

Protein Graph Repository

Wajdi Dhifli, Rabie Saidi

FSJ – Université de Jendouba, Tunisie

wajdidhifli@gmail.com

saidi@isima.fr

<http://www.isima.fr/~saidi>

Résumé. Protein Graph Repository (PGR) est un outil bioinformatique sur le web permettant d'obtenir une nouvelle représentation de protéines sous la forme de graphes d'acides aminés, une représentation plus simple et plus facile à étudier par les moyens informatiques et statistiques dédiés aux graphes. La génération des graphes est faite à partir d'un parseur appliqué sur des fichiers des protéines PDB extraits de la base Protein Data Bank et en précisant les paramètres et la méthode à utiliser. Les graphes générés sont ensuite enregistrés dans un entrepôt doté de moyens de recherche, de filtrage et de téléchargement. PGR peut être provisoirement consulté à l'adresse <http://www.enode-edition.com/pgr/>, il est spécialement dédié aux recherches intéressées à l'étude de données protéiques sous la forme de graphes et permettra donc de fournir des échantillons pour des travaux expérimentaux.

1 Description technique

PGR a été testé dans 2 environnements différents :

- Environnement WAMP sous Windows XP PRO(SP3) avec un ordinateur de bureau Intel Pentium D équipé d'un processeur Dual Core 2 GHz et RAM 1Go DDR2.
- Environnement LAMP sous Linux Ubuntu 9 avec un serveur dédié d'hébergement Intel Core 2 duo 2.66 GHz et 2 Go de RAM DDR2.

Les langages de développement utilisés sont XHTML, CSS2, JavaScript (ajax), PHP5, JAVA, MySQL. PGR est implémenté en utilisant les techniques de développement Web2 les plus récentes (modèle MVC (modèle/vue/contrôleur), Framework JavaScript JQUERY, ..) tout en respectant les normes de standardisation du W3C.

2 Description fonctionnelle

Le modèle fonctionnel de PGR est décrit par la figure 1.

2.1.1 Parseur

Cet outil permet de transformer les fichiers de protéines PDB (Berman et al., 2000) en fichiers de graphes. Plusieurs formats peuvent être générés facilitant leur exploitation à l'aide d'autres outils existant tels que BioLayout (Theocharidis et al, 2009), GraphClust (Reforgiato et al, 2009)... Plusieurs méthodes de construction de graphes sont possibles, ces méthodes sont implémentées en java et détaillées dans (Saidi et al 2009). L'utilisation est très simple ;

l'utilisateur soumet sa liste des fichiers PDB, spécifie la méthode, ses paramètres et le format de graphe désiré et valide ses choix. Une description plus détaillée est reportée sur le site.

2.1.2 Entrepôt

L'entrepôt représente une banque de graphes de protéines accessible directement par les utilisateurs de PGR. Cet entrepôt est accompagné d'un outil de filtrage basé sur plusieurs critères permettant à l'utilisateur de cibler une population spécifiée de graphes de protéines. L'entrepôt est continuellement alimenté chaque fois qu'un utilisateur ou un administrateur utilise le parseur.

2.1.3 Extensions

Les travaux sur PGR sont encore en progression. Plusieurs autres extensions sont en cours de développement tel qu'un outil d'alignement, un outil d'extraction de motifs et de classification.

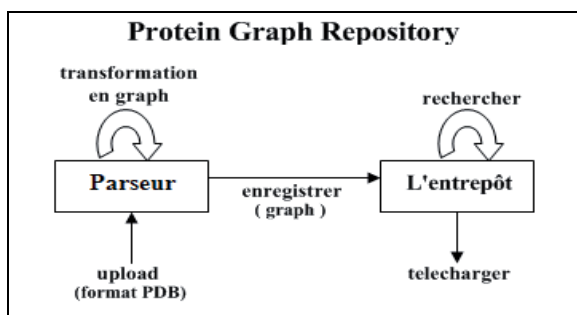


FIG. 1 – Représentation générale du fonctionnement de PGR

Références

- Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. & Bourne P.E. (2000). *The Protein Data Bank*. *Nucleic Acids Res*, 28 (1): 235-242.
- Diego Reforgiato & Rodrigo Gutierrez & Dennis Shasha, 2008. *GraphClust: A Method for Clustering Database of Graphs*. *Journal of Information & Knowledge Management (JIKM)*, World Scientific Publishing Co. Pte. Ltd., vol. 7(04): 231-241.
- Saidi, R., Maddouri, M., & Mephu Nguifo, E (2009). *Comparing graph-based representations of protein for mining purposes*. In *Proceedings of the Kdd-09 Workshop on Statistical and Relational Learning in Bioinformatics*, pp. 35-38.
- Theocharidis A., van Dongen S., Enright A.J. and Freeman T.C. *Network Visualisation and analysis of Gene Expression Data using BioLayout Express3D*. *Nature Protocols*; vol. 4 (10): 1535-50.