

PROJET LONG

Modélisation comparative par satisfaction de contraintes spatiale

Projet proposé par Jean-Christophe Gelly

Projet

Connaître la structure des protéines est indispensable pour comprendre leur évolution et leurs fonctions à un niveau moléculaire. Plus de 57 ans après la résolution expérimentale de la première structure protéique (Kendrew et al., Nature, 1958), la prédiction de la structure à partir de sa seule séquence est encore aujourd'hui un problème majeur des sciences biologiques. En effet la résolution de la structure des protéines de manière expérimentale est encore complexe et exige de gros investissement tant financièrement qu'en terme de temps. De plus les approches expérimentales ne rencontrent pas systématiquement le succès. Ainsi le nombre de séquences de protéines dans les banques croissent à une vitesse bien plus importante que la vitesse de résolution expérimentale de ces structures. Une méthode théorique capable de prédire la structure des protéines serait donc d'un intérêt majeur. Selon le journal Science, il s'agit d'un des 125 plus importants problèmes irrésolus des sciences (Norman et al, Science, 2005). Les applications de la prédiction de structure sont en effet considérables : en biologie fondamentale (pour comprendre les mécanismes responsables de la vie à l'échelle moléculaire), en médecine (afin d'élucider les modes d'action des pathologie), en pharmacologie (pour la conception rationnelle de médicament par exemple) ou encore en biotechnologie (pour la conception et l'ingénierie de nouvelles enzymes).

La modélisation comparative est actuellement la stratégie la plus utilisée et la plus efficace pour la prédiction de structure de protéines. L'objectif de la modélisation comparative est de prédire la structure d'une protéine d'intérêt en se basant sur une structure déjà résolue (provenant d'une protéine homologue). Les deux principales étapes sont (i) la détection d'une structure support parmi toutes les structures résolues expérimentalement (ii) la modélisation de la structure de la protéine d'intérêt en utilisant la structure support déterminer lors de l'étape I.

L'étape de la détection de cette structure support, étape appelée reconnaissance de repliement, est une étape primordiale de ce processus. Nous avons développé une nouvelle méthode de reconnaissance de repliement nommée ORION (Optimized fold RecognitION) (Ghouzam et al., Bioinformatics, 2015). Peu après notre méthode a bénéficié de l'incorporation de nouvelles améliorations qui ont permis d'augmenter encore la sensibilité (Ghouzam et al., Scientific reports, 2016).

La prédiction de la structure des protéines est un enjeu scientifique majeur et est l'objet d'une compétition internationale intense. Dans ce contexte une compétition internationale de prédiction de structure a été mise en place par la communauté scientifique : le Critical Assessment of protein Structure Prediction (CASP) qui se déroule tous les deux ans depuis 1994 et vise à évaluer les performances des meilleures équipes internationales.

Nous avons participé avec succès (1er groupe français) à deux éditions de la compétition (CASP 2014 et 2018) en utilisant ORION. Lors de la dernière édition nous avons réussi à déterminer la structure à la résolution quasi-atomique de cinq protéines et avons terminé premier sur ces cinq cibles parmi toutes les équipes internationales ayant participées. Le succès de notre stratégie se base sur l'identification de structures supports homologues lointaines ou partielles de la protéine d'intérêt grâce à ORION et à notre toute dernière méthode : nano-ORION.

Toutefois, un des problèmes de notre méthodologie est la construction du modèle de la structure d'intérêt à partir de la structure support. Actuellement la stratégie mise en œuvre repose principalement sur la modélisation comparative et se base sur les logiciels Modeller (Sali et al. J. Mol. Biol., 1993) et ROSETTA (Simons et al., J. Mol. Biol. 1997 ; Leaver-Fay et al., Methods Enzymol. 2011). Ces deux outils sont les plus utilisés dans le domaine de la prédiction de structures protéiques. Mais leur licence d'exploitation est restrictive et ils se trouvent limités face à certaines difficultés.

L'objectif du projet est donc de concevoir un nouveau programme de modélisation comparative. A partir d'un alignement de séquence entre une séquence cible à modéliser et une ou plusieurs séquences supports dont les structures sont connues, l'objectif du programme est de construire un modèle 3D de la protéine cible en utilisant les structures supports. Le programme reposera sur GROMACS (Lindahl et al., Molecular modeling annual 2001 ; Berendsen et al., Computer Physics Communications, 1995), le logiciel majeur dans le domaine de la mécanique moléculaire. Le programme conçu devra s'interfacer avec GROMACS pour réaliser les calculs complexes de mesure énergétique. GROMACS a l'avantage d'une licence libre (GNU Lesser General Public License, version 2.1.), d'un développement très soigné et de performances très élevées. Actuellement GROMACS ne permet pas de faire de la modélisation comparative, le stagiaire recruté sur ce projet aura pour objectif d'intégrer la modélisation comparative dans GROMACS (avec un niveau d'intégration qui dépendra de l'avancée du projet).

Nous avons déjà une préversion de la méthode qu'il est nécessaire de revoir.

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 série de partitions emboîtées 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".

L'objectif du projet est de proposer une nouvelle méthode de *Protein Peeling* en utilisant non pas une segmentation hiérarchique mais en recherchant itérativement les régions de la protéine les plus compactes et les plus indépendantes.

Informations supplémentaires

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