République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieure et de la Recherche Scientifique
Université de Médéa
Faculté des Sciences et de la Technologie

Mémoire
Présenté pour obtenir le diplôme de Master

En chimie

Spécialité : Normes et contrôle de qualité

Par :
Mr AMRAOUI Fethi

Mr BELZEKKOUR Hamid

Thème :

Prédiction de l'activité biologique des dérivés de 1,2- dithiole -3-thione

Soutenue publiquement le : 14 /06 /2016. Devant le jury composé de :

Mr HADIDI Noureddine	Université de Médéa	Président
Mr RAHAL Sofiane	Université de Médéa	Examinateur
Mr BITAM Said	Université de Médéa	Encadreur

Année universitaire 2015/2016

Remerciement

En premier lieu, je tiens à remercier notre Dieu, qui m'a donné la force pour accomplir ce travail.

J'adresse mes vifs remerciements à mon promoteur consultant, Dr: BETTAM Said pour m'avoir dirigé tout au long de ce travail.

Je remercie également Mr: GHRABA Mustapha, pour ses explications, remarques judicieuses et conseils qui m'ont été précieux pour la réalisation de ce travail.

Je présente mes chaleureux remerciements aux membres de jury qui ont consulté mon travail et aussi aux enseignants pour leur aides et orientation durant mes études.

Je remercie également tous les membres qui m'ont beaucoup aidé à réaliser ce travail dans de bonnes conditions.

Mes derniers remerciements et ce ne sont pas les moindres, vont à tous ceux qui ont participé de près ou de loin pour l'aboutissement de ce travail.

Dédicace

Je dédie ce travail à :

** En premier lieu à mes parents qui ont consenti beaucoup de sacrifices pour me permettre de réaliser mes objectifs. Qu'ils trouvent ici toute ma reconnaissance et ma gratitude.*

** A mes frères*

** A ma famille*

** A tous mes copains.*

** A tous mes voisins de la ferme école sans exception.*

**A tout le groupe de' chimie.*

Amraoiu fethi

Dédicace

Je dédie ce travail à :

** En premier lieu à mes parents qui ont consenti beaucoup de sacrifices pour me permettre de réaliser mes objectifs. Qu'ils trouvent ici toute ma reconnaissance et ma gratitude.*

** A mes frères*

** A ma famille*

** A tous mes copains.*

** A tous mes voisins de la ferme école sans exception.*

**A tout le groupe de' chimie.*

Belzekkour hamid

Résumé :

Une étude QSAR a été effectuée sur seize molécules analogues de 1,2-dithiole-3 thione. La régression linéaire multiple (MLR) a été utilisée pour quantifier les relations entre les descripteurs moléculaires et les propriétés de chimio prévention des dérivés du 1,2-dithiole-3-thione. La prédiction des modèles obtenus a été confirmé par la méthode de validation. Une forte corrélation a été observée entre les valeurs expérimentales et prédites des activités biologiques de l'activité spécifique de quinone réductase QR, se qui indique la validité et la qualité des modèles QSAR obtenus.

Mots Clés : 1,2-dithiole-3-thiones, modélisation moléculaire, QSAR, MLR.

Abstract :

QSAR studies have been performed on sextine molecules of 1,2-dithiole-3-thione analogues. A multiple linear regression (MLR) procedure was used to design the relationships between molecular descriptor and chemo-prevention properties of the 1,2-dithiole-3-thione derivatives. The prediction models obtained was confirmed by the method of validation. High correlation between experimental and predicted activity values was observed, indicating the validation and the good quality of the derived QSAR models.

Keywords : 1,2-dithiole-3-thiones, molecular modelling, QSAR, MLR.

الملخص:

دقل أتيرجادلارةسد ***QSAR*** ىلع ةتسد رشعة زجيئ نم لاثاممت *1,2-dithiole-3-thiones* . اانلمعتسد طقةير ادحذلاار

اي طخلا ادعتملد ميمكتل الارابتطات نيب اولاافصت اةيئيزجلا و اصئاصخلا اةيئايميكلا –اةيئاقولا ائيزجلت -*1, 2-dithiole-3-thione* تم قتدير ابنتلؤمنللوذج اصحتملل هيلع بواسطة اةينقتل . قد لوحظ وجود ارتباط اعل يين ايقلم اجتلرةيبي

واتملوةعة اشنللط ايبلولويج نلإلزيم ***QR***

الكلمات الجوهرية: 1,2-dithiole-3-thiones, نمذجة جزيئية, QSAR, MLR

Sommaire

Introduction générale……. 1

I. Généralités sur les 1 ,2-Dithiole-3-thiones

I.1. Introduction ……..... 3

I.2. Propriétés physico-chimiques……... 4

I.3. Les propriétés thérapeutiques des 1,2-dithiole-3-thiones .....… 5

I.4. Préparation des dithiolethiones …… 6

I.4.1. La méthode générale de la préparation des dithiolethiones ……..... 7

I.4.2. Préparation des dithiolethiones à partir de cyclanone …… 7

I.4.3. Méthodes spécifiques pour préparer les 1,2-dithiole-3-thiones 8

I.4.3.1. Préparation du 1,2-dithiole-3-thione à partir de l'acétophénone ……. 8

I.4.3.2. Préparation du fluorure de 1,2-dithiole-3-thione…… 8

I.5. La chimio-prévention et les mécanismes enzymatiques de QR et GST…… 9

I.5.1. Introduction…… 9

I.5.2. Mécanisme enzymatique…..……...…… 10

I.5.2.1. Le système du glutathion (GSH)…… 10

I.5.2.2. La glutathion S-transférase……. 11

I.5.2.3. Quinone réductase (QR)…….……. 12

I.5.3. Historique ……..….… 13

I.5.4. Bio activation par NQO1……..…. 14

I.5.4.1. Réactions de Bio-activation……...… 14

I.5.4.2. Agents anti tumoraux dirigés par enzyme ……...… 14

I.5.5. Mode d'action …….. 15

II. Modélisation moléculaire ……..... 19

II.1. Introduction…….. 19

II.2. Définition de QSAR ……..... 20

II.3.Histoire de QSAR…..…… 20
II.4.Les Modèles QSAR/QSPR……...…… 22
II.5. Les applications de l'étude QSAR……..… 22
II.6.Les descripteurs moléculaires……..... 23
II.6.1.Les types des descripteurs moléculaires……..... 23
II.6.1.1.Les descripteurs 1D ……...… 23
II.6.1.2.Les descripteurs 2D …… 23
II.6.1.3.Les descripteurs 3D …… 24
II.6.2.Réduction du nombre de variables ……..… 25
II.6.3.Sélection des variables pertinentes…… 25
II.7.Méthodes utilisées pour le développement de modèles QSAR/QSPR 26
II.7.1.Introduction …… 26
II.7.2.Méthodes de régressions linéaire et multilinéaire ……..... 27
II.7.2.1.Aperçu général……... 27
II.7.2.2.Evaluation préliminaire des données 28
II.7.2.3. Modèles statistiques…… 29
II.8. Régression linéaire multiple…… 29
II.8.1.Description de la méthode……..…… 30
II.8.2.Confiance accordée aux résultats ……... 31
II.8.3.Test de la signification globale de la régression…… 31
II.8.4.Validations des modèles……..… 34
III. Etude quantitative des relations structure-activité (QSAR) des dérivés de 1,2-dithiole-3-thione……... 35
III.1.Introduction……... 35
III.2. Données expérimentales …….. 36
III.3.Préparation de base de données …… 39
III.4. Sélection des variables …… 40

III.5.Les caractéristiques du modèle obtenu ……... 41

III.6. Quantification des descripteurs…… 45

Conclusion générale…… 46

Liste des figures

Chapitre I :

Figure I.1: Formule générale des 1,2-dithiole-3-thiones……..……...…... 3

Figure I.2: 4,5-diméthyl-3*H*-1,2-dithiole-3-thione……...…… 3

Figure I.3: Dérivés de 1,2-dithiole-3-thione : (1) Oltipraz. ; (2) ADT……..……...… 4

Figure I.4: La méthode générale de la préparation des dithiolethiones……..…….....…..... 7

Figure I.5 : Préparation des dithiolethiones à partir de cyclanone…….……..…. 7

Figure I.6 : Préparation du 1,2-dithiole-3-thione à partir de l'acétophénone……... 8

Figure I.7 : Préparation du fluorure de 1,2-dithiole-3-thione……..…..….. 8

Figure I.8 : La structure cristalline de GST……...…….…….…… 12

Figure I.9 : La structure cristalline de quinone réductase (QR)……..…….....…. 13

Figure I.10 : Mécanisme enzymatique de l'induction des enzymes de phase II par les dithiolethiones……..……. 18

Chapitre III :

Figure III.1: Modèle d'étude quantitative des relations structure/activité (QSAR)……..….. 36

Figure III.2 : Corrélation entre les activités biologiques expérimentales et prédites……..... 44

Figure III.3 : Représentation graphique des valeurs résiduelles en fonction des valeurs expérimentales de pCQR…….…….. 44

Liste des tableaux

Chapitre I

Tableau I.1 : Les propriétés thérapeutiques de quelques 1,2-dithiole-3-thiones…….…..... 6

Chapitre II

Tableau II.1 : Table d'analyse de variance…….. 33

Chapitre III

Tableau III.1 : Structures chimiques des dérives de 1,2-dithiole-3-thione……...…….. 37

Tableau III.2 : L'activité biologique des dérivés de 1-2, dithiole-3-thione……..... 39

Tableau III.3 : montre les 2 descripteurs retenus pour construire les modèles de QSAR…….. 41

Tableau III.4 : matrice de corrélation entre les descripteurs obtenus…….....**.....** 42

Tableau III.5 :Valeurs des activités prévues et résiduelles de l'ensemble de dérives de 1,2-dithiole-3-thione……..…….**.....** 43

Introduction générale :

Les ordinateurs sont devenus des outils indispensables en chimie moderne.

Leur rôle est essentiel, tant au niveau de la découverte de nouveau médicaments que du développement de ceux-ci. Les progrès rapides réalisés dans les logiciels et dans le matériel qui les accompagnent fait que la plus part des opérations qui étaient, jadis uniquement, réalisables par des informaticiens avertis peuvent maintenant être exécutées par des pharmaco-chimistes, avec des ordinateurs couramment employés aux laboratoires, pour autant qu'ils possèdent les notions élémentaires de mécanique quantique et autres équations qui ont trait aux molécules [1].

La modélisation moléculaire regroupe les techniques de visualisation, de manipulation, d'analyse et de calcul de la structure spatiale [2-4]. Les techniques de graphisme moléculaire permettent de représenter sur un écran la structure en 2D et en 3D d'une molécule, de la manipuler (rotations, translations, superpositions,...) de façon interactive ou d'analyser certains paramètres géométriques (distances, angles, surfaces,…).

Ces structures peuvent être obtenues à partir d'éléments tirés de la bibliographie (atome, résidus peptidiques, …), extrais d'une banque de données cristallographiques (*Cambridge Structural Database*, *Brookhaven Protéine Databank*) ou enfin construites sur la base d'observation expérimentale (RMN, RX). Le deuxième volet de la modélisation moléculaire est la mise en œuvre de méthodes permettant de réaliser des calculs d'énergie, des optimisations de géométrie, des simulations de mouvement moléculaire, Par opposition à la chimie théorique pure, les termes de chimie informatique ou chimie computationnelle résument clairement les objectifs de cette technique à savoir l'emplois de moyens informatiques pour la modélisation quantitative du comportement physicochimique d'espèce chimiques par un formalisme mathématique[5].

Les méthodes théoriques qui permettent de calculer les données relatives à la structure et aux propriétés d'une molécule peuvent être subdivisées en deux catégories : la mécanique moléculaire et la mécanique quantique (semi-empirique, ab initio et DFT) [6].

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

Les techniques usuelles reposent sur la caractérisation des molécules par un ensemble de descripteurs, nombres réels mesurés ou calculés à partir des structures moléculaires. Il est alors possible d'établir une relation entre ces descripteurs et la grandeur modélisée [7].

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

Le principal objectif de ce travail est l'application de la modélisation moléculaire dans le développement des méthodes QSAR des propriétés chimio-préventives pour prédire l'activité biologique attendue dans des nouvelles molécules bioactives pour de certains dérives de 1,2- dithiole-3-thione.

En plus d'une introduction et d'une conclusion générales, notre travail comporte en trois chapitres, dans le premier chapitre nous présenterons une étude bibliographique, qui s'articule sur deux points. Le premier consiste des généralités sur 1,2-dithiole-3-thione. Le deuxième point nous présenterons la notion de chimio-prévention et le mécanisme enzymatique de QR et GST.

Dans le deuxième chapitre nous présenterons des généralités sur les méthodes QSAR et la régression linéaire.

Le troisième chapitre de ce travail, présente les résultats obtenus par l'étude QSAR de l'activité anticancéreuse des dé rivés de 1,2-dithiole-3-thione et leurs discussions.

I. Généralités sur les 1 ,2-Dithiole-3-thiones :

I.1. Introduction :

Les 1,2-dithiole-3-thiones (dithiolethiones) sont des hétérocycles sulfurés dont la structure générale est présentée sur la figure I.1.

Figure I.1: Formule générale des 1,2-dithiole-3-thiones.

La première synthèse de cette famille de composé a été réalisée en 1884 par A.G. Barbaglia, chimiste italien. Il a isolé et purifié une substance de formule $C_5H_6S_3$ obtenue par une réaction entre le soufre et l'isovaléraldéhyde (3-méthylbutanal)[9]. Barbaglia n'a pas déterminé sa structure, mais il a mesuré son point de fusion et a donné la méthode de synthèse, il a stipulé que cette substance serait le 4,5-diméthyl-3*H*-1,2-dithiole-3-thione (figure I.2).

Figure I.2: 4,5-diméthyl-3*H*-1,2-dithiole-3-thione.

Le plus important des composés de cette famille est le 3*H*-1,2-dithiole-3-thione (D3T, R1=R2=H) a été synthétisé en 1948 [10]. Nombreux dérivés ont été préparés entre les années 1940 et 1950, notamment en France, dont les utilisations potentielles de ces composés seraient dans les domaines industriels et pharmaceutiques [11].

Le 5-(4-méthoxyphényl)-3*H*-1,2-dithiole-3-thione (ADT, composé 2; figure I.3) et de 4-méthyl-5-pyrazinyle-3*H*-1,2-dithiole-3-thione (Oltipraz, composé 1; figure I.3) ont été testés en matière d'efficacité thérapeutique chez l'homme. ADT est utilisé comme stimulant de la salive [12] et oltipraz a été utilisé dans le traitement de la schistosomiase [13].

1 2

Figure I.3: Dérivés de 1,2-dithiole-3-thione : (1) Oltipraz. ; (2) ADT

Dans les années 1980, l'intérêt de cette classe de composés a été fortement stimulé par les travaux du groupe de Bueding, qui a indiqué que l'oltipraz et ADT, lorsqu'ils sont administrés à des souris, les niveaux hépatiques augmentent significativement en glutathion et glutathion-S-transférase (GST) et neutralisent les effets toxiques de paracétamol et du tétrachlorure de carbone [13].

De nombreuses études sur les dithiolethiones ont été réalisées depuis lors. Les deux composés, oltipraz et ADT ont fait l'objet de plusieurs essais de chimio prévention chez l'homme, de nombreux nouveaux analogues ont été synthétisés et évalués, et de nouveaux mécanismes de chimio prévention de dithiolethiones ont été identifiés.

I.2. Propriétés physico-chimiques :

Les 1,2-dithiole-3-thiones sont généralement inodores et possèdent un goût amer. Les petits dérivés alkyles de 1,2-dithiole-3-thiones ont une odeur caractéristique.

La configuration électronique des 1,2-dithiole-3-thiones inhabituelle se traduit par un nombre de ces propriétés physiques. Par conséquent, ils sont tous de couleur, variant généralement du jaune au rouge. Les 1,2-dithiole-3-thiones absorbent dans les domaines UV et visible. Les spectres UV de plusieurs dérivés de 1,2-thiole-3-thione ont été rapportés.

Challenger et son équipe rapporte les valeurs de λ_{max} de 1,2-dithiole-3-thione dans le cyclohexane vers 230, 254, 336 et 415 nm [14].

Spindt, Steevens, Baldwin, Landis et Hamilton ont montré que les substituants alkyles provoquent un faible déplacement des bandes d'absorption, tandis que, Luttringhaus et Cleve montrent des déplacements importants avec les substituants aryles [14].

Les spectres RMN ^{13}C des 1,2-dithiole-3-thiones présentent des déplacements chimiques vers 216.7, 140.2 et 155.1 ppm pour les carbones C3 et C4 et C5 respectivement [15].

Mecke et Luttringhaus déterminent les spectres infrarouges d'un certain nombre de dithiolethiones. Les spectres IR présentent le caractère aromatique du cycle de ces molécules [14].

Les 1,2-dithiole-3-thiones possèdent des points de fusion souvent élevés. Ces composés sont insolubles dans l'eau et solubles dans certains solvants organiques usuels comme le chloroforme. Tandis que, les 1,2-dithiole-3-thiones sont solubles dans les acides minéraux, dans ces conditions, ces composés présentent une grande stabilité résulte de la formation des sels stables dont la charge positive se répartie dans le cycle.

Les 1,2-dithiole-3-thiones forment des complexes instables avec les halogènes; Lorsqu'ils sont mélangés avec les halogènes dans le benzène. Challenger et son équipe ont prouvé que l'halogène ne s'additionne pas sur la double liaison carbone-carbone. Le traitement du produit obtenu par l'addition du brome sur 1,2-thiole-3-thiones avec de l'eau régénère le composé original [14].

I.3. Les propriétés thérapeutiques des 1,2-dithiole-3-thiones :

Les dithiolethiones jouent un rôle très important dans plusieurs domaines. Ils sont utilisés comme des promoteurs dans le champ de la pharmacologie où la sulfarlem (ADT) a été utilisée pendant longtemps comme médicament thérapeutique, ainsi que l'oltipraz (OPZ) qui possède la propriété d'un inhibiteur efficace au HIV-1. Aussi, le composé diclofenac a été Utilisé pendant longtemps comme anti-inflammatoire et dans le traitement des maladies du rhumatisme surtout chez les personnes âgées. Un grand nombre des dithiolethiones sont connus comme des agents chimio protecteurs, et peuvent être accomplis par l'usage de produits chimiques naturels et synthétiques pour retarder, bloquez ou renversez le processus cancérigène. Nous citons quelques dithiolethiones qui sont utilisées actuellement comme des médicaments efficaces pour un grand nombre de maladies (tableau I.1) [16].

Tableau I.1 : Les propriétés thérapeutiques de quelques 1,2-dithiole-3-thiones.

Composés	***Utilisation Thérapeutique***
Oltipraz	• Traitement de sida : Inhibiteur efficace de la reproduction du HIV-1 par inactivation de la transcriptase inverse. • Contre la fibrose du foie. • Agents chimio-protecteurs, utilisés pour traiter une grande gamme de cancer, peau touchante, foie, poitrine, poumon • Susceptible de contrer la résistance à l'insuline dans les cas de maladies du foie. • Traitement de Schistosome ou bilharzia :(parasite chronique qui provoque les cancers Vésicaux). • Antioxydant (substance capable de neutraliser ou de réduire les dommages causés par les radicaux libres dans l'organisme).
Sulfarlem anetholedithione ADT	• Comme cholérétique : toute substance qui augmente le débit de la sécrétion biliaire (active la sécrétion de la bile) utilisé depuis longtemps. • Agents chimio-protecteurs
S-diclofenac	• Anti-inflammatoire.
D3T	• Agents chimio-protecteurs. • Antioxydant. • Possibilité de protéger les cellules nerveuses.

I.4. Préparation des dithiolethiones :

Vu l'importance accordée aux dithiolethiones et ses dérivés, plusieurs méthodes sont citées pour la préparation des dithiolethiones. Nous décrivons quelques méthodes ci dessous:

I.4.1. La méthode générale de la préparation des dithiolethiones :

Les dithiolethiones et ses dérivés ont attiré l'attention depuis plusieurs années de leurs activités biologiques considérables. Cette importance nécessite des méthodes de préparation efficace.

Les procédures habituelles pour la synthèse des dithiolethiones sont l'utilisation de la réaction d'un 3-oxoester avec P_4S_{10} et le soufre, dans le toluène bouillant ou xylène [17].

Figure I.4: La méthode générale de la préparation des dithiolethiones.

I.4.2. Préparation des dithiolethiones à partir de cyclanone :

A.THUILLIER et J.VIALLE ont montré qu'il était possible de condenser une molécule de sulfure de carbone, avec une cyclanone (1) et d'obtenir après méthylation le composé α [bis (methylthio)-méthyléne]-cyclanone (2). Par action du penta sulfure de phosphore sur la cétone précédente (2), ils obtiennent les 1,2- dithiole-3-thione (3) [18].

Figure I.5 : Préparation des dithiolethiones à partir de cyclanone.

I.4.3. Méthodes spécifiques pour préparer les 1,2-dithiole-3-thiones :

I.4.3.1. Préparation du 1,2-dithiole-3-thione à partir de l'acétophénone :

La réaction d'acétophénone avec CS_2 en présence de 2 équivalents de KH dans une solution de THF/DMPU (N,Nj-dimethylpropyleneurea) puis par le HMDT (hexaméthyldisilathiane) et C_2Cl_6, donne des dithiolethiones avec un bon rendement [19].

Figure I.6 : Préparation du 1,2-dithiole-3-thione à partir de l'acétophénone.

I.4.3.2. Préparation du fluorure de 1,2-dithiole-3-thione :

Dans cette méthode de préparation on utilise les perfluoroketenes diethylthioacetal (1) de manganèse suivie d'une sulfuration à 210 °C, le rendement arrive jusqu'à 99% [20].

Figure I.7 : Préparation du fluorure de 1,2-dithiole-3-thione.

Le chauffage de perfluoroketene diethylthioacetal (1) avec $MgBr_2$ + S_8 à 210°C pour une durée supérieure à 5 min donne un produit insoluble, c'est le 4-fluoro-5-trifluoromethyl- 1,2-dithiole-3-thione (3). Cette réaction passe par un intermédiaire de k-bromo-k-dithiocrotonic du trifluoromethyl (2), qui devient après 5 minutes de réaction le 4-fluoro-5-trifluoromethyl-1,2-dithiole-3-thione.

Le composé (2) peut être obtenu avec un bon rendement en chauffant le mélange à 180°.

I.5. La chimio-prévention et les mécanismes enzymatiques de QR et GST :

I.5.1. Introduction :

La chimio prévention du cancer consiste à utiliser des agents chimiques pour prévenir ou inhiber le développement du processus de carcinogenèse. Cette intervention peut être envisagée à tous les stades de la carcinogenèse depuis l'apparition des premières anomalies moléculaires dans des cellules encore morphologiquement normales jusqu'au stade de tumeur invasive. La chimio prévention peut aussi être utilisée pour tenter de diminuer l'apparition de récidives tumorales ou de nouvelles tumeurs chez des patients déjà traités pour un cancer.

Les composés administrés à des populations à risque, voire à toute la population doivent être efficaces tout en présentant une toxicité minime pour pouvoir être largement utilisés. Au sens large, la prévention nutritionnelle fait aussi partie de ce concept, mais elle n'est pas considérée comme une chimio prévention au sens strict sauf lorsqu'elle est appliquée via des éléments isolés (vitamines, calcium, composés spécifiques...) [21].

Le développement de ces agents repose sur des bases classiques:

- Phase I : appréciant la toxicité et la pharmacodynamie avec des études en dose unique puis en doses répétées sur des périodes pouvant aller de quelques jours à un an.

- Phase II : avec études randomisées en double aveugle évaluant l'efficacité du produit sur des marqueurs intermédiaires corrélés à l'apparition des tumeurs comme par exemple l'étude de la prolifération cellulaire colique.

- phase III : avec des études randomisées en double aveugle évaluant l'efficacité réelle du produit sur l'apparition ou la récurrence des tumeurs [22].

I.5.2. Mécanisme enzymatique :

Une fois entrés dans l'organisme, les xéno-biotiques sont susceptibles d'interagir avec toutes les molécules qu'ils rencontrent. La toxicité de la plupart des contaminants est relativement peu élevée dans leur état natif mais les interactions chimiques avec les biomolécules entraînent généralement la formation de composés plus réactifs et nocifs pour l'organisme ; c'est le phénomène de bio-activation.

Les propriétés des xéno-biotiques bioactifs et donc les réactions spécifiques qu'ils peuvent avoir avec différents composés cellulaires font que les dommages exercés par un contaminant auront lieu préférentiellement sur certains types de biomolécules, appelées molécules-cibles. Cela dépend de l'accessibilité de la biomolécule, de sa réactivité mais également de son rôle dans la cellule, dont l'altération sera fortement dommageable s'il est critique pour le fonctionnement.

Le mécanisme de bio-activation aux effets les plus délétères est la production de radicaux libres, c'est-à-dire de molécules possédant des électrons non appariés et présentant un fort potentiel oxydant. En effet, ces espèces réactives peuvent oxyder et modifier de façon covalente les composants cellulaires, avec des conséquences potentiellement dramatiques dans le cas des macromolécules comme les protéines et l'ADN, mais aussi des lipides membranaires. Comme déjà évoqué, les processus de métabolisation des xéno-biotiques sont une source importante de radicaux libres ; certains composés peuvent générer des espèces réactives en découplant le transfert d'électron du NAD(P) H vers l'oxygène au niveau des microsomes [23].

I.5.2.1. Le système du glutathion (GSH) :

Le glutathion (GSH) est un tri-peptide ubiquitaire, qui constitue le thiol le plus abondant dans les cellules. Il a pu être montré qu'il joue un rôle central dans la défense cellulaire contre le stress oxydant et les xéno-biotiques [24]. Le maintien d'un taux de GSH normal pourrait donc être impliqué dans la résistance à la toxicité de divers oxydants, dont les effets peuvent entraîner une mortalité cellulaire rapide ou une modification des cascades de transduction du signal.

Les défenses cellulaires contre les espèces réactives de l'oxygène mettent en jeu des enzymes anti-oxydantes: le super-oxyde dismutase et la catalase convertissent respectivement O^{-}_{2} en H_2O_2 et H_2O_2 en H_2O, par exemple.

Mais le cycle redox du GSH est également très important puisqu'il permet de diminuer les taux des espèces réactives de l'oxygène, ou des produits de la peroxydation lipidique en les transformant en alcools lipidiques non toxiques [25].

I.5.2.2. La glutathion S-transférase :

Est une protéine dimérique (homo- ou hétéro-dimérique avec des sous-unités de 23 à 28 kDa) appartenant à une grande famille d'iso-enzymes multifonctionnelles, les GST, qui font partie de la classe II des enzymes de détoxification. La réactivité du groupe nucléophile thiol du GSH est une voie importante du métabolisme de nombreux xéno-biotiques lipophiles et de composés endogènes possédant un centre électrophile. La conjugaison peut, suivant la nature du substrat électrophile, être non enzymatique ou être catalysée par les GST. Ces enzymes, en conjuguant de nombreux xéno-biotiques au glutathion, rendent ces composés hydrosolubles, et permettent leur détoxification. Mais les GST séquestrent également des toxines par une liaison de forte affinité [26]. Le rôle des GST est donc clairement celui de protéger les cellules contre la cyto-toxicité de divers composés. Cette superfamille comprend quatre classes d'enzymes cytosoliques (A, M, P et T) et une classe d'enzymes microsomales. La distribution des iso-enzymes de la GST est tissu-spécifique.

❖ Les GSTA : sont le produit de deux gènes, *GSTA* 1 et *GSTA2,* qui sont surtout exprimés dans le foie, le rein ou l'intestin grêle,

❖ La GSTTI : est abondante dans les érythrocytes. Récemment un polymorphisme de délétion a été retrouvé pour la OST Tl-l leucocytaire qui explique que l'enzyme soit absente chez 25 à 40 % de la population [27]. L'absence de la OST Tl-l pourrait être associée à l'augmentation de la génotoxicité de certains composés (oxyde d'éthylène, dichlorométhane ...) dans les lymphocytes [28].

❖ La GSTMl-l : est exprimée de façon variable dans les différents tissus humains mais on estime que 40 à 50% de la population a un phénotype nul pour cette isoenzyme à cause d'un polymorphisme de délétion, et n'exprime donc pas la OSTMl-l [29].

❖ La GSTPl-l : produit du gène *GSTP* 1, est exprimée comme la OST majoritaire dans le poumon, le sein, le gros intestin ou la vessie [30]. Elle présente des polymorphismes de la région codante au niveau des codons 105 et 114, qui lui confèrent des activités catalytiques différentes [31].

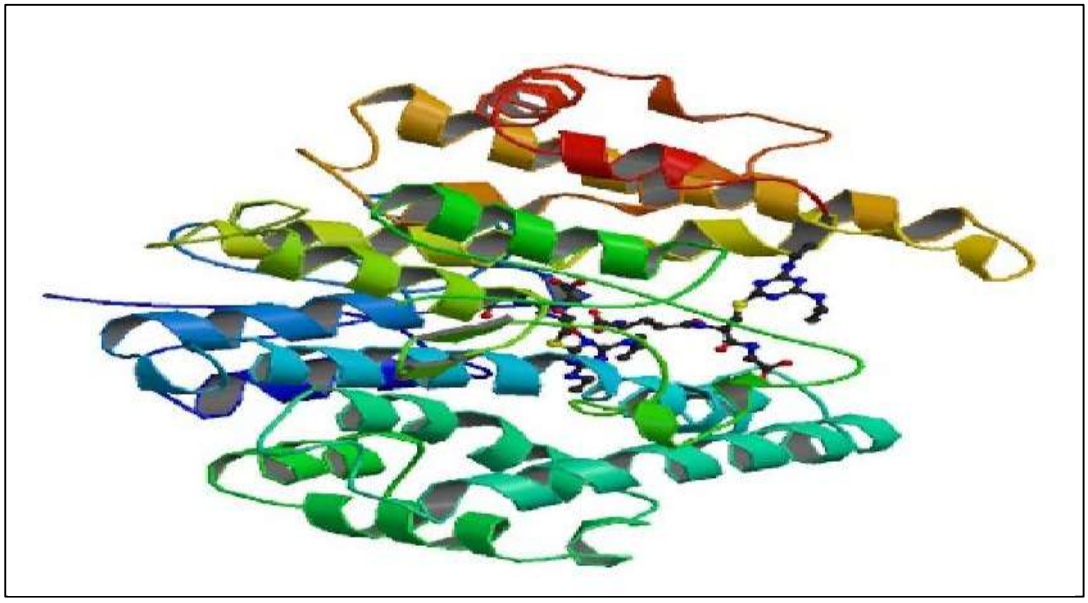


Figure I.8: La structure cristalline de GST [32].

I.5.2.3. Quinone réductase (QR):

La quinone réductase (QR), une des enzymes représentatives de la phase II, est un flavo-protéine qui catalyse la réduction et la désintoxication des quinones et des dérivés électrophiles de quinone. QR a été montré pour protéger les cellules mammifères contre le recyclage redox, l'effort oxydant, et la néoplasie en convertissant les quinines toxiques en hydroquinones et en induisant l'autre enzyme de détoxication telle que le glutathion, le glucuronidyl-, et les enzymes de sulfo-transférase en réponse aux xéno-biotiques, antioxydants, oxydants, métaux lourds, et aux rayonnements [33] .

Par conséquent, une régulation de QR est vraisemblablement un bio-marquer utile d'activité métabolique de la phase II et d'élimination cancérogène [34].

La mesure de l'activité d'induction de QR en cellules Hepa1c1c7 cultivées est une méthode de dépistage simple pour détecter la capacité des composés d'agir en tant que chimio-préventeurs potentiels [35].

Les quinones comportent une grande classe des composés aromatiques, trouvée largement aux usines, sous la forme générale (de benzoyle) de $C_6H_4O_2$, mais ont également trouvé dans toutes les organisations comme flavonoïdes, coenzymes d'électron-transport et produits finaux métaboliques d'oxydation [36,37].

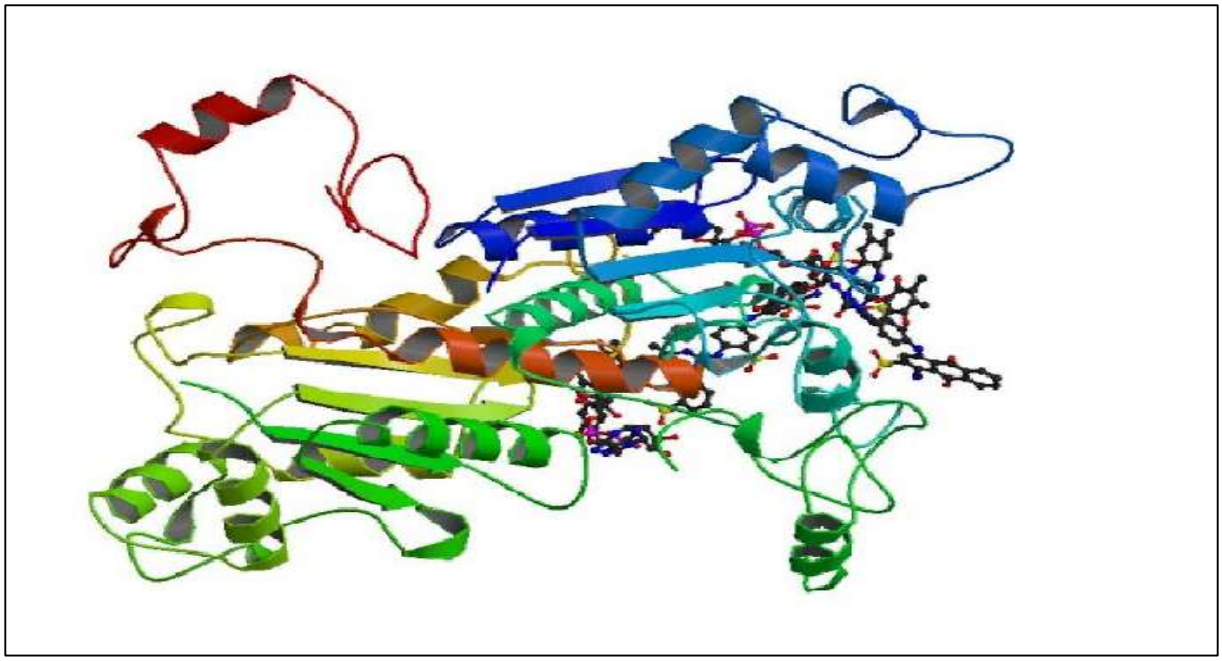


Figure I.9 : La structure cristalline de quinone réductase (QR)[38].

I.5.3. Historique :

Le NAD (P) H : on pense que donc la famille de gènes de l'oxydoréductase d'accepteur de quinone (NQO) est antique - probablement entre 2 et plus de 3 milliards d'années.

Les flavo-enzymes cytosoliques de NQO catalysent la réduction salutaire de deux-électrons de quinones aux hydroquinones. La réaction de deux-électron empêche la réduction de quinones par les réductases d'un-électron, qui auraient comme conséquence la formation des espèces réactives de l'oxygène (ROS) produites par le recyclage redox des semi-quinones en présence d'oxygène moléculaire [36,37].

Chez les animaux, en plus du rôle probable dans la détoxication de beaucoup de quinones diététiques dérivées des usines, des enzymes de NQO ont été montrées pour catalyser l'activation réductrice des quinones chimio thérapeutiques telles que des mitomycins et des aziridinylbenzoquinones [39,40]. Les hydroquinones dérivées de ces agents semblent être chimiquement plus réactifs, en raison de l'auto-oxydation ou de la remise en ordre rapide qui, dans l'un ou l'autre cas, produiront des espèces réactives qui peuvent tuer cellules malines [41,42].

I.5.4. Bio activation par NQO1 :

I.5.4.1. Réactions de Bio-activation :

Les propriétés chimiques de l'hydroquinone formée après que la réduction négociée par NQO1 d'une quinone détermine si NQO1 catalyse l'activation ou la désactivation. Non, toutes les hydroquinones sont chimiquement écurie et dans certains cas le métabolisme par NQO1 rapporte un produit plus actif qui peut s'auto oxyder pour produire les espèces réactives de l'oxygène ou subir la remise en ordre pour produire des espèces de alkylation. A titre d'exemple : la réduction et l'activation des dérivés nitrés ont trouvé en nourritures cuites telles que 4 nitroquinoline-1-oxide (4NQO). Sugimura et collègues ont rapporté l'activation de 4NQO à 4 hydroxy-aminoquinoline-1-oxide par une enzyme de foie de rat [43].

L'enzyme prédominante responsable de ceci était NQO1 [44]. Comme avec la plupart des substrats, qui, bio-activés par NQO1, les naphtho-quinones (NQ) et les dinitropyrenes (DNP) peuvent être activés ou désactivés par NQO1 selon des groupes de substituant et leur endroit. Induction de NQO1 avec de l'hydroxy-anisole butylé, rats protégés des effets toxiques de 2 methyl-1,4-NQ mais renforcés la toxicité de 2 hydroxy-1,4-NQ [45].

Pour DNPs qui sont trouvés en échappement diesel, le positionnement des groupes nitro influence considérablement le clasto-génique potentiel des composés activés. En outre NQO1 augmente la muta-génicité de 1.6 - mais non de 1.3 - ou de 1,8-DNP dans l'analyse de salmonella typhimurium d'Ames [46].

I.5.4.2. Agents anti tumoraux dirigés par enzyme :

Les médicaments qui peuvent produire les métabolites d'alkylation après la réduction, sont nommés les agents d'alkylation bio réactive [47]. Le résultat du bio-réduction est la production des espèces d'alkylation ou des espèces actives de l'oxygène selon les propriétés chimiques des composés subissant la réduction enzymatique. La toxicité sélective des agents d'alkylation bio-réductive aux tumeurs a été basée sur les lieux, que les cellules hypoxiques près des noyaux nécrotiques des tumeurs auraient une plus grande propension pour le métabolisme réducteur [47] . Le développement

anti tumoral enzyme-dirigé de drogue, exploite les enzymes bioactives qui sont exprimées aux niveaux élevés en à tissu impliqué relatif aux tumeurs [48,49].

NQO1 est exprimé à un niveau élevés dans beaucoup de tumeurs pleines, et est un candidat possible pour une approche enzyme-dirigée. Un examen détaillé de découverte enzyme-dirigée de drogue s'est concentré sur des quinones a été récemment édité [50].

Puisque NQO1 est présent dans les tissus impliqués aussi bien que la tumeur du tissu humain, il est possible que le tissu normal de toxicité puisse être une issue dans la thérapie avec des quinones anti tumorales de NQO1-dirigées. L'ADN est vraisemblablement la cible des agents d'alkylation enzyme-dirigés, ainsi la fraction de croissance plus élevée des tumeurs peut encore procurer l'occasion à la toxicité sélective [51].

I.5.5. Mode d'action [52]:

La recherche sur l'activité Chimio-préventive de cancer de l'oltipraz a été lancée au commencement sur la base de sa capacité d'induire les enzymes cytoprotective et d'élever le glutathion, comme mentionné ci-dessus. Les études suivantes se sont concentrées presque exclusivement sur l'ampleur de la modulation du cytoprotection par l'oltipraz, le mécanisme fondamental et le point auxquels la modulation du cytoprotection explique l'activité Chemopreventive de cancer de l'oltipraz.

Oltipraz a été montré pour induire des enzymes de la phase II importante pour la désintoxication des carcinogènes et des oxydants, comme NQO1, sous-unités multiples de GST, synthétase de cystéine de glutamate, hydrolase d'époxyde, et UDP-glucuronosyl- transférase. Un mécanisme important par lequel l'oltipraz induit des enzymes de la phase II est l'activation du facteur de la transcription Nrf2. Nrf2 est normalement lié par son répresseur Keap1 et visé pour la dégradation par le système ubiquitin-proteosome.

Les inducteurs des enzymes de la phase II, y compris l'oltipraz perturbent le Nrf2- Keap1 complexe et font transférer Nrf2 au noyau où il dimerizes avec ses associés tels que Maf, lie à l'élément antioxydant de réponse (ordre de noyau : 5 le `- le `TGACnnnGC-3) dans la région d'instigateur des gènes de la phase II, et stimule la transcription de gène. Le coup de grâce de Nrf2 a rendu des enzymes de la phase II non sensibles à l'oltipraz et a causé la perte de son efficacité chimio-préventive de cancer

chez les modèles animaux du cancer de réservoir souple et de cancer de l'estomac. L'activation de la signalisation Nrf2 est donc critique pour l'inhibition du développement de cancer.

Cependant, il n'est pas clair dans quelle mesure le rôle préventif de cancer de Nrf2 dépend de l'induction des enzymes de la phase II ou qui l'enzyme spécifique de la phase II est d'importance primordiale, parce que Nrf2 règle également beaucoup les gènes de la non-phase II, et la réponse de ces gènes à l'oltipraz sous le coup de grâce Nrf2 est aussi bien bloquée probablement. La modification des résidus critiques de cystéine de Keap1 par des inducteurs et des facteurs de force, qui libère Nrf2 de Keap1, a été identifiée comme mécanisme principal de l'activation Nrf2. Bien que la réaction directe de l'oltipraz avec des thiols de la cystéine Keap1 n'ait pas été encore montrée, les dithiolethiones sont réactifs avec des thiols.

D'ailleurs, l'oltipraz et les D3T ont été montrés pour produire des espèces réactives de l'oxygène en cellules qui peuvent plus loin contribuer à l'activation Nrf2, pendant que les espèces réactives de l'oxygène peuvent oxyder les thiols de cystéine de Keap1 et/ou activer d'autres protéines impliquées dans l'activation Nrf2 telle que la protéine kinase C .

Des mécanismes additionnels peuvent être impliqués dans le règlement de certaines enzymes de la phase II par l'oltipraz. L'instigateur NQO1 humain est connu pour contenir un accepteur d'AP-1- [élément de réponse de TPA (`d'ordre 5 de noyau - `TGACTCA-3).

En cellules humaines d'adénocarcinome des deux points HT-29, l'oltipraz a élevé de manière significative l'ADN messagère les niveaux d'AP-1 factorise aussi bien que celui du redox factor-1, qui augmente l'attache AP-1, et du déplacement de l'emplacement AP-1 dans l'instigateur NQO1 dans une construction de journaliste a eu comme conséquence la perte de 65% de l'induction par l'oltipraz. UDP-glucuronosyl - transférase 1A6 de rat porte un élément xéno-biotique de réponse (`d'ordre 5 de noyau - `TGCGTG-3), qui a été montré pour négocier au moins en partie l'induction de cette protéine par l'oltipraz dans les hépatocytes primaires de rat.

L'élément xéno-biotique de réponse est identifié par un complexe hétéro-dimerique d'hydrocarbure arylique translocator nucléaire de récepteur et de récepteur d'hydrocarbure d'aryle.

Le traitement des cellules humaines de la tumeur hépatique HepG2 avec l'oltipraz a mené à la plus grande attache du récepteur arylique d'hydrocarbure et à la translocation nucléaire de récepteur d'hydrocarbure d'aryle à l'élément xéno-biotique de la réponse UDP- glucuronosyl-transférase 1A6. Oltipraz a été également montré pour se protéger contre certains carcinogènes en empêchant le cytochrome P450s (CYPs ou phase I enzymes).

L'activation métabolique d'AFB1 par des enzymes de la phase I est exigée pour son activité cancérogène, et l'oltipraz a été montré pour empêcher l'activation AFB1 en agissant en tant qu'inhibiteur concurrentiel et irréversible de CYP1A2 et de CYP3A4 exprimés en bactéries et dans les hépatocytes humains primaires et de CYP2B en hépatocytes primaires de rat.

Les expériences de rat ont prouvé que l'oltipraz a favorisé la dégradation de CYP2B1 et ont empêché le métabolisme de CYP1A2-mediated d'AFB1. D'ailleurs, une dose simple de mg 125 chez l'homme a semblé réduire l'activité CYP1A2 de f 75%, comme mesuré par le dégagement de la caféine par démethylation, et l'administration quotidienne de l'oltipraz de mg 125 pendant 8 jours a semblé avoir comme conséquence davantage d'inhibition de CYP1A2. Cependant, l'effet de l'oltipraz sur CYPs n'est apparemment pas toujours inhibiteur.

Oltipraz a été montré pour induire de manière significative CYP1A2 et CYP3A2 dans le foie de rat en vie, CYP1A dans le rein et le poumon de rat, et CYP1A1 en cellules de la tumeur hépatique H4IIE de rat et cellules humaines du carcinome Caco-2 de deux points dans la culture. Intéressant, la structure de pyrazinyl est probablement principalement, sinon entièrement, responsable de l'effet inductif de l'oltipraz sur CYPs, parce que le composé non substitué, D3T, est en grande partie exempt d'un tel effet. En plus de la phase de modulation I et II enzymes, une étude a également indiqué l'activité anti-angiogenique de l'oltipraz, montrant en vie l'inhibition de formation de micro-vessel, et des analyses, qui était accompagnée de bystrong de xéno-graft d'angiosarcome chez les souris nues.

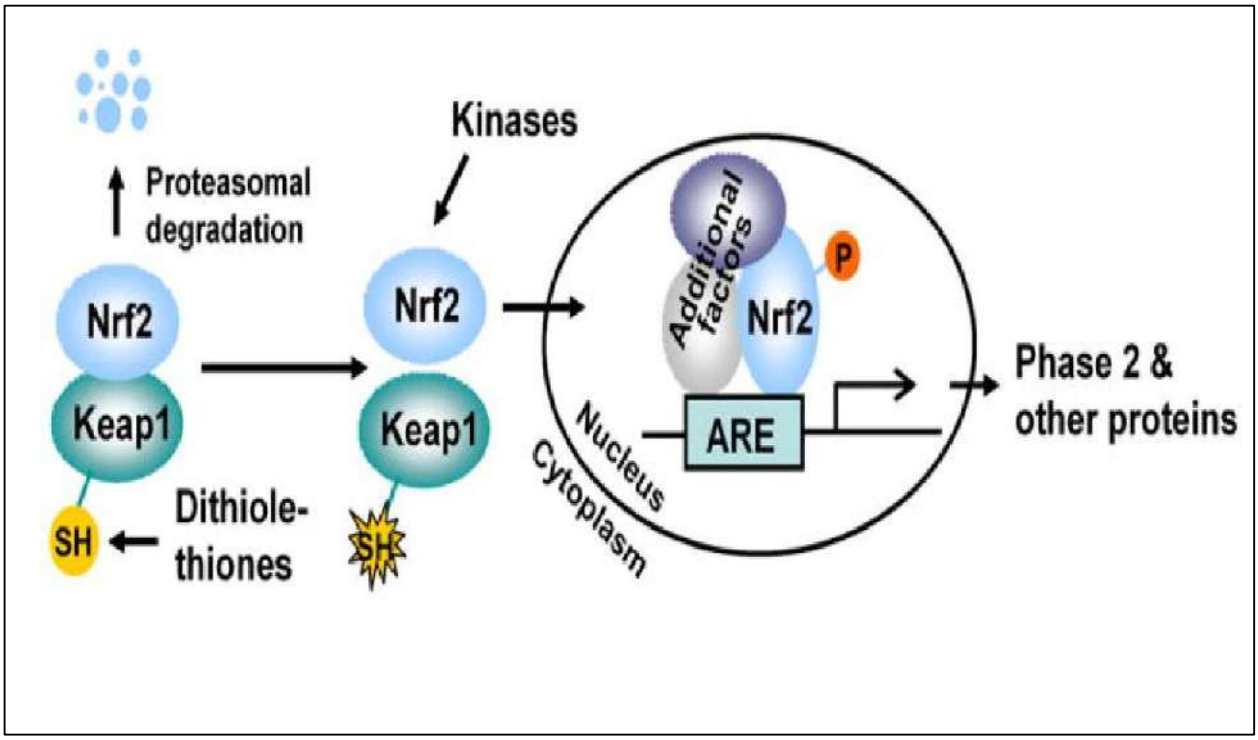


Figure I.10 : Mécanisme enzymatique de l'induction des enzymes de phase II par les dithiolethiones.

II.1. Introduction

La modélisation moléculaire peut être considéré comme un ensemble de techniques informatiques basées sur des méthodes de chimie théorique et les données expérimentales qui peuvent être utilisés soit pour analyser les molécules et les systèmes moléculaires ou à prédire les propriétés moléculaires, chimiques et biochimiques [53]. Il sert de pont entre la théorie et expérience pour:

1. Extrait résultats pour un modèle particulier.

2. Comparer les résultats expérimentaux du système.

3. Comparer les prédictions théoriques du modèle.

4. Aide comprendre et interpréter les observations expérimentales.

5. Corrélation entre détails microscopiques au niveau atomique et moléculaire et les propriétés macroscopiques.

6. Fournir de l'information non disponible à partir d'expériences réelles.

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

En effet les systèmes moléculaires qui sont amènes à être étudies tendent à devenir de plus en plus complexes. Cette complexité est bien sur liée à la taille des molécules envisagées (plusieurs centaines de milliers d'atomes pour les molécules biologiques par exemple) ainsi qu'à la structure intrinsèque des atomes eux-mêmes, mais aussi au degré de précision exige pour le calcul de certaines grandeurs physiques [54].

La modélisation moléculaire consiste en la construction de modèles tridimensionnels a partir des données. Elle trouve sa raison d'être d'une part dans les limitations expérimentales des méthodes de détermination de structure de protéines et d'autre part dans l'incapacité actuelle de prédire la structure 3D a partir de la seule information de séquence. En outre, elle permet d'investiguer les changements de conformations liées à des mutations à partir de structures 3D expérimentales.

La modélisation moléculaire est une application des méthodes théoriques et des méthodes de calcul pour résoudre des problèmes impliquant la structure moléculaire et la réactivité chimique [55].

Ces méthodes peuvent être relativement simples et utilisables rapidement ou au contraire elles peuvent être extrêmement complexes et demander des centaines d'heures de temps d'ordinateur, même sur un super-ordinateur. En plus, ces méthodes utilisent souvent des moyens infographiques très sophistiques qui facilitent grandement la transformation de

quantités impressionnantes de nombres en quelques représentations graphiques facilement interprétables [56].

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

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

II.2. Définition de QSAR :

Une Relation Quantitative Structure-Activité (en anglais : Quantitative structure-activity relationship ou encore QSAR) est le procédé par lequel une structure chimique est corrélée avec un effet bien déterminé comme l'activité biologique ou la réactivité chimique.

Ainsi par exemple l'activité biologique peut être exprimée de manière quantitative, comme pour la concentration de substance nécessaire pour obtenir une certaine réponse biologique [57].

De plus lorsque les propriétés ou structures physio-chimiques sont exprimées par des chiffres, on peut proposer une relation mathématique, ou relation quantitative structure à activité, entre les deux. L'expression mathématique obtenue peut alors être utilisée comme moyen prédictif de la réponse biologique pour des structures similaires.

II.3.Histoire de QSAR :

Les premiers essais de modélisation d'activités de molécules datent de la fin du 19ème siècle, lorsque Crum-Brown et Frazer [58] postulèrent que l'activité biologique d'une molécule est une fonction de sa constitution chimique C (II.1).

$$Ø = f(C) \qquad \text{(II.1)}$$

Richet [59] a découvert que la toxicité des composés organiques suit inversement leur solubilité dans l'eau. Un tel rapport correspond à l'eq II.1, où les $\Delta\phi$ représentent des différences entre les valeurs des activités biologiques, qui sont causées par les changements de ces propriétés chimiques et particulièrement les propriétés physico-chimiques ΔC.

$$\Delta\phi = f(\Delta C) \qquad \text{(II.2)}$$

Aujourd'hui il n'y a aucune méthode qui applique l'eq. II.1 pour traiter les données biologiques. Toutes les équations QSAR correspondent à l'eq. II.2, parce que les différences dans

l'activité biologique sont seulement quantitativement corrélées avec des changements de lipophilie et/ou d'autres propriétés physico-chimiques des composés.

On peut considérer l'année 1964 comme l'année de naissance de la méthodologie QSAR moderne. Deux articles ont été publiés, un par Hansch et Fujita intitulé " method for the correlation of biological activity and chemical structure" [60], l'autre par Free et Wilson portant pour titre "A mathematical contribution to structure activity studies" [61].

Les deux contributions ont commencé par élaborer deux nouvelles méthodes pour quantifier. La relation Activité Biologique / Structure (QSAR) appelée "Hansch analysis" (linear free energy-related approach, extrathermodynamic approach) et Free "Wilson analysis", respectivement.

L'approche QSAR résulte de la combinaison de différents paramètres physico-chimiques de façon linéaire additive (eq. II.3 ; log (1/C) est le logarithme de l'inverse de la dose molaire qui produit ou empêche une certaine réponse biologique, logP est le logarithme du coefficient de partage de n-octanol/ eau). D'autres méthodes utilisent un paramètre connu sous le nom de paramètre de lipophilie calculée π (II.4), il est employé au lieu des valeurs mesurées de logP (comme les valeurs σ de Hammett sont employées au lieu des constantes d'équilibre des réactions organiques), et la formulation d'une équation parabolique pour la description quantitative non-linéaire des rapports lipophilie-activité (II.5). [62 ,63].

$$\log 1/C = a \log P + b\,\sigma + \ldots + \text{const.} \qquad \text{(II.3)}$$

$$\pi x = \log P_{R-X} - \log {}^{\cdot\cdot}P_{R-H} \qquad \text{(II.4)}$$

$$\log 1/C = a (\log P)^2 + b \log P + c\,\sigma + \ldots + \text{const} \qquad \text{(II.5)}$$

D'après la contribution signifiée par Fujita et Ban [64]. Le modèle libre de Wilson est défini par l'eq. II.6, où a_{ij} est la contribution du groupe substituant X_I en position j, $\boldsymbol{\mu}$ est la valeur de l'activité biologique (théorique) d'un composé référence dans la série ; toutes les contributions de groupe a_{ij} des différents substituants X_I se rapportent aux substituants correspondants (le plus souvent l'hydrogène) pour ce composé de référence.

$$\text{Log } 1 / c = \sum \text{aij} + \boldsymbol{\mu} \qquad \text{(II.6)}$$

II.4.Les Modèles QSAR/QSPR :

Au cours des décennies passées, les Relations Quantitatives Structure- Activité/ Propriété (QSAR) sont devenues un puissant outil théorique, alternatif à la mécanique quantique, pour la description et la prédiction des propriétés des systèmes moléculaires complexes dans différents environnements. L'approche QSAR procède de l'hypothèse d'une correspondance univoque entre n'importe quelle propriété physique, affinité chimique, ou activité biologique d'un composé chimique et sa structure moléculaire [65]. Cette dernière peut être représentée par la composition chimique, la connectivité des atomes, la surface d'énergie potentielle, et la fonction d'onde électronique d'un composé. Différents descripteurs moléculaires physico- chimiques reflétant la structure peuvent être déterminés empiriquement où en utilisant des méthodes théoriques et computationnelles de différentes complexités. Il est à souligner que la connaissance de la constitution chimique exacte et/ou de la structure moléculaire tridimensionnelle des composés chimiques étudiés est un pré-requis à l'application de l'approche QSAR.

Le succès de l'approche QSAR dépend de façon critique de la définition précise et de l'utilisation appropriée des descripteurs moléculaires.

II.5. Les applications de l'étude QSAR :

Certaines études QSAR semblent être un peu plus d'études universitaires, il existe un grand nombre d'applications de ces modèles tels que [66] :

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

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

II.6.Les descripteurs moléculaires :

De nombreuses recherches ont été menées, au cours des dernières décennies, pour trouver la meilleure façon de représenter l'information contenue dans la structure des molécules, et ces structures elles-mêmes, en un ensemble de nombres réels appelés descripteurs ; une fois que ces nombres sont disponibles, il est possible d'établir une relation entre ceux-ci et une propriété ou activité moléculaire, à l'aide d'outils de modélisation classiques.

II.6.1.Les types des descripteurs moléculaires :

Nous allons présenter les descripteurs moléculaires les plus courants, en commençant par les descripteurs les plus simples, qui nécessitent peu de connaissances sur la structure moléculaire, mais véhiculent peu d'informations. Nous verrons ensuite comment les progrès de la modélisation moléculaire ont permis d'accéder à la structure 3D de la molécule, et de calculer des descripteurs à partir de cette structure.

II.6.1.1.Les descripteurs 1D :

Sont accessibles à partir de la formule brute de la molécule (par exemple C_6H_6O pour le phénol), et décrivent des propriétés globales du composé. Il s'agit par exemple de sa composition, c'est-à-dire les atomes qui le constituent, ou de sa masse molaire. On peut remarquer que ces descripteurs ne permettent pas de distinguer les isomères de constitution.

II.6.1.2.Les descripteurs 2D :

Sont calculés à partir de la formule développée de la molécule. Ils peuvent être de plusieurs types.

- Les ***indices constitutionnels*** caractérisent les différents composants de la molécule.

Il s'agit par exemple du nombre de liaisons simples ou multiples, du nombre de cycles.

- Les ***indices topologiques*** peuvent être obtenus à partir de la structure 2D de la molécule, et donnent des informations sur sa taille, sa forme globale et ses ramifications. Les plus fréquemment utilisés sont l'indice de Wiener [67], l'indice de Randić [68], l'indice de connectivité de valence de Kier-Hall [69] et l'indice de Balaban [70]. L'indice de Wiener permet de caractériser le volume moléculaire et la ramification d'une molécule : si l'on appelle distance topologique entre deux atomes le plus petit nombre de liaisons séparant ces deux atomes, l'indice de Wiener est égal à la somme de toutes les distances topologiques entre les différentes paires d'atomes de la molécule. L'indice de Randić est un des descripteurs les plus utilisés ; il peut être interprété comme une mesure de l'aire de la molécule accessible au

solvant. Ces descripteurs 2D reflètent bien les propriétés physiques dans la plupart des cas, mais sont insuffisants pour expliquer de façon satisfaisante certaines propriétés ou activités, telles que les activités biologiques. Des descripteurs, accessibles à partir de la structure 3D des molécules, ont pu être calculés grâce au développement des techniques instrumentales et de nouvelles méthodes théoriques.

II.6.1.3.Les descripteurs 3D :

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

- **Les descripteurs géométriques** les plus importants sont le volume moléculaire, la surface accessible au solvant, le moment principal d'inertie.
- **Les descripteurs électroniques** permettent de quantifier différents types d'interactions inter- et intramoléculaires, de grande influence sur l'activité biologique de molécules. Le calcul de la plupart de ces descripteurs nécessite la recherche de la géométrie pour laquelle l'énergie stérique est minimale, et fait souvent appel à la chimie quantique. Par exemple, les énergies de la plus haute orbitale moléculaire occupée et de la plus basse vacante sont des descripteurs fréquemment sélectionnés.

Le moment dipolaire, le potentiel d'ionisation, et différentes énergies relatives à la molécule sont d'autres paramètres importants.

- ***Descripteurs spectroscopiques*** : les molécules peuvent être caractérisées par des mesures spectroscopiques, par exemples par leurs fonctions d'onde vibrationnelles.

En effet, les vibrations d'une molécule dépendent de la masse des atomes et des forces d'interaction entre ceux-ci ; ces vibrations fournissent donc des informations sur la structure de la molécule et sur sa conformation. Les spectres infrarouges peuvent être obtenus soit de manière expérimentale, soit par calcul théorique, après recherche de la géométrie optimale de la molécule. Ces spectres sont alors codés en vecteurs de descripteurs de taille fixe. Le descripteur EVA [71] est ainsi obtenu à partir des fréquences de vibration de chaque molécule. Les descripteurs de type *MoRSE* [72] *(Molecule Representation of Structures based on Electron diffraction)* sont calculés à partir d'une simulation du spectre infrarouge ; ils font appel au calcul des intensités théoriques de diffraction d'électrons.

Par ailleurs, le calcul de certains descripteurs demande une étape préliminaire d'alignement des molécules, pour les placer dans une orientation commune.

II.6.2.Réduction du nombre de variables :

Comme nous l'avons rappelé, un grand nombre de descripteurs différents sont collectés pour la modélisation d'une grandeur donnée, car les facteurs déterminants du processus étudié ne sont a priori pas connus. Cependant, les descripteurs envisagés n'ont pas tous une influence significative sur la grandeur modélisée, et les variables ne sont pas toujours mutuellement indépendantes. De plus, le nombre de descripteurs, c'est-à-dire la dimension du vecteur d'entrée, détermine la dimension du vecteur des paramètres à ajuster. Si cette dimension est trop importante par rapport au nombre d'exemples de la base d'apprentissage, le modèle risque d'être surajusté à ces exemples, et incapable de prédire la grandeur modélisée sur de nouvelles observations.

Il est donc nécessaire de réduire la dimensionnalité des variables d'entrée. Plusieurs approches sont possibles pour résoudre ce problème :

- réduire la dimension de l'espace des entrées ;
- remplacer les variables corrélées par de nouvelles variables synthétiques, obtenues à partir de leurs combinaisons ;
- sélectionner les variables les plus pertinentes.

Nous allons maintenant décrire les méthodes les plus fréquemment utilisées.

II.6.3.Sélection des variables pertinentes :

L'analyse en composantes principales a pour but de réduire les corrélations entre les variables, mais cette étape de réduction est indépendante de la grandeur modélisée. Or, les variables calculées n'ont pas nécessairement une influence sur cette grandeur. Il est nécessaire d'éliminer celles dont l'influence est inférieure à celle du bruit, et de sélectionner uniquement les plus pertinentes d'entre elles. Il est possible, à partir de p descripteurs, de former $2p$ sous-ensembles de variables, donc $2p$ modèles, dont il faudrait comparer les performances pour déterminer le meilleur jeu de descripteurs. Cette méthode permet d'envisager toutes les combinaisons possibles de descripteurs, mais sa mise en œuvre est très lourde. Des approches alternatives ont donc été développées.

La ***sélection progressive*** consiste à incorporer les variables au modèle une à une, en sélectionnant, à chaque étape, la variable dont la corrélation partielle avec la grandeur modélisée est la plus élevée. À l'inverse, lors de l'***élimination progressive***, on débute la

modélisation avec l'ensemble des descripteurs, en les éliminant un par un jusqu'à obtenir le meilleur jeu de composantes.

La ***sélection pas à pas*** est une combinaison des deux méthodes évoquées précédemment. Les variables sont incorporées une à une dans le modèle, par sélection progressive. Cependant, à chaque étape, on vérifie que les corrélations partielles des variables précédemment introduites sont encore significatives.

Une méthode plus efficace est celle du ***descripteur sonde*** [73, 74]. Cette sélection est généralement effectuée à partir d'un modèle linéaire en ses paramètres, par exemple un modèle polynomial. En effet, la pertinence d'une variable est indépendante du modèle

II.7.Méthodes utilisées pour le développement de modèles QSAR/QSPR :

II.7.1.Introduction :

L'application pratique des gammes des descripteurs moléculaires dans le développement de modèles QSAR/QSPR n'est pas une tâche aisée [75]. Tout d'abord, un très grand nombre (>3000) de descripteurs moléculaires, de différentes complexités et de conceptions diverses ont été imaginés et proposés au cours des (60 dernières) années. Ensuite, pendant ce temps, aucune règle stricte n'a été établie, ni même proposée, pour la sélection de descripteurs adaptés parmi la myriade de descripteurs disponibles. Ce choix a souvent été basé sur l'intuition chimique des chercheurs, ou en se pliant à la tradition.

Une autre difficulté dans la sélection des descripteurs QSAR découle de la non standardisation des gammes de descripteurs. Les gammes empiriques des constantes d'induction, de résonance et d'effet stérique des constituants, ou les échelles empiriques d'effets de solvant comportent des erreurs intrinsèques liées aux erreurs respectives des mesures expérimentales. Par ailleurs, les méthodes quanto- mécaniques appliquées aux calculs des descripteurs moléculaires et aux distributions de charges liés aux OM sont souvent basées sur différents paramètres semi- empiriques, ou l'utilisation de différents ensembles de base dans les calculs ab- initio. Naturellement, un descripteur construit à l'aide de différentes méthodes expérimentales ou théoriques, pour divers composés, ne peut être utilisé pour le calcul d'un modèle QSAR unique. Une approche systématique pour la sélection de gammes de descripteurs pour le calcul de modèles QSAR est basée sur la discrimination statistique entre de larges ensembles de descripteurs.

Dans ce qui suit nous passerons en revue diverses approches utilisées pour le développement des " meilleures" équations QSPR dans de grands espaces de descripteurs.

En dernier ressort, les modèles QSAR peuvent être développés selon des modèles mathématiques différents, généralement en relation avec l'analyse statistique multi variée.

Le premier modèle, et le plus largement utilisé, consiste en une équation (multi) linéaire obtenue par régression des données expérimentales en fonction d'un ensemble de descripteurs pré- sélectionnés (ou d'un seul), en utilisant la méthode des moindres carrés ordinaires (MCO).

Dans quelques cas, les modèles physiques ou chimiques connus du phénomène étudié laissent prévoir certaines formes mathématiques non linéaires (exponentielles ou logarithmiques) de la dépendance entre les données expérimentales et les descripteurs moléculaires.

Les modèles QSAR peuvent alors être établis à l'aide de la technique de régression par les moindres carrés non linéaires. D'autres modèles ont été développés en utilisant l'analyse factorielle ou l'analyse en composantes principales. L'intérêt de ces méthodes est qu'elles évacuent le problème de multi colinéarité inhérent aux méthodes de régression linéaires. Cependant, l'interprétation des équations QSAR est alors entravée par la nature formelle des facteurs ou des composantes principales. Une alternative aux méthodes très classiques de régression linéaire multiple (RLM) et d'analyse en composantes principales (ACP) est la technique de régression par les moindres carrés partiels (MCP ou PLS) [76, 77].

II.7.2.Méthodes de régressions linéaire et multilinéaire :

II.7.2.1.Aperçu général :

Comme signalé auparavant, l'investigateur choisit dans chaque cas un ou plusieurs descripteurs supposé(s) refléter les interactions physiques ou chimiques à la base de la propriété moléculaire ou de la caractéristique du phénomène étudié. Ce choix, encore une fois, est habituellement fondé sur l'intuition chimique, la tradition, ou simplement la disponibilité du descripteur. Néanmoins, cinq principes peuvent aider à la sélection de descripteurs moléculaires convenables pour l'établissement de modèles QSAR. Ce sont:

a) Un nombre maximal de données expérimentales (de préférence toutes) doivent être caractérisées par des valeurs de descripteurs originaux complémentaires.

b) Les valeurs des descripteurs doivent être obtenues de la même source et, de préférence, mesurées selon le même protocole expérimental ou calculées en utilisant le même logiciel.

c) Le nombre de descripteurs dans les modèles de régression multiples doit être minimisé, sans perte d'information, ce que mettent en évidence les critères statistiques (valeurs des tests t et F…).

d) Dans les modèles RLM, les descripteurs utilisés doivent être statistiquement orthogonaux.

e) Pourvu que les autres critères soient similaires, la nature physique ou chimique du descripteur sélectionné doit être la plus proche de la propriété ou du phénomène étudié.

En réalité, il est difficile de se conformer pratiquement aux 5 principes énoncés. Cependant, la négligence de plusieurs d'entre eux peut conduire à des équations inutiles sans aucun pouvoir prédictif sinon très limité.

II.7.2.2.Evaluation préliminaire des données :

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

On distingue habituellement dans un tel pré- traitement des données des analyses uni variées et des analyses bi variées [76, 78].

Dans l'analyse uni variée, il est recommandé de vérifier la conformité des données à la distribution normale. Une précaution particulière doit être prise lors de la procédure de régression subséquente si les valeurs de la propriété étudiée, ou d'un descripteur, ne suivent pas la loi de Laplace- Gauss.

Pour un ensemble de descripteurs différents, il est nécessaire d'effectuer une analyse des données bivariée, c'est-à-dire de calculer le coefficient de corrélation linéaire R entre chacune des paires de l'ensemble des descripteurs. Si R est statistiquement significatif (R>0,95), ces deux descripteurs ne peuvent être utilisés simultanément lors de l'analyse par RLM.

II.7.2.3. Modèles statistiques :

L'objectif de l'analyse statistique est justement de « démêler » ces descripteurs et d'identifier ceux qui sont corrélés à la variable cible, qui produisent du signal, de ceux qui ne le sont pas, qui produisent du bruit. L'analyse statistique permet également d'identifier les descripteurs qui sont corrélés entre eux pour ne garder que les principaux et réduire ainsi la redondance d'informations [79].

Dans la terminologie des statistiques, la propriété étudiée est appelée variable cible ou variable dépendante ou encore variable Y. C'est la caractéristique que l'on cherche à expliquer et à optimiser.

Les descripteurs, en l'occurrence chimiques, sont appelés variables explicatives, variables indépendantes ou encore variables X.

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

Les principaux outils statistiques pour obtenir un modèle sont :

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

II.8. Régression linéaire multiple :

Une analyse de régression où la variable dépendante *Y* dépend linéairement de plusieurs variables indépendantes *X*1, *X*2, . . ., *Xj* est appelée régression linéaire multiple.

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

$Y = f(X_1, X_2, \ldots, X_j)$ où $f(X_1, X_2. \ \ldots X_j)$ est une fonction linéaire de X_1, X_2, X_j [80].

Le modèle mathématique peut se présenter sous la forme :

$$y^i = a + \sum_{j=1}^{j=p} b_j \ x_j^i \qquad \text{(II.7)}$$

Où b_j sont les coefficients de régression et (a) la constante de régression.

II.8.1.Description de la méthode :

La régression multilinéaire est la méthode la plus simple et la plus employée pour le développement de modèles prédictifs [81]. Elle repose sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante Y (ici, la propriété) et une série de p variables indépendantes Xi (ici, les descripteurs). L'objectif est d'obtenir une équation de la forme suivante :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_p X_p + \varepsilon \qquad \text{(II.8)}$$

Où βi sont les coefficients de la régression.

La détermination de l'équation (II.8) se fait alors à partir d'une base de données de n échantillons pour laquelle a la fois les variables dépendantes et la variable indépendante sont connues. Il s'agit donc de considérer un système de p équations.

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

. . . . . .

. . . . . .

. . . . . .

. . . . . .

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

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

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

Le modèle s'écrit alors sous forme $Y = X\,b + \varepsilon$.

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

En premier lieu, les variables indépendantes Xi comme leur nom l'indique, sont supposées indépendantes entre elles et leur incertitude est négligeable. Ensuite, les différents échantillons Yi sont supposes indépendants entre eux. Enfin, par nature, la dépendance de Y vis-à-vis des Xi est supposée linéaire.

La valeur prédite de la variable dépendante est alors :

$$\hat{Y} = \beta_0 + \hat{B}_1 X_{1,j} + \ldots + \hat{B}_p X_{n,j} \quad \text{(II.9)}$$

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

$$\varepsilon = Y_i - \hat{Y}_i \quad \text{(II.10)}$$

Il s'agit alors de trouver les coefficients afin de minimiser la somme des carrés de ces résidus (*RSS*) pour l'intégralité de la base de données.

II.8.2.Confiance accordée aux résultats :

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

- ✓ tester la signification globale de la régression,
- ✓ tester l'hypothèse nulle β = 01 et a calculer l'intervalle de confiance de β ,1
- ✓ tester l'hypothèse nulle β = 00 et a calculer l'intervalle de confiance de β ,0
- ✓ calculer la précision de l'estimation de Y pour la moyenne et pour une observation individuelle.

II.8.3.Test de la signification globale de la régression :

a. Coefficient de détermination (R^2) :

Le coefficient de détermination R^2 est la mesure de degré de liaison entre Y_n et X_j. R^2 déterminé par différentes relations suivantes :

$$R^2 = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} \quad \text{(II.11)}$$

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

b. Coefficient de corrélation (R) :

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

c. Test Fischer-Snedecor (F) :

Le test Fischer permet de justifier la liaison globale entre Yn et X,j une version dérivée de cet indicateur peut juger du degré de pertinence des variables du modèle. Il s'agit de vérifier pour chaque variable *Xj* si, lorsqu'on passe du modèle complet à *p* prédicateurs au modèle simplifié obtenu en *t* mesure en fait l'apport marginal de variable *Xj* à l'explication de *Yn*.

On peut définir quelques paramètres utilisés dans la régression multilinéaire.

Somme des carrés totaux : $TSS = \sum (Y_{obs} - \bar{y})^2$ (II.12)

Somme des carrés expliqués : $ESS = \sum (Y_{cal} - \bar{y})^2$ (II.13)

Somme des carrée résiduels : $RSS = \sum (Y_{obs} - Y_{cal})^2$ (II.14)

Ainsi, $TSS = ESS + RSS$

F peut s'écrire comme suit :

$$F = \frac{ESS}{p} \frac{n-p-1}{RSS} \quad \text{(II.15)}$$

La forme de l'équation (II.15) représente le nombre de degrés de liberté associé avec chaque paramètre. Le *ESS* associé avec *p* degrés de liberté et le *RSS* associé n-p-1 degrés de liberté.

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

Intuitivement, nous rejetterons l'hypothèse nulle lorsque la somme des carrés expliquée par la régression est grande. En d'autres termes, la région critique de ce test est de la forme (F > seuil). Si la quantité F observée dépasse le seuil, on rejette l'hypothèse H0 dans le cas contraire, on conserve H_0.

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

Le p-value est le niveau de significativité du test de Fischer-Snedecor, c'est-a-dire la probabilité de dépasser le F observé si l'hypothèse nulle est vraie. On compare la p-value au risque α choisi (par exemple α=0.05).

Si p-value ≤ α, alors on rejette l'hypothèse nulle $\beta_1 = ... = \beta_p = 0$.

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

Tableau II.1 : Table d'analyse de variance

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

***c.* Ecart type (s) :**

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

$$S = \sqrt{\frac{RSS}{n-p-1}} \qquad \text{(II.16)}$$

Dont *p* est le nombre de variables indépendants.

d- Coefficient de prédiction (Q^2)

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

$$Q^2 = 1 - \frac{PRESS}{SSY} \qquad \text{(II.17)}$$

II.8.4.Validations des modèles :

Quel que soit le but ultime du modèle QSAR, il doit être obligatoirement validé avant d'être interprété ou utilisé à fins prédictives. Il existe différentes solutions pour s'assurer de validité d'un modèle. La validation croisée *LOO* (*leave-one-out*) est un processus qui permet de tester la précision prédictive d'un modèle, cette méthode basée sur le calcul de quelques paramètres statistiques tels que :

la somme des carrée résiduels RSS (PRESS), Somme des carrés totaux TSS, qualité d'ajustement R^2_{adj}, coefficient de corrélation de validation croisée R^2_{CV}, la validation standard des erreurs de prédiction (SPRESS) et l'erreur de prédiction (PE). Ces paramètres statistiques sont calculés à partir des relations suivantes :

$$PRESS = \sum (y_{obs} - y_{cal})^2 \quad \text{(II.18)}$$

$$TSS = \sum (Y_{obs} - \bar{y})^2 \quad \text{(II.19)}$$

$$R^2_{ajs} = (1 - r^2)\left(\frac{n-1}{n-p-1}\right) \quad \text{(II.20)}$$

$$R^2_{CV} = 1 - \frac{PRESS}{TSS} \quad \text{(II.21)}$$

$$S_{PRESS} = \sqrt{\frac{PRESS}{n}} \quad \text{(II.22)}$$

$$PE = 0.6745\ (1 - r2)/\sqrt{n} \quad \text{(II.23)}$$

Avec :

PRESS : somme des carrées des écarts entre les valeurs observées et les valeurs prévues à partir du modèle.

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

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

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

PE : L'erreur de prédiction du coefficient de corrélation est utilisée pour déterminer le pouvoir prédictif des modèles proposés.

III. Etude quantitative des relations structure-activité (QSAR) des dérivés de 1,2-dithiole-3-thione :

III.1.Introduction :

Le principe des méthodes QSAR (Quantitative Structure Activity Relationship) consiste comme leurs noms l'indiquent à mettre en place une relation mathématique à l'aide des méthodes d'analyse de données, reliant des propriétés moléculaires microscopiques appelées descripteurs, à un effet expérimental (activité biologique, toxicité, affinité pour un récepteur), pour une série de composés chimiques similaires. Le point de départ de telles méthodes se construit sur la définition des descripteurs moléculaires empiriques ou théoriques. Ces dernières prennent en compte des informations sur la structure et les caractéristiques physico-chimiques des molécules.

Le choix de la base des données expérimentales de référence est décisif dans une étude QSAR. Elle doit être composée de données expérimentales fiables obtenues en suivant un protocole expérimental unique. En effet, la robustesse du modèle dépend fortement de la base sur lequel il se fonde.

Enfin, le lien entre les descripteurs et la base de données est déterminé grâce à des outils d'analyse comme les régressions multi-linéaires (MLR), les régressions aux moindres carrés partiels (PLS), les arbres de décisions, les réseaux de neurones, et les algorithmes génétiques.

Les informations extraites à partir des résultats d'étude de QSAR peuvent être utilisées pour obtenir une meilleure connaissance des structures moléculaires et probablement le mode d'action au niveau moléculaire. Ces informations peuvent alors être utilisées pour prévoir les propriétés physicochimiques et les activités biologiques de nouveaux composés ainsi que pour concevoir de nouvelles structures (figure III.1) [82].

D3T (3*H*-1,2-dithiole-3-thione) et ses dérivés ont attiré une attention considérable depuis plusieurs années en raison de leurs activités biologiques importantes. Par exemple, D3T est connu par son activité de la détoxication des agents carcinogènes de l'environnement, la prévention contre les néoplasies et l'excitation des autres effets protecteurs par l'induction des enzymes antioxydantes de la phase II [83]. Plusieurs dérivés de D3T possèdent une propriété de l'induction des enzymes de la phase II, parmi ces dérivés on trouve Olipraz (4-méthyl-5-pyrazinyl-3*H*-1 ,2-dithiole-3-thione) [84]. Cette molécule possède une activité inhibitrice remarquable des virus VIH-1 (SIDA) [85].

D'autres dérivés ont des propriétés fongitoxiques et bactériostatiques [86]. Burgot et son équipe ont rapporté une relation entre les propriétés de détoxication d'une série de dérivés de 1,2-dithiole-3-thione et le paramètre de lipophilie log P [87].

Suite à notre intérêt dans ce domaine, notre présente étude visait à développer des modèles QSAR pour ces composés par rapport à leur activité spécifique de quinone réductase.

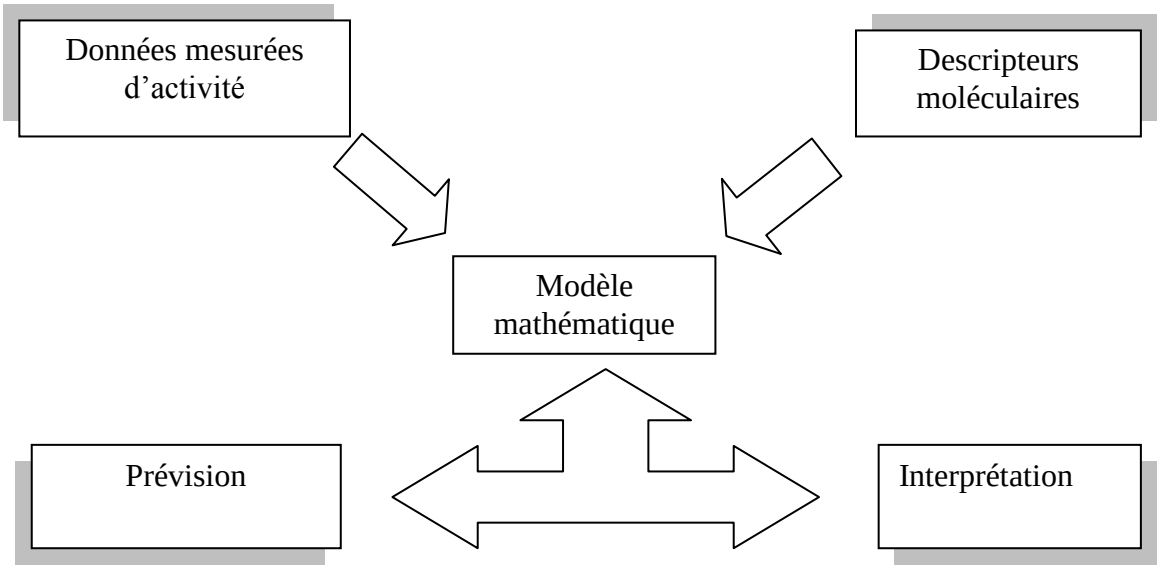


Figure III.1: Modèle d'étude quantitative des relations structure/activité (QSAR).

III.2. Données expérimentales :

Dans notre travail nous avons utilisé le paramètre biologique: l'activité spécifique de la quinone réductase (QR).

Les études ***QSAR*** ont été réalisées sur une sérié des 16 dérivés de 1,2-dithiole-3-thione qui sont utilisées comme des agents chimio-protecteurs. Le tableau (III.1) montre les structures chimiques de ces dérivés.

En effet, les dérivés ont une propriété d'induction sur les enzymes détoxication, en l'occurrence quinone réductase ***(QR)*** dont l'étude précitée est basée sur l'activité biologique de cet enzyme ainsi que les descripteurs moléculaires.

L'activité spécifique de ces dérivés est décrire par logarithme négative de la dose efficace médiane mesuré expérimentalement (tableau III.2).

Les résultats biologiques de ces 16 dérivés de 1,2-dithiole-3-thione sont exploités par la méthode de régression linéaire multiple (*MLR = multiple linear regression*) dans le but d'établir les modèles QSAR.

Tableau III.1 : Structures chimiques des dérives de 1,2-dithiole-3-thione.

Molécules	Dithiolethione	Structure
1	5-methyl-3H-1,2-dithiole-3-thione	
2	5-ethyl-3H-1,2-dithiole-3-thione	
3	5,6-dihydrocyclopenta[c][1,2]dithiole-3(4H)-thione	
4	4-phenyl-3H-1,2-dithiole-3-thione	
5	4-methyl-5-(methylthio)-3H-1,2-dithiole-3-thione	
6	4-methyl-3H-1,2-dithiole-3-thione	
7	4-ethyl-3H-1,2-dithiole-3-thione	
8	3-thioxo-3H-1,2-dithiole-5-carboxamide	

9	3-thioxo-3H-1,2-dithiole-4-carboxamide	
10	3H-1,2-dithiole-3-thione	
11	(E)-5-(2-(dimethylamino)vinyl)-4-methyl-3H-1,2-dithiole-3-thione	
12	5-(tert-butyl)-3H-1,2-dithiole-3-thione	
13	5-(dimethylamino)-3H-1,2,4-dithiazole-3-thione	
14	4-methyl-5-(pyrazin-2-yl)-3H-1,2-dithiole-3-thione	
15	4,5,6,7-tetrahydro-3H-benzo[c][1,2]dithiole-3-thione	
16	5-phenyl-3H-1,2-dithiole-3-thione	

Tableau III.2 : L'activité biologique des dérivés de 1-2, dithiole-3-thione.

Molécules	Activité biologique(QR)
1	4.2
2	2.7
3	3.5
4	6.1
5	20.8
6	10.2
7	5.9
8	3.7
9	8.4
10	10.8
11	5.1
12	1.2
13	7.3
14	7.8
15	9.5
16	5.4

III.3.Préparation de base de données :

Afin de procéder à la préparation de notre base de données, nous avons suivie les étapes suivantes :

- Sélectionner seize dérivés des molécules de 1,2-dithiole-3-thione (ces molécules ont été téléchargées à partir de la base de données chemspider).
- Optimiser leurs structures chimiques par ChemDraw pour donner la forme la plus stable.

- Transformer, a l'aide d'un logiciel Openbabel les structures chimiques du format .mol au format SDF
- Organiser les données dans un fichier Excel.
- Calculer les descripteurs moléculaires par le logiciel Padel (le logiciel utilisé a donné un nombre important de ces descripteurs (> 400)).

III.4. Sélection des variables :

Un descripteur moléculaire peut être considéré comme la conséquence d'un processus logique et mathématique, appliqué à l'information chimique codifiée à travers la représentation d'une molécule [88].

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

L'information codée d'un descripteur moléculaire dépend du type de représentation moléculaire employée et de l'algorithme défini pour son calcul.

Le nombre relativement important de caractéristiques ou de descripteurs moléculaires ainsi calculé pour le codage peut manquer de pertinence sur l'interprétation structurale. En plus, Il existe un problème de choix de variable. Ces contraintes se résoudre par des méthodes d'apprentissage, appliquées comme étapes de prétraitement ou de post-traitement :

1er Etape :

Dans cette étape nous avons éliminé tous les descripteurs qui :

- Possèdent des valeurs nulles.
- Possèdent des valeurs identiques.
- Possèdent plus de trois valeurs nulles.

Après cette étape, nous avons pu obtenir environ de 368 descripteurs.

2eme Etape :

L'analyse statistique ***RML*** détermine et quantifie les corrélations entre les descripteurs et la variable cible. Elle indique également la contribution relative de chaque descripteur dans l'explication globale de l'activité.

La méthode ***RML*** est utilisée pour modéliser la relation linéaire entre une variable dépendante ***Y*** est les variables indépendantes ***X*.** Le modèle QSAR obtenue est ajusté de telle sorte que l'écart entre les valeurs observées et prédites est minimisé. L'étude statistique par la méthode ***RML*** est réalisée à l'aide logiciel statistique *SPSS 19* [89].

Dans notre travail, nous avons basé sur deux méthodes dont le but de minimiser au maximum le nombre des descripteurs.

Par la première méthode « **Entré »**, nous avons obtenu une matrice de corrélation. Dans cette dernière, les descripteurs qui ont une corrélation entre eux supérieur à 0.8 ($r > 0.8$) ont été éliminé.

La deuxième méthode appelée « **pas à pas »** ; les variables sont incorporées une à une dans le modèle, par une sélection progressive.

Par cette dernière, les variables qui présentent une forte corrélation entre eux ont été éliminés automatiquement. Alors on obtient un modèle avec deux descripteurs.

III.5.Les caractéristiques du modèle obtenu :

Dans la première étape de notre étude, les différents dithiolethiones substitués (**Tableau III.1**) ont été évalués pour leur activité de l'induction des enzymes de la phase II.

Un paramètre biologique Quinone Réductase (QR) a été introduit dans cette étude et les résultats sont illustrés dans le **Tableau III.2**. Afin de déterminer le rôle des caractéristiques structurelles des descripteurs, une étude QSAR a été réalisée.

Une série de 16 dérivés de 1,2-dithiole-3- thione ont été utilisés pour la production du modèle de régression multilinéaire. Différents descripteurs physico-chimiques ont été utilisés comme variables indépendantes et ont été corrélées avec l'activité biologique.

La matrice de corrélation de Pearson a été effectuée sur tous les descripteurs en utilisant le logiciel SPSS. L'analyse de la matrice révélé deux descripteurs pour le développement du modèle QSAR. Les valeurs des descripteurs sélectionnés pour le modèle QSAR sont présentées dans le **Tableau III.3**.

Tableau III.3 : Les 2 descripteurs retenus pour construire les modèles de QSAR.

Composé	AATS3p	ATSC3m
1	2,6387	-1024,1991
2	2,0667	7,3581
3	2,0018	875,0090
4	2,1874	632,3705
5	3,3438	706,0463
6	2,7589	150,8820
7	2,0862	869,1458

8	2,4868	-415,8605
9	2,5995	-14,3284
10	2,8182	-479,4217
11	2,1279	269,3333
12	1,9698	-355,4679
13	2,2501	441,8233
14	2,4534	504,5492
15	1,8332	1079,8093
16	2,1863	200,6620

La matrice de corrélation entre les descripteurs obtenus par l'analyse ***MLR*** et l'activité biologique est reportée dans le **Tableau III.4**.

Tableau III.4 : Matrice de corrélation entre les descripteurs obtenus

Variables	**AATS3p**	**ATSC3m**	**QR**
AATS3p	**1,0000**	-0,2946	0,7510
ATSC3m	-0,2946	**1,0000**	0,3106
QR	0,7510	0,3106	**1,0000**

La corrélation entre l'activité biologique (QR) et les descripteurs est exprimés par la relation suivante :

$$QR = -19.25666+10.70580*AATS3p+0.00462*ATSC3m \quad (20)$$

Avec : n= 16**; r** =0.9347**; r²**= 0,8737

La valeur de la fraction de la variance peut varier entre 0 et 1. Le modèle QSAR doit considérer un $R^2 > 0.6$ pour qu'il soit valide [90]. Les valeurs R = 0.9347 (R^2 = 0,8737) nous a permis d'indiquer fermement la corrélation entre les différents descripteurs utilisés et l'activité biologique QR.

Le calcul des valeurs prédites (log $(1/CDQR)_{pred}$) et résiduelles (log $(1/CDQR)_{resd.}$) de l'activité biologique des dithiolethiones est un autre moyen pour estimer le pouvoir prédictif des modèles développés (**Tableau III.5**).

Tableau III.5 : Valeurs des activités prédites et résiduelles de l'ensemble de dérives de 1,2-dithiole-3-thione.

Composés	QR	Préd(QR)	Résidu
1	4,2000	4,2586	-0,0586
2	2,7000	2,9031	-0,2031
3	3,5000	6,2185	-2,7185
4	6,1000	7,0838	-0,9838
5	20,8000	19,8044	0,9956
6	10,2000	10,9765	-0,7765
7	5,9000	7,0951	-1,1951
8	3,7000	5,4441	-1,7441
9	8,4000	8,5069	-0,1069
10	10,8000	8,6991	2,1009
11	5,1000	4,7688	0,3312
12	1,2000	0,1885	1,0115
13	7,3000	6,8748	0,4252
14	7,8000	9,3411	-1,5411
15	9,5000	5,3595	4,1405
16	5,4000	5,0772	0,3228

La courbe de régression linéaire (**Figure III.2**) montre l'existence d'une relation mathématique entre les valeurs prédites et leurs valeurs expérimentales de l'activité biologique de ***QR.*** La courbe des modèles ***QR*** montre que les points du courbe de corrélation

de l'activité biologique sont tous proche de la droite idéale, ce qui traduit l'importance des descripteurs choisis à la prédiction d'une nouvelle activité biologique.

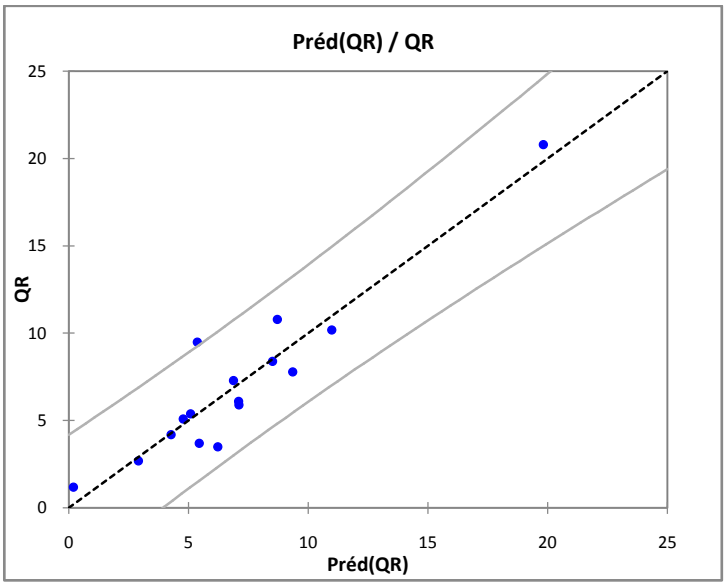


Figure III.2 : Corrélation entre l'activité biologique expérimentales et prédites

Pour vérifier l'existence d'une erreur systématique dans les modèles ***QSAR***, les résidus des valeurs prédites de l'activité biologique ***pCQR*** ont été tracés en fonction des valeurs expérimentales, comme le montre la **figure III.3**.

La propagation des résidus des deux côtés de zéro indique qu'il n'y a aucune erreur systémique, ce qui indique que ces modèles peuvent être appliqués avec succès pour cette série des molécules afin de prédire l'activité biologique.

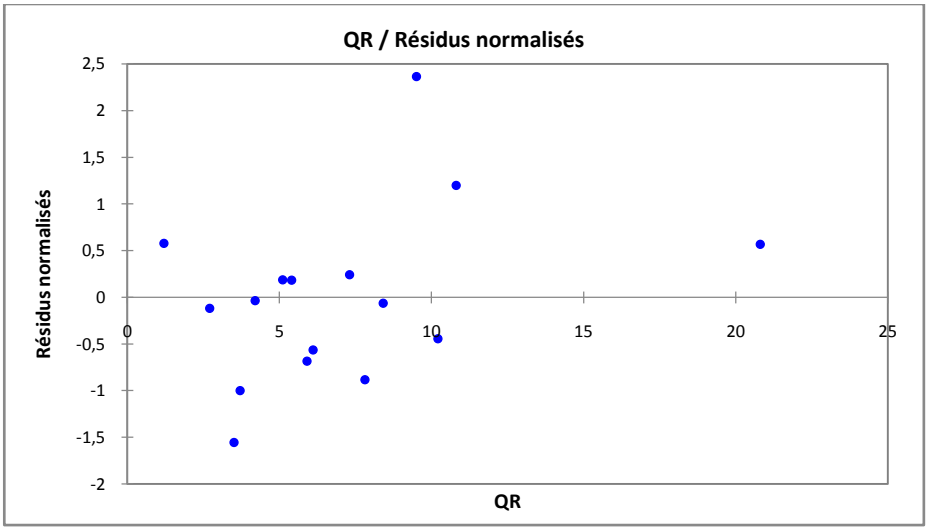


Figure III.3: Représentation graphique des valeurs résiduelles en fonction des valeurs expérimentales de pCQR .

Du point de vue des performances statistiques, ce modèle présente un coefficient de corrélation R^2 significatif. La valeur de R^2 est égale à 0,8737 ; indique que ce modèle peut être appliqué avec succès pour prédire l'activité spécifique de la quinone réductase pour les dithiolethiones étudiés.

Afin de tester la validité du modèle QSAR construit, la valeur élevée de R^2_{adj} est utilisée comme critère de validation [91].

Les résultats obtenus de (R^2_{adj}) pour le modèle QSAR est R^2_{adj}= 0,8543.

(R^2_{adj} >0.6) signifie une bonne qualité du modèle construit.

III.6. Quantification des descripteurs :

L'analyse statistique ***RML*** détermine et quantifie les corrélations entre les descripteurs et la variable cible. Elle indique également la contribution relative de chaque descripteur dans l'explication globale de l'activité, dans l'équation (20). .

Les descripteurs présentés dans l'équation ***(20)***; qui jouent un rôle important sur les dérivés de 1,2-dithiole-3-thiones utilisés dans l'étude ***QSAR*** pour donner une bonne activité de l'induction des enzymes de détoxification de la phase II (***QR).***

Conclusion générale

Les dérivés de 1,2-dithiole-3-thiones utilisés dans l'étude QSAR sont caractérisés par son activité de l'induction des enzymes de détoxification de la phase II.

La modélisation moléculaire et l'analyse QSAR ont été effectuées pour déterminer quantitativement l'effet de la structure moléculaire des composés étudiés sur leur activité biologique. Un modèle mathématique a été développé pour prédire l'activité spécifique de la quinone réductase pour certains dérivés de 1,2-dithiole-3-thione.

La validité de modèle obtenu à été établie par la détermination des paramètres statistiques appropriés. Le modèle mis en place a été utilisé pour prédire l'activité biologique des dithiolethiones étudiés. Les valeurs prédites sont plus proches avec celles des valeurs expérimentales pour l'activité biologiques QR.

Les faibles valeurs résiduelles de l'activité biologique et la grande valeur du coefficient de validation ajusté (R^2_{adj}) obtenu suggèrent une bonne capacité prédictive de modèl QSAR développés.

En termes de notre travail, on peut conclure que l'activité biologique QR pour la série étudiée de dérivés de 1,2-dithiole-3-thione peut être modélisé avec succès en utilisant des différents descripteurs moléculaires

Références bibliographiques :

[1] A.R. Leach, Molecular Modelling: Principals and Applications, Addison Wesley Longman Ltd, Edimbourg, Essex, England, 1996.

[2] H. Chermette, chemical abstract, *Spectra*, **22**, 15 (1993).

[3] G.R. Marshall, Burger's Medicinal Chemistry and Drug Discovery, tm. 1, chap. 15, John Wiley and Sons, New York, 1995.

[4] L.P. Graham, Chimie pharmaceutique, de boeck, Paris, 2003.

[5] A. Fortuné, Thèse de doctorat, Techniques de Modélisation Moléculaire Appliquées à L'Etude et à l'Optimisation de Molécules Immunogènes et de Modulateurs de la Chimiorésistance, Université Joseph Fourier, Gronoble I, 2006.

[6] P. Bultinck, Computational Medicinal Chemistry for Drug Discovery, Dekker, New York, 2004.

[7] G.S.A. Aurélie, Thèse de doctorat, Une nouvelle méthode d'apprentissage de données structurées : applications à l'aide à la découverte de médicaments, Université de Paris 6, 2008.

[8] N. Goudarzi, M. Goodarzi, T. Chen, *Med. Chem. Res*., **21**, 437 (2012). :

[9] Y. Zhang, R. Munday, *Mol. Cancer Ther.*, **7**, 3470 (2008).

[10] A. Lüttringhaus, H.B. König, B. Böttcher. *Liebigs Ann. Chem*., **560**, 201 (1948).

[11] N. Lozac'h, J. Vialle, The chemistry of the 1,2-dithiole ring. In: Kharasch H, Meyers CY, editors. The chemistry of organic sulfur compounds (volume 2), Pergamon Press, Oxford, 1966.

[12] J. B. Ep stein, W.E. Decoteau, A. Wilkinso, *Oral. Surg. Oral. Med. Oral. Pathol*., **56**, 495,(1983).

[13] S.S. Ansher, P. Dolan, E. Bueding, *Hepatology*, **3**, 932 (1983).

[14] D. S. Breslow, H. Skolnik, The Chemistry of Heterocyclic Compounds, Multi-Sulfur and Sulfur and Oxygen Five- and Six-Membered Heterocycles, Interscience publishers, New York, pp. 347, (**1966).**

[15] N. Plavac, I. W. J. Still, M. S. Chauhan, and D. M. McKinnon, *Can. J. Chern.*, pp.53, 836 **(1975).**

[16] M.Boukebbous khaled, *Etude de complexes métalliques des ligands soufres, azotes et d'autres molécules halogènes,* mémoire présente pour l'obtention du diplôme de magister en chimie, Universite Mentouri Constantine, (2009).

[17] Xin He, Anne McElwee Reeve, Umesh R. Desai, Glen E.Kellogg, and Kevin A. Reynolds, Antimicrobial agents and chemotherapy, Aug., pp. 3093–3102, **(2004).**

[18] H.Pelster and E.Muchbauer – Brevet allemand n° 1 , pp. 181 -205 **(1965).**

[19] Thomas J. Curphey, and Adam H. Libby, Dianions of 3-oxodithioic acids: preparation and conversion to 3H-1,2-dithiole-3 –thiones. Tetrahedron Letters 41 (2**000).**

[20] Vadim M. Timoshenko, Jean-Philippe Bouillon, Yuriy G. Shermolovich and Charles Portella, Fluorinated ketene dithioacetals. Part 9: Synthesis and some chemical properties of new fluorinated 3H-1,2-dithiole-3-thiones. Tetrahedron Letters 43, pp. 5809–5812, **(2002).**

[21] Janne, PA, Mayer RJ. – Chemoprevention of coloectal cancer. N Engl J Med, pp. 342, 1960, **(2000).**

[22] Kelloff GJ, Johnson JR, Crowell JA, Boone CW, Degeorge JJ, Steele VE, Mehta MU, Temeck JW, Schmidt WJ, Burke G, et al. Approaches to the development and marketing approval of drugs that prevent cancer. Cancer Epidemiol Biomarkers Prev, pp. 4, 1-10, **(1995).**

[23] Livingstone, D. R., 1988. Response of microsomal NADPH-cytochrome c reductase activity and cytochrome P-450 in digestive glands of *Mytilus edulis* and *Littorina littorea* to environmental and experimental exposure to pollutants. Marine Ecology Progress Series, pp.46, 37-43, **(2000).**

[24] Choi J., Liu RM., Forman HJ. Adaptation to oxidative stress: quinonemediated protection of signaling in rat lung epithelial L2 cells. Biochem. Pharmacol., pp. 53, 987-993, **(1997).**

[25] Brown L.A. Glutathione protects signal transduction in type II cells under oxidant stress. Am.1. Physiol., pp. 266, L172-L177, **(1994).**

[26] Schipper D.L., Wagenmans M.J.H., Wagener D.J.T., Peters W.H.M. Glutathione STransferases and cancer. lnt. J. Oncol., pp. 10, 1261-1264, **(1997).**

[27] Schroder K.R., Hallier E., Peter E., Boit H.M. Dissociation of a new glutathione Stransferase activity in human erythrocytes. Biochem. Pharmacol., pp. 43, 1671-1674, **(1992).**

[28] Hallier E., Langhof T., Dannappel D., Lentbecher M., Schroder K, Georgens H.W., Muller A., BoIt H. M. Polymorphism of glutathione conjugation of methyl bromide, ethylene oxide and dichloromethane in human blood : influence of the induction of sister chromatid exchanges in lymphocytes. Arch. Toxicol., pp.67, 173-178, **(1993).**

[29] Tew K.D. Glutathione-associated enzymes in anticancer drug resistance. Cancer Res., pp. 54,4313-4320, **(1994).**

[30] Commandeur J.N.M., Stjntjes G.J., Vermeulen N.P.E. Enzymes and transport systems involved in the formation and disposition of glutathione Sconjugates. Pharmacol. Rev., pp. 47, 271-330, **(1995).**

[31] Stanulla M.,Schrappe M.,Müller Brechlin A., Zimmermann M., WeIteK.Polymorphisms within glutathione S-transferase genes *(GSTM1, GSTT1, GSTP1)* and risk of relapse in childhood B-cell precursor acute lymphoblastic leukemia: a case-control study. Blood, pp. 95, 1222-1228, **(2000).**

[32] Structures of herbicides in complex with their detoxifying enzyme glutathione Stransferase - explanations for the selectivity of the enzyme in plants. Prade, L., Huber, R., Bieseler, B.**,** Structure **6,** pp. 1445-1452, **(1998)**

[33] Talalay P, Fahey JW, Holtzclaw WD, Prestera T and Zhang Y. Chemoprotection against cancer by phase 2 enzyme induction. Toxicol Lett, pp. 82-83, 173-179, **(1995).**

[34] Moon YJ, Wang X and Morris ME. Dietary flavonoids: effects on xenobiotic and carcinogen metabolism. Toxicol in Vitro, pp. 20, 187-210, **(2006).**

[35] Prochaska HJ. Screening strategies for the detection of anticarcinogenic enzyme inducers. J Nutr Biochem, pp. 5, 360-368, **(1994).**

[36] Ernster, L. 'DT-diaphorase', Methods Enzymol. Vol. 10, pp. 309 – 317, **(1967).**

[37] Lind, C., Cadenas, E., Hochstein, P. et al. 'DT-diaphorase. Purification, properties and function', Methods Enzymol. Vol. 186, pp. 287 – 301, (**1990).**

[38] The three-dimensional structure of NAD(P)H:quinone reductase, a flavoprotein involved in cancer chemoprotection and chemotherapy: mechanism of the two-electron reduction. Li, R., Bianchet, M.A., Talalay, P., Amzel, L.M. Proc.Natl.Acad.Sci.USA **92:** 8846-8850, **(1995).**

[39] Siegel, D., Gibson, N.W., Preusch, P.C. et al. , 'Metabolism of mitomycin C by DTdiaphorase: Role in mitomycin C-induced DNA damage and cytotoxicity in human colon carcinoma cells', Cancer Res. Vol. 50, pp. 7483 – 7489, **(1990).**

[40] Gibson, N.W., Hartley, J.A., Butler, J. et al., 'Relationship between DT-diaphorasemediated metabolism of a series of aziridinylbenzoquinones and DNA damage and cytotoxicity', Mol. Pharmacol. Vol. 42, pp. 531 – 536, **(1992).**

[41] Ross, D., Siegel, D., Beall, H. et al., 'DT-diaphorase in activation and detoxification of quinones. Bioreductive activation of mitomycin C', Cancer Metastasis Rev. Vol. 12, pp. 83 –101, **(1993).**

[42] Ross, D., Beall, H., Traver, R.D. et al., 'Bioactivation of quinines by DT-diaphorase. Molecular, biochemical, and chemical studies', Oncol. Res. Vol. 6, pp. 493 – 500, **(1994).**

[43] T. Sugimura, K. Okabe, M. Nagao, The metabolism of 4-nitroquinoline-1-oxide, a carcinogen byan enzyme catalyzing the conversion of 4-nitroquinoline-1-oxide to 4-hydroxylaminoquinoline-1-oxide in rat liver and hepatomas, Cancer Res. pp. 26 ,1717–1721, **(1966).**

[44] A.M. Benson, Conversion of 4-nitroquinoline 1-oxide (4NQO) to 4-hydroxyaminoquinoline 1-oxide by a dicumarol-resistant hepatic 4NQO nitroreductase in rats and mice, Biochem. Pharmacol., pp. 46 , 1217-1221, **(1993).**

[45] R. Munday, B.L. Smith, C.M. Munday, Effect of inducers of DT-diaphorase on the toxicity of 2-methyl- and 2-hydroxy-1,4-naphthoquinone to rats, Chem. Biol. Interact., pp. 123, 219-237, **(1999).**

[46] A.K. Hajos, G.W. Winston, Purified NAD(P)H-quinone oxidoreductase enhances the mutagenicity of dinitropyrenes in vitro, J. Biochem. Toxicol.,vol. 6, pp. 277–282, **(1991).**

[47] A.C. Sartorelli, Therapeutic attack of hypoxic cells of solid tumors: presidential address, Cancer Res. pp.48 ,775–778, **(1988).**

[48] D. Ross, Quinone reductases, in: F.P. Guengerich (Ed.), Comprehensive Toxicology, vol. 3. Biotransformation, Pergamon, New York, pp. 179–198, **(1997).**

[49] D. Ross, H.D. Beall, D. Siegel, R.D. Traver, D.L. Gustafson, Enzymology of bioreductive drug activation, Br. J. Cancer 74 (Suppl. XXVII) S1–S8, **(1996).**

[50] H.D. Beall, S.L. Winski, Mechanisms of action of quinone-containing alkylating agents: NAO1-directed drug development, Front. Biosci. 5 D639–648, **(2000).**

[51] D.C. Liebler, K.L. Kaysen, T.A. Kennedy, Redox cycles of vitamin E: hydrolysis and ascorbic acid dependent reduction of 8a-alkyldioxytocopherones, Biochemistry vol., 28, pp. 9772–9779, **(1989).**

[52] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2597301

[53] H.-D. Höltje, G. Folkeis, Molecular Modeling: Basic Principles and Applications. VCH, New-York, 1997.

[54] C. Audouze, Vers une parallelisation par bandes en chimie quantique, Laboratoire de Mathematique, UMR CNRS 8628, Universite Paris-Sud, 2003.

[55] D. Liotta, Advances in Molecular Modeling, 1, JAI Press, Greenwich, 1988.

[56] C.S. Tsai, An Introduction to Computational Biochemistry, Wiley-Liss, New York, 2002.

[57].http://fr.wikipedia.org/wiki/Relation_quantitative_structure_à_activité.consullté le 06/09/2010.

[58]. A. Crum-Brown, T. Frazer, On the connection between chemical constitution and physiological action. *Transactions of the Royal Society of Edinburgh*, 25,151-203 (1868-69).

[59]. M. C. Richet, 3D QSAR in drug design :theory methods and aplications , *Compt. Rend. Soc. Biol. (Paris) 45***,** 775-776 (1893).

[60]. C. Hansch, T. Fujita, , *J. Am. Chem,* 3D QSAR in drug design : recent advences. *Soc. 86***,** 1616-1626 (1964).

[61]. Jr. Free, S. M, J. W. Wilson, the organic chemestry of drug : design and drug action.*Med. Chem. 7***,** 395-399 (1964).

[62]. C. Hansch, principles for evaluating chemicals in the environment. ***Ace.*** *Chem. Res.* 2**,** 232-239 (1969).

[63]. C. Hansch , J. M. Clayton, *J. Pharm* ,physical proprety prediction in organic chemistry. *Sci. 62***,** 1-21 (1973).

[64]. T.Fujita, T.Ban, , *J. Med,* towards in information theory of complex network : statistical methods and applications. *Chem.44,* 148-152 (1971).

[65]. M. Karelson. Molecular descriptors in QSAR/QSPR. Wiley- Interscience, 385 (2000).

[66]. Jerzy leszczynski, Challenges and advances in computational chemistry and physics, department of chemistry, jackson state university, U.S.A, (2010).

[67]. Wiener, H. Structural determination of paraffin boiling points. *Journal of Chemical* Information and Computer Sciences, 69, p. 17-20, (1947).

[68]. Randić, M. On characterization of molecular branching. *Journal of the American Chemical Society*, 97, p. 6609-6614, (1975).

[69].. Kier, L.B., et Hall, L.H. Molecular connectivity in chemistry and drug research. New- York : Academic Press, 1976.

[70].. Balaban, A.T. Highly discriminating distance-based topological index. *Chemical Physics Letters*, **89**, p. 399-404, (1982).

[71]. Heritage, T.W., *et al.* EVA : A novel theoretical descriptor for QSAR studies.*b Perspectives in Drug Discovery and Design*, **9-11** (0), p. 381-398, (1998).

[72]. Schuur, J.H., Selzer, P., et Gasteiger, J. The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations

and studies of biological activity. *Journal of Chemical Information and Computer Sciences*, **36** (2), p. 334-344, (1996).

[73]. Stoppiglia, H., *et al.* Ranking a random feature for variable and feature selection. *Journal of Machine Learning Research*, **3**, p. 1399-1414,(2003).

[74]. Stoppiglia, H. Méthodes statistiques de sélection de modèles neuronaux ; applications Financières et bancaires [thèse en ligne]. Paris : Université Pierre et Marie Curie, 1997. Disponible sur : http://www.neurones.espci.fr.

[75]. M. Karelson. Molecular descriptors in QSAR/QSPR. Wiley- Interscience, 385 (2000).

[76]. B. Kowalski, R. Gerlach, H. Wold. Systems under Indirect Observation (K. Jörenskog , H. Wold, eds.), North Holland, Amsterdam, 191-206 (1982).

[77]. P. Dagnélie. Statistique Théorique Et Appliquée. Tomes 1 et 2. De Boeck & Larcier s. a. (1998).

[78]. A. R. Katritzky, V. S. Lobanov, M. Karelson. CODESSA Reference Manual. University of Florida, Gainesville (1994).

[79]. A. Fortuné, Thèse de doctorat, Techniques de Modélisation Moléculaire Appliquées à l'Etude et à l'Optimisation de Molécules Immunogènes et de Modulateurs de la Chimiorésistance, Université Joseph Fourier – Grenoble I, (2006).

[80]. Y. Dodge, Statistique Dictionnaire encyclopédique, Springer-Verlag France, Paris, (2007).

[81]. I. Charif, Elaboration des corrélations quantitatives structure-activité des acides carbonés. Etude théorique des effets de solvants sur les équilibres libres cétoénoliques des composés b-dicarbonyles cycliques, Thèse de doctorat, Université de Tlemcen, (2012).

[82] H.T. Phuong, Thèse de doctorat, Synthèse et étude des relations structure/activité quantitatives (QSAR/2D) d'analogues Benzo[c]phénanthridiniques, Université d'Angers, (2007).

[83] A.R. Karuri, , Y. Huang, S. Bodreddigari, C.H. Sutter, B.D. Roebuck, T.W. Kensler, and T.R. Sutter, *J. Pharmac. Exp. Therap.*, **317,** 61 (2006).

[84] S.S. Ansher, P. Dolan, E. Bueding, *Hepatology*, **3**, 932 (1983).

[85] H.J. Prochaska, Y. Yeh, P. Baron, B. Polsky, *P. Natl. Acad. Sci. USA.*, **90**, 3953 (1993).

[86] C.T. Pedersen, *ADV. Heterocyl. Chem.*, **31**, 63 (1982).

[87] G. Burgot, N. Bona, M. V. Christen, and J.L. Burgot, *Int. J. Pharm.*, **129**,

[88] H. Gonzalez-Diaz, E. Olazabal, L. Santana, E. Uriarte, Y. Gonzalez-Diaz, N.

Castanedo, *Bioorg. Med. Chem.*, **15**, 962 (2007).129 (1996).

[89]: SPSS 19 for Windows, SPSS Incorporation, http://www.spss.com.

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

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