

Rational drug design

Introduction

Drug = petite molécule qui lie, l'agit et modifie l'action de R spécifique

Récepteurs : la + part du temps des protéines qui lient et l'agissent acc d'6 molécules pour réaliser de nombreuses fonctions requises pour maintenir la vie. Peut-être à l'intérieur ou à la surface

Differentes étapes sont requises avant la découverte d'une molécule active

- Identifier la molé cible : celle-ci est un récepteur biologique impliquée dans un mécanisme biologique que l'on essaye d'activer, d'inhiber ou de moduler. La molé cible peut être protéine, gène ou ARN. La cible doit être efficace, sûre, être en accord avec les besoins cliniques et commerciaux et être "druggable".

Druggable : la cible doit être accessible à la drogue potentielle et modifiable/celle-ci et la liaison de la drogue doit entraîner une réponse biologique mesurable *in vitro* et *in vivo*.

Pour identifier une molé cible : - identifier une molécule cible pertinente pathologiquement

- déterminer sa séquence
- élucider la fonction et les mécanismes de la protéine
- prouver le concept thérapeutique dans les organismes

Molécules cibles et mécanismes d'action des drugs

- Enzymes affectées par des inhibiteurs réversibles et irréversibles. Biocatalyste endogène qui convertit ou plus substrats en un ou plus produits
- Récepteurs affectés par des agonistes ou des antagonistes. R = une protéine soluble, liée à la RH ou un complexe protéique qui a un effet physio via plus étape après avoir lié un agoniste. Agoniste = un ligand qui medie la réponse d'un R (effet intrinsèque). Antagoniste = un ligand qui empêche l'action d'un agoniste de manière directe (compétitive) ou indirecte (allostérique)
- Agonist partiel : un antagoniste à haute affinité qui a plus ou moins d'activité.
- Canaux ioniques affectés par des bloqueurs ou des "openers".
- Transporteurs affectés par l'assimilation d'inhibiteurs. Un inhibiteur est un ligand qui empêche la liaison d'un substrat à son enzyme de manière directe (compétitif) ou indirecte (allostérique), réversible ou irreversible.
- ADN affecté par les agents alkylant, intercalant, liant le sillon minime

Rational drug design

Découverte de médicaments

But : découvrir des "hits" qui pourraient devenir des "leads". Un "hit" est une molé qui déclenche une interact° avec le R

Les méthodes expérimentales utilisées à large échelle dans l'industrie pharmaceutique sont le criblage haut débit, la chimie combinatoire et les méthodes assistées par ordinateurs

- △ A prendre en compte pour améliorer le design des drug et la stratégie
1. Les propriétés pharmacodynamiques : permettent d'optimiser l'interact° de la drug avec sa cible. Détermine ce que la drug fait au corps.
 2. Propriétés pharmacocinétiques : Détermine ce que le corps fait à la drug. permet d'améliorer la capacité de la drug à atteindre sa cible et de lui conférer un temps de demi-vie acceptable.

Lead : Propriétés requises

- Doit avoir certaine acti biologique désirable (m si faible et non sélective)
- Doit exister des analogues s'y rapportant. Indiquant que les modifications structurelles vont moduler l'activité biologique et d'autres propriétés
- Sa structure ne peut être extrêmement polaire ou lipophile (pt causer prob d'1 pt de rme de la biodisponibilité)
- Ne peut contenir de groupes toxic ou de groupes produisant des métabolites toxics

Etude du lead trouvé

Etude du SAR : composant synthétique pour lequel un groupe fonctionnel particulier de la molécule est retiré ou altéré.

Le but est de découvrir les groupes de la molécule qui sont essentiels pour l'activité biologique et ceux qui ne le sont pas

Differents groupes fonctionnels retrouvés dans les leads

- Alcool, amines, esters, amides, acides carboxyliques, cétones peuvent agir avec des sites de liaisons via ponts hydrogènes
- Amines et acides carboxyliques peuvent agir par lien ionique
- Alkènes et cycles aromatiques agissent via interact° de Van der Waals
- Substituts alkyles et le squelette carbonés peuvent agir avec les régions hydrophobiques du site de liaison
- Les groupes réactifs fonctionnels comme les halogénés alkyles peuvent faire des liens covalents irréversibles (et le lead et la cible)

RDD

Découverte de médicaments

Lead : optimisation

- Structure doit avoir un certain poids moléculaire et appartenir à un certain range de lipophilicité

Recommandations: Poids moléculaire < 350 (+ 1 molé est grande + elle passera difficilement la mu)

• Lipophilicité exprimée par $\log P \leq 3$. Le $\log P$ est une mesure de la partition de la molé (elle use octanol et eau)

$P = m\text{-octanol/water partition coeff}$ positif et élevé = + soluble ds octanol que ds eau

• Les 5 règles de Lipinski: drug doit avoir un poids moléculaire < 500, lipophilicité $\log P \leq 5$, pas plus de 5 liens hydrogènes donneurs, pas plus de 10 N et O dans la molécule

! Il y a un risque élevé de mauvaise biodisponibilité si 2 ou + de ces conditions ne sont pas respectées.

Mais il existe des exceptions pour lesquelles de mauvais leads peuvent être optimisés pour devenir des drugs valables.

→ L'optimisation de la structure d'un lead est une procédure évolutive. Chaque amélioration mineure ou majeure de certaines propriétés amène à un nouvel analogue qui peut être optimisé jusqu'à ce que le dernier candidat possède toutes les propriétés désirées pour commencer son investigation clinique.

Recherche via "Me Too Research"

Principe: copie de drug déjà existante avec quelques variations mineures.

Permet d'améliorer: biodisponibilité, la largeur du spectre, contourner une résistance, de diminuer les effets secondaires (p.e. Antihistaminique), d'ajouter la spécificité. Méthode de moins en moins utilisée car les brevets sont de + en + limitant

RDD

Recherche de leads via méthodes expérimentales : HTS et chimie combinatoire

Drug design qui favorise l'approche des grands nombres

But : accélérer le processus de découverte en ciblant une grande librairie de composants à une vitesse \rightarrow milliers de composants/jour ou semaine

High Throughput Screening (HTS) (1997)

Utilise des robots pour cibler une grande librairie de composants sur une protéine cible isolée, une cellule ou un tissu. Afin d'identifier les molécules capables de les lier.

MAIS : - Nécessite l'utilisation d'automates complexes de labo, sans connaissance préalable de la nature du type clinique qui pourrait avoir une activité à la protéine cible.

\exists les analyses biochimiques et les analyses basées sur la gène + partielles des analyses biochimiques, principalement inhibition d'enzyme et analyse de la liaison du ligand au récepteur. Limitation car toutes les cibles ne peuvent être purifiées.

- A prendre en compte :
- la pertinence pharmacologique de l'analyse
 - la reproductibilité
 - le coût
 - la qualité de l'analyse
 - les effets des composants

\rightarrow la chimie ne sait pas suivre le HTS \Rightarrow développement de la chimie combinatoire

Chimie Combinatoire

Synthèse d'un grand nombre de composants. Bibliothèque combinant de manière systématique les représentations de 2 ou plus de famille de composants (p.e. aldéhydes +, ac amines +, acides carboxyliques +)

Cependant assay décevant : souvent les composants trouvés sont nombreux mais ils ne sont pas drug like (trop grand ou lipophile)

Evolution vers une chimie combinatoire + restrictive qui prend en compte les règles (5) de Lipinsky \rightarrow + petites librairies mais + pures, composant à + grand intérêt biologique

Utilisation de réactions à composants multiples générant une multitude de scuils synthétisant des librairies diverses

mais la chimie en phase solide ne permet pas d'utiliser toutes les synthèses accessibles en solut^o \rightarrow limite de cette approche

RDD

Design de drug et instrument?

Développement de nouvelles approches basées sur la structure de la cible en combinaison avec les outils de rétrosynthèse afin de guider la synthèse automatique → introduction de rationalité dans les méthodes basée sur la "chance".

Designs basés sur la structure

Nécessite des informations structurales précises.
Les structures peuvent être déterminées pour un complexe ligand - R ou pour montrer l'interact° de certains résidus du R avec le ligand.
Via NMR ou diffraction à rayon X

Méthodes computationnelles de découverte de drug (CADD)

Les méthodes assistées par ordinateurs sont utilisées dans la recherche pharmacochimique pour identifier les ligands de cibles orphelines (pas de ligand connu) ou pour identifier de nouveaux ligands pour les cibles non orphelins.

Méthodes rapides et efficaces pour identifier des ligands "lead" potentiel. N'est pas un substitut aux méthodes expérimentales mais limite de manière rationnelle l'espace chimique.

HTS est limité par le temps de ciblage et la diversité des bibliothèques chimiques.

L'espace chimique a été estimé à 10^{60} moléculas la proba qu'une bibliothèque (à chimie combinatoire) couvre cet espace est faible.

Méthodes computationnelles → la taille de l'ensemble de moléculas qu'il faut vraiment tester et propose de nouveaux chemins pour de nouveaux designs. Prend part à presque toutes les étapes pour le processus de design de ligand → la partie

Ordinateurs → recherche à travers une DB, des molécules qui pourraient correspondre aux besoins p/r à la cible et l'activité cherchée

- après HTS utilisat° d'analyse de similarité en ciblant DB

de ligands que l'on compare aux molécules positives trouvées par HTS

- après le début de la phase de la synthèse chimique, utilisation de QSAR (quantitative structure activity relationship) + ciblage de pharmacophore

- une fois la stct de la cible déterminée ainsi que la stct de complexes potentiels avec qq ligands synthétisés, l'optimisation et la synthèse de nouveaux ligands peut commencer

RDD

computer aided
drug discovery

1 Méthodes computationnelles de découverte de drug (CADD)

2 familles différentes :

- Basées sur le ligand : pas d'utilisation de la structure de la cible.
- Basées sur la structure de la cible : utilisation de la stct de la cible

1.1 Basées sur le ligand (LB-CADD)

Utilise la connaissance de molécules actives et inactives à travers

- la recherche de similarité chimique
- la construction de modèles prédictifs QSAR

④ Utilisé lorsqu'on a peu d'informations structurales (surtout le cas pour prot.)

Implique : 1. Choix de ligands connus pour interagir avec la cible d'intérêt
2. Dévise des structures de référence à prédire ces ligands
3. Analyse de leurs structures 2D ou 3D

But : représentation de ces composants avec leurs propriétés physicochimiques les + importantes pour leurs interactions voulues.

- Éliminer les info non pertinentes pour les interactions

Deux approches fondamentales :

- QSAR
- Screening virtuel

1.1.1 Construction d'un modèle QSAR

Modèle qui prédit l'activité biologique à partir de la structure chimique. Donne un poids aux caractéristiques de la structure chimique des composants en fonction de leur influence sur l'activité biologique d'intérêt.

QSARs (Quantitative Structure Activity Relationship) dérivés des équations qui permettent de prédire l'activité biologique à prédire la stct des ligands.

Essayer d'établir un lien quantitatif entre la variat° de l'activité biologiques du ligand et ses descripteurs moléculaires.

Descripteurs moléculaires : caractéristiques électriques, spectroscopiques, hydrophobicité, ... Tant plus important ! Physicochimiques et structurelles. L'acti biol et les propriétés physicochimiques sont connectées par une fonction mathématique. Acti biol = F (propriétés physicochimiques)

Principe : identifier cette relation et l'appliquer à de nouvelles entités chimiques

RDI

CADD

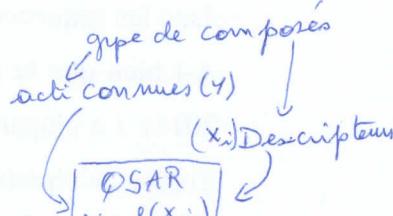
Basée sur le ligand

QSAR

La manière dont une molécule se comporte dépend de sa structure (p.e. nbre de Carbone détermine l'état)

QSAR Nécessite :

- un nombre de molé p. r. lesquelles l'acti est connue
- des descripteurs physicochimiques pour ces molécules
- une méthode appropriée pour obtenir un modèle



Les propriétés des descripteurs (selon leur stat et descript^o physicochimique) sont p.e:

- - le pdm moléculaire
- la géométrie
- le V, la surface
- l'électroégativité
- polarité
- la compor^t de groupes fonctionnels
- les propriétés de solvation

Classification des descripteurs moléculaires (n l'étude on favorise un type de descript)

1D: Les propriétés physicochimiques scalaires // pdm moléculaires

2D: descripteurs dérivés de la nature moléculaire. Presque plus utilisé. Prédicte unique par ligands connus proches des

3D: descripteurs dérivés de la conformation moléculaire

4D: extension de QSAR - 3D: traite chaque molé comme un ensemble de différentes conformations.

La prédiction des propriétés physicochimiques est une étape critique dans le développement de descripteurs moléculaires efficaces. Pt è simple p. e. des propriétés telle que le poids moléculaire, le nbre de lien doux ou d'H... Et + complexe p.e. p. la solubilité.

Utilisat^o d'un certain nombre de ligands pour générer l'équat^o mathématique. Ensuite le reste des ligands sont à tester l'équat^o générée.

RDD

CADD

Basée sur le ligand

QSAR

Méthodes pour générer des équations mathématiques (QSAR 2D)

- Régression linéaire simple (droite des moindres carrés) : activité = $a_0 + a_1(V_{mole_i})$
↳ jamais aussi simple
- Régression linéaire multiple (+ siens covariables = + siens descripteurs 2D)
 $acti = a_0 + a_1(V_{mole_i}) + a_2(\log P) + a_3(\mu_i) \dots$
↳ moment dipolaire
- Analyse en composante principale = Régression partielle des moindres carrés
(p à QSAR 3D)
 $acti = a_0 + a_1(PC_1) + a_2(PC_2) + a_3(PC_3) \dots$
- Algorithmes génétiques
- Réseaux de neurones

Méthodes pour générer équations mathématique pour QSAR 3D :

CoMFA: comparative field molecular analysis

④ Descripteurs moléculaires bcp + sophistiqués : ce ne sont plus des valeurs scalaires mais des champs qui décrivent champs stériques, champs électrostatiques.. Tous les champs intéressants pour prédire l'acti biol.

Méthodologie

1. Choix d'un set d'entraînement : set de ligands représentatifs pour décrire l'équat°
2. Info 3D : définir la conformation bioactive de chq molécule (difficile car pas forcément connue). Une molé a + siens conformat° ms en général 1 seule conformat° qui permet de se lier à une proté donnée. Chaque molé (ligand) a un certain nbre de degrés de liberté. nbre de cycles + ou - de liberté. Les cycles sont reliés par des "rotatable bond" autour desquelles on peut tourner librement. Pas facile à prédire si on ne connaît pas la stct du complexe molé-ligand.
3. Aligner ou superposer toutes les confo bioactives de chq ligand, en superposant les parties communes et les ligands analogues. Principe : si ces molé se lient ds le m̄ site du R cible, on suppose qu'elles vont ḡer m̄ types d'act → mise en évidence des groupes importants pour liaison au R.
4. Les molécules sont placées dans un V que l'on va griller
5. A chaque point de la grille/du réseau on va calculer des champs (stériques, électrostatiques (p.e. via charge partielle)). En chq point on considère un atome sonde qui porte certaines caractéristiques et on calcule l'interact° de ces caractères avec les mailles du réseau. Pour chq maille on aura donc une descript° de la contribut° (+, -, ..) qui permet de calculer champs. Ces champs pourront être introduits dans l'équat° que l'on va paramétriser

RDD

CADD : LB

QSAR

- CoMFA (3D)

Méthodo

6. Importe toutes ces valeurs d'E décivant les champs dans une table
7. Identifie corrélation entre les variat° de l'activité biologique avec les variation dans ces champs. Elaborat° d'une équation mathématique qui permettra d'infirer l'activité d'un molé à partir de ses propriétés

Succès dépend de :

sur le R qui nous intéresse

- Qualité du groupe de composés actifs/inactifs initial. De + ou f l' H_1 que le pôle de ligands actifs se lie au même endroit de R lors de l'alignement.
- Du choix des descripteurs. + ou f l' H_1 que les contribut° enthalpiques et les + importantes et les m° pr les composés.
- la capacité à générer une relation mathématique appropriée

Désavantages et avantages

- ⊖ Alignement est compliqué et s'il est mauvais impact sur le résultat
- ⊖ Difficile de définir confo bioactive
- ⊖ Prend du temps car on calcule les chps ds 1 grille d'une certaine taille (ms que pr Stat based R)
- ⊕ On obtient info qualitative
- ⊕ On reste ds les m° familles de ligands que ceux de départ
- ⊕ On peut avoir diversité moléculaire un peu + importante
- ⊕ On peut f une visualisation graphique des résultats
- ⊕ On peut optimiser relativement rapidement les séries de ligands que l'on va développer

- 4D

QSAR 3D pour lequel on tient compte d'un ensemble de conformations des L

RDD

CADD : LB

Sélection de nouveaux Composants basée sur la similarité chimique

Utilisation de mesures de similarité à des composés actifs pour sélectionner de nouveaux composants. On part du principe: actif // non actif // pas très vrai!

Le screening virtuel reprend des méthodes qui permettent de sélectionner les candidats appropriés.

1. On donne en input des structures chimiques avec certaines propriétés calculées de composants.

2. Le screening virtuel recherche dans des bibliothèques de presque toutes les tailles des structures qui partagent des similarités ou peut définir certains critères qu'on veut (pds molé, logP) ou qu'on veut pas (toxique)

○ On ainsi le nombre de structures à tester en définissant des priorités
La recherche de similarité peut se faire sur la sous structure 2D, la sous structure 3D ou l'appariement de pharmacophore.

Recherche de pharmacophore

Pharmacophore ≠ molécule. Contient des informations concernant les groupes fonctionnels qui agissent avec la cible, des info concernant le types d'act° non covalentes, les distances atomiques et les groupes actifs. Déterminer les caractéristiques que l'on pense importantes pour l'act° des ligands connus avec la cible

Ensuite ciblage virtuel dans les DB de molécules ayant ces caractéristiques

▷ à trouver équilibre et donner trop de caractéristiques (trop restrictif) en donnant trop peu (trop de molé correspond)

1. Sélect° de molécules représentatives

2. Choix d'un molé de référence

3. Définir sa conformation bioactive

4. Superposit° de toutes les molé sur la référence

5. Extraire les propriétés pharmacophore: donneur/accepteur de ptH, cycle aromatique, group hydrophile/phobe

6. Design une nouvelle molécule

Screening virtuel

Introduction de caractéristiques = stabilité, solubilité et métabolisme par l'influence sur l'absorption, la distribution, le métabolisme et l'excrétion (ADME) de la drug.

• Sélection Positive: pr sélectionner molécules.

Utilisé des 5 règles de Lipinski pr évaluer la biodisponibilité. Pas + de 5 H donneurs pds moléculaire ≤ 500 g/mol ; logP > 5 ; accepteurs d'H ≤ 10

• Sélection Négative: pr éliminer molécules.

Exclusion de composés connus pr leur toxicité. REOS ("Rapid Elimination Of Stupid") permet de guider la recherche vers ligands pas seulement actifs mais aussi potentiellement de bons ligands d'un pt de vue pharmacochimique

▷ Rapide \oplus Facile $\ominus \oplus$ les L se lient ds m site \ominus Pas de scoring \ominus m famille de molé

RDD

CADD

(cours 4) Basée sur la structure du récepteur (SB)

Basée sur la connaissance de la structure de la protéine cible (3D) afin de sélectionner des composants en se basant sur leurs \mathcal{E} de liaison

④ utilisé lorsqu'on dispose de données de hautes résolut° concernant la structure de la protéine cible (p.e. Proté solubles/facilement cristallisables)
Rmq: si pas on peut se baser sur modèle

Permet de concevoir et évaluer la qualité en termes d'affinité d'une série de ligands.

L'affinité s'évalue via le calcul de la différence d' \mathcal{E} $L + R$ en solut° et l'énergie du complexe formé $L + R \rightarrow L + R$

1. Info structurales concernant la cible (si pas dispo on peut se baser sur modélisation)

2. Recherche des interactions favorables formées entre les ligands et la protéine spécifique

3. Design de nouveaux composés ayant ces interactions favorables à travers l'analyse minutieuse du site de liaison de la protéine.

Deux approches fondamentales:

- le docking de ligand
- fragment based design

Ces 2 approches utilisent des algorithmes de docking moléculaires pour positionner les molécules ou les fragments dans le site de liaison du R. Ceci génère toute une série de complexes.

utilisent une fonction de scoring pour évaluer les énergies d'interactions et ainsi classifier les complexes.

La liaison d'un ligand = processus dynamique durant lequel L et R peuvent changer de conformations.

- la flexibilité du ligand est prise en compte sauf si trop grand nbre de L qui doivent être dockés.

- très vite impraticable de prendre en compte toute la flexibilité de la protéine. Il faut donc de nombreuses approches pour modéliser la flexibilité du R
Via la dynamique moléculaire et Monte Carlo MAIS prend beaucoup de temps

Via des bibliothèques de rotamères et de conformations de protéines

On calcule toutes les confor possibles de la molé en confor cis haute $\mathcal{E} \chi = 0^\circ$ et trans basse \mathcal{E} p.ex. le butane

Evalue les rotations des chaînes latérales

↳ 2CH₃ se font face

CADD : SB

RDD

Docking de ligand ("entier")

1. Docking : une série de molécules sont positionnées dans un récepteur
(le mieux : avoir une idée du site où positionner le ligand)

Utilisation de "Structure-based high-throughput screening (SB-vHTS)

On utilise des bases de données de structure 3D de ligands que l'on va docker dans la structure 3D du R.

2. Après le ciblage on filtre via une fonction de scoring qui permet d'évaluer l'affinité des différents complexes générés.

(3.) Tester expérimentalement via $IC_{50} = [L]$ d'inhibiteur qu'il faut mettre pour avoir 50% d'inhibition. Ex: Hsp90 cible pr thérapie de oncologie
Myeloperoxidase cible contre maladie inflammatoire,

Fragment-based Design

1. Construction d'une molécule à partir de "briques" moléculaires positionnées dans le site de liaison des R. Construct^o in situ des L, fragment par fragment

- Via la technique de croissance séquentielle

La molécule croît dans le site de liaison. Un 1^{er} fragment ayant le meilleur score est positionné dans le site de liaison. Apd du 1^{er} fragment on ajoute un nouveau fragment qui a un bon score et peut faire une liaison avec le 1^{er} fragment. La croissance de la molé dans le site de liaison est contrôlée par un algo de recherche qui évalue chaque possibilité de croissance avec une fonction de score

- Via placement des fragments puis liaison

Le site de liaison est mappé pour identifier les points d'ancrages possibles pour des groupes fonctionnels. Ces groupes sont ensuite liés ensemble pour former la molécule entière

p.e. Thrombine

NFB

RDD

CADD : SB

Cette méthode nécessite un logiciel de docking qui doit avoir :

- une méthode de recherche de position efficace
- une fonction de scoring adéquate (le + difficile)

Méthode de docking moléculaire (que ce st L ou fragments en théorie !!)

- Utilisée pour prédire les structures des complexes protéine - ligand
- Le logiciel fournit différents modes de liaison pour un ligand
- La méthode de docking est combinée à une méthode d'évaluation pour classer les complexes en fonction de leurs affinités prédictives
- La fonction de scoring peut être utilisée pour classer les différentes positions et orientation du ligand. Ou pour classer différents ligands.

2 types de méthodes de docking moléculaire :

- Méthode géométrique

- Fonction d'optimisation

Méthodes géométriques

Basée sur la complémentarité de forme (espace) et sur les contraintes chimiques tel R et le L (donneur et accepteurs de pt H, sites hydrophobes)

1. Identification et caractérisation du site de liaison du R pour identifier les sites chimiques d'intérêt et les zones accessibles séparément. Les points d'intérêt sont ainsi placés sur la protéine et sur le ligand.
2. Superposition des points correspondant à la fois le L et le R
Le ligand est positionné dans le site de liaison du R de manière à ce que sa géométrie corresponde à celle du R

Dock : Crée une image "négative" du site de liaison à tel la S moléculaire de la cible.

1. Cette image est composée de sphères superposées (représentant atome sans H) de différents rayon qui sont en contact avec la surface du R de manière à être situé entre 2 points de cette surface.
2. Les centres des sphères sont liés de manière à générer des familles de positions pour lesquelles les $|d|$ entre les centres de sphères correspondent à des distances tel les atomes de ligands contenus ds des DB. Sélect de molé qui correspondent à la $|d|$ calculée tel les atomes.