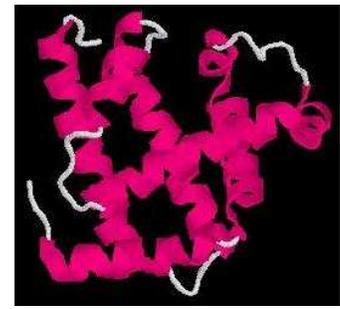


Structure des protéines :  
descripteurs de formes  
et applications

Proposition de stage de DEA  
2003-2004



**Titre du stage :** Structure des protéines : descripteurs de formes et applications

**Mots-clés :** Biologie structurale, Protéines, Structure Tertiaire, Repléments, Géométrie Algorithmique

**Durée :** 4-6 mois, à partir de Février 2004

**Encadrant :** Frédéric Cazals

**Fonction :** Chargé de Recherches INRIA, projet Geometrica

**Laboratoire :** Project Geometrica, INRIA Sophia-Antipolis, 2004 route des Lucioles, – BP 93, F-06902 Sophia-Antipolis

**Téléphone :** +33 (0)4 92 38 71 88

**Télécopie :** +33 (0)4 92 38 76 43

**Email :** Frederic.Cazals@inria.fr

**www** <http://www.inria.fr/prisme>, <http://www.inria.fr/prisme/personnel/cazals>

## 1 Contexte

L'un des grands problèmes de la biologie structurale est la prédiction du repliement des protéines — à partir de la séquence d'acides aminés, et une telle prédiction nécessite une bonne compréhension des structures tridimensionnelles admissibles de chaînes polypeptidiques.

Si les éléments de structure secondaire sont connus depuis longtemps, la modélisation géométrique de leurs interactions au sein de la structure tertiaire est plus récente. On citera par exemple les tableaux de contact entre éléments de structure secondaire [L95], les matrices d'aires SAS [AT97, MW03], certains invariants topologiques de la théorie des noeuds [BMT02, E03], ainsi que des notions plus géométriques liées à la définition d'un voisinage tubulaire autour du *backbone* de la protéine [GM99].

## 2 Description détaillée du travail

L'objet de ce stage sera :

- d'envisager de façon plus précise la définition de motifs structuraux, tout en proposant un classement multi-résolution de ceux-ci.
- d'étudier la similarité entre protéines de façon plus locale — en prenant en compte les éventuels contacts entre chaînes latérales,
- de replacer les outils développés dans le contexte du repliement des protéines.

Les outils algorithmique et mathématiques utilisés dériveront des fonctions distances et de l'étude de leurs singularités — théorie de Morse, diagrammes de Voronoï.

### 3 Bibliographie

- L95** A Lesk, Systematic representation of protein folding patterns ; Journal of Molecular Graphics, 13, 1995.
- GM99** Oscar Gonzalez and John H. Maddocks, Global curvature, thickness, and the ideal shapes of knots ; PNAS (96) 1999.
- BMT02** JR Banavar and C Micheletti and A Trovato, Geometry and physics of proteins, Proteins—structure function and genetics ; May 15 2002 ; v.47.
- P71** Porteous, The normal singularities of a submanifold, J. Diff. Geom., 5, 543-564, 1971.
- E03** Michael A. Erdmann, Protein Similarity from Knot Theory and Geometric Convolution, TR-CMU-CS-03-181, September 2003.
- AT97** R.A. Abagyan and M.M. Totrov, Contact Area Difference (CAD) : A robust measure to Evaluate accuracy of protein models, JMB, 268, 1997.
- MW03** J. Mintseris and Z. Weng, Atomic contact vectors in Protein-Protein recognition, Proteins : structure function and genetics (2003).

### 4 Environnement informatique

**Système, machines** PC Linux

**Géométrie** bibliothèque CGAL —[www.cgal.org](http://www.cgal.org)