

Levy et al proposent en 2012 le concept de *stickiness* [1]. La *stickiness* reflète la propension de chaque type d'acide aminé à être impliqué dans des interfaces protéiques qu'elles soient spécifiques ou non spécifiques.

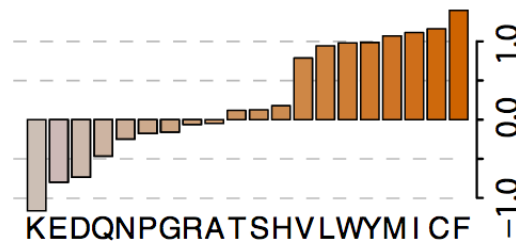


Figure 1 : échelle de *stickiness*. Extrait de Levy et al, 2012 [1]

Formule de *stickiness* : $\text{stickiness}(\text{AA}) = \log(\frac{\text{freq}(\text{AA interface})}{\text{freq}(\text{AA surface})})$

En effet, certains acides aminés sont plus enclins à l'interaction (généralement les acides aminés hydrophobes) que d'autres et peuvent alors promouvoir des interactions non spécifiques (non « désirées ») avec les autres protéines du compartiment cellulaire. On parle alors d'acides aminés *sticky* tandis que les acides aminés non *sticky* regroupent majoritairement des acides aminés chargés et polaires (voir échelle de *stickiness* en figure 1). Ainsi, une protéine avec une surface très *sticky* va être plus encline à interagir de façon non spécifique et pourrait s'avérer délétère pour la cellule si elle est présente en grande quantité. En effet, elle pourrait conduire à des agrégats protéiques fortement délétères pour l'organisme. Dans ce projet, on se propose d'étudier la corrélation entre l'abondance des protéines de *S. cerevisiae* et la *stickiness* de leurs surfaces. Vous disposez des abondances dans le fichier « 4932-WHOLE_ORGANISM-integrated .txt » avec les identifiants « gene name ».

Partie 1 : Analyse

Dans un premier temps, récupérez sur UNIPROT toutes les protéines dont la structure 3D a été caractérisée. La correspondance entre les identifiants « gene name » et PDB

est renseignée dans le fichier « uniprot-yeast-filtered-organism.txt » qui contient l'information de tous les gènes de *S. cerevisiae*. Le « gene name » est renseigné après le pattern « OrderedLocusNames= ». Si la structure 3D a été caractérisée, le code PDB est indiqué par le pattern « PDB; ». Lorsque plusieurs fichiers PDB sont disponibles, prendre celui de meilleure résolution (i.e. la plus petite valeur de résolution renseignée après le pattern X-ray; »).

Exemple :

DR PDB; 2IW3; X-ray; 2.40 Å; A/B=2-981.

DR PDB; 2IWH; X-ray; 3.00 Å; A/B=2-981.

Dans cet exemple, prendre le PDB 2IW3 car sa résolution est meilleure ($2.4 < 3.00$).

Ensuite récupérer les fichiers PDB correspondants.

Identifiez les résidus de surface de chaque PDB et en déduire les résidus enfouis. Vous pouvez pour cela utiliser le logiciel freeSASA qui permet de calculer l'accessibilité au solvant des résidus d'une protéine à partir d'un fichier PDB. On considère qu'un résidu est exposé lorsque son accessibilité est $> 25\%$.

Calculez la *stickiness* de chaque protéine (pour les résidus de surface uniquement, pour les résidus enfouis et pour tous les résidus). La *stickiness* d'un ensemble de résidus correspond simplement à la somme des valeurs de *stickiness* des acides aminés de cet ensemble. Les valeurs sont données dans la table « stickiness.txt ».

Calculez la corrélation entre l'abondance et la *stickiness* de la surface protéique, du cœur (résidus enfouis) et de la totalité des résidus. Que concluez-vous ?

Partie 2 : Prédiction

On se propose ensuite d'utiliser le critère de *stickiness* pour prédire le(s) site(s) d'interaction d'une protéine de structure connue. Le principe consiste à identifier des « *patches* » enrichis en résidus *sticky* sur sa surface. Proposez et implémentez une stratégie permettant d'identifier ces *patches*. Pour information, les sites d'interaction des protéines impliquent généralement une trentaine de résidus. Vous avez à disposition un jeu test de complexes protéiques vous permettant de tester et évaluer votre méthode.

Références

[1] Levy ED, De S, Teichmann SA. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. Proc Natl Acad Sci USA 2012;109:20461–6. doi:10.1073/pnas.1209312109.