

Cours de bioinformatique (Par Mlle Ilham Boulhissa):

On regroupe sous le terme de **bioinformatique** un champ de recherche multi-disciplinaire où travaillent de concert biologistes, informaticiens, mathématiciens et physiciens, dans le but de résoudre un problème scientifique posé par la biologie en énonçant des hypothèses (à valider expérimentalement). Tout comme l'informatique peut être considérée, suivant les situations, comme une science ou comme une technologie, le terme bioinformatique peut décrire toutes les applications informatiques résultant de ces recherches. Cela va de la modélisation moléculaire, à la reconstruction d'arbres phylogénétiques (phylogénie) en passant par l'analyse des séquences nucléiques et/ou protéiques.

Autrement dit, la bioinformatique est constituée par l'ensemble des concepts et des techniques nécessaires à l'interprétation de l'information génétique (séquences) et structurale (repliement 3D). C'est le décryptage de la " bio-information ". La bioinformatique est donc une branche théorique de la biologie.

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

La bioinformatique des séquences : traitant les données issues de l'information génétique contenue dans les séquences nucléiques (ADN et/ou ARN) ou protéiques, et organisant cette énorme masse d'information et la rendre disponible à l'ensemble de la communauté des chercheurs grâce à différentes bases de données, accessibles en lignes, comme GenBank, UniProt, PDB,...*etc.*

L'alignement de séquences permet de trouver les ressemblances entre deux séquences et déterminer leurs éventuelles homologues. Les alignements sont à la base de la construction de parentés (arbres phylogénétiques) suivant des critères moléculaires, ou encore de la reconnaissance de motifs particuliers dans une protéine à partir de la séquence de celle-ci.

La bioinformatique structurale : La modélisation moléculaire permet une construction intellectuelle d'un modèle moléculaire à partir des données indirectes issues principalement d'analyses cristallographiques (étude des figures de diffraction des rayons X par un cristal), ou de résonance magnétique nucléaire (RMN). Le modèle moléculaire obtenu ensuite est un ensemble de coordonnées atomiques dans l'espace. L'informatique intervient dans toutes les

étapes conduisant de l'expérimentation au modèle, puis ensuite dans l'analyse du modèle par la visualisation moléculaire (voir les protéines en 3D).

Elle est utilisée par exemple pour étudier les sites actifs d'une enzyme, mettre au point informatiquement une série d'inhibiteurs possibles pour cette enzyme, et ici on parle de la méthode de docking moléculaire qui permet la simulation (prédiction) du mode d'interaction le plus favorable d'un ligand au sein de son récepteur (protéine ou ADN). Cette discipline 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 basées 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. La découverte des médicaments commence par l'identification de la cible thérapeutique, suivie par l'identification des composés « Hits » par le criblage à haut débit « HTS » et / ou criblage *in silico –in vitro*, la transformation de Hits en Leads et leur optimisation, puis les essais précliniques et cliniques jusqu'au médicament approuvé. Ce parcours prend d'une moyenne d'environ 12 à 14 ans et un coût global estimé est d'au moins 800 millions de dollars.

Identification et validation d'une cible thérapeutique

On doit identifier une cible thérapeutique et la valider en vérifiant sa relation avec une pathologie par des expériences. Les cibles thérapeutiques sont souvent des protéines dont l'activation ou l'inhibition donne un effet thérapeutique vis-à-vis d'une maladie.

Identification de Hits

Cette étape consiste à identifier des molécules appelées « Hits » qui sont des molécules ayant la capacité d'interagir avec la cible choisie et susceptibles de moduler ses effets sur le processus biologique en question.

En premier temps, l'industrie pharmaceutique a appliqué la méthode de criblage à haut débit (« HTS » en anglais) ou le criblage expérimental pour identifier des molécules possédant une activité élevée envers la cible, cette méthode permet de tester des milliers de composés

chimiques divers envers des milliers de cibles potentielles pour identifier une nouvelle combinaison médicament-cible.

Dans un deuxième temps, l'industrie pharmaceutique a intégré une méthode informatique « criblage virtuel », qui est couramment utilisée de façon alternative ou complémentaire au criblage à haut débit, 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, et élaborer enfin un nombre réduit de molécules adressées au test *in vitro*, ces molécules sont très spécifiques à la cible. L'avantage de la technique est de traverser l'espace chimique en explorant de nouvelles molécules actives et en gagnant du temps et du coût.

Optimisation des Hits en Leads :

L'optimisation des candidats Hits concernant la puissance, la sélectivité et les propriétés ADMET (absorption, métabolisme, excrétion et toxicité) est d'une importance centrale dans le processus de découverte de médicaments (Venkatesh & Lipper, 2000). Le manque d'efficacité (29%) et la faible biodisponibilité et toxicité (39%) représentent la majorité des candidats médicaments défaillants (Venkatesh & Lipper, 2000, Lipinski *et al.*, 2001). L'optimisation des Hits est rendu possible 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 de celles d'un médicament. Cette étape est souvent prise en charge à l'aide de techniques de docking, par exemple l'analyse du mode de liaison et la prédiction ADMET *in silico*.

Tests de sécurité pré-cliniques.

Les molécules optimisées seront testée *in vivo* sur des animaux, afin de prouver leur efficacité et leur innocuité *in vivo* c.-à-d. ayant un bon profil ADME-Toxicité chez l'animal.

- ✚ 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)

Phases cliniques :

Beaucoup de candidats médicaments sont éliminés avant ce stade atteint par seulement un médicament sur quinze. Ces tests sont menés sur l'homme en trois phases principales :

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.

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.

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. Criblage virtuel :

Le criblage virtuel est un ensemble de techniques computationnelles ayant pour objectif l'exploration de bases de composés à la recherche de molécules d'intérêt. Il englobe deux familles, le criblage virtuel « structure-based » et le criblage virtuel « ligand-based ».

-On parle du **criblage basé sur la structure « structure-based »** si la structure tridimensionnelle de la protéine est connue (structure-based design), l'amarrage (encore appelé docking en anglais) peut être utilisé pour prédire l'affinité des ligands pour un récepteur donné.

Cette stratégie de criblage permet potentiellement d'identifier de nouvelles classes de molécules actives, elle peut être réalisée en 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.

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, qui se réfère à la construction d'un modèle d'une protéine « cible » à partir de sa séquence d'acides aminés et d'une structure expérimentale tridimensionnelle d'une protéine homologue.

-On parle du **criblage basé sur le ligand « ligand-based »**, une méthode alternative qui peut être envisagée, si aucune donnée de protéine n'est disponible. Par conséquent, on utilise les ligands connus comme guides pour définir les propriétés responsables de leur affinité pour la cible et identifier des molécules similaires. Le criblage virtuel basé sur le ligand, également connu sous le nom de recherche de comportement de voisinage, a tendance à trouver des résultats étroitement liés aux produits actifs connus, dans lesquels des bases de données de structures chimiques sont recherchées pour trouver des composés similaires à des actifs connus (recherche de similitude) ou possédant un pharmacophore ou sous-structure en commun avec un actif connu (recherche de sous-structure de pharmacophore). Une recherche de sous-structure de similarité peut être effectuée en se référant à la structure 2D ou 3D de divers composés à cribler. Méthodes basées sur les ligands basées sur le principe de similitude, c'est-à-dire que des composés similaires sont supposés produire des effets similaires. Dans ce cas, si un ou plusieurs produits chimiques actifs sont connus, il est possible de rechercher dans une base de données des molécules similaires mais plus puissantes.

Tableau 1: Classification des méthodes de criblage virtuel

| | Ligand (s) connus | Ligands inconnus |
|---|---|---|
| Structure connue de la cible ou d'un homologue proche | Criblage virtuel basé sur la structure : Docking protéine-ligand | Criblage virtuel basé sur la structure <i>de novo</i> : Docking protéine-ligand |
| Structure de cible inconnue | Criblage virtuel basé sur le ligand : Quelques ligands : recherches de similarité Plusieurs ligands : recherches de pharmacophore | Criblage virtuel ne peut pas être appliqué |

Aujourd'hui, les techniques de criblage virtuel sont devenues des outils indispensables de la chimie médicinale et sont utilisées de manière quotidienne aussi bien dans les laboratoires de recherche publics que dans les grands laboratoires pharmaceutiques.

L'amarrage moléculaire ou le docking moléculaire est la méthode de SBVS la plus largement utilisée qui recherche un ligand pour s'adapter au site de liaison d'une protéine cible.

3. Docking moléculaire

Le terme de docking moléculaire signifie une méthode qui permet d'évaluer l'affinité entre deux entités virtuellement, en faisant des calculs menées par des logiciels spécialisés, et qui se subdivise en deux étapes :

La première étape est de déterminer le mode d'interactions d'un complexe formé de ces deux entités, en cherchant les orientations dans l'espace et des conformations favorables pour la fixation d'un ligand à un récepteur, et par la suite de trouver les interactions biochimiques favorables dans le complexe. Le logiciel de docking travaille avec des algorithmes de recherche pour traiter la flexibilité d'un ligand.

La deuxième étape est la prédiction de l'énergie d'interaction entre les deux entités en utilisant des fonctions de scores.

Donc, le programme de docking moléculaire cherche le mode de liaison le plus approprié du ligand avec le site de liaison de la structure protéique, avec l'énergie d'interaction la plus faible.

3.1. Applications de docking moléculaire

Le docking moléculaire peut être utile dans l'étude et l'analyse des interactions des complexes (protéine-ligand) ou encore protéine-médicament. Il amène aussi à l'identification des hits principal lorsque seule la structure d'une cible et son site de liaison sont disponibles, et ensuite le criblage des chimiothèques ou libraires de composés chimiques. Par ailleurs, ça peut être un outil d'optimisation de leads lorsque les modifications des structures actives connues peuvent être rapidement testées dans des modèles informatiques avant la synthèse des composés.

3.2. Principaux programmes de docking moléculaire

Plus de 60 programmes de docking et plus de 30 fonctions de score ont été introduits. Les principaux logiciels d'amarrage sont listés dans le tableau suivant:

Tableau 2: Principaux programmes de docking moléculaire et leurs sites internet.

| Nom du programme | Site Internet |
|------------------|---|
| AutoDock | http://www.scripps.edu/mb/olson/doc/autodock/ |
| Dock | http://dock.compbio.ucsf.edu/ |
| FlexX | http://www.biosolveit.de/FlexX/ |
| Fred | http://www.eyesopen.com/products/applications/fred.html |
| Glide | http://www.schrodinger.com/Products/glide.html |
| Gold | http://www.ccdc.cam.ac.uk/products/life_sciences/gold/ |
| ICM | http://www.molsoft.com/products.html |
| LigandFit | http://www.accelrys.com/cerius2/c2ligandfit.html |
| Surflex | http://www.biopharmics.com/products.html |

3.3. Types de docking

1/ Le **docking rigide** a été utilisé par les premiers logiciels de docking développés, il se base sur le modèle « clé-serrure », le récepteur et le ligand sont maintenus fixes (rigides) et l'amarrage est exécuté.

2/ Plus tard, le **docking semi-flexible** a permis d'obtenir des résultats plus précis, le récepteur est maintenu rigide alors que le ligand est flexible.

3/ Aujourd'hui, les chercheurs sont arrivés à appliquer le **docking flexible** qui considère le ligand et le récepteur flexibles sur le plan de la conformation.

3.4. Les algorithmes de recherche

Il existe trois types d'algorithmes pour le traitement de la flexibilité du ligand : les méthodes systématiques, les méthodes stochastiques ou aléatoires, et les méthodes de simulation.

3.4.1. Recherche systématique

Ces méthodes ont pour but d'explorer la totalité des degrés de liberté du ligand. Pour cela, toutes les torsions du ligand subissent des rotations de 0 à 360°, en utilisant un pas incrémental fixe, afin de générer la totalité des conformations possibles. Rapidement, ceci provoque une augmentation très importante du nombre de conformations créées que l'on nomme explosion combinatoire, c'est pourquoi ces techniques ne sont pas employées dans leur forme originelle, mais en association avec un certain nombre de contraintes destinées à réduire le nombre de solutions.

Les méthodes de reconstruction incrémentale sont les algorithmes de recherche systématique les plus couramment appliqués. Celles-ci séparent la molécule en fragments rigides, et arriment ces fragments de façon indépendante avant de faire grandir étape par étape le ligand dans le site actif en rajoutant les différents fragments sur le premier fragment arrimé qui sert de base à la reconstruction. Ce type d'algorithme est notamment employé dans les programmes FlexX et les versions flexibles du programme Dock. Certains programmes utilisent des ensembles de conformations préétablies afin de contrer le phénomène d'explosion combinatoire, comme FLOG.

3.4.2. Recherche aléatoire

Ces méthodes fonctionnent en crossovers) opérant des changements aléatoires sur un ligand ou une population de ligands, lesquels sont acceptés ou rejetés sur la base d'une fonction de probabilité « fitness ».

Les algorithmes génétiques utilisent des procédés inspirés de la théorie de l'évolution. En partant d'une population de conformations définies par un ensemble de variables

assimilées à des gènes, ils appliquent des opérateurs génétiques (mutations, permettant de faire varier les conformations jusqu'à l'obtention d'une population finale qui satisfait la fonction de « *fitness* » choisie. Ce type d'algorithme est utilisé notamment dans les programmes Gold et Autodock. L'algorithme de recherche tabou explore l'espace conformationnel en interdisant les espaces déjà explorés et accepte les nouvelles solutions sur la base de leur RMSD par rapport aux précédentes conformations. Il est utilisé par exemple dans le programme PRO_LEADS.

3.4.3. Méthode de simulation

Ces méthodes sont basées sur la résolution des équations du mouvement de Newton. Elles comprennent notamment les techniques de dynamique moléculaire et des algorithmes de minimisation. Les premières ne sont jamais utilisées pour générer de la flexibilité sur le ligand car elles demandent un temps de calcul non compatible avec la gestion de bases de molécules. Les secondes en revanche sont parfois utilisées dans les programmes de docking, en complément d'un autre algorithme de recherche, afin d'atteindre une conformation de basse énergie. C'est le cas par exemple dans les versions récentes du programme Dock, et dans le programme ICM.

4. Les interactions protéine-ligand :

La connaissance des données structurales et des mécanismes de liaison protéine-ligand pour optimiser le processus de recherche de nouveaux médicaments, une compréhension approfondie de la nature de la reconnaissance / interaction moléculaire est également d'une grande importance pour faciliter la découverte, la conception et le développement de médicaments.

4.1. La liaison hydrogène

La liaison hydrogène est une liaison chimique non covalente de type dipôle-dipôle entre deux molécules (intermoléculaire) ou entre deux groupements d'une molécule (intramoléculaire). Elle consiste essentiellement dans l'interaction entre deux molécules dont l'une possède un atome donneur d'électrons (O, N, F) et l'autre possède un atome H accepteur

d'électrons (OH, NH₂). Ce type de liaison résulte d'un transfert partiel d'un électron célibataire sur le groupement H.

Dans certains cas particuliers, le groupe C-H peut se comporter comme donneur de liaisons hydrogène. Cependant, les liaisons hydrogène impliquant ces atomes sont souvent très faibles. Les liaisons hydrogènes sont plus fortes que les interactions de Van der Waals ; leurs énergies sont estimées entre 3 et 9 kcal/mol.

4.2. Les interactions de Van der Waals

Ces forces résultent de l'interaction des nuages électroniques d'atomes ou de molécules proches les uns des autres. Ces nuages fluctuent, soit de façon permanente, c'est le cas dans les molécules polaires qui se comportent comme un dipôle permanent, soit de façon temporaire, sous l'action d'un champ électrique induit par le voisinage d'un ion, d'une molécule faisant apparaître un moment dipolaire dit induit.

Elles reposent sur les interactions entre les dipôles constitués par les molécules. On distingue trois types de ces interactions :

- Interaction dipôle permanent – dipôle permanent ou effet d'orientation de Keesom : Ce type d'interaction se développe entre deux molécules polaires.
- Interaction dipôle permanent – dipôle instantané ou effet d'induction de Debye : Ce type d'interaction se développe entre une molécule polaire et une molécule quelconque (polaire ou apolaire).
- Interaction dipôle instantané – dipôle instantané ou dispersion de London : Ce type d'interaction se développe entre deux molécules quelconques (polaires ou apolaires).

Ces interactions sont très faibles mais dans le cas des macromolécules, leur nombre élevé va produire au totale une force importante.

4.3. Les Interactions électrostatiques

Les interactions électrostatiques s'établissent entre des molécules portant des charges électriques opposés. Elles peuvent être de type ion-ion, ion-dipôle ou dipôle-dipôle.

4.4. L'effet hydrophobe

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

5. Les outils de criblage virtuel (docking moléculaire) :

5.1. La structure 3D de la protéine cible:

Il est possible de télécharger la structure 3D des différentes protéines à partir de la banque de donnée PDB (« *Protein Data Bank* », en anglais), qui renferme en plus des structures protéiques, des structures d'ADN et ARN. Ce sont des structures élaborées par différents techniques tels que la cristallographie au rayon X, RMN, et microscopie électronique. La structure protéique peut être téléchargée sous différents formats : PDB,

Si aucune donnée disponible de la structure protéique, on suit la modélisation par homologie.

5.2. La chimiothèque (ligands) :

La chimiothèque est une banque de données de molécules. Elle peut contenir de plusieurs dizaines à plusieurs millions de composés chimiques. Il y a deux façons pour ramener la chimiothèque à cribler :

1/ Utiliser une chimiothèque téléchargée à partir de leur site internet à condition, d'être prête pour le criblage, sinon une étape de sa préparation doit être effectuée.

2/ Construire une chimiothèque en utilisant des programmes de construction moléculaire comme : Arguslab, Titan, Sybyl (payant), Marvin, ChemDraw...etc. Après le dessin des molécules et leur minimisation d'énergie, on peut les enregistrer sous différents formats : SMILES, SDF, MOL2 et PDB.

Tableau 3: Les bases de données chimiques généralement examinées.

| Chimiothèque | Nombre de composés | Lien |
|--------------|--------------------|---|
| ZINC | ≈ 35 million | https://zinc.docking.org/ |
| eMolecules | ≈ 7 million | https://www.emolecules.com/ |
| ChemSpider | ≈ 58 million | http://www.chemspider.com/ |
| ChEMBL | ≈ 2 million | https://www.ebi.ac.uk/chembl/ |
| PubChem | ≈ 92 million | https://pubchem.ncbi.nlm.nih.gov/ |
| ChemDiv | ≈ 1.6 million | https://www.chemdiv.com/ |
| Enamine | ≈ 1.7 million | https://enamine.net/ |
| ChemBridge | ≈ 1 million | https://www.chembridge.com/ |

5.3. Programme de docking moléculaire :

Il faut choisir un programme de docking moléculaire adapté avec la cible en question, en plus un test de fiabilité est recommandé avant d'entamer les calculs.

5.3.1. Test de fiabilité du programme de docking moléculaire :

1. Test RMSD (Root mean square deviation).

Souvent les performances d'un programme de docking moléculaire sont évaluées en termes de capacité à reproduire le mieux possible des complexes expérimentaux. Cette capacité est habituellement jugée au moyen de la déviation quadratique moyenne ou RMSD qui représente la moyenne des différences de position du ligand simulée par le programme vis-à-vis de celle déterminée expérimentalement existant au niveau de la PDB. La prédiction est jugée correcte lorsque la valeur du RMSD obtenue est inférieure ou égale à 2 Å. Ce test peut être réalisé sur une centaine de complexes protéine-ligand choisis à partir de la PDB (de préférence de la cible), en faisant le docking de chaque ligand au sein de sa protéine native avec notation du pourcentage des complexes qui ont un $\text{RMSD} \leq 2\text{\AA}$.

2. Test du coefficient de corrélation linéaire (R) :

Ce test consiste à déterminer le degré de corrélation entre l'activité biologique de l'inhibiteur (IC_{50}) et son énergie d'interaction avec l'enzyme. Le coefficient R est calculé en traçant la droite de régression linéaire entre les valeurs de l'énergie d'interaction (ΔG) des inhibiteurs des protéines cibles générées par le programme de docking et les valeurs de leur

log IC50. La valeur du coefficient de corrélation est située entre -1 et 1. Plus le coefficient de corrélation est proche des valeurs extrêmes -1 et +1, plus la corrélation linéaire entre les variables est forte.

3. Test d'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.

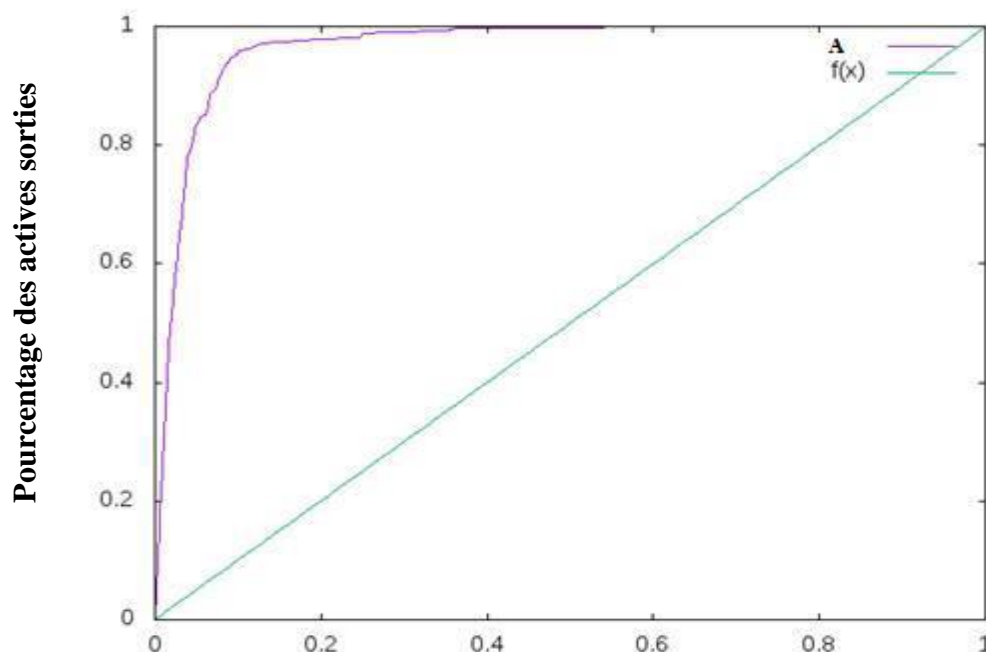
A l'heure actuelle, plusieurs banques d'évaluation ont été mises en place aidant à la réalisation du test d'enrichissement pour évaluer un criblage virtuel. Celle la plus utilisée est connue sous le nom DUD-E (*Directory of Useful Decoys Enchanced*) comprenant des molécules actives vis-à-vis de 102 protéines.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



Classement des molécules de la chimiothèque totale

Courbe 1 : courbe d'enrichissement représentant 100% de la chimiothèque totale

5.4. Programmes de construction moléculaire :

Parmi les programmes de construction moléculaire générant les structures 3D des molécules, on peut citer: Arguslab, Titan, Sybyl (payant), Marvin, ChemDraw...etc. Après le dessin des molécules et la minimisation d'énergie de leur géométrie, on peut les enregistrer sous différents formats : SMILES, SDF, MOL2 et PDB.

5.5. Programmes de visualisation :

On l'utilise dans l'étape de préparation de la protéine et la chimiothèque de criblage virtuel, il renferme des options pour ajuster leurs structures, par exemple l'option de supprimer les molécules indésirables dans la structure protéique (molécules d'eau), ou d'ajouter des hydrogènes qui manquent dans la structure.

Il Permet aussi la visualisation des poses issues de docking moléculaire en indiquant les interactions établies entre une protéine et un ligand.

On peut citer parmi ces programmes: Viewerlite, VMD, Biovia Discovery Visualizer, Maestro....etc

5.6. Programmes de conversion des formats :

Ces programmes deviennent indispensables lorsque le format de protéine ou du ligand est incompatible avec celui qui est demandé par le logiciel de docking, à titre d'exemple le logiciel Open babel.

6. Les étapes de criblage virtuel :

6.1. Préparation de la cible :

Cette étape se traduit par des modifications apportées à la structure protéique :

- Garder une seule chaîne parmi les chaînes identiques de la protéine.
- Ajouter les atomes d'hydrogène manquants dans la structure.
- Éliminer les molécules hétérogènes (molécules utilisées dans la cristallographie) qui ne font pas partie de la protéine.
- Éliminer les molécules d'eau, sauf celles qui sont présentes dans le site actif de la protéine.

6.2. Préparation de la chimiothèque :

La préparation de la chimiothèque renferme aussi l'ajout des atomes d'hydrogène manquants, la génération des tautomères ainsi que de prendre en compte l'état d'ionisation de chaque molécule de la chimiothèque.

Il est préférable aussi de faire une filtration de la chimiothèque selon des propriétés « drug like » pour éviter un éventuel échec au cours de l'étude in vivo du profil ADME-T.

6.3. 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 (molécule active *in silico*, inactive *in vitro* ou *in vivo*) soit bien classée doit diminuer car, avec leurs fonctions de score différentes, les logiciels ont peu de chance de commettre la même erreur.

6.4. Analyse visuelle et sélection des composés à tester expérimentalement

Un biais bien connu, est par exemple le bon classement des ligands avec un poids moléculaire élevé, 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, ...etc.).
- Interaction avec les résidus clés de la cible.
- Diversité structurale
- Disponibilité des molécules réelles dans la chimiothèque.

7. Etude *in silico* des propriétés ADME-Toxicité

Les indicateurs physico-chimiques sont de plus en plus utilisés au cours des premières étapes de la découverte de médicaments pour fournir une compréhension globale des propriétés clés qui affectent les fonctions biologiques (ADME).

De nombreuses définitions des propriétés des composés « drug like » ont été publiées, sur la base d'une analyse des propriétés moléculaires simples de médicaments efficaces. Celles-ci sont généralement présentées comme des règles qui indiquent quand les propriétés d'un composé diffèrent amplement de celles de la majorité des médicaments, ce qui peut indiquer un risque élevé de mauvais résultats pour la pharmacocinétique ou l'innocuité *in vivo*. Alternativement, les composés peuvent être classés en fonction de leur similitude avec les

médicaments commercialisés en utilisant une mesure continue de ressemblance médicamenteuse « drug-likeness ». Une biodisponibilité orale élevée est souvent une considération importante pour le développement de molécules bioactives en tant qu'agents thérapeutiques. Les paramètres pharmacocinétiques *in vivo* sont fortement influencés par les propriétés physicochimiques d'un médicament.

La première analyse approfondie des propriétés ADME a été réalisée par Lipinski (Lipinski *et al.*, 1997), et a abouti à la fameuse «règle de 5», qui fait valoir qu'une bonne absorption ou perméation est plus probable lorsque:

- ✓ Le poids moléculaire (MW) ≤ 500 g/mol
- ✓ Le nombre de donneurs de liaison hydrogène (HBD) ≤ 5 (en comptant la somme de tous les groupes NH et OH)
- ✓ Le coefficient de partage octanol / eau Log P ≤ 5
- ✓ Le nombre d'accepteurs de liaison hydrogène (HBA) ≤ 10 (en comptant tous les N et O atomes).

LogP : est égal au logarithme du rapport des concentrations de la substance étudiée dans l'octanol et dans l'eau (LogP = Log Coct/Ceau). 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.

On considère toute molécule qui remplit au moins trois de ces critères respecte la règle.

Ainsi Veber (Veber *et al.*, 2002) a ajouté deux autres critères (règle de Veber):

- ✓ Le nombre de liaisons rotatives (NBR) < 10
- ✓ La surface topologique polaire « *Topological Polar Surface Area* » (TPSA) $\leq 140 \text{ \AA}^2$

En sachant que le TPSA peut être obtenu en calculant les surfaces totales occupées par les atomes d'oxygène et d'azote et l'hydrogène attachés à ces molécules. Cela met en évidence la relation directe du potentiel de la liaison hydrogène avec la valeur TPSA du composé.

La perméabilité et la solubilité du médicament sont des paramètres critiques qui affectent le profil pharmacocinétique et pharmacodynamique du médicament, à partir du site d'administration, de l'absorption dans la circulation systémique, des mouvements dans le sang et de l'excrétion.

Ces paramètres peuvent être évalués en utilisant des serveurs en ligne tels que : Molinspiration, SwissADME.

D'autres filtres pharmacologiques peuvent être mis en jeu pour étudier le profil ADME, rajoutés aux règles précédentes :

- ✓ La pénétration de la barrière hémato-encéphalique (BHE).
- ✓ L'absorption gastro- intestinale (AGI).
- ✓ Les enzymes du cytochrome sont principalement impliquées dans le métabolisme du médicament pour l'élimination et / ou la biotransformation. Des interactions médicamenteuses majeures sont signalées en raison de l'inhibition des enzymes CYP et, par conséquent, la co-administration du médicament peut s'accumuler à un niveau toxique en raison de l'inhibition des enzymes CYP.

On peut calculer ces paramètres ainsi d'autres paramètres de toxicité comme le test d'Ames (pouvoir mutagène), la cancérogénicité, la toxicité aigüe...etc, en consultant les serveurs SwissADME, et PreADMET.

Lexique : (pour savoir)

Minimisation d'énergie : Le principe est de déplacer les atomes afin d'obtenir la conformation la plus stable possible (qui possède une énergie plus basse que le point de départ).

Dynamique moléculaire :

Dynamique moléculaire : est une technique de simulation numérique permettant de modéliser l'évolution d'un système de particules au cours du temps

En pratique, la dynamique moléculaire consiste à simuler le mouvement d'un ensemble de quelques dizaines à quelques milliers de particules dans un certain environnement (température, pression, champ électromagnétique, conditions aux limites, ...) au cours du temps, dans le cadre de la mécanique newtonienne.