant d'une classe définie. Autrement dit c'est une fonction de la fonction.

L'objectif principal de la DFT était de remplacer la fonction d'onde poly-électronique de $4N$ variables par la densité électronique avec $f(x,y,z)$ seulement.

❖ Premier théorème de Hohenberg et Kohn

« L'énergie moléculaire, la fonction d'onde et toutes les autres propriétés électroniques de l'état fondamental sont déterminées à partir de la densité de l'état fondamental $\rho_0(x,y,z)$ »(++)

$$\int \rho_0(r) dr = n \quad (35)$$

$\rho_0(r)$ exprime la densité ponctuelle au point r , l'intégrale sur tout l'espace donne le nombre d'électrons.

❖ **Potentiel externe**

L'hamiltonien d'un système polyélectronique est :

$$H = -\frac{1}{2} \sum_i^n \Delta_i + \sum_{i>j}^n \frac{1}{r_{ij}} + \sum_i^n v(r_i) \quad (36)$$

Avec
$$v(r_i) = -\sum_\alpha \frac{Z_\alpha}{r_{i\alpha}} \quad (37)$$

$v(r_i)$: est le potentiel externe de l'électron i.

❖ **Energie**

$$\begin{aligned} E_0 &= E_0[\rho_0] \\ &= V_{ne}[\rho_0] + T[\rho_0] + V_{ee}[\rho_0] \end{aligned} \quad (38)$$

Pour calculer E_0 il faut connaître les 3 fonctionnelles.

$$V_{ne}[\rho_0] = \int \rho_0(r) v(r) dr \text{ Potentiel externe exercé au point } r.$$

La fonctionnelle de l'énergie peut s'écrire comme :

$$E_0[\rho] = \int \rho_0(r) v(r) dr + F[\rho_0] \quad (39)$$

Avec
$$F[\rho_0] = T[\rho_0] + V_{ee}[\rho_0] \quad (40)$$

Dans ce cas la fonctionnelle $F[\rho_0]$ est inconnue.

❖ **Deuxième théorème de Hohenberg et Kohn**

« Pour n'importe quelle fonctionnelle de densité électronique d'essai $\rho'(r)$, qui satisfait les relations $\rho'(r) \geq 0$ et $\int \rho'(r) dr = n$ on aura la relation suivante $E_0[\rho'] \geq E_0[\rho]$ »

C'est un théorème équivalent au principe variationnel. Son importance réside dans le fait qu'il y a une seule densité qui correspond à l'état fondamental le plus stable.

3.2. Méthodologie de Kohn et Sham

Le théorème de Hohenberg et Kohn ne propose pas une méthode pour déterminer la fonctionnelle de densité ρ et calculer ainsi l'énergie E_0 , sans déterminer au préalable la fonction d'onde. Kohn et Sham ont proposé en 1965 [13] une méthode indirect en utilisant des orbitales d'un système fictif de référence dénoté (s), constitué des n électrons non interagissant.

Le système fictif de référence est choisi de façon à avoir

$$\rho_s(\mathbf{r}) = \rho_0(\mathbf{r}) \quad (41)$$

L'hamiltonien du système s'écrit donc :

$$\hat{H}_s = \sum_{i=1}^n \left[-1/2 \nabla_i^2 + v_s(\mathbf{r}_i) \right] = \sum_{i=1}^n h_i^{\text{KS}} \quad (42)$$

$$\text{Avec} \quad h_i^{\text{KS}} = -1/2 \nabla_i^2 + v_s(\mathbf{r}_i) \quad (43)$$

Par conséquent les orbitales de Kohn et Sham θ_i^{KS} pour l'électron i, seront :

$$h_i^{\text{KS}} \theta_i^{\text{KS}} = \varepsilon_i^{\text{KS}} \theta_i^{\text{KS}} \quad (44)$$

❖ Terme d'échange-corrélation

ΔT : différence de l'énergie cinétique des électrons entre le système réel et le système fictif.

$$\Delta T = T[\rho] - T_s[\rho] \quad (45)$$

ΔV_{ee} : différence entre la vraie répulsion électron-électron et la répulsion coulombienne entre 2 charges ponctuelles.

$$\Delta V = V_{ee}[\rho] - 1/2 \iint \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 \quad (46)$$

L'énergie s'écrit alors :

$$E_v[\rho] = \int \rho(\mathbf{r})v(\mathbf{r})d\mathbf{r} + T_s[\rho] + 1/2 \iint \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 + \Delta T[\rho] + \Delta V_{ee}[\rho]$$

47)

La fonctionnelle d'énergie d'échange-corrélation est définie comme suit :

$$E_{xc}[\rho] = \Delta T[\rho] + \Delta V_{ee}[\rho] \quad (48)$$

$$E_v[\rho] = \int \rho(r)v(r)dr + T_s[\rho] + 1/2 \iint E_{xc}[\rho] \quad (49)$$

La difficulté Dans cette formule est de trouvé une bonne approximation pour le calcul de l'énergie d'échange-corrélation E_{xc} .

A partir du système de référence ; le orbitales KS permettent le calcul de la densité électronique ρ_0 suivant :

$$\rho_0 = \rho_s = \sum_{i=1}^n |\theta_i^{KS}|^2 \quad (50)$$

et de calculer la cinétique du système T_s . Donc l'énergie peut s'écrire comme :

$$E_0 = -\sum_{\alpha} Z_{\alpha} \int \frac{\rho(r_1)}{r_{1\alpha}} dr_1 - 1/2 \sum_{i=1}^n \langle \theta_i^{KS}(1) | \nabla_1^2 | \theta_i^{KS}(1) \rangle + 1/2 \iint \frac{\rho(r_1)\rho(r_2)}{r_{12}} d r_1 dr_2 + E_{xc}[\rho] \quad (51)$$

L'équation aux valeurs propres correspondante est de la forme :

$$\left[-1/2 \nabla_1^2 - \sum_{\alpha} \frac{Z_{\alpha}}{r_{1\alpha}} + \int \frac{\rho(r_2)}{r_{12}} dr_2 + v_{xc}(1) \right] \theta_i^{KS}(1) = \varepsilon_i^{KS} \theta_i^{KS}(1) \quad (52)$$

$$\text{Avec :} \quad v_{xc}(r) = \frac{\partial E_{xc}[\rho(r)]}{\partial \rho(r)} \quad (53)$$

V_{xc} : le potentiel d'échange-corrélation, définit comme la dérivée de l'énergie d'échange-corrélation E_{xc} par rapport à la densité électronique $\rho(r)$.

Le terme V_{xc} représente la principale différence entre les différentes fonctionnelle DFT. Plusieurs approximations pour déterminer ce terme ont été proposé par la suite, s'inscrivant dans le cadre du formalise de Kohn et Sham.

3.3. L'approximation locale LDA

L'approximation de la densité locale LDA stipule qu'en première approximation que la densité peut être localement constante. L'énergie d'échange-corrélation est définit des lors par la formule :

$$E_{xc}^{LDA}[\rho] = \int \rho(r)\varepsilon_{xc}(\rho)dr \quad (54)$$

\mathcal{E}_{xc} : l'énergie d'échange-corrélation par électron dans le modèle du gaz homogènes d'électrons (Gélium), exprimée en étant la somme de deux contributions.

$$\mathcal{E}_{xc}(\rho) = \mathcal{E}_x(\rho) + \mathcal{E}_c(\rho) \quad (55)$$

Avec $\mathcal{E}_x(\rho)$: l'énergies d'échange proposée par Dirac [14] comme approximation, selon la

formule :

$$\mathcal{E}_x(\rho) = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} (\rho(r))^{1/3} \quad (56)$$

Donc :

$$E_x^{LDA} = \int \rho(r) \mathcal{E}_x dr = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \int [\rho(r)]^{4/3} dr \quad (57)$$

Et le terme $\mathcal{E}_c(\rho)$ de l'énergie d'échange est donné par la formule de Vosko, Wilk et Nusair (VWN)[15].

3.4. La méthode X_α

Appelée aussi méthode Hartree-Fock-Slater proposé avant les théorèmes de Hohenberg et Kohn proposée par Slater [16]. Elle néglige le terme de correction et s'écrit suivant la formule :

$$E_{xc} \approx E_x^{X_\alpha} = -\frac{9}{8} \left(\frac{3}{\pi} \right)^{1/3} \alpha \int [\rho(r)]^{4/3} dr \quad (58)$$

α : un paramètre ajustable $\in [\frac{1}{3}, 1]$.

3.5. Approximation de la densité de spin local LSDA

Dans la méthode LDA les électrons de spin opposés ont les mêmes orbitales de Kohn et Sham θ^{KS} ce qui rend cette méthode applicable pour les systèmes à couche fermées. Pour les systèmes à couches ouvertes et les géométries de molécules près de leur état de dissociation, la méthode LSDA donne des meilleurs résultats. La LSDA distingue entre les orbitales des électrons de spins opposés.

$$E_{xc}^{LSDA} = E_{xc}[\rho^\alpha, \rho^\beta] \quad (59)$$

3.6. Approximation du gradient généralisé GGA

Les méthodes LDA et LSDA sont des approximations locales. Elles sont utilisées pour les systèmes ou la densité électronique varie très lentement avec la position, tel que les systèmes periodique (solide, matériau, polymère...). La GGA a apporté une correction en introduisant les gradients de densité de spin ρ_α et ρ_β . L'énergie échange-corrélation s'écrit :

$$E_{xc}^{GGA} [\rho^\alpha, \rho^\beta] = \int f(\rho^\alpha(r), \rho^\beta(r), \nabla \rho^\alpha(r), \nabla \rho^\beta(r)) dr \quad (60)$$

Où : f est la fonction des densités de spin et leurs gradients

Et E_{xc}^{GGA} est l'énergie d'échange-corrélation

$$E_{xc}^{GGA} = E_x^{GGA} + E_c^{GGA} \quad (61)$$

De nombreuses fonctionnelles ont été développées depuis, parmi les plus connues :

- Fonctionnelle d'échange de Becke(B88) [17] et de Perdew et Wang(PW91) [18].
- Fonctionnelle de corrélations il y a la fonctionnelle de Perdew(P86) [19], Lee Yang et Parr (LYP) [20] et de Perdew et Wang (PW91) [21].

3.7. Fonctionnelles DFT hybrides B3LYP

La fonctionnelle B3LYP (Becke-3parametres-Lee-Young-Parr) [22] est une fonctionnelle a trois paramètres d'ajustement combinant les fonctionnelles locale d'échange de Becke d'échange HF avec les fonctionnelles de corrélation locale (VWN) et corrigée du gradient de Lee-Young-Parr (LYP).

$$E_{xc}^{B3LYP} = (1 - a_0 - a_x) E_x^{LSDA} + a_0 E_x^{HF} + a_x E_x^{B88} + (1 - a_c) E_c^{VWN} + a_c E_c^{LYP} \quad (62)$$

Les 3 paramètres a_0 , a_1 , et a_2 ont été choisi pour reproduire les meilleurs lissages avec des valeurs expérimentales des énergies d'atomisation des molécules.

$$a_0 = 0.20 \quad a_1 = 0.72 \quad a_2 = 0.81$$

3.8. Processus SCF d'un calcul DFT

Etape 1 : la densité de départ est obtenue en superposant les densités électroniques calculées des atomes individuels, à une géométrie moléculaire bien choisie. A partir de cette densité une valeur initiale du potentiel d'échange-corrélation V_{xc} est obtenue. Cette valeur du potentiel d'échange-corrélation est remplacée dans l'équation de Kohn et Sham (eq :52) en utilisant l'approximation OM-CLOA.

$$\theta_i^{KS} = \sum_{r=1}^N C_{ri} \varphi_r$$

Donc la détermination des orbitales moléculaires de Kohn et Sham θ^{KS} se ramène à la détermination des coefficients.

En utilisant la méthode de variation le système séculaire similaire à celui de Roothan est obtenu.

$$\sum_{s=1}^N C_{si} \left(h_{rs}^{KS} - \varepsilon_i^{KS} S_{rs} \right) = 0 \quad r = 1, 2, \dots, N \quad (63)$$

Etape2 : les orbitales moléculaire de Kohn et Sham obtenues dans l'étape précédente sont utilisées pour calculer la nouvelle densité dans l'équation (eq. 50).

La densité améliorée va être utilisée pour calculer V_{xc} amélioré dans le système séculaire (eq. 63) pour obtenir les coefficients et les orbitale θ^{KS} améliorées. Un processus itératif est déclenché jusqu'à convergence a un champ auto-cohérent.

4. Bases d'orbitales atomiques

Le déclanchement du processus itératif auto-cohérent de la méthode de Roothan qui utilise l'approximation OM-CLOA nécessite au préalable le calcul de toutes les intégrales mono et biélectronique sur une base d'orbitales atomique OA. Ce qui permet de définir les coefficients C_{kr} de départ et les injecter dans la matrice de Fock.

En 1928 les fonctions de Hartree-Fock ont été résolues numériquement pour traiter les atomes à plusieurs électrons. Les NHFO reproduisent correctement la fonction d'onde exacte près et à grande distance des noyaux. Leur défaut essentiel est de ne pas posséder de forme analytique de sorte que le calcul des intégrales moléculaire est quasi alourdi.

En 1931 les STO en été élaboré par Slater [23] dans le but de reproduire les NHFO, par un lissage de la courbe. L'expression numérique d'une STO est donnée par la formule :

$$\Psi_{nlm} = N_n \cdot r^{n^*-1} \cdot e^{-\zeta r} \cdot Y_{lm}^{(\theta, \varphi)} \quad (64)$$

n^* : nombre quantique effectif

N_n : constante de normalisation

ζ : exposant de Slater (l'exponentiel orbitale, détermine la taille de l'orbitale)

$Y_{lm}^{(\theta, \varphi)}$: les harmoniques sphériques

Les STO présentent une forme analytique simple mais le lissage qui concerne la partie radiale, rend le calcul d'intégrales complexe et difficile à réaliser.

En 1951 Boys [24] a proposé la base GTO qui utilise les des fonctions de Gauss cartésiennes. Dans le but de reproduire les STO et par conséquent faciliter le calcul des intégrales et réduire le temps de calcul. Ces fonctions sont des puissances de (x,y,z) multiplié par $e^{(-\alpha.r^2)}$, α est une constante qui détermine l'extension radiale de la fonction.

$$\Phi_{GTO} = N. x^i. y^j. z^k. e^{(-\alpha.r^2)} \quad (65)$$

i,j,k : nombres entiers simulant les nombres atomiques n,l,m , la somme i,j,k définie le type de l'orbitale.

$$\begin{cases} i + j + k = 0 & \text{GTO type } s \\ i + j + k = 1 & \text{GTO type } p \\ i + j + k = 2 & \text{GTO type } d \end{cases}$$

Les GTO facilitent le calcul d'intégrales multacentriques. Elles possèdent la propriété mathématique dite le produit de multiplication : « Le produit de deux gaussiennes centrées en deux points A et B est équivalent à une gaussienne centrée au point C ». Par conséquent une intégrale à deux centres se réduit à une intégrale monocentrique.

La base minimale STO-3G signifie qu'il faut contracter trois gaussiennes primitives pour reproduire une STO.

$$G^{CGTO} = \sum_{\lambda=1}^k d_{\lambda} \cdot g_{\lambda}^{PGTO} \quad (66)$$

d_{λ} : coefficient de contraction de la gaussienne primitive g_{λ}

K : degré de contraction

Les couches de valences responsables de la liaison chimique doivent être considérées avec une meilleur précision que les orbitales internes. D'où le niveau split valence développé par Pople et collaborateurs [25].

La base Split Valence-Double Zeta (SV-DZ) telles que 3-21G, 4-31G, 6-31G ou le premier chiffre représente le nombre de gaussiennes utilisé pour représenter les orbitales interne. Les orbitales de la couche de valence sont représentées quand a elles avec deux fonctions de bases, qui sont composées du nombre de gaussiennes indiqué dans la seconde partie de la dénomination de la base, ainsi une base 6-31G pour la molécule H₂O comprendra 13 fonctions de bases et 22 gaussienne primitive.

La base Split Valence-Triple Zeta (SV-TZ) tel que 6-311G considère 6 gaussiennes primitives pour les orbitales de cœur (internes) et 3 fonctions de bases pour la couche de valence composées des contractions de 3,1 et 1 gaussiennes primitives respectivement.

Les orbitales atomiques dans la molécule subissent une déformation de leurs nuages électroniques et des distorsions dues à leur environnement. Pour une plus grande flexibilité des orbitales de polarisations sont rajoutées. Dénommé par l'introduction d'un astérisque ou la lettre d ; exemple 6-31G* / 6-31G(d) qui désigne d'un jeu de fonctions d'un nombre quantique supérieur a été introduit à tous les atomes sauf l'Hydrogène. La notation double astérisque ou (d,p) désigne l'ajout des orbitales de polarisation pour tous les atomes y compris l'Hydrogène. Ainsi une base 6-31G** pour la molécule H₂O comprendra 25 fonctions de bases et 42 gaussienne primitive. Cette technique permet une bonne description de grandeurs telles que l'énergie de dissociation, moment di ou multipolaire et la localisation des états de transitions.

Une bonne description des régions éloignées des noyaux et au-delà des couches de valences nécessite l'introduction des orbitales dites, fonctions diffuses. Notamment pour les systèmes ayants des doublets libres et les espèces chargées (anions) ou même les bimères à liaison d'hydrogène. Les orbitales diffuses ont les mêmes nombres quantiques des orbitales de valence elles sont caractérisées par des exposants α très petits. La notation [++] désigne l'ajout des orbitales diffuses pour tous les atomes et [+] l'ajout des orbitales diffuses pour tous les atomes lourds (excepté les hydrogènes). L'exemple de la molécule H₂O une base double zêta 6-31G++(d,p) signifie qu'elle comprend 31 fonctions de bases et 48 gaussiennes primitives.

5. Méthodes semi-empiriques

Les méthodes semi empiriques sont des méthodes quantiques dans lesquelles une partie des intégrales est calculée d'une manière exacte et autres parties soit négligées soit paramétrées grâce à des données expérimentales. Les intégrales biélectronique sont traitées de différentes manières. Et un nombre d'approximations est appliqué.

Une première approximation qui prend en considération les électrons de la couche de valence. Les électrons internes forment avec les noyaux réels le cœur moléculaire ce qu'on appelle noyaux effectifs. Les électrons de cœur sont pris en compte par une fonction de répulsion en même temps que la répulsion nucléaire. Dans le but de faciliter les calculs d'intégrales et par conséquent réduire le temps de calcul.

Il existe deux grandes familles de méthodes semi empiriques :

Méthodes semi-empiriques basées sur l'approximation NDO

C'est une méthode proposée par Parr en 1952 [26] dans le but de réduire le nombre d'intégrales à calculer dans les équations de Roothan. Le recouvrement différentiel dS correspondant à l'élément de volume.

$$S = \int_{esp} dS = \int_{esp} \phi_r \phi_s d\tau \quad (67)$$

Avec
$$dS_{rs} = \phi_r \phi_s d\tau = \delta_{rs} = \begin{cases} 1 & \text{si } r = s \\ 0 & \text{si } r \neq s \end{cases}$$

Parmi les principales méthodes semi-empiriques basées sur l'approximation NDO on trouve la CNDO/2, INDO (1966), MINDO/3 (1973) et CNDO/2-4.

La CNDO (complet NDO) proposée par Pople en 1965 [27]. Elle est valable pour les 3 premières périodes. Elle présente un défaut majeur c'est qu'elle ne distingue pas entre les états et triplets des systèmes. Elle ne peut pas être utilisée pour les couches ouvertes.

Méthodes semi-empiriques basées sur l'approximation NDDO

Cette approximation considère que le recouvrement n'est nul que si les orbitales atomiques ϕ_r et ϕ_s appartiennent à deux atomes différents. Si ϕ_r et ϕ_s appartiennent au même atome le recouvrement est pris en compte, et donc les intégrales mono et bioélectroniques ne sont pas négligées. Car la répulsion entre deux électrons appartenant au même atome est importante et ne peut pas être négligée. Les NDDO sont donc plus précises que les NDO.

Parmi les principales méthodes semi-empiriques basées sur l'approximation NDDO on trouve la MNDO, CNDO/U, AM1, PM3 et PM6.

5.1 MNDO

La méthode MNDO est une méthode semi-empirique proposée par Dewar et Thiel en 1977 [28,29]. Elle était paramétrisée pour les atomes de carbone, d'hydrogène d'oxygène et d'azote. Par la suite elle s'est étendue à un plus grand nombre d'atomes.

5.2 AM1

Pour résoudre les problèmes des liaisons hydrogènes, défaut de l'approximation MNDO, Dewar et al ont développé la méthode (Austin Model 1) [30]. Pour représenter les interactions noyau-noyau ils ont introduit des fonctions gaussiennes à la méthodologie MNDO.

5.3 PM3

C'est une méthode semi-empirique, proposée par Stewart et al. En 1989 [31]. Elle est considérée comme la troisième version de la MNDO, nommée « Parametrized Model 3 ». La paramétrisation atomique dans cette méthodologie a été effectuée par groupe éléments. Elle utilise deux fonctions gaussiennes par atome pour le calcul des répulsions cœur-cœur. Des paramètres en plus sont ajoutés pour les éléments du groupe d.

5.4 PM6

En s'appuyant sur la méthodologie MP3, Stewart et al. [32]. Ont développé en 2007 une nouvelle méthode nommée PM6. Ils ont incorporé un nouveau paramétrage d'interactions cœur-cœur en utilisant la fonction de Voityuk [33]. Ce paramètre permet de résoudre la répulsion de deux atomes non chargés par l'injection d'un terme diatomique. De plus des paramètres pour le traitement des orbitales d ont été ajouté. Ce qui permet le traitement des composés à intérêt biologique tels que les métalloprotéines [34], puisque désormais, 80 atomes ont été paramétrisés. Quoique la méthode PM6 reste déficiente dans la description des interactions non-covalentes, notamment en ce qui concerne la dispersion et la représentation sous-estimée des liaisons hydrogènes [35]. Depuis les 2009 plusieurs méthodologies ont été présentées afin de pallier ces problèmes dont PM6-HD+ [36] et PM7 [37].

Références Bibliographiques

- [1] E. Schrodinger, Ann. Phys. Leipzig. 1926, 76, 361.
- [2] M. Born, J. R. Oppenheimer, Ann. Phys. 1927, 84, 457.
- [3] V. Minkine, B. Simkine, R. Minaev, « Théorie de la structure moléculaire » Ed. Mir, Moscou, 1982.
- [4] W. Pauli, Phys. Rev. 1940. 58, 719.
- [5] V. Fock, Z. Physik., 1930, 61, 126.
- [6] J. C Slater, Phys. Rev., 1929, 34, 1293; 1931, 38, 38.
- [7] C. C. Roothaan, Rev. Mod. Phys., 1951, 23, 69.
- [8] Löwdin, P.O. Avan. Chem. Phys. 1959, 2, 207.
- [9] I. Shavitt, « *Methods of Electronic Structure Theory* », H. F. Shaefer, Ed., Plenum Press, New-York, 1977, 189.
- [10] C. Moller, M. S. Plesset, *Phys. Rev.*, 1934, 46, 618.
- [11] (a) L.H. Thomas, Proc. Cambridge Phil. Soc. 1927, 23, 452.
(b) E. Fermi, Rend. Accad. 1927, 6, 602.
(c) E. Fermi, Z. Phys. 1928, 48, 73.
(d) E. Fermi, Rend. Accad. 1928, 7, 342
- [12] P. Hohenberg, W. Kohn, Phys. Rev. A 1964, 136, 864
- [13] W. Kohn, L. Sham, Phys. Rev. A 1965, 140, 1133
- [14] Dirac, P.A.M. Proc. Cambridge Philos. Soc. 1930, 26, 376
- [15] S. H. Vosko, L. Wilk, M. Nusair, Phys. Rev. A. 1980, 58, 1200
- [16] J. C. Slater, Phys. Rev., 1951, 81, 385.
- [17] A.D. Becke, Phys. Rev. A 1988, 38, 3098
- [18] J. P. Perdew, Electronic Structure of Solids '91, P. Ziesche and H. Eschrig, Akademie Verlag, Berlin, 1991
- [19] J. P. Perdew, Phys. Rev. B 1986, 33, 8822
- [20] C. Lee, w. Yang and R.G. Parr, Phys., 1993, 98, 5648
- [21] J. P. Doucet, J. Weber, J. computer-aided molecular design, Acad. Press, London, 1996.
- [22] A. D. Becke, J. Chem. Phys., 1993, 98, 5648.
- [23] J.C. Slater, Phys. Rev. 1930, 36, 57
- [24] (a) Boys, S.F. Proc. Roy. Soc. (London), 1950, A200, 542.
(b) Boys, S.F.; Bernardi, F. Molec. Phys. 1970, 19, 553.
- [25] (a) Hehre, W.J.; Stewart, R.F.; Pople, J.A. J. Chem. Phys. 1969, 2657.
(b) Binkley, J.S.; Pople, J.A.; Hehre, W.J. J. Am. Chem. Soc. 1980, 102, 939.
- [26] R. G. Parr, J. Chem. Phys., 1952, 20, 1499.
- [27] J. A. Pople and G. A. Segal. Journal of chemical physics. 43, s136, (1965)
- [28] M. J. S. Dewar, W. Thiel, J. Am. Chem. Soc. 1977, 99, 4899–4907.
- [29] M. J. S. Dewar, W. Thiel, J. Am. Chem. Soc. 1977, 99, 4907–4917.
- [30] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, J. Am. Chem. Soc. 1985, 107, 3902–3909.
- [31] J. J. P. Stewart, J. Comput. Chem. 1989, 10, 209–220.
- [32] J. J. P. Stewart, J. Mol. Model. 2007, 13, 1173–1213.
- [33] A. A. Voityuk, N. Rösch, J. Phys. Chem. A 2000, 104, 4089–4094.
- [34] J. J. P. Stewart, J. Mol. Model. **2007**, 13, 1173–1213.
- [35] J. Rezac, K. J. Fanfrl k, D. Salahub, P. Hobza, J. Chem. Theory. Comput. **2009**, 5, 1749–1760.
- [36] M. Korth, J. Chem. Theory. Comput. **2010**, 6, 3808–3816.
- [37] J. P. Stewart, J. Mol. Model. **2003**, 19, 1–32.

CHAPITRE II

*Bases Théoriques : Méthodes d'analyse
statistique*

Introduction

Une méthode d'analyse est importante pour la mise au point d'un tel modèle QSPR/QSAR. Pour ce faire, différents types d'outils statistiques peuvent être employés :

- Régressions multi-linéaires (*Multi-linear regression* MLR) [1-5], Régressions aux moindres carrés partielles (PLS) [6], Arbres de décision [7], Réseaux de neurones [8], Algorithmes génétiques [9].

Les méthodes les plus utilisées sont les méthodes de régression linéaire. Le principe de ces méthodes est de décrire et d'évaluer une relation linéaire entre une variable (dite variable à expliquer ou variable dépendante) souvent notée Y, et une (ou plusieurs) variable(s), dite(s) variable(s) explicative(s) ou indépendante (s) notée (s) X.

1. Régression linéaire

Définitions :

- **Espérance** : désigne la moyenne des valeurs prises par X, et pondérées par leurs probabilités de réalisation.

$$\mu = E(X) = \sum_{i=1}^{i=k} x_i P(x_i = X)$$

- **Variance** : est la moyenne des écarts par rapport à la moyenne arithmétique, en terme mathématique elle peut être considérée comme une mesure servant à caractériser la dispersion d'une distribution. Elle est donnée par :

$$V(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \text{ avec } \bar{X} = \frac{\sum_{i=1}^n X_i}{n} \text{ la moyenne de la variable X}$$

- **Covariance** : elle permet d'étudier les variations simultanées de deux variables X et Y par rapport à leurs moyennes respectives.

$$Cov(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

- **Ecart-type** : est une mesure de la dispersion d'une variable aléatoire, c'est la racine carrée de la variance.

$$\sigma = \sqrt{V(X)}$$

La relation linéaire entre les deux variables, explicative (X) et à expliquer (Y) est de la forme suivante :

$$Y = a + bX$$

Puisque la relation qu'on étudie est une simplification de la réalité, que l'on a forcément oublié des facteurs explicatifs de Y, on introduit un terme aléatoire (perturbation, appelé encore résidu) noté ε à cette relation déterministe. Ainsi le modèle que l'on étudie est le suivant :

$$Y = \beta_0 + \beta_1 X_i + \varepsilon \quad (1)$$

Avec β_0 (appelé *Intercepte*) et β_1 sont les paramètres inconnus, appelés coefficients de régression, que l'on cherche à déterminer. ε , le terme d'erreur, qui est supposé être une variable aléatoire et vérifie :

- La variance des résidus est constante, $V(\varepsilon_i) = \text{constante}$.
- La covariance des résidus est nulle, $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ (si $i \neq j$), ce qui implique la non corrélation des résidus.

L'objectif est de déterminer des valeurs pour les paramètres β_0 et β_1 à partir des n observations sur les variables X et Y.

Graphiquement, il s'agit de trouver les paramètres de la droite de régression, qui passe au milieu du nuage de points dessiné dans le plan (X, Y).

On remarque que les paramètres β_0 et β_1 sont les mêmes pour toutes les observations, autrement dit que l'influence de X sur Y est la même pour toutes les observations.

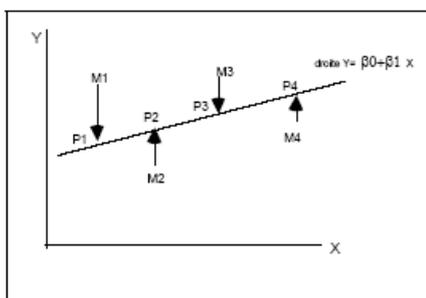


Figure.1. Représentation graphique de la régression

Pour trouver la droite qui passe « au plus près » de tous les points il faut se donner un

Critère d'ajustement. On projette les points de M1 à M4 parallèlement à l'axe Y. Sur la droite on obtient les points P1 à P4, comme le montre la Figure 1. Le critère retenu pour déterminer la droite passant au plus près de tous les points sera tel que : « La somme des carrés des écarts (SCE) des points observés M_i à la droite solution soit minimum »

La droite solution sera appelée droite de **régression de Y sur X**. Le critère est le «critère des Moindres Carrés Ordinaires » (Ordinary Least Squares).

1.1- Méthode d'estimation des paramètres β_0 , β_1 ...

Critère des Moindres Carrés Ordinaires (MCO, Ordinary Least Squares)

- **Cas d'une régression simple** (une seule variable explicative)

La Somme des Carrés des Ecart (SCE) est donnée par :

$$S = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \quad (2)$$

La valeur de cette fonction S est minimum lorsque les dérivées de S par rapport à β_0 et β_1 s'annulent. La solution est obtenue en résolvant le système :

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \quad (3)$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^n X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \quad (4)$$

$\hat{\beta}_0$ et $\hat{\beta}_1$ sont des estimateurs de β_0 et β_1

Ces dérivées s'annulent pour deux valeurs $\hat{\beta}_0$ et $\hat{\beta}_1$ solutions des deux équations à deux inconnues :

L'équation (3) nous donne

$$\sum Y_i - n\hat{\beta}_0 - n\hat{\beta}_1 \sum X_i = 0 \quad (5)$$

En utilisant la formule de la moyenne $\bar{X} = \sum_i \frac{X_i}{n}$ et en divisant par n nous obtenons

l'équation ci-dessous

$$\hat{\beta}_0 = \bar{Y} - \beta_1 \bar{X} \quad (6)$$

L'équation (4) nous donne

$$\sum Y_i X_i - \hat{\beta}_0 \sum X_i - \hat{\beta}_1 \sum X_i^2 = 0 \quad (7)$$

On remplace $\hat{\beta}_0$ obtenu dans l'équation (6) dans l'équation (7) nous aboutissons à l'équation ci-dessous

$$\sum Y_i X_i - (\bar{Y} - \hat{\beta}_1 \bar{X}) \sum X_i - \hat{\beta}_1 \sum X_i^2 = 0 \quad (8)$$

Cette équation nous permet d'obtenir la formule du deuxième coefficient de l'équation de régression

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \quad (9)$$

En remplaçant l'expression de $\hat{\beta}_1$ dans l'équation (6) on obtient

$$\hat{\beta}_0 = \bar{Y} - \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \bar{X}$$

$$\hat{Y} = \beta_0 + \beta_1 X \quad (10)$$

\hat{Y} est l'estimation de Y obtenue à partir de l'équation de régression.

L'écart entre cette prédiction \hat{Y} et Y est appelé résidu

$$Y_i - \hat{Y}_i = \varepsilon_i \quad (11)$$

la variation résiduelle appelée σ^2 est donnée par :

$$S^2 = \frac{1}{n-2} \sum \varepsilon_i^2 \quad (12)$$

- Cas d'une régression multiple (plusieurs variables explicatives)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (13)$$

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

β_p .

Il s'agit donc de considérer un système d'équations :

$$\begin{aligned}
 Y_1 &= \beta_0 + \beta_1 X_{1,1} + \beta_2 X_{2,1} + \dots + \beta_p X_{n,1} + \varepsilon_1 \\
 Y_2 &= \beta_0 + \beta_1 X_{1,2} + \beta_2 X_{2,2} + \dots + \beta_p X_{n,2} + \varepsilon_2 \\
 &\dots \\
 Y_n &= \beta_0 + \beta_1 X_{1,p} + \beta_2 X_{2,p} + \dots + \beta_p X_{n,p} + \varepsilon_n
 \end{aligned}
 \tag{14}$$

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

Ce système d'équation peut être donné sous la forme matricielle suivante :

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

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

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

Les coefficients β_p sont obtenu comme suit

$$\beta = (X'X)^{-1}X'Y$$

1.2- Décomposition de la variance et qualité de la régression

A partir de l'équation de la droite de régression (modèle retenu), on peut pour tout point i d'abscisse X_i calculer son estimation (ordonnée) \hat{Y}_i . A partir des équations (6) et (10) on obtient :

$$\hat{Y}_i = \bar{Y} + \hat{\beta}_1 (X_i - \bar{X})
 \tag{15}$$

Ou encore

$$\hat{Y}_i - \bar{Y} = \hat{\beta}_1 (X_i - \bar{X})
 \tag{16}$$

$$Y_i - \hat{Y}_i = (Y_i - \bar{Y}) - (\hat{Y}_i - \bar{Y})
 \tag{17}$$

On élève les deux membres au carré et on somme sur les observations i on obtient

$$\sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \bar{Y})^2 + (\hat{Y}_i - \bar{Y})^2 - 2\sum (Y_i - \bar{Y})(\hat{Y}_i - \bar{Y}) \quad (18)$$

En utilisant l'équation (16) on aboutit à

$$\sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \bar{Y})^2 + (\hat{Y}_i - \bar{Y})^2 - 2\sum (Y_i - \bar{Y})\hat{\beta}_1(\hat{X}_i - \bar{X}) \quad (19)$$

En utilisant la transformation de l'équation (9) $\hat{\beta}_1 \sum (X_i - \bar{X})^2 = \sum (X_i - \bar{X})(Y_i - \bar{Y})$ on obtient

$$\sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \bar{Y})^2 + \sum (\hat{Y}_i - \bar{Y})^2 - 2\hat{\beta}_1^2 \sum (X_i - \bar{X})^2 \quad (20)$$

En utilisant l'équation (16) on aboutit à la relation

$$\sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \bar{Y})^2 + \sum (\hat{Y}_i - \bar{Y})^2 - 2\sum (\hat{Y}_i - \bar{Y})^2 \quad (21)$$

Finalement on obtient la relation fondamentale de l'analyse de la variance

$$\sum_i (Y_i - \bar{Y})^2 = \sum_i (\hat{Y}_i - \bar{Y})^2 + \sum_i (Y_i - \hat{Y}_i)^2 \quad (20)$$

Ces diverses sommes des carrés sont définies comme suit :

* La quantité SST (Sum of Squares Total) représente la somme des carrés des écarts à la moyenne (SCE total).

$$SST = \sum (Y_i - \bar{Y})^2 \quad (21)$$

Cette variation totale à expliquer, peut se décomposer en :

-Somme des carrés expliquée par le model, ou plus précisément par la variable (X)

$$SSR = \sum (\hat{Y}_i - \bar{Y})^2 \quad (22)$$

-Somme des carrés des résidus (partie que le modèle n'explique pas)

$$SSE = \sum \varepsilon_i^2 = \sum (Y_i - \hat{Y}_i)^2 \quad (23)$$

On a alors l'équation d'analyse de la variance suivante :

$$SST = SSR + SSE \quad (24)$$

Cette dernière formule montre que les variations de Y de sa moyenne, c'est-à-dire SCE totale (SS Total) peuvent être expliquées par le modèle grâce à SCE (Modèle) (SS Model ou SS régression) et ce que ne peut être expliqué par le modèle est dans SCE (Erreurs) (SS Error).

1.3. Représentation de l'analyse de la variance

Toutes ces sommes sont présentées dans un tableau (table d'analyse de variance ou table d'ANOVA)

Tableau .1. Analyse de la variance

Source de Variation	Degrés de liberté DF	Somme des carrés SS	Moyenne des carrées MS
Model	p	$\sum (\hat{Y}_i - \bar{Y})^2$	$\frac{SSR}{p}$
Error	n-1-p	$\sum (Y_i - \hat{Y}_i)^2$	$\frac{SSE}{n - p - 1}$
Total	n-1	$\sum_i^n (Y_i - \bar{Y})^2$	$\frac{SSR}{p} + \frac{SSE}{n - p - 1}$

DF : degré de liberté

SS : Sum of squares : somme carrée des écarts

MS : Mean square, est le rapport SS/DF

MSE: Mean Square Error

1.4. Qualité de la régression linéaire

La qualité d'une relation linéaire entre Y et Xi est vérifiée par plusieurs coefficients :

1.4.1 Coefficient de détermination R²

Pour évaluer la précision avec laquelle la dépendance trouvée décrit la variance de la variable dépendante (c'est à dire la qualité de l'ajustement statistique), le carré du coefficient de détermination (R²) est utilisé.

On définit alors le coefficient de détermination, qui mesure la part de la variance expliquée par le modèle dans la variance totale.

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{25}$$

$$R^2 = \frac{\sum (Y_i - \hat{Y})^2}{\sum (Y_i - \bar{Y})^2} = \frac{b_1^2 \sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2} = \frac{\text{cov}(x_i, y_i)^2}{\text{var}(x_i) \text{var}(y_i)} = r_{xy}^2 \tag{26}$$

1.4.2 Déviation standard (SD)

La fiabilité de la prédiction de la variable dépendante peut être évaluée également par la valeur de l'erreur type d'estimation "s" ou déviation standard (SD).

$$s^2 = MSE = \frac{\sum_i^n (Y_i - \hat{Y}_i)^2}{n-p-1} = \frac{SSE}{n-p-1} \quad (27)$$

$$\text{Root MSE} = \text{SD} = s \quad (29)$$

1.5. Confiance à accorder aux résultats

Test de la signification globale de la régression :

1.5.1. Test de Fisher-Snedecor

Ce test a surtout un intérêt dans le cadre de la régression multiple, c'est-à-dire avec p régresseurs X_1, X_2, \dots, X_p :

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

On teste l'hypothèse nulle : $H_0: \beta_1 = \dots = \beta_p = 0$ contre H_1 : il existe au moins un β_j parmi β_1, \dots, β_p non égal à 0.

$$F = \frac{MS_{\text{model}}}{MS_{\text{error}}} \quad (30)$$

$$\text{Avec} \quad MS_{\text{model}} = \frac{SS_{\text{Model}}}{p} \quad (31)$$

$$\text{Et} \quad MS_{\text{error}} = \frac{SS_{\text{error}}}{n-p-1} \quad (32)$$

Une équation de régression est considérée comme statistiquement significative si la valeur observée de F est supérieure à une valeur tabulée pour le niveau de signification choisi (typiquement 95%) et les degrés de liberté correspondants de F. Les degrés de liberté de F sont égaux à p et n-p-1.

Si $p\text{-value} \leq \alpha$ alors on rejette l'hypothèse nulle

La statistique F est liée au coefficient de détermination par la relation suivante

$$F = \frac{R^2}{1-R^2} \frac{n-p-1}{p} \quad (33)$$

1.5.2. Test de Student

- Statistiques liées aux paramètres β_1, β_0

On calcule alors le T-test

$$t\text{-observé} = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)} \quad (34)$$

$s(\hat{\beta}_1)$ est l'erreur type du paramètre β_1

$$t\text{-observé} = \frac{\hat{\beta}_0}{s(\hat{\beta}_0)} \quad (35)$$

$s(\hat{\beta}_0)$ est l'erreur type du paramètre β_0

Remarque :

En pratique si on choisit le risque $\alpha = 5\%$ et si n est grand ($n > 30$) pour approcher la loi de Student par la loi normale la valeur critique de t-test oscille le plus souvent autour de 2 (1,96) [10, 11].

1.6. Sélection de variables et choix du modèle [12-18]

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

1.6.1. Coefficient de détermination ajusté $R^2_{\text{ajusté}}$

Ce coefficient est utilisé en régression multiple par ce qu'il tient compte du nombre de paramètres (descripteurs) du modèle.

$$R^2_{\text{adj}} = 1 - \frac{(n - \text{intercept})(1 - R^2)}{n - p} \quad (36)$$

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

1.6.2. Critère de validation croisée : PRESS

La somme des erreurs quadratiques de prédiction « *prediction sum of squares* » (PRESS) est définie par

$$PRESS = \sum_{i=1}^n \varepsilon_{(i)}^2 \quad (37)$$

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

1.6.3. Cross validated R_{cv}^2

La procédure statistique cross-validation peut être utilisée pour évaluer le pouvoir prédictif des modèles QSAR.

Le coefficient qui décrit la validation est donné par l'équation ci-dessous

$$R_{cv}^2 = 1 - \frac{\sum_i (Y_i^{pred} - Y_i^{obs})^2}{\sum (Y_i^{obs} - Y^{mean})^2} \quad (38)$$

Ce coefficient peut être calculé à partir de PRESS comme suit

$$R_{cv}^2 = 1 - \frac{PRESS}{SS_{total}}$$

Références Bibliographiques

- [1] S. Chatterjee and A. S. Hadi, fourth Edition, a John Wiley & Sons, Inc., Hoboken, New Jersey, **2006**.
- [2] S. Weisberg, thirth Edition, John Wiley & Sons, IncHoboken, New Jersey **2005**.
- [3] A. C. Rencher, G. B. Schaalje, Second Edition, John Wiley & Sons, Inc., Hoboken, New Jersey, **2008**.
- [4] P. A. Cornillon, E.M.atznerLøber, Springer Verlag France, Paris **2007**.
- [5] M. Lejeune, "Statistiques : Springer Verlag, Paris **2004**.
- [6] P. Geladi, B.R. Kowalski, *Anal. Chim. Acta*, **1986**, *185*, 1.
- [7] A.F. Duprat, T. Huynh, G. Dreyfus, *J. Chem. Inf. Comput. Sci.*, **1998**, *38*, 586-594
- [8] I.V. Tetko, A.E.P. Villa, D.J. Livingstone, *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, 794-803.
- [9] A.F. Siegel "Practical Business Statistics", IRWIN, **1997** (3rd edition)
- [10] A.F. Siegel "Practical Business Statistics", IRWIN, **1997** (3rd edition)
- [11] R. D. Cook, S .Weisberg, "An introduction to Regression Graphics", Wiley Series in Probability and Statistics, **1994**.
- [12] J. JACQUES, "Modélisation Statistique" [http ://labomath.univ-lille1.fr/~jacques](http://labomath.univ-lille1.fr/~jacques)
- [13] P. Besse. "Pratique de la modélisation statistique , Publications du laboratoire de statistique et Probabilités", **2003**. Disponible sut [http:// www.math.univ-toulouse.fr](http://www.math.univ-toulouse.fr)
- [14] P. Besse. "Apprentissage Statistique Data mining , Publications du laboratoire de statistique et Probabilités", **2009**.
- [15] G.J.Mclachlan, "Discriminant analysis and Statistical Pattern Recognition. Wiley, New-York, **1992**.
- [16] J.P.Nakache et J.Confais, "Statistique explicative appliquée", Edition Technip **2003**.
- [17] G.Saporta, "Probabilité, Analyse de données et statistique", 2^{ème} Edition, Edition Technip, **2006**.
- [18] D.Laffly , "Régression multiple : principes et exemples d'application", **2006**.

CHAPITRE III

Résultats et discussion

Introduction

Les méthodes QSAR représentent un domaine de recherche très important de la chimie-physique. L'objectif est d'établir une relation ou une équation mathématique reliant des descripteurs structurels d'une série de composés étudiés aux activités biologiques.

Ces méthodes ont permis de nouvelles avancées vers la compréhension des données biologiques en utilisant des descripteurs simples et ciblés.

Vu les dégâts impliqués par la propagation des radicaux libres dans l'organisme causant ainsi de nombreuses maladies plus ou moins graves, l'activité anti-oxydante se présente parmi les activités biologiques qui suscitent le plus d'intérêt ces dernières années. Reis et al. [1] ont étudié l'activité anti-oxydante de 41 composés phénoliques, ils ont trouvé que l'activité dépend de deux descripteurs électroniques, le potentiel d'ionisation et la charge sur l'atome d'oxygène de la liaison O-H du phénol. Un bon modèle a été obtenu avec ces deux descripteurs. Cependant la stabilité, la validation externe (critères de Tropsha) et le domaine d'applicabilité de ce modèle n'ont pas été discutés et analysés.

L'objectif de ce travail est d'étudier et de prédire l'activité anti-oxydante de cette même série constituée de 41 composés phénoliques simples, en utilisant des descripteurs ciblés qui peuvent expliquer le mécanisme de cette activité étudiée et en respectant toute la méthodologie d'une étude QSAR et tous les critères d'OCDE [2].

1. Méthodologie

1.1. Base de données

Le potentiel redox déterminé expérimentalement à pH 7 (E_7) a été couramment utilisé comme mesure directe de l'activité anti-oxydante par plusieurs auteurs [1, 3-6]. Les 41 composés utilisés dans cette étude, ainsi que les valeurs du potentiel redox expérimental (E_7) sont présentés dans le **tableau 1**. La structure modèle de ces molécules est représentée sur le **schéma 1**.

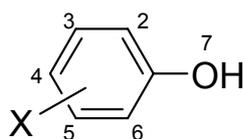


Schéma 1 : structure des composés phénoliques

Tableau1 : Base de données [1]

N°	Nom du composé	Substituant X	E ₇ (v)
1	4-nitorphenl	4-NO ₂	1,23
2	3,5-dichlorophenol	3,5-Cl ₂	1,15
3	4- (trifluoromethyl)phenol	4-CF ₃	1,13
4	3-nitrophenol*	3-NO ₂	1,13
5	4-benzoylphenol	4-PhCO	1,12
6	3-cyanophenol	3-CN	1,11
7	4-hydroxybenzoic acid	4-COOH	1,04
8	3-acetylphenol	3-CH ₃ CO	0,98
9	phenol	4-H	0,97
10	4-bromophenol*	4-Br	0,96
11	4-chlorophenol	4-Cl	0,94
12	4-fluorophenol	4-F	0,93
13	Tyrosine	Tyrosine	0,89
14	4-acetyl-3-hydroxyphenol*	3-OH,4-COCH ₃	0,89
15	4-methylphenol	4-CH ₃	0,87
16	3,5-dimethoxyphenol	3,5-(CH ₃ O) ₂	0,85
17	3-methylphenol	3-CH ₃	0,85
18	3-hydroxy-5-methoxyphenol*	3-OH,5-OCH ₃	0,84
19	3,5-dimethylphenol	3,5-(CH ₃) ₂	0,84
20	4-phenylphenol	4-Ph	0,84
21	2-methylphenol	2-CH ₃	0,82
22	3-hydroxyphenol*	3-OH	0,81
23	2-methoxyphenol	2-OCH ₃	0,77
24	4-methoxyphenol*	4-OCH ₃	0,73
25	3,4-dimethoxyphenol	3,4-(CH ₃ O) ₂	0,67
26	3,4,5-trimethoxyphenol	3,4,5-(CH ₃ O) ₃	0,66
27	Sesamol*	Sesamol	0,62
28	3,4-dihydroxybenzoic acid	2-OH, 4-COOH	0,60
29	2,6-dimethoxyphenol	2,6-(CH ₃ O) ₂	0,58

30	2,3-hydroxyphenol	2,3-(OH) ₂	0,58
31	2,3,5-trihydroxybenzoatede methyl	2,3-(OH) ₂ , 5-COOCH ₃	0,56
32	3,4-dihydrocynnamic acid	3,4-Dihydrocynnamic acid	0,54
33	2-hydroxyphenol	2-OH	0,53
34	2-hydroxy-4-methylphenol	2-OH, 4-CH ₃	0,52
35	hydroxyphenol	4-OH	0,46
36	4-aminophenol	4-NH ₂	0,41
37	4-cyanophenol	4-CN	1,17
38	4-acetylphenol*	4-COCH ₃	1,06
39	4-t-buthylphenol	4-t-Bu	0,80
40	2,6-dimethylphenol	2,6-(CH ₃) ₂	0,77
41	2-methoxy-4-methylphenol	2-OCH ₃ , 4-CH ₃	0,68

*(Série de Test)

1.2. Calcul des descripteurs

Les géométries d'équilibre des composés neutres ont été effectuées en utilisant le programme Gaussian G09 [7] aux différents niveaux de calcul : AM1, PM6, RHF/6-31G(d,p), DFT/B3LYP/6-31++G(d,p) et B3LYP/6-311++G(d,p). Tandis que les géométries des structures radicalaires ont été optimisées au moyen des méthodes UHF et UB3LYP/6-31++G(d,p) et UB3LYP/6-311++G(d,p).

D'après les recherches bibliographiques [5, 8-10], plusieurs descripteurs issus des propriétés électroniques des composés phénoliques peuvent expliquer l'activité antioxydante.

Dans ce travail, les propriétés électroniques les plus pertinentes en relation avec l'activité étudiée et son mécanisme sont :

L'affinité électronique (*EA*), le potentiel d'ionisation *IP*, la *BDE-OH*, la charge *Mulliken* (*MK*) et la *densité du spin* de Mulliken (*MSD*).

L'affinité électronique (*EA*) a été calculée comme la différence d'énergie entre la forme neutre et la forme anion correspondante.

$$EA = E_{neutre} - E_{anion} \quad (1)$$

Le potentiel d'ionisation IP a été calculé suivant trois méthodes. IP_1 comme la différence entre l'enthalpie de formation du cation radical et l'enthalpie de formation de la molécule neutre.

Formules d' IP_1

$$IP_1 = \Delta H_f(\text{ArOH}^{\cdot+}) - \Delta H_f(\text{ArOH}) \quad (2)$$

Avec ArOH composé phénolique

Et Ar OH $^{\cdot+}$ Cation radical correspondant

IP_2 comme la différence d'énergie entre le cation radical et la molécule neutre.

$$IP_2 = E(\text{ArOH}^{\cdot+}) - E(\text{ArOH}) \quad (3)$$

La BDE-OH a été calculée en utilisant les formules ci-dessous :

$$\text{BDE-OH} = [\Delta H_f(\text{PhOH}) + \Delta H_f(\text{ArO}^{\cdot})] - [\Delta H_f(\text{PhO}^{\cdot}) + \Delta H_f(\text{ArOH})] + \text{BDE-OH}_{\text{exp}}(\text{PhOH}) \quad (4)$$

Dans cette étude nous avons introduit un descripteur de partage, log P, qui reflète le phénomène de transfert entre deux phases lipidique et aqueuse.

Ce descripteur est obtenu en utilisant le programme ACD [11] et Chemspider [12].

1.3. Analyse statistique

Des modèles QSAR ont été générés en utilisant des régressions simples SLR puis multiples MLR sur le programme Minitab version 17 [13]. Le E_7 est pris comme la variable dépendante, tandis que les variables indépendantes sont les descripteurs calculés :

E_{HOMO} , BDE-OH, IP_2 , logP EA, MSD.

2. Résultats et discussion

La base de données constituée de 41 composés phénoliques est répartie en deux sous séries. Une série d'apprentissage constituée de 31 composés et une série de test de 10 composés.

Plusieurs modèles sont élaborés en utilisant les descripteurs calculés aux différents niveaux. Seuls les descripteurs calculés avec la méthode **B3LYP6/311++G(d,p)** sont retenus parce qu'ils ont donnés les meilleures corrélations avec l'activité.

L'analyse des résidus standardisés a montré l'existence de quatre valeurs incohérentes (*outliers*) (structures en **gras** sur le **tableau 1**) avec des valeurs résiduelles standardisées supérieure à **2,5 unités** (seuil exigé par Minitab).

Nous présentons uniquement les modèles après élimination de ces structures

2.1. Modèles QSAR avec un seul descripteur

Tableau 2 : Modèles QSAR simples

<i>N°</i>	<i>Modèle</i>	<i>R²</i>	<i>R²_{adj}</i>	<i>SD</i>
1	$E_7 = -2,37 + 0,0376 \text{ BDE-OH}$	26,21%	23,48%	0,190834
2	$E_7 = 0,8393 + 0,1844 \text{ AE}$	32,16%	29,65%	0,182983
3	$E_7 = 0,5603 + 0,1536 \log P$	34,23%	31,79%	0,180170
4	$E_7 = -1,431 + 5,76 \text{ MSD}$	45,98%	43,98%	0,163288
5	$E_7 = -1,441 + 0,01198 \text{ IP}_2$	69,39%	68,26%	0,122910
6	$E_7 = -1,383 - 9,35 E_{\text{HOMO}}$	69,91%	68,80%	0,121856

Les 2 meilleurs modèles obtenus avec 1 seul descripteur sont ceux avec :

E_{HOMO} : $R^2 = 69,91\%$ $SD = 0,121856$

IP_2 : $R^2 = 69,39\%$ $SD = 0,122910$

Dans le but d'obtenir un modèle plus explicatif des modèles avec 2 et 3 descripteurs sont élaborés.

2.2. Modèles QSAR avec 2 descripteurs

Avant d'élaborer les modèles avec deux et trois descripteurs, la matrice de corrélation doit être établie afin de vérifier les colinéarités des descripteurs.

Tableau 3 : Matrice de corrélation

	E_7	E_{HOMO}	BDE	IP_2	$LogP$	SDM
E_{HOMO}	-0,836					
BDE-OH	0,512	-0,601				
IP_2	0,833	-0,907	0,726			
LogP	0,585	-0,418	0,122	0,329		
SDM	0,678	-0,727	0,582	0,849	0,367	
AE	0,567	-0,599	0,543	0,371	0,371	0,361

L'analyse de la matrice de corrélation montre que le IP_2 et E_{HOMO} sont fortement corrélés, donc les modèles MLR ne doivent pas contenir les 2 descripteurs en même temps. Le **tableau 4** regroupe tous les modèles avec deux descripteurs avec $R^2 > 50\%$

Tableau 4 : modèles QSAR avec deux descripteurs

N°	Modèle	R^2	R^2_{adj}	SD	F
1	$E_7 = -2,202 + 0,03282 BDE + 0,1393 LogP$	53,94%	50,40%	0,153639	15,23
2	$E_7 = -1,451 - 9,25 E_{HOMO} + 0,00108 BDE$	69,93%	67,61%	0,124149	30,23
3	$E_7 = -1,592 - 8,14 E_{HOMO} + 1,26 MSD$	70,96%	68,72%	0,122010	31,76
4	$E_7 = -1,217 - 8,66 E_{HOMO} + 0,0334 AE$	70,59%	68,33%	0,122771	31,21
5	$E_7 = -0,610 - 0,0143 BDE + 0,01402 IP_2$	71,19%	68,98%	0,121508	32,13
6	$E_7 = -1,195 - 8,02 E_{HOMO} + 0,0750 LogP$	76,65%	74,86%	0,109391	42,68
7	$E_7 = -1,284 + 0,01033 IP_2 + 0,0916 LogP$	80,26%	78,74%	0,100593	52,85

D'après les résultats présentés dans le **tableau 4**, il s'avère que le meilleur modèle obtenu avec deux descripteurs est celui avec IP_2 et $LogP$

2.3. Modèles avec trois descripteurs

Tableau 5 : Modèles QSAR avec trois descripteurs

<i>N°</i>	<i>Modèle</i>	<i>R²</i>	<i>R²_{adj}</i>	<i>SD</i>	<i>F</i>
1	$E_7 = -0,839 + 0,0814 \text{ LogP} + 3,92 \text{ MSD} + 0,0928 \text{ AE}$	65,61%	61,48%	0,135	15,90
2	$E_7 = -1,506 - 8,23 E_{\text{HOMO}} - 0,0015 \text{ BDE} + 1,32 \text{ MSD}$	70,98%	67,50%	0,124	20,38
3	$E_7 = -0,438 - 0,0150 \text{ BDE} + 0,01571 \text{ IP}_2 - 1,11 \text{ MSD}$	71,66%	68,26%	0,122	21,07
4	$E_7 = -0,460 - 0,0130 \text{ BDE} + 0,01267 \text{ IP}_2 + 0,0485 \text{ AE}$	72,75%	69,48%	0,120	22,24
5	$E_7 = -1,536 - 7,46 E_{\text{HOMO}} + 0,00549 \text{ BDE} + 0,0781 \text{ LogP}$	77,00%	74,24%	0,110	27,90
6	$E_7 = -1,183 + 0,00984 \text{ IP}_2 + 0,0875 \text{ logP} + 0,0234 \text{ AE}$	80,60%	78,27%	0,101	34,63
7	$E_7 = -0,811 - 0,00828 \text{ BDE-OH} + 0,01158 \text{ IP}_2 + 0,0877 \text{ logP}$	80,84%	78,54%	0,101	35,16
8	$E_7 = -1,050 + 0,01294 \text{ IP}_2 + 0,0973 \text{ logP} - 1,88 \text{ MSD}$	81,58%	79,37%	0,099	36,92

Le **tableau 5** montre que le meilleur modèle avec trois descripteurs est celui obtenu avec la combinaison d'IP₂, logP avec MSD.

$$E_7 = -1,050 + 0,01294 \text{ IP}_2 + 0,0973 \text{ logP} - 1,88 \text{ SDM} \quad (5)$$

$$R^2=81,58\% \quad SD= 0,099 \quad F= 36,92$$

Il s'avère qu'il ' y a une petite amélioration de R² par rapport au modèle obtenu avec deux descripteurs (modèle # 5 du tableau 4).

On voulait connaître s'il s'agit d'une contribution réelle du MSD/ (BDE, AE), ou bien cette amélioration est due seulement à l'**ajout d'un troisième descripteur** (effet du nombre de descripteurs).

Pour cette raison on était obligé d'analyser les autres paramètres statistiques à savoir le T-test et P-value de ce modèle (**Eq.5**).

Tableau 6 : Coefficient, S.E coefficient, T-test-P-value des descripteurs du modèle Eq.5

	<i>Coef</i>	<i>SE Coef</i>	<i>T-test</i>	<i>P-Value</i>
Constante	-1,050	0,295	-3,56	0,002
IP ₂	0,01294	0,00234	5,53	0,000
LogP	0,0973	0,0242	4,02	0,000
SMD	-1,88	1,40	-1,34	0,192

On remarque que le T-test pour MSD est égal à -1,34 la valeur absolue est inférieure à la valeur seuil (1,96), la valeur de P-value justifie aussi la non contribution de ce descripteur : P-Value=0,192 > α , ($\alpha=5\%$). Le même résultat est obtenu avec la BDE et l'AE.

Pour cette raison le meilleur modèle retenu est celui avec 2 descripteurs nommés IP₂ et logP.

Pour confirmer ce choix nous avons utilisé la méthode *Stepwise*, tous les descripteurs sont introduits avec l'activité, le meilleur modèle choisi par cette méthode est celui présenté par l'Eq.6

2.4. Etude du meilleur modèle

$$E_7 = -1,284 + 0,01033 IP_2 + 0,0916 \text{ LogP} \quad (6)$$

$$R^2 = 80,26\% \quad R_{adj}^2 = 78,74\% \quad SD = 0,100593 \quad F = 52,85$$

La forme standardisée donne l'équation ci-dessous :

$$E_{7 \text{ stand}} = 0,7183 IP_{2 \text{ stand}} + 0,3490 \log P_{\text{ stand}} \quad (7)$$

Pour voir l'ordre de contribution de chaque descripteur, la forme standardisée est recommandée. Cette dernière a donné le même ordre obtenu dans le **tableau 7**.

Tableau 7 : Coefficients, SE. Coefficient des variables, T-test et p-value des descripteurs du meilleur modèle Eq.6

	<i>Coef</i>	<i>SE Coef</i>	<i>T-test</i>	<i>P-Value</i>
Constant	-1,284	0,241	-5,33	0,000
IP ₂	0,01033	0,00133	7,79	0,000
logP	0,0916	0,0242	3,78	0,001

On remarque que le T-test augmente en passant de logP à IP₂. Ce qui signifie que la contribution d'IP dans l'explication de l'activité est meilleure à celle de logP.

D'après le **tableau 7**, T-test est supérieur à la valeur seuil (1,96) et P-value est inférieur à 0,05 (risque $\alpha=5\%$) pour les deux descripteurs de notre meilleur modèle. Ce qui explique **l'existence** de ce modèle. Ce dernier peut expliquer 80% de la variance de l'activité anti-oxydante étudiée.

Tableau 8 : Valeurs des descripteurs, d'E₇ calculé/expérimental et des résidus

<i>IP(ev)</i>	<i>logP</i>	<i>E₇ (exp)</i>	<i>E₇ (calculé)</i>	<i>Résiduel</i>
Série d'apprentissage				
9.43088304	1.57	1.23	1.11184535	0.1179711
9.06730728	3.33	1.15	1.18679853	-0.0369953
9.08639014	1.93	1.13	1.11227047	0.01754689
8.22403408	2.87	1.12	0.94324586	0.17661509
9.20567641	1.70	1.11	1.07216173	0.03766469
8.99409961	1.42	1.04	0.99510232	0.04474199
8.71447429	1.39	0.98	0.92448986	0.05537148
8.60545734	1.48	0.97	0.90978502	0.06008013
8.43025309	2.43	0.94	0.95133322	-0.01147547
8.52505245	1.77	0.93	0.91559634	0.01426822
8.13772174	1.94	0.87	0.83786913	0.03201474
8.23928295	1.43	0.85	0.81627285	0.03361483
8.42816059	1.94	0.85	0.90877827	-0.05891158
8.23633059	2.40	0.84	0.90396324	-0.06409409
7.52816124	3.20	0.84	0.81002262	0.02987154
8.2251987	1.94	0.82	0.85956626	-0.03968764
7.67060926	3.17	0.80	0.89891245	-0.09903992
7.23267753	1.19	0.77	0.65814376	0.11178155
7.17838402	2.40	0.77	0.82834415	-0.05845668

8.25640393	1.65	0.68	0.63132435	0.04860875
7.13497751	1.25	0.67	0.55856763	0.11138198
7.94453937	1.07	0.66	0.52839965	0.13155677
7.59404651	1.16	0.60	0.7944468	-0.1945546
7.96717868	0.77	0.58	0.49134029	0.08862426
8.24784072	0.29	0.58	0.642534	-0.06260743
8.77499089	1.42	0.54	0.72781964	-0.18791056
7.92851771	0.88	0.53	0.70224504	-0.17233128
7.92198442	1.34	0.52	0.66444417	-0.14451997
7.38818124	-0.29	0.41	0.39111535	0.01887051
Série de Test				
9.12471981	2.82	1.13	1.14799662	-0.04498943
8.7988388	1.42	1.06	0.94212547	0.08557663
8.38296032	2.49	0.96	0.94106717	0.03826523
8.51746402	1.74	0.89	0.90440855	-0.02003254
8.44918088	0.91	0.84	0.81211419	0.010179
8.59572542	0.76	0.81	0.83328393	-0.05454234
8.02107651	1.31	0.73	0.74677146	0.00432942
7.69235469	1.52	0.62	0.68769956	-0.01878598

Le tableau 8 montre que la molécule 36 possédant le plus petit E_7 (calculé) (la plus grande activité antioxydante) est caractérisée par une forte hydrophilie (-0.29) et le plus petit potentiel d'ionisation.

2.4.1. Validation du meilleur modèle QSAR

1) Validation interne

Un modèle QSAR peut être obtenu par le hasard, pour cette raison, il faut toujours s'assurer de sa stabilité. Deux méthodes sont utilisées pour l'étude de la stabilité de notre meilleur modèle Eq.6.

- **LOO Leave one out :**

La procédure LOO (mettre une molécule à part) a donné une valeur moyenne $R_{cv}^2 = 76,11$, cette valeur est supérieure à 50%, en plus R_{cv}^2 est proche de $R^2 = 80\%$. Ce résultat montre que notre le meilleur modèle QSAR n'est pas sensible à cette opération de mettre à part une molécule et de la remettre dans la série d'apprentissage. C'est une première indication sur la stabilité du modèle QSAR retenu.

- Y-Randomisation

Tableau 9 : 20 premières itérations de Y-randomisation

Itération	R_r^2	$R_{r\ cv}^2$	Itération	R_r^2	$R_{r\ cv}^2$
1	2,07	0.00	11	2,27	0.00
2	19,48	6,81	12	16,86	0.00
3	0,75	0.00	13	4,82	0.00
4	3,50	0.00	14	15,95	0.00
5	6,69	0.00	15	20,37	5,83
6	0,86	0.00	16	7,29	0.00
7	4,14	0.00	17	1,62	0.00
8	1,00	0.00	18	4,27	0.00
9	5,61	0.00	19	10,90	0.00
10	15,55	0.00	20	7,97	0.00

Les résultats des 20 premières itérations de l'Y-randomisation sont donnés dans le **tableau 9**, il s'avère que les modèles élaborés avec les valeurs randomisées de $Y(E_7)$ ont des R_r^2 inférieur à 25%. D'autre part Roy [14] a proposé un paramètre, nommé R_p^2 , pour juger que le modèle n'est pas dû à **une chance**.

$$R_p^2 = R^2(R^2 - R_r^2)^{\frac{1}{2}} \quad (7)$$

Avec R_r^2 est la valeur moyenne des R_r^2 des modèles obtenus avec l'activité randomisée.

Ce paramètre doit être supérieur à 0,5 (50%).

Pour notre modèle QSAR la valeur de ce paramètre $R_p^2 = 0.695 > 0,5$, ce résultat indique que notre meilleur modèle existe réellement et il **n'est pas dû à une chance**.

D'après les résultats de la validation interne on peut conclure que ce modèle est **stable** et a un pouvoir **explicatif très acceptable**.

La **figure1**, représente la variation des résidus standardisés en fonction de l'activité expérimentale, on constate qu'il 'y a une répartition plus au moins homogène des résidus autour de 0.

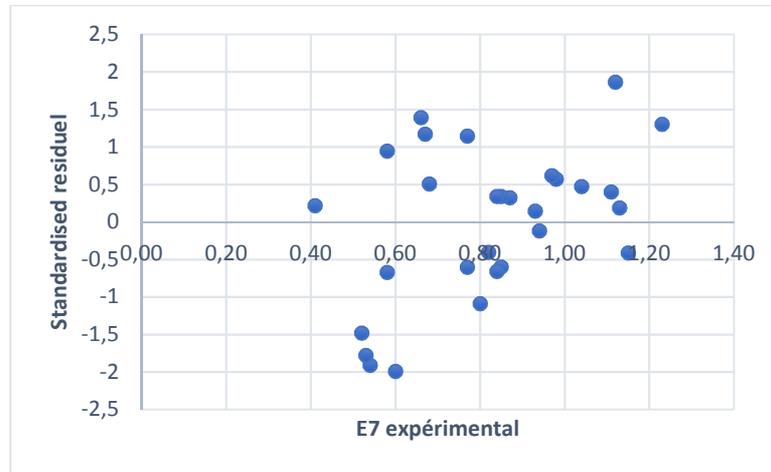


Figure 1 : *variation des résidus standardisés avec le potentiel Redox expérimental*

2) Validation externe

Cette validation se fait en deux étapes :

- Comparaison des paramètres statistiques
- Vérification des critères de Tropsha.

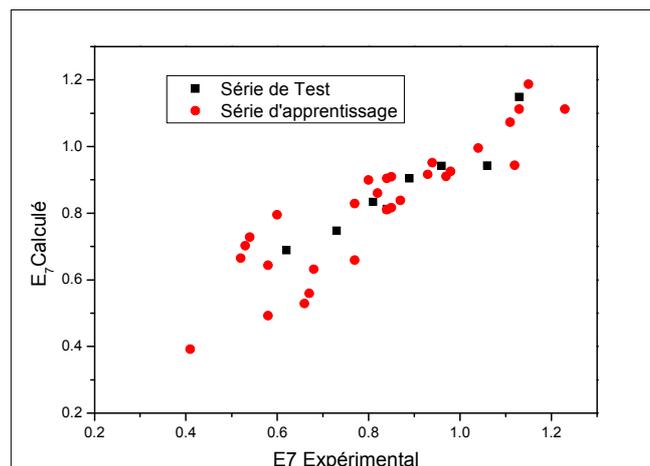


Figure 2 : Variation de E₇ calculé en fonction de E₇ expérimental

- Critères de tropsha

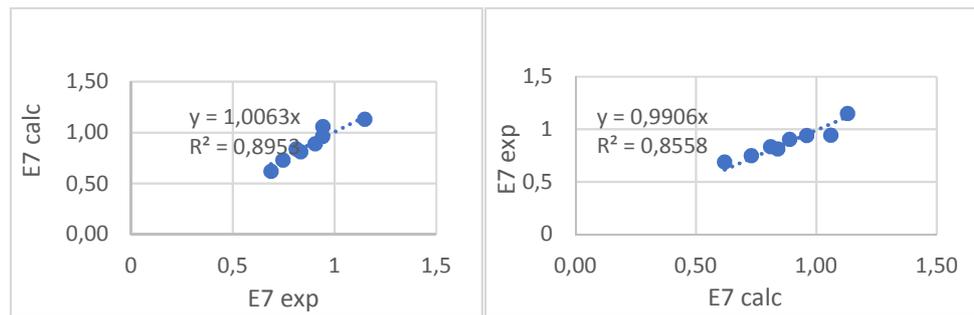


Figure 3: a- Variation de E_7 calc en fonction de E_7 exp (avec *intercepte* =0)

b- Variation de E_7 exp en fonction de E_7 calc (avec *intercepte* =0)

Tableau 10 : Validation externe du meilleur modèle

Série de test				Critères de Tropsha			
N	R ²	R ² _{cv}	SD	R ₀ ²	k	R ² -R ₀ ² /R ²	R ² -R ₀ ²
8	90,58	77,49	0,0471	89,53	1,0063	0,0011	0,0105

D'après Tropsha [15-17]

- $R^2 > 0.7$ (critère 1)
- $R^2_{cv} > 0.6$ (critère 2)
- $\frac{R^2 - R_0^2}{R^2} < 0.1$ et $0.85 \leq k \leq 1.15$ (critère 3)
- $\frac{R^2 - R_0'^2}{R^2} < 0.1$ et $0.85 \leq k' \leq 1.15$ (critère 4)
- $|R^2 - R_0^2| \leq 0.3$ (critère 5)

D'après le **tableau 10**, qui regroupe les résultats de la validation externe, il s'avère que :

- ✓ Les paramètres statistiques pour la série de test sont très satisfaisants $R^2(\text{test})=0.9058 > R^2(\text{modèle})$, $R^2_{cv}(\text{test})=0,7749 > R^2_{cv}(\text{modèle})$, $SD(\text{test})=0,047 < SD(\text{modèle})$.

- ✓ Tous les critères de Tropsha sont vérifiés
- ✓ La valeur de $k=1.006$ ($k'=0,990$) est très voisine de 1 ce qui indique que l'activité calculée, en utilisant l'équation de notre meilleur modèle, est très proche à la valeur expérimentale : $E_7(\text{calc})= k. E_7(\text{exp})$ et $(E_7(\text{exp})= k. E_7(\text{calc}))$.
- ✓ Les résultats de la validation externe montrent que le meilleur modèle QSAR **Eq.6** est très prédictif, il peut être utilisé pour la prédiction de l'activité anti-oxydante, quantifiée par E_7 , de nouveaux composés phénoliques.

2.4.2. Domaine d'applicabilité du meilleur modèle QSAR

Un modèle n'est pas universel, il est élaboré pour un nombre donné de composés. Pour cela il faut déterminer un domaine d'applicabilité pour chaque modèle.

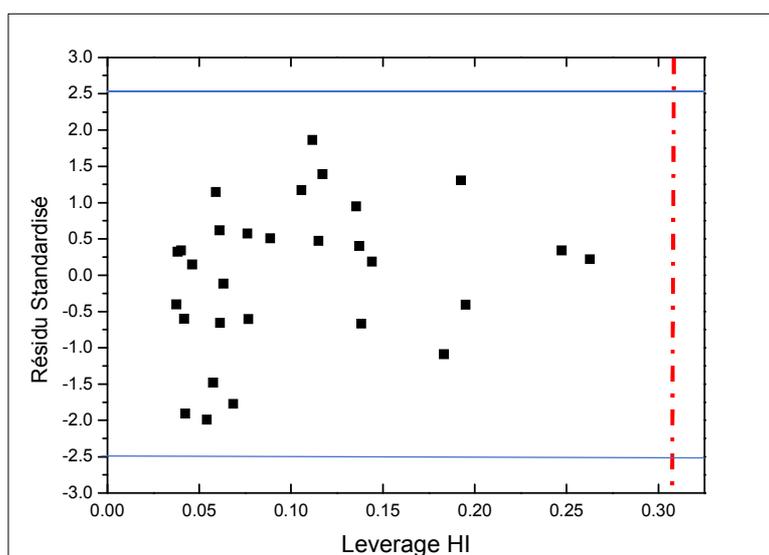


Figure 3 : Domaine d'applicabilité du meilleur modèle

La **figure 3** montre que toutes les observations ont des résidus standardisés comprises entre $[-2,2]$. En plus les « *leverages* » obtenus sont inférieurs à la valeur seuil **0,31** ($h^*=3p/n$), ce qui prouve l'absence d'observations aberrantes (*outliers*).

Les résultats de la validation externe et le domaine d'applicabilité montrent que le modèle QSAR élaboré dans ce travail peut être utilisé pour des objectifs prédictifs pour les composés

ayant des potentiel Redox (E_7) et des valeurs des descripteurs appartenant aux gammes des composés de la série d'apprentissage.

2.4.3. Interprétation mécanistique du meilleur modèle

Le meilleur modèle QSAR contient deux caractéristiques importantes déterminantes de l'activité anti-oxydante, à savoir, la facilité d'ioniser la molécule et la lipophilie de cette dernière.

Le coefficient positif obtenu pour les deux descripteurs dans le meilleur QSAR, appuie le concept que le **potentiel redox** (E_7) augmente avec l'augmentation du potentiel d'ionisation et le caractère hydrophobe de la molécule anti-oxydante.

Il s'avère qu'il y a une augmentation du **pouvoir antioxydant** quand la molécule est hydrophile et facile à ioniser (petit IP).

Conclusion

L'activité anti-oxydante de 41 composés phénoliques a été étudiée en respectant toute la méthodologie QSAR et tous les critères exigés par OECD [2].

Le meilleur modèle QSAR obtenu a montré non seulement une signification statistique, mais aussi une capacité prédictive (R^2_{test} est élevé et tous les critères de la validation externe « Tropsha » sont vérifiés) et a révélé que les faibles valeurs d'IP (DFT/B3LYP/6-311++G(d,p)) combinées avec le grand caractère hydrophile, entraînent une augmentation de l'activité anti-oxydante (diminution du potentiel Redox E_7) et une interprétation chimique du mécanisme. Ici, il est intéressant de mentionner que la variable IP (DFT) est un descripteur électronique, ce qui renforce que les effets électroniques ont un rôle très important dans le **mécanisme** par lequel un antioxydant peut désactiver un radical libre par transfert d'électrons. Le meilleur modèle contient également un descripteur de lipophilie/hydrophilie (logP) ce qui suggère que la différence de solubilité entre les phases, aqueuse et hydrophobe, joue un rôle crucial sur le pouvoir antioxydant de ces composés phénoliques.

Références Bibliographiques

- [1] M. Reis, B. Lobato, J. Lameira, A. S. Santos, C. N. Alves, *Europ. J. Med. Chem.*, **2007**, 42, 440.
- [2] OECD <http://www.oecd.org/chemicalsafety/risk-assessment>
- [3] E. S. Lien, S. Ren, H. H. Bui, R. Wang, *Free Radic. Biol. Med.* **1999**, 26, 285.
- [4] S. Steenken, P. Neta, *J. Phys. Chem.* **1982**, 86, 3661
- [5] S.V. Jovanovic, S. Steenken, M.G. Simic, Y. Hara, C.A. Rice-Evans, L. Packers (Eds.), *Flavonoids in Health and Disease*, Marcel Dekker, Inc., New York, **1998**.
- [6] F.G. Bordwell, J.P. Cheng, *J. Am. Chem. Soc.* **1991**, 113, 1736.
- [7] Gaussian 09, Revision D.01, Gaussian, INC, Wallingford CT, **2009**.
- [8] M. Filipovic, Z. Markovic, J. Đorovic, J. Dimitric, C. Markovic, B. Lucic, D. Amic. *C. R. Chem.* **2015**, 18, 492
- [9] L.V.B. Hoelz, B. A.C. Horta, J.Q. Araújo, M.G. Albuquerque, R. Bicca de Alencastro, J.F.M. Silva, *J. Chem. Pharm. Res.*, **2010**, 2, 291.
- [10] B. Modak, M.L. Contreras, F. Gonza 'lez-Nilo, R. Torres, *Bioorg. Med.Chem. Lett.* **2005**, 15, 309.
- [11] ACD/Labs, Release 12. **2009**, <http://www.acdlabs.com>
- [12] ChemSpider <http://www.chemspider.com/Search.aspx?rid=c107ad4e-3a23-4237-bd1a-b318f495dfb6>
- [13] MINITAB, State College, PA Minitab, Inc. **2006**.
- [14] P.P. Roy, S. Paul, I. Mitra, K. Roy, *Molecules.* **2009** 14, 1660.
- [15] A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* **2002**, 20, 269.
- [16] T. M. Martin, P. Harten, D. M. Young, E.N. Muratov, A. Golbraikh, H. Zhu and A. Tropsha, *J. Chem. Inf. Model.* **2012**, 52, 2570.
- [17] I.V. Tetko, I. Sushko, A.K. Pandey, H. Zhu, A. Tropsha, E. Papa, T. Oberg, R. Todeschini, D. Fourches, A. Varnek. *J. Chem. Inf. Model.* **2008**, 48, 1733.

Conclusion générale

Conclusion générale

Ce présent travail s'inscrit dans le cadre de l'étude de l'activité anti-oxydante des composés phénoliques en utilisant la méthode de modélisation QSAR.

L'objectif est de développer des modèles QSAR stables fiables et prédictifs pour la prédiction du pouvoir antioxydant d'une série constituée de 41 composés phénoliques simples.

QSAR se trouve à l'intersection de la chimie, les statistiques et la biologie, pour cette raison les deux premiers chapitres de cette thèse ont été consacrés à la présentation de l'activité anti-oxydante, la méthodologie QSAR/QSPR ainsi tous les outils nécessaires pour la mise en place d'un modèle QSAR. L'outil quantique nécessaire pour le calcul des descripteurs moléculaires à l'aide de la structure chimique et l'outil statistique pour lier l'activité étudiée aux descripteurs moléculaires.

Ces outils ont été utilisés pour élaborer des modèles QSAR en utilisant des descripteurs électroniques calculés par la méthode DFT/B3LYP/6-311++G(d,p) et un descripteur qui reflète l'hydrophobie/l'hydrophilie des antioxydants étudiés.

Le meilleur modèle obtenu en utilisant la méthode d'analyse de données la plus transparente MLR, a montré que la combinaison du potentiel d'ionisation IP_2 et le paramètre d'hydrophobie, facteur de transport, peut expliquer 80% de la variation de l'activité anti- oxydante, quantifiée par le potentiel Redox E_7 , des 41 composés étudiés.

La stabilité du meilleur modèle QSAR a été vérifiée en utilisant la validation croisée (R_{cv}^2), la Y-Randomisation et le calcul du paramètre de Roy.

Le modèle QSAR élaboré a montré un grand pouvoir prédictif ceci est confirmé par la validation externe ($R_{cv}^2(\text{test})$) et les critères de Tropsha). Ce modèle est considéré comme validé et applicable pour l'exploitation de cette série de composés phénoliques, le domaine d'applicabilité a montré l'absence de toute observation aberrante (*outier*).

Deux caractéristiques importantes déterminantes de l'activité anti-oxydante, à savoir, la facilité d'ioniser la molécule et la lipophilie de cette dernière sont contenues dans le meilleur QSAR.

Le meilleur modèle a révélé que les faibles valeurs d'IP (DFT/B3LYP/6-311++G(d,p) combinées avec le grand caractère hydrophile, entraînent une augmentation de l'activité anti-oxydante (diminution du potentiel Redox E_7) et une interprétation chimique du mécanisme. Ici, il est intéressant de mentionner que la variable IP (DFT) est un descripteur électronique, ce qui renforce que les effets électroniques ont un rôle très important dans l'explication du **mécanisme** par lequel un antioxydant peut désactiver un radical libre par transfert d'électrons.

IL est difficile d'obtenir un modèle QSAR parfait et de stabiliser un mécanisme de l'activité anti-oxydante pour plusieurs raisons :

- La différence des structures et les effets attracteurs/donneurs dans une base de données.
- La nature complexe de ce processus antioxydant.

Les perspectives de ce présent travail sont nombreuses. D'une part, nous envisageons d'utiliser le même modèle pour la même série de composés avec des activités mesurées avec d'autres méthodes.

D'autre part nous projetons d'élaborer des modèles QSAR pour d'autres bases de données d'antioxydants.

ملخص

في السنوات الأخيرة، الاهتمام بالمواد المضادة للأكسدة قد ازداد بشكل ملحوظ لعلاقتها بعدد من الخصائص العلاجية، الصيدلانية وغيرها...

في هذه الدراسة، تم تحليل النشاط المضاد للأكسدة لـ 41 مركب فينولي باستخدام طريقة $DFT/B3LYP/6-311++G^*$ لحساب واصفات كيميائية الكم. تم تطوير نماذج QSAR باستخدام تقنية الانحدار الخطي المتعدد (MLR). الاستقرار والقوة التفسيرية والقوة التنبؤية، لأفضل نموذج QSAR المحصل عليه تم تقييمهم بواسطة التحقق الداخلي، تقنية التوزيع العشوائي لـ Y والتحقق الخارجي، المجال الكيميائي الهيكلي تم التحقق منه بواسطة تقنية "leverage". أظهرت النتائج أن نموذج QSAR المطور قابل للتفسير، لديه استقرار جيد، قوة وقدرة كبيرة على التنبؤ.

RESUME

Ces dernières années, l'intérêt porté aux antioxydants, en relation avec leurs propriétés thérapeutiques, pharmacologiques... a augmenté considérablement.

Dans la présente étude, l'activité anti-oxydante de 41 composés phénoliques a été analysée à l'aide de la méthode $DFT / B3LYP/6-311 ++ G (d, p)$ pour calculer les descripteurs chimiques quantiques. Les modèles QSAR sont développés à l'aide de la technique de régression linéaire multiple (MLR). La stabilité, le pouvoir explicatif et le pouvoir prédictif, du meilleur QSAR modèle obtenu ont été évalués par validation interne, la technique de randomisation Y et la validation externe, leur domaine chimique structural a été vérifié par l'approche de « leverage ». Les résultats montrent que le modèle QSAR développé est interprétable, a une bonne stabilité, une robustesse et a un grand pouvoir prédictif.

ABSTRACT

In recent years, the interest in antioxidants, in relation to their therapeutic and pharmacological properties, has increased considerably.

In the present study, the antioxidant activity of 41 phenolic compounds was analyzed using the $DFT / B3LYP/6-311 ++ G (d, p)$ method to calculate the quantum chemical descriptors. QSAR models are developed using the Multiple Linear Regression (MLR) method. The stability, and the predictive power of the best QSAR model obtained were evaluated by internal validation, Y randomization and external validation, their structural chemical domain was verified by the leverage approach. The results show that the developed QSAR model is interpretable, has good stability, robustness and has great predictive power.