BIOINFORMATICS APPLICATIONS NOTE Vol. 22 no. 14 2006, pages 1796–1799
doi:10.1093/bioinformatics/btl138

Structural bioinformatics

Analysis assistant for single-molecule force spectroscopy data on membrane proteins—MPTV

Frank Mueller[2], Daniel J. Muller[1] and Dirk Labudde[1,*]

[1]Department of Cellular Machines, Center of Biotechnology, TU Dresden, Tatzberg 47-51, 01307 Dresden, Germany and [2]University of Applied Sciences Lausitz, Grossenhainer Strasse 57, 01968 Senftenberg, Germany

Received on January 3, 2006; revised on February 24, 2006; accepted on April 4, 2006
Advance Access publication April 10, 2006
Associate Editor: Alfonso Valencia

ABSTRACT

Summary: Force-distance (F-D) curves of single membrane proteins reveal information on inter- and intramolecular interactions occurring within a protein and between proteins. However, the analysis of single-molecule force spectroscopy data is a time consuming and complex process requiring objective criteria. In most cases the user requires additional information to interpret F-D curves. Therefore we developed a software assistant representing the force or molecular interaction pattern and the topology or the 3D structure of the membrane protein. This representation establishes a basis for detailed interpretation of the protein structure and its underlying molecular interactions. Various integrated bioinformatic features further assist in the interpretation of measured and assigned molecular interactions that determine membrane protein folding, structure, stability and function. Web queries and programs about the topology are directly linked. Motifs, helix types, representation of Venn diagrams and the complete functionality of the program Jmol belong to it.

Availability: The program MPTV is freely available from the website at http://www.bioforscher.de/mptv.htm/

Contact: dirk.labudde@biotec.tu.dresden.de

1 INTRODUCTION

The detection of molecular forces that stabilize single proteins, form individual receptor–ligand bonds or control antibody–antigen binding became possible with ultra-sensitive force probe methods like the atomic force microscope (Janshoff *et al*., 2000). In force spectroscopy experiments, a single-molecule or receptor–ligand pair is tethered between the tip of a micro-fabricated cantilever and a supporting surface. During continuous stretching of the molecule, the applied forces are measured by the deflection of the cantilever and plotted against extension yielding a characteristic Force-distance (F-D) curve that is formed by subsequent events of molecular interactions. These interactions can represent unbinding, or unfolding events, or ligand-binding of the protein. In the near future, force spectroscopy assays will permit screening of physiologically relevant parameters (e.g. pH, electrolyte concentration and temperature), which alter the inter- and intramolecular interactions that determine, for example, the folding, structure, stability, function and assembly of biological macromolecules (Muller *et al*., 2002, Janovjak *et al*., 2003, Kedrov *et al*., 2005, Sapra *et al*., 2006). In order to obtain statistically relevant results, several hundred to thousand single-molecule experiments have to be performed, each resulting in a unique F-D curve. This F-D curve exhibits a certain pattern, which contains information about strength and location of molecular forces (i.e. interactions) established inside and between individual biomolecules, about intermediates and reaction pathways, and the probability with which these occur. To draw solid statistical conclusions on these molecular interactions (e.g. the magnitude/strength of these interactions), on their structural location, their probability, whether they are independent or occur only with other events, one must be able to analyze a statistically relevant number of force curves by identical objective procedures. Thus, there is an increasing demand for data analysis techniques that offer fully automated processing of many datasets applying identical scientific criteria. To evaluate F-D curves showing various specific and non-specific interactions and different interaction pathways, classification and pattern recognition algorithms are needed. So far, various off-line software packages have been developed to analyze single-molecule force spectroscopy data (Kuhn *et al*., 2005a, b; Carl *et al*., 2001, http://membres.lycos.fr/punias/).

Therefore, we have used a software solution allowing high-throughput classification and statistical analysis of force spectra (Kuhn *et al*., 2005a, b). The algorithms not only include the fitting of polymer extension models to force peaks, but also apply alignment and pattern recognition procedures to reveal different reaction pathways. To demonstrate these capabilities, we have chosen individual single molecule unfolding curves recorded on bacteriorhodopsin mutant (bRP50A) as examples of a membrane protein that follows different unfolding pathways. These unfolding pathways lead to a (F-D) pattern: which peaks contain information on the strength and the location of molecular interactions within the protein. Prior to pattern recognition, each single force curve (Fig. 1A) was smoothed. Then individual force peaks (i.e. the molecular interactions) were detected and fitted using the worm like chain model. This model allows to determine the length of the stretched polymer and thus to locate the molecular interaction within the protein structure. Then, multiple force spectra were aligned by converting them to a sequence of normal distributions and maximizing the overlap. The common (F-D) pattern became visible (Fig. 1D). In a last feature not shown here, combining the procedures of fitting and aligning the force spectra allowed the identification of common peaks being across all recorded spectra. The hierarchical tree of F-D curves recorded from single bRP50A (Fig. 1B) form separate areas which differ clearly and at the same

*To whom correspondence should be addressed.

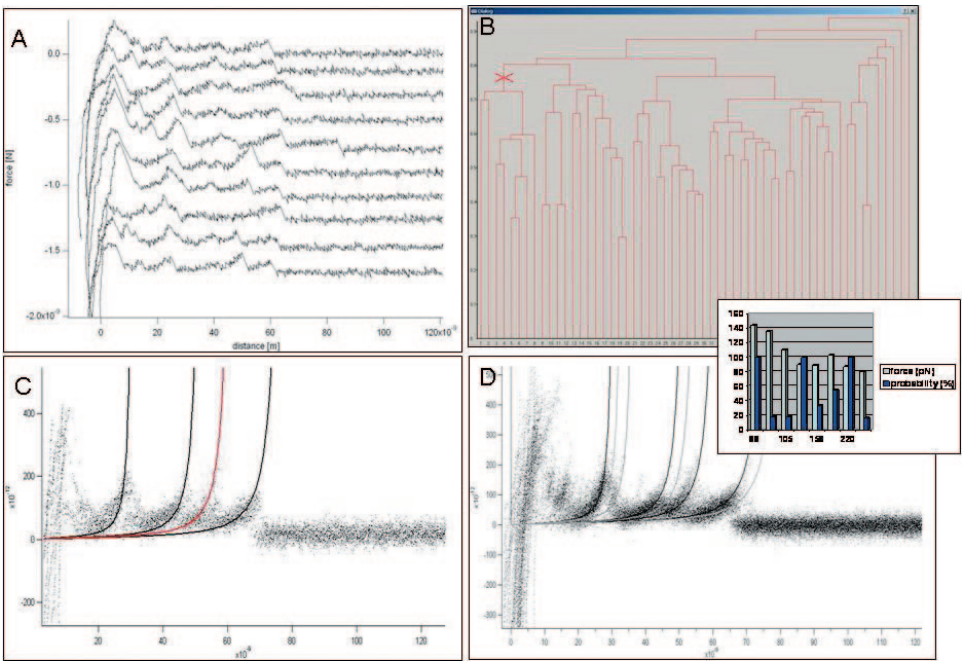

Fig. 1. Pattern recognition and analysis of single molecule force spectra. (**A**) Single F-D curves were recorded while unfolding mutant bacteriorhodopsin P50A. (**B**) Resulting hierarchical tree of F-D curves after fitting and calculating similarities on the basis of resulting WLC fits. (**C**) Overlay of each curve of the highlighted (red marker) cluster. (**D**) Overlay of all included F-D curves. The resulting F-D pattern is shown as histogram (top of D) with the average forces and probabilities of the observed unfolding barriers.

time had certain common features. The approach presents one way to analyze a large number of F-D curves. Certainly, more refined alignment methods will ultimately improve the quality of the alignment and the ability of the software to explore the complexity of molecular interactions within and between biomolecular systems.

However, tools for the visualization and interpretation of molecular interactions correlated to the protein structure are currently not available. The process of understanding force pattern requires additional knowledge and information. Useful information is the topology of the membrane protein, such as structural or functional motifs, secondary structural elements, properties of the involved amino acids and H-bond pattern. For structurally solved membrane proteins this information is included in the 3D structure and available form the PDB (http://www.rcsb.org/pdb/) (Berman *et al*., 2000). In the case where the structure of the membrane protein has not been solved, a variety of bioinformatics tools may be applied for structure prediction. The analysis of single-molecule force spectra recorded upon unfolding membrane proteins requires the so-called membrane compensation. This procedure considers that a molecular interaction established within the protein may be located in the hydrophobic core of the membrane. A similar compensation must be applied for α-helices sticking out of the membrane. The aim of this work was to develop a software tool for visualization and interpretation of the subsequent unfolding or unbinding events.

2 RESULTS AND DISCUSSION

The program MPTV (Membrane Protein Topology Viewer) was designed as an assistant to interpret F-D pattern (of a whole tree or of clusters) and to map interactions on the protein structure. A major task is the concurrent representation of the F-D pattern and the topology of the membrane protein. Lower frame of Figure 2 shows a schematic representation of such a force pattern.

The topology of an investigated membrane protein may be generated in three different ways. For proteins with known 3D structure the topology is provided by the PDB file. Alternatively, DSSP the standard program (Kabsch and Sander, 1983) to identify secondary structures on the basis of a PDB file may be applied. In case of a membrane protein with unknown structure, the program MPTV offers the possibility of topology prediction. In the current MPTV version, the TMHMM (Krogh *et al*., 2001) (http://www.cbs.dtu.dk/services/TMHMM) and the HMMTOP (http://www.enzim.hu/hmmtop) (Tusnády and Simon, 2001) servers are integrated for web queries. In the literature the user may find additional information to include structurally unsolved membrane proteins into this interpretative approach. On the basis of the force pattern the user can include this additional information to manipulate the predicted topology. Therefore, a number of features have been integrated in the tool (extending, combining, reducing or splitting of helices, changing the type of helices, defining transmembrane regions and setting helical breaks).

The sequence information of the investigated membrane protein can be represented as Venn diagrams. The properties of each amino acid are shown at the bottom of the mean window and are highlighted in different colors along the sequence. The definition of amino acid properties takes place in an option window. Alternatively three different types of helical wheels (α-, 3.10-, Π- helix) may be shown and further analyzed.

If the contour length deduced from the F-D pattern indicates that the stretched peptide is not anchored at the membrane surface,

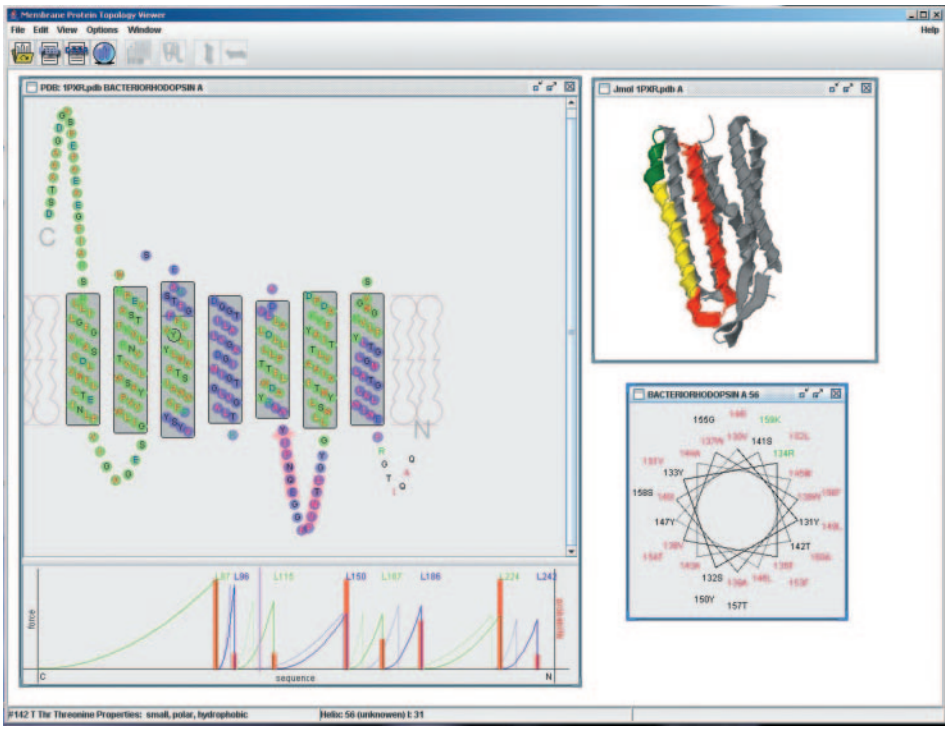

Fig. 2. Representation of the GUI for the program MPTV. On the left side the topology for the membrane protein bRP50A (bacteriorhodopsin mutant Pro50Ala, PDB_ID 1PXR) is shown. On the right side the 3D structure of bR with highlighted membrane compensated contour lengths and a helix wheel (PDB nomenclature, counted N-terminal) is depicted. The expected force pattern calculated from the structural information of the protein is given at the bottom of the left side. The pattern is shown without (transparent) and with (highlighted) membrane compensation.

the thickness of the lipid bilayer (in MPTV changeable) must be considered to correctly obtain the anchoring point of the peptide. The influence of the thickness of the membrane is likewise included. We distinguish three cases for the automated calculation: (1) the anchoring point resides in the middle of the helix, and the anchoring point is located (2) in the N-terminal or (3) in the C-terminal end (detailed information on the web page).

Structural and functional motifs of proteins often give clues of the structure–function relationship. These motifs are available from the server (http://motif.genome.jp) and highlighted directly along the sequence. A special feature for membrane proteins with known structures involves the 3D structure viewer Jmol (http://jmol.sourceforge.net/). The start and the utilization of the Jmol were directly connected with topology of the investigated protein. The user receives information about possible structural influences on the force pattern. The hydrogen bonds are advantageous for the interpretation of selected and highlighted contour lengths. A variable set of options enables the user to adapt the program to individual settings. The settings of the amino acid properties, the colors marking individual force pattern segments, the colors or shadings for structural and functional motifs and miscellaneous window options are freely adjustable by the user. An ongoing project can be interrupted and saved at any given time. For this purpose we have written an own data format, which stores every content of the windows including the settings. For a good documentation, the user can save all input values and predicted information. Images of the figures (including the 3D structure) can be exported in jpeg, png, bmp and eps formats.

The program is written in Java and requires Java 1.5 for flawless and safe runs. More detailed information is presented on the web page (www.bioforscher.de/mptv.htm).

ACKNOWLEDGEMENTS

The authors are thankful to K. T. Sapra for providing the data for bR P50A. The authors are grateful for the financial support of this work by the BMBF (project no. 03I4044B), the free state of Saxony, the European Union (EU) and the Deutsche Forschungsgemeinschaft (DFG).

Conflict of Interest: none declared.

REFERENCES

Berman,H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

Carl,P. *et al.* (2001) Forced unfolding modulated by disulfide bonds in the Ig domains of a cell adhesion molecule. *Proc. Natl Acad. Sci. USA*, **98**, 1565–1570.

Janshoff,A. *et al.* (2000) Force spectroscopy of molecular systems-single molecule spectroscopy of polymers and biomolecules. *Angew. Chem. Int. Ed. Engl.*, **39**, 3212–3237.

Janovjak,H. *et al.* (2003) Unfolding pathways of native bacteriorhodopsin depend on temperature. *EMBO J.*, **22**, 5220–5229.

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

Kuhn,M. *et al.* (2005a) Automated alignment and pattern recognition of single-molecule force spectroscopy data. *J. Microsc.*, **218**, 125–32.

Kuhn,M. *et al.* (2005b) Pattern recognition of single-molecule force spectroscopy Data. *Biophotonics International*, **12**, 50–51.

Krogh,A. *et al*. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol*., **305**, 567–580.

Kedrov,A. *et al*. (2005) Locating ligand binding and activation of a single antiporter. *EMBO Rep*., **6**, 668–674.

Kedrov,A. *et al*. (2006) Observing folding pathways and kinetics of a single membrane protein. *J. Mol. Biol*., **355**, 2–8.

Muller,D.J. *et al*. (2002) Stability of bacteriorhodopsin α-helices and loops analyzed by single-molecule force spectroscopy. *Biophys. J*., **83**, 3578–3588.

Sapra,T. *et al*. (2006) Characterizing molecular interactions in different bacteriorhodopsin assemblies by single-molecule force spectroscopy. *J. Mol. Biol*., **355**, 640–650.

Tusnády,G.E. and Simon,I. (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics*,, **17**, 849–850.