

M1 BIBS

Séance 8

CONTEXTE BIOLOGIQUE

En biologie structurale le RMSD (Root Mean Square Deviation) est utilisé pour mesurer la distance moyenne séparant les atomes de deux protéines superposées. Il rend ainsi compte de la déviation structurale entre deux structures protéiques alignées. Plus le RMSD est petit, plus les structures des protéines alignées se ressemblent.

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2}$$

où N correspond au nombre total de PAIRES d'atomes (une paire comprenant un atome de la structure 1 et son atome correspondant dans la structure 2) utilisés pour le calcul (généralement sur les carbones α (CA), mais le RMSD peut aussi être calculé sur les atomes du squelette peptidique N, CA, C,O par exemple ou n'importe quel jeu d'atomes) et δ_i correspond à la distance spatiale séparant les deux atomes numéro i de la structure 1 et de la structure 2 (paires d'atomes équivalents entre les deux structures).

Le principe consiste donc à superposer les deux structures d'intérêt puis à calculer la distance spatiale séparant chaque paire d'atomes superposés. Le RMSD correspond à la racine de la moyenne des distances élevées au carré.

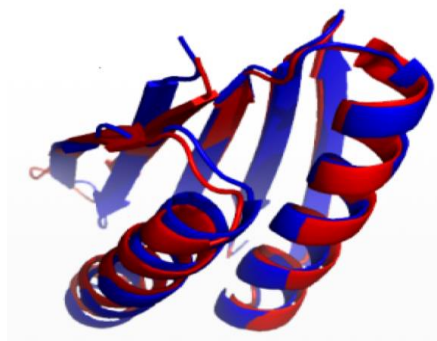
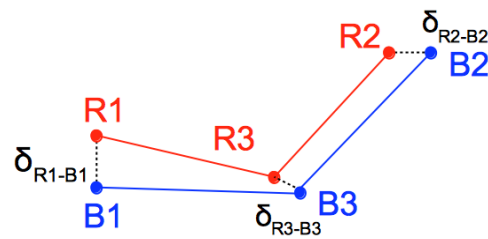


Figure 1 : calcul du RMSD sur les CA des deux structures rouge et bleue. Les deux structures sont alignées, puis la distance spatiale entre chaque paire d'atome $\text{CA}_{\text{rouge}}\text{-CA}_{\text{bleu}}$ est calculée. Pour une protéine de 110 résidus par exemple, le RMSD calculé sur les CA, sera calculé à partir des 110 distances séparant les couples de $\text{CA}_{\text{bleu}}\text{-CA}_{\text{rouge}}$.

PARTIE 1 :



Dans cet exemple artificiel, on cherche à calculer le RMSD entre les 3 atomes (1,2 et 3) de ces deux structures rouges et bleues. Les coordonnées atomiques sont stockées dans les deux fichiers rouge.pdb et bleu.pdb. Ecrire un petit script permettant de calculer le RMSD entre ces deux structures. Vérifier que l'on trouve un RMSD de $\sim 0.32 \text{ \AA}$.

PARTIE 2 :

Généraliser le script pour pouvoir calculer le RMSD entre deux protéines entières (utilisez les deux structures superposées stockées dans les fichiers 2LQ7_1.pdb et 2LQ7_2.pdb). On se limitera au cas où les deux protéines alignées sont de même taille et de même séquence. Seule leur conformation 3D diffère et nous voulons justement calculer le RMSD entre ces deux conformations DEJA alignées. Votre programme concerne UNIQUEMENT la partie « calcul du RMSD » à partir des fichiers pdb contenant les coordonnées 3D des protéines alignées. Votre programme devra en particulier :

- Prendre en argument de façon explicite, le nom des deux fichiers pdb contenant les coordonnées 3D des deux structures déjà alignées (encore une fois, vous ne vous préoccupez pas d'aligner les structures !!!) ainsi que le nom des atomes à partir desquels le RMSD sera calculé. La valeur par défaut sera calcul sur les CA uniquement (essayez aussi sur les CB, puis sur l'ensemble des N, CA, C, O).
- Comprendre une fonction usage() permettant d'afficher un descriptif du programme et de ses arguments. Cette fonction sera appelée automatiquement dès qu'un des 2 arguments obligatoires (-pdb1, -pdb2) ne sera pas appelé ou mal utilisé par l'utilisateur.
- Afficher en sortie, le nombre d'atomes sur lesquels aura été calculé le RMSD, la valeur du RMSD et le nombre de résidus de la protéine.

PARTIE 3 :

On veut maintenant pouvoir calculer le RMSD entre toutes les paires de structures stockées dans le répertoire « structures ». Créer un nouveau script permettant de calculer automatiquement le RMSD pour l'ensemble de structures stockées dans le répertoire « structure ». Il fournira en sortie un fichier de 3 colonnes contenant respectivement le nom des deux pdb utilisés pour le calcul et le RMSD

correspondant.

PARTIE 4 :

1/ On veut voir comment la déviation structurale se répartie le long de la séquence. Modifier le script de la partie 2 de façon à calculer le RMSD indépendamment pour chaque résidu de la protéine. Votre programme fournira en sortie un fichier de deux colonnes correspondant respectivement à la position du résidu et au RMSD entre les deux conformations de ce résidu observées dans les deux fichiers d'entrée.

2/ Votre programme devra en sortie fournir un graphique représentant la distribution des valeurs de RMSD de chaque résidu le long de la séquence protéique.

Exemple de sortie :

