**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.

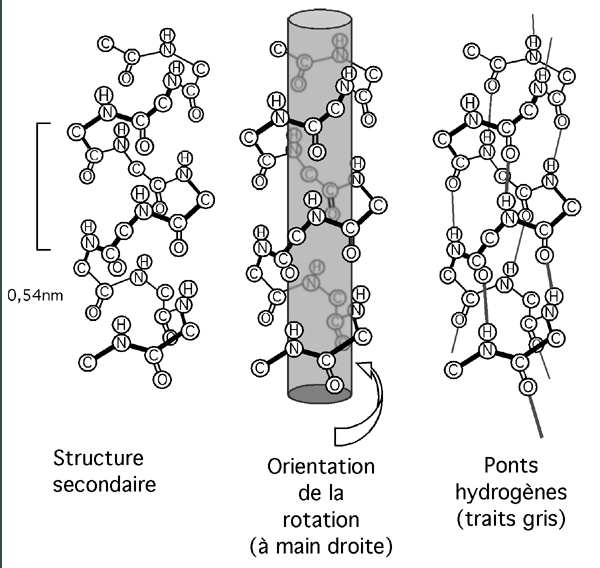
L'hélice alpha (α-hélice) a une conformation en spirale à droite, dans laquelle chaque groupe N-H du squelette donne une liaison hydrogène au groupe C=O du squelette de l'acide aminé quatre résidus avant lui dans la séquence.

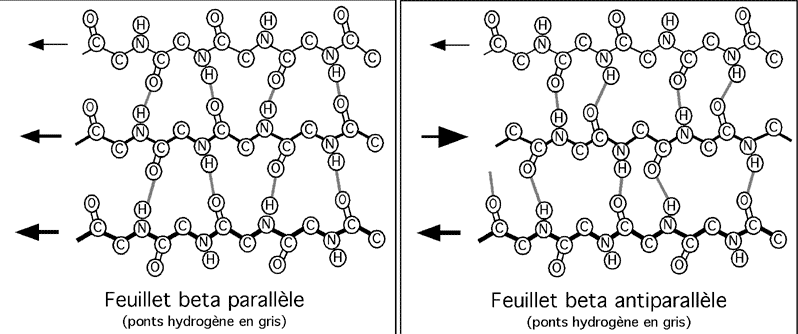
L'autre type courant de structure secondaire est le brin bêta. Un brin bêta (β-brin) est un tronçon de chaîne polypeptidique, généralement long de 3 à 10 acides aminés, dont le squelette est dans une conformation presque entièrement étendue. Deux ou plusieurs chaînes polypeptidiques adjacentes parallèles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

* 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**
2. 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 logiciel HBPlus[1] qui 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 est le langage de programmation le plus utilisé aujourd’hui en bioinformatique, notamment pour l’analyse de données.

1. L’utilisation de PyMol :

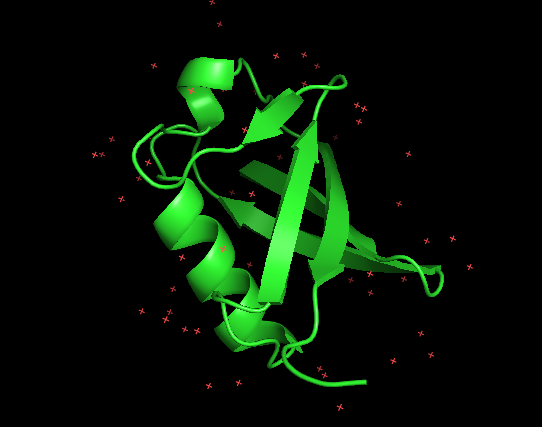
PyMol nous a permet de visualiser la structure secondaire de la structure en 3D, d’où on peut voir visuellement (sans des coordonnées ni des emplacements des résidus ) , les types de structures présents dans notre molécule d’intérêt.

1. **Résultats et discussion**
2. Structure Hélice alpha :

Capture d’écran de résultat

1. Structure feuillet beta :

Capture d’écran



Résultat issu à partir du logiciel Pymol :

On peut clairement voir la présence d’une grande hélice et d’une autre plus petite.

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>

<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>

Readme

L’Input pour notre programme est un fichier qu’on a extrait en utilisant le logiciel HBPlus :

Ref : <http://www.uoxray.uoregon.edu/local/manuals/hbplus/hbplus.txt>

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

1. Installation de HBPlus :

Pour pouvoir installer HBPlus, Vous devez être un utilisateur académique et avoir envoyé un accord de confidentialité signé aux auteurs.

HBPLUS est disponible par ftp anonyme sous la forme d'un fichier 'crypté', hbplus.tar.Z.cr.

Sur un système unix, utilisez :

unix> crypt [password] < hbplus.tar.Z.cr > ! hbplus.tar.Z

unix> décompresser hbplus.tar.Z

unix> tar xf hbplus.tar

unix> make

Le mot de passe est envoyé par e-mail comme réponse au fichier de confidentialité.

Après l’installation du logiciel, on le lance par la commande : ./hbplus et finalement on ouvre notre fichier PDB. Qu’on a déjà nettoyé par la commande "^ATOM" 1ubq.pdb > new\_1ubq.pdb avec 1ubq le nom de fichier PDB de la protéine qu’on souhaite analyser.

1. L’Output de HBPlus :

Le logiciel HBPlus nous permet d’extraire les liaisons d’hydrogènes présentes dans une protéine, à partir d’un fichier PDB.

A l’entête, il y’a la version du logiciel utilisé, le temps de la récupération du fichier, les noms des auteurs..

Après, on a les liaisons d’hydrogènes avec l’emplacement(les atomes engagés dans cette liaisons) , la distance et l’angles entre les atomes ..

Language de programmation utilisé : Python

Explication du code :

Le code exploité est divisé en cinq parties :

1. Première partie : Lecture du fichier .hb2 et création de deux listes : liste HDonor : qui contient les indices des résidus qui ont donné la molécule d’hydrogène dans la liaison et la liste HAcceptor qui contient les indices des résidus qui ont reçu la molécule d’hydrogène.
2. Deuxième partie : Création d’une liste « turn » qui permet de déterminer les couples (donneur et accepteur) qui forment le 4\_turn.

C’est quoi le 4\_turn ? D’après l’article

1. Troisième partie : Détermination de l’hélice :

Création d’une boucle qui en détectant 4\_turn consécutifs, va assigner une hélice