

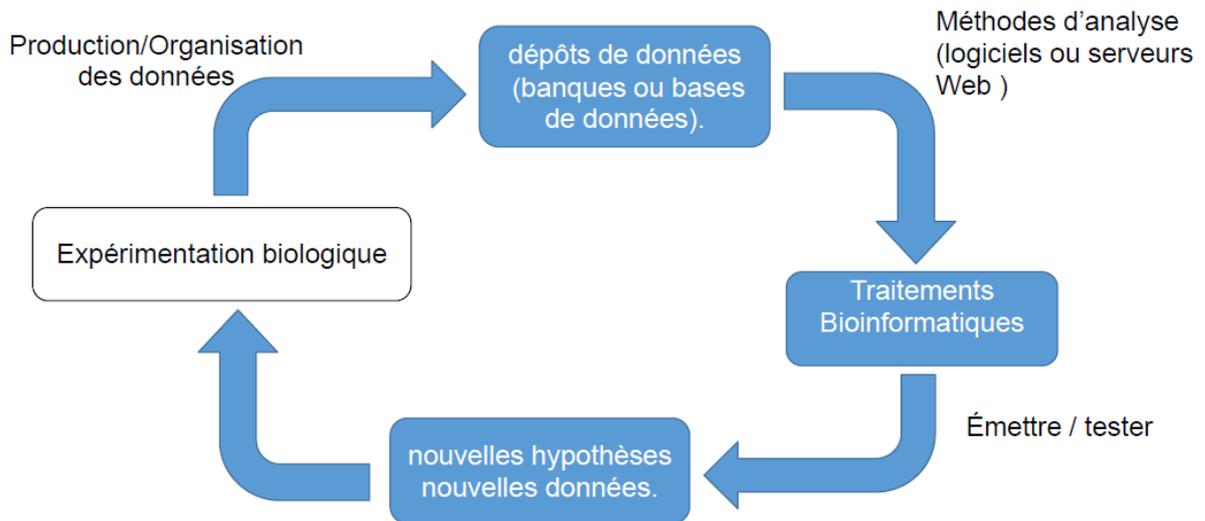
Introduction et rappels bibliographiques

La **bio-informatique** est un domaine de recherche pluridisciplinaire visant à traiter et à analyser les informations biologiques (bibliographie, structure chimique, séquences nucléiques ou protéiques...etc.) en vue de résoudre certains problèmes scientifiques en énonçant des hypothèses (à valider expérimentalement). Cela va de la modélisation moléculaire, à la reconstruction d'arbres phylogénétiques (phylogénie) en passant par l'analyse des séquences nucléiques et/ou protéiques.

Pour certains, la bio-informatique est une branche théorique de la biologie alors que pour d'autres, elle se situe clairement au carrefour de l'informatique et de la biologie car le traitement de l'information en biologie nécessite certainement l'utilisation de l'outil informatique. Ainsi, cette discipline constitue la «biologie *in silico*».

À l'heure actuelle, plusieurs champs d'application ou sous-disciplines de la bio-informatique existent dont les deux principaux sont :

- **La bio-informatique des séquences**, traitant les données issues de l'information génétique contenue dans les séquences nucléiques (ADN et/ou ARN) ou protéiques. Cette branche s'intéresse en particulier à l'identification des ressemblances entre les séquences, à l'identification des gènes ou de régions biologiquement pertinentes dans les protéines, en se basant sur l'enchaînement ou séquence de leurs composants élémentaires (nucléotides, acides aminés).
- **La bio-informatique structurale**, permettant la simulation (prédiction) du mode d'interaction le plus favorable d'un ligand au sein de son récepteur (protéine ou ADN). cette discipline, aussi connue sous le nom du « docking moléculaire » est très utilisée au cours des premières phases du *drug-design*.



1- Introduction : *drug-design*.

Le développement d'un nouveau médicament, aussi appelé drug design, peut être défini par un ensemble de processus inventifs capables de trouver de nouveaux médicaments basés sur les connaissances de cibles biologiques. Le médicament ou la molécule thérapeutique est généralement une petite molécule activant ou inhibant la fonction d'une biomolécule telle que les protéines et entraînant un bénéfice thérapeutique pour le patient. Au cours des phases successives, des milliers de molécules doivent être triées et sélectionnées, pour aboutir à un nombre très restreint de candidats. Pour chaque nouveau médicament, le temps passe depuis la première étape, au cours de laquelle une cible biologique pertinente dans un processus pathologique donné est identifiée, jusqu'à la mise sur le marché d'un composé, tourne autour d'une moyenne d'environ 12 à 14 ans. Le coût global estimé est d'au moins 800 millions de dollars

1-1 Recherche et développement :

1-1-1 Recherche, identification et validation d'une cible thérapeutique

Les maladies se développent lorsque les fonctions de l'organisme sont altérées ou ne répondent pas correctement. Pour développer un médicament, il est important de comprendre en détail (à l'échelle moléculaire) les mécanismes physio-pathologiques installés, afin de pouvoir cibler et corriger le processus anormal. Les cibles thérapeutiques sont souvent des

protéines dont l'activation ou l'inhibition donne un effet thérapeutique vis-à-vis d'une maladie.

L'étape de l'identification d'une cible est une phase appelée "pré-découverte", puisqu'elle précède l'étape de découverte du médicament proprement dit.

1-1-2 Identification des hits (génération des têtes de série)

Cette étape consiste à identifier des molécules têtes de série (hits), c'est-à-dire des molécules ayant la capacité d'interagir avec la cible choisie et susceptibles de moduler ses effets sur le processus biologique en question. Ces molécules peuvent être d'origine naturel ou synthétique.

Dans le passé, la méthode expérimentale établie pour la recherche des hits s'appelle le criblage à haut débit (« HTS » en anglais) où l'activité (ou plutôt l'affinité) d'un très grand nombre de composés est systématiquement testée *in vitro*. Afin de réduire les coûts et le temps du développement d'un nouveau médicament, les compagnies pharmaceutiques utilisent actuellement le criblage virtuel qui consiste à simuler et à prédire *in silico* l'activité biologique de grandes chimiothèques (de 10 000 jusqu'à des millions de composés) vis-à-vis de la cible étudiée. Grâce à cette stratégie, les compagnies pharmaceutiques peuvent aboutir plus rapidement à des listes réduites de molécules présentant l'activité souhaitée (souvent entre 100 et 1000) lesquelles seront testées expérimentalement. Le criblage virtuel permet, donc, d'ajuster au mieux le nombre de tests expérimentaux « *in-vitro*», ce qui est plus rapide et moins coûteux que les criblages expérimentaux.

1-1-3 Optimisation des hits en leads (candidat médicament)

À l'issue du processus de criblage, les têtes de série identifiées doivent souvent être optimisées en ajoutant ou en supprimant certains de leurs groupements chimiques afin d'améliorer davantage leur activité biologique tout en conférant aux molécules optimisées (leads) des propriétés physico-chimiques (ADME) très proches à celles d'un médicament (les rendre moins toxique en améliorant leur pénétration dans la cellule). Il est important de signaler que les approches bio-informatiques peuvent aussi contribuer à l'optimisation des hits en leads.

C'est à ce stade que les données scientifiques et techniques sur le composé candidat, telles que sa structure moléculaire et ses effets, sont généralement enregistrées ou brevetées, afin que le composé soit protégé en tant que propriété intellectuelle.

1-2 Tests de sécurité pré-cliniques.

Les tests de sécurité non cliniques sont effectués chez l'animal. Ces études permettent non seulement de déterminer le profil de sécurité chez l'animal (toxicité), mais elles fournissent également des informations importantes sur :

- L'absorption (façon dont le médicament pénètre dans l'organisme)
- La distribution (façon dont le médicament circule dans l'organisme)
- Le métabolisme (décomposition du médicament par l'organisme)
- L'excrétion (élimination du médicament par l'organisme)

Toutes ces informations sont utilisées pour décider si le composé candidat peut faire l'objet d'une première étude (clinique) chez l'homme et dans l'affirmative, quelles doses devront être utilisées.

1-3 Etude clinique

Beaucoup de candidats médicaments sont éliminés avant ce stade atteint par seulement un médicament sur quinze. Les molécules retenues font, ensuite, l'objet de tests cliniques afin de démontrer leur efficacité. Ces tests sont menés sur l'homme en trois phases principales dans des conditions bien réglementées (dans un cadre juridique).

1-3-1 Phase 1 (étude de la tolérance)

Des doses croissantes de la molécule à l'étude sont administrées à des volontaires **sains** différents sous surveillance étroite au sein de structures particulières agréées, jusqu'à ce qu'apparaissent les premiers signes d'intolérance, repérés grâce à la surveillance clinique et au suivi des constantes biologiques. Cette phase comprend les premières études de pharmacocinétique permettant de déterminer la biodisponibilité, la résorption, la diffusion, le métabolisme et l'élimination du médicament. Elle doit aussi préciser la posologie à appliquer en phase 2.

1-3-2 Phase 2 (étude de l'efficacité)

La phase 2 s'effectue sur des patients modérément atteints par la pathologie cible du médicament candidat. Elle permet de préciser les connaissances de pharmacocinétique et le métabolisme du produit, de recenser ses propriétés pharmacologiques, d'établir les courbes de relation entre sa concentration et les effets obtenus, de préciser la dose optimale pour laquelle l'effet thérapeutique est le meilleur pour le moins d'effets secondaires entraînés.

1-3-3 Phase 3 (essai comparatif)

La phase III concerne un large groupe de malades (plusieurs milliers). Elle consiste à comparer le médicament en développement à un autre médicament ayant déjà fait ses preuves ou à un placebo (un médicament dénué d'activité thérapeutique). L'objectif est de montrer l'efficacité et d'évaluer le rapport efficacité/tolérance.

2- Reconnaissance moléculaire (Interaction Protéine-Ligand)

Dans la conception rationnelle d'un principe actif, la reconnaissance moléculaire est un critère essentiel impliquant une **complémentarité stérique** (spatiale) entre le ligand et sa cible. Cette complémentarité est importante car elle permet de laisser entrer le ligand mais elle ne suffit pas pour qu'il reste au sein de la protéine. Dans ce contexte, une complémentarité chimique Protéine-ligand est indispensable faisant intervenir le plus souvent des **interactions non covalentes** telles que : la liaison hydrogène, ionique, Pi-Pi, hydrophobes...etc.

Les liaisons non covalentes sont des interactions qui ont lieu pour des distances plus élevées que les liaisons covalentes. Bien que ces interactions soient énergétiquement beaucoup plus faible par rapport aux liaisons covalentes (C-C ou C-N par exemple), leur portée est bien plus grande car assurant l'évolution des biomolécules, et notamment les changements conformationnels. En effet, les liaisons non covalentes peuvent être créées et rompues plus facilement.

2-1 Liaison Hydrogène

La liaison hydrogène est une force attractive qui s'opère entre deux groupes d'atomes impliquant un atome d'hydrogène. Elles sont dues à l'attraction d'un atome d'hydrogène lié à un atome électronégatif donneur d'hydrogène (OH, NH ou FH) par un autre atome électronégatif accepteur d'hydrogène (O, N ou F).

Les liaisons hydrogène sont souvent décrites par deux paramètres : la longueur de la liaison et l'angle formé par les trois atomes. Les valeurs des angles et des distances varient respectivement de 90 à 180 degrés et de 1.2 à 4.0 Å. Ces deux paramètres déterminent la force des liaisons hydrogènes variant de 1 à 4 Kcal/mol.

On distingue trois catégories répertoriées dans le tableau

Longueur Å	Angle (°)	Energie (kccal/mol)	description
2.2 à 4.0	90 à 150	0.24 à 1	Faible intensité
1.5 à 3.2	130 à 180	1 à 4	Modérée
1.2 à 2.5	175 à 180	4 à 10	Forte

2-2 Interaction ionique

L'interaction ionique résulte de l'attraction entre deux ions de charges opposées. Sous l'effet de l'attraction électrostatique, les deux ions se rapprochent jusqu'à une distance d'équilibre et forment alors une liaison ionique. La valeur énergétique de la liaison ionique est plus importante à celle de la liaison hydrogène dépassant ainsi les 4 Kcal/mol.

2-3 Les interactions hydrophobiques

Les molécules ou groupes d'atomes non polaires ne sont pas capables de former des liaisons hydrogènes car ne pouvant pas s'hydrater. L'effet hydrophobe est la tendance que possèdent ces groupes à se rassembler de façon à minimiser les contacts avec l'eau. Les interactions hydrophobiques ont lieu pour des distances plus élevées que les liaisons hydrogènes et ioniques (4Å). Certes, elles ont une intensité moindre (inférieure à 1Kcal/mol) mais leur portée est bien plus grande (nombre élevé dans les complexes protéine-ligand).

2-4 Interactions Pi-Pi (π - π stacking ou interactions aromatiques)

Ce sont des interactions hydrophobiques entre deux cycles aromatiques offrant plus de stabilité pour le bon positionnement d'un ligand au sein de son récepteur.

Il existe différents types de positionnement entre deux cycles aromatiques :

- La forme en T ou un cycle se positionne perpendiculairement au plan et dans l'axe du centre du second
- La position parallèle
- La position parallèle déplacée, équivalent au parallèle mais avec un décalage de l'un des deux cycles

Lors de l'interaction Protéine-ligand, l'orientation des deux cycles aromatiques est dépendante du contexte structural du ligand comme de la protéine.

3 Le docking moléculaire

Le docking moléculaire (ancrage, amarrage ou arrimage moléculaire en français) est une approche *in silico* consistant à prédire et à simuler la position la plus favorable d'un ligand au sein de son récepteur (cible souvent protéique). A l'heure actuelle, il existe plus que 40 programmes (logiciels) de docking moléculaire. Bien que ces programmes reposent le plus souvent sur des algorithmes spécifiques, leur protocole est composé de 2 étapes complémentaires :

La première étape dite **docking** permet au ligand d'adopter plusieurs conformations et plusieurs positions au niveau de son récepteur afin de retenir celle la plus favorable.

La deuxième étape, dite **scoring**, permet d'évaluer l'affinité entre le ligand et la protéine et de donner un score aux poses obtenues lors de la phase de docking. Ce score permettra de retenir la meilleure pose parmi toutes celles proposées.

3-1 Applications

Il y a au moins deux applications principales au programme de docking.

La plus ancienne est la prédiction du mode d'interaction qui consiste à déterminer le positionnement correct du ligand par rapport à son récepteur.

La deuxième utilisation réside à rechercher par criblage virtuel de nouvelles molécules actives envers une cible thérapeutique donnée.

Le criblage virtuel *in silico* est une approche permettant de simuler et prédire le mode d'interaction d'un très grand nombre de molécules (collectées en chimiothèques) vis-à-vis un site actif ciblé. Autrement dit, le criblage virtuel est un docking moléculaire de plusieurs milliers de molécules envers un seul récepteur.

3-2 Du docking rigide vers l'introduction de la flexibilité

Les premiers logiciels de docking développés au début des années 80 considéraient le récepteur et le ligand rigides en se basant sur le modèle « clé-serrure ». C'est sans aucun doute la catégorie la plus simple et rapide de docking.

Mis en marche plus tard, le docking semi-flexible a permis d'obtenir des résultats plus précis. Le récepteur est considéré comme corps rigide mais tient compte de la flexibilité du ligand.

Aujourd'hui, la puissance de calcul des ordinateurs permet, pour certains logiciels de docking, de traiter la flexibilité du ligand et de façon partielle le récepteur.

3-3 Outils du docking moléculaire (criblage virtuel)

3-3-1 Récepteur

Les structures 3D des différentes protéines sont disponibles et accessibles gratuitement dans la banque de données PDB. (« *Protein Data Bank* », en anglais). Il s'agit de la plus grande archive de données structurales de macromolécules biologiques (protéines, ADN, ARN ...etc.). Cette banque de données a été établie en 1971 et contenait à l'époque sept structures. Actuellement, la PDB compte plus de 146 000 structures dont 131 000 ont été résolues par cristallographie au rayon X, 12 500 ont été définies par RMN et 2500 qui restent par microscopie électronique. Il est à souligner que 93% des structures disponibles dans la PDB sont des protéines.

Les coordonnées atomiques des structures se trouvent dans un format de référence, appelé *pdb*, lu par différents logiciels de visualisation, de modélisation, de criblage virtuel...etc. Donc il suffit de consulter la PDB et d'inscrire le nom de la structure pour télécharger le fichier du récepteur au format *pdb* indispensable à l'opération de docking moléculaire.

Chaque fichier *pdb* contient diverses informations sur la protéine en question. Par exemple : le nom du récepteur, l'équipe qui a résolu la structure, la méthode expérimentale, ...etc. On y retrouve aussi des informations sur la structure primaire, les hétéroatomes (ligand, métaux, résidus modifiés, ...etc.), la structure secondaire et les coordonnées atomiques X, Y et Z qui déterminent la position exacte de chaque atome dans une conformation donnée.

Il est important de signaler que certaines protéines ne sont pas encore disponibles dans cette banque de données et si cette dernière contient une protéine avec des séquences similaires, il devient alors possible de construire la structure 3D de la cible souhaitée, en faisant appel à la modélisation par homologie.

3-3-2 Ligands (chimiothèques)

Pour un docking moléculaire, le ligand doit être également sous forme 3D. Pour y obtenir, il existe deux moyens :

Le premier moyen consiste à utiliser des programmes de construction moléculaire 3D tels que ChemDraw, Arguslab, Titan, Marvin, Sybyl...etc. Chacun de ces programmes contient une banque d'atomes permettant de dessiner le ligand tout en tenant compte de l'état d'hybridation de chaque atome et de types de liaisons simple, double ou triples. Ensuite, la géométrie du ligand doit être optimisée à l'aide d'un algorithme semi-empirique. Le ligand obtenu est enregistré dans un fichier sous un des formats 3D pouvant être utilisé et lu par un programme de docking moléculaire : *pdb, mol, mol2, SDF*....etc.

Le second moyen d'obtenir un ligand 3D est souvent d'aspect commercial, consistant à consulter des banques de données de ligands appelées chimiothèques.

Les chimiothèques virtuelles comportent un ensemble d'informations plus ou moins organisées et hiérarchisées regroupant des données (structure, activité, propriétés physico-chimiques....) de différents ligands. Dans ce type de chimiothèques, la structure 3D de chaque composé est enregistrée dans de fichiers avec les formats SMILES, SDF, MOL2 et PDB. Parmi les chimiothèques virtuel nous citons : La chimiothèque Nationale Française, la chimiothèque, ZINC, PubChem.....etc.

La Chimiothèque Nationale Française :

Elle regroupe à l'heure actuelle environ 70 000 molécules de synthèse et produits naturels issus des laboratoires académiques français. Cette collection est répertoriée dans une base de données et possède plusieurs modes de stockage physique : des solutions dans le DMSO (diméthylsulfoxyde) au format microplaque et des poudres dans des piluliers.

Pubchem :

Est une banque de données de molécules chimiques gérée par le *National Center for Biotechnology Information* (NCBI), branche de la Bibliothèque nationale de médecine des États-Unis sous l'autorité de la *National Institutes of Health* (NCI). La chimiothèque PubChem répertorie plus de 94 millions de composés en mettant en ligne, gratuitement, pour chaque substance une grande quantité de données de divers ordres : chimique, biochimique, pharmacologique, production, toxicologie, environnemental, etc.

Zinc :

ZINC est gratuit pour tout le monde à utiliser et à télécharger sur le site Web zinc.docking.org. Cette base de données et le service est fourni par le laboratoire Shoichet dans le Département de chimie pharmaceutique à l'Université de Californie à San Francisco (UCSF). Elle comporte environ 35 millions de composés qui peuvent simplement être achetés.

3-3-3 Programmes de docking

Le Docking moléculaire est l'ensemble de méthodes algorithmiques et géométriques permettant la prédiction *in silico* de la conformation (position et orientation relative) la plus favorable d'un ligand au sein de son récepteur. A présent, un bon nombre de programmes du docking moléculaire (commerciaux ou non) est disponible. Les plus fréquemment cités sont respectivement AutoDock (27%), Gold (15%), FlexX (11%), Dock (6%) ou encore IcmDocking (6%)

4 Algorithmes de docking

Les algorithmes de docking ont été conçus pour rechercher de façon exhaustive, rapide et efficace les modes d'association récepteur-ligand les plus favorables. Actuellement, il existe une panoplie d'algorithmes de docking qu'on peut répartir en deux grandes catégories selon la méthode utilisée :

4-1 L'ajustement de fragments (matching)

Ces algorithmes sont dits aussi de fragmentation/reconstruction (fragmentation incrémentale) car, dans un premier temps, la molécule est découpée en parties rigides et en liaisons flexibles. Par la suite, les parties rigides sont positionnées en premier et le ligand est reconstruit de proche en proche, explorant diverses conformations lors de la mise en place des liaisons flexibles. Des logiciels tels que DOCK, Surflex et FlexX utilisent cette approche qui permet notamment un criblage rapide de vastes bibliothèques de composés

4-2 Simulation par trajectoire

La seconde approche, par trajectoire, est plus précise : à partir d'une position initiale aléatoire, à l'extérieur du site actif, le ligand considéré dans son entier explore le site étudié par la répétition successive de mouvements et d'évaluations de l'interaction ligand-récepteur. Les mouvements sont effectués par des opérations de translation, de rotation et de changement de

conformation et bien entendu, l'algorithme s'arrête lorsqu'il trouve la position idéale du ligand dans le récepteur. Ces techniques sont plus lentes que celles par *matching* mais prennent mieux en compte la flexibilité du ligand et permettent l'exploration de régions plus vastes. Dans cette catégorie nous pouvons citer GLIDE, GOLD et AutoDock.

5- Programmes de construction moléculaire

Les programmes Arguslab, Titan, Syby (payant) 1, Marvin, ChemDraw....etc. sont des programmes permettant de construire des modèles 3D de ligands tout en optimisant leur géométrie (algorithme de minimisation de l'énergie).

5-1 Programmes de visualisation moléculaire

Viewerlite (4.2, 2001) et VMDsont des programmes libres (gratuits) de visualisation, permettant un affichage 3D d'une structure de molécule biologique. Ils lisent plusieurs formats de fichier dont les formats *.pdb .mol .mol2 .sdf*. Ces programmes proposent plusieurs fonctions telles que : la présentation des liaisons chimiques (liaison hydrogène), la mesure de distances interatomiques, l'annotation des acides aminés (nom, numéro), créations de surfaces, choix de couleur (selon les atomes, les structures...), la capacité de cacher et puis afficher à nouveau les différentes molécules.

5-2 Programmes de conversion des formats

Il existe différents formats appropriés pour les structures dans leur conformation 1, 2 ou 3 D de molécules. Le format le plus simple (1D) est : SMILES (*.smi Simplified Molecular Input Line Entry Specification*). Il permet de représenter les molécules sur une seule ligne de caractères. Les tirets simples ou doubles représentent des liaisons de même nature, et un système de parenthèses permet de repérer les embranchements au niveau de la molécule. Les chiffres permettent d'identifier les connections de cycle. L'avantage principal de ce format est qu'il est léger, permettant le stockage de bases de molécules de grande taille dans un espace mémoire restreint.

Cependant, d'autres informations sont souvent nécessaires, notamment la structure 3D de la molécule. Des formats comme PDB, SDF ou mol2 contiennent les coordonnées cartésiennes de chaque atome, et éventuellement leur charge partielle (mol2), ainsi que toutes les liaisons formant la molécule.

Le format pdb est un fichier texte ouvrable par des éditeurs texte comme Wordpad. Il contient des les coordonnées X, Y et Z des atomes qui constituent une molécule donnée. Le format pdb contient également des métadonnées pouvant par exemple être : la structure primaire de la molécule, ses éventuelles structures secondaires, la méthode expérimentale qui a permis d'obtenir les coordonnées des atomes, etc.

Open Babel (2.0.2, 1991) est un programme libre, visant à faciliter l'inter conversion des données chimiques d'un format à un autre de fichiers de divers types. Les formats de fichier que « Open Babel » prend en charge comprennent : PDB, MOL, MOL2, SDF, XYZ, PC, SMI...etc. Il existe d'autres convertisseurs de formats en ligne à savoir : Molinspiration WebME Editor et RPBS JME Editor,

6 Application du docking moléculaire dans le *drug design* (criblage virtuel)

Suivant la nature de l'information expérimentale disponible, on distingue deux approches du criblage virtuel. La première reposant sur la connaissance d'un nombre suffisant d'informations concernant une ou plusieurs molécules actives de référence, est appelée "*ligand-based virtual screening*". La seconde, basée sur la structure de la cible est connue sous le nom de "*structure-based virtual screening*". Bien que ces deux approches soient surtout utilisées de manière exclusive (souvent parce que la nature des données de départ ne laisse qu'un seul choix possible), leur combinaison lors d'une campagne de criblage permet de maximiser les chances de succès pour identifier de nouvelles touches.

6-1 Le criblage virtuel « ligand-based » (pharmacophore)

Cette démarche repose sur les connaissances acquises concernant des molécules de référence interagissant expérimentalement avec la cible d'intérêt biologique. Ces molécules peuvent être utilisées afin d'en dériver un pharmacophore qui définit un ensemble de points caractéristiques que doit posséder une molécule pour être active. Autrement dit, un modèle pharmacophorique est un arrangement spatial de points résultant de l'alignement de molécules dont le mode de liaison à un récepteur est connu ou supposé. A chaque point de ce modèle correspond des coordonnées 3D ainsi que des propriétés physico-chimiques (lipophilie, donneur/accepteur de liaisons hydrogène). Son élaboration passe tout d'abord par la sélection de composés actifs pour la cible en question. Des conformations sont ensuite générées. Un alignement des poses sélectionnées permet par la suite de déduire les points significatifs responsables de l'affinité avec le récepteur. L'intérêt de coupler cette technique au docking

est d'une part de pouvoir éliminer directement les molécules indésirables et donc d'économiser du temps de calcul. Par ailleurs, une telle stratégie permet de guider la construction incrémentale du ligand au sein du site actif.

6-2 Le criblage virtuel « structure-based » :

En utilisant la structure de la cible biologique, la conception de molécules ayant une grande affinité sélective à la cible peut être réalisée suivant deux méthodes :

- a- La première catégorie s'intéresse à concevoir *in silico*, un ligand ayant une complémentarité spatiale avec la cible étudiée en y engageant le maximum possible de liaisons (conception *de novo*).
- b- La deuxième catégorie consiste, quant à elle, à cribler un grand nombre de ligands sans tenir compte des contraintes structurales dues à la poche de liaison.

Ainsi, contrairement à l'approche basée sur des ligands de référence, cette approche peut potentiellement identifier de nouvelles classes de molécules actives.

6-3 Les différentes étapes du criblage virtuel de vastes chimiothèques

Le criblage virtuel est largement utilisé pour identifier de nouvelles substances bio-actives. Cette approche est souvent composée de plusieurs étapes essentielles :

a- Préparation de la cible

La préparation de la structure 3D d'une cible thérapeutique commence par choisir la meilleure structure cristallographique disponible dans la PDB. Ce choix s'effectue selon la valeur de la résolution définissant la qualité de la structure. Il est à souligner que les structures 3D des protéines provenant de la PDB nécessitent une préparation préalable avant leur utilisation dans des expériences de docking moléculaire. En effet, une seule chaîne doit être gardée dans le cas des protéines contenant plusieurs sous-unités identiques et ce pour une utilisation plus aisée lors du docking. L'inhibiteur **de référence** et d'autres éléments appelés hétéroatomes (**alpha-l-fucose, le 1,2-Ethandiol et le N-Acetyl-d-glucosamine**) ayant servi à la cristallisation doivent également être supprimés pour ne conserver que la structure seule de la chaîne A de la protéine. Par la suite, les hydrogènes manquants doivent être rajoutés tandis que toutes les molécules d'eau doivent être supprimées sauf ceux faisant partie de la cavité catalytique de l'enzyme. Les états de protonation de chaque résidu dans la cavité de l'enzyme

doivent être vérifiés et l'énergie d'interaction intramoléculaire de la protéine doit être minimisée.

b- Préparation de la chimiothèque

Dans certains cas, il est préférable de procéder à un nettoyage de la chimiothèque avant de lancer les calculs de criblage virtuel. Cette étape consiste à ne garder que les composés présentant une structure adéquate avec un bon potentiel de candidat médicament. Pour ce faire, il est nécessaire d'appliquer, sur la chimiothèque en question, des filtres dits « ADMET ». Les composés de la chimiothèque nettoyé (ou pas) doivent être préparés en générant les différentes formes ionisées et les tautomères car dans les conditions physiologiques, une molécule peut prendre différentes formes en subissant des phénomènes de tautomérie ou d'ionisation.

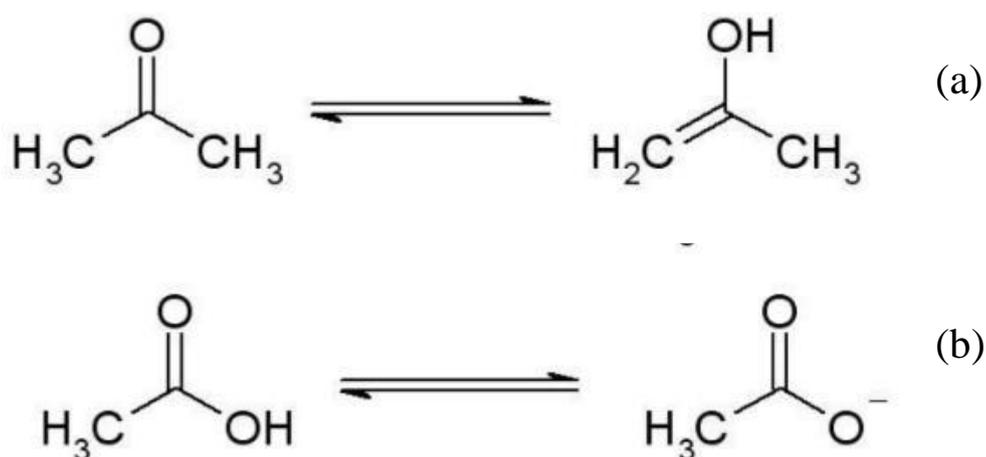


Figure. Exemple de formes chimiques tautomères (a) ou états d'ionisation différents (b) (Empereur-Mot, C. 2017).

Après génération des tautomères et des formes ionisées, les hydrogènes manquants doivent être rajoutés et la géométrie de chaque composé doit être optimisée.

c- Criblage virtuel proprement dit

Le criblage virtuel doit être effectué en utilisant un ou plusieurs logiciels. Malgré l'existence d'un grand nombre de programmes de docking, aucun ne peut être appliquée de façon générale sur la totalité des systèmes biologiques car chacun présente des avantages mais également des imperfections. Ainsi, l'utilisation de plusieurs programmes de docking suivie par l'application d'une méthode consensus est recommandée pour combiner et tirer le meilleur profit des logiciels de docking utilisés dans le cadre d'une même étude. Autrement

dit, cette approche permet de combiner les informations issues des différentes fonctions de score afin de compenser leurs imperfections individuelles ce qui améliore la qualité des résultats obtenus. L'hypothèse sous-jacente est que la probabilité qu'une molécule soit active doit augmenter si cette dernière est associée à de bons scores d'affinité selon plusieurs fonctions de score. De même, la probabilité qu'une molécule fautive/positive soit bien classée doit diminuer car, avec leurs fonctions de score différentes, les logiciels ont peu de chance de commettre la même erreur.

d- Analyse visuelle et sélection des composés à tester expérimentalement

Bien que l'application d'un protocole de criblage virtuel permet de supprimer automatiquement un grand nombre de composés indésirables, on ne peut pas substituer l'intervention humaine. Un biais bien connu, est par exemple le bon classement des ligands de haute masse moléculaire, qui en raison d'un nombre d'atomes plus important, sont tout simplement capables de créer un plus grand nombre d'interactions même-ci au niveau l'entrée voir à l'extérieur de la cavité étudiée. Désormais, l'inspection visuelle du mode d'interaction des composés ayant obtenus les meilleurs scores est une étape primordiale afin de faire ressortir une liste réduite de molécules prometteuses à tester expérimentalement. Lors de cette inspection, plusieurs critères de sélection doivent être pris en considération :

- Positionnement des composés dans la cavité étudiée
- Le mode d'interaction (nombre optimal de liaisons H, interactions ioniques, hydrophobes, Pi-Pi...etc.).
- Interaction avec les résidus clés de la cible.
- Diversité structurale
- Disponibilité des molécules réelles dans la chimiothèque.

7- Succès du criblage virtuel

De très nombreux exemples de réussite de criblages virtuels sont disponibles dans la littérature scientifique. Le tableau 1 récence 31 exemples de médicaments commercialisés entre 1995 et 2009 pour lesquels le criblage virtuel, qu'ils soient « *structure-based* » et/ou « *ligand-based* », a joué un rôle crucial.

Tableau 1. Quelques exemples de médicaments mis sur le marché pour lesquels la mise en œuvre de criblages virtuels à joué un rôle important.

Nom du principe actif	Cible	indication	Référence
Dorzolamide	Anyhdrase carbonique	Glaucome	Baldwin, J.J. et <i>al</i> 1989
Saquinavir	Protéase du VIH	SIDA	Roberts, N.A. et <i>al</i> 1990
Ritonavir	Protéase du VIH	SIDA	Kempf, D. J. et <i>al</i> 1998
Indinavir	Protéase du VIH	SIDA	Dorsey, B.D. et <i>al</i> 1994
Brinzolamide	Anyhdrase carbonique	Glaucome	Silver, L.H. 2000
Nelfinavir	Protéase du VIH	SIDA	Kaldor, S.W. et <i>al</i> 1997
Amprenavir	Protéase du VIH	SIDA	Kim, E.E. et <i>al</i> 2004
Lopinavir	Protéase du VIH	SIDA	Sham, H.L. et <i>al</i> 1998
Zanamivir	Neuraminidase	Grippe	von Itzstein, M. et <i>al</i> 1993
Oseltamivir	Neuraminidase	Grippe	Kim, C.U. et <i>al</i> 1998
Imatinib	Kinase BCR-ABL	Leucémie myéloïde chronique	Zimmermann, J. et <i>al</i> 1997
Gefitinib	EGFR	Cancer du poumon à grandes cellules	Hennequin ,L. F. et <i>al</i> 1999
Atazanavir	Protéase du VIH	SIDA	Robinson, B.S. et <i>al</i> 2000
Fosamprenavir	Protéase du VIH	SIDA	Becker, S. et <i>al</i> 2004
Erlotinib	EGFR	Cancer du poumon à grandes cellules	Pollack, V.A.

			<i>et al</i> 1999
Sorafenib	VEGFR	Cancer du rein	Wilhelm, S.M. <i>et al</i> 2004
Tipranavir	Protéase du VIH	SIDA	Turner, S.R. <i>et al</i> 1998
Udenafil	Phosphodiesterase 5	Dysfonctionnements érectiles	Ahn, B.O. <i>et al</i> 2003
Sunitinib	Kinases variées	Cancer du rein	Mendel, D.B. <i>et al</i> 2003
Darunavir	Protéase du VIH	SIDA	Ghosh, A.K. <i>et al</i> 2007
Vorinostat	Histone désacétylase	Lymphome T cutané	Kelly, W.K. <i>et al</i> 2005
Dasatinib	Tyrosine kinases variées	Leucémie myéloïde chronique	Lombardo, L.J. <i>et al</i> 2004
Nilotinib	Kinase BCR-ABL	Leucémie myéloïde chronique	Weisberg, E. <i>et al</i> 2005
Aliskiren	Rénine	Hypertension artérielle	Rahuel, J. <i>et al</i> 2000
Lapatinib	Kinases variées	Cancer du sein	Rusnak, D.W. <i>et al</i> 2001
Rivaroxaban	Facteur X	Thromboses veineuses	Roehrig, S. <i>et al</i> 2005
Dabigatran	Thrombine	Thromboses veineuses	Sorbera, L.A. <i>et al</i> 2005
Etravirine	Transcriptase inverse	SIDA	Das, K. <i>et al</i> 2004
Pazopanib	Kinases variées	Cancer des ovaires	Sleijfer, S. <i>et al</i> 2009
Ximelagatran	Thrombine	Embolie veineuse	Gustafsson, D. <i>et al</i> 2004
Boceprevir	protéase HCV	Hépatite C	Njoroge, F.G. <i>et al</i> 2008

8- Filtrage ADMET

Les principales raisons de l'échec des molécules en cours de développement d'un médicament sont le manque d'efficacité chez l'homme et les problèmes de pharmacocinétique. Actuellement, plusieurs critères physicochimiques et pharmacocinétiques régissant les propriétés d'absorption, de distribution, de métabolisme, d'excrétion et de toxicité d'une molécule peuvent être estimés à partir de sa structure chimique, ce qui contribue à réduire les échecs dus à ces mauvaises propriétés d'ADME-T.

8-1 Propriétés physicochimiques

8-1-1 Règle de Lipinski

Suite à un criblage virtuel (ou même avant) il est très important d'appliquer la méthode de filtre ADME qui est basée sur la règle de cinq, énoncé en 1997 par Christopher Lipinski. Cette règle a été définie à partir de l'analyse de près de 3500 candidats médicaments ayant passé avec succès les tests cliniques. Selon Lipinski, un composé est capable d'être administré par voie orale s'il remplit au moins trois de ces cinq critères :

- La masse moléculaire doit être inférieure à 500 g/mol.
- Le nombre de donneurs de liaisons hydrogène (somme des groupements OH et NH) doit être inférieur ou égal à 5.
- Le nombre d'accepteurs de liaisons hydrogène (somme des atomes O et N) doit être inférieur ou égal à 10.
- Le nombre de liaisons flexibles doit être inférieur ou égal à 15.
- Le logarithme décimal du coefficient de partage 1-octanol/eau, noté LogP, doit être compris entre -2 et 5.

LogP : est égal au logarithme du rapport des concentrations de la substance étudiée dans l'octanol et dans l'eau ($\text{LogP} = \text{Log } C_{\text{oct}}/C_{\text{eau}}$). Cette valeur permet d'appréhender le caractère hydrophile ou hydrophobe d'une molécule. En effet, une molécule doit, pour parvenir dans l'organisme jusqu'à son lieu d'action, se dissoudre dans des phases aqueuses, traverser des membranes lipidiques, des phases protéiques et osidiques. La solubilité dans l'eau et dans les lipides ainsi que le coefficient de partage jouent dans ce transport, un rôle

fondamental. Une molécule dotée d'une action plus rapide doit avoir un coefficient de partage qui favorise le plus son transport.

Ces paramètres peuvent être évalués en utilisant des serveurs en ligne tel que ; jussieu, monilspiration, SuissADME, PreADMET.

8-1-2 Règle de Veber

Par la suite, d'autres critères ont été mis en place pour compléter et ajuster les règles de Lipinski dans la sélection de composés « drug-like ». Ainsi, Veber et ses collaborateurs (Veber, D.F. et *al* 2002) ont proposé une simplification des règles de « Lipinski » après analyse des données pharmaco-cinétiques chez le rat pour 1100 candidats médicaments. Dans cette nouvelle méthode de prédiction, quatre critères sont pris en considération à savoir :

- L'aire de surface polaire (PSA, polar surface area) doit être inférieure ou égal à 140 Å².
- Le nombre de liaisons flexibles doit être inférieur ou égal à 15.
- Le nombre d'accepteurs de liaisons hydrogènes doit être inférieur ou égal à 10.
- Le nombre de donneurs de liaisons hydrogène doit être inférieur ou égal à 5.

8-1-3 Solubilité dans l'eau

La solubilité dans l'eau est l'un des paramètres les plus importants que doit posséder un candidat-médicament pour atteindre sa cible et ce avec la concentration souhaitée. Elle représente un moyen de référence pour prédire la biodisponibilité et l'absorption des médicaments administrés par voie orale ou à l'usage parentéral.

8-1-4 Accessibilité à la synthèse

Des modifications structurales doivent être menées sur composés actifs lors de leur optimisation afin de les rendre plus actifs, à faciliter leur pénétration en cellules et le cas échéant, à les rendre moins toxiques. Ces modifications structurales ne peuvent pas avoir lieu avec des hits dont la synthèse chimique est très difficile à réaliser. Bien que l'accessibilité synthétique ne soit pas le seul facteur à prendre en compte lors de la définition des priorités pour le suivi, elle constitue l'un des facteurs les plus importants et décisifs dans le succès ou

l'échec de nombreux efforts d'optimisation. L'accessibilité à la synthèse chimique peut être simulée via le serveur SwissADME en valeurs allant de 1 (très facile à synthétiser) jusqu'à 10 (très difficile à synthétiser).

8-2 Propriétés pharmacocinétiques

Outre une affinité qui lui confère une activité biologique, une molécule puissante doit atteindre sa cible dans le corps avec une concentration suffisante et y rester sous une forme bioactive suffisamment longtemps pour que les événements biologiques attendus se produisent. Un médicament, une fois entré dans l'organisme, rencontre une série d'obstacles divers sur son chemin vers la cible. Ces obstacles incluent principalement les barrières buccale, gastrique, intestinale, pulmonaire, hépatique, rénale, hémato-encéphalique et de la peau. Actuellement, de nouvelles approches de prédiction sont mises à disposition via des serveurs en ligne tels que SwissADME (<http://www.swissadme.ch>), et PreADMET (<https://preadmet.bmdrc.kr>). Elles permettent d'évaluer le caractère « *drug-like* » des composés en décrivant plusieurs critères à savoir : l'Absorption gastro-intestinale, perméabilité hémato-encéphalique, l'inhibition des Cytochromes P450 ainsi que la perméabilité cellulaire.

8-3 Toxicité potentielle

La toxicité d'un hit est l'un des paramètres les plus redoutés lors de son optimisation en *lead* (candidat-médicament). À l'heure actuelle, de nombreuses méthodes *in silico* ont été mises en place pour tenter de prédire la toxicité ainsi que les effets secondaires des hits. Ces méthodes présentent non seulement un potentiel économique, mais engendrent également une série d'avantages écologiques : réduction des modèles animaux et essais cliniques risqués. Le serveur PreADMET (<https://preadmet.bmdrc.kr/toxicity>) emploie des modèles QSAR (Quantitative Structure-Activity Relationship) afin de simuler à partir de la structure 3D des hits, la toxicité potentielle selon plusieurs tests à savoir : Le test d'Ames (pouvoir mutagène), la cancérogénicité, la toxicité aigüe... etc.

9- Tests de fiabilité des programmes de docking moléculaire :

Des tests de fiabilité des programmes de docking moléculaire doivent être réalisés après un docking. Actuellement il existe plusieurs tests de fiabilité, les plus fréquemment cités sont :

- L'écart quadratique moyen ou le RMSD (*root mean square deviation*).
- Le coefficient de corrélation linéaire (r).
- L'enrichissement.

9-1 RMSD

Correspond à la moyenne de la déviation de chacun des atomes d'une position d'un ligand simulée (après docking) par rapport à ceux de la pose de référence du même ligand (obtenue par cristallographie). Souvent, les performances d'un programme de docking moléculaire sont évaluées en termes de capacité à reproduire le mieux possible des complexes expérimentaux. Le mieux possible signifie que la valeur du RMSD entre la pose du ligand calculée par le logiciel et la conformation dans le complexe expérimental est la plus petite possible.

La norme actuelle pour évaluer les performances d'un programme de docking est de faire un test à partir de plusieurs centaines de complexes protéines-ligands cristallisés. Selon les références, un programme de docking moléculaire est jugé performant Si les valeurs de RMSD obtenues sont inférieures à 2 Å.

9-2 Le coefficient de corrélation linéaire (r).

Souvent, les ligands obtenus par construction moléculaire (issus de la littérature) ne sont pas co-cristallisés (ne sont pas disponibles dans la PDB). Dans ce cas, nous ne pouvons pas appliquer le test RMSD pour évaluer la fiabilité des programmes quant au docking de ces molécules. L'application d'autres tests de fiabilité devient nécessaire.

Le coefficient de corrélation linéaire permet d'évaluer l'intensité et le sens de la relation linéaire entre deux séries de données provenant de l'échantillonnage de deux variables métriques. Ce coefficient prend des valeurs situées entre -1 et 1. S'il n'y a pas de relation linéaire entre les deux séries de données, le coefficient de corrélation est très proche de zéro, et on dit que les deux variables ne sont pas corrélées. Plus le coefficient de corrélation linéaire s'approche de 1 ou -1 plus il ya une forte corrélation entre les deux variables étudiés.

Pour juger les performances d'un programme de docking moléculaire, ce test consiste à évaluer le degré de corrélation existant entre les scores de docking des ligands simulés avec le logiciel (trentaine) et les valeurs de leur IC50 déterminées expérimentalement.

Plus la valeur absolue du coefficient de corrélation obtenu se rapproche de 1, plus le programme est performant pour simuler les interactions de ce groupe de ligands vis-à-vis la protéine étudiée.

9-3 Enrichissement

Lors du processus du drug-design, notamment lors de la recherche des hits par criblage virtuel, il est important de juger la fiabilité des programmes utilisés. Le test par enrichissement est désormais parmi les meilleurs moyens permettant de juger la fiabilité des résultats issus d'un criblage virtuel.

A l'heure actuelle, plusieurs banques d'évaluation ont été mises en place aidant à la réalisation du test d'enrichissement pour évaluer un criblage virtuel. Celle la plus utilisée est connue sous le nom DUD-E (*Directory of Useful Decoys Enchanced*) comprenant des molécules actives vis-à-vis de 102 protéines. (chimiothèques d'évaluation).

Le test par enrichissement consiste à cribler virtuellement une chimiothèque totale en réalité constituée de deux chimiothèques combinées :

- une chimiothèque d'évaluation contenant une centaine de molécules réellement actives sur la cible en question
- la chimiothèque de molécules à cribler sur la cible étudiée contenant des milliers de molécules différentes dont l'activité envers la cible est inconnue.

Autrement dit, la chimiothèque appropriée pour le criblage est enrichie au préalable par des molécules réellement actives avant de lancer le criblage.

Après un criblage virtuel d'une chimiothèque donnée, avec laquelle intégrée une chimiothèque d'évaluation, il est possible de classer les composés selon leurs scores. A partir de ces données, des courbes d'enrichissement (encore appelées courbes d'accumulation) sont déduites, représentant le pourcentage de molécules biologiquement actives récupérées durant le criblage en fonction de leur classement sur l'ensemble total des molécules criblées. Selon les auteurs, un programme de criblage virtuel est aussi performant s'il est capable de bien

prédire les molécules actives en les classant dans les premières portions du classement. De ce fait, l'allure de la courbe d'enrichissement doit être hyperbolique au-dessus du *random* ($f(x)=x$).