



HAL
open science

Conception d'un logiciel de docking et applications dans la recherche de nouvelles molécules actives

Aurélien Grosdidier

► **To cite this version:**

Aurélien Grosdidier. Conception d'un logiciel de docking et applications dans la recherche de nouvelles molécules actives. Sciences pharmaceutiques. 2007. dumas-01115990

HAL Id: dumas-01115990

<https://dumas.ccsd.cnrs.fr/dumas-01115990>

Submitted on 12 Feb 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il n'a pas été réévalué depuis la date de soutenance.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact au SICD1 de Grenoble : **thesebum@ujf-grenoble.fr**

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

UNIVERSITÉ JOSEPH FOURIER
FACULTÉ DE PHARMACIE DE GRENOBLE

Année : 2007

N°

Conception d'un logiciel de docking et applications dans la recherche de nouvelles molécules actives.

THÈSE

PRÉSENTÉE POUR L'OBTENTION DU TITRE DE DOCTEUR EN PHARMACIE

DIPLÔME D'ÉTAT

Aurélien Grosdidier

Né le 27 novembre 1977 à Die

THÈSE SOUTENUE PUBLIQUEMENT À LA FACULTÉ DE PHARMACIE DE GRENOBLE

Le 27 juin 2007

DEVANT LE JURY COMPOSÉ DE

Président du jury : Pr. Emmanuel Drouet

Membres:

Pr. Olivier Michielin, Directeur de thèse

Pr. Wim Burmeister

Dr. V. Zoete

La faculté de Pharmacie de Grenoble n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions sont considérées comme propres à leurs auteurs.

*À Celui qui,
alors que nous ne pouvons qu'espérer les prolonger,
offre un sens à nos vies.*

1 UFR de Pharmacie

1.1 Professeurs

BARKI	Aziz	Pharmacie Galénique et industrielle, Formulation et Procédés Pharmaceutiques (G.R.N.R., I.N.S.E.R.M. E03 40)
BURMEISTER	Wilhelm	Physique (I.V.M.S.)
CALOP	Jean	Pharmacie Clinique (C.H.U.)
DANEL	Vincent	SAMU-SMUR et Toxicologie (C.H.U.)
DECOUT	Jean-Luc	Chimie Bio-Inorganique (D. P. M.)
DEMENGE	Pierre	Physiologie / Pharmacologie
DROUET	Emmanuel	Immunologie / Microbiologie / Biotechnologie (I.V.M.S.)
FAVIER	Alain	Biochimie (L.A.N.)
GODIN-RIBUOT	Diane	Physiologie / Pharmacologie (H.P.2)
GOULON	Chantal	Physique (E.S.R.F.)
GRILLOT	Renée	Parasitologie / Mycologie Médicale (C.H.U.)
LECLERC	Gérard	Chimie Organique
MARIOTTE	Anne-Marie	Pharmacognosie (D.P.M.)
PEYRIN	Eric	Chimie Analytique (D.P.M.)
RIBUOT	Christophe	Physiologie / Pharmacologie (H.P.2)

ROUSSEL	Anne-Marie	Biochimie (N.V.M.C.)
SEIGLE MURANDI	Françoise	Émérite
STEIMAN	Régine	Biologie Cellulaire (T.I.M.C.)
WOUESSIDJEWÉ	Denis	Pharmacotechnie et Vectorisation (D.P.M.)

1.2 Professeurs associés

CHAMPON	Bernard	Pharmacie Clinique
RIEU	Isabelle	Qualitologie

1.3 Maître de conférences

ALDEBERT	Delphine	Parasitologie / Mycologie (L.A.P.M./U.M.R. C.N.R.S. 5163)
ALLENET	Benoît	Pharmacie Clinique (C.H.U. / ThEMAS T.I.M.C.-I.M.A.G.)
BARTOLI	Marie-Hélène	Pharmacologie et Physiologie
BATANDIER	Cécile	Nutrition et Physiologie (I.N.S.E.R.M. E-0221) Bioénergétique Fondamentale et Appliquée)
BOUMENDJEL	Ahcène	Pharmacognosie (D.P.M.)
BRETON	Jean	Biologie Moléculaire / Biochimie (L.A.N.)
BUDAYOVA SPANNO	Monika	Biophysique Structurale (I.V.M.S.)
CHOISNARD	Luc	Pharmacotechnie et Vectorisation (D.P.M.)

COLLE	Pierre Emmanuel	Anglais
DELETRAZ- DELPORTE	Martine	Droit Pharmaceutique Économie Santé (Lyon)
DEMEILLIERS	Christine	Biochimie (N.V.M.C.)
DESIRE	Jérôme	Chimie Bio-Inorganique (D.P.M.)
DURMORT- MEUNIER	Claire	Virologie (I.B.S.)
ESNAULT	Danielle	Chimie Analytique (D.P.M.)
FAURE	Patrice	Biochimie (H.P.2)
GEZE	Annabelle	Pharmacotechnie et Vectorisation (D.P.M.)
GERMI	Raphaële	Microbiologie (I.V.M.S. / C.H.U.)
GILLY	Catherine	Chimie Thérapeutique (D.P.M.)
GROSSET	Catherine	Chimie Analytique (D.P.M.)
GUIRAUD	Pascale	Biologie Cellulaire et Génétique (T.I.M.C.)
HININGER- FAVIER	Isabelle	Biochimie (N.V.M.C.)
JOYEUX-FAURE	Marie	Physiologie / Pharmacologie (H.P.2)
KRIVOBOK	Serge	Botanique / Cryptogamie (L.E.C.A.)
MORAND	Jean-Marc	Chimie Thérapeutique (D.P.M.)
NICOLLE	Edwige	Chimie Organique (D.P.M.)

PINEL	Claudine	Parasitologie / Mycologie Médicale (C.H.U. / C.I.B.)
RACHIDI	Walid	Biochimie (L.A.N.)
RAVEL	Anne	Chimie Analytique (D.P.M.)
RAVELET	Corinne	Chimie Analytique (D.P.M.)
SEVE	Michel	Biotechnologie (C.H.U. / C.R.I. I.A.B.)
TARBOURIECH	Nicolas	Biophysique (I.V.M.S.)
VANHAVERBEKE	Cécile	Chimie Bio-Inorganique (D.P.M.)
VILLEMAIN	Danielle	Biostatistiques (Radio pharmaceutiques Bio cliniques, I.N.S.E.R.M. E03 40)
VILLET	Annick	Chimie Analytique (D.P.M.)

1.4 Enseignants Anglais

FITE	Andrée
GOUBIER	Laurence

1.5 ATER

½	TRAVIER Laetitia	Immunologie
½	SACCONE Patrick	Mycologie
1	MICHALET Serge	Pharmacognosie

- 1 KHALEF Nawel Pharmacie Galénique et Industrielle, Formulation et Procédés
Pharmaceutiques (G.R.N.R., I.N.S.E.R.M. E03 40)
- 1 BEGUIN Pauline Physiologie

1.6 Professeurs Agrégé (PRAG)

ROUTABOUL Christel Chimie Générale (D.P.M.)

2 Remerciements

Tout d'abord, je tiens à remercier M. le Prof. Emmanuel Drouet, sans lequel ce travail n'aurait pas été possible. Sans y avoir contribué directement, son appui tout au long de mon cursus à Grenoble ainsi que lors de mon départ vers la Suisse a été d'une aide immense. Sans lui, la filière recherche de laquelle je suis issu n'aurait probablement pas existé, ou pas de cette façon, et en tant qu'ancien étudiant, je souhaite lui rendre hommage.

Je remercie également chaleureusement M. le Prof. Olivier Michielin, pour m'avoir donné la possibilité de rejoindre son groupe de recherche, pour son soutien et sa confiance quelles qu'aient été les difficultés scientifiques ou techniques. Un énorme merci à mon co-superviseur et ami Vincent Zoete, pour sa très importante contribution à ce travail de thèse. C'est un plaisir de travailler avec une personne aussi talentueuse, droite et infatigable. Je tiens également à remercier les membres du groupe de recherche dans lequel ce travail s'est déroulé, par ordre chronologique : Theres Fagerberg, Pierre Chodanowski, Antoine Leimgruber, Michel Cuendet, Hamid Hussain-Kahn, Simon Bernèche, Ute Röhrig, Justyna Iwaszkiewicz, Mathias Ferber et Thierry Schüpbach. Merci à tous pour toutes les discussions que nous avons eues, votre soutien, ces bons moments passés ensemble, et les préparations sans fin de nos manuscrits.

Travailler avec vous tous a été un grand plaisir.

Ce travail n'aurait pas été possible sans l'équipe VITAL-IT de l'institut Suisse de Bioinformatique et le projet Cluster versus Cancer, qui ont fourni les ressources informatiques dont nous avons besoin. EADock est un bel outil, et l'utiliser demande de la puissance de calcul. Merci à vous tous !

Pendant ces années, les communautés internet ont également été très utiles, et je remercie la communauté DLFP pour sa réactivité extraordinaire, qu'il s'agisse d'une question sérieuse, d'un troll, ou d'un FUD.

3 Table des matières

1 UFR de Pharmacie.....	4
1.1 Professeurs.....	4
1.2 Professeurs associés.....	5
1.3 Maître de conférences.....	5
1.4 Enseignants Anglais.....	7
1.5 ATER.....	7
1.6 Professeurs Agrégé (PRAG).....	8
2 Remerciements.....	9
3 Table des matières.....	10
4 Introduction.....	12
4.1 Vers de nouveaux médicaments.....	13
4.2 Le docking.....	15
4.2.1 Énoncé.....	15
4.2.2 Problème théorique et réponses méthodologiques.....	18
4.2.3 Modélisation des interactions moléculaires.....	24
4.3 État de l'art du docking moléculaire.....	27
4.3.1 Implémentations courantes.....	27
4.3.2 Performance.....	27
4.4 Deux applications des logiciels de docking dans la conception de nouveaux médicaments.....	28
4.4.1 Le criblage virtuel.....	29
4.4.2 Conception de médicament basée sur le docking de fragments.....	31
4.5 Conclusion générale.....	36
5 EADock.....	37
5.1 Algorithme de docking.....	38
5.1.1 Génération de la population initiale : « seeding ».....	40
5.1.2 Sélection.....	40
5.1.3 Génération de diversité.....	43
5.1.4 Traitement final.....	45

5.1.5 Visualisation des résultats.....	46
5.2 Complexes utilisés pour la validation d'EADock.....	47
5.3 Validation de l'algorithme.....	50
5.4 Discussion.....	53
5.4.1 Évaluation des algorithmes de docking.....	53
5.4.2 Notre approche.....	54
6 Applications.....	58
6.1 Aperçu.....	58
6.2 Matériel et méthode.....	58
6.2.1 Aperçu.....	58
6.2.2 EADock.....	61
6.2.3 Champ de force.....	61
6.3 L'intégrine.....	61
6.3.1 Contexte biologique.....	62
6.3.2 Validation d'EADock.....	63
6.3.3 Conception d'inhibiteurs de l'intégrine $\alpha 5\beta 1$	64
6.3.4 Conclusion.....	67
6.4 Conception de ligands peptidiques de PPAR α	68
6.4.1 Contexte biologique.....	68
6.4.2 Validation d'EADock.....	68
6.4.3 Conception de ligands peptidiques du hPPAR α	69
6.4.4 Conclusion.....	72
7 Conclusion.....	73
8 Bibliographie.....	76
9 Serment des apothicaires.....	90

4 Introduction

Alors que les questions relatives à l'origine et au sens de la vie sont très personnelles et que nos sociétés ne proposent pas de réponses claires, ces dernières ont toujours cherché à assumer les conséquences pratiques de la vie, de la compétition inévitable avec les autres formes de vie jusqu'à la lutte contre les maladies et la mort.

Des batailles historiques ont été remportées grâce à la science, par exemple suite à la découverte et à l'utilisation des antibiotiques. Devant de tels succès, beaucoup pensent désormais que le pire (voire le seul) ennemi de l'être humain n'est autre que l'être humain lui-même. Il est en effet clair que nos comportements personnels peuvent avoir un impact négatif sur notre santé. Leurs effets peuvent être directs : la sédentarité, par exemple, est un facteur de risque des pathologies cardiovasculaires. Ils peuvent également être indirects, via les molécules que nous produisons et qui s'accumulent jusqu'à atteindre un seuil toxique (les fibres d'amiante, liées au cancer du poumon, par exemple).

Ces risques interpellent nos comportements, tant individuels que collectifs, mais ne doivent pas cacher que la mort elle-même fait partie intégrante de la dynamique de la vie. Tout se paie, comme l'illustre l'incidence en augmentation des pathologies neurodégénératives en raison de la progression de l'espérance de vie dans les pays développés. Pire : chaque victoire reste fragile, à l'image du problème de santé publique causé par la dissémination de souches bactériennes multirésistantes, particulièrement dans les hôpitaux.

De nouveaux médicaments plus efficaces peuvent apporter des éléments de réponse à ces défis, tout autant qu'à de nouvelles difficultés encore inconnues mais qui ne manqueront pas de se présenter au cours des années qui viennent. Gardons bien à l'esprit que cette course à l'armement thérapeutique est une fuite devant la mort, d'ores et déjà perdue, et que nous ne pouvons que nous concentrer sur les maladies, avec une grande humilité, dans l'espoir de soulager voire guérir ceux qui souffrent et, pourquoi pas, de prolonger encore un peu nos vies.

4.1 Vers de nouveaux médicaments

Le développement d'un nouveau médicament est un processus long et coûteux, allant de l'identification d'une cible biologique d'intérêt thérapeutique jusqu'au patient, dans lequel les essais cliniques succèdent au développement pré-clinique (voir Table I).

	Développement préclinique	Essais cliniques			Pharmacovigilance
		Phase I	Phase II	Phase III	Phase IV
Durée (an)	3.5	1	2	3	Surveillance après la mise sur le marché
Population test	Laboratoire et études animales	20 à 80 volontaires sains	100 à 300 patients volontaires	1000 à 3000 patients volontaires	
Objectif	Sécurité et activité biologique	Sécurité et dosage	Efficacité, effets indésirables	Efficacité, effets indésirables à long terme	
Succès	5,000 molécules évaluées	5 entrent en clinique			

Table I : Développer des médicaments : plusieurs étapes. Adapté de Dale E. Wierenga and C. Rovert Eaton, Office of Research and Development, http://www.allp.com/drug_dev.htm

L'identification et la mise au point d'une molécule possédant l'activité biologique recherchée se fait lors du développement pré-clinique. Elles nécessitent, entre autres, l'investigation de ses propriétés pharmacodynamiques et pharmacocinétiques, et se basent de plus en plus sur des approches informatisées (*in silico*), dont les résultats se retrouvent *in vitro* et *in vivo*, dans des cellules voire des organismes modèles.

Les essais cliniques commencent lorsque cette première molécule atteint une activité et une sécurité d'utilisation satisfaisantes, et après une phase de formulation galénique. La première phase de ces essais permet de vérifier que le futur médicament n'est pas toxique. La relation dose-effet est étudiée en profondeur lors de la seconde phase, avant de tester le futur médicament sur une cohorte de 1000 à 3000 patients volontaires (troisième phase).

Une fois sur le médicament sur le marché commence une quatrième phase, dite de pharmacovigilance, afin de détecter, évaluer, comprendre et prévenir les effets indésirables qui n'auraient pas été identifiés lors des essais cliniques.

En plus d'un total de douze ans d'efforts, ce long processus de développement de nouveaux médicaments est de plus en plus coûteux. De \$359 millions de dollars US en 1993 (Congressional Office of Technology Assessment¹), il est passé à \$897 millions à la fin des années 90 [101]. L'année dernière, une étude du Tufts Center for the Study of Drug Development a même chiffré le développement d'un nouveau produit biotechnologique à \$1.2 milliards de dollars US². Cette étude indique également que le coût est réparti en parts approximativement égales entre le développement pré-clinique et les essais cliniques (respectivement \$625 et \$615 millions de dollars US).

Alors que ces coûts augmentent, l'industrie pharmaceutique a rencontré plusieurs difficultés ces deux dernières années [47], qui se sont traduites par de nombreuses fusions/acquisitions guidées principalement par des impératifs à court terme [33]. Ces pratiques ne doivent pas cacher que la solution à long terme la plus vraisemblable est l'amélioration du processus de développement des médicaments [46].

Une telle optimisation doit permettre de raccourcir le temps entre la définition d'un besoin thérapeutique et la délivrance du médicament correspondant au patient [75], en gagnant du temps sur les projets les moins raisonnables, pour se focaliser sur les plus prometteurs [45]. Des temps de développement raccourcis pourraient également permettre de commercialiser des médicaments plus accessibles financièrement, de réagir plus rapidement face aux difficultés sanitaires à venir, ainsi que de diminuer le nombre de maladies orphelines.

La recherche translationnelle³ est sans doute l'une des clés de cette optimisation, en permettant au savoir biologique et médical, qui passe habituellement uniquement du laboratoire de recherche vers le patient, d'aller également dans l'autre sens. Elle nécessite un environnement de recherche multidisciplinaire dans lequel l'information retourne de l'*in vivo* à l'*in vitro* à l'*in silico*. L'une des implémentations les plus développées à l'heure actuelle est de permettre, via des phases cliniques précoces (phase

1 http://www.allp.com/drug_dev.htm

2 <http://csdd.tufts.edu/NewsEvents/NewsArticle.asp?newsid=69>

3 <http://nihroadmap.nih.gov/clinicalresearch/overview-translational.asp>

0), l'évaluation des propriétés pharmacodynamiques et pharmacocinétiques d'un médicament encore en développement, administré à des doses infinitésimales, afin de détecter d'éventuels problèmes [82]. Sans être l'ultime solution, une meilleure communication entre les différents intervenants impliqués dans la conception d'un nouveau médicament est essentielle afin de permettre un stockage, une gestion et une exploitation utile et rationnelle des énormes quantités d'informations (et du bruit qui leur est associé) générées par les méthodes à haut débit, de plus en plus recommandées et utilisées à chacune des étapes de recherche et développement [75].

Ces dix dernières années, de telles méthodes ont été utilisées principalement dans deux directions. Elles permettent l'identification de cibles biologiques d'intérêt thérapeutique [83], en particulier grâce au développement des techniques liées aux micro puces dans les domaines de la génomique et de la protéomique. L'explosion des cibles potentielles nécessite la mise au point en aval d'autres méthodes capables de concevoir plus efficacement de nouvelles molécules actives, en impliquant plus en profondeur la pharmaco- et la toxicogénomique, la médecine expérimentale [83] et, bien évidemment, la chimie computationnelle. Quelques-uns des différents rôles que cette dernière peut jouer lors de la conception de médicaments sont présentés dans la référence [73]. On trouve le criblage virtuel (CV), la conception *de novo* de molécules actives, évaluation de la ressemblance à un principe actif, et la détermination des interactions moléculaires (docking) entre le principe actif et sa cible biologique, le plus souvent une protéine.

4.2 Le docking

4.2.1 Énoncé

Les interactions entre molécules sont à la base de la plupart des mécanismes biologiques. Les détails de ces interactions, au niveau moléculaire, sont donc d'un très grand intérêt et peuvent être étudiés par cristallographie aux rayons X ou résonance magnétique nucléaire (RMN). Malheureusement, ces techniques ne permettent pas de détailler la totalité de ces interactions étant donné le nombre incalculable de molécules différentes au sein d'une seule cellule.

Le docking moléculaire *in silico* vise à prédire la structure d'un complexe moléculaire à partir des molécules isolées, ce qui est considérablement plus facile à mettre en oeuvre, moins cher et plus rapide que l'utilisation d'une des méthodes expérimentales mentionnées ci-dessus. Les logiciels de docking sont donc des outils très utiles en biologie, pharmacie et médecine, car la plupart des principes actifs sont de petites molécules (ligand) qui interagissent avec une cible biologique d'intérêt thérapeutique, généralement protéique (récepteur), afin d'influencer le mécanisme dans lequel cette protéine est impliquée. Ce travail de thèse se termine d'ailleurs par des applications du docking dans la conception de nouvelles molécules actives.

Après avoir énoncé le problème du docking, cette introduction présente quelques-unes des approches les plus couramment utilisées pour le résoudre, ainsi que leurs performances actuelles. Deux applications du docking dans le cadre de la mise au point de nouveaux médicaments sont ensuite présentées.

La procédure typique à mettre en place pour prédire les interactions entre un ligand et son récepteur est présentée en Figure 1. La première étape est l'obtention d'une structure pour le récepteur, par cristallographie aux rayons X, par RMN ou par modélisation. L'interaction prédite sera d'autant plus informative, précise et utile que la description de cette structure sera précise (Figure 2). Elle doit donc être inspectée avec précaution. Tout d'abord, elle doit correspondre à une conformation pertinente sur le plan biologique. Il est donc nécessaire, entre autres, de vérifier la présence de contacts cristallins, ou la présence et l'éventuelle interaction avec d'autres molécules du cristal. Ensuite, la qualité de la structure doit être vérifiée au niveau des atomes eux-mêmes. Par exemple, le docking d'un ligand ne pourra pas être pertinent si des atomes manquent dans la région du site d'interaction, si l'identité de séquence avec la structure utilisée comme modèle est trop basse (si la structure de la cible a été obtenue par modélisation par homologie), ou si la région d'interaction contient des acides aminés flexibles (identifiés par leurs B-factor élevés pour les structures rayons X, ou par de multiples conformations différentes pour celles identifiées par RMN).

Une fois identifiés, de tels problèmes doivent être pris en compte lors de la préparation de

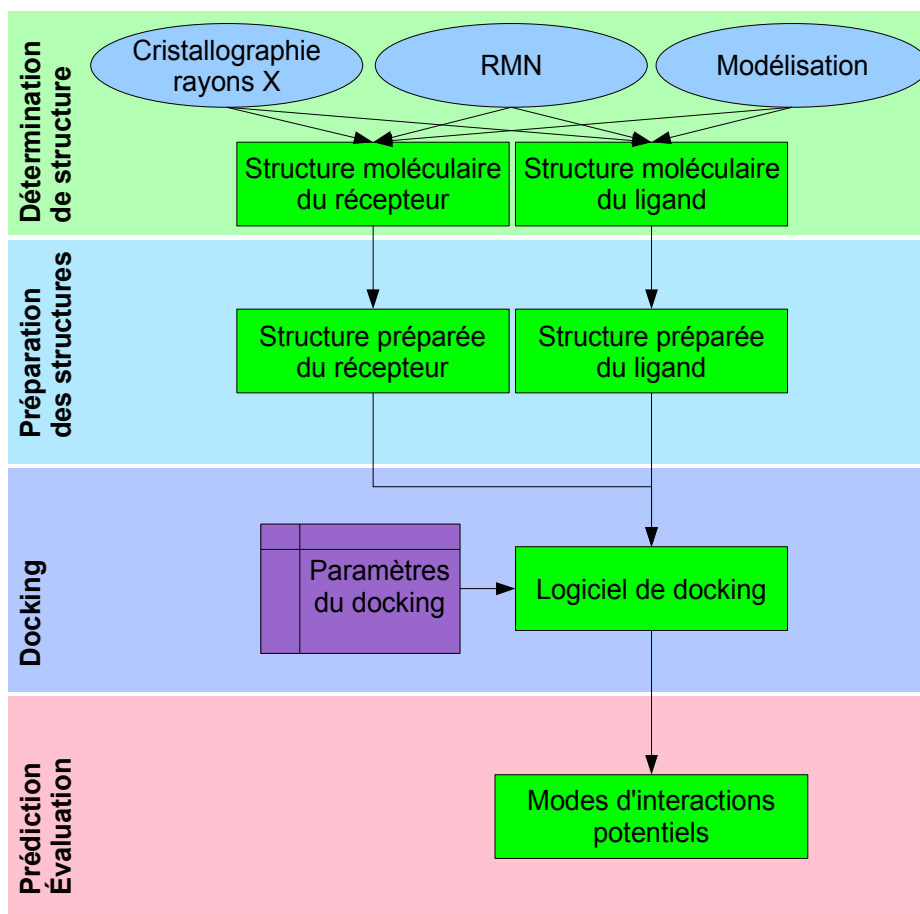


Figure 1 : Étapes typiques d'un docking. Voir détails dans le texte.

la structure en vue du docking. Il faut veiller à résoudre les problèmes de clashes stériques, ainsi qu'aux états de protonation. La conformation initiale du ligand est généralement modifiée et optimisée durant le docking, et n'est donc en principe pas très importante. Une fois les deux structures préparées, les paramètres *ad hoc* sont donnés au logiciel de docking, et ce dernier propose un ou plusieurs modes d'interactions potentiels qui peuvent être ensuite utilisés d'une multitude de façons différentes.

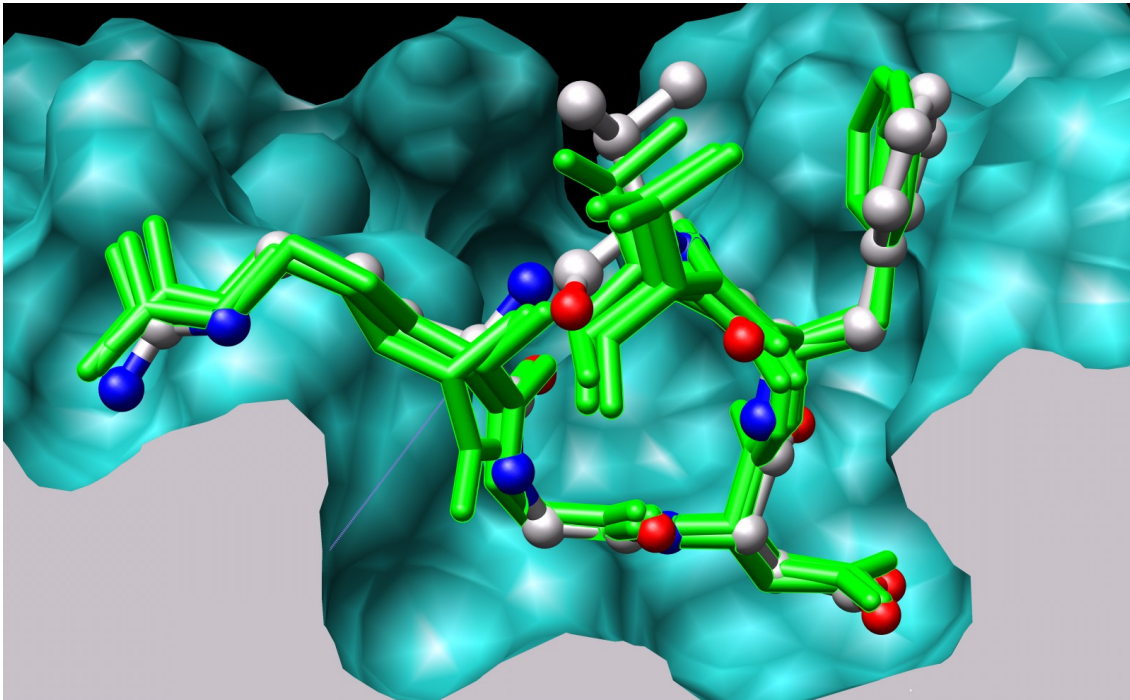


Figure 2 : Coupe de l'interaction (déterminée expérimentalement) entre le pentapeptide cyclisé du Cilengitide (boules/bâtons) en contact avec la surface de sa cible biologique, l'intégrine $\alpha V\beta 3$. Les modes d'interactions prédits par l'algorithme présenté dans cette thèse sont colorés en vert. La distance (root mean square deviation, RMSD) moyenne entre les structures prédites et la structure expérimentale est de 1.17 Å. Une telle précision ouvre la voie de l'optimisation rationnelle de molécules actives en se basant sur des critères structuraux.

4.2.2 Problème théorique et réponses méthodologiques

Pour prédire correctement un mode d'interaction, un programme de docking doit être capable d'en générer plusieurs, et de reconnaître parmi ceux-ci lequel est le bon. De ce fait, le docking peut être considéré comme l'optimisation de critères structuraux et énergétiques, décrits par une fonction objective, étant donné un ensemble de degrés de liberté décrivant la position et la conformation du ligand et du récepteur. Cette formulation relativement simple permet de mettre en évidence deux difficultés. Tout d'abord, étant donné la taille de l'espace de recherche, qui s'accroît considérablement avec le nombre de degrés de liberté (ddl), une exploration exhaustive de ce dernier n'est pas réalisable. Tout programme de docking repose donc sur une heuristique

d'échantillonnage générant des modes d'interactions. La seconde difficulté est de définir une fonction objective capable d'identifier parmi tous ceux générés, le mode d'interaction correspondant à la vraie interaction.

4.2.2.1 Heuristiques d'échantillonnages

Ces heuristiques sont nécessaires au regard de la taille de l'espace de recherche [76]. Elles peuvent être classées en trois familles distinctes : celle des approches dérivées d'échantillonnages systématiques, celle des techniques de dynamique moléculaire, et celle des méthodes stochastiques.

Un échantillonnage exhaustif est irréalisable, mais plusieurs heuristiques utilisent une approche exhaustive combinée avec des techniques de filtrage [48] ou avec une reconstruction incrémentale du ligand. Le principe général consiste à couper le ligand en fragments rigides et flexibles : un [40] [29] [99] ou plusieurs [133] [20] fragments rigides sont positionnés à la surface de la protéine à partir desquels le ligand est ensuite reconstruit.

Les méthodes basées sur les techniques de simulations par mécanique moléculaire (dynamique moléculaire et minimisation) sont séduisantes au regard des principes physiques sous-jacents. Cependant, elles sont coûteuses et leur capacité à franchir des barrières d'énergie libre élevées dans des temps de calcul réalistes est très limitée [28]. Néanmoins, l'échantillonnage peut être amélioré en abaissant ces barrières, via une réduction des composantes répulsives des interactions électrostatique et de van der Waals [135] [129], par exemple.

Les méthodes stochastiques (Monte Carlo, algorithme évolutionnaire (AE) et recherche taboue) sont des méthodes d'optimisation très générales et sans fondement physique, mais ont l'avantage de pouvoir explorer l'espace de recherche en ignorant les barrières d'énergie.

Les algorithmes évolutionnaires (AE) sont des méthodes d'optimisation génériques et itératives, reposant sur un processus adaptatif similaire à celui de l'évolution naturelle. Ils

sont classés, avec les réseaux de neurones et les systèmes à logique floue, parmi les techniques d'intelligence artificielle. À l'inverse de la plupart des techniques d'optimisation, ils sont capables de gérer plusieurs solutions simultanément, regroupées dans une « population » (voir Figure 3).

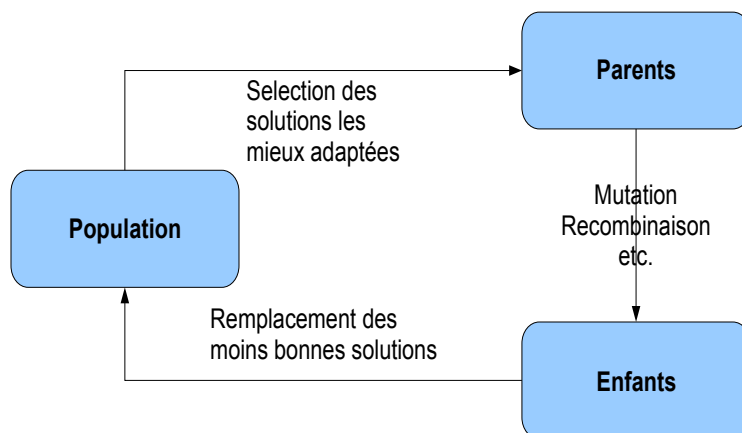


Figure 3 : représentation schématique du cycle évolutif typique d'un AE. Voir détails dans le texte.

Cette population est exposée à une pression de sélection, implémentée dans une fonction objective qui décrit le problème à optimiser et dans laquelle plusieurs ddl interviennent. L'ensemble des ddl permettant de définir une solution est appelé « génome ».

La sélection des solutions les plus performantes d'après la fonction objective (appelées « parents ») est contrebalancée par la génération de nouvelles solutions (« enfants »), afin de maintenir la diversité de la population. Les enfants sont générés par modification du génome parental par des opérateurs, classés suivant le nombre de parents sur lesquels ils s'appliquent. Les opérateurs unaires sont appelés « mutations », les opérateurs binaires « recombinaisons ». Dans les AE Lamarckiens, une recherche locale supplémentaire est réalisée à proximité de tout génome nouvellement créé, de façon à ce que l'enfant correspondant se place dans un minimum local pertinent. Une telle stratégie améliore sensiblement la recherche [102]. Cette astuce peut être vue comme des adaptations du phénotype introduites dans le génome.

Une fois les enfants créés, ils prennent la place des plus mauvaises solutions de la

population. Cette dernière est ensuite de nouveau exposée à la pression de sélection implémentée par la fonction objective, et l'évolution continue par une nouvelle itération (génération). Au fur et à mesure des générations, des solutions de mieux en mieux adaptées au problème décrit par la fonction objective émergent. Ces étapes se poursuivent jusqu'à convergence de la population, ou après un nombre déterminé de générations. Appliquée au problème du docking, cette dernière cherche à décrire les interactions entre le ligand et le récepteur, et contient les ddl correspondant aux positions, orientations et conformations du ligand et du récepteur.

De telles méthodes sont générales car elles peuvent s'appliquer à n'importe quel type de problème sans adaptation algorithmique particulière, même lorsque les fonctions objectives sont discontinues et/ou bruitées. Une telle versatilité rend les AE très intéressants pour des problèmes très différents, dans lesquels ils interviennent comme « boîtes noires ». Cependant, de telles implémentations naïves peuvent également être combinées avec des connaissances spécifiques du problème à résoudre, afin de diriger l'échantillonnage avec plus d'efficacité, par exemple via l'utilisation d'opérateurs semi-stochastiques ou via une recherche locale adaptée à la surface d'énergie sous-jacente. Ces algorithmes évolutionnaires hybrides ont montré de meilleures aptitudes que les implémentations naïves [34].

Les AE reposent sur un équilibre entre la génération de diversité (par les opérateurs) et la pression de sélection (par la fonction objective). Il permet de contrôler la distribution des solutions dans l'espace de recherche, de façon à spéculer efficacement sur les nouvelles solutions à explorer en priorité car, au regard de celles déjà évaluées, elles pourraient être plus adaptées. Une grande diversité combinée avec un renouvellement lent de la population rend les AE robustes mais lents, relativement similaires à une recherche par Monte-Carlo. À l'inverse, une diversité réduite combinée avec un renouvellement rapide de la population risque de faire converger la population prématurément sur un minimum local inintéressant [53]. Ce biais d'échantillonnage, contrôlé par les paramètres évolutionnaires, est un aspect très puissant des AE, car ils peuvent ainsi être ajustés à la complexité du problème à résoudre.

Dans le cadre du docking, deux limites ne doivent pas être oubliées. D'abord, ce biais d'échantillonnage ne suit pas une statistique de Boltzmann, et de ce fait ne permet pas d'estimer des propriétés thermodynamiques du système telles que l'énergie libre de liaison. Ensuite, les AE ont une large composante stochastique et, de ce fait, il n'est pas certain de trouver une solution optimale dans un laps de temps raisonnable. Les AE hybrides [34], plus efficaces car spécifiquement adaptés au docking, sont désormais largement utilisés [102] [72] [124]. La plupart d'entre eux combinent un échantillonnage stochastique permettant de franchir les barrières énergétiques afin d'obtenir des minima raisonnables, raffinés ensuite par une recherche locale, par exemple par minimisation d'énergie [102]. Plusieurs publications récentes [51] [105] [3] font également état d'une décomposition du docking en deux étapes. D'abord, les poches à la surface de la protéine susceptibles d'accueillir le ligand sont identifiées, puis la conformation du ligand dans ces dernières est optimisée. De telles approches se sont montrées très efficaces dans le cadre du criblage virtuel (approche consistant à trier plusieurs milliers de petites molécules provenant généralement d'une base de données en vue d'isoler celles qui ont le plus de chance d'avoir l'activité souhaitée) [27], mais ce travail a récemment été très critiqué [110].

4.2.2.2 Fonctions objectives

En plus des problématiques liées aux heuristiques d'échantillonnages, le problème des logiciels de docking le plus souvent documenté concerne les performances des fonctions objectives qui sont responsables de diriger cet échantillonnage, par le classement des modes d'interactions générés [76]. La fonction objective idéale serait celle qui donnerait, pour un mode d'interaction donné, l'énergie libre de liaison entre le ligand et son récepteur [119] [79]. En effet, elle rendrait également possible le classement par affinités de différents ligands, d'une importance toute particulière dans le cadre du CV. Pour la prédiction du mode d'interaction d'un ligand unique, les énergies libres de liaison relatives seraient suffisantes. Malheureusement, de tels calculs sont trop coûteux pour être utilisés tels quels dans un logiciel de docking car la plupart de ces derniers ont été mis au point dans l'optique CV, ce qui implique le docking de plusieurs dizaines de milliers

de ligands d'une base de données afin d'en identifier les plus prometteurs. Leurs fonctions objectives doivent donc être très rapides, ce qui implique des approximations. À l'inverse, la conception rationnelle de molécules nécessite une fonction objective malheureusement coûteuse, afin de prédire avec une grande exactitude le mode d'interaction de quelques dizaines de ligands. Une bonne fonction objective se doit d'être à la fois efficace et sélective [129]. Elle doit pouvoir diriger l'échantillonnage d'une façon aussi continue que possible (c'est à dire qu'elle doit fournir un gradient d'énergie), tout en étant capable d'identifier le mode d'interaction correct parmi un ensemble de leurres (c'est à dire que ce mode d'interaction correct doit avoir une énergie plus favorable que les leurres, qui sont des modes d'interactions qui ne correspondent pas à la structure expérimentale).

Les fonctions objectives se classent en trois familles : les fonctions empiriques, celles basées sur l'observation de structures déjà résolues, et celles reposant sur des champs de force.

Les fonctions objectives empiriques somment en les pondérant des termes décrivant les différents types d'interactions moléculaires, comme les liaisons hydrogène, les interactions ioniques ou de van der Waals [66] [108]. La pondération est généralement optimisée grâce à une base de données de complexes dont les structures et les énergies libres de liaisons sont connues. Pour cette raison, leurs transférabilités hors de cette base de données sont plus limitées que celles des fonctions objectives basées sur des champs de force.

La seconde famille de fonctions objectives se base sur des potentiels de force moyenne (PMF), calculés à partir de nombreuses structures tridimensionnelles résolues expérimentalement [15] [14] [63] [104] [103].

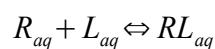
La troisième famille repose sur les champs de force de mécanique moléculaire, qui somment les énergies internes et d'interaction des deux partenaires, idéalement en prenant en compte l'effet du solvant. Si la protéine reste rigide durant le docking, son énergie interne est constante et peut donc être ignorée afin d'accélérer l'évaluation d'un mode d'interaction par la fonction objective. Ces dernières sont généralement sensibles

aux petites variations de coordonnées atomiques, ce qui limite leurs champs d'applications [43]. Cet inconvénient trouve une réponse partielle dans l'utilisation de potentiels de van der Waals adoucis, dans lesquels la contribution de la composante répulsive est limitée afin d'autoriser quelques clashes stériques sans trop pénaliser le mode d'interaction correspondant. Malheureusement, la sélectivité s'en ressent [23]. Ces fonctions objectives n'étant pas entraînées sur des structures existantes, leurs transférabilités à des applications réelles sont meilleures. DARWIN [124] et CDOCKER [135] sont deux logiciels de docking qui reposent sur de telles fonctions objectives, calculées par le logiciel CHARMM [18].

Aucune fonction objective n'est parfaite (cf. ci-dessous), mais l'utilisation combinée de plusieurs d'entre elles permet de réduire considérablement les modes d'interactions irréalistes qui peuvent avoir une énergie favorable selon l'une ou l'autre [26].

4.2.3 Modélisation des interactions moléculaires

L'association non covalente et réversible d'un récepteur (R) et d'un ligand (L) pour former un complexe récepteur-ligand (RL), dans un milieu aqueux contenant des électrolytes peut être décrit par l'Équation 1:



Équation 1

À l'équilibre, cette réaction est déterminée par l'énergie libre de liaison ΔG° . Cette quantité est liée (Équation 2) aux constantes d'association (K_A) et de dissociation (K_d), qui peuvent être déterminées expérimentalement (Équation 3). Elle contient une composante enthalpique (ΔH°) et entropique ($T\Delta S^\circ$) (T étant la température absolue).

$$\Delta G^\circ = -RT \ln K_A = \Delta H^\circ - T \Delta S^\circ$$

Équation 2

$$K_A = K_D^{-1} = \frac{[RL]}{[R][L]}$$

Équation 3

Le champ de force de mécanique moléculaire de CHARMM [96] est une fonction d'énergie potentielle empirique pour la modélisation moléculaire et l'étude dynamique des protéines. Cette fonction d'énergie inclut des termes liés et non liés décrits dans la Table II.


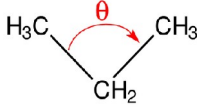
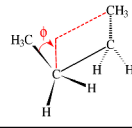
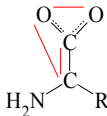

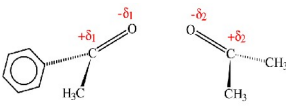
Terme énergétique	Illustration	Équation
Longueur des liaisons		$E_{bond} = \sum_{bonds} k_r (r - r_0)^2$
Angles des liaisons		$E_{angle} = \sum_{angles} k_\theta (\theta - \theta_0)^2$
Interactions liées		
Angles dièdres		$E_{torsion} = \sum_{dihedrals} k_\phi [1 + \cos(n\phi - \delta)]$
Angles impropres		$E_{impropers} = \sum_{angles} k_\omega (\omega - \omega_0)^2$
van der Waals		$E_{vdW} = \sum_{i,j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$
Interactions non liées		
Électrostatique		$E_{elec} = \sum_{i,j} \frac{q_i q_j}{\epsilon r_{ij}}$

Table II : Termes énergétiques utilisés dans CHARMM.

Les variations de longueur des liaisons covalentes sont modélisées par un potentiel harmonique, dans lequel r est la longueur de la liaison, r_0 la longueur à l'équilibre, et k_r la constante de force de liaison. Les variations d'angles entre liaisons covalentes sont également modélisées par un potentiel harmonique, avec θ représentant l'angle, θ_0 l'angle à l'équilibre, et k_θ la constante de force de l'angle. La torsion des angles dièdres

est décrite par une fonction cosinus, où ϕ est l'angle dièdre, k_ϕ la constante de force associée, n la périodicité et δ la phase. Les interactions entre deux atomes non liés i et j sont modélisées par un potentiel électrostatique de Coulomb et par un potentiel de Lennard-Jones pour les interactions de van der Waals, avec q_i et q_j les charges atomiques des atomes i et j , ϵ_{ij} le minimum du potentiel de Lennard-Jones, σ_{ij} le diamètre de Lennard-Jones, r_{ij} la distance et ϵ la constante diélectrique.

La fonctionnelle complète de l'enthalpie de CHARMM est une sommation de ces termes, voir Équation 4.

$$E = \sum_{\text{bonds}} k_r (r - r_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{impropers}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{dihedrals}} k_\phi [1 + \cos(n\phi - \delta)] + \sum_{i,j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i,j} \frac{q_i q_j}{\epsilon r_{ij}}$$

Équation 4

L'énergie de solvation électrostatique $\Delta G_{elec,solv}$ peut être calculée par le modèle de Born généralisé analytique GB-MV2 [88] [87]. Il est capable de reproduire les énergies de solvation électrostatique calculées par l'équation de Poisson, la référence des modèles de solvation implicite, avec une précision de 1 %. GB-MV2 est particulièrement intéressant car il est 20 fois plus rapide que la résolution de cette équation, et permet donc de calculer $\Delta G_{elec,solv}$ pour un grand nombre de structures. L'équation de Born généralisée prend la forme suivante, avec ϵ la permittivité relative du solvant, q_i et q_j les charges atomiques des atomes i et j , r_{ij} la distance entre les deux particules,

$$\alpha_i \text{ et } \alpha_j \text{ les rayons de Born, et } D_{ij} = \frac{r_{ij}^2}{(K_s \alpha_i \alpha_j)} \quad (K_s \text{ étant une constante égale à } 8).$$

$$\Delta G_{(elec,solv)} = \frac{-1}{2} \left(1 - \frac{1}{\epsilon}\right) + \sum_i \sum_j \frac{q_i q_j}{(r_{ij}^2 + \alpha_i \alpha_j e^{-D_{ij}})^{0.5}}$$

Équation 5

4.3 État de l'art du docking moléculaire

4.3.1 Implémentations courantes

Plus de 30 logiciels de docking sont actuellement disponibles [121] et la plupart sont conçus pour le CV. Les cinq les plus fréquemment cités pèsent 65 % du total de citations trouvées dans la littérature : AutoDock (27 %) [102], GOLD (15 %) [72], FlexX (11 %) [115], DOCK (6 %) [40] et ICM (6 %) [1]. Il est intéressant de remarquer que les deux logiciels les plus cités, AutoDock et GOLD, reposent sur une fonction objective basée sur un champ de force combiné avec un AE, ces derniers permettant facilement d'implémenter l'heuristique d'échantillonnage d'un côté et la fonction objective de l'autre. La combinaison de ces éléments semble donc remporter un succès certain.

4.3.2 Performance

Une évaluation de 37 fonctions objectives et de 10 logiciels de docking (et de plusieurs protocoles de docking pour chacun d'entre eux), a été réalisée sur 8 cibles pertinentes sur le plan pharmacologique [131]. En considérant le protocole le plus efficace et l'ensemble des modes d'interactions proposés par les logiciels de docking, au moins l'un d'entre eux a été capable de positionner correctement plus de 40 % des ligands connus et cristallisés à moins de 2 Å de RMSD de la structure cristalline pour 7 des 8 cibles. Pour certaines de ces cibles, le taux de succès monte jusqu'à 90 % voire 100 % en prenant en compte les prédictions à moins de 2 Å ou 4 Å. Ces résultats prometteurs ont été obtenus dans des conditions honnêtes mais optimales, en impliquant des chimistes computationnels experts dans chacune des cibles, ou experts dans l'un des logiciels de docking testés. La conclusion des auteurs de cette étude est que les logiciels de docking sont capables d'échantillonner l'espace de recherche suffisamment bien pour générer des modes d'interactions corrects, malgré les différences entre les cibles visées. Devant la baisse de

performance observée lorsqu'ils ne considèrent que les modes d'interactions les plus performants au regard des fonctions objectives, les auteurs constatent que ces dernières ont un succès moindre.

Globalement, il semble qu'aucun logiciel de docking ne soit universel [131] [109], même si certains d'entre eux (comme GOLD [80]) aboutissent régulièrement à des prédictions plus fiables que les autres. Dans des applications réelles, il est donc souhaitable d'utiliser plusieurs d'entre eux et de recouper leurs prédictions. Le lecteur doit également garder en tête qu'une comparaison rigoureuse entre différents logiciels impliquerait une validation statistique des différences de précisions, ce qui n'a pour l'instant jamais été le cas [30].

4.4 Deux applications des logiciels de docking dans la conception de nouveaux médicaments

Comme il a été mentionné plus haut, les logiciels de docking sont des outils utiles en pharmacie et en médecine, car la plupart des PA sont de petites molécules (ligands) conçues pour interagir sur le plan moléculaire avec leurs protéines cibles (récepteurs), afin d'agir sur les voies biologiques dans lesquelles ces dernières sont impliquées. L'identification d'une nouvelle molécule active est une étape importante sur le chemin entre le besoin d'une activité biologique et la délivrance d'un médicament au patient. Deux des applications les plus courantes des logiciels de docking destinées à cette identification sont présentées ci-dessous.

Les efforts et les améliorations des techniques cristallographiques se traduisent en pratique par une croissance exponentielle du nombre de structures résolues, parmi lesquelles des cibles potentielles pour l'industrie pharmaceutique. Les approches basées sur ces structures ont donc le vent en poupe depuis quelques années. La plupart de ces approches essaient de limiter le nombre d'expériences nécessaires à la validation de l'activité de futurs PA potentiels en identifiant *in silico* les molécules ayant les activités, biodisponibilités et toxicités souhaitées. Un tel filtrage est très efficace puisque 5000 molécules actives sont nécessaires pour avoir la chance que l'une d'entre elles devienne

un PA (Table I). Les PA des médicaments disponibles sur le marché sont généralement proches des premières molécules identifiées (les « hits ») avec l'activité désirée [114]. Les caractéristiques de ces dernières sont donc très importantes, et il est donc plus prudent de chercher à identifier plusieurs familles chimiques différentes pour l'activité recherchée [75] [131]. L'échantillonnage de l'espace chimique est donc un paramètre crucial lors de la recherche de hits. Les méthodes *in silico* sont particulièrement douées sur ce plan, en particulier lorsque l'on considère le CV ou la conception rationnelle de molécules [75]. Ces méthodes sont désormais largement utilisées [75] et peuvent également générer des profits substantiels [125]. Les logiciels de docking occupent une place centrale dans ces deux approches. On les retrouve le plus fréquemment dans le CV, qui essaie de trier plusieurs milliers de petites molécules (provenant généralement d'une base de données) en fonction de leur intérêt vis-à-vis d'une cible pharmacologiquement pertinente. La conception/optimisation rationnelle de molécules est complémentaire du CV, car elle propose des modifications d'un hit initial en fonction de la structure de son récepteur, et permet également la mise au point de ces hits initiaux, comme vous pourrez le constater dans le dernier chapitre de cette thèse. Ces deux applications des logiciels de docking sont présentées ci-dessous.

4.4.1 Le criblage virtuel

4.4.1.1 Aperçu

Les PA sont généralement de petites molécules organiques de moins de 30 atomes. L'espace chimique correspondant est estimé à environ 10^{60} molécules [38]. En plus de ce critère de taille, quatre autres règles permettent de caractériser la plupart des PA [95] : ils doivent avoir moins de 5 donneurs et 10 accepteurs de liaison hydrogène, un poids moléculaire inférieur à 500 g/mol, et un ClogP inférieur à 5. Malheureusement, les méthodes expérimentales à haut débit restent très loin de pouvoir tester toutes les molécules répondant à ces critères. Par contre, une fois que des molécules représentatives de cet espace sont identifiées, en prenant en compte leurs diversités et leurs ressemblances à des PA [73], les logiciels de docking peuvent identifier celles qui

ont le plus de chance d'être actives [86] (voir Figure 4). Ce filtrage permet de restreindre le nombre de molécules pour lesquelles une mesure d'affinité est nécessaire par la suite. Le CV est plus rapide, moins cher, et permet un meilleur échantillonnage de l'espace chimique que les méthodes à haut-débit expérimentales [77]. Ces deux techniques sont néanmoins complémentaires car, à l'instar de toutes les approches *in silico*, le CV nécessite une validation expérimentale *a posteriori* (voir Figure 4).

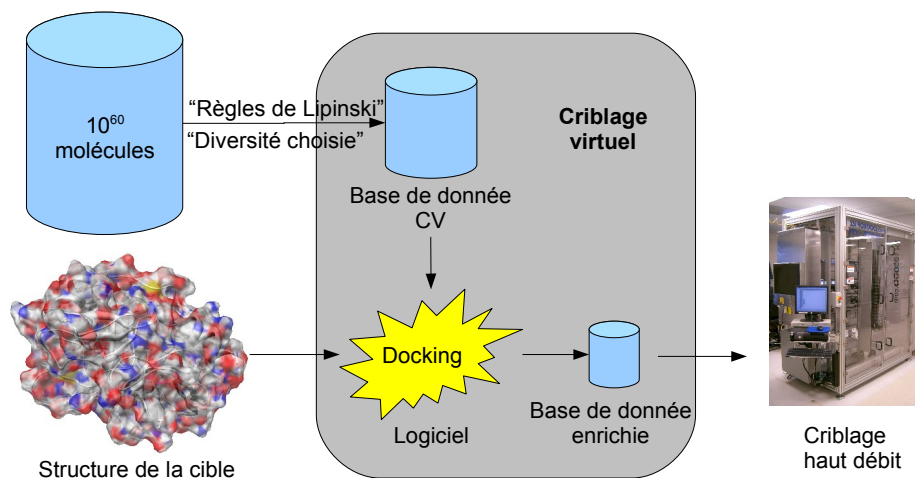


Figure 4 : Aperçu du criblage virtuel, voir détails dans le texte.

4.4.1.2 État de l'art

Une évaluation récente de 10 logiciels de docking et de 37 fonctions objectives sur 8 cibles biologiques a permis de conclure que le CV est capable d'identifier des molécules actives, parmi lesquelles plusieurs hits potentiels [131]. Cette étude a également permis de vérifier que des molécules inactives mais chimiquement proches de ces hits sont disqualifiées à juste titre, et que pour toutes les cibles sauf une, un enrichissement significatif a été permis par au moins une combinaison de logiciel de docking/fonction.

Ces résultats prometteurs ne doivent cependant pas cacher les performances souvent très variables du CV [131] [50]. Une autre évaluation de 10 fonctions objectives et de 4 logiciels de docking a montré que la discrimination des fonctions objectives actuelles est limitée par la précision des modes d'interactions générés par les heuristiques d'échantillonnages [81]. Cette observation met en exergue leurs importances trop souvent

négligées.

Une étude récente [74] a comparé les enrichissements par CV pour cinq cibles de structures connues, en utilisant soit les cinq structures correspondantes, soit cinq modèles par homologie (donc dérivés de structures proches), ou même en utilisant la structure des modèles eux-mêmes. Les enrichissements réalisés à partir des modèles par homologie n'ont pas été influencés par la qualité de la structure utilisée comme modèle, et ont été comparables voire meilleurs que ceux obtenus à partir des cinq structures cristallines [74]. Étant données les différences moléculaires entre la structure des cibles et celles des structures apparentées, allant au delà de la simple différence de conformation, il est encore plus surprenant de constater que les enrichissements réalisés à partir des structures apparentées elles-mêmes sont similaires aux autres. Ces observations semblent indiquer qu'il y a au mieux une corrélation extrêmement limitée entre la précision du logiciel de docking et l'enrichissement obtenu, comme cela a été mentionné par ailleurs [131] [86] [12]. Cette conclusion peut sembler surprenante au regard du développement du CV ces dix dernières années. Pour autant, il ne faut pas oublier que le problème du docking lui-même n'est pas résolu de façon universelle et définitive, et qu'il reste pour les années à venir de grands défis qui concernent la prédiction et la reconnaissance de modes d'interactions de plus en plus précis, ainsi que l'évaluation de leur énergie libre de liaison. D'autres études récentes par des auteurs reconnus vont même jusqu'à suggérer que l'enrichissement par CV ne serait pas dû à la discrimination positive des molécules intéressantes, mais au filtrage de celles qui ne le sont pas [86] [126]. D'un point de vue scientifique, cette conclusion est peu satisfaisante même si l'utilité du CV a été montrée par ailleurs [131].

4.4.2 Conception de médicament basée sur le docking de fragments

4.4.2.1 Aperçu

Les performances très variables du CV sont très probablement dues aux contraintes sur les vitesses de calculs, qui doivent rester compatibles avec le criblage de larges bases de données, et qui se répercutent dans les implémentations sous forme d'approximations

délétères. Il est possible de prendre ce problème autrement, en partant de l'observation suivante : la plupart des PA sont faits de régions relativement rigides, reliées entre elles par des connexions (Figure 5), les uns et les autres étant des fragments de molécules. Il est possible de positionner ces fragments à la surface de la cible, puis de les optimiser, de les connecter entre eux, de les agrandir ou de les relier, afin d'obtenir un hit, et dans le cas idéal, un nouveau PA (Figure 6, Figure 7)

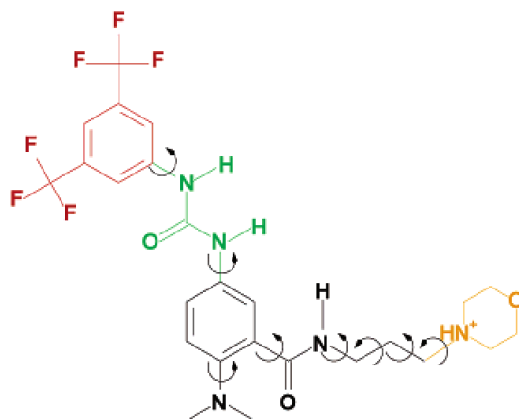


Figure 5 : décomposition en fragments colorés d'un inhibiteur de la β -Secretase [69] [78].

Essayer de mettre au point de nouveaux PA en partant de ces fragments permet de diminuer considérablement la taille des molécules à prendre en compte, tout en fournissant de bons points de départ pour une optimisation structurale [39], ces derniers peuvent avoir une affinité de l'ordre du millimolaire ou du micromolaire [24] [31].

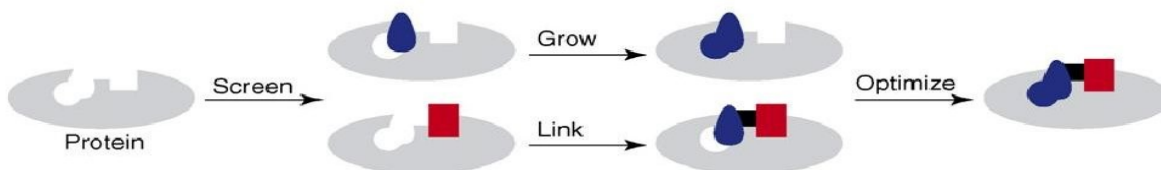


Figure 6 : principes de base de la conception de PA par fragments : l'ovale bleu et le rectangle rouge représentent des fragments qui se lient à la protéine cible. Voir détails dans le texte. [37]

Conséquence intéressante de cette approche basée sur des fragments, les molécules actives qu'elle permet d'identifier sont susceptibles de donner des hits bénéficiant d'une

meilleure énergie de liaison par atome, appelée efficacité du ligand (EL) [37] [24] [67], cette dernière étant d'un intérêt tout particulier (Figure 7).

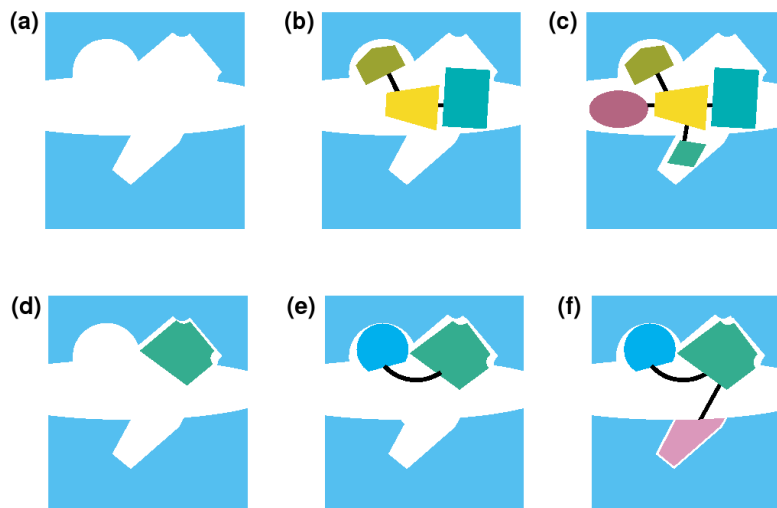


Figure 7 : Représentation schématique de recherches de nouveau hit par méthodes expérimentales haut débit et approche par fragment **(a)** représentation du site actif d'une protéine, composée de trois poches offrant de bons sites d'interactions aux inhibiteurs **(b)** représentation d'une molécule issue de méthodes expérimentales haut débit. Ce hit est déjà complexe et fait plusieurs interactions de qualité médiocres. Son affinité est la somme de toutes les interactions faites par la molécule complète. **(c)** représentation du hit obtenu en (b) après optimisation. Le site actif est mieux rempli, mais le poids moléculaire et la complexité sont encore augmentés. De tels ligands optimisés ont peu de chance de ressembler à des médicaments au regard des règles de Lipinski **(d)** représentation d'un fragment fixé par l'intermédiaire d'un petit nombre d'interactions très efficaces avec l'une des trois poches du site actif. À cause de leur petite taille, l'affinité de tels fragments est difficilement détectable expérimentalement. **(e)** représentation d'un hit faisant des interactions de bonne qualité dans le site actif, qui peut être facilement identifié à partir de (d) et d'un autre fragment faisant de bonnes interactions (en bleu). **(f)** représentation d'un hit optimisé, dérivé de (e) et encore amélioré, faisant toutes les interactions importantes avec les poches du site actif. ([24])

4.4.2.2 Définition des fragments

La définition des fragments peut se faire de plusieurs façons [78] [11] [10] [89] [56] [71] [130]. La tendance générale a conduit à définir des fragments suivant trois critères, corollaires des règles de Lipinski pour les molécules entières [31]. Ils ont en général un poids moléculaire inférieur à 300g/mol, un maximum de trois donneurs ou accepteurs de liaisons hydrogène, et un ClogP maximum de 3.

4.4.2.3 Méthodes *in silico*

La littérature récente contient plusieurs revues sur des approches expérimentales par fragments [37] [36]. Les approches *in silico* ont sur ces dernières les mêmes avantages et les mêmes inconvénients qu'a le CV sur le criblage expérimental à haut débit. Elles sont plus rapides, moins chères, plus faciles à mettre en oeuvre, mais doivent être suivies d'une validation expérimentale tout au long ou à la fin du processus d'optimisation. Comme auparavant, ces approches *in silico* permettent également un meilleur échantillonnage de l'espace chimique [75]. Elles sont aussi plus à même d'identifier un fragment dans une région intéressante car elles se concentrent sur les modes d'interactions les plus favorables, en souffrant moins de la sensibilité limitée des méthodes expérimentales lorsque les affinités sont faibles (voir chapitre Applications). L'identification de plusieurs modes d'interactions géométriquement superposés ou incompatibles est un autre avantage sur la cristallographie par rayon X, qui risque de ne proposer que le mode d'interaction pour le fragment dont l'affinité est la plus grande.

Les approches *in silico* comportent généralement deux étapes : tout d'abord, un squelette carboné est généré [93] [92] [91] [90] [94] [85] [128] [49]. Les types d'atomes sont ensuite assignés de façon à optimiser les contacts électrostatiques ou hydrophobes avec la protéine [127] [25] [8] [7] [6] [5] [4]. La synthèse de la molécule obtenue peut malheureusement ne pas être possible selon la méthode utilisée pour choisir/combiner les fragments.

Les fragments peuvent être positionnés à la surface de la cible biologique par des programmes tels que Ludi [13], GRID [54], X-CITE [84], SEED [97] ou MCSS [100]. Ils

peuvent ensuite être liés par des programmes tels que HOOK [35] et DLD [18] [123]. Ce dernier connecte de petits groupes fonctionnels dont les positions ont été déterminées expérimentalement ou par le logiciel Multiple Copy Simultaneous Search (MCSS, [100]). Sa fonction objective permet la sélection des fragments qui interagissent favorablement avec la cible, tout en respectant la longueur des liaisons, les angles entre elles, ainsi que les angles dièdres. Les molécules reconstruites sont ensuite optimisées par un échantillonnage par Monte-Carlo et un recuit simulé.

Une suite logicielle intéressante a été développée dans le groupe du Prof. Caflisch : les fragments, identifiés par DAIM [78] sont positionnés par le logiciel SEED [97], puis utilisés par FFLD [20] comme point d'ancrage pour proposer des modes d'interactions vraisemblables pour des molécules plus complexes. Cette approche a permis d'identifier des inhibiteurs micromolaires de la β -Secretase [69], une cible thérapeutique intéressante dans la maladie d'Alzheimer.

Un autre succès intéressant a été présenté dans la référence [16] : le CV d'une base de données de fragments, combiné avec un criblage à haut débit biaisé puis à une optimisation rationnelle des molécules intéressantes en prenant en compte leur environnement structural a permis l'identification d'inhibiteurs de la DNA gyrase. Le meilleur d'entre eux est dix fois plus puissant que l'inhibiteur de référence, la novobiocine.

4.4.2.4 Conclusion

La plupart des travaux de recherche de nouvelles molécules actives publiés ces cinq dernières années ont été menés dans l'industrie [125] [38], par des entreprises telles que Vertex, Sareum, Astex Technology ou Sunesis Pharmaceutical. Le positionnement des fragments, leurs connexions et optimisations *in silico* pourraient beaucoup se développer si des programmes de docking capables de prédire avec précision les positions et orientations des fragments apparaissent. Pour l'instant, aucune réponse définitive n'est apportée à la question du docking [86], mais les calculs demandés par ces approches, moins nombreux que ceux requis par le CV, permettent d'envisager des fonctions objectives plus précises, et des heuristiques d'échantillonnage plus efficaces.

4.5 Conclusion générale

Comme exprimé clairement dans la référence [121] : « Despite the very promising picture drawn, molecular docking still holds several hidden weaknesses, and the so-called docking problem is far from being solved. The lack of a suitable scoring function, able to efficiently combine both accuracy and speed, is perhaps the most detrimental weakness. The results of a docking experiment should therefore not be taken as the final result of a structural study, but rather as a good starting point for a deeper and more accurate analysis. In this sense, docking must be necessarily fast, enabling large quantities of data to be considered, and reasonable and coherent solutions to be generated. However, the final result (geometry, binding free energy) should always be determined by a more accurate and precise methodology, naturally slower ».

Le but principal de ce travail de thèse a été de concevoir un logiciel de docking utilisable, assez précis pour la conception rationnelle de médicaments, et pouvant également servir de boîte à outils pour proposer les nouvelles réponses et les nouveaux développements nécessaires afin d'obtenir des prédictions de modes d'interactions de meilleure qualité [86] : EADock.

5 EADock

La plupart des logiciels de docking ont été conçus dans l'optique du CV. Il en résulte une contrainte de vitesse qui implique que leurs fonctions objectives doivent être rapides (et donc approximatives) et/ou que l'échantillonnage soit limité aux alentours du site de fixation supposé sur la cible. Le logiciel de docking dédié à la conception rationnelle de molécules présenté dans ce travail, Evolutionary Algorithm for Docking (EADock), se présente comme une synthèse de quatre approches jusqu'alors présentées séparément [135] [124] [68] [9] [107].

Premier élément, l'heuristique d'échantillonnage implémentée dans EADock s'inspire des AE et combine deux fonctions objectives. La première, qui néglige les effets du solvant, est utilisée afin de diriger la recherche vers les minima locaux grâce à son efficacité et sa vitesse. Une fois identifiés, ces minima locaux sont ensuite évalués par une seconde fonction objective, plus sélective mais également plus coûteuse, qui prend en compte l'énergie libre de solvatation. Cette combinaison repose sur la correspondance des minima des deux fonctions objectives, même si leurs rangs peuvent être différents [120].

Second élément, un système inspiré par les recherches taboues limite dynamiquement l'espace de recherche au fur et à mesure de l'évolution, en conservant une liste des modes d'interactions défavorables visités afin d'éviter de les échantillonner à nouveau. Ce mécanisme biaisant l'exploration de l'espace conformationnel exploré permet d'éviter que l'évolution converge sur des modes d'interactions favorables selon la première fonction objective, mais pas selon la seconde, plus sélective.

Troisième élément, l'échantillonnage est réalisé par des opérateurs qui combinent une recherche globale et une recherche locale de l'espace conformationnel. Certains de ces opérateurs, gérant rotations et translations, font intervenir une composante stochastique. D'autres, appelés « opérateurs intelligents » permettent de franchir les barrières énergétiques en modifiant temporairement la surface d'énergie des fonctions objectives, d'une façon déterministe.

Quatrième élément, la manipulation des coordonnées et les calculs d'énergies sont réalisés par le logiciel CHARMM, connecté à EADock par l'intermédiaire d'un connecteur réalisé pour l'occasion. EADock est donc capable d'utiliser les dernières améliorations disponibles dans CHARMM, en particulier les modèles de solvation sophistiqués tels que GB-MV2 [88] [87].

La performance d'EADock a été évaluée sur 37 complexes utilisés dans une précédente étude [22]. Une prédiction correcte est définie par un RMSD entre le mode d'interaction prédit et la structure cristalline, en ne prenant en compte que les atomes lourds du ligand (nommé simplement RMSD par la suite), inférieur à 2 Å. Malgré des conditions initiales défavorables, un espace de recherche énorme et une évolution très courte, de tels modes d'interactions ont été identifiés avec succès pour 92 % des complexes, les classant au premier rang des modes d'interactions possibles dans 68 % des cas. L'un des échecs s'explique par la présence d'une liaison covalente entre le ligand et le récepteur, qui ne peut pas être prise en compte par notre algorithme. Les autres échecs peuvent s'expliquer par la présence de contacts cristallins entre le ligand et un complexe voisin de la maille cristalline.

5.1 Algorithme de docking

La Figure 8 donne un aperçu de l'algorithme implémenté dans EADock. Il est initialisé avec des paramètres relatifs au docking (telles que les coordonnées de référence pour la protéine et pour le ligand, ainsi qu'une liste de dièdres libres) ainsi qu'au processus évolutionnaire lui-même (comme la taille de la population et le nombre de générations). Bien évidemment, la structure cristalline n'est jamais utilisée en aucune façon pour introduire un biais ou guider le logiciel. La région d'intérêt (RI) est définie par une sphère qui peut, suivant son centre et son rayon, être limitée autour de la poche de fixation sur la cible, ou contenir totalement cette dernière.

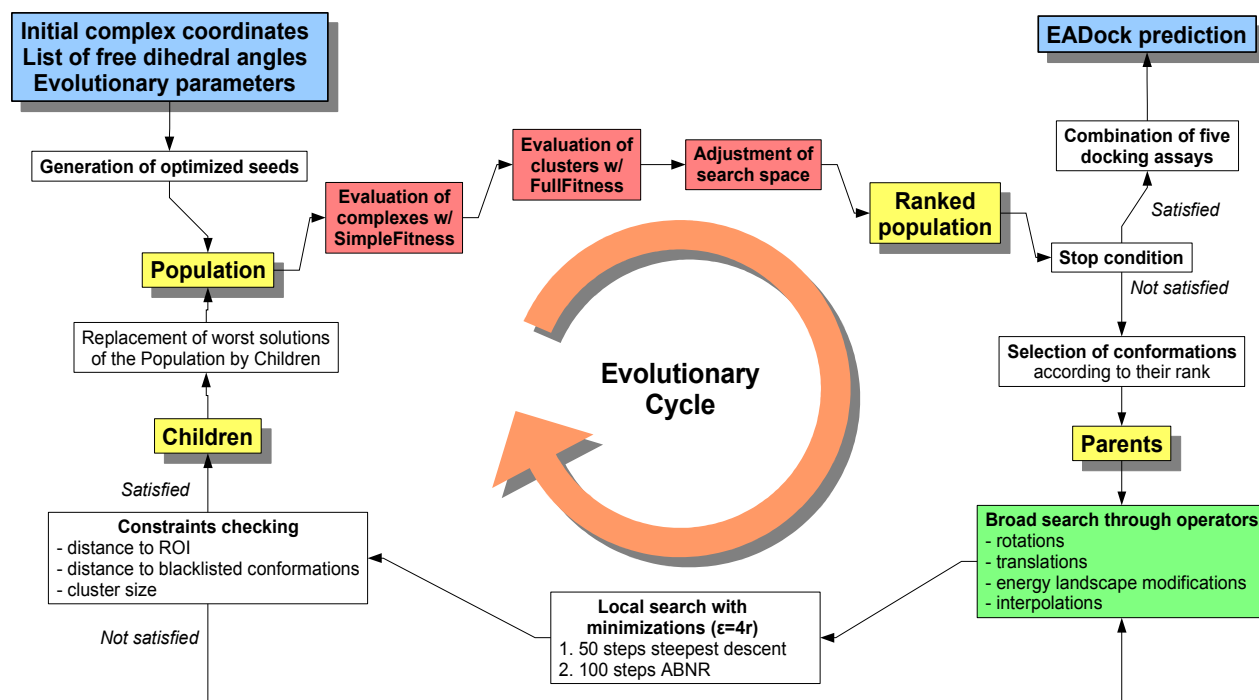


Figure 8 : Principales étapes de l'algorithme de docking implémenté dans EADock. Des paramètres typiques sont une taille de population de 250 complexes, 400 générations, un rayon de clustering de 2 Å, et une taille maximum de cluster de 8 éléments.

Sur le plan technique, EADock est implémenté en Java. Il se compose d'un moteur évolutionnaire générique, d'un code dédié au docking, lui-même interfacé avec un moteur de mécanique moléculaire. Pour ce travail, c'est le logiciel CHARMM qui a été choisi (Figure 9).

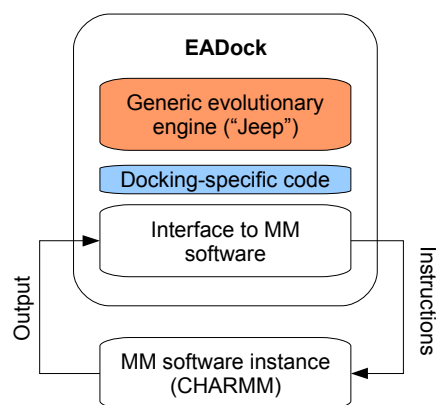


Figure 9 : Conception du logiciel

5.1.1 Génération de la population initiale : « seeding »

La première population, correspondant à la génération 0, est formée à partir de leurres générés à partir des coordonnées de référence du ligand : les seeds. Chaque seed est générée par une translation suivie d'une rotation aléatoire du ligand dans sa conformation de référence, elle-même suivie d'une optimisation séquentielle de chacun des dièdres définis par l'utilisateur. Les complexes obtenus sont ensuite optimisés par une routine appelée *smartAttractor*. Dans cette procédure, le ligand est translaté jusqu'à un minimum de l'énergie d'interaction, au contact de la surface du récepteur, le long d'un axe défini par leurs deux plus proches atomes. Ce minimum est identifié par une translation progressive du ligand vers la protéine par un pas de 0.1 Å suivie d'une minimisation de l'énergie.

Chaque angle dièdre est ensuite optimisé séquentiellement par l'opérateur *OptRot* (voir ci-dessous). Finalement, le ligand est minimisé à nouveau par 50 pas de steepest descent (SD) suivis par 100 pas d'Adopted Basis Newton-Raphson (ABNR). De cette façon, il est certain qu'aucune conformation trop irréaliste n'est utilisée comme seed, et que les angles dièdres, les longueurs et les angles des liaisons de chacun de ces modes d'interactions initiaux sont optimaux.

5.1.2 Sélection

Une fois la population générée, les deux fonctions objectives sont successivement utilisées pour identifier les complexes les plus intéressants. Ils sont d'abord triés par une fonction objective simple et rapide, la *SimpleFitness*. Ensuite, les *clusters de complexes* qui émergent (reflétant les minima locaux de la *SimpleFitness*) sont identifiés et évalués par la *FullFitness*, plus précise mais plus lente, et les rangs dans la population des centres de ces clusters sont ajustés afin d'influencer la sélection des parents de la génération suivante.

La *SimpleFitness* est égale à l'énergie totale du système calculée avec le champ de force de mécanique moléculaire CHARMM22, une constante diélectrique de 1, et sans troncature des interactions non liées:

$$SimpleFitness = E_{intra}^{ligand} + E_{intra}^{recept.} + E_{inter}$$

E_{intra}^{ligand} et $E_{intra}^{recept.}$ sont les énergies internes du ligand et du récepteur. Elles somment les interactions liées (liaisons, angles, etc.) et non liées (interactions électrostatiques et de van der Waals). Lorsque le récepteur est fixe, l'énergie interne de ce dernier, $E_{intra}^{recept.}$, est constante. E_{inter} , l'énergie d'interaction entre le ligand et le récepteur est égale à la somme des énergies d'interactions électrostatiques et de van der Waals. La *SimpleFitness* est rapide, mais ne tient pas compte de l'effet du solvant, bien que ce dernier contribue significativement à l'énergie libre de liaison. Cette approximation permet néanmoins d'identifier des solutions raisonnables [120].

Les clusters de modes d'interactions sont ensuite évalués par la *FullFitness*, après avoir été identifiés sur la base de la matrice de RMSD entre tous les complexes de la population. Le complexe le plus favorable est utilisé comme centre du premier cluster. Les modes d'interactions au voisinage de ce dernier (définis par un seuil de RMSD), sont inclus dans ce cluster initial. Le complexe restant le plus favorable est ensuite choisi comme centre du second cluster, peuplé par ses voisins suivant le même principe. Cette procédure se répète jusqu'à ce que toute la population soit découpée en clusters. Lorsqu'au moins trois clusters ont atteint leur taille maximale, leur *FullFitness* est calculée.

Cette *FullFitness* correspond à la moyenne du percentile 30 des énergies effectives de ses éléments, afin de ne pas pénaliser un cluster par quelques complexes énergétiquement défavorables. Cette énergie effective se compose de l'énergie totale du système à laquelle est ajouté un terme de solvatation. En négligeant la contribution entropique du soluté, il est possible d'écrire:

$$G_{eff} = E_{intra}^{ligand} + E_{intra}^{recept.} + E_{inter} + \Delta G_{elec, solv} + \sigma \times SASA$$

en calculant comme précédemment E_{intra}^{ligand} , $E_{intra}^{recept.}$, et E_{inter} . L'énergie de solvation se compose d'un terme électrostatique, $\Delta G_{elec,solv}$, et de contributions non polaires. Ces dernières peuvent être décrites par un terme de cavité et un terme de van der Waals entre soluté et solvant, couramment considéré proportionnel à la surface accessible au solvant (SASA) [65] [2]. Nous avons utilisé une valeur 0.0072 kcal/(mol Å²) pour le paramètre σ [52] [62] [122], et la SASA est calculée analytiquement par CHARMM.

$\Delta G_{elec,solv}$ est calculé par le modèle analytique Generalized Born Molecular Volume (GB-MV2) implémenté dans CHARMM, et environ 20 fois plus rapide que la résolution de l'équation de Poisson. Des études récentes ont montré que la déviation entre les énergies de désolvatation calculées par GB-MV2 et par Poisson-Boltzmann (PB) est constante pour une série de différentes conformations d'un même complexe, ce qui veut dire que les rangs des complexes calculés par GB-MV2 seront identiques à ceux calculés par PB [138].

Chaque fonction objective modifie le rang des complexes de la population, et ce rang est utilisé pour sélectionner les parents de la génération suivante, les centres des clusters dont la *FullFitness* est la plus favorable étant placés plus favorablement que tous les autres complexes. Inversement, les centres des clusters dont la *FullFitness* est la moins favorable sont retirés de la population et ajoutés à la liste des conformations interdites.

Les parents sont choisis de façon à raffiner et peupler les clusters identifiés, afin qu'ils puissent entrer en compétition face à la *FullFitness* aussi souvent que possible, pour limiter le risque de perdre des complexes intéressants à cause de la sélectivité limitée de la *SimpleFitness*. Les parents sont sélectionnés parmi la moitié des complexes de la population dont le rang est le plus favorable, d'abord en fonction de leur rang à l'intérieur du cluster auquel ils appartiennent, puis ensuite du rang de ce cluster parmi les autres clusters. Le meilleur élément de chaque cluster est sélectionné en premier, puis le second de chaque cluster, et ainsi de suite. Cette sélection se poursuit avec les éléments de rangs inférieurs jusqu'à ce que suffisamment de parents aient été sélectionnés. De cette façon, les membres des petits clusters peuvent être sélectionnés plusieurs fois.

Pour résumer, EADock utilise deux fonction objectives sur deux niveaux différents : les

complexes sont classés d'après leur *SimpleFitness* (rapide, efficace) afin de garantir des interactions électrostatiques et de van der Waals raisonnables. Les clusters de complexes sont ensuite évalués par la *FullFitness* (lente, sélective) qui prend en compte l'effet du solvant, et l'espace de recherche est ajusté en conséquence. Les deux fonctions objectives modifient le rang des complexes dans la population, afin de permettre une sélection optimale des parents de la génération suivante. Après un nombre de générations défini par l'utilisateur, l'évolution s'arrête.

5.1.3 Génération de diversité

Un ou deux complexes « parents » sont nécessaires à la création d'un « enfant ». Il(s) est (sont) sélectionné(s) comme décrit ci-dessus, puis modifié(s) par un opérateur. Les opérateurs implémentés dans EADock combinent une recherche globale (voir ci-dessous) suivie d'une recherche locale par minimisation d'énergie. Il a été montré précédemment que cette dernière permet d'accélérer la convergence tout en améliorant la qualité des prédictions [102] en résolvant les petits problèmes stériques qui peuvent être introduits par la recherche globale, et par l'ajustement des angles de valence et de la longueur des liaisons. Une fois qu'un enfant a été généré dans l'espace de recherche et confronté à la liste de complexes interdits, il est inclus dans la population. S'il se rattache à un cluster déjà plein, il remplace l'élément possédant la *SimpleFitness* la moins favorable. Sinon, c'est la moins bonne solution de la population qui est remplacée. Ce mécanisme permet d'éviter une convergence prématurée de la population sur un minimum unique.

Les nouveaux complexes peuvent être générés par plusieurs opérateurs. Quatre d'entre eux optimisent la position et l'orientation du ligand par rapport à la protéine, faisant des rotations et des translations aléatoires. Pour chacun, deux jeux de paramètres permettent de limiter l'amplitude de ces mouvements (rotations jusqu'à 40° et translations jusqu'à 2.5 Å), ou de faire des sauts importants lors de la recherche (par des rotations libres et des translations jusqu'à 10 Å). Ces derniers opérateurs sont appelés par la suite « opérateurs à longue distance », les précédents et tous les autres étant des « opérateurs à courte distance ». La routine *SmartAttractor* est appelée après chaque opérateur longue distance, car ces derniers sont susceptibles de modifier en profondeur le complexe parent,

généralisant des clashes stériques ou aboutissant à des complexes sans interactions entre le ligand et le récepteur.

L'opérateur *OptRot* optimise les angles dièdres du ligand laissés libres par l'utilisateur. À chaque appel, un de ces angles dièdres est aléatoirement choisi puis optimisé : deux groupes d'atomes sont d'abord identifiés, de part et d'autre de la liaison covalente centrale du dièdre (qui définit l'axe de rotation). Le premier groupe d'atomes est fixé, et le second subit des rotations incrémentales par pas de 60°. Les poses générées sont minimisées par 50 pas de SD suivis de 100 pas de ABNR, et sont évaluées par la *SimpleFitness*. La conformation du second groupe d'atomes avec la *SimpleFitness* la plus favorable est sélectionnée. Les rotations sont ensuite répétées en inversant les deux groupes d'atomes afin de faire tourner le premier, et de la même façon, le rotamère dont la *SimpleFitness* est la plus favorable est conservé.

Trois opérateurs, *ElectrostaticOptimizer*, *VanDerWaalsOptimizer* et *SoftLigand*, ont été conçus pour modifier le mode d'interaction du parent par l'intermédiaire d'une minimisation d'énergie dans un champ de force temporairement altéré. D'une certaine façon, cette stratégie peut être comparée à l'utilisation des potentiels de van der Waals adoucis décrits précédemment [135] [129], qui lissent également la surface d'énergie. Ces trois opérateurs reposent sur un principe de fonctionnement commun : d'abord la contribution relative d'un des termes énergétiques est artificiellement augmentée ou diminuée, ensuite une minimisation d'énergie dans ce champ de force altéré est réalisée. Finalement, les contributions relatives des différents termes énergétiques sont restaurées et une minimisation additionnelle est réalisée pour relaxer le ligand dans le champ de force correct.

ElectrostaticOptimizer et *VanDerWaalsOptimizer* augmentent transitoirement les contributions électrostatiques ou de van der Waals d'un facteur cinq. Les deux minimisations correspondant à la recherche et à la relaxation sont identiques : 50 pas de SD, suivis par 100 pas de ABNR. Dans l'opérateur *SoftLigand*, c'est l'énergie interne du ligand qui est diminuée d'un facteur quatre. Cette altération du champ de force permet au ligand de se déformer afin de franchir les barrières d'énergies et d'améliorer son

interaction avec la surface de la protéine par minimisation (150 pas d'ABNR). 500 pas d'ABNR sont ensuite réalisés pour relaxer le système.

Le dernier opérateur, *Interpolator*, part de deux complexes parents, distants de plus de 0.2 Å et de moins de 5 Å. Une série de conformations interpolées est générée et chacune est optimisée par le *SmartAttractor* (voir 4.1.1). Celle dont l'énergie est la plus favorable est conservée.

Après avoir sélectionné aléatoirement un complexe parent, un opérateur est choisi en fonction de sa probabilité d'être utilisé, cette dernière étant ajustée automatiquement en fonction des améliorations de *SimpleFitness* qu'il a permis au cours des cinq générations précédentes. Ce mécanisme, appelé automatic operator scheduling, a été décrit précédemment [34]. En quelques mots : lorsqu'un enfant est créé, l'opérateur appliqué est crédité de la différence de *SimpleFitness* entre l'enfant et le parent. Toutes les cinq générations, la probabilité des opérateurs est ajustée en fonction de ce crédit. De plus, selon la taille du cluster auquel le parent appartient, un biais est introduit. Si ce cluster a déjà atteint sa taille maximale, les opérateurs à longue distance ont plus de chance d'être choisis afin d'explorer une région distante de ce minimum déjà identifié. Inversement, les opérateurs à courte distance sont appliqués plus fréquemment lorsque le cluster du parent est presque vide, afin de le peupler par de nouveaux éléments raffinés, pour en évaluer rapidement la *FullFitness*.

5.1.4 Traitement final

La fiabilité de chaque essai de docking est améliorée en combinant plusieurs évolutions indépendantes (généralement cinq). Les complexes desquels l'énergie effective a été calculée sont regroupés dans une population optimisée, dont les minima sont identifiés par clustering. Les nouveaux clusters identifiés sont classés par leur *FullFitness*, comme décrit précédemment.

Une prédiction est considérée comme correcte lorsque le RMSD entre le cluster le plus favorablement classé et la structure cristalline est inférieur à 2 Å. Si aucune conformation

respectant ce critère n'a été échantillonnée, l'erreur est causée par l'échantillonnage. Si une conformation correcte a été générée, mais perdue avant son évaluation par la *FullFitness*, l'erreur est due à la *SimpleFitness*. Si un cluster correct a été évalué par la *FullFitness*, mais perdu par la suite à cause d'un score défavorable, l'erreur provient de la *FullFitness*.

5.1.5 Visualisation des résultats

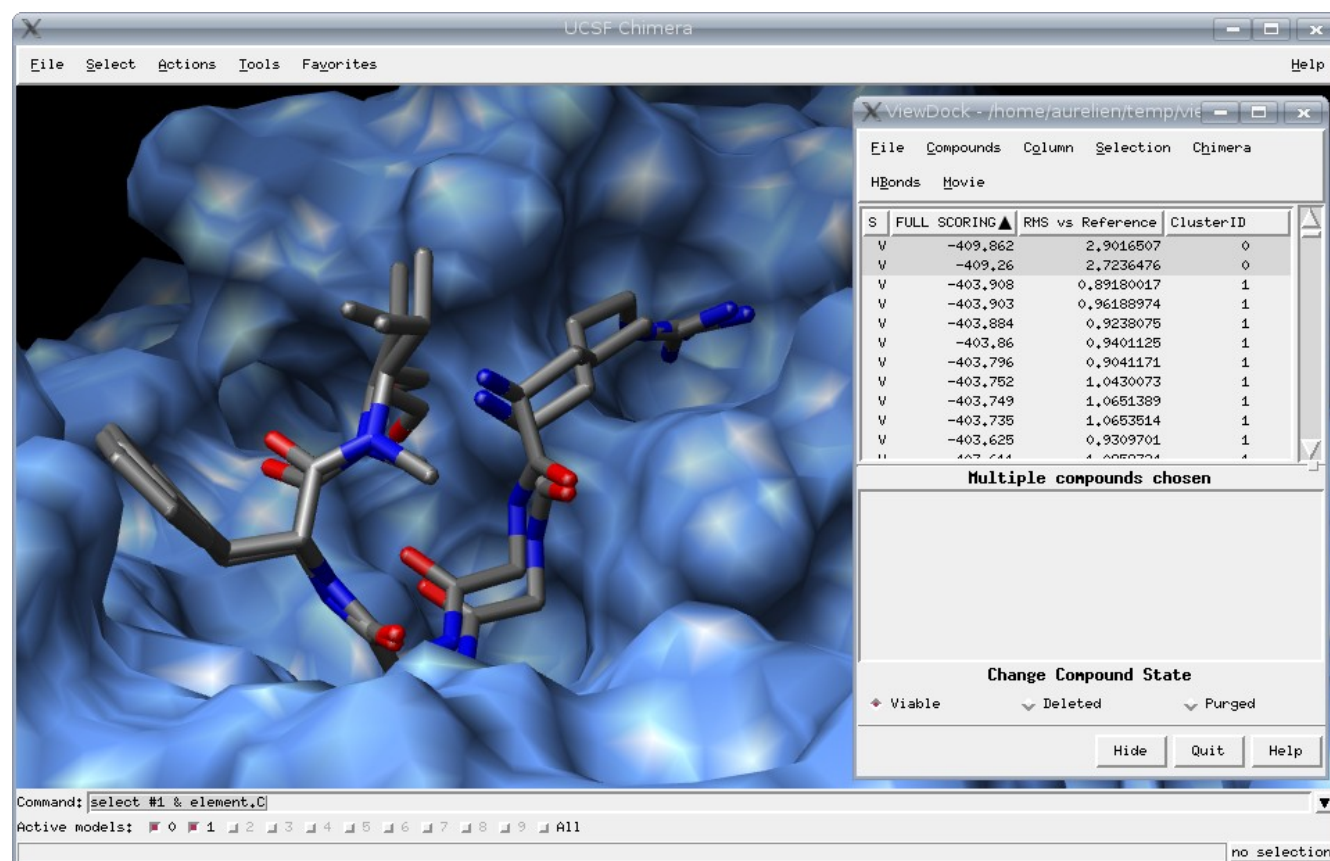


Figure 10 : Prédictions réalisées par EADock et visualisées avec UCSF Chimera et son greffon ViewDock.

Les prédictions de modes d'interactions d'EADock sont créées sous forme de fichiers CRD au format CHARMM, avec un en-tête contenant le détail des termes énergétiques, le rang dans la population, le cluster d'appartenance, et le RMSD aux coordonnées de référence du ligand. Ces fichiers CRD peuvent facilement être convertis au format DOCK 4.0 afin d'être visualisés par le greffon ViewDock de UCSF Chimera (Figure 10). Une telle

interface permet de rendre très accessible l'interprétation des prédictions.

5.2 Complexes utilisés pour la validation d'EADock

Afin de ne pas introduire de biais [30], la validation d'EADock s'est faite sur des complexes utilisés par une étude précédente [22]. Ils sont présentés dans la Table III.

Accessibilité du site de fixation	Complexe	q	DDL	Hb A.	Hb D.	Masse	% B. Sur.
Accessible	<i>Anhydrase carbonique</i>						
	1cil	-1	3	6	2	323.4	85.1
	1okl	0	2	4	1	249.3	87.7
	1cnx	0	10	6	3	331.4	74.2
	<i>Neuraminidase</i>						
	1nsc	-1	4	9	6	308.3	92.0
	1nsd	-1	4	8	5	290.3	92.6
	1nnb	-1	4	8	5	290.3	89.7
	<i>Ribonucléase</i>						
	1gsp	0	2	9	3	360.3	80.2
	1rhl	-2	3	10	4	361.2	78.1
	1rls	-2	3	10	4	361.2	79.2
	<i>Trypsine</i>						
	3ptb	1	1	0	2	121.2	94.6
	1tng	1	1	0	1	114.2	91.6
	1tnj	1	2	0	1	122.2	92.4
	1tnk	1	3	0	1	136.2	91.0
	1tni	1	4	0	1	150.2	85.6
	1tnl	1	1	0	1	134.2	92.7
1tpp	0	2	3	2	206.2	86.9	
1pph	1	7	3	3	429.6	69.9	
Peu accessible	<i>Carbocypeptidase</i>						
	1cbx	-1	3	4	1	207.2	98.2
	3cpa	0	4	4	3	238.2	97.7
	6cpa	-1	9	8	2	477.4	82.3
	<i>Penicillopepsine</i>						
	1apt	1	17	6	5	501.7	85.9
	1apu	0	15	6	4	485.7	85.0
	<i>Thermolysine</i>						
	3tmn	0	5	3	3	303.4	73.0
	5tn	-1	7	5	3	320.3	79.8
	6tmn	-1	11	8	3	471.5	73.2
	<i>ε-Thrombine</i>						
	1etr	0	7	6	4	504.6	87.9
1ets	1	7	4	4	522.7	88.3	
1ett	1	5	3	3	429.6	88.2	
Enfoui	<i>Cytochrome P-450cam</i>						
	1phf	0	1	1	1	144.2	100.0
	1phg	0	3	3	0	226.3	100.0
	2cpp	0	0	1	0	152.2	100.0
	<i>FABP intestinal</i>						
	1icm	-1	11	2	0	227.4	95.6
	1icn	0	14	2	1	282.5	96.0
	2ifb	-1	13	2	0	255.4	96.9
	<i>L-Arabinose</i>						
1abe	0	0	5	4	150.1	100.0	
1abf	0	0	5	4	164.2	100.0	
5abp	0	1	6	5	180.2	100.0	

Table III : EADock a été testé sur les 37 complexes présentés dans cette table. Pour chacun, elle donne le code PDB, la charge du ligand (q), le nombre d'angles dièdres laissés libres (DDL), le nombre de donneurs et d'accepteurs de liaisons hydrogène (Hb. A. and Hb. D.), la masse du ligand (Masse) et la fraction de la surface du ligand occultée lors de la complexation (% B. Sur.)

Les ligands impliqués dans ces complexes sont très divers par rapport à leurs charges, leurs poids moléculaires, le nombre de d'angles dièdres qu'ils contiennent, leurs nombres de donneurs et d'accepteurs de liaison hydrogène, et leurs ClogP.

Les sites de fixations sur le récepteur peuvent être classés suivant leurs accessibilités. Si plus de 95 % de la surface de tous les ligands d'un même récepteur est occultée lors de la complexation, le site de fixation correspondant est « enfoui ». C'est le cas pour le cytochrome P450, la L-arabinose et le FABP intestinal. En fonction de leurs formes, les autres sites de fixations peuvent être classés en « très accessibles » (trypsine, neuraminidase, ribonucléase, anhydrase carbonique), ou « peu accessibles » (ϵ -thrombine, thermolysine, pénicillopepsine, carbocypeptidase).

La protonation des fonctions chimiques titrables a été définie à pH neutre. La protonation des histidines s'est faite en fonction de leurs environnements locaux. Les protéines et les ions ont été modélisés par le champ de force CHARMM22 [96]. Les hydrogènes manquants de la structure cristalline ont été ajoutés par la fonction HBUILD [19] implémentée dans CHARMM. Les paramètres CHARMM22 manquants pour le ligand ont été dérivés du Merck Molecular Force Field (MMFF) [61] [60] [59] [58] [57], en prenant le terme modélisant les angles dièdres tel quel, mais uniquement la partie quadratique modélisant la longueur des liaisons et les angles entre celles-ci. Les charges partielles et les paramètres de van der Waals des atomes du ligand ont été repris de MMFF. Tous les hydrogènes des ligands ont été modélisés.

Avant le docking, les structures cristallines ont été minimisées par 100 pas de SD avec le modèle de solvatation implicite GB-MV2, sans troncature des interactions non liées. Cette courte minimisation permet de résoudre les petits problèmes des structures de départ sans en altérer les conformations, ainsi que de placer les atomes d'hydrogène correctement. Le RMSD entre les conformations avant et après minimisation, calculés en prenant en compte tous les atomes lourds, a toujours été inférieur à 0.15 Å. Le ligand a ensuite été retiré avant le docking.

5.3 Validation de l'algorithme

EADock a été testé en conditions réelles sur les 37 complexes décrits ci-dessus. Les complexes de la génération initiale ont été générés entre 3 Å et 10 Å de RMSD à la structure cristalline pour les sites de fixation accessibles (les complexes trop proches de la structure cristalline ont été explicitement retirés de cette génération afin de ne pas introduire de biais favorables dans nos résultats, voir ci-dessous). La RI a été définie par une sphère de 15 Å de rayon centrée sur le centre de masse du ligand dans la structure cristalline. Pour les sites de fixations occultés, les complexes de la génération initiale ont été générés entre 3 Å et 5 Å de RMSD, et la sphère définissant la RI s'est limitée à un rayon de 5 Å afin d'éviter l'échantillonnage de l'extérieur du site de fixation. L'ensemble des prédictions est détaillé en Table IV.

Dans 68 % des cas, les éléments du cluster le plus favorable d'après EADock sont à moins de 2 Å de RMSD de la structure cristalline. Ce taux de succès monte à 78 % en considérant les cinq clusters les plus favorables, et même 92 % lorsque l'ensemble des clusters présents dans la dernière population de l'évolution est pris en compte. La comparaison avec les logiciels évalués dans la référence [22] ne peut se faire que sur les 11 complexes correspondant à des docking natifs. Malgré l'exclusion explicite des complexes proches de la structure cristalline dans la génération initiale, le RMSD moyen calculé entre les modes d'interactions prédits par EADock et les structures cristallines est seulement de 0.75 Å. Cette valeur est significativement meilleure que celles reportées pour ICM (1.04 Å), AutoDock (2.46 Å), GOLD (3.31 Å), FlexX (3.85 Å) et DOCK (3.87 Å).

Initialisation	Accessibilité du site de fixation	Codes PDB des complexes	%Rec.Acc.Surf dans la RI	Premier cluster acceptable			Explication possible	
				Rang	RMSD moyen (Å)	Δ FullFitness (kcal/mol)		
3-10 Å RMSD entre les seeds et le mode d'interaction cristallin	Accessible	<i>Carbonic anhydrase</i>					Description imparfaite de l'interaction entre le ligand et l'atome de zinc	
		1cil	10.0%	49	1.08	26.78		
		1okl	9.7%	58	1.57	34.53		
		1cnx	12.3%	46	1.65	55.28		
		<i>Neuraminidase</i>						
		1nsc	8.1%	1	0.52	-	-	
		1nsd	8.1%	1	0.69	-	-	
		1nnb	8.8%	1	0.74	-	-	
		<i>Ribonuclease</i>						
		1gsp	31.1%	1	0.85	-	-	
	1rhl	32.0%	1	1.12	-	-		
	1rls	32.4%	1	0.98	-	-		
	<i>Trypsin</i>							
	3ptb	15.9%	1	0.49	-	-		
	1tnng	16.1%	1	0.21	-	-		
	1tnj	17.5%	1	0.86	-	-		
	1tnk	17.2%	1	1.19	-	-		
	1tni	17.1%	2	1.98	2.96	Contact cristallin		
	1tnl	17.0%	1	0.99	-	-		
	1tpp	15.0%	1	0.35	-	-		
1pph	15.4%	1	0.49	-	-			
Peu accessible	<i>Carbocypeptidase</i>							
	1cbx	9.7%	1	0.58	-	-		
	3cpa	9.6%	1	0.85	-	-		
	6cpa	10.8%	2	0.98	0.88	Contact cristallin		
	<i>Penicillopepsin</i>							
	1apt	13.3%	Erreur d'échantillonnage			-		
	1apu	12.9%	6	0.68	12.03	-		
	<i>Thermolysin</i>							
	3tmn	11.5%	1	0.57	-	-		
	5tln	11.1%	7	1.86	28.98	Contact cristallin		
6tmn	10.1%	Erreur de la fonction objective			Intéactions via molécules d'eau			
<i>ϵ-Thrombin</i>								
1etr	12.6%	Erreur de la fonction objective			Contact cristallin			
1ets	12.0%	1	1.16	-	-			
1ett	12.4%	1	0.79	-	-			
3-5 Å RMSD entre les seeds et le mode d'interaction cristallin	Enfoui	<i>Cytochrome P-450cam</i>					Liaison covalente entre le ligand et le récepteur	
		1phf	0.3%	2	1.83	1.62		
		1phg	0.2%	3	1.65	0.97		
		2cpp	0.2%	1	0.19	-	-	
		<i>Intestinal FABP</i>						
		1icm	1.4%	1	0.66	-	-	
		1icn	1.0%	1	1.87	-	-	
		2ifb	0.8%	1	0.76	-	-	
		<i>L-Arabinose</i>						
		1abe	0.3%	1	0.18	-	-	
1abf	0.3%	1	0.68	-	-			
5abp	0.3%	1	0.64	-	-			

Table IV : Résumé des prédictions pour chaque complexe de test, présentant pour chacun la fraction de la surface du récepteur incluse dans la RI, le rang du premier cluster acceptable, et le RMSD moyen entre ses membres et la structure cristalline. Si le premier cluster acceptable n'est pas classé premier, la différence de FullFitness avec ce dernier est donnée dans l'avant dernière colonne, ainsi qu'une explication possible.

La seule erreur imputable à l'échantillonnage à été observée pour pénicillopepsine/1apt, dont le site de fixation est peu accessible. Cependant, lorsque le complexe correspondant à la structure cristalline est généré, il est identifié avec succès par les deux fonctions objectives (données non fournies).

Concernant le cas de la trypsine/1tni, le cluster avec l'énergie la plus favorable est rejeté à cause d'un RMSD à la structure cristalline de 2.1 Å, à peine supérieur au seuil définissant un succès de 2 Å. Les différentes conformations que peut prendre le cycle du phényl en contact avec la protéine peuvent expliquer cette variation. Il est également intéressant de noter que les B-factors des atomes correspondants dans le fichier PDB original sont très élevés : 47.1 Å² en moyenne, à comparer aux 17.6 Å² des autres atomes de la structure, ce qui suggère que cette partie du ligand est très flexible. De plus, un contact cristallin existe entre le cycle du phényl et un complexe adjacent de la maille cristalline.

Des contacts cristallins similaires peuvent également être observés pour les structures de la ε-thrombine/1etr, de la carbocypeptidase/6cpa et de la thermolysine/5tln. Dans ce dernier cas, il existe même une interaction ionique entre le ligand et un complexe adjacent dans la maille cristalline.

Les ligands du cytochrome/1phf et du cytochrome/1phg sont liés de façon covalente au fer du hème du récepteur. Une telle interaction ne peut pas être décrite par le champ de force de nos fonctions objectives. Ces interactions étaient discutées dans l'article original [113], mais absentes de l'étude faite par Bursulaya et al [22].

Aucun programme évalué dans cette étude [22], à l'exception d'ICM, n'arrive à prédire avec succès les modes d'interactions des complexes de la famille de l'anhydrase carbonique. Ces échecs pourraient être dus à une modélisation trop imparfaite des interactions entre le ligand et l'ion zinc de la poche de fixation [43]. Il faut également noter que la RI utilisée avec ICM n'est pas documentée, et qu'il n'est donc pas possible d'exclure qu'elle soit très réduite et limite ainsi l'espace des modes d'interactions accessibles seulement à ceux qui sont acceptables.

Les 37 complexes utilisés pour notre validation ne sont pas assez nombreux pour évaluer l'impact de l'utilisation du modèle de solvataion implicite GB-MV2 (*FullFitness*) par rapport à $\epsilon=1$ (*SimpleFitness*). Toutefois, pour certains complexes, une influence très encourageante a pu être mise en évidence : 3tmn, 1ets, 1etr et 3cpa. Par exemple, trois molécules d'eau sont impliquées dans un réseau de liaisons hydrogène entre le ligand et la protéine du complexe 3tmn. Deux d'entre elles sont nécessaires pour que la *SimpleFitness* (donc *in vacuo*), évalue le cluster correspondant à l'interaction observée dans le cristal plus favorablement que tous les autres. Sans elles, ce cluster n'est qu'un minimum local de la *SimpleFitness*. Il est néanmoins assez marqué pour permettre l'émergence d'un cluster lors du processus évolutif, qui se retrouve classé très favorablement lorsqu'il est évalué par la *FullFitness*. Il est également intéressant de mentionner que le docking échoue lorsque la liste des conformations interdites n'est pas utilisée, ce qui souligne l'importance de ne pas laisser l'échantillonnage se focaliser sur des régions dont la *FullFitness* n'est pas suffisamment favorable. Dans l'étude précédente, seul ICM [22] a été capable d'identifier le bon mode d'interaction, mais sa RI n'ayant pas été décrite, son échantillonnage peut avoir été très limité autour de la zone d'interaction.

5.4 Discussion

5.4.1 Évaluation des algorithmes de docking

Deux questions sont posées lors de l'évaluation d'un algorithme de docking : 1) sa capacité à échantillonner la bonne solution, et 2) son aptitude à voir cette solution reconnue par sa fonction objective. Ces deux questions sont liées, car des limites de fonctions objectives peuvent ne pas être identifiées dès lors que les bonnes solutions ne sont pas même générées. Pire : la structure cristalline est probablement un minimum au moins raisonnable des fonctions objectives. De ce fait, si, dans un exercice d'évaluation, les algorithmes sont initialisés avec des complexes trop proches de (voire incluant) cette structure cristalline, ou si la RI est trop limitée, un algorithme dont l'heuristique d'échantillonnage est défaillante pourra retrouver l'interaction même si sa fonction objective est également défaillante, puisque des modes d'interactions distants et

physiquement non pertinents, mais qui peuvent correspondre au minimum global de cette fonction, ne seront même pas échantillonnés. Dans ce cas, la recherche convergera simplement vers le minimum local le plus proche, très probablement proche de la structure cristalline. La qualité de l'algorithme n'est dans ce cas pas évaluée correctement, car les prédictions correctes peuvent être attribuées soit à une bonne fonction objective, soit à une déficience de l'échantillonnage [30]. De plus, le peuplement de la génération initiale par des complexes extrêmement proches de la structure cristalline implique que cette dernière soit identifiée *a priori*, ce qui n'est pas le cas pour une prédiction réelle [30].

5.4.2 Notre approche

Afin d'éviter ces problèmes lors de l'évaluation d'EADock, nous avons forcé une distance minimale de 3 Å RMSD entre les complexes de la génération initiale et la structure cristalline. De cette façon, un succès ne peut que refléter la performance de l'heuristique d'échantillonnage qui doit dans ce cas précis identifier un minimum énergétique plus favorable que ceux correspondant à ces complexes initiaux, par ailleurs déjà bien optimisés.

Des conditions si difficiles rendent impossible la comparaison directe entre EADock et les autres programmes évalués en [22] sur les mêmes complexes, et ce pour deux raisons. Tout d'abord, dans cette dernière, le mode d'interaction trouvé dans le cristal pouvait être présent dès la génération initiale. Ensuite, la RI utilisée pour l'évaluation d'EADock correspond à 15 % de la surface des récepteurs, en moyenne, alors que dans l'étude précédente, les RI incluaient seulement 10 % (AutoDock), 4.6 % (Gold) et 1.7 % (DOCK et FlexX) de cette dernière. Si de telles RI ne permettent pas de comparer différents programmes (voir ci-dessus), elles peuvent néanmoins se révéler intéressantes pour les cas où la zone d'interaction est précisément définie *a priori*.

Malgré des conditions initiales défavorables, EADock converge en seulement 50000 évaluations par la *SimpleFitness* et 15000 par la *FullFitness* en moyenne, suivant le chemin emprunté par l'évolution. Cette performance met en évidence l'efficacité de notre

heuristique d'échantillonnage, par comparaison avec les 2500000 modes d'interactions nécessaires à AutoDock [22]. Une autre étude récente utilisant d'autres complexes pour l'étape de validation reportée, suivant les logiciels, entre 200000 et 400000 évaluations avant de converger [107]. Ces valeurs sont comparables à d'autres évaluations réalisées [111]. Notre bonne performance est également mise en évidence par le fait qu'EADock est capable de docker correctement tous les ligands si le RMSD à la structure cristalline est utilisé comme fonction objective, et ce même avec une génération initiale très dégénérée (données non fournies). La limite actuelle d'EADock semble bien être les fonctions objectives utilisées, mais il est intéressant de voir que la plupart de ces échecs peuvent s'expliquer par la présence de contacts cristallins, voire par une liaison covalente entre le ligand et le récepteur.

L'estimation de l'énergie libre de liaison nécessite une prédiction aussi précise que possible de l'interaction entre le ligand et le récepteur. En considérant les prédictions correctes, le RMSD moyen entre les structures cristallines et les modes d'interactions prédits par EADock est seulement de 0.75 Å. Ce gain par rapport aux approches existantes peut certainement jouer un rôle important dans l'optimisation rationnelle de ligand.

Afin de démontrer l'efficacité de notre approche en conditions extrêmes, dans le cadre d'une application réelle, le mode d'interaction entre l'intégrine $\alpha V\beta 3$ et son inhibiteur le Cilengitide (un pentapeptide cyclique) a été prédit par EADock. Les paramètres précédents ont été utilisés, excepté la RI, étendue à une sphère de 25 Å de rayon centrée sur le site d'interaction (soit 65000 Å³). Les conformations initiales ont été générées loin de la zone d'interaction, entre 15 Å et 25 Å de RMSD à la structure cristalline. Malgré ces conditions difficiles, EADock est capable de prédire correctement l'interaction, avec un RMSD de 1.17 Å seulement (Figure 11).

Cette performance met en évidence l'efficacité de notre heuristique de recherche et de notre combinaison de fonctions objectives dans le cadre d'une application réelle, et ouvre la voie de l'optimisation rationnelle de molécules actives dérivées du Cilengitide, dont l'intérêt clinique est indiscutable [21].

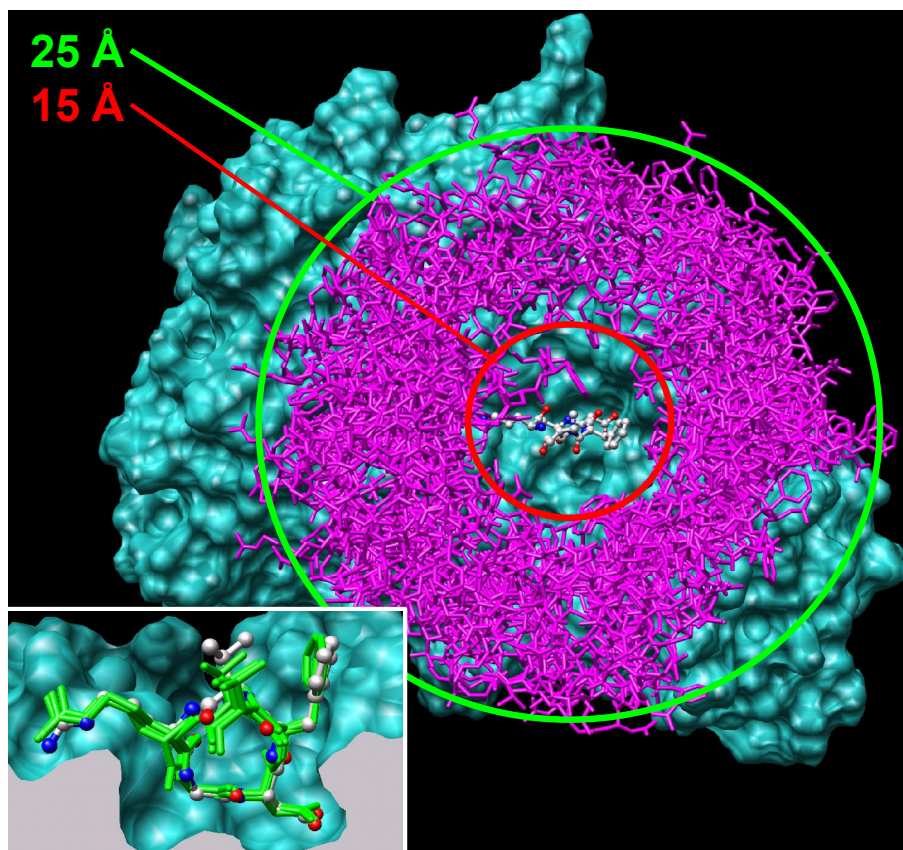


Figure 11 : Docking du pentapeptide cyclique $c(-RGDf[NMe]V-)$ sur l'intégrine $\alpha V\beta 3$. Les positions initiales des ligands (bâtonnets rouges) ont été générées entre 15 Å et 25 Å de RMSD à la structure cristalline (cercles rouges et verts). Les paramètres de docking standard ont été utilisés, et la RI a été définie par une sphère de 25 Å de rayon dont le volume s'élève à 65449 Å³. Le mode d'interaction du ligand dans la structure cristalline est représenté dans l'insert (boule/bâton), et correspond au cluster le plus favorable prédit par EADock (bâtonnets verts). Le RMSD moyen des trois éléments les plus favorables de ce cluster à la structure cristalline est de 1.17 Å.

Nos résultats ont été obtenus par la combinaison d'une optimisation multi-objective prenant en compte l'énergie libre de solvation du complexe, avec une heuristique d'échantillonnage performante, capable de converger en évaluant peu de complexes. La prise en compte de l'énergie libre de solvation par une fonction objective basée sur un champ de force a trois inconvénients majeurs : elle est très sensible à de petites variations des coordonnées atomiques, elle ne fournit pas de gradient d'énergie utilisable, et elle est très coûteuse en temps de calcul. Notre algorithme de docking propose une solution pour

chacun de ces problèmes. D'abord, les impacts des variations de coordonnées atomiques sont lissés par un moyennage sur plusieurs modes d'interactions proches, définis par un même cluster. Ensuite, un gradient d'énergie est introduit dans le processus évolutionnaire par l'intermédiaire de la *SimpleFitness*, qui permet de filtrer les modes d'interactions raisonnables. Finalement, l'évaluation d'un complexe par la *SimpleFitness* est assez rapide, et fournit une approximation acceptable de la *FullFitness*. Ces trois solutions permettent à EADock d'utiliser une fonction objective incluant l'énergie libre de solvation, calculée par le logiciel CHARMM. Son universalité (elle n'est pas limitée aux protéines) et sa description moyenne de l'effet du solvant sont deux avantages sur les autres fonctions objectives. En effet, nous pouvons nous attendre à une bonne transférabilité de nos résultats à d'autres familles de récepteurs et de ligands, comme à l'ADN par exemple, tout autant qu'à d'autres stratégies, comme les approches basées sur le docking de fragments de molécules (voir Introduction et Applications).

6 Applications

6.1 Aperçu

La porte de la conception de PA *in silico* s'ouvre avec la possibilité d'identifier le mode d'interaction le plus favorable d'une petite molécule avec une cible d'intérêt thérapeutique. En parallèle des développements méthodologiques présentés dans le chapitre précédent, plusieurs collaborations ont été menées avec des chercheurs expérimentaux dans des domaines variés. Deux d'entre elles sont présentées dans ce chapitre et visent justement la conception et l'optimisation de nouvelles molécules actives, avec pour cible le Peroxisome Proliferator-Activated Receptor α (PPAR α) et l'intégrine $\alpha 5\beta 1$. Le premier est la cible des fibrates hypolipidémiants et la seconde est intéressante en oncologie : un de ses inhibiteurs est actuellement en phase clinique [64]. Une approche basée sur le docking de fragments (voir Introduction) a été développée avec Vincent Zoete. Elle repose sur EADock, combiné avec plusieurs autres composants logiciels développés pour l'occasion. Elle est brièvement présentée ci-dessous.

L'utilisation d'EADock dans le cadre de collaborations, dès les premières phases de son développement, a eu des conséquences très positives en termes de fonctionnalités, de stabilité et d'ergonomie. Sur le plan technique, l'implémentation elle-même a été réorganisée plusieurs fois en fonction des besoins toujours nouveaux des chercheurs expérimentaux, jusqu'à atteindre une versatilité suffisante aujourd'hui.

Jusqu'aux prochains défis.

6.2 Matériel et méthode

6.2.1 Aperçu

Une approche basée sur le docking de fragments a été utilisée afin de mettre au point de nouveaux ligands peptidiques de l'intégrine $\alpha 5\beta 1$ et du récepteur nucléaire PPAR α humain. Les grandes lignes de cette approche sont présentées en Figure 12, partant de la

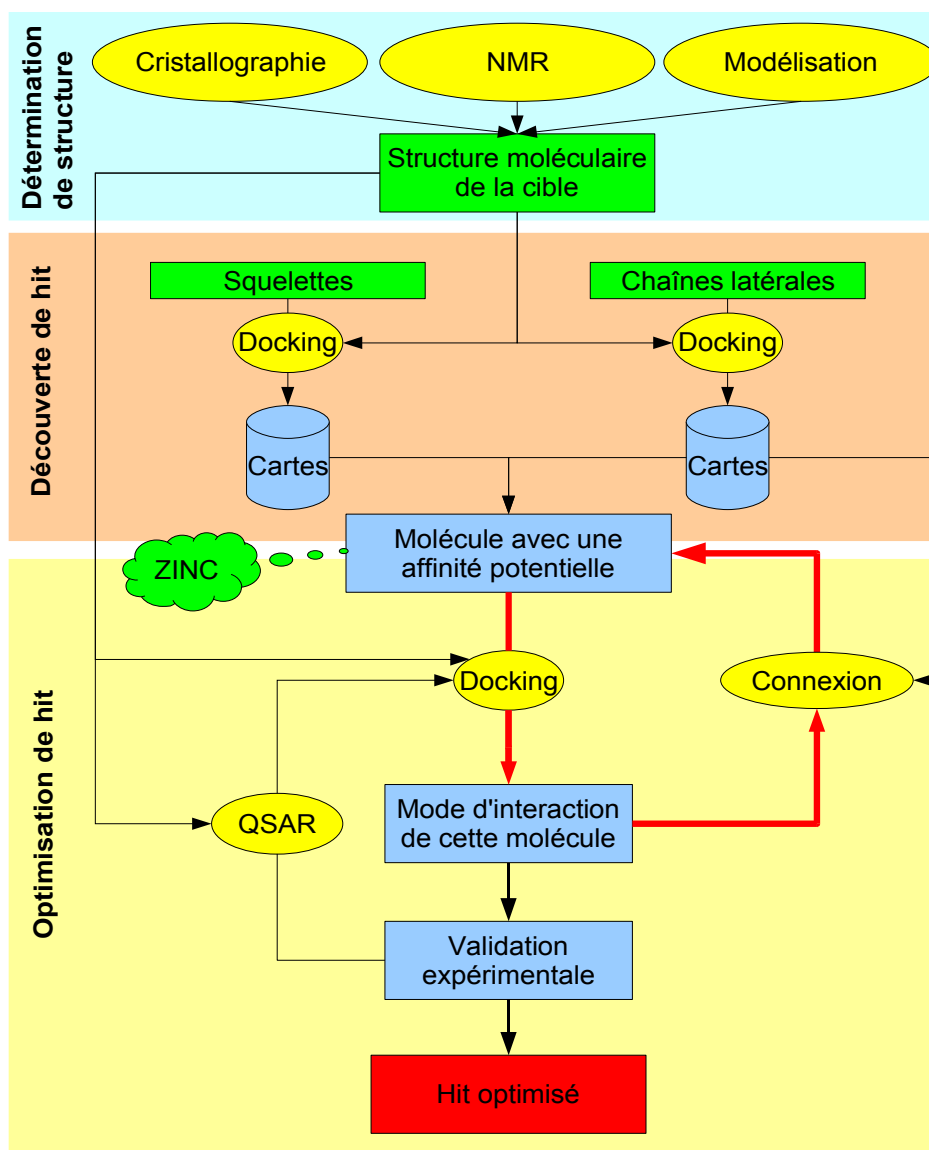


Figure 12 : Aperçu de la méthode générique de conception/optimisation de molécules actives. Les données sont colorées en bleu, sauf les données de départ (vert) et celles d'arrivée (rouge). Les méthodes sont colorées en jaune. Le cycle d'optimisation est matérialisé par les flèches rouges. Voir détails dans le texte.

structure moléculaire d'une cible biologique, et aboutissant à une molécule optimisée.

La première étape de cette approche est de réaliser avec EADock une carte des positions et orientations de chaque fragment de molécules, sur la surface de la protéine cible (Figure 12). Ces fragments sont classés en « squelette » et en « chaînes latérales » [11]

[10]. Dans notre approche, les « squelettes » peuvent être des fragments rigides d'inhibiteurs connus, ou des hits virtuels (molécules conçues *ab initio* et présentant un mode d'interaction calculé par EADock reproductible et compatible avec la structure de la poche de fixation à la surface de la protéine). Les « chaînes latérales » sont de petits fragments chimiques plus flexibles qui peuvent être ensuite reliés aux « squelettes ».

Ici appliquée à la conception de peptides, cette approche utilise les chaînes des acides aminés naturels comme fragments, exceptées la glycine (qui n'a pas de chaîne latérale) et la proline. Une fois que leurs positions les plus favorables sont identifiées par EADock, les possibilités de connexion entre elles sont évaluées suivant des critères chimiques et géométriques (Figure 13).

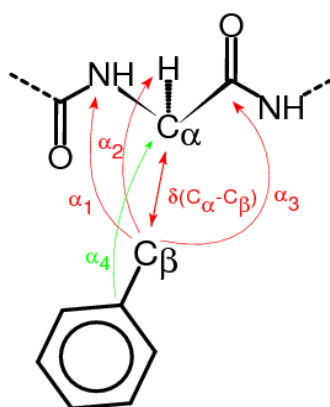


Figure 13 : Distance et angles mesurés pour évaluer la possibilité de connexion entre un backbone peptidique (en haut) et une chaîne latérale potentielle (en bas).

Dans le cas de la conception de peptide, la distance entre un C_{α} du backbone et le C_{β} d'une chaîne latérale potentielle doit être de l'ordre de 1.54 Å, et les différents angles $C_{\alpha}-C_{\beta}-C_{\gamma}$, $N-C_{\alpha}-C_{\beta}$, $H-C_{\alpha}-C_{\beta}$ et $C-C_{\alpha}-C_{\beta}$ doivent être proches de 109-110°. En fonction des mutations virtuellement possibles, une liste de ligands peptidiques potentiels est construite de façon combinatoire. Le mode d'interaction de chacun de ces ligands potentiels est ensuite prédit *in silico* par EADock. Les ligands les plus intéressants en termes d'interactions avec la protéine cible sont retenus comme points de départ pour l'itération suivante de l'optimisation. Ce redocking des ligands les plus intéressants

permet un ajustement de leurs backbones de façon à mieux choisir les mutations pouvant être envisagées par la suite. Cette procédure peut être répétée plusieurs fois (voir Figure 12, « cycle d'optimisation »), jusqu'à obtenir une liste de molécules à tester expérimentalement. Les affinités qui peuvent alors être mesurées sont susceptibles de permettre la mise au point d'une relation structure activité (QSAR). Cette dernière peut à son tour être utilisée pour prédire l'affinité des nouveaux ligands potentiels lors des itérations suivantes.

6.2.2 EADock

EADock est utilisé lors des dockings *in silico* afin d'identifier le mode d'interaction le plus favorable d'un ligand [55], ainsi que les cartes des interactions les plus favorables pour les squelettes et les chaînes latérales, avec les paramètres suivants : la taille maximale des clusters est limitée à 5 éléments, et leurs rayons à 1.5 Å. La taille de la population est portée à 500 complexes, et la liste de conformations interdites n'est pas utilisée afin de conserver les modes d'interactions les plus favorables rencontrés durant le processus évolutionnaire.

6.2.3 Champ de force

L'ensemble des calculs de mécanique moléculaire est fait par le programme CHARMM (version c31b1) avec les paramètres de champ de force CHARMM22. L'état de protonation standard à pH neutre est attribué à chaque fonction titrable. La protonation des histidines est faite par inspection visuelle de leurs environnements locaux.

6.3 L'intégrine

Ce travail a été réalisé par Vincent Zoete et Aurélien Grosdidier, en collaboration avec Curzio Rüegg et Gian-Carlo Alghisi du Laboratoire de recherche sur l'angiogénèse des tumeurs et le mélanome, du Centre Hospitalier Universitaire Vaudois. Le manuscrit correspondant est en préparation.

6.3.1 Contexte biologique

La survie d'une cellule dépend de nombreux facteurs, parmi lesquels l'attachement aux composants de la matrice extra-cellulaire (ECM), par l'intermédiaire de molécules d'adhésions cellulaires comme les intégrines. Les intégrines sont des molécules transmembranaires $\alpha\beta$ -hétérodimériques. 18 sous unités α et 8 β ont déjà été identifiées, et peuvent former 24 dimères connus, dont les expressions dépendent du type cellulaire. Elles sont le sujet de plusieurs revues récentes [70] [118] [132]. Chaque sous-unité de l'intégrine a un grand domaine extra-cellulaire, un court domaine transmembranaire, et un petit domaine intracellulaire. Elles sont les principaux récepteurs des protéines de l'ECM comme le collagène, la fibronectine et la laminine. Les interactions cellule-ECM qu'elles permettent sont essentielles au développement embryonnaire, à la prolifération, la survie, l'adhésion, la différenciation et la migration des cellules [132] [17] [117] [116] [136] (voir Figure 14).

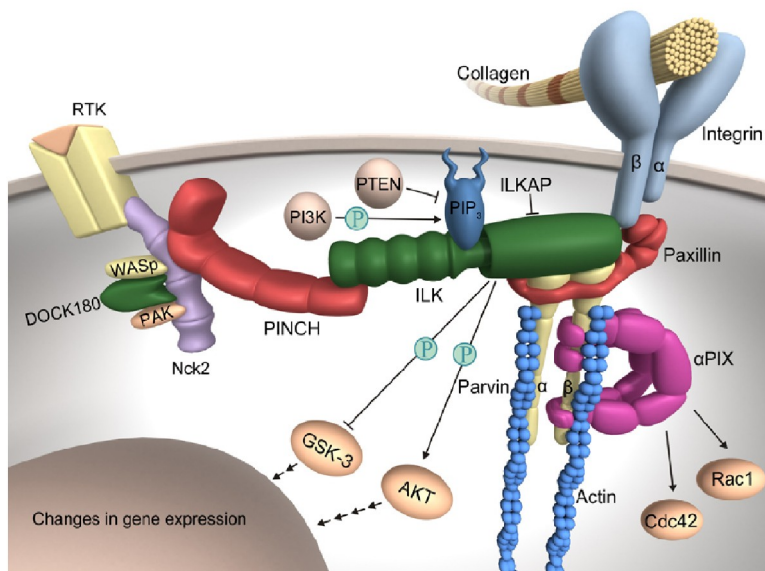


Figure 14 : Aperçu des voies métaboliques reliées aux intégrines [41].

Les cellules tumorales ont développé plusieurs mécanismes afin d'échapper à l'inhibition de contact médiée par des molécules telles que les intégrines. La surexpression ou la perte des intégrines contribue à plusieurs types de tumeurs, et les changements de niveau d'expression des intégrines influencent le caractère invasif et métastatique des tumeurs

[32] [42] [106] [112] [134].

De par leur rôle central dans plusieurs mécanismes de survie cellulaire, les intégrines sont des cibles particulièrement intéressantes dans le domaine de l'oncologie, et plusieurs ligands ont déjà été conçus [64]. Le Cilengitide est l'un des plus prometteurs. Ce pentapeptide cyclique c(-RGDf[NMe]V-) contient un motif RGD également présent dans la fibronectine, connu pour son affinité pour la plupart des intégrines. Il a déjà démontré un caractère inhibiteur de $\alpha V\beta 3$ et de $\alpha 5\beta 1$ (à un niveau moindre cependant), qui limite le développement de la tumeur et l'angiogénèse. Suivant les indications, il est actuellement en essais cliniques de phase I ou II. Notre approche basée sur le positionnement et la reconnexion de fragments a été utilisée pour proposer de nouveaux inhibiteurs peptidiques, en partant du motif RGD.

6.3.2 Validation d'EADock

La capacité d'EADock à positionner correctement des peptides à la surface d'une intégrine a été évaluée en essayant de reproduire le mode d'interaction entre le Cilengitide et l'intégrine $\alpha V\beta 3$, complexe pour lequel une structure cristalline est disponible (code 1L5G dans la PDB). Le mode d'interaction expérimental du Cilengitide a été reproduit avec un RMSD de 1.17 Å, en dépit d'une population initiale distante de ce dernier de 15 Å à 25 Å (voir chapitre précédent). La capacité d'EADock à positionner correctement des fragments (et donc à réaliser des cartes d'interactions favorables pertinentes) a été évaluée sur l'intégrine $\alpha V\beta 3$. En particulier, les cartes de l'aspartate et de l'arginine ont été réalisées et comparées à la position expérimentale de l'aspartate et de l'arginine du Cilengitide. Le choix de ces deux chaînes latérales a été dicté par leurs appartenances au motif RGD du Cilengitide, également présent dans presque tous les ligands naturels des intégrines, et particulièrement importants pour leurs interactions avec ces dernières. La Figure 15 montre le motif RGD du Cilengitide lié à l'intégrine $\alpha V\beta 3$, ainsi que la carte réalisée pour la chaîne latérale de l'aspartate (Figure 15A). La position la plus favorable d'après EADock est presque parfaitement superposée à la position de l'aspartate inclus dans le motif RGD. De façon similaire, les positions les plus favorables pour la chaîne latérale de l'arginine font des interactions ioniques avec les aspartates 150

et 218 de la chaîne α , comme ceux de l'arginine du Cilengitide. La sixième position la plus

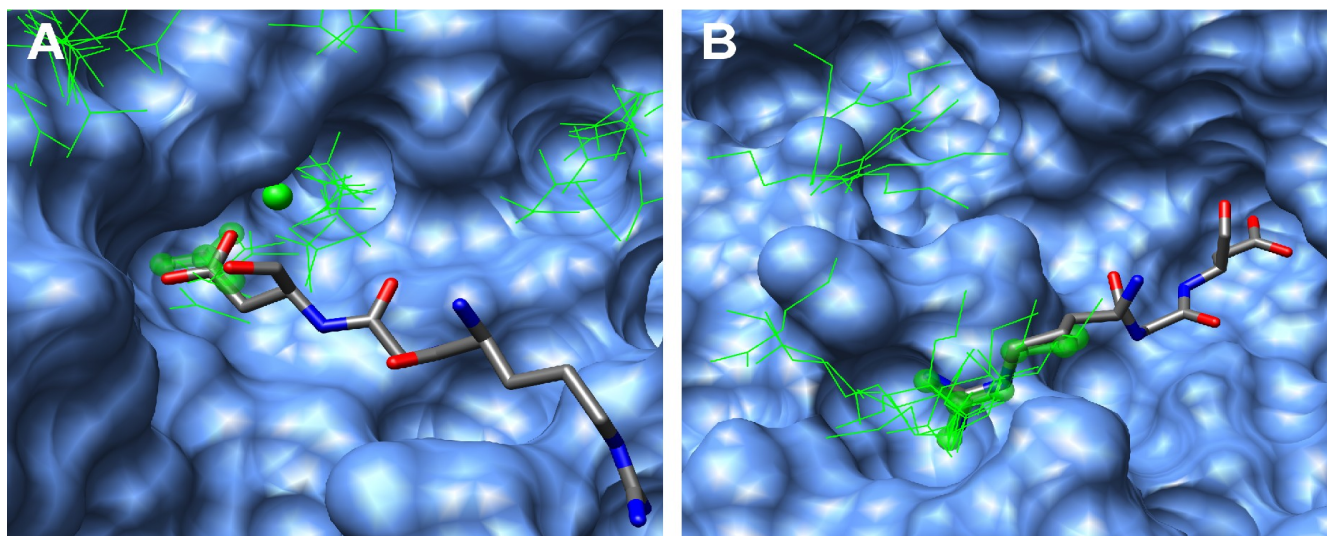


Figure 15 : (A) Le motif RGD du Cilengitide est représenté en bâtonnets et coloré suivant le type d'atome, sur la surface de l'intégrine $\alpha 5\beta 3$. Le reste du ligand n'est pas représenté pour plus de clarté. Les positions favorables pour le fragment acétate sont représentées en fines lignes vertes, excepté le plus favorable en représentation boule-bâton vert transparent. (B) même représentation que (A) concernant la chaîne latérale de l'arginine. Le sixième mode d'interaction le plus favorable adopte une représentation boule-bâton vert transparent

favorable est même superposée à l'arginine de ce dernier (voir Figure 15B). Ces résultats illustrent la capacité d'EADock à positionner des ligands complets aussi bien que des fragments de ceux-ci à la surface de l'intégrine.

6.3.3 Conception d'inhibiteurs de l'intégrine $\alpha 5\beta 1$.

Aucune structure expérimentale de l'intégrine $\alpha 5\beta 1$ n'étant disponible, un modèle par homologie a été réalisé grâce à MODELLER [44], en se basant sur la structure de l'intégrine $\alpha V\beta 3$ liée au Cilengitide (1L5G dans la PDB, 52 % d'identité de séquence). Le modèle obtenu a été minimisé par 30 pas de SD et 30 pas d'ABNR, puis à nouveau par 100 pas de SD dans le modèle de solvation implicite GB-MV2. En nous inspirant de la séquence et de la structure de la fibronectine (1TTG dans la PDB), nous sommes partis du peptide linéaire RGDSP comme hit virtuel afin d'initier le cycle d'optimisation. Ce peptide

a d'abord été positionné sur l'intégrine $\alpha 5\beta 1$ par EADock. La Figure 16 montre le mode d'interaction calculé le plus favorable.

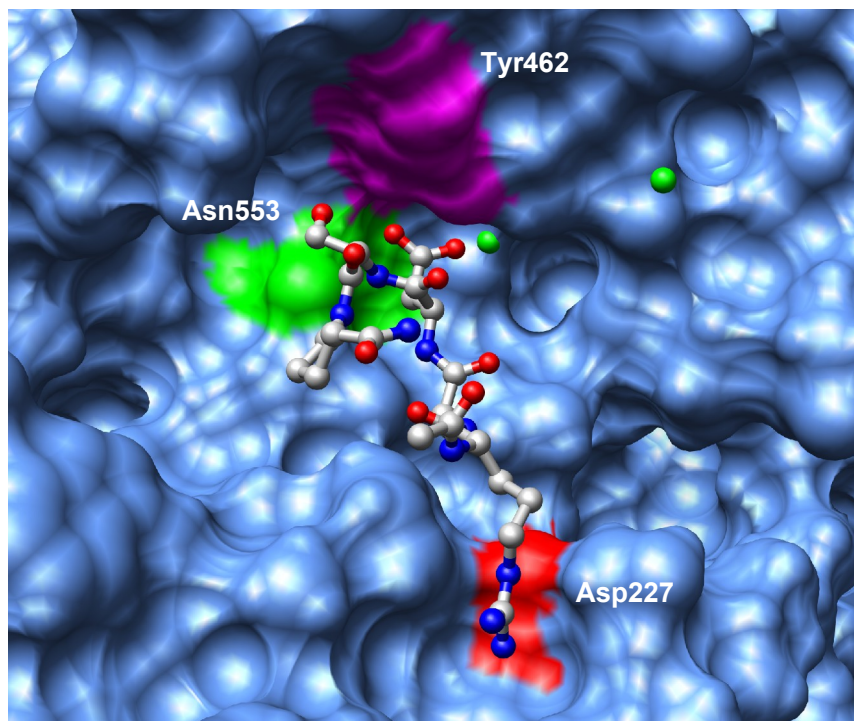


Figure 16 : Interaction de RGDSP avec l'intégrine $\alpha 5\beta 1$

La position du motif RGD est très proche de celle observée dans le mode d'interaction du Cilengitide sur $\alpha V\beta 3$ déterminé expérimentalement. L'arginine fait des interactions ioniques avec la chaîne latérale d'Asp227 sur la chaîne α , l'aspartate avec le backbone de Asn553 et Tyr462 (chaîne β) ainsi qu'avec un ion Mg^{2+} . Des liaisons hydrogène supplémentaires sont également présentes entre l'atome NH du backbone et la fonction hydroxyle de la sérine du peptide d'un côté, et la fonction carbonyle du backbone de l'Asn553 de la chaîne β de l'intégrine de l'autre côté. Comme la N-méthylValine du Cilengitide, la proline ne fait pas d'interactions avec l'intégrine. Le mode d'interaction calculé par EADock étant stable et reproductible, il a été utilisé pour démarrer l'optimisation de séquence.

À partir de la séquence linéaire RGDSP, l'optimisation itérative décrite précédemment a permis de suggérer plusieurs autres pentapeptides afin de les tester expérimentalement.

Leur liste est donnée dans la Table V.

Peptide	IC ₅₀ ($\alpha 5\beta 1$)
Cilengitide, i.e. c(RGDf[Met]V)	5.5 μ M
RGD	650 μ M
RGDSP	45 μ M
RGDLP	175 μ M
RGDFP	17.5 μ M
RGDWP	12 μ M

Table V : Séquence et IC₅₀ des peptides testés expérimentalement sur l'intégrine $\alpha 5\beta 1$.

Par exemple, les positions les plus favorables calculées par EADock pour la chaîne latérale du tryptophane sont visibles dans la Figure 17. La plus favorable de toutes fait plusieurs interactions de van der Waals avec Leu512, Pro515 et Cys516, ainsi qu'une liaison hydrogène avec la chaîne latérale de Ser518 (sur la chaîne β). Sa position lui permet de remplacer le quatrième résidu de notre inhibiteur virtuel potentiel RGDSP. La structure de la séquence RGDWP est donc automatiquement générée et le peptide correspondant est repositionné par EADock. Comme le montre la Figure 17B, le tryptophane de RGDWP, dont le mode d'interaction est proposé par EADock, opte pour une position similaire à celle du fragment correspondant à la chaîne latérale isolée du tryptophane. L'approche utilisée semble donc valide, car un fragment de molécule faisant plusieurs interactions favorables avec la protéine cible et dont le positionnement est compatible avec une connexion à un peptide déjà positionné permet d'obtenir un nouveau peptide dont le mode d'interaction calculé est conservé tant pour la chaîne latérale greffée que pour le reste du peptide.

Comme attendu du fait qu'elle ne fait aucun contact avec l'intégrine, aucune mutation n'a pu être proposée pour la proline. Plusieurs modifications ont été par contre suggérées pour la sérine, telles que des mutations en leucine, phénylalanine ou tryptophane. Les quatre pentapeptides linéaires correspondants ont été synthétisés et testés expérimentalement avec le tripeptide RGD (comme référence). La Table V montre que l'affinité de ce dernier est basse, avec une IC₅₀ de seulement 650 μ M. RGDSP, dont la séquence a été inspirée par la fibronectine, montre une meilleure IC₅₀ de 45 μ M. Parmi les trois peptides dérivés par notre approche *in silico*, deux montrent une IC₅₀ meilleure que

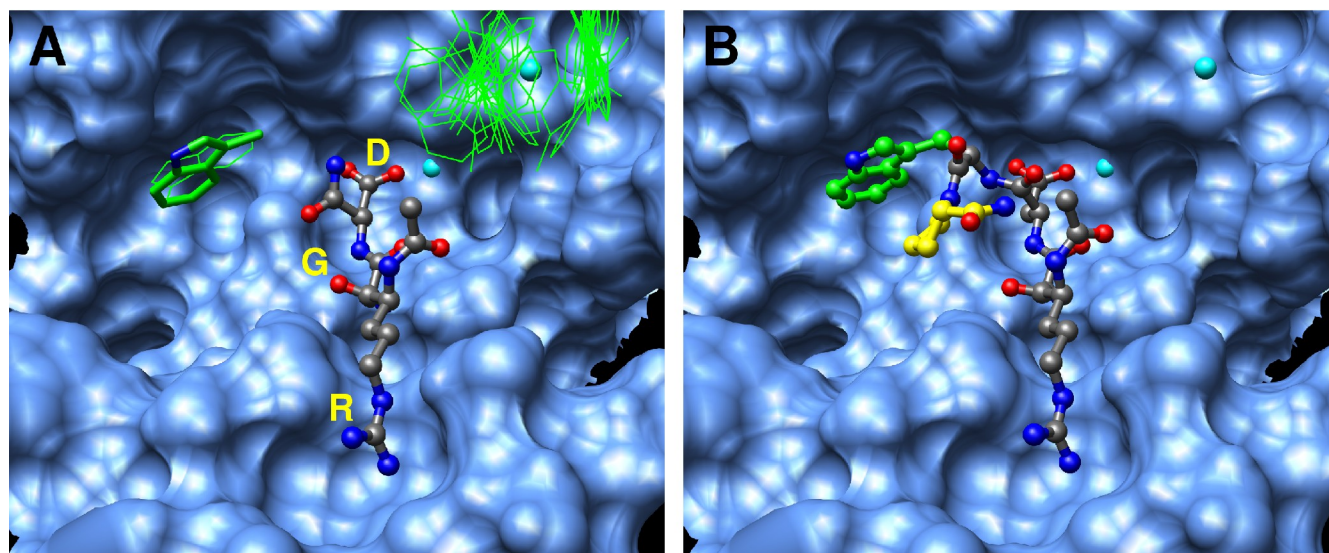


Figure 17 : (A) le peptide RGDSP (boule/bâton) positionné à la surface de l'intégrine $\alpha 5\beta 1$. Les résidus S et P ne sont pas montrés par souci de clarté de la figure. Les positions favorables de la chaîne latérale du tryptophane sont visibles sous forme de fine ligne verte, excepté le plus favorable en représentation bâtonnets verts. (B) Mode d'interaction calculé par EADock pour le peptide linéaire RGDWP.

RGDSP : RGDFP et RGDWP. Ce dernier, le plus actif, se rapproche même de l'IC₅₀ du Cilengitide.

6.3.4 Conclusion

Notre approche basée sur le positionnement de fragments a permis la découverte de plusieurs pentapeptides linéaires montrant une activité significative pour l'intégrine $\alpha 5\beta 1$. Leurs activités mesurées sont supérieures à celles du motif RGD et du point de départ de l'optimisation RGDSP, et pour l'un d'entre eux, elle est même comparable à celle du Cilengitide. Cette étude montre que les approches *in silico* peuvent permettre de découvrir de nouveaux ligands de protéines cibles.

Ces résultats préliminaires ne sont qu'une étape dans la conception d'inhibiteurs de $\alpha 5\beta 1$. Plusieurs itérations de cette optimisation de séquence vont être réalisées afin d'améliorer encore l'affinité de ces trois peptides. De plus, une augmentation de leurs demi-vies peut facilement être envisagée, par exemple par l'introduction d'acides aminés D dans la

séquence, voire par la cyclisation de la molécule. Cette dernière opération pourrait du même coup augmenter significativement l'affinité des peptides, car elle réduirait considérablement leurs espaces conformationnels accessibles, par rapport à leurs contreparties linéaires, par limitation de la pénalité entropique induite par la complexation. Comme le montre la Figure 17B, une telle cyclisation semble possible au regard de la proximité entre les extrémités N- et C-terminales du peptide dans sa conformation bioactive prédite par EADock.

6.4 Conception de ligands peptidiques de PPAR α

Ce travail a été réalisé par Vincent Zoete, Lina Yip-Sondernegger et Aurélien Grosdidier, en collaboration avec les groupes de Liliane Michalik et de Walter Wahli de l'Université de Lausanne - Centre Intégratif de Génomique. Le manuscrit est en préparation.

6.4.1 Contexte biologique

PPAR α est un récepteur nucléaire exprimé dans le foie, le rein, le coeur et les muscles, et participe à l'activation des gènes impliqués dans le maintien de l'homéostasie du métabolisme des acides gras et du métabolisme des lipoprotéines, par recapture et oxydation. PPAR α est la cible des fibrates hypolipémiants.

6.4.2 Validation d'EADock

Tous les calculs ci-dessous ont été réalisés à partir de la structure cristalline du PPAR α humain (hPPAR α) lié à un de ses antagonistes GW409544 (code 1K7L dans la PDB). Ce ligand a bien sûr été retiré de la structure avant tout calcul.

La capacité d'EADock à positionner correctement des molécules dans le site actif des PPARs a été validée précédemment [98]. Sa capacité à positionner des fragments de molécules favorables a également été évaluée dans une étude antérieure (données non fournies).

6.4.3 Conception de ligands peptidiques du hPPAR α .

Le PPAR α ne disposant d'aucun ligand peptidique connu, la première étape a été la conception d'un hit virtuel potentiel. Ce dernier est un peptide qui n'a pas été testé expérimentalement et dont nous ne savons donc pas s'il est actif ou pas. Cependant, son mode d'interaction prédit *in silico* par EADock est bien défini et reproductible. Il peut donc être utilisé comme point de départ de l'optimisation de séquence. L'analyse des ligands organiques synthétiques montre qu'ils ont en commun un certain nombre de caractéristiques : ils possèdent une tête polaire (généralement une fonction carboxylate ou un cycle thiazolidinedione), interagissant avec la partie polaire du site actif (Ser280, Tyr314, His440 et Tyr464), ainsi qu'une longue queue hydrophobe qui fait des interactions de van der Waals avec le reste du site actif, lui-même hydrophobe [137]. Le hit peptidique virtuel potentiel que nous avons choisi sur cette base a la séquence ADAAA. Ce peptide été positionné dans le site actif de hPPAR α par EADock. Le mode d'interaction observé était celui attendu, avec la chaîne latérale de l'aspartate participant à un réseau de 4 liaisons hydrogène, et les alanines remplissant la partie hydrophobe de site actif (voir Figure 18).

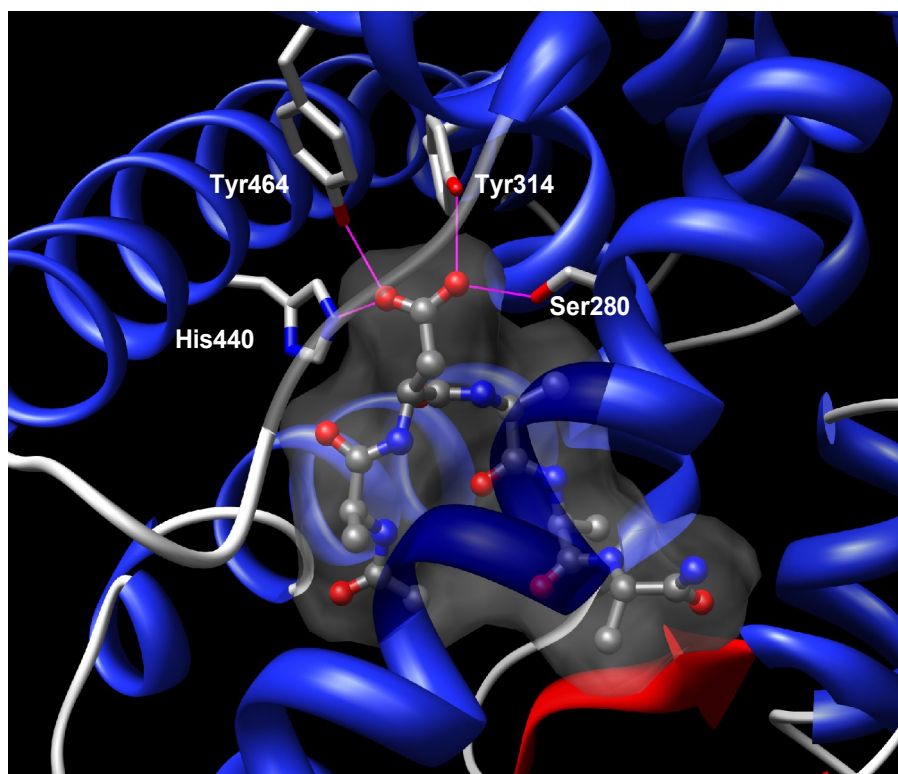


Figure 18 : Mode d'interaction calculé pour le ligand potentiel virtuel ADAAA. La fonction carboxylate de l'aspartate participe à un réseau de liaisons hydrogène avec Ser280, Tyr314, His440 et Tyr464.

Deux cycles d'optimisation ont été réalisés à partir du hit peptidique virtuel potentiel ADAAA. Les cartes d'interactions favorables réalisées n'ont suggéré aucune mutation pour l'alanine N-terminale. L'aspartate en position 2 pourrait être remplacé par un glutamate, et une mutation de l'alanine en position 3 vers la serine pourrait permettre la création d'une liaison hydrogène supplémentaire avec la chaîne latérale de la Ser280. L'espace disponible autour de l'alanine en position 4 est très réduit, et aucune mutation n'a donc pu être proposée. Enfin, l'alanine C-terminale semblerait pouvoir être remplacée par un acide aminé hydrophobe. La Figure 19 illustre quelques positions favorables de la chaîne latérale du tryptophane, dont une pouvant être liée au reste du peptide et faisant des interactions de van der Waals favorables avec Cys275, Val332 et Ala333), ainsi qu'une interaction π - π avec la chaîne latérale de Tyr334 (T-stacking).

L'affinité de plusieurs peptides a été réalisée expérimentalement.

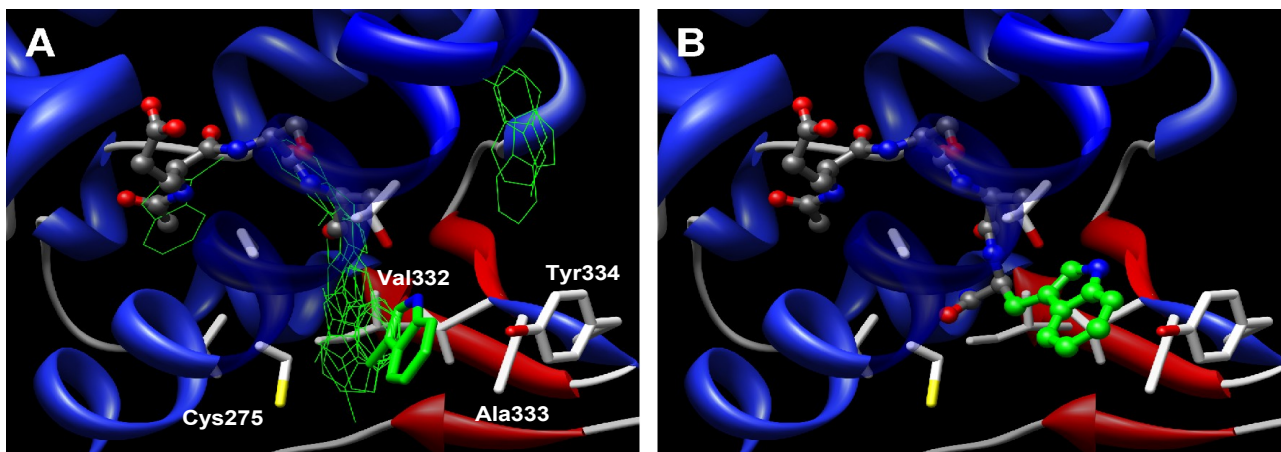


Figure 19 : (A) Mode d'interaction calculé pour un peptide template (boule/bâton). Les acides aminés N et C terminaux ne sont pas représentés par souci de clarté. Les positions favorables de la chaîne latérale du tryptophane sont représentées par de fines lignes vertes, sauf celle permettant une reconnexion avec le reste du peptide par l'intermédiaire de son acide aminé C-terminal. (B) Mode d'interaction calculé pour le peptide modifié. La chaîne latérale du tryptophane est colorée en vert.

Aucune activité détectable n'a pu être mise en évidence pour ADA33. Bien que surprenant, ce résultat est intéressant car il confirme qu'il n'est pas nécessaire qu'un ligand ait une activité détectable pour alimenter le cycle d'optimisation de séquence, dès lors que son mode d'interaction prédit *in silico* est bien défini et reproductible. Nos résultats préliminaires montrent que l'un des peptides a une activité similaire à celle du ligand organique Wy 14,643. Ces premiers résultats sont très encourageants au regard de l'absence de point de départ expérimental de l'optimisation, et en gardant à l'esprit que l'optimisation de séquence s'est faite intégralement *in silico*, montrant ainsi l'intérêt clair de ces méthodes dans la conception de nouvelles molécules actives.

6.4.4 Conclusion

L'approche basée sur des fragments utilisée dans ce travail a conduit à plusieurs tétrapeptides et pentapeptides dont certains ont montré une activité sur le PPAR α détectable expérimentalement. L'activité mesurée de l'une des molécules testées est similaire à celle du Wy14,643, ligand organique bien connu de PPAR α .

D'autres cycles d'optimisation de séquence vont être réalisés afin d'améliorer encore l'affinité de ces peptides. Plusieurs modifications peuvent être envisagées afin de limiter leurs sensibilités aux protéases, comme l'introduction d'acides aminés D dans leurs séquences.

7 Conclusion

Les approches théoriques permettant la prédiction du mode d'interaction d'un ligand avec son récepteur sont complémentaires des expériences *in vitro* et *in vivo*, et permettent parfois leur interprétation. Elles donnent un éclairage irremplaçable sur les interactions au niveau moléculaire et sont, de ce fait, le socle à partir duquel une conception ou une optimisation rationnelle de molécules actives, prenant en compte des critères structuraux, peut être envisagée. De plus, elles permettent de réduire significativement les expériences *in vitro* en focalisant la recherche sur une petite fraction intéressante d'une base de données de molécules par ailleurs trop volumineuse pour être explorée expérimentalement. Les données expérimentales sont néanmoins absolument nécessaires au développement des outils théoriques et à leurs validations. Cette synergie entre les approches théoriques et expérimentales est cruciale pour une évolution optimale des connaissances en biologie, pharmacie et médecine. Un tel environnement de recherche est très certainement l'une des clés vers un chemin de plus en plus court entre l'identification d'un besoin thérapeutique à la délivrance du médicament *ad hoc* au patient.

THÈSE SOUTENUE PAR : AURÉLIEN GROSDIDIER

TITRE : Conception d'un logiciel de docking et applications dans la recherche de nouvelles molécules actives.

CONCLUSION

Les récentes difficultés de l'industrie pharmaceutique ne semblent pouvoir se résoudre que par l'optimisation de leur processus de développement de médicaments. Cette dernière implique de plus en plus de techniques dites "haut-débit", particulièrement efficaces lorsqu'elles sont couplées aux outils informatiques permettant de gérer la masse de données produite. Désormais, les approches *in silico* telles que le criblage virtuel ou la conception rationnelle de nouvelles molécules actives sont utilisées couramment. Toutes deux reposent sur la capacité à prédire les détails de l'interaction moléculaire entre une molécule ressemblant à un principe actif (PA) et une protéine cible ayant un intérêt thérapeutique. Les comparatifs de logiciels s'attaquant à cette prédiction sont flatteurs, mais plusieurs problèmes subsistent. La littérature récente tend à remettre en cause leur fiabilité, affirmant l'émergence d'un besoin pour des approches plus précises du mode d'interaction. Cette précision est essentielle au calcul de l'énergie libre de liaison, qui est directement liée à l'affinité du PA potentiel à la protéine cible, et indirectement liée à son activité biologique. Une prédiction précise est d'une importance toute particulière pour la découverte et l'optimisation de nouvelles molécules actives.

Cette thèse présente un nouveau logiciel, EADock, mettant en avant une telle précision. Cet algorithme évolutionnaire hybride utilise deux pressions de sélections, combinées à une gestion de la diversité sophistiquée. EADock repose sur CHARMM pour les calculs d'énergie et la gestion des coordonnées atomiques. Sa validation a été effectuée sur 37 complexes protéine-ligand cristallisés, incluant 11 protéines différentes. L'espace de recherche a été étendu à une sphère de 15 Å de rayon autour du centre de masse du ligand cristallisé, et contrairement aux comparatifs habituels, l'algorithme est parti de solutions déjà optimisées et présentant un RMSD jusqu'à 10 Å par rapport à la structure cristalline. Cette validation a permis de mettre en évidence l'efficacité de notre heuristique de recherche car des modes d'interactions présentant un RMSD inférieur à 2 Å par rapport à la structure cristalline ont été classés premier pour 68% des complexes. Lorsque les cinq meilleures solutions sont prises en compte, le taux de succès grimpe à 78%, et 92% lorsque la totalité de la dernière génération est prise en compte. La plupart des erreurs de prédiction sont imputables à la présence de contacts cristallins.

Depuis, EADock a été utilisé pour comprendre les mécanismes moléculaires impliqués dans la régulation de la Na,K-ATPase et dans l'activation du peroxisome proliferator-activated receptor α (PPAR α). Il a également permis de décrire l'interaction de polluants couramment rencontrés sur PPAR γ , ainsi que l'influence de la métabolisation de l'Imatinib (PA anticancéreux) sur la fixation à la kinase BCR-ABL. Une approche basée sur la prédiction des interactions de fragments moléculaires avec protéine cible est également proposée. Elle a permis la découverte de nouveaux ligands peptidiques de PPAR α et de l'intégrine $\alpha 5\beta 1$. Dans les deux cas, l'activité de ces nouveaux peptides est comparable à celles de ligands bien établis, comme le Wy14,643 pour le premier, et le Cilengitide (PA anticancéreux) pour la seconde.

VU ET PERMIS D'IMPRIMER

Grenoble, le 23 mai 2007

LE DOYEN

Professeur Renée GRILLO



LE PRESIDENT DE LA THÈSE

A handwritten signature in black ink, consisting of a long vertical stroke followed by a series of loops and a final horizontal stroke.

8 Bibliographie

- 1: ABAGYAN R., TOTROV M., Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins., J Mol Biol, 1994, **235** : 983-1002
- 2: AMIDON G. L., YALKOWSKY S. H., ANIK S. T., VALVANI S. C., Solubility of nonelectrolytes in polar solvents. V. Estimation of the solubility of aliphatic monofunctional compounds in water using a molecular surface area approach, J. Phys. Chem., 1975, **79** : 2239 - 2246
- 3: AN J., TOTROV M., ABAGYAN R., Pocketome via comprehensive identification and classification of ligand binding envelopes., Mol. Cell. Proteomics, 2005, **4** : 752-761
- 4: BARAKAT M. T., DEAN P. M., The atom assignment problem in automated de novo drug design. 1. Transferability of molecular fragment properties., J Comput Aided Mol Des, 1995, **9** : 341-350
- 5: BARAKAT M. T., DEAN P. M., The atom assignment problem in automated de novo drug design. 2. A method for molecular graph and fragment perception., J Comput Aided Mol Des, 1995, **9** : 351-358
- 6: BARAKAT M. T., DEAN P. M., The atom assignment problem in automated de novo drug design. 3. Algorithms for optimization of fragment placement onto 3D molecular graphs., J Comput Aided Mol Des, 1995, **9** : 359-372
- 7: BARAKAT M. T., DEAN P. M., The atom assignment problem in automated de novo drug design. 4. Tests for site-directed fragment placement based on molecular complementarity., J Comput Aided Mol Des, 1995, **9** : 448-456
- 8: BARAKAT M. T., DEAN P. M., The atom assignment problem in automated de novo drug design. 5. Tests for envelope-directed fragment placement based on molecular similarity., J Comput Aided Mol Des, 1995, **9** : 457-462
- 9: BAXTER C. A., MURRAY C. W., CLARK D. E., WESTHEAD D. R., ELDRIDGE M. D.,

Flexible docking using Tabu search and an empirical estimate of binding affinity., *Proteins*, 1998, **33** : 367-382

10: BEMIS G. W., MURCKO M. A., Properties of known drugs. 2. Side chains., *J. Med. Chem.*, 1999, **42** : 5095-5099

11: BEMIS G. W., MURCKO M. A., The properties of known drugs. 1. Molecular frameworks., *J. Med. Chem.*, 1996, **39** : 2887-2893

12: BISSANTZ C., FOLKERS G., ROGNAN D., Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations., *J Med Chem*, 2000, **43** : 4759-4767

13: BOEHM H. J., LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads., *J Comput Aided Mol Des*, 1992, **6** : 593-606

14: BOEHM H. J., Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs, *J. Comput. Aided Mol. Des.*, 1998, **12** : 309-323

15: BOEHM H. J., The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure, *J. Comput. Aided Mol. Des.*, 1994, **8** : 243-256

16: BOEHM H. J., BOEHRINGER M., BUR D. et al, Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening., *J Med Chem*, 2000, **43** : 2664-2674

17: BRAKEBUSCH C., FAESSLER R., The integrin-actin connection, an eternal love affair., *EMBO J*, 2003, **22** : 2324-2333

18: BROOKS B. R., BRUCCOLERI R. E., OLAFSON B. D., STATES D. J., SWAMINATHAN S., KARPLUS M., CHARMM: a program for macromolecular energy minimization, and dynamics calculations., *J. Comput. Chem.*, 1983, : 187-217

19: BRUNGER A. T., KARPLUS M., Polar hydrogen positions in proteins: empirical energy

placement and neutron diffraction comparison., *Proteins*, 1988, **4** : 148-156

20: BUDIN N., MAJEUX N., CAFLISCH A., Fragment-Based flexible ligand docking by evolutionary optimization., *Biol. Chem.*, 2001, **382** : 1365-72

21: BURKE PATRICIA A, DENARDO SALLY J, MIERS LAIRD A, LAMBORN KATHLEEN R, MATZKU SIEGFRIED, DENARDO GERALD L, Cilengitide targeting of alpha(v)beta(3) integrin receptor synergizes with radioimmunotherapy to increase efficacy and apoptosis in breast cancer xenografts., *Cancer Res*, 2002, **62** : 4263-4272

22: BURSULAYA B. D., TOTROV M., ABAGYAN R., BROOKS C. L., Comparative study of several algorithms for flexible ligand docking., *J. Comput. Aided. Mol. Des.*, 2003, **17** : 755-63

23: CAMACHO C. J., VAJDA S., Protein docking along smooth association pathways., *Proc. Natl. Acad. Sci. U S A*, 2001, **98** : 10636-10641

24: CARR R., CONGREVE M., MURRAY C. W., REES D. C., Fragment-based lead discovery: leads by design., *Drug Discov Today*, 2005, **10** : 987-992

25: CHAN S. L., CHAU P. L., GOODMAN J. M., Ligand atom partial charged assignment for complementary electrostatic potentials, *Journal of computer-aided molecular design*, 1992, **6** : 461-474

26: CHARIFSON P. S., CORKERY J. J., MURCKO M. A., WALTERS W. P., Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins., *J Med Chem*, 1999, **42** : 5100-5109

27: CHEN H., LYNE P. D., G. F., LOVELL T., LI J., On evaluating molecular-docking methods for pose prediction and enrichment factors., *J. Chem. Inf. Model.*, 2006, **46** : 401-415

28: CHO A. E., WENDEL J. A., VAIDEHI N. et al, The MPSim-Dock hierarchical docking algorithm: application to the eight trypsin inhibitor cocrystals., *J. Comput. Chem.*, 2005, **26** : 48-71

29: CLAUSSEN H., BUNING C., RAREY M., LENGAUER T., FlexE: efficient molecular

- docking considering protein structure variations., *J. Mol. Biol.*, 2001, **308** : 377-95
- 30: COLE J. C., MURRAY C. W., WILLEM J., NISSINK M., TAYLOR R. D., TAYLOR R., Comparing protein-ligand docking programs is difficult., *Proteins*, 2005, **60** : 325-332
- 31: CONGREVE M., CARR R., MURRAY C., JHOTI H., A 'rule of three' for fragment-based lead discovery?, *Drug Discov Today*, 2003, **8** : 876-877
- 32: D'ARDENNE A. J., RICHMAN P. I., HORTON M. A., MCAULAY A. E., JORDAN S., Coordinate expression of the alpha-6 integrin laminin receptor sub-unit and laminin in breast cancer., *J Pathol*, 1991, **165** : 213-220
- 33: DANZON P. M., EPSTEIN A., NICHOLSON S., Mergers and Acquisitions in the Pharmaceutical and Biotech Industries, NBER Working Paper, 2004, **10536** :
- 34: DAVIS L., *Handbook of Genetic Algorithms*, 1991
- 35: EISEN M. B., WILEY D. C., KARPLUS M., HUBBARD R. E., HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site., *Proteins*, 1994, **19** : 199-221
- 36: ERICKSON J. A., JALAIIE M., ROBERTSON D. H., LEWIS R. A., VIETH M., Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy., *J. Med. Chem.*, 2004, **47** : 45-55
- 37: ERLANSON D. A., Fragment-based lead discovery: a chemical update., *Curr Opin Biotechnol*, 2006, **17** : 643-652
- 38: ERLANSON D. A., MCDOWELL R. S., O'BRIEN T., Fragment-based drug discovery., *J Med Chem*, 2004, **47** : 3463-3482
- 39: ERLANSON D. A., WELLS J. A., BRAISTED A. C., Tethering: fragment-based drug discovery., *Annu Rev Biophys Biomol Struct*, 2004, **33** : 199-223
- 40: EWING T. J., MAKINO S., SKILLMAN A. G., KUNTZ I. D., DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases., *J. Comput. Aided. Mol. Des.*, 2001, **15** : 411-428

- 41: FAESSLER R., Molekulare Medizin, 2004
- 42: FELDING-HABERMANN B., Integrin adhesion receptors in tumor metastasis., Clin Exp Metastasis, 2003, **20** : 203-213
- 43: FERRARA P., GOHLKE H., PRICE D. J., KLEBE G., BROOKS C. L., Assessing scoring functions for protein-ligand interactions., J. Med. Chem., 2004, **47** : 3032-47
- 44: FISER A., DO R. K., SALI A., Modeling of loops in protein structures., Protein Sci, 2000, **9** : 1753-1773
- 45: FRANTZ S., Study reveals secrets to faster drug development., Nat Rev Drug Discov, 2006, **5** : 883
- 46: FRANTZ S., Pipeline problems are increasing the urge to merge., Nat Rev Drug Discov, 2006, **5** : 977-979
- 47: FRANTZ S., Pharma faces major challenges after a year of failures and heated battles, Nat Rev Drug Discov, 2007, **6** : 5-7
- 48: FRIESNER R. A., BANKS J. L., MURPHY R. B. et al, Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy., J. Med. Chem., 2004, **47** : 1739-49
- 49: GILLET V. J., NEWELL W., MATA P. et al, SPROUT: recent developments in the de novo design of molecules., J Chem Inf Comput Sci, 1994, **34** : 207-217
- 50: GILSON M. K., ZHOU H.-X., Calculation of Protein-Ligand Binding Affinities., Annu Rev Biophys Biomol Struct, 2007, **36** : 21-42
- 51: GLASER F., MORRIS R. J., NAJMANOVICH R. J., LASKOWSKI R. A., THORNTON J. M., A method for localizing ligand binding pockets in protein structures., Proteins, 2006, **62** : 479-488
- 52: GOHLKE H., KIEL C., CASE D. A., Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes, J. Mol. Biol., 2003, **330** : 891-913

- 53: GOLDBERG D. E., Genetic Algorithms in Search Optimization and Machine Learning, 1989
- 54: GOODFORD P. J., A computational procedure for determining energetically favorable binding sites on biologically important macromolecules., J Med Chem, 1985, **28** : 849-857
- 55: GROSDIDIER A., ZOETE V., MICHIELIN O., EADock: docking small molecules into protein active sites with a multiobjective evolutionary optimization, Proteins: structure, function and bioinformatics, 2007, **67** : 1010-1025
- 56: HAJDUK P. J., BURES M., PRAESTGAARD J., FESIK S. W., Privileged molecules for protein binding identified from NMR-based screening., J Med Chem, 2000, **43** : 3443-3447
- 57: HALGREN T. H., Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94, J. Comput. Chem, 1996, **17** : 490-519
- 58: HALGREN T. H., Merck Molecular Force Field. II. MMFF94 van der Waals and Electrostatic Parameters for Intermolecular Interactions, J. Comput. Chem, 1996, **17** : 520-552
- 59: HALGREN T. H., Merck Molecular Force Field. III. Molecular Geometries and Vibrational Frequencies for MMFF94, J. Comput. Chem, 1996, **17** : 553-586
- 60: HALGREN T. H., Merck Molecular Force Field. IV. Conformational Energies and Geometries for MMFF94, J. Comput. Chem, 1996, **17** : 587-615
- 61: HALGREN T. H., Merck Molecular Force Field. V. Extension of MMFF94 Using Experimental Data, Additonal Computational Data, and Empirical Rules, J. Comput. Chem, 1996, **17** : 616-641
- 62: HASEL W., HENDRIKSON T. F., STILL W. C., A rapid approximation to the solvent accessible surface areas of atoms, Tetrahedron Comput. Methodol., 1988, **1** : 103-116
- 63: HEAD R. D., SMYTHE M. L., OPREA T. I., WALLER C. L., GREEN S. M., MARSHALL G. R., VALIDATE: a new method for the Receptor-Based prediction of binding affinities of Novel ligands, J. Am. Chem. Soc., 1996, **118** : 3959-3969

- 64: HEHLGANS S., HAASE M., CORDES N., Signalling via integrins: implications for cell survival and anticancer strategies., *Biochim Biophys Acta*, 2007, **1775** : 163-180
- 65: HERMANN R. B., Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area, *J. Phys. Chem.*, 1972, **76** : 2754 - 2759
- 66: HOLLOWAY M. K., A priori Prediction of Activity for HIV-1 Protease Inhibitors Employing Energy Minimization in the Active Site, *J. Med. Chem.*, 1995, **38** : 305-317
- 67: HOPKINS A. L., GROOM C. R., ALEX A., Ligand efficiency: a useful metric for lead selection., *Drug Discov Today*, 2004, **9** : 430-431
- 68: HUANG D., LUETHI U., KOLB P. et al, Discovery of cell-permeable non-peptide inhibitors of beta-secretase by high-throughput docking and continuum electrostatics calculations., *J Med Chem*, 2005, **48** : 5108-5111
- 69: HUANG D., LUETHI U., KOLB P., CECCHINI M., BARBERIS A., CAFLISCH A., In silico discovery of beta-secretase inhibitors., *J Am Chem Soc*, 2006, **128** : 5436-5443
- 70: HYNES R. O., Integrins: bidirectional, allosteric signaling machines., *Cell*, 2002, **110** : 673-687
- 71: JACOBY E., D. J., BLOMMERS M. J. J., Design of small molecule libraries for NMR screening and other applications in drug discovery., *Curr Top Med Chem*, 2003, **3** : 11-23
- 72: JONES G., WILLETT P., GLEN R. C., LEACH A. R., TAYLOR R., Development and validation of a genetic algorithm for flexible docking., *J. Mol. Biol.*, 1997, **267** : 727-48
- 73: JORGENSEN W. L., The many roles of computation in drug discovery., *Science*, 2004, **303** : 1813-1818
- 74: KAIRYS V., FERNANDES M. X., GILSON M. K., Screening drug-like compounds by docking to homology models: a systematic study., *J. Chem. Inf. Model.*, 2006, **46** : 365-379
- 75: KESERU G. M., MAKARA G. M., Hit discovery and hit-to-lead approaches., *Drug Discov Today*, 2006, **11** : 741-748
- 76: KITCHEN D. B., DECORNEZ H., FURR J. R., BAJORATH J., Docking and scoring in

virtual screening for drug discovery: methods and applications., Nat. Rev. Drug. Discov., 2004, **3** : 935-49

77: KLEBE G., Virtual ligand screening: strategies, perspectives and limitations., Drug Discov Today, 2006, **11** : 580-594

78: KOLB P., CAFLISCH A., Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking., J Med Chem, 2006, **49** : 7384-7392

79: KOLLMAN P., Free Energy Calculations: Applications to Chemical and Biochemical Phenomena, Chemical Reviews, 1993, **93** : 2395-2417

80: KONTOYIANNI M., MCCLELLAN L. M., SOKOL G. S., Evaluation of docking performance: comparative data on docking algorithms., J Med Chem, 2004, **47** : 558-565

81: KONTOYIANNI M., SOKOL G. S., MCCLELLAN L. M., Evaluation of library ranking efficacy in virtual screening., J Comput Chem, 2005, **26** : 11-22

82: KUMMAR S., KINDERS R., RUBINSTEIN L. et al, Compressing drug development timelines in oncology using phase '0' trials, Nat Rev Cancer, 2007, **7** : 131-139

83: LARSON R. S., Bioinformatics And Drug Discovery, 2005

84: LASKOWSKI R. A., THORNTON J. M., HUMBLET C., SINGH J., X-SITE: use of empirically derived atomic packing preferences to identify favourable interaction regions in the binding sites of proteins., J Mol Biol, 1996, **259** : 175-201

85: LEACH A. R., KILVINGTON S. R., Automated molecular design: a new fragment-joining algorithm., J Comput Aided Mol Des, 1994, **8** : 283-298

86: LEACH A. R., SHOICHET B. K., PEISHOFF C., Prediction of Protein-Ligand Interactions. Docking and Scoring: Successes and Gaps., J Med Chem, 2006, **49** : 5851-5855

87: LEE M. S., FEIG M., SALSBURY F. R., BROOKS C. L., New analytic approximation to the standard molecular volume definition and its application to generalized born

calculations, *J. Comput. Chem.*, 2003, **24** : 1348-1356

88: LEE M. S., SALSBUURY F. R., BROOKS C. L., Novel generalized Born methods, *J. Chem. Phys.*, 2002, **116** : 10606-10614

89: LEWELL X. Q., JUDD D. B., WATSON S. P., HANN M. M., RECAP-retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry., *J Chem Inf Comput Sci*, 1998, **38** : 511-522

90: LEWIS R. A., Automated site-directed drug design: a method for the generation of general three-dimensional molecular graphs., *J Mol Graph*, 1992, **10** : 131-143

91: LEWIS R. A., Automated site-directed drug design: approaches to the formation of 3D molecular graphs., *J Comput Aided Mol Des*, 1990, **4** : 205-210

92: LEWIS R. A., DEAN P. M., Automated site-directed drug design: the concept of spacer skeletons for primary structure generation., *Proc R Soc Lond B Biol Sci*, 1989, **236** : 125-140

93: LEWIS R. A., DEAN P. M., Automated site-directed drug design: the formation of molecular templates in primary structure generation., *Proc R Soc Lond B Biol Sci*, 1989, **236** : 141-162

94: LEWIS R. A., ROE D. C., HUANG C., FERRIN T. E., LANGRIDGE R., KUNTZ I. D., Automated site-directed drug design using molecular lattices., *J Mol Graph*, 1992, **10** : 66-78, 106

95: LIPINSKI C. A., LOMBARDO F., DOMINY B. W., FEENEY P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Advanced Drug Delivery Reviews*, 2001, **46** : 3-26

96: MACKERELL A. D., BASHFORD D., BELLOTT M. et al, All-atom empirical potential for molecular modeling and dynamics studies of proteins, *Journal of Physical Chemistry B*, 1998, **102** : 3586-3616

97: MAJEUX N., SCARSI M., APOSTOLAKIS J., EHRHARDT C., CAFLISCH A., Exhaustive

- docking of molecular fragments with electrostatic solvation., *Proteins*, 1999, **37** : 88-105
- 98: MICHALIK ., ZOETE V., KREY G. et al, Combined simulation and mutagenesis analyses reveal the involvement of key residues for PPARalpha helix 12 dynamic behavior., *J Biol Chem*, 2007, **282** : 9666-9677
- 99: MILLER M. D., KEARSLEY S. K., UNDERWOOD D. J., SHERIDAN R. P., FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure., *J. Comput. Aided. Mol. Des.*, 1994, **8** : 153-174
- 100: MIRANKER A., KARPLUS M., Functionality maps of binding sites: a multiple copy simultaneous search method., *Proteins*, 1991, **11** : 29-34
- 101: MORAN M., Cost of Bringing New Drugs To Market Rising Rapidly, *Psychiatr News*, 2003, **38** : 25
- 102: MORRIS G. M., GOODSELL D. S., HALLIDAY R. S. et al, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function., *J. Comput. Chem.*, 1998, **19** : 1639-1662
- 103: MUEGGE I., A knowledge-based scoring function for protein-ligand interactions: Probing the reference state, *Perspectives in Drug Discovery and Design*, 2000, **20** : 99-114
- 104: MUEGGE I., MARTIN Y., A general and fast scoring function for protein-ligand interactions: a simplified potential approach, *J. Med. Chem.*, 1999, **42** : 791-804
- 105: NAYAL M., HONIG B., On the nature of cavities on protein surfaces: Application to the identification of drug-binding sites., *Proteins*, 2006, **63** : 892-906
- 106: PARISE L. V., LEE J., JULIANO R. L., New aspects of integrin signaling in cancer., *Semin Cancer Biol*, 2000, **10** : 407-414
- 107: PEI J., WANG Q., LIU Z., LI Q., YANG K., LAI L., PSI-DOCK: Towards highly efficient and accurate flexible ligand docking., *Proteins*, 2006, **62** : 934-946
- 108: PEREZ C., PASTOR M., ORTIZ A. R., GAGO F., Comparative binding energy analysis

of HIV-1 Protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptor-based drug design, *J. Med. Chem.*, 1998, **41** : 836-852

109: PEROLA E., CHARIFSON P. S., Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding., *J. Med. Chem.*, 2004, **47** : 2499-510

110: PEROLA E., WALTERS W. P., CHARIFSON P. J., Comments on the Article "On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors", *J. Chem. Inf. Model.*, 2007, **47** : 251-253

111: PEROLA E., WALTERS W. P., CHARIFSON P. S., A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance., *Proteins*, 2004, **56** : 235-49

112: PETERSEN O. W., RØNNOV-JESSEN L., HOWLETT A. R., BISSELL M. J., Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells., *Proc Natl Acad Sci U S A*, 1992, **89** : 9064-9068

113: POULOS T.L., HOWARD A.J., Crystal structures of metyrapone- and phenylimidazole-inhibited complexes of cytochrome P-450cam., *Biochemistry*, 1987, **26** : 8165-8174

114: PROUDFOOT J. R., Drugs, leads, and drug-likeness: an analysis of some recently launched drugs., *Bioorg Med Chem Lett*, 2002, **12** : 1647-1650

115: RAREY M., KRAMER B., LENGAUER T., KLEBE G., A fast flexible docking method using an incremental construction algorithm., *J Mol Biol*, 1996, **261** : 470-489

116: SCHWARTZ M. A., Integrin signaling revisited., *Trends Cell Biol*, 2001, **11** : 466-470

117: SCHWARTZ M. A., ASSOIAN R. K., Integrins and cell proliferation: regulation of cyclin-dependent kinases via cytoplasmic signaling pathways., *J Cell Sci*, 2001, **114** : 2553-2560

118: SCHWARTZ M. A., GINSBERG M. H., Networks and crosstalk: integrin signalling spreads., *Nat Cell Biol*, 2002, **4** : E65-E68

- 119: SIMONSON T., ARCHONTIS G., KARPLUS M., Free energy simulations come of age: protein-ligand recognition., *Acc. Chem. Res.*, 2002, **35** : 430-437
- 120: SIROCKIN F., SICH C., IMPROTA S. et al, Structure activity relationship by NMR and by computer: a comparative study., *J. Am. Chem. Soc.*, 2002, **124** : 11073-11084
- 121: SOUSA S. F., FERNANDES P. A., RAMOS M. J., Protein-ligand docking: current status and future challenges., *Proteins*, 2006, **65** : 15-26
- 122: STIL W. C., TEMPCZYK A., HAWLEY R. C., HENDRICKSON T., Semianalytical treatment of solvation for molecular mechanics and dynamics, *J. Am. Chem. Soc.*, 1990, **112** : 6127 - 6129
- 123: STULTZ C. M., KARPLUS M., Dynamic ligand design and combinatorial optimization: designing inhibitors to endothiapepsin., *Proteins*, 2000, **40** : 258-289
- 124: TAYLOR J. S., BURNETT R. M., DARWIN: a program for docking flexible molecules., *Proteins*, 2000, **41** : 173-91
- 125: THIEL K. A., Structure-aided drug design's next generation., *Nat Biotechnol*, 2004, **22** : 513-519
- 126: TIRADO-RIVES J., JORGENSEN W. L., Contribution of conformer focusing to the uncertainty in predicting free energies for protein-ligand binding., *J Med Chem*, 2006, **49** : 5880-5884
- 127: TODOROV N. P., DEAN P. M., A branch-and-bound method for optimal atom-type assignment in de novo ligand design., *J Comput Aided Mol Des*, 1998, **12** : 335-349
- 128: TODOROV N. P., DEAN P. M., Evaluation of a method for controlling molecular scaffold diversity in de novo ligand design., *J Comput Aided Mol Des*, 1997, **11** : 175-192
- 129: VIETH M., HIRST J. D., KOLINSKI A., BROOKS C. L., Assessing energy functions for flexible docking., *J. Comput. Chem.*, 1998, **19** : 1612-1622
- 130: VIETH M., SIEGEL M. G., HIGGS R. E. et al, Characteristic physical properties and structural fragments of marketed oral drugs., *J Med Chem*, 2004, **47** : 224-232

- 131: WARREN G., ANDREWS C., CAPELLI A.-M. et al, A Critical Assessment of Docking Programs and Scoring Functions., J Med Chem, 2006, **49** : 5912-5931
- 132: WATT F. M., Role of integrins in regulating epidermal adhesion, growth and differentiation., EMBO J, 2002, **21** : 3919-3926
- 133: WELCH W., RUPPERT J., JAIN A. N., Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites., Chem. Biol., 1996, **3** : 449-462
- 134: WILHELMSSEN K., LITJENS S. H. M., SONNENBERG A., Multiple functions of the integrin alpha6beta4 in epidermal homeostasis and tumorigenesis., Mol Cell Biol, 2006, **26** : 2877-2886
- 135: WU G., ROBERTSON D. H., BROOKS C. L., VIETH M., Detailed analysis of grid-based molecular docking: A case study of CDOCKER-A CHARMM-based MD docking algorithm., J. Comput. Chem., 2003, **24** : 1549-62
- 136: ZAHIR N., WEAVER V. M., Death in the third dimension: apoptosis regulation and tissue architecture., Curr Opin Genet Dev, 2004, **14** : 71-80
- 137: ZOETE V., GROSDIDIER A., MICHIELIN O., Peroxisome proliferator-activated receptor structures: Ligand specificity, molecular switch and interactions with regulators., Biochim Biophys Acta, 2007, :
- 138: ZOETE V., MEUWLY M., KARPLUS M., Study of the insulin dimerization from binding free energy calculations and per-residue free energy decomposition, Proteins, 2005, **61** : 79-93

9 Serment des apothicaires

Je jure, en présence des maîtres de la faculté, des conseillers de l'ordre des pharmaciens et de mes condisciples :

D'honorer ceux qui m'ont instruit dans les préceptes de mon art et de leur témoigner ma reconnaissance en restant fidèle à leur enseignement.

D'exercer, dans l'intérêt de la santé publique, ma profession avec conscience et de respecter non seulement la législation en vigueur, mais aussi les règles de l'honneur, de la probité et du désintéressement.

De ne jamais oublier ma responsabilité et mes devoirs envers le malade et sa dignité humaine ; en aucun cas je ne consentirai à utiliser mes connaissances et mon état pour corrompre les mœurs et favoriser des actes criminels.

Que les hommes m'accordent leur estime si je suis fidèle à mes promesses. Que je sois couvert d'opprobre et méprisé de mes confrères si j'y manque.

Conception d'un logiciel de docking et applications dans la recherche de nouvelles molécules actives.

Les récentes difficultés de l'industrie pharmaceutique semblent ne pouvoir se résoudre que par l'optimisation de leur processus de développement de médicaments. Cette dernière implique de plus en plus de techniques dites "haut-débit", dont des approches in silico telles que le criblage virtuel ou la conception rationnelle de nouvelles molécules. Toutes deux reposent sur la capacité à prédire l'interaction moléculaire entre un principe actif potentiel et une protéine cible ayant un intérêt thérapeutique. Cette thèse présente un nouveau logiciel réalisant cette prédiction: EADock. Cet algorithme évolutionnaire hybride utilise deux pressions de sélections et repose sur CHARMM. Son efficacité a été mise en évidence sur 37 complexes protéine-ligand cristallisés. Il a également permis la découverte de nouveaux ligands peptidiques de PPAR α et de l'intégrine $\alpha 5\beta 1$. Dans les deux cas, l'activité de ces nouveaux peptides est comparable à celles de ligands bien établis.

Mots-clés : docking, ligand, récepteur, algorithme évolutionnaire, EADock, inhibiteur, fragment

Contact : Aurélien Grosdidier, aurelien.grosdidier@gmail.com, Parc Audra, 26150 Die (France)

Soutenu le: 27 juin 2007 à la Faculté de Pharmacie de Grenoble

Devant le jury composé de:

Pr. Emmanuel Drouet, président du jury

Pr. Olivier Michielin, Directeur de thèse

Pr. Wim Burmeister, expert externe

Dr. V. Zoete, expert interne