

## 1. Introduction

Comment comprendre l'intérêt thérapeutique des molécules ? Comment trouver les molécules actives parmi des milliers de molécules découvertes chaque année dans l'ensemble des laboratoires ? Pour tous les chercheurs qui mettent au point de nouvelles molécules qui peuvent être biologiquement actives ou des nouveaux médicaments ou cherchent des outils pour comprendre des fonctions complexes de ces interactions dans un organisme, ces questions sont essentielles. Des technologies nouvelles ont permis d'accélérer les recherches dans ce domaine. Parmi elles, le recours aux chimiothèques et au criblage moléculaire.

Tout criblage virtuel se décompose en trois étapes d'égale importance:

- (1) la mise au point de la chimiothèque de départ,
- (2) le criblage proprement dit,
- (3) la sélection d'une liste de touches virtuelles.

Il est à noter que toute erreur à chacune de ces trois étapes aura des conséquences importantes se traduisant généralement par une augmentation du taux de faute. Il convient donc d'être très attentif à chacune d'entre elles.

Nous avons donc décidé d'effectuer une étude qui consiste à prédire à la fois la conformation et l'orientation relative de chaque dérivé de diamides par rapport à la cible d'intérêt. Cette étude a été effectuée selon les étapes suivantes : la préparation de la chimiothèque, une étude de similarité et le docking moléculaire.

## Chimiothèque de ligands

Deux types de chimiothèques peuvent être utilisés à des fins de criblage virtuel : des collections de criblages (molécules disponibles) et des collections virtuelles (molécules à synthétiser).

Pour avoir un aperçu sur le premier type de chimiothèque, un chimiste peut se référer à une liste assez exhaustive décrite sur le site : (<http://www.warr.com/links.html#chemlib>), Il est également invité à consulter les documentations relatives à chacune de ces chimiothèques pour un descriptif de leur valorisation et de leur originalité. Nous proposons ci-dessous des liens vers divers, outils de manipulation des chimiothèques (**Figure II.1**)

## CHAPITRE II. CRIBLAGE VIRTUEL DE DIAMIDES

Nom	Editeur	Site Internet	Fonction
ISIS/base	MDL	<a href="http://www.mdli.com">http://www.mdli.com</a>	archivage
ChemOffice	Cambridge Soft.	<a href="http://www.cambridgesoft.com">http://www.cambridgesoft.com</a>	archivage
Filter	Openeyes	<a href="http://www.eyesopen.com">http://www.eyesopen.com</a>	filtrage
Cliff	Molecular Network	<a href="http://www.mol-net.de">http://www.mol-net.de</a>	filtrage
Pipeline Pilot	SciTegic	<a href="http://www.scitegic.com">http://www.scitegic.com</a>	filtrage et automatisation de procédures
Marvin	Chemaxon	<a href="http://www.chemaxon.com">http://www.chemaxon.com</a>	archivage
Ligprep	Schrodinger	<a href="http://www.schrodinger.com">http://www.schrodinger.com</a>	filtrage

**Figure II . 1:** Principaux outils de conception / gestion de chimiothèques électroniques.

Au vue de l'importance des composés azotés et plus particulièrement de la fonction amide et en raison des améliorations réalisées ces dernières années pour synthétiser des diamides <sup>[1-5]</sup>, ces composés sont devenus un outil largement utilisé en synthèse. En effet, ils offrent de multiples possibilités d'introduire des fonctions azotées ce qui constitue un challenge constant en synthèse puisque la quasi-totalité des produits biologiquement actifs en possède au moins une.

Il existe peu de méthodes directes permettant l'accès facile à ces diamides. Elles sont donc la plupart du temps synthétisées en plusieurs étapes, Dans ce contexte, nous avons envisagé de réaliser une réaction de Ugi 4-composants, elle permet l'accès à différents types de diamides en une seule étape. Pour cette raison, les réactions multicomposants sont des méthodes de choix pour la synthèse de produits biologiquement actifs.

Afin d'étendre l'intérêt de cette méthode, nous nous intéressons au profilage de base de données de molécules (deuxième type de chimiothèque) contre une cible ou une série de cible pour établir leur profile biologique. La conception d'une chimiothèque de seize dérivés de diamides a été effectuée selon la disponibilité des réactifs au niveau du laboratoire, dans le but de ne sélectionner que des composés présentant entre eux une certaine diversité chimique tout en contenant les groupements fonctionnels de motif de base, en vue d'extraire un sous-ensemble de la chimiothèque, pour la synthèse de quelques dérivés de diamides, dont aucun travail précédent n'a rapporté la synthèse de ces derniers.

Après avoir créé la librairie, Nous avons utilisé le ChemOffice pour dessiner Les structures de nos molécules mentionnées dans la figure (II.2) et les transformées en format 3D.

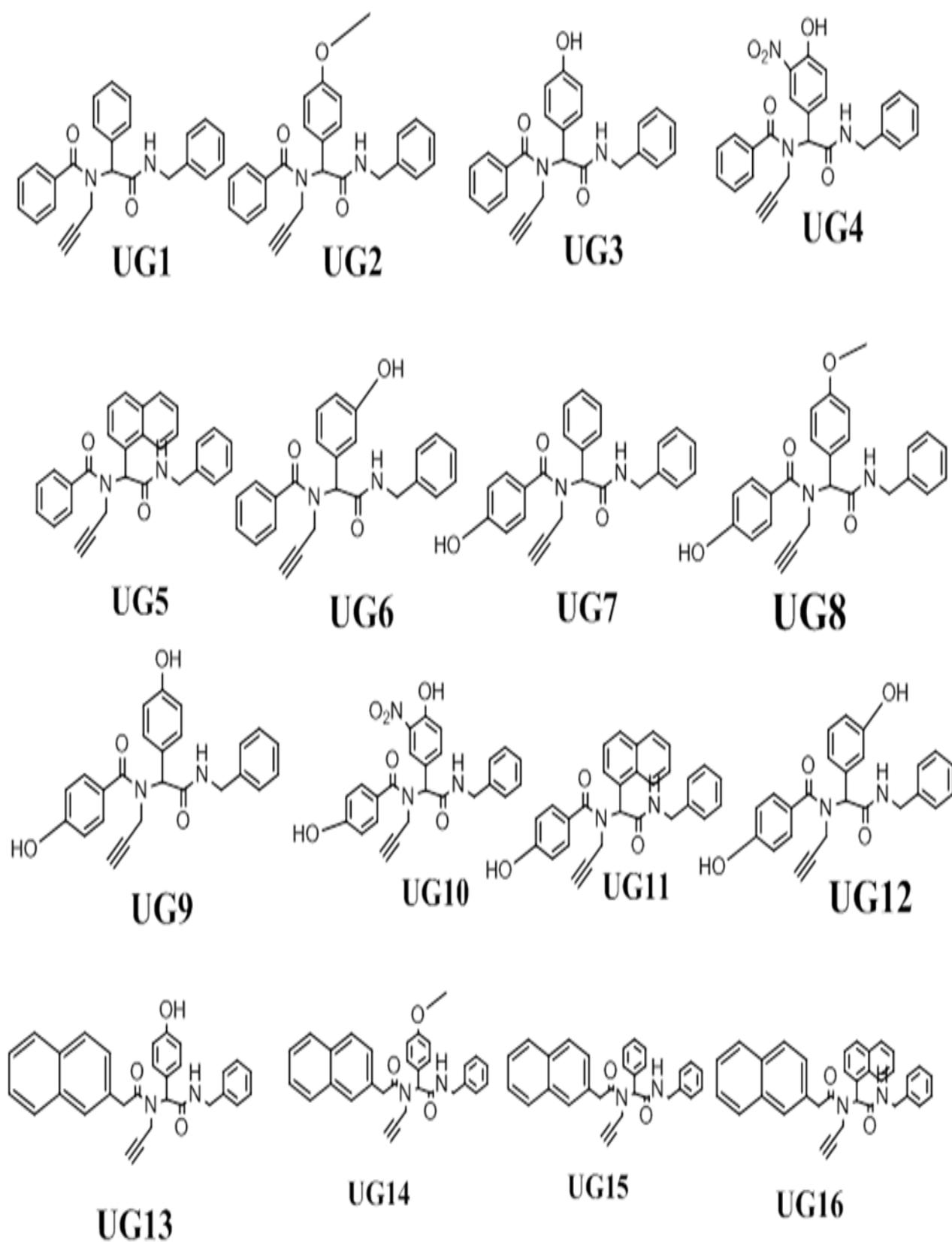


Figure II . 2: Chimiothèque des molécules cible.

### Une étude de similarité

Le pharmacophore est défini comme étant l'agencement tridimensionnel des caractéristiques permettant à une molécule d'avoir une activité biologique spécifique <sup>[6]</sup>. En d'autres termes, un pharmacophore est l'ensemble des points fonctionnels (donneurs/accepteurs de liaison hydrogène, hydrophobes...) d'une molécule selon la conformation spatiale existant pendant sa fixation dans le site de liaison de la cible.

Cette définition est à la base de son utilisation en criblage *in silico*, en s'appuyant sur le postulat selon lequel les molécules ayant le même pharmacophore auront une activité similaire sur la cible, pouvant permettre de s'affranchir du châssis moléculaire de la molécule servant de référence, ce qui n'est pas le cas, par exemple, des méthodes de similarité 2D<sup>[7]</sup>. Quand plusieurs molécules sont connues comme étant actives sur la cible d'intérêt, les points fonctionnels communs à l'ensemble (ou à la majorité) d'entre elles sont retenus pour constituer le pharmacophore sur lequel seront criblées les molécules de test<sup>[8]</sup>.

La construction d'un pharmacophore nécessite beaucoup d'informations à la fois sur la structure des ligands de référence et sur ces activités, afin de déterminer quelles sont les parties impliquées dans cette activité et qui serviront donc pour identifier de nouvelles molécules actives<sup>[9]</sup>.

### 3.1. Similarité

Ces méthodes s'intéressent directement à la structure des molécules de référence et des composés à cribler. Comme pour les pharmacophores, l'idée sous-jacente est que les molécules qui ont une structure similaire vont vraisemblablement partager des propriétés similaires<sup>[10]</sup>.

Les méthodes de similarité 3D sont non seulement basées sur les structures 2D des molécules, mais aussi sur leurs conformations. Cependant, la conformation bioactive d'un ligand est, le plus souvent, connue seulement si la structure expérimentale de la protéine cible est connue également<sup>[11]</sup>.

Une ou plusieurs conformations tridimensionnelles de cette molécule peuvent être générées (avec Molegro, par exemple), servant d'hypothèses pour le criblage des molécules testées. Lorsque plusieurs ligands sont connus, une première recherche conformationnelle est faite afin d'avoir un alignement unique conduisant à un modèle prédictif d'une activité biologique<sup>[12]</sup>.

## CHAPITRE II. CRIBLAGE VIRTUEL DE DIAMIDES

---

La recherche de similarité 3D a deux composantes principales :

1. La similarité de conformation des deux molécules, qui peut être prise en compte en tentant de minimiser les différences entre les deux surfaces moléculaires ou de maximiser le recouvrement du volume des deux molécules,
2. La similarité des points pharmacophoriques qui peut se retrouver dans l'utilisation de « couleurs ».

Les méthodes de recherche de similarité précédentes s'appuient sur le principe selon lequel des molécules de structures similaires auront plus de chances d'avoir des propriétés similaires<sup>[10]</sup> et de se lier au même groupe de protéines<sup>[13]</sup>. Cependant, la similarité chimique seule ne peut pas expliquer l'activité d'un composé sur une voie de signalisation ou un groupe de voies de signalisations.

Kauvar et al<sup>[14]</sup> se sont basés sur les interactions observées entre un certain nombre de composés et plusieurs protéines pour construire des « empreintes d'affinité » et prédire ainsi les affinités de nouveaux composés. De la même manière, Fliri et al<sup>[15,16]</sup> ont prédit des interactions simultanées de molécules à l'aide d'empreintes.

Il existe trois principaux types de métriques de similarité : les coefficients de distance (distance de Hamming), d'association (Jaccard/Tanimoto, Simpson...) et de corrélation (Pearson, Stiles...).

Les coefficients d'association vont, quant à eux, quantifier le degré de similarité entre deux composés. Le coefficient de Tanimoto est le plus souvent utilisé dans les études cherchant à évaluer la similarité entre différents composés<sup>[17-19]</sup>.

$$S_{A,B} = \frac{c}{a + b - c}$$

L'équation représente la similarité entre deux molécules A et B, avec

*a* : le nombre d'empreintes dans le composé A.

*b* : le nombre d'empreintes dans le composé B.

*c* : le nombre d'empreintes qui apparaissent dans les deux composés.

## CHAPITRE II. CRIBLAGE VIRTUEL DE DIAMIDES

Dans cette partie, un premier processus de criblage virtuel basé sur le calcul de coefficient de Tanimoto a été effectué. Les structures de nos diamides ont été construites avec le programme de construction moléculaire ChemDraw, elles sont enregistrées sous format (.mol.2) et en suite sont convertis en format (.sdf).

Pour que nous pouvons calculer le coefficient de Tanimoto, ce calcul a été effectué en utilisant la banque de données TTD (Target Therapeutic database (<http://bidd.nus.edu.sg/group/ttd/ttd.asp>)). Un premier essai a été impliqué sur un noyau de base des molécules étudiées, les résultats sont présentés dans le tableau (II.1).

**Tableau II . 1** : La similarité de noyau de base des molécules étudiées avec des médicaments connus.

ID	Nom du médicament	Coefficient de Tanimoto	Taux de similarité
1 <u>DAP001010</u>	Erlotinib	0.997742	Très similaire
2 <u>DCL000642</u>	SSR125543	0.996697	Très similaire
3 <u>DAP000579</u>	Selegiline	0.996642	Très similaire
4 <u>DAP000606</u>	Ethinamate	0.996623	Très similaire
5 <u>DAP001107</u>	Rasagiline	0.995361	Très similaire
6 <u>DAP001361</u>	Clorgyline	0.995349	Très similaire
7 <u>DAP000580</u>	Pargyline	0.994783	Très similaire
8 <u>DAP000163</u>	Ethchlorvynol	0.986611	Très similaire

Cet essai montre que la molécule introduite à la base de données présente une similarité qui dépasse les 99% avec plusieurs médicaments dont la meilleure similarité a été remarquée avec Erlotinib avec un taux de 99.77%. La même procédure de calcul a été appliquée sur les autres molécules de notre chimiothèque, les résultats sont mentionnés dans le tableau.

## CHAPITRE II. CRIBLAGE VIRTUEL DE DIAMIDES

---

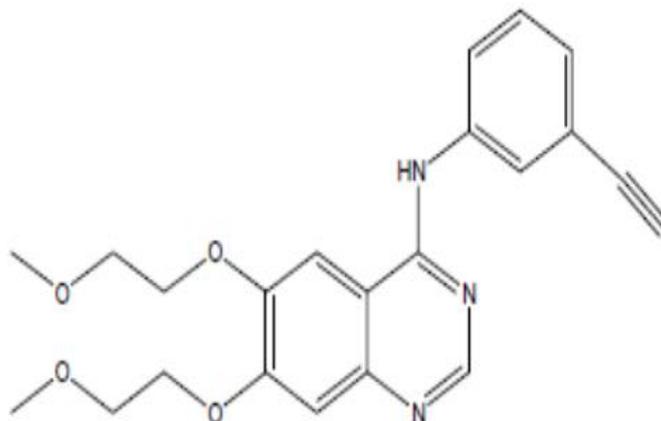
**Tableau II . 2:** Le coefficient de Tanimoto des molécules cible.

La molécule	Coefficient de Tanimoto
UG1	0.997742
UG2	0.997742
UG3	0.997742
UG4	0.997742
UG5	0.997742
UG6	0.997755
UG7	0.997761
UG8	0.997813
UG9	0.997476
UG10	0.996738
UG11	0.997198
UG12	0.997460
UG13	0.997330
UG14	0.997040
UG15	0.997151
UG16	0.994886

---

Le tableau montre que dans tous les cas, les valeurs de coefficient de Tanimoto sont supérieures à 99%. Ces résultats sont fortement appuyés par l'analyse de similarité basé sur le comportement de voisinage «Molécules structurellement similaires ont généralement des propriétés similaires, donc une activité biologique similaire ».

Donc nous pouvons déduire que les molécules de notre chimiothèque sont toutes similaires à l'Erlotinib, dont sa structure est indiquée dans la figure (II.3).



**Figure II . 3:** La structure de l'Erlotinib.

L'erlotinib (TARCEVA®) <sup>[20]</sup>, connu aussi sous le nom de CP-358-774 et OSI-774, l'Erlotinib provoque l'inhibition de EGFR (Epidermal Growth Factor Receptor). Ce résultat nous mène à une étude plus approfondie de l'interaction de nos molécules avec l'enzyme cible (EGFR).

L'EGFR est un récepteur à activité tyrosine kinase des facteurs de croissance de fibroblastes <sup>[21-22]</sup>, il existe sous forme homodimérique. La région de fixation du ligand contient deux régions riches en cystéine qui peuvent se lier avec différents ligands: l'EGF et sept autres ligands, appelés EGF-like: l'HB-EGF (Heparin-binding EGF), l'amphiréguline, le TGF- $\alpha$  (Transforming Growth Factor  $\alpha$ ), la betacelluline, l'épiréguline, l'épigène et le crypto.

La surexpression des EGFR est impliquée dans un certain nombre de cancers <sup>[23-30]</sup>. L'EGFR (HER-1) est impliqué, entre autre dans les cancers du sein, des ovaires, des poumons ainsi que dans des glioblastomes. L'*erbB2* (HER-2) est également impliqué dans des cancers du sein, des ovaires, des poumons, mais aussi dans des cancers de l'estomac, du colon et des voies intestinales. L'*erbB3* (HER-3) est impliqué dans les cancers du sein. Enfin, l'*erbB4* (HER-4) est impliqué dans les cancers du sein et de la granulosa.

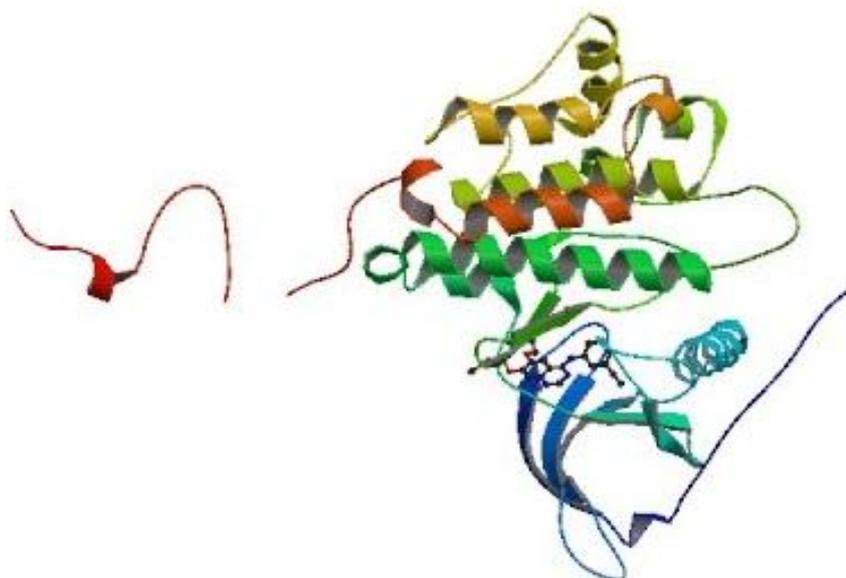
### 4. Docking

L'intérêt des analogues de diamides auxquels nous nous sommes intéressés n'ont pas encore montré une activité inhibitrice vis-à-vis de l'EGFR. Une approche a pu être envisagée grâce à outil *in silico* « docking moléculaire », pour étudier l'interaction de nos diamides tant que des inhibiteurs de l'EGFR. Cet enzyme a été sélectionné comme cible biologique pour réaliser notre criblage qui vise le développement de nouveaux dérivés d'inhibiteurs prometteurs <sup>[31-40]</sup>. d'EGFR en tant que substance médicamenteuses candidates.

### 4.1. Préparation de la protéine

Le fichier du récepteur utilisé, l'EGFR est téléchargé sous forme PDB via le banque de la protéine ( <http://www.pdb.org> ). *Protein Data Bank* ou PDB est une collection mondiale de données sur la structure tridimensionnelle (structure 3D) de macromolécules biologiques : protéines essentiellement, et acides nucléiques. Ces structures sont essentiellement déterminées par cristallographie aux rayons X ou par spectroscopie RMN.

Parmi plusieurs complexes proposés par la PDB, nous avons choisi le complexe avec le codes 1M17, car il s'agit des structures correctement définie avec une bonne résolution, contenant l'EGFR (Epidermal Growth Factor Receptor tyrosine kinase) et l'Erlotinib comme ligand, le complexe a été téléchargé au format pdb après l'introduction de son code. (Figure II.4)



**Figure II . 4:** Structure cristalline du complexe erlotinib/EGFR (PDB).

La préparation de récepteur pour l'amarrage s'effectue comme suit : les molécules d'eau, les cofacteurs, et les ions ont été exclus; les états de protonation (les ligands et la protéine) ont été placés au pH physiologique ; les ligands ont été placés pour être flexible ;et l'enzyme a été préparée comme corps rigide.

### 4.2. Préparation des ligands

Les structures des molécules retenues pour le docking ont été enregistrées sous format (.mol2) et en ensuite sont convertis en format (. Pdb). Ce type de préparation ne peut se faire sans une étape de minimisation importante de l'énergie potentielle, durant laquelle le

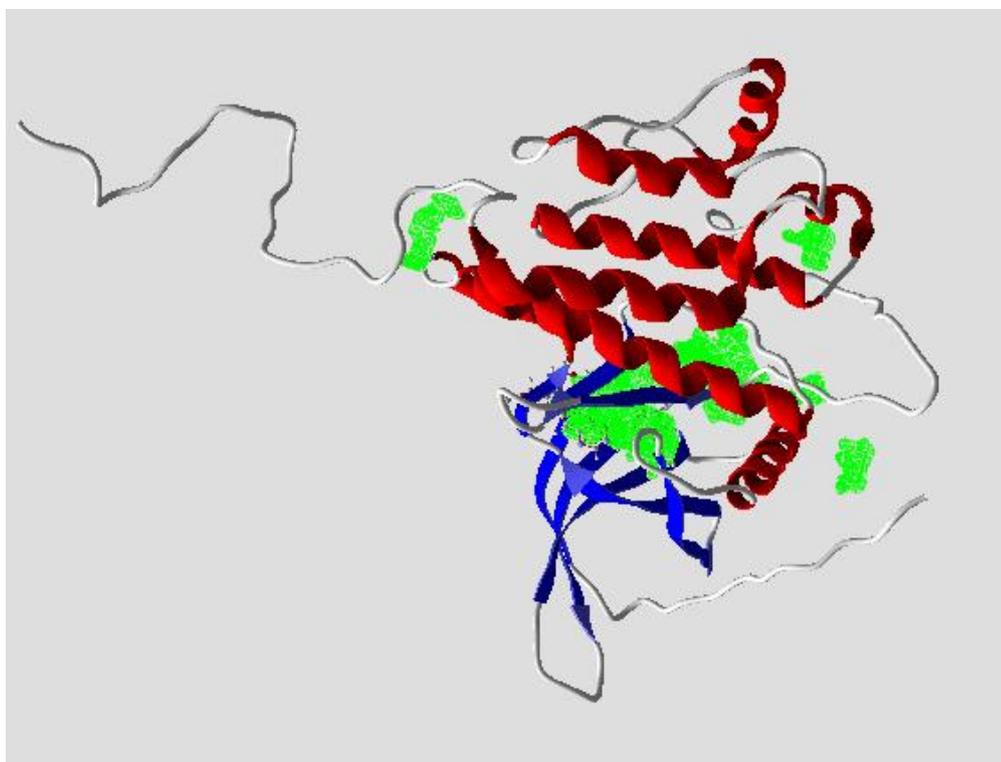
choix du champ de force est primordial. Pour nos diamides, l'optimisation a été réalisée pour l'obtention d'une conformation de basse énergie pour chaque ligand à l'aide du logiciel Hyprechem7 en utilisant la méthode semi-empirique PM3. Les calculs d'optimisation structurale ont été exécutés avec un PC portable Core i5.

### 4.3. Le processus du docking moléculaire

Toutes les simulations de docking effectuées lors de cette étude ont été assistées par Molegro Virtual Docker (MVD 2012.5.5.0, Molegro ApS) dont notre choix de logiciel est justifié par sa performance prouvée [41-43]. Pour une meilleure visualisation après les simulations du docking, les résultats obtenus ont été stockés et visualisés avec l'outil de visualisation moléculaire PyMol [44]. Les ligands choisis pour le docking seront importer dans l'espace de recherche, en utilisant la commande suivante : "Import Molecules".

#### 4.3.1. Détection des cavités

Les cavités présentes dans le récepteur sont identifiées en utilisant l'algorithme intégré pour la détection de cavité. La maille de la grille est fixée par défaut à 0.80Å. En utilisant cet algorithme, nous avons détecté cinq cavités. (Figure II.5)



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

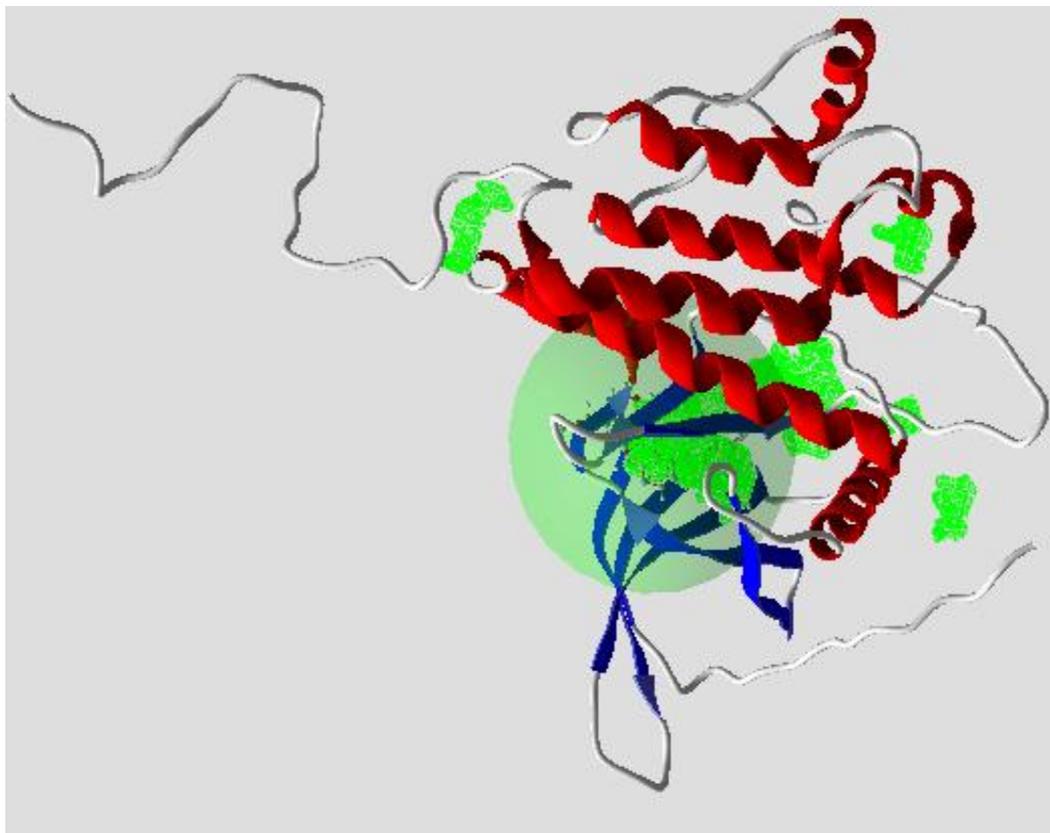
## CHAPITRE II. CRIBLAGE VIRTUEL DE DIAMIDES

Les cavités détectées sont classées selon leurs volumes (tableau II.3).

**Tableau II . 3:** les cavités détectées avec l'algorithme de détection de MVD pour 1M17

Cavité détectée	Volume (Å <sup>3</sup> )	Surface (Å <sup>2</sup> )
Cavité 1	275.97	885.76
Cavité 2	23.55	96
Cavité 3	15.87	70.4
Cavité 4	15.60	70
Cavité 5	14.84	60.16

Nous avons remarqué que le ligand de référence (Erlotinib) est co-cristallisé dans la cavité 1 et entouré par les résidus catalytiques. Nous avons donc décidé de conserver cette cavité comme un centre de recherche (research center) pour le processus du docking. Cette sphère de recherche couvre tous les résidus du site actif avec un rayon de 12 Å.



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

### 4.3.2. Protocole et validation

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

- Nombre de tours (runs) : 100.
- Taille de population : 150.
- Itération maximale de l'algorithme: 2000.
- Taux de croisement = 0,9.
- Fin de recherche : basée sur la variation «variation-based».

### 4.3.3. Fonction de score

Lors de la simulation du docking plusieurs poses candidates peuvent être obtenues pour chaque ligand, ici, nous sélectionnons les meilleures poses avec les énergies les plus basses selon la fonction de score basée sur une grille «MolDock Score [Grid]». La fonction de score,  $E_{score}$ , est définie par les termes d'énergie suivants:

$$E_{score} = E_{inter} + E_{intra}$$

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

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

Toutes ces énergies sont stockées et regroupées dans un fichier destiné à être inspecté dans le «work space» du MVD.

### 4.3.4. Validation du protocole de docking

Pour vérifier si le protocole que nous venons de développer est correct, nous avons jugé utile d'évaluer d'abord la performance du programme Molegro virtuel Docker (MVD) par les deux tests suivants :

- L'écart quadratique moyen ou le RMSD (root mean square deviation)
- La superposition des ligands.

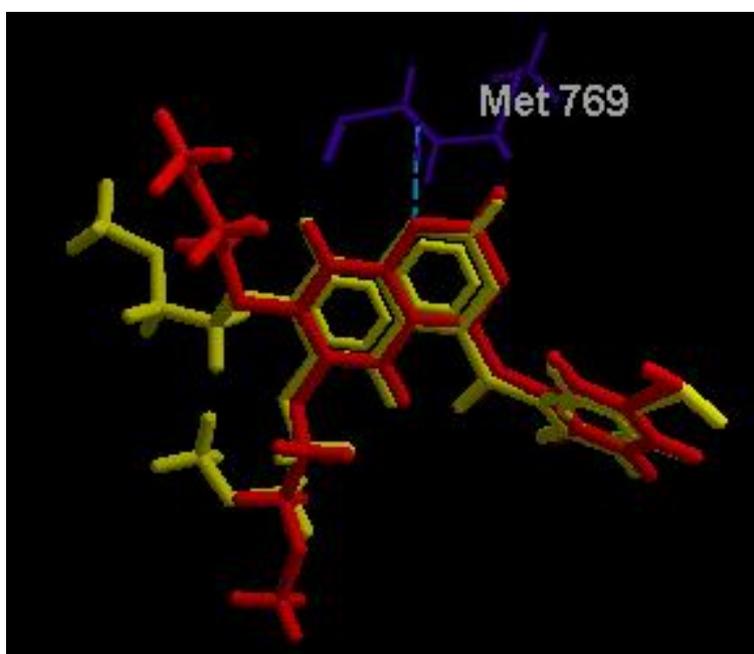
Plusieurs méthodes sont utilisées pour évaluer les performances des différents programmes de docking pour chaque application. L'aptitude d'un algorithme à trouver l'emplacement correct du ligand par rapport à son récepteur est habituellement déterminée au moyen de la déviation quadratique moyenne ou RMSD <sup>[45]</sup> (root-mean-square deviation) du modèle conçu par le logiciel vis-à-vis de la structure du cristal. La valeur admise est une

## CHAPITRE II. CRIBLAGE VIRTUEL DE DIAMIDES

différence maximale de 2 angströms au-delà de laquelle la prédiction est considérée comme non adéquate [46-47].

Le test RMSD a été réalisé sur le complexe (1M17) disponibles dans la PDB. Il s'agit plus précisément, de complexe formés entre l'enzyme EGFR et l'Erlotinib comme ligand. Pour notre étude, La simulation du re-docking a généré un RMSD équivalent à 1.8 Å, inférieur du seuil, ce qui indique que notre protocole a réussi et qu'il est applicable pour le reste de cette étude.

La superposition de l'Erlotinib et sa meilleur pose générée (Figure II.7), visualisée par Molegro, confirme la précision du positionnement et la génération d'une conformation quasi similaire à celle cristallographique originale.



**Figure II . 7:** La superposition de la structure cristallographique du ligand (en jaune) et la pose obtenue après re-docking (en rouge), RMSD =1.8Å

### 4.4. Criblage de la chimiothèque

Le criblage débute par l'importation de tous les ligands de notre chimiothèque dans le «work space» du MVD, une fois la chimiothèque est importée, en appliquant le protocole précédemment décrit et validé. Le «scoring» des poses obtenues pour chaque ligand (après 100 tours de recherche) mène à un classement des meilleures poses basées sur leurs énergies d'interaction «Moldock score».

## CHAPITRE II. CRIBLAGE VIRTUEL DE DIAMIDES

Ainsi, pour chaque élément de la chimiothèque, une seule pose prédite est sélectionnée, et donc seize poses prédites sont retenues dans le tableau II.4.

**Tableau II . 4:** Les énergies d'interaction Moldock score des meilleures poses obtenues pour chaque ligand de la chimiothèque.

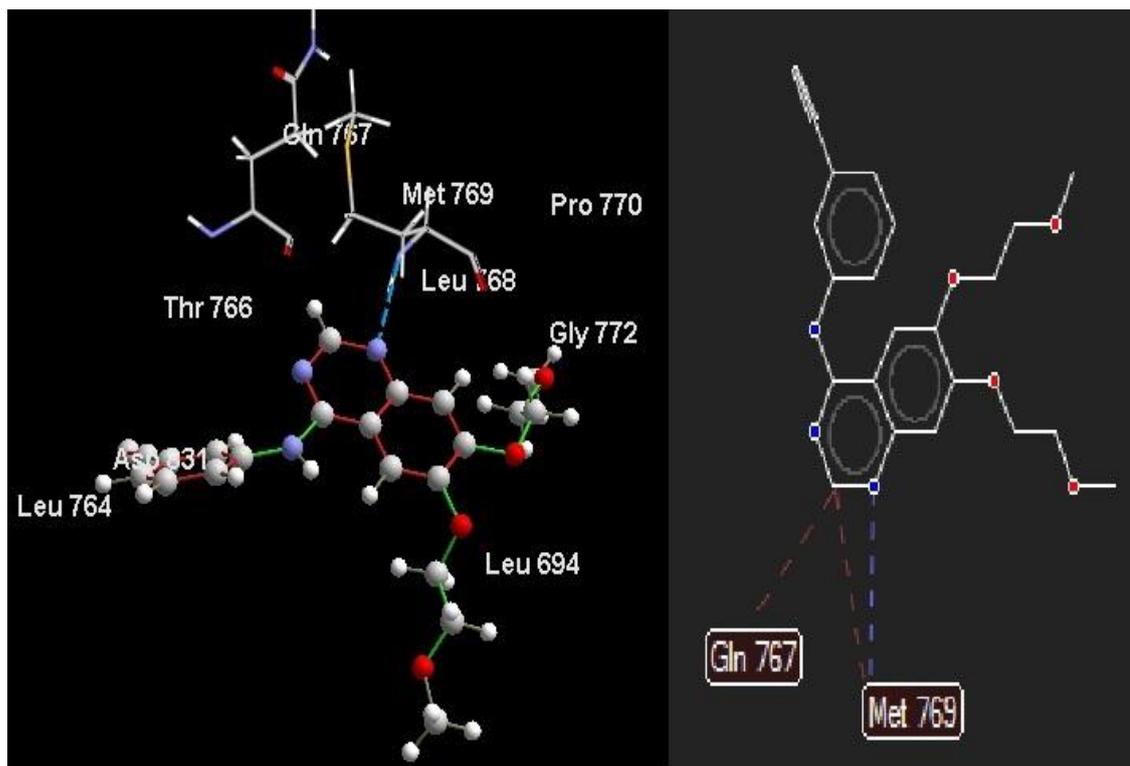
Molécule	Energie totale (Kcal/mol)	L'énergie d'interaction (Kcal/mol)	Interaction de type van der Waals (Kcal/mol)	L'énergie L.H (Kcal/mol)
Erlotinib	-137,86	-150,84	-148,34	-2,50
UG1	-110,403	-132,396	-128 ,374	-4,022
UG2	-101,775	-132,72	-128,280	-4.44
UG3	-118,49	-142,31	-134,31	-7,41
UG4	-117,15	-141,37	-138,96	-2,41
UG5	-124,36	-147,14	-136,17	-10,98
UG6	-110,02	-136,43	-131,34	-5,09
UG7	-113,85	-146,84	-140,41	-6,42
UG8	-114,92	-138	-135,65	-2,35
UG9	-124,93	-148,08	-135,65	-12,43
UG10	-108,5	-139,57	-133,87	-5,69
UG11	-114,23	-137,77	-130,89	-6,88
UG12	-117,15	-133,60	-125,96	-7,63
UG13	-120,273	-151,430	-146,712	-4,718
UG14	-120,010	-147,036	-147,036	0
UG15	-114,727	-147,392	-144,893	-2,49
UG16	-134,131	-147,228	-147,228	0

## CHAPITRE II. CRIBLAGE VIRTUEL DE DIAMIDES

Les valeurs d'énergie d'interaction Moldock score des dérivés de diamides sélectionnés varient de -134,13 à -101,77 Kcal/mol, tandis que celle de l'erlotinib est de -137,86 Kcal/mol, ce qui indique que ces composés pourraient être d'une affinité comparable à celle du référence ce qui confère une stabilité meilleur dans le site actif. Les énergies des liaisons hydrogène et de VDW sont également importantes en comparaison avec l'Erlotinib.

### 4.5. Analyse virtuelle des interactions

La structure cristalline de l'EGFR en présence de deux conformères de l'erlotinib a été déterminée (Figure II.8). Sur cette structure, nous remarquons que l'atome d'azote N1 du cycle quinazoline est engagé dans une liaison hydrogène ( $d = 2,7 \text{ \AA}$ ) avec la fonction amine du résidu Met 769 du récepteur (ligne en pointillé bleu). L'autre atome d'azote de la quinazoline, N3 n'est pas à une distance compatible avec la formation d'une liaison hydrogène avec la chaîne latérale du résidu Thr 766. Une liaison stérique avec le résidu Glu 767 a eu lieu (ligne en pointillé rouge).



**Figure II . 8:** Interactions entre Erlotinib et les résidus du site d'interaction.

Les recuits simulés des différents complexes EGFR-ligand sont réalisés à partir des complexes obtenus par minimisation d'énergie.