RIBFIND: a web server for identifying rigid bodies in protein structures and to aid flexible fitting into cryo EM maps

Arun Prasad Pandurangan[*] and Maya Topf[*]

Department of Crystallography/Biological Sciences, Institute of Structural and Molecular Biology, Birkbeck College, University of London, London WC1E 7HX, UK

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: To better analyze low-resolution cryo electron microscopy maps of macromolecular assemblies, component atomic structures frequently have to be flexibly fitted into them. Reaching an optimal fit and preventing the fitting process from getting trapped in local minima can be significantly improved by identifying appropriate rigid bodies (RBs) in the fitted component.

Results: Here we present the RIBFIND server, a tool for identifying RBs in protein structures. The server identifies RBs in proteins by calculating spatial proximity between their secondary structural elements.

Availability: The RIBFIND web server and its standalone program are available at http://ribfind.ismb.lon.ac.uk.

Contact: a.pandurangan@mail.cryst.bbk.ac.uk

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Studying the structure and function of macromolecular assemblies using cryo electron microscopy (cryo EM) techniques is becoming increasingly popular (Orlova and Saibil, 2011). Interpreting the resulting 3D maps, which are typically at a resolution range of ~5 to 25 Å, is almost always performed by first fitting atomic structures of assembly components (e.g. proteins, domains and RNA segments) into them (Beck *et al.*, 2011). However, when the assembly components in the density map are in a different conformation from their corresponding atomic structures (Orlova and Saibil, 2011), fitting the components rigidly may not result in a satisfying pseudo-atomic model. To alleviate this problem, assembly components can be treated flexibly during the fitting process (flexible fitting) (Beck *et al.*, 2011), i.e. by allowing changes in their conformation. In principle, flexible fitting can be applied in a hierarchical fashion at any level of representation of the structure, such as domains, secondary structure elements (SSEs), amino acids and even individual atoms (Topf *et al.*, 2008). In practice, due to the increasing size of the search space, as the level of representation becomes finer, identifying appropriate rigid bodies (RBs) can be vital in reaching the optimal fit. A prior knowledge of the rigid and flexible regions of the components can prevent

the fitting process from getting trapped into many of the local minima (Ahmed and Gohlke, 2006; Pandurangan and Topf, 2012).

There have been many program tools reported in the literature to identify rigid and flexible regions in proteins. Most of these tools attempt to identify 'compact units' in proteins with strong spatial relationship within them (Lesk and Rose, 1981). These compact units can be arranged hierarchically, which can lead to the prediction of a comprehensive set of protein dynamic pathways (Lesk and Rose, 1981). One such program uses a graph theory-based approach to model the atomic structure of the molecule as a bond-bending network and applies a fast combinatorial algorithm to determine the rigidity of this network, thereby the RBs of the molecule (FIRST) (Jacobs *et al.*, 2001). Other approaches to predict flexible hinge regions in proteins include elastic network models (e.g. HingePort) (Emekli *et al.*, 2008), energetic interactions within protein fragments (e.g. FlexOracle) (Flores and Gerstein, 2007) and the identification of rigid blocks in proteins by comparing inter residue distances of two different conformations (e.g. RigidFinder) (Abyzov *et al.*, 2010). We recently introduced RIBFIND, a method that identifies RBs by using spatial proximity between SSEs (Pandurangan and Topf, 2012). Using the method to aid flexible fitting of protein components in low-resolution (15 Å) cryo EM maps was shown to improve the accuracy of the initial fits by ~41% on average (Pandurangan and Topf, 2012; Topf *et al.*, 2008). Here, we present a web server and a standalone tool for RIBFIND.

2 METHODS

The RIBFIND method treats alpha helices and beta sheets as the basic units and clusters them based on their spatial proximities. Following that the group of SSEs in each cluster and the loops connecting them are treated as one RB. SSEs that do not fall into any cluster are treated as independent RBs. The clustering (which is independent of the sequence continuity) is governed by two parameters—the residue 'contact distance' and the 'cluster cutoff'. The contact distance is the distance between the centroids of any two side chains. Its default value is 6.5 Å (Pandurangan and Topf, 2012). The cluster cutoff is the percentage of residues within two SSEs that are in contact. Its value can vary between 1 and 100%. As the value of the cluster cutoff increases, more tightly packed clusters are likely to be identified. For a given value of contact distance, RIBFIND computes a set of RBs at each possible cluster cutoff value in increments of 1%, these sets of RBs can aid the user in selecting the appropriate RBs for flexible fitting into cryo EM density maps.
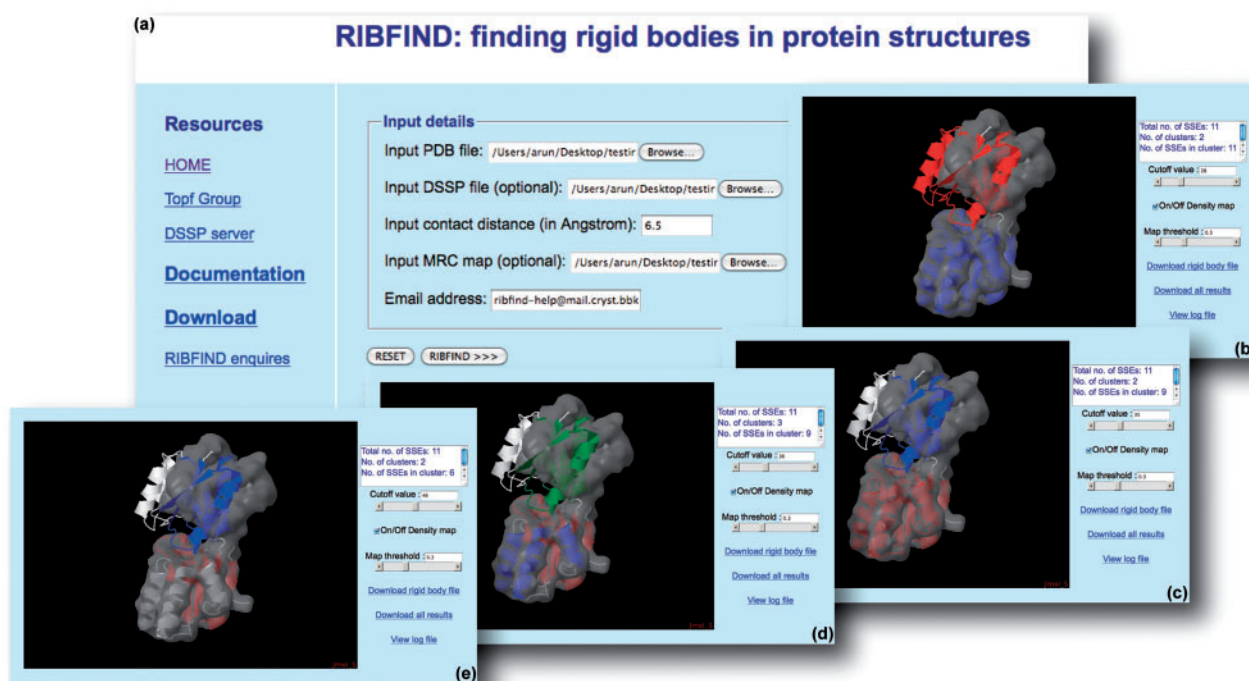
Fig. 1. Screenshots of the RIBFIND webserver. (**a**) The input page. (**b–e**) The results' pages for four different sets of RBs. The total number of SSEs in the protein, the number of RBs and the total number of SSEs that are present in RBs are shown in the right panel

To use RIBFIND, a web server has been set up. It accepts the following inputs (Fig. 1a): (i) the protein coordinates in PDB file format (a single chain or multiple chains with the residues numbered continuously); (ii) a description of the protein SSEs in DSSP (Kabsch and Sander, 1983) format (optional, if not provided, the server automatically runs DSSP); (iii) the value of the contact distance parameter; (iv) the email address of the user and (v) a density map in MRC format (optional). The RIBFIND client interface reads the input, validates it and submits it to the server for execution. After completing the job the server sends an automatic email to the user with a link to the results page. This page contains a Jmol applet (http://jmol.sourceforge.net) showing a cartoon of the user PDB file with each RB uniquely coloured. All SSEs and loops that are not part of a cluster are coloured white. The RB set that is displayed by default is the one that contains the maximal number of clusters identified by RIBFIND [we previously showed that using this set in cryo EM flexible fitting produced the best results in most cases (Pandurangan and Topf, 2012)]. Using the slider control on the result page, the user can view different sets of RBs generated for each cluster cutoff and save the corresponding RB file in a text format [which can be used, e.g., by Flex-EM (Topf *et al.*, 2008)]. The standalone version of RIBFIND is provided on the webpage. It also includes a script to aid the visualization of the RBs using Chimera (www.cgl.ucsf.edu/chimera). The program is written in python and it requires the Biopython (http://biopython.org/) and Numpy (http://numpy.scipy.org/) packages to be installed locally. It takes about 5 min for a RIBFIND run to be completed for a protein of ~800 residues on a single Intel Xeon 3-GHz processor.

3 CASE STUDY

We present the results of the RIBFIND server using the 'close' and 'open' conformations of the *Escherichia coli* ribose-binding protein (271 residues) with the corresponding PDB ids: 2DRI

and 1URP, respectively. Based on DSSP, the structure has a total of 11 SSEs including nine helices and two beta sheets. Running RIBFIND with a contact distance of 6.5 Å resulted in four different sets of RBs for the following cluster cutoff values: 26–33%, 34–35%, 36–37% and 38–50% (below 26% all SSEs fall into one cluster and above 50% no clusters are found) (Fig. 1b–e). The maximal number of three clusters was obtained for a cluster cutoff of 36 and 37% (Fig. 1d). Using the latter set of RBs during flexible fitting into the 10 Å resolution density map of the open conformation (simulated from 1URP), starting from the close conformation of the protein (2DRI), was helpful in reducing the Cα Root Mean Square Deviation (RMSD) between the two conformations from 7.55 to 1.71 Å (Pandurangan and Topf, 2012). It is worth noting that in cases where the close conformation is highly compact, identifying RBs by RIBFIND may be difficult and may not result in the open conformation during flexible fitting into the corresponding open map (e.g. actin subunit; Pandurangan and Topf, 2012).

4 CONCLUSION

The RIBFIND web server offers an automatic way of identifying different sets of RBs in protein structures based on clustering of SSEs. In addition to providing a tool for analyzing polypeptide chains by shedding light on their flexibility, it can be useful for the refinement of structures in the context of cryo EM maps.

REFERENCES

Abyzov,A. *et al.* (2010) RigidFinder: a fast and sensitive method to detect rigid blocks in large macromolecular complexes. *Proteins*, **78**, 309–324.

Ahmed,A. and Gohlke,H. (2006) Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins*, **63**, 1038–1051.

Beck,M. *et al.* (2011) Exploring the spatial and temporal organization of a cell's proteome. *J. Struct. Biol.*, **173**, 483–496.

Emekli,U. *et al.* (2008) HingeProt: automated prediction of hinges in protein structures. *Proteins*, **70**, 1219–1227.

Flores,S.C. and Gerstein,M.B. (2007) FlexOracle: predicting flexible hinges by identification of stable domains. *BMC Bioinformatics*, **8**, 215.

Jacobs,D.J. *et al.* (2001) Protein flexibility predictions using graph theory. *Proteins*, **44**, 150–165.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Lesk,A.M. and Rose,G.D. (1981) Folding units in globular proteins. *Proc. Natl. Acad. Sci. USA*, **78**, 4304–4308.

Orlova,E.V. and Saibil,H.R. (2011) Structural analysis of macromolecular assemblies by electron microscopy. *Chem. Rev.*, **111**, 7710–7748.

Pandurangan,A.P. and Topf,M. (2012) Finding rigid bodies in protein structures: application to flexible fitting into cryoEM maps. *J. Struct. Biol.*, **177**, 520–531.

Topf,M. *et al.* (2008) Protein structure fitting and refinement guided by cryo-EM density. *Structure*, **16**, 295–307.