



## Review

# Mechanistic studies of the biogenesis and folding of outer membrane proteins *in vitro* and *in vivo*: What have we learned to date?



Lindsay M. McMorran, David J. Brockwell, Sheena E. Radford\*

Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK  
 School of Molecular and Cellular Biology, University of Leeds, Leeds LS2 9JT, UK

## ARTICLE INFO

## Article history:

Received 10 January 2014  
 and in revised form 16 February 2014  
 Available online 5 March 2014

## Keywords:

Protein folding  
 Outer membrane protein  
 Periplasmic chaperone  
 BAM complex  
 $\Phi$ -Value analysis  
 Protein stability

## ABSTRACT

Research into the mechanisms by which proteins fold into their native structures has been on-going since the work of Anfinsen in the 1960s. Since that time, the folding mechanisms of small, water-soluble proteins have been well characterised. By contrast, progress in understanding the biogenesis and folding mechanisms of integral membrane proteins has lagged significantly because of the need to create a membrane mimetic environment for folding studies *in vitro* and the difficulties in finding suitable conditions in which reversible folding can be achieved. Improved knowledge of the factors that promote membrane protein folding and disfavour aggregation now allows studies of folding into lipid bilayers *in vitro* to be performed. Consequently, mechanistic details and structural information about membrane protein folding are now emerging at an ever increasing pace. Using the panoply of methods developed for studies of the folding of water-soluble proteins. This review summarises current knowledge of the mechanisms of outer membrane protein biogenesis and folding into lipid bilayers *in vivo* and *in vitro* and discusses the experimental techniques utilised to gain this information. The emerging knowledge is beginning to allow comparisons to be made between the folding of membrane proteins with current understanding of the mechanisms of folding of water-soluble proteins.

© 2014 Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

## Principles of protein folding

### How water-soluble proteins fold

The biological activity of many proteins is reliant on their ability to adopt a specific, three-dimensional structure. Failure to achieve this structure can have serious consequences, as evidenced by the prevalence of diseases caused by protein misfolding and aggregation [1,2]. The information required for a polypeptide chain to attain its native structure was shown to be contained within its primary sequence by Anfinsen's experiments on ribonuclease A [3,4]. This small, globular protein was completely unfolded in chemical denaturants and shown to regain its native conformation spontaneously upon dialysis [3]. This seemingly simple conclusion has led to decades of scientific research to determine how the amino acid sequence of a protein confers its native structure.

If folding were to occur by random sampling of all possible conformations of the polypeptide chain, finding the native state would take an astronomically long time. Proteins, however, fold

on biologically relevant timescales [5]. Levinthal suggested that this apparent paradox could be resolved if proteins fold *via* defined pathways [6]. Following this conclusion, several mechanisms were proposed to describe the pathways traversed by a protein *en route* to the native state. Analysis of the refolding kinetics of ribonuclease A revealed two distinct phases [7] and led to the suggestion of a “framework” mechanism of folding whereby secondary structural elements of proteins are formed prior to their docking to form the tertiary structure [8–10]. Further investigation of the refolding of ribonuclease A revealed that one of the phases observed was not, in fact, due to the presence of an observable folding intermediate, but arose as a consequence of proline *cis-trans* isomerisation [11]. This realisation, alongside the characterisation of the folding of chymotrypsin inhibitor 2 which revealed a simple two-state folding mechanism [12–14], saw the framework mechanism become disfavoured due to its implication that folding intermediates should be present [10,15]. To explain folding in the absence of detectable intermediates, the nucleation-condensation mechanism was proposed by Fersht and co-workers [12]. This mechanism involves the formation of a small nucleus of structure stabilised by weak, possibly long-range, contacts and the subsequent rapid collapse around this folding nucleus to yield the native state [10,12,15]. A combination of simulation and experimental data

\* Corresponding author at: Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK.

E-mail address: [s.e.radford@leeds.ac.uk](mailto:s.e.radford@leeds.ac.uk) (S.E. Radford).

on members of the homeodomain-like super-family of proteins revealed that the seemingly contrasting framework and nucleation-condensation folding models could be considered as two extremes of a single mechanism. In this unified model, the relative stability of the secondary and tertiary structure determines whether these elements are formed in series or in parallel [16] and, thus, two decades of conflicting evidence were resolved.

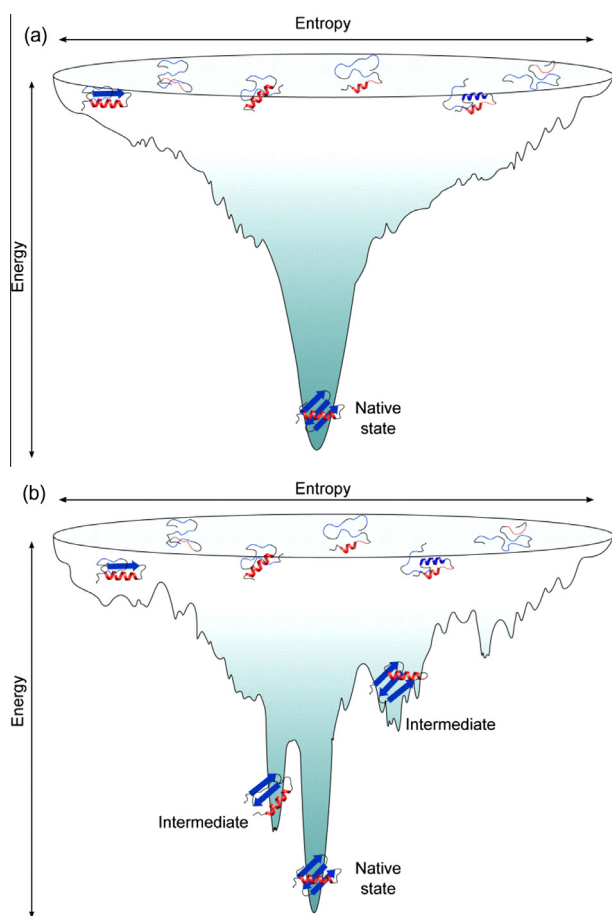
More recently it has been realised that proteins fold *via* a collection of parallel pathways which make up a funnel-shaped energy landscape (Fig. 1) [17–20]. At the top of the funnel, the unfolded state represents a large ensemble of high-entropy conformations of the polypeptide chain. While unstructured, the polypeptide chain may be biased by weak, residual interactions which initiate folding [21]. Indeed, an unfolded variant of the bacterial immunity protein Im7 has been studied recently under non-denaturing conditions, revealing conformational restriction in the regions of the protein sequence which ultimately form the native helices, emphasising the importance of such interactions in the initiation of folding [22]. Similar conclusions have been drawn from other proteins and protein fragments under different denaturing conditions [23–28]. As folding progresses, the polypeptide chain undergoes many reorganisations aiding the formation of stabilising interactions between side-chains, the protein backbone and the solvent as the protein approaches the native state [17]. The landscape view is an attractive one as it does not place restrictions

on whether secondary structure must form before, or at the same time, as the tertiary structure. Additionally, the funnel-shaped landscape predicts the experimentally observed robustness of the folding process to destabilising mutations: if the final fold remains the most stable state relative to the unfolded ensemble, a mutation may block some of the pathways to the native state but alternative folding pathways can be utilised [17].

The landscape theory of protein folding also predicts the existence of “downhill folding”, that is, folding in the absence of significant energy barriers [17]. Even with the most powerful single molecule techniques available it is still difficult to experimentally identify barrier-less folding unequivocally [29]. Exciting developments in simulation methods have recently enabled the first visions of protein folding in all-atom detail for small, single domain proteins [30]. These simulations have suggested that at least 9 of the 12 rapidly-folding proteins studied experience energy barriers that slow folding [30]. For proteins in which only the native and unfolded states are stably populated, known as a two-state folding mechanism, the folding energy landscape is relatively smooth (Fig. 1a) [31]. With the advent of more rapid triggering methods (ultra-rapid mixing, temperature jump, *etc.* [31]) and more sensitive detection methods (e.g. single molecule techniques [32–34]), such a simple folding scenario is rarely observed, even for small proteins. Instead, the folding landscape is often rugged with local energy minima representing the population of one or more folding intermediates *en route* to the native state (Fig. 1b) [31,35].

Rugged folding landscapes result as a consequence of the need to form the thousands of weak, stabilising interactions which are present in the native state (reviewed in Ref. [31]). During the structural reorganisations that are required for the protein to traverse the folding landscape, it is not always possible to minimise the energy of all of these interactions simultaneously, leading to “frustration” in the landscape [19,36]. Theory predicts that a rough landscape will lead to slower folding by limiting the rate at which the protein can find the native state [19]. Recently the effect of solvent viscosity on the folding and unfolding kinetics of homologous domains of  $\alpha$ -spectrin was studied, revealing the first explicit experimental evidence that differences in internal friction, and hence frustration, can indeed influence the folding kinetics of structurally homologous proteins [37]. It has been suggested previously that frustration, and hence ruggedness of energy landscapes, arises due to conflicting pressures to evolve protein sequences which can reliably fold to a stable native structure and that are capable of carrying out a defined biological role [19,36]. While this may be the case for some proteins, the differences in unfolding rates of structurally and functionally similar  $\alpha$ -spectrin domains were shown to be a consequence of different degrees of landscape ruggedness, with slower unfolding occurring as a result of increased frustration [37]. This result suggests that frustration has been selected to slow unfolding in this protein domain, which has an unusually long half-life *in vivo* [37]. Folding of the Trp-cage mini protein was also shown to be slowed when its folding intermediate was destabilised by the presence of chemical denaturant or a helix-breaking mutation [38]. This suggests that local energetic minima in the folding landscape can act to separate the conformational search into multiple, smaller problems and hence accelerate the conformational search process (reviewed in Ref. [35]). Whether folding intermediates act to disrupt or promote the folding process for many proteins is still under debate. The utility of intermediate species in providing insights into the structural regulation of biological function as well as their role in the initiation of protein aggregation and recognition by molecular chaperones, makes them important targets of study [39].

This brief overview of the folding of small water-soluble proteins has outlined how modern advances in experimental and computational techniques have provided significant progress



**Fig. 1.** Schematic representation of funnel-shaped folding landscapes. Example of (a) a smooth folding landscape expected for a two-state folding mechanism where only the native and unfolded states are stably populated and (b) a rugged landscape wherein the polypeptide chain populates one or more intermediate structures which represent local energy minima. Reprinted from [31] by permission from Macmillan Publishers Ltd: Nat. Struct. Mol. Biol. © 2009.

towards understanding the folding and assembly mechanisms of this class of proteins. Many fundamental questions remain unanswered about this complex and important biological process, particularly regarding the folding of more complicated systems such as multi-domain proteins, folding in the cellular milieu and membrane protein folding [40–44]. Current research into understanding the latter problem, focussing on bacterial outer membrane proteins (OMPs), is covered in the following sections.

### The membrane protein folding problem

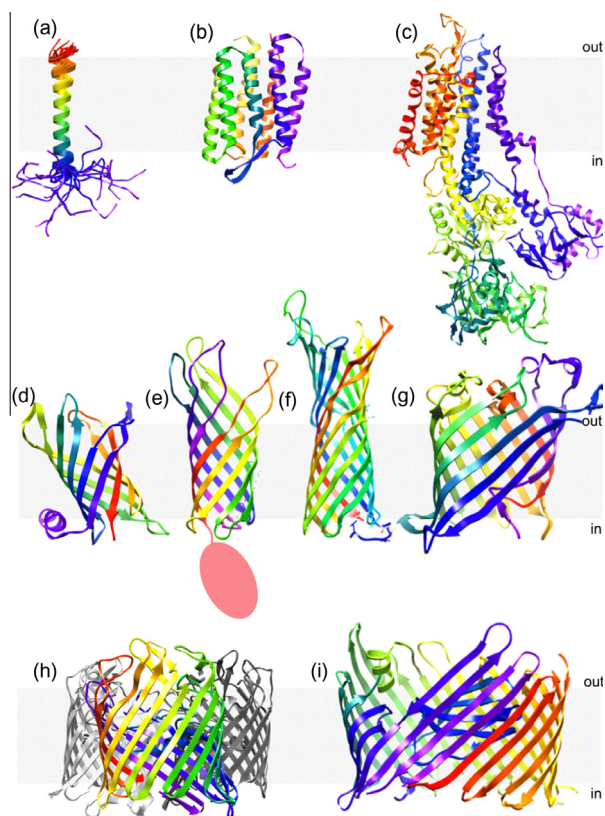
Understanding the folding, stability and function of membrane proteins is an important area of research as these proteins represent 60% of current drug targets and have vital roles in the cell, including signalling, transport and biogenesis [45,46]. In contrast with the wealth of information available about the folding of small, water-soluble proteins, the field of membrane protein folding has

lagged significantly [47]. For water-soluble proteins, folding is driven by the need to bury hydrophobic side-chains in order to prevent aggregation and to facilitate the formation of stable structures [40]. In addition to the attainment of the native state, the folding of integral membrane proteins is complicated by the need to insert the polypeptide chain into a lipid membrane [47]. Following membrane insertion, most of the surface of an integral membrane protein is in contact with the membrane's hydrophobic interior. Hydrophilic residues are either restricted to regions which contact the polar head groups of the membrane lipids or are exposed to the aqueous environment on either side of the membrane [47]. Additionally, the membrane environment *in vivo* is highly dynamic and heterogeneous with regions of varying lipid composition [48]. Recreating this environment *in vitro* has proved more difficult than the simple aqueous environment needed to fold water-soluble proteins and this, in part, has limited studies on the folding of integral membrane proteins [47,48].

### Classes of membrane proteins

The proteins present in biological membranes can be categorised into two families: the lipid-anchored proteins, which have a covalently-bound fatty-acid moiety through which a water-soluble protein is attached to a membrane, and the integral membrane proteins, which contain membrane-spanning regions. Only the folding mechanism of the latter will be described here. In contrast with lipid-associated proteins, the integral membrane proteins are constrained by the need to compensate for the energetic cost of burying peptide bonds in the lipid bilayer [49], estimated to be 1.2 kcal/mol per peptide bond [50]. As a consequence, it was predicted that membrane spanning regions would form regular secondary structural elements in order to maximise the hydrogen bonding potential of the peptide backbone [49]. Formation of secondary structure reduces the energetic cost of incorporation of peptide bonds into a bilayer by  $\approx 0.4$  kcal/mol per peptide bond for  $\alpha$ -helical structure and  $\approx 0.5$  kcal/mol per peptide bond for  $\beta$ -sheet structure [50]. The first  $\alpha$ -helical membrane protein structure was solved in 1975 by Henderson and Unwin using electron microscopy to generate a three-dimensional image of the purple membrane of *Halobacterium salinarum* [51]. The resulting 7 Å resolution image revealed the structure of bacteriorhodopsin (bR)<sup>1</sup> to be a seven helical, transmembrane bundle [51]. The structural information available about bR has since been increased by structures at higher resolution obtained using electron microscopy (3 Å, [52]) and X-ray diffraction (1.43 Å, [53]). Since the structural elucidation of bR, a wide variety of helical transmembrane structures have been solved and deposited in the Protein Data Bank (PDB) [54]. These show a diverse range of size and function across the kingdoms of life. Some examples are depicted in Fig. 2a–c.

In contrast with the ubiquitous distribution of  $\alpha$ -helical transmembrane proteins, membrane proteins containing  $\beta$ -sheet secondary structure are found only in the outer membranes of Gram-negative bacteria, mitochondria and chloroplasts [55,56].



**Fig. 2.** Example structures of integral membrane proteins. Structures of (a) the transmembrane segment of a glycoporphin A monomer from human erythrocyte membranes solved by NMR spectroscopy (1AFO [228]); (b) bacteriorhodopsin, a seven-helical bundle from the purple membrane of *Halobacterium salinarum* (1C3W [229]); (c) calcium ATPase 1 from the sarcoplasmic reticulum membrane of *Oryctolagus cuniculus*, a ten-helical bundle with a large cytoplasmic domain (1IWO [230]); (d) PagP, an 8-stranded palmitoyl transferase enzyme from *E. coli* (1THQ [206]); (e) the 8-stranded transmembrane domain of OmpA, an ion channel from *E. coli* (1BXW [231]), with the C-terminal periplasmic domain (structure currently not determined) represented by a red circle; (f) the 10-stranded OM protease, OmpT, from *E. coli* (1I78 [232]); (g) the 12-stranded, colicin-secreting phospholipase A, OmpLA, from *E. coli* (1QD5 [233]); (h) the OmpF porin, a trimer comprised of three 16-stranded  $\beta$ -barrels, from *E. coli* (2ZFG [234]) and (i) the 24-stranded translocation domain of PapC from *E. coli* (3FIP [235]). Unless otherwise specified, all structures were solved using X-ray crystallography. Proteins are coloured rainbow: violet (N-terminus) to red (C-terminus). In (h), a single OmpF monomer is coloured, while the remaining monomers are shown in greyscale. The approximate position of the membrane is indicated in all images with grey shading. All images were generated from the PDB files using the accession numbers given in brackets using UCSF Chimera molecular visualisation application [236].

<sup>1</sup> Abbreviations used: bR, bacteriorhodopsin; PDB, protein data bank; hVDAC, human mitochondrial voltage-dependent anion-selective channel; NMR, nuclear magnetic resonance; IM, inner membrane; OM, outer membrane; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; LPS, lipopolysaccharide; OMP, outer membrane protein; PPIase, parvulin-like peptidyl-prolyl isomerase; BAM,  $\beta$ -barrel assembly machinery; MD, molecular dynamics; POTRA, polypeptide transport-associated; CD, circular dichroism; SUV, small unilamellar vesicle; LUV, large unilamellar vesicle; C<sub>x</sub>yPZ, a phospholipid where x indicates the number of carbon atoms in the acyl chain, y indicates the number of double bonds per acyl chain and Z represents the head-group moiety; PC, phosphatidylcholine; LDAO, N-lauryl-N,N-dimethylammonium-N-oxide; HT PagP, C-terminally His-tagged PagP; SDS-PAGE, sodium dodecyl sulphate polyacrylamide gel electrophoresis; GuHCl, guanidine hydrochloride.

In order to satisfy all of the possible hydrogen bonds in the peptide backbone, each  $\beta$ -strand associates laterally with its neighbours resulting in an overall cylindrical topology, known as a  $\beta$ -barrel [55]. With predominantly non-polar side chains exposed to the hydrophobic membrane interior and each of the backbone hydrogen bonds satisfied, the resulting  $\beta$ -barrel structures have high thermodynamic stability [50,57]. The size of the  $\beta$ -barrel is highly variable, with known structures containing between 8 and 24  $\beta$ -strands, and the proteins may contain periplasmic or extracellular domains [55,56]. The regions between  $\beta$ -strands often alternate between tight turns on the periplasmic side of the membrane and longer, more flexible loops on the outer surface, which are exposed to the external environment (Fig. 2d–i) [55]. One almost entirely conserved structural feature across this family of protein structures is an even number of  $\beta$ -strands, with only one known exception to this rule discovered thus far: the 19-stranded human mitochondrial voltage-dependent anion-selective channel (hVDAC) [56,58].

As more and more research is focused on the determination of membrane protein structure, a total of over 430 unique structures have now been reported in the membrane protein structure (mpstruc) database with both  $\alpha$ -helical and  $\beta$ -barrel proteins represented [59]. Structure determination has been most successful using X-ray diffraction, although 105 of the reported structures were solved using nuclear magnetic resonance (NMR) spectroscopy and a further 14 structures have been solved using electron diffraction, demonstrating the utility of multiple techniques in the membrane protein structure determination tool box [60,61]. Almost 30 years after the first structure was reported, integral membrane proteins still account for only 1% of the protein structures in the PDB [62]. Significant progress is being made in this area, however, with 64 structures determined in 2012 and 50 in 2013 to date [59].

## Biological membranes

Lipids in cells have three main functions: energy storage, signal transduction and forming the matrix of biological membranes – the approximately 30 Å thick layer which encloses the cell and organelles within eukaryotic cells [63]. Cellular membranes are composed of polar lipids, which self-associate into bilayers to shield the hydrophobic regions from the aqueous environment in a process driven entropically by water molecules [63]. The structure and composition of lipid bilayers varies greatly even within an organism, allowing the properties of different membranes to be tailored to a specific function. Modulation of lipid composition to adapt to different functional requirements implies the evolutionary advantage of an extensive and complex lipid repertoire [63]. In addition to varying lipid composition, membranes can have either a symmetrical or asymmetrical distribution of lipids between the two leaflets of the bilayer. Within a leaflet, favourable interactions between some of the lipid components can generate domains of specific lipid composition, known as lipid rafts, which are thought to be involved in localising membrane proteins [63]. The variable properties of biological membranes demonstrate the highly dynamic, heterogeneous and complex nature of the lipid environment in which integral membrane proteins must insert, fold and function.

## The cell envelope of Gram-negative bacteria

A hallmark of Gram-negative bacteria, such as *Escherichia coli*, is the cell envelope, which is composed of two lipid membranes enclosing an aqueous compartment called the periplasm (Fig. 3a) [64,65]. The inner membrane (IM) is a symmetric phospholipid

bilayer composed of approximately 70% phosphatidylethanolamine (PE, Fig. 3d), 25% phosphatidylglycerol (PG, Fig. 3e) and 5% or less cardiolipin (Fig. 3f) and forms the barrier between the cytoplasm and the periplasm (Fig. 3a) [64–66]. There are two types of protein associated with the IM; lipoproteins which undergo lipid modifications of an N-terminal cysteine residue to anchor them to the periplasmic face of the IM and the  $\alpha$ -helical integral membrane proteins (reviewed in Ref. [65]). IM proteins are responsible for many cellular processes, including lipid synthesis and small molecule transport. This class of proteins is discussed at length elsewhere in this issue.

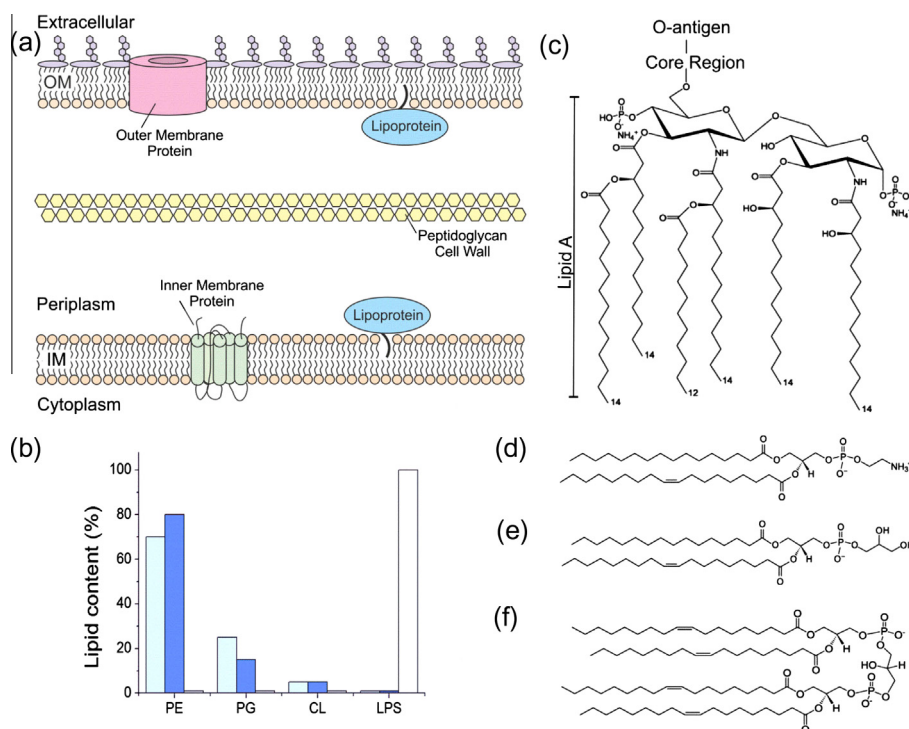
The periplasm is the compartment between the IM and the outer membrane (OM), which comprises around 10% of the total cell volume and contains soluble proteins, as well as a peptidoglycan cell wall. The peptidoglycan layer plays important roles in maintaining the shape of the cell and preventing lysis, while periplasmic proteins are involved in maintaining the integrity of the cell envelope. The processes which take place in the periplasm are independent of nucleotide hydrolysis, since no ATP is present in this compartment [65,67]. When energy is required, the cell relies on complex coupling mechanisms which often use the proton-motive force across the inner membrane as the primary energy source [65].

The inner leaflet of the asymmetric OM is comprised of phospholipids and is similar in composition to the IM, although the PE content is enriched compared with the IM (Fig. 3b) [65,66]. The outer leaflet of the OM consists of lipopolysaccharide (LPS), a glycolipid typically consisting of lipid A, a core oligosaccharide and an O-antigen (Fig. 3c) [64,68]. The structure of the O-antigen is highly variable, even within a species, with approximately 170 variants being recorded in *E. coli* [69]. The high number of fatty acid chains on LPS compared with phospholipids, and the fact that these chains are saturated, facilitates tight packing of the LPS molecules in the OM [68]. This confers low fluidity to the membrane and it is this property that is responsible for the low permeability of the OM [68]. As a consequence, rapid diffusion of small, hydrophobic molecules across the OM is prevented and Gram-negative bacteria thus tend to be less susceptible to hydrophobic antibiotics than their Gram-positive counterparts [68]. As with the IM, integral membrane proteins and lipoproteins are associated with the OM, but in this membrane the integral outer membrane proteins (OMPs) usually have  $\beta$ -barrel structures [55,65]. The barrel is formed by membrane-spanning  $\beta$ -strands which are held together by a lateral hydrogen bond network [55]. One reported exception, the polysaccharide translocon Wza of *E. coli*, has a barrel structure composed of laterally associated  $\alpha$ -helices [70]. Several examples of OMP structures are shown in Fig. 2d–i.

## OMP biogenesis *in vivo*

### *Periplasmic chaperones assisting OMP biogenesis*

Following their synthesis in the cytosol, OMPs are targeted to the SecYEG translocon by the SecB chaperone, whereupon they are translocated across the IM through SecYEG in an unfolded state [71,72]. The unfolded OMPs must be protected from aggregation and must be able to traverse the periplasm, including the peptidoglycan layer, and then correctly fold and insert into the OM [73]. These requirements suggest that transport across the periplasm and membrane insertion may be facilitated processes and, indeed, a number of periplasmic and OM-associated proteins have been implicated in the OMP assembly pathway [73]. These proteins can be roughly grouped into three categories: proteases; chaperones which stabilise unfolded and non-native conformations of their client proteins; and folding catalysts, which catalyse rate-limiting steps in folding (Fig. 4a and b) [73].



**Fig. 3.** The cell envelope of Gram-negative bacteria. (a) The cytoplasm of *E. coli* is surrounded by the inner membrane (IM), the periplasm and the outer membrane (OM). The IM is a symmetric phospholipid (shown in orange) bilayer containing integral  $\alpha$ -helical membrane proteins. The OM is an asymmetric bilayer of phospholipid and lipopolysaccharide (LPS, shown in purple) and contains  $\beta$ -barrel integral membrane proteins. The periplasm is the aqueous compartment between the two membranes in which the peptidoglycan cell wall is found. Both membranes have associated lipoproteins on their periplasmic faces. (b) The lipid composition of the IM (light blue), inner leaflet of the OM (dark blue) and outer leaflet of the OM (white) in *E. coli* (percentages based on those reported in [64–66]). Structures of (c) LPS, (d) phosphatidylethanolamine, (e) phosphatidylglycerol and (f) cardiolipin are shown.

Given the range of essential functions carried out by OMPs [56], it is unsurprising that the presence of unfolded OMPs in the periplasm is a strong inducer of the envelope stress response [73]. Exposed C-terminal residues of misfolded OMPs are recognised by the PDZ domain of the inner membrane-associated protease, DegS (Fig. 4a), activating the protease domain [74]. Activated DegS triggers a proteolytic cascade which leads to induction of the  $\sigma^E$  heat-shock response [73,74]. Using the assumption that depletion of periplasmic chaperones will cause an increase in misfolded OMPs, identification of many of the known periplasmic folding factors arise by genetic studies of bacterial strains showing high  $\sigma^E$  activity [67]. Some of the key periplasmic assistants of OMP folding, including SurA, Skp, DegP and FkpA (Fig. 4), are described in the following sections.

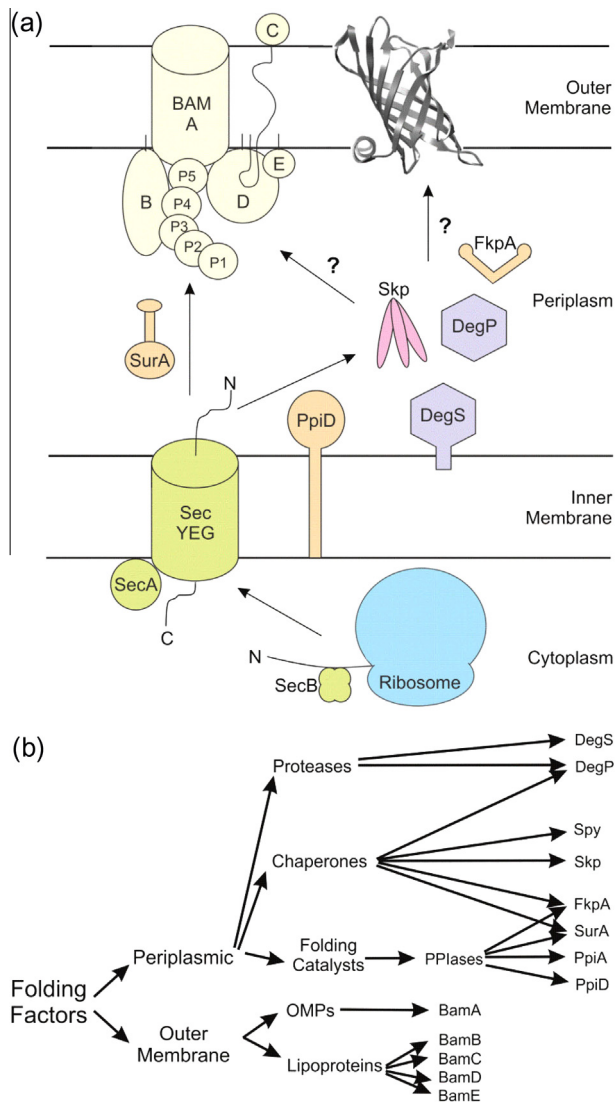
#### SurA

SurA was first identified in 1990 when it was shown to be required for the survival of *E. coli* in the stationary phase [75]. Initial characterisation described SurA as a parvulin-like peptidyl-prolyl isomerase (PPIase) involved in the proper assembly of major OMPs [67,76,77]. Trypsin digestion of OMPs in *surA* deletion mutants showed that not all OMPs have increased rates of proteolysis, leading to the conclusion that SurA is not an essential folding factor [76]. The amounts of FadL, LamB, OmpA, OmpC, OmpF, OmpX and LptD, however, were all found to be reduced in *surA* deletion strains [78]. Similar results were obtained using a proteomic analysis, which also revealed an upregulation of proteins under the control of the  $\sigma^E$  regulon in *surA* deletion strains [79].

Crystallisation of SurA revealed a four-domain protein with two PPIase domains (P1 and P2) sandwiched between the N- and C-terminal domains (Fig. 5a) [80]. PPIase domain P1 is packed against

the core structure of the N- and C-terminal domains and does not show significant activity, while the more active P2 domain extends away from the core structure [77,80,81]. The PPIase activity of P2 has been shown to be increased in the presence of the adjacent chaperone domain, presumably as this domain facilitates substrate binding close to the active site of P2 [82]. Deletion of both PPIase domains, however, did not cause a significant loss of SurA function *in vivo* and the isolated PPIase domains failed to complement activity in *surA* deletion mutants [81]. This led to the conclusion that SurA functions mainly as a chaperone [81]. Interestingly, mutations which would be expected to cause a loss of PPIase function in the P1 domain, if this domain were active, destabilised SurA *in vitro* but increased chaperone activity *in vivo* [83]. This result suggests a regulatory function of the P1 domain, explaining its lack of significant PPIase activity [83].

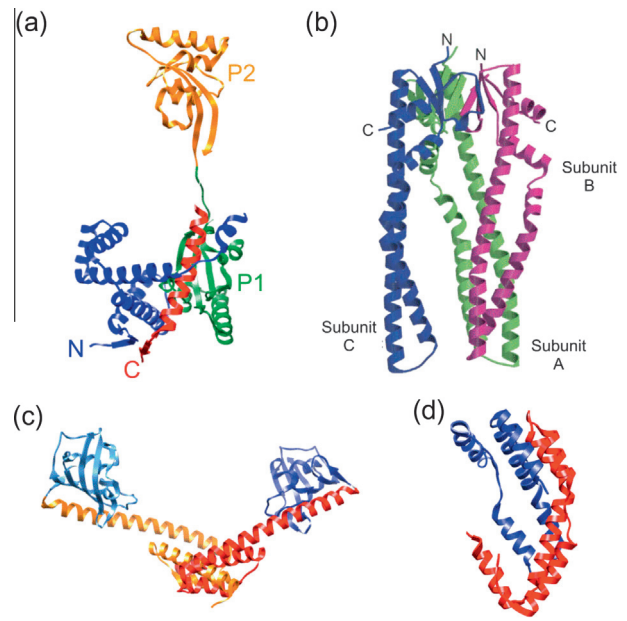
Binding studies using peptide sequences have been carried out on SurA *in vitro*, revealing a preference for aromatic-rich sequences with an Ar-X-Ar motif [84,85]. Peptide sequences containing this motif bind to SurA with dissociation constants in the micromolar range [84], suggesting that SurA binds to OMPs by recognition of this motif, which is frequently found in OMP C-terminal regions [78,86]. The C-terminal Y-X-F motif is responsible for activating DegS and triggering the  $\sigma^E$  stress response [73,78] and, interestingly, it has been reported that SurA does not show significant binding affinity to this motif [87]. In contrast with the peptide binding data, only a few studies have reported the binding of full length OMPs to SurA [88–91] and one study failed to detect stable SurA:OMP complexes [92]. The available data, however, suggest that unfolded OMPs are bound by SurA in preference to unfolded soluble proteins, while folded proteins do not bind [89]. While the polypeptide binding site of SurA has not been identified conclusively, an extended crevice, located in



**Fig. 4.** Schematic of the current model of biogenesis and chaperoning of OMPs in *E. coli*. (a) OMPs are synthesised on the ribosome before post-translational translocation across the inner membrane by the SecYEG translocon. Unfolded OMPs are then chaperoned across the periplasm to the  $\beta$ -barrel assembly machinery (BAM) complex, which aids folding and insertion into the OM. BAM complex proteins are labelled A–E, and the periplasmic polypeptide transport-associated (POTRA) domains of BamA are labelled P1–5. Horizontal black lines indicate the approximate position of the inner and outer membranes. (b) Flow diagram of the periplasmic and outer membrane-anchored proteins which may be implicated in OMP biogenesis. (a) was adapted from [92] with permission from Elsevier, © 2013, while (b) was reproduced from [73] with permission from John Wiley and Sons, © 2005.

the core region, was observed in the crystal structure of SurA. This is thought to be the site through which SurA:client interaction occurs [80].

Sklar and co-workers demonstrated that depletion of SurA causes a loss of OM density, suggestive of OMP assembly defects, that was not observed when the other periplasmic folding factors, Skp and DegP, were depleted [93]. Additionally, SurA has been shown to be localised at the OM [84] and is the only periplasmic folding factor to have been successfully cross-linked to BamA of the  $\beta$ -barrel assembly machinery (BAM) complex *in vivo* [93]. It is not yet known, however, if SurA binds directly to BamA, or if this interaction is mediated through SurA-bound substrates [86]. Together, these observations have led to the hypothesis that SurA is the main chaperone for OMP transport *in vivo* [93,94], although



**Fig. 5.** Crystallographic structures of selected periplasmic chaperones. (a) Ribbon diagram of SurA coloured as follows N-terminal domain (blue), PPlase domain P1 (green), PPlase domain P2 (orange) and C-terminal domain (red) (1M5Y [80]). (b) Ribbon diagram of Skp trimer with the subunits A, B and C coloured in green, magenta and blue, respectively, (1U2M [104]). The tips of the  $\alpha$ -helices in subunits A and B have been modelled. (c) Ribbon diagram of the FkpA dimer showing the N-terminal chaperone domains (red and orange) through which dimerisation occurs and the C-terminal PPlase domains (blue) (1Q6H [109]). (d) Ribbon diagram of the Spy dimer with the monomers coloured individually in red and blue (3O39 [114]). (a), (c) and (d) were generated from PDB files using the accession numbers given in brackets using UCSF Chimera molecular visualisation application [236]. (b) was reproduced from [104] with permission from Elsevier, © 2004.

there is clearly much to learn about the function of this protein at both cellular and molecular levels.

### Skp

Although Skp had previously been suggested to be a DNA-binding protein, an OMP and an LPS binding protein, it is now known that Skp is a periplasmic protein, as evidenced by the N-terminal signal sequence which targets Skp for translocation across the inner membrane [73]. Depletion of Skp *in vivo* led to a moderate reduction of OmpC, OmpF, OmpA and LamB in the OM fraction [95]. These cells, however, were still viable demonstrating that Skp is not essential. Further genetic studies revealed that both the *skp surA* deletion mutant and the *degP surA* deletion mutant result in a synthetic lethal phenotype, leading to the prevailing hypothesis that the periplasmic chaperones Skp and DegP act on redundant pathways separate to SurA (Fig. 4a) [94]. Proteomic analysis of a *skp* deletion mutant suggested that none of the OMPs identified were affected significantly; however, depletion of SurA in the *skp* null strain reduced the levels of almost all OMPs, consistent with the hypothesis of parallel chaperone pathways [96]. While this hypothesis explains many of the observations above, the accumulation of protein aggregates in the periplasm of *skp degP* deletion mutants suggests that Skp may have a role in maintaining the solubility of at least some OMPs prior to folding [97]. Additionally, *skp fkpA* deletion mutants showed defects in the folding of LptD, an essential OMP involved in LPS synthesis [98]. Over-expression of SurA could not ameliorate these effects [98]. Together, these results suggest that the chaperone pathways in the periplasm may be inter-dependent.

The chaperone activity of Skp in OMP folding was confirmed recently by the finding that Skp binds with high affinity ( $K_d \sim$  nM) to

unfolded OMPs [95,99,100]. Furthermore, Skp binding to OmpA was shown to occur via the transmembrane domain of OmpA [95]. NMR spectroscopy confirmed that the transmembrane domain of OmpA is Skp bound, while the periplasmic domain is free to fold independently in solution [100]. Interaction of unfolded OMPs with Skp is thought to occur early after translocation across the IM, as evidenced by the ability to cross-link Skp to the OMP PhoE at the periplasmic side of the IM in spheroplasts [101]. Skp has also been shown to be required for the release of newly-translocated OmpA from the periplasmic side of the IM in spheroplasts [97], suggesting that the N-terminal residues of the client protein may interact with Skp as they emerge from the translocon. This hypothesis was supported by Förster resonance energy transfer (FRET) experiments which showed that the N-terminal residues of OmpC enter the Skp cavity first [102]. Molecular dynamics (MD) simulations yielded similar results [102].

Skp is a trimeric protein in solution with a structure that has been described as “jellyfish-like” (Fig. 5b) [103,104]. A  $\beta$ -barrel domain forms the body of the jellyfish, with long  $\alpha$ -helical “tentacles” protruding from the  $\beta$ -barrel [103,104]. Skp is a basic protein, with a theoretical isoelectric point (pI) around 9.5 [73]. The charge distribution of Skp is notable, with many basic residues clustered around the tips of the  $\alpha$ -helical tentacles resulting in a large macrodipole moment [103,104]. This unusual distribution of charges suggests a role of Skp in substrate delivery to negatively charged membranes. Additionally, a putative LPS-binding site was identified on the outer edge of the Skp  $\alpha$ -helices, similar to an LPS-binding motif found on the OMP FhuA [104]. This finding led to the suggestion that Skp may act as a stress chaperone, acting to rescue OMPs which veer off the SurA chaperone pathway [105].

The  $\alpha$ -helical domains of Skp define a central cavity enriched with hydrophobic residues, which has been shown to be involved in sequestering unfolded substrates to prevent their misfolding and aggregation [88,90,100,102–104]. In contrast with SurA, no common binding motif has been identified for the interaction of Skp with its substrates [73,90] and both electrostatic and hydrophobic interactions have been implicated as being important for binding [99,102,106]. In the absence of a specific binding motif, the high affinity interaction between Skp and its substrates has been shown to arise from the formation of a large number of simultaneous weak interactions that exist between the chaperone and its substrate [90]. Despite a low sequence identity, Skp shares a striking structural resemblance to prefoldin [104], an ATP-independent, cytosolic chaperone in eukaryotes and archaea, which protects partially folded proteins from aggregation and passes clients to other chaperones to promote folding [103,104]. The transient nature of the multitude of weak, local interactions between Skp and its substrates has also been hypothesised to facilitate handover of substrates to other chaperones [90], which could point to functional, as well as structural similarities, between these ATP-independent chaperones. Conformational analysis of Skp-bound OmpX revealed a highly flexible, but compact, ensemble which lacks secondary structure [90], in contrast with the molten globule-like conformations of substrates bound by ATP-dependent chaperones such as GroEL [107]. This difference in substrate binding provides further evidence that ATP-independent chaperones, such as Skp, rely on a thermodynamic gradient for substrate release and subsequent folding [90,108]. In this model, the large free energy of folding of OMPs provides a “thermodynamic sink” that drives the release of these client proteins from Skp and enables their subsequent folding in the OM [108]. The ability of Skp to prevent aggregation of water-soluble proteins has also been shown using clients including lysozyme [104] and single chain antibodies [106], revealing that Skp is able to chaperone soluble proteins as well as OMPs. Skp may, therefore, act as a universal chaperone assisting the folding of both soluble proteins and OMPs in the periplasm.

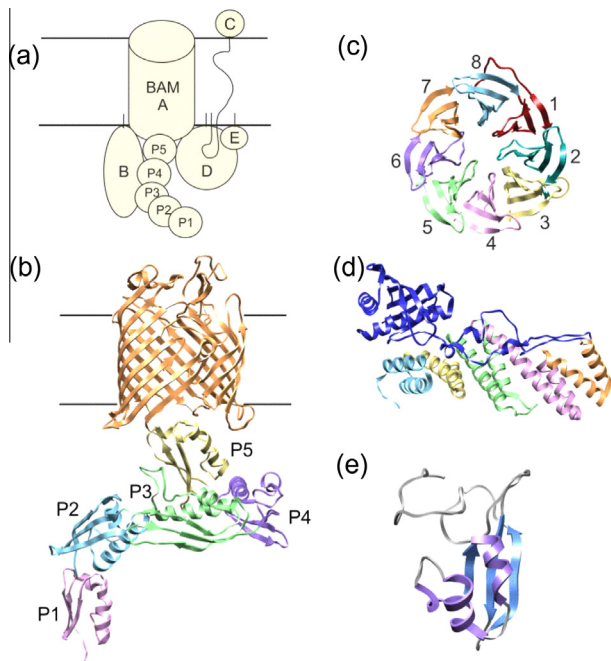
#### Other periplasmic folding factors

While SurA and Skp are currently the most studied of the periplasmic folding factors, other proteins have been shown to participate in OMP biogenesis. Three other periplasmic PPIase enzymes are currently known: PpiA, PpiD and FkpA [64]. While PpiA has been shown to have the highest activity of these proteins *in vitro*, its deletion had no detectable effect on the assembly of OMPs *in vivo* and, as yet, no evidence has directly linked PpiA to OMP biogenesis [64,73]. In contrast, deletion of the inner membrane-anchored PpiD (Fig. 5a) was reported to cause a reduction in the levels of OmpC, OmpF, OmpA and LamB in the OM of the mutant cells and a *ppiD surA* deletion mutant was reported to be lethal, implicating PpiD in OMP folding [64,73]. Later studies did not replicate these findings, however, and the role of PpiD in OMP assembly remains an open question [64].

FkpA, like SurA, is an example of a dual PPIase-chaperone folding factor and *fkpA* null mutants have been shown to be viable, but to display increased OM permeability and an up-regulation of periplasmic proteases [67]. Structural studies revealed that FkpA has a C-terminal PPIase domain, appended to an N-terminal chaperone domain that mediates dimerisation to form a v-shaped cleft (Fig. 5c) [109]. It has been hypothesised that substrate binding occurs in this cleft [109], but no data are currently available to support this. Until recently, chaperone activity of FkpA had been reported only for soluble protein substrates and the involvement of FkpA in OMP biogenesis was not well supported [64,109]. Creation of a *skp fkpA* deletion mutant by Schwalm and co-workers, however, showed that folding of LptD is compromised in this strain, providing the first evidence that FkpA may indeed chaperone OMPs in the periplasm [98].

Another example of a dual-function chaperone in the periplasm is the protease-chaperone DegP which, alongside DegS, belongs to the HtrA family of proteases [73]. Temperature changes were initially thought to cause the switch between the two activities of DegP, with the chaperone activity dominating at 28 °C and the protease activity becoming dominant at 42 °C [73]. Structure-function analysis of DegP concluded that the resting state of this protein is a hexamer, in which the interactions between subunits block the protease sites and the central cavity is large enough to accommodate unfolded substrates [110]. Similarly to DegS, DegP recognises the C-terminal residues of misfolded OMPs via its PDZ domains and this was thought to cause structural reorganisation to larger proteolytically-active oligomers comprised of 12 or 24 DegP monomers [78,111,112]. Cryo-electron microscopy of DegP in the presence of OmpC revealed a tetrahedral arrangement of DegP trimers with the inner cavity filled by a cylindrical area of electron density into which the structure of folded OmpC could be modelled [111]. This led to the conclusion that the fate of DegP-bound substrates lies in their ability to adopt their native structure within the cavity, as only unfolded substrates can be degraded [111]. Recent data using cage-deficient mutants of DegP, which can only associate into trimers, revealed that these variants are able to bind and degrade substrates without the need to form higher order oligomers, suggesting that cage formation in response to substrate binding may be linked solely to the chaperone function of this protein [113].

The most recently discovered periplasmic chaperone, Spy, was identified by its over-expression in bacterial strains expressing unstable periplasmic proteins [114]. *In vitro* characterisation of Spy revealed it to be an effective chaperone, suppressing aggregation and protecting substrates from inactivation by tannins [114]. Spy has a novel  $\alpha$ -helical cradle structure, which is unlike that of any known chaperone (Fig. 5d) [114]. It is not known, currently, whether Spy is involved in OMP biogenesis, but the discovery of a previously unknown chaperone only two years ago highlights that there is much still to be learned about the chaperone network



**Fig. 6.** Structure of the BAM complex. (a) Schematic of the *E. coli* BAM complex with BAM proteins labelled A–E and POTRA domains labelled P1–5 (reproduced from [92] with permission from Elsevier, © 2013). (b) Crystal structure of *N. gonorrhoeae* BamA (4K3B [120]). The  $\beta$ -barrel domain is shown in orange. POTRA domains are labelled as in (a) and are shown in pink, blue, green, purple and yellow. (c) Crystal structure of *E. coli* BamB (3P1L [130]). The blades of the  $\beta$ -propeller structure are coloured individually and labelled 1–8. (d) Crystal structure of the N-terminal domain of *E. coli* BamC (dark blue) bound to BamD (2YHC [125]). The five TPR motifs of BamD coloured in light blue, yellow, green, pink and orange. (e) Lowest energy structure of *E. coli* BamE solved by NMR spectroscopy (2KXX [237]). All images were generated from the PDB files using the accession numbers given in brackets using UCSF Chimera molecular visualisation application [236].

in the *E. coli* periplasm. In addition to the complexity evident in a single species, studies on the periplasmic chaperones of *Neisseria meningitidis* show that SurA deletion causes no detectable defects in OMP assembly, while deletion of Skp caused lower levels of some porins, but not all OMPs [115]. Complementation of the *skp* null mutant was not possible using *E. coli* Skp, suggesting that periplasmic chaperones may act in a species-specific manner, making the derivation of generic principles of periplasmic chaperoning difficult to achieve [115].

#### Insertion into the outer membrane

Following chaperone-assisted transport across the periplasm, unfolded OMPs must insert and fold into the OM, a process which is assisted by the BAM complex [116]. The BAM complex in *E. coli* (Fig. 6a) is comprised of the outer membrane protein BamA (previously YaeT, Fig. 6b) and four accessory lipoproteins, BamB (YfgL, Fig. 6c), BamC (NlpB, Fig. 6d), BamD (YfiO, Fig. 6d) and BamE (SmpA, Fig. 6e) [105,116]. BamA is an essential protein, the depletion of which causes an accumulation of aggregated OMPs in the periplasm, leading to cell death [117]. BamA belongs to the conserved Omp85 super-family first identified in *N. meningitidis* [118]. Homologues are also found in mitochondria (Sam50) and chloroplasts (Toc75) [119]. All of the BamA homologues identified thus far have a C-terminal transmembrane  $\beta$ -barrel domain and a water-soluble N-terminal region [116]. The structure of the BamA  $\beta$ -barrel in two different organisms has been elucidated recently,

revealing it to be 16-stranded with a conserved VRGY motif in loop L6 which is thought to be functionally important (Fig. 6b) [120].

The N-terminal soluble region of BamA comprises different numbers of polypeptide transport-associated (POTRA) domains, dependent on the identity of the organism. Bacterial BamA homologues tend to have five POTRA domains. Toc75 in chloroplasts, however, has three POTRA domains and Sam50 in mitochondria has only one POTRA domain [116]. Crystallisation of the four N-terminal POTRA domains (POTRAs 1–4) of *E. coli* revealed that despite their low sequence identity, the overall fold (a 3-stranded  $\beta$ -sheet with 2  $\alpha$ -helices) is conserved (Fig. 6b) [119]. Previous studies of the binding of OMPs revealed a direct interaction of BamA with unfolded OMPs and peptides mimicking the conserved aromatic-rich C-terminal sequence of OMPs [121]. Evidence of  $\beta$ -augmentation between POTRA domains of neighbouring proteins in the crystal structure of *E. coli* POTRA domains 1–4 suggests a possible mechanism for substrate binding [119]. Deletion of individual POTRA domains from BamA demonstrated that one of the functions of these soluble domains is to provide a scaffold for the association of the lipoproteins which make up the BAM complex with BamA [119]. Interestingly, while POTRAs 3–5 are essential for cell viability in *E. coli*, only POTRA 5 is essential in *N. meningitidis* [116], suggesting differences in the roles of these conserved domains between bacterial species.

Of the four accessory lipoproteins in the BAM complex, only BamD has been shown to be essential for cell viability in *E. coli*, implying an important role for this lipoprotein in OMP biogenesis [116]. BamD is a predominantly  $\alpha$ -helical protein with 5 tetratricopeptide repeat motifs (Fig. 6d), sharing structural similarity with other proteins which recognise C-terminal targeting sequences [122]. BamD interacts with BamA via POTRA domain 5 [116] and, accordingly, a mutation in POTRA 5 of BamA that is lethal at 37 °C causes dissociation of BamA from the BamCDE subcomplex [123]. Recent experiments *in vivo* using point mutations in BamD and BamA have suggested that BamD regulates the activity of BamA [123,124]. Co-crystallisation of BamD with the non-essential lipoprotein BamC (Fig. 6d) revealed that a binding pocket in the N-terminal region of BamD is the site of interaction of these two lipoproteins [125]. A truncated variant of BamD was used to demonstrate binding of this lipoprotein to synthetic peptides based on the OMP C-terminal targeting sequence, and this interaction occurred at the same binding pocket as the interaction with BamC [116,125]. Together, these data have led to the hypothesis that BamC may act as a regulator of BamD substrate binding [116], although the precise role of BamC (which has been shown to span the OM with parts of its sequence exposed to the extracellular environment (Fig. 6a) in OMP assembly remains unclear [126].

BamE (Fig. 6e) is the most recently discovered accessory lipoprotein of the BAM complex. This lipoprotein associates with the BAM complex via interaction with BamD and has been shown to cause only minor defects in OMP assembly when depleted [116,127]. NMR spectroscopy revealed that BamE binds preferentially to PG lipids [127]. These lipids have been shown to enhance the insertion of OMPs into liposomes *in vitro* [116], suggesting that a role of BamE may be to promote OMP insertion into the OM. Interestingly, in *AbamE* and BamD R197L expressing strains of *E. coli*, BamA becomes more susceptible to degradation by proteinase K, suggesting that BamE might also be involved in conformational modulation of BamA [124,128]. While BamC and BamE are associated with BamA via interaction of BamD with the soluble POTRA domains, BamB (Fig. 6c) [129–132] is associated directly with BamA via POTRAs 2–5 (Fig. 6a) [116,119]. Although not essential for cell viability, deletion of BamB results in reduced OMP assembly and a phenotype similar to SurA depletion mutants [133]. The OMPs most affected by deletion of BamB are those with larger  $\beta$ -barrels and, thus, it has been suggested that BamB has a role in



substrate delivery to BamA, perhaps by increasing the substrate binding capacity of the BAM complex [116]. Indeed, BamB and BamD have been shown recently to bind to unfolded OmpA and unfolded BamA in the absence of the other BAM components [134]. Interestingly, both BamB and BamD accelerate BamA assembly into liposomes, however, OmpA folding was not similarly assisted by either lipoprotein, suggesting that one role of the BAM lipoproteins is to facilitate assembly of BamA [134].

Exciting recent work by Kahne and colleagues has demonstrated that the BAM complex can be reconstituted in a functional form *in vitro* [135]. The complex was found to have a 1:1:1:1 ratio of BamA:B:C:D, but the stoichiometry of BamE could not be determined due to its small size [135]. The activity of the BAM complex was monitored using the folding of the OM  $\beta$ -barrel peptidase enzyme, OmpT, which can be measured directly by cold SDS-PAGE or indirectly by following OmpT enzymatic activity [135,136]. Compared with experiments in which the intact BAM complex was present, OmpT folding efficiency was reduced by approximately 6-fold in the absence of BamB [135,136]. Reduced OmpT folding efficiency was also observed in the absence of SurA, even when the intact BAM complex was present [135], supporting the previous suggestion that BamB and SurA have similar, but not redundant, roles *in vivo* [116,135]. These results also demonstrate that the BAM complex is able to complete the OMP assembly cycle without the aid of additional cellular components [136].

The mechanism of action of the BAM complex during OMP folding and membrane insertion is still under debate and several models have been proposed [116]. These include substrate translocation across the OM through BamA, followed by folding into the OM from outside the cell. Alternatively, the BamA  $\beta$ -barrel may act as a scaffold for  $\beta$ -sheet formation for the folding OMP or formation of BamA multimers may create a pore in the OM through which the substrate can insert [116]. Recent elucidation of the crystal structure of BamA from two different organisms, however, shows a large cavity in the BamA  $\beta$ -barrel within which a substrate could be accommodated [120]. Additionally, MD simulations point to the existence of a lateral gate in the BamA  $\beta$ -barrel, potentially allowing substrates to move from the barrel interior into the membrane [120]. The observation of two conformations of BamA with different resistance to proteinase K has been observed previously [124,128]. These data and the observation that the conserved L6 loop is only accessible for labelling in one conformation [124], suggest that opening of the lateral gate may be modulated by movements of the L6 loop. Indeed, structural studies of BamA [120] and two homologues, FhaC [137] and TamA [138], revealed that the L6 loop adopts a different conformation in the FhaC crystal structure to that observed in the TamA and BamA crystal structures. This may suggest that L6 loop movements are a conserved functional feature across the Omp85 super-family. Lateral opening in TamA [138] has also been suggested MD simulations suggest, however, that FhaC does not share this functional feature [120]. These differences may be reflective of differences in substrate handling by these OMPs, with FhaC exporting substrates across the OM, while TamA and BamA most likely insert substrates directly into the OM. In contrast with the wealth of information about the mechanism of action of many cytosolic chaperones and chaperonins [41,139,140], there is clearly much to learn about the mechanism of action of the BAM complex and its associated periplasmic chaperones in facilitating the folding and membrane insertion of OMPs.

#### Application of different biophysical methods to the study of OMP folding into lipid bilayers

In recent years, many techniques have been developed to study the folding of water-soluble proteins and together these have

yielded a near-atomistic view of the folding landscapes of many such proteins [31]. By contrast, the experimental toolbox for the study of membrane protein folding is more limited [141]. Much of the information available about OMP folding into lipid bilayers has resulted from the use of techniques such as cold SDS-PAGE (where samples are not boiled prior to loading, resulting in the differential migration of the folded and unfolded conformations of the OMP [142]) and tryptophan fluorescence emission and far-UV circular dichroism (CD) spectroscopies on a limited subset of proteins, as summarised below and in Table 1. In spite of the challenges faced, the development of new methodologies and the application of existing methods to study OMP folding are now beginning to yield knowledge of OMP folding mechanisms both in the presence and absence of chaperones.

As well as the methods mentioned above, the quenching of tryptophan fluorescence has been utilised successfully to follow the folding and membrane insertion of OmpA [143,144]. In this technique, lipids which are brominated at different positions in the acyl chain are introduced into liposomes. The depth of membrane insertion of an OMP during a folding reaction is measured by following the kinetics of tryptophan fluorescence quenching by the bromine atoms within the bilayer [144]. Additionally, the use of mutagenesis to create OmpA variants with only a single tryptophan residue allowed the insertion of different regions of the OmpA  $\beta$ -barrel into the bilayer to be monitored. For example, varying the location of a single tryptophan residue can yield information on whether individual secondary structural elements are inserted sequentially or simultaneously [143]. Alongside other spectroscopic techniques, these experiments provided the first evidence of how an OMP folds, revealing that the folding and membrane insertion of OmpA is a concerted process [143,144].

NMR studies on unfolded water-soluble proteins has provided evidence that the starting point of folding is not a random coil structure [22,23,28,145,146]. Indeed, residual structure in the unfolded state has been suggested to be important in initiating folding by facilitating the collapse of the polypeptide chain into conformations able to fold efficiently to the native state [22]. In contrast with helical IM proteins, OMPs can be denatured in urea or guanidine hydrochloride (GuHCl), enabling studies of their folding using classic Anfinsen-style experiments [43]. Analysis of the urea denatured state of OmpX has shown that the protein is globally unfolded, but has two regions of non-random structure: one a hydrophobic cluster and the other a helical region [147]. Analysis of peptides corresponding to these regions showed independent binding of the clusters to detergent micelles, suggesting a role of residual structure in the unfolded state in the initiation of OMP folding and membrane insertion [148]. The application of NMR spectroscopy to folded membrane proteins is complicated by the need to find a suitable mimic of the membrane environment. Detergent micelles are widely used, but problems with long term stability and the maintenance of proteins in a functional state are commonly encountered [149]. As a result, development of alternative non-micellar systems such as bicelles, nanodiscs and amphipols, to stabilise the folded state of membrane proteins has become an active research area [149–151] and is reviewed elsewhere in this issue.

Detailed information about the folding mechanisms of water-soluble proteins has been obtained by measuring the folding and unfolding kinetics of the protein of interest, using a spectroscopic probe to monitor the reaction time-course in the presence of varying concentrations of denaturant. The logarithm of the kinetic rate constants acquired from these data can be plotted against the denaturant concentration, giving a characteristic v-shaped chevron plot [152]. If the limbs of the chevron remain linear over denaturant concentrations from 0 M to the highest concentration used (highly denaturing) and the free energies associated with the

**Table 1**  
Summary of selected *in vitro* folding studies of outer membrane proteins (OMPs) into lipid bilayers.

Protein	Notes	Techniques used	Selected References
OmpA	Folding studies of OmpA are described in this review and elsewhere.	Cold SDS–PAGE, far-UV CD, Trp fluorescence	Kleinschmidt (2006) [171]; Otzen (2013) [58]
PagP	Folding studies of a C-terminally his-tagged construct of PagP (HT PagP) and untagged construct (PagP) are described in detail in this review.	Cold SDS–PAGE, far-UV CD, Trp fluorescence, $\Phi$ -value analysis	Bishop et al. (2000) [239]; Ahn et al. (2004) [206]; Huysmans et al. (2010) [157]
hVDAC	Human voltage-dependent anion-selective channel (hVDAC) can be folded to the native state in LUVs composed of lipids of varying acyl chain length ( <i>diC</i> <sub>10:0</sub> PC to <i>diC</i> <sub>18:1</sub> PC). Folding yield was estimated to be 94% in <i>diC</i> <sub>12:0</sub> PC LUVs. Secondary structure content of hVDAC in <i>diC</i> <sub>12:0</sub> PC LUVs was not affected by changing pH from 7.0 to 3.0.	Cold SDS–PAGE, far-UV CD, Trp fluorescence, sucrose density gradient centrifugation and proteolysis	Shanmugavadivu et al. (2007) [225]
FomA	FomA can be folded to the native state in both <i>diC</i> <sub>10:0</sub> PC and <i>diC</i> <sub>18:1</sub> PC (LUVs and SUVs). Kinetic analysis of folding into <i>diC</i> <sub>10:0</sub> PC and <i>diC</i> <sub>18:1</sub> PC SUVs suggested that FomA folds <i>via</i> parallel pathways into both lipids. The folding half-time is dependent on acyl chain length and reaction temperature.	Cold SDS–PAGE, Trp fluorescence, far-UV CD	Pocanschi et al. (2006) [202]
OmpG	The porin OmpG reconstituted in native <i>E. coli</i> lipids is gated by conformational changes in extracellular loops in a pH-dependent manner (closed at pH 5.0). Unfolding under force reveals each $\beta$ -hairpin unfolds individually. Refolding from this mechanically unfolded state also proceeds by sequential folding of individual $\beta$ -hairpins.	Atomic force microscopy (AFM)	Sapra et al. (2009) [240]; Damaghi et al. (2010) [241]; Mari et al. (2010) [242]; Damaghi et al. (2011) [243]
OmpF	Refolding of urea-solubilised OmpF into <i>diC</i> <sub>14:0</sub> PC SUVs occurred at only 15% yield. Refolding kinetics were biphasic but much slower than OmpA.	Cold SDS–PAGE, Trp fluorescence, far-UV CD	Surrey et al. (1996) [244]

folding and unfolding events equate to the equilibrium unfolding free energy, the data suggest that the protein under study probably folds *via* a two-state mechanism [13,14]. Deviations from linearity, termed rollover, in either limb can occur and have been interpreted to indicate the presence of a folding intermediate [152,153], movement of the transition state [154] or aggregation [155]. As well as being an indicator of folding mechanism (alongside equilibrium folding analysis), chevron plot analysis also yields information about the compactness of the transition state (or any populated intermediates) and their position on the folding reaction coordinate [152]. This detailed kinetic analysis is routine in the study of water-soluble protein folding and its recent application to study the folding of OmpA [156] and PagP [157] will be discussed later in this review.

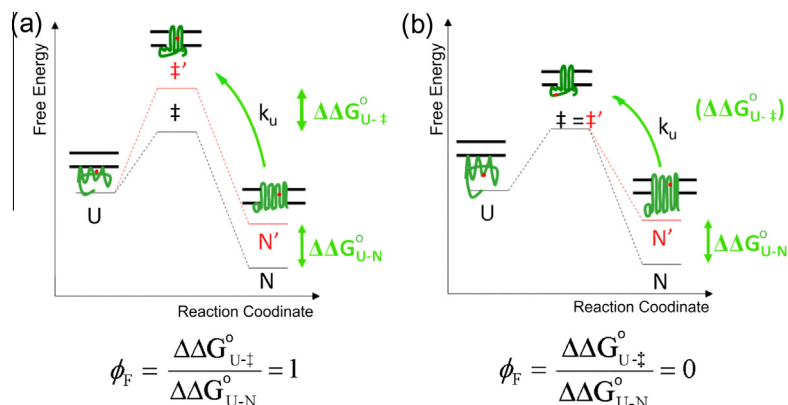
Protein engineering methods coupled to kinetic analysis have been the most successful tool to probe the folding mechanisms of membrane proteins and have been applied to the IM proteins bR [44,158,159] (Fig. 2b), DsbB [160] and the OMP, PagP (Fig. 2d) [157].  $\Phi$ -Value analysis is a powerful protein engineering approach, which can be used to map the formation of contacts in the transition states and populated intermediates formed during protein folding (Fig. 7) [161–163]. In this approach, specific side-chain interactions are deleted by mutation, and the effect on the kinetic and thermodynamic parameters of the variant protein is measured and compared with wild-type [163]. The amino acid substitution may cause a change in the free energy of activation (kinetic), the equilibrium (thermodynamic) free energy of the folding reaction, or both. The ratio of these changes is the  $\Phi$ -value, which is usually between 0 and 1, and gives a measure of the change of stability of the transition state (or intermediate), compared with the change in stability of the native state [163]. Both stabilities are measured relative to the denatured state, whose free energy is assumed to be unaffected by the mutation [163].  $\Phi$ -Values close to zero indicate that the transition state (or intermediate) is unstructured in the region of the amino acid substitution, while higher  $\Phi$ -values indicate that the region is structured in the transition state (or intermediate) [161,162]. Partial  $\Phi$ -values are also observed and can be interpreted in several ways. For example, partial  $\Phi$ -values may result if native contacts are partially formed in the transition state (or folding intermediate) or if multiple folding routes exist [163].  $\Phi$ -Values have

provided a wealth of information on the structure of partially folded intermediates and transition states of water soluble proteins, revealing detailed information about their folding mechanisms [163]. Furthermore, the use of experimental  $\Phi$ -values as constraints in MD simulations allows atomistic models of these ensembles to be created [164]. The main stumbling block for the application of  $\Phi$ -value analysis to OMPs is the need to find experimental conditions that enable reversible folding for the protein of interest, which has proved difficult for many OMPs [58,156,165].

### Folding studies of OMPs *in vitro*

#### OmpA

The study of OMP folding *in vitro* has been pioneered by work on the monomeric, 8-stranded ion channel OmpA (Fig. 2e) from *E. coli*. Initial experiments using a combination of cold SDS–PAGE and protease digestion were conducted by Schweizer et al. as early as 1978 [166], which demonstrated that OmpA can fold in the presence of LPS and Triton X-100 detergent. It was not until over a decade later that Surrey and Jähnig reported the folding of OmpA into lipid bilayers [167], opening the door to studies of membrane protein folding in a more native-like environment. OmpA, solubilised in 8 M urea, was shown by these authors to insert spontaneously into small unilamellar vesicles (SUVs) of *diC*<sub>14:0</sub>PC upon rapid dilution from urea [167]. Detailed kinetic studies were carried out by these, and other, authors on the refolding of urea-unfolded OmpA into liposomes using cold SDS–PAGE, far-UV CD and tryptophan fluorescence emission [143,168,169]. Kinetic measurements of the formation of secondary structure by far-UV CD and tertiary structure by tryptophan fluorescence emission revealed that these structural elements form concomitantly [170]. Additionally, the rate of quenching of single tryptophan mutants of OmpA by brominated lipids suggested that the 8  $\beta$ -strands penetrate the membrane simultaneously [144]. Based on these data a scheme for the refolding pathway of OmpA was proposed, beginning with collapse of the protein in aqueous solution, followed by adsorption to the membrane surface and folding to the native state by progressive penetration deeper into the membrane as the  $\beta$ -barrel forms (Fig. 8) [143,144,168,169]. These experiments provided the first



**Fig. 7.** Schematic of the principles of  $\Phi$ -value analysis. (a) A mutation (shown as a red dot) is made in a region of the protein which is native-like in the transition state ( $\ddagger$ ) leading to equal destabilisation of  $\ddagger$  and the native state (N) resulting in a  $\Phi$ -value of 1 or (b) a mutation is made in a region of the protein which is unfolded in  $\ddagger$  but structured in the native state, leading to destabilisation of N only and a  $\Phi$ -value of 0. It is assumed that the mutation does not affect the free energy of the unfolded ensemble (U). Abbreviations:  $\Delta\Delta G_{U-N}^\circ$  refers to the difference in the free energy of folding upon mutation and  $\Delta\Delta G_{U-\ddagger}^\circ$  refers to the difference in the free energy between U and  $\ddagger$  upon mutation. This image was adapted with permission from G.H.M. Huysmans.

evidence of the concerted nature of Omp folding and membrane insertion [143,144,168,169,171].

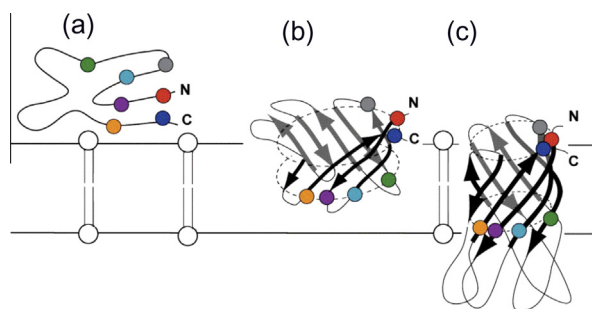
More recently, measurement of the stability of OmpA has been achieved in several membrane mimetic environments [156,172,173]. Finding experimental conditions under which membrane protein folding is fully reversible and reaches equilibrium on experimentally tractable timescales *in vitro* is challenging and continues to impede studies of membrane protein folding and stability [174]. Folding of OmpA into SUVs composed of 92.5% C<sub>16:0</sub>C<sub>18:1</sub>PC and 7.5% C<sub>16:0</sub>C<sub>18:1</sub>PG was reported to be reversible [172]. Good agreement of the rate of folding when measured by tryptophan fluorescence emission and cold SDS-PAGE implied a two-state folding mechanism [172]. Introduction of guest lipids into this reversible system allowed the effects of bilayer thickness and the extent of lipid acyl chain saturation on OmpA stability to be measured, revealing that forces within the bilayer, such as curvature stress, affect thermodynamic stability [172]. Kinetic analysis of OmpA refolding into large unilamellar vesicles (LUVs) of varying diameter confirmed the role of curvature elastic stress in modulating the folding rate, suggesting that the irreversibility of folding observed for OmpA into LUVs may be a kinetic trapping effect [175].

Establishing conditions under which OmpA folds reversibly allowed double mutant cycle analyses of the role of side-chain-side-chain interactions in the protein folding mechanism [176,177]. After investigating residues involved in OmpA function [176], this method was used to probe how the interactions

between aromatic residues in the intrafacial region of the  $\beta$ -barrel contribute to stability [177]. In both  $\alpha$ -helical and  $\beta$ -sheet membrane proteins, aromatic residues are enriched in the intrafacial region forming “aromatic girdles” [178]. This common feature of OMPs has been hypothesised to stabilise the  $\beta$ -barrel and to fix the location of proteins in the lipid bilayer, based on the partitioning energies of small model peptides into a model bilayer from aqueous solution [179]. Indeed, double mutant cycle analysis of OmpA provided the first evidence that clustering of aromatic residues provides a driving force for Omp folding and stability [177].

Analysis of OmpA folding into octyl maltoside micelles showed that, although OmpA folding remained irreversible when urea was used as a denaturant, reversibility could be obtained using GuHCl as an alternative denaturant [156]. A full kinetic characterisation of the folding mechanism was undertaken using chevron plot analysis and revealed that OmpA folds *via* a three-state mechanism to a native-like protein that has a thermodynamic stability much higher than that measured in SUVs ( $\Delta G_{\text{octylmaltoside}}^\circ \approx -65.2$  kJ/mol,  $\Delta G_{\text{SUVs}}^\circ \approx -14.2$  kJ/mol) [156,172]. This result was confirmed in a subsequent study of OmpA folding into detergent micelles of *N*-lauryl-*N,N*-dimethylammonium-*N*-oxide (LDAO,  $\Delta G_{\text{LDAO}}^\circ \approx -60$  kJ/mol) [173]. By contrast, the thermodynamic stability of OmpA in the amphipol A8-35 was found to be lower than that in detergent SUVs ( $\Delta G_{\text{A8-35}}^\circ \approx -8$  kJ/mol), and the time taken to reach equilibrium during (un)folding in A8-35 was reduced to 25 days, compared with 52 days in LDAO [173]. These studies demonstrate the utility of different bilayer mimics to study membrane protein stability and function, allowing kinetic reversibility on an experimentally tractable timescale *in vitro*.

OmpA was also used in the first *in vitro* studies of the effect of a periplasmic chaperone on OMP folding into bilayers. Skp (Fig. 5b) was shown to maintain unfolded OmpA in a soluble state at low urea concentrations ( $\approx 0.4$  M), to accelerate OmpA folding into negatively charged SUVs, but to retard folding into zwitterionic SUVs [180]. Additionally, pre-binding of the Skp:OmpA complex to LPS prior to addition of lipid was shown to promote folding into either negatively charged or zwitterionic SUVs suggesting that LPS may play a role in OMP assembly [180]. Incorporation of full length BamA into liposomes was found to increase the folding rate constant of the OmpA  $\beta$ -barrel domain into LUVs of *diC*<sub>12:0</sub>PC [181]. The transmembrane domain of BamA alone gave a small, twofold increase in the folding rate constant, which is comparable to the rate increase when FomA, an OMP not implicated in OMP assembly, is incorporated into the liposomes. This suggests that



**Fig. 8.** Proposed mechanism of folding and insertion of the OmpA  $\beta$ -barrel domain into lipid bilayers *in vitro*. (a) Depicts an unfolded, membrane-bound state, (b) depicts a partially folded and inserted state, and (c) depicts the native state. Coloured circles indicate the location of tryptophan residues in the OmpA structure. Reproduced from [238] with permission from Elsevier, © 2011.

the rate increase observed with the BamA  $\beta$ -barrel alone results from non-specific perturbations of the local lipid environment [181]. This study also showed that incorporation of 20%  $diC_{12:0}PE$  into the  $diC_{12:0}PC$  LUVs has an inhibitory effect on OmpA folding which was overcome by the incorporation of BamA into the liposomes. These experiments suggest how OMPs may overcome the barrier to folding *in vivo* by destabilisation of the OM, which is composed of approximately 80% PE lipids, in the presence of BamA [181]. Finally, the previously observed retardation of OmpA folding into zwitterionic bilayers in the presence of Skp was shown to be ablated in the presence of BamA. This is possibly mediated through the interaction of negatively charged patches on the surface of the POTRA domains with positively charged Skp [181,182]. Together, the available data on OmpA folding has provided valuable insights into the forces that govern the folding of OMPs. Lack of such extensive data on other OMPs, however, makes general conclusions hard to draw.

#### Comparative studies of OMP folding

The study of homologous proteins has been utilised widely in the field of protein folding [183,184]. Application of this approach to water soluble proteins includes the homeodomain-like superfamily [16,185],  $\alpha$ -spectrin domains [37,184,186–189], bacterial immunity proteins [190–192], ribosomal S6 proteins [193,194] and immunoglobulin-like domains [195–199]. These studies have yielded valuable insights into the role of protein topology, amino acid sequence and secondary structure propensity on folding mechanisms [183,184]. Recently, Burgess et al. adopted a screening approach using cold SDS–PAGE analysis to determine the folding yield of nine different OMPs (OmpX, OmpW, OmpA, PagP, OmpT, OmpLA, FadL, Omp85 (BamA) and OmpF) under different conditions [200]. The effects of pH, temperature, vesicle size and lipid composition were investigated, revealing that, in general, folding yield was increased by using pH values of 8–10, lipids with shorter acyl chains and smaller vesicle diameter [200]. These results accord with previous work on the effects of membrane thickness and curvature on the folding of OmpA [170]. A single condition to maximise the folding yield of all nine OMPs could not, however, be identified [200].

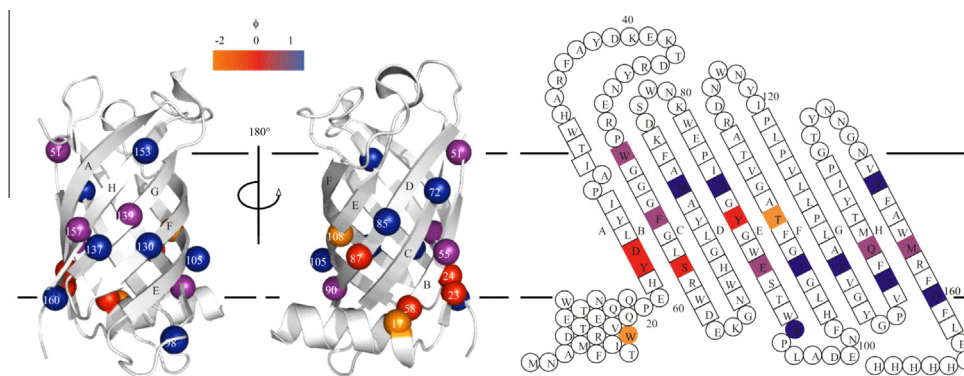
High temperature has been reported to increase the folding yield of OmpA [169,201], FomA [202] and OmpX [201], but increasing temperature had varying effects on the folding yield of the OMPs in the Burgess study [200]. This variation was proposed to arise from the different aggregation propensities of the OMPs studied at higher temperature [200]. The aggregation propensity of the nine OMPs in the urea-unfolded state was investigated using sedimentation velocity analytical ultra-centrifugation, revealing that self-association did not correlate well with folding efficiency [203]. Some trends were observed, however, showing that addition of salt (50–400 mM) and lower pH (values 6–7) increase self-association, while the presence of  $\geq 4$  M urea kept all the OMPs in a monomeric state [203]. The relatively low aggregation propensity of OmpA has since been attributed to the presence of the periplasmic domain, which has been shown to fold independently of the  $\beta$ -barrel domain [204]. A thorough screen to establish conditions for optimal folding of OmpLA was then conducted using the results of previous studies as a guide [200,203], revealing that OmpLA folds reversibly into LUVs  $diC_{12:0}PC$  at pH 3.8, 37 °C [165]. These conditions also allowed reversible folding of PagP and OmpW, but not the other OMPs studied [108]. The difficulty in finding conditions suitable for the folding of different OMPs is thus a challenge, making comparative studies of these homologous proteins challenging. It is therefore difficult to draw out similarities and differences in the folding mechanism of a single protein class in the context of OMP folding mechanisms.

#### PagP

PagP is an OMP from *E. coli* that forms an 8-stranded  $\beta$ -barrel with a 19-residue amphipathic  $\alpha$ -helix at its N-terminus (Fig. 2d) [205–207]. The  $\beta$ -barrel is tilted by approximately 25° to the membrane normal [205,206,208], stabilised in this position by the interactions of the aromatic girdles with the membrane intrafacial region [209]. PagP is a palmitoyl transferase enzyme, which transfers a palmitate chain from a phospholipid to hexa-acylated lipid A and thus helps to reinforce the structure of the outer membrane [210]. Folding studies on a C-terminally His-tagged construct of PagP (HT PagP) revealed that the protein folds into both detergent micelles and liposomes *in vitro* [211]. Unlike OmpA, a high concentration of urea (7 M) was required to solubilise the protein and prevent aggregation prior to insertion [143,211]. Far-UV CD was used to follow the formation both  $\beta$ -sheet structure and tertiary structure (*via* a Cotton effect between tyrosine-26 and tryptophan-66) [211,212]. In accordance with previous results on OmpA, secondary and tertiary structure were found to form concomitantly during HT PagP folding into either cyclofos-7 micelles or  $diC_{12:0}PC$  liposomes (SUVs and LUVs) [170]. Mutants of HT PagP were then created to investigate the role of the N-terminal  $\alpha$ -helix (residues 1–19) in folding, since the possession of a periplasmic  $\alpha$ -helix is an unusual feature of an OMP [211]. This study demonstrated that the helix increases the stability of folded HT PagP in liposomes, but this effect was not so pronounced in detergent, illustrating the importance of the lipid bilayer in the stability of the native state and the importance of developing membrane-like mimics for studies of OMP stability [211]. One mutant in which a conserved residue in the  $\alpha$ -helix, tryptophan-17, was replaced with alanine, unfolded fifty times more rapidly than the wild-type [211]. Moreover, in a HT PagP helix deletion construct, mutation of arginine-59, located in the intrafacial region of the  $\beta$ -barrel domain, to tryptophan restored the folding and unfolding kinetics in liposomes to rates similar to those of the wild type protein, demonstrating the importance of the aromatic girdle in the folding and stability of PagP [213].

By systematically varying the protein concentration and studying the folding of HT PagP under a range of lipid-to-protein ratios (LPRs), conditions were established under which the HT PagP unfolding transition is completely reversible in  $diC_{12:0}PC$  LUVs [157]. Equilibrium stability studies and kinetic chevron plot analysis of HT PagP (un)folding revealed that the protein folds *via* a two-state mechanism over the range of urea concentrations studied (7.8–10 M). A  $\Phi$ -value analysis was then undertaken for HT PagP using point mutants of 19 residues spread throughout the protein structure [157]. These experiments provided the first insights into the structural features of a transition state for OMP folding, suggesting a polarised transition state in which the N-terminal half of the protein remains largely unstructured, whilst the C-terminal half of the protein is native-like (Fig. 9) [157]. Interestingly, two negative  $\Phi$ -values were observed, providing evidence for stabilisation of the transition state by non-native interactions [157]. The resulting mechanism of tilted insertion is consistent with the concerted folding and insertion suggested for OmpA [143,144,170]. It remains to be seen whether this mechanism is observed for other OMPs.

Kinetic analysis of HT PagP (un)folding over a range of urea concentrations (7.8–10 M) revealed that folding into  $diC_{12:0}PC$  LUVs is best described by a burst phase followed by a single exponential phase [157,211,214]. Lowering the lipid-to-protein ratio, the HT PagP concentration, or inclusion of  $diC_{12:0}PE$  in the  $diC_{12:0}PC$  LUVs, resulted in a burst phase followed by two exponential phases [214]. Interrupted refolding experiments, in which folding is allowed to proceed for varying lengths of time before unfolding is re-initiated by dilution into high urea concentrations, were also



**Fig. 9.**  $\Phi$ -Value analysis of HT PagP.  $\Phi_F$ -values determined from kinetic analysis of HT PagP variants are mapped onto a ribbon diagram (left) and a topology model (right). Regions with  $\Phi_F$ -values close to 0 are shown in red, regions with  $\Phi_F$ -values close to 1 are shown in blue, intermediate  $\Phi_F$ -values are shown in purple,  $\Phi_F$ -values less than 1 are shown in orange and undetermined  $\Phi_F$ -values are grey. Reproduced with permission from [157].

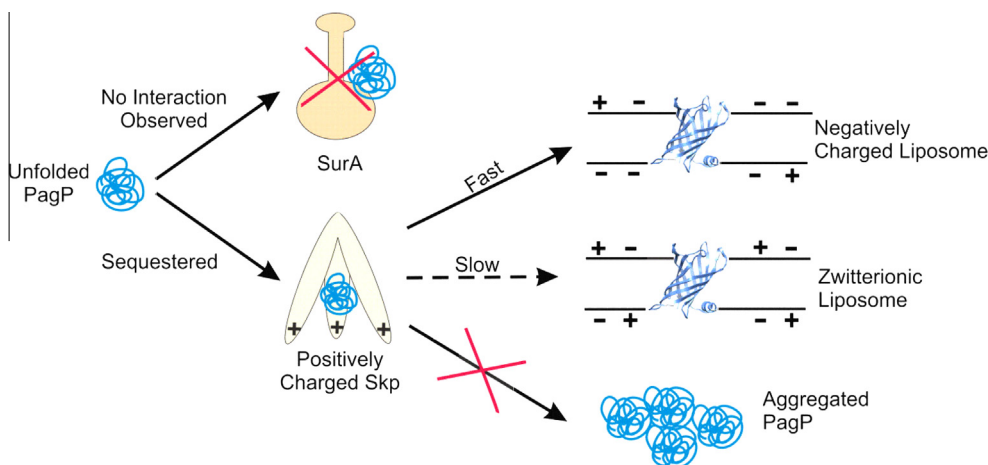
carried out. These experiments revealed that the rate of formation of the native state is best described by two rate constants indicative of parallel folding pathways [214], a feature also seen in the folding of some water-soluble proteins [145,197,215–218]. Since the unfolding kinetics of HT PagP are well described by single exponential kinetics under all conditions tested, there was insufficient evidence for an alternative native-like state of HT PagP [214]. The slower pathway most likely arises from a second population of folding-competent HT PagP molecules in solution ready to adsorb onto the lipid upon exposure of free membrane surface [214].

More recently, kinetic analysis of an untagged variant of PagP has been undertaken in the presence of SurA and Skp [92]. This PagP construct, initially reported by Burgess et al. [200], was chosen for folding assays in the presence of soluble chaperones as it had been reported to be folding competent in urea concentrations as low as 1 M, in stark contrast with the high urea concentrations required for efficient folding of HT PagP [200,211]. Interestingly, however, folding of PagP is not fully reversible under conditions which promote reversible folding of HT PagP, again highlighting the difficulty in generating OMPs suitable for equilibrium denaturation studies [92]. Nonetheless, the folding kinetics of PagP into both zwitterionic and negatively charged liposomes were investigated in the presence and absence of Skp or SurA (Fig. 10) [92]. These experiments showed that membrane composition and ionic

strength of the buffer strongly influences the effect that Skp has on PagP folding, suggesting that electrostatic interactions play an important role in the mechanism of action of this chaperone [92], consistent with the previous results on Skp-mediated folding of OmpA [92,182]. SurA, however, did not affect the observed folding rates of PagP, in contrast with the results observed for OmpT refolding [135], but consistent with the view that Skp and SurA may act by distinct mechanisms in partially redundant chaperone pathways [78,105]. The ability of Skp to prevent the aggregation of HT PagP was also investigated, revealing that even under conditions in which aggregation is strongly favoured, Skp can rescue the folding and membrane insertion of HT PagP [92]. Together, these studies indicate the power of combining different methods to study the folding mechanism of an OMP and set the scope for future investigations into how OMPs fold both unassisted and assisted by folding factors *in vitro* and *in vivo*.

### Summary and outlook

Early experiments on the folding of water-soluble proteins focused on small, single domain proteins as these provide relatively simple folding models [3,11,13,14]. In the intervening decades, many details on the folding of these proteins have been elucidated as new experimental methods and computational techniques have



**Fig. 10.** Schematic of the effects of Skp and SurA on the refolding of PagP. SurA and PagP do not interact stably under the conditions of the refolding assay. Skp readily interacts with PagP, retarding the PagP folding rate into zwitterionic liposomes, but accelerating the folding rate of PagP into negatively charged liposomes in a manner dependent on the ionic strength of the buffer. Additionally, the holdase activity of Skp was demonstrated by its ability to rescue the folding and membrane insertion of HT PagP under conditions which strongly favour aggregation of this construct. This figure was adapted from [92] with permission from Elsevier, © 2013.

been developed [31]. More complicated folding scenarios are beginning to be investigated experimentally, including studies of multidomain proteins, oligomeric proteins, folding in the presence of chaperones and in the cellular milieu [40,41,139,219]. Even fifty years after Anfinsen's experiments, the first simulations of protein folding in all-atom detail have only been reported in the last two years and these initial studies were limited to small (<100 residues), single domain proteins [30,220,221]. More recently, the characterisation of the folding and assembly of a dimeric protein in atomistic detail using equilibrium MD simulations has paved the way for simulation studies of the folding of oligomeric proteins [222]. Application of simulation techniques to study integral membrane proteins firstly requires the protein under study to be correctly positioned within the bilayer as this information is not obtained when membrane protein structures are determined by X-ray crystallography or NMR [223]. This is often achieved by first applying a coarse-grained approach to correctly position the protein within the membrane and using these results to guide atomistic simulations to obtain detailed information on lipid-protein interactions and protein dynamics [223]. The multiscale simulation approach has been successful in probing the lipid interactions of aquaporins, ion channels and G-protein coupled receptors, among others, but simulating the complexities of lipid organisation within membranes remains a challenge in this field [223,224].

Similar to the evolution of studies of water-soluble protein folding, analysis of OMP folding mechanisms, thus far, have focused on a small subset of relatively simple proteins that are amenable to the array of kinetic and thermodynamic assays required to determine folding mechanisms [58,159]. Recent trends, however, suggest that the field of membrane protein folding is already progressing towards more complex folding systems. Folding models utilised thus far include OmpA and Omp85 (a BamA homologue) both of which have soluble, periplasmic domains, the trimeric porin, OmpF and the large, 19-stranded hVDAC [108,200,203,225]. Moreover, the use of periplasmic chaperones to aid folding and membrane insertion has now been applied to several different folding systems [92,108,135,182].

Membrane protein folding is complicated by the need to recreate a suitable membrane mimetic environment *in vitro* into which the protein can fold [47]. While early studies concentrated on simple micellar systems, these often do not provide a good mimic for the complex, heterogeneous environment of a membrane *in vivo* and can lead to inactivation or aggregation of the protein of interest [226]. The use of lipid-based mimics, usually synthetic liposomes, has provided a wealth of information based on the ability to introduce guest lipids to modulate surface charge, lateral pressure and membrane fluidity and to examine their effect on OMP folding and stability. Studies to date have focused on relatively simple lipid mixtures, often with short acyl chains ( $\sim C_{12}$ ), as these have been shown to promote the folding of OMPs *in vitro*, and at lipid-to-protein molar ratios (often 400:1 to 3200:1, or up to approximately 100:1 by weight) which far exceed those encountered in biological membranes (in the range 1:1 to 1:3 by weight [227]). While much has been learned about the interactions between OMPs and their surrounding lipid environment from these studies, these simple mimics do not accurately represent the folding situation *in vivo*. Indeed, the membrane environment of OMPs *in vivo* is characterised by variable lipid composition and asymmetric distributions of lipids, as well as being densely packed with the many proteins associated with, and inserted into, the bilayer [63].

As the experimental toolbox for the folding of OMPs continues to expand, the level of complexity attainable in *in vitro* models is likely to increase. Additionally, as the gap in understanding between the folding of OMPs and water-soluble proteins continues to decrease, the similarities and differences between the folding mechanisms of these two protein classes will emerge. This

knowledge will be critical in understanding some of the fundamental biophysical questions which remain to be answered, such as the relationship between the primary sequence of a protein and its native structure in the context of water or lipid as the solvent, and how chaperones assist folding *in vitro* and *in vivo*. Moreover, these concepts will be critical in refining current knowledge of protein folding mechanisms and working towards a universal folding theory which encompasses all of nature's proteins.

## Acknowledgments

We thank G.H.M. Huysmans, A.I. Bartlett, B. Schiffrin and S.A. Baldwin for helpful discussions and many critical insights. We also acknowledge, with grateful thanks, members of the Radford and Brockwell groups as well as our collaborators past and present. L.M.M. is supported by the UK Biotechnology and Biological Sciences Research Council (BB/F01614X/1).

## References

- [1] A. Naeem, N. Fazili, *Cell Biochem. Biophys.* 61 (2011) 237–250.
- [2] M. Vendruscolo, T.P.J. Knowles, C.M. Dobson, *Cold Spring Harb. Perspect. Biol.* 3 (2011) a010454.
- [3] C.B. Anfinsen, E. Haber, M. Sela, F.H. White, *Proc. Natl. Acad. Sci. USA* 47 (1961) 1309–1314.
- [4] C.B. Anfinsen, *Science* 181 (1973) 223–230.
- [5] C. Levinthal, *Univ. Illinois Bull.* 67 (1969) 22–24.
- [6] C. Levinthal, *J. Chim. Phys.* 65 (1968) 44–45.
- [7] T.Y. Tsong, R.L. Baldwin, E.L. Elson, *Proc. Natl. Acad. Sci. USA* 69 (1972) 1809–1812.
- [8] O.B. Ptitsyn, *Dokl. Akad. Nauk SSSR* 210 (1973) 1213–1215.
- [9] P.S. Kim, R.L. Baldwin, *Ann. Rev. Biochem.* 51 (1982) 459–489.
- [10] A.R. Fersht, *Nat. Rev. Mol. Cell Biol.* 9 (2008) 650–654.
- [11] J.F. Brandts, H.R. Halvorson, M. Brennan, *Biochemistry* 14 (1975) 4953–4963.
- [12] L.S. Itzhaki, D.E. Otzen, A.R. Fersht, *J. Mol. Biol.* 254 (1995) 260–288.
- [13] S.E. Jackson, A.R. Fersht, *Biochemistry* 30 (1991) 10428–10435.
- [14] S.E. Jackson, A.R. Fersht, *Biochemistry* 30 (1991) 10436–10443.
- [15] A.R. Fersht, *Curr. Opin. Struct. Biol.* 7 (1997) 3–9.
- [16] S. Gianni, N.R. Guydosh, F. Khan, T.D. Caldas, U. Mayor, G.W.N. White, M.L. DeMarco, V. Daggett, A.R. Fersht, *Proc. Natl. Acad. Sci. USA* 100 (2003) 13286–13291.
- [17] J.N. Onuchic, P.G. Wolynes, *Curr. Opin. Struct. Biol.* 14 (2004) 70–75.
- [18] J.D. Bryngelson, P.G. Wolynes, *Proc. Natl. Acad. Sci. USA* 84 (1987) 7524–7528.
- [19] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, P.G. Wolynes, *Proteins* 21 (1995) 167–195.
- [20] A. Schug, J.N. Onuchic, *Curr. Opin. Pharmacol.* 10 (2010) 709–714.
- [21] L.J. Smith, K.M. Fiebig, H. Schwalbe, C.M. Dobson, *Fold Des.* 1 (1996) R95–R106.
- [22] C.L. Pashley, G.J. Morgan, A.P. Kalverda, G.S. Thompson, C. Kleantsh, S.E. Radford, *J. Mol. Biol.* 416 (2012) 300–318.
- [23] V. Ozenne, J. Noel, P.O. Heidarsson, S. Brander, F.M. Poulsen, M.R. Jensen, B.B. Kragelund, M. Blackledge, J. Danielsson, *J. Mol. Biol.* 426 (2014) 722–734.
- [24] B. Anil, Y. Li, J.-H. Cho, D.P. Raleigh, *Biochemistry* 45 (2006) 10110–10116.
- [25] Y. Li, F. Picart, D.P. Raleigh, *J. Mol. Biol.* 349 (2005) 839–846.
- [26] D. Shortle, M.S. Ackerman, *Science* 293 (2001) 487–489.
- [27] C.J. Francis, K. Lindorff-Larsen, R.B. Best, M. Vendruscolo, *Proteins: Struct., Funct., Bioinf.* 65 (2006) 145–152.
- [28] J.A. Marsh, J.D. Forman-Kay, *J. Mol. Biol.* 391 (2009) 359–374.
- [29] B. Schuler, W.A. Eaton, *Curr. Opin. Struct. Biol.* 18 (2008) 16–26.
- [30] K. Lindorff-Larsen, S. Piana, R.O. Dror, D.E. Shaw, *Science* 334 (2011) 517–520.
- [31] A.I. Bartlett, S.E. Radford, *Nat. Struct. Mol. Biol.* 16 (2009) 582–588.
- [32] B. Schuler, H. Hofmann, *Curr. Opin. Struct. Biol.* 23 (2013) 36–47.
- [33] G. Žoldák, M. Rief, *Curr. Opin. Struct. Biol.* 23 (2013) 48–57.
- [34] A.C.M. Ferreon, A.A. Deniz, *Biochim. Biophys. Acta Proteomics* 1814 (2011) 1021–1029.
- [35] D.J. Brockwell, S.E. Radford, *Curr. Opin. Struct. Biol.* 17 (2007) 30–37.
- [36] D.U. Ferreira, J.A. Hegler, E.A. Komives, P.G. Wolynes, *Proc. Natl. Acad. Sci. USA* 104 (2007) 19819–19824.
- [37] B.G. Wensley, S. Batey, F.A.C. Bone, Z.M. Chan, N.R. Tumelty, A. Steward, L.G. Kwa, A. Borgia, J. Clarke, *Nature* 463 (2010) 685–688.
- [38] H. Neuweiler, S. Doose, M. Sauer, *Proc. Natl. Acad. Sci. USA* 102 (2005) 16650–16655.
- [39] M. Tsytlonok, L.S. Itzhaki, *Arch. Biochem. Biophys.* 531 (2013) 14–23.
- [40] F.U. Hartl, M. Hayer-Hartl, *Nat. Struct. Mol. Biol.* 16 (2009) 574–581.
- [41] R.M. Vabulas, S. Raychaudhuri, M. Hayer-Hartl, F.U. Hartl, *Cold Spring Harb. Perspect. Biol.* 2 (2010).
- [42] H.-X. Zhou, *FEBS Lett.* 587 (2013) 1053–1061.
- [43] P.J. Booth, J. Clarke, *Proc. Natl. Acad. Sci. USA* 107 (2010) 3947–3948.
- [44] P.J. Booth, *Curr. Opin. Struct. Biol.* 22 (2012) 469–475.
- [45] A.L. Hopkins, C.R. Groom, *Nat. Rev. Drug Discov.* 1 (2002) 727–730.

- [46] H.J. Nam, J. Jeon, S. Kim, *Bmb Rep.* 42 (2009) 697–704.
- [47] P.J. Booth, P. Curnow, *Curr. Opin. Struct. Biol.* 19 (2009) 8–13.
- [48] J.U. Bowie, *Nature* 438 (2005) 581–589.
- [49] S.J. Kennedy, *J. Membr. Biol.* 42 (1978) 265–279.
- [50] S.H. White, A.S. Ladokhin, S. Jayasinghe, K. Hristova, *J. Biol. Chem.* 276 (2001) 32395–32398.
- [51] R. Henderson, P.N.T. Unwin, *Nature* 257 (1975) 28–32.
- [52] Y. Kimura, D.G. Vassylyev, A. Miyazawa, A. Kidera, M. Matsushima, K. Mitsuoka, K. Murata, T. Hirai, Y. Fujiyoshi, *Nature* 389 (1997) 206–211.
- [53] B. Schobert, J. Cupp-Vickery, V. Hornak, S. Smith, J. Lanyi, *J. Mol. Biol.* 321 (2002) 715–726.
- [54] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *Nucleic Acids Res.* 28 (2000) 235–242.
- [55] W.C. Wimley, *Curr. Opin. Struct. Biol.* 13 (2003) 404–411.
- [56] J.W. Fairman, N. Noinaj, S.K. Buchanan, *Curr. Opin. Struct. Biol.* 21 (2011) 523–531.
- [57] T. Haltia, E. Freire, *Biochim. Biophys. Acta* 1228 (1995) 1–27.
- [58] D.E. Otzen, K.K. Andersen, *Arch. Biochem. Biophys.* 531 (2013) 34–43.
- [59] S.H. White, <<http://blanco.biomol.uci.edu/mpstruc/listAll/list>>, 2013 (accessed 03.12.13).
- [60] D.E. Warschawski, <<http://www.drorlist.com/nmr/MPNMR.html>>, 2013 (accessed 03.12.13).
- [61] P. Raman, V. Cherezov, M. Caffrey, *Cell. Mol. Life Sci.* 63 (2006) 36–51.
- [62] I. Maslennikov, S. Choe, *Curr. Opin. Struct. Biol.* 23 (2013) 555–562.
- [63] G. van Meer, D.R. Voelker, G.W. Feigenson, *Nat. Rev. Mol. Cell Biol.* 9 (2008) 112–124.
- [64] M.P. Bos, V. Robert, J. Tommassen, *Annu. Rev. Microbiol.* 61 (2007) 191–214.
- [65] N. Ruiz, D. Kahne, T.J. Silhavy, *Nat. Rev. Microbiol.* 4 (2006) 57–66.
- [66] D.L. Diedrich, E.H. Cota-Robles, *J. Bacteriol.* 119 (1974) 1006–1018.
- [67] D. Missiakas, J.M. Betton, S. Raina, *Mol. Microbiol.* 21 (1996) 871–884.
- [68] N. Ruiz, D. Kahne, T.J. Silhavy, *Nat. Rev. Microbiol.* 7 (2009) 677–683.
- [69] C.R.H. Raetz, C. Whitfield, *Annu. Rev. Biochem.* 71 (2002) 635–700.
- [70] C. Dong, K. Beis, J. Nesper, A.L. Brunkan-LaMontagne, B.R. Clarke, C. Whitfield, J.H. Naismith, *Nature* 444 (2006) 226–229.
- [71] A.J.M. Driessen, N. Nouwen, *Annu. Rev. Biochem.* 77 (2008) 643–667.
- [72] T.A. Rapoport, *Nature* 450 (2007) 663–669.
- [73] J.E. Mogensen, D.E. Otzen, *Mol. Microbiol.* 57 (2005) 326–346.
- [74] J. Sohn, R.A. Grant, R.T. Sauer, *Structure* 17 (2009) 1411–1421.
- [75] A. Tormo, M. Almiron, R. Kolter, *J. Bacteriol.* 172 (1990) 4339–4347.
- [76] S.W. Lazar, R. Kolter, *J. Bacteriol.* 178 (1996) 1770–1773.
- [77] P.E. Rouviere, C.A. Gross, *Genes Dev.* 10 (1996) 3170–3182.
- [78] M. Merdanovic, T. Clausen, M. Kaiser, R. Huber, M. Ehrmann, *Annu. Rev. Microbiol.* 65 (2011) 149–168.
- [79] D. Vertommen, N. Ruiz, P. Leverrier, T.J. Silhavy, J.-F. Collet, *Proteomics* 9 (2009) 2432–2443.
- [80] E. Bitto, D.B. McKay, *Structure* 10 (2002) 1489–1498.
- [81] S. Behrens, R. Maier, H. de Cock, F.X. Schmid, C.A. Gross, *EMBO J.* 20 (2001) 285–294.
- [82] A.-J. Geitner, E. Varga, M. Wehmer, F.X. Schmid, *J. Mol. Biol.* 425 (2013) 4089–4098.
- [83] D.P. Ricci, J. Schwalm, M. Gonzales-Cope, T.J. Silhavy, *mBio* 4 (2013). e00540-00513.
- [84] G. Hennecke, J. Nolte, R. Volkmer-Engert, J. Schneider-Mergener, S. Behrens, *J. Biol. Chem.* 280 (2005) 23540–23548.
- [85] E. Bitto, D.B. McKay, *J. Biol. Chem.* 278 (2003) 49316–49322.
- [86] S. Behrens-Kneip, *Int. J. Med. Microbiol.* 300 (2010) 421–428.
- [87] X. Xu, S. Wang, Y.-X. Hu, D.B. McKay, *J. Mol. Biol.* 373 (2007) 367–381.
- [88] B.M. Burmann, S. Hiller, *Chimia* 66 (2012) 759–763.
- [89] E. Bitto, D.B. McKay, *FEBS Lett.* 568 (2004) 94–98.
- [90] B.M. Burmann, C. Wang, S. Hiller, *Nat. Struct. Mol. Biol.* 20 (2013) 1265–1272.
- [91] S. Wu, X. Ge, Z. Lv, Z. Zhi, Z. Chang, X.S. Zhao, *Biochem. J.* 438 (2011) 505–511.
- [92] L.M. McMorran, A.I. Bartlett, G.H.M. Huysmans, S.E. Radford, D.J. Brockwell, *J. Mol. Biol.* 425 (2013) 3178–3191.
- [93] J.G. Sklar, T. Wu, D. Kahne, T.J. Silhavy, *Genes Dev.* 21 (2007) 2473–2484.
- [94] A.E. Rizzitello, J.R. Harper, T.J. Silhavy, *J. Bacteriol.* 183 (2001) 6794–6800.
- [95] R. Chen, U. Henning, *Mol. Microbiol.* 19 (1996) 1287–1294.
- [96] K. Denoncin, J. Schwalm, D. Vertommen, T.J. Silhavy, J.-F. Collet, *Proteomics* 12 (2012) 1391–1401.
- [97] U. Schafer, K. Beck, M. Muller, *J. Biol. Chem.* 274 (1999) 24567–24574.
- [98] J. Schwalm, T.F. Mahoney, G.R. Soltes, T.J. Silhavy, *J. Bacteriol.* 195 (2013) 3734–3742.
- [99] J. Qu, C. Mayer, S. Behrens, O. Holst, J.H. Kleinschmidt, *J. Mol. Biol.* 374 (2007) 91–105.
- [100] T.A. Walton, C.M. Sandoval, C.A. Fowler, A. Pardi, M.C. Sousa, *Proc. Natl. Acad. Sci. USA* 106 (2009) 1772–1777.
- [101] N. Harms, G. Koningstein, W. Dontje, M. Muller, B. Oudega, J. Luirink, H. de Cock, *J. Biol. Chem.* 276 (2001) 18804–18811.
- [102] Z.-X. Lyu, Q. Shao, Y.Q. Gao, X.S. Zhao, *PLoS One* 7 (2012) e46068.
- [103] I.P. Korndorfer, M.K. Dommel, A. Skerra, *Nat. Struct. Mol. Biol.* 11 (2004) 1015–1020.
- [104] T.A. Walton, M.C. Sousa, *Mol. Cell* 15 (2004) 367–374.
- [105] C.L. Hagan, T.J. Silhavy, D. Kahne, *Annu. Rev. Biochem.* 80 (2011) 189–210.
- [106] K.C. Entzinger, C. Chang, R.O. Myhre, K.C. McCallum, J.A. Maynard, *Biochemistry* 51 (2012) 4822–4834.
- [107] R. Horst, E.B. Bertelsen, J. Fiaux, G. Wider, A.L. Horwich, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* 102 (2005) 12748–12753.
- [108] C.P. Moon, N.R. Zaccai, P.J. Fleming, D. Gessmann, K.G. Fleming, *Proc. Natl. Acad. Sci. USA* 110 (2013) 4285–4290.
- [109] F.A. Saul, J.P. Arie, B. Vulliez-le Normand, R. Kahn, J.M. Betton, G.A. Bentley, *J. Mol. Biol.* 335 (2004) 595–608.
- [110] T. Krojer, M. Garrido-Franco, R. Huber, M. Ehrmann, T. Clausen, *Nature* 416 (2002) 455–459.
- [111] T. Krojer, J. Sawa, E. Schafer, H.R. Saibil, M. Ehrmann, T. Clausen, *Nature* 453 (2008) 885–890.
- [112] J. Sawa, A. Heuck, M. Ehrmann, T. Clausen, *Curr. Opin. Struct. Biol.* 20 (2010) 253–258.
- [113] S. Kim, R.T. Sauer, *Proc. Natl. Acad. Sci. USA* 109 (2012) 7263–7268.
- [114] S. Qian, P. Koldewey, T. Tapley, N. Kirsch, K.M. Ruane, J. Pfizenmaier, R. Shi, S. Hofmann, L. Foit, G. Ren, U. Jakob, Z. Xu, M. Cygler, J.C.A. Bardwell, *Nat. Struct. Mol. Biol.* 18 (2011) 262–269.
- [115] E.B. Volokhina, J. Grijpstra, M. Stork, I. Schilders, J. Tommassen, M.P. Bos, *J. Bacteriol.* 193 (2011) 1612–1621.
- [116] K.H. Kim, S. Aulakh, M. Paetzel, *Protein Sci.* 21 (2012) 751–768.
- [117] R. Voulhoux, M.P. Bos, J. Geurtsen, M. Mols, J. Tommassen, *Science* 299 (2003) 253–258.
- [118] R. Voulhoux, J. Tommassen, *Res. Microbiol.* 155 (2004) 129–135.
- [119] S. Kim, J.C. Malinverni, P. Sliz, T.J. Silhavy, S.C. Harrison, D. Kahne, *Science* 317 (2007) 961–964.
- [120] N. Noinaj, A.J. Kuszak, J.C. Gumbart, P. Lukacik, H. Chang, N.C. Easley, T. Lithgow, S.K. Buchanan, *Nature* 501 (2013) 385–390.
- [121] V. Robert, E.B. Volokhina, F. Senf, M.P. Bos, P. Van Gelder, J. Tommassen, *PLoS Biol.* 4 (2006) 1984–1995.
- [122] C.M. Sandoval, S.L. Baker, K. Jansen, S.I. Metzner, M.C. Sousa, *J. Mol. Biol.* 409 (2011) 348–357.
- [123] D.P. Ricci, C.L. Hagan, D. Kahne, T.J. Silhavy, *Proc. Natl. Acad. Sci. USA* 109 (2012) 3487–3491.
- [124] N.W. Rigel, D.P. Ricci, T.J. Silhavy, *Proc. Natl. Acad. Sci. USA* 110 (2013) 5151–5156.
- [125] K.H. Kim, S. Aulakh, M. Paetzel, *J. Biol. Chem.* 286 (2011) 39116–39121.
- [126] C.T. Webb, J. Selkrig, A.J. Perry, N. Noinaj, S.K. Buchanan, T. Lithgow, *J. Mol. Biol.* 422 (2012) 545–555.
- [127] T.J. Knowles, D.F. Browning, M. Jeeves, R. Maderbocus, S. Rajesh, P. Sridhar, E. Manoli, D. Emery, U. Sommer, A. Spencer, D.L. Leyton, D. Squire, R.R. Chaudhuri, M.R. Viant, A.F. Cunningham, I.R. Henderson, M. Overduin, *EMBO Rep.* 12 (2011) 123–128.
- [128] N.W. Rigel, J. Schwalm, D.P. Ricci, T.J. Silhavy, *J. Bacteriol.* 194 (2012) 1002–1008.
- [129] A. Heuck, A. Schleiffer, T. Clausen, *J. Mol. Biol.* 406 (2011) 659–666.
- [130] K.H. Kim, M. Paetzel, *J. Mol. Biol.* 406 (2011) 667–678.
- [131] N. Noinaj, J.W. Fairman, S.K. Buchanan, *J. Mol. Biol.* 407 (2011) 248–260.
- [132] C. Dong, X. Yang, H.-F. Hou, Y.-Q. Shen, Y.-H. Dong, *Acta Crystallogr. D Biol. Crystallogr.* 68 (2012) 1134–1139.
- [133] A.R. Ureta, R.G. Endres, N.S. Wingreen, T.J. Silhavy, *J. Bacteriol.* 189 (2007) 446–454.
- [134] C.L. Hagan, D.B. Westwood, D. Kahne, *Biochemistry* 52 (2013) 6108–6113.
- [135] C.L. Hagan, S. Kim, D. Kahne, *Science* 328 (2010) 890–892.
- [136] C.L. Hagan, D. Kahne, *Biochemistry* 50 (2011) 7444–7446.
- [137] B. Clantin, A.-S. Delattre, P. Rucktooa, N. Saint, A.C. Méli, C. Loch, F. Jacob-Dubuisson, V. Villeret, *Science* 317 (2007) 957–961.
- [138] F. Gruss, F. Zähringer, R.P. Jakob, B.M. Burmann, S. Hiller, T. Maier, *Nat. Struct. Mol. Biol.* 20 (2013) 1318–1320.
- [139] F.U. Hartl, *Nat. Med.* 17 (2011) 1206–1210.
- [140] Y.E. Kim, M.S. Hipp, A. Bracher, M. Hayer-Hartl, F. Ulrich Hartl, *Annu. Rev. Biochem.* 82 (2013) 323–355.
- [141] P.L. Clark, *J. Mol. Biol.* 406 (2011) 543–544.
- [142] K.B. Heller, *J. Bacteriol.* 134 (1978) 1181–1183.
- [143] J.H. Kleinschmidt, T. den Blaauwen, A.J.M. Driessen, L.K. Tamm, *Biochemistry* 38 (1999) 5006–5016.
- [144] J.H. Kleinschmidt, L.K. Tamm, *Biochemistry* 38 (1999) 4996–5005.
- [145] S.E. Radford, C.M. Dobson, P.A. Evans, *Nature* 358 (1992) 302–307.
- [146] C.S. Le Duff, S.B.M. Whittaker, S.E. Radford, G.R. Moore, *J. Mol. Biol.* 364 (2006) 824–835.
- [147] H. Tafer, S. Hiller, C. Hilty, C. Fernández, K. Wüthrich, *Biochemistry* 43 (2004) 860–869.
- [148] S. Hiller, G. Wider, L.L. Imbach, K. Wüthrich, *Angew. Chem. Int. Ed.* 47 (2008) 977–981.
- [149] T. Raschle, S. Hiller, M. Etzorn, G. Wagner, *Curr. Opin. Struct. Biol.* 20 (2010) 471–479.
- [150] J.M. Kielec, K.G. Valentine, A.J. Wand, *Biochim. Biophys. Acta Biomembr.* 1798 (2010) 150–160.
- [151] J.M. Glück, M. Wittlich, S. Feuerstein, S. Hoffmann, D. Willbold, B.W. Koenig, *J. Am. Chem. Soc.* 131 (2009) 12060–12061.
- [152] A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, Macmillan, 1999.
- [153] R.L. Baldwin, *Fold Des.* 1 (1996) R1–R8.
- [154] T. Ternström, U. Mayor, M. Akke, M. Oliveberg, *Proc. Natl. Acad. Sci. USA* 96 (1999) 14854–14859.
- [155] M. Silow, M. Oliveberg, *Proc. Natl. Acad. Sci. USA* 94 (1997) 6084–6086.
- [156] K.K. Andersen, H. Wang, D.E. Otzen, *Biochemistry* 51 (2012) 8371–8383.
- [157] G.H.M. Huysmans, S.A. Baldwin, D.J. Brockwell, S.E. Radford, *Proc. Natl. Acad. Sci. USA* 107 (2010) 4099–4104.
- [158] P. Curnow, P.J. Booth, *Proc. Natl. Acad. Sci. USA* 106 (2009) 773–778.

- [159] P. Curnow, N.D. Di Bartolo, K.M. Moreton, O.O. Ajoje, N.P. Saggese, P.J. Booth, *Proc. Natl. Acad. Sci. USA* 108 (2011) 14133–14138.
- [160] D.E. Otzen, *Protein Eng. Des. Sel.* 24 (2011) 139–149.
- [161] A.R. Fersht, S. Sato, *Proc. Natl. Acad. Sci. USA* 101 (2004) 7976–7981.
- [162] A. Matouschek, J.T. Kellis, L. Serrano, A.R. Fersht, *Nature* 340 (1989) 122–126.
- [163] A.R. Fersht, A. Matouschek, L. Serrano, *J. Mol. Biol.* 224 (1992) 771–782.
- [164] C.T. Friel, D.A. Smith, M. Vendruscolo, J. Gsponer, S.E. Radford, *Nat. Struct. Mol. Biol.* 16 (2009) 318–324.
- [165] C.P. Moon, S. Kwon, K.G. Fleming, *J. Mol. Biol.* 413 (2011) 484–494.
- [166] M. Schweizer, I. Hindennach, W. Garten, U. Henning, *Eur. J. Biochem.* 82 (1978) 211–217.
- [167] T. Surrey, F. Jähnig, *Proc. Natl. Acad. Sci. USA* 89 (1992) 7457–7461.
- [168] T. Surrey, F. Jähnig, *J. Biol. Chem.* 270 (1995) 28199–28203.
- [169] J.H. Kleinschmidt, L.K. Tamm, *Biochemistry* 35 (1996) 12993–13000.
- [170] J.H. Kleinschmidt, L.K. Tamm, *J. Mol. Biol.* 324 (2002) 319–330.
- [171] J.H. Kleinschmidt, *Chem. Phys. Lipids* 141 (2006) 30–47.
- [172] H.D. Hong, L.K. Tamm, *Proc. Natl. Acad. Sci. USA* 101 (2004) 4065–4070.
- [173] C.L. Pocanschi, J.-L. Popot, J.H. Kleinschmidt, *Eur. Biophys. J.* 42 (2013) 103–118.
- [174] A.M. Stanley, K.G. Fleming, *Arch. Biochem. Biophys.* 469 (2008) 46–66.
- [175] C.L. Pocanschi, G.J. Patel, D. Marsh, J.H. Kleinschmidt, *Biophys. J.* 91 (2006) L75–L77.
- [176] H. Hong, G. Szabo, L.K. Tamm, *Nat. Chem. Biol.* 2 (2006) 627–635.
- [177] H. Hong, S. Park, R.H. Flores Jiménez, D. Rinehart, L.K. Tamm, *J. Am. Chem. Soc.* 129 (2007) 8320–8327.
- [178] J.A. Killian, G. von Heijne, *Trends Biochem. Sci.* 25 (2000) 429–434.
- [179] W.C. Wimley, S.H. White, *Nat. Struct. Mol. Biol.* 3 (1996) 842–848.
- [180] P.V. Bulieris, S. Behrens, O. Holst, J.H. Kleinschmidt, *J. Biol. Chem.* 278 (2003) 9092–9099.
- [181] G.J. Patel, J.H. Kleinschmidt, *Biochemistry* 52 (2013) 3974–3986.
- [182] G.J. Patel, S. Behrens-Kneip, O. Holst, J.H. Kleinschmidt, *Biochemistry* 48 (2009) 10235–10245.
- [183] A.A. Nickson, J. Clarke, *Methods* 52 (2010) 38–50.
- [184] A.A. Nickson, B.G. Wensley, J. Clarke, *Curr. Opin. Struct. Biol.* 23 (2013) 66–74.
- [185] W. Banachewicz, T.L. Religa, R.D. Schaeffer, V. Daggett, A.R. Fersht, *Proc. Natl. Acad. Sci. USA* 108 (2011) 5596–5601.
- [186] B.G. Wensley, M. Gartner, W.X. Choo, S. Batey, J. Clarke, *J. Mol. Biol.* 390 (2009) 1074–1085.
- [187] B.G. Wensley, L.G. Kwa, S.L. Shammass, J.M. Rogers, S. Browning, Z. Yang, J. Clarke, *Proc. Natl. Acad. Sci. USA* 109 (2012) 17795–17799.
- [188] B.G. Wensley, L.G. Kwa, S.L. Shammass, J.M. Rogers, J. Clarke, *J. Mol. Biol.* 423 (2012) 273–283.
- [189] K.A. Scott, S. Batey, K.A. Hooton, J. Clarke, *J. Mol. Biol.* 344 (2004) 195–205.
- [190] A.P. Capaldi, C. Kleanthous, S.E. Radford, *Nat. Struct. Mol. Biol.* 9 (2002) 209–216.
- [191] C.T. Friel, A.P. Capaldi, S.E. Radford, *J. Mol. Biol.* 326 (2003) 293–305.
- [192] C.T. Friel, G.S. Beddard, S.E. Radford, *J. Mol. Biol.* 342 (2004) 261–273.
- [193] M.O. Lindberg, E. Haglund, I.A. Hubner, E.I. Shakhnovich, M. Oliveberg, *Proc. Natl. Acad. Sci. USA* 103 (2006) 4083–4088.
- [194] M. Olofsson, S. Hansson, L. Hedberg, D.T. Logan, M. Oliveberg, *J. Mol. Biol.* 365 (2007) 237–248.
- [195] S.J. Hamill, A. Steward, J. Clarke, *J. Mol. Biol.* 297 (2000) 165–178.
- [196] E. Cota, A. Steward, S.B. Fowler, J. Clarke, *J. Mol. Biol.* 305 (2001) 1185–1194.
- [197] C.F. Wright, K. Lindorff-Larsen, L.G. Randles, J. Clarke, *Nat. Struct. Mol. Biol.* 10 (2003) 658–662.
- [198] C.D. Geierhaas, E. Paci, M. Vendruscolo, J. Clarke, *J. Mol. Biol.* 343 (2004) 1111–1123.
- [199] I. Lappalainen, M.G. Hurley, J. Clarke, *J. Mol. Biol.* 375 (2008) 547–559.
- [200] N.K. Burgess, T.P. Dao, A.M. Stanley, K.G. Fleming, *J. Biol. Chem.* 283 (2008) 26748–26758.
- [201] S.R. Maurya, D. Chaturvedi, R. Mahalakshmi, *Sci. Rep.* 3 (2013).
- [202] C.L. Pocanschi, H.J. Apell, P. Puntervoll, B. Hogh, H.B. Jensen, W. Welte, J.H. Kleinschmidt, *J. Mol. Biol.* 355 (2006) 548–561.
- [203] A.E. Tan, N.K. Burgess, D.S. DeAndrade, J.D. Marold, K.G. Fleming, *Macromol. Biosci.* 10 (2010) 763–767.
- [204] E.J. Danoff, K.G. Fleming, *Biophys. Chem.* 159 (2011) 194–204.
- [205] P.M. Hwang, W.-Y. Choy, E.I. Lo, L. Chen, J.D. Forman-Kay, C.R.H. Raetz, G.G. Privé, R.E. Bishop, L.E. Kay, *Proc. Natl. Acad. Sci. USA* 99 (2002) 13560–13565.
- [206] V.E. Ahn, E.I. Lo, C.K. Engel, L. Chen, P.M. Hwang, L.E. Kay, R.E. Bishop, G.G. Prive, *EMBO J.* 23 (2004) 2931–2941.
- [207] J.A. Cuesta-Seijo, C. Neale, M.A. Khan, J. Mokhtar, C.D. Tran, R.E. Bishop, R. Pomès, G.G. Privé, *Structure* 18 (2010) 1210–1219.
- [208] K. Cox, M.S.P. Sansom, *Mol. Membr. Biol.* 26 (2009) 205–214.
- [209] G.E. Schulz, *Biochim. Biophys. Acta Biomembr.* 1565 (2002) 308–317.
- [210] R.E. Bishop, *Mol. Microbiol.* 57 (2005) 900–912.
- [211] S.G.H.M. Huysmans, S.E. Radford, D.J. Brockwell, S.A. Baldwin, *J. Mol. Biol.* 373 (2007) 529–540.
- [212] M.A. Khan, C. Neale, C. Michaux, R. Pomes, G.G. Prive, R.W. Woody, R.E. Bishop, *Biochemistry* 46 (2007) 4565–4579.
- [213] G.H.M. Huysmans, Ph.D. thesis, University of Leeds, UK, 2008.
- [214] G.H.M. Huysmans, S.E. Radford, S.A. Baldwin, D.J. Brockwell, *J. Mol. Biol.* 416 (2012) 453–464.
- [215] A.R. Lam, J.M. Borreguero, F. Ding, N.V. Dokholyan, S.V. Buldyrev, H.E. Stanley, E. Shakhnovich, *J. Mol. Biol.* 373 (2007) 1348–1360.
- [216] S. Gianni, C. Travaglini-Allocatelli, F. Cutruzzola, M. Brunori, M.C. Shastri, H. Roder, *J. Mol. Biol.* 330 (2003) 1145–1152.
- [217] C.F. Wright, A. Steward, J. Clarke, *J. Mol. Biol.* 338 (2004) 445–451.
- [218] A.L. Mallam, S.E. Jackson, *J. Mol. Biol.* 359 (2006) 1420–1436.
- [219] E. Braselmann, J.L. Chaney, P.L. Clark, *Trends Biochem. Sci.* 38 (2013) 337–344.
- [220] S. Piana, K. Lindorff-Larsen, D.E. Shaw, *Proc. Natl. Acad. Sci. USA* 110 (2013) 5915–5920.
- [221] S. Piana, K. Lindorff-Larsen, D.E. Shaw, *Proc. Natl. Acad. Sci. USA* 109 (2012) 17845–17850.
- [222] S. Piana, K. Lindorff-Larsen, D.E. Shaw, *J. Phys. Chem. B* 117 (2013) 12935–12942.
- [223] P.J. Stansfeld, M.S.P. Sansom, *Structure* 19 (2011) 1562–1572.
- [224] P.J. Stansfeld, E.E. Jefferys, M.S.P. Sansom, *Structure* 21 (2013) 810–819.
- [225] B. Shanmugavadivu, H.J. Apell, T. Meins, K. Zeth, J.H. Kleinschmidt, *J. Mol. Biol.* 368 (2007) 66–78.
- [226] J.-L. Popot, *Annu. Rev. Biochem.* 79 (2010) 737–775.
- [227] G. Guidotti, *Arch. Intern. Med.* 129 (1972) 194–201.
- [228] K.R. MacKenzie, J.H. Prestegard, D.M. Engelman, *Science* 276 (1997) 131–133.
- [229] H. Luecke, B. Schobert, H.-T. Richter, J.-P. Cartailleur, J.K. Lanyi, *J. Mol. Biol.* 291 (1999) 899–911.
- [230] C. Toyoshima, H. Nomura, *Nature* 418 (2002) 605–611.
- [231] A. Pautsch, G.E. Schulz, *Nat. Struct. Mol. Biol.* 5 (1998) 1013–1017.
- [232] L. Vandeputte-Rutten, R.A. Kramer, J. Kroon, N. Dekker, M.R. Egmond, P. Gros, *EMBO J.* 20 (2001) 5033–5039.
- [233] H.J. Snijder, I. Ubarretxena-Belandia, M. Blaauw, K.H. Kalk, H.M. Verheij, M.R. Egmond, N. Dekker, B.W. Dijkstra, *Nature* 401 (1999) 717–721.
- [234] E. Yamashita, M.V. Zhalmnina, S.D. Zakharov, O. Sharma, W.A. Cramer, *EMBO J.* 27 (2008) 2171–2180.
- [235] H. Remaut, C. Tang, N.S. Henderson, J.S. Pinkner, T. Wang, S.J. Hultgren, D.G. Thanassi, G. Waksman, H. Li, *Cell* 133 (2008) 640–652.
- [236] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, *J. Comput. Chem.* 25 (2004) 1605–1612.
- [237] K.H. Kim, H.S. Kang, M. Okon, E. Escobar-Cabrera, L.P. McIntosh, M. Paetzl, *Biochemistry* 50 (2011) 1081–1090.
- [238] J.H. Kleinschmidt, P.V. Bulieris, J. Qu, M. Dogterom, T. den Blaauwen, *J. Mol. Biol.* 407 (2011) 316–332.
- [239] R.E. Bishop, H.S. Gibbons, T. Guina, M.S. Trent, S.I. Miller, C.R. Raetz, *EMBO J.* 19 (2000) 5071–5080.
- [240] K.T. Sapra, M. Damaghi, S. Köster, Ö. Yildiz, W. Kühlbrandt, D.J. Muller, *Angew. Chem. Int. Ed.* 48 (2009) 8306–8308.
- [241] M. Damaghi, C. Bippes, S. Köster, Ö. Yildiz, S.A. Mari, W. Kühlbrandt, D.J. Muller, *J. Mol. Biol.* 397 (2010) 878–882.
- [242] S.A. Mari, S. Köster, C.A. Bippes, Ö. Yildiz, W. Kühlbrandt, D.J. Muller, *J. Mol. Biol.* 396 (2010) 610–616.
- [243] M. Damaghi, S. Köster, C.A. Bippes, Ö. Yildiz, D.J. Müller, *Angew. Chem. Int. Ed.* 50 (2011) 7422–7424.
- [244] T. Surrey, A. Schmid, F. Jähnig, *Biochemistry* 35 (1996) 2283–2288.