Understanding of SMFS barriers by means of energy profiles

Frank Dressel[1], Annalisa Marsico[1], Anne Tuukkanen[1], Michael Schroeder[1] and Dirk Labudde[1]
[1]Biotechnologisches Zentrum, TU Dresden, 01062 Dresden, Germany

Abstract: In the last years, Single Molecule Force Spectroscopy was more and more used to gain insight into the fundamental principles behind protein structure and stability. Nevertheless, the interpretation of the experimental findings is not so easy and additional computational approaches are needed to interpret them. Here, we proposed an approach based on interaction patterns between amino acids to explain the emergence of SMFS unfolding barriers in the experiment. With our approach, we can predict around 64% of the experimentally detectable barriers.

1 Introduction

Integral membrane proteins play an essential role in cellular processes, including photosynthesis, transport of ions and small molecules, signal transduction and light harvesting. Mutations can reduce the function of the proteins dramatically and thus, lead to severe diseases when happening in membrane protein ([SM04]).

Recently, single-molecule force spectroscopy (SMFS) has proved to be a novel tool for detecting and locating inter- and intra-molecular forces on a single molecule level (see e.g. [PSK$^+$07]). In Fig. 1 A, a schematic representation of the force spectroscopy instrumentation is shown. While unfolding proteins with SMFS, the force applied to the protein is measured by the deflection of the cantilever. Moving the cantilever away from the protein yields a characteristic force-distance curve (F-D curve) (see Fig. 1 C). The so called Worm-like-chain model (see [BMSS94], [KP49]) relates the position of the peaks in the F-D curve to a position in the sequence of the protein. Such a peak defines a barrier with respect to mechanical unfolding, in the following referred to as unfolding barrier.

The reason for peaks in the F-D curve can be separated in two classes:

- Geometrical constraints (e.g. a whole helix has to be rotated/bended/kinked or steric hindrance).
- Energetic constraints: Some residues are stronger bounded than the rest.

However, it is not easy to differentiate these two types. Together with geometric constraints e.g., a weak interaction can be the reason for a barrier.

For an energetic caused barrier, not only single residues but patches of strongly interacting residues are needed. With the knowledge of interaction energy per residue, the so called

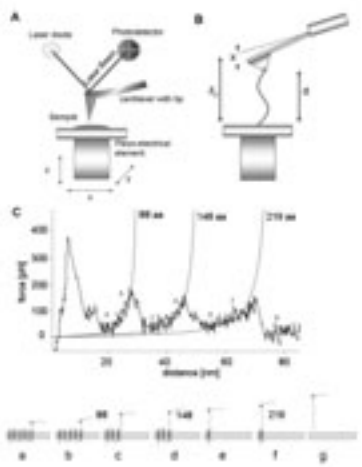


Figure 1: Experimental settings (A and B) and force-distance curve (C) for a SMFS experiment. The picture is taken from [MLS$^+$07].

energy profile, such residues are detectable and possible stabilizing regions can be identified. We derived an interaction scheme for amino acids within membranes based on a statistical potential. With this interaction scheme we are able to predict at least 60 % of the unfolding barriers detected by SMFS measurements for the used set of transmembrane proteins.

1.1 Analysed transmembrane proteins

We used a set of important membrane proteins, which covers different functions and topologies. The used proteins are listed in the following:

Bacteriorhodopsin is a membrane-bound energy-transducing enzyme, which uses light energy to transport protons across the bacterial membrane against the proton electrochemical gradient [Oes98]. It belongs to a special class of archaeal transmembrane proteins and consists of seven closely packed α-helices named A through G [KVM$^+$97],[LSR$^+$99]. The seven helices are arranged in two arcs, which form a membrane spanning cavity. Bacteriorhodopsin from $Halobacterium\ Salinarum$ forms trimers in the native purple membrane. However, trimerization does not seem to have any functional meaning and a bacteriorhodopsin monomer is capable of proton pumping. For the analysis, we used the PDB structure 1BRR.

Halorhodopsin is a light-driven chloride pump found in haloarchaea [Oes98, KBEO00]. In addition to chloride it transports bromide, iodide and nitrate using light energy to work

Protein name	Experimental barriers	references
Bacteriorhodopsin	(12 - 20) (24 - 32) (69-77) (86 - 94) (96 - 104) (139-147) (151 - 156) (157 - 163)	[JKO$^+$03]
Rhodopsin	(16-24) (22-30) (33-41) (56-64) (62-80) (84-92) (94-108) (105-113) (196-206) (239-247) (252-260) (269-277) (282-290) (306-314) (320-328)	[SPF$^+$06]
Antiporter	(362-370) (342-350) (324-332) (293-301) (277-284) (259-267) (221-229) (182-290) (159-167) (126-134) (66-74) (56-64)	[KJZ$^+$06]
Halorhodopsin	(243-252) (237-244) (223-231) (186-194) (182-190) (173-181) (168-176) (123-131) (110-118) (90-98) (47-55) (33-41)	[COM05]
Aquaporin	(249-257) (236-262) (199-207) (180-188) (157-165) (109-117) (82-90) (62-70) (35-43)	[MFS$^+$03]

Table 1: Experimentally determined position of the barriers in the sequence of the used proteins.

against their electrochemical gradients [SL82]. Halorhodopsin belongs to the subfamily of seven α helical archaeal rhodopsins. It has only 31 % sequence identity with bacteriorhodopsin from $H.\ salinarum$, but the transmembrane region is structurally well conserved. Halorhodopsin functions as a monomer, but occurs naturally as a trimer. For the analysis, we used the PDB structure 1E12.

Rhodopsins are members of a large family of G protein - coupled receptors (GPCRs). All GPCRS have a structure comprising of seven transmembrane α-helices [FTPS03]. Rhodopsins have the same arrangement of seven α-helices as archaeal rhodopsins. Rhodopsins have been revelead to work as dimers [FLF$^+$03]. For the analysis, we used the PDB structure 1F88.

Na^+/H^+ antiporters regulate sodium and proton concentration of cells as well as cell volume [PVGD01] and their activity is highly controlled by pH. NhaA comprises of 12 mostly α-helical transmembrane segments and forms a functional dimer. For the analysis, we used the PDB structure 1ZCD.

Aquaporins are integral membrane proteins which facilitate flow of water molecules across cellular membranes [BNEA99, dGG05]. They have been shown to take part in a large variety of physiological functions and to transport also glycerol, urea and carbon dioxide [BNEA99, ULSK03]. The structural core of aquaporins consists of six transmembrane α-helices [RRC$^+$01]. Human aquaporin-1 forms tetramers $in\ vivo$, but a single monomer has been shown to work as a channel. For the analysis, we used the PDB structure 1H6I.

For all these proteins, SMFS data are available (see Table 1).

2 Methods

2.1 Interaction energies for amino acids

We designed an interaction scheme, which is robust in terms of small structural deviations. The robustness circumvent a detailed dependency of the interaction scheme on the distance. Thus, we decide to look at the average shielding of an amino acid from the surrounding. This property includes already possible interactions with the environment as well as interactions with other amino acids. A similar approach was done by Wertz et al. [WS78] for determining hydrophobicity.

A handcrafted set[1] of membrane proteins from the PDB was scanned with respect to residues being inside or on the surface of the protein (outside, means facing the phospholipids, inside means surrounded by other amino acids). The investigated PDB structures in this study were not contained in the this set. We did this by first detecting the membrane position and secondly selecting just the residues in the membrane. The membrane was detected by a clustering approach, where the end points of the largest secondary structure elements are divided into two classes. The largest secondary structure elements span in first order between both membrane sides and thus can be used as hints for the membrane position. A least square fit was performed to get two planes defined by the two clusters of secondary structure end points. Everything between these two planes was assigned as "inside the membrane".

The amino acid specific property inside or outside was defined in the following way:

$$\left\{ \begin{array}{c} inside \\ outside \end{array} \right\} if \left\{ \begin{array}{c} \left|\vec{C}_\alpha - \vec{c}\right| < 5 \; or \; \left(\vec{C}_\alpha - \vec{C}_\beta\right) \cdot \left(\vec{C}_\alpha - \vec{c}\right) < 0 \\ else \end{array} \right\} \tag{1}$$

$\vec{C}_{\alpha/\beta}$ are the positions of the $C_{\alpha/\beta}$ atoms, $\vec{c}$ is the center of mass of all amino acids in a surrounding sphere with radius 10 Å. Only C_α's are considered for the center of mass.

The energy for bringing a single amino acid i from outside to inside can be calculated by a simple approach from statistical physics:

$$e_i \propto \ln\left(\frac{n_{i,in}}{n_{i,out}}\right), \tag{2}$$

where $n_{i,in}$ and $n_{i,out}$ are the number of inside and outside occurrence of amino acid i, respectively (see e.g. [WS78]).

A interaction e_{ij} between two amino acids i and j can then be described by the sum over the single amino acid solvation energy

$$e_{ij} = (e_i + e_j). \tag{3}$$

[1] 1BY5, 1E54, 1EHK, 1FEP, 1FX8, 1HXX, 1K24, 1KMO, 1L9H, 1M0K, 1NQE, 1OKC, 1P4T, 1QJ8, 1QJP, 1THQ, 1U7G, 1UYN, 1V9M, 2POR

It should be noted, that this ansatz gives both repulsive and attractive interactions for the used set of proteins. We used this surface definition, because it provides us with a simple binary classification. While using other measures like accessible surface area, one has to define (amino acid specific) cutoffs to define inside/outside.

3 Calculating the energy profiles

We assumed an interaction between two residues, if their two C_β atoms are less than $8\mathring{A}$ apart. For every interaction, we add the solvation term to the energy of each involved amino acids:

$$E_i = \frac{e_i}{2} \sum_{<i,j>} 1, \tag{4}$$

where $< i, j >$ denotes combinations of amino acids i and j, which have a distance less than $8\mathring{A}$ between the respective C_β's. In the end, this procedure results in an energy value for every amino acid i of a protein.

For detecting barriers we defined a threshold of $E_{threshold} = -4$ [a.u.]. Then, each profile is scanned and a barrier is assigned, if at least 3 consecutive residues have an energy below the threshold.

4 Results

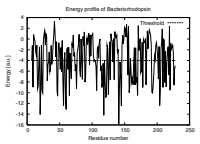
The experimentally determined barriers for all used proteins are shown in Table 3. The predicted ones are shown in Fig. 2(c), 3(c), 4(c), 5(c) and 6(c) respectively.

For comparison with the experiment, we considered the published barrier position from the experiment plus $\Delta_{exp} = \pm 4$ amino acids. This is due to systematic experimental errors like intrinsic movement of the cantilever and others. This errors are reported to be in the range of 3-7 amino acids (see e.g. [KZJ$^+$04]), depending on the used cantilever.

Name	Correctly predicted	Number of barriers in experiment
Bacteriorhodopsin	6	8
Halorhodopsin	6	12
Rhodopsin	7	15
Antiporter	8	12
Aquaporin	6	9

Table 2: Predicted barriers vs. experimental ones

Table 2 shows the number of predicted barriers for all the five analysed proteins, compared with the number of barriers detected in SMFS experiments. Around 60% of the barriers



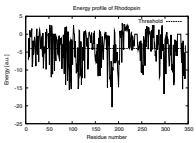
(a)

(b)

Predicted barriers
16 - 18 31 - 33 74 - 76
105 - 107 113 - 122 **141 - 145**
160 - 164 168 - 170 192 - 195

(c)

Figure 2: Energy profile (a) and location of barriers in Bacteriorhodopsin (b). The barriers detected by SMFS and energy profile are colored in red. Barriers detected by SMFS only are shown in cyan. The ones detected by energy profile only are colored in pink. Predicted barriers for Bacteriorhodopsin are shown in subfigure (c). The start / stop position of the barriers are given by the amino acid numbers in the primary sequence of the protein. The predicted barriers that correspond to experimentally detected barriers are shown in bold.



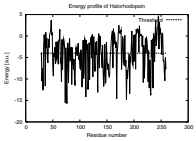
(a)

(b)

Predicted barriers
92 - 94 108 - 111 120 - 122
149 - 153 181 - 188 **195 - 202**
251 - 253 279 - 281 297 - 299
308 - 310 314 - 317 340 - 343

(c)

Figure 3: Energy profile (a) and location of barriers in Rhodopsin (b). Predicted barriers for Rhodopsin are shown in subfigure (c).



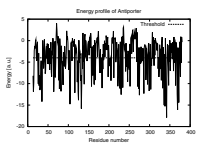
(a)

(b)

Predicted barriers
30 - 32 50 - 53 64 - 66
87 - 91 98 - 101 103 - 105
143 - 145 150 - 156 **184 - 194**
228 - 231

(c)

Figure 4: Energy profile (a) and location of barriers in Halorhodopsin (b). Predicted barriers for Halorhodopsin are shown in subfigure (c).



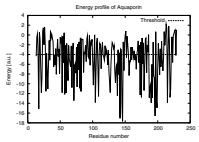
(a)

(b)

Predicted barriers
28 - 36 84 - 87 103 - 105
122 - 125 130 - 133 180 - 183
201 - 206 **225 - 227 270 - 273**
303 - 305 340 - 342 371 - 373

(c)

Figure 5: Energy profile (a) and location of barriers in Antiporter (b). Predicted barriers for Antiporter are shown in subfigure (c).



(a)

(b)

Predicted barriers
20 - 22 30 - 32 **40 - 42**
64 - 66 71 - 74 100 - 102
116 - 118 125 - 128 134 - 138
155 - 158 187 - 192 194 - 196
198 - 200 207 - 209

(c)

Figure 6: Energy profile (a) and location of barriers in Aquaporin (b). Predicted barriers for Aquaporin are shown in subfigure (c).

are detected by our method. A match between a predicted barrier and an experimentally determined one is established when they overlap or their boundaries are less than $\Delta_{profile}$ amino acids apart. To incorporate geometric constraints, we used a $\Delta_{profile}$ of ± 3 amino acids (corresponding to the smallest experimental error). In most cases, the predicted barriers fits with the experimental ones with $\Delta'_{profile} = \pm 1$ amino acids or better.

A remarkable result is the following: all the structural barriers for all proteins that, according to SMFS measurements, are found having 100% of probability of occurrence (unlike the side peaks that occur only on some force-distance traces) are correctly predicted by our method. This suggests that major unfolding barriers in the experiments are due to energetic reasons. Minor events, whose origin is not yet clear, are more difficult to detect with this kind of approach. The weakest prediction, in terms of correctly predicted barriers, is obtained for Rhodopsin. This can be easily explained by considering the fact, that SMFS experiments on Rhodopsin already showed a high variability in the distribution of detected peaks among the experiments.

The values for precision and recall are given in Table 4. We applied our method also to 3D-structures taken from a molecular dynamics simulation run. X-ray crystallography provides a time-averaged snapshot of the structure, whereas MD simulations can be used to study the dynamics and stability of protein structures. We used the MD simulations to get a conformational sampling for the used proteins. The resultant MD structures were clustered based on their backbone C_α atom RMSDs during one nanosecond equilibrium simulation. Representatives where taken from 4 found clusters. There was no significant change in predicted barriers between the MD and the original PDB structures.

Name of Protein	Experimental barriers	References
Bacteriorhodopsin	**12 - 20 24 - 32 69-77** 86 - 94 **96 - 104 139-147** 151 - 156 **157 - 163**	[JKO$^+$03]
Rhodopsin	16-24 22-30 33-41 56-64 62-80 **84-92 94-108 105-113 196-206** 239-247 **252-260** 269-277 **282-290 306-314** 320-328	[SPF$^+$06]
Antiporter	**362-370 342-350** 324-332 **293-301 277-284 259-267 221-229 182-190** 159-167 **126-134** 66-74 56-64	[KJZ$^+$06]
Halorhodopsin	243-252 237-244 **223-231 186-194 182-190** 173-181 168-176 123-131 110-118 **90-98 47-55 33-41**	[COM05]
Aquaporin	249-257 236-262 **199-207 180-188 157-165 109-117** 82-90 **62-70 35-43**	[MFS$^+$03]

Table 3: Experimentally detected barriers for the used set of membrane proteins. The barriers, which could be predicted with our energy profile-based method are shown in bold.

5 Discussion

While pulling proteins out of the membrane, different unfolding events occur and are detectable using force-distance curves. Around 60% of these patterns can be explained by a pure energetic based approach presented in this paper, there the barriers correspond to strongly interacting regions.

Although the energetics-only view is quite successful, the incorporation of geometry would enhance the understanding of mechanical unfolding a lot. Not every patch of strongly interacting residues is able to be a barrier just by geometric reasons, whereas on the other hand not every barrier has an energetic reason. Another source of errors is the simplicity of the energy function used. Nevertheless, this effect should be small due to the averaging over at least 3 consecutive amino acids.

Why are patches of strongly interacting residues responsible for barriers? If one residue is moved due to an applied force it forces neighboring residues to move too. Thus a whole bunch of residues has to be moved and stabilizing bonds have to be broken. If (all) these residues are strongly interacting, this results in a pronounced barrier and our approach is able to find them.

6 Summary

We calculated energy profiles for membrane proteins with a wide range of function and structure. For this set, we are able to assign for at least 64 % of the barriers an energetic

Protein name	Recall in %	Precision in %
Bacteriorhodopsin	75	67
Halorhodopsin	50	56
Rhodospin	44	58
Antiporter	67	73
Aquaporin	73	57

Table 4: Precision and recall for the used proteins with respect to the experimental data.

reason. On the other hand, there is a large space left for geometric origin of barriers. To our knowledge, this is the first time that such an energetic approach, based on a coarse grained model, is used to predict with a reasonable accuracy barriers for all the proteins probed by SMFS.

References

[BMSS94] C Bustamante, J F Marko, E D Siggia, and S Smith. Entropic elasticity of lambda-phage DNA. *Science*, 265(5178):1599–1600, Sep 1994. Letter.

[BNEA99] M. Borgnia, S. Nielsen, A. Engel, and P. Agre. Cellular and molecular biology of the aquaporin water channels. *Annu Rev Biochem*, 68:425–458, 1999.

[COM05] D. Cisneros, D. Oesterhelt, and D. J. Muller. Probing origins of molecular interactions stabilizing the membrane proteins halorhodopsin and bacteriorhodopsin. *Structure*, 13(2):235–242, 2005.

[dGG05] Bert L de Groot and Helmut Grubmller. The dynamics and energetics of water permeation and proton exclusion in aquaporins. *Curr Opin Struct Biol*, 15(2):176–183, Apr 2005.

[FLF+03] Dimitrios Fotiadis, Yan Liang, Slawomir Filipek, David A Saperstein, Andreas Engel, and Krzysztof Palczewski. Atomic-force microscopy: Rhodopsin dimers in native disc membranes. *Nature*, 421(6919):127–128, Jan 2003.

[FTPS03] Slawomir Filipek, David C Teller, Krzysztof Palczewski, and Ronald Stenkamp. The crystallographic model of rhodopsin and its use in studies of other G protein-coupled receptors. *Annu Rev Biophys Biomol Struct*, 32:375–397, 2003.

[JKO+03] Harald Janovjak, Max Kessler, Dieter Oesterhelt, Hermann Gaub, and Daniel J Muller. Unfolding pathways of native bacteriorhodopsin depend on temperature. *EMBO J*, 22(19):5220–5229, 2003.

[KBEO00] M. Kolbe, H. Besir, L. O. Essen, and D. Oesterhelt. Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution. *Science*, 288(5470):1390–1396, May 2000.

[KJZ+06] Alexej Kedrov, Harald Janovjak, Christine Ziegler, Werner Kuhlbrandt, and Daniel J Muller. Observing folding pathways and kinetics of a single sodium-proton antiporter from Escherichia coli. *J Mol Biol*, 355(1):2–8, 2006.

[KP49] O Kratky and G Porod. Roentgenuntersuchung geloester Fadenmolekle. *Rec. Trav. Chim. Pays-Bas.*, 68:1106, 1949.

[KVM+97] Y. Kimura, D. G. Vassylyev, A. Miyazawa, A. Kidera, M. Matsushima, K. Mitsuoka, K. Murata, T. Hirai, and Y. Fujiyoshi. Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature*, 389(6647):206–211, Sep 1997.

[KZJ+04] Alexej Kedrov, Christine Ziegler, Harald Janovjak, Werner Kuhlbrandt, and Daniel J. Muller. Controlled unfolding and refolding of a single sodium-proton antiporter using atomic force microscopy. *J Mol Biol*, 340(5):1143–1152, 2004.

[LSR+99] H. Luecke, B. Schobert, H. T. Richter, J. P. Cartailler, and J. K. Lanyi. Structure of bacteriorhodopsin at 1.55 Å resolution. *J Mol Biol*, 291(4):899–911, Aug 1999.

[MFS+03] C. Moller, D. Fotiadis, K. Suda, A. Engel, M. Kessler, and D. J. Muller. Determining molecular forces that stabilize human aquaporin-1. *J Struct Biol*, 142(3):369–378, 2003.

[MLS+07] Annalisa Marsico, Dirk Labudde, Tanuj Sapra, Daniel J Muller, and Michael Schroeder. A novel pattern recognition algorithm to classify membrane protein unfolding pathways with high-throughput single-molecule force spectroscopy. *Bioinformatics*, 23(2):231–236, Jan 2007.

[Oes98] D. Oesterhelt. The structure and mechanism of the family of retinal proteins from halophilic archaea. *Curr Opin Struct Biol*, 8(4):489–500, Aug 1998.

[PSK+07] Paul S-H Park, K Tanuj Sapra, Michal Kolinski, Slawomir Filipek, Krzysztof Palczewski, and Daniel J Muller. Stabilizing effect of Zn2+ in native bovine rhodopsin. *J Biol Chem*, 282(15):11377–11385, Apr 2007.

[PVGD01] E. Padan, M. Venturi, Y. Gerchman, and N. Dover. Na^+/H^+ antiporters. *Biochim Biophys Acta*, 1505(1):144–157, May 2001.

[RRC+01] G. Ren, V. S. Reddy, A. Cheng, P. Melnyk, and A. K. Mitra. Visualization of a water-selective pore by electron crystallography in vitreous ice. *Proc Natl Acad Sci U S A*, 98(4):1398–1403, Feb 2001.

[SL82] B. Schobert and J. K. Lany. Halorhodopsin is a light-driven chloride pump. *J Biol Chem*, 257(17):10306–10313, Sep 1982.

[SM04] Charles R Sanders and Jeffrey K Myers. Disease-related misassembly of membrane proteins. *Annu Rev Biophys Biomol Struct*, 33:25–51, 2004.

[SPF+06] K. T. Sapra, P. S.-H. Park, S. Filipek, A. Engel, D. J. Muller, and K. Palczewski. Detecting Molecular Interactions that Stabilize Native Bovine Rhodopsin. *J Mol Biol*, 358(1):255–269, 2006.

[ULSK03] Norbert Uehlein, Claudio Lovisolo, Franka Siefritz, and Ralf Kaldenhoff. The tobacco aquaporin NtAQP1 is a membrane CO_2 pore with physiological functions. *Nature*, 425(6959):734–737, Oct 2003.

[WS78] D H Wertz and H A Scheraga. Influence of water on protein structure. An analysis of the preferences of amino acid residues for the inside or outside and for specific conformations in a protein molecule. *Macromolecules*, 11(1):9–15, 1978.