

Mini projet 3 ; La construction des profils

Professeur Tom Lenaerts

Assistant Catharina Olsen et Elisa Cilia

Information additionnelle sur :

http://www.ulb.ac.be/di/map/tlenaert/Home_Tom_Lenaerts/INFO-F-208.html

Les données

Le but de ce projet est de construire un profil (PSSM) pour un ensemble de domaines:

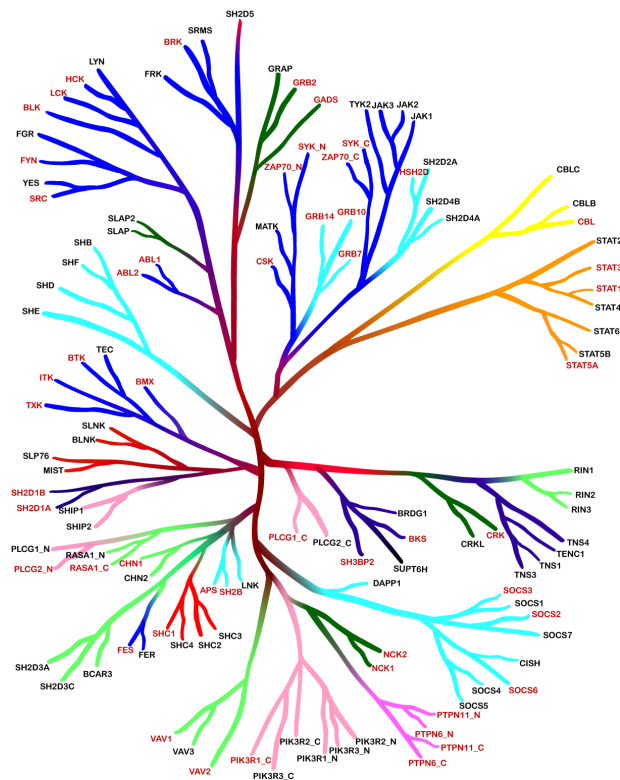


Figure 1: La famille des domaines SH2 trouvé dans des protéines humaines. Les feuilles contiennent les noms des protéines. Toutes informations concernant ces domaines peuvent être trouvées sur le site <https://sites.google.com/site/sh2domain/>

Cherchez les séquences de tous les domaines SH2 sur le site web Uniprot (<http://www.uniprot.org/>) et insérez toutes les séquences en format FASTA dans un seul fichier avec le nom `to-be-aligned.fasta`.

Les outils

Alignez les séquences au sein de ce fichier en utilisant deux des outils suivants :

1. CLUSTAL Omega : <http://www.ebi.ac.uk/Tools/msa/clustalo/>
2. TCooffee : <http://www.ebi.ac.uk/Tools/msa/tcofee/>

3. MUSCLE : <http://www.ebi.ac.uk/Tools/msa/muscle/>

Enregistrez les résultats de ces deux alignements en format FASTA dans un fichier nommé `msaresults-<nom d'outil MSA>.fasta`. Ce fichier contient maintenant l'alignement entre toutes les séquences qui était produit par un des deux outils que vous avez choisis (CLUSTAL, TCOFFEE ou MUSCLE).

Implémentez un logiciel qui construit un profil pour chaque fichier (donc 1 PSSM par outil d'alignement). Regardez les slides 23-34 dans « L7 Alignement de plusieurs séquences ». N'oubliez pas d'utiliser les *pseudo-counts*.

Assurez-vous que le code puisse être utilisé pour d'autres groupes de séquences. Faites le teste avec un autre ensemble de séquences ; **regardez le site PFAM, <http://pfam.sanger.ac.uk>, et prenez une des famille de ce site.**

Validation et présentation

Quand vous avez construit les deux PSSM, il faut les analyser et comparer.

- 1) Construisez un Weblogo pour la famille SH2 et la famille que vous avez choisis. Comparez ce Weblogo avec l'information dans vos PSSM. Quelles sont les positions conservées ? Est-ce que les acides aminés qui font parti des structures secondaires sont bien conservés ?
- 2) Comparez aussi vos résultats avec l'information que vous trouvez sur le site PFAM pour les deux familles.

Sur votre wiki vous devez ajouter une introduction sur les deux familles. Quels sont les types de protéines dans lequel vous pouvez trouver ces deux domaines ? Quelles sont leurs fonctionnalités ? Est-ce qu'elles sont impliqués dans certaines maladies ? Quelle structure 3D est-ce qu'ils ont ? Avec quels peptides ou protéines est-ce que ces domaines aimes interagir ? etc. **Essayez de donner une explication pédagogique pour chaque famille.**