DockStar: A Novel ILP Based Integrative Method for Structural Modelling of Multimolecular Protein Complexes (Extended Abstract)

Naama Amir, Dan Cohen, and Haim J. Wolfson(✉)

Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel
wolfson@tau.ac.il, naamaamir@mail.tau.ac.il

Introduction. Atomic resolution modelling of large multimolecular protein complexes is a key task in Structural Cell Biology. A single cell consists of hundreds of different functional complexes. To understand how these complexes operate and to develop strategies to modulate them for therapeutic purposes, we need to describe their 3D structure. However, high resolution experimental techniques, such as X-ray crystallography and NMR spectroscopy cannot handle very large complexes, whereas electron microscopy or mass spectrometry produce low resolution data [1]. It is becoming clear that integration of data derived from a variety of bio-physical techniques at multiple levels of resolution is essential for the structural analysis of large complexes.

We introduce DockStar, a novel Integer Linear Programming (ILP) based algorithm for modelling protein complexes which integrates both low and high resolution experimental data. The optimal assembly of the subunits is formulated as an ILP task, thus, enabling simultaneous assignment of a position per subunit in the complex. This enables efficient handling of relatively large assemblies. The method demonstrated good results both in bound and unbound cases and significantly outperformed other leading multimolecular docking methods.

Methods. The algorithm accepts as input atomic resolution structures of the individual subunits obtained from X-ray, NMR or homology modelling, and low resolution data of interaction between the complex subunits and of cross-links data obtained by mass spectrometry. The algorithm first generates a set of candidate transformations for each subunit by one of three methods: (i) a pairwise (soft) docking algorithm for neighbouring subunits,(ii) aligning the subunits to a homologue complex with known 3D structure or (iii) fitting the subunits to a cryo-EM map [2]. Then, for each two candidate transformations of different subunits the resulted subunit interaction is scored according to a knowledge-based potential and satisfaction

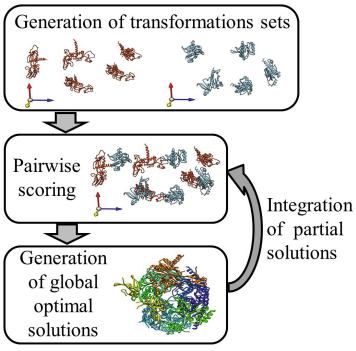

Fig. 1. Flowchart of the DockStar algorithm

T. Przytycka (Ed.): RECOMB 2015, LNBI 9029, pp. 13–15, 2015.
DOI: 10.1007/978-3-319-16706-0_2

of the cross-linking restraints. Finally, globally optimal multimolecular complex hypotheses are assembled by formulating the task as an ILP. When the method of choice for generating candidate transformations sets is docking, the algorithm is limited to complexes which interaction graph has a star shaped spanning tree. In such cases, the resulted solution might not cover the whole complex. Therefore, top solutions of intermediate star shaped subcomplexes are integrated to produce a solution which covers the whole assembly. The integration is done by translating the top intermediate solutions to transformations sets and repeating the two last steps of the algorithm (Fig. 1).

Results. The method was tested on several representative complexes, both in the bound and unbound cases. It placed correctly most of the subunits of multimolecular complexes of up to 16 subunits (Table 1). DockStar was compared with the state of the art Haddock [4] and CombDock [3] multimolecular assembly algorithms and proved to be significantly more time efficient than the other methods, while exhibiting better performance.

Table 1. Summary of the DockStar's Results

Target Complex	B(Bound)/ U(Unbound)	Units Num.	Rank	Global Cα-RMSD[a]	Contacts Num.[b]	Predicted Contacts[c]	Run Time HH:MM
PP2A	B	3	1	0.68	2	2	00:34
	U	3	1	6.9	2	2	00:42
Beef Liver	B	4	1	0.85	3	3	02:51
Catalase	U	4	1	2.7	3	3	03:53
RNA	B	11	1	7.9	10	9	04:53
polII	U	11	3	4.8	10	8	04:51
Yeast	B	10	1	5.1	9	7	10:34
Exosome	U	10	12	6.0	9	4	11:22

[a] Global Cα-RMSD between the predicted and the native assemblies including only predictions with i-RMSDbb $\leq 8.0\mathring{A}$

[b] Number of contacts in the spanning tree of the complex interaction graph.

[c] Num. of contacts in the spanning tree of the complex interaction graph of the predicted complex that have i-RMSDbb$\leq 8.0\mathring{A}$ (in most cases the i-RMSDbb$\leq 4.0\mathring{A}$).

Funding: This research was supported by the Israel Science Foundation (grant No. 1112/12), the I-CORE program of the Budgeting and Planning Committee and the Israel Science Foundation (center No. 1775/12), and by the Minkowski Minerva Geometry Center. N.A. acknowledges the E.J. Safra Bioinformatics Center fellowship.

References

1. Alber, F., et al.: Determining the architectures of macromolecular assemblies. Nature **450**(7170), 683–694 (2007)
2. Cohen, D., et al.: 3D-Mosaic: An efficient method for integrative modeling of large multimolecular complexes (to be submitted, 2015)

3. Inbar, Y., et al.: Prediction of multimolecular assemblies by multiple docking. Journal of Molecular Biology **349**(2), 435–447 (2005)
4. Karaca, E., et al.: Building macromolecular assemblies by information-driven docking introducing the Haddock multibody docking server. Molecular & Cellular Proteomics **9**(8), 1784–1794 (2010)

http://www.springer.com/978-3-319-16705-3

Research in Computational Molecular Biology
19th Annual International Conference, RECOMB 2015,
Warsaw, Poland, April 12-15, 2015, Proceedings
Przytycka, T.M. (Ed.)
2015, XVII, 368 p. 110 illus., 98 illus. in color., Softcover
ISBN: 978-3-319-16705-3