

## OFFRE de THÈSE

### Sujet de thèse : Modelling Protein Structure

#### Description :

The protein 3D structure is the spatial conformations in which the protein sequence folds. The biological function of a protein depends on its 3D structure and its knowledge is highly important in medicine (for example, in drug design) and biotechnology (for example, in the design of novel enzymes). Many of available structures are available in the Protein Data Bank as sets of the atoms coordinates, determined by costly techniques such as X-ray crystallography. Prediction of the structure of proteins is one of the most important goals pursued by bioinformatics and theoretical chemistry. The set of atoms coordinates contains the raw information but is not the more adequate for automatic prediction or characterizations based on the structures of proteins. In the team we have been able to get good or promising results by using alternative representations, such as protein contact maps, for the comparison of structures and their automatic classification [1] and pattern discovery and grammatical inference [2]. Protein contact maps represent structures of proteins by a binary relation between the residues which are close enough. The subject of the thesis is to continue this new line of research and to study how this kind of new representations of the structure can offer a good level of abstraction (without loss of important information) allowing to improve the quality of the result and the efficiency of the programs on protein structures.

**Département :** D7 - Gestion des données et de la connaissance

**Equipe :** SYMBIOSE ; <http://www.irisa.fr/symbiose/positions>

**Directeur de thèse :** Jacques Nicolas

**Encadrant(s) :** François Coste

**Contact :** [francois.coste@irisa.fr](mailto:francois.coste@irisa.fr) ; tel (33)0 299 847 491

**Début des travaux :** Dès que possible

#### Bibliographie :

[1] Maximum Contact Map Overlap Revisited, Rumen Andonov, Noël Malod-Dognin, Nicola Yanev, Journal of Computational Biology 18, 1 (2011) 1-15

[2] Towards pattern discovery and grammatical inference in protein folds, François Coste, Fabio Cunial, techn. rapport, Symbiose, 2010.

[3] Vassura M, Margara L, Di Lena P, Medri F, Fariselli P, Casadio R (2008). "Reconstruction of 3D Structures From Protein Contact Maps". IEEE/ACM Transactions on Computational Biology and Bioinformatics 5 (3): 357–367

[4] Lo A, Chiu YY, Rødland EA, Lyu PC, Sung TY, Hsu WL. (2009). "Predicting helix-helix interactions from residue contacts in membrane proteins". Bioinformatics 25 (8): 996–1003