MVP-Fit: A Convenient Tool for Flexible Fitting of Protein Domain Structures with Cryo-Electron Microscopy Density Map

Dong Xu and Yang Zhang*
Department of Computational Medicine & Bioinformatics, University of Michigan, 100 Washtenaw Ave, Ann Arbor, MI 48109

(December 17, 2010)

ABSTRACT

Summary: MVP-Fit is a visualization tool for interactive fitting atomic protein domain structures with the cryo-Electron Microscopy density map. To achieve the best match, the program can conveniently adjust the loop and tail outliers of individual domains to accommodate the local conformational changes from the rigid-body rotation and translation of protein domains. **Availability:** The MVP-Fit program and the user manual are freely available at http://zhanglab.ccmb.med.umich.edu/MVP-Fit.

1. INTRODUCTION

Three-dimensional atomic structure of protein molecules can be determined by the experimental techniques, including X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. However, these techniques often have difficulty in determining the so-called quaternary structure of multiple-molecule complexes. Many of the big protein complexes have only structural information in the form of 3D density maps, which are constructed from projected images of cryo-Electron Microscopy (cryo-EM) (Frank, 2006). The density maps don't contain atomic details of the molecules due to the low resolution of cryo-EM images.

Given the tertiary structure of each component domains, researchers often need to fit the monomer structures into the cryo EM density maps to assist their study of structure and function of the complexes. Only a small portion of monomers have their experimental structures deposited in the Protein Data Bank (PDB) (Berman, et al., 2000), while the acquirement of other proteins have to rely on the computer-based structure prediction (Baker and Sali, 2001; Zhang, 2008). The quality of the structure predictions usually depends on whether homologous templates are available in the PDB. Although there were arguments that the current PDB is complete for single-domain protein structures (Zhang and Skolnick, 2005), only around 70% of globular protein targets could have the templates correctly identified by the state-of-the-art threading algorithms (Zhang, 2009). For multiple domain proteins, in particular, the templates of individual domains are often derived from different proteins, where relative orientations of the domains are usually unknown.

Until now, manual guides are still needed for fitting the protein domains into complex density maps since automatic structural fitting requires tremendous computational time and correctness of the final models cannot be guaranteed. However, the automatic search has the advantage in local structure refinement and energy minimization. One optimal solution to the issue is first to generate a rough estimation of the domain positions by heuristic manual fitting and then to perform automatic fitting searches near the manual estimations. UCSF Chimera is one of the most often-used programs for visualizing cryo-EM density maps (Goddard, et al.,

*To whom correspondence should be addressed. Email: zhng@umich.edu.

2007). It supports manual segmentation of density maps and rigid-body fitting of atomic protein models. UROX is another interactive tool for atomic fitting which can do normal-modes calculation and fit one density map to another (Siebert and Navaza, 2009). However, most of these programs do not allow flexible structural adjustments which can make structural fitting difficult, because unbound protein models often have structural outliers in the loop and tail regions which need to be relocated in order to achieve the optimal fitting. Meanwhile, a number of existing EM fitting packages rely on external programs for visual graphics (e.g. Pymol and VMD), which can prevent real-time interactive fitting between EM map and model-derivation electron density calculations.

MVP (Macromolecular Visualization and Processing) is a recently developed visualization system which generates quickly and accurately triangulated isosurfaces for density maps (Xu and Zhang, 2009). Based on MVP, here we present a new algorithm, MVP-Fit, for cryo-EM structure fitting by the combination of both global rigid-body movements and local flexible accommodations.

2. METHODS

After loading a density map, MVP-Fit represents the map by triangulated meshes instead of semitransparent isosurface or point cloud, which makes the inner cavities be clearly seen. As described in the previous paper (Xu and Zhang, 2009), the generated mesh surface is very smooth and contains no singularity. The total number of triangles is only half of that by the traditional method, which makes the interactive operation more fluent. The density threshold can be adjusted by users with the consequent new meshes regenerated efficiently.

The following two kinds of movements are carried out for the atomic models. The reduced cartoon model or Cα backbone model are used to represent each atomic model.

2.1.Global rigid-body movements

There are six degrees of freedom when moving one domain model, which includes 3 translations and 3 rotations along 3 perpendicular axes. One also needs to decide whether the movement is forward or backward when doing the translation or rotation. Hence, totally 12 hotkeys are adopted for rigid-body movements. MVP-Fit also supports fitting multiple models in the EM data. Each time, one model is moved while others keep frozen (see http://zhanglab.ccmb.med.umich.edu/MVP-Fit/UserGuide.pdf).

Human knowledge can be incorporated to guide the movements. One is based on the shape similarity between mesh surface of the EM density map and the contour of the atomic models. The fitting is more favorable when the atomic models are filling the inner of EM isosurface. Another is based on the solvent accessibility. Hydrophobic residues tend to lie in the core of the EM density map. Furthermore, if there is more than one monomer in the EM data, MVP-Fit can make use of the known interacting residues between monomers to guide the fitting.

2.2. Local flexible movements

Since most protein structures are still unsolved, predicted tertiary models are often used for EM fitting. If a model contains multiple domains, it is difficult to decide the relative domain orientations when there is no template for the whole chains. Even for the domain of experimental structure, most were solved in the unbound form, which may not exactly fit with the EM shape. Thus, when doing rigid-body fitting, users may find that some domain cannot be put inside the

EM isosurface. MVP-Fit can move the domains into the surface by changing the designated loop regions between domains. There are two types of local movements to one region:

(1)Translate one region along three axes (X, Y, Z). Each direction has backward and forward movement options.

(2)Rotate one region by changing the main-chain Φ/ψ torsion angles. There are two rotation modes: One uses the starting residue as the pivot and another uses the ending residue. The user can specify whether to increase or decrease the torsion angles.

After translation or rotation of one region, the whole-chain model will have one or two broken parts. A post-movement is carried out afterwards using the LMProt algorithm (da Silva, et al., 2004). It tries to satisfy all restrictions of bond lengths and bond angles by perturbing all the main-chain atoms in the loop region. After using the perturbation to rebuild the main chain, the clashed pairs of Cα atoms are recalculated. The movement is accepted only when all the bond lengths and bond angles are in physical regions and there is no backbone clash. All other atoms are reconstructed based on their relative positions to the main chain atoms.

3. RESULTS

As an illustration of using MVP-Fit, we apply it to the model construction for phosphorylase kinase (phk), which is a hexadecameric complex of 4 groups, each having four different domains (alpha, beta, gamma, delta). This represents a typical and realistic case of EM-fitting where only part of the domain structures is available and both high- and low-resolution predictions are needed for the domain modeling (more complete set of examples and illustrations can be seen at http://zhanglab.ccmb.med.umich.edu/MVP-Fit).

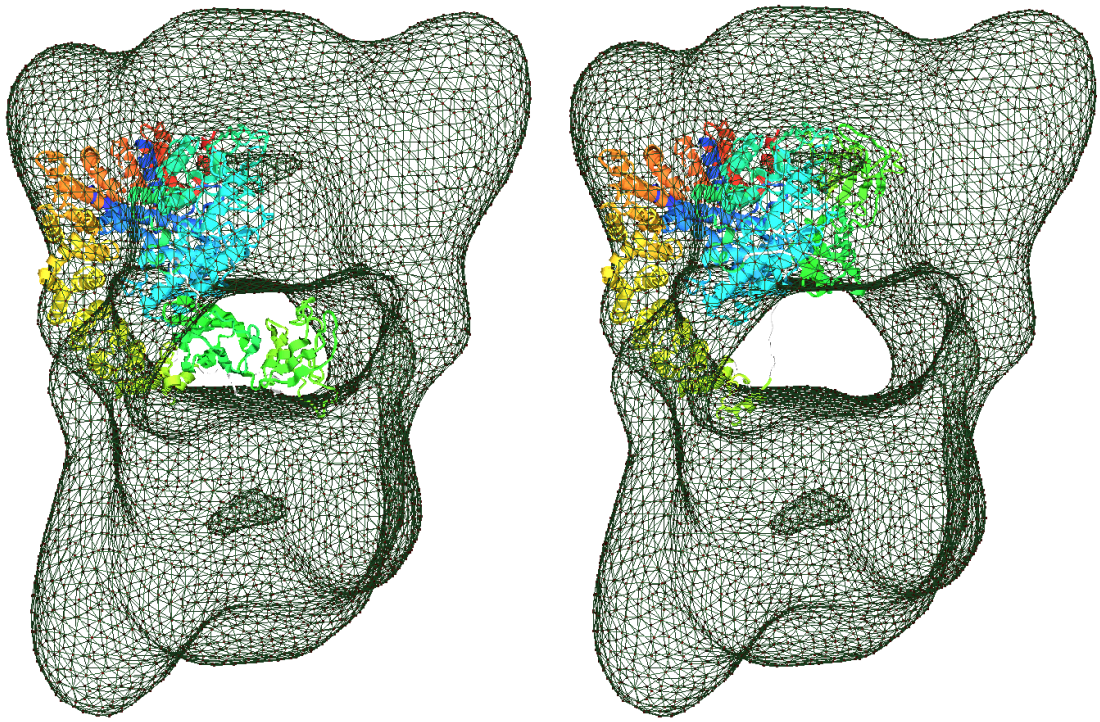

Figure 1. Manual fitting of predicted models of beta subunit into phosphorylase kinase using MVP-Fit. Left panel shows the result with rigid-body fitting while the right shows that with further local flexible fitting.

It is believed that the regulatory beta subunit locates in the bridge (Nadeau, et al., 2002) which stabilizes the global complex. The beta subunit contains four domains. Only the first domain has homologous templates and we model the structure using the I-TASSER threading structure assembly algorithm (Roy, et al., 2010). The structure of other three domains is generated by the QUARK *ab initio* folding program (Xu and Zhang, 2010a; Xu and Zhang, 2010b), which was shown in CASP9 to be the best method for *free model* (FM) structural prediction (Grishin, 2010). The four domain structures are then assembled together by MVP-Fit to form the complete model for the beta subunit.

Figure 1 shows the modeling results by fitting the beta subunits into the phk EM data. Most domains of the subunit fit well with the density map after the rigid-body movements, except for the second domain which has loop and tail outliers have steric clashes with other domains no matter what orientation is assigned (left panel). MVP-Fit moves the domain by a few rounds of local rotation and translation with the loop/tail outliers repacked. As shown in the right panel, the local flexible refinements make the domain dock to the first domain and fit well with the EM isosurface.

ACKNOWLEDGEMENTS

We thank Drs. O Nadeau and GM Carlson for providing EM data for phosphorylase kinase and guidance on the subunit positions.

Funding: The project is supported by the Alfred P. Sloan Foundation, NSF Career Award 0746198 and the National Institute of General Medical Sciences Grant GM083107 and GM084222.

Conflict of interest: none declared.

REFERENCES

Baker, D. and Sali, A. (2001) Protein structure prediction and structural genomics, *Science*, **294**, 93-96.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank, *Nucleic Acids Res*, **28**, 235-242.

da Silva, R.A., Degreve, L. and Caliri, A. (2004) LMProt: an efficient algorithm for Monte Carlo sampling of protein conformational space, *Biophys J*, **87**, 1567-1577.

Frank, J. (2006) *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. Oxford University Press, NewYork

Goddard, T.D., Huang, C.C. and Ferrin, T.E. (2007) Visualizing density maps with UCSF Chimera, *J Struct Biol*, **157**, 281-287.

Grishin, N.V. (2010) Assessment of Free Modeling in CASP9 Experiment.

Nadeau, O.W., Carlson, G.M. and Gogol, E.P. (2002) A Ca(2+)-dependent global conformational change in the 3D structure of phosphorylase kinase obtained from electron microscopy, *Structure*, **10**, 23-32.

Roy, A., Kucukural, A. and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction, *Nat Protoc*, **5**, 725-738.

Siebert, X. and Navaza, J. (2009) UROX 2.0: an interactive tool for fitting atomic models into electron-microscopy reconstructions, *Acta Crystallogr D Biol Crystallogr*, **65**, 651-658.

Xu, D. and Zhang, Y. (2009) Generating triangulated macromolecular surfaces by Euclidean distance transform, *PLoS One*, **4**, e8140.

Xu, D. and Zhang, Y. (2010a) QUARK ab initio protein structure prediction I: Method developments, Submitted.
Xu, D. and Zhang, Y. (2010b) QUARK ab initio protein structure prediction II: results of benchmark and blind tests, Submitted.
Zhang, Y. (2008) Progress and challenges in protein structure prediction, *Curr. Opin. Struct. Biol.*, **18**, 342-348.
Zhang, Y. (2009) Protein structure prediction: when is it useful?, *Curr Opin Struct Biol*, **19**, 145-155.
Zhang, Y. and Skolnick, J. (2005) The protein structure prediction problem could be solved using the current PDB library, *Proc. Natl. Acad. Sci. USA*, **102**, 1029-1034.