

Estimation de l'affinité entre deux molécules à l'aide des réseaux de fonctions de coûts

Thématique : Bioinformatique, Problème de Satisfaction de Contraintes, Optimisation combinatoire

Équipe d'accueil : Statistique et Algorithmique pour la Biologie

Laboratoire d'accueil : Mathématiques et Informatiques Appliquées Toulouse, Institut National de la Recherche Agronomique

Lieu : Auzeville-Tolosane (près de Toulouse), France

Encadrants : Thomas Schiex (tschiex@toulouse.inra.fr Tel : 05 61 28 54 28)

Gratification : environ 400 euros / mois

Contexte

Dans le domaine des biotechnologies, une partie importantes des efforts visent à la conception d'enzymes (des protéines définies par une séquence d'acides aminés, donc une séquence de caractères) permettant de catalyser des réactions chimiques (bio-énergies, médicaments...). Ces enzymes sont en effet respectueux de l'environnement, bio-dégradables et capables de fonctionner à des pressions et températures ambiantes, donc économes en énergie.

Le "Computational Protein Design" vise en particulier à faciliter le processus de conception expérimentale de telles protéines en modélisant les protéines et en essayant d'optimiser la probabilité qu'elles adoptent une structure tridimensionnelle choisie, qui définit leur fonction biochimique. En particulier, dans le cas d'interactions entre deux protéines (ou entre une protéine et un ligand), il est important de pouvoir estimer rapidement l'affinité entre ces molécules, un problème qui se ramène à un problème de dénombrement de solutions dans un réseau de fonctions de coûts (Seydou et al., 2013).

L'équipe d'accueil mène des travaux sur les réseaux de fonctions de coûts, *Weighted Constraint Satisfaction Problem* (WCSP) (Cooper et al, 2010), et met en oeuvre leur intégration dans une plate-forme logicielle *open-source* C++ *toulbar2* (<https://mulcyber.toulouse.inra.fr/projects/toulbar2/>) ayant remporté plusieurs compétitions (*UAI 2008, 2010, and 2011 Challenges* <http://www.cs.huji.ac.il/project/UAI10/> et <http://www.cs.huji.ac.il/project/PASCAL/board.php> *ficolofo*).

Le laboratoire d'accueil héberge également la plate-forme bioinformatique de la Génopole Toulouse Midi-Pyrénées (GENOTOUL <http://bioinfo.genotoul.fr/>) qui offre un support matériel (cluster 2000 cœurs) et informatique pour accompagner des projets en bioinformatique.

Sujet

Informatiquement, le problème de calcul d'affinité se réduit à un problème de dénombrement de solutions (un problème $\#-P$ complet), difficile. Un algorithme, appelé K^* (Georgiev et al, 2008), a été proposé par les biologistes structuraux pour accélérer ce calcul. Le but du stage est d'évaluer la capacité des techniques de filtrage par cohérence locale des réseaux de fonctions de coûts (ou CSP pondérés, (Cooper et al, 2010)) pour accélérer l'algorithme K^* dans un cas simplifié, en faisant l'hypothèse d'un squelette de protéine rigide, pour aboutir à une preuve de concept.

Des expérimentations seront menées sur des données issues d'une collaboration en cours avec des biologistes structuraux de l'INSA Toulouse.

Bibliographie

Cooper, M.C., de Givry, S., Sanchez, M., Schiex, T., Zytnicki, M., and Werner, T., Soft Arc-consistency revisited, *Artificial Intelligence*, 2010

Traoré, S., Allouche, D., André, I., de Givry, S., Katsirelos, G., Schiex, T. Barbe, S. A new framework for computational protein design through cost function network optimization. *Bioinformatics*, 29(17), p. 2129-2136, 2013.

Georgiev, I., Lilien, Ryan H, Donald, Bruce R, The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles. *Journal of Computational Chemistry*. 29, 10. p. 1527-1542, 2008.