

République Algérienne Démocratique et Populaire  
Ministère de l'enseignement supérieur et de la recherche scientifique

Université Saleh Boubnider Constantine III

Faculté de médecine  
Département de pharmacie



Mémoire de fin d'études en vue de l'obtention du diplôme de Docteur  
en Pharmacie

## Thème

**Conception in silico de nouveaux inhibiteurs anti-EGFR  
en vue d'une application anticancéreuse**

**Réalisé et présenté par :**

ABDELLI Khadidja  
AGOUDJIL Naoufel  
BOURENANE Amina

**Encadré par :**

Dr. ALLAOUA Amina

**Membres du jury :**

**Présidente :** Dr. HAMLAOUI Ikram

**Examineur :** Dr. GUEROUI Mehdi

Année universitaire : 2020/2021

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

فَالْوَيْلُ لِمَنْ حَانَ أَلْعَامُ لَنَا الْإِمَامُ عَلِيُّنَا  
إِنِّي أَنْبَأْتُ الْعَالِمَ الْحَكِيمَ

صِدْقَ اللَّهِ الْعَظِيمِ

## **REMERCIEMENTS**

En tout premier lieu, nous remercions le bon Dieu, tout puissant pour toute la volonté, le courage et la patience qu'il nous a donné pour l'achèvement de ce travail, ainsi que l'audace pour dépasser toutes les difficultés.

A notre encadreur Docteur ALLAOUA Amina :

**Maître-assistante Hospitalo-universitaire en Chimie Thérapeutique**

Pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter notre réflexion.

Aux membres de jury

Dr. HAMLAOUI Ikram et Dr.GUEROUI Mehdi

Nous vous remercions pour l'intérêt que vous avez porté à notre travail en acceptant de juger notre mémoire et de l'enrichir par vos propositions.

Ce travail représente la fin d'un parcours des études riches en expériences humaines et académiques, il nous apparaît opportun de remercier chaleureusement toute l'équipe des professeurs du département de pharmacie, faculté de médecine de Constantine, qu'ont partagé leurs savoirs avec nous au cours de ces six années d'étude, avec tout notre respect et notre profonde gratitude.

## Dédicace

Je tiens c'est avec grande plaisir que je dédie ce Modest travail :

A l'être le plus cher de ma vie ma mère **DEGHIMA Zahia**

Que je fasse ou que je dise je ne serais point vous remercier comme il se doit, votre affection me couvre, votre surveillance me guide et vos prières été ma source de force pour affrété les obstacles

A celui qui ma fait de moi la femme plaine de morale et principes, vous êtes toujours a mes cotés pour me soutenir et m'encourager :a mon cher père **Mostefa**

A ma sœur **Khaoula**

Tu es la plus douce personne que je connu, la plus sage et plaine de talent et a ton mari **MADI Yakoub** qui ma aider pour terminer ce travail avec tous ce qu'il peut

A mon frère **Idriss** la source de ma confiance au soi le plus charment

A mon frère **Med Amine** le dernier de la lignée et la source de notre joie

A la mémoire de mon grand père **Bouazza**

A mes grands parents **DEGHIMA Amar** , **NASSRI Erabeh** et **GUENDOUIZI Hafsa** que dieu vous protège et te préserve

Et toute ma famille

A mes amies M. **Nour El Houda**, **Hayam** , **Iman**, **Nousseiba**, **Meriem** et **Farah** : vous étés toujours présentes et prêtes pour m'aider, les bonne souvenir de cite universitaires sont a cause de vous

A l'équipe du laboratoire centrale de l'hôpital militaire :

En particulier **DR Zouiten**

*ABDELLA Khadidja*

## **Dédicace**

*Je dédie ce travail :*

*A mes très chers parents,*

*Qui n'ont jamais cessé, de formuler des prières à mon égard de me soutenir et de m'épauler pour que je puisse atteindre mes objectifs. Aucun hommage ne pourrait être à la hauteur de vos sacrifices, de l'amour et de l'affection dont vous n'avez jamais cessé de m'entourer toutes au long de ces années d'étude.*

*A celle que j'aime beaucoup et qui m'a soutenue tout au long de ce projet à la femme de ma vie : ma fiancée Rayane.*

*A mes chers frères (Abderraouf, Mahia, Abdennour, Zineddine et Atef) qui je le sais, ma réussite est très importante à leurs yeux. Que dieu vous garde pour moi.*

*A mes meilleurs amis (Wail, Samy, Okba et Mouadh) avec qui j'ai partagé de bons moments, que notre amitié dure plus longtemps que l'éternité.*

*A ma belle-famille, et mes belles sœurs.*

*Aux petits (Ayham, Nourcine et Belkis).*

*Et enfin à ceux qui sont présents dans mon cœur, ceux qui j'aime et ceux qui m'aiment.*

***Naoufel***

## **Dédicace**

Je dédie ce travail de fin d'études :

A mes très chers parents, qui m'ont accompagné durant tous mon cursus scolaire jusqu'à ce jour, qui ont investis leurs confiances en moi, et qui m'ont entouré et orienté tout au long de ma vie, Que le bon DIEU les garde et les protège.

A mes frères et mes sœurs, les mots ne suffisent pour exprimer l'attachement et l'affection que je porte pour vous, je vous aime beaucoup.

A toute ma famille pour leur soutien.

A mes adorables amies et a tous ceux qui ont contribué à mon bien être et ceux qui mon soutenu de près ou de loin

**Amina.**

# TABLE DES MATIERES

Liste des Abréviations .....	i
Liste des Annexes.....	iv
Liste des Equations.....	v
Liste des Figures.....	vi
Liste des Tableaux.....	ix
INTRODUCTION :.....	1
Partie bibliographique	
CHAPITRE 1 : CANCER & TYROSINE KINASE	
I. CANCER :.....	3
I.1. DEFINITION : .....	3
I.2. EPIDEMIOLOGIE : .....	3
I.3. CARACTERISTIQUES DU CANCER: .....	3
I.4. PROGRESSION DU CANCER.....	5
I.4.1. Cancérogenèse.....	7
I.4.2. Tumorigenèse .....	8
I.5. SIGNALISATION CELLULAIRE ET CANCER.....	9
I.6. TRAITEMENT DU CANCER.....	11
I.6.1. Avances thérapeutiques.....	12
II. TYROSINE KINASE .....	13
I.1. GENERALITE .....	13
II.1. CLASSIFICATION DES PK.....	13
II.1.1. Les PTK cytosoliques (CTK).....	14
II.1.2. Les PK membranaires ou Récepteur Tyrosine Kinase (RTK).....	15
II.2. TK ET CANCER.....	21
CHAPITRE 2 : EGFR	
I. GENERALITE .....	22

II.	BIOLOGIE MOLECULAIRE, CELLULAIRE ET STRUCTURALE.....	22
II.1.	Structure primaire et organisation .....	22
II.2.	Structure tridimensionnelle .....	24
II.3.	Rôles physiologiques .....	26
II.4.	Activation des récepteurs .....	28
III.	Ciblage thérapeutique des récepteurs ERBB .....	30
III.1.	Anticorps.....	31
III.2.	Inhibiteurs de tyrosine kinase .....	32
IV.	MUTATIONS DE L'EGFR.....	33
IV.1.	Conséquences fonctionnelles des mutations EGFR.....	35
CHAPITRE 3 : CRIBLAGE VIRTUEL		
I.	GENERALITE : .....	36
II.	MODELES PHARMACOPHORIQUES.....	37
III.	APPROCHES DE LA CONCEPTION DE MEDICAMENTS :.....	38
III.1.	Approche <<Ligand-based >> .....	38
III.2.	Approche << Structure-based>>.....	38
III.2.1.	Approche basée sur le récepteur.....	39
III.2.2.	Approche basée sur le complexe ligand-récepteur :.....	39
IV.	DOCKING MOLECULAIRES : .....	42
IV.1.	Docking ligand rigide .....	42
IV.2.	Docking « semi-flexible » .....	43
IV.3.	Docking ligand flexible .....	45
IV.3.1.	Algorithmes de recherche .....	45
V.	SCORING : .....	53
V.1.	Fonctions de score .....	53
V.2.	Autres paramètres importants à prendre en compte pour réaliser un docking .....	56
V.2.1.	Importance des structures de protéines .....	56



V.2.2. Importance du solvant.....	57
Partie pratique	
OBJECTIF .....	58
MATERIEL ET METHODES	
I. Matériel : .....	59
I.1. Microordinateur : .....	59
I.2. logiciels: .....	60
I.2.1. Biovia Discovery Studio 2019 : .....	60
I.2.2. AutoDock : .....	60
I.2.3. AutoDock Vina : .....	61
I.2.4. Excel : .....	61
I.3. Serveur Online .....	62
I.3.1. SwissADME : (98) .....	62
I.3.2. PreADMET : (99).....	62
I.4. Banques des Données : .....	62
I.4.1. « PDB » (Protein Data Bank) : (100).....	62
I.4.2. « PubChem »: (101).....	62
II. Méthodes : .....	63
II.1. Docking moléculaire (D.M) : .....	63
II.1.1. Préparation du récepteur et ligands: .....	63
II.1.2. Recherche des similaires .....	64
II.1.3. Docking moléculaires: .....	65
II.1.4. Visualisation 2D : .....	66
II.1.5. Filtrage ADME/tox : .....	66
II.2. Tests de fiabilité du programme AutoDock Vina : .....	69
II.2.1. Analyse visuelle : .....	69
II.2.2. Coefficient de corrélation.....	69

II.2.3. RMSD :	69
<b>RESULTATS ET DISCUSSION</b>	
I. Tests de fiabilité du programme AutoDock Vina :	70
I.1. RMSD :	70
I.2. Analyse visuelle :	70
I.3. Coefficient de corrélation (r) :	72
II. DOCKING MOLECULAIRE :	74
II.1. La recherche bibliographique	74
II.2. Recherche des similaires au gefitinib:	81
II.3. Etude des interactions 4WKQ-Gefetinib :	85
II.4. Etude des interactions 4WKQ-les 12 meilleures molécules	88
II.4.1. Visualisation des interactions :	88
II.4.2. Conclusion de visualisation :	99
II.5. Résultat filtrage ADME/tox :	100
II.5.1. Propriétés physico-chimiques :	100
II.5.2. Propriétés pharmacocinétiques :	102
II.5.3. Toxicité:	103
CONCLUSION ET PERSPECTIVES :	105
<b>Bibliographie :</b>	
<b>Annexes</b>	

## Liste des Abréviations

ADAM : a disintegrin and métallo-proteas

ADME/Tox : Absorption Distribution Métabolisme Excrétion/toxicité

ADN : Acide désoxyribonucléique

ATP : Adénosine-Triphosphate

BBB permeability : Blood–Brain Barrier perméabilité

COSMIC: catalog of somatic mutations in cancer

CR1,CR2 : cysteins -rich 1 et 2

CYP 1A2 et 2C9 et 3A4 : Cytochrome P1A2 et 2C9 et 3A4

DS : discovery studio visualizer

EGF : epidermal growth factor

EGFR : epidermal growth factor receptor

EMM : énergie de la force mécanique moléculaire

ERBB 2 et ERBB 3 et ERBB 4 : homologues de la même famille d'EGfr

GI absorption : Gastro-intestinal absorption

GIST: tumeurs stromales gastro-intestinales

GRB2 : growth factor receptor bound protein 2

HER-1,HER-2,HER-3,HER-4 : human epidermal growth factor receptor 1 et 2

et 3 et 4

hERG : The human ether-a-go-go-related gene

IC 50 : Inhibitory Concentration 50.

IGF1 : insulin growth factor 1

ITK : inhibiteur tyrosine kinase

IUPAC : international union of pure and applied chemistry

Ki : Constante d'inhibition.

KRAS : v-ki-ras 2 kisten rat sarcoma viral oncogene homolog

Log p : Coefficient de partage Eau/Octanol

LMC : leucémie myéloïde chronique

MAP : mitogen activated protein

MET : Mesenchymal Epithelial Transition

nRTK: non receptor tyrosine kinase

OMS : Organisation mondiale de la santé

PDB : Protein Data Bank.

Pdbqt : Protein Data Bank, Partial Charge (Q), & Atom Type (T)

PK: Protéine(s) Kinase

Pm : poids moléculaire

PTK : Protéine(s) Tyrosine Kinase

QSAR : relation structure activité quantitative

RMN : résonance magnétique nucléaire

RMSD : Root Mean Square Deviation (écart quadratique moyen)

RNA: ribonucleic acid

RTK: récepteur tyrosine kinase

R&D : Recherche et Développement

SH2,SH3 : Domaine d'homologie à Src de type 2, de type 3

SHC: Src homology 2 domain containing

Sdf : Spatial Data File

siRNA :short interfering ribonucleic acid

TGFa : transforming growth factor a

TKD : domaine tyrosine kinase

TPSA : Surface polaire

VADS : voies aérodigestives supérieures

VEGF: Vascular Endothelial Growth Factor

VEGFR: Vascular Endothelial Growth Factor Receptor

3D : Trois dimensions

$\Delta G$  : Energie d'interaction

## **Liste des Annexes**

**Annexe N° I :** Les scores des similaires obtenus suite au docking avec l'EGFR (4WKQ) :

## Liste des Equations

Équation 1: Probabilité P d'acceptation d'une nouvelle conformation ( <i>avec <math>E_{new}</math> : énergie de la nouvelle conformation, <math>E_{old}</math> : énergie de l'ancienne conformation, <math>k</math> : constante de Boltzmann, <math>T</math> : température de simulation</i> ) (70).....	48
Équation 2 : <i>Equation du mouvement de Newton pour un système atomique (<math>F</math> : force, <math>m</math> : masse, <math>a</math> : accélération)</i> .....	51
Équation 3 : <i>Calcul de la force (<math>F</math>) de chaque atome (<math>E</math> : énergie potentiel, <math>r</math> : distance entre les deux positions)</i> .....	51
Équation 4: <i>Calcul des positions atomiques à chaque intervalle de temps (<math>d^2r_i/dt^2</math>), avec <math>F</math> : force et <math>m</math> : masse atomique)</i> .....	52
Équation 5 : <i>Calcul de l'énergie potentielle d'un système (EMM) par les champs de force de mécanique moléculaire. Avec <math>K_r</math>, <math>K_\theta</math> et <math>K_\varphi</math> : facteurs de pénalité pour les liaisons, les angles et les angles dièdres ; <math>r</math> et <math>r_{ref}</math> : longueurs des liaisons mesurées et de référence <math>\vartheta</math> et <math>\vartheta_{ref}</math> : valeurs des angles mesurées et de référence ; <math>\phi</math> et <math>\phi_{ref}</math> : valeurs des angles dièdres mesurées et de référence ; <math>A_{ij}</math> et <math>B_{ij}</math> : constantes attractives et répulsives mesurées expérimentalement ; <math>R_{ij}</math> : distance entre les atomes <math>i</math> et <math>j</math> ; <math>q_i</math> et <math>q_j</math> : charges des atomes <math>i</math> et <math>j</math> ; <math>\epsilon</math>: constante diélectrique</i> (79) .....	53
Équation 6 : <i>Calcul de l'énergie libre de liaison (<math>\Delta G_{bind}</math>) prenant en compte uniquement la contribution enthalpique évaluée par les potentiels de Lennard-Jones et de Coulomb</i> ...	54
Équation 7 : <i>Forme simplifiée d'une fonction de score empirique. (Avec <math>\Delta G_{binding}</math> : l'affinité de liaison, <math>f_i</math> correspond à un terme d'une interaction <math>i</math> entre le ligand et la protéine et <math>\Delta G_i</math> le coefficient associé)</i> (70).....	54

## Liste des Figures

Figure 1: L'évolution des causes de décès au cours des prochaines décennies	3
Figure 2 : Illustration Parallèles des caractéristiques du cancer entre le cancer et la cicatrisation des plaies	5
Figure 3:La signalisation dans le microenvironnement tumoral au cours de la progression maligne	10
Figure 4:Ciblage thérapeutique des caractéristiques du cancer	12
Figure 5:Représentation des familles de PTK cytosoliques ou nRTK (15)	14
Figure 6:Transduction de signal activée par ligand	15
Figure 7:Représentation des familles RTKs et de leurs membres (15)	16
Figure 8:du mécanisme d'activation des RTKs	18
Figure 9:Représentation d'une Kinase (complexe 1R0P/KSA) sous sa conformation active (motifDFG-in).	20
Figure 10:Représentation schématique de l'EGFR	22
Figure 11: Structure de l'EGFR démontrant les domaines pertinents	24
Figure 12: Structure cristalline de l'EGFR	24
Figure 14:Récepteurs : EGFR (HER-1), HER-2, HER-3 et HER-4	26
Figure 13:Représentation schématique de l'activation d'EGFR sous l'effet et d'un ligand	26
Figure 15:Représentation schématique de l'autophosphorylation	28
Figure 16:Principales voies sous-jacentes et fonctions d'EGFR	29
Figure 17:Cytotoxicité et internalisation du récepteur EGFR induites par le cétuximab	31
Figure 18:Molécules inhibitrices du récepteur EGFR	32
Figure 19:Répartition des principales mutations d'EGFR.	33
Figure 20:Représentation tridimensionnelle du domaine tyrosine kinase d'EGFR et de ses principales mutations	35
Figure 21: Un exemple de pharmacophore généré avec le logiciel LigandScout 4.0, représentant le ligand S58 (1CX2), un inhibiteur sélectif de COX-2 (flèches rouges : groupements accepteurs de liaisons hydrogène, sphères jaunes : groupements hydrophobes)	37
Figure 22 Liaison ionique(66)	40
Figure 23 : Forces de van der Waals(67).	40
Figure 24 :Pont hydrogène(68).	41
Figure 25 : Interactions hydrophobes (65).	41



Figure 26:Schéma illustrant le principe de « lock and key »	42
Figure 27:Principe du docking rigide illustré par le logiciel DOCK	43
Figure 28:Illustration de la méthode de recherche exhaustive employée par EUDOC	46
Figure 29:Protomol généré par Surflex pour la cyclooxygénase 2 (1cx2)	47
Figure 30:Schéma illustrant la méthode de reconstruction incrémentale (70).	47
Figure 31:Schéma illustrant la méthode de Monte-Carlo	49
Figure 32:Schéma illustrant la méthode utilisant les algorithmes génétiques	50
Figure 33:Schéma illustrant la méthode de recherche tabou	51
Figure 34 : Processus de correspondance entre les atomes et la densité électronique pour produire un modèle cristallographique. (93)	57
Figure 35:informations générales du micro ordinateur utiliser dans le travail	59
Figure 36:L'interface du Biovia Discovry Studio 2019	60
Figure 37:L'interface d'AutoDock	60
Figure 38:L'interface du PyRx	61
Figure 39:La séparation du ligand de son récepteur	63
Figure 40:les charges partielles des atomes de la cible de type"gesteiger"	64
Figure 41:Détection des liaisons rotatoires	64
Figure 42:Détermination des cordonnés(XYZ) du site actif d'une manière manuelle	65
Figure 43:Résultats du docking	65
Figure 44:Structures 3D des 4 complexes issues de la PDB	70
Figure 45:Comparaison des conformations obtenues par rayon-X (colorés en vert) et par docking moléculaire avec AutoDock vina (colorés en bleu).	71
Figure 46:Corrélation entre l'activité biologique (Log IC50) de 34 ligands et leur énergie d'interaction vis-à-vis de l'EGFR	72
Figure 47:Structures 3D des complexes EGFR issues de la PDB	79
Figure 48:les cordonnées du site d'interaction du complexe 4WKQ	79
Figure 49: L'enzyme 4WKQ non simplifié	80
Figure 50: L'enzyme 4WKQ simplifié.	80
Figure 51:La séparation du ligand «gefitinib » de son récepteur « le complexe 4WKQ »	81
Figure 52:Recherche des similaires au gefitinib sur PubChem	81
Figure 53:Les structures des meilleurs similaires	84
Figure 54:Représentation du positionnement du Gefetinib dans le site actif de la 4WKQ	85
Figure 55:Mode d'interaction du Gefetinib dans le site actif de la 4WKQ.	85
Figure 56:Mode de liaison du gefitinib a l'EGFR (160)	87

Figure 57:Mode d'interaction de la molécule n°1 dans le site actif de la 4WKQ	88
Figure 58:Mode d'interaction de la molécule n°2 dans le site actif de la 4WKQ	89
Figure 59:Mode d'interaction de la molécule n°4 dans le site actif de la 4WKQ	90
Figure 60:Mode d'interaction de la molécule n°5 dans le site actif de la 4WKQ	91
Figure 61:Mode d'interaction de la molécule n°6 dans le site actif de la 4WKQ	92
Figure 62:Mode d'interaction de la molécule n°7 dans le site actif de la 4WKQ	93
Figure 63:Mode d'interaction de la molécule n°8 dans le site actif de la 4WKQ	94
Figure 64:Mode d'interaction de la molécule n°9 dans le site actif de la 4WKQ	95
Figure 65:Mode d'interaction de la molécule n°10 dans le site actif de la 4WKQ	96
Figure 66:Mode d'interaction de la molécule n°11 dans le site actif de la 4WKQ	97
Figure 67:Mode d'interaction de la molécule n°12 dans le site actif de la 4WKQ	98

## Liste des Tableaux

Tableau I:Les facteurs de croissance de la famille de l'EGF.....	23
Tableau II: Modèles de souris transgéniques.....	27
Tableau III:Exemples de molécules ciblant les récepteurs ErbB.....	30
Tableau IV: Interprétation par rapport à la valeur du coefficient de corrélation .....	69
Tableau V:Les cordons des sites d'interaction (récepteur-ligand) (XYZ).....	71
Tableau VI:Corrélation entre les énergies d'interaction des 30 inhibiteurs de la EGFR et leur activité biologique (Log IC50).....	73
Tableau VII:Code PDB des complexes " EGFR-inhibiteur " et leurs caractéristiques .....	74
Tableau VIII: Score des meilleurs inhibiteurs obtenus suite au docking moléculaire de la collection de similaires du gefitinib .....	82
Tableau IX:Types d'interaction et acides aminés impliqués dans l'inhibition du EGFR (PDB ID : 4WKQ) avec Gefitinib .....	86
Tableau X:Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°1 .....	88
Tableau XI :Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°2.....	89
Tableau XII:Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°4.....	90
Tableau XIII:Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°5.....	91
Tableau XIV:Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°6.....	92
Tableau XV:Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°7.....	93
Tableau XVI:Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°8.....	94
Tableau XVII:Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°9.....	95
Tableau XVIII:Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°10.....	96
Tableau XIX:Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°11 .....	97

Tableau XX :Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°12.....	98
Tableau XXI: Propriétés physicochimiques .....	100
Tableau XXII: Propriétés pharmacocinétiques .....	102
Tableau XXIII :Résultats de la toxicité .....	103

## INTRODUCTION :

La communication cellulaire est indispensable à la vie des organismes pluricellulaires. Les informations transmises d'une cellule à une autre correspondent à six grands types de consignes à exécuter, antithétiques deux à deux : se reproduire ou se différencier ; demeurer attaché ou migrer ; survivre ou mourir. La cellule cancéreuse est une cellule génétiquement instable, capable d'explorer les fonctions de l'ensemble du génome et de mettre à profit tout avantage prolifératif ou migratoire pour le sélectionner et le transmettre à sa descendance. Toutes les voies de signalisation impliquées dans la prolifération et la différenciation, dans l'adhésion et la migration, dans la survie et la mort, pourront servir de support à des altérations oncogéniques. Ces altérations peuvent précisément faire l'objet d'un ciblage thérapeutique: c'est de cette constatation qu'est né le concept de thérapie ciblée. Les inhibiteurs de kinases sont des médicaments antinéoplasiques très importants et largement utilisés en clinique ; Parmi les récepteurs à activité tyrosine kinase, les récepteurs de la famille ERBB sont les plus impliqués dans l'oncogenèse de plusieurs cancers épithéliaux: sein, poumon, côlon, voies aérodigestives supérieures (VADS), etc. ; ils sont donc des cibles très recherchées dans la conception et la découverte des médicaments. Cette dernière implique l'utilisation des techniques de modélisation moléculaire telles que la modélisation de pharmacophore, la dynamique moléculaire, le criblage virtuel et le Docking moléculaire pour expliquer l'activité de biomolécules, définir les déterminants moléculaires pour l'interaction avec la cible du médicament et concevoir des médicaments candidats plus efficaces.

C'est dans ce même contexte que s'inscrit ce travail en mettant en avant le développement de nouveaux inhibiteurs de l'EGFR potentiellement plus puissants en faisant appel à la modélisation par criblage virtuel basé sur le ligand d'une collection des molécules d'une chimiothèque. Ces molécules ont fait l'objet d'un docking moléculaire suivi d'une analyse visuelle de l'interaction EGFR-inhibiteurs et la détermination des propriétés ADME Tox pour pouvoir en fin sélectionner les meilleurs candidats médicaments pour des études ultérieures.

Ce manuscrit est composé de deux grandes parties ;Une recherche bibliographique : divisée en 3 chapitres: le premier chapitre est une étude bibliographique se rapportant au cancer et tyrosine kinase. Par la suite, un deuxième chapitre dédié à la présentation de la cible EGFR et on termine par le troisième chapitre qui est consacré à la présentation des méthodes in silico, les modèles pharmacophoriques et les études de docking qui seront particulièrement détaillés. Dans la partie pratique on va mettre en exergue la méthodologie ou expérimentation utilisée dans ce Travail ; l'analyse et l'interprétation des résultats expérimentaux puis la discussion de ces résultats.

Enfin une conclusion générale clôturant ce travail est tirée

---

# **Partie**

# **bibliographique**

---

**CHAPITRE 1 : CANCER  
&  
TYROSINE KINASE**



## I. CANCER :

### I.1. DEFINITION :

Le mot «cancer» est un terme générique désignant un grand groupe de maladies pouvant toucher n'importe quelle partie de l'organisme. On parle aussi de tumeurs malignes ou de néoplasmes. L'un des traits caractéristiques du cancer est la multiplication rapide de cellules anormales à la croissance inhabituelle, qui peuvent envahir des parties voisines de l'organisme, puis migrer vers d'autres organes. On parle alors de métastases, lesquelles constituent la principale cause de décès par cancer.

### I.2. EPIDEMIOLOGIE :

Le cancer est la principale cause de décès dans les pays économiquement développés et la deuxième cause de décès dans les pays en développement.

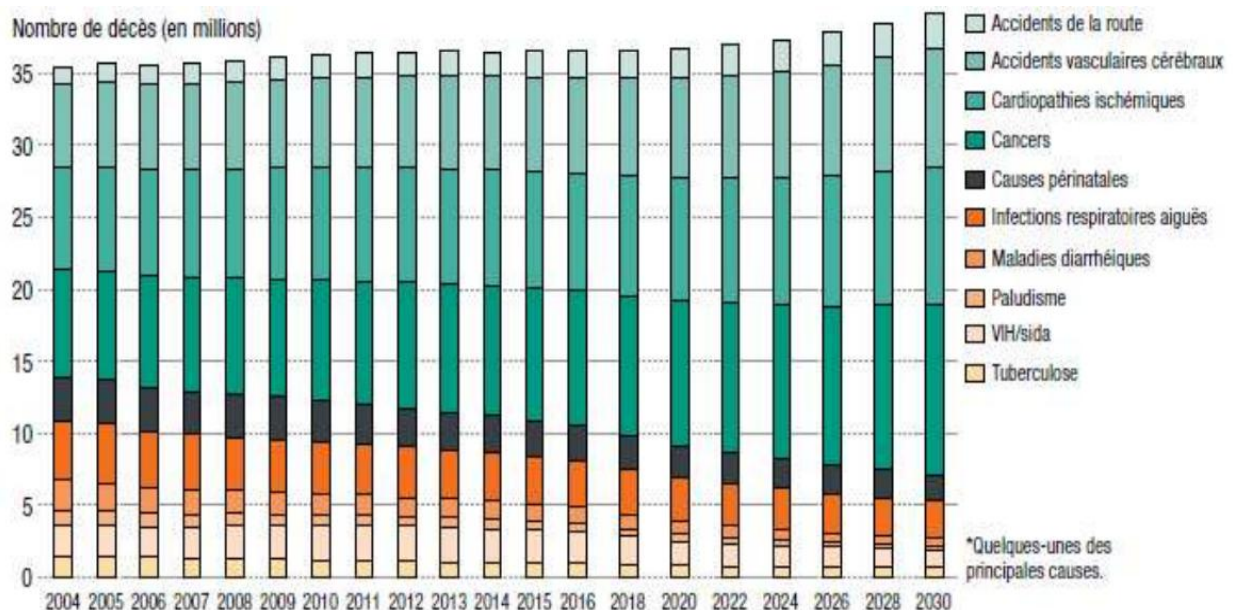


Figure 1: L'évolution des causes de décès au cours des prochaines décennies

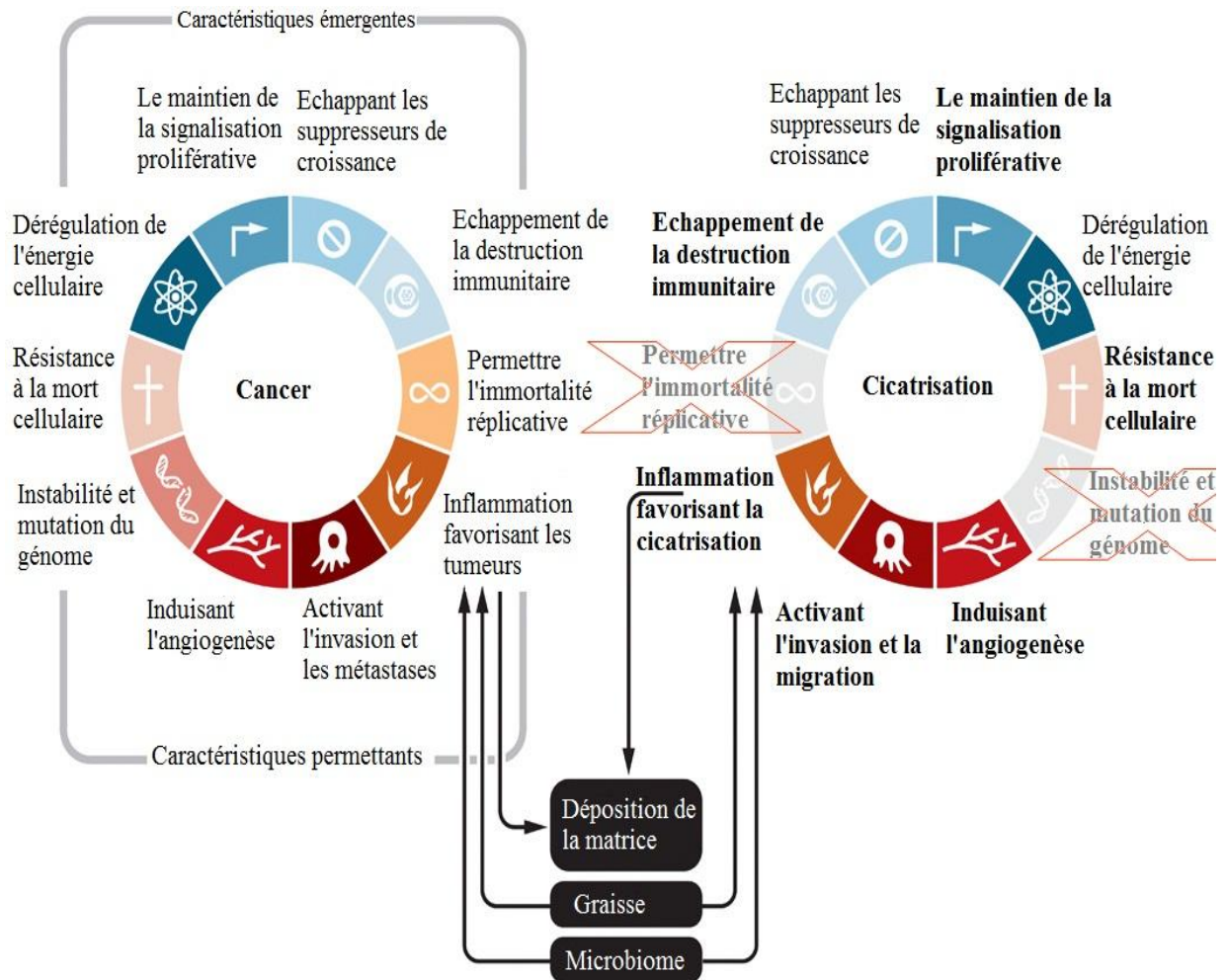
### I.3. CARACTERISTIQUES DU CANCER:

lorsque les cellules normales évoluent progressivement vers un état néoplasique, elles acquièrent une succession de ces capacités caractéristiques, et que le processus en plusieurs étapes de la pathogenèse des tumeurs humaines pourrait être rationalisé par le besoin des cellules cancéreuses naissantes d'acquies les caractéristiques qui leur permettent de devenir tumoro-gènes et finalement malignes.

Selon les caractéristiques originales du cancer, six capacités différentes conduisent au développement et à la progression des cancers (1)

L'évolution de la compréhension conceptuelle de la biologie du cancer au cours de la dernière décennie a conduit à suggérer récemment que la reprogrammation du métabolisme et l'évitement de la destruction immunitaire soient acceptés comme des caractéristiques supplémentaires (2).

Nous proposons que l'altération épigénétique soit considérée comme un autre signe distinctif du cancer, et donc comme un axe d'étude supplémentaire pour la prochaine génération de thérapies anticancéreuses. Le cancer se caractérise par une croissance cellulaire incontrôlée et l'acquisition de propriétés métastatiques. Dans la plupart des cas, l'activation d'oncogènes et/ou la désactivation de gènes suppresseurs de tumeurs entraînent une progression incontrôlée du cycle cellulaire et l'inactivation des mécanismes apoptotiques. Contrairement aux tumeurs bénignes, les cancers malins développent des métastases, ce qui est dû en partie à la régulation négative des récepteurs d'adhésion cellulaire nécessaires à l'attachement cellule-cellule spécifique au tissu, et à la régulation positive des récepteurs qui augmentent la motilité cellulaire. En outre, l'activation des métalloprotéases membranaires fournit aux cellules cancéreuses métastatiques une voie physique de propagation. Il existe différents mécanismes par lesquels ces changements génétiques et cellulaires se produisent. Les mécanismes canoniques sont la mutation, la translocation ou la délétion chromosomique, et l'expression ou l'activité dérégulée des voies de signalisation. Ces événements peuvent activer des gènes qui favorisent un cycle cellulaire dérégulé et/ou inactivent les voies apoptotiques. Ces processus sont bien décrits dans la littérature existante, et de nombreuses et excellentes revues sont disponibles sur chaque sujet(3) (4)



**Figure 2 : Illustration Parallèles des caractéristiques du cancer entre le cancer et la cicatrisation des plaies**

### I.4. PROGRESSION DU CANCER

Lorsqu'une division et/ou prolifération cellulaire anarchique atteint des cellules du tissu de l'organisme, on dit que ce dernier est atteint de cancer (5) (6) . De nouvelles cellules résultant de ce trouble des mécanismes du renouvellement cellulaire se forment. Elles sont connues sous le nom de tumeurs, néoplasmes, carcinomes ou encore métastases (tumeurs filles).

Comme c'est une pathologie liée aux troubles de la reproduction cellulaire, le cancer est une maladie qui se caractérise par des modifications, causées par des facteurs externes ou internes, des gènes.

L'altération constitutionnelle d'un certain type de gène constitue la cause des cancers dits familiaux qui peuvent se transmettre à la descendance.

Les gènes altérés en question se distinguent en 4 catégories :

- **Les oncogènes** (état muté des proto-oncogènes) : considérés comme des catalyseurs, ils régulent positivement la prolifération ou croissance cellulaire anarchique.

- **Les gènes suppresseurs de tumeurs** : au contraire de la première catégorie, ils inhibent la prolifération ou croissance cellulaire anarchique.

- **Les gènes angiogéniques** : la cellule cancéreuse se divise rapidement et a besoin d'un grand apport en oxygène véhiculé par l'hémoglobine. Ce type de gène contrôle l'apport sanguin des cellules. Son altération favorise donc l'angiogénèse des cellules, dont les cancéreuses.

- **Les gènes métastatiques** : la plupart des cellules sont assignées à des places bien précises dans les tissus s'entourant de cellules voisines qui aident à la définition de la localisation. Une fois les cellules délocalisées de leurs tissus, elles s'arrêtent de proliférer et meurent. Les cellules cancéreuses peuvent acquérir, par le biais de mutations, la faculté de se déplacer sans mourir permettant ainsi leur diffusion dans l'organisme. Les gènes métastatiques permettent le contrôle de la diffusion des cellules cancéreuses (métastases).

la maladie est la conséquence d'un ou plusieurs dysfonctionnement(s) cellulaire(s) tel(s) que :

- Une perte de contrôle du cycle cellulaire
- Insensibilité à l'apoptose
- Irrégularité dans la réparation de l'ADN

Le cancer est considéré comme une pathologie multi génique (plusieurs gènes sont altérés), contrairement aux maladies mono géniques comme la mucoviscidose, certaines hémophilies et les myopathies. Cette altération peut toucher de 10 à 20 gènes de manière successive, non-aléatoire et pouvant s'étaler sur une longue période (5 à 20 ans).

Chaque type de cancer possède son propre mécanisme d'évolution de la maladie. A titre d'exemple, dans le cancer du colon, le phénomène débute par l'inactivation du gène suppresseur de tumeur APC, la mutation de l'oncogène Haras et l'inactivation du gène suppresseur de tumeur P53. Cette spécificité et cette chronologie bien précise de chaque type de cancer, montre que la diversité génétique est la base de l'hétérogénéité de la pathologie cancéreuse.

L'évolution de la maladie du cancer comporte deux étapes successives : la cancérogenèse et la tumoro-genèse (6).

#### **I.4.1. Cancérogenèse**

Il existe des agents carcinogènes entraînant des altérations dans la séquence de l'ADN.

À ce stade là, le cancer n'altère pas encore les fonctions de l'organisme et n'est cliniquement pas observable. Ce phénomène précurseur est appelé carcinogenèse, généralement observé dans les cancers dérivant des tissus épithéliaux. De manière générale, l'exposition à un agent carcinogène n'est pas un phénomène irréversible, en terme d'évolution de la maladie. Les agents carcinogènes peuvent être désactivés et/ou éliminés par le biais des agents enzymatiques qui assurent leur métabolisme. S'ils ne sont pas expulsés, s'ensuit alors l'étape

de l'initiation du processus tumoral. Conséquence directe de la carcinogenèse, cette étape est causée par l'accumulation des anomalies génétiques, en particulier les mutations des gènes maintenant l'intégrité et les caractéristiques des différents types cellulaires. Il en résulte l'activation des oncogènes, responsables de la multiplication cellulaire, et l'inhibition des gènes suppresseurs de tumeurs (P53) responsables de la mort cellulaire programmée. Les cellules résultant de l'ensemble de ces altérations génétiques sont dites « cellules initiées » et ne sont pas encore cancéreuses(7) . Leur prolifération, via d'autres agents carcinogènes, plus ou moins contrôlée, marque le début de la seconde étape de la cancérogenèse, la promotion.

Au cours de cette multiplication, d'autres mutations sont engendrées et transmises aux cellules filles des cellules initiées. Le passage de la maladie au stade clinique est déclenché par l'instabilité accrue et l'accumulation des anomalies génétiques au niveau des composantes de la cellule qui apportent nutriments et soutien à leur développement (le stroma)..

### I.4.2. Tumorigenèse

Il s'agit de la progression des cellules précancéreuses d'abord localement, puis allant en grossissant. Le stade clinique de tumeur est alors atteint. Leur délocalisation par voies lymphatiques ou veineuses au niveau des tissus avoisinants donne lieu à des métastases. La circulation sanguine joue un rôle catalyseur dans l'invasion cellulaire ou l'angiogénèse. En effet, les cellules tumorales ayant besoin d'un milieu favorisant leur croissance et leur déplacement, acquièrent un pouvoir d'angiogénèse (système sanguin nourricier) par stimulation des cellules endothéliales des capillaires voisins. L'acquisition de cette capacité à susciter l'angiogénèse des cellules cancéreuses est la réaction cellulaire normale face à l'hypoxie ou le résultat d'une mutation génétique.

Certaines tumeurs primaires peuvent progresser vers un envahissement plus global de l'organisme par échappement de cellules tumorales issues de cette tumeur primaire : on parle alors de métastase.

Les métastases, qui sont l'évolution et la mobilité des cellules cancéreuses vers un ou plusieurs lieux, sont de 2 types :

**Métastases ganglionnaires ou carcinomes** : se déplacent par les canaux lymphatiques, arrivées au niveau du ganglion, elles peuvent :

Être détruites suite à une inflammation

Rester latentes sans activité directe

Se fixer aux ganglions et s'y multiplier, entraînant une grosseur cliniquement palpable

Traverser le ganglion se dirigeant vers d'autres canaux lymphatiques

Se faufiler dans tout le circuit lymphatique, signe évident de l'agressivité de la tumeur

**Métastases sanguines ou sarcomes** : pénètrent les vaisseaux sanguins soit de manière directe ou par le biais des canaux lymphatiques. Elles infiltrent préférentiellement les organes où passe gros volume de sang (foie, poumons... etc.) ainsi que les os (métastases osseuses).

## I.5. SIGNALISATION CELLULAIRE ET CANCER

La cellule cancéreuse est une cellule génétiquement instable. capable d'explorer les fonctions de l'ensemble du génome et de mettre à profit tout avantage prolifératif ou migratoire pour le sélectionner et le transmettre à sa descendance. Une tumeur est une néoformation nécessitant une multiplication cellulaire toujours active ; le support de la malignité des cancers est lié à leur aptitude à disséminer dans l'organisme : la capacité de survie est une nécessité pour les cellules tumorales. Toutes les voies de signalisation impliquées dans la prolifération et la différenciation, dans l'adhésion et la migration, dans la survie et la mort, pourront servir de support à des altérations oncogéniques. On a pu dire ainsi que le cancer était une maladie de la signalisation cellulaire.

Les voies de signalisation sont constituées de protéines interagissant les unes avec les autres en séquences ordonnées. Certaines de ces protéines ont des fonctions catalytiques (en particulier des phosphorylations et des déphosphorylations, assurées respectivement par des kinases et des phosphatases) et sont plus accessibles au pharmacologue que celles dont le rôle est simplement d'interagir avec une autre protéine. Une difficulté particulière vient du fait que les voies de signalisation sont redondantes et interconnectées il arrive qu'un message que l'on empêche d'aboutir en bloquant une voie donnée pourra quand même être transmis par l'intermédiaire d'une autre voie ou en empruntant un chemin détourné.(8)

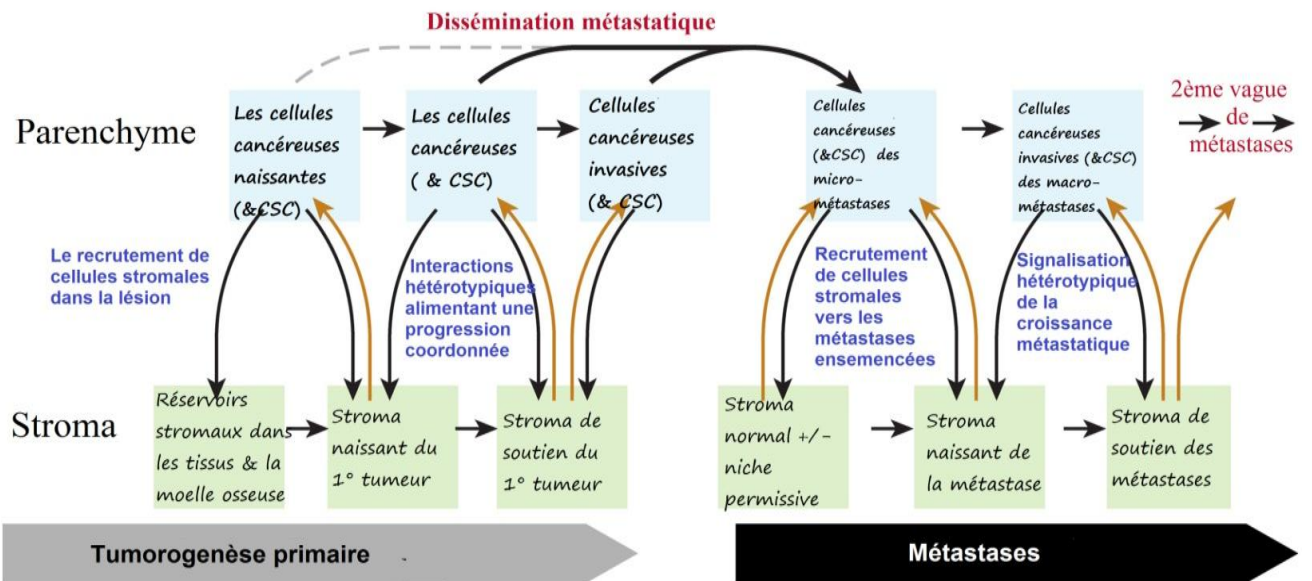
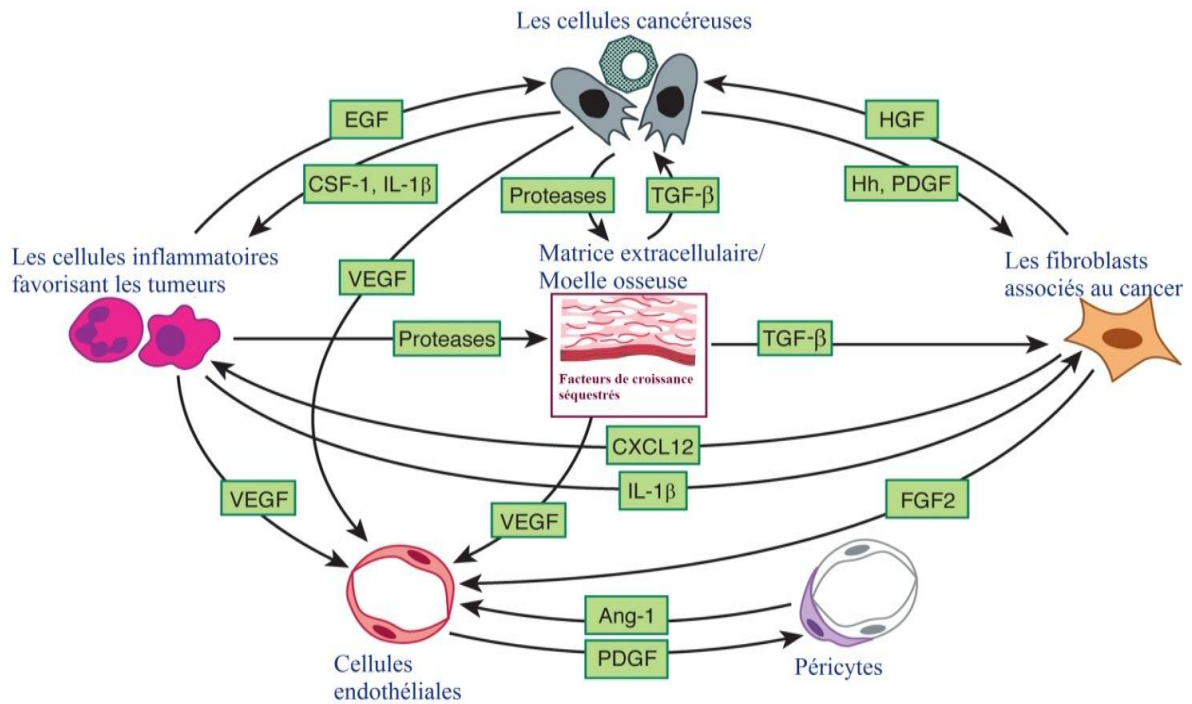


Figure 3: La signalisation dans le microenvironnement tumoral au cours de la progression maligne



## I.6. TRAITEMENT DU CANCER

Le traitement nécessite d'avoir un diagnostic de certitude et de connaître le type du cancer, d'évaluer son extension locale, régionale et la présence ou non de métastases ainsi que l'état général du patient (âge, fonctions cardiaque et rénale, présence d'autres maladies) suivant les cas, il repose sur :

- L'exérèse (l'ablation) chirurgicale large de la tumeur quand cela est possible, large voulant dire que le bistouri du chirurgien passe uniquement par des tissus sains.
- Une radiothérapie, une irradiation de la tumeur par des rayons x ou  $\gamma$  permettant de faire diminuer voire disparaître celle-ci.
- Les greffes de la moelle osseuse : il ne s'agit pas d'un traitement direct mais elles sont parfois utilisées pour renforcer le système de production des cellules sanguines.
- Une chimiothérapie qui correspond à la prescription de médicaments s'attaquant au cancer et à ses métastases. L'inconvénient principal c'est le manque de sélectivité (la chimiothérapie n'attaque pas seulement les cellules cancéreuses mais touche aussi des cellules saines).

La connaissance des mécanismes de l'oncogenèse et celle des perturbations des voies de signalisation dans les cancers a permis l'émergence du concept de thérapie ciblée. Alors que les médicaments de la chimiothérapie classique ont pour cible les mécanismes impliqués dans la multiplication cellulaire, qu'elle soit normale ou néoplasique, les thérapeutiques ciblées s'adressent aux mécanismes mêmes de l'oncogenèse et devraient donc avoir une spécificité importante pour les cellules cancéreuses. Ce que l'on cible, dans les thérapeutiques ciblées, ce sont les protéines impliquées dans le contrôle de la prolifération et de la mort cellulaires et non celles impliquées dans l'exécution de ces programmes, c'est la cause de l'anomalie à l'origine d'un cancer et non plus l'effet qui en résulte. Ces thérapies ont émergé lorsque la compréhension de la biologie des cancers a atteint un niveau suffisant et que les outils permettant d'interférer avec la transformation maligne ont été disponibles. (9) (10)

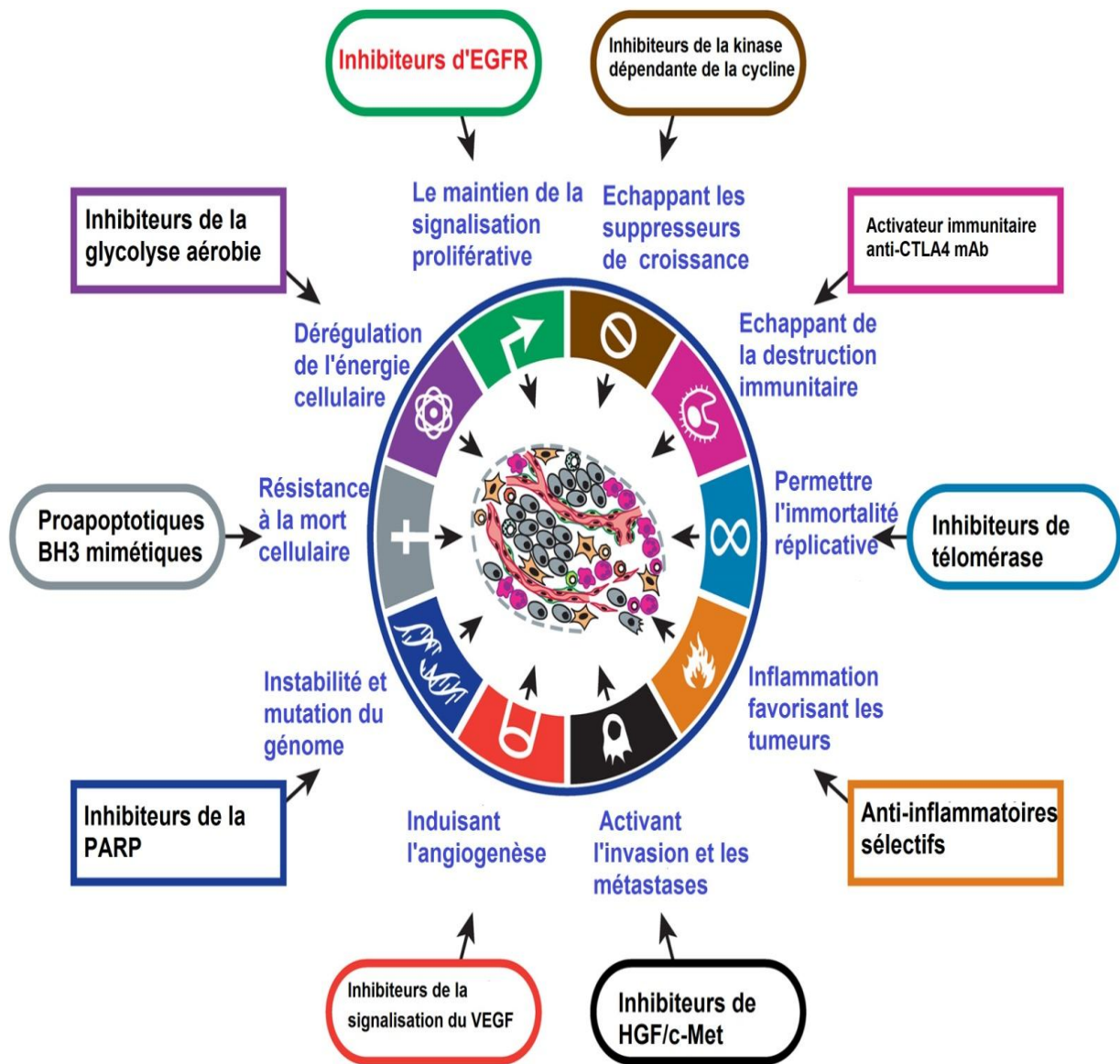


Figure 4: Ciblage thérapeutique des caractéristiques du cancer

### I.6.1. Avances thérapeutiques

Les avancées dans le domaine des découvertes des structures tridimensionnelles des protéines, de la résolution du séquençage du génome humain ainsi que dans le développement des connaissances des maladies et tout leur système de signalisation ont contribué à la croissance des bases de données des molécules à vertus thérapeutiques. Ce progrès constant des moyens utilisés en R&D du secteur pharmaceutique, contribue à rendre ce domaine vorace en argent et en temps d'investissement. Pour toutes ces raisons, l'avènement des approches *in silico* dans le domaine de la R&D apparaît comme une option de plus en plus prisée, dans le milieu pharmaceutique en particulier.

## II. TYROSINE KINASE

### I.1. GENERALITE :

Les protéines-tyrosine kinases (PTK) sont des régulateurs importants des voies de transduction des signaux intracellulaires qui interviennent dans le développement et la communication multicellulaire chez les métazoaires. Leur activité est normalement étroitement contrôlée et régulée.

La perturbation de la signalisation des PTK par des mutations et autres altérations génétiques entraîne une dérégulation de l'activité des kinases et une transformation maligne.

L'étude séminale de Cohen et al. en 1965 (11) postulait qu'un récepteur membranaire était responsable de la réponse proliférative des cellules épithéliales à un ligand soluble à savoir le facteur de croissance épidermique (EGF).

Par la suite, au milieu des années 1980, deux groupes différents ont élucidé la séquence primaire de ce récepteur membranaire, appelé récepteur du facteur de croissance épidermique ou EGFR(12,13), ce qui a déclenché des recherches intensives sur ces RTK et d'autres au cours des trois dernières décennies. Cette revue évalue les différences structurelles entre les RTKs distincts à travers l'évaluation de modèles spécifiques, ainsi que les implications pour la transduction du signal, la fonction cellulaire et la maladie.

### II.1. CLASSIFICATION DES PK

De manière générale, la classification des protéines kinases peut s'effectuer selon leurs structures (leur séquence) ou en fonction de leur emplacement dans la cellule. La classification la plus connue est celle de Hank & Hunter (14) qui a été mise à jour avec la découverte d'autres PK. En effet, le nombre de classes est passé de 6 à 9. Classées en fonction des résidus phosphorylés au cours de la catalyse, les 9 familles appartiennent à 3 grandes classes. À cette classification s'ajoute une autre classe regroupant le reste des kinases

Notre intérêt se porte essentiellement sur la famille des Tyrosine Kinases « TK ». Cette famille tire son nom du fait que la phosphorylation des protéines cibles se fait sur des résidus de type Tyrosine grâce à la catalyse d'une protéine membre de cette famille. Les protéines appartenant à la famille des TK se subdivisent en 2 groupes par rapport à leur localisation dans la cellule :

II.1.1. Les PTK cytosoliques (CTK)

Appelées également les Non-Récepteurs Tyrosine Kinase (nRTK). Cette classification concerne les protéines évoluant dans le cytosol de la cellule (phase liquide du cytoplasme), plus précisément, au niveau de la surface interne de la membrane plasmique. Cette subdivision qui fut découverte avant les Récepteurs Tyrosine Kinases (RTK) comporte les 10 sous-familles suivantes : ABL, ACK, CSK, FAK, FES, FRK, JAK, SRC, SYK et TEC. Ces tyrosines sont impliquées dans les interactions de type protéine-protéine, protéine-lipide et protéine-ADN.

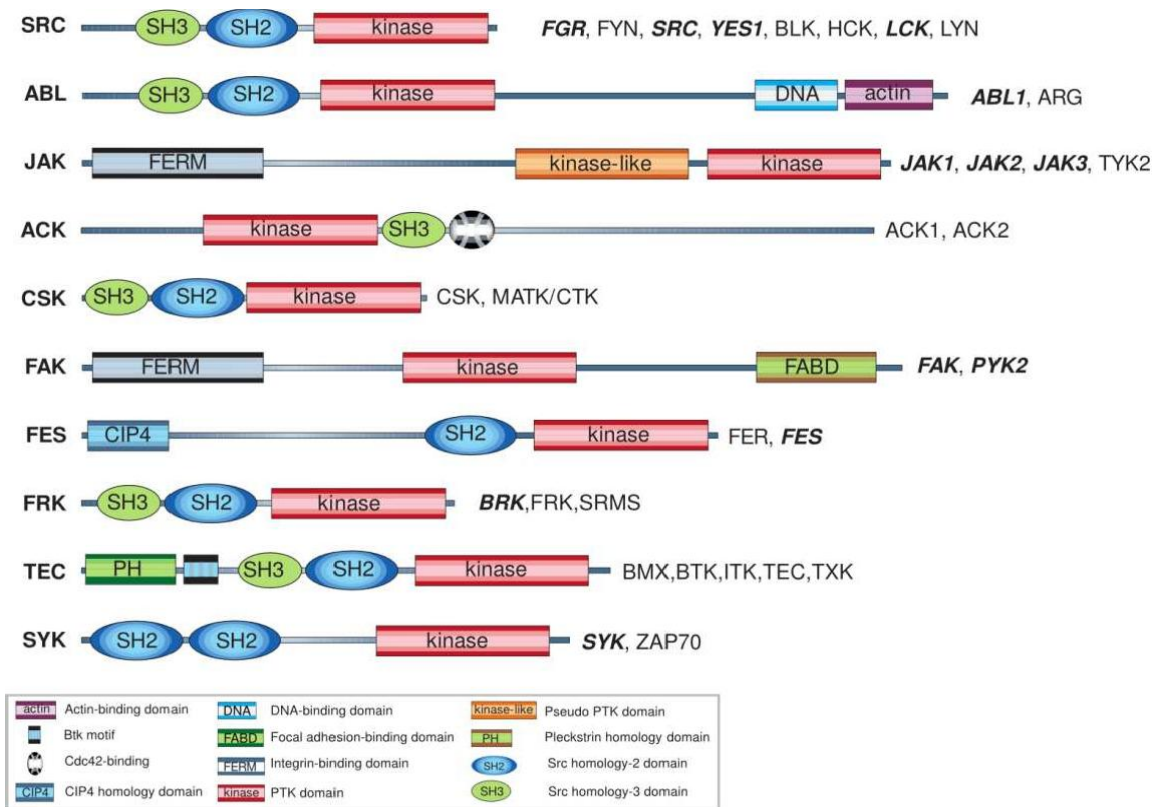
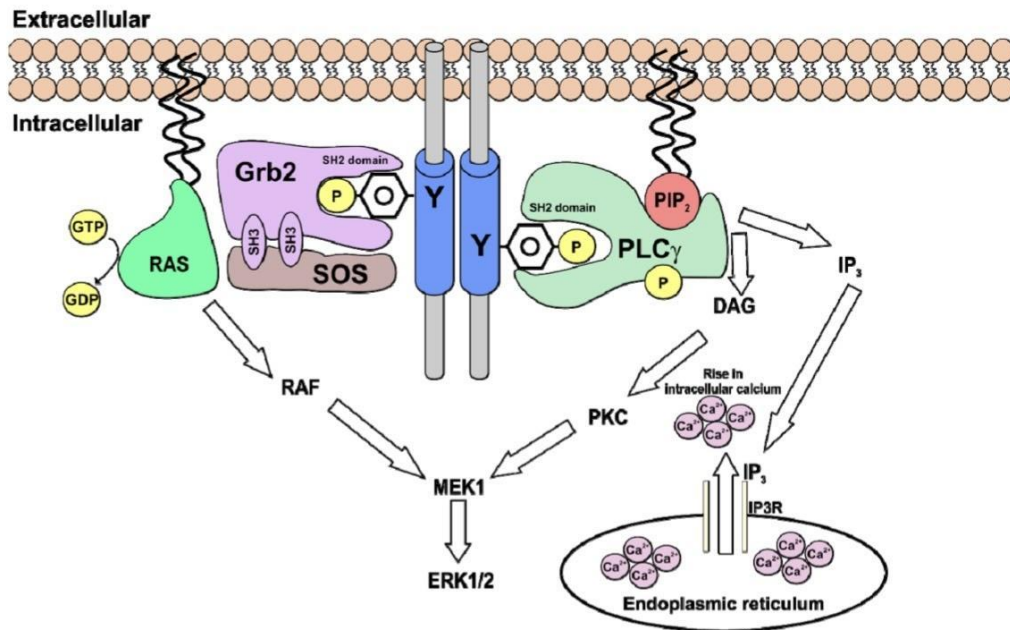


Figure 5: Représentation des familles de PTK cytosoliques ou nRTK (15)

Elles jouent un rôle majeur dans les voies de signalisation et doivent se trouver capables de détecter la phosphorylation d'une autre protéine en amont ainsi que son changement conformationnel. Les domaines de ces protéines activant les cascades signalétiques intercellulaires se présentent comme :

-Le domaine **SH2** : constitué d'une centaine d'acides aminés, il possède 2 poches de fixation édifiées en structure secondaire de feuillets  $\beta$ . La première poche est profonde et possède une arginine conservée formant l'interaction avec la tyrosine phosphorylée. L'autre poche en aval de la première se trouve être spécifique de la protéine phosphorylée. En d'autres termes, ce domaine est capable d'identifier les séquences peptidiques possédant une tyrosine phosphorylée.



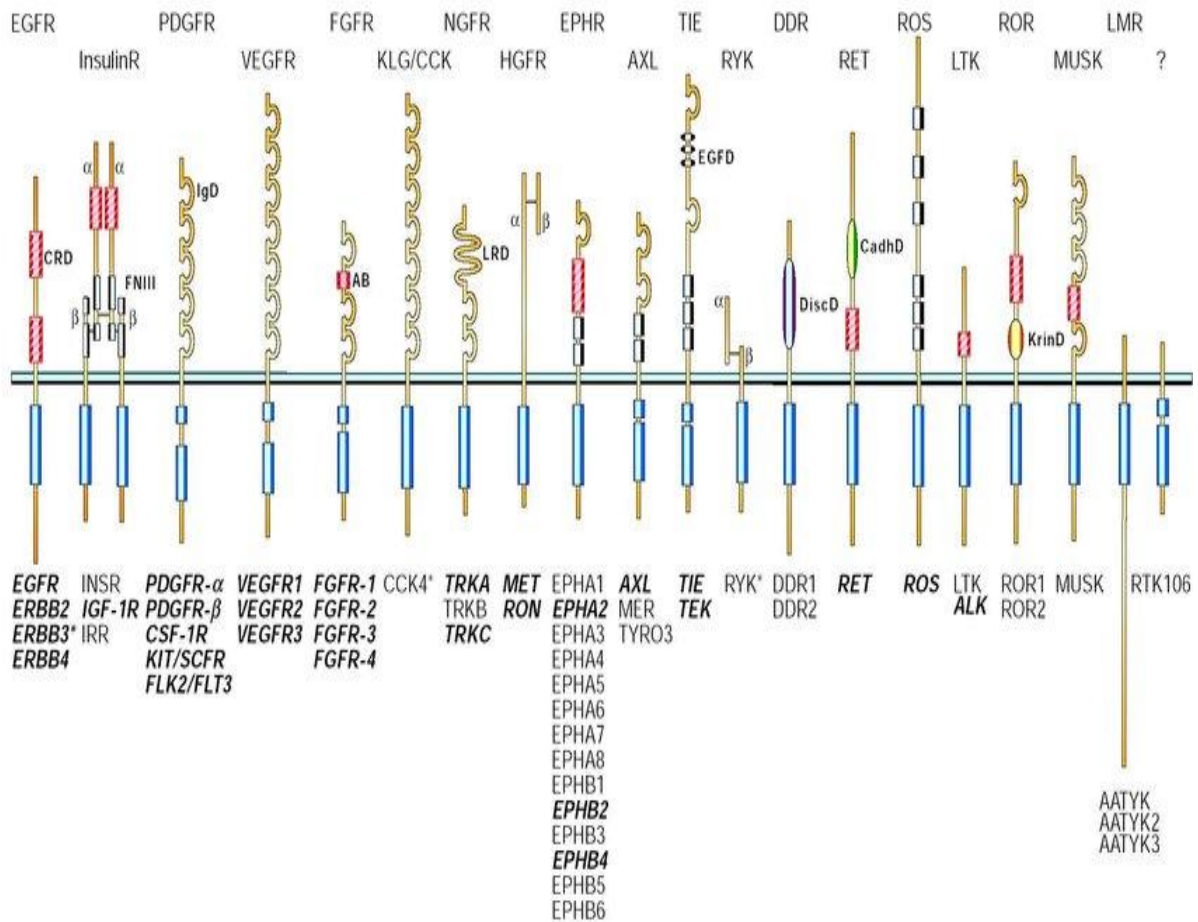
**Figure 6: Transduction de signal activée par ligand**

-Le domaine **SH3** : constitué de 3 sous-sites, S0, S1 et S2, comprenant près de 60 acides aminés, ce domaine affiche une spécificité pour des peptides possédant des prolines. Afin d'établir une interaction, le sous-site S2 fixe une arginine près de la région contenant les prolines.

-Le domaine **PH (Pleckstrin Homology)** : afin d'encadrer les protéines vers la membrane plasmique, ce domaine permet la fixation aux lipides phosphatidylinositol.

### II.1.2. Les PK membranaires ou Récepteur Tyrosine Kinase (RTK)

La majorité des RTK présentent une structure similaire (Figure 10), comprenant un domaine extracellulaire de liaison au ligand suivi d'un seul domaine transmembranaire, d'une région juxta membranaire, d'un domaine tyrosine kinase cytosoliques (TKD) et d'une queue C-terminale flexible(16)



**Figure 7:Représentation des familles RTKs et de leurs membres (15)**

Les domaines extracellulaires des RTK varient selon les sous-familles, avec différents motifs de liaison et modules qui spécifient la reconnaissance et l'assemblage du ligand. Grâce à cette diversité, les récepteurs sont capables de se lier à des ligands avec une grande spécificité, empêchant ainsi l'amplification indésirable du signal et augmentant sa précision. Les régions extracellulaires se dimérisent lors de la liaison du ligand, formant des complexes d'ordre supérieur grâce à de multiples interactions électrostatiques (16–18) Le modèle classique d'activation des RTK consiste en l'assemblage de polypeptides récepteurs monomères en un complexe homodimérique de signalisation active lors de l'interaction avec le ligand. avec des changements conformationnels qui activent le domaine tyrosine kinase intracellulaire (19).

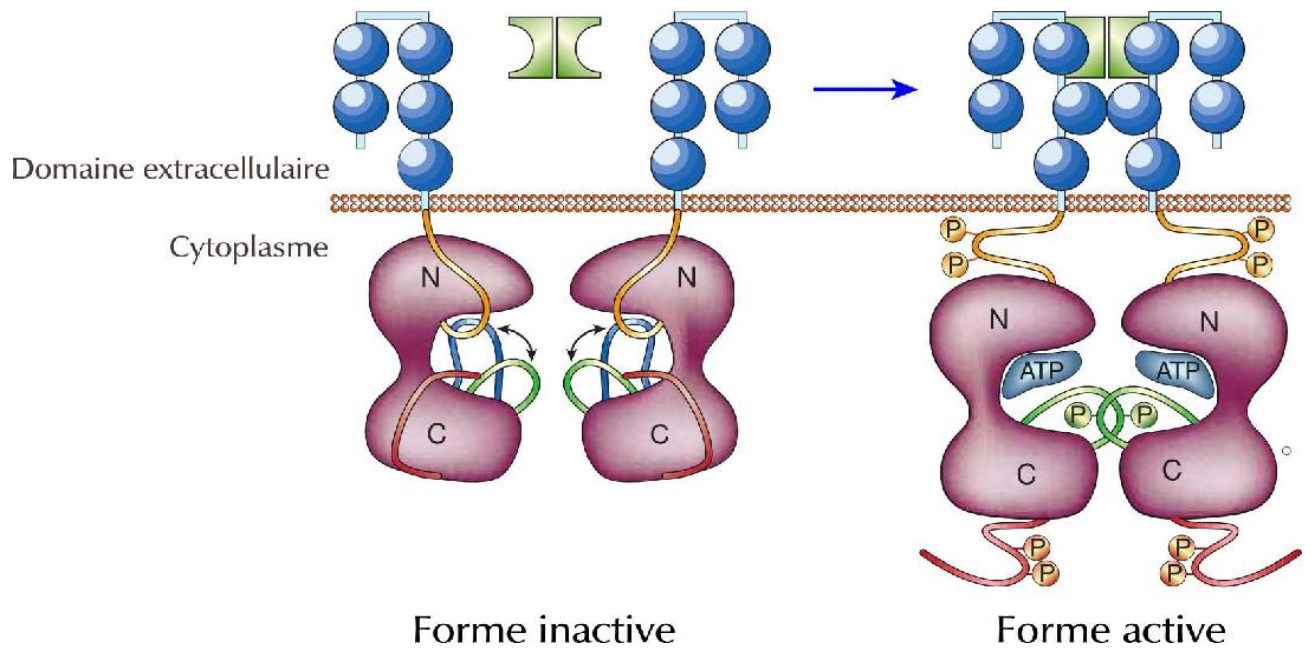
## II.1.2.1. ACTIVATION

L'activité de la majorité des kinases dépend de la phosphorylation de leur boucle d'activation par le biais d'une succession d'événements bien définis. Il n'y a que l'origine du déclenchement de cette activation, conduisant à un changement conformationnel de la protéine, qui varie d'une kinase à une autre.

Parmi les événements déclencheurs, on peut citer l'interaction de la kinase avec un ligand (exemple de fixation de l'insuline par son récepteur) ou l'interaction de la kinase avec un autre partenaire protéique.

Suite à la phosphorylation de la boucle d'activation et du changement conformationnel de la protéine, les 2 lobes (N et C terminaux) se rapprochent. Cette conformation permet alors un bon alignement d'acides aminés essentiels à la catalyse, un positionnement approprié de l'ATP et une exposition du site de reconnaissance du substrat. L'ATP maintenu dans la poche du petit lobe adjacente à la boucle d'activation, se trouve stabilisé par l'ion magnésium ( $Mg^{2+}$ , appelé également agent cofacteur). Il aide à la libération du  $\gamma$ -P induisant la phosphorylation spécifique de quelques résidus de la boucle catalytique. Cette boucle adopte ainsi une conformation tournée vers l'extérieur de la protéine, prête à accueillir le substrat, appelé DFG-in, et mettant ainsi la kinase sous sa forme active. On dit, selon l'appartenance de la PK à une famille ou une autre, que la kinase est capable soit de phosphoryler des substrats ou ligands spécifiques (phosphorylation en trans ou intermoléculaire) ou bien de s'autophosphoryler (phosphorylation en cis ou intramoléculaire)





**Figure 8:du mécanisme d'activation des RTKs**

Le changement de conformation de la kinase induit par la fixation du ligand (vert) favorise le positionnement de l'ATP dans le site catalytique. L'ATP permet la phosphorylation aussi bien de la boucle d'activation (en vert) rendant la kinase active, que la phosphorylation de la boucle coté C-terminal (en rouge) favorisant ainsi l'interaction avec d'autres partenaires (protéines) impliqués dans les voies de signalisation. (adapté de Blume-Jensen et Hunter (15))

Les RTKs catalysent le transfert d'un groupe phosphate de l'ATP à un substrat donné. En effet, l'activité enzymatique de ces protéines membranaires est effective lors de la fixation du ligand conduisant à la formation de dimères, provoquant ainsi la phosphorylation d'une tyrosine (autophosphorylation). Cette activité peut être aussi exercée sur des substrats, de type protéine dans le cytoplasme, qui transmettent à leur tour le signal de phosphorylation en cascade au sein de la cellule



## II.1.2.2. AUTO INHIBITION

À l'état de repos, chacun RTK présente encore une activité kinase enzymatique résiduelle ou basale. Le complexe RTK au repos est supposé " respirer " ou fluctuer entre des états dynamiques ; même en l'absence de ligand, des tyrosines kinases solubles (par exemple c-Src) pourraient phosphoryler le TKD et ainsi déclencher la transduction du signal basal (16).

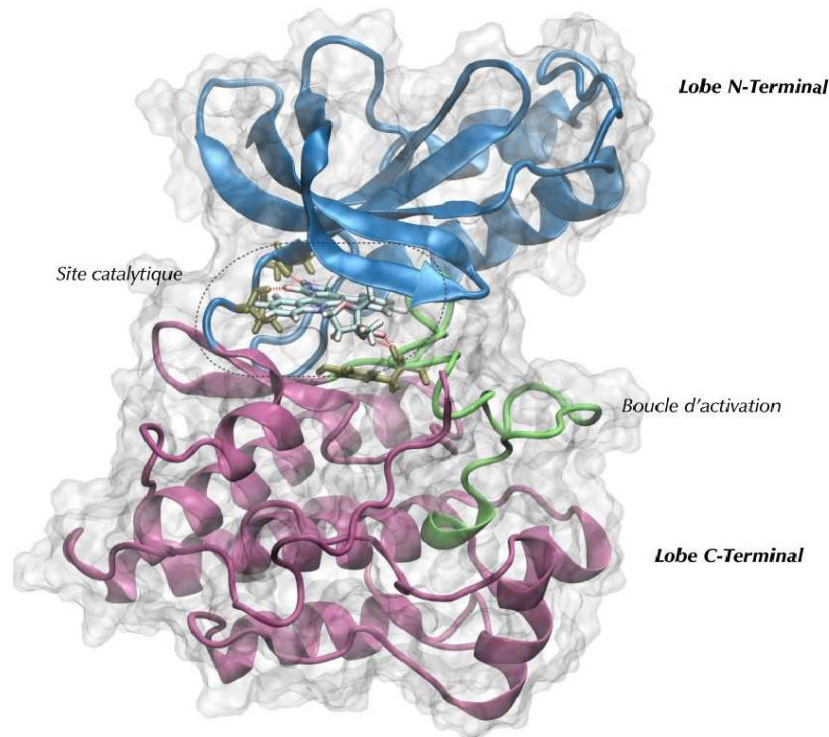
Pour empêcher la trans-autophosphorylation spontanée, les RTK contiennent un certain nombre d'éléments auto-inhibiteurs, la plupart des TKD étant inhibées de manière unique (20).

## II.1.2.3. Constituants du domaine catalytique des Kinases

Dans les études portant sur la connaissance de la flexibilité ou plasticité des protéines réceptrices, l'intérêt se porte essentiellement sur le domaine catalytique de ces dernières. Ce domaine, lieu de la catalyse est composé de 250 à 300 acides aminés partagés entre 12 domaines assez conservés (21)(I à XII). Commun à l'ensemble des membres d'une même famille, le domaine est généralement composé des constituants suivants :

Le lobe N-terminal ou « petit lobe » : contenant les sous domaines I à IV, il est essentiellement impliqué dans le positionnement et la liaison du substrat nucléotidique(ATP) dans le site catalytique. Il est majoritairement constitué des feuilletts  $\beta$  ainsi que de l'hélice  $C\alpha$ .

Le lobe C-terminal ou « petit lobe » : contenant les sous-domaines VI à XI, sièges du site de liaison à l'ion métal cofacteur ( $Mg^{2+}$ ) et de la boucle kinase catalytique avec le motif DFG. Il est majoritairement constitué des hélices  $\alpha$  de la protéine et des boucles d'activation. L'assemblage des acides aminés contenus dans ce lobe permet la reconnaissance du substrat et du positionnement du groupement hydroxyle des résidus tyrosine, sérine et thréonine du substrat. La région charnière : contenant le sous-domaine V, il s'agit de la boucle qui relie les 2 lobes constituant la kinase et se trouve être l'un des sites de liaison du complexe protéine/ligand. En effet, le site de liaison en question est connu comme le site de liaison ATP ou phosphoryle donneur. Il est situé dans la région charnière



**Figure 9:Représentation d'une Kinase (complexe 1R0P/KSA) sous sa conformation active (motifDFG-in).**

#### II.1.2.4. Les voies de signalisation médiées par les RTKs

Les RTKs modulent une large variété de processus biologiques, plus spécialement les signaux envoyés de la membrane cellulaire vers des cibles intracellulaires. Ils entrent également en jeu dans la coordination de fonctions biologiques complexes telle que la croissance cellulaire.

Les signaux intracellulaires sont ciblés par l'activation du domaine tyrosine kinase à travers la phosphorylation de multiples substrats. Les sous-familles des RTKs sont catégorisées par la similarité des aminoacides constituant leur séquence et leur structure extracellulaire. C'est pour cette raison que les membres d'une même sous-famille se lient de la même manière à certains ligands et induisent donc une cascade de signalisations complexes dont la compréhension se simplifie au gré de l'évolution des nouvelles méthodes de modélisation (22).

D'une manière générale, nous pouvons décrire sommairement les deux principales voies de signalisation transduites par la majorité des RTKs..

Voie RAS/MAPK : c'est une voie qui concerne directement la croissance et le développement cellulaire. La prolifération cellulaire qui en découle rend cette voie de signalisation capitale dans la compréhension des phénomènes liés au processus cancéreux dans le cas où son activation devient incontrôlable.

Voie PI3K/PKB (ou AKT) : cette voie régit la plupart des phénomènes liés au métabolisme tels que la synthèse de glycogène, d'acides gras ainsi que la translocation des agents transporteurs de glucose (23).

## II.2. TK ET CANCER

Un certain nombre de mutations distinctes dans les gènes codants pour les RTK sont liées à divers types de cancer. Elles sont documentées dans la liste des allèles connus liés au cancer dans une base de données appelée Catalogue of Somatic Mutations in Cancer (COSMIC) (24) et dans les mutations par gain de fonction dans les ectodomains des RTK entraînent une augmentation de la stabilisation, de l'activation et de la signalisation du dimère des RTK en l'absence de ligand (16,25). D'autres polymorphismes alléliques des RTK dans les domaines juxta membranaire et kinase provoquent une transformation oncogène en favorisant un état de signalisation constitutif et en altérant la régulation auto-inhibitrice, comme le montrent les études sur les tumeurs stromales gastro-intestinales (GIST) (26). Les polymorphismes du gène RTK au sein des exons codant pour la liaison au ligand, la trans-membrane, la juxta-membrane, la TKD, la boucle d'activation et la queue C-terminale sont également liées aux hémopathies et aux tumeurs solides malignes (27).

L'angiogénèse pathologique est une caractéristique du cancer et, comme tous les tissus, les tumeurs ont besoin d'un apport sanguin sain pour fournir des nutriments et de l'oxygène, ainsi que pour permettre l'élimination des déchets métaboliques et du dioxyde de carbone (28).

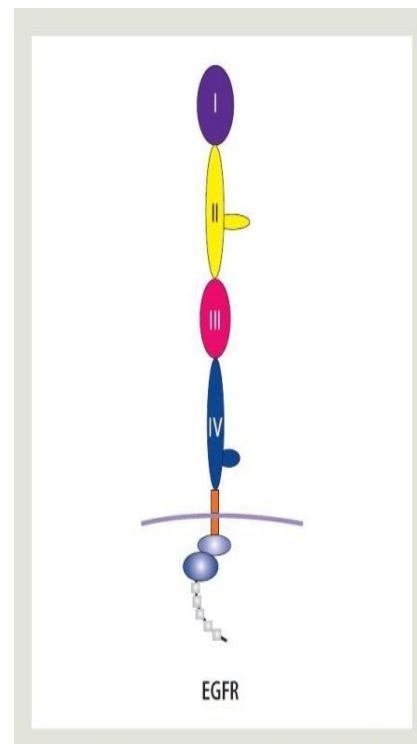
# **CHAPITRE 2 : EGFR**

## I. GENERALITE

La majorité des tumeurs épithéliales sont caractérisées par une activation fonctionnelle de facteurs de croissance agissant sur des récepteurs de la famille EGFR (epidermal growth factor receptor). L'EGFR est un récepteur transmembranaire de type tyrosine-kinase, membre d'une famille de quatre protéines structurellement apparentées. A ce jour, plusieurs ligands du récepteur EGFR ont été identifiés, permettant son activation, sa dimérisation, aboutissant finalement à l'autophosphorylation de son extrémité intracellulaire.

Cette modification biochimique est à la base d'une cascade de signalisations intracellulaires, régulant la prolifération tumorale, sa capacité d'invasion locale et à distance, sa résistance à l'apoptose ainsi que sa néo-angiogénèse. Les premières données publiées évaluant la possibilité d'inhiber la fonction du récepteur EGFR dans les cancers humains datent de 1984.

Depuis lors, plusieurs molécules inhibitrices de l'EGFR, distinctes dans leur structure et leur fonction, ont été testées cliniquement.



**Figure 10: Représentation schématique de l'EGFR**

## II. BIOLOGIE MOLECULAIRE, CELLULAIRE ET STRUCTURALE

### II.1. Structure primaire et organisation

L'EGF et son récepteur sont les prototypes de familles de protéines qui ont fait l'objet de duplications et de diversification au cours de l'évolution animale. Chez l'homme, on trouve quatre récepteurs homologues et onze ligands.

Les ligands de la famille de l'EGF peuvent être classés en fonction de leur sélectivité pour les récepteurs (29). Tous ces ligands sont des petites protéines solubles qui dérivent par protéolyse de précurseurs membranaires. Leur structure tridimensionnelle est caractérisée par la conservation de l'existence de trois ponts disulfures. Le clivage des précurseurs membranaires de ces facteurs de croissance est dû à l'activation de métalloprotéases appartenant aux familles ADAM et matricielles, en réponse à divers stimuli, en particulier à l'activation de récepteurs couplés aux protéines G.

Ce phénomène d'activation indirecte, semble jouer un rôle important dans le pouvoir oncogénique de ces récepteurs

Ils jouent un rôle très important dans la croissance, la différenciation et le contrôle du métabolisme cellulaire. (30–32) .

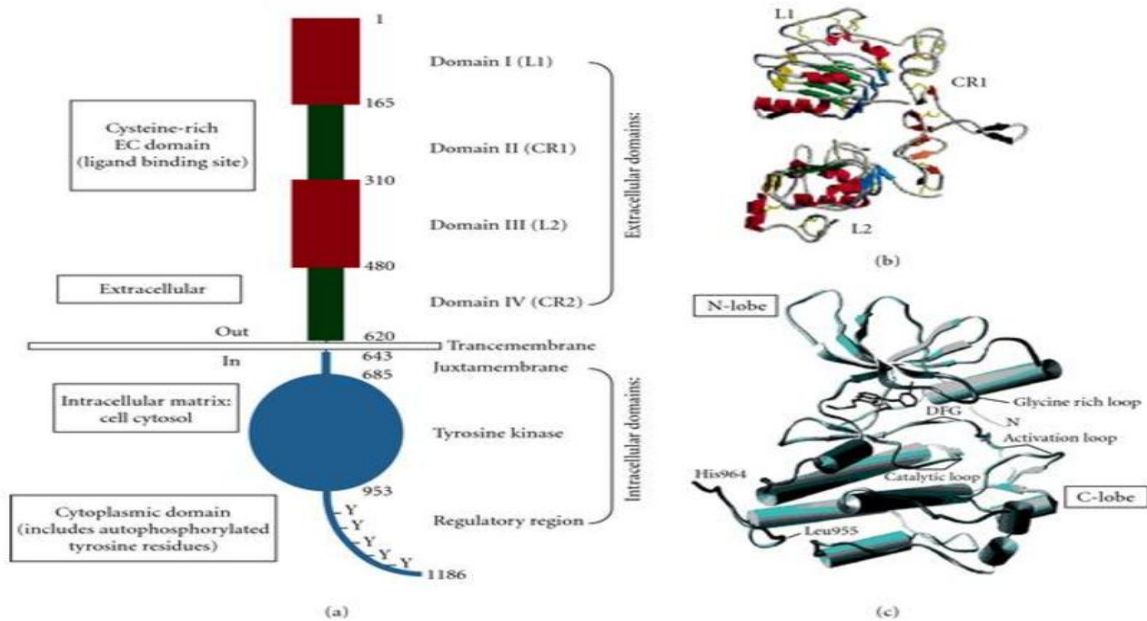
Les récepteurs ErbB partagent une forte homologie de structure primaire (40-45 %) qui se traduit par une organisation structurale commune caractéristique (Figure 10): la région extracellulaire ou ectodomaine est composée de quatre sous-domaines (I à IV) répétés deux à deux et se succédant en alternance. Les domaines I et III (L1 et L2) sont très semblables et analogues à un domaine du récepteur de l'IGF1 (insulin-like growth factor I), un autre récepteur à activité

tyrosine kinase. Les domaines II et IV sont, eux, caractérisés par leur abondance en résidus cystéines, d'où leur appellation CR1 et 2 (CR : cysteine-rich). La région extracellulaire est suivie du domaine trans membranaire, courte séquence de 23-24 acides aminés, principalement caractérisée par la très grande hydrophobie des acides aminés la composant.

Cette séquence d'ancrage membranaire est suivie du côté cytoplasmique par un domaine juxta membranaire dont le rôle n'est pas entièrement élucidé, un domaine à activité tyrosine kinase, très conservé aussi bien pour la famille des récepteurs à tyrosine kinase que pour les tyrosine kinases solubles et les protéines, et lipides-kinases en général. Enfin, on décrit un domaine C-terminal, riche en sites de phosphorylation, très important pour la transduction du signal « EGF »(29,33).

**Tableau I: Les facteurs de croissance de la famille de l'EGF**

Facteurs de croissance	Récepteurs reconnus	Année de découverte
EGF ( <i>epidermal growth factor</i> ; = urogastrone)	EGFR	1962
Amphiréguline (AR)	EGFR	1988
TGF- $\alpha$ ( <i>transforming growth factor</i> $\alpha$ )	EGFR	1982
Epigène (EPG)	EGFR	2001
Bêtacelluline (BTC)	EGFR, ErbB4	1993
HB-EGF ( <i>heparin-binding EGF</i> )	EGFR, ErbB4	1991
Epiréguline (EPR)	EGFR, ErbB4	1995
Neurégulines (NRG, 4 gènes, nombreuses isoformes)	ErbB3, ErbB4	1992
Récepteurs de facteurs de croissance	Ligands reconnus	Année de découverte
EGFR ( <i>EGF receptor</i> )= ErbB1, HER1	Tous sauf NRG	1978, cloné en 1984
ErbB2 ou HER2, neu (oncogène de rat)	Aucun	cloné en 1984
ErbB3 ou HER3	NRG1 et 2	cloné en 1993
ErbB4 ou HER4	BTC, HB-EGF, EPR, NRG1 à 4	cloné en 1993



**Figure 11: Structure de l'EGFR démontrant les domaines pertinents**

## II.2. Structure tridimensionnelle

La détermination de la structure des protéines membranaires en général est particulièrement difficile. Si elles représentent environ le tiers des gènes dans les génomes publiés à ce jour, moins de 1 % des structures déposées dans la banque Protein Data Bank (<http://www.rcsb.org/pdb/>) sont des protéines membranaires.

Pour les récepteurs de la famille ErbB plusieurs structures cristallographiques ont été établies au début des années 1990 : le domaine kinase intracellulaire de l'EGFR complexé ou non avec un inhibiteur, ainsi que les domaines extracellulaires de l'EGFR en présence et en absence de ligand, d'ErbB2 et d'ErbB3 en l'absence de ligand(39).



**Figure 12: Structure cristalline de l'EGFR**

Les structures cristallographiques de l'ectodomaine de l'EGFR lié à l'EGF ou au TGF $\alpha$  ont confirmé l'implication des domaines I (L1) et III (L2) dans la fixation de ligands. Elles ont surtout révélé l'existence d'interactions directes entre les domaines II (CR1) de chaque monomère, au niveau de ce qui a été baptisé « bras d'interaction ». (Figure 11)

En l'absence de ligand, les structures cristallographiques des ectodomains de l'EGFR et d'ErbB3 ont montré que ce bras de dimérisation est bloqué par l'existence d'interactions entre les domaines II et IV ; on dit que le récepteur adopte alors une configuration « contrainte » ou « fermée »(34)

La structure cristallographique révèle l'existence de monomères dont la configuration ressemble à celle de l'état activé de l'EGFR. Il n'y a pas d'interactions entre les domaines II et IV, et le bras de dimérisation du domaine II est exposé. Cette structure expliquerait pourquoi ErbB2 est le partenaire privilégié des interactions d'hétérodimérisation, puisque sa structure est en quelque sorte « prête » pour interagir avec un autre récepteur en configuration ouverte. De plus, cette structure est effectivement incompatible avec la liaison d'un ligand du fait de l'éloignement des domaines I et III.

La structure du domaine tyrosine kinase de l'EGFR a été déterminée en présence ou non d'un inhibiteur de petite taille. Ce domaine kinase a une structure globale tout à fait comparable à celle de toutes les autres kinases cristallisées .Il est constitué de deux lobes, l'un, petit, du côté N-terminal qui comprend des feuilletts bêta et une hélice alpha, et l'autre, plus grand, du côté C-terminal qui comprend surtout des hélices alpha. Les deux lobes sont séparés par un sillon où se fixe l'ATP ou ses analogues. Le lobe C-terminal comprend la boucle activatrice qui, pour la plupart des récepteurs à tyrosine kinase, recouvre le sillon ATP en l'absence de ligand, et change de conformation après phosphorylation d'une ou plusieurs tyrosines(35). Dans le cas de l'EGFR, cette boucle activatrice ne contient qu'une seule tyrosine qui ne semble pas participer au mécanisme d'activation, puisque sa mutation est sans effet. Dans la structure cristalline, cette boucle adopte en fait une conformation similaire à celle de la boucle activatrice phosphorylée du récepteur de l'insuline. La régulation de l'activité kinase semble donc dépendre d'autres mécanismes que l'habituelle phosphorylation de la boucle d'activation, probablement de changements conformationnels consécutifs à la dimérisation(36).



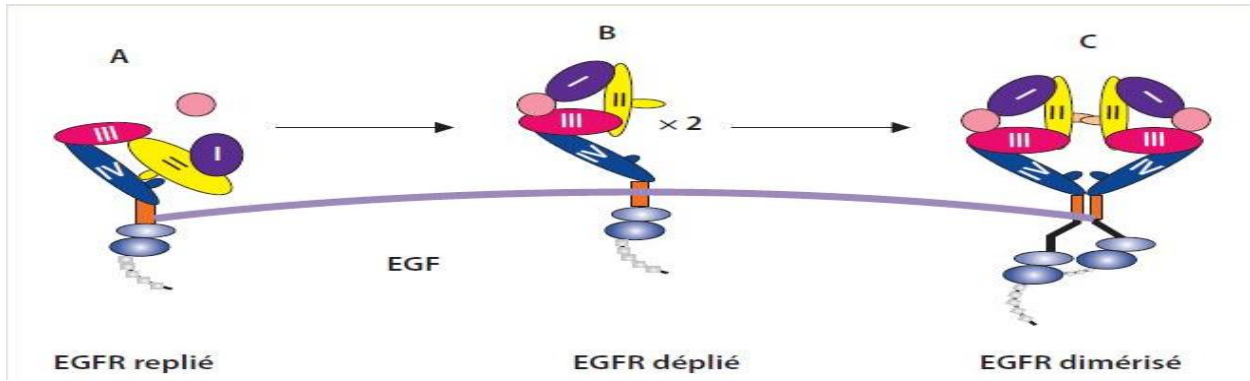


Figure 13: Représentation schématique de l'activation d'EGFR sous l'effet et d'un ligand

### II.3. Rôles physiologiques

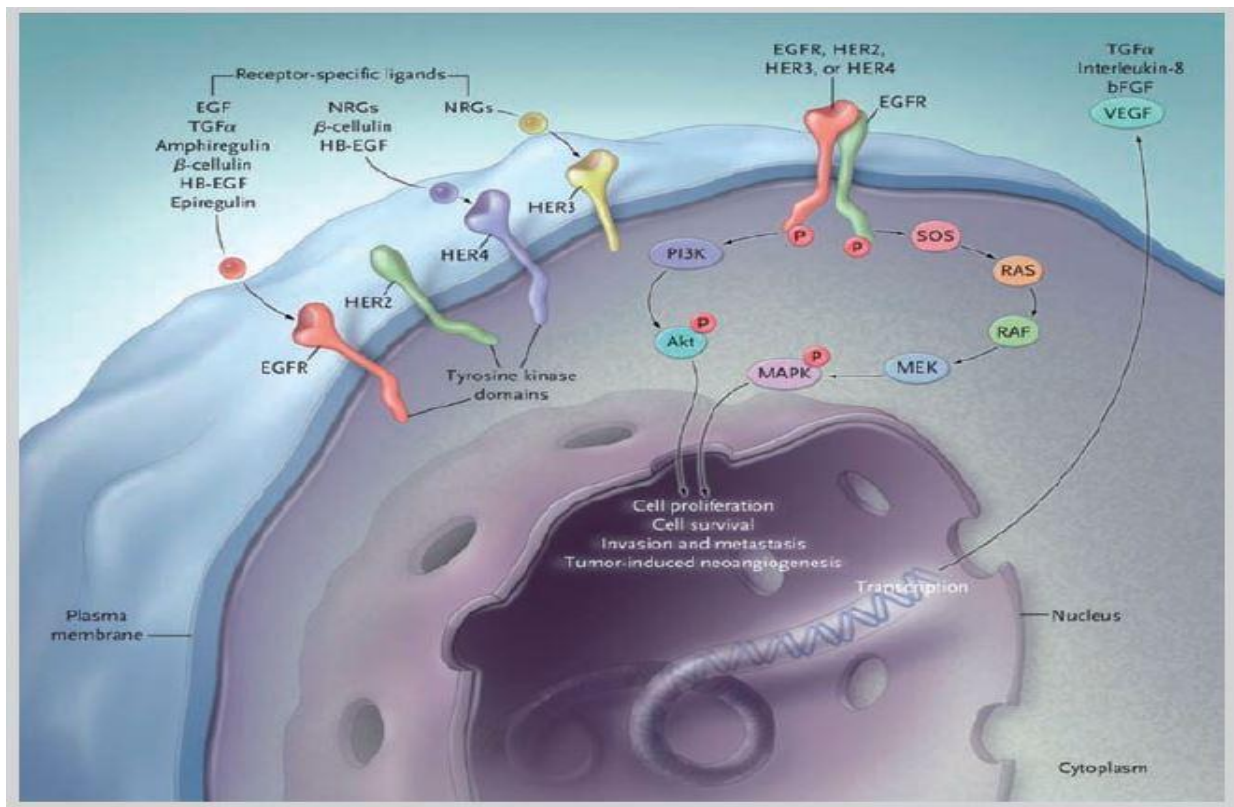


Figure 14: Récepteurs : EGFR (HER-1), HER-2, HER-3 et HER-4

Au niveau cellulaire, les voies de signalisation activées par les récepteurs ErbB participent à la prolifération, à la migration et à la différenciation de beaucoup de types de cellules. Elles ont aussi un effet anti-apoptotique. La répartition tissulaire des récepteurs ErbB et de leurs ligands, ainsi que des expériences d'inactivation ciblant ces gènes chez la souris (Tableau II), indiquent un rôle très important dans le développement du système cardiovasculaire (ErbB2 en particulier), du système nerveux (ErbB3, ErbB4 et les neurégulines), des glandes mammaires, ainsi que des épithéliums en général(33,37). Chez l'adulte, ces récepteurs restent aussi impliqués dans le fonctionnement normal de ces organes et tissus.

**Tableau II: Modèles de souris transgéniques**

<b>Inactivation du gène (knock out)</b>	
EGF	phénotype normal.
TGFa	développement normal, mais anomalies peau e fourrure.
Neuréguline	† < E11. Anomalies développement systémene et cœur.
EGFR	embryons petits à E10-17, retard de développement des épithéliums (yeux,poumons) † < 8 jours.
ErbB2	† < E11. Anomalies développement systémene et cœur.
ErbB3	† < E13. Anomalies développement système nerveux.
ErbB4	† < E11. Anomalies développement du cœur.
<b>Transgénèse</b>	
Il existe de nombreux modèles de cancer, en particulier du sein surexpression d'Erb-B2. Des <i>knock out</i> conditionnels ont aussi confirmé l'importance des récepteurs ErbB dans le développer du cœur, du muscle et du système nerveux.	
† : décès	

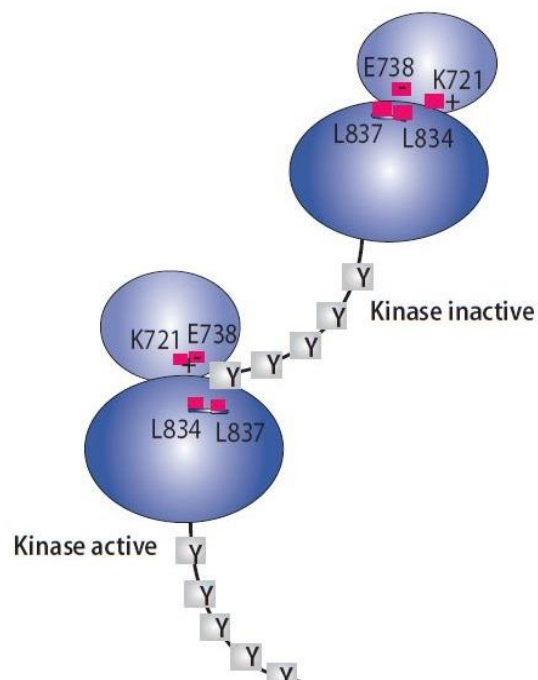
Ce dernier point est illustré par les effets indésirables des thérapeutiques ciblant ces récepteurs. Le trastuzumab (Herceptin®), anticorps dirigé contre le domaine extracellulaire d'ErbB2, a une toxicité cardiaque. Les inhibiteurs de l'activité kinase de l'EGFR provoquent souvent des rashes cutanés caractéristiques.

## II.4. Activation des récepteurs

Le mécanisme de l'activation des RTK, en particulier ceux de la famille ERBB, est maintenant bien connu(36) :il fait intervenir une dimérisation du récepteur et une autophosphorylation croisée de résidus tyrosine du domaine C-terminal du récepteur. Les résidus phospho-tyrosine ainsi formés sont par eux-mêmes porteurs d'information et constituent le message intracellulaire. Les résidus phospho-tyrosine sont reconnus par des protéines adaptatrices, dotées de motifs de reconnaissance spécifiques appelés "motifs SH2", et le changement de leur conformation, lié à la fixation de ce motif sur un résidu phospho-tyrosine, leur confère la possibilité d'activer d'autres protéines, et de transmettre ainsi l'information

La dimérisation du récepteur est ainsi la première étape de son activation(38). Cette dimérisation, est induite par le ligand, EGF(ou un autre facteur de croissance de la famille): celui-ci reconnaît un motif situé à cheval sur les sous-domaines I et III ; sa fixation à ce niveau va provoquer un changement au niveau de la partie extracellulaire du récepteur (Figure 13B) .Un motif d'interaction homophile de ce sous-domaine, qui pourra se lier à un motif identique d'une autre molécule de récepteur ayant elle aussi fixé une molécule de ligand(Figure 13C).

L'autophosphorylation du récepteur est permise par cette dimérisation(38) .L'activité catalytique de l'un puisse s'exercer sur les résidus tyrosine C-terminaux de l'autre( **Figure 15**) Le domaine intracellulaire est fait de 2 lobes, N et C, entre lesquels se trouve une fente appelée "poche à ATP" Le phosphate en  $\gamma$  sera fixé sur une tyrosine du récepteur adjacent. L'entrée de la poche à ATP est conditionnée par l'état d'activation du récepteur (fermée au repos et s'ouvre lors du changement de conformation induit par la dimérisation, permettant ainsi à la réaction enzymatique de se produire(39).



**Figure 15:Représentation schématique de l'autophosphorylation**

Au niveau intracellulaire, les résidus tyrosine des différents récepteurs, au nombre d'une dizaine pour chaque récepteur, ne jouent pas le même rôle une fois qu'ils ont été phosphorylés, parce qu'ils ne sont pas reconnus par les mêmes protéines dotées d'un domaine SH2(40).

C'est ainsi que plusieurs phosphotyrosines de l'EGFR sont reconnues par les protéines GRB2, SHC et PLC $\gamma$  principalement, alors que les phosphotyrosines d'ERBB3 sont reconnues surtout par la protéine p85, associée à la sous-unité catalytique de la PI3 kinase ; il en résulte une certaine spécificité dans la nature des voies de signalisation activées par les différents récepteurs : la voie des MAP kinases, en aval de GRB2, est activée en réponse à l'activation d'EGFR, lui-même active entre autres par l'EGF ; alors que la voie de la PI3 kinase, en aval de p85, est activée en réponse à l'activation d'ERBB3 dans le cadre de son hétérodimérisation. Dans une cellule qui reçoit le message apporté par un facteur de croissance donne, la disponibilité relative de tel ou tel récepteur de la famille ERBB peut ainsi aboutir à l'activation de voies de signalisation distinctes : de la sorte, des cellules différentes n'obéissent pas aux mêmes ordres.

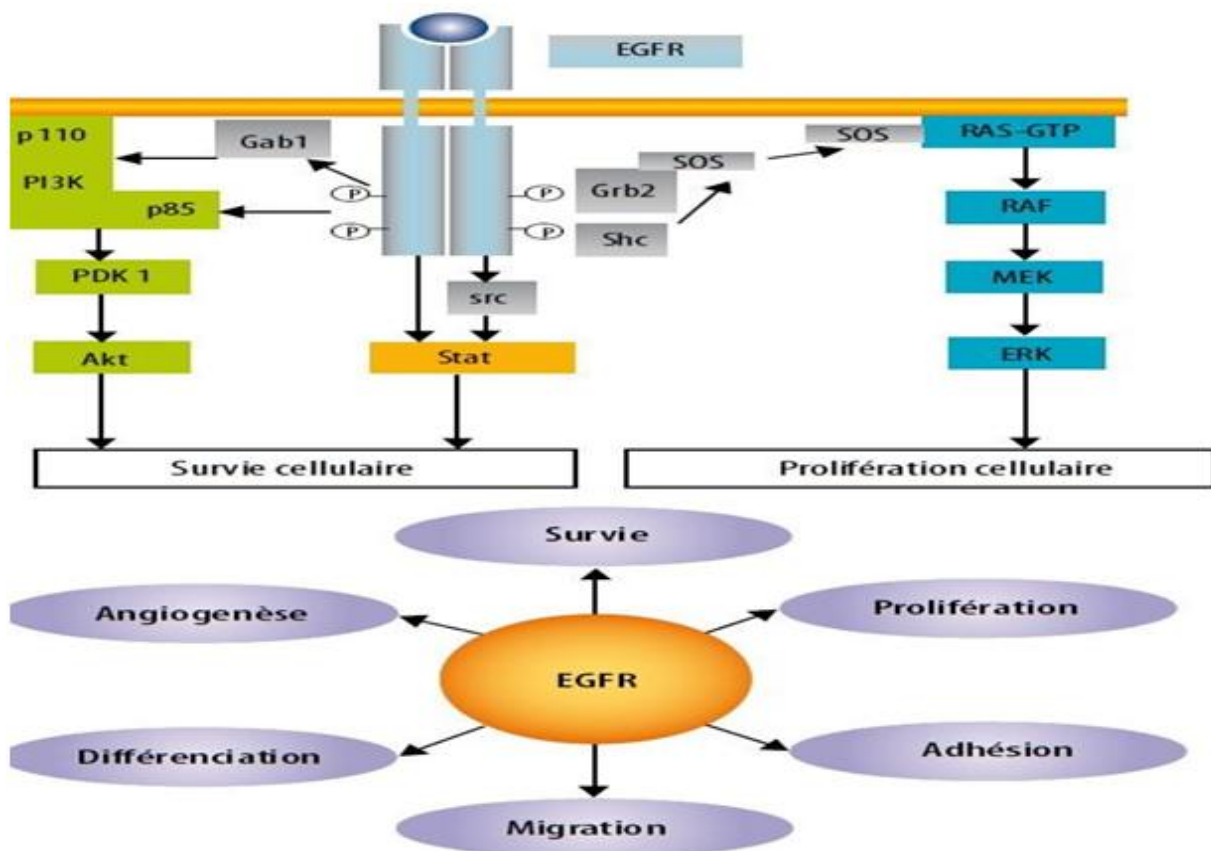


Figure 16: Principales voies sous-jacentes et fonctions d'EGFR

### III. Ciblage thérapeutique des récepteurs ERBB

Deux types d'approche thérapeutique peuvent être conçues pour atteindre les récepteurs ERBB dans les cancers épithéliaux : des anticorps dirigés contre la partie extracellulaire du récepteur(41) ; et des petites molécules capables d'inhiber la fonction catalytique des récepteurs : les inhibiteurs de tyrosine kinase (ITK) (42,43).

**Tableau III:Exemples de molécules ciblant les récepteurs ErbB**

Molécule	Cible	Cancer	Laboratoire	Statut
<b>Anticorps</b>				
Trastuzumab (Herceptin®)	ErbB2	Sein	Genentech/Roche	Approuvé
Pertuzumab (Omnitarg)	ErbB2	Ovaire, sein	Genentech	Phase II
Cetuximab (Erbix®)	EGFR	Côlon	ImClone Systems / Merck	Approuvé
Matuzumab (EMD72000)	EGFR	Rein	Merck	Phase II-III
Panitumumab (ABX-EGF)	EGFR	Côlon, rein	Abgenix	Phase II
<b>Inhibiteurs de TK</b>				
Gefitinib (Iressa®)	EGFR	Poumon	Astra Zeneca	Approuvé
Erlotinib (OSI 774, Tarceva®)	EGFR	Poumon	Roche/Genentech/OSI	Approuvé
EKB569 (irréversible)	EGFR + ErbB2	Côlon, poumon	Wyeth-Ayerst	Phase II-III
Lapatinib (GW2016)	EGFR + ErbB2	Sein	GlaxoSmithKline	Phase III
Canertinib (CI1033) (irréversible)	Pan-ErbB	Sein, poumon	Pfizer	Phase I/II
AEE788	EGFR + ErbB2, VEGF-R		Novartis	Phase I

### III.1. Anticorps

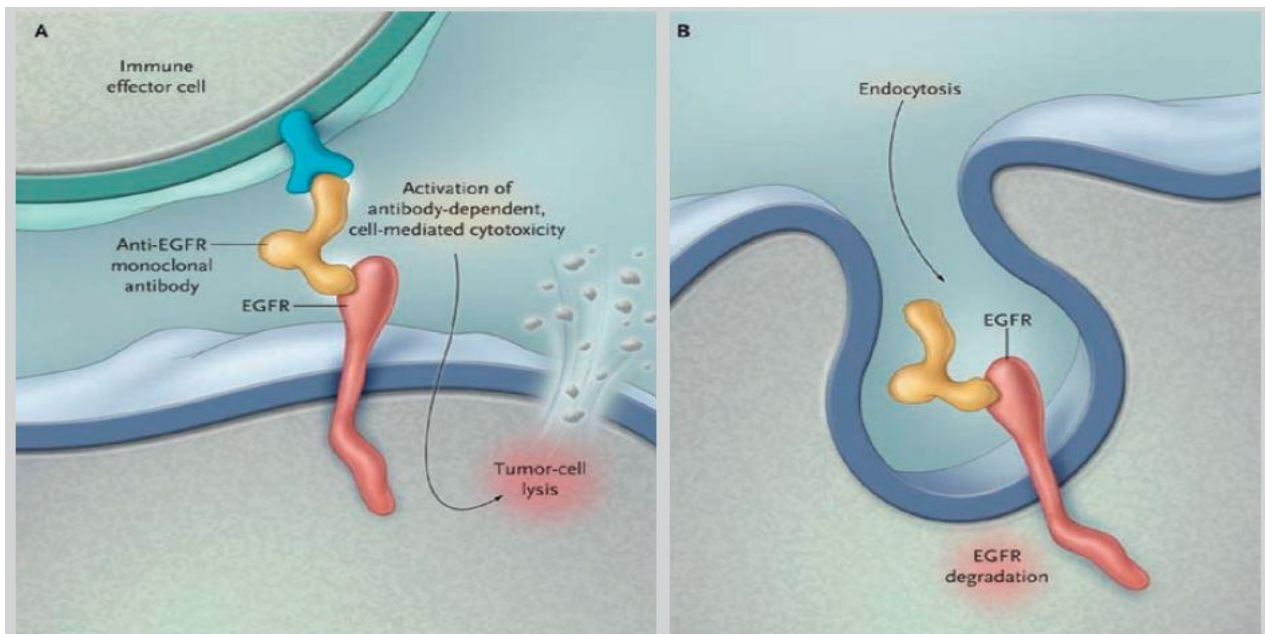
Des anticorps ont été développés contre les domaines extracellulaires des récepteurs de la famille ERBB. Le cetuximab (ErbixR) est disponible pour le traitement des cancers colorectaux métastatiques et celui des cancers des VADS localement avancés ou récidivants, et le panitumumab (VectibixR) pour le traitement des cancers colorectaux métastatiques

L'expression du récepteur n'est pas corrélée à l'activité de ces anticorps. Ces 2 anticorps, ainsi que le matuzumab reconnaissent des épitopes localisés au niveau du domaine III, empêchant le dépliement. Et interrompent la signalisation en aval de l'EGFR ; comme le cetuximab est une immunoglobuline G1, cela lui confère en outre une possibilité d'action immunologique grâce à sa reconnaissance par les récepteurs Fc $\gamma$ , alors que le panitumumab est une immunoglobuline G2.

En ce qui concerne le récepteur ERBB2, le trastuzumab (HerceptinR) est utilisé dans les cancers du sein en situation métastatique.

L'amplification du gène ERBB2 dans les cancers gastriques les rend sensibles au trastuzumab.

Ces anticorps reconnaissent des épitopes distincts : le trastuzumab reconnaît le domaine IV et serait efficace contre les seuls homodimères, alors que le pertuzumab, dirigé contre le domaine II, qui contient le bras de dimérisation, inhiberait l'hétéro-dimérisation. Les 2 anticorps ont une action synergique *in vivo* du fait de cette complémentarité,

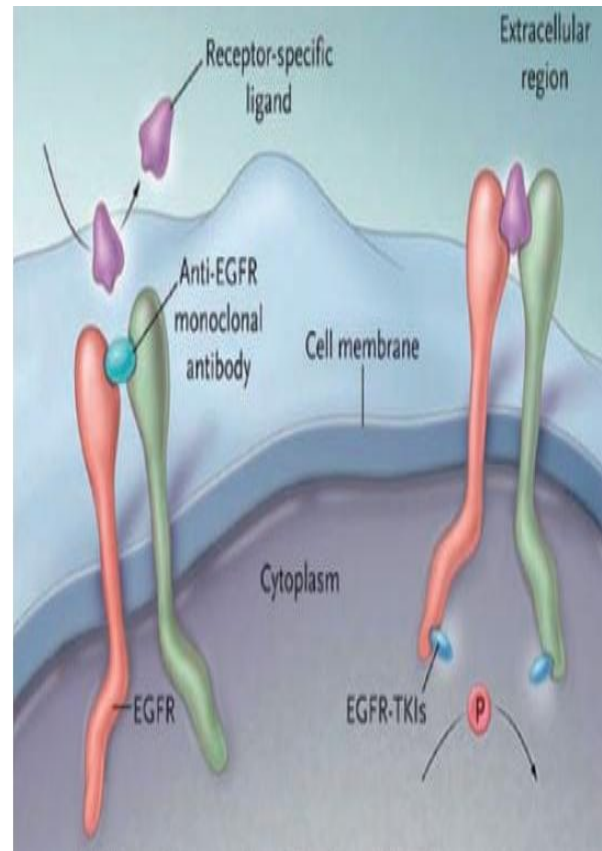


**Figure 17: Cytotoxicité et internalisation du récepteur EGFR induites par le cetuximab**



### III.2. Inhibiteurs de tyrosine kinase

Deux inhibiteurs de tyrosine kinase spécifiques d'EGFR, caractérisés par la présence d'un noyau quinazoline, ont été développés initialement et sont maintenant disponibles pour le traitement des cancers du poumon non à petites cellules : le gefitinib (Iressa®) et l'erlotinib (Tarceva®). Ce sont des ATP-mimétiques prenant la place de l'ATP dans la "poche" destinée à le recevoir. On s'est rapidement aperçu qu'ils n'étaient actifs que dans les cancers présentant une mutation activatrice du domaine kinase : ces mutations, en effet, outre leur rôle activateur donnant au récepteur une activité constitutive, sensibilisent les tumeurs aux ITK en raison précisément de l'accessibilité de la poche à ATP du récepteur mute. Par ailleurs, sous traitement, on peut voir émerger des clones résistants aux ITK, caractérisés par une mutation supplémentaire de l'EGFR, T790M, qui modifie le site où se loge



**Figure 18: Molécules inhibitrices du récepteur EGFR**

L'ATP et en interdit l'accès à l'inhibiteur. L'industrie pharmaceutique développée actuellement des ITK capables de contourner ce mécanisme de résistance.

Des inhibiteurs moins spécifiques du récepteur EGFR ont été synthétisés. Le premier, le lapatinib (Tyverb®), est considéré comme bispécifique EGFR-ERBB2 et peut être prescrit dans le traitement des cancers du sein métastatiques ; d'autres, comme le neratinib ou l'afatinib, sont en cours de développement. Enfin, des inhibiteurs irréversibles (car capables de se lier de façon covalente aux récepteurs) et susceptibles de reconnaître tous les récepteurs ERBB ("pan-ERBB") sont en cours d'étude : citons par exemple le dacomitinib.

#### IV. MUTATIONS DE L'EGFR

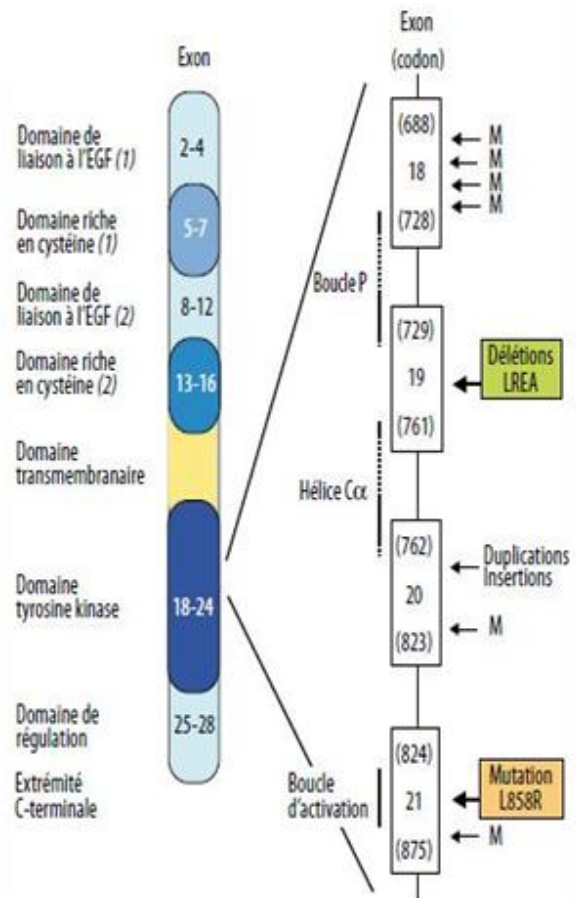
Le récepteur de l'EGF a été le premier récepteur membranaire clairement impliqué dans la cancérogenèse, avec la découverte de son homologie avec l'oncogène vErbB qui code pour une version tronquée et constitutivement active de ce récepteur.

On a découvert depuis que les récepteurs ErbB sont impliqués par différents mécanismes dans beaucoup de cancers humains, en particulier des tumeurs solides. Les altérations observées sont des mutations, des surexpressions (avec ou sans amplification du gène) ou une stimulation anormale par leurs ligands(45,46).

La mutation la plus fréquemment retrouvée est une délétion du domaine extracellulaire de l'EGFR, appelée type III (vIII, commune dans les glioblastomes), qui provoque une activation de l'activité kinase en l'absence de ligand. La surexpression de l'EGFR,

particulièrement lorsqu'elle est associée à une co-expression d'un de ses ligands, est un facteur de mauvais pronostic pour différents cancers. Elle est fréquemment décrite dans les cancers de la tête et du cou, du poumon, du côlon, du pancréas, du rein, etc.

Une mutation activatrice d'ErbB2 a aussi été découverte dans des neuro-blastomes chimiquement induits chez le rat. Elle est située dans le domaine transmembranaire et provoque, elle aussi, une activation constitutive de l'activité kinase. Elle n'a jamais été retrouvée dans des cancers humains. L'implication d'ErbB2 dans la cancérogenèse est le fait de sa surexpression, souvent due à une amplification génique. Des surexpressions d'ErbB2 sont décrites dans les cancers du sein et de l'ovaire, des glandes salivaires, de l'estomac, etc.



**Figure 19: Répartition des principales mutations d'EGFR.**



La surexpression d'ErbB3 a été observée dans des cancers du sein, du côlon, de l'estomac et dans d'autres carcinomes. Son expression avec ErbB2 est, entre autres, un facteur de mauvais pronostic pour les cancers du sein.

Bien que le traitement par ITK permette dans certains cas des réponses spectaculaires, la plupart des patients finissent par présenter une rechute de la maladie, suite à l'acquisition d'une résistance au traitement. Les bases moléculaires de la rechute peuvent être multiples : activation d'autres voies de signalisation de prolifération cellulaire, recrutement de molécules sous-jacentes à EGFR indépendamment de son activation ou apparition d'une mutation secondaire de résistance. Cette dernière hypothèse a été confirmée pour la première fois par Kobayashi et al. chez un patient atteint d'un adénocarcinome bronchique avec une délétion dans l'exon 19 d'EGFR et qui avait été traité par gefitinib, avec une excellente réponse durant 24 mois, avant que la maladie ne progresse de nouveau(47). L'analyse moléculaire de la tumeur après rechute a permis de détecter une seconde mutation du domaine tyrosine kinase, T790M (exon 20), qui empêche l'ancrage des ITK. Une mutation analogue a été mise en évidence dans les LMC résistantes à l'imatinib.

Récemment, Engelman et al. ont observé un autre mécanisme de résistance(48). Les auteurs ont mis en évidence, sur des lignées cellulaires résistantes au gefitinib, une amplification de l'oncogène MET responsable de l'activation de la voie PI3 Kinase via la surexpression de HER3. Ainsi, la voie PI3-kinase, et donc la survie cellulaire, est activée indépendamment d'EGFR.

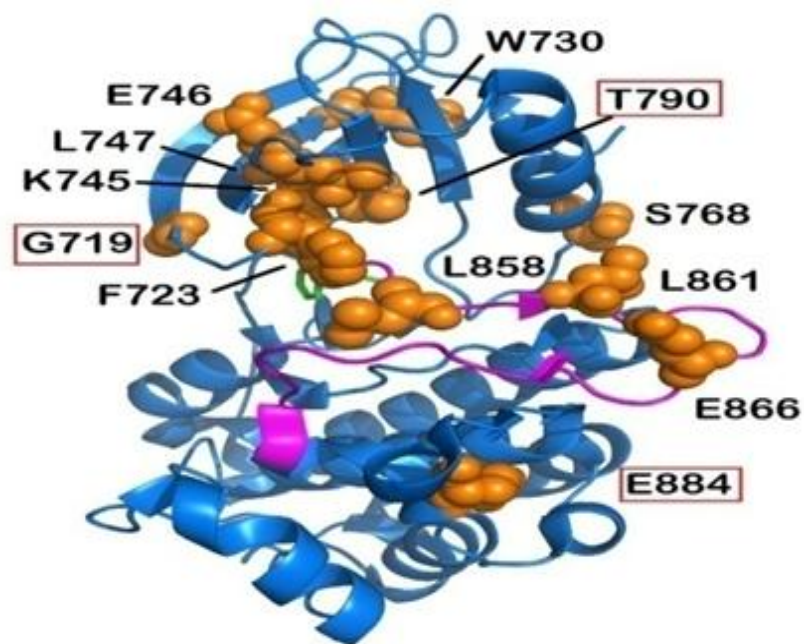
Enfin, rappelons que d'autres mutations sont associées à une résistance au traitement par ITK, notamment les mutations de KRAS, qui surviennent dans 10 % à 20 % des CBNPC(49).

#### IV.1. Conséquences fonctionnelles des mutations EGFR

Les délétions en cadre dans l'exon 19 surviennent juste en aval d'un résidu lysine en position 745 (K745), essentiel pour la fixation de l'ATP (Figure 20). La disparition de quelques acides aminés adjacents pourrait perturber la configuration du site catalytique d'EGFR et la fixation de l'ATP. La mutation L858R survient quant à elle à côté du motif très conservé DFG (acide aspartique, phénylalanine, glycine) dans la boucle d'activation. Ces mutations peuvent engendrer des changements conformationnels responsables d'une augmentation de l'activité du récepteur et de la sensibilité aux ITK.

Les travaux sur lignées cellulaires montrent que les mutants EGFR engendrent une réponse accrue et prolongée après stimulation par EGF, ainsi qu'une activation de voies sous-jacentes différentes par rapport aux réponses observées avec le récepteur sauvage (50). Par ailleurs, la survie des lignées porteuses de la mutation EGFR est dépendante de l'activité du mutant EGFR

Ces observations ont contribué à étayer le concept d'«addiction» au mutant EGFR : les cellules porteuses d'une mutation seraient constitutivement dépendantes de l'activité d'EGFR pour leur survie. Cette hypothèse est confortée par la constatation que l'inhibition de l'expression de l'EGFR par si RNA (*short-interfering RNA*) ou l'inhibition pharmacologique



**Figure 20: Représentation tridimensionnelle du domaine tyrosine kinase d'EGFR et de ses principales mutations**

# **CHAPITRE 3 :**

# **CRIBLAGE VIRTUEL**

## I. GENERALITE :

Le domaine du << drug design >> ou << conception de médicaments >> peut être exploré par la modélisation moléculaire et la conception de nouvelles molécules à l'aide des outils informatiques.

Il est aussi appelé drug design rationnel et peut être défini par un ensemble de processus inventifs capables de trouver de nouveaux médicaments basés sur les connaissances de cibles biologiques. Le médicament ou la molécule thérapeutique est généralement une petite molécule activant ou inhibant la fonction d'une biomolécule telle que les protéines et entraînant un bénéfice thérapeutique pour le patient. Fondamentalement, la conception de médicament inclut la conception de petites molécules de formes et charges complémentaires à la cible moléculaire avec laquelle elles interagissent. La conception de médicaments repose fréquemment sur les techniques de modélisation sur ordinateur dites aussi << in silico >>(51).

Ce type de modélisation est connu sous le nom de conception de médicaments assistée par ordinateur.

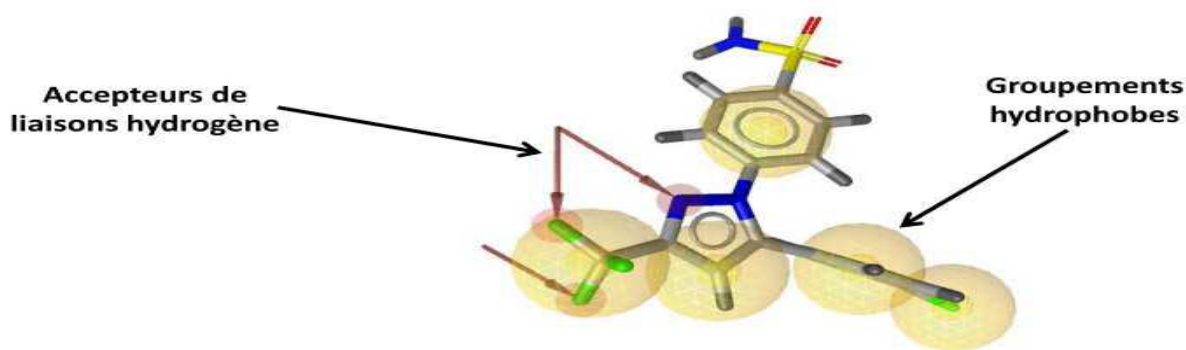
Les techniques de modélisation tentent de prédire l'affinité entre le ligand et sa cible et, à ce jour, les résultats sont globalement satisfaisants. Cependant, il existe d'autres propriétés qui doivent être optimisées pour que le ligand devienne sûr et efficace autant qu'il puisse l'être un médicament. Ces caractéristiques moléculaires que sont l'absence d'effets secondaires, la biodisponibilité, la demi-vie métabolique, la toxicité etc., sont par ailleurs assez difficiles à optimiser avec les techniques de conception de médicaments rationnelles.

Une cible médicamenteuse est une clé moléculaire qui se trouve dans un métabolisme particulier et évolue dans un cheminement de signalisation spécifique à une pathologie ou une infection, ou à la survie de pathogènes microbiens. Il existe des approches qui tentent d'inhiber le fonctionnement du cheminement de la maladie par l'intervention d'une molécule clé qui arrêterait ce fonctionnement en question. Des médicaments se liant à la région active et inhibant cette partie clé de la molécule peuvent être alors conçus. Une autre approche vise à améliorer le cheminement normal par la promotion de molécules spécifiques, qui affectent ce cheminement en question, dans un état pathologique. En plus de tout cela, les médicaments conçus doivent répondre à la propriété de sélectivité. Cette sélectivité est un terme relatif qui correspond au rapport de l'affinité du médicament pour la cible visée, dans un objectif thérapeutique, sur son affinité pour d'autres cibles. En d'autres termes, si le médicament ne répond pas à la caractéristique de la sélectivité, cela provoquerait de lourds effets secondaires(52).

## II. MODELES PHARMACOPHORQUES

Le concept de pharmacophore a été instauré pour la première fois en 1909, par Paul Ehrlich,(53) qui rapporte que certains groupes ou fonctions chimiques portés par une molécule sont responsables de l'effet biologique et que des molécules qui ont des effets similaires ont des fonctions chimiques similaires(54). Cinquante ans plus tard, le terme « pharmacophore » a été introduit par Schueler(55) qui le définit comme un système moléculaire qui porte (phoros) les caractéristiques essentielles responsables de l'activité biologique d'un médicament (pharmacon)(54,56,57).

La définition officielle donnée par l'IUPAC (International Union of Pure and Applied Chemistry) date de 1997 et indique qu'« un pharmacophore possède l'ensemble des propriétés stériques et électroniques d'une molécule nécessaires pour assurer l'établissement d'interactions supramoléculaires optimales avec une cible biologique spécifique et engendrer ou bloquer une réponse biologique » (58). Un pharmacophore ne représente pas une molécule ou une association de groupes fonctionnels mais un concept abstrait qui prend en compte les capacités d'un groupe de composés à réaliser des interactions moléculaires communes avec la structure de leur cible (54,57). Ces interactions sont définies comme les points pharmacophoriques : les donneurs et accepteurs de liaisons hydrogène, les groupements cationiques et anioniques réalisant des interactions ioniques, les groupements aromatiques conduisant à du  $\pi$ -stacking et les groupements hydrophobes donnant des interactions hydrophobes (54)(Figure 21).



**Figure 21: Un exemple de pharmacophore généré avec le logiciel LigandScout 4.0, représentant le ligand S58 (1CX2), un inhibiteur sélectif de COX-2 (flèches rouges : groupements accepteurs de liaisons hydrogène, sphères jaunes : groupements hydrophobes)**

Les approches pharmacophoriques ont été largement utilisées dans les processus de découverte de nouveaux médicaments, et ce notamment dans la phase d'optimisation(56).

Un modèle pharmacophorique peut être établi soit par une approche « ligand-based » en superposant des molécules actives et en extrayant les caractéristiques physicochimiques communes essentielles à leur activité biologique, soit par l'approche « structure-based » en recherchant les points d'interaction entre la cible et les ligands (56).

### III. APPROCHES DE LA CONCEPTION DE MEDICAMENTS :

Deux types d'approches de drug design sont connus. La première fait référence au ligand et se nomme <ligand-based drug design > et la deuxième se réfère à la cible moléculaire et se nomme <structure-based drug design >

#### III.1. Approche <<Ligand-based >>

Cette démarche repose sur les connaissances acquises concernant d'autres molécules interagissant avec la cible d'intérêt biologique. Ces autres composés peuvent être utilisés afin d'en dériver un pharmacophore qui définit le minimum de caractéristiques que doit posséder la molécule afin d'interagir avec le système biologique d'intérêt (59). Ici, le manipulateur détermine une relation de structure-activité quantitative (QSAR) dans laquelle une corrélation entre les propriétés des molécules calculées et leur activité biologique, expérimentalement déterminée, est établie. À leur tour, ces relations QSAR peuvent être utilisées pour la prédiction de nouveaux analogues.

#### III.2. Approche << Structure-based >>

Elle repose sur les connaissances de la structure tridimensionnelle de la cible biologique. Ces connaissances sont obtenues par des méthodes expérimentales telles que la cristallographie par rayons X et la spectroscopie par Résonance Magnétique Nucléaire (RMN). Ceci est possible si la structure de la cible est connue. Par contre, pour toutes les structures non déterminées expérimentalement il est possible d'obtenir un modèle, construit avec la modélisation par homologie, basé sur la structure expérimentale d'une protéine proche (même classe voire même famille). En utilisant la structure de la cible biologique, la conception de molécules ayant une grande affinité sélective à la cible se fait à l'aide d'outils graphiques et interactifs, de procédures de calculs automatisés et en comptant aussi sur l'intuition d'un chimiste médicinal. Avec le développement des méthodes expérimentales (Cristallographie aux Rayons X, spectroscopie RMN), la quantité d'informations des structures 3D des cibles biomoléculaires connaît elle aussi un grand essor. En même temps, les informations relatives

aux propriétés électroniques et à la dynamique structurale des ligands sont en continuelle expansion. Tous ces développements se répercutent positivement sur l'évolution des approches << structure-based >> de la conception des médicaments. (60)

Les pharmacophore générés dits « structure-based » s'appuient sur les informations structurales 3D de la protéine seule (approche basée sur le récepteur) ou Co-cristallisée avec un ligand (approche basée sur le complexe ligand-récepteur).

### **III.2.1. Approche basée sur le récepteur**

Dans cette approche, seules les propriétés du site actif du récepteur sont étudiées. Étant donné que la structure 3D de la protéine est dépourvue de ligand, il est nécessaire d'utiliser des ligands actifs connus afin de valider les modèles. Pour générer un modèle pharmacophorique, la démarche est la suivante(61): Après avoir déterminé le site actif de la protéine, (62) une image négative du site actif est créée.

Les images négatives présentent un très grand nombre de propriétés rendant quasi-impossible la génération d'un seul modèle pharmacophorique. Une analyse avec des ligands actifs connus permet de sélectionner les propriétés principales participant aux interactions ligand-récepteur. Pour cela différents modèles avec un plus petit nombre de propriétés sont générés et testés sur les ligands actifs pour choisir quel(s) modèle(s) utiliser pour cribler des chimiothèques.

### **III.2.2. Approche basée sur le complexe ligand-récepteur :**

Dans cette approche, la structure 3D de la protéine co-cristallisée avec un ligand est utilisée. Cette approche fournit des informations directes sur les interactions s'établissant entre le ligand et son récepteur. Un modèle pharmacophorique unique peut être directement généré avec la structure de complexe. Cependant, un seul modèle pharmacophorique généré avec une seule structure n'est pas assez fiable et il est donc recommandé d'utiliser les informations d'un ensemble de complexes différents. A partir de l'alignement de ces différents complexes, les groupes fonctionnels qui se retrouvent bien alignés permettent l'élaboration d'un modèle pharmacophorique commun à toutes ces structures. Lorsque plusieurs modes de liaisons existent pour un site actif, plusieurs modèles pharmacophoriques peuvent être générés. Pour générer un modèle pharmacophorique(63) la première étape consiste à préparer la structure de la protéine en tenant compte de la flexibilité de la protéine. Cette flexibilité peut être réalisée à partir de différentes conformations du complexe protéine-ligand obtenues par dynamique moléculaire ou par docking (64).

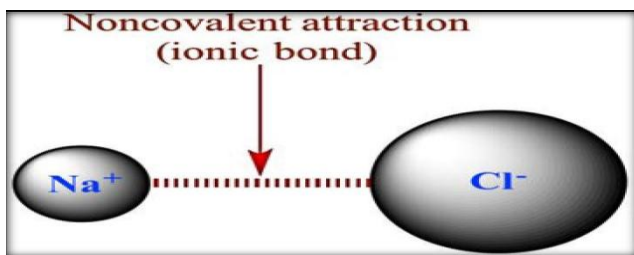
L'étape suivante consiste à déterminer le site de liaison grâce à la position du ligand co-cristallisé puis à identifier et sélectionner des points pharmacophoriques.

Pour cela, en partant du ligand co-cristallisé, les points pharmacophoriques réalisant des interactions favorables avec la protéine sont rapidement identifiés et localisés.

Les liaisons faibles permettent le repliement des molécules stables en raison du grand nombre de liaisons intramoléculaires accumulées. Elles interviennent aussi bien entre atomes d'une même molécule, qu'entre atomes de molécules voisines, ou encore entre molécules et solvant. Au niveau du site actif, le processus d'interaction se fait par liaison complémentaire du ligand avec l'architecture protéique. Cette association est assurée grâce à plusieurs types de liaisons faibles (65).

### III.2.2.1. Liaisons ioniques

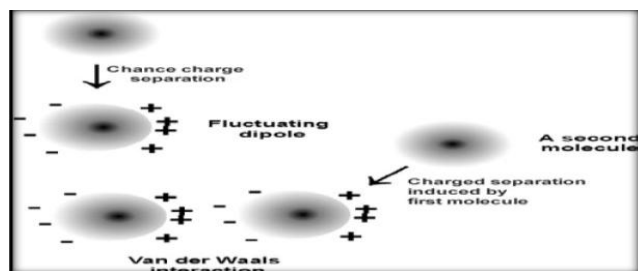
Une liaison ionique se crée entre deux atomes d'électronégativités très différentes (Figure 22). Un atome cède un ou plusieurs électrons de valence à l'atome le plus électronégatif, qui porte alors une charge entière négative. Celui qui a cédé le ou les électrons portés une charge entière positive (66).



**Figure 22 Liaison ionique(66)**

### III.2.2.2. Forces de Van Der Waals

Les interactions de Van Der Waals sont extrêmement faibles et se manifestent à des distances courtes. Les interactions entre les nuages électroniques de deux atomes adjacents conduisent à la présence d'une force attractive pour une distance de 4 Å (Error! Reference source not found.).



**Figure 23 : Forces de van der Waals(67).**

De part leur grand nombre, ces interaction jouent un rôle important dans la stabilisation de la structure des protéines et favorise le compactage(67).Les interactions de van der Waals, contribuant à la stabilisation du système avec une énergie comprise entre 0.4 et 4 kJ/mol (68).

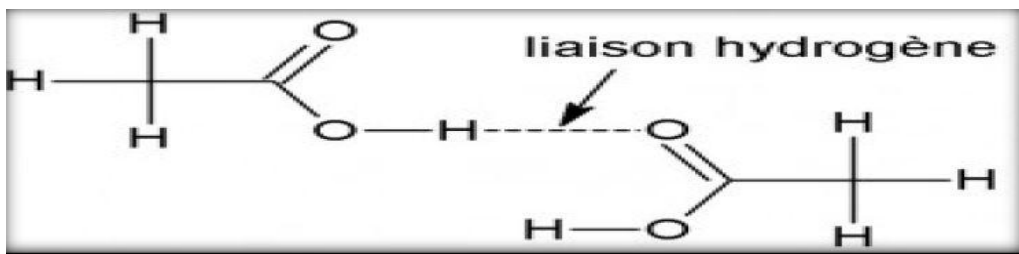


## III.2.2.3. Liaison hydrogène

Cette liaison, également appelée pont hydrogène, intervient lorsqu'un atome d'hydrogène lié par covalence à un atome électronégatif (le donneur) est attiré par un autre atome électronégatif (l'accepteur, généralement un atome d'azote ou d'oxygène) (Figure 24).

Ce type de liaison, pouvant être intra ou intermoléculaire, explique en grande partie l'arrangement spatial des macromolécules mais aussi l'importance de l'effet structurant de l'eau. Sa force est de l'ordre de 12 à 30 kJ/mol, Elle est assez élevée pour avoir des conséquences importantes et assez faible pour être réversible (68).

La liaison hydrogène est donc un élément essentiel dans la structure et la dynamique des complexes protéine-ligand, elle contribue à leur stabilité et joue un rôle important à la détermination de leur énergie (69).

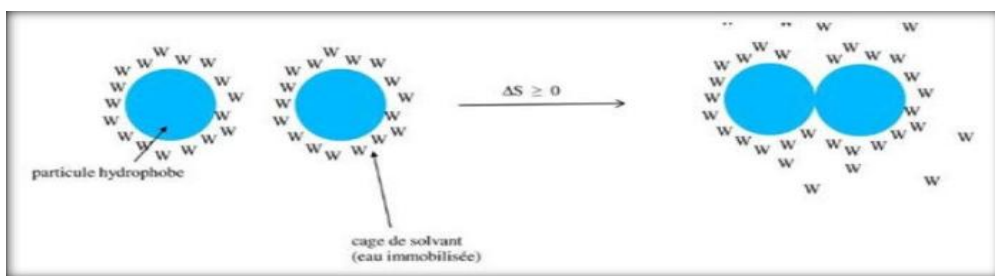


**Figure 24 :Pont hydrogène(68).**

## III.2.2.4. Interactions hydrophobes

Les molécules ou groupes non polaires ne sont pas capables de former de liaisons hydrogène et ne peuvent donc pas s'hydrater : on les nomme pour cette raison substances hydrophobes. L'effet hydrophobe est la tendance qu'ont ces groupes à se rassembler par coalescence de façon à minimiser les contacts avec l'eau. La force des interactions hydrophobes est de l'ordre de 20 à 30 kJ/mol.

Une molécule hydrophobe rompt localement le réseau de liaisons hydrogène du solvant et crée un état énergétiquement défavorable en diminuant le désordre. Les molécules d'eau se réorganisent alors autour des molécules hydrophobes de manière à minimiser l'énergie(65).



**Figure 25 : Interactions hydrophobes (65).**

## IV. DOCKING MOLECULAIRES :

L'action d'une molécule dans une protéine est gouvernée par des principes de reconnaissance mutuelle entre la molécule et sa cible. Les caractéristiques générales des interactions ligands - protéines résident en une complémentarité stérique, hydrophobe et polaire entre les deux structures et une conformation d'énergie favorable du ligand qui privilégie une bonne affinité de liaison(70). Les méthodes de docking ont pour but de prédire la conformation et l'orientation d'une molécule lors de sa liaison à son récepteur (71). Elles permettent de réaliser une analyse détaillée des modes de liaison supposées, nommées « poses », adoptés par le ligand dans le site actif de la protéine. Ainsi, l'objectif de cet outil est non seulement de discriminer les « vrais » ligands des molécules qui ne sont pas capables de se lier au récepteur, mais aussi de révéler leur mode de liaison avec la protéine cible (70).

Ainsi, les méthodes de docking peuvent être classées en trois catégories selon la prise en compte de la flexibilité du ligand, les algorithmes utilisés pour rechercher les poses possibles d'un ligand et les fonctions de score permettant d'évaluer l'affinité de liaison de chacune de ces poses (72).

Ainsi, la première catégorie considère le ligand et la protéine comme étant rigides, la seconde considère le ligand comme étant flexible tout en conservant la protéine rigide et la dernière catégorie considère la protéine « semi-flexible » et le ligand flexible (70,73)

### IV.1. Docking ligand rigide

Cette méthode reflète le principe de la clé et de la serrure (« lock and key ») énoncé par Emile Fisher en 1890 (74) où seul le ligand (représentant la clé) qui possède la bonne taille et la bonne forme peut ouvrir une serrure, la protéine (Figure 26) (75).

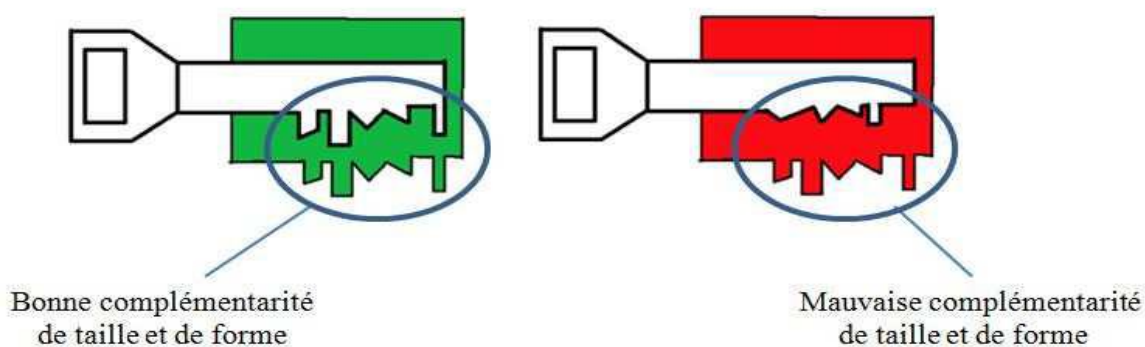
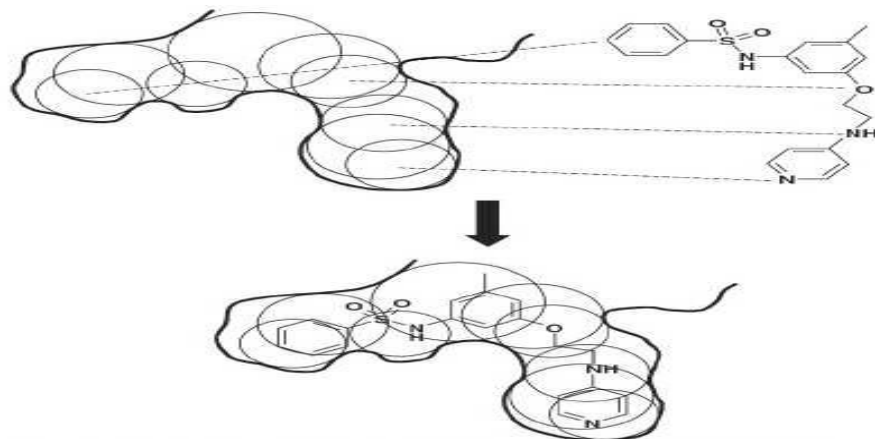


Figure 26:Schéma illustrant le principe de « lock and key »

Les premiers logiciels de docking se sont donc inspirés de ce concept et considéraient le ligand et le récepteur comme deux entités rigides. Dans cette approche, seuls les mouvements de translations et de rotations des ligands sont étudiés pour trouver les différentes poses au sein d'un site de liaison (Figure 27) (73).



**Figure 27: Principe du docking rigide illustré par le logiciel DOCK**

Le site de liaison de la protéine est distingué par l'utilisation des sphères de tailles différentes. Les correspondances entre les atomes du ligand et ces sphères sont recherchées et permettent de révéler une ou des orientations du ligand dans le site actif(72,76).

Pour améliorer la prédiction des modes de liaison, la prise en compte de la flexibilité Conformationnelle du ligand a été introduite. Ainsi, de nos jours, le docking ligand rigide est de moins en moins utilisé en faveur des méthodes de docking ligand flexible (73).

#### IV.2. Docking « semi-flexible »

En règle générale, la flexibilité des protéines est souvent négligée dans les méthodes de docking car ce concept est encore de nos jours mal défini et le temps de calcul est beaucoup trop long (73). Or, les différentes conformations que peuvent adopter une protéine doivent être prise en compte car une conformation particulière d'une protéine lui confère une fonction spécifique (73). En général, les protocoles de docking utilisent des complexes de protéine co-cristallisée avec un ligand. Cependant, ces complexes ne représentent qu'une petite partie de la flexibilité de la protéine et d'autres conformations pourraient se révéler être plus intéressantes. De plus, les changements conformationnels de la protéine tiennent un rôle majeur dans l'ajustement du ligand dans son site actif et la mise en place des interactions entre le ligand et son récepteur (73).

La capacité de simuler ces changements conformationnels avec des temps de calculs raisonnables est devenu un véritable challenge (73).

De nombreuses approches (73), connues sous le nom de docking « semi-flexible », ont été développées permettant une flexibilité partielle de la protéine.

La prise en compte de la flexibilité peut alors se réaliser à différents moments du docking :

#### **Avant le docking**

Les approches qualifiées d' « ensemble » considèrent un groupe de structures de protéine dans différentes conformations(70,77,78) obtenues par cristallographie aux rayons X ou RMN. Ensuite, le docking des ligands peut être réalisé dans chaque conformation séparément en ne gardant que le meilleur score pour chaque ligand, ou en utilisant quelques ou toutes les structures disponibles en un seul docking (77).

Une autre approche est d'utiliser des méthodes de dynamique moléculaire ou de Monte Carlo(71,73,77–79). La dynamique moléculaire permet de capturer des images instantanées de la protéine des différentes conformations observées au cours du temps. Ces dernières peuvent être ensuite utilisées pour réaliser un docking.

#### **Au cours du docking**

Ces approches consistent à utiliser des bibliothèques de rotamères(71,73,78–80). Ces bibliothèques explorent les degrés de liberté des chaînes latérales des résidus du site actif de la protéine pour trouver la combinaison optimale entre les chaînes latérales des résidus et l'orientation du ligand. Cependant, d'autres approches constituent des bibliothèques de rotamères en se limitant aux mouvements de torsions des résidus pour trouver la complémentarité optimale du ligand au cours du docking. Enfin certaines approchent combinent la flexibilité des chaînes latérales et celle du squelette des résidus de la protéine pour constituer ces bibliothèques (77).

#### **Après le docking**

L'approche « soft-docking » permet d'induire de petits changements conformationnels dans le site actif de la protéine grâce à une fonction de score plus indulgente (« soft ») qui autorise des clashes stériques entre les atomes du ligand et les atomes de la protéine. Cette fonction donne alors une certaine estimation de la « plasticité » du récepteur (78). L'objectif est de trouver la bonne orientation du ligand et des résidus dans le site actif de la protéine (77,78). L'approche soft-docking a été introduite dans de nombreux programmes de docking (77,80) et peut être utilisée en fin de docking lors de l'attribution de scores en utilisant la fonction de score « soft ».

### IV.3. Docking ligand flexible

L'un des grands enjeux des méthodes de docking flexibles (70) est d'explorer l'espace conformationnel des ligands dans le but de déterminer d'une part les bonnes poses du ligand dans le site actif de la protéine et d'autre part de trouver les modes de liaison correspondant à une énergie libre de liaison basse (70).

#### IV.3.1. Algorithmes de recherche

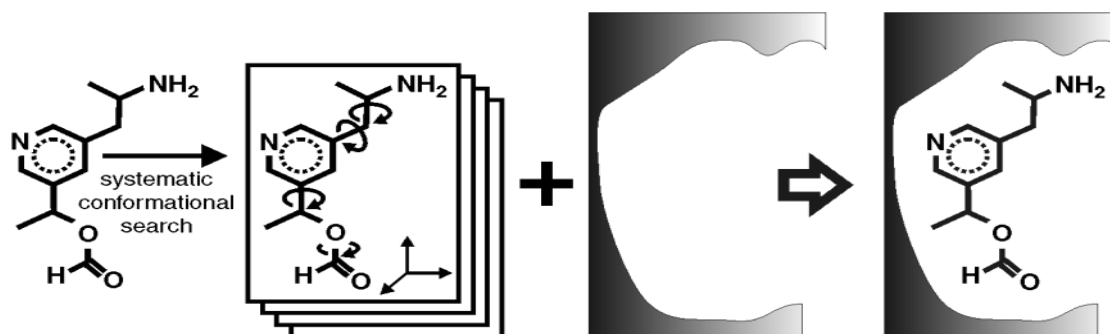
Les algorithmes de recherche étudiant la flexibilité conformationnelle des ligands sont généralement classés en trois catégories : les algorithmes de recherche systématique ou combinatoire, les algorithmes de recherche aléatoire ou stochastique et les algorithmes de recherche déterministe ou de simulation (70,79).

##### IV.3.1.1. Algorithmes de recherche systématique ou combinatoire

Ces algorithmes prennent en compte tous les degrés de liberté d'une molécule en explorant toutes les liaisons rotatives possibles. Ainsi, en fonction du degré de flexibilité de la molécule, le nombre de conformations générées peut rapidement s'élever et entraîner le phénomène d'explosion combinatoire (71).

Pour éviter l'explosion combinatoire, deux méthodes ont été développées : la méthode de recherche exhaustive et la méthode de fragmentation reconstruction.

La méthode de recherche exhaustive passe par une définition de limites imposées soit par le site de liaison, soit par des contraintes géométriques et/ou chimiques imposées aux conformations initiales des ligands à docker. Dans le premier cas, les mouvements de translation et de rotation systématiques des ligands sont générés dans un site de liaison putatif et prédéfini, limitant ainsi l'espace de recherche conformationnelle du ligand. Cette méthode est employée par le logiciel EUDOC (Figure 28) (73). Une autre méthode utilisée dans le docking ligand rigide consiste à générer une bibliothèque de conformations du ligand en utilisant des programmes tel qu'OMEGA. Ensuite, chaque conformation subit des mouvements de rotation et translation dans le site actif de la protéine de telle sorte qu'à chaque mouvement, le déplacement d'atomes ne dépasse pas une valeur seuil prédéfinie. Là aussi, grâce à cette délimitation, le risque d'explosion combinatoire est réduit (Figure 28) (73)



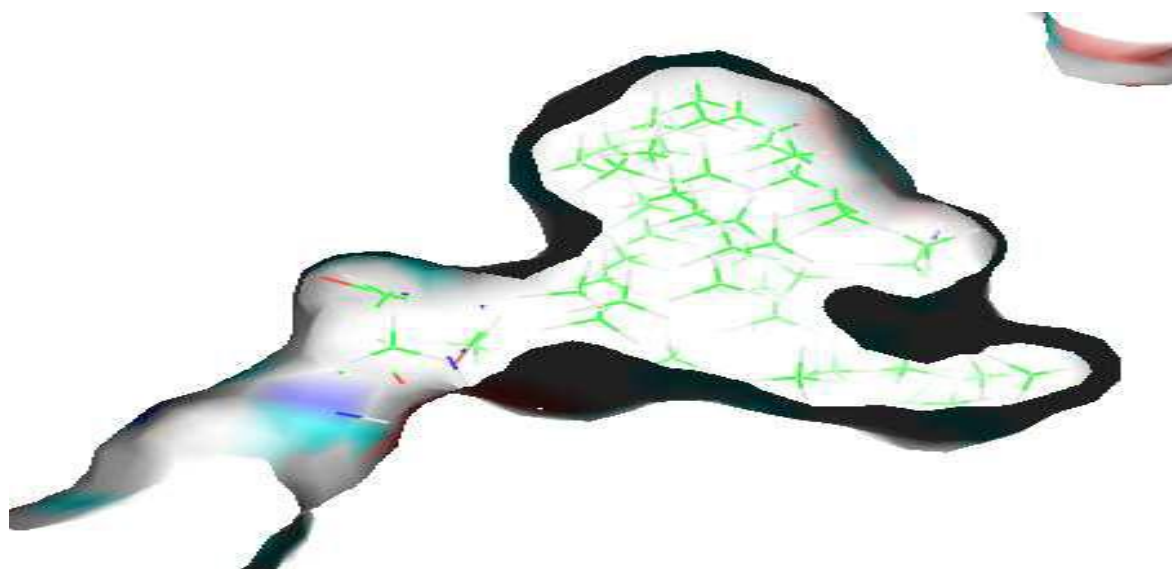
**Figure 28: Illustration de la méthode de recherche exhaustive employée par EUDOC**

Les méthodes de fragmentation-reconstruction consistent à diviser la molécule en fragments qui sont soit rigides soit représentés par un petit ensemble conformationnel [173]. Ensuite ces fragments peuvent être traités de deux manières : 1) placement des fragments un à un dans le site de liaison ou méthode de reconstruction incrémentale, 2) placement de tous les fragments de manière simultanée ou méthode « place and join » (73). L'objectif est ensuite de les lier dans une orientation favorable. Ces méthodes permettent de limiter le nombre de conformations à générer. En effet, en considérant une molécule avec 7 liaisons rotatives, si chaque liaison à 6 rotamères, le nombre de conformations à générer s'élève à 67 ( $> 250000$ ) conformations. Alors qu'en rompant une liaison rotative pour obtenir deux fragments, cela élimine la recherche conformationnelle pour une liaison rotative et les conformations des deux fragments sont indépendantes ce qui réduit le nombre de conformations à  $63 + 63 = 432$  conformations (81).

La méthode de reconstruction incrémentale est généralement constituée de 3 étapes :

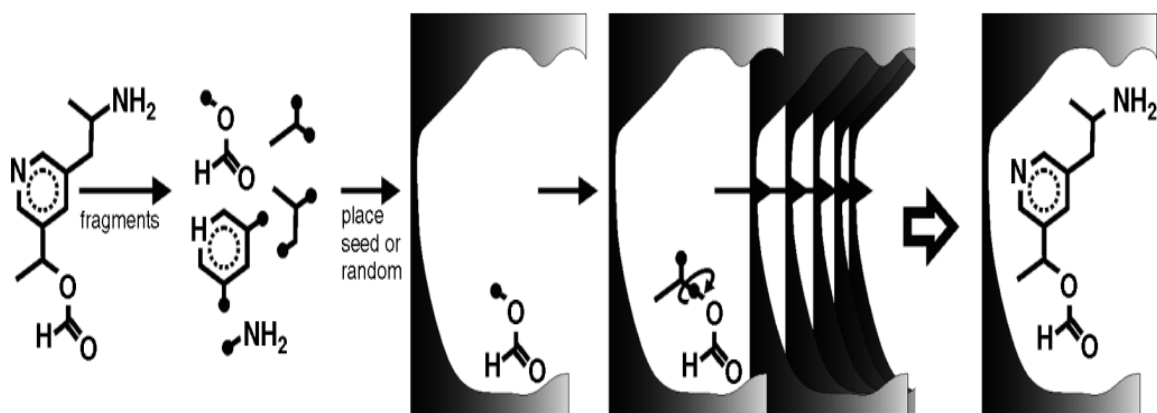
- 1) sélection de fragments rigides,
- 2) placement des fragments rigides dans le site actif,
- 3) construction incrémentale (73).

De nombreux logiciels se basent sur cette méthode (73), notamment le logiciel Surflex-Dock (81). Le logiciel commence par générer un « protomol », un ligand idéal qui occuperait tout l'espace du site actif de la protéine. Pour cela, 3 types de fragments ( $\text{CH}_4$ ,  $\text{C}=\text{O}$ , et  $\text{NH}$ ) sont placés dans le site actif, optimisés pour permettre l'établissement d'interactions avec le site de liaison et scorés. Seuls les fragments ayant obtenus les meilleurs scores sont retenus pour construire le protomol (Figure 29).



**Figure 29: Protomol généré par Surfex pour la cyclooxygénase 2 (1cx2)**

Une fois le protomol généré, les molécules sont fragmentées de sorte à ne conserver que les liaisons aromatiques (81). Puis, pour chaque fragment, un ensemble de conformations (en général 20) (81) est généré. Enfin dans l'étape de construction incrémentale, chaque fragment est aligné avec le protomol de sorte à maximiser la similarité avec le protomol (81). Les fragments ayant obtenus un bon score, les « heads », sont optimisés localement pour s'adapter au mieux dans le site actif. A partir de ceux-ci, les conformations du fragment suivant (« tails ») sont alignées par rapport au protomol de sorte à être liées au premier fragment (« heads ») (81). Cet alignement tient compte des contraintes imposées par le « head » et seules les conformations respectant ces contraintes et démontrant une similarité avec le protomol sont retenues (81). Le protocole continue jusqu'à la reconstruction totale de la molécule de départ. Enfin pour chaque pose déterminée, un score est attribué (81) (Figure 30).



**Figure 30: Schéma illustrant la méthode de reconstruction incrémentale (70).**

La méthode « place and join » décrite par DesJarlais et collaborateurs (82) consiste dans un premier temps à représenter le site de liaison par un ensemble de sphères. Ensuite les ligands sont fragmentés et dockés séparément dans le site actif. C'est à cette étape que la flexibilité des fragments est prise en compte. Puis, ils sont rassemblés de sorte à reformer le ligand initial en prenant en compte des contraintes de distances. En considérant deux fragments, il faut que les atomes de chaque fragment soient suffisamment proches pour pouvoir être joints ensemble. Cette étape détermine les différentes orientations des fragments et leur assemblage génère ainsi les différentes poses. Enfin l'énergie de chaque pose est minimisée(73,79,82).

#### IV.3.1.2. Algorithmes de recherche aléatoire ou stochastique

Ces algorithmes prennent en compte les mouvements de translation, rotations et torsions réalisés de manière aléatoire (71). Les changements sont acceptés ou rejetés selon une fonction de probabilité. Les approches les plus connues sont les méthodes de Monte-Carlo, les algorithmes génétiques et la recherche tabou (71).

Les méthodes de Monte-Carlo consistent à chercher les différentes poses des ligands en explorant de manière séquentielle les mouvements de translation, rotations et torsions. Ainsi, chaque degré de liberté est déterminé un à un et de manière aléatoire (73). Après chaque changement, la nouvelle conformation est retenue si elle présente une énergie plus basse que la conformation précédente, sinon elle est soumise au critère de Métropolis (70,71,73).

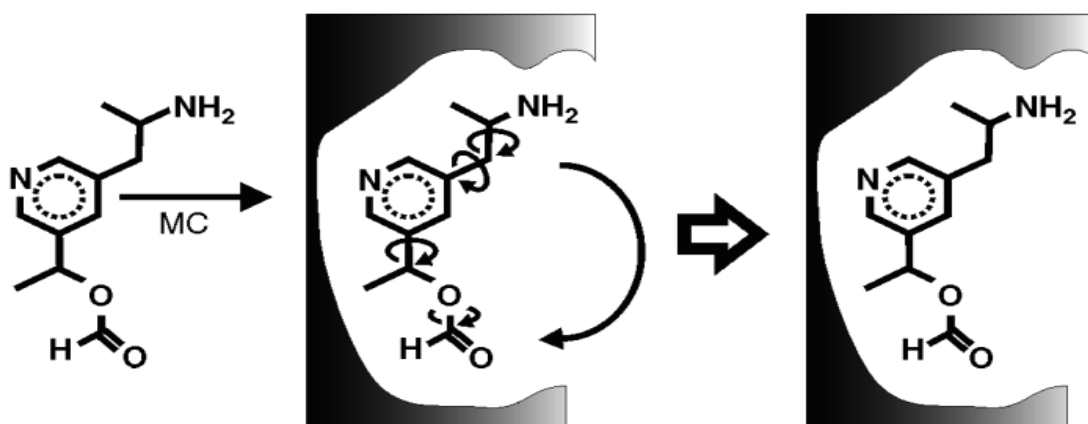
Ce critère est basé sur la fonction de probabilité de Boltzmann (Équation 1).

$$P = e^{\frac{E_{new} - E_{old}}{kT}}$$

**Équation 1: Probabilité P d'acceptation d'une nouvelle conformation (avec  $E_{new}$  : énergie de la nouvelle conformation,  $E_{old}$  : énergie de l'ancienne conformation,  $k$  : constante de Boltzmann,  $T$  : température de simulation) (70)**

Le processus est répété un certain nombre de fois, nombre qui définit le pas de Monte-Carlo, jusqu'à ce que le nombre de conformations désiré soit atteint (71). De nombreux logiciels (73,79) utilisent cette méthode tels que ICM(83) et GLIDE(84). Ce dernier utilise dans un premier temps un algorithme de recherche systématique pour trouver les poses des ligands et les poses ayant obtenu un bon score sont dans un second temps optimisées par la méthode de Monte-Carlo (73) (Figure 31).

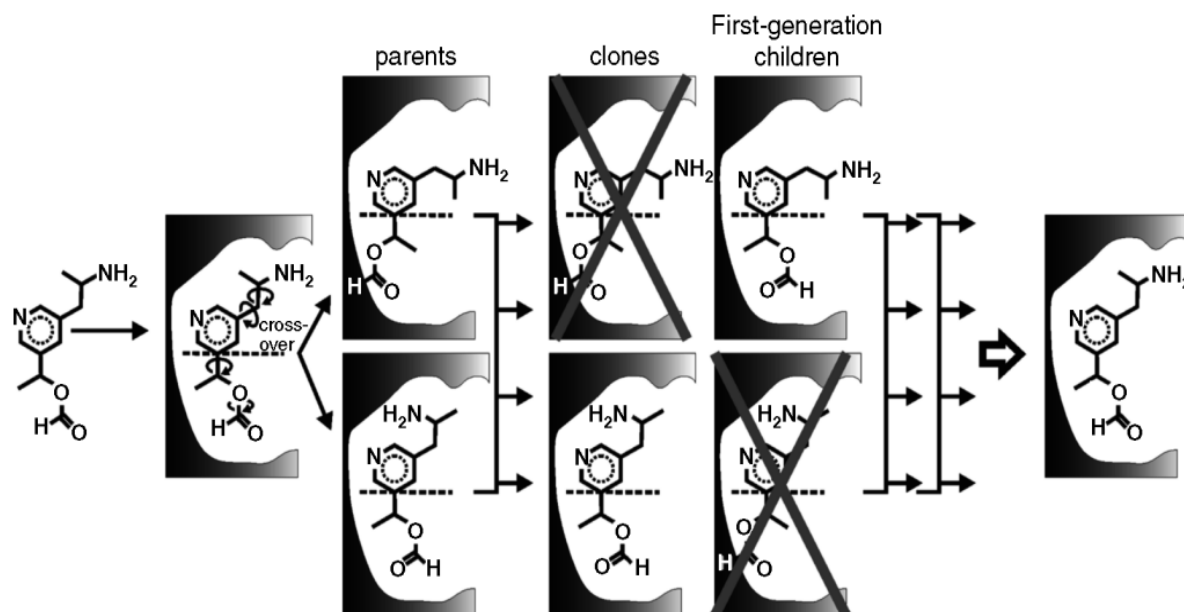




**Figure 31: Schéma illustrant la méthode de Monte-Carlo**

#### IV.3.1.3. Les algorithmes génétiques :

Ont pour but de mimer les principes de l'évolution naturelle afin de réduire le problème d'explosion combinatoire. Pour cela, ils utilisent des opérateurs génétiques tels que la mutation, la recombinaison et la sélection et une fonction de survie (ou « fitness function ») (85). Les algorithmes génétiques sont constitués d'une population de conformations où chaque membre représente un chromosome. Les gènes de ces chromosomes codent pour différentes variables, c'est-à-dire pour la translation du ligand, la rotation et les angles de torsions (73). La première génération de solutions appelée « population parent » évolue par recombinaison ou mutation en agissant respectivement sur chacun des deux parents (recombinaison) ou sur un seul (mutation). Les résultats de ces changements génèrent la population « enfants ». Une fonction de survie est ensuite utilisée pour sélectionner les meilleures conformations qui ont le plus de chance de survivre et de passer la génération suivante. Ces conformations sélectionnées constituent alors la nouvelle population « parent » qui subira à son tour les changements des divers opérateurs génétiques (73). Ce processus est répété jusqu'à obtenir une population finale qui répond à un seuil prédéfini par la fonction de survie (80). De nombreux logiciels utilisent cette méthode (73,80), l'un des plus populaires est GOLD(86) (Figure 32)



**Figure 32:**Schéma illustrant la méthode utilisant les algorithmes génétiques

#### IV.3.1.4. Les algorithmes de recherche tabou :

S'inspirent de l'idée d'éviter de réévaluer un espace conformationnel déjà exploré par l'algorithme de recherche (73). Pour cela des espaces de restrictions sont organisées sous forme de « liste tabou » constituant les conformations déjà obtenues (73,87). La recherche commence à partir d'une seule conformation d'un ligand puis différentes conformations sont générées grâce à une procédure similaire à la mutation qui réalise des changements de manière aléatoire. Ces conformations sont scorées et ordonnées selon une fonction d'énergie (87) : les conformations d'énergie plus basse qui ne sont pas encore contenues dans la liste tabou (« non-tabou ») sont ajoutées à cette liste, celles qui sont très similaires à la liste ne sont intégrées dans la liste que si et seulement si elles présentent une énergie plus basse (87). Enfin, les autres conformations ne répondant pas à ces critères, sont rejetées (71,73). Le processus continue en commençant par la dernière conformation retenue et s'arrête jusqu'à atteindre la limite d'itérations ou jusqu'à ce qu'aucun changement acceptable ne soit obtenu (87). Le logiciel PRO\_LEADS (88) est l'un des logiciel les plus connus utilisant cette approche (Figure 33). (73)

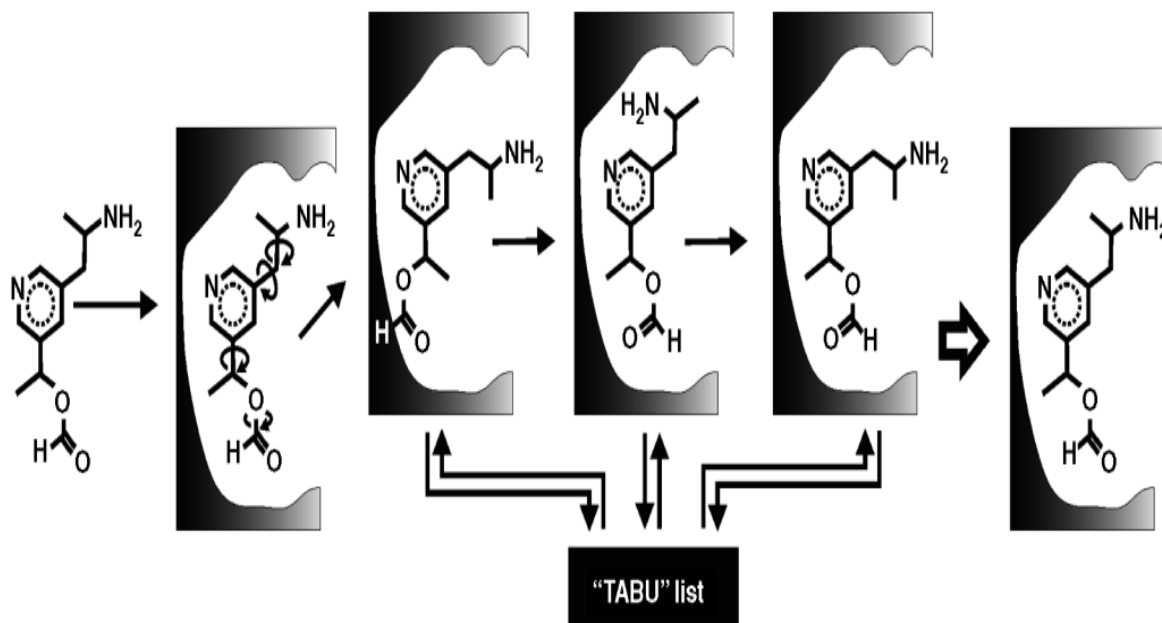


Figure 33: Schéma illustrant la méthode de recherche tabou

#### IV.3.1.5. Algorithmes de recherche déterministe ou de simulation

Deux méthodes principales les caractérisent : les méthodes de dynamique moléculaire et les méthodes de minimisation d'énergie.

Les méthodes de dynamique moléculaire se basent sur les équations de mouvements de Newton (Équation 2) (71,73) :

$$F_i = m_i a_i$$

**Équation 2 :** Equation du mouvement de Newton pour un système atomique ( $F$  : force,  $m$  : masse,  $a$  : accélération)

La force de chaque atome est déterminée à partir d'un changement dans l'énergie potentielle entre la position actuelle et une nouvelle position (Équation 3) (71) :

$$F_i = -\frac{dE}{r_i}$$

**Équation 3 :** Calcul de la force ( $F$ ) de chaque atome ( $E$  : énergie potentiel,  $r$  : distance entre les deux positions)

Les positions atomiques sont ensuite déterminées à des intervalles de temps très courts permettant d'obtenir la trajectoire des changements des positions atomiques au cours du temps (Équation 4) (71).

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i}$$

**Équation 4:** *Calcul des positions atomiques à chaque intervalle de temps ( $d^2r_i/dt^2$ ), avec  $F$  : force et  $m$  : masse atomique)*

Les méthodes de dynamique moléculaire présentent des temps de calculs trop longs et arrivent difficilement à franchir les barrières de hautes énergies (73).

En conséquence, l'identification du minimum global d'énergie se révèle être très difficile voire quasi-impossible avec ces méthodes et les conformations des ligands obtenues sont généralement piégées dans des minimaux locaux d'énergie (71,73,79). De plus, les résultats finaux dépendent grandement de la conformation initiale. Malgré tout, de nombreuses approches ont été développées pour pallier à ce problème (71,73,79). Une des méthodes consiste à réaliser des simulations de différentes parties du complexe ligand-récepteur en variant la température (89) tandis que d'autres proposent de lancer la simulation de dynamique moléculaire à partir de différentes positions initiales du ligand [171] afin d'augmenter les chances de trouver un minimum local d'énergie (73,79).

Les méthodes de minimisation d'énergie recherchent uniquement si des minimaux locaux d'énergie peuvent être atteints. Elles sont en général utilisées en complément d'autres méthodes (71).

## V. SCORING :

Les fonctions de score ont pour but de prédire l'affinité de liaison entre la protéine et chaque pose générée lors du docking (70,80). En effet, même si les conformations sont correctement prédites, il est primordial de pouvoir différencier les poses correctes des poses incorrectes (71). Le rôle principal des fonctions de score est donc d'évaluer et de classer les poses (70) pour identifier le « vrai » mode de liaison des composés (80) et les ligands les plus prometteurs (71).

### V.1. Fonctions de score

Les fonctions de score sont divisées en trois catégories : les fonctions de scores basées sur les champs de force, les fonctions de score empiriques et les fonctions de score basées sur les connaissances (71).

Les fonctions de score basées sur les champs de force caractérisent l'énergie d'une pose par la somme de deux termes énergétiques (71): l'une représentant l'énergie d'interaction ligand-récepteur et l'autre l'énergie interne du ligand. En général, ces fonctions de score se basent sur les champs de force de mécanique moléculaire qui permettent de calculer l'énergie potentielle du système (EMM). EMM est calculée par une somme de quatre termes (Equation 5) : les trois premiers termes estiment l'énergie interne du ligand en pénalisant les déviations de longueurs et d'angles de liaisons du ligand par rapport aux valeurs de références. Le quatrième terme reflète la contribution enthalpique estimée par les potentiels de Lennard-Jones et de Coulomb. Le potentiel de Lennard-Jones évalue les interactions de Van der Waals et le potentiel de Coulomb détermine les interactions électrostatiques (79)(Équation 5).

$$E_{MM} = \sum_{bond} K_r (r - r_{ref})^2 + \sum_{angles} K_\theta (\theta - \theta_{ref})^2 + \sum_{dihedrals} K_\phi [1 + \cos(n\Phi - \Phi_0)]^2 + \sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

**Équation 5 : Calcul de l'énergie potentielle d'un système (EMM) par les champs de force de mécanique moléculaire. Avec  $K_r$ ,  $K_\theta$  et  $K_\phi$  : facteurs de pénalité pour les liaisons, les angles et les angles dièdres ;  $r$  et  $r_{ref}$  : longueurs des liaisons mesurées et de référence  $\theta$  et  $\theta_{ref}$  : valeurs des angles mesurées et de référence ;  $\phi$  et  $\phi_{ref}$  : valeurs des angles dièdres mesurées et de référence ;  $A_{ij}$  et  $B_{ij}$  : constantes attractives et répulsives mesurées expérimentalement ;  $R_{ij}$  : distance entre les atomes  $i$  et  $j$  ;  $q_i$  et  $q_j$  : charges des atomes  $i$  et  $j$  ;  $\epsilon$ : constante diélectrique (79)**

En règle générale, les fonctions de score basées sur les champs de force simplifient l'équation EMM et se concentrent uniquement sur les interactions de Van der Waals et électrostatiques (Équation 6).

$$\Delta G_{bind} = \sum_{i=1}^{ligand} \sum_{j=1}^{protéine} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

**Équation 6 : Calcul de l'énergie libre de liaison ( $\Delta G_{bind}$ ) prenant en compte uniquement la contribution enthalpique évaluée par les potentiels de Lennard-Jones et de Coulomb**

Ainsi plusieurs champs de force de mécanique moléculaire tels qu'AMBER, CHARMM... ont été développés et introduits dans de nombreux logiciels (GOLD, AutoDock...) (73).

L'avantage de l'utilisation des fonctions de score basées sur les champs de force réside en leur rapidité de calcul mais leur grand inconvénient est qu'elles ne prennent pas en compte les effets entropiques et de solvation (71,79,80) ce qui a été amélioré dans différents logiciels (GOLD, AutoDock (71,80), HINT, (73), les versions de DOCK récentes (5 et 6) (73).

Les fonctions de score empiriques ont pour but de reproduire les données expérimentales d'énergie de liaison et utilisent une somme de plusieurs termes individuels non corrélés pondérés par des coefficients (Équation 7) (71,73,80). Les coefficients des différents termes sont calculés par des analyses de régression utilisant les données expérimentales d'affinité et les données structurales de cristallographie (80).

$$\Delta G_{binding} \approx \sum \Delta G_i f_i$$

**Équation 7 : Forme simplifiée d'une fonction de score empirique. (Avec  $\Delta G_{binding}$  : l'affinité de liaison,  $f_i$  correspond à un terme d'une interaction  $i$  entre le ligand et la protéine et  $\Delta G_i$  le coefficient associé) (70)**

L'avantage d'utiliser les fonctions de score empiriques est que leur forme est plus simple que celles basées sur les champs de force, leurs termes sont facilement déterminés (71) et les temps de calculs sont relativement rapides (80).

Cependant, cette méthode reste dépendante des données expérimentales (70,71) et les valeurs des coefficients varient entre différentes fonctions de scores empiriques pour un terme donné. En conséquence, les termes provenant de différentes fonctions de scores ne peuvent pas être combinés en une seule fonction (71).

Les fonctions de scores basées sur les connaissances ou « knowledge-based » ont pour but de reproduire les données expérimentales en utilisant des structures de complexes ligand-récepteur (71,80). L'idée provient du fait que dans un complexe obtenu par cristallographie, les atomes du ligand sont placés de manière optimale par rapport aux atomes de la protéine. Des analyses statistiques permettent de déterminer la fréquence et la distribution des paires d'atomes entre le ligand et la protéine et ainsi de définir des potentiels d'interaction (71,80). En effet, il a été postulé que les interactions les plus fréquemment observées contribuent positivement au processus général de mode de liaison tandis que les interactions les moins communes sont considérées comme globalement répulsives (73). Les avantages de ces fonctions « knowledge-based » résident sur leur simplicité et leur rapidité en temps de calcul. Cependant, leurs inconvénients, similaires aux méthodes empiriques, sont leur dépendance sur la qualité et la variété des données expérimentales (73).

De nombreuses fonctions de score basées sur la connaissance ont été développées. Les plus connues sont PMF, Drug Score, SMOG (71,73,80).

Les fonctions de scores sont souvent critiquées (70) puisqu'elles ne reflètent qu'une approximation de l'énergie de liaison libre du complexe ligand-protéine. De plus, ces fonctions utilisent des termes additifs, elles ont donc tendance à favoriser les molécules ayant un poids moléculaire élevé (> 250 g/mol) en leur attribuant un haut score. Des logiciels ont alors amélioré leurs fonctions de score en supprimant la corrélation entre le score et le nombre d'atomes (FLOG) (70).

De même, les fonctions de score pénalisent rarement les interactions défavorables et des améliorations ont été apportées en prenant en compte les clashes stériques (70). Cependant, d'autres points restent à améliorer tels que la prise en compte incomplète de l'occupation stérique du ligand dans la cavité du site actif et l'espace restant inoccupée, et la reconnaissance de ligands inactifs (70).

De nouvelles approches doivent donc être établies afin d'optimiser les fonctions de score (70). Le « consensus scoring » ou fonction de score consensus a été développé en se basant sur l'hypothèse que la combinaison de plusieurs fonctions pourrait dominer les faiblesses individuelles de chacune des fonctions et démontrer une meilleure performance en terme de scoring (73).

Trois principales méthodes de « consensus scoring » peuvent être décrites : les méthodes « Rank by vote » (attribution d'un « vote » pour une pose, pour chaque fonction de score du consensus qui la prédit dans les n premiers %), les méthodes « Rank by Rank » (calcul de la moyenne des rangs prédits par chaque fonction de score) et les méthodes « Rank by number » (calcul de la moyenne des scores prédits par chaque fonction de score) (73).

Les fonctions de score consensus ont été utilisées dans de nombreuses études révélant des résultats intéressants (90–92). Cependant, cette méthode présente un inconvénient majeur : utiliser différentes fonctions de score qui sont corrélées risque d'amplifier les erreurs de calculs plutôt que de les compenser (71,80). Il faut donc trouver de nouvelles stratégies réduisant ce problème et favoriser le développement de nouvelles fonctions de score consensus.

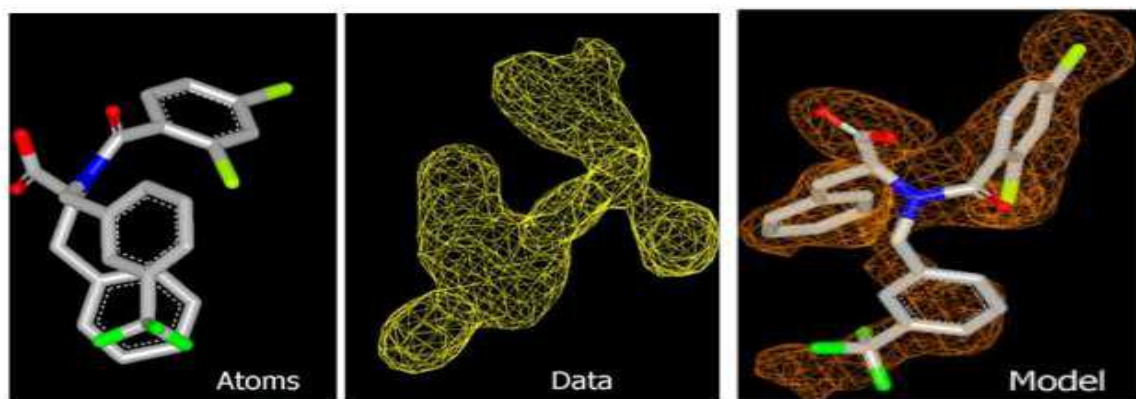
## V.2. **Autres paramètres importants à prendre en compte pour réaliser un docking**

### V.2.1. **Importance des structures de protéines**

L'un des éléments essentiels à considérer avant de réaliser un docking est la disponibilité et la qualité de la structure des protéines étudiées (73). En général, les sources d'information de ces structures proviennent majoritairement de la cristallographie aux rayons X et de la RMN.

La structure résolue par cristallographie aux rayons X est en fait un modèle construit en cherchant des correspondances entre les atomes et la densité électronique (93) (Figure 34). La qualité de la structure obtenue doit donc être évaluée à l'aide de différents paramètres. La résolution mesure la précision avec laquelle le modèle cristallographique reproduit les données expérimentales. En général, une structure cristallographique présentant une résolution inférieure à 2Å est considérée comme une source de donnée fiable (73). Un autre paramètre intéressant est le facteur R ou Rfree qui détermine le pourcentage de différence entre le cristal et le modèle. Enfin le facteur de température ou « B-factor » mesure le déplacement des atomes [193], et donne des indications sur la mobilité locale de la structure de protéine (93) et de la fiabilité du modèle (94). Les limites de la RMN résident dans la taille de la protéine (80-100 kDa) et les problèmes de solubilité limitent cette technique.





**Figure 34 : Processus de correspondance entre les atomes et la densité électronique pour produire un modèle cristallographique. (93)**

D'autres critères sont à prendre en compte concernant l'évaluation de la qualité des structures de protéines, notamment la position des hydrogènes, les états tautomériques et d'ionisation des résidus de la protéine (en particulier du site actif) et du ligand (73).

#### **V.2.2. Importance du solvant**

L'eau joue un rôle essentiel dans les interactions ligand-protéine (73). Les molécules d'eau peuvent notamment servir de pont entre la protéine et le ligand permettant de stabiliser le complexe. La disposition des molécules d'eau dans le site actif peut aussi permettre à la protéine d'adopter une conformation permettant d'accueillir le ligand. Beaucoup de logiciels de docking prennent en compte l'eau de manière implicite, par les effets de solvation (70) mais il existe un réel besoin d'explicitement l'eau. De nombreuses approches ont été développées dans ce but (Consolv du logiciel SLIDE (95), FlexX (96)). Leur objectif est de déterminer quelles molécules d'eau sont nécessaires à l'établissement des interactions ligand-protéine et devraient être retenues et lesquelles devraient être déplacées.

Les méthodes de docking, tout comme les modèles pharmacophoriques, sont très utilisées pour guider les scientifiques dans le développement de nouveaux médicaments. Le docking tente de distinguer « les vrais » ligands des ligands alternatifs grâce à une prédiction du mode de liaison. Cette prédiction se veut être la plus précise possible en reflétant au maximum les données expérimentales. Ainsi les algorithmes de recherche et les fonctions de score qui constituent les programmes de docking ne cessent d'être améliorés : introduction des molécules d'eau, qualité des modèles cristallographiques, prise en compte de la flexibilité de la protéine, fonction de score consensus... faisant de ces méthodes un outil remarquable, capable d'identifier des molécules intéressantes.

---

# **Partie pratique**

---

## **OBJECTIF**

L'objectif principal de cette étude est :

Rechercher des nouvelles molécules capables de cibler le récepteur du facteur de croissance épidermique (EGFR) par un criblage virtuel basé sur le ligand

L'objectif secondaire est:

Analyser in silico leurs propriétés physico-chimiques, pharmacocinétiques et toxicité potentielle.

# **MATERIEL ET METHODES**

## I. Matériel :

### I.1. Microordinateur :

Nous avons utilisé un micro-ordinateur de la marque **ASUS**® ayant les caractères suivants :

- Mémoire installée (RAM) : 16 Go
- Processeur : AMD Ryzen7 3800X 8-Core Processor
- Carte graphique : NVIDIA GeForce GTX970 (4Go)
- Système d'exploitation : Windows 10 Professionnel ; 64bits

#### Informations système générales

##### Édition Windows

Windows 10 Professionnel

© Microsoft Corporation. Tous droits réservés.



##### Système

Processeur : AMD Ryzen 7 3800X 8-Core Processor 3.89 GHz  
Mémoire installée (RAM) : 16.0 Go  
Type du système : Système d'exploitation 64 bits, processeur x64  
Stylet et fonction tactile : La fonctionnalité d'entrée tactile ou avec un stylet n'est pas disponible sur cet écran.

##### Paramètres de nom d'ordinateur, de domaine et de groupe de travail

Nom de l'ordinateur : DESKTOP-SQL7BIF  
Nom complet : DESKTOP-SQL7BIF  
Description de l'ordinateur :  
Groupe de travail : WORKGROUP

 [Modifier les paramètres](#)

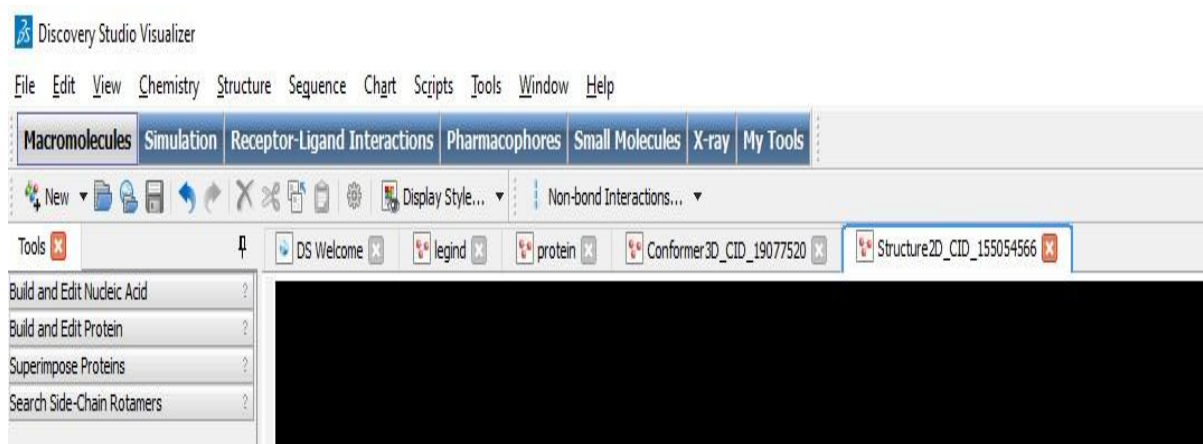
##### Activation de Windows

Windows est activé. [Lire les termes du contrat de licence logiciel Microsoft](#)

**Figure 35:informations générales du micro ordinateur utiliser dans le travail**

## I.2. logiciels:

### I.2.1. Biovia Discovery Studio 2019 :

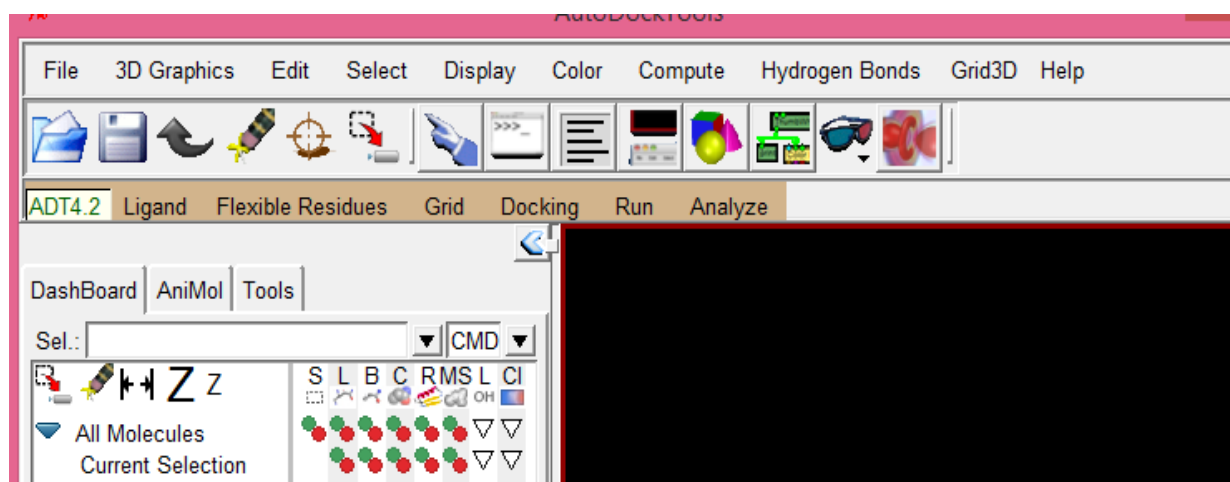


**Figure 36:L'interface du Biovia Discovery Studio 2019**

- Utiliser afin de séparer les complexes [Ligands-récepteur] en format pdb.
- Comparaison conformationnelle
- Représentation du Mode d'interaction des ligands dans le site actif de la protéine

### I.2.2. AutoDock :

L'outil graphique AutoDock mg Tools version 1.5.6 et ses accessoires (autodocksuite-4.2.5.1 et autodocksuite-4.2.6) est conçu pour faire le docking moléculaires, pour notre étude nous avons l'utilisé pour préparer les récepteurs et les ligands et les fournir en format pdbqt ; aussi pour définir Les cordones de site d'interaction récepteur-ligand (XYZ).



**Figure 37:L'interface d'AutoDock**

### I.2.3. AutoDock Vina :

AutoDock Vina est un programme open source pour faire de l'amarrage moléculaire et la prédiction considérablement précise du mode de liaison de ligands au site actif de la protéine (97). L'AutoDock Vina est intégré dans l'interface graphique « Pyrex »

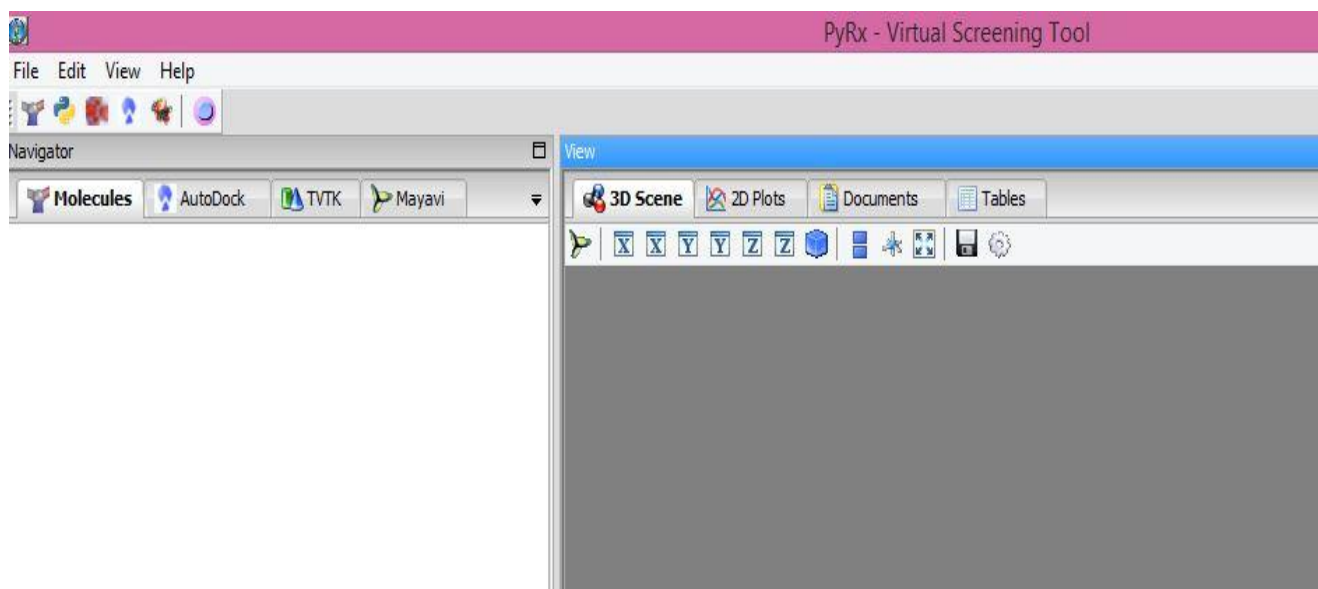


Figure 38:L'interface du PyRx

### I.2.4. Excel :

L'utilisation de ce logiciel « Excel » nous a permis de tracer et calculer respectivement la droite et le coefficient de corrélation linéaire.

### I.3. Serveur Online

#### I.3.1. SwissADME : (98)

Ce site Web nous permet de calculer des descripteurs physico-chimiques ainsi que de prédire les paramètres ADME, les propriétés pharmacocinétiques, la nature médicamenteuse et la convivialité de la chimie médicinale d'une ou plusieurs petites molécules pour soutenir la découverte de médicaments. (<http://www.swissadme.ch/>)

#### I.3.2. PreADMET : (99)

PreADMET est un serveur en ligne permettant de prédire les données ADMET. Dans le présent travail, ce serveur est utilisé pour prédire plusieurs critères régissant la toxicité potentielle de nos composés par deux tests : Test d'Ames et test de cancérogénicité chez les rongeurs. <https://preadmet.bmdrc.kr/>

### I.4. Banques des Données :

#### I.4.1. « PDB » (Protein Data Bank) : (100)

Cette ressource est alimentée par les informations d'archives de la Protein Data Bank sur les formes 3D des protéines, des acides nucléiques et des assemblages complexes qui aident les étudiants et les chercheurs à comprendre tous les aspects de la biomédecine et de l'agriculture, de la synthèse des protéines à la santé et à la maladie.

Le RCSB PDB s'appuie sur les données en créant des outils et des ressources pour la recherche et l'enseignement en biologie moléculaire, biologie structurale, biologie computationnelle et au-delà.

#### I.4.2. « PubChem »: (101)

PubChem est la plus grande collection au monde d'informations chimiques librement accessibles. Elle nous permet de rechercher des produits chimiques par nom, formule moléculaire, structure et autres identifiants ; et trouver des informations concernant ces molécules comme leurs propriétés chimiques et physiques, des activités biologiques, des informations sur la sécurité et la toxicité, des brevets, des citations de la littérature et plus encore.



## II. Méthodes :

### II.1. Docking moléculaire (D.M) :

#### II.1.1. Préparation du récepteur et ligands:

##### II.1.1.1. Recherche bibliographique :

Sur la « protein data bank » on fait une recherche des EGFR, à partir des résultats apparus on choisit les complexes dont leurs résolutions est  $< 2\text{Å}$  et on les classe dans un tableau qui contient les informations suivantes : les complexes, leur ligand, la résolution et l'IC50

##### II.1.1.2. Sélection du complexe :

La sélection du complexe de choix et le ligand de référence on prenant en considération les meilleurs résultats entre les complexes étudiées précédemment

##### II.1.1.3. Téléchargement :

Le complexe sélectionné est ensuite téléchargé en format PDB à partir de la banque des données

##### II.1.1.4. Séparation :

On sépare le ligand de son récepteur puis on enregistre chacun sous format PDB

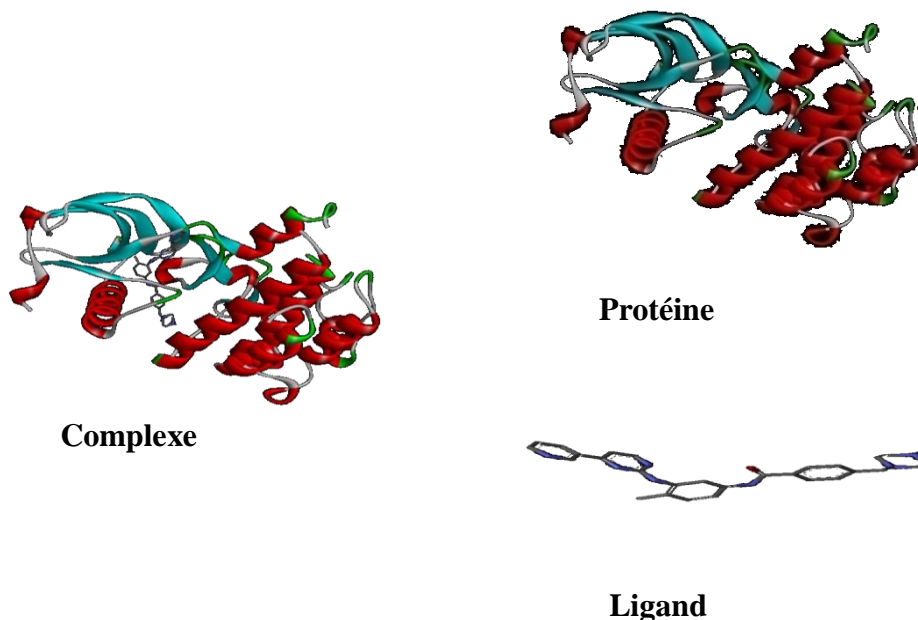
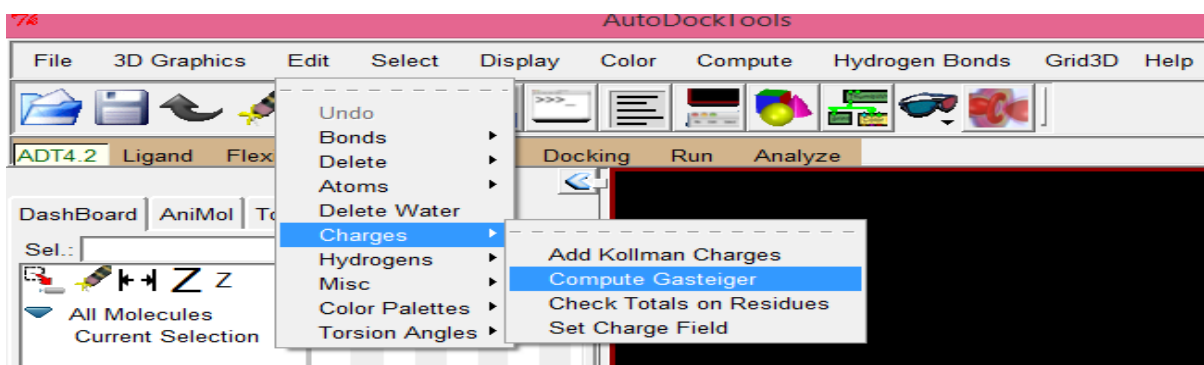


Figure 39: La séparation du ligand de son récepteur

## II.1.1.5. Préparation de la cible :

La cible est ensuite préparée en définissant les charges partielles de ses atomes par l'intermédiaire (AutoDock Tools) qui attribue des charges de type "gasteiger"

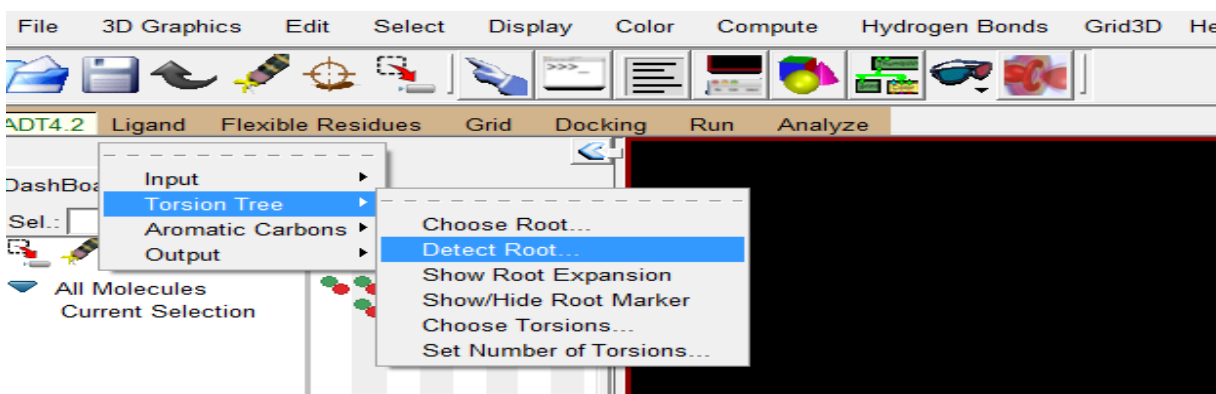


**Figure 40: les charges partielles des atomes de la cible de type "gasteiger"**

Il est aussi nécessaire de rajouter à ce récepteur les hydrogènes dits polaires. Le fichier de la protéine est ensuite converti par AutoDock au format pdbqt (format propre à Auto Dock Vina) qui correspond aux données des charges

## II.1.1.6. Préparation du ligand

Déjà enregistré en format PDB sera préparé en détectant les liaisons rotatoires « Roots ». La molécule ainsi obtenue est enregistrée au format pdbqt.



**Figure 41: Détection des liaisons rotatoires**

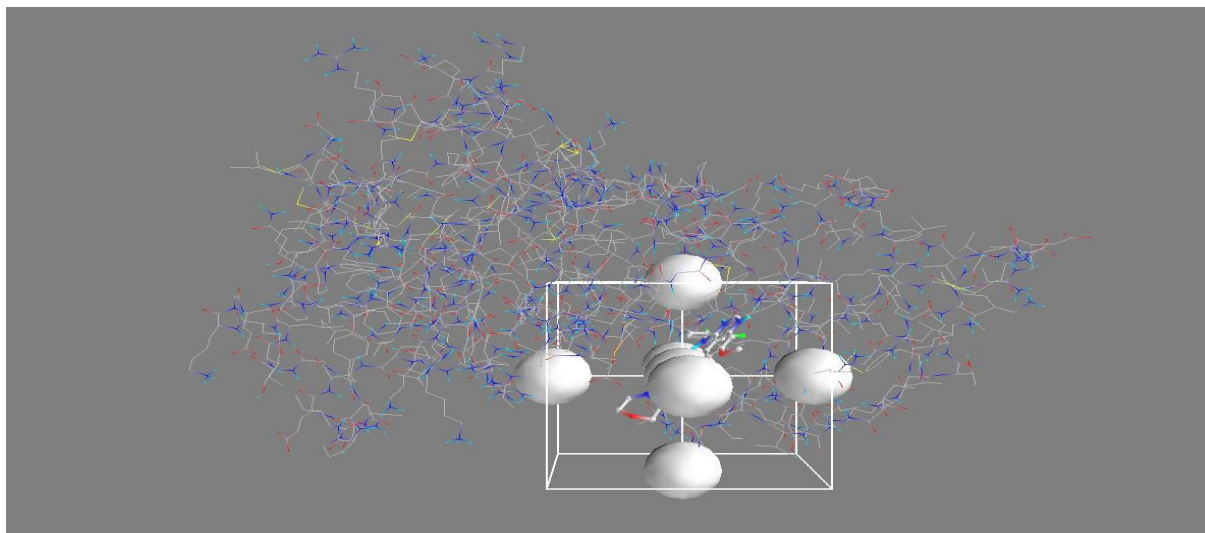
## II.1.2. Recherche des similaires

Afin d'établir les molécules similaires au ligand de choix en vue de leur utilisation au docking, une recherche sur la banque des données PubChem est réalisée et toutes les molécules sont téléchargées sous format SDF

Ces molécules sont ensuite enregistrées sous format PDB par DS2019 puis pdbqt par AutoDock Tools.

### II.1.3. Docking moléculaires:

Le docking est réalisée avec le programme AutoDock Vina intégré dans PyRx, nécessitant le format pdbqt. En ce qui concerne le site actif, il a été défini de manière manuelle en se basant sur les coordonnées obtenus par AutoDock Tools, dans le but d'englober la totalité de l'espace couvert par le ligand de référence dans le site d'interaction



**Figure 42: Détermination des coordonnées(XYZ) du site actif d'une manière manuelle**

Le docking moléculaire nous permet de générer les meilleures positions (conformations) de chaque ligand étudié au sein du site actif et les classe selon leur énergie d'interaction croissante de chaque pose.

Ligand	Binding Affinity (kcal/mol)	Mode	RMSD lower bound	RMSD upper bound
protein_legend	-8.2	0	0.0	0.0
protein_legend	-8.2	1	2.275	5.551
protein_legend	-7.9	2	3.111	5.821
protein_legend	-7.8	3	3.366	6.323
protein_legend	-7.7	4	1.985	5.286
protein_legend	-7.7	5	1.567	1.996
protein_legend	-7.5	6	3.775	7.617
protein_legend	-7.3	7	2.994	5.723
protein_legend	-7.2	8	3.58	6.595

**Figure 43: Résultats du docking**

#### II.1.4. Visualisation 2D :

On fait une visualisation des interactions entre les molécules choisies en les comparant avec les interactions du ligand de référence

#### II.1.5. Filtrage ADME/tox :

##### II.1.5.1. Propriétés physico-chimiques :

Chaque médicament éventuel doit se conformer à plusieurs propriétés de base, tels son faible coût de production, être soluble, stable mais doit aussi se conformer à des barèmes associés à ses propriétés pharmacologiques d'absorption, de distribution, de métabolisme, d'excrétion et de toxicité. En 1997 Lipinski et al ont proposés « la règle de 5 » également appelée « règle de LIPINSKI » permettant de déterminer si un composé chimique possède des caractéristiques médicamenteuses précises.

Selon cette règle, un composé possède toutes les chances d'être disponible oralement s'il respecte au moins 3 des 5 critères suivantes :

1. Possède une masse moléculaire inférieure à 500 daltons.
2. Un coefficient de partition (logP) ou une lipophilie entre -2 et 5.
3. Ne possède pas plus de 5 donneurs de ponts hydrogène (OH et NH).
4. Ne possède pas plus de 10 accepteurs de ponts hydrogène (O et N).
5. Le nombre de fonctions rotatives (flexibles et linéaires) inférieur à 15.

Des compléments ont été apportés, notamment par Veber et al (2002), suggérant que les composés qui répondent uniquement aux deux critères suivants, favoriseraient une bonne Biodisponibilité orale. :

1. 10 liaisons rotatives ou moins.
2. Une surface polaire égale ou inférieure à 140 Å<sup>2</sup>.

La vérification de la solubilité dans l'eau et l'accessibilité à la synthèse chimique, a été également établie.(102)

## II.1.5.2. Propriétés pharmacocinétiques :

Pour déterminer la voie d'administration d'un médicament, la présence de plusieurs barrières biologiques dans le corps humain doit être prise en compte. Les barrières biologiques incluent les barrières buccale, gastrique, intestinale, pulmonaire, nasale, vaginale, testiculaire, oculaire, hépatique, rénale, placentaire, hémato encéphalique et de la peau.

La molécule doit être absorbée pour se retrouver dans la circulation systémique et être distribuée dans les tissus et les organes.

La phase d'essais précliniques demande l'utilisation de modèles expérimentaux *in silico*, *in vitro* et *in vivo* comme outils de prédiction de l'absorption chez l'humain. Les modèles animaux sont les plus fiables pour évaluer l'absorption des médicaments(103).

Les tests effectués sur les animaux peuvent être accompagnés ou précédés par les tests BBB-PAMPA, qui utilisent un mélange de phospholipides infusés dans des microfibrilles lipophiliques avec des lipides chargés négativement, un système de membrane artificielle performante et représentative des propriétés des membranes lipidiques du cerveau. Les données sont utilisées pour prédire les cinétiques de passage des membranes du cerveau.(104)

L'inhibition d'un ou plusieurs iso formes (CYP1A2, CYP2C9, CYP3A4...) de cytochromes P450 par un composé, cause une diminution du métabolisme des autres molécules biotransformées par les isoformes inhibées. Cette diminution provoque un ralentissement de la détoxification et élimination des molécules et une accumulation dans l'organisme. La concentration de ces molécules peut augmenter jusqu'à arriver à des niveaux toxiques.

En clinique, les conséquences de l'inhibition des cytochromes P450 sont généralement négatives, et elles provoquent des symptômes graves chez les patients, Mais il y des cas où l'inhibition des cytochromes P450 peut être considérée comme utile quand elle permet d'obtenir une augmentation de la biodisponibilité de médicaments, et d'améliorer son efficacité.(105)

## II.1.5.3. Tests de toxicité:

La mutagénèse a pour objectif de détecter d'éventuelles modifications du matériel génétique induites par le médicament (effet mutagène ou clastogène). Le test d'Ames (1975) est le test de génotoxicité le plus répandu. On y utilise des souches de bactéries (*Salmonella typhimurium*) qui ont perdu par mutation la faculté de synthétiser un acide aminé, l'histidine ; on cultive ces bactéries sur des milieux sans histidine, en présence de substance à tester (native ou traitée avec des enzymes hépatiques). Si la substance n'est pas mutagène, le développement des bactéries n'a pas lieu ; dans le cas contraire, la substance provoque une mutation "réverse", qui redonne aux bactéries la faculté de synthétiser l'histidine et donc de se développer.

La cancérogenèse est une étape permettant de détecter un éventuel pouvoir cancérigène du produit, qui sera administré au long cours (2ans ou plus), chez 2 espèces (généralement rat et souris) avec 3 niveaux de doses. La voie d'administration sera identique à celle utilisée en clinique. Il est nécessaire d'apporter la preuve de l'exposition au produit par des mesures répétées de concentration plasmatique.

Les tests d'inhibition du gène hERG, (tests pour les nouveaux médicaments réservés aux essais). Ce test sert habituellement à révéler le mécanisme qui sous-tend différents cas d'arythmie associée aux canaux ioniques et à dévoiler l'inhibition du courant ionique comme un effet primaire ou secondaire du produit pharmacologique.

Également, des tests utilisant des espèces aquatiques ; poissons ex : (*Daphnia*, *Medaka*, *Minnnow*) et algues peuvent être utiles pour vérifier la toxicité d'un candidat médicament. Les paramètres généralement mesurés lors de ces essais sur les différentes espèces sont : la mobilité, croissance et reproductibilité...etc.

Dans notre travail, ces paramètres ont été calculés et déterminés en utilisant les deux serveurs : SwissADME et PreADMET. Il faut tenir compte que les composés doivent être convertis au format smile pour le premier serveur et au format mol pour le second, afin d'établir le filtrage ADME/tox.

## II.2. Tests de fiabilité du programme AutoDock Vina :

L'étude in silico de l'inhibition de l'EGFR par docking moléculaire nécessite une évaluation préalable de la performance du programme AutoDock vina par des tests différents :

### II.2.1. Analyse visuelle :

Ayant pour but la comparaison visuelle de la superposition des ligands séparés de ces récepteurs téléchargés de la PDB avec ceux qui résultent du selfdock.

### II.2.2. Coefficient de corrélation

Le calcul de coefficient de corrélation entre les énergies d'interaction obtenu par le docking et les logIC50 des inhibiteurs de la littérature.

**Tableau IV: Interprétation par rapport à la valeur du coefficient de corrélation**

	INTERPRETATION	
Corrélation	Négative	Positive
Faible	de -0,5 à 0	de 0 à +0,5
Forte	de -1 à -0,5	de 0,5 à +1

### II.2.3. RMSD : L'écart quadratique moyen ou (root mean square deviation).

Un moyen largement utilisé pour comparer les structures de biomolécules ou de corps solides consiste à translater et à faire pivoter une structure par rapport à l'autre pour minimiser la déviation quadratique moyenne (RMSD).

# **RESULTATS ET DISCUSSION**



## I. Tests de fiabilité du programme AutoDock Vina :

### I.1. RMSD :

Le programme sélectionné est Vina, principalement utilisé pour les calculs de criblage virtuel, il ne contient aucun script de calcul RMSD. Pour cette raison, nous avons choisi de vérifier la fiabilité de vina en réalisant deux tests : l'analyse visuelle, et de tests de corrélation linéaire

### I.2. Analyse visuelle :

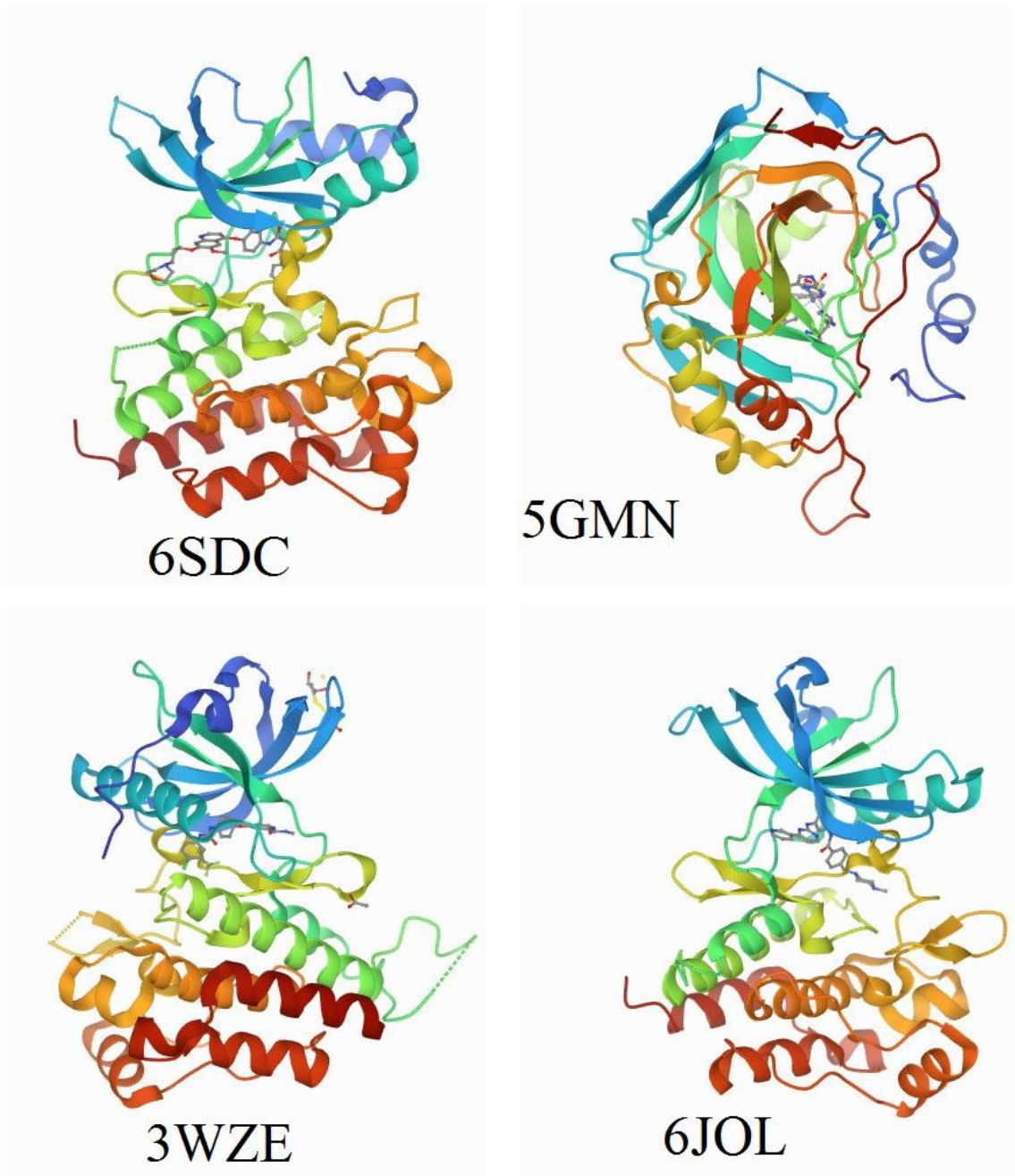
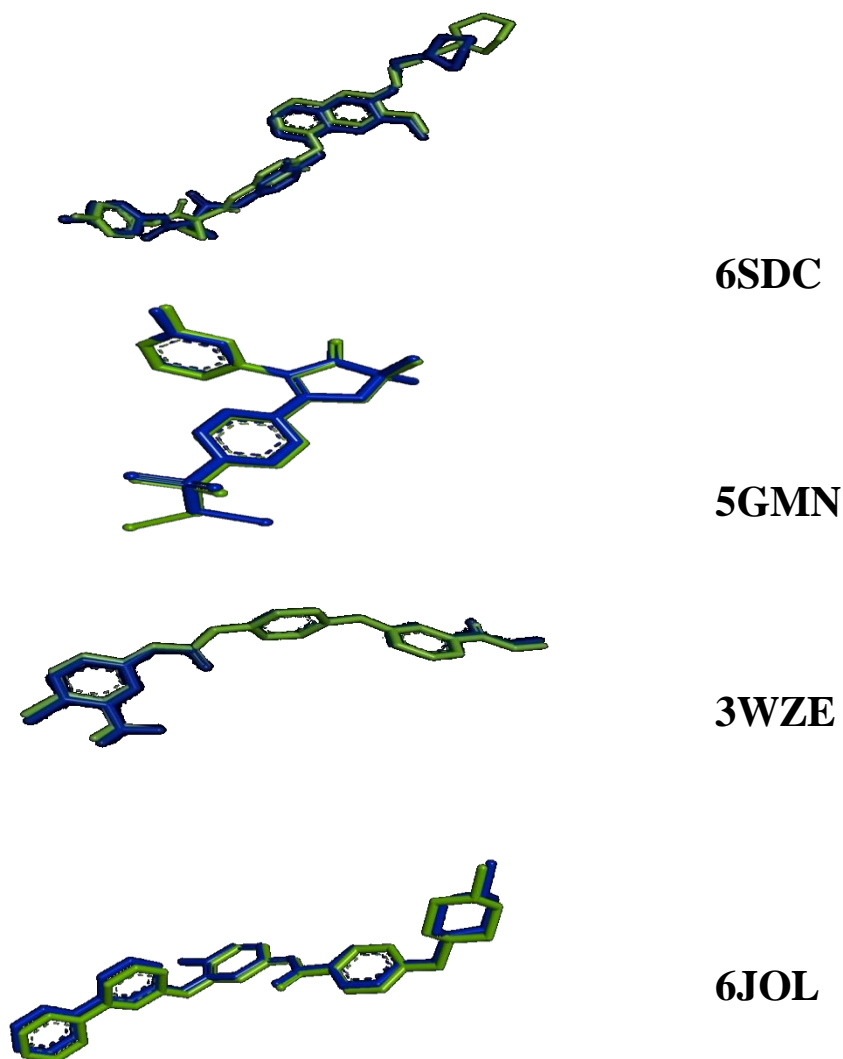


Figure 44: Structures 3D des 4 complexes issues de la PDB

Tableau V: Les cordonnées des sites d'interaction (récepteur-ligand) (XYZ).

	X	Y	Z	Références
<b>6SDC</b>	9,019	-20,746	-26,937	(106)
<b>5GMN</b>	9,834	1,495	12,734	(107)
<b>3WZE</b>	21,492	25,061	35,952	(108)
<b>6JOL</b>	-30,439	156,489	-0,398	(109)



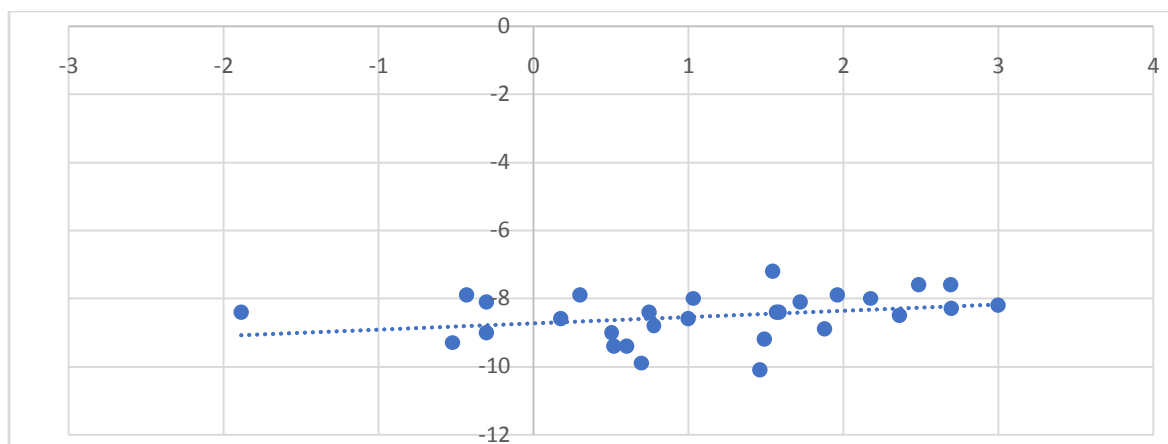
**Figure 45:** Comparaison des conformations obtenues par rayon-X (colorés en vert) et par docking moléculaire avec AutoDock vina (colorés en bleu).

Dans notre cas, la **Figure 45** montre qu'on a une bonne superposition des poses simulées par Vina (colorés en bleu) et celles du ligand naturel (colorées en verts).

### I.3. Coefficient de corrélation (r) :

Ce test consiste à calculer le coefficient de corrélation linéaire en traçant la droite de régression linéaire entre les valeurs des énergies d'interaction ( $\Delta G$ ) des inhibiteurs de l'EGFR calculés par AutoDock vina et les valeurs de leur IC50 déterminées expérimentalement dans la littérature par des essais in vitro.

Dans le présent travail, nous avons utilisé 34 inhibiteurs de l'EGFR issus de la littérature. Les résultats des énergies d'interaction des inhibiteurs étudiés ainsi que leur Log IC50 sont représentés dans le **Tableau VI**.



**Figure 46:Corrélation entre l'activité biologique (Log IC50) de 34 ligands et leur énergie d'interaction vis-à-vis de l'EGFR**

L'analyse par régression linéaire réalisée entre les énergies d'interaction et les IC50 (Log 10 IC50) des inhibiteurs étudiés donne un nuage de points (Figure 46) avec un coefficient de corrélation égal à **0,311283115**, témoignant que ces deux variables évoluent dans le même sens et donc qu'il existe une bonne corrélation linéaire entre elles. Ces résultats apportent la preuve, encore une fois, que le programme AutoDock Vina est performant.

**Tableau VI:Corrélation entre les énergies d'interaction des 30 inhibiteurs de la EGFR et leur activité biologique (Log IC50).**

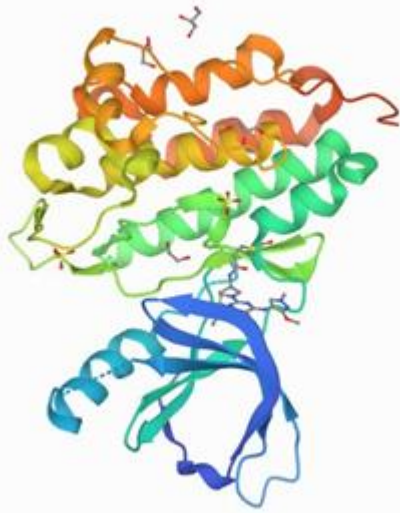
		Y		X	
	COMPLEXE	Binding energy	IC50	logIC50	Références
1	protein_liferafenib	-10,1	29	1,462397998	(110)
2	protein_icotinib	-9,9	5	0,698970004	(111)
3	protein_sapitinib	-9,4	4	0,602059991	(112)
4	protein_abivertinib	-9,4	3,3	0,51851394	(113)
5	protein_zorifertinib	-9,3	0,3	-0,522878745	(114)
6	protein_tesevatinib	-9,3	0,3	-0,522878745	(115)
7	protein_nazartinib	-9,2	31	1,491361694	(116)
8	protein_poziotinib	-9	3,2	0,505149978	(117)
9	protein_allitinib	-9	0,5	-0,301029996	(118)
10	protein_lazertinib	-8,9	76	1,880813592	(119)
11	protein_dacometinib	-8,8	6	0,77815125	(120)
12	protein_olmutinib	-8,6	10	1	(121)
13	protein_canertinib	-8,6	1,5	0,176091259	(122)
14	protein_naquotinib	-8,5	230	2,361727836	(123)
15	protein_ibrutinib	-8,4	5,6	0,748188027	(124)
16	protein_pyrotinib	-8,4	0,013	-1,886056648	(125)
17	protein_pelitinib	-8,4	38,5	1,58546073	(126)
18	protein_gefetinib	-8,4	37	1,568201724	(127)
19	protein_vandetanib	-8,3	500	2,698970004	(128)
20	protein_rezivertinib	-8,2	1000	3	(129)
21	protein_selatinib	-8,1	52,67	1,721563318	(130)
22	protein_afatinib	-8,1	0,5	-0,301029996	(131)
23	protein_lapatinib	-8	10,8	1,033423755	(132)
24	protein_brigatinib	-8	150,3	2,176958981	(133)
25	protein_neratinib	-7,9	92	1,963787827	(134)
26	protein_erlotinib	-7,9	2	0,301029996	(135)
27	protein_almonertinib	-7,9	0,37	-0,431798276	(136)
28	protein_osimertinib	-7,6	493,8	2,693551086	(137)
29	protein_mavelertinib	-7,6	307	2,487138375	(138)
30	protein_mobocertinib	-7,2	35	1,544068044	(139)

## II. DOCKING MOLECULAIRE :

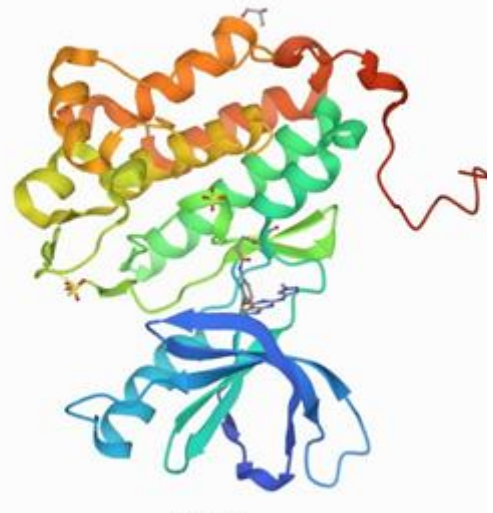
### II.1. La recherche bibliographique

Tableau VII :Code PDB des complexes " EGFR-inhibiteur " et leurs caractéristiques

	Code PDB	Résolution (Å)	Code de ligand	Nom du ligand	Mutation du récepteur	nombre des chaines	Nombre des acides aminés	ic50 de l'inhibiteur (nM)	Références
1	5UG9	1,33	8AM		Oui	1	329	/	(140)
2	5HG8	1,42	634		Oui	1	329	/	(141)
3	6TG0	1,5	N78		Oui	2	333	/	(142)
4	6TFY	1,5	N7Z		Oui	2	333	/	(143)
5	6TFV	1,5	N7Q		Oui	2	333	/	(144)
6	5HG5	1,52	633		Oui	1	329	/	(145)
7	6TG1	1,6	N82		Oui	2	333	/	(146)
8	6V66	1,79	QP1		Oui	4	327	/	(147)
9	6TFZ	1,8	N7B		Oui	2	333	/	(148)
10	5GNK	1,8	80U		Oui	1	297	/	(149)
11	4I24	1,8	1C9		Oui	2	329	/	(150)
12	5X2A	1,85	7XO		Oui	4	331	/	(151)
13	5HG7	1,85	630		Oui	1	329	/	(152)
14	4WKQ	1,85	IRE	Gefitinib	<b>Non</b>	1	330	37	(153)
15	7A2A	1,9	7G9	Spebrutinib	Oui	2	333	<0,5	(154)
16	6P8Q	1,9	J01		Oui	4	327	/	(155)
17	3W2S	1,9	W2R		<b>Non</b>	1	330	/	(156)
18	5YU9	1,95	1 E 8	Ibrutinib	Oui	4	331	35,7 +/-0,1	(157)



**5UG9**



**5HG8**

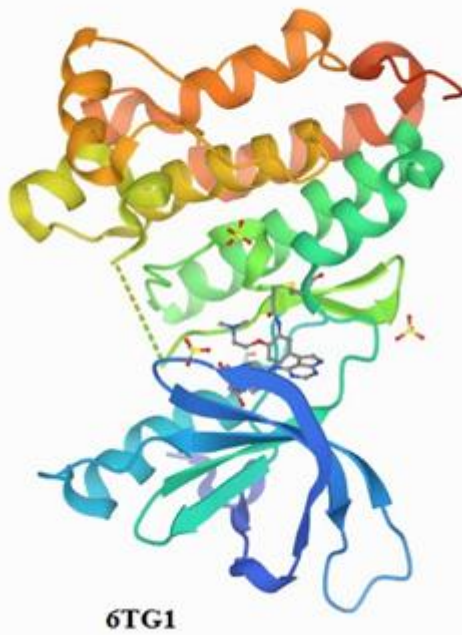
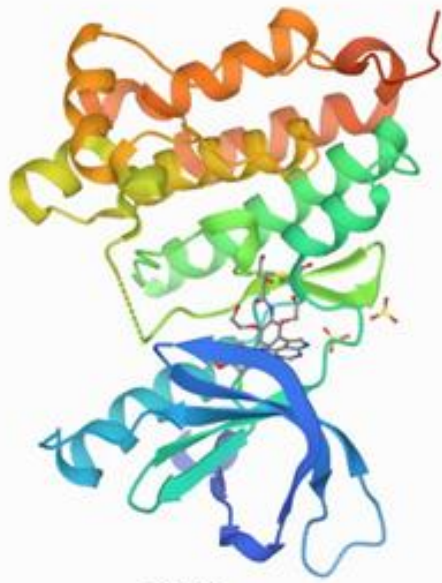


**6TG0**



**6TFY**



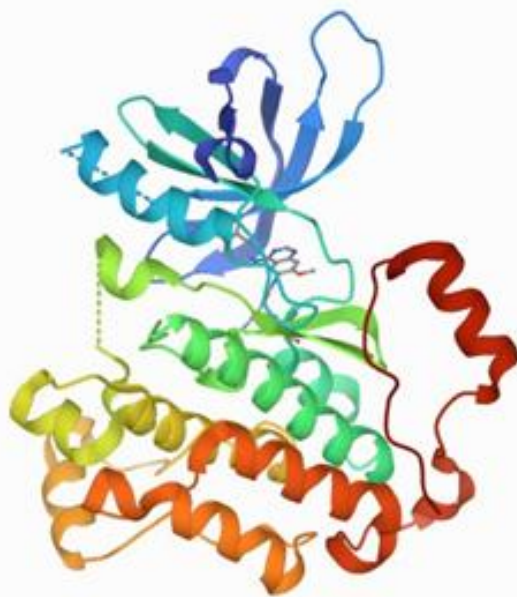




6TFZ



5GNK



4I24

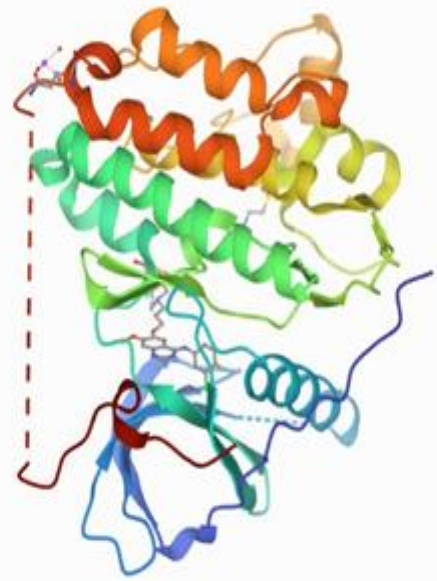


5X2A

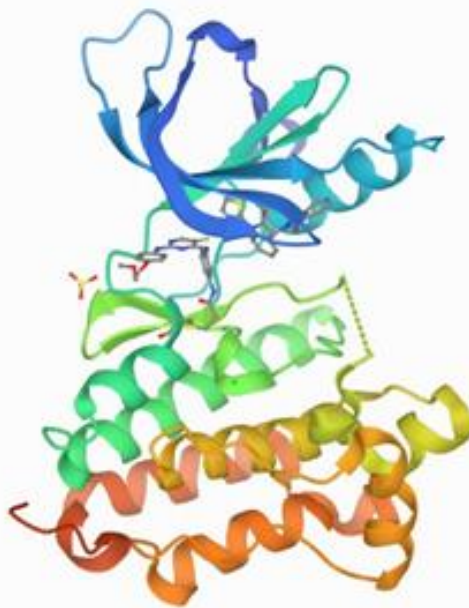




5HG7



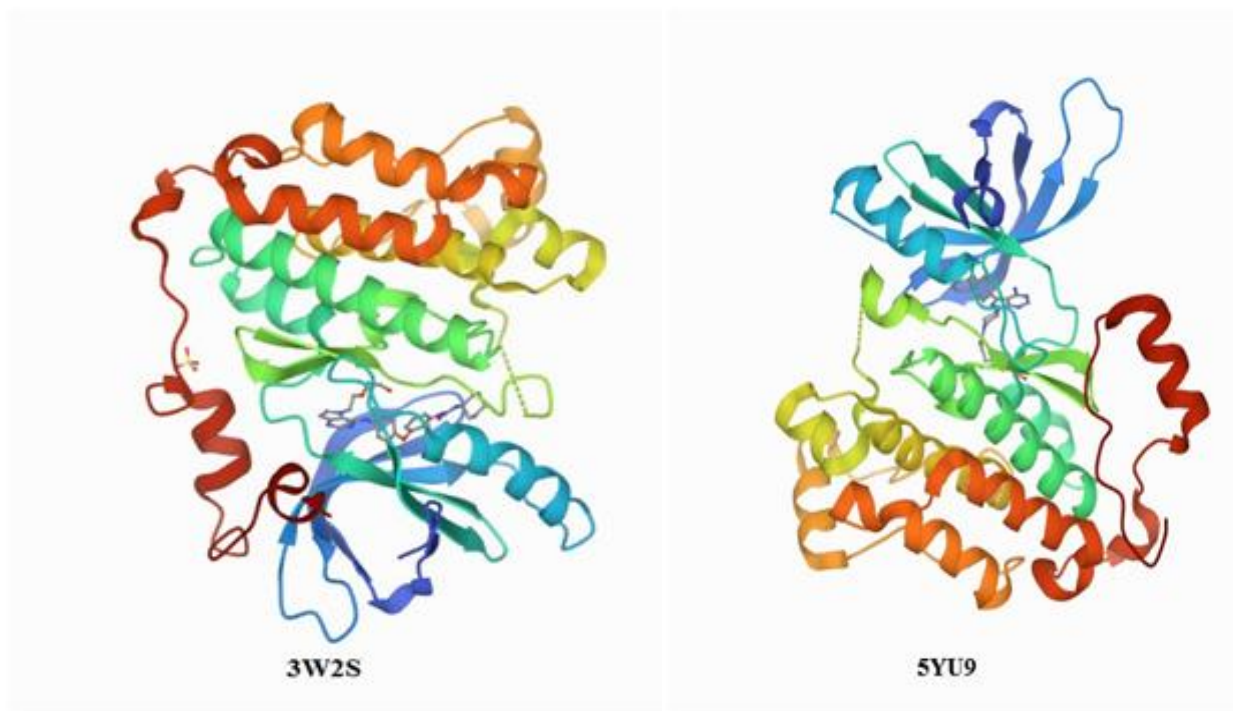
4WKQ



7A2A



6P8Q



**Figure 47: Structures 3D des complexes EGFR issues de la PDB**

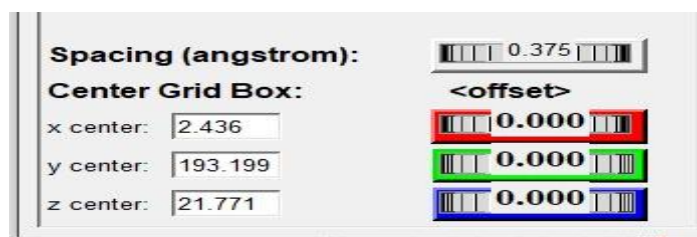
-Le complexe de choix est le **4WKQ** car il s'agit d'une résolution  $<2$  ( $=1.85\text{\AA}$ ),  $IC_{50}=37$  sans mutation et son ligand est le **gefitinib** (code IRE) qu'est une molécule très efficace contre les tumeurs pulmonaires non à petites cellules et présente une efficacité même en cas des mutations, et cette molécule est approuvée par l'OMS (voir page 35)

-Le téléchargement du complexe **4WKQ** sous forme PDB.

-La préparation du complexe au docking, mais d'abord la séparation du ligand de son récepteur et la détermination des coordonnées du site actif (XYZ)

**Tableau VIII: coordonnées de site d'interaction du complexe 4WKQ**

	X	Y	Z
<b>4WKQ</b>	2,436	193,199	21,771



**Figure 48: les coordonnées du site d'interaction du complexe 4WKQ**

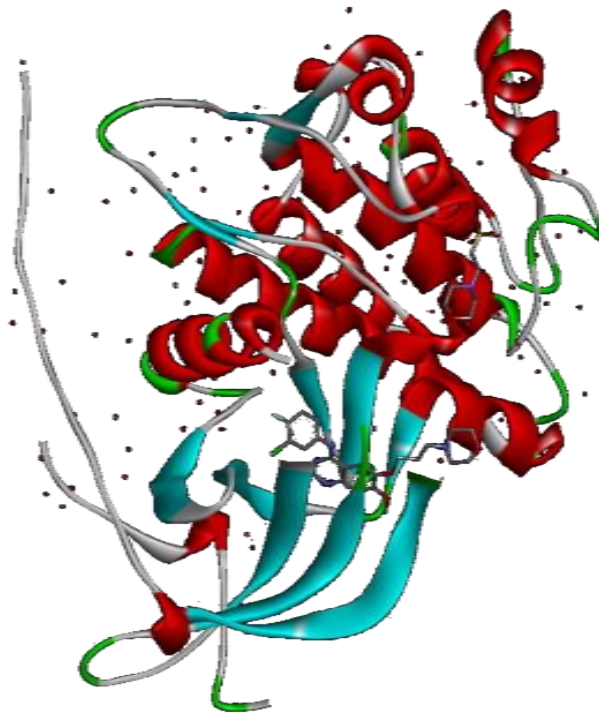


Figure 49: L'enzyme 4WKQ non simplifié

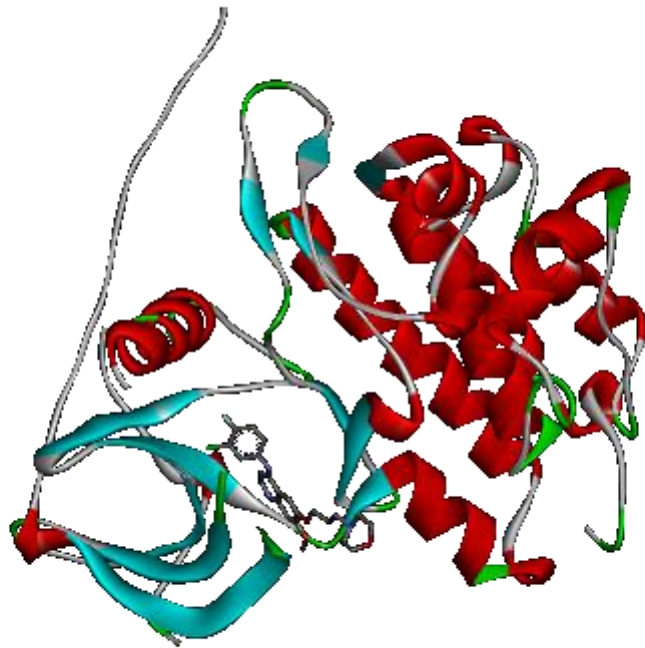


Figure 50: L'enzyme 4WKQ simplifié.

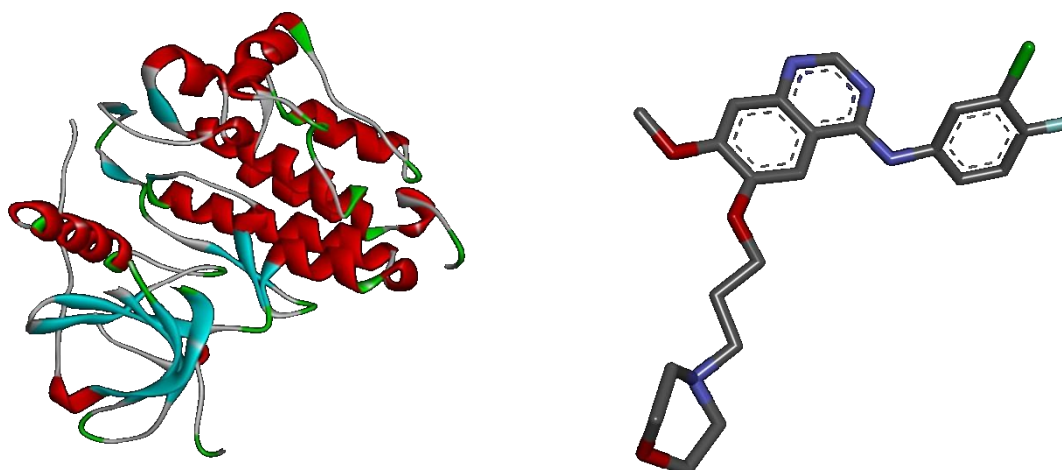


Figure 51: La séparation du ligand « gefitinib » de son récepteur « le complexe 4WKQ »

## II.2. Recherche des similaires au gefitinib:

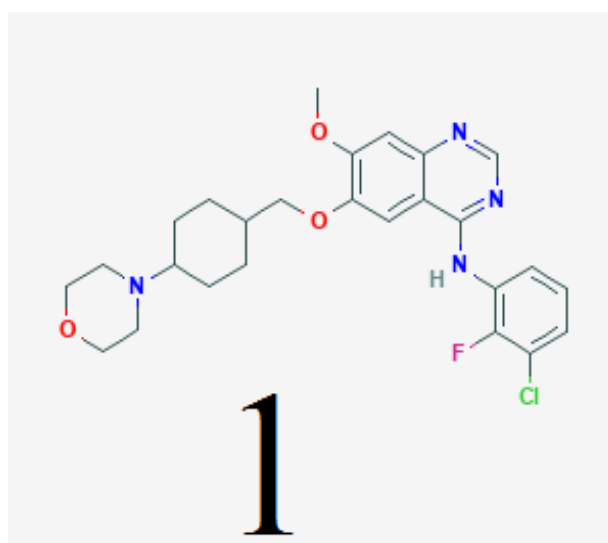
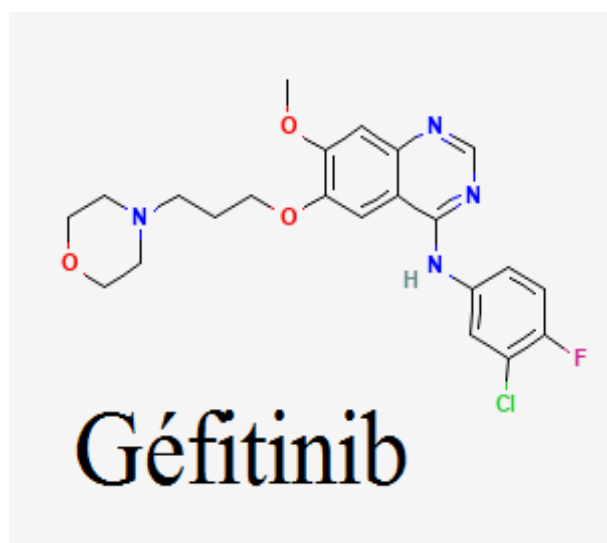
Figure 52: Recherche des similaires au gefitinib sur PubChem

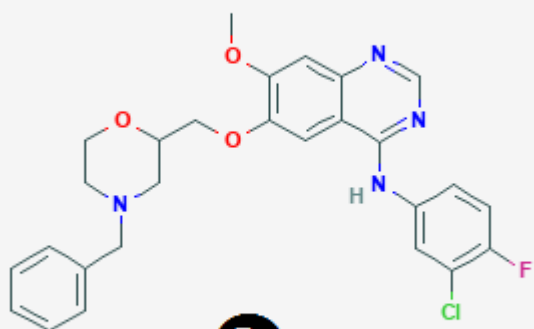
Afin d'établir de nouvelles molécules ayant une activité inhibitrice mieux que le Gefetinib, nous avons fait appel à la PubChem pour générer d'autres structures similaires de ce dernier. La PubChem nous a fournis une collection de 394 structures avec un taux de similarité de 97% ; dont 41 sont répétés.

Le docking moléculaire de ces similaires avec la **4WKQ**, réalisé avec AutoDock Vina a permis de calculer leur énergie d'interaction (**annexe N°1**)

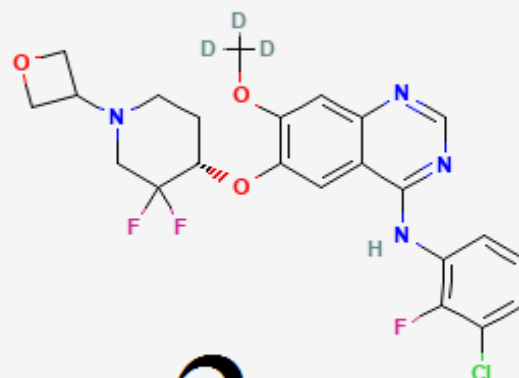
**Tableau VIII: Score des meilleurs inhibiteurs obtenus suite au docking moléculaire de la collection de similaires du gefitinib**

	Ligand	Energie d'interaction (Kcal/mol)
1	protein_Conformer3D_CID_146844011	-9,8
2	protein_Conformer3D_CID_142675009	-9,6
3	protein_Conformer3D_CID_137455797	-9,6
4	protein_Conformer3D_CID_137455767	-9,6
5	protein_Conformer3D_CID_24875330	-9,5
6	protein_Conformer3D_CID_145577260	-9,5
7	protein_Conformer3D_CID_118343039	-9,5
8	protein_Conformer3D_CID_145577297	-9,4
9	protein_Conformer3D_CID_71617554	-9,3
10	protein_Conformer3D_CID_71572791	-9,3
11	protein_Conformer3D_CID_151031088	-9,3
12	protein_Conformer3D_CID_10453762	-9,3
.....	.....	.....
295	protein_ligand(Gefetinib)	-8,2

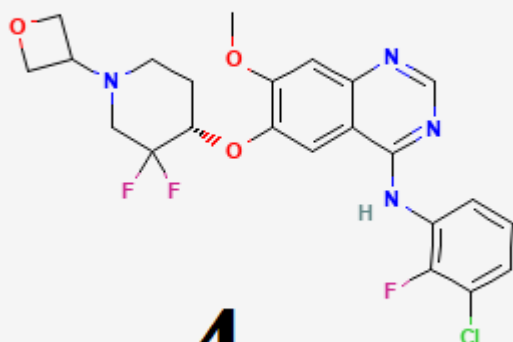




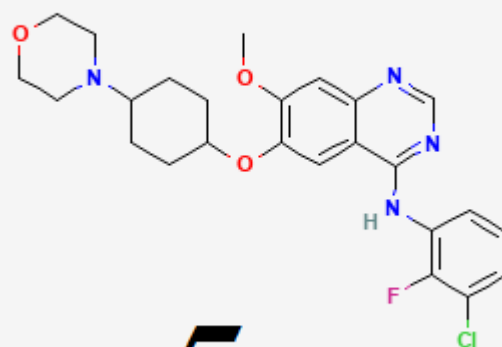
2



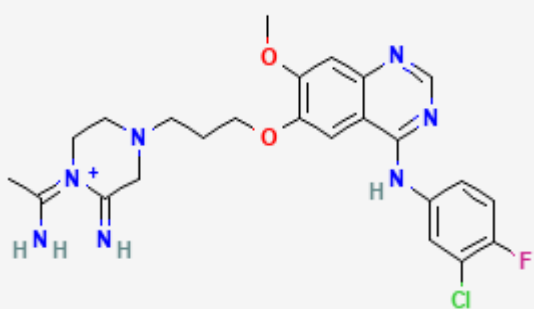
3



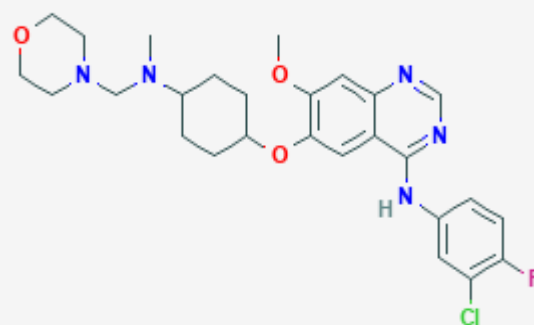
4



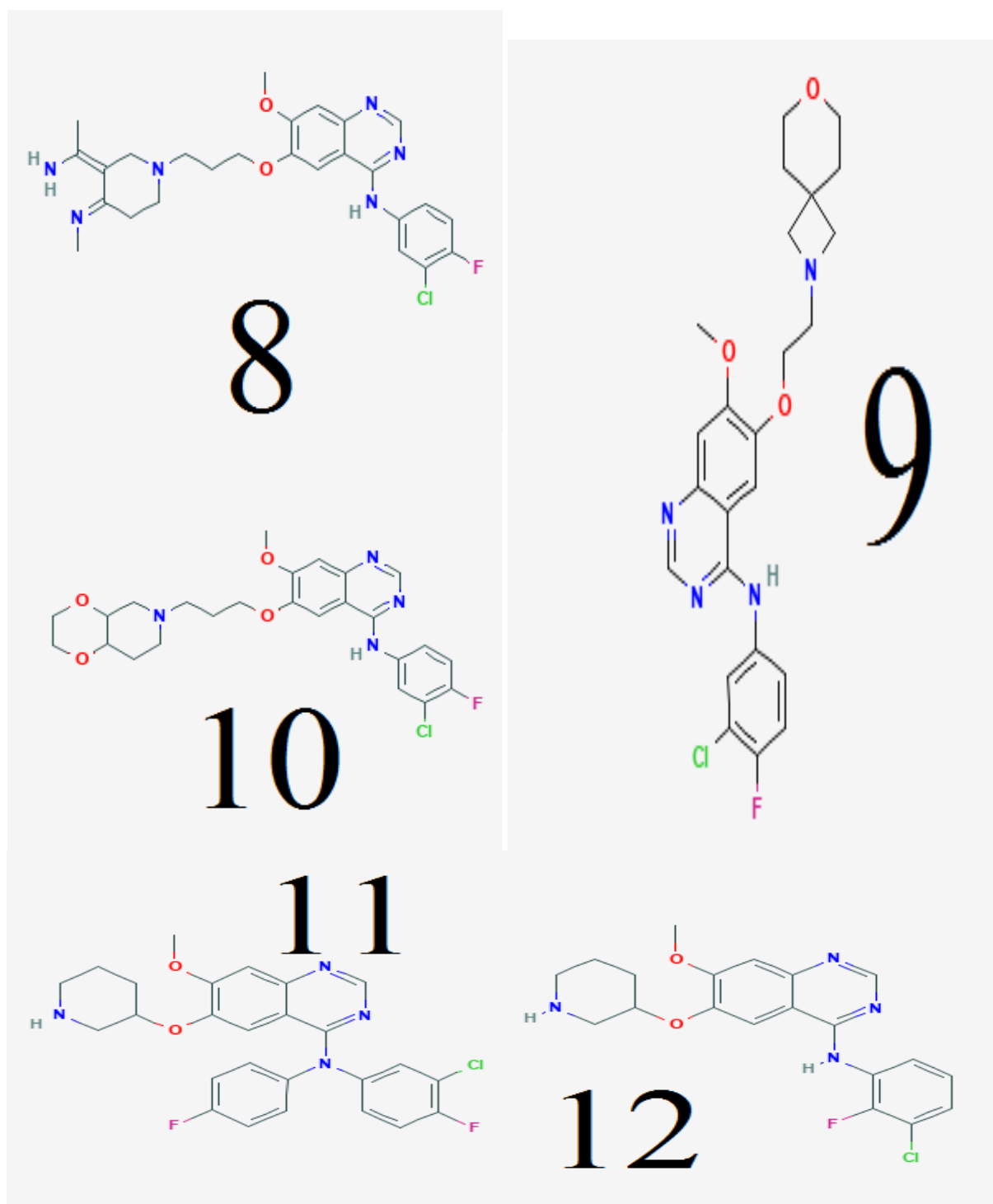
5



6



7



**Figure 53:** Les structures des meilleurs similaires

La molécule n°3 a la même structure que la molécule n°4 sauf qu'elle est marquée donc elle est éliminée des études suivantes



### II.3. Etude des interactions 4WKQ-Gefetinib :

Dans le but de contribuer au développement de nouveaux inhibiteurs de la 4WKQ, il nous a semblé important d'étudier, en premier lieu les mécanismes impliqués dans l'inhibition de cette éventuelle enzyme.

Pour ce faire, nous avons sélectionné le Gefetinib qui nous avons utilisé précédemment dans le test de performance par l'analyse visuelle et qui a révélé le meilleur score (-8,4 Kcal/Mole) lors du docking moléculaire. L'interaction du Gefetinib avec la 4WKQ après docking a été visualisée par DS.

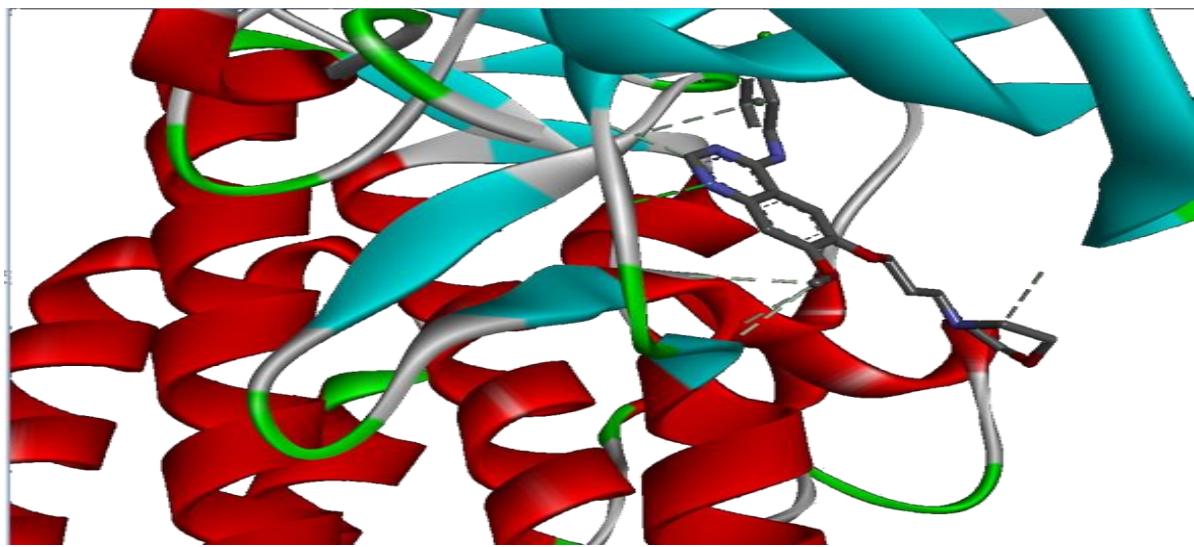


Figure 54: Représentation du positionnement du Gefetinib dans le site actif de la 4WKQ

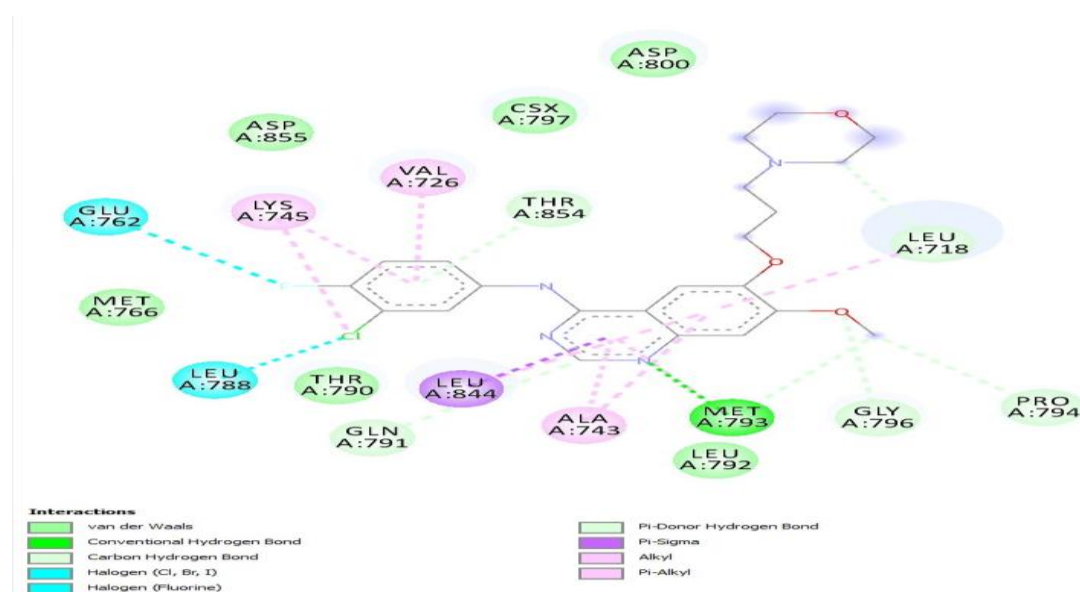


Figure 55: Mode d'interaction du Gefetinib dans le site actif de la 4WKQ.



**Tableau IX: Types d'interaction et acides aminés impliqués dans l'inhibition du EGFR (PDB ID : 4WKQ) avec Gefitinib**

	Van Der Waals	Hydrogène	Longueur d'interaction H (Å )	Interaction carbone-Hydrogène/Pi-Donneur	Halogène	Pi-Sigma	Alkyl et Pi-Alkyl
<b>4WKQ-Gefitinib</b>	ASP800 CSX797 ASP855 MET766 THR790 LEU792	MET793	2,97764	PRO794 GLY796 MET793 LEU718 GLN791 THR854	GLU762 LEU788	LEU844	VAL726 LYS745 LYS745 LEU844 ALA743 ALA743 LEU718

L'analyse visuelle montre que le Gefitinib du complexe 4WKQ est stabilisé par de nombreuses interactions

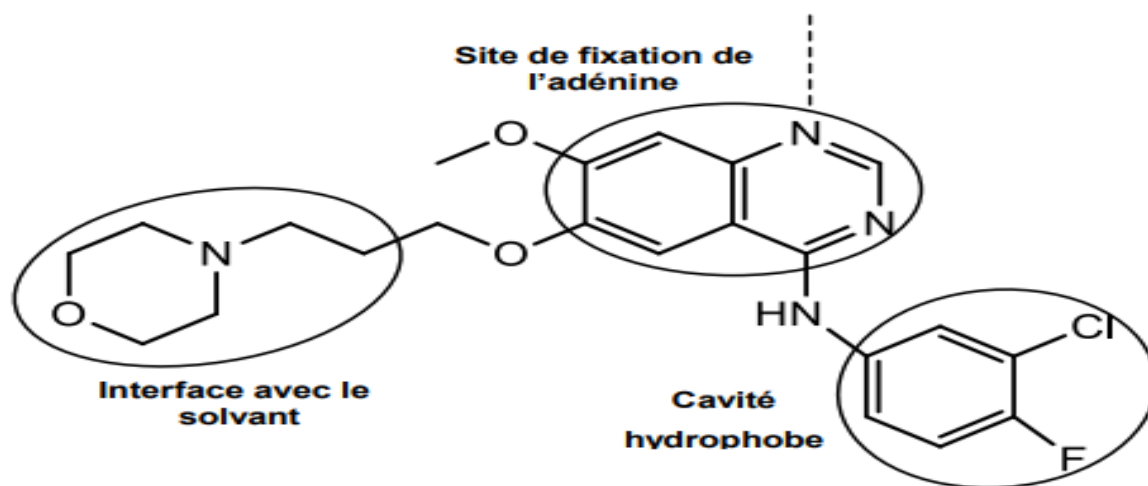
Le gefitinib présente sur un noyau 4- anilinoquinazoline les éléments structuraux nécessaires à l'activité des dérivés de cette série (158)

L'azote en position 1 de la quinazoline est orienté vers la région charnière qu'est un motif de liaison à l'ATP et établit une liaison hydrogène avec MET793 (159)

L'introduction d'un atome de fluor sur l'aniline en position para permet d'améliorer la stabilité et donc l'activité de la molécule in vivo (158)

Les atomes de chlore et de fluor présents sur l'aniline sont entourés respectivement par les résidus THR790, LEU788, MET766, GLU762 , LYS745, ASP855, VAL726 et THR854 de la cavité hydrophobe (159)

La chaîne latérale propoxymorpholine, peu structurée, est exposée au solvant, d'où son rôle dans l'amélioration des propriétés pharmacocinétiques de la molécule. Le gefitinib n'établit donc qu'une liaison hydrogène avec le domaine kinase (159)



**Figure 56: Mode de liaison du gefitinib à l'EGFR (160)**

Donc la présence d'une liaison hydrogène avec MET793 et plusieurs d'interaction au niveau de la cavité hydrophobe, sont deux critères essentiels de choix des molécules susceptibles d'être actives comme inhibitrices d'EGFR.

## II.4. Etude des interactions 4WKQ-les 12 meilleures molécules

## II.4.1. Visualisation des interactions :

## II.4.1.1. Molécule n°1

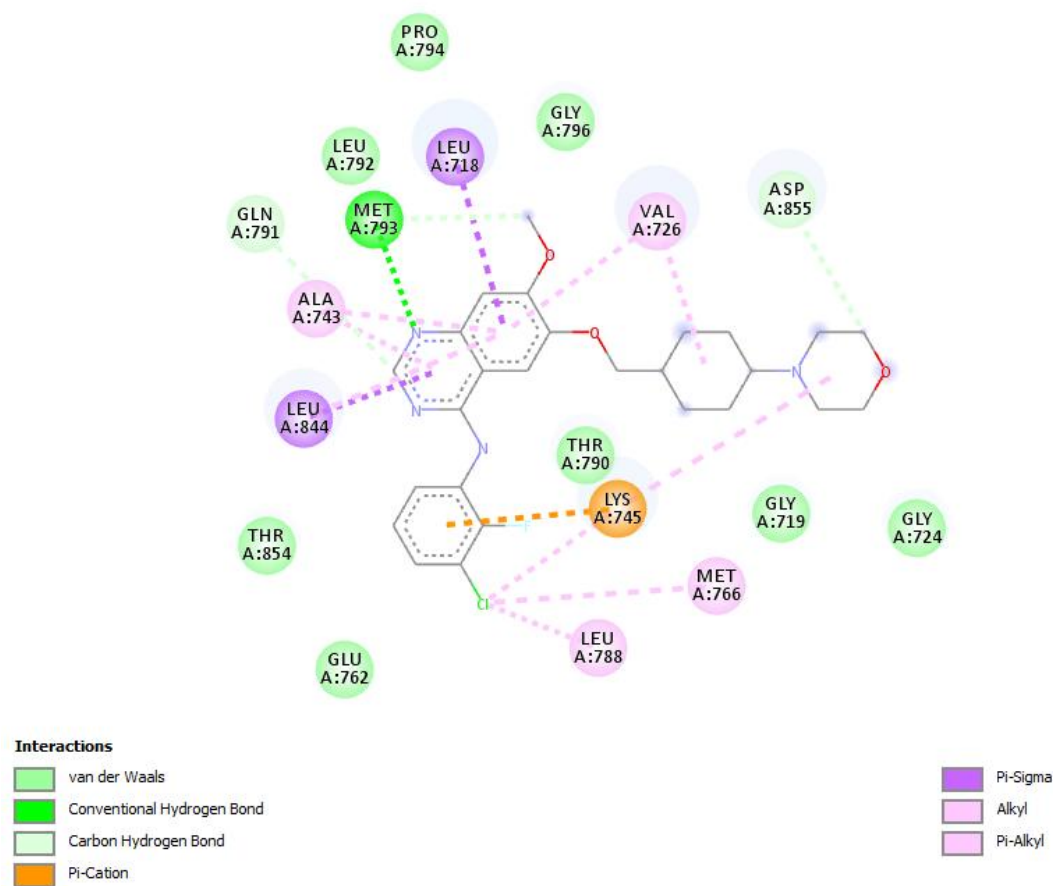


Figure 57: Mode d'interaction de la molécule n°1 dans le site actif de la 4WKQ

Tableau X: Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°1

	Van Der Waals	Hydrogène	Longueur d'interaction H (Å)	Interaction carbone-Hydrogène	Pi-Cation	Pi-Sigma	Alkyl et Pi-Alkyl
<b>4WKQ- molécule n°1</b>	GLU762 THR854 GLY724 GLY719 THR790 LEU792 GLY796 PRO794	MET793	2,17313	GLN791 ASP855	LYS745	LEU718 LEU844	MET766 LEU788 LYS745 LYS745 VAL726 VAL726 LAU844 ALA743 ALA743

## II.4.1.2. Molécule n°2

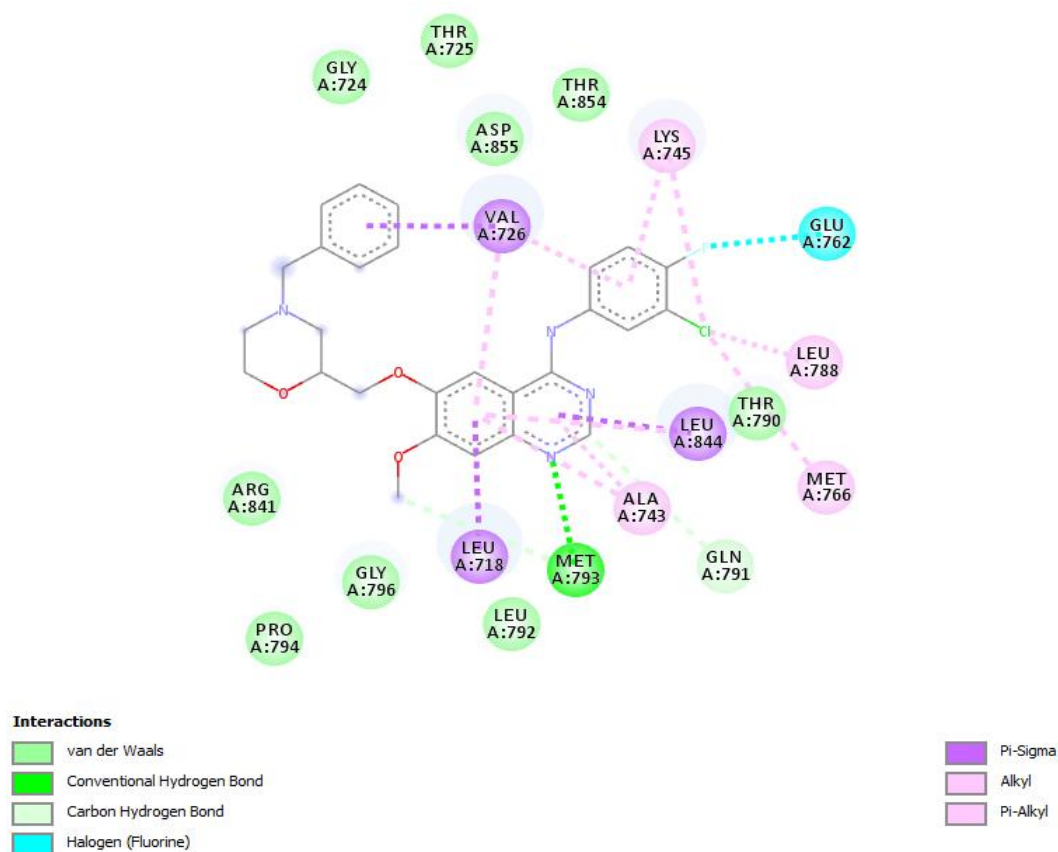


Figure 58: Mode d'interaction de la molécule n°2 dans le site actif de la 4WKQ

Tableau XI : Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°2

	Van Der Waals	Hydrogène	Longueur d'interaction H (Å)	Interaction carbone-Hydrogène	Halogène	Pi-Sigma	Alkyl et Pi-Alkyl
<b>4WKQ- molécule n°2</b>	GLY724 THR725 ASP855 THR854 THR790 ARG841 PRO794 GLY796 LEU792	MET793	2,07493	GLN791 MET793	GLU762	VAL726 LEU718 LEU844	LYS745 LYS745 LEU788 MET766 LEU844 ALA743 ALA743 VAL726 VAL726

## II.4.1.3. Molécule n°4

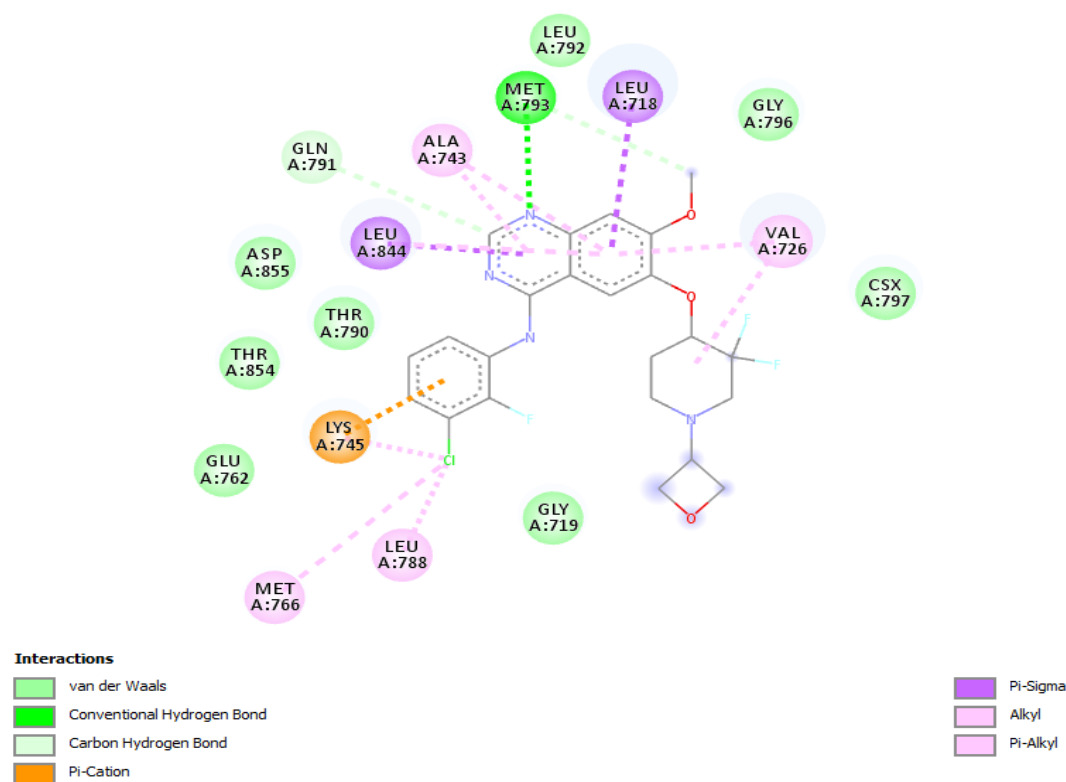


Figure 59: Mode d'interaction de la molécule n°4 dans le site actif de la 4WKQ

Tableau XII: Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°4

	Van Der Waals	Hydrogène	Longueur d'interaction H (Å)	Interaction carbone-Hydrogène	Pi-Cation	Pi-Sigma	Alkyl et Pi-Alkyl
4WKQ- molécule n°4	GLY719 GLU762 THR854 THR790 ASP588 LEU792 GLY796 CSX797	MET793	2,22226	GLN791 MET793	LYS745	LEU718 LEU844	VAL726 VAL726 LEU788 MET766 LYS745 LEU844 ALA743 ALA743

## II.4.1.4. Molécule n°5

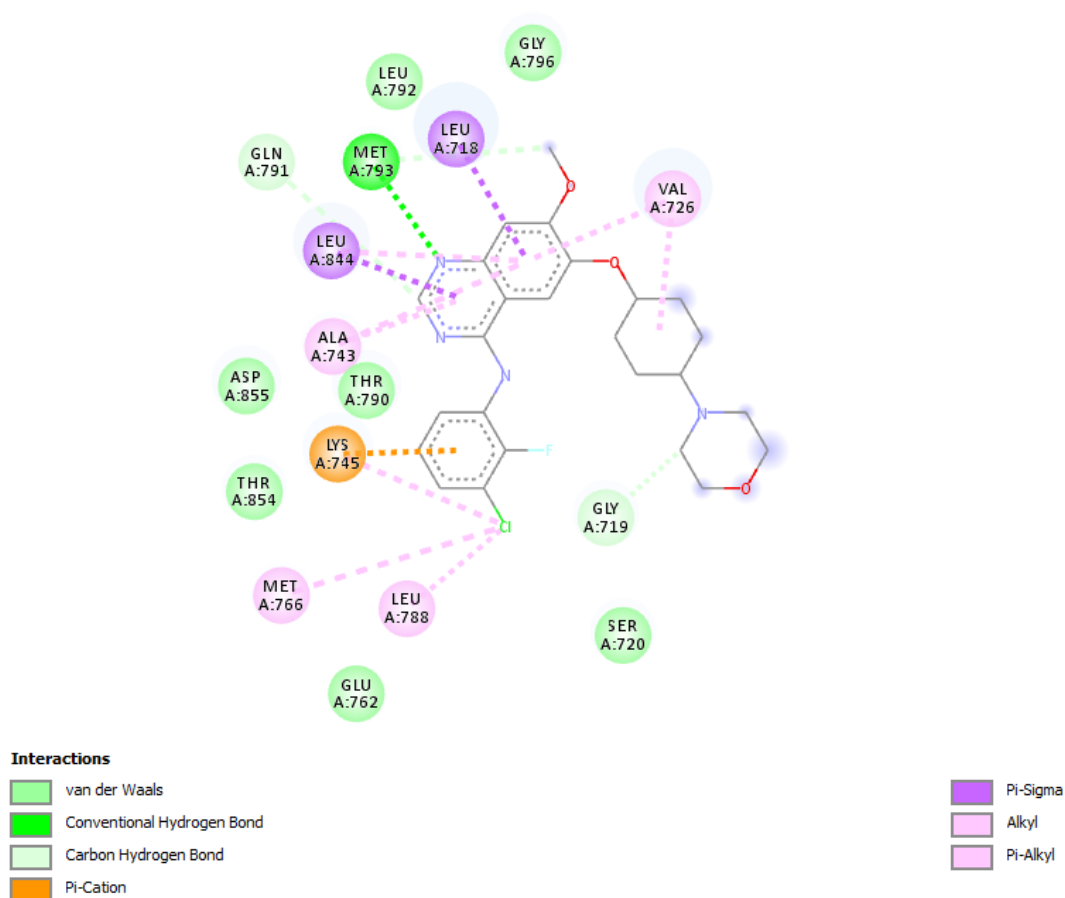


Figure 60: Mode d'interaction de la molécule n°5 dans le site actif de la 4WKQ

Tableau XIII: Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°5

	Van Der Waals	Hydrogène	Longueur d'interaction H (Å)	Interaction carbone-Hydrogène	Pi-Cation	Pi-Sigma	Alkyl et Pi-Alkyl
4WKQ- molécule n°5	GLY796 LEU792 ASP855 THR790 AHR854 GLU762 SER720	MET793	2,18596	GLN791 GLY719 MET793	LYS745	LEU718 LEU844	LEU788 MET766 LYS745 ALA743 ALA743 LEU844 VAL726 VAL726

## II.4.1.5. Molécule n°6

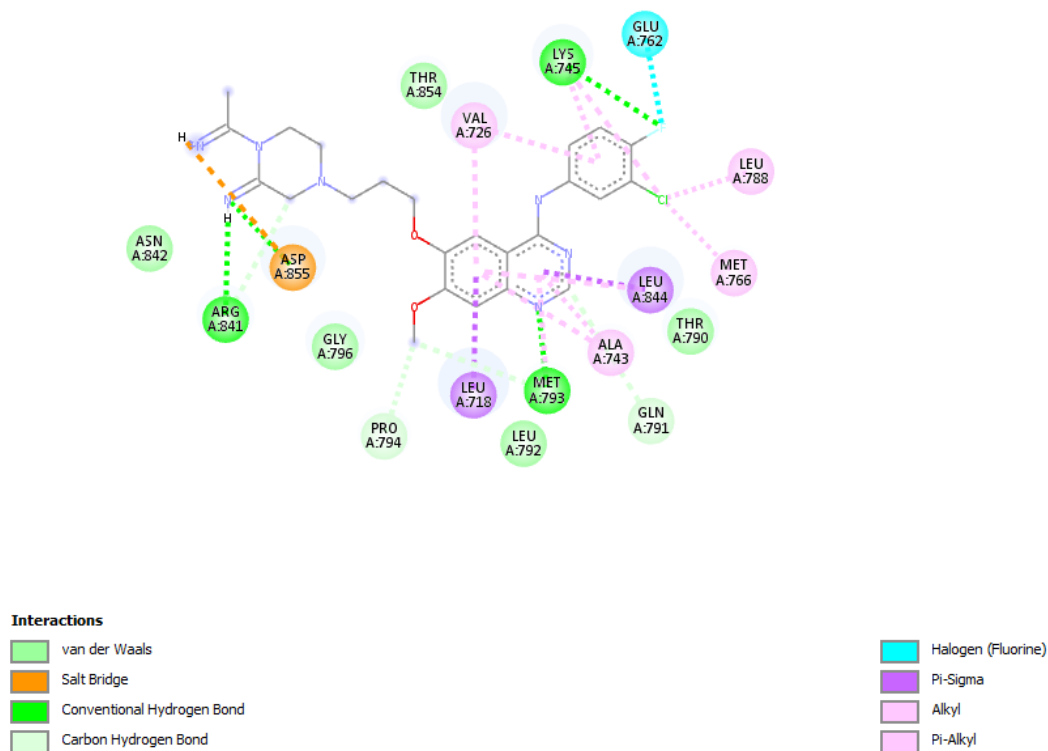


Figure 61: Mode d'interaction de la molécule n°6 dans le site actif de la 4WKQ

Tableau XIV: Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°6

	Van Der Waals	Hydrogène	Longueur d'interaction H (Å)	Pont du sel	Interaction Carbone Hydrogène	Halogène	Pi-Sigma	Alkyl et Pi Alkyl
4WKQ- molécule n°6	ASN842 GLY796 THR854 LEU792 THR790	ARG841 ASP855 LYS745 MET793	2,3454 2,95539 2,41058 1,9158	ASP855	ARG841 PRO794 MET793 GLN791	GLU762	LEU718 LEU844	VAL726 VAL726 LYS745 LYS745 LEU788 MET766 LEU844 ALA743 ALA743 MET793

## II.4.1.6. Molécule n°7

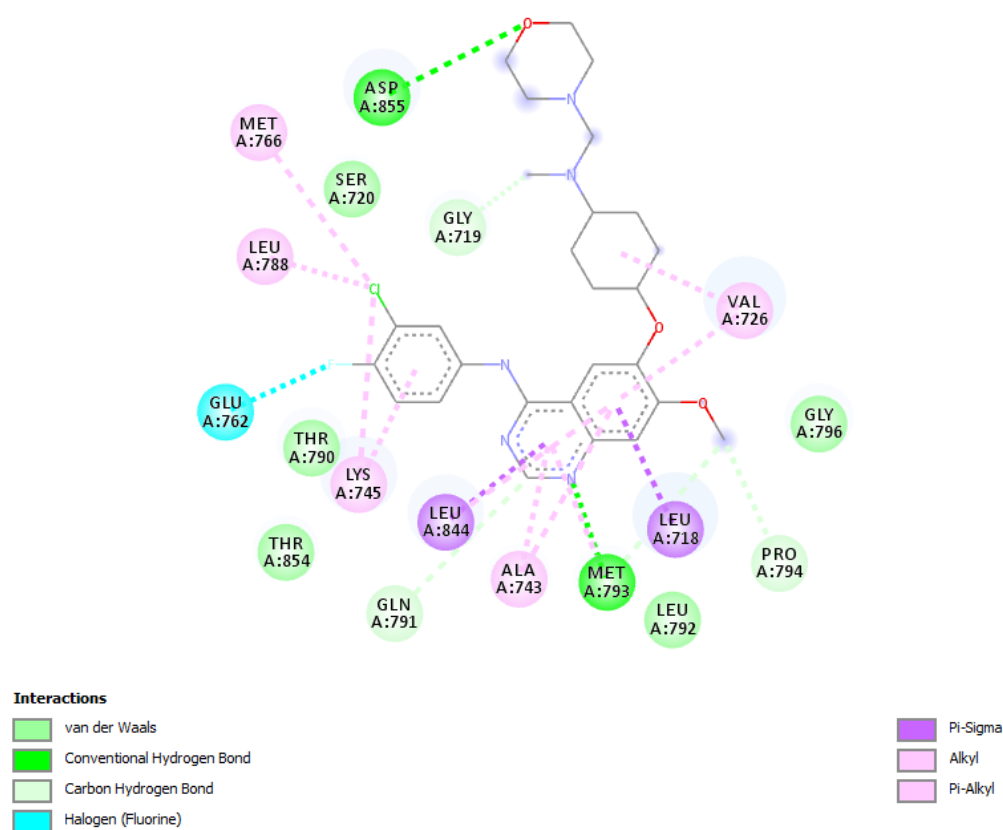


Figure 62: Mode d'interaction de la molécule n°7 dans le site actif de la 4WKQ

Tableau XV: Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°7

	Van Der Waals	Hydrogène	Longueur d'interaction H (Å)	Interaction Carbone Hydrogène	Halogène	Pi-Sigma	Alkyl et Pi-Alkyl
4WKQ- molécule n°7	SER720 THR790 THR854 LEU792 GLY796	ASP855 MET793	3,18785 2,05191	GLY719 PRO794 MET793 GLN791	GLU762	LEU718 LEU844	MET766 LEU788 LYS745 LYS745 LEU844 VAL726 VAL726 MET793 ALA743 ALA743



## II.4.1.7. Molécule n°8

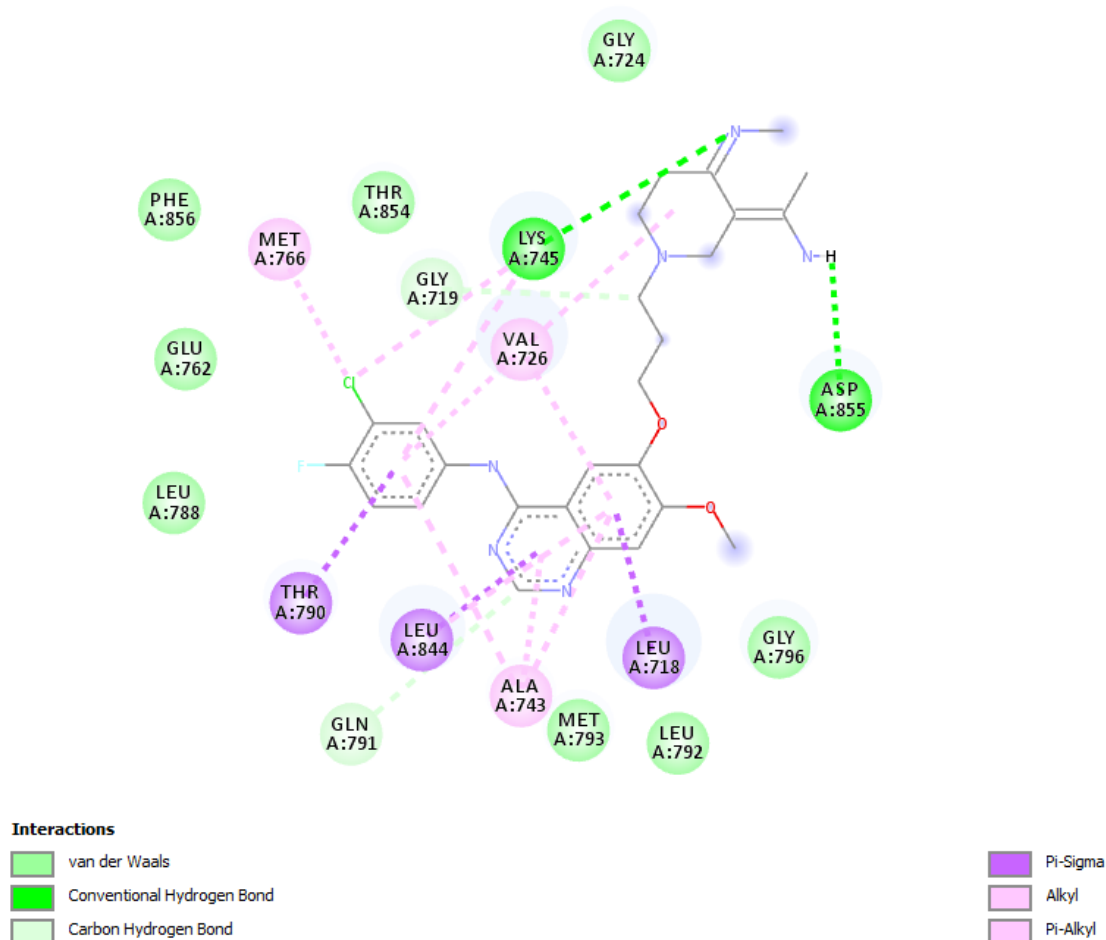


Figure 63: Mode d'interaction de la molécule n°8 dans le site actif de la 4WKQ

Tableau XVI: Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°8

	Van Der Waals	Hydrogène	Longueur d'interaction H (Å)	Interaction Carbone-Hydrogène	Pi-Sigma	Alkyl et Pi-Alkyl
4WKQ- molécule n°8	PHE856	LYS745	2,61355	GLY719	THR790	MET766
	THR854	ASP855	7,44178	GLN791	LEU844	LYS745
	GLY724				LEU718	LYS745
	GLU762					VAL726
	LEU788					VAL726
	MET793					VAL726
	LEU792					ALA743
	GLY796					ALA743
						ALA743
						LEU844

## II.4.1.8. Molécule n°9

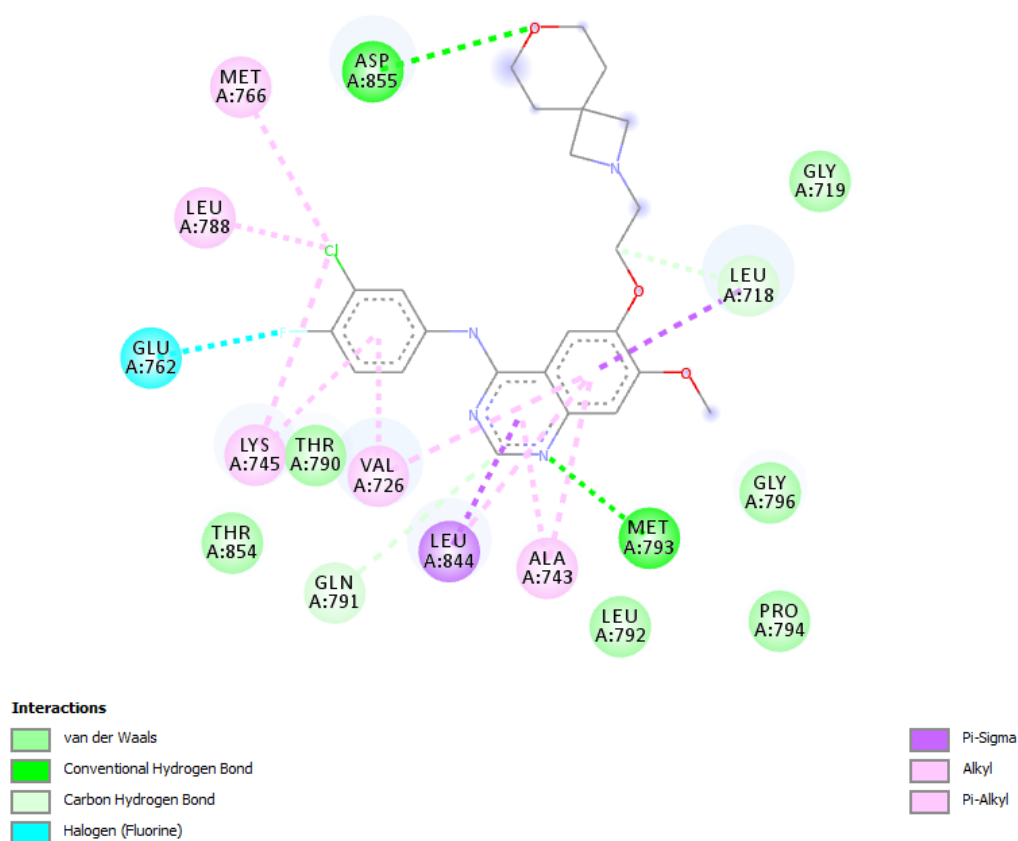


Figure 64: Mode d'interaction de la molécule n°9 dans le site actif de la 4WKQ

Tableau XVII: Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°9

	Van Der Waals	Hydrogène	Longueur d'interaction H (Å)	Interaction Carbone-hydrogène	Halogène	Pi-Sigma	Alkyl et Pi-Alkyl
4WKQ- molécule n°9	THR854 THR790 LEU792 PRO794 GLY796 GLY719	ASP855 MET793	3,29213 2,16851	LEU718 GLN791	GLU762	LEU844 LEU718	MET766 LEU788 LYS745 LYS745 VAL726 VAL726 LEU844 ALA743 ALA743

## II.4.1.9. Molécule n°10

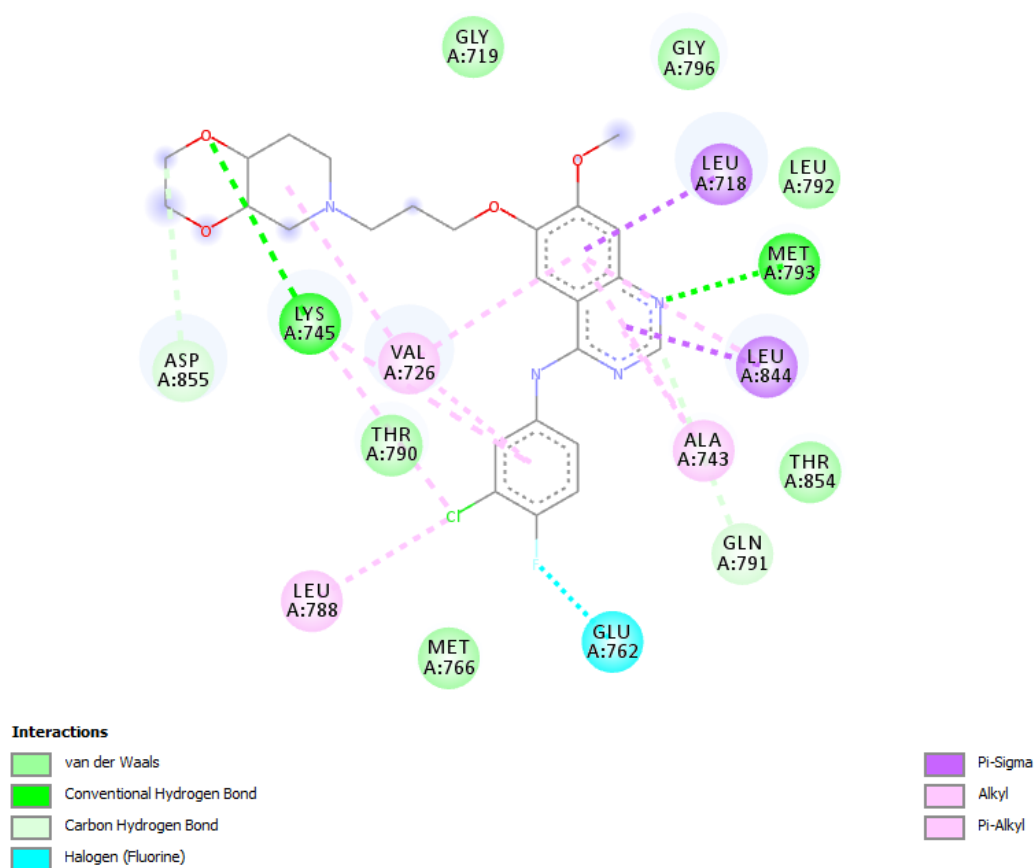


Figure 65: Mode d'interaction de la molécule n°10 dans le site actif de la 4WKQ

Tableau XVIII: Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°10

	Van Der Waals	Hydrogène	Longueur d'interaction H (Å)	Interaction Carbone-hydrogène	Halogène	Pi-Sigma	Alkyl et Pi-Alkyl
4WKQ- molécule n°10	GLY719 GLY796 LEU792 EHR790 THR854 MET766	LYS745 MET793	2,20872 2,23577	ASP855 GLN791	GLU762	LEU718 LEU844	LEU844 ALA743 VAL726 VAL726 VAL726 LYS745 LYS745 LEU788

## II.4.1.10. Molécule n°11

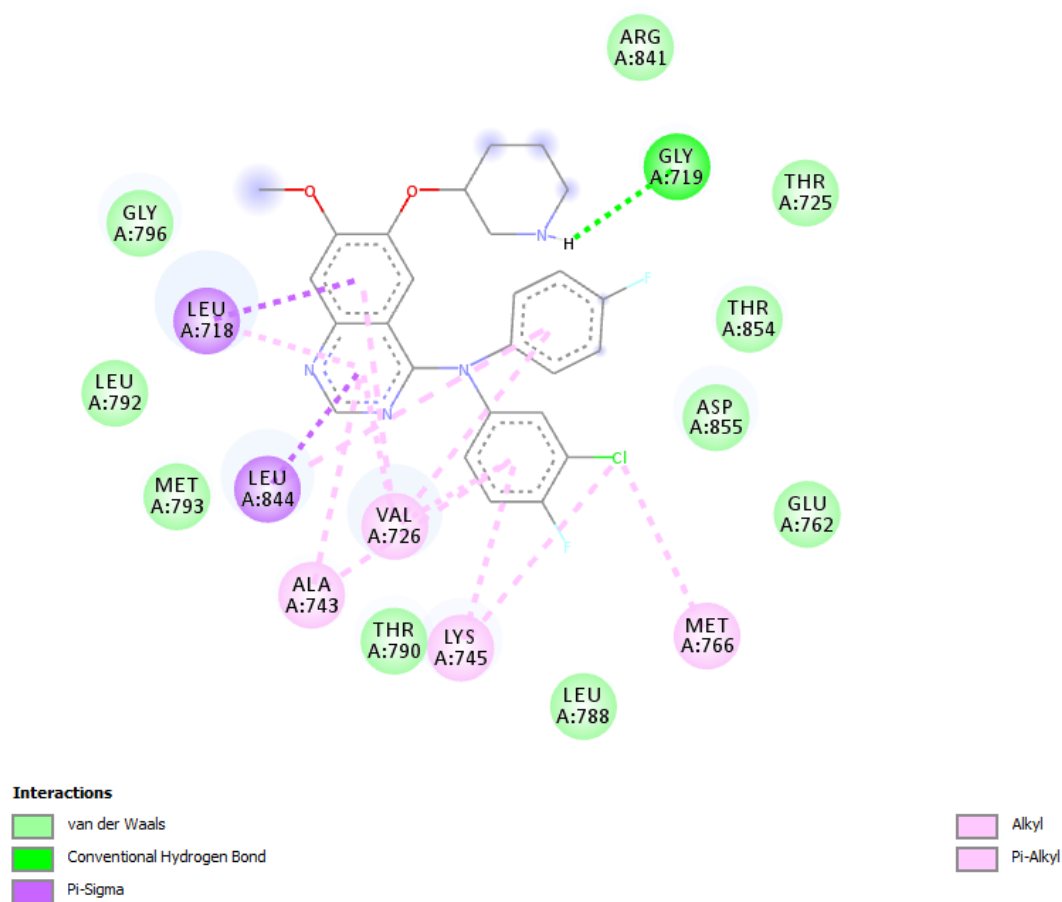


Figure 66: Mode d'interaction de la molécule n°11 dans le site actif de la 4WKQ

Tableau XIX: Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°11

	Van Der Waals	Hydrogène	Longueur d'interaction H (Å)	Pi-Sigma	Alkyl et Pi-Alkyl
4WKQ- molécule n°11	GLY796 LEU792 MET793 THR790 LEU788 ARG841 THR725 THR854 ASP855 GLU762	GLY719	2,32759	LEU718 LEU844	MET766 LYS745 LYS745 VAL726 VAL726 ALA743 ALA743 LEU844 LEU718

## II.4.1.11. Molécule n°12

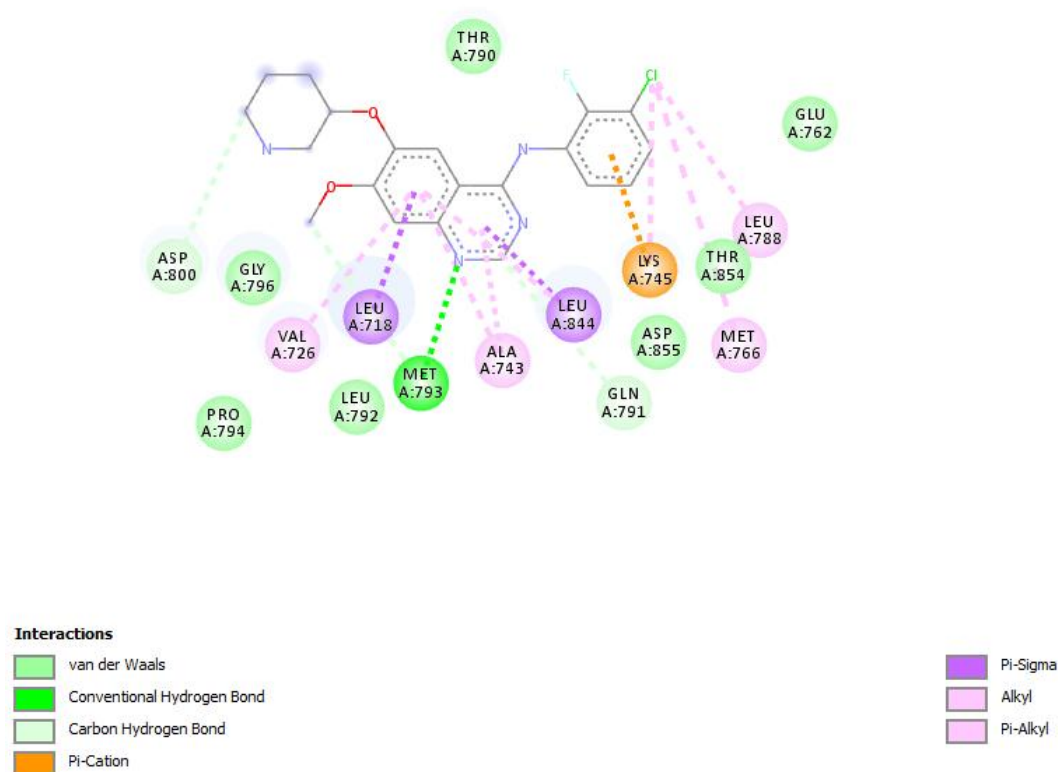


Figure 67: Mode d'interaction de la molécule n°12 dans le site actif de la 4WKQ

Tableau XX : Types d'interaction et acides aminés impliqués dans l'inhibition du 4WKQ avec la molécule n°12

	Van Der Waals	Hydrogène	Longueur d'interaction H (Å)	Interaction Carbone-hydrogène	Pi-Cation	Pi-Sigma	Alkyl et Pi-Alkyl
4WKQ- molécule n°12	THR790 GLU762 THR854 ASP855 LEU792 PRO794 GLY796	MET793	2,14861	ASP800 MET793 GLN791	LYS745	LEU844 LEU718	VAL726 ALA743 ALA743 LEU844 LYS745 MET766 LEU788

#### II.4.2. Conclusion de visualisation :

En comparant l'interaction des 11 similaires avec l'EGFR (4WKQ) par rapport à celle du gefitinib de référence, on peut constater que :

-L'analyse visuelle révèle pour les molécules (n°8 et n°11) l'absence de la liaison hydrogène avec la MET793 ce qu'ils rendent inactifs comme inhibiteurs d'EGFR donc sont éliminées des études ADME/Tox

-Et pour la boucle hydrophobe, toutes ces molécules sont stabilisées par différents types d'interaction :

- Une liaison de type Pi-cation avec la LYS 745 et trois liaisons de type alkyle avec (LYS745, MET766 et LEU788) pour les molécules (n°1, n°4, n°5 et n°12) ; en association avec d'autre interaction :
  - Trois liaisons Van der Waals avec (THR854, GLU762 et THR790) pour les molécules (n°1, n°4 et n°5) et une autre liaison Van der Waals avec ASP855 avec les molécules (n°4 et n°5)
  - Trois liaisons Van der Waals avec (THR854, ASP855 et GLU762) pour la molécule n°12
- Une liaison halogène avec GLU762 (présente aussi dans l'interaction avec le gefitinib) , 3 liaisons alkyle (LEU788 et 2 avec LYS745) et une liaison Van der Waals avec THR790 pour les molécules (n°2, n°6, n°7, n°9 et n°10) , en association avec d'autre interaction :
  - Une liaison alkyl avec MET766 et une liaison Van der Waals avec THR845 pour les molécules (n°2, n°6 et n°7)
  - Une liaison alkyl avec VAL726 pour les molécules (n°2, n°6, n°9 et n°10)
  - Une liaison hydrogène avec LYS745 pour la molécule n°6
  - Une liaison de Van der Waals avec ASP855 et MET766 pour les molécules n°2 et n°10 respectivement.

## II.5. Résultat filtrage ADME/tox :

Avant de proposer les composés comme étant de nouveaux inhibiteurs plus puissants de l'EGFR, il était indispensable de compléter notre travail par l'application de la règle de 5 de Lipinski et celle de Veber. On a également déterminé les propriétés pharmacocinétiques, ainsi des tests de toxicité aiguë ont été établit in silico. Les mêmes propriétés ont été également prédites pour la Gefetinib afin d'établir une comparaison. Les résultats sont montrés dans les tableaux ci-dessous

### II.5.1. Propriétés physico-chimiques :

**Tableau XXI: Propriétés physicochimiques**

	<b>Pm (G/Mol)</b>	<b>Tpsa Å<sup>2</sup></b>	<b>Règle de Lipinski</b>	<b>Règle de Veber</b>	<b>Solubilité dans l'eau</b>	<b>Accessibilité a la synthèse</b>
Gefetinib	446,9	68,74	OUI :0 infraction	OUI	Peu soluble	3,28
molécule n°1	500,99	68,74	OUI :1infraction : MW>500	OUI	Peu soluble	4,48
molécule n°2	508,97	68,74	OUI :1infraction : MW>500	OUI	Peu soluble	3,97
molécule n°4	494,89	68,74	OUI :0 infraction	OUI	Peu soluble	4,15
molécule n°5	486,97	68,74	OUI :0 infraction	OUI	Peu soluble	4,37
molécule n°6	500,96	112,19	OUI :1infraction : MW>500	OUI	Peu soluble	3,73
molécule n°7	530,03	71,98	OUI :1infraction : MW>500	OUI	Peu soluble	4,66
molécule n°9	472,94	68,74	OUI :0 infraction	OUI	Peu soluble	3,87
molécule n°10	502,97	77,97	OUI :1infraction : MW>500	OUI	Peu soluble	4,38
molécule n°12	402,85	68,3	OUI :0 infraction	OUI	Peu soluble	3,55

Selon Lipinski, tous les composés ne validant pas au moins trois des 5 conditions sont susceptibles de poser des problèmes d'absorption par voie orale. Dans notre cas, les 9 composés ainsi que le Gefetinib, répondent aux critères de la règle de Lipinski.

Ces composés, ont également respecté les 2 critères de la règle de Veber, concernant la surface polaire qui est largement inférieure à 140 Å<sup>2</sup>, ainsi que le nombre de liaisons rotatives qui n'a pas dépassé 10. Ce qui favorise une bonne biodisponibilité orale.

On note que ces composés ont présentés une solubilité dans l'eau plus au moins bonne. Ce qui leurs favorise le passage dans la circulation sanguine.

Par rapport à l'accessibilité à la synthèse, tous les composés cités dans le tableau peuvent être chimiquement synthétisables comme les résultats le montrent dont les valeurs n'ont pas dépassé l'échelle de 10.



**II.5.2. Propriétés pharmacocinétiques :**

Il était utile de compléter le travail par une analyse des propriétés pharmacocinétiques afin d'affirmer notre sélection et d'aller vers le ou les composés ayant le plus de chance d'être un candidat médicament

Comme critère de choix ce/ces médicament(s) doit être administrer par voie orale donc présenter par une bonne absorption GI, les cancers du cerveau ne sont pas ciblés donc pas de perméabilités BBB.

**Tableau XXII: Propriétés pharmacocinétiques**

	Absorption GI	BBB Perméabilité	Substrat gluco prot P	Inhibiteur de CYP1A2	Inhibiteur de CYP2C19	Inhibiteur de CYP2C9	Inhibiteur de CYP2D6	Inhibiteur de CYP3A4
Gefetinib	Haute	OUI	NON	NON	OUI	OUI	OUI	OUI
molécule n°1	Haute	NON	OUI	NON	NON	NON	OUI	OUI
molécule n°2	Haute	NON	OUI	OUI	NON	OUI	OUI	OUI
molécule n°4	Haute	NON	OUI	NON	NON	OUI	OUI	OUI
molécule n°5	Haute	NON	OUI	NON	NON	OUI	OUI	OUI
molécule n°6	Haute	NON	OUI	NON	OUI	OUI	OUI	OUI
molécule n°7	Haute	NON	OUI	NON	NON	NON	OUI	OUI
molécule n°9	Haute	OUI	OUI	NON	OUI	OUI	OUI	OUI
molécule n°10	Haute	NON	OUI	NON	NON	OUI	OUI	OUI
molécule n°12	Haute	OUI	OUI	OUI	NON	OUI	OUI	OUI

Comme pour le gefitinib, une absorption gastro-intestinale, haute a été constatée et une perméabilité hémato-encéphalique pour les molécules (n°9 et n12).

## II.5.3. Toxicité:

Dans le but de vérifier la toxicité des composés proposés, plusieurs paramètres ont été déterminés in silico. Parmi eux, le test d'Ames qui nous renseigne sur la génotoxicité.

Tableau XXIII : Résultats de la toxicité

	algae_ at	Ames test	Carcino Mouse	Carcino Rat	daphnia_ at	hERG inhibition	medaka_ at	minnow_ at
Gefitinib	0.0068 2372	mutagen	positive	negative	0.01804 3	Medium risk	0.00079 114	0.00101 369
molécule n°1	0,0022 22272	mutagen	negative	negative	0,00474 107	Medium risk	0,31973 e-005	7,08891e -005
molécule n°2	0,0023 1621	mutagen	negative	negative	0,00240 382	Medium risk	1,89935 e-005	9,13298e -005
molécule n°4	0,0047 6514	Non mutagen	positive	negative	0,01316 78	Medium risk	0,00043 7164	0,00042 9837
molécule n°5	0,0026 4954	Non mutagen	positive	negative	0,00698 966	Medium risk	0,00013 1414	0,00012 9852
molécule n°6	0,0084 7423	Non mutagen	negative	negative	0,05697 68	Medium risk	0,00843 755	0,01054 74
molécule n°7	0,0023 1906	mutagen	Negative	Negative	0,00569 807	Medium Risk	9,3058e- 005	0,00010 0706
molécule n°9	0,0047 3822	Mutagen	Negative	Negative	0,01308 97	Medium Risk	0,00044 2188	0,00064 9405
molécule n°10	0,0057 0626	Mutagen	Negative	Negative	0,01867 7	Medium Risk	0,00090 312	0,00126 019
molécule n°12	0,0080 2699	Mutagen	Negative	Negative	0,01676 82	Medium risk	0,00065 832	0,00063 4729

-En comparaison avec le gefitinib, les molécules (n°1, n°2, n°7, n°9, n°10 et n°12) présentent un effet mutagène qui se traduit par leur possibilité à induire des modifications au niveau du matériel génétique, comme le gefitinib contrairement aux molécules (n°4, n°5 et n°6), qui révèlent un effet non mutagène.

-Le pouvoir cancérigène des composés a été également vérifié. On note un résultat négatif total concernant le test sur les rats et positive pour le gefitinib et les molécules (n°4 et n°5)

-Concernant le test d'inhibition du gène hERG, les mêmes résultats a été démontré pour la totalité des composés, témoignant ainsi un risque moyen vis-à-vis de l'arythmie cardiaque.

-Les résultats des tests utilisant les espèces aquatiques cités dans le Tableau XXIII, montrent que la molécule n°6 présente un taux supérieur à celui révélé par le gefitinib.

En conclusion, les molécules n°4, n°5 et n°6 proposées pour le filtrage ADME / tox répondent favorablement à l'ensemble de critères, ce qui ouvre la voie à l'optimisation de ces molécules comme candidats médicaments.

## CONCLUSION ET PERSPECTIVES :

Afin de contribuer au développement *in silico* de nouveaux inhibiteurs des récepteurs du facteur de croissance épidermique (EGFR), cible enzymatique récente jouant un rôle clé dans le traitement du cancer. Nous avons effectué un criblage virtuel basé sur le ligand.

Au préalable, on a jugé utile de vérifier la fiabilité du programme de docking choisit (AutoDock Vina intégrée dans PyRx) par deux tests différents : L'analyse visuelle et l'étude de la corrélation linéaire. Après la validation de notre protocole de docking, nous avons sélectionné un ligand de départ issu de la littérature (gefitinib), en complexe avec le récepteur EGFR (4WKQ). A partir de squelette de base du gefitinib, la banque de donnée PubChem nous a donné une collection de 353 molécules avec une similarité de 97%. Les calculs de docking moléculaire ont permis de ressortir les 12meilleurs molécules avec des scores entre (-9,3 et -9,8 Kcal/Mole) ; c'est à dire une énergie d'interaction considérablement inférieure à celle du composé de référence qui s'évalue à -8,2 Kcal/Mole.

Enfin, l'application des règles de Lipinski et Veber, la vérification des propriétés physicochimique et les paramètres pharmacocinétiques, ainsi que les tests de toxicité potentielle *in silico* via les serveurs en ligne SWISS ADME et PREADMET ; nous ont orienté vers les molécules n°4, n°5 et n°6 (CID\_137455767, CID\_24875330 et CID\_145577260respectivement). Ces trois molécules sont les candidats médicament qui se présentent comme des nouveaux inhibiteurs potentiellement plus puissants envers l'EGFR.

En termes de perspectives d'avenir, ce travail pourra être poursuivi par des études expérimentales *in vitro* et/ou *in vivo* pour vérifier l'admissibilité de ces molécules aux phases suivantes de développement des médicaments.

## Bibliographie :

- [1] Hanahan D, Weinberg RA. The Hallmarks of Cancer. *Cell*. 7 janv 2000;100(1):57-70.
- [2] Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. *Cell*. 4 mars 2011;144(5):646-74.
- [3] Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1 juin 1990;61(5):759-67.
- [4] Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med*. août 2004;10(8):789-99.
- [5] Mudduluru G, Medved F, Grobholz R, Jost C, Gruber A, Leupold JH, et al. Loss of programmed cell death 4 expression marks adenoma-carcinoma transition, correlates inversely with phosphorylated protein kinase B, and is an independent prognostic factor in resected colorectal cancer. *Cancer*. 2007;110(8):1697-707.
- [6] Que sait-on du cancer ? [Internet]. [cité 17 juin 2021]. Disponible sur: </produit/137/9782759808700/Que%20sait-on%20du%20cancer%20>
- [7] Les bases du cancer - oncoblog! [Internet]. [cité 17 juin 2021]. Disponible sur: <http://sarah7476.over-blog.com/article-21900727.html>
- [8] de Cremoux P, Robert J. Signalisation cellulaire et cancer : caractérisation de cibles thérapeutiques. *Pathol Biol*. août 2012;60(4):217-22.
- [9] Bolteau R. Conception, synthèse et évaluation pharmacologique d'antagonistes des récepteurs A2A et de ligands duaux ciblant les récepteurs A2A et mGlu5. 2020;350.
- [10] Dufort et al. - 2007 - 053 Vectorisation de biomolécules pour l'imagerie .pdf [Internet]. [cité 18 juin 2021]. Disponible sur: [https://tel.archives-ouvertes.fr/tel-00531238/file/these\\_Sandrine\\_Dufort.pdf](https://tel.archives-ouvertes.fr/tel-00531238/file/these_Sandrine_Dufort.pdf)
- [11] Cohen S (1965) The stimulation of epidermal proliferation by a specific protein (EGF). *Dev Biol*12: 394-407.

- [12] Cohen S, Carpenter G, Lembach KJ (1975) Interaction of epidermal growth factor (EGF) with cultured fibroblasts. *Adv Metab Disord* 8: 265–284.
- [13] Carpenter G, King L Jr., Cohen S (1978) Epidermal growth factor stimulates phosphorylation in membrane preparations in vitro. *Nature* 276: 409–410.
- [14] Moitessier, N., et al., Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *British journal of pharmacology*, 2008. 153 Suppl: p. S7-S26.
- [15] Coupez, B. and R.A. Lewis, Docking and scoring - theoretically easy, practically impossible? *Current medicinal chemistry*, 2006. 13: p. 2995-3003.
- [16] Lemmon MA, Schlessinger J (2010) Cell signaling by receptor tyrosine kinases. *Cell* 141: 1117–1134.
- [17] Lemmon MA, Schlessinger J, Ferguson KM (2014) The EGFR family: not so prototypical receptor tyrosine kinases. *Cold Spring Harb Perspect Biol* 6: a020768.
- [18] Kovacs E, Zorn JA, Huang Y, et al. (2015) A Structural Perspective on the Regulation of the Epidermal Growth Factor Receptor. *Annu Rev Biochem*. E-pub ahead of print.
- [19] Ullrich A, Schlessinger J (1990) Signal transduction by receptors with tyrosine kinase activity. *Cell* 61: 203–212.
- [20] Nolen B, Taylor S, Ghosh G (2004) Regulation of protein kinases; controlling activity through activation segment conformation. *Mol Cell* 15: 661–675.
- [21] Cui, J.J., Inhibitors targeting hepatocyte growth factor receptor and their potential therapeutic applications. *Expert opinion on therapeutic patents*, 2007. 17: p. 1035- 1045.
- [22] Underiner, T.L., T. Herberitz, and S.J. Miknyoczki, Discovery of small molecule c-Met inhibitors: Evolution and profiles of clinical candidates. *Anti-cancer agents in medicinal chemistry*, 2010. 10: p. 7-27.
- [23] Asses, Y., et al., Analysis of c-Met kinase domain complexes: a new specific catalytic site receptor model for defining binding modes of ATP-competitive ligands. *Chemical biology & drug design*, 2009. 74: p. 560-570.

- [24] Forbes SA, Tang G, Bindal N, et al. (2010) COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res* 38:D652–657.
- [25] Yuzawa S, Opatowsky Y, Zhang Z, et al. (2007) Structural basis for activation of the receptor tyrosine kinase KIT by stem cell factor. *Cell* 130: 323–334.
- [26] Corless CL, Heinrich MC (2008) Molecular pathobiology of gastrointestinal stromal sarcomas. *Annu Rev Pathol* 3: 557–586.
- [27] Toffalini F, Demoulin JB (2010) New insights into the mechanisms of hematopoietic cell transformation by activated receptor tyrosine kinases. *Blood* 116: 2429–2437.
- [28] Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144: 646–674.
- [29] The ErbB receptors and their role in cancer progression - PubMed [Internet]. [cité 21 juin 2021]. Disponible sur: <https://pubmed.ncbi.nlm.nih.gov/12648469/>
- [30] Schlessinger J. Cell signaling by receptor tyrosine kinases. *Cell*. 13 oct 2000;103(2):211-25.
- [31] Blume-Jensen P, Hunter T. Oncogenic kinase signalling. *Nature*. 17 mai 2001;411(6835):355-65.
- [32] Bennisroune A, Gardin A, Aunis D, Crémel G, Hubert P. Tyrosine kinase receptors as attractive targets of cancer therapy. *Crit Rev Oncol Hematol*. avr 2004;50(1):23-38.
- [33] T H, Ne H. ErbB receptors: directing key signaling networks throughout life. *Annu Rev Pharmacol Toxicol* [Internet]. 2004 [cité 21 juin 2021];44. Disponible sur: <https://pubmed.ncbi.nlm.nih.gov/14744244/>
- [34] Ferguson KM. Active and inactive conformations of the epidermal growth factor receptor. *Biochem Soc Trans*. nov 2004;32(Pt 5):742-5.
- [35] Hubbard SR, Till JH. Protein tyrosine kinase structure and function. *Annu Rev Biochem*. 2000;69:373-98.

- [36] Jorissen RN, Walker F, Pouliot N, Garrett TPJ, Ward CW, Burgess AW. Epidermal growth factor receptor: mechanisms of activation and signalling. *Exp Cell Res.* 10 mars 2003;284(1):31-53.
- [37] Casalini P, Iorio MV, Galmozzi E, Ménard S. Role of HER receptors family in development and differentiation. *J Cell Physiol.* sept 2004;200(3):343-50.
- [38] Schlessinger J. Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell.* 20 sept 2002;110(6):669-72.
- [39] An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor - PubMed [Internet]. [cité 21 juin 2021]. Disponible sur: <https://pubmed.ncbi.nlm.nih.gov/16777603/>
- [40] EGF R Protein - ACROBiosystems [Internet]. [cité 21 juin 2021]. Disponible sur: [https://www.acrobiosystems.com/L-192-EGF%20R.html?gclid=CjwKCAjwzruGBhBAEiwAUqMR8LO1ibdBh9pgAAqpsxeYo6-UTjpTWd4ttbk9o9bhzv2IZv9s2EimHBoCSOYQAvD\\_BwE](https://www.acrobiosystems.com/L-192-EGF%20R.html?gclid=CjwKCAjwzruGBhBAEiwAUqMR8LO1ibdBh9pgAAqpsxeYo6-UTjpTWd4ttbk9o9bhzv2IZv9s2EimHBoCSOYQAvD_BwE)
- [41] ErbB 3 (ERBB3) Rabbit Polyclonal Antibody – TA590106 | OriGene [Internet]. [cité 21 juin 2021]. Disponible sur: [https://www.origene.com/catalog/antibodies/primary-antibodies/ta590106/erbb-3-erbb3-rabbit-polyclonal-antibody?gclid=CjwKCAjwzruGBhBAEiwAUqMR8FQgM4ZBG1749GiqJsdmRwfNcVQ8AormgS2E3DHPi8xj5wxWsVTXjhoCr0MQAvD\\_BwE](https://www.origene.com/catalog/antibodies/primary-antibodies/ta590106/erbb-3-erbb3-rabbit-polyclonal-antibody?gclid=CjwKCAjwzruGBhBAEiwAUqMR8FQgM4ZBG1749GiqJsdmRwfNcVQ8AormgS2E3DHPi8xj5wxWsVTXjhoCr0MQAvD_BwE)
- [42] Ciardiello F, Tortora G. EGFR antagonists in cancer treatment. *N Engl J Med.* 13 mars 2008;358(11):1160-74.
- [43] Biologics & Small Molecules Research [Internet]. Sartorius. [cité 21 juin 2021]. Disponible sur: <https://www.sartorius.com/en/applications/life-science-research/label-free-detection/biologics-small-molecules-research>
- [44] Robert J. Signalisation cellulaire et cancer Un manuel pour les étudiants et les oncologues [Internet]. 2010 [cité 25 avr 2021]. Disponible sur: <https://doi.org/10.1007/978-2-8178-0028-8>



- [45] Gschwind A, Fischer OM, Ullrich A. The discovery of receptor tyrosine kinases: targets for cancer therapy. *Nat Rev Cancer*. mai 2004;4(5):361-70.
- [46] Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol*. févr 2001;2(2):127-37.
- [47] EGFR mutation and resistance of non-small-cell lung cancer to gefitinib - PubMed [Internet]. [cité 21 juin 2021]. Disponible sur: <https://pubmed.ncbi.nlm.nih.gov/15728811/>
- [48] Engelman JA, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, Park JO, et al. MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science*. 18 mai 2007;316(5827):1039-43.
- [49] Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst*. 2 mars 2005;97(5):339-46.
- [50] Sordella R, Bell DW, Haber DA, Settleman J. Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. *Science*. 20 août 2004;305(5687):1163-7.
- [51] Drug Discovery | Cryo EM | Thermo Fisher Scientific - DZ [Internet]. [cité 29 juin 2021]. Disponible sur: [https://www.thermofisher.com/dz/en/home/electron-microscopy/life-sciences/drug-discovery.html?cid=cmp-03309-y8k7&utm\\_source=google&utm\\_medium=digital-paid&utm\\_campaign=2020-ls-learning-center&gclid=CjwKCAjwieuGBhAsEiwA1Ly\\_nVnRxwZ\\_mMC3iUNQARaOCEFKNLjPTfTJxPQRtoof\\_QzjNiDjgirfBRoCM7kQAvD\\_BwE](https://www.thermofisher.com/dz/en/home/electron-microscopy/life-sciences/drug-discovery.html?cid=cmp-03309-y8k7&utm_source=google&utm_medium=digital-paid&utm_campaign=2020-ls-learning-center&gclid=CjwKCAjwieuGBhAsEiwA1Ly_nVnRxwZ_mMC3iUNQARaOCEFKNLjPTfTJxPQRtoof_QzjNiDjgirfBRoCM7kQAvD_BwE)
- [52] Livage J. Chimie de la matière condensée. L'annuaire Collège Fr Cours Trav. 1 mars 2010;(109):213-8.
- [53] Über den jetzigen Stand der Chemotherapie - Ehrlich - 1909 - Berichte der deutschen chemischen Gesellschaft - Wiley Online Library [Internet]. [cité 29 juin 2021]. Disponible sur: <https://chemistry-europe.onlinelibrary.wiley.com/doi/10.1002/cber.19090420105>

- [54] Qing X, Lee XY, Raeymaecker JD, Tame JR, Zhang KY, Maeyer MD, et al. Pharmacophore modeling: advances, limitations, and current utility in drug discovery. *J Recept Ligand Channel Res.* 11 nov 2014;7:81-92.
- [55] *Chemobiodynamics and Drug Design.* By F. W. SCHUELER. The Blakiston Division, McGrawHill Book Co., Inc., 330 West 42nd St., New York 36, N. Y., 1960. 15 × 23 cm. Price \$19.50. xiv + 638 pp. *J Pharm Sci.* 1961;50(1):92-92.
- [56] Yang S-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov Today.* juin 2010;15(11-12):444-50.
- [57] Güner OF, Bowen JP. Setting the record straight: the origin of the pharmacophore concept. *J Chem Inf Model.* 27 mai 2014;54(5):1269-83.
- [58] Wermuth CG, Ganellin CR, Lindberg P, Mitscher LA. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl Chem.* 1 janv 1998;70(5):1129-43.
- [59] Pharmacophore Perception, Development and Use in Drug Design [Internet]. [cité 29 juin 2021]. Disponible sur: <https://www.mdpi.org/molecules/html/50700987.htm>
- [60] Schneider G, Fechner U. Computer-based de novo design of drug-like molecules. *Nat Rev Drug Discov.* août 2005;4(8):649-63.
- [61] Dror O, Shulman-Peleg A, Nussinov R, Wolfson HJ. Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design. *Curr Med Chem.* janv 2004;11(1):71-90.
- [62] The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure | SpringerLink [Internet]. [cité 29 juin 2021]. Disponible sur: <https://link.springer.com/article/10.1007/BF00126743>
- [63] Sanders MPA, McGuire R, Roumen L, Esch IJP de, Vlieg J de, Klomp JPG, et al. From the protein's perspective: the benefits and challenges of protein structure-based pharmacophore modeling. *MedChemComm.* 1 janv 2012;3(1):28-38.

- [64] GmbH JB. Fragment Screen [Internet]. Jena Bioscience. [cité 29 juin 2021]. Disponible sur: <https://www.jenabioscience.com/crystallography-cryo-em/screening/fragment-screen>
- [65] Biochimie GARRETT , GRISHAM De boeck superieur - 9782744500206 [Internet]. [cité 29 juin 2021]. Disponible sur: <https://www.unitheque.com/biochimie/de-boeck-superieur/Livre/934>
- [66] Atkins P.W.1998. Eléments de chimie physique. Paris : De Boeck Université, 510p.
- [67] Rappe A.K., Casewit C.J. 1997. Molecular Mechanics Across Chemistry. Eds: University Science books Sausalito, California. 444 p.
- [68] Gilbas C., Jambeck P. 2001. Introduction à la bioinformatique. France : O'Reilly. 375p.
- [69] Honig B.J. 1999. Protein folding: from the Levinthal paradox to structure prediction. J.Mol Biol. 293, 283-293.
- [70] Sotriffer, C., Klebe, G., Stahl, M., and Böhm, H.-J., Docking and Scoring Functions/Virtual Screening in Burger's Medicinal Chemistry and Drug Discovery., ed. I. John Wiley & Sons. 2003. 281–331.
- [71] Kitchen, D.B., Decornez, H., Furr, J.R., and Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. Nat. Rev. Drug Discovery, 2004. 3(11): p. 935-949.
- [72] N, O.B. Protein-Ligand Docking. 2010; Available from: <http://fr.slideshare.net/baoilleach/proteinligand-docking>.
- [73] Spyraakis, F., Cozzini, P. and Kellogg, G. E. , Docking and Scoring in Drug Discovery in Burger's Medicinal Chemistry and Drug Discovery. . 2010. 601–684.
- [74] Lichtenthaler, F.W., 100 Years "Schlüssel-Schloss-Prinzip": What Made Emil Fischer Use this Analogy? Angew. Chem. Int. Ed., 1995. 33(23-24): p. 2364-2374.
- [75] Vuorinen, A. and Schuster, D., Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. Methods (Amsterdam, Neth.), 2015. 71: p. 113-134.

- [76] Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., and Ferrin, T.E., A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.*, 1982. 161(2): p. 269- 288.
- [77] Barril, X. and Fradera, X., Incorporating protein flexibility into docking and structurebased drug design. *Expert Opin. Drug Discovery*, 2006. 1(4): p. 335-349.
- [78] Carlson, H.A. and McCammon, J.A., Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.*, 2000. 57(2): p. 213-218.
- [79] Brooijmans, N. and Kuntz, I.D., Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.*, 2003. 32: p. 335-373.
- [80] Sousa, S.F., Fernandes, P.A., and Ramos, M.J., Protein-ligand docking: current status and future challenges. *Proteins: Struct., Funct., Bioinf.*, 2006. 65(1): p. 15-26.
- [81] Jain, A.N., Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem*, 2003. 46(4): p. 499-511.
- [82] DesJarlais, R.L., Sheridan, R.P., Dixon, J.S., Kuntz, I.D., and Venkataraghavan, R., Docking flexible ligands to macromolecular receptors by molecular shape. *J. Med. Chem.*, 1986. 29(11): p. 2149-53.
- [83] Abagyan, R., Totrov, M., and Kuznetsov, D., ICM - a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.*, 1994. 15(5): p. 488-506.
- [84] Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., Shaw, D.E., Francis, P., and Shenkin, P.S., Glide: A new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *J. Med. Chem.*, 2004. 47(7): p. 1739- 1749.
- [85] Lameijer, E.-W., Baeck, T., Kok, J.N., and Ijzerman, A.P., Evolutionary algorithms in drug design. *Nat. Comput.*, 2005. 4(3): p. 177-243.
- [86] Jones, G., Willett, P., Glen, R.C., Leach, A.R., and Taylor, R., Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, 1997. 267(3): p. 727-748.

- [87] Westhead, D.R., Clark, D.E., and Murray, C.W., A comparison of heuristic search algorithms for molecular docking. *J. Comput.-Aided Mol. Des.*, 1997. 11(3): p. 209-228.
- [88] Baxter, C.A., Murray, C.W., Clark, D.E., Westhead, D.R., and Eldridge, M.D., Flexible docking using tabu search and an empirical estimate of binding affinity. *Proteins: Struct., Funct., Genet.*, 1998. 33(3): p. 367-382.
- [89] Di Nola, A., Roccatano, D., and Berendsen, H.J.C., Molecular dynamics simulation of the docking of substrates to proteins. *Proteins: Struct., Funct., Genet.*, 1994. 19(3): p. 174-82.
- [90] Oda, A., Tsuchida, K., Takakura, T., Yamaotsu, N., and Hirono, S., Comparison of Consensus Scoring Strategies for Evaluating Computational Models of Protein-Ligand Complexes. *J. Chem. Inf. Model.*, 2006. 46(1): p. 380-391.
- [91] Charifson, P.S., Corkery, J.J., Murcko, M.A., and Walters, W.P., Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of ThreeDimensional Structures into Proteins. *J. Med. Chem.*, 1999. 42(25): p. 5100-5109.
- [92] Huang, S.-Y., Grinter, S.Z., and Zou, X., Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys.*, 2010. 12(40): p. 12899-12908.
- [93] Hawkins, P.C.D., Warren, G.L., Skillman, A.G., and Nicholls, A., How to do an evaluation: pitfalls and traps. *J. Comput.-Aided Mol. Des.*, 2008. 22(3-4): p. 179-190.
- [94] Davis, A.M., Teague, S.J., and Kleywegt, G.J., Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew. Chem., Int. Ed.*, 2003. 42(24): p. 2718-2736.
- [95] Raymer, M.L., Sanschagrin, P.C., Punch, W.F., Venkataraman, S., Goodman, E.D., and Kuhn, L.A., Predicting conserved water-mediated and polar ligand interactions in protein using a K-nearest-neighbors genetic algorithm. *J. Mol. Biol.*, 1997. 265(4): p. 445-464.

- [96] Rarey, M., Kramer, B., Lengauer, T., and Klebe, G., A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.*, 1996. 261(3): p. 470-489.
- [97] [vinna.scioos.edu](http://vinna.scioos.edu).
- [98] <http://www.swissadme.ch/>.
- [99] <https://preadmet.bmdrc.kr>.
- [100] <https://www.rcsb.org/>.
- [101] <https://pubchem.ncbi.nlm.nih.gov/>.
- [102] Veber D. F., Johnson S. R., Cheng H. Y., Smith B. R., Ward K. W., Kopple K. D. Molecular Properties That Influence Chem. 2002, 45, 2615-2623.
- [103] Akik W. Etude de la perméabilité intestinale des médicaments par la reconstitution du transporteur BCRP/ABCG2 dans des protéoliposomes. Mémoire en pharmacie : université de montreal.2017.65p.
- [104] INFORBAL CHIMIE, [Enligne]. <http://www.inforlab-chimie.fr>.
- [105] Fontana E. Les inhibiteurs "suicides" des Cytochromes p450 : Etablissement d'une banque de données, mise au point d'un test de screening et études structures-activité concernant des substrats furaniques du CYP 3A4.these de doctorat : Université René Descartes. PARIS V.2005.207p.
- [106] <https://www.rcsb.org/structure/6SDC>.
- [107] <https://www.rcsb.org/structure/5GMN>.
- [108] <https://www.rcsb.org/structure/3WZE>.
- [109] <https://www.rcsb.org/structure/6JOL>.
- [110] Lifirafenib (BGB-283) |  $\geq 99\%$ (HPLC) | Selleck | Raf inhibitor [Internet]. [selleckchem.com](https://www.selleckchem.com). [cité 11 sept 2021]. Disponible sur: <https://www.selleckchem.com/products/bgb-283-bgb283.html>

- [111] Icotinib (BPI-2009H) |  $\geq 99\%$ (HPLC) | Selleck | EGFR inhibitor [Internet].  
selleckchem.com. [cité 11 sept 2021]. Disponible sur:  
<https://www.selleckchem.com/products/icotinib.html>
- [112] Sapitinib (AZD8931) |  $\geq 99\%$ (HPLC) | Selleck | HER2 inhibitor [Internet].  
selleckchem.com. [cité 11 sept 2021]. Disponible sur:  
<https://www.selleckchem.com/products/AZD8931.html>
- [113] Aperçu Scopus - Scopus - Bienvenue dans Scopus [Internet]. [cité 11 sept 2021].  
Disponible sur: <https://www.scopus.com/home.uri>
- [114] Zorifertinib (AZD3759) | Inhibiteur d'EGFR | MedChemExpress [Internet]. [cité 11  
sept 2021]. Disponible sur: <https://www.medchemexpress.com/AZD3759.html>
- [115] Tesevatinib, inhibiteur [Internet]. [cité 11 sept 2021]. Disponible sur:  
<https://www.mybiosource.com/inhibitor/tesevatinib/5764422>
- [116] Nazartinib (EGF816) | EGFR Inhibitor | MedChemExpress [Internet].  
MedchemExpress.com. [cité 11 sept 2021]. Disponible sur:  
<https://www.medchemexpress.com/EGF816.html>
- [117] Poziotinib (HM781-36B) |  $\geq 99\%$ (HPLC) | Selleck | EGFR inhibitor [Internet].  
selleckchem.com. [cité 11 sept 2021]. Disponible sur:  
<https://www.selleckchem.com/products/poziotinib-hm781-36b.html>
- [118] Allitinib (AST-1306) |  $\geq 99\%$ (HPLC) | Selleck | EGFR inhibitor [Internet].  
selleckchem.com. [cité 11 sept 2021]. Disponible sur:  
<https://www.selleckchem.com/products/ast-1306.html>
- [119] Lazertinib |  $\geq 99\%$ (HPLC) | Selleck | EGFR inhibitor [Internet]. selleckchem.com. [cité  
11 sept 2021]. Disponible sur: [https://www.selleckchem.com/products/lazertinib-  
yh25448-gns-1480.html](https://www.selleckchem.com/products/lazertinib-yh25448-gns-1480.html)
- [120] Dacomitinib (PF-00299804) |  $\geq 99\%$ (HPLC) | Selleck | EGFR inhibitor [Internet].  
selleckchem.com. [cité 11 sept 2021]. Disponible sur:  
<https://www.selleckchem.com/products/pf299804.html>

- [121] Olmutinib (HM61713, BI 1482694) | EGFR Inhibitor | MedChemExpress [Internet]. MedchemExpress.com. [cité 11 sept 2021]. Disponible sur: <https://www.medchemexpress.com/Olmutinib.html>
- [122] Canertinib (CI-1033) |  $\geq 99\%$ (HPLC) | Selleck | EGFR inhibitor [Internet]. selleckchem.com. [cité 11 sept 2021]. Disponible sur: [https://www.selleckchem.com/products/CI-1033\(Canertinib\).html](https://www.selleckchem.com/products/CI-1033(Canertinib).html)
- [123] Naquotinib (ASP8273) | EGFR Inhibitor | MedChemExpress [Internet]. [cité 11 sept 2021]. Disponible sur: <https://www.medchemexpress.com/Naquotinib.html>
- [124] Ibrutinib (PCI-32765) |  $\geq 99\%$ (HPLC) | Selleck | Target Protein Ligand chemical [Internet]. selleckchem.com. [cité 11 sept 2021]. Disponible sur: <https://www.selleckchem.com/products/pci-32765.html>
- [125] Pyrotinib (SHR-1258) dimaleate |  $\geq 99\%$ (HPLC) | Selleck | EGFR inhibitor [Internet]. selleckchem.com. [cité 11 sept 2021]. Disponible sur: <https://www.selleckchem.com/products/pyrotinib.html>
- [126] Pelitinib (EKB-569) |  $\geq 99\%$ (HPLC) | Selleck | EGFR inhibitor [Internet]. selleckchem.com. [cité 11 sept 2021]. Disponible sur: <https://www.selleckchem.com/products/Pelitinib.html>
- [127] Gefitinib (ZD1839) |  $\geq 99\%$ (HPLC) | Selleck | EGFR inhibitor [Internet]. selleckchem.com. [cité 11 sept 2021]. Disponible sur: <https://www.selleckchem.com/products/Gefitinib.html>
- [128] Vandetanib (ZD6474) |  $\geq 99\%$ (HPLC) | Selleck | VEGFR inhibitor [Internet]. selleckchem.com. [cité 11 sept 2021]. Disponible sur: <https://www.selleckchem.com/products/Vandetanib.html>
- [129] Nagasaka M, Zhu VW, Lim SM, Greco M, Wu F, Ou S-HI. Beyond Osimertinib: The Development of Third-Generation EGFR Tyrosine Kinase Inhibitors For Advanced EGFR+ NSCLC. *J Thorac Oncol*. 1 mai 2021;16(5):740-63.
- [130] Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL database in 2017. *Nucleic Acids Res*. 4 janv 2017;45(D1):D945-54.



- [131] Afatinib (BIBW2992) |  $\geq 99\%$ (HPLC) | Selleck | EGFR inhibitor [Internet].  
selleckchem.com. [cité 11 sept 2021]. Disponible sur:  
<https://www.selleckchem.com/products/BIBW2992.html>
- [132] Lapatinib (GW-572016) |  $\geq 99\%$ (HPLC) | Selleck | EGFR inhibitor [Internet].  
selleckchem.com. [cité 11 sept 2021]. Disponible sur:  
<https://www.selleckchem.com/products/lapatinib.html>
- [133] Brigatinib (AP26113) |  $\geq 99\%$ (HPLC) | Selleck | ALK inhibitor [Internet].  
selleckchem.com. [cité 11 sept 2021]. Disponible sur:  
<https://www.selleckchem.com/products/brigatinib-ap26113.html>
- [134] Neratinib (HKI-272) |  $\geq 99\%$ (HPLC) | Selleck | HER2 inhibitor [Internet].  
selleckchem.com. [cité 11 sept 2021]. Disponible sur:  
[https://www.selleckchem.com/products/Neratinib\(HKI-272\).html](https://www.selleckchem.com/products/Neratinib(HKI-272).html)
- [135] Erlotinib (OSI-774) |  $\geq 99\%$ (HPLC) | Selleck | EGFR inhibitor [Internet].  
selleckchem.com. [cité 11 sept 2021]. Disponible sur:  
<https://www.selleckchem.com/products/erlotinib.html>
- [136] Almonertinib (HS-10296) | EGFR-sensitizing/T790M Inhibitor | MedChemExpress  
[Internet]. MedchemExpress.com. [cité 11 sept 2021]. Disponible sur:  
<https://www.medchemexpress.com/hs-10296.html>
- [137] Osimertinib (AZD9291) |  $\geq 99\%$ (HPLC) | Selleck | EGFR inhibitor [Internet].  
selleckchem.com. [cité 11 sept 2021]. Disponible sur:  
<https://www.selleckchem.com/products/azd9291.html>
- [138] mavelertinib | Graphiques d'activité des ligands | Guide IUPHAR/BPS de  
PHARMACOLOGIE [Internet]. [cité 11 sept 2021]. Disponible sur:  
[https://www.guidetomalariapharmacology.org/GRAC/LigandActivityRangeVisForward  
?ligandId=9765](https://www.guidetomalariapharmacology.org/GRAC/LigandActivityRangeVisForward?ligandId=9765)
- [139] Mobocertinib (TAK-788) | EGFR/HER2 Mutations Inhibitor | MedChemExpress  
[Internet]. MedchemExpress.com. [cité 11 sept 2021]. Disponible sur:  
<https://www.medchemexpress.com/mobocertinib.html>

- [140] <https://www.rcsb.org/structure/5UG9>.
- [141] <https://www.rcsb.org/structure/5HG8>.
- [142] <https://www.rcsb.org/structure/6TG0>.
- [143] <https://www.rcsb.org/structure/6TFY>.
- [144] <https://www.rcsb.org/structure/6TFV>.
- [145] <https://www.rcsb.org/structure/5HG5>.
- [146] <https://www.rcsb.org/structure/6TG1>.
- [147] <https://www.rcsb.org/structure/6V66>.
- [148] <https://www.rcsb.org/structure/6TFZ>.
- [149] <https://www.rcsb.org/structure/5GNK>.
- [150] <https://www.rcsb.org/structure/4I24>.
- [151] <https://www.rcsb.org/structure/5X2A>.
- [152] <https://www.rcsb.org/structure/5HG7>.
- [153] <https://www.rcsb.org/structure/4WKQ>.
- [154] <https://www.rcsb.org/structure/7A2A>.
- [155] <https://www.rcsb.org/structure/6P8Q>.
- [156] <https://www.rcsb.org/structure/3W2S>.
- [157] <https://www.rcsb.org/structure/5YU9>.
- [158] Barker AJ, Gibson KH, Grundy W, Godfrey AA, Barlow JJ, Healy MP, et al. Studies leading to the identification of ZD1839 (IRESSA): an orally active, selective epidermal growth factor receptor tyrosine kinase inhibitor targeted to the treatment of cancer. *Bioorg Med Chem Lett*. 2001;11(14):1911-4.

- [159] Yun C-H, Boggon TJ, Li Y, Woo MS, Greulich H, Meyerson M, et al. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell*. 2007;11(3):217-27.
- [160] Wu P, Nielsen TE, Clausen MH. FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol Sci*. 2015;36(7):422-39.

## Annexes :

**Annexe N° I :** Les scores des similaires obtenus suite au docking avec l'EGFR (4WKQ) :

	Ligand	Binding Affinity
1	protein_Conformer3D_CID_146844011	-9,8
2	protein_Conformer3D_CID_142675009	-9,6
3	protein_Conformer3D_CID_137455797	-9,6
4	protein_Conformer3D_CID_137455767	-9,6
5	protein_Conformer3D_CID_24875330	-9,5
6	protein_Conformer3D_CID_145577260	-9,5
7	protein_Conformer3D_CID_118343039	-9,5
8	protein_Conformer3D_CID_145577297	-9,4
9	protein_Conformer3D_CID_71617554	-9,3
10	protein_Conformer3D_CID_71572791	-9,3
11	protein_Conformer3D_CID_151031088	-9,3
12	protein_Conformer3D_CID_10453762	-9,3
13	protein_Conformer3D_CID_71616634	-9,2
14	protein_Conformer3D_CID_70880600	-9,2
15	protein_Conformer3D_CID_70870428	-9,2
16	protein_Conformer3D_CID_59557983	-9,2
17	protein_Conformer3D_CID_57802650	-9,2
18	protein_Conformer3D_CID_44416719	-9,2
19	protein_Conformer3D_CID_20819742	-9,2
20	protein_Conformer3D_CID_145577303	-9,2
21	protein_Conformer3D_CID_10173248	-9,2
22	protein_Conformer3D_CID_10070557	-9,2
23	protein_Conformer3D_CID_10002824	-9,2
24	protein_Conformer3D_CID_9980933	-9,1
25	protein_Conformer3D_CID_71616745	-9,1
26	protein_Conformer3D_CID_71616744	-9,1
27	protein_Conformer3D_CID_71572790	-9,1
28	protein_Conformer3D_CID_60036694	-9,1

29	protein_Conformer3D_CID_57802832	-9,1
30	protein_Conformer3D_CID_57802680	-9,1
31	protein_Conformer3D_CID_57489414	-9,1
32	protein_Conformer3D_CID_44416461	-9,1
33	protein_Conformer3D_CID_150196854	-9,1
34	protein_Conformer3D_CID_129391266	-9,1
35	protein_Conformer3D_CID_11742676	-9,1
36	protein_Conformer3D_CID_9980764	-9
37	protein_Conformer3D_CID_90294712	-9
38	protein_Conformer3D_CID_89664884	-9
39	protein_Conformer3D_CID_72544698	-9
40	protein_Conformer3D_CID_58208325	-9
41	protein_Conformer3D_CID_57802897	-9
42	protein_Conformer3D_CID_57489413	-9
43	protein_Conformer3D_CID_44416673	-9
44	protein_Conformer3D_CID_25253125	-9
45	protein_Conformer3D_CID_20819748	-9
46	protein_Conformer3D_CID_20819741	-9
47	protein_Conformer3D_CID_20819731	-9
48	protein_Conformer3D_CID_151757353	-9
49	protein_Conformer3D_CID_143966615	-9
50	protein_Conformer3D_CID_141307494	-9
51	protein_Conformer3D_CID_137455760	-9
52	protein_Conformer3D_CID_137455756	-9
53	protein_Conformer3D_CID_137455753	-9
54	protein_Conformer3D_CID_137455750	-9
55	protein_Conformer3D_CID_12094954	-9
56	protein_Conformer3D_CID_10408840	-9
57	protein_Conformer3D_CID_10343270	-9
58	protein_Conformer3D_CID_10216188	-9
59	protein_Structure2D_CID_155054566	-8,9

60	protein_Conformer3D_CID_68919703	-8,9
61	protein_Conformer3D_CID_68908614	-8,9
62	protein_Conformer3D_CID_67959110	-8,9
63	protein_Conformer3D_CID_57489412	-8,9
64	protein_Conformer3D_CID_21925181	-8,9
65	protein_Conformer3D_CID_19077508	-8,9
66	protein_Conformer3D_CID_154677763	-8,9
67	protein_Conformer3D_CID_148722116	-8,9
68	protein_Conformer3D_CID_145577403	-8,9
69	protein_Conformer3D_CID_142675010	-8,9
70	protein_Conformer3D_CID_142674991	-8,9
71	protein_Conformer3D_CID_142674981	-8,9
72	protein_Conformer3D_CID_139270404	-8,9
73	protein_Conformer3D_CID_123603727	-8,9
74	protein_Conformer3D_CID_122484763	-8,9
75	protein_Conformer3D_CID_11429574	-8,9
76	protein_Conformer3D_CID_91130311	-8,8
77	protein_Conformer3D_CID_90873973	-8,8
78	protein_Conformer3D_CID_72545386	-8,8
79	protein_Conformer3D_CID_72545385	-8,8
80	protein_Conformer3D_CID_72544697	-8,8
81	protein_Conformer3D_CID_70853646	-8,8
82	protein_Conformer3D_CID_60023716	-8,8
83	protein_Conformer3D_CID_57802743	-8,8
84	protein_Conformer3D_CID_57802740	-8,8
85	protein_Conformer3D_CID_25253122	-8,8
86	protein_Conformer3D_CID_25252965	-8,8
87	protein_Conformer3D_CID_21925262	-8,8
88	protein_Conformer3D_CID_142778087	-8,8
89	protein_Conformer3D_CID_142675023	-8,8
90	protein_Conformer3D_CID_142675013	-8,8

91	protein_Conformer3D_CID_142674972	-8,8
92	protein_Conformer3D_CID_142674965	-8,8
93	protein_Conformer3D_CID_129391320	-8,8
94	protein_Conformer3D_CID_123596536	-8,8
95	protein_Conformer3D_CID_11476336	-8,8
96	protein_Conformer3D_CID_11443766	-8,8
97	protein_Conformer3D_CID_11430359	-8,8
98	protein_Conformer3D_CID_10391639	-8,8
99	protein_Conformer3D_CID_101772339	-8,8
100	protein_Conformer3D_CID_10127173	-8,8
101	protein_Conformer3D_CID_89664885	-8,7
102	protein_Conformer3D_CID_72945803	-8,7
103	protein_Conformer3D_CID_72546351	-8,7
104	protein_Conformer3D_CID_72545144	-8,7
105	protein_Conformer3D_CID_72544917	-8,7
106	protein_Conformer3D_CID_59779526	-8,7
107	protein_Conformer3D_CID_59772910	-8,7
108	protein_Conformer3D_CID_57802587	-8,7
109	protein_Conformer3D_CID_57794533	-8,7
110	protein_Conformer3D_CID_56656058	-8,7
111	protein_Conformer3D_CID_44416595	-8,7
112	protein_Conformer3D_CID_25252809	-8,7
113	protein_Conformer3D_CID_22624991	-8,7
114	protein_Conformer3D_CID_22478398	-8,7
115	protein_Conformer3D_CID_21925242	-8,7
116	protein_Conformer3D_CID_21925173	-8,7
117	protein_Conformer3D_CID_19077514	-8,7
118	protein_Conformer3D_CID_19077507	-8,7
119	protein_Conformer3D_CID_19077486	-8,7
120	protein_Conformer3D_CID_154688616	-8,7
121	protein_Conformer3D_CID_148972513	-8,7

122	protein_Conformer3D_CID_142675051	-8,7
123	protein_Conformer3D_CID_142675007	-8,7
124	protein_Conformer3D_CID_141515866	-8,7
125	protein_Conformer3D_CID_123811683	-8,7
126	protein_Conformer3D_CID_123660423	-8,7
127	protein_Conformer3D_CID_12094953	-8,7
128	protein_Conformer3D_CID_118710441	-8,7
129	protein_Conformer3D_CID_11676330	-8,7
130	protein_Conformer3D_CID_11524799	-8,7
131	protein_Conformer3D_CID_101772335	-8,7
132	protein_Conformer3D_CID_9981727	-8,6
133	protein_Conformer3D_CID_90713942	-8,6
134	protein_Conformer3D_CID_88965602	-8,6
135	protein_Conformer3D_CID_72545884	-8,6
136	protein_Conformer3D_CID_72545147	-8,6
137	protein_Conformer3D_CID_72545146	-8,6
138	protein_Conformer3D_CID_72545145	-8,6
139	protein_Conformer3D_CID_72544474	-8,6
140	protein_Conformer3D_CID_72544473	-8,6
141	protein_Conformer3D_CID_66954893	-8,6
142	protein_Conformer3D_CID_59779527	-8,6
143	protein_Conformer3D_CID_59772884	-8,6
144	protein_Conformer3D_CID_57794528	-8,6
145	protein_Conformer3D_CID_57721234	-8,6
146	protein_Conformer3D_CID_57572804	-8,6
147	protein_Conformer3D_CID_54586507	-8,6
148	protein_Conformer3D_CID_46244894	-8,6
149	protein_Conformer3D_CID_46244893	-8,6
150	protein_Conformer3D_CID_46244807	-8,6
151	protein_Conformer3D_CID_46244721	-8,6
152	protein_Conformer3D_CID_46244633	-8,6



153	protein_Conformer3D_CID_46244365	-8,6
154	protein_Conformer3D_CID_46244271	-8,6
155	protein_Conformer3D_CID_46244101	-8,6
156	protein_Conformer3D_CID_46244017	-8,6
157	protein_Conformer3D_CID_46243935	-8,6
158	protein_Conformer3D_CID_46243933	-8,6
159	protein_Conformer3D_CID_46243783	-8,6
160	protein_Conformer3D_CID_44416610	-8,6
161	protein_Conformer3D_CID_44158623	-8,6
162	protein_Conformer3D_CID_25253428	-8,6
163	protein_Conformer3D_CID_22624941	-8,6
164	protein_Conformer3D_CID_21925199	-8,6
165	protein_Conformer3D_CID_21925183	-8,6
166	protein_Conformer3D_CID_21081995	-8,6
167	protein_Conformer3D_CID_20819750	-8,6
168	protein_Conformer3D_CID_18753410	-8,6
169	protein_Conformer3D_CID_153160696	-8,6
170	protein_Conformer3D_CID_151594033	-8,6
171	protein_Conformer3D_CID_147048046	-8,6
172	protein_Conformer3D_CID_143021800	-8,6
173	protein_Conformer3D_CID_142962285	-8,6
174	protein_Conformer3D_CID_142677274	-8,6
175	protein_Conformer3D_CID_142674970	-8,6
176	protein_Conformer3D_CID_141620760	-8,6
177	protein_Conformer3D_CID_140435224	-8,6
178	protein_Conformer3D_CID_138964152	-8,6
179	protein_Conformer3D_CID_137553842	-8,6
180	protein_Conformer3D_CID_118710446	-8,6
181	protein_Conformer3D_CID_11719266	-8,6
182	protein_Conformer3D_CID_11604752	-8,6
183	protein_Conformer3D_CID_10388898	-8,6

184	protein_Conformer3D_CID_101772342	-8,6
185	protein_Conformer3D_CID_91020854	-8,5
186	protein_Conformer3D_CID_72545633	-8,5
187	protein_Conformer3D_CID_72544915	-8,5
188	protein_Conformer3D_CID_72544475	-8,5
189	protein_Conformer3D_CID_71616849	-8,5
190	protein_Conformer3D_CID_71616746	-8,5
191	protein_Conformer3D_CID_68904834	-8,5
192	protein_Conformer3D_CID_59557982	-8,5
193	protein_Conformer3D_CID_57777497	-8,5
194	protein_Conformer3D_CID_54091225	-8,5
195	protein_Conformer3D_CID_51352660	-8,5
196	protein_Conformer3D_CID_46244895	-8,5
197	protein_Conformer3D_CID_46244722	-8,5
198	protein_Conformer3D_CID_46244720	-8,5
199	protein_Conformer3D_CID_46244719	-8,5
200	protein_Conformer3D_CID_46244635	-8,5
201	protein_Conformer3D_CID_46244634	-8,5
202	protein_Conformer3D_CID_46244550	-8,5
203	protein_Conformer3D_CID_46244549	-8,5
204	protein_Conformer3D_CID_46244548	-8,5
205	protein_Conformer3D_CID_46244467	-8,5
206	protein_Conformer3D_CID_46244368	-8,5
207	protein_Conformer3D_CID_46244366	-8,5
208	protein_Conformer3D_CID_46244273	-8,5
209	protein_Conformer3D_CID_46244187	-8,5
210	protein_Conformer3D_CID_46244186	-8,5
211	protein_Conformer3D_CID_46244185	-8,5
212	protein_Conformer3D_CID_46244184	-8,5
213	protein_Conformer3D_CID_46244102	-8,5
214	protein_Conformer3D_CID_46244100	-8,5

215	protein_Conformer3D_CID_46244019	-8,5
216	protein_Conformer3D_CID_46243934	-8,5
217	protein_Conformer3D_CID_46243932	-8,5
218	protein_Conformer3D_CID_46243784	-8,5
219	protein_Conformer3D_CID_46243753	-8,5
220	protein_Conformer3D_CID_46243752	-8,5
221	protein_Conformer3D_CID_44416407	-8,5
222	protein_Conformer3D_CID_44159309	-8,5
223	protein_Conformer3D_CID_44158626	-8,5
224	protein_Conformer3D_CID_44158624	-8,5
225	protein_Conformer3D_CID_22478193	-8,5
226	protein_Conformer3D_CID_19077520	-8,5
227	protein_Conformer3D_CID_18753452	-8,5
228	protein_Conformer3D_CID_16091478	-8,5
229	protein_Conformer3D_CID_152774002	-8,5
230	protein_Conformer3D_CID_152321503	-8,5
231	protein_Conformer3D_CID_143966622	-8,5
232	protein_Conformer3D_CID_142811676	-8,5
233	protein_Conformer3D_CID_142674994	-8,5
234	protein_Conformer3D_CID_137553840	-8,5
235	protein_Conformer3D_CID_136655855	-8,5
236	protein_Conformer3D_CID_125323009	-8,5
237	protein_Conformer3D_CID_123134766	-8,5
238	protein_Conformer3D_CID_10159951	-8,5
239	protein_Conformer3D_CID_72545636	-8,4
240	protein_Conformer3D_CID_72545634	-8,4
241	protein_Conformer3D_CID_72545384	-8,4
242	protein_Conformer3D_CID_72545143	-8,4
243	protein_Conformer3D_CID_72544919	-8,4
244	protein_Conformer3D_CID_72544916	-8,4
245	protein_Conformer3D_CID_69933632	-8,4

246	protein_Conformer3D_CID_59772904	-8,4
247	protein_Conformer3D_CID_59772857	-8,4
248	protein_Conformer3D_CID_51352594	-8,4
249	protein_Conformer3D_CID_49849312	-8,4
250	protein_Conformer3D_CID_49849311	-8,4
251	protein_Conformer3D_CID_4913175	-8,4
252	protein_Conformer3D_CID_46244981	-8,4
253	protein_Conformer3D_CID_46244806	-8,4
254	protein_Conformer3D_CID_46244805	-8,4
255	protein_Conformer3D_CID_46244636	-8,4
256	protein_Conformer3D_CID_46244551	-8,4
257	protein_Conformer3D_CID_46244466	-8,4
258	protein_Conformer3D_CID_46244465	-8,4
259	protein_Conformer3D_CID_46244464	-8,4
260	protein_Conformer3D_CID_46244367	-8,4
261	protein_Conformer3D_CID_46244272	-8,4
262	protein_Conformer3D_CID_46244270	-8,4
263	protein_Conformer3D_CID_46244099	-8,4
264	protein_Conformer3D_CID_46244018	-8,4
265	protein_Conformer3D_CID_46244016	-8,4
266	protein_Conformer3D_CID_46243782	-8,4
267	protein_Conformer3D_CID_46243767	-8,4
268	protein_Conformer3D_CID_46243766	-8,4
269	protein_Conformer3D_CID_46243765	-8,4
270	protein_Conformer3D_CID_46243764	-8,4
271	protein_Conformer3D_CID_46243754	-8,4
272	protein_Conformer3D_CID_44629454	-8,4
273	protein_Conformer3D_CID_44159198	-8,4
274	protein_Conformer3D_CID_44159197	-8,4
275	protein_Conformer3D_CID_147265553	-8,4
276	protein_Conformer3D_CID_146786393	-8,4

277	protein_Conformer3D_CID_142811067	-8,4
278	protein_Conformer3D_CID_142675015	-8,4
279	protein_Conformer3D_CID_142637805	-8,4
280	protein_Conformer3D_CID_123631	-8,4
281	protein_Conformer3D_CID_11167092	-8,4
282	protein_Conformer3D_CID_91600019	-8,3
283	protein_Conformer3D_CID_69532515	-8,3
284	protein_Conformer3D_CID_68906695	-8,3
285	protein_Conformer3D_CID_46244808	-8,3
286	protein_Conformer3D_CID_46243936	-8,3
287	protein_Conformer3D_CID_44300050	-8,3
288	protein_Conformer3D_CID_44159308	-8,3
289	protein_Conformer3D_CID_44158625	-8,3
290	protein_Conformer3D_CID_20819745	-8,3
291	protein_Conformer3D_CID_152664589	-8,3
292	protein_Conformer3D_CID_148405083	-8,3
293	protein_Conformer3D_CID_145577395	-8,3
294	protein_Conformer3D_CID_141322940	-8,3
295	protein_ligand	-8,2
296	protein_Conformer3D_CID_59825276	-8,2
297	protein_Conformer3D_CID_59772861	-8,2
298	protein_Conformer3D_CID_5329020	-8,2
299	protein_Conformer3D_CID_23228252	-8,2
300	protein_Conformer3D_CID_20819740	-8,2
301	protein_Conformer3D_CID_18445190	-8,2
302	protein_Conformer3D_CID_141515877	-8,2
303	protein_Conformer3D_CID_122485650	-8,2
304	protein_Conformer3D_CID_118884417	-8,2
305	protein_Conformer3D_CID_10411719	-8,2
306	protein_Conformer3D_CID_89773039	-8,1
307	protein_Conformer3D_CID_89330073	-8,1

308	protein_Conformer3D_CID_5329022	-8,1
309	protein_Conformer3D_CID_5329021	-8,1
310	protein_Conformer3D_CID_5329019	-8,1
311	protein_Conformer3D_CID_44300029	-8,1
312	protein_Conformer3D_CID_44299956	-8,1
313	protein_Conformer3D_CID_22478296	-8,1
314	protein_Conformer3D_CID_72545632	-8
315	protein_Conformer3D_CID_68695162	-8
316	protein_Conformer3D_CID_57794562	-8
317	protein_Conformer3D_CID_19077501	-8
318	protein_Conformer3D_CID_19077498	-8
319	protein_Conformer3D_CID_151908485	-8
320	protein_Conformer3D_CID_148482023	-8
321	protein_Conformer3D_CID_146751853	-8
322	protein_Conformer3D_CID_142677250	-8
323	protein_Conformer3D_CID_141320482	-8
324	protein_Structure2D_CID_155301878	-7,9
325	protein_Conformer3D_CID_59772864	-7,9
326	protein_Conformer3D_CID_59772830	-7,9
327	protein_Conformer3D_CID_23105628	-7,9
328	protein_Conformer3D_CID_20819703	-7,9
329	protein_Conformer3D_CID_18445189	-7,9
330	protein_Conformer3D_CID_150131053	-7,9
331	protein_Conformer3D_CID_142642690	-7,9
332	protein_Conformer3D_CID_126709552	-7,9
333	protein_Conformer3D_CID_117740745	-7,9
334	protein_Conformer3D_CID_71579401	-7,8
335	protein_Conformer3D_CID_44300005	-7,8
336	protein_Conformer3D_CID_23228249	-7,8
337	protein_Conformer3D_CID_19077513	-7,8
338	protein_Conformer3D_CID_19077488	-7,8

339	protein_Conformer3D_CID_137553843	-7,8
340	protein_Conformer3D_CID_118884440	-7,8
341	protein_Conformer3D_CID_118884422	-7,8
342	protein_Conformer3D_CID_18753446	-7,7
343	protein_Conformer3D_CID_149845388	-7,7
344	protein_Conformer3D_CID_132140632	-7,7
345	protein_Conformer3D_CID_126516980	-7,7
346	protein_Conformer3D_CID_56954235	-7,6
347	protein_Structure2D_CID_151707829	-7,5
348	protein_Conformer3D_CID_70937892	-7,5
349	protein_Conformer3D_CID_44299983	-7,5
350	protein_Conformer3D_CID_19077504	-7,5
351	protein_Conformer3D_CID_126489053	-7,5
352	protein_Conformer3D_CID_143733544	-7,3
353	protein_Conformer3D_CID_150917137	-7,2
354	protein_Structure2D_CID_155092292	-4,6

**RESUME :**

Les voies de signalisation cellulaire impliquant les enzymes tyrosine kinase dont l'EGFR représentent une cible thérapeutique innovante pour le traitement du cancer. La conception in silico des médicaments est une nouvelle approche impliquant l'utilisation des techniques de modélisation moléculaire telles que le criblage virtuel et le Docking moléculaire.

L'objectif de ce travail est de rechercher des nouvelles molécules inhibitrices de l'EGFR par un criblage virtuel basé sur le ligand.

Le programme de docking utilisé est l'AutoDock vina intégré dans Pyrex dont on a vérifié préalablement la fiabilité. Le géfitinib est sélectionné comme ligand de référence et le complexe ligand-récepteur choisi de la PDB est le 4WKQ.

Une série de 354 molécules issues de PubChem avec une similarité de 97% au géfitinib ont été sujettes au docking moléculaire envers le site actif de l'EGFR. Les 12 meilleures molécules dont l'énergie de liaison est comprise entre -9,3 et -9,8 Kcal/Mole (géfitinib : -8,2 Kcal/Mole) ont été prises pour l'étude des interactions avec le récepteur en comparaison avec celles du géfitinib de référence. On a trouvé que 9 molécules ont eu des interactions intéressantes. Par la suite, l'application des règles de Lipinski et Veber, l'analyse des paramètres physicochimique et pharmacocinétiques ainsi que les tests de toxicité potentielle, nous ont conduit aux molécules n°4, n°5 et n°6 (CID\_137455767, CID\_24875330 et CID\_145577260 respectivement).

En conclusion, cette étude nous permet de proposer les molécules CID\_137455767, CID\_24875330 et CID\_145577260 comme des nouveaux candidats inhibiteurs de l'EGFR, et que ces molécules méritent de passer aux phases ultérieures du développement des médicaments.

**Mots clés :** Cancer, EGFR, Docking moléculaire, AutoDock vina, Energie de liaison.

**ABSTRACT :**

Cell signaling pathways involving tyrosine kinase enzymes including EGFR represent an innovative therapeutic target for the treatment of cancer. In silico drug design is a new approach that uses molecular modeling techniques such as virtual screening and molecular docking.

The objective of this work is to search for new inhibitory molecules of EGFR by a ligand based virtual screening.

The docking program used is AutoDock vina integrated into Pyrex, which has been previously checked for reliability. Gefitinib is selected as the reference ligand and the chosen ligand-receptor complex for PDB is 4WKQ.

A series of 354 molecules from PubChem with 97% similarity to gefitinib were subjected to molecular docking into the active site of EGFR. The 12 best molecules whose binding energy is between -9.3 and -9.8 Kcal / Mole (gefitinib: -8.2 Kcal / Mole) were taken for the study of interactions with the receptor in comparison with those of the reference gefitinib. It was found that 9 molecules had interesting interactions. Subsequently, the application of the rules of Lipinski and Veber, the analysis of the physicochemical and pharmacokinetic parameters as well as the tests of potential toxicity, led us to molecules n ° 4, n ° 5 and n ° 6 (CID\_137455767, CID\_24875330 and CID\_145577260 respectively).

In conclusion, this study allows us to propose the molecules CID\_137455767, CID\_24875330 and CID\_145577260 as new EGFR inhibitor candidates, and that these molecules deserve to move on to the next phases of drug development.

**Keywords:** Cancer, EGFR, Molecular docking, AutoDock vina, Binding energy.