

الجمهورية الجزائرية الديمقراطية الشعبية وزارة التعليم العالي والبحث العلمي

République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieure et de la recherche Scientifique

جامعة محمد الصديق بن يحيى جيجل



Université de Jijel

Mémoire de fin d'Etudes

En vue de l'obtention du diplôme : Master Académique

Faculté des sciences de la nature et de la vie

Département : Biologie Moléculaire et Cellulaire

Filière : Science Biologique

Option : Biochimie

Thème

Approches et méthodes de bioinformatique pour l'étude des maladies chroniques

Membre de jury :

Présidente : Dr. Medouri Asma

Examinatrice : Dr. Brihoum Hadjer

Encadrante : Dr. Benchouieb Ilham

Présenté par :

M^{elle} Kaddour Rima

M^{elle} Nibouche Nassiha

M^{elle} Zouaoui Dounia

Année Universitaire :2022 /2023



REMERCIEMENTS



Avant tous nous remercions « **Allah** », le tout puissant de nous avoir donné la volonté, la force, le courage et la patience nécessaire pour accomplir ce travail.

Nous tenons à exprimer nos profonds remerciements à notre encadrant, Dr **Ilham Benchouieb** pour sa confiance, sa gentillesse, ses conseils et sa disponibilité tout au long de la réalisation de ce mémoire.

On tient particulièrement à remercier les membres de jury Dr. **Asma Medouri** et Dr. **Hadjer Brihoum** d'avoir accepté de jurer ce travail.

En fin, Nos remerciements les plus sincères aux personnes qui nous apporté leur aide et qui ont contribué à l'élaboration de ce mémoire ainsi qu'à la réussite de cette formidable année. Merci à tous et à toutes.

Nous tenons aussi à remercier tous nos collègues de la promotion biochimie 2022-2023 et nous leur souhaitons beaucoup de réussite.



Dédicaces

Je remercie Dieu le tout Puissant, qui m'aider et m'a donné la force, la santé et le courage pour réaliser ce travail.

Je dédie ce mémoire :

A mes chers parents pour tous leur sacrifice et leur soutien tout au long de mon parcours pour suivre le chemin de la réussite afin de réaliser mon but.

À ma chère sœur Anissa et mon cher frère Khaled, pour leur soutien et leur encouragement continu.

À mes proches Aya, Yasmin, Amina que j'aime

À mes binômes Nassiha et Rima dont ont partagée des moments difficiles mais agréables lors de la réalisation de ce travail.

À tout ce qui m'ont apportée de l'aide de près ou de loin.

A toute les familles Zouaoui et Fates.





Dédicaces

Je remercie avant tous ALLAH. Le Tout puissant, de m'avoir donné la force pour survivre, la compétence et le courage pour dépasser toutes les difficultés et Finaliser ce modeste travail de recherche

À ma chère mère

Tu as toujours été la star qui illumine ma vie. Tu n'arrêtes jamais de m'encourager, tu as veillé sur moi. Il n'y pas de mots pour exprimer mes sentiments envers toi ; je t'aime très fort. Je te souhaite une longue et belle vie et Que Dieu te protège pour nous

À mon cher père

Aucun dédicace ne e saurait exprimer mon amour, l'estime et le respect qui j'ai toujours pour toi merci pour les efforts fournis jour et nuit pour mon éducation et mon bien être. Ce travail est le fruit de tes sacrifices que tu as consentis pour mon éducation et ma formation le long de ces Années

À mes binômes Dounia et Nassiha dont ont partagée des moments difficiles mais agréables lors de la réalisation de ce travail.

Je le dédie aussi à tous les enseignants de notre faculté qui ont toujours guidé tout au long de mon parcours éducatif A mes amis qui ont cru en moi et qui ont toujours encouragé, et avec qui j'ai passé des années inoubliables





Dédicaces

Je remercie Dieu le tout Puissant, qui m'aider et m'a donné la force, la santé et le courage pour réaliser ce travail.

Du fond de mon cœur et avec l'intensité de mes émotions, Je dédie ce travail :

À mes parents qui m'ont enseigné la persévérance dans mes études, qui m'ont toujours été d'un grand secours par leur soutien et leur encouragement pendant les moments difficiles, qu'ils trouvent ici le fruit de leur labeur et le témoignage de mon grand amour

Que le dieu me les protège.

À mes très chers frères Ismail et Mouhammed Amine, Ainsi à mes sœurs Soulima, Kawtar, pour leurs amours et leurs encouragements

Je formule les plus sincères vœux de bonheur et d'un avenir radieux À mon fiancé.

À mes tînômes Dounia et Rima dont ont partagée des moments très difficiles mais agréables lors de la réalisation de ce travail.



Nassiha

Table des matières

Remerciment

Dédicaces

Liste des abréviations

Liste des figures

Liste des tableaux

INTRODUCTION 1

Chapitre □ : Introduction de la bio-informatique

I . 1. Définition de la bio-informatique 3

I . 2. Les types de la bio-informatique 4

I . 2.1. La bio-informatique structurelle 4

I . 2.2. La bio-informatique des séquences 4

I . 2.3. La bio-informatique des réseaux 4

I . 3. Les principaux composants de la bio-informatique 4

3.1 Création de données ou bases de données 5

3.2 Développements des algorithmes 5

3.3 Analyses et interprétation des données 5

I . 4. Historique de la bio-informatique 6

I . 5. Développement de la bio-informatique 8

I . 6. Les domaines d'applications de la bio-informatique 8

6.1 La médecine 9

6.2 La médecine personnelle 9

6.3 Découverte de médicaments 9

6.4 L'agriculture 11

6.5 Industries laitières.....	12
6.6 Chemoinformatique.....	12
I . 7. Les différents études « omiques » et diverses applications de bio-informatique	12
7.1 La Génomique	12
7.2 La transcriptomique	12
7.3 Métabolomique	13
7.4 Protéomique.....	13
I . 8. Les bases de données en biologies.....	14
I . 8.1 Définitions d'une base de données.....	14
I . 8.2 Définition des bases de données biologiques.....	14
I . 8.3 L'importance des bases de données	14
I . 8.4 Les types des bases de données biologiques	15
I . 8.4.1 Les bases de données primaires (généralistes).....	15
I . 8.4.1.1 Bases de données de séquences de nucléotides primaires	15
A. The National Center for Biotechnology Information GenBank (NCBI GenBank)	16
B. DNA Data Bank of Japan (DDBJ)	16
C. miRBase (mi- croRNA Registry)	17
I . 8.4.1.2. Séquences de protéines et bases de données sur les structures	18
A. Swiss-Prot	18
B. Protein Information Resource (PIR)	18
C. Protein Databank (PDB).....	18
I . 8.4.1.3 Bases de données sur les microréseaux et la génomique fonctionnelle.....	19
A. Gene Expression Omnibus (GEO).....	19
B. ArrayExpress Archives.....	20
I . 8.4.2 Les bases de données secondaires (spécialisées).....	21
I . 8.4.3 Bases de données composites.....	22
I . 8.5 Différence entre base de données et banque de données	23

I . 9. Les algorithmes utilisés en bio-informatique	25
9.1 L'algorithme	25
9.2 Alignement de séquences	26
9.3 Les types d'alignement	26
9.3.1 L'alignement global	26
9.3.2 L'alignement local.....	28

Chapitre □ : Les méthodes employées dans la bioinformatique

II .1. Le séquençage de l'ADN	29
1.1 Les techniques de séquençage de l'ADN	29
1.2 Les applications du séquençage.....	30
II . 2.Docking moléculaire	31
1.3 Les types du docking moléculaire	32
1.4 Les étapes de docking moléculaire	32
1.5 Les différents types d'interactions	33
1.6 Les applications du docking moléculaire	34
II .3. Modélisation moléculaire	35
II .3.1. Les méthodes de la modélisation moléculaires	37
II .3.2. Les applications de la modélisation moléculaire	38
II 4. Drug design	39
II 4.1. Les étapes de Drug design	39
II 4.2. Les études ADMET.....	42
II 4.3. Les application de Drug design	44

Chapitre □ : L'étude In silico des maladies chroniques (Analyse d'articles)	
Méthodologie de l'étude.....	47
III.1. Objectif de l'étude.....	47
III.2. Schéma de l'étude.....	47
III.2-3. Analyse des articles sélectionnés.....	47
Analyse d'articles.....	48
-Objectif	
-Résultats	
-Discussion	
-Conclusion	
CONCLUSION	92
Références bibliographiques.....	93
Résumé	

Liste des abréviations

AC : Activité biologique

ADMET : Absorption, Distribution, Métabolisme, Excrétion et Toxicité

ADN : Acide DésoxyriboNucléique

AGPI-LC : Acide Gras Polyinsaturé à Longue Chaîne

ANP : Ester Adénylate de l'acide Phospho,aminophosphonique.

ARN-Seq : Séquençage de l'ARN

ASM : Alignements de Séquences Multiples

ATP: Adénosine-Triphosphate

BLAST: Basic Local Alignment Search Tool

CDD: Conserved Domain Database

CF: *Coccidioides immitis*

CPNPC : Cancer du Poumon Non à Petite Cellules

DBRD : Disponibilité de Brookhaven Raster Display

DDBJ: The DNA Data Bank of Japan.

DFT : Théorie de la Fonctionnelle de la Densité

DM : Dynamique Moléculaire

DSC : Dérivées de la traduction de toutes les séquences codantes

EBI Genomes: The European Bioinformatics Institute Genomes

ECA: Energie de Contact Atomique

EGFR: Epithelial Growth Factor

EMBL: European Molecular Biology Laboratory

ENA: European Nucleotide Archive

EXPASY: Expert Protein Analysis System

FASTA: Fast Alignment Sequence Test for Application

FDA: Food and Drug Administration

FTO: Fat mass and Obesity-associated protéin

GenBank: Genetic Sequence Databank

GEO: Gene Expression Omnibus

GSK-3 β : Glycogène Synthase Kinase-3 bêta

ING : Institut National de Génétique

INSDC : International Sequence Database Collaboration

KEGG: Kyoto Encyclopedia of Genes and Genomes

MAOB : Monoamine Oxydase B

Mi RBase: MicroRNA Registry

MM : Mécanique Moléculaire

MP : Maladie de Parkinson

MQ : Mécanique Quantique

NBRF: National Biomedical Research Foundation

NCBI GenBank:The National Center for Biotechnology Information GenBank

OMIM: Online Mendelian Inheritance in Man

OMS : Organisation Mondiale de la Santé

ONS : Office National des Statistiques

PDB : Protein Data Bank

PHA : Protéine Alpha-Amylase

PIR : Protein Information Resource

PR : Polyarthrite Rhumatoïde

QSAR : Quantitative structure Relationship

RCPG : Récepteurs couplés aux protéines G

SGBD : Système de Gestion de Base de Données

SQL: Structure Query Langage

TNF: Tumor Necrosis Factor Alpha

UniProtKB: UniProt Knowledgebase

Liste des figures

Figure 1 : Les domaines liées à la bio-informatique.....	3
Figure 2 : Le criblage virtuel dans le contexte de la découverte de nouvelles molécules actives.....	11
Figure 3 : L'interaction entre différentes données omiques et diverses applications bio-informatiques.....	13
Figure 4 : Interface de la base de données GenBank.....	16

Figure 5 : Interface de la base de données DNA Data Bank of Japan (DDBJ).....	17
Figure 6 : Interface de la base de données miRBase.....	18
Figure 7 : Interface de la base de données Protein Databank (PDB).....	19
Figure 8 : Interface de la base de données Gene Expression Omnibus (GEO).....	20
Figure 9 : Interface de la base de données Array Express Archives.....	21
Figure 10 : Représentation d'une matrice de similarité (Dot-Plot) pour deux séquences d'ADN.....	27
Figure 11 : Processus de l'opération BLAST.....	27
Figure 12 : Alignement global et local.....	28
Figure 13 : Les techniques de séquençage de l'ADN.....	29
Figure 14 : Interaction protéine- ligand.....	31
Figure 15 : Principe général de docking moléculaire.....	31
Figure 16 : Représentation du repliement d'une immunoglobuline.....	37
Figure 17 : Les phases de découverte et développement d'un médicament.....	41
Figure 18 : Flux de travail pour la découverte de médicaments basés sur la structure.....	43
Figure 19 : Confirmation de liaison du ligand avec le domaine de liaison ATP des tyrosines de l'EGFR.....	49
Figure 20 : 4JJ7 docké avec le stigmastérol montrant diverses interactions avec les acides aminés - ILE333, TYR334, PHE399, GLU396, LEU401, MET403 et THR469.....	54
Figure 21 : 6E28 docké avec le stigmastérol montrant diverses interactions avec les acides aminés UNK0, TRP11 et ARG18.....	54
Figure 22 : 5GJH docké avec le stigmastérol montrant diverses interactions avec les acides aminés HIS98, PHE105, LYS106 et TRP118.....	55
Figure 23 : 5NI9 docké avec le stigmastérol montrant diverses interactions avec les acides aminés VAL129, LYS139, VAL142, LEU147 et VAL159.....	55
Figure 24 : 5N86 ancré avec le stigmastérol montrant diverses interactions avec les acides aminés SER1, TYR78 et TRP120.....	55

Figure 25 : 3CZF arrimé au stigmastérol montrant diverses interactions avec les acides aminés PHE8, TYR26, TYR27, TYR63, PRO235, ARG239 et PHE241	56
Figure 26 : 5BXU docké avec le stigmastérol montrant diverses interactions avec les acides aminés TYR536, ASP556, TYR569 et HIS571.....	56
Figure 27 : 4FKL docké avec le stigmastérol montrant diverses interactions avec les acides aminés - ILE10, VAL18, GLU81, LEU83, LYS129 et LEU134.....	56
Figure 28 : 4EK4 arrimé au stigmastérol montrant diverses interactions avec les acides aminés TYR15, ILE35, PRO45, PRO155 et ARG157.....	57
Figure 29 : 4RI5 arrimé au stigmastérol montrant diverses interactions avec les acides aminés LYS666 et ASN826.....	57
Figure 30 : Structure 3D du plomb 1. acide ascorbique 2. Bêta sitostérol 3. Sesquiterpène 4. Tocophérol 5. cétirizine 6.acide salicylique 7.ibuprofène et 8. Célécoxib.....	59
Figure 31 : Protéine cible Récepteur de l'histamine 1 avec le code PDB 3RZE.....	60
Figure 32 : Protéine cible Prostaglandine H2 synthases avec le code PDB 1IGX.....	61
Figure 33 : Protéine cible Cyclooxygénase 2 avec le code PDB 6COX.....	61
Figure 34 : Protéine cible Cyclooxygénase I avec le code PDB 3KK6.....	61
Figure 35 : Poches de liaison possibles du ligand à la surface de la cible contre le récepteur de l'histamine 1 avec le code PDB 3RZE. Poches calculées par GHECOM. 1. acide ascorbique 2.bêta sitostérol3.sesquiterpène 4.tocophérol et 5. Cetirizine.....	62
Figure 36 : Poches de liaison possibles du ligand à la surface de la cible contre le récepteur de la prostaglandine H2 synthases avec le code PDB 1IGX calculées par GHECOM.1.acide ascorbique 2.bêta sitostérol 3.sesquiterpène 4.tocophérol et 5. Acide salicylique.....	63
Figure37 : Poches de liaison possibles du ligand à la surface de la cible contre le récepteur de la cyclo-oxygénase I avec le code PDB 3KK6. Par GHECOM.1.acide ascorbique 2.bêta sitostérol 3.sesquiterpène 4. tocophérol et 5. Ibuprofène.....	64
Figure 38 : Poches de liaison possibles du ligand à la surface de la cible contre le récepteur de la cyclo-oxygénase II avec le code PDB 6COX.....	66
Figure 39 : 3 LFM (protéine FTO).....	68
Figure 40 :1LV6 (Resistin).....	68
Figure 41 :1AX8 (Leptin).....	68
Figure 42 : Interaction de la protéine FTO avec la daidzéine.....	71
Figure 43 : Interaction de la protéine FTO avec la catéchine.....	71
Figure 44 : Interaction de la résistine avec la quercétine.....	71
Figure 45 : Interaction de la leptine avec la catéchine.....	72
Figure 46 : Interaction de la leptine avec la daidezine.....	72

Figure 47 : (a) Forme tridimensionnelle de la structure secondaire de la protéine Mono amino Oxidase B (MAOB). (b-v) Forme tridimensionnelle du complexe ancré de divers inhibiteurs de la MAOB et de la protéine MAOB.....	75
Figure 48 : Structure chimique du morine et de la chroménone	79
Figure 49 : Diagramme de Ramachandran pour la protéine modèle d'homologie.	81
Figure 50 : Structure chimique des 12 principaux dérivés de chroménone.....	83
Figure51 : Présentation des modèles 2D et 3D des interactions entre le composé 1 et la protéine modélisée (GSK-3 β).....	84
Figure 52 : Présentation des modèles 2D et 3D des interactions entre le composé 2 et la protéine modélisée (GSK-3 β).....	85
Figure53 : Présentation des modèles 2D et 3D des interactions entre le composé 3 et la protéine modélisée (GSK-3 β).....	86
Figure 54 : Structure 2D et 3D de la liaison de la berbérine avec la protéine PDB ID : 4X9Y.....	89

Liste des tableaux

Tableau 1 : Les étapes d'évolution de la bio-informatique.....	7
Tableau 2 : Les Banques de données biologique.....	24
Tableau 3 : Les principaux programmes de docking moléculaire.....	34
Tableau 4 : Résultats de l'amarrage moléculaire.....	50

Tableau 5 : Sources et structures des six substances phytochimiques les plus recherchées...	51
Tableau 6 : Principales conformations d'amarrage indiquant l'énergie de liaison des structures amarrées.....	53
Tableau 7 : Propriétés des ligands des molécules de plomb sélectionnées.....	59
Tableau 8 : Résumé des études d'ancrage moléculaire des composés contre le récepteur de l'histamine 1.....	62
Tableau 9 : Résumé des études d'ancrage moléculaire des composés contre les synthèses de la prostaglandine H2.....	63
Tableau 10 : Résumé des études d'amarrage moléculaire des composés contre la cyclooxygénase I.....	64
Tableau 11 : Résumé des études d'amarrage moléculaire des composés contre la cyclooxygénase II.....	65
Tableau 12 : Interaction de FTO avec les ligands.....	69
Tableau 13 : Interaction de la résistine avec les ligands.....	70
Tableau 14 : Interaction de la leptine avec des ligands.....	70
Tableau 15 : Le nom du composé, le numéro d'accession à la banque de médicaments et la valeur de l'énergie de contact atomique (ECA) (Kcal/Mol) pour les figures (b-v).....	76
Tableau 16 : Résultats de la simulation d'amarrage moléculaire pour les composés sélectionnés et le récepteur modélisé (GSK-3 β).....	82

INTRODUCTION

Les maladies chroniques sont des pathologies de longue durée qui évoluent et s'aggravent souvent avec le temps. Leurs impacts sur la vie des individus qui en sont atteints et leur entourage sont énormes. D'après l'Organisation Mondiale de la Santé (OMS), les maladies chroniques sont responsables de 63 % des décès annuels, ce qui représente la première cause de mortalité dans le monde **(Wei Li et al., 2023)**.

Ces maladies menacent les Algériens. Les chiffres de l'office nationale des statistiques (ONS) sont effarants : au moins 10% de la population algérienne, un peu plus de 3 millions de personnes, sont touchées par les maladies chroniques. Ça prévalence augmente de manière sensible avec l'âge, notamment auprès des personnes âgées de 60 ans au plus. Les maladies les plus répandues sont l'hypertension artérielle (41.8%), le diabète (20.1%), les maladies cardiovasculaires (16.5%), l'asthme (11.5%), les maladies articulaires (10.6%) **(APS, 2021)**.

Dans le monde de la recherche, les modèles « in vivo » et « in vitro » sont utilisés dans les essais cliniques, les études scientifiques et les nouvelles procédures médicales. Les tests in vivo sont souvent utilisés plutôt qu'in vitro car ils sont mieux adaptés pour observer les effets globaux d'une expérience sur un sujet vivant. Bien qu'il existe de nombreuses raisons de croire que les études in vivo ont le potentiel d'offrir des informations concluantes sur la nature de la médecine et de la maladie, il existe un certain nombre de façons dont ces conclusions peuvent être trompeuses **(Cherid et Cheraitia, 2022)**.

Les essais cliniques sont souvent moins clairs, mal conçus et donnent des résultats ambigus ou inutiles. Le recrutement d'un nombre suffisant de patients pour obtenir une taille d'échantillon correcte, les préjugés et les perceptions lors du recrutement, le retrait des patients d'un essai, le manque de transparence ou l'absence d'un examen suffisant par des experts sont des facteurs très réels qui peuvent brouiller l'étude et rendre difficile l'obtention de conclusion claires et nettes **(Cherid et Cheraitia, 2022)**.

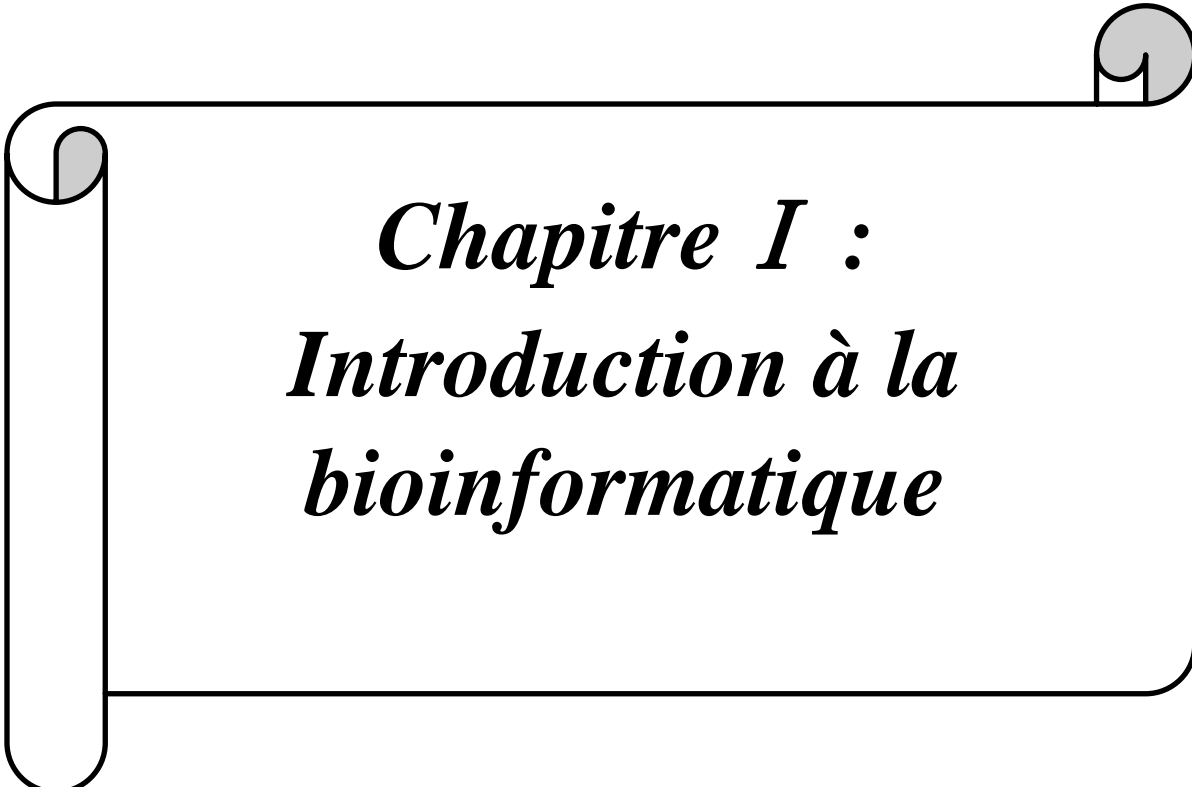
Comme nous le savons, la bio-informatique est reconnue comme la science du 21^e siècle et possède un énorme potentiel pour décoder des systèmes biologiques complexes via l'analyse et l'intégration de données multiomiques **(pathak et al., 2022)**. La bio-informatique est un domaine de recherche qui vise à rechercher des connaissances sur des données biologiques, plus spécifiquement des biomolécules, à travers des modèles et des algorithmes issus de l'informatique. Elle couvre la collecte, le stockage, la récupération, la manipulation et la modélisation des données pour l'analyse, la visualisation ou la prédiction grâce au développement d'algorithmes et de stratégies de calcul. Il utilise également des connaissances

en physique, chimie, statistique et mathématiques dans la résolution de problèmes biologiques (Almeida Paiva et al., 2022).

Les études « in silico » représentent une voie de recherche relativement nouvelle, elles ont commencé à être largement utilisées dans les études qui prédisent comment les médicaments interagissent avec le corps et avec les agents pathogènes. Les approches conventionnelles dans le domaine biologique ont été prises en charge par les progrès dans le domaine de l'informatique et avec l'avènement des stratégies in silico, les protocoles ont été convertis en stratégies de calcul à haut débit afin d'économiser du temps, de l'argent et de l'énergie (Mitra et al., 2022).

L'objectif de notre travail est d'étudier les différentes méthodes employées dans la bioinformatique pour l'étude des maladies chroniques, en ce précisant sur l'importance du docking moléculaire qui vise la modélisation de la structure d'un complexe protéine-ligand permettant de comprendre les interactions entre un composé potentiel « ligand » et sa cible thérapeutique « protéine » dans le but de concevoir des molécules thérapeutiques.

Notre travail se divise en 3 chapitres, le 1^{er} chapitre : Introduction à la bioinformatique, le 2^{ème} chapitre : les méthodes employées dans la bioinformatique, le 3^{ème} : Analyse d'articles.



Chapitre I :
Introduction à la
bioinformatique

I .1. Définition de la bioinformatique

La bioinformatique est une science multidisciplinaire, elle est située au carrefour entre la biologie, l'informatique, la chimie et beaucoup d'autre discipline (**Hammad, 2020**) qui résout des problèmes biologiques (**Rather et al., 2023**). C'est la discipline de l'analyse « in silico » de l'information biologique contenue dans les séquences nucléotidiques (séquences de nucléotides) et protéique (séquences d'acides aminées) (**Coutouly et al., 2012**).

En 1970, Paulien Hogeweg et Ben Hesper ont développé le terme "bioinformatique", qui combine les deux sciences, la science biologique et l'informatique. Elle offre une plateforme stimulante pour le développement, l'extension et l'application de nouveaux algorithmes et outils logiciels. Pour aider à la compréhension des processus biologiques pour l'agriculture et l'industrie et les secteurs connexes y compris l'homme (**Rather et al., 2023**).

Selon les définitions données à la bioinformatique, nous pouvons résumer les activités de cette discipline en trois points principaux :

- ✓ Organisation et acquisition des données biologiques.
- ✓ Conception de logiciels pour l'analyse, la comparaison et la modélisation des données.
- ✓ Analyse des résultats produits par les logiciels (**David ,2004**).

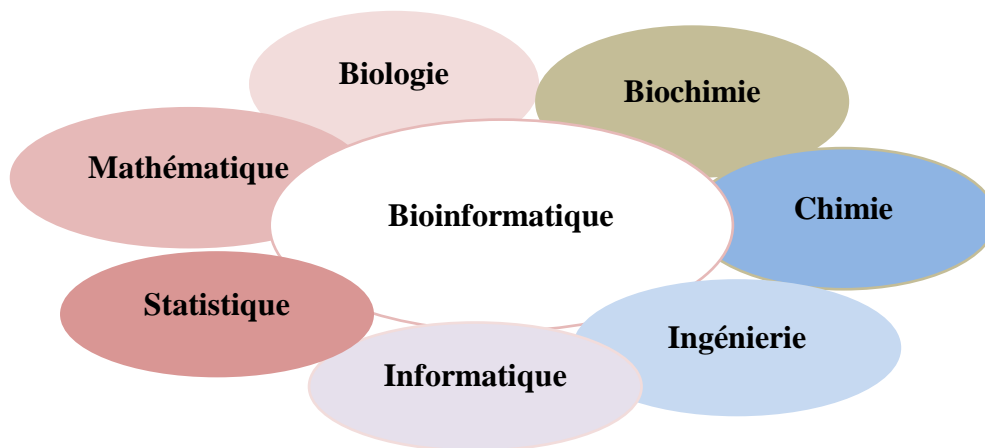


Figure 1 : Les domaines liées à la bioinformatique (**Parul, 2018**).

I.2. Les types de la bioinformatique

I.2.1 La bioinformatique structurelle

La bioinformatique structurelle est une branche de la bioinformatique portant principalement sur les problèmes de biologie structurelle, auxquels le mot "structurel".

Elle était aussi appelée "biologie structurelle computationnelle " utilisant les techniques distinctives des stimulations moléculaires computationnelle. Ses intérêts de recherche étaient principalement axés sur l'analyse et la prédiction des structures tridimensionnelle et des fonctions connexes des macromolécules biologiques telles que les protéines, l'ARN et l'ADN (Wei et al., 2015).

I.2.2 La bioinformatique des séquences

La bioinformatique des séquences, qui traite l'analyse des données issues de l'information génétique contenue dans la séquence de l'ADN et/ou ARN. Cette branche s'intéresse en particulier à l'identification des ressemblances entre les séquences, à l'identification des gènes ou de régions biologiquement pertinentes dans l'ADN ou dans les protéines, en se basant sur l'enchaînement ou séquence de leurs composants élémentaires (nucléotides, acides aminés).

I 2.3 La bioinformatique des réseaux

La bioinformatique des réseaux qui s'intéresse aux interactions entre gènes, protéines, cellules, organismes, en essayant d'analyser et de modéliser les comportements collectifs d'ensembles de briques élémentaires du Vivant. Cette partie de la bioinformatique se nourrit en particulier des données issues de technologies d'analyse à haut débit comme la protéomique ou la transcriptomique pour analyser des flux génétiques ou métaboliques (Cheghi et Benfadel, 2020).

I .3. Les principaux composants de la bioinformatique

- Création de données ou bases de données.
- Développements des algorithmes.
- Analyses et interprétation des données.

I 3.1 Création de données ou bases de données

La collecte, l'organisation, le stockage et la gestion des collections de données biologiques font partie de la bioinformatique. Les bases de données biologiques sont des dépôts en ligne de données biologiques issues d'études scientifiques, de la littérature publiée, de l'expérimentation à haut débit et de l'analyse informatique. Les chercheurs et les universitaires du monde entier peuvent accéder aux bases de données pour obtenir des informations générales et en soumettre de nouvelles.

Une base de données peut être utilisée pour découvrir un gène en termes d'informations évolutives, informations structurales, informations fonctionnelles, etc.

Des exemples bien connus de bases de données biologiques sont la GenBank, la base de données sur les gènes, la base de données sur l'ADN European Molecular Biology Laboratory (EMBL) (<https://www.embl.org/>) et la DNA Data Bank of Japan (<https://www.ddbj.nig.ac.jp/>) (Rather et al., 2023).

I 3.2 Développements des algorithmes

Une procédure ou une technique simple, étape par étape, qui prend des données d'entrée et génère une sortie est appelée algorithme. Il est couramment utilisé pour le traitement des données, le calcul et d'autres activités informatiques et mathématiques.

Un algorithme peut être utilisé pour modifier des données de différentes manières, y compris l'introduction de nouvelles données, la recherche de données spécifiques, etc.

Le développement d'algorithmes permet d'obtenir des logiciels et des outils particuliers qui peuvent être utilisés pour évaluer de grandes quantités de données, tels que les données de séquençage de l'ADN de nouvelle génération, les données de séquençage des protéines, le séquençage de l'ARN (RNA-Seq), etc. (Rather et al., 2023).

I 3.4 Analyses et interprétation des données

L'application correcte des composantes de la bio-informatique implique l'évaluation des données et l'interprétation des résultats d'une manière biologiquement pertinente, telles que les données issues du séquençage de l'ADN de séquençage des protéines, séquençage de l'ARN (RNA-Seq), profils d'expression génique, voies métaboliques et marqueurs moléculaires.

Les logiciels/outils les plus utilisés pour les données biologiques volumineuses sont Strand NGS, CLC Genomics Workbench, Lasergene Genomics Suite, NextGENe, Partek Genomics Suite, Genomatix, FreeBayes, Nanopore, GenomeComb, SIFT, PANTHER, Genome Track Analyzer, etc (**Rather et al., 2023**).

I .4. Historique de la bioinformatique

La bioinformatique est apparue au début des années 1960, une décennie avant que le séquençage de l'ADN et de l'ARN ne devienne réalisable. Wright-Fisher a été le premier scientifique à utiliser la technologie informatique en science pour résoudre une équation différentielle afin de calculer le changement de fréquence des gènes par unité de temps sous sélection dans un modèle de diffusion (**Mitra et al., 2022**).

Tableau 1 : Les étapes d'évolution de la bio-informatique (Cheghi et Benfadel, 2020).

Année	Les événements
1953	Structure en double hélice de l'ADN (Watson et Crick).
1956	-Séquence en acides aminés de la première protéine : insuline. -Fortran ((FORmulaTRANslation) : Premier langage informatique de haut niveau.
1958	Première structure 3D de protéine « myoglobine » (John Kendrew).
1955-1965	Premiers langages informatiques, premier ordinateur commercial.
1961	-Sidney Brenner, François Jacob, Matthew Meselson identifiés l'ARN messenger. -Nirenberg et Matthaei déchiffrent le code génétique.
1965	Jacques Monod et François Jacob découvrent les mécanismes de la régulation génétique impliqués dans le dogme central de la biologie Ordinateur IBM/360.
1971	-Premier microprocesseur INTEL PDB. - Protein Data Bank (structures 3D macromolécules).
1974	Programme de prédiction de structures secondaires des protéines (Chou et Fasman): "Prediction of Protein Conformation".
1977	Mise au point des techniques de séquençage de l'ADN.
1980	Création de la banque EMBL : banque européenne généraliste de séquences nucléiques.
1984	Développement de la réaction de polymérisation en chaîne (PCR) par Mullis Création de la banque NBRF : banque américaine généraliste de séquences protéiques Logiciel d'analyse de séquence (UW GCG) Devereux.
1987	-1ère carte génétique du génome humain. -Apparition de la technologie des puces à ADN. Genbank, EMBL et DDBJ s'échangent leur contenu et adoptent un système de conventions communes (The DDBJ/EMBL/Genbankfeature Table Definition).
1989	L'apparition de l'internet.
1992	-L'utilisation du terme bioinformatique comme une nouvelle discipline. -Séquençage complet du chromosome III de la levure. <i>Saccharomyces cerevisiae</i> . -13 gènes de maladies génétiques ont été identifiés par clonage positionnel.
1998	Séquençage du 1er organisme pluricellulaire (premier animal), <i>Caenorhabditis elegans</i> .
2001	-Publication des travaux de séquençage du génome humain presque complet. - Premier clone d'embryon humain.
2005	Séquençage à haut débit (high-throughput sequencing en anglais HTS).
2007	-L'apparition de la première génération des appareils de séquençage à haut débit : Le pyroséquençage. -Le séquençage avec des terminateurs réversibles. Et le séquençage par ligation.
2012	-Plus de 3040 génomes eucaryotes et procaryotes séquencés et des milliers en projet (GenomesOnLine). -Plus de 393 milliards de nucléotides.
Nos jours	- Plus 3000 articles publiés concernant la bioinformatique. Plus 1400 bases de données biologiques.

I .5. Développement de la bioinformatique

Le terme "bioinformatique" a été utilisé pour la première fois en 1989 par Daniel R. Masys cependant, la recherche en bioinformatique remonte aux années 1960. Depuis lors, la bioinformatique a évolué de façon spectaculaire.

Cependant, l'évolution de la bioinformatique au cours des 50 dernières années ne s'est pas faite de manière uniforme. Presque 50 ans ne s'est pas produite de manière linéaire, mais plutôt par étapes, et a été étroitement liée aux progrès des sciences de la vie et de l'informatique.

Les années 1970 et 1980 ont été les décennies au cours desquelles les bases de données de séquences de protéines et d'acides nucléiques ont été créées et le besoin croissant de récupération et d'analyse des données contenues dans les bases de données a conduit au développement d'outils d'accompagnement. Cependant, la détermination des données et l'utilisation de ces outils étaient décourageantes pour beaucoup, car les systèmes et les programmes informatiques n'étaient pas conviviaux.

L'émergence du projet du génome humain a fait des années 1985 et 1986 des années particulièrement importantes dans l'histoire de la bioinformatique, mais surtout, le développement d'interfaces web et la connectivité croissante à l'Internet dans les années 1990 ont commencé à rendre la bioinformatique plus conviviale, l'augmentation de la connectivité à Internet dans les années 1990 ont commencé à rendre les données plus accessibles. Entre-temps, en 1995, la première séquence du génome d'un organisme vivant libre, « Haemophilus influenza » a été publiée en 2001, et les progrès des technologies de séquençage au cours des 16 années qui ont précédé la mise en place du projet ont dépassé l'imagination de quiconque. Au cours de la dernière décennie, les progrès des techniques modernes à haut débit dans les sciences de la vie, couplées aux développements de la technologie informatique, ont ouvert une nouvelle ère pour la bioinformatique, qui s'est rapidement transformée en discipline.

Aujourd'hui, l'ampleur des données biologiques disponibles à organiser et analyser est impressionnante, mais elle offre des possibilités passionnantes qui pourraient apporter des réponses à de nombreuses questions de la recherche en sciences de la vie (**Tastan et al., 2014**).

I .6. Les domaines d'applications de la bio-informatique

La bioinformatique peut être largement appliquée à de nombreux domaines tels que la médecine, la médecine personnelle, la découverte des médicaments, l'agriculture, l'industries laitières et la chemoinformatique.

I . 6.1 La médecine

Les études assistées par ordinateur sont utiles pour détecter les maladies génétiques à un stade précoce de la vie. Dans certains cas, elles peuvent contribuer à la guérison des maladies génétiques. L'analyse du pedigree permet de conseiller les futurs parents sur la prévention de certaines maladies génétiques (**Kumar, 2019**).

I . 6.2 La médecine personnalisée

La médecine personnelle est la personnalisation des soins de santé en fonction du patient grâce à l'utilisation de l'information génétique cette dernière peuvent être largement déterminées à l'aide d'outils bioinformatiques et de bases de données (**Parul, 2018**).

La médecine personnalisée aussi appelée « médecine de précision » ou « médecine génomique » est un concept encore flou aujourd'hui. Parmi ses outils, les possibilités de séquençage de l'acide désoxyribonucléique (ADN) occupent une place centrale qui semble offrir de belles perspectives en matière de santé publique. Déjà pratiquée dans certains domaines médicaux tels que la cancérologie et la pharmacologie, elle cherche désormais aussi à agir sur la prévention des maladies chroniques en se basant sur les profils génétiques de risque. Ainsi, moyennant quelques cellules buccales, les patients peuvent déjà accéder (via des sites Internet) à leur probabilité de développer au cours de leur vie certaines maladies chroniques (diabète, certains cancers, maladies cardiovasculaires et inflammatoires). Cette approche est encore balbutiante et sujette à caution mais, par exemple, un consortium anglais et australien a pu développer un score polygénique qui augmente sensiblement la qualité de prédiction pour les maladies coronariennes (**Cardinaux et al., 2021**).

I . 6.3 Découverte de médicaments

L'un des problèmes les plus courants est la production de médicaments peu coûteux et efficaces pour une pathologie. Qui peut être résolu grâce à la bioinformatique et à la médecine personnalisée (**Sulochana, 2020**).

La découverte de médicaments est un processus qui vise à identifier un composé thérapeutiquement utile pour guérir et traiter une maladie. Ce processus implique l'identification de candidats, la synthèse, la caractérisation, la validation, l'optimisation, le criblage et les tests d'efficacité thérapeutiques. Une fois qu'un composé a démontré son importance lors de ces recherches, il entame le processus de développement de médicaments avant les essais cliniques. Le processus de développement d'un nouveau médicament doit passer par plusieurs étapes afin de produire un médicament sûr, efficace et répondant à toutes les exigences réglementaires (**Deore et al., 2019**).

Le criblage virtuel in-silico est une tâche de classification. Les éléments à classer sont des molécules et l'on est en présence de deux classes : les molécules actives sur la protéine ciblée et les molécules inactives. En situation réelle, la classe de chaque molécule est inconnue à l'avance. Aussi, en fonction du jeu de données utilisé lors de la validation scientifique d'une méthode, il se peut que seules certaines molécules aient été testées expérimentalement et détectées comme actives. Les molécules inactives sont souvent des erreurs qui ont été générés automatiquement à partir des molécules actives. Certaines erreurs se révéleraient probablement actifs s'ils étaient testés expérimentalement.

Lors de la recherche de nouvelles molécules thérapeutiques, le criblage virtuel peut être utilisé en début du processus afin de sélectionner un petit ensemble de molécules qui vont être testées expérimentalement. Typiquement, sur les millions de molécules d'une chimiothèque, entre 50 et 200 molécules seulement seront sélectionnées (**Bérenger, 2016**).

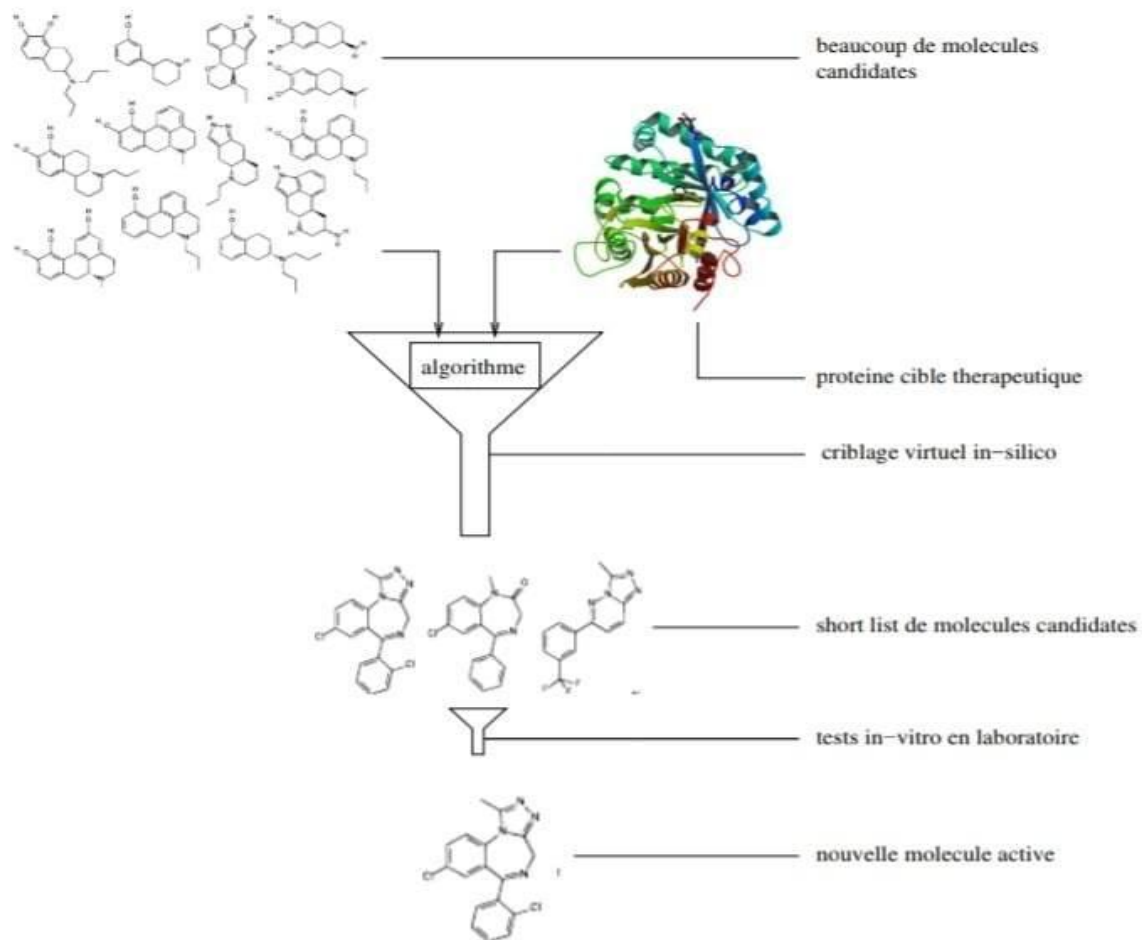


Figure 2 : Le criblage virtuel dans le contexte de la découverte de nouvelles molécules actives. Des méthodes informatiques permettent de prioriser certaines molécules afin qu’elles soient testées in-vitro en laboratoire (Bérenger, 2016).

I . 6.4 L’agriculture

Les tâches les plus critiques en bio-informatique consiste à trouver des gènes dans les séquences d'ADN de divers organismes, l'élaboration de méthodes pour prédire la structure et la fonction des séquences de protéines et d'ARN découvertes, le regroupement de familles de séquences apparentées, développement de modèles de protéines pour l'amélioration des gènes des plantes et des animaux. Les connaissances génétiques pourraient être utilisées pour produire des corps plus résistants aux insectes et améliorer la qualité du bétail, ce qui les rendrait plus sains, plus résistants aux maladies et plus productifs (Sethi, Behera, 2012).

I . 6.5 Industries laitières

Lactococcus lactis (bactérie) est l'un des micro-organismes les plus importants de l'industrie laitière. Il est utilisé dans la fabrication de produits tels que le fromage, le yaourt et le babeurre. Par conséquent, grâce aux bases de données impliquées dans la génétique des organismes, ces micro-organismes peuvent être étudiés pour améliorer la valeur nutritionnelle et la qualité des produits (**Parul, 2018**).

I . 6.6 Chemoinformatique

L'informatique chimique se concentre sur le stockage, l'indexation, la recherche, l'extraction et l'application d'informations sur les produits chimiques. Elle implique l'organisation des données chimiques sous une forme logique afin de faciliter la recherche des propriétés chimiques, des structures et de leurs relations. Grâce à la bio-informatique, il est possible d'identifier et de modifier la structure d'un produit naturel, de concevoir un composé avec des propriétés chimiques, des structures et des relations structurelles d'un produit naturel et d'évaluer ses effets thérapeutiques. L'analyse chimio-informatique comprend des analyses telles que la recherche de similarités, le regroupement, la modélisation QSAR (Quantitative Structure Relationship), le criblage virtuel (**Kumar, 2019**).

I.7. Les différents études « omiques » et diverses applications de bio-informatique

I . 7.1 La Génomique

La génomique vise à caractériser et à quantifier collectivement les gènes et comprend le séquençage, la cartographie, l'édition et l'analyse des génomes afin de comprendre mécanisme et le chemin de l'évolution d'une espèce individuelle, d'assembler et d'analyser la structure et la fonction de l'ensemble des génomes, et de déchiffrer le mécanisme de l'expression des gènes (**Aslam et al, 2016**). Les plateformes vont générer de nombreuses données génotypiques et phénotypiques. Leur exploitation optimale nécessite une connexion des données digitales, des données de santé du quotidien par le dossier patient informatisé, des données environnementales, et doit s'envisager dans le cadre d'un entrepôt national de données de santé health data hub (**Laviolle et al., 2019**).

I . 7.2 La transcriptomique

Le terme "transcriptome" désigne l'ensemble des gènes activés, des ARNm dans un tissu donné à un moment donné. L'application de la bio-informatique dans le domaine de la

transcriptomique a permis d'ajouter une dimension supplémentaire aux données génomiques actuelles en aidant les chercheurs à déterminer plus précisément les niveaux quantitatifs d'expression des gènes et dans les études d'association à l'échelle du génome avec les progrès des nouvelles technologies de séquençage et divers outils bio-informatiques (**Prasad Gajula1. 2016**).

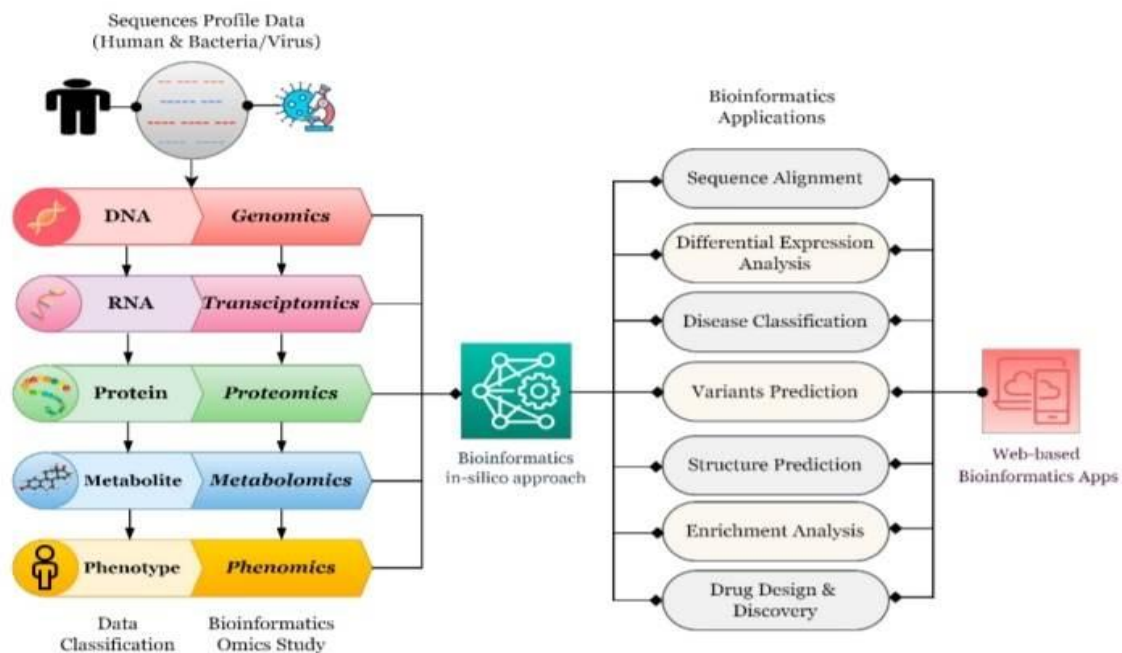


Figure 3 : L'interaction entre différentes données omiques et diverses applications bio-informatiques (**Iqbal et Kumar, 2023**).

I . 7.3 Métabolomique

La métabolomique traite de l'analyse des métabolites de petites molécules et des polymères. Au niveau de l'application, la métabolomique implique l'identification et la caractérisation d'une large gamme de métabolites en se référant à l'analyse biochimique quantitative. L'importance réside dans l'identification quantitative des métabolites qui sont des indicateurs directs du phénotype souhaité. Les outils bio-informatiques sont essentiels à chaque étape, depuis le criblage jusqu'à la sauvegarde des données, car les métabolites ne sont pas toujours bien identifiés (**Prasad Gajula1. 2016**).

I . 7.4 Protéomique

La protéomique implique l'application de technologies pour l'identification et la quantification de l'ensemble des protéines présentes dans une cellule, un tissu ou un organisme. Elle

complète les autres technologies "omiques" telles que la génomique et la transcriptomique pour expliquer l'identité des protéines d'un organisme et pour connaître la structure et les fonctions d'une cellule, d'un tissu ou d'un organisme (Aslam et al, 2016).

I .8. Les bases de données en biologies

En général, les données biologiques sont collectées et stockées dans le monde entier et la bio-informatique joue un rôle important dans la gestion des données.

Les bases de données et les programmes permettent d'accéder aux informations existantes et de comparer ces données pour trouver des similitudes et des différences. Les bases de données peuvent être consultées ou faire l'objet de références croisées sur l'intérêt ou en utilisant des versions téléchargées sur des ordinateurs locaux ou des réseaux informatiques par plusieurs utilisateurs (Reeti, 2014).

I . 8.1 Définitions d'une base de données

Une base de données est définie comme une collection organisée de données ou d'informations qui sont stockées électroniquement et accessibles à partir d'un système informatique.

La nature organisée de la base de données facilite l'accès, la gestion, la mise à jour périodique et la recherche rapide des informations requises à partir d'un système informatique approprié. (Anuj,2020).

I . 8.2 Définition des bases de données biologiques

Les bases de données biologiques sont des bibliothèques d'informations sur les sciences de la vie, recueillies à partir d'expériences scientifiques, de publications, de la technologie des expériences à haut débit et de l'analyse computationnelle et d'analyses informatiques (Kamble et khairkar, 2016).

I . 8.3 L'importance des bases de données

L'importance des bases de données biologiques peut être comprise à partir des points suivants :

- Grâce aux progrès rapides de la biologie moléculaire, de la protéomique et des technologies de séquençage du génome à haut débit et à faible coût, d'énormes quantités d'informations biologiques, telles que les données brutes, sont désormais disponibles, des données brutes de séquençage, les protéomes sont générées à un rythme très rapide. Ainsi, le stockage et le traitement de ces informations stupéfiantes sont les principaux défis de l'ère génomique actuelle.
- Outre la génération, l'analyse des données et l'élaboration de conclusions significatives sont également des éléments importants de toute recherche scientifique cela nécessite souvent le partage des données au sein de la communauté scientifique. Dans ce contexte, la base de données biologique permet aux scientifiques d'accéder à et d'extraire les données biologiques pertinentes, y compris les données brutes, les séquences génomiques, les ensembles de données analysées et les annotations dans des formats faciles à gérer et à organiser.
- Les bases de données biologiques permettent également d'indexer les données et de supprimer la redondance des données.
- À l'heure actuelle, les bases de données biologiques sont devenues l'élément central de la bio-informatique. Grâce aux divers outils d'exploration de données, il est possible d'accéder facilement à toutes les informations biologiques, ce qui permet d'économiser du temps, des ressources et des efforts (Anuj,2020).

I . 8.4. Les types des bases de données biologiques

I . 8.4.1 Les bases de données primaires (généralistes)

Les bases de données primaires, également connues sous le nom de « bases de données d'archives », contiennent essentiellement des ensembles de données dérivées expérimentales telles que les séquences de nucléotides et de protéines ainsi que les informations structurales des macromolécules. Ces informations de base peuvent être accompagnées d'une annotation fonctionnelle, bibliographies et liens vers d'autres bases de données. Les données de la base de données primaire sont directement soumises par les chercheurs. Une fois soumises, les données reçoivent un numéro d'accès, qui est permanent et fait partie du dossier scientifique.

Les exemples suivants sont des bases de données primaires :

I . 8.4.1.1. Bases de données de séquences de nucléotides primaires

The European Nucleotide Archive (ENA), The National Center for Biotechnology Information GenBank (NCBI GenBank), The DNA Data Bank of Japan (DDBJ) et mi RBase(mi-cro RNA Registry)(Anuj,2020).

A. The National Center for Biotechnology Information GenBank (NCBI GenBank)

GenBank (Genetic Sequence Databank) est l'une des banques de séquences de nucléotides connues dont la croissance est la plus rapide. Les fichiers contiennent des données sur les séquences, les numéros d'accès les noms de gènes, les classifications phylogénétiques, les références à la littérature publiée. Elle est développée et maintenue au NCBI. GenBank fait partie de International Sequence Database Collaboration (INSDC) qui comprend la base de données ENA, la base de données DDBJ et GenBank au NCBI. Ces organisations échangent des données quotidiennement (Tanizawa et al., 2023).

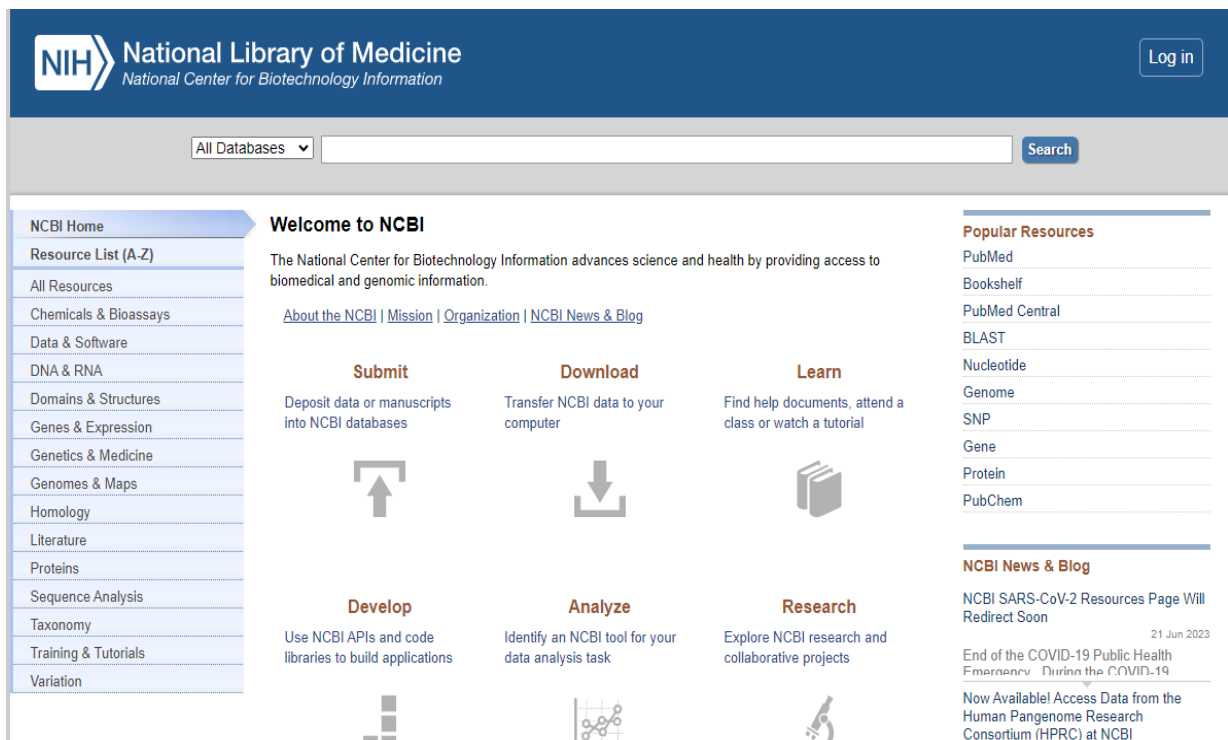


Figure 4 : Interface de la base de données GenBank <https://www.ncbi.nlm.nih.gov/> (Anuj,2020).

B. DNA Data Bank of Japan (DDBJ)

La banque de données ADN du Japon (DDBJ) est une base de données publique de séquences de nucléotides établie au centre d la bioinformation et DDBJ Center (DDBJ Center ; <https://www.ddbj.nig.ac.jp>) de l'Institut National de Génétique (ING). La DDBJ accepte les séquences de nucléotides annotées, numéros d'accession et les distribue en collaboration avec GenBank au NIG (Tanizawa et al., 2023).

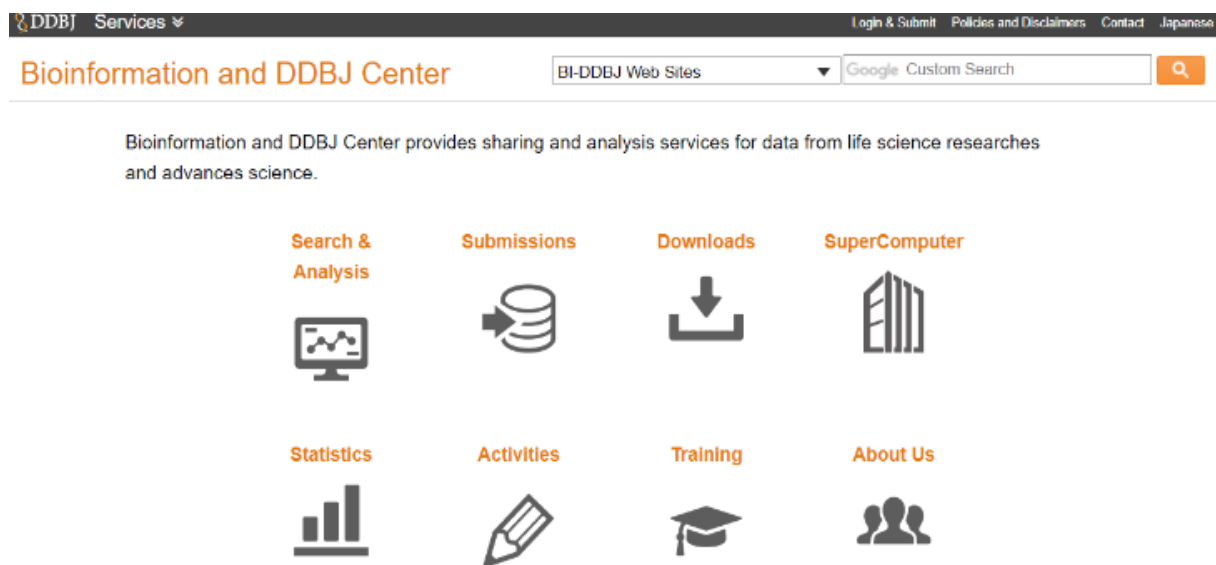


Figure 5 : Interface de la base de données DNA Data Bank of Japan (DDBJ) <https://www.ddbj.nig.ac.jp/> (Anuj,2020).

C. miRBase (mi-croRNA Registry)

La banque de données miRBase est le principal dépôt public et la source en ligne de séquences et d'annotations de microARN. Elle est Créée en 2002 et responsable de la nomenclature des gènes de microARN. La base de données miRBase regroupe les fonctions de dénomination des gènes et de base de données de séquences précédemment remplies par les microARN (Kozomara et al., 2018).

miRBase: the microRNA database

miRBase provides the following services:

- The [miRBase database](#) is a searchable database of published miRNA sequences and annotation. Each entry in the miRBase Sequence database represents a predicted hairpin portion of a miRNA transcript (termed *mir* in the database), with information on the location and sequence of the mature miRNA sequence (termed *miR*). Both hairpin and mature sequences are available for [searching](#) and [browsing](#), and entries can also be retrieved by name, keyword, references and annotation. All sequence and annotation data are also [available for download](#).
- The [miRBase Registry](#) provides miRNA gene hunters with unique names for novel miRNA genes prior to publication of results. Visit the [help pages](#) for more information about the naming service.

To receive email notification of data updates and feature changes please subscribe to the [miRBase announcements mailing list](#). Any queries about the website or naming service should be directed at mirbase@manchester.ac.uk.

miRBase is managed by the [Griffiths-Jones lab](#) at the [Faculty of Biology, Medicine and Health, University of Manchester](#) with funding from the [BBSRC](#). miRBase was previously hosted and supported by the [Wellcome Trust Sanger Institute](#).

References

If you make use of the data presented here, please cite the following articles in addition to the primary data sources:

[miRBase: from microRNA sequences to function.](#)
Kozomara A, Birgaoanu M, Griffiths-Jones S.
Nucleic Acids Res 2019 47:D155-D162

[miRBase: annotating high confidence microRNAs using deep sequencing data.](#)
Kozomara A, Griffiths-Jones S.

Figure 6: Interface de la base de données miRBase (mi-croRNA Registry) <http://www.mirbase.org/> (Anuj,2020).

I . 8.4.1.2. bases de données de séquences et structures de protéines

Swiss-Prot et Protein Information Resource (PIR) pour les séquences de protéines, Protein Databank (PDB) pour la structure des protéines.

A. Swiss-Prot

Swiss-Prot est une base de données de séquences de protéines qui s'efforce de fournir un niveau élevé d'annotations (telles que la description de la fonction d'une protéine, la structure de ses domaines, les modifications post-traductionnelles, les variantes, etc.), un niveau minimal de redondance et un niveau élevé d'intégration avec d'autres bases de données.

Les développements récents de la base de données comprennent : une augmentation du nombre et de l'étendue des organismes modèles, des références croisées avec deux bases de données supplémentaires, une variété de nouveaux fichiers de documentation et la création de TrEMBL, un supplément annoté par ordinateur à Swiss-Prot. Ce supplément se compose d'entrées dans un format similaire à celui de Swiss-Prot, dérivées de la traduction de toutes les séquences codantes (DSC) de la base de données de séquences nucléotidiques de l'EMBL, à l'exception des DSC déjà incluses dans Swiss-Prot (Bairoch et Apweiler , 1998).

B. Protein Information Resource (PIR)

PIR (Protein Information Resource) est une ressource bioinformatique publique destinée à soutenir la recherche en protéomique et les différentes études scientifiques. Il a été créé en 1984 par le NBRF (National Biomedical Research Foundation) aux États-Unis. La base de données offre une autre ressource pour soutenir la distribution et la cohérence des annotations de protéines, telles que PIRSF, ProClass et ProLINK (Sapundzhi, 2022).

C. Protein Data Bank (PDB)

La PDB est une archive universelle gratuite pour les données structurales des macromolécules biologiques. Elle a été créée en 1971 au Brookhaven National Laboratory sous la direction de de Walter Hamilton et ne contenait à l'origine que 7 structures de protéines. Deux raisons majeures ont été à l'origine de la création de la PDB, la première était l'augmentation de données relatives à la structure des protéines, une autre raison était la disponibilité de Brookhaven Raster Display (DBRD) qui permet d'envisager les structures des protéines en 3D. Il contient des structures tridimensionnelles de protéines, de fragments d'acides nucléiques, de molécules d'ARN et de protéines, de grands peptides et de structures complexes de protéines et d'acides nucléiques (Aditi et al., 2017).

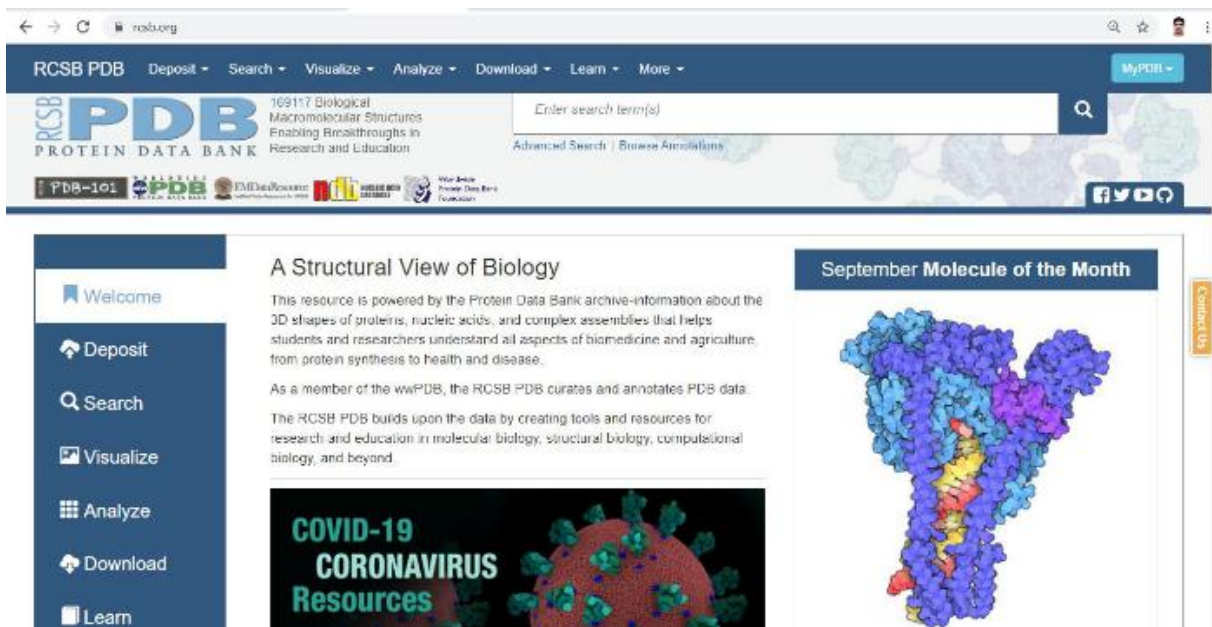


Figure 7 : Interface de la base de données Protein Databank (PDB) <https://www.rcsb.org/> (Anuj, 2020).

I . 8.4.1.3. Bases de données sur les microréseaux et la génomique fonctionnelle :

Gene Expression Omnibus (GEO) et ArrayExpress Archives.

A. Gene Expression Omnibus (GEO)

La base de données GEO est un dépôt public international qui archive et distribue gratuitement des ensembles de données d'expression génique à haut débit et d'autres données de génomique fonctionnelle. Créée en 2000 en tant que ressource mondiale pour les études d'expression génique, GEO a évolué avec les changements rapides des technologies et accepte désormais des données à haut débit pour de nombreuses autres applications de données, y compris celles qui examinent la méthylation du génome, la structure de la chromatine et les interactions génome-protéine. GEO prend en charge les normes de rapport dérivées de la communauté qui spécifient la fourniture de plusieurs éléments d'étude essentiels, notamment les données brutes, les données traitées et les métadonnées descriptives. La base de données permet non seulement d'accéder aux données de dizaines de milliers d'études, mais offre également divers outils et stratégies basés sur le web qui permettent aux utilisateurs de localiser les données correspondant à leurs intérêts spécifiques, ainsi que de visualiser et d'analyser les données (Clough et Barrett, 2016).

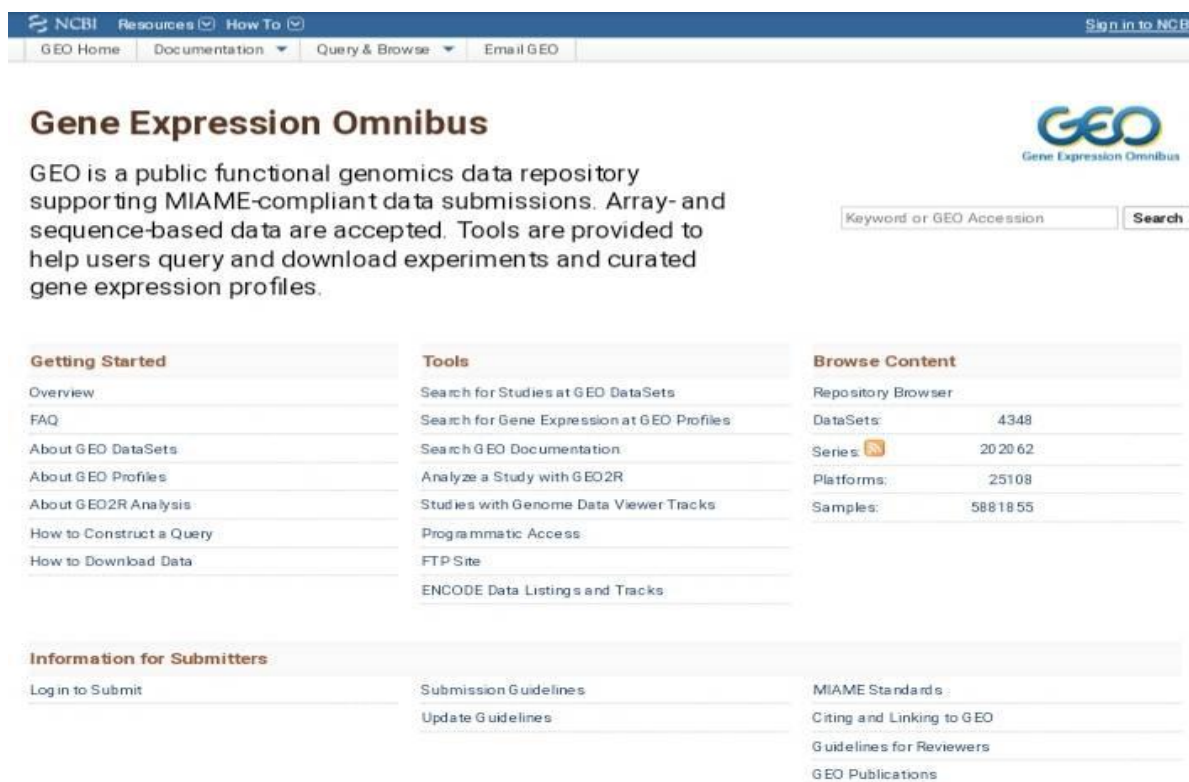


Figure 8: Interface de la base de données Gene Expression Omnibus (GEO) <http://www.ncbi.nlm.nih.gov/geo/>(Clough et Barrett, 2016).

B. ArrayExpress Archives

La base de données Array Express Archives est une archive de données de génomique fonctionnelle à l'EMBL- EBI, créé en 2002, initialement en tant qu'archive pour les données de microréseaux liées à des publications et a été de microarray liées à des publications, avant d'accepter des données basées sur le séquençage. Au cours de la dernière décennie, une part croissante biologique fait appel à de multiples technologies pour tester différentes modalités biologiques, telles que l'épigénétique et l'expression de l'ARN et des protéines (**Sarkans et al., 2021**). ArrayExpress se compose de deux parties : le référentiel ArrayExpress, qui est une archive publique de support MIAME de données de puces à ADN, et l'entrepôt de données (**Parkinson et al., 2007**).

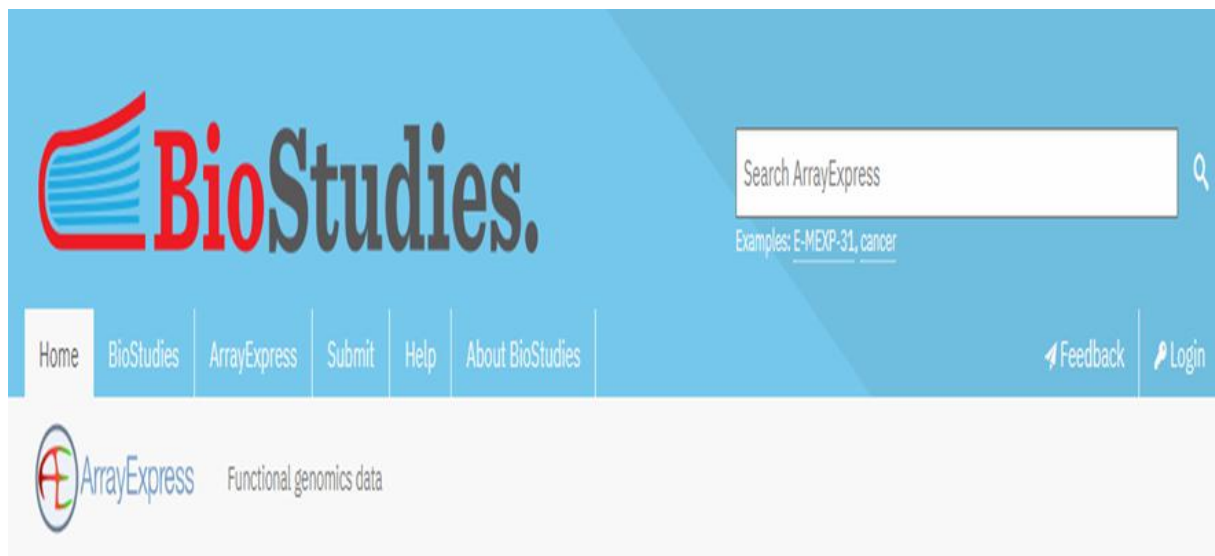


Figure 9 : Interface de la base de données Array Express Archives <https://www.ebi.ac.uk/arrayexpress> (**Sarkans et al., 2021**).

I . 8.4.2 Les bases de données secondaires (spécialisées)

Les bases de données secondaires sont celles qui contiennent des informations dérivées des bases de données primaires. Elles contiennent des informations telles que la séquence conservée, la séquence de signature et les résidus du site actif des familles de protéines, etc. Certaines bases de données comme SCOP développée à l'université de Cambridge, CATH, développée à l'université de Californie et eMOTIF à Standford, sont créées et hébergées par des chercheurs individuels dans leurs laboratoires (**Kambel et khairkar, 2016**).

Les exemples suivants sont des bases de données secondaires :

- **Bases de données de structures et familles de protéines**
InterPro, PROSITE, SCOP, CATH et NCBI Conserved Domain Database (CDD).
- **Bases de données d'informations fonctionnelles et des Séquences de protéines**
UniProt Knowledgebase (UniProtKB).
- **Bases de données d'annotation et des séquences de nucléotides (gènes/génomés)**
NCBI UniGene, The European Bioinformatics Institute Genomes (EBI Genomes), et Ensembl, (Anuj,2020).

I . 8.4.3 Bases de données composites

Les bases de données composites sont des collections de plusieurs (généralement plus de deux) ressources de base de données primaires. Cela permet d'alléger la tâche fastidieuse de recherche dans plusieurs bases de données se référant aux mêmes données.

L'approche utilisée, par exemple l'algorithme de recherche employé, diffère considérablement d'une base de données à l'autre. Par exemple : DrugBank offre des détails sur les médicaments et leurs cibles, BioGraph incorpore des connaissances assorties de la science biomédicale et Bio Model est un entrepôt de modèles computationnels des développements biologiques (Aditi et al., 2017).

- **DrugBank**

Est une ressource web complète, librement accessible, contenant des informations détaillées sur les médicaments, leurs cibles, leur action et leurs effets sur les médicaments approuvés par la Food and Drug Administration (FDA) ainsi que sur des médicaments expérimentaux en cours d'approbation par la FDA. Le contenu riche, de haute qualité et de source primaire a permis à la DrugBank de devenir l'une des références les plus utilisées au monde en matière de médicaments. Des ressources de référence sur les médicaments les plus utilisées au monde. Elle est régulièrement utilisée par le grand public, les éducateurs, les pharmaciens, les pharmacologues, les chimistes médicaux, les chercheurs de l'industrie pharmaceutique (Wishart, 2018).

- **BioGraph**

C'est une application web qui permet d'interroger, de visualiser et d'analyser des données biologiques appartenant à plusieurs sources disponibles en ligne. BioGraph est construit sur notre base de données graphique précédemment développée appelée BioGraphDB, qui intègre et stocke des ressources biologiques hétérogènes et les rend accessibles au moyen d'une structure commune et d'un langage d'interrogation unique. BioGraph met en œuvre des technologies de pointe et fournit scénarios bio-informatiques précompilés, ainsi que la possibilité d'effectuer des requêtes personnalisées et d'obtenir une visualisation interactive et dynamique des résultats (Messina et al, 2018).

I . 8.5. Différence entre base de données et banque de données

Une banque de données est un ensemble de fichiers textes sans relation entre eux (on parle de fichiers « plat »).

Une base de données est un ensemble de relations entre des données gérées avec un système de gestion de base de données (SGBD) et interrogeable par SQL (Structure Query Langage) (Deléage et Gouy, 2013).

Tableau 2 : Les banques de données biologiques (Coutouly et al., 2012).

Nom	Commentaire	Adresse
Portails		
NCBI (National Center for Biotechnology Information)	Portail américain de bioinformatique. Banques de séquences nucléotidiques, BLAST banques génomiques. PubMed	http://www.ncbi.nlm.nih.gov
EBI (European Bioinformatics Institute)	Portail européen de bioinformatique. Banques de séquences nucléotidiques et protéiques, FASTA, ClustalW.	http://www.ebi.ac.uk

KEGG (Kyoto Encyclopedia of Genes and Genomes)	Portail japonais de bioinformatique. Voies métaboliques, banque de biomolécules	http://www.genome.ad.jp/kegg
EXPASY (Expert Protein Analysis System)	Portail de l'institut Suisse de bioinformatique. Protéines, enzymes, structures 3D	http://www.expasy.ch
Infobiogen	Portail français de bioinformatique. Annuaire des outils, documentation	http://www.infobiogen.fr
Bioweb Pasteur	Portail de bioinformatique de l'Institut Pasteur. Logiciels de bioinformatique en ligne	http://Bioweb.Pasteur.fr
Protéine		
Swiss – prot	Séquences bien annotées et faiblement redondantes	http://ca.expasy.org/sprot
PIR (Protein Information Resource)	Séquences et outils d'analyses	http://pir.georgetown.
TrEMBL	Séquences déduites à partir des séquences codantes des banques nucleotidiques EMBL / GenBank / DDBJ	http://www.ebi.ec.uk/trembl
Uniprot	Consortium regroupant Swiss - Prot - TrEMBL – PIR	http://www.ebi.uniprot.org
PDB (Protein Data Bank)	Structure 3D des protéines au format PDB	http://www.rcsb.org
Banque de données spécialisées		
Mapviewever	Génomes complets, localisation des gènes	http://www.ncbi.nlm.nih.gov/mapview
OMIM(Online Mendelian Inheritance in Man)	Informations sur les maladies génétiques chez l'homme	http://www3.ncbi.nlm.nih.gov/Omim
Orphanet	Base de données sur les maladies rares	http://orpha.net
VectorDB	Base de données des vecteurs utilisés en biologie	http://seq.yeastgenome.org/vectordb
Bibliographie		
PubMed – Medline	Articles , publications en biologie et en médecine	http://www.ncbi.nlm.nih.gov/

Connectsciences	Service du CNRS. Articles, thèses...	http://connectsciences/inist.fr
Google Scholar	Publications scientifiques avec Google	http://scholar.google.com

I .9.Les algorithmes utilisés en bio-informatique

La bio-informatique, ou biologie informatique, est un domaine d'étude polyvalent permettant de comprendre les données biologiques à l'aide d'un logiciel de traitement de l'information. L'importance de ce nouveau champ d'investigation se développera au fur et à mesure que nous continuerons à produire et à intégrer d'énormes quantités de données génomiques, protéiques et autres. Un domaine d'étude fascinant de la bio-informatique est l'application et le développement des algorithmes d'apprentissage automatique, pour expliquer les difficultés biologiques. Les informaticiens et les statisticiens élaborent des algorithmes et des méthodes pour le stockage, la gestion, le traitement et l'étude des données biologiques (**Raghuvanshi et al., 2020**).

9.1 L'algorithme

Un algorithme est une séquence d'instructions non ambiguës permettant de résoudre un problème, c'est-à-dire d'obtenir une sortie requise pour toute entrée légitime en un temps fini. Il donne une description illustrative de la relation entre le problème, l'algorithme, l'entrée et la sortie d'un algorithme et les données. C'est la complexité temporelle qui mesure l'efficacité des algorithmes. L'algorithme d'Euclide est plus efficace que l'algorithme naïf parce qu'il a une faible complexité temporelle (**Raghuvanshi et al., 2020**).



Du problème vers l'algorithme

9.2 Alignement de séquences

Un alignement de séquences est une façon d'arranger les séquences d'ADN, d'ARN ou de protéines pour d'identifier les régions de similarité qui peuvent être la conséquence des

relations fonctionnelles, structurelles ou évolutives entre les séquences alignées de nucléotides ou d'acides aminés sont généralement représentées sous forme de lignes dans une matrice. Les alignements de séquences sont également utilisés pour séquences non biologiques, telles que celles présentes dans le langage naturel ou dans les données financières.

En bio-informatique, l'alignement de deux ou plusieurs séquences d'ADN ou de protéines est une tâche très courante. Cela s'explique par le fait qu'un objectif important de la génomique est de déterminer si une séquence particulière est identique à une autre. L'alignement des séquences peut donc être pour comparer un gène séquencé dont la fonction est inconnue avec des séquences similaires provenant d'une base de données afin d'identifier la fonction du gène. (Essie et al., 2016).

9.3 Les types d'alignement

9.3.1 L'alignement global

Alignement de deux séquences sur la totalité de leur longueur en tenant compte de tous les résidus. Si les longueurs des séquences sont différentes, des insertions / délétions sont introduites pour aligner les deux extrémités des deux séquences. Il permet de mesurer le degré de similitude entre deux séquences connues (figure 12) (Coutouly et al., 2012).

Les alignements de séquences multiples (ASM) jouent un rôle primordial dans plusieurs domaines de la bio-informatique moderne, tels que la prédiction de la structure 3D, la fonction des protéines, l'inférence de la phylogénie, la fonction moléculaire, les interactions intermoléculaires et de nombreuses autres tâches courantes dans l'analyse des séquences. Un grand nombre d'algorithmes et de techniques innovantes ont été proposées pour améliorer la qualité des ASM, mais aucun des outils ASM ne peut reconstruire des alignements précis pour tous les types d'ensembles de données. (Babar et al., 2014)

□ Les diagrammes en points (Dot-Plot)

En bio-informatique, l'utilisation d'une matrice de similarité, appelée "dot plot", est une façon de visualiser la similarité entre deux séquences d'acides nucléiques ou de protéines.

Un diagramme de points peut être défini comme une méthode graphique qui permet de comparer de deux séquences biologiques afin d'identifier les régions de étroite similarité entre elles. En d'autres termes, une dot plot est une matrice utilisée pour détecter visuellement les alignements (figure 10). Il a été introduit par Gibbs et McIntyre (1970) et constitue l'un

des moyens les plus anciens et les plus simples de comparer deux séquences (Essie et al, 2016).

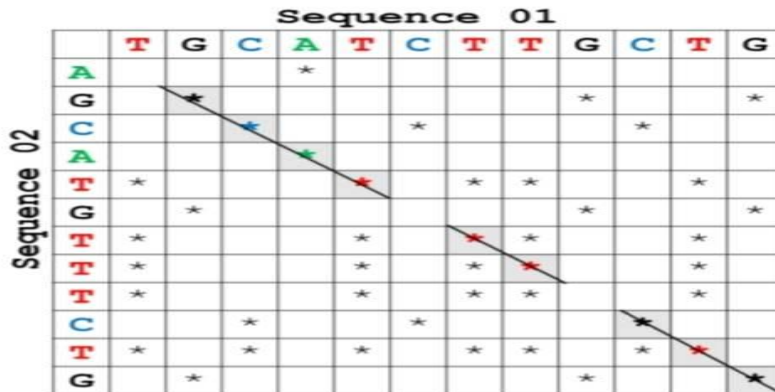


Figure 10 : Représentation d’une matrice de similarité (Dot-Plot) pour deux séquences d’ADN (Diniz et Canduri, 2017).

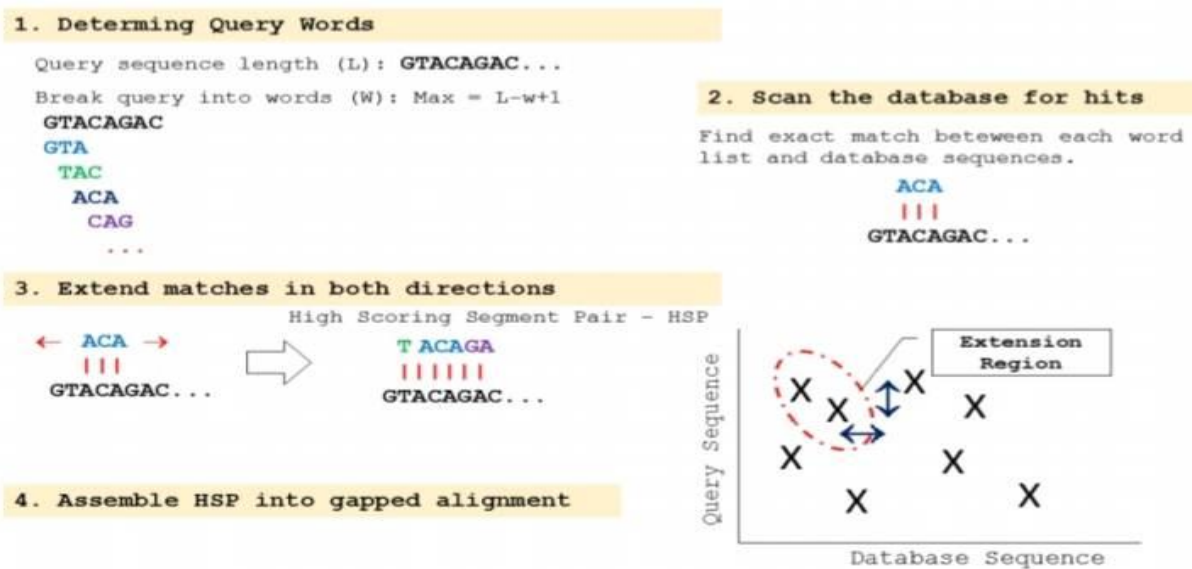


Figure 11 : Processus de l’opération BLAST (Diniz et Canduri, 2017).

9.3.2 L'alignement local

Alignement de deux séquences portant sur des régions isolées et permettant de trouver des séquences qui ont un haut degré de similitude (figure 12). Cette propriété en fait un outil idéal, rapide et efficace, de recherche dans les bases de données en comportant une séquence inconnue avec les séquences de la banque (Coutouly et al., 2012).

Dans cette approche, les algorithmes de programmation dynamique, l'analyse de la matrice de points et la méthode des k-tuple sont mis en évidence. La méthode de programmation

dynamique est basée sur le principe d'optimalité de Bellman qui propose que la solution à des problèmes complexes soit résolue par ses différents sous-problèmes.

Cette méthodologie peut être appliquée pour produire des alignements globaux et locaux grâce aux algorithmes de Needleman-Wunsch et de Smith-Waterman, respectivement. Pour l'alignement, un système de notation est nécessaire pour les correspondances et les non-concordances, pour les acides aminés ou les nucléotides, et une valeur de pénalité pour les lacunes. Ainsi, l'algorithme calculera l'alignement optimal entre les séquences.

□ **BLAST (Basic Local Alignment Search Tool)**

BLAST (Basic Local Alignment Search Tool) est un algorithme d'alignement local spécifique dérivé de l'algorithme de Smith-Waterman qui présente un score d'alignement maximal de deux séquences. Outre la programmation dynamique découlant de l'algorithme susmentionné, BLAST utilise une heuristique basée sur la méthode k-tuple pour rechercher les séquences dans la base de données (figure 11) (Diniz et Canduri, 2017).

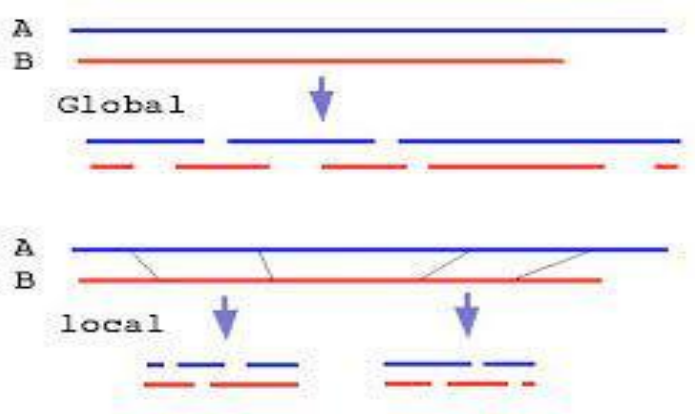
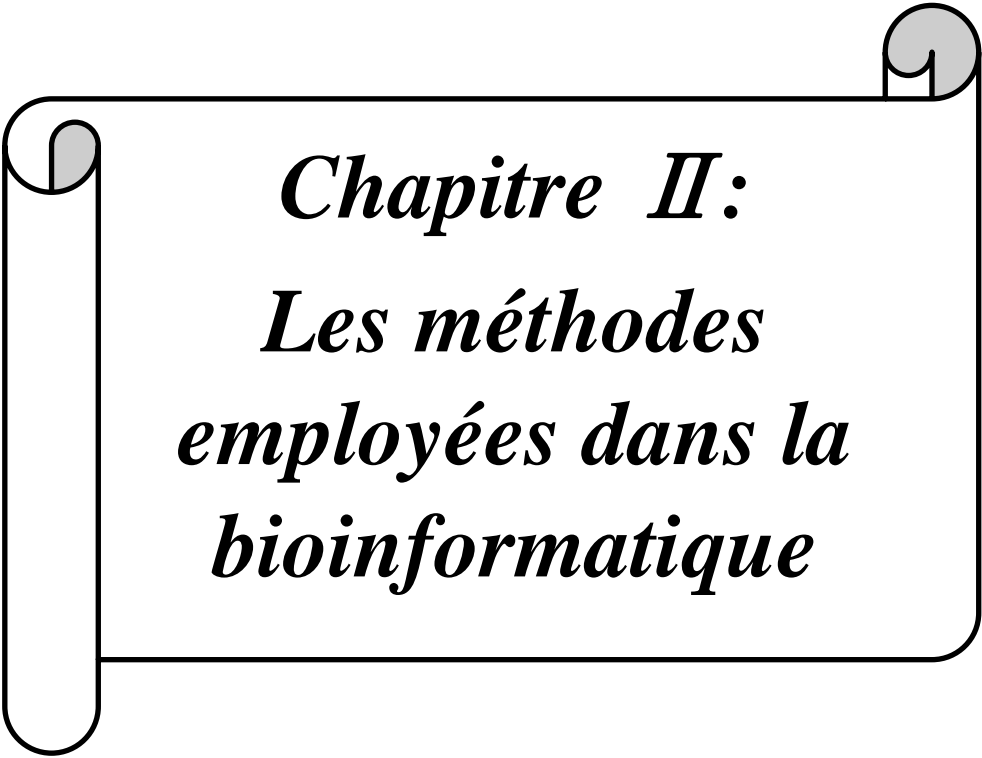


Figure 12 : Alignement global et local (Coutouly et al., 2012).



*Chapitre II:
Les méthodes
employées dans la
bioinformatique*

La bio-informatique est une discipline à l'interface de la biologie et de l'informatique qui recouvre l'ensemble des technologies et des méthodes tels que le séquençage de l'ADN, le docking moléculaire, la modélisation moléculaire et le drug design permettant de collecter, de stocké, d'analyser et d'interpréter les données biologiques.

II.1. Le séquençage de l'ADN

Le séquençage de l'ADN a été inventé dans la deuxième moitié des années 1970. Deux méthodes ont été développées indépendamment, l'une par l'équipe de Walter Gilbert, aux États Unis, et l'autre par celle de Frederick Sanger en Grande-Bretagne. Ces deux méthodes sont fondées sur des principes diamétralement opposés : l'approche de Sanger est une méthode par synthèse enzymatique sélective, tandis que celle de Maxam et Gilbert est une méthode par dégradation chimique sélective (El fahim et Ennaji, 2007).

II.1.1. Les techniques de séquençage de l'ADN

Les deux premières techniques de séquençage de l'ADN, celle de Maxam-Gilbert et celle de Sanger ont été décrites en 1977. La technique de Sanger a révolutionné le monde de la biologie moléculaire en permettant de décrypter différents génomes, tel que celui du génome humain complètement déchiffré en 2006 ou d'autres génomes, le génome bactérien, par exemple, le premier d'entre eux étant celui d'*Haemophilus influenzae*, complètement décrit en 1995. La méthode de Sanger continue d'être la méthode la plus employée dans le monde, à l'heure actuelle (Lamoril et al., 2008).

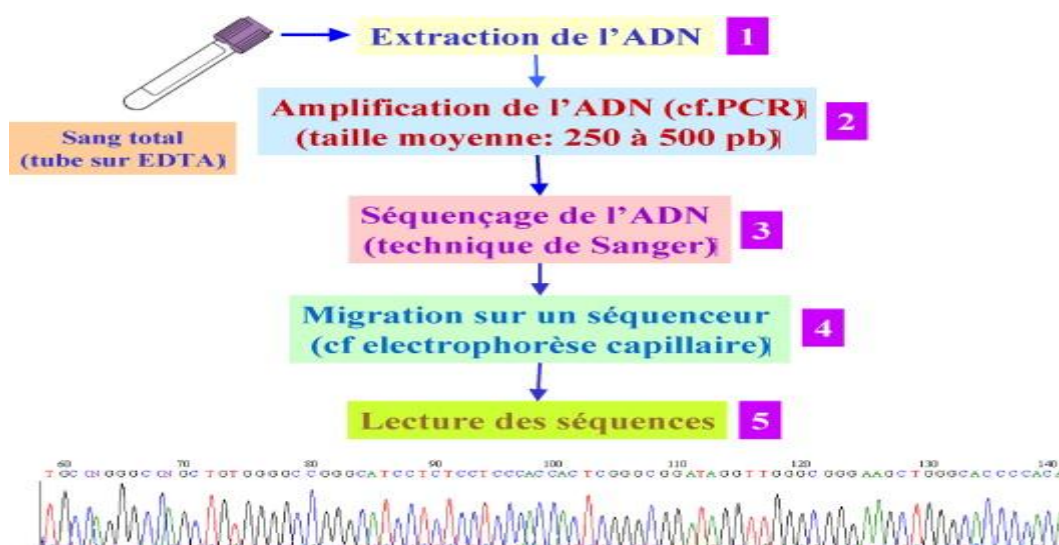


Figure 13 : Les techniques de séquençage de l'ADN (Lamoril et al., 2008).

II.1.2. Les applications du séquençage

- ❖ Diagnostic et traitement de nombreuses maladies humaines (exemples : cancers, maladies infectieuses, maladies héréditaires...)
- ❖ Informations sur le génome (structure, fonction, évolution) et étude des variations du génome (polymorphismes, insertions, délétions) réarrangement de gènes, variation du nombre de copies de gènes, duplication.
- ❖ Variants génétiques associés à une pathologie (par exemple, le diabète).
- ❖ Analyse de méthylation du génome (études épigénétiques et méthylome).
- ❖ Analyse microbiologique (identification d'espèces, taxonomie, études épidémiologiques, génotypage à but pronostique et/ou thérapeutique).
- ❖ Tests de paternité et médecine légale.
- ❖ Police scientifique.
- ❖ Pharmacogénétique.
- ❖ Études anthropologiques (**Lamoril et al., 2008**).

II.2. Docking moléculaire

Le docking moléculaire (Molecular Docking, en anglais), ou encore (Ancrage ou Amarrage moléculaire), est une technique qui consiste à faire interagir une petite molécule organique, appelée "ligand", dans le site actif d'une macromolécule (Protéine), appelée « récepteur »(figure 14) (**Hammd, 2020**). Le docking moléculaire vise à optimiser la conformation et la direction relative entre la protéine et le ligand (**Das et al.,2020**)

L'amarrage moléculaire est une méthode in-silico qui permet de prédire de petites molécules ou de ligands dans le site actif de leur protéine cible (récepteur). Elle est principalement utilisée pour estimer avec précision les modes de liaison les plus favorables et les bio-affinités des ligands avec leur récepteur, Actuellement, elle a été largement appliquée au criblage virtuel pour l'optimisation des composés principaux (**Surabhi et al., 2018**).

Le docking moléculaire a pour objectif essentiel de prédire la conformation définie par la position la plus favorable et l'orientation relative la plus favorable du ligand au sein de son récepteur au niveau du site actif (**Soudani et al., 2022**).

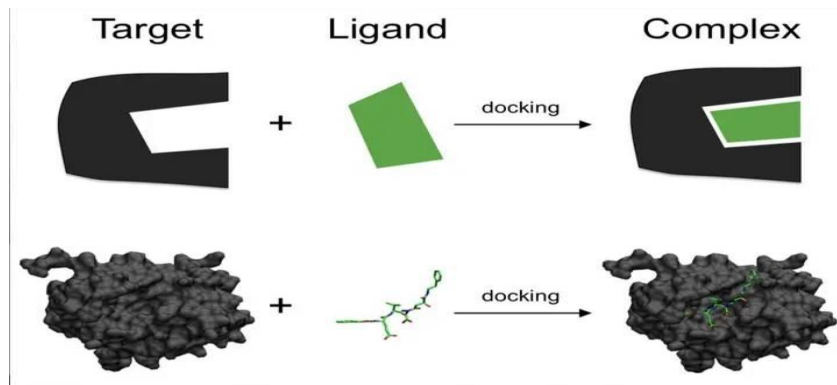


Figure14 : Interaction protéine- ligand lors du docking moléculaire (Poudyal, 2021).

➤ Principe du docking moléculaire

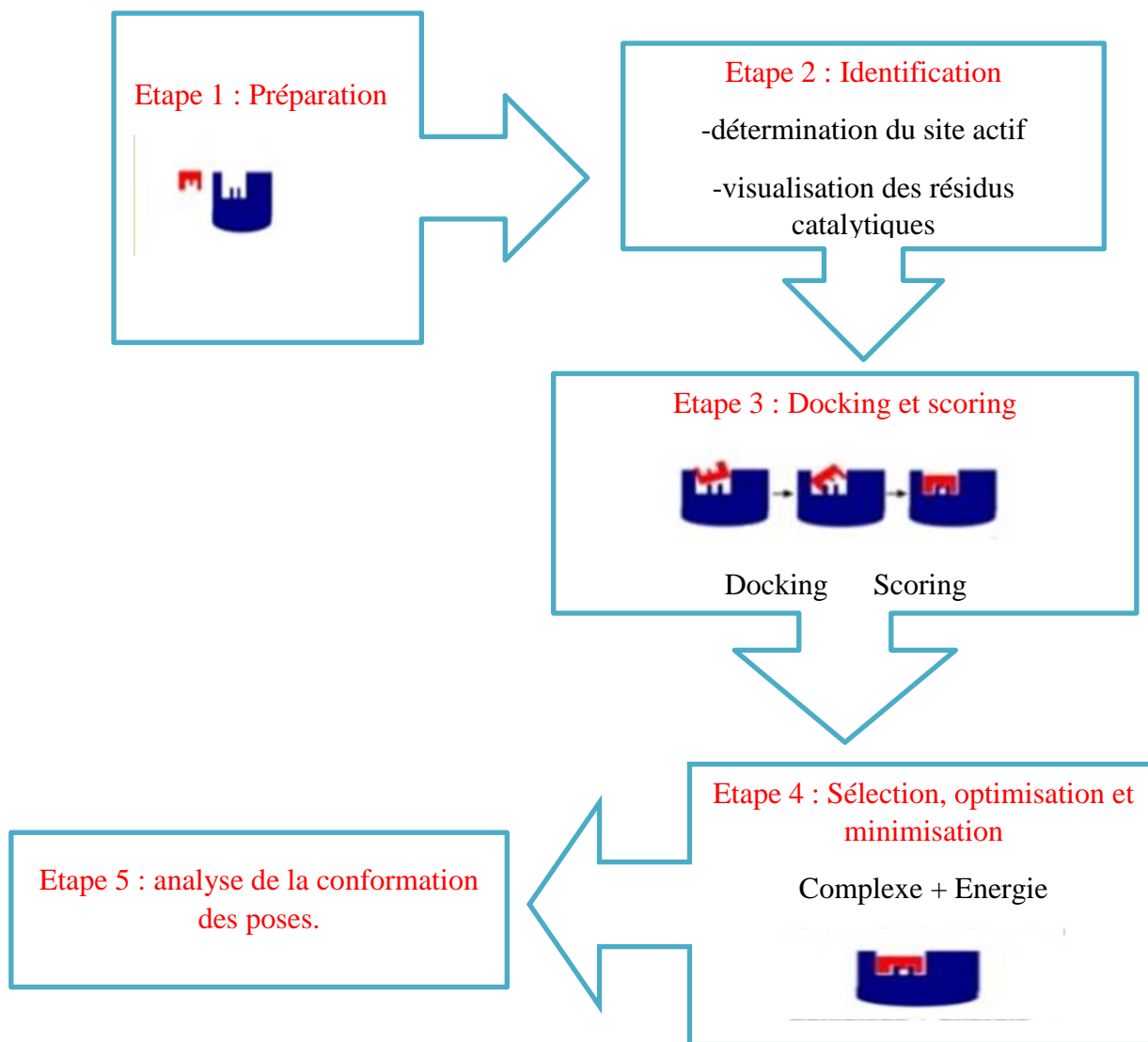


Figure 15 : Principe général du docking moléculaire (**Bouchagra, 2018**).

II.2.1. Les types du docking moléculaire

✓ **Docking rigide**

Le docking rigide, au cours duquel la protéine et le ligand sont tous deux traités comme entièrement rigides. Ainsi, seuls les degrés de liberté translationnels et rotationnels du ligand relativement au récepteur sont explorés. Cette simplification s'apparente à considérer un modèle de liaison de type « serrure-clé » où ni le ligand ni le récepteur ne subissent de réarrangements conformationnels suite à leur interaction.

Le docking rigide est généralement employé pour l'amarrage entre deux macromolécules (docking protéine-protéine ou protéine/acide nucléique structuré) (**Chevrollier, 2019**).

✓ **Docking semi-flexible**

Le docking semi-flexible est "asymétrique" et généralement utilisé pour l'amarrage protéine-ligand, le ligand étant considéré comme flexible, la protéine gardée rigide (**Bouchagra, 2018**).

✓ **Docking flexible**

Le ligand et le récepteur sont conformationnellement flexibles, à chaque rotation, l'occupation des cellules de la surface et l'énergie sont calculées, ensuite, la pose la plus optimale est sélectionnée (**Chaudhary et Mishra, 2016**).

II.2.2. Les étapes de docking moléculaire

□ **Préparation de la protéine**

A partir de la banque de données structurales (PDB) télécharger la structure 3D de la protéine. Ensuite, la structure téléchargée doit être prétraitée. De l'élimination des molécules d'eau, la stabilisation des charges, les résidus manquants par remplissage, l'ajout de chaînes latérales et la génération d'atomes d'hydrogène (**Bhagat et al., 2021**).

□ **Prédiction du site actif**

Après la préparation de la protéine, le site actif de la protéine doit être prédit, le récepteur peut posséder de nombreux sites actifs, il suffit de choisir celui qui nous intéresse, la plupart du temps, les molécules d'eau et les hétéroatomes sont enlevés s'ils sont présents.

Préparation du ligand

Les ligands peuvent être récupérés de plusieurs bases de données telles que ZINC, Pub Chem ou peuvent être esquissés à l'aide de l'outil Chem sketch. Lors du choix du ligand, la règle des 5 de LIPINSKY doit être respectée.

La règle des 5 de LIPINSKY permet de distinguer les ligands qui ne ressemblent pas à des médicaments de ceux qui ressemblent à des médicaments :

- + Moins de cinq donneurs de liaisons hydrogène.
- + Moins de dix accepteurs de liaisons hydrogène.
- + Masse moléculaire inférieure à 500 Da.
- + Haute lipophilie (exprimée par un Log P inférieur ou égal à 5) (**Chaudhary et Mishra, 2016**).

Docking

Les interactions entre le ligand et la protéine sont analysées, le meilleur score d'amarrage doit être sélectionné (**Bhagat et al.,2021**).

II.2.3. Les différents types d'interactions

Les forces d'interaction sont généralement réparties en quatre catégories :

- Forces électrostatiques : dipôle-dipôle, charge-dipôle et charge-charge.
- Forces électrodynamiques : interaction de Van der Waals.
- Forces stériques : causées par l'entropie.
- Forces liées aux solvants : liaison hydrogène et interactions hydrophobes (**Chaudhary et Mishra, 2016**).

II.2.4. Les programmes du docking moléculaire

Le tableau suivant représente les différents programmes du docking moléculaire

Tableau 3 : les principaux programmes de docking moléculaire (Karima, 2022).

Nom	Editeur	Site internet
AutoDock	Scripps	http://www.scripps.edu/mb/olson/doc/autodock/
Dock	UCSF	http://dock.compbio.ucsf.edu/
FlexX	BioSolveIT	http://www.biosolveit.de/FlexX/
Fred	OpenEyes	http://www.eyesopen.com/products/application/fred.html
Glide	Schrodinger	http://www.schorodinger.com/products/glide.html
Gold	CCDC	http://ccdc.cam.ac.uk/products/life_sciences/gold/
ICM	Molosoft	http://www.molsoft.com/products.html
LigandFit	Accelrys	http://www.accelrys.com/cerius2/c2ligandfit.html
Surflex	Biopharmics	http://www.biopharmics.com/products.html

II.2.5. Les applications du docking moléculaire

○ Applications de l'amarrage moléculaire dans le développement de médicaments

Le docking moléculaire est le plus souvent utilisé pour la découverte de médicaments, car la majorité des médicaments sont composés de minuscules composées organiques. Le docking peut être utilisé pour :

- **Identification des résultats**

Le docking moléculaire, associé à une fonction de score, permet de cribler rapidement de vastes bases de données de médicaments possibles in silico, afin de trouver des composés capables de se lier à une cible d'intérêt (Raval et Ganatra, 2022).

- **Optimisation des pistes**

L'amarrage moléculaire peut prédire une orientation optimisée du ligand sur sa cible. Il peut prédire différents modes de liaison du ligand dans le sillon de la molécule cible. Cela peut être

utilisé pour développer des candidats médicaments plus puissants, plus sélectifs et plus efficaces (Dar et Mir, 2017).

- **Remédiation**

En outre, l'amarrage protéine-ligand peut être utilisé pour prévoir quels contaminants sont dégradables par les enzymes. Il peut être utilisé pour déterminer le niveau de dégradation des contaminants par les enzymes. L'amarrage moléculaire peut être utilisé pour identifier les enzymes et leur mode d'action. Il peut également être utilisé pour déterminer les relations entre les protéines (Raval et Ganatra, 2022).

- **Prédiction de l'activité biologique (AC)**

Sur la base de leurs caractéristiques chimiques, un modèle Quantitatif structure Relationship (QSAR) validé peut être utilisé pour prédire l'activité biologique de nouvelles substances. En outre, les composés peuvent être modifiés afin d'accroître leur action. En règle générale, tous les composés nouvellement développés sont classés comme devant faire l'objet de tests expérimentaux sur la base des projections QSAR (Kumar, 2021).

- **Autres applications**

- Détermination des effets secondaires de l'utilisation d'une autre molécule.
- Conception de médicaments.
- Etude de la géométrie d'un complexe particulier.
- Etude des interactions médicament – ADN en établissant une corrélation entre la structure moléculaire des médicaments et leurs effets cytotoxiques.
- Démonstration de la faisabilité d'une réaction biochimique, car il est effectué avant toute investigation expérimentale.
- Caractérisation des petites molécules, comme principale application.
- Détermination de la spécificité d'un médicament proposé contre des protéines homologues.
- Identification des interactions protéines (Poudyal, 2021).

II.3. Modélisation moléculaire

La modélisation moléculaire est un ensemble de méthodes physiques théoriques et de techniques informatiques qui tentent de mimer virtuellement le comportement des molécules

(**Brillouet et al., 2010**). Elle est maintenant utilisée dans plusieurs domaines d'applications telles que la chimie, l'industrie pharmaceutique et la biologie (**Kettani et Mikou, 2008**), pour analyser les molécules et les systèmes moléculaires ou pour prédire les réactions moléculaires, chimiques et biochimiques (**Mokhtari, 2021**).

La modélisation moléculaire joue un rôle important dans l'optimisation des structures obtenues expérimentalement par cristallographie rayon X ou par résonance magnétique nucléaire (**Brillouet et al., 2010**). Cette technique a pour but de représenter, de prévoir la structure et la réactivité des molécules ou d'un système moléculaire.

Elle sert de pont entre la théorie et l'expérience pour :

1. Extraire des résultats pour un modèle particulier.
2. Comparer les résultats expérimentaux du système.
3. Comparer les prévisions théoriques du modèle.
4. Aider à comprendre et à interpréter les observations expérimentales.
5. Corrélation entre les détails microscopiques au niveau atomique et moléculaire et les propriétés macroscopiques.
6. Fournir des informations qui ne sont pas disponibles dans le cadre d'expériences réelles (**Mokhtari, 2021**).

La modélisation moléculaire comprend deux axes : le graphisme moléculaire et la chimie Computationnelle.

- ✓ Le graphisme moléculaire ou infographie moléculaire pour le design et la représentation en 3 dimensions des modèles avec différentes options : couleurs, volume, prescriptive, zoom, rotation... autant d'outils qui permettent de mettre en évidence les propriétés physico-chimiques du modèle (zone hydrophobe, polaire, chargée, site actif...)
- ✓ La chimie « Computationnelle » : il s'agit de reproduire les propriétés physico-chimiques en utilisant différents algorithmes suivant la taille du système. Ainsi les calculs ab initio qui tiennent compte des propriétés des électrons ne peuvent être menés que sur des petites molécules (100 atomes maximum) compte tenu du temps de calcul long. De même, les calculs semi empiriques, bien que moins précis, ne permettent de traiter que les molécules

de près de 1000 atomes. Aussi, pour réduire le temps de calcul sur les Marco systèmes biologique (protéines, ADN, complexes) (Kettani et Mikou, 2008).

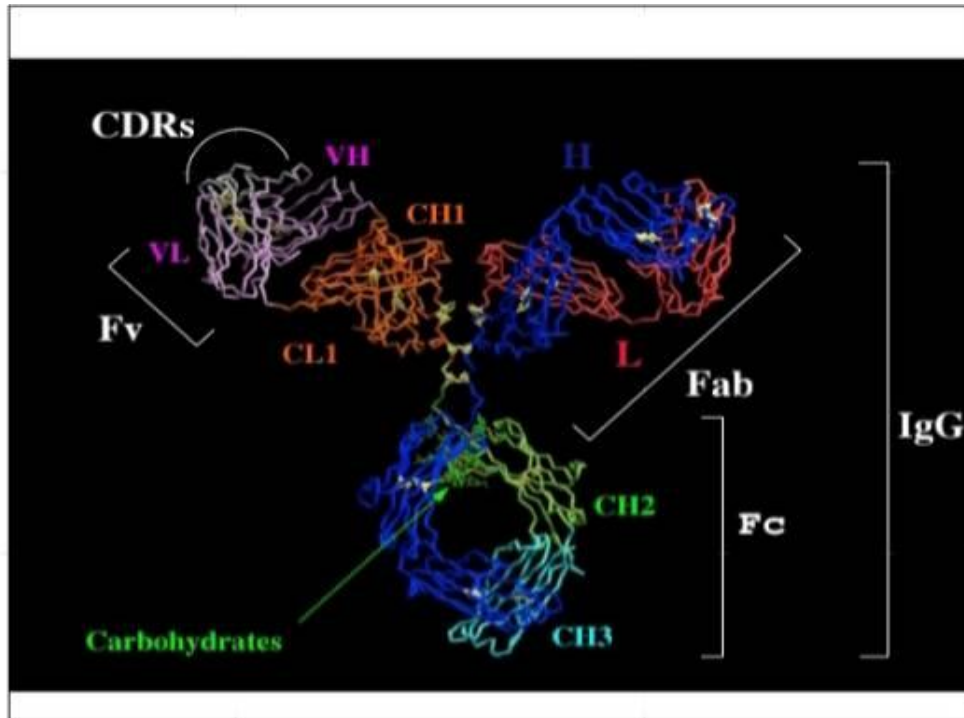


Figure 16 : Représentation du repliement d'une immunoglobuline (pellequer, 1999).

II .3.1 Les méthodes de la modélisation moléculaires

Les modélisations moléculaires comprennent trois méthodes : Les méthodes quantiques, la mécanique moléculaire et la dynamique moléculaire.

II .3.1.1 Mécanique Quantique

La mécanique quantique (MQ) a pour objectif la détermination de l'énergie, ainsi que la distribution électronique d'un système moléculaire. Elle consiste à résoudre l'équation de Schrödinger, qui impliquent un opérateur linéaire appelé hamiltonien, qui décrit l'évolution des états du système physique dans le temps (Hammad, 2020).

II .3.1.2 Mécanique Moléculaire

La mécanique moléculaire (MM) ou mécanique newtonienne, est une méthode empirique qui permet d'obtenir des résultats de géométries et d'énergies moléculaires, en se basant sur la mécanique classique, en utilisant comme outil, le « Champ de force » (Hammad, 2020) est le moyen le plus simple et le plus rapide d'évaluer les systèmes moléculaires.

II.3.1.3 Dynamique moléculaire

La dynamique moléculaire (DM) est l'étude de la façon dont les molécules se déplacent, se déforment et interagissent au fil du temps. La prédiction ou l'interprétation de ces changements est essentielle en chimie, en physique, en biologie, en ingénierie et dans d'autres domaines (**Mokhtari, 2021**). Réalisée par un ordinateur, durant laquelle le système subit des changements conformationnels et cinétiques, décrivant l'évolution au cours du temps des positions (trajectoires) atomiques (**Hammad, 2020**).

❖ Calcul de la dynamique moléculaire

La dynamique moléculaire est généralement simulée selon les étapes suivantes :

- **Minimisation**

Cette étape consiste à trouver l'énergie minimale globale par rapport à la position des atomes des chaînes latérales qui représente la géométrie des arrangements particuliers d'atomes dans laquelle la force d'attraction nette sur chaque atome atteint un maximum.

- **Chauffer le système**

Dans la phase de chauffage, les vitesses initiales sont assignées à chaque atome du système pendant la minimisation de l'énergie et les équations du mouvement de Newton qui représentent l'évolution temporelle du système sont intégrées numériquement. À de courts intervalles prédéfinis, de nouvelles vitesses sont attribuées correspondant à une température légèrement plus élevée et la simulation se poursuit jusqu'à ce que la température souhaitée soit atteinte.

- **Equilibrage**

L'étape d'équilibrage est utilisée pour équilibrer les énergies cinétique et potentielle, c'est-à-dire distribuer l'énergie cinétique "pompée" dans le système. Dans la simulation explicite du solvant, les positions des protéines sont fixes et les eaux se déplacent en conséquence.

Une fois le solvant équilibré, les contraintes sur la protéine peuvent être supprimées et l'ensemble du système (protéine-solvant) peut évoluer dans le temps (**Mokhtari, 2021**).

- **Production**

Dans cette étape qui suit la phase d'équilibration, les conformations et les énergies du système, sont sauvegardées à des intervalles réguliers et prédéfinies, qui seront utilisées par la suite, pour calculer les caractéristiques relatives au système moléculaire.

- **Interprétation des résultats / Analyses usuelles en dynamique moléculaire**

Différents outils, sont utilisés pour analyser les résultats obtenus lors de la dynamique moléculaire, afin d'évaluer la pertinence des mouvements et interactions intramoléculaire, ainsi que de statuer sur la stabilité du système moléculaire (**Hammad, 2020**).

II .3.2 Les applications de la modélisation moléculaire

- La conception de nouvelles molécules.
- Le drug design (**Brillouet et al., 2010**).
- Le développement de médicaments : Elle est utilisée pour évaluer les inconvénients potentiels produits par des relations avec d'autres protéines, telles que les protéases, le cytochrome P450 et autres. (**Raval et Ganatra, 2022**).

II .4. Drug design

Drug design, également appelée conception rationnelle de médicaments, est le processus inventif de recherche de nouveaux médicaments basé sur la connaissance d'une cible biologique.

La conception de médicaments définit la conception de molécules dont la forme et la charge sont complémentaires à la cible biomoléculaire avec laquelle elles interagissent et qui se lieront donc à elle. Le médicament est le plus souvent une petite molécule organique qui active ou inhibe la fonction d'une biomolécule telle qu'une protéine, ce qui se traduit par un bénéfice thérapeutique pour le patient (**Takagi, 2021**).

II .4.1. Les étapes de drug design

- **Étape 1 : Identification de la cible**

Une cible est un terme général qui peut s'appliquer à toute une série d'entités biologiques, comme les protéines, les gènes et l'ARN. Une bonne cible doit être efficace, sûre, répondre aux besoins cliniques et commerciaux et surtout être "médicamentable". Une cible

"médicamentable" est accessible à la molécule médicamenteuse présumée, qu'il s'agisse d'une petite molécule ou d'une molécule biologique plus importante dont la liaison déclenche une réponse biologique qui peut être mesurée à la fois *in vitro* et *in vivo*. On sait aujourd'hui que certaines classes de cibles se prêtent mieux à la découverte de petites molécules, par exemple les récepteurs couplés aux protéines G (RCPG), tandis que les anticorps bloquent plus facilement les interactions protéine/protéine. Une bonne identification et validation des cibles permet de renforcer la confiance dans la relation entre la cible et la maladie et nous permet d'explorer si la modulation de la cible conduira à des effets secondaires basés sur le mécanisme (Shultana, 2021).

▪ **Étape 2 : Validation de la cible**

La validation de nouvelles cibles est la base de l'exploration de nouveaux médicaments et l'étape initiale de la découverte de médicaments. Les chercheurs peuvent identifier les composés qui ont un effet sur la cible sélectionnée.

Des tests sont effectués pour confirmer que l'interaction avec la cible des médicaments est associée à un changement souhaité dans le comportement de la cellule malade. Les chercheurs analysent et comparent chaque cible de médicament à d'autres en fonction de leur association avec une maladie spécifique et de leur capacité à réguler les composés biologiques et chimiques dans l'organisme.

Le processus de validation des cibles peut comprendre six étapes :

- ✓ Découverte d'une molécule intéressante.
- ✓ Évaluation de son potentiel en tant que cible.
- ✓ Conception d'un essai biologique pour mesurer l'activité biologique.
- ✓ Construction d'un criblage à haut rendement.
- ✓ Effectuer un criblage pour trouver des résultats positifs.
- ✓ Évaluation des résultats (Kumar Singh et al., 2021).

▪ **Étape 3 : Identification des pistes**

Les pistes peuvent être identifiées à l'aide de techniques telles que la conception basée sur la structure. À ce stade, la structure de la protéine cible en complexe avec la molécule principale peut s'avérer extrêmement utile pour suggérer des moyens d'améliorer l'affinité de la molécule principale pour la cible. Les molécules utilisées dans ce cas peuvent être loin d'être parfaites

et doivent donc être optimisées afin d'accroître leur affinité pour les sites cibles. L'optimisation peut être obtenue en modifiant leurs caractéristiques structurales (Naikwadi et al., 2021).

▪ **Étape 4 : Optimisation de la piste**

L'optimisation de la piste est un processus complexe et non linéaire qui consiste à d'affiner la structure chimique de la molécule confirmée pour d'améliorer ses caractéristiques médicamenteuses dans le but de produire un candidat-médicament préclinique.

Le processus d'optimisation est très interactif. Les pistes sont évaluées lors d'essais pharmacologiques afin de déterminer leur "similitude avec un médicament". Généralement, un ou plusieurs résultats confirmés sont évalués dans les essais secondaires, ainsi qu'un ensemble de composés apparentés appelés analogues, sont synthétisés et criblés (Saha et Hait, 2011).

▪ **Étape 5 : Développement préclinique et clinique**

Le développement préclinique comprend les éléments suivants : élaboration d'une synthèse à grande échelle, études d'innocuité sur les animaux, tests de cancérogénicité, administration du médicament, études d'élimination et de métabolisme, expériences de formulation du médicament, études d'évaluation de la dose chez l'animal, Métabolisme, expériences de formulation de médicaments, études de dosage chez animaux (prakash et Devangi, 2010).

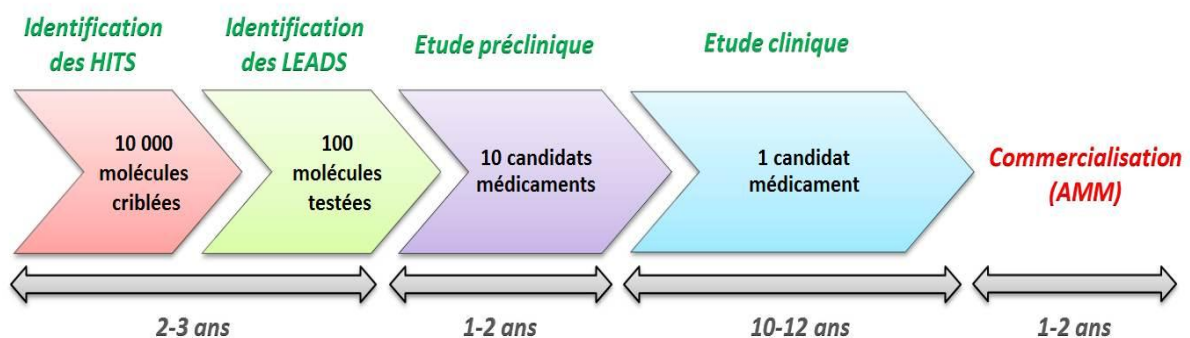


Figure 17 : Les phases de découverte et développement d'un médicament (Journot et al., 2017).

- **Les essais cliniques pharmaceutiques sont généralement classés en quatre phases**
- **Phase 0** : Une désignation récente pour les essais exploratoires, les premiers chez l'homme, destinés à accélérer le développement d'agents thérapeutiques prometteurs en établissant rapidement si l'agent se comporte chez les sujets humains comme prévu par les études précliniques.
 - **Phase 1** : Un petit groupe de volontaires en bonne santé (20-80) est sélectionné pour évaluer l'innocuité, la tolérabilité et l'efficacité de l'agent sélectionné pour évaluer la sécurité, la pharmacocinétique et la pharmacodynamie d'un traitement.
 - **Phase 2** : Réalisée sur des groupes plus importants (20-300), elle est conçue pour évaluer l'activité du traitement, les effets secondaires.
 - **Phase 3** : Essais contrôlés randomisés sur de grands groupes de patients visant à évaluer définitivement l'efficacité de la nouvelle thérapie par rapport à la thérapie standard.
 - **Phase 4** : Suivi de la sécurité après le lancement et assistance technique continue d'un médicament peuvent être mandatés ou initiés par l'entreprise pharmaceutique afin de détecter les effets indésirables rares ou à long terme sur un grand nombre de patients (prakash et Devangi, 2010).

II.4.2. Les études ADMET

ADMET signifie Absorption, Distribution, Métabolisme, Excrétion et Toxicologie des médicaments. Cette méthode est également appelée la modélisation pharmacocinétique basée sur la physiologie. Elle est utilisée dans recherche et développement pharmaceutiques et dans l'évaluation pour prédire l'ADMET d'un composé chez l'homme. Les paramètres ADMET influencent la performance et l'activité pharmacologique du composé en tant que médicament en fonction de la cinétique de leur exposition aux tissus et la façon dont ils seront traités ou acceptés par l'organisme, ce qui donne un aperçu de la nature d'un composé pharmaceutique dans un organisme.

Cette approche joue un rôle crucial dans la découverte et le développement de médicaments, car elle permet de sélectionner les composés dès les premières étapes de la recherche et réduit considérablement l'attrition tardive des médicaments au cours des essais préclinique/clinique, s'avérant ainsi une technique efficace en termes de coûts et de temps (Agrawal, 2013).

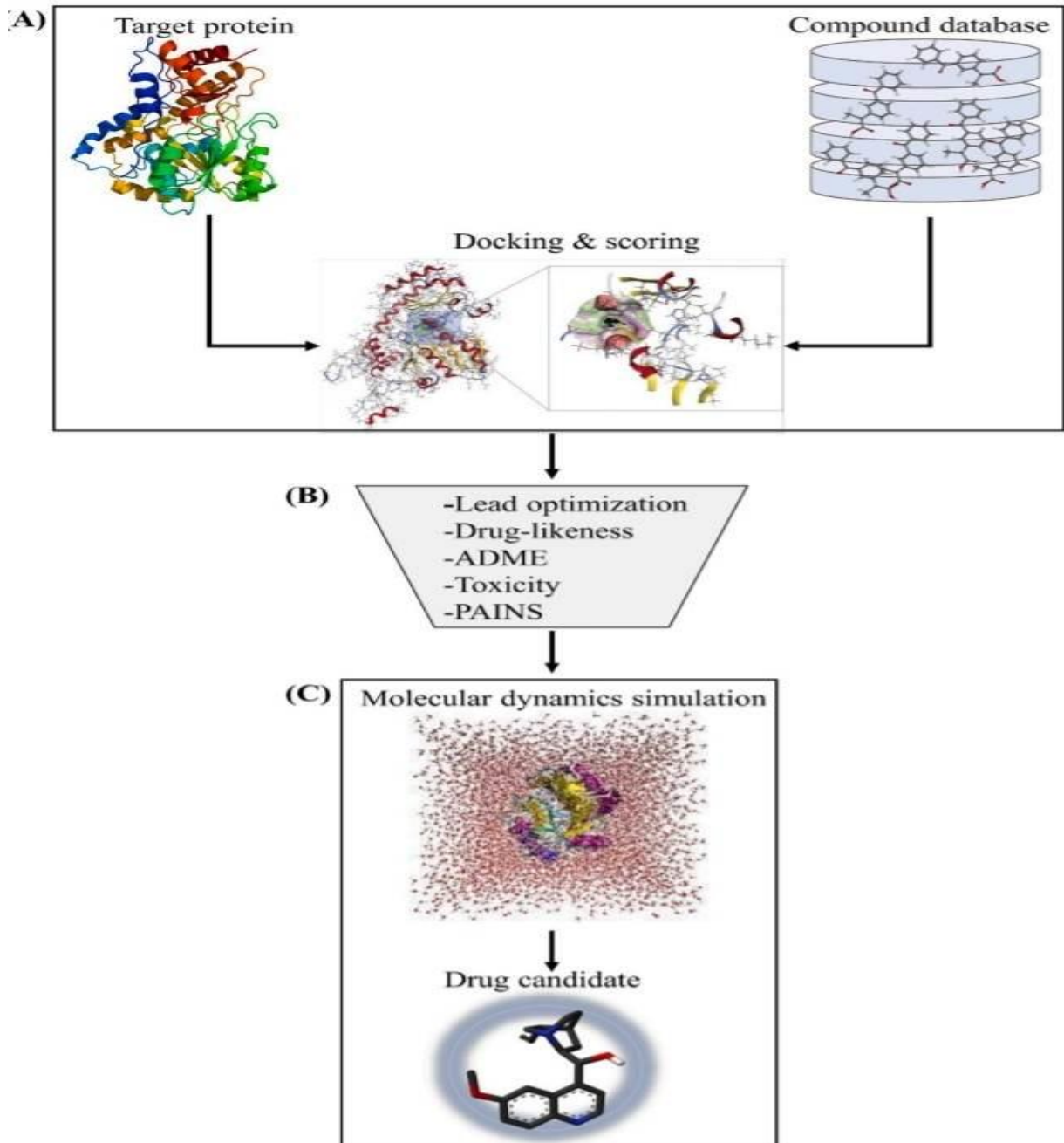


Figure 18 : Flux de travail pour la découverte de médicaments basés sur la structure.

(A) Les composés sont ancrés sur une protéine cible et sont classés par ordre de priorité en fonction de leur affinité et de leur mode de liaison. **(B)** Les composés retenus sont optimisés

sur le plan structurel et filtrés en fonction de leurs propriétés physiochimiques en déterminant l'ADME-Tox et d'autres propriétés pharmacologiques. (C) Des simulations supplémentaires de la dynamique moléculaire sont effectuées pour affiner les produits chimiques conçus et enfin les médicaments candidats sont sélectionnés. Ces candidats sont ensuite soumis à une validation expérimentale. (Shaker, 2021).

II.4.3. Les Application de drug design

- **Contrôler l'expression des gènes**

- Étude des réseaux de régulation.
- Découverte de médicaments, mécanisme d'action
- Diagnostic, diagnostic des tumeurs

- **Hybridations de l'ADN génomique**

- Exploration de la diversité microbienne.
- Comparaisons génomiques globales, évolution du génome
- Identification des sites de fixation de l'ADN.
- Diagnostic, diagnostic des tumeurs

- **Pharmacogénomique**

La pharmacogénomique est l'étude du rôle du génome dans la réponse aux médicaments. Son nom reflète la combinaison de la pharmacologie et de la génomique. La pharmacogénomique analyse la manière dont la génétique d'un individu affecte sa réponse aux médicaments.

La liste ci-dessous présente quelques-unes des applications les plus connues de la pharmacogénomique :

- Améliorer la sécurité des médicaments et réduire les effets indésirables des médicaments.
- Adapter les traitements à la prédisposition génétique unique des patients, en identifiant le dosage optimal.
- Améliorer la découverte de médicaments ciblés sur les maladies humaines.
- Améliorer la preuve de principe pour les essais d'efficacité.

-La pharmacogénomique peut être appliquée à plusieurs domaines de la médecine, notamment la gestion de la douleur, la cardiologie, l'oncologie et la psychiatrie.

-La pharmacogénétique peut également être utilisée dans le traitement du cancer, des troubles cardiovasculaires, des troubles psychiatriques, etc. **(Senthil et al, 2023)**.



Chapitre III:
Analyse d'articles

Ces dernières années, des études d'amarrage moléculaire *in silico* sur des produits naturels ont permis des approches de criblage computationnel pour évaluer leur potentiel thérapeutique. Ces études utilisent des techniques bio-informatiques et peuvent être utilisées pour découvrir comment les médicaments candidats provoquent une activité thérapeutique en prédisant les interactions entre les médicaments et les protéines et en analysant l'influence sur les voies et les fonctions biologiques (**Koshak et koshak , 2020**).

Les essais cliniques randomisés restent l'étalon-or pour évaluer l'efficacité et l'innocuité des nouveaux médicaments. Ces essais complexes et coûteux peuvent cependant exposer de larges populations de patients à des thérapies non éprouvées pendant plusieurs années. Les difficultés de sélection, de recrutement, de suivi et de rétention des patients dans les études cliniques conventionnelles contribuent toutes à des taux d'échec importants. De plus, si les essais contrôlés codifiés sont conçus pour évaluer des médicaments grand public, ils sont moins adaptés pour tester de nouveaux traitements pour des maladies chroniques hétérogènes complexes ou des indications de maladies rares (**Moingeon et al., 2023**).

Pour ces raisons, les simulations « *in silico* » suscitent un vif intérêt en tant qu'approche complémentaire aux essais cliniques classiques dans l'évaluation de médicaments expérimentaux ou de dispositifs médicaux (**Moingeon et al., 2023**).

La maladie chronique est un terme général désignant des maladies qui ne constituent pas une infection et qui accumulent à long terme des dommages sous forme de maladie, telles que le diabète, les maladies cardiovasculaires, le cancer, les maladies respiratoires, etc. Elles sont progressivement devenues des problèmes majeurs de santé publique qui se caractérisent par une longue durée, généralement à progression lente et non transmissibles d'une personne à l'autre. La survenue et le développement des maladies chroniques sont liés à de nombreux facteurs tels que l'hérédité, l'environnement, les mauvais comportements et le mode de vie (**Wei Li et al., 2023**).

Avec les progrès d'une société vieillissante dans le monde, la fréquence des maladies chroniques ne cesse d'augmenter et il a été étudié que la mortalité causée par les maladies chroniques représente plus de 60 % des décès dans le monde. Près de 300 millions de personnes (moins de 65 ans occupés pour 50%) en Chine souffrent de maladies chroniques et la prévalence de la dyslipidémie, le diabète et l'hypertension sont de 40,4 %, 9,7 % et 25,2 %, respectivement, et maintiennent une tendance à la hausse d'une année sur l'autre. La prévalence de ces maladies dans les pays américains et européens semble être plus élevée

puisque 60 % des personnes atteintes d'une maladie chronique aux États-Unis et 56,3 % des personnes atteintes de maladies chroniques multiples en Suède ont été signalées (**Wei Li et al., 2023**).

Par ailleurs, la prévalence des maladies chroniques en Algérie a augmenté avec l'âge. Certaines personnes souffrent diverses maladies chroniques, les plus courants étaient l'hypertension artérielle, le diabète, les maladies cardiovasculaires, l'asthme et les maladies cardiaques articulations (**APS, 2021**).

De nos jours, les maladies chroniques principalement traitées avec des médicaments chimiques entraînent toujours une lourde charge financière et des effets secondaires potentiels. Il convient de noter qu'un nombre croissant d'enquêtes ont indiqué qu'un régime alimentaire contenant des substances bioactives naturelles pourrait être lié à un risque moindre de maladies chroniques (**Wei Li et al., 2023**).

III.1. Objectif de l'étude

Cette partie comprenant l'analyse de plusieurs articles relatifs au thème étudié a pour objectif de rassembler un ensemble d'études *in silico* autour des diverses maladies chroniques les plus répandues dans le monde et en Algérie, en donnant un exemple sur les différentes approches qui permettent d'aborder et de réaliser ce type d'étude, sur chaque type de maladie : cancer, diabète, maladies cardiovasculaires, maladie de Parkinson, obésité, polyarthrites, urticaire...

III.2. Schéma de l'étude

III.2-1. Critères d'inclusion des articles

Les articles qui répondent aux critères suivants sont inclus :

- Publication en anglais.
- Publication de type recherche originale.
- Date de publication de 2016 jusqu'à maintenant.
- Toute étude en relation avec le thème choisi.

III.2-3. Analyse des articles sélectionnés

L'analyse a été entreprise tout en précisant :

- L'objectif principal.
- La méthodologie d'étude.
- Les principaux résultats et discussion.
- Une conclusion.

Analyse d'articles

III.1. Article de Nand et al., 2016

In Silico identification of novel EGFR Tyrosine Kinase Inhibitors associated with Non-small Cell Lung Cancer Form phytochemical library

Objectif :

Identifier des inhibiteurs végétaux puissants contre le domaine de liaison à l'ATP de la tyrosine kinase de l'EGFR.

Méthode :

➤ **Extraction de la structure des protéines**

La structure cristalline tridimensionnelle du co-cristal cible de 1,85 Angström du domaine kinase de l'EGFR avec le Gefitinib (PDB id : 4WKQ) a été extraite de la banque de données sur les protéines. Divers ligands et autres hétéroatomes tels que l'eau ont été retirés de la structure de la protéine pour préparer l'ancrage à l'aide du logiciel Chimera 1.8.

➤ **Préparation de la structure du ligand**

La structure tridimensionnelle des molécules de ligand a été extraite de ChemSpider et de la base de données Zinc. Ensuite, les structures des ligands au format mol ont été converties en différents formats moléculaires tels que mol2, sdf en utilisant le logiciel Avogadro 10.

➤ **Amarrage moléculaire**

L'amarrage computationnel de tous les ligands a été réalisé dans le site actif du domaine kinase de l'EGFR à l'aide d'AutoDockVina dans l'interface graphique PyRx version 0.8. Avant l'arrimage, des atomes d'hydrogène polaires ont été ajoutés à la protéine, puis Kollman a unifié les atomes. Ensuite, les charges partielles des atomes unis de Kollman ont été attribuées. L'amarrage a été réalisé à l'aide de l'algorithme génétique lamarckien14 avec des protocoles d'amarrage standard. Les coordonnées de la protéine avec le ligand ont été visualisées à l'aide de la chimère UCSF et les interactions du complexe protéine-ligand ont été étudiées à l'aide de LigPlot+15.

➤ **ADMET**

Les propriétés pharmacologiques telles que l'absorption, la distribution, le métabolisme, l'excrétion et la toxicité des molécules retenues ont été prédites à l'aide du serveur admetSAR16. Dans ce serveur, les données les plus récentes et les plus complètes pour divers

produits chimiques associés à des profils ADMET connus ont été comparés aux composés de la requête.

➤ **Élucidation des pharmacophores**

-Les pharmacophores communs des ligands ont été analysés à l'aide du serveur web PharmaGist. Les neuf molécules actives obtenues lors d'un essai biologique (AID : 256664) ont été utilisées comme référence pour l'analyse du pharmacophore des molécules cibles. Ensuite, le pharmacophore commun aux ligands sélectionnés à l'issue des trois processus ci-dessus ont également été analysés.

Résultats :

Dans cette étude, 50 composés phytochimiques ont été ciblés contre la protéine EGFR afin d'analyser leur effet inhibiteur sur le CPNPC et les résultats ont été comparés à l'efficacité inhibitrice de l'inhibiteur standard (Gefitinib). Un modèle d'apprentissage automatique a été préparé pour sélectionner les molécules actives 28 molécules sur 50 ont été jugées actives. Une analyse d'amarrage plus poussée a révélé que 18 composés ont une affinité très élevée pour la protéine cible et une affinité similaire à celle de la molécule standard.

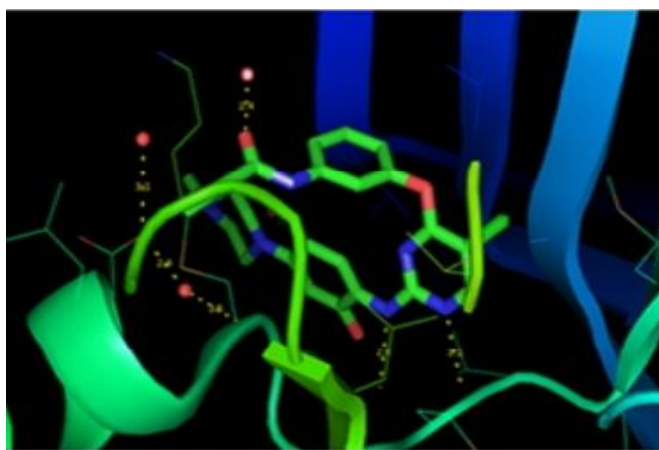


Figure 19 : Confirmation de liaison du ligand avec le domaine de liaison ATP des tyrosines de l'EGFR (Nand et al., 2016).

Des prédictions ADME et de toxicité ont finalement été effectuées pour évaluer la pharmacocinétique et la toxicité des 15 composés phytochimiques. 6 ligands ont été






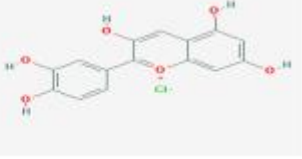

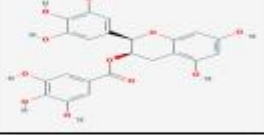

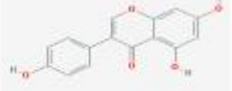


sélectionnés en analysant les résultats de l'ADMET, car ils ne présentent aucune propriété de mutagenicité ou de cancérogénicité et de cancérogénicité (Voir tableau 4).

Les composés actifs présentaient quatre atomes accepteurs de liaison hydrogène, un atome donneur de liaison hydrogène, un atome hydrophobe adjacent, deux anneaux aromatiques, tandis que le pharmacophore commun de 6 molécules cibles présentait trois accepteurs de liaison hydrogène, un donneur de liaison hydrogène, un groupe hydrophobe et zéro cycle aromatique.

Tableau 4 : Résultats de l'amarrage moléculaire (Nand et al., 2016).

S.N	Name of Phytochemicals	Docking Score (kcal/mol)	admetSAR Results			
			Human Intestinal Absorption	Caco-2 Permeability	Blood Brain Barrier	Carcinogens
i.	Reference	-4.1	HIA+	Caco2+	BBB+	Non-carcinogen
ii.	Apigenin	-8.2	HIA+	Caco2+	BBB+	Non-carcinogen
iii.	Curcumin	-5.0	HIA+	Caco2+	BBB+	Non-carcinogen
iv.	Cyanidin	-5.6	HIA+	Caco2+	BBB+	Non-carcinogen
v.	Epigallocatechin gallate	-5.0	HIA+	Caco2+	BBB+	Non-carcinogen
vi.	Genistein	-4.9	HIA+	Caco2+	BBB+	Non-carcinogen
vii.	Gingerol	-3.9	HIA+	Caco2-	BBB+	Non-carcinogen
viii.	Glucosinolate	-3.2	HIA+	Caco2-	BBB-	Non-carcinogen
ix.	Daidzein	-6.6	HIA-	Caco2-	BBB+	Non-carcinogen
x.	Isoflavone	-6.7	HIA+	Caco2-	BBB+	Non-carcinogen
xi.	Phenoxodiol	-5.9	HIA+	Caco2-	BBB+	Non-carcinogen
xii.	Benzopyran	-4.9	HIA+	Caco2-	BBB-	Non-carcinogen
xiii.	Cephalotaxine (A)	-5.9	HIA+	Caco2+	BBB+	Non-carcinogen
xiv.	Resveratrol	-5.8	HIA-	Caco2-	BBB+	Non-carcinogen
xv.	Rohitukine	-3.1	HIA+	Caco2-	BBB+	Non-carcinogen
xvi.	Flavopiridol	-4.1	HIA-	Caco2-	BBB+	Non-carcinogen

Tableau 5 : Sources et structures des six composés phytochimiques les plus recherchées (Nand et al., 2016).

Sl. NO	Name of Compounds	Sources		Structure
1	Apigenin	<i>Moringa oleifera</i>		
2	Curcumin	<i>Curcuma longa L</i>		
3	Cyanidin	Grapes, blackberry, cranberry, raspberry, apples, plums, red cabbage. And red onion.		
4	Epigallocatechin gallate	Green tea		
5	Genistein	lupine, fava beans, Soybeans, kudzu, psoralea, <i>Flemingia vestita</i> , and coffee.		
6	Cephalotaxine	Cephalotaxus		

Discussion :

La protéine EGFR est oncogénique joue un rôle important dans l'apparition et le développement descancers.

L'étude complète a été menée in silico pour explorer les potentiels thérapeutiques des substances phytochimiques d'origine naturelle contre l'EGFR. L'analyse globale suggère que 6 composés phytochimiques [tableau 5] présentent une affinité plus élevée pour la protéine EGFR. L'étude fournit un indice pour le développement de nouvelles molécules principales ayant une importance thérapeutique qui pourraient être utilisées pour la prévention ou la guérison du CPNPC.

Conclusion :

L'inhibition de l'EGFR en ciblant son domaine tyrosine kinase est un traitement mondialement reconnu du CPNPC. Les composés d'origine végétale dérivés de plantes présentent un excellent potentiel en tant qu'inhibiteurs de la croissance cancéreuse.

III.2. Article de Kose et al., 2019**Association of herbal plant in Rheumatoid Arthritis studies using in silico approach.****Objectif :**

Vise à trouver un composé phytochimique qui est Stigmastérol montrant une bonne affinité de liaison vers les récepteurs cibles, en utilisant approche d'amarrage pour découvrir une molécule de médicament potentielle contre la protéine TNF (Tumor Necrosis Factor alpha) qui provoque une inflammation.

Méthode :**➤ Sélection des protéines**

Au total, 10 protéines PR telles que 4JJ7, 6E28, 5gjh, 4FKL, 4EK4, 3CZF, 5NI9, 4RI5, 5BXU et 5N86 ont été sélectionnées comme cibles. De plus les fonctions, les mécanismes et les voies de ces protéines ont été analysés à l'aide de la base de données UniProt, les structures tridimensionnelles des protéines PR ont été extraites de la base de données PDB.

➤ Sélection des ligands

Au total, 6 substances phytochimiques ont été sélectionnées comme ligands, à savoir les phytostérols tels que le stigmastérol, les lignanes tels que l'arctigénine, les substances phytochimiques telles que les coumarines, la stéroïdalsaponine, l'hydroxytyrosol et la thymoquinone ont été vérifiés pour la règle de Lipinski de 5 dans PubChem (NCBI).

➤ Minimisation de l'énergie

Les protéines sont des macromolécules et doivent constituer une structure stable pour l'amarrage moléculaire. La minimisation de l'énergie permet de stabiliser la structure par le logiciel GROMACS 5.1.1.

➤ **Études de docking**

Les protéines dont l'énergie a été minimisée ont été arrimées à des substances phytochimiques sélectionnées afin de trouver un médicament puissant contre la PR, qui agisse comme un agent anti-inflammatoire, en utilisant le logiciel AutoDock 4.2.6.

➤ **Visualisation**

L'interaction entre les ligands (composés phytochimiques) et les protéines a été visualisée à l'aide du visualiseur Discovery Studio 19.1.

Résultats :

➤ **Résultats du docking**

Cette étude a permis de constater que le score d'amarrage du stigmastérol avec les 10 protéines présentait l'affinité de liaison la plus élevée et l'énergie de liaison obtenue étant la plus faible en tant qu'énergie de liaison obtenue dans les valeurs négatives indiquent que les scores obtenus sont très efficaces. Le tableau suivant montre les meilleures poses d'amarrage classées en fonction de leur énergie de liaison.

Tableau 6 : Principales conformations d'amarrage indiquant l'énergie de liaison des structures amarrées (Kose et al., 2019).

S no.	PDB ID	Phytochemical	Binding energy (Kcal/mol)
1	4JJ7	Stigmastérol	-8.26
2	6E28		-7.28
3	5GJH		-7.37
4	5NI9		-7.58
5	5N86		-8.51
6	3CZF		-10.13
7	5BXU		-7.35
8	4FKL		-7.51
9	4EK4		-7.91
10	4RI5		-8.78

➤ **Résultats de la visualisation :**

-Les figures 20 à 29 montrent les interactions protéine-ligand visualisées dans Discovery Studio 19.1 où les résidus d'acides aminés sont marqués en vert et le ligand est coloré en vert.

-Les substances phytochimiques les mieux classées (tableau 6) ont été sélectionnées pour montrer leur interaction avec les protéines de la PR.

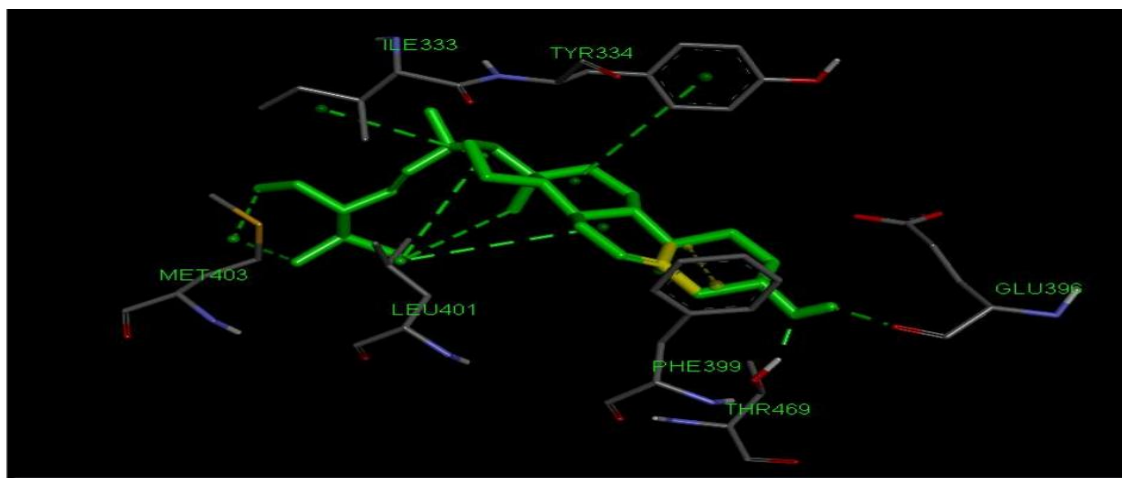


Figure 20 : 4JJ7 docké avec le stigmasterol montrant diverses interactions avec les acides aminés - ILE333, TYR334, PHE399, GLU396, LEU401, MET403 et THR469 (Kose et al., 2019).

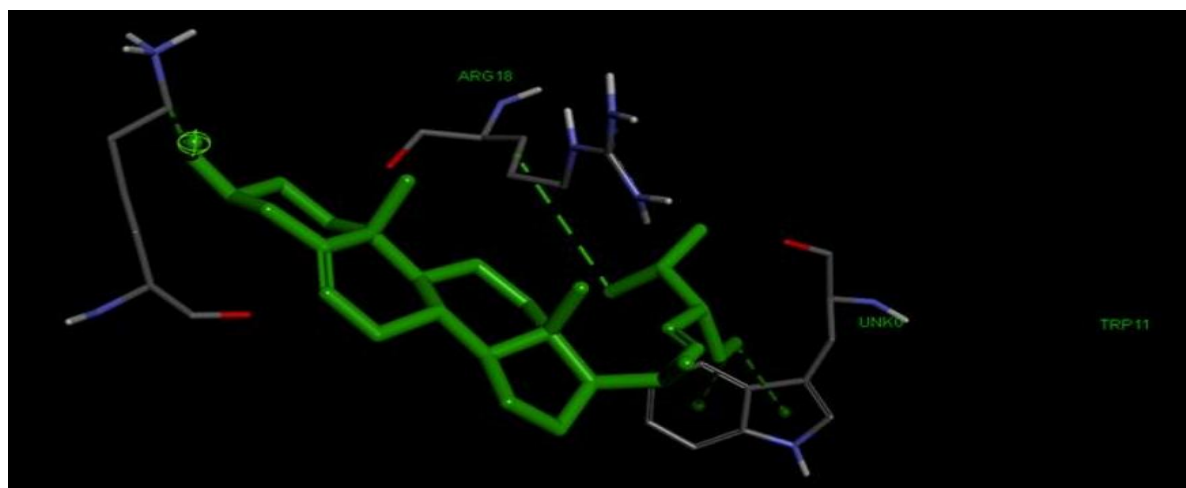


Figure 21 : 6E28 docké avec le stigmasterol montrant diverses interactions avec les acides aminés UNK0, TRP11 et ARG18 (Kose et al., 2019).

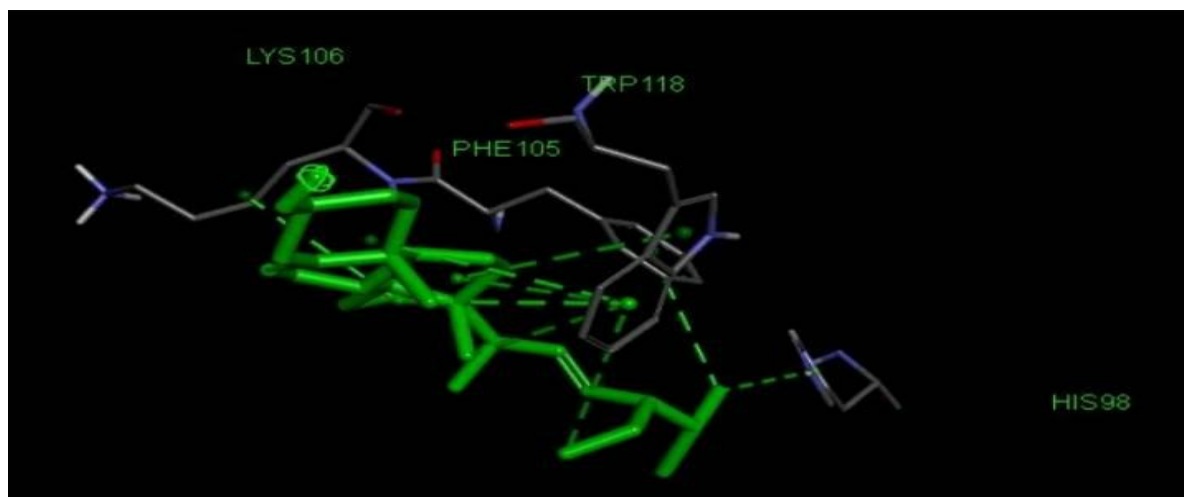


Figure 22 : 5GJH docké avec le stigmasterol montrant diverses interactions avec les acides aminés HIS98, PHE105, LYS106 et TRP118 (Kose et al., 2019).

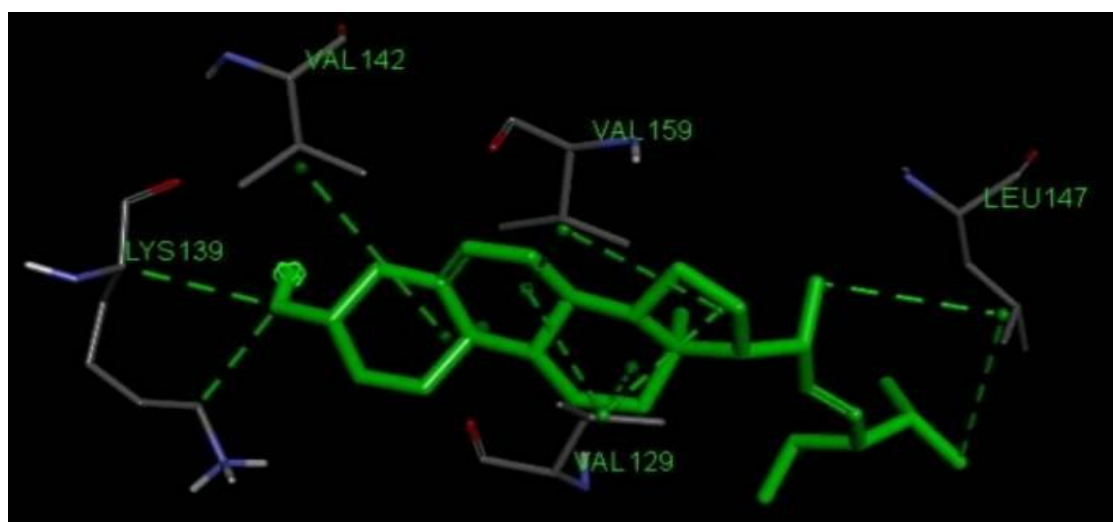


Figure 23 : 5NI9 docké avec le stigmasterol montrant diverses interactions avec les acides aminés VAL129, LYS139, VAL142, LEU147 et VAL159 (Kose et al., 2019).

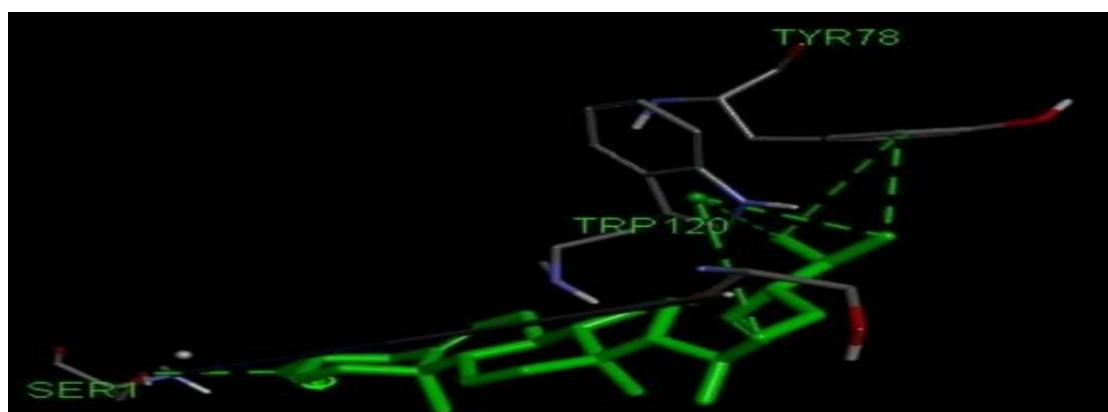


Figure 24 : 5N86 ancré avec le stigmasterol montrant diverses interactions avec les acides aminés SER1, TYR78 et TRP120 (Kose et al., 2019).

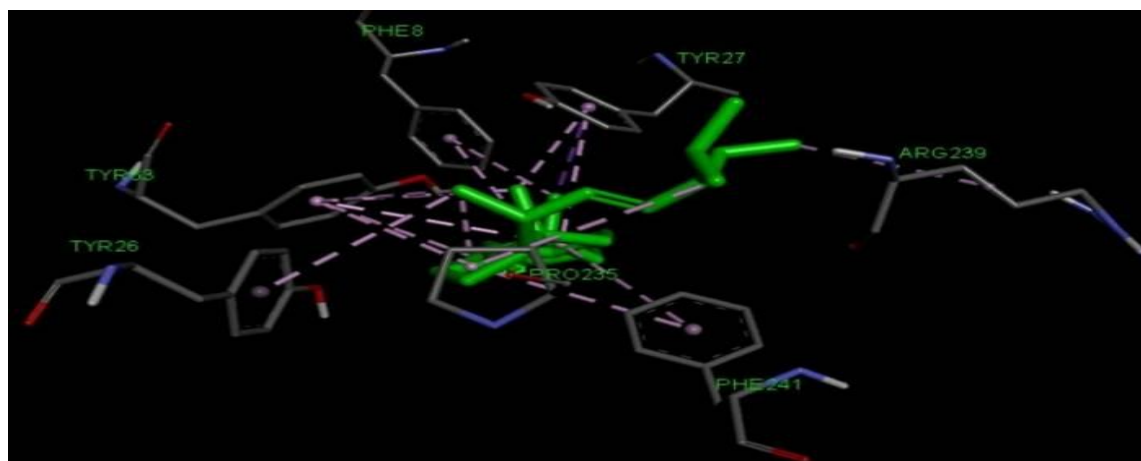


Figure 25 : 3CZF arrimé au stigmasterol montrant diverses interactions avec les acides aminés PHE8, TYR26, TYR27, TYR63, PRO235, ARG239 et PHE241 (Kose et al., 2019).

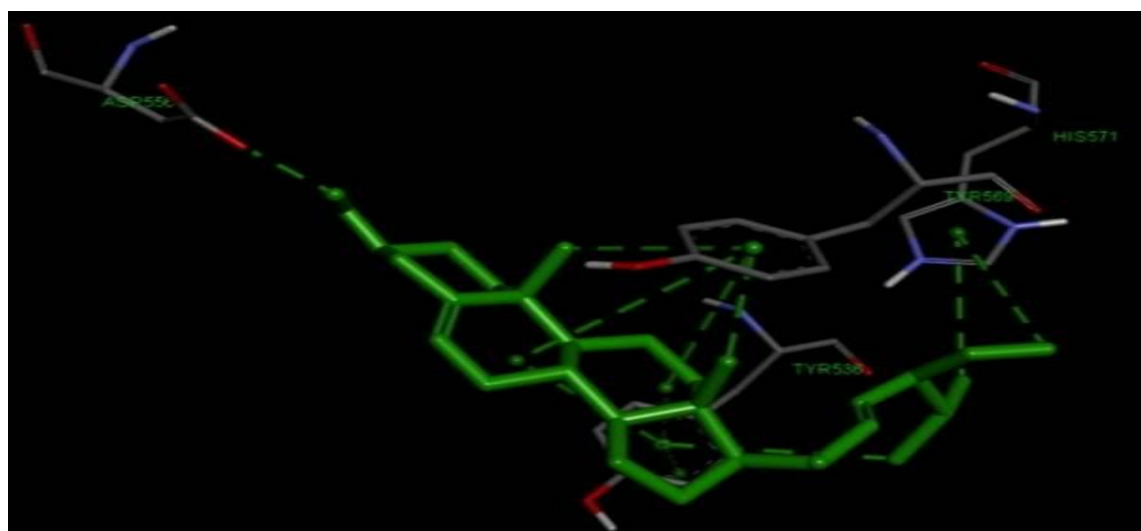


Figure 26 : 5BXU docké avec le stigmasterol montrant diverses interactions avec les acides aminés TYR536, ASP556, TYR569 et HIS571 (Kose et al., 2019).

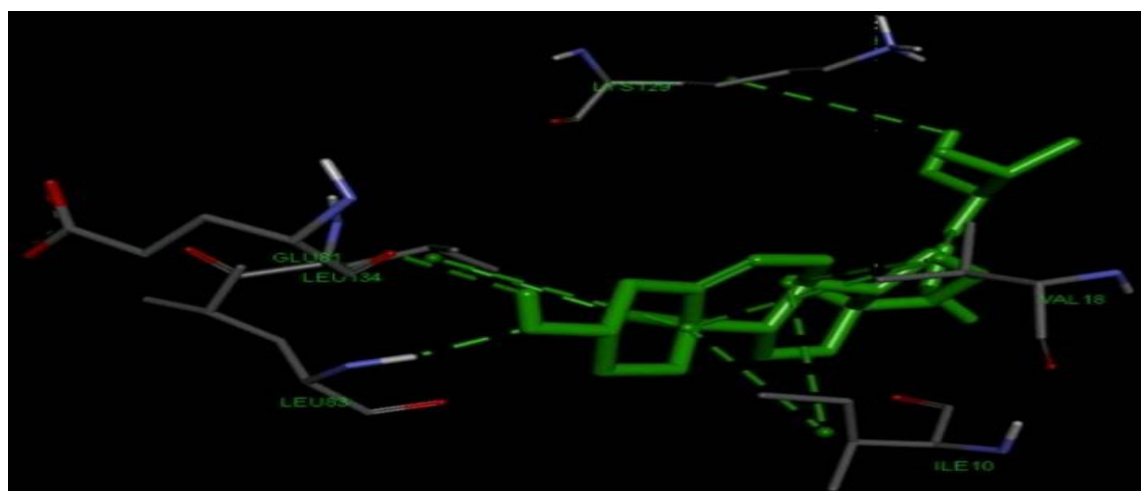


Figure 27 : 4FKL docké avec le stigmasterol montrant diverses interactions avec les acides aminés - ILE10, VAL18, GLU81, LEU83, LYS129 et LEU134 (Kose et al., 2019).

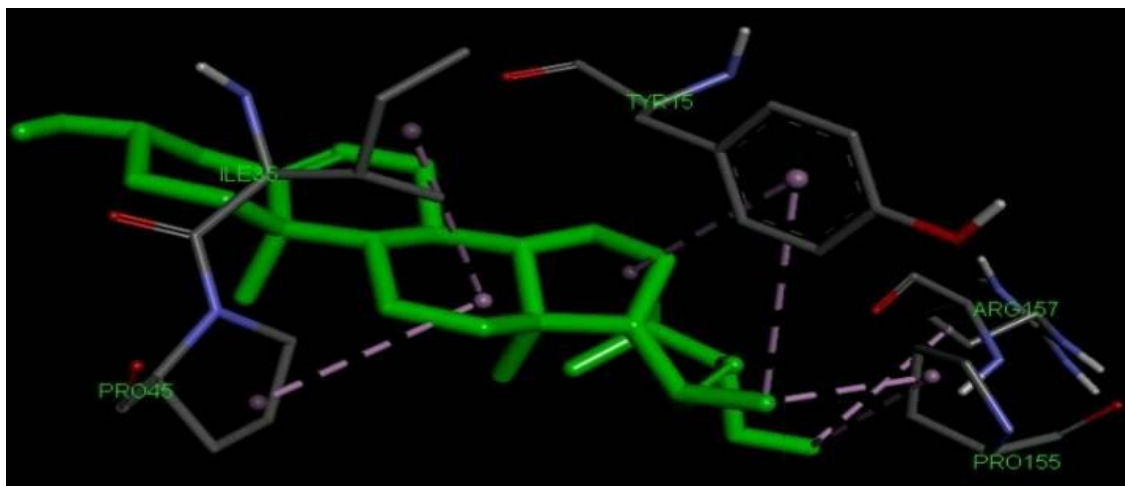


Figure 28 : 4EK4 arrimé au stigmastérol montrant diverses interactions avec les acides aminés TYR15, ILE35, PRO45, PRO155 et ARG157 (Kose et al., 2019).

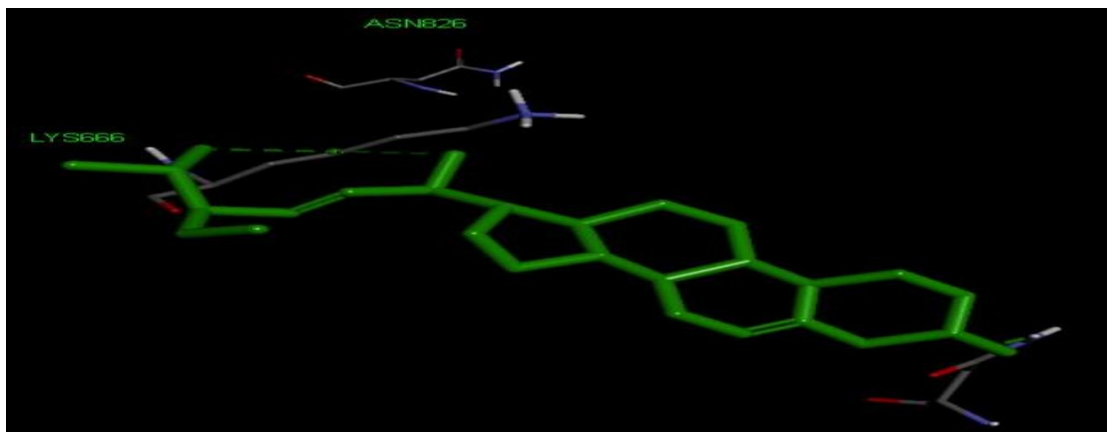


Figure 29 : 4RI5 arrimé au stigmastérol montrant diverses interactions avec les acides aminés LYS666 et ASN826 (Kose et al., 2019).

Discussion :

Dans cette étude, l'amarrage moléculaire a été utilisé pour traiter l'inflammation en ciblant les récepteurs des protéines 4JJ7, 6E28, 5GJH, 4FKL, 4EK4, 3CZF, 5NI9, 4RI5, 5BXU et 5N86, qui provoque une inflammation due à l'augmentation du niveau de la protéine TNF alpha à l'intérieur des articulations.

Des phytostérols comme le stigmastérol, des lignanes comme l'arctigénine, des composés phytochimiques comme les coumarines, les saponines stéroïdienne, l'hydroxytyrosol et la thymoquinone. Parmi ces composés, l'énergie de liaison la plus faible avec la cible n'est observée que dans le cas du stigmastérol, allant de -7,00 Kcal/mol à -10,13 Kcal/mol (tableau 6). Ces composés phytochimiques peuvent être considérées comme des médicaments puissants contre le rhumatisme articulaire (Kose et al., 2019).

Conclusion :

L'approche d'amarrage moléculaire a été utilisée pour déterminer les composés phytochimiques présentant une bonne affinité de liaison avec les protéines du PR à l'aide de l'outil AutoDock. Le stigmastérol peut être utilisé comme médicament potentiel pour l'activité anti-inflammatoire dans la PR, qui peut inhiber l'activité des protéines du PR comme la protéine 4JJ7, et qui supprime la protéine TNF alpha responsable de l'inflammation sans effets secondaires.

III.3. Article de Karthic et al., 2019**In-silico molecular docking analysis of potential phytotherapeutics from the medicinal herb *Corallocarpus Epigaeus* for treating urticaria.****Objectif :**

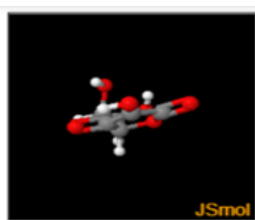
Evaluer le potentiel antiallergique des composés phytothérapeutiques bioactifs de l'herbe *Corallocarpus Epigaeus* contre le récepteur de l'histamine 1 avec PDB 3RZE, Prostaglandine H2 synthases- PDB 1IGX, Cyclooxygénase I- PDB 3KK6, Cyclooxygénase 2- PDB 6COX.

Méthode :**✓ Préparation des ligands**

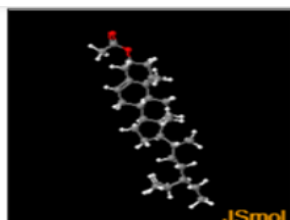
-Les ligands tels que l'acide ascorbique, le bêta sitostérol, sesquiterpène et tocophérol, ainsi que les ligands standard suivants : Cetirizine, acide salicylique, Ibuprofène et Celecoxib ont été élaborés à l'aide de Chemskech et optimisés à l'aide d'Outil web en ligne du serveur d'amarrage tel qu'illustré dans la figure 29. Pour les études d'amarrage en utilisant la méthode d'optimisation de la géométrie MMFF94 et le calcul de la charge a été effectué à l'aide de la méthode de Gasteiger au PH 7, comme indiqué dans le tableau 7.

Tableau 7 : Propriétés des ligands des molécules de plomb sélectionnées (Karthic et al., 2019).

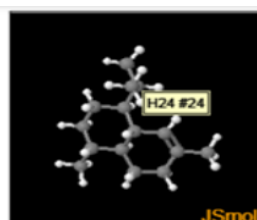
S.No	Name of the Compounds	Molar weight g/mol	H Bond Donor	H Bond Acceptor	Rotatable bonds
1	Ascorbic Acid	176.124 g/mol	4	6	2
2	Beta Sitosterol	414.718 g/mol	1	1	6
3	Sesquiterpene	480.777 g/mol	0	2	20
4	Tocopherol	430.717 g/mol	1	2	12
5	Cetirizine	461.808 g/mol	1	5	8
6	Salicylic acid	138.122 g/mol	2	3	1
7	Celecoxib	381.373 g/mol	1	7	3
8	Ibuprofen	206.285 g/mol	1	2	4



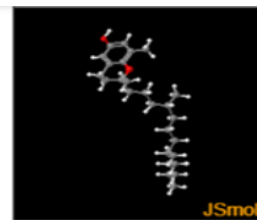
1



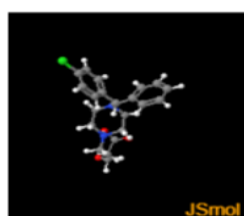
2



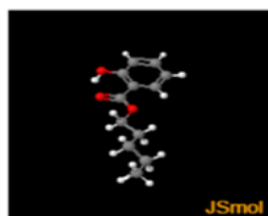
3



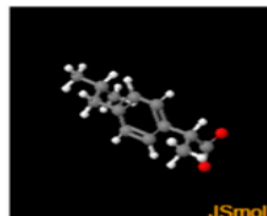
4



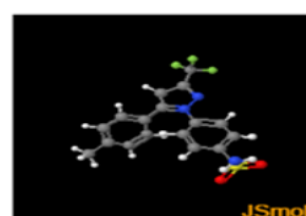
5



6



7



8

Figure 30 : Structure 3D du plomb 1.acide ascorbique 2. Bêta sitostérol 3. sesquiterpène 4. Tocophérol 5.cétirizine 6.acide salicylique 7.ibuprofène et 8. Célécoxib (Karthic et al., 2019).

✓ **Prédiction du site actif**

Le site actif de l'enzyme a été obtenu par le serveur web LIGSITE en utilisant l'identification automatique des poches à la surface de la protéine à partir des coordonnées 3D de la protéine. De plus, les sites de liaison potentiels des ligands dans Récepteur de l'histamine1, Prostaglandine H2 synthases, Cyclooxygénase I, Cyclooxygénase II est identifié en utilisant un espace de grille de 1 et une sonde de 0,5 mm. La prédiction du site du ligand a été effectuée en utilisant l'outil en ligne GHECOM et les calculs de poches respectives.

✓ **Méthodologie de docking**

Les calculs d'empilement ont été effectués à l'aide de Docking Server et les charges partielles de Gasteiger ont été ajoutées aux atomes du ligand. Les atomes d'hydrogène non polaires ont été fusionnés et les liaisons rotatives ont été définies. De plus les calculs d'amarrage ont été effectués sur la base de l'énergie libre sur les composés suivants : acide ascorbique, bêta sitostérol, sesquiterpène et le tocophérol, ainsi que la cétirizine, l'acide salicylique, l'ibuprofène, le célécoxib et leur affinité de liaison vers la protéine cible avec le récepteur de l'histamine 1, Prostaglandine H2 synthases, Cyclooxygénase I, Cyclooxygénase II, comme le montrent les figures 31 à 34. Les atomes d'hydrogène essentiels, le type d'atome uni de Kollman, les charges et les paramètres de solvation ont été ajoutés à l'aide des outils Auto Dock et les cartes d'affinité (grille) de points de grille Å et un espacement de 0,375 Å ont été générés à l'aide du programme Autogrid.



Figure 31 : Protéine cible récepteur de l'histamine 1 avec le code PDB 3RZE (Karthic et al., 2019).

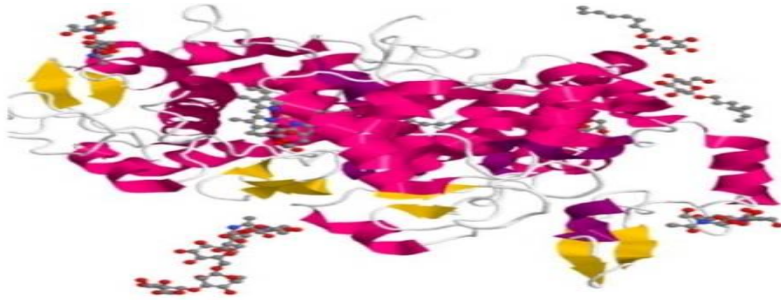


Figure 32 : Protéine cible Prostaglandine H2 synthases avec le code PDB 1IGX (Karthic et al., 2019).

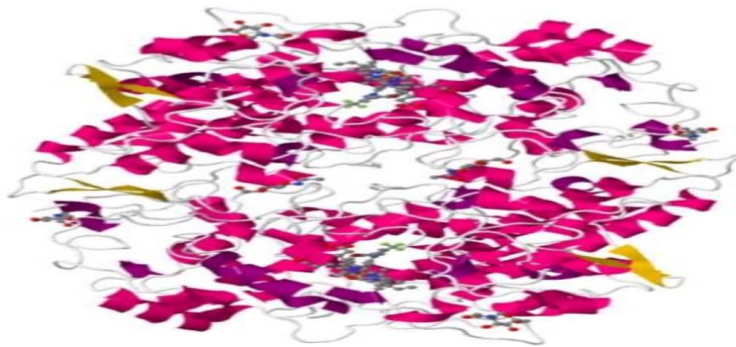


Figure 33 : Protéine cible Cyclooxygénase 2 avec le code PDB 6COX (Karthic et al., 2019).

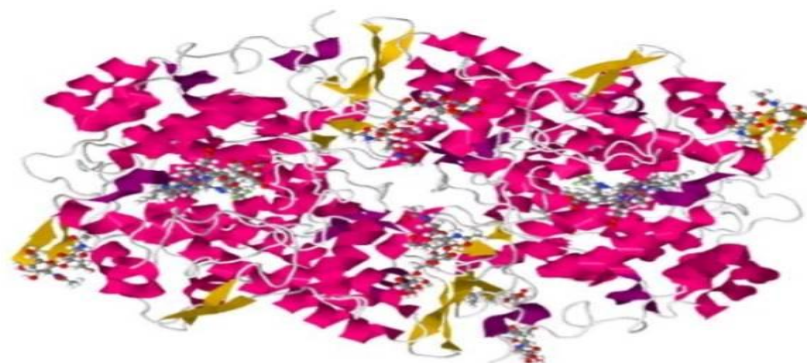


Figure 34 : Protéine cible Cyclooxygénase I avec le code PDB 3KK6 (Karthic et al., 2019).

Résultats :

✓ Score de docking des ligands avec les récepteurs de l'histamine 1

L'acide aminé 428 TRP est le résidu le plus important dans la médiation de l'activité du récepteur de l'histamine 1. La liaison des composés principaux avec ce résidu central peut avoir une activité de blocage du récepteur H1 sur les quatre composés suivants : L'acide ascorbique, le bêta-sitostérol et le tocophérol ont tendance à se lier à l'acide aminé 428 TRP, similaire à celle de la cétirizine standard. Ces composés possèdent donc des propriétés prometteuses de blocage des récepteurs de l'histamine 1. Alors que le Sesquiterpène n'a pas pu se lier à l'acide aminé 428 TRP et n'a donc pas d'activité de blocage des récepteurs de l'histamine 1.

Le score d'amarrage en ce qui concerne l'énergie libre de liaison, la constante d'inhibition y compris la surface d'interaction totale, sont répertoriés dans le tableau 8 et illustré dans la figure 35.

Tableau 8 : Résultats des études d'ancrage moléculaire des composés contre le récepteur de l'histamine 1(Karthic et al., 2019).

Compound	Docking Score (Est Free Binding Energy k Cal/mol)	Inhibition constant Ki μ M (*mM)(** nm)	Electrostatic energy Kcal/mol	Intermolecular energy Kcal/mol	Total Interaction Surface
Ascorbic Acid	-4.85	278.87	-0.14	-4.7	427.58
Beta Sitosterol	-8.31	811.62**	-0.01	-10.08	773.35
Sesquiterpene	-7.44	3.51	0	-7.74	634
Tocopherol	-6.05	36.65	-0.04	-9.72	994.11
Cetirizine	-5.58	81.77	-0.19	-8.95	782.54

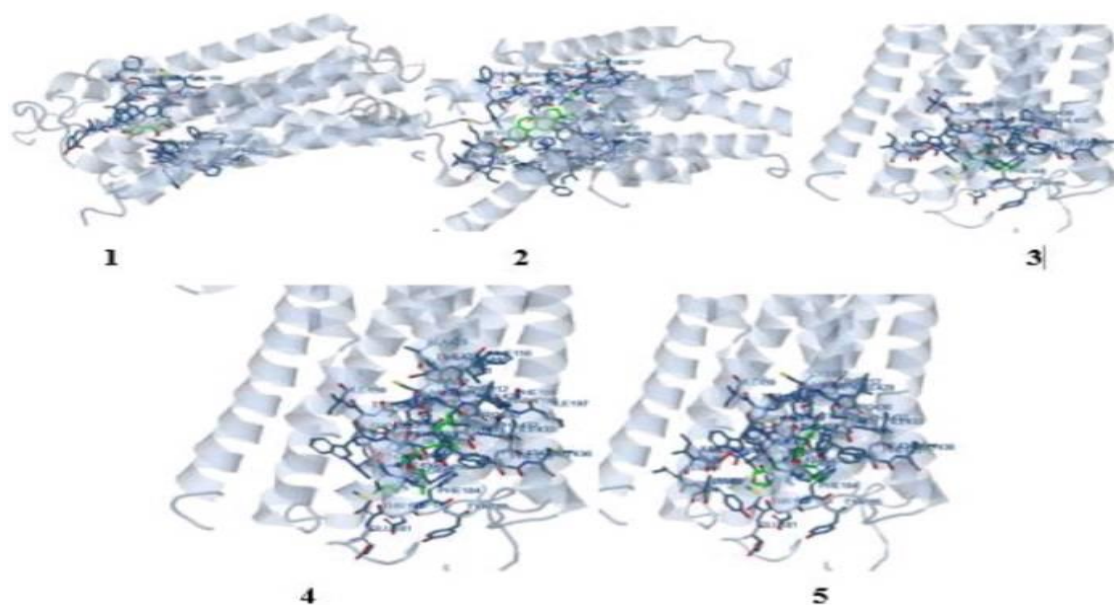


Figure 35 : Poches de liaison possibles du ligand à la surface de la cible contre le récepteur de l'histamine 1 avec le code PDB 3RZE. calculées par GHECOM. 1. acide ascorbique 2. bêta sitostérol 3. sesquiterpène 4. tocophérol et 5. Cetirizine (Karthic et al., 2019).

✓ **Score d'ancrage des ligands avec les synthèses de la prostaglandine H2 synthases**

Les acides aminés tels que 202 ALA, 206 THR, 385 TYR, 338 HIS et 390 LEU sont les principaux résidus qui interviennent dans la médiation de l'activité enzymatique de la prostaglandine synthase. Sur les quatre composés, le bêta-sitostérol et le sesquiterpène ont 6 interactions similaires à celles de l'acide salicylique standard et l'acide ascorbique et le tocophérol ont respectivement 5 et 4 interactions similaires à celles de l'étalon. Ces composés possèdent donc des propriétés d'inhibition de l'enzyme prostaglandine synthase.

Le score d'amarrage en ce qui concerne l'énergie libre de liaison, la constante d'inhibition y compris la surface d'interaction totale ont été répertoriés dans le tableau 9 et illustré dans la figure 36.

Tableau 9 : Résultats des études d'ancrage moléculaire des composés contre les synthèses de la prostaglandine H2 (Karthic et al., 2019).

Compounds	Docking Score (Est Free Binding Energy kcal/ mol)	Inhibition constant Ki μM (*mM)(**nm)	Electrostatic energy Kcal/mol	Intermolecular energy Kcal/mol	Total Interaction Surface
Ascorbic Acid	-5.39	111.03	-0.07	-5.37	459.97
Beta Sitosterol	-5.65	72.64	0	-9.11	922.79
Sesquiterpene	-8.74	398.87**	0	-9.04	594.84
Tocopherol	-7.15	5.72	0.01	-10.22	883.92
Salicylic acid	-5.24	145.22	-0.04	-5.82	395.57

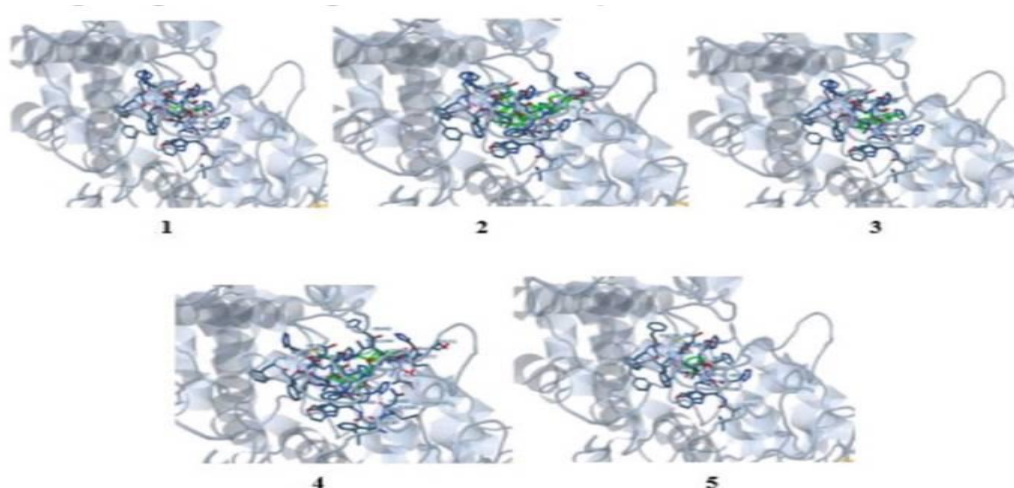


Figure 36 : Poches de liaison possibles du ligand à la surface de la cible contre le récepteur de la prostaglandine H2 synthases avec le code PDB 1IGX calculées par GHECOM. 1.acide ascorbique 2.bêta sitostérol 3.sesquiterpène 4.tocophérol et 5. Acide salicylique (**Karthic et al., 2019**).

✓ Score d'ancrage des ligands avec la cyclooxygénase I

Les acides aminés tels que 192 GLN, 352 LEU et 523 ILE sont les principaux résidus impliqués dans la médiation de l'activité enzymatique de la cyclo-oxygénase I. Sur les quatre composés, L'acide ascorbique et le bêta-sitostérol ont deux interactions similaires à celles de l'ibuprofène standard. D'autres composés tels que l'acide ascorbique et le bêta-sitostérol ont une interaction similaire à celle de l'ibuprofène standard. Par conséquent, ces composés possèdent une activité prometteuse d'inhibition de l'enzyme Cyclooxygénase I prometteuse.

Le score d'amarrage en ce qui concerne l'énergie libre de liaison, la constante d'inhibition, y compris la surface d'interaction totale, sont répertoriés dans le tableau 10 et illustrés dans la figure 37.

Tableau 10 : Résumé des études d'amarrage moléculaire des composés contre la cyclooxygénase I (**Karthic et al., 2019**).

Compounds	Docking Score (Est Free Binding Energy k Cal/ mol)	Inhibition constant Ki μM (*mM) (**nm)	Electrostat ic energy Kcal/mol	Intermolecul ar energy Kcal/mol	Total Interaction Surface
Ascorbic Acid	-4.71	350.89	-0.02	-4.74	490.52
Beta Sitosterol	-3.4	3.23*	0.02	-10.6	960.29
Sesquiterpene	-8.02	1.33	-0.01	-8.31	623.25
Tocopherol	-8.04	1.28	0.03	-11.94	1022.49
Ibuprofen	-6.51	16.99	0	-7.71	648.81

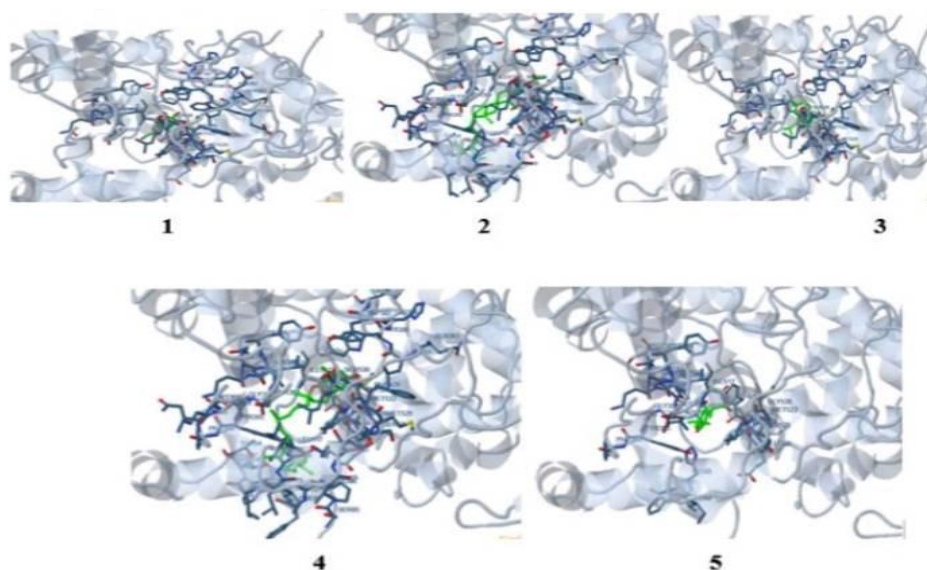


Figure 37 : Poches de liaison possibles du ligand à la surface de la cible contre le récepteur de la cyclo-oxygénase I avec le code PDB 3KK6 par GHECOM. 1.acide ascorbique 2.bêta sitostérol 3.sesquiterpène 4. tocophérol et 5. Ibuprofène (Karthic et al., 2019).

✓ Score de Docking des ligands avec la cyclooxygénase II

Les acides aminés tels que 90 HIS, 352 LEU, 353SER et 387 TRP sont les principaux résidus impliqués dans la médiation de l'activité enzymatique de la cyclooxygénase II.

D'après les résultats obtenus par l'analyse d'amarrage actuelle, il a été observé qu'aucun des composés n'a tendance à se lier au résidu central susmentionné et il a donc été conclu

qu'aucun des quatre composés n'a tendance à inhiber l'enzyme cyclooxygénase II par rapport au célécoxib standard.

Le score d'amarrage en ce qui concerne l'énergie libre de liaison, la constante d'inhibition, y compris la surface d'interaction totale, sont répertoriés dans le tableau 11 et illustré dans la figure 38.

Tableau 11: Résumé des études d'amarrage moléculaire des composés contre la cyclooxygénase II (Karthic et al., 2019).

Compounds	Docking Score (Est Free Binding Energy k Cal/ mol)	Inhibition constant Ki μM (*mM) (**nm)	Electrosta tic energy Kcal/mol	Intermolecu lar energy Kcal/mol	Total Interaction Surface
Ascorbic Acid	-4.52	485.94	-0.03	-4.48	490.45
Beta Sitosterol	-4.6	425.55	0.04	-6.42	640.33
Sesquiterpene	-7.87	1.69	-0.01	-8.17	623.53
Tocopherol	-5.54	87.08	0	-9.03	955.53
Celecoxib	-7.86	1.73	-0.04	-9.7	775.42

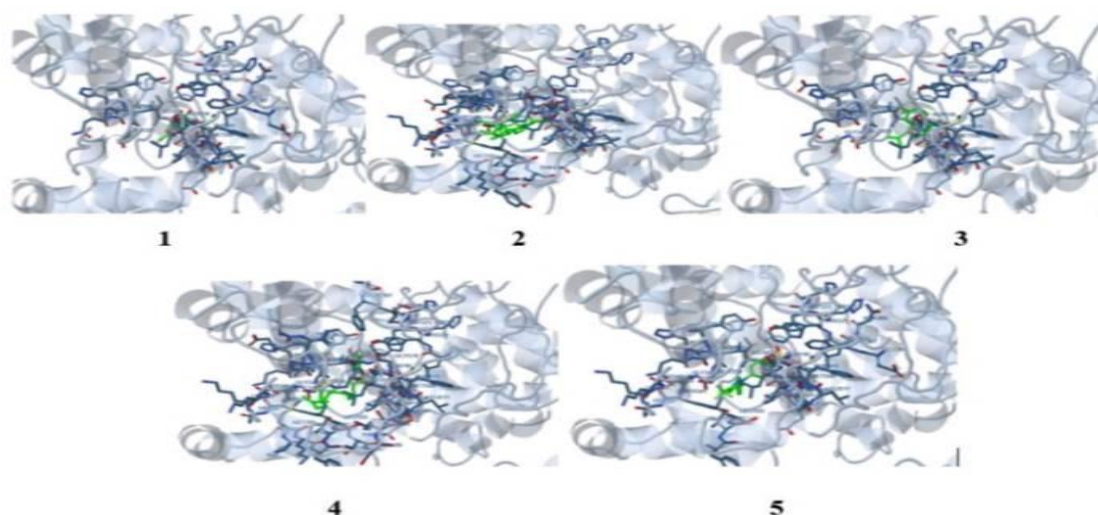


Figure38 : Poches de liaison possibles du ligand à la surface de la cible contre le récepteur de la cyclooxygénase II avec le code PDB 6COX calculées par GHECOM. 1.acide ascorbique 2.bêta sitostérol 3.sesquiterpène 4.tocophérol et 5. Celecoxib (**Karthic et al., 2019**).

Discussion :

Cyclooxygénase I et cyclooxygénase II ont des structures et des activités catalytiques similaires. Les séquences d'acides aminés des sites de liaison du substrat et du site catalytique sont identiques, mais la valine est substituée à l'isoleucine en positions 434 et 523. Sur les quatre composés, l'acide ascorbique et le bêta sitostérol ont deux interactions similaires à celles de l'ibuprofène standard. D'autres composés tels que l'acide ascorbique et le bêta sitostérol ont une interaction similaire à celle de la norme sur le récepteur de la cyclooxygénase I. En outre, Les résultats obtenus à partir de l'analyse d'amarrage actuelle ont permis d'observer qu'aucun des composés n'a tendance à se lier au résidu central susmentionné, ce qui permet de conclure qu'aucun des quatre composés n'a tendance à inhiber l'enzyme cyclooxygénase II par rapport au célécoxib standard.

Comme l'enzyme prostaglandine synthase est impliquée dans la catalysation de la production de prostaglandine qui sensibilise la terminaison nerveuse pour moduler l'amplitude des réponses à la douleur. Il a été observé, d'après les résultats de la présente étude que le bêta sitostérol et le sesquiterpène ont des interactions similaires à celles de l'acide salicylique standard. D'autres composés tels que l'acide ascorbique et le tocophérol ont des interactions 5 et 4 similaires à celles de l'affinité de liaison standard avec la prostaglandine H2 synthases.

L'histamine déclenche considérablement l'épisode de démangeaison qui est la marque pathologique la plus importante de l'urticaire. L'utilisation d'un antihistaminique entraîne des effets secondaires potentiels qui comprennent la somnolence, la fatigue, les maux de tête, les

nausées et la sécheresse de la bouche. Les résultats de la présente étude ont révélé que les molécules principales telles que l'acide ascorbique, le bêta-sitostérol et le tocophérol ont tendance à se lier à l'acide aminé 428 TRP, de façon similaire à la cétirizine. Par conséquent, ces composés possèdent une activité prometteuse de blocage des récepteurs de l'histamine 1.

Le docking computationnel minimise le temps d'analyses moléculaires pour sélectionner un ligand approprié pour les recherches en laboratoire (Karthic et al., 2019).

Conclusion :

L'amarrage moléculaire des quatre composés phytothérapeutiques tels que l'acide ascorbique, le bêta sitostérol, le sesquiterpène et le tocophérol présent dans *Corallocarpus Epigaeus* ont une affinité significative pour les récepteurs tels que le récepteur de l'histamine 1 la prostaglandine H2 synthase et la cyclooxygénase I, tandis qu'aucun des composés n'exercerait l'affinité de liaison attendue sur le récepteur de la cyclooxygénase II. De ce fait, il a été que l'herbe *Corallocarpus Epigaeus* pourrait constituer une meilleure piste thérapeutique pour la gestion clinique des symptômes liés à l'urticaire.

III.4. Article de Akhila et al., 2020

An in-silico approach to curtail the action of obesity proteins with phytoconstituents.

Objectif :

Chercher à limiter l'action des protéines en amarrage aux phytoconstituants.

Méthode :

- **Cibles protéiques**

Trois protéines ciblées ont été soumises à l'étude. Les protéines sélectionnées étaient les protéines associées à la masse grasse et à l'obésité FTO (ID PDB : 3LFM), résistine (ID PDB : 1LV6) et leptine (ID PDB : 1AX8). De plus toutes les protéines ont été téléchargées de la banque de données sur les protéines et représenté dans les figures 39, 40 et 41.

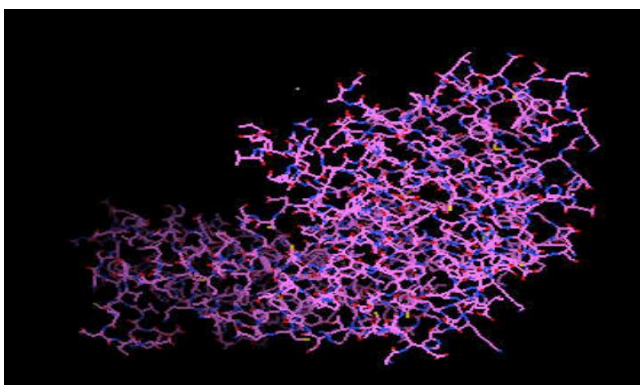


Figure 39 : 3 LFM (protéine FTO) (Akhila et al., 2020).

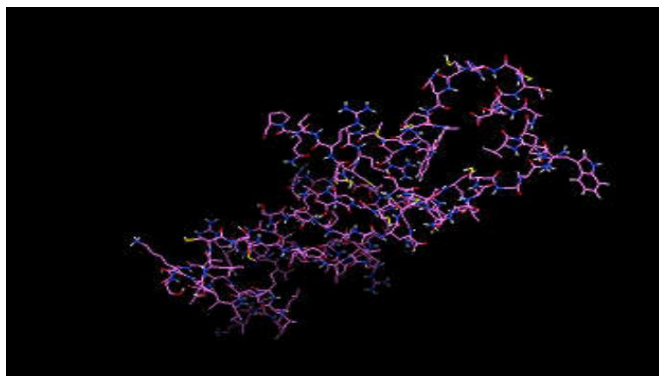


Figure 40 :1LV6 (Resistin) (Akhila et al., 2020).

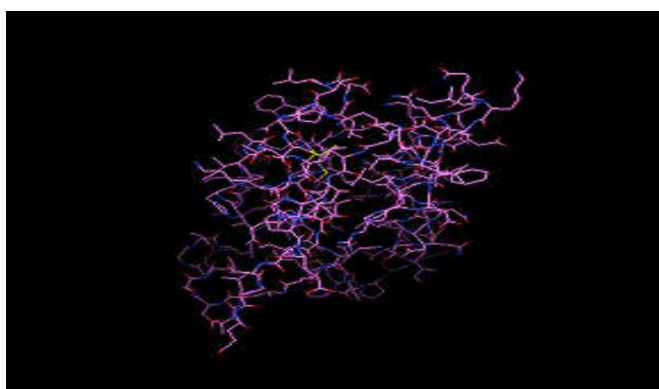


Figure 41 :1AX8 (Leptin) (Akhila et al., 2020).

- **Préparation des ligands**

Les phytoconstituants sélectionnés pour l'étude ont été recueillis après un examen approfondi des ingrédients alimentaires, généralement des fruits, des légumes et des plantes et des graines utilisés dans la vie quotidienne, et dont on dit qu'ils sont préférés dans le régime anti-obèse. Tous les ligands utilisés pour l'étude ont été dessinés à l'aide de chem sketch et convertis en fichiers pdb à l'aide de l'outil en ligne CORINA.

- **Analyse d'amarrage**

Auto dock tools 1.5.6 a été utilisée pour l'amarrage moléculaire et la liaison du ligand à trois protéines différentes a été calculée. Les fichiers PDBQT des protéines préparées et des ligands préparés ont été utilisés pour préparer l'autogrille et les points de grille dans les axes X, Y et Z ont été fixés à 60Å. L'algorithme génétique lamarckien a été utilisé pour docking conforme 60 et le docking a été réalisé sur la base de l'algorithme génétique de Lamarck.

Résultats :

Les résultats ont été comparés à ceux de l'orlistat. L'énergie de liaison obtenue pour les phytoconstituants par rapport à chaque protéine FTO, résistine et leptine sont présentées dans les tableaux 12, 13 et 14 respectivement.

-L'interaction des constituants avec les protéines est représentée dans les figures 42,43, 44, 45 et 46, respectivement.

Tableau 12 : Interaction de FTO avec les ligands (Akhila et al., 2020).

Ligands	Binding energy	No: of H bonds formed
Caffeic acid	-5.36	2
Coumaric acid	-4.28	1
Quercetin	-5.13	2
Genistein	-6.18	1
Daidzein	-6.84	1
Apigenin	-5.92	1
Luteolin	-5.47	2
Catechin	-5.45	3
Malvidin	-5.16	2
Capsaicin	-5.02	1
Nicotine	-5.48	2
Resveratrol	-5.1	1
Orlistat (synthetic drug)	-7.15	0

Tableau 13 : Interaction de la résistine avec les ligands (Akhila et al., 2020).

Ligands	Binding energy	No: of H bonds formed
Caffeic acid	-3.01	1
Coumaric acid	-3.76	3
Quercetin	-5.19	4
Genistein	-4.43	1
Daidzein	-5.81	2
Apigenin	-5.18	3
Luteolin	-4.9	4
Catechin	-4.23	1
Malvidin	-3.93	1
Capsaicin	-3.36	0
Nicotine	-5.48	1
Resveratrol	-3.49	2
Orlistat (synthetic drug)	-7.16	0

Tableau14 : Interaction de la leptine avec des ligands (Akhila et al., 2020).

Ligands	Binding energy	No: of H bonds formed
Caffeic acid	-3.20	2
Coumaric acid	-3.61	1
Quercetin	-5.13	3
Genistein	-5.25	1
Daidzein	-5.99	1
Apigenin	-5.93	0
Luteolin	-4.62	1
Catechin	-5.15	5
Malvidin	-4.69	4
Capsaicin	-4.19	1
Nicotine	-5.15	2
Resveratrol	-5.46	1
Orlistat (synthetic drug)	-5.31	1

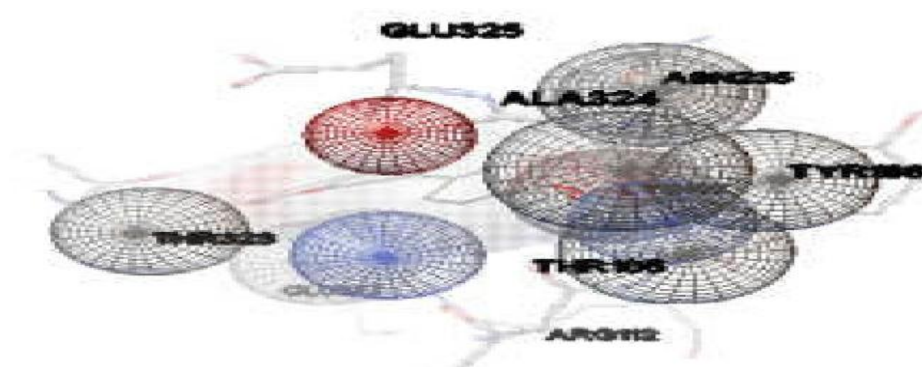


Figure 42 : Interaction de la protéine FTO avec la daidzéine (Akhila et al., 2020).

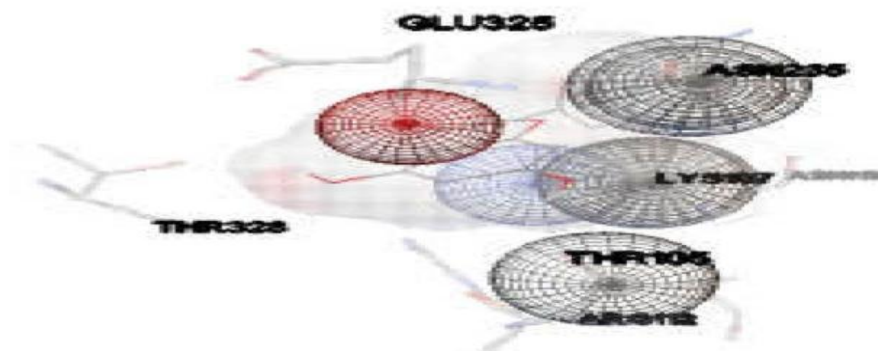


Figure 43 : Interaction de la protéine FTO avec la catéchine (Akhila et al., 2020).

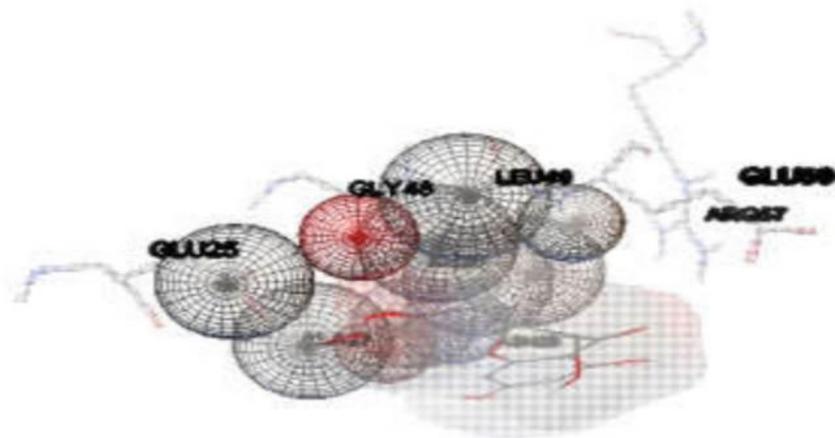


Figure 44 : Interaction de la résistine avec la quercétine (Akhila et al., 2020).



Figure 45 : Interaction de la leptine avec la catéchine (Akhila et al., 2020).

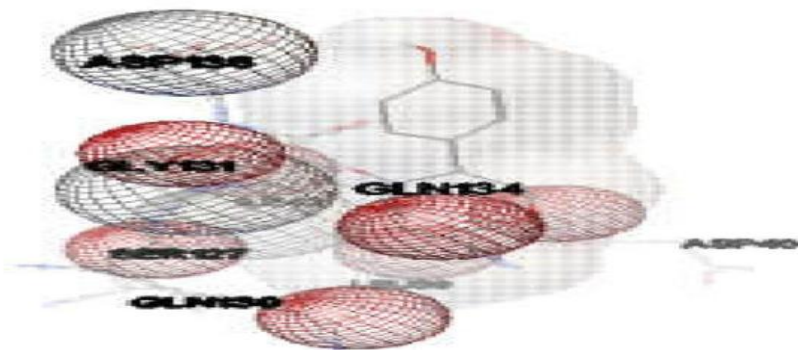


Figure 46 : Interaction de la leptine avec la daidezine (Akhila et al., 2020).

Discussion :

FTO (fat mass and obesity-associated protéin) et la résistine et la leptine sont les protéines associées à la masse grasse et à l'obésité, la résistine est FOT joue un rôle important dans la régulation de l'accumulation de graisse dans l'organisme et la leptine est une hormone qui joue un rôle important en ajustant à la fois la sensation de faim et l'ajustement de la dépense énergétique.

La présente étude a donc pour but de cribler certains composés phytoconstituants appartenant à différentes classes et communément observés dans les fruits et légumes contre trois différentes protéines cibles. Cela a permis d'identifier les composés phytoconstituants qui agissent mieux contre l'obésité avec des mécanismes différents. Les phytoconstituants tels que la daidzéine, qui est un isoflavonoïde, se sont avérés être les meilleurs phytoconstituants pour se lier avec les trois protéines et prévenir l'obésité. Les flavanols comme la quercétine et les catéchines ont également montré une bonne efficacité de liaison et un assez grand nombre de liaisons H, ce qui permet de prévenir l'obésité (Akhila et al., 2020).

Conclusion :

Les phytoconstituants utilisés dans l'étude indique clairement son bénéfique dans la lutte contre l'obésité et confirme l'importance des molécules. De nouveaux médicaments peuvent être conçus en utilisant ces constituants cliniquement sûrs qui agissent par différents mécanismes pour atteindre le même objectif.

III.5. Article de Poornima et Sumathy, 2022

In Silico Interaction Studies of Various Antioxidant Molecules with MAO-B Protein–A Potent Secondary Drug Target in Parkinson's Disease

Objectif :

Comparer, analyser et de déterminer le meilleur médicament adjuvant pour le traitement de la maladie de Parkinson.

Méthode :

- **Recherche de données sur les médicaments anti-MP utilisés dans le cadre d'une thérapie adjuvante**

Dans cette étude, plusieurs médicaments adjuvants anti-MP utilisés comme inhibiteurs de la MAOB ont été extraits de la littérature et des bases de données des banques de médicaments (<https://go.drugbank.com>). Les médicaments approuvés par la FDA qui ont été utilisés comme adjuvants de la lévodopa et également utilisés en monothérapie aux premiers stades de la MP ont été choisis. De plus les bases de données des banques de médicaments ont été utilisées pour recueillir des informations complètes sur les médicaments adjuvants anti-MP utilisés de manière conventionnelle. Plusieurs autres molécules d'origine naturelle ayant des propriétés antioxydantes potentielles ont été choisies pour cette étude et les formes 3D des molécules antioxydantes présélectionnées et des médicaments existants ont été extraits de NCBI (<http://www.ncbi.nlm.nih.gov>).

- **Modélisation et visualisation de la protéine cible**

La base de données Ensembl, principale source de séquences du génome humain, a été utilisée pour obtenir des informations complètes sur le gène de la protéine MAOB cible (<http://www.ensembl.org/>). Cette base de données a été reliée par hyperlien à la base de données uniport qui fournit les séquences complètes d'acides aminés au format Fast Alignment Sequence Test for Application (FASTA) (<https://www.uniport.org/uniport>). De plus CPHmodels-3.0 a été utilisé pour l'étude de modélisation des protéines (<https://www.cbs.dtu.dk/services/CPHmodels>). Il s'agit d'un serveur de modélisation d'homologie à distance à haute définition. Cet outil a prédit un modèle tridimensionnel précis et sans faille de MAOB. Le logiciel Discovery Studio V 2.5, l'outil de visualisation moléculaire, a été utilisé pour visualiser la structure 3D de la MAOB protéine.

- **Arrimage des médicaments et visualisation 3D**

PatchDock a été adopté pour un arrimage facile et précis (<https://bioinfo3d.cs.tau.ac.il/patchdock>). Ce serveur d'amarrage de médicaments moléculaires est entièrement automatisé. Les médicaments adjuvants anti-MP prescrits en clinique et les molécules antioxydants choisies ont été amarrés pour obtenir les scores de liaison des médicaments. De plus le logiciel Discovery Studio a été utilisée pour générer la forme 3D de la carte d'interaction protéine-ligand.

Résultats :

Un serveur automatisé de modélisation homologique CPHmodels-3.0 a prédit une forme 3D parfaite de l'enzyme. La forme 3D prédite de la MAOB a été visualisée à l'aide du logiciel

Discovery studio (figure 47 a). L'énergie de Contact Atomique (ECA) fait référence aux scores d'affinité d'interaction entre le médicament et la protéine cible. Le complexe ancré d' α -Tocophérol et de MAOB (figure 47 k) présentait la conformation la plus stable avec une valeur ECA de -373,83Kcal/Mol. Après le tocophérol, l'acide gras polyinsaturé à longue chaîne (AGPI-LC) a établi une conformation plus stable avec de bonnes valeurs de liaison d'affinité.

Les valeurs ECA des AGPI-LC ω -3 et ω -6 étaient plus élevées que celles des médicaments conventionnels tels que la Rasagiline, la sélégiline, la pargyline, la toloxatone, le moclobémide, la phénelzine, l'isocarboxazide et la tranylcypromine. La carte d'interaction de la tranylcypromine (figure 47 o), un inhibiteur de la MAOB classiquement utilisé, a montré le complexe de conformation le plus faible avec la valeur ECA de -174,00Kcal/Mol. Le complexe de Safinamide et de MAOB (figure 47 d) a révélé la conformation la plus stable parmi les inhibiteurs de la MAOB. Le médicament adjuvant anti-MP utilisé de manière conventionnelle. Il a présenté une valeur ACE de -302,00 Kcal/Mol, mais son affinité de liaison est moindre que celle de l' α - tocophérol, de l'acide docosahexaénoïque (DHA) et de l'acide di-hommo gamma linoléinique (DGLA).

Parmi les acides gras ω -6, le DGLA et la protéine MAOB (figure 47 g) ont montré une conformation plus stable avec une valeur ECA de -314,60 Kcal/Mol et les acides gras ω -3, le DHA (figure 47 j) a révélé un excellent score d'affinité de liaison de -322,77 Kcal/Mol.

Si l'on considère les AGPI-LC, les acides gras individuels respectifs (y compris les ω -3 et ω -6) présentent le score d'affinité de liaison dans l'ordre hiérarchique suivant : ADH (-322,77) > ADGL (-314,60) > AEP (-294,01) > AAR (-273,48) > AAL (-269,60) > AL (-240,46) > AGL(-222,0).

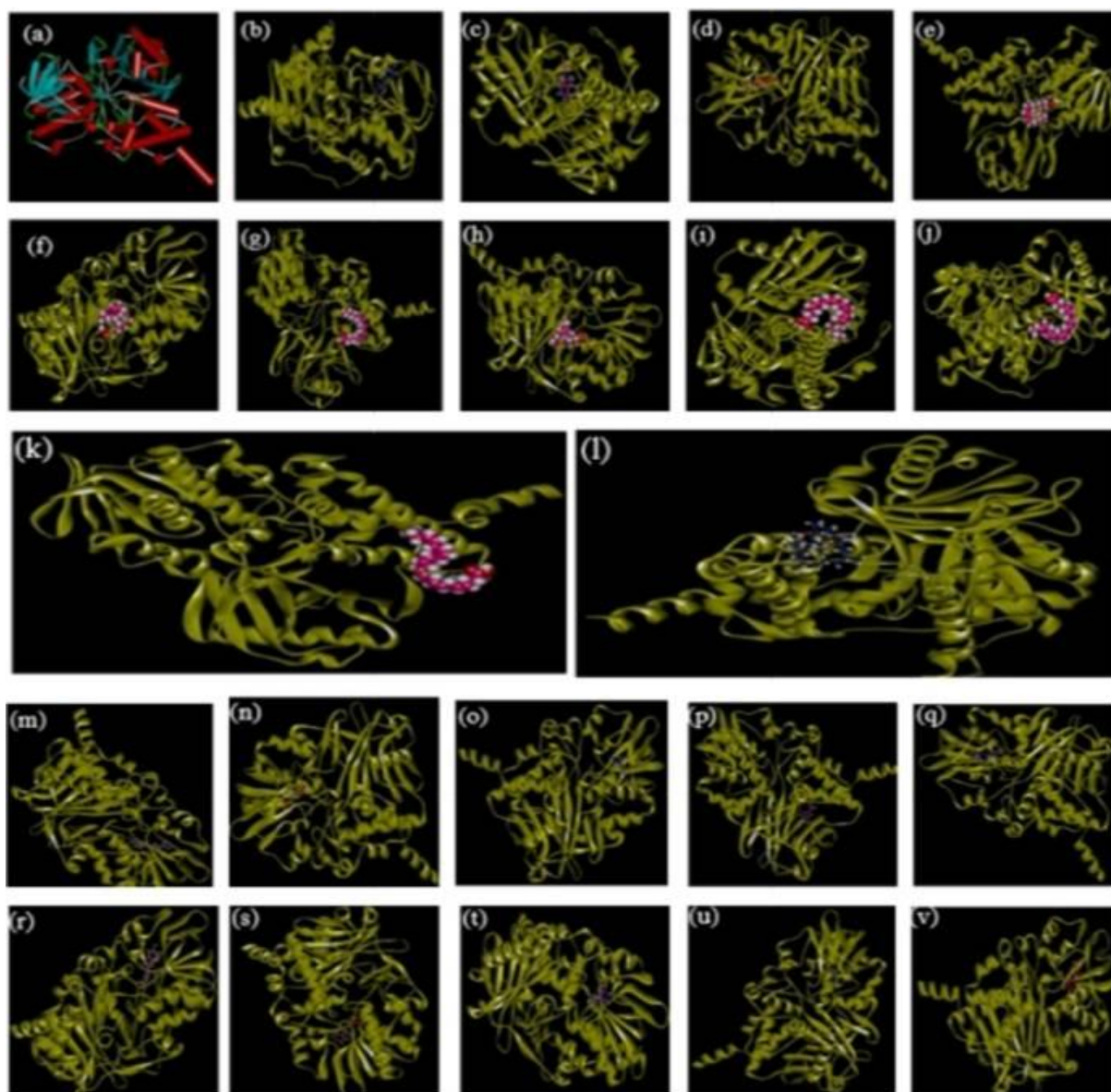


Figure 47 : (a) Forme tridimensionnelle de la structure secondaire de la protéine MAOB, (b-v) Forme tridimensionnelle du complexe ancré de divers inhibiteurs de la protéine MAOB.

Tableau 15 : Le nom du composé, le numéro d'accèsion à la banque de médicaments et la valeur de l'énergie de contact atomique (ECA) (Kcal/Mol) pour les figures (b-v).

Nom du composé	Numéro d'accèsion à la banque de médicaments	La valeur de l'énergie de contact atomique (ECA) (Kcal/Mol)
(b) Sélégiline	(DB01037)	-233,91
(c) Rasagiline	(DB01367)	-251,96
(d) Safinamide	(DB06654)	-302
(e) Acide linoléique (AL)	(DB14104)	-240,46

(f) Acide alpha-linolénique (AAL)	(DB00132)	-269,60
(g) Acide gamma-linolénique (AGL)	(DB13854)	-222,00
(g) Acide di-homo gamma linoléique (ADGL)	(DB00154)	-314,60
(h) Acide arachidonique (AAR)	(DB04557)	-273,48
(i) Acide eicosapentaénoïque (AEP)	(DB00159)	-294,01
(j) Acide docosahexaénoïque (ADH)	(DB03756)	-322,77
(k) Alpha tocophérol	(DB1125)	- 373,83
(l) Glutathione	(DB00143)	-284.61
(m) Isocarboxazide	(DB01247)	-267,55
(n) Phénelzine	(DB00780)	-199,55
(o) Tranlycypromine	(DB00752)	-174,70
(p) Génistéine	(DB01645)	-272,52
(q) Daidzéine	(DB13182)	-266,87
(r) Moclobémide	(DB01171)	-266,73
(s) Pargyline	(DB01626)	-203,89
(t) Safrazine	(DB09253)	-225,99
(u) Toloxatone	(DB09245)	-225,24

Discussion :

La monoamine oxydase B (MAO-B) est une protéine majeure du métabolisme des monoamines, oxyde facilement la dopamine, ce qui la rend insuffisante pour traiter les patients atteints de la maladie de Parkinson.

De nombreux chercheurs ont révélé l'efficacité de l' α -tocophérol dans la protection du système nerveux central, L' α -tocophérol a donc un fort potentiel de protection du système nerveux central.

Le cerveau est un organe riche en lipides et les AGPI-LC sont réputés pour protéger le système nerveux. C'est pourquoi les AGPI-LC ont été pris en compte de manière significative et ont fait l'objet d'une attention particulière dans la présente étude. Les AGPI-LC piègent les radicaux libres de l'oxygène et sont les principaux précurseurs des leucotriènes, des prostaglandines, une carence en ces molécules inévitables pourrait être à l'origine de l'incidence idiopathique de la MP. D'autre part, le manque de molécules antioxydants à fort potentiel pourrait être la raison sous-jacente de l'augmentation du stress oxydatif dans le système nerveux, conduisant à une MP induite par une mutation. De ce fait les AGPI-LC constituent un complément alimentaire efficace dans le traitement de la MP, qui atténue efficacement le développement de la MP.

Des recherches antérieures ont porté sur diverses méta-analyses des études in vitro et des essais cliniques qui ont recommandé la consommation nécessaire d'AGPI oméga 3 pour réduire le risque de la MP. L'amarrage moléculaire des AGPI et leurs scores de liaison avec la MAOB sont restés inconnus. Par conséquent, l'étude actuelle a montré la comparaison détaillée des molécules individuelles d'AGPI avec la protéine cible MAOB (**Poornima et Sumathy, 2022**).

Conclusion :

Comme le nombre de cas de MP augmente, les neurologues recherchent une thérapie médicamenteuse efficace pour soulager l'inconfort des patients. Cette méthode in-silico s'est avérée bénéfique pour comparer, prédire et déterminer le meilleur composé naturel avec d'excellents scores de liaison d'affinité. La consommation d'AGPI oméga 3 et d' α -tocophérol a été considérée comme sûre et efficace pour traiter les stades initiaux et modérés de la thérapie de la MP. Les résultats obtenus dans le cadre de ces in-silico doivent être corrélés avec des études analytiques et in-vivo.

III.6. Article de Zhang et al., (2022)

Chromenone-based GSK-3 β inhibitors as potential therapeutic targets for cardiovascular diseases: In silico study, molecular dynamics, and ADMET profiles.

Objectif :

Etudier l'évaluation d'inhibiteurs naturels/synthétiques spécifiques et sélectifs de GSK-3 β pour ouvrir la voie au traitement de la régénération cardiaque, de l'inflammation et de leur impliquait des mécanismes sous-jacents via l'amarrage moléculaire.

Méthode :

▪ Dépistage virtuel

L'amarrage moléculaire a été utilisé pour cribler la base de données de composés PubChem afin d'identifier les molécules interagissant avec GSK-3 β en appliquant le logiciel AutoDock.

▪ Préparation de la structure cible

La modélisation par homologie a été appliquée pour construire un récepteur 3D optimisé de GSK-3 β . De plus la séquence protéique de GSK-3 β dans les mitochondries cardiaques humaines (Acc :HGNC : 4617) est extraite de la base de données d'ensemble (<https://www.ensembl.org>) au format FASTA et considérée comme la séquence GSK-3 β dans cette étude. Utilisation de quatre modèles similaires (1J1B-1A, 1J1B-B, 6HOU, 3M1S).

▪ Préparation de ligands

-La structure 3D des composés naturels et synthétiques a été recherchée sur la base de la structure similaire de la morine (Figure 48) En tant qu'échafaudage de chroménone et de leurs caractéristiques médicinales.

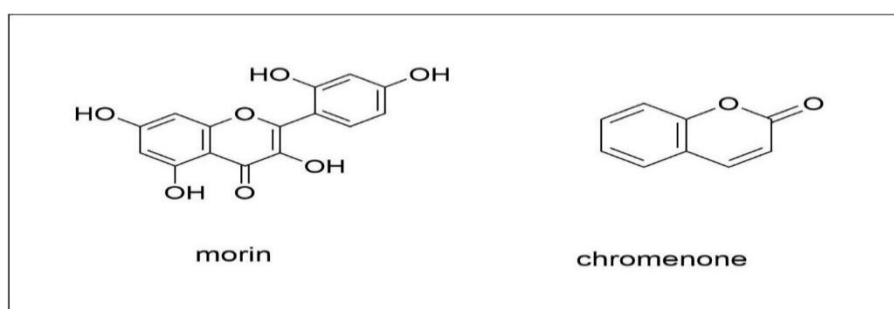


Figure 48 : Structure chimique du morine et de la chroménone (Das et al., (2023).

-5640 composés ont été obtenus à partir de deux chimiothèques comme la base de données PubChem (<https://www.pubchem.ncbi.nlm.nih.gov>) et les structures ont été récupérées au format SDF et converties au format PDB. Tous les composés collectés ont été optimisés par la théorie de la fonctionnelle de la densité (DFT) et l'énergie minimisée à la fonctionnelle B3LYP et 6-31 + G ** ensemble de bases en utilisant le paquet Gaussian 09.

▪ Amarrage moléculaire

La protéine modélisée a été préparée par les étapes suivantes :

-Ajouter des charges partielles d'atomes de Kollman et polaires et l'élimination de la molécule d'eau et des atomes d'hydrogène non polaires fusionnés.

- une boîte de grille de 60x60x60 avec un espacement des points de 0,375 Å a été calculée autour du site actif dans GSK-3 β pour fournir le module autogrille.

-150 analyses GA ont été mises en œuvre à l'aide de l'algorithme génétique lamarckien.

- La combinaison Maestro 11.0 Schrodinger a été utilisée pour visualiser la démonstration 2D et 3D.

▪ Dépistage virtuel

Pour étudier de nouveaux composés naturels et synthétiques potentiels pour inhiber GSK-3 β , une simulation d'amarrage moléculaire in silico a été réalisée. Pour valider l'amarrage moléculaire précis de la structure protéique modélisée et le comparer à l'affinité de liaison expérimentale, nous avons exécuté la procédure de redocking. L'ANP (ester d'adénylate d'acide aminophosphonique phosphoreux) en tant que ligand apparenté a été utilisé pour évaluer le protocole d'amarrage moléculaire et les résultats de réamarrage ont indiqué une valeur RMSD de 1,46 Å.

Sur la base de la structure de similarité avec la morine en tant qu'inhibiteur de GSK-3 β , 5640 composés ont été extraits de la base de données PubChem et tous les composés sélectionnés suivent la règle des cinq de Lipinski (tableau 16). Pour étudier l'inhibition potentielle des composés à interagir avec la poche de liaison de GSK-3 β , une simulation d'amarrage moléculaire a été réalisée. Par conséquent, sur la base de l'énergie libre de liaison la plus faible, les 13 meilleures molécules ancrées ont été sélectionnées pour une analyse plus approfondie (Figure 50).

Résultats :

Comme illustré sur la figure 49

90,5 % des résidus d'acides aminés étaient dans la région favorisée, 8,8 % du total des résidus dans la région autorisée et seulement 0,6 % des résidus d'acides aminés dans la région non autorisée.

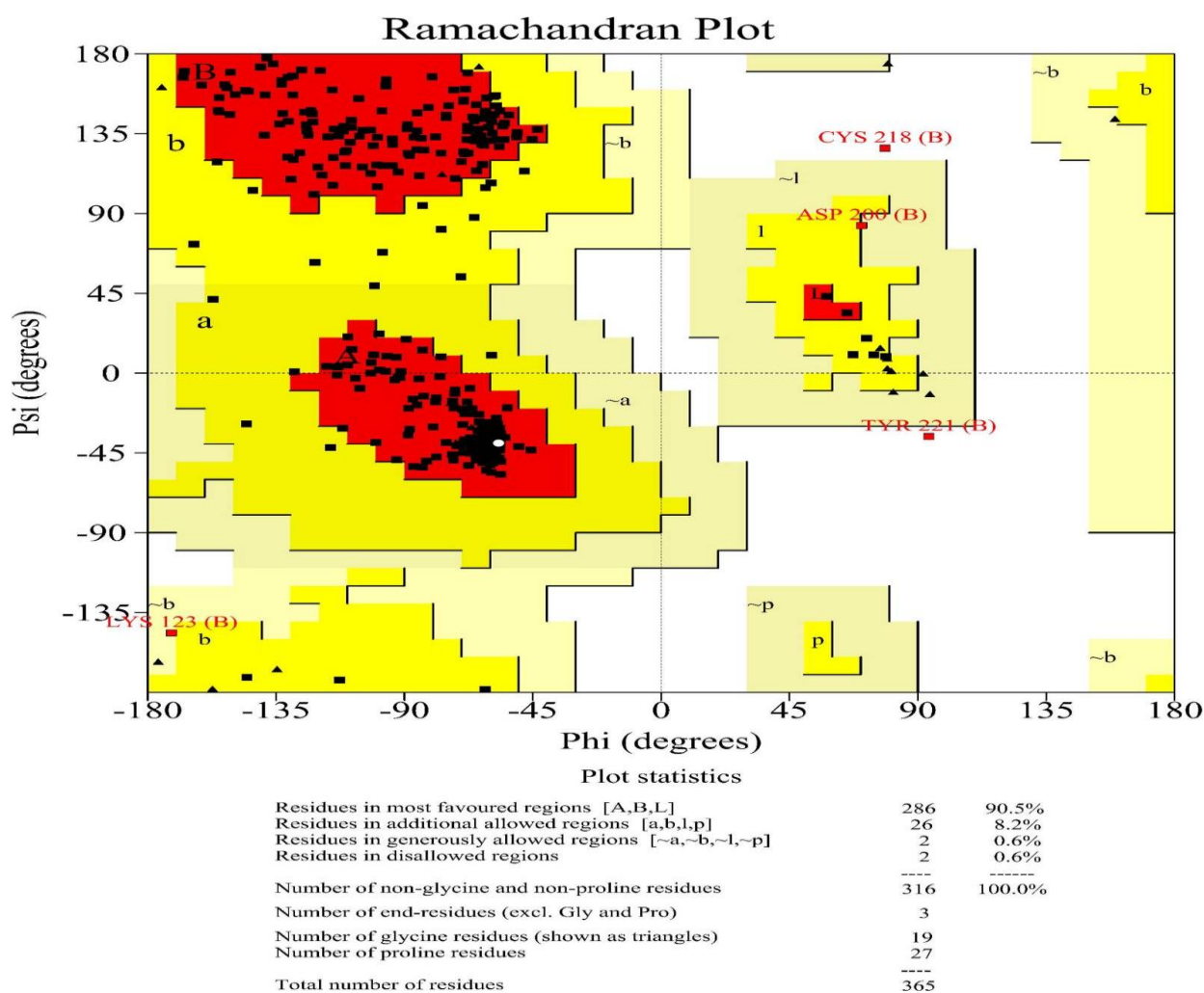


Figure 49 : Diagramme de Ramachandran pour la protéine modèle d'homologie (Das et al., (2023)).

Tableau 16 : Résultats de la simulation d'amarrage moléculaire pour les composés sélectionnés et le récepteur modélisé (GSK-3 β) (Das et al., (2023)).

Composé		ID APB : GSK3B	
NON	Pubchem-CID	EB (kcal/mol)	Ki (μ M)
1	127039278	-11,4	3.1
2	44378184	-11,1	3.9
3	1608168	-10,9	4.3
4	15478904	-10,2	5.1
5	12019530	-10,1	5.4

Composé		ID APB : GSK3B	
NON	Pubchem-CID	EB (kcal/mol)	Ki (μ M)
6	14309760	-9,9	6.2
7	44258296	-9,8	6.3
8	52949348	-9,7	6.7
9	10617404	-9,5	7.2
10	4020240	-9,4	7.8
11	14583584	-9,3	8.0
12	53305954	-9,1	8.4
13	5281670	-9,8	6.1

Abréviation : BE : Binding Energy (énergie de liaison)

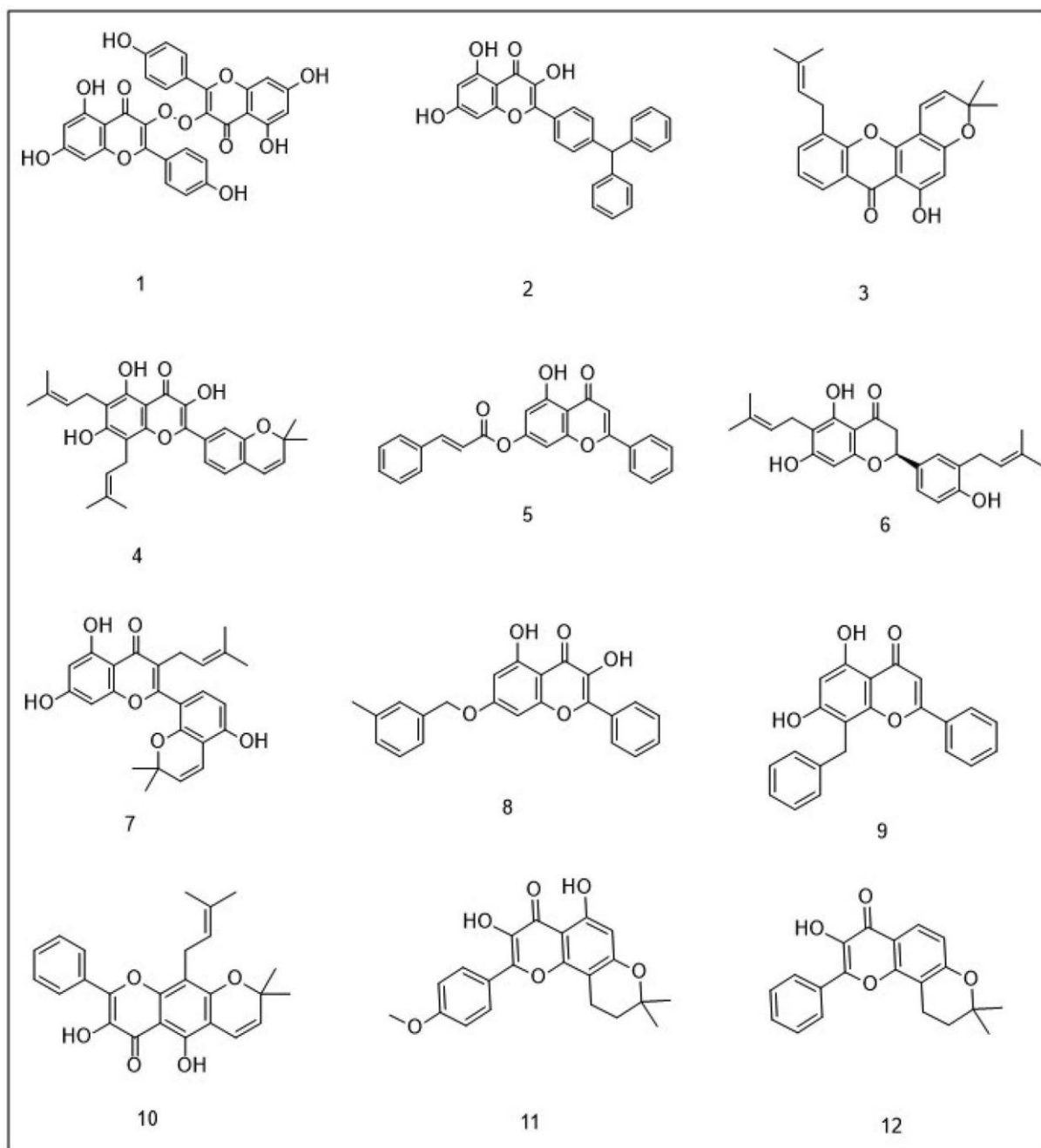


Figure 50 : Structure chimique des 12 principaux dérivés de chroménone (**Das et al., (2023)**).

-Comme présenté dans le tableau 16

13 composés supérieurs avec l'affinité de liaison la plus faible ont interagit avec la poche de liaison de GSK-3 β par des interactions non covalentes et montre que le composé 1 en tant que dérivé du chroménone avait l'énergie libre de liaison la plus faible.

-Comme le montre la figure 51

- Le composé 1 a interagi avec des résidus d'acides aminés comme Leu132, Cys199, Ala83, Val70, Val87, Val69, Phe67 et Leu188 par interaction hydrophobe. De plus, le composé 1 a été établi dans le site actif de la cible par sept interactions de liaison H. Le groupe carbonyle de la chroménone du composé 1 a interagi avec Lys85 à une distance de 2,88 Å, et le composé a également formé une liaison H avec les résidus Ser66, Lys183, Gln185, Val69 et Lys86 avec le groupe hydroxyle du récepteur à une distance de 2,76, 2,52, 3,14, 2,77, 2,99 et 2,81 Å, respectivement. Le composé a également interagi avec les résidus de Asp133, Asp200, Lys85, Lys86 et Lys183 en utilisant des interactions chargées. La fraction chroménone du composé 1 est couverte par différentes interactions non covalentes indiquant que ces interactions jouent un rôle important dans l'inhibition de GSK-3β.

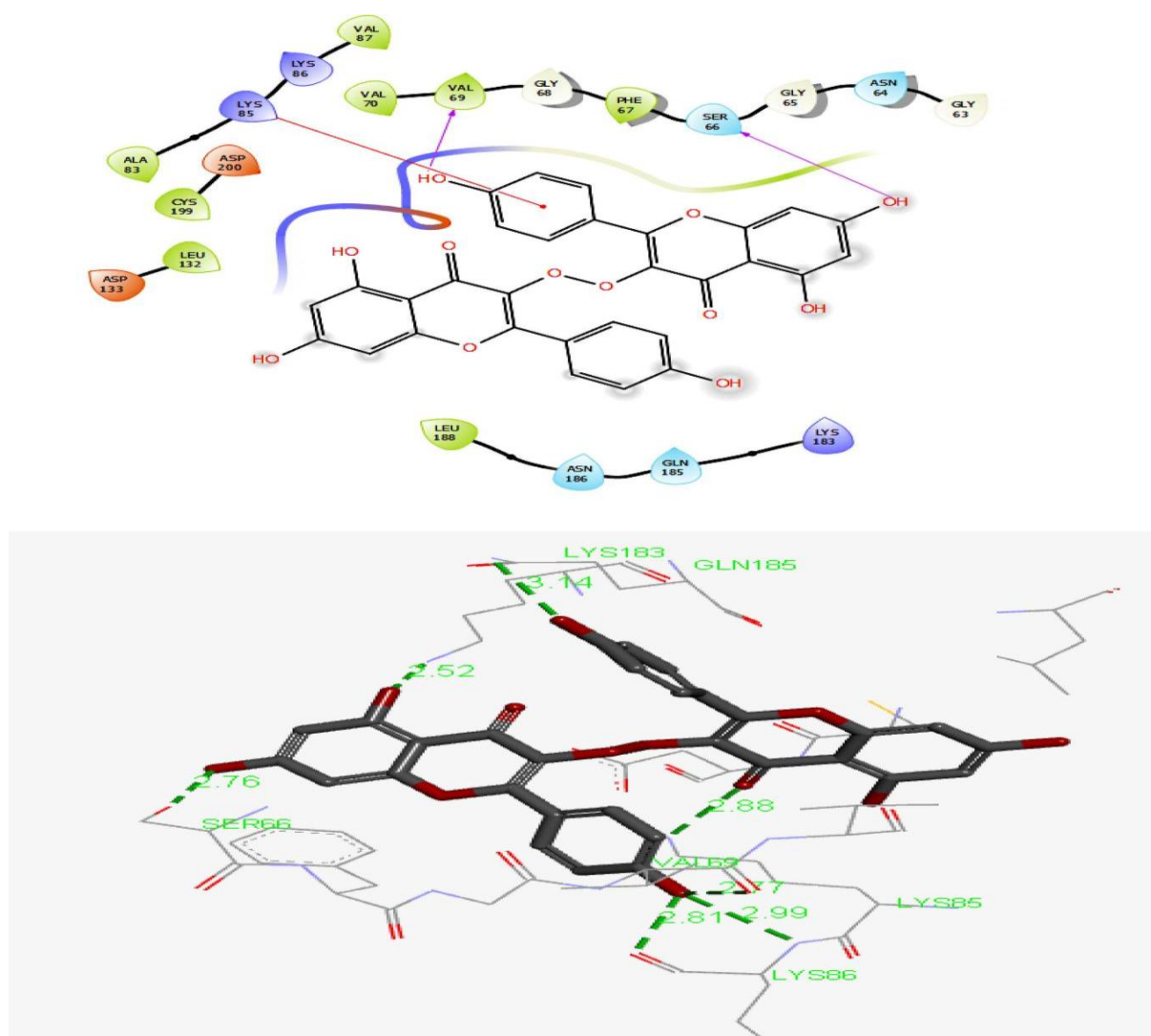


Figure 51 : Présentation des modèles 2D et 3D des interactions entre le composé1 et la protéine modélisée (GSK-3β) (Das et al., (2023)).

-Comme le montre la figure 52

L'analyse d'amarrage moléculaire a révélé que le composé 2, constitué d'un fragment chroménone, présentait un excellent score d'amarrage de -11,1 kcal/mol et a présenté une interaction hydrophobe avec des résidus d'acides aminés tels que Val70, Phe67, Ile62, Cys199, Ala83, Val110, Leu132, Tyr134, Val135, Pro136 et Leu188 dans le site actif de la cible. De plus, trois hydrogènes liés entourent le fragment chroménone du composé dans le site de liaison du récepteur. Les résidus d'acides aminés tels que Lys183, Asn186 et Gly68 interagissent avec le composé à des distances de 3,29, 2,85 et 2,79 Å, respectivement. Ce composé a une affinité notable pour la structure de GSK-3 β avec les résidus Asp200, Asp133, Glu137, Arg141 et Lys85 par le biais d'interactions électrostatiques. Par conséquent, il pourrait être proposé comme inhibiteur potentiel de la GSK-3 β .

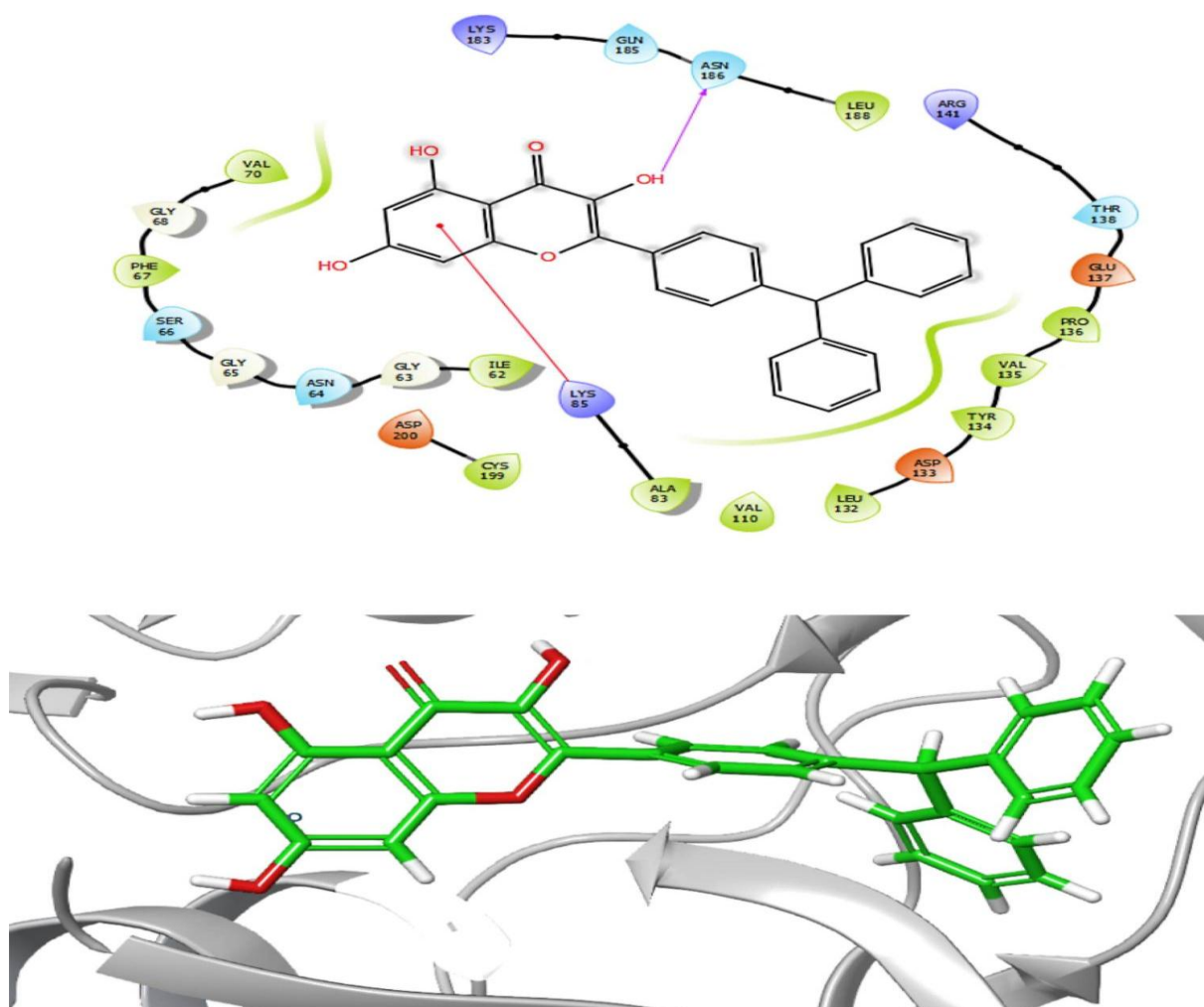


Figure52 : Présentation des modèles 2D et 3D des interactions entre le composé 2 et la protéine modélisée (GSK-3 β) (Das et al., (2023)).

Comme le montre la figure 53

Le composé 3 était établi dans le site actif, des résidus comme Pro136, Val135, Tyr134, Leu132, Ala83, Val110, Leu188, Cys199 et Ile62 ont formé une interaction hydrophobe avec le ligand. De plus, une énergie libre de liaison de -10,9 kcal/mol est présentée dans le tableau 16 interagissant avec la liaison hydrogène avec les résidus Asp133 et Val135 avec le groupe hydroxyle de la chroménone du composé à une distance de 2,92 et 2,98 Å, respectivement. Le composé 3 a interagi avec des résidus tels que Glu137, Lys85, Asp200, Arg141 et Lys85 en utilisant des interactions électrostatiques.

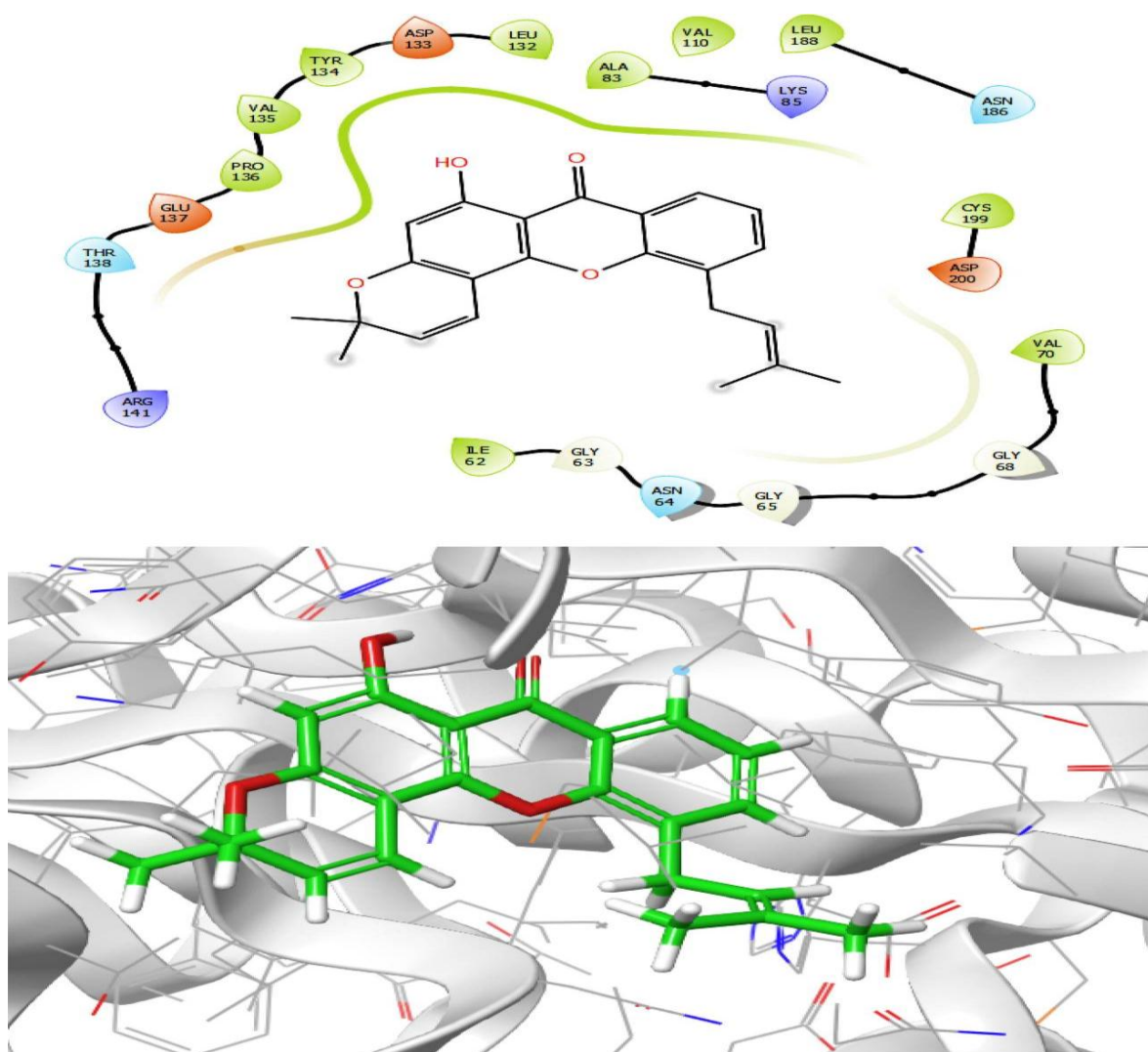


Figure 53 : Présentation des modèles 2D et 3D des interactions entre le composé 3 et la protéine modélisée (GSK-3 β) (Das et al., (2023)).

Discussion :

Sur la base d'enquêtes antérieures, l'inhibition de GSK-3 β agit comme un facteur sous-jacent dans les maladies cardiovasculaires. Nous implémentons le graphique de Ramachandran pour évaluer la qualité de la protéine modélisée et l'intégrité structurelle. De plus, cette structure 3D a été analysée à l'aide du serveur PROCHECK pour estimer la qualité stéréochimique du récepteur, indiquant que 99,4 % du total des résidus d'acides aminés avaient une géométrie précise. Par conséquent, la structure protéique modélisée générée sur la base de la méthode de modélisation par homologie est fiable et peut être utilisée pour la conception de médicaments et la recherche de candidats-médicaments prometteurs. Par conséquent, l'analyse de la structure chimique a indiqué que les composés constitués du groupement chroménone avaient une affinité de liaison élevée pour la poche de liaison de la protéine modélisée et occupaient le site actif de la cible. De plus, ces composés inhibent les résidus d'acides aminés clés dans le mode de liaison de la protéine à travers les différents types d'interactions (électrostatique, polaire, liaison hydrogène et hydrophobe).

Fait intéressant, la plupart des composés présentés dans le tableau 16 sont plus convaincants que la morine en tant que ligand de référence (Pubchem-CID : 5281670). Il semble que cet effet des composés mentionnés soit associé à l'existence de substitués de donneurs d'électrons ou d'extenseurs de résonance électronique dans la chroménone fragment, en particulier dans le cycle pyrone. De plus, en comparant la structure des composés, certains points sont importants dans leur potentiel d'inhibition de la GSK-3 β : la présence de groupement hydroxyle ou de substitution oxygénée dans les carbones 5 et 7 de la chroménone, pour être des groupements hydroxyle libres, la résonance électronique doit être facilement accessible par stéréochimie appropriée des substituants. Par conséquent, les composés constitués d'un groupement chroménone pourraient être considérés comme des inhibiteurs potentiels prometteurs de GSK-3 β (Das et al., (2023)).

Conclusion :

Le glycogène synthase kinase-3 (GSK-3) est une sérine/thréonine protéine kinase cytoplasmique identifiée comme un régulateur essentiel dans plusieurs voies de signalisation différentes de la pathogenèse comme le diabète de type 2, la maladie d'Alzheimer et le métabolisme cardiaque, en particulier du glycogène. GSK-3 β agit comme un inhibiteur compétitif qui inhibe efficacement l'ATP poche de liaison de la protéine. L'inhibition de GSK-3 β dans le tissu cardiaque entraîne une diminution de l'hypertrophie et de la nécrose des

cardiomyocytes dans les cardiopathies ischémiques, la régénération cardiaque et l'insuffisance cardiaque.

III.7. Article de Das et al., 2023

In silico studies and evaluation of in vitro antidiabetic activity of berberine from ethanol seed extract of *Coscinium fenestratum* (Gaertn.) Colebr.

Objectif :

Evaluer la puissante activité antidiabétique de l'extrait éthanolique de *Coscinium fenestratum* (CF) après confirmation par une étude d'amarrage moléculaire in silico.

Méthode :

La méthode Soxhlet a été utilisée pour la préparation de l'extrait éthanolique. En outre, un criblage phytochimique et une CCM ont été effectués pour détecter le composé responsable de l'activité antidiabétique. La structure accessible de l'extrait de plante a été utilisée dans une enquête d'amarrage in silico basée sur cette prédiction. Pour l'étude de l'amarrage du diabète, la protéine alpha-amylase (APH) avec PDB ID : 4X9Y a été choisie. Après cela, des techniques de test d'inhibition de l'alpha-amylase et de l'alpha-glucosidase ont également été utilisées pour tester l'activité antidiabétique in vitro. La berbérine a été détectée et identifiée par CCM avec un Rf de 0,37 et s'est avérée responsable de l'activité antidiabétique.

▪ **L'étude d'amarrage in silico :**

○ **Préparation de la cible**

-La protéine alpha-amylase pancréatique humaine (APH) a été obtenue auprès du RCSB PDB, une banque de données sur les protéines et a été sélectionnée pour les études d'amarrage du diabète. Il a été obtenu au format pdb et la protéine a été convertie en cible par élimination des molécules d'eau, minimisation de l'énergie, ajout de charges telles que Kollmann et les atomes manquants de la protéine ont été ajoutés. Pour la minimisation de l'énergie, le logiciel MOE2018 a été utilisé et pour les études d'amarrage moléculaire, AutoDock4.2 a été utilisé.

○ **Préparation de ligand**

-La berbérine a été obtenue à partir du Chem draw 3D et enregistrée au format sdf. De plus Openbabelsoftware a été utilisé pour convertir les ligands du format sdf au format pdb. Plus

tard, en utilisant AutoDock4.2, le critère d'aromaticité et plusieurs torsions ont été définis et enregistrés au format pdbqt.

Résultats :

L'énergie de liaison de la berbérine s'est avérée être de -7,84. De plus, des liaisons hydrogène se sont formées entre la molécule cible et le ligand. Pour la berbérine, il existe cinq liaisons hydrogène et trois liaisons alkyles (figure 54).

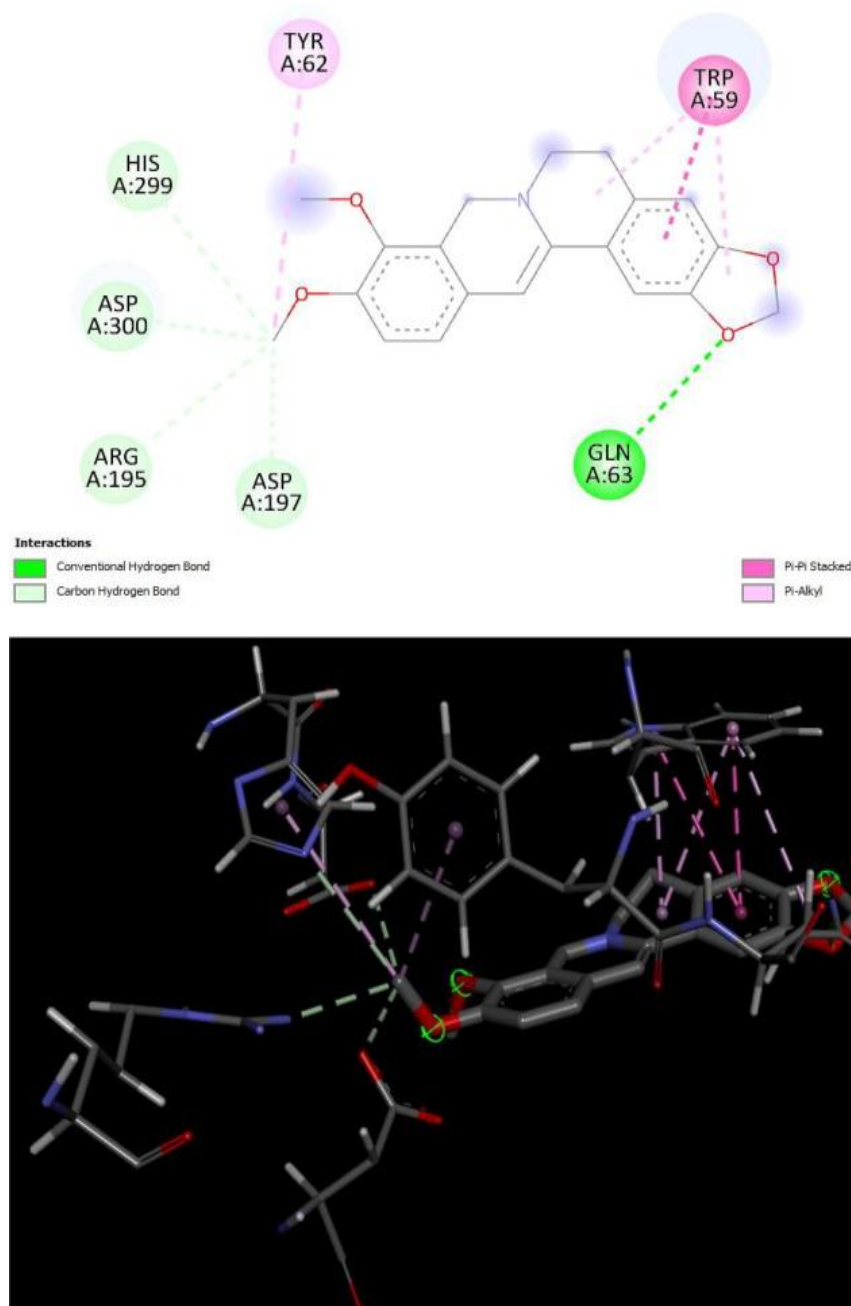


Figure 54 : Structure 2D et 3D de la liaison de la berbérine avec la protéine PDB ID : 4X9Y (Das et al., (2023)).

Discussion :

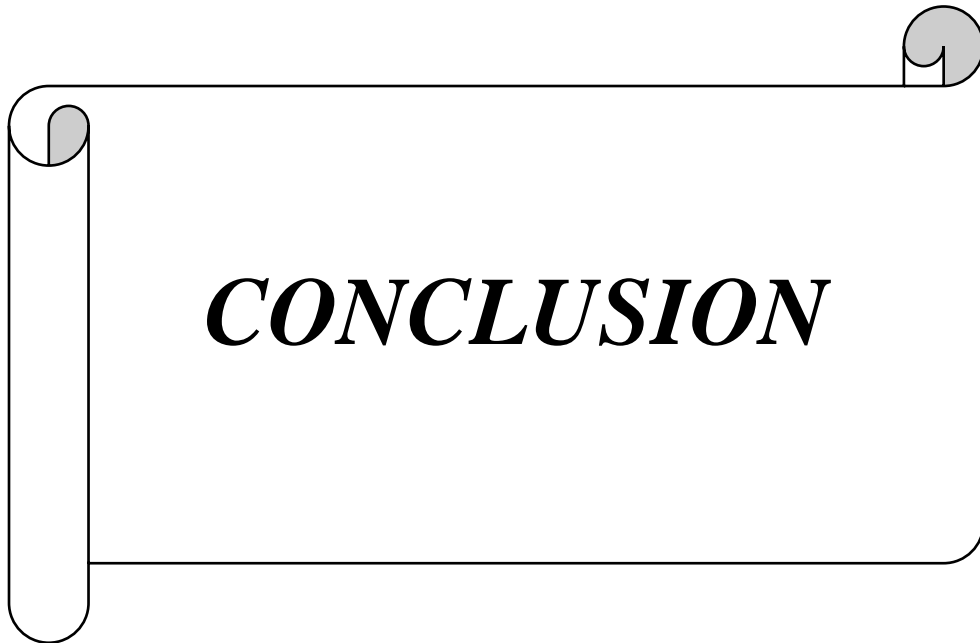
α Amylase est une endoglucosidase qui hydrolyse les polymères d' α -D-glucose au niveau des liaisons 1-4. Sécrétée majoritairement par les glandes salivaires et le pancréas, cette enzyme a pour rôle la digestion de l'amidon présent dans le bol alimentaire

L'un des traitements les plus efficaces contre le diabète consiste à réduire la quantité de glucides absorbés après avoir mangé. Pour être absorbés dans le duodénum et le jéjunum supérieur, les amidons complexes, les oligosaccharides et les disaccharides doivent d'abord être décomposés en monosaccharides par l' α -amylase et les α -glucosidases. La plupart des produits naturels agissent en inhibant ces enzymes nécessaires à la décomposition des polysaccharides en monosaccharides, prévenant ainsi l'hyperglycémie postprandiale. Les effets antidiabétiques sont présents dans les plantes en raison des flavonoïdes, des terpénoïdes, des saponines, des caroténoïdes, des alcaloïdes et des glycosides. Plus précisément, des substances bioactives individuelles telles que la palmatine, la berbérine, l'honokiol, les amorfrutines, la trigonelline, les acides gymnémiques, la gurmarine et la phlorizine ont démontré un potentiel antidiabétique. La CCM et des tests chimiques spécifiques ont confirmé la présence de berbérine, le principal ingrédient actif de l'extrait de graines de FC. Il a été démontré que la berbérine faisait partie des alcaloïdes.

Des études antérieures ont soutenu l'idée que la berbérine joue un rôle important dans la prévention du diabète par de multiples mécanismes. Ces mécanismes comprennent une utilisation accrue de l'insuline, une modulation du microbiote intestinal sans activité anti-infectieuse systémique, et la stimulation de l'absorption du glucose de façon marquée. La structure chimique de la berbérine et des autres alcaloïdes isoquinoléiques est très différente de celle des autres médicaments hypoglycémisants couramment utilisés, tels que les sulfonylurées, les biguanides, les thiazolidinediones ou l'acarbose. La berbérine et les alcaloïdes isoquinolines similaires sont connus pour avoir des propriétés antidiabétiques. Il est possible que l'amélioration du métabolisme du glucose induite par la berbérine soit due à la stimulation de la glycolyse. De plus, la berbérine a le potentiel d'inhiber l' α -glucosidase. Les activités de la disaccharidase ont été inhibées et la quantité de glucose transportée à travers l'épithélium intestinal a été réduite (Das et al., 2023).

Conclusion :

L'extrait de graines à l'éthanol de CF peut avoir un effet antidiabétique en raison de la présence de berbérine dans l'extrait comme l'un des composés phytochimiques. Cependant, des recherches supplémentaires sont nécessaires pour déterminer les effets synergiques ou antagonistes potentiels entre les divers constituants de l'extrait de graines de CF.



CONCLUSION :

Notre travail avait pour objectif d'acquérir des compétences en docking moléculaire afin d'étudier de différentes cibles thérapeutiques pour le traitement des maladies chroniques.

D'après les résultats de l'amarrage moléculaires nous permettant d'identifier différentes molécules ciblées tels que les composants bioactifs, les substances phytochimiques et les phytoconstituants en déterminant les interactions entre deux molécules et pour trouver la meilleure orientation du ligand qui formera un complexe avec une énergie globale minimale, l'amarrage moléculaire des composants a été validé avec succès contre les récepteurs cibles.

Le docking moléculaire est devenue plus en plus un outil important pour la découverte de médicaments. Ces méthodes in-silico sont avérées bénéfiques pour comparer, prédire et déterminer le meilleur composé naturel avec d'excellents scores de liaison d'affinité. Les études in silico peuvent utiliser avant les études in vivo et in vitro afin de prouver que l'expérience est valable ou après pour confirmer les résultats de l'étude. Certaines études in silico doivent être corrélées avec des études in vitro et in vivo car ils sont complémentaires à l'expérimentation pour prouver l'efficacité des ressources naturelles de médication.



***Références
bibliographiques***

A

Aditi, S., Rahul, S., Pavan Kumar A., Icx K. (2017). Bioinformatics Database Resources. In book: Biotechnology (pp.84-119).

Agrawal, P. (2013). Conception de médicaments basée sur la structure. Journal of Pharmacovigilance. 1 (04), 1-2.

Akhila, S., Anu, V., Preeja, G, P. (2020). An in-silico approach to curtail the action of obesity proteins with phytoconstituents. International Journal of Current Advanced Research. Vol 9, Issue 03(A), pp 21479-21481.

Almeida Paiva, V., Souza Gomes, I., Monteiro, C. R., Mendonça, M. V., Martins, P. M., Santana, C. A., Azevedo Silveira, S. (2022). Protein structural bioinformatics: An overview. Computers in Biology and Medicine. 105695.

Anuj, T. (2020). Biological Databases. In book: e-Training on Biocomputational Interventions to Analyse Canine & Livestock Genomes.

Aslam, B., Basit, M., Nisar, M. A., Khurshid, M., & Rasool, M. H. (2016). Proteomics: technologies and their applications. Journal of chromatographic science. 1-15.

B

Babar, M., Pervez, M., Nadeem, A., Hussain, T., Aslam, N. (2014). Multiple Sequence Alignment Tools: Assessing Performance of the Underlying Algorithms. Journal of Applied Environmental and Biological Sciences. 4(8S)76-80.

Bairoch, A., & Apweiler, R. (1998). The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1998. *Nucleic Acids Research*, 26(1). 38-42.

Bérenger, F. (2016). Nouveaux logiciels pour la biologie structurale computationnelle et la chémoinformatique (Doctoral dissertation, Paris, CNAM).

Bouchagra, S. (2018). Modélisation des interactions protéine-petites molécules : étude de la relation structure- fonction dans le cas des lipases (Thèse de doctorat). Université Badji Mokhtar. Annaba.

Brillouet, S., Dorbes, S., Silvente-Poirot, S., Mestre-Voegtlé, B., Picard, C., Poirot, M., & Courbon, F. (2010). Ciblage peptidique en oncologie nucléaire : intérêt de la modélisation moléculaire. *Médecine nucléaire*, 34(5), 289-294.

C

Cardinaux, R., Cohidon, C., Guessous, I., Chenal, R., Widmer, D., Cerqui, D., ... & Desvergne 1, B. (2021). Médecine personnalisée et prévention des maladies chroniques : l'attitude des médecins généralistes. *Santé publique*, 33(1), 121-126

Chaudhary, KK., Mishra, N. (2016). Une revue sur l'amarrage moléculaire : nouvel outil pour la découverte de médicaments. *Bases de données*, 3 (4), 1029.

Cherghi, Y., Benfadel, D. (2020). Système de classification automatique des insectes. Université des frères Mentouri Constantine.

Cherid, F., Cheraitia, N. (2022). Rôle des méthodes in Silico, in Vitro et in Omic dans l'évaluation de la pharmacovigilance des médicaments (Doctoral dissertation, Université de Jijel).

Chevrollier, N. (2019). Développement et application d'une approche de docking par fragments pour modéliser les interactions entre protéines et ARN simple-brin (Doctoral dissertation, Université Paris-Saclay (ComUE)).

Coutouly, G., Klein, E., Barbieri, E., Kriat, M., (2012). Travaux dirigés de biochimie, biologie moléculaire et bioinformatique. 3^{ème} édition. France.

Chowdhary, M., Rani, A., Parkash, J., Shahnaz, M., & Dev, D. (2016). Bioinformatics: an overview for cancer research. *Journal of Drug Delivery and Therapeutics*, 6(4), 69-72.

D

Dar, A. M., Mir, S. (2017). Molecular docking: approaches, types, applications and basic challenges. *J Anal Bioanal Tech*, 8(2), 1-3.

Das, D., Kumar, D., Kumar, P., Edash, B. (2020). L'amarrage moléculaire et son application à la recherche d'agent antifalciforme à partir de *Carica papaya*. *Journal of Applied Biology and Biotechnology*, 8 (1), 105-116.

Das, K., Iyer, K. R., Orfali, R., Asdaq, S. M. B., Alotaibi, N. S., Alotaibi, F. S., Ghoneim, M. (2023). In silico studies and evaluation of in vitro antidiabetic activity of berberine from ethanol seed extract of *Coscinium fenestratum* (Gaertn.) Colebr. *Journal of King Saud University-Science*, 35(5), 102666.

David, W. (2004). *Bioinformatics: sequence and genome analysis*. Cold Spring Harbor Laboratory press.

Deore , AB, Dhumane, JR, Wagh, R. et Sonawane, R. (2019). Les étapes du processus de découverte et de développement de médicaments. *Journal asiatique de recherche et de développement pharmaceutiques*, 7 (6), 62-67.

Diniz, W., Canduri, F. (2017). *Bioinformatique : un aperçu et ses applications*. *Genet Mol Res*, 16 (1), 17.

E

Essie, O., Mary, E., Anthony, J., Peter, N. (2016). Sequence retrieval and analysis: A useful tool in bioinformatics- A review *European journal of pharmaceutical and medical research*. 3(1), 76-83.

El Fahime, E., Ennajie, M, M. (2007). Évolution des techniques de séquençage. *Les technologies de laboratoire*, 2(5).

G

Gilbert, D., Manolo, G. (2013). *Bioinformatique cour et cas pratique*. 2ème édition. France.

H

Hammad, S. (2020). Conception par bioinformatique et modélisation moléculaire de nouvelles molécules bioactives dans le domaine du cancer (Doctoral dissertation, Université Mouloud Mammeri Tizi-Ouzou).

I

Iqbal, N., Kumar, P. (2023). De la science des données à la bioscience : ère émergente des applications, outils et défis bioinformatiques. *Procedia Computer Science*, 218, 1516-1528.

J

Journot, G., C, Copin., S, Couillard. (2017). Le Développement Chimique (1/3) : de la découverte d'un principe actif à la commercialisation d'un médicament.

K

Kamble, A., Khairkar, R. (2016). Basics of bioinformatics in biological research. International Journal of Applied Sciences and Biotechnology, 4(4), 425-429.

Karthic, V. M., Poongodi, B., Shanmugapriya, P., & Sivaraman, D. (2019). In-silico molecular docking analysis of potential phytotherapeutics from the medicinal herb Corallocarpus Epigaeus for treating urticaria. Int. J. Trans. Res. Ind. Med, 1, 5-12.

Kettani, A., Mikou, A. (2008). La Modélisation moléculaire, un outil de laboratoire précieux. Les technologies de laboratoire, 3(9).

Kose, A. H. P., Patel, P., & Wani, S. (2019). Association of herbal plant in Rheumatoid Arthritis studies using in silico approach.

Koshak, A. E., Koshak, E. A. (2020). Nigella sativa L as a potential phytotherapy for coronavirus disease (2019): A mini review of in silico studies. Current Therapeutic Research, 93, 100602.

Kumar Singh, S. (2019). Bioinformatics: Concepts and Applications. In book: Advances in Horticultural Crop Management and Value Addition (Eds 1), Chapter 12. India, Laxmi Publications Pvt Ltd, pp 127-136

Kumar, P. (2021). Interaction of Protein-Ligand: Molecular Docking, A novel computational biology tool. Annals of RSCB, 25(6), 20763-20775.

Kumar S., Yadav, P., Mishra, P., Shameem, A., Kumar, Y, D. (2021). Drug Discovery. International Journal of Creative Research Thoughts. Volume 9, 2320-2882

L

Lamoril, J., Ameziane, N., Deybach, J, C., Bouizegarène, P., & Bogard, M. (2008). Les techniques de séquençage de l'ADN : une révolution en marche. Première partie. Immuno-analyse & Biologie Spécialisée, 23(5), 260-279.

Li, W., Chen, H., Xu, B., Wang, Y., Zhang, C., Cao, Y., Xing, X. (2023). Research progress on classification, sources and functions of dietary polyphenols for prevention and treatment of chronic diseases. *Journal of Future Foods*, 3(4), 289-305.

Laviolle, B., Perche, O., Gueyffier, F., Bégué, É., Bilbault, P., Esperou, H., Varoqueaux, N. (2019). Apport de la génomique dans la médecine de demain, applications cliniques et enjeux. *Therapies*, 74(1), 1-8.

M

Mitra, D., Mitra, D., Bensaad, M. S., Sinha, S., Pant, K., Pant, M., ...& Mohapatra, P. K. D. (2022). Evolution of Bioinformatics and its impact on modern bio-science in the twenty-first century: special attention to pharmacology, plant science and drug discovery. *Computational Toxicology*, 100248.

Mokhtari, M. (2021). Modélisation moléculaire d'un composé de la famille Dithioléthione 4-p-tolyl-1,2-dithiol-3-thione est un dérivé (Thèse de doctorat, Université Kasdi Merbah Ouargla).

Moingeon, P., Chenel, M., Rousseau, C., Voisin, E., Guedj, M. (2023). Patients virtuels, jumeaux numériques et modèles de maladie causale : préparer le terrain pour des essais cliniques in silico. *la découverte de médicaments aujourd'hui*, 7, 103605.

N

Naikwadi, J, J., M, N, Desai., S, Kumar., V, Patil. (2021). Computer aided drug design- an overview. *International Journal of Creative Research Thoughts (IJCRT)*. Vol. 9. 2320-2882.

Nand, M., Maiti, P., Chandra, S., Pande, V. (2016). In Silico identification of novel Egfer Tyrosine Kinase Inhibitors associated with Non-small Cell Lung Cancer Form phytochemical library. *International research journal of pharmacy*. 7 (3). 2230 – 8407.

P

Parul, A. (2018). Bioinformatics-tools and applications. Conférence: 5th international conférence n comuting for sustainable global devvelopment. India. 5044-5047.

Pathak, R. K., Singh, D. B., & Singh, R. (2022). Introduction to basics of bioinformatics. In Bioinformatics (pp. 1-15). Academic Press.

Pellequer, J. L. (1999). Modélisation moléculaire par homologie des protéines : ses applications en Biologie et en Bioinformatique (Doctoral dissertation, Faculté des sciences de Luminy).

Poornima, S. I., & Sumathy, V. J. H. (2022). In Silico Interaction Studies of Various Antioxidant Molecules with MAO-B Protein—A Potent Secondary Drug Target in Parkinson's Disease. *Indian Journal of Science and Technology*, 15(19), 923-926.

Poudyal, B. (2021). Technique d'amarrage moléculaire. *Nat. J.Pharm. Sei* , 1 , 8-14.

Prakash, N., Devangi, P. (2010). *Journal of Antivirals and Antiretrovirals* Volume 2(4): 063-068.

Prasad Gajula, M, N, V., Anuj Kumar, E, A., Siddiq, A, K. (2016). Polumetla.Role of Bioinformatics in Agriculture. *International Journal of Economic Plants*.3(2):077-080.

R

Raghuvanshi, D., Solanki, V., Arora, N., Hashmi, F. 2020). **Computational of Bioinformatics.** *International Journal of Trend in Scientific Research and Development (IJTSRD)*.4(4). 2456-6470.

Rather, MA., Agarwal, D., Bhat, TA., Khan, IA, Zafar, I., Kumar, S., Qadri, T. (2023). Approches bioinformatiques et opportunités d'analyse de mégadonnées pour améliorer la pêche et l'aquaculture. *Journal international des macro molécules biologiques*, 123549.

Raval, K., Ganatra, T. (2022). Bases, types et applications de l'amarrage moléculaire : une revue. *IP International Journal of Comprehensive and Advanced Pharmacology*, 7 (1), 12-16.

S

Saha, D., Hait, M. (2011). Lead Optimization: Research Pathways in Medicinal Chemistry. *Asian Journal of Research in Chemistry*, 4(4), 509-510.

Sapundzhi, FI., Dzimbova, TA (2022). L'importance des bases de données biologiques dans la modélisation de la relation structure-activité. *Interprétation*, 7-11.

Shaker, B., Ahmad, S., Lee, J., Jung, C., & Na, D. (2021). Méthodes et outils in silico pour la découverte de médicaments. *Informatique en biologie et médecine*, 137, 104851.

Senthil, S., T Malathi., M, Surendrakumar. (2023). Novel Approaches in Drug Design. *Journal of developing drugs*. Vol. 12.

Sethi, PC., Behera, PK (2012). Bioinformatique : applications et enjeux. Communication CSI.

Soudani, W., Bouachrine, M., Zaki, H., & Hadjadj-Aould, F. Z. (2022). Docking moléculaire de 2-chloroéthylnitrososulfamides à potentiel anticancéreux cytotoxique alkylant. *RHAZES : Green and Applied Chemistry*, 15, 24-26.

Shultana, S., Maraz, KM., Ahmed, A., Sultana, T., Khan, RA. (2021). Conception, découverte et développement de médicaments et leur innocuité ou leur efficacité sur le corps humain. *CGC Sciences Biologiques et Pharmaceutiques*, 17 (2), 113-122.

Sulochana, M. (2020). Advancement in Applications of Bioinformatics in Different Medical Science Field. *Journal of Proteomics Bioinformatics*. Vol.15(4).15-581.

T

Tanizawa, Y., Fujisawa, T., Kodama, Y., Kosuge, T., Mashima, J., Tanjo, T., Nakamura, Y. (2023). Rapport de mise à jour 2022 de la Banque de données ADN du Japon (DDBJ). *Recherche sur les acides nucléiques*, 51 (D1), D101-D105.

Takagi, T. (2021). Drug Design Techniques and Pharmacology. *International Journal of Drug Development and Research*. Vol.13.

Tastan, B, Ö., Adebisi, EF., Alzohairy, AM., Everett, D., Ghedira, K., Ghouila, A. (2015). Enseignement de la bioinformatique perspectives et défis hors d'Afrique. *Briefings en bioinformatique* , 16 (2), 355-364.

W

Wei, D., Xu, Q., Zhao, T. et Dai, H. (éd.). (2015). Avancée en bio-informatique structurale. Dordrecht: Springer.

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., ...& Wilson, M. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research*, 46(D1), D1074-D1082.

Z

Zhang, M., Zhou, S., Obaid, N. H., Altimari, U. S., Mohammed, M. A., Aldulaim, A. K. O., ... & Rahman, M. L. (2022). Chromenone-based GSK-3 β inhibitors as potential therapeutic targets for cardiovascular diseases: In silico study, molecular dynamics, and ADMET profiles. *Arabian Journal of Chemistry*, 15(12), 104288.

<https://www.aps.dz/sante-science-technologie/116791-plus-de-50-des-deces-en-algerie-dus-aux-maladies-chroniques>

Réalisé par M^{elle} : Kaddour Rima Nibouche Nassiha Zouaoui Dounia	Membre de jury : Présidente : Dr. Medouri A Examinatrice : Dr. Brihoum H Encadrant : Dr. Benhouieb I
--	---

Thème
Approches et méthode bio-informatique pour l'étude des maladies chroniques

Résumé

Résumé :

La bio-informatique est un domaine interdisciplinaire situé au carrefour de l'informatique, des mathématiques et de la biologie. C'est la multi-discipline théorique de l'analyse « in silico » de l'information biologique contenue dans les séquences nucléotidiques et protéique. Elle couvre la collecte, le stockage, la récupération, la manipulation et la modélisation des données pour l'analyse, la visualisation ou la prédiction grâce au développement d'algorithmes et de stratégies de calcul. La présente étude vise à établir les différentes méthodes employées pour l'étude de maladies chroniques telles que le cancer des poumons, polyarthrite rhumatoïde, l'urticaire, l'obésité, la maladie de parkinson, les maladies cardiovasculaires et le diabète. Pour ce faire, une analyse de quelques travaux pertinents publiés dans ce contexte est faite. Sept articles répondant aux critères d'inclusion de l'étude ont été analysés. Les résultats de l'analyse d'articles sont avérés bénéfiques pour comparer, prédire et déterminer les meilleures composées naturelles avec d'excellents scores de liaison d'affinité dans le but de concevoir des molécules thérapeutiques

Mots clés : la bio-informatique, in silico, les maladies chroniques.

Abstract:

Bioinformatics is an interdisciplinary field at the crossroads of computer science, mathematics and biology. It is the theoretical multi-discipline of "in silico" analysis of biological information contained in nucleotide and protein sequences. It covers the collection, storage, retrieval, manipulation and modeling of data for analysis, visualization or prediction through the development of algorithms and computational strategies. The present study aims to establish the different methods employed for the study of chronic diseases such as lung cancer, rheumatoid arthritis, urticaria, obesity, Parkinson's disease, cardiovascular disease and diabetes. To this end, a review of a few relevant works published in this context is carried out. Seven articles meeting the study's inclusion criteria were analyzed. The results of the article analysis proved beneficial for comparing, predicting and determining the best natural compounds with excellent affinity binding scores, with the aim of designing therapeutic molecules.

Keywords: bioinformatics, in silico, chronic diseases.

الساخص

المعلوماتية الحيوية هي مجال متعدد التخصصات يقع في مفترق طرق علوم الكمبيوتر والرياضيات وعلم الأحياء. إنها النظرية متعددة التخصصات لتحليل "في السيليكو" للمعلومات البيولوجية الموجودة في متواليات النيوكليوتيدات والبروتينات. ويغطي جمع وتخزين واسترجاع ومعالجة ونمذجة البيانات للتحليل أو التصور أو التنبؤ من خلال تطوير الخوارزميات والاستراتيجيات الحسابية. تهدف هذه الدراسة إلى تحديد الطرق المختلفة المستخدمة لدراسة الأمراض المزمنة مثل سرطان الرئة والتهاب المفاصل الروماتويدي والشرى والسمنة ومرض باركنسون وأمراض القلب والأوعية الدموية والسكري. للقيام بذلك، تم إجراء تحليل لبعض الأعمال ذات الصلة المنشورة في هذا السياق. تم تحليل سبع مقالات تلبية معايير الاشتمال للدراسة. أثبتت نتائج تحليل المقالة أنها مفيدة في المقارنة والتنبؤ وتحديد أفضل المركبات الطبيعية ذات درجات الربط الممتازة لغرض تصميم الجزيئات العلاجية.

الكلمات المفتاحية: المعلوماتية الحيوية، في السيليكو، الأمراض المزمنة