W310–W314 *Nucleic Acids Research, 2018, Vol. 46, Web Server issue* *Published online 21 May 2018*
doi: 10.1093/nar/gky424

FragFit: a web-application for interactive modeling of protein segments into cryo-EM density maps

Johanna K.S. Tiemann[1,2,†], Alexander S. Rose[1,†], Jochen Ismer[1], Mitra D. Darvish[1], Tarek Hilal[1], Christian M.T. Spahn[1] and Peter W. Hildebrand[1,2,*]

[1]Institute of Medical Physics and Biophysics, Charité University Medicine Berlin, Berlin 10117, Germany and
[2]Institute of Medical Physics and Biophysics, Medical University Leipzig, Leipzig, Sachsen 04107, Germany

Received January 30, 2018; Revised April 25, 2018; Editorial Decision April 28, 2018; Accepted May 10, 2018

ABSTRACT

Cryo-electron microscopy (cryo-EM) is a standard method to determine the three-dimensional structures of molecular complexes. However, easy to use tools for modeling of protein segments into cryo-EM maps are sparse. Here, we present the FragFit web-application, a web server for interactive modeling of segments of up to 35 amino acids length into cryo-EM density maps. The fragments are provided by a regularly updated database containing at the moment about 1 billion entries extracted from PDB structures and can be readily integrated into a protein structure. Fragments are selected based on geometric criteria, sequence similarity and fit into a given cryo-EM density map. Web-based molecular visualization with the NGL Viewer allows interactive selection of fragments. The FragFit web-application, accessible at http://proteinformatics.de/FragFit, is free and open to all users, without any login requirements.

INTRODUCTION

Due to recent technical advances in development of direct electron detectors, cryo-electron microscopy (cryo-EM) has become a key technology in structural biology (1) that now even allows *de novo* modeling of side chains in well-resolved parts (2,3). In cryo-EM density maps resolved at sub-nanometer resolution (4), secondary structure elements or backbone traces can be identified and modeled (5,6). Therefore, fast and easy to use methods for modeling loops, helices or sheets into cryo-EM density maps are in great demand.

Here, we present the FragFit web-application for modeling of missing segments into cryo-EM density maps of proteins. FragFit employs a classical fragment-based approach for modeling of segments in proteins (7–9) and uses the local fit to a given cryo-EM density map for re-scoring. FragFit works very well for a broad spectrum of resolutions, but provides best results for maps with resolutions of at least 12 Å (10). Test cases are available for high (<4 Å), medium (6 Å) and low resolution (8.9 Å). FragFit can be used to model or remodel parts of proteins for which cryo-EM density maps are available. It has been proven to guide modeling of poorly resolved flexible loops in ribosome bound initiation factor-2, which cryo-EM density map was resolved at a global map resolution of 3.7 Å (11). Moreover, FragFit can be readily integrated into modeling approaches, where conformational changes of proteins only affect a substructure of the protein or a single domain, while the general fold remains unchanged (12). In these cases, flexible fitting of the complete structure or complex is not required. Instead, the structure can be dissembled into its different domains which are rigidly fitted (13,14). FragFit can then be used to reconnect these domains or to re-model the hinge regions.

A great advantage of using FragFit compared to other methods such as Coot (15), RosettaES (16), EM-Fold (17,18), Segger (19) or VolRover (20) is ease of usage. While the latter tools are powerful or even allow *de novo* modeling of backbone and side chains of regions resolved at high resolution, they require specialized knowledge (21). In contrast, FragFit can be used instantly and no installation is required.

The quality of the modeled structure depends on the type of secondary structure, resolution and presence of fragmentations or artifacts within a map. Since most maps feature fragmentations and local variations in resolution, a fully automated approach is challenging. Near native conformations are, however, regularly found in the top five results list of FragFit (10) and the 100 top hits can be visualized and selected in the FragFit web-application. Here, we use the NGL Viewer (22,23) for integrated and interactive visualization. The NGL Viewer adopts capabilities of modern web browsers, such as WebGL for molecular graphics, allowing comprehensive molecular visualization even of huge trajectories derived from molecular dynamics simulations without the need to install additional software. The user

*To whom correspondence should be addressed. Tel: +49 341 97 15 712; Fax: +49 30 450 524 138; Email: peter.hildebrand@medizin.uni-leipzig.de
†The authors wish to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
Present address: Alexander S. Rose, RCSB Protein Data Bank, San Diego Supercomputer Center, University of California, San Diego, CA 92093, USA.

can directly upload the cryo-EM density map and available structure coordinates into the NGL Viewer, handle a search and visually inspect and validate fragment candidates. The fragment database used is regularly updated. The web-service is freely available at http://proteinformatics.de/FragFit.

WEB SERVER CONSTRUCTION

Server workflow

The primary purpose of FragFit is to model missing segments such as loops, helices or β-sheets of up to 35 amino acids length into protein structures. The server employs a fragment-based approach for modeling of fragments into cryo-EM density maps. As a distinctive feature, FragFit provides a powerful visualization to allow interactive validation and selection by the user. A schematic of the workflow is shown in Figure 1. In the initial step, fragment-based prediction (FragSearch) is performed by a fast hierarchical search algorithm to detect suitable fragments (7–9). In the second step (FragFit), cross-correlation between the simulated and the experimentally determined density map is calculated for re-scoring to obtain fitting fragments. Interactive visualization and selection provided by the NGL Viewer (22,23) is required to select suitable fragments especially for longer segments.

Template dataset

Two fragment database options are provided by FragFit: LIP ('Loops In Proteins') contains all overlapping fragments of 3–35 residues length extracted from all protein entries of the PDB, while LIMP ('Loops In Membrane Proteins') only composes from loops derived from membrane proteins (7,8,24). The LIMP database takes into account the specifics of the lipid bilayer (25) and is accordingly intended for modeling of loops in membrane proteins. With the latest regular update, the LIP database was extended from 900 million (December 2015) to more than 1 billion (November 2017) protein fragments. For each fragment the amino acid sequence, PDB ID, PDB chain label, the residue numbers of N- and C-terminal stem atoms and a geometrical fingerprint is stored in the databases.

Implementation details

The FragFit server integrates two major steps—the fragment search (FragSearch) and the fitting of these fragments (FragFit) into cryo-EM density (see Figure 1).

FragSearch is based on the hierarchical approach implemented by Superlooper and SL2, which allows fast and efficient searches of huge databases (7,8). Briefly, FragSearch selects fragment candidates of the same length as the queried segment and with a similar distance d of N- and C-terminal stem residues as in the gap of the protein structure ($\Delta d < 0.75$ Å). The resulting candidate fragments are ranked by a geometrical fingerprint score that takes geometric matching of stem residues of fragment and the protein chain at the gap and sequence similarity into account. Fragments of highly similar three dimensional structures are filtered out from the top 1000 list to maximize the conformational space represented by this list (see (9)). The top 100 list of 'suitable fragments' is finally provided and subsequently re-ranked by the local fit to cryo-EM density maps (FragFit).

FragFit employs pre-processed cryo-EM density maps for re-scoring of the top 100 list of suitable fragments. For pre-processing a minimal box enclosing the density of the queried segment is extracted from the uploaded cryo-EM density map before densities occupied by other parts of the structure are deleted. These pre-processing steps maximize segment prediction quality and minimize calculation time (10). Simulated density maps of the backbone of the fragments are generated and filtered to the resolution of the experimental cryo-EM density map using a Butterworth low pass filter. At last, to re-rank the top 100 list of suitable fragments, the Pearson cross-correlation coefficient is calculated between the simulated density maps and experimentally determined cryo-EM density maps with SPIDER (26). FragFit applies well to high or medium resolution maps. Even for resolutions above 12 Å, the overall shape of the map can restrict the space of possible fragment conformations sufficiently to guide modeling (10). In the latter case, visual inspection and control is, however, obligatory.

Technical details

FragFit is free to access from http://proteinformatics.de/FragFit and no login procedure is required. The web application is written in JavaScript and compatible with any modern web browser (Mozilla Firefox (>v.29), Google Chrome (>v.27), Microsoft Internet Explorer (≥v.11), Apple Safari (≥v.8)) and requires no plug-ins to be installed.

The web server is implemented as a combination of several Python modules running on a dedicated Linux server. The job management and scheduling is handled by a Python job server using the Flask framework (http://flask.pocoo.org/). Each job is run as a Python script performing preprocessing of the input, running external scripts and programs (Python, SPIDER and DELPHI) and preparing the output. The structure and the top 100 list of fitting fragments are visualized for interactive selection by the NGL Viewer (22,23).

UTILITY AND WEB INTERFACE

The FragFit web application provides an intuitive graphical user interface (GUI) integrated into the NGL GUI (Figure 1). In the following, the main GUI elements for input and output are described in detail. A descriptive guide and further information are provided in the online documentation, including a method section, example results, documentation and a list of frequently asked questions.

Required server input and input modifications

The user has to provide a cryo-EM density map in MRC or CCP4 (MAP) format and the atomic coordinates of a protein structure in PDB format v3.x as input for the web interface of FragFit. Both are automatically loaded into the NGL Viewer for interactive inspection. Note that in order to speed up upload to the server, huge maps should be cropped to contain the domains enclosing the region

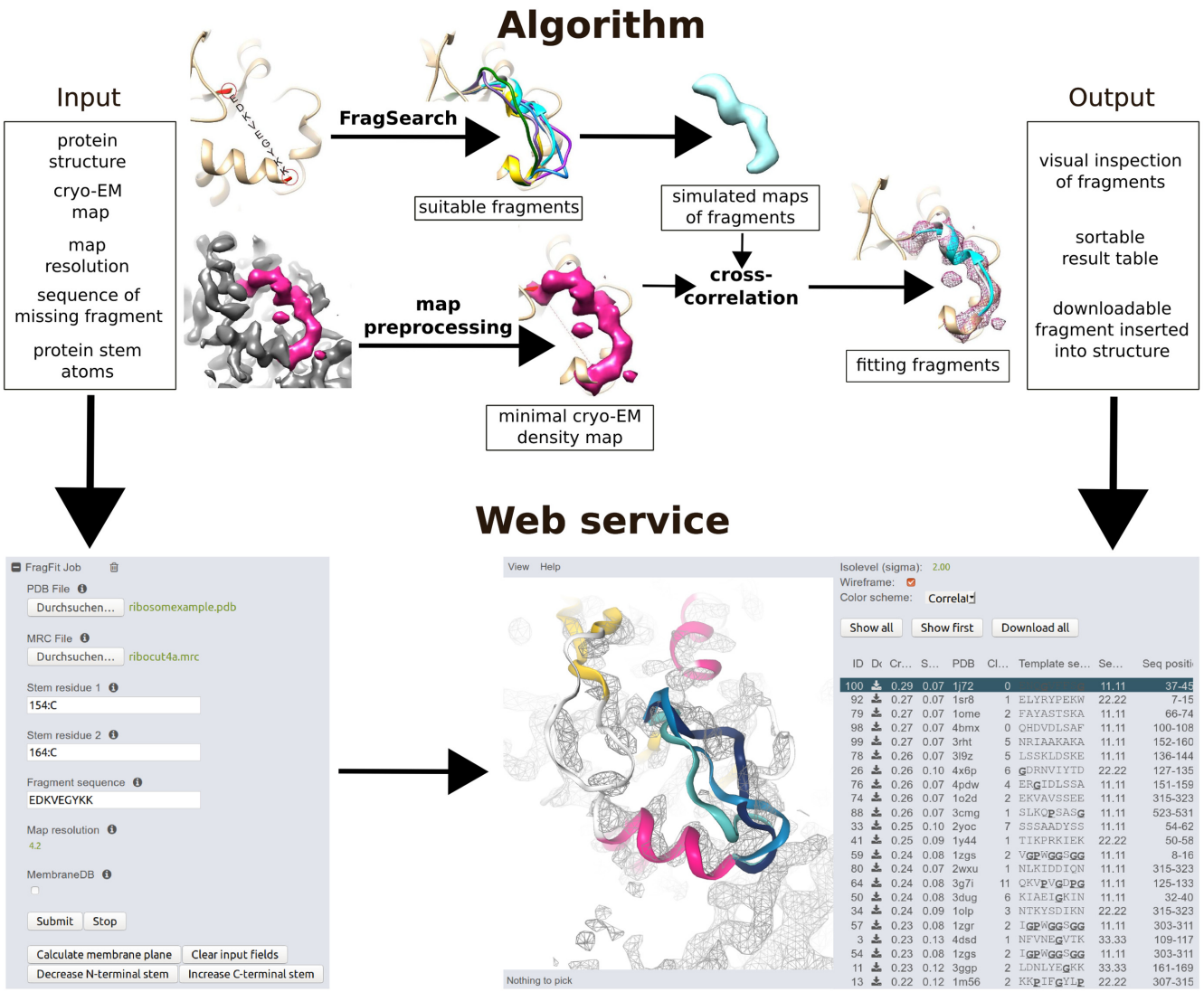

Figure 1. Workflow, input panel and result table of the FragFit web-application. The top part gives an overview about the FragFit workflow, the lower part presents a screenshot of the web service with input panel (left), molecular visualization and result panel (right). As input, the PDB structure ('PDB file'), the cryo-EM density map ('Density/MRC file'), its resolution ('Map resolution'), the sequence of the queried segment ('Fragment sequence') and the stem residues flanking the queried segment (Stem residue 1/2) must be provided. As soon as a search is initiated, suitable fragments are selected from the database by geometric and sequence criteria (FragSearch). The list of suitable fragments is automatically re-scored by the cross-correlation coefficient between the simulated map of the fragment and the minimal cryo-EM density map, which is extracted during pre-processing. The output (bottom right) allows visual inspection and selection of the most appropriate fitting fragment by the NGL Viewer from the top-100 hit list (bottom center). A sortable result table (bottom right) provides information about FragSearch and FragFit scores used for ranking, sequence similarity, sequence and origin of the template and backbone clashes. If no appropriate fragment is found, the input data can be modified e.g. the search window can be extended in N- or C-terminal directions. Finally, selected fragments are integrated into the protein structure and are available for download.

of missing fragments. The density map and the structure file have to be aligned. Number and chain-ID of the stem residues (where a fragment should be modeled in) must either be typed into the respective form (Figure 1) or can be selected by directly clicking onto the stem residues within the viewer canvas. The sequence of the queried protein segment and the resolution of the cryo-EM density map need to be specified as well. For membrane proteins, planes indicating the lipid bilayer can be automatically obtained from the TMDET (27) web-service to guide modeling of membrane protein loops in the NGL Viewer (22,23).

After the job has been started, the input interface remains open to provide the possibility for input modifications after initial result inspection. If no suitable candidate is found, a new search can be started with the search window extended in N- or C-terminal directions. In this case, the amino acid sequence of the queried segment is automatically expanded.

Server output

Depending on the fragment length, finishing a search may take some seconds up to several minutes. The page is automatically updated (see Figure 1). Fitting fragments can be selected from a result list and directly visualized within

the cryo-EM density map by the NGL Viewer. The interactive result table is initially ranked according to the cross-correlation score, which evaluates the local fit of a fragment to the cryo-EM density map. For each candidate the FragSearch and FragFit score, PDB ID, the origin and sequence of the template protein, the sequence identity and the number of clashes is listed. Of note, the user can sort the results list by each of these columns. A selected loop candidate is automatically integrated into the structure with its target sequence and side-chains (not energetically minimized) added and can be downloaded for further usage. Since the fragments are taken from PDB structures which have undergone several steps of quality control, the fragments do not necessarily have to be refined, only the side chain rotamers might have to be edited.

CONCLUSION

Using a hierarchical search algorithm and efficient preprocessing of cryo-EM density maps, FragFit is able to quickly and effectively model protein segments into cryo-EM density maps. Providing visualization by an interactive web application carried out by the NGL Viewer (22), loop candidates can be inspected and controlled directly within the same application. The utility of FragFit has been proven to guide modeling of poorly resolved regions (11), but may also serve for approaches where protein domains are reconnected after rigid fitting or more generally to re-model flexible hinge regions. In summary, FragFit offers structural biologists easy access to modeling of missing segments in cryo-EM structures.

DATA AVAILABILITY

The web server is freely available and no login is required.

ACKNOWLEDGEMENTS

We thank Andrean Goede for his help in handling the fragment database.

FUNDING

Deutsche Forschungsgemeinschaft [SFB740/B6, HI 1502/1–2, BI 893/8 to P.W.H, SFB740/Z1 to C.M.T.S.]. Berlin Institute of Health [to P.W.H.]. Stiftung Charité [to P.W.H]. Einstein Center Digital Future [to P.W.H]. Funding for open access charge: Deutsche Forschungsgemeinschaft [SFB740/B6].
Conflict of interest statement. None declared.

REFERENCES

1. Callaway,E. (2015) The revolution will not be crystallized: a new method sweeps through structural biology. *Nature*, **525**, 172–174.
2. Brown,A., Long,F., Nicholls,R.A., Toots,J., Emsley,P., Murshudov,G. and Acta Crystallogr,D.B.C. (2015) Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D Biol. Crystallogr.*, **71**, 136–153.
3. Cassidy,C.K., Himes,B.A., Luthey-Schulten,Z. and Zhang,P. (2018) CryoEM-based hybrid modeling approaches for structure determination. *Curr. Opin. Microbiol.*, **43**, 14–23.
4. 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.* (2011) EMDataBank.org: unified data resource for CryoEM. *Nucleic Acids Res.*, **39**, D456–D464.
5. Baker,M.L., Baker,M.R., Hryc,C.F., Ju,T. and Chiu,W. (2012) Gorgon and pathwalking: macromolecular modeling tools for subnanometer resolution density maps. *Biopolymers*, **97**, 655–668.
6. Si,D., Ji,S., Nasr,K.A. and He,J. (2012) A machine learning approach for the identification of protein secondary structure elements from electron cryo-microscopy density maps. *Biopolymers*, **97**, 698–708.
7. Ismer,J., Rose,A.S., Tiemann,J.K.S., Goede,A., Preissner,R. and Hildebrand,P.W. (2016) SL2: an interactive webtool for modeling of missing segments in proteins. *Nucleic Acids Res.*, **44**, W390–W394.
8. Hildebrand,P.W., Goede,A., Bauer,R.A., Gruening,B., Ismer,J., Michalsky,E. and Preissner,R. (2009) SuperLooper: a prediction server for the modeling of loops in globular and membrane proteins. *Nucleic Acids Res.*, **37**, 571–574.
9. Michalsky,E., Goede,A. and Preissner,R. (2003) Loops In Proteins (LIP)–a comprehensive loop database for homology modelling. *Protein Eng.*, **16**, 979–985.
10. Ismer,J., Rose,A.S., Tiemann,J.K.S. and Hildebrand,P.W. (2017) A fragment based method for modeling of protein segments into cryo-EM density maps. *BMC Bioinformatics*, **18**, 475.
11. Sprink,T., Ramrath,D.J., Yamamoto,H., Yamamoto,K., Loerke,J., Ismer,J., Hildebrand,P.W., Scheerer,P., Burger,J., Mielke,T. *et al.* (2016) Structures of ribosome-bound initiation factor 2 reveal the mechanism of subunit association. *Sci. Adv.*, **2**, e1501502.
12. Chapman,B.K., Davulcu,O., Skalicky,J.J., Brüschweiler,R.P. and Chapman,M.S. (2015) Parsimony in protein conformational change. *Structure*, **23**, 1190–1198.
13. 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.
14. van Zundert,G.C.P., Trellet,M., Schaarschmidt,J., Kurkcuoglu,Z., David,M., Verlato,M., Rosato,A. and Bonvin,A.M.J.J. (2017) The DisVis and PowerFit web Servers: explorative and integrative modeling of biomolecular complexes. *J. Mol. Biol.*, **429**, 399–407.
15. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
16. Frenz,B., Walls,A.C., Egelman,E.H., Veesler,D. and Di Maio,F. (2017) RosettaES: A sampling strategy enabling automated interpretation of difficult cryo-EMmaps. *Nat. Methods*, **14**, 797–800.
17. Lindert,S., Staritzbichler,R., Wötzel,N., Karakaş,M., Stewart,P.L. and Meiler,J. (2009) EM-Fold: de novo folding of α-Helical proteins guided by intermediate-resolution electron microscopy density maps. *Structure*, **17**, 990–1003.
18. Lindert,S., Alexander,N., Wötzel,N., Karakaş,M., Stewart,P.L. and Meiler,J. (2012) EM-Fold: De novo atomic-detail protein structure determination from medium-resolution density maps. *Structure*, **20**, 464–478.
19. Pintilie,G.D., Zhang,J., Goddard,T.D., Chiu,W. and Gossard,D.C. (2010) Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.*, **170**, 427–438.
20. Zhang,Q., Bettadapura,R. and Bajaj,C. (2012) Macromolecular structure modeling from 3D EM using VolRover 2.0. *Biopolymers*, **97**, 709–731.
21. Pintilie,G. and Chiu,W. (2012) Comparison of Segger and other methods for segmentation and rigid-body docking of molecular components in Cryo-EM density maps. *Biopolymers*, **97**, 742–760.
22. Rose,A.S. and Hildebrand,P.W. (2015) NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.*, **43**, W576–W579.
23. Rose,A.S., Bradley,A.R., Valasatava,Y., Duarte,J.M., Prlić,A. and Rose,P.W. (2016) Web-based molecular graphics for large complexes. *Proc. 21st Int. Conf. Web3D Technol. - Web3D '16*, 185–186.
24. Rose,P.W., Bi,C., Bluhm,W.F., Christie,C.H., Dimitropoulos,D., Dutta,S., Green,R.K., Goodsell,D.S., Prlic,A., Quesada,M. *et al.* (2013) The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Res.*, **41**, D475–D482.
25. Hildebrand,P.W., Preissner,R. and Frömmel,C. (2004) Structural features of transmembrane helices. *FEBS Lett.*, **559**, 145–151.

26. Frank,J., Radermacher,M., Penczek,P., Zhu,J., Li,Y., Ladjadj,M. and Leith,A. (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.*, **116**, 190–199.

27. Tusnády,G.E., Dosztányi,Z. and Simon,I. (2005) TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. *Bioinformatics*, **21**, 1276–1277.