

SUJET DOCKING

CONTEXTE

Objectif du docking : identifier à partir des structures 3D de deux monomères que l'on sait interagir, la structure 3D du complexe qu'ils forment.

Principe du docking : le récepteur (en bleu ou orange sur les figures de gauche et de droite respectivement) est maintenu fixé pendant tous les calculs (i.e. ses coordonnées 3D sont maintenues inchangées pendant tous les calculs). Afin d'explorer la surface du récepteur, **on discrétise la surface du récepteur** selon des couples d'angles θ, ϕ . A chaque couple (θ, ϕ) correspond donc **une position** du ligand – en général, on génère environ 200 positions de départ du ligand. **Pour chaque position** du ligand, nous aurons **plusieurs orientations** de ce dernier (1000 environ). On a donc au total (200x1000 conformations possibles). En effet, ce dernier peut lui aussi tourner sur lui même (voir flèche verte, figure de gauche). Pour ce faire, on discrétise aussi la surface du ligand cette fois par des couples d'angles (θ', ϕ') , ce qui permettra d'explorer **pour une position du ligand, toute sa surface**. **Chacune des orientations de chaque position du ligand** (soit chacune des 100 000 conformations possibles) est ensuite **évaluée par une fonction de score**, qui généralement basée sur le score d'énergie du complexe. Néanmoins, il est fréquent d'y ajouter d'autres critères empiriques (tels que la conservation des acides-aminés impliqués dans l'interface du complexe, la taille de l'interface formée par le complexe etc) afin d'améliorer les prédictions. La plupart des programmes ne prennent pas en compte la flexibilité du ligand et du récepteur pendant le calcul de docking car cela est beaucoup trop coûteux. Les coordonnées internes du ligand et du récepteur sont donc inchangées pendant le calcul. C'est la limite principale de ces méthodes qui peinent à trouver la bonne solution lorsque l'un des deux partenaires subit des réarrangements conformationnels importants lors de l'assemblage. Néanmoins, ces méthodes s'avèrent efficaces lorsque les conformations des deux monomères sont relativement rigides lors de l'assemblage.

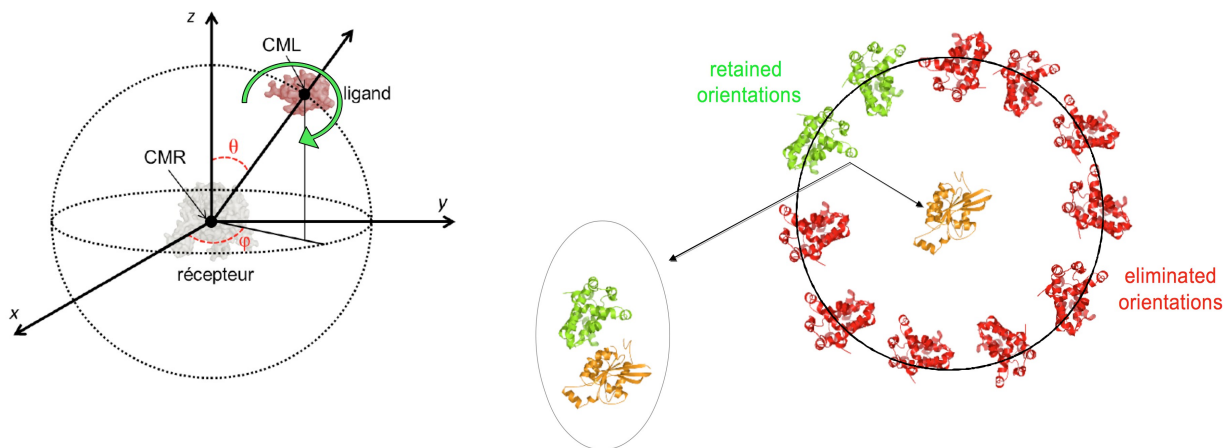


Figure 1 : **gauche** : échantillonnage de l'espace conformationnel du récepteur et du ligand ; **droite** : score des solutions et identification des solutions les plus probables.

PROJET

Score et sélection des solutions : le but de ce projet consiste à écrire un programme permettant d'évaluer un sous-ensemble (948 solutions que nous avons présélectionnées) des solutions générées par un algorithme de docking afin de « scorer » les différentes solutions et prédire la structure 3D du complexe formé par les deux monomères de l'étude.

Dans un premier temps, il faudra implémenter la fonction de score impliquant les termes non-liés de la fonction d'énergie décrite dans *Cornell et al, JACS 1995* ainsi que les paramètres associés afin d'évaluer énergétiquement toutes les solutions possibles. Calculez le score de chaque solution avec votre programme. Votre programme devra créer un répertoire « `scoring_Cornell` » et stocker dans ce répertoire un fichier contenant le classement des solutions avec leurs énergies associées ainsi que la structure 3D de la meilleure solution dans un fichier pdb nommé « `complexe_predit_score1.pdb` ».

Dans un second temps, il s'agit d'incorporer lors de l'étape de « `scoring` » (soit directement en ajoutant un ou des nouveaux termes dans la fonction de score, soit en filtrant les meilleures solutions énergétiques, à vous de voir) une ou plusieurs nouvelles contraintes pour voir évaluer vos solutions. Pour cette étape, vous avez la liberté totale ! Vous pouvez vous aider de la bibliographie de ce qui est connu sur l'interaction barnase-barstar (contrainte obligeant certains résidus clés à l'interaction à se trouver à l'interface par ex.) ou de ce que font d'autres méthodes de docking (taille des interfaces, hydrophobicité des interfaces, conservation des acides-aminés présents dans l'interface, ou autre idée...). Votre programme devra créer un second répertoire et y stocker le nouveau classement ainsi que la structure 3D de la meilleure solution dans un fichier pdb nommé « `complexe_predit_score2.pdb` ».

Votre programme devra être générique et fonctionner pour n'importe quel couple de deux protéines, néanmoins, dans le cadre de ce projet, vous appliquerez votre programme au couple barnase-barstar pour lequel nous avons déjà généré les conformations (i.e. les différentes solutions).

Evaluation des solutions : en fait la structure 3D du complexe est déjà connue (fichier `1brs_native.pdb`). On aimerait évaluer la performance de votre programme dans sa capacité à prédire la structure 3D du complexe. Pour cela, vous devez écrire un second programme indépendant du premier qui prendra en argument la vraie solution (i.e. le fichier `1brs.pdb` dans votre cas) et la meilleure solution identifiée par le premier programme. Ce nouveau programme devra calculer le RMSD entre la vraie structure 3D et votre solution. En particulier, 3 RMSD seront à fournir : RMSD calculé sur le complexe entier, RMSD calculé uniquement sur le ligand et RMSD calculé sur les résidus de l'interface. Enfin, votre programme devra aussi calculer le nombre de contacts « natifs » identifiés dans votre structure prédite. C'est à dire, votre programme devra calculer le nombre de contacts impliquant des paires de résidus de la protéine 1 ET de la protéine 2 retrouvés dans l'interface native et dans celle que vous aurez prédite. Ces 4 critères permettront de rendre compte de la pertinence de votre prédiction.

Données : le répertoire `allatom_conf` contient les fichiers pdb correspondant aux coordonnées du ligand pour chacune des 948 solutions de docking que nous avons présélectionnées. Le fichier `1BRS_rec.pdb` correspond aux coordonnées du récepteur qui a été maintenu fixé pendant le calcul de docking (c'est pourquoi, il y a un seul fichier). Le fichier `1BRS_A_1BRS_B.tra` contient les coordonnées 3D du centre de masse du ligand pour chacune des solutions. Cela peut éventuellement vous aider à avoir une idée de la distribution des positions du ligand autour du récepteur. Vous pouvez les visualiser en l'ouvrant sous pymol et en chargeant aussi la structure du récepteur. Enfin, le fichier `1brs_native.pdb` contient les coordonnées 3D du complexe natif.

Remarques générales : votre programme devra être commenté, générique. La clarté et la modularité du code seront évaluées. Les paramètres modifiables devront être donnés en argument. Une fonction `help()/usage()` sera la bienvenue.