Charles-Antoine Lanoix remis le 29 octobre 2019  
Félix Adam

TP2 Bio-informatique

Repliement d’ARN en tige-boucle

1. **À quoi correspondent les cellules de M sur l’anti-diagonale (M[i, |S| − i])? Faut-il remplir toute la table?**

Les cases de l’anti-diagonale représentent deux nucléotides qui se suivent dans la séquence. Par conséquent, elles ne peuvent pas être appariées. Les cases de l’anti-diagonale n’influencent pas le score de l’alignement. Elles seront initialisées avec une étoile on pourra remplir la partie supérieure à celles-ci en commençant par la partie supérieure gauche.

1. **Donnez les équations d’initialisation et de récurrence pour remplir la table M.**
2. **Décrivez comment on peut retrouver un repliement en tige-boucle maximisant les appariements de nucléotides.**

Les cases étoiles représentent des points de départ au « back-tracking » à partir desquelles il est possible de se déplacer dans toutes les directions. Des diagonales de cases étoiles pourront être ajoutées si on veut trouver des tiges-boucle de plusieurs nucléotides. Il faut partir de n’importe quelle case étoile sur l’anti-diagonale en s’assurant que la première case visitée qui possède un score soit maximum.

1. **Appliquez votre algorithme à : GCGUGCUUGCGUGCACG. On vous demande la table de programmation dynamique, le score, ainsi que le repliement.**
2. **Pour être stable, la boucle de la tige-boucle doit contenir au moins 3 nucléotides. Décrivez une façon de modifier votre algorithme afin d’avoir des tige-boucles avec au moins 3 nucléotides dans la boucle.**

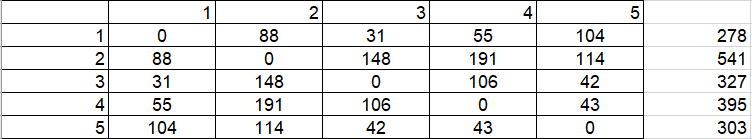
Il faudra rajouter des anti-diagonales de cases étoiles en haut de l’anti-diagonale de base et commencer sur l’anti-diagonale en s’assurant que la première case avec score soit maximal.

1. **Décrivez une modification de votre l’algorithme qui permet les appariements "G-U"**

Il faut modifier les équations de récurrence.

Alignement multiple de séquences

1. **Grâce à l’algorithme d’alignement global, calculez la matrice des scores de similarité entre toutes les paires de séquences.**



1. **En déduire la séquence centrale S\* de S (justifiez ce choix).**

La séquence centrale S\* de S est S2 puisque c’est celle ou la somme de ses scores par rapport aux autres séquences est la plus élevé (541)

1. **Construire un alignement multiple de S en utilisant la méthode de la séquence centrale. Vous devez illustrer les différentes étapes pour la construction de votre alignement.**
2. **Quelle est le score SP de votre alignement ?**
3. **Donnez la séquence consensus Z de l’alignement et son pourcentage d’identité avec chacune des séquences de S. Notez que le pourcentage d’identité entre deux séquences alignées Sx et Sy désigne le pourcentage de positions comprenant des acides aminés identiques.**

Outils de bioinformatique

1. **En utilisant BLASTP, identifier le nom du gène, l’espèce d’origine et la fonction de chacune des protéines de sequences.fasta.**

Protéine #1

Nom : hémoglobine 2

Espèce d’origine : Arabidopsis thaliana

Fonction : L'hémoglobine, couramment symbolisée par Hb, parfois Hgb, est une métalloprotéine contenant du fer, présente essentiellement dans le sang des vertébrés au sein de leurs globules rouges, ainsi que dans les tissus de certains invertébrés. Elle a pour fonction de transporter l'oxygène O2 depuis l'appareil respiratoire (poumons, branchies) vers le reste de l'organisme. [Wikipedia]

Protéine #2

Nom : cytoglobine

Espèce d’origine : Bos taurus

Fonction : La cytoglobine (ou histoglobine) est une globine respiratoire constituée, chez l'homme et la souris, par 190 acides aminés. [Wikipedia]

Protéine #3

Nom : sous-unité alpha A d’hémoglobine

Espèce d’origine : Gallus gallus

Fonction : L'hémoglobine possède une structure quaternaire caractéristique de nombreuses protéines à sous-unités globulaires. La plupart de ses résidus d'acides aminés sont engagés dans des hélices α reliées entre elles par des segments non hélicoïdaux. [Wikipedia]

Protéine #4

Nom : myoglobine

Espèce d’origine : construction synthétique

Fonction : La myoglobine, couramment symbolisée par Mb, est une métalloprotéine contenant du fer présente dans les muscles des vertébrés, et particulièrement des mammifères. Elle est apparentée structurellement à l'hémoglobine, mais a pour fonction de stocker l'oxygène O2 plutôt que de le transporter.

Protéine #5

Nom : neuroglobine

Espèce d’origine : Rattus norvegicus

Fonction : La neuroglobine est une globine de stockage et de transport de l'oxygène dans le système nerveux. À ce titre, elle présente beaucoup d'analogies avec la myoglobine dont une structure monomérique et une forte affinité pour l'oxygène. Ainsi, la neuroglobine facilite le transport d'oxygène vers les mitochondries des cellules nerveuses et joue un rôle fondamental dans la protection des neurones en conditions d'hypoxie.

1. **Utilisez le programme Clustal (vous pouvez aussi vous servir de ce site http://www.ebi.ac.uk/Tools/msa/clustalo/) avec les paramètres par défaut pour faire un alignement multiple des séquences, puis comparez le score SP entre cet alignement et celui que vous avez obtenu à la section précédente.**