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 2011

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 :	Cazals Frederic

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 developing new algorithms to handle

cryo-EM maps and probability density maps.

GOALS

In cryo-electron microscopy, a density map is a 3D matrix with one intensity per voxel, encoding the local density of matter in the voxel. In molecular modeling, a probability density maps is a 3D matrix with one probability per voxel, encoding the probability to have this voxel covered by a protein. 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.

On the theoretical side, new techniques to analyze, partition and model maps will be developed, based on recent results in computational topology (Morse theory) [C-11] and geometric optimization algorithms [C-10,L-11].

On the applied side, the techniques will be used for docking atomic resolution models within a map, a complex 3D jigsaw puzzle. Tests will be conducted on transcription complexes, in collaboration with the group of P. Schultz, at the Institut of Genetique et Biologie Moleculaire et Cellulaire, in Strasbourg, and on viral complexes, in collaboration with the group of F. Rey at Institut Pasteur Paris.

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.

[C-10] F. Cazals and T. Dreyfus; Multi-scale Geometric Modeling of Ambiguous Shapes with Toleranced Balls and Compoundly Weighted alpha-shapes; Symposium on Geometry Processing, 2010.

[C-11] F. Cazals and D. Cohen-Steiner; Reconstructing 3D compact sets; Computational Geometry Theory and Applications, 2011, to appear.

[L-11] S. Loriot and S. Sachdeva and K. Bastard and C. Prevost and F.Cazals; On the

Characterization and Selection of Diverse Conformational Ensembles; IEEE/ACM Transactions on Computational Biology and Bioinformatics; 8(2), 2011.

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

-- full description of the PhD thesis:
ftp://ftp-sop.inria.fr/abs/fcazals/positions/thesis11_large_assemblies.pdf

URL : ftp://ftp-sop.inria.fr/abs/fcazals/positions/thesis11_large_assemblies.pdf