

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université MOHAMED Boudiaf – M'Sila

Faculté des sciences

Département de chimie

Mémoire de Fin d'Études

En vue de l'obtention du diplôme de

Master (LMD)

Spécialité : chimie organique

Présenté et soutenu par :

Ben Fredj Amal et Lachache Sabira

Intitulé :

**Etude 3DQSAR, Docking moléculaire et criblage
ADMET pour des inhibiteurs potentiels de la
Cathepsine B**

Jury :

Dr. Mokhtar Djehiche

Président Université de Msila

Dr .Nabila Bennabila.

Examineur Université de Msila

Dr. Samir Bouacha

Rapporteur Université de Msila

Année universitaire : 2020 /2021

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



DEDICACES

Nous nous engageons sur ce travail :

Dédié à nos parents, merci

Pour tous leurs sacrifices, leur amour,

Leur tendresse, Leur soutien et leurs prières

Tout au long de notre processus d'apprentissage,

Pour tout Leur aide et leur présence dans nos vies.

Merci à nos frères et sœurs pour leurs encouragements

Soutien permanent et moral

A tous nos amis et familles

REMERCIEMENTS

Tout d'abord à remercier Allah qui nous a donné le courage, la patience ET la volonté qui nous ont permis de reprendre les études universitaires, après une longue rupture et pour finir ce travail.

*Nous remercions les plus profonds et les plus vifs à notre encadreur **SAMIR BOUACHA** pour ses conseils précieux, ses orientations, sa patience qui ont donné vie à ce travail Nos remerciements les plus vifs à Dr **MOKHTER DJECHICHE** pour avoir accepté de présider le jury de ce mémoire. Nos vifs remerciements également L'examinatrice **NABILA BENNABILA** pour ses conseils et pour avoir d'examiner ce travail et l'enrichir par ses propositions.*

Enfin, on espère que Ce mémoire sera à la hauteur de ses attentes et pourra compenser au moins une partie des

ملخص

في هذه الدراسة، تم استخدام الأساليب الأكثر استعمالاً في استراتيجيات اكتشاف جزيئات علاجية جديدة. من خلال هذا العمل، اعتمدنا على نموذج 3DQSAR وطريقة الإرساء وفحص ADMET من أجل توجيه و تخليق مشتقات الشالكون كمثبطات لكتبسين ب. لوحظ عند استخدام 3DQSAR أن هناك علاقة قوية بين النشاط التجريبي والنشاط المتوقع، مما يشير إلى الجودة العالية لنموذج 3DQSAR الذي تم الحصول عليه. استعملنا طريقة الإرساء من أجل فهم طريقة الارتباط بين المركبات النشطة المختارة والبروتين. عند فحص ADMET للمركبات النشطة، كانت النتائج متوافقة مع القيم المطلوبة.

الكلمات المفتاحية: كتبسين، نشاط بيولوجي، شالكون ، 3DQSAR ، الإرساء ، ADMET.

Résumé

Dans cette étude, les méthodes les plus couramment utilisées dans les stratégies de découverte de nouvelles molécules thérapeutiques sont utilisées. A travers ce travail, nous nous sommes appuyés sur le modèle 3DQSAR, la méthode d'amarrage et le dosage ADMET afin de diriger et prioriser la synthèse des dérivés de chalcone comme des inhibiteurs de la cathepcine B. Il a été observé lors de l'utilisation de 3DQSAR qu'il y avait une forte corrélation entre l'activité expérimentale et prédite, indique la haute qualité du modèle 3DQSAR obtenu. La méthode d'amarrage était de comprendre la méthode de liaison entre les composés actifs sélectionnés et la protéine. En vérifiant ADMET pour les composés actifs, les résultats étaient en accord avec les valeurs requises.

Mots clé : cathepcine B, Activité biologique, Chalcone, 3DQSAR, Docking, ADMET.

ABSTRACT :

In this study, the methods most commonly used in the strategies for discovering new therapeutic molecules are used. Through this work, we relied on the 3DQSAR model, the docking method and the ADMET assay in order to direct and prioritize the synthesis of chalcone derivatives as cathepcine B inhibitors. It was observed when using 3DQSAR that there was a strong correlation between experimental and predicted activity, indicating the high quality of the 3DQSAR model obtained. The docking method was to understand the method of binding between the selected active compounds and the protein. When checking ADMET for active compounds, the results were in agreement with the required values.

Keywords: cathepsine B, Biological activity, Chalcone, 3DQSAR, Docking, ADMET.

Table des matières

Dédicaces

Remerciements

Sommaire

Liste des figures

Liste des tableaux

Liste des abréviations

Introduction générale

CHAPITRE I : CANCER ET CATHEPSINE B

| | |
|--|-----------|
| I.1.1. épidémiologie du mot cancer..... | 19 |
| I.1.2. Généralité sur le cancer..... | 19 |
| I.1.3. Les causes de cancer..... | 20 |
| I.1.4. Division cellulaire..... | 21 |
| I.2. L'implication de la cathepsine B dans la progression deMs cancers..... | 22 |
| I.2.1. Généralités sur les protéases..... | 22 |
| I.2.2. Cathepsine B..... | 22 |
| I.2.2.1. Structure..... | 22 |
| I.2.2.2. Fonction biologique | 23 |
| I.2.2.3. Régulation de cathepsine B | 24 |
| I.2.2.4. role de la cathepsine B dans la développement de cancer et la métastase | 24 |
| I.3. Inhibiteurs naturels de cathepsine..... | 25 |

| | |
|---|-----------|
| I.3.1. Les chalcones..... | 25 |
| I.3.1. Définition des chalcones..... | 25 |
| I.3.2. Propriétés biologique..... | 26 |
| I.3.2.1. Activité antibactérienne..... | 26 |
| I.3.2.2. Activité anticancéreuse..... | 27 |
| I.3.2.3. Activité anti-inflammatoire..... | 27 |
| I.3.2.4. Activité anti-oxydant..... | 28 |
| I.3.2.5. Activité anti-HIV..... | 29 |
| I.3.2.6. Activité antipaludique..... | 29 |

| | |
|---|-----------|
| Références bibliographiques..... | 31 |
|---|-----------|

CHAPITER II : Méthodes de criblage virtuel in silico

| | |
|---|-----------|
| II.1.le criblage..... | 34 |
| II.1.1.Généralités de criblage..... | 34 |
| II.1.2.Criblage virtuel..... | 34 |
| II.1.2.1.Criblage virtuel « Ligand-based »..... | 35 |
| II.1.2.2.Criblage virtuel « Structure-based »..... | 35 |
| II.1.3.Filtrage ADMET-TOX..... | 36 |
| II.2. Docking moléculaire..... | 38 |
| II.2.1.Définition..... | 38 |
| II.2.2.Principe de docking..... | 39 |
| II.2.3.Les outils du docking moléculaire..... | 39 |
| II.2.4.Type de docking moléculaire..... | 39 |
| II.2.4.1.Docking rigide..... | 39 |

| | |
|---|-----------|
| II.2.4.2. Docking semi-flexible..... | 39 |
| II.2.4.3. Docking flexible..... | 40 |
| II.2.5. Scoring..... | 40 |
| II.2.5.1. Principe..... | 40 |
| II.2.5.2. Fonction de score..... | 40 |
| II.3. 3D de la relation structure activité quantitative (3D QSAR)..... | 41 |
| II.3.1. définition..... | 41 |
| II.3.2. Objectif de QSAR..... | 41 |
| II.3.3. Progrès dans les approches 3D-QSAR | 41 |
| II.3.4. Etudes 3D QSAR..... | 42 |
| II.3.5. Méthodes statistiques..... | 43 |
| Références bibliographiques..... | 45 |

CHAPITRE III : Résultats et discussion

| | |
|--|-----------|
| III.1. Introduction..... | 49 |
| III.2. Méthodologie de calculs..... | 52 |
| III.2.1. Base de donnée..... | 52 |
| III.2.2. Préparation de protéine..... | 52 |
| III.2.3. Génération de la grille du récepteur..... | 52 |
| III.2.4. Préparation de ligand..... | 53 |
| III.2.5. Docking..... | 53 |
| III.3. Résultats et discussion..... | 53 |
| III.3.2. Model3D-QSAR field based..... | 53 |
| III.3.1. Docking | 60 |
| III.3.3. Propriétés ADMET..... | 66 |

Références bibliographiques.....68

Conclusion générale.....69

Liste des figures

Chapitre I

| | |
|---|----|
| Figure I.1: Caractéristiques des cellules tumorales..... | 20 |
| Figure I. 2: Représentation schématique des étapes du processus cancérogène..... | 17 |
| Figure I. 3: La boucle d'occlusion est un élément structurel flexible qui dicte l'activité de la cathepsine B..... | 23 |
| Figure I. 6: Numérotation du squelette chalcone (a) et flavanone (b)..... | 26 |
| Figure I. 7: Chalcones à activité anticancéreuse..... | 28 |
| Figure I. 8: Chalcones à activité anti-inflammatoire..... | 28 |
| Figure I. 9: Chalcones à activité antioxydant..... | 29 |
| Figure I. 10: Chalcone à activité anti-HIV..... | 30 |
| Figure I. 11: Chalcone à activité antipaludique..... | 30 |

Chapitre II

| | |
|---|----|
| Figure II. 1: Illustration schématique des approches ligand-based et structure-based..... | 35 |
| Figure II. 2: Représentation schématique de devenir d'un médicament dans l'organisme | 37 |
| Figure II. 3: Docking moléculaire..... | 38 |
| Figure II. 4: Modèle de l'étude de relation structure activité..... | 42 |

Chapitre III

| | |
|---|----|
| Figure III. 1: Dérivés de chalcone & activité IC ₅₀ | 52 |
| Figure III. 2: graphique de corrélation ente PIC ₅₀ expérimental et prédictif des ensembles d'apprentissage et de test..... | 56 |
| Figure III. 3. a: Aires du champ gaussien stérique pour ligand 17 (ligand de référence)..... | 58 |
| Figure III. 3. b: Aires du champ gaussien d'hydrophobe pour ligand 17..... | 59 |
| Figure III. 3. C: Aires du champ gaussien d'accepteur de liaison d'hydrogène pour ligand 17..... | 59 |
| Figure III. 3. d: Aires du champ gaussien électrostatique pour le ligand 17..... | 60 |
| Figure III.4a : Interactions 2D et 3D du ligand 6 dans le site actif du récepteur..... | 62 |
| Figure III.4b : Interactions 2D et 3D du ligand 14 dans le site actif du récepteur..... | 63 |
| Figure III.4c : Interactions 2D et 3D du ligand 17 dans le site actif du récepteur..... | 64 |

Figure III.4d : Interactions 2D et 3D du ligand 18 dans le site actif du récepteur.....65

Liste des Tableaux

Chapitre III

| | |
|---|----|
| Tableaux III.1: Données d'activité prédite pour l'apprentissage et le test de l'ensemble de dérivés de chalcone..... | 55 |
| Tableaux III.2: Valeurs statistiques pour le modèle 3D-QSAR généré par PLS..... | 56 |
| Tableaux III.3: Détails des calculs du modèle Field-based QSAR..... | 56 |
| Tableaux III.4: Résultats de l'analyse d'amarrage des ligands sélectionnés..... | 60 |
| Tableaux III.5: Criblage in silico ADMET pour des ligands sélectionnés. | 66 |

Liste des abréviations

ROS: Reactive Oxygen Species

HIV: Virus de l'Immunodéficience Humaine

API: Protéine 1

NF- κ P: Facteur Nucléaire.

ADN: Acide DésoxyriboNucléique

RMSD: Root Mean Square Deviation, représente l'écart entre deux conformations

HITS: HIgh Throughput Screening

RMN: Résonance Magnétique Nucléaire

ADMET: Absorption Distribution Métabolisme Elimination Toxicité

VPH: Human Pappilloma Virus

RER: Rough Endoplasmic Reticulum

ARNm: Acide RiboNucléique messenger

ECM: Matrice ExtraCellulaire

UPA: Activateur Plosminogène Urokinase

TPA: Activateur Plosminogène Tissulaire

QSAR: Quantitative Structure-Activity Relation

RI: Receptor Indoependent

PHASE: PHarmacophore Alignment and Scoring Engine

PDB: Protein Data Bank

PLS: Partial Least Squares

CoMSA: Comparative Molecular Surface Analysis

CoMFA: Comparative Molecular Field Analysis

IC50: la moitié de la Concentration maximale Inhibitrice

THS : Criblage à haut débit (High-throughput screening)

RSA : Relation structure-activité

R&D : Recherche et de développement

VDW : Van Der Waals

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

PH : Plekstrin Homology.

ΔG : Énergie d'activation

Les acides aminés

Ile : Isoleucine

The : Threonine

His : Histidine

Asp : acide Aspartique

Arg : Arginine

Introduction Générale

Selon l'OMS (Organisation Mondiale de la Santé), les cancers sont une des causes majeures de décès dans le monde. Ces maladies sont à l'origine de 7,6 millions de décès en 2008, soit 13 % de la mortalité mondiale . D'après les estimations, la mortalité due aux cancers va continuer à augmenter pour dépasser 13,1 millions de décès en 2030. Le cancer est une maladie complexe qui peut atteindre tous les types de tissus du corps humain. Il est d'abord caractérisé par une prolifération importante et incontrôlée des cellules au sein d'un tissu normal ce qui est nocif pour son homéostasie et sa survie [1].

La cathepsine B est une protéase lysosomale dont le rôle majeur est de dégrader les protéines en fin de vie. Dans les cancers colorectaux, l'épithélium colique possède une importante activité cathepsine B dès l'apparition d'adénomes précoces, mais aussi dans des tumeurs avancées et métastatiques. Par ailleurs, une expression et une activité importante de cette enzyme sont des indices de mauvais pronostics de survie chez les patients atteints de cancers colorectaux. [2]

Les inhibiteurs de cathepsine qui bloquent l'activité de certaines cathepsines ont fait l'objet d'études qui testent le traitement d'un large éventail de maladies humaines.

La chemoinformatique est un outil de choix pour réduire le temps et le coût de développement d'un médicament. La bio-informatique apporte elle aussi sa participation à l'étude des interactions protéine-ligand par des méthodes comme « dynamique moléculaire » et « protéine-ligand docking »

Le docking moléculaire in silico vise à prévoir la structure d'un complexe moléculaire à partir des composées isolées, ce qui est notablement plus facile à mettre en œuvre, moins cher et plus rapide que l'utilisation des méthodes expérimentales in vitro. Les logiciels de docking sont donc des instruments très utiles en biologie, pharmacie et médecine, car la majorité des principes actifs sont de petites molécules (ligand) qui interagissent avec une cible biologique d'intérêt thérapeutique, habituellement protéique (récepteur), afin d'influencer le mécanisme dans lequel cette protéine est impliquée.

L'étude 3D-QSAR et d'amarrage (docking) pourrait offrir plus d'informations pour comprendre les caractéristiques structurales du site de liaison de la protéine et le détail des interactions protéine-ligand pour diriger la conception de nouvelles molécules potentielles.

Dans la présente étude, les méthodes les plus utilisées dans les stratégies de découverte de nouvelles molécules à visée thérapeutiques sont utilisées. Au cours de ce travail, nous nous basons sur le modèle 3D QSAR, la méthode d'amarrage (docking) et le criblage in silico ADMET afin de guider et de prioriser la synthèse des molécules comme des inhibiteurs de la cathepsine B.

Pour faire ce travail d'une manière convenable on a divisé en trois chapitres :

- Le premier chapitre est présenté un aperçu sur le cancer et un rappel bibliographique sur La cathepsine B et les inhibiteurs les plus significatifs de dérivés de chalcone.
- Le deuxième chapitre est consacré de l'intérêt du criblage dans le cadre de la recherche pharmaceutique pour répondre aux besoins de santé publique actuels en détaillant sur les différentes méthodes de criblage virtuel (Docking moléculaire, 3D-QSAR) et ADMET).
- Le troisième chapitre (partie pratique) regroupe les données relatives aux matériels et à méthodes utilisées au cours de ce travail. Puis nous exposerons le criblage par l'étude 3D-QSAR, docking moléculaire sur une chimiothèque virtuelle de dérivés de chalcone les plus significatives dans le but de sélectionner les composés prometteurs qui réagissent plus favorablement avec la cathepsine B.

Enfin, cette mémoire se termine par une conclusion générale.

Références

- [1] R. J. Van Vuuren, M. H. Visagie, A. E. Theron, and A. M. Joubert, “Antimitotic drugs in the treatment of cancer,” *Cancer Chemother. Pharmacol.*, vol. 76, no. 6, pp. 1101–1112, 2015, doi: 10.1007/s00280-015-2903-8.
- [2] Fonović M, Turk B. Cysteine cathepsins and extracellular matrix degradation. *Biochim. Biophys. Acta* 2014; 1840: 2560–2570.

CHAPTER I

Cancer & Cathepsine B

I.1.Le cancer

I.1.1. Epidémiologie du mot cancer

Hippocrate (460-370 avant J.-c), le père de la médecine, a décrit précisément le cancer et a utilisé les mots « carcinos » ou « carcinoma » pour définir des renflements qui apparaissaient comme des tumeurs malignes. Le mot carcinos en grec signifie : (1) crabe (2) cancer comme signe zodiacal, (3) ulcère, (4) tragédien attique et (5) tout ressemblant au pince de crabe (e. g. pince, compas). Le médecin Romain Celsus (28 avant J.-C. -50 après J.-C) traduisit le mot grec « carcinos » en latin par « cancer ». Il a également introduit le mot grec « carcinoma » dans la langue latine.

Le terme « cancer » désignant des ulcères d'allure maligne avec pénétration profonde tandis que le terme « carcinoma » désignait des lésions malignes plus superficielles. Par la suite, peu de nouveaux termes sont introduits pour désigner les tumeurs. Au début du 19^e siècle, le terme grec « carcinoma » est devenu synonyme du terme « cancer » et la terminaison «-orna » est utilisée pour désigner certaines lésions cancéreuses [1].

I.1.2.Généralité sur le cancer

D'après le DR BODIN, la définition du cancer serait : « prolifération importante, anarchique et indéfinie de cellules anormales d'un tissu ou d'un organe qui ont la capacité d'envahir et de détruire les tissus sains et de se disséminer dans l'organisme et dont l'évolution spontanée est la mort de l'individu porteur » [2].

Le cancer n'est pas une seule maladie ; le terme cancer recouvre en fait un vaste ensemble de maladies, parfois très différentes les unes des autres, cataloguées selon les cellules, les tissus, les organes à partir desquels les cancers se forment ; il existe donc presque un cancer pour chaque organe ou tissu. HANAHAN et WEINBERG résumant ainsi les caractéristiques des cellules cancéreuses (**figure.I.1**).

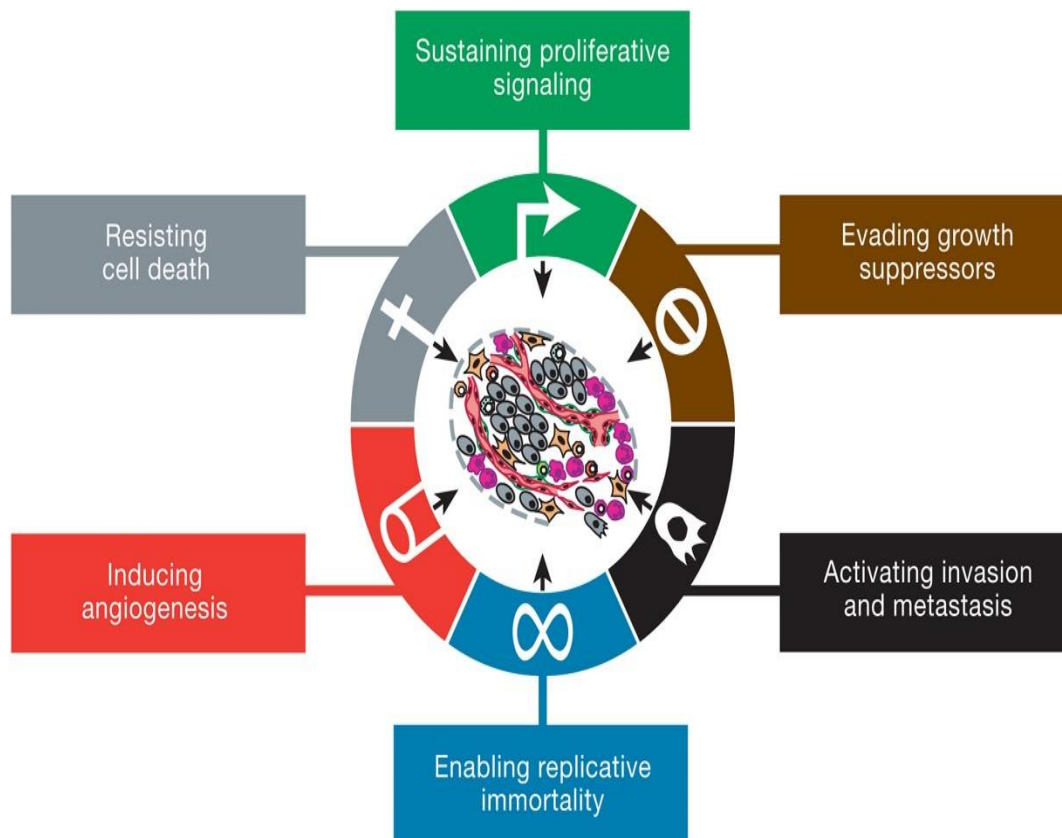


Figure I.1 : caractéristiques des cellules tumorales [3]

En résumé, la cancérisation est souvent liée aux phénomènes suivants :

- Altération des systèmes de réparation de l'ADN qui favorise la survenue de lésions génétiques.
- Altération des gènes impliqués dans la surveillance du génome et la régulation de l'apoptose (mort cellulaire programmée) qui autorise la survie de cellules anormales.

Enfin, la cancérogenèse est influencée par des facteurs de risque qui peuvent être internes (prédispositions et mutations génétique, facteurs hormonaux) ou externes (exposition à des substances cancérogènes, risque environnementaux...); mais ces risques ne sont pas toujours bien définis ou bien identifiables.

I.1.3. Les causes du cancer

Ce que l'on sait sur les causes du cancer provient d'une part de l'étude de l'évolution du cancer dans les populations humaines et, d'autre part, de l'observation des processus d'induction de tumeurs chez l'animale de laboratoire traité par des agents cancérogènes. Le tabac, l'amiante, les

aflatoxines et les rayons ultraviolets font partie des cancérogènes les plus importants pour l'homme. Par ailleurs, près de 20 des cancers sont associés à des infections chroniques, dont les principales sont dues aux virus de l'hépatite (VHB, VHC) et du papillome humain (VPH) [4], et à la bactérie *helicobacter pylori*. Il est de plus en plus largement reconnu que les facteurs liés au mode de vie, comme l'alimentation, l'activité physique et la consommation d'alcool, jouent un rôle dans le développement du cancer. La prédisposition génétique peut aussi fortement influencer les risques liés aux expositions environnementales.

Présentement, les chercheurs déploient beaucoup d'effort afin d'élucider les parties subissant des modifications génétiques au niveau de l'ADN et étant à l'origine du cancer. Les gènes responsables d'un cancer sont appelés oncogènes. Lors que les oncogènes subissent une modification et sont activés de façon inappropriée, ils peuvent transformer une cellule normale en cellule cancéreuse. En général, les oncogènes se forment à partir de gènes normaux qui régulent la croissance et la développement, les proto-oncogènes. Les proto-oncogènes accomplissent des fonctions normales dans toutes les cellules jusqu'à ce qu'ils subissent une perturbation maligne. Ils peuvent également être activés anormalement à la suite d'un réarrangement des chromosomes.[5]

I.1.4. Division cellulaire

La division cellulaire est le mode de multiplication de toutes cellules. Elle est un processus fondamental dans le monde du vivant puisqu'elle est nécessaire à la reproduction, à la formation et au maintien de tous les organismes eucaryotes. La division cellulaire est cyclique (cycle cellulaire) et comprend 4 phases successives : G₁ (intervalle de croissance un), S (synthèse ADN), G₂ (intervalle de croissance deux) et M (mitose). La phase G₀ de la cellule est l'état où la cellule ne se divise pas et correspond à son état de repos. Pendant cette période, la cellule accomplit les fonctions normales qui lui sont prédestinées. La phase G₁ est la phase du cycle où la cellule croît en taille et synthétise les protéines cytoplasmiques nécessaires à la division cellulaire. La durée de la phase G₁ est la plus longue et correspond à la préparation de la division cellulaire. La phase S est la phase où la cellule transcrit activement les gènes des protéines nécessaires à leur survie et au maintien des fonctions spécialisées. La durée de cette phase varie de quelques minutes à plusieurs heures selon le type des cellules. Pendant la phase S, la cellule dédouble aussi les centrosomes, nécessaires à la migration des chromosomes. Les deux centrosomes migrent autour du noyau et se positionnent de façon diamétralement opposées. La phase G₂ est une phase de contrôle de la bonne transcription du matériel génétique, impliquant des gènes et des protéines importants pour le bon déroulement de la mitose. Elle est très souvent

souvent courte (quelques heures au maximum). Il s'agit d'une phase automatique : les cellules entrées dans la phase S aboutissent toujours à la phase M. La phase M est la phase finale de la division cellulaire. Elle est sous-divisée en 5 phases : la prophase, la métaphase A, l'anaphase B et la télophase [6].

I.2. l'implication de la cathepsine B dans la progression des cancers

I.2.1. généralités sur les protéases

Les protéases (peptidase ou enzymes protéolytiques) [7] sont des enzymes hydrolytiques qui catalysent le clivage des liaisons peptidiques clivant les protéines en fragments peptidiques plus petits [8] elles peuvent être spécifiques conduisant à un clivage sélective des protéines pour la modification et la maturation post-traductionnelle [7].

Les protéases sont capables de dégrader in vitro un nombre multiple de substrats et, la plupart du temps, plusieurs protéases ont la capacité de dégrader in vitro un même substrat. Cette caractéristique découle du fait que la spécificité d'une protéase n'intervient généralement peu ou pas au niveau de la conformation du substrat, mais plutôt au niveau de la liaison peptidique à cliver [9].

I.2.2. Cathepsine B

La cathepsine B une cystéine protéase lysosomales qui est structurellement similaire à l'enzyme papaine, est l'une des 11 cystéines cathepsines humaines (B, C, F, H, L, K, O, S, V, W, X/Z). La cathepsine B est exprimée de manière constitutive et liée au renouvellement général des protéines dans les lysosomes. La cathepsine B est synthétisée sur le réticulum endoplasmique rugueux (RER) sous forme de préproenzyme de 339 acides aminés avec un peptide signal de 17 acides aminés. Le peptide signal dirige la protéine dans la lumière du RER où le peptide signal est éliminé et une forme précurseur inactive de 43/46 KDa, la procathepsine B, est formée. La procathepsine B est ensuite transportée à travers le RER vers l'appareil de Golgi où elle est glycosylée au niveau de deux résidus d'asparagine par des oligosaccharides contenant du mannose avec des résidus de mannose phosphorylés. La protéine phosphorylée se lie aux récepteurs du mannose-6-phosphate dans le réseau trans-Golgi et est transportée vers les lysosomes via des vésicules de transport [10].

I.2.2.1. Structure

La cathepsine B est une protéine bilobale, le site actif et la fente de liaison au substrat sont situés à l'interface entre les deux lobes. La cathepsine B possède un élément structurel

supplémentaire qui le distingue du reste des cathepsines à cystéine (**figure.2**). Cet élément s'appelle l'occlusion boucle et comprend 20 acides aminés (Ile105-Thr125). dans la conformation fermée, la boucle est maintenue sur le corps de l'enzyme par deux ponts salins (His110-Asp22 et Arg116-Asp224) qui ferment le haut du site actif fente et empêcher la liaison de grands substrats d'endopeptidase dans les sous-sites S2' et S3' [11]. Cela confère sur la cathepsine B une activité endopeptidase faible par rapport à celles des autres cystéine cathepsines. Au en même temps, His 111, qui est situé en haut de la boucle, interagit avec le carboxylate C-terminal du peptide substrats, potentialisant ainsi l'exopeptidase (peptidyldipeptidase) activité de l'enzyme [12].

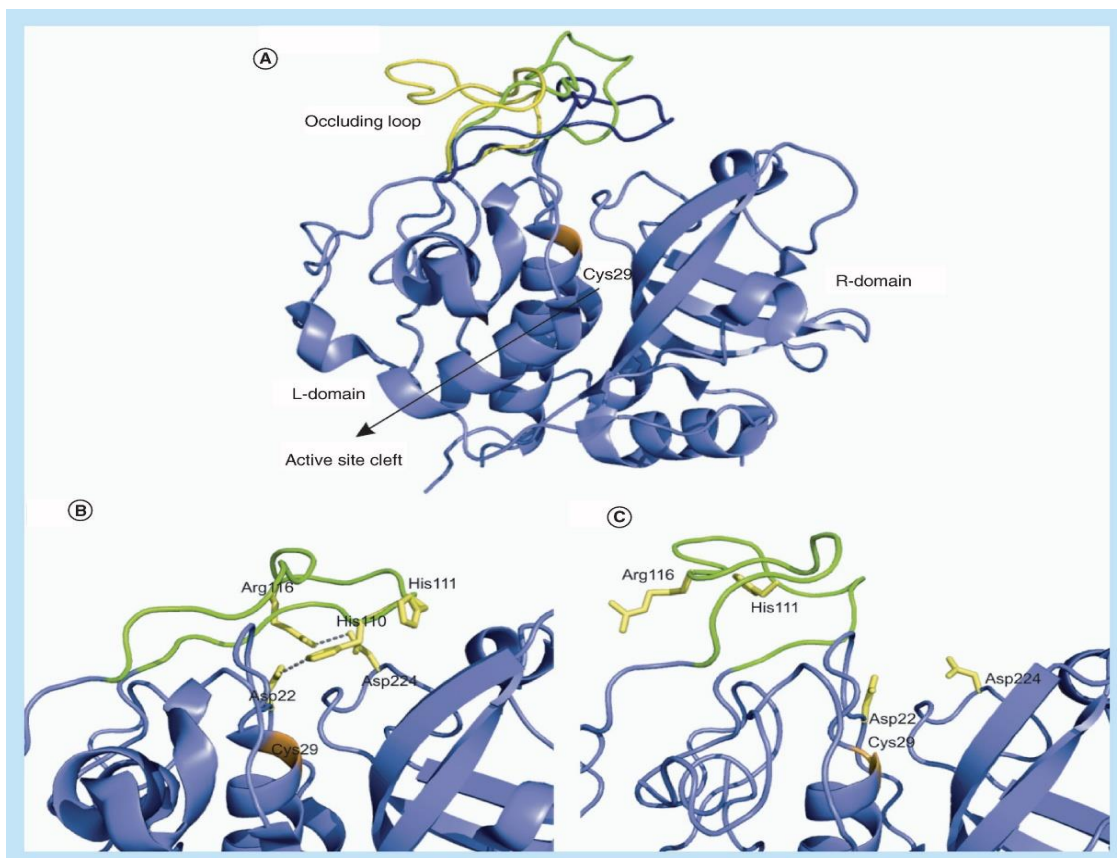


Figure.I.2 :: la boucle d'occlusion est un élément structurel flexible qui dicte l'activité de la cathepsine B

I.2.2.2. Fonction biologique

La distribution omniprésente de la cathepsine B suggère qu'il remplit une fonction essentielle. Bien qu'aucune expérience de désactivation de gène testerait directement cette hypothèse aurait été publiée, action et localisation de la cathepsine B dans les lysosomes impliquent qu'il est important dans dégradation des protéines. Les endopeptidases initient dégradation des protéines par clivages des liaisons peptidiques dans la séquence de la protéine substrat, loin des

terminaisons amino et carboxy. Cela conduit à une augmentation de nombre de nouveaux terminaisons, permettant aux exopeptidases de fonctionner avec une plus grande efficacité car leur substrat la concentration est augmentée. Avec sa double action comme endopeptidase et peptidyle dipeptidase, la cathepsine B est équipée pour participer aux deux stades précoces et tardifs de la protéine lysosomale panne. Dans certaines situations, la cathepsine B peut être remplissant un rôle plus spécifique, comme dans le traitement protéolytique de leur précurseur, la thyroglobuline [13].

I.2.2.3. Régulation de cathepsine B

L'action de la cathepsine B est régulée de différentes manières :

-Au niveau de l'expression, les variantes d'épissage d'ARNm diffèrent dans leur stabilité et leur taux de traduction [14]. Les variantes d'épissage alternatives sans exon 2 et 3 codent des formes tronquées de cathepsine B entraînant une altération de localisation. La cathepsine B tronquée manque une cascade d'événements entraînant la mort cellulaire [15].

-La cathepsine B est synthétisée sous forme de zymogène inactif qui doit être activé [16].

-Compartmentalisation endosomale / lysosomale les vésicules limitent l'accès à ses substrats.

-Le rôle de la cathepsine B diffère selon sa localisation cellulaire.

-L'optimum de pH de l'activité exopeptidase de la cathepsine B est de 5.0 et, d'activité endopeptidase, au-dessus de pH 7.0 [17]. A l'intérieur des lysosomes, la cathepsine B agit principalement comme une exopeptidase et, dans la membrane plasmique et dans le cytosol, en tant qu'endopeptidase. L'activité endopeptidase est typique de la dégradation pathologique de protéines dans l'ECM, donc la découverte de nouveaux inhibiteurs de la cathepsine B comme antitumoral potentiel les médicaments devraient être axés sur les composés qui peuvent inhiber spécifiquement l'activité endopeptidase tout en préservant l'activité « physiologique » ;

-les glycosaminoglycanes potentialisent l'activité endopeptidase de la cathepsine B [17] ; les inhibiteurs endogènes sont d'importants régulateurs d'activité de la cathepsine B.

I.2.2.4. Rôle de la cathepsine B dans le développement de cancer et la métastase

La cathepsine B joue un rôle important dans la protéolyse intracellulaire ; les niveaux de cathepsine B sont régulés à la hausse dans de nombreux microenvironnements tumoraux différents. La cathepsine B peut améliorer l'activité d'autres protéases, y compris la

métalloprotéinase matricielle, l'urokinase et la cathepsine D. En outre, elle occupe une position cruciale dans le processus protéolytique des composants de la matrice extracellulaire, dans l'interruption de la communication intercellulaire et l'inhibition d'une autre expression de protéases. Ainsi, il est impliqué dans le remodelage et la dissolution de la membrane basale et du tissu conjonctif dans les processus de croissance tumorale en invasion et en métastase, ce qui peut entraîner une dégradation et une invasion de l'ECM via des lysosomes sécrétés.

La cathepsine B est également impliquée dans l'autophagie et le catabolisme, qui sont des processus bénéfiques dans la malignité tumorale. La cathepsine D est l'une des protéases responsables de la conversion de la pro-cathepsine B en cathepsine B active. la cathepsine B est déclenchée par des protéases différentes comme la cathepsine G, l'activateur du plasminogène de type urokinase(UPA), l'activateur du plasminogène de type tissulaire (TPA) et l'élastase. L'interaction entre la cathepsine B et l'UPA n'est pas unique ; en fait, la cathepsine B peut également activer UPA. Un tel cas de réciprocité souligne l'évaluation selon laquelle les interactions protéolytiques ne sont pas une cascade unidirectionnelle plus simple mais font partie d'un réseau complexe.[18]

I.3. Inhibiteurs naturels de cathepsine

Plusieurs produits naturels sont des inhibiteurs de cathepsine, et les produits naturels sont souvent le premier domaine d'étude pour ceux qui examinent dans les candidats potentiels de médicament pour un large éventail de maladies. Les inhibiteurs de cathepsine peuvent être extraits des centrales, des champignons et des bactéries. Les inhibiteurs naturels de cathepsine comprennent :

- **Aspergillus**
- **Les dérivés d'acides aminés**
- **Bifalvones**
- **Dérivés de chalcone**

Notre travail est basé sur les dérivés de chalcone, donc on fait un rappel sur les chalcones et leurs utilisations.

I.3.1.les chalcones

I.3.1.1.Définition des Chalcones

Les chalcones (**figure.I.1**) sont uniques au sein de la famille des flavonoïdes. Dépourvus du cycle C central, les deux cycles A et B sont reliés par une chaîne tri carbonée cétonique α , β insaturée, ces

composés sont rarement substitués sur le cycle B. les positions 2',4' et 6' du cycle peuvent être hydroxylées ou méthoxylées [19].

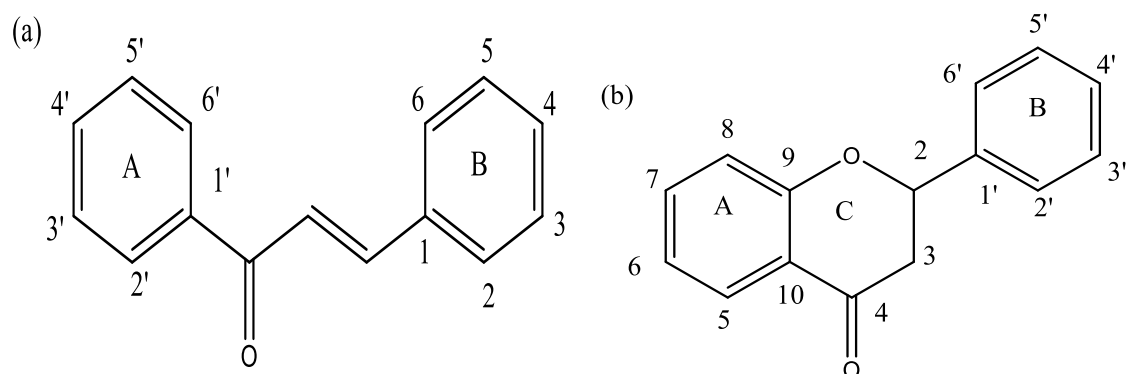


Figure I.3 : Numérotation du squelette chalcone (a) et flavanone (b).

La présence d'une double liaison conjuguée confère aux chalcones une couleur jaune. Elles peuvent exister sous forme de deux stéréo-isomères Z et E, les stéréo-isomères E sont les plus abondants.

Les chalcones présentent un groupe très intéressant parmi les produits naturels. Elles forment les intermédiaires essentiels dans la biosynthèse des autres flavonoïdes [20].

I.3.1.2. Propriétés biologique

Les molécules contenant des unités de la chalcone présentent plusieurs activités biologique parce que existe la double liaison et du groupe carbonyle dans les chalcones[21]. Elles sont donc très largement décrites dans la littérature et montrent des potentiels pharmacologiques multiples. Parmi ces activités on peut citer :

I.3.2.1. Activité antibactérienne

L'activité antibactérienne des chalcones est reliée à la présence à la fois de la double liaison et de la fonction carbonyle. Des tests in vivo ont montré que la 3'-nitro-4'-hydroxy-2-méthoxychalcon possède une action antibactérienne élevée. Les chalcones substituées par des halogènes (chlore, brome, iode et fluor) sur les cycles aromatiques ont également des activités antibactériennes intéressantes [22].

I.3.2.2. Activité anticancéreuse :

Les dérivés de la chalcone sont utilisés pour lutter contre les différents types de cancers. En effet, ils sont capables d'empêcher la prolifération des cellules tumorales et d'inhiber l'activité des interleukines-5 (qui stimulent les globules blancs). Ils se révèlent particulièrement efficaces contre les cancers des ovaires, de l'utérus et du sein [23].

L'isolate retigenin **4a**, le flavokawain **4b** et l'hamilcone **4c**, sont des exemples de composé anticancéreux.

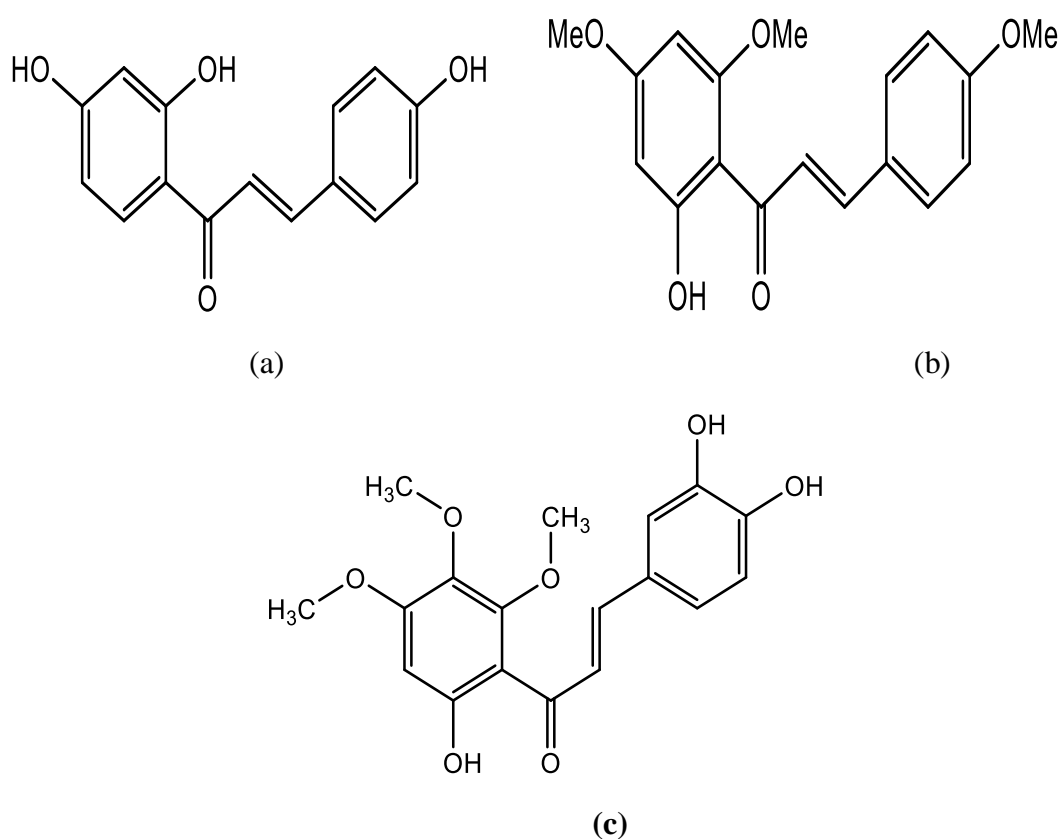


Figure I.4 : chalcones à activité anticancéreuse.

I.3.2.3. Activité anti-inflammatoire :

La 4-hydroxycorodoine **5a** qui est rapporté à se produire dans lonchocarpus neuroscapha possède un effet anti-inflammatoire qui réduit l'activation du facteur nucléaire KB (NF-KB) et active les voies de la protéine-1 (AP-1). Le Flavokawain **5b** est un puissant anti-inflammatoire qui a été isolé à partir d'Alpinia pricei. [24-25]

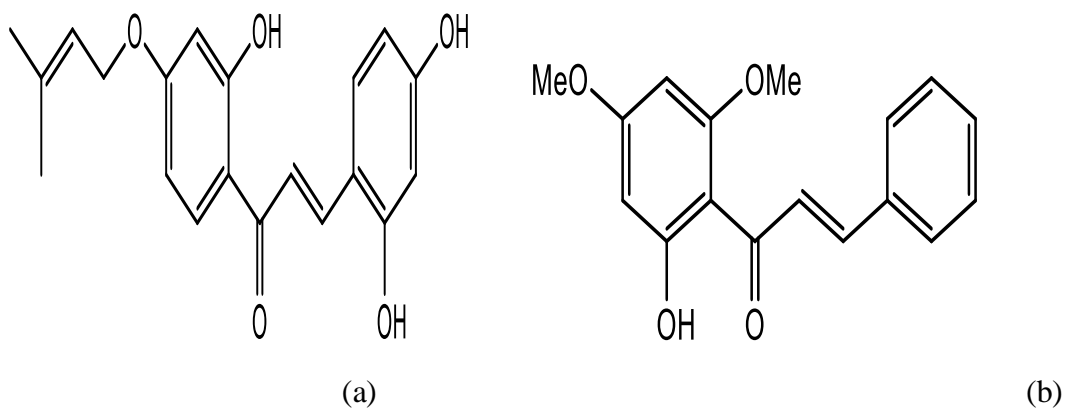


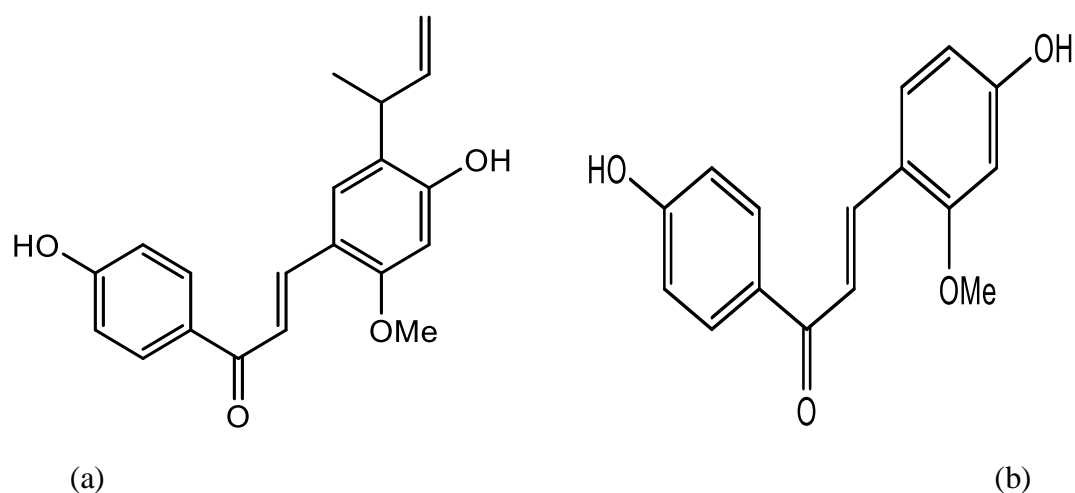
Figure I.5 : chalcones à activité anti-inflammatoire.

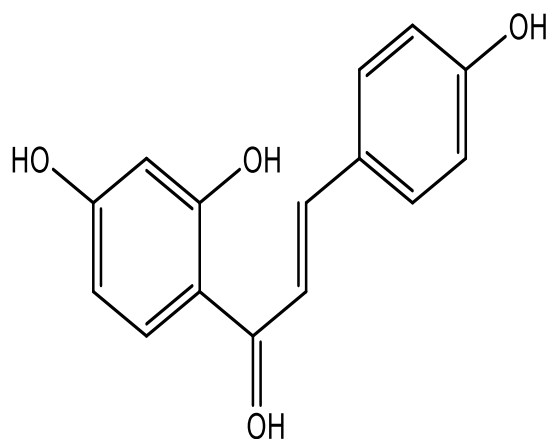
I.3.2.4. Activité antioxydant :

Les mécanismes de l'action d'un antioxydant peuvent comprendre [26]:

- Le piégeage direct des espèces réactives oxygénées (ROS).
- L'inhibition des enzymes et la chélation des traces métalliques responsables de la production de ROS.
- La protection des systèmes de défense antioxydants.

Les chalcones possédant des activités antioxydants telles que la licochalcone **6a**, l'échination **6b** et l'isoliquiritigénine **6c** ont été isolées de la plante *Glycyrrhiza inflata*. [27]





(c)

Figure I.6 : chalcones à activité antioxydant

I.3.2.5. Activité anti-HIV:

Parmi les chalcones synthétiques biologiquement actives, on peut citer quelques types, le premier est représenté par une chalcone fluorée **7** qui a montré une bonne activité contre le virus de l'immunodéficience humaine HIV. [28]

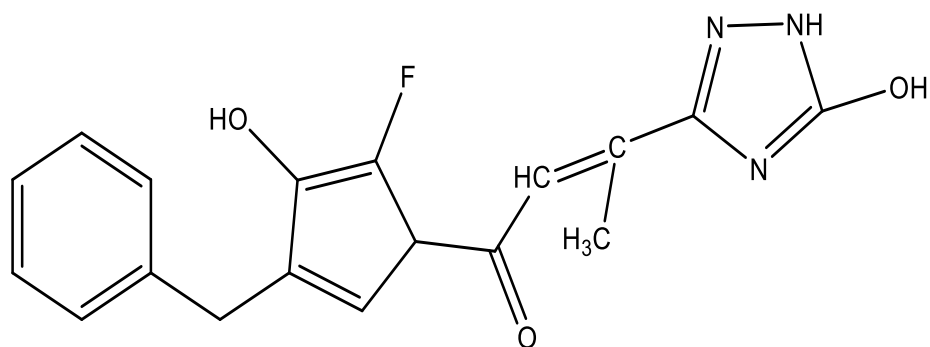


Figure I.7: chalcone à activité anti-HIV.

I.3.2.6. Activité antipaludique :

La série de chalcones synthétiques, contenant le noyau quinoléique, a révélé une activité antipaludique intéressante [29].

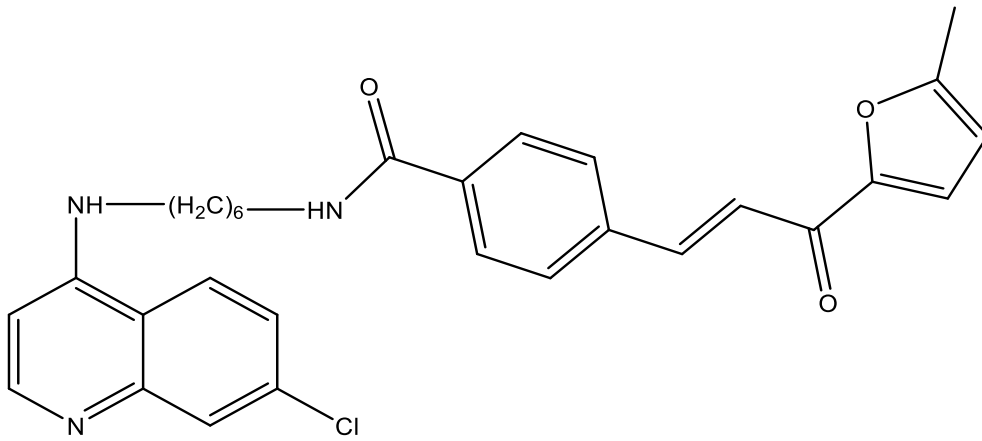


Figure I.8 : Chalcone à activité antipaludique.

Références

- [1].Pettit, G. R., & Lippert, J. W. Antineoplastic agents 429. Syntheses of the combretastatin A-1 and combretatin B-1 prodrugs. *Anticancer Drug design*, **2000**, 15(3), 2003-2016.
- [2].Sanchez, C. Développement de vecteurs de pénétration intracellulaire pour un adressage d'inhibiteurs de la cathepsine D (Doctoral dissertation, université Montpellier), **2016**, P32.
- [3].Hanahan, D., & Weinberg, R. A. Hallmarks of cancer. The next generation. *Cell*, **2011**, 144(5), 646-674.
- [4].Matsuyama, A., Shimazu, T., Sumida, Y., Saito, A., Yoshimatsu, Y., Seigneurim-Berny, D., ... &Yashida, M. (). In vivo destabilization of dynamic microtubules by HDAC6-mediated deacetylation. *The EMBO journal*, **2002**, 21(24), 6820-6831.
- [5].Fortin, E., Fortin, S., Desjardins, M., Rousseau, J. L., Petitclerc, E., C.-Gaudreault, R. *Bioorg. Med. Chem. Lett.* **2007**, 17, 2000.
- [6].Meijer, L. Le cycle de division cellulaire et sa régulation. *ONCOLOGIE-PARIS-*, **2003**, 5(7/8), 311-326.
- [7].souza, P. M. D., Bittencourt, M. L. D. A., Caprara, C. C., Freitas, M. D., Almeida, R. P. C. D., silveira, D., ... & Magalhães, P. O. Abiotechnology perspective of fungal proteases. *Brazilian Journal of Microbiology*, **2015**, 46(2), 337-346.
- [8].Radha, S., Babu, R. H., Sridevi, A., Prasad, N. B. L., &Narasimha, G. Development of mutant fungal strains of *Aspergillus niger* for enhanced production of acid protease in submerged and solid state fermentation. *European Journal of Experimental Biology*, **2012**, 2, (5), 1517-1528.
- [9].Authier, F., Taupin, V., Posner, B. I., & Bergeron, J. J. M. Les protéases multifonnelles de famille des insulinasés, **1997**, P601.
- [10].Aggarwal, N., & Slone, B. F. Cathepsin B: multiple roles in cancer. *PROTEOMICS-Clinical Applications*, **2014**, 8, 427-437.
- [11].Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R & Katunuma, N. The refined 2. 15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *The EMBO journal*, **1991**, 10(9), 2321-2330.
- [12].Dubowchik, G. M., & Firestone, R. A. Cathepsin B-sensitive dipeptide prodrugs. 1. A model study of structural requirements for efficient release of doxorubicin. *Bioorganic medicinal chemistry letters*, **1998**, 8 (23), 3341-3346.
- [13].Mort, J. S., & Buttle, D. J.(1997). Cathepsin b. *The international journal of biochemistry & cell biology*, 29(5), 715-720.
- [14].Mohmed, M. M., & Sloane, B. F. Multifunctional enzymes in cancer. *Nature Reviews Cancer*, **2006**, 6(10), 764-775.
- [15].Baici, A., Muntener, K., Willimann, A., & Zwicky, R. . Regulation of human cathepsin B by alternative ARNm splicing: homeostasis, fatal errors and cell death, **2006**.

- [16].Pungercar, J. R., Caglic, D., Sajid, M., Dolinar, M., Vasiljeva, O., Pozgan, U & Turk, B. . Autocatalytic processing of procathepsin B is triggered by proenzyme activity. The FEBS journal, **2009**, 276(3), 660-668.
- [17].Almeida, P. C., Nantes, I. L., Chagas, J. R., Rizzi, C. C., Faljoni-Alario, A., Carmona, E., ... & Tersariol, I. L. .Cathepsin B activity regulation: heparin-like glycosaminoglycans protect human cathepsin B from alkaline pH-induced inactivation. Journal of Biological Chemistry, **2001**, 276(2), 944-951.
- [18]. BOOK, cancer- leading proteases: structure, function, inhibition , P17-18.
- [19]. BENEDITE PORTET, recherche bioguidée de molécules antipludiques d'une plante guyanaise piper hostmannianum var. berbicense, **2007**.
- [20].DRIDI DELEL, synthèse de dérivés coumariniques d'intérêts biologique et antioxydants, **2015**.
- [21].P. SINGH, A. ANAND, V. KUMAR, EUR. J. MED. Chem., **2014**, 85, 758-777,.
- [22].REDOUANE MOKRINI, Mécanismes radicalaires dans la dégradation de composés phénoliques en chimie sous rayonnement : radiolyse gamma des chalcones et de l'acide ferulique en solutions alcooliques, **2006**.
- [23].HORDE P..Chalcone-définition. Journal des femmes santé, **2015**.
- [24].NIELSENS. F., CHRISTENSENS. B., CRUCIANI G, KHARZMI A., LILJEFORS T., Antileish manial chalcones : Statistical design, synthesis, and three- dimensional quantitative structure- activity relation shpanalysis. J Med chem. **1998**, 41, 4819-4832.
- [25].XIAOLIN Z.I., ANNE R., Simoneau. Flavokawian A, a Novel Chalcone from kava extract, Induces apoptosis in bladder cancer cells by involvement of bax protein- dependent and mitochondria-dependent apoptotic pathway and suppresses tumor growth in mice. Cancer Research, **2005**, 65, 3479-3486.
- [26].HALLIWELL, B. Free radicals and antioxidants: a personal view. Nutrition Reviews. **1994**, 52 :253-265.
- [27].H.HARAGUCHI, H.ISHIKAWA, K. MIZUTANI, Y.TAMURA, T.KINOSHITA, Bioorg. Med. Chem., **1998**, 6, 339-347.
- [28].M. UENAK, K. KAWATA, M. NAGAI, T. E. ENDOH, Chem. Abstr ., **2001**.134,29421.
- [29].F.J.SMIT, D.D. N'DA, Bioorg. Med, Chem., **2014**, **22**, 1128-1138.

CHAPTER II

Méthodes de criblage virtuel in silico

II.1.Le criblage

II.1.1.Généralités sur le criblage

La découverte de nouvelles drogues médicamenteuses dépend d'un procédé long et coûteux. Le temps compté entre l'identification d'un nouveau composé potentiellement actif sur une cible à son approbation et sa mise sur le marché varie entre 9 et 12 ans [1], oublié qu'un grand nombre de molécules sont rejetées au fur et à mesure des phases cliniques. Ce procédé peut être scindé en plusieurs étapes, allant de la découverte du (composé médicament) à sa production [2].

II.1.2.Criblage virtuel

Le criblage virtuel est la stratégie *in silico* la plus utilisée pour l'identification de touche ("hits") dans le cadre de la recherche de nouveaux médicaments [3, 4,5]. Celui-ci fait désormais partie intégrant de la plupart des programmes de recherche décomposés bioactifs, qui ceux-ci se déroulent en milieu académique ou industriel, car il constitue un complètement essentiel au criblage biologique haut-débit [6, 7,8].

Le criblage virtuel permet l'exploration de larges chimiotèque ($> 10^6$ molécules) à la recherche de composés actifs vis-à-vis d'une cible thérapeutique donnée. Ce processus vise à réduire significative la chimiotèque 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 ("hit rate"), tandis qu'une sélection aléatoire de molécules de la chimiothèque saurait fournir un tel enrichissement. Ainsi, le temps aussi bien que les couts de l'identification de nouvelles touches peuvent être réduits de façon significative [6].

Le criblage virtuel se décompose en trois étapes :

- la mise en point de la chimiotèque de départ. Celle-ci peut être soit une collection de criblage (molécules disponibles) soit une collection virtuelle (molécule à synthétiser)
- de criblage proprement dit, qui consiste à prédire à la fois la conformation active et l'orientation relative de chacune des molécules de la chimiotèque, par rapport à la cible d'intérêt ;
- la sélection d'une liste de touches virtuelles : après différents traitement permettant d'éliminer les « faux positifs », la sélection finale des molécules à évaluer passe par un examen individuel des interactions 3D de chaque hit virtuel avec le récepteur.

Les méthodes de criblage sont séparées en deux grandes familles : « ligand-based » et « structure-based », en fonction du type de données.

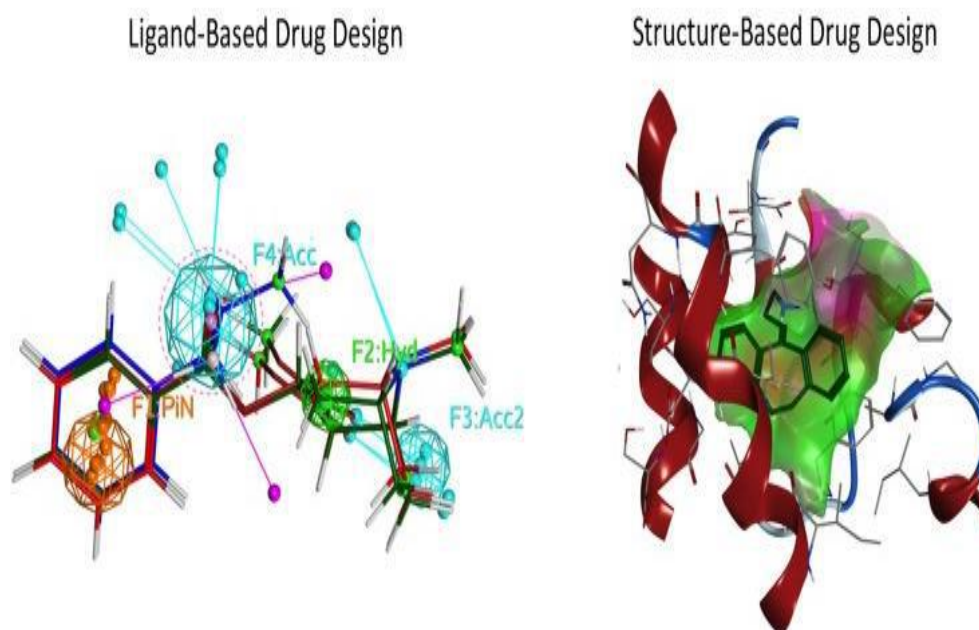


Figure (II.1) : Illustration schématique des approches ligand-based et structure-based.

II.1.2.1. Criblage virtuel « Ligand-based »

Les méthodes de criblage « ligand-based » reposent sur la connaissance préalable de ligands ayant une activité sur la cible thérapeutique. Il sera ainsi possible d'utiliser ces ligands comme une première base de « hits » afin d'identifier d'autres composés similaires, présentant des caractéristiques d'activité comme aux ligands connus de la cible [8,9]. Différents types de descripteurs moléculaires pourront être calculés pour quantifier la similitude entre composés. Suivant le nombre de ligands connus de la cible thérapeutique. Plusieurs méthodes peuvent être employées : la recherche de similarité, le criblage de pharmacophores, ou les approches QSAR. [9]

II.1.2.2. Criblage virtuel « Structure-based »

Le criblage virtuel « structure-based » repose sur la connaissance de la structure de la cible thérapeutique pour l'identification de composés actifs. La structure 3D d'une protéine peut être déterminée principalement par les méthodes de cristallographie à rayon X ou résonance magnétique nucléaire (RMN) [10]. Les structures obtenues par cristallographie à rayons X ont l'avantage d'être d'une bonne résolution (78% de structures ont une résolution inférieure à 2.5 Å) [10]

Cependant, cette technique décrit un état cristallin de la protéine, qui peut parfois différer largement de son état *in vivo* en phase aqueuse. Inversement, la RMN permet d'obtenir un ensemble de conformation illustrant la dynamique et la flexibilité d'une protéine en phases aqueuse.

II.1.3.Filtrage ADMET-TOX

Les échecs dans le développement de médicament étaient principalement la conséquence de mauvaise performance pharmacocinétique. Ce problème a été en grand partie résolu grâce à l'utilisation d'un filtrage procédant l'utilisation de chimiotéque, Résidant les taux d'échec dans les phases de développement.[11]

Le filtre par ADMET-TOX (Absorption. Distribution. Métabolisme. Elimination et toxicité) sont rapidement devenus populaires. Ils reposent sur plusieurs critères déterminant les propriétés pharmacocinétiques potentielles des molécules et sont des désormais largement utilisés pour réduire le nombre de composés d'une chimiotéque en sélectionnant les plus aptes à devenir des candidats médicament, avant tout processus de criblage. L'utilisation de ces filtres a donné de très bons résultats :

II.1.3.1.Absorption

L'absorption est le processus par lequel le médicament inchangé passe de son site d'administration à la circulation générale la voie d'administrations du médicament influence cette première phases : la voie intra-veineuse est la voie de référence puisque par définition, à la différence des autres voies (orale par exemple) toute la dose administrée atteint la circulation général.

II.1.3.2.La distribution

La distribution est fonction de la capacité du médicament à franchir les barrières qui peuvent être franchies par diffusion passive ou par l'action de transporteurs.

II.1.3.3.Métabolisme

Le terme de métabolisme fait référence à la transformation, par une réaction enzymatique d'un médicament en ou plusieurs autres composées actifs ou inactifs au plan pharmacologique. De nombreux tissus peuvent réaliser cette transformation (peau, poumon, rien, intestin...). Néanmoins le principal site de biotransformation est situé au niveau hépatique, dans les enzymes des microsome.

II.1.3.4. Elimination

Elimination se définit comme le volume de plasma contenant la quantité de médicament éliminée par unité de temps. Si, stricto sensu, la clairance d'un médicament peut être le rénale sont, habituellement, les deux principalement à prendre en compte dans l'étude des médicaments.

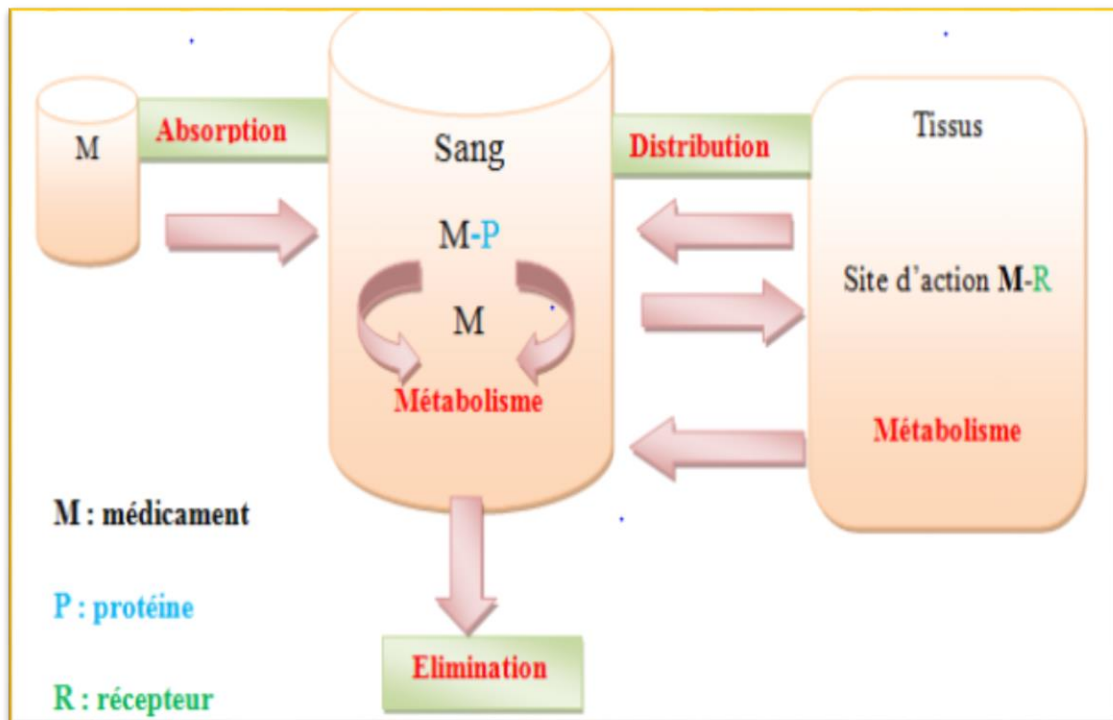


Figure II.2 : Représentation schématique de devenir d'un médicament dans l'organisme.

II.1.3.5. Toxicologie

La toxicité d'un candidat-médicament est l'un des paramètres les plus redoutés par les développeurs, surtout lorsqu'il entre phases cliniques. En effet, un candidat-médicament présentant des effets secondaires graves lors des phases cliniques. Est immédiatement abandonné, et ce généralement de manière définitive. Il est donc très important de tenter de prédire la toxicité d'un composé à partir de sa structure lors des phases précoces de développement R&D. la toxicité peut être classifiée selon l'organe affecté ou selon le mécanisme de toxicité.[12]

II.2. Docking moléculaire

II.2.1. Définition :

Le docking est une technique informatique qui permet de prédire les interactions probables entre des ligands (substrat, activateur ou inhibiteur) et la structure d'une protéine. Il se déroule en deux étapes :

- une étape de positionnement du ligand dans le site choisi de la protéine.
- une étape d'évaluation des interactions énergétique potentielles entre le ligand et la protéine.

Les méthodes utilisées pour ces deux étapes différentes en fonction du programme de docking utilisé [13]. On peut dire que le docking est une méthode qui prédit l'orientation d'une molécule par rapport à une autre pour avoir le complexe le plus stable. Il est fréquemment utilisé sur l'étude de la cible moléculaire des médicaments et réduire les essais expérimentaux. En générale il Ya deux approches, l'un basée sur la complémentarité des surfaces et l'autre basée sur le calcul de l'énergie du complexe.

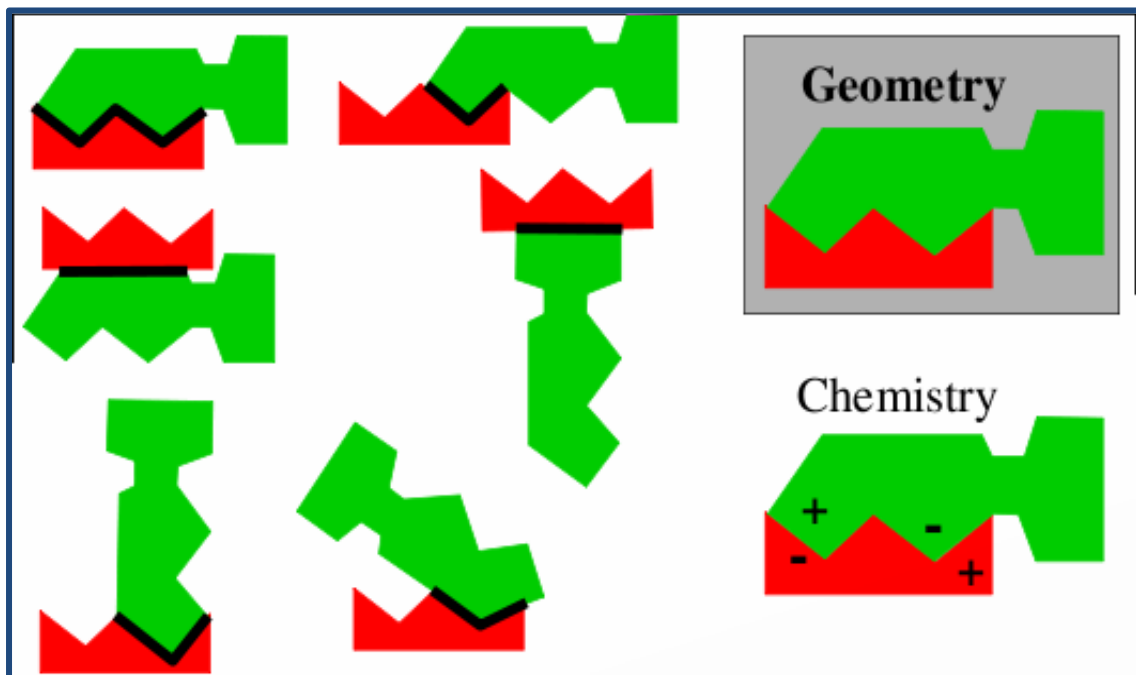


Figure II.3 : Docking moléculaire.

Il est l'ensemble des mécanisme et interactions intervenant lors de la formation de complexes moléculaire. il à des applications pratiques dans l'ingénierie des protéines et la conception de nouveaux médicaments. Les différents types de docking sont : ligand-protéine, protéine-protéine, glucide-protéine et ADN-protéine. [14]

II.2.2.Principe de docking

Le plus important problème pour l'étape de docking est de parcourir le mieux possible l'espace conformationnel. En fonction du nombre de degrés de liberté de translation, de rotation en plus des conformations de départ possibles du ligand. Pour éviter les calculs que les machines ne peuvent résoudre ou seulement dans des temps bien trop importants.[15].

II.2.3.Les outils du docking moléculaire

Les logiciels de docking sont des outils utiles en pharmacie et en médecine, car la plupart des principes actifs sont des petites molécules (ligand) conçues pour interagir sur le plan moléculaire avec leurs protéines cibles (récepteur), pour agir sur les voies biologiques dans lesquelles ces dernières sont impliquées [16]

II.2.4.Docking avec ligand rigide, semi-flexible et flexible

II.2.4.1.avec ligand rigide

Le ligand et la protéine sont tous les deux considérés comme étant rigides. Seuls les 6 degrés de liberté de translation et rotation sont explorés pour le ligand, le subir des translations et rotations dans chacune des trois dimensions, la recherche conformationnelle commence en énumérant toutes les translations et rotations possible avec au minimum 1 Å et 1.5 Å de RMSD, respectivement, entre chaque changement. Chaque pose sans clash est notée à l'aide d'une fonction de score. Les poses ayant les meilleurs scores sont affinées par le programme, qui teste les rotations et translations proches (RMSD supérieure à 0.5 Å et 0.75 Å, respectivement) [17]. D'autres programmes, tels que DOCK [18] font aussi docking rigide.

La précision de ces méthodes peut être améliorée en générant plusieurs conformés pour chaque ligand. Bien que moins précises que les méthodes de docking flexibles, leur rapidité et leur relative efficacité les rendent intéressants pour les protocoles de criblage hiérarchique. Elles peuvent constituer un premier filtre, éliminant de la chimiotèque les molécules trop grandes ou présentant une mauvaise complémentarité avec le site de liaison [19-20]

II.2.4.2.avec ligand semi-flexible

Lorsque l'espace conformationnel des ligands est exploré, le nombre de degrés de liberté de l'espace de recherche peut être conséquent dans le cas de molécule très flexible, d'autres algorithmes dit de fragmentation, sont employés pour construire de façon incrémentielle le ligand au sien du site actif de la protéine, l'espace des conformations du ligand est alors

restreint au voisinage d'un ensemble initial d'états simplifiés. Cette stratégie de recherche par construction, qui se présente sous diverses variantes [21], est notamment adoptée par les programmes DOCK [22-23], FLE_{XX} [24] et Hammerhead [25].

Les programmes de docking semi-flexible considérés comme les plus efficaces emploient des méthodes de recherche aléatoires ou stochastiques.

II.2.4.3.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 :

- La première étape correspond à une exploration de l'espace conformationnel de manière à retrouver, parmi les conformations proposées, les conformations bioactives.
- la deuxième étape, une fonction de score évalue ces conformations [26]. Il existe plusieurs types d'algorithmes de recherche conformationnelle.

II.2.5.Scoring

II.2.5.1.Principe

Le score est une donnée numérique utile pour quantifier le degré avec lequel un ligand se complexifie à un récepteur. C'est globalement une approximation de l'énergie libre résultant du passage de la forme libre de la protéine et du ligand à l'association sous forme de complexe. Le principe thermodynamique suivant : [27].

$$\Delta G = \Delta G_{\text{complexe}} - \Delta G_{\text{ligand}} - \Delta G_{\text{protéine}}$$

L'établissement d'une bonne fonction de score est un important problème du docking. Il arrive souvent que la solution évaluée comme étant la plus probable ne soit pas la forme native attendue. Ceci peut être dû au fait que le complexe natif n'est pas forcément celui qui présente la plus grande surface d'accès, ou encore le plus grand nombre de liaisons hydrogène disponibles [28].

II.2.5.2.Fonction de score

La procédure de docking permet de générer une liste de complexes représentant les modes d'associations favorables entre le ligand et le récepteur macromoléculaire. Ensuite l'évaluation

des complexes, pour trouver celui ou ceux les plus susceptibles de reproduire au mieux le mode d'association réel.

L'association entre protéine et ligands est gouvernée par plusieurs paramètres thermodynamiques :

- Les interactions hydrophobes
- Les interactions électrostatiques
- Les liaisons hydrogène

II.3.3D QSAR

II.3.1. Définition

Les relations quantitatives structure-activité (QSAR) sont des relations mathématiques liant la structure chimique et l'activité pharmacologique de manière quantitative pour une série de composés. Les méthodes qui peuvent être utilisées dans QSAR comprennent diverses techniques de régression et de reconnaissance de formes.[29]

II.3.2. Objectifs du QSAR

La plupart des méthodes QSAR se concentrent sur les objectifs suivants:

- Corréler et récapituler quantitativement les relations entre les tendances des altérations de la structure chimique et les changements respectifs des paramètres biologiques pour comprendre quelles propriétés chimiques sont les déterminants les plus probables de leurs activités biologiques
- Optimiser les leads existants pour améliorer leurs activités biologiques
- Prédire les activités biologiques de composés non testés et parfois encore indisponibles.[29]

II.3.3. Progrès dans les approches 3D-QSAR

3D-QSAR est un terme large englobant toutes ces méthodes QSAR qui corrélient les propriétés macroscopiques des cibles avec des descripteurs calculés basés sur des atomes dérivés de la représentation spatiale (tridimensionnelle) des structures moléculaires. La méthodologie est apparue comme une extension naturelle des approches QSAR classiques

mises au point par Hansch et Free-Wilson. L'inconvénient majeur des techniques 3D-QSAR est qu'elles reposent toutes sur diverses hypothèses qui sont décrites dans la section suivante.[29] .

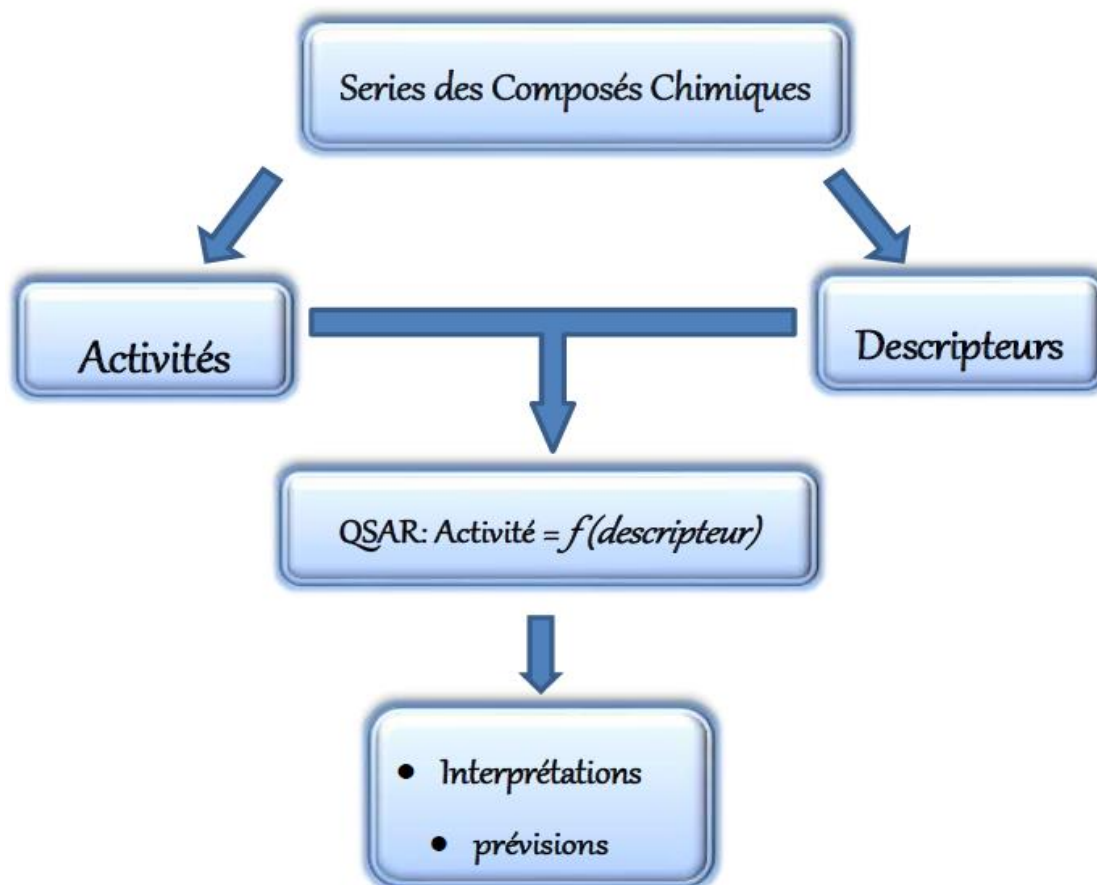


Figure II.4 : Modèle de l'étude de relation structure activité

II.3.4. Études 3D-QSAR

Des modèles 3D-QSAR dit « receptor dependent » (RD-QSAR) peuvent être mis en œuvre lorsque la structure 3D de la cible biologique est résolue [30]. Les approches 3D-QSAR ont été développées pour corréler l'activité biologique d'une série de composés actifs de référence avec l'arrangement spatial de nombreuses propriétés de la molécule telles que les propriétés stériques, lipophiles et électroniques et pour fournir des indications pour l'optimisation par pharmacomodulation et la conception de nouveaux composés avec des profils d'activité améliorés [31].

En 1979, La première approche 3D-QSAR a été proposée en décrivant les propriétés de champs moléculaires de composés, les calculer sur une grille régulière puis les corréler à leur

activité biologique par analyse en «composante principale» (PCA) [32]. Actuellement, différentes méthodes 3D-QSAR sont utilisées tel que CoMFA (Comparative Molecular Field Analysis) et CoMSIA (Comparative Molecular Similarity Indices Analysis). Il est à noter que toutes ces méthodes nécessitent un alignement minutieux des ligands de référence. Cependant, ces approches sont considérées à la fois « ligand-based » et « structure-based ».

Le CoMFA (Comparative Molecular Field Analysis) , une méthode qui reflète l'interaction sans liaison entre le récepteur et le ligand, est largement utilisé dans la conception de médicaments. Les énergies potentielles stérique (Lennard-Jones) et électrostatique (coulombique) des champs de force implémentés dans logiciel SYBYL ont été évaluées par CoMFA. Pour chaque ligand d'alignement de pharmacophore, un réseau cubique 3D avec un espacement de grille de 1,0 Å dans les directions x, y et z a été généré pour enfermer l'agrégat de molécule. Un atome de carbone sp³ avec une charge de +1,0 et un rayon de Van der Waals de 1,52 Å a été utilisé comme sonde ; cet atome a été placé à chaque point du réseau pour calculer divers champs stériques et électrostatiques. Une valeur de coupure d'énergie de 30 kcal / mol a été imposée à tous les calculs de CoMFA pour éviter des valeurs d'énergie excessivement élevées et irréalistes dans la molécule. Ensuite, l'analyse des moindres carrés partiels (PLS) a été appliquée pour obtenir le modèle final.[33]

Le CoMSIA (Comparative Molecular Similarity Indices Analysis) a utilisé une fonction de type gaussien basée sur la distance. Ainsi, CoMSIA peut être capable d'obtenir des modèles plus stables que CoMFA dans les études 3D-QSAR. Le modèle CoMSIA construit a fourni des informations sur les champs stériques, électrostatiques, hydrophobes, donneurs de liaisons hydrogène et accepteurs de liaisons hydrogène.[33]

II.3.5. Modèles statistiques

L'objectif de l'analyse statistique est justement de « démêler » ces descripteurs et d'identifier ceux qui sont corrélés à la variable cible, qui produisent du signal, de ceux qui ne le sont pas, qui produisent du bruit. L'analyse statistique permet également d'identifier les descripteurs qui sont corrélés entre eux pour ne garder que les principaux et réduire ainsi la redondance d'informations [34].

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

la variable cible en fonction de la somme des valeurs pondérées des descripteurs. Les principaux outils statistiques pour obtenir un modèle sont :

- la régression linéaire multi variée (*Multivariate Linear Regression* - MLR).
- la régression en composante principale (*Principale Component Regression* - PCR).
- la régression des moindres carrés partiels (*Partial Least Squares* - PLS) [35].
- les réseaux de neurones artificiels (*Artificial Neural Network* - ANN).

II.3.5. 1. Régression linéaire multiple

Une analyse de régression où la variable dépendante Y dépend linéairement de plusieurs variables indépendantes X_1, X_2, \dots, X_j est appelée régression linéaire multiple. L'équation de régression linéaire multiple est de la forme :

$Y = f(X_1, X_2, \dots, X_j)$ où $f(X_1, X_2, \dots, X_j)$ est une fonction linéaire de X_1, X_2, \dots, X_j [34].

II.3.5. 2. Test de la signification globale de la régression

- **Coefficient de corrélation (R^2)**

C'est le coefficient de corrélation de Bravais Pearson entre Y et Y' , c'est à dire entre valeurs observées et prédites par le modèle de régression, il est noté R , sa valeur varie entre 0 et 1. Une bonne corrélation entre l'activité cible et l'activité initiale si r est plus proche de 1.

- **Test Fisher-Snedecor (F)**

Le test de Fisher-Snedecor est un test statistique qui compare les variances de deux échantillons statistiques.

- **Coefficient de prédiction (Q^2)**

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

Références

- [1]. Warner, T.D. & Mitchell, J.A. Cyclooxygenase-3 (COX-3) : Filling in the gaps toward a COX continuum . Proc. Natl. Acad. Sci. **2002**, 99,13371-13373
- [2]. Overington, J.P., Al-lazikani, B. & Hopkins, A. L. How many drug targets are there ,Nat. Rev. Drug Discov. **2006**, 5,993-996.
- [3]. Waszkowycz, B.; Perkins, T.D.J.; Sykes, R.A.; Li, J.Large-scale virtual screening for discovering leads in the postgenomic era. IBM Systems journal **2001**, 40 (2), 367.
- [4]. Opera, T.I.; Matter, H. Integrating virtual screening in lead discovery. Current Opinion in Chemical Biology **2004**, 8 (4), 349-358.
- [5]. Bajorath, J. intergration of virtual and high-throughput screening. Nat. Rev. Drug Discovery **2002**, 1 (11), 882-894.
- [6]. Mastres, J. virtual screening: a real screening complement to high-throughput screening. Biochem Soc Trans **2002**, 30 (4), 797-9.
- [7]. Stahura, F.L., Bajorath, J. virtual screening methods that complement HTS. Comp chem High Throughput screen **2004**, 7 (4), 259-69.
- [8]. Schulz, K-P. Concepts and application of Molecular Similarity. Berichte der Bunsengesellschaft für Phys. Chemie **1992**, (96), 1087.
- [9]. Hawkins, P.C.D.; Skillman, A.G. & Nicholls, A. Comparison of Shape-Matching and Docking as virtual screening Tools. J Med Chem **2007**, 50.
- [10]. Berman, H. M. et al. The Protein Data Bank. Nucleic Acids Res. **2000**, (28) 235-242.
- [11]. Kola, I. & Landis, J. Can the pharmaceutical industry reduce attrition rates? Nat. Rev Drug discov .**2004** ,43, 711-716 .
- [12]. Valerio, L.G.; Jr. In silico toxicology for the pharmaceutical sciences Toxicol Appl Pharmacol, **2009**, 241(3).
- [13]. Thèse de doctorat de l'université de Paris Sud, Développement d'une nouvelle méthode de docking basée sur les mécanismes enzymatiques et guidée par des groupes prosthétiques, p(31), François Martz,24,11,**2014**.
- [14].Thèse de master de l'université Frère Mentouri, Modélisation par docking moléculaire de l'interaction enzyme-inhibiteur : cas de la tyrosine kinase-JAK2 en tant-que cible thérapeutique impliquée dans les pathologies du sang, p(21,24),Dekkiche Zakia, Benaiche Soulef, 03,07,**2018**.
- [15].Thèse de master de l'université d'EL-Oued, Etude des interactions enzymae-ligand. Cas des inhibiteurs de l'acétylcholinestérase, p(46), Méme Laouar Iman, 2014-2015.
- [16].McGANN,M.FRED pose prediction and virtual screening accuracy. Journal of chemical Information and Modeling, **2011**,51,578-96.

- [17].Kuntz, ID ; BLANEY, JM; OATLEY, SJ ET AL.A geoemtric approach to Macromolecule-ligand interactions. *Journal of Molecular Biology*, 1982, 161,269-88.
- [18].HIRAYAMA, K; AOKI, S, NISHIKAWA, K ET AL.Identification of novel chemical inhibitors for ubiquitin C-terminal hydrolase-L3 by virtual screening. *Bioorganic & Medicinal chemistry*, **2007**, 15, 6810-8.
- [19].HAMZA, A ; ZHAO, X ; TONG, M ET AL.Novel human mPGES-1 inhibitors identified through structure-based virtual screening. *Bioorganic & Medicinal chemistry*, **2007**, 15, 6810-8.
- [20]. Brouijmans, N.; Kuntz, I.D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol.struct.* **2003**,32,335-373.
- [21]. Kuntz, I.D.; Blaney, J.M.; Oatley, S.J.; Langridge, R.; Ferrin, T.E.A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, 161 (2), 296-288.
- [22]. Ewing, T. J.; Makino, S.; Skillman, A.G.; Kuntz, I.D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. -Aided Mol. Des.* **2001**, 15 (5), 441-428.
- [23]. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G.A fast flexible docking method using an incremental construction algorithm. *J.Mol. Biol.***1996**, 3(6),449-462.
- [24]. Welch, W.; Ruppert, J.; Jain, A.N. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* **1996**, 3 (6).
- [25]. Metropolis, N.; Rosenbluth, A.E.; Rosenbluth, M. N.; Taller; A. H.; Taller, E. Equatoin of state calculations by fast computing machines. *Journal of chemical physics* **1953**, 21, 1087-1092.
- [26]. Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Hou, S.; Chong, L.; Lee, T.; Duan, Y.; Wang, W.; Domini, O.; Cieplak, P.; Srinivasan, J.; Case, D.A.; Cheatham, T.E.; Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum, models *Acc. Chem. Res.* **2000**, 33, 889-897.
- [27]. Thèse de doctorat de l'université de Toulouse, Nouvelle approche méthodologique pour la prise en compte de la flexibilité dans les interactions entre molécules biologiques : Les Modes Statiques, P(34), Marie Brut, 05, 03, **2009**.
- [28]. Thèse de master, l'universite de Abou-Bekr Belkaid-Tlemcen, Contribution à la modélisation des biomolécules et leurs interaction ; au cours de l'x fragile cas de la FMRP et du récepteur AMPA, p(26), Melle Lazar Bochra, 02, 07,**2017**.
- [30]. J. Verma, V. Khedkar, and E. Coutinho, "3D-QSAR in Drug Design - A Review," *Curr. Top. Med. Chem.*, vol. 10, no. 1, pp. 95–115, **2010**.
- [31]. T. Scior, J. L. Medina-Franco and Q. T. Do, *Curr Med Chem*, **2009**, 16, 4297-4313.
- [32]. W. Sippl, T. L. Puzyn and K. Cronin, *M.T.D., Editor*, **2010**, 414.
- [33]. R. D. Cramer and M. Milne, *American Chemical Society Meeting*, **1979**, Computer Chemistry Section.

[34]. X. Zhao, M. Chen, B. Huang, H. Ji, and M. Yuan, “Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) Studies on α 1A-Adrenergic Receptor Antagonists Based on Pharmacophore Molecular Alignment,” *Int. J. Mol. Sci.*, vol. 12, no. 10, pp. 7022–7037, **2011**.

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

[36]. P. Khadikar, M. Jaiswal, M. Gupta, D. Mandloi and R. S. Sisodia, *Bioorg. Med. Chem. Lett.*, **15**, 1249 **2005**.

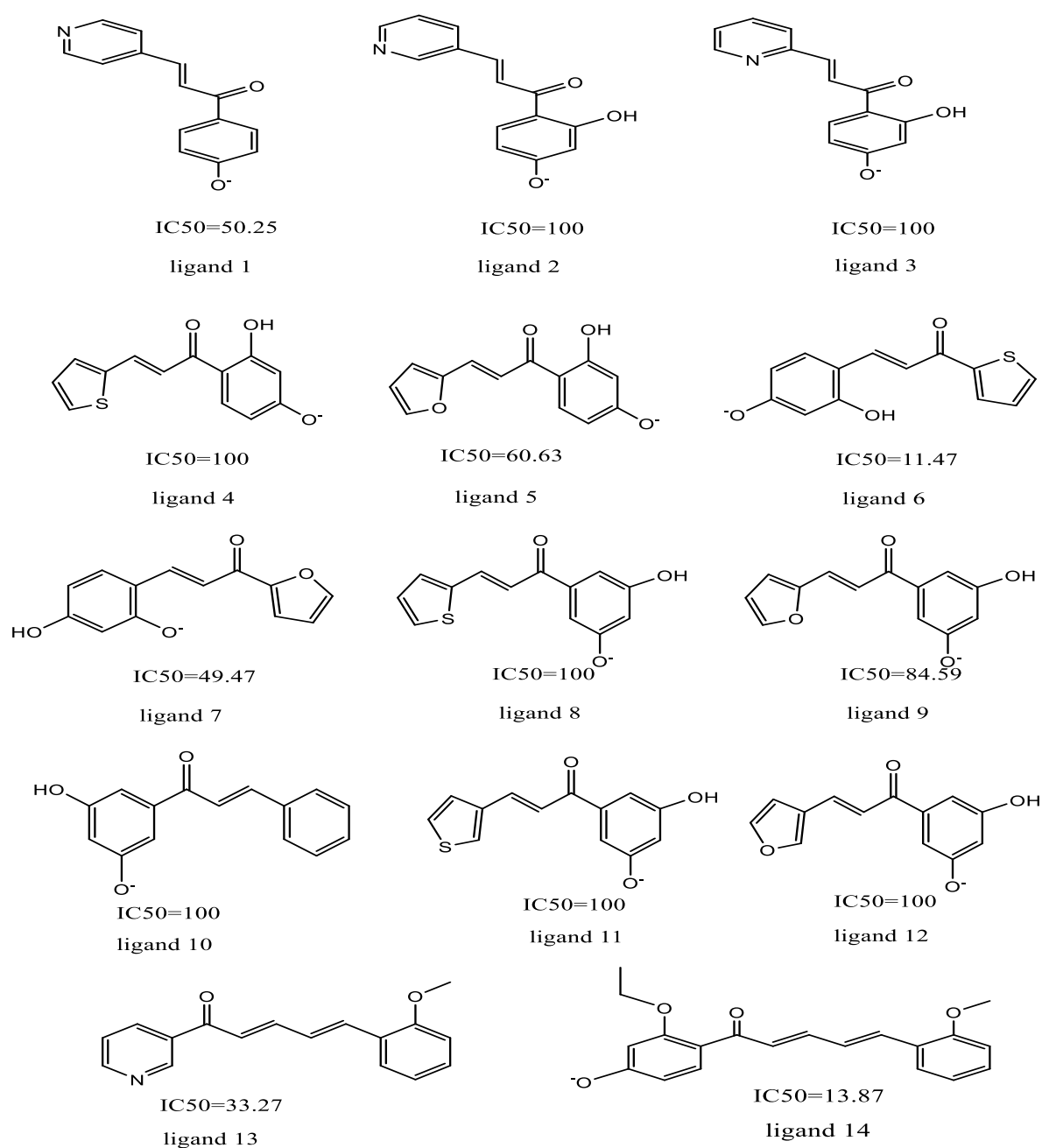
CHAPTER III

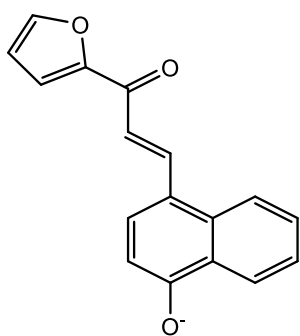
Résultats et discussion

III.1.Introduction

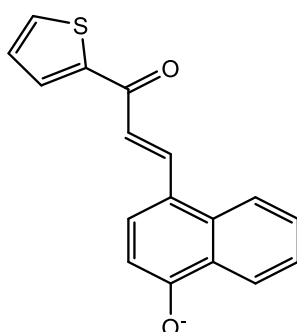
L'interaction entre une protéine et son substrat est la première étape de la plupart des réactions biologiques. Comprendre son mode de fonctionnement et définir quels sont les résidus mis en jeu : l'affinité entre deux molécules, les distances entre les acides aminés du site actifs de l'enzyme et ceux des inhibiteurs et l'énergie d'interaction.

Dans la présente étude, un modèle 3D-QSAR basé sur le ligand et des études docking moléculaire sur des dérivés de chalcone (les molécules ci-dessous) **la figure III.1**, ont été menées pour élucider les caractéristiques structurales des inhibiteurs de la Protéase cathepsine B.

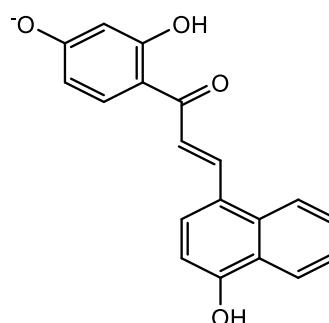




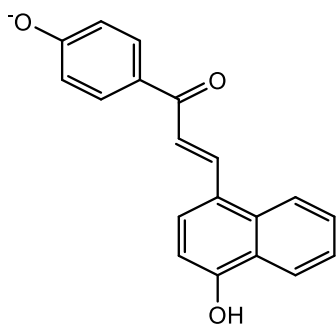
IC₅₀=50
ligand 15



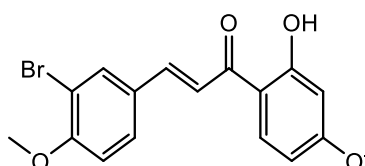
IC₅₀=50
ligand 16



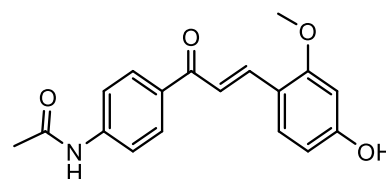
IC₅₀=6.34
ligand 17



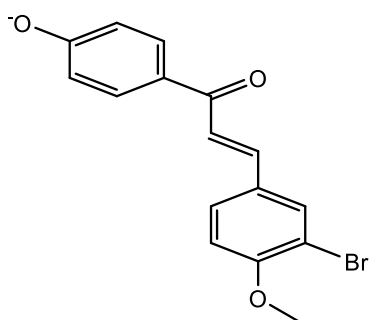
IC₅₀=14.81
ligand 18



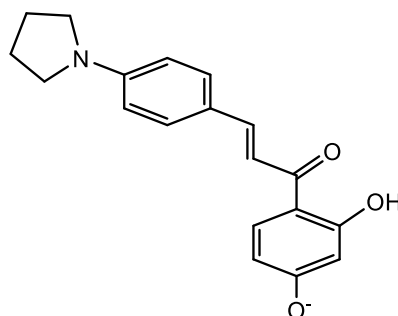
IC₅₀=100
ligand 19



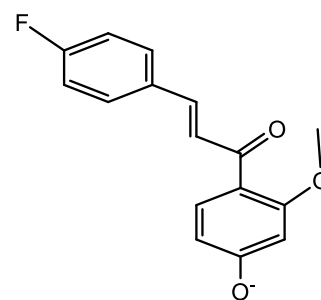
IC₅₀=46.81
ligand 20



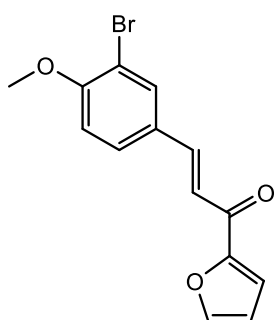
IC₅₀=100
ligand 21



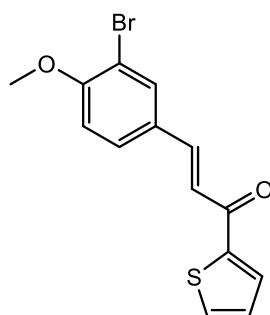
IC₅₀=28.67
ligand 22



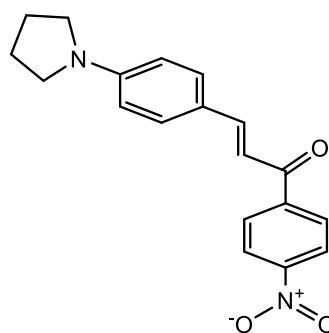
IC₅₀=95.09
ligand 23



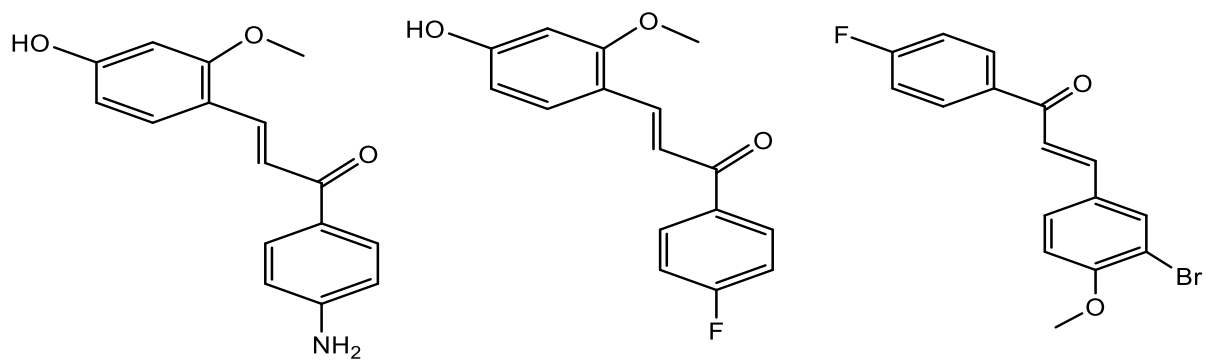
IC₅₀=100
ligand 24



IC₅₀=100
ligand 25



IC₅₀=28.95
ligand 26



IC₅₀=76.47

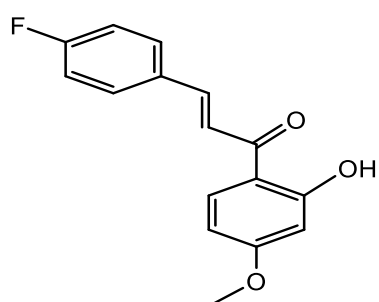
ligand 27

IC₅₀=100

ligand 28

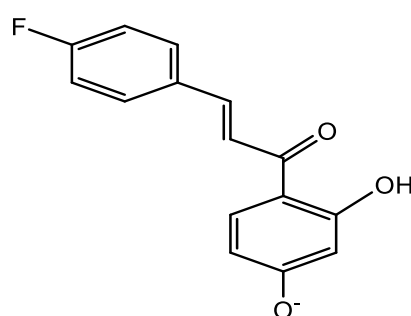
IC₅₀=100

ligand 29



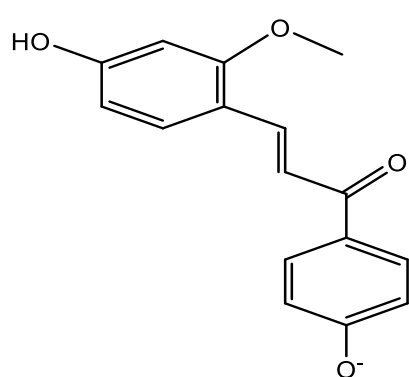
IC₅₀=100

ligand 30



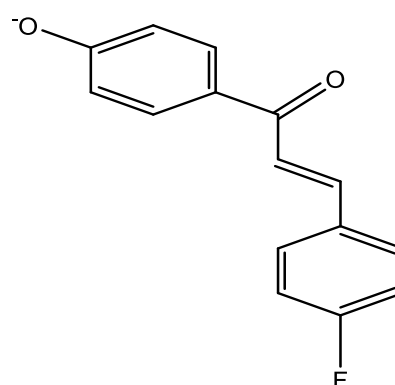
IC₅₀=100

ligand 31



IC₅₀=57.68

ligand 32



IC₅₀=100

ligand 33

Figure III.1: Dérivés de chalcone & activité IC₅₀.

III.2.Méthodologie de calcules

III.2.1.basse de donnée

Un ensemble de 33 composés a été recherché dans la base de données pubchem et la littérature [1, 2] sur les inhibiteurs de la protéase de la cathepsine B. l'ensemble de données a été divisé au hasard en un ensemble d'apprentissage et un ensemble de test, en considérant 75% des totales molécules dans l'ensemble d'apprentissage (25 composés) et 25% dans l'ensemble de test (8 composés). L'ensemble de d'apprentissage a été utilisé pour générer des modèles 3D-QSR, tandis que l'ensemble de test a été utilisé pour valider la qualité du modèle. Toutes les activités biologiques utilisées dans cette étude ont été exprimées par **PIC50 = -log10IC50** (voir le tableau (III.1)).

III.2.2.Préparation de protéine

Structure cristalline tridimensionnelle de la cathepsine B (PDB ID : 3ai8) a été téléchargée dans PDB format de la banque de données sur les protéines [3]. Après cela, la structure a été préparée et affinée en utilisant l'assistant de préparation de protéines de Schrödinger-Maestro v11.8. Des charges et des ordres de liaison ont été attribués, des hydrogènes ont été ajoutés aux atomes lourds et toutes les eaux ont été supprimées. La minimisation de l'énergie a été réalisée à l'aide du champ de force OPLS 2005 en réglant le RMSD (root-mean-square-déviation) de l'atome lourd à 0,30 Å.

III.2.3.Génération de la grille du récepteur

Le module Glide [4] du logiciel Schrödinger a été utilisé pour générer des grilles de récepteurs. Récepteur les grilles ont été déterminées pour les protéines préparées de telle sorte que diverses poses de ligand se lient dans le site actif prévu pendant l'amarrage. Les grilles ont été produites en conservant les paramètres par défaut de facteur d'échelle de van der Waals 1,00 et coupure de charge 0,25 sous le champ de force OPLS 2005. La boîte cubique de dimensions spécifiques focalisée autour du centroïde des résidus du site actif (site actif du ligand de référence, chalcone) a été créée pour le récepteur. Le cadre de sélection a été défini sur 10 Å × 10 Å × 10 Å pour les expériences d'amarrage.

III.2.4.Préparation de ligand

Toutes les molécules ont été construites dans Maestro version 11.8 [5] et préparées à l'aide de LigPrep (module implémenté dans le logiciel Schrödinger) pour convertir la structure bidimensionnelle en 3D, générer un stéréoisomère, déterminer l'état d'ionisation le plus probable en utilisant les conditions par défaut à pH $7,0 \pm 2,0$. L'utilisation le sous-programme epike pour neutraliser les structures chargées, ajouter de l'hydrogène et générer les conformères bioactifs les plus minimisés de l'énergie en appliquant le champ de force OPLS (potentiels optimisés pour les simulations de liquides) 2005.

III. 2. 5. Amarrage (docking)

Quand la grille de récepteur a été produite ; les ligands ont été amarrés à la protéine en utilisant le protocole d'amarrage de Glide. Le ligand flexible et le mode XP (Précision Extra) ont été utilisés dans la procédure d'amarrage. Les conformères amarrés ont été évalués à l'aide de docking Score [6]. Le protocole « Glide XP » déduit des termes énergétiques tels que les interactions : liaisons d'hydrogènes, électrostatiques, hydrophobe et pi – pi (π - π).

III. 3. Résultats et discussion

III.3.Model 3D-QSAR Field based

Field-Based QSAR (dans le logiciel Schrodinger) est un outil pour construire un modèle pour la relation entre les valeurs d'activité connues et les caractéristiques 3D d'un ensemble de composés alignés (similaire à l'analyse comparative de champ moléculaire CoMFA / analyse comparative de l'indice de similarité moléculaire CoMSIA). Il pourrait être possible de visualiser le modèle QSAR dans l'espace pour évaluer qualitativement, ce qui nous permet d'ajouter ou de supprimer des groupes fonctionnels et d'utiliser le modèle pour prédire quantitativement les activités d'autres molécules. Le modèle 3D-QSAR a été généré en PHASE en d'utilisant 25 composés comme d'apprentissage (training), tandis que le modèle a été validé en utilisant 8 composés en tant que composés de test avec le facteur PLS. Un modèle à quatre facteurs PLS avec de bonnes statistiques et une capacité prédéterminée a été généré pour l'ensemble de données. Les valeurs statistiques ont été répertoriées dans le **tableau III.2**. Le résultat QSAR montre le meilleur coefficient de régression de R^2 (0.876), une grande valeur de F (36.5) et une valeur de P (8.09e-009), une petite valeur de l'écart , type RMSE (0.16) et des valeurs élevées de pearson-P (0.796) indiquent un modèle de

régression statistiquement significatif. Le modèle a été validé par un coefficient de corrélation à validation croisée Q^2 (0.507). Le graphique de l'activité biologique prévue des ensembles d'apprentissage et de test est illustré à la **figure (III. 2)**.

Tableau III.1. Données d'activité prédite pour l'apprentissage et le test de l'ensemble de dérivés de chalcone.

| Numéros de ligands | QSAR set | Activité Exp | PLS facteurs | Activité prédite |
|--------------------|-----------------|--------------|--------------|------------------|
| 1 | Training | 4.298 | 4 | 4.275 |
| 2 | Training | 4.00 | 4 | 4.11 |
| 3 | Test | 4.00 | 4 | 4.21 |
| 4 | Training | 4.00 | 4 | 4.23 |
| 5 | Training | 4.217 | 4 | 4.133 |
| 6 | Training | 4.940 | 4 | 4.893 |
| 7 | Training | 4.305 | 4 | 4.357 |
| 8 | Training | 4 | 4 | 3.873 |
| 9 | Test | 4.072 | 4 | 4.264 |
| 10 | Training | 4 | 4 | 4.047 |
| 11 | Training | 4 | 4 | 3.993 |
| 12 | Test | 4 | 4 | 4.035 |
| 13 | Training | 4.477 | 4 | 4.470 |
| 14 | Training | 4.857 | 4 | 5.010 |
| 15 | Training | 4.301 | 4 | 4.302 |
| 16 | Test | 4.301 | 4 | 4.247 |
| 17 | Training | 5.197 | 4 | 4.826 |
| 18 | Training | 4.829 | 4 | 4.747 |
| 19 | Training | 4.000 | 4 | 3.999 |
| 20 | Training | 4.329 | 4 | 4.363 |

| | | | | |
|----|----------|-------|---|-------|
| 21 | Training | 4.000 | 4 | 4.986 |
| 22 | Test | 4.542 | 4 | 4.605 |
| 23 | Training | 4.000 | 4 | 3.977 |
| 24 | Training | 4.000 | 4 | 3.829 |
| 25 | Test | 4.000 | 4 | 3.972 |
| 26 | Test | 4.538 | 4 | 4.436 |
| 27 | Training | 4.116 | 4 | 4.128 |
| 28 | Training | 4.021 | 4 | 4.120 |
| 29 | Training | 4.000 | 4 | 3.992 |
| 30 | Training | 4.000 | 4 | 3.933 |
| 31 | Training | 4.000 | 4 | 4.050 |
| 32 | Training | 4.238 | 4 | 4.469 |
| 33 | Test | 4.000 | 4 | 4.317 |

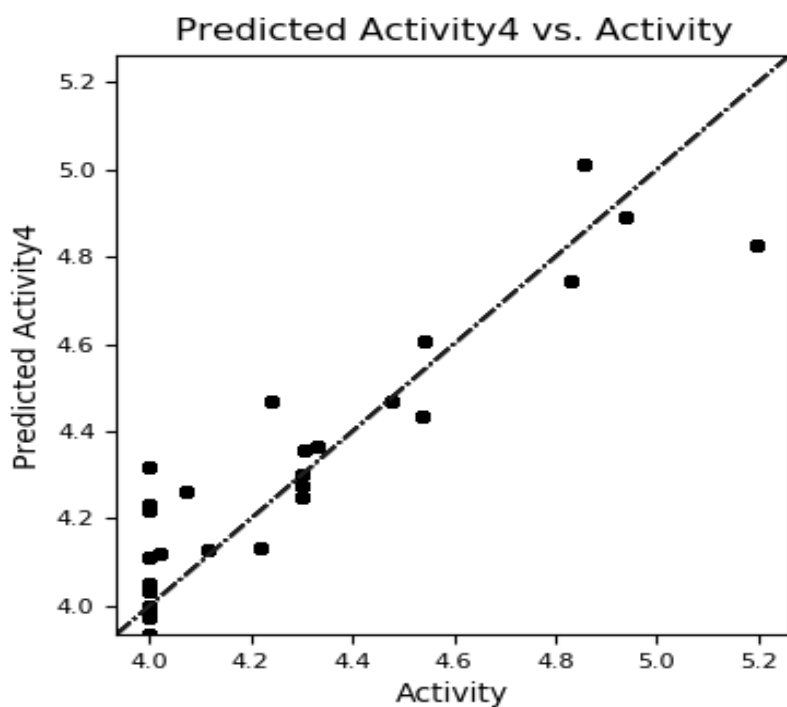


Figure III.2: Graphique de corrélation ente PIC50 expérimental et prédictif des ensembles d'entraînement et de test.

Les contributions relatives des champs stérique, électrostatique, hydrophobe, accepteur de liaison hydrogène et donneur de liaison hydrogène sont respectivement de 0.345, 0.169, 0.177, 0.214 et 0.094 ((voir le **tableau (III.3)** pour le meilleur PLS). La validité effective du modèle est indiquée par sa capacité à prédire l'activité biologique de nouvelles molécules. Une analyse approfondie des différents tests de validité indique que le modèle crée par l'analyse QSAR basée sur le champ est très bon. Les activités expérimentales et prévues pour l'ensemble d'apprentissage et l'ensemble de test sont données dans le **tableau (III.2)**. D'après le **tableau (III.3)**, il apparaît que le champ stérique contribue de façon remarquable sur l'activité par rapport aux champs accepteur de liaison d'hydrogène, hydrophobe, électrostatique et donneur de liaison d'hydrogène, respectivement.

Tableau (III.2) : Valeurs statistiques pour le modèle 3D-QSAR généré par PLS.

| # Factors | SD | R ² | F | P | RMSE | Q2 | Pearson- r |
|--------------|-------|----------------|------|-----------|------|-------|---------------|
| 1 | 0.274 | 0.414 | 16.3 | 0.00052 | 0.20 | 0.245 | 0.618 |
| 2 | 0.238 | 0.578 | 15.1 | 7.38e-005 | 0.23 | 0.008 | 0.780 |
| 3 | 0.174 | 0.785 | 25.6 | 3.28e-007 | 0.18 | 0.383 | 0.792 |
| 4 | 0.135 | 0.876 | 36.5 | 8.09e-009 | 0.16 | 0.507 | 0.796 |

Tableau (III.3) : Détails des calculs du modèle Field-based QSAR

| #Factors | Guassian Steric | Guassian Electostatic | Guassian Hydrophobic | Hbond Acceptor | Hbond Donor |
|----------|--------------------|--------------------------|-------------------------|-------------------|----------------|
| 1 | 0.376 | 0.190 | 0.156 | 0.201 | 0.074 |
| 2 | 0.364 | 0.199 | 0.163 | 0.195 | 0.077 |
| 3 | 0.332 | 0.182 | 0.185 | 0.208 | 0.090 |
| 4 | 0.345 | 0.169 | 0.177 | 0.214 | 0.094 |

Les données d'interaction ainsi obtenues permettent la construction d'un modèle de régression PLS comprenant un grand nombre de coefficients. Afin de faciliter la compréhension de ce modèle, les résultats sont généralement représentés sous la forme d'aires de contour indiquant les régions favorables et défavorables aux interactions stériques, électrostatiques, hydrophobique, accepteur de liaison d'hydrogène et donneur de liaison d'hydrogène.

Les contours 3QSAR basés sur le champ gaussien pour les interactions stérique, hydrophobique, accepteur de liaison d'hydrogène et électrostatiques sont représentés sur les figures III.3a, III.3b, III.3c et III.3d respectivement. Dans les aires de contour, chaque contour coloré correspond à des propriétés spécifiques telles que les contours verts pour les régions à haute influence stérique, jaune pour une faible influence stérique. Les contours bleus représentent les régions électrostatiques favorisées, et les contours rouges pour les régions électrostatiques défavorisées. Les contours orange représentent les régions hydrophobiques favorisées et les contours bleus pour les régions hydrophobiques défavorisées. Les contours rouge et violet indiquent des régions favorables et défavorables pour l'accepteur de liaison d'hydrogène, respectivement.

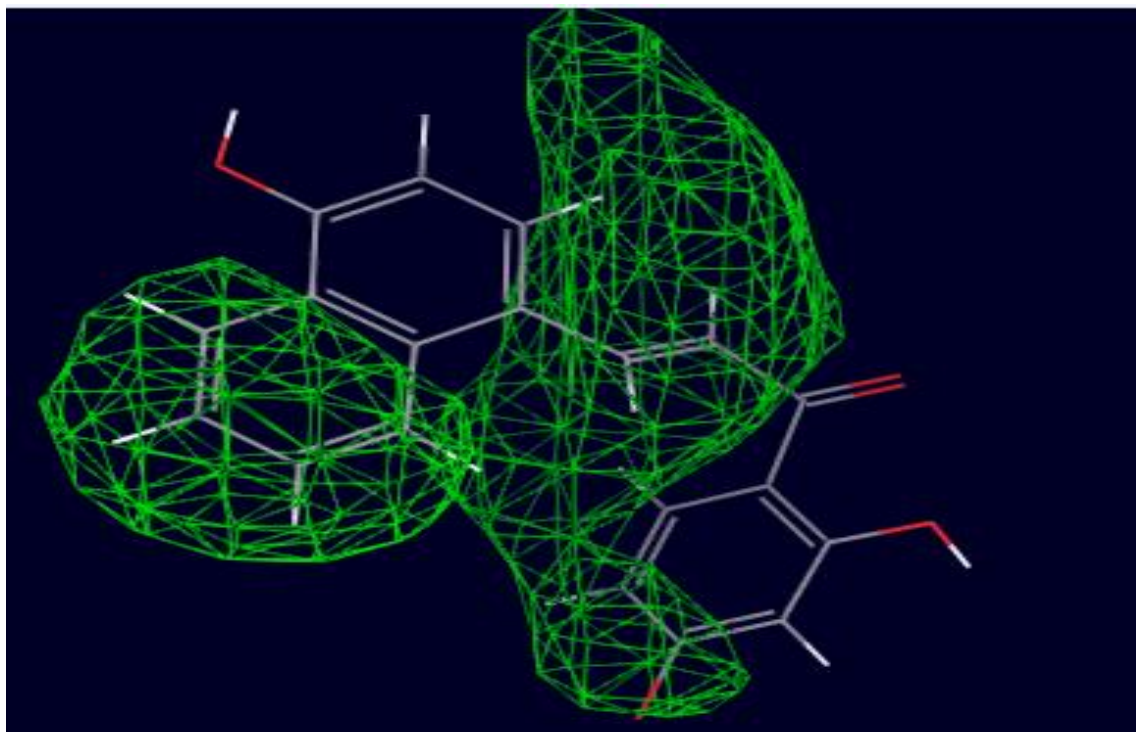


Figure III.3a: Aires du champ gaussien stérique pour ligand 17 (ligand de référence).

Le QSAR basé sur le champ sur la **figure (III.3a)** ne montre aucune présence de contours jaunes sur la molécule indiquant l'absence de régions défavorables. Les grands contours vert situés sur un cycle de naphthalène, la double liaison de la chaîne linéaire et l'atome d'oxygène chargé négativement indiquent une influence stérique élevée dans ces zones (la zone où le groupe volumineux favorise l'activité).

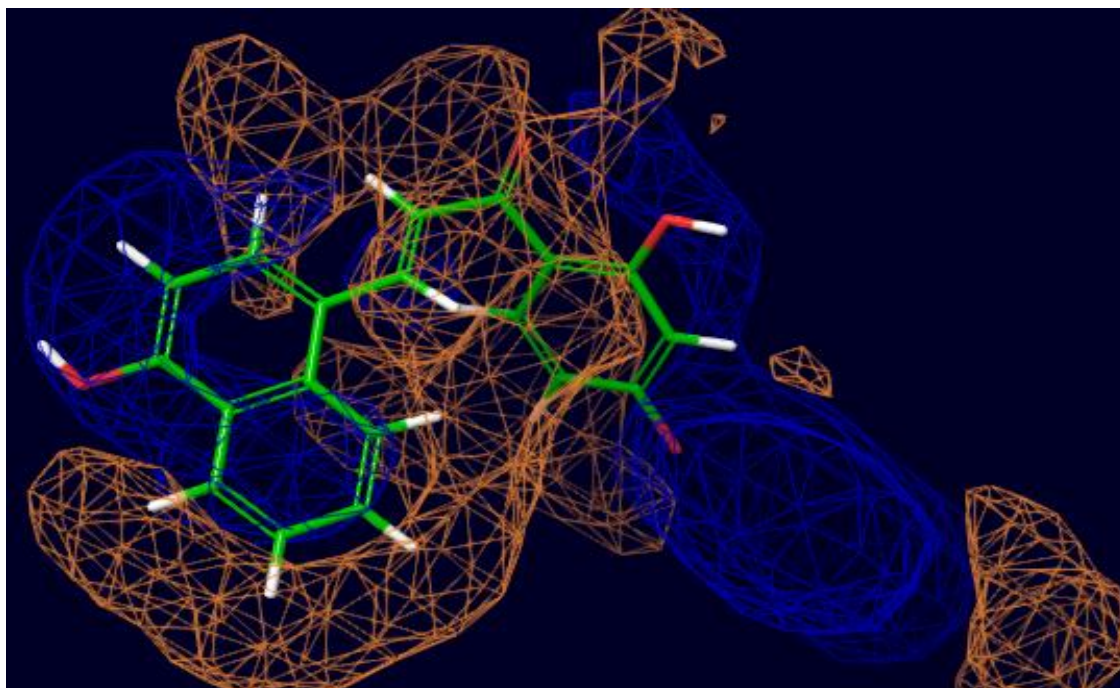


Figure III.3c: Aires du champ gaussien d'hydrophobe pour ligand 17.

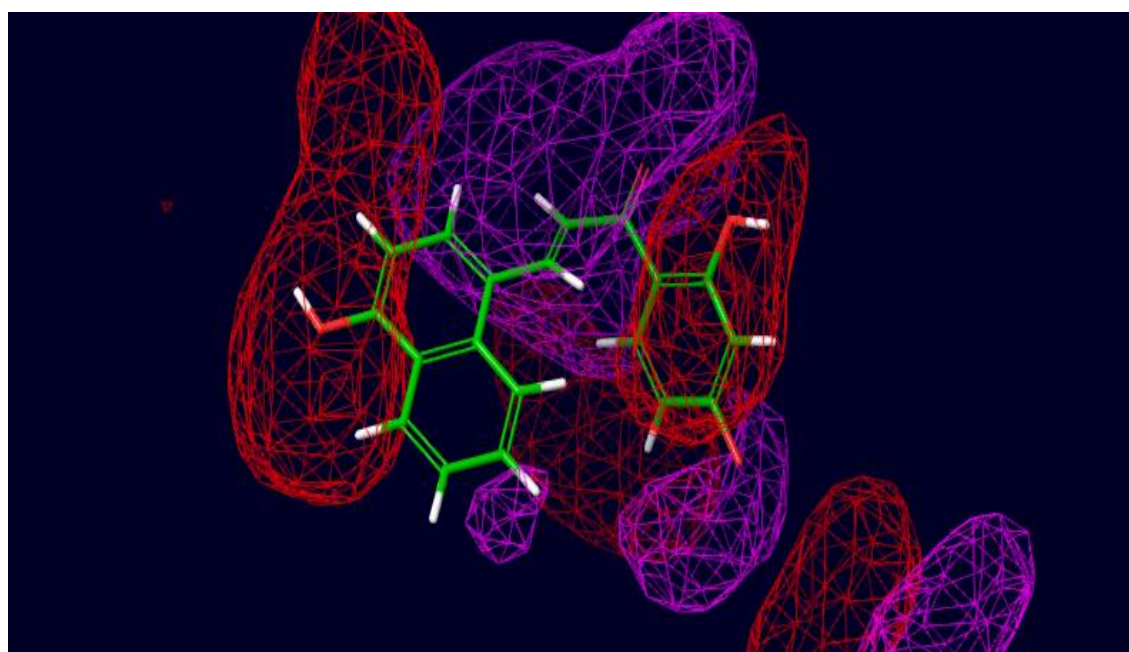


Figure III.3d: Aires du champ gaussien d'accepteur de liaison d'hydrogène pour ligand 17.

Les aires de contours hydrophobes sont affichées sur **la figure III.3b**. Les contours d'orange indiquent des régions hydrophobes favorables situées sur : une partie dans le cycle benzène (les positions non substituées), la zone de la fonction cétone, la double liaison de la chaîne linéaire et les positions non substituées du cycle qui ne contient pas des groupe hydroxyle de naphtalène. Les contours bleus des régions hydrophobiques défavorables reliées à la position qui contient le groupe OH et le cycle non substituées de naphtalène.

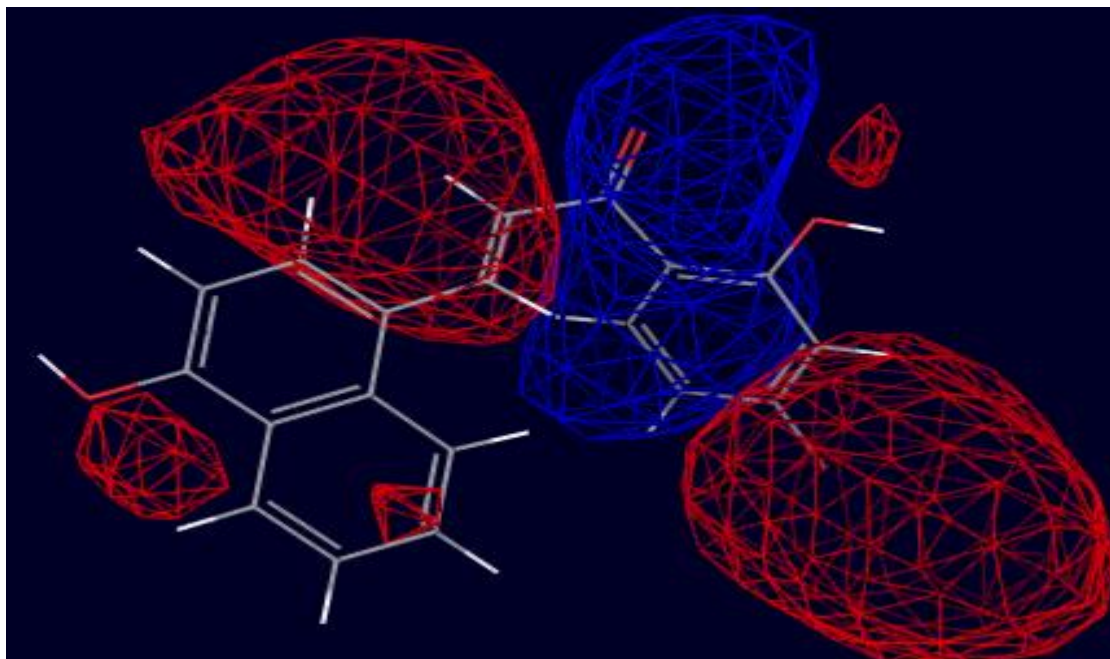


Figure III.3.d: Aires du champ gaussien électrostatique pour le ligand 17.

Les aires de contours d'accepteur de liaison d'hydrogène sont indiquées par des contours rouges et violet sur **la figure III.3c**. Les contours rouges représentent des régions où les accepteurs de liaisons hydrogène sont favorables et les contours magenta indiquent des régions où les accepteurs de liaisons hydrogène sont défavorables pour l'activité. Dans cette position (rouge), tout substituant contenant un groupe accepteur augmente l'activité. La région où liaison hydrogène accepteur est favorable dépend au cycle benzène et l'entourage de la groupe hydroxyle (OH) du cycle de naphtalène et les régions défavorables (violet) situées sur la zone de la double liaison de la chaîne linéaire.

Les aires de contours électrostatiques sont indiquées par des contours rouges et bleu sur **la figure III.3d**. Les contours rouges représentent des régions favorables situées sur la double

liaison de la chaîne linéaire et le groupe OH de naphthalène. Les contours bleus indiquent des régions défavorables pour l'activité.

II.2.5. Docking

Quand la grille du récepteur a été produite ; les ligands ont été ancrés à la protéine en utilisant le protocole d'amarrage (docking) Glide. Le mode flexible et le mode XP (Extra Précision), ont été utilisés dans la procédure de l'amarrage. Les conformères ancrés ont été évalués en utilisant un score de docking. Le «Glide XP» protocole déduit des termes énergétiques tels que les interactions des liaisons d'hydrogène, l'interaction électrostatique, hydrophobe et l'interaction π – π et le reste des paramètres ont été conservés par défaut pour le score [6].

L'amarrage a été effectué pour étudier le mode de liaison des composés actifs avec le récepteur : 3ai8 et obtenir des informations pour de nouvelles améliorations structurelles.

Tableau III.4 : Résultats de l'analyse d'amarrage des ligands sélectionnés.

| Ligands | Score Docking | Energie Glide | Activité PIC ₅₀ |
|-------------|---------------|---------------|----------------------------|
| L-6 | -2.710 | -25.68 | 7.301 |
| L-22 | -4.175 | -39.20 | 6.207 |
| L-23 | -4.015 | -40.73 | 6.346 |
| L-30 | -3.165 | -33.31 | 6.309 |

Le score d'amarrage obtenu pour les inhibiteurs potentiels L-6, L-6, L-14, L-17 et L-18 sont -2.71, -4.175, -4.015 et -3.165, respectivement (Voir le tableau III.4). D'après le tableau III.4, le complexe L-14/ cathepcine B est plus stable que les complexes L-17/ cathepcine B, L-18/ cathepcine B et L-6/ cathepcine B, respectivement. Parmi ces inhibiteurs étudiés, l'inhibiteur L-14 a un score d'amarrage plus élevé par rapport à tous les inhibiteurs actifs sélectionnés.

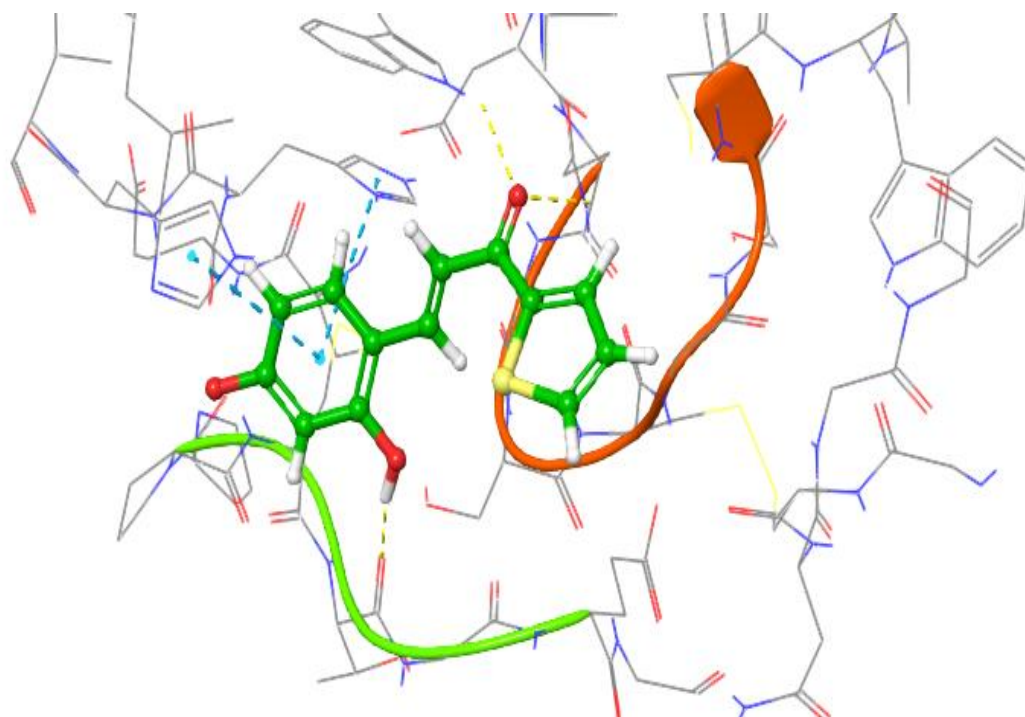
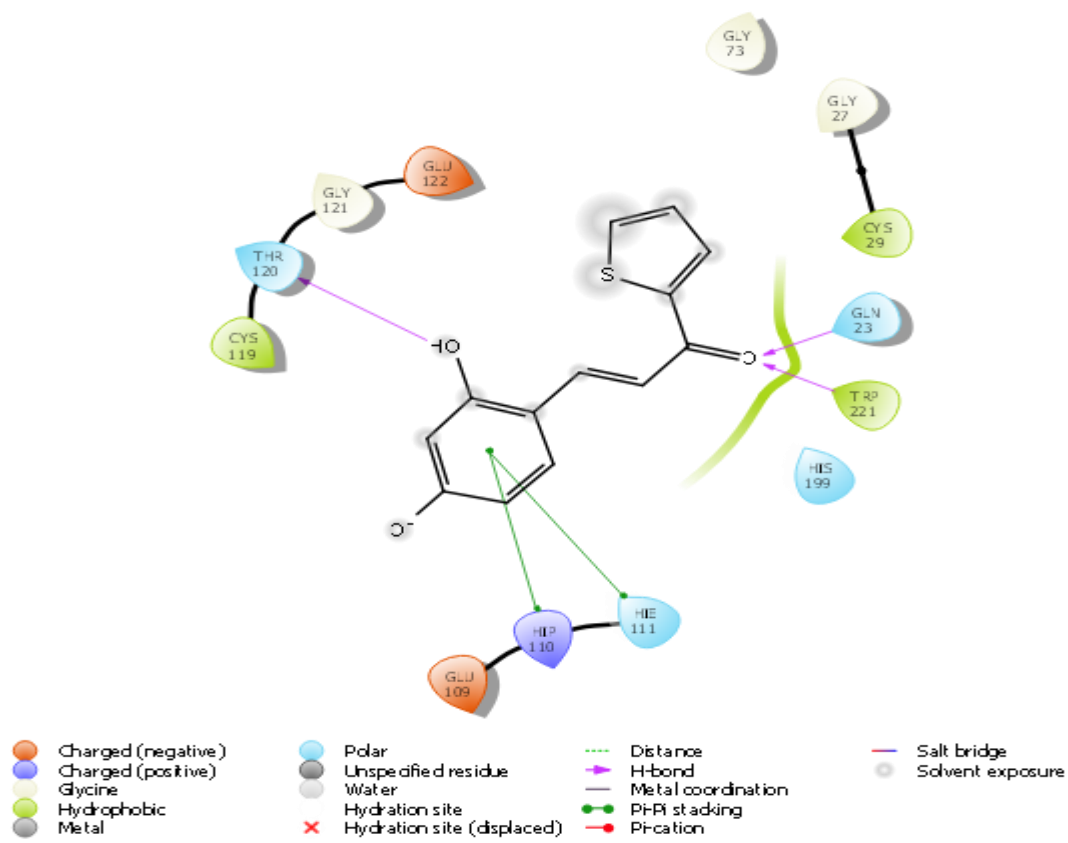


Figure III.4a : Interactions 2D et 3D du ligand 6 dans le site actif du récepteur.

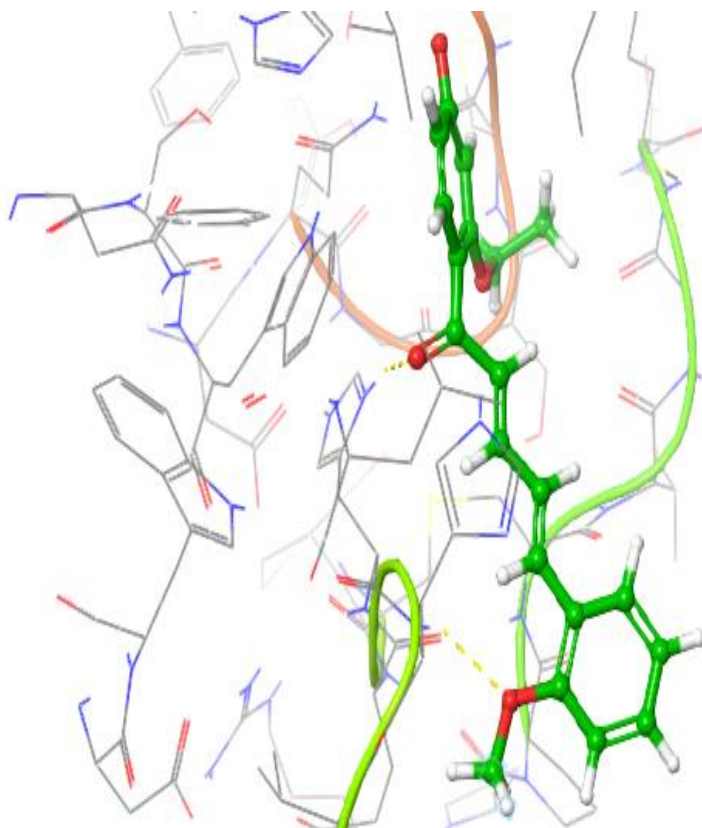
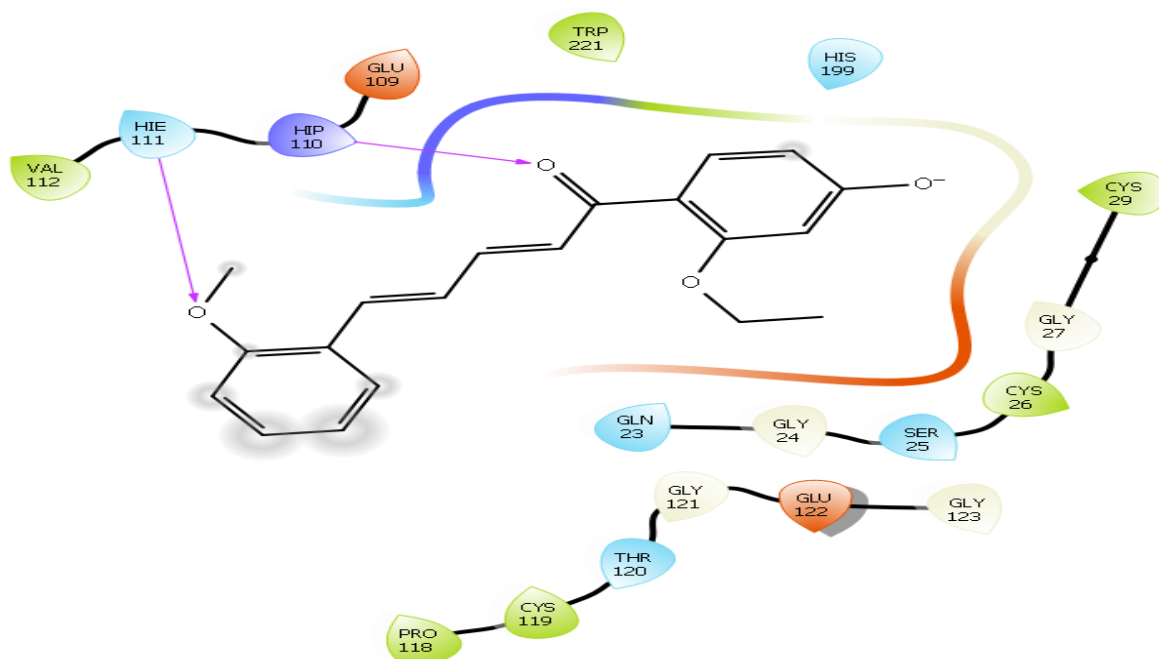


Figure III.4b : Interactions 2D et 3D du ligand 14 dans le site actif du récepteur.

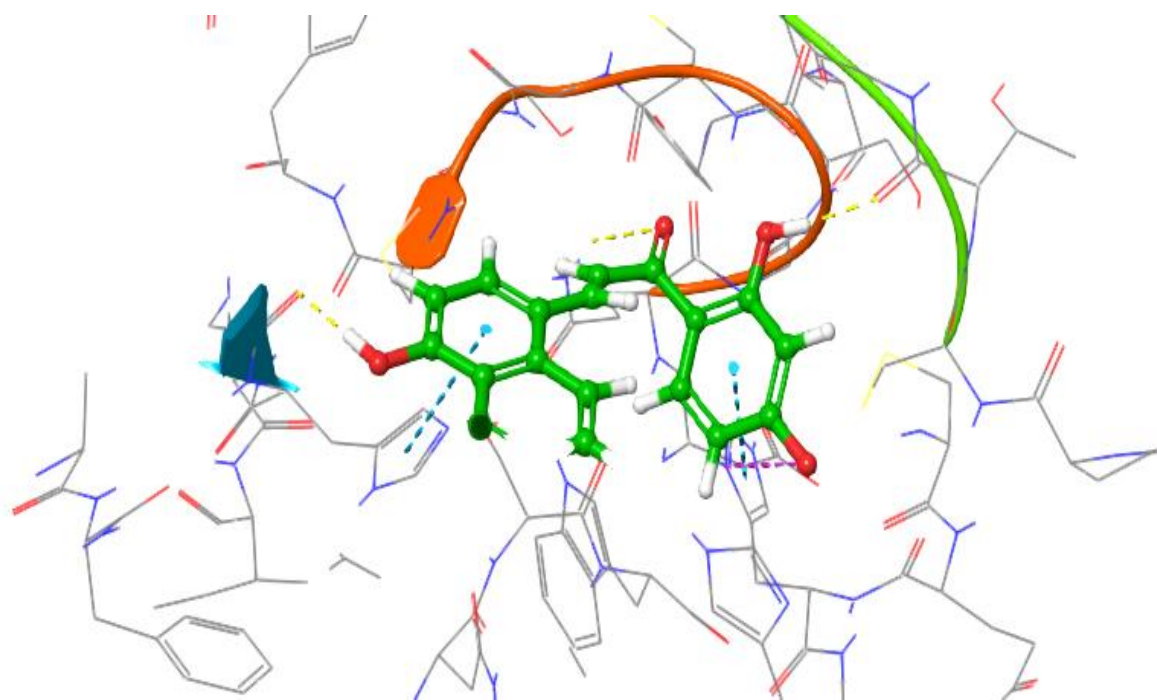
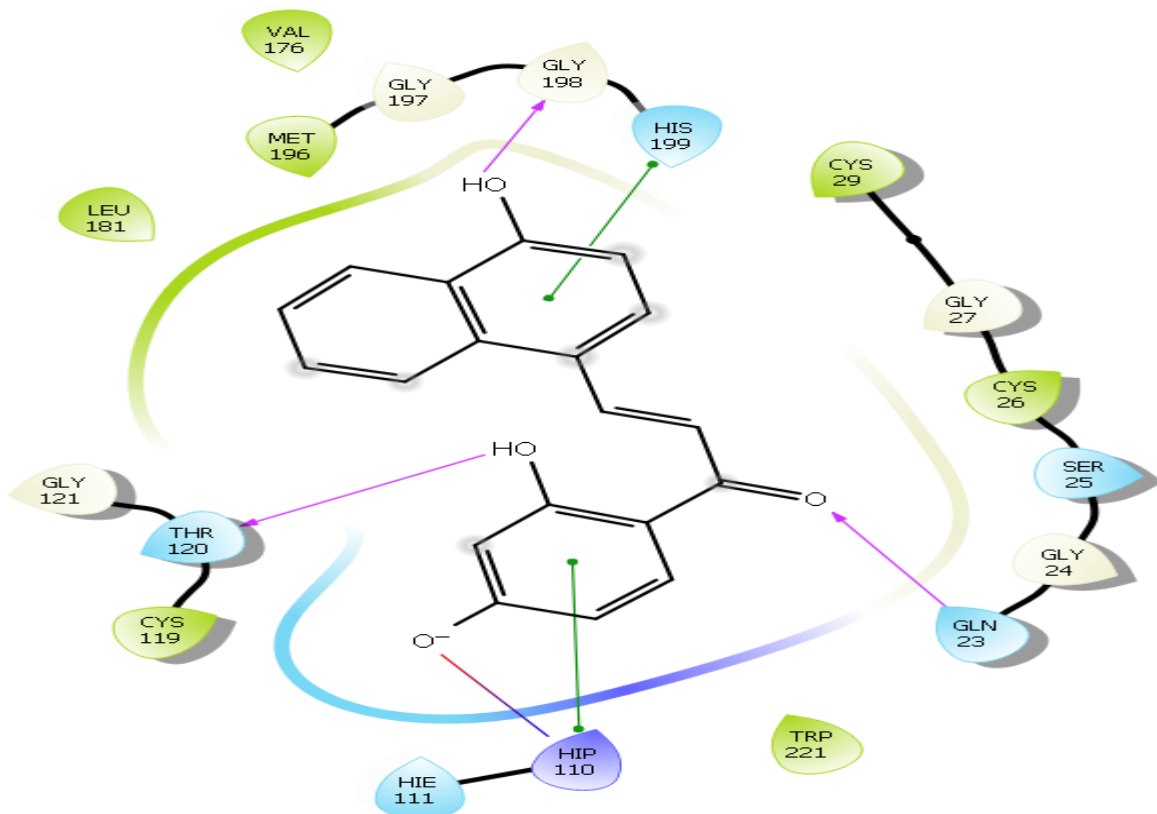


Figure III.4c : Interactions 2D et 3D du ligand 17 dans le site actif du récepteur.

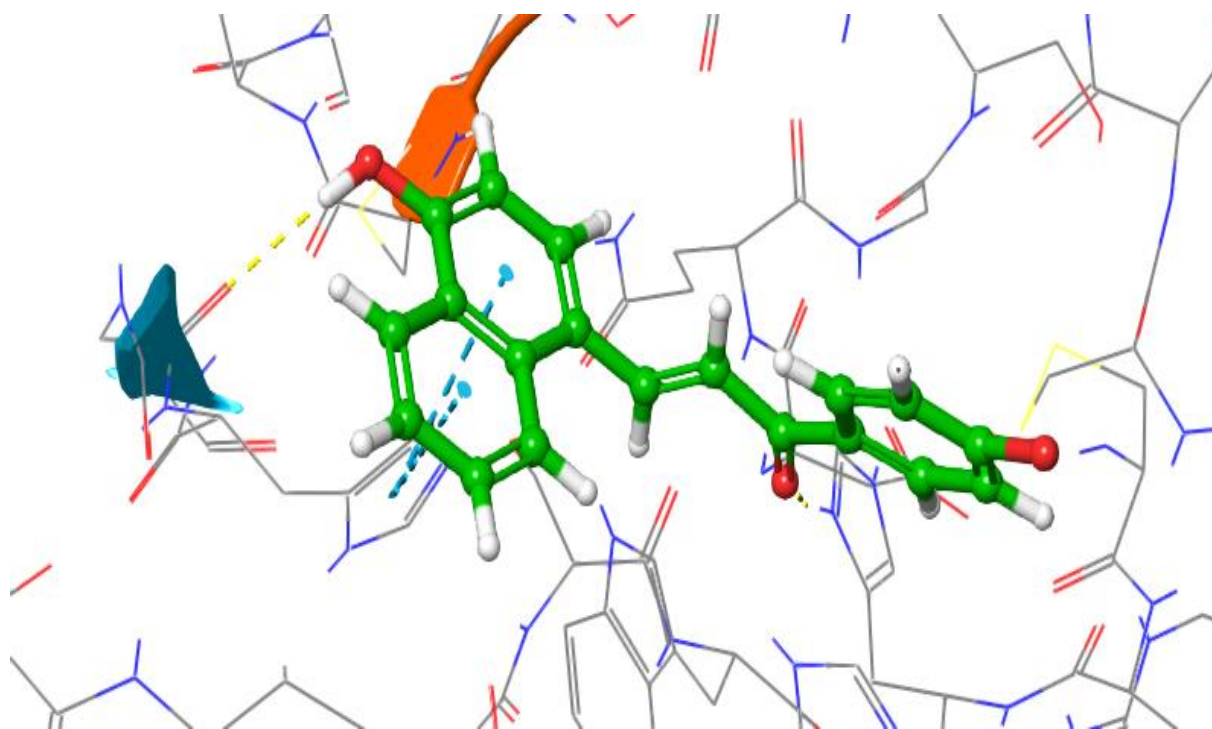
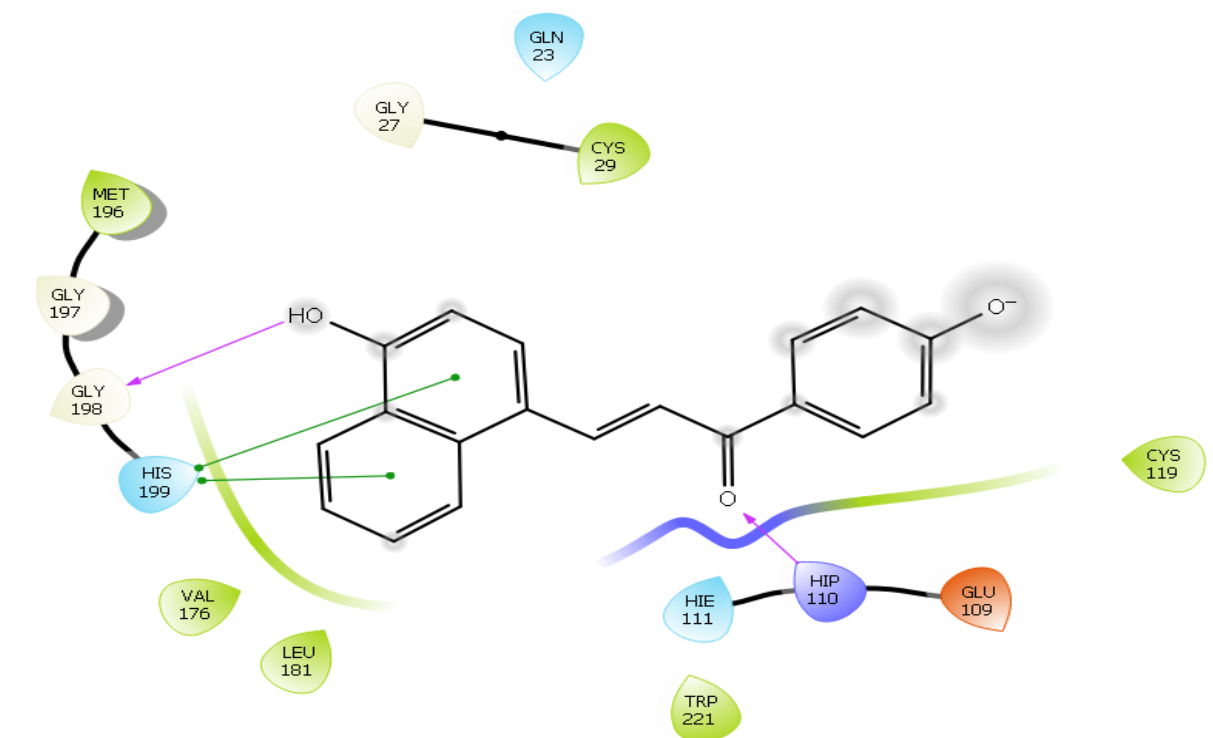


Figure III.4d : Interactions 2D et 3D du ligand 17 dans le site actif du récepteur.

L'étude d'amarrage du **ligand 6** (Figure III.4a) montre les interactions suivantes :

- l'atome d'oxygène du groupe (OH) dans le cycle benzénique agit comme donneur et crée une liaison hydrogène avec **THR :120**.
- La fonction cétone agit comme accepteur et effectue une liaison d'hydrogène avec des acides aminés **TRP: 221** et **GLN: 23**, respectivement.
- Le cycle du benzène effectue des interactions $\pi - \pi$ (π -stacking) avec les résidus **HIE :111** et **HIP :110** , respectivement.

Pour le **ligand 14**, l'atome d'oxygène du groupe (OCH₃) agit comme un accepteur de liaison d'hydrogène et crée une liaison hydrogène avec **HIE : 111** et l'atome d'oxygène de la fonction cétone agit comme un accepteur de liaison d'hydrogène et crée une liaison hydrogène avec le résidu **HIP : 110** (Figure III.4b).

L'étude d'amarrage du **ligand 17** (Figure III.4c) montre les interactions suivantes :

- L'atome d'oxygène du groupe (OH) de naphthalène agit comme un donneur de liaison d'hydrogène et crée une liaison hydrogène avec **GLY : 198**.
- L'atome d'oxygène du groupe (OH) du cycle de benzène agit comme un donneur et effectue une liaison hydrogène avec **THR : 120**.
- L'un du cycle de naphthalène et le cycle de benzène effectuent des interactions $\pi - \pi$ (π -stacking) avec les résidus **HIS : 196** et **HIP : 110**, respectivement.
- La fonction cétone agit comme accepteur et crée une liaison hydrogène avec **GLN : 23**.

Pour le **ligand 18**, l'atome d'oxygène du groupe (OH) de naphthalène agit comme un donneur de liaison d'hydrogène et crée une liaison hydrogène avec **GLY : 198**, les deux cycles de naphthalène effectuent des interactions $\pi - \pi$ (π -stacking) avec le résidu **HIS : 199**. La fonction cétone agit comme accepteur et crée une liaison d'hydrogène avec **HIP: 110**(Figure III.4d).

Ces interactions globales avec HIP : 110 pourraient être la raison de score de docking élevé et par conséquent il permet d'être puissant et aussi sélectif pour l'activité biologique étudiée.

III.3.ADMET propriétés

Les propriétés d'analyse ADMET (absorption, distribution, métabolisme, élimination, toxicité) des composés peuvent être déterminées in-silico en utilisant le module qikprop de Schrödinger suite 2018. Nous avons évalué les descripteurs physico-chimiques et les propriétés pharmaceutiques pertinentes des ligands actifs pour analyser les propriétés médicamenteuses (tableau III.5). Le nombre calculé de réactions métaboliques probables (#metab) est 3 pour tous les ligands sélectionnés. . Tous les ligands actifs ont montré de bonnes valeurs de coefficient de partage (QP log Po / w) (2à 4.25), qui étaient essentielles pour l'absorption et la distribution des médicaments. Le nombre stimulé de liaisons hydrogène qui seraient données par le soluté aux molécules d'eau dans une solution aqueuse des composés est compris entre 1 et 3. Le nombre estimé de liaisons hydrogène qui seraient acceptées par le soluté à partir de molécules d'eau dans une solution aqueuse des composés est compris entre 3 et 5. Le nombre de violations de la règle de Lipinski de cinq est de 0. Les composés ont un% d'absorption orale humaine compris entre 82 et100%. Le coefficient de partage cerveau / sang estimé (QPlogBB) est compris entre -1 et -2. À partir de ces résultats, nous pouvons dire que les ligands actifs sélectionnés peuvent être utilisés dans les essais cliniques en raison de bonnes propriétés ADMET. Les détails des propriétés ADMET pour les ligands actifs sont indiqués dans le **Tableau III.5**.

Tableau.III.5 : criplage in silico ADMET pour des ligands sélectionnés.

| Ligands | QPlogBB | QPlogHERG | QPlogKP | QPlog caco | QPlogMDCK |
|---------------------------|--------------------|-----------------|---------------|---------------------------------|---------------------------------|
| 6 | -1.033 | -5.03 | -2.66 | 402.63 | 263.5 |
| 14 | -1.06 | -6.03 | -1.26 | 1216.6 | 611.5 |
| 17 | -1.55 | -5.24 | -3.215 | 151.92 | 64.53 |
| 18 | -1.55 | -5.24 | -3.215 | 151.925 | 64.53 |
| Recommended values | -3.0 to 1.0 | Above -5 | -8 to1 | ≤ 25: poor; ≥ 500: great | ≤ 25: poor; ≥ 500: great |

| Ligands | Donor HB | Accept HB | QlogP o/w | # metab | QPlog Khsa | Rule of Five | % Human Absorption | Oral |
|---------------------------|------------|-------------|-------------------|--------------|------------------|--------------|--|------|
| 6 | 2 | 3.5 | 2 | 3 | -0.210 | 0 | 85.26 | |
| 14 | 1 | 3.5 | 4.25 | 3 | 0.423 | 0 | 100 | |
| 17 | 2 | 4.25 | 2.76 | 3 | 0.215 | 0 | 82.2 | |
| 18 | 2 | 3.25 | 2.76 | 3 | 0.215 | 0 | 82.2 | |
| Recommended values | ≤ 5 | ≤ 10 | -2.0 – 6.5 | 1 – 8 | -1.5– 1.5 | max 4 | >80% is high <25% is poor | |

Références

- [1].National Center for Biotechnology Information. PubChem Database
<https://pubchem.ncbi.nlm.nih.gov> .:
- [2]. S.-H. Kim et al. Chalcones, inhibitors for topoisomerase I and cathepsin B and L, as potential anti-cancer agents , Bioorg. Med. Chem. Lett. **2013**, 23, 3320–3324
- [3].Protein Data Bank (<https://www.rcsb.org>).
- [4].GLIDE, Schrodinger, LLC, New York, USA, **2018**.
- [5]. Schrodinger suite, Maestro version 11.8, **2018**.
- [6]. R. A. Friesner et al., “Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy,” J. Med. Chem., **2004**, 47, 7, 1739–1749.

Conclusion générale

Dans cette étude, des analyses de modèle QSAR basé sur le champ ont été utilisées pour prédire l'activité contre la cathepsine B d'un ensemble de dérivés de chalcone.

Les modèles QSAR ont donné de bons résultats statistiques en termes de valeurs r^2 et Q^2 ($r^2=0.87$, $Q^2=0.506$). Le modèle QSAR basé sur le champ a fourni une corrélation significative du champ stérique, hydrophobe, accepteur de liaison d'hydrogène et électrostatique avec les valeurs d'activité biologique. Les champs du modèle 3D QSAR au niveau des molécules alignées sur l'activité biologique ont été expliqués en analysant les aires de contour.

L'amarrage a été effectué pour étudier le mode de liaison des ligands actifs sélectionnés sur le récepteur afin d'obtenir des informations pour une optimisation plus poussée de la structure. Le ligand L-14 affiche le meilleur score d'amarrage.

Le criblage insilico ADMET de ces ligands actifs a également été effectué et les valeurs de toutes les propriétés sont dans les valeurs recommandées.

Les informations obtenues dans cette étude fournissent une méthodologie pour diriger la conception de nouvelles molécules potentielles de dérivés de chalcone pour inhiber la cathepsine B.

