MultiFit: a web server for fitting multiple protein structures into their electron microscopy density map

Elina Tjioe[1,2], Keren Lasker[1,2,3], Ben Webb[1,2], Haim J. Wolfson[3,*] and Andrej Sali[1,2,*]

[1]Department of Bioengineering and Therapeutic Sciences, [2]Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences (QB3), University of California, San Francisco, CA, USA and [3]Blavatnik School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel

ABSTRACT

Advances in electron microscopy (EM) allow for structure determination of large biological assemblies at increasingly higher resolutions. A key step in this process is fitting multiple component structures into an EM-derived density map of their assembly. Here, we describe a web server for this task. The server takes as input a set of protein structures in the PDB format and an EM density map in the MRC format. The output is an ensemble of models ranked by their quality of fit to the density map. The models can be viewed online or downloaded from the website. The service is available at; http://salilab.org/multifit/ and http://bioinfo3d. cs.tau.ac.il/.

SIGNIFICANCE

Macromolecular assemblies are involved in nearly all cellular processes. Determining the structures of these biological machines is crucial for deciphering their function. Recent advances established electron microscopy as a central technique for studying the structures of macromolecular assemblies in different functional states *in vitro* and *in vivo*. Because the resolution of an electron microscopy density map is relatively low, fitting of atomic resolution component structures into the density map of the whole assembly is essential. MultiFit is the first web server for achieving this task.

INTRODUCTION

Recent advances have established electron microscopy (EM) as a central technique for studying the structures of macromolecular assemblies in different functional

states *in vitro* and *in vivo* (1). The resolution of an EM density map is typically better than 25 Å, and can be as high as ~4 Å for highly symmetric structures (2,3). In most cases, however, the resolution is insufficient to construct a full atomic model of a protein complex. To this end, fitting of atomic resolution structures into an EM density map of the whole assembly is essential (4–8).

In the past decade, different algorithms have been developed for fitting a single protein subunit into its density map (9–20). Most methods use a variant of the cross-correlation coefficient as the quality-of-fit measure (21). The position of a protein subunit inside the density map is sampled either exhaustively or by matching precalculated geometric features. Methods for fitting multiple components of large assemblies have also been recently described (22–25). In particular, we have developed the MultiFit module of the Integrative Modeling Platform (IMP, http://www.salilab.org/imp/) software package (23,26). MultiFit simultaneously positions protein subunits into a density map of a protein assembly by combining geometric criteria commonly used in molecular docking and quality-of-fit criteria commonly used in EM fitting. The method was validated in the 2010 EM modeling challenge (http://ncmi.bcm.edu/challenge/).

Here, we present a web interface to MultiFit. The server takes as input a set of protein structures in the PDB format and an EM density map in the MRC format. The output is an ensemble of models ranked by their quality of fit to the density map. The models can be viewed online or downloaded from the website.

THE MULTIFIT METHOD

MultiFit is a method for simultaneously fitting atomic-resolution protein structures into their assembly density

map at resolutions as low as 25 Å. The input is a set of atomic structures of proteins and an EM density map of their assembly. The component positions and orientations are optimized with respect to a scoring function that includes the quality-of-fit of components in the map, the protrusion of components from the map envelope and the shape complementarity between pairs of components. The scoring function is optimized by an exact inference optimizer DOMINO (Discrete Optimization of Multiple INteracting Objects) that efficiently finds the global minimum within a discrete sampling space. Specifically, the optimization algorithm is composed of four stages, each sampling assembly models at increasingly higher resolution and accuracy. In 'anchor graph segmentation' stage, an unlabeled segmentation of the density map into regions is calculated using a Gaussian mixture model; the segmented regions correspond approximately to the subunits in the complex. In 'fitting-based assembly configuration' stage, a set of coarse assembly models is found by an enumeration over possible assignments of subunits to regions, followed by simultaneous local fitting of the subunits in the corresponding regions. In 'docking-based pose refinement' stage, each of the models found in the 'configuration' stage is refined by simultaneous local

optimization of the interfaces between pairs of interacting subunits as sampled by local pairwise docking. In 'rigid body minimization' stage, each of the models found in the 'refinement' stage is further refined using a local Monte Carlo/conjugate gradients minimization procedure. The default run of the MultiFit web server omits the final refinement stage. Users can explore the ensemble of solutions generated by the first three stages and then refine a subset of the ensemble using a downloaded version of MultiFit. For cyclic symmetric complexes, the symmetry is imposed within the optimization procedure for improved efficiency, such that only symmetric models are sampled. In particular, in 'fitting-based assembly configuration' and 'docking-based pose refinement', only cyclic symmetric models consistent with the symmetry of the density map are sampled (26).

WEB SERVER

Input

The MultiFit web server requires as input a set of protein structures in the PDB format, an EM density map of their assembly in the MRC format, and a few

Figure 1. Snapshots of the MultiFit web server. (**A**) Input page. The inputs are divided into three parts: (i) general information, (ii) density map information and (iii) protein complex information. Seven copies of the GroEL chaperon monomer [PDB entry 1oel (33)] are simultaneously fitted to its ring density map at 11.5-Å resolution [EMDB entry 1080 (34)] using cyclic symmetry mode. The input subunit PDB file and the input assembly density map used for this example can be obtained from the MultiFit web server help page. The input parameters for the resolution, spacing, contour level and symmetry order obtained from EMDB site are 11.5, 2.7, 0.852 and 7, respectively. The optional parameters for *X*, *Y* and *Z* origins are set to –50, –50 and –50, respectively. (**B**) Output page. The top 20 assembly models of the GroEL chaperon complex are ranked according to the quality-of-fit score from top left to bottom right. The user can click on the model thumbnail to open it using Chimera for further analysis. The PDB files and the transformation output file can be downloaded. Job results will be available for 6 days. (**C**) Top scored structural model of the GroEL ring fitted into the density map.

parameters (Figure 1). The parameters for the density map include: (i) resolution (Å) (27); (ii) voxel spacing on the grid representing the map (Å); and (iii) the contour level that results in the volume accommodating the molecular mass of the complex. These parameters are included for maps deposited in the EM Data Bank (EMDB) (28).

The MultiFit web server operates in two modes: cyclic-symmetric and non-symmetric. In the cyclic-symmetric mode, the symmetry order should be provided (2 for dimer, 3 for trimer, etc.). If the arrangement of the input monomers in its native complex follows a different type of symmetry, the user should use the downloaded version of MultiFit. In the non-symmetric mode, a list of subunit PDB files and the number of copies of each subunit are required. The input density should be pre-segmented to contain only the input set of proteins.

The server also has an optional input parameter specifying an e-mail address to which a link to the results page will be sent once the job is completed. Alternatively, the user can bookmark a web link to the results page at the time of data submission. The status of the job (queued, running or finished) can be accessed on the queue page.

Output

The computation is performed in real time and the server page is updated once the calculation has finished. The typical running time is about 20 min for assemblies with tens of thousands of atoms. The web server output page displays a table of the top 20 assembly models that best fit the assembly density map, along with their quality-of-fit scores ranked from top left to bottom right (Figure 1). MultiFit lists the optimal as well as suboptimal solutions; when the latter have good scores and are different from the optimal solution, the user should be skeptical about all solutions and further analyze the ensemble.

Each model can be saved as a PDB file and can also be directly opened with UCSF Chimera (19). A compressed file containing all models is available for download. Moreover, the MultiFit output text file can be downloaded. Row *i* lists the transformation applied to each of the subunits, the model quality-of-fit score, and the geometric complementarity score for model *i*. This output file can be used as input to IMP for further refinement and analysis. It can also be used as input for refining symmetric complexes using the SymmRef method (29).

CONCLUSIONS

With the growing number of macromolecular assemblies characterized by EM, integrative modeling techniques are becoming increasingly useful for a mechanistic understanding of these assemblies (6,30–32). The MultiFit web server was designed to provide a user-friendly web interface to the MultiFit module in the IMP package, for fitting multiple protein structures into their assembly density map.

REFERENCES

1. Frank,J. (2006) *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*, 2nd edn. Oxford University Press, New York.
2. Zhang,J., Baker,M.L., Schroder,G.F., Douglas,N.R., Reissmann,S., Jakana,J., Dougherty,M., Fu,C.J., Levitt,M., Ludtke,S.J. *et al.* (2010) Mechanism of folding chamber closure in a group II chaperonin. *Nature*, **463**, 379–383.
3. Chen,J.Z., Settembre,E.C., Aoki,S.T., Zhang,X., Bellamy,A.R., Dormitzer,P.R., Harrison,S.C. and Grigorieff,N. (2009) Molecular interactions in rotavirus assembly and uncoating seen by high-resolution cryo-EM. *Proc. Natl Acad. Sci. USA*, **106**, 10644–10648.
4. Fabiola,F. and Chapman,M.S. (2005) Fitting of high-resolution structures into electron microscopy reconstruction images. *Structure*, **13**, 389–400.
5. Baker,M.L., Baker,M.R., Hryc,C.F. and Dimaio,F. (2011) Analyses of subnanometer resolution cryo-EM density maps. *Methods Enzymol.*, **483**, 1–29.
6. Lasker,K., Phillips,J.L., Russel,D., Velazquez-Muriel,J., Schneidman-Duhovny,D., Tjioe,E., Webb,B., Schlessinger,A. and Sali,A. (2010) Integrative structure modeling of macromolecular assemblies from proteomics data. *Mol. Cell Proteomics*, **9**, 1689–1702.
7. Forster,F. and Villa,E. (2011) Integration of cryo-EM with atomic and protein-protein interaction data. *Methods Enzymol.*, **483**, 47–72.
8. Lindert,S., Stewart,P.L. and Meiler,J. (2009) Hybrid approaches: applying computational methods in cryo-electron microscopy. *Curr. Opin. Struct. Biol.*, **19**, 218–225.
9. Volkmann,N. and Hanein,D. (1999) Quantitative fitting of atomic models into observed densities derived by electron microscopy. *J. Struct. Biol.*, **125**, 176–184.
10. Roseman,A.M. (2000) Docking structures of domains into maps from cryo-electron microscopy using local correlation. *Acta Crystallogr. D Biol. Crystallogr.*, **56**, 1332–1340.
11. Rossmann,M.G., Bernal,R. and Pletnev,S.V. (2001) Combining electron microscopic with X-ray crystallographic structures. *J. Struct. Biol.*, **136**, 190–200.
12. Jiang,W., Baker,M.L., Ludtke,S.J. and Chiu,W. (2001) Bridging the information gap: computational tools for intermediate resolution structure interpretation. *J. Mol. Biol.*, **308**, 1033–1044.
13. Chacon,P. and Wriggers,W. (2002) Multi-resolution contour-based fitting of macromolecular structures. *J. Mol. Biol.*, **317**, 375–384.
14. Suhre,K., Navaza,J. and Sanejouand,Y.H. (2006) NORMA: a tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps. *Acta Crystallogr. D Biol Crystallogr.*, **62**, 1098–1100.

15. Birmanns,S. and Wriggers,W. (2007) Multi-resolution anchor-point registration of biomolecular assemblies and their components. *J. Struct. Biol.*, **157**, 271–280.
16. Navaza,J., Lepault,J., Rey,F.A., Alvarez-Rua,C. and Borge,J. (2002) On the fitting of model electron densities into EM reconstructions: a reciprocal-space formulation. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 1820–1825.
17. Topf,M., Baker,M., John,B., Chiu,W. and Sali,A. (2005) Structural characterization of components of protein assemblies by comparative modeling and electron cryo-microscopy. *J. Struct. Biol.*, **149**, 191–203.
18. Lasker,K., Dror,O., Shatsky,M., Nussinov,R. and Wolfson,H.J. (2007) EMatch: discovery of high resolution structural homologues of protein domains in intermediate resolution cryo-EM maps. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **4**, 28–39.
19. Goddard,T.D., Huang,C.C. and Ferrin,T.E. (2007) Visualizing density maps with UCSF Chimera. *J. Struct. Biol.*, **157**, 281–287.
20. Garzón,J.I., Kovacs,J.A., Abagyan,R. and Chacón,P. (2007) ADP_EM: Fast exhaustive multi-resolution docking for high-throughput coverage. *Bioinformatics*, **23**, 427–33.
21. Vasishtan,D. and Topf,M. (2011) Scoring functions for cryoEM density fitting. *J. Struct. Biol.*, **174**, 333–343.
22. Kawabata,T. (2008) Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a gaussian mixture model. *Biophys. J.*, **95**, 4643–4658.
23. Lasker,K., Topf,M., Sali,A. and Wolfson,H. (2009) Inferential optimization for simultaneous fitting of multiple components into a cryoEM map of their assembly. *J. Mol. Biol.*, **388**, 180–194.
24. Birmanns,S., Rusu,M. and Wriggers,W. (2010) Using Sculptor and Situs for simultaneous assembly of atomic components into low-resolution shapes. *J. Struct. Biol.*, **173**, 428–435.
25. Zhang,S., Vasishtan,D., Xu,M., Topf,M. and Alber,F. (2010) A fast mathematical programming procedure for simultaneous fitting of assembly components into cryoEM density maps. *Bioinformatics*, **26**, i261–i268.
26. Lasker,K., Sali,A. and Wolfson,H.J. (2010) Determining macromolecular assembly structures by molecular docking and fitting into an electron density map. *Proteins*, **78**, 3205–3211.
27. Penczek,P.A. (2010) Resolution measures in molecular electron microscopy. *Methods Enzymol.*, **482**, 73–100.
28. Lawson,C.L., Baker,M.L., Best,C., Bi,C., Dougherty,M., Feng,P., van Ginkel,G., Devkota,B., Lagerstedt,I., Ludtke,S.J. *et al.* (2010) EMDataBank.org: unified data resource for CryoEM. *Nucleic Acids Res.*, **39**, D456–D464.
29. Mashiach-Farkash,E., Nussinov,R. and Wolfson,H.J. (in press) SymmRef: A flexible refinement method for symmetric multimers. *Proteins: Structure, Function, and Bioinformatics*.
30. Sali,A., Glaeser,R., Earnest,T. and Baumeister,W. (2003) From words to literature in structural proteomics. *Nature*, **422**, 216–225.
31. Robinson,C., Sali,A. and Baumeister,W. (2007) The molecular sociology of the cell. *Nature*, **450**, 973–982.
32. Alber,F., Forster,F., Korkin,D., Topf,M. and Sali,A. (2008) Integrating diverse data for structure determination of macromolecular assemblies. *Annu. Rev. Biochem.*, **77**, 443–477.
33. Braig,K., Adams,P.D. and Brunger,A.T. (1995) Conformational variability in the refined structure of the chaperonin GroEL at 2.8 Å resolution. *Nat. Struct. Biol.*, **2**, 1083–1094.
34. Ludtke,S.J., Jakana,J., Song,J.L., Chuang,D.T. and Chiu,W. (2001) A 11.5 A single particle reconstruction of GroEL using EMAN. *J. Mol. Biol.*, **314**, 253–262.