Introduction

Désordre protéique

Épissage alternatif

Isoforme

Matériel et méthodes

Base de données

Clustal Omega [1]

Clustal Omega est un algorithme d’alignement de séquences, utilisés lorsqu’il est nécessaire d’aligner plus de deux séquences nucléiques ou protéiques. Dans ce cas d’application, étant donné que ces sont des protéines résultant de l’épissage alternatif qui sont étudiées, il semble approprié d’utiliser ce type d’alignement. L’algorithme de Clustal Omega est utilisé en local, et avec différents scripts pour obtenir les résultats d’alignement.

En premier lieu, un script python de traitement de fichiers permet de découper une base de données. En effet, les bases de données disponibles au format fasta sur uniprot contiennent l’ensemble des protéines avec leurs isoformes concaténés en un fichier. Pour permettre un meilleur traitement via Clustal, la base de données est séparée en différents fichiers, contenant chacun une protéine avec l’ensemble de ses isoformes. Un script bash permet ensuite d’appliquer l’algorithme Clustal à l’ensemble de ces fichiers.

Enfin, pour vérifier le bon fonctionnement de l’algorithme, un script R permet de vérifier le nombre de gap moyen par séquence inséré par séquence via l’algorithme, ainsi que leur longueur. Étant donné que les protéines étudiées sont des isoformes, il serait logique d’observer un faible nombre de gap par séquence, avec une longueur moyenne élevée.

PSIPRED [2]

PSIPRED est un algorithme de prédiction de structures secondaires (principalement les trois types de structures secondaires majoritaires, à savoir les boucles, les hélices α et les brin β), se basant sur le calcul de matrice PSSM à l’aide de Psi-BLAST. On estime que les résultats obtenus via PSIPRED sont fiables à environ 80%. L’algorithme va ici être utilisé pour déterminer si l’épissage alternatif semble modifier le type de structure secondaire prédit sur les isoformes, et dans quelle proportion. Dans ce projet, PSIPRED a été utilisé en local,

Disopred [3]

Résultats

Clustal Omega

Psipred

Disopred

Discussion

Conclusion

Bibliographie

[1] Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Research. 2019 Jul;47(W1):W636-W641. DOI: 10.1093/nar/gkz268.

[2] Jones DT. (1999) Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292: 195-202.

[3] Jones DT, Cozzetto D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics. 2015 Mar 15;31(6):857-63. doi: 10.1093/bioinformatics/btu744. Epub 2014 Nov 12.