

Compte rendu de l'avancement des travaux - 1ère année de thèse

David Mignon

29 octobre 2013

Sujet de thèse : Computational protein design (CPD) : un outil pour l'ingénierie des protéines et la biologie synthétique.

Le CPD est en développement au sein de notre laboratoire depuis déjà quelques années, avec plusieurs succès à son actif. Ce sont ces résultats prometteurs qui fondent notre motivation à aller de l'avant en améliorant nos outils, en enrichissant nos programmes pour progresser encore dans nos résultats. En particulier, mon travail se concentre sur le problème de l'exploration de l'espace des séquences d'acides aminées et son contrôle.

Une optimisation de type Monte Carlo ayant été ajoutée début 2012 à proteus (notre programme de recherche de séquences), ma thèse a débuté par une évaluation de l'apport que représente cette nouvelle fonction. Pour cela l'optimisation classique de proteus a été comparée avec une série de protocoles Monte Carlo variants selon différents paramètres (en premier lieu, la température). Puis, une tentative d'amélioration de l'état non plié de la protéine a été engagée, à partir d'une structure étirée. Malheureusement, cette approche s'est révélée non exploitable dans notre cadre.

Cependant, cela a mis en évidence des limites de proteus notamment dans la gestion d'états multiples d'une protéine. Alors, une nouvelle version du programme a été écrite. Elle permet de faire de l'optimisation avec différentes copies de parties du système étudié, via un mécanisme de groupes et de poids. Elle contient également une gestion plus fine de la mémoire, des calculs de l'énergie améliorés, etc. Cette version de proteus a été intégrée à la dernière version du projet qui a fait l'objet de la publication suivante :

Simonson T, Gaillard T, Mignon D, Schmidt Am Busch M, Lopes A, Amara N, Polydorides S, Sedano A, Druart K, Archontis G. 2013. Computational protein design : The Proteus software and selected applications. *Journal of Computational Chemistry*

Une série de comparaisons de plusieurs générations de séquences ont été effectuées, en particulier avec la protéine 1CKA. Ces générations ayant été obtenues avec des variations du modèle ou des variations sur la fonction d'énergie (présence ou non de ligand, différentes pondérations de l'affinité protéine-ligand versus la stabilité du complexe, etc).

Maintenant, une autre partie du projet est en travaux. Nous avons entrepris de modifier le programme X-plor (qui calcule les énergies de nos structures) pour y augmenter les possibilités de gestion des coordonnées d'atomes. Puis, X-plor va être parallélisé pour rendre possible le traitement de gros système et le backbone flexible. Ensuite, les multicopies de backbone et un nouveau déplacement Monte Carlo associé vont être ajoutés dans proteus afin que lui aussi puisse travailler avec un backbone flexible.