2D Visualisation of SMFS data on membrane proteins

Annalisa Marsico[1], Kerstin Vanselow[1], Jing Wang[2], and Dirk Labudde[1]

[1]Department of Bioinformatics, Center of Biotechnology,
Dresden University of Technology, Germany
{annalisa.marsico, kerstin.vanselow, dirk.labudde}@biotec.tu-dresden.de

[2]Department of Information Systems, Massey University,
Palmerston North, New Zealand
j.w.wang@massey.ac.nz

Abstract

Misfolding of membrane proteins plays an important role in many human diseases such as retinitis pigmentosa, heredity deafness and diabetes insipidus. Single molecule force spectroscopy is a novel technique which measures the force necessary to pull a protein out of a membrane. Such force distance curves contain valuable information on the proteins structure. High-throughput force spectroscopy experiments generate hundreds of curves. These curves are then evaluated and aligned to generate a two-dimensional plot of unfolding patterns representing the distribution of force and contour length through a Gaussian distribution.

Keywords: Single molecule force spectroscopy, bacteriorhodopsin, force-distance curve.

1 Introduction

Integral membrane proteins play essential roles in cellular processes, including photosynthesis, transport of ions and small molecules, signal transduction and light harvesting. Despite the central importance of transmembrane proteins, the number of high resolution structures remains small due to the practical difficulties in crystallising them (Bowie, 2005). Many human disease-linked point mutations occur in transmembrane proteins. These mutations cause structural instabilities in a transmembrane protein, causing it to unfold or to fold in an alternative conformation (Filipek et al., 2003; Mirzadegan et al., 2003).

Protein folding is described by multidimensional energy landscapes or folding funnels, as a result of complex inter- and intra-molecular interactions (Onuchic and Wolynes, 2004).

Atomic force microscopy (AFM) is mostly known for its imaging capabilities, but it also provides a novel tool for detecting and locating forces on a single molecule level, such as the inter- and intra-molecular interactions that stabilise protein structures (Janshoff et al., 2000).

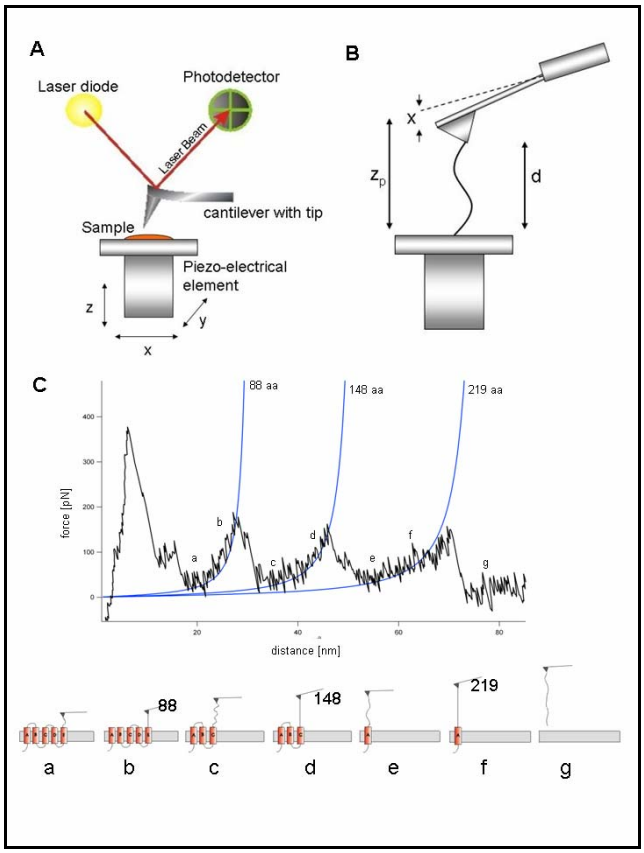

Fig. 1: A) Schematic representation of atomic force microscopy. The sample is mounted on a piezoelectrical element and scanned under a sharp tip attached to the cantilever. The voltage difference of the photo detector is proportional to the deflection of the cantilever. B) Unfolding of a transmembrane protein. A single molecule is kept between the tip and the sample while the tip-sample separation is continuously increased. C) Typical spectrum obtained from an experiment of unfolding of bacteriorhodopsin with the main peaks fitted by a hyperbolic function (worm-like chain model) and correlated to the unfolding of secondary structure elements (as cartoon on the button).

Single molecule force spectroscopy (SFMS) experiments allow measuring the stability of membrane proteins, and also probing the energy landscapes (Janoviak et al., 2004). In Figure 1A we show a schematic representation of the force spectroscopy instrumentation. Molecules with complex three-dimensional structures, such as proteins, can be unfolded in a controlled way. Titin and *bacteriorhodopsin* (bR) are examples of proteins that have been intensively studied (Oesterhelt et al., 2000; Janoviak et al., 2004; Sapra et al., 2006). When transmembrane proteins are unfolded in force spectroscopy experiments, during continuous stretching of the molecule, the applied forces are measured by the deflection of the cantilever and plotted against extension (tip-sample separation), yielding a characteristic *force-distance curve* (F-D curve) for the specific molecule under investigation, see Figure 1B. The F-D curve is the result of subsequent events of molecular interactions (Zhuang and Rief, 2003; Oesterhelt et al., 2000).

From the analysis of single molecule force spectra it is possible to associate the peaks to single potential barriers stabilising segments within membrane proteins. These segments can be represented by transmembrane helices, polypeptide loops or fragments. It is not yet clear how these interactions are established. Currently it is assumed that single amino acids as well as grouped amino acids can stabilise or destabilise membrane proteins (Faham et al., 2004). For a given molecule under study, the force-distance curves exhibit certain patterns, which contain information about strength and location of molecular forces established inside the molecule, about stable intermediates and reaction pathways, and the probability with which they occur. For membrane proteins the sequence of the unfolding peaks follows the amino acid sequence of the protein (Muller et al., 2002). For each peak the number of already unfolded amino acids can be determined from the length of the unfolded part of the polypeptide chain, obtained from a fit to a hyperbolic function, the *worm-like chain model* (WLC), of the given peak (Rief, et al., 1997). Consequently, with the peaks and the predicted secondary structure, it is possible to associate the peaks with structural domains, see Figure 1C (Muller et al., 2002).

The light-driven proton pump bR was chosen as a model system for this study since it represents one of the most extensively studied membrane proteins (Haupts et al., 1999; Oesterhelt, 1998). bR converts the energy of light (λ=500–650 nm) into an electrochemical proton gradient, which in turn is used for adenosine triphosphate (ATP) production by the cellular ATP synthase. Its structural analysis has revealed the photoactive retinal embedded in seven closely packed transmembrane α-helices (Belrhali et al., 1999; Essen et al., 1998; Grigorieff et al., 1996; Luecke et al., 1999; Mitsuoka et al., 1999), which builds a common structural motif among a large class of related G protein-coupled receptors (Baldwin, 1993; Helmreich and Hofmann, 1996; Kolbe et al., 2000; Palczewski et al., 2000; Royant et al., 2001). The bR helices are lettered A, B, C, D, E, F, and G, to which the C-terminal end is connected. (For illustration see also Figure 7.) With increasing knowledge of its structural and functional properties, bR has become a paradigm for α-helical membrane proteins in general and for ion transporters in particular (Lanyi, 1999). Together with adjacent lipids bR molecules assemble into trimers, which are packed into a two-dimensional hexagonal lattice of the purple membrane as a chemically distinct domain of the cell membrane.

The main unfolding pathway of bacteriorhodopsin is shown in Figure 1C and it is characterized by the presence of three main peaks that suggest a pairwise unfolding of the transmembrane helices. On the other hand, unfolding and analyzing many single bR molecules, it was found that, besides the main three peaks that occur on every F-D curve, other peaks, referred to as side peaks, occur with different probabilities besides the main peaks, indicating that bR exhibit sometimes different unfolding intermediates (Muller et al., 2002).

The course of an F-D curve represents conformational changes in the protein during the process of unfolding. In order to understand if specific point mutations affect the mechanical stability of bacteriorhodopsin, in this work we present a 2D visualisation and interpretation of SMFS experiments on the wild type bR and mutant P50A (Faham et al, 2004).

2 F-D pattern generation

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.

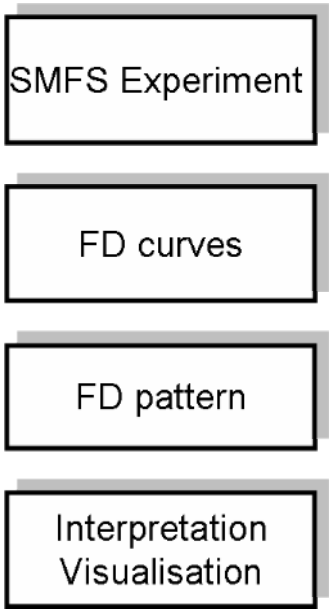

Fig. 2: Schematic representation of the F-D pattern building. From single molecule force spectroscopy derive force distance curves. For all given curves a pattern is found to interpret and visualize the data.

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 of their occurrence. 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 analyse 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. The single stages to derive an F-D pattern are shown in Figure 2. SMFS measurements as described in Figure 1A lead to a set of F-D curves. After the analyses of the whole set an F-D pattern of the investigated membrane protein can be derived, approximating a significant count of the underlying F-D curves. The last step is the visualisation and interpretation of the unfolding pathways.

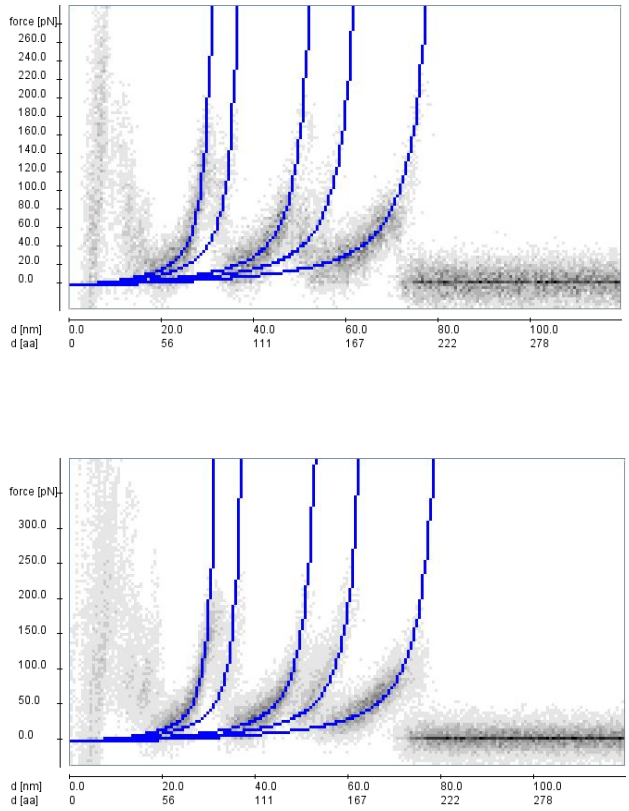

Fig. 3: A) Alignment of 26 F-D curves for bR wild type. The blue curves show the resulting worm-like chain model with its five peaks corresponding to the gray scale figure of the aligned F-D curves. B) Alignment and worm-like chain model of 54 F-D curves for mutant P50A.

Our procedure for data analysis is made up of three steps: data preparation; curve alignment with dynamic programming; peak detection and fitting. In the first stage we automatically identify and remove curves that do not contain unfolding signals, curves that exhibit an overall length indicative of partial or multiple unfolding events or curves showing peaks due to unspecific interactions (corrupted curves).

In the second stage of our method the curves are aligned using global multiple sequence alignment with dynamic programming (Marsico et al, 2006). For the detailed description of the overall spectra alignment procedure see Marsico (2006). The key reason for using a sequence alignment technique is the meaningful definition of matches/ mismatches, insertions and deletions. Matches and mismatches reward or penalise more or less fitting parts of the force curves. Insertions and deletions are important, as peaks in the curves may vary by up to six residues and as peaks may be missing completely between two curves.

In the third stage of our procedure we fit every peak of every aligned F-D curve with the worm-like chain model with a persistence length of 0.4nm and a monomer length of 0.36nm (Rief et al., 1997). The number of extended amino acids at each peak is then calculated using the contour length obtained from the WLC fitting. This allows us to assign unfolding events to structural segments as described in (Rief et al., 1997). To measure the unfolding force and probability of unfolding for each individual structural segment, every event of each curve was analysed. We show the automated superimposition of 26 F-D curves for bR wild type and 54 for mutant P50A. The alignments are shown as so called grey scale figures, see Figure 3.

The results show that bR wild type and its mutants exhibit the same unfolding steps (peaks at the same positions). We can see in Table 1 that the unfolding forces are also the same within experimental errors. Furthermore, we can observe that, while the probability of occurrence of the main peaks does not change from the wild type to the mutant, the probability of occurrence of the side peaks slightly vary. However, this variation is not so strong to support the idea that mutants exhibit different unfolding pathways from the wild type, as there are no additional or missing peaks in any of the analysed mutants.

WLC (L[aa])	Force [pN]		SD [pN]		Probability [%]	
	wild type	mutant P50A	wild type	mutant P50A	wild type	mutant P50A
88	150,28	173,46	24,1	23,9	100	100
104	112,00	122,85	21,7	28,3	45	37
148	105,53	112,27	30,3	34,7	100	100
175	75,46	108,55	25,9	43,2	64	79
220	93,68	103,51	33,5	26,7	100	100

Table 1: Listed representation of the results from the alignment and calculation of WLCs for our given samples. Rows two and three show the average force for the indicated WLC position. Rows four and five show the standard deviation, the last two rows are a proportional indication of the peaks to occur over all curves.

3 Visualisation

Figure 4 corresponds to the variation of all single unfolding events over the contour lengths L. Based on the evaluation of all F-D curves for wild type and mutant we can calculate the average unfolding force and the standard deviation, see Table 1.

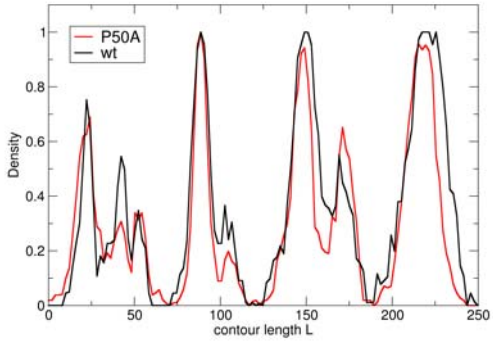

Fig. 4: Normalized density plot for both wild type and mutant F-D curves and their underlying peaks.

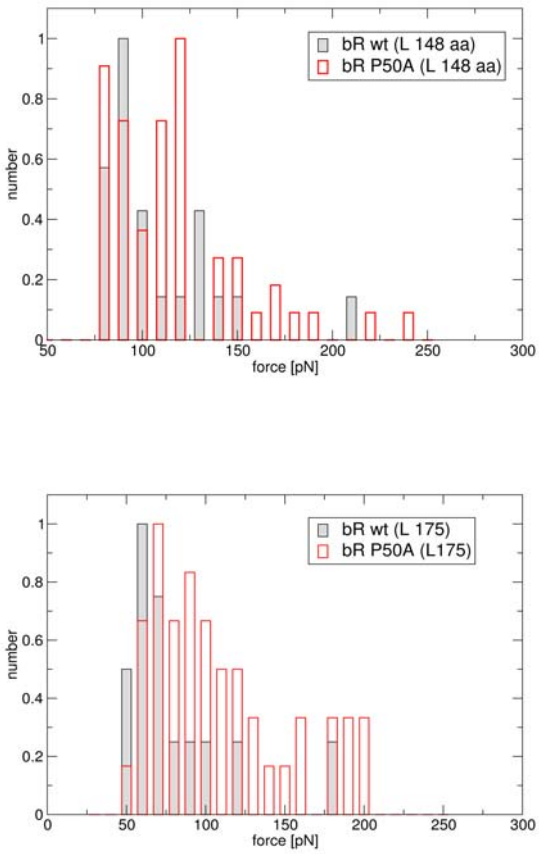

Fig. 5: A) Force distribution of the unfolding steps with contour length L=148. B) Force distribution of the unfolding steps with contour length L=175.

The analysis of the force distributions provides more information. The contour lengths L=220 includes the mutation place (P50A) and describes the unfolding of helices B and C. Details are shown in Figure 1C.

Figure 5 shows the force distribution of the unfolding step with contour length L=148 and L=175 for the wild type and mutant. Note that the measured forces at contour point L rather are the forces to pull the membrane protein to length L+1.

We can use both distribution types for a new way of describing the whole unfolding process. A two-dimensional plot shows the unfolding patterns from bR wild type and mutant P50A. The distributions and their variations throughout all the curves are illustrated with the mean (the average value of all force values for the given peak) and the standard deviation (represented as a coloured circle).

Figure 6 shows a dummy two-dimensional plot as intended to derive from the distributions from the results.

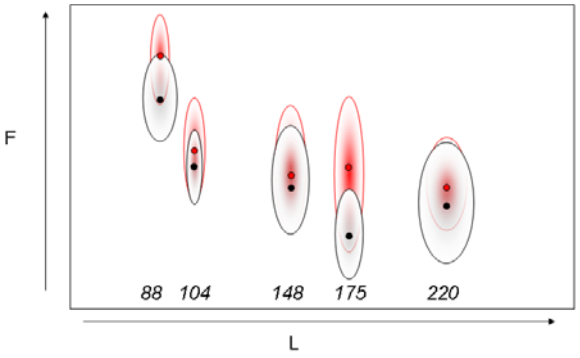

Fig. 6: Two-dimensional plot of unfolding patterns from bR wild type (black) and mutant P50A (red). The distributions are represented as Gaussian distribution for both force and contour length L (mean and standard deviation).

4 Discussion and outlook

On the basis of force and contour length distribution we derived a two-dimensional F-D pattern for bR wild type and mutant P50A. We can locate the greatest differences at contour length L=88 and L=175. At these positions the average applied force differs by more then 20pN between wild type and mutant.

We perform structural alignments between wild type and mutant P50A (Berman et al., 2002) using the DaliLite program for pairwise structure comparison and structure database search (Holm and Park, 2000). The program output reports Z-scores, which are similarity scores between two protein data bank (PDB) structures, normalized to their structure size. We determine Z-scores for all the four structural alignments of the wild type with the mutants. The root mean square deviation (RMSD) of pairwise alignment is 0.7 and the Z-score is 36.4, indicating a good super-imposition between wild type and mutant P50A. The main backbone structure is conserved within the mutants with minimal structural changes.

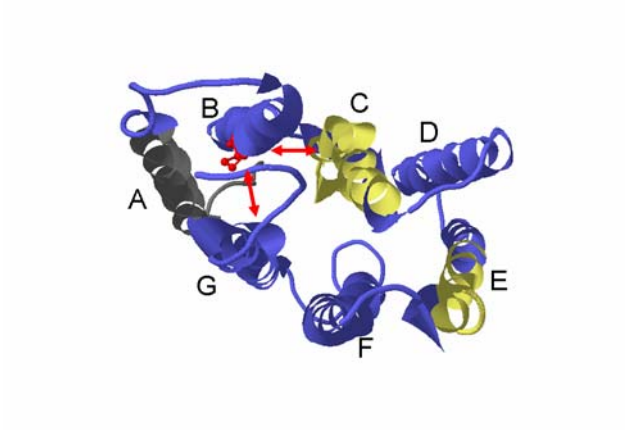

Fig. 7: Topview on bR wild type with clockwise lettered helices. Yellow and blue markings show sequential unfolding steps. The mutation is located in helix B (marked red). The WLC peak of 88 corresponds with G/F and whereas the WLC peak of 175 corresponds with C.

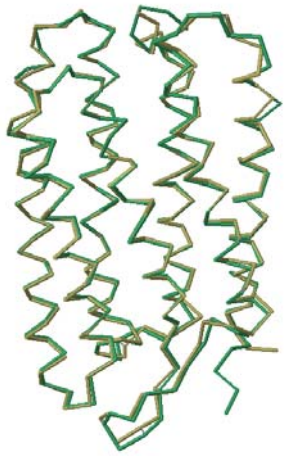

Fig. 8: Representation of wild type and mutant P50A through DaliLite. The illustration shows how close both membrane proteins are, but also points out the locations of variation between both wild type and mutant.

Analysing both the DaliLite and the Topview we can assume the interpretation that interaction patterns of bR wild type and the mutant change locally but not over the whole structure, see Figures 7 and 8.

The determined bottleneck of SMFS measurements is the analyses and alignment of the F-D curves. The better these tasks are solved, the better the peaks can be distinguished. These peaks can then be associated to single potential barriers stabilising segments within membrane proteins.

The new two-dimensional plot provides additional information on protein structure, and will be helpful in analysing membrane proteins and mutants that have been investigated before.

5 References

Baldwin, J.M. (1993): *The probable arrangement of the helices in G protein-coupled receptors*. EMBO J. 12, 1693-1703.

Belrhali, H., Nollert, P., Royant, A., Menzel, C., Rosenbusch, J.P., Landau, E.M. and Pebay-Peyroula, E. (1999): *Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 A resolution*. Struct. Fold. Des. 7, 909-917.

Berman, H. M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J.D., Zardecki, C. (2002): *The Protein Data Bank*. Acta Crystallogr D Biol Crystallogr. 58, 899-907.

Bowie, J.U. (2005): *Solving the membrane protein folding problem*. Nature 438(7068), 581-589.

Essen, L-O., Siegert, R., Lehmann, W.D., and Oesterhelt, D. (1989): *Lipid patches in membrane protein oligomers: Crystal structure of the bacterorhodopsin-lipid complex*. Proc. Natl. Acad. Sci. U.S.A. 95, 11673-11678.

Faham, S., Yang, D., Bare, E., Yohannan, S., Whitelegge, J.P., and Bowie, J.U. (2004): *Side-chain contributions to membrane protein structure and stability*. J Mol Biol, 335(1), 297-305.

Filipek, S., Teller, K. Palczewski, and Stenkamp R. (2003): *The crystallographic model of rhodopsin and its use in studies of other g-protein-coupled receptors*. Annu Rev Biophys Biomol Struct 32, 375-397.

Grigorieff, N., Ceska, T.A., Downing, K.H., Baldwin, J.M. and Henderson, R. (1996): *Electron-crystallographic refinement of the structure of bacteriorhodopsin*. J. Mol. Biol. 259, 393-421.

Helmreich, E.J.M., and Hofmann, K.P., (1996): *Structure and function of proteins in G protein coupled signal transfer*. Biochem. Biophys. Acta. 1286, 285-322.

Holm, L. and Park, J. (2000): *Dalilite Workbench for protein structure comparison*. Bioinformatics, 16(6), 566-567.

Janovjak, H., Struckmeier, J., Hubain, M., Kedrov, A., Kessler, M., and Muller, D.J. (2004): *Probing the energy landscape of the membrane protein bacteriorhodopsin*. Structure 12(5), 871-879.

Janshoff, A., Neitzert, M., Oberdorfer, Y. and Fuchs, H. (2000): *Force spectroscopy of molecular systems-single molecule spectroscopy of polymers and biomolecules*. Angew Chem Int Ed Engl 39(18), 3212-3237.

Kolbe, M., Besir, H., Essen, L.O. and Oesterhelt, D. (2000): *Structure of the light-driven chloride pump halorhodopsin at 1.8 resolution*. Science. 288, 1390-1396.

Lanyi, J.K. (1999): *Progress toward an explicit mechanistic model for the light-driven pump, bacteriorhodopsin*. FEBS Lett. 464, 103-107.

Luecke, H., Schobert, B., Richter, H.T., Cartailler, J.P. and Lanyi, J.K. (1999) *Structure of bacteriorhodopsin at 1.55 resolution*. J. Mol. Biol. 291, 899-911.

Marsico, A., Sapra, T., Muller, D., Labudde, D., and Schroeder, M. (2006): *A novel pattern recognition algorithm to classify membrane protein unfolding pathways with high-throughput single molecule force spectroscopy.* Bioinformatics. 23, 231-236.

Mirzadegan, T., Benko, G., Filipek, S. and Palczewski, K. (2003): *Sequence analyses of g-protein coupled receptors: similarities to rhodopsin*. Biochemistry 42(10), 2759-2767.

Mitsuoka, K., Hirai, T., Murata, K., Miyazawa, A., Kidera, A., Kimura, Y., and Fujiyoshi, Y. (1999): *The structure of bacteriorhodopsin at 3.0 resolution based on electroncrystallography: implication of the charge distribution*. J. Mol. Biol. 286, 861-882.

Muller, D. J., Kessler, M., Oesterhelt, F., Moller, C., Oesterhelt, D., and Gaub, H. (2002): *Stability of bacteriorhodopsin alpha-helices and loops analyzed by single-molecule force spectroscopy*. Biophys J 83(6), 3578-3588.

Oesterhelt, F., Oesterhelt, D., Pfeiffer, M., Engel, A., Gaub, H., and Muller, D.J. (2000): Unfolding pathways of individual bacteriorhodopsins. Science 288(5463), 143-146.

Onuchic, J.N. and Wolynes P.G. (2004): *Theory of protein folding*. Current Opinion in Structural Biology 14, 70-75.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M., and Miyano, M. (2000): *Crystal structure of rhodopsin: a G protein-coupled receptor*. Science 289, 739-45.

Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M., and Gaub H.E. (1997): *Reversible unfolding of individual titin immuno-globulin domains by AFM*. Science 276(5315).

Royant, A., Nollert, P., Edman, K., Neutze, R., Landau, E.M., Pebay-Peyroula, E., and Navarro, J. (2001): *X-ray structure of sensory rhodopsin II at 2.1-resolution*. Proc. Natl. Acad. Sci. U.S.A. 98, 10131 - 10136.

Sapra, K.T., Besir, H., Oesterhelt, D., and Muller D.J. (2006): *Characterizing molecular interactions in different bacteriorhodopsin assemblies by single-molecule force spectroscopy*. J Mol Biol 355(4), 640-650.

Zhuang, X. and Rief, M. (2003): *Single-molecule folding*. Curr Opin Struct Biol 13(1), 88-97.