PROJET LONG

Université de Paris

M2 Biologie Informatique

CONTACT
Jean-Christophe Gelly

WEB
www.dsimb.inserm.fr/~gelly

Next Generation Protein Peeling algorithm using Convolutional Neural Network for Protein Structure Analysis

Projet proposé par Jean-Christophe Gelly

Introduction

Le "*Protein Peeling*" est une méthode permettant de découper une structure protéique en petites unités compactes, nommées "unités protéiques" (UPs). Les UPs sont définies comme des fragments protéiques très compacts d'au moins une vingtaine d'acides aminés. La méthode de "Protein Peeling" se base exclusivement sur la matrice de contact des distances inter-Calpha. Les distances sont traduites en probabilités au moyen d'une fonction logistique. Le découpage emploie le principe d'une segmentation hiérarchique classique pour construire une sérié de partitions emboitées de la structure 3D. Les UPs les plus compactes, définies selon un coefficient de compacité, sont composées de structures secondaires régulières contenant de nombreux contacts internes. Les UPs peuvent aussi être considérées comme un échelon supplémentaire dans la description de l'anatomie des protéines".

Projet

En utilisant la même première étape (génération de la carte de probabilités de contact) que celle décrite dans les publications détaillées en références, votre objectif est d'implémenter une nouvelle version du programme qui se base sur l'identification des régions de la protéine les plus compactes et les plus indépendantes.

L'objectif du projet est de proposer une nouvelle méthode de *Protein Peeling* qui se base sur la carte de probabilités de contacts et un algorithme d'identification basé sur un réseau *convolutionnel (figure 1)*. L'idée est le réseau "apprenne" ce qu'est une unité protéique sur des exemples existants obtenues par l'algorithme de *Protein Peeling* classique.

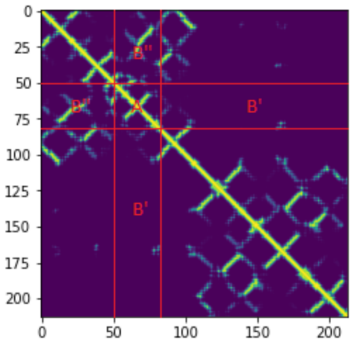

Figure 1. Principe d'identification d'une Unité protéique A à partir de la carte de probabilité de contact. Les régions B correspondent aux contacts "externes" qui doivent être le plus faible possible tandis que le nombre de contact "interne" (région A) doivent être le plus important possible. B' contient peu de contact tandis que B" en contient plus.

Vous comparerez les résultats obtenus avec cette nouvelle méthode avec les résultats de la méthode classique (www.dsimb.inserm.fr/dsimb_tools/peeling3) sur les exemples présentés dans les articles.
Vous comparerez également aux découpages obtenus au niveau domaine issu de CATH et SCOP et accessible à l'adresse suivante : http://www.bio.ifi.lmu.de/SCOPCath

Informations supplémentaires

Vous devez prendre contact pour avoir des détails sur le projet et des informations supplémentaires.

Références

1- 2011 ; Gelly, J. C. ; de Brevern, A. G. Protein Peeling 3D: new tools for analyzing protein structures BIOINFORMATICS ; 27 ; 132--133 ;

2 - 2006 ; Gelly, J. C. ; de Brevern, A. G. ; Hazout, S. 'Protein Peeling': an approach for splitting a 3D protein structure into compact fragments BIOINFORMATICS ; 22 ; (2) 129-33 ;

3- 2006 ; Gelly, J. C. ; Etchebest, C. ; Hazout, S. ; de Brevern, A. G. Protein Peeling 2: a web server to convert protein structures into series of protein units NUCLEIC ACIDS RES ; 34 ; (Web Server issue) W75-8 ;

4. 2017 ; Postic, G. ; Ghouzam, Y. ; Chebrek, R. ; Gelly, J. C. An ambiguity principle for assigning protein structural domains SCIENCE ADVANCES ; 3 ; (1) e1600552