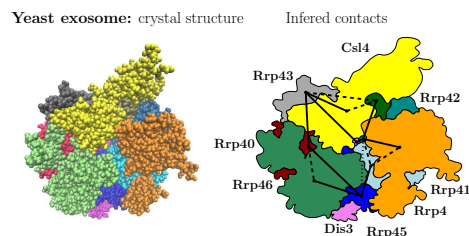Lab.: Inria Sophia-Antipolis-Méditerranée
Group: Algorithms-Biology-Structure
Supervisor: Dorian.Mazauric@inria.fr
Co-supervisor: Frederic.Cazals@inria.fr
Web: `http://team.inria.fr/abs`

Yeast exosome: crystal structure   Infered contacts

PHD THESIS PROPOSAL

COMBINATORIAL OPTIMIZATION TECHNIQUES LEVERAGE
THE RECONSTRUCTION OF ATOMIC RESOLUTION MODELS OF LARGE PROTEIN ASSEMBLIES

Context. An antibody-antigen complex or an enzyme-inhibitor complex are example protein complexes involved in key biological functions, respectively the immune response and catalysis. The central problem in structural biology is to understand how the function of these molecular machines emerge from their structure and dynamics. These studies are especially challenging since they require structural data, which are pretty scarce: as of today, the protein data bank contains circa 100,000 structures, while millions of coding sequences are known. Even worse, few of these structures concern protein complexes–the vast majority concern isolated molecules.

Goals. The goal of this thesis is to develop novel methods to design atomic resolution models of molecular assemblies involving from 3 to 10 subunits. In such cases, two complementary sources of information are usually available [ADV+07]: on the one hand, high resolution (i.e. atomic resolution) crystal structures of the isolated subunits; on the other hand, low resolution data for the whole assembly, typically coming from electron microscopy or mass spectrometry [ACCC15]. A natural strategy therefore consists of combining these sources of information [ACW15], to hopefully reach atomic resolution for the whole assembly. Doing so involves two mains steps, namely computing conformations of subunits compatible with low resolution data [CDM+15, RDRC16], and assembling these conformations to build the assembly [ACW15]. Both steps can be phrased as optimization and enumeration problems involving graphs, namely problems in the realm of graphical problems, and so graph algorithms techniques will be developed.

The goal of this thesis will be to contribute to such methods, and to validate them on systems involving from 3 to 10 subunits.

Background. Master in theoretical computer science, or bioinformatics/biophysics, or applied mathematics.

References

[ACCC15] D. Agarwal, C. Caillouet, D. Coudert, and F. Cazals. Unveiling contacts within macro-molecular assemblies by solving minimum weight connectivity inference problems. *Molecular and Cellular Proteomics*, 14:2274–2282, 2015.

[ACW15] N. Amir, D. Cohen, and H. J. Wolfson. Dockstar: a novel ilp-based integrative method for structural modeling of multi-molecular protein complexes. *Bioinformatics*, 31(17):2801–2807, 2015.

[ADV+07] F. Alber, S. Dokudovskaya, L. M. Veenhoff, W. Zhang, J. Kipper, D. Devos, A. Suprapto, O. Karni-Schmidt, R. Williams, B.T. Chait, M.P. Rout, and A. Sali. Determining the Architectures of Macromolecular Assemblies. *Nature*, 450(7170):683–694, Nov 2007.

[CDM+15] F. Cazals, T. Dreyfus, D. Mazauric, A. Roth, and C.H. Robert. Conformational ensembles and sampled energy landscapes: Analysis and comparison. *Journal of Computational Chemistry*, 36(16):1213–1231, 2015.

[RDRC16] A. Roth, T. Dreyfus, C.H. Robert, and F. Cazals. Hybridizing rapidly growing random trees and basin hopping yields an improved exploration of energy landscapes. *Journal of Computational Chemistry*, 37(8):739–752, 2016.