Biochimica et Biophysica Acta 1837 (2014) 656–663

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbabio

Review

Retinal proteins as model systems for membrane protein folding ☆

Oznur Tastan [a,1], Arpana Dutta [b,1], Paula Booth [c], Judith Klein-Seetharaman [d,*]

[a] Department of Computer Engineering, Bilkent University, Ankara, Turkey
[b] Department of Physiology and Biophysics, Case Western Reserve University, Cleveland, USA
[c] School of Biochemistry, University of Bristol, UK
[d] Division of Metabolic and Vascular Health, Warwick Medical School, University of Warwick, Coventry, UK

ARTICLE INFO

Article history:
Received 30 July 2013
Received in revised form 19 November 2013
Accepted 28 November 2013
Available online 12 December 2013

Keywords:
Rhodopsin
Bacteriorhodopsin
Membrane protein folding
Denatured states

ABSTRACT

Experimental folding studies of membrane proteins are more challenging than water-soluble proteins because of the higher hydrophobicity content of membrane embedded sequences and the need to provide a hydrophobic milieu for the transmembrane regions. The first challenge is their denaturation: due to the thermodynamic instability of polar groups in the membrane, secondary structures in membrane proteins are more difficult to disrupt than in soluble proteins. The second challenge is to refold from the denatured states. Successful refolding of membrane proteins has almost always been from very subtly denatured states. Therefore, it can be useful to analyze membrane protein folding using computational methods, and we will provide results obtained with simulated unfolding of membrane protein structures using the Floppy Inclusions and Rigid Substructure Topography (FIRST) method. Computational methods have the advantage that they allow a direct comparison between diverse membrane proteins. We will review here both, experimental and FIRST studies of the retinal binding proteins bacteriorhodopsin and mammalian rhodopsin, and discuss the extension of the findings to deriving hypotheses on the mechanisms of folding of membrane proteins in general. This article is part of a Special Issue entitled: Retinal Proteins—You can teach an old dog new tricks.

1. Introduction

Membrane proteins (MPs) constitute typically more than 20% of genomes, yet compared to soluble proteins, the principles by which they assume their three-dimensional structures are largely unknown. This is predominantly due to the fact that it is experimentally difficult to study MPs, especially under denaturing conditions, the "bread and butter" of protein folding studies. Among MPs, two retinal proteins have model system status for MP folding studies, mammalian rhodopsin (MR) and bacteriorhodopsin (BR). Retinal proteins are excellent models with which to perform such experiments because the retinal chromophore acts as a natural reporter of the final folded state. Experimental efforts to denature and renature these two retinal proteins are therefore reviewed here, together with a comparison of experimental and computational studies aimed at understanding the molecular properties of denatured states and pathways of folding. Their possible significance for understanding the principles of folding of other MPs is also discussed.

BR is a light-driven proton pump in the purple membrane of *Halobacteria salinaria*. It consists of seven transmembrane (TM) α-helices that are connected by short loops and a retinal chromophore covalently bound to a lysine residue. Ground-breaking early work by the late H. Gobind Khorana and his co-workers established BR as a folding paradigm by fully denaturing and refolding BR *in vitro* [1]. Their subsequent seminal studies of its fragments led to the predominant MP folding theory proposed by Popot and Engelman: the two-stage hypothesis [2,3], see below.

MR is the dim light photoreceptor and a prototypical G protein coupled receptor. It also consists of seven TM helices connected by cytoplasmic and extracellular loops. Mutations in MR have been linked to the retinal degenerative disease, Retinitis Pigmentosa [4]. This is an inherited disorder that causes night blindness and leads to progressive loss of vision in later life due to a gradual reduction of rod and cone photoreceptor cells. Although rare, it is the main inherited retinal degeneration disease, with about 1 in 4000 people affected worldwide [5]. According to the human gene mutation database, more than 150 MR mutations are known to cause this disease [6]. Khorana and co-workers have shown that most of these mutations lead to misfolding and/or instability of MR as a result of which the receptor protein is retained in the endoplasmic reticulum and is incapable of binding its chromophore *11-cis* retinal [7–9]. Thus, absence of correctly folded MR and presence of misfolded MR in the rod outer segments contribute to the major causes of death of rod cells in autosomal dominant cases [10]. An understanding of folding mechanisms of MR may help design effective strategies to combat Retinitis Pigmentosa by providing deeper insights into the underlying causes of misfolding. In order to

☆ This article is part of a Special Issue entitled: Retinal Proteins—You can teach an old dog new tricks.

* Corresponding author at: Biomedicine and Systems Biology, Division of Metabolic and Vascular Health, Warwick Medical School, Rm. B037, Gibbet Hill, University of Warwick, Coventry CV4 7AL, UK. Tel.: +44 02476 573 806; fax: +44 02476 574 637.
E-mail address: j.klein-seetharaman@warwick.ac.uk (J. Klein-Seetharaman).

[1] Joint first co-authors.

0005-2728/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbabio.2013.11.021

begin to understand the folding mechanism of MR, denaturation and stability studies in the cell, *in vitro* and in silico have been carried out (reviewed in [11–19]).

2. Chemical denaturation studies of membrane proteins

Critical to the understanding of protein folding is the reverse process, denaturation from the native, folded states. This is because chemical denaturation studies can provide a useful measure of protein stability, which is intricately linked to folding. The method has been extensively used for water-soluble proteins, for which a denaturation curve is generated typically by the addition of urea. Re-folding from the urea-denatured state then gives a renaturation curve. In the simplest case the denaturation and renaturation curves overlap and the reaction is a two-state equilibrium between the unfolded and folded states. In this case, the equilibrium constant and free energy of unfolding can be determined and plotted against denaturant concentration to reveal a linear free energy relationship that enables extrapolation to a free energy value in the absence of a denaturant [20]. This latter free energy is taken as the intrinsic, thermodynamic stability of the protein. The chemical denaturation approach has been successfully applied to α-helical MPs, however it has proved decidedly challenging to refold such proteins and demonstrate equilibrium refolding (see reviews [21–23]). Furthermore, there are a number of complications arising from the membrane mimetic used in the experiments [24]. Following the first application of this thermodynamic approach to *Escherichia coli* diacylglycerol kinase, the method has been successfully applied to BR. In both cases, linear free energy relationships have been observed against the concentration of the denaturant, sodium dodecyl sulfate (SDS) [25,26]. The resulting free energy of BR in the absence of a denaturant provides a useful reference measure of the protein's inherent stability. Moreover, it also provides the basis for obtaining information on the strength of interhelical hydrogen bonds and the folding transition state using ϕ value analysis [27–29]. ϕ analysis involves introducing a mutation into the protein and determining the change this mutation induces in the free energy of folding and the activation energy for the reaction. Thus the change in free energy of the transition state can be compared to that in the folded state, which in turn allows transition state structure to be inferred. BR was the focus of the first ϕ value study of the folding transition state for a MP [27]. Interestingly, helix formation in the transition state correlates with sequence position, with the second helix being largely formed in the transition state while the final C-terminal helix is predominantly unstructured [28]. This polarized transition state structure also correlates with the order of TM insertion into the cell membrane, showing that *in vitro* measurements, correspond well to the cellular situation [24].

In the cell, the folding of MPs may be more complex due to the presence of chaperone functions. For example, the *Drosophila* protein NinaA is a known chaperone for rhodopsin, but its role in folding of rhodopsin is not clear [30]. More generally, MPs pass through and are partially retained by the translocon, a channel in the endoplasmic reticulum (ER) membrane that allows entry of proteins into the ER for posttranslational modifications. Although it is still not clear whether MPs begin to fold in the translocon and if they are partially or completely folded before entering into the ER membrane, studies using an *in vitro* transcription–translation system to investigate the role of the translocon in the mechanism of MP folding indicate that the translocon takes on the role of a chaperone during MP folding [31,32]. It appears that TM helices 1–4 translocate to the membrane sequentially after each TM helix is synthesized, TM helices 5–7 are retained in the translocon until opsin synthesis is complete. It was shown that retention of the TM helix depends on the properties of the helix since replacement of TM7 with TM3 resulted in spontaneous exit of TM3 after its synthesis. This raises the question, why is it necessary for some of the TM segments to be retained in the translocon until opsin synthesis is complete? It is tempting to speculate that long-range interactions may be required for proper folding to occur *in vivo*. Furthermore, additional chaperone proteins participate in MP folding *in vivo* [33].

3. Denaturation and refolding of bacteriorhodopsin

The chemically denatured state is the closest possible equivalent to a nascent polypeptide chain *in vivo* that can be achieved *in vitro*. Finding conditions under which a protein can be denatured allows subsequent in depth analysis of the molecular characteristics of the unfolded states, including identification of residual structure (e.g. [34]). Ideally, the denatured conformational ensemble is devoid of secondary and tertiary structure — as much as possible since a true random coil may not exist [35]. However, the random coil ideal is achieved for MPs even less than for soluble proteins. Due to the presence of membranes and membrane mimetics, denaturation is exceedingly difficult for MPs. In fact, BR is one of only two MPs that have been fully denatured to date (the other MP being CopA).

Refolding from the denatured states is the subsequent logical step in the study of folding mechanisms, and has been achieved successfully for many soluble proteins. In contrast, very few MPs can be refolded, even from only partially denatured states. Again, BR is one of the few exceptions. After using trifluoroacetic acid (TFA) as a denaturant, BR was subsequently successfully refolded into phospholipids via an initial transfer into SDS micelles [1]. The possibility of refolding BR *in vitro* from denatured states established BR as a model system for studying the thermodynamics and kinetics of folding of helical MPs [36], as outlined above. However, these studies involve refolding BR from a partially SDS denatured state which retains about half of the native helical structure [26]. Additional kinetic studies of refolding BR from an SDS-denatured state into lipid/detergent micelles have enabled detection of folding intermediates in its folding pathway [11,37–41].

4. Denaturation and efforts to refold rhodopsin

Identification of conditions under which denatured states of MR can be experimentally studied has proven more difficult than for BR. Unlike BR, it has not been possible so far to fully denature MR, and even partially denatured states have not been amenable to refolding. An extreme case demonstrating irreversibility is shown for illustrative purposes in Fig. 1, where MR is denatured by high temperatures [42]. While increasing the temperature progressively shows the denaturation of the protein as evidenced by the change in ellipticity at 222 nm, the subsequent reverse lowering of the temperature does not result in a refolding curve that traces the denaturation curve. This indicates that the protein is not

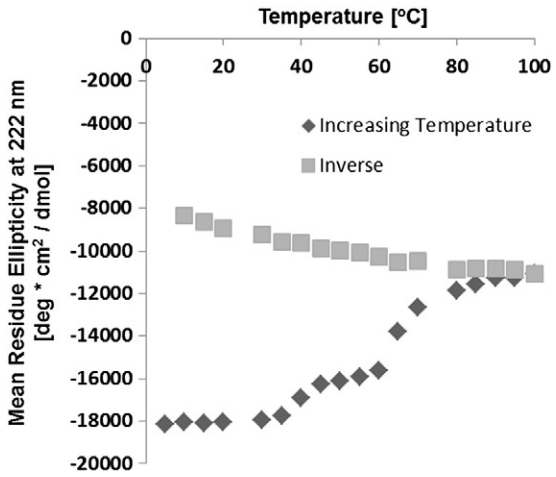

Fig. 1. Irreversibility of unfolding of MR as evidenced by thermal melting curves. Secondary structure is measured by circular dichroism. Closed symbols are for the heating cycle (5 °C–100 °C). Open symbols are for the subsequent cooling cycle (100 °C–5 °C). A concentration of 2.5 μM of MR in 0.05% dodecyl maltoside micelles was used.

refolding, typically due to the irreversible aggregation of the protein. The presence of aggregates of MR at high temperatures has been shown in [18].

Early studies used the denaturants guanidine hydrochloride (GuHCl) and urea to understand the stability of MR in native membranes [43,44]. The total decrease in mean residue ellipticity (MRE) at 222 nm, which reports on the amount of helicity, was ~50% for GuHCl. MR in its native environment was shown to refold. In another report, GuHCl denaturation of MR in 2% digitonin was carried out [43]. No changes in secondary structure on the addition of urea and GuHCl were reported, and the focus was therefore on retinal binding and hydrolysis of the Schiff base. The effect of urea on absorbance at 500 nm of MR showed that even at 8 M concentration, urea has no effect on rhodopsin [45]. It was reported that opsin, the apo form of MR, is more unstable towards urea denaturation than MR since urea does begin to denature opsin at the low concentration of 1 M [43]. When compared to GuHCl, urea denatures opsin at lower concentration than GuHCl, with the latter inducing denaturation at 3.5 M [43]. Recently, denaturation of opsin in phospholipid bicelles with urea was reported [46]. It was shown that denaturation of opsin in phospholipid/detergent mixed micelles of DMPC/CHAPS with 4 M urea leads to an irreversible unfolding, corresponding to a decrease of 50% in helical content of opsin.

A recent survey of different denaturing conditions, including urea, GuHCl, detergents and combinations of different denaturants showed that only high concentrations of SDS and combinations of SDS and urea are suitable for studying denaturation of MR in detergent micelles without aggregation [18]. These studies have opened the door to in-depth characterization of MR denatured states [17]. However, refolding from any of the conditions investigated has not been possible even with extensive variation of different lipid and detergent systems (unpublished results).

5. Comparison of rhodopsin and bacteriorhodopsin using FIRST

The experimental conditions used to (de)stabilize and study BR and MR are different, making a direct comparison difficult. To complement such experimental studies, computational approaches that only require the input of a crystal structure have been applied [16,47,48]. Although crystal structures are static data, it was shown repeatedly that computational methods can predict protein folding nuclei from native state structures [49]. A common approach is normal mode analysis where each amino acid is represented as a bead, and beads are connected by springs based on a cut-off distance. The movement of the beads relative to each other can be described by a matrix version of Newton's second law of motion, which when solved produces a protein's intrinsic motions, called normal modes (with the respective frequencies). Although normal modes describe the motions accessible to a protein, they are themselves static properties that are directly computed from crystal structure coordinates. However, the low frequency/large amplitude motions define the softest directions (that are easiest to excite) on the energy landscape, therefore conformational transitions favorably happen in accordance with those easily accessible directions. Static structure defines the relevant reaction coordinates to monitor a kinetic event and stores sufficient information to find where the softest directions are. Of course, charting those directions, by itself, does not suffice to generate the exact time evolution trajectory of a system. A predicted pathway then does not refer to the complete time trajectory, but to the most likely path to be taken.

The algorithm we deploy, the so-called Floppy Inclusions and Rigid Substructure Topography (FIRST) method [50], is an all atom analysis of the rigidity and flexibility of protein structures. In FIRST, the protein is modeled as a constraint network where each atom is a node and the nodes are subject to constraints defined by the covalent bonds, hydrogen bonds, hydrophobic tethers and salt bridges [50]. Once the constraint network is constructed, FIRST calculates for every bond whether it is part of a rigid cluster or part of an underconstrained region. Through the dilution of hydrogen bond and salt bridge contacts and analyzing the constraint network after each breakage of a hydrogen bond or salt bridge (in the order of their strength), an unfolding pathway is obtained. The applicability of FIRST shows that protein unfolding is inherently similar to the melting of structural glasses, upon which the method is based. The melting of glasses is a second order, continuous transition, thus describing a dynamic process. In its application to proteins, the method was specifically developed to track loss of structural stability as a measure of unfolding and to identify a "folding core" as the rigid cluster that resists dissolution the longest. In an extensive comparison of experimentally determined folding cores in 29 diverse proteins, it was shown that FIRST predicted folding cores provide statistically significant enhancements over random correlation [49]. We therefore set out to investigate if a folding core can be observed in FIRST simulated unfolding of BR [48] and/or MR [47].

Most structures when analyzed with FIRST start with one large cluster, interrupted only by flexible regions within loops, visible here also for BR (Fig. 2A) and MR (Fig. 2B). The vertical axis refers to the hydrogen bond energy and the horizontal axis shows the residues along the protein sequence. Each line in the dilution plot indicates which residues are rigid and flexible at the corresponding hydrogen bond energy cutoff. Thin black lines represent residues with a flexible backbone; whereas each colored block identifies the rigid clusters a residue belongs to with different colors signifying different rigid clusters. The red cluster indicates the largest rigid cluster. Lines are shown only when there is a change in the backbone rigid clusters. For orientation purposes, the positions of the TM helices are outlined by thick black lines in the dilution plots. The simulated unfolding of BR (Fig. 2A) is dominated by interactions within individual TM helices which can be seen by the relatively rapid fragmentation of the largest rigid cluster (red) into distinct rigid clusters corresponding to each helix. By contrast, in the denaturation of MR, the largest rigid cluster is observed to contain segments from multiple helices and loops for most of the dilution (Fig. 2B). This persistent rigidity results from the interconnectivity of structural elements in MR and portrays a nonlocal cooperativity as opposed to the individuality of helices observed in BR denaturation.

In the dilution plot, the folding core intuitively corresponds to the most stable residues that resist denaturation the longest. Thus in previous studies, it was found that a reasonable definition for the folding core is the lowest line in the dilution plot where at least three consecutive residues are mutually rigid with at least three consecutive residues of another secondary structural element [50]. A residue is considered rigid if at least two of the backbone atoms are present in the rigid cluster. The lines of the folding cores of MR and BR according to the above definition are marked by arrows, coinciding with -4.7 and -2.3 kcal/mol in hydrogen bond energy cutoff, respectively. The rigid clusters at the folding core line are mapped onto the BR and MR structures shown in Fig. 2. In BR (Fig. 2A), we see the coincidence of the folding units with the positions of the TM helices. In MR (Fig. 2B), at the folding core energy cutoff the largest rigid cluster (red) constitutes a core of interconnected residues at the interface between the TM and extracellular domains, lining the retinal binding pocket towards the extracellular side. This folding core is characterized by long-range interactions involving amino acids close in space but distant in sequence comprising positions from both extracellular loop and TM regions.

Thus, the comparison of BR and MR using FIRST yields very different conclusions on the folding of these two structurally related proteins. In the case of BR (Fig. 2A), early during the simulated denaturation process, individual clusters break off, that correlate with the positions of the TM helices. This observation is consistent with the vast amount of experimental studies that demonstrate helices to be the major folding units in BR, the hallmark of the 2-stage hypothesis [2,3]. In this model, secondary structure elements form first, followed by tertiary interactions to form the helical bundle. By contrast, when FIRST is applied to MR (Fig. 2B), the initial largest cluster gradually becomes smaller until a region remains with residues at the interface between the TM and

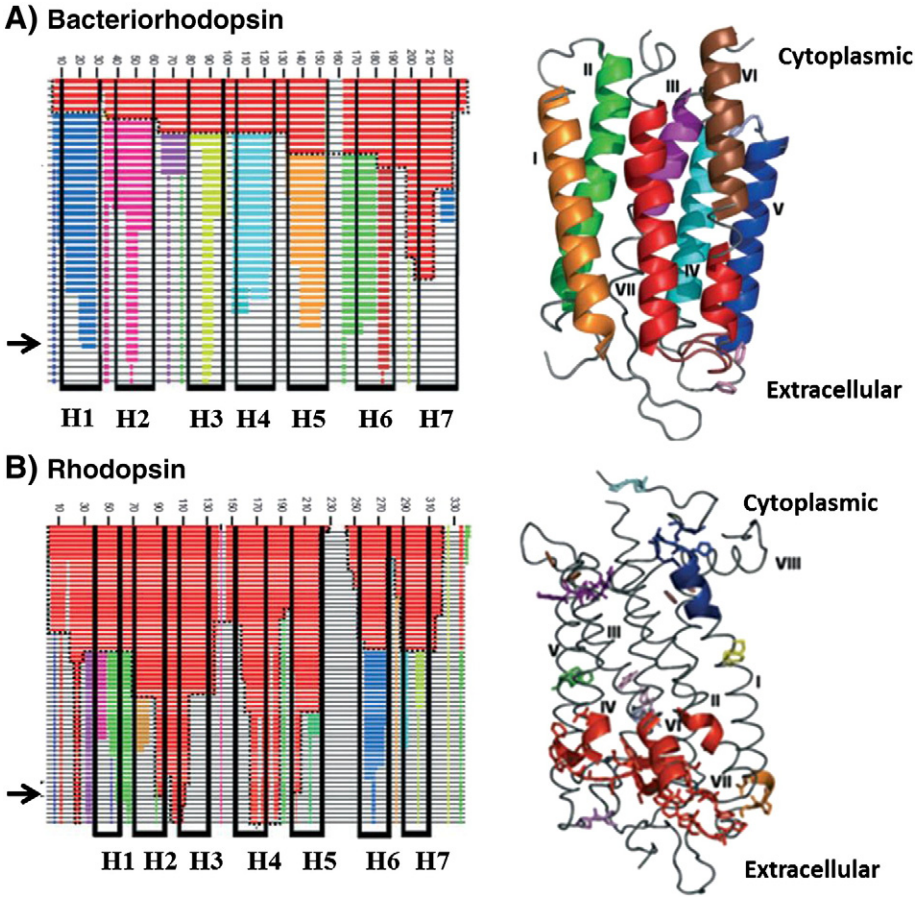

Fig. 2. FIRST unfolding of bacteriorhodopsin and mammalian rhodopsin. Simulated thermal denaturation plot and folding core clusters for A. BR (PDB ID code: 1c3w) and B. MR (PDB ID code: 1l9h). The vertical axis refers to the hydrogen bond energy and the horizontal axis shows the residues along the protein sequence. Each line in the dilution plot indicates which residues are rigid and flexible at the corresponding hydrogen bond energy cutoff. Lines are shown only when there is a change in the backbone rigid clusters. Thin black lines represent residues with a flexible backbone; whereas each colored block identifies the rigid clusters that a residue belongs to with different colors signifying different rigid clusters. Red color represents the largest rigid cluster. Arrows indicate the positions of the predicted folding cores, as defined by the last line in which three residues from independent secondary structure elements are part of the same cluster. The clusters at the folding core energy lines are mapped on the 3D structures of BR (A, right) and MR (B, right). The MR folding core is at the interface between the extracellular and TM domain and it includes the critical disulfide bond. BR does not reveal such a core, and the rigid clusters correspond to helical regions.

extracellular domain, the folding core. Two additional small clusters remain at the intracellular side. Neither of these clusters correlates with the positions of any helices, indicating that the loop regions in MR may be much more important to the folding of MR than to that of BR. This emphasizes the role of tertiary interactions at the early stages during folding, which has led to the proposal of a long-range interactions model for MP folding [16]. It is these long-range interactions that may contribute to the difficulty in experimentally refolding MR as opposed to BR.

6. Comparison of rhodopsin and bacteriorhodopsin folding in experiments

One major difference between BR and MR is the presence of a disulfide bond between the bottom of TM helix III and the second extracellular loop in MR (Fig. 2B), while there is no disulfide bond in BR and related structures. This disulfide bond is known to be critical for stability of MR and a wrong disulfide bond leads to misfolding of MR [51]. However, the presence or absence of this disulfide bond can be ruled out as a determinant for the differences between MR and BR structures in FIRST simulated unfolding, because a control calculation in which the disulfide bond was removed from the structure prior to application of FIRST [47] yielded the same result. Thus, while the disulfide bond once formed contributes to additional stability of the folded protein, it is important to distinguish a folding core from a stability core, that may form during later stages in the folding process. The folding core proposed here refers to a core of structural stability that resists denaturation and is assumed to form first in the reverse process. This hypothesis of course requires experimental validation.

While direct experimental proof for the presence of a folding core comprising loop and helix elements in MR is still missing, there is some preliminary evidence supporting the notion of a folding core. Using single-molecule dynamic force spectroscopy, a core of rigid structural segments was observed in MR but not BR [19]. In this study, breakpoints within helices were detected, suggesting that not all helices act as independent single folding units. Furthermore, segments containing loop and TM portions were observed, suggesting the involvement of loop residues in assisting secondary structure stabilization. There has also been a forced unfolding study of proteorhodopsin displaying characteristics similar to BR [52]. The second line of evidence is studies from our lab in collaboration with Wayne Hubbell (Arpana Dutta, Christian Altenbach, Sheryll Mangahas, Wayne L. Hubbell and Judith Klein-Seetharaman, unpublished results), in which we attached EPR spin labels in proximity to the folding core and found some restriction of the EPR labels in conformational space in the unfolded states, consistent with the idea of a folding core.

Indirect evidence supporting the observation that folding of MR is more complex than that of BR comes from extensive early studies on the role of loops in BR and MR folding through split receptor experiments (in detail reviewed in [53]). In these experiments, cuts are introduced in a loop connecting two helices and the resulting fragments are isolated and then reconstituted. In the case of BR, all possible

combinations of cuts have been made and no loop was found to be essential for the formation of a three-dimensional structure able to bind retinal. However, reconstituted fragments did display lower stability, and loop mutations, substitutions and deletions do cause changes in BR stability. Similar studies were also carried out for MR (see [53] for references). However, in contrast to BR, it is not possible to purify fragments of MR and reconstitute them later to yield a chromophore binding three-dimensional structure. Instead, split genes were expressed in cells and were characterized subsequent to *in vivo* reconstitution. In this setup, it is possible to introduce cuts, but only when placed in the loop connecting helices 4 and 5 and the loop connecting helices 5 and 6 in the cytoplasmic domain. No cut in extracellular loops allow formation of a folded structure, consistent with the predicted folding core located at the interface between the extracellular and TM domains.

7. Simulated unfolding of non-retinal binding proteins

Given the structural similarity of seven TM helices and retinal species covalently bound to the protein, it is remarkable to find the above described differences in simulated unfolding of BR versus MR. This poses the question how relevant these findings are for the folding of other MPs. To address this question, we applied FIRST to all helical MPs with known structure. The list of proteins was retrieved from the PDBTM [54]. Theoretical models and duplicate chains were removed and only proteins with resolution better than 3 Å were chosen. This resulted in 237 protein chains. These were clustered using a 35% sequence identity cutoff with BlastClust [55]. Clustering resulted in 52 protein chain groups. When MR and BR related proteins (i.e. MR, BR, sensory rhodopsins I & II, halorhodopsin) were also removed, 47 chain groups remained.

Visual inspection of the dilution plots indicated that some proteins behaved like BR, and some like MR, while some showed characteristics of both, with individual helices breaking off early but a core involving residues from other helices and loops remaining until late. Quantification of the folding core energies exhibited a spectrum of values (Fig. 3). MR and BR are at the two ends of the spectrum (indicated by arrows in Fig. 3). The dilution of long-range interactions [48] also showed a range of values (data not shown).

Finally, we developed a new way to quantify the dilution plots, shown in Fig. 4A. In order to be able to assess to what extent a helix unfolds independently of the other helices, we defined a metric, $\Delta H_i = H_u - H_s$. For a helix, H_s is defined as the energy level in the denaturation process where the helix (80% of its residues) becomes independent of other helices. H_u, on the other hand, marks the energy level where the helix unfolds completely. We define that a helix unfolds independently if it satisfies one of the two conditions: i) if $\Delta H_s >= -3$ kcal/mol and $\Delta H_i >= -1$ kcal/mol ii) it satisfies $\Delta Hs > -3$ kcal/mol and $\Delta Hi <= -2$ kcal/mol. The definitions were set such that the average number of helices that fold independently was high for BR-like structures and low for MR-like structures. Based on the above definition, 85% of helices in each of the resolved BR-like structures and only 17% of helices in each of the MR-like structures were estimated to unfold independently. Based on the folding behavior of the helices we labeled each of the structures as "BR-like", "MR-like" or "Mixed". Therefore, if at least 70% of the helices in a protein are unfolding independently, we classify the unfolding process as BR-like. If less than 30% of the helices in a protein were independently unfolding, we labeled it as MR-like, otherwise we labeled the structure as mixed. In this way, having labeled each structure within each protein chain group, the group then assigned an overall label based on the labels of the structures within the group. This resulted in five different groups of MP structures labeled as: MR-like, BR-like, mixed, mixed-BR and mixed-MR. MR-like, BR-like and Mixed means all structures in this group are MR-like, BR-like or Mixed, respectively. BR-mixed means the group has structures both BR and Mixed in it and MR-mixed means the group has structures both MR and Mixed in it. The numbers of protein chain groups within each of these five major groups is shown in Fig. 4B. Approximately half of the MP protein chain groups (25/47) are assigned BR-like, while only 4 are MR-like. The remainder (18/47) shows characteristics of both. A comparison between this classification and the folding core energy is shown in Fig. 4C. MP groups are ranked by folding core energy (analogous to Fig. 3) and their class label is shown (the color coding used corresponds to Fig. 4B). As one can see all but two groups with less negative values are classified as BR, while the mixed and MR-like groups also have lower folding core energies. This analysis indicates that most helical MPs contain helices as independent folding units, but frequently, there are some regions within the protein structure spanning multiple helices and including loop regions that show high cooperativity in unfolding. This suggests that the contribution of loops to areas of greatest stability is also a common theme among the folding of MPs.

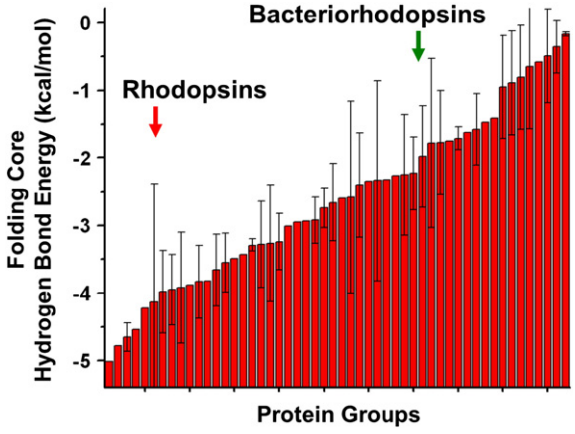

Fig. 3. FIRST predicted folding core energies of membrane proteins with known structure. Folding core energies were extracted from FIRST simulated unfolding plots as shown in Fig. 3. The averages and standard deviations for all members in the 52 groups of clustered MPs (see text) are plotted in order of decreasing folding core energy. The MR and BR groups are indicated by arrows.

As with all studies that take published crystal structures as input, our analysis may reflect crystallization bias. It is very well possible that BR-like behavior promotes ease of crystallization. While only a diversification of crystallized MP structures will help address this question, at this stage we can at least rule out resolution as contributing factor. When we ranked the different MP structure groups according to predicted folding core energy, there was no correlation with the associated resolution of the structures (Supplementary Fig. S1). We also investigated qualitatively if the groups may be correlated with membrane origin (bacterial gram negative inner membrane, bacterial gram positive plasma membrane, endoplasmic reticulum membrane, mitochondrial inner membrane, thylakoid membrane, archaebacterial membrane and eukaryotic plasma membrane (Supplementary Fig. S2). We found no correlation. Finally, we began investigating protein functional differences between protein groups (e.g. channels vs. transporters vs. enzymes) and found no correlation (data not shown). However, a rigorous hypothesis testing, e.g. using Gene Ontology functional terms, should be carried out to investigate this question further.

8. Extent of denaturation and refolding of non-retinal binding membrane proteins

Testing the predictions made by FIRST experimentally will require studies with denatured states of MPs. There are a number of proteins beyond the model systems BR and MR, for which conditions under which their denatured states can be studied have been established. Bacterial helical MPs for which refolding has been possible include: KcsA, from a trifluoroethanol (TFE)-denatured state [56]; DAGK from SDS [57], urea and GuHCl [58], DsbB from SDS denatured state [59], EmrE from a SDS + urea denatured state [60], CopA from a GuHCl

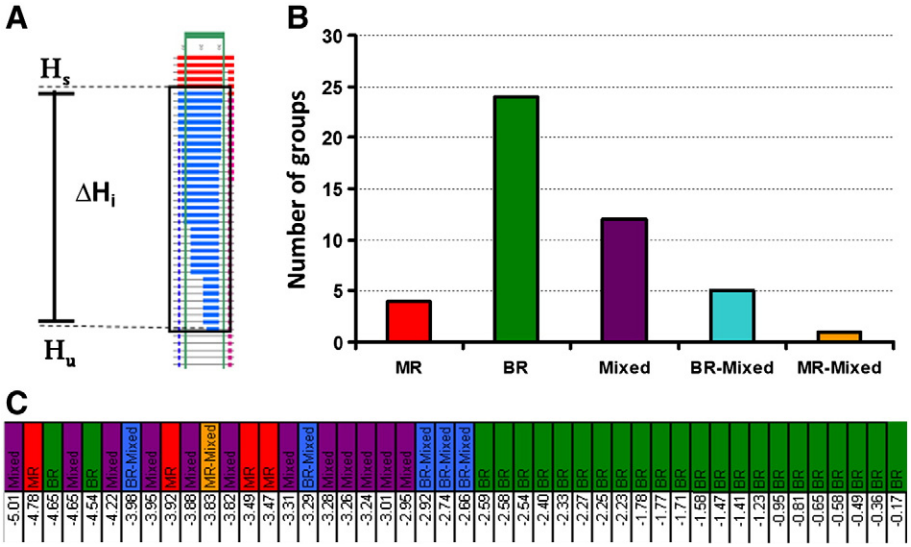

Fig. 4. Independent helix analysis of FIRST simulated unfolding results. A. Definition of parameters quantified from the simulated unfolding plot: H_s is the energy level of a helix before it starts to unfold; H_u is the energy level when a helix becomes independent of other helices; ΔH_i is the difference between H_u and H_s. In each case, a helix is defined by 80% of all residues in the helix. B. Helix-based classification to identify BR-like and MR-like behavior in non-retinal membrane proteins. We classified a helix as an "helix that unfolds independently of other helices" if it satisfies $\Delta H_s >= -3$ kcal/mol and $\Delta H_i >= -1$ kcal/mol or it satisfies $\Delta H_s < -3$ kcal/mol and $\Delta Hi >= -2$ kcal/mol. Then if at least 70% of the helices in a protein are unfolding independently, we classify the unfolding process of that protein as BR-like. If less than 30% of the helices in a protein were independently unfolding, we labeled it as MR-like, otherwise we labeled the structure as mixed. In each group, the majority vote was taken. When there was a tie, both labels were assigned. This gives rise to five groups, MR-like, BR-like, mixed, BR-mixed and MR-mixed. Each group is color coded. C. Relationship between folding core energies and helix-based classification. The 47 groups of MPs (which exclude all retinal protein structures, including MR, BR, sensory rhodopsin, halorhodopsin) were classified according to their predicted folding cores (Fig. 3) and color coded according to their classification from Fig. 4B. As one can see, the less negative folding core values are BR-like folders, while the MR and mixed groups have more negative folding core values.

denatured state [61] and LHCII complex in higher plants from a SDS-denatured state [62,63]. In all cases except CopA, refolding was accomplished only from a partially unfolded, near native state. GuHCl denaturation of CopA that led to a substantial decrease in MRE at 222 nm by about 75% could be refolded into the native state [61]. The only eukaryotic helical MP that has been purified in folded form, denatured and refolded so far is the Human Peripheral Myelin Protein 22 (PMP22), a four-helix bundle, whose misfolding has been implicated in Charcot-Marie-Tooth disease. PMP22 displays poor stability even in the absence of denaturants and folding studies were done both in the absence and presence of glycerol which acts as a stabilizing osmolyte. PMP22 was denatured in mixed micelles of dodecylphosphocholine and n-lauroyl sarcosine and renaturation was achieved by dilution of n-lauroyl sarcosine [64]. A number of other eukaryotic MPs have also been refolded from denatured states, although the details of the denatured states were not a focus of these studies: G protein coupled receptors have been expressed extensively in *E. coli* and these studies are becoming increasingly successful [65,66].

Refolding studies on LHCII, a helical MP, provides a good example of MP folding study where a structural insight into folding mechanisms of MP was obtained. EPR spectroscopy was used to investigate the details of refolding LHCII from a 50% lithium dodecyl sulfate denatured state [63]. In this report, distance measurements between two pairs of residues derivatized with EPR labels were carried out during equilibrium and kinetic unfolding and refolding studies. One pair was placed on either end of a TM helix and the other pair was placed in the luminal side of two different TM helices. Refolding kinetics showed that formation of tertiary contacts between the TM helices occurs after formation of TM helices. These findings are in accordance with the two-stage model of folding [2,3]. However, TM helix formation during LHCII refolding was observed to extend beyond the first stage of folding thereby not completely following the two-stage model.

9. Discussion and future studies

Analysis of protein folding is greatly facilitated when a protein can be reversibly unfolded, i.e. its native structure can be disrupted and subsequently restored. A microscopically reversible, dynamic equilibrium folding reaction allows a thermodynamic analysis of the data, although strictly this requires the equilibrium to be solvent independent, which is generally not the case. Thermodynamic studies help to link the findings of the refolding experiments and denatured states to the biological situation. There are few reports of successful recovery of fully folded, functional MPs from denatured states. In part this reflects the small number of research laboratories that are studying MP folding and endeavoring to establish defined, controlled refolding conditions. However, there are also inherent experimental difficulties compounded by a lack of knowledge of the nature of the denatured state as well as the key structural elements and interactions that are required for correct folding. α-Helical MPs pose considerable challenges, many of which arise from their inherent hydrophobicity and the problems involved in trying to re-create their natural, complex membrane environment. Moreover, some of these proteins are large with flexible structures, making the task of satisfying the requirements for folding of dynamic membrane-embedded, aqueous-exposed and lipid headgroup contacting regions even more difficult. There are four main problems that hinder the re-folding of α-helical MPs *in vitro*. Firstly, the precise nature of the denatured state is unknown. Hence, the residual structure, protein–denaturant interactions and ensemble of conformations are unknown as well as whether low-order aggregation states are present. This makes it difficult to identify existing unfavorable interactions that could hinder folding through formation of aggregates or incorrectly folded states. A second problem facing folding studies is the lack of information on key structural interactions that are required for correct folding. Thus, it is challenging to optimize the correct conditions for the folding of certain structural elements. It does appear that critical core and/or helical structure aids correct folding [36]. A third problem is the solvent. Frequently, detergents or mixtures of detergents and different lipids are used for practicality. These solvents are not accurate mimics of the native environment since a membrane is composed of many different lipids, with asymmetric monolayer composition and lateral heterogeneity together with the presence of other MPs. Moreover, there are differing solvent environments at either side of the membrane. The types of

environments used in experiments *in vitro* only partially reproduce the native characteristics. Furthermore, it is likely that different solvent properties are necessary to optimize different stages of folding [67]. A fourth issue for MP re-folding *in vitro*, is tweaking conditions to attain a fully functional state. While regaining the native secondary structure can be proved possible it is often challenging to completely recover the tertiary structure and native ligand binding or function. This final stage could require further optimization of solvent conditions. All four problems are compounded by the lack of experimental methods to probe protein and solvent structures and interactions during the folding process.

Future studies could aim at a molecular description of the denatured states. Since experimental conditions under which MR can be denatured without aggregation are now known [18], it will be possible to characterize unfolded states analogous to what has been done with soluble proteins [34]. This could directly test experimentally the predicted folding core. Future improvements in computational methods to study MP folding will need to include lipids. The FIRST method was applied without taking the membrane into account, which is clearly a limitation. For example, hydrogen bonds are ranked based on the structures as input only, without considering the presence of the membrane. While the strength of side chain hydrogen bonds in helical MPs appears to be similar to that of soluble proteins [68], backbone hydrogen bonds in membranes will be stronger, and this may affect the ranking of the bonds considered during the dilution step in FIRST. Other methods should also be explored. While molecular dynamics simulations have been applied extensively to lipid systems, none have tackled MP unfolded states. Recent developments in coarse grained modeling of lipids and membranes using the MARTINI force field [69] may provide alternative avenues for future modeling of MP folding in the presence of lipids.

Acknowledgements

We would like to thank Professor Charles Sanders of Vanderbilt University for asking his insightful question “Why can't rhodopsin be refolded?” at the Membrane Protein Folding Meeting of the Biophysical Society in Seoul, South Korea on May 19–22, 2013, inspiring the theme of this article. Thanks also to Fernanda Balem for the preparation of Fig. 2. Research by the authors reviewed in this article has been funded in part by NSF CAREER grant CC044917 and NSF EAGER grant IIS-1144281, National Institutes of Health Grant NLM108730, a DAAD-Helmholtz Fellowship and a Royal Society Wolfson Merit Award, Leverhulme Research Fellowship as well as BBSRC grants (BB/G002037/1 and BB/F013183/1) to PJB.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbabio.2013.11.021.

References

[1] K.S. Huang, et al., Refolding of an integral membrane protein. Denaturation, renaturation, and reconstitution of intact bacteriorhodopsin and two proteolytic fragments, J. Biol. Chem. 256 (8) (1981) 3802–3809.
[2] J.L. Popot, D.M. Engelman, Membrane protein folding and oligomerization: the two-stage model, Biochemistry 29 (17) (1990) 4031–4037.
[3] J.L. Popot, S.E. Gerchman, D.M. Engelman, Refolding of bacteriorhodopsin in lipid bilayers. A thermodynamically controlled two-stage process, J. Mol. Biol. 198 (4) (1987) 655–676.
[4] Retnet, http://www.sph.uth.tmc.edu/Retnet/disease.htm#03.202d.
[5] D.T. Hartong, E.L. Berson, T.P. Dryja, Retinitis pigmentosa, Lancet 368 (9549) (2006) 1795–1809.
[6] M.P. Krebs, et al., Molecular mechanisms of rhodopsin retinitis pigmentosa and the efficacy of pharmacological rescue, J. Mol. Biol. 395 (5) (2010) 1063–1078.
[7] C.H. Sung, C.M. Davenport, J. Nathans, Rhodopsin mutations responsible for autosomal dominant retinitis pigmentosa. Clustering of functional classes along the polypeptide chain, J. Biol. Chem. 268 (35) (1993) 26645–26649.
[8] C.H. Sung, et al., Rhodopsin mutations in autosomal dominant retinitis pigmentosa, Proc. Natl. Acad. Sci. U. S. A. 88 (15) (1991) 6481–6485.
[9] S. Kaushal, H.G. Khorana, Structure and function in rhodopsin. 7. Point mutations associated with autosomal dominant retinitis pigmentosa, Biochemistry 33 (20) (1994) 6121–6128.
[10] H.F. Mendes, et al., Mechanisms of cell death in rhodopsin retinitis pigmentosa: implications for therapy, Trends Mol. Med. 11 (4) (2005) 177–185.
[11] P.J. Booth, S. High, Polytopic membrane protein folding and assembly *in vitro* and *in vivo*, Mol. Membr. Biol. 21 (2004) 163–170.
[12] J.P. Chapple, et al., Unfolding retinal dystrophies: a role for molecular chaperones, Trends Mol. Med. 7 (2001) 414–421.
[13] C. McKibbin, et al., Opsin stability and folding: modulation by phospholipid bicelles, J. Mol. Biol. 374 (2007) 1319–1332.
[14] C. McKibbin, et al., Opsin stability and folding: the role of Cys 185 and abnormal disulfide bond formation in the intradiscal domain, J. Mol. Biol. 374 (2007) 1309–1318.
[15] R.S. Saliba, et al., The cellular fate of mutant rhodopsin: quality control, degradation and aggresome formation, J. Cell Sci. 115 (2002) 2907–2918.
[16] J. Klein-Seetharaman, Dual role of interactions between membranous and soluble portions of helical membrane receptors for folding and signaling, Trends Pharmacol. Sci. 26 (4) (2005) 183–189.
[17] A. Dutta, et al., Characterization of membrane protein non-native states. 2. The SDS-unfolded states of rhodopsin, Biochemistry 49 (30) (2010) 6329–6340.
[18] A. Dutta, et al., Characterization of membrane protein non-native states. 1. Extent of unfolding and aggregation of rhodopsin in the presence of chemical denaturants, Biochemistry 49 (30) (2010) 6317–6328.
[19] K.T. Sapra, et al., Mechanical properties of bovine rhodopsin and bacteriorhodopsin: possible roles in folding and function, Langmuir 24 (4) (2008) 1330–1337.
[20] A. Fersht, Enzyme Structure and Mechanism, 2nd ed. W.H. Freeman & Co., New York, 1985.
[21] P.J. Booth, P. Curnow, Folding scene investigation: membrane proteins, Curr. Opin. Cell Biol. 19 (2009) 8–13.
[22] P.J. Booth, et al., *In vitro* studies of membrane protein folding, Crit. Rev. Biochem. Mol. Biol. 36 (2001) 501–603.
[23] H. Hong, et al., Methods for measuring the thermodynamic stability of membrane proteins, Methods Enzymol. 455 (2009) 213–236.
[24] P.J. Booth, A successful change of circumstance: a transition state for membrane protein folding, Curr. Opin. Struct. Biol. 22 (2012) 469–475.
[25] S. Faham, et al., Side-chain contributions to membrane protein structure and stability, J. Mol. Biol. 335 (2004) 297–305.
[26] P. Curnow, P.J. Booth, Combined kinetic and thermodynamic analysis of alpha-helical membrane protein unfolding, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 18970–18975.
[27] P. Curnow, P.J. Booth, The transition state for integral membrane protein folding, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 773–778.
[28] P. Curnow, et al., Stable folding core in the folding transition state of an alpha-helical integral membrane protein, Proc. Natl. Acad. Sci. U. S. A. 108 (34) (2011) 14133–14138.
[29] N.H. Joh, et al., Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins, Nature 453 (7199) (2008) 1266–1270.
[30] R. Webel, et al., Role of asparagine-linked oligosaccharides in rhodopsin maturation and association with its molecular chaperone, NinaA, J. Biol. Chem. 275 (32) (2000) 24752–24759.
[31] N. Ismail, et al., Specific transmembrane segments are selectively delayed at the ER translocon during opsin biogenesis, Biochem. J. 411 (3) (2008) 495–506.
[32] S.L. Meacock, et al., Different transmembrane domains associate with distinct endoplasmic reticulum components during membrane integration of a polytopic protein, Mol. Biol. Cell 13 (12) (2002) 4114–4129.
[33] D. Athanasiou, M. Aguila, D. Bevilacqua, S.S. Novoselov, D.A. Parfitt, M.E. Cheetham, The cell stress machinery and retinal degeneration, FEBS Lett. 587 (2013) 2008–2017.
[34] J. Klein-Seetharaman, et al., Long-range interactions within a nonnative protein, Science 295 (5560) (2002) 1719–1722.
[35] R.L. Baldwin, B.H. Zimm, Are denatured proteins ever random coils? Proc. Natl. Acad. Sci. U. S. A. 97 (23) (2000) 12391–12392.
[36] P.J. Booth, P. Curnow, Membrane proteins shape up: understanding *in vitro* folding, Curr. Opin. Struct. Biol. 16 (4) (2006) 480–488.
[37] P.J. Booth, A. Farooq, Intermediates in the assembly of bacteriorhodopsin investigated by time-resolved absorption spectroscopy, Eur. J. Biochem. 246 (3) (1997) 674–680.
[38] P.J. Booth, A. Farooq, S.L. Flitsch, Retinal binding during folding and assembly of the membrane protein bacteriorhodopsin, Biochemistry 35 (18) (1996) 5902–5909.
[39] P.J. Booth, et al., Intermediates in the folding of the membrane protein bacteriorhodopsin, Nat. Struct. Biol. 2 (2) (1995) 139–143.
[40] M.L. Riley, et al., Slow alpha helix formation during folding of a membrane protein, Biochemistry 36 (1) (1997) 192–196.
[41] S.J. Allen, et al., Folding kinetics of an alpha helical membrane protein in phospholipid bilayer vesicles, J. Mol. Biol. 342 (4) (2004) 1279–1291.
[42] F. Balem, N. Yanamala, J. Klein-Seetharaman, Additive effects of chlorin e6 and metal ion binding on the thermal stability of rhodopsin *in vitro*, Photochem. Photobiol. 85 (2) (2009) 471–478.
[43] R. Hubbard, Absorption spectrum of rhodopsin: 500 nm absorption band, Nature 221 (5179) (1969) 432–435.
[44] H. Shichi, Conformational aspects of phodopsin associated with disc membranes, Exp. Eye Res. 17 (6) (1973) 533–543.
[45] E.W. Abrahamson, S.E. Ostroy, The photochemical and macromolecular aspects of vision, Prog. Biophys. Mol. Biol. 17 (1967) 179–215.
[46] C. McKibbin, et al., Urea unfolding of opsin in phospholipid bicelles, Photochem. Photobiol. 85 (2) (2009) 494–500.
[47] A.J. Rader, et al., Identification of core amino acids stabilizing rhodopsin, Proc. Natl. Acad. Sci. U. S. A. 101 (19) (2004) 7246–7251.

[48] O. Tastan, et al., Comparison of stability predictions and simulated unfolding of rhodopsin structures, Photochem. Photobiol. 83 (2) (2007) 351–362.
[49] A.J. Rader, I. Bahar, Folding core predictions from network models of proteins, Polymer 45 (2004) 659–668.
[50] D.J. Jacobs, et al., Protein flexibility predictions using graph theory, Proteins 44 (2) (2001) 150–165.
[51] J. Hwa, J. Klein-Seetharaman, H.G. Khorana, Structure and function in rhodopsin: mass spectrometric identification of the abnormal intradiscal disulfide bond in misfolded retinitis pigmentosa mutants, Proc. Natl. Acad. Sci. U. S. A. 98 (9) (2001) 4872–4876.
[52] Klyszejko, et al., Folding and assembly of proteorhodopsin, J. Mol. Biol. 376 (2008) 35–41.
[53] O. Tastan, J. Klein-Seetharaman, H. Meirovitch, The effect of loops on the structural organization of alpha-helical membrane proteins, Biophys. J. 96 (6) (2009) 2299–2312.
[54] G.E. Tusnady, Z. Dosztanyi, I. Simon, PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank, Nucleic Acids Res. 33 (2005) D275–D278(Database issue).
[55] S.F. Altschul, et al., Basic local alignment search tool, J. Mol. Biol. 215 (3) (1990) 403–410.
[56] F.N. Barrera, et al., Unfolding and refolding *in vitro* of a tetrameric, alpha-helical membrane protein: the prokaryotic potassium channel KcsA, Biochemistry 44 (43) (2005) 14344–14352.
[57] F.W. Lau, J.U. Bowie, A method for assessing the stability of a membrane protein, Biochemistry 36 (19) (1997) 5884–5892.
[58] J.K. Nagy, W.L. Lonzer, C.R. Sanders, Kinetic study of folding and misfolding of diacylglycerol kinase in model membranes, Biochemistry 40 (30) (2001) 8971–8980.
[59] D.E. Otzen, Folding of DsbB in mixed micelles: a kinetic analysis of the stability of a bacterial membrane protein, J. Mol. Biol. 330 (4) (2003) 641–649.
[60] D. Miller, et al., *In vitro* unfolding and refolding of the small multidrug transporter EmrE, J. Mol. Biol. 393 (4) (2009) 815–832.
[61] E.A. Roman, J.M. Arguello, F.L. Gonzalez Flecha, Reversible unfolding of a thermophilic membrane protein in phospholipid/detergent mixed micelles, J. Mol. Biol. 397 (2) (2010) 550–559.
[62] F.G. Plumley, G.W. Schmidt, Reconstitution of chlorophyll a/b light-harvesting complexes: xanthophyll-dependent assembly and energy transfer, Proc. Natl. Acad. Sci. U. S. A. 84 (1) (1987) 146–150.
[63] C. Dockter, et al., Refolding of the integral membrane protein light-harvesting complex II monitored by pulse EPR, Proc. Natl. Acad. Sci. U. S. A. 106 (44) (2009) 18485–18490.
[64] J.P. Schlebach, et al., Reversible folding of human peripheral myelin protein 22, a tetraspan membrane protein, Biochemistry 52 (19) (2013) 3229–3241.
[65] T. Dahmane, et al., Amphipol-assisted *in vitro* folding of G protein-coupled receptors, Biochemistry 48 (27) (2009) 6516–6521.
[66] K. Michalke, et al., Mammalian G protein-coupled receptor expression in *Escherichia coli*: II. Refolding and biophysical characterization of mouse cannabinoid receptor 1 and human parathyroid hormone receptor 1, Anal. Biochem. 401 (1) (2010) 74–80.
[67] P.J. Booth, Sane in the membrane: designing systems to modulate membrane proteins, Curr. Opin. Struct. Biol. 15 (2005) 435–440.
[68] J.U. Bowie, Membrane protein folding: how important are hydrogen bonds? Curr. Opin. Struct. Biol. 21 (1) (2011) 42–49.
[69] S.J. Marrink, et al., The MARTINI force field: coarse grained model for biomolecular simulations, J. Phys. Chem. B 111 (27) (2007) 7812–7824.