

PROJETS COURTS

UE Programmation et Gestion de Projet

2023-2024

Contact

Dr. Tatiana Galochkina
tatiana.galochkina@u-paris.fr

Pr. Jean-Christophe Gelly
jean-christophe.gelly@u-paris.fr



SUJET : REPLIEMENT D'UN MODELE SIMPLIFIE DE PROTEINE PAR UN ALGORITHME DE MONTE-CARLO ET ECHANGE DE REPLIQUES

Objectif : Réalisez un programme reprenant la méthode décrite dans l'article. Dans un premier temps l'algorithme de Monte-Carlo simple sera implémenté, ensuite le RE. Une séquence arbitraire de protéine pourra être soumise au programme afin de calculer son repliement par ce modèle.

:

Référence : Thachuk C, Shmygelska A, Hoos HH. A replica exchange Monte Carlo algorithm for protein folding in the HP model. BMC Bioinformatics. 2007 Sep 17;8:342. PuProfessional Address:



SUJET : ALIGNEMENT HEURISTIQUE RAPIDE POUR LA RECHERCHE DE SEQUENCES DANS UNE

BANQUE : BLAST

Objectif : Réalisez un programme reprenant la méthode décrite dans l'article.

Référence : Stephen Altschul, Warren Gish, Webb Miller; Eugene Myers; David J. Lipman (1990). "Basic local alignment search tool". Journal of Molecular Biology. 215 (3): 403-410. doi:10.1016/S0022-2836(05)80360-2.



SUJET : CALCUL DE LA SURFACE ACCESSIBLE AU SOLVANT D'UNE PROTEINE

Objectif : Objectif : Réalisez un programme permettant de calculer la surface accessible au solvant (absolue et relative) à partir des coordonnées d'une protéine issue d'un fichier PDB.

Étapes :

- Extraction des coordonnées de chaque atome
- A partir de chaque atome contenu caractérisant la protéine, créez un nuage de points uniformément sur la surface d'une sphère centrée sur l'atome. La sphère aura pour valeur de rayon le rayon de Van der Waals de l'atome + rayon de la sonde (solvant = rayon d'un atome d'oxygène) (l'algorithme de Saff et Kuijlaars pourra être utilisé).
- A partir de chaque point de la sphère recherchez s'il existe un point appartenant à une autre sphère à une distance inférieure au rayon de la sonde
- A partir des points sans contact, calculez la surface accessible totale en terme de points puis la convertir en \AA^2 de chaque résidu de la protéine. Calculez ensuite la surface accessible relative, et enfin le pourcentage d'accessibilité au solvant.
- Comparez et évaluez la méthode par rapport à NACCESS

Référence : Shrake, A; Rupley, JA. (1973). "Environment and exposure to solvent of protein atoms. Lysozyme and insulin". *J Mol Biol* **79** (2): 351-71. doi: 10.1016/0022-2836(73)90011-9



SUJET : ALIGNEMENT MULTIPLE HEURISTIQUE PAR LA METHODE CLUSTAL

Objectif : Réalisez un programme reprenant la méthode décrite dans l'article. L'algorithme de construction de l'arbre pourra être remplacé par la construction d'un arbre par embranchement séquentiel. L'algorithme heuristique d'alignement séquentiel pourra être remplacé par un alignement basé sur la programmation dynamique.

Référence : Desmond G. Higgins , Paul M. Sharp, Fast and sensitive multiple sequence alignments on a microcomputer, *Bioinformatics*, Volume 5, Issue 2, April 1989, Pages 151–153, <https://doi.org/10.1093/bioinformatics/5.2.151>



SUJET : ASSIGNATION DES STRUCTURES SECONDAIRE DE PROTEINES : DSSP

Objectif : Réalisez un programme reprenant la méthode décrite dans l'article. Implémentation de la méthode DSSP pour l'assignation des structures secondaires. Dans un premier temps, il est possible de s'intéresser qu'aux structures secondaires régulières (alpha Hélice, et brin bêta).

Référence : Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Kabsch W, Sander C, Biopolymers. 1983 22 2577-2637. PMID: 6667333; UI: 84128824.



SUJET : ASSIGNATION ET DETECTION DES PARTIES TRANSMEMBRANAIRES D'UNE PROTEINE

Objectif : Réalisez un programme reprenant la méthode décrite dans l'article : concevoir un outil permettant de déterminer les zones transmembranaires d'une protéine.

Etapes:

- Choisir des protéines membranaires et globulaires de référence (simple à analyser, voir banque OPM)
- Calcul de la surface accessible au solvant de chaque acide aminé avec NACCESS
- Lecture du PDB et extraction des coordonnées des Calpha
- Calcul du centre de masse
- Détermination des droites passant par le centre de masse et quadrillant avec suffisamment de précision toutes les directions
- Déplacement d'une tranche de 1 Angstrom normale à une droite et calcul de l'hydrophobicité relative des résidus dans les zones exposées dans la tranche
- Calcul de la position de la membrane par la moyenne de l'hydrophobicité relative et en comparant ces valeurs selon les différentes droites

Référence : Tusnády GE, Dosztányi Z, Simon I. Transmembrane proteins in the Protein Data Bank: identification and classification. Bioinformatics. 2004 Nov 22;20(17):2964-72. Epub 2004 Jun 4.

Remarques :

- Il n'est pas demandé de vérifier si la protéine est de taille suffisante ou si le fichier contient effectivement une protéine (on prendra donc soin de vérifier son programme sur des chaîne unique de protéines membranaire et de protéines globulaire)
- Il n'est pas demandé de reconstruire le modèle biologique de la molécule (à partir du champ BIOMOLECULE)
- Le programme NACCESS ou DSSP sera utilisé pour calculer la surface accessible au solvant



SUJET : CONCEPTION D'UN PROGRAMME D'EVALUATION DE LA QUALITE D'UN MODELE 3D DE PROTEINE

Objectif : Réalisez un programme reprenant la dernière partie de la méthode décrite dans l'article 1. Ce programme prend en entrée un modèle de structure de protéine (au format mmCIF ou PDB) et calcule, à partir des potentiels statistiques DOPE et du calcul des distances entre carbone alpha une pseudo-énergie (ou score) de chaque résidu de la protéine. La somme de ces pseudo-énergies sera calculée et un z-score calculé à partir d'une série de mélange aléatoire de la nature des résidus sur la même structure.

Les potentiels statistiques DOPE précalculés sont disponibles à l'adresse : www.dsimb.inserm.fr/~gelly/data/dope.par. La première colonne est la nature du résidu 1, la deuxième l'identifiant de l'atome du résidu 1, la 3^{ème} la nature du résidu 2, la 4^{ème} l'identifiant de l'atome du résidu 2. Les autres colonnes représentent la valeur d'énergie pour une classe de distance donnée.

Référence : Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci. 2006 Nov;15(11):2507-24.



SUJET : CONCEPTION D'UN PROGRAMME DE *THREADING* PAR DOUBLE PROGRAMMATION DYNAMIQUE

Objectif : Réalisez un programme reprenant la méthode décrite dans l'article 3 basé sur la double programmation dynamique (pour plus d'information 4). Le *threading* (enfilage) [1,2,3] est une stratégie pour rechercher des séquences compatibles avec une structure. Seul les carbones α de la protéine seront considérés. Vous utiliserez les potentiels statistiques DOPE. Les potentiels statistiques DOPE précalculées sont disponibles à l'adresse [5].

Références :

- 1- Jones, D.T., Taylor, W.R. & Thornton, J.M. (1992) A new approach to protein fold recognition. *Nature*. 358, 86-89.
- 2- Jones, D.T., Miller, R.T. & Thornton, J.M. (1995) Successful protein fold recognition by optimal sequence threading validated by rigorous blind testing. *Proteins*. 23, 387-397.
- 3- Jones, D.T. (1998) *THREADER : Protein Sequence Threading by Double Dynamic Programming*. (in) *Computational Methods in Molecular Biology*. Steven Salzberg, David Searls, and Simon Kasif, Eds. Elsevier Science. Chapter 13.
- 4- [Protein Structure Comparison Using SAP - Springer](#)
- 5- <http://www.dsimb.inserm.fr/~gelly/doc/dope.par>



SUJET : *PROTEIN INTERACTIONS CALCULATOR* IMPLEMENTATION

Objectif : Réalisez un programme reprenant la méthode décrite dans l'article : *Protein Interactions Calculator* (PIC; <http://pic.mbu.iisc.ernet.in>) est un serveur web particulièrement intéressant qui permet le calcul des différents types d'interactions importantes pour la structure protéique : les liaisons disulfure, les interactions hydrophobes, les interactions ioniques, les liaisons hydrogènes, les interactions aromatiques-aromatiques, les interactions aromatiques-soufre et les interactions cation- π (Tina *et al*, *Nucleic Acid Res*, 2007). Il a aussi d'autres propriétés mais cela est en dehors de ce projet.

Le projet est de ré-implémenter ces calculs à partir d'une structure correcte (pas de problème spécifique à gérer), et générer une sortie similaire (et utilisable localement) mettant en avant les interactions entre atomes / résidus.

Référence : Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Kabsch W, Sander C, *Biopolymers*. 1983 22 2577-2637. PMID: 6667333; UI: 84128824.