

Rapport projet: Modélisation Comparative

Introduction:

La modélisation d'une structure protéique basée sur une structure homologue(modélisation comparative) est une méthode utilisée en biologie structurale. Dans ce processus, un alignement d'une séquence de protéine cible sur la structure d'un ou de plusieurs supports est utilisé comme entrée dans un programme qui construit un modèle 3D. L'alignement de la séquence cible avec le support ainsi que le choix de la ou des meilleures structures support sont importants dans la modélisation comparative.

L'objectif de ce projet est de réaliser un programme de modélisation comparative. A partir d'un alignement de séquence entre une séquence cible à modéliser et une ou plusieurs séquences supports dont les structures sont connues,on construit un modèle 3D de la protéine cible en utilisant les structures supports.

Il existe de nombreux programmes de modélisation comparative tels que Modeller, SegMod/ENCAD, SWISS-MODEL, 3D-JIGSAW, nest et Builder [1].

Dans ce projet on s'inspire de l'algorithme de Segmod/Encad [2] et de ProMod (Swiss-Model) [3].

L'originalité du projet va se baser sur l'utilisation d'une plus grande banque de fragments comme la base de donnée HOMSTRAD ou la base de donnée qui contient les 500 meilleures structures utilisées pour les distributions Ramachandran-plot :

<http://kinemage.biochem.duke.edu/databases/top500.php> .

Matériel et méthode:

Le programme est écrit en Python 3 et divisé en différents étapes:

1. Dans un premier temps, on garde uniquement les carbones α de la structure support. A partir du fichier pdb de la protéine support, on récupère les coordonnées des atomes C alpha ainsi que le nom et le numéro des résidus correspondants.
2. Ensuite, on fait correspondre la nature des acides aminés dans la structure support à partir de l'alignement entre la structure support et la séquence à modéliser. On effectue alors un alignement deux à deux.Cette étape permet aussi de générer la structure initiale constituée uniquement de carbone alpha.
3. Génération des boucles non conservées et des brèches : On cherche dans une banque de 500 structures(<http://kinemage.biochem.duke.edu/databases/top500.php>) des fragments de structures compatibles avec les parties manquantes de la protéine(les gaps ou les insertions) en utilisant comme contrainte la distance entre carbones α des extrémités N et C terminale de la boucle et du nombre d'acide aminées à modéliser dans la boucle. S'il n'existe aucun fragment compatible, on relaxe les

contraintes de distance entre carbones α des extrémités N et C terminales pour trouver un fragment compatible.

4. Génération des atomes de la chaîne principale : Jusqu'ici, on crée un modèle de structure constitué uniquement de carbones alpha. Il faut construire le reste des atomes de la chaîne principale (N, C, O). Pour ce faire, pour chaque fragment de 5 carbones α la structure initiale, on recherche dans la sous-banque obtenue à partir de PISCES les fragments les plus proches, en terme de distance inter carbone alpha. A chaque position de la structure initiale, on superpose le fragment le plus proche des 5 carbones α centré sur la position d'intérêt. Ainsi on ajoute les coordonnées des atomes de la chaîne principale (N, C, O) sur 3 positions.
5. Enfin la génération des chaînes latérales sera réalisée grâce au logiciel SCWRL de Roland Dunbrack.
6. Une étape de minimisation peut être aussi réalisé avec GROMACS .
7. A la fin on compare les résultats avec ceux de Modeller.

Perspectives:

Le programme n'est pas abouti jusqu'à la fin des étapes. Jusqu'ici, le programme permet de générer la structure initiale constituée uniquement de coordonnées des carbones alpha correspondant entre la protéine cible et la structure support. Pour avoir la structure avec les autres atomes de la chaîne principale; on doit tenir en compte des gap et des insertions et chercher dans une banque de fragments ceux qui sont compatibles. La résolution de cette étape a pris beaucoup de temps et n'a pas permis de finir le projet.

La génération des chaînes latérales reste à faire aussi dans ce projet.

Bibliographie:

[1] : Wallner B, Elofsson A. All are not equal: a benchmark of different homology modeling programs. Protein Sci. 2005

May;14(5):1315-27.

[2] : Levitt M. Accurate modeling of protein conformation by automatic segment matching. J.

Mol. Biol. 1992;226:507-

533.

[3]: Peitsch, M. C. (1995) Protein modeling by E-mail Biochemical Society Transaction 24:274-279