



République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Med Khider Biskra

Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie

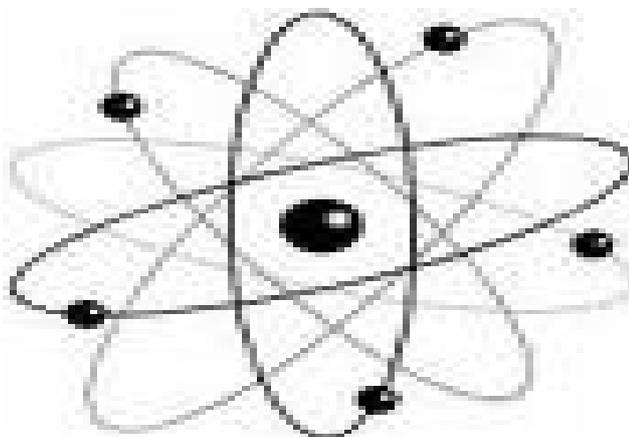


Département des Sciences de la Matière

Domaine : Sciences de la Matière

Filière : Chimie

Spécialité : Chimie pharmaceutique



Mémoire de fin d'étude en Master

Intitulé :

***Etude par modélisation moléculaire l'activité
cytotoxiques d'une série de dérivés de benzimidazole
(QSAR et Docking moléculaire)***

Présenté par :

Abiza Fiala

Devant le Jury :

DAOUD Ismail

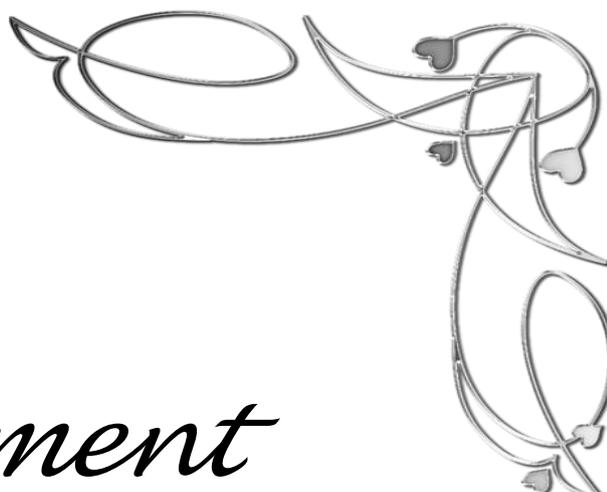
MELKEMI Nadjib

HARKATI Dalal

Président

Rapporteur

Examinatrice



Remerciement

Je tiens à remercier DIEU le tout puissant pour toute la volonté et le courage qu'il je ouvre a offert pour réaliser ce travail.

Ce mémoire a été dirigé par **MELKEMI Nadjib** maître de conférences A au Département des Sciences de la Matière, Faculté des Sciences exactes et Sciences de la nature et de la vie, Université **Mohamed Khider de Biskra**. Ce travail a été réalisé au sein de Laboratoire de Chimie Moléculaire et Environnement (LMCE) domicilié à l'université de **Mohamed Khider de Biskra**

Je tiens particulièrement à remercier Monsieur **MELKEMI Nadjib** pour son encadrement scientifique. Je tiens à lui témoigner sincèrement toute ma gratitude pour son esprit critique et ses orientations qui ont été d'une aide précieuse pour mon travail, ainsi que pour la patience et la gentillesse dont il a fait preuve à mon égard. Je me dois également de le remercier pour ses efforts qui m'ont permis d'aboutir dans mon travail de mémoire.

Je remercie également les membres de jury Monsieur **DAOUD Ismail** et **HARKATI Dalal** pour l'honneur qu'ils m'ont fait en acceptant de juger notre travail.

Je remercie vivement toute l'équipe de chimie informatique et pharmaceutique au laboratoire de Chimie Moléculaire et Environnement (LMCE) de Biskra.

Merci



DEDICACES

Je dédie cette thèse

A mes chers parents

*pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et
leurs prières tout au long de mes études*

A ma chère sœur

A mes chers frères

A mes grands- mères

A la mémoire de mon grand père ' paix à son âme'

À mes copines Torkia, zineb, wahiba, jiji, Les halogènes, Zina, Dossa,

A tous les membres de ma famille, petits et grands

pour leur soutien tout au long de mon parcours universitaire,

Merci d'être toujours là pour moi.

Abiza Fiala

Sommaire

Liste de tableau.....	i
Liste des figures	ii
Liste des principales abréviations.....	iii
Introduction générale.....	1
Références bibliographique.....	3

Chapitre I : Comporte une recherche bibliographique sur les maladies cancéreuses et l'activité cytotoxique

I.1.1 Introduction.....	4
I.1.2 Historique	4
I.1.3 Tumeur ou cancer	4
I.1.3.1 Une tumeur bénigne	4
I.1.3.2 Une tumeur maligne	5
I.1.3.3 La relation entre le tumeur et cancer.....	5
I.1.4 Caractéristiques d'une cellule cancéreuse.....	5
I.1.5 Etapes de la cancérogenèse.....	6
I.1.5.1 Initiation.....	6
I.1.5.2 Promotion.....	6
I.1.5.3 Progression.....	6
I.1.5.4 Invasion.....	7
I.1.5.5 Angiogenèse.....	7
I.1.5.6 Métastases ganglionnaires.....	7
I.1.5.7 Métastases sanguines.....	8
I.1.5.8 La différence entre la cellule normale et la cellule cancéreuse	8
I.1.6 Types de cancer	8
I.1.7 Cancers les plus fréquents.....	9
I.1.8 Les cancers les plus fréquents chez l'homme et la femme.....	9
I.1.9 Symptômes du cancer.....	9
I.1.9.2 Déshydratation.....	10
I.1.9.3 Fatigue.....	10

I.1.9.4 Nausées et vomissements.....	10
I.1.9.5 Troubles de la bouche.....	10
I.1.9.6 Constipation	10
I.1.9.7 Difficulté à respirer.....	10
I.1.9.8 Troubles de la peau.....	11
I.1.10 Causes de cancer	11
I.1.11 Les différents traitements du cancer.....	11
I.1.11.1 La chirurgie oncologique	11
I.1.11.2 La radiothérapie, ou irradiation.....	12
I.1.11.3 L'hormonothérapie	12
I.1.11.4 La chimiothérapie	12
I.1.11.5 Immunothérapie.....	12
I.2.1 Introduction	13
I.2.2 Les interactions biomoléculaires (récepteur-ligand).....	13
I.2.2.1 Une espèce reconnue par le récepteur.....	14
I.2.2.2 Un récepteur	14
I.2.3 Epidermal Growth Factor Receptor (EGFR).....	14
I.2.4 Michigan Cancer Foundation-7 (MCF7).....	16
I.2.5 Carcinome hépatocellulaire (HepG2).....	16
I.2.6 Cytologie.....	17
I.2.7 Cytostatique	17
I.2.8 Cytotoxique.....	17
I.2.9 Mécanisme d'action des agents cytotoxiques.....	17
I.2.9.1 Classification suivant leurs propriétés pharmacologiques.....	17
I.2.9.1.1 ADN.....	19
I.2.9.1.1.1 Les agents alkylants	19
I.2.9.1.1.2 Les agents intercalants.....	19
I.2.9.1.1.3 Les inhibiteurs de topoïsomères.....	20
I.2.9.1.1.4 Les antimétabolites	20
I.2.9.1.2 Fuseau mitotique.....	20
I.2.9.1.2.1 Les poisons du fuseau.....	20
I.2.9.1.2.2 Inhibition de la polymérisation	21
I.2.9.1.2.3 Inhibition de la dépolymérisation.....	21

I.2.9.1.3 Protéines.....	21
I.2.9.1.3.1 Les tyrosines kinases.....	21
I.2.9.1.3.2 Les inhibitrices protéines kinase	21
I.2.9.1.4 Anticorps monoclonaux	22
I.2.9.1.4.1 Les anticorps.....	22
I.2.9.1.4.2 Anticorps monoclonaux	22
I.2.10 Le test de cytotoxicité in vitro	23
I.3.1 Benzimidazole.....	24
I.3.1.1 Définition	24
I.3.1.2 L'activité biologique générale	24
I.3.1.3 L'activité biologique sur le cancer.....	24
Références bibliographique.....	26

Chapitre II : Aspect théorique sur les méthodes de calculs utilisées

II.1.1 Introduction.....	31
II.1.2 Modalisation moléculaire	32
II.1.3 Méthodes de la modélisation moléculaire.....	32
II.1.3.1 La mécanique quantique (MQ)	32
II.1.3.1.1 Equation de Schrödinger	33
II.1.3.1.2 L'approximation de Born-Oppenheimer	34
II.1.3.1.3 La théorie de la fonctionnelle de la densité (DFT)	35
II.1.3.1.4 Les bases d'orbitales.....	35
II.1.3.1.5 Méthodes ab-initio.....	36
II.1.3.1.6 Les méthodes semi-empiriques.....	36
II.1.3.2 La mécanique moléculaire (MM)	37
II.1.3.2.1 Fonction d'énergie potentielle	37
II.1.3.2.2 Les différentes énergies	38
II.1.3.2.2.1 Energies des interactions liées	38
II.1.3.2.2.2 Energies des interactions intermoléculaires - "non - liees"	40
II.1.3.2.3 Champ de force	41
II.1.3.2.3.1 Différents champs de force en mécanique moléculaire.....	42

II.1.3.3 Dynamique Moléculaire	42
II.1.3.3.1 Principe de Dynamique Moléculaire.....	43
II.1.3.3.2 Mise en oeuvre d'un calcul de DM	44
II.1.3.3.3 Méthodes empirique	45
II.1.4 Domaine d'application de la modélisation moléculaire.....	45
II.2.1 Introduction	47
II.2.2 Définition.....	47
II.2.3.1 Les équations QSAR prennent la forme générale suivante.....	49
II.2.4 Les applications de l'étude QSAR.....	49
II.2.5 Les modèles statistiques.....	49
II.2.5.1 La régression multilinéaire RML.....	50
II.2.5.2 Formulation matricielle de la régression linéaire	50
II.2.5.3 La signification du modèle.....	51
II.2.5.4 Le choix du meilleur modèle	51
II.2.6 Validation du modèle	52
II.2.6.1 Validation interne	52
II.2.6.2 Validation externe	53
Références bibliographique.....	54

Chapitre III : Investigation des relations structure/activité cytotoxique des dérivés de
bezimidazole

III.1.1 Introduction	59
III.1.2 Les propriétés QSAR caractérisant la structure moléculaire	60
III.1.2.1 Les paramètres amphiphiliques	60
III.1.2.2 Les paramètres quantique/électronique.....	62
III.1.2.3 Les propriétés « drug-like ».....	63
III.1.2.4 Surface moléculaire polaire (PSA).....	64
III.1.2.5 Liaison rotative (Ntrob).....	64
III.1.2.6 L'efficacité du ligand (LE).....	64
III.1.2.7 L'efficacité lipophile de ligand (LLE).....	64
III.1.3 Programmes de modélisation utilisés	65
III.2.1 Etude de la relation structure-propriétés des dérivés de Benzimidazole...	66
III.2.1.1 Etude des propriétés physicochimiques des dérivés de	69

Benzimidazole.....	
III.2.1.2 Étude des propriétés « drug-like ».....	72
III.2.1.2.1 Règle de lipinski	72
III.2.1.2.2 Règle de viber.....	73
III.2.1.3 L'efficacité du ligand (LE)	74
III.2.1.4 Efficacité lipophile de ligand (LLE).....	75
III.2.2 Etude quantitative des propriétés QSAR d'une série des dérivés de Benzimidazole	76
III.2.2.1 Développement des modèles QSAR	79
III.2.2.2 Interprétation.....	80
III.2.2.3 Validation des modèles QSAR.....	83
III.2.3 Conclusion.....	87
Références bibliographique.....	88
<u>Chapitre IV</u> : Etude par docking moléculaire des interactions benzimidazoles/enzymes (1M17- 3EQM – 3GCM)	
IV.1.1 Introduction	91
IV.1.2 Docking moléculaire	91
IV.1.3 Les programmes d'assemblage moléculaire	93
IV.1.3.1 Recherche globale des assemblages potentiels	93
IV.1.3.2 Reclassement des complexes trouvés	94
IV.1.4 Protocole Générale de Docking.....	94
IV.2.1 Etude par docking moléculaire des propriétés cytotoxiques des benzimidazoles	96
IV.2.1.1 Préparation du récepteur et ligands.....	96
IV.2.1.1.1 préparation de l'enzyme.....	96
IV.2.1.1.2 Préparation du ligand.....	98
IV.2.1.2 Résultats de docking moléculaire.....	98
IV.2.1.2.1 Etapes de docking moléculaire.....	98
IV.2.1.2.2 Détection des cavités	100
IV.2.1.2.3 Énergies d'interactions	101
IV.2.2 Conclusion	108
Références bibliographique.....	109
Conclusion générale	110

Liste des tableaux

Tableau I.1 : Facteurs de risque de cancer.....	11
Tableau III.1 : Structures chimiques des dérivés de Benzimidazole.....	67
Tableau III.2 : Paramètres physicochimiques des dérivés de Benzimidazole....	69
Tableau III.3 : Les paramètres de Lipinski des dérivés de Benzimidazole.....	73
Tableau III.4 : Les paramètres de Veber des dérivés de Benzimidazole.....	74
Tableau III.5 : Efficacité de ligand (LE) des dérivés de Benzimidazole.....	75
Tableau III.6 : Efficacité lipophile de ligand (LLE) des dérivés de Benzimidazole.....	76
Tableau III.7 : Descripteurs utilisés dans le développement des modèles QSAR.....	78
Tableau III.8 : Les signes de différents descripteurs dans les modèles QSAR obtenus.....	81
Tableau III.9 : Matrice des corrélations de 1 ^{ère} modèle.....	82
Tableau III.10 : Matrice des corrélations de 2 ^{ème} modèle.....	82
Tableau III.11 : Matrice des corrélations de 3 ^{ème} modèle.....	82
Tableau III.12 : Paramètres de validation croisée.....	83
Tableau III.13 : Les valeurs expérimentales, prédites et résiduelles de (P) de série bioactive de benzimidazole.....	84

Liste des figures

Figure I.1 : Les six caractéristiques d'un cancer.....	5
Figure I.2 : Différentes étapes du cancer.....	6
Figure I.3 : La différence entre la cellule normale et la cellule cancéreuse.....	8
Figure I.4 : Cancers les plus fréquents.....	9
Figure I.5 : Différents partenaires impliqués dans les interactions biomoléculaires	14
Figure I.6 : Epidermal growth factor receptor (EGFR).....	15
Figure I.7 : Structure de l'EGFR.....	15
Figure I.8 : Cellules MCF-7.....	16
Figure I.9 : Cellules HepG2.....	17
Figure I.10 : Cible des anticancéreux.....	18
Figure I.11 : Classification des agents cytotoxiques.....	18
Figure I.12 : Réaction chimique entre un nucléophile et un électrophile formation d'une liaison covalente.....	19
Figure I.13 : Intercalation directe avec l'ADN.....	19
Figure I.14 : Activité de la topoisomérase I et de la topoisomérase II.....	20
Figure I.15 : Effet de poisons du fuseau sur la mitose.....	21
Figure I.16 : Mode d'action des inhibiteurs de tyrosine kinase.....	22
Figure I.17 : Mode d'action des Anticorps monoclonaux.....	23
Figure I.18 : Structure de Benzimidazole.....	24
Figure II.1 : Energie de déformation des liaisons.....	38
Figure II.2 : Energie de rotation.....	39
Figure II.3 : Energie de déformation des liaisons.....	39
Figure II.4 : Energie de déformation des liaisons.....	40
Figure II.5 : Modèle d'étude quantitative des relations structure / activité (QSAR)	
Figure III.1 : Structure de Benzimidazole.....	48
Figure III.2 : Les sites donneurs et accepteurs dans les composés 10 et 15.....	70
Figure III.3 : L'effet de lipophile dans la membrane cellulaire.....	71
Figure III.4 : Les courbes de la régression linéaire des valeurs prédites en fonction des valeurs expérimentales de pHepG2, pEGFR et pMCF7.....	85
Figure III.5 : Représentations graphique des valeurs résiduelles en fonction des valeurs expérimentales pour les modèles 1, 2, 3.....	86

Figure IV.1 : Etapes typique d'un docking moléculaire.....	92
Figure IV.2 : Etapes typique d'un docking moléculaire.....	93
Figure IV.3 : Schéma du protocole général de docking moléculaire.....	95
Figure IV.4 : Enzyme tyrosine kinase.....	97
Figure IV.5 : Enzyme proprotéine convertase subtilisine/kexine de type 9 "PCSK9".....	97
Figure IV.6 : Enzyme aromatase cytochrome p450.....	98
Figure IV.7 : Etapes de Docking Moléculaire.....	99
Figure IV.8: Cavités détectés dans les enzymes étudiées.....	100
Figure IV.9 : Interaction hydrogène et stérique entre les ligands et les résidus du site actif de la cavité1.....	105

Liste des principales abréviations

EGF	Epidermal Growth Factor.
EGFR	Epidermal Growth Factor Receptor.
MCF7	Michigan Cancer Foundation-7
HepG2	Carcinome hépatocellulaire
IC50	Concentration inhibitrice 50.
nM	Nanomètre.
hCA	Heterocyclic amine.
μM	Micromètre.
Ψ	Fonction d'onde de la molécule.
H	Hamiltonien totale.
E_{TOT}	Energie totale.
DFT	Théorie de la fonctionnelle de la densité.
B3LYP	Becke 3-paramètres Lee-Yang-Parr.
CNDO/2	Complete Neglect of Differential Overlap.
INDO	Intermediate Neglect of Differential Overlap.
NDDO	Neglec of Diatomic Differential Overlap.
MNDO	Modified Neglected of Differential Overlap.
AM1	Austin Model 1.
PM3	Parametric Method 3.
QSPR	Quantitative Structure-Property Relationships.
QSAR	Quantitative Structure-Activity Relationships.
MLR	Régressions multi-linéaires.
PLS	Partiel least squares regression.
CSPA	Charged partial surface area.
LOO, LMO	Leave-one out, Leave-many out.
PRESS	prediction sum of squares.
RMN	Résonance magnétique nucléaire.
Å°	Ångström (unité de longueur).
UMA	Unité de masse atomique.
d	Densité

Liste des principales abréviations

Kcal	Kilocalorie.
HBD	Liaisons hydrogène donneurs.
HBA	Liaisons hydrogène accepteurs.
PSA	Surface polaire atomique.
%ABS	pourcentage d'absorption.
LE	Efficacité de ligand.
LiPE	Efficacité lipophile.
NH	Nombre des atomes lourds.
OM	Les orbitals Moléculaires.
eV	Electron volt.
ADME	Absorption, Distribution, Métabolisme et Excrétion.
PDB	Protein Data Bank.
MVD	Molegro Vertual Docker.

Introduction générale

La chimie pharmaceutique est une discipline qui s'intéresse à la conception, à la préparation et à l'interprétation du mode d'action des médicaments. C'est une science qui tente, entre autres, d'établir des relations entre la structure chimique des drogues et leur activité biologique [1].

Le cancer correspond à la multiplication anarchique de cellules anormales qui échappent au mécanisme de contrôle cellulaire. Ces cellules prolifèrent pour former une masse tissulaire appelée tumeur qui envahit les tissus normaux avoisinant; d'où la malignité [2].

Selon l'organisation mondiale de la santé, le cancer est aujourd'hui la deuxième cause de mortalité après les maladies cardiovasculaires dans les pays développés ce qui représente plus de dix millions de décès dans le monde.

Dans les pays développés, la mortalité globale due au cancer est plus de deux fois supérieure à celle des pays en voie de développement. Cela s'explique par la forte consommation de tabac, tout comme l'exposition aux cancérogènes professionnels, au mode de vie et à l'alimentation.

Dans les pays en voie de développement, l'OMS estime que vingt cinq millions de personnes vivent avec le cancer et prédit une progression d'environ 50% des nouveaux cas dans ces pays d'ici à 2020 [3].

La relation quantitative structure-activité (QSAR) est l'un des outils les plus pratiques en physico-chimie computationnelle. L'idée fondamentale de QSAR consiste en la possibilité d'une relation entre un ensemble de descripteurs, et une réponse moléculaire. QSAR peut être considérée comme une règle calculée par ordinateur qui décrit quantitativement l'activité biologique en termes de descripteurs chimiques, il a été fréquemment utilisé pour prédire les activités biologiques des nouveaux composés [4].

La prédiction de l'activité biologique est basée sur des calculs mathématiques et statistiques [5]. La recherche de 'bonne' structure est cruciale dans la conception de nouveau médicament. Ou encore, selon l'association de 2 protéines, le signal déclenché lors de leur association peut être différent. D'un point de vue informatique, la prédiction de structure se traduit par un problème d'optimisation. A chaque position relative possible est

associe une énergie (également appelé score, fitness, etc.). La structure la plus probable est alors celle qui minimise cette énergie.

Le premier objectif de notre travail est l'étude quantitative des relations structure-activité des propriétés cytotoxique d'une série de benzimidazole. Cette étude permet de développer des modèles QSAR utilisés dans la prédiction de l'activité cytotoxique des dérivés de benzimidazole.

Le second objectif est la détermination des meilleurs ligands qui forment des complexes stables avec les enzymes (tyrosine kinase, Aromatase cytochrome p450 et proprotéine convertase subtilisine/kexine de type 9), ainsi que, les types d'interactions entre les ligands et les enzymes étudiées.

Pour réaliser ce travail d'une manière convenable, nous avons divisé notre travail en quatre chapitres. Le premier chapitre comporte une recherche bibliographique sur les maladies de cancer, l'activité cytotoxique et les paramètres biologiques. Le deuxième chapitre est consacré à la description des méthodes de calcul utilisé. Le troisième chapitre est étude quantitative des relations structure/activité des propriétés cytotoxique des dérivés de benzimidazole. Le dernier chapitre comporte une étude par docking moléculaire des propriétés cytotoxiques de benzimidazoles.

Références bibliographique

- [1] : Lavrado, João, et al. "Antitrypanosomal and cysteine protease inhibitory activities of alkyldiamine cryptolepine derivatives." *Bioorganic & medicinal chemistry letters* 22.19 (2012): 6256-6260.
- [2] : GNOULA, Charlemagne. ETUDES DES PROPRIETES CYTOTOXIQUES ET ANTIRADICALAIRES D'EXTRAITS DE FEUILLES ET DE GALLES. Diss. UNIVERSITE D'OUAGADOUGOU, 2009.
- [3] : Organisation mondiale de la santé (OMS ,2015)
- [4] : Wermuth, Camille G. "Designing prodrugs and bioprecursors." *The Practice of Medicinal Chemistry (Third Edition)*. 2008. 721-746.
- [5] : Goudarzi, Nasser, Mohammad Goodarzi, and Tao Chen. "QSAR prediction of HIV inhibition activity of styrylquinoline derivatives by genetic algorithm coupled with multiple linear regressions." *Medicinal Chemistry Research* 21.4 (2012): 437-443.

I.1.1 Introduction

L'étude du cancer se nomme oncologie. Des Cancers ont été connus à l'humanité depuis des époques antiques. Le Cancer commence quand les cellules dans une partie du fuselage commencent à se développer hors du contrôle. Plusieurs différentes parties du fuselage peuvent être affectées par le cancer. [1]

I.1.2 Historique

Une Partie de la preuve du cancer la plus précoce est trouvée parmi des tumeurs osseuses fossilisées dans les momies humaines en Egypte antique, et des références à la même chose a été trouvées en manuscrits antiques.

Bien Que le cancer de mot n'ait pas été utilisé, la description la plus ancienne de la maladie est l'Egypte et remonte environ à 3000 BC. Elle est appelée le Papyrus d'Edwin Smith et est une copie d'une partie d'un manuel Égyptien antique sur la chirurgie de traumatisme. Elle décrit 8 cas des tumeurs ou ulcères du sein qui ont été traités par cautérisation avec un outil appelé l'exercice contre l'incendie.

- ✓ Hippocrate (460-370 BC) le médecin Grec. Il est considéré le « Père du Médicament.» Hippocrate avait l'habitude :
 - les carcinos pour décrire la tuméfaction bénigne
 - le carcinome pour décrire le cancer entraînant la mort
 - le squirrhus pour décrire le cancer curable
- ✓ Celsus (28-50 BC) un médecin Romain, a traduit la condition Grecque en cancer, le mot Latin pour le crabe.
- ✓ Galen (130-200 AD), un autre médecin Romain, qui avait l'habitude les oncos de condition (Grecs pour le gonflement) pour décrire des tumeurs. Oncos est le mot de racine pour l'oncologie ou l'étude des cancers. [1]

I.1.3 Tumeur ou cancer

Est un terme générique correspondant au développement d'un tissu nouvellement formé au sein d'un tissu normal. Elle est provoquée par le dysfonctionnement du développement cellulaire. ona deux types de tumeurs : [2]

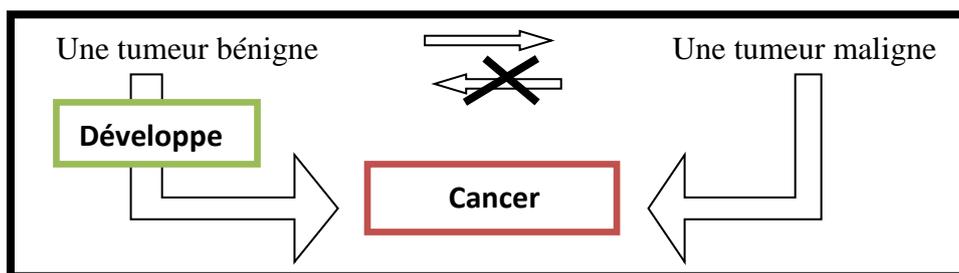
I.1.3.1 Une tumeur bénigne

Est une tumeur sans gravité, c'est-à-dire ne pouvant donner lieu à des métastases et n'étant pas mortelle. Ce n'est pas un cancer. [3]

I.1.3.2 Une tumeur maligne

Le cancer est le nom donné à tout type de tumeur maligne qui a une évolution rapide et se fait vers l'atteinte du reste de l'organisme. Le cancer est du à une anomalie de fonctionnement des cellules qui entraîne un développement anarchique de celles-ci. Il se développe d'abord dans un organe et forme une ou des tumeurs ; mais il arrive que les cellules cancéreuses atteignent d'autres parties du corps. Ce sont les métastases. On parle alors de cancer primitif et de cancer secondaire. [4]

I.1.3.3 La relation entre le tumeur et cancer



I.1.4 Caractéristiques d'une cellule cancéreuse

Pour devenir cancéreuse, une cellule doit accumuler certaines altérations précises de sa physiologie qui se résume en six caractéristiques. [5]

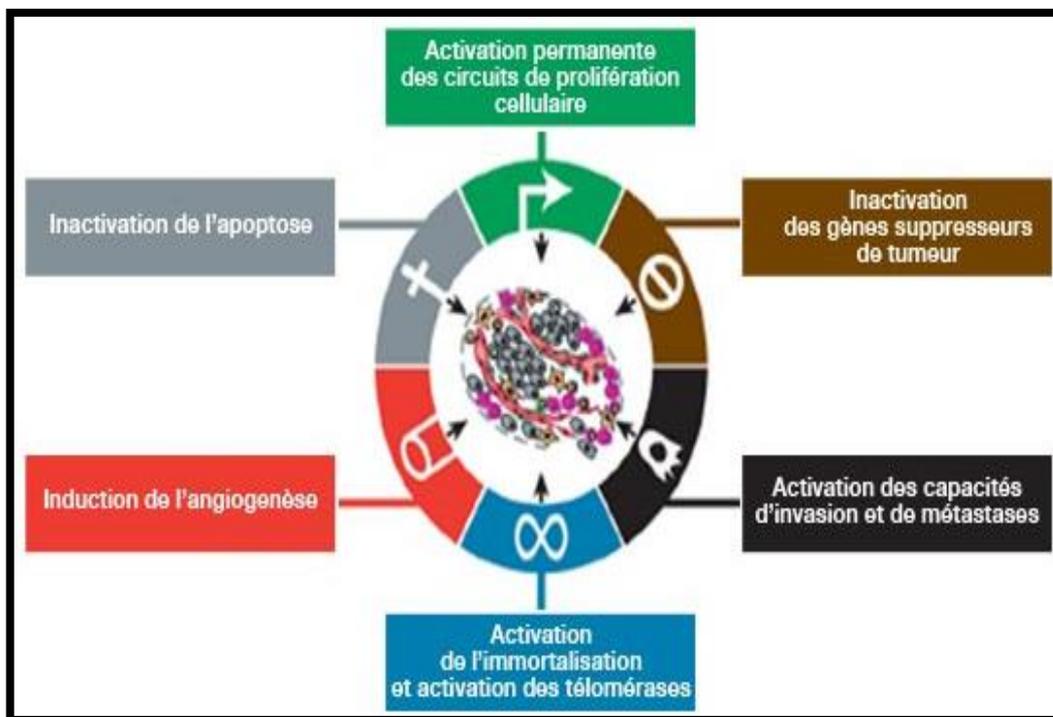


Figure I.1 : Les six caractéristiques d'un cancer

I.1.5 Etapes de la cancérogenèse

La cellule devrait passer par un processus de cancérogenèse se résumant en plusieurs étapes :

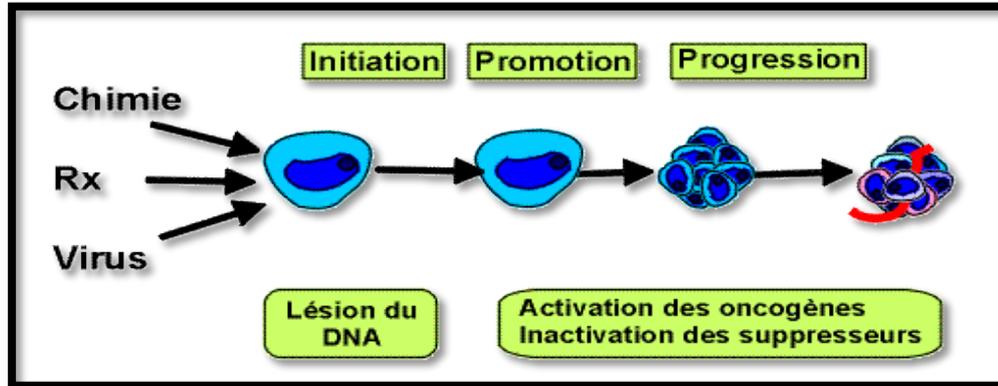


Figure I.2 : Différentes étapes du cancer

I.1.5.1 Initiation

Elle correspond à une lésion rapide et irréversible de l'ADN après une exposition à un agent carcinogène tel que les produits chimiques, les virus et les radiations. L'initiation consiste donc, en une dérégulation au niveau de certains gènes aboutissant à une transformation cellulaire. Cette étape cible les oncogènes, les gènes suppresseurs de tumeurs, les gènes de réparation de l'ADN et les gènes de l'apoptose. [6]

I.1.5.2 Promotion

Cette étape consiste en une prolifération plus ou moins contrôlée des cellules initiées en processus d'acquies les caractéristiques lui permettant de générer un cancer. Elle résulte d'une cascade d'interactions entre les cytokines et leurs récepteurs, ce qui amène à la perte de l'homéostasie tissulaire et l'émergence de clones cellulaires transformées. [7]

I.1.5.3 Progression

Durant l'étape de la progression, la prolifération des cellules tumorales est incontrôlable et le cancer devient cliniquement détectable (plus de 10^9 cellules). Cependant, du fait du nombre très important de mitoses, les cellules finissent par acquies certaines propriétés d'autonomie de croissance, une adaptabilité métabolique et une pharmacorésistance. Durant la progression, une perte de fonction caractérise les cellules cancéreuses qui deviennent alors incapables d'effectuer leurs fonctions normales au sein d'un tissu. De même, l'instabilité génétique conduit les cellules cancéreuses à l'acquisition de nouvelles propriétés d'indépendance et de nouvelles caractéristiques fonctionnelles. [8]

I.1.5.4 Invasion

Contrairement aux cellules normales, les cellules cancéreuses ne respectent plus les relations structurelles et tissulaires et s'infiltrent puis envahissent les tissus voisins. Le mécanisme le plus important dans l'invasion tumorale est la perte d'adhésion des cellules entre elles à cause de la perte ou la modification des glycoprotéines de surface membranaire qui assurent à l'état normal la reconnaissance des cellules d'un même tissu. De même, la constitution d'épithélium qui est bien organisé, empêchent la congruence des cellules entre elles et préparent les cellules cancéreuses à traverser les membranes basales (structures qui séparent les couches cellulaires de surface et le tissu conjonctif, tissu de soutien) en essayant de les dissoudre soit par elles-mêmes, soit par la stimulation des cellules mésenchymateuses avoisinantes, cela afin d'envahir les tissus voisins [9]. La rupture et le franchissement de la membrane basale représentent un critère formel pour distinguer les cancers invasifs des cancers in situ. Durant ce stade, les cellules vont également acquérir une certaine mobilité pour s'approcher des vaisseaux sanguins. [6]

I.1.5.5 Angiogenèse

Une fois les cellules cancéreuses franchissent les membranes basales, elles passent aux vaisseaux du tissu conjonctif (artères, veines lymphatiques) et aux organes adjacents. La sécrétion des facteurs impliqués dans l'angiogenèse et des facteurs protéolytiques durant cette étape permet aux cellules cancéreuses de constituer leur propre réseau de capillaires nourriciers indispensables à leur prolifération. La stimulation des cellules endothéliales des capillaires voisins par les cellules cancéreuses permet de former des boucles vasculaires en leur direction, pour l'apport en nutriments et en oxygène [10]. Les cellules cancéreuses vont également stimuler les cellules mésenchymateuses normales, aboutissant à ce qu'on appelle le stroma réaction qui permet le développement d'un tissu de soutien non cancéreux à la tumeur et qui est nécessaire à sa survie. [11]

I.1.5.6 Métastases ganglionnaires

La gravité d'un cancer réside dans sa capacité de développer des métastases à distance. Les cellules cancéreuses utilisent les mêmes mécanismes pour envahir les vaisseaux lymphatiques et sanguins. Ainsi, le drainage lymphatique des tissus normaux explique leur envahissement rapide par les cellules cancéreuses qui sont emportées par le courant lymphatique vers les sinus corticaux du premier relais ganglionnaire. L'arrivée des cellules cancéreuses dans les ganglions est souvent accompagnée d'un état réactionnel, sous forme de lymphadénite chronique non spécifique. [12]

I.1.5.7 Métastases sanguines

Dans cette étape les cellules cancéreuses gagnent de multiples organes via le cœur après leur passage dans la circulation veineuse ou lymphatique.

En plus de la dissémination hématogène et lymphatique, il existe d'autres modes de dissémination qui sont la dissémination par voie cavitaire, connue aussi sous le nom de voie séreuse, la dissémination anatomique le long des conduits naturels par voie canalaire et enfin, la dissémination par le liquide céphalorachidien a été reconnue responsable des métastases leptoméninges ou des métastases dans les ventricules. [13]

I.1.5.8 La déférence entre la cellule normale et la cellule cancéreuse

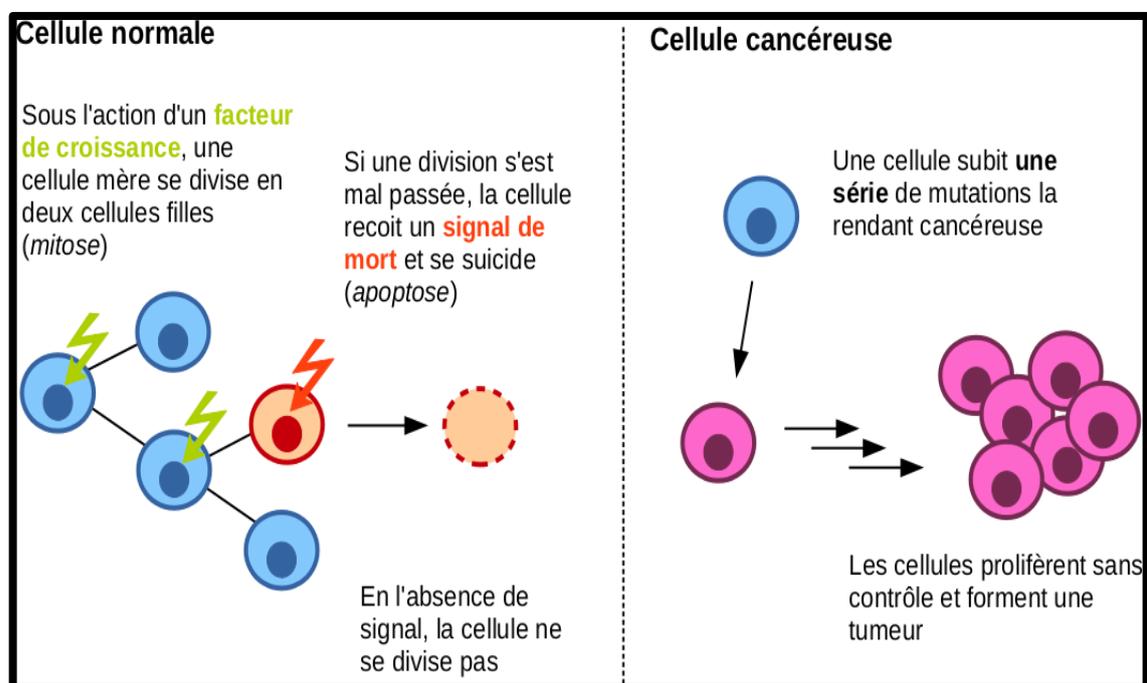


Figure I.3 : La déférence entre la cellule normale et la cellule cancéreuse

I.1.6 Types de cancer

Il existe plus de 100 différents types de cancer. La maladie peut toucher presque n'importe quel organe du corps. On distingue 6 principales catégories de cancers : [14]

- ✓ carcinomes sont les tumeurs qui prennent naissance dans le revêtement extérieur ou intérieur des organes internes (appelé tissu épithélial) et sur la surface extérieure du corps.
- ✓ leucémies sont les cancers des éléments constitutants du sang.
- ✓ lymphomes sont les tumeurs qui se forment dans le système lymphatique;
- ✓ mélanomes sont les cancers qui commencent dans les cellules qui donnent la pigmentation à la peau.

- ✓ sarcomes sont les tumeurs qui prennent naissance dans le tissu conjonctif, comme les muscles, les os et le cartilage.
- ✓ types mixtes de cancers possédant les caractéristiques de plus d'une catégorie de cancer.

I.1.7 Cancers les plus fréquents

Un homme sur deux et une femme sur trois sont concernés par le cancer. La fréquence du cancer a été multipliée par 4 en 10 ans chez les femmes de 35 à 45 ans. Les principaux cancers sont : [15]

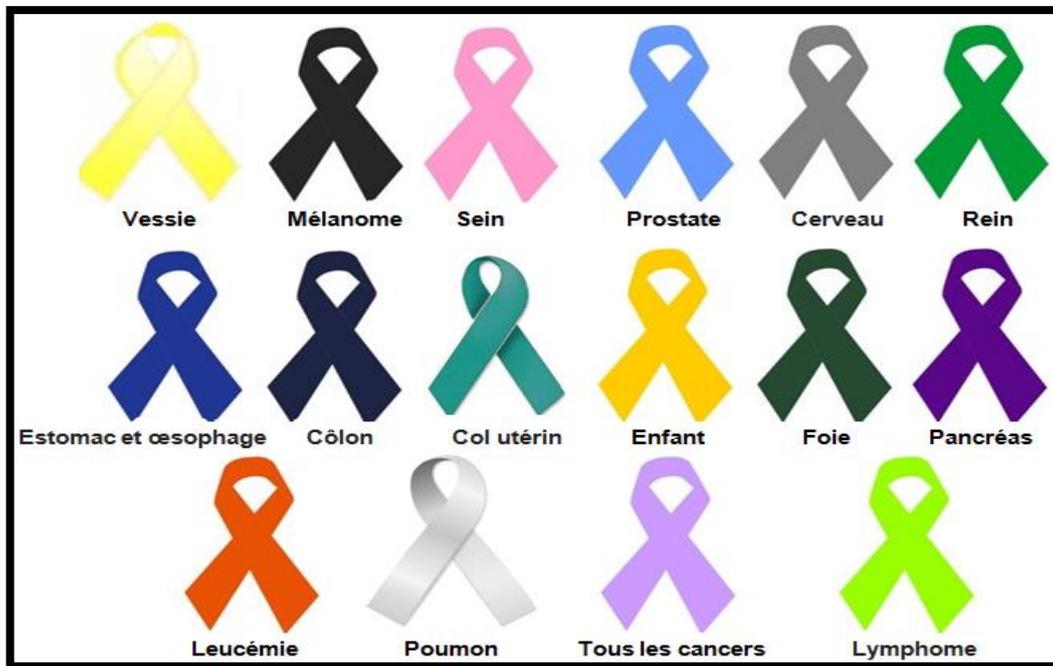


Figure I.4 : Cancers les plus fréquents

I.1.8 Les cancers les plus fréquents chez l'homme et la femme

✓ Homme

Les cancers les plus fréquents chez l'homme sont le cancer de la prostate, le cancer du poumon et le cancer colorectal.

✓ Femme

Les cancers les plus fréquents chez la femme sont les cancers du sein, le cancer du côlon, le mélanome, le cancer des ovaires, le cancer de l'utérus et le cancer du poumon. [15]

I.1.9 Symptômes du cancer

Il arrive souvent que les personnes qui ont un cancer avancé éprouvent des symptômes inquiétants. Les symptômes physiques dépendent habituellement du type spécifique de cancer dont une personne est atteinte et des régions où il s'est propagé. [16]

I.1.9.1 Perte d'appétit et perte de poids

Les personnes atteintes d'un cancer avancé deviennent généralement très minces. C'est peut-être parce qu'elles trouvent difficile de manger ou de boire ou qu'elles ont perdu l'appétit, même pour les aliments qu'elles avaient l'habitude d'aimer. Cela mène à la malnutrition et à la perte de poids.

I.1.9.2 Déshydratation

Il arrive qu'une personne atteinte d'un cancer avancé soit incapable de boire suffisamment de liquide parce qu'elle ne peut pas avaler ou qu'elle n'a pas envie de boire. Cela risque de mener à la déshydratation.

I.1.9.3 Fatigue

La fatigue est très fréquente chez les personnes atteintes d'un cancer avancé. Elles se fatiguent souvent très vite et ont moins d'énergie. Elles peuvent être somnolentes et moins alertes et actives.

I.1.9.4 Nausées et vomissements

Beaucoup de personnes atteintes d'un cancer avancé peuvent avoir des nausées et des vomissements chroniques. Les nausées et vomissements peuvent être causés par les médicaments antidouleur, la constipation, le cancer qui évolue, un blocage dans l'intestin ou un ralentissement de la digestion.

I.1.9.5 Troubles de la bouche

Il est possible qu'une personne atteinte d'un cancer avancé ait la bouche sèche ou la bouche et la gorge endolories. Ces problèmes peuvent se manifester si elle respire par la bouche, si elle boit moins et si elle reçoit certains types de traitement. Si le système immunitaire est affaibli à cause du traitement ou de la maladie qui est avancée, une personne risque d'avoir des infections dans la bouche.

I.1.9.6 Constipation

La constipation est courante chez les personnes atteintes d'un cancer avancé. La prise de médicaments antidouleur comme les opioïdes, moins manger et moins boire et être moins actif peuvent causer la constipation.

I.1.9.7 Difficulté à respirer

Certaines personnes atteintes d'un cancer avancé peuvent avoir de la difficulté à respirer ou être essouffées. L'essoufflement peut être causé par :

- ✓ une accumulation de liquide dans les poumons (épanchement pleural)
- ✓ l'anémie

- ✓ une accumulation de liquide autour du cœur (épanchement péricardique)
- ✓ une infection pulmonaire (pneumonie)

I.1.9.8 Troubles de la peau

Il est possible que les personnes atteintes d'un cancer avancé souffrent de troubles de la peau comme la sécheresse, les démangeaisons ou une rougeur au-dessus des articulations. La peau peut devenir mince et très fragile. Ces problèmes peuvent être causés par une moins grande consommation de liquides et d'aliments et une perte de poids.

I.1.10 Causes de cancer

Certains facteurs de risque, dits « internes », sont liés à notre âge ou à notre histoire familiale. Certaines personnes présentent plus de risques de développer un cancer que d'autres parce qu'à leur naissance, elles portent certaines mutations dans un ou plusieurs de leurs gènes.

D'autres facteurs de risque, dits « externes », sont liés à nos comportements ou à notre environnement. [17]

Tableau I.1 : facteurs de risque de cancer

FACTEURS LIÉS À NOS MODES DE VIE	FACTEURS LIÉS À NOTRE ENVIRONNEMENT
<ul style="list-style-type: none"> ○ Le tabac ○ L'alcool ○ Le surpoids et l'obésité, liés à une alimentation déséquilibrée et/ou un manque d'exercice ○ Une exposition excessive aux rayonnements UV ○ L'exposition à certains virus ou bactéries ○ Hormonale 	<ul style="list-style-type: none"> ○ Le contact avec certains produits physiques et chimiques, (amiante, benzène, poussière de bois...) ○ L'exposition à certaines substances polluantes présentes dans l'air, la terre ou l'eau ○ Les rayonnements (d'origine naturelle, comme le radon, ou artificielle comme les rayons X) ○ Infection Chronique

I.1.11 Les différents traitements du cancer

Il existe différents traitements du cancer utilisés seuls ou associés entre eux :

I.1.11.1 La chirurgie oncologique

Est l'ablation d'une tumeur par un chirurgien lors d'une opération. Dans la plupart des cas, le chirurgien enlève la tumeur et des tissus qui l'entourent. L'élimination de ces tissus

voisins aide à empêcher la réapparition de la tumeur. Le chirurgien peut aussi enlever quelques ganglions lymphatiques proches. [18]

I.1.11.2 La radiothérapie, ou irradiation

Est un traitement anticancéreux qui consiste à utiliser des rayons à haute énergie pour détruire ou endommager les cellules cancéreuses.

I.1.11.3 L'hormonothérapie

Est un traitement qui consiste à bloquer l'action ou la production d'hormones naturelles afin d'empêcher le développement des cellules cancéreuses. Contrairement à la chimiothérapie ou à la radiothérapie, qui cherchent à tuer rapidement les cellules cancéreuses, l'hormonothérapie vise à entraîner leur mort à plus long terme en créant un milieu hormonal qui leur est défavorable.

I.1.11.4 La chimiothérapie

Est une méthode thérapeutique utilisée essentiellement contre le cancer. La chimiothérapie consiste à administrer des médicaments qui détruisent les cellules cancéreuses. Elle est utilisée en association avec de la chirurgie et de la radiothérapie. [19]

I.1.11.5 Immunothérapie

Utilisés en cancérologie, les traitements par immunothérapie n'agissent pas directement sur les cellules malignes. Ils ciblent au contraire le système immunitaire dans son ensemble. Objectif : le stimuler pour renforcer les défenses du malade contre les cellules cancéreuses. [20]

I.2.1 Introduction

La chimie pharmaceutique est une discipline qui s'intéresse à la conception, à la préparation et à l'interprétation du mode d'action des médicaments. C'est une science qui tente, entre autres, d'établir des relations entre la structure chimique des drogues et leur activité biologique.

Parmi ces drogues, on trouve les agents cytotoxiques qui sont des médicaments utilisés comme des produits anticancéreux. [21]

Le cancer est une maladie dévastatrice; Cependant, plusieurs avancées thérapeutiques ont récemment été réalisées, dans lesquelles l'EGFR, MCF7 et HepG2 sont devenus des biomarqueurs utiles et des cibles thérapeutiques.

EGFR, une glycoprotéine transmembranaire est un membre de la superfamille des tyrosines kinases du récepteur ERBB. L'EGFR se lie à son ligand apparenté EGF, qui induit en outre la phosphorylation de la tyrosine et la dimérisation du récepteur avec d'autres membres de la famille, conduisant à une prolifération non contrôlée accrue.

MCF-7 est le nom de la lignée de cellules tumorales mammaires largement étudiée dérivée de l'adénocarcinome du sein. Les cellules MCF7 peuvent être utilisées pour détecter l'implication PI3K et MAPK, ainsi que la détection facile de la phosphorylation ERK et Akt. En outre, via les récepteurs d'oestrogènes cytoplasmiques, ces cellules ont la capacité de traiter l'œstradiol.

Les cellules HepG2 sont un système de modèle in vitro viable pour l'étude des hépatocytes humains polarisés. Les cellules HepG2 présentent une différenciation morphologique et fonctionnelle robuste avec une formation contrôlable des domaines de surface cellulaire apical et basolatéral.

Plusieurs traitements anti-EGFR, MCF7 et HepG2 tels que les anticorps monoclonaux et les inhibiteurs de la tyrosine kinase, aromatase cytochrome p450 et proprotéine convertase subtilisine/kexine de type 9 ont été développés, ce qui a permis aux cliniciens d'identifier et de traiter des cohortes de patients spécifiques. [22-23]

I.2.2 Les interactions biomoléculaires (récepteur-ligand)

Les interactions moléculaires sont à la base de tous les systèmes vivants. Elles interviennent de façon fondamentale dans la plupart des processus vitaux comme le métabolisme, la réplication ou la croissance. Ces interactions impliquent un contact étroit et donc une complémentarité structurale entre deux molécules. Les deux partenaires impliqués dans cette reconnaissance moléculaire sont définis de la façon suivante : [24]

I.2.2.1 Une espèce reconnue par le récepteur

La nature de cette espèce peut être très variable et le terme général de **ligand** (Sont des molécules de signalisation qui en se fixant sur des récepteurs spécifications Déterminent une réponse cellulaire). Est le plus souvent employé pour désigner cette cible. [24, 25]

I.2.2.2 Un récepteur

Sont des protéines soit exprimés à la surface des cellules –cibles, soit sont des protéines intracellulaires du cytosol ou du noyau des Cellules – cibles. [25]

La reconnaissance moléculaire entre les deux entités résulte de leur complémentarité fonctionnelle, chimique et géométrique. L'image souvent utilisée pour illustrer cette complémentarité est le système « clé-serrure ».

Les interactions en biologie peuvent s'effectuer en solution comme dans le cas d'une enzyme soluble et son substrat mais elles opèrent le plus souvent aux interfaces comme dans le cas de l'adhésion des cellules à la matrice extracellulaire ou dans les processus d'adhésion des virus et des bactéries aux cellules eucaryotes. [24]

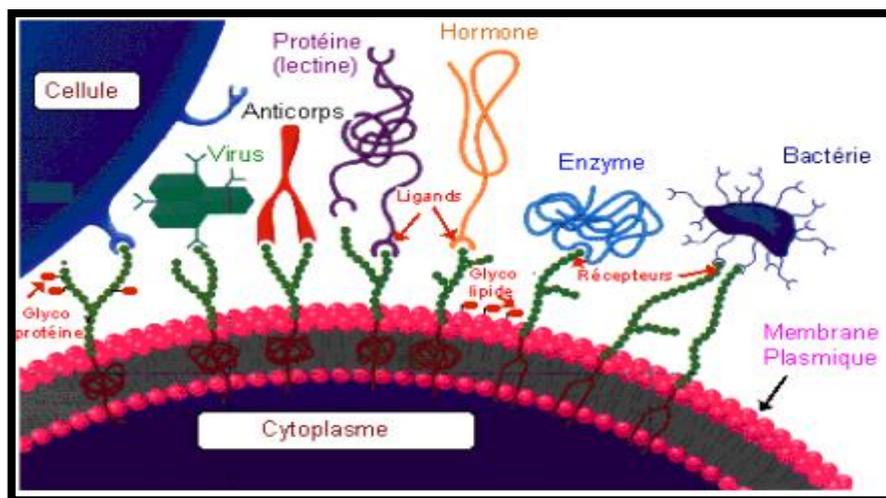


Figure I.5 : Différents partenaires impliqués dans les interactions biomoléculaires

I.2.3 Epidermal Growth Factor Receptor (EGFR)

Le récepteur de l'EGF (Epidermal Growth Factor) ou EGFR est une protéine monomérique transmembranaire qui transduit le signal consécutif à sa liaison au facteur de croissance épidermique. C'est une protéine à activité tyrosine kinase intrinsèque. Il présente des similitudes avec le récepteur de l'insuline. Il appartient à la RTK famille des récepteurs à activité tyrosine kinase. Son gène est l'EGFR porté par le chromosome 7 humain. [26]

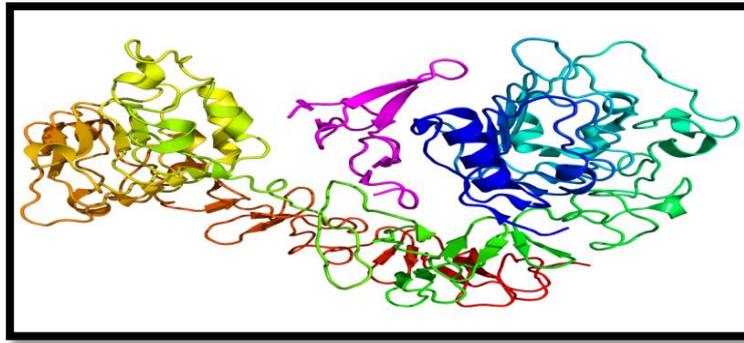


Figure I.6 : Epidermal growth factor receptor (EGFR)

Les récepteurs de la famille de l'EGFR sont ancrés dans la membrane cytoplasmique et se composent de trois domaines. Ils possèdent un domaine extracellulaire riche en cystéine et subdivisé en 4 sous domaines (I, II, III, IV). Les domaines I et III sont impliqués dans la liaison des ligands alors que le domaine II est important dans l'interaction entre les récepteurs, indépendante des ligands. Une courte région hydrophobe transmembranaire relie le domaine extracellulaire à un domaine tyrosine kinase intra cytoplasmique, très conservé dans les récepteurs de la famille EGFR, possédant des résidus tyrosines qui sont phosphorylés suite à l'activation des récepteurs. [27]

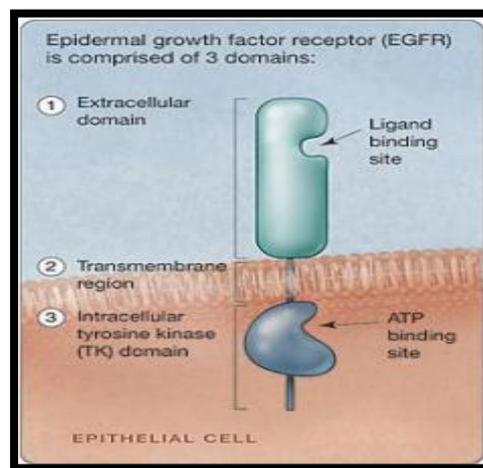


Figure I.7 : Structure de l'EGFR. L'EGFR possède un domaine extracellulaire, une région transmembranaire et un domaine tyrosine kinase intracellulaire.

Le gène de l'Epidermal Growth Factor Receptor (EGFR) ou récepteur du facteur de croissance épidermique est connu depuis longtemps. Il code une glycoprotéine transmembranaire appartenant à la famille des récepteurs tyrosine kinase comptant 4 membres : egfr ou erbB1, erbB2 ou HER-2/neu (Human Epidermal Receptor), erbB3 ou

HER-3 et erbB4 ou HER-4. Ces récepteurs sont exprimés dans de nombreux tissus d'origine mésodermique et ectodermique. [27]

I.2.4 Michigan Cancer Foundation-7 (MCF7)

MCF-7 est une lignée cellulaire de cancer du sein humain avec des récepteurs d'oestrogène, de progestérone et de glucocorticoïde. Il est dérivé de l'épanchement pleural d'un cancer du sein métastatique caucasien de 69 ans (adénocarcinome) en 1970 par le Dr Soule de la Michigan Cancer Foundation, Detroit, MI. Les cellules MCF-7 sont utiles pour les études mammaires in vitro car elles conservent plusieurs caractéristiques idéales particulières à l'épithélium mammaire, comme le traitement des œstrogènes, sous forme d'œstradiol, via les récepteurs œstrogéniques (RE) dans le cytoplasme des cellules. C'est la première lignée cellulaire de cancer du sein répondant aux hormones. Ses propriétés uniques sont utiles pour les thérapeutiques expérimentales. Ils sont également sensibles à la cytokératine. Lorsqu'elles sont cultivées in vitro, les cellules sont capables de former des dômes et les cellules de type épithélial se développent en monocouches. La figure 4.4 montre une vue des cellules MCF-7. [28]

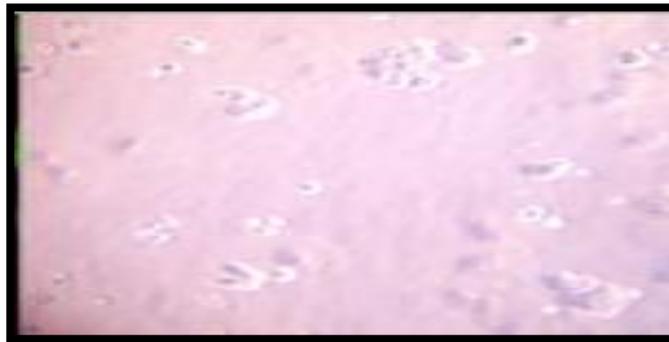


Figure I.8 : Cellules MCF-7

I.2.5 Carcinome hépatocellulaire (HepG2)

HepG2 est une lignée cellulaire immortalisée constituée de cellules de carcinome hépatique humain, dérivées du tissu hépatique d'un homme caucasien de 15 ans ayant un carcinome hépatocellulaire bien différencié. Le carcinome hépatocellulaire est le cinquième cancer le plus commun dans le monde. La morphologie des cellules HepG2 est épithéliale et contient 55 paires de chromosomes. Les cellules HepG2 peuvent être cultivées avec succès à grande échelle et sécréter de nombreuses protéines plasmatiques, telles que la transferrine, le fibrinogène, le plasminogène et l'albumine. Ils peuvent être stimulés avec l'hormone de croissance humaine. Les cellules HepG2 sont des cellules épithéliales adhérentes qui se développent en monocouches et en petits agrégats. [29]

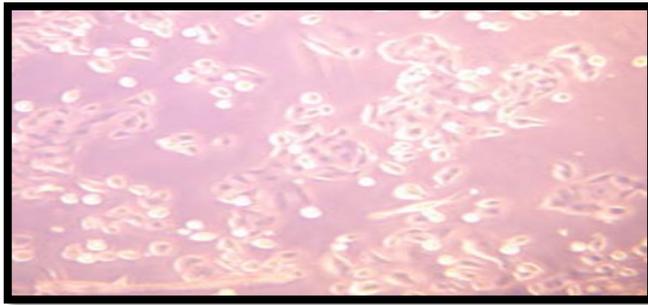


Figure I.9 : Cellules HepG2

I.2.6 Cytologie

Partie de la biologie qui étudie la cellule vivante sous tous ses aspects (kutos = cellule).

I.2.7 Cytostatique

Se dit des substances ayant la propriété de bloquer la synthèse, le fonctionnement ou la multiplication cellulaire. Les médicaments administrés dans le cadre d'une chimiothérapie sont appelés cytostatiques. [30]

I.2.8 Cytotoxique

La cytotoxicité est une des stratégies majeures employées par le système immunitaire pour combattre les agressions extérieures et les dysfonctionnements de l'organisme. [31] Les médicaments cytotoxiques inhibent ou empêchent le bon fonctionnement des cellules [31]. On les utilise surtout pour traiter le cancer, souvent dans le cadre d'une chimiothérapie. Dernièrement, leur utilisation s'est étendue au traitement de certaines maladies de la peau comme le psoriasis, de l'arthrite rhumatoïde, de l'arthrite rhumatoïde juvénile, de même que des problèmes musculaires résistant aux stéroïdes [33]. Les médicaments cytotoxiques les plus connus sont les antinéoplasiques. Le terme antinéoplasique est d'ailleurs parfois utilisé au lieu de médicament cytotoxique. [34]

I.2.9 Mécanisme d'action des agents cytotoxiques :

Leur activité cytotoxique résulte de leur interaction avec certains substrats cellulaires indispensables à la vie ou à la division cellulaire : ADN, ARN, protéines, enzymes.

I.2.9.1 Classification suivant leurs propriétés pharmacologiques

Les agents cytotoxiques peuvent être classés schématiquement selon leurs cibles cellulaires et/ou leur mécanisme d'action. [35]

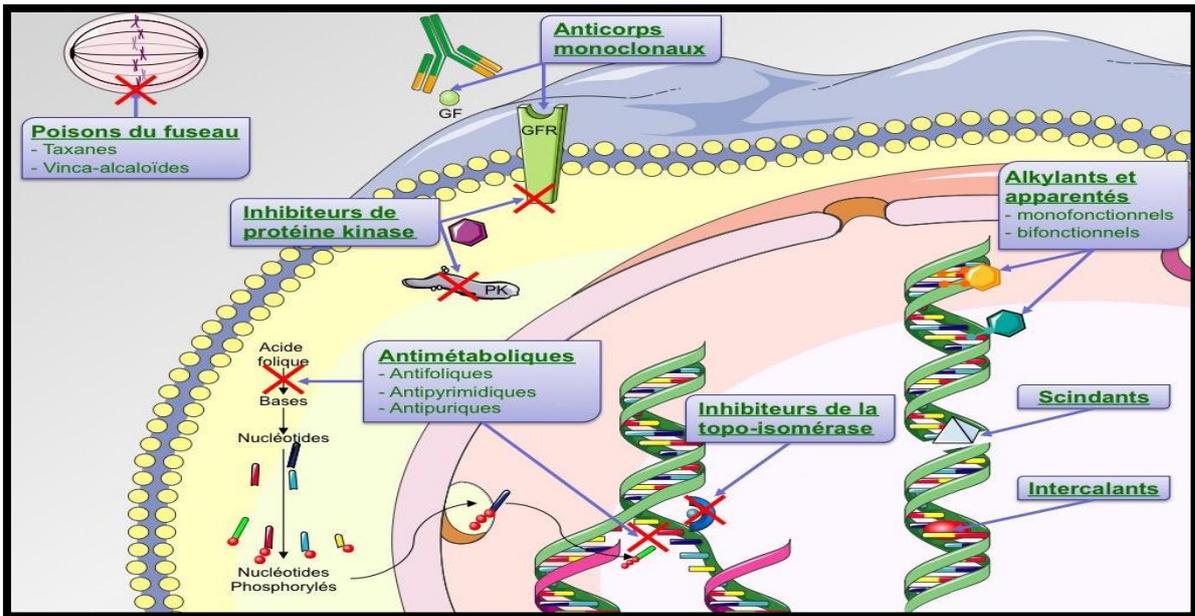


Figure I.10 : Cible des anticancéreux. GF: Growth Factor = Facteur de croissance, GFR: Growth Factor Receptor = Récepteur au Facteur de croissance, PK: Protéine Kinase

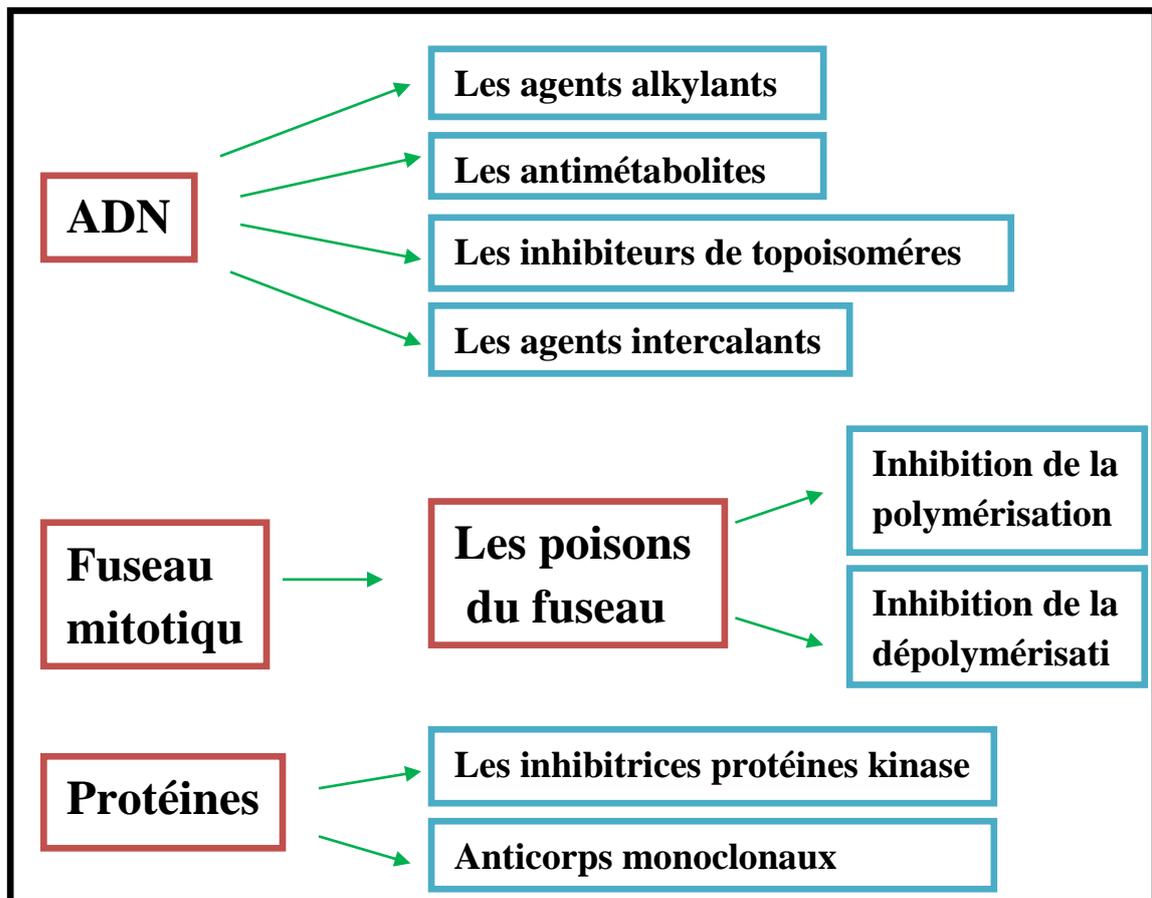


Figure I.11 : Classification des agents cytotoxiques

I.2.9.1.1 ADN

I.2.9.1.1.1 Les agents alkylants

Un agent alkylant agit en substituant un radical alkyl réactif à un proton d'une structure nucléophile hautement réactive de l'ADN. [36]

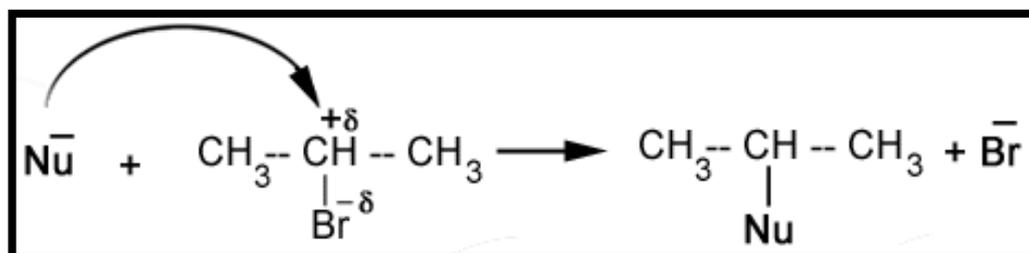


Figure I.12 : Réaction chimique entre un nucléophile et un électrophile formation d'une liaison covalente.

Il en découle la formation de ponts inter ou intracaténaux non réparables. On a donc une altération de structure de l'ADN ainsi qu'une interférence lors du transfert de l'information entre ADN et ARN. [36]

I.2.9.1.1.2 Les agents intercalants

Un agent intercalant est une molécule polycyclique (caractérisées par plusieurs noyaux aromatiques) plane capable de se glisser entre les plateaux que constituent 2 paires de bases contiguës de l'ADN entraînant ainsi un éloignement de ces paires de bases et une désspiralisation de l'ADN [37]. Donc un empêchement de la progression des ARN et ADN polymérases ainsi qu'une inhibition de la réplication et de la transcription. [38]

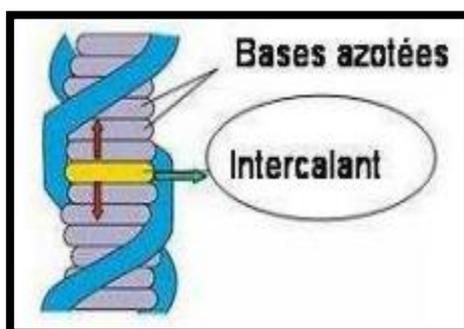


Figure I.13: Intercalation directe avec l'ADN

La présence de substance intercalée entre les deux brins de l'ADN peut perturber aussi l'action de l'ADN topoisomérases I et II et provoquer des cassures mono- et bicaténales. [38]

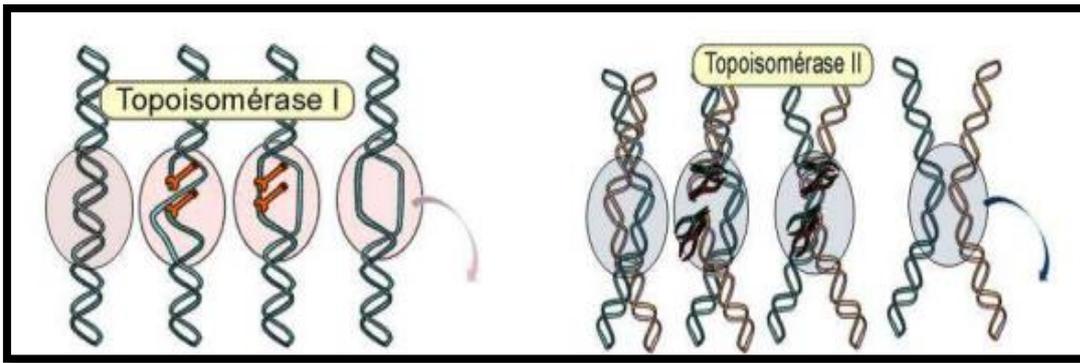


Figure I.14 : Activité de la topoisomérase I et de la topoisomérase II

I.2.9.1.1.3 Les inhibiteurs de topoisomères

Les topoisomérases sont des enzymes assurant la condensation / décondensation de l'ADN après avoir créé des coupures transitoires de l'un (topoisomérase I) ou des deux (topoisomérase II) brins, puis leur ligation, permettant une relaxation des forces de torsion générées au moment de la réplication. [39]

Les inhibiteurs de topoisomérase II empêchent cette religation en stabilisant le complexe (ADN-enzyme) normalement transitoire bloquant la progression de la fourche de réplication : l'ADN reste cassé.

I.2.9.1.1.4 Les antimétabolites

Ces composés vont bloquer la synthèse des acides nucléiques en phase S du cycle cellulaire, soit en inhibant les protéines nécessaires au métabolisme des bases puriques et pyrimidiques, soit en se substituant aux bases azotées, soit encore en hydrolysant la lasparagine, qui est un précurseur de la synthèse protéique. [39]

I.2.9.1.2 Fuseau mitotique

I.2.9.1.2.1 Les poisons du fuseau

Ces agents d'origine végétale [40] sont des agents inhibant les microtubules, qui peuvent être des composés naturels ou synthétiques avec des structures variées, incluent les composés qui interfèrent avec soit la polymérisation, soit la dépolymérisation des microtubules, supprimant ainsi la dynamique des microtubules et conduisant la cellule vers l'apoptose. [41-52]

Grâce à leur activité antitumorale potentielle, ces composés ciblant les microtubules sont parmi les plus fréquemment prescrits en tant qu'agents cytotoxiques. [43]

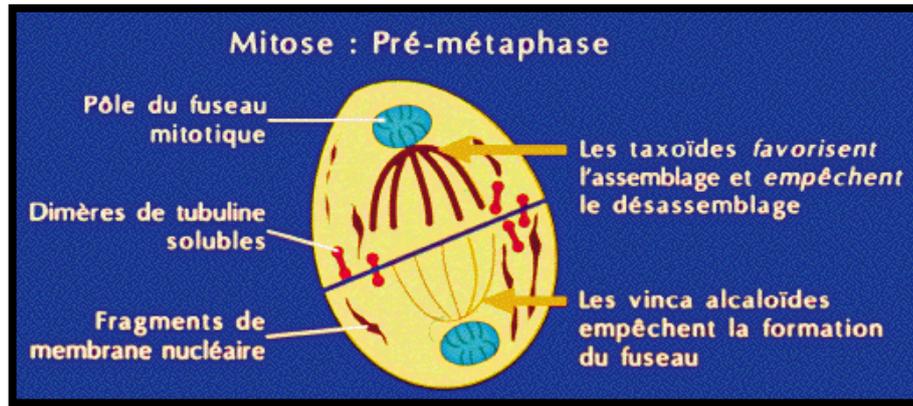


Figure I.15 : Effet de poisons du fuseau sur la mitose

I.2.9.1.2.2 Inhibition de la polymérisation

Empêchent la polymérisation du fuseau mitotique.

I.2.9.1.2.3 Inhibition de la dépolymérisation

Facilitent la polymérisation de la tubuline et en inhibent la dépolymérisation. [36]

I.2.9.1.3 Protéines

I.2.9.1.3.1 Les tyrosines kinases

Les tyrosines kinases sont des enzymes, qu'il s'agisse de récepteurs ou non, qui jouent un rôle majeur dans la signalisation cellulaire en aval des facteurs de croissance. Les tyrosines kinases assurent le transfert d'un groupement phosphate de l'adénosine triphosphate (ATP) vers une protéine effectrice impliquée dans de nombreux processus de régulation cellulaire.

Les récepteurs à activité tyrosine kinase sont des protéines transmembranaires impliquées dans la transduction intracytoplasmique du signal émanant du niveau extracellulaire. À l'intérieur de la cellule, d'autres protéines tyrosine kinases jouent également un rôle essentiel dans la transduction du signal. [44]

I.2.9.1.3.2 Les inhibitrices protéines kinase

L'activation de ces protéines, récepteurs ou protéines intracellulaires, permet d'induire la prolifération et la croissance cellulaire tumorale, de réprimer l'apoptose, et de promouvoir l'angiogenèse et la diffusion métastatique.

Les inhibiteurs de tyrosine kinase se fixent de manière compétitive sur les sites de liaisons de l'ATP et bloquent ainsi l'activation des sites tyrosine kinase. Par voie de conséquence, la signalisation cellulaire en aval est interrompue, rétablissant ainsi le contrôle de la prolifération de la survie cellulaire. Dans les tumeurs solides, les inhibiteurs

de tyrosine kinase induisent également une inhibition de l'angiogenèse et de la diffusion métastatique. [44]

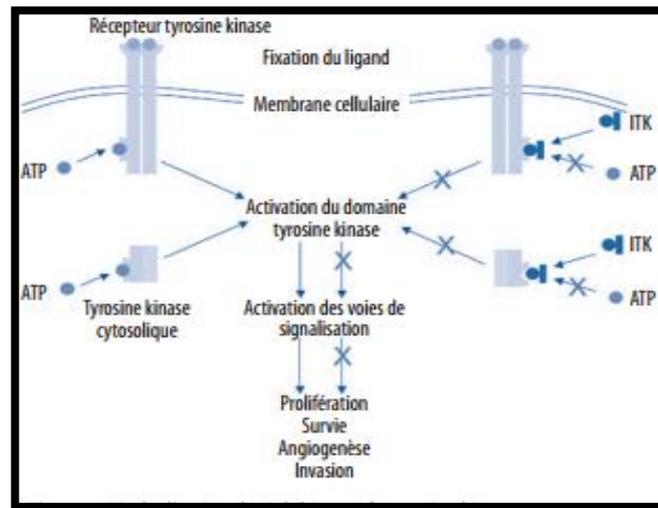


Figure I.16 : Mode d'action des inhibiteurs de tyrosine kinase

I.2.9.1.4 Anticorps monoclonaux

I.2.9.1.4.1 Les anticorps

Les anticorps sont des protéines en forme de Y et sont des éléments essentiels du système immunitaire. Ils sont produits par les cellules B. Le corps les produit en réponse aux intrusions de bactéries, de virus ou d'autres parasites pour les marquer puis les détruire. Ces anticorps sont composés de deux chaînes d'acides aminés lourdes et de deux légères reliées les unes aux autres. [45]

I.2.9.1.4.2 Anticorps monoclonaux

Köhler entreprit de fusionner des globules blancs (cellules B) sains avec des cellules tumorales bien définies dans le but de produire des anticorps monoclonaux. La cellule ainsi produite fabriquera non seulement des anticorps comme le font les cellules B mais elles se multiplieront à l'infini comme le font les cellules tumorales. Toutes les cellules-filles sont absolument identiques à la cellule-mère.

Les avantages des anticorps monoclonaux sont surtout prédominants dans le traitement du cancer. Tout traitement cancéreux a pour but de détruire la majeure partie des cellules atteintes par le cancer tout en ayant le moins d'effets secondaires possibles. Lorsqu'une caractéristique particulière se trouve sur la cellule d'une tumeur, un anticorps monoclonal pourra être développé exactement contre celle-ci. Les anticorps reconnaissent ainsi de façon idéale les cellules tumorales et donnent le signal de destruction. Les anticorps ne reconnaissant que la caractéristique de la cellule tumorale, ils ne pourront guère affecter

d'autres cellules. Les anticorps pourront alors atteindre les colonies de cellules tumorales du corps, les plus petites comme les invisibles, et les détruire presque sans effets secondaires. [46]

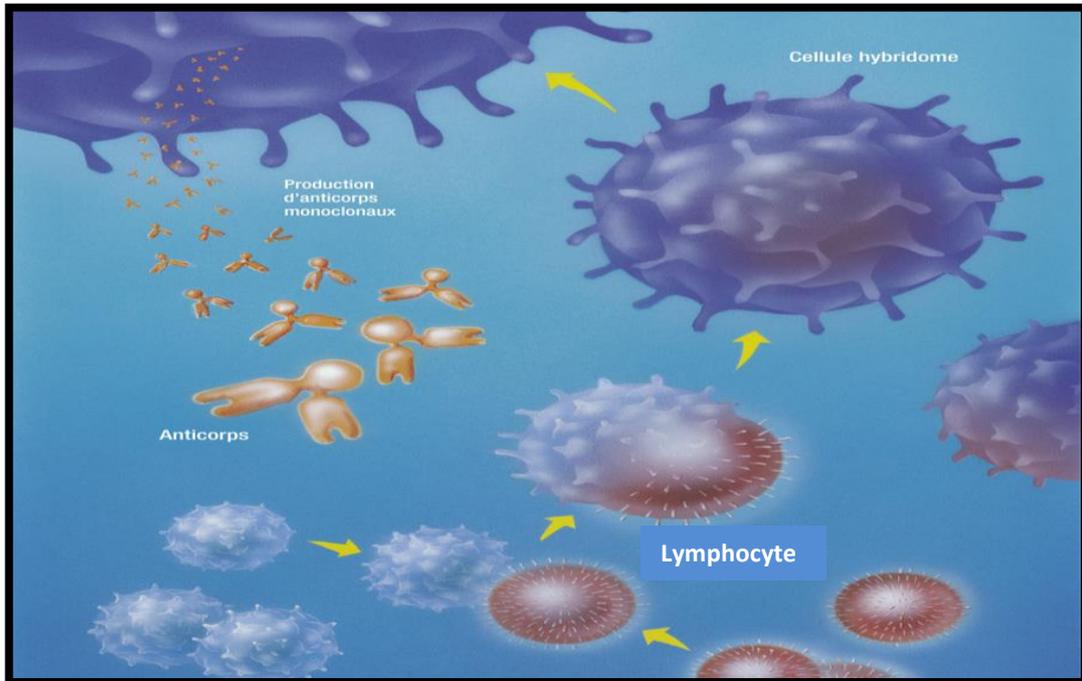


Figure I.17 : Mode d'action des Anticorps monoclonaux

I.2.10 Le test de cytotoxicité in vitro

Le test de cytotoxicité permet la recherche d'un effet toxique sur un type cellulaire par une méthode évaluant un paramètre donné (croissance cellulaire, intégrité cellulaire, activité fonctionnelle cellulaire...). Son but est de déterminer la concentration inhibitrice 50% (IC50) qui est la concentration qui inhibe de moitié la prolifération cellulaire globale. Plusieurs méthodes d'étude de la cytotoxicité existent à ce jour ; entre autres les méthodes fondées sur des perturbations de la perméabilité membranaire et les méthodes fondées sur des altérations de la prolifération cellulaire. [47]

I.3.1 Benzimidazole

I.3.1.1 Définition

Le Benzimidazole est un composé hétérocyclique aromatique résultant de la fusion d'un cycle benzénique et d'un cycle imidazole. Ce composé bi-cyclique possède un caractère basique ($pK_a = 5.68$) moindre que celui de l'imidazole et un caractère acide du NH plus prononcé ($pK_a = 12.75$). [48]

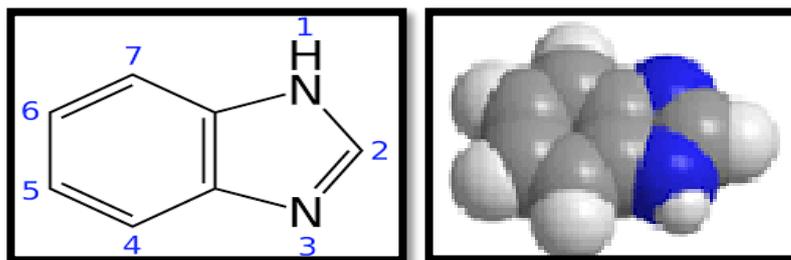


Figure I.18 : structure de Benzimidazole

I.3.1.2 L'activité biologique générale

Les benzimidazoles rapportés pour leurs activités anti-inflammatoires, analgésiques, anticancéreux, antinéoplasiques, anticonvulsantes, antidiabétiques et antihelminthiques ainsi que pour leurs effets sur le système cardiovasculaire et le niveau de cholestérol. [49]

I.3.1.3 L'activité biologique sur le cancer

Les Benzimidazoles sont structurellement apparentés à l'indole et sont des ligands des récepteurs de la sérotonine (5-HT), des récepteurs de l'histamine (H4), des récepteurs de la bradykinine (B2) et des récepteurs de la dopamine (D4) [50]. De plus, le benzimidazole étant un isostère de nucléosides puriques et un échafaudage important dans diverses molécules biologiquement actives est largement exploré pour le développement d'agents anticancéreux [51-52]. Certains des médicaments en cours d'essais cliniques contenant des fractions benzimidazole et oxadiazole et des composés ciblés rationnellement conçus sont représentés.

En 2012, Rashid et al [53] ont synthétisé l'oxadiazole substitué par le benzimidazole-2 et les composés ont montré une activité anticancéreuse remarquable contre la plupart des lignées cellulaires testées avec des valeurs de GI50 comprises entre 0,79 et 17,8 μ M. Salahuddin et al.

En 2014[54]. Ont synthétisé des dérivés de benzimidazole-1-oxadiazole et trouvé un composé actif contre les lignées cellulaires de cancer du sein (MDA-MB-468) et ont montré 72,85% de croissance (GP). Le présent travail est orienté vers la conception et la

synthèse de dérivés d'oxadiazole liés au benzimidazole en utilisant des produits chimiques facilement accessibles, des voies de synthèse faciles et l'évaluation de l'EGFR et du test d'inhibition erbB-2 en utilisant le géfitinib comme témoin positif. Les données de la littérature concernant l'activation de l'EGFR ont été associées au développement et à la progression des tumeurs humaines à savoir. Poitrine, foie, poumons et kératinocyte humain immortalisé. Ainsi, des études de cytotoxicité ont été déterminées sur cinq lignées cellulaires tumorales humaines (MCF7 "breast", HEPG2 "foie", MDA-MB231 "breast", et A549 "carcinome pulmonaire humain") en utilisant un dosage MTT. [54]

Références bibliographique

- [1]: Ananya Mandal. Histoire de Cancer, New Medical Life Sciences .Oct 30, 2017.
- [2]: Sante – médecine. Journal des femmes.fr-tumeur de finition .
- [3]: Site : https://fr.wikipedia.org/wiki/Tumeur_b%C3%A9nigne
- [4]: Sante – médecine. Journal des femmes.fr-tumeur malinge.
- [5]: Chafia, T. I. G. R. I. N. E. Effets anticancéreux et chimioprotecteur de l'extrait polyphénolique, riche en flavonoïdes, des feuilles de Cléome arabica. Diss. Université Ferhat Abbas de Sétif 1, 2014.
- [6]: SHERBET, Gajanan V. et LAKSHMI, M. S. The genetics of cancer: genes associated with cancer invasion, metastasis and cell proliferation. Academic Press, 1997.
- [7]: King, R.J.B., Cancer biology. Edition Harlow, Essex (England), 1996, 227p
- [8]: Liotta, L.A., 1992. Cancer cell invasion and metastasis. Scientific American 266, 54-59.
- [9]: Weinberg, R.A., One renegade cell: how cancer begins, Basic Books Publisher (New York), 1998, 170p
- [10]: MCCARTHY, Nicola. Angiogenesis: Blood born killer?. Nature Reviews Cancer, 2008, vol. 8, no 3, p. 164.
- [11]: Berstein, L.M., Santen. R., 2008. Innovative endocrinology of cancer. In Advances in Experimental Medicine & Biology, 630.
- [12]: Fajardo, E., Geller, T., Koenig, H.M., et Kleine, M. L., 1982. Chronic meningitis, polyarthritis, lymphadenitis, and pulmonary hemosiderosis. The Journal of pediatrics 1, 738-740.
- [13] : Bettaieb, A., 2009. Mécanismes d'induction de l'apoptose par le choc thermique et effet protecteur de la thermotolérance induite à 40°C. Thèse de Doctorat en Biologie, Université du Québec, Montréal.
- [14]:Site : <http://santecheznous.com/condition/getcondition/cancer#facts>

- [15]: Sante-médecine.journal des femmes.fr/contents/118-cancer-définition-symptômes-et-traitement
- [16]: Shuster, John L., et al., eds. When the focus is on care: Palliative care and cancer. Amer Cancer Society, 2005.
- [17]: Institut national du cancer -Principaux facteurs de risque de cancer
- [18]: fondation contre le cancer -traitements-du-cancer
- [19]: Sante – médecine. Journal des femmes.fr-chimiothérapie-définition-et-effets-secondaires
- [20]: de notre envoyé spécial au 47ème congrès de l'American Society of Clinical Oncology(ASCO), Chicago, 3-7 juin 2011 - Manuel Merck, 4e Edition
- [21]: Pirotte, Bernard, et al. " Chimie pharmaceutique (Partie II)." European journal of medicinal chemistry Université de Liège
- [22]: SESHACHARYULU, Parthasarathy, PONNUSAMY, Moorthy P., HARIDAS, Dhanya, et al. Targeting the EGFR signaling pathway in cancer therapy. Expert opinion on therapeutic targets, 2012, vol. 16, no 1, p. 15-31.
- [23]: Fagan, Dedra H., et al. "Acquired tamoxifen resistance in MCF-7 breast cancer cells requires hyperactivation of eIF4F-mediated translation." Hormones and Cancer 8.4 (2017): 219-229.
- [24]: SANDRIN, Ludivine. Caractérisation des interactions biomoléculaires entre des ligands peptidiques immobilisés sur une surface et des récepteurs cellulaires. 2009. Thèse de doctorat. Université Joseph-Fourier-Grenoble I.
- [25]: Benahmad, Interaction ligand-récepteur .2015/2016.Faculte Médecine d'Annaba .
- [26]: Site : https://fr.wikipedia.org/wiki/R%C3%A9cepteur_de_l%27EGF
- [27]: HYNES, Nancy E. et LANE, Heidi A. ERBB receptors and cancer: the complexity of targeted inhibitors. Nature Reviews Cancer, 2005, vol. 5, no 5, p. 341.
- [28]: Sundararajan, Raji, ed. Electroporation-based therapies for cancer: from basics to clinical applications. Elsevier, 2014

- [29]: Lin, Chih-Li, Hsiu-Chen Huang, and Jen-Kun Lin. "Theaflavins attenuate hepatic lipid accumulation through activating AMPK in human HepG2 cells." *Journal of lipid research* 48.11 (2007): 2334-2343
- [30]: Site : <https://www.futura-sciences.com/sante/definitions/biologie-cytostatique-2657/>
- [30]: VÉLY, Frédéric et VIVIER, E. Mécanismes moléculaires de la cytotoxicité des cellules NK. 1996.
- [32]: CONNOR, Thomas H. et MCDIARMID, Melissa A. Preventing occupational exposures to antineoplastic drugs in health care settings. *CA: a cancer journal for clinicians*, 2006, vol. 56, no 6, p. 354-365.
- [33]: Cytotoxic Drugs and Related Waste Guide (2008) préparé par le Cytotoxic drugs working party for WorkCover, gouvernement de la Nouvelle-Galles du Sud, Australie
- [34]: Cytotoxic Drugs, (Nov. 1999) Occupational Health and Safety Division, gouvernement de la Saskatchewan.
- [35]: Site: <https://pharmacomedicale.org/74-actualites/232-mise-a-jour-2017-du-site-du-cnpm>
- [36]: Anticancéreux : classification et mécanismes d'action, principes de leur utilisation thérapeutique et traitements associés
- [37]: RAPPEL PHARMACOLOGIQUE maj 2010
- [38]: BINASCHI, Monica, ZUNINO, Franco, et CAPRANICO, Giovanni. Mechanism of action of DNA topoisomerase inhibitors. *Stem Cells*, 1995, vol. 13, no 4, p. 369-379.
- [39]: LENGLET, Gaëlle. Mécanisme d'action de nouveaux agents alkylants ciblant l'ADN ou les protéines. 2010. Thèse de doctorat. Université du Droit et de la Santé-Lille II.
- [40]: Traitement médical du cancer : la chimiothérapie./ Dr.Lalami (Bordet) Formation postuniversitaire AFPHB La chimiothérapie des cancers,2003
- [41]: JORDAN, Mary Ann et WILSON, Leslie. Microtubules and actin filaments: dynamic targets for cancer chemotherapy. *Current opinion in cell biology*, 1998, vol. 10, no 1, p. 123-130.
- [42]: BHATTACHARYYA, Bhabatarak, PANDA, Dulal, GUPTA, Suvroma, et al. Anti-mitotic activity of colchicine and the structural basis for its interaction with

- tubulin. Medicinal research reviews, 2008, vol. 28, no 1, p. 155-183. [40]: CHEN, Si-Meng, MENG, Ling-Hua, et DING, Jian. New microtubule-inhibiting anticancer agents. Expert opinion on investigational drugs, 2010, vol. 19, no 3, p. 329-343.
- [43]: MERLIN, Jean-Louis. Les inhibiteurs de tyrosine kinase en oncologie. La Lettre du pharmacologue, 2008, vol. 22, no 2, p. 51-62.
- [44]: RENTIER, Bernard. Le système immunitaire: principes généraux. Journal de Pharmacie de Belgique, 1988, vol. 43, no 3, p. 185-211.
- [45]: Anticops monoclonaux :attaque des anticorps contre le cancer ,Centre didactiquebiotech
- [46]: GNOULA, Charlemagne. ETUDES DES PROPRIETES CYTOTOXIQUES ET ANTIRADICALAIRES D'EXTRAITS DE FEUILLES ET DE GALLES. 2009. Thèse de doctorat. UNIVERSITE D'OUAGADOUGOU.
- [47]: Eicher, T.; Hauptmann, S.; Speicher, A. "The Chemistry of Heterocycles: Structure, Reactions, Syntheses, and Applications";Eds. Wiley-VCH, Weinheim, Germany, 2003, pp.175-178.
- [48]: KLIMEŠOVÁ, Věra, KOČÍ, Jan, WAISSER, Karel, et al. New benzimidazole derivatives as antimycobacterial agents. Il Farmaco, 2002, vol. 57, no 4, p. 259-265.
- [49]: CHEN, Li-Hsun, CHANG, Chia-Mao, SALUNKE, Deepak B., et al. Divergent synthesis of unsymmetrical annulated biheterocyclic compound libraries: benzimidazole linked indolo-benzodiazepines/quinoxaline. ACS combinatorial science, 2011, vol. 13, no 4, p. 391-398.
- [50]: KAMAL, Ahmed, KUMAR, P. Praveen, SREEKANTH, K., et al.Synthesis of new benzimidazole linked pyrrolo [2, 1-c][1, 4] benzodiazepine conjugates with efficient DNA-binding affinity and potent cytotoxicity. Bioorganic & medicinal chemistry letters, 2008, vol. 18, no 8, p. 2594-2598.
- [51]: SHAO, Kun-Peng, ZHANG, Xu-Yao, CHEN, Peng-Ju, et al.Synthesis and biological evaluation of novel pyrimidine–benzimidazol hybrids as potential anticancer agents. Bioorganic & medicinal chemistry letters, 2014, vol. 24, no 16, p. 3877-3881.

[52]: RASHID, Mohd, HUSAIN, Asif, et MISHRA, Ravinesh. Synthesis of benzimidazoles bearing oxadiazole nucleus as anticancer agents. *European journal of medicinal chemistry*, 2012, vol. 54, p. 855-866.

[53]: SHAHARYAR, Mohammad, MAZUMDER, Avijit, AHSAN, Mohamed Jawed, et al. Synthesis, characterization and anticancer evaluation of 2-(naphthalen-1-ylmethyl/naphthalen-2-ylloxymethyl)-1-[5-(substituted phenyl)-[1, 3, 4] oxadiazol-2-ylmethyl]-1H-benzimidazole. *Arabian Journal of Chemistry*, 2014, vol. 7, no 4, p. 418-424.

[54]: AKHTAR, Md Jawaid, SIDDIQUI, Anees Ahmad, KHAN, Ahsan Ahmed, et al. Design, synthesis, docking and QSAR study of substituted benzimidazole linked oxadiazole as cytotoxic agents, EGFR and erbB2 receptor inhibitors. *European journal of medicinal chemistry*, 2017, vol. 126, p. 853-869.

II.1.1 Introduction

La chimie informatique est maintenant devenue un outil très important dans les domaines industriel et académique. Avec utilisation des méthodes théoriques calcul qui sont devenus de plus en plus sophistiquées [1].

Chimie théorique ou modélisation moléculaire est une discipline qui permet de prévoir la structure et la réactivité des molécules ou des systèmes de molécules aux moyens de programmes informatiques. Le développement des outils informatiques a permis de mettre au point des techniques de calculs de plus en plus poussées, rendant ainsi possible l'étude des systèmes de plus en plus complexes [2, 3]. Son but est l'obtention, par le calcul, de résultats appropriés aux problèmes rencontrés en chimie tels les propriétés thermodynamiques, les grandeurs géométriques, les fréquences de vibration, les constantes de couplage, les chemins réactionnels et d'autres grandeurs qui ne sont pas accessibles par l'expérience comme par exemple les propriétés de l'état de transition [3].

La recherche et la synthèse de nouveaux composés chimiques sont aujourd'hui souvent associées à une étude par modélisation moléculaire. La modélisation moléculaire est une technique permettant, non seulement de représenter les propriétés et les réactions chimiques mais aussi de manipuler les modèles des structures en deux ou trois dimensions [4].

La modélisation moléculaire est une application des méthodes théoriques et des méthodes de calcul pour résoudre des problèmes impliquant la structure moléculaire et la réactivité chimique [5]. Ces méthodes peuvent être relativement simples et utilisables rapidement ou au contraire elles peuvent être extrêmement complexes et demander des centaines d'heures de temps d'ordinateur, même sur un super-ordinateur. En plus, ces méthodes utilisent souvent des moyens infographiques très sophistiqués qui facilitent grandement la transformation de quantités impressionnantes de nombres en quelques représentations graphiques facilement interprétables [6].

Différentes approches sont envisageables dans le cadre des outils de modélisation moléculaire. Si ceux de mécanique classique, économiques en termes de temps de calcul, permettent de traiter des systèmes moléculaires de grande taille, les méthodes quantiques (ab initio, semi-empiriques ou théorie de la fonctionnelle de la densité) sont, quant à elles, capables de calculer les propriétés électroniques des systèmes. C'est pour cette raison que ces approches ont été employées dans le cadre de cette étude [7].

II.1.2 Modalisation moléculaire

La modélisation moléculaire est un outil destiné aux chercheurs préoccupés par la structure et la réactivité des molécules. La connaissance de la structure des édifices moléculaires permet de comprendre ce qui est réalisé dans une transformation physique, chimique ou biologique [8]. Elle peut permettre aussi de prévoir de telles transformations. La compréhension comme la prévision sont considérablement facilitées lorsque l'on peut visualiser les structures. Une molécule est correctement décrite par sa géométrie et ses propriétés thermodynamiques. La visualisation doit rendre compte de l'ensemble de ces caractéristiques. La question essentielle est de représenter une molécule sur l'écran de la façon la plus proche possible de la "réalité" [9].

II.1.3 Méthodes de la modélisation moléculaire

La modélisation moléculaire implique l'utilisation des méthodes de calcul théoriques (mécanique moléculaire, dynamique moléculaire, mécanique quantique ab initio ou semi-empirique,...) permettant de déterminer la représentation graphique de la géométrie ou de la configuration des atomes d'une molécule et d'évaluer les propriétés physico-chimiques de la molécule étudiée [10]. Ces méthodes peuvent être relativement simples et utilisables rapidement ou au contraire elles peuvent être extrêmement complexes et demander des centaines d'heures de temps d'ordinateur, même sur un super-ordinateur. En plus, ces méthodes utilisent souvent des moyens infographiques très sophistiqués qui facilitent grandement la transformation de quantités impressionnantes de nombres en quelques représentations graphiques facilement interprétables [11].

La modélisation moléculaire a pour but de prévoir la structure et la réactivité des molécules ou des systèmes de molécules. Les méthodes de la modélisation moléculaire peuvent être rangées en trois catégories [12]:

- ✓ La mécanique quantique (MQ).
- ✓ La mécanique moléculaire (MM).
- ✓ La dynamique moléculaire (DM).

II.1.3.1 La mécanique quantique (MQ) :

Les méthodes de la mécanique quantique, qui font appel à la distribution des électrons répartis en orbitales autour de la molécule, impliquent des temps de calcul souvent élevés qui limitent leur usage à des petites molécules ou nécessitent le recours à de nombreuses approximations. Elles sont particulièrement adaptées au calcul des charges et des potentiels électrostatiques, à l'approche des mécanismes réactionnels ou à la polarisabilité. L'objectif de

la mécanique quantique est principalement de déterminer l'énergie et la distribution électronique [13].

La chimie quantique définit la structure moléculaire comme un noyau autour duquel gravitent des électrons, qui sont décrits par leur probabilité de présence en un point et représentés par des orbitales [14].

La chimie quantique applique les principes de la mécanique quantique aux systèmes moléculaires pour tenter de résoudre l'équation de Schrödinger [15]. En effet, le comportement électronique et nucléaire des molécules, étant responsable des propriétés chimiques, peut être décrit de façon réaliste à partir de cette équation. Différentes méthodes de résolution ont alors été développées. En particulier, le développement grandissant des moyens informatiques ont permis le développement de ces méthodes [16].

II.1.3.1.1 Equation de Schrödinger :

Le calcul de l'énergie totale d'un système composé d'ions et d'électrons en interaction est obtenu dans le cas général par la résolution de l'équation de Schrödinger des états stationnaires [17].

$$H \psi = E \psi$$

- H l'Hamiltonien.
- ψ une fonction d'onde décrivant l'état du système.
- E son énergie totale.

Généralement, l'Hamiltonien H est donné par : $\mathbf{H} = \mathbf{T}_e + \mathbf{T}_N + \mathbf{V}_{ee} + \mathbf{V}_{eN} + \mathbf{V}_{NN}$

- T_e l'énergie cinétique des électrons :

$$\hat{T}_e = -\sum_{k=1}^{2n} \frac{\hbar^2}{2m_e} \nabla_k^2$$

- T_N l'énergie cinétique des noyaux :

$$\hat{T}_N = -\sum_{A=1}^N \frac{\hbar^2}{2M_A} \nabla_A^2$$

- V_{ee} l'énergie potentielle d'interaction entre les électrons :

$$\hat{V}_{ee} = \sum_{k=1}^{2n} \sum_{l>k}^{2n} \frac{e^2}{4\pi\epsilon_0} \frac{1}{r_{kl}}$$

- V_{eN} l'énergie potentielle d'interaction entre électrons-noyaux:

$$\hat{V}_{Ne} = - \sum_{k=1}^{2n} \sum_{A=1}^N \frac{e^2}{4\pi\epsilon_0} \frac{Z_A}{r_{kA}}$$

- VNN l'énergie potentielle d'interaction entre les noyaux :

$$\hat{V}_{NN} = \sum_{A=1}^N \sum_{B>A}^N \frac{e^2}{4\pi\epsilon_0} \frac{Z_A Z_B}{R_{AB}}$$

Dans le cas d'un système composé de n électrons en interaction avec N noyaux et décrits par la fonction d'onde poly-électronique $\Psi(r_1, r_2, \dots, r_n)$, d'énergie E, l'hamiltonien total du système est défini en unités atomiques par :

$$\hat{H} = - \sum_{k=1}^{2n} \frac{\hbar^2}{2m_e} \nabla_k^2 - \sum_{k=1}^{2n} \sum_{A=1}^N \frac{e^2}{4\pi\epsilon_0} \frac{Z_A}{r_{kA}} + \sum_{k=1}^{2n} \sum_{l>k}^{2n} \frac{e^2}{4\pi\epsilon_0} \frac{1}{r_{kl}} - \sum_{A=1}^N \frac{\hbar^2}{2M_A} \nabla_A^2 + \sum_{A=1}^N \sum_{B>A}^N \frac{e^2}{4\pi\epsilon_0} \frac{Z_A Z_B}{R_{AB}}$$

Dans le cas général des systèmes d'intérêt chimique, qui sont le plus souvent polyatomiques et multi-électroniques, cette équation ne peut être résolue analytiquement. Des approximations ont donc été proposées [18].

II.1.3.1.2 L'approximation de Born-Oppenheimer :

Les noyaux ayant une masse beaucoup plus importante que celle des électrons, il est alors possible, selon l'approximation de Born-Oppenheimer [19], de considérer leurs mouvements comme étant très lents par rapport à celui des électrons. Les électrons se déplacent alors dans un champ de noyaux fixes. Dans ce cadre, l'énergie cinétique des noyaux peut être supposée constante et nulle et la répulsion entre les différentes paires de noyaux considérées également comme constante.

Nous pouvons donc définir un nouveau Hamiltonien, c'est celui des électrons He donné par

$$\mathbf{H}_e = \mathbf{T}_e + \mathbf{V}_{ee} + \mathbf{V}_{eN}$$

L'équation de Schrödinger s'écrit alors :

$$\left[- \frac{\hbar^2}{2m_e} \sum_i \nabla_i^2 + \frac{1}{2} \sum_{i,j} \frac{e^2}{|r_i - r_j|} - \sum_{i,\alpha} \frac{e^2 Z_\alpha}{|r_i - R_\alpha|} \right] \psi_e = E_e \psi_e$$

À ce niveau, il existe plusieurs méthodes de résolution de l'équation (II-5) dont les premières sont celles de Hartree [20] et Hartree-Fock [21] basées sur l'hypothèse des électrons libres (individuels) interagissant avec les noyaux et tous les autres électrons du système. Ces méthodes sont plus utilisées en chimie quantique pour traiter les atomes et les molécules, mais pour les solides, elles sont moins précises. Cependant il existe une méthode moderne et certainement plus puissante qui est la théorie de la fonctionnelle de la densité (DFT)

II.1.3.1.3 La théorie de la fonctionnelle de la densité (DFT) :

Les bases de la théorie de la fonctionnelle de la densité (DFT) ont été élaborées en 1927 par Thomas et Fermi qui calculèrent l'énergie d'un atome en représentant son énergie cinétique en fonction de la densité électronique [22]. En 1928, Dirac introduit le terme d'échange prédit par Hartree mais il n'y a toujours aucune prise en compte de la corrélation électronique qui fût finalement ajoutée par Wigner. Dans ce modèle, les n électrons dépendants de $3n$ coordonnées d'espace sont remplacés par leur densité $\rho(r)$ qui ne dépend plus que de 3 variables.

L'état fondamental de ce système est décrit par la fonction d'onde $\Psi_0(r_1, r_2, \dots, r_n)$ qui correspond à une unique densité électronique $\rho(r)$. Cette fonction d'onde, et l'énergie E_0 qui lui est associée, sont déterminées par la minimisation de l'énergie totale du système. Le potentiel externe $V_{\text{ext}}(r)$ créé par les N noyaux du système est alors complètement déterminé et fixe donc l'hamiltonien. Ainsi, le nombre d'électron n et le potentiel $V_{\text{ext}}(r)$, définissent toutes les propriétés de l'état fondamental [23].

La DFT est capable de déterminer avec précision les propriétés moléculaires géométriques, les énergies de liaisons et différents types de spectre pour des molécules aussi complexes que des composés de coordination [24]. De plus, les surfaces d'énergie potentielle des réactions chimiques prédites par les fonctionnelles corrigées par le gradient et/ou les fonctionnelles hybrides sont aussi précises que celles prédites par les niveaux de théories ab initio les plus sophistiqués [25].

II.1.3.1.4 Les bases d'orbitales :

Le choix de la base d'orbitales est aussi un critère très important pour une bonne estimation des données, bien qu'il soit moins important pour la DFT que pour les autres méthodes ab initio, en particulier post Hartree Fock, la base représente la description des orbitales atomiques.

- ✓ **La base 6-31 G*** : décrit une orbitale atomique par la combinaison de six fonctions gaussiennes par orbitale de cœur, de trois autres pour la description des électrons de

valence et d'une dernière pour description des électrons de valence les plus éloignés du noyau (externe). L'astérisque signifie l'utilisation d'orbitales de polarisation pour les atomes lourds c'est à dire d pour les orbitales p, f pour les orbitales d, etc.

- ✓ **La base 3-21G** : base telles que chaque orbitale atomique interne est représentée par une gaussienne contractée qui est elle-même une combinaison linéaire de 3 gaussiennes primitives. Alors que chaque orbitale de valence est représentée par deux fonctions gaussiennes contractées l'une est combinaison linéaire de 2 primitives et l'autre 1 gaussienne simple.
- ✓ **La base 6 -311G** : représente une base du type triple zêta. Dans cette base l'orbitale interne est représentée par une seule STO combinaison linéaire de 6 GTO et les orbitales de valences sont représentées par 3 STO

II.1.3.1.5 Méthodes ab-initio :

Les méthodes ab initio sont des méthodes non empiriques, toutes les intégrales sont rigoureusement calculées et il n'y a pas d'approximation à faire sauf celle de Born Oppenheimer et l'approximation OM-CLOA [26]. Dans les méthodes ab-initio, toutes les particules (noyau et électrons) sont traitées explicitement. On n'utilise aucun paramètre empirique dans le calcul de l'énergie.

Les méthodes ab initio se divisent en deux sous familles : les méthodes Hartree - Fock (HF, RHF, UHF, ROHF) (Hartree, 1928 ; Fock, 1930), et les méthodes post Hartree-Fock, (MPn, CAS, ...) (Moller, 1934) [27].

La méthode de Hartree-Fock [2] permet une résolution approchée de l'équation de Schrödinger d'un système quantique à n électrons et N noyaux dans laquelle la fonction d'onde poly-électronique Ψ_{HF} est écrite sous la forme d'un déterminant de Slater composé de spin-orbitales mono-électroniques qui respecte l'antisymétrie de la fonction d'onde :

$$\Psi(1,2,\dots,N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1\alpha(1) & \phi_1\beta(1) & \dots & \phi_{N/2}\alpha(1) & \phi_{N/2}\beta(1) \\ \phi_1\alpha(2) & \phi_1\beta(2) & \dots & \phi_{N/2}\alpha(2) & \phi_{N/2}\beta(2) \\ & & \dots & & \\ & & & & \\ \phi_1\alpha(n) & \phi_1\beta(n) & \dots & \phi_{N/2}\alpha(n) & \phi_{N/2}\beta(n) \end{vmatrix}$$

II.1.3.1.6 Les méthodes semi-empiriques :

Ces méthodes utilisent une approximation de l'énergie d'interaction sous la forme d'une combinaison linéaire de termes empiriques et quantiques. Il s'agit le plus souvent d'une approximation correcte pourtant toutes les forces impliquées dans la formation du complexe

ne sont pas additives. Les calculs semi-empiriques sont eux développés sur la même structure générale que les calculs HF, mais certaines parties de l'information sont sujettes à approximation ou même complètement omises, afin de les rendre moins exigeants en termes de temps de calcul [28].

Les méthodes semi empiriques sont généralement utilisées pour traiter de grosses molécules, afin d'obtenir une première description de ses propriétés moléculaires (structure, réactivité). On distingue plusieurs variantes : **CNDO**, **INDO**, **MINDO/3**, **MNDO**, **AM 1**, **PM 3**, **SAM 1** [29-30].

- **AM 1** : (Austin Model 1) Proposée par Dewar en 1985. Il a tenté de corriger les défauts de MNDO.
- **PM 3** : (Parametric Method 3) Proposée par Stewart en 1989. Présente beaucoup de points en commun avec AM1, D'ailleurs il existe toujours un débat concernant les mérites relatifs de paramétrisation de chacune d'elles.

II.1.3.2 La mécanique moléculaire (MM) :

La mécanique moléculaire est apparue en 1930 [31], mais s'est développée à partir des années 1960, avec les progrès d'accessibilité et de performance des ordinateurs.

La mécanique Moléculaire (MM) ne prend en compte la structure électronique qu'implicitement et les atomes sont alors représentés par des sphères ayant des masses et des rayons différents et possédant une charge. Les liaisons sont décrites comme des ressorts de différentes constantes de force. Ces charges et constantes de force peuvent être obtenues par divers moyens (données cristallographiques ou spectroscopie vibrationnelle, calcul ab initio) et sont contenues dans le champ de force. Ce dernier représente l'ensemble des paramètres permettant de calculer l'énergie potentielle du système. La Mécanique Moléculaire permet le calcul de l'énergie d'interaction d'un système en fonction des seules positions des noyaux, en ignorant ainsi le mouvement des électrons. L'approximation de Born-Oppenheimer est aussi utilisée, cependant, le mouvement des noyaux n'est plus décrit par un Hamiltonien quantique comme en Mécanique Quantique, mais par les descriptions de la mécanique classique ou les atomes sont assimilés à des points massiques (éventuellement chargés) et les liaisons chimiques à des ressorts mécaniques [32].

II.1.3.2.1 Fonction d'énergie potentielle :

La fonction énergie potentielle est représentée comme une hyper-surface à $3N-6$ dimensions. Elle est définie empiriquement comme étant la somme des différentes contributions traduisant les interactions physiques inter et intramoléculaires du système étudié. La fonction d'énergie

est généralement décrite en termes internes reliés directement aux liaisons covalentes des atomes (interactions liantes) et en termes externes traduisant les interactions entre atomes non liés de manière covalente (interactions non liantes). L'ensemble de ces termes et ces paramètres sont utilisés pour décrire chaque type d'atome rencontré qui se traduit par un champ de forces qui comporte généralement cinq contributions principales [33]

$$E_{\text{atomes-liés}} = E_{\text{liaisons}} + E_{\text{angles}} + E_{\text{diédres}} \quad (13)$$

$$E_{\text{MM}} = E_{\text{liaisons}} + E_{\text{angles}} + E_{\text{torsion}} + E_{\text{vdW}} + E_{\text{électrostatique}} + (E_{\text{croisés}}) \quad (14)$$

$$E_{\text{liés}} \qquad \qquad \qquad E_{\text{non-lié}}$$

II.1.3.2.2 Les différentes énergies :

II.1.3.2.2.1 Energies des interactions liées :

a) Energie d'élongation

L'expression de l'énergie de liaison (élongation) est basée sur la loi de Hooke (par premier approximation) et mesure la quantité d'énergie mise en jeu par les variations de longueur des liaisons après extension ou compression de ressort, à partir de leurs valeurs d'équilibre [1].

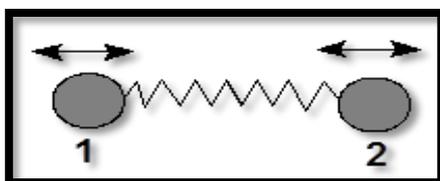


Figure II.1 : Energie de déformation des liaisons.

Ainsi, l'énergie d'élongation est calculée en considérant l'ensemble des liaisons de la structure moléculaire selon la formule suivante :

$$E(L) = \sum_n \frac{1}{2} k_l (L - L_0)^2$$

K_i : est la constante.

L_0 : la longueur de la liaison de référence (à l'équilibre).

L : la longueur de la liaison dans le modèle.

n : nombre de liaisons dans la molécule.

b) Energie de rotation (flexion)

L'expression de l'énergie de flexion détermine la quantité d'énergie impliquée par les variations d'angle entre deux liaisons contigües à partir d'un angle de liaison à l'équilibre [1].

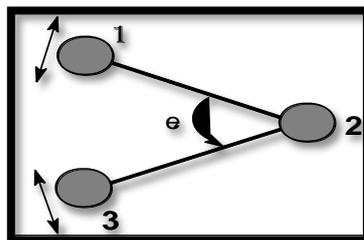


Figure II.2 : Energie de rotation.

L'énergie de flexion est calculée en considérant l'ensemble des angles de la structure moléculaire selon la formule suivante :

$$E_{\theta} = \frac{1}{2} \sum_n k_{\theta} (\theta - \theta_0)^2$$

θ_0 : désigne la valeur de l'angle à l'équilibre entre 3 atomes (en °).

θ : Angle de valence dans la molécule

k_{θ} la constante de force de flexion associée (en kcal .mol⁻¹, θ désigne la valeur d'angle de liaison entre 3 atomes.

n : est un nombre d'angles dans la molécule.

c) Energie de torsion

Elle concerne l'angle dièdre formé par les atomes 1-2-3-4. Elle rend notamment compte de la structure 3D de la molécule [1].

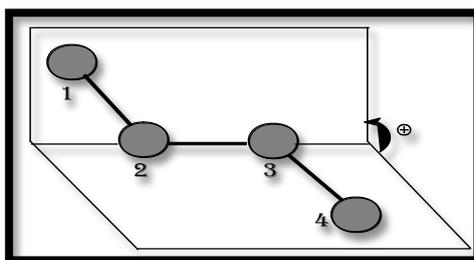


Figure II.3 : Energie de déformation des liaisons.

L'expression mathématique la plus commune permettant de décrire comment l'énergie d'une molécule évolue en fonction de l'angle de torsion est la série de Fourier dans AMBER (Assisted Molecular Building and Energy Refinement), le terme de torsion est défini de la manière suivante

$$E_{tors} = \sum_{n \text{ dièdres}} \frac{V_n}{2} (1 + \cos(n\phi - \gamma)) = \sum_{n \text{ dièdres}} C_n (\cos(\phi))^n$$

V_n et C_n : sont des constantes paramétriques (sont les valeurs de la barrière de rotation d'ordre n (en kcal.mol⁻¹).

ϕ : Représente l'angle de torsion autour de la liaison 2-3.

γ : est le facteur de phase qui représente l'angle où l'énergie de torsion passe par un minimum (en °).

II.1.3.2.2 Energies des interactions intermoléculaires - "non - liées" :

a) Energie de Van der Waals:

La fonction d'énergie de van der Waals, $E_{van_der_Waals}$, traduit l'attraction et la répulsion entre deux atomes constituant un dipôle. Elle s'exprime sous la forme d'un potentiel de Lennard Jones, constitué d'un terme d'énergie attractive en $1/r_{ij}^6$ et d'un terme d'énergie répulsive en $1/r_{ij}^{12}$.

Les interactions de van der Waals ont donc lieu entre atomes relativement proches. Lorsque la distance qui sépare les deux atomes est inférieure à la somme des rayons de van der Waals, c'est le terme répulsif qui prédomine et inversement pour une distance supérieure, c'est le terme attractif qui agit principalement :

$$E_{van-der-waals} = \sum_{i,j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right]$$

A_{ij} , B_{ij} : paramètres dépendant de la nature des atomes, i et j distants de r_{ij} .

b) Energie électrostatique :

L'énergie des interactions électrostatiques, s'écrit sous la forme d'un potentiel coulombien entre les atomes considérés comme porteurs d'une charge nette.

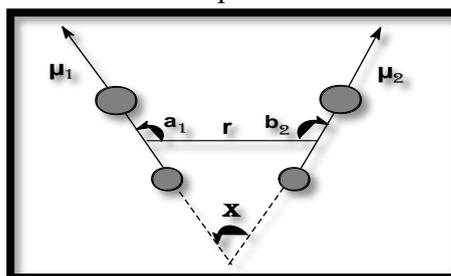


Figure II.4: Energie de déformation des liaisons.

r : distance entre les milieux des deux liaisons

μ_1, μ_2 : représentent respectivement les moments dipolaires des deux liaisons.

X : l'angle entre les deux vecteurs moments.

a_1, b_2 : angles formés respectivement entre μ_1 et r et μ_2 et r .

$$E_{\text{électrostatique}} = \sum_{i,j} \frac{q_i q_j}{\epsilon_{ij} r_{ij}}$$

q_i, q_j : charges partielles des atomes i et j , distants de r_{ij} .

ϵ_{ij} : constante diélectrique rendant compte de l'atténuation de l'interaction électrostatique par l'environnement (solvant ou molécule elle-même).

c) Energie des liaisons hydrogène

La liaison hydrogène est une interaction d'importance intermédiaire (8 à 20 KJ/mol) entre un hydrogène déficient en électrons et un atome de forte densité électronique portant un doublet d'électrons libres. Le modèle électrostatique, vrai lorsque la distance A-B est grande, n'est pas suffisant pour décrire ces interactions particulières. A plus courte distance, les phénomènes de répulsion et de délocalisation électronique interviennent. Plusieurs types de fonctions d'énergie potentielle ont été développés pour tenir compte de la directivité de la liaison hydrogène. Actuellement, les fonctions les plus utilisées permettant d'exprimer ces interactions dans des systèmes moléculaires importants sont souvent simplifiées :

La fonction EH = $A/r_{ij}^{12} - B/r_{ij}^{10}$

La fonction EH = $A'/r_{ij}^{12} - B'/r_{ij}^6$

Les coefficients A, B, A', B' sont spécifiques des liaisons hydrogène [13].

r_{ij} : distance séparant entre deux atomes i et j .

II.1.3.2.3 Champ de force :

Le champ de force est une expression que la mécanique moléculaire a empruntée à la spectroscopie vibrationnelle, en modifiant légèrement sa signification [34]. Sous ce terme sont en fait regroupés deux éléments : d'une part l'expression des différentes fonctions contribuant au calcul énergétique et d'autre part les valeurs des différentes constantes paramétrant ces fonctions. Ces paramètres sont identifiés à partir de données expérimentales (IR, micro-onde...) ou évalués théoriquement.

L'énergie totale du système est donc une énergie potentielle multidimensionnelle décrivant les interactions intramoléculaires ou interactions liantes (impliquant des atomes reliés par des liaisons explicites) et les interactions intermoléculaires ou interactions non liantes (impliquant des atomes non liés directement par des liaisons explicites) :

$$E_{\text{Totale}} = E_{\text{Intermoléculaire}} + E_{\text{Intramoléculaire}}$$

II.1.3.2.3.1 Différents champs de force en mécanique moléculaire :

Différents champs de force utilisent le même type de termes énergétiques mais paramétrés de manières différentes [35]. Chacun a un domaine d'application spécifique de sorte que le choix d'un champ de force dépend des propriétés et de l'application du système que l'on veut étudier.

Type de composé : carbohydrate, complexe métallique.

Environnement : gaz, solution.

- Type d'interaction à étudier : liaison hydrogène,
- Les champs de forces en MM peuvent être groupés en trois classes principales [36]:
- Champs de force contenant uniquement les termes harmoniques.
- Champs de force utilisant les termes d'ordre supérieur (cubique, quadratique, ...).
- Champs de force suggérés par Allinger et col [37] ne considérant pas que les termes de la mécanique moléculaire classique mais aussi les effets chimiques comme l'électronégativité.
- Dans ce qui suit on va présenter les principaux champs de forces utilisés dans les calculs de la MM.
- MM2/MM3/MM4 : MM2 est le premier champ de force développé par Allinger et col [38-39]. Il a été conçu au début pour les molécules simples (alcane, alcène, alcyne non conjugués, amines...), mais ses versions améliorées MM3 (1989) [40] et MM4 (1996) [41] permettent de traiter des molécules organiques de plus en plus complexes.
- AMBER : Champ de forces pour les protéines et les acides nucléiques
- CHARMM : Champ de forces pour les protéines, les acides nucléiques et les lipides.
- MM3PRO : Version de MM3 dont les paramètres sont adaptés à l'étude des protéines.
- OPLS : Champ de forces pour les protéines ainsi que pour certaines classes de molécules organiques. Ce champ de forces est de type « united-atom » c.-à-d. que les atomes d'hydrogènes sont inclus dans l'atome qui les portent, par exemple un type d'atome représentera un -CH₃.

II.1.3.3 Dynamique Moléculaire :

Les premiers pas de la dynamique moléculaire n'ont pu se faire que grâce à l'arrivée des premiers ordinateurs (1957) [39]. Mais les premières réelles simulations ont été faites par Rahman [40], grâce à ses travaux sur la simulation de l'argon liquide en 1964 avec un temps de simulation de 10-11 s, puis de l'eau liquide [42] en 1971.

La mécanique moléculaire permet de calculer la position des atomes et des molécules dans leurs états énergétiques minima, à 0° K c'est à dire sans la moindre

vibration ni le moindre mouvement.

La dynamique moléculaire tente à simuler le mouvement des atomes et des molécules au cours du temps pour une température supérieure à 0° K. Elle consiste à étudier la trajectoire d'une molécule en appliquant les lois de la mécanique classique Newtonienne et elle permet de simuler les mouvements intramoléculaires que l'on peut visualiser ensuite en temps réel. Ces mouvements correspondent à des vibrations autour d'un minimum, ou au passage d'un minimum à un autre minimum d'énergie.

II.1.3.3.1 Principe de Dynamique Moléculaire :

La dynamique moléculaire utilise la seconde loi de Newton pour décrire le mouvement d'une molécule en fonction du temps.

$$\vec{F}_i = m_i \vec{a}_i = m_i \frac{d^2 \vec{r}_i(t)}{dt^2}$$

\vec{F}_i : Vecteur force agissant sur l'atome i.

m_i : masse de l'atome i.

\vec{a}_i : Vecteur accélération de l'atome i.

\vec{r}_i : La position de l'atome i.

Grâce aux vitesses et aux positions de chaque atome dans le temps, il est possible d'évaluer les données macroscopiques, comme l'énergie cinétique et la température. L'énergie cinétique se calcule selon la formule :

$$E_c = \sum_{i=1}^N \frac{|p_i|^2}{2m_i}$$

Où p_i est la quantité de mouvement de l'atome i.

La température s'obtient à partir de l'énergie cinétique par :

$$E_c = \frac{k_b T}{2} (3N - N_c)$$

k_b : constante de Boltzmann

N_c : nombre de contrainte

$3N - N_c$: nombre totale de degré de liberté.

La force F_i qui s'exerce sur un atome i se trouvant en position $r_i(t)$ est déterminée par dérivation de la fonction potentielle :

$$\vec{F}_i = -\frac{d\vec{E}(r_1, \dots, r_n)}{dr_i(t)}$$

E : fonction de l'énergie potentielle totale d'interaction.

r_i : coordonnées cartésiennes de l'atome i.

Les vitesses de chaque atome sont calculées à partir de la connaissance des accélérations atomiques :

$$\vec{a}_i = \frac{d\vec{v}_i}{dt}$$

Et les positions des atomes sont calculées à partir des vitesses atomiques par la relation :

$$\vec{v}_i = \frac{d\vec{r}_i}{dt}$$

L'intégration de ces équations se fait en subdivisant la trajectoire en une série d'états séparés par des intervalles de temps très courts dont la longueur définit le pas d'intégration t , ce qui conduit à une trajectoire en fonction du temps. Connaissant la vitesse et l'accélération de l'atome i à l'instant t , on peut connaître sa position à l'instant $t + \Delta t$:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i \Delta t + \frac{1}{2} \mathbf{a}_i \Delta t^2$$

II.1.3.3.2 Mise en oeuvre d'un calcul de DM :

Avant de commencer une simulation de DM, il est important d'avoir un modèle de départ proche de la structure réelle.

Minimisation par mécanique moléculaire : on part, en général d'une conformation optimisée par la MM.

Dynamique moléculaire : en général, un calcul de DM se décompose en trois phases :

- ✓ Thermalisation : il s'agit de chauffer le système pour l'amener à la température souhaitée (généralement 300 K), cette montée en température demande 2 à 10 ps de simulation (2000 à 10000 pas d'intégration).
- ✓ Equilibration : est une phase importante au cours de laquelle, on stabilise la T du système. Il y a alors un échange important entre l'énergie potentielle et l'énergie cinétique. Cette

étape peut durer entre 10 et 20 ps. La température est calculée après chaque étape et ramenée à la T de référence par la formule :

$$\frac{dT}{dt} = \tau^{-1} (T_0 - T)$$

Avec :

τ : Temps de relaxation, la valeur de τ doit être assez petite pour permettre un couplage efficace (généralement égale à 10 fs).

T_0 : température de référence.

T : température du système.

- ✓ Dynamique ou production : c'est la phase réellement exploitable qui dure en moyenne de 60 à 100 ps. Au cours de cette simulation, les conformations sauvegardées toutes les 0.05 ou 0.1ps.

II.1.3.3 Méthodes empirique

Les méthodes empiriques essaient d'estimer l'énergie d'interaction en réunissant de manière intuitive les termes que l'expérience a définis comme fondamentaux. Classiquement ces termes regroupent les liaisons hydrogène, les interactions ioniques, la surface de contact ligand -récepteur et les contributions entropiques. Ces fonctions doivent autant que possible équilibrer l'influence des deux forces principales que sont les interactions ioniques et hydrophobes pour pouvoir s'appliquer au plus grand nombre.

II.1.4 Domaine d'application de la modélisation moléculaire

Il existe un nombre important des travaux qui utilisent la modélisation moléculaire comme technique d'analyse des argiles minérales. Deux types d'études peuvent être différenciés :

- 1- l'étude de la structure des argiles et leurs interaction avec des contreions et leurs hydratations [43-44].
- 2- l'étude des systèmes composites à base d'argile et des polymères [45-46].

D'autre part, quelques études concernent la modélisation de l'adsorption des composés organiques, d'aminoacides, de protéines ou des substances humiques sur diverses surfaces minérales comme la beidellite, la muscovite, la hydrotalcite, la pyrophyllite ou la montmorillonite.

Néanmoins, aucune étude par modélisation moléculaire n'a été reportée précédemment pour l'étude de l'adsorption des exopolysaccharides bactériens sur la montmorillonite sodique [47-48].

On peut diviser l'application de la MM en trois catégories :

- ✓ Soit pour obtenir une géométrie à laquelle on attache de l'intérêt. Cette situation se présente lorsque la modélisation guide l'interprétation des résultats provenant des études de structure par rayons X ou par diffraction électronique, ou lorsqu'il s'agit de modéliser une molécule pour les besoins de l'infographie.
- ✓ Dans l'interprétation des effets stériques sur la réactivité ou bien de la stabilité relative des isomères en tant qu'énergie stérique ou de tension.
- ✓ Quand aucune liaison n'est rompue, ni formée et qu'aucun intermédiaire chargé n'intervient, l'inter conversion conformationnelle se prête particulièrement bien à une description par la MM. On peut obtenir grâce à cette analyse des informations structurales sous forme d'un profil énergétique (en fonction d'un angle dièdre par exemple) ou des cartes énergétiques 3D.
- ✓ En conclusion, on peut dire que la mécanique moléculaire aujourd'hui est à la porte de tous les chercheurs.

II.2.1 Introduction

Le développement de nouvelles techniques de modélisation a permis la mise en place de nombreuses méthodes RQSP (en anglais QSPR : Quantitative Structure Property Relationships) et RQSA (en anglais QSAR : Quantitative Structure-Activity Relationships) ; elles reposent pour la plupart sur « la recherche d'une relation entre un ensemble de nombres réels, appelés descripteurs moléculaires, et la propriété ou l'activité que l'on souhaite prédire ». Ces méthodes permettent de justifier les données expérimentales disponibles et de prédire les propriétés/activités pour des ~~nouveaux composés~~ ou des composés pour lesquels les données expérimentales ne sont pas disponibles. Dans ce chapitre, une étude bibliographique sur les différentes méthodologies RQSA/RQSP a été présentée, les différentes étapes de développement, de validation et d'application de ces méthodes sont aussi mises en œuvre [49].

II.2.2 Définition

~~Une relation quantitative structure à activité (en anglais : Quantitative structureactivity relationship ou QSAR, parfois désignée sous le nom de relation quantitative structure à propriété en anglais : quantitative structure property relationship ou QSPR) est le procédé par lequel une structure chimique est corrélée avec un effet bien déterminé comme l'activité biologique ou la réactivité chimique [50]~~

~~Certain projet de recherche mettent en jeu des données biologiques et physicochimiques appropriées. Ces données peuvent être représentées et analysées de diverses manières. Le groupement et la classification des composés, basés sur leurs propriétés, sont les éléments principaux lors d'études de similarité moléculaire. Les études de régression ou de corrélation entre les données biologiques et chimiques sont généralement utiles pour rationaliser les relations structure-activité. Deux types principaux d'études peuvent être définis:~~

~~(1) par corrélation entre les données biologiques et physicochimiques en utilisant les techniques de régression.~~

~~(2) par classification des composés ou méthodes d'identification de modèle.~~

Les équations QSAR prennent la forme générale suivante :

$$\text{Activité biologique} = \text{fonction} \{ \text{paramètre(s)} \}$$

Dans l'équation, l'activité biologique est exprimée comme $\text{Log} [1/C]$ où C est généralement la concentration minimum requise pour causer une réponse biologique définie. La relation mathématique entre un paramètre ou plusieurs paramètres physico-chimiques et l'activité biologique d'un composé peut être exprimée à travers l'équation de Hansch [51]:

$$\text{Log } 1/C = -K_1 (\log P)^2 + K_2 \log P + K_3 \sigma + K_4 E_s + K_5$$

Cette équation est obtenue par analyse de régression. La valeur du coefficient r obtenue après cette analyse donne des informations sur la concordance des paramètres qui sont utilisés pour former l'équation de Hansch. Si le résultat obtenu pour r est $> 0,9$ cela signifie que les paramètres utilisés sont valides et que l'équation peut être exploitée pour le reste des composés.

L'équation de Hansch est utilisée pour prévoir l'activité biologique de composés dont la structure est semblable à ceux utilisés pour former cette équation. Cependant, dans le cas où la valeur prévue est différente de celle déterminée par l'expérimentation, cela peut signifier que l'activité biologique réelle des composés n'est pas prise en compte par les paramètres mis en jeu dans le processus de construction de l'équation. L'équation de Hansch peut être également utilisée pour montrer le degré d'importance de différents paramètres sur l'activité biologique [52].

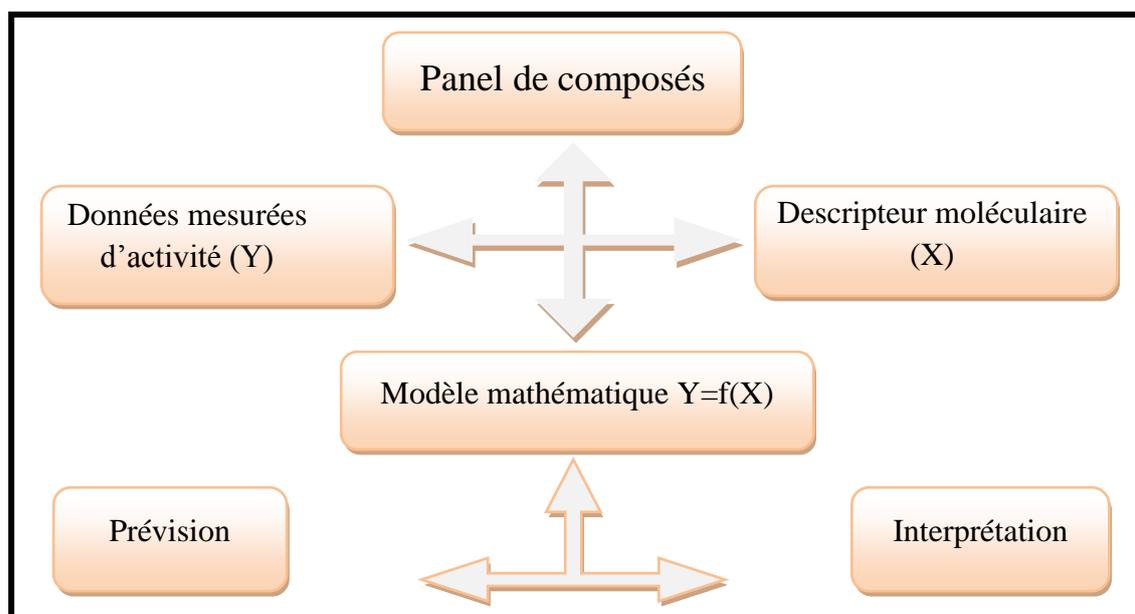


Figure II.5: modèle d'étude quantitative des relations structure / activité (QSAR)

II.2.4 Les applications de l'étude QSAR

Certaines études QSAR semblent être un peu plus d'études universitaires, il existe un grand nombre d'applications de ces modèles tels que [53]:

- ✓ L'optimisation de l'activité pharmacologique.
- ✓ La conception rationnelle de nombreux autres produits tels que des agents tensioactifs, des parfums, des colorants et des produits chimique fins.
- ✓ L'identification des composés dangereux dans les premiers stades de développement de produits ou la projection des stocks de composés existants.

- ✓ La prédiction de la toxicité et les effets secondaires de nouveaux composés.
- ✓ La prédiction de la toxicité pour les espèces environnementales.
- ✓ La sélection des composés ayant des propriétés pharmacocinétiques optimales, que ce soit la stabilité ou la disponibilité dans les systèmes biologiques.
- ✓ La prédiction d'une variété de propriétés physico-chimique des molécules.
- ✓ La prédiction du devenir des molécules qui sont libérées dans l'environnement.
- ✓ La prédiction des effets conjugués de molécules, que ce soit dans des mélanges ou des formulations.

II.2.5 Les modèles statistiques

L'élaboration de modèle QSAR n'est pas une chose facile. On a deux problèmes importants

1- la différence d'échelles existant entre les données à corrélérer.

2- le traitement de données.

Toutes les caractéristiques d'une molécule ne sont pas liées à son activité biologique.

Les descripteurs chimiques sont souvent corrélés entre eux et donnent la même information.

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

- ✓ La régression linéaire multivariée (Multivariate Linear Regression - MLR).
- ✓ La régression en composante principale (Principale Component Regression - PCR).
- ✓ La régression des moindres carrés partiels (Partial Least Squares - PLS).
- ✓ Les réseaux de neurones artificiels (Artificial Neural Network - ANN).

Le choix de la méthode dépend principalement de la question qui est posée et de la nature des données à traiter.

II.2.5.1 La régression multilinéaire RML : [53]

Le problème de la régression linéaire est de déterminer les paramètres d'un modèle reliant Y à X, à partir de n mesures $\{(Y_1, X_1), (Y_2, X_2) \dots (Y_n, X_n)\}$. Le cas simple du modèle du premier ordre avec une seule variable s'écrit :

$$\hat{Y} = \beta_0 + \beta_1 X \quad (1)$$

Le calcul des paramètres $[\beta_0 \beta_1]$ est réalisé en minimisant l'erreur commise sur leur estimation.

La méthode des moindres carrés consiste à chercher les valeurs des paramètres β_0 et β_1 qui rendent minimale la somme des carrés PRESS des écarts résiduels entre les valeurs observées Y_i et les valeurs calculées de \hat{Y}_i :

$$\text{PRESS} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Avec : PRESS est la somme des carrés des résidus

En remplaçant \hat{Y}_i :

$$\text{PRESS} = \sum_{i=1}^n (\beta_0 + \beta_1 X_i - Y_i)^2$$

Cette relation fait apparaître la somme des carrés des écarts comme une fonction des paramètres β_0 et β_1 . Lorsque cette fonction est minimale, les dérivées par rapport à ces paramètres s'annulent. L'annulation des dérivées de PRESS conduit aux équations suivantes, dites les équations normales :

$$\begin{cases} \beta_0 n + \beta_1 \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i \\ \beta_0 \sum_{i=1}^n X_i + \beta_1 \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i \end{cases} \quad (2)$$

Avec : β_0 et β_1 sont alors déterminés au sens des moindres carrés.

II.2.5.2 Formulation matricielle de la régression linéaire : [53]

Le système des équations (2) s'écrit sous forme matricielle (Σ sur i de 1 à n) :

$$\begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} \quad (3)$$

A B = C

En introduisant les matrices X et le vecteur Y tels que :

$$\mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & X_n \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}$$

A et C s'exprime alors en fonction de X et Y :

$$\mathbf{A} = \mathbf{X}^t \mathbf{X} \quad \mathbf{C} = \mathbf{X}^t \mathbf{Y}$$

Le système (3) s'écrit alors en fonction de l'échantillon $\{X \ Y\}$:

$$\mathbf{X}^t \mathbf{X} \mathbf{B} = \mathbf{X}^t \mathbf{Y}$$

L'estimateur des moindres carrés de B s'exprime alors :

$$\hat{\mathbf{B}} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}$$

II.2.5.3 La signification du modèle

Après avoir déterminé le modèle QSAR, la question qui se pose est : La liaison globale entre Y et X_1, \dots, X_k est-elle significative ?

Modèle : $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$

Test : $H_0 : \beta_1 = \dots = \beta_k = 0$ (Y = β_0 ne dépend pas des X)

H_1 : Au moins un $\beta_k \neq 0$ (Y dépend d'au moins un X)

Donc on doit rejeter l'hypothèse de H_0 par le teste de signification où le paramètre sig est la probabilité conditionnelle qu'une relation aussi forte que celle observée dans les données puisse exister si l'hypothèse nulle était vraie. Aussi appelée valeur p, en règle générale une valeur inférieure à 0.05 est considérée comme significative.

II.2.5.4 Le choix du meilleur modèle : [53]

Pour choisir le meilleur modèle on doit calculer quelque paramètre tel que F, SE, r et Q.

Ces paramètres sont calculés à partir des équations mathématiques où:

F : est le rapport de deux carrés de moyennes, quand la valeur de F est importante l'hypothèse nulle peut être rejetée, d'autres termes F est un test statistique utilisé pour tester l'hypothèse nulle que l'introduction d'une variable supplémentaire n'entraîne pas un accroissement significatif du R^2 .

SE : (standard error of the estimate) c'est une mesure de variation de la valeur d'un statistique test, d'un échantillon à l'autre. Il s'agit de l'écart-type de la distribution de l'échantillon pour une statistique.

R : c'est un coefficient de corrélation entre les valeurs observées et les prévisions de la variable dépendante, sa valeur variée entre 0 et 1.

Q : facteur de qualité d'ajustement suggère la puissance de prédiction.

II.2.6 Validation du modèle : [53]

II.2.6.1 Validation interne

Quel que soit le but ultime du modèle QSAR, il doit être obligatoirement validé avant d'être interprété ou utilisé à fins prédictives. Il existe différentes solutions pour s'assurer de validité d'un modèle. La validation croisée LOO (leave-one-out) est un processus qui permet de tester la précision prédictive d'un modèle, cette méthode basée sur le calcul de quelques paramètres statistiques tels que : **PRESS**, **SSY**, **Spress**, **R^2_{cv}** , **R^2_{adj}** et **PE** ; ces paramètres sont calculés à partir des équations suivantes :

$$\text{PRESS} = \sum (Y_{\text{obs}} - Y_{\text{calc}})^2$$

$$\text{SSY} = \sum (Y_{\text{obs}} - Y_{\text{mean}})^2$$

$$\text{SPRESS} = \sqrt{\frac{\text{PRESS}}{n}}$$

$$r^2_{\text{CV}} = 1 - \frac{\text{PRESS}}{\text{SSY}}$$

$$r^2_{\text{adj}} = ([1 - (r^2)] \left(\frac{n-1}{n-p-1} \right))$$

$$\text{PE} = 0.6745 (1 - r^2) / \sqrt{n}$$

Avec : **n** : nombre de molécules de la série étudiée.

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

SSY : somme des carrés des distances des valeurs observées pour une variable par rapport à la moyenne de cette variable, la somme des carrés permet de mesurer la variation totale dans une variable.

Spres: la capacité prédictive des modèles est évaluée par l'erreur de racine-carrée-moyenne.

r²_{adj} : Estimation de la qualité d'ajustement du modèle à la population.

r²_{cv} : est la variation de la statistique du R² obtenue en ajoutant ou en enlevant une variable indépendante.

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

II.2.6.2 Validation externe

Cette méthode consiste à prédire la propriété/activité d'une série de molécules appelée généralement série de test qui ne sont pas dans la série de développement du modèle, cette validation est caractérisée par le paramètre R² test. Récemment plusieurs études [55,56] ont montré l'insuffisance des paramètres R², R²_{CV} pour vérifier le pouvoir prédictif des modèles RQSA/RQSP. Par conséquent, d'autres paramètres doivent être vérifiés pour cet objectif. Ces paramètres sont connus sous le nom «critères de validation externe» ou souvent appelés «critères de Trophsa» [55].

Références bibliographique

- [1]: BELAIDI, SALAH. Thèse de doctorat, Université de Batna, 2002
- [2]: RIVAIL, Jean-Louis. Éléments de chimie quantique à l'usage des chimistes ,CNRS Edition , PARIS , 1999.
- [3]: MCQUARRIE, Donald A. et SIMON, John D. Chimie physique: approche moléculaire: cours et exercices corrigés. Dunod, 2000.
- [4]: FAYET, Guillaume. Développement de modèles QSPR pour la prédiction des propriétés d'explosibilité des composés nitroaromatiques. 2010. Thèse de doctorat. Chimie ParisTech.
- [5]: LIOTTA, Dennis (ed.). Advances in molecular modeling. Jai Pr, Greenwich, 1988.
- [6]: TSAI, C. Stan. An introduction to computational biochemistry. Wiley-Liss, New York, 2002
- [7]: HARKATI, Dalel. Etude de la structure et des propriétés physico-chimiques associées, de quelques molécules bioactives à intérêt pharmaceutique. 2015. Thèse de doctorat. Université Mohamed Khider-Biskra.
- [8]: H.-D. Höltje, G. Folkeis, Molecular Modeling: Basic Principles and Applications. VCH, New-York, 1997.
- [9]: LERBET-SERENI, Frédérique. Expériences de la modélisation, modélisation de l'expérience. Editions L'Harmattan, 2004.
- [10]: BELAIDI, S. OMARI, M , LANEZ, T et A . Dibi, J. Soc.Alger. Chim., 2004, 14, 27.
- [11]: TSAI, C. Stan, An Introduction to Computational Biochemistry, Wiley-Liss, New York, 2003
- [12]: MOSTEFAOUI, Larbi. Contribution à la description et à la compréhension de la solvation des biomolécules. 2011.
- [13] : DUGAS, Hermann. Principes de base en modélisation moléculaire. Aspects théoriques et pratiques, quatrième édition, Librairie de l'Université de Montréal, 1996.
- [14]: BOYD, Donald B. et LIPKOWITZ, Kenny B. Molecular mechanics: The method and its underlying philosophy. Journal of Chemical Education, 1982, vol. 59, no 4, p. 269.

- [15]: Cancès, E. C. Le Bris. Y. Maday, Méthodes mathématiques en chimie quantique. Une introduction, Springer Science & Business, 2006.
- [16]: Hladik; J. Chrysos; M. P.-E. Hladik. L.U. Ancarani, Mécanique quantique, Masson, 1997.
- [17]: HASSAINE, Roumaïssa. Etude QSAR basée sur la DFT conceptuelle des propriétés cytotoxiques de certains dérivés de chalcone .2016. Mémoire de fin d'étude en Master
- [18]: BOUNEHAS, RiIMA. APPROCHE QUALITATIVE DE LA RELATION STRUCTURE ACTIVITE DANS DES DERIVES BIOACTIVES DE 1,2,3-TRIAZINE.2014. Mémoire de fin d'étude en Master
- [19]: BORN, Max et OPPENHEIMER, Robert. Zur quantentheorie der molekeln. Annalen der physik, 1927, vol. 389, no 20, p. 457-484.
- [20]: HARTREE, Douglas R. The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods. In : Mathematical Proceedings of the Cambridge Philosophical Society. Cambridge University Press, 1928. p. 89-110.
- [21]: FOCK, V. „Selfconsistent field “mit Austausch für Natrium. Zeitschrift für Physik, 1930, vol. 62, no 11-12, p. 795-805.
- [22]: P. A. M. Dirac, Proc. Roy. Soc. London A117 (1928) 610-624. (b) P. A. M. Dirac, Proc. Roy. Soc. London A118 (1928) 351-361.
- [23]: POPLE, John A., SANTRY, D. Po, et SEGAL, Gerald A. Approximate self-consistent molecular orbital theory. I. Invariant procedures. The Journal of Chemical Physics, 1965, vol. 43, no 10, p. S129-S135.
- [24]: CHERMETTE, H. Density functional theory: a powerful tool for theoretical studies in coordination chemistry. Coordination chemistry reviews, 1998, vol. 178, p. 699-721.
- [25]: VON RAGUÉ SCHLEYER, Paul. Encyclopedia of computational chemistry. Chichester, UK : Wiley, 1998.
- [26]: G. Boucekkine, Méthodes de la chimie quantique. Techn.Ing., AF6050, (2007).
- [27]: PIRON, Constantin. Mécanique quantique: Bases et applications. PPUR presses polytechniques, 1998.

- [28]: MAROT, Christophe. Conception de nouvelles molécules à activité sérotoninergique par des méthodes QSAR et des études de dynamique moléculaire de complexe ligands/récepteur. 1995. Thèse de doctorat. Université d'Orléans.
- [29]: POPLE, J. A., BEVERIDGE, D. L., et DOBOSH, P. A. Approximate Self-Consistent Molecular-Orbital Theory. V. Intermediate Neglect of Differential Overlap. The Journal of Chemical Physics, 1967, vol. 47, no 6, p. 2026-2033.
- [30]: BENHAMED, Karima. CONTRIBUTION A L'ETUDE STRUCTURALE PAR DIFFRACTION DES RX ET PAR MODELISATION MOLECULAIRE DU COMPLEXE DE Ni (II) DERIVE DE CINNAMALDEHYDE THIOSEMICARBAZONE. 2010.
- [31]: ANDREWS, Donald H. The relation between the Raman spectra and the structure of organic molecules. Physical Review, 1930, vol. 36, no 3, p. 544.
- [32]: LEBBAD, Fatima. Etude par modélisation moléculaire des mécanismes de complexation. Thèse de doctorat. Université Aboubeker Belkaid de Tlemcen.
- [33]: M. LEDECQ, Thèse de doctorat, Université de Notre-Dame de la Paix, France, 2004
- [34]: LOMAS, J. La mécanique moléculaire, une méthode non quantique pour le calcul de la structure et de l'énergie d'entités moléculaires. L'actualité chimique, 1986, vol. 5, p. 7-22.
- [35]: HEISENBERG, Werner. The physical principles of the quantum theory. Courier Corporation, 1949.
- [36]: SMITH, Jeremy C. et KARPLUS, Martin. Empirical force field study of geometries and conformational transitions of some organic molecules. Journal of the American Chemical Society, 1992, vol. 114, no 3, p. 801-812.
- [37]: SHIMANOUCI, T. Force constants of small molecules. Pure and Applied Chemistry, 1963, vol. 7, no 1, p. 131-146.
- [38]: Frankie Tristram, Viviane Durier, Gérard Vergoten, Journal of Molecular Structure., 378, 249-2561996.
- [39] : ALDER, B. J. et WAINWRIGHT, Tef. Phase transition for a hard sphere system. The Journal of chemical physics, 1957, vol. 27, no 5, p. 1208-1209.

- [40]: RAHMAN, A. Correlations in the motion of atoms in liquid argon. *Physical Review*, 1964, vol. 136, no 2A, p. A405.
- [41]: STILLINGER, Frank H. et RAHMAN, Aneesur. Improved simulation of liquid water by molecular dynamics. *The Journal of Chemical Physics*, 1974, vol. 60, no 4, p. 1545-1557.
- [42]: Rahman. A, Stillinger. F.H, *J. Chem. Phys.*, 5, 3336, 1971.
- [43]: TEPPEN, Brian J., RASMUSSEN, Kjeld, BERTSCH, Paul M., et al. Molecular dynamics modeling of clay minerals. 1. Gibbsite, kaolinite, pyrophyllite, and beidellite. *The Journal of Physical Chemistry B*, 1997, vol. 101, no 9, p. 1579-1587.
- [44]: ČAPKOVÁ, Pavla, POSPÍŠIL, Miroslav, et WEISS, Zdeněk. Combination of modeling and experiment in structure analysis of intercalated layer silicates. *Journal of molecular modeling*, 2003, vol. 9, no 3, p. 195-205.
- [45]: POSPÍŠIL, M., KALEDOVÁ, Alena, ČAPKOVÁ, Pavla, et al. Structure analysis of intercalated layer silicates: combination of molecular simulations and experiment. *Journal of colloid and interface science*, 2004, vol. 277, no 1, p. 154-161.
- [46]: YU, Ching-Hsing, NEWTON, Susan Q., NORMAN, Mya A., et al. Molecular dynamics simulations of adsorption of organic compounds at the clay mineral/aqueous solution interface. *Structural Chemistry*, 2003, vol. 14, no 2, p. 175-185
- [47]: YU, Ching-Hsing, NEWTON, Susan Q., NORMAN, Mya A., et al. Molecular dynamics simulations of the adsorption of methylene blue at clay mineral surfaces. *Clays and Clay Minerals*, 2000, vol. 48, no 6, p. 665-681.
- [48]: N. L. Allinger, F. Li and L. Yan, *J. Am. Chem. Soc.* , 1990, 117, 868.
- [49]: CHTITA, Samir. Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments. 2017. Thèse de doctorat. Université Moulay Ismaïl, Meknès.
- [50]: David R. Bevan. QSAR and Drug Design. Department of biochemistry and Anaerobic Microbiology. Virginia Polytechnic Institute and State University. Blacksburg, VA 24061-0308 USA, 2010.
- [51]: THOMAS, G. The SAR and QSAR approaches to drug design. *Fundamentals of Medicinal Chemistry*, 2003, p. 71-92.

- [52]: SCHNEIDER, Gisbert, CLÉMENT-CHOMIENNE, Odile, HILFIGER, Laurence, et al. Virtual screening for bioactive molecules by evolutionary de novo design. *Angewandte Chemie International Edition*, 2000, vol. 39, no 22, p. 4130-4133.
- [53]: Ziani Tarek, Prédiction de l'effet inhibiteur des dérivés d'imidazole sur l'enzyme Cox-2, Université Med Khider Biskra, 2012-2013.
- [54]: SARI MOHAMED, Souad. Etude de l'Interaction Protéine-Protéine par Modélisation Moléculaire. Application aux Processus Biologiques. 2013. Thèse de doctorat.
- [55]: PRATIM ROY, Partha, PAUL, Somnath, MITRA, Indrani, et al. On two novel parameters for validation of predictive QSAR models. *Molecules*, 2009, vol. 14, no 5, p. 1660-1701.
- [56]: GOLBRAIKH, Alexander et TROPSHA, Alexander. Beware of q²!. *Journal of molecular graphics and modelling*, 2002, vol. 20, no 4, p. 269-276.

III.1.1 Introduction

La recherche et la synthèse de nouveaux composés chimiques et biochimiques sont aujourd'hui souvent associées à une étude par modélisation moléculaire.

Grace au développement informatique de ces dernières années et à l'essor du calcul parallèle intensif en particulier, la modélisation moléculaire est devenue un véritable enjeu [1].

La modélisation moléculaire peut être définie comme une application de l'informatique pour créer, manipuler, calculer et prédire les structures moléculaires et les propriétés associées. L'informatique pour créer, manipuler, calculer et prédire les structures moléculaires et les propriétés associées [2].

La Relation Quantitative Structure et Activité (QSAR) est un outil permet d'expliquer l'interaction des composés chimiques avec les sujets actuels, le principe des méthodes QSAR est comme leur nom l'indique, de mettre en place une relation mathématique reliant de manière quantitative des propriétés moléculaires, appelées descripteurs, avec une observable macroscopique (activité biologique, toxicité, propriété physico-chimique, etc.), pour une série des molécules biologiquement actives à l'aide des méthodes d'analyse de données .

Les informations extraites à partir des résultats d'étude de QSAR peuvent être utilisées pour obtenir une meilleure connaissance des structures moléculaires et probablement le mode d'action au niveau moléculaire. Ces informations peuvent alors être utilisées pour prévoir les propriétés physicochimiques et les activités biologiques de nouveaux composés ainsi que pour concevoir de nouvelles structures [3].

Dans ce travail, nous intéressons à l'étude des propriétés physico-chimiques d'une série dérivée de bezimidazole qui permettent de prévoir les paramètres physicochimiques qui sont influents sur l'activité biologique, afin de prédire l'activité biologique de nouvelles molécules, c'est pour cette raison on a met le choix d'une série de vingt dérivés de benzimidazole avec des différents descripteurs physico-chimiques.

III.1.2 Les propriétés QSAR caractérisant la structure moléculaire

III.1.2.1 Les paramètres amphiphiliques

Entrent en jeu dans l'activité d'une molécule sont :

Génère un module Hyperchem qui vous permet de calculer un certain nombre de QSAR propriétés physicochimiques des molécules (les descripteurs). Ces calculs généralement somment des contributions atome et sont très rapides [4].

Les propriétés fournies par l'hyperchem sont :

a) **Surface moléculaire₂**: (SAG)

C'est une enveloppe entourant les atomes localisés à la périphérie, qui explique la surface de contact Ligand-Récepteur c'est-à-dire l'affinité de liaison L-R [5-6]. L'HyperChem permet de calculer la surface grille (SAG), c'est une méthode de grille où une méthode plus rapide plus approximative; quelque soit la zone accessible au solvant ou la surface de van der Waals [4]. Dans cette théorie, chaque atome de la molécule est représenté par une sphère. La surface extérieure de toutes les sphères atomiques définit la surface de Van der Waals [7].

b) **Volume moléculaire** : (V)

C'est un volume occupé par un corps, c'est un volume spécifique (à température et pression normales). Son calcul est très similaire à celui de la Surface (Grid), il emploie une méthode de grille décrite par Bodor et al.

Le volume est défini par la relation suivante [8] :

$$V = \frac{W}{d}$$

Où:

W: est la masse moléculaire.

d: est la densité.

c) **Energie d'hydratation** : (HE)

L'énergie d'hydratation est le facteur clé, déterminant la stabilité des différentes conformations moléculaires [9] . Les liaisons hydrogènes constituent un cas particulier d'une interaction dipôle-dipole.

Les groupements donneurs sont F-H, O-H, N-H, Cl-H, P-H et par voie C-H.les groupements accepteurs doivent obligatoirement porter un doublet disponible : F, O, N, mais aussi Cl, Set P et les entités chargés tell que l'anion carboxylate, par exemple. L'énergie de liaison hydrogène est généralement de 2à 10 kcal /mol et les longueurs des liaisons hydrogène sont intermédiaires, elles sont plus longue que les liaisons covalentes mais plus courtes que les liaisons de Van Der Waals. [10]

d) Coefficient de partage : (Log P)

La lipophilie est une propriété qui a un effet important sur la solubilité, l'absorption, la distribution, la métabolisation et l'excrétion des médicaments. Hansch et Leo ont estimé que les molécules qui portent une grande lipophilie seront conservées à l'intérieur des lipides membranaires [9].

La meilleure méthode pour estimer la capacité d'un composé à se dissoudre dans les milieux aqueux et dans les milieux organiques est la mesure de la lipophile.

La lipophilie c'est le coefficient de partage d'une molécule médicamenteuse dans deux phases, organique et aqueuse.

$$P = \frac{[\text{concentration du médicament}]_{\text{Octanol}}}{[\text{concentration du médicament}]_{\text{Eau}}}$$

Le solvant organique couramment utilisé pour mesurer expérimentalement le coefficient de partage des molécules médicamenteuses est le n-octanol.

La phase aqueuse est habituellement tamponnée à pH = 7,4.

Une valeur positive pour log P, ($P > 1$), indique que le composé est lipophile.

Une valeur négative pour log P, ($P < 1$) indique que le composé est hydrophile.

Ce paramètre peut être utilisé pour estimer la biodisponibilité de l'administration par voie orale d'une molécule médicamenteuse.

- ✓ $0 < \text{Log } P < 3$: Activité biologique optimale (perméabilité, solubilité).
- ✓ $\text{Log } P < 0$: Composés trop hydrophiles (mauvaise perméabilité de bicouche lipidique).
- ✓ $\text{Log } P > 3$: Composés trop lipophiles (mauvaise solubilité aqueuse).

e) L'hydrophobicité

Le caractère hydrophobe d'une drogue est crucial en ce qui concerne la facilité avec laquelle elle traverse les membranes cellulaires et peut également être un facteur important lors de ses interactions avec le récepteur.

f) Réfractivité molaire: (MR)

La réfractivité est un cas particulier de volume moléculaire, c'est une réfractivité ramenée à une quantité de matière, ainsi que la réfractivité est la qualité de réfringence d'un corps, elle est utilisée dans la radioélectricité, en biologie... [11]. En chimie, la réfractivité moléculaire est un critère important pour mesurer le facteur stérique, elle est importante dans le cas où le substituent possède des électrons π ou des doublets libres [9]. La réfractivité moléculaire est définie par la relation suivante :

$$MR = V \frac{n^2 - 1}{n^2 + 1}$$

Où :

V : volume moléculaire.

n: est l'indice de réfraction.

g) Polarisabilité : (Pol)

La polarisabilité c'est, la facilité avec la quelle son nuage électronique se déforme par un champ électrique externe. La polarisabilité d'une molécule peut améliorer sa solubilité aqueuse. Cette caractéristique joue un rôle important dans la modélisation de nombreuses propriétés moléculaires [9].

La polarisabilité moléculaire est définie par la relation suivante :

$$P(e) = \epsilon_0 \alpha E$$

Où :

P(e) : Coefficient de polarisabilité.

ϵ_0 : Constante diélectrique.

α : Moment dipolaire électrique induit.

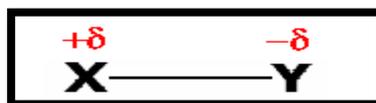
III.1.2.2 Les paramètres quantique/électronique

a) Moment dipolaire

La notion de moment dipolaire en physique et en chimie repose sur l'existence de dipôles électrostatiques. Il s'agit d'une répartition hétéroclite de charges électriques telles que le barycentre des charges positives ne coïncide pas avec celui des charges négatives. Le dipôle le plus simple est donc un couple de deux charges, de signes opposés, séparées par une distance non nulle. Un moment dipolaire est un vecteur noté généralement P ou D.

Deux charges égales en valeur absolue mais de signe opposé, +q et -q, séparées par une distance r, constituent un dipôle. On le caractérise un vecteur de module égal au produit $q \times r$, orienté, par convention, de la charge négative à la charge positive (du - vers le +).

Par exemple, dans une molécule diatomique polaire, chaque atome porte une charge partielle δ ou $-\delta$, où $0 < \delta < |Q_e|$: [12]



Le moment dipolaire est alors un vecteur qui se dirige vers l'atome le moins électronégatif:



III.1.2.3 Les propriétés « drug-like »**a) Règles de Lipinski**

Règles de Lipinski : sont un ensemble de valeurs de propriété qui ont été tirés de la classification des propriétés physicochimiques des composés clés de la drogue, etc.

La règle de Lipinski [13] est la plus utilisée pour caractériser les composés « drug-like ». Nous rappelons que cette règle a pour but d'identifier et sélectionner les composés posant des problèmes d'absorption et de perméabilité, et qu'elle a été établie à partir d'une liste de composés ayant passé avec succès les tests cliniques de phase II. D'autres règles ont depuis été introduites.

Les propriétés les plus importantes qui affectent la perméabilité intestinale passive d'une molécule sont la lipophilicité, l'hydrophobicité, la taille et la charge. L'effet individuel de chaque facteur est difficile ou même impossible à déterminer parce qu'ils se produisent simultanément.

Ces règles, communément appelées « règles de Lipinski » ou « la règle de 5 », comportent cinq critères physico-chimiques qui décrivent la molécule :

- ✓ Le poids moléculaire du composé ne doit pas être supérieur à 500 u.m.a.
- ✓ Le logarithme décimal du coefficient de partage eau / 1-octanol, noté logP, doit être inférieur à 5.
- ✓ Le nombre de donneurs de liaisons hydrogène doit être inférieur à 5.
- ✓ Le nombre d'accepteurs de liaisons hydrogène doit être inférieur à 10.
- ✓ nombre de liaisons retables (LF) ≤ 15

Les composés dont les propriétés physico-chimiques ne satisfont pas au moins 2 des règles sont fortement susceptibles de présenter des problèmes d'absorption ou de perméabilité.

La « règle des 5 » a été mise au point à partir de composés administrables par voie orale ayant passé avec succès la phase II des tests cliniques. Ce n'est donc pas une méthode pour distinguer les composés étant potentiellement des médicaments de ceux n'en étant pas, mais plutôt une méthode pour identifier les composés ayant une faible absorption ou une faible perméabilité.

b) Les règles de Veber :

Veber a identifié que la majorité de 1100 composés étudiés avec une biodisponibilité orale avaient moins de 10 liaisons rotatives et une surface polaire (PSA) de moins de 140 Å².

Les caractéristiques de Veber sont : [14]

- ✓ La surface polaire (PSA) < 140 Å²

✓ Les liaisons rotatives : NRB <10.

III.1.2.4 Surface moléculaire polaire (PSA)

La surface moléculaire polaire (PSA) est un paramètre très utile pour la prédiction des propriétés de transport du médicament. La surface polaire est définie comme la somme des surfaces des atomes polaires (généralement les atomes d'oxygène, les atomes d'azote et les atomes d'hydrogène attachés) dans une molécule. Il a été montré que ce paramètre est très bien corrélé avec l'absorption intestinale humaine, la perméabilité des monocouches de Caco-2 et la pénétration de la barrière hémato-encéphalique. [15]

Egalement, le PSA a été utilisé pour calculer le pourcentage d'absorption (%ABS) selon l'équation : $\%ABS = 109 \pm 0,345x PSA$ [16] .

III.1.2.5 Liaison rotative (Ntrob)

Ce paramètre topologique simple est une mesure de la flexibilité moléculaire. Il s'est avéré être un très bon descripteur de la biodisponibilité orale des médicaments.

La liaison rotative est définie comme toute liaison non annulaire unique, liée à un atome lourd non terminal (c'est-à-dire non hydrogène). Les liaisons CN Amide ne sont pas prises en compte en raison de leur haute barrière énergétique. [15]

III.1.2.6 L'efficacité du ligand (LE)

L'efficacité du ligand est une mesure de l'énergie de liaison par atome d'un ligand à son partenaire de liaison, tel qu'un récepteur ou une enzyme. [16]

À pour objectif d'atténuer la tendance à trop mettre l'accent sur l'optimisation de l'activité au détriment d'autres propriétés nécessaires.

Efficacité de ligand provient de l'observation ce que les composés les plus petits ont tendance à avoir les meilleures propriétés physico-chimiques et ADME que les grands composés. [14]

Le calcul de ce paramètre est réalisé comme suit : [14]

$$LE = \Delta G / (NH) \quad \text{Avec : } \Delta G = -RT \ln K_i \text{ est l'énergie libre de fixation.}$$

: NH est le nombre d'atomes lourds.

$$LE = (1.4 \text{ pIC}_{50}) / NH \quad \text{Avec : pIC}_{50} \text{ est l'activité biologique.}$$

III.1.2.7 L'efficacité lipophile de ligand (LLE)

LLE est un paramètre qui combine à la fois l'activité et le caractère lipophile, il est défini comme une mesure de l'efficacité d'un ligand exploitant son caractère lipophile pour se lier à une cible donnée. En bref, dans une série d'optimisation (les composés du plomb), il y a une grande probabilité d'obtenir de bonnes performances in vivo lorsque l'activité peut être augmentée sans augmenter les valeurs de log P ou log D.

Pour une LLE de 5 à 7 ou même plus, on peut dire que ces composés optimisés devraient être plus sélectifs. [14]

Le calcul de ce paramètre est réalisé comme suit :

$$\text{LLE} = \text{pIC}_{50} - \log P$$

III.1.3 Programmes de modélisation utilisés

Notre travail de recherche a été effectué au sein de l'équipe de chimie informatique et pharmaceutique du laboratoire LMCE (Laboratoire de chimie moléculaire et environnement) à l'université de Biskra.

L'étude des propriétés électroniques et structurales de benzimidazole et ses dérivés, a été réalisée par la modélisation moléculaire (mécanique moléculaire et QSAR), en utilisant le logiciel HyperChem (8.0.7) et le logiciel Gaussian (09).

la base de calcul c'est : **B3LYP/6-31G(d,p)/chelpg**

L'étude qualitative pour la surface polaire (PSA) et le nombre des liaisons rotatives en utilisant le logiciel molinspiration [15]

Nous avons utilisé également les logiciels Molegro Virtuel Docker (MVD) (5.5) [16] dans l'étude par docking moléculaire les interactions entre les benzimidazoles et les enzymes appropriés de leur activité cytotoxique.

La méthode de régression multilinéaire (RML) a été utilisée pour développer des modèles QSAR en utilisant le logiciel SPSS 20 [17].

Autres logiciels ont été utilisés dans notre travail à savoir : GaussView (5.0) [18], ChemDraw Ultra (8.0).

III.2.1 Etude de la relation structure-propriétés des dérivés de Benzimidazole

Dans notre travail nous avons choisi une série bioactive des dérivés de benzimidazole.

Ces molécules sont caractérisées par une structure de base de benzimidazole (figure III.1).

La série choisi comporte vingt molécules sont caractérisées par son activité cytotoxique (tableau III.1).

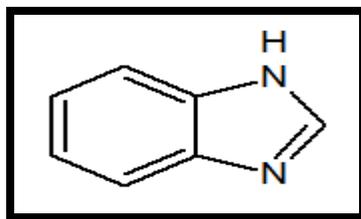
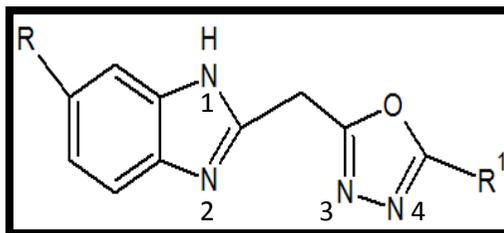
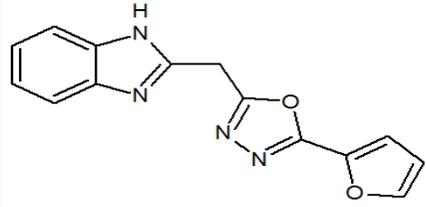
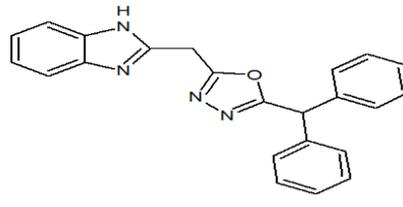
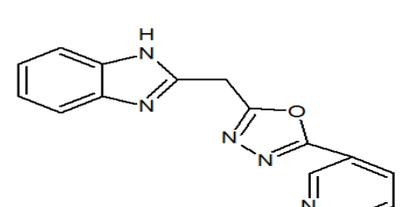
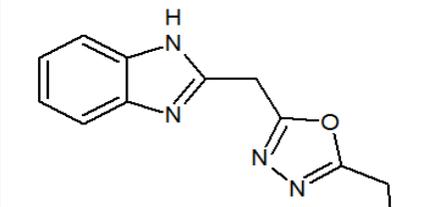
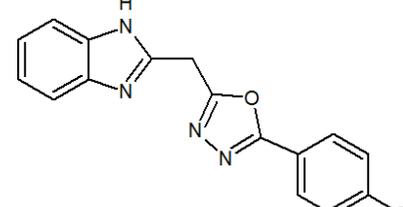
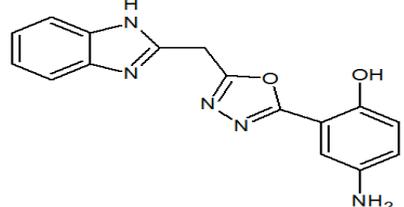
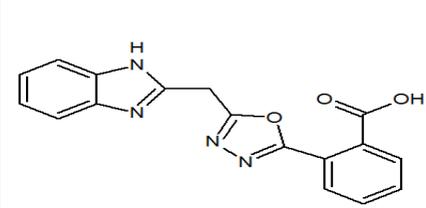
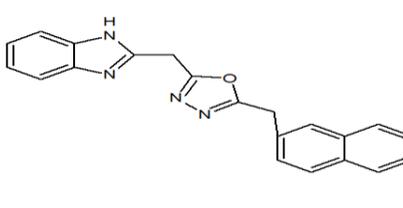
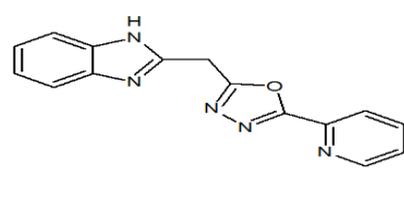
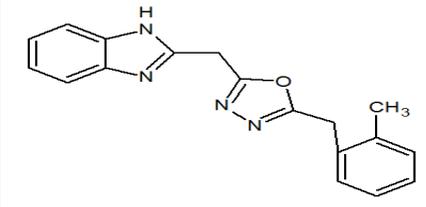
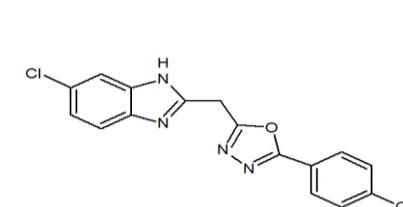


Figure III.1 : structure de Benzimidazole

Tableau III.1 : Structures chimiques des dérivés de Benzimidazole



composé	Structure	composé	Structure	composé	Structure
1		2		3	
4		5		6	
7		8		9	

10		11		12	
13		14		15	
16		17		18	
19		20			

III.2.1.1 Etude des propriétés physicochimiques des dérivés de Benzimidazole :

Dans ce travail, nous avons étudié sept propriétés physicochimiques de vingt des dérivés de benzimidazole (tableau. III.2) par rapport à leur activité. Tout d'abord, le vent- neuf dérivés sélectionnés de benzimidazole ont été optimisés par la méthode de mécanique moléculaire (champ de force MM+). En utilisant un logiciel HyperChem (8.0.7) pour calculer les paramètres suivants: **la polarisabilité (Pol)**, **le coefficient de partage octanol/eau (log P)**, **l'énergie d'hydratation (HE)**, **le volume molaire (MV)**, **la surface moléculaire (SAG)** et **le masse moléculaire (MW)**. Les résultats de calculs des paramètres physicochimiques de la série sélectionnées sont représentés dans le tableau suivant :

Tableau III.2 : Paramètres physicochimiques des dérivés de Benzimidazole

Composé	Mass (UMA)	Surface (Å ²)	Volume (Å ³)	Polarisabilité (Å ³)	Energie d'hydratation (Kcal/mol)	Refractivité (Å ³)	Log P
1	310,74	528,85	842,12	32,60	-12,09	92,15	1,06
2	345,19	543,54	878,67	34,53	-11,02	96,87	0,83
3	290,32	541,15	856,52	32,51	-12,13	91,99	1,21
4	276,30	503,10	789,96	30,68	-12,46	87,44	1,28
5	292,30	513,11	818,65	31,31	-19,05	89,04	0,25
6	310,74	522,02	835,55	32,60	-11,39	92,15	1,06
7	290,32	522,34	839,57	32,51	-10,45	91,72	1,43
8	336,35	578,45	943,89	35,62	-11,59	100,19	-0,71
9	282,32	483,98	765,28	30,20	-12,98	84,73	-0,11
10	266,26	474,12	745,83	27,83	-14,50	78,29	-0,45
11	366,42	632,28	1051,25	42,17	-11,30	120,80	2,23
12	277,29	493,6	783,31	29,97	-13,17	83,87	-0,2
13	239,24	455,13	707,62	24,70	-17,13	68,45	0,33
14	294,29	500,29	805,69	30,58	-11,98	87,57	0,68
15	307,31	528,22	846,29	32,66	-19,79	92,59	-1,47
16	320,31	531,46	856,99	33,23	-12,82	93,44	0,66
17	340,38	592,93	975,16	38,69	-12,59	110,18	1,29
18	277,29	497,71	785,41	29,97	-13,14	84,14	1,31
19	304,35	552,28	899,75	34,35	-10,89	96,27	1,36
20	345,19	551,66	885,37	34,53	-11,55	96,87	0,38

Grande valeur : **rouge** ; Petite valeur : **vert**

A lumière de ces résultats, on remarque que les valeurs du volume sont proportionnelles aux valeurs des surfaces. Le terme MM/p définit un volume, [20] suivant ce terme là on voit que le volume moléculaire est proportionnel aux valeurs de la masse molaire, malgré ça

L'ordre du volume de nos résultats est un peu différent de celui de la masse molaire pour la majorité des composés; cela est dû essentiellement au repliement de la structure chimique de ces substances ou les interactions électrostatiques surtout intramoléculaire.

La polarisabilité et la réfractivité molaire augmentent relativement avec la taille et le poids moléculaire des dérivés de benzimidazole étudiés (**Tableau. III.2**). Alors, ces résultats sont en accord avec la formule de Lorentz-Lorenz [21].

Cette relation montre que la réfractivité molaire et la polarisabilité augmentent avec le volume et la masse moléculaire. Par exemple, les composés 11 et 17 portent des substituant volumineux ont des valeurs importantes de la polarisabilité (42.70 Å³), (38.69 Å³) et de la réfractivité molaire (120.80Å³), (110.18Å³) respectivement. Contrairement, les composés 13 et 10 sont des petites molécules dans la série étudiée de benzimidazole, elles ont des petites valeurs de polarisabilité (24.70Å³) et (27.83 Å³) et de réfractivité molaire (68.45 Å³) et (78.29Å³).

L'énergie d'hydratation en valeurs absolues, les plus importantes sont celle des composés 7 (-10.45Kcal/mol), 19 (-10.89Kcal/mol) et la petite valeur est celle de composé 15 (-19.79Kcal/mol) (**Tableau. III.2**).

Les groupements hydrophobes dans les structures des dérivés de benzimidazole induit une diminution de l'énergie d'hydratation.

Le composé 10 possède un site donneur de proton (NH) et six sites accepteurs de proton (3N, NH, 2O). Par contre Le composé 15 possède 4 sites donneurs (NH, NH₂, OH) et sept sites accepteurs (3N, NH, O, OH et NH₂).

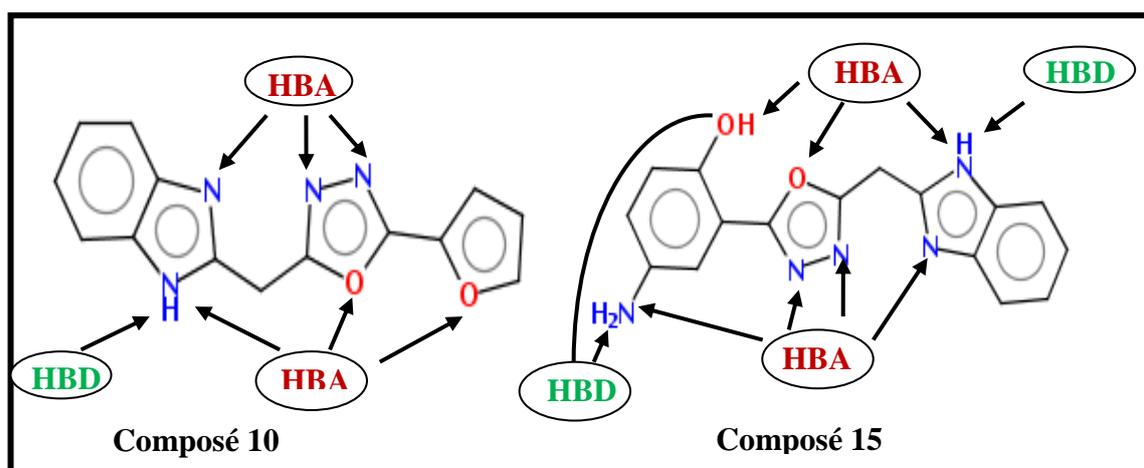


Figure III.2 : Les sites donneurs et accepteurs dans les composés 10 et 15

En effet, dans les milieux biologiques, les molécules polaires sont entourées par les molécules d'eau où les liaisons hydrogènes peuvent être établies entre les molécules d'eau et

ces molécules. Les sites donneurs de protons interagissent avec l'atome d'oxygène de l'eau et les sites accepteurs de protons interagissent avec l'atome d'hydrogène. Ces molécules hydratées sont partiellement déshydratées avant et lors de leurs interactions. Ces interactions de faible énergie et généralement réversibles, que l'on observe généralement avec les différents récepteurs biologiques.

La lipophilie est un facteur important dans les processus de solubilité, d'absorption, de distribution, de métabolisme et d'excrétion, ainsi que, l'activité pharmacologique. Hansch et Leo ont démontrés que les molécules très lipophiles se répartissent et sont conservées à l'intérieur des couches lipidiques des membranes cellulaires [22].

Pour une bonne biodisponibilité orale, le $\log P$ doit être supérieur à zéro et inférieur à 3 ($0 < \log P < 3$). En effet, Pour un $\log P$ trop élevé, le médicament présente une faible solubilité aqueuse. Dans le cas d'un $\log P$ très faible, le médicament présente une difficulté de se pénétrer dans les bicouches lipidiques des membranes cellulaires [23].

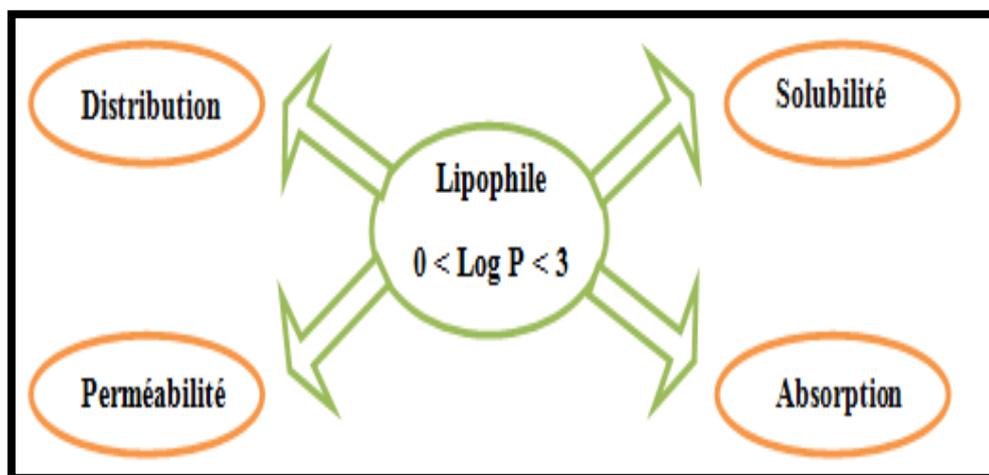


Figure III.3 : L'effet de lipophile dans la membrane cellulaire

La lipophilie augmente proportionnellement avec le caractère hydrophobe des groupes substituant. Les valeurs de $\log P$ des composés 7,11 sont dans le domaine des valeurs optimales ($0 < \log P < 3$) (1.43, 2.23) donc on peut dire que ces composés ont une activité biologique optimale (perméabilité, solubilité).

Ces composés possèdent généralement une bonne absorption intestinale, en raison d'un bon équilibre entre la solubilité et la perméabilité par diffusion passive ; l'élimination métabolique est faible, en raison de faibles liaisons avec les enzymes métaboliques. En outre, ces molécules ont tendance à être libérées par les reins, en raison de leur polarité.

On trouve aussi quelques valeurs de $\log P < 0$ comme pour les composés 8,15 qui présentent des faibles coefficients de partition -0.71 et -1.47 respectivement. Ces composés donnent une

bonne solubilité, mais une faible absorption et pénétration, en raison de leur mauvaise perméabilité (trop hydrophiles) par la diffusion passive. Ces composés ont une tendance élevée d'éliminer par les reins [23]. Ces composés présentent une meilleure tolérance gastrique.

III.2.1.2 Étude des propriétés « drug-like »

III.2.1.2.1 Règle de lipinski :

Drug-like apparaît comme un paradigme prometteur pour coder l'équilibre entre les propriétés moléculaires d'un composé qui influence ses pharmacodynamique et la pharmacocinétique et optimise finalement, leur absorption, distribution, métabolisme et excrétion (ADME) dans le corps humain comme une drogue. Les conditions empiriques pour satisfaire la règle de Lipinski et manifester une bonne biodisponibilité orale impliquent un équilibre entre la solubilité aqueuse d'un composé et sa capacité à diffuser passivement à travers les différentes barrières biologiques. Ces paramètres permettent à en tant que certain la perméabilité orale d'absorption ou de membrane qui se produit quand la molécule évaluée suit la règle de Lipinski.

Molécules qui violent plusieurs de ces règles peuvent avoir des problèmes avec la biodisponibilité. Par conséquent, cette règle établit certains paramètres structurels pertinents pour la prédiction théorique du profil de biodisponibilité orale, et est largement utilisé dans la conception de nouveaux médicaments [19].

Tous les composés de la série ont des masses molaires inférieures à 500 Da et tous les composés sont vraisemblablement solubles et traversent facilement les membranes cellulaires et ils conformes aux règles 2 et 3 pour tous les composés. Ils sont sans doute peu polaires et facilement absorbés; et ils répondent à la règle 4 sont solubles en solution aqueuse et donc capables d'atteindre la surface des membranes et ont une tolérance gastrique. Les médicaments les plus efficaces sont donc d'habitude le fruit d'un compromis : ils ne sont ni trop lipophiles, ni trop hydrophiles. [11]

Les résultats de calcul (tableau III.3) prouvent que tous les composés satisfont aux règles de Lipinski, suggérant que ces composés théoriquement n'aient pas des problèmes avec la disponibilité biologique orale.

Tableau III.3: Les paramètres de Lipinski des dérivés de Benzimidazole

Composé	Mass	Log P	HBA	HBD	Ntrob	N° de violation
1	310,74	1,06	5	1	3	0
2	345,19	0,83	5	1	3	0
3	290,32	1,21	5	1	4	0
4	276,30	1,28	5	1	3	0
5	292,30	0,25	6	2	3	0
6	310,74	1,06	5	1	3	0
7	290,32	1,43	5	1	3	0
8	336,35	-0,71	7	1	5	0
9	282,32	-0,11	5	1	3	0
10	266,26	-0,45	6	1	3	0
11	366,42	2,23	5	1	5	0
12	277,29	-0,2	6	1	3	0
13	239,24	0,33	6	1	3	0
14	294,29	0,68	5	1	3	0
15	307,31	-1,47	7	4	3	0
16	320,31	0,66	7	2	4	0
17	340,38	1,29	5	1	4	0
18	277,29	1,31	6	1	3	0
19	304,35	1,36	5	1	4	0
20	345,19	0,38	5	1	3	0

HBA : accepteur d'hydrogène HBD : donneur d'hydrogène

III.2.1.2.2 Règle de viber :

On remarque que tous les composés de la série ayant le nombre de liaisons inférieure à 10, donc tous les composés sont flexibles et capable de changé leur conformation Et on remarque aussi que tous les composés de notre série, ont des valeurs de la surface polaire sont en accord au la règle 2^{ème} de Veber c'est-à-dire tous les valeurs de la surface polaire inférieure à 140 Å² et cette résultat qui montre une bonne corrélation avec le transport moléculaire passive à travers les membranes, donc, il permet l'estimation des propriétés de transport des médicaments.

On a aussi les valeurs de pourcentage d'absorption (%ABS), on peut observer que tous les composés présente un grand pourcentage d'absorption allant 72.18% à 85.76% Cela signifie que ces composés avoir une bonne perméabilité membranaire cellulaire.

Tableau III.4: Les paramètres de Veber des dérivés de Benzimidazole

Composé	PSA	Ntrob	%ABS	Score de Viber
1	67,61	3	85,67	0
2	67,61	3	85,67	0
3	67,61	4	85,67	0
4	67,61	3	85,67	0
5	87,83	3	78,70	0
6	67,61	3	85,67	0
7	67,61	3	85,67	0
8	86,08	5	79,30	0
9	67,61	3	85,67	0
10	80,75	3	81,14	0
11	67,61	5	85,67	0
12	80,50	3	81,23	0
13	91,40	3	77,47	0
14	67,61	3	85,67	0
15	113,86	3	69,72	0
16	104,91	4	72,81	0
17	67,61	4	85,67	0
18	80,50	3	81,23	0
19	67,61	4	85,67	0
20	67,61	3	85,67	0

TPSA : Surface polaire %ABS : Pourcentage d'absorption

III.2.1.3 L'efficacité du ligand (LE)

Dans le tableau III.5 on remarque que si le nombre d'atomes lourds augmente, les valeurs de PIC₅₀ diminuent avec l'efficacité de ligand (LE), donc l'efficacité de ligand varié proportionnellement avec les valeurs de PIC₅₀.

Dans le tableau III.5 on observe que le composé 11 comporte un grand nombre d'atomes lourds (N= 28). Ce composé est caractérisé par des petites valeurs de LE, 0.22 pour EGFR et MCF7, 0.21 pour HepG2. Par contre, le composé 20 comporte le plus petit nombre d'atomes lourds (N=23). Ce composé est caractérisé par une grande valeur de l'efficacité de ligand, 0.37, 0.32 et 0.30 pour EGFR, MCF7 et HepG2 respectivement. Ces composés ont tendance à avoir les meilleures propriétés physico-chimiques et ADME. Mais une grande valeur de LE qui définit cette tendance est du composé 1 (LE= 0.45, 0.34, 0.31) avec N égale à 22, donc cette valeur de LE explique que le composé 1 a aussi une tendance à avoir des meilleures propriétés physico-chimiques et ADME.

Tableau III.5: Efficacité de ligand (LE) des dérivés de Benzimidazole

Composé	PIC ₅₀ (EGFR)	PIC ₅₀ (MCF7)	PIC ₅₀ (HepG2)	N	LE (EGFR)	LE (MCF7)	LE (HepG2)
1	7,1	5,3	4,9	22	0,45	0,34	0,31
2	5,4	5,2	4,3	23	0,33	0,32	0,26
3	4,6	4,6	4,3	22	0,29	0,29	0,27
4	4,5	4,2	4,2	21	0,30	0,28	0,28
5	4,0	4,5	4,1	22	0,25	0,29	0,26
6	5,3	4,7	4,2	22	0,34	0,30	0,27
7	4,4	4,3	4,0	22	0,28	0,27	0,25
8	4,2	4,5	4,5	25	0,24	0,25	0,25
9	4,0	4,2	4,0	20	0,28	0,29	0,28
10	4,0	4,2	4,3	20	0,28	0,29	0,30
11	4,3	4,4	4,2	28	0,22	0,22	0,21
12	4,0	4,1	4,0	21	0,27	0,27	0,27
13	4,0	4,1	4,0	18	0,31	0,32	0,31
14	4,0	4,5	4,3	22	0,25	0,29	0,27
15	4,1	4,3	4,1	23	0,25	0,26	0,25
16	4,0	4,4	4,1	24	0,23	0,26	0,24
17	5,2	4,6	4,6	26	0,28	0,25	0,25
18	4,1	4,3	4,0	21	0,28	0,29	0,27
19	4,3	4,4	4,0	23	0,26	0,27	0,24
20	6,0	5,3	4,9	23	0,37	0,32	0,30

Grande valeur : rouge ; Petite valeur : vert

III.2.1.4 Efficacité lipophile de ligand (LLE) :

Comme la lipophilie est le principal facteur de la promiscuité des composés, LipE devraient être plus sélectifs pour les composés optimisés. Il est suggéré de cibler un LipE dans une gamme de 5-7 ou même plus élevé [24] pour maximiser la puissance tout en maintenant la lipophilie aussi faible que possible, en raison de l'association entre une lipophilie élevée et plusieurs problèmes, y compris une faible solubilité, perméabilité membranaire, stabilité métabolique, ...etc

Dans la série étudiée LipE est entrain de changer au cours de l'optimisation (tableau III.6). Le composé 1 le plus hydrophile atteint un LipE est de 6.03 pour EGRF suivi par le composé 15 qui possède des valeurs de LipE, 5.75 et 5.77 pour HepG2 et MCF7 respectivement.

D'un autre côté, ces composés sont situés dans la gamme suggérée de 5-7, ce qui indique que ces composés ont été optimisés avec succès. par contre les valeur de l'efficacité lipophile de ligand pour le composé 11 ayant des valeurs d'efficacité de ligand lipophile les plus faible

dans cette série de composés, elles sont inférieures à 5 (2.12, 2.17,1.97) pour les trois récepteurs EGRF,MCF7 et HepG2 respectivement, donc dans ce cas , le égard d'affinité s'accompagne d'une augmentation de la lipophilie , donc les composés non optimisés avec succès.

Tableau III.6: Efficacité lipophile de ligand (LLE) des dérivés de Benzimidazole

Composé	Pic50 (EGFR)	Pic50 (MCF7)	Pic50 (HepG2)	Log P	LLE (EGFR)	LLE (MCF7)	LLE (HepG2)
1	7,1	5,3	4,9	1,06	6,03	4,24	3,84
2	5,4	5,2	4,3	0,83	4,60	4,37	3,47
3	4,6	4,6	4,3	1,21	3,39	3,39	3,09
4	4,5	4,2	4,2	1,28	3,23	2,92	2,92
5	4,0	4,5	4,1	0,25	3,75	4,25	3,85
6	5,3	4,7	4,2	1,06	4,29	3,64	3,14
7	4,4	4,3	4	1,43	2,97	2,87	2,57
8	4,2	4,5	4,5	-0,71	4,92	5,21	5,21
9	4,0	4,2	4	-0,11	4,11	4,31	4,11
10	4,0	4,2	4,3	-0,45	4,45	4,65	4,75
11	4,3	4,4	4,2	2,23	2,12	2,17	1,97
12	4,0	4,1	4	-0,2	4,20	4,30	4,20
13	4,0	4,1	4	0,33	3,67	4,43	3,67
14	4,0	4,5	4,3	0,68	3,32	3,82	3,62
15	4,1	4,3	4,1	-1,47	5,75	5,77	5,75
16	4,0	4,4	4,1	0,66	3,34	3,74	3,44
17	5,2	4,6	4,6	1,29	3,95	3,31	3,31
18	4,1	4,3	4	1,31	2,82	2,99	2,69
19	4,3	4,4	4	1,36	2,94	3,04	2,64
20	6,0	5,3	4,9	0,38	5,63	4,92	4,52

Grande valeur : rouge ; Petite valeur : vert

III.2.2 Etude quantitative des propriétés QSAR d'une série des dérivés de Benzimidazole

Dans ce travail, nous nous intéressons à l'étude des propriétés physico-chimiques et électroniques de benzimidazole et ses dérivés qui permettent de prévoir les paramètres physicochimiques qui sont influents sur l'activité biologique, afin de prédire l'activité biologique de nouvelles molécules, c'est pour cette raison on a met le choix d'une série de vingt dérivés de benzimidazole avec utilisation de différents descripteurs physico-chimiques et électroniques.

Les processus qui mènent à la découverte de nouveaux médicaments sont longs et fastidieux, et les taux de succès sont relativement faibles. L'identification de candidats par le biais de tests expérimentaux s'avère coûteuse, et nécessite de connaître en profondeur les mécanismes d'action de la protéine visée afin de mettre en place des essais efficaces. La modélisation de propriétés et d'activités de molécules (QSAR) peut considérablement accélérer ces processus en permettant une évaluation rapide de plusieurs milliers de molécules afin de déterminer lesquelles sont les plus susceptibles de se lier à une activité biologique ou propriété physique/chimique d'une molécule. L'un des challenges de la chémo-Informatique est d'être capable de décrire de manière simple des molécules afin de pouvoir les utiliser dans des études de similarité ou de pouvoir prédire leur activité en se basant sur les informations contenues dans les composés déjà connus. [25]

Tableau III.7: Descripteurs utilisés dans le développement des modèles QSAR

com	PIC50 EGFR	PIC50 MCF7	PIC50 HepG2	MW (UMA)	MV (Å3)	SAG (Å2)	POL (Å3)	REF (Å3)	Log P	HBA	HBD	nrotb	MD (Debye)	qN1	qN2	qO	qN3	qN4
1	7,1	5,3	4,9	310,74	842,12	528,85	32,60	92,15	1,06	5	1	3	5,0366	-0,506	-0,616	-0,334	-0,262	-0,400
2	5,4	5,2	4,3	345,19	878,67	543,54	34,53	96,87	0,83	5	1	3	3,2407	-0,513	-0,619	-0,349	-0,296	-0,349
3	4,6	4,6	4,3	290,32	856,52	541,15	32,51	91,99	1,21	5	1	4	4,2969	-0,489	-0,616	-0,406	-0,246	-0,463
4	4,5	4,2	4,2	276,30	789,96	503,10	30,68	87,44	1,28	5	1	3	4,7315	-0,519	-0,619	-0,343	-0,278	-0,393
5	4	4,5	4,1	292,30	818,65	513,11	31,31	89,04	0,25	6	2	3	5,6189	-0,510	-0,672	-0,364	-0,301	-0,387
6	5,3	4,7	4,2	310,74	835,55	522,02	32,60	92,15	1,06	5	1	3	5,0366	-0,506	-0,616	-0,334	-0,262	-0,400
7	4,4	4,3	4	290,32	839,57	522,34	32,51	91,72	1,43	5	1	3	4,5159	-0,499	-0,629	-0,352	-0,385	-0,269
8	4,2	4,5	4,5	336,35	943,89	578,45	35,62	100,19	-0,71	7	1	5	6,7241	-0,506	-0,630	-0,281	-0,262	-0,415
9	4	4,2	4	282,32	765,28	483,98	30,20	84,73	-0,11	5	1	3	4,9477	-0,512	-0,628	-0,385	-0,279	-0,389
10	4	4,2	4,3	266,26	745,83	474,12	27,83	78,29	-0,45	6	1	3	5,0874	-0,499	-0,620	-0,395	-0,248	-0,429
11	4,4	4,4	4,2	366,42	1051,25	632,28	42,17	120,80	2,23	5	1	5	4,1068	-0,527	-0,625	-0,379	-0,389	-0,271
12	4	4,1	4	277,29	783,31	493,6	29,97	83,87	-0,2	6	1	3	3,7742	-0,499	-0,612	-0,339	-0,258	-0,404
13	4	4,1	4	239,24	707,62	455,13	24,70	68,45	0,33	6	1	3	2,7412	-0,478	-0,607	-0,364	-0,223	-0,446
14	4	4,5	4,3	294,29	805,69	500,29	30,58	87,57	0,68	5	1	3	3,4133	-0,510	-0,616	-0,347	-0,274	-0,399
15	4,2	4,3	4,1	307,31	846,29	528,22	32,66	92,59	-1,47	7	4	3	7,1803	-0,504	-0,631	-0,383	-0,342	-0,312
16	4	4,4	4,1	320,31	856,99	531,46	33,23	93,44	0,66	7	2	4	5,6370	-0,489	-0,618	-0,367	-0,244	-0,421
17	5,2	4,6	4,6	340,38	975,16	592,93	38,69	110,18	1,29	5	1	4	4,5559	-0,513	-0,621	-0,400	-0,247	-0,461
18	4,1	4,3	4	277,29	785,41	497,71	29,97	84,14	1,31	6	1	3	6,6396	-0,516	-0,606	-0,362	-0,283	-0,397
19	4,3	4,4	4	304,35	899,75	552,28	34,35	96,27	1,36	5	1	4	4,3382	-0,488	-0,625	-0,357	-0,229	-0,457
20	6	5,3	4,9	345,19	885,37	551,66	34,53	96,87	0,38	5	1	3	4,7994	-0,479	-0,606	-0,345	-0,281	-0,388

III.2.2.1 Développement des modèles QSAR :

Le développement d'un modèle QSAR exige un ensemble de diverses données, pour cet effet, un nombre considérable de descripteurs doivent être considérés dans l'étude QSAR. Les descripteurs sont des valeurs numériques qui codent les différentes caractéristiques structurales des molécules. La matrice de corrélation de Pearson a été effectuée sur tous les descripteurs en utilisant le logiciel SPSS [26]. L'analyse de la matrice révèle seize descripteurs pour le développement du modèle QSAR. Les valeurs des descripteurs sélectionnés pour le modèle QSAR sont présentées dans le tableau III.7.

La corrélation entre l'activité biologique et les descripteurs, exprimés par les relations mathématiques suivantes :

$$\begin{aligned} \text{pHepG2} = & 9.991 + 0.027 \text{ SAG} - 0.186 \text{ LOG P} - 1.078 \text{ POL} + 0.313 \text{ REF} - 0.426 \\ & \text{HBD} - 0.505 \text{ ntrob} + 8.405 \text{ N1} - 0.146 \text{ MD} + 16.234 \text{ N3} + 11.908 \text{ N4} - 3.494 \text{ O} \end{aligned} \quad (1)$$

$$n = 20 \quad R = 0.928 \quad R^2 = 0.862 \quad SE = 0.160 \quad F = 4.531 \quad Q = 5.782$$

$$\begin{aligned} \text{pEGFR} = & 2.178 - 0.05 \text{ VOL} + 0.122 \text{ SAG} - 1.282 \text{ ntrob} + 28.909 \text{ N3} + \\ & 18.876 \text{ N4} - 0.272 \text{ HBD} \end{aligned} \quad (2)$$

$$n = 20 \quad R = 0.901 \quad R^2 = 0.812 \quad SE = 0.436 \quad F = 9.349 \quad Q = 2.066$$

$$\text{pMCF7} = 0.839 - 0.023 \text{ VOL} + 0.035 \text{ SAG} + 0.018 \text{ MAS} - 0.142 \text{ HBA} \quad (3)$$

$$n = 20 \quad R = 0.87 \quad R^2 = 0.757 \quad SE = 0.204 \quad F = 11.714 \quad Q = 4.268$$

Où :

2.178, 0.839, 9.991: l'ordonnée à l'origine ou de la valeur de base de pIC₅₀.

R² : coefficient de détermination

SE : erreur standard de l'estimation

F : statistique Fischer

n : indique le nombre d'observations.

III.2.2.2 Interprétation :

Les valeurs de la fraction de la variance peuvent varier entre 0 et 1. Le modèle QSAR doit considérer un $R^2 > 0.6$ pour qu'il soit valide.

Les valeurs $R = 0.928$ ($R^2 = 0.862$) (modèle 1), $R = 0.901$ ($R^2 = 0.812$) (modèle 2) et $R = 0.870$ ($R^2 = 0.757$) (modèle 3) nous permis d'indiquer fermement la corrélation entre les différents descripteurs utilisés et les activités biologiques dans HepG2, EGFR et MCF7 respectivement.

La valeur F est jugé statistiquement significative au niveau de 95%, étant donné que la valeur F calculée est supérieure par rapport aux valeurs du tableau critiques. La valeur positive de facteur de qualité (Q) suggère son haut pouvoir prédictif. [27-28]

Dans les trois modèles, les coefficients positifs montrent que toute augmentation dans les valeurs de ces paramètres entraîne une augmentation de l'activité biologique. Par contre les coefficients négatifs l'augmentation des paramètres provoquent l'augmentation de l'activité cytotoxique (tableau III.8)

Dans les trois équations l'augmentation de la surface moléculaire en accord avec l'augmentation de surface de contact ligand-récepteur, donc on observe le coefficient positif qui montre que toute augmentation de SAG entraîne une augmentation de l'activité inhibitrice.

Dans l'équation 1, le coefficient positif de la masse explique qu'il entraîne une augmentation de l'activité biologique de ces composés, c'est-à-dire notre noyau avec une grande valeur mieux puisque il ne besoin d'une diffusion dans les tissu.

Le coefficient positif de l'énergie d'hydratation et négatif de Log P indique que les dérivés hydrophiles donnent une bonne activité biologique.

Dans l'équation 1 le coefficient négatif de log P explique que toute augmentation de la lipophile des molécules entraîne une diminution de l'activité biologique de sorte que les molécules hydrophiles sont plus importantes pour cette activité biologique.

Dans les équations 1 et 2 le nombre de liaison rotatif avec le coefficient négatif obtient de diminution de l'activité biologique c'est-à-dire pour avoir une bonne activité cytotoxique, la molécule doit être moins flexible [14].

Les paramètres HBD, et les charge atomiques dans les modèles 1, 2 et 3 sont responsables aux interactions électrostatiques de dérivés de benzimidazoles avec les résidus d'acides aminés dans les sites actifs de récepteur biologique. Les liaisons hydrogène donneur et la charge atomique O empêchent l'activité biologique. Par contre, les charges atomiques N1, N3 et N4

participent à la formation des liaisons de nature électrostatique avec les sites actifs en augmentant l'activité cytotoxique des benzimidazoles.

Tableau III.8 : Les signes de différents descripteurs dans les modèles QSAR obtenus.

Descripteur	SAG	VOL	MAS	POL	REF	LOGP	LHA	LHD	ntrob	MD	N1	N3	N4	O
Modèle 1	+			-	+	-		-	-	-	+	+	+	-
Modèle 2	+	-						-	-			+	+	
Modèle 3	+	-	+				-							

Les matrices de corrélation entre les descripteurs obtenus par l'analyse MLR et l'activité biologique sont reportée dans les tableaux **III.9**, **III.10** et **III.11**

Tableau III.9 : Matrice des corrélations de 1^{ère} modèle

	PIC ₅₀	SAG	LOGP	POL	REF	MD	LHD	ntrob	NH	N	NR	O
PIC ₅₀	1											
SAG	0,393	1										
LOGP	0,032	0,381	1									
POL	0,357	0,981	0,403	1								
REF	0,355	0,977	0,412	0,998	1							
MD	0,044	0,117	-0,412	0,110	0,109	1						
LHD	-0,197	-0,013	-0,551	-0,006	0,004	0,556	1					
NTROB	0,110	0,779	0,228	0,704	0,696	0,129	-0,108	1				
NH	0,060	-0,287	-0,238	-0,384	-0,415	-0,184	0,025	-0,119	1			
N	0,147	-0,354	0,238	-0,420	-0,436	-0,154	-0,280	-0,055	0,475	1		
NR	-0,162	0,250	0,084	0,327	0,338	0,110	0,290	-0,047	-0,426	-0,965	1	
O	0,219	0,048	-0,125	0,010	-0,004	0,121	-0,215	0,066	-0,023	0,046	0,075	1

Tableau III.10 : Matrice des corrélations de 2^{ème} modèle

	PIC ₅₀	MAS	VOL	SAG	HBA
PIC ₅₀	1				
MAS	0,604	1			
VOL	0,357	0,906	1		
SAG	0,380	0,903	0,997	1	
HBA	-0,368	-0,131	-0,144	-0,150	1

Tableau III.11 : Matrice des corrélations de 3^{ème} modèle

	PIC ₅₀	VOL	SAG	ntrob	N3	N1	HBD
PIC ₅₀	1						
VOL	0.267	1					
SAG	0.299	0.997	1				
ntrob	-0.134	0.782	0.779	1			
N3	-0.012	0.263	0.250	-0.047	1		
N1	-0.012	-0.361	-0.354	-0.055	-0.965	1	
HBD	-0.219	-0.012	-0.013	-0.108	0.290	-0.280	1

Les tableaux **III.9**, **III.10** et **III.11** montrent que les descripteurs obtenus sont corrélés entre eux et avec l'activité biologique. Les tableaux **III.9** et **III.11** montrent que le descripteur surface (SAG) est le paramètre le plus important (39.3% et 29.9%) dans la corrélation entre les descripteurs sélectionnés et les activités biologique des dérivés de benzimidazole. Le tableau **III.10** montre que le descripteur masse (MAS) est le paramètre le plus important (60.4 %) dans la corrélation entre les descripteurs sélectionnés et l'activité biologique des dérivés de benzimidazole

III.2.2.3 Validation des modèles QSAR :

Afin de tester la validité de la puissance prédictive des modèles RML sélectionnés (Eq. 1, 2 et 3), la technique de LOO (leave-one-out) a été utilisée. Les modèles développés ont été validés par le calcul des paramètres statistiques suivants: PRESS, SSY, Spress, r^2_{adj} , r^2_{CV} et PE (tableau **III.12**).

Tableau **III.12**: Paramètres de validation croisée

modèle	PRESS	SSY	PRESS/SSY	Spress	r^2_{cv}	r^2_{adj}	6PE
1	0.206	1.490	0.139	0.101	0.862	0.672	0.125
2	2.472	13.140	0.188	0.351	0.812	0.725	0.056
3	0.623	2.570	0.242	0.176	0.757	0.693	0.053

résiduelle prédictive des carrés (PRESS, Predictive Residual Sum of Squares) est une mesure des capacités prédictives d'un modèle. La petite valeur de PRESS signifie la meilleure prévisibilité du modèle [29].

Dans le (tableau **III.12**) les valeurs de PRESS sont données par : 0.206, 2.472 et 0.623 et les valeurs de SSY sont 1.490, 13.140 et 2.570 pour les modèles 1,2 et 3 respectivement.

D'après les résultats présentés dans le tableau **III.12**, le modèle 1, 2 et 3 sont statistiquement significatifs. En outre, pour les modèles QSAR raisonnables, les rapports PRESS / SSY doivent être inférieur à 0,4 [29]. Les valeurs présentées dans le tableau **III.12** indiquent que les modèles développés ont des rapports de 0.139, 0.188 et 0.242 pour les modèles 1, 2 et 3 respectivement ; toutes ces valeurs sont inférieures à 0.4.

Spress de ces trois modèles ont des petites valeurs de 0.101, 0.351 et 0.176 pour les modèles 1, 2 et 3 respectivement. Ce résultat confirme le pouvoir prédictif de ces trois modèles.

Les valeurs de r^2_{cv} et r^2_{adj} des trois modèles QSAR 1, 2 et 3 sont (0.862, 0.812, 0.757) et (0.672,0.725 ,0.693) respectivement. Les valeurs élevées du r^2_{cv} et r^2_{adj} sont des critères essentiels pour la meilleure qualification des modèles QSAR.

L'erreur de prédiction du coefficient de corrélation (PE) est un autre paramètre utilisé pour déterminer le pouvoir prédictif des modèles proposés. Nous avons calculé la valeur de r pour les modèles obtenus et ils sont présentés dans le tableau III.12. Pour les trois modèles la condition $r > 6PE$ est satisfaite et donc ils ne peuvent être considérés comme ayant un bon pouvoir prédictif [30].

L'activité anticancéreuse expérimentale, prédite et résiduelle de benzimidazole et ses dérivés sont représentés dans le tableau III.13. Le résidu est dû d'une part à l'incertitude de la mesure expérimentale et d'autre part à l'imperfection du modèle, donc si l'activité prédite est plus proche de l'activité expérimentale, le modèle peut être applicable [31].

Tableau III.13: Les valeurs expérimentales, prédites et résiduelles de (P) de série bioactive de benzimidazole

Com	p HepG2 exp	pHepG2 pred	pHepG2 resd	pEGFR exp	pEGFR pred	pEGFR resd	pMCF7 exp	pMCF7 pred	pMCF7 resd
1	4,9	4,657	0,279	7,1	5,896	1,195	5,3	4,851	0,449
2	4,3	4,286	0,037	5,4	5,861	-0,429	5,2	5,150	0,050
3	4,3	4,303	0,051	4,6	4,679	-0,077	4,6	4,573	0,027
4	4,2	4,208	-0,040	4,5	5,002	-0,493	4,2	4,526	-0,326
5	4,1	4,046	-0,041	4	3,980	0,020	4,5	4,361	0,139
6	4,2	4,471	-0,283	5,3	5,386	-0,039	4,7	4,763	-0,063
7	4	3,982	-0,034	4,4	4,143	0,259	4,3	4,304	-0,004
8	4,5	4,477	-0,048	4,2	4,071	0,144	4,5	4,411	0,089
9	4	3,882	0,139	4	3,932	0,068	4,2	4,539	-0,339
10	4,3	4,404	0,046	4	3,831	0,169	4,2	4,207	-0,007
11	4,2	4,234	0,034	4,4	4,380	-0,033	4,4	4,641	-0,241
12	4	4,079	-0,144	4	4,539	-0,539	4,1	4,222	-0,122
13	4	3,960	0,062	4	3,805	0,195	4,1	3,931	0,169
14	4,3	4,282	-0,050	4	3,879	0,121	4,5	4,390	0,110
15	4,1	4,156	-0,023	4,2	4,146	0,009	4,3	4,383	-0,083
16	4,1	3,985	0,109	4	4,048	-0,048	4,4	4,487	-0,087
17	4,6	4,562	0,057	5,2	5,139	0,105	4,6	4,550	0,050
18	4	4,029	-0,006	4,1	4,348	-0,214	4,3	4,318	-0,018
19	4	4,127	-0,189	4,3	4,501	-0,200	4,4	4,213	0,187
20	4,9	4,870	0,043	6	6,220	-0,211	5,3	5,280	0,020

P exp: Activité expérimentale
P pred. : Activité prédite
P resd. : Activité résiduelle

La figure III.3 ci-dessous montre les courbes de régression linéaire des valeurs prédites par rapport aux valeurs expérimentales de l'activité cytotoxique des denzimidazole. Les courbes des modèles 1,2 et 3 représentent pour être plus pratique avec $r^2 = 0.862$, 0.812 et 0.757 respectivement. Il indique que les modèles 1, 2 et 3 peuvent être appliqués avec succès pour prédire l'activité cytotoxique de cette série de molécules.

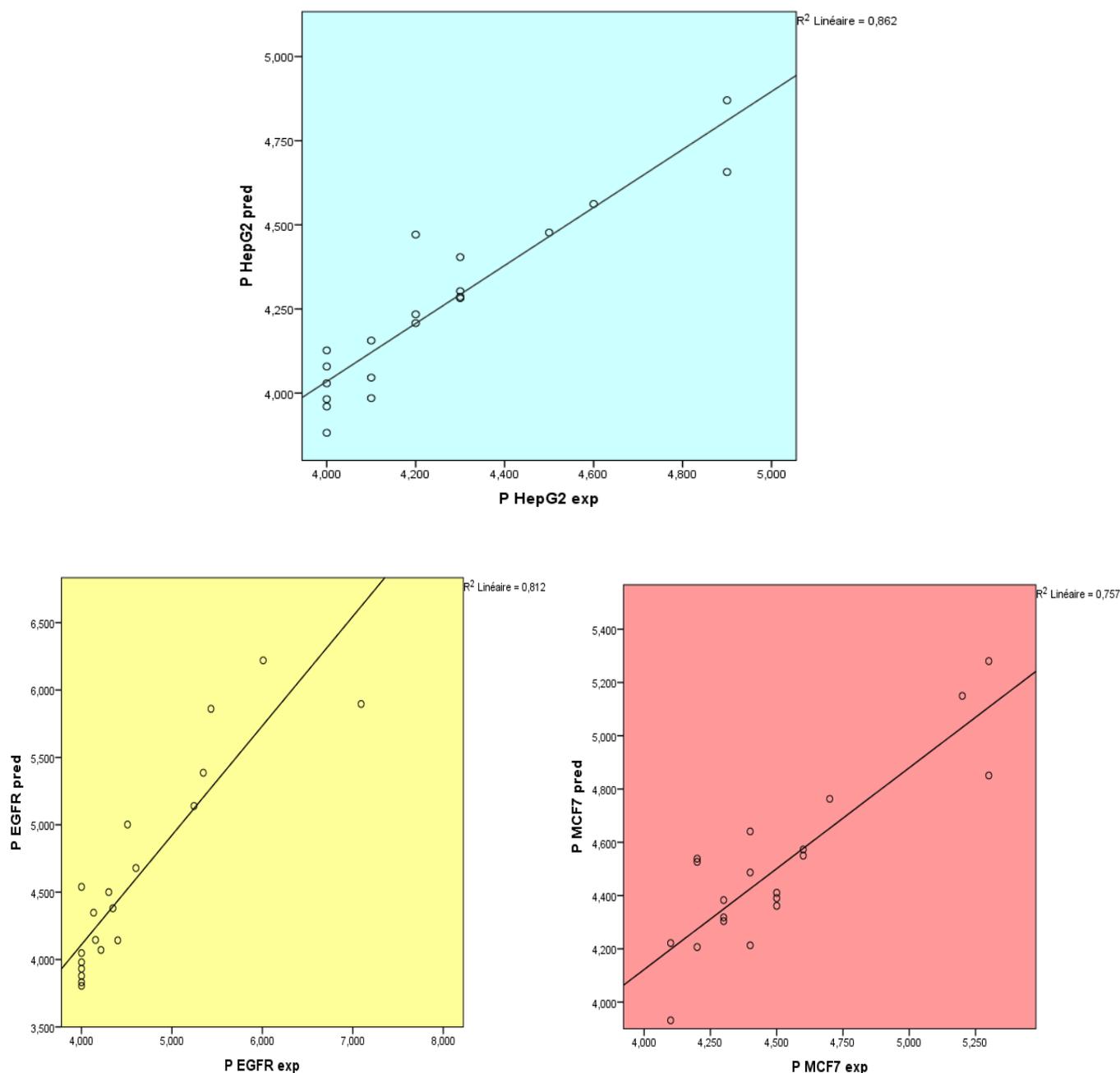


Figure III.5: Les courbes de la régression linéaire des valeurs prédites en fonction des valeurs expérimentales de pHepG2, pEGFR et pMCF7.

Pour rechercher la présence d'une erreur systématique dans l'élaboration des modèles QSAR, les valeurs résiduelles des trois activités biologiques (P) ont été tracées en fonction des valeurs expérimentales, comme le montre la figure III.5

La répartition des résidus dans les deux côtés de l'axe des abscisses indique l'absence de l'erreur systématique [32]. Il indique que ces trois modèles peuvent être appliqués avec succès pour prédire les activités biologiques pour la série de molécules utilisée dans le développement des modèles QSAR.

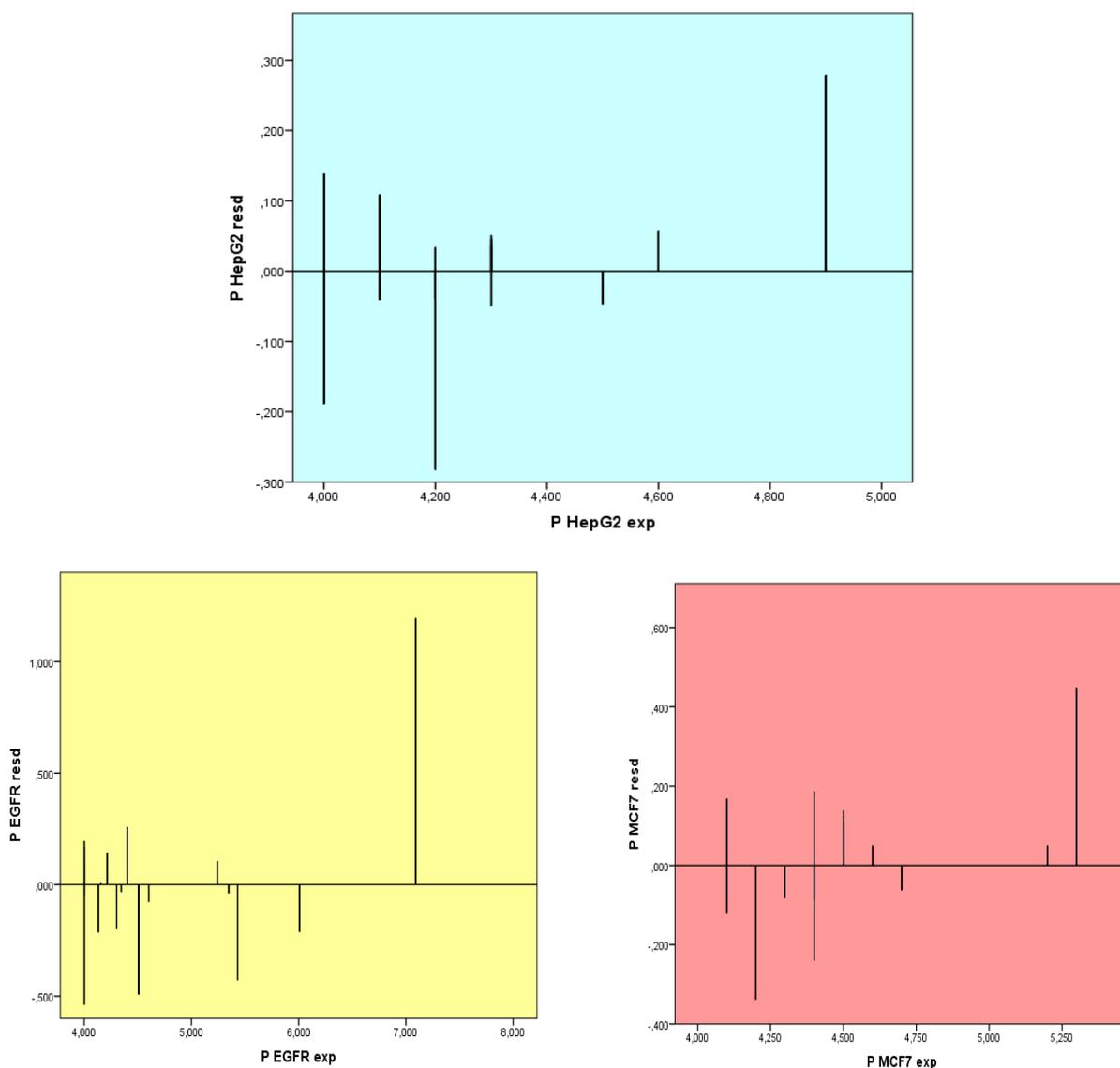


Figure III.6: Représentations graphique des valeurs résiduelles en fonction des valeurs expérimentales pour les modèles 1, 2, 3.

III.2.3 Conclusion

La modélisation moléculaire et l'analyse QSAR ont été effectuées pour déterminer quantitativement l'effet de la structure moléculaire des composés étudiés sur leur activité biologique. Des modèles mathématiques ont été développés pour prédire l'activité cytotoxique de benzimidazole de trois lignées cellulaires, EGFR (cancer de la peau), MCF7 (cancer du sein), HepG2 (cancer du foie).

La validité des modèles obtenus ont été établie par la détermination des paramètres statistiques appropriés. Le modèle mis en place a été utilisé pour prédire l'activité biologique étudiée. Les valeurs prédites sont plus proches que les valeurs expérimentales pour les trois activités biologiques.

Les faibles valeurs résiduelles de l'activité biologique et les grandes valeurs du coefficient de validation croisé (R^2_{CV}) obtenus suggèrent une bonne capacité prédictive des modèles QSAR développés. Il indique que les trois activités biologiques pour la série étudiée de dérivés de benzimidazole peuvent être modélisées avec succès en utilisant des différents descripteurs moléculaires.

Références bibliographique

- [1] : HUYNH, Thi Ngoc Phuong. Synthèse et études des relations structure/activité quantitatives (QSAR/2D) d'analyse benzo [c] phénanthridiniques. 2007. Thèse de doctorat. Université d'Angers.
- [2]: CRAMER, Christopher J. Essentials of computational chemistry: theories and models. John Wiley & Sons, 2004.
- [3] : FAYET, Guillaume. Développement de modèles QSPR pour la prédiction des propriétés d'explosibilité des composés nitroaromatiques. 2010. Thèse de doctorat. Chimie ParisTech.
- [4] : HyperChem help.
- [5] : Abhilash Thakur. Mamta Thakur, Total QSAR study on Phenyl Acridine Derivatives in Reference to DNA Binding Affinity, 3rd International conference on Medical, Biological and Pharmaceutical Sciences, 2013
- [6] : Kubinyi, Hugo, Gerd Folkers, and Yvonne C. Martin, eds. 3D QSAR in Drug Design: Volume 2: Ligand-Protein Interactions and Molecular Similarity. Vol. 2. Springer Science & Business Media, 1998
- [7]. Ivanciuc, Ovidiu. "3D QSAR models." QSPR/QSAR Studies by Molecular Descriptors (2001): 233-280.
- [8] : Cohen, Y. "Pharmacologie moléculaire." (1978).
- [9] : Belaidi, S., N. Melkemi, and D. Bouzidi. "Molecular geometry and structure-property relationships for 1, 2-dithiole-3-thione derivatives." Int J Chem Res 4.2 (2012): 134-139.
- [10] : Hocquet, Alexandre, and Morten Langgård. "An evaluation of the MM+ force field." Molecular modeling annual 4.3 (1998): 94-112.
- [11] : Xavier, M. In silico Screening Ligand-based, 2009.
- [12]. <https://www.futura-sciences.com/sciences/definitions/physique-moment-dipolaire-9262/>
- [13] : Lipinski, Christopher A., et al. "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings." Advanced drug delivery reviews 64 (2012): 4-17.

- [14] : Salah, Toufik, et al. "Molecular Geometry, Electronic Properties, MPO Methods and Structure Activity/Property Relationship Studies of 1, 3, 4-Thiadiazole Derivatives by Theoretical Calculations." *Reviews in Theoretical Science* 3.4 (2015): 355-364
- [15] : <http://www.molinspiration.com/>
- [16] : Molegro, A. P. S. "MVD 5.0 Molegro Virtual Docker." *DK-8000 Aarhus C, Denmark* (2011).
- [17] : Priyatno, Duwi. "Cara kilat belajar analisis data dengan SPSS 20." *Yogyakarta: Andi Offset* (2012).
- [18] : Dennington, Roy D., Todd A. Keith, and John M. Millam. "GaussView 5.0. 8." *Gaussian Inc* (2008).
- [19] : Kuntz, I. D., et al. "The maximal affinity of ligands." *Proceedings of the National Academy of Sciences* 96.18 (1999): 9997-10002.
- [20] : Graham, L., Patrick; « chimie pharmaceutique » ; 2^{ème} édition ; de boeck ; Paris; 2003; 206, 266, 270, 271.
- [21] : Dimitrov, V., and T. Komatsu. "An interpretation of optical properties of oxides and oxide glasses in terms of the electronic ion polarizability and average single bond strength." *J. Univ. Chem. Technol. Metall* 45.3 (2010): 219-250.
- [22] : Pajouhesh, Hassan, and George R. Lenz. "Medicinal chemical properties of successful central nervous system drugs." *NeuroRx* 2.4 (2005): 541-553.
- [23] : Di, Li, and Edward H. Kerns. *Drug-like properties: concepts, structure design and methods from ADME to toxicity optimization*. Academic press, 2015.
- [24] : Pathak, Jitendra, et al. "Theoretical Raman and FTIR vibrational analysis of 2-phenyl-1H-indene-1, 3 (2H)-dione by ab initio method." *J Atomic Mol Sci* 3.2 (2012): 95-105.
- [25] : Elkaïm, Judith. *Drug design in silico: criblage virtuel de protéines à visée thérapeutique*. Diss. Bordeaux 1, 2011.
- [26] : Melkemi, N. *Etude des corrélations entre les paramètres structuraux et les propriétés bioactives de quelques composés organiques*. Diss. Université Mohamed Khider Biskra, 2013.

- [27] : Almi, Zineb, et al. "Structure Activity Relationships, QSAR Modeling and Drug-like calculations of TP inhibition of 1, 3, 4-oxadiazoline-2-thione Derivatives." *International Letters of Chemistry, Physics and Astronomy* 18 (2014): 113-124
- [28] : Dermeche, Keltoum, et al. "Qualitative Structure-Activity Relationships and 2D-QSAR Modeling of TNF- α Inhibition by Thalidomide Derivatives." *Journal of Bionanoscience* 9.5 (2015): 395-400.
- [29] : Podunavac-Kuzmanović, Sanja O., Dragoljub D. Cvetković, and Dijana J. Barna. "QSAR analysis of 2-amino or 2-methyl-1-substituted benzimidazoles against *Pseudomonas aeruginosa*." *International journal of molecular sciences* 10.4 (2009): 1670-1682.
- [30] : Srivastava, A. K., and Neerja Shukla. "Quantitative structure activity relationship (QSAR) studies on a series of imidazole derivatives as novel ORL1 receptor antagonists." *Journal of Saudi Chemical Society* 17.3 (2013): 321-328.
- [31] : Clark, Matthew, and Richard D. Cramer. "The probability of chance correlation using partial least squares (PLS)." *Molecular Informatics* 12.2 (1993): 137-145.
- [32] : Jalali-Heravi, Mehdi, and Anahita Kyani. "Use of computer-assisted methods for the modeling of the retention time of a variety of volatile organic compounds: a PCA-MLR-ANN approach." *Journal of chemical information and computer sciences* 44.4 (2004): 1328-1335.

IV.1.1 Introduction

Le docking moléculaire consiste à déterminer la position relative de deux molécules (un ligand et un récepteur) qui sont associés pour donner une réponse biologique. Cette technique de modélisation moléculaire peut être à la conception de nouveau médicament.

D'un point de vue informatique, la prédiction de structure se traduit par un problème d'optimisation. A chaque position relative possible est associée une énergie (également appelé score, fitness, etc.). La structure la plus probable est alors celle qui minimise cette énergie en donnant un système stable.

IV.1.2 Docking moléculaire

Les interactions entre les molécules et son récepteur sont la base de la plupart des mécanismes biologiques. Les détails de ces interactions, au niveau moléculaire, sont donc d'un très grand intérêt et peuvent être étudiés par la cristallographie aux rayons X ou par la résonance magnétique nucléaire (RMN). [1]

Le docking moléculaire vise à prédire la structure d'un complexe moléculaire à partir des molécules isolées, ce qui est considérablement plus facile à mettre en œuvre, moins cher et plus rapide. [2]

Les logiciels de docking moléculaire sont donc des outils très utiles en biologie, pharmacie et médecine, car la plupart des principes actifs sont de petites molécules (ligands) qui interagissent avec une cible biologique d'intérêt thérapeutique, généralement protéique (récepteur), afin d'influencer le mécanisme dans lequel cette protéine est impliquée [1].

La procédure typique à mettre en place pour prédire les interactions entre un ligand et son récepteur est présentée dans la figure IV.1. La première étape est l'obtention d'une structure pour le récepteur, par DRX, par RMN ou par modélisation moléculaire [1].

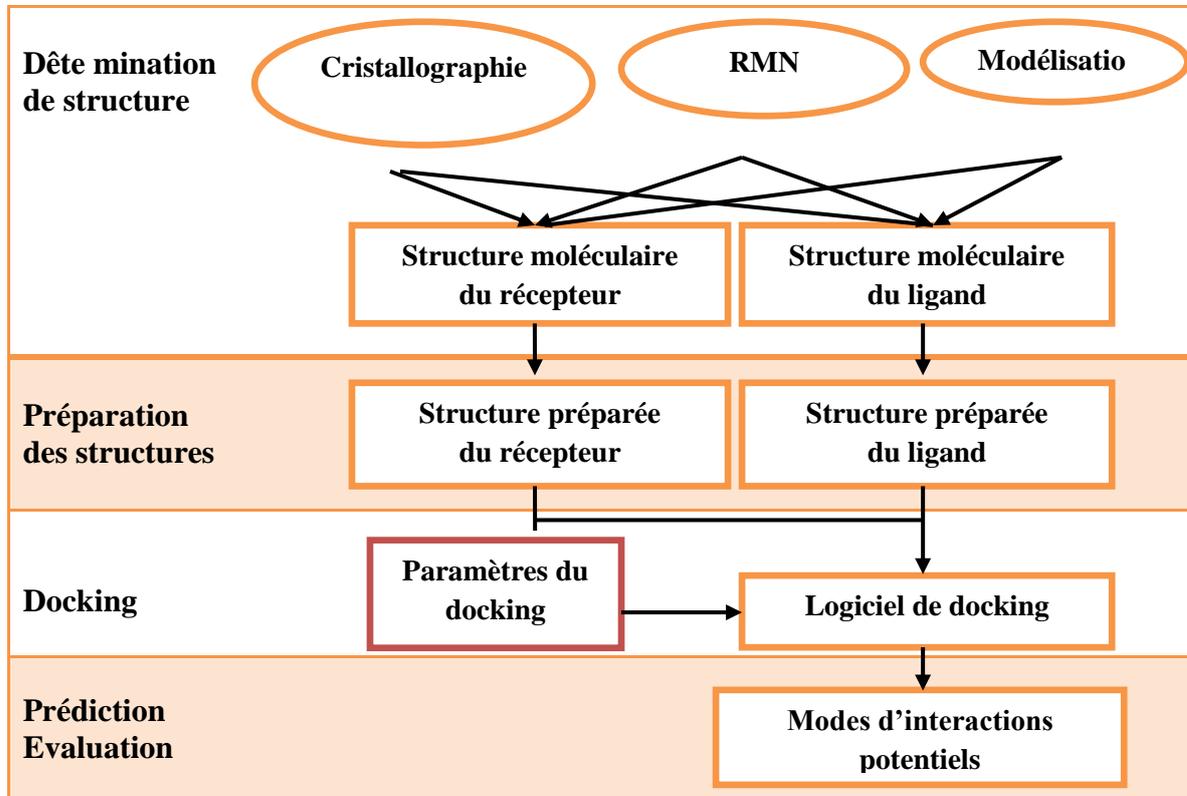


Figure IV.1 : Etapes typique d'un docking moléculaire

La conformation initiale du ligand est généralement modifiée et optimisée durant le docking. Une fois les deux structures préparées, les paramètres sont donnés au logiciel de docking, et ce dernier propose un ou plusieurs modes d'interactions potentiels qui peuvent être ensuite utilisés d'une multitude de façons différentes [1].

Le docking moléculaire a pour but de déterminer le mode d'interaction d'un complexe formé de deux ou plusieurs molécules, en cherchant des orientations dans l'espace et des conformations favorables pour la fixation d'un ligand à un récepteur [3].

Une simulation de docking comprend essentiellement deux étapes : le docking proprement dit et le scoring (Figure IV.2).

- ✓ Le docking (la première étape) est l'étape de sélection, consistant à placer le ligand dans le site actif de la protéine et à échantillonner les conformations, positions et orientations

(poses) possibles, en ne retenant que celles qui représentent les modes d'interactions les plus favorables.

- ✓ Le Scoring (la deuxième) est l'étape de classement, qui consiste à évaluer l'affinité entre le ligand et la protéine et de donner un score aux poses obtenues lors de la phase de docking. Ce score permettra de retenir la meilleure pose parmi toutes celles proposées. [4].

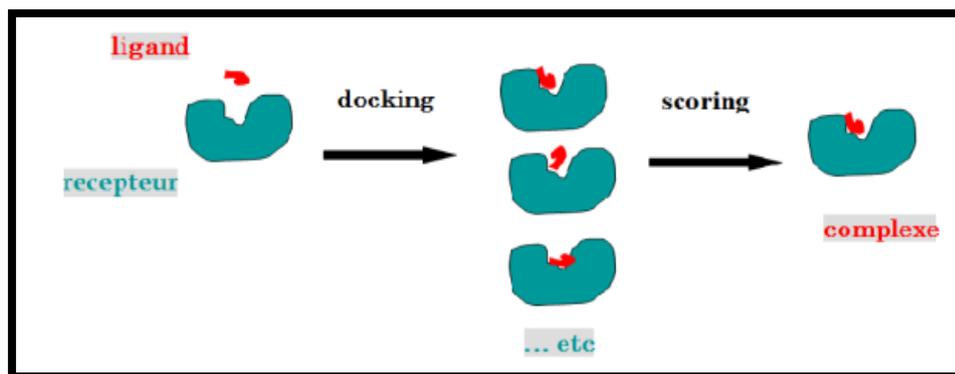


Figure IV.2 : Etapes typique d'un docking moléculaire

IV.1.3 Les programmes d'assemblage moléculaire

Avant de détailler les divers programmes de docking existants, nous allons présenter les étapes de base de ces programmes. Il existe d'ailleurs plusieurs publications de synthèse très intéressantes sur le sujet. Plus de 30 logiciels de docking sont actuellement disponibles.

Le principe des programmes de docking est d'essayer d'assembler deux structures macromoléculaires, protéines et/ou acides nucléiques, afin d'obtenir un complexe [1]. Ce processus se déroule en plusieurs étapes :

IV.1.3.1 Recherche globale des assemblages potentiels

Cette première étape est basée sur la complémentarité de forme entre les partenaires. En règle générale, la plus grosse molécule (le récepteur) est fixée dans l'espace tandis que la plus petite molécule (le ligand) explore l'espace des conformations autour du récepteur. Nous pouvons noter 3 grandes techniques pour effectuer cette recherche.

La plus largement utilisée est la transformée de Fourier rapide ou (Fast Fourier Transform)-FFT. Avec cette méthode, la molécule est discrétisée en petit cube unitaire d'une grille. Cette technique a l'avantage de fournir des solutions très rapidement. De nombreux programmes de docking utilisent cette approche : comme ZDock, SmoothDock, MolFit, DOT, FTDock, 3D Dock, GRAMM ou encore BiGGER . Une autre approche possible est d'utiliser les

harmoniques sphériques pour le calcul de la surface moléculaire et la recherche rapide de complexes potentiels : il s'agit de la transformée de Fourier polaire ou Polar Fourier FT qui est utilisée dans le programme Hex. [1].

IV.1.3.2 Reclassement des complexes trouvés

Toutes ces techniques permettent d'obtenir de quelques centaines à quelques milliers d'assemblages potentiels. Il faut alors pouvoir discriminer ce type de complexe des solutions potentielles. Il est important de rajouter des informations comme le potentiel électrostatique à la surface des partenaires, la prise en compte de la formation de liaisons hydrogène ou des termes de désolvatation dans le but de reclasser les assemblages potentiels. Tous ces paramètres sont le plus souvent inclus dans des fonctions de score. Deux autres phases cruciales du docking sont : l'incorporation d'informations (expérimentales ou issues de la littérature) et le traitement de la flexibilité. Chaque programme peut néanmoins utiliser ces étapes de différentes manières, par exemple, en combinant la recherche de candidats potentiels avec l'utilisation de fonctions de score pour guider les recherches, ou en modélisant la flexibilité avant, après ou pendant cette même phase de recherche [1].

IV.1.4 Protocole Générale de Docking

Les approches utilisées actuellement sont exclusivement calculatoires et évaluées par des outils de visualisation. Ces approches peuvent être décomposées en quatre à cinq phases successives (**Figure IV.3**) :

- ✓ Choix du mode de représentation des protéines (tout atome, pseudo-atome, grille, etc.).
- ✓ Exploration conformationnelle (corps-rigide position/orientation du ligand et/ou flexible position/orientation/forme du ligand).
- ✓ Minimisation de la fonction d'évaluation de l'énergie d'interaction (ou fonction de score) des conformations issues de l'exploration.
- ✓ Regroupement par ressemblances et classification par évaluation plus fine du score, accompagnée d'une étape non automatique d'évaluation visuelle des résultats lorsque le score ne permet pas de discriminer la conformation native des différentes conformations générées.
- ✓ Une étape optionnelle d'affinement des complexes sélectionnés par minimisation ou dynamique moléculaire.

- ✓ Un algorithme de recherche pour explorer les possibilités de modes de liaison, un mécanisme pour placer le ligand dans le site de liaison et une fonction de score pour classer les différents modes de liaison [4].

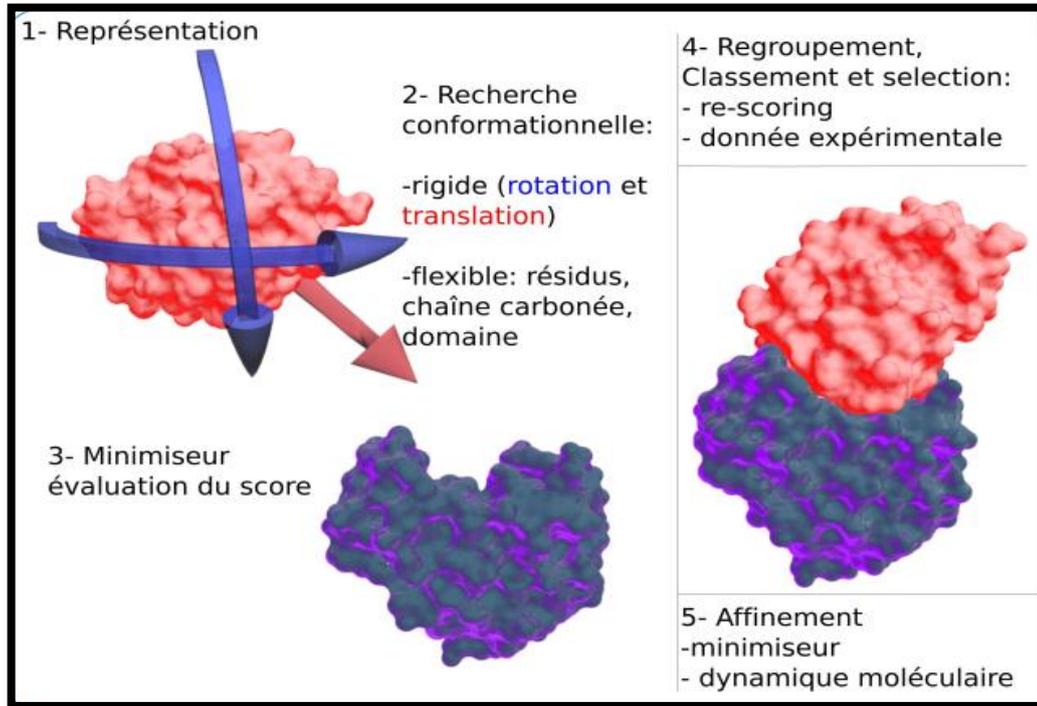


Figure IV.3 : Schéma du protocole général de docking moléculaire

IV.2.1 Etude par docking moléculaire des propriétés cytotoxiques des benzimidazoles

L'interaction entre une protéine et son substrat est la première étape de la plupart des réactions biologiques. Comprendre son mode de fonctionnement c'est définir quels sont les résidus mis en jeu:

- ✓ L'affinité entre deux molécules.
- ✓ Les distances entre les acides aminés du site actif de l'enzyme et ceux des inhibiteurs.
- ✓ L'énergie d'interaction.

De même, la découverte de nouvelles drogues activant ou inhibant l'activité biologique d'une protéine ne peut se faire qu'en prédisant leur affinité respective. C'est dans ce but que des techniques de modélisation moléculaire, regroupées sous le nom d' "Amarrage" ou « Docking moléculaire » ont été développées.

La flexibilité des molécules biologiques résulte de la haute dimensionnalité des systèmes, mais aussi de l'intervention d'un grand nombre d'interactions faibles (Van der Waals, liaison hydrogène, effet hydrophobe, ...). Il est connu qu'elle fait partie intégrante de la dynamique des macromolécules et de leur bon fonctionnement. La négligence de ce paramètre dans les calculs introduit des erreurs, il convient donc de déterminer la flexibilité intrinsèque d'une molécule mais aussi de pouvoir l'introduire, au moins partiellement, dans une procédure de docking. [5-6- 7- 8]

IV.2.1.1 Préparation du récepteur et ligands**IV.2.1.1.1 préparation de l'enzyme**

- ✓ téléchargement des enzymes à partir de la base de données Bookhaven Protein Data Bank
- ✓ Simplification une des deux chaînes.
- ✓ Elimination des molécules d'eau.
- ✓ Elimination des inhibiteurs de la Co-cristallisation

Les enzymes utilisés dans cette étude sont : tyrosine kinase (EGFR), Aromatase cytochrome p450 (MCF7) et proprotéine convertase subtilisine/kexine de type 9 "PCSK9" (HepG2). Ces enzymes ont été téléchargées à partir d'une base de données RCSB Protéine Data Bank [2] avec le code d'accès sont : 1M17, 3EQM et 3GCW respectivement.

Les structures 3D de la tyrosine kinase, Aromatase cytochrome p450 et proprotéine convertase subtilisine/kexine type 9 ont été obtenues par diffraction en rayon X avec une résolution de 2.6 Å, 2.9 Å et 2.7 Å respectivement.

Les structures des enzymes sont associées par autres molécules, à savoir : les molécules d'eau et les cofacteurs. Il est nécessaire d'éliminer ces espèces avant de procéder le docking moléculaire.

- ✓ Le ligand de référence de tyrosine kinase est : 4-anilinoquinazoline
- ✓ Le ligand de référence d'aromatase cytochrome p450 est : androstenedione
- ✓ Aucun ligand de référence n'a été donné pour PCSK9.

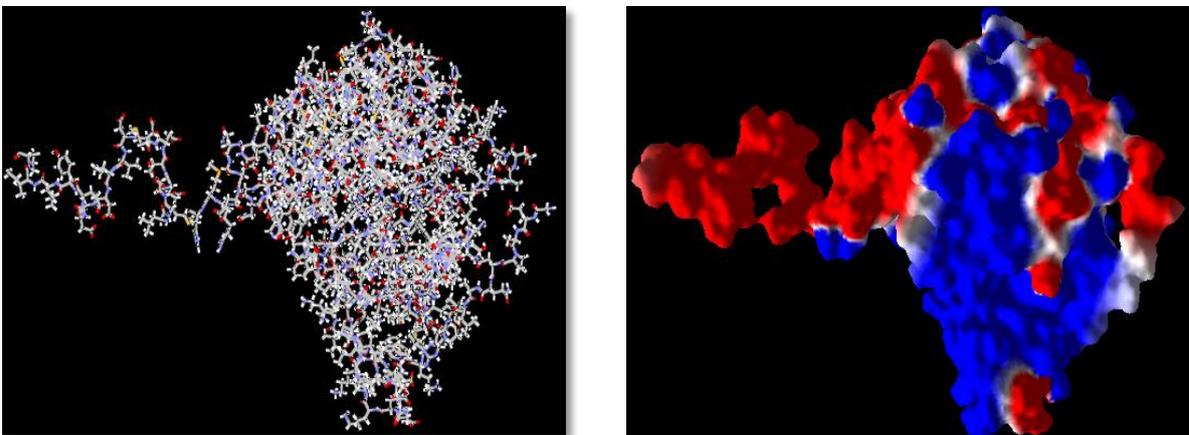


Figure IV.4 : Enzyme tyrosine kinase (1M17)

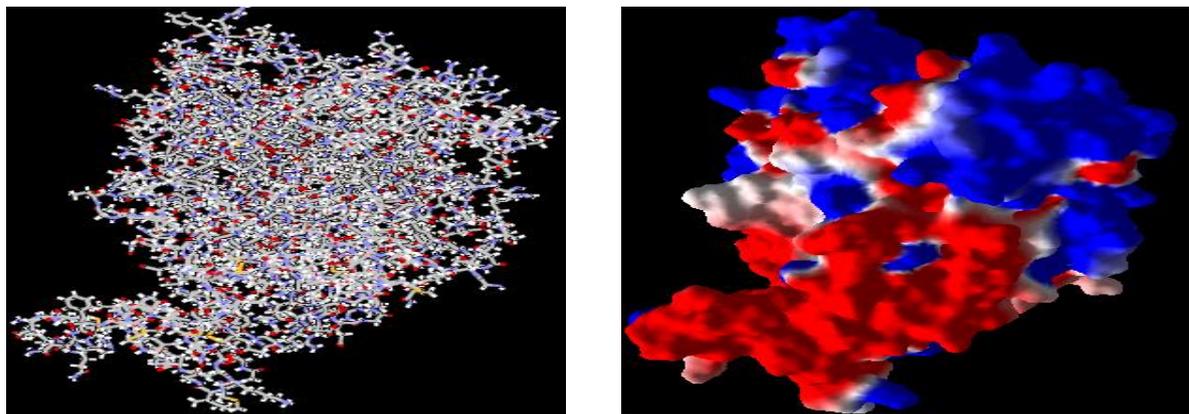


Figure IV.5 : Enzyme proprotéine convertase subtilisine/kexine de type 9 "PCSK9" (3EQM)

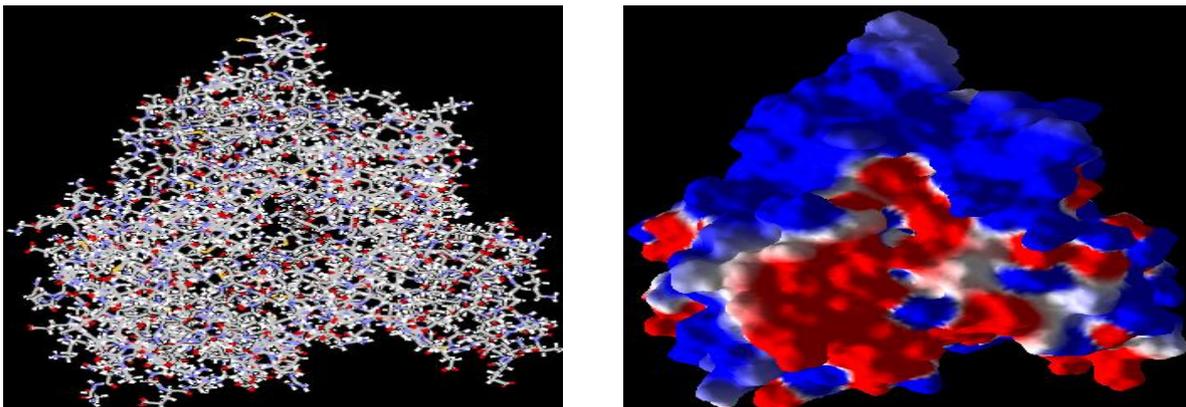


Figure IV.6 : Enzyme aromatase cytochrome p450 (3GCW)

IV.2.1.1.2 Préparation du ligand

Les ligands (inhibiteurs) ont été dessinés par le logiciel Chemdraw8.0.

L'optimisation de la géométrie du ligand a été effectuée à l'aide du champ de forces (MM+) implanté dans le logiciel Hyperchem 7.5 version professionnelle pour déterminer la conformation la plus stable. Le docking moléculaire permet de déterminer les molécules qui ont une grande affinité à lier avec l'enzyme. Cette affinité se traduit par le pouvoir de la molécule d'inhiber l'enzyme.

IV.2.1.2 Résultats de docking moléculaire

Tous les calculs par docking effectués lors de cette étude ont été assistés par Molegro Virtual Docker (MVD 2012.5.5.0, Molegro ApS). Notre choix de ce logiciel est justifié par sa meilleure performance pour étudier et prédire le mode d'interaction ligand-récepteur en comparaison avec d'autres logiciels utilisés.

IV.2.1.2.1 Etapes de docking moléculaire

La première étape, consiste au téléchargement des structures chimiques des cibles à traiter (Enzyme dans notre cas), pour cela il est nécessaire d'aller directement à la Bank PDB et déterminer où sont déposées les structures de ces cibles.

En suit la PDB contient plusieurs milliers de structures protéiques obtenues soit par cristallographie (rayons X), soit par RMN. Si la cible n'est pas encore déposée au niveau de la Bank, et cette dernière contient une protéine avec des séquences similaires, la modélisation par homologie intervient afin de construire la structure 3D de la cible souhaitée.

Après le téléchargement de la cible (PDB), nous utilisons un logiciel de visualisation pour voir avec quels ligands l'enzyme est Co-cristallisé (eau, ligands, ion,...).

La **deuxième étape**, concerne les structures du (ou des) ligand(s) utilisé(s) lors du docking moléculaire. Il y a deux grandes bases de données de structures chimiques des ligands.

- ✓ La première représente ces structures par les programmes d'informatique de modélisation moléculaire, où les différentes structures sont générées par optimisation de géométrie.
- ✓ Le deuxième cas, elles sont obtenues à partir des bases de données comme PDB (Protein Data Bank).

La **troisième étape**, est l'étape de sélection, consistant à placer le ligand dans le site actif de la protéine et à échantillonner les conformations, positions et orientations (poses) possibles, en ne retenant que celles qui représentent les modes d'interactions les plus favorables.

La **quatrième étape**, est l'étape de classement, qui consiste à évaluer l'affinité entre le ligand et la protéine et de donner un score aux poses obtenues lors de la phase de docking. Ce score permettra de retenir la meilleure pose parmi toutes celles proposées.

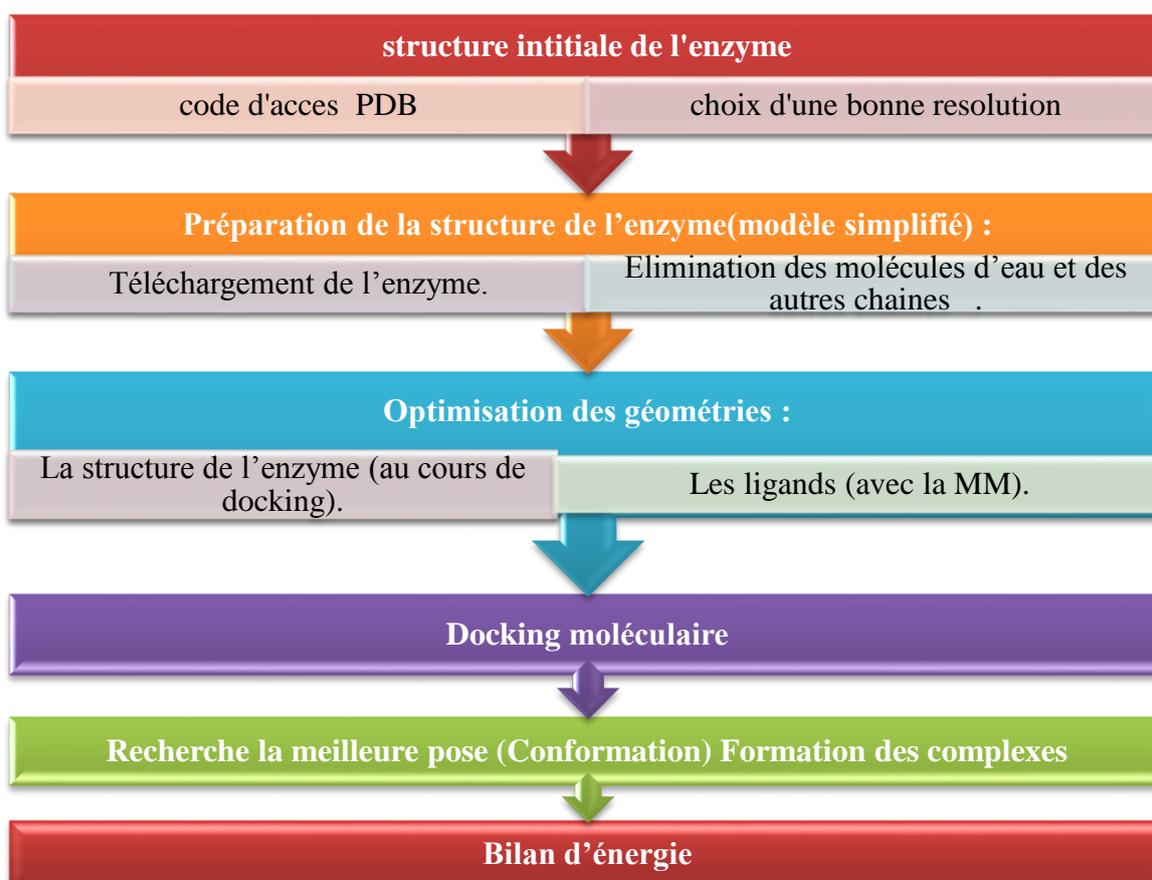


Figure IV.7 : Etapes de Docking Moléculaire

IV.2.1.2.2 Détection des cavités

Pour obtenir de meilleurs sites de liaison potentiels dans chaque enzyme (tyrosine kinase, Aromatase cytochrome p450 et PCSK9), un maximum de cinq cavités a été détecté en utilisant des paramètres par défaut (logiciel). Le tableau IV.1 représente le volume et la surface des 5 cavités détectées pour chaque enzyme. On remarque que le ligand de référence benzimidazole se trouve dans la cavité 1. Ceci nous a permis de faire l'étude sur la cavité 1, cette cavité possède un volume plus important que les autres cavités détectées.

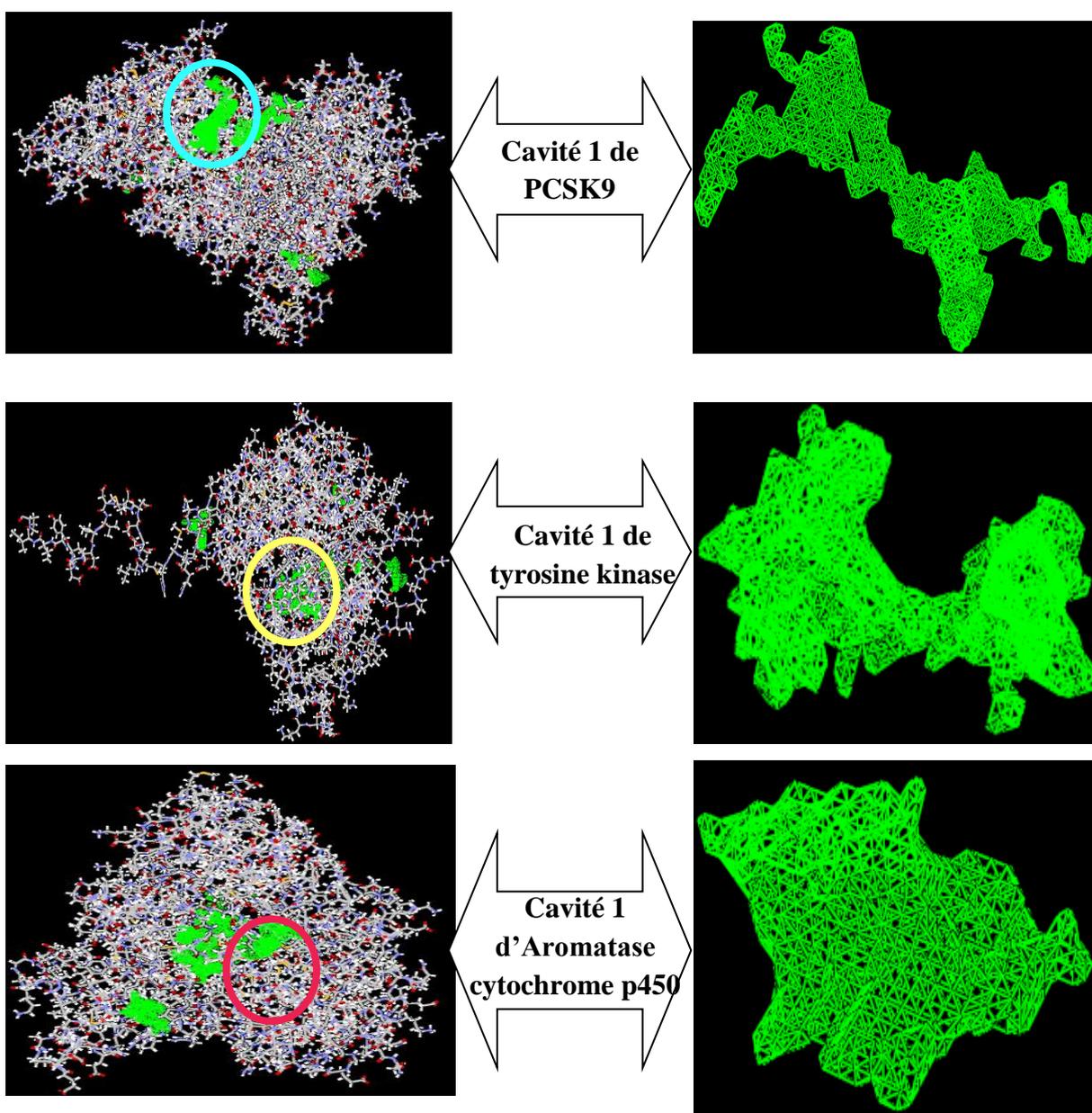


Figure IV.8: Cavités détectées dans les enzymes étudiées.

Tableau IV.1 : Volume et surface des cavités détectées.

	Cavités	Volume (Å ³)	Surface (Å ²)
3GCW	Cavité 1	276.992	849.92
	Cavité 2	40.448	158.72
	Cavité 3	37.888	131.84
	Cavité 4	31.744	139.52
	Cavité 5	24.064	96
1M17	Cavité 1	250.88	800
	Cavité 2	29.184	128
	Cavité 3	22.528	92.16
	Cavité 4	18.432	81.92
	Cavité 5	16.896	75.52
3EQM	Cavité 1	276.992	849.92
	Cavité 2	40.448	158.72
	Cavité 3	37.888	131.84
	Cavité 4	31.744	139.52
	Cavité 5	24.064	96

IV.2.1.2.3 Énergies d'interactions

Les différentes énergies d'interactions entre les enzymes (1M17, 3EQM et 3GCW) et les dérivés de benzimidazole sont calculées au cours de docking moléculaire. Elles sont regroupées dans le tableau IV.2.

Tableau IV.2 : Résultats de docking moléculaire des ligands avec la cavité 1 des enzymes

Ligand	3GCW					1M17					3EQM				
	Mol Dock Score	Interaction	Liaison-H	E (stérique)	E (VdW)	Mol Dock Score	Interaction	Liaison-H	E (stérique)	E (VdW)	Mol Dock Score	Interaction	Liaison-H	E (stérique)	E (VdW)
1	110.103	-118.803	-5.165	8.553	89.131	-102.660	-109.542	-1.207	6.677	87.936	-144.793	-154.160	-9.533	8.096	88.672
2	-121.298	-122.601	-4.314	1.267	49.871	-110.256	-110.488	-1.760	0.084	49.460	-148.141	-151.000	-4.409	1.989	50.093
3	-111.730	-116.997	-1.827	3.789	58.274	-105.069	-112.945	-1.939	6.519	58.639	-146.416	-153.044	-5.473	3.552	52.038
4	-108.115	-113.375	-2.310	2.061	49.145	-107.015	-110.079	-0.598	2.912	49.042	-139.183	-142.732	-2.499	2.090	48.619
5	-120.206	-122.571	-9.701	1.785	48.791	-104.888	-106.703	-1.704	1.716	49.085	-139.927	-145.286	-11.989	4.651	50.765
6	-110.113	-111.031	-1.680	0.891	49.541	-107.758	-109.941	0	1.722	48.841	-141.995	-143.700	-5.786	1.421	48.985
7	-109.019	-114.896	-4.324	5.639	58.021	-106.328	-110.447	-0.292	3.823	55.409	144.821	-149.260	-8.778	4.097	56.473
8	-117.092	-128.356	-5.483	10.239	66.279	-110.764	-122.619	-3.484	9.457	67.618	-154.370	-168.389	-6.275	10.460	91.746
9	-111.245	-106.363	-3.800	-5.582	23.834	-106.448	-106.445	-4.794	-3.235	25.052	-139.643	-137.760	-5.869	-3.816	24.409
10	-116.685	-111.839	-6.776	-6.589	23.320	-105.098	-100.048	0	-5.821	24.009	-138.666	-137.162	-5.298	-2.436	26.875
11	-130.452	-140.432	-1.523	6.561	84.173	-122.565	-138.258	-6.817	8.282	93.911	-164.292	-178.050	-2.241	10.488	97.422
12	-111.840	-115.556	-1.706	2.746	47.237	-109.012	-115.992	-2.497	6.749	50.644	-138.197	-143.188	-3.991	2.873	46.775
13	-114.829	-116.599	-9.886	-1.167	27.909	-96.8534	-94.653	-0.462	-2.330	25.029	-123.117	-120.824	-5.962	-2.592	24.87
14	-115.192	-118.675	-4.588	3.456	50.257	-106.994	-108.785	-1.823	1.634	49.522	-150.525	-156.862	-3.909	2.325	51.806
15	-114.448	-116.459	-4.991	1.640	50.355	-117.010	-119.849	-5.644	2.765	50.086	-153.206	-159.079	-10.496	3.904	53.180
16	-125.768	-134.399	-6.048	7.812	91.162	-119.990	-125.234	-5.878	5.18741	88.9255	-155.090	-160.160	-9.817	5.479	88.796
17	-121.510	-130.347	-4.469	7.416	77.591	-116.665	-126.352	-3.139	7.803	75.735	-168.815	-180.762	-9.089	4.914	74.439
18	-106.620	-110.321	-3.638	3.079	47.054	-100.045	-103.799	-1.463	3.661	47.269	-132.994	-140.273	-3.530	6.019	49.869
19	-109.754	-120.119	-5.053	6.030	61.609	-109.677	-115.011	-4.814	2.990	56.914	-149.597	-156.989	-6.094	6.192	59.016
20	-118.032	-122.813	-3.246	2.734	50.609	-105.612	-106.485	-0.905	0.624	49.807	-147.401	-150.596	-2.496	0.651	49.308
Co crist	-102.259	-112.009	-5.098	9.751	56.970	-111.595	-107.650	-1.635	-5.921	74.390					

Tableau IV.3 : Les termes d'énergie calculés avec la fonction de score MolDock Optimizer

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

Remarque :

- ✓ Lors de la simulation du docking plusieurs poses candidates peuvent être obtenus pour chaque ligand, ici, nous sélectionnons les meilleures poses avec les énergies les plus basses par le biais de la fonction de score basée sur une grille «MolDock Score [Grid]».
- ✓ Mol Dock Score calculé par la somme d'interaction externe et interne de ligand (interaction protéine-ligand) en utilisant le Viewer Moléculaire Virtuelle (1.2.0).
- ✓ Énergie d'interaction totale entre la pose et les molécules cibles (s).

D'après les résultats obtenus, on remarque pour que les complexes formés par les ligands L₁₁, L₁₆, L₁₇ et L₂ possèdent une énergie (MolDock Score) le plus faible possible (-130.452,-125.768,-121.510 et-121.298 respectivement) par rapport aux autres ligands pour l'enzyme PCSK9.

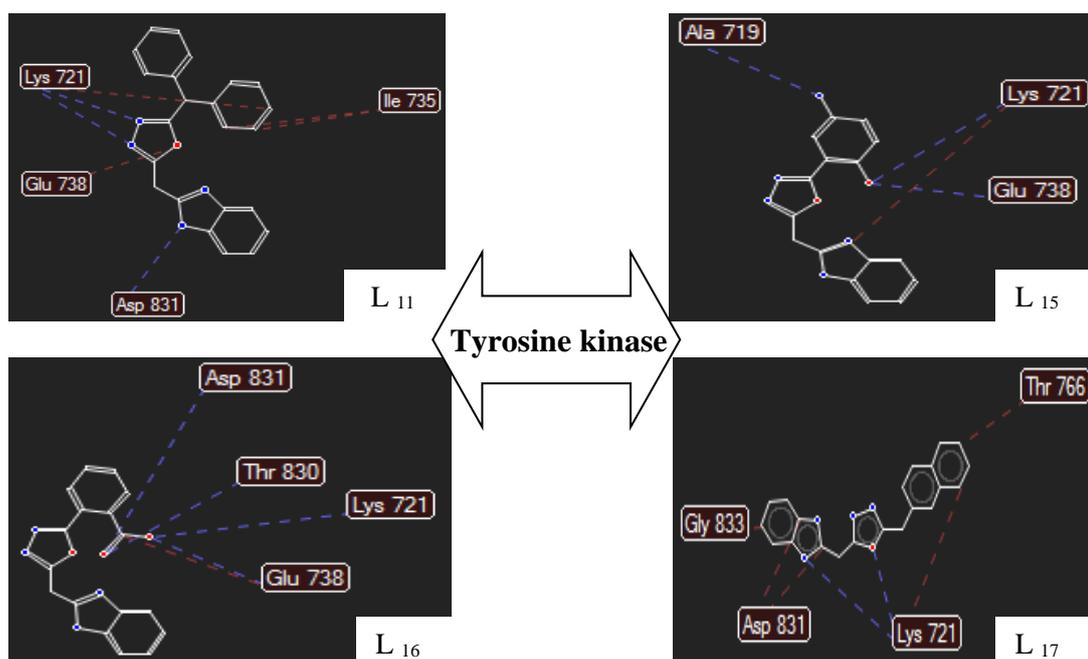
Pour l'enzyme Tyrosine kinase on remarque que les complexes formés par les ligands L₁₁, L₁₆, L₁₅ et L₁₇ possèdent des énergies les plus faibles (-122.565,-119.990, -117.010 et -116.665 respectivement) par dans l'enzyme aromatasé cytochrome p450 les complexes formés par les ligands L₁₇, L₁₁, L₁₆ et L₈ possèdent des énergies les plus faibles (-168.815,-164.292, -155.090 et -154.370 respectivement).

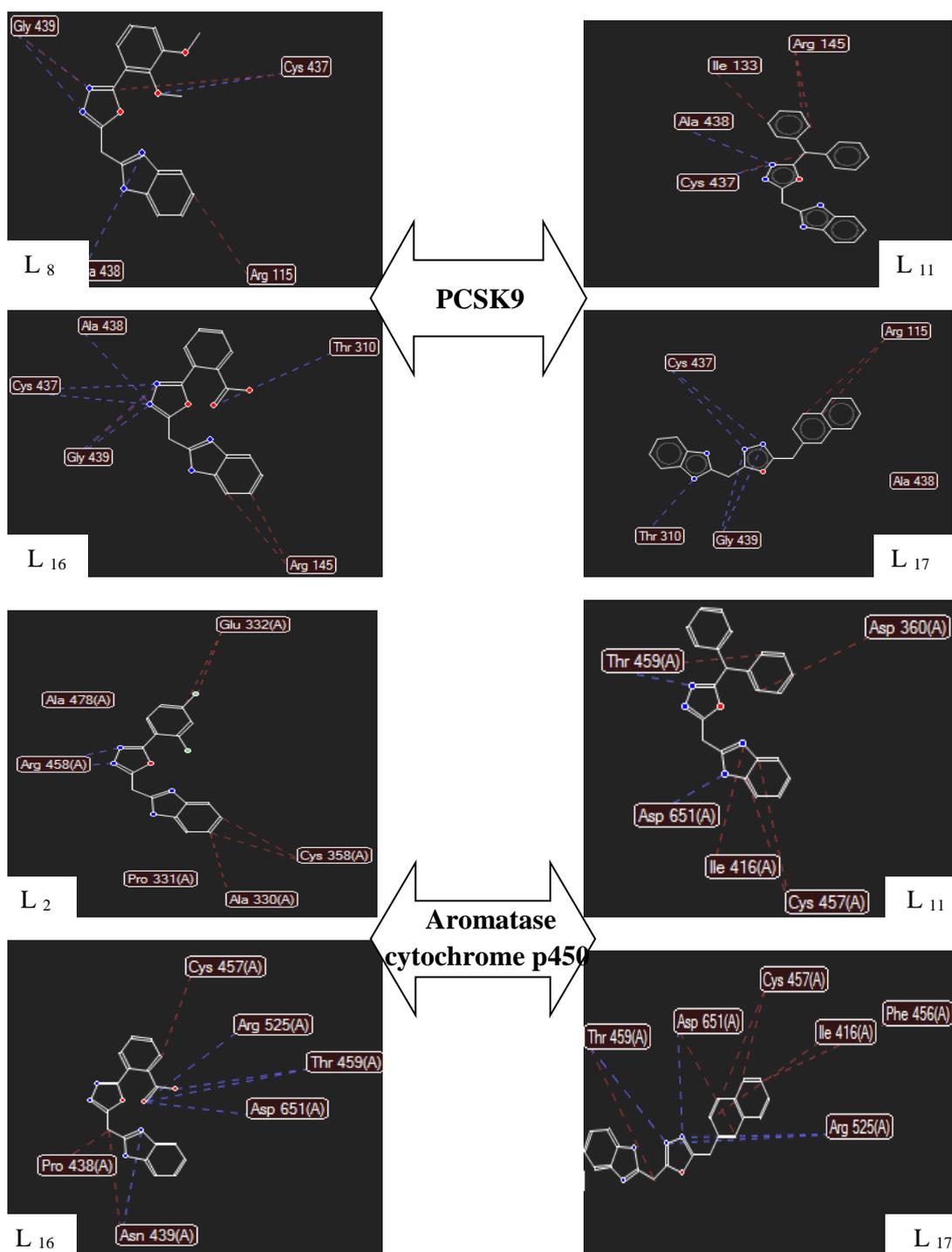
Tableau IV.4 : Résultats des énergies des meilleurs complexes formés pour chaque enzyme.

	Ligand	Mol Dock Score	Interaction	Liaison-H	E (stérique)	E (VdW)
3GCW	11	-130.452	-140.432	-1.523	6.561	84.173
	16	-125.768	-134.399	-6.048	7.812	91.162
	17	-121.510	-130.347	-4.469	7.416	77.591
	2	-121.298	-122.601	-4.314	1.267	49.871
	Co crist	-102.259	-112.009	-5.098	9.751	56.970
1M17	11	-122.565	-138.258	-6.817	8.282	93.911
	16	-119.990	-125.234	-5.878	5.18741	88.9255
	15	-117.010	-119.849	-5.644	2.765	50.086
	17	-116.665	-126.352	-3.139	7.803	75.735
	Co crist	-111.595	-107.650	-1.635	-5.921	74.390
3EQM	17	-168.815	-180.762	-9.089	4.914	74.439
	11	-164.292	-178.050	-2.241	10.488	97.422
	16	-155.090	-160.160	-9.817	5.479	88.796
	8	-154.370	-168.389	-6.275	10.460	91.746

D'après les résultats obtenus, on peut conclure que les interactions mises en jeu entre les résidus de site actif et les ligands peuvent former des complexes stables.

La figure IV.9 représente les différentes interactions entre les résidus de site actif de chaque enzyme et les 4 ligands L2, L11, L16 et L17.





-----: Interactions stériques

-----: Liaisons hydrogène

Figure IV.9 : Interaction hydrogène et stérique entre les ligands et les résidus du site actif de la cavité 1.

Dans la figure **IV.9**, les interactions stériques (rouge) et hydrogènes (bleu) conduisant la formation des complexes stables.

D'après les résultats obtenus, on peut conclure que les interactions mises en jeu entre les résidus de site actif et les ligands peuvent former des complexes stables.

Les distances mesurées entre ces résidus et les quatre ligands (L2, L11, L16, L17) sont regroupés dans les tableaux suivants :

Tableau IV.5 : Distances de liaison hydrogènes entre les résidus de site actif et les ligands.

Tyrosine kinase			PCSK9			Aromatase cytochrome p450		
Résidus	L ₁₁ -Atomes	Liaison-H	Résidus	L ₈ -Atomes	Liaison-H	Résidus	L ₂ -Atomes	Liaison-H
Asp 831	N	3.01	Ala 438	N	3.23	Arg 478	N	2.60/3.35
Lys 721	N	3.00/3.01	Cys 437	O	2.75		L₁₁-Atomes	
	L₁₅-Atomes		Gly 439	N	3.04/2.87	Thr 459	N	3.40
Ala 719	N	3.09	Cys 437	N	3.10	Asp 651	N	2.64
Glu738	O	2.61	Cys 437	C	2.38		L₁₆-Atomes	
Lys 721	O	3.11		L₁₁-Atomes		Asp 651	O	3.01
	L₁₆-Atomes		Ala 438	N	3.10	Arg 525	O	2.79
Asp 831	O	3.18	Cys 437	N	3.17	Thr 459	O	3.40/2.79
Thr 830	O	2.60		L₁₆-Atomes		Asn 439	O	3.13
Lys 721	O	2.98	Ala 438	N	3.09		L₁₇-Atomes	
Glu 738	O	2.60	Thr 310	O	3.10	Thr 459	N	3.15
	L₁₇-Atomes		Cys 437	N	2.81/3.10	Asp 651	N	3.11
Lys 721	N	3.04	Gly 439	N	3.30/2.60	Arg 525	N	2.84/3.11
Lys 721	O	3.31		L₁₇-Atomes				
			Cys 437	N	2.60/2.60			
			Thr 310	N	2.94			
			Gly439	N	3/2.87			

Les distances entre les résidus du site actif et les ligands varient entre 2.38 Å et 3.40 Å, dans ce cas on peut remarquer que d'après Anne Imberty et col, les interactions ayant des distances comprises entre 2.5Å et 3.1Å sont considérés comme fortes, celles ayant des distances comprises entre 3.1Å et 3.55Å sont supposées moyennes et quand leurs distances sont supérieures à 3.55Å, elles sont considérées faibles [9].

Le tableau **IV.6** présente les distances entre les distances des interactions stériques entre les enzymes étudiées et les ligands L11, L16 et L17.

La formation d'un complexe stable dépend de la fixation du ligand dans le site active.

D'après les résultats obtenus, montrent qu'il existe plusieurs interactions entre les ligands et les résidus du site actif, ce qui signifie la formation des complexes stables, et par la suite, une meilleure fixation de ces ligands au niveau du site actif.

Tableau IV.6 : Distances des interactions stériques entre les résidus de site actif et les ligands.

Tyrosine kinase			PCSK9			Aromatase cytochrome p450		
Résidus	L ₁₁ -Atomes	Interaction stérique	Résidus	L ₈ -Atomes	Interaction stérique	Résidus	L ₂ -Atomes	Interaction stérique
Lys 721	C	3.08	Arg 115	C	2.99	Cys 358	C	3.04/3.16
Ile 735	C	3.17/3.17	Gly 439	N	3.13	Ala 330	C	3.12
Gly 439	O	2.88	Cys 437	C	3.02	Glu 332	Cl	2.88/2.89
	L₁₅-Atomes			L₁₁-Atomes			L₁₁-Atomes	
Lys 721	N	3.13	Ile 133	C	3.05	Thr 459	C	3.10
	L₁₆-Atomes		Cys 437	C	3.16	Asp 360	C	3.09
Gly 439	C	3.15	Arg 145	C	2.87/2.98	Ile 416	N	2.35
	L₁₇-Atomes			L₁₆-Atomes		Cys 457	c	3.07/3.14
Asp 831	C	3.09	Gly 439	N	3.12		L₁₆-Atomes	
Lys 721	C	3.03	Arg 145	C	3.02/2.80	Pro 331	C	3.06
Thr 766	C	3.12	Cys 437	O	2.90	Asn 439	C	2.83
				L₁₇-Atomes		Cys 437	C	2.83
			Arg 115	C	2.99/3.15		L₁₇-Atomes	
						Thr 459	C	3.09
						Asp 651	C	2.68
						Cys 457	C	3.12/3.07
						Phe 416	C	3.15
						Ile 456	C	2.38

IV.2.2 Conclusion

Une étude par docking moléculaire a été utilisée sur une série de dérivés de benzimidazole et trois enzymes EGFR, MCF7, PCSK9. Cette étude permet de prédire les meilleurs ligands qui donnent une bonne activité cytotoxique, ainsi que, les différentes interactions entre ces ligands et les résidus des sites actifs des enzymes étudiées.

Parmi les meilleurs composés en terme d'énergie d'interactions « top-scoring » nous avons sélectionné 4 ligands (L11 ,L15 ,L16 et L17) pour les enzymes tyrosine kinase ,(L2, L11 ,L16 et L17) pour PCSK9 et (L8 ,L11 ,L16 et L17) pour aromatasé cytochrome p450. L'analyse des résultats obtenus permet de déterminer deux types d'interactions responsables de l'affinité de ces ligands à ces enzymes, les liaisons hydrogènes et interaction de VDW. Cette affinité traduit la stabilité du ligand étudié dans le site actif et favorisent leur fixation ce qui aurait pour résultat une inhibition potentielle des enzymes cibles.

Références bibliographique

- [1] : SARI MOHAMED, Souad. Etude de l'Interaction Protéine-Protéine par Modélisation Moléculaire. Application aux Processus Biologiques. Diss. 2013.
- [2] : LEBBAD, Fatima. Etude par modélisation moléculaire des mécanismes de complexation. Diss. Université Aboubeker Belkaid de Tlemcen.
- [3] : Corbeil, Christopher R., Pablo Englebienne, and Nicolas Moitessier. "Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0." *Journal of chemical information and modeling* 47.2 (2007): 435-449.
- [4] : Férey, Nicolas, et al. "Docking de protéines en réalité virtuelle Une approche hybride et multimodale." *Technique et science informatiques* 28.8 (2009): 983-1015.
- [5] : Carlson, Heather A., and J. Andrew McCammon. "Accommodating protein flexibility in computational drug design." *Molecular pharmacology* 57.2 (2000): 213-218.
- [6] : Carlson, Heather A. "Protein flexibility and drug design: how to hit a moving target." *Current opinion in chemical biology* 6.4 (2002): 447-452.
- [7] : Carlson, Heather A. "Protein flexibility is an important component of structure-based drug discovery." *Current Pharmaceutical Design* 8.17 (2002): 1571-1578.
- [8] : Teague, Simon J. "Implications of protein flexibility for drug discovery." *Nature reviews Drug discovery* 2.7 (2003): 527.
- [9] : Imberty, Anne, et al. "Molecular modelling of protein-carbohydrate interactions. Docking of monosaccharides in the binding site of concanavalin A." *Glycobiology* 1.6 (1991): 631-642.

Conclusion générale

Dans ce travail, nous avons appliqué les méthodes de la chimie informatique sur d'une série de benzimidazole.

Ce travail est une étude propriétés de la relation structure-activité d'une série de benzimidazole bioactives que basé sur la méthode QSAR, consiste à trouver une corrélation entre une activité biologique mesurée pour un panel des composés et certains descripteurs moléculaires.

L'étude qualitative de la relation structure-propriétés a été réalisée sur les dérivés. Les molécules utilisées dans cette étude possèdent des propriétés pharmacologiques. La nature des groupements qui se lient sur le noyau de base des molécules étudiées affecte sur leurs propriétés physicochimiques et par conséquent sur leurs propriétés pharmacologiques. L'étude quantitative de la relation structure-propriétés (QSAR) a été effectuée sur une série de dérivés de benzimidazole. Trois modèles QSAR ont été établis en utilisant la méthode de régression multilinéaire (MLR).

Le pouvoir prédictif des modèles obtenus a été confirmé par la méthode de validation croisée LOO. Une forte corrélation a été observée entre les valeurs expérimentales et prédites des activités biologiques, ce qui indique la validité et la qualité des modèles QSAR obtenus.

Le docking moléculaire a été procédé sur les 20 molécules obtenues à partir de screening virtuel. Cette étude permet de déterminer les molécules qui ont une grande affinité à se lier avec les trois enzymes : tyrosine kinase (EGFR), Aromatase cytochrome p450 (MCF7) et proprotéine convertase subtilisine/kexine de type 9 "PCSK9" pour former des complexes stables, ainsi que, les types des interactions entre les ligands et les résidus de site actif de l'enzyme.

L'étude montre que les ligands L11, L16 et L17 sont probablement les meilleurs ligands pour inhiber le fonctionnement les trois enzymes. Ces ligands possèdent les plus basses énergies par rapport aux autres et en plus ils ont des fortes interactions avec les résidus de site actif es trois enzymes, et par conséquent ces ligands pouvant former des complexes stables.

Résumé

Une étude des relations structure-activité cytotoxique a été effectuée sur vingt dérivés de benzimidazole envers trois lignées cellulaires : EGFR, MCF7 et HepG2. Trois modèles QSAR ont été développés en utilisant la méthode de régression multilinéaire (RML).

Le pouvoir prédictif des modèles obtenus ont été confirmé par la méthode de validation croisée LOO. Une forte corrélation a été observée entre les valeurs expérimentales et prédites des activités biologiques, ce qui indique la validité et la qualité des modèles QSAR obtenus. Une étude des propriétés « drug-like » montre que la série de benzimidazole étudiée ont vérifiées les règles de Lipinski et Veber et sont caractérisés par leurs biodisponibilité pour l'administration par voie orale. Enfin, une étude par docking moléculaire a été effectuée sur une série de la même série de benzimidazole. Les résultats obtenus montre que les ligands L11, L16 et L17 donnent des fortes interactions avec les résidus de site actifs de trois enzymes EGFR, MCF7 et HepG.

Mots-clés : Benzimidazole, , Activité cytotoxique, QSAR, Docking moléculaire , RML

Abstract

A study of the cytotoxic structure-activity relationship was carried out on twenty benzimidazole derivatives against three cell lines: EGFR, MCF7 and HepG2. Three QSAR models have been developed using the multilinear regression (RML) method.

The predictive power of the obtained models was confirmed by the LOO cross validation method. A strong correlation was observed between the experimental and predicted values of biological activities, indicating the validity and quality of the QSAR models obtained. A study of "drug-like" properties shows that the benzimidazole series studied have verified the rules of Lipinski and Veber and are characterized by their bioavailability for oral administration. Finally, a molecular docking study was performed on a series of the same series of benzimidazole. The results obtained show that ligands L11, L16 and L17 give strong interactions with the active site residues of three EGFR, MCF7 and HepG enzymes.

Keywords: Benzimidazole, Cytotoxic activity, QSAR, Molecular Docking, RML