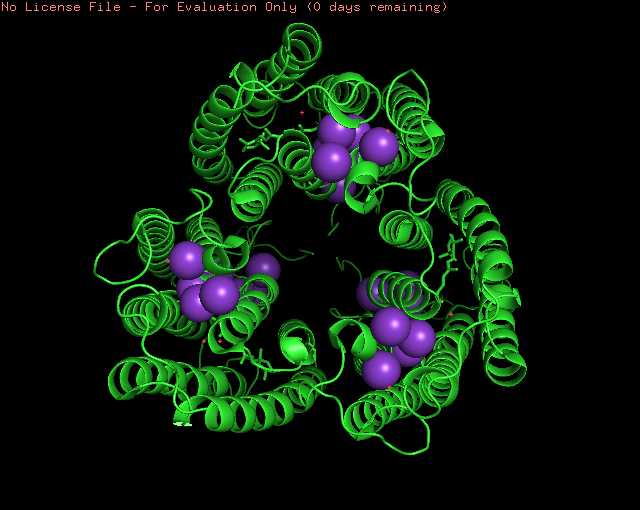
Master Bio-informatique

2024-2025



Rapport : Projet court

Dictionary of Secondary Structure of Proteins  
(DSSP)



Introduction :

Le repliement des protéines est le mécanisme par lequel une chaîne d'acides aminés adopte sa structure tridimensionnelle naturelle, indispensable à son rôle fonctionnel. Cette forme est dictée par la séquence des acides aminés et maintenue grâce à plusieurs types d'interactions. Nous pouvons notamment citer les interactions Hydrogène, qui certes sont des interactions dites faible. Malgré la faible énergie nécessaire pour les rompre, elles restent néanmoins primordiales dans le repliement protéique. En effet leur nombre les rends incontournable pour prédire le repliement protéique.

Dans cette optique une part importante de la recherche de repliement protéique s’est concentré sur le développement d’algorithme de prédiction. Parmi les outils de prédiction nous pouvons parler du DSSP (Dictionary of protein secondary structure Pattern) qui a fait le sujet d’un article de Biopolymers en 1983. Dans l’article les auteurs expliquent utiliser les liaisons hydrogène et des caractéristiques géométriques. Au cours de notre projet nous allons tenter d’implémenter cet algorithme.

Matériel et Méthode :

Dans un premier temps nous avons sélectionné 10 protéines sur Protein Data Bank (PDB). Nous avons fait en sorte de prendre des protéines de class différentes afin de représenter au mieux la diversité de structure des protéines. Nous avons donc téléchargé un fichier PDB pour chaque protéine. Afin de pouvoir comparer notre algorithme à celui d’origine nous avons soumis toutes les protéines à cette première méthode.

Une image contenant texte, Police, capture d’écran, nombre

Description générée automatiquementDans un second temps nous avons utilisé l’algorithme que nous avons programmer. Nous avons d’abord utilisé un programme de calcul des liaison hydrogène écrit par I.K MacDonald, D Naylor, D Jones et J.M Thornton à l’University College London. Celui-ci s’appelle HBplus. Il permet de calculer et d’ajouter les hydrogènes à un fichier PDB ainsi que de calculer les liaisons hydrogènes. Par default ce programme renvoi un fichier au format hb2 qui s’organise selon le format suivant.

Nous pouvons donc récupérer les toute ces informations pour sélectionner les liaisons qui sont pertinente. En effet nous avons pu sélectionner les interactions ayant lieux entre deux atomes de la chaine main.

Après ce premier traitement nous avons pu lancer la prédiction des structures secondaires. Nous avons tenté d’implémenter la prédiction des hélices alpha et de feuillet beta. Pour cela nous avons rechercher les liaisons entre les atomes i et i+4 pour les hélices et les interactions entre i et i+ (5 à 120) pour les feuillets. Nous nous sommes limités à ces deux recherches par manque de temps.

Nous avons ensuite voulu comparer les deux méthodes. Pour ce faire nous avons comparé les prédictions pour chaque résidu. Cependant comme nous n’avons pas implémenté toute les structure secondaire nous n’avons pas comptabiliser les structures autre que les helice alpha et les brin beta.

Résultat :

Pour des raisons de non-compatibilité d’environnement entre deux machines le code n’a pas pu être lancer. Pour le moment nous ne pouvons donc pas présenter de résultat

Discussion :

Les difficultés que nous avons rencontrées ont été principalement d’organisation et de segmentation des taches. En effet ce projet a été le premier projet important que nous avons réalisé. Nous avons donc perdu beaucoup de temps à savoir comment procéder.

De plus Nous avons travailler sur VScode ce qui était également une première fois. Nous avons donc renconter beaucoup de problème d’environement non compatible entre plusieurs machines.

Nous avons fait le choix de ne comptabiliser que les structures les plus simple dans nos comparaisons. Ce choix affecte notre accuracy étant donner que certain résidu pourrai être placer dans différentes catégorie en remplacement.

Bibliographie :

I.K. McDonald and J.M. Thornton (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.*, **238**, 777-793. [PubMed id: [**8182748**](http://www.ncbi.nlm.nih.gov/pubmed/8182748)].

Lien github https://github.com/Gcluzeau/projet\_court\_Gcluzeau