

وزارة التعليم العالي و البحث العلمي

BADJI MOKHTAR - ANNABA UNIVERSITY

UNIVERSITE BADJI MOKHTAR - ANNABA



جامعة باجي مختار – عنابة

Année 2022

Faculté des Sciences  
Département de Chimie

## THESE

Présentée en vue de l'obtention du diplôme de  
DOCTORAT EN SCIENCES

Option : Chimie Organique Appliquée

## THEME

*Etude théorique de l'activité inhibitrice de quelques  
enzymes par le  
Criblage virtuel et le docking moléculaire*

Présentée Par:

Mme BENOUIS Sabrina

Devant le jury :

Présidente :	Mme ATTOUI YAHIA Ouassila	Pr	Université de Annaba
Directeur de thèse :	M. FERKOUS Fouad	Pr	Université de Annaba
Examineurs :	M. KHORIEF NACEREDDINE Abdelmalek	Pr	ENSET Skikda
	M. SERIDI Achour	Pr	Université de Guelma
	M. BENDJEFFAL Hassen	MC A	ENSET Skikda
	M. BOUKACHABIA Mourad	MC A	Université de Annaba
Membre invité :	M. KRAIM Khaireddine	MC B	ENSET Skikda

## Résumé

Notre travail de thèse a été consacré au criblage virtuel des chimiothèques en se basant sur le docking moléculaire et les études des relations quantitatives structure-activité (QSAR). Notre travail a été réalisé en deux parties : dans la première partie, des études de docking moléculaire ont été réalisées sur des analogues de gingérol et de l'arbutine extraits de la base de données ZINC. Les résultats obtenus à partir de docking indiquent que certains analogues ont montré un score de Moldock plus élevé (en termes d'énergie négative) par rapport au gingérol, à l'arbutine et aux inhibiteurs de tyrosinase connus expérimentalement. De plus, ces hits ancrés respectent la règle des cinq de Lipinski, affirmant que ces molécules n'ont aucun problème de biodisponibilité. Dans la deuxième partie du travail, un modèle QSAR a été développé pour prédire quantitativement l'activité anti-inflammatoire pour un ensemble de 47 dérivés d'acide 2-[4-(thiazole-2-yl) phényl] propionique. Le modèle proposé a une bonne stabilité, robustesse et prédictivité lorsqu'il est vérifié par validation interne (validation croisée par LOO et randomisation Y) et également par validation externe. Ce dernier a été utilisé pour cribler la chimiothèque PubChem. Parmi les 100 meilleurs composés classés en termes d'activité inhibitrice de cox2 ( $pI_{50}$ ), il existe certains d'entre eux qui possèdent, théoriquement, des valeurs intéressantes de  $pI_{50}$  supérieures à celles de l'inhibiteur de référence l'indométacine.

**Mots clés :** Tyrosinase, cyclooxygénase-2, arbutine, gingérol, Inibition, criblage virtuel, docking moléculaire, QSAR.

**Abstract:**

Our thesis work was devoted to the virtual screening of chemical libraries based on molecular docking and quantitative structure-activity relationships (QSAR) studies. Our work was carried out in two parts: in the first part, molecular docking studies were performed on gingerol and arbutin analogues extracted from the ZINC database. The results obtained from docking indicate that some analogs showed a higher Moldock score (in terms of negative energy) compared to gingerol, arbutin and experimentally known tyrosinase inhibitors. Moreover, these grounded hits adhere to Lipinski's rule of five, stating that these molecules have no bioavailability issues. In the second part of the work, QSAR model was developed to predict quantitatively the anti-inflammatory activity for a set of 47 2-[4-(thiazol-2-yl) phenyl] propionic acid derivatives. The proposed model has good stability, robustness and predictivity when verified by internal validation (cross-validation by LOO and Y randomization) and also by external validation. The latter was used to screen the PubChem chemical library. Among the top 100 compounds ranked in terms of cox2 inhibitory activity ( $pI_{50}$ ), there are some which theoretically possess interesting  $pI_{50}$  values higher than that of the reference inhibitor indomethacin.

**Key words:** Tyrosinase, cyclooxygenase-2, arbutine, gingerol, Inhibition, virtual screening, molecular docking, QSAR.

## الملخص

تم تكريس عمل أطروحتنا للفحص الافتراضي للمكتبات الكيميائية بناءً على دراسات الالتحام الجزيئي ودراسات العلاقة بين التركيب والنشاط الكمي (QSAR). يتكون هذا العمل من جزأين: خلال الجزء الأول، أجريت دراسات الالتحام الجزيئي على مشتقات جينجيرول و أربوتين المستخرجة من قاعدة بيانات ZINC. النتائج التي تم الحصول عليها من الالتحام تشير إلى أن بعض المشتقات أظهرت درجة مولدوك أعلى (من حيث الطاقة السلبية) مقارنة مع جينجرول، أربوتين ومثبطات التيروسيناز المعروفة تجريبياً. بالإضافة إلى ذلك، تحترم هذه النتائج قاعدة ليبينسكي المكونة من خمسة، مشيرة إلى أن هذه الجزيئات لا تعاني من مشكلة التوافر البيولوجي. خلال الجزء الثاني من العمل، تم تطوير نموذج "QSAR" للتنبؤ الكمي بالنشاط المضاد للالتهابات لمجموعة مكونة من 47 مشتق [2 -4] (ثيازول -2-يل) فينيل]حمض بروبيونيك. يتميز النموذج المقترح باستقرار وقوة وتوقع جيد عندما يتم التحقق منه عن طريق التحقق الداخلي (التحقق المتبادل من خلال LOO و Y العشوائية) وأيضاً عن طريق التحقق الخارجي. تم استخدام هذا الأخير لفحص مكتبة المواد الكيميائية "PubChem". من بين أفضل 100 مركب تم تصنيفها من حيث النشاط المثبط لأنزيم "cox2"، هناك بعض المركبات التي تمتلك نظرياً قيماً مثيرة للاهتمام من "pI50" أعلى من مثبط الإندوميتاسين المرجعي.

**الكلمات المفتاحية:** التيروسيناز، السيكلوأوكسيجيناز، جينجيرول، الأربوتين، تثبيط، الالتحام الافتراضي الجزيئي، الفحص الافتراضي QSAR.

# *Dédicaces*

*A la mémoire de ma mère*

*A mon père*

*A ma sœur Houđa et à mon frère sofiane*

*A mes filles zeineb et roukaiia*

*A mon époux youcef*

*A toute ma belle Famille.*

## ***Remerciements***

Ce travail de thèse a été réalisé au sein du Laboratoire de Chimie Organique Appliquée (LCOA), sous la direction du Professeur Fouad FERKOUS à l'université de Annaba.

*Avant tout, je remercie le Seigneur mon Dieu de m'avoir honorée en m'octroyant l'occasion, la force et la patience pour réaliser ce travail.*

J'adresse mes sincères remerciements à Monsieur le Professeur FERKOUS Fouad pour son engagement et d'avoir accepté de diriger ma thèse, pour sa gentillesse, pour ses conseils précieux et surtout pour sa disponibilité durant de longues années. Qu'il soit assuré de ma profonde gratitude.

Mes remerciements vont également à Madame le Professeur ATTOUI YAHIA Wassila de l'Université de Annaba d'avoir accepté de présider le jury de ma soutenance de thèse.

J'aimerais également remercier Messieurs les Professeurs KHORIEF NACREDINE Abdelmalek de l'ENSET de Skikda et SERIDI Achour de l'Université de Guelma pour avoir accepté de faire partie de mon jury de thèse.

J'adresse également mes remerciements à Monsieur BOUKACHABIA Mourad Maître de conférence à l'Université de Annaba et à Monsieur le Maître de conférence BENDJEFAL Hacem maître de conférence à ENSET Skikda pour avoir accepté de juger ce travail.

Je tiens à remercier vivement mes amies Karima, Farida, Sihem.

Merci Youcef pour ton aide, ta compréhension, ton soutien et ta patience.

Et pour finir, merci aussi à tous ceux qui j'ai pu oublier.

## Liste des abréviations

**1D** : une dimension

**2D** : Deux dimensions

**3D** : Trois dimensions

**ADMETox**: Absorption Distribution Métabolisme Elimination – Toxicité

**AMM** : Autorisation de Mise sur le Marché

**Coef** : Coefficients.

**E**: Energie.

**EC** : Enzyme classification

**Eq** : Equation.

**F<sub>obs</sub>** : test de fisher observé.

**H<sub>Bond</sub>**: Hydrogen Bond: liaison hydrogène

**HIV** : Human Immunodeficiency Virus: virus de l'immunodéficience humaine.

**HTS** : High Throughput Screening : Criblage à haut débit

**IC<sub>50</sub>** : la concentration d'inhibiteur pour laquelle la vitesse initiale de formation de produit diminue jusqu'à la moitié de sa valeur mesurée en absence d'inhibiteur.

**ICM** : Internal Coordinate Mechanic interactions

**IUPAC** : International Union of Pure and Applied Chemistry

**MDV**: Molegro Virtual Docker.

**Met** :Méthionine

**N** : nombre d'atome

**NA** : Non Actif

**PDB** : Protein Data Bank

**Phe** : Phénylalanine

**AM1**: semi empirique

**PMF** : Potential of Mean Force : potentiel de force moyenne

**QSAR** : Quantitative Structure-Activity Relationship: Relation quantitative structure-activité.

**Rdt** : rendement

**R&D** : Recherche et Développement

**RD-QSAR** : Receptor Dependent Quantitative Structure-Activity Relationship.

**RMSD**: root-mean-square derivation : Ecart quadratique moyen.

**RX**: rayons X

**SDF** : Structure Data File

**Ser** : Sérine

**SRA** : Structure-Activity Relationship : Relations Structure-Activité

**Thr** : Thréonine

**Tyr**: Tyrosine

**Val** : Valine

**VDW** : Van Der Waals

## Lexique

**Chimiothèque** : librairie de substances chimiques

**Drug discovery** : Ensemble des processus menant à l'identification de molécules actives pouvant être commercialisées en tant que médicament.

**Drug-like (molécules)** : Molécules dont les propriétés drug-like sont satisfaisantes pour une application comme candidat médicament.

**Drug-like (propriétés)** : Caractéristiques des molécules actives ayant un impact sur leur propriétés pharmacocinétiques (ADME) et/ou sur leur éventuelle toxicité.

**Hits** : «touches» : des substances ayant la capacité d'interagir avec la cible biologique et moduler ses effets.

**Leads** : «têtes de séries» : des composés plus actifs et plus sélectifs que les hits dont ils sont issus, tout en présentant des propriétés pharmacocinétiques optimales.

**Pose** : Agencement 3D d'un composé dans un complexe, conformation et position d'un ligand au sein du site actif d'une protéine.

**Screening** : Criblage

**Top-scoring** : Têtes de liste obtenues via le scoring d'un ensemble de molécules.

# Liste des tableaux

## Partie I : Recherche bibliographique

### Chapitre I : Généralités sur les enzymes

**Tableau 1** : Comparaison de différentes méthodes d'immobilisation des enzymes. 26

**Tableau 2** : Enzymes utilisées dans le domaine thérapeutique 28

### Chapitre II : Méthodes de conception de médicaments assistées par ordinateur

**Tableau 1**: Bases de données des petites molécules les plus connues et les plus importantes. 40

**Tableau 2** : liste des blocs de descripteurs fournis par le logiciel Dragon. 46

**Tableau 3** : Différents logiciels de docking. 57

## Partie II : Résultats et discussions

### Chapitre I : Criblage par Docking de molécules analogues d'arbutine et de gingérol

**Tableau 1**: Cavités détectées avec l'algorithme de détection de MVD pour 2Y9X. 71

**Tableau 2** : Type d'atomes contribuant à la formation d'une liaison hydrogène. 76

**Tableau 3** : Les paramètres correspondant à l'énergie de torsion. 77

**Tableau 4** : Les termes d'énergie calculés avec la fonction de score MolDock Optimizer. 76

**Tableau 5** : Les dix meilleurs hits classés par docking moléculaire avec inhibiteurs de tyrosinase connus expérimentalement. 82

**Tableau 6**: structure du ligand, résidus présents, énergie d'interaction ligand-protéine et distances d'interaction pour les meilleurs résultats. 84

**Tableau 7** : Valeurs des paramètres de la règle de Lipinski pour les top hits et le gingérol. 89

**Tableau 8** : Score d'amarrage pour les dix premiers hits classés avec des inhibiteurs de tyrosinase connus expérimentalement. 91

**Tableau 9** : Structure du ligand, résidus présents, énergie d'interaction ligand-protéine et distances d'interaction pour les meilleurs résultats. 92

**Tableau 10** : Valeurs des paramètres de la règle de Lipinski pour les top hits et l'arbutine. 97

### Chapitre 3 : Etude de l'activité anti-inflammatoire par criblage virtuel à base de QSAR

**Tableau 1** : Valeurs expérimentales et calculées de  $pI_{50}$  avec les descripteurs sélectionnés de l'ensemble d'apprentissage (TSET) et de l'ensemble de validation (VSET). 109

**Tableau 2** : Série des modèles optimums à différentes dimensions sélectionnées par la 112

méthode des algorithmes génétiques	
<b>Tableau3</b> : Résultats de l'analyse de la variance (ANOVA)	114
<b>Tableau4</b> : Tableau des coefficients	115
<b>Tableau5</b> : Matrice de corrélation de l'équation 1	117
<b>Tableau6</b> : Valeurs des facteurs d'inflation de la variance et des facteurs de tolérance pour les descripteurs significatifs.	118
<b>Tableau 7</b> : $R^2$ et $Q^2_{cv-100}$ issus du test de randomisation	120
<b>Tableau 8</b> : Comparaison des coefficients du modèle QSAR avec ceux exigés par Tropsha.	121
<b>Tableau 9</b> : Les cinq descripteurs moléculaires qui constituent le modèle QSAR	123
<b>Tableau 10</b> : Structures 2D des 10 meilleurs composés sélectionnés par criblage virtuel.	127
<b>Tableau 11</b> : Les 100 meilleurs composés classés par ordre décroissant des valeurs de pI50.	129

# Liste des figures

## Partie I : Recherche bibliographique

### Chapitre I : Généralités sur les enzymes

<b>Figure 1</b> : Structure tertiaire d'une protéine.	7
<b>Figure 2</b> : Complémentarité stérique.	8
<b>Figure 3</b> : Ajustement induit.	9
<b>Figure 4</b> : Représentation graphique d'une cinétique michaélienne ; la tangente de la courbe représente la vitesse initiale de la réaction.	14
<b>Figure 5</b> : L'effet de pH sur deux enzymes différentes.	15
<b>Figure 6</b> : Effet de la température sur la vitesse d'une (a) réaction non catalysée et (b) une réaction enzymatique.	16
<b>Figure 7</b> : Modèle d'inhibition compétitive.	17
<b>Figure 8</b> : mécanisme d'inhibition compétitive.	17
<b>Figure 9</b> : Graphique de Lineweaver-Burk d'inhibition compétitive	18
<b>Figure 10</b> : Mécanisme d'inhibition incompétitive.	18
<b>Figure 11</b> : Graphique de Lineweaver-Burk d'inhibition incompétitive	19
<b>Figure 12</b> : mécanisme d'inhibition non compétitive.	20
<b>Figure 13</b> : Graphique de Lineweaver-Burk d'inhibition non compétitive	20
<b>Figure 14</b> : Mécanisme d'inhibition irréversible	21
<b>Figure 15</b> : Immobilisation par liaison covalente.	22
<b>Figure 16</b> : Immobilisation par réticulation.	23
<b>Figure 17</b> : Immobilisation par co-réticulation.	23
<b>Figure 18</b> : Immobilisation par Inclusion.	24
<b>Figure 19</b> : Immobilisation par adsorption.	25

### Chapitre II : Méthodes de la conception de médicaments assistées par ordinateur

<b>Figure 1</b> : Etapes du processus de recherche et développement des médicaments.	35
<b>Figures 2</b> : Classification des méthodes de criblage virtuel.	41
<b>Figure 3</b> : Etapes effectuées pour le développement d'un modèle QSAR	43
<b>Figure 4</b> : Dessin et optimisation des structures moléculaires.	44
<b>Figure 5</b> : Division des molécules collectées	45
<b>Figure 6</b> : Génération de descripteurs moléculaires indispensables pour établir la relation QSAR	45
<b>Figure 7</b> : Procédures de sélection de descripteurs moléculaires.	47
<b>Figure 8</b> : construction du modèle mathématique.	48

<b>Figure 9 :</b> Méthodes de validation du modèle QSAR.	49
<b>Figure 10 :</b> Flexibilité moléculaire en docking protéine-ligand	52
<b>Figure 11 :</b> Reconstruction incrémentale (programme Dock) : la molécule est divisée en fragments puis dockée et reconstruite dans le site actif.	54

## **Partie II : Résultats et discussions**

### **Chapitre I : Criblage par Docking de molécules analogues d'arbutine et de gingérol**

<b>Figure 1 :</b> schémas des activités monophénolase et catéchol oxydase de la tyrosinase.	66
<b>Figure 2 :</b> Exemples du phénomène d'hyperpigmentation de la peau.	66
<b>Figure 3 :</b> Exemples d'inhibiteurs de la tyrosinase connus expérimentalement.	67
<b>Figure 4 :</b> La tyrosinase, extraite du champignon <i>Agaricus Bisporus</i> (PDB : 2Y9X).	68
<b>Figure 5 :</b> Illustration des cinq cavités du Tyrosinase (grilles en vert) par l'algorithme de détection de MVD	71
<b>Figure 6 :</b> Illustration de l'espace de recherche.	72
<b>Figure 7 :</b> Principe de l'algorithme évolutionnaire différentiel de MDV	74
<b>Figure 8 :</b> Matrice de RMSD générée par MDV après re-docking	79
<b>Figure 9 :</b> La superposition de la structure cristallographique du ligand (en jaune) et la pose obtenue après re-docking (en rose), RMSD =0,78Å.	79
<b>Figure 10 :</b> La cavité de liaison de l'enzyme tyrosinase avec les résidus Phe264, Phe90, His94, His244, Phe292, Ala286, Val283, Gly281, Met280, Asn260, Ser282, His85, His61, His263, His259, His296 et les ions cuivre représentés par deux boules argentées.	80
<b>Figure 11 :</b> Structures moléculaires des dix meilleurs hits classés par docking moléculaire et celle du gingérol.	81
<b>Figure 12 :</b> Tracé linéaire ajusté de la corrélation entre l'énergie d'interaction et le poids moléculaire.	83
<b>Figure 13 :</b> a) Top dix hits ancrés dans le site actif de la tyrosinase et b) vue de surface électrostatique de la tyrosinase montrant les dix premiers hits d'amarrage à l'intérieur du site de liaison.	86
<b>Figure 14:</b> Interactions prédites des liaisons H (lignes pointillées soufflées) reliant ZINC13377891 (rouge) aux résidus Ala323, His244, Asn260, Met280 de l'enzyme tyrosinase.	87
<b>Figure 15:</b> Interactions prédites des liaisons H (lignes pointillées soufflées) entre ZINC13377888 (rouge) et Ala323, His85, Asn260, Met280 résidus de l'enzyme tyrosinase.	87
<b>Figure 16:</b> Interactions prédites des liaisons H (lignes pointillées soufflées) entre le gingérol (rouge) et le résidu Met280 de l'enzyme tyrosinase.	88

<b>Figure 17 :</b> Structures moléculaires des dix meilleurs hits classés par docking moléculaire et celle de l'arbutine.	90
<b>Figure 18.</b> Tracé linéaire ajusté de la corrélation entre l'énergie d'interaction et le poids moléculaire.	92
<b>Figure 19.</b> a) Top dix hits ancrés dans le site actif de la tyrosinase et b) vue de surface électrostatique de la tyrosinase montrant les dix premiers hits d'amarrage à l'intérieur du site de liaison.	94
<b>Figure 20.</b> Interactions prédites des liaisons H (lignes pointillées bleues) entre les résidus Zinc03978768 (rouge) et His263, His61, His85, Met280 de l'enzyme tyrosinase.	95
<b>Figure 21.</b> Interactions prédites des liaisons H (lignes pointillées bleues) reliant Zinc31160316 (rouge) au résidu Met280 de l'enzyme tyrosinase.	96
<b>Figure 22.</b> Interactions prédites des liaisons H (lignes pointillées bleues) entre l'arbutine (rouge) et les résidus Met280, His85, Asn260, His244 de l'enzyme tyrosinase.	96
<b>Chapitre 3 : Etude de l'activité anti-inflammatoire par criblage virtuel à base de QSAR</b>	
<b>Figure 1 :</b> Structure générale des inhibiteurs de la cyclooxygénase (COX2) dérivés de l'acide 2-[4-(thiazole-2-yl)phényl]propionique.	103
<b>Figure 2 :</b> Etapes du protocole de criblage virtuel utilisé pour identification des nouveaux inhibiteurs de cyclooxygénase2.	104
<b>Figure 3 :</b> Structures développées des dérivés d'acide 2-[4-(thiazole-2-yl)phényl]propionique	108
<b>Figure 4 :</b> Représentation des coefficients de détermination, obtenus par GA, en fonction du nombre de descripteurs.	113
<b>Figure 5 :</b> Les différentes méthodes utilisées pour la validation de notre modèle QSAR.	119
<b>Figure 6 :</b> Valeurs de $Q^2_{cv-100}$ en fonction de $R^2$ issues du test de randomisation	120
<b>Figure 7 :</b> Valeurs observées en fonction des valeurs prédites des activités pour les ensembles d'apprentissage et de prédiction.	122
<b>Figure 8 :</b> Calcul du descripteur F03[C-S] pour la molécule 1 de l'ensemble étudié	123
<b>Figure 9 :</b> Structure des molécules 20, 21, 22, 23 et 24 montrant l'influence du descripteur B03[N-O]	124
<b>Figure 10 :</b> Structure de la molécule 8 montrant l'influence du descripteur F03[C-S]	125
<b>Figure 11 :</b> Structure des molécules 2, 3 et 4 de l'ensemble étudié.	126

# **Table des matières**

## Tables des matières

Résumé	
Abstract	
الملخص	
Dédicaces	
Remerciements	
Liste des abbréviations	XIV
Lexique	VII
Liste des tableaux	IX
Liste des figures	XI
Table des matières	XIV

### Partie I

#### Chapitre I : Généralités sur les enzymes

<b>1. Structure des enzymes et fonction</b>	6
1.1 Mécanisme d'action des enzymes	7
1.1.1 Modèle de Fisher (modèle de serrure et de clé)	8
1.1.2 Modèle d'ajustement induit	8
1.1.3 Catalyse covalente	9
<b>2. Classification des enzymes</b>	9
<b>3. Cinétique enzymatique</b>	10
3.1. Facteurs influençant la vitesse d'une réaction enzymatique	14
3.1.1. Effet du pH	14
3.1.2. Effet de la température	15
3.1.3. Inhibition enzymatique	16
3.1.3.1. Inhibitions réversibles	17
3.1.3.1.1. Inhibition compétitive	17
3.1.3.1.2. Inhibition incompétitive	18
3.1.3.1.3. Inhibition non-compétitive	19
3.1.3.2. Inhibition irréversible	21
<b>4. Immobilisation des enzymes</b>	21
4.1. Méthodes d'immobilisation	22
4.1.1. Immobilisation par liaison covalente	22

4.1.2. Immobilisation par réticulation et co-réticulation	23
4.1.3. Inclusion ou piégeage	24
4.1.4. Immobilisation par adsorption	25
4.2. Avantages et désavantages des méthodes d'immobilisation	26
<b>5. Domaine d'application des enzymes</b>	27
5.1 Applications des enzymes dans le domaine agro-alimentaire	27
5.1.1 Raffinage enzymatique d'huiles alimentaires	27
5.1.2 Enzymes utilisées dans la panification	27
5.2 Applications analytiques et thérapeutiques	28
5.2.1. Enzymes utilisées dans les méthodes d'analyses de biologie moléculaire	28
5.2.2. Enzymes utilisées en thérapie et analyse.	28
5.3. Applications industrielle	29
5.3.1. Industrie du papier	29
5.3.2. Industrie des détergents.	29
<b>6. Conclusion</b>	30
<b>7. Références</b>	31

## **Chapitre II : Méthodes de conception de médicaments assistées par ordinateur.**

<b>1. Introduction sur de la découverte des médicaments</b>	34
1.1. Le choix d'une cible thérapeutique	35
1.2. Identification des touches ( <i>hits</i> )	35
1.3. Génération et optimisation de tête de série (leads)	36
1.4. Les tests précliniques et cliniques	37
1.4.1 Tests précliniques	37
1.4.2. Tests cliniques	37
1.4.2.1. Les essais de la phase I	38
1.4.2.2. Les essais de la phase II	38
1.4.2.3. Les essais de la phase III	38
1.4.2.4. Les essais de la phase IV	38
<b>2. Le principe du criblage virtuel (<i>in silico</i>)</b>	39
2.1. Généralités	39
2.2. Préparation des chimiothèques	39
2.3. Préfiltrage de la chimiothèque	40

<b>3. Les différentes méthodes du criblage virtuel</b>	41
3.1. Méthodes « ligand-based virtual screening »	42
3.1.1. Recherche de similarité	42
3.1.2. Développement de Modèles des relations quantitatives structure –active (QSAR)	42
3.1.2.1. Collecte et traitement de l'ensemble des données	43
3.1.2.2. Utilisation des descripteurs moléculaires	44
3.1.2.3. Sélection des descripteurs	47
3.1.2.4. Construction du modèle QSAR	48
3.1.2.5. Validation du modèle	49
3.1.2.5.1. Validation interne	49
3.1.2.5.2. Validation externe	50
3.2. Méthodes « Structure -based virtual screening »	51
3.2.1. Méthodes de docking moléculaire	51
3.2.1.1. Flexibilité moléculaire dans le docking des protéines-ligands	52
3.2.1.2. Docking avec ligand rigide	52
3.2.1.3. Docking avec ligand flexible	53
3.2.1.4. Algorithmes de recherche conformationnelle	53
3.2.1.4.1. Recherche systématique	53
3.2.1.4.2. Recherche aléatoire	54
3.2.1.4.3. Méthodes de simulation	55
3.2.1.5. Fonctions de score	55
3.2.1.5.1. Fonctions de score basées sur les champs de force	55
3.2.1.5.2. Fonctions de score empiriques	56
3.2.1.5.3. Fonctions de scores basées sur les connaissances	57
3.2.1.5.4. Fonction de score consensus	57
3.2.1.6. Principaux logiciels de docking	57
<b>4. Conclusion</b>	58
<b>5. Références</b>	59

## Partie II : Résultats et discussions

### Chapitre I : Criblage par Docking de molécules analogues d'arbutine et de gingérol

<b>1. Introduction</b>	65
------------------------	----

<b>2. Méthodes de recherche</b>	68
2.1. Choix de la cible (Enzyme)	68
2.2. Chimiothèque de ligands : recherche de similarité chimique	69
2.3. Processus du docking moléculaire	69
2.3.1. Préparation de la cible et les ligands par MDV	69
2.3.2. Détection des cavités et définition de la région du site actif	70
2.4. Protocole et validation	72
2.4.1. Algorithme de recherche conformationnelle	72
2.4.2. Fonction de score	74
2.4.3. Validation du protocole de docking	77
<b>3. Criblage des chimiothèques</b>	79
3.1. Criblages de la chimiothèques des analogues du gingérol	79
3.2. Criblage de la chimiothèque des analogues de l'arbutine	89
<b>4. Conclusion</b>	98
<b>5. Références</b>	99

**Chapitre 3 : Etude de l'activité anti-inflammatoire par criblage virtuel à base de QSAR**

<b>1. Introduction</b>	103
<b>2. Protocole de criblage virtuel</b>	104
<b>3. Création et préparation de la chimiothèque</b>	105
<b>4. Développement du modèle QSAR</b>	105
4.1. Méthodes expérimentales	105
4.1.1. Ensemble de molécules et activités biologiques	105
4.1.2. Dessin et optimisation des structures	110
4.1.3. Génération des descripteurs	110
4.1.4. Sélection des descripteurs	110
4.1.5. Construction du modèle	111
4.2. Résultats et discussion	112
4.2.1. Analyse de la justesse du modèle optimal	113
4.2.2. Tests de colinéarité et Multicolinéarité	117
4.2.3. Validation interne et externe du modèle de QSAR obtenu	119
4.2.3.1. Validation interne	119
4.2.3.2. Validation externe	121

4.3. Interprétation physico-chimique des descripteurs	122
4.3.1. Influence des descripteurs B03[N-O] et F03[C-S]	123
4.3.2. Influence de TE2	125
4.3.3. Influence des descripteurs H4e et R7e+	126
<b>5. Calcul des activités des molécules de la chimiothèque : Application du modèle QSAR</b>	127
5.1. Visualisation des meilleures structures moléculaires	129
<b>6. Conclusion</b>	131
<b>7. Références</b>	132
<b>Conclusion générale</b>	134

# **Introduction générale**

Anciennement, un grand nombre de médicaments ont été découverts tout simplement grâce à l'identification de principes actifs extraits de substances naturelles historiquement utilisées dans la médecine non-conventionnelle, ou même par hasard. Mais plus le nombre de médicaments connus augmente et plus les probabilités de faire une telle découverte sont faibles. Par la suite, les avancées dans le domaine de la synthèse chimique et de la pharmacologie ont conduit à une démarche de recherche systématique permettant l'élaboration de nouveaux médicaments de plus grande efficacité<sup>1,2</sup>.

Le processus de recherche et développement de nouveaux médicaments est extrêmement long et coûteux. Au cours des phases successives, des milliers de molécules doivent être triées et sélectionnées, afin d'obtenir un nombre très limité de candidats<sup>3</sup>. Pour chaque nouveau médicament, le temps passé depuis la première étape, au cours de laquelle une cible biologique pertinente dans un processus pathologique donné est identifiée, jusqu'à la mise sur le marché d'un composé tourne autour d'une moyenne d'environ 12 à 14 ans. Le coût global estimé est de 800 millions de dollars par composé mis sur le marché<sup>1-3</sup>.

Le criblage virtuel est la stratégie *in silico* la plus utilisée pour l'identification de *touches* (*hits*) dans le cadre de la recherche de nouveaux médicaments<sup>4-6</sup>. Celui-ci fait désormais partie intégrante de la plupart des programmes de recherche de composés bioactifs, que ceux-ci se déroulent en milieu académique ou industriel, car il constitue un complément essentiel au criblage biologique haut-débit<sup>7-9</sup>. Le criblage virtuel permet l'exploration de larges chimiothèques (> 10<sup>6</sup> molécules) à la recherche de composés actifs vis-à-vis d'une cible thérapeutique donnée. Ce processus vise à réduire de façon significative la chimiothèque de départ à une liste limitée de composés jugés les plus prometteurs. Cette approche conduit souvent à une nette amélioration de la concentration de molécules actives pour la cible, tandis qu'une sélection aléatoire de molécules de la chimiothèque ne saurait fournir un tel enrichissement. Ainsi, le temps aussi bien que les coûts de l'identification de nouvelles touches peuvent être réduits de façon significative<sup>10</sup>.

Dans nos travaux de thèse nous présentons des études de criblage virtuel basées sur le docking moléculaire et la relation quantitative structure-activité, effectuées sur deux cibles thérapeutiques différentes : la première est la tyrosinase impliquée dans le processus de la production de la mélanine responsable de la

## Introduction générale

---

coloration de la peau, des yeux et des cheveux, la seconde cible est la cyclooxygénase2 responsable de l'inflammation.

L'objectif de nos études et de découvrir de nouvelles molécules inhibitrices de ces deux cibles afin d'enrichir les classes thérapeutiques des anti-hyperpigmentation et des anti-inflammatoires.

Ce manuscrit est divisé principalement en deux parties : une première partie bibliographique et une deuxième partie consacrée aux résultats et discussions.

Le premier chapitre de la partie bibliographique présente une introduction sur les enzymes, leur définition, leur mode d'action, la cinétique enzymatique ainsi que les principales applications industrielles des enzymes.

Le deuxième chapitre présente les différentes techniques informatiques utilisées dans le développement de nouvelles substances médicamenteuses, notamment la méthode QSAR et le docking moléculaire.

La deuxième partie « résultats et discussions » est constituée de deux chapitres : le premier chapitre est consacré à l'étude de l'inhibition de l'enzyme tyrosinase par le docking moléculaire, criblage virtuel, aux résultats obtenus et à leurs discussions. Le deuxième chapitre représente une étude quantitative structure-activité inhibitrice de l'enzyme cyclooxygénase2 et son application dans le criblage virtuel des chimiothèques.

Enfin, nous terminerons ce manuscrit par une conclusion générale et des perspectives.

## Références

1. K.C. Nicolaou, *Angewandte Chemie*, 53(35), **2014**, 9128-9140.
2. H.Beck, M.Härter, B.Haß, C. Schmeck, L. Baerfacker, *Drug discovery today*, 27(6), **2022**, 1560-1574.
3. R. E. Hubbard, *Molecular BioSystems*, 1, **2005**, 391-406.

4. M.J. Keiser, V. Setola, J.J. Irwin, C. Laggner, A.I. Abbas, S.J. Hufeisen, N.H. Jensen, M.B. Kuijer, R.C. Matos, T.B. Tran, R. Whaley, R. a Glennon, J. Hert, K.L.H. Thomas, D.D. Edwards, B.K. Shoichet, B.L. Roth, *Nature*, 462, **2009**, 175–181.

5. B.K. Shoichet, *Nature*, 432, **2004**, 862–865.

6. J. Waszkowycz, B. Perkins, T. D. J. Sykes, R. A. Li, *IBM Syst. J*, 40(2), **2001**, 360–376.

7. J. Bajorath, *Nat. Rev. Drug Discov*, 1(11), **2002**, 882–894.

8. F. Stahura, J. Bajorath, *Comb Chem High Throughput Screen*, 7(4), **2004**, 259–69.

9. K. Shivani, *In Silico Drug Design*, **2019**, 161–189.

10. M.Gupta R. Sharma, M. Singh, A. Kumar, *Computational Biology and Chemistry*, 26, **2018**, 210-217.

# **Partie I**

## **Recherche bibliographique**

# **Chapitre I**

## **Généralités sur les enzymes**

Les enzymes sont des protéines qui allient la capacité de reconnaître très spécifiquement des molécules<sup>1</sup>, au pouvoir de catalyser efficacement leur transformation en composées utiles au métabolisme des organismes vivants.

Les réactions chimiques dans les systèmes biologiques se font rarement en l'absence d'enzymes.

Une enzyme ne favorise pas seulement une réaction chimique donnée, elle empêche aussi la présence de réactions secondaires gênantes.

Au cours de ces dernières années, le nombre d'enzymes connues a considérablement augmenté et leurs applications sont multiples, en chimie analytique, en pharmacologie, en toxicologie, dans l'industrie agro-alimentaire, dans le secteur biomédical et en chimie organique.

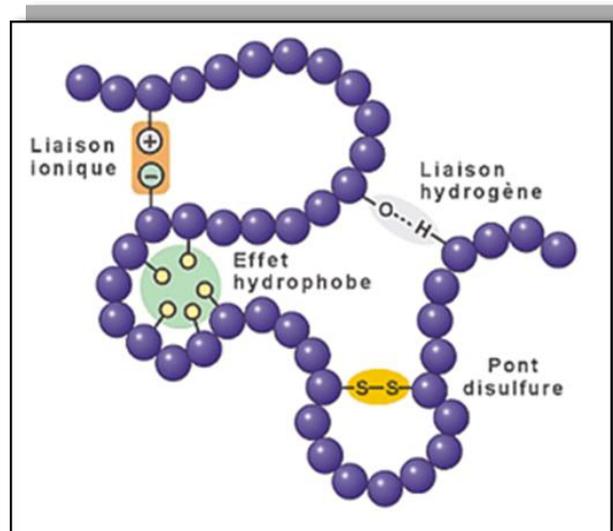
## 1. Structure des enzymes et fonction

Comme toute protéine l'enzyme est constituée d'enchaînement d'acides aminés<sup>2</sup>, qui sont reliés les uns aux autres par des liaisons peptidiques qui se forment entre le groupement carboxylique du premier acide aminé et le groupement aminé suivant.



Cette chaîne d'acides aminés détermine la structure primaire de l'enzyme. La structure secondaire des protéines est classiquement définie par le repliement de la chaîne polypeptidique en hélice  $\alpha$  ou en feuillet  $\beta$ .

La structure tertiaire est due à la formation de différentes liaisons et interactions, dont les ponts disulfures, les liaisons hydrogène, liaisons ioniques et les interactions hydrophobes (**Figure 1**).



**Figure 1** : Structure tertiaire d'une protéine.

### 1.1 Mécanisme d'action des enzymes

Le mécanisme d'action est basé sur une réaction chimique, dans laquelle l'enzyme se lie au substrat et forme finalement un complexe enzyme-substrat.<sup>3</sup> Cette réaction a lieu dans une zone relativement petite de l'enzyme appelée site actif ou catalytique. En d'autres termes, le mécanisme d'action enzymatique est basé sur la nature de l'interaction enzyme-substrat, qui explique la spécificité de réaction des catalyseurs biologiques. Le site actif d'une enzyme est constitué de plusieurs acides aminés, situés à distance les uns des autres dans la chaîne peptidique. Ces acides aminés sont rapprochés par le repliement résultant de la structure secondaire et tertiaire des enzymes. Les chaînes latérales de résidus d'acides aminés au niveau du site actif fournissent des groupes pour la liaison avec des groupes spécifiques du substrat. Les cofacteurs assistent la catalyse. Le substrat forme des liaisons avec des résidus d'acides aminés dans le domaine de liaison au substrat du site actif. La liaison induit une réaction conformationnelle dans le site actif. Au cours de la réaction, l'enzyme forme un complexe d'état de transition. Au fur et à mesure que les produits de la réaction se dissocient, l'enzyme revient à son état d'origine. Deux différents modèles postulés pour le mécanisme d'action enzymatique sont donnés ci-dessous.

### 1.1.1 Modèle de Fisher (modèle de serrure et de clé)

Il s'agit d'un modèle rigide du site catalytique, proposé par Emil Fischer en 1894<sup>4</sup>. Le modèle explique l'interaction entre un substrat et une enzyme en termes d'analogie serrure et clé. Dans ce modèle, le site catalytique est supposé être préformé. Le substrat s'adapte comme une clé dans une serrure. L'inconvénient de ce modèle est la rigidité implicite du site catalytique. Le modèle ne peut pas expliquer les changements dans la structure enzymatique en présence de modulateurs allostériques.

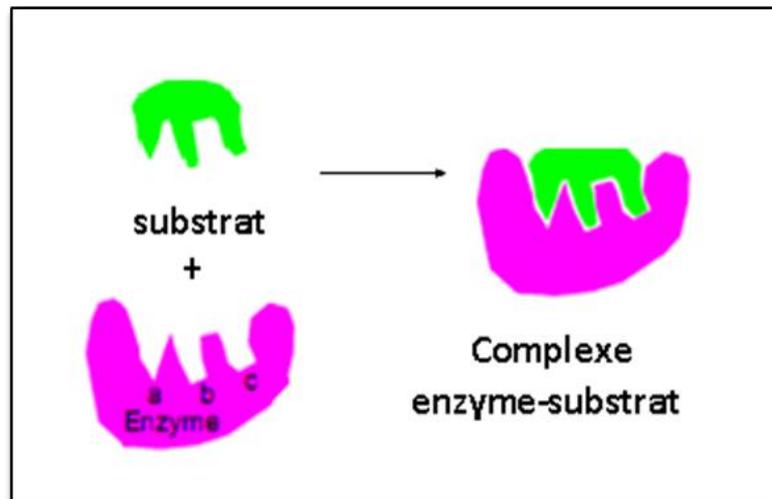
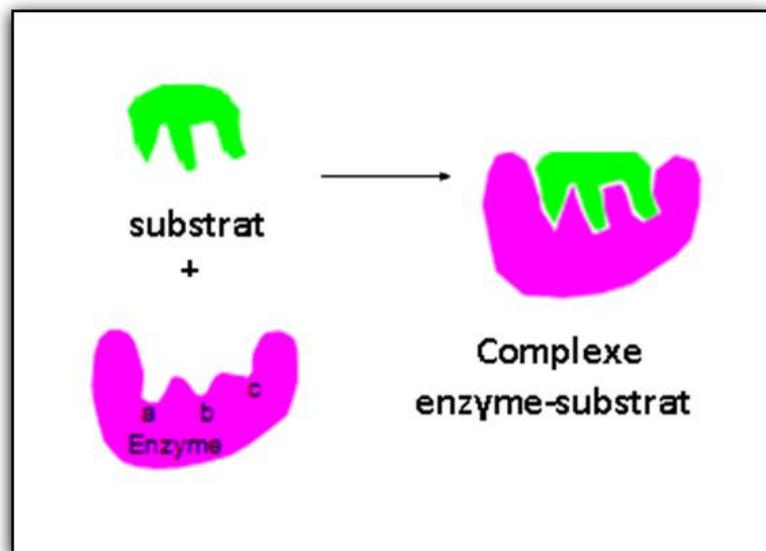


Figure 2 : Complémentarité stérique.

### 1.1.2 Modèle d'ajustement induit

Contrairement à la méthode ci-dessus, ce modèle suggère un mode flexible pour le site catalytique. Pour surmonter les problèmes du modèle de serrure et de clé dus au site catalytique rigide, Koshland<sup>5-7</sup> a proposé un modèle d'ajustement induit en 1963. La caractéristique importante de cette procédure est la flexibilité du site actif. Dans le modèle d'ajustement induit, le substrat induit un changement conformationnel dans le site actif de l'enzyme de sorte que le substrat s'insère dans le site actif de la manière la plus pratique afin de favoriser la réaction chimique. Cette méthode suggère une inhibition compétitive, une modulation allostérique et une inactivation des enzymes lors de la dénaturation.

La théorie de Michaelis-Menten de l'action enzymatique<sup>6</sup> offre la base de la plupart des recherches actuelles sur le mécanisme de l'action enzymatique. Ce concept de schéma complexe enzyme-substrat suppose la combinaison de l'enzyme et du substrat dans la première phase (parfois appelée phase de transition) de l'activité enzymatique et la libération de l'enzyme et des produits de la catalyse dans la phase deux de la réaction.



**Figure 3** : Ajustement induit.

### 1.1.3 Catalyse covalente

La catalyse covalente est mise en évidence dans des enzymes capables de former des liaisons covalentes entre la substance et le groupement catalytique du site actif<sup>7</sup>. Un certain nombre d'enzymes réagissent avec leurs substrats pour former des complexes enzyme-substrat très instables, liés par covalence, qui subissent une réaction supplémentaire pour donner des produits beaucoup plus facilement que dans une réaction non catalysée.

## 2. Classification des enzymes

On trouve encore, à l'heure actuelle, la dénomination ancienne de certaines enzymes telle que trypsine ou pepsine. Cependant, dans la mesure où l'on s'intéresse moins à la nature qu'à la fonction de l'enzyme, on utilise plutôt la nomenclature de

l'Union Internationale de Biochimie qui recommande l'utilisation du suffixe « ase » couplé parfois avec le substrat (uréase, phosphatase...), mais généralement avec le type de réaction catalysée (hydrolase, oxydase, isomérase...).

Selon les réactions catalysées, les enzymes peuvent être classées en six grandes catégories:

- 1. Oxydoréductases** : Elles catalysent des réactions d'oxydation ou de réduction : oxygénation de la liaison C-H, C-C, C=C, ou l'élimination ou l'addition d'un atome d'hydrogène équivalent.
- 2. Transférases** : Il s'agit d'enzymes responsables de transfert d'un groupement fonctionnel aldéhyde, cétone, acyle, amine, sucre, ou d'un radical d'une molécule à une autre (méthyle).
- 3. Hydrolases** : Elles catalysent des réactions d'hydrolyse enzymatique (coupure des liens C-C, C-O, C-N et autres par de l'eau).
- 4. Lyases** : qui catalysent l'addition ou l'élimination de groupes sur des liaisons doubles C=C, C=N, C=O.
- 5. Isomérases** : qui catalysent le transfert de groupes dans une même molécule pour produire des formes isomères.
- 6. Ligases** : Favorisent la formation des liaisons C-C, C-S, C-O et C-N.

### 3. Cinétique enzymatique<sup>8</sup>

La cinétique enzymatique traite la vitesse des réactions enzymatiques, fournissant une information indirecte sur le mécanisme d'une réaction, sur la spécificité de l'enzyme ainsi que sur d'autres paramètres précisant les propriétés physiques.

Une réaction enzymatique est caractérisée par ses constantes de vitesse.

En 1913, L. Michaelis et M. Menten ont proposé une théorie pour décrire les réactions enzymatiques. Dans leur théorie, les réactions enzymatiques se déroulent

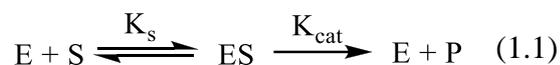
en deux étapes: d'abord, un complexe enzyme-substrat (E-S) se forme rapidement, puis ce complexe est lentement converti en produit.

Dans l'étape limitante de la réaction on admet que l'enzyme (E) et le substrat (S) restent à tout moment en équilibre thermodynamique avec le complexe.

Les valeurs de la constante de Michaelis  $K_m$  et de la vitesse maximale  $V_m$  d'une enzyme peuvent, fournir certains renseignements sur son activité. Pour la plupart des enzymes,  $K_m$  est approximativement égale à la constante de dissociation du complexe enzyme-substrat. La valeur de  $K_m$  est inversement proportionnelle à l'affinité du substrat pour l'enzyme. Pour les concentrations en substrat inférieures ou égales à  $K_m$ , la vitesse de la réaction dépend de la vitesse d'association entre enzyme et substrat.

### Equation de Michaelis-Menten

La réaction enzymatique peut être représentée sous la forme simplifiée suivante :



Cette première forme de l'équation Michaelis-Menten permet d'exprimer la vitesse de la réaction en fonction des concentrations de l'enzyme et du substrat.

$$v = k_{cat} \frac{[E_t][S]}{K_s + [S]} \quad (1.2)$$

Où [S] la concentration du substrat.

[ $E_t$ ] la concentration totale de l'enzyme qui est reliée aux concentrations d'enzyme libre [E] et de complexe enzyme-substrat [ES] par l'équation ;

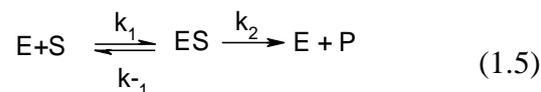
$$[E]_t = [E] + [ES] \quad (1.3)$$

$K_S$  est la constante de dissociation du complexe ES, ou complexe Michaelis-Menten et est représentée par:

$$K_S = \frac{[E][S]}{[ES]} \quad (1.4)$$

$K_{cat}$  la constante catalytique.

Cette forme générale de l'équation Michaelis-Menten rend compte de nombreuses réactions enzymatiques.



Ce modèle de Michaelis-Menten a été amélioré par Briggs et Maldane qui ont introduit l'hypothèse de l'état stationnaire qui implique que la concentration  $[ES]$  reste constante soit :  $d[ES]/dt=0$ .

L'expression de la vitesse initiale devient alors (1.6).

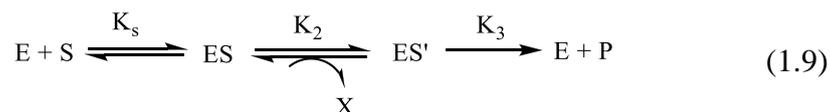
$$v = \frac{V_m [S]}{K_M + [S]} \quad (1.6)$$

Avec  $V_m$  = vitesse initiale maximale :  $V_m = k_2[E]_t$  (1.7)

et  $K_M$  la constante Michaelis-Menten qui est donné par l'équation (1.8)

$$K_M = \frac{k_{-1} + k_2}{k_1} \quad (1.8)$$

Une troisième forme de l'équation de Michaelis-Menten est utilisée dans laquelle on postule la formation d'un complexe covalent entre l'enzyme et le substrat ( $ES'$ ), avec une consommation du substrat (x).



L'équation qui décrit l'état stationnaire pour ES' est :

$$\frac{d[ES']}{dt} = k_2 [ES] - k_3 [ES'] = 0 \quad (1.10)$$

Retenons que :

$$K_S = \frac{[E][S]}{[ES]} \quad (1.11)$$

Et la concentration totale de l'enzyme est  $[E]_t = [E] + [ES] + [ES']$ . Puis on a :

$$E_t = [ES] \left[ \frac{K_S}{[S]} + 1 + \frac{k_2}{k_1} \right] \quad (1.12)$$

La vitesse de formation du produit de la réaction est :

$$v = k_2[ES] \quad (1.13)$$

En introduisant la valeur de [ES] de l'équation (1.12) dans l'équation (1.13), on obtient la loi des vitesses à l'état stationnaire.

$$v = [E]_t [S] \left[ \frac{(k_1 k_2) / (k_2 + k_3)}{(K_S k_3) / (k_2 + k_3) + [S]} \right] \quad (1.14)$$

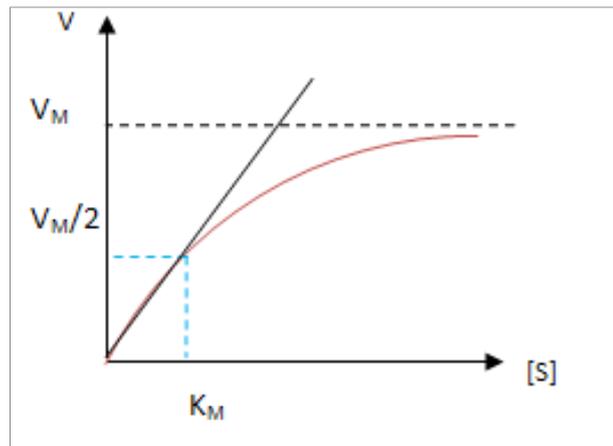
L'équation est de forme identique à celle de Michaelis-Menten obtenue auparavant, d'où:

$$K_M = K_S \frac{k_3}{k_2 + k_3} \quad (1.15)$$

Et:

$$k_{cat} = \frac{k_2 k_3}{k_2 + k_3} \quad (1.16)$$

Les trois formulations de l'équation de vitesse de Michaelis-Menten illustrent un principe général de portée essentielle: de la forme de l'équation de vitesse d'une réaction, on ne peut pas déduire le mécanisme impliqué dans la réaction, des mécanismes différents pouvant présenter une cinétique identique.



**Figure 4** : Représentation graphique d'une cinétique michaelienne; la tangente de la courbe représente la vitesse initiale de la réaction.

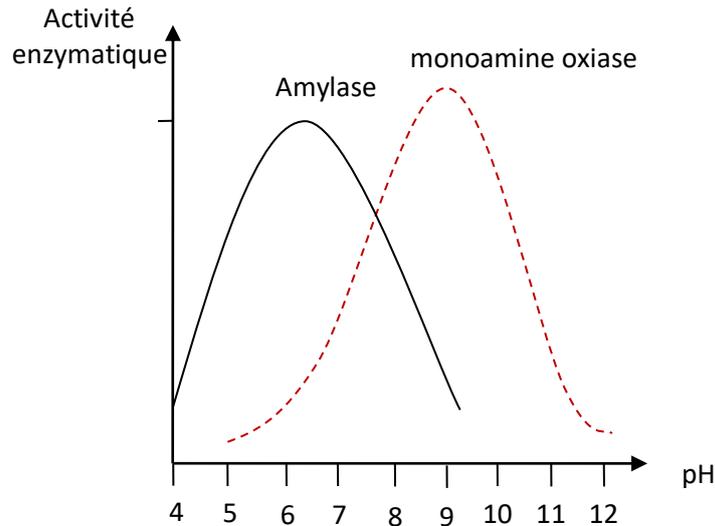
### 3.1. Facteurs influençant la vitesse d'une réaction enzymatique

Nous allons passer en revue l'effet des différents facteurs réactionnels sur l'activité de l'enzyme.

#### 3.1.1. Effet du pH

La plupart des enzymes sont actives dans un domaine limité de pH. On peut attribuer ce fait à la stabilité de la protéine enzymatique dans une certaine zone de pH.

On explique simplement la sensibilité au pH des activités enzymatiques par le fait que les substrats et les sites actifs des enzymes portent souvent des groupes fonctionnels acides ou basiques dont l'ionisation varie avec le pH. Ce dernier peut agir sur les paramètres cinétiques tels que  $K_{cat}$ ,  $K_M$  et  $K_{cat} / K_M$ . Ces effets de pH diffèrent d'une enzyme à l'autre (**Figure 5**).



**Figure 5** : L'effet de pH sur deux enzymes différentes.

### 3.1.2. Effet de la température

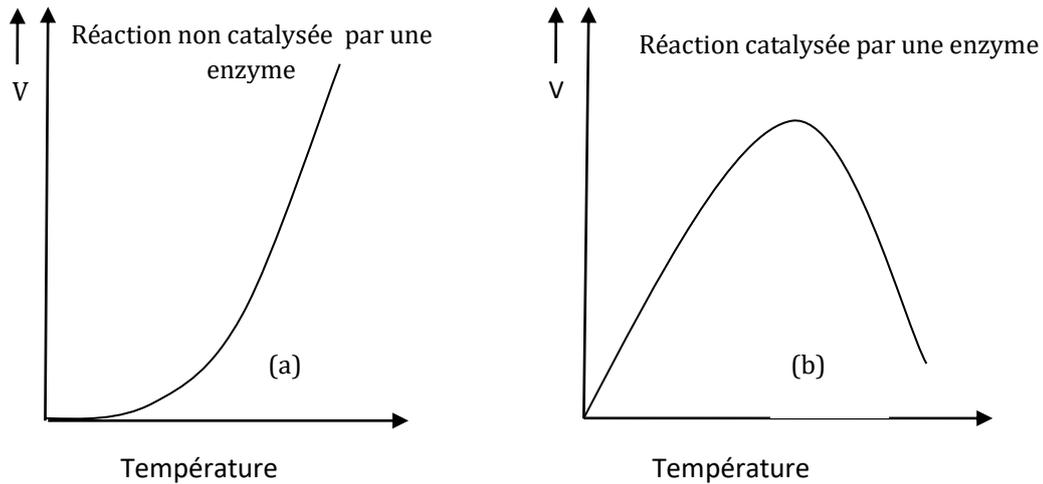
Il est connu que la vitesse de la plupart des réactions chimiques s'élève quand la température s'accroît. Les réactions enzymatiques n'échappent pas à cette règle.

L'effet de la température sur la vitesse est illustré dans la **figure 6** pour deux types de réactions avec et sans enzyme. Dans le cas de réaction enzymatique, la vitesse d'une réaction s'élève jusqu'à une valeur maximale, puis décroît rapidement, car la chaleur dénature l'enzyme, et la vitesse de sa dénaturation augmente avec la température.

Il existe une température optimale qui permet d'obtenir une vitesse constante pendant toute la durée de l'expérience. Il en résulte l'existence d'une température critique  $T_c$ , au-delà de laquelle il se produit le phénomène de dénaturation.

D'autre part, la résistance de l'enzyme vis-à-vis de la dénaturation thermique dépend d'autres paramètres tels que le pH ou la force ionique et de la présence de ligands. La fixation du substrat protège généralement l'enzyme.

Les enzymes de faible masse moléculaire portant une seule chaîne polypeptidique et des liaisons disulfures sont plus sensibles à la dénaturation que les enzymes à grande masse moléculaire. En général, elles sont plus stables dans l'extrait brut en présence d'autres protéines ou immobilisées à l'état insoluble.



**Figure 6 :** Effet de la température sur la vitesse d'une (a) réaction non catalysée et (b) une réaction enzymatique.

Il est donc nécessaire d'optimiser les conditions d'utilisation d'une enzyme dans un environnement donné notamment les changements de pH et de température.

### 3.1.3. Inhibition enzymatique

L'inhibition des enzymes signifie leur désactivation, par ralentissement ou arrêt total de l'activité enzymatique, sous l'action d'un composé chimique appelé inhibiteur.

L'inhibition enzymatique joue un rôle important dans la régulation des voies métaboliques. De nombreux composés chimiques bioactifs, comme des médicaments, des pesticides, etc., ne sont autres que des inhibiteurs d'enzymes ciblées.

L'action de l'inhibiteur a lieu par sa fixation à l'enzyme qui peut s'effectuer de deux manières :

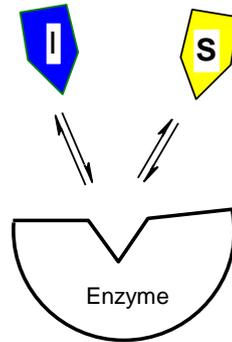
- sa liaison au site actif de l'enzyme, empêchant ainsi la fixation du substrat,
- sa liaison dans une autre région de l'enzyme provoquant la déformation de la protéine et donc sa désactivation partielle ou totale.

On distingue généralement les inhibiteurs *réversibles*, qui se lient à l'enzyme par des liaisons de faible énergie, et les inhibiteurs *irréversibles*, qui se fixent de manière covalente.

### 3.1.3.1. Inhibitions réversibles

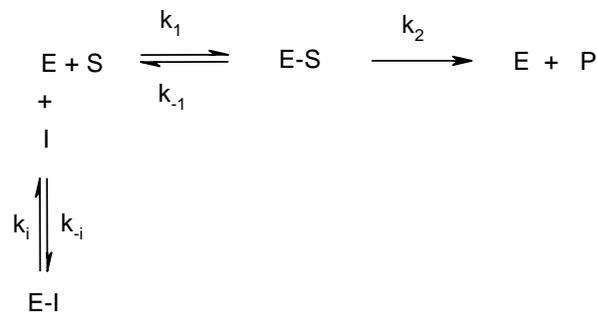
#### 3.1.3.1.1. Inhibition compétitive

Dans ce cas l'inhibiteur (I) se lie au site actif de l'enzyme et empêche le substrat (S) de s'y lier. Le substrat et l'inhibiteur sont donc en compétition pour se lier à l'enzyme :



**Figure 7** : Modèle d'inhibition compétitive.

En terme cinétique (michaélienne), cette inhibition est représentée par les équilibres suivants :



**Figure 8** : mécanisme d'inhibition compétitive.

Ces équilibres expriment la compétition du substrat et de l'inhibiteur pour l'enzyme E.

Par rapport aux équations citées plus haut ayant aboutie à l'établissement de l'équation de Michaelis-Menten, l'intervention de l'inhibiteur introduit les données suivantes :

$$[E]_0 = [E] + [E-S] + [E-I] \quad \text{et} \quad K_i = \frac{[E][I]}{[E-I]}$$

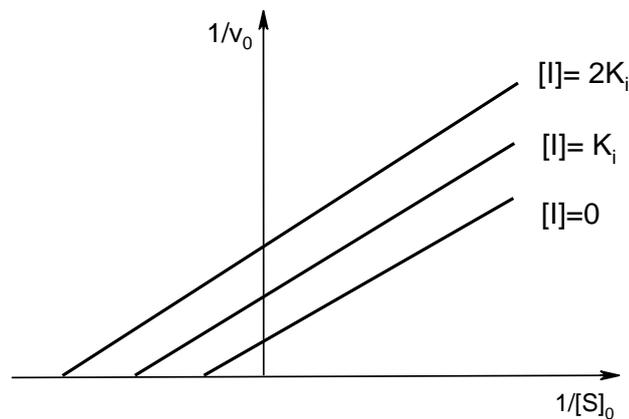


En tenant compte de ces égalités dans l'équation de Michaelis-Menten, la vitesse initiale de formation du produit P devient :

$$v_0 = \frac{v_m [S]_0 / (1 + \frac{[I]}{K_i})}{[S]_0 + K_m / (1 + \frac{[I]}{K_i})}$$

Dans ce type de compétition,  $V_m$  et  $K_m$  sont tous les deux diminués d'un facteur  $(1 + [I]/K_i)$ .

Le graphique de Lineweaver-Burk correspondant à cette inhibition est alors de la forme :

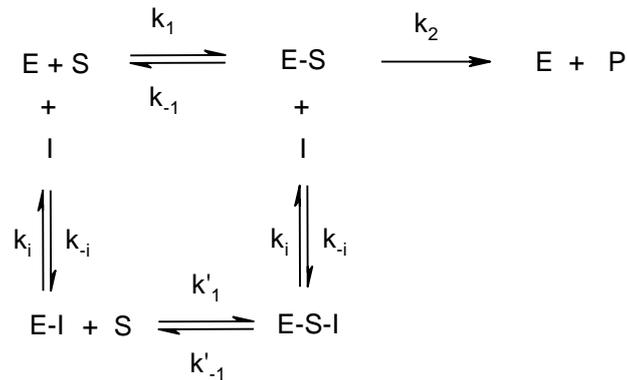


**Figure 11** : Graphique de Lineweaver-Burk d'inhibition incompétitive.

### 3.1.3.1.3. Inhibition non-compétitive

C'est la combinaison des deux précédents types de compétition. L'inhibiteur se lie à l'enzyme et au complexe enzyme-substrat.

Le système d'équations correspondant à cette situation est le suivant :

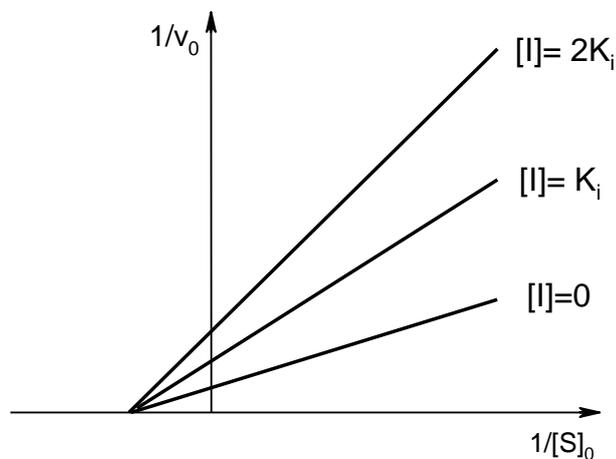


**Figure 12:** Mécanisme d'inhibition non compétitive.

L'équation exprimant la vitesse initiale de la formation du produit P est alors :

$$v_0 = \frac{v_m [S]_0}{[S]_0 + K_m \left(1 + \frac{[I]}{K_i}\right)}$$

Ce qui se traduit en représentation de Lineweaver-Burk comme suit :



**Figure 13 :** Graphique de Lineweaver-Burk d'inhibition non compétitive.



#### 4.1. Méthodes d'immobilisation<sup>10-11</sup>

Les principales méthodes d'immobilisation sont classées en quatre catégories<sup>9-12</sup> :

- Adsorption
- Liaison covalente
- réticulation
- inclusion

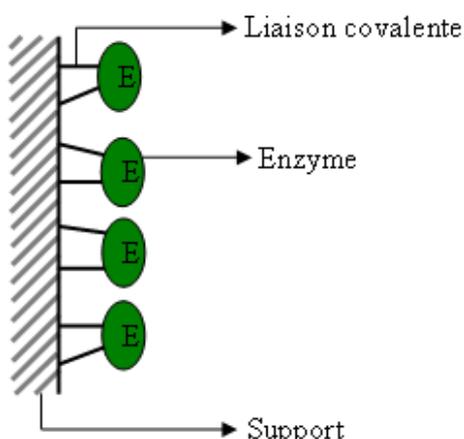
##### 4.1.1. Immobilisation par liaison covalente

L'immobilisation par liaison covalente conduit à la formation des liaisons très solides entre enzyme et support.

Le couplage covalent d'enzyme sur support nécessite la présence d'un groupe fonctionnel sur la surface de celui-ci (**Figure 15**). Ces groupements sont en général –COOH, –NH<sub>2</sub>, –OH, ou –SH ; ils sont peu réactifs chimiquement et il convient de les activer pour qu'ils réagissent dans des conditions douces avec des groupements fonctionnels de l'enzyme.

Différents supports tels que la silice<sup>13</sup>, Eupergit<sup>14</sup> ont été activés et exploités pour cette technique.

L'immobilisation par liaison covalente est utilisée afin d'augmenter la stabilité et l'activité de l'enzyme.<sup>15</sup>



**Figure 15** : Immobilisation par liaison covalente.

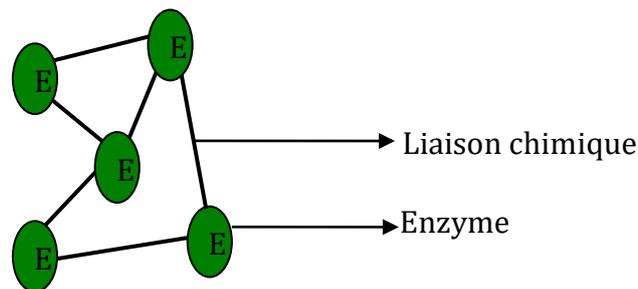
#### 4.1.2. Immobilisation par réticulation et co-réticulation

La réticulation consiste, à l'aide d'un agent réticulant, à créer des liaisons chimiques qui renforcent la cohésion de la membrane, **(Figure 16)**.

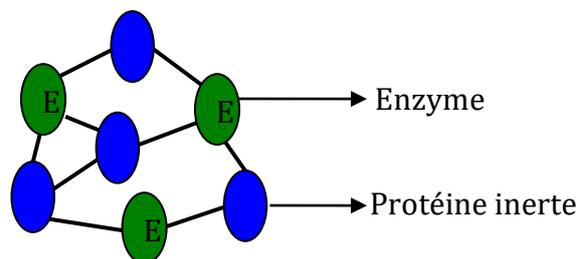
Quant à la co-réticulation, elle est basée sur le même principe que la réticulation en utilisant à la fois un agent réticulant plus souvent en présence du glutaraldéhyde et une protéine inerte pour faciliter ou améliorer la réticulation **(Figure 17)**.

L'utilisation de cette protéine permet, par une meilleure répartition des masses des différentes protéines, une meilleure maîtrise de l'activité enzymatique sans altérer les propriétés mécaniques des membranes obtenues.

Cependant, à cause de la réaction chimique, une perte d'activité enzymatique de la membrane co-réticulée peut être observée pendant les mesures.



**Figure 16** : Immobilisation par réticulation.



**Figure 17** : Immobilisation par co-réticulation.

### 4.1.3. Inclusion ou piégeage

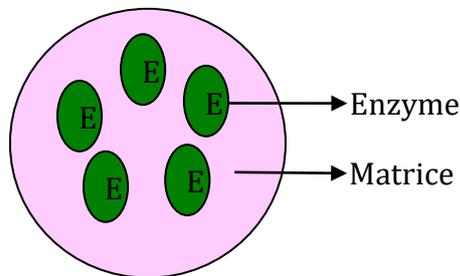
La technique consiste à emprisonner l'enzyme dans les mailles du réseau d'une matrice polymérique ou inorganique : gel de polyacrylamide, d'amidon, de gélatine, d'alginate.

Le procédé est très simple et universel, mais les risques de fuite ne sont pas négligeables si les mailles du réseau sont larges.

La maille de la matrice assure de manière purement physique la rétention de l'enzyme tout en permettant la diffusion du substrat jusqu'au site actif de l'enzyme grâce à une porosité du gel suffisante (**Figure 18**).

Toutefois cette technique est parfois limitée par la taille des pores du gel qui favorise le relargage des enzymes de faible poids moléculaire.

L'activité de l'enzyme est dépendante du micro-environnement local de l'enzyme immobilisé (pH, force ionique, diffusion moléculaire, etc.)



**Figure 18** : Immobilisation par Inclusion.

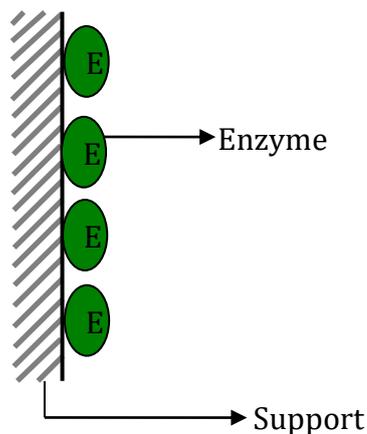
#### 4.1.4. Immobilisation par adsorption

L'adsorption est la méthode la plus économique et la plus simple d'immobilisation des enzymes<sup>16</sup>. Dans ce procédé, l'enzyme est retenue à la surface d'un corps adsorbant, minéral ou organique, grâce à des liaisons de Van der Waals et des interactions homophiles (hydrophobes et/ou hydrophiles) **(Figure 19)**.

Les supports adsorbants utilisés sont très variés, tant du point de vue de leur structure chimique que du point de vue de leurs propriétés physiques (densité, granulométrie, porosité, etc.).

Différents supports tel que chitosane<sup>17</sup>, kaolin<sup>18</sup>, céelite<sup>19</sup>, carbonate de calcium<sup>20</sup> ont été exploités pour l'immobilisation des enzymes.

Les quantités d'enzyme immobilisé par adsorption sont extrêmement variables et sont évidemment fonction de la nature du support.



**Figure 19** : Immobilisation par adsorption.

## 4.2. Avantages et désavantages des méthodes d'immobilisation

Le Tableau1 résume les avantages généraux et les inconvénients de chaque technique<sup>21</sup>

**Tableau 1** : Comparaison de différentes méthodes d'immobilisation des enzymes.

<b>Méthode d'immobilisation</b>	<b>Avantages</b>	<b>Inconvénients</b>
<b>Liaison covalente</b>	<ul style="list-style-type: none"> <li>• une large variété de supports disponible</li> <li>• stabilité à long terme</li> <li>• Liaison covalente forte entre l'enzyme et le support</li> </ul>	<ul style="list-style-type: none"> <li>• Modification chimique de l'enzyme</li> <li>• Procédure longue</li> <li>• Toujours couteuse</li> <li>• peut limiter le transfert de masse</li> </ul>
<b>Réticulation et co-réticulation</b>	<ul style="list-style-type: none"> <li>• haute activité et stabilité</li> <li>• efficace pour l'ensemble de cellules</li> <li>• liaison solide</li> </ul>	<ul style="list-style-type: none"> <li>• modification chimique de l'enzyme peut diminuer l'activité</li> <li>• peut limiter le transfert de masse</li> </ul>
<b>Inclusion ou piégeage</b>	<ul style="list-style-type: none"> <li>• Pas de modification chimique de l'enzyme</li> <li>• Peut être simple</li> <li>• efficace pour l'ensemble de cellules</li> </ul>	<ul style="list-style-type: none"> <li>• les modifications de l'environnement peuvent perturber le réseau et provoquer une fuite</li> <li>• peu ou pas de stabilisation</li> <li>• souvent limite le transfert de masse</li> </ul>
<b>Adsorption</b>	<ul style="list-style-type: none"> <li>• Simple</li> <li>• Facile à effectuer</li> <li>• Pas de modification chimique de l'enzyme</li> <li>• Moins couteuse</li> </ul>	<ul style="list-style-type: none"> <li>• Faible liaison entre l'enzyme et le support</li> <li>• fixation non spécifique</li> <li>• Risque de désorption</li> <li>• peut limiter le transfert de masse</li> </ul>

## 5. Domaines d'application des enzymes

Actuellement, les enzymes sont souvent utilisées pour un large éventail d'applications<sup>22</sup> telles que : l'industrie des détergents (35 % d'utilisation), fabrication de textile (10 %), l'industrie du cuir (3 %), l'industrie du papier; protection de l'environnement; production alimentaire, transformation et conservation des aliments (45 %) et applications médicales.

### 5.1 Applications des enzymes dans le domaine agro-alimentaire

Les enzymes sont utilisées depuis des siècles dans la production alimentaire<sup>23</sup>. Leur utilisation dans l'alimentaire ne se restreint pas à l'exploitation des propriétés catalytiques des enzymes, elle regroupe également les opérations visant à contrôler voire inhiber des enzymes pouvant altérer les qualités organoleptiques d'un aliment ou d'un constituant alimentaire. A titre d'exemple les brunissements enzymatiques observés dans certains aliments (exemple de la pomme de terre ou de la banane tranchée) dues à la formation des quinones issus de l'oxydation des polyphénols.<sup>24</sup>

#### 5.1.1 Raffinage enzymatique d'huiles alimentaires

Le dégomme enzymatique est la méthode la plus récente pour dégommer les huiles végétales, y compris l'huile de soja. Cette technique a initialement été développée par la compagnie allemande LurgiGmbH, elle est également connue sous le nom de procédé Enzymax®. Le but de ce procédé est de convertir grâce à **une phospholipase**<sup>23</sup>, les phospholipides non hydratés en une forme hydratée avec pour avantages un accroissement du rendement en huile, des coûts de fabrication réduits ainsi que la diminution des effluents.

#### 5.1.2 Enzymes utilisées dans la panification

De nombreuses fabriques de pain utilisent des émulsionnants chimiques à savoir les monoglycérides, afin de retarder le durcissement du pain<sup>25</sup>. On obtient avec l'amylase un pain à l'aspect plus frais qu'avec les agents chimiques. Cependant, le développement de nouvelles méthodes de production plus performantes et l'utilisation de nouvelles sources d'enzymes d'origine microbienne utilisées dans l'industrie agroalimentaire ont conduit à l'utilisation d'enzymes plus complexes.

Leurs domaines d'application sont très variés et on les retrouve non seulement en panification mais également en pâtisserie et en viennoiserie.

## 5.2 Applications analytiques et thérapeutiques

Les enzymes sont utilisées en médecine, dans les recherches scientifiques et dans l'industrie pharmaceutique.

### 5.2.1. Enzymes utilisées dans les méthodes d'analyses de biologie moléculaire

La Taq polymérase est une ADN polymérase utilisée pour l'amplification de l'ADN dans la PCR (polymerase chain reaction).<sup>26</sup> Cette enzyme est extraite de la bactérie *Thermus aquaticus*, un microorganisme thermophile vivant près des sources chaudes. La Taq polymérase peut résister à de très hautes températures, elle est utilisée dans la PCR car elle résiste à des températures où l'ADN est dénaturé : 95°C.

### 5.2.2. Enzymes utilisées en thérapie et analyse.

L'utilisation des enzymes dans le domaine médicale est de plus en plus fréquente, aussi bien d'un point de vue analytique que thérapeutique.<sup>27</sup>

Le tableau 2 regroupe les principales applications des enzymes dans le domaine thérapeutique et analyse.

**Tableau 2** : Enzymes utilisées dans le domaine thérapeutique

Enzyme	EC	Reaction	Appllication
Asparginase	3.5.1.1	L-Asparagine H <sub>2</sub> O → L-aspartate + NH <sub>3</sub>	Leukaemia
Collagenase	3.4.24.3	Collagen hydrolysis	Skin ulcers
Glutaminase	3.5.1.2	L-Glutamine H <sub>2</sub> O → L-glutamate + NH <sub>3</sub>	Leukaemia
Hyaluronidase	3.2.1.35	Hyaluronale hydrolysis	Heart attack
Lysozyme	3.2.1.17	Bacterial cell wall hydrolysis	Antibiotic
Rhodanase	2.8.1.1	S <sup>2</sup> O <sub>3</sub> <sup>2+</sup> + CN <sup>-</sup> → SO <sub>3</sub> <sup>2+</sup> + SCN <sup>-</sup>	Cyanide poisoning
Ribonuclease	3.1.26.4	RNA hydrolysis	Antiviral
B-Lactamase	3.5.2.6	Penicillin → penicilloate	Penicillin

Sterptokinase	3.4.22.10	Plasminogen → plasmin	Blood clots
Trypsine	3.4.21.4	Protein hydrolysis	Inflammation
Uricase	1.7.3.3	Urate + O <sup>2</sup> → allantoin	Gout
Urokinase	3.4.21.31	Plasminogen → plasmin	Blood clots

### 5.3. Applications industrielles

#### 5.3.1. Industrie du papier

L'utilisation de xylanases dans l'étape de blanchiment de la pâte kraft permet d'éviter le chlore et de réduire les déchets toxiques<sup>28</sup>, d'augmenter le degré de blanc de la pâte et de diminuer les coûts de l'opération, notamment dans les usines qui utilisent de grandes quantités de dioxyde de chlore.

Les lipases, les estérases, les pectinases, les hemicellulases, les cellulases et les enzymes lignitiques sont également utilisées dans le désencrage du papier ce qui permet le recyclage et la réutilisation du papier pollué dans la fabrication du carton.

#### 5.3.2. Industrie des détergents

L'industrie des détergents utilise le plus gros volume d'enzymes, c'est-à-dire 45 % du total du marché. Les enzymes utilisées dans ce secteur sont les protéases, bactériennes et fongiques, les amylases, les cellulases et les lipases.<sup>29</sup> Les enzymes les plus présentes sur le marché sont les protéases bactériennes, dont il existe actuellement une grande variété. Ces enzymes possèdent des propriétés nettoyantes croissantes et une grande stabilité face aux oxydants. Elles sont également actives dans l'élimination des particules de poussière et de terre car elles éliminent les microfibrilles des fibres de coton, ce qui produit de plus un effet de brillantage de la couleur. Il faut signaler que les enzymes utilisées dans les détergents ont un impact sur l'environnement positif. Elles entraînent en effet des économies d'énergie dues à la réduction des températures du lavage, permettent la réduction de la teneur en produits chimiques des détergents, et elles sont biodégradables.

## **6. Conclusion**

Dans ce chapitre nous avons présenté un aperçu général sur les enzymes, leur définition, leur mode d'action, la cinétique enzymatique ainsi que l'intérêt de leurs applications dans divers domaines industriels. Les enzymes dont le rôle est de catalyser les réactions chimiques du vivant sont privilégiées en industrie car elles permettent de contourner les inconvénients des produits chimiques et améliorent les relations coûts-efficacité des procédés.

## 7. Références

1. D.Voet, J. G. Voet, Biochimie, De Boeck Université, Bruxelles, **1998**.
2. A. Robert, Mechanism, and Data Analysis, **2000** by Wiley.
3. R.drapon, Station, L'Actualité Chimique, 1, **1973**.
4. F. Cramer, *Biochemical correctness*, 69(4), **1995**, 193–203.
5. R.U. Lemieux and U. Spohr, *Carbohydr. Chem.Biochem*, 50, **1994**, 1–20.
6. P. Csermely, R. Palotai, R. Nussinov, *Biochem. Sci*, 35, **2010**, 539–46.
7. D. E. Koshland, *Chem. Int. Ed. Engl*, 33, **1995**, 2375–8.
8. L. Michaelis, M. L. Menten, *Biochem. Z*, 49, **1913**, 333–369.
9. W.A.M. Alloue, M.Aguedo, J. Destain, H. Ghalfi, C. Blecker, J.P. Wathelet, P.Thonart, *Biotechnol .Agon.Soc.Environ*, 12(1), **2008**, 57-68.
- 10 Z. D. Knežević, S.S. Šiler-Marinković, Lj.V. Mojović, *APTEFF*, 35, **2004**, 151-164.
11. E. BROWN, J-F. BIELLMANN, Techniques de l'Ingénieur. Ttraité Génie des procédés. J 1 240.
12. P. J. Worsfold, *Pure &Appl.Chem*, 67(4), **1995**, 597-600.
13. Y.-X. Bai, Y.-F. Li, Y- Y. Liu-Xiang Yi, *Journal of biotechnology*, 125, **2006**, 574-582.
- 14., Z. Knezevic, N. Milosavic, D. Bezbradica, Z. Jakovljevic, R. Prodanovic, *Biochemical Engineering Journal*, 30, **2006**, 269–278.
15. M. H. Vermue, J. Tramper, *Pure and Appl.Chemistry*, 67(2), **1995**, 345-373.
16. S.K. Khare, M. Nakajima, *Food chemistry*, 68, **2000**, 153-157.
17. M.L. Foresti, *Enzyme and Microbial Technology*, 40(4), **2007**, 769-777.
18. M. B. Abdul Rahman, S.M. Tajudin, M. Z. Hussein, R. N. Z. Abdul Rahman, A. B. Salleh, M. Basri, *Applied Clay Science*, 29, **2005**, 111– 116.
19. W.Kaewthong, S. sirisaneeyakul, P. Prasertsan, A. H-Kittikun, *Process Biochemistry*, 40, **2005**, 1525-1530.
20. Y. Gargouri, H. Ghamgui, N. Miled, M. Karra-chaâbouni, *Biochemical Engineering Journal*, 37(1), **2007**, 34–41.
21. B. Chen, Immobilization of *Candida Antartica* lipase B on resine and Flats Films of effects of physical properties and surface chemistry of Matrice, thèse de doctorat, Polytechnic University, **2007**.

22. J-M. Choi, S-S. Han, H-S. Kim, *Biotechnology Advances*, 33(7), **2015**, 1443-1454.
23. H.E. SPINLER, Techniques de l'Ingénieur, Procédés chimie-bio-agro| Agroalimentaire, BIO650 V1, **2013**.
24. D. Rigal, (**2001**). Recherches sur l'inhibition du brunissement enzymatique : utilisation de préparations enzymatiques, substituts aux sulfites.
25. G. LINDEN, Techniques de l'Ingénieur. Ttraité Génie des procédés, F3700 V1, **1998**.
- 26.F. Uhel, L. Zafrani et Commission de la recherche translationnelle de la Société de réanimation de langue française, *Méd. Intensive Réa*, 28(6), **2019**, 464-472.
27. G. K. Meghwanshi, K. Navpreet, V. Swati, N. K. Dabi, Charan. V. A, Purohit. P. D, Praveen; H.S, Bhandari, N. Bhojak, K. Rajender, *Biotechnology and Applied Biochemistry*, **2020**.
28. G.Singh, N.Capalash, K.Kaur, S.Puri, P.Sharma, *Agro-Industrial Wastes as Feedstock for Enzyme Production*, **2016**, 157-172.
29. F.Hasan, A. Shah, S. Javed et Abdul Hameed, *African Journal of Biotechnology*, 9, **2010**, 4836-4844.

## **Chapitre II**

# **Méthodes de conception de médicaments assistées par ordinateur**

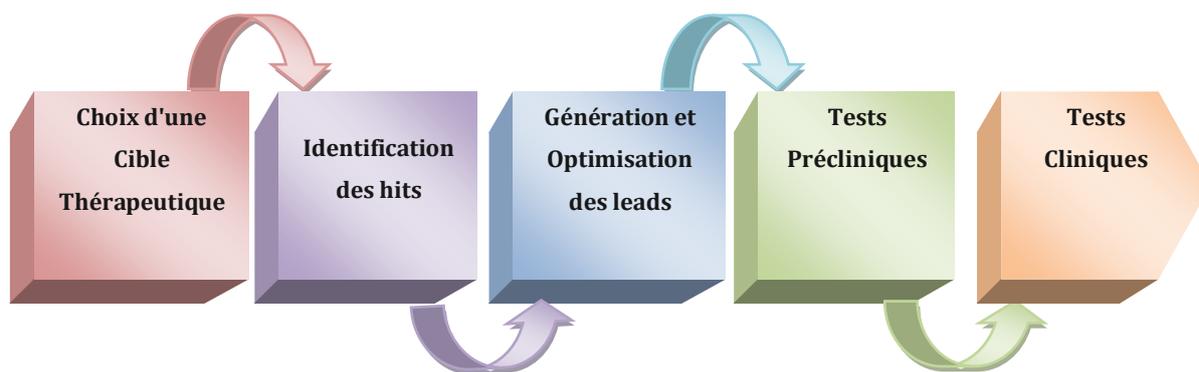
### 1. Introduction sur de la découverte des médicaments

Les maladies et les « essais thérapeutiques » remontent à l'existence de l'homme sur Terre. De la préhistoire jusqu'au début du XXe siècle, la grande majorité des médicaments provenaient de l'identification de principes actifs extraits de substances naturelles. Les premiers efforts sérieux étaient réalisés en vue d'isoler et de purifier les principes actifs de ces remèdes (c'est-à-dire les substances chimiques qui sont responsables de leur action pharmacologique. Dans les premiers temps, il n'y avait aucune possibilité de comprendre la nature de la maladie<sup>1</sup>.

Cependant, le début du XIXe siècle marque un tournant historique dans la recherche de nouveaux médicaments grâce à l'isolement de principes actifs de substances naturelles précédemment utilisées dans la médecine traditionnelle. En 1803, Friedrich Wilhelm isole la morphine à partir du pavot, qui a été commercialisée par les laboratoires Merck à partir de l'année 1827.<sup>2</sup> Cependant, la nécessité de développer d'autres sources de médicaments est rapidement apparue et a permis de nombreuses avancées tout au long du XXe siècle et jusqu'à aujourd'hui.

Plusieurs grandes disciplines ont permis la révolution de la recherche de nouveaux médicaments et ont abouti aux processus de R&D (Recherche et développement) mis en œuvre actuellement, les différentes étapes sont illustrées dans la Figure 1.<sup>3</sup>

Au cours des phases successives, des milliers de molécules doivent être triées et sélectionnées, afin d'obtenir un nombre très limité de candidats. Pour chaque nouveau médicament, le temps passé depuis la première étape, au cours de laquelle une cible biologique pertinente dans un processus pathologique donné est identifiée, jusqu'à la mise sur le marché d'un composé tourne autour d'une moyenne de 12 à 14 ans. Le processus de recherche et de développement de nouveaux médicaments est très long et coûteux.



**Figure 1:** Etapes du processus de recherche et développement des médicaments.

Les nouvelles méthodes permettant la découverte de nouveaux médicaments se doivent donc d'innover afin de mettre en évidence des molécules encore inconnues ayant un certain potentiel d'activité sur des cibles biologiques connues<sup>4</sup>. Les outils mis en place doivent être capables de guider les chimistes médicinaux dans le choix des molécules à cribler et à synthétiser<sup>1,5</sup>.

### 1.1. Le choix d'une cible thérapeutique

Le choix d'une cible thérapeutique revient à sélectionner, à l'aide de données issues de la littérature, une biomolécule (en général une protéine ou un complexe protéique) impliquée dans un certain processus pathologique.

Une cible thérapeutique peut être découverte (ou mise en évidence) à partir d'un mécanisme d'action de médicaments déjà connue, de ligands naturels, d'une compréhension des processus cellulaires et physiologiques, ou bien d'une façon aléatoire.

Avant de rechercher des molécules capables d'agir sur la cible biologique identifiée, il est nécessaire de procéder à la validation de cette cible. Cette validation consiste d'une part à s'assurer de l'effet bénéfique de la modulation de la cible sur la pathologie étudiée tout en vérifiant d'autre part que les conséquences de ces altérations ne seront pas néfastes. Ces paramètres régulent ce que l'on appelle la « druggabilité » de la cible<sup>7-9</sup>.

### 1.2. Identification des touches (*hits*)

La seconde étape consiste à identifier des touches (*hits*), c'est à dire des molécules ayant la capacité d'interagir avec la cible choisie et susceptibles de moduler ses effets

sur le processus biologique en question. Ces molécules sont sélectionnées à partir de bibliothèques de ligands naturels, ou de bibliothèques, ciblées ou non, le plus souvent à l'aide des techniques de criblage à haut débit (HTS pour High Throughput Screening).

Cependant, les tests expérimentaux, qui nécessitent le plus souvent une connaissance approfondie des systèmes étudiés, ne sont pas toujours aisés à mettre en œuvre, notamment à grande échelle. L'automatisation des procédures serait très coûteuse<sup>10</sup>. Ces difficultés ont un impact particulièrement important lorsque l'on se situe dans la phase d'identification des touches (*hits*) qui nécessite le criblage de milliers ou centaines de milliers de molécules.

Les stratégies de criblage virtuel, ou *in silico*, sont donc depuis quelques années employées en tant qu'alternative ou de façon complémentaire. Ces techniques sont en général assez faciles à mettre en place, pour un coût bien moindre que les criblages expérimentaux<sup>11</sup>. De plus, l'évolution technologique constante de ces dernières décennies a permis d'accélérer considérablement les temps de calculs nécessaires à la simulation de systèmes complexes ou de bases de données de plusieurs milliers de molécules. Le criblage virtuel est donc aujourd'hui employé dans de nombreux projets, afin de sélectionner, au sein de vastes bibliothèques de molécules, un nombre restreint de composés à cribler expérimentalement<sup>5</sup>.

### 1.3. Génération et optimisation de tête de série (leads)

À ce niveau, un certain nombre de touches (*hits*) vont être transformés en leads, qui vont ensuite permettre de créer des séries de composés ayant une activité démontrée sur la cible choisie.

Le choix porte sur des composés qui possèdent un motif structural de base sur lequel il est possible de greffer un certain nombre de substituants permettant de faire varier les propriétés des ligands formés. Cette étape est souvent considérée comme critique pour la sélection de composés de qualité<sup>12</sup>.

Les séries de leads sont ensuite optimisées d'une part afin de présenter une activité pour la cible la plus forte possible tout en conservant une spécificité suffisante vis-à-vis d'autres protéines<sup>13</sup>, mais également afin de répondre à un certain nombre de critères régissant les propriétés ADMETox et « *drug-like* » de la molécule, qui déterminent la capacité de celle-ci à être un bon médicament<sup>9,14,15</sup>.

Cette phase correspond à l'expression « *drug design* » (conception de médicaments) souvent utilisée a tort comme un équivalent de « *drug discovery* » (découverte de médicaments).

### 1.4. Les tests précliniques et cliniques

Après avoir trouvé et défini un nouveau candidat médicament, une série de tests liés à l'innocuité, la posologie, la distribution, aux paramètres pharmacologiques et toxicologiques doit être effectuée.

#### 1.4.1 Tests précliniques

La phase de recherche préclinique est une étape fondamentale dans la fabrication d'un médicament. Elle permet de comprendre le mode d'action du médicament et de tester l'efficacité et la toxicité des molécules sur les animaux avant son administration à l'homme.

Les études permettent non seulement de déterminer le profil de sécurité chez l'animal, mais elles fournissent également des informations importantes sur :

- L'absorption (façon dont le médicament pénètre dans l'organisme)
- La distribution (façon dont le médicament circule dans l'organisme)
- Le métabolisme (décomposition du médicament par l'organisme)
- L'excrétion (élimination du médicament par l'organisme)

Ces quatre facteurs sont regroupés sous l'acronyme ADME.

A l'issue de l'ensemble de ces études précliniques, et au cas où le produit semble prometteur et non toxique, sera alors décidé le passage aux essais cliniques chez l'homme.

#### 1.4.2. Tests cliniques

Une étude clinique est une étude scientifique réalisée en thérapeutique médicale humaine pour évaluer l'efficacité et la tolérance d'un médicament.

### 1.4.2.1. Les essais de la phase I

Les essais de phase I permettent une première évaluation de la tolérance du médicament en fonction de sa dose. Le candidat-médicament est administré à l'homme pour la première fois. L'objectif premier est d'évaluer la sécurité du produit et de déterminer l'échelle des doses tolérées. La molécule est testée sur une courte période, afin de déterminer la dose minimale active, les paramètres pharmacocinétiques et le déclenchement d'effets indésirables.

### 1.4.2.2. Les essais de la phase II

Les essais de phase II concernent des malades atteints de l'affection que le produit est destiné à soulager. Différentes quantités du médicament sont administrées à environ 100 personnes qui souffrent du trouble cible, afin de confirmer l'efficacité thérapeutique de l'indication envisagée, et à affiner la posologie. Dans les études de phase II on tente également de mettre en évidence des effets indésirables à court terme, ainsi que des facteurs capables de modifier la cinétique du produit.

### 1.4.2.3. Les essais de la phase III

Les essais thérapeutiques proprement dits commencent avec la phase III. Ils portent sur un nombre beaucoup plus étendu de malades, le produit est administré selon un protocole précis, les résultats obtenus chez un groupe traité sont comparés à ceux d'un groupe témoin. L'objectif de cette étude est de déterminer le profil thérapeutique et le devenir pharmacocinétique du produit, c'est une phase décisive de l'Autorisation de Mise sur le Marché (AMM).<sup>9</sup>

### 1.4.2.4. Les essais de la phase IV

Cette phase commence après la commercialisation du médicament, et concerne les études à très grandes échelles. Les essais de la phase IV ont pour but d'affiner les indications thérapeutiques, établir l'innocuité du médicament, cibler des populations particulières et vérifier l'absence d'interactions médicamenteuses.

## 2. Le principe du criblage virtuel (*in silico*)

### 2.1. Généralités

Le criblage virtuel est la stratégie *in silico* la plus utilisée pour l'identification de touches (*hits*) dans le cadre de la recherche de nouveaux médicaments<sup>13,16,17</sup>. Celui-ci fait désormais partie intégrante de la plupart des programmes de recherche de composés bioactifs, que ceux-ci se déroulent en milieu académique ou industriel, car il constitue un complément essentiel au criblage biologique haut-débit<sup>18-20</sup>. Le criblage virtuel permet l'exploration de larges chimiothèques (> 10<sup>6</sup> molécules) à la recherche de composés actifs vis-à-vis d'une cible thérapeutique donnée. Ce processus vise à réduire de façon significative la chimiothèque de départ à une liste limitée de composés jugés les plus prometteurs. Cette approche conduit souvent à une nette amélioration de la concentration de molécules actives pour la cible, tandis qu'une sélection aléatoire de molécules de la chimiothèque ne saurait fournir un tel enrichissement. Ainsi, le temps aussi bien que les coûts de l'identification de nouvelles touches peuvent être réduits de façon significative<sup>18</sup>.

### 2.2. Préparation des chimiothèques

Ces bases de données peuvent être seulement des collections de structure (2D, 3D et autres formats) ou fournir des données supplémentaires telles que la bioactivité mesurée des composés et leurs cibles protéiques ainsi que les maladies ciblées, mais très peu de ces bases de données tentent de lier les données sur les petites molécules à des informations sur leurs caractéristiques biologiques, leurs cibles, ainsi que leurs données d'essai disponibles.

L'importance de la chimiothèque employée est la première condition pour le succès d'un criblage virtuel, bien avant celle des algorithmes utilisés pour la recherche de touches au sein de la chimiothèque<sup>6</sup>. En effet, seule une librairie de composés suffisamment diverse peut garantir une exploration satisfaisante de l'espace chimique maximisant ainsi les chances de découvrir de nouvelles touches<sup>21</sup>.

**Tableau 1:** Bases de données des petites molécules les plus connues et les plus importantes.

Chimiothèque	Type de Licence	Éditeur
NCI Open database	Disponible publiquement	National Cancer Institute
MDDR	Commerciale	Accelrys Inc
Binding DB	Disponible publiquement	University of Maryland USA
Relibase	Commerciale	Cambridge crystallographic data center
Chem spider	Disponible publiquement	Royal society of chemistry ,UK
DrugBank	Disponible publiquement	University of Alberta Canada
PubChem	Disponible publiquement	National center for biotechnology information
Therapeutic Target Database	Disponible publiquement	National University of Singapore, Singapore
ZINC	Disponible publiquement	University of California San Francisco, USA
ChEMBL	Disponible publiquement	European bioinformatics Institute, KU
World Drug Index	Commerciale	Thomson Reuters

### 2.3. Préfiltrage de la chimiothèque

Pour éviter de perdre du temps avec des molécules possédant des caractéristiques incompatibles avec celles de composés d'intérêt pharmaceutique, le processus de criblage comporte généralement une étape préliminaire de filtrage. Plusieurs critères de base sont pris en compte tels que le faible coût de production, la stabilité mais aussi les propriétés pharmacologiques, d'absorptions, de distributions, de métabolismes, d'excrétions, et de toxicités (ADME-Tox)<sup>22</sup>.

Cette tâche, qui peut être prise en charge par des programmes spécialisés, consiste à exclure les composés toxiques ou supposés tels ou comportant des groupements jugés trop réactifs. Ensuite, ne sont retenus que les composés obéissant à des définitions empiriques simples du profil de molécule active (caractère dit « *drug-like* »), telle que la populaire « règle des 5 » de Lipinski<sup>14</sup> et ses extensions.

Durant leur phase d'optimisation, la complexité, le poids moléculaire et le caractère lipophile des touches ont généralement tendance à augmenter. Par conséquent, des filtres physico-chimiques additionnels peuvent être employés pour restreindre la chimiothèque aux composés ayant un profil de tête de série, dits « *lead-like* ».<sup>12</sup> On estime alors que les risques d'échec aux tests cliniques sont minimisés, ce

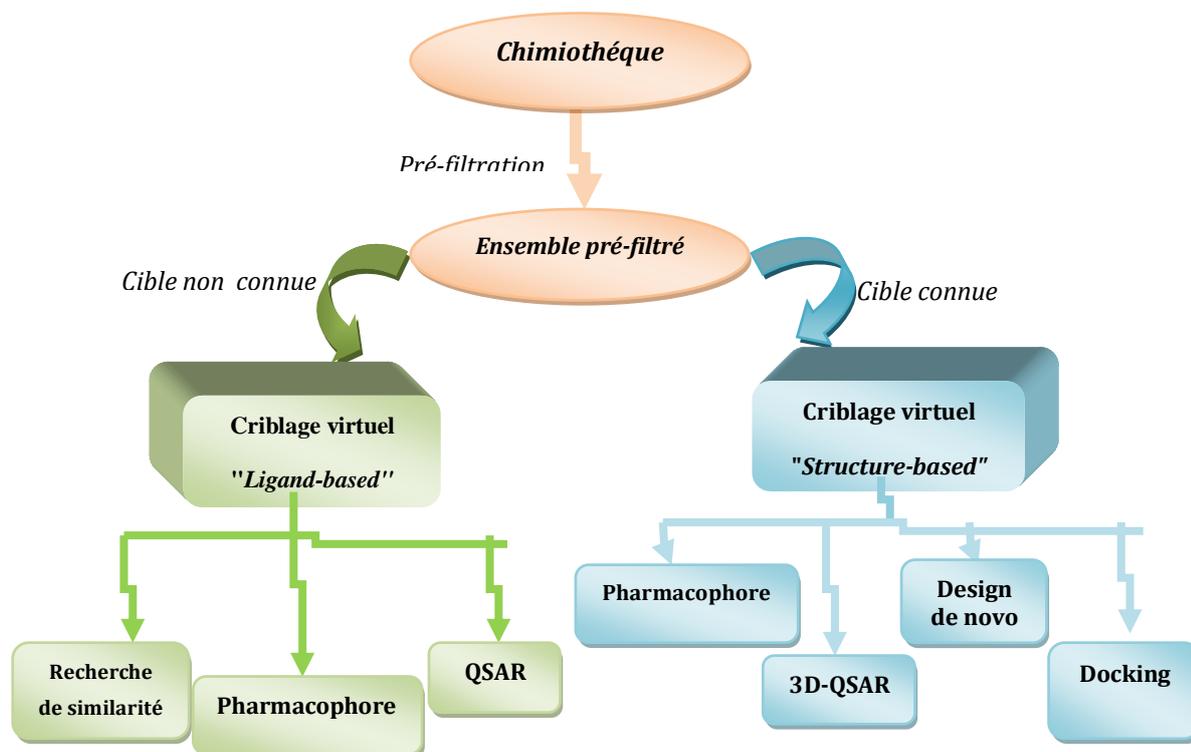
qui est souvent un impératif économique. Toutefois, appliquées de façon stricte, ces règles de pré-filtrage peuvent assez facilement conduire à l'exclusion de molécules d'intérêt.

Ainsi, de nombreux médicaments commercialisés ne respectent pas les règles de Lipinski, c'est pourquoi ces règles empiriques sont souvent employées de façon plus ou moins permissive.

### 3. Les différentes méthodes du criblage virtuel

Selon la nature de l'information expérimentale disponible, on distingue la méthode de criblage *in silico* en deux approches, la première se base sur la structure de la cible afin d'identifier les interactions possibles au sein du site de liaison et ainsi sélectionner des molécules capables de s'y arrimer et est connue sous le nom de « *structure-based virtual screening* »<sup>23,24</sup>. La seconde utilise les ligands connus comme guides pour définir les propriétés responsables de leur affinité pour la cible et identifier des molécules similaires, est appelée « *ligand-based virtual screening* ». <sup>18,25</sup>

Bien que ces deux approches soient surtout utilisées de manière exclusive (souvent parce que la nature des données de départ ne laisse qu'un seul choix possible), leur combinaison lors d'une campagne de criblage permet de maximiser les chances de succès pour identifier de nouvelles touches<sup>26</sup>.



### Figures 2 : Classification des méthodes de criblage virtuel.

Dans ce chapitre nous nous limiterons à décrire en détail les deux méthodes utilisées dans le travail de cette thèse :

- Les méthodes de la relation quantitative structure-activité (QSAR) en tant que méthode basée sur la structure du ligand et,
- Le docking moléculaire en tant que méthode basée sur la structure de la cible (enzyme, récepteur, ...).

#### 3.1. Méthodes « *ligand-based virtual screening* »

Lorsqu'au moins un ligand de la cible étudiée est connu, un criblage virtuel basé sur les ligands peut être mis en œuvre. Le principe de base commun à toutes les méthodes basées sur les ligands est que des molécules similaires vont avoir tendance à présenter des profils d'activité similaires<sup>27</sup>. La similarité des molécules peut se mesurer par recherche de propriétés communes, qui seront utilisées comme descripteurs de similarité. En fonction du nombre de ligands de référence pour la cible et du type de descripteurs, différentes méthodes peuvent être employées : la recherche de similarité, le criblage à l'aide de pharmacophore et les méthodes QSAR.

##### 3.1.1. Recherche de similarité

La recherche de similarité est la méthode à employer lorsque très peu de ligands ont été rapportés pour la cible biologique choisie.

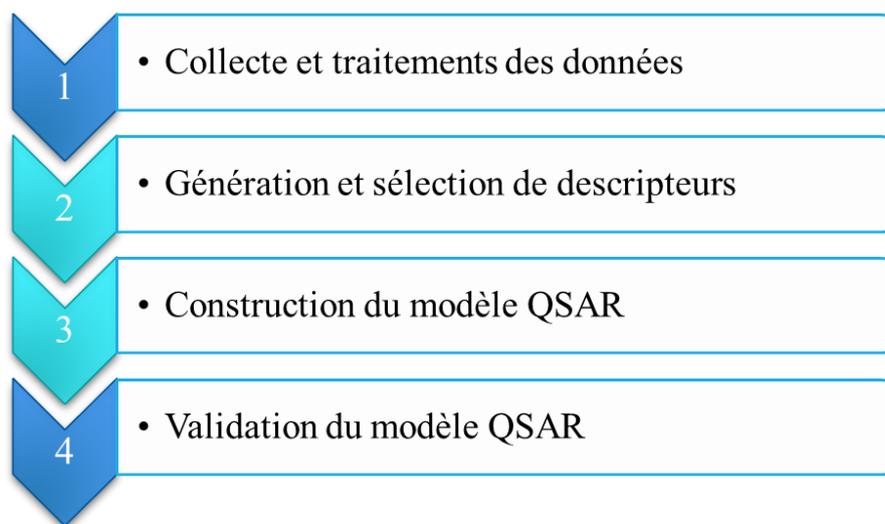
Cette méthode permet de définir au sein d'une base de données quels ligands sont plus ressemblants des ligands actifs connus expérimentalement. Elle repose donc sur l'utilisation de descripteurs et de métriques de similarité (coefficients de distance, d'association et de corrélation) permettant de comparer des molécules à cribler à un ou plusieurs ligands de référence pour prédire leur profil d'activité.

##### 3.1.2. Développement de Modèles des relations quantitatives structure-actives (QSAR)

Les modèles QSAR représentent des équations mathématiques corrélant la réponse des composés chimiques (activité) avec leurs informations structurales et physico-chimiques sous la forme de quantités numériques, c'est-à-dire les descripteurs moléculaires. Des méthodes statistiques appropriées sont étendues pour

développer un modèle robuste et riche en informations impliquant un grand nombre de variables.

En général, pour prédire une activité biologique ou une propriété par la méthode QSAR (QSPR), la construction du modèle mathématique est effectuée en différentes étapes principales **Figure 3**.



**Figure 3** : Etapes effectuées pour le développement d'un modèle QSAR.

L'ensemble des données (l'ensemble des molécules) utilisées dans l'étude de QSAR sont collectées soit directement à partir des mesures expérimentales, extraites de la littérature ou à partir de bases de données.

L'ensemble des données collectées est divisé en deux sous-ensembles :

- Ensemble d'apprentissage « *Training Set* » : utilisé pour la construction du modèle.
- Ensemble de test « *Test Set* » : utilisé pour valider et mesurer la qualité des modèles formés.

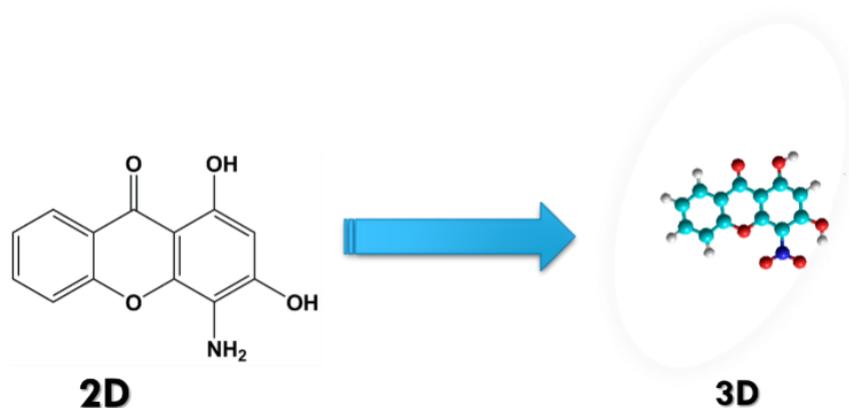
### 3.1.2.1. Collecte et traitement de l'ensemble des données

Les modèles QSAR sont très dépendant des données expérimentales utilisées pour leur construction. On doit donc tenir compte des données à modéliser ainsi qu'au choix de l'ensemble de données qui représente une étape très importante dans la construction des modèles QSAR.

Les données devraient, idéalement, être de grande qualité, fiables et homogènes, présentant des incertitudes faibles afin de limiter les barres d'erreurs expérimentales. On doit s'assurer que les tests menés sont effectués selon le même protocole d'un seul ou plusieurs laboratoires, car les conditions expérimentales ont, généralement, une forte influence sur les résultats obtenus.

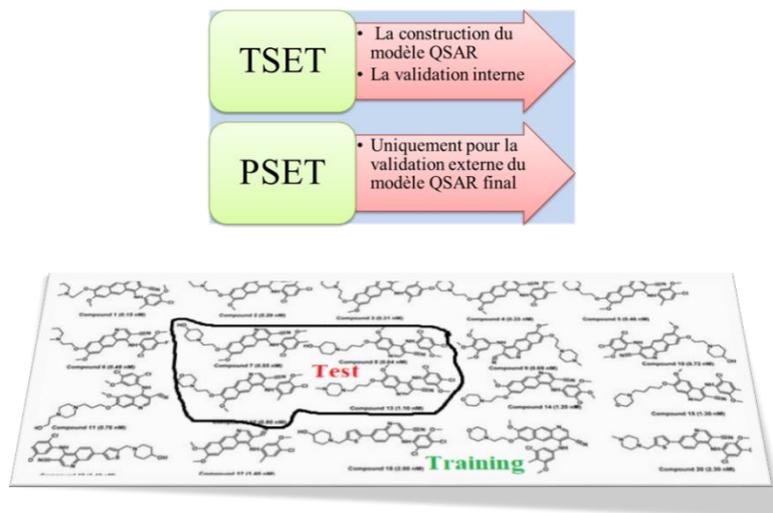
L'efficacité d'un modèle QSAR dépend également du type de molécules qui y sont incluses, plus le modèle présentera des composés de structures proches et similaires, plus il aura de chance d'être performant. La présence de molécules de comportement déviant (bizarre) peut influencer de manière importante le modèle final.

Les structures moléculaires sont généralement téléchargées à partir des bases de données telles que ZINC database, PubChem ou tirées à partir des travaux disponibles dans la littérature puis dessinées à l'aide des logiciels. Une optimisation et minimisation d'énergie par des méthodes empiriques ou quantiques est nécessaire pour les calculs 3D. Plusieurs logiciels peuvent être employés pour le dessin et l'optimisation tels que HyperChem, ChemDraw et Gaussian.



**Figure 4 :** Dessin et optimisation des structures moléculaires.

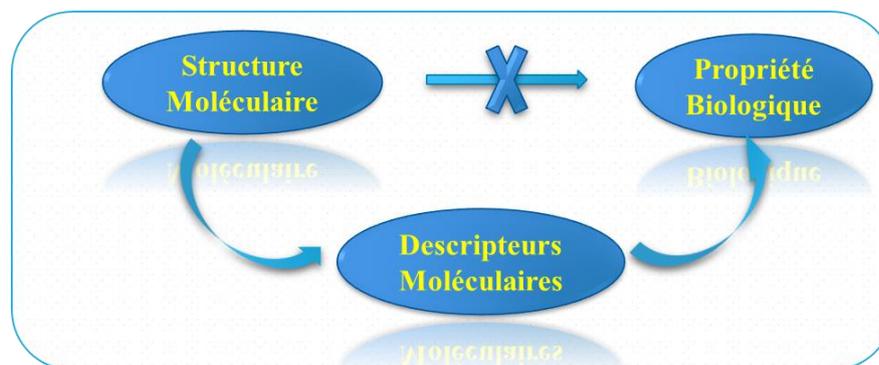
L'ensemble des molécules à modéliser doit être divisé en deux sous-ensembles : un sous ensemble d'apprentissage constitué de 70 à 80% des molécules, utilisé pour la construction du modèle QSAR et un sous-ensemble de test d'environ 20 à 30% des molécules réservées pour la validation du modèle construit. <sup>28</sup>



**Figure 5 :** Division des molécules collectées.

### 3.1.2.2. Utilisation des descripteurs moléculaires

L'activité biologique ne peut pas être reliée directement avec la structure moléculaire, cette dernière est codée par des grandeurs qui représentent d'une manière quantitative les informations contenues dans la structure moléculaire telles que les caractéristiques physico-chimiques et structurales. Ces grandeurs sont appelées descripteurs.



**Figure 6 :** Génération de descripteurs moléculaires indispensables pour établir la relation QSAR.

On en dénombre aujourd'hui des milliers de descripteurs, qui peuvent être calculés ou obtenus de manière empirique. De nombreux logiciels ont été développés pour calculer les différents descripteurs moléculaires tels que : l'OASIS, ADAPT, CODESSA, DRAGON, etc.

Dans ce travail de thèse nous avons utilisé le logiciel DRAGON.6.0 qui est capable de générer plus de 4800 descripteurs moléculaires qui sont répartis en 29 blocs logiques (**Tableau.2**).

**Tableau.2** : liste des blocs de descripteurs fournis par le logiciel Dragon.6.0

N°	Nom du bloc	N°	Nom du bloc
1	Constitutional descriptors	16	RDF descriptors
2	Ring descriptors	17	3D-MoRSE descriptors
3	Topological indices	18	WHIM descriptors
4	Walk and pathcounts	19	GETAWAY descriptors
5	Connectivity indices	20	Randicmolecular profiles
6	Information indices	21	Functional group counts
7	2D matrix-baseddescriptors	22	Atom-centred fragments
8	2D autocorrelations	23	Atom-type E-state indices
9	Burdeneigenvalues	24	CATS 2D
10	P_VSA-likedescriptors	25	2D Atom Pairs
11	ETA indices	26	3D Atom Pairs
12	Edgeadjacency indices	27	Charge descriptors
13	Geometricaldescriptors	28	Molecular properties
14	3D matrix-baseddescriptors	29	Drug-like indices
15	3D autocorrelations		

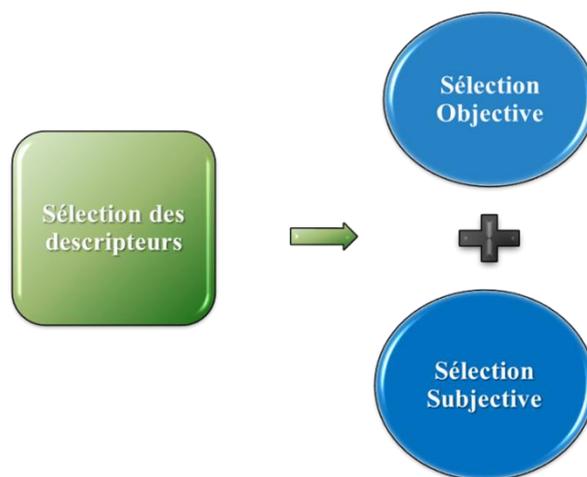
---

On peut calculer, avec Dragon, non seulement les types les plus simples de descripteurs (constitutionnels), mais aussi plusieurs descripteurs topologiques et géométriques tels que : le volume et la surface moléculaire et d'autres propriétés comme l'alerte de *LIPINSKI* (appelée aussi : la règle de cinq) et certains indices de drogues utilisés dans la sélection de composés pour le criblage et /ou de la conception de médicaments.

### 3.1.2.3. Sélection des descripteurs

Lorsqu'un grand nombre de descripteurs est introduit, certains d'entre eux peuvent contenir des informations redondantes, entraînant un problème de colinéarité. De plus les descripteurs calculés n'ont pas nécessairement une influence sur l'activité à modéliser, il est donc nécessaire d'éliminer les descripteurs dont l'influence est inférieure à celle du bruit, et de sélectionner uniquement les plus pertinents d'entre eux. De manière générale, un descripteur est retenu lorsqu'on constate que son retrait entraîne une décroissance significative de la performance du modèle.<sup>29-35</sup>

La procédure de sélection et de réduction des variables (descripteurs) peut être effectuée en deux étapes : sélection objective et sélection subjective.



**Figure 7** : procédures de sélection des descripteurs moléculaires.

- **Sélection objective**

La première étape de cette procédure consiste à exclure tous les descripteurs ayant un pourcentage élevé de valeurs identiques pour l'ensemble des composés (variance non significative). Cela permet de s'assurer que de tels descripteurs ne sont pas inclus par chance dans le modèle final. De même, lorsque deux descripteurs sont fortement corrélés et leur combinaison possède un coefficient de détermination supérieur au seuil requis ( $R^2 > 0,90$ ), seul celui présentant la plus grande variance est retenu. Ces procédures évitent l'introduction de descripteurs inappropriés et elles rendent la suite de l'analyse moins coûteuse en terme de temps de calcul.

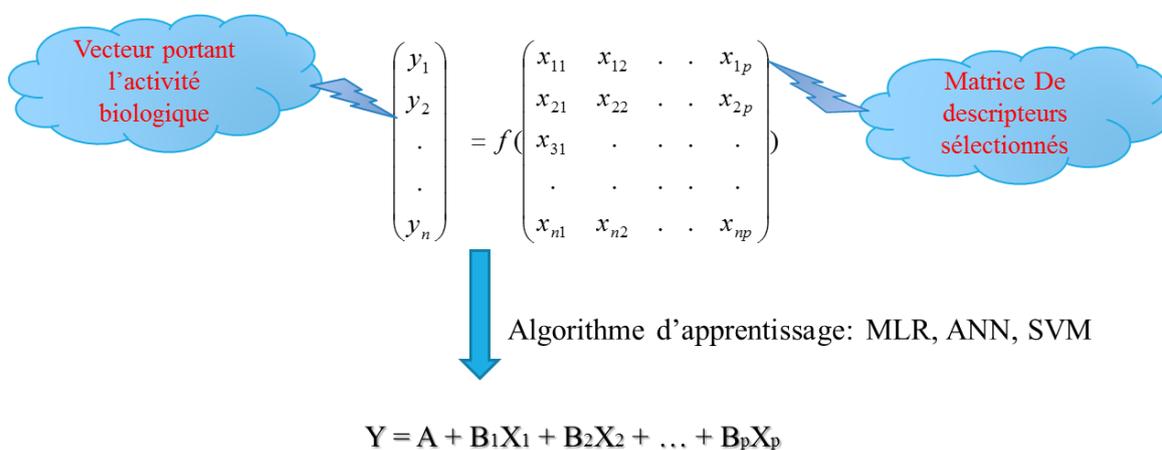
- **Sélection subjective**

Cette méthode fait participer la variable dépendante (l'activité biologique) pour la sélection et la recherche des meilleurs sous-ensembles de descripteurs en utilisant des méthodes statistiques et informatiques tels que : les méthodes informatiques (les algorithmes génétiques et réseaux de neurones) et les méthodes statistiques (la sélection pas à pas).

### 3.1.2.4. Construction du modèle QSAR

L'objectif d'une étude QSAR est d'établir un modèle mathématique permettant de relier de manière quantitative ou qualitative une activité avec la structure moléculaire<sup>36</sup>.

Il existe plusieurs méthodes d'apprentissage statistiques et heuristiques utilisées pour optimiser et modéliser les activités biologiques ou les propriétés physico-chimiques, on trouve à titre d'exemple : la régression linéaire multiple (MLR), régression des moindres carrés partiels (PLS) et la régression par les machines à vecteur de support (SVM) et réseaux de neurones.



**Figure 8** : construction du modèle mathématique.

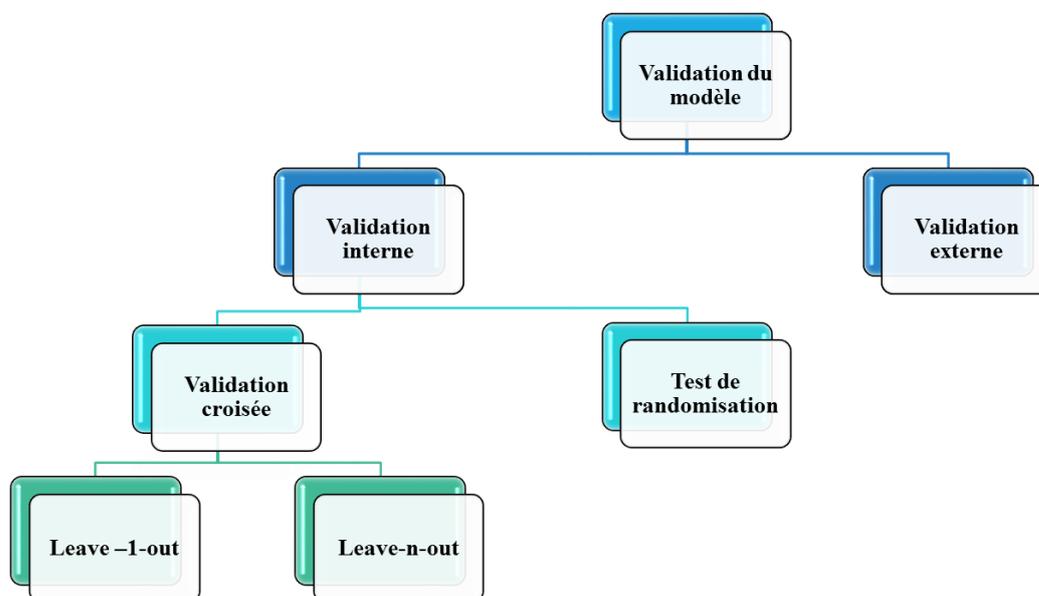
La régression linéaire multiple est la plus simple méthode statistique de modélisation et la plus appliquée dans les études de la relation structure-activité. Elle repose sur l'hypothèse qu'il existe une relation linéaire entre l'activité biologique et les descripteurs moléculaires (eq.02).

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad eq.02$$

Où  $a_i$  ( $i = 0, 1, \dots, n$ ) sont les coefficients de la régression.

### 3.1.2.5. Validation du modèle

La modélisation vise à fournir un modèle qui soit non seulement ajusté aux données expérimentales, mais aussi d'être généralisé sur de nouveaux exemples<sup>36</sup>. Pour ce faire, plusieurs méthodes de validation, telles que la validation interne, la validation externe et le test de randomisation, sont employés pour estimer la fiabilité du modèle QSAR et pour déterminer sa pertinence pour une application donnée. Ainsi, différents indicateurs statistiques sont employés pour déterminer la qualité d'un modèle QSAR, nous citerons ci-dessous les plus répandus entre eux.



**Figure 9 :** méthodes de validation du modèle QSAR.

#### 3.1.2.5.1. Validation interne

La validation interne est souvent la technique la plus employée, dans les études QSAR pour déterminer la stabilité du modèle et tester l'influence de chaque échantillon de l'ensemble d'apprentissage sur le modèle final. Pour ce faire, on emploie les techniques de la validation croisée (cross validation CV) leave-one-out ou leave-n-out ainsi que le test de randomisation.<sup>37, 38</sup>

La performance des modèles de régression est estimée à l'aide des paramètres statistiques de la validation croisée décrits par les équations ci-dessous<sup>39</sup>:

L'écart type de la validation croisée  $S_{CV}$  :

$$S_{CV} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_{icv})^2} \quad eq. 15$$

Le coefficient de détermination de la validation croisée  $Q_{CV}^2$  :

$$Q_{CV}^2 = \frac{\sum_{i=1}^N (y_i - \bar{y})^2 - \sum_{i=1}^N (y_i - y_{icv})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad eq. 16$$

Où

$y_{icv}$  : Représente la valeur de l'activité de la molécule  $i$  calculée par la méthode de la validation croisée,

$\bar{y}$  : La valeur moyenne de l'activité,

$y_i$  : Les valeurs observées de l'activité.

Le test de randomisation est utilisé par le modélisateur pour affirmer que les bonnes corrélations, entre les descripteurs et l'activité, présentées par le modèle QSAR ne sont pas dues à la chance. Pour se faire, les observations sont aléatoirement désorganisées, par exemple dix fois, en changeant aléatoirement la colonne des activités, mais les colonnes des descripteurs restent inchangées. Si la randomisation des observations conduit à des modèles de prévisions faibles cela signifie que les capacités prédictives du modèle QSAR construit ne sont pas dues aux corrélations de chance.

### 3.1.2.5.2. Validation externe

Afin de tester de manière fiable le pouvoir prédictif du modèle QSAR, il est nécessaire d'employer un ensemble de validation externe, non employé pour le développement du modèle. Il suffit d'appliquer le modèle QSAR aux molécules qui le composent et de déterminer la corrélation existant entre les activités calculées et celles expérimentales. Plus cette corrélation est importante, plus le modèle est capable de prédire les activités pour des molécules hors l'ensemble d'apprentissage<sup>40</sup>.

$$Q_{ext}^2 = 1 - \frac{\sum_{i=1}^N (y_i - y_{iext})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad eq. 17$$

### 3.2. Méthodes « *Structure-based virtual screening* »

Les méthodes « *structure-based* » utilisent la structure expérimentale tridimensionnelle de la protéine pour prédire quelles molécules vont s'y lier.

Différentes approches peuvent être employées pour réaliser ce criblage : la construction de modèles pharmacophoriques, l'établissement de modèles 3D-QSAR, la conception de novo « *de novo design* » et les méthodes de docking moléculaire qui sont les plus populaires.

#### 3.2.1. Méthodes de docking moléculaire

La reconnaissance moléculaire est la capacité des biomolécules à reconnaître d'autres molécules et à interagir sélectivement avec elles afin de promouvoir des événements biologiques fondamentaux tels que la transduction du signal, le transport, la régulation, la catalyse enzymatique...etc.

Le docking moléculaire est le processus qui consiste à placer des molécules dans des configurations appropriées pour interagir avec un récepteur.

Dans la modélisation moléculaire, le terme «docking moléculaire» fait référence à l'étude de la façon dont deux ou plusieurs structures moléculaires interagissent entre elles.

Le docking moléculaire *in silico* a pour but de prédire la structure d'un complexe moléculaire à partir des molécules isolées, ce qui est considérablement plus facile à mettre en œuvre, moins cher et plus rapide que l'utilisation des méthodes expérimentales.

Les logiciels de docking sont donc des outils très utilisés en chimie, biologie, pharmacie et en médecine car la plupart des principes actifs sont des petites molécules (ligands) qui interagissent avec une cible biologique d'intérêt thérapeutique, généralement protéique (récepteur, enzyme,...ect). Ces méthodes de docking visent à prévoir la capacité ou non d'une molécule à se lier au site actif d'une protéine en se basant pour cela sur la prédiction de la conformation, de la position et de l'orientation de la molécule lors de sa liaison au site du récepteur.

A cet effet, les méthodes de docking combinent l'utilisation d'un algorithme de recherche conformationnel, permettant de générer des modes de liaisons du ligand dans le site du récepteur, ou « poses », et d'une fonction de score, employée pour classer les différentes poses selon un score prédit d'affinité<sup>41</sup>. Les méthodes de

docking s'attachent donc, d'une part, à identifier les molécules qui sont des ligands véritables du récepteur parmi toutes celles étudiées, mais aussi d'autre part, à déterminer les poses correctes soit les conformations adoptées par les ligands lors de la liaison au récepteur.

### 3.2.1.1. Flexibilité moléculaire dans le docking des protéines-ligands

L'adaptation mutuelle d'un ligand avec son récepteur est cruciale pour comprendre la liaison du ligand et la fonction des protéines. L'un des défis majeurs du docking moléculaire est de savoir comment prendre en compte cette adaptation dans les calculs de docking.

Par ordre croissant de complexité, le docking moléculaire peut être classé selon la flexibilité comme suit :

1. Docking de corps rigide : ignore la flexibilité des molécules et les traite comme des objets rigides
2. Récepteur rigide - docking de ligand flexible: seul le ligand est traité comme flexible, le récepteur est rigide
3. Récepteur flexible - docking de ligand souple: la protéine et le ligand sont traités comme flexibles.

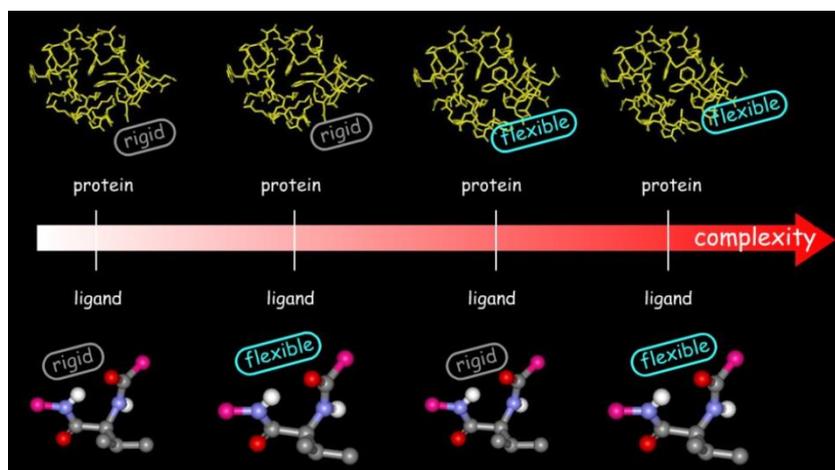


Figure 10 : Flexibilité moléculaire en docking protéine-ligand.

### 3.2.1.2. Docking avec ligand rigide

Le mécanisme de liaison d'un ligand à son récepteur a été envisagé comme un processus statique dans lequel le ligand constituait une clé de forme complémentaire à celle de la serrure qu'il était capable « d'ouvrir », le récepteur (modèle « clé-serrure »)

ou « *lock-and-key* » model) <sup>42</sup>. Pour tenter de reproduire ce modèle, les premiers logiciels de docking, parmi lesquels le logiciel DOCK<sup>43</sup> considérait donc le ligand et le site de liaison comme deux entités rigides. C'est ce qu'on appelle le docking avec ligand rigide.

Dans cette approche, la recherche de la pose optimale se limite au positionnement en énumérant toutes les rotations et translations possibles pour un ligand à l'intérieur du site de liaison. Ensuite une image négative du site de liaison est utilisée pour éliminer toutes les poses incompatibles avec le site actif (« clash », distance). Enfin, les poses sélectionnées se voient attribuées un score et les meilleures sont optimisées<sup>44</sup>.

Bien que moins précises que les méthodes de docking flexible, leur rapidité d'exécution permet de les employer lors des criblages virtuels comme un premier filtre permettant de ne pas retenir des molécules aberrantes (trop volumineuses, mauvaise complémentarité avec le site de liaison, ...) <sup>45</sup>.

### 3.2.1.3. Docking avec ligand flexible

Lorsque les méthodes de docking prennent en compte la flexibilité du ligand, deux étapes sont effectuées successivement pendant toute la durée du docking<sup>46</sup>.

Il existe plusieurs types d'algorithmes pour le traitement de la flexibilité du ligand : les méthodes systématiques (fragmentation /reconstruction), les méthodes aléatoires, et les méthodes de simulation (dynamique moléculaire) <sup>46,47</sup>.

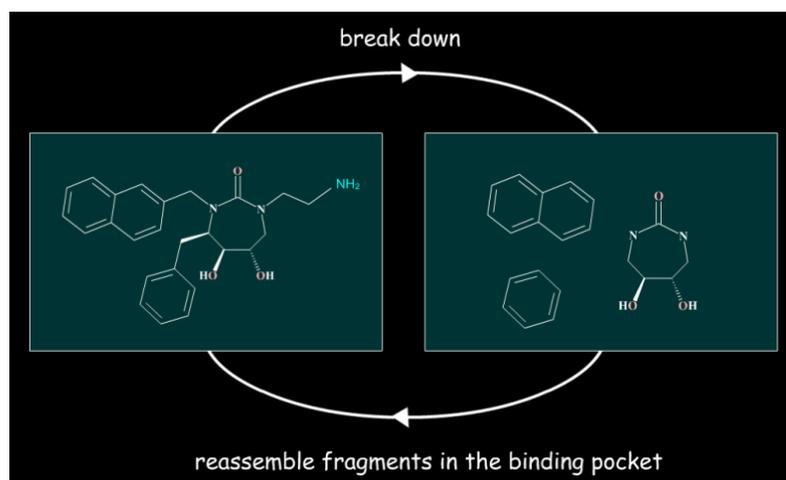
### 3.2.1.4. Algorithmes de recherche conformationnelle

#### 3.2.1.4.1. Recherche systématique

Les algorithmes de recherche systématique ont pour but d'explorer la totalité des degrés de liberté du ligand. Pour cela, toutes les torsions du ligand subissent des rotations de 0 à 360°, en utilisant un pas incrémental fixe, afin de générer la totalité des conformations possibles. Rapidement, ceci provoque une augmentation très importante du nombre de conformations créées que l'on appelle l'explosion combinatoire<sup>47</sup>.

Pour pallier ce problème, les méthodes de recherche systématique utilisent donc souvent un algorithme de fragmentation reconstruction qui consiste à diviser le ligand en plusieurs fragments pour le reconstruire dans le site de liaison. Par exemple

le ligand est divisé en fragments rigides et flexibles. Tous les fragments rigides sont placés dans le site de liaison et les fragments flexibles sont ajoutés de manière incrémentée<sup>46,47</sup>. Ce type d'algorithme est notamment employé dans les programmes FlexX<sup>48,49</sup> et les versions flexibles du programme Dock<sup>50,51</sup> (**Figure 11**).



**Figure 11** : Reconstruction incrémentale (programme Dock) : la molécule est divisée en fragments puis dockée et reconstruite dans le site actif<sup>51</sup>.

### 3.2.1.4.2. Recherche aléatoire

Ces méthodes procèdent à des changements aléatoires sur la conformation d'un ligand ou d'une population de ligands, lesquels sont acceptés ou rejetés sur la base d'une fonction de probabilité (fitness)<sup>52,53</sup>.

Les algorithmes génétiques utilisent des procédés inspirés de la théorie évolutionnaire de Darwin. En partant d'une population de conformations définies par un ensemble de variables assimilées à des gènes, ils appliquent des opérateurs génétiques (reproduction, croisement, mutations) permettant de faire varier les conformations jusqu'à l'obtention d'une population finale qui satisfait la fonction de « fitness » choisie<sup>47</sup>. Ce type d'algorithme est utilisé notamment dans les programmes Gold<sup>54,55</sup> et Autodock<sup>56</sup>. L'algorithme de recherche tabou explore l'espace conformationnel en interdisant les espaces déjà explorés et accepte les nouvelles solutions sur la base de leur RMSD par rapport aux précédentes conformations<sup>47</sup>. Il est utilisé par exemple dans le programme PRO\_LEADS<sup>57</sup>.

### 3.2.1.4.3. Méthodes de simulation

Ces méthodes sont basées sur la résolution des équations du mouvement de Newton. Elles comprennent les techniques des algorithmes de minimisation (par exemple dans les versions récentes du programme Dock et dans le programme ICM)<sup>50-53</sup> et de dynamique moléculaire. Ces dernières ne sont jamais utilisées pour générer de la flexibilité sur le ligand car elles demandent un temps de calcul très lent non compatible avec la gestion de bases de molécules.

### 3.2.1.5. Fonctions de score

Les fonctions de score sont utilisées pour estimer mathématiquement l'affinité de liaison entre un récepteur et chacune des poses générées pendant le docking<sup>58</sup>. Dans les programmes de docking les plus couramment utilisés, une seule fonction de score primaire existe, et elle est souvent équivalente à la (ou une lorsqu'il en existe plusieurs) fonction secondaire<sup>59</sup>. La distinction est donc souvent inexistante, mais leur rôle étant sensiblement différent, il est important de différencier les deux processus de score.

En effet, s'il est possible d'effectuer un rescoring afin de compenser d'éventuelles défaillances dans la fonction de score secondaire utilisée par le programme de docking, la qualité des poses générées dépend de façon critique de la capacité de la fonction primaire à distinguer les bonnes poses des mauvaises<sup>60,61</sup>.

Les fonctions de score peuvent être divisées en quatre classes : les fonctions basées sur des champs de force, les fonctions de scores empiriques, les fonctions de scores basées sur les connaissances et les fonctions de score consensus<sup>46,47</sup>.

#### 3.2.1.5.1. Fonctions de score basées sur les champs de force

Dans ce type de fonctions, le score est la somme de l'énergie d'interaction ligand-récepteur, de l'énergie interne du ligand et de l'énergie interne de la protéine, l'énergie interne de la protéine étant négligeable lorsque la conformation de celle-ci ne varie pas. Les énergies sont calculées à partir d'un terme de type Van Der Waals souvent (représentée par un potentiel de Lennard-Jones) et d'un terme électrostatique (potentiel de Coulomb) assorti d'une fonction diélectrique dépendante de la distance qui permet d'atténuer les contributions des interactions charge-charge. De nombreux logiciels de docking utilisent des fonctions de score basées sur les champs de force. Cependant, les valeurs des champs de force ne sont

pas toujours les mêmes. Ainsi, les fonctions G-Score<sup>48</sup> (basée sur le champ de force de Tripos<sup>48</sup>) et celle implémentée dans AutoDock<sup>63</sup> (basée sur le champ de force AMBER<sup>64</sup>) sont des exemples de ce type de fonction de score.

Ces fonctions sont limitées par l'absence de termes de solvation et d'entropie, ainsi que la nécessité d'introduire des distances de cutoff pour le traitement des interactions non liantes à longue distance<sup>46,47</sup>. D'autre part, les modèles PB/SA (Poisson-Boltzman/Surface Area)<sup>63</sup> et GB/SA (Generalized Born/Surface Area)<sup>65</sup> permettent de simuler le solvant de façon implicite mais les temps de calculs associés ne sont pas compatibles avec des études de docking à large échelle. Cependant, ces modèles peuvent être utilisés pour affiner la prédiction d'énergie de liaison ou le classement des ligands sur un petit nombre de composés, en combinaison avec une fonction de score ou après l'étape d'attribution de score<sup>65</sup>.

### 3.2.1.5.2. Fonctions de score empiriques

Ce type de fonction de score approxime l'énergie libre de liaison en sommant de façon pondérée différents termes d'interaction dérivés de paramètres structuraux. Les différents poids de la fonction de score sont ajustés pour reproduire en priorité des données expérimentales, telles que les constantes de liaison tirées d'un jeu d'entraînement de complexes protéine-ligand<sup>66</sup>. La plupart des programmes de docking implémentent ce type de fonction de score témoignant de leur efficacité (en termes de rapport précision/rapidité). Cependant, le principal désavantage de ces fonctions empiriques est leur forte dépendance envers les ensembles de données utilisés lors de leur calibration<sup>9,46,47</sup>, la plupart des fonctions de score empiriques existantes ayant été entraînées sur des training set d'une centaine de complexes au mieux. Cependant, l'augmentation constante du nombre de complexes pour lesquels il existe des données structurales et d'affinité permet d'envisager l'utilisation d'un grand nombre de complexes, ce qui devrait théoriquement permettre la conception d'une fonction empirique assez généraliste<sup>67</sup>. Parmi les principales fonctions de score empiriques, on peut citer : ChemScore<sup>68</sup>, PLP<sup>69</sup> et LigScore<sup>70</sup>.

### 3.2.1.5.3. Fonctions de scores basées sur les connaissances

Ces fonctions de score sont construites à partir de règles fondées sur une analyse statistique des complexes protéine-ligand résolus expérimentalement. Elles

partent du principe que les distances interatomiques les plus représentées statistiquement dans les complexes constituent des contacts énergétiques favorables et, qu'à l'inverse, les plus rares représentent des interactions moins stables.

Ainsi, leur paramétrisation dépend de la quantité d'informations expérimentales disponibles et on doit leur apparition à la profusion de données structurales accessibles dans des bases de données telles que la *Protein Data Bank*. Les exemples populaires de ces fonctions de scores sont PMF<sup>71</sup>, SmoG<sup>72</sup> et DrugScore<sup>73</sup>.

### 3.2.1.5.4. Fonction de score consensus

Ces fonctions hybrides combinent les résultats issus de diverses fonctions de score. On estime qu'il est possible de compenser partiellement les faiblesses intrinsèques de chacune des fonctions de score employées, évitant leurs erreurs individuelles et ainsi d'augmenter la probabilité d'identifier des composés actifs<sup>74</sup>. Cependant, si les termes des différentes fonctions de score sont fortement corrélés, l'intérêt du consensus devient limité car il peut entraîner une amplification des erreurs, au lieu de les atténuer. Ces fonctions consensus ont récemment fait l'objet d'une revue<sup>75</sup>.

### 3.2.1.6. Principaux logiciels de docking

De nombreux logiciels de docking présentant des algorithmes de recherche et des fonctions de score variées ont été développés au cours du temps (**Tableau 3**).

**Tableau 3** : Différents logiciels de docking.

Programme	Traitement du ligand	Recherche conformationnelle	Algorithme de recherche	Fonction de score
<b>AutoDock</b> <sup>55</sup>	Flexible	Stochastique	Génétique	Basée sur les champs de force
<b>Dock</b> <sup>75</sup>	Flexible	Systématique	Fragmentation / reconstruction	Basée sur les champs de force
<b>FlexX</b> <sup>48</sup>	Flexible	Systématique	Fragmentation / reconstruction	Basée sur les connaissances
<b>FRED</b> <sup>43</sup>		Systématique	Recherche	Basée sur les

			exhaustive	connaissances
<b>Glide</b> <sup>76</sup>	Flexible	Stochastique	Monte Carlo	Empirique
<b>Gold</b> <sup>54</sup>	Flexible	Stochastique	Génétique	Basée sur les champs de force
<b>ICM</b> <sup>52</sup>	Flexible	Stochastique	Monte Carlo	Basée sur les champs de force
<b>Surflex-Dock</b> <sup>77</sup>	Flexible	Systématique	Hammerhead modifié	Empirique
<b>PRO_LEADS</b> <sup>56</sup>	Flexible	Stochastique	Tabou	Empirique

### 4. Conclusion

Dans ce chapitre nous avons présenté un aperçu général sur la conception de médicaments ainsi que les différentes étapes essentielles pour la recherche et le développement d'un nouveau médicament.

Nous avons décrit le principe du criblage virtuel qui consiste en la recherche de nouveaux composés actifs biologiquement en exploitant les gigantesques chimiotèques. Cette stratégie vise à réduire de façon significative le temps et les coûts de l'identification de nouvelles molécules. Nous avons également présenté brièvement les différentes méthodes assistées par ordinateur, utilisées dans le criblage virtuel : les méthodes « *structure based* » basées sur la structure de la cible thérapeutique et « *ligand based* » basées sur la structure du ligand, particulièrement la méthode de docking moléculaire et les études QSAR qui seront utilisées dans la partie expérimentale de notre travail de thèse.

## 5. Références

1. W. Sneader, Drug discovery: a history. *John Wiley & Sons*, **2005**.
2. Schmitz, R, *Pharmacy in history*, 27(2), **1985**, 61-74.
3. S. M. Paul, D. S. Mytelka, C. T. Dunwiddie, C. C. Persinger, B. H. Munos, S. R. Lindborg, & A. L. Schacht, *Nature reviews Drug discovery*, 9(3), **2010**, 203-214
4. J.A. Dimasi, R.W. Hansen, H.G. Grabowski, *J Heal. Econ*, 22, **2003**, 151-185.
5. W.P. Walters, M.T. Stahl, M.A. Murcko, *Drug Discov. Today*, 3, **1998**, 160-178.
6. K.H. Bleicher, H. Bohm, K. Muller, A.I. Alanine, *Nature reviews Drug discovery*, 2, **2003**, 369-378.
7. A.C. Cheng, R.G. Coleman, K.T. Smyth, Q. Cao, P. Souldard, D.R. Caffrey, A.C. Salzberg, E.S. Huang, *Computational-Biology*, 25, **2007**, 71-75.
8. J. Owens, *Nat. Rev. Drug Discov*, 6, **2007**, 187.
9. G. Vistoli, A. Pedretti, B. Testa, *Drug Discov. Today*, 13, **2008**, 285-294.
10. T. Tuccinardi, CombChem High Throughput Screen, 12, **2009**, 303-14.
11. N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, C.R. Corbeil, *Br. J. Pharmacol*, 153, **2008**, 7-26.
12. T.I. Oprea, H. Matter, *Curr. Opin. Chem. Biol*, 8, **2004**, 349-358.
13. M.J. Keiser, V. Setola, J.J. Irwin, C. Laggner, A.I. Abbas, S.J. Hufeisen, N.H. Jensen, M.B. Kuijjer, R.C. Matos, T.B. Tran, R. Whaley, R. a Glennon, J. Hert, K.L.H. Thomas, D.D. Edwards, B.K. Shoichet, B.L. Roth, *Nature*, 462, **2009**, 175-181.
14. C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, *Adv. Drug Deliv. Rev*, 23, **1997**, 3-25.
15. P.D. Leeson, B. Springthorpe, *Nat. Rev. Drug Discov*, 6, **2007**, 881-890.
16. B.K. Shoichet, *Nature*, 432, **2004**, 862-865.
17. J. Waszkowycz, B. Perkins, T. D. J. Sykes, R. A. Li, *IBM Syst. J*, 40, **2001**, 360-376.
18. J. Bajorath, *Nat. Rev. Drug Discov*, 1, **2002**, 882-894.
19. J. Mestres, *Biochem Soc Trans*, 30, **2002**, 797-9.
20. F. Stahura, J. Bajorath, *Comb Chem High Throughput Screen*, 7, **2004**, 259-69.
21. J. Perez, *Chem. Soc. Rev*, 34, **2005**, 143-152.
22. M.P. Gleeson, A. Hersey, S. Hannongbua, *Curr. Top. Med. Chem*, 11, **2011**, 358-381.
23. S. Ghosh, A. Nie, J. An, Z. Huang, *Curr. Opin. Chem. Biol*, 10, **2006**, 194-202.
24. P.D. Lyne, *Drug Discov. Today*, 7, **2002**, 1047-1055.

25. T. Lengauer, C. Lemmen, M. Rarey, M. Zimmermann, *Drug Discov.Today*, 9, **2004**, 27–34.
26. I. Muegge, S. Oloff, *Drug Discov. TodayTechnologies*, 3, **2006**, 405–411.
27. G.M. Johnson, M.A. Maggiora, *Wiley Ed*, **1990**.
28. K. Roy, P. P. Roy, and J. T. Leonard, *Chem. Cent. J*, **2008**, 2(S1) p. P9.
29. N. Naceiri, Mrabti and M. Elhallaoui, *Univ. Sci*, 11(1), **2017**, 18–39.
- 30.X. Yang, B. Albijanic, Y. Zhou, Y. Zhou, and X. Zhu, *Miner. Eng*, 129, **2018**, 112–119.
- 31.R. Guha and E. Willighagen, *Curr. Top. Med. Chem*, 12(18), **2012**, 1946–1956.
- 32.R. Todeschini and V. Consonni, *Handbook of molecular descriptors*, 11, **2008**, John Wiley & Sons.
- 33.D. J. Livingstone, *J. Chem. Inf. Comput. Sci*, 40(2), **2000**, 195–209.
- 34.M. Dehmer, K. Varmuza, and D. Bonchev, 2, **2012**.
- 35.L. J. de Campos and E. B. de Melo, *J. Mol. Struct*, 1141, **2017**, 252–260
36. K. Roy and R. Narayan Das, *Curr. Drug Metab*, 15 (4), **2014**, 346–379.
37. J. G. Topliss and R. J. Costello, *J. Med. Chem*, 15(10), **1972**, 1066–1068.
- 38.S. Wold, *Technometrics*, 20( 4), **1978**, 397–405.
- 39.I. Mitra, A. Saha, and K. Roy, *Mol. Simul*, 36 (13), **2010**,1067–1079.
40. A. Golbraikh and A. Tropsha, *J. Mol. Graph. Model*, 20(4), **2002**, 269–276.
41. X. Barril, R. Soliva, *Mol. BioSyst*, 2, **2006**, 660–681.
42. E. Fischer, Einfluss der Configuration auf die Wirkung der Enzyme, *Berichte Der Dtsch. Chem. Gesellschaft*, 27(3), **1894**, 2985–2993.
43. I. Kuntz, J. Blaney, S. Oatley, R. Langridge, T. Ferrin, *J Mol Biol*, 161, **1982**, 269–88.
44. M. McGann, *J. Chem. Inf. Model*, 51, **2011**, 578–596.
- 45.M.A. Miteva, W.H. Lee, M.O. Montes, B.O. Villoutreix, *J. Med. Chem*, 48, **2005**, 6012–6022.
- 46.D. Kitchen, H. Decornez, J. Furr, J. Bajorath, *Nat Rev Drug Discov*, 3, **2004**, 935–949.
47. S. Sergio Filipe, F. PerdoAlexandrino, R. Maria Joao, *Proteins, Proteins*, 65, **2006**, 15–26.
48. B. Kramer, M. Rarey, T. Lengauer, *Proteins*, 37, **1999**, 228–241.
49. M. Rarey, B. Kramer, T. Lengauer, E. Al, *J Mol Biol*, 261, **1996**, 470–89.
50. T.J.A. Ewing, S. Makino, A.G. Skillman, I.D. Kuntz, *J. Comput.Aided.Mol. Des*, 15, **2001**, 411–428.

- 51.D. Moustakas, P. Lang, S. Pegg, E. Pettersen, I. Kuntz, N. Brooijmans, R. Rizzo, *Development J, Comput, Aided Mol. Des*, 20, **2006**, 601–619.
- 52.R. Abagyan, S. Batalov, T. Cardozo, M. Totrov, J. Webber, Y. Zhou, *Proteins Struct. Funct. Genet*, 29, **1997**, 29–37.
53. R. Abagyan, M. Totrov, D. Kuznetsov, *J. Comput. Chem*, 15, **1994**, 488–506.
54. G. Jones, P. Willett, R. Glen, *J. Mol. Biol*, 245, **1995**, 43–53.
- 55.G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, *J. Mol. Biol*, 267, **1997**, 727–48.
56. G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, *J. Comput. Chem*, 19, **1998**, 1639–1662.
57. C.A. Baxter, C.W. Murray, D.E. Clark, D.R. Westhead, M.D. Eldridge, *Proteins Struct. Funct. Genet*, 33, **1998**, 367–382.
58. R. Dias, W.F.J. de Azevedo, *Curr Drug Targets*, 9, **2008**, 1040–7.
59. M. Stahl, M. Rarey, *J. Med. Chem*, 44, **2001**, 1035–1042.
60. V. Mohan, A.C. Gibbs, M.D. Cummings, E.P. Jaeger, L. Renee, *Curr. Pharm.Des*, 11, **2005**, 323–333.
61. E. Perola, W.P. Walters, P.S. Charifson, *Proteins Struct. Funct. Genet*, 56, **2004**, 235–249.
62. D. Goodsell, G. Morris, A. Olson, *J. Mol. Recognit*, 9, **1996**, 1–5.
- 63.W. Rocchia, S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, B. Honig, *J. Comput. Chem*, 23, **2002**, 128–137.
64. X. Liu, H. Kuntz, I.D. Zou, *J. Phys. Chem. B*, 108, **2004**, 5453–5462.
65. E. Yuriev, M. Agostino, P.A. Ramsland, *J. Mol. Recognit*, 24, **2011**, 149–164.
- 66.A. Di Nola, D. Roccatano, H.J.C. Berendsen, *Proteins Struct. Funct. Bioinforma*, 19, **1994**, 174–182.
67. S.-Y. Huang, S.Z. Grinter, X. Zou, *Phys. Chem. Chem. Phys*, 12, **2010**, 12899–12908.
68. M.D. Eldridge, C.W. Murray, T.R. Auton, G. V Paolini, R.P. Mee, *J. Comput. Aided. Mol. Des*, 11, **1997**, 425–445.
- 69.D.K. Gehlhaar, G.M. Verkhivker, P.A. Rejto, C.J. Sherman, D.R. Fogel, L.J. Fogel, S.T. *Chem. Biol*, 2, **1995**, 317–324.
- 70.A. Krammer, P.D. Kirchhoff, X. Jiang, C.M. Venkatachalam, M. Waldman, *J. Mol. Graph. Model*, 23, **2005**, 395–407.
71. I. Muegge, Y.C. Martin, *J. Med. Chem*, 42, **1999**, 791–804.
72. R.S. DeWitte, E.I. Shakhnovich, *J. Am. Chem. Soc*, 118, **1996**, 11733–11744.

73. H. Gohlke, M. Hendlich, G. Klebe, *J. Mol. Biol.*, 295, **2000**, 337–356.

74. P.S. Charifson, J.J. Corkery, M.A. Murcko, W.P. Walters *J. Med. Chem.*, 42, **1999**, 5100–5109.

75. M. Feher, *Drug Discov. Today*, 11, **2006**, 421–428.

76. P. Lang, S. Brozell, S. Mukherjee, E. Pettersen, *Rna*, 15(6), **2009**, 1219–1230.

77. R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, *J. Med. Chem.*, 47, **2004**, 1739–1749.

# **Partie II**

## **Résultats et discussions**

# **Chapitre I**

## **Criblage par Docking de séries de molécules dérivées d'arbutine et de gingérol**

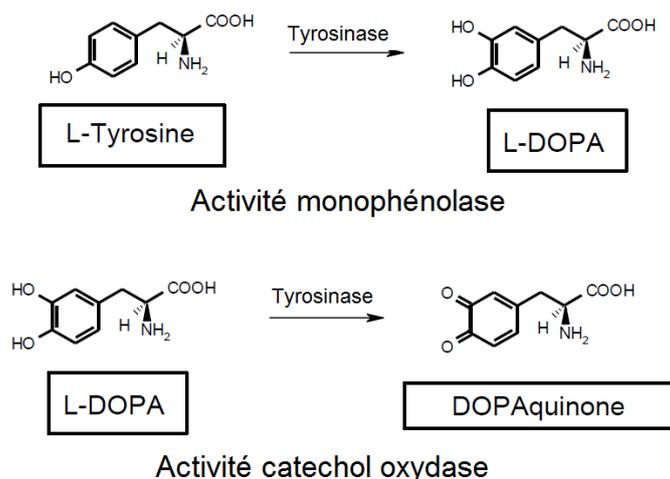
### 1. Introduction

La peau est le tissu qui représente l'interface entre notre organisme et l'environnement extérieur. Elle joue le rôle d'une barrière contre les agressions extérieures et est le reflet de notre santé. De ce fait, les pathologies et les imperfections cutanées sont traitées avec attention de plus en plus grande au même titre que les dysfonctionnements des autres organes du corps humain. Parmi les dysfonctionnements cutanés, les hyperpigmentations et leur traitement tiennent une place importante dans la recherche en dermocosmétique. Ces hyperpigmentations ont pour origine, un dérèglement dans la production de mélanines, qui sont les pigments responsables de la coloration cutanée. Par conséquent, ces hyperpigmentations sont traitées par des produits dépigmentants inhibant la mélanogénèse.

La mise au point de tels produits nécessite une bonne compréhension du phénomène de production de la mélanine (mélanogénèse), mais également des régulations et des dysfonctionnements possibles de la synthèse. Pour inhiber la mélanogénèse, plusieurs mécanismes peuvent être mis en œuvre tels que le blocage de la synthèse de mélanine à un stade donné, en mettant au point des inhibiteurs de la tyrosinase, qui est une enzyme jouant un rôle primordial dans la synthèse des mélanines.

La tyrosinase est une métalloprotéine qui catalyse les deux premières étapes communes de la mélanogénèse et s'avère être ainsi l'enzyme limitante<sup>1-3</sup>. Son absence ou des mutations de son gène conduisent à une diminution voire à un arrêt de la pigmentation (l'albinisme oculocutané)<sup>4</sup>.

Chez les mammifères, la tyrosinase catalyse l'hydroxylation de la L-tyrosine en L-DOPA (il s'agit de l'activité monophénol hydroxylase) et l'oxydation de la L-DOPA en Dopaquinone (activité catéchol oxydase)<sup>5</sup>.



**Figure 1** : schémas des activités monophénolase et catéchol oxydase de la tyrosinase.

Chez les insectes, les tyrosinases sont impliquées dans le processus de sclérotisation (durcissement de la chitine) et dans les mécanismes de défense<sup>6,7</sup>.

Dans l'agroalimentaire, la tyrosinase est connue comme l'enzyme responsable du changement de la couleur des fruits et de certains végétaux (brunissement). Ce phénomène est dû à l'oxydation de composés diphénoliques en quinone<sup>7</sup>.

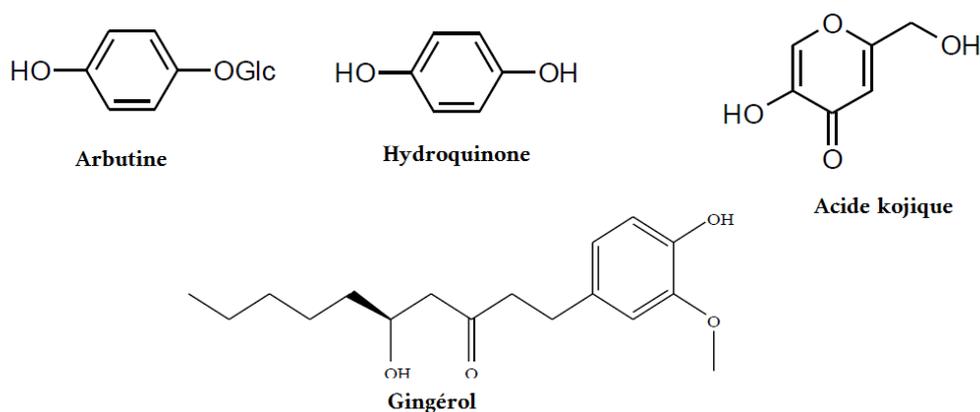
Au cours de la vie, des anomalies pigmentaires ou dyschromies peuvent survenir par apparition des taches d'hyperpigmentation marquées par un assombrissement de la peau, des taches d'hypopigmentations marquées par une perte de la coloration naturelle de la peau ou des affections dyschromiques marquées par une coloration anormale de la peau (**Figure 2**)<sup>8</sup>.



**Figure 2** : Exemples du phénomène d'hyperpigmentation de la peau.

Ces hyperpigmentations, qui résultent souvent d'une augmentation de la synthèse de mélanine, peuvent être particulièrement inesthétiques et motivées chez certaines personnes l'utilisation de composés à activité dépigmentante qui peuvent freiner la production de mélanines. La majorité des produits dépigmentants agissent en inhibant l'activité de la tyrosinase. Ces produits peuvent être des analogues de substrats, des inhibiteurs compétitifs ou encore des chélateurs de cuivre, cofacteur indispensable à l'activité tyrosinase.

De nombreux composés ont été décrits comme agents dépigmentants tels que : l'hydroquinone, l'une des premières molécules dépigmentantes dont l'efficacité a été prouvée, l'arbutin qui est une prodrogue de l'hydroquinone largement utilisée en cosmétique et l'acide kojique qui est très répandu dans les formulations dermocosmétiques (**Figure 3**)<sup>9-11</sup>.



**Figure 3** : Exemples d'inhibiteurs de la tyrosinase connus expérimentalement.

Les recherches sur la mélanogenèse et le traitement des hyperpigmentations ont beaucoup progressé ces dernières années. Par conséquent, les inhibiteurs de la tyrosinase sont devenus des constituants importants des produits cosmétiques et dermatologiques ayant pour cible les hyperpigmentations cutanées. De ce fait, la recherche de molécules naturelles, d'origine végétale, inhibitrices de tyrosinase s'est considérablement développée, car les plantes constituent une source importante de molécules bioactives souvent exemptes d'effets secondaires. Nous avons choisi dans cette partie de travail de thèse d'étudier deux familles de molécules naturelles : les analogues de l'arbutine et ceux du gingérol, molécule extraite du gingembre « *Zingiber officinale roscoe* »<sup>12</sup>.

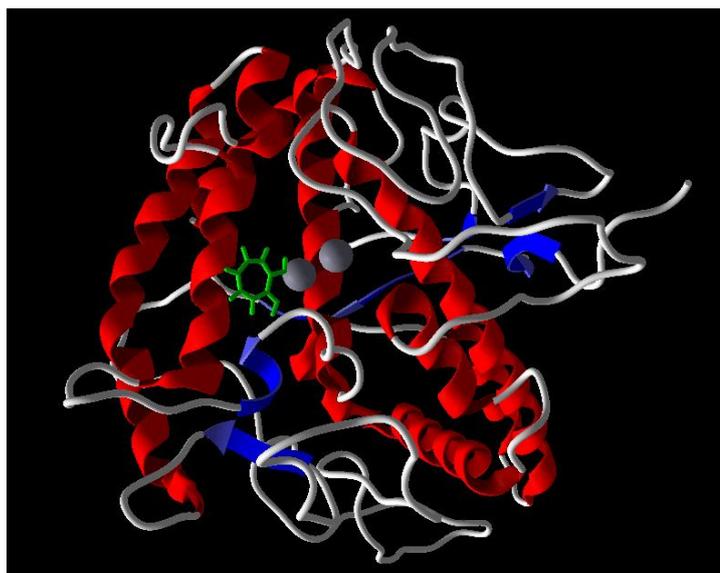
Différentes molécules naturelles ou non, appartenant à ces deux familles, seront criblées virtuellement et évaluées par docking moléculaire sur notre cible la tyrosinase afin d'enrichir la classe thérapeutique des inhibiteurs de cette enzyme.

## 2. Méthodes de recherche

### 2.1. Choix de la cible (Enzyme)

Dans le cadre de notre thématique de recherche qui porte sur le développement de nouveaux inhibiteurs de la tyrosinase, l'enzyme tyrosinase, extraite du champignon *Agaricus Bisporus*, a été sélectionnée comme enzyme cible pour réaliser notre criblage qui vise le développement de nouveaux dérivés d'arbutine et de gingérol inhibiteurs de cette enzyme en tant que substances médicamenteuses candidates. La structure tridimensionnelle de l'enzyme cible a été extraite de la Protein Data Bank (PDB ID : 2Y9X)<sup>13</sup>. La cible sélectionnée (ID : 2Y9X) représente les structures cristallines de l'octamère L4H4 (quatre dimères L-H) de la tyrosinase du champignon *Agaricus Bisporus*<sup>14</sup> avec une résolution de 2,78 Å.

Le domaine tyrosinase représenté par la sous-unité H1, avec une séquence d'une longueur de 392 acides aminés, contenant le site de liaison binucléaire du cuivre (**Figure4**) est sélectionné pour réaliser les études de docking moléculaire<sup>15</sup>.



**Figure 4** : La tyrosinase, extraite du champignon *Agaricus Bisporus* (PDB : 2Y9X).

### 2.2. Chimiothèque de ligands : recherche de similarité chimique

Notre choix s'est porté sur deux familles de ligands cibles ayant prouvé leur effet inhibiteur sélectif vis-à-vis de l'enzyme tyrosinase : les dérivés de gingérol et les dérivés d'arbutine. Une sélection de ces dérivés a mené à la construction de deux chimiothèques virtuelles à cribler comprenant 1800 analogues de gingérol et 1268 analogues d'arbutine qui n'ont pas encore montré une activité inhibitrice vis-à-vis de la tyrosinase. Les structures tridimensionnelles de l'arbutine (Zinc04104676) et du gingérol (ZINC1531846) ainsi que leurs analogues ont été extraites de la base de données publique gratuite Zinc<sup>16</sup>. Nous avons adopté le coefficient tanimoto comme paramètre de recherche avec 70% comme seuil de similarité. Toutes les structures obtenues sont enregistrées sous forme de fichiers "mol2" pour nos études d'amarrage. La bibliothèque est ensuite préparée avec le programme MVD.

### 2.3. Processus du docking moléculaire

Lorsque la structure tridimensionnelle est connue, l'affinité de liaison entre le ligand et l'enzyme est traitée avec le docking moléculaire. Les études de docking moléculaire ont été réalisées avec le programme Molegro Virtual Docker (MVD 2012.5.5.0, Molegro ApS). Le programme est basé sur trois algorithmes principaux : la fonction de score Moldock, l'algorithme de recherche conformationnelle et l'algorithme de prédiction de cavité. Notre choix de ce logiciel est justifié par sa meilleure performance pour étudier et prédire le mode d'interaction ligand-récepteur en comparaison avec d'autres logiciels fréquemment utilisés tels que Autodock, Superflex, Flex et Gold<sup>17-19</sup>.

Pour une meilleure visualisation après les simulations du docking, les résultats obtenus ont été stockés et traités avec le logiciel PyMol<sup>20</sup>.

#### 2.3.1. Préparation de la cible et les ligands par MDV

Avant d'effectuer le docking moléculaire avec Molego, le complexe ligand-protéine ainsi que son environnement doit être préparé à l'aide du même logiciel en éliminant toutes les molécules d'eau en ne gardant que l'enzyme, le ligand de référence et le

cofacteur. L'enzyme est considéré comme un corps rigide en gardant les résidus fixes et les ligands sont flexibles. La structure de l'enzyme a été préparée et corrigée en effectuant les principales opérations suivantes:

- Ajout des hydrogènes explicites
- Assignement des charges
- Correction des types des liaisons (simple ou double).
- Détection des torsions et les liaisons rotatives.

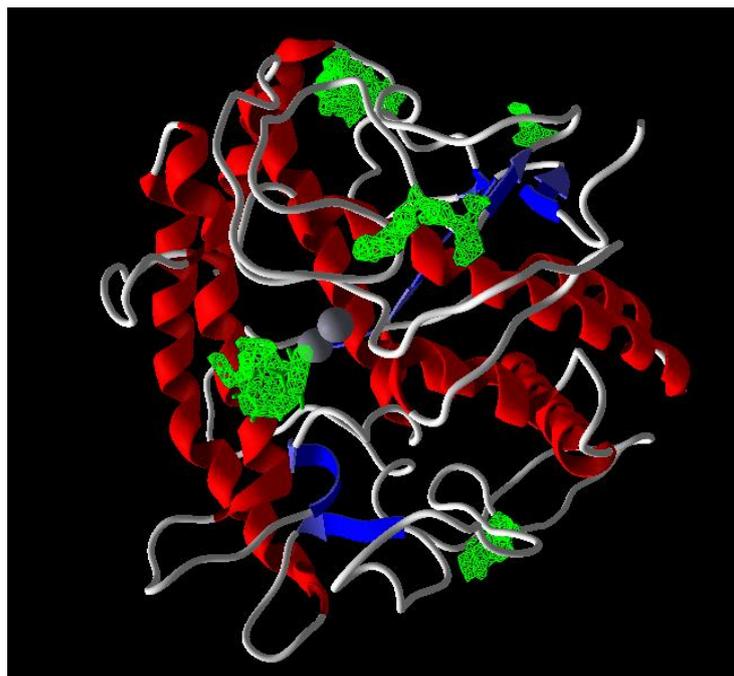
### 2.3.2. Détection des cavités et définition de la région du site actif

Molegro virtual Docker (MVD) possède un algorithme de détection de cavités moléculaires dans le volume tridimensionnel du récepteur afin de prédire les sites d'interactions possibles. Cet algorithme basé sur une grille «grid-based » fonctionne comme suit :

Une grille discrète d'une résolution de 0,8 Å couvrant tout le volume de la protéine est créée. A chaque point de cette grille, une sphère de rayon 1,4 Å est placée en vérifiant l'intersection de ces sphères avec la surface de protéine. Les points de grille qui chevauchent avec les atomes de protéine sont définis en tant que points inaccessibles et les autres sont considérés points accessibles.

Chaque point accessible de cette grille sera analysé afin de vérifier s'il fait partie de la cavité par l'insertion des lignes de directions aléatoires à partir de ce point jusqu'à franchir les limites de la grille. Si la majorité des directions franchissent d'autres points accessibles au cours de leurs chemin (à revoir défaut au moins 12 sur 16 directions différentes testés), ce point est considéré faisant partie d'une cavité. Ceci est répété un certain nombre de fois pour définir tous les points accessibles des cavités. La dernière étape consiste à déterminer les régions de cavités en connectant les points voisins. Les cavités trouvées sont ensuite classées en fonction de leur volume.

Par le biais de cet algorithme, nous avons détecté cinq cavités pour la tyrosinase, ce nombre de cavités est donné par défaut par le logiciel. Un nombre plus élevé de cavités peut être défini par le logiciel selon le besoin (**Figure5**).



**Figure 5 :** Illustration des cinq cavités du Tyrosinase (grilles en vert) par l'algorithme de détection de MVD.

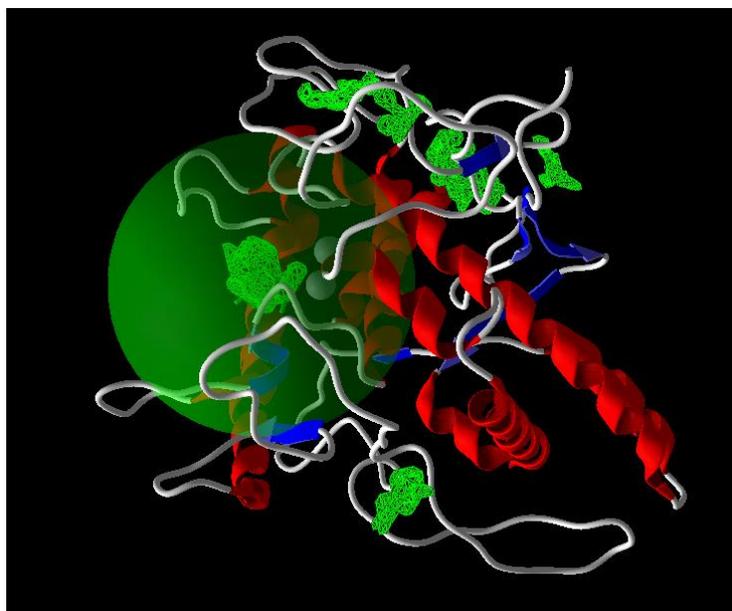
Les cavités sont classées selon leur volume (**Tableau 1**).

**Tableau 1:** Cavités détectées avec l'algorithme de détection de MVD pour 2Y9X

Cavité détectée	Coordonnées (X,Y,Z)	Surface (Å <sup>2</sup> )	Volume (Å <sup>3</sup> )
Cavité 1	(3,76 ; -6,28 ; -33,61)	245,76	85,50
<b>Cavité 2</b>	<b>(-9,84 ; -28,39 ; -44,56)</b>	<b>160,00</b>	<b>44,54</b>
Cavité 3	(-24,81 ; -13,06 ; -43,43)	131,84	30,72
Cavité 4	(-15,36 ; 28,16 ; 98,56)	98,56	28,16
Cavité 5	(-11,88 ; -5,41 ; -26,71)	66,56	14,34

Lors de notre investigation des cavités candidates pour un éventuel site d'interaction de notre étude docking, à l'aide de l'algorithme de prédiction de cavités, nous avons remarqué que le ligand de référence (Tropolone) est Co-cristallisé dans la cavité 2 et entouré par les résidus catalytiques. Nous avons donc décidé de conserver cette cavité pour la considérer comme centre de recherche (research center) pour le

processus du docking. Cette sphère de recherche (en vert) couvre tous les résidus du site actif avec un rayon de 14 Å (**Figure 6**).



**Figure 6** : Illustration de l'espace de recherche.

### 2.4. Protocole et validation

#### 2.4.1. Algorithme de recherche conformationnelle

Moldock Optimizer, utilisé comme algorithme de recherche par le programme MVD, est une technique d'optimisation itérative inspirée de la théorie de l'évolution de Darwin introduite par Storn et Price<sup>21</sup>. L'algorithme de Moldock génère aléatoirement la première population. Ensuite, les mauvaises solutions sont éliminées et les bonnes solutions candidates sont sélectionnées et modifiées à l'aide d'opérations de croisement et de mutation entre les solutions candidates (poses)<sup>22</sup>.

La principale idée innovante dans le Moldock Optimizer est la création de solutions dites "enfants" à partir d'une différence par rapport aux individus initiaux "parents" de la population.

-Représentation des individus de la population :

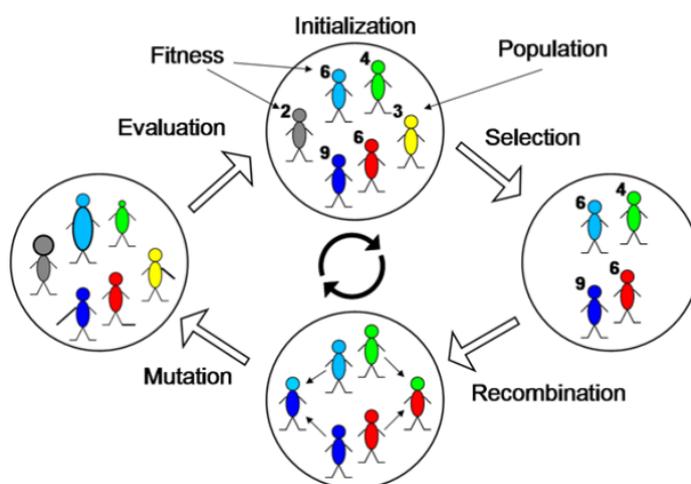
Seules les propriétés des ligands sont représentées pour les individus puisque la protéine reste rigide lors du docking. Ainsi, chaque solution est encodée en représentant sa position, son orientation et sa conformation avec des coordonnées cartésiennes, un vecteur de rotation, un angle de rotation et un angle de torsion pour chaque liaison flexible.

Ces paramètres sont attribués initialement au ligand de manière aléatoire pour définir les individus "parents" de la population. Pour chaque individu de la population initiale, une position aléatoire à l'intérieur de l'espace de recherche (déjà défini) est attribuée. L'initialisation de l'orientation se déroule en choisissant simplement l'axe d'orientation (entre -1,0 et +1,0, suivi par la normalisation des valeurs pour former un vecteur unitaire) et l'angle de rotation compris entre  $-180^\circ$  et  $+180^\circ$ . Les angles flexibles de torsion (le cas échéant) se voient attribuer également un angle aléatoire entre  $-180^\circ$  et  $+180^\circ$ .

-Principe de l'algorithme :

Une des raisons pour lesquelles l'algorithme évolutionnaire différentielle est attribué au MVD c'est sa capacité à nous permettre l'exploitation d'une diversité importante de la population : la progression des solutions parentales, initialement générées de manière aléatoire avec une grande variation, entrainera la création de nouvelles solutions "enfants" largement différentes.

Cet algorithme converge vers les meilleures solutions évaluées à l'aide d'une fonction de score, ce qui a pour résultat la diminution de la taille de la population et la création de meilleures solutions enfants. Ensuite, les solutions enfants remplacent les parents, si et seulement si elles sont plus adaptées. Dans le cas contraire, le parent survit et est transmis à la prochaine génération (appelé itération de l'algorithme) (**Figure 7**). Ce processus est exécuté aussi longtemps que la condition de fin de recherche n'est pas remplie.



**Figure 7** : Principe de l'algorithme évolutionnaire différentiel de MDV.

Nous avons utilisé Moldock optimizer comme algorithme de recherche avec le réglage algorithmique suivant :

- Nombre de tours (runs) : 100
- Taille de population : 100
- Itération maximale de l'algorithme : 10000
- Taux de croisement = 0,9
- Fin de recherche : basée sur la variation « variation-based »

### 2.4.2. Fonction de score

Lors de la simulation du docking, l'algorithme de recherche génère plusieurs poses (complexe ligand-protéine) pour chaque ligand, nous sélectionnons alors les meilleures poses avec les énergies les plus basses par le biais de la fonction de score «MolDock Score» qui calcule les valeurs d'énergies d'interaction des poses générées et les classe par ordre croissant selon leurs valeurs de l'énergie de score. Cette fonction de score utilisée par MVD est un dérivé amélioré des fonctions de score PLP (piecewise linear potential) initialement proposées par Gehlhaar et al.<sup>24,25</sup> et développées plus tard par Yang et al.<sup>26</sup> avec l'insertion de nouveaux termes de liaisons hydrogène et de charges.

La fonction de score est composée de deux termes représentant respectivement l'énergie intramoléculaire (l'interaction entre atomes d'un même ligand) et l'énergie d'interaction externe (ligand-protéine et ligand-cofacteur). L'énergie de pose (enzyme –ligand) est représentée par la somme de ces deux termes énergétiques et leurs expressions sont définies comme suit :

$$E_{score} = E_{inter} + E_{intra} \quad eq.01$$

Avec :  $E_{inter}$ : énergie d'interaction ligand-protéine.

$E_{intra}$ : énergie d'interaction intramoléculaire du ligand

L'énergie d'interaction ligand-protéine  $E_{inter}$  est donnée par l'équation suivante:

$$E_{inter} = \sum_{i \in \text{ligand}} \sum_{j \in \text{protéine}} \left[ E_{PLP}(r_{ij}) + 332.0 \frac{q_i q_j}{4r_{ij}^2} \right] \quad eq. 02$$

Cette équation prend en compte tous les atomes lourds ( $i$ ) du ligand et les atomes lourds ( $j$ ) de la protéine ainsi que les atomes des cofacteurs et des molécules d'eau (le cas échéant).

-  $E_{PLP}$ : le potentiel linéaire PLP

-  $332.0 \frac{q_i q_j}{4r_{ij}^2}$ : décrit les interactions électrostatiques entre les atomes chargés.

La valeur numérique 332.0 est donnée pour fixer l'unité de l'énergie électrostatique en Kcal/mol.

-  $D(r) = 4r_{ij}$ : constante diélectrique

-  $q_i$  et  $q_j$ : représentent respectivement les charges des atomes du ligand et des atomes de la protéine.

Le potentiel  $E_{PLP}$  prend en considération principalement deux paramètres; le premier optimise les interactions stériques (Van der Waals) entre les atomes, le second optimise les potentiels des liaisons hydrogènes.

Une liaison est considérée comme une liaison hydrogène si l'un des atomes peut donner un atome d'hydrogène et l'autre atome peut l'accepter. Les types d'atomes qui peuvent contribuer à la formation de liaisons hydrogène sont classés selon **le tableau 2**.

**Tableau 2** : Type d'atomes contribuant à la formation d'une liaison hydrogène.

Atomes	Types
N et O sans H liés	Accepteur

N et S avec un ou plusieurs H liés	Donneur
O avec H attachés	Donneur et Accepteur

L'équation suivante représente l'énergie d'interaction intramoléculaire du ligand  $E_{intra}$  :

$$E_{intra} = \sum_{i \in \text{ligand}} \sum_{j \in \text{ligand}} E_{PLP}(r_{ij}) + \sum_{\text{liaisons flexibles}} A [1 - \cos(m \cdot \theta - \theta_0)] + E_{clash} \quad \text{eq. 03}$$

Le premier terme de cette équation correspond à la somme des énergies d'interaction entre toutes les paires d'atomes du ligand. Le second terme représente l'énergie de torsion, cette énergie dépend essentiellement de l'hybridation des atomes liés (tableau 3).  $\theta$  est l'angle de torsion de la liaison, cet angle ne correspond pas nécessairement à une seule torsion, une moyenne est définie dans le cas où plusieurs torsions sont déterminées. Le dernier terme,  $E_{clash}$ , correspond à une constante de pénalité égale à 1000 attribuée à l'équation afin d'éliminer les conformations du ligand dans le cas où un atome lourd du ligand est situé en dehors de la région du site d'interaction (l'espace de recherche), ou si la distance entre deux atomes lourds est inférieure à 2,0 Å.

**Tableau 3** : Les paramètres correspondant à l'énergie de torsion.

Liaison	$\theta_0$	m	A
sp <sup>2</sup> -sp <sup>3</sup>	0.0	6	1.5
sp <sup>3</sup> -sp <sup>3</sup>	$\pi$	3	3.0
sp <sup>2</sup> -sp <sup>2</sup>	0.0	2	3.0

A la fin du processus du docking, un ensemble de poses prometteuses est généré et les termes d'énergie calculés (**tableau 4**) sont stockés et regroupés dans un fichier de format «*mvd results*» destiné à être inspecté dans le «*work space*» du MVD ou avec Molegro Data Modeller.

**Tableau 4** : Les termes d'énergie calculés avec la fonction de score MolDock Score.

Terme d'énergie	Signification
$E_{\text{Total}}$	Energie totale MolDock Score
$E_{\text{Inter (protein - ligand)}}$	Energie d'interaction MolDock Score totale entre le ligand et la protéine).
$E_{\text{Steric}}$	Energie d'interaction stérique entre la protéine et le ligand (calculé par PLP).
$H_{\text{Bond}}$	Energie de liaison hydrogène entre la protéine et le ligand (calculé par PLP).
<b>Electro</b>	Energie d'interaction électrostatique protéine-ligand pour courte distance interatomique ( $r < 4.5\text{\AA}$ )
<b>ElectroLong</b>	Energie d'interaction électrostatique protéine-ligand pour longue distance interatomique ( $r > 4.5\text{\AA}$ )
$E_{\text{Intra}}$	Energie total MolDock Score interne du ligand (interactions intramoléculaires).
$E_{\text{Intra (steric)}}$	Energie d'interaction stérique du ligand (calculé par PLP).
$E_{\text{Intra (hbond)}}$	Energie de liaison hydrogène du ligand (calculé par PLP).
$E_{\text{Intra (elec)}}$	Energie d'interaction électrostatique du ligand
$E_{\text{Intra (tors)}}$	Energie de torsion du ligand

### 2.4.3. Validation du protocole de docking

Pour valider notre protocole de docking basé sur une combinaison de l'algorithme de recherche Moldock Optimizer et la fonction de score Moldock Score, nous avons évalué la précision de positionnement de la meilleure pose (top-scoring) du ligand de référence "Tropolone". Cette pose est obtenue, parmi plusieurs poses générées, après application du protocole de docking sur le complexe Tropolone-Tyrosinase. La meilleure est choisie selon l'énergie totale d'interaction (MolDock Score) la plus basse d'une valeur de -61,024 Kcal /mol. Cette énergie sera également utilisée comme énergie de référence avec les autres ligands de la chimiothèque à cribler. Cette validation nécessite deux étapes.

La première étape est le «Re-docking», qui a pour but la détermination de la capacité du protocole du docking à générer une pose avec un positionnement, dans le site d'interaction identique ou très proche à celui de la structure cristallographique originale. Cette méthode s'intéresse à la déviation quadratique moyenne «RMSD» (RootMean Square Deviation)<sup>27</sup>, qui est exprimée par la distance entre les positions atomiques du ligand de référence Co-cristallisé expérimentalement (de coordonnées cartésiennes  $x_{ligand}, y_{ligand}, z_{ligand}$ ) par rapport à celles de la structure prédite ligand (pose) obtenues après docking dans le site actif. La déviation RMSD est donnée par l'équation suivante :

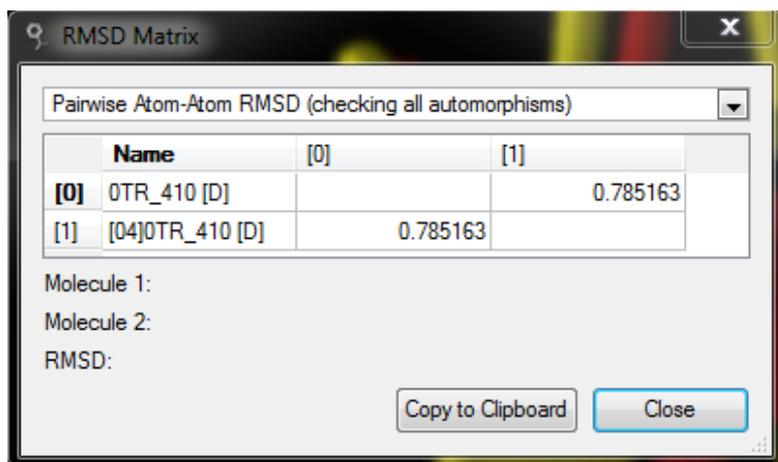
$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{cristal_i} - x_{pose_i})^2 + (y_{cristal_i} - y_{pose_i})^2 + (z_{cristal_i} - z_{pose_i})^2} \quad eq. 4$$

Avec :

- n : le nombre total d'atomes lourds.
- $(x_{cristal_i}, y_{cristal_i}, z_{cristal_i})$  et  $(x_{pose_i}, y_{pose_i}, z_{pose_i})$  sont respectivement les coordonnées cartésiennes des atomes identiques de la structure expérimentale et de la structure prédite.

Il est considéré que les meilleurs protocoles du docking génèrent des valeurs RMSD inférieurs à 2,0Å, cette valeur est un seuil au-dessus duquel la précision du positionnement est jugée mauvaise, entraînant l'invalidation du protocole<sup>28</sup>.

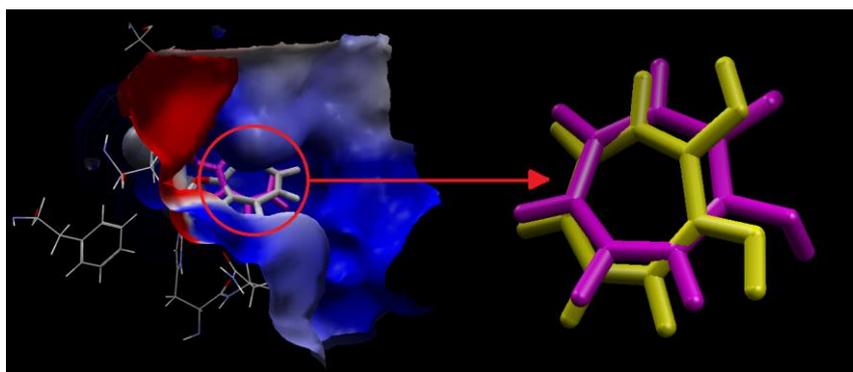
Pour notre étude, le calcul de la valeur RMSD est effectué avec MVD en appliquant le protocole du docking décrit précédemment avec définition du Tropolone comme un ligand de référence. La simulation du re-docking a généré un RMSD équivalent à 0,78 Å (**Figure 8**), largement inférieur au seuil, ce qui indique que notre protocole a réussi et qu'il est applicable pour le reste de cette étude.



	Name	[0]	[1]
[0]	0TR_410 [D]		0.785163
[1]	[04]0TR_410 [D]	0.785163	

**Figure 8** : Matrice de RMSD générée par MDV après re-docking.

La superposition du tropolone et sa meilleure pose générée (**Figure 9**) confirme la précision du positionnement et la génération d'une conformation quasi similaire à celle Co-cristallisée.



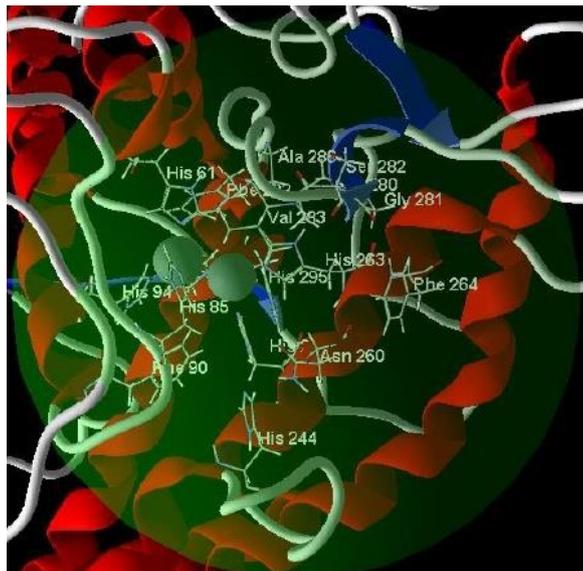
**Figure 9** : La superposition de la structure cristallographique du ligand (en jaune) et la pose obtenue après re-docking (en rose), RMSD =0,78Å.

### 3. Criblage des chimiothèques

#### 3.1. Criblages de la chimiothèques des analogues du gingérol

Le complexe de structure protéique de l'inhibiteur de référence avec l'enzyme tyrosinase (PDB ID : 2Y9X) a été obtenu à partir de la base de données Protein.

La cavité de liaison potentielle au ligand a été prédite à l'aide de Molegro Virtual Docker pour la tyrosinase de champignon qui a un volume de 50,69 Å<sup>3</sup> et une surface de 197,12 Å<sup>2</sup> (**Figure 10**).



**Figure 10 :** La cavité de liaison de l'enzyme tyrosinase avec les résidus Phe264, Phe90, His94, His244, Phe292, Ala286, Val283, Gly281, Met280, Asn260, Ser282, His85, His61, His263, His259, His296 et les ions cuivre représentés par deux boules argentées.

Le criblage débute par l'importation de tous les ligands de notre chimiothèque de 1800 analogues du gingérol dans le «work space» du MVD, les ligands subissent une post-préparation avec l'outil de préparation automatique du logiciel afin de vérifier et de s'assurer que les propriétés structurales de chaque ligand sont prises en compte par le logiciel.

Tous les ligands seront orientés vers l'espace de recherche déjà défini en appliquant le protocole précédemment décrit et validé. Le «*scoring*» des poses obtenues pour chaque ligand (après 100 tours de recherche) mène à un classement des meilleures poses basé sur leurs énergies d'interaction «*Moldock score*». Ainsi, pour chaque élément de la chimiothèque, une seule pose prédite est sélectionnée (**Figure 11**).

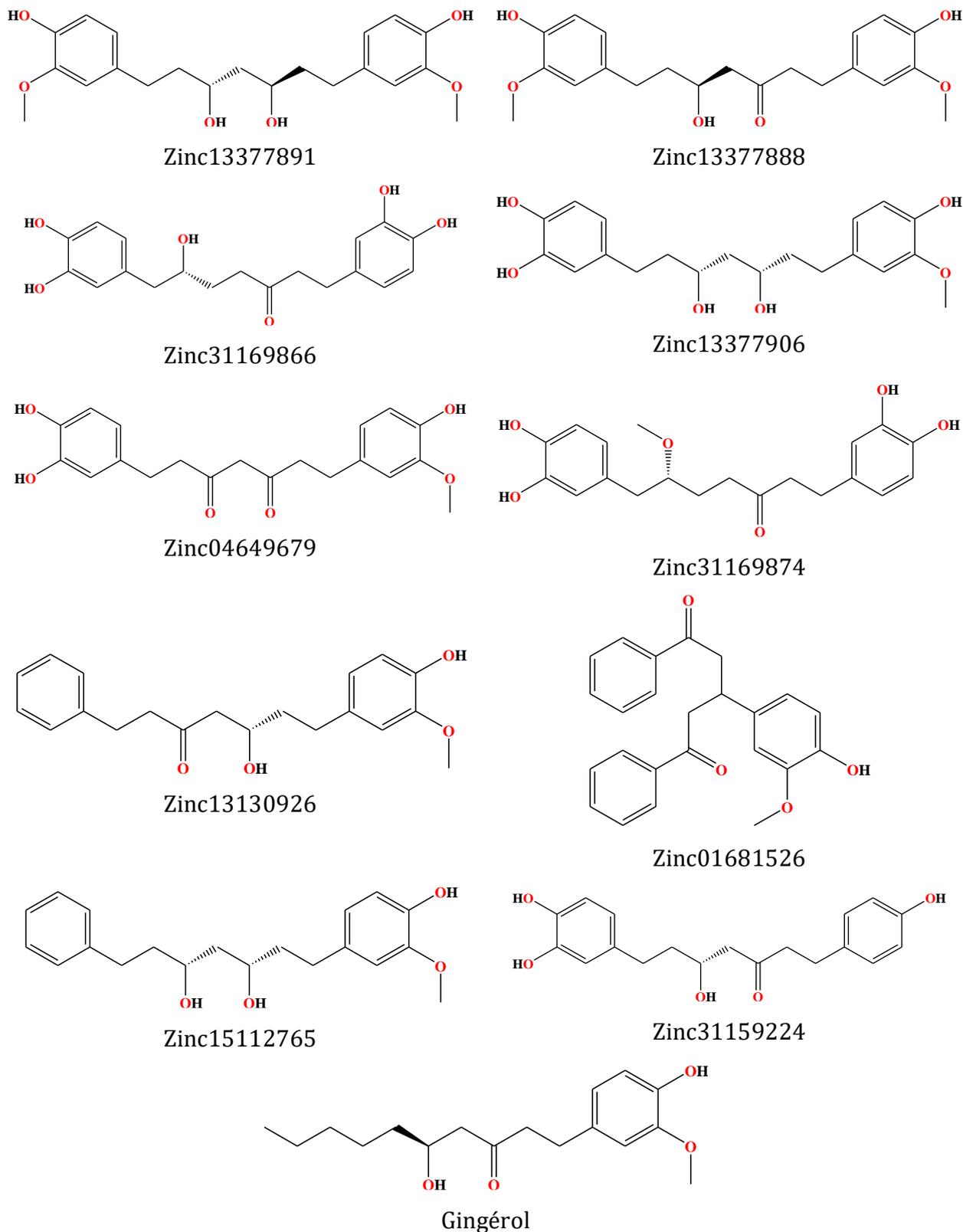


Figure 11 : Structures moléculaires des dix meilleurs hits classés par docking moléculaire et celle du gingérol.

## Chapitre I : Criblage par Docking de molécules dérivées d'arbutine et de gingérol

Le docking moléculaire a été effectué et les dix meilleures poses ancrées dans la région du site actif de la tyrosinase ont été classées dans un ordre croissant de l'énergie d'interaction MolDock (**Tableau 5**).

**Tableau 5** : Les dix meilleurs hits classés par docking moléculaire avec inhibiteurs de tyrosinase connus expérimentalement.

Compound ID	E-Inter (cof - lig) <sup>a</sup> (kcal.mol <sup>-1</sup> )	E-Inter (pro - lig) <sup>b</sup> (kcal.mol <sup>-1</sup> )	E-Inter total <sup>c</sup> (kcal.mol <sup>-1</sup> )	MolDock score (kcal.mol <sup>-1</sup> )	H-bond energy <sup>d</sup> (kcal.mol <sup>-1</sup> )
ZINC13377891	-3.93	-164.30	-168.23	-157.15	-8.12
ZINC13377888	-4.81	-162.68	-167.49	-155.86	-4.55
ZINC31169866	-5.14	-157.80	-162.94	-150.75	-15.09
ZINC13377906	-5.077	-158.57	-163.65	-149.68	-16.16
ZINC04649679	-2.47	-153.36	-155.83	-142.06	-7.87
ZINC31169874	-4.38	-149.35	-153.73	-141.75	-10.62
ZINC13130926	-4.79	-146.99	-151.78	-139.13	-6.20
ZINC01681526	-3.39	-142.74	-146.13	-137.53	-3.06
ZINC15112765	-4.42	-155.01	-159.43	-137.49	-6.80
ZINC31159224	-4.37	-150.55	-154.93	-135.97	-11.83
Gingerol	-4.18	-130.67	-134.85	-130.67	-1.43
Kojic acid	-1.85	-81.91	-81.76	-81.26	-4.63
Hydroquinone	-3.09	-63.70	65.87	-58.25	-4.28
tropolone	-2.18	-75.94	-79.04	-77.09	-3.11

<sup>a</sup>: Energie d'interaction totale ligand-cofacteur.

<sup>b</sup>: Energie d'interaction totale pose-protein.

<sup>c</sup>: Energie d'interaction totale.

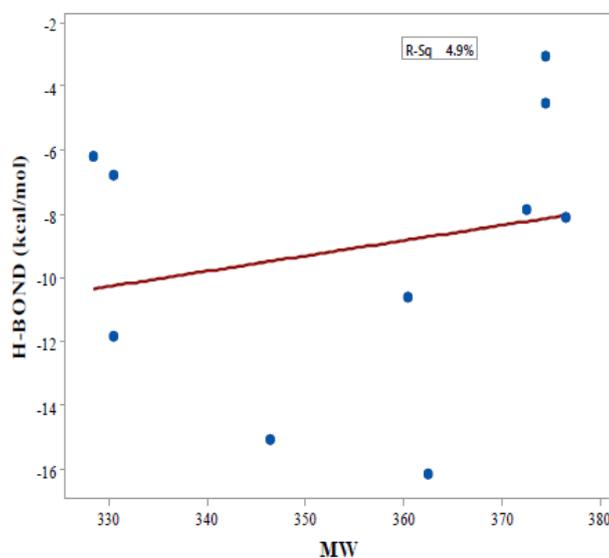
<sup>d</sup>: : Energie de liaison Hydrogen.

D'après le résultat d'amarrage obtenu, on remarque que les analogues du gingérol (zinc13377891, zinc13377888, zinc31169866, zinc13377906, zinc04649679, zinc31169874, zinc13130926, zinc01681526, zinc15112765, zinc31159224) ont des scores MolDock plus élevés (en termes d'énergie négative) que le gingérol et les inhibiteurs de tyrosinase connus expérimentalement.

Ces dix ligands représentent, du point de vue énergétique, une bonne affinité avec les résidus du site actif (**tableau 6**). Cette affinité est due essentiellement à la présence des interactions de liaisons hydrogène représentées par une valeur énergétique «H<sub>bond</sub> Score», et des interactions hydrophobiques de type Van Der Walls (VDW).

Les analogues du gingérol les mieux classés présentent un niveau élevé d'énergie d'interaction ligand-protéine par rapport au gingérol et aux inhibiteurs de référence de la tyrosinase, comme l'acide kojique<sup>29</sup>, l'hydroquinone<sup>30</sup> et la tropolone<sup>31</sup>.

La corrélation entre l'énergie de liaison hydrogène et le poids moléculaire pour les composés les mieux classés est calculée et représentée sur la **figure 12**. Une faible valeur de coefficient de détermination (0,049) a été obtenue, ce qui indique que l'énergie de liaison hydrogène prédite était principalement due aux caractéristiques de structure et pas à cause de sa taille moléculaire.



**Figure 12** : Tracé linéaire ajusté de la corrélation entre les valeurs de l'énergie d'interaction et le poids moléculaire.

Afin d'avoir un aperçu structurel approfondi de l'interaction entre les dix principaux composés sélectionnés et notre cible, une analyse plus approfondie basée à la fois sur les liaisons d'interaction et hydrogène ont été effectuées (**Tableau 6**).

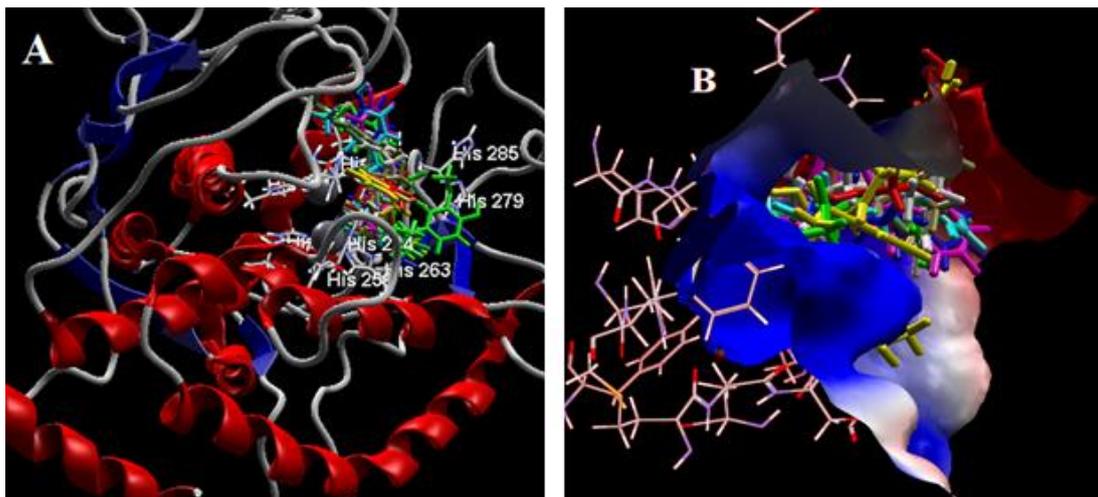
**Tableau 6:** Valeurs d'énergie d'interaction et distances des liaisons hydrogène entre les ligands et les résidus présents dans l'enzymes pour les meilleurs résultats.

Compound ID	Résidus	Distance d'interaction (Å)	Energie d'interaction (kcal.mol <sup>-1</sup> )
Zinc13377891	Ala323	2.86	-0.64
	Asn260	2.95	-1.46
	His240	3.39	-1.03
	His240	2.74	-2.50
	Met280	2.99	-2.50
Zinc13377888	Ala323	3.05	-2.50
	Asn260	2.96	-1.17
	Asn260	3.40	-0.99
	His85	3.49	-0.54
	Met280	2.60	-2.50
	Ser282	3.57	-0.02
Zinc31169866	Glu322	3.01	-2.30
	Glu322	3.11	-2.19
	Glu322	3.23	-1.48
	Thr84	3.53	-0.11
	Asn320	2.60	-0.59
	Val283	3.20	-2.00
	Met280	2.77	-2.50
	Asn260	3.20	-1.42
	Glu256	3.09	-2.50
Zinc13377906	Val283	3.20	-1.99
	Met280	2.72	-2.50
	Asn260	2.89	-1.82
	His244	3.00	-2.46
	His244	2.86	-2.50
	Cys83	2.60	-2.38
Zinc04649679	His85	3.04	-2.50
	Asn260	2.87	-0.37

## Chapitre I : Criblage par Docking de molécules dérivées d'arbutine et de gingérol

	Cys83	2.75	-2.50
	His85	3.05	-2.50
	Met280	2.99	-2.50
Zinc31169874	Asn260	2.98	-0.83
	Val283	3.56	-0.19
	Asn81	3.06	-2.50
	Cys83	2.60	-2.50
	Cys83	3.33	-1.37
	Met280	2.76	-2.08
	Zinc13130926	Asn260	3.07
Asn260		2.77	-2.50
Met280		2.60	-2.50
Ser282		3.55	-0.02
Zinc01681526	Asn260	2.91	-0.93
	Ser282	3.41	-0.93
	Val283	2.92	-2.13
Zinc15112765	His244	3.10	-2.50
	Asn260	3.10	-1.79
	Met280	2.60	-2.46
	Glu256	2.99	-0.04
	Ser282	3.52	-0.02
Zinc31159224	Gly245	3.05	-2.50
	Asn260	3.01	-1.85
	His244	2.98	-2.50
	Met280	3.09	-2.50
	Met280	2.60	-2.47
Gingérol	Met280	3.28	-1.59

Le **tableau 7** répertorie toutes les principales interactions prédites entre les dix top hits et les résidus du site actif de la tyrosinase, présentant l'énergie d'interaction ligand-protéine, les résidus présents et les distances d'interaction. Les modes de liaison des top hits révèlent que ces derniers se sont avérés ancrés dans le site de liaison de la tyrosinase (**Figure 13**).

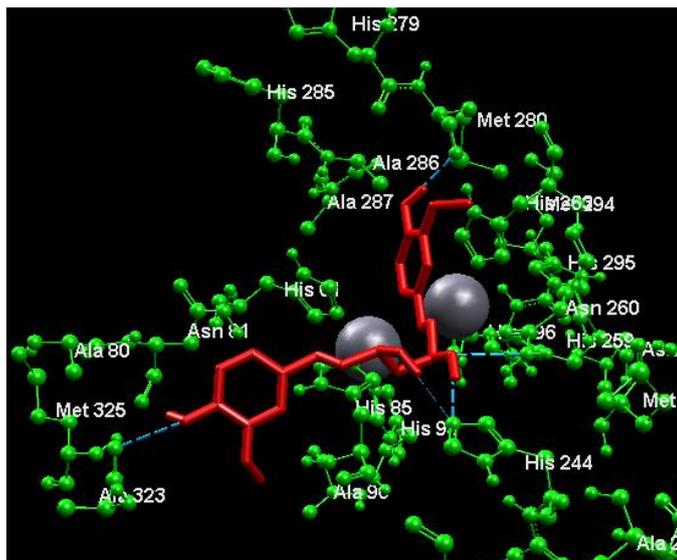


**Figure 13 :** a) Top dix hits ancrés dans le site actif de la tyrosinase et b) vue de surface électrostatique de la tyrosinase montrant les dix premiers hits d'amarrage à l'intérieur du site de liaison.

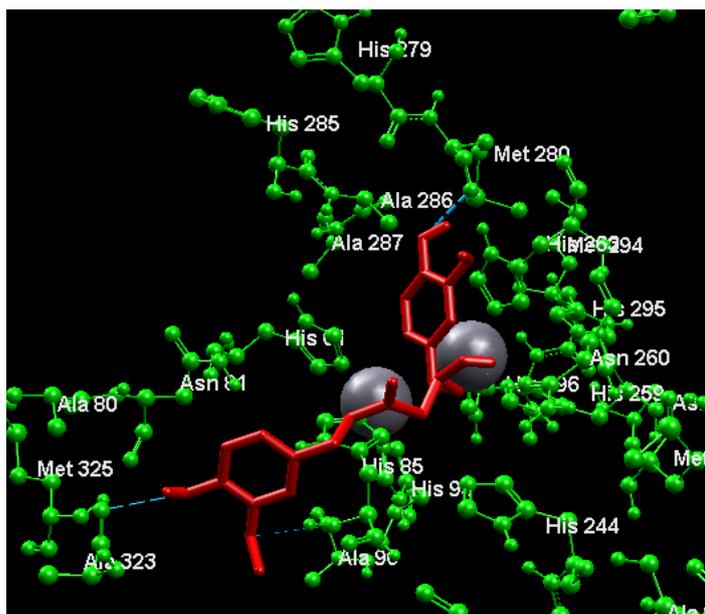
Selon les différents modes de liaison des dix top hits, on observe que les interactions avec les acides aminés Met280 et Asn260 du site actif de la tyrosinase sont les plus prédominantes parmi les top hits. On observe également que les deux premiers hits Zinc13377891 et Zinc13377888 ont une interaction commune avec les résidus Met280, Ala323 et Asn260 de la tyrosinase.

La pose la mieux classée Zinc13377891 forme cinq liaisons hydrogène avec quatre résidus d'acides aminés (**Figure 14**). L'atome d'oxygène (ID 28) du groupe hydroxyle a médié une liaison hydrogène avec l'atome N (ID2129) d'Asn260 à une distance de 2,59 Å. Le même atome de la pose la mieux notée est également impliqué dans une interaction de liaison hydrogène avec l'atome d'azote (ID 2018) de His244 à une distance de 2,74 Å. L'atome d'oxygène (ID27) du groupe hydroxyle a formé une liaison hydrogène avec l'atome d'azote (ID 2018) de His244 à une distance de 3,39 Å.

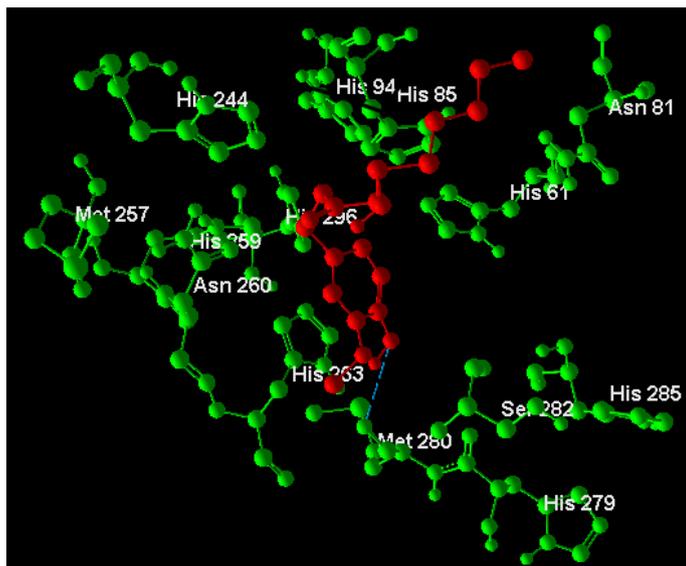
Enfin, deux liaisons hydrogène se sont formées entre les atomes d'oxygène (ID8 et 26) des atomes d'oxygène des groupes carbonyle (ID 2635 et ID2228) de Ala323 et Met280 respectivement. Les positions dans le site actif de la tyrosinase et les modes de liaison des deux principaux composés (ZINC13377891 et ZINC13377888) et du gingérol sont illustrés respectivement à la **figure 14**, à la **figure 15** et à la **figure 16**. La comparaison entre les modes de liaison de ces composés révèle que l'affinité de liaison du ZINC13377891 et du ZINC13377888 avec les résidus du site actif de la tyrosinase est plus favorable que celle du gingérol.



**Figure 14:** Interactions prédites des liaisons H (lignes pointillées soufflées) reliant ZINC13377891 (rouge) aux résidus Ala323, His244, Asn260, Met280 de l'enzyme tyrosinase.



**Figure 15:** Interactions prédites des liaisons H (lignes pointillées soufflées) entre ZINC13377888 (rouge) et Ala323, His85, Asn260, Met280 résidus de l'enzyme tyrosinase.



**Figure 16:** Interactions prédites des liaisons H (lignes pointillées soufflées) entre le gingérol (rouge) et le résidu Met280 de l'enzyme tyrosinase.

La règle de cinq, décrite par Lipinski et al<sup>32,33</sup>, stipule que la plupart des molécules de type médicament ont des valeurs de LogP (le logarithme du coefficient de partage octanol/eau) <5, MW (poids moléculaire) <500, HBA (nombre d'hydrogène accepteurs de liaisons) <10 et HBD (nombre de donneurs de liaisons hydrogène) ≤5. Les molécules ne respectant pas ces règles peuvent avoir des problèmes de biodisponibilité<sup>34</sup>.

Le tableau 7 représente les valeurs des paramètres de la règle de cinq pour les hits les mieux classés, le gingérol et les inhibiteurs de tyrosinase connus expérimentalement. Selon les valeurs indiquées dans **le tableau 8**, tous les hits les mieux classés utilisés dans ce travail respectent la règle de Lipinski et peuvent être des composés actifs.

**Tableau 8 :** Valeurs des paramètres de la règle de Lipinski pour les top hits et le gingérol.

Compound ID	LogP <sup>a</sup>	MW <sup>b</sup>	HBA <sup>c</sup>	HBD <sup>d</sup>	TPSA <sup>e</sup>	ROT-B <sup>f</sup>
ZINC13377891	2.05	376.443	6	4	99.38	10
ZINC13377888	1.97	374.428	6	3	96.22	10
ZINC31169866	1.52	346.374	6	5	118.22	8
ZINC13377906	1.83	362.417	6	5	110.38	9
ZINC04649679	4.16	372.412	4	2	66.76	10

## Chapitre I : Criblage par Docking de molécules dérivées d'arbutine et de gingérol

ZINC31169874	1.89	360.401	6	2	93.06	9
ZINC13130926	1.75	328.402	6	4	107.22	9
ZINC01681526	2.77	374.429	4	2	66.76	8
ZINC15112765	2.85	330.418	4	3	69.92	9
ZINC31159224	3.21	330.375	4	1	55.76	8
gingerol	3.23	294.386	4	2	66.00	10

<sup>a</sup>: Le logarithme du coefficient de partage octanol/eau.

<sup>b</sup>: Le poids moléculaire.

<sup>c</sup>: Le nombre d'accepteurs de liaisons hydrogène.

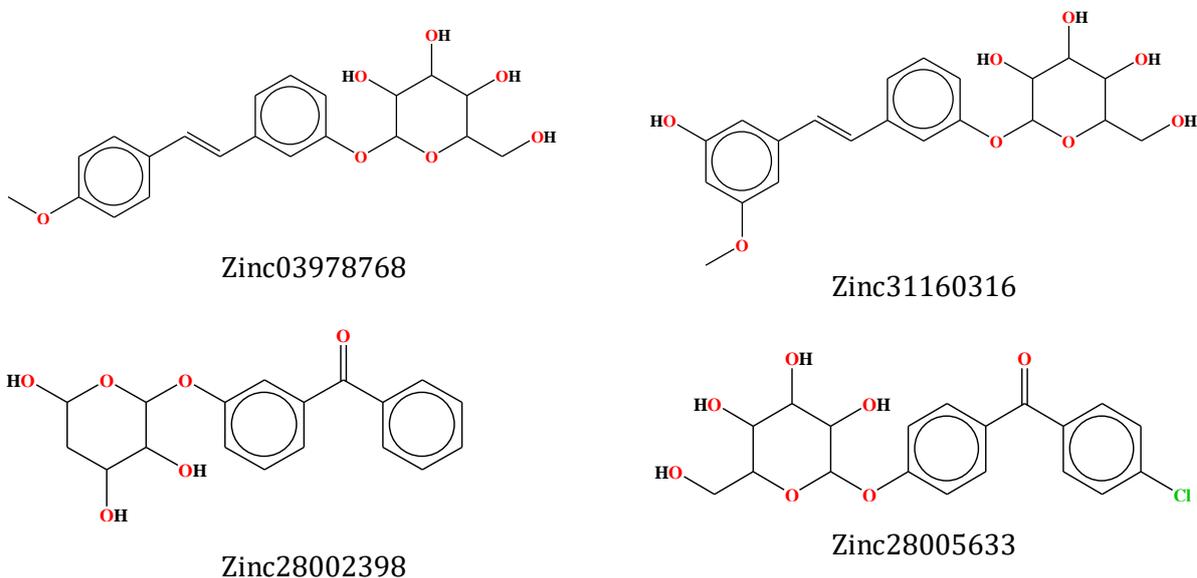
<sup>d</sup>: le nombre de donneurs de liaisons hydrogène.

<sup>e</sup>: La surface polaire topologique.

<sup>f</sup>: Le nombre de liaisons rotatives.

### 3.2. Criblage de la chimiothèque des analogues de l'arbutine :

Après l'importation de tous les ligands de chimiothèque (1268 structures) dans le logiciel MVD, les mêmes étapes réalisées dans le cas de criblage des analogues du gingérol sont répétées pour les analogues de l'arbutine (**Figure 17**).



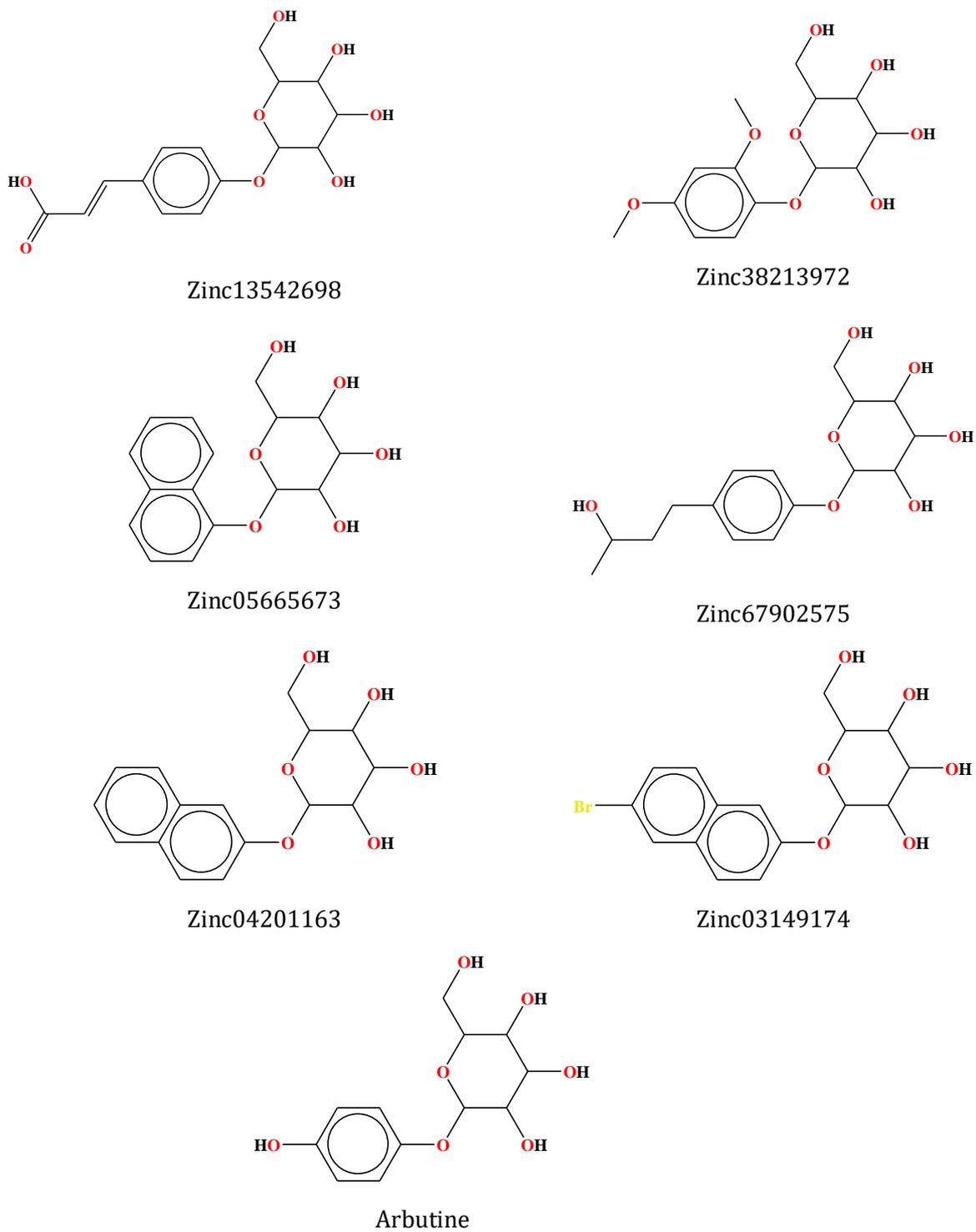


Figure 17 : Structures moléculaires des dix meilleurs hits classés par docking moléculaire et celle de l'arbutine.

Le docking moléculaire a été effectué et les meilleures poses ancrées dans le site actif de la tyrosinase ont été classées en fonction de l'énergie du score MolDock (**tableau 8**).

D'après le résultat d'amarrage obtenu, nous remarquons que les analogues de l'arbutine (Zinc03978768, Zinc31160316, Zinc28002398, Zinc28005633, Zinc13542698, Zinc38213972, Zinc05665673, Zinc67902575, Zinc04201163, Zinc03149174) ont des scores d'énergie MolDock plus élevés que celles des inhibiteurs de la tyrosinase connus expérimentalement.

**Tableau 8** : Score d'amarrage pour les dix premiers hits classés avec des inhibiteurs de tyrosinase connus expérimentalement.

CompoundID	E-Inter(cof-lig) <sup>a</sup> (kcal.mol <sup>-1</sup> )	E-Inter(pro-lig) <sup>b</sup> (kcal.mol <sup>-1</sup> )	E-Inter total <sup>c</sup> (kcal.mol <sup>-1</sup> )	MolDockscore (kcal.mol <sup>-1</sup> )	H-bond energy <sup>d</sup> (kcal.mol <sup>-1</sup> )
Zinc03978768	-3.451	-155.717	-159.168	-138.301	-8.659
Zinc31160316	-5.056	-149.988	-155.044	-127.429	-5.000
Zinc28002398	-3.984	-142.775	-146.758	-125.788	-4.923
Zinc28005633	1.193	-147.886	-146.693	-123.878	-9.988
Zinc13542698	0.723	-134.729	-134.007	-114.494	-11.609
Zinc38213972	-4.865	-125.419	-130.284	-114.467	-8.652
Zinc05665673	0.000	-133.558	-148.413	-114.272	-14.855
Zinc67902575	-4.684	-131.776	-136.460	-113.987	-10.00
Zinc04201163	-2.389	-130.830	-133.219	-109.393	-9.147
Zinc03149174	-1.369	-123.643	-131.544	-107.582	-7.901
Arbutin	-2.411	-111.299	-113.709	-89.877	-12.752
Kojicacid	-5.043	-78.238	-83.282	-70.911	-6.219
Tropolone	0.269	-70.428	-70.159	-61.024	-2.104

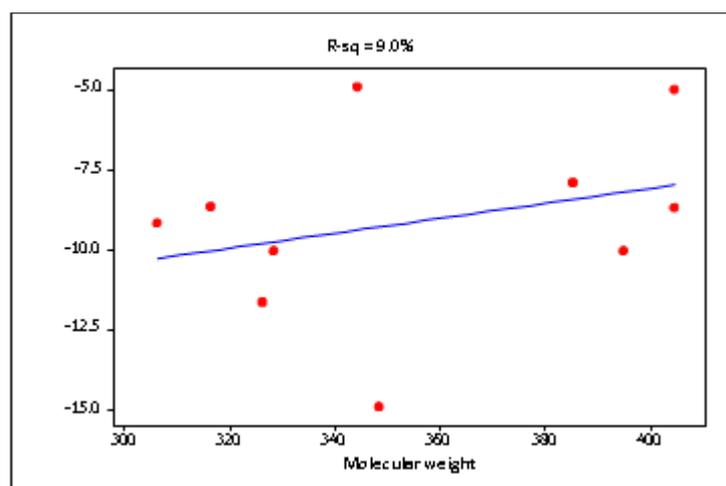
a : L'énergie totale d'interaction ligand-cofacteur.

b : L'énergie totale d'interaction pose-protéine.

c : L'énergie d'interaction totale.

d : énergie de liaison hydrogène.

La corrélation entre l'énergie de liaison hydrogène et le poids moléculaire pour les composés les mieux classés est calculée et illustrée à la **figure 18**. Une faible valeur de coefficient de détermination (0,09) a été obtenue, ce qui indique que l'énergie de liaison hydrogène prévue était due aux caractéristiques de structure et pas à cause de sa taille moléculaire.



**Figure 18 :** Tracé linéaire ajusté de la corrélation entre l'énergie d'interaction et le poids moléculaire.

Afin d'avoir un aperçu structurel approfondi de l'interaction entre les dix principaux composés sélectionnés et notre cible, une analyse plus approfondie basée à la fois sur la liaison électrostatique et hydrogène a été effectuée pour les meilleures poses des composé classés. Les résultats de cette analyse sont regroupés dans le **tableau 9**.

**Tableau 9 :** Structure du ligand, résidus présents, énergie d'interaction ligand-protéine et distances d'interaction pour les meilleurs résultats.

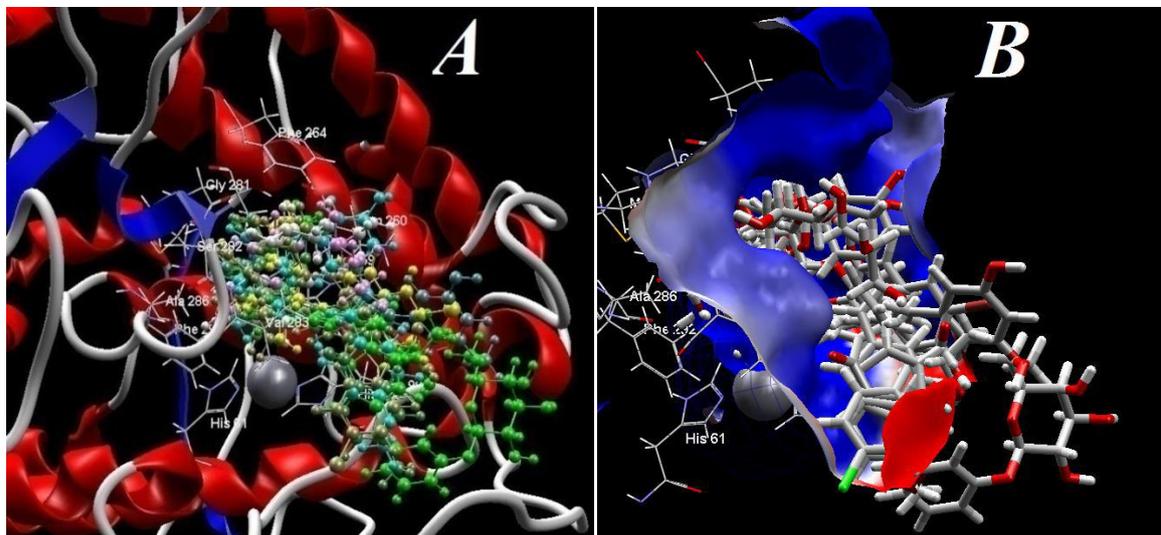
CompoundID	Résidus	Distance d'interaction (Å)	Energie d'interaction (kcal.mol <sup>-1</sup> )
Zinc03978768	Met280	3.00	-2.50
	Met280	2.55	-2.12
	His263	2.94	-0.31
	His61	3.35	-1.23
	His85	3.10	-2.50
Zinc31160316	Met280	2.82	-2.50
	Met280	2.86	-2.50
Zinc28002398	Met280	3.12	-2.42
	Met280	2.60	-2.50
Zinc28005633	Met280	3.02	-2.50
	Met280	2.60	-2.50
	His85	2.85	-2.50
	His61	3.10	-2.49

## Chapitre I : Criblage par Docking de molécules dérivées d'arbutine et de gingérol

Zinc13542698	Glu322	2.95	-2.50
	His85	3.24	-1.82
	His85	3.47	-0.63
	His61	2.87	-2.50
	Ser282	3.34	-0.43
	Val283	2.96	-1.47
	Met280	2.27	-0.25
	Asn260	3.10	-2.50
Zinc38213972	Met280	3.36	-1.19
	Val283	3.11	-2.46
	Asn260	2.93	-2.50
	Gly281	3.10	-2.50
Zinc05665673	Gly86	3.42	-0.92
	Gly86	3.48	-0.61
	Met319	2.64	-2.50
	Arg321	2.58	-2.30
	Arg321	3.28	-1.03
	Ala246	2.84	-2.50
	Glu322	2.67	-2.50
	Glu322	3.09	-2.50
Zinc67902575	Met280	2.90	-2.50
	Met280	2.83	-2.50
	Glu322	2.68	-2.50
	His85	2.69	-2.50
Zinc04201163	Glu256	2.63	-2.50
	Asn260	3.40	-1.02
	Asn260	3.04	-2.50
	Asn260	3.41	-0.96
Zinc03149174	His244	3.17	-2.17
	His85	2.60	-2.50
	His61	3.48	-0.59
	His263	3.25	-1.76
	Met280	2.37	-0.55
Arbutine	Asn260	3.03	-2.50
	Asn260	2.95	-2.50
	Met280M	3.50	-0.25
	et280	2.61	-2.50
	His85	2.60	-2.50
	His244	2.96	-2.50
	His244	3.10	-2.50

Le **tableau 9** rassemble les principales interactions prédites entre les dix top hits et les résidus du site actif de la tyrosinase, exposant l'énergie d'interaction

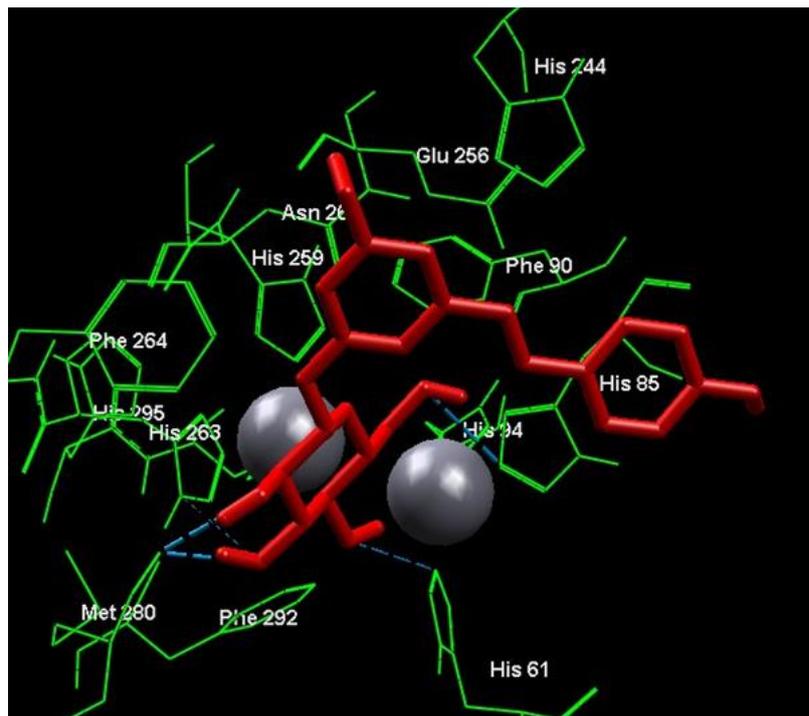
ligand-protéine, les résidus présents et les distances d'interaction. Les modes de liaison des top hits révèlent que ces derniers se sont avérés ancrés dans le site de liaison de la tyrosinase (**Figure 19**).



**Figure 19** : a) Top dix hits ancrés dans le site actif de la tyrosinase et b) vue de surface électrostatique de la tyrosinase montrant les dix premiers hits d'amarrage à l'intérieur du site de liaison.

Selon les différents modes de liaison des dix top hits, on observe que les interactions avec les acides aminés Met280, His85, His61 et Asn260 du site actif de la tyrosinase sont les plus prédominantes parmi les top hits. On observe également que les quatre premiers hits Zinc03978768, Zinc28005633, Zinc13542698 et Zinc03149174 ont une interaction commune avec les résidus Met280, His85 et His61 de la tyrosinase.

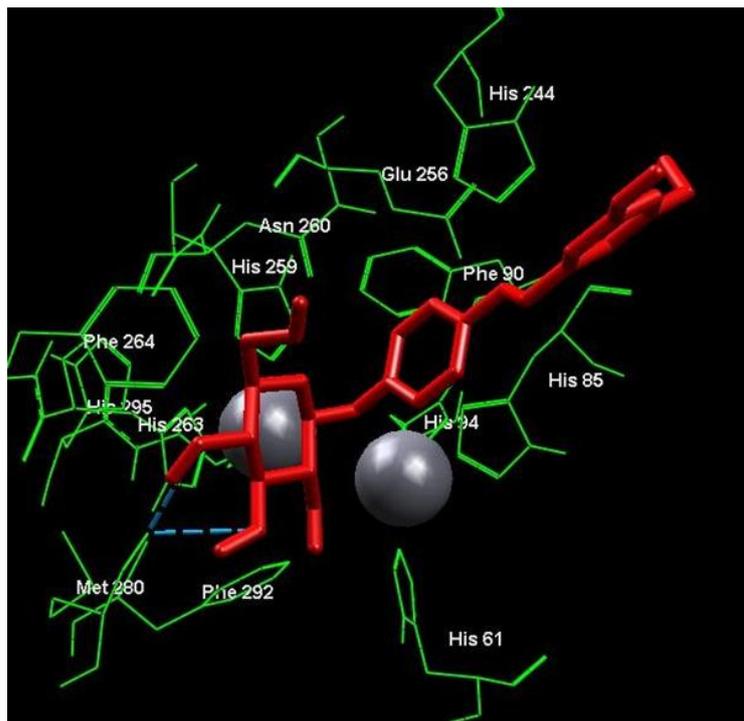
La pose la mieux notée Zinc03978768 forme cinq liaisons hydrogène avec quatre résidus d'acides aminés (**Figure 20**).



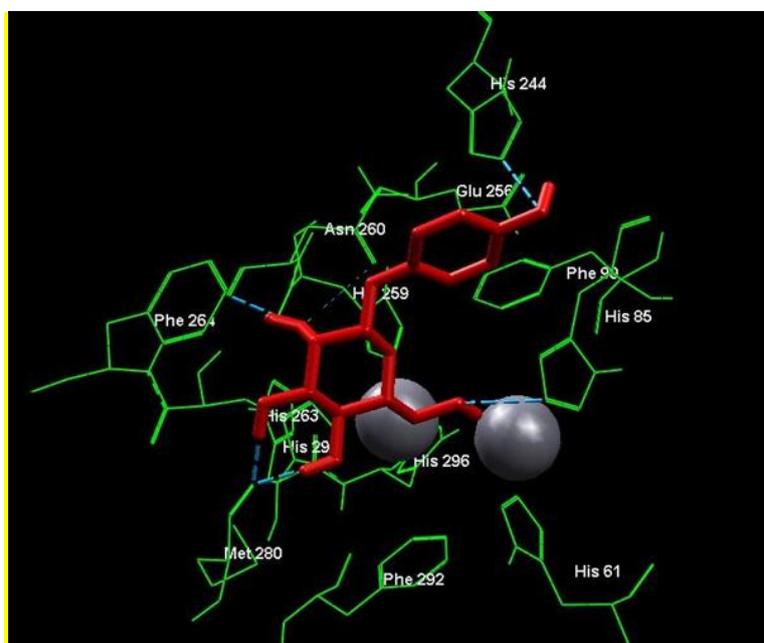
**Figure 20** : Intéractions prédites des liaisons H (lignes pointillées bleues) entre les résidus Zinc03978768 (rouge) et His263, His61, His85, Met280 de l'enzyme tyrosinase.

L'atome d'hydrogène (ID 50) du groupe hydroxyle (ID 31) a formé une liaison hydrogène avec l'atome d'oxygène (ID 2280) de Met280 à une distance de 2,55 Å. L'atome d'oxygène du même groupe hydroxyle (ID 31) de la pose est également impliqué dans une interaction de liaison hydrogène avec l'atome d'Hydrogène (ID 5250) de His263 à une distance de 2,94 Å. L'atome d'Hydrogène (ID 51) du groupe hydroxyle (ID 32) a formé une liaison hydrogène avec l'atome d'oxygène (ID 2280) de Met280 à une distance de 3,00 Å. Enfin, deux liaisons hydrogène ont été formées entre les atomes d'Hydrogènes (ID 48 et 49) avec des atomes d'azote (ID 484 et ID 684) des résidus His61 et His85 respectivement.

Les positions dans le site actif de la tyrosinase et les modes de liaison des deux principaux composés (Zinc03978768 et Zinc31160316) et de l'arbutine sont illustrés respectivement aux **figures 20, 21** et **22**. La comparaison entre les modes de liaison de ces composés révèle que l'affinité de liaison de Zinc03978768 et Zinc31160316 avec les résidus du site actif de la tyrosinase est plus favorable que celle de l'arbutine.



**Figure 21** : Interactions prédites des liaisons H (lignes pointillées bleues) reliant Zinc31160316 (rouge) au résidu Met280 de l'enzyme tyrosinase.



**Figure 22** : Interactions prédites des liaisons H (lignes pointillées bleues) entre l'arbutine (rouge) et les résidus Met280, His85, Asn260, His244 de l'enzyme tyrosinase.

Le **tableau 10** représente les valeurs des paramètres de la règle de cinq pour les hits les mieux classés, l'arbutine et les inhibiteurs de tyrosinase connus expérimentalement. Selon les valeurs indiquées dans le **tableau 10**, tous les hits les mieux classés utilisés dans ce travail respectent la règle de Lipinski et peuvent être des composés actifs.

**Tableau 10** : Valeurs des paramètres de la règle de Lipinski pour les top hits et l'arbutine.

CompoundID	LogP <sup>a</sup>	MW <sup>b</sup>	HBA <sup>c</sup>	HBD <sup>d</sup>	TPSA <sup>e</sup>	ROT-B <sup>f</sup>
Zinc03978768	1.74	404.415	8	5	129	6
Zinc31160316	1.74	404.415	8	5	129	6
Zinc28002398	1.98	344.363	6	3	96	4
Zinc28005633	1.80	394.807	7	4	116	5
Zinc13542698	-0.36	325.293	8	4	140	5
Zinc38213972	-0.29	316.306	8	4	118	5
Zinc05665673	1.40	348.351	7	4	108	5
Zinc67902575	0.10	328.361	7	5	120	6
Zinc04201163	0.85	306.314	6	4	99	3
Zinc03149174	1.64	385.21	6	4	99	3
Arbutin	-0.81	272.253	7	5	119	3

a : Le logarithme du coefficient de partage octanol/eau.

b : Le poids moléculaire.

c : Le nombre d'accepteurs de liaisons hydrogène.

d : le nombre de donneurs de liaisons hydrogène.

e : La surface polaire topologique.

F : Le nombre de liaisons rotatives.

### 4. Conclusion

Dans cette partie de travail nous avons appliqué un criblage virtuel, basé sur une étude de docking moléculaire, sur des analogues de gingérol et de l'arbutin extraits de la base de données ZINC afin d'étudier leurs mode d'interaction vis-à-vis de notre cible d'intérêt biologique (la tyrosinase) et d'enrichir cette classe thérapeutique. La structure tridimensionnelle de cette enzyme de souche fongique est obtenue à partir de la banque des données PDB avec l'identifiant 2Y9X en complexe avec le « tropolone ». Le processus du docking est assisté par le programme Molegro Virtual Docker (MVD) avec un protocole basé essentiellement sur l'algorithme évolutionnaire différentiel « MolDock Optimizer » et la fonction de score « MolDock Score » avec une population de 100 individus durant 100 tours de docking. La validation de ce protocole a été réalisée par un re-docking du ligand de référence co-cristallisé avec un RMSD de 0.78Å.

Les résultats obtenus à partir du docking moléculaire indiquent que certains analogues ont montré un score de Moldock plus élevé (en termes d'énergie négative) par rapport au gingérol et aux inhibiteurs de tyrosinase connus expérimentalement. Les dix top hits des analogues du gingérol ont montré une meilleure interaction que le gingérol, montrant une interaction moléculaire commune avec les résidus Met280 et Asn260 de l'enzyme tyrosinase.

Les dix top hits des analogues de l'arbutin ont montré une meilleure interaction que l'arbutine, montrant une interaction moléculaire commune avec les résidus Met280, His85, His61 et Asn260 de l'enzyme tyrosinase. De plus, les dix hits ancrés, de deux familles étudiées, respectent la règle des cinq de Lipinski, affirmant que ces molécules n'ont aucun problème de biodisponibilité.

## 5. Références

1. B. Seo, J. Yun, S. Lee, M. Kim, K. Hwang, J. Kim, ; K. R. Min, Y. Kim, D. Moon, *Planta Med*, 65, **1999**, 683-686.
2. K. Shimizu, R. Kondo, K. Sakai, S-H. Lee, H. Sato, *Planta Med*, 64, **1998**, 408-412.
3. H. Zhao, D. J. Eling, E. E. Medrano, R. E. Boissy, *J. Invest. Dermatol*, 106, **1996**, 744-752.
4. R.A. Spritz, L. Ho, M. Furumura, *J. Invest. Dermatol*, 109, **1997**, 207-212.
5. V. J. Hearing, *Methods in Enzymology*, 142, **1987**, 154-167.
6. C. Gedermann, C. Eicken, B. Krebs, *Acc. Chem. Res*, 35, **2002**, 183-191.
7. H-S. Lee, *J. Agric. Food Chem*, 50, **2002**, 1400-1403.
8. S. OKOMBI, Recherche et étude de molécules à activité antityrosinase et leur utilisation comme agents dépigmentants en dermocosmétique, Thèse doctorat, Université Joseph FOURIER-Grenoble 1 - UFR de Pharmacie, France.
9. Barbarit I- les dépigmentants : des corticoïdes à l'acide kojique. Thèse Doct. Pharm, **1999**, Nantes
10. A. Rescigno, F. Sollai, B. Pisu, A. Rinaldi, E. Sanjust, *J. Enzyme. Inhib. and Med. Chem*, 17 (4), **2002**, 207-218.
11. C-S Chen, K-J. Liu, Y-H. Lou, C-J. J. Shieh, *Sci. Food Agric*, 82, **2002**, 601-605.
12. H-C. HUANG, S-H. CHIU, and T-M. CHANG, 75(6), **2011**, 1067-1072.
13. F. Yu, Z. Pan, Qu. Bingliang, Yu. Xiangyong, Xu. Kaihang, D. Yuewen, L. Feilong, *Gene*, 656(1), **2018**, 1-8.
14. K. U. Zaidi, S. A. Ali and A. S. Ali, *J. Microsc. Ultrastruct*, 5(2), **2017**, 82-89.
15. [www.rcsb.org](http://www.rcsb.org).
16. W. T. Ismaya, H. J. Rozeboom, A. Weijn, J. J. Mes, F. Fusetti, H. J. Wichers and B. W. Dijkstra, *Biochemistry*, 50, **2011**, 5477-548.
17. W. F. D. A. G. Heberlé, *Curr. Med. Chem*, 18, **2011**, 1339 -1352.
18. M. H. C. R. Thomsen, *J Med Chem*, 11, **2006**, 3315-3321.
19. J. A. L. J. Q. Araújo, A. da C. Pinto, R. B. de Alencastro, M. G. Albuquerque and J. Mol. Model, 17, **2011**, 1401-1412.

20. T. P. M. G. *System, ed. L. Schrodinger*, 1.8 edn.
21. R. Storn and K. Price, Tech-report, International Computer Science Institute, Berkley, **1995**.
22. zinc15.docking.org
23. W. T. Ismayaa, O. M. Tandrasasmitaa, S. Sundaria, A. Diana, X. Lai, D. S. Rentnoningrum, B.W. Dijkstra, R. R. Tjandrawinata and H. Rachmawati, *Int. J. Biol. Macromol*, 102, **2017**, 308-314.
24. G. V. D. K. Gehlhaar, P. A. Rejto, D. B. Fogel, L. J. Fogel, S. T. Freer, **1995**, 615-627.
25. D. B. D. K. Gehlhaar, P. A. Rejto, **1998**, 449-461.
26. J. M. Yang and C. C. Chen, *Proteins*, 55, **2004**, 288-304.
27. J. Kirchmair, P. Markt and S. Distinto, *J Comput Aided Mol Des*, 22, **2008**, 213-228.
28. J. L. B. R. A. Friesner, R. B. Murphy, T. A. Halgren, J. J. Klicic, *J Med Chem*, 47, **2004**, 1739-1749.
29. R. Thomsen and M. Christensen, *J. Med. Chem*, 49, **2006**, 3315.
30. B. K. Singh, S. H. Park, H. B. Lee, Y. H. Goo, H. S. Kim, S. H. Cho, J. H. Lee, G. W. Ahn, J. P. Kim, S. M. Kang and E. K. Kim, *Ann. Dermatol*, 28, **2016**, 555.
31. T. Pillaiyar, M. Manickam and V. Namasivayam, *J. Enzyme. Inhib. Med. Chem*, 32(1), **2017**, 403-425.
32. R. Nairn, W. Cresswell and J. Nairn, *Biochem. Mol. Biol. Educ*, 43, **2015**, 370.
33. C. A. Lipinski, *J. Pharmacol. Toxicol. Methods*, 44, **2000**, 235.
34. C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug. Del. Rev.*, 23(3), **1997**, 3-25.
35. M. P. Freitas, *Org. Biomol. Chem*, 4, **2006**, 1154.

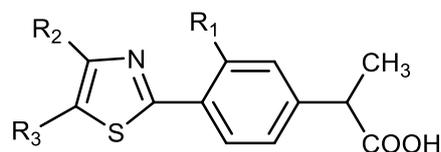
# **Chapitre II**

**Etude de l'activité anti-  
inflammatoire par criblage virtuel  
à base de QSAR**

## 1. Introduction

Les anti-inflammatoires non stéroïdiens (AINS) sont encore les médicaments couramment prescrits dans le monde pour le traitement des maladies inflammatoires telles que la polyarthrite rhumatoïde, l'arthrose, les blessures orthopédiques, les douleurs postopératoires, les myalgies aiguës, etc.<sup>1</sup>. Vane<sup>2</sup>, Samuelson et Bergstrom ont réussi à montrer que les anti-inflammatoires non stéroïdiens (AINS) exercent des effets anti-inflammatoires et analgésiques par l'inhibition de la synthèse des prostaglandines (PG), en bloquant l'activité de la cyclooxygénase (COX). De nombreux chercheurs ont travaillé sur la synthèse et l'évaluation de nouveaux composés<sup>3,4</sup>. Dans les composés hétérocycliques moléculaires, la thiazolidine est un échafaudage reconnu pour les médicaments et candidats potentiels. Les activités anticonvulsivantes, sédatives, antidépressives, anti-inflammatoires, antihypertensives, antihistaminiques et antiarthritiques ne sont que quelques-unes parmi de nombreuses autres réponses biologiques<sup>5</sup> de ces composés hétérocycliques.

Dans cette partie de travail nous présentons une étude de criblage virtuel basée sur l'application d'un modèle QSAR, construit à partir d'un ensemble de dérivées de l'acide 2-[4-(thiazole-2-yl)phényl]propionique (**Figure 1**) en tant qu'agents anti-inflammatoires.



**Figure.1** Structure générale des inhibiteurs de la cyclooxygénase (COX2) dérivés de l'acide 2-[4-(thiazole-2-yl) phényl]propionique.

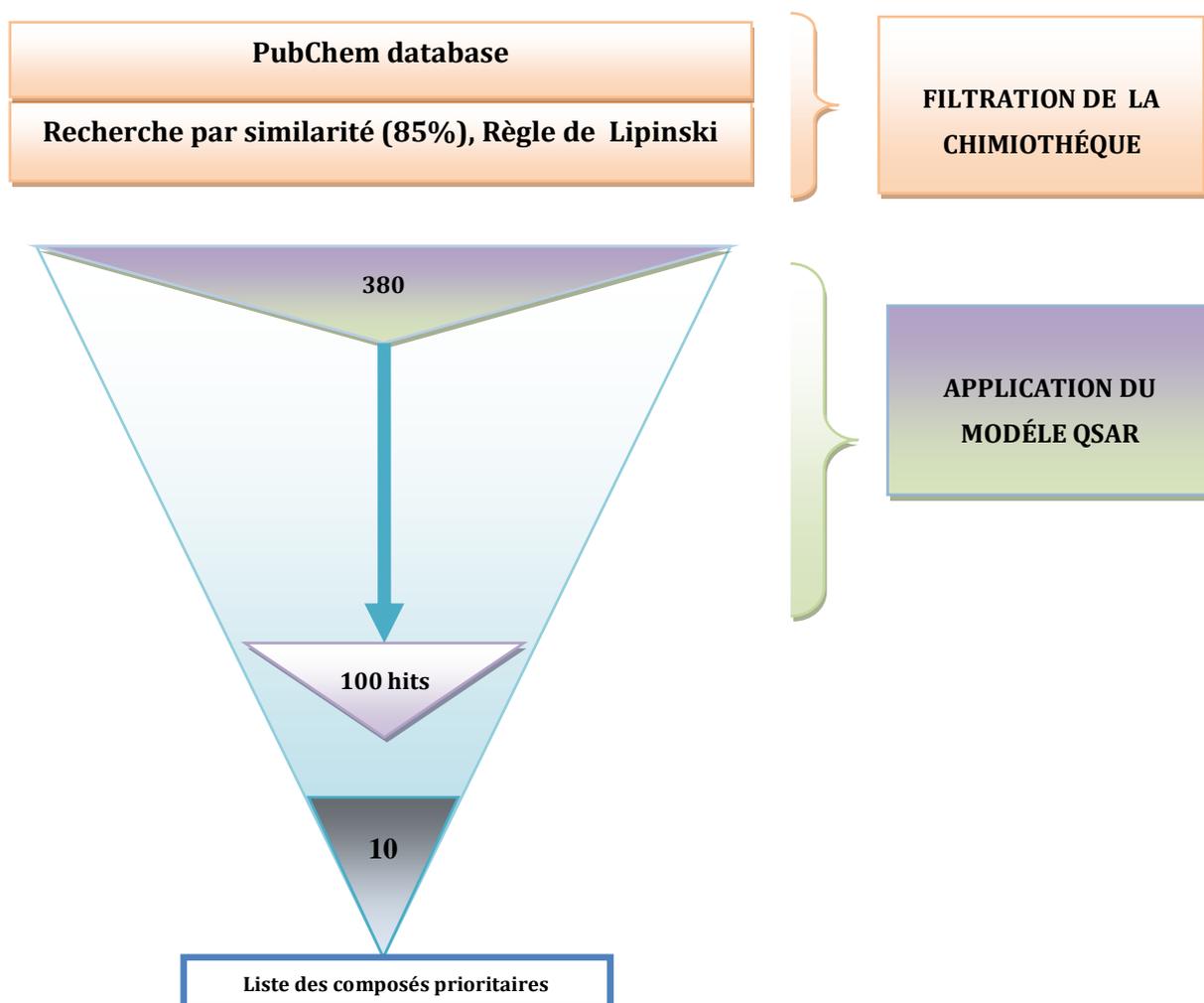
Les études de relation quantitative structure-activité (QSAR) présentent une méthode puissante pour la conception de composés bioactifs, la prédiction de l'activité en fonction des propriétés physiques et chimiques et une meilleure compréhension des caractéristiques structurales de ces types de composés, à l'aide de descripteurs moléculaires

Afin d'obtenir une bonne équation de régression linéaire pour prédire l'activité anti-inflammatoire des dérivés d'acide 2-[4-(thiazole-2-yl) phényl]

propionique, nous avons utilisé une analyse de régression linéaire multiple couplée aux algorithmes génétiques comme méthode de sélection de variables.

## 2. Protocole de criblage virtuel

Dans ce travail, le criblage virtuel que nous avons réalisé est basé sur la structure du ligand « ligand-based » en appliquant un modèle QSAR sur les composés de la base de données PubChem, qui vise à la recherche de nouvelles structures moléculaires potentiellement actives biologiquement vis-à-vis de l'enzyme cyclooxygénase et d'enrichir ainsi la classe thérapeutique des anti-inflammatoires. Le processus de criblage virtuel peut être schématisé en quatre étapes, dont les enjeux sont décrits ci-dessus (**Figure 2**).



**Figure 2** : Etapes du protocole de criblage virtuel utilisé pour l'identification des nouveaux inhibiteurs de cyclooxygénase2.

### **3. Création et préparation de la chimiothèque**

Les structures tridimensionnelles des molécules composant la chimiothèque de criblage (380 structures) sont téléchargées de la base de données PubChem avec un pourcentage de similarité de 85%. PubChem est une base de données de chimie gratuite (libre) aux National Institutes of Health (NIH)<sup>6</sup>. Elle contient principalement, depuis son lancement en 2004, de petites molécules (plus de 111 millions de composés), mais également des molécules plus grosses telles que des nucléotides, des glucides, des lipides, des peptides et des macromolécules modifiées chimiquement. Elle dispose, également, des informations sur les structures chimiques, les identifiants, les propriétés chimiques et physiques, les activités biologiques, les brevets, la santé, la sécurité, les données de toxicité et bien d'autres.

Nous avons adopté le coefficient tanimoto comme paramètre de recherche avec 85% comme seuil de similarité par rapport au squelette de base de l'ensemble d'apprentissage du modèle QSAR, représenté par l'acide 2-[4-(thiazole-2-yl) phényl] propionique. Toutes les structures obtenues sont enregistrées sous forme de fichiers "sdf". Puis préparées et filtrées avec le programme « *ligand preparation* » du logiciel STUDIO et enregistrées en format MOL2. La géométrie de toutes les structures a été optimisée par le logiciel HyperChem par la méthode semi empirique AM1. Les descripteurs moléculaires des structures sont ensuite générés à l'aide du logiciel Dragon 6.0 et enregistrés en format Texte<sup>7</sup>.

## **4. Développement du modèle QSAR**

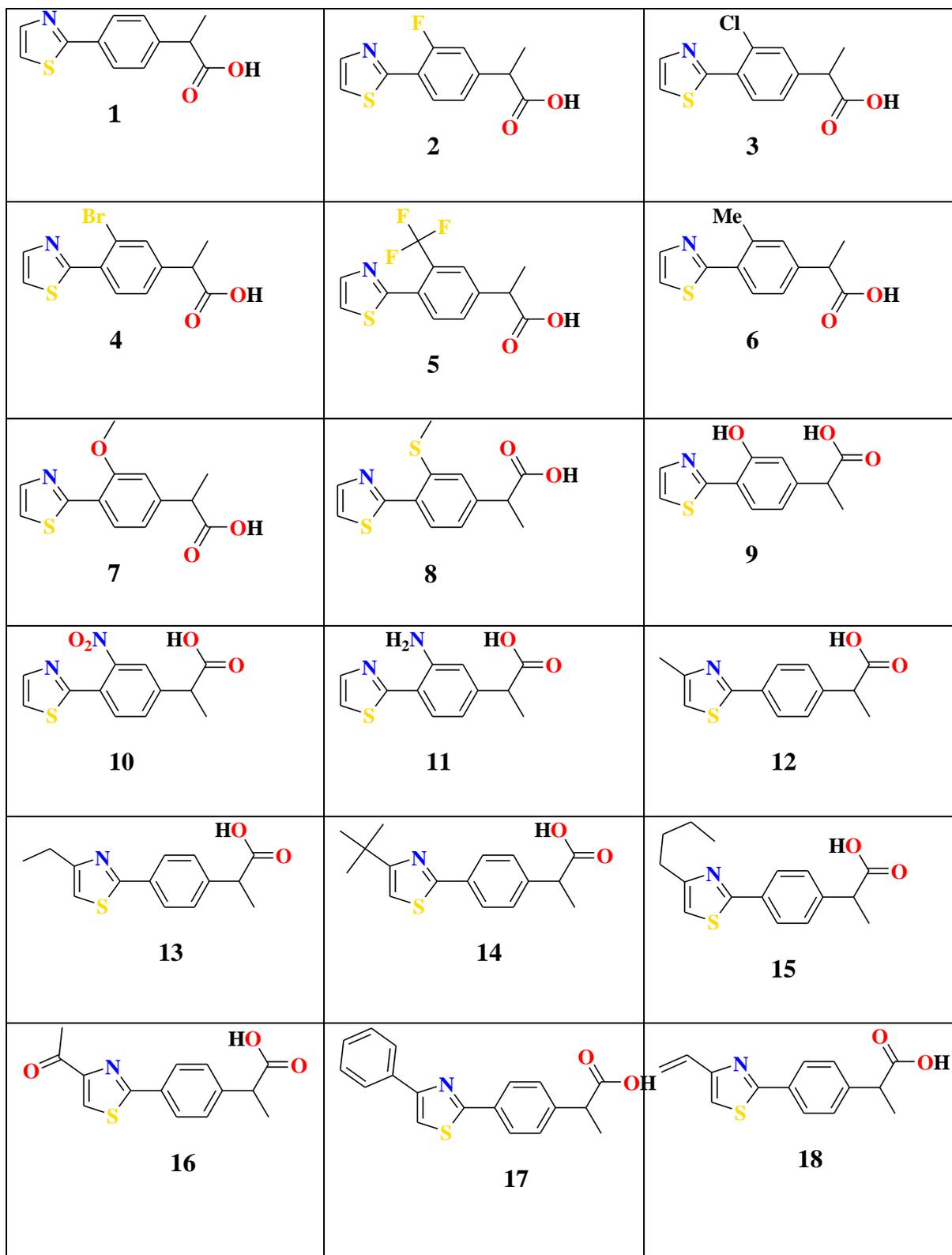
### **4.1. Méthodes expérimentales**

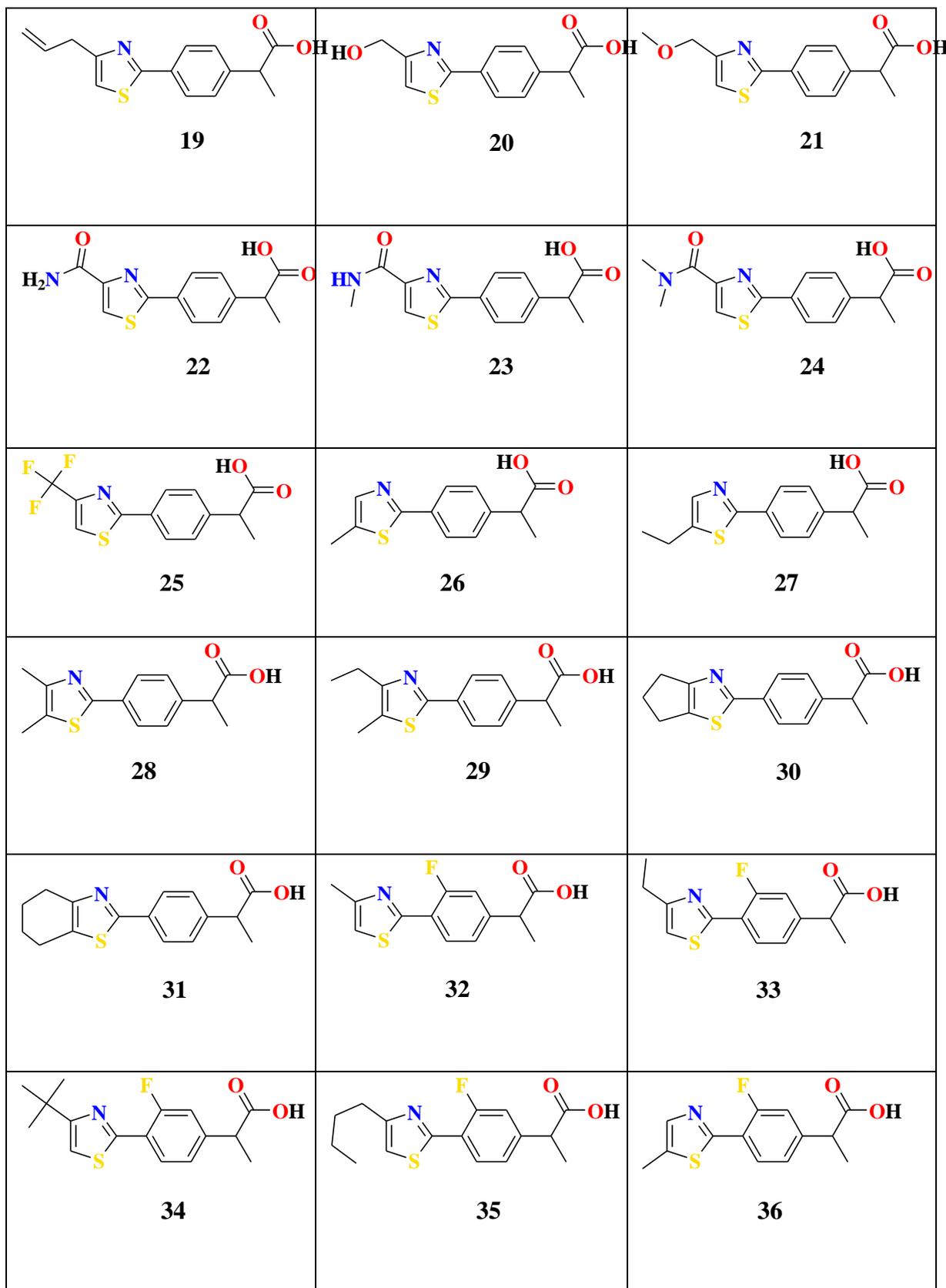
#### **4.1.1. Ensemble de molécules et activité biologique**

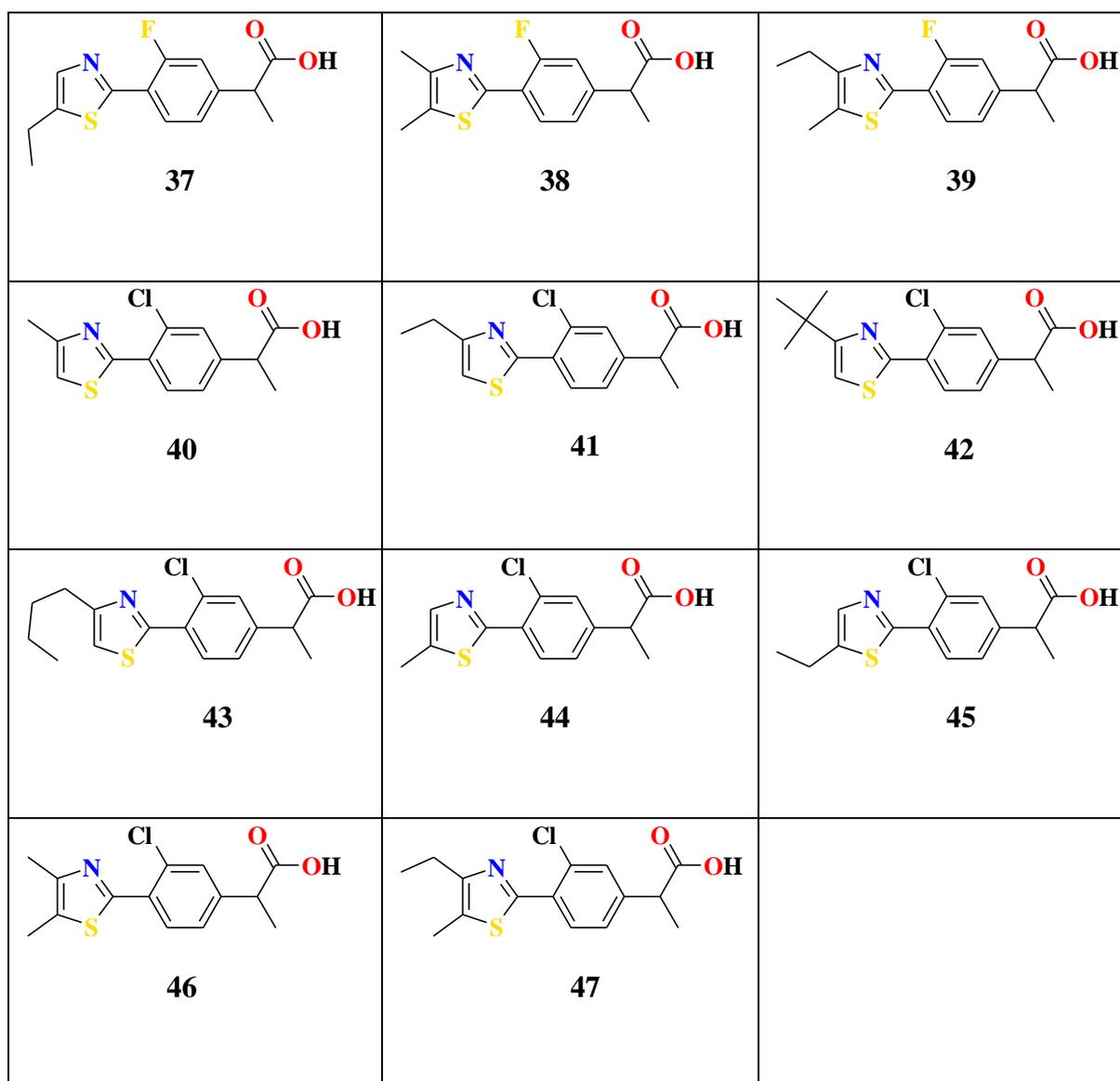
L'ensemble étudié dans ce travail est constitué de 47 molécules dérivées de l'acide 2-[4-(thiazole-2-yl) phényl] propionique (**Figure 3**) décrit par Naito et al<sup>8</sup>. Les molécules de cet ensemble contiennent dans leurs structures des atomes de C, N, O, S, H et trois halogènes (F, Cl, Br) avec un nombre total d'atomes variant de 27 à 39 et un poids moléculaire variant de 233,31 à 323,87 g/mol. Cet ensemble est divisé aléatoirement en deux sous-ensembles : un sous ensemble de 11 molécules ( $\approx 30\%$ ) choisies au hasard, en tant qu'ensemble de validation (VSET), pour évaluer le pouvoir

## Chapitre II : Etude de l'activité anti-inflammatoire par criblage virtuel à base de QSAR.

prédicatif des modèles proposés, et le reste des molécules, en nombre de 36, ( $\approx 70\%$ ) constitue l'ensemble d'apprentissage (Training Set: TSET) utilisé dans la construction du modèle de QSAR.







**Figure 3.** Structures développées des dérivés d'acide 2-[4-(thiazole-2-yl) phényle] propionique.

Selon le travail de Naito et al<sup>8</sup>, la réponse biologique est donnée par la grandeur  $I_{50}$  qui mesure la concentration molaire, en molécule inhibitrice, nécessaire pour inhiber 50% de l'activité enzymatique. Pour notre cas nous avons exprimé l'activité inhibitrice par le rapport logarithmique  $pI_{50}$  [ $\text{Log}(1/I_{50})$ ] qui exprime directement le potentiel d'inhibition. Les valeurs  $pI_{50}$  pour l'ensemble de molécules varient de  $pI_{50} = 2,991$  à  $7,222$ . Les valeurs  $pI_{50}$  expérimentales et calculées des structures moléculaires sont illustrées dans le Tableau 1.

**Tableau 1 :** Valeurs expérimentales et calculées de  $pl_{50}$  avec les descripteurs sélectionnés de l'ensemble d'apprentissage (TSET) et de l'ensemble de validation (VSET).

Molécule	$pl_{50}^{exp.}$	$pl_{50}^{pred.}$	H4e	R7e+	B03[N-O]	F03[C-S]	TE2
<b>Ensemble d'apprentissage</b>							
<b>No.</b>							
1	5.56067	5.75083	0.801	0.027	0	2	5.239
3	6.69897	6.75730	0.755	0.041	0	2	5.134
5	6.52288	6.27576	0.909	0.059	0	2	6.480
6	5.69897	5.82208	0.805	0.038	0	2	5.789
7	5.28400	5.38202	0.672	0.033	0	2	6.005
8	6.52288	6.75316	0.661	0.030	0	5	6.673
10	5.53760	5.65041	0.783	0.053	0	2	6.805
13	5.52288	5.14911	0.937	0.021	0	3	6.055
16	3.97881	3.82528	0.909	0.021	1	3	6.660
19	4.88606	4.96292	1.037	0.021	0	3	6.156
20	4.79588	5.05002	0.857	0.026	1	3	5.853
21	4.52288	4.62188	1.325	0.022	1	3	5.694
22	3.12494	2.90410	1.063	0.021	1	3	7.398
23	2.99140	3.11936	1.091	0.022	1	3	7.237
24	3.03621	2.92947	1.731	0.023	1	3	7.020
26	5.25964	5.57568	0.938	0.027	0	2	5.304
28	6.00000	5.59648	0.817	0.020	0	3	5.672
29	4.99140	5.02543	0.880	0.019	0	3	6.096
30	6.15490	6.28317	0.842	0.020	0	4	5.720
31	6.15490	5.73198	1.086	0.018	0	4	5.943
32	7.00000	6.53381	0.810	0.038	0	3	5.828
33	5.92082	5.75132	0.997	0.035	0	3	6.247
34	4.46852	4.51862	1.179	0.028	0	3	6.858
35	4.29243	4.17621	1.658	0.031	0	3	7.006
36	6.52288	6.52487	0.910	0.046	0	2	5.520
37	6.04576	6.35475	1.092	0.042	0	3	6.020
38	6.69897	6.42412	0.815	0.037	0	3	5.869
39	5.37675	5.66579	0.976	0.034	0	3	6.284
40	7.00000	6.87005	0.783	0.037	0	3	5.482
41	6.15490	6.06535	0.992	0.034	0	3	5.906
42	4.46852	4.84462	1.154	0.027	0	3	6.520
43	4.20761	4.42554	1.659	0.029	0	3	6.664
44	7.22185	6.73441	0.870	0.043	0	2	5.187
45	6.69897	6.53969	1.080	0.039	0	3	5.690
46	6.09691	6.79229	0.805	0.037	0	3	5.538
47	5.92082	5.95370	0.991	0.033	0	3	5.953
<b>Ensemble de validation</b>							
<b>No.</b>							
2	6.60206	6.585	0.821	0.045	0	2	5.471
4	7.00000	6.446	0.822	0.041	0	2	5.373
9	6.22185	5.626	0.773	0.039	0	2	6.047
11	6.00000	5.171	0.747	0.043	0	2	6.708
12	6.14267	5.834	0.812	0.023	0	3	5.626

14	4.78781	4.278	1.134	0.020	0	3	6.661
15	3.97881	3.819	1.586	0.021	0	3	6.822
17	4.44370	4.706	1.144	0.020	0	3	6.261
18	5.67778	5.557	0.840	0.022	0	3	5.804
25	5.08619	5.676	0.881	0.034	0	3	6.341
27	5.49485	5.655	1.045	0.026	0	3	5.796

#### **4.1.2. Dessin et optimisation des structures**

Le dessin et l'optimisation des structures des molécules dérivées de l'acide 2-[4-(thiazole-2-yl) phényl] propionique sont effectués à l'aide du logiciel HyperChem (version 7.5)<sup>9</sup>. Nous avons utilisé la méthode semiempirique (AM1) en tant que méthode d'optimisation en employant l'algorithme Polack-Ribiere avec un gradient énergétique égal à 0.01 kcal/mol.Å°.

Le programme de calcul est exécuté sur un PC (hp pro) Intel (R) CORE(TM) i3CPU 3,30 GHz 3,30 GHz.

#### **4.1.3. Génération des descripteurs**

Les descripteurs utilisés sont ceux calculés par le programme Dragon version 6.0, application utilisée pour le calcul de descripteurs moléculaires, qui comprend 4887 descripteurs répartis en 29 blocs<sup>10</sup>.

#### **4.1.4. Sélection des descripteurs**

La sélection de descripteurs (variables) représente la partie la plus importante lors de la construction d'un modèle de QSAR. Elle sert à réduire le nombre de descripteurs et à sélectionner les meilleurs d'entre eux qui assurent une bonne corrélation avec l'activité biologique. Cette procédure est effectuée selon deux étapes successives : la sélection objective et la sélection subjective.

##### **a. Sélection objective**

Dans cette étape, on travaille avec la matrice des descripteurs (variables indépendantes) sans l'utilisation de la variable dépendante, qui est représentée par le vecteur de l'activité biologique  $pI_{50}$ . Tous les descripteurs jugés non utiles seront éliminés. En effet :

- Tous les descripteurs présentant les mêmes valeurs pour toutes les molécules sont éliminés.
- Pour toute paire de descripteurs présentant un coefficient de corrélation supérieur ou égal à 0,9 ( $R \geq 0,9$ ), celui qui possède la petite variance sera éliminé.

### **b. Sélection subjective**

Cette procédure sert à réduire, encore une fois, le nombre de descripteurs en utilisant cette fois-ci la variable dépendante ( $pI_{50}$ ) pour définir le meilleur sous-ensemble de descripteurs riches en informations. Le choix du sous ensemble optimal de descripteurs est réalisé par les algorithmes génétiques (AG).

Les conditions de simulation par les AG appliquées dans ce contexte sont :

- Nombre générations égal à  $10^4$
- Nombre de croisement égal à  $5 \cdot 10^3$
- Facteur de flexibilité égal à 1
- La probabilité de mutation pour ajouter un nouveau terme est égale à 50%
- La taille population contient 100 modèles.

Le procédé des algorithmes génétiques est répété plusieurs fois pour confirmer que les descripteurs choisis constituent le sous-ensemble optimal pour expliquer l'activité biologique.

### **4.1.5. Construction du modèle**

Nous avons utilisé la régression linéaire multiple comme algorithme d'apprentissage supervisé pour la construction du modèle, en utilisant le sous-ensemble d'apprentissage (TSET). Les paramètres statistiques tels que : les valeurs des coefficients de détermination  $R^2$ , l'écart type « s » et le tableau d'ANOVA sont générés automatiquement en utilisant le programme de statistiques MINITAB (version 15)<sup>11</sup>.

**4.2. Résultats et discussion**

Le modèle de QSAR développé est constitué principalement de descripteurs calculés théoriquement à l'aide du logiciel Dragon sur un ensemble de molécules dérivées de d'acide 2-[4-(thiazole-2-yl) phényl] propionique et testées expérimentalement contre la cyclooxygénase2 (**Tableau 2**).

Vu le grand nombre de descripteurs contenu dans la matrice des descripteurs, il est nécessaire de réduire la taille de la matrice et sélectionner le meilleur sous-ensemble qui présente la bonne corrélation avec l'activité biologique  $PI_{50}$ . Pour ce faire, nous avons fait appel à la méthode Algorithmes Génétiques. Cette dernière nous permet de définir les meilleurs sous-ensembles avec différentes dimensions.

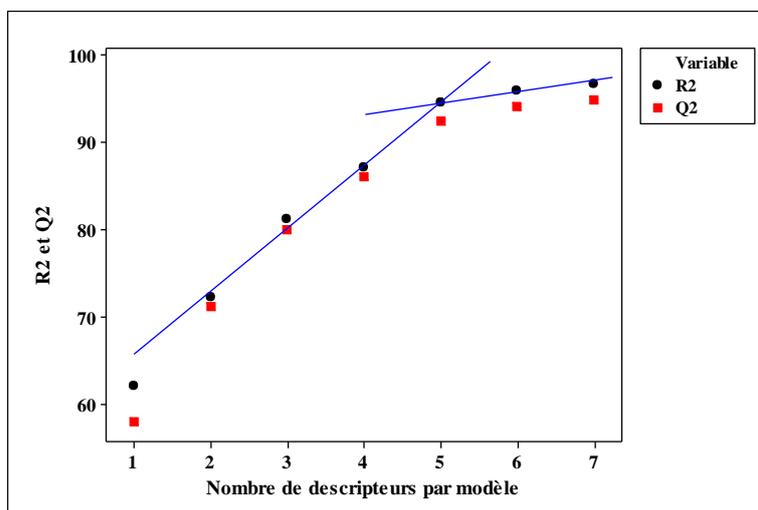
Nous avons établi consécutivement plusieurs modèles de différentes dimensions allant de 1 jusqu'à 7 descripteurs par modèle. Le tableau 2 regroupe le meilleur sous-ensemble de descripteurs pour chaque dimension, ainsi que leurs coefficients de détermination  $R^2$  et de validation  $Q^2$  qui varient de 0,622 à 0,967 et de 0,5810 à 0,9493 respectivement.

**Tableau 2** : Série des modèles optimums à différentes dimensions sélectionnées par la méthode des algorithmes génétiques.

<b>Dimension du modèle</b>	<b>Descripteurs constituant le modèle</b>	<b><math>R^2</math></b>	<b><math>Q^2_{cv-100}</math></b>
7	Me; MATS7e; Mor08e; H5e; B03[N-O]; F03[C-S]; TE2	96.67	94.93
6	H5e R5s nArX B03[N-O] F03[C-S] TE2	95.99	94.15
<b>5</b>	<b>H4e; R7e+; B03[N-O]; F03[C-S]; TE2</b>	<b>94.59</b>	<b>92.45</b>
4	R7e+; F03[C-S]; TE2; LLS_01	87.11	86.08
3	PW5 ; X3Av ; RDF010m	81.21	80.03
2	nCbH ; LLS_01	72.31	71.16
1	H3m	62.15	58.07

Le choix de la meilleure dimension de notre modèle est défini par la technique du point de rupture<sup>12</sup>, qui consiste à représenter graphiquement les paramètres statistiques ( $R^2$  et  $Q^2_{cv-100}$ ) en fonction du nombre de descripteurs.

La figure 4 montre la projection de  $R^2$  et  $Q^2_{cv-100}$  en fonction du nombre de descripteurs en utilisant les algorithmes génétiques. La dimension qui correspond au point de virage est égale à cinq ( $n= 5$ ) comme indiqué par la (Figure 4). D'après la technique du point de virage, le nombre optimal de descripteurs est fixé à cinq, avec  $R^2 = 0,9459$  et  $Q^2_{cv-100} = 0,9245$ .



**Figure 4 :** Représentation des coefficients de détermination, obtenus par GA, en fonction du nombre de descripteurs.

### 4.2.1. Analyse de la justesse du modèle optimal

Après détermination de la dimension du sous-ensemble optimal sélectionné par les algorithmes génétiques, nous avons construit le modèle en utilisant la régression linéaire multiple (RLM) à l'aide du logiciel MINITAB.15. Le modèle QSAR reliant l'activité biologique exprimée par la grandeur  $pI_{50}$  aux cinq descripteurs (ISH, R6e+, GATS2m, RDF105m et R8m) accompagné par ses paramètres statistiques, coefficient de détermination  $R^2$ , coefficient de corrélation multiple  $R$ , et l'écart type  $s$ , est décrit ci-dessous par l'équation 1.

$$pI_{50} = 8,90 - 0,762 H4e - 0,686 B03[N-O] + 0,758 F03[C-S] - 1,09 TE2 + 61,2 R7e+ \quad (\text{eq.1})$$

$$N = 36 ; R = 97,25\% ; R^2 = 94,59\% ; R^2_{cv-1-out} = 92,45\% ; s = 0,286$$

La valeur élevée du coefficient de corrélation multiple  $R = 0,973$  nous indique qu'il existe une forte corrélation entre les valeurs expérimentales de l'activité biologique et celles calculées par le modèle de la régression linéaire multiple.

Le coefficient de détermination  $R^2$  montre, que 94,59 % de la variance des valeurs expérimentales est expliquée par les cinq descripteurs (H4e ; R7e+ ; B03[N-O] ; F03[C-S] ; TE2).

La valeur, faible, de l'écart type  $s=0,286$  nous renseigne sur la bonne qualité de la dispersion des données.

Les valeurs élevées des coefficients mentionnés ci-dessus résument la corrélation des cinq descripteurs avec l'activité biologique. Cette corrélation est vérifiée en examinant le tableau d'ANOVA et le tableau des coefficients.

### A. Analyse de la variance (ANOVA)

L'analyse de la variance (ANOVA) est réalisée à partir du tableau d'ANOVA (**Tableau 3**) où la présence ou l'absence de corrélation entre l'ensemble de descripteurs et l'activité biologique est vérifiée en examinant la statistique de Fisher  $F_{obs}$ . Pour réaliser cette analyse nous avons fait appel aux deux hypothèses : hypothèse nulle et hypothèse alternative.

**Tableau3** : Résultats de l'analyse de la variance (ANOVA)

Source	DF <sup>a</sup>	SS <sup>b</sup>	MS <sup>c</sup>	F <sub>obs</sub>	P <sup>d</sup>
Régression	5	42,826	8,565	104,94	0,000
Erreur Résiduelle	30	2,449	0,0816		
Total	35	45,275			

a: degré de liberté ; b : somme des carrés ; c : écart moyen ; d : probabilité.

- *L'hypothèse nulle ( $H_0$ )* : « aucun descripteur n'est lié à l'activité biologique :  $\beta_j = 0$  avec ( $j=0 ; 1 ; 2, 3, 4$ ),  $\beta_j$  : coefficient correspondant au descripteur  $j$  »<sup>13</sup>.

Cette hypothèse est acceptée si la valeur de la statistique de Fischer observée est inférieure à la valeur  $F_{(0,05 ; 5 ; 30)}$ .

- *L'hypothèse alternative ( $H_1$ )* : « il existe au moins un descripteur corrélé avec l'activité biologique »<sup>13</sup>.

Cette hypothèse est acceptée lorsque la valeur de la statistique de Fischer observée est supérieure à la valeur  $F_{(0,05; 5; 30)}$ .

D'après le tableau d'ANOVA, la statistique de Fischer observée ( $F_{obs}=104,94$ ) est supérieure à  $F_{(0,05; 5; 30)}=2,53$  ce qui nous permet d'accepter l'hypothèse alternative et de confirmer qu'il existe au moins un coefficient différent de zéro c'est-à-dire un descripteur corrélé avec l'activité inhibitrice exprimée par  $pl_{50}$ .

### B. L'examen du tableau des coefficients

Nous avons ensuite examiné le tableau des coefficients pour vérifier la signification de chaque descripteur et sa contribution dans l'explication de l'activité biologique. L'utilisation des valeurs de la statistique  $t$  de Student, affichées dans le **(Tableau 4)** pour chaque descripteur, nous permet de vérifier la présence ou l'absence de corrélation entre chaque descripteur et l'activité biologique en se basant sur les deux hypothèses : hypothèse nulle et hypothèse alternative.

**Tableau 4 :** Tableau des coefficients

Var. Ind. <sup>a</sup>	Coef <sup>b</sup>	E. T (Coef) <sup>c</sup>	$t_{obs}^d$	$p^e$
intercepte	8,8987	0,6035	14,74	0,000
H4e	-0,7617	0,2234	-3,41	0,002
B03[N-O]	-0,6861	0,1567	-4,38	0,000
F03[C-S]	0,7580	0,09413	8,05	0,000
TE2	-1,0893	0,1059	-10,29	0,000
R7e+	61,218	6,431	9,52	0,000

a: variables indépendantes ; b : coefficients ; c : erreur type ; d : test de student observé ; e : la valeur de probabilité.

- *L'hypothèse nulle ( $H_0$ )* : « le descripteur n'est pas lié à l'activité biologique :  $\beta_j = 0$  avec ( $j=0 ; 1 ; 2 ; 3 ; 4$ ),  $\beta_j$  : coefficient correspondant au descripteur  $j$  »

Cette hypothèse est acceptée si la valeur de la statistique de Student  $t$  observée est inférieure à la valeur  $t_{(0,025; 30)}$  à un niveau de confiance  $\alpha = 0,05$ .

- *L'hypothèse alternative ( $H_1$ )* : « le descripteur est corrélé avec l'activité biologique :  $\beta_j \neq 0$  ».

Cette hypothèse est acceptée lorsque la valeur de la statistique de Student  $t$  observée est supérieure à la valeur  $t_{(0,05,30)} = 1,697$  en valeur absolue

D'après le tableau 4, les valeurs de la statistique observées  $t_{obs}$  sont plus élevées par rapport à celle de la distribution  $t_{(0,05,30)}$ . Cela nous permet de rejeter l'hypothèse nulle, c'est-à-dire que les coefficients inclus dans le modèle diffèrent considérablement de zéro. Ce jugement est consolidé par les faibles valeurs de probabilité ( $p \leq 10^{-3}$ ) pour les six paramètres (constante + 5 coefficients) de l'équation (eq.1).

L'intervalle de confiance pour chaque coefficient est calculé avec l'expression suivante :

$$\beta_i = \hat{\beta}_j \pm t_{(n-p-1, \alpha/2)} \cdot E.T.(\hat{\beta}_j)$$

Avec :

E.T. : l'erreur type.

$n-p-1 = 36-5-1=30$  : degrés de liberté du modèle.

$\alpha$  : le niveau de confiance.

Les intervalles de confiance pour les quatre paramètres (avec  $\alpha = 0,05$ ) sont établis comme suit :

Constante :  $7,070095 < \text{constante} < 10,727305$

RDF105m :  $-1,438602 < H4e < -0,084798$

ISH :  $-1,160901 < B03[N-O] < -0,211299$

R6e+ :  $0,4728061 < F03[C-S] < 1,0432339$

R8m :  $-1,410177 < TE2 < -0,768423$

GATS2m :  $41.73207 < R7e+ < 80.70393$

Nous constatons que tous les coefficients ne renferment plus la valeur zéro dans leurs intervalles de confiance.

#### **4.2.2. Tests de colinéarité et Multicolinéarité**

Afin d'éviter la redondance de l'information fournie par les descripteurs figurant dans le modèle construit, il faut vérifier deux paramètres : la colinéarité et la multicolinéarité. Elles sont expliquées respectivement par la présence d'une forte corrélation d'un descripteur avec un autre ou avec un ensemble de descripteurs. Les descripteurs corrélés entre eux contiennent une grande partie de la même information provoquant ainsi des problèmes calculatoires lors de la génération de la matrice inverse  $(X'X)^{-1}$ .

Statistiquement, nous avons testé la colinéarité par l'examen de la matrice de corrélation en calculant les coefficients de corrélation pour toutes les combinaisons paires des cinq descripteurs (**Tableau 5**).

**Tableau5** : Matrice de corrélation de l'équation 1

	<b>H4e</b>	<b>B03[N-O]</b>	<b>F03[C-S]</b>	<b>TE2</b>	<b>R7e+</b>
<b>H4e</b>	1				
<b>B03[N-O]</b>	0,304	1			
<b>F03[C-S]</b>	0,133	0,099	1		
<b>TE2</b>	0,497	0,418	0,304	1	
<b>R7e+</b>	-0,262	-0,410	-0,505	-0,191	1

Selon le tableau 5, le modèle QSAR construit ne présente aucun problème de colinéarité entre les descripteurs. En effet, toutes les valeurs calculées des coefficients de corrélation, limitées entre 0,099 et 0,505 en valeurs absolues, sont inférieures à la valeur 0,9. Il est à noter que le problème de colinéarité apparaît lorsque les valeurs des coefficients de corrélation sont supérieures ou égales à 0,9.

La multicolinéarité est testée par le facteur d'inflation de la variance (FIV)<sup>14</sup> exprimée par l'équation 2 :

$$FIV(x_k) = \frac{1}{1 - R_k^2} \quad \text{(eq.2)}$$

Avec :

$FIV(x_k)$  : représente la valeur du facteur d'inflation de la variance pour le descripteur  $x_k$  ( $k=1 ; 2 ; 3 ; 4 ; 5$ ).

$R_k^2$  : représente le coefficient de corrélation carrée résultant de la régression du descripteur  $x_k$  sur le reste des descripteurs.

Une valeur de  $FIV \geq 10$  signifie la présence d'une forte colinéarité entre les descripteurs, l'absence de la multicollinéarité est signifiée par une valeur de  $FIV < 10$ .

On peut aussi confirmer la présence ou l'absence de la forte multicollinéarité par une autre grandeur, similaire au  $FIV$ , appelée facteur de tolérance<sup>12</sup>  $FT$  donnée par l'équation (eq.3) :

$$TF(x_k) = \frac{1}{VIF(x_k)} = 1 - R_{x_k}^2 \quad (\text{eq.3})$$

Les valeurs de  $FT$  varient dans l'intervalle  $0 < FT < 1$ . Les valeurs de  $FT > 0,5$  et  $FT < 0,5$  correspondent à l'absence et à la présence de fortes multicollinéarité entre les descripteurs respectivement.

Les valeurs correspondantes aux  $FIV(x_k)$  et aux  $FT$  calculés pour les cinq descripteurs constituant le modèle QSAR sont réunies dans le Tableau 6.

**Tableau 6** : Valeurs des facteurs d'inflation de la variance et des facteurs de tolérance pour les descripteurs significatifs.

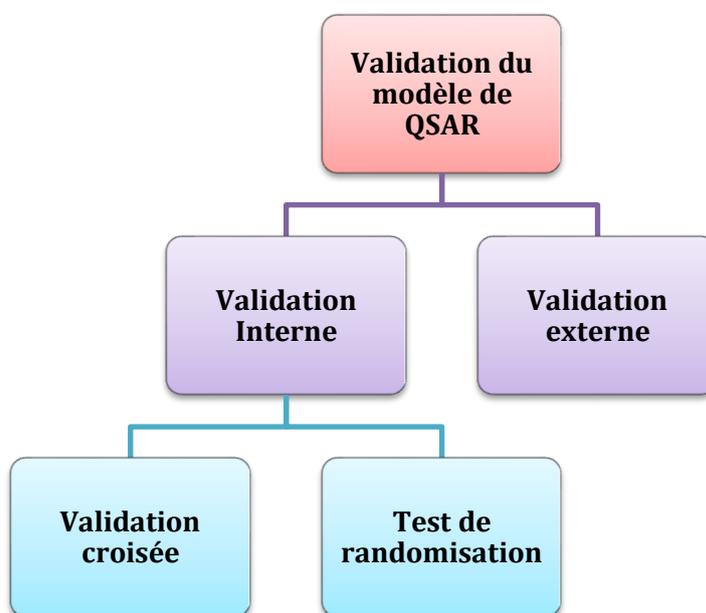
Variable	FIV	FT
H4e	1,4	0,71
B03[N-O]	1,5	0,67
F03[C-S]	1,6	0,63
TE2	1,7	0,59
R7e+	1,7	0,59

D'après le tableau 6, les cinq descripteurs ne présentent aucun problème de multicollinéarité avec une valeur maximale de  $FIV_{Glv} = 1,7 < 10$  et les valeurs de FT toutes supérieures à 0,5.

### 4.2.3. Validation interne et externe du modèle de QSAR obtenu

#### 4.2.3.1. Validation interne

La validation du modèle QSAR représente une étape indispensable pour l'estimation de la fiabilité du modèle et pour la détermination de sa pertinence et sa robustesse. Pour ce faire nous avons employé plusieurs méthodes de validation, telles que la validation interne, la validation externe et le test de randomisation (**Figure 5**).



**Figure 5:** Les différentes méthodes utilisées pour la validation de notre modèle QSAR.

La validation croisée est effectuée par la procédure leave-one-out : on retire successivement chaque molécule de l'ensemble d'apprentissage. Cette procédure est répétée  $n$  fois ( $n$  est le nombre des molécules qui constituent l'ensemble d'apprentissage) afin de prédire les propriétés de toutes les molécules.

La valeur élevée du coefficient de détermination ( $Q^2_{cv-100} = 0,9245$ ) et la petite valeur de l'écart type ( $s_{cv-100} = 0,286$ , issues de la validation croisée du modèle obtenu, prouvent la puissance prédictive de cette approche et la stabilité du modèle.

Pour confirmer la robustesse de notre modèle, ce dernier est reconstruit dix fois et à chaque fois, le classement des composantes du vecteur  $pI_{50}$  est randomisé en fixant la matrice X constituée des cinq descripteurs.

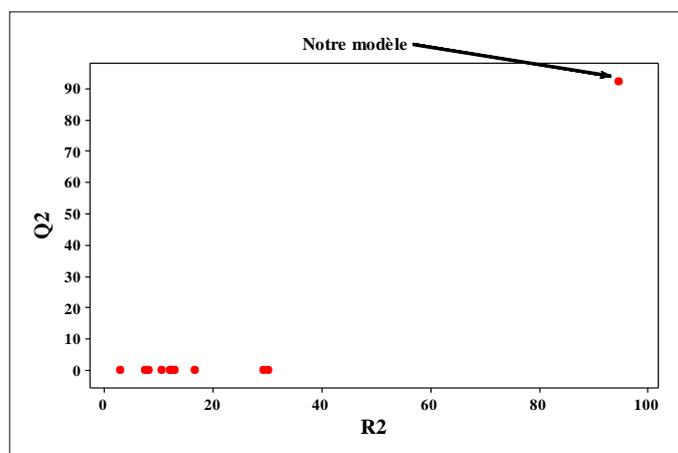
Les paramètres statistiques issus de chaque reconstruction du modèle et ceux de l'eq.1 sont cités dans le tableau 7.

**Tableau 7** :  $R^2$  et  $Q^2_{cv-100}$  issus du test de randomisation

Iteration	$R^2$	$Q^2_{cv-100}$
1	16,5	0,0
2	12,2	0,0
3	29,2	0,0
4	10,5	0,0
5	30,1	0,0
6	02,9	0,0
7	12,0	0,0
8	08,1	0,0
9	07,4	0,0
10	12,9	0,0
11*	94,59	92,45

\*: les paramètres statistiques issus de l'eq.1

La représentation graphique des valeurs du coefficient de détermination de la validation croisée  $Q^2_{cv-100}$  en fonction du coefficient de détermination (**Figure 6**) confirme que le modèle présenté par l'eq.1 est plus robuste et n'est pas dû au hasard.



**Figure 6** : Valeurs de  $Q^2_{cv-100}$  en fonction de  $R^2$  issues du test de randomisation.

#### 4.2.3.2. Validation externe

La validation externe est utilisée pour tester de manière fiable le pouvoir prédictif de notre modèle QSAR. Pour cela, nous avons utilisé l'équation du modèle (eq.1) pour prédire l'activité inhibitrice de la cyclooxygénase d'un ensemble de validation externe.

Les valeurs prédites de l'activité de molécules de l'ensemble externe sont affichées dans la troisième colonne du tableau1 (page 109). La valeur du coefficient de détermination issue de la validation externe est égale à  $Q_{ext}^2 = 0,7523$ .

Pour confirmer le pouvoir prédictif de notre modèle de QSAR nous avons vérifié certaines conditions exigées par Tropsha<sup>15</sup> (**Tableau 8**).

**Tableau 8** : comparaison des coefficients du modèle QSAR avec ceux exigés par Tropsha.

	$Q_{cv-loo}^2$	$R^2$	$R_0^2$	$k$
Selon Tropsha	<b>&gt; 0,50</b>	<b>&gt;0,6</b>	$\frac{R^2 - R_0^2}{R^2} < 0.1$	$0,85 \leq k \leq 1,15$
Notre modèle	<b>0,9245</b>	<b>0,9459</b>	<b>0,0316</b>	<b>1,01</b>

Avec :

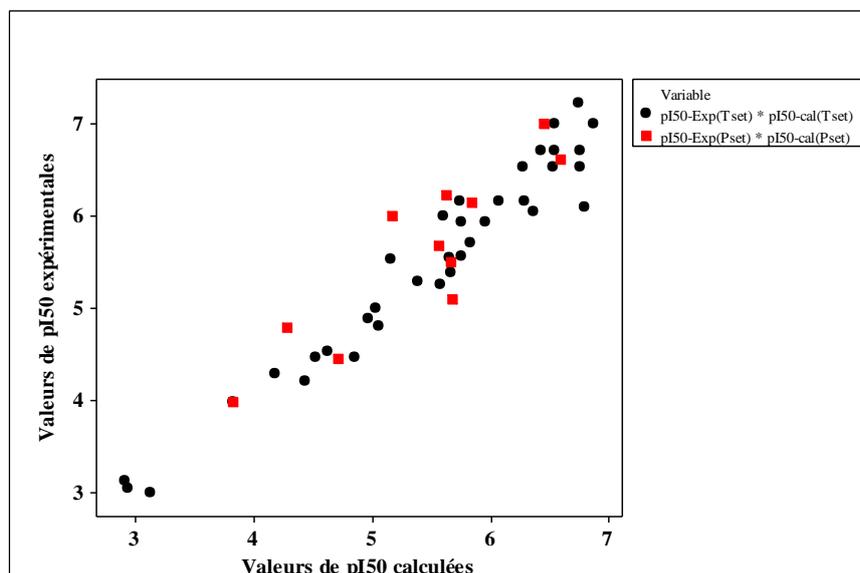
$R^2$  : représente le coefficient de détermination entre les valeurs expérimentales et celles calculées par le modèle QSAR (seulement pour l'ensemble d'apprentissage : 36 molécules).

$R_0^2$ : représente le coefficient de détermination issu de la régression linéaire des valeurs expérimentales de l'activité en fonction de celles calculées pour tout l'ensemble (47 molécules) pour notre cas  $R_0^2 = 0.916$ .

k: représentent la pente de la ligne de régression qui passent par l'origine.

La **figure 7** représente la représentation linéaire des valeurs calculées en fonction des valeurs expérimentales pour les deux sous-ensembles d'apprentissage et de validation (TSET et VSET). Les résultats obtenus montrent que la technique

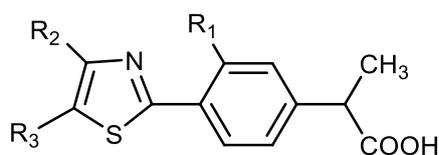
linéaire de MLR combinée avec les algorithmes génétiques en tant que méthode de sélection de variables est adéquate pour produire un modèle QSAR efficace, capable de modéliser et de prédire l'activité inhibitrice de la cyclooxygénase.



**Figure 7 :** Valeurs observées en fonction des valeurs prédites des activités pour les ensembles d'apprentissage et de prédiction.

### 4.3. Interprétation physico-chimique des descripteurs

De point de vue structural, les composés de l'ensemble étudié sont tous caractérisés par le même squelette de base de l'acide 2-[4-(thiazole-2-yl) phényl] propionique (**Figure 1**), diversement substitués aux positions R<sub>1</sub>, R<sub>2</sub> et R<sub>3</sub>.



**Figure.1** Structure générale des inhibiteurs de la cyclooxygénase (COX) dérivés de l'acide 2-[4-(thiazole-2-yl) phényl]propionique.

D'après l'équation du modèle QSAR obtenu (eq.1), nous constatons que l'activité inhibitrice de l'enzyme cyclooxygénase est influencée par les cinq descripteurs H4e, R7e+, TE2, B03[N-O] et F03[C-S] qui sont définis dans le **tableau 9**. Les descripteurs H4e ; B03[N-O] et TE2 possèdent des effets négatifs sur l'activité inhibitrice tandis que le F03[C-S] et le R7e+ sont positivement corrélés avec l'activité.

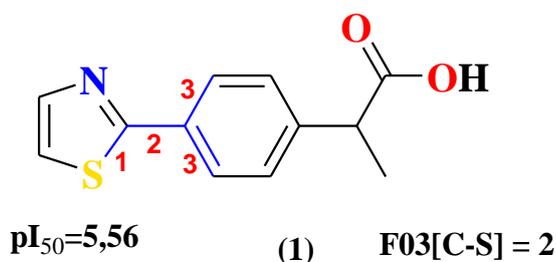
**Tableau 9** : les cinq descripteurs moléculaire qui constituent le modèle QSAR

Nom du descripteur	Désignation	Famille
<b>H4e</b>	H autocorrelation of lag4/pandéré par l'électronégativité de Sanderson	GETAWAY descriptors
<b>B03[N-O]</b>	Presence/Absence of N-O at topological distance 3.	2D Atom Pairs
<b>F03[C-S]</b>	Frequency of C-S at topological distance 3.	2D Atom Pairs
<b>TE2</b>	Topographic electronic descriptor (bond restricted)	Charge descriptors
<b>R7e+</b>	R maximal autocorrelation of lag 7/ pandéré par l'électronégativité de Sanderson	GETAWAY descriptors

#### 4.3.1. Influence des descripteurs B03[N-O] et F03[C-S]

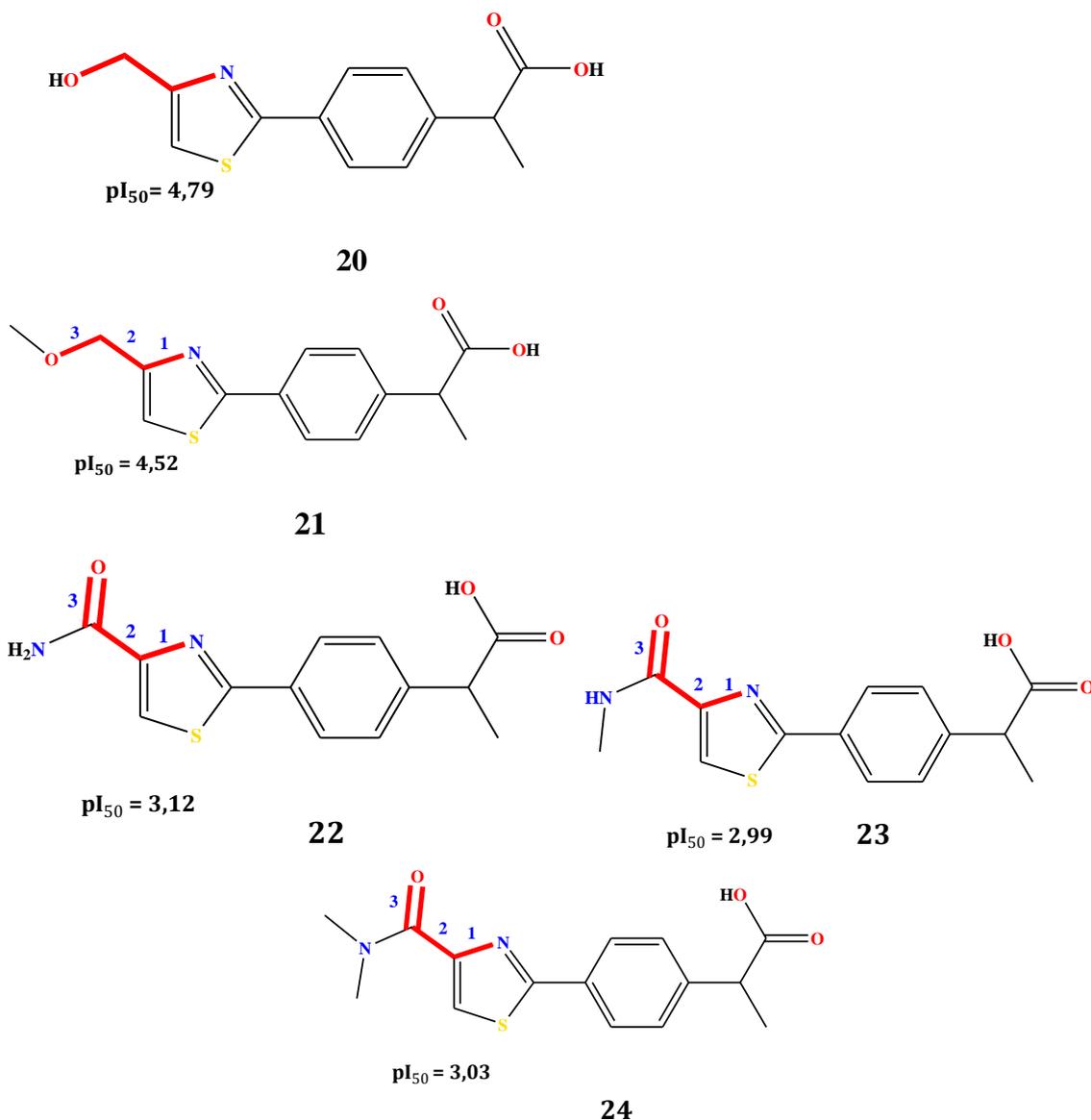
Les descripteurs B03[N-O] et F03[C-S] appartiennent à la famille « 2D Atom Pairs ». Le descripteur B03[N-O] désigne la présence ou l'absence d'une connexion entre un atome d'azote et un oxygène à une distance topologique égale à 3 (une distance topologique est représentée par une arête ou liaison). Ce descripteur prend seulement deux valeurs, la valeur 1 dans le cas de présence d'une paire N-O connectée à cette distance sinon il prend la valeur 0.

Le descripteur F03[C-S] représente le nombre de paires d'atomes C-S connectés à 3 distances topologiques. Par exemple  $F03[C-S] = 2$  si la molécule possède deux paires d'atomes C-S connectés à trois distances topologiques<sup>16</sup> (cas de la molécule 1, **Figure 8** ).



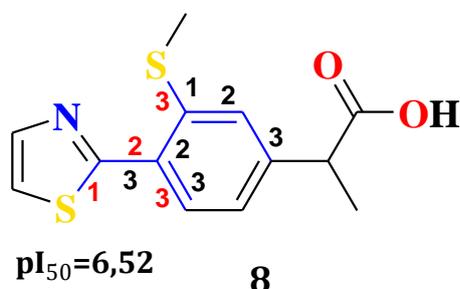
**Figure 8** : Calcul du descripteur F03[C-S] pour la molécule 1 de l'ensemble étudié.

Le descripteur **B03[N-O]** influe négativement l'activité, c'est-à-dire la présence de deux atomes de N et O connectés à une distance topologique égale à 3 diminue le pouvoir inhibiteur des molécules. Ce qui est exprimé par la diminution du pouvoir inhibiteur des molécules 20, 21, 22, 23 et 24 substituées par les groupements :  $-\text{CH}_2\text{OH}$  ;  $-\text{CH}_2\text{Ome}$  ;  $-\text{CONH}_2$  ;  $-\text{CONHMe}$  ;  $-\text{CONMe}_2$  respectivement en position  $\text{R}_2$ . La présence de ces groupements a diminué l'activité de  $\text{pI}_{50} = 5,56$  (molécule 1) jusqu'à  $\text{pI}_{50} = 2,99$  (molécule 23). Afin d'améliorer l'activité inhibitrice on doit donc éviter la présence d'un atome d'oxygène à 3 distances topologiques avec l'azote du cycle thiazol (**Figure 9**).



**Figure 9 :** Structure des molécules 20, 21, 22, 23 et 24 montrant l'influence du descripteur B03[N-O].

Le descripteur **F03[C-S]** compte le nombre de connexion C-S à 3 distances topologiques. Ce descripteur est positivement corrélé avec l'activité (selon le modèle eq1), plus le nombre de connexion C-S augmente plus le pouvoir inhibiteur est important. La molécule 8 qui contient cinq connexions C-S à trois distances topologique possède un pouvoir inhibiteur  $pI_{50} = 6,52$  plus important que celui de la molécule 1 qui ne possède que deux connexions C-S avec un  $pI_{50} = 5,56$  (**Figure 10**).



**Figure 10** : Structure de la molécule 8 montrant l'influence du descripteur F03[C-S].

#### 4.3.2. Influence de TE2

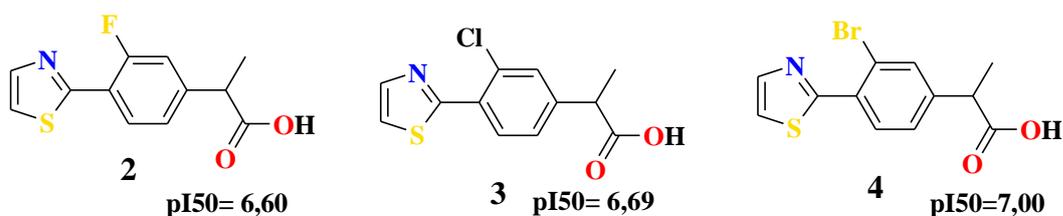
Le descripteur TE2, appartient à la famille des descripteurs électroniques de charge. Ces derniers sont définis en termes de charges atomiques et utilisés pour décrire les aspects électroniques à la fois de la molécule entière et de régions particulières, telles que les atomes, les liaisons, les fragments moléculaires.

Le descripteur électronique topographique (TE2) est calculé à partir des charges atomiques partielles  $q$  comme suit :

$$TE2 = \sum_{i=1}^{nAT-1} \sum_{j=i+1}^{nAT} \left| \frac{q_i - q_j}{r_{ij}} \right| \quad (\text{eq. 4})$$

$r_{ij}$  : la distance entre les deux atomes

où la somme s'étend sur seulement les paires d'atomes connectés (les atomes déconnectés ne sont pas considérés) <sup>17,18</sup>. Le TE2 est inversement corrélé avec l'activité inhibitrice de la cyclooxygénase (selon le modèle QSAR), ceci signifie que la présence d'atomes très polaires provoque la génération de charges partielles importantes, ce qui peut être expliqué par les structures des molécules 2, 3, 4 substituées par les atomes F, Cl et Br respectivement dans la position R<sub>1</sub>. L'activité biologique diminue progressivement avec l'augmentation de la polarité des atomes d'halogène (**Figure 11**).



**Figure 11** : Structure des molécules 2, 3 et 4 de l'ensemble étudié.

#### 4.3.3. Influence des descripteurs H4e et R7e+

Les deux descripteurs H4e et R7e+ appartiennent à la famille des descripteurs GETAWAY (GEometry, Topology, and Atom-Weights Assembly)<sup>19,20</sup>. Ces descripteurs sont calculés à partir des formules générales (eq.5) et (eq.6) respectivement :

$$Hkw = \sum_{i=1}^{nAT-1} \sum_{j>i} h_{ij} \cdot w_i \cdot w_j \cdot \delta(k, d_{ij}; h_{ij}) \quad k = 0, 1, \dots, 8 \quad (\text{eq. 5})$$

$$Rkw+ = \max_{ij} \left( \frac{\sqrt{h_{ii} \cdot h_{jj}}}{r_{ij}} \cdot w_i \cdot w_j \cdot \delta(k, d_{ij}) \right) \quad i \neq j \text{ et } k = 1, 2, \dots, 8 \quad (\text{eq. 6})$$

Où : nAT est le nombre d'atomes de la molécule ;  $d_{ij}$  est la distance topologique entre les atomes  $i$  et  $j$  ;  $w_i$  est une propriété physico-chimique atomique (dans le cas de R7e+ et He4 :  $w$  représente l'électronégativité de sanderson et  $k=7$ ) ;  $\delta(k, d_{ij})$  est une fonction ( $\delta = 1$  si  $d_{ij} = k$ , sinon 0) ;  $\delta(k; d_{ij}; h_{ij})$  est une autre fonction ( $\delta = 1$  si  $d_{ij} = k$  et  $h_{ij} > 0$ , sinon  $\delta = 0$ ) ;  $r_{ij}$  est leur distance interatomique.

Ces descripteurs sont issus de la matrice d'influence moléculaire (Molecular Influence Matrix), notés H et définis comme suit :

$$H = M \cdot (M^T \cdot M)^{-1} \cdot M^T$$

Où M est la matrice moléculaire constituée des coordonnées cartésiennes centrées x, y, z des atomes de la molécule (hydrogènes inclus) dans une conformation choisie, et l'exposant T fait référence à la matrice transposée. Les coordonnées atomiques sont supposées calculées par rapport au centre géométrique de la molécule.

Les éléments diagonaux  $h_{ii}$  de la matrice d'influence moléculaire, appelés leviers, vont de 0 à 1 et codent des informations atomiques liées à "l'influence" de chaque atome de la molécule dans la détermination de la forme globale de la molécule ; en effet, les atomes loins ont toujours des valeurs  $h_{ii}$  plus grandes que les atomes proches du centre de la molécule. Comme dérivées de la géométrie de la

molécule, les valeurs de levier sont effectivement sensibles aux changements conformationnels significatifs et aux longueurs de liaison qui tiennent compte des types d'atomes et de la multiplicité des liaisons.

Chaque élément  $h_{ij}$  hors diagonale représente le degré d'accessibilité du  $j$ ème atome aux interactions avec l' $i$ ème atome, ou, en d'autres termes, l'attitude des deux atomes considérés à interagir entre eux. Un signe négatif pour les éléments hors diagonale signifie que les deux atomes occupent des régions moléculaires opposées par rapport au centre, d'où le degré de leur accessibilité mutuelle faible.

L'interprétation des effets de ces deux descripteurs exige une analyse approfondie des valeurs de levier issues de la matrice d'influence H et des valeurs de l'électronégativité de Sanderson.

L'effet négatif du descripteur H4e sur l'activité (eq.1) peut être expliqué par la présence d'atomes polaires dans certaines molécules de l'ensemble étudié à une distance topologique égale à 4 (cas des molécules 20, 21, 22, 23, 24 et 25) où la présence d'atomes d'oxygène en position  $R_2$  diminue l'activité. L'effet de ce descripteur est corrélé positivement avec celui du B03[N-O].

## **5. Calcul des activités des molécules de la chimiothèque : Application du modèle QSAR**

Une fois les descripteurs des structures composant la chimiothèque générés on applique le modèle QSAR développé précédemment, représenté par l'équation 1 :

$$pI_{50} = 8.90 - 0.762 H4e - 0.686 B03[N-O] + 0.758 F03[C-S] - 1.09 TE2 + 61.2 R7e+ \quad (\text{eq.1})$$

Les 100 meilleures molécules classées, par ordre décroissant en termes d'activité inhibitrice de l'enzyme cyclooxygénase, sont illustrées dans le **tableau 10**.

**Tableau 10 :** Les 100 meilleurs composés classés par ordre décroissant des valeurs de  $pI_{50}$ .

N°	Molécule	Activité prédite $pI_{50}$	N°	Molécule	Activité prédite $pI_{50}$	N°	Molécule	Activité prédite $pI_{50}$
01	19698732	10.055	35	62651716	7.329	69	80315742	6.664
02	61220635	9.478	36	15163888	7.297	70	28343225	6.660

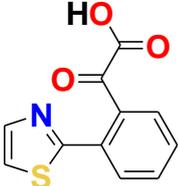
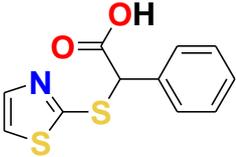
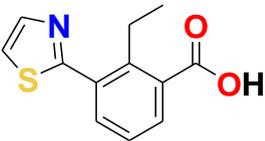
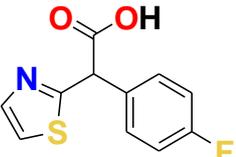
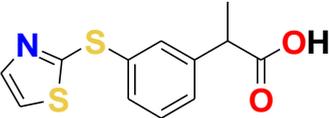
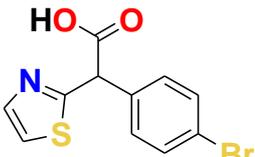
**Chapitre II : Etude de l'activité anti-inflammatoire par criblage virtuel à base de QSAR.**

03	118167276	8.897	37	137731769	7.287	71	104572707	6.649
04	115041951	8.674	38	90124074	7.276	72	80315933	6.646
05	20364693	8.487	39	89463211	7.194	73	146246448	6.642
06	154387846	8.428	40	82023539	7.182	74	89036522	6.638
07	84431301	8.414	41	84431129	7.155	75	79824290	6.624
08	115460897	8.261	42	124523015	7.155	76	82026629	6.613
09	84431302	8.247	43	40481274	7.146	77	15163893	6.589
10	67258063	8.161	44	28343223	7.106	78	58407251	6.581
11	63097704	8.051	45	15163889	7.086	79	80314552	6.572
12	57532085	7.978	46	21075087	7.059	80	70977821	6.559
13	63100004	7.973	47	82290099	7.040	81	137731807	6.557
14	12785130	7.968	48	152283546	7.018	82	119091954	6.554
15	115041953	7.952	49	19979384	7.010	83	140093434	6.537
16	105345135	7.939	50	67258370	7.005	84	23510471	6.501
17	105345069	7.935	51	137518964	7.001	85	15163949	6.497
18	153861563	7.908	52	82023545	6.990	86	15163890	6.483
19	90124076	7.892	53	25150333	6.954	87	122439186	6.470
20	11241445	7.888	54	105345136	6.943	88	15163902	6.469
21	119091942	7.872	55	115039550	6.934	89	58873071	6.469
22	84432031	7.872	56	137734023	6.922	90	42555835	6.466
23	64564771	7.832	57	134279354	6.894	91	59604477	6.454
24	21313634	7.748	58	82290101	6.893	92	82023541	6.432
25	115006319	7.709	59	84432901	6.834	93	15163947	6.423
26	117727893	7.677	60	157298022	6.828	94	23510389	6.417
27	22630869	7.648	61	118167275	6.818	95	15163950	6.415
28	80315349	7.637	62	113705810	6.770	96	71010155	6.411
29	80315545	7.634	63	62651340	6.767	97	70977822	6.388
30	115041952	7.561	64	82026628	6.749	98	15163913	6.384
31	57212451	7.523	65	80314745	6.745	99	81476149	6.373
32	15163912	7.366	66	80314946	6.712	100	79015695	6.314
33	25149912	7.364	67	80315543	6.706	101	<b>Indométacine</b>	7.000
34	15163887	7.358	68	155911699	6.671	102	<b>Ibuprofène</b>	5.996

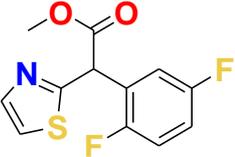
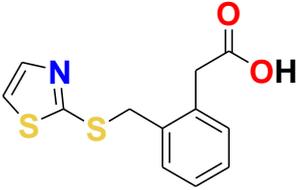
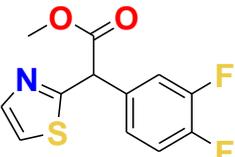
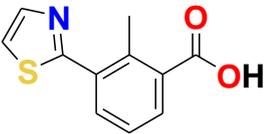
### 5.1. Visualisation des meilleures structures moléculaires

Parmi les 100 composés classés précédemment, il y a 51 composés possédant des valeurs de l'activité anti-inflammatoire supérieures à celles des inhibiteurs de référence tels que l'indométacine et l'ibuprofène. Les structures des dix meilleurs composés classés par le criblage virtuel sont illustrés dans le tableau 11.

**Tableau 11 :** Structures 2D des 10 meilleurs composés sélectionnés par criblage virtuel.

N°	Molécule ID	Structure	Nomenclature	Poids moléculaire (g/mol)
1	19698732		2-oxo-2-[2-(1,3-thiazol-2-yl)phenyl]acetic acid	233,24
2	61220635		2-Phenyl-2-(1,3-thiazol-2-yl)sulfanyl acetic acid	251,30
3	118167276		2-Ethyl-3-(1,3-thiazol-2-yl)benzoic acid	233,29
4	115041951		2-(4-Fluorophenyl)-2-(1,3-thiazol-2-yl)acetic acid	237,25
5	20364693		2-[3-(2-Thiazolylthio)phenyl]propionic acid	265,40
6	154387846		2-(4-Bromophenyl)-2-(1,3-thiazol-2-yl)acetic acid	298,16

**Chapitre II : Etude de l'activité anti-inflammatoire par criblage virtuel à base de QSAR.**

7	84431301		Methyl 2-(2,5-difluorophenyl)-2-(1,3-thiazol-2-yl) acetate	<b>269,27</b>
8	115460897		2-[2-(1,3-Thiazol-2-ylsulfanylmethyl) phenyl] acetic acid	<b>265,40</b>
9	84431302		Methyl 2-(3,4-difluorophenyl)-2-(1,3-thiazol-2-yl) acetate	<b>269,27</b>
10	67258063		2-Methyl-3-(1,3-thiazol-2-yl) benzoic acid	<b>216,26</b>

## **6. Conclusion**

Ce chapitre a été consacré à l'étude de criblage virtuel basé sur l'application de la méthodologie de QSAR, pour modéliser l'activité anti-inflammatoire en inhibant l'enzyme cyclooxygénase2 par les dérivés de l'acide 2-[4-(thiazole-2-yl) phényl] propionique. La méthode de régression linéaire multiple et les algorithmes génétiques sont utilisés dans le développement du modèle en tant que méthode d'apprentissage et de sélection respectivement.

La dimension du modèle développé est déterminé par la technique du point de virage et est constitué de cinq descripteurs moléculaires : H4e, B03+[N-O], F03[C-S], TE2 et R7e+. Les valeurs élevées des coefficients de détermination et la faible valeur de l'écart type confirment la bonne corrélation entre l'activité et les descripteurs moléculaires constituant le modèle. La stabilité, la robustesse et le pouvoir prédictif du modèle sont évalués par la validation croisée, le test de randomisation et la validation externe. Les résultats obtenus de la validation confirment la fiabilité et la robustesse du modèle obtenu.

Une chimiothèque a été collectée et préparée à partir de la base de données PubChem. Le modèle QSAR est appliqué pour prédire l'activité des molécules de la chimiothèque et les structures des 100 molécules ayant les meilleures activités prédites sont classées par ordre décroissant par rapport à leurs valeurs de pouvoir anti-inflammatoire.

## **7. Références**

1. J. R. Vane, *Nature New Biol*, 231, **1971**, 232- 235.
2. L. J Marnett, A. S. Kalgutkar, *Trends Pharm. Sci*, 20, **1999**, 465-469.
3. S. Steven, L. Wesolowski, L. William, *Bioorg. Med. Chem. Lett*, 12, **2002**, 267–270.
4. M. Negwer, *Organic-Chemical Drugs and Their Synonyms*, **1994**, 7th edn. Akademie Verlag, VCH Publishers, New York.
5. Y. Naito, Y. Yamaura, Y. Inoue, C. Fukaya, K. Yokoyama, Y. Nakagawa, T. Fujita , *Eur J Med Chem*, 27,**1992**, 645.
6. <https://pubchem.ncbi.nlm.nih.gov/>.
7. [http://www.talete.mi.it/products/dragon\\_molecular\\_descriptors.htm](http://www.talete.mi.it/products/dragon_molecular_descriptors.htm).
8. Y. Naito, T. Goto, F. Akahoshi, S. Ono, H. Yoshitomi, T. Okada, N. Sugiyama, S. Abe, S. Hanada, M. Hirata, M. Watanabe, C. Fukaya, K. Yokoyama, T. Fujita , *Chem. Pharm. Bull*, **1991**, 39:2323.
9. [www.hyper.com](http://www.hyper.com)
10. [www.talete.mi.it](http://www.talete.mi.it)
11. [www.minitab.com](http://www.minitab.com)
12. D. Liao, R. Valliant. *Techniques d'enquête*, 38(1), **2012**, 57-67.
13. K. KRAIM, « *Etudes de QSAR sur des activités biologiques utilisant des produits d'origines naturels* », Thèse de doctorat, Université de Guelma, Algérie, **2009**.
14. R. Rakotomalala, « *Pratique de la Régression Linéaire Multiple Diagnostic et sélection de variables* », Université Lumière Lyon 2, Version 2.0, **2011**.
15. A. Golbraikh, A. Tropsha, *Molecular Diversity*, 5, **2002**, 231-243.
16. R.E. Carhart et al, *J. Chem. Inf. Calcul. Sci*, 25, **1985**, 64-73.
17. K. Osmialowski, J. Halkiewicz, A. Radecki, R. Kaliszan, *J. Chromat*, 346, **1985**, 53-60.
18. A.R. Katritzky, E.V. Gordeeva, *J.Chem.Inf.Comput.Sci*, 33, **1993**, 835-857.
19. V. Consonni, R. Todeschini, M. Pavan, *J. Chem. Inf. Calcul. Sci*, 42, **2002**, 682-692.
20. V. Consonni, R.Todeschini, M.Pavan, P. Gramatica, *J. Chem. Inf. Calcul. Sci*, 42, **2002**, 693-705.

# **Conclusion générale**

## Conclusion générale

---

Nos travaux de thèse ont porté sur des études de criblage virtuel basées sur le docking moléculaire et la relation quantitative structure-activité, effectuées sur deux cibles thérapeutiques différentes : la première est la tyrosinase impliquée dans le processus de la production de la mélanine responsable de la coloration de la peau, des yeux et des cheveux, la seconde cible est la cyclooxygénase2 responsable de l'inflammation.

Les études de docking moléculaire ont été réalisées sur des analogues de gingérol et de l'arbutine extraits de la base de données ZINC. Les résultats obtenus à partir de docking indiquent que certains analogues ont montré un score de Moldock plus élevé (en termes d'énergie négative) par rapport au gingérol, à l'arbutine et aux inhibiteurs de tyrosinase connus expérimentalement. Les dix top hits, dans le cas des analogue du gingérol, ont montré une meilleure interaction que le gingérol, montrant une interaction moléculaire commune avec les résidus Met280 et Asn260 de l'enzyme tyrosinase. De plus, ces dix hits ancrés respectent la règle des cinq de Lipinski, affirmant que ces molécules n'ont aucun problème de biodisponibilité. Dans le cas des analogues de l'arbutine, les dix résultats les mieux classés ont montré une meilleure interaction que l'arbutine, montrant une interaction moléculaire commune avec les résidus Met280, His85, His61 et Asn260 de l'enzyme tyrosinase. De plus, ces molécules respectent la règle des cinq de Lipinski, affirmant qu'elles n'ont aucun problème de biodisponibilité.

Un modèle QSAR a été développé pour prédire quantitativement l'inhibition anti-inflammatoire avec un haut degré de précision en utilisant la méthode de régression linéaire multiple. Dans une première étape, l'anti-inflammatoire pour un ensemble de 47 dérivés d'acide 2-[4-(thiazole-2-yl) phényl] propionique a été modélisé avec succès par analyse de régression linéaire multiple, en utilisant les algorithmes génétiques en tant que méthode de sélection variables. Les résultats suggèrent que des variables comme  $R7e+$  ;  $H4e$  ;  $B03[N-O]$  ;  $F03[C-S]$  et  $TE2$  jouent un rôle important dans la définition de l'activité inhibitrice des dérivés d'acide 2-[4-(thiazole-2-yl)phényl] propionique contre l'enzyme COX2. Le modèle proposé a une bonne stabilité, robustesse et prédictivité lorsqu'il est vérifié par validation interne (validation croisée par LOO et randomisation Y) et également par validation externe.

## Conclusion générale

---

Le modèle QSAR construit a été appliqué dans le criblage virtuel de la base PubChem dans le but de proposer de nouvelles structures possédant une activité anti-inflammatoire intéressante et d'enrichir ainsi cette classe thérapeutique. 100 structures moléculaires ont été classées, théoriquement, par ordre décroissant en termes de pouvoir anti-inflammatoire. Parmi ces dernières 51 composés ont prouvé théoriquement un pouvoir anti-inflammatoire supérieur à celui de l'inhibiteur de référence, l'indométacine.

Il serait intéressant de valider ces résultats théoriques *in vitro* et *in vivo*.