

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université Mentouri de Constantine
Faculté des Sciences de la Nature et de la Vie
Département de Biochimie - Microbiologie

N° d'ordre : 020/Mag/2012
N° de série : 006/SN/2012

MEMOIRE

Présenté par

Hioual Khadidja Soulef

Pour obtenir le diplôme de MAGISTERE EN BIOCHIMIE

Option

«Technologies des explorations biochimiques»

THEME

Etude comparative de l'efficacité de deux programmes de
docking et application à l'inhibition de la neuraminidase

Soutenu le : 27/02/2012

Devant le jury :

Président : Mr BENSEGUENI A.	Maître de conférences	Université Mentouri Constantine
Rapporteur : Mr CHIKHI A.	Maître de conférences	Université Mentouri Constantine

Examineurs :

Mr KHELIFI D.	Professeur	Université Mentouri Constantine
Mr BOUDAH A.	Maître de conférences	Université Mentouri Constantine

❧ Remerciements ❧

Le grand merci revient encore et toujours à notre DIEU le tout puissant qui m'a donné le courage et la patience et qui a éclairé mon chemin pour achever ce travail.

La réalisation de ce mémoire est indissociable de son contexte familial, scientifique et matériel. Nombreux sont ceux qui m'ont soutenu, aidé ou supporté. Je suis reconnaissante à tous d'avoir été là quand il le fallait, de s'être investi, d'avoir partagé, d'avoir contribué autant que possible, pour que ce mémoire se fasse.

Je tiens tout d'abord à remercier :

- Monsieur. CHIKHI Abdelouahab, Maître de Conférences à l'université Mentouri de Constantine, qui a tout d'abord accepté de m'encadrer, puis a soutenu, orienté, enrichi, critiqué, bref, dirigé... ma recherche. Je lui dois évidemment aussi beaucoup pour le soutien moral et quotidien qu'il m'a apporté, ainsi que la patience qu'il a su montrer.

-Les membres de mon Jury :

- ★ Monsieur BENSEGVENI Abderrahmane, Maître de Conférences à l'université Mentouri de Constantine, qui a bien voulu présider ce jury.*
- ★ Monsieur KHELIFI Douadi, Professeur à l'université Mentouri de Constantine, et Monsieur BOUDAH Abdennacer, Maître de Conférences à l'université Mentouri de Constantine, auxquels je suis très reconnaissante d'avoir accepté d'être examinateurs de ce travail.*

Khadidja Soulef.H

❧ Dédicaces ❧

Merci mon DIEU de m'avoir permis d'arriver jusqu'ici et de m'avoir donné l'aptitude d'achever ce modeste travail que je dédie particulièrement à mes très chers et adorables parents qui m'ont inculqué toutes les bases de mon savoir, que DIEU me les garde.

Je dédie aussi ce travail à mes frères et sœurs : Mohamed Saleh, Noureddine, Ghania, Souad, Ouassila et Ouided

A Hemmam Sofiane , Amouche Ilhem. & Nezzar Sorya

A Mes adorables Kamel Ouassim, Louai Eddine & Imen

A mes très chères ami(e) s : Karim, Imen, Houda, Hanene, Sabrina, Sihem, Aicha, Ouahiba, Naziha, Nouzha, Amina, Sabah, Khadidja, Lyes, Abdelmadjid.

A mes collègues : Amina, Hanene, Meriem, Camélia, Rym , Soumiya , Afaf, Mohsen, Bilel.

Je n'oublierai pas de dédier cet ouvrage à tous ceux ou celles qui ont ne serai-ce qu'un moment partagé ma vie.

Khadidja Soulef .H

TABLE DES MATIERES

LISTE des figures	
LISTE des tableaux	
LISTE des abréviations	
INTRODUCTION	1
CHAPITRE 1. Les interactions en Biologie : défis et complications	3
1. Enjeux socio-économiques	4
1.1. La recherche biologique comme composante d'un espace d'innovation	4
1.2. L'importance de la biochimie, une science récente	4
1.3. La réponse de la science aux questions de la biochimie.....	5
1.3.1. Approche expérimentale	5
1.3.2. Approche théorique.....	6
1.4. Effets sur la société	7
2. INTERACTIONS.....	8
2.1. Les liaisons fortes	8
2.1.1 La liaison covalente.....	8
2.1.2 La liaison ionique.....	9
2.2 Les liaisons faibles.....	10
2.2.1 Les forces de van der Waals.....	10
2.2.2 La liaison hydrogène.....	11
2.2.3 L'effet hydrophobe.....	13
CHAPITRE 2. L'informatique : une nouvelle méthode de recherche scientifique	15
1. La modélisation moléculaire et ses applications	16
2. Le <i>docking</i>	17
2.1. Principe	17
2.2. Représentation de la surface des protéines	18
3. L'évaluation de l'affinité d'un ligand pour une protéine donnée	19
3.1. Les fonctions de score basées sur un champ de force	19
3.2. Les fonctions de score empiriques	20
3.3. Les fonctions de score basées sur des connaissances statistiques	20
3.4. Les fonctions consensus	21

3.5. Autres types de fonctions de score	21
CHAPITRE 3. Méthodes de recherche conformationnelle et algorithmes de <i>docking</i>	22
1. Le <i>docking</i> corps rigide multi-conformationnelle.....	23
2. Le ligand flexible	23
2.1. Le ligand flexible : La recherche systématique.....	23
2.2. Le ligand flexible : La recherche stochastique	27
2.2.1. Les méthodes Monte Carlo	28
2.2.2 Les algorithmes génétiques	28
2.2.3. La recherche Tabou	29
2.3. Le ligand flexible : La recherche déterministe	30
3. Le récepteur flexible	30
CHAPITRE 4. Matériel et Méthode	32
1. Matériel	33
1.1. Ordinateur	33
1.2. Programmes utilisés	33
1.2.1. GOLD 5.0.1.....	33
1.2.2. FlexX 1.3.0.....	35
1.2.3. Open Babel 2.0.2.....	36
1.3. Protein Data Bank ou PDB.....	37
2. Méthode.....	37
2.1. Les ensembles de test	37
2.2. La procédure GOLD	38
2.2.1. Initialisation des protéines et des ligands	38
2.2.2. Détermination du site de liaison du ligand	39
2.2.3. La flexibilité du ligand	39
2.2.4. La flexibilité de la protéine	39
2.2.5. Arrangement des contraintes	39
2.2.6. <i>Docking</i> avec GOLD	39
2.3. La procédure FlexX	40
2.3.1. Le ligand.....	40
2.3.2. Le récepteur	41
2.3.3. Définir le site de liaison du ligand	41
2.3.4. Le <i>docking</i> avec FlexX	41

2.4. Les paramètres de comparaison étudiées	42
2.4.1. RMSD ou Root Mean Square Deviation	42
2.4.2. Les liaisons rotatives	42
2.4.3. La vitesse	42
2.4.4. Les faux positifs	43
CHAPITRE 5. Résultats et Discussion	44
1. La précision du processus du <i>docking</i>	45
2. La précision dans le classement des solutions.....	49
3. Les liaisons rotatives et la performance du processus du <i>docking</i>	50
4. Le temps d'exécution et la performance du processus du <i>docking</i>	52
5. Les faux positifs dans les programmes de <i>docking</i>	53
6. Rôle des molécules d'eau dans la performance du processus du <i>docking</i>	56
CHAPITRE 6. EXEMPLE D'APPLICATION : Etude des inhibiteurs de la neuraminidase	60
1. Mises aux points bibliographiques.....	61
1.1. La neuraminidase : origine, structure et fonction	61
1.2. Inhibiteurs de la neuraminidase	64
1.3. Changements antigéniques et résistance	66
2. Matériel et méthode	68
2.1. Choix de la cible	68
2.2. FlexX 1.3.0	69
2.3. PubChem	70
2.4. La règle de Christopher A. Lipinski (1997)	70
3. Résultat et discussion	71
3.1. Vérification de la règle de Lipinski	71
3.2. Le <i>docking</i> par FlexX 1.3.0	73
Conclusion	77
Bibliographie	
ANNEXE 1. Résultats-FlexX	
ANNEXE 2. Résultats-GOLD	
ANNEXE 3. Résultats-FlexX-GOLD	
Résumés	

LISTE DES FIGURES

1.1	Exemple d'éléments identifiables par l'expérience sur une molécule biologique et de la technique associée. <i>Source Angew. Chem. Int. Ed.</i> [18].....	7
1.2	Processus de formation d'une liaison ionique illustré avec l'exemple de NaCl. ...	10
1.3	La liaison hydrogène est une interaction entre deux dipôles, l'un formé par le donneur et l'atome d'hydrogène, l'autre par l'accepteur et le carbone auquel il est lié.....	11
1.4	Structures secondaires des protéines [31].....	12
1.5	Exemple d'interactions intervenant dans le repliement d'une chaîne polypeptidique.....	13
2.1	Principales étapes des méthodes de docking.	18
2.2	Surface du site actif d'une protéine dont la représentation est basée sur les harmoniques sphériques.....	19
3.1	A) Définition du site de liaison dans DOCK. B) Variation de l'angle de torsion dans DOCK.....	24
3.2	Technique utilisée pour placer le fragment. Les trois centres d'interaction du ligand (en gris). Les trois points d'interactions du site actif (en noir).....	26
3.3	Représentation schématique du cycle évolutif typique d'un GA.....	29
3.4	Méthodes automatisées de docking les plus connues et exemples d'algorithmes (en gris) fréquemment cités dans la littérature [74].....	31
4.1	Illustration de la construction de solution basée sur le point de fixation.....	33

5.1	Le complexe 3CKZ.....	47
5.2	Le complexe 1BSZ.....	47
5.3	Le complexe 2ZR1.....	48
5.4	Le complexe 3NRB.....	51
5.5	Le complexe 1JSN.....	52
5.6	Le résultat du docking du complexe 1EUS par GOLD.....	55
5.7	Le résultat du docking du complexe 2VW0 par GOLD.....	56
5.8	Le ligand n° 19 :2DVD.....	57
6.1	Structure du Virus de la grippe	62
6.2	Structure tridimensionnelle de la zone globulaire de la neuraminidase [103].....	63
6.3	Structure chimique de l'acide sialique : acide N-acétyl neuraminique (Neu5Ac) ou (NANA).....	64
6.4	Structures chimiques dans l'ordre du : <i>DANA</i> , <i>Zanamivir</i> , <i>Oseltamivir</i> , <i>Peramivir</i>	64
6.5	La surface moléculaire de la neuraminidase N1 liée avec le zanamivir (le complexe 3CKZ).....	69
6.6	Neuraminidase N1-zanamivir (Le complexe 3CKZ).....	74
6.7	Neuraminidase N1-le similaire n°26.....	75
6.8	Structure du similaire n°26 du zanamivir	76

LISTE DES TABLEAUX

1.1	Les trois catégories de liaison hydrogène [30].	12
4.1	Paramètres pris en compte dans le docking par GOLD.....	40
5.1	Temps d'exécution du processus d'arrimage pour les deux programmes GOLD et FlexX	52
6.1	Les similaires de zanamivir répondant aux critères de la règle de Lipinski	72
6.2	Résultats du docking du zanamivir et ses similaires à 90% dans la neuraminidase N1	73

LISTE DES ABREVIATIONS

ADME/Tox	Absorption, Distribution, Métabolisme, Excrétion, Toxicité
AMM	Autorisation de Mise sur le Marché
CADD	<i>Computer-Aided Drug Design</i>
CCDC	<i>Cambridge Crystallographic Data Centre</i>
CPU	<i>Central Processing Unit</i>
DANA	Acide 2-Deoxy-2,3-Dihydro-Nacetylneuraminique
GA	<i>Genetic Algorithm</i>
GOLD	<i>Genetic Optimisation For Ligand Docking</i>
HA	Hémagglutinine
INAs	Inhibiteurs de la Neuraminidase
MC	Monte Carlo
MD	<i>Molecular Dynamic</i>
NA	Neuraminidase
NANA	Acide N-acétyl neuraminique
NCBI	<i>National Center for Biotechnology Information</i>
NCI	<i>National Institutes of Health</i>
PDB	Protein Data Bank
PDF	Peptide Deformylase
PSI	<i>Protein Structure Initiative</i>
RMN	Résonance Magnétique Nucléaire
RMSD	Root Mean Square Deviation
RSGI	<i>Riken Structural Genomics/Proteomics Initiative</i>
SPINE	<i>Structural Proteomics in Europe</i>

Introduction

Aujourd'hui, la protéomique et les multiples techniques expérimentales ou calculatoires qui l'accompagnent, prennent le relais pour extraire la fonction des protéines codées, *via* la connaissance de leur structure et de la manière dont elles interagissent avec d'autres molécules. Puisque le Projet du Génome Humain (*The Human Genome Project* [1]) a permis de décrypter la séquence de presque 30 000 gènes mais n'a pas donné d'information directe sur leur fonction.

Les interactions intra et intermoléculaires sont décisives lors de la formation des complexes biomoléculaires et pour assurer la viabilité des cellules. En considérant par exemple qu'une protéine peut établir en moyenne 5 interactions différentes, on peut donc estimer à 150 000 au moins le nombre d'interactions dans les cellules humaines [2]. Comprendre et prédire les mécanismes d'interactions macromoléculaires ainsi que leurs possibles modifications avec le temps, le milieu, ou encore les maladies, représente donc un enjeu inestimable dans le domaine de la santé et de la recherche pharmaceutique.

En se maintenant au niveau moléculaire, la biologie s'appuie dans un premier temps sur les méthodes expérimentales pour identifier des protéines, leur structure et leur affinité [3]. Une telle approche ne peut cependant pas convenir à l'étude d'un grand nombre d'interactions, toutes les interactions moléculaires ne pouvant être identifiées par des études expérimentales, longues et coûteuses. La voie informatique présente une bonne alternative aux limites expérimentales : la recherche *in silico* est actuellement extrêmement active, son essor laisse présager son importance et son caractère incontournable dans les années à venir.

Les méthodes informatiques utilisées à l'heure actuelle ont deux objectifs : d'une part étudier la structure tridimensionnelle et les changements conformationnels d'une molécule, et d'autre part simuler les mécanismes d'amarrage moléculaire, ou *docking*.

Le processus du *docking* est itératif et chaque passe de calcul s'articule en deux étapes. Tout d'abord, une portion limitée de l'espace des conformations du complexe protéine-ligand est explorée afin d'améliorer la pose (orientation et/ou conformation) du ligand dans le site actif de la protéine. L'étape suivante fait intervenir une fonction de score qui évalue la qualité de la pose générée [4]. Cette estimation *in silico* de l'affinité du ligand pour la cible est basée sur un examen simplifié des interactions entre les deux partenaires. La répétition du cycle de recherche associé à l'estimation du score guidant l'exploration de l'espace doit assurer la convergence de l'algorithme vers un état représentatif du minimum global de l'énergie libre d'association.

Etant donnée la grande diversité des programmes de *docking*, ce mémoire a pour but, dans un premier temps, de faire une comparaison entre deux logiciels de docking GOLD et FlexX à travers différents paramètres : l'écart quadratique moyen rmsd ou *root mean square deviation*, le nombre de liaisons rotatives (pouvant tourner), le temps d'exécution et le taux de faux positifs afin de savoir lequel des deux logiciels est le plus performant et le plus fiable et, dans un deuxième temps, appliquer les résultats de cette comparaison pour trouver de nouveaux inhibiteurs de la neuraminidase à travers les similaires tirés de la PubChem. Cette enzyme constitue un des composants de l'enveloppe du virus de la grippe, épidémie pouvant se transformer en une pandémie.

Nous présentons nos travaux dans ce mémoire au cours de deux parties :

Une partie théorique avec trois chapitres : Le premier chapitre introduit l'approche de la biologie à l'heure actuelle, à travers l'étude des événements atomiques et moléculaires. Dans un second chapitre, nous nous intéressons à une nouvelle méthode de recherche scientifique qui est le *docking*, son principe ainsi que la notion de score. Le troisième chapitre est consacré aux méthodes de recherche conformationnelle et aux algorithmes de *docking*.

Une partie pratique, elle-même divisée en trois : une partie consacrée au matériel utilisé dans notre travail et la méthode suivie, une deuxième qui traite les résultats obtenus et les discussions apportées et la dernière est une application des résultats précédents pour proposer de nouveaux inhibiteurs de la neuraminidase en utilisant les similaires d'un de ces inhibiteurs tirés de la PubChem.

CHAPITRE 1

Les interactions en Biologie : défis et complications

Chaque être vivant est constitué d'unités structurales élémentaires : les cellules [5]. Chaque cellule est elle-même une entité vivante et fonctionnelle, autonome mais coordonnée avec son environnement. Elle a la capacité de grandir, de se reproduire, de recevoir et d'émettre des signaux, de traiter des informations... Toute son activité est régie par de multiples réactions chimiques et de fines interactions inter et intramoléculaires.

C'est à ce niveau qu'interviennent la biologie moléculaire, la biologie structurale et la biochimie : dans la compréhension du vivant, par l'étude d'événements moléculaires, depuis la formation, la structure et la fonction des molécules biologiques jusqu'à leur implication dans les processus chimiques et les assemblages macromoléculaires.

1. Enjeux socio-économiques

1.1 La recherche biologique comme composante d'un espace d'innovation

Les innovations scientifiques et les techniques de pointe élaborées par les biologistes et les biochimistes dans les laboratoires pour les avancées de leur discipline initient des processus industriels nouveaux, notamment dans l'industrie pharmaceutique ou agrochimique. Ceci tend à produire d'autres interrelations entre science (biologie moléculaire, génétique, biochimie) et technologies, et de nouvelles configurations de réseaux d'innovation associant aux laboratoires de recherche fondamentale et de recherche-développement, des groupes pharmaceutiques et des services cliniques [6].

Le document "Biologie 1990 - Enjeux et problématiques" (1987) fixait à la recherche biologique quatre enjeux majeurs pour la société : la santé, l'agro-alimentaire, l'utilisation des micro-organismes, et l'environnement.

1.2. L'importance de la biochimie, une science récente

On peut situer le début de la biochimie en 1828, lorsque Wöhler synthétise pour la première fois de l'urée, une substance organique sécrétée par les êtres vivants à partir de cyanate d'argent. La théorie du vitalisme [7], jusque là admise, entame son déclin. Distinguant les êtres vivants de la matière inerte en leur associant une force vitale différente de l'âme, et en donnant ainsi à la biologie une indépendance par rapport aux autres sciences, le vitalisme laisse place à une conception plus matérialiste de la vie qui lève l'indétermination : on admet désormais que les règles physico-chimiques des êtres vivants sont les mêmes que celles régissant la matière inanimée.

Cette nouvelle vision modifie totalement le rapport de l'homme aux sciences de la vie, et par conséquent, leur place dans notre société et nos préoccupations [8]. Le XIX^{ème} siècle assiste ainsi à la rationalisation des savoirs et des observations. Les progrès de la physique et de la chimie permettent le triomphe des méthodes expérimentales. En 1864, Claude Bernard publie L'introduction à la médecine expérimentale et confirme la déroute du vitalisme. Il propose l'existence de lois naturelles rigoureuses gouvernant la physiologie tout comme la physique et la chimie, et étant la cause de l'existence de tout phénomène vivant.

Aujourd'hui, avec l'avancée des technologies modernes d'exploration et de conception, ainsi que des moyens de calculs, de nombreuses questions trouvent peu à peu leur réponse : Quelles sont les structures chimiques et tridimensionnelles des molécules biologiques ? Dans quelle mesure leur fonction change-t-elle avec cette structure ? Quels sont les mécanismes de reconnaissance, d'assemblage ? Quels mécanismes inter ou intramoléculaires leur permettent de fonctionner ? Par exemple, comment une enzyme réalise-t-elle la catalyse ? De quelle façon son activité est-elle régulée ? Comment une protéine peut-elle recevoir ou émettre un signal ? Comment peut-on exprimer et transmettre l'information génétique ?

1.3. La réponse de la science aux questions de la biochimie

1.3.1. Approche expérimentale

Depuis le milieu du XX^{ème} siècle, avec la découverte de la structure de l'ADN puis le séquençage du génome, la quantité de molécules répertoriées ne cesse de croître. On estime qu'il existe environ 100 000 protéines chez l'homme, dont un millier possède une structure tridimensionnelle connue [9], et dont seulement une centaine est employée comme cible pour les médicaments. Les techniques expérimentales permettent d'identifier à haut débit les structures tridimensionnelles des molécules biologiques. Elles contribuent ainsi à l'effort mené sur l'étude de la fonction de ces molécules, de leurs mécanismes d'interactions et de leur implication dans l'activité normale ou pathologique des cellules.

Deux méthodes sont généralement utilisées pour déterminer les structures tridimensionnelles à haute résolution des macromolécules biologiques: la cristallographie par diffraction des rayons X [10] et la spectroscopie par résonance magnétique nucléaire (RMN) [11]. De très nombreuses structures sont mises à la disposition de l'utilisateur : la Protein Data Bank (PDB) [9] par exemple, recensa en 2009 près de 55000 structures de molécules, dont 45000 ont été résolues par rayons X et 7500 par RMN [12]. Plus de 80% des structures sont résolues par diffraction des rayons X. Cette méthode, contrairement à la spectroscopie RMN, n'est pas limitée par la taille des molécules. Elle nécessite néanmoins une cristallisation préalable qui peut présenter des difficultés. La RMN quant à elle peut fournir des réponses sur la dynamique. Notons que la résolution de complexes de macromolécules demeure limitée : en 2006, elle ne représentait que 1.5% des structures disponibles [13].

De multiples obstacles doivent être franchis par l'expérimentateur, notamment dans le choix de la méthode qui doit convenir à la propriété recherchée, ainsi qu'à l'échelle de temps du phénomène (qui peut aller de la femtoseconde (10^{-15} s) à l'heure [14]). Il existe donc de nombreuses autres méthodes expérimentales selon la propriété structurale ou physico-chimique que l'on cherche à identifier [15] ; la Figure 1.1 présente un exemple d'éléments recherchés sur une protéine ainsi que la technique associée.

1.3.2. Approche théorique

Dans le but de renforcer les connaissances de la communauté scientifique et d'aider à la compréhension de la relation séquence-structure-fonction, des consortiums de génomique structurale tentent d'organiser les efforts au niveau international pour déterminer la structure 3D d'un maximum de protéines différentes (PSI (*Protein Structure Initiative*), RSGI (*Riken Structural Genomics/Proteomics Initiative*) et SPINE (*Structural Proteomics in Europe*) [16]). Cependant, en raison de difficultés techniques et du coût humain élevé nécessaire à la résolution expérimentale des structures, le fossé entre le nombre de séquences protéiques connues et le nombre de structures continue de se creuser.

Dans ce contexte, les méthodes de bioinformatique structurale ont un rôle particulièrement important à jouer et doivent fournir des méthodes alternatives permettant de réduire cet écart. Selon les travaux d'Åqvist, "toute l'information nécessaire pour obtenir la conformation native d'une protéine dans un environnement donné est contenue dans l'enchaînement des acides aminés [17]. Cette hypothèse est tout à fait fondamentale et fondatrice pour les méthodes de bioinformatique structurale visant à prédire la structure tertiaire d'une protéine à partir de sa séquence. Par ailleurs, selon Levinthal, la formation simultanée de petits noyaux structurés dans plusieurs régions d'une même chaîne polypeptidique initierait et accélérerait le repliement. Ainsi, cette seconde hypothèse, aujourd'hui considérée comme complémentaire de celle d'Åqvist, met en avant l'importance des interactions à courte distance et la formation de petites structures locales au cours du processus de repliement.

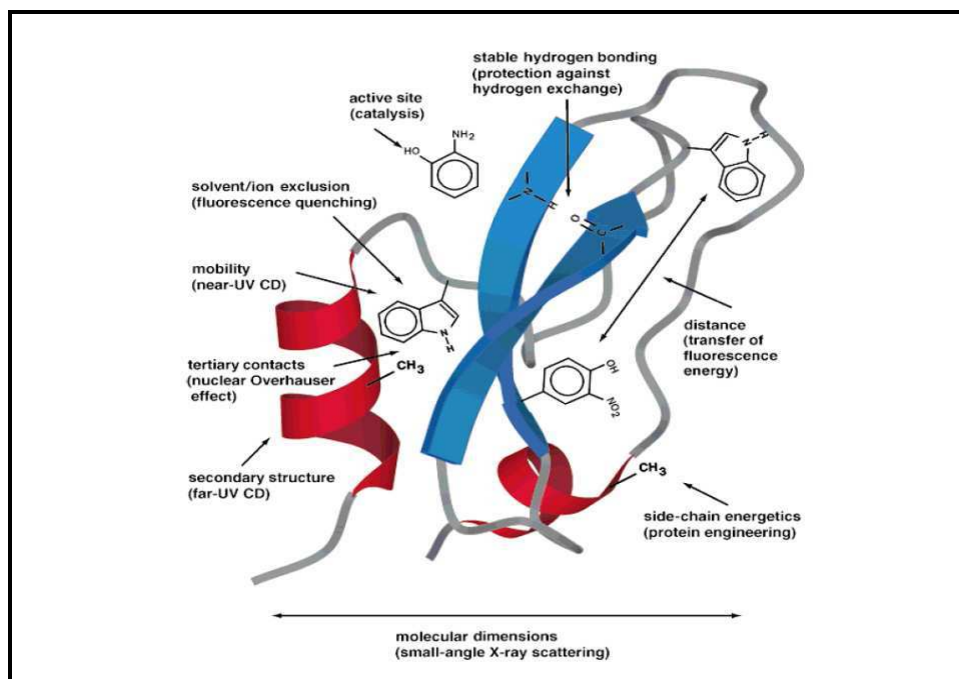


Figure 1. 1 Exemple d'éléments identifiable par l'expérience sur une molécule biologique et de la technique associée. *Source Angew. Chem. Int. Ed.* [18].

1.4. Effets sur la société

Ces développements débouchent sur des progrès quotidiens de la biologie dont l'impact sur la société est de ce fait considérable. La compréhension des mécanismes moléculaires de maladies permet la conception et l'optimisation de produits pharmaceutiques, la compréhension de maladies conformationnelles (Alzheimer, Parkinson, maladies à prions, amyloses...), le dépistage d'anomalies génétiques, mais aussi la modification d'êtres vivants avec des applications dans le domaine agro-alimentaire, le clonage [19].

Ces avancées deviennent également la source de nouvelles technologies en plein essor : les nanobiosciences [20]. Elles se situent à l'interface de la physique, la chimie, la biologie et l'ingénierie et présentent de nombreuses perspectives dans des domaines tels que la médecine (diagnostic avec les biopuces, thérapie, organes artificiels), la visualisation et la manipulation de molécules uniques (nanoparticules magnétiques), l'imagerie biologique (nanoparticules fluorescentes).

Aujourd'hui, les outils bioinformatiques font parler les structures recueillies par les expérimentateurs afin d'en inférer des fonctions biologiques présentant un intérêt

(économique, industriel, santé publique, agroalimentaire). Les biomolécules entrent alors en jeu pour la fonctionnalisation de surface (biopuces, biocapteurs...), la détection de cibles impliquées dans les fonctions biologiques (diagnostic précoce). Tous ces dispositifs nécessitent l'élaboration de molécules, parfois même leur conception *de novo* par la biologie de synthèse [21] pour arriver à des systèmes présentant de nouvelles propriétés physico-chimiques adaptées à une fonction dans la nanoingénierie.

2. Interactions

Le développement de la biologie moléculaire, ainsi que de la biologie structurale, depuis le milieu du siècle dernier a permis de comprendre, par la voie de l'expérience et de l'analyse, les processus biologiques à une échelle de plus en plus petite [22]. Les outils de simulation permettent depuis le début des années 1980 d'étudier ces processus sous la forme d'interactions entre biomolécules jusqu'à l'échelle atomique. Grâce à ces avancées scientifiques et techniques, de nombreuses perspectives industrielles ont vu le jour, notamment dans le domaine de la médecine.

Dans cette partie, ces forces intermoléculaires seront décrites, mais également les forces intramoléculaires qui maintiennent la molécule dans un état stable. Parmi ces forces d'interaction, deux classes se distinguent : les liaisons chimiques, qui assurent la stabilité de la molécule, et les liaisons physiques, qui vont régir l'évolution de la molécule dans son milieu [23].

2.1. Les liaisons fortes

Dans l'organisme, les molécules intervenant dans les processus biologiques peuvent être définies comme des groupements d'atomes ayant une stabilité suffisante pour conserver leur identité lorsqu'elles sont en interaction avec l'environnement [24]. Cette stabilité est assurée par les interactions intramoléculaires qui sont les liaisons chimiques : liaisons covalentes et liaisons ioniques.

2.1.1. La liaison covalente

La liaison covalente est le résultat de l'interaction entre deux atomes qui se partagent de manière équiprobable leur paire d'électrons de valence [25]. Ce sont des interactions qui ont lieu à courte distance (typiquement entre 0.5 et 2 angströms) et elles possèdent une énergie relativement élevée comparée aux valeurs des autres interactions ou de l'agitation thermique.

A titre d'exemple, les valeurs de l'énergie de ces liaisons varient de 24 fois à 350 fois kT (avec $1\text{kJ/mol} \sim 0.4kT$ par liaison à 300K) [23].

La liaison covalente est le type commun de liaison où la différence d'électronégativité entre les atomes liés est petite ou nulle. Dans ce dernier cas, on parle parfois de liaison purement covalente. Lorsque les atomes se rapprochent, leurs nuages électroniques se recouvrent et les atomes partagent alors leurs électrons de valence. Un équilibre s'établit entre forces attractive et répulsive entre les atomes. Suivant le nombre de paires d'électrons partagé entre les atomes, la liaison sera plus ou moins forte. Les atomes peuvent partager 1, 2, 3 paires d'électrons et peuvent ainsi former des liaisons simple, double ou triple. Par exemple, dans le cas de la liaison O-O, la liaison simple possède une énergie de 145 kJ/mol , alors que celle de la liaison double est de 498 kJ/mol [26].

2.1.2. La liaison ionique

La liaison ionique est une liaison chimique similaire à la liaison covalente, impliquant deux atomes dont la différence d'électronégativité est supérieure à 1,7. L'atome le moins électronégatif cède un ou plusieurs électrons à l'atome le plus électronégatif, formant ainsi des ions positifs et négatifs respectivement, tous deux possédant alors une configuration électronique stable [25]. Sous l'effet de l'attraction électrostatique, les deux ions se rapprochent jusqu'à une distance d'équilibre et forment alors une liaison ionique. La liaison ionique résulte de l'attraction entre deux ions de charges opposées, alors que dans le cas de la liaison covalente, les atomes sont liés par le partage d'électrons. Mais en réalité, il n'existe pas de liaison ionique « pure », car théoriquement il faudrait une différence d'électronégativité infinie entre les atomes. Les liaisons ioniques se distinguent par leur pourcentage de covalence.

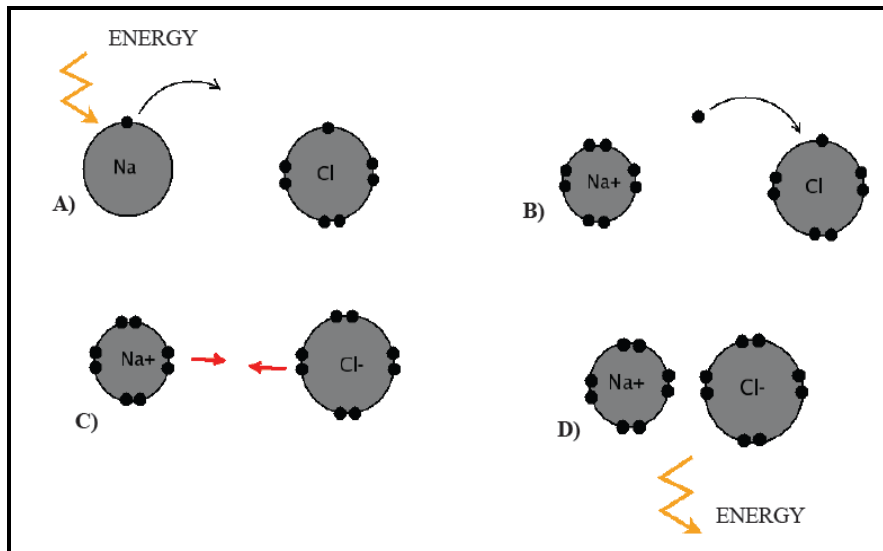


Figure 1.2 Processus de formation d'une liaison ionique illustré avec l'exemple de NaCl. A) suite à un apport externe d'énergie, l'atome de sodium perd son électron célibataire. B) l'électron libre est alors récupéré par le chlore. C) les deux ions de signes opposés résultant de ce transfert de charge ont alors leur couche de valence remplie et s'attirent. D) on observe alors la formation d'une liaison ionique, exothermique.

2.2. Les liaisons faibles

Les liaisons physiques sont des interactions qui ont lieu pour des distances plus élevées que les liaisons chimiques. Pour la plupart d'entre elles, elles ont une intensité moindre que les liaisons chimiques, mais leur portée est bien plus grande [23]. Cela permet aux molécules de réagir avec d'autres éléments éloignés du milieu. Ce sont ces interactions qui régissent l'évolution des biomolécules, et notamment les changements conformationnels. A l'intérieur des cellules par exemple, le repliement des protéines est régi par ces interactions avec le milieu [27]. On va donc décrire dans la section qui suit un certain nombre de liaisons physiques qui jouent un rôle important en biologie.

2.2.1. Les forces de van der Waals

Les interactions électrostatiques sont à la base des phénomènes physiques entre les atomes chargés. Le principe de Pauli prévoit qu'à mesure que deux molécules s'approchent l'une de l'autre, et que leurs orbitales électroniques se chevauchent, il y a répulsion si les électrons ne peuvent partager leur sphère électronique [28]. Deux molécules stables devront donc respecter un certain volume impénétrable, nommé le rayon van der Waals. L'énergie générée par le

rapprochement entre deux molécules est prédite par la loi de Coulomb, qui résulte du produit des deux charges divisé par la distance entre elles. Si les charges sont de signes opposés, l'énergie (E) diminue à mesure que les molécules s'approchent et donc, l'interaction est favorable. Si les charges sont de même signe, l'énergie augmente considérablement et devient défavorable.

Les interactions électrostatiques de type Van der Waals sont des liens d'assez faible énergie, entre $1-7 \text{ kcal mol}^{-1}$ [28], et la force de l'interaction est alors proportionnelle à l'électronégativité des molécules.

2.2.2. La liaison hydrogène

La liaison hydrogène est une force attractive qui s'opère entre deux groupes d'atomes impliquant un atome d'hydrogène. C'est une liaison physique qui se situe entre la liaison covalente et l'interaction électrostatique. Le caractère covalent de la liaison hydrogène a été démontré à la fin des années 1990, mais la force d'interaction prédominante reste électrostatique [29]. Cette liaison ne s'effectue pas avec n'importe quel atome. Il faut un donneur et un accepteur (sous entendu de proton) :

- Le donneur est l'atome d'hydrogène lié de manière covalente à un atome électronégatif ; typiquement azote, oxygène, fluor.
- L'accepteur est un autre atome électronégatif ; uniquement azote, oxygène ou fluor. La liaison hydrogène est donc le résultat d'une interaction dipôle-dipôle (cf. Figure 1.3)

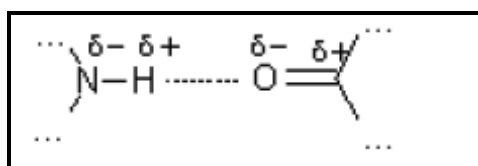


Figure 1.3 La liaison hydrogène est une interaction entre deux dipôles, l'un formé par le donneur et l'atome d'hydrogène, l'autre par l'accepteur et le carbone auquel il est lié.

Les liaisons hydrogène sont souvent décrites par deux paramètres : la longueur de la liaison et l'angle formé par les trois atomes. Les valeurs des angles et des distances varient respectivement de 90 à 180 degrés et de 1.2 à 4.0 Å.

Ces paramètres caractéristiques qui dépendent essentiellement du milieu vont déterminer la force des liaisons hydrogènes. On distingue trois catégories répertoriées dans le Tableau 1.1

Tableau 1. 1 Les trois catégories de liaison hydrogène [30]. La longueur est la distance entre l'atome d'hydrogène et l'accepteur ; l'angle est formé par les trois atomes formant la liaison hydrogène : l'accepteur, l'hydrogène et le donneur.

Longueur (Å)	Angle (°)	Energie (kcal/mol)	Description
2.2 à 4.0	90 à 150	0.24 à 1	Faible intensité, comparable à Van der Waals. Liaison de type électrostatique
1.5 à 3.2	130 à 180	1 à 4	Liaisons modérées, situées entre les deux extrêmes
1.2 à 2.5	175 à 180	4 à 10	Liaison forte de type covalente

Ces interactions qui s'établissent entre deux groupes d'atomes, peuvent s'effectuer entre deux molécules différentes (extramoléculaires), mais également au sein d'une même molécule (intramoléculaires). Dans ce dernier cas, elles vont tenir un rôle prépondérant dans la structure de la molécule. Dans le cas des protéines par exemple, les structures secondaires possèdent un réseau bien particulier de liaisons hydrogène qui contribuent à maintenir la molécule dans cette conformation (cf. Figure 1.4).

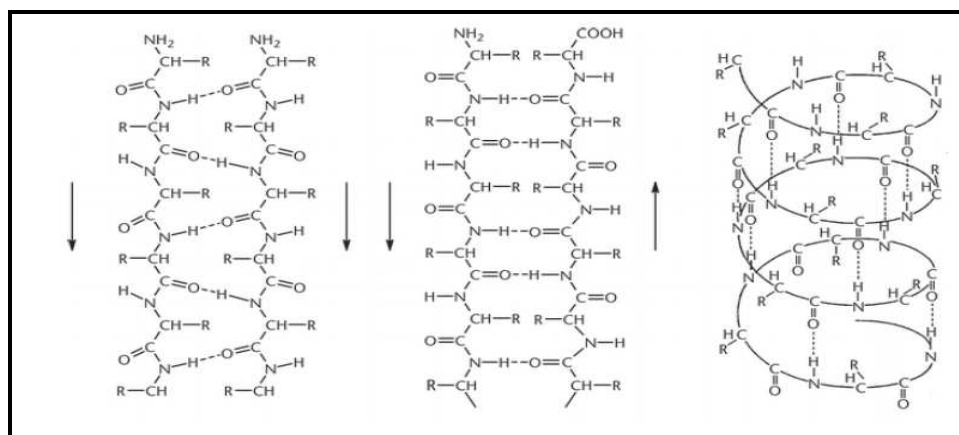


Figure 1.4 Structures secondaires des protéines. De gauche à droite : feuillet β antiparallèle ; feuillet β parallèle ; hélice α . □ □ Les liaisons hydrogènes sont représentées en pointillé. Les flèches indiquent le sens N-terminal vers C-terminal [31].

Ces structures déterminées par Pauling et ses collaborateurs en 1951 [32] se trouvent être des conformations très répandues. En effet, parmi toutes les structures de protéines connues à l'heure actuelle, plus de la moitié des acides aminés sont dans une de ces conformations [9].

2.2.3. L'effet hydrophobe

Les molécules ou groupes d'atomes non polaires ne sont pas capables de former des liaisons hydrogènes et ne peuvent donc pas s'hydrater : pour cette raison, on les nomme 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. Les liaisons hydrophobes possèdent une énergie de l'ordre de 7 kcal/mol [30].

Une molécule hydrophobe rompt localement le réseau de liaisons hydrogènes du solvant et crée un état énergétiquement défavorable (perte d'entropie). Les molécules d'eau se réorganisent alors autour des substances hydrophobes de manière à minimiser l'énergie. D'un point de vue thermodynamique, la diminution du nombre de molécules d'eau autour des groupes non-polaires et l'agrégation de ces derniers sous l'effet des forces de dispersion, c'est-à-dire la séparation de deux phases distinctes, organique et aqueuse, s'accompagne d'un gain d'entropie de l'ensemble du solvant et de la molécule non polaires. Pour cette raison, les régions non polaires sont souvent enfouies à l'intérieur des molécules. Cet effet a d'importantes répercussions sur la conformation et la stabilité des molécules. Elle est à l'origine du processus de repliement des protéines par exemple, ou encore de l'autoassemblage de biomolécules [33].

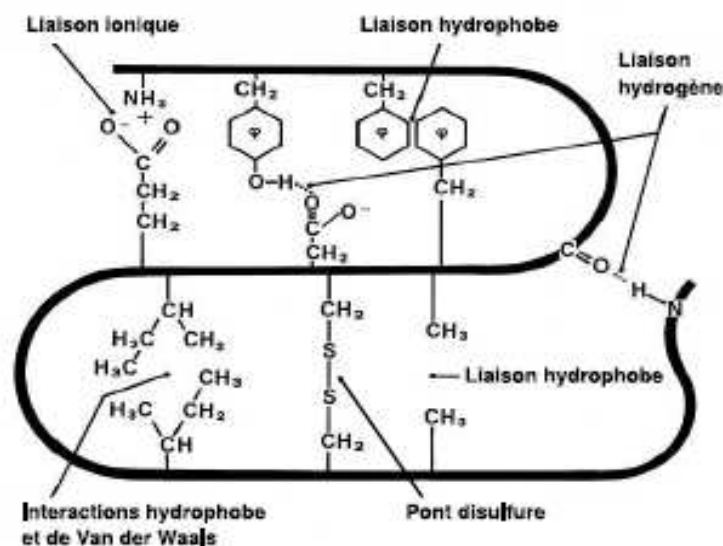


Figure 1.5 Exemple d'interactions intervenant dans le repliement d'une chaîne polypeptidique.

Dans ce chapitre, nous avons décrit rapidement les intérêts actuels de la biologie structurale : identifier et comprendre la fonction biologique (et donc pharmacologique) d'une biomolécule *via* la connaissance de sa structure, de ses mécanismes et de son réseau d'interactions.

Aujourd'hui, si les outils expérimentaux et calculatoires permettent d'ouvrir la voie à l'étude des interactions inter et intra moléculaires, ce projet n'en demeure pas moins ambitieux. L'expérimentation ne permet pas à ce jour d'apporter toutes les réponses et reste lente et coûteuse, malgré ses indiscutables progrès.

Une nouvelle voie s'ouvre alors à nous : celle du calcul, qui, en se basant sur les lois usuelles de la chimie et de la physique, est tout à fait à même de s'appliquer aux problèmes de la biologie. Le calcul se doit néanmoins d'être rapide et rentable, d'une part pour traiter des données abondantes, et d'autre part pour modéliser les systèmes extrêmement complexes de la biologie.

CHAPITRE 2

L'informatique : une nouvelle méthode de recherche scientifique

La conception de molécules d'intérêt thérapeutique a bénéficié ces dernières décennies des progrès issus de diverses disciplines scientifiques telles que la biologie, la pharmacochimie et l'informatique. Ainsi la recherche, qui consistait autrefois à synthétiser et tester les composés sélectionnés sur la base de l'intuition et de l'expérience du chimiste médicinal, a radicalement évolué. L'essor de l'outil informatique a particulièrement changé la donne, en conduisant à l'émergence d'une nouvelle discipline pouvant participer aux étapes initiales de la recherche pharmaceutique en complément des méthodes expérimentales déjà reconnues. On parle alors de conception de médicaments *in silico* – c'est-à-dire assistée par ordinateur – qui correspond à un ensemble de techniques informatiques spécifiques souvent désigné par l'acronyme CADD [34] (pour "*Computer-Aided Drug Design*"). Bien que ces outils aient un large champ d'application dans le processus de recherche de nouveaux médicaments, nous nous limiterons à la description des méthodes utilisées pour ce travail, à savoir l'arrimage moléculaire, ou "*docking*" [35] [36].

1. La modélisation moléculaire et ses applications

La modélisation moléculaire est un domaine scientifique multidisciplinaire : elle exploite des lois de la chimie, de la physique et de la biologie dans des programmes informatiques spécifiques afin de calculer structures et propriétés d'entités chimiques et biochimiques (protéines, acides nucléiques, complexes moléculaires, solides, cristaux etc.). Le but est la compréhension ou la prédiction des phénomènes auxquels s'intéressent ces disciplines [37].

Depuis les années 80, la modélisation moléculaire connaît un essor continu, non seulement comme outil scientifique précieux pour des domaines fondamentaux (tels que la chimie et la biologie) ou appliqués (tels que la pharmacie et les ingénieries), mais aussi comme une branche scientifique à part entière, consacrée à l'amélioration de la robustesse des logiciels et de l'efficacité des processeurs. Ce développement peut être attribué à deux facteurs principaux : le premier est l'évolution de l'informatique au cours des dernières décennies. Ceci a permis l'intégration de concepts de chimie théorique dans les algorithmes et le développement de machines de plus en plus puissantes, capables d'exécuter les lourds calculs numériques requis et de stocker les données générées. Le deuxième facteur est l'élucidation des structures tridimensionnelles de nombreuses protéines, grâce aux progrès de la génomique, et des techniques de cristallisation et résolution structurale par diffraction au rayon X. En effet, les structures tridimensionnelles de ces biomolécules sont le point de départ pour des simulations visant à comprendre leurs interactions, entre elles ou avec des petites molécules, ce qui est à la base d'un grand nombre de phénomènes chimiques, biologiques et biochimiques [38]. Ainsi, la modélisation moléculaire trouve de nos jours d'importantes applications, parmi lesquelles trois exemples classiques sont :

- L'étude des propriétés des matériaux : à partir des simulations moléculaires, un lien peut être établi entre les résultats de calculs menés à l'échelle microscopique (sur un échantillon représentatif d'atomes en interaction) et des propriétés mesurables à l'échelle macroscopique (sur les systèmes matériels réels que l'on souhaite maîtriser). Ainsi, peuvent être étudiées les propriétés mécaniques des polymères, les propriétés électroniques de solides cristallins ou encore les propriétés thermodynamiques et spectroscopiques d'une large gamme de composés d'intérêt scientifique et technologique [39].

- Le développement de nouveaux médicaments : le mécanisme d'action de nombreux médicaments consiste à agir comme inhibiteur (ligand) d'une enzyme (récepteur) impliquée dans le développement de la maladie, que ce soit une protéine d'un microorganisme pathogène ou du propre organisme humain. L'élucidation de la structure tridimensionnelle de protéines impliquées dans plusieurs pathologies a permis, *via* des simulations informatiques, la découverte d'inhibiteurs puissants pour ces protéines, en réduisant considérablement le nombre d'essais de screening nécessaires pour aboutir à un nouveau médicament (*rational drug design*) [40, 41].
- La rationalisation de l'ingénierie enzymatique : le fondement est le même que celui du développement de médicaments : en permettant d'étudier les interactions entre les substrats (ligand) et les enzymes (récepteurs), la modélisation moléculaire apporte une meilleure compréhension de la sélectivité enzymatique, au niveau moléculaire. Ceci peut rendre possible, par la suite, le développement d'approches prédictives pour la sélectivité des enzymes natives vis-à-vis d'un substrat ou d'une classe de substrats donnée (*rational process design*), ou encore, d'identifier des acides aminés dont la mutation ponctuelle permettrait de modifier la sélectivité ou la spécificité enzymatiques (*site-directed mutagenesis*) [42].

2. Le *docking*

Ce mémoire se limitera ici à une description assez générale des techniques de *docking* afin de situer le contexte de notre travail.

2.1. Principe

Le *docking* est utilisé pour prédire la structure du complexe intermoléculaire résultant de l'association entre au moins deux molécules. Quand il s'agit de deux protéines, on parle de *docking* protéine- protéine [43], par opposition au *docking* protéine-ligand que nous avons utilisé pour ce travail.

Tout d'abord, une portion limitée de l'espace des conformations du complexe protéine-ligand est explorée afin d'améliorer la pose (orientation et/ou conformation) du ligand dans le site actif de la protéine. L'étape suivante fait intervenir une fonction de score qui évalue la qualité de la pose générée [4]. Cette estimation *in silico* de l'affinité du ligand pour la cible est basée sur un examen simplifié des interactions entre les deux partenaires (cf. Figure 2.1).

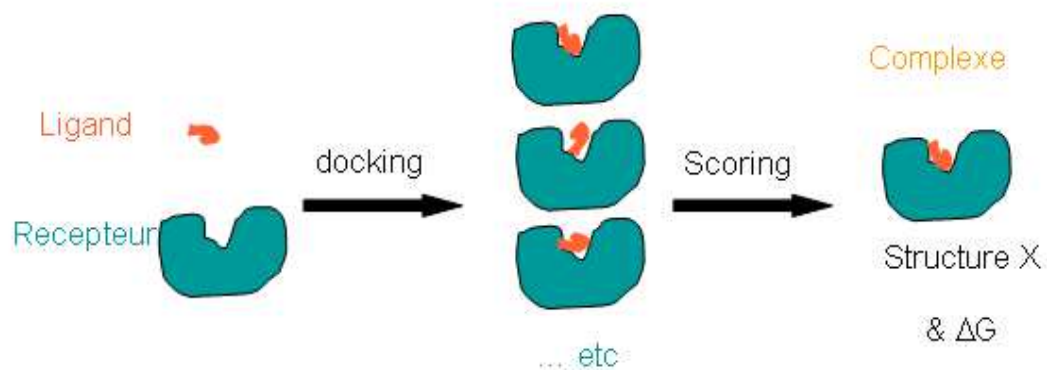


Figure 2.1 Principales étapes des méthodes de docking

2.2. Représentation de la surface des protéines

Un des problèmes des programmes de *docking* réside dans la représentation de la surface des protéines. En effet, peu de méthodes de *docking* utilisent une représentation explicite des chaînes latérales lors des premières étapes de recherche, en raison du coût important que cela engendre en terme de temps de calcul. Généralement, le choix de représentation des chaînes latérales est également lié à l'algorithme de recherche conformationnelle, comme c'est le cas pour les programmes de recherche sur grilles [44], ces derniers représentant la structure des protéines en les projetant sur une grille tridimensionnelle. Certains programmes modélisent la surface des protéines par des harmoniques sphériques ou encore par une représentation simplifiée basée sur les surfaces de Conolly (cf. Figure 2.2). Ces représentations de la surface des protéines peuvent être "adoucies" pour permettre une certaine prise en compte de la flexibilité des chaînes latérales de façon implicite, en autorisant par exemple un certain degré de recouvrement des surfaces. Enfin, d'autres approches utilisent des approximations, telle qu'une représentation simplifiée des chaînes latérales par des centroïdes [45].

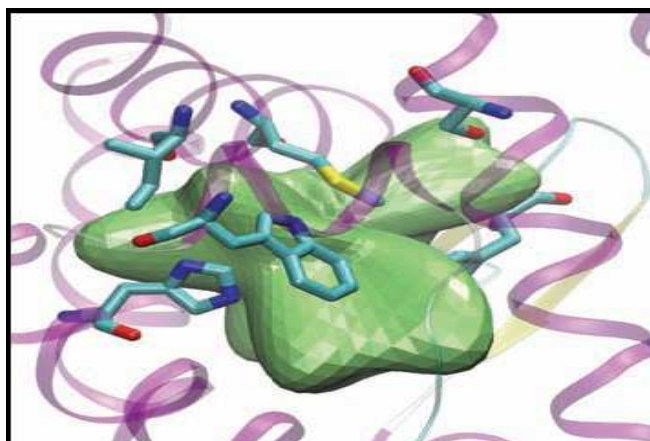


Figure 2.2 Surface du site actif d'une protéine dont la représentation est basée sur les harmoniques sphériques.

3. L'évaluation de l'affinité d'un ligand pour une protéine donnée

Des techniques basées sur le calcul d'énergie libre ont été développées et constituent un moyen quantitatif rigoureux pour estimer l'affinité de liaison d'un ligand pour une protéine [46]. Bien qu'elles soient très précises, leur coût (en temps, en contraintes et en expertise) ne les rend pas appropriées pour une application de *docking* dans le cadre d'un criblage virtuel. Pour représenter le meilleur compromis entre vitesse et précision, les fonctions de score implémentées dans les programmes de *docking* sont donc basées sur la simplification des phénomènes impliqués dans la reconnaissance moléculaire, en particulier de ceux qui sont délicats à évaluer en dehors des calculs d'énergie libre (p. ex. l'entropie).

Les différentes fonctions de score implémentées dans les programmes de *docking* ont fait l'objet de nombreuses publications [4] [47] [48]. Elles sont généralement classées suivant trois catégories : les fonctions de score basées sur un champ de force, les fonctions de score empiriques et celles basées sur des connaissances statistiques. Enfin, plusieurs fonctions de score de nature différente peuvent être combinées pour former des fonctions de score dites de consensus.

3.1 Les fonctions de score basées sur un champ de force

Les champs de force dans leur forme standard évaluent la somme de deux énergies : l'énergie entre atomes liés au sein d'une molécule donnée (énergie interne) et l'énergie entre atomes non-liés. Ce deuxième terme correspond au terme principal de l'énergie d'interaction protéine-ligand dans le cas d'une application à un problème de *docking*. La plupart du temps,

les fonctions de score basées sur un champ de force ne considèrent qu'une conformation donnée de la protéine. Ainsi, si l'on compare l'activité de deux ligands par rapport à cette même conformation, cela permet de faire abstraction du terme d'énergie interne de la protéine qui s'annule dans l'expression de la différence d'énergie libre d'interaction entre les deux ligands.

De telles fonctions de score présentent certaines limitations qui s'ajoutent à celles induites par la représentation du système (p. ex. la non représentation explicite du solvant). En particulier, les effets d'entropie, qui peuvent varier d'un ligand à l'autre pour un site actif donné, aussi bien que pour un ligand donné d'un site actif à l'autre, ne sont pas pris en compte. Seule la contribution enthalpique de l'énergie libre d'interaction est ainsi prise en compte.

Les fonctions G-Score [49] (basée sur le champ de force de Tripos [49]) et celle implémentée dans AutoDock [50] (basée sur le champ de force AMBER) sont des exemples de ce type de fonction de score.

3.2. Les fonctions de score empiriques

Ce type de fonction de score approxime l'énergie libre de liaison en sommant de façon pondérée différents termes d'interaction dérivés de paramètres structuraux. Les différents poids de la fonction de score sont ajustés pour reproduire en priorité des données expérimentales, telles que les constantes de liaison tirées d'un jeu d'entraînement de complexes protéine-ligand.

La plupart des programmes de *docking* implémentent ce type de fonction de score témoignant de leur efficacité (en terme de rapport précision/rapidité). Cependant, le principal inconvénient de ces fonctions empiriques est leur forte dépendance aux données utilisées pour les calibrer qui, en cas de mauvaise paramétrisation, peut limiter leur transférabilité sur des systèmes différents. Parmi les principales fonctions de score empiriques, on peut citer : LigScore [51].

3.3. Les fonctions de score basées sur des connaissances statistiques

Ces fonctions de score sont construites à partir de règles fondées sur une analyse statistique des complexes protéine-ligand résolus expérimentalement. Elles partent du principe que les distances interatomiques les plus représentées statistiquement dans les complexes constituent des contacts énergétiques favorables et, qu'à l'inverse, les plus rares représentent des

interactions moins stables. Ainsi, leur paramétrisation dépend de la quantité d'informations expérimentales disponibles et on doit leur apparition à la profusion de données structurales accessibles dans des bases de données telles que la PDB. Comme exemple populaire de ces fonctions de scores : DrugScore [52].

3.4. Les fonctions consensus

Ces fonctions hybrides combinent les résultats issus de diverses fonctions de score. On estime qu'il est possible de compenser partiellement les faiblesses intrinsèques de chacune des fonctions de score employées, évitant leurs erreurs individuelles et ainsi d'augmenter la probabilité d'identifier des composés actifs [53]. Cependant, si les termes des différentes fonctions de score sont fortement corrélés, l'intérêt du consensus devient limité car il peut entraîner une amplification des erreurs, au lieu de les atténuer. Ces fonctions consensus ont récemment fait l'objet d'une revue [54].

3.5. Autres types de fonctions de score

Les méthodes de *docking* reposant sur des surfaces utilisent des fonctions de score qui sont adaptées et principalement centrées sur la complémentarité géométrique surface-surface et auxquelles il peut être adjoit une estimation d'interactions sur le modèle des fonctions de score plus conventionnelles. Par exemple, LigandFit [55] génère, par Monte Carlo, les conformations du ligand dont les formes sont ensuite comparées à celle du site actif. D'autres programmes, tels que FRED [56], compare la forme de chacun des conformères, générés au préalable, à la forme du site actif de la protéine.

L'utilisation des programmes de *docking* a conduit à de nombreux succès dans le domaine de la découverte de nouvelles molécules bioactives [57] ; néanmoins, leurs algorithmes sont toujours en maturation. Les axes principaux de leur amélioration méthodologique visent essentiellement à considérer la totale flexibilité de la protéine pendant le processus de *docking* et à prendre en compte les effets d'entropie et de solvant dans l'évaluation de l'affinité du complexe protéine-ligand par la fonction de score.

De façon plus générale, le choix d'un programme donné pour un criblage virtuel par *docking* doit répondre à la question suivante : « Que veut-on obtenir, en combien de temps, et avec quelle précision ?

CHAPITRE 3

Méthodes de recherche conformationnelle et algorithmes de *docking*

Diverses méthodes de *docking-scoring* ont été rapportées dans la littérature. La complexité du *docking* moléculaire implique plusieurs approximations, du *docking* corps rigide, au *docking* (pseudo)-flexible (où le récepteur est maintenu de manière rigide et le ligand est partiellement flexible) au *docking* flexible (où la flexibilité des deux récepteur et ligand est considérée). Les algorithmes traitants la flexibilité peuvent être divisés en trois types, à savoir systématique, recherches stochastiques et déterministes (par exemple, la minimisation de l'énergie et la dynamique moléculaire) [36].

1. Le *docking* corps rigide multi-conformationnelle

Parmi les différents programmes de *docking* ; de nombreux programmes du *docking* corps rigide ont été rapporté pour arrimer des conformations préalablement générées par correspondance des points d'interaction du site du récepteur avec les atomes du ligand. Dans les méthodes du *docking* corps rigide, une recherche orientationnelle du ligand dans la poche de liaison de la protéine est réalisée alors que le récepteur et le ligand restent rigides.

Un des premiers des programmes du *docking* moléculaire pour l'interaction protéine-petite molécule impliquant le *docking* corps rigide était DOCK qui a été développé par Kuntz et ses collaborateurs [58]. Le programme DOCK génère une image négative des sphères du récepteur qui remplissent la poche de liaison et représentent les sites d'interaction potentielle. L'algorithme DOCK tente de superposer les atomes du ligand sur les centres des sphères.

Un autre programme du *docking* corps rigide FRED (<http://www.eyesopen.com>) [56] applique une fonction Gaussienne d'ajustement de forme pour optimiser la surface de contact entre le ligand et la protéine qui permet une procédure du *docking* rigide extrêmement rapide. FRED filtre l'ensemble des poses en rejetant celles qui s'affrontent avec la protéine en utilisant une image négative du site actif. Les poses raffinées peuvent ensuite être classées par diverses fonctions de score.

Malgré les limites évidentes, les méthodes du *docking* corps rigide sont intéressantes car ils sont beaucoup plus rapides que les algorithmes de *docking* flexible. Des logiciels tels que FRED peuvent arrimer jusqu'à 10 composés par seconde dans un mono-processeur standard sous Linux [56]. La vitesse et la précision relative aux méthodes du *docking* corps rigides les rendent attractives [59].

2. Le ligand flexible

2.1. Le ligand flexible : La recherche systématique

Les algorithmes de la recherche systématique essayent d'explorer tous les degrés de liberté. Afin de réduire le nombre des évaluations à effectuer, le critère de terminaison est défini pour empêcher l'algorithme de faire face à l'explosion combinatoire. Les algorithmes de fragmentation / reconstruction (méthodes de construction incrémentale) divisent généralement le ligand en petits noyaux rigides et des jonctions flexibles (cf. Figure 3.1). Les fragments du

noyau rigides sont amarrés d'abord dans le site de liaison et les parties flexibles sont ajoutées progressivement pour reconstruire le ligand complet.

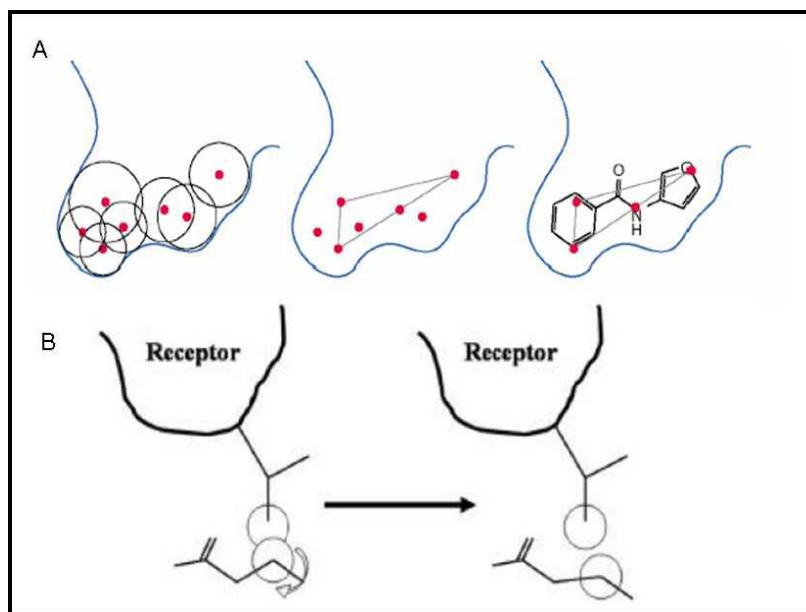


Figure 3.1 A) Définition du site de liaison dans DOCK. B) Variation de l'angle de torsion dans DOCK.

L'approche générale dans le programme DOCK (<http://dock.compbio.ucsf.edu/>) pour le traitement de la flexibilité du ligand [60] (la méthode ancrer-et-croître) est divisée en trois étapes principales. Premièrement, la détermination d'un ensemble de sphères qui se chevauchent en contact avec la surface du site du récepteur. Ces sphères remplissent la surface moléculaire du site de liaison et représentent une image négative du site de la cible. Deuxièmement, le centre de ces sphères est lié avec les atomes du ligand *via* l'utilisation d'un algorithme d'appariement de graphes. Troisièmement, une fonction de score est utilisée pour évaluer la pertinence des poses de docking par approximation de l'énergie de liaison protéine / ligand.

Le programme Surflex (www.biopharmics.com) [61] est basé sur un programme développé auparavant nommé Hammerhead [62]. Il utilise le même concept de détecteur de poche et la définition des investigations du site de liaison (protomols) mais il est caractérisé par une construction incrémentale innovante du ligand et récemment une fonction de score raffinée. Le programme crée d'abord un site de liaison idéalisé ; le protomol qui sert de cible à laquelle les ligands ou les fragments du ligand sont alignés sur la base de la similitude moléculaire.

Chaque ligand est fragmenté, résultant en 1 à 10 fragments moléculaires, dont chacun peut avoir certaines liaisons rotatives. Chaque fragment est ensuite recherché conformationnellement et chaque conformation de chaque fragment est alignée sur le protomol pour céder les poses qui maximisent la similitude moléculaire au protomol. Les termes de fonction de score impliquent par ordre de signification, la complémentarité hydrophobe, la complémentarité polaire, des termes entropiques et des termes de solvation.

Le logiciel FlexX (<http://www.biosolveit.de/>) [63] arrime des ligands flexibles aux récepteurs rigides en utilisant une approche incrémentale et certains concepts présents dans le programme LUDI [64]. L'approche peut être divisée en trois zones: la flexibilité conformationnelle, les interactions protéine-ligand et la fonction de score. La flexibilité conformationnelle du ligand est modélisée par un ensemble discret et préféré d'angles de torsion au niveau des liaisons simples acycliques et plusieurs conformations pour les systèmes d'anneau. En tenant compte du schéma d'interaction, FlexX repose sur la détection des interactions géométriquement restrictives telles que les liaisons hydrogènes, les interactions hydrophobiques spécifiques tels que les doublets phényl-méthyl, ou les surfaces sphériques qui sont issus des distances favorisées de l'interaction. L'algorithme de *docking* est divisé en trois phases: la sélection du fragment de base, le placement du fragment de base, et la construction du complexe, où le ligand est construit progressivement à partir du fragment de base. Le classement des ligands est réalisé par la fonction de score empirique modifiée de Böhm. Cette fonction comprend plusieurs termes (pondérés): un terme fixé, un terme tenant compte de la perte d'entropie lors de la fixation du ligand en raison de l'empêchement de rotation de liaisons dans le ligand, la liaison hydrogène, l'interaction ionique, l'interaction aromatique et l'interaction lipophile.

➤ Sélection du fragment de base : [63] [65]

La première étape du *docking* par FlexX est de choisir une partie du ligand qui va interagir en premier avec la protéine, cette partie est appelée « le fragment de base ». La technique de la sélection du fragment de base est la suivante :

- la première étape est la division du ligand en plusieurs composants par l'ouverture des liaisons simples acycliques non terminales. Par définition un fragment est dit valide lorsqu'il constitue la partie qui va se connecter à la protéine, et n'ayant pas plus de 30 conformations [63] [9].

- la deuxième étape consiste au classement des fragments valides.

➤ Placement de base :

Les algorithmes de placement sont basés sur la technique de « *geometric hashing* » [65] [66] [67]. Ces algorithmes sont le « *triangle algorithme* » (triplet) et le « *line algorithme* » (paire). Le premier algorithme est utilisé lorsqu'il y a trois interactions compatibles entre le ligand et le récepteur, il s'agit d'un triplet (sous forme de triangle) (cf. Figure 3.2) ; le second est utilisé lorsqu'il y a deux interactions compatibles, dans ce cas il s'agit d'une paire.

Pour des raisons géométriques les surfaces d'interactions du site actif sont présentées par des points (cette idée est similaire à celle développée par Böhm en 1992) [64].

La première étape de cette technique est de générer tous ces points d'interaction dans des cercles où la distance point-point est fixe. Sachant que la distance de la liaison hydrogène et les sels pontés est égale à 1.2 Å et que celle des interactions hydrophobiques est égale à 1.6 Å.

La technique « *geometric hashing* » traite toutes les paires de points qui ont une distance dans l'intervalle de 0.5 à 10 Å comme suit : le type d'interaction du 1^{er} point, le type d'interaction du 2^{ème} point et la distance. Elle génère tous les triplets (ou paires) des centres d'interaction du fragment et définit tous les triplets compatibles. On dit que les deux triangles définis précédemment sont compatibles seulement s'ils ont les mêmes longueurs d'arêtes et que leurs sommets ont des types d'interaction compatibles [63] (cf. Figure 3.2).

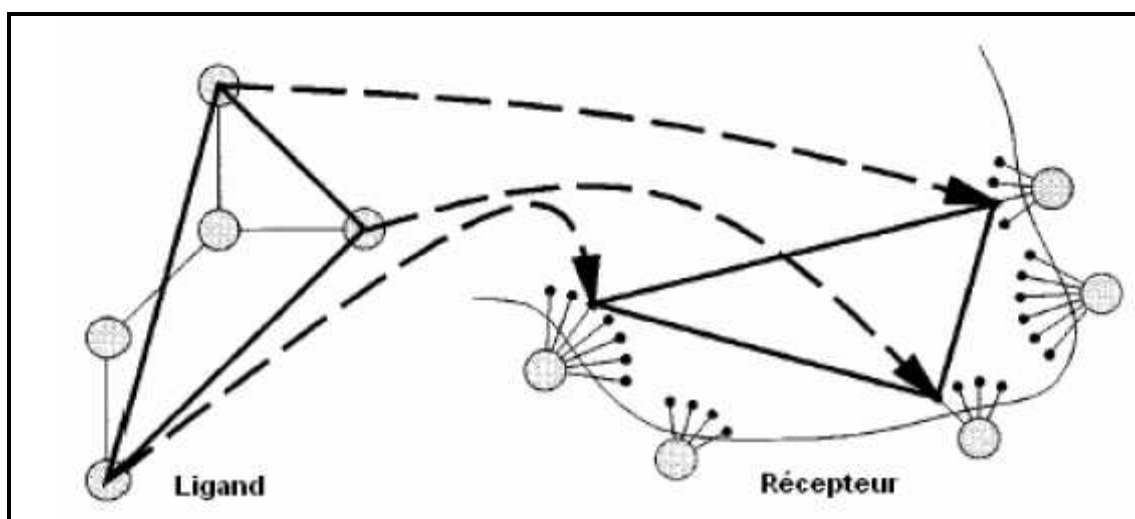


Figure 3.2 Technique utilisée pour placer le fragment. Les trois centres d'interaction du ligand (en gris). Les trois points d'interactions du site actif (en noir).

La deuxième étape est l'ajustement du placement en considérant :

- Les contraintes angulaires ou géométriques (le centre d'interaction du ligand coïncide avec la surface d'interaction du récepteur et *vice-versa*).
- Le ligand et la protéine (récepteur) ne doivent pas se chevaucher. La liste des placements (positions) obtenus par cette étape contient plusieurs interactions similaires pour deux raisons :
 - Changer un point par un autre proche dans la même surface d'interaction change légèrement l'interaction.
 - Le fragment peut faire plus de trois interactions simultanément.

La troisième étape sert à regrouper tous les placements qui ont un rms de déviation minimale, (rms : root mean square=0.7 Å). Ces placements vont converger vers une seule solution, celle qui correspondra à l'orientation finale du ligand [63] [68, 69].

La dernière étape est un test de chevauchement (l'algorithme ne prend en compte que les placements qui ne se chevauchent pas).

- Construction du complexe :

Le processus de la construction incrémentielle commence juste après avoir placé le fragment de base [63].

Cette technique est représentée par un arbre « technique arborescente » où chaque point représente une position ou un placement d'une partie du ligand [63]. Dans le premier niveau de cet arbre, on trouve les différentes positions du fragment de base, chaque niveau suivant contient des positions alternatives pour les autres fragments du ligand.

2.2. Le ligand flexible : La recherche stochastique

Les algorithmes de recherche stochastique (ou aléatoire) impliquent des changements aléatoires pour modifier la position du ligand (translation et rotation) ainsi que les angles de torsion afin de générer différentes conformations. Les principales méthodes de recherche stochastique sont Monte Carlo (MC), Algorithmes Génétiques (GA) et la recherche Tabou.

2.2.1. Les méthodes Monte Carlo

MC peut générer un ensemble de conformations qui sont statistiquement compatibles à une température donnée. Les perturbations aléatoires des positions atomiques sont appliquées afin d'explorer l'espace conformationnel du système moléculaire. Une fonction d'énergie évalue si l'énergie de la conformation nouvellement générée est soit inférieure à celle de la précédente étape ou, si elle est supérieure, se situe dans une gamme d'énergie définie par le facteur que l'on appelle de Boltzmann.

LigandFit (www.accelrys.com) [55] a des utilités pour prédire / définir le site de liaison en fonction de la forme de protéines (algorithme de remplissage en inondation). Une méthode de MC est employée pour la recherche conformationnelle du ligand. Lors de cette recherche, les longueurs de liaison /les angles de liaison sont intacts alors que les angles de torsion sont aléatoires. Multiples changements structurels peuvent donc se produire au même temps au cours de cette étape. Une fois une nouvelle conformation pour le ligand est générée, le montage du composé dans la poche de liaison est réalisée (procédure de recherche de similitude de forme), éventuellement suivie par la minimisation de corps rigide. Si la forme est similaire puis le ligand est amarré dans le site de liaison et que son énergie de liaison est évaluée à travers une fonction d'énergie appelé DockScore [55] impliquant un terme de Van Der Waals et un terme électrostatique avec une constante diélectrique dépendante de la distance et, éventuellement, l'énergie interne du ligand. Plusieurs fonctions de score sont disponibles, y compris, par exemple, Ludi [64], LigScore [51].

2.2.2. Les algorithmes génétiques

GA sont vaguement modélisés sur des concepts empruntés de l'évolution Darwinienne. Plusieurs conformations et positions du ligand sont générées, formant une population de solutions (cf. Figure 3.3). Ce premier état de population est évolué comme étant la plus basse énergie des positions. Cette population est soumise à un croisement et des mutations aléatoires ou biaisés afin de former la population suivante. L'algorithme maintient une pression sélective envers une solution optimale, avec une information randomisée d'échange permettant l'exploration de l'espace de recherche. Après des étapes successives d'évolution, les meilleures positions et conformations du ligand sont conservées résultants dans la pose du ligand de la plus basse énergie. Ces algorithmes sont utilisés dans de nombreux programmes de docking tels que GOLD [69], AutoDock [50]. Par exemple, dans GOLD, le mécanisme du

placement du ligand est basé sur les points de fixation. Le programme ajoute les points de fixation aux groupes de liaison hydrogène sur la protéine et le ligand, et cartographie les points accepteurs dans le ligand sur les points donneurs dans la protéine et *vice versa*. L'algorithme génétique optimise des dièdres flexibles du ligand, les géométries en cercle du ligand, les dièdres des groupes OH et NH des protéines, et les mappages des points de fixation.

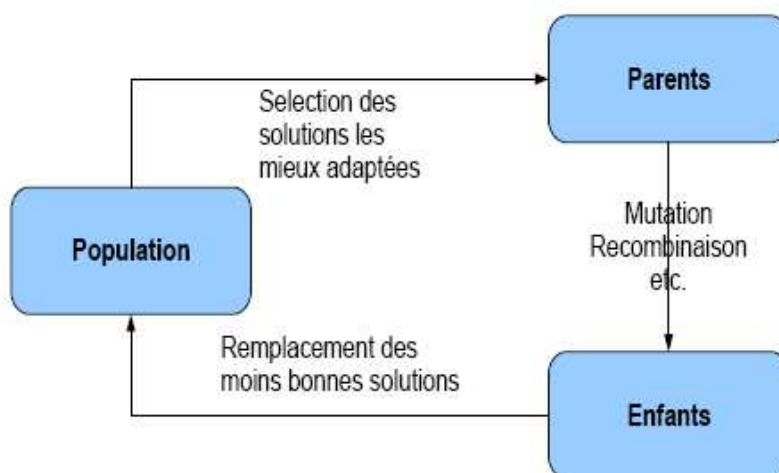


Figure 3.3 Représentation schématique du cycle évolutif typique d'un GA.

2.2.3. La recherche Tabou

La recherche Tabou combine une procédure de minimisation avec des restrictions sur le chemin de recherche, tels que la solution est forcée dans des régions inexplorées auparavant de l'espace de recherche. Il procède par étape à partir d'une solution initiale, tout en maintenant une liste des solutions précédentes. La liste des solutions précédentes fournit à la fois un classement des solutions et un enregistrement partiel des régions explorées de l'espace de recherche. L'algorithme Tabou génère un ensemble de N nouvelles solutions à partir d'une solution précédente et l'une des N solutions est maintenue. Une solution est ajoutée à la liste si elle est la meilleure solution à présent, ou la solution explore une nouvelle région de l'espace de recherche. Un algorithme de recherche Tabou est par exemple mis en œuvre dans PRO_LEADS [70].

2.3. Le ligand flexible : La recherche déterministe

Dans la recherche déterministe, l'état initial détermine le changement qui peut être fait pour générer l'état suivant, qui est généralement énergétiquement préféré par rapport à l'état initial. Un problème avec ces approches est qu'elles peuvent être piégées dans le minimal local. Les méthodes déterministes sont la minimisation de l'énergie et la dynamique moléculaire (MD) méthodes permettant une flexibilité du site de liaison du récepteur. La dynamique moléculaire par rapport aux simulations de MC, ne peut pas traverser facilement les barrières de haute énergie dans des délais raisonnables de simulation et à température ambiante [35, 36] comme exemple d'algorithmes combinant la dynamique moléculaire et le recuit simulé : SDOCKER [71].

3. Le récepteur flexible

Diverses stratégies sont proposées dans la littérature pour prendre en considération la flexibilité du récepteur. La façon la plus pertinente, mais apparemment la plus consommable en terme du temps est d'appliquer les méthodes déterministes comme les simulations de MD (voir ci-dessus). Différentes structures de protéines peuvent être générés par les simulations de MD ou de MC [72]. La méthodologie à base du mode normal a également été montrée pour incorporer la flexibilité du récepteur dans le *docking* du ligand et le criblage virtuel.

Cavasotto et Abagyan ont proposé ICM (www.molsoft.com) [73] algorithme de *docking* avec récepteur flexible (IFREDA) pour tenir en compte la flexibilité de la protéine lors du criblage virtuel. Le programme ICM est basé sur une simulation de MC qui repose sur l'optimisation globale de la fonction d'énergie du ligand flexible dans le champ du récepteur (ligand flexible / approche du récepteur de grille) (chaînes latérales du récepteur / flexibilité de la chaîne principale). Une procédure de minimisation de MC dans l'espace de coordonnées internes est employée pour rechercher le minimum global de la fonction d'énergie. Chaque étape de l'algorithme consiste en un changement aléatoire de deux types, de torsion ou de position, suivie par une minimisation locale. Les changements torsionnelles des chaînes latérales des acides aminés à l'interface peuvent également être appliqués en utilisant une méthodologie de probabilité biaisée. La fonction de score utilisée dans ICM se compose de l'énergie interne du champ de force du ligand et l'énergie d'interaction ligand / récepteur avec éventuellement un terme pour tenir compte de la taille du site de liaison / ligand. L'énergie d'interaction ligand / récepteur comprend plusieurs termes pondérés: Van Der Waals, un terme d'hydrophobicité basé sur la surface accessible au solvant enterrés sur la liaison, un

terme de solvation électrostatique calculé en utilisant une solution aux limites des éléments de l'équation de Poisson, les interactions de liaison hydrogène et un terme entropique proportionnelle au nombre de torsions flexibles dans le ligand.

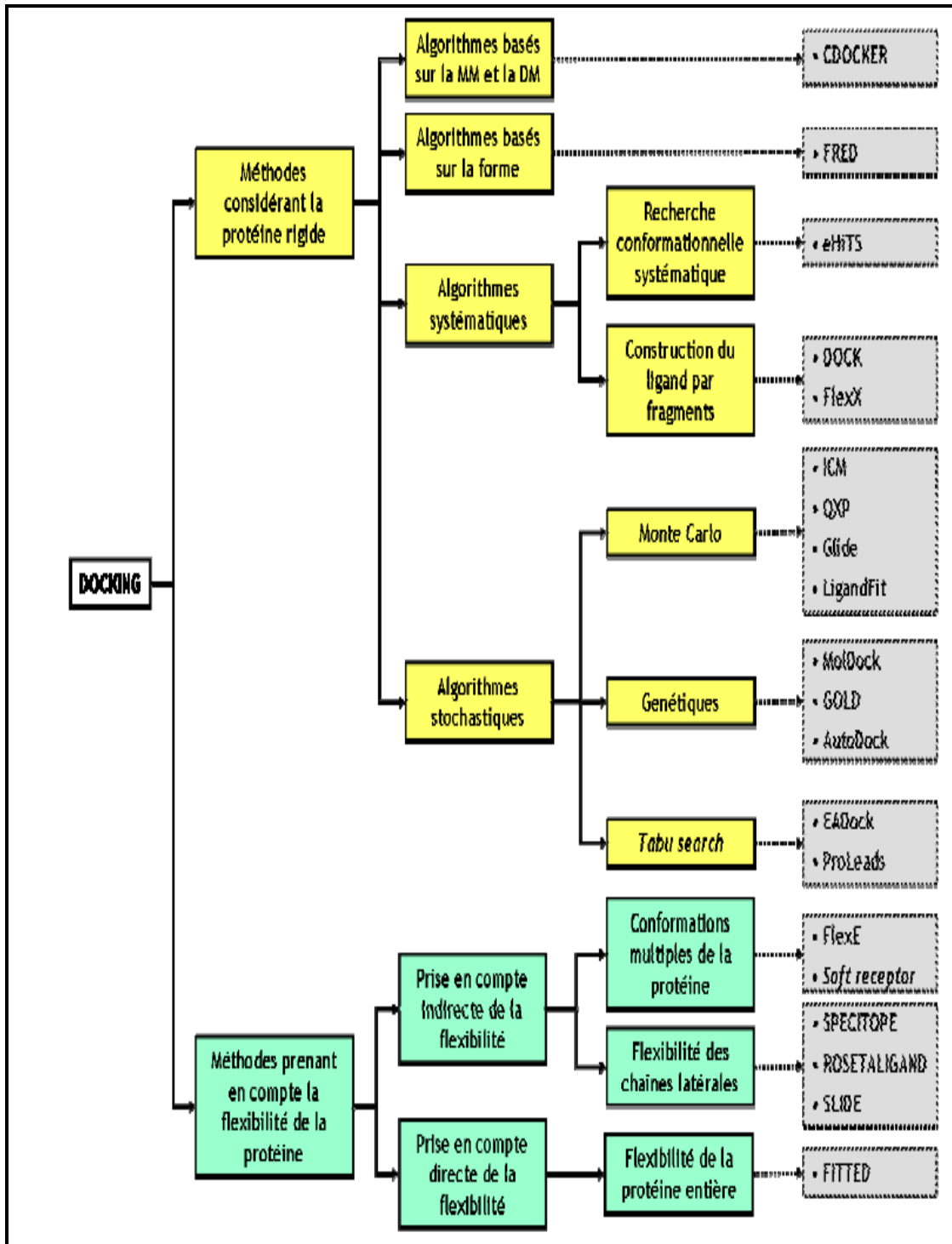


Figure 3.4 Méthodes automatisées de *docking* les plus connues et exemples d'algorithmes (en gris) fréquemment cités dans la littérature. [74]

CHAPITRE 4

Matériel et Méthode

Ces dernières années, un certain nombre d'articles explorant la performance des programmes de *docking* dans la recherche des bases de données ont été publiés [75] [53] [76, 77]. Ces rapports enquêtaient sur l'exactitude des fonctions de score, après que l'amarrage soit terminé, faisant ainsi l'hypothèse sous-jacente que la procédure de *docking* a réussie à identifier la pose expérimentale avec précision et si oui ou non un programme de *docking* aura atteint les taux de succès d'une méthodologie de criblage *in silico*. Toutefois, avant de répondre à la question ci-dessus, on a besoin de connaître à quel point est la précision des programmes de *docking* les plus connus dans la découverte des solutions expérimentales pour les complexes cible-ligand d'une manière compréhensive et comparative. Ainsi, une étude comme référence comparant les forces et les limites des programmes d'amarrage est toujours manquante.

Par conséquent dans cette étude, nous avons tenté d'évaluer la performance de deux programmes de *docking* GOLD 5.0.1 et FlexX 1.3.0, ce sont les dernières versions disponibles jusqu'au début de notre recherche, le premier est celui du Janvier 2011 et le deuxième est celui d'Octobre 2010. Ces deux programmes sont accessibles au public et ont été obtenus par nous sous deux licences académiques.

1. Matériel

1.1. Ordinateur

L'ordinateur utilisé pour cette étude est un Pentium 3, avec une vitesse de 1.10GHZ et une capacité de 376 Mo de RAM.

1.2. Programmes utilisés

1.2.1. GOLD 5.0.1

GOLD (*Genetic Optimisation For Ligand Docking*) est l'un des plus réussis et largement utilisés des programmes de docking [78]. Un opérateur basé sur l'algorithme génétique (GA) est utilisé pour l'échantillonnage des conformations favorables des complexes protéine-ligand guidé par l'une des nombreuses fonctions de *scoring* disponibles.

GOLD utilise un modèle d'île et applique une technique de niche pour maintenir une certaine diversité au sein des populations. Le nombre d'opérations GA, ainsi que la mutation, le croisement et les probabilités de migration sont déterminés automatiquement par une heuristique qui est basée sur la structure du ligand à amarrer et la polarité du site de liaison de la protéine [78]. GOLD utilise un modèle unique basé sur le point de fixation pour tenter de créer des solutions qui contiennent des produits des interactions chimiques plausibles des complexes protéine-ligand, à partir des structures de départ et au cours du processus de *docking*. (cf. Figure 4.1)

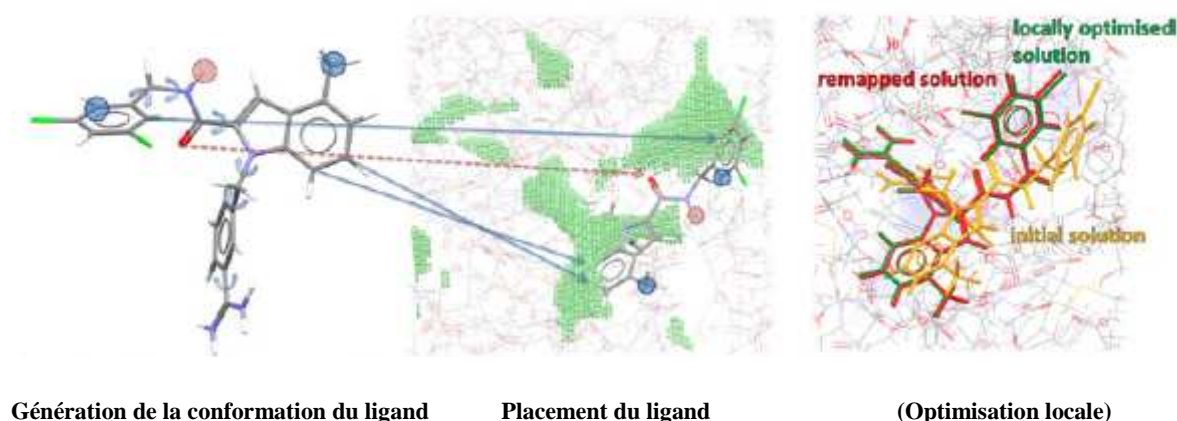


Figure 4.1 Illustration de la construction de solution basée sur le point de fixation. Une conformation du ligand est construite à partir des angles de torsion internes (petites flèches) et placée dans le site de liaison à l'aide de la tâche mise en évidence par le solide (liaison hydrophobe) et les flèches en pointillés (liaison hydrogène). Les cercles représentent les atomes non attribués. L'approche proposée dans ce travail peut en outre appliquer une étape d'optimisation locale continue et redéfinit la solution à l'espace d'apport discret.

-Attribution des points de fixation et génération de la conformation :

Les sites de liaison hydrogène sont utilisés pour attribuer les points de fixation du ligand à la protéine. Pour chaque point donneur dans le ligand, un point accepteur est attribué au hasard. Le point accepteur est soit un site pair seul approprié sur la protéine ou un «*dummy*» site ou un site «model». Les *dummy* sites indiquent qu'un donneur donné ne doit pas être utilisé dans le calcul initial de la matrice de rotation et de traduction du placement du ligand dans le site de liaison. Une approche similaire est également utilisée pour l'attribution des points accepteurs dans le ligand aux points donneurs dans la protéine, et pour les points hydrophobes dans le ligand aux points hydrophobes dans la protéine. Une partie du chromosome de GOLD code les angles de torsion discrets pour toutes les liaisons rotatives dans la structure du ligand et de la protéine. Avant de placer le ligand dans le site de liaison, les conformations de la protéine et du ligand à la fois sont construites à partir de ces valeurs.

-Placement du Ligand et Score :

En utilisant les paires de points attribués, «*Procruste rotation* » [79] ou rotation procuste est utilisée pour générer une transformation qui cartographie la conformation dans le site de liaison tels que les points de fixation se chevauchent au meilleur degré. Suivant cette première étape, la transformation est recalculée en utilisant uniquement les paires de points qui se trouvent à 2°A l'une de l'autre après la première rotation. Cette étape secondaire fait focaliser l'objet de l'algorithme sur l'optimisation des meilleures interactions dans le site de liaison. Enfin, l'affinité de liaison du complexe protéine-ligand généré est appréciée par l'une des nombreuses fonctions de score.

Quand l'algorithme génétique aura fini, la meilleure solution trouvée est localement optimisée utilisant l'algorithme de Nelder Mead Simplex. Comme indiqué plus haut, GOLD peut utiliser différentes fonctions de score pour le processus d'optimisation. L'approche a été initialement conçue pour une utilisation avec GoldScore, une fonction de score basée sur un champ de force. Les points de fixation ont été conçus pour faire correspondre les géométries favorables des liaisons hydrogènes de cette fonction de score et, par conséquent, il est possible que l'approche ne réussisse pas aussi bien avec d'autres fonctions de score.

La fonction de score Goldscore est une fonction de type mécanique moléculaire qui permet de classer les différents modes de liaison ; elle est exprimée par quatre termes:

$$\text{Goldscore ou GOLD Fitness} = S_{\text{hb-ext}} + S_{\text{vdw-ext}} + S_{\text{hb-int}} + S_{\text{vdw-int}}$$

Où : $S_{\text{hb-ext}}$: est le score de la liaison hydrogène du complexe protéine-ligand.

$S_{\text{vdw-ext}}$: est le score de la liaison Van Der Waals du complexe protéine-ligand.

$S_{\text{hb-int}}$: est la contribution à la *fitness* en raison des liaisons hydrogènes intramoléculaires dans le ligand ; ce terme est nul dans tous les calculs présentés dans ce travail (par défaut GOLD, donne généralement le meilleur des résultats)

$S_{\text{vdw-int}}$: est la contribution due à la tension intramoléculaire dans le ligand

1.2.2 FlexX 1.3.0

FlexX est un programme informatique pour prédire les interactions protéine-ligand. Pour une protéine donnée et un ligand, FlexX prédit la géométrie du complexe ainsi qu'une estimation de la force de liaison, la protéine est supposée être rigide. Ainsi, la protéine doit être donnée dans une conformation qui est similaire à l'état lié. L'algorithme de *docking* dans FlexX fonctionne sans intervention manuelle. Néanmoins, dans certains cas plus d'informations sur le ligand ou même le complexe peuvent être introduites. On peut intégrer cette connaissance dans les calculs avec FlexX en effectuant des étapes simples manuellement.

Depuis 1993, il était en cours d'élaboration à l'Institut pour les algorithmes et le calcul scientifique (SCAI) [63] avec des partenaires de coopération différents de l'industrie pharmaceutique. BioSolveIT GmbH (Sankt Augustin) ont fait les marchés du logiciel à travers le monde. Scientifiquement, FlexX est actuellement une partie de LeadIT et est développé à trois endroits, à BioSolveIT GmbH (Sankt Augustin), à la ZBH de l'Université de Hambourg (Hambourg) et à l'Institut Max-Planck for Computer Science (Sarrebruck).

La fonction de score de FlexX est la fonction développée par Böhm pour le programme de conception de novo LUDI [64] avec quelques modifications mineures :

$$\Delta G = \Delta G_0 + \Delta G_{\text{rot}} \times N_{\text{rot}} \quad (1)$$

$$+ \Delta G_{\text{hb}} \sum_{\text{neutral H-bonds}} f(\Delta R, \Delta \alpha) \quad (2)$$

$$+ \Delta G_{\text{io}} \sum_{\text{ionic int.}} f(\Delta R, \Delta \alpha) \quad (3)$$

$$+ \Delta G_{\text{aro}} \sum_{\text{aro. int.}} f(\Delta R, \Delta \alpha) \quad (4)$$

$$+ \Delta G_{\text{lipo}} \sum_{\text{lipo.cont.}} f^*(\Delta R) \quad (5)$$

La fonction peut être divisée en trois parties :

- La première partie (1) se compose d'un terme ΔG_0 qui est fixe et le terme $\Delta G_{\text{rot}} \times N_{\text{rot}}$ prenant compte de la perte d'entropie lors de la liaison du ligand.
- La deuxième partie (2-4) contient les contributions pour les groupes d'interaction appariés, comme les liaisons hydrogènes (2), ponts salins et les liaisons hydrogènes chargées (3), et les interactions aromatiques (4). Chacun de ces termes se compose d'une contribution fixée par interaction (ΔG_{hb} , par exemple) multiplié par une fonction de pénalité $f(\Delta R, \Delta \alpha)$. Les fonctions de pénalité sont des fonctions linéaires par morceaux; mises à l'échelle de la contribution d'une interaction avec le respect de sa géométrie.
- La troisième partie (5) représente les taux des contacts atome-atome entre la protéine et le ligand comme les contacts hydrophobes et redoutablement les contacts proches (affrontements).

La deuxième et la troisième partie de la fonction de score sont appelées, respectivement, le match score et contact score.

1.2.3. Open Babel 2.0.2

Le logiciel libre OpenBabel est un système expert principalement utilisé dans des applications de chemo-informatique pour la conversion de fichiers de structures chimiques. [80]

L'objectif du projet scientifique Open Babel est d'offrir aux utilisateurs et aux développeurs un ensemble de programmes et de bibliothèques logicielles multi plate-forme conçus pour la modélisation moléculaire, la chimie et de nombreux domaines connexes, tels que la

conversion de format de fichiers et de données. Ce dernier est obtenu à partir de cette adresse électronique : <http://openbabel.org>.

1.3. La Protein Data Bank ou PDB

La Protein Data Bank ou PDB [9] est une banque mondiale de données de structures tridimensionnelles de macromolécules biologiques, essentiellement des protéines et des acides nucléiques. Ces structures sont principalement déterminées par les deux méthodes la cristallographie à rayons X et la RMN. Les structuralistes du monde entier sont tenus d'y déposer leurs données expérimentales, de cette façon elles appartiennent au domaine public. La consultation des structures est gratuite et peut se faire *via* l'internet.

Son adresse électronique est : <http://www.rcsb.org/pdb>.

2. Méthode

Il est souvent difficile de juger quel est le meilleur programme de *docking* pour une protéine cible donnée. Cela a mené à un intérêt de l'évaluation de la performance des deux programmes de *docking* GOLD et FlexX par rapport à un grand ensemble de complexes protéine-ligand.

Nous avons téléchargé à partir de la PDB 153 complexes protéine-ligand de 3 cibles différentes : l'hémagglutinine (HA) ; la neuraminidase (NA) et la peptide deformylase (PDF), ces complexes sont de différentes tailles : grandes ; moyennes et petites.

2.1. Les ensembles de test

Selon Cole *et al.* [81], l'ensemble de test idéal ne doit pas être biaisé vers une famille de protéines donnée. Au lieu de cela, il doit être suffisamment grand pour enjamber les complexes représentant une haute résolution, organisé pour enlever ou réparer des erreurs telles que des affrontements stériques. Les complexes avec des résidus manquants ou des liaisons covalentes doivent être évités. Les données de liaison (comme les K_D ou IC_{50}) devraient être disponibles pour chaque complexe

Pour une comparaison équitable, tous les programmes devraient avoir une définition comparable de l'espace de recherche. Les définitions de la région d'intérêt explorée par les différents programmes doivent être similaires. Si elle est trop petite, aucune conclusion claire ne peut être tirée à propos de la performance de l'heuristique d'échantillonnage [81], et les

erreurs de la fonction de score peuvent aussi être cachées, comme un échantillonnage plus approfondi aurait pu conduire à des modes de liaison avec un score plus favorable. Tous les programmes devraient faire face à un ensemble similaire de degrés de liberté, tel que les rotations, les angles diédraux, les longueurs de liaison et les angles de liaison. La même information structurale telle que la présence de molécules d'eau ou des états de protonation spécifiques, le cas échéant, devrait être utilisée par tous les programmes quand il faut en tenir compte. Tous les programmes devraient être comparés dans leur dernière version disponible et le nombre de poses évalué et le temps CPU ou *Central Processing Unit Time* devrait être considéré.

2.2. La procédure GOLD

GOLD est un programme pour le calcul des modes de *docking* de petites molécules dans les sites de liaison des protéines et est fourni dans le cadre de GOLD Suite, un ensemble de programmes pour la visualisation et la manipulation de structure (Hermes), pour l'amarrage protéine-ligand (GOLD) et pour le post-traitement (GoldMine) et la visualisation des résultats de docking. Hermes agit comme un centre pour de nombreux produits de *Cambridge Crystallographic Data Centre* (CCDC) [82].

GOLD est caractérisé par le 'wizard' qui regroupe un ensemble d'étapes de *docking* complètes.

2.2.1. Initialisation des protéines et des ligands

Après avoir téléchargé le complexe protéine-ligand sous format .pdb dans GOLD ; les molécules d'eau et les ions sont éliminés et les atomes d'hydrogène sont ajoutés, et incluant ceux nécessaires pour définir l'ionisation correcte et l'état tautomérique des résidus tels que l'acide aspartique (Asp), l'acide glutamique (Glu) et l'histidine (His).

Une fois que tous les types d'atomes et tous les types de liaisons sont corrects ; les atomes d'hydrogène sont placés sur les atomes corrigés ; les angles de torsion de la sérine (Ser) ; la thréonine (Thr), et groupements hydroxyles de la tyrosine (Tyr) seront optimisés par GOLD de sorte que leurs positions n'ont pas d'importance.

Spécifiquement, chaque OH de la Ser, Thr et Tyr sera autorisé à tourner pour optimiser ses liaisons hydrogènes au ligand ; les groupements NH₃⁺ de la Lysine sont similairement optimisés à moins qu'ils soient fixés par des liaisons H à des résidus voisins de la protéine.

La description 2D du ligand est convertie en structure 3D, Tous les atomes d'hydrogène sont ajoutés pour s'assurer que toutes les valences des atomes lourds sont satisfaites. GOLD ne tient pas compte des charges d'atomes, à la fois formelles et partielles. Par conséquent, nous n'avons pas besoin de nous soucier de la charge des ligands. Enfin, nous sauvegardons les ligands en tant que fichier MOL2, le format sdf du ligand est aussi toléré.

2.2.2. Détermination du site de liaison du ligand

Cela peut être effectué de plusieurs manières, dans notre travail on a choisi de déterminer le site de liaison par le ligand de référence à 5Å°.

2.2.3. La flexibilité du ligand

Lors de l'initialisation du ligand, un groupement amide du ligand sera mis à la conformation trans. En fait, la liaison amide dans le ligand semble basculer entre la forme plane-cis et plane-trans dans la liaison à la protéine. Par conséquent, nous considérons que la liaison amide peut être tournée. Nous considérons aussi un tournement de coin si le ligand a un cycle non-aromatique.

2.2.4. La flexibilité de la protéine

Durant le processus du *docking* par GOLD, la protéine est maintenue rigide, cette simplification permet d'effectuer le *docking* de manière plus efficace ce qui est essentiellement crucial lors du criblage des bases de données.

2.2.5. Arrangement des contraintes

Après avoir fait une série d'opérations de *docking*, on peut trouver qu'un certain atome d'hydrogène dans la liaison hydrogène de la protéine peut former une liaison hydrogène avec le ligand. Par conséquent, nous précisons qu'un atome particulier de la protéine devrait être lié par liaison hydrogène au ligand. Ainsi, GOLD est influencé en vue de trouver une solution dans laquelle la contrainte spécifiée est satisfaite. Par conséquent, nous pouvons optimiser la solution GOLD.

2.2.6. Docking avec GOLD

Les paramètres de GA ont une incidence directe sur le temps d'exécution et la probabilité de trouver l'optimum global. Les principaux paramètres qui affectent le temps d'exécution (CPU)

et la précision sont le nombre de *docking* et le nombre d'opérations de GA dans chaque opération de *docking*. Selon l'étude de f. Jones et al, les résultats sont obtenus avec 2, 5, 10, et 20 GA, et les auteurs cités concluent que GOLD exige généralement moins de 20 GA pour reproduire un mode de liaison. Dans notre étude, nous avons choisi d'exécuter le docking par GOLD avec un résultat de 10 GA, les autres paramètres sont résumés dans le tableau 4.1:

Tableau 4.1 Paramètres pris en compte dans le *docking* par GOLD

Number of Operations	Population Size	Selection Pressure	Number of Islands	Crossover	Mutate	Migrate	Niche Size	Hydrogen Bonding	Van der Waals
100000	100	1.1	5	95	95	10	2	4	2.5

Dans GOLD, le ligand est d'abord placé dans le site de liaison de la protéine sur la base des points de fixation; puis, un GA est utilisé pour explorer l'espace conformationnel des ligands sur la base de *fitness score*. La population initiale des poses possibles du ligand est mise en place au hasard. Chaque membre de la population est codé comme un chromosome, qui contient des informations sur la liaison entre la protéine et le ligand.

2.3. La procédure FlexX

Avant de décrire les résultats obtenus il est nécessaire de donner la démarche de la préparation des données et ceci dans le GUI mode.

2.3.1. Le ligand

Les étapes suivantes sont celles du prétraitement du ligand pour le *docking*. Tout d'abord, le ligand est extrait du fichier pdb et transformé en format de fichier SYBYL mol2. Les types d'atomes corrigés y compris les états d'hybridation [83] ainsi que les types de liaison correctes sont définis. Les atomes d'hydrogène sont ajoutés à tous les atomes avec des géométries raisonnables, des charges formelles sont attribuées à chaque atome. Le ligand est centré et une minimisation de l'énergie est réalisée avec le champ de force Tripos [83]. Le format .sdf du ligand est aussi accepté en GUI mode

FlexX analyse la structure du ligand et détecte les symétries topologiques locales à des liaisons simples dont l'angle de torsion peut être changé en moins de 360 ° de telle sorte que

la même conformation en résulte, la computation de la déviation rms considère aussi cette symétrie locale.

2.3.2. Le récepteur

Dans le GUI mode, le récepteur ou la protéine est téléchargée sous format pdb, le complexe protéine-ligand est téléchargé tel qu'il est dans FlexX puis la chaîne dans laquelle se trouve le ligand à amarrer et les cofacteurs qui entrent dans le processus du *docking* sont ensuite déterminés par l'utilisateur. Comme dans GOLD les molécules d'eau peuvent ou ne pas être prises en considération. Dans notre travail les molécules d'eau ont été éliminées.

2.3.3. Définir le site de liaison du ligand

Lorsque tout le complexe protéine-ligand est utilisé pour définir la protéine ; le site de liaison peut être défini par le ligand de référence. Dans notre cas tous les atomes de la protéine avec une distance de 6.5\AA de l'atome du ligand connu jusqu'à sa position cristalline sont utilisés pour définir le site de liaison.

2.3.4. Le docking avec FlexX

Depuis la publication de l'algorithme complet de *docking* [84], la préparation des données d'entrée a été modifiée et adaptée au futur développement de FlexX. Par conséquent, nous résumons l'ensemble du processus à nouveau. Les fichiers contenant le ligand d'entrée ont été générés avec SYBYL [83]. D'abord, la structure du ligand contenant uniquement des atomes non hydrogène est extraite du fichier pdb. Il s'agit de la structure de référence qui est utilisée pour le calcul des valeurs RMSD plus tard. Par la définition des types d'atomes correctes dont les Etats d'hybridation, ainsi que les types de liaisons correctes, en ajoutant des atomes d'hydrogène, affectation des charges formelles de chaque atome, et enfin la minimisation de l'énergie de la structure, nous obtenons les fichiers d'entrée ligand. La minimisation de l'énergie garantit une conformation de basse énergie avec des distances de liaison appropriées et angles. En principe elle garantit également qu'il y a une toute nouvelle géométrie avec aucune information de *docking* de la structure pdb. En règle générale, tous les groupes d'acides carboxyliques et les acides phosphoriques ont été ionisés, et tous les groupements amines, amidines, et guanidines, mais pas de groupements amide, ont été protonés.

Les données d'entrée géométriques des atomes de récepteur qui sont prises à partir des fichiers pdb sont traitées rigide lors de la procédure de *docking*. La préparation des données d'entrée du récepteur exige la définition des atomes du récepteur (*via* des identifiants de la chaîne et les groupes hétéro), la résolution des ambiguïtés dans le fichier pdb (les indicateurs d'emplacement en alternance, etc), la détermination des positions des atomes d'hydrogène essentielles, et la définition des atomes du site actif. Les attributions des positions d'hydrogène sont faites sur la base de règles par défaut, sauf pour la définition des angles de torsion sur les groupes hydroxyles des acides aminés sérine, la thréonine, la tyrosine et la position d'hydrogène à l'intérieur de la chaîne latérale histidine. Ici, les angles de torsion convenable et l'état tautomérique optimal de l'histidine, respectivement, ont été sélectionnés par une inspection visuelle de la protéine. Les résidus de lysine et de l'arginine sont protonés et les groupements acides d'acide aspartique de et l'acide glutamique sont ionisés.

2.4. Paramètres de comparaison étudiés

Pour essayer d'effectuer une comparaison entre GOLD 5.0.1 et FlexX 1.3.0 plusieurs paramètres ont été pris en considération :

2.4.1. RMSD ou Root Mean Square Deviation

C'est la capacité d'un programme de reproduire les modes de liaison expérimentales d'un ligand et pour cela, un ligand dans sa structure obtenu par cristallographie aux rayons X est amarré dans son site de liaison puis le mode de liaison amarré est comparé à celui obtenu expérimentalement et un RMSD entre les deux est calculé. Un bon RMSD est celui inférieur ou égal à 2 Å.

2.4.2. Les liaisons rotables

Plus le nombre de liaisons rotables est petit plus les résultats sont bons parce que le ligand n'admettra pas beaucoup de conformations.

2.4.3. La vitesse

Le temps CPU ou le temps d'exécution du programme de *docking* est très important surtout lors du criblage des milliers de chimiothèques.

2.4.4. Les faux positifs

On trouve parfois parmi les meilleures solutions générées par les programmes de *docking* des résultats où il y a un large RMSD avec un large score ce sont les faux positifs, la performance du logiciel peut être évaluée en calculant le pourcentage de ces derniers.

CHAPITRE 5

Résultats et Discussion

Depuis que chaque outil de *docking* nécessite de combiner un moteur de *docking* avec une fonction de score rapide, la littérature récente est pleine de repères répondant à trois questions possibles : la capacité d'un algorithme de *docking* de reproduire la pose cristallographique des ligands de petit poids moléculaire sélectionnés [55, 56] [62] [69] [70]; la tendance des fonctions de score rapides à prédire l'énergie libre de liaison à partir de la meilleure pose classée [52] [4] [70] [75, 76, 77]; la discrimination des liants connus à partir des molécules choisies au hasard dans des expériences de criblage virtuel [53] [55] [75] [77].

L'objectif de l'étude en cours est de fournir des repères indépendants pour les programmes de *docking* largement utilisés. Comme il serait presque impossible de les examiner tous, les programmes de *docking* ont été sélectionnés en fonction de trois critères [85] :

- La disponibilité.
- L'utilisation des formats de fichiers conventionnels comme entrée (exemple pdb, sdf, mol2).
- Une application facile pour le criblage virtuel.

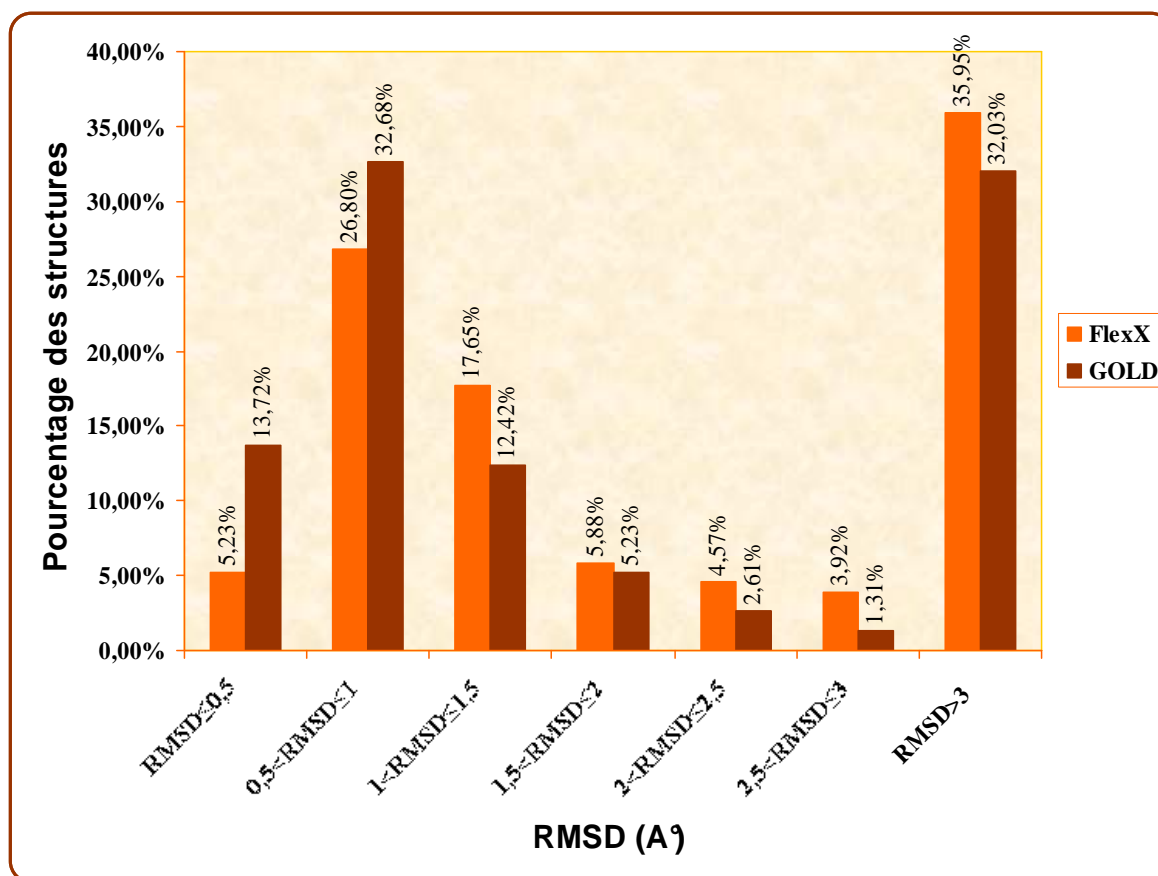
1. La précision du processus du *docking*

La caractéristique clé d'un bon programme de *docking* est sa capacité à reproduire les modes de liaison expérimentaux des ligands [86]. Pour tester cela, un ligand est extrait de la structure cristallographique aux rayons X de son complexe protéine-ligand et amarré dans son site de liaison. Le mode de liaison amarré est ensuite comparé au mode de liaison expérimentale, et l'écart quadratique moyen ou Root Mean Square Deviation (RMSD) entre les deux est calculé; une prédiction d'un mode de liaison est considérée réussie si le RMSD est au dessous d'une certaine valeur (généralement 2,0 Å) [86]. Récemment, Nissink *et al.* soulignaient que pour établir le taux de réussite d'un programme de *docking*, un grand ensemble des complexes protéine-ligand est nécessaire et doivent être soigneusement construits [87].

Plusieurs intervalles de RMSD ont été utilisés : $\leq 0,5\text{Å}$; $\leq 1,0\text{Å}$; $\leq 1,5\text{Å}$; $\leq 2,0\text{Å}$; $\leq 2,5\text{Å}$; $\leq 3,0\text{Å}$ et $>3,0\text{Å}$, dans cette étude la prédiction des modes de liaison dans les complexes protéine-ligand est classée [88] selon le critère suivant: bonne ($\text{RMSD} \leq 2,0\text{Å}$), acceptable ($\text{RMSD} > 2\text{Å}$ et $\leq 3,0\text{Å}$), imprécise lorsque la solution du docking est loin du site actif ou dans une position inversée ou incorrecte ($\text{RMSD} > 3\text{Å}$).

Le Graphe 5.1 montre l'évaluation des programmes de *docking* pour leur capacité de prédire le mode de liaison correcte d'un ligand donné dans le site actif de sa protéine cible. Les pourcentages représentent la meilleure pose selon le RMSD pour chaque ligand amarré. .

Grphe 5.1 La performance des deux programmes de *docking* selon la meilleure pose générée



La majorité des bonnes RMSD se situe dans l'intervalle] 0,5Å-1,0Å] pour GOLD et FlexX.

Nous remarquons d'après ces résultats que le programme GOLD reproduit bien les données expérimentales, et à un degré moindre FlexX. En effet, 64,05% des valeurs de RMSD sont inférieures à 2Å pour le premier, 55,56% pour le deuxième.

Notre résultat confirme des résultats obtenus par Zaheer *et al.*, en 2010 [88] où six programmes de docking ont été utilisés : FRED, GOLD, MOE, AutoDock, Surflex-Dock et FlexX pour une étude comparative afin de déterminer leur capacité de reproduire les poses expérimentales *via* le RMSD en utilisant 26 complexes de l'Acetyl cholinestérase, FRED était le meilleur suivi de Surflex-Dock et GOLD, les autres programmes comme FlexX, MOE et AutoDock ont montré une performance un peu moindre dans la génération des poses.

Michael *et al.*, [89] ont évalué aussi dans la même année la performance des quatre programmes GOLD, AutoDock, Surflex-Dock et FRED par le calcul du RMSD en utilisant les inhibiteurs du sarco endoplasmic reticulum calcium ATPase, les meilleurs résultats ont été obtenus par GOLD et FRED.

Les résultats du *docking* par GOLD sont illustrés par les trois complexes (à gauche) : 3CKZ (le plus petit RMSD), 1BSZ (RMSD moyen), 2ZR1 (le plus mauvais RMSD), dans les figures ci-dessous. La meilleure conformation du ligand générée par le *docking* (mauve) est superposée à la pose expérimentale du ligand naturel (vert).

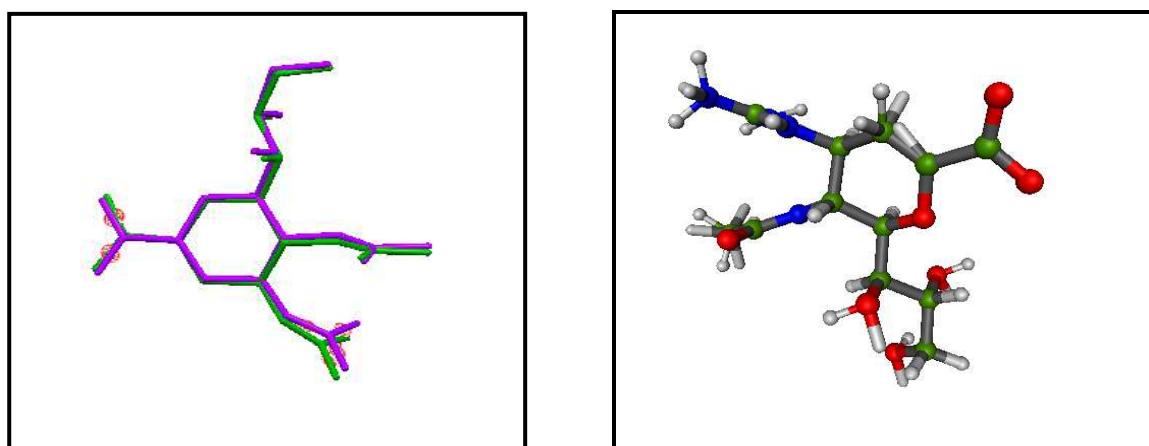


Figure 5.1 Le complexe 3CKZ

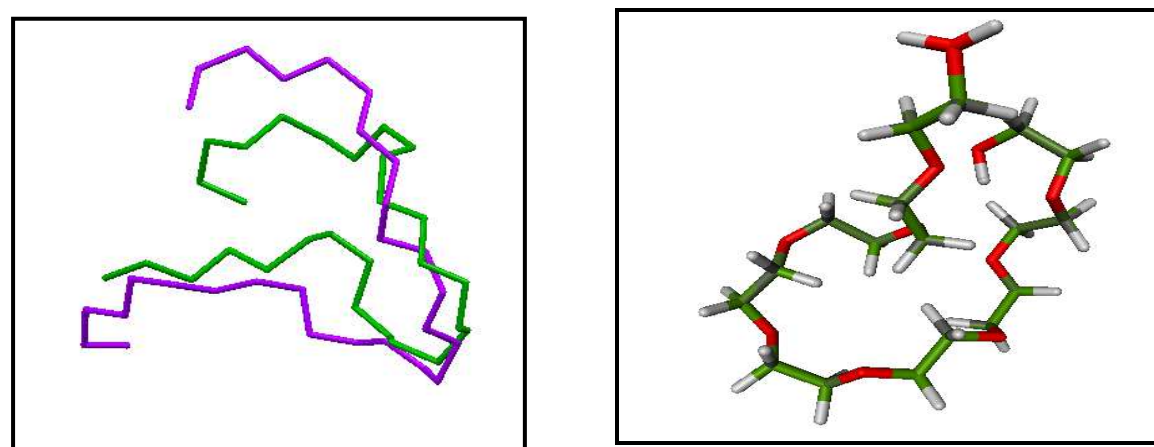


Figure 5.2 Le complexe 1BSZ

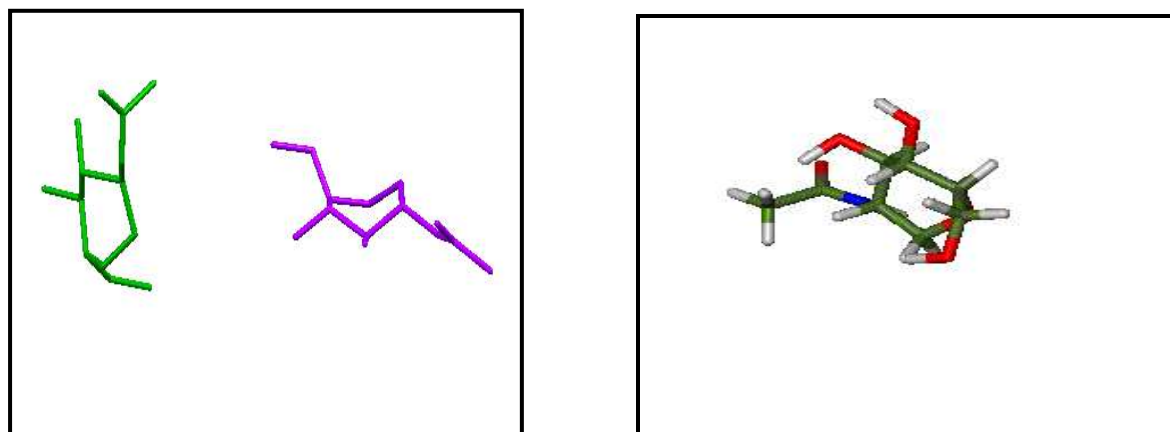


Figure 5.3 Le complexe 2ZR1

Dans le complexe 3CKZ avec un RMSD de 0.2337 Å, la superposition est presque parfaite; dans le complexe 1BSZ avec un RMSD de 3.7690 Å, elle l'est moins; alors que dans le troisième cas avec le complexe 2ZR1 avec un RMSD de 9.8637 Å elle ne l'est pas du tout.

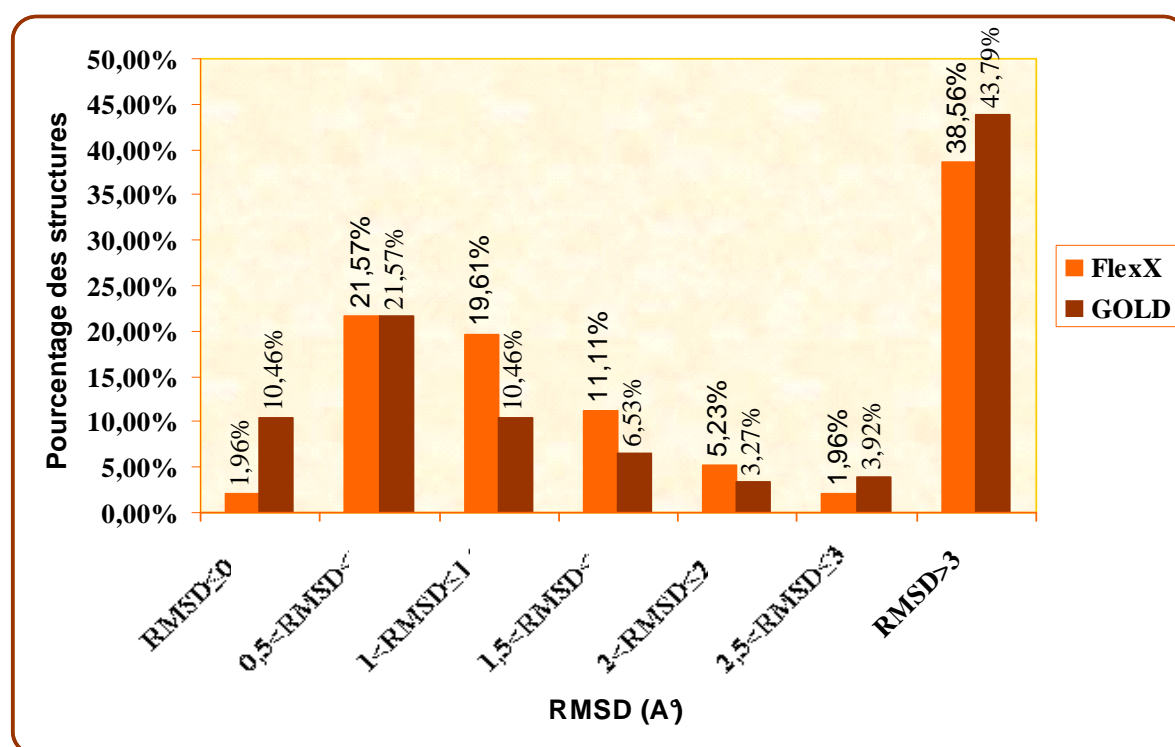
A droite, les résultats du *docking* des mêmes complexes par FlexX avec les RMSD suivants : le complexe 3CKZ avec un RMSD de 0,5837Å, le complexe 1BSZ avec un RMSD de 5,1791Å et le complexe 2ZR1 avec un RMSD de 4,5561Å, ceci pour mettre en évidence la différence entre les deux programmes dans la prédiction des modes de liaison des complexes protéine-ligand mais aussi pour faire une inspection visuelle sur la superposition de ces complexes pour les deux programmes de *docking*.

Un certain nombre d'évaluations comparatives des programmes de *docking*, menées au cours des dernières années, indiquant que les programmes de *docking* sont généralement en mesure de générer les poses du ligand qui sont similaires à celle déterminée expérimentalement ont été publiées [90]. Ces études montrent que la bonne performance de reproduction des modes de liaison déterminée expérimentalement ne transmet pas le succès dans le criblage virtuel et la corrélation entre les scores de *docking* et l'affinité du ligand est encore mise en question [90].

2. La précision dans le classement des solutions

La performance des deux programmes GOLD et FlexX au niveau des 153 complexes protéine–ligand étudiés dans le classement de diverses poses générées pour chaque ligand arrimé dans le site actif de la protéine cible est aussi étudiée [85]. Elle est obtenue en comparant la solution classée en premier (top-pose) par les deux programmes et celle déterminée par cristallographie appréciée par le RMSD [91]. (cf. Graphe 5.2)

Graphe 5.2 La performance des deux programmes de *docking* selon la meilleure pose classée.



Malgré qu'il y a égalité de proportions dans la capacité de classement de solutions prédites entre les deux programmes GOLD et FlexX dans l'intervalle de RMSD] 0,5Å-1,0Å] avec une valeur de 21,57% et malgré que GOLD dépasse FlexX avec une valeur de 10,46% dans les RMSD comprises entre 0 et 0,5Å, FlexX a réussi à localiser 54,25% des ligands à des valeurs de RMSD $\leq 2,0\text{\AA}$.

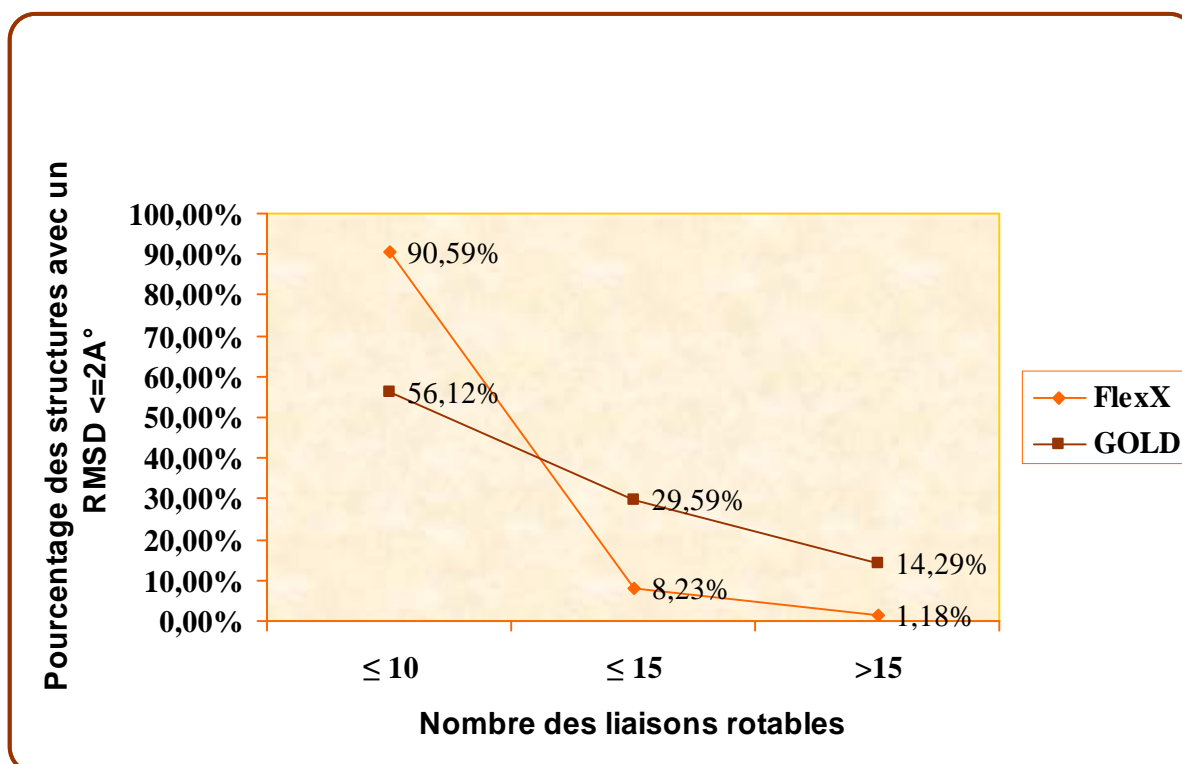
Si on compare FlexX à GOLD, ce dernier n'a pu classer que 49,02% des ligands à des RMSD $\leq 2,0\text{\AA}$, cela nous permet de dire que GOLD n'a pas bien réussi à classer les solutions, résultat confirmé par des études antérieures réalisées par Kontoyianni *et al.*, en 2004 [90] et

une remarque déjà faite par les développeurs de GOLD, Jones *et al.*, en 1997 [69] affirmant que GOLD est plus performant en amarrant qu'en classant.

3. Les liaisons rotables et la performance du processus du *docking*

L'effet des caractéristiques du ligand sur l'exactitude du processus du *docking* est un autre type d'analyse que nous avons menées (cf. Graphe 5.3). Il est bien connu que quand le nombre de liaisons rotables du ligand augmente, la précision du *docking* diminue puisqu'un espace conformationnel beaucoup plus large doit être échantillonné [91]. Les rotations autour de ces liaisons entraînent des déviations à partir de la géométrie idéale. Il en résulte une petite pénalité énergétique due à ces déviations dans les longueurs de liaison et les angles de liaison. [92]

Graphe 5. 3 Les liaisons rotables du ligand en relation avec la précision du processus du *docking*



Les complexes dans cette étude ont été divisés en trois groupes, les ligands avec des liaisons rotables ≤ 10, les ligands avec des liaisons rotables ≤ 15 et ceux avec des liaisons rotables > 15 [91]. Les résultats confirment des travaux déjà réalisés. En effet toutes les procédures de

docking échouent habituellement dans le placement du ligand dans le site actif quand la flexibilité du ligand augmente [88].

En tenant compte de la complexité des ligands, GOLD semble être le moins sensible, en effet FlexX reproduit jusqu'à 90,59% des poses expérimentales lorsque le nombre de liaisons rotatives des ligands est inférieur ou égal à 10, ce pourcentage augmente encore plus et atteint 98,82% lorsque le ligand a 15 liaisons rotatives ou moins. Notre constatation est en accord avec celle de Kramer *et al.*, [49] qui ont déjà fait des analyses précédentes avec FlexX en 1999.

Voici des exemples de deux complexes amarrés par GOLD avec des liaisons rotatives différentes et les RMSD correspondants.

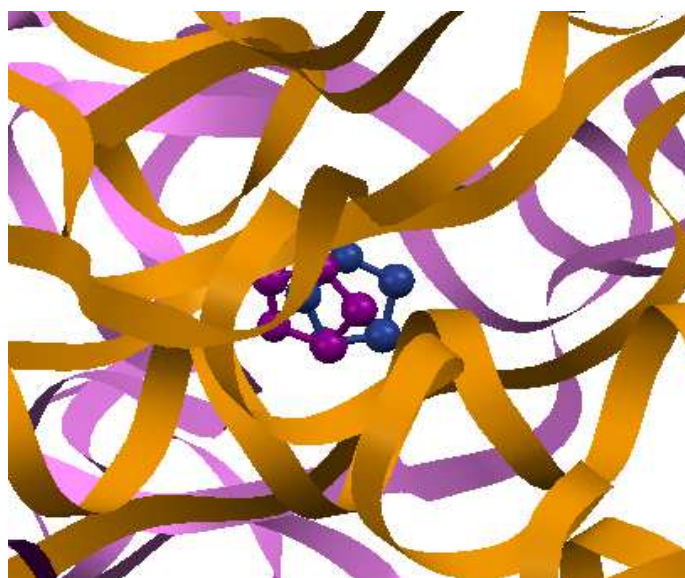


Figure 5.4 Le complexe 3NRB

Le ligand dans le complexe 3NRB a 0 liaison rotative et un RMSD de 1,3692 Å

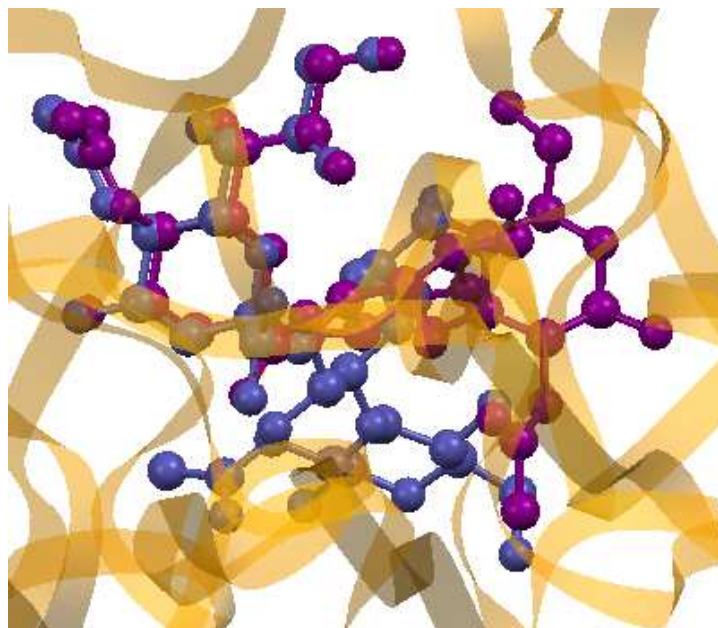


Figure 5.5 Le complexe 1JSN

Le ligand dans le complexe 1JSN possède 24 liaisons rotables et un RMSD de 0,4035 Å

Ces deux exemples sont pour montrer que GOLD n'est pas beaucoup sensible au nombre de liaisons rotables comparé à FlexX.

4. Le temps d'exécution et la performance du processus du *docking*

Nous avons essayé de définir certains paramètres de *docking* pour parvenir à une comparaison équitable des deux programmes de *docking* GOLD et FlexX spécialement en terme de vitesse du processus du *docking*.

L'analyse du temps d'exécution du processus d'arrimage ou CPU time d'après trois intervalles de temps : minimum, moyen et maximum indique clairement des différences significatives (cf. Tableau 5.1)

Tableau 5.1 Temps d'exécution du processus d'arrimage pour les deux programmes GOLD et FlexX

		Minimum	Moyen	Maximum
Temps d'exécution (en secondes)	FlexX	11,88	109,16	323,91
	GOLD	150,0012	2248,0016	12010,0005

D'après les résultats du Tableau 5.1, FlexX est remarquablement le plus rapide du fait qu'il est capable d'amarrer un ligand dans moins d'une minute, tandis que GOLD atteint une distribution remarquable du temps d'exécution (150s-12010s) pourtant il est beaucoup moins sensible que FlexX en terme de la flexibilité du ligand (qui est un facteur essentiel influençant la démarche du *docking* [85]), ces résultats s'accordent bien avec ceux obtenus par Badry *et al.*, en 2003 [93] et Kellenberge *et al.*, en 2004 [85] montrant que GOLD est beaucoup plus lent que FlexX dans le processus de *docking*.

Une utilisation importante des programmes de *docking* protéine-ligand est dans le criblage virtuel, dans lequel de grandes bibliothèques de composés sont amarrées dans le site actif de la protéine cible et l'affinité du ligand pour la protéine est ensuite évaluée par le score. A cet effet, les dockings ont besoin d'être rapide [86]. L'accélération d'un protocole de *docking* se fait souvent par le coût de l'échantillonnage d'un nombre moindre de modes de liaison, et, par conséquent, le taux de réussite est réduit. Il est donc important que les paramètres de recherche soient choisis pour donner des vitesses de *docking* avantageuses dans les applications du criblage virtuel (dans notre cas, jusqu'à 2 min / composé sur un seul processeur), avec une perte acceptable de la précision de *docking*.

Selon les résultats discutés ci-dessus, FlexX est un bon programme destiné pour le criblage virtuel et pour participer dans les processus de découverte de nouvelles molécules bio-actives.

5. Les faux positifs dans les programmes de *docking*

Les fonctions de score sont nécessaires pour deux raisons: pendant le processus de *docking*, elles servent de fonctions de *fitness* dans l'optimisation de l'orientation et la conformation du ligand et, pour la comparaison avec d'autres molécules, elles sont utilisées comme des estimations de l'affinité de liaison pour la molécule complètement amarrée [88]. Ainsi il y a différents critères pour la qualité d'une fonction de score : sa capacité à identifier le bon mode de liaison d'un ligand en dehors des solutions de *docking* alternatives, sa capacité de classer les ligands liés à l'égard de leur affinité de liaison, et sa capacité à sélectionner des inhibiteurs en dehors d'une grande base de données de composés inactifs. Ici nous allons nous concentrer sur le troisième critère, qui est la question centrale dans le criblage virtuel.

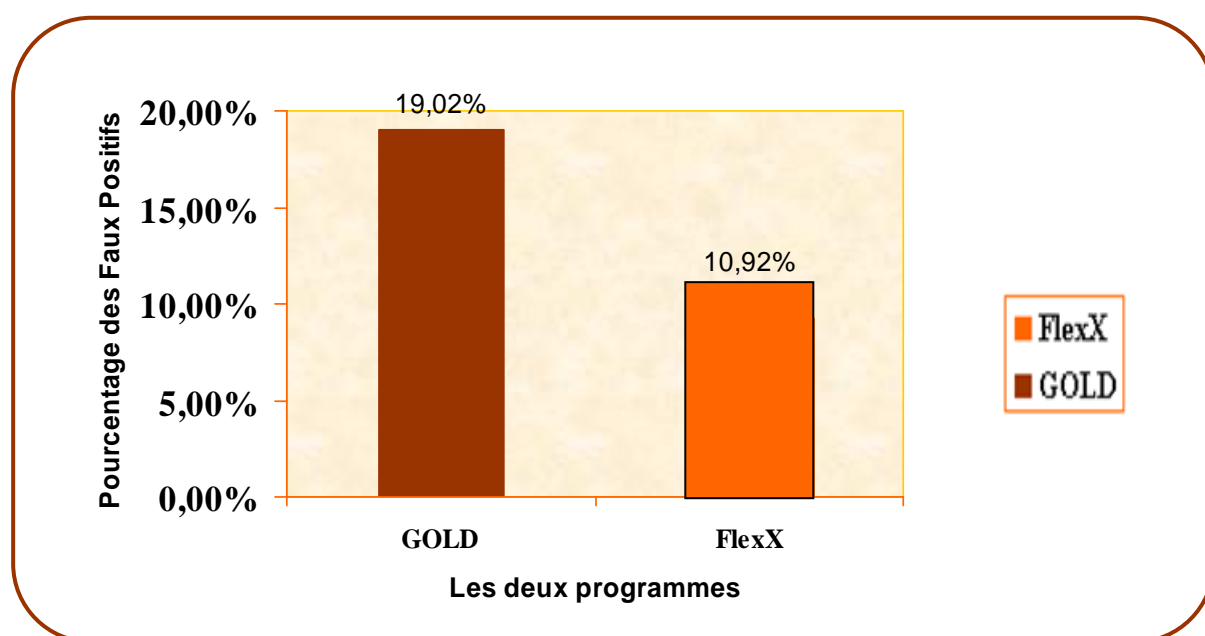
Ce paramètre est analysé en tenant compte seulement des dix meilleures solutions classées par les deux programmes GOLD et FlexX qui peuvent montrer ou pas quelques structures avec de larges valeurs de RMSD. Il n'est cependant pas rare que les solutions presque optimales

soient généralement perdues parmi un ensemble de solutions avec des scores élevés et de grandes valeurs de RMSD ; c'est ce qu'on appelle les faux positifs. [94]

Pour chacun des deux programmes on essaye de trouver de manière approximative dans la meilleure liste de chaque ligand amarré, le pourcentage des solutions ayant un score élevé mais aussi un grand RMSD.

Le Graphe 5.4 montre le pourcentage approximatif des faux positifs obtenu pour les deux programmes.

Graphe 5.4 Pourcentage des Faux Positifs dans les dix meilleures solutions pour l'ensemble des ligands arrimés pour chaque programme.



Les résultats montrent clairement que GOLD produit plus d'erreurs que FlexX dans le classement des dix meilleures poses. Ceci s'explique aisément par la différence entre les deux programmes dans les fonctions de scores.

L'expression de *fitness* score de GOLD est incapable de distinguer entre les complexes obtenus et la structure du complexe dans sa forme native [94], ceci s'explique par le fait que dans certains cas, il a été observé que les ligands ont été placés dans les sites actifs de la protéine, mais n'ont pas adopté la conformation qui leur permet d'être correctement orienté dans le site actif résultant en des valeurs élevées de RMSD, par conséquent la pose résultante a été incorrectement classée [95].

On peut expliquer la notion de faux positifs à travers les deux exemples suivants :

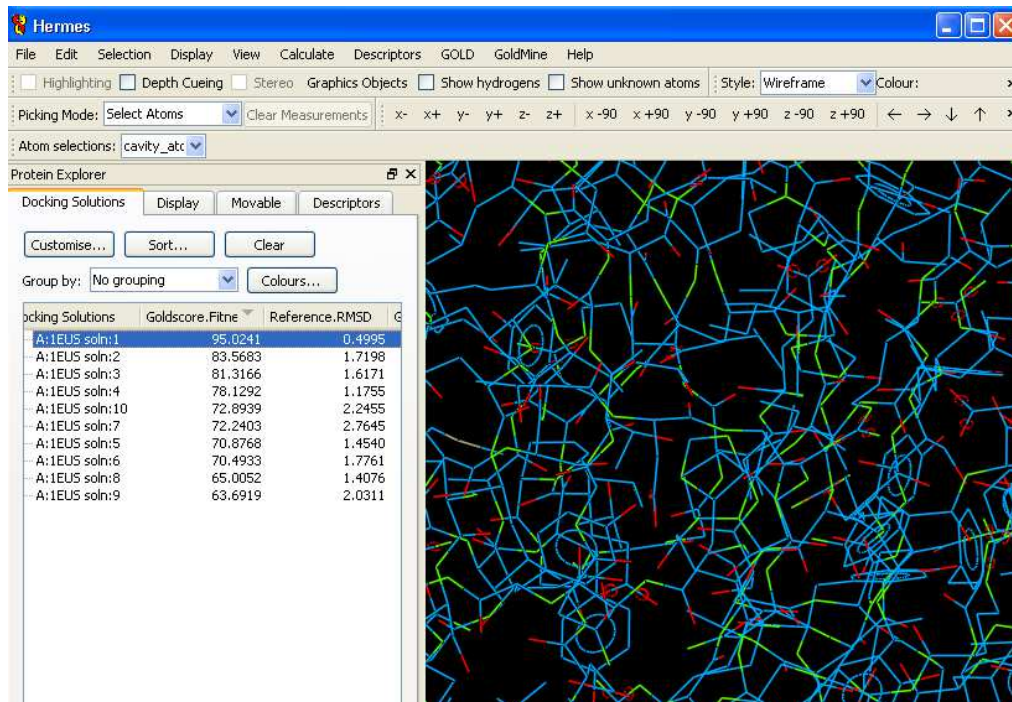


Figure 5.6. Le résultat du *docking* du complexe 1EUS par GOLD

La Figure 5.6 montre que les solutions 10 et 7 peuvent être considérées comme faux positifs et donc donnent le pourcentage de 20%.

La solution 10 et 7 ont une large *fitness* score avec un large RMSD.

Un autre exemple est celui du complexe 2VW0

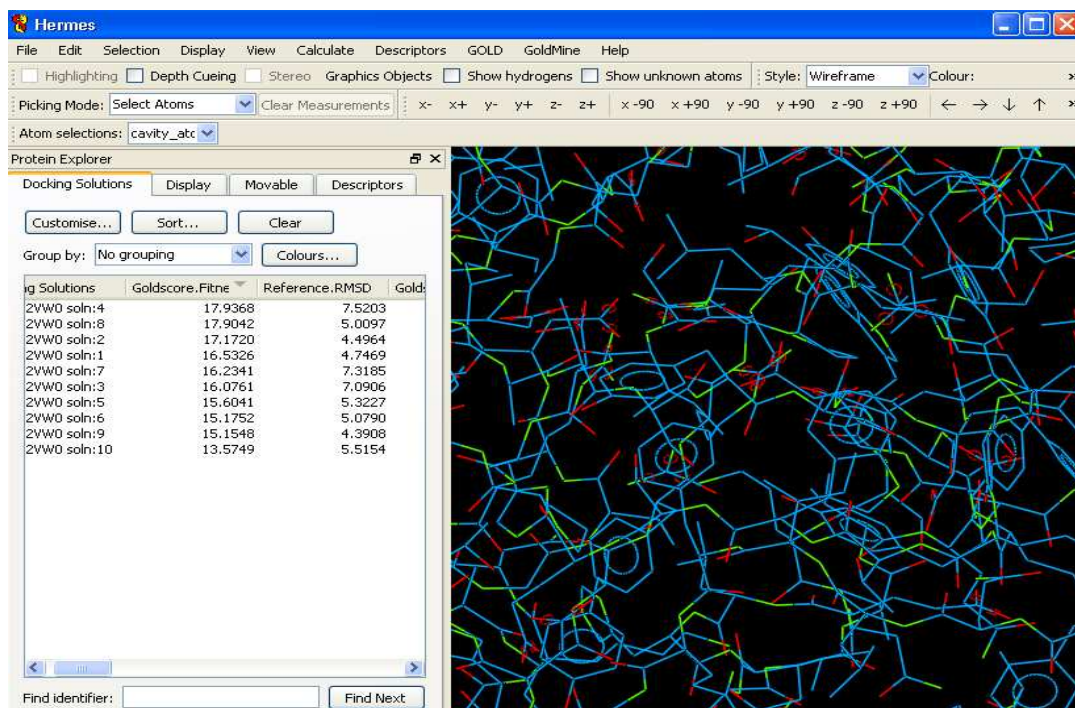


Figure 5.7. Le résultat du *docking* du complexe 2VW0 par GOLD

Dans cet exemple, toutes les solutions ont des petites *fitness* score avec des RMSD extrêmement larges ce qui est évident, dans ce cas le pourcentage de faux positifs est de 0%.

6. Rôle des molécules d'eau dans la performance du processus du *docking*

La présence des molécules d'eau joue un rôle important dans la précision des prédictions du *docking* protéine-ligand. Les molécules d'eau peuvent être impliquées dans la reconnaissance protéine-ligand soit en formant des liaisons hydrogènes de médiation entre la protéine et le ligand ou en étant déplacées par le ligand; ces deux mécanismes ont été dévoilés d'une importance pour la découverte de médicaments [96].

Afin d'étudier le rôle que les molécules d'eau jouent dans la prédiction des poses dans le *docking*, les molécules d'eau ont été rajoutées dans le site actif des protéines avant le *docking*

FlexX 1.3.0 a donné les résultats résumés dans le Tableau 1.3 de l'ANNEXE1. Résultats-FlexX:

D'après le Tableau 1.3, la performance de FlexX pour générer de bonnes poses a été significativement diminuée en raison de l'absence d'interactions médiatisées par les

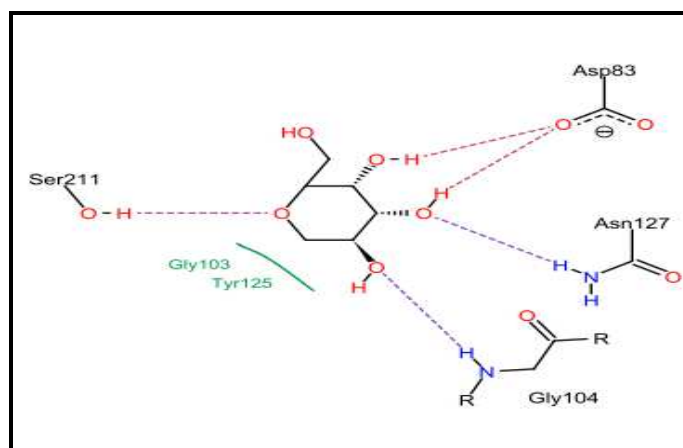
molécules d'eau conservées c'est le cas par exemple des ligands n° :19, 41, 51 et 101 où les scores ont nettement changés.

Par contre d'autres ligands ont gardé le même score, ce sont ceux qui n'ont pas les molécules d'eau dans le site actif de leur protéine cible c'est le cas par exemple des ligands n° :13, 75 et 84.

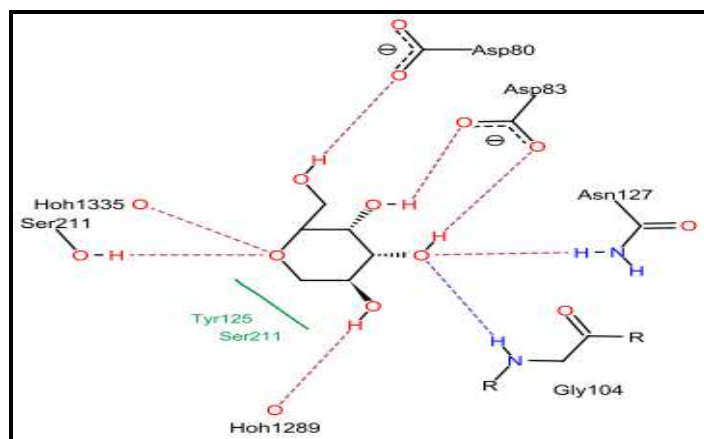
Cela n'est pas surprenant que les programmes de *docking* échouent généralement à produire des solutions correctes en absence d'interactions avec les molécules d'eau [96]. Toutefois, GOLD s'est avéré insensible à la présence des molécules d'eau dans l'identification des poses correctes.

La capacité de FlexX à générer des poses correctes a été réduite en absence d'interactions avec les molécules d'eau. Cela est dû au fait qu'en absence des molécules d'eau l'algorithme n'est pas en mesure de compléter le bon cheminement de la construction [49].

Voici en images un exemple de ligand arrimé par FlexX 1.3.0 en absence et en présence des molécules d'eau



A. En absence des molécules d'eau

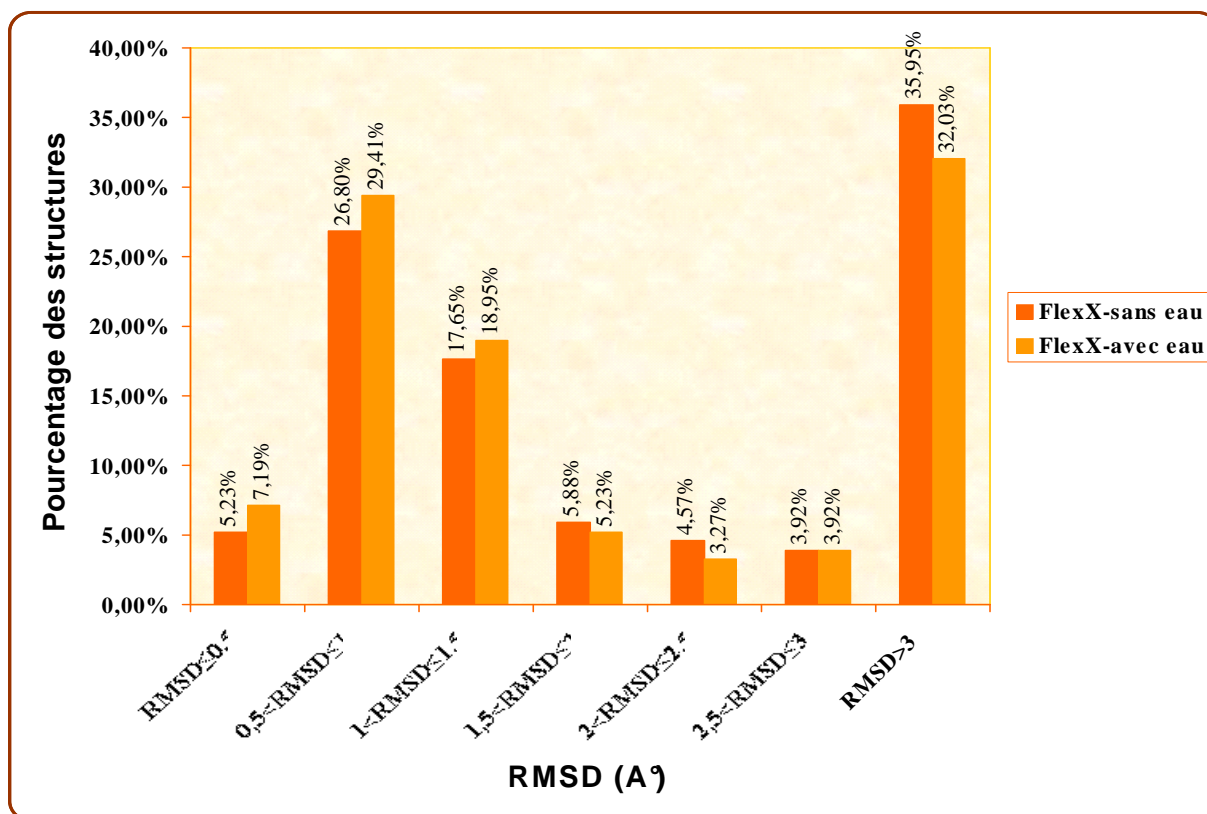


B. En présence des molécules d'eau

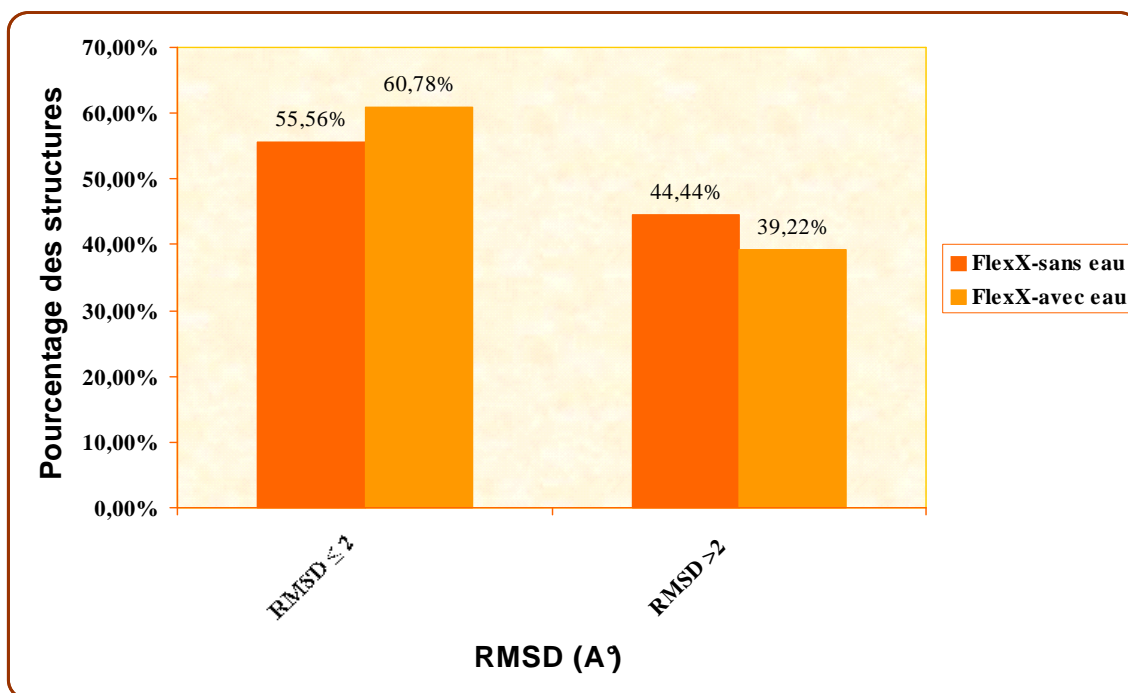
Figure 5.8 Le ligand n° 19 :2DVD

On peut résumer les résultats des RMSD du Tableau 1.3 de l'ANNEXE1 Résultats-FlexX dans le graphe suivant :

Graph 5.5 Résultats du *docking* par FlexX 1.3.0 en absence et en présence d'eau en fonction de la meilleure pose générée.



Graphe 5.6 Pourcentage des structures avec $\text{RMSD} \leq 2\text{\AA}$ et $\text{RMSD} > 2\text{\AA}$ amarrées par FlexX sans eau et avec eau.



D'après le Graphe 5.6 on observe la nette différence dans les pourcentages des structures avec un $\text{RMSD} \leq 2\text{\AA}$ en absence et en présence d'eau.

Les résultats de ce premier travail nous orientent vers le choix de FlexX 1.3.0 pour compléter notre étude, laquelle consiste à chercher de nouveaux inhibiteurs de la neuraminidase plus efficaces à travers les similaires du zanamivir tirés de la banque PubChem.

CHAPITRE 6

Exemple d'application : Etude des inhibiteurs de la neuraminidase

Nous avons tous eu un jour la grippe et souhaiterions ne pas l'avoir à nouveau. Une forte fièvre, des maux de tête, de la fatigue et bien d'autres symptômes, malheureusement, nous ne pouvons pas être sûrs d'y échapper la saison suivante (5 à 15 % de la population souffre d'infections des voies respiratoires supérieures chaque hiver). Décrite dès l'Antiquité et au Moyen-Age, les épidémies annuelles de grippe sont également une cause importante de mortalité, en particulier chez les personnes âgées et celles atteintes de maladies chroniques (3 à 5 millions de cas graves et 250 000 à 500 000 décès par an dans le monde). Enfin, de sévères pandémies de grippe – trois au cours du dernier siècle : grippe espagnole en 1918 (40 millions de morts), grippe asiatique en 1957 et grippe de Hong Kong en 1968 – ont coûté la vie à de nombreuses personnes sur une courte période de temps, touchant même les plus résistants [97].

A la lumière des résultats obtenus précédemment, le programme de *docking* FlexX a été choisi pour rechercher de nouveaux inhibiteurs d'une des cibles thérapeutiques du virus de la grippe qui est la neuraminidase.

1. Mises aux points bibliographiques

1.1. La neuraminidase : origine, structure et fonction

La famille des *Orthomyxoviridae* comprend trois genres : le premier regroupe les influenzavirus de type A et B, le second les influenzavirus de type C et enfin le dernier genre les *Thogotovirus*. La classification des influenzavirus en type s'appuie sur les différences antigéniques des protéines virales de nucléocapside et de matrice [98].

Seuls les influenzavirus de type A sont véritablement inféodés à diverses espèces animales. Ils ont été isolés chez des oiseaux (sauvages et domestiques), des mammifères terrestres (Homme, porcins, équidés, mustélidés, félidés) et des mammifères marins (baleines et phoques) [99]. Les influenzavirus de type B et C sont globalement inféodés à l'homme bien que des influenzavirus de type B aient également été isolés chez des phoques et que des influenzavirus de type C aient également été isolés chez des porcs.

Les influenzavirus de type A sont des virus enveloppés, de forme sphérique ou filamenteuse, d'un diamètre variant de 80 à 120 nm. Ils sont sensibles à la chaleur (30 minutes à 56°C), aux acides (pH 3) et aux solvants lipidiques mais sont particulièrement résistants dans les tissus et dans l'environnement, notamment dans l'eau [98].

Leur génome est constitué de huit segments d'ARN monocaténaire, codant pour dix protéines virales. Parmi elles, deux glycoprotéines d'enveloppe, l'hémagglutinine (notée HA) et la neuraminidase (notée NA), sont les principaux inducteurs d'anticorps protecteurs chez l'hôte infecté. L'hémagglutinine permet l'attachement de l'influenzavirus à un récepteur cellulaire de l'hôte par reconnaissance des acides sialiques que ce récepteur porte [100]. La neuraminidase permet l'hydrolyse du récepteur cellulaire ce qui favorise la libération des particules virales.

Jusqu'à récemment, quinze sous-types d'hémagglutinine (notés de H1 à H15) et neuf sous-types de neuraminidase (notés de N1 à N9), antigéniquement distincts, étaient répertoriés mais un seizième sous-type d'hémagglutinine a été isolé en 2005 chez des goélands, en Suède [100]. Un influenzavirus possède un sous-type d'hémagglutinine et un sous-type de neuraminidase, combinés de manière apparemment aléatoire.

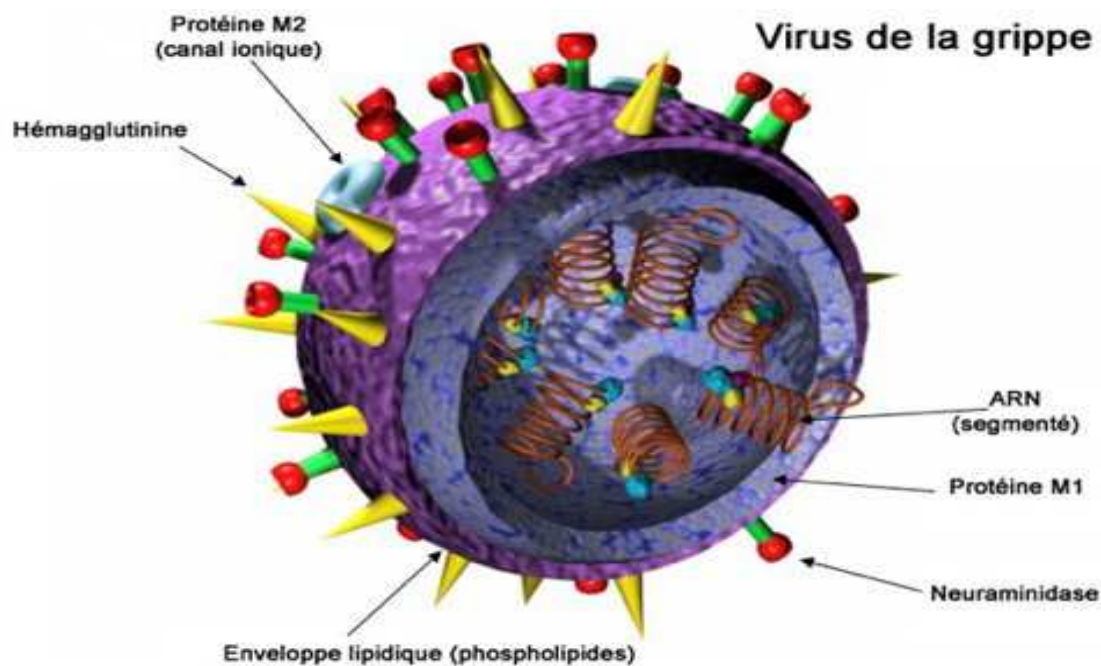


Figure 6.1 Structure du Virus de la grippe

Actuellement, seul un nombre limité de virus intéressent le genre humain : les virus dits saisonniers A (H1N1) et A (H3N2) responsables d'épidémies hivernales, le virus pandémique A (H1N1)2009 qui, depuis avril 2009 circule sur toute la planète et un certain nombre de virus d'origine aviaire comme les virus A (H5N1), A (H7N7), A (H9N2) responsables d'infections humaines parfois sévères, principalement en Asie du Sud-Est [101].

La NA des virus influenza de type A et B est une glycoprotéine membranaire de type II, avec une activité enzymatique de type N acétyl-neuraminosyl-glycohydrolase (EC 3.2.1.18). La NA est un homotétramère ancré à la membrane virale par une séquence hydrophobe unique de quelques 29 acides aminés près de l'extrémité N-terminale. Des images de cette glycoprotéine, obtenues par coloration négative, montrent une morphologie de type « champignon » avec une partie supérieure globulaire reliée à une tige mince dont la partie distale est enchâssée dans la membrane virale. La zone globulaire est composée de quatre structures identiques. Chaque structure ou monomère est majoritairement composée de structures secondaires en feuillet β ; cette zone a été subdivisée en six parties telle une hélice (nommées $\beta 1$ à $\beta 6$) (cf. Figure 6.2). Le site catalytique de la NA grippale, constitué de résidus

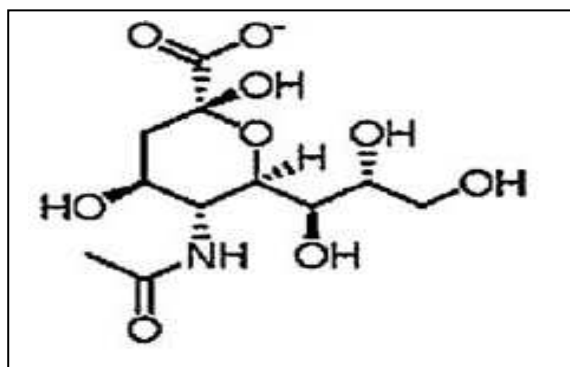
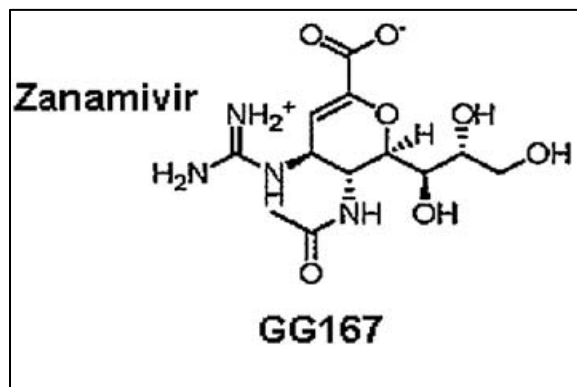
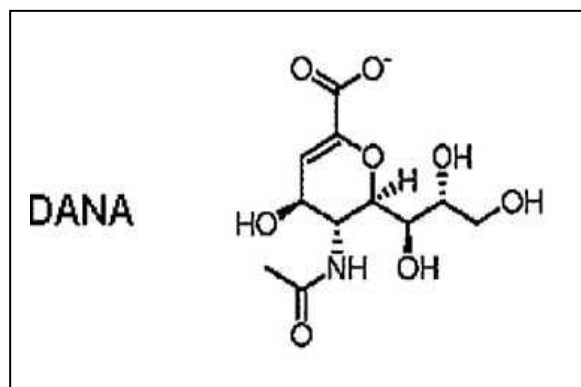


Figure 6.3 Structure chimique de l'acide sialique : acide N-acétyl neuraminique (Neu5Ac) ou (NANA)

1.2 Inhibiteurs de la neuraminidase

Les premiers inhibiteurs de la neuraminidase (INAs) ont été élaborés en fonction de l'état de transition du substrat durant l'hydrolyse de ce dernier. Le principal analogue obtenu fût l'acide 2-deoxy-2,3-dihydro-Nacetylneuraminique (DANA). Puis, à partir de la structure de la NA de virus influenza de type A et B déterminée par rayons X et de son complexe avec l'acide sialique ou le DANA, deux nouvelles molécules efficaces ont été synthétisées, le zanamivir et l'oseltamivir carboxylate. Le zanamivir (GG167, 2,4-didéoxy-2,3- didéhydro-4-guanidine sialic acid) est un analogue synthétique de l'acide sialique, analogue à la structure du DANA, avec la substitution d'un groupement guanidinium sur la position 4 du cycle. L'oseltamivir (GS4104 ester éthylique (prodrogue), GS4071 (médicament actif) ; éthyl-4-acétamido-5-amino-3-(1-ethylpropoxy)-1-cyclohexane-1-carboxylate)est,lui, analogue à un cyclohexane avec la substitution d'un groupement aminé en position 4 sur le cycle et un groupement hydrophobe pentyl éther qui remplace la chaîne latérale poly hydroxylée en position 6. (cf. Figure 6.4) [105].



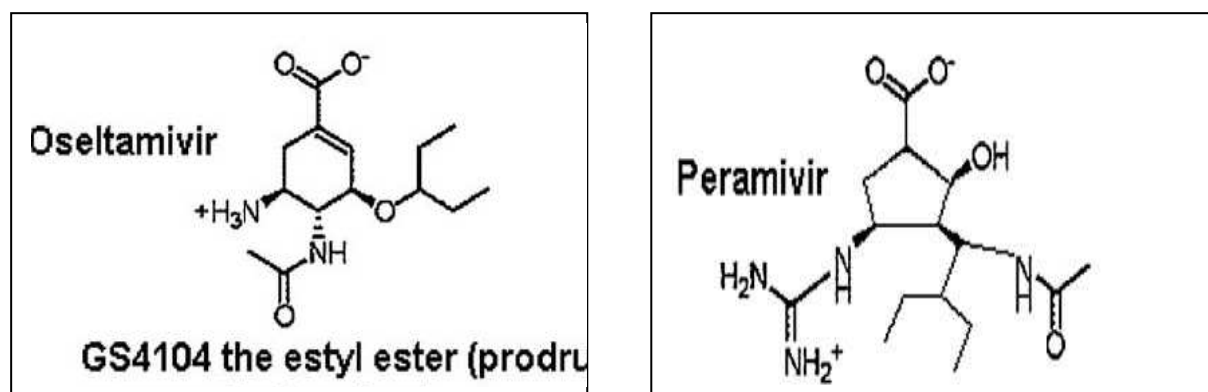


Figure 6.4 Structures chimiques dans l'ordre du : *DANA*, acide 2-deoxy-2,3-dehydro-N acétylneuraminique (Neu5Acen). *Zanamivir*, 4-guanidino-Neu5Ac2en. *Oseltamivir*, 4-N acetyl-5-amino-3-(1-ethylprooxy)-1-cyclohexane-1-carboxylic acid. *Peramivir*

Les inhibiteurs de la neuraminidase (INAs) sont des analogues de l'acide sialique. Ils inhibent par compétition spécifique l'action de la neuraminidase en bloquant son site actif. Il s'ensuit un blocage du relargage des virions qui s'agrègent à la surface de la cellule hôte et entre eux, par liaison de leur hémagglutinine aux résidus de l'acide sialique non clivé. La dissémination virale se retrouve diminuée.

- le zanamivir (Relenza®) mis au point en 1993 et ayant obtenu son autorisation de mise sur le marché (AMM) française en juillet 1999
- l'oseltamivir (Tamiflu®) mis au point en 1997 et ayant obtenu une AMM européenne en juillet 2002.

L'oseltamivir est une prodrogue rapidement absorbée et transformée en métabolite actif (=carboxylate d'oseltamivir) par les estérases hépatiques et intestinales. Environ 80% de la dose orale atteint la circulation systémique sous forme de métabolite actif. Celui-ci est décelé dans le plasma dans les trente minutes suivant la prise. Le métabolite est distribué dans tous les tissus. La demi-vie est de six à dix heures. L'élimination se fait par voie rénale.

Cependant, il ne faut pas ignorer que l'usage de cet antiviral peut mener au développement de résistances. La résistance à l'oseltamivir est liée à diverses mutations de la neuraminidase.

Récemment, il a été montré que la résistance du virus H5N1 à l'oseltamivir était liée à la mutation H274Y [106].

Le zanamivir a une faible biodisponibilité orale et se présente sous la forme d'une poudre sèche inhalée à l'aide d'un diskhaler. Environ 10 à 20% du médicament atteignent les voies respiratoires basses, 70 à 80% se déposent dans l'oropharynx. Moins de 20% passent dans la

circulation systémique et 90% du médicament absorbé sont excrétés sous forme inchangée dans l'urine. La demi-vie sérique est de trois à cinq heures. Il est actif sur le virus H5N1 présentant une mutation H274Y. Bien que cela n'ait pas encore été prouvé, la résistance liée à la neuraminidase est plus probable de se développer contre l'oseltamivir que contre le zanamivir [106].

L'association de ces deux molécules peut être envisagée à condition que les mutations conférant la résistance à chacune des molécules ne se chevauchent pas.

D'autres inhibiteurs de la neuraminidase tels que le peramivir (RWJ270201) ont été décrits. Ces molécules sont efficaces contre cinq variants résistants au zanamivir et six variants résistants à l'oseltamivir [107].

Cependant, les premières études montraient une moindre efficacité du peramivir par voie orale contre le virus de type A. Le peramivir a l'avantage d'une demi-vie plus longue, ce qui permet de réduire les posologies [107]. Les études suivantes se sont intéressées à l'administration intraveineuse du peramivir.

1.3. Changements antigéniques et résistance

La grande variabilité du virus influenza permet de générer de nombreux variants pouvant être à l'origine de nouvelles pandémies. Contrairement à ce que l'on observe avec certains autres virus respiratoires, on ne peut être immunisé à vie contre influenza même après infection. Cela s'explique par la grande facilité avec laquelle les domaines antigéniques des glycoprotéines HA et NA du virus peuvent muter.

En effet, ces changements antigéniques peuvent se faire de deux façons : (1) lentement mais de manière continue (glissement antigénique), ou (2) très soudainement mais de manière occasionnelle (cassure antigénique) [108].

Les **glissements antigéniques** (ou « *drift* ») sont des changements lents qui s'opèrent à la surface du virus au cours du temps et qui résultent en l'accumulation de petites mutations au niveau des gènes codants pour les glycoprotéines. En effet, par l'absence d'enzyme de relecture, l'ARN polymérase qui réplique le génome viral fait des erreurs : 1 tous les 10 000 nucléotides (~ taille du génome d'influenza). Une grande majorité des virus nouvellement synthétisés s'avèrent donc être des mutants. Ce mécanisme est commun à un grand nombre de virus à ARN et complique considérablement la mise au point de vaccins. Ces glissements

antigéniques ont ainsi permis la génération de nombreux sous-types du virus influenza, principalement présents chez les oiseaux, totalisant 16 antigènes HA et 9 NA bien distincts.

Les **cassures antigéniques** (ou « *shift* ») sont le résultat de réassortiments de gènes viraux quand une cellule se retrouve infectée par deux ou plusieurs souches différentes. La cellule va alors produire des réassortants qui combineront une partie du génome de chaque souche et si un de ces réassortants peut se répliquer, il sera alors amplifié. Ces changements génétiques très rapides sont favorisés par la fragmentation du génome viral en 8 chromosomes ; ils permettent l'émergence de nouvelles souches recombinées pouvant être à l'origine de pandémie. Ces cassures antigéniques n'arrivent que quelques fois au cours d'un siècle, mais les conséquences peuvent en être désastreuses.

Les mutations observées sur la NA ont mis en évidence un taux plus élevé de mutation après traitement à l'oseltamivir par rapport au zanamivir. Cinq mutations (E119 V, D198N, H274Y, R292K et N294S) ont ainsi été identifiées après traitement par l'oseltamivir, tandis que seulement deux (Q136K et R152K) l'ont été après traitement par le zanamivir. D'un point de vue fonctionnel, les mutations R292K, H274Y et N294S semblent bloquer le changement conformationnel du site actif nécessaire à la fixation de l'oseltamivir dans ce dernier. Toutes ces mutations réduisent l'affinité de liaison entre l'oseltamivir et le site actif de l'enzyme sans pour autant bloquer la fixation et l'hydrolyse du substrat [101].

Les mutations R292K et E119 V sont principalement observées sur les NAs de sous-type N2, la mutation H274Y principalement sur les NAs de sous-type N1. Les sites actifs des N1 et N2 diffèrent principalement par la présence d'une cavité adjacente au site actif observable uniquement dans le groupe N1 [101]. L'influence de cette variabilité structurelle sur la sensibilité aux inhibiteurs a été rapportée durant les programmes de surveillance de la sensibilité aux molécules antivirales.

Govorkova *et al.* ont montré que les NAs N2, N3, N6, N7 et N9 étaient plus sensibles au zanamivir qu'à l'oseltamivir tandis que les NAs N1, N4, N5 et N8 étaient plus sensibles à l'oseltamivir qu'au zanamivir.

L'émergence d'une mutation induisant une diminution de la sensibilité aux inhibiteurs dépend de l'inhibiteur et du groupe structurel auquel appartient la NA. L'incidence d'une mutation sur la sensibilité aux inhibiteurs ainsi que sur l'activité et la stabilité de l'enzyme dépend

également du sous-type considéré. La réduction de l'activité neuraminidasique et de la stabilité de l'enzyme, associée à une diminution de la capacité répliquative du virus muté démontre que la réduction de l'affinité de liaison aux inhibiteurs est le plus souvent associée à une diminution de la fonction enzymatique et donc du fitness du virus.

2. Matériel et méthode

2.1. Choix de la cible

La NA a été choisie comme cible antivirale, en raison de son rôle majeur dans la propagation des virus influenza [104] et en raison de la conservation des résidus d'acides aminés du site actif qui interagissent directement ou indirectement avec le substrat dans les virus influenza de type A et B.

N1 et N2 sont les sous-types représentatifs qui circulent actuellement dans les virus de la grippe humaine. Les virus de la pandémie grippale H1N1 et H5N1 actuels appartiennent au sous-type N1 [101]. La structure cristallographique de la NA de la grippe aviaire H5N1 révèle que le groupe-1 et le groupe-2 de la NA ne sont pas seulement génétiquement distincts mais aussi structuralement distincts les uns des autres. Le groupe-1 des NAs N1, N4, N5, N8 possède une boucle flexible (connue sous le nom de la boucle-150) et une cavité (la cavité 150) adjacente au site actif conservé. Le groupe-1 des NAs peut changer la conformation de la forme «ouverte» à la forme «fermée» une fois liées avec les inhibiteurs, tandis que le groupe-2 des NAs assume la forme «fermé» à tous les temps. Les médicaments actuels tels que le zanamivir et l'oseltamivir sont à l'origine conçus sur la base des structures cristallines du groupe-2 des NAs (N2 et N9).

Compte tenu du fait que le site actif du groupe-1 des NAs change de la forme «ouverte» à la forme «fermée» lors de la liaison avec l'inhibiteur, l'étude de telles transformations conformationnelles gagnerait des idées très utiles pour concevoir des inhibiteurs efficaces spécifiquement contre les NAs du groupe-1 en inhibant le site actif conservé et la cavité -150, simultanément. Ceci est important pour concevoir des médicaments plus efficaces parce un site de multi liaison des inhibiteurs peut-être plus étanche en terme des affinités et des interactions qu'un site unique de liaison des inhibiteurs.

La structure cristalline de la NA est téléchargée à partir de la PDB [9], le choix est tombé sur le complexe 3CKZ qui possède la cavité 150 adjacente au site de liaison (cf. Figure 6.5).qui

est une caractéristique typique du sous-type N1 de NA. Il a été montré par les recherches antérieures [109] que la réplication virale était significativement inhibée par la liaison d'un composé à la cavité 150 de la N1 de NA.

2.2. FlexX 1.3.0

Dans cette étude, nous avons appliqué le programme de *docking* FlexX [63], qui utilise une méthode de construction incrémentale efficace afin d'optimiser l'interaction entre un ligand flexible et un site de liaison rigide. Dans cette méthodologie, une fonction de score empirique, qui est optimisée pour reproduire les affinités de liaisons expérimentales et les conformations de liaisons pour diverses complexes cristallographiques protéine / ligand résolus, est utilisée pour prédire l'énergie libre de liaison (ΔG_{bind}). Dans toutes les expériences de *docking* présentées dans cette étude, une fonction de score avec les paramètres par défaut a été utilisée.

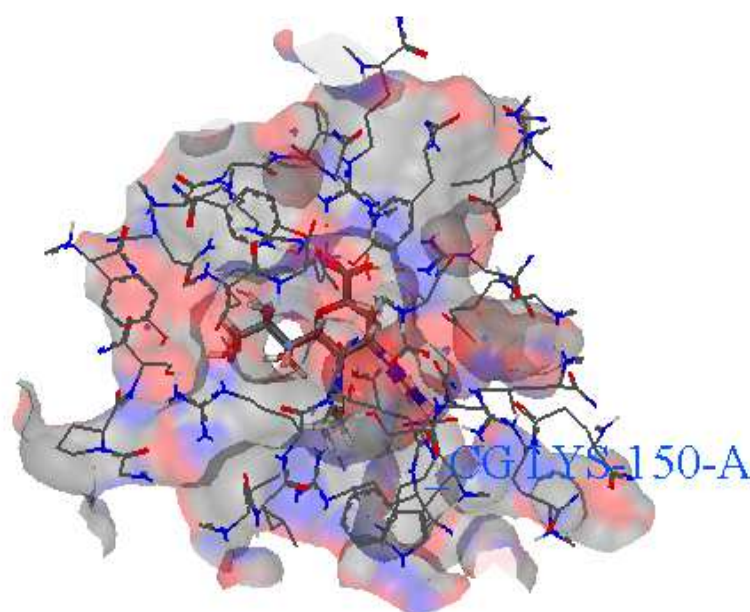


Figure 6.5 La surface moléculaire de la neuraminidase N1 liée avec le zanamivir (le complexe 3CKZ)

2.3. PubChem

PubChem [110] est une banque de données américaine de molécules chimiques gérée par le *National Center for Biotechnology Information* (NCBI), branche de la Bibliothèque nationale de médecine des États-Unis sous l'autorité de la *National Institutes of Health* (NCI). PubChem répertorie actuellement plusieurs millions de composés dont la structure et les propriétés physico-chimiques sont gratuitement accessibles *via* Internet.

Cette banque a été utilisée pour faire ressortir des similaires structuraux de notre ligand le zanamivir à un pourcentage de 90% afin de les amarrer par notre programme FlexX 1.3.0 et de rechercher parmi ces derniers un qui se lie plus fortement que le zanamivir à notre cible la neuraminidase N1 et de le proposer comme un nouveau inhibiteur plus efficace pour la neuraminidase.

2.4. La règle de Christopher A. Lipinski (1997)

Chaque médicament éventuel doit de se conformer à plusieurs critères de base, tels son faible coût de production, être soluble, stable, brevetable, mais doit aussi se conformer à des barèmes associés à ses propriétés pharmacologiques d'absorption, de distribution, de métabolisme, d'excrétion et de toxicité [41]. La liste des aspects à inclure est très longue mais on admet communément que les molécules validant au moins deux conditions de la règle de Lipinski sont des candidats potentiels. Ces règles sont :

- Un poids moléculaire maximum de 500g/mol
- Un maximum de 5 liaisons H-donneurs
- Un coefficient de partition (logP) ou une lipophilie ≤ 5 .
- Un maximum de 10 liaisons H-accepteurs

LogP est égal au logarithme du rapport des concentrations de la substance étudiée dans l'octanol et dans l'eau. $\text{LogP} = \text{Log} (C_{\text{oct}}/C_{\text{eau}})$. Cette valeur permet d'appréhender le caractère hydrophile ou hydrophobe (lipophile) d'une molécule.

Diverses études statistiques ont mis en évidence les valeurs optimales de logP pour qu'un composé (typiquement un médicament) soit absorbé par le corps humain :

- Pénétration au cerveau : 2,0
- Absorption orale : 1,8
- Absorption sub-linguale : 5
- Absorption percutanée : 2,6

En se basant sur cette simple règle de Lipinski, le criblage virtuel à haut débit pourra, par exemple, exclure d'une banque de composés initiaux de 1 milliard d'éléments, 900 000 éléments qui ne possèdent pas les propriétés communes aux médicaments administrables oralement. C'est la raison pour laquelle nous avons évalué ces paramètres pour les différents inhibiteurs étudiés.

Un autre aspect de sélection très important dans le domaine pharmacologique est l'élimination des drogues manifestant des propriétés toxiques *in vivo*. Indice d'absorption, de distribution, de métabolisme, d'excrétion (ADME) et de toxicité sont généralement appliqués aux ensembles de tests utilisés pour caractériser les composés selon leur absorption par les intestins, distribution dans l'organisme, métabolisme par le foie, excrétion par les reins et profil de toxicité. Le criblage ADME/Tox dans la conception et le développement de médicament joue un rôle important dans le processus étant donné le nombre important de composés "lead" (modèles) et de drogues qui détiennent des propriétés toxiques.

En effet, en complément de ces règles, d'autres critères sont pris en compte dans la sélection de molécules potentiellement candidates :

- nombre de fonctions rotatives ≤ 15
- nombre d'halogènes ≤ 7
- chaînes alkyles $\leq -(CH_2)_6-CH_3$
- nombre de cycles ≤ 6
- pas de grands cycles de plus de 7 membres
- au moins un atome d'azote ou d'oxygène dans la molécule.

3. Résultats et discussion

3.1. Vérification de la règle de Lipinski

Tous les similaires de notre inhibiteur le zanamivir étudiés répondent à la règle de Lipinski comme indiqué dans le tableau ci-dessous (Tableau 6.1)

Tableau 6.1 Les similaires de zanamivir répondant aux critères de la règle de Lipinski

N°	Ligand	Poids Moléculaire (g/mol)	LogP	n Hacc	n Hdon	n Liais Rot
1	CID44574175	416,42622	-1,8	9	5	11
2	CID 10112455	414,4534	-1,4	8	5	8
3	CID 5270784	316,26734	-0,7	9	5	6
4	CID 5273235	316,26734	-0,7	9	5	6
5	CID10041431	316,26734	-0,7	9	5	6
6	CID14802591	316,26734	-0,7	9	5	6
7	CID14802593	316,26734	-0,7	9	5	6
8	CID21157879	316,26734	-0,7	9	5	6
9	CID5273237	318,32298	-4,1	8	5	6
10	CID44371716	290,26982	-4,4	7	5	4
11	CID16095341	458,50596	-0,2	9	5	14
12	CID10180973	428,47998	-0,9	8	5	9
13	CID10181866	442,50656	-0,3	8	5	10
14	CID10205443	470,55972	0,7	8	5	12
15	CID10300514	498,61288	1,8	8	5	14
16	CID46215567	329,30586	-1,3	9	5	7
17	CID23380404	372,37366	-1,2	9	4	10
18	CID502292	328,36416	-0,6	6	4	7
19	CID502293	342,39074	-0,1	6	4	8
20	CID10088846	344,3205	0,3	9	5	7
21	CID5273230	370,39754	-2,8	8	5	10
22	CID10740676	417,45404	-3,2	9	5	11
23	CID10740677	417,45404	-3,4	9	5	9
24	CID11724157	330,29392	-0,4	9	4	7
25	CID502295	412,52364	2,4	6	4	13
26	CID5278155	326,34828	-1,1	6	4	4
27	CID10781794	318,32298	-2	8	5	7
28	CID502294	370,4439	0,8	6	4	10
29	CID10038864	272,25784	-2	6	5	4
30	CID10264569	260,24384	-4,4	7	5	4
31	CID10265401	275,25518	-1,4	7	5	5
32	CID10497603	344,3205	0,1	9	4	8

3.2. Le docking par FlexX 1.3.0

Une fois que la règle de Lipinski est vérifiée pour tous les similaires étudiés, le docking par FlexX 1.3.0 de ces derniers ainsi que notre inhibiteur d'origine a donné les résultats suivants : (Tableau 6.2)

Tableau 6.2 Résultats du docking du zanamivir et ses similaires à 90% dans la neuraminidase

N1

N°	Ligand	Score	Match	Lipo	Ambig	Clash	Rot	#Match
	3CKZ	-74.9698	-78.1127	-5.787	-11.9453	2.8751	12.6000	16
1	CID 44574175	-67.0328	-71.7204	-6.3155	-12.4883	2.6914	15.4000	15
2	CID 10112455	-38.1617	-45.5758	-6.1217	-10.9531	6.4889	12.6000	15
3	CID 5270784	-39.7879	-52.8120	-2.8300	-5.4216	4.6756	11.2000	13
4	CID 5273235	-45.4008	-57.7234	-2.8738	-6.8157	5.4122	11.2000	10
5	CID 10041431	-49.0993	-57.6054	-2.2210	-7.2202	1.3473	11.2000	11
6	CID14802591	-55.8186	-63.2724	-2.1901	-7.4754	0.5194	11.2000	12
7	CID14802593	-49.0904	-57.9034	-2.4892	-6.8011	1.5032	11.2000	10
8	CID 21157879	-48.4832	-59.6385	-3.0475	-6.6192	4.2220	11.2000	13
9	CID 5273237	-38.7517	-52.6517	-3.3567	-8.4927	9.1494	11.2000	11
10	CID 44371716	-49.1148	-56.5160	-4.3498	-6.8784	3.4294	9.8000	11
11	CID 16095341	-66.7145	-74.4605	-6.8622	-12.6514	2.2596	19.6000	17
12	CID 10180973	-37.5495	-45.1258	-6.9662	-12.6318	7.7743	14.0000	14
13	CID 10181866	-37.5329	-45.2882	-7.7349	-12.6514	7.3416	15.4000	14
14	CID 10205443	-39.6282	-51.5136	-8.5857	-10.7491	7.6201	18.2000	15
15	CID 10300514	-38.3600	-51.5424	-10.1696	-10.6770	7.6290	21.0000	15
16	CID 46215567	-21.1643	-33.9521	-3.2307	-6.3054	4.3239	12.6000	11
17	CID 23380404	-57.3460	-64.1054	-3.3540	-7.7486	1.2621	11.2000	14
18	CID 502292	-68.0551	-67.8805	-6.4697	-10.8048	3.2999	8.4000	13
19	CID 502293	-69.3773	-66.6381	-8.2999	-12.8015	3.1623	9.8000	14
20	CID 10088846	-58.0467	-66.1309	-5.9218	-10.8032	6.8092	12.6000	14
21	CID 5273230	-44.1653	-57.4363	-4.9959	-8.8595	4.9264	16.8000	12
22	CID 10740676	-73.7973	-81.6385	-6.4655	-11.7167	3.8234	16.8000	16
23	CID 10740677	-61.1877	-70.9431	-6.8893	-11.1366	8.3813	14.0000	13
24	CID 11724157	-26.3543	-39.8034	-3.4699	-6.7579	5.6769	12.6000	13
25	CID 502295	-66.3935	-69.0091	-8.4919	-14.4196	3.3271	16.8000	14
26	CID 5278155	-74.4958	-68.7768	-6.7076	-10.9616	2.3502	4.2000	14
27	CID 10781794	-37.1979	-47.1481	-5.7340	-8.2900	5.9742	12.6000	14
28	CID 502294	-67.4514	-68.1714	-8.6600	-14.1970	5.5770	12.6000	14
29	CID 10038864	-69.2587	-66.4422	-4.7206	-10.9010	1.8051	5.6000	14
30	CID 10264569	-68.1993	-70.6802	-4.8834	-8.2759	3.2403	7.0000	14
31	CID 10265401	-44.2669	-57.0326	-2.7569	-5.8545	6.1771	9.8000	14
32	CID 10497603	-25.6356	-36.2767	-5.4260	-9.2352	5.9023	14.0000	12

Nous avons trouvé intéressant de tester ces similaires, de comparer leur affinité à la cible par rapport au ligand initial et de trouver ainsi le meilleur inhibiteur de l'enzyme.

Parmi les 32 similaires testés, seul le similaire n° 26 (cf. Tableau 6.2) établit une interaction proche de celle de notre inhibiteur d'origine le zanamivir (3CKZ), qui reste, cependant, le meilleur inhibiteur de la neuraminidase N1 avec une énergie de $-74,9698$ KJ/mol et un ensemble de 15 liaisons hydrogène (cf. Figure 6.6).

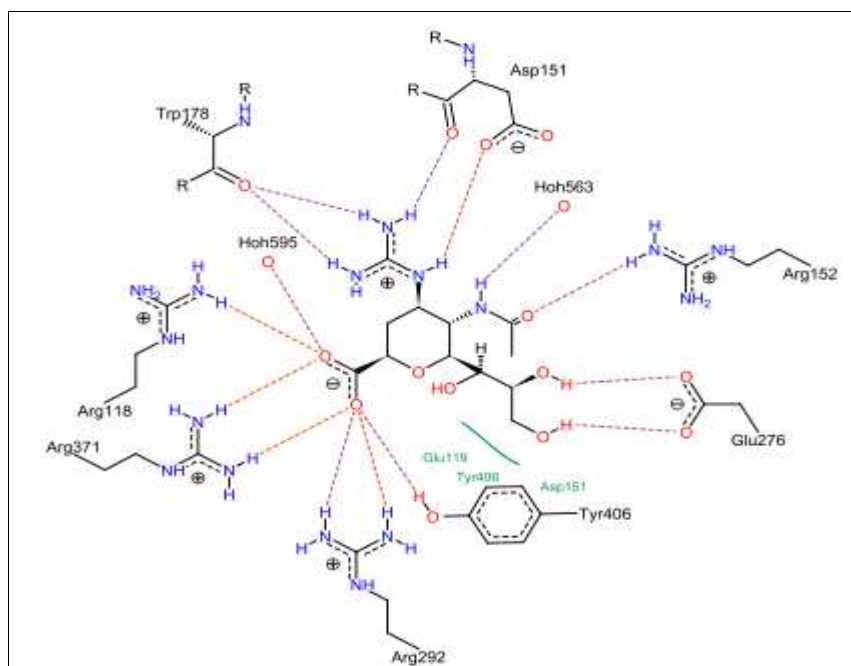


Figure 6.6 Neuraminidase N1-zanamivir (Le complexe 3CKZ)

Malgré que le similaire n° 26 établit moins de liaisons hydrogène avec la neuraminidase N1 que le zanamivir avec seulement un total de 13 liaisons hydrogène (cf. Figure 6.7) ce qui se traduit par une légère baisse de l'énergie d'interaction qui est de $-74,4958$ KJ/mol pour ce dernier, ce similaire possède des propriétés pharmacocinétiques plus intéressantes que le zanamivir.

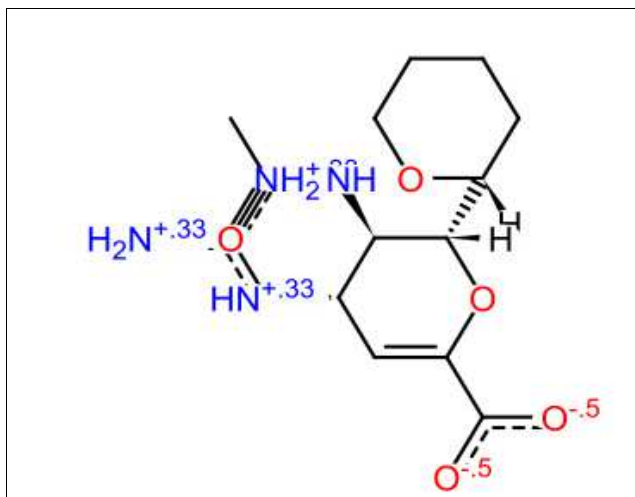


Figure 6.8 Structure du similaire n°26 du zanamivir

A partir des 3 remarques indiquées ci-dessus, on peut proposer le similaire n°26 comme un autre inhibiteur de la neuraminidase N1. Il serait également intéressant de faire des substitutions sur ce dernier pour faire augmenter d'avantage son énergie d'interaction. En plus des tests *in vitro* et *in vivo* sont recommandés pour confirmer l'efficacité de notre programme FlexX.

Conclusion

La compréhension et le traitement des interactions inter et intramoléculaires en biologie représentent un enjeu majeur pour l'exploration et l'utilisation des biomolécules à des fins thérapeutiques ou technologiques.

Dans ce contexte, notre travail avait pour but de tenter une comparaison entre deux programmes de *docking* GOLD 5.0.1 et FlexX 1.3.0 à travers différents paramètres en utilisant 3 cibles l'hémagglutinine, la neuraminidase et la peptide deformylase formant un ensemble de 153 complexes et d'appliquer les résultats obtenus pour proposer de nouveaux inhibiteurs d'une des cibles thérapeutiques qui est la neuraminidase à partir des similaires du zanamivir tirés de la PubChem.

Notre démarche consistait dans un premier temps, à explorer lequel des deux programmes utilisés trouve des solutions expérimentales pour les complexes protéine-ligand étudiés en comparant la déviation par le RMSD entre la structure simulée et celle obtenue par cristallographie, avec des valeurs du RMSD inférieures à 2\AA . Les performances des approches incrémentales (FlexX) et par algorithme génétique (GOLD) sont nettement prouvées et parfaitement établies pour les différentes structures choisies dans cette étude avec, pour GOLD, un taux de 64,05% alors que pour FlexX 55,56% seulement.

Nous avons aussi exploré le taux de réussite de ces deux programmes pour ces familles de récepteurs en cherchant le classement du mode de liaison expérimentale du ligand quand un programme de *docking* est en mesure de le reproduire, les résultats ont montré que FlexX a réussi à classer 54.25% des modes de liaison correctement prédits tandis que GOLD n'a pu classer que 49.02% des ligands avec un $\text{RMSD} \leq 2 \text{\AA}$.

Avec les besoins de l'industrie pour le criblage virtuel et la nécessité pour les fonctions de score qui peuvent discriminer les composés actifs de ceux inactifs, ce constat est assez encourageant. Toutefois, il faut prendre en compte non seulement la précision et la performance mais également la vitesse d'exécution. GOLD tend à être très lent avec une moyenne de 2248 secondes alors que FlexX est plus rapide avec une moyenne de 109 secondes pour les complexes étudiés ici.

Parfois les solutions presque optimales sont perdues parmi un ensemble de solutions avec des scores élevés et de grandes valeurs de RMSD c'est ce qu'on a appelé faux positifs ; d'après les résultats obtenus il est bien clair que FlexX produit moins d'erreurs lors du classement des solutions avec un pourcentage de 10,92% alors que GOLD échoue dans 19,02% des cas.

Dans une tentative à comprendre quels sont les facteurs de l'échec des programmes de *docking*, nous avons décidé d'analyser les résultats en terme de complexité des ligands ; FlexX semble être le plus sensible car il est capable de reproduire 90.59% des complexes avec un $\text{RMSD} \leq 2\text{\AA}$ quand les ligands ont 10 liaisons rotatives ou moins tandis que GOLD n'a pu générer que 56.12%.

Dans un deuxième temps, nous avons jugé nécessaire d'introduire les molécules d'eau dans le processus de *docking* pour voir à quel point ce dernier joue un rôle dans la reconnaissance protéine-ligand. Puisque des publications ont montré que GOLD est indifférent à la présence ou l'absence d'eau dans l'identification des poses, le traitement avec l'eau ne s'est appliqué que pour FlexX qui a montré une différence dans la capacité de générer de bonnes poses allant jusqu'au taux de 60,78% en présence d'eau ; ce qui n'était pas le cas en absence de ce dernier avec seulement la valeur de 55,56%.

Les résultats de comparaison obtenus entre les deux programmes à travers les paramètres cités et les cibles thérapeutiques traitées nous ont permis de choisir FlexX comme un outil pour accomplir la troisième partie de notre travail qui consiste à proposer de nouveaux inhibiteurs de la neuraminidase N1 par le biais des similaires de son inhibiteur naturel issus de la PDB. La simulation par ordinateur du zanamivir et ses similaires nous a permis de suggérer le similaire CID 5278155 comme un nouvel inhibiteur potentiel de la neuraminidase N1 avec une énergie d'interaction comparable à celle du zanamivir mais avec des propriétés pharmacocinétiques plus intéressantes.

Malgré que la comparaison entre les programmes de *docking* soit un processus délicat, cette étude mérite d'être poursuivie :

- Par l'analyse d'autres paramètres de comparaison à savoir l'enrichissement.
- Par l'utilisation des fonctions de rescoring pour diminuer les erreurs issues des fonctions de score.
- En faisant intervenir la flexibilité de la protéine qui est un processus difficile ; cependant la flexibilité du site actif a apporté dans plusieurs études une amélioration dans la précision du *docking*.
- En procédant à des substitutions au niveau du zanamivir et de son similaire proposé CID 5278155 afin d'améliorer l'énergie d'interaction.

BIBLIOGRAPHIE

- [1] Human Genome Project. [en ligne]. Dernière mise à jour Lundi 13 Décembre 2010 http://www.ornl.gov/sci/techresources/Human_Genome/home.-.shtml. Consulté le 3 Mars 2011.
- [2] Figeys D., 2003, Novel approaches to map protein interactions. *Curr. Opin. Biotechnol*, **14**, 119- 125.
- [3] Ameziane N., Bogard M., Lamoril J., Principes de biologie moléculaire en biologie clinique. *Elsevier Masson*, 2006. 705p
- [4] Gohlke H.; Klebe G., 2002, Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed. Engl*, **41** (15), 2644-2676.
- [5] Lodish H., Berk A., Matsudaira P., Kaiser C.A., Krieger M., Scott M.P., Zipursky S.L., Darnell J., chap. 1 In *Biologie moléculaire de la cellule*. De Boeck Université, 2005.
- [6] Julian-Reynier C., Moatti J.P., Bourret P., 1996, Vers une colonisation génétique de la médecine ? *Sociologie et Sociétés*, **28**(2), 141-155.
- [7] Bergson H., L'évolution créatrice. *Presses Universitaires de France*, 1907, 86ème Ed, 2007.
- [8] Monod J., Le hasard et la nécessité : essai sur la philosophie naturelle de la biologie moderne. Paris, *Seuil*, rééd. Point-Essais n°43, 1970. 244p
- [9] Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E., 2000, The Protein Databank. *Nucl. Ac. Res*, **28**, 235-242.
- [10] Cavarelli J., 2000, Cristallographie des macromolécules biologiques. *Techn. Ing*, **2**, 1090.
- [11] Malliavin T., Dardel F., 2002, Structure des protéines par RMN. *Techn. Ing*,
- [12] Berman H., Henrick K., Nakamura H., Markley J.L., 2007, The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucl. Ac. Res*, **35**, 301-303.
- [13] Radaev S., Li S., Sun P.D., 2006, A survey of protein-protein complex crystallizations. *Acta Cryst*, **62**, 605-612.

- [14] McCammon J.A., Harvey S.C., Dynamics of proteins and nucleic acids. *Cambridge University Press*, 1987.
- [15] Sali A., Glaeser R., Earnest T., Baumeister W., 2003, From words to literature in structural proteomics. *Nature*, **422**, 216-225.
- [16] Berry I.M., Dym O., Esnouf R.M., Harlos K., Megeed R., Perrakis A., Sussman J.L., Walter T.S., Wilson J., Messerschmidt A., 2006, SPINE high-throughput crystallization, crystal imaging and recognition techniques: current state, performance analysis, new technologies and future aspects. *Acta Crystallogr D Biol Crystallogr*, **62**, 1137-1149.
- [17] Anfinsen C.B., 1973. Principles that govern the folding of protein chains. *Science*, **181**, 223-230.
- [18] Dobson C.M., Sali A., Karplus M., 1998, Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed.*, **37**, 868-893.
- [19] Ronsin C., L'histoire de la biologie moléculaire : héros et pionniers. De Boeck Université, 2005.
- [20] Grognet J M., 2006, Nanobiotechnologies : présent, promesses et enjeux. *MURS*, **47**, 38-48.
- [21] Synthetic Biology. [en ligne].<http://syntheticbiology.org/> consulté le 20 Mars 2011
- [22] Astbury W., 1961. Molecular Biology or Ultrastructural Biology? *Nature*, **190**, 1124.
- [23] Gershel A., Liaisons intermoléculaires. Inter Editions/CNRS Editions. S.L : Savoirs Actuels, 1995.
- [24] Goh C S., Milburn D., Gerstein M., 2004, Conformational Changes Associated with Protein-Protein Interactions. *Curr. Op. Struct. Biol*, **14**, 104-109.
- [25] Langmuir I., 1919, The Arrangement of Electrons in Atoms and Molecules. *J. Am. Chem. Soc*, **41** (6), 868- 934.
- [26] Clark J., Edexcel IGCSE CHEMISTRY Syllabus. Editions: S.L: Pearson, 2009.
- [27] Jahn T.R., Radford S.E., 2008, Folding versus Aggregation: Polypeptide conformations on Competing Pathway. *Arch. Biochem. Biophys*, **469**(1), 100-117.
- [28] Creighton T. E., Proteins: Structure and Molecular Properties, New – York: Freeman W. H, 2nd edition, 1993.
- [29] Isaacs E.D., Shukla A., Platzman P .M., Hamann D.R., Barbielleini B.R., Tulk C.A., 1999, Covalency of the Hydrogen Bond in Ice: A direct X-Ray Measurement. *Phys. Rev. Lett*, **82**, 600-603.

- [30] Schulz G.E., Schirmer R.H., Principles of Protein Structure. New York, Berlin: Springer-Verlag, 1979.
- [31] Li H., Carter J.D., LaBean T.H., 2009, Nanofabrication by DNA Self-Assembly. *Mater. Today*, **12**, 5.
- [32] Pauling L., Corey R.B., 1951, Configurations of Polypeptide Chains with Favoured Orientations Around Single Bonds: Two Pleated Sheets. *Proc. Natl. Ac. Sci. USA*, **37**, 729-740.
- [33] Lin M.S., Chen L.Y., Tsai H.T., Wang S.S., Chang Y., Higuchi A., Chen W.Y., 2008, Investigation of the Mechanism of Beta-Amyloid Fibril Formation by Kinetic and Thermodynamics Analyses. *Langmuir*, **24**, 5802-5808.
- [34] Tang, Y., Zhu W., Chen K., Jiang H., 2006, New technologies in computer-aided drug design: Toward target identification and new chemical entity discovery. *Drug. Discov. Today*, **3** (3), 307-313.
- [35] Kitchen D. B., Decornez H., Furr J. R., Bajorath J., 2004, Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov*, **3** (11), 935-949.
- [36] Brooijmans N., Kuntz I. D., 2003, Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct*, **32**, 335-373.
- [37] Leach A.R., Molecular Modelling: Principles and Applications, New Jersey: Prentice Hall, 2001.
- [38] Mala J.G.S., Tekeuchi S., 2008, Understanding structural features of microbial lipases-an overview. *Anal. Chem. Insights*, **3**, 9-19.
- [39] Toulhoat H., 2007, Modélisation Moléculaire –Bases théoriques : des propriétés microscopiques aux propriétés macroscopiques. *Techn. Ing*, 1013.
- [40] Joseph-McCarthy D., 1999, Computational approaches to structure-based ligand design. *Pharmacol. Ther*, **84**, 179-191.
- [41] Lipinski C.A., Lombardo F., Dominy B.W., Feeny P.J., 2001, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev*, **46**, 3-26.
- [42] Damborsky J., Brezovsky J., 2009, Computational tools for designing and engineering biocatalysts. *Curr. Opin. Chem. Biol*, **13**, 26-34.
- [43] Mendez R., Leplae R., Lensink M. F., Wodak S. J., 2005, Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures. *Proteins: Struct. Funct. Bioinf*, **60** (2), 150-169.
- [44] Katchalski-Katzir E., Shariv I., Eisenstein M., Friesem A. A., Aflalo C., Vakser I. A., 1992, Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci U S A*, **89**, 2195-2199.

- [45] Gray J. J., Moughon S., Wang C., Schueler-Furman O., Kuhlman B., Rohl C. A., Baker D., 2003, Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol*, **331**, 281-299.
- [46] Simonson T., Archontis G., Karplus M., 2002, Free energy simulations come of age: protein-ligand recognition. *Acc. Chem. Res*, **35** (6), 430-437.
- [47] Ferrara P., Gohlke H., Price D. J., Klebe G., Brooks III, C. L., 2004, Assessing scoring functions for protein-ligand interactions. *J. Med. Chem*, **47**, 3032-3047.
- [48] Wang R., Lu Y., Fang X., Wang S., 2004, An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes. *J. Chem. Inf. Comput. Sci*, **44** (6), 2114-2125.
- [49] Kramer B., Rarey M., Lengauer T., 1999, Evaluation of the FlexX incremental construction algorithm for protein-ligand docking. *Proteins: Struct, Funct, Genet*, **37** (2), 228-241.
- [50] Goodsell D. S., Morris G. M., Olson A. J., 1996, Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit*, **9** (1), 1-5.
- [51] Krammer A., Kirchhoff P. D., Jiang X., Venkatachalam C. M., Waldman M., 2005. LigScore: a novel scoring function for predicting binding affinities. *J.Mol. Graph. Model*, **23** (5), 395-407.
- [52] Gohlke.H., Hendlich. M., Klebe.G., 2000, Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol*, **295** (2), 337-356.
- [53] Charifson P. S., Corkery J. J., Murcko M. A., Walters W. P., 1999, Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem*, **42** (25), 5100-5109.
- [54] Feher M., 2006, Consensus scoring for protein-ligand interactions. *Drug Discov. Today*, **11** (9-10), 421- 428.
- [55] Venkatachalam C. M., Jiang X., Oldfield T., Waldman M., 2003, LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graph. Model*, **21** (4), 289-307.
- [56] McGann M.R., Almond H.R, Nicholls A., Grant J.A., Brown F.K., 2003, Gaussian Docking Functions. *Biopolymers*, **68**, 76-90.
- [57] Ghosh. S., Nie. A., An, J., Huang, Z., 2006, Structure-based virtual screening of chemical libraries for drug discovery. *Curr. Opin. Chem. Biol*, **10** (3), 194-202.
- [58] Kuntz I.D., Blaney J.M., Oatley S.J., Langridge R., Ferrin T.E., 1982, A Geometric Approach to Macromolecule-Ligand Interactions. *J Mol Biol*, **161**, 269-288.

- [59] Mozziconacci J.C., Arnoult E, Bernard P., Do Q.T., Marot C., Morin-Allory L., 2005, Optimization and Validation of a Docking-scoring Protocol; Application to Virtual Screening for Cox-2 Inhibitors. *J Med Chem*, **48**, 1055-1068.
- [60] Makino S., Kuntz. I.D., 1997, Automated Flexible Ligand Docking Method and Its Application for Database Search. *J Comput Chem*, **18**, 1812-1825.
- [61] Jain A.N., 2003, Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. *J Med Chem*, **46**, 499-511.
- [62] Welch W., Ruppert J., Jain A.N., 1996, Hammerhead: Fast, Fully Automated Docking of Flexible Ligands to Protein Binding Sites, *Chem Biol*, **3**, 449-462.
- [63] Rarey M., Kramer B., Lengauer T., Klebe G., 1996, A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J Mol Biol*, **261**, 470-489.
- [64] Böhm H.J., 1992, The Computer Program Ludi: A New Method for the De Novo Design of Enzyme Inhibitors. *J Comput Aided Mol Des*, **6**, 61-78.
- [65] Rarey M., Kramer B., Lengauer T., 1999, Docking of hydrophobic ligands with interaction-based matching algorithms. *Bioinformatics*, **15** (3): 243-250.
- [66] Rarey M., Wefing S., Lengauer T., 1996, Placement of medium-sized molecular fragments into active sites of proteins. *J. Comput. Aided Mol. Des.*, **10**, 41-54.
- [67] Linnainmaa S., Harwood D., Davis L.S., 1988, Pose determination of a three dimensional object using triangle pairs. *IEEE Trans. Pattern Anal. Mach. Intell*, **10** (5), 634-646.
- [68] Kabsch W., 1976, A solution for the best rotation to relate two sets of vectors. *Acta. Cryst. Sect. A*, **32**, 922-923.
- [69] Jones G., Willett P., Glen R.C., Leach A.R., Taylor R., 1997, Development and validation of genetic algorithm for flexible docking. *J. Mol. Biol*, **267**, 727-748.
- [70] Baxter C.A., Murray C.W., Clark D.E., Westhead D.R., Eldridge M.D., 1998. Flexible Docking Using Tabu Search and an Empirical Estimate of Binding Affinity. *Proteins*, **33**, 367-382.
- [71] Wu G., Vieth M., 2004, Sdock: A Method Utilizing Existing X-Ray Structures to Improve Docking Accuracy. *J Med Chem*, **47**, 3142-3148.
- [72] Hart T.N., Read R.J., 1992, A Multiple-Start Monte Carlo Docking Method. *Proteins*, **13**, 206-222.
- [73] Abagyan R., Totrov M., 1994. Biased Probability Monte Carlo Conformational Searches and Electrostatic Calculations for Peptides and Proteins, *J Mol Biol*, **235**, 983- 1002.

- [74] Bursulaya B.D., Trotov M., Abagyan R., Brooks III C.L., 2004, Comparative study of several algorithms for flexible ligand docking. *J. Comput-Aided Mol. Des*, **17**, 755-763.
- [75] Bissantz C., Folkers G., Rognan D., 2000, Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem*, **43**, 4759- 4767.
- [76] Clark R. D., Strizhev A., Leonard J. M., Blake J. F., Matthew J. B., 2002, Consensus scoring for ligand/protein interactions. *J. Mol. Graph. Model*, **20**, 281-295.
- [77] Stahl M., Rarey M., 2001, Detailed analysis of scoring functions for virtual screening. *J. Med. Chem*, **44**, 1035-1042.
- [78] Moitessier N., Englebienne P., Lee D., Lawandi J., Corbeil C.R, 2008, Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol*, **153**(S1), S7–S26
- [79] Digby G.N., Kempton R.A., 1987, *Multivariate Analysis of Ecological Communities*, Chapman and Hall, London, 112–115.
- [80] Rajarshi G., Michael T. H., Geoffrey R. H., Peter M.R., Henry R., Christoph S., Jörg K. W., Egon L. W., 2006, The Blue Obelisk-Interoperability in Chemical Informatics. *J. Chem. Inf. Model*, **46** (3), 991–998
- [81] Cole J. C., Murray C. W., Willem J., Nissink M., Taylor R. D., Taylor R., 2005 , Comparing protein-ligand docking programs is difficult.. *Proteins*, **60**, 325-332.
- [82] Jun M., Bin C., Identification of Falcipain-2 binding sites using docking methods. Report for I619 final project,.2009
- [83] TRIPOS Associates, Inc St. Louis Missouri, USA. 1994, SYBYL Molecular Modeling Software, Version 6.x,
<http://www.chem.ac.ru/Chemistry/Soft/SYBYL.en.html>
- [84] Rarey M., Kramer B., Lengauer T., 1997, Multiple automatic base selection: protein-ligand docking based on incremental construction without manual intervention. *J Comput. Aided Mol Des*, **11**, 369–384.
- [85] Kellenberger E., Rodrigo J., Muller P., Rognan D., 2004, Comparative evaluation of eight docking tools for docking and virtual screening accuracy, *Proteins*, **57**(2), 225–242.
- [86] Verdonk M.L., Cole J.C., Hartshorn M.J., Murray C.W., Taylor R.D., 2003, Improved protein–ligand docking using GOLD. *Proteins*, **52**, 609–623.
- [87] Nissink J.W.M., Murray C.W., Hartshorn M.J., Verdonk M.L., Cole J.C., Taylor R.A., 2002, New test set for validating predictions of protein– ligand interaction. *Proteins*, **49**, 457–471.

- [88] Zaheer U.H., Sobia A.H, Reaz U., Jeffry D., Madurab B., 2010, Benchmarking docking and scoring protocol for the identification of potential acetyl cholinesterase inhibitors, *J. Mol. Graph. Model*, **28**, 870–882.
- [89] Michael L., Christopher E., Stefan P., 2010, Comparison of current docking tools for the simulation of inhibitor binding by the transmembrane domain of the sarco/endoplasmic reticulum calcium ATPase. *Biophys. Chem*, **150**, 88-97.
- [90] Kontoyianni M., McClellan L.M., Sokol G.S., 2004, Evaluation of docking performance: comparative data on docking algorithms, *J. Med. Chem*, **47**, 558–565.
- [91] Chikhi A., Bensegueni A., 2008, Docking Efficiency Comparison of Surflex, a Commercial Package and Arguslab, a Licensable Freeware. *J Comput Sci Syst Biol*, **1**, 081-086.
- [92] Klebe G., 2006, Virtual ligand screening: strategies, perspectives and limitations. *Drug. Discov. Today*, **11**, 580–594.
- [93] Badry D. B., Maxim T., Ruben A., Brooks III. C.L., 2003. Comparative study of several algorithms for flexible ligand docking. *J.Comput-Aided. Mol. Des*, **17**, 755–763.
- [94] Becue A., Meurice N., Leherte L., Vercauteren D.P.,. Protein-Protein Docking Using Three-Dimensional Reduced Representations and Based on a Genetic Algorithm.. Boeyens J.C.A and Ogilvie J.F. *Models, Mysteries, and Magic of Molecules*, Springer: Guildford, 2008, 301-324.
- [95] Verkhivker G.M., Bouzida D., Gehlhaar D.K., Rejto P.A, Arthurs S., Colson A.B., Freer S.T., Larson V., Luty B.A., Marrone T., Rose P.W., 2000, Deciphering common failures in molecular docking of ligand–protein complexes, *J. Comput.-Aided. Mol. Des*, **14**, 731–751.
- [96] Roberts B.C., Mancera R.L., 2008, Ligand-protein docking with water molecules, *J.Chem. Inf. Model*, **48**, 397–408.
- [97] Oxford J. S., 2000, Influenza A pandemics of the 20th century with special reference to 1918: virology, pathology and epidemiology. *Rev Med Virol*, **10** (2), 19-33.
- [98] Formosa S., Episodes de grippe aviaire à Hong Kong en 1997 et 1999 : conséquences épidémiologiques. Thèse Méd. Vét., Toulouse, 2004, n° 42, 96.
- [99] Ebrahim G. J., 2004, Avian Flu and Influenza Pandemics in Human Populations. *J. Trop. Pediatr*, **50**, (4), 192-4.
- [100] Fouchier R., Munster V., Wallensten A., 2005, Characterization of a novel Influenza A virus Hemagglutinin subtype (H16) obtained from Black-Headed Gulls. *J. Virol*, **79**, 2814-22.
- [101] Ferraris O., Escuret V., Bouscambert-Duchamp M., Lina B., Morfin F., 2010, Intérêts des inhibiteurs de la neuraminidase dans la prise en charge des infections dues aux virus influenza. *Pathol. Biol*, **58**, 69–78.

- [102] Air G.M., Laver W.G., 1989. The neuraminidase of influenza virus. *Proteins*, **6**, 341–56.
- [103] Varghese J.N., Laver W.G., Colman P.M., 1983, Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature*, **303**, 35–40.
- [104] Moscona A., 2005, Neuraminidase inhibitors for influenza. *N Engl J Med*, **353**, 1363–73.
- [105] Taylor N.R., Von Itzstein M., 1994, Molecular modeling studies on ligand binding to sialidase from influenza virus and the mechanism of catalysis. *J Med Chem*, **37**, 616–24.
- [106] Nzamba. T., L'influenza aviaire A, virus A (H5N1) chez les oiseaux domestiques et sauvages. Thèse de doctorat, Université Henri Poincaré –Nancy 1, 2009, 161.
- [107] Sidwell R.W., Smee D.F., Huffman J.H., Barnard D.L., Morrey J.D., Bailey K.W., Feng W.C., Babu Y.S., Bush K., 2001, Influence of virus strain, challenge dose, and time of therapy initiation on the in vivo influenza inhibitory effects of RWJ- 270201. *Antiviral Res*, **51**, 179-187.
- [108] Wagner E.K., Hewlett M.J., Bloom D.C., Camerini D., Basic virology. Malden: Blackwell Publishing, 3rd ed, 2008, 550.
- [109] An J., Lee D.C., Law A.H., Yang C.L., Poon L.L., Lau A.S., Jones S.J., 2009. A novel small molecule inhibitor of the avian influenza H5N1 virus determined through computational screening against the neuraminidase. *J. Med. Chem.*, **52**, 2667–2672.
- [110] National Center For Biotechnology Information. (NCBI). PubChem [en ligne]. Date de création 2004.<http://pubchem.ncbi.nlm.nih.gov/> consulté le 10 Août 2011

ANNEXE1. Résultats-FlexX

Tableau 1.1 Résultats du *docking* par FlexX 1.3.0 : RMSD et Liais-Rotables

N°	Ligand	Score	Match	Lipo	Ambig	Clash	Rot	RMSD	Liais-Rotables
1	1JLX	-1.3913	-10.3676	-3.6031	-3.8348	2.6143	8.4000	0.8035	6
2	1JSM	-0.2456	-9.3221	-2.0649	-3.9625	2.7039	7.0000	4.8550	5
3	1JSN	-13.0303	-21.7657	-5.0586	-6.4792	2.2732	12.6000	0.6093	9
4	1JSO	-3.7493	-17.8970	-5.3666	-5.8606	5.9749	14.0000	1.1776	10
5	1KJ1	-7.6876	-16.5719	-1.8499	-3.8125	0.7466	8.4000	1.1571	6
6	1OUW	-9.4498	-12.6731	-2.4588	-3.9179	0.0000	4.2000	0.3018	3
7	1QFU	-7.3764	-14.1773	-3.5318	-4.1438	3.4764	5.6000	4.4262	4
8	1RD8	-0.0170	-7.5209	-2.6205	-4.2892	3.4476	5.6000	5.1504	4
9	1RUZ	-2.5657	-14.7150	-1.7468	-3.0202	3.1163	8.4000	2.6885	6
10	1RV0	-2.9395	-11.8625	-2.1671	-3.8333	1.1234	8.4000	4.7399	6
11	1RVT	-1.3786	-13.3623	-3.5370	-3.9485	1.4692	12.6000	1.3980	9
12	1SBD	-9.4049	-17.9877	-4.6660	-5.3547	6.2035	7.0000	1.4575	5
13	1SBF	-14.2197	-18.1653	-4.8999	-4.4115	2.2570	5.6000	0.9996	4
14	1WGC	-5.7047	-8.0135	-1.6216	-3.6398	0.9795	12.6000	4.8324	9
15	2CCV	-6.6959	-12.9595	-2.8512	-5.8262	1.1410	8.4000	0.5462	6
16	2CGZ	-4.7365	-12.5730	-3.5505	-7.4124	0.7993	12.6000	0.8031	9
17	2CWG	-1.8578	-11.7853	-3.0355	-3.3037	1.9823	12.6000	5.4625	9
18	2DV9	-8.9503	-14.9436	-3.4753	-3.0403	1.5089	5.6000	0.9483	4
19	2DVD	-7.8540	-12.9102	-4.7070	-3.6809	1.0442	7.0000	0.3851	5
20	2IBX	-6.7231	-12.4764	-4.5437	-4.2646	2.1615	7.0000	4.0003	5
21	2Q3N	-2.3253	-10.6886	-2.1550	-3.5872	1.7055	7.0000	3.6720	5
22	2WR0	-1.5579	-8.9275	-2.9562	-3.4457	4.4873	7.0000	5.0130	5
23	2X3T	-4.8963	-12.0007	-6.1265	-4.6869	2.7178	9.8000	0.8383	7
24	2X52	-13.8477	-36.5851	-14.1958	-15.0080	7.3412	39.2000	1.8968	28
25	2ZB5	-0.2879	-10.2665	-2.0316	-2.9463	2.5566	7.0000	4.9487	5
26	2ZB6	-1.3564	-9.5537	-3.3506	-4.9076	4.0554	7.0000	4.9820	5
27	2ZR1	-2.7024	-16.2974	-3.2141	-3.8024	8.2115	7.0000	4.5561	5
28	3AH1	-7.4038	-14.1186	-8.2703	-6.7945	2.3797	14.0000	0.4713	10
29	3AH4	-1.8655	-13.1723	-3.5906	-3.6167	4.7141	8.4000	1.0689	6
30	3BT6	-0.3095	-6.4538	-2.7969	-3.9719	0.5131	7.0000	5.8572	5
31	3C9Z	-0.3109	-10.7333	-1.7244	-3.5818	3.3287	7.0000	4.5549	5
32	3CA0	-3.0695	-10.0491	-3.6270	-4.9264	3.1330	7.0000	4.4973	5
33	3CA3	-1.2028	-11.4977	-4.0996	-5.1903	5.7847	8.4000	2.2282	6
34	3CA4	-1.6556	-3.8909	-1.2225	-1.0088	2.3778	0.0000	2.5439	0
35	3CA5	-3.9489	-0.8924	-0.5963	-0.4318	0.4694	0.0000	2.6738	0
36	3CA6	-8.7037	-10.3455	-5.0662	-6.4078	0.7158	7.0000	0.4560	5
37	3EYK	-0.5465	-6.6854	-3.0903	-1.4705	1.0997	4.2000	1.9558	3
38	3EYM	-3.0062	-6.9837	-5.5381	-2.3750	2.2906	4.2000	1.1119	3
39	3INB	-1.8000	-12.6206	-2.2702	-4.3743	5.0651	7.0000	4.7412	5
40	3LLZ	-8.1228	-13.2440	-6.8064	-4.8105	2.9382	8.4000	0.7043	6
41	1A4G	-57.8843	-63.5301	-5.1697	-8.9701	3.1856	11.2000	0.5634	8
42	1A4Q	-42.0860	-48.8992	-7.9262	-6.5238	6.0631	9.8000	1.5320	7
43	1A14	-2.7944	-6.9129	-1.6059	-3.0650	1.9783	7.0000	5.8771	5
44	1B9S	-25.1269	-27.4180	-5.4302	-6.2523	4.3736	4.2000	0.9639	3
45	1B9T	-41.8714	-45.0956	-5.7295	-8.0750	6.0287	5.6000	0.8333	4
46	1B9V	-19.3453	-28.2181	-3.6157	-7.4131	4.7016	9.8000	5.4444	7
47	1E8V	-17.7084	-32.3125	-3.6852	-5.8853	6.1746	12.6000	2.0839	9
48	1EUS	-43.0313	-49.0267	-6.2085	-7.7382	1.9420	12.6000	0.8306	9
49	1EUU	-1.0572	-9.1387	-1.9374	-4.6010	0.8199	8.4000	2.2632	6
50	1F8B	-30.7520	-38.7798	-4.3415	-6.3429	2.1121	11.2000	0.6382	8
51	1INF	-33.6797	-36.1801	-4.5642	-5.5974	7.2620	0.0000	1.0814	0

52	1ING	-15.7017	-15.0830	-4.7808	-6.1794	2.1416	2.8000	0.6784	2
53	1INH	-0.7225	-9.0302	-4.7547	-4.6475	5.3098	7.0000	4.1205	5
54	1IVC	-9.2081	-14.8320	-6.2331	-5.0397	4.4967	7.0000	5.8447	5
55	1IVD	-24.8314	-28.7590	-3.6059	-5.6316	6.3652	1.4000	1.0844	1
56	1IVE	-0.7766	-10.2227	-1.9386	-4.2063	3.1910	7.0000	6.2100	5
57	1IVG	-0.9625	-5.4883	-5.0887	-5.1242	4.2637	7.0000	5.8227	5
58	1L7F	-53.4371	-58.3759	-6.0679	-8.8692	3.2759	11.2000	0.9329	8
59	1L7H	-44.8914	-48.2240	-6.1820	-8.8077	1.7223	11.2000	0.9268	8
60	1MWE	-38.3407	-48.0678	-5.0538	-6.2310	1.6119	14.0000	0.7017	10
61	1NCA	-0.3900	-10.1271	-4.5403	-4.5601	6.4375	7.0000	5.6977	5
62	1NCB	-7.6198	-13.7971	-4.3996	-5.0167	3.1935	7.0000	5.2400	5
63	1NCC	-1.2237	-7.3042	-3.3454	-3.6900	3.1633	7.0000	5.3136	5
64	1NCD	-3.2121	-6.0832	-4.4203	-6.1846	7.5001	7.0000	5.4867	5
65	1NMA	-16.3112	-24.7561	-1.2385	-3.2447	0.5282	7.0000	4.0455	5
66	1NMB	-1.6731	-6.7520	-2.4396	-3.6331	2.0978	7.0000	4.2709	5
67	1NMC	-1.5371	-9.6475	-3.7506	-4.3076	3.7686	7.0000	5.0032	5
68	1NN2	-1.5408	-6.1834	-4.1864	-3.9866	3.4972	7.0000	2.0267	5
69	1NNB	-22.7784	-33.3567	-3.9901	-6.7430	3.3114	12.6000	1.3363	9
70	1NNC	-24.8110	-39.0732	-3.7429	-7.4887	8.8939	11.2000	4.3722	8
71	1NSB	-0.8527	-8.1357	-2.1055	-3.5228	0.5113	7.0000	5.5154	5
72	1NSC	-44.9033	-57.4987	-4.2011	-8.2803	5.6768	14.0000	1.3867	10
73	1NSD	-3.5529	-8.5039	-3.6517	-4.7498	8.0583	7.0000	5.0998	5
74	1SLI	-44.7991	-44.8582	-4.5866	-8.4525	3.4982	4.2000	0.7187	3
75	1USR	-48.0466	-55.7629	-3.2926	-7.9844	0.9934	12.6000	0.8429	9
76	1USX	-27.1033	-39.3345	-4.1875	-5.8352	4.2539	12.6000	1.8420	9
77	1V2I	-7.5027	-12.6337	-5.3788	-6.0552	4.1650	7.0000	4.9296	5
78	1V3B	-15.5139	-19.7104	-4.5598	-5.6336	1.9899	7.0000	4.3053	5
79	1V3C	-20.3198	-31.8865	-3.1524	-6.7294	2.0485	14.0000	1.3839	10
80	1V3D	-27.0519	-36.0337	-4.0329	-6.8842	3.2990	11.2000	0.8187	8
81	1VCJ	-14.5115	-16.1026	-7.7913	-8.5084	4.0908	8.4000	1.2486	6
82	1W0O	-11.5686	-24.3829	-3.6007	-5.8740	2.8891	14.0000	0.5753	10
83	1W0P	-6.8241	-10.4726	-2.5418	-2.6839	0.6742	2.8000	3.6102	2
84	1W8N	-53.1499	-61.5953	-6.1297	-7.5801	2.7552	14.0000	0.3190	10
85	1W8O	-3.0602	-12.4661	-0.9432	-2.4596	0.4088	7.0000	1.5875	5
86	1WCQ	-57.1538	-63.4265	-6.5066	-8.8753	2.2546	14.0000	0.4435	10
87	1XOE	-27.0142	-36.6747	-5.1870	-6.1629	5.8104	9.8000	1.9639	7
88	1XOG	-2.4443	-10.9841	-1.4863	-4.1763	1.8025	7.0000	3.0828	5
89	1Z4Y	-3.8843	-8.8675	-4.4503	-4.4869	1.5205	7.0000	4.3124	5
90	2AEP	-0.8201	-9.2754	-3.2448	-4.4424	3.7425	7.0000	5.3965	5
91	2AEQ	-0.9944	-8.7113	-2.2739	-3.1348	0.7257	7.0000	3.0760	5
92	2BAT	-26.4413	-34.6186	-5.1246	-8.2487	2.1506	14.0000	0.4221	10
93	2BER	-50.6255	-54.8018	-5.8339	-9.6329	1.6431	12.6000	0.5231	9
94	2HT5	-1.6854	-8.5305	-1.9727	-2.3881	0.7766	8.4000	3.9616	6
95	2HT7	-33.3935	-38.2744	-5.5882	-7.2124	5.2814	7.0000	0.6663	5
96	2HTQ	-46.0519	-53.8984	-4.6737	-8.2761	4.1963	11.2000	0.7537	8
97	2HTU	-41.0844	-45.5565	-5.9296	-10.2679	4.0696	11.2000	1.0395	8
98	2HTV	-1.6274	-11.8304	-1.7710	-3.4147	1.5886	8.4000	4.8837	6
99	2QWA	-1.2352	-8.6871	-2.0849	-3.0728	2.6800	7.0000	4.8145	5
100	2QWB	-25.6598	-37.0641	-3.5040	-6.2182	1.7266	14.0000	1.4201	10
101	2QWC	-16.8204	-29.9145	-3.1998	-5.1358	4.8297	11.2000	2.0932	8
102	2QWD	-43.2413	-49.7406	-4.2576	-6.5025	2.0594	9.8000	0.4197	7
103	2QWE	-34.7962	-40.9091	-5.9316	-7.8682	3.3126	11.2000	1.4264	8

104	2QWF	-33.5361	-38.3247	-6.7270	-8.3612	6.0769	8.4000	0.9006	6
105	2QWG	-36.5125	-38.0300	-5.6305	-6.9823	3.1302	5.6000	0.9299	4
106	2QWK	-52.5465	-53.1143	-7.0752	-7.4986	2.7416	7.0000	0.5337	5
107	2SIM	-36.7901	-45.4672	-4.8510	-8.1535	5.0815	11.2000	1.6102	8
108	2SLI	-41.4108	-46.5459	-4.8854	-8.6580	3.4785	9.8000	0.5494	7
109	2VVZ	-37.3142	-45.8381	-4.5202	-8.6953	5.1393	11.2000	0.7872	8
110	2VW0	-1.3351	-9.4000	-1.1644	-1.8929	1.3924	7.0000	2.5720	5
111	2VW1	-47.9040	-50.3478	-5.8501	-10.5436	2.2375	11.2000	0.9340	8
112	2VW2	-3.7417	-14.1000	-0.8799	-2.1414	0.9795	7.0000	1.7883	5
113	2W20	-3.4648	-10.9441	-1.4919	-2.1378	1.5091	4.2000	3.9859	3
114	3B7E	-52.9288	-53.0041	-6.4916	-11.9651	1.9319	11.2000	0.5301	8
115	3BEQ	-0.5429	-8.1071	-1.3411	-2.8640	0.4551	7.0000	2.8924	5
116	3CKZ	-64.1348	-70.7955	-5.4483	-9.9342	4.0432	12.6000	0.5837	9
117	3CL0	-25.9751	-30.1110	-5.7140	-6.7830	2.8329	8.4000	1.0667	6
118	3GVL	-4.0736	-12.6445	-1.6469	-2.2465	1.2115	14.0000	1.0853	10
119	3SLI	-49.8199	-53.2398	-5.3740	-8.1026	1.6965	9.8000	0.5732	7
120	5NN9	-0.4553	-8.1462	-1.9930	-3.1433	1.3379	7.0000	5.0118	5
121	6NN9	-8.2332	-2.8571	-1.9802	-1.8914	2.5619	7.0000	4.5349	5
122	1BS4	-6.4327	-34.5581	-8.7187	-9.7225	3.3667	37.8000	8.5784	27
123	1BSJ	-11.3794	-19.6883	-11.2738	-9.7866	7.1693	16.8000	4.1711	12
124	1BSK	-4.5224	-21.2122	-9.9822	-6.9576	11.4295	16.8000	4.6192	12
125	1BSZ	-17.4569	-41.9073	-11.6358	-11.1705	4.0566	37.8000	5.1791	27
126	1G27	-17.4761	-26.7536	-9.4497	-8.5201	7.8473	14.0000	1.2388	10
127	1ICJ	-15.7475	-43.8005	-10.4739	-11.9223	7.2492	37.8000	2.6432	27
128	1IX1	-19.3833	-21.2129	-13.2875	-10.2703	4.5875	15.4000	1.4660	11
129	1LM4	-8.2934	-9.9944	-1.5899	-3.3276	1.2185	0.0000	4.2912	0
130	1LM6	-6.2701	-8.7356	-1.3014	-2.8049	1.1718	0.0000	4.5354	0
131	1LQY	-24.7276	-31.4642	-9.2882	-10.2688	5.4937	15.4000	0.6998	11
132	1LRU	-23.4196	-26.7109	-9.5854	-11.0334	3.1101	15.4000	0.8754	11
133	1LRY	-25.7263	-32.7193	-9.1334	-11.6561	5.5824	16.8000	1.2462	12
134	1N5N	-8.9602	-16.7844	-2.1990	-3.6636	1.2868	7.0000	4.1225	5
135	1Q1Y	-20.8071	-25.8041	-10.7369	-10.8901	5.8240	15.4000	2.1163	11
136	1RL4	-14.6438	-18.8535	-7.0284	-6.5708	2.6090	9.8000	0.9242	7
137	1SV2	-2.5388	-7.1145	-5.2665	-5.7641	3.2062	7.0000	3.6864	5
138	1SZZ	-24.0997	-28.5736	-11.2001	-9.4373	4.3113	15.4000	1.1267	11
139	1VEV	-12.5170	-12.7984	-7.2731	-6.5390	4.4934	4.2000	2.0129	3
140	1VEZ	-13.5994	-18.2362	-2.8797	-5.2947	3.2113	4.2000	1.1735	3
141	1WS1	-21.5571	-28.6355	-10.7396	-10.7931	6.4111	16.8000	0.8431	12
142	1Y6H	-8.9292	-12.5147	-0.9704	-1.7138	0.8697	0.0000	1.4071	0
143	2AI7	-6.9067	-12.6900	-6.7373	-5.2036	5.3242	7.0000	1.2040	5
144	2AI8	-8.9854	-12.7157	-7.9152	-5.6390	4.8844	7.0000	1.0746	5
145	2EW6	-31.3725	-28.3724	-12.6559	-8.9902	4.8460	8.4000	0.9230	6
146	2OKL	-13.5856	-23.9304	-6.3461	-9.0781	4.9689	15.4000	4.2822	11
147	2OS3	-7.8760	-21.7600	-7.0099	-8.4092	8.5031	15.4000	3.8596	11
148	3CMD	-15.2041	-18.1021	-1.8190	-3.9910	0.5079	2.8000	0.9729	2
149	3E3U	-12.4358	-16.1209	-12.6255	-7.1225	6.8331	11.2000	0.6694	8
150	3G5K	-18.6823	-26.1403	-10.6033	-10.1618	6.0230	16.8000	0.6784	12
151	3K6L	-27.4220	-33.3819	-7.6383	-9.4693	5.0675	12.6000	1.2544	9
152	3NRB	-0.6848	-3.7076	-1.2679	-1.9578	0.8485	0.0000	1.3545	0
153	3O1L	-6.5554	-14.1000	-1.8578	-3.5650	0.5674	7.0000	1.6603	5

Tableau 1.2 Résultats du *docking* par FlexX 1.3.0 : Temps et Faux Positifs

N°	Ligand	Temps en secondes (Run Time)	% Faux Positifs
1	1JLX	29,52	0
2	1JSM	71,89	0
3	1JSN	150,5	10
4	1JSO	131,2	40
5	1KJ1	80,88	30
6	1OUW	45,45	20
7	1QFU	200,66	10
8	1RD8	140,45	20
9	1RUZ	79,2	30
10	1RV0	180,82	10
11	1RVT	200,62	0
12	1SBD	113,28	0
13	1SBF	161,22	20
14	1WGC	85,25	0
15	2CCV	47,28	0
16	2CGZ	150,44	0
17	2CWG	88,62	10
18	2DV9	171,2	20
19	2DVD	79,62	20
20	2IBX	149,41	10
21	2Q3N	231,85	0
22	2WR0	88,15	10
23	2X3T	51,2	30
24	2X52	211,86	10
25	2ZB5	145,62	0
26	2ZB6	81,35	20
27	2ZR1	159,21	20
28	3AH1	97,21	30
29	3AH4	209,91	10
30	3BT6	123,44	20
31	3C9Z	135,37	20
32	3CA0	141,04	0
33	3CA3	231,22	10
34	3CA4	167,28	10
35	3CA5	79,22	20
36	3CA6	148,33	0
37	3EYK	64,12	30
38	3EYM	144,26	40
39	3INB	78,59	10
40	3LLZ	90,25	10
41	1A4G	79,44	20
42	1A4Q	153,49	10
43	1A14	224,52	0
44	1B9S	79,15	0
45	1B9T	82,46	20

46	1B9V	103,93	20
47	1E8V	70,44	10
48	1EUS	187,3	30
49	1EUU	49,61	10
50	1F8B	79,85	0
51	1INF	207,87	0
52	1ING	135,65	10
53	1INH	81,48	10
54	1IVC	149,04	0
55	1IVD	57,51	20
56	1IVE	115,29	0
57	1IVG	159,66	10
58	1L7F	107,33	0
59	1L7H	94,56	20
60	1MWE	122,42	40
61	1NCA	98,17	30
62	1NCB	118,14	10
63	1NCC	37,7	0
64	1NCD	216,91	0
65	1NMA	156,2	20
66	1NMB	79,72	0
67	1NMC	154,86	10
68	1NN2	62,45	0
69	1NNB	159,07	20
70	1NNC	166,39	10
71	1NSB	52,42	10
72	1NSC	148,63	30
73	1NSD	43,95	10
74	1SLI	132,75	0
75	1USR	153,94	0
76	1USX	195,66	0
77	1V2I	72,63	0
78	1V3B	64,35	10
79	1V3C	135,94	20
80	1V3D	121,09	30
81	1VCJ	176,89	0
82	1W0O	117,27	10
83	1W0P	55,62	40
84	1W8N	182,29	10
85	1W8O	28,12	0
86	1WCQ	323,91	0
87	1XOE	63,36	0
88	1XOG	36,58	20
89	1Z4Y	51,57	0
90	2AEP	51,52	10
91	2AEQ	56,07	0
92	2BAT	153,49	10
93	2BER	133,83	20
94	2HT5	84,24	10

95	2HT7	229,28	0
96	2HTQ	148,72	0
97	2HTU	78,88	0
98	2HTV	66,96	0
99	2QWA	31,68	10
100	2QWB	179,5	0
101	2QWC	141,88	0
102	2QWD	175,05	0
103	2QWE	150,88	30
104	2QWF	212,44	10
105	2QWG	142,42	0
106	2QWK	216,76	0
107	2SIM	147,28	0
108	2SLI	149,33	20
109	2VVZ	108,27	10
110	2VW0	65,65	10
111	2VW1	132,88	20
112	2VW2	38,96	10
113	2W20	19,39	20
114	3B7E	135,22	0
115	3BEQ	43,42	20
116	3CKZ	135,88	10
117	3CL0	76,59	20
118	3GVL	96,39	10
119	3SLI	140,44	0
120	5NN9	22,27	10
121	6NN9	20,02	0
122	1BS4	99,85	0
123	1BSJ	107,77	0
124	1BSK	101,2	20
125	1BSZ	104,4	10
126	1G27	115,16	0
127	1ICJ	94,32	10
128	1IX1	82,58	0
129	1LM4	12,91	20
130	1LM6	74,92	30
131	1LQY	43,6	20
132	1LRU	96,93	10
133	1LRY	84,82	0
134	1N5N	17,32	0
135	1Q1Y	45,72	40
136	1RL4	50,62	20
137	1SV2	179,91	10
138	1SZZ	98,46	10
139	1VEV	86,85	20
140	1VEZ	47,34	50
141	1WS1	56,92	0
142	1Y6H	11,88	0
143	2AI7	22,45	0

144	2AI8	26,73	20
145	2EW6	106,42	10
146	2OKL	113,55	0
147	2OS3	76,36	0
148	3CMD	41,76	20
149	3E3U	70,88	10
150	3G5K	104,62	0
151	3K6L	86,98	0
152	3NRB	30,69	10
153	3O1L	36,81	0

Tableau 1.3 Résultats du *docking* par FlexX 1.3.0 en absence et en présence des molécules d'eau

N°	Ligand		Score	RMSD		Score	RMSD
1	1JLX	En absence des molécules d'eau	-1.3913	0.8035	En présence des molécules d'eau	-2.7458	0.6987
2	1JSM		-0.2456	4.8550		-1.5213	5.4238
3	1JSN		-13.0303	0.6093		-13.7699	0.6123
4	1JSO		-3.7493	1.1776		-3.8434	1.0065
5	1KJ1		-7.6876	1.1571		-7.5182	2.6770
6	1OUW		-9.4498	0.3018		-9.3206	0.3018
7	1QFU		-7.3764	4.4262		-3.1112	4.4262
8	1RD8		-0.0170	5.1504		-0.0170	5.1504
9	1RUZ		-2.5657	2.6885		-2.5657	2.6885
10	1RV0		-2.9395	4.7399		-1.3768	4.0199
11	1RVT		-1.3786	1.3980		-1.6678	1.6393
12	1SBD		-9.4049	1.4575		-9.4049	1.4575
13	1SBF		-14.2197	0.9996		-14.2197	0.9996
14	1WGC		-5.7047	4.8324		-5.7047	4.8324
15	2CCV		-6.6959	0.5462		-6.6959	0.5462
16	2CGZ		-4.7365	0.8031		-4.7365	0.8031
17	2CWG		-1.8578	5.4625		-2.2048	5.9492
18	2DV9		-8.9503	0.9483		-9.0325	0.9483
19	2DVD		-7.8540	0.3851		-21.4714	0.3011
20	2IBX		-6.7231	4.0003		-6.7231	4.0003
21	2Q3N		-2.3253	3.6720		-2.3253	3.6720
22	2WR0		-1.5579	5.0130		-2.3826	3.1875
23	2X3T		-4.8963	0.8383		-4.8963	0.8383
24	2X52		-13.8477	1.8968		-0.9833	2.5965
25	2ZB5		-0.2879	4.9487		-0.2879	4.9487
26	2ZB6		-1.3564	4.9820		-1.3564	4.9820
27	2ZR1		-2.7024	4.5561		-2.7024	4.5561
28	3AH1		-7.4038	0.4713		-7.4038	0.4713
29	3AH4		-1.8655	1.0689		-7.4390	1.1508
30	3BT6		-0.3095	5.8572		-0.3095	5.8572
31	3C9Z		-0.3109	4.5549		-0.3109	4.5549

32	3CA0		-3.0695	4.4973		-6.9677	5.9165
33	3CA3		-1.2028	2.2282		-4.9869	2.2769
34	3CA4		-1.6556	2.5439		-1.2282	2.5439
35	3CA5		-3.9489	2.6738		-0.8469	1.8191
36	3CA6		-8.7037	0.4560		-8.7252	0.4560
37	3EYK		-0.5465	1.9558		-0.0064	1.9558
38	3EYM		-3.0062	1.1119		-3.0062	1.1119
39	3INB		-1.8000	4.7412		-1.8000	4.7412
40	3LLZ		-8.1228	0.7043		-9.1148	1.1748
41	1A4G		-57.8843	0.5634		-64.0111	0.3218
42	1A4Q		-42.0860	1.5320		-51.7133	0.9168
43	1A14		-2.7944	5.8771		-2.7944	5.8771
44	1B9S		-25.1269	0.9639		-29.5086	1.1873
45	1B9T		-41.8714	0.8333		-42.0046	1.0587
46	1B9V		-19.3453	5.4444		-19.2397	5.4186
47	1E8V		-17.7084	2.0839		-17.7084	2.0839
48	1EUS		-43.0313	0.8306		-37.4903	0.7996
49	1EUU		-1.0572	2.2632		-1.0572	2.2632
50	1F8B		-30.7520	0.6382		-21.5823	0.7594
51	1INF		-33.6797	1.0814		-52.8734	0.6904
52	1ING		-15.7017	0.6784		-15.7017	0.6784
53	1INH		-0.7225	4.1205		-0.7225	4.1205
54	1IVC		-9.2081	5.8447		-9.2081	5.8447
55	1IVD		-24.8314	1.0844		-28.4090	1.1615
56	1IVE		-0.7766	6.2100		-0.7766	6.2100
57	1IVG		-0.9625	5.8227		-0.9242	5.8227
58	1L7F		-53.4371	0.9329		-54.5927	0.8173
59	1L7H		-44.8914	0.9268		-51.2964	1.1441
60	1MWE		-38.3407	0.7017		-39.7623	0.5280
61	1NCA		-0.3900	5.6977		-0.3900	5.6977
62	1NCB		-7.6198	5.2400		-7.6198	5.2400
63	1NCC		-1.2237	5.3136		-1.2237	5.3136
64	1NCD		-3.2121	5.4867		-3.2121	5.4867
65	1NMA		-16.3112	4.0455		-16.3112	4.0455
66	1NMB		-1.6731	4.2709		-1.6731	4.2709
67	1NMC		-1.5371	5.0032		-1.5371	5.0032
68	1NN2		-1.5408	2.0267		-1.5408	2.0267
69	1NNB		-22.7784	1.3363		-22.7784	1.3363
70	1NNC		-24.8110	4.3722		-58.9153	0.3979
71	1NSB		-0.8527	5.5154		-2.8646	4.9665
72	1NSC		-44.9033	1.3867		-44.9033	1.3867
73	1NSD		-3.5529	5.0998		-2.7759	5.0998
74	1SLI		-44.7991	0.7187		-47.1780	0.7187
75	1USR		-48.0466	0.8429		-48.0466	0.8429
76	1USX		-27.1033	1.8420		-27.1033	1.8420
77	1V2I		-7.5027	4.9296		-7.5027	4.9296
78	1V3B		-15.5139	4.3053		-15.5139	4.3053
79	1V3C		-20.3198	1.3839		-20.3198	1.3839
80	1V3D		-27.0519	0.8187		-27.0519	0.8187

En absence des molécules d'eau

En présence des molécules d'eau

81	1VCJ		-14.5115	1.2486		-14.5115	1.3562
82	1W0O		-11.5686	0.5753		-11.5686	0.5753
83	1W0P		-6.8241	3.6102		-7.4487	1.8384
84	1W8N		-53.1499	0.3190		-53.1499	0.3190
85	1W8O		-3.0602	1.5875		-0.8614	1.0645
86	1WCQ		-57.1538	0.4435		-57.1538	0.4435
87	1XOE		-27.0142	1.9639		-27.0384	1.9639
88	1XOG		-2.4443	3.0828		-2.4443	3.0828
89	1Z4Y		-3.8843	4.3124		-2.7355	4.9005
90	2AEP		-0.8201	5.3965		-10.1395	5.2604
91	2AEQ		-0.9944	3.0760		-0.9944	3.0760
92	2BAT		-26.4413	0.4221		-31.2856	0.6522
93	2BER		-50.6255	0.5231		-50.6255	0.5231
94	2HT5		1.6854	3.9616		-13.2770	2.9423
95	2HT7		-33.3935	0.6663		-33.3935	0.6663
96	2HTQ		-46.0519	0.7537		-54.3414	0.5957
97	2HTU		-41.0844	1.0395		-41.0844	1.0395
98	2HTV		-1.6274	4.8837		-7.9006	3.4291
99	2QWA		-1.2352	4.8145		-6.2246	5.2368
100	2QWB		-25.6598	1.4201		-36.0847	0.8413
101	2QWC		-16.8204	2.0932		-40.1002	0.3773
102	2QWD		-43.2413	0.4197		-52.1423	0.6576
103	2QWE		-34.7962	1.4264		-52.5281	0.5315
104	2QWF		-33.5361	0.9006		-46.9160	0.8763
105	2QWG		-36.5125	0.9299		-47.7921	0.8143
106	2QWK		-52.5465	0.5337		-62.8413	0.4288
107	2SIM		-36.7901	1.6102		-51.3180	0.5785
108	2SLI		-41.4108	0.5494		-51.1557	0.5489
109	2VVZ		-37.3142	0.7872		-37.3142	0.7872
110	2VW0		-1.3351	2.5720		-6.0699	5.6789
111	2VW1		-47.9040	0.9340		-47.9040	0.9340
112	2VW2		-3.7417	1.7883		-1.9374	0.7281
113	2W20		-3.4648	3.9859		-3.4648	3.9859
114	3B7E		-52.9288	0.5301		-52.9288	0.5301
115	3BEQ		-0.5429	2.8924		-3.1634	1.8953
116	3CKZ		-64.1348	0.5837		-74.9698	0.3973
117	3CL0		-25.9751	1.0667		-34.9374	0.5434
118	3GVL		-4.0736	1.0853		-0.3973	0.9537
119	3SLI		-49.8199	0.5732		-50.8107	0.5621
120	5NN9		-0.4553	5.0118		-0.4553	5.0118
121	6NN9		-8.2332	4.5349		-8.2332	4.5349
122	1BS4		-6.4327	8.5784		-13.8522	3.1290
123	1BSJ		-11.3794	4.1711		-11.0612	4.2598
124	1BSK		-4.5224	4.6192		-4.5224	4.6418
125	1BSZ		-17.4569	5.1791		-10.7954	4.1622
126	1G27		-17.4761	1.2388		-17.4761	1.2388
127	1ICJ		-15.7475	2.6432		-15.7475	2.6432
128	1IX1		-19.3833	1.4660		-22.9507	0.9581
129	1LM4		-8.2934	4.2912		-10.9453	4.1625

En absence des molécules d'eau

En présence des molécules d'eau

130	1LM6		-6.2701	4.5354		-17.0017	0.8103
131	1LQY		-24.7276	0.6998		-24.7276	0.6998
132	1LRU		-23.4196	0.8754		-25.7316	1.0539
133	1LRY		-25.7263	1.2462		-25.8679	1.2462
134	1N5N		-8.9602	4.1225		-2.5297	1.6788
135	1Q1Y		-20.8071	2.1163		-17.9404	1.2575
136	1RL4		-14.6438	0.9242		-14.4083	0.9913
137	1SV2		-2.5388	3.6864		-3.8783	1.4553
138	1SZZ		-24.0997	1.1267		-24.0997	1.1267
139	1VEV		-12.5170	2.0129		-12.5170	2.0129
140	1VEZ		-13.5994	1.1735		-14.4735	1.3672
141	1WS1		-21.5571	0.8431		-23.9375	0.7994
142	1Y6H		-8.9292	1.4071		-8.9677	1.4071
143	2AI7		-6.9067	1.2040		-6.9067	1.2040
144	2AI8		-8.9854	1.0746		-8.9854	1.0746
145	2EW6		-31.3725	0.9230		-31.1151	0.5345
146	2OKL		-13.5856	4.2822		-15.4774	4.0561
147	2OS3		-7.8760	3.8596		-19.5021	0.9874
148	3CMD		-15.2041	0.9729		-16.0429	1.0618
149	3E3U		-12.4358	0.6694		-14.2614	0.8377
150	3G5K		-18.6823	0.6784		-22.6909	0.9244
151	3K6L		-27.4220	1.2544		-27.4220	1.2544
152	3NRB		-0.6848	1.3545		-0.6848	1.3545
153	3O1L		-6.5554	1.6603		-2.0600	1.3008

ANNEXE2. Résultats-GOLD

Tableau 2.1 Résultats de docking par GOLD 5.0.1 : RMSD et Liais-Rotables

N°	Ligand	Fitness	S(hb_ext)	S(vdw_ext)	S(hb_int)	S(int)	intcor	RMSD	Liais Rot
1	1JLX	56,17	3,41	41,41	0	-4,17	-9,15	0,8102	15
2	1JSM	18,67	4,05	16,3	0	-7,79	-8,21	4,3162	13
3	1JSN	57,13	18,94	37,92	0	-13,96	-20,83	0,4035	24
4	1JSO	54,27	17,86	26,74	0	-0,35	-9,24	0,6128	11
5	1KJ1	30,91	11,96	15,71	0	-2,65	-4,02	1,6424	6
6	1OUW	23,15	9,93	11,38	0	-2,43	-2,39	0,5288	3
7	1QFU	31,65	9,95	18,82	0	-4,17	-4,75	3,6606	12
8	1RD8	43,31	11,74	26,45	0	-4,79	-29,28	5,0995	19
9	1RUZ	37,66	14,04	19,99	0	-3,86	-4,9	1,6327	7
10	1RV0	25,77	11,34	12,35	0	-2,55	-4,9	4,2564	7
11	1RVT	46,55	10,81	27,56	0	-2,14	-7,45	1,2771	16
12	1SBD	38,77	12,8	24,72	0	-8,01	-12,62	0,3532	12
13	1SBF	45,63	14,13	23,92	0	-1,39	-9,7	0,3755	12
14	1WGC	25,34	11,16	12,24	0	-2,65	-20,26	4,7223	23
15	2CCV	42,17	9,48	24,52	0	-1,02	-4,39	0,7867	7
16	2CGZ	39,76	9,97	21,82	0	-0,21	-7,4	0,9598	11
17	2CWG	55,72	11,47	35,93	0	-5,15	-41,25	7,714	34
18	2DV9	45,87	11,51	26,09	0	-1,52	-6,71	0,3479	12
19	2DVD	44,65	17,25	20,11	0	-0,26	-6,66	0,3966	12
20	2IBX	32,06	5,07	24,77	0	-7,05	-9,91	4,4597	13
21	2Q3N	19,13	13,84	6,28	0	-3,35	-2,12	6,166	7
22	2WR0	33,03	7,11	20,97	0	-2,91	-7,75	3,567	13
23	2X3T	41,89	11,85	27,26	0	-7,44	-9,01	0,9738	11
24	2X52	60,92	14,08	52,04	0	-24,72	-49,88	2,6885	30
25	2ZB5	30,48	2,52	20,8	0	-0,64	-4,08	8,5457	6
26	2ZB6	35,62	12,59	20,45	0	-5,08	-6,74	8,4003	13
27	2ZR1	32,39	0,44	23,93	0	-0,95	-3,16	9,8637	6
28	3AH1	50,59	6,64	31,98	0	-0,02	-4,68	0,7656	11
29	3AH4	34,57	13,89	16,07	0	-1,42	-4,17	0,6901	6
30	3BT6	22,04	4,37	16,66	0	-5,24	-10,84	0,6901	6
31	3C9Z	27,98	5,68	19,77	0	-60,85	-18,56	3,0788	18
32	3CA0	34,34	13,95	16,81	0	-2,72	-90,55	3,7188	40
33	3CA3	23,09	9,83	11,25	0	-2,22	-4,01	6,7446	7
34	3CA4	13,13	0	9,55	0	0	-1,14	1,7769	1
35	3CA5	10,02	0	7,45	0	-0,23	-1,03	4,503	1
36	3CA6	38,6	13,39	23,53	0	-7,15	-6,51	1,9386	10
37	3EYK	30,17	8,21	16,99	0	-1,4	-12,93	4,5031	6
38	3EYM	32,77	1,38	23,69	0	-1,19	-15,79	0,7875	6
39	3INB	26,66	13,23	12,55	0	-3,82	-6,48	3,981	6
40	3LLZ	55	18,5	29,76	0	-4,42	-13,95	0,478	13
41	1A4G	97,95	49,04	38,76	0	-4,39	-28,85	0,6194	10
42	1A4Q	102,72	37,51	49,77	0	-3,23	-11,49	0,8391	12
43	1A14	21,74	9,38	10,2	0	-1,66	-2,1	5,7848	6
44	1B9S	48,14	15,01	26,6	0	-3,45	-9,17	1,393	9
45	1B9T	75,88	29,3	36,01	0	-2,93	-4,98	0,8186	7
46	1B9V	82,01	27,37	40,1	0	-0,5	-11,46	0,7823	12
47	1E8V	61,71	29,42	25	0	-2,1	-7,29	0,5063	10
48	1EUS	95,02	49,75	33,99	0	-1,28	-7,56	0,4995	10
49	1EUU	37,76	16,96	15,96	0	-1,14	-4,36	2,3364	6
50	1F8B	82,3	46,75	29	0	-4,33	-7,08	1,0184	10
51	1INF	66,41	38,68	23,96	0	-5,21	-6,36	0,5532	4
52	1ING	52,82	22,07	25,02	0	-3,66	-9,63	1,1779	6
53	1INH	47,62	2,91	34,93	0	-3,32	-44,61	3,6506	12

54	1IVC	42,53	8,06	27,22	0	-2,96	-7,31	4,4897	13
55	1IVD	51,98	24,94	20,08	0	-0,56	-11,61	0,9856	5
56	1IVE	40,58	4,57	31,9	0	-7,86	-17,48	5,6399	19
57	1IVG	37,78	5,73	27,74	0	-6,09	-9,21	5,1259	13
58	1L7F	101,95	44,67	42,28	0	-0,85	-19,11	0,4618	11
59	1L7H	97,28	42,52	40,89	0	-1,47	-16,31	0,4151	11
60	1MWE	85,25	44,05	31,88	0	-2,63	-9,14	0,7952	11
61	1NCA	24,93	4,14	15,66	0	-0,75	-2,83	7,1619	6
62	1NCB	35,98	13,2	17,18	0	-0,85	-44,07	5,1598	37
63	1NCC	25,22	4,05	15,59	0	-0,26	-4,52	6,5353	6
64	1NCD	23,41	5,72	13,75	0	-1,21	-2,72	6,0041	6
65	1NMA	60,6	15,84	38,16	0	-7,72	-38,45	2,9705	37
66	1NMB	65,93	16,67	45,86	0	-13,8	-51,56	1,137	42
67	1NMC	27,08	2,73	18,42	0	-0,98	-3,11	6,2184	6
68	1NN2	29,43	7,54	16,58	0	-0,91	-3,82	6,5763	6
69	1NNB	78,59	40,6	28,57	0	-1,29	-2,58	0,7955	10
70	1NNC	101,46	51,06	37,89	0	-1,7	-11,48	0,4136	10
71	1NSB	21,32	1,44	14,67	0	-0,3	-6,14	3,7225	6
72	1NSC	93,41	54,51	29,89	0	-2,21	-11,34	0,9761	11
73	1NSD	21,85	1,96	14,81	0	-0,47	-4,16	3,6733	6
74	1SLI	80,44	44,68	27,01	0	-1,38	-1,61	1,0885	9
75	1USR	77,04	40,82	29,67	0	-4,59	-8,41	0,5476	10
76	1USX	64,6	29,69	28,15	0	-3,8	-7,44	0,7757	10
77	1V2I	18,68	5	25,08	0	-58,16	-25,63	7,3517	31
78	1V3B	35,36	1,08	26,51	0	-2,17	-9,62	4,3189	13
79	1V3C	57,24	31,71	23,36	0	-6,6	-12,06	0,8019	11
80	1V3D	71,89	36,19	28,08	0	-2,91	-6,8	0,7568	10
81	1VCJ	81,61	21,1	44,22	0	-0,29	-13,35	0,7745	12
82	1W0O	46,5	17,26	21,38	0	-0,17	-6,68	0,7307	11
83	1W0P	20,31	9,11	8,64	0	-0,69	-8,44	2,3317	3
84	1W8N	84,72	48,12	31,61	0	-6,87	-18,97	0,6016	12
85	1W8O	32,07	14,28	14,25	0	-1,81	-1,69	0,5701	5
86	1WCQ	100,7	50,79	39,28	0	-4,11	-23,04	0,4847	12
87	1XOE	69,15	30,77	33,73	0	-7,99	-14,63	0,656	11
88	1XOG	26,38	6,95	14,54	0	-0,56	-3,4	3,6145	6
89	1Z4Y	15,43	4,31	9,52	0	-1,97	-3,96	4,8622	6
90	2AEP	48,69	5,62	39,86	0	-11,74	-42,52	5,4567	31
91	2AEQ	22,17	0,92	17,25	0	-2,48	-3,52	6,1927	6
92	2BAT	82,07	45,99	28,61	0	-3,25	-8,36	0,4148	11
93	2BER	84,51	42,6	31,95	0	-2,01	-6,48	0,3053	10
94	2HT5	18,98	2,99	15,06	0	-4,73	-4	4,9149	7
95	2HT7	74,78	35,38	29,57	0	-1,26	-10,2	0,8031	10
96	2HTQ	80,01	36,76	39,87	0	-11,57	-19,07	0,8321	10
97	2HTU	101,18	44,19	41,64	0	-0,27	-12,11	0,4913	11
98	2HTV	22,31	2,43	15,06	0	-0,83	-4,42	6,1569	7
99	2QWA	28,3	7,98	21,58	0	-9,35	-12,22	6,0462	13
100	2QWB	74,2	38,63	27,5	0	-2,24	-8,96	0,864	11
101	2QWC	62,2	32,87	23,77	0	-3,36	-6,05	0,8097	10
102	2QWD	75,38	43,04	25,94	0	-3,33	-10,53	0,7571	10
103	2QWE	83,96	40,37	35,09	0	-4,66	-13,54	0,9246	10
104	2QWF	85,76	30,68	40,68	0	-0,86	-9,84	0,3219	10
105	2QWG	75,16	34,99	31,16	0	-2,67	-11,6	0,6824	10
106	2QWK	86,96	38,24	36,23	0	-1,1	-10,97	0,4727	10
107	2SIM	75,74	40,99	25,85	0	-0,79	-7,1	0,7728	10

108	2SLI	84,19	48,08	28,48	0	-3,04	-4,41	0,6888	8
109	2VVZ	75,66	33,04	33	0	-2,76	-6,99	0,7475	10
110	2VW0	15,15	2,14	9,68	0	-0,29	-3,22	4,3908	5
111	2VW1	75,71	40,35	31,6	0	-8,1	-8,57	0,6237	10
112	2VW2	25,51	6,37	14,01	0	-0,12	-3,82	1,7948	5
113	2W20	36,07	2,88	25,9	0	-2,42	-8,79	1,2099	3
114	3B7E	101,19	46,57	40,95	0	-1,69	-6,96	0,7205	10
115	3BEQ	19,49	6	10,34	0	-0,72	-3,22	3,4192	5
116	3CKZ	100,72	46,29	39,93	0	-0,48	-7,5	0,2337	10
117	3CLO	85,09	38,27	34,97	0	-1,27	-13,58	0,6644	10
118	3GVL	43	24,51	14,22	0	-1,06	-11,52	4,7692	11
119	3SLI	84,53	46,46	30,22	0	-3,48	-3,18	0,315	8
120	5NN9	27,1	0,94	19,76	0	-1,01	-6,96	0,315	8
121	6NN9	24,8	3,9	15,98	0	-1,07	-4,57	7,4263	6
122	1BS4	51,81	18,94	37,07	0	-18,1	-42,35	3,3213	27
123	1BSJ	76,17	10,93	49,05	0	-2,2	-21,34	1,034	14
124	1BSK	89,13	25,75	50,33	0	-5,83	-18,85	0,7842	14
125	1BSZ	61,61	11,79	48,22	0	-16,48	-32,48	3,796	27
126	1G27	65,76	7	42,92	0	-0,26	-14,91	1,0801	18
127	1ICJ	60,77	8,72	41,67	0	-5,24	-36,05	2,3334	27
128	1IX1	74,36	15,43	47,63	0	-6,56	-9,02	1,0822	16
129	1LM4	31,68	13,83	13,75	0	-1,06	-0,79	4,7606	3
130	1LM6	32,71	16,74	12,49	0	-1,2	-2,51	5,3588	4
131	1LQY	75	12,99	45,27	0	-0,25	-15,69	0,4071	16
132	1LRU	81,68	19,39	45,9	0	-0,82	-17,02	1,1353	17
133	1LRY	75,46	22,33	41,7	0	-4,22	-20,23	0,8724	17
134	1N5N	27,66	6,6	15,59	0	-0,37	-3,24	1,8067	5
135	1Q1Y	67,47	14,25	43,18	0	-6,15	-11,16	1,126	16
136	1RL4	43,03	6,22	28,38	0	-2,22	-5,19	0,8447	9
137	1SV2	39,85	5,8	27,11	0	-3,23	-27,73	2,4893	6
138	1SZZ	65,2	17,5	39,59	0	-6,74	-9,55	1,3149	16
139	1VEV	63,8	18,63	33,36	0	-0,71	-2,68	0,8705	3
140	1VEZ	26,39	13,6	11	0	-2,34	-5,45	1,7894	6
141	1WS1	67,24	4,63	47,44	0	-2,63	-7,76	0,6604	15
142	1Y6H	15,73	7,99	6,37	0	-1,03	-0,74	0,6992	1
143	2AI7	40,24	2,68	29,63	0	-3,18	-0,88	1,4255	6
144	2AI8	51,34	2,96	38,19	0	-4,14	-1,85	0,7444	6
145	2EW6	59,49	5,02	41,58	0	-2,7	-7,11	0,3834	8
146	2OKL	55,24	8,3	36,5	0	-3,25	-9,47	1,3486	16
147	2OS3	63,49	3,95	48,2	0	-6,74	-10,7	1,2898	16
148	3CMD	36,63	21,17	11,98	0	-1,01	-0,73	1,0974	2
149	3E3U	61,72	8,57	45,07	0	-8,83	-7,36	1,1537	11
150	3G5K	67,06	9,69	47,04	0	-7,31	-8,74	0,7822	17
151	3K6L	63,76	7,11	41,45	0	-0,34	-18,35	0,8602	18
152	3NRB	16,04	1,93	10,26	0	0	0	1,3692	0
153	3O1L	28,66	10,37	14,74	0	-1,98	-3,41	1,6241	5

Tableau 2.2 Résultats de *docking* par GOLD 5.0.1 : Temps et Faux Positifs

N°	Ligand	Temps en (secondes)	% Faux Positifs
1	1JLX	3325,0003	0
2	1JSM	2129,0004	0
3	1JSN	5547,0017	0

4	1JSO	1668,001	50
5	1KJ1	812,0018	50
6	1OUW	219,0011	10
7	1QFU	1811,0017	0
8	1RD8	4780,0009	0
9	1RUZ	1391,0014	70
10	1RV0	1204,0008	0
11	1RVT	3387,0005	10
12	1SBD	2343,0005	60
13	1SBF	2087,0009	10
14	1WGC	3693,0005	0
15	2CCV	861,0006	30
16	2CGZ	1499,0007	40
17	2CWG	7850,0009	0
18	2DV9	2182,0011	40
19	2DVD	2217,0013	60
20	2IBX	2550,0015	0
21	2Q3N	810,0015	0
22	2WR0	2452,0009	0
23	2X3T	1885,0008	70
24	2X52	6582,0018	0
25	2ZB5	816,0011	0
26	2ZB6	2274,0013	0
27	2ZR1	971,0015	0
28	3AH1	1744,001	30
29	3AH4	696,0014	70
30	3BT6	2291,0004	0
31	3C9Z	3476,0014	0
32	3CA0	9498,001	0
33	3CA3	919,0011	0
34	3CA4	163,0005	30
35	3CA5	150,0012	0
36	3CA6	1510,0008	70
37	3EYK	872,0009	0
38	3EYM	1244,0014	70
39	3INB	873,0013	0
40	3LLZ	2493,0017	30
41	1A4G	2349,0005	30
42	1A4Q	2784,0014	20
43	1A14	755,0002	0
44	1B9S	1697,0011	50
45	1B9T	1335,0011	50
46	1B9V	2535,0013	30
47	1E8V	2139,0005	30
48	1EUS	1789,0003	20
49	1EUU	806,0012	0
50	1F8B	1757,0009	20
51	1INF	802,001	20
52	1ING	1127,0017	30
53	1INH	2593,0005	0
54	1IVC	2839,0011	0
55	1IVD	867,0011	30
56	1IVE	5123,0006	0
57	1IVG	2738,0016	0

58	1L7F	2553,0006	40
59	1L7H	2523,0013	20
60	1MWE	2216,0011	10
61	1NCA	753,0009	0
62	1NCB	7249,0008	0
63	1NCC	764,0013	0
64	1NCD	763,0014	0
65	1NMA	9368,0014	0
66	1NMB	12010,0005	20
67	1NMC	709,0017	0
68	1NN2	850,001	0
69	1NNB	1867,001	10
70	1NNC	2107,0016	20
71	1NSB	742,0008	0
72	1NSC	2424,0008	10
73	1NSD	804,0012	0
74	1SLI	2043,001	30
75	1USR	2134,0013	20
76	1USX	2367,0003	30
77	1V2I	8495,0013	0
78	1V3B	2874,0011	0
79	1V3C	2304,0013	50
80	1V3D	2121,0007	20
81	1VCJ	2585,0012	10
82	1W0O	1962,0007	60
83	1W0P	294,0011	0
84	1W8N	2423,0012	30
85	1W8O	599,0014	40
86	1WCQ	2089,0009	10
87	1XOE	2033,0019	10
88	1XOG	627,0012	0
89	1Z4Y	658,001	0
90	2AEP	8465,001	0
91	2AEQ	753,0015	0
92	2BAT	2061,0012	10
93	2BER	1932,0013	20
94	2HT5	817,0016	0
95	2HT7	1777,001	0
96	2HTQ	2388,0007	50
97	2HTU	2523,0013	40
98	2HTV	801,0011	0
99	2QWA	2103,0015	0
100	2QWB	2060,0016	40
101	2QWC	1929,0005	30
102	2QWD	2057,0013	40
103	2QWE	2393,0005	20
104	2QWF	2047,0001	10
105	2QWG	1806,0004	30
106	2QWK	2010,0006	10
107	2SIM	1975,0006	10
108	2SLI	1709,0004	30
109	2VVZ	2341,0005	10
110	2VW0	391,0009	0
111	2VW1	2174,0007	20

112	2VW2	590,0009	90
113	2W20	513,0017	30
114	3B7E	2228,0003	20
115	3BEQ	367,0013	0
116	3CKZ	2258,0016	40
117	3CLO	1978,0014	10
118	3GVL	1594,0009	0
119	3SLI	1880,0007	30
120	5NN9	666,0005	0
121	6NN9	655,0014	0
122	1BS4	5204,0011	0
123	1BSJ	2910,0011	30
124	1BSK	2770,0012	0
125	1BSZ	4893,0016	0
126	1G27	3441,0009	10
127	1ICJ	4858,0005	0
128	1IX1	3369,0012	40
129	1LM4	275,001	0
130	1LM6	323,0003	0
131	1LQY	3509,001	50
132	1LRU	3966,0013	20
133	1LRY	3519,0015	20
134	1N5N	535,0008	70
135	1Q1Y	3503,001	20
136	1RL4	1584,0005	70
137	1SV2	768,0015	0
138	1SZZ	3614,0001	40
139	1VEV	492,0008	40
140	1VEZ	721,0012	40
141	1WS1	2910,0011	0
142	1Y6H	155,0011	10
143	2AI7	824,0006	30
144	2AI8	988,0009	50
145	2EW6	1473,0009	50
146	2OKL	3416,0016	30
147	2OS3	3228,0003	10
148	3CMD	176,0009	10
149	3E3U	2132,0009	30
150	3G5K	3954,0003	0
151	3K6L	3920,0003	20
152	3NRB	159,0012	20
153	3O1L	566,0012	40

ANNEXE3.Résultats FlexX-GOLD

Tableau 3.1 Résultats du *docking* par FlexX et GOLD : précision du classement

N°	Ligand	GoldScore,Fitness	RMSD(GOLD)	Rank(GOLD)	Score(FlexX)	RMSD(FlexX)	Rank(FlexX)
1	1JLX	56,157	0,8102	10	-12,3154	0,8554	1
2	1JSM	28,9504	4,708	3	-5,2534	5,1443	1
3	1JSN	63,4987	0,6236	9	-13,0303	0,6093	2
4	1JSO	54,2747	0,6128	1	-9,368	1,6865	24
5	1KJ1	33,0992	1,7881	7	-11,2729	1,3324	3
6	1OUW	26,1579	1,4945	8	-9,4498	0,3018	1
7	1QFU	31,651	3,6606	9	-8,9757	4,9887	3
8	1RD8	45,5173	5,1166	10	-13,1664	5,7019	18
9	1RUZ	45,9889	5,1453	2	-10,1934	3,2315	13
10	1RV0	35,5806	7,0419	7	-7,7983	5,0088	34
11	1RVT	49,9586	1,282	5	-5,3603	1,429	29
12	1SBD	51,8331	2,0877	1	-9,4049	1,4575	2
13	1SBF	47,1118	2,1465	1	-14,2197	0,9996	1
14	1WGC	28,3201	6,9009	8	-6,7165	5,3257	79
15	2CCV	42,1741	0,7867	1	-9,3012	0,695	2
16	2CGZ	49,5147	7,4751	2	-5,5776	1,1072	10
17	2CWG	62,7146	8,7395	1	-9,5978	6,0951	246
18	2DV9	47,9111	2,3132	2	-13,2727	1,0278	2
19	2DVD	44,6457	0,3966	2	-10,6263	0,5268	12
20	2IBX	38,4651	6,8692	4	-6,7231	4,0003	19
21	2Q3N	27,6549	9,0892	1	-2,6919	3,9679	44
22	2WR0	37,9065	7,0969	1	-1,9839	5,1806	116
23	2X3T	50,8773	4,7568	1	-12,0921	0,9148	3
24	2X52	63,0256	2,993	6	-14,2005	2,2388	4
25	2ZB5	33,5231	11,7606	5	-0,2879	4,9487	71
26	2ZB6	38,5365	11,5146	1	-3,9929	5,1574	67
27	2ZR1	37,5058	12,4655	2	-9,3826	5,0969	1
28	3AH1	50,5925	0,7656	1	-8,2706	0,8852	45
29	3AH4	34,5671	0,6901	1	-9,8797	1,4745	7
30	3BT6	42,2707	6,2396	10	-3,8066	6,201	10
31	3C9Z	32,8274	3,215	2	-5,6326	4,5821	9
32	3CA0	53,7743	6,3315	6	-10,9436	4,606	10
33	3CA3	30,7076	7,8646	2	-5,9925	2,2485	12
34	3CA4	16,4464	4,1699	2	-3,7112	2,8811	26
35	3CA5	12,6792	5,9381	1	-4,0687	4,5836	33
36	3CA6	52,0166	2,8898	1	-15,26	0,7034	5
37	3EYK	37,8188	11,8278	6	-0,5465	1,9558	9
38	3EYM	38,1503	4,7455	2	-3,5285	2,5009	98
39	3INB	33,7088	10,3881	1	-8,0836	5,2311	3
40	3LLZ	54,9974	0,478	1	-11,3996	1,0633	2
41	1A4G	97,9454	0,6194	1	-57,8843	0,5634	1
42	1A4Q	102,7175	0,8391	1	-42,086	1,532	3
43	1A14	28,3782	6,9571	3	-2,8071	5,8926	122
44	1B9S	56,7051	1,5641	1	-28,3833	1,2694	10
45	1B9T	75,8819	0,8186	1	-42,9876	0,862	2
46	1B9V	82,0089	0,7823	1	-19,3453	5,4444	123
47	1E8V	70,7869	1,4235	1	-23,9259	2,1764	81
48	1EUS	95,0241	0,4995	1	-46,3597	0,8616	1
49	1EUU	41,0645	3,6733	1	-3,4951	2,3456	93
50	1F8B	82,2966	1,0184	1	-34,4821	1,1113	11
51	1INF	66,4112	0,5532	1	-35,0879	1,3606	6

52	1ING	54,2	2,699	1	-25,3921	0,8167	5
53	1INH	53,4556	5,3724	3	-3,4216	4,3984	28
54	1IVC	44,6074	11,2871	4	-9,2081	5,8447	4
55	1IVD	55,4356	5,0955	1	-25,3829	1,5127	22
56	1IVE	58,1303	6,0247	3	-0,989	6,2214	156
57	1IVG	44,2911	7,5142	7	-3,6649	6,1945	23
58	1L7F	101,9539	0,4618	1	-54,9032	0,9724	2
59	1L7H	97,2751	0,4151	1	-45,8292	0,942	1
60	1MWE	85,247	0,7952	1	-42,4994	0,9307	2
61	1NCA	25,9872	7,7741	5	-0,7086	6,0002	166
62	1NCB	49,0141	8,2111	1	-7,6198	5,3941	1
63	1NCC	25,8445	7,7138	8	-3,5783	7,3520	136
64	1NCD	31,5159	7,9633	1	-3,8148	5,8122	158
65	1NMA	72,832	3,2467	5	-16,3112	4,0455	1
66	1NMB	84,8616	7,5975	7	-3,0973	4,7397	1
67	1NMC	33,4706	8,0369	1	-4,0570	5,4498	10
68	1NN2	37,04	8,4895	4	-7,3278	2,3649	16
69	1NNB	78,5894	0,7955	1	-26,0268	1,3522	23
70	1NNC	101,4572	0,4136	1	-24,8110	4,3722	13
71	1NSB	28,5512	10,329	10	-1,2461	5,5178	20
72	1NSC	93,407	0,9761	1	-46,8339	1,5878	1
73	1NSD	27,6614	5,8924	3	-5,5389	5,3877	36
74	1SLI	80,4417	1,0885	2	-47,2568	0,8973	4
75	1USR	77,038	0,5476	1	-48,0466	0,8429	1
76	1USX	64,6015	0,7757	1	-30,5448	1,9344	12
77	1V2I	46,6538	10,3274	2	-10,1537	4,9487	1
78	1V3B	53,695	6,1348	1	-19,2016	4,4428	1
79	1V3C	58,6688	5,0398	3	-31,0691	1,5740	9
80	1V3D	71,8928	0,7568	1	-43,2652	1,5116	1
81	1VCJ	60,96	0,7745	1	-23,3109	1,2548	1
82	1W0O	46,5012	0,7307	2	-12,1013	0,7562	31
83	1W0P	22,056	5,2958	2	-7,1565	3,6875	29
84	1W8N	84,718	0,6016	1	-53,1499	0,3190	2
85	1W8O	32,0729	0,5701	1	-8,1139	1,8741	10
86	1WCQ	100,696	0,4847	1	-57,1538	0,4435	2
87	1XOE	72,6529	1,2026	2	-28,0087	2,3070	4
88	1XOG	26,3767	3,6145	8	-3,0239	3,6442	92
89	1Z4Y	21,5794	6,1989	5	-6,9686	5,0079	16
90	2AEP	60,001	7,4736	3	-6,7494	5,7706	38
91	2AEQ	25,8108	8,5392	1	-0,9944	3,0760	149
92	2BAT	82,0733	0,4148	1	-45,6993	0,8012	2
93	2BER	84,5139	0,3053	1	-59,1473	0,6006	1
94	2HT5	21,5172	9,457	2	-2,0891	4,5228	51
95	2HT7	74,7773	0,8031	1	-37,9181	0,7611	1
96	2HTQ	80,0056	0,8321	1	-50,5404	0,9084	3
97	2HTU	101,1805	0,4913	1	-43,6394	1,7663	2
98	2HTV	24,4881	7,6001	3	-1,6274	4,8837	56
99	2QWA	36,3781	36,3781	1	-1,9894	5,2353	150
100	2QWB	74,1977	0,864	1	-33,9022	1,5528	1
101	2QWC	76,5265	1,2242	1	-17,6531	2,5980	76
102	2QWD	75,3818	0,7571	1	-46,2950	0,7937	1
103	2QWE	83,9612	0,9246	1	-40,1386	1,6294	1

104	2QWF	85,7612	0,3219	1	-39,0974	1,3695	1
105	2QWG	75,1564	0,6824	1	-41,3956	1,0893	35
106	2QWK	86,9606	0,4727	1	-52,5465	0,5337	2
107	2SIM	85,9664	1,1748	1	-36,7901	1,6102	1
108	2SLI	89,9549	1,1039	1	-42,7858	0,8867	11
109	2VVZ	81,7622	1,099	1	-38,3722	1,0308	10
110	2VW0	17,9368	7,5203	3	-2,0711	3,8878	31
111	2VW1	77,8195	1,8669	3	-60,2196	1,0664	1
112	2VW2	29,6529	4,1682	6	-3,7417	1,7883	287
113	2W20	39,7413	2,3141	2	-7,1421	4,8117	3
114	3B7E	101,1926	0,7205	1	-69,7189	0,6333	1
115	3BEQ	27,7628	5,246	1	-3,5731	3,4935	29
116	3CKZ	100,7179	0,2337	1	-64,1348	0,5837	1
117	3CL0	85,0883	0,6644	1	-45,7026	1,2174	6
118	3GVL	47,9858	6,3259	2	-4,0736	1,0853	315
119	3SLI	84,5308	0,315	1	-49,8199	0,5732	1
120	5NN9	30,4478	7,2056	1	-2,1077	5,5218	50
121	6NN9	25,5812	8,0399	1	-8,5589	5,7817	174
122	1BS4	78,5267	6,6015	3	-8,7705	8,7651	3
123	1BSJ	76,4509	1,3599	1	-18,6494	4,2974	1
124	1BSK	98,3951	1,7017	4	-9,0256	5,1640	207
125	1BSZ	76,0396	6,5889	9	-18,5954	5,2825	16
126	1G27	65,7572	1,0801	1	-23,3741	1,3248	10
127	1ICJ	74,4196	3,8506	1	-16,7765	2,6709	4
128	1IX1	74,8123	4,4565	1	-20,4949	1,5265	14
129	1LM4	29,5003	5,2871	10	-10,7679	4,6020	5
130	1LM6	34,8463	6,6345	3	-8,7622	4,7431	22
131	1LQY	74,9955	0,4071	1	-26,7658	0,7959	1
132	1LRU	81,6798	1,1353	1	-25,5030	1,0646	14
133	1LRY	78,6507	1,5284	1	-26,4078	1,2492	28
134	1N5N	33,0976	4,5398	1	-10,9261	4,2165	2
135	1Q1Y	71,7408	2,8313	4	-21,2402	2,4882	77
136	1RL4	47,6694	3,763	1	-15,9605	1,2110	189
137	1SV2	46,3487	2,8374	1	-11,3643	3,8174	3
138	1SZZ	77,7657	1,6585	2	-26,8877	1,2953	27
139	1VEV	63,7966	0,8705	1	-14,6955	2,1260	8
140	1VEZ	32,0238	2,3401	1	-14,4529	1,3269	53
141	1WS1	68,7464	0,7245	1	-26,1947	0,9734	6
142	1Y6H	19,0453	1,5107	1	-9,0337	1,4073	1
143	2AI7	44,1933	1,472	3	-10,5645	1,4149	2
144	2AI8	53,6939	1,3436	2	-11,3673	1,2312	4
145	2EW6	59,4936	0,3834	1	-32,0040	1,2247	1
146	2OKL	62,9143	3,5897	7	-14,6364	4,6339	81
147	2OS3	63,4903	1,2898	1	-10,1535	4,1162	227
148	3CMD	42,3835	1,6464	1	-15,2041	0,9729	5
149	3E3U	65,628	1,8393	7	-18,3239	0,9477	62
150	3G5K	67,0624	0,7822	6	-20,2617	0,7913	5
151	3K6L	63,7633	0,8602	3	-28,8176	1,3597	11
152	3NRB	17,384	2,7418	1	-1,6405	1,5812	9
153	3O1L	30,5713	1,7511	2	-8,4463	1,6708	6

عنوان: دراسة مقارنة لفعالية برنامجين لـ *docking* وتطبيق على تثبيط الـ Neuraminidase

ملخص:

حاليا هناك اهتمام كبير بمقارنة برامج *docking* بروتين-رابط. مجلة لمقارنات أخيرة يُبين أنه من الصعب استخلاص استنتاجات للتطبيق العام. القياسات الرقمية مثل الـ RMSD يجب أن تُفسر بحذر و يمكن استكمالها بشكل مفيد بقياسات مبنية على أساس التفاعلات و التفتيش البصري للـ *docking* مجموعات الاختبار يجب أن تكون متنوعة وذات موثوقية تجريبية جيدة. لم يكن في نيتنا في الدراسة الحالية اقتراح تسلسل هرمي لبرنامجي الـ *docking* المتوفرين GOLD و FlexX ولكن استخلاص الايجابيات و السلبيات للأدوات المختارة وذلك في إطارات مختلفة. الأهداف العلاجية المستخدمة في هذا العمل المتمثلة في الـ ايمافلوتينين, النورامينيداز و البيبتيد ديفورميلاز أشارت إلى أن GOLD و FlexX قادرين على إنتاج أنماط الروابط التجريبية بنسبة 64.05 % بالنسبة للأول و 55.56 % فقط بالنسبة للثاني. و العكس FlexX قادر على تحديد و ترتيب 54.25 % من أنماط الروابط الملاحظة في حين GOLD لا يستطيع ترتيب الـ 49.02 % من هته الأخيرة . وقد أظهرت دراسات إضافية أن FlexX هو الأسرع وأكثر حساسية من GOLD فيما يخص عدد الروابط التي تدور, وكذلك الأقل في ارتكاب الأخطاء من خلال ترتيب الحلول و المسماة الايجابيات الخاطئة مع نسبة 10.92 % فقط أما GOLD فشلت بنسبة 19.02 % من الحالات هذا ما شجّعنا على مواصلة عملنا مع FlexX و إدخال الماء في عملية الـ *docking* لمعرفة دور هذا الأخير وملاحظة التحسن الواضح في دقة التنبؤات. أخيرا استعمال البرنامج FlexX سمح لنا باقتراح المشابه سي اي دي 5278155 للزناناميفير كمثبط جديد محتمل للنورامينيداز ن1, الاثنان يملكان طاقات تفاعل متشابهة لكن خصائص دوائية أكثر أهمية بالنسبة للمشابه.

الكلمات المفتاحية: برامج الـ *docking* بروتين-رابط, مقارنات, DMSR, GOLD, FlexX, دقة, *docking*, الروابط التي تدور, الايجابيات الخاطئة, نورامينيداز ن1, زاناميفير, مشابه

Title: Comparative study of the efficiency of two docking programs and application to the inhibition of neuraminidase.

Abstract:

Currently there is great interest in comparing the programs of protein-ligand docking. A review of recent comparisons shows that it is difficult to draw conclusions for a general application. The numerical measures such as RMSD should be interpreted with caution and can be usefully complemented by measures based on the interaction and visual inspection of the docking. The test sets must be of appropriate diversity and good experimental reliability. It was not our intention in the current study to propose a hierarchy of the two available docking programs GOLD and FlexX but to relieve the advantages and disadvantages of selected tools in different contexts. Therapeutic targets used in this work namely the hemagglutinin, neuraminidase and peptide deformylase indicated that GOLD and FlexX are able to reproduce the experimental binding modes with a rate of 64.05% for the first and only 55.56 % for the second. However FlexX is able to identify and classify 54.25% of the observed binding modes while GOLD can only rank 49.02% of these latter's. Further studies showing that FlexX is faster and more sensitive than GOLD on the number of rotatable links, and also, at least to make mistakes by classifying the solutions labeled false positives with only a percentage of 10.92% while GOLD fails in 19.02% of cases encouraged us to continue our work with FlexX and introducing water into the docking process to see the role of the latter and notice the significant improvement in the accuracy of predictions. Finally the use of the software FlexX allowed us to propose the similar CID5278155 of zanamivir as a new potent inhibitor of neuraminidase N1, both having comparable interaction energies but pharmacokinetic properties more interesting for the similar.

Key words: programs for protein-ligand docking, comparisons, RMSD, GOLD, FlexX, precision, docking, rotatable links, false positives, neuraminidase N1, zanamivir, similar.

