

Bioinformatique structurale niveau II

TP de B. Offmann (séance 3 et 4)

Classification des motifs pentapeptidiques à l'aide d'une carte topologique de Kohonen et applications.

Vous travaillerez sur les fichiers DSSP dérivés du sous-ensemble A. Vous vous aiderez de l'article « Local structural motifs of protein backbones are classified by self-organizing neural networks ».

- 1) A l'aide de scripts réalisés dans l'environnement R ou d'un programme que vous écrirez en C, effectuer une classification non-supervisée des angles phi et psi associés à tous les pentapeptides constitutifs des structures du sous-ensemble A. Pour cela, vous utiliserez une fenêtre coulissante correspondant à 5 résidus consécutifs pour extraire à chaque fois un vecteur de 6 angles dièdres (psi1, phi2, psi2, phi3, psi3, phi4, psi4, phi5). Vous définirez une carte de kohonen de 100 cellules (10 x 10) où chaque cellule comprend un vecteur de 8 angles dièdres dont les valeurs fluctuent entre -180° et +180°. Vous initialiserez aléatoirement cette matrice.
- 2) Vous indiquerez les conditions de l'apprentissage non-supervisé que vous mettrez en œuvre (taux d'apprentissage, taux de propagation autour de la cellule vainqueur, calcul de distance, critère de convergence, nombre de cycles...)
- 3) Lorsque l'apprentissage aura convergé, vous fournira une représentation graphique de la matrice 10 x 10 faisant clairement apparaître les valeurs d'angles phi et psi. Chaque cellule ainsi « stabilisée » représente un prototype structural définit par un vecteur de 6 angles.
- 4) Vous effectuerez alors une analyse statistique de la représentation de ces prototypes ainsi dans les structures du sous-ensemble A. Il s'agit d'effectuer un comptage du nombre de fois que ces prototypes sont « trouvés » dans les structures de départ. Déterminer le pourcentage de pentapeptides représentés par les 26 cellules les plus peuplées.
- 5) Nommer ces cellules par les lettres A à T. Cet ensemble constitue un alphabet structural. Ecrire un script qui transforme les fichiers PDB en suite de lettres tirées de cet alphabet selon les valeurs d'angles phi et psi de leurs pentapeptides constitutifs.
- 6) Créer un script qui rajoute aux fichiers DSSP une colonne supplémentaire indiquant l'alphabet structural associé à l'acide aminé central de chaque pentapeptide d'une structure.