

# Large Macromolecular Assembly Modeling by Computational Integration of Data from Various Experimental Sources

Haim J. Wolfson

Tel Aviv University, Israel, [wolfson@post.tau.ac.il](mailto:wolfson@post.tau.ac.il)

Modeling of multimolecular assemblies is crucial for the understanding of cellular function. Nevertheless, most of the structures in the PDB are either monomers or dimers. The yeast cell, for example, contains approximately 800 distinct core complexes of 5 proteins, on the average, most of which have not yet been structurally characterized. In parallel, the various Structural Genomics efforts and the improvement in homology modeling techniques provide a wealth of single protein atomic resolution structures, on the one hand, and recent developments in experimental techniques, such as cryo-EM or SAXS provide low resolution information on large macromolecular complexes, on the other hand. Development of efficient algorithms, which integrate high and low resolution data, in order to model macromolecular complexes at atomic resolution, is a key Structural Bioinformatics task.

We have developed several integrative algorithms to tackle the multi-molecular assembly task, among them **Combinatorial Docking** with distance constraints, and **Ematch**, which fits high resolution structures into medium resolution cryo-EM density maps. Currently, we are actively developing the **MultiFit** algorithm, which accepts as input a low resolution ( $\sim 20\text{\AA}$ ) density map of the multi-molecular assembly and either atomic resolution structures of the participating proteins or their homology models. By applying a cascade of (multi-resolution) shape fitting, protein-protein docking and graph model based optimization algorithms the multi-molecular "puzzle" is assembled. A prototype of the algorithm has been implemented and performed well on a benchmark of simulated structures as well as on few available experimental structures. Ongoing challenges include the handling of significant conformational flexibility, "missing" parts, ranking of the candidate solutions, and exploitation of systems biology interaction network data as well as other experimental sources.