Titre : Mesurer l’écart entre deux structures de protéine

Contexte : Un protéine est composée d’une séquence d’acides aminés. Pour exercer sa fonction, il ne suffit pas que ces acides aminés soient simplement disposés le long d’une chaîne, la protéine doit se replier, adopter une certaine forme dans l’espace. Si l’on veut interférer avec cette protéine, pour des raisons thérapeutiques par exemple, il faut connaître parfaitement sa forme.

Comment déterminer sa forme ?

* Il existe des méthodes de détermination par l’expérience longues et coûteuses comme la cristallographie aux rayons X.
* Plus récemment, on a vu des algorithmes capables de plus ou moins déterminer comment la protéine allait se replier à partir de sa séquence d’acides aminés. Cela fait l’objet de la compétition informatique CASP. Les candidats doivent deviner la forme d’une molécule déjà connue et leur résultat est ensuite comparé par la forme obtenue par l’expérience. **Nous allons étudier différents outils de comparaison.**

Etapes :

* Cas simple : comparaison de 2 nuages de points dans l’espace (voire un nuage de points reliés)
* Mesure(s) de comparaison par un indice de corrélation entre 0 et 1 (comme pour une régression linéaire)
* Etude du format PDB OU mise en place d’un format plus simple (beaucoup d’informations inutiles dans un fichier PDB) :
* manipuler une protéine simple
* représenter dans l’espace (logiciel adapté comme Rasmol ou sur python)
* représenter 2 molécules dans un même espace
* Cas plus complexe : point = atome de carbone, centre d’un acide aminé ; points reliés par des liaisons ET prise en compte des sous-structures (comme les feuillets alpha où les points forment des « tires-bouchons » et beta )

Outil utilisé par l’organisation du CASP : Global Distance Test

Méthode (Test de distance global - https://fr.abcdef.wiki/wiki/Global\_distance\_test) :

A revérifier

1. Superposition des structures (orientation et alignement)
2. On définit une distance seuil d, (où d = 0,5 ; 1 ; 1,5 ;… ; 10 Angstrom)
3. R = nbr de carbones alpha dont la distance avec la ref >d / nbr total
4. Moyenne des R