**Sujet N°3 : Assignation des structures secondaire de protéines**

1. **Introduction :**

Presque toutes les fonctions des êtres vivants dépendent des protéines. Elles représentent 50 % de la masse sèche des cellules et jouent un rôle dans tout ce que fait un organisme. Aussi diverses qu'elles puissent être, elles sont toutes composées des mêmes 20 acides aminés. En formant des liaisons peptidiques entre les groupes amino et carboxyle de deux acides aminés différents, de grandes chaînes polypeptidiques peuvent être créées.

La structure secondaire des protéines résulte des liaisons hydrogène formées entre les atomes du squelette polypeptidique. Les liaisons hydrogène se forment entre l'atome d'oxygène partiellement négatif et l'atome d'azote partiellement positif. Deux types extrêmement courants en biochimie sont l'hélice alpha et le feuillet bêta.

L'hélice alpha (α-hélice) a une conformation en spirale à droite, dans laquelle chaque groupe N-H du squelette donne une liaison hydrogène au groupe C=O du squelette de l'acide aminé quatre résidus avant lui dans la séquence.

L'autre type courant de structure secondaire est le brin bêta. Un brin bêta (β-brin) est un tronçon de chaîne polypeptidique, généralement long de 3 à 10 acides aminés, dont le squelette est dans une conformation presque entièrement étendue. Deux ou plusieurs chaînes polypeptidiques adjacentes parallèles ou antiparallèles de brin bêta stabilisées par des liaisons hydrogène forment un feuillet bêta.

Pour ce fait, l’article « Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features» montre la méthode de détermination de la structure secondaire d’une protéine en utilisant la méthode DSSP, donc on va se baser sur cet article pour comprendre et identifier les types de la structure d’une protéine d’intérêt.

* La détermination de la structure secondaire d’une protéine par la méthode DSSP :

Comme c’était déjà mentionné, la structure secondaire des protéines est déterminée par le modèle de liaison hydrogène. Un grand nombre de serveurs et d'outils sont utilisés pour prédire l'analyse de la structure secondaire. DSSPcont et STRIDE sont des outils en ligne utilisés pour comprendre la structure secondaire. Dans le programme d'analyse de simulation de dynamique moléculaire, le DSSP (Dictionary of Protein Secondary Structure) est utilisé pour créer, visualiser le tracé de la structure secondaire. Cela nous permet de comprendre le changement structurel de la structure de la protéine.

Define Secondary Structure of Proteins (DSSP) est l'outil standard pour l'annotation des éléments de structure secondaire à partir des structures protéiques. Basé principalement sur les modèles de liaison hydrogène et certaines contraintes géométriques, il attribue chaque résidu à l'un des huit états possibles.

1. **Matériels et Méthodes**

Le principe du programme est de parcourir la liste des résidus impliqués dans la formation des liaisons d’hydrogène dans le but de dénoter la position des structures dans un ordre d’évaluation, commençant par les turns et les hélices, puis les bridges , las ladders et finalement les feuillets.

1. Détermination des Turns
2. Détermination des Hélices alpha
3. Détermination des Bridges : Parallèles et Antiparallèles
4. Détermination des Ladders
5. Détermination des feuillets

**HBPlus**

Fichier PDB Fichier .hb2 🡺

1. Le fichier de l’Input :

Pour étudier la structure secondaire d’une protéine, on a besoin d’un fichier de format PDB (Protein Data Bank) [2]. Ce fichier-là, Contient des informations bibliographiques attachées à la structure, sur la résolution et les paramètres cristallographiques, la séquence et parfois la structure secondaire. 2ème partie: Elle contient les coordonnées atomiques Dans cette partie les atomes désignés par ATOM se situent sur la chaine protéique, tandis que les atomes désignés par HETATM (HETeroAToM group) font partie des molécules cofacteurs, substrats, ions ou autres groupes qui sont liés par une liaison covalente à la chaîne protéique. On peut obtenir le fichier PDB, en mettant le nom de la protéine sur laquelle on souhaite travailler sur le site RCSB.

Le fichier PDB ne contient pas les liaisons d’hydrogène. Pour ce fait, on a eu recours au programme HBPlus[1]. Il nous a permet d’avoir un fichier .hb2 qui contient toutes les liaisons d’hydrogènes qui se trouvent dans la protéine avec les emplacements et les noms des résidus impliqués.

Alors l’Input dans note code est un fichier .hb2 issu de HBPlus.

1. Langage de Programmation :

Python : le langage de programmation le plus utilisé aujourd’hui en bioinformatique, notamment pour l’analyse de données.

L’interface : Visual Code

1. L’utilisation de PyMol :

PyMol nous a permet de visualiser la structure secondaire de la structure en 3D, d’où on peut voir visuellement (sans des coordonnées ni des emplacements des résidus) , les types de structures présents dans notre molécule d’intérêt.

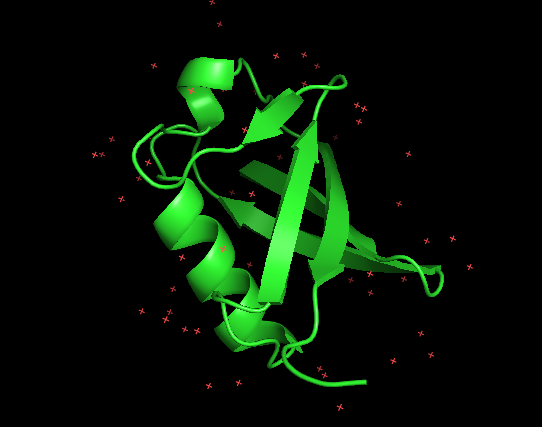
1. DSSP : Le output obtenu par DSSP va servira comme une référence sur laquelle on se base pour comparer les résultats du programme DSSP et les résultats obtenus par ce programme.

En effet, Le programme DSSP fonctionne en calculant l'affectation de la structure secondaire la plus probable compte tenu de la structure 3D d'une protéine. Pour ce faire, il lit la position des atomes dans une protéine, puis calcule l'énergie de la liaison H entre tous les atomes. L'algorithme écarte tous les hydrogènes présents dans la structure d'entrée et calcule les positions optimales des hydrogènes en les plaçant à 1,000 Å du N du squelette dans la direction opposée à la liaison C=O du squelette. Les deux meilleures liaisons H pour chaque atome sont ensuite utilisées pour déterminer la classe la plus probable de structure secondaire pour chaque résidu de la protéine [4].

1. Tout le travail est basé sur les conditions énoncées dans l’article « Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features »
2. **Résultats et discussion**

#### La molécule d’intérêt est la BIQUITIN (1ubq)

1. Résultats de Pymol



* on peut clairement voir la présence d’une longue hélice et d’une autre moins petite.
* Le nombre des feuillets présents dans la protéine est deux : un feuillet avec deux ladders et un autre avec trois.

Structure secondaire de la protéine 1ubq visualisée par Pymol

1. Résultats du programme

L’implémentation effectuée et testée avec la protéine 1ubq a montré la présence d’une hélice et d’un feuillet dans les positions suivantes :

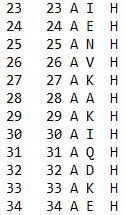






1. Résultats fournis par DSSP :

Le output de DSSP (qui se trouve dans l’annexe), montre la présence d’une hélice entre le résidu num°23 et le résidu num°34



🡺Le code implémenté et le DSSP ont donné presque le même résultat avec

une différence au niveau de 3 résidus : l’efficacité du programme pour la

détection des hélices est de 75 %.

Les algorithmes pour déterminer les sheets ont été implémentés mais ne fournissent pas de résultat qui semblent attendus

Référence :

[1] : <http://www.uoxray.uoregon.edu/local/manuals/hbplus/hbplus.txt>

[2] <https://wiki.ubuntu.com/kmezhoud/Bioinformatics?action=AttachFile&do=get&target=structure-3D-prediction.pdf>

[3] <https://biochimiedesproteines.espaceweb.usherbrooke.ca/2b.html>

[4] <https://swift.cmbi.umcn.nl/gv/dssp/DSSP_3.html>

<https://github.com/DSSP-github/DSSP/blob/master/AS_DSSP.py>

<https://git.pagelibre.org/richard/projet_court/src/master/src/geom.py>

<https://hal.archives-ouvertes.fr/hal-03264103v2/document>