Ecole Doctorale STIC

Sciences & Technologies de l'Information et de la Communication

ED STIC - Proposition de Sujets de Thèse

pour la campagne d'Allocation de thèses 2015

Axe Sophi@Stic :	BioSanté|
Titre du sujet :	Towards nano-molecular design : advanced algorithms for modeling large protein assemblies
Mention de thèse :	Informatique
HDR Directeur de thèse inscrit à l'ED STIC :	Frederic Cazals

Co-encadrant de thèse éventuel :

Nom :	
Prénom :	
Email :	
Téléphone :	

Email de contact pour ce sujet :	frederic.cazals@inria.fr
Laboratoire d'accueil :	INRIA

Description du sujet :

Voir description en Anglais.

English version:

CONTEXT

A number of key biological functions are accomplished by large molecular machines, involving from tens to hundreds of macro-molecules. Constructing models of such machines is a key endeavor, with applications in fundamental biology and in nano molecular design. Yet, no atomic resolution model is known to date for most large assemblies. Typically, the experimental data available are either large scale (they concern the whole assembly) but low resolution, or local

scale (they concern a sub-system) and high resolution. In the former category, density maps are of special interest, and the goal of this thesis will consist of leveraging the reconstruction of models from such maps.

GOALS

In cryo-electron microscopy, a density map is a 3D matrix with one intensity per voxel, encoding the electron density in the neighborhood of the voxel. Due to the plasticity of assemblies (which do not always contain the same set of proteins) and the flexibility of the constituting proteins, these maps feature a low signal to noise ratio, so that their processing typically involves greedy region growing algorithms [A-07] and/or to clustering algorithms [L-09]. The goal of this thesis will be twofold.

First, we shall develop novel techniques, combining Morse theory and image coding theory, to reconstruct secondary structure elements (SSE) in density maps. SSE indeed provide anchor points to embed
proteins with known crystal structures into a map [L-09]. SSE will also be used to rule out false positive contacts inferred from mass spectrometry data [A-15].

Second, we shall work on the problem of sampling conformations of a protein, to optimize its location in a map upon performing the aforementioned registration. This step will exploit novel sampling techniques for energy landscapes of macro-molecules [C-15].

BACKGROUND

The PhD candidate should have a strong background in theoretical computer science or applied mathematics or biophysics or bioinformatics, and a genuine interest for (structural) biology.

BIBLIOGRAPHY

[A-07] F. Alber et al; Determining the architectures of macromolecular assemblies; Nature 450, 2007.

[L-09] K. Lasker and M. Topf and A. Sali and H.J. Wolfson; Inferential Optimization for Simultaneous Fitting of Multiple; Components into a CryoEM Map of Their Assembly; Journal of Molecular Biology, 2009.

[A-15] D. Agarwal and C. Caillouet and F. Cazals and D. Coudert,
Unveiling Contacts within Macro-molecular assemblies by solving Minimum Weight Connectivity Inference Problems,
Molecular and Cellular Proteomics, 2015.

[C-15] F. Cazals and T. Dreyfus and D. Mazauric and A. Roth and C.H. Robert, Conformational Ensembles and Sampled Energy Landscapes: Analysis and Comparison, Journal of Computational Chemistry, 2015.

MISC

The Algorithms-Biology-Structure group from INRIA Sophia-Antipolis-Méditerranée: http://www-sop.inria.fr/abs