

## **Application à la comparaison de structures de protéines de la recherche du plus grand sous-graphe isomorphe**

Professeur encadrant : Pascal Vallaëys, professeur de mathématiques au lycée Faidherbe

### Positionnement thématique

INFORMATIQUE, MATHÉMATIQUES (Géométrie, Théorie des graphes)

Mots-clés

Mots-clés(français)

Mots-clés(anglais)

Isomorphisme/automorphisme de graphes      Graph isomorphism/automorphism

Identification de sous-graphes isomorphes Subgraph isomorphism

Repliement des protéines

Protein folding

### Bibliographie commentée

La comparaison de graphes est un problème d'informatique classique, qui, en plus d'avoir des utilités pratiques, s'avère être un problème théoriquement très intéressant du point de vue informatique.

Scott Fortin fait état de la théorie sur le problème dans son papier [3], où il présente notamment une manière, liée à l'algèbre, de comparer les graphes établie par McKay [4].

Un pilier dans le domaine est Lazlo Babai, qui a proposé en 2015, le meilleur algorithme répondant au problème, de complexité quasi-polynomiale ( $O(\ln(n)^c)$ ).

Un problème plus spécifique, adapté à la comparaison de structures de deux graphes  $G$  et  $H$  [2], est la recherche des plus grand sous-graphe induits de  $G$  et  $H$  isomorphes [1].

Une fois encore, des hypothèses fortes sur les graphes (de degré borné, planaires, typage des sommets ...) vont permettre de résoudre en temps polynomial, sans lesquelles la complexité est exponentielle.

### Problématique retenue

Il s'agit de comparer les structures de protéines lorsqu'elles ont formé leurs liaisons et se sont formées dans l'espace, et déterminer leur proximité tant sur la forme que sur la position.

Un outil de comparaison permettrait de mesurer la précision des prédictions informatiques d'un modèle, par rapport à un modèle référent dont la structure a été déterminée par une méthode longue et coûteuse (par rayons X par exemple).

### Objectifs

Je me propose:

- d'étudier des modèles simples, une suite de points sans ramifications par exemple, et déterminer leur proximité par différentes méthodes
- de me familiariser avec la structure des protéines et le format pdb,

- de proposer des moyens de comparer des protéines (graphes), en commençant par écrire un algorithme naïf de détection d'isomorphismes de graphes, améliorer en utilisant des invariances entre les sommets.

Puis plus élaboré, notamment en utilisant la méthode McKay [4], en appréhendant les concepts mathématiques des preuves, pour finalement adapter ces concepts à la recherche du plus grand sous-graphe isomorphe.

( Bonus : recherche des plus grands-sous graphes isomorphes -> couverture maximale des 2 graphes par plusieurs sous-graphes isomorphes (plus difficile), cela permettrait d'amortir un phénomène de "cassure" )

- d'établir un indice de comparaison, en fonction de la ressemblance de structure et de position des atomes dans l'espace

- d'en estimer l'utilité et de comparer avec les moyens actuellement mis en oeuvre, notamment à la compétition CASP (Global Distance Test)

### Bibliographie

[1] Subgraph Isomorphism in Planar Graphs and Related Problems, David Eppstein, 1999

[2] Measuring the similarity of labeled graphs, Pierre-Antoine Champin, Christine Solnon, 2003

[3] The Graph Isomorphism Problem, Scott Fortin, 1996

[4] B. McKay. Practical graph isomorphism. Congressus Numerantium, 1981