



PROTÉINES : MODÉLISATION, PRÉDICTION ET CONCEPTION

Markus Lindström

Département d'Informatique

Les protéines sont des macromolécules aux rôles multiples dans l'organisme : anticorps, enzymes de digestion, etc.

Ce sont des chaînes de blocs élémentaires appelés *acides aminés* dont il existe 20 sortes ; il est possible de caractériser n'importe quelle protéine en donnant sa *structure primaire*, autrement dit une suite de tels acides.

La structure primaire ne fournit malheureusement aucune information sur la fonction de la protéine dans l'organisme. Il faut pour cela déterminer la *conformation* de celle-ci, c'est-à-dire sa structure tridimensionnelle.

Dans le sillage du *Projet Génome Humain* notamment, il est de plus en plus utile de savoir prédire la conformation d'une protéine donnée, ou comment elle se *replie* sur elle-même. En effet, la détermination expérimentale de celle-ci est onéreuse, et le séquençage de l'ADN humain nous a fourni un nombre énorme de structures primaires à analyser.

Nous présenterons brièvement deux méthodes de prédiction de la structure d'une protéine : *ab initio* (ou *de novo*) se basant uniquement sur les principes fondamentaux de la physique, et la *modélisation comparative* exploitant des données obtenues expérimentalement sur les protéines.

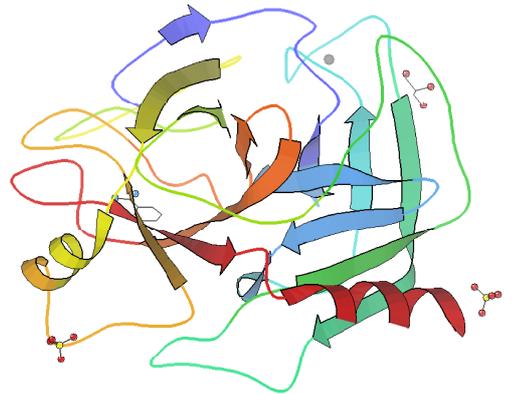


Illustration d'une protéine typique, la trypsine cationique (2BY7).

ab initio / de novo

Ces approches se basent sur l'idée qu'une protéine a naturellement tendance à vouloir atteindre sa structure native en minimisant son *énergie libre*. Il faut *modéliser* la protéine pour en simplifier la représentation, et se définir une *fonction potentielle* associant à chaque conformation du modèle une énergie libre.

Sur base de ces fondations, on peut se livrer à une *recherche conformationnelle* qui consiste à trouver la conformation native.

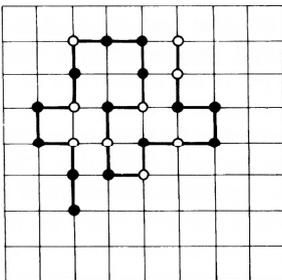
Cette approche idéaliste ne fonctionne actuellement que rarement en pratique dès que la structure primaire compte plus de quelques acides aminés.

Modélisation comparative

Comme son nom le suggère, cette approche consiste à comparer une séquence primaire inconnue avec des protéines dont celle-ci et la conformation sont bien déterminées par expérimentation, contrairement aux méthodes *ab initio*.

L'idée générale est de quantifier à quel point la séquence dont on souhaite obtenir la conformation « ressemble » à une autre. En effet, des similitudes entre les structures primaires impliquent en général des conformations proches.

Des algorithmes bien connus pour résoudre ce problème sont BLAST et FASTA, qui exploitent tous deux des *bases de données structurales* regroupant sous forme électronique toutes les conformations connues.



○ = hydrophilic ● = hydrophobic

Exemple de modèle complet utilisé en prédiction *ab initio*, le modèle HP de Chan et Dill.

D'accord, mais quel intérêt ?

Les méthodes décrites plus haut permettent d'approcher le problème inverse : si nous souhaitons obtenir une protéine avec certaines fonctions biologiques bien choisies, comme par exemple l'interaction avec certains agents infectieux, pouvons-nous obtenir une structure primaire menant à celle-ci ?

Cette question est au cœur de la recherche concernant la *conception de protéines*, un domaine aux applications multiples telles le développement de nouvelles thérapies biotechnologiques, de matériaux, et de systèmes à l'échelle nanométrique.

