An evolutionarily conserved glycine-tyrosine motif forms a folding core in outer membrane proteins

Marcin Michalik[1,2], Marcella Orwick-Rydmark[1], Michael Habeck[3,4], Vikram Alva[5], Thomas Arnold[2,6], Dirk Linke[1,2]*

1 Department of Biosciences, University of Oslo, Oslo, Norway, **2** Previous affiliation: Department of Protein Evolution, Max Planck Institute for Developmental Biology, Tübingen, Germany, **3** Statistical inverse problems in Biophysics, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany, **4** Felix Bernstein Institute for Mathematical Statistics in the Biosciences, University of Göttingen, Göttingen, Germany, **5** Department of Protein Evolution, Max Planck Institute for Developmental Biology, Tübingen, Germany, **6** Boehringer Ingelheim Veterinary Research Center, Hannover, Germany

* dirk.linke@ibv.uio.no

Abstract

An intimate interaction between a pair of amino acids, a tyrosine and glycine on neighboring β-strands, has been previously reported to be important for the structural stability of auto-transporters. Here, we show that the conservation of this interacting pair extends to nearly all major families of outer membrane β-barrel proteins, which are thought to have originated through duplication events involving an ancestral ββ hairpin. We analyzed the function of this motif using the prototypical outer membrane protein OmpX. Stopped-flow fluorescence shows that two folding processes occur in the millisecond time regime, the rates of which are reduced in the tyrosine mutant. Folding assays further demonstrate a reduction in the yield of folded protein for the mutant compared to the wild-type, as well as a reduction in thermal stability. Taken together, our data support the idea of an evolutionarily conserved 'folding core' that affects the folding, membrane insertion, and thermal stability of outer membrane protein β-barrels.

Introduction

Outer membranes (OM) of Gram-negative bacteria and of mitochondria and chloroplasts that evolved from prokaryotic organisms [1,2] mainly contain transmembrane proteins with a β-barrel structure, called outer membrane proteins (OMPs). Such β-barrel membrane proteins are not found in other membrane systems, presumably because they form a cylindrical pore that allows for the passage of ions by diffusion even in the smallest of examples [3–5]. Known β-barrel structure sizes range from 8 to 26 β-strands, and the strand number–in prokaryotes at least–is always a multiple of 2, with both N- and C-termini residing in the periplasm. Such OMPs have very diverse functions, including active and passive transport of nutrients and protein export (as part of various secretion systems), and are classified as enzymes, receptors,

structural proteins, or adhesins [6]. Their specific function is often based on their pore structure, or on their additional periplasmic or surface-exposed domains.

All OMPs are evolutionarily related and presumably evolved by duplication events of β-hairpin repeats [4,7]. This evolutionary relationship and sequence features, resulting from functional as well as structural constraints, have been exploited for the prediction of OMP β-barrels from sequence alone. There are two main features: an alternating pattern of hydrophilic and hydrophobic residues indicative of transmembrane β-strands, and a band of aromatic residues (tryptophan, phenylalanine and tyrosine) which faces the lipid-water-interface region in both leaflets of the membrane [8].

Membrane insertion and folding of OMPs *in vivo* is a much-studied but poorly understood process. OMPs, like all other bacterial proteins, are synthesized by ribosomes in the cytoplasm. Their sequences include an N-terminal signal peptide that is recognized by the Sec translocation system which transports the unfolded chain through the inner membrane (IM) [9,10]. After the signal peptide has been cleaved by the signal peptidase, the protein is bound by periplasmic chaperones (*SurA*, *Skp*, *DegP*), to protect the unfolded state against aggregation, misfolding, and proteolytic cleavage [11]. As there is no external source of energy available in the periplasm in the form of ATP or a proton gradient, the process of trafficking an OMP across the periplasm must be exergonic, and the energy may come from the folding of transported products [12]. The folding of OMPs into the OM is mediated by the β-barrel-assembly machinery (BAM) complex, which *in E. coli* consists of the BamA protein (itself an OMP) and four additional lipoproteins. The exact mechanism by which OMPs are inserted and folded into the OM remains unclear, although high-resolution structures and many *in vitro* and *in vivo* studies are available [13–16]. The most popular hypothesis currently suggests that the BAM complex initially recognizes the unfolded chaperone-bound OMP, which is subsequently threaded through the pore of BamA where insertion of β-strands into the membrane occurs through the lateral opening of BamA strands 1 and 16 [14,17,18]. It is worth noting however, that at least small OMPs are able to spontaneously insert into lipid membranes *in vitro* without help by any other factors [19], suggesting that the BAM complex only facilitates a process whose driving force is already encoded into the OMP sequence.

Beyond their general sequence features as described above, OMPs include specific sequence motifs for their function, *i.e.* a C-terminal motif that plays a role in BAM recognition [20,21]. In autotransporters, another recently recognized motif includes a conserved glycine and tyrosine pair, where the glycine can also be substituted for a threonine, and the tyrosine for a phenylalanine or tryptophan. This pair is present on neighboring β-strands and forms a clamp inside the barrel. This motif, described as a mortise-tenon motif, has been suggested to provide structural stability to autotransporters by aligning and locking the neighboring β-strands [22]. A similarly conserved tyrosine-glycine pair had been found earlier in alignments of 8-stranded OMPs [23], but its possible function in protein structure or folding was not discussed further.

Using bioinformatics tools, we assessed the evolutionary conservation of this motif amongst OMPs. Further, to understand its functional role, we used a combination of mutagenesis and a variety of biophysical tools to study the folding kinetics and structural stability of the prototypical 8-stranded β-barrel bacterial OMP, OmpX [24]. The tyrosine-glycine motif in OmpX involves residues Y80 and G112 (Fig 1). The aromatic ring of Y80 is oriented towards the lumen of the barrel and is located over G112 on the neighboring strand, the small size of which allows for Y80 to position in a planar manner directly over it (Fig 1), forming an apparent mortise-tenon motif analogous to that described in autotransporters [22]. The biological function of OmpX is not fully understood, but it is believed to play a role in neutralizing host defense mechanisms [24]. Previous studies have shown that it is a very robust protein that can still fold *in vivo* and *in vitro*
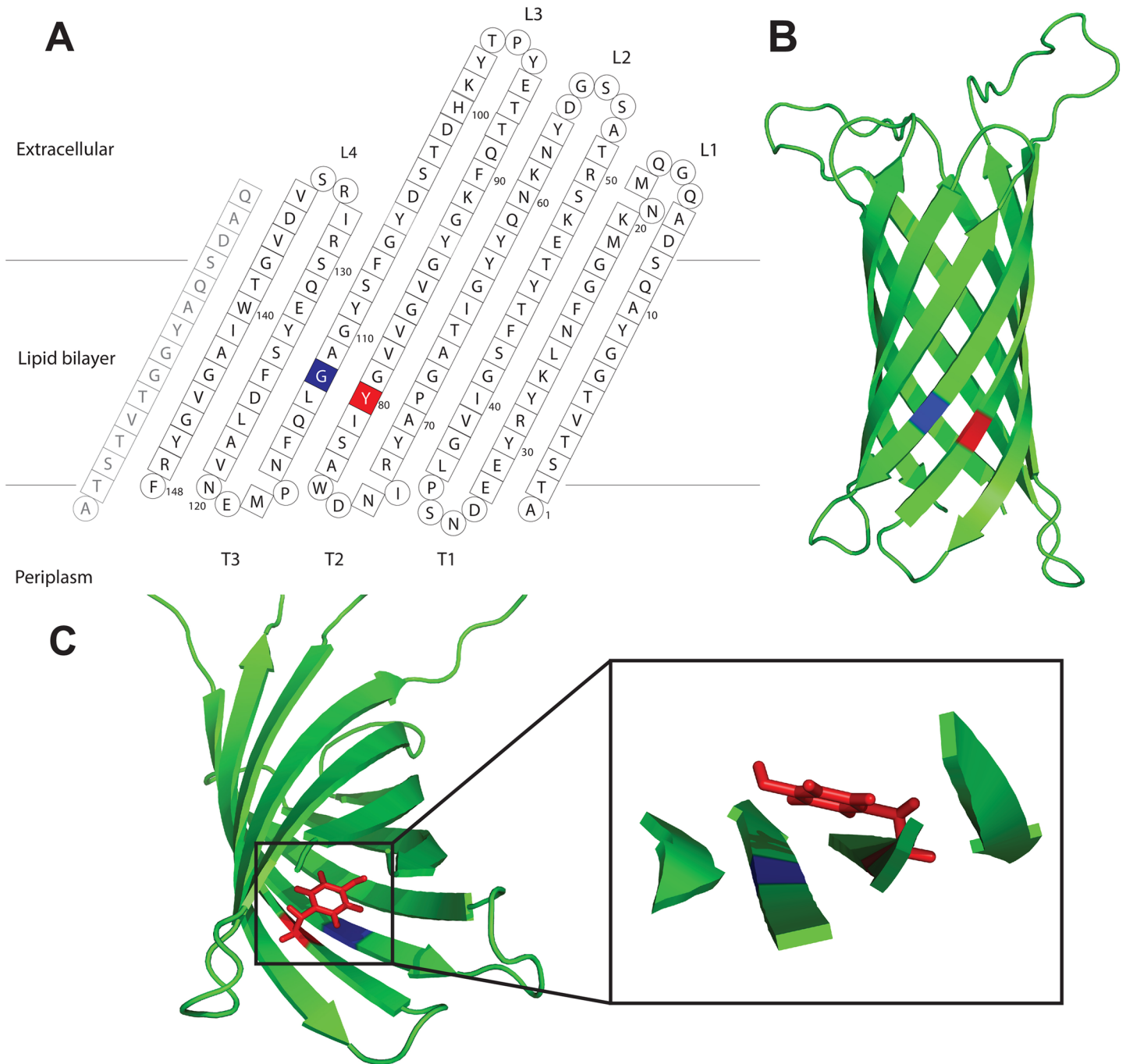
Fig 1. OmpX structure. (A) The topology model of OmpX with the highlighted mortise-tenon motif in red (Y80) and blue (G112) on the 5th and 6th β-strands. Residues forming the β-strands are represented as squares, while loop residues are shown as circles (adapted from [24]). **(B)** The OmpX solution NMR structure in 3D representation using PyMOL [PDB code 2MNH]. **(C)** The mortise-tenon motif involves residues Y80 (red) and G112 (blue). The tyrosine's aromatic side chain is positioned in a planar manner over the glycine in the neighboring strand (G112).

https://doi.org/10.1371/journal.pone.0182016.g001

after the introduction of mutations [4,25] and that it can easily be refolded from inclusion bodies into detergent micelles or liposomes [26,27].

Results

Bioinformatics

To investigate how widespread mortise-tenon motifs are across all OMPs, we decided to search for them in a comprehensive set representative of all known OMP types. For this, we chose the Structural Classification of Proteins (SCOPe) database [28], which classifies proteins of known structure into hierarchies based on evolutionary and structural criteria. SCOPe classifies OMPs into 7 superfamilies, comprising a total of 15 families. Using sequence and structural analysis, we searched for mortise-tenon motifs in these 15 families, as described in the Methods section. We detected highly conserved mortise-tenon motifs in 13 of these families (Table 1), showing that they not restricted to only autotransporters and eight-stranded barrels, but that they are in fact more widespread among OMPs. We analyzed these results in more detail for the three major classes of OMPs with multiple exemplars of known structure: the 8-stranded, 16-stranded, and 22-stranded OMPs.

For 8-stranded β-barrels, which are grouped into a superfamily (SCOPe f.4.1) with 4 families in SCOPe, there is one motif located on the 5th and 6th β-strand which in most cases consists of a glycine-tyrosine pair (Fig 2). While this motif is highly conserved in three of these families, one family (SCOPe f.4.1.3) contains no widely conserved motifs. In 16-stranded β-barrels (SCOPe family f.4.3.1), we found multiple mortise-tenon pairs, with the most conserved pairs located on strands 1 and 2, 3 and 4, and 15 and 16 (Fig 3). Here, the aromatic residue is somewhat less conserved compared to 8-stranded β-barrels, with more examples of phenylalanine and tryptophan replacing tyrosine (Fig 3A). Finally, in 22-stranded-β- barrels (SCOPe family f.4.3.3), we also found multiple mortise-tenon pairs, with the most conserved ones located between β-strands 9 and 10 and 15 and 16 (**Fig 4**). Similar to the 16-β-stranded barrels, the tyrosine can be substituted for a phenylalanine or tryptophan (**Fig 4A**).

Table 1. Summary of results into the conservation of the mortise-tenon motif across all families of OMPs as classified by the SCOPe database. The mortise-tenon motif is found to be conserved in 13 of the 15 OMP families.

SCOPe ID	SCOP Family	PFAM ID	PFAM Family	Strand-number	Mortise-Tenon motif location
f.4.1.1	Outer membrane protein	PF01389	OmpA_membrane	8	OmpA; 1QJP; Y94 –G126
f.4.1.2	Outer membrane enzyme PagP	PF07017	PagP	8	PagL; 2ERV; F78 –G111
f.4.1.3	GNA1870 immunodominant domain-like	PF01298	TbpB_B_D	8	no widely conserved motifs
f.4.1.4	PsbO-like	PF01716	MSP	8	PsbO; 3WU2, chain o; F146—G195
f.4.2.1	Outer membrane phospholipase A (OMPLA)	PF02253	PLA1	12	OMPLA; 1QD5; G139 –Y159,
f.4.3.1	Porin	PF00267	Porin_1	16	OmpK36; 1OSM; Y14 –G44, Y58 –G87, Y310 –G335
f.4.3.2	Maltoporin-like	PF02264	LamB	18	LamB; 1AF6; F366 –G414, W101 –G128
f.4.3.3	Ligand-gated protein channel	PF00593	TonB_dep_Rec	22	ButB; 1NQE; G314 –Y341; Y436 –G463
f.4.3.4	Outer membrane protein transport protein	PF03349	Toluene_X	14	none
f.4.4.1	Outer membrane protease OMPT	PF01278	Omptin	10	OMPT; 1I78; G133 –Y180, Y248 –G292
f.4.4.2	Outer membrane adhesin/invasin OpcA	PF07239	OpcA	10	OpcA; 2VDF; Y154 –G198, W158—G194, F212 –G246
f.4.5.1	Autotransporter	PF03797	Autotransporter	12	EspP; 2QOM; G1081 –Y1108
f.4.6.1	Tsx-like channel	PF03502	Channel_Tsx	12	tsx; 1TLW; G126 –Y144
f.4.7.1	OccD-like	PF03573	OprD	18	OpdC; 3SY9; G159 –W193, G195 –Y209
f.4.7.2	OccK-like	PF03573	OprD	18	OpdF; 3SZD; G153 –F184, G186 –W201

A

```
OmpX         91  GPAYRIN---DWASIYGVVGVGYGKFQTTEYPT----------------YKHDTSDYGFSYGAGLQFNPMENVALDFSY  150
HypoProt_1  107  KASYPLAFISDDLDIYARAGAFVHITKWKQPHHS----------------GSSETNVSPVFALGLDYRLNDDFSTRLEY  169
HypoProt_2   96  DAVYKLP----ESDLYLFSGIKRFNNTRS-------------------YNALNVGAGYGFQVNERFSVYTEA  144
OmpA        103  KLGYPIT---DDLDIYTRLGGMVWRADTKSNVYG---------------KNHDTGVSPVFAGGVEYAITPEIATRLEY  162
OmpW        117  MAQWYFGDAQSKVRPYVGAGINYTTFFNEDFNDTGKA--------AGLSDLSLKDSWGAAGQVGLDYLINRDWLLNMSV  187
PagL_1       88  VFVYEFA--GDSIKPFIEAGIGVAAFSGTRVGD----------------QNLGSSLNFEDRIGAGLKFANGQSVGVRA  147
PagL_2       97  IPMFRWW---ANDRFYLEAGIGATVFSSTSFAD----------------KRIGSAFQFGDHVGLGFLLTQNNRIGLRY  155
Ail         102  GPVFRIN---EYVSLYGLLGAGHGKAKVSSIFG----------------QSESRSKTSLAYGAGLQFNPHPNFVIDASY  161
NspA         86  SAIYDFDT-QSPVKPYLGARLSLNRASVDLGGSDS-------------FSQTSTGLGVLAGVSYAVTPNVDLDAGY  147
OprF         96  DAIYHFGTPGVGLRPYVSAGLAHQNITNINSDSQ--------------GRQQMTMANIGAGLKYYFTENFFAKASL  157
OmpP5       111  KGSYEVL---DGLDVYGKAGVALVRSDYKFYEDANGT--------RDHKKGRHTARASGLFAVGAEYAVLPELAVRLEY  178
MotB        121  VTRFGLS---EDFELFAKAGSFYWDGDNQGPYS---------------RNSDSGWAPMLGAGLAYQISPSWVARLEY  179
OmpF         95  NLLYYFGKSGDRLRPYAAFGVGDGRWEFDNT-------------------DWVETDVNLGGGVRYELAKNVSLRADA  152
OmpA-Like   112  KLSYPIT---DDFHIYTQLGGMMFWDNLSFKKDLQN------------IFTEKSRLIPNVSLGAEYIFNKKFITRLDY  174
PagC         95  GPAWRLN---DYISLYALAGLSHSTMERNARDYDYYG--------HPERINFSDRSNAFAYGAGVIFNVTEHVTLNAGY  162
Tail-Prot   102  GPSVRVN---EWFSAYAMAGVAYSRVSTFSGDYIRVTDNKGKTHDVLTGSDDGRHSNTSLAWGAGVQFNPTESVAIDIAY  178
Vir-Prot     97  GPAYRIN---SLISLYAVGGFARTKATGNTQRQMNNN--------SIKRRALYGRETSFAYGAGIIINPRDYLTVNLGY  164
Inv-prot    103  GPTYRAN---KYFSAYALIGFASGTANVRYSGYGYN----------LSGSESDSNIAVGAGVQFDITDRFALDASY  165
Consensus_ss:    eeeee  eeeeeeeeeeeeeeeeee          eeeeeeeeeeeeee   eeeeeee
```
                    5th strand   6th strand

B

NspA   OmpA   OmpX   OprF   PagL
Y81 - G113   Y94 - G126   Y80 - G112   Y87 - G118   F78 - G111
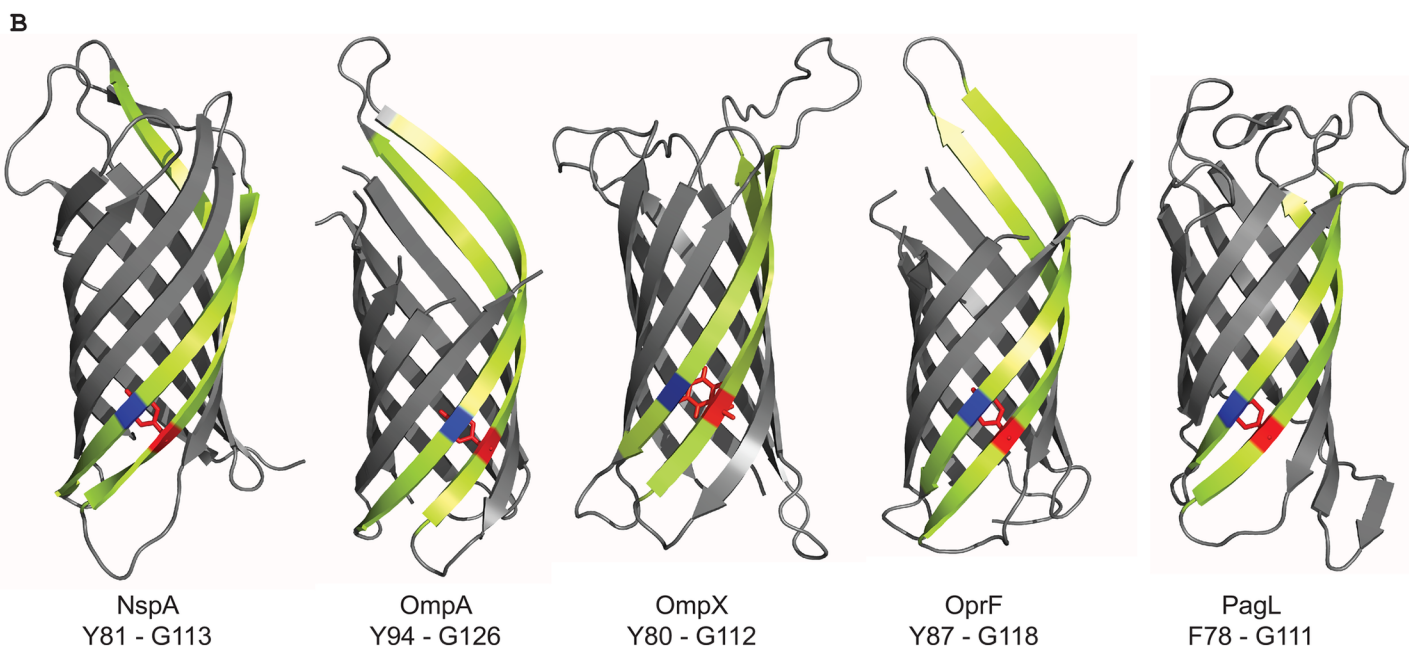
Fig 2. (A) Alignments of the mortise-tenon regions in 8-stranded OMP β-barrels. There is 100% conservation of the glycine (dark blue, strand 6th) and very high conservation of the tyrosine (red, strand 5th) in the mortise-tenon motif. Alignments were prepared using PROMALS3D web server; the predicted β-strands are colored in blue, and the consensus secondary structure prediction is given below the alignment. **OmpX**—APC51088.1 Outer membrane protein X [*Escherichia coli* str. K-12 substr. W3110]; **HypoProt_1**—WP_039105847.1 hypothetical protein [*Frischella perrara*]; **HypoProt_2**—WP_044834085.1 hypothetical protein [*Thalassomonas actiniarum*]; **OmpA**—BAA35715.1 Outer membrane protein A [*Escherichia coli* str. K-12 substr. W3110]; **OmpW**—CDO14140.1 Outer membrane protein W [*Klebsiella pneumoniae*]; **PagL_1**—BAR70092.1 PagL [*Pseudomonas aeruginosa*]; **PagL_2**—AGW82166.1 PagL [*Bordetella avium*]; **Ail**—AAB36601.1 Ail [*Yersinia pseudotuberculosis*]; **NspA**—CAX50548.1 NspA [*Neisseria meningitidis* 8013]; **OprF**—AFM37279.1 OprF [*Pseudomonas aeruginosa*]; **OmpP5**—AAA03346.1 outer membrane protein P5 [*Haemophilus influenzae*]; **MotB**—WP_012154185.1 flagellar motor protein MotB [*Shewanella pealeana*]; **OmpF**—WP_013260126.1 Outer membrane porin F [*gamma proteobacterium* HdN1]; **OmpA**-Like—WP_011053841.1 OmpA-like protein [*Buchnera aphidicola*]; **PagC**—WP_067708083.1 PagC [*Erwinia* sp. ErVv1]; **TailProt**—CSH02213.1 Tail assembly protein [*Shigella sonnei*]; **VirProt**—WP_038626525.1 virulence protein [*Pantoea* sp. PSNIH2]; **InvProt**—SDG65145.1 attachment invasion locus protein [*Vibrio xiamenensis*]. **(B)** Structures of different 8-stranded β barrel proteins. The 5th and 6th β-strands are highlighted in yellow and the mortise-tenon motif is highlighted in blue and red accordingly. PDB codes: NspA [1P4T]; OmpA[1QJP]; OmpX[2MNH]; OprF[4RLC]; PagL[2ERV].

A

```
OmpF          1  -MKRNILAVVIPALLVAGAA-NAAEIYNKDGNKLDLYGKVDARHQFSDNAGQ----------------DGDETYVRFGFKGETQI  67
OmpD          1  -MKLKLVAVAVTSLLAAGAV-NAAEVYNKDANKLDIYGKVHAQHYFSDDTSS----------------DGDKTYARLGFKGETQI  67
MProt_1       1  -MNNSALAIFTAALALTSTA-QAAEVYNKDGNKLDFYGKVKAMHYLSDTASN---------------DGDKSYVRIGFKGQTQI  67
MProt_2       1  -MKRKALAILLPSLLISGAA-GAAEIYNKNGNKVDFYGKMVGEHIMTDTDEGNR-------------NSEDTSYARFGVKGETQI  70
MProt_3       1  -MKRRILSLMIPALLVAGSA-SAAEVYNKDGNKLDLYGLVDGLHYFSSDKGS----------------DGDQSYIRLGFKGETQI  67
MProt_4       1  MMKRSVVALAVVLATAASMA-NAAEIYNKDGNKLDLYGRVNAKHLFSDNKSA---------------DGDATYVRFGFKGETQI  68
MProt_5       1  -MKKSTLAIVVSALALTSAA-QAAEVYNKDGNKLDFYGKVKAEHYFSDNAGA---------------NGDKSYVRIGFKGQTQI  67
Porin_1       1  -MKKSLTALAAL--AVTGLA-SA-------QSSVTLFGVVDAGVSYQSATSRDPLTGATSKQSQWSLANSGYNSSRIGFRGTEDL  74
Porin_2       1  -MKKSLIALAVL--AASGAA-MA-------QSSVTMFGIVDVGVGHVSATNSV----------SGVTQSGNSSSRLGFRGVEDL  63
Porin_3       1  -MKKSLIALAVL--AASGAA-MA-------QSSVTLFGIVDAGVSYVDNADGNS--------SNYGMHNSGNATSRLGLRGTEDL  66
Porin_4       1  MMKRNVLAVVIPALLAAGAA-NAAEVYNKDGNKLDLYGKVEVQHKFANNKSGE---------------DGDNSFARLGFKGETQV  69
Porin_5       1  -MKRNILAVVIPALLVAGAA-NAAELYNKDGNKLSLYGKVDARYKFTNDKSD---------------DGDHTYARFGFKGETQI  67
Porin_6       1  -MKKSLIALAVL--AASGAA-MA-------QSSVTLFGIVDTNVAYVNKDAAGD--------SRYGLGTSGASTSRLGLRGTEDL  66
OmpC_1        1  -MQWWVVLGLIATSTYFPSV-NAAEIYNSDGNKLDLYGRFDAMHWFSGDNGI----------------DGDQSYLRFGFKGETQI  67
OmpC_3        1  -MKVKVLSLLVPALLVAGAA-NAAEIYNKDGNKLDLYGKVDGLHYFSDDDSQ---------------DGDQTYMRLGFKGETQV  67
PhoE_1        1  -MKYNKLAVVVSSLLLVSAGIHAAEIYNKDGNKLDLYGKVDARHQFSDDKGA---------------DGDQTYVRLGFKGETQI  67
PhoE_2        1  MMKRNILAVVIPALLAAGAA-NAAEIYNKDGNKVDIYGKAVGLHYFSKDNGANS--------YGG---NGDQTYARLGFKGETKI  73
PhoE_3        1  -MKRRILAAVIPALLAAATA-NAAEIYNKDGNKLDLYGKAVGRHVWTTTGDSK-------------NADQTYAQIGFKGETQI  68
PhoE_4        1  -MKKSTLALVVMGIVASASI-QAAEIYNKDGNKLDVYGKVKAMHYMSDNASK---------------DGDQSYIRFGFKGETQI  67
Consensus ss:    hhhhhhhhhhhh               eeeeeeeeeeeeeeee         ee  e      eeeeeeeeee
                                                     1st strand                        2nd strand
```

```
OmpF         68  TDQLTGYGQWEYNIQANHAE--SQSTE---GNKTRLGFAGLKFANYGSFDYGRNYGVIYDVNAWTDMLPVFGGDS--------LS  139
OmpD         68  NDQLVGFGQWEYEFKGNRSE--DQGSE---ADKTRLAFAGLKFAEYGSFDYGRNYGVAYDIGAWTDVLPEFGGDT--------WT  139
MProt_1      68  NELMTGYGQWEYHLPASNAE--SDGTA---GTKTRLGFAGLKFGDYGSFDYGRNYGLIYDVEAFTDMMPEFGATG-------YT  139
MProt_2      71  NSDLTGYGRFEYNMKANKPE----GDQ---GSATRLAFAGLKYASYGSFDYGRNYGAVYDAAGYTDMLVEWGGDG--------LV  140
MProt_3      68  TDQLTGYGQWEYQVQANQAE--SEDNR----SFTRYGFAGLKFGDYGSFDYGRNNGVLYDVAAWTDLQPEFDGST---------Y  137
MProt_4      69  NSDLTGYGQWEYNVQANHAE--SAGTE---GTKTRLGFAGLKFSQYGSFDYGRNYGVLYDVESYTDMLPEFGGDT--------YT  140
MProt_5      68  NDMMTGYGQWEYEYKANNSEGGSDFNN----GNKTRLGFAGLKFSQYGSFDYGRNYGVLYDVESLTDVFPEFGGDG--------TA  141
Porin_1      75  GGGLAASFWLEAPISNDDGA----NGI---SSFNRRSTVSLSG-GFGEVRLGRDYTSTFWNDTVFDPFGTNGSGTNVISTVGASS  151
Porin_2      64  GGGLKAGFWLEGAISPDTGTGAGANNAGPGFAFSRRSTVSLAG-NFGEVRLGRELAAGYDKPAISYDPFGQV----------GLA  137
Porin_3      67  GGGLKAGFWLEGAVLNDVGG------G---FDFRRESTVRLSG-NFGEVRLGRETTPTFRAGLKYDLFGATGIGKNQGFSDWAGA  141
Porin_4      70  NDQLTGFGRWEYNIKGNRAE--GTDEE----TKTRLAFAGLKLSNYGSLNYGRNYGVVYDVNAWTDVLPVFGGDS--------MA  140
Porin_5      68  NSELTGYGQWEAEAKAKKPE--SDSGN----FKTRLGFAGLKFADYGSLDYGRNYGVVYDVEAWTDVLPVFGGDT--------YA  138
Porin_6      67  GGGLKAGFWLEGEIFGDDGN----ASG---FNFKRRSTVSLSG-NFGEVRLGRDLVPTSQKLTSYDLFSATGIGPFMGFRNWAAG  143
OmpC_1       68  NTQLIGYGQWEYQANLNRPE---SEDN---KNFTRVGFAGLKLQDYGSIDYGRNYGVLYDVANWTDVMPEFGGNT---------Y  137
OmpC_3       68  NDQLTGYGQWEYQIQGNSGE--NENN-----SWTRVAFAGLKFGDAGSFDYGRNYGVVYDVTSWTDVLPEFGGDT--------Y  136
PhoE_1       69  NDLLTGYGQWEYNMQANNTE--DSGNQ----AWTRLAFAGLKRAGDYGSADFGRNYGVNYDIEAWTDLLPVFGGDS--------YT  139
PhoE_2       74  NDQLTGYGQWEYNFQGNNSEGGTDAQS---GNKTRLAFAGLKFGDFGSFDYGRNYGVVYDVDAIGVTDMLPEFGGDT---------A  146
PhoE_3       69  NTDLTGFGQWEYRTKADRAE--GEQQN---SNLVRLAFAGLKYAEVGSIDYGRNYGIVYDVESYTDMAPYFSGET------WGGA  142
PhoE_4       68  NDQLTGYGRWEAEFAGNKAE--SDTAQ----QKTRLAFAGLKYKDLGSFDYGRNLGALYDVEAWTDMFPEFGGDS--------SA  138
Consensus ss:    eeeeeeeeeee        eeeeeeeeeee     eeeeeeeee   hheeeeeee
                 3rd strand                 4th strand
```

```
OmpF        302  AT-GAPIGDQ-DLVKYVSVGSYYYFNKNMSTYVDYKINLLDKNDFTVANG------------------INTDDVVGVGLVYQF  364
OmpD        303  ------YGDQ-DLVEYVDLGATYYFNKNMSTFVDYKINLLDDNNFTNAAK------------------VSTDNIVAVGMTYQF  360
MProt_1     311  G-----VGDA-DIYKFADVGATYYFNKNMSLFVDYKINLLDMDND---NKLA---------------QNTDDIVAVGVTYQF  366
MProt_2     296  G----GRGNA-EIMNYLELGTWYYFNKNFNVYTAYKFNLLDEKDRAVTG-------------------AATDDQFALGITYQF  354
MProt_3     310  DRLGNNYDDQ-DLVKFVDLGATYYFNKNMSTYVDYKINLLDDNQFTQNAG------------------INTDNVTAVGLVYQF  373
MProt_4     310  ------IGDV-DLVKYIDIGATYYFNKNMYTYVDYKINQLNDD---NKLK-----------------LNTDNIVAASLTYRF  364
MProt_5     318  G-----IGNE-DLVKYIDVGATYYFNKNMSTFVDYKINQLDKD---NKLA-----------------LNNDDVVAVGLTYQF  373
Porin_1     310  I-----TGES-PRVNKFALGYVHNLSKRTALYATVARVNNRNDAAYTGDLTAAST-TGYGSTGVGYTGLPRSSTGYDFGIRHAF  386
Porin_2     281  -----------NDSNALAIGYVVNLSKRTAVYGTYAHMNNKNAAARNLNGVSGSF-NSNVT-----VGAGENVDGYQVGIRHAF  347
Porin_3     289  -----------AKAHQFSLGYVHNLSKRTALYGTVAYLKNKEGSFMALDAKGVSNTVGLKD-----SGAARNQTGVQVGIRHAF  356
Porin_4     326  DD-GKESSSR-DLVKYISVGSYYYFNKNLTTYVDYKINLLDSNEFTKATG-----------------TSTDNVVGVGMIYQF  388
Porin_5     301  ------YGDKVDVVKYVDLGATYYFNKNMSTYVDYKINLLDNNDFTNAAG-----------------ISTDNIVGVGMIYQF  359
Porin_6     288  -----------SKAHQITLGYVHNLSKRTALYGNLAFLKNKDASTLGLQAKGVY----AGG-----VQAGESQTGVQVGIRHAF  351
OmpC_1      312  Y------GDQ-DIYKDINIGVFYYFNSNMYAYADYQINLLHNNNFTRDAH-----------------IATDNMVAVALNYQF  369
OmpC_3      305  G-----YGDQ-DLLKYVDVGATYYFNKNMSTYVDYKINLLDDKEFTRNAG-----------------ISTDDIVALGLVYQF  363
PhoE_1      307  A-----RSSE-DLVKYIEVGSSYYFNKNLATYVDYKINLLDKNDLG------------------IKTDNVVGAGIVYQF  361
PhoE_2      323  G-----IGSE-DYINYVEVGATYYFNKNMSTYVDYIINQIDKD---NKLG-----------------VGSDDTVAVGIVYQF  377
PhoE_3      305  A-----GGSA-DLAKYIQAGATYYFNKNMNVWVDYRFNLLDENDYS-SSY-----------------VGTTDDQAAVGITYQF  362
PhoE_4      297  ------IGDE-DLVNYIDVGATYYFNKNMSAFVDYKINQLDSD---NKLN-----------------INNDDIVAVGMTYQF  351
Consensus ss:    eeeeeeeeeee       eeeeeeeeeeee                     eeeeeeeeee
                 15th strand              16th strand
```

B

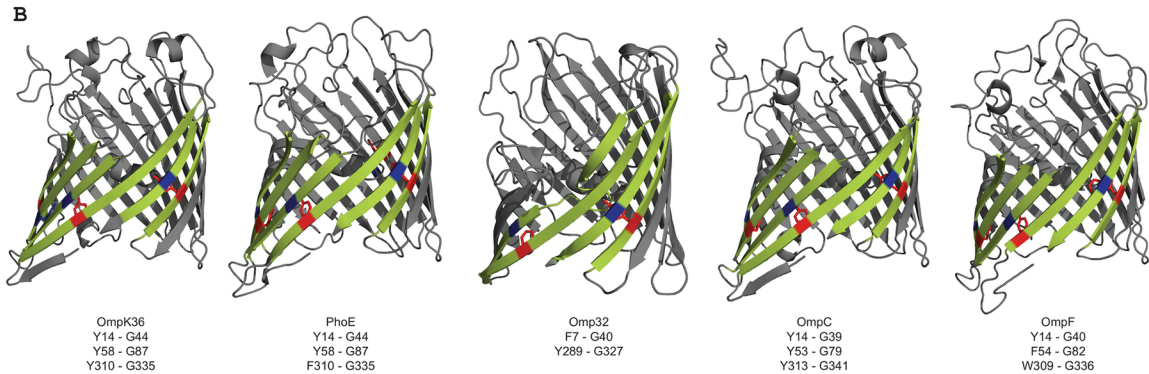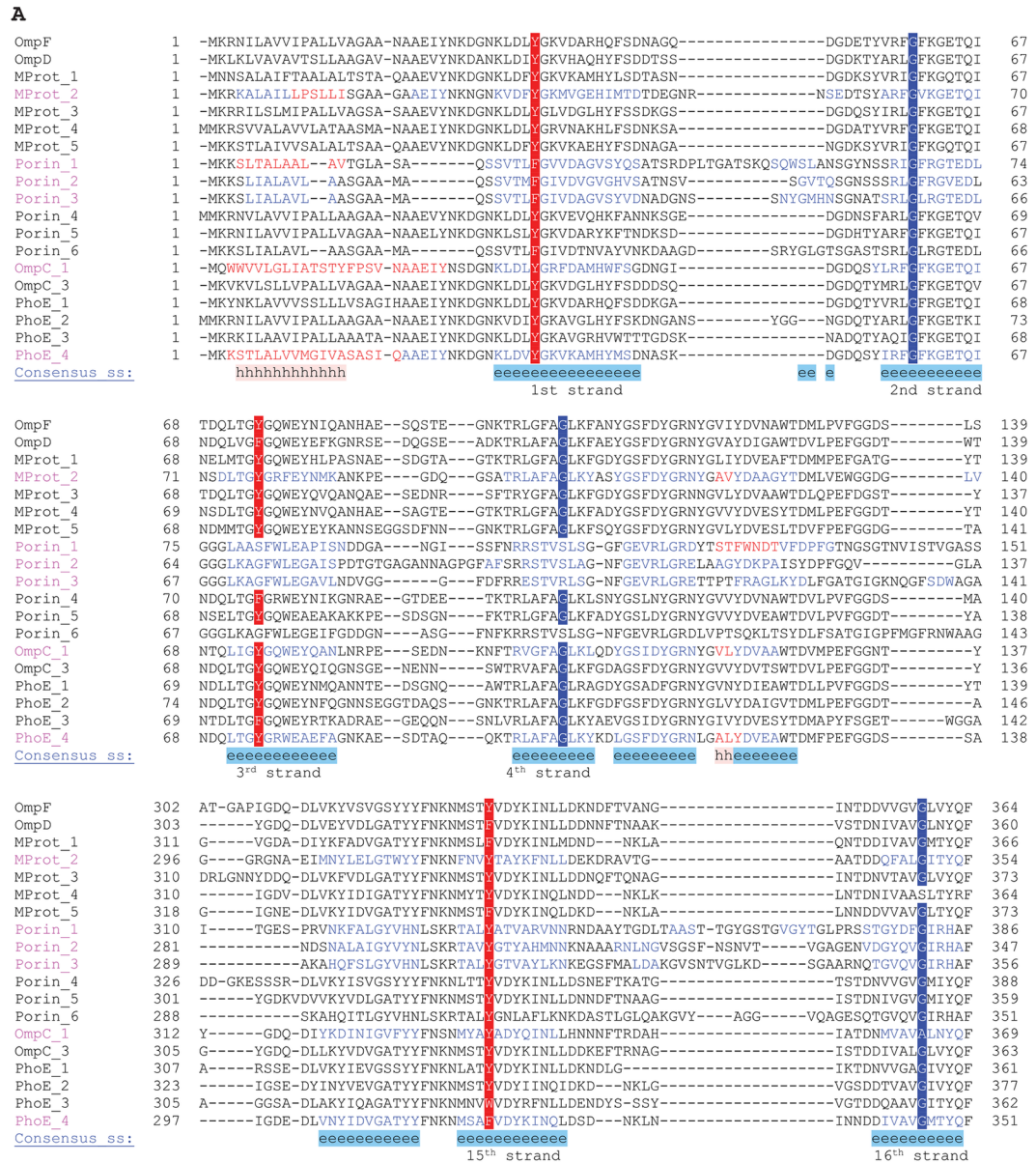|				
OmpK36	PhoE	Omp32	OmpC	OmpF
Y14 - G44	Y14 - G44	F7 - G40	Y14 - G39	Y14 - G40
Y58 - G87	Y58 - G87	Y289 - G327	Y53 - G79	F54 - G82
Y310 - G335	F310 - G335		Y313 - G341	W309 - G336

Fig 3. (A) Alignments of the mortise-tenon motifs in 16-stranded OMP β-barrels. Three conserved motifs are present on the 1st and 2nd β-strands, 3rd and 4th β-strands and on the 15th and 16th β-strands. The color scheme is the same as in Fig 2. **OmpF**—

AJW81160.1 OmpF [*Yersinia ruckeri*]; **OmpD**—GAL44027.1 outer membrane porin OmpD [*Citrobacter werkmanii* NBRC 105721]; **MProt_1**—WP_052899673.1 membrane protein [*Erwinia iniecta*]; **MProt_2**—WP_045859316.1 membrane protein [*Raoultella terrigena*]; **MProt_3**—WP_034791683.1 membrane protein [*Ewingella americana*]; **MProt_4**—WP_024486820.1 membrane protein [*Serratia fonticola*]; **MProt_5**—WP_028724068.1 membrane protein [*Pantoea ananatis*]; **Porin_1**—SCX53975.1 Outer membrane protein (porin) [*Variovorax sp.* EL159]; **Porin_2**—WP_005794534.1 porin [*Acidovorax delafieldii*]; **Porin_3**—WP_003058791.1 porin [*Comamonas testosteroni*]; **Porin_4**—WP_011146030.1 porin [*Photorhabdus luminescens*]; **Porin_5**—WP_035609637.1 porin [*Edwardsiella ictaluri*]; **Porin_6**—WP_034397082.1 porin [*Delftia acidovorans*]; **OmpC_1**—WP_019081949.1 porin OmpC [*Yersinia enterocolitica*]; **OmpC_3**—WP_058683256.1 porin OmpC [*Enterobacter cloacae*]; **PhoE_1**—WP_002439127.1 phosphoporin PhoE [*Shimwellia blattae*]; **PhoE_2**—WP_017456205.1 phosphoporin PhoE [*Kosakonia sacchari*]; **PhoE_3**—WP_072077160.1 phosphoporin PhoE [*Salmonella enterica*]; **PhoE_4**—WP_032085013.1 phosphoporin PhoE [*Escherichia coli*]. **(B)** Structures of different 16-stranded β barrel proteins. The 1st-2nd, 3rd-4th and 15th-16th β-strands are highlighted in yellow and the mortise-tenon motifs are highlighted in blue and red accordingly. The amino-acid pairs forming the motif are given under the names of the proteins. PDB codes: OmpK36 [1OSM]; PhoE [1PHO]; Omp32 [2FGQ]; OmpC [2J1N]; OmpF [3NSG].

The presence of this motif in most families of OMP β–barrels and its extreme positional conservation strongly suggests that it is evolutionarily old, and presumably has a very basic function that is independent of the different transport or enzymatic functions of this broad family of proteins. Our previous studies on OMP evolution using OmpX as a model system showed that OmpX can fold in almost all permutations of added or deleted β-hairpins, as long as the central hairpin is kept [4]. The fact that this hairpin contains the conserved tyrosine residue described here led us to the hypothesis that the mortise-tenon pair presents a conserved 'folding core' for β-barrel proteins.

Folding kinetics

To evaluate the role of the conserved mortise-tenon motif in OMP folding, stopped-flow fluorescence experiments to measure folding rates were carried out on wt OmpX as well as a mutant in which Y80 was mutated to a glycine to remove the aromatic side chain completely (Y80G). OmpX contains two tryptophan residues at positions 76 and 140, and the refolding kinetics were monitored using the change in fluorescence intensity after folding, using a fluorescence detector with an excitation wavelength of 295 nm to avoid contribution from other aromatic residues, and a cutoff filter at 320 nm to capture maximal fluorescence increase without contributions of scattered excitation light. Measuring tryptophan fluorescence intensity over time is very useful in measuring kinetics of OMPs: tryptophan is sensitive to the polarity of the local environment, and during protein folding shows a characteristic red-shift in its emission spectrum. In the case of OmpX this is coupled to an increase in the fluorescence intensity [29–31].

Rapid refolding was carried out by diluting the urea-unfolded OmpX into buffered detergent-rich solution (C8POE) in a stopped-flow setup. The raw data were subjected to multi-exponential fitting procedures using a self-written software that is based on a maximum likelihood approach (see methods section). The refolding trace was fit to single, double, and triple exponential functions to determine the best fit based on standard deviations and chi squared values (Table 2, and Fig 5). For the single exponential fit, a comparatively high chi-square value was found (Fig 5A, Fig 5B), while for the triple exponential fit the standard deviations were high compared to the double exponential fit. We therefore used a double exponential function to fit the refolding trace (Fig 5C). In wt OmpX, the first detected folding event is approximately 10-fold faster than the second folding event. When performing the same experiment with the Y80G mutant, we found the rate of folding to be significantly reduced compared to wt OmpX (Table 2). The first, fast step is reduced to 78% in Y80G OmpX compared to wt OmpX, while the second slower step is reduced to 60% of the rate observed for wt OmpX.
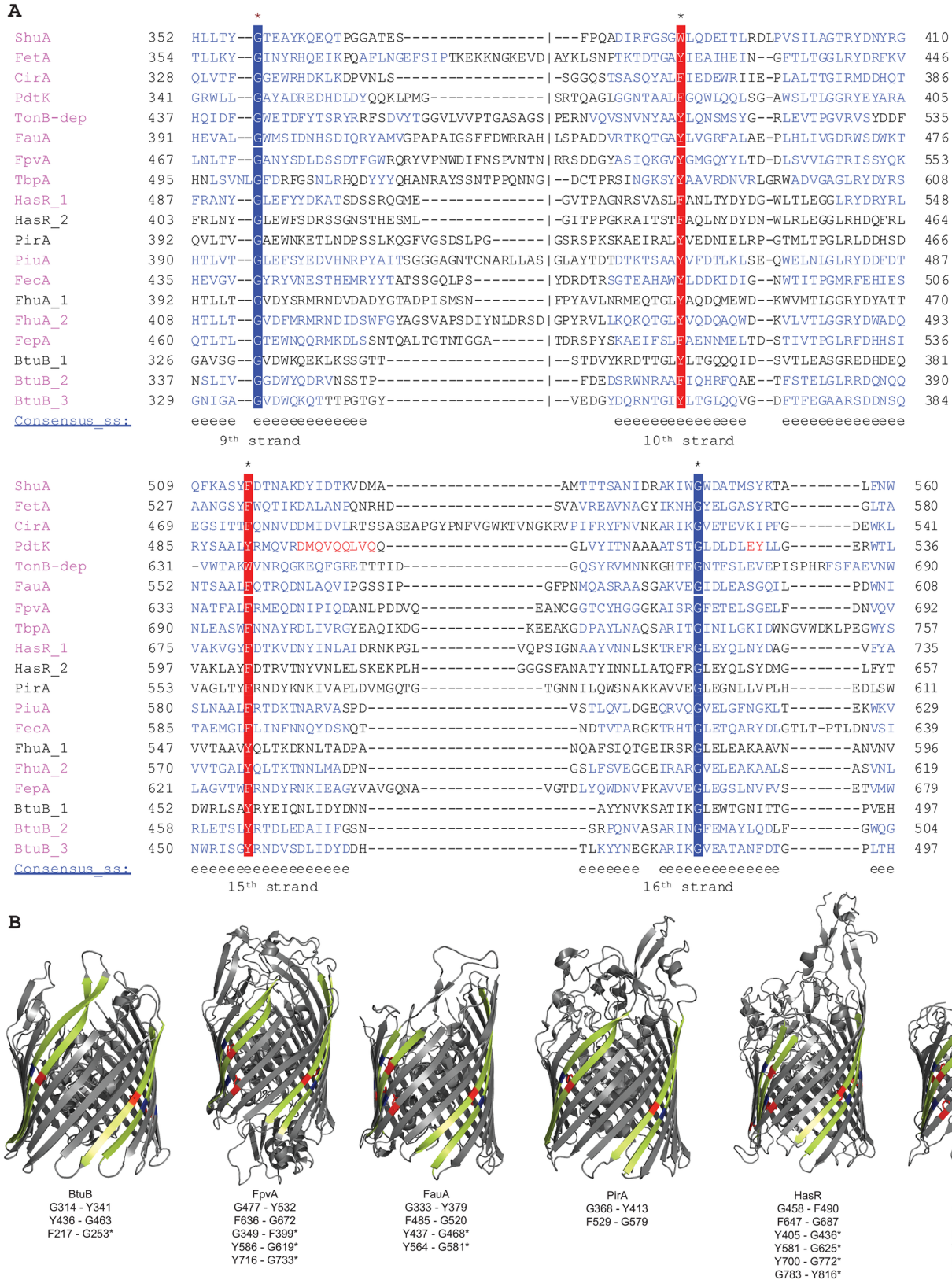
A

```
                      *                                              *
ShuA      352  HLLTY--GTEAYKQEQTPGGATES----------------|---FPQADIRFGSGWLQDEITLRDLPVSILAGTRYDNYRG  410
FetA      354  TLLKY--GINYRHQEIKPQAFLNGEFSIPTKEKKNGKEVD|AYKLSNPTKTDTGAYIEAIHEIN--GFTLTGGLRYDRFKV  446
CirA      328  QLVTF--GGEWRHDKLKDPVNLS---------------|-SGGQSTSASQYALFIEDEWRIIE-PLALTTGIRMDDHQT  386
PdtK      341  GRWLL--GAYADREDHDLDYQQKLPMG------------|SRTQAGLGGNTAALFGQWLQQLSG-AWSLTLGGRYEYARA  405
TonB-dep  437  HQIDF--GWETDFYTSRYRRFSDVYTGGVLVVPTGASAGS|PERNVQVSVNVNYAAYLQNSMSYG--RLEVTPGVRVSYDDF  535
FauA      391  HEVAL--GWMSIDNHSDIQRYAMVGPAPAIGSFFDWRRAH|LSPADDVRTKQTGAYLVGRFALAE-PLHLIVGDRWSDWKT  476
FpvA      467  LNLTF--GANYSDLDSSDTFGWRQRYVPNWDIFNSPVNTN|RRSDDGYASIQKGVYGMGQYYLTD-DLSVVLGTRISSYQK  553
TbpA      495  HNLSVNLGFDRFGSNLRHQDYYYQHANRAYSSNTPPQNNG|--DCTPRSINGKSYYAAVRDNVRLGRWADVGAGLRYDYRS  608
HasR_1    487  FRANY--GLEFYYDKATSDSSRQGME--------------|-GVTPAGNRSVASLFANLTYDYDG-WLTLEGGLRYDRYRL  548
HasR_2    403  FRLNY--GLEWFSDRSSGNSTHESML--------------|-GITPPGKRAITSTFAQLNYDYDN-WLRLEGGLRHDQFRL  464
PirA      392  QVLTV--GAEWNKETLNDPSSLKQGFVGSDSLPG------|GSRSPKSKAEIRALYVEDNIELRP-GTMLTPGLRLDDHSD  466
PiuA      390  HTLVT--GLEFSYEDVHNRPYAITSGGGAGNTCNARLLAS|GLAYTDTDTKTSAAYVFDTLKLSE-QWELNLGLRYDDFDT  487
FecA      435  HEVGV--GYRYVNESTHEMRYYTATSSGQLPS--------|YDRDTRSGTEAHAWYLDDKIDIG--NWTITPGMRFEHIES  506
FhuA_1    392  HTLLT--GVDYSRMRNDVDADYGTADPISMSN--------|FPYAVLNRMEQTGLYAQDQMEWD--KWVMTLGGRYDYATT  470
FhuA_2    408  HTLLT--GVDFMRMRNDIDSWFGYAGSVAPSDIYNLDRSD|GPYRVLLKQKQTGLYVQDQAQWD--KVLVTLGGRYDWADQ  493
FepA      460  QTLTL--GTEWNQQRMKDLSSNTQALTGTNTGGA------|TDRSPYSKAEIFSLFAENNMELTD-STIVTPGLRFDHHSI  536
BtuB_1    326  GAVSG--GVDWKQEKLKSSGTT-----------------|--STDVYKRDTTGLYLTGQQQID--SVTLEASGREDHDEQ  381
BtuB_2    337  NSLIV--GGDWYQDRVNSSTP------------------|---FDEDSRWNRAAFIQHRFQAE--TFSTELGLRRDQNQQ  390
BtuB_3    329  GNIGA--GVDWQKQTTTPGTGY-----------------|--VEDGYDQRNTGIYLTGLQQVG--DFTTFEGAARSDDNSQ  384
Consensus_ss:  eeeee  eeeeeeeeeeee          eeeeeeeeeeeeee   eeeeeeeeeeeee
               9th strand                 10th strand
```

```
                   *                                                *
ShuA      509  QFKASYFDTNAKDYIDTKVDMA-------------------AMTTTSANIDRAKIWGWDATMSYKTA--------LFNW  560
FetA      527  AANGSYFWQTIKDALANPQNRHD----------------SVAVREAVNAGYIKNHGYELGASYRTG--------GLTA  580
CirA      469  EGSITTFQNNVDDMIDVLRTSSASEAPGYPNFVGWKTVNGKRVPIFRYFNVNKARIKGVETEVKIPFG-------DEWKL  541
PdtK      485  RYSAALYRMQVRDMQVQQLVQQ--------------------GLVYITNAAAATSTGLDLDLEYLLG-------ERWTL  536
TonB-dep  631  -VWTAKWVNRQGKEQFGRETTTID------------------GQSYRVMNNKGHTEGNTFSLEVEPISPHRFSFAEVNW  690
FauA      552  NTSAALFQTRQDNLAQVIPGSSIP---------------GFPNMQASRAASGAKVEGIDLEASGQIL-------PDWNI  608
FpvA      633  NATFAIFRMEQDNIPIQDANLPDDVQ-------------EANCGGTCYHGGGKAISRGFETELSGELF-------DNVQV  692
TbpA      690  NLEASWFNNAYRDLIVRGYEAQIKDG------------KEEAKGDPAYLNAQSARITGINILGKIDWNGVWDKLPEGWYS  757
HasR_1    675  VAKVGYFDTKVDNYINLAIDRNKPGL-----------VQPSIGNAAYVNNLSKTRFRGLEYQLNYDAG--------VFYA  735
HasR_2    597  VAKLAYFDTRVTNYVNLELSKEKPLH-----------GGGSFANATYINNLLATQFRGLEYQLSYDMG--------LFYT  657
PirA      553  VAGLTYFRNDYKNKIVAPLDVMGQTG--------------TGNNILQWSNAKKAVVEGLEGNLLVPLH-------EDLSW  611
PiuA      580  SLNAALFRTDKTNARVASPD---------------------VSTLQVLDGEQRVQGVELGFNGKLT-------EKWKV  629
FecA      585  TAEMGIFLINFNNQYDSNQT----------------------NDTVTARGKTRHTGLETQARYDLGTLT-PTLDNVSI  639
FhuA_1    547  VVTAAVYQLTKDKNLTADPA----------------------NQAFSIQTGEIRSRGLELEAKAAVN-------ANVNV  596
FhuA_2    570  VVTGALYQLTKTNNLMADPN-------------------GSLFSVEGGEIRARGVELEAKAALS-------ASVNL  619
FepA      621  LAGVTWFRNDYRNKIEAGYVAVGQNA-------------VGTDLYQWDNVPKAVVEGLEGSLNVPVS-------ETVMW  679
BtuB_1    452  DWRLSAYRYEIQNLIDYDNN--------------------AYYNVKSATIKGLEWTGNITTG--------PVEH  497
BtuB_2    458  RLETSLYRTDLEDAIIFGSN--------------------SRPQNVASARINGFEMAYLQDLF--------GWQG  504
BtuB_3    450  NWRISGYRNDVSDLIDYDDH--------------------TLKYYNEGKARIKGVEATANFDTG--------PLTH  497
Consensus_ss:  eeeeeeeeeeeeeeeee          eeeeeee eeeeeeeeeeeee           eee
               15th strand                16th strand
```

B

BtuB	FpvA	FauA	PirA	HasR	ShuA
G314 - Y341	G477 - Y532	G333 - Y379	G368 - Y413	G458 - F490	G316 - W352
Y436 - G463	F636 - G672	F485 - G520	F529 - G579	F647 - G687	F478 - G509
F217 - G253*	G349 - F399*	Y437 - G468*		Y405 - G436*	F142 - G155*
	Y586 - G619*	Y564 - G581*		Y581 - G625*	Y270 - G294*
	Y716 - G733*			Y700 - G772*	F407 - G456*
				G783 - Y816*	G565 - Y593*

Fig 4. (A) Alignments of the mortise-tenon motifs in 22-stranded OMP β-barrels. Two conserved motifs are present on the 9th and 10th β-strands and on the 15th and 16th β-strands. Compared to 8-stranded proteins, the glycine in the motif is also completely conserved, while the tyrosine in the

motif is less conserved, with the aromatic residues phenylalanine and tryptophan also present at a higher frequency. The color scheme is the same as in Fig 2. **ShuA**—AIN19718.1 outer membrane heme receptor ShuA [*Yersinia kristensenii*]; **FetA**—AHW74795.1 FetA [*Neisseria meningitidis*]; **CirA**—CDO15561.1 cirA [*Klebsiella pneumoniae*]; **PdtK**—ABC68350.1 PdtK [*Pseudomonas putida*]; **TonB-dep**—WP_043878013.1 TonB-dependent receptor [*Pectobacterium atrosepticum*]; **FauA**—AAD26430.1 ferric alcaligin siderophore receptor FauA [*Bordetella pertussis*]; **FpvA**—AOX26958.1 FpvA [*Pseudomonas aeruginosa*]; **TbpA**—AAF81744.1 transferrin-binding protein A [*Neisseria meningitidis*]; **HasR_1**—CAE46936.1 hasR [*Serratia marcescens*]; **HasR_2**—AIN17416.1 hasR protein [*Yersinia kristensenii*]; **PirA**—CFU88752.1 ferric enterobactin receptor PirA [*Pseudomonas aeruginosa*]; **PiuA**—SCM64637.1 PiuA putative outer membrane ferric siderophore receptor [*Pseudomonas aeruginosa*]; **FecA**—AAL08456.2 FecA [*Shigella flexneri* 2a]; **FhuA_1**—CDO15921.1 fhuA [*Klebsiella pneumoniae*]; **FhuA_2**—ANK05474.1 fhuA [*Escherichia coli* O25b:H4]; **FepA**—ADB98042.1 FepA [*Escherichia coli*]; **BtuB_1**—CDO16333.1 btuB [*Klebsiella pneumoniae*]; **BtuB_2**—AEV65209.1 BtuB [*Pseudomonas fluorescens* F113]; **BtuB_3**—EFE98543.1 btuB [*Escherichia coli* FVEC1412]. **(B)** Structures of different 22-stranded OMP β-barrels. The 9th-10th and 15th-16th β-strands are highlighted in yellow and the mortise-tenon motifs are highlighted in blue and red accordingly. The amino acid pairs forming the motif are given under the names of the proteins. Observed mortise-tenon pairs that were at non-conserved positions are marked with an astrix. PDB codes: BtuB [1NQE]; FpvA [2O5P]; FauA [3EFM]; PirA [5FR8]; HasR [3CSN]; ShuA [3FHH].

https://doi.org/10.1371/journal.pone.0182016.g004

The initial folding of OmpX into detergent micelles from urea is likely to include at least three rapid events: Upon urea dilution, OmpX will collapse from an extended form, adsorb to detergent micelles, and simultaneously or shortly thereafter fold and insert β-strands into the detergent micelles. In the later folding stages, loops and a stable, hydrogen-bonded network will form. This final, slow folding event has been followed by NMR and HD exchange on the time-scale of seconds to hours in the detergents DPC and LDAO, but those experiments were not sensitive to the rapid initial folding events measured here [32]. Our stopped-flow experiments are sensitive to the rapid, initial folding events.

It is possible that the two events observed here are directly linked to the environments of the two tryptophan residues present in the structure, suggesting that the Y80G mutation affects different parts of the folding process in slightly different ways, as was observed previously for other membrane proteins with multiple Trp residues, e.g. with the help of single-tryptophan mutants [31]. In the folded state, W140 is located on the membrane- or detergent-exposed hydrophobic surface of the barrel of OmpX on β-strand 8, at the site where the barrel closes by interacting with strand 1, while W76 is located in a more water-accessible periplasmic turn. Y80 is in close proximity to W76 in both sequence and in space in both the folded and unfolded state [33,34]; based on this, we suggest that the faster folding component observed in our experiments is contributed mainly by W76 –as the structural changes between folded and unfolded state are minimal and thus potentially fast. The slower component would then mainly represent the slower process of barrel closure, where W140 is a sensitive indicator as it is located on strand 8 close to the C-terminus (see also Discussion section).

It is interesting to note that in NMR studies on unfolded OmpX in 8M urea, two regions of the protein retained residual structure mediated by a network on hydrogen bonds [34]. These

Table 2. Data from stopped-flow experiments of OmpX refolded into C8POE. Each individual stopped-flow measurement was fitted to a single, double and triple exponential function. The individual rates, amplitudes and chi-square values were then averaged and are shown here with standard deviations for wt and Y80G OmpX. The double exponential function consistently gave better results than single or higher exponential functions and the kinetic constants for the double-exponential fit are shown here (see also Methods section). For the estimation of wt OmpX and Y80G OmpX kinetic parameters, 429 and 438 single experiments were performed, respectively. The raw data of the experiments is provided in S1 File.

	wt OmpX	Y80G OmpX
Average rate 1 (in s^{-1})	8.95 ± 0.40	7.02 ± 0.19
Average amplitude 1	91.7 ± 0.68	91.7 ± 0.42
Average rate 2 (in s^{-1})	0.85 ± 0.24	0.51 ± 0.04
Average amplitude 2	8.3 ± 0.68	8.3 ± 0.42
Sample size	429	438
Avg. chi^2	0,51 ± 0,16	1,67 ± 0,27

https://doi.org/10.1371/journal.pone.0182016.t002
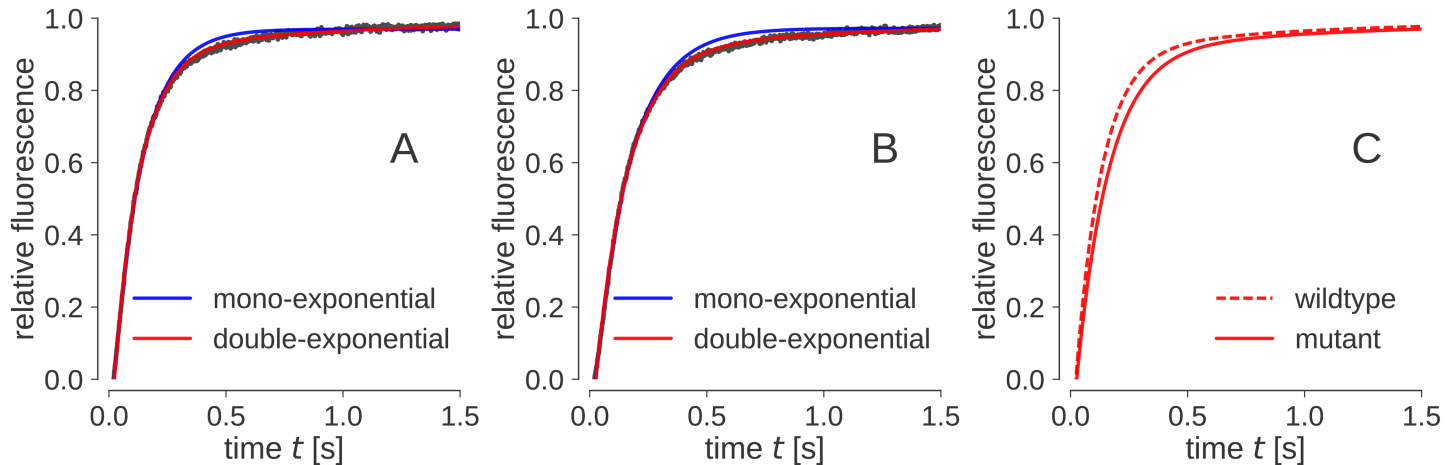
Fig 5. Raw stopped-flow data and exponential fits for wt (panel A) and Y80G OmpX (panel B). The raw data is shown in black, the single-exponential fit in blue, and the double-exponential fit in red. Clearly, a single exponential fit is not sufficient to describe the data. Panel C shows the difference between wt and Y80G OmpX double exponential fits (see also **Table 2**).

two regions coincide with the location of the two Trp residues observed here: one is the periplasmic turn around W76 and intriguingly includes Tyr80 of the mortise-tenon motif; the other is the C-terminal region including W140. The C-terminal region is known to be essential for recognition by the Bam machinery *in vivo* [20,21], but we can only speculate if residual structure of the C-terminal motif is important for membrane insertion. Our *in vitro* folding data suggests that the periplasmic turn region around W76 is a folding core, in line with previous findings that this region is the only part of engineered, larger barrel structures that cannot be deleted [4]. Based on this, we suggest that the interaction between strand 5 and 6 mediated by the mortise-tenon motif of Y80 and G112 is indeed the initial event in 8-stranded barrel
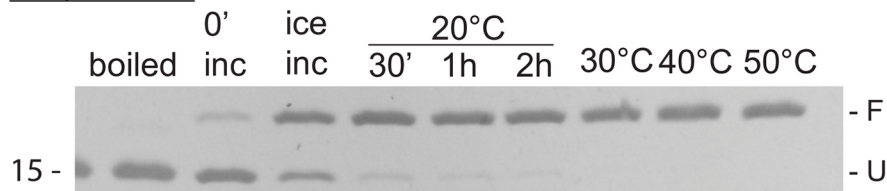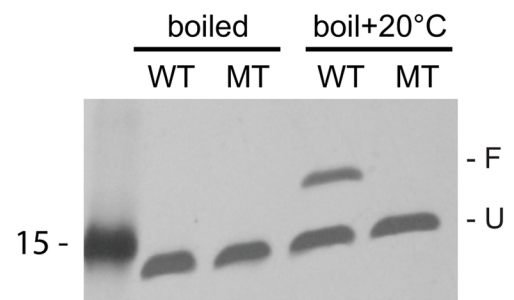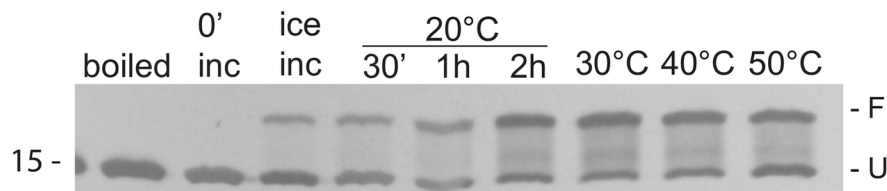
Fig 6. *In vitro* folding experiments for wt (WT) and Y80G OmpX (MT) at different temperatures. Protein samples in urea were diluted 1:20 into detergent-containing buffer, and incubated for 2h on ice, and at 20, 30, 40 and 50˚C. The negative controls (boiled) in the gels are detergent-refolded samples which were boiled excessively at 100˚C, while heated refolded samples (boil+20˚C) are samples that after heating to 100˚C for 1h to denature were shifted to 20˚C for another hour to test for reversible refolding. The folded fraction (F) is represented by the upper band, while the unfolded fraction (U) is the lower band.
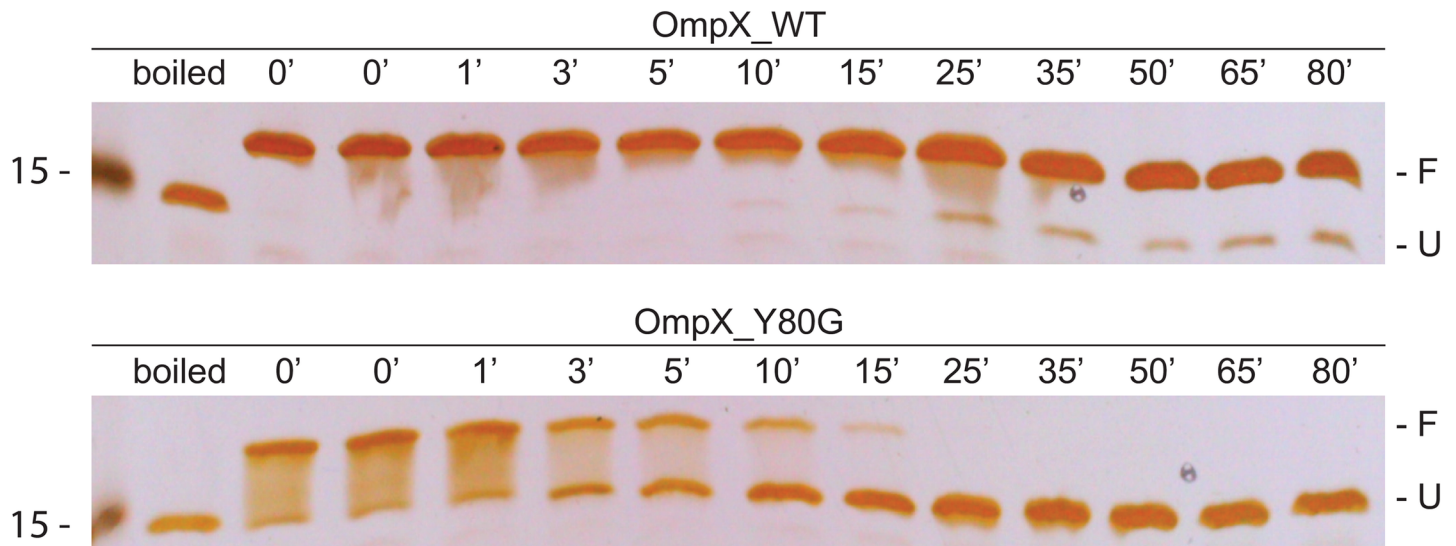
Fig 7. Heat stability for wt and Y80G OmpX monitored by gel shifts. Proteins were refolded *in vitro* prior to incubation at 85°C for the indicated time. The denatured OmpX appears as the lower band (U) on the gel, while folded OmpX is shifted to the upper band (F). An estimation of folded/unfolded yields of proteins can be made by comparing the intensities of the two bands. In the second lane of both gels OmpX samples were boiled at 100°C for 30 min before loading as a control.

formation at least *in vitro*, after collapse of the region from an extended state. We speculate that the motif is specifically important for aligning the strands in the proper register, so that the barrel can close efficiently and without mismatch. The extreme evolutionary conservation of the mortise-tenon motif suggests that it plays an equally important role for initiation of folding events *in vivo*.

Thermodynamics: Folding yield and thermal stability

The *in vitro* folding efficiency for OMPs into lipid bilayers and detergents can vary with temperature [35,36]. In SDS sample buffer, folded OMPs retain their native state, unless they are heated (boiled) or treated with denaturants such as urea; they then migrate on an SDS-PAGE gel differently when compared to unfolded samples, an effect called a gel shift [37,38]; this can be used to assay folding yields and thermodynamic stability. Usually, smaller folded OMPs run at an increased molecular weight compared to their unfolded form, while for larger OMPs the effect is reversed [37,38]. OmpX, with eight β-strands, is considered a small OMP, and accordingly, the migration of the folded state appears at a higher molecular weight than that of the unfolded state [4]. For wt OmpX, the folding efficiency into micelles of the detergent SB-12 was found to be over 90% from ice-cold buffer to 50°C (Fig 6). In contrast, the Y80G mutant folded to only a negligible degree on ice (less than 5%, using identical buffer conditions), while increasing the refolding temperature to RT resulted in ⁓50% folding efficiency (Fig 6, see also Materials and Methods). Despite the increased temperature necessary for folding, the similar gel shift for folded wt and Y80G OmpX does suggest that the final folded states of both wt and mutant are the same in terms of secondary and tertiary structure.

The difference in the total yield of properly folded wt vs Y80G OmpX supports the view that the presence of the mortise-tenon motif promotes a folding energy landscape where most protein molecules end up in a properly folded form, whereas the lack of this motif results in fewer protein molecules being rescued from pathways that can ultimately lead to misfolding. The observed increased efficiency in folding with the increase in temperature (for both wt and
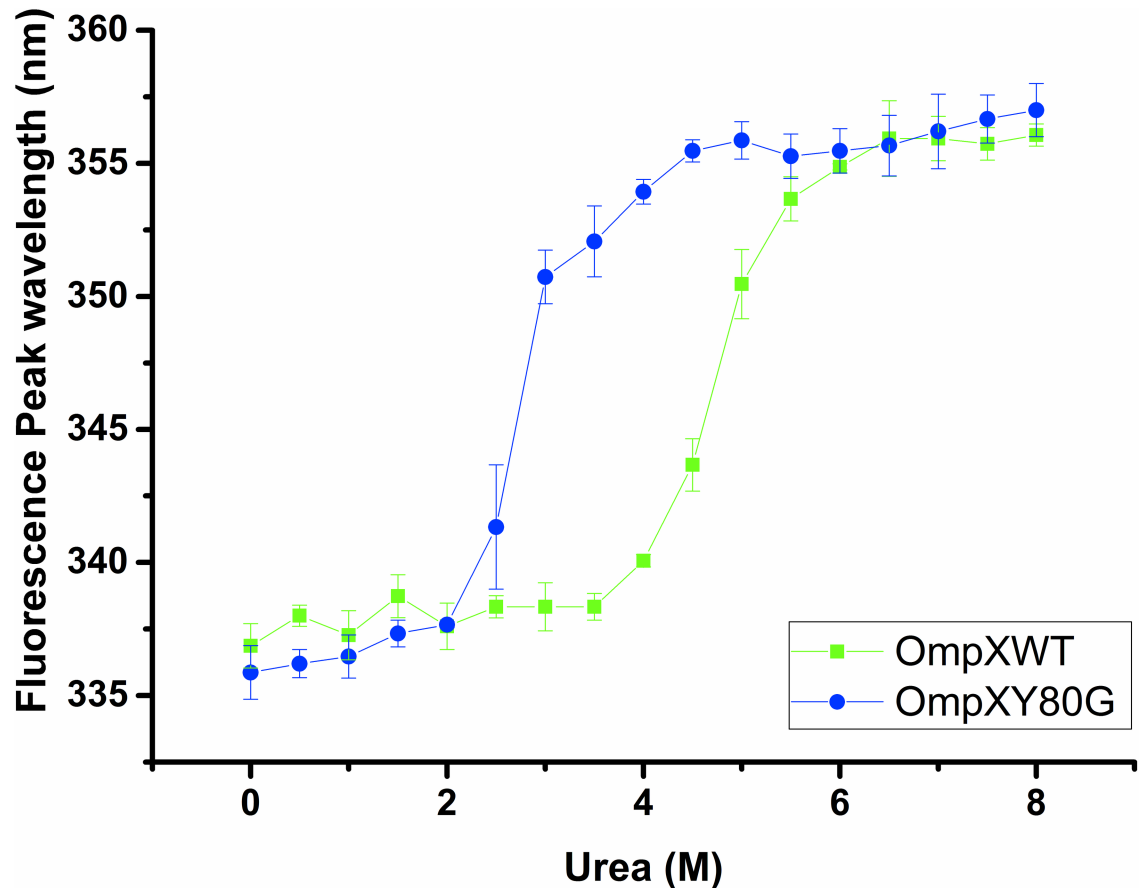
Fig 8. Stability of wt and Y80G OmpX as a function of urea concentration. As OmpX unfolds, the 2 tryptophan residues are exposed to a hydrophilic environment, resulting in a shift of the fluorescence maximum to higher wavelengths (336 nm to 356 nm), indicating that the protein has unfolded. Three technical replicates were used, and the averaged results are shown with standard deviation.

Y80G OmpX) can be explained by the activation energy required for OmpX folding into SB-12. Differences in the SB-12 micellar size and/or water-headgroup interactions at the different temperatures might also play a role. In general, detergent micelles become slightly smaller at increased temperatures, and water-head group interactions can also vary as a function of temperature [39–41]; we thus cannot exclude that temperature-dependent differences in SB-12 properties also affect OmpX folding efficiency.

We also performed a recovery experiment where following thermal denaturation of wt and Y80G OmpX at 100˚C, the protein was incubated on ice to assess whether the protein could reversibly refold (Fig 6, gel 3). While ~90% of wt OmpX was able to reversibly refold, OmpX Y80G was unable to refold under these conditions. While this to some extent may be an effect of the reduced hydrophobicity of the mutant, we believe that the recovery after denaturation, as well as the changed activation energy requirements for folding, are both more likely a consequence of the deleted mortise-tenon motif. This motif might act to thermodynamically stabilize transiently folded states by a zipper-like mechanism that brings together neighboring β-strands.

To study the effect of the Y80G mutation on the thermal stability of folded OmpX, heat stability assays using SDS-PGE gels were carried out for both wt and Y80G OmpX (Fig 7). Refolded wt and Y80G OmpX were heated to 85˚C for different time points, and then run on

an SDS-PAGE gel to determine the levels of folded vs. unfolded protein (Fig 7). Y80G OmpX is clearly less thermostable than wt OmpX—while approximately 90% of wt OmpX is still folded after 80 minutes, Y80G OmpX is fully denatured after 25 minutes.

In addition, we have tested the stability of both wt and Y80G OmpX in the chaotropic reagent urea (Fig 8). Measuring the fluorescence spectra for both wt OmpX and Y80G mutant reveals that wt OmpX is stably folded in urea concentrations to up to 4M, while the Y80G mutant begins to denature at 2.5M urea.

Thus, along with the observed effects on protein folding kinetics, folding yield and refolding efficiency, the Y80G mutation also leads to decreased thermal stability of folded OmpX, and to decreased resistance towards chemical perturbations by chaotropic reagents. This is in accordance with studies on other OMPs where single point mutations can also stabilize the structure against the same type of perturbations [42]. We thus suggest that while Y80G OmpX maintains a native-like fold that is supported within detergent micelle, the final structure is nonetheless compromised, likely because strands 5 and 6 are not stably locked together, leading to an overall less stable secondary and tertiary fold.

Discussion

In this study, we investigate the effect of a widely conserved mortise-tenon motif on the folding of beta-barrel proteins, and show that mutating the tyrosine of this motif in OmpX, leads to a change in folding kinetics, activation energy of folding (based on the requirement for higher temperature for refolding of the mutant compared to the wt protein, Fig 6), and stability of the folded state (Fig 7). Y80 in OmpX is present in a previously proposed hydrophobic cluster involving residues 73–82 of OmpX. This cluster was shown by NMR to interact transiently with DHPC micelles and to retain a non-random α-helical structure in 8M urea [33,34]. While W76 has been previously implicated to have a role in the formation and stability of this hydrophobic cluster, as well as in the kinetic stability of OmpX [43], this is the first time to our knowledge that the effect of mutating the tyrosine, and thus the mortise-tenon motif in this region, on OmpX folding has been studied.

Our study shows *in vitro* folding. *In vivo*, OmpX and other OMPs are bound to periplasmic chaperones prior to membrane insertion. In the structure of OmpX bound to the *E. coli* chaperone Skp, no specific regions of OmpX were recognized by Skp, and it was found that the protein was overall highly flexible. Compared to urea-unfolded OmpX, no residual secondary structure was present and the protein was found to be in a more compact state compared to urea-unfolded OmpX [44]. Importantly, the authors argued that the lack of specific interactions supports the idea that Skp is able to interact with a variety of β-barrel substrates. It has also been observed in the autotransporter Esp, that the mortise-tenon motif is not interacting either with Skp or BamA [45]. We would thus suggest that the role of Y80 in OmpX, and the corresponding mortise-tenon motifs that we describe for other OMP families, is not specific to chaperone-bound states *in vivo*, but stabilizes a conformational intermediate that increases the overall rate and efficiency of folding, and influences initial protein-lipid interactions that catalyze protein folding. The potential for protein-lipid interactions with Y80 is further supported by the observed hydrophobic interactions between DHPC micelles and the partially structured regions between residues 73–82 and residues 137–145 in urea-unfolded OmpX [33]. In the presence of the DHPC micelles, these two clusters of amino acids were found to have helical-like conformations with the hydrophobic side chains, including that of Y80, pointing outwards. The outward-facing conformation of the hydrophobic side chains would likely promote the observed interactions between these two regions and the DHPC micellar tails, which is

supported by the fact that deletion of either W76 or W140 led to a loss in DHPC interactions with this two clusters, as well these region's structural propensity.

Like for all enzymes, the role of BamA is to change a reaction pathway to remove kinetic barriers, rather than to change the thermodynamics of the reaction. Thus, while the expected mechanisms for the folding of OmpX *in vivo* vs. *in vitro* will be quite different in detail, we can assume that the start- and endpoints of the folding process are similar (Fig 9). Both *in vivo* and *in vitro*, it is likely that the G112-Y80 motif in OmpX plays a role in initiation of folding: by Y80 interacting with the membrane or micelle surface (panels B), by increasing the rate of protein folding (by helping to find the registered for the folding strands and hairpins, panels C), and/or by stabilizing the folded protein (by clamping neighboring strands together, panels D). We speculate that *in vivo*, this region and the C-terminal β-signal (which is essential for BAM recognition) [46] interact with the periplasmic leaflet of the OM upon release from Skp, similarly to the interactions observed between these regions and DHPC micelles.

To further access the conservation of this motif within non-bacterial β-barrels, we analyzed the protein sequence of the voltage dependent anion channel (VDAC). VDAC is an essential pore in the eukaryotic outer mitochondrial membrane that functions to transport substrates across the mitochondrial OM [47–50], and is the only eukaryotic β-barrel membrane protein whose structure is solved to date to high resolution, with structures of isoform 1 from both human and mouse [51–53], and isoform 2 from zebrafish [54]. All structures show the existence of the mortise-tenon motif on the 13th and 14th β-strands where the pair is formed most commonly between a threonine and tyrosine, in contrast to bacterial β-barrel OMPs where more commonly the pair is formed between a glycine and tyrosine (Fig 10, top panel). To further access the conservation of this motif, we aligned the sequences of VDAC channels from a diverse range of organisms, from yeast to plant to human, with sometimes very low pairwise sequence identity (less than 30%), and found near 100% conservation of this motif (Fig 10, upper panel). The high conservation of this motif in VDAC further supports the importance of this motif in all transmembrane β-barrels and may indeed be direct proof of an evolutionary link to prokaryotic β-barrels, which previous statistical analysis of sequence conservation [7] has not provided with high confidence.

All β-barrel OMPs seem to have evolved from a single, ancestral ββ hairpin through duplication events [7], which is further supported by the sequence similarity between individual β-hairpins [55]. Indeed, refolding studies with engineered OmpX variants have shown that by a simple duplication event, larger folded barrel structures can be formed in a single genetic operation. Intriguingly, constructs lacking a central hairpin containing β-strands 4 and 5, where Y80 is located on β-strand 5, were unable to fold in all variants tested, further supporting the idea of a central folding core within OmpX [4]. Conserved protein building blocks have been shown to exist throughout nature [56], and specifically within coiled coils there are local motifs that are required for triggering protein folding [57]. We propose that the mortise-tenon motif is important in folding and stabilizing one such conserved structural building block, a transmembrane ββ hairpin in β-barrel OMPs, with an additional important function in stabilizing folding intermediates and annealing neighboring strands in larger barrel structures. Extremely efficient insertion and folding of OMPs is required in fast-growing bacteria to avoid protein misfolding and subsequent negative effects such as inclusion body formation and the induction of the periplasmic stress response [58–60]. At the same time, the stability of folded proteins can be important in withstanding adverse environmental conditions such as elevated temperatures [14]. Together, both the kinetic and thermodynamic benefits of the mortise-tenon motif explain its extreme conservation in nearly all OMP families.
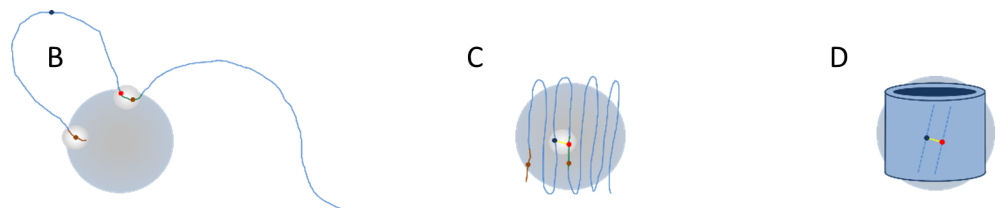
In vivo
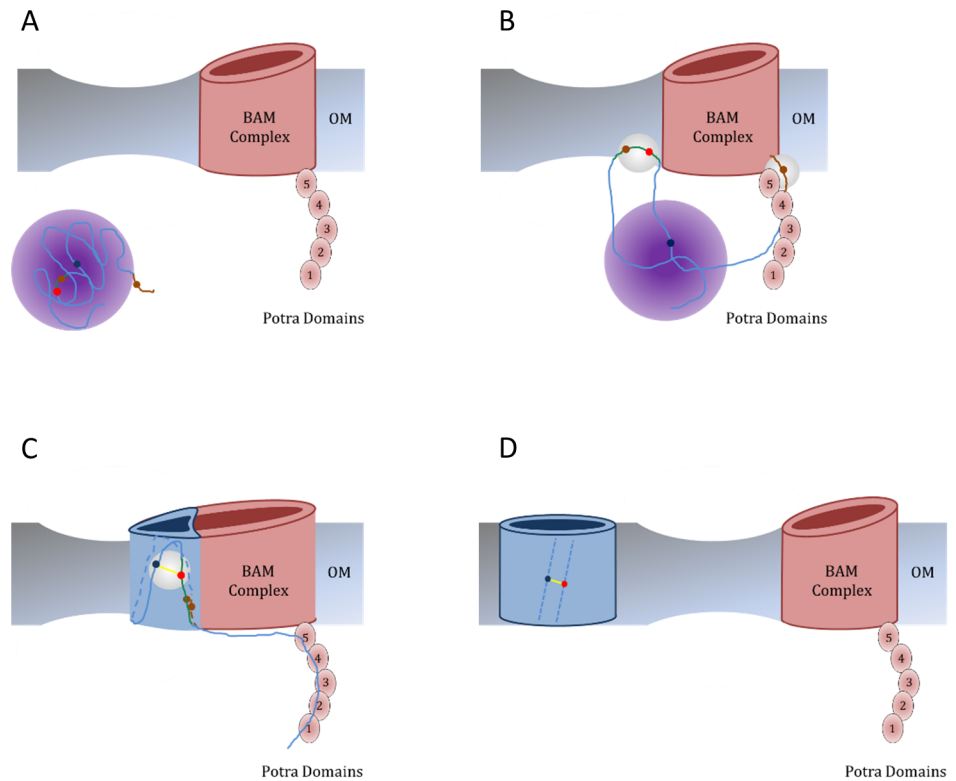
In vitro

Fig 9. The folding of OmpX *in vivo* and *in vitro*. Highlighted in white spheres in both folding models are where similarities between the two folding processes may be present. ***In vivo***: The BAM and Skp-assisted folding of OmpX into the OM. **(A)** OmpX is held in a flexible, relatively packed state by the Skp complex. **(B)** Recognition of the β-signal by the BAM complex leads to its release from Skp, and delivery to the BAM Complex. Possible interactions between the periplasmic leaflet and the β-signal and region 73–82 may also occur. **(C)** The BAM-assisted folding and insertion of OmpX into the OM is stabilized by the G112-Y80 motif by aligning strands 5 and 6. **(D)** OmpX is released from the BAM complex, and the fully-folded protein is thermally stabilized by the G112-Y80 motif. ***In vitro***: **(B)** OmpX is expected to be extended in 8M urea compared to the Skp-bound form. Residual secondary structure and micelle interactions are present in the β-signal and region 73–82. **(C)** Upon urea dilution, OmpX will collapse and simultaneously fold into the DHPC micelle. This process is mediated by Y80G-Y112 motif, which kinetically increases the speed of folding by aligning strands 5 and 6. Strands 1 and 8 come together after folding of the protein core, followed by loop formation and a stable hydrogen bond network. **(D)** The final structure is thermally stabilized by the G112-Y80 motif.

A

```
                          *              *
Mouse_VDAC      240  ETSKSRVTQSNFAVGYKTDEFQLHTNVND-GTEFGGSIYQKVNKKL-ETAVNLAWTAG--  295
Human_VDAC_1    158  ETAKSRVTQSNFAVGYKTDEFQLHTNVND-GTEFGGSIYQKVNKKL-ETAVNLAWTAG--  213
Human_VDAC_2    169  DSAKSKLTRNNFAVGYRTGDFQLHTNVND-GTEFGGSIYQKVCEDL-DTSVNLAWTSG--  224
Human_VDAC_3    158  DTAKSKLSQNNFALGYKAADFQLHTHVND-GTEFGGSIYQKVNEKI-ETSINLAWTAG--  213
Elephant_VDAC   158  ETAKSRVTQSNFAVGYKTDEFQLHTNVND-GTEFGGSIYQKVNKKL-ETAVNLAWTAG--  213
Platypus_VDAC   159  DTAKSKLAQNNFALGYKAADFQLHTNVND-GTEFGGSIYQKVNEKV-ETSVNLAWTAG--  214
Croc_VDAC       158  ETTKSRVTQSNFAVGYKTDEFQLHTNVND-GTEFGGSIYQKVNDKL-ETAVNLAWTAG--  213
Gecko_VDAC      197  ETTKSRITQSNFAVGYKTDEFQLHTNVND-GTEFGGSIYQKVNDKL-ETAVNLAWTAG--  252
Mallard_VDAC    158  ETAKSRVTQSNFAVGYKTDEFQLHTNVND-GTEFGGSIYQKVNDKL-ETAVNLAWTAG--  213
Zebrafish_VDAC  158  DTAKSKMTQNNFAVGYKTGDFQLHTNVND-GSEFGGSIYQKVSDKL-ETAVNLAWTAG--  213
Yeast_VDAC      212  DISAGSISRYAMALSYFAKDYSLGATLNN-EQITTVDFHQNVNAFLQVGAKATMNCKLP-  269
Cobra_VDAC      169  DTAKSKLSQNNFALGYKAADFQLHTNVND-GTEFGGSIYQKVNNKV-ETAVSLAWTAG--  224
Nematode_VDAC   136  WCFQTFLQSLSFAFGRVGTDYVLHSYVND-AREFGGSFYHSVAHNL-EIGAMFSWMTG--  191
Rice_VDAC       148  DTATGNLTKYNAGLSFSNDDLIASLNLNNKGDSLTASYYHIVNHS-A-TAVGAELTHSFS  205
Consensus_ss:        ee     eeeeeeee   eeeeeee    eeeeeeee    eeeeeeee
                                                 13th strand    14th strand
```

B

Mouse VDAC1
Y208 - G217

Human VDAC
Y195 - G204

Zebrafish VDAC2
Y195 - G204
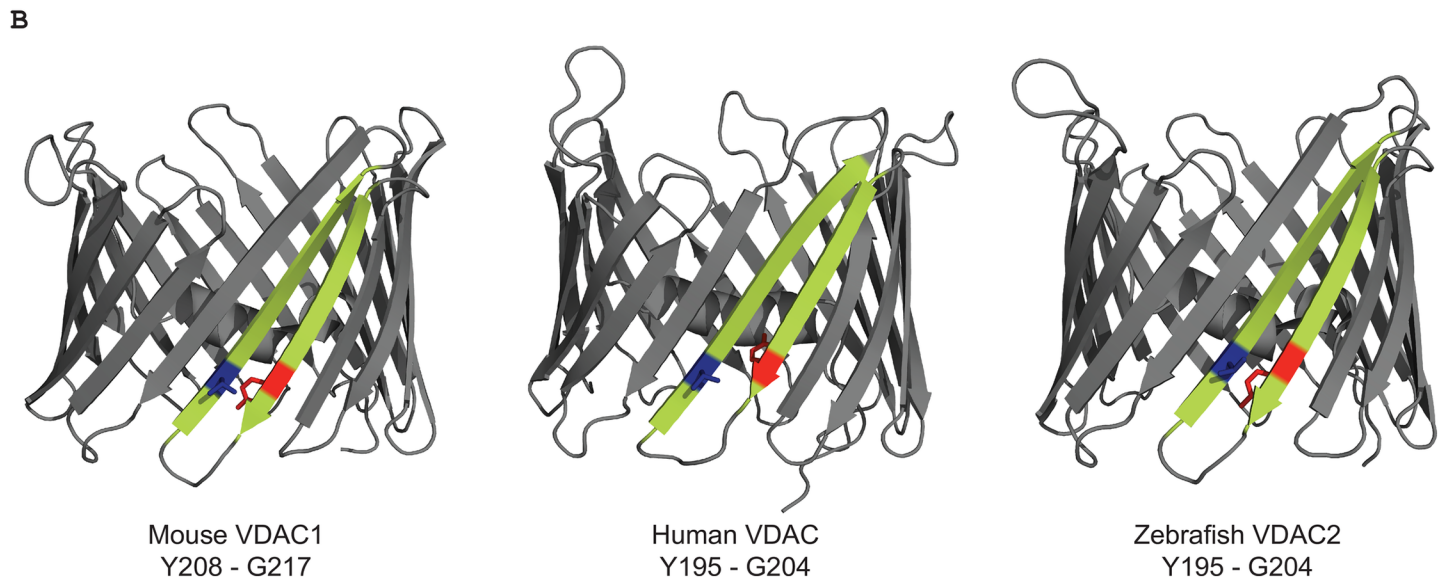
Fig 10. **(A)** Alignments of the mortise-tenon motif in VDAC proteins. The conserved motif is located on the 13th and 14th β-strands, and most commonly consists of a threonine and a tyrosine. The color scheme is the same as in Fig 2. **Mouse_VDAC**—AAB47777.1 isoform 1 [*Mus musculus*]; **Human_VDAC_1**— AAA61272 isoform 1 [*Homo sapiens*]; **Human_VDAC_2**—AAB59457.1 isoform 2 [*Homo sapiens*]; **Human VDAC_3**—AAB93872.1 isoform 3 [*Homo sapiens*]; **Elephant_VDAC**—XP_010588919.1 isoform 1 [*Loxodonta africana*]; **Platypus_VDAC**—XM_007668866.2 isoform 1 [*Ornithorhynchus anatinus*]; **Croc_VDAC** —XP_019390612.1 isoform 1 [*Crocodylus porosus*]; **Gecko_VDAC**—AAW79047.1. isoform 1 [*Gekko japonicus*]; **Mallard_VDAC**—XP_005011296.1 isoform 1 [*Anas platyrhynchos*]; **Zebrafish_VDAC**—AAH42329.1 isoform 2 [*Danio rerio*]; **Yeast_VDAC**—AAA35208.1 isoform 1 [*Saccharomyces cerevisiae* (strain ATCC 204508 / S288c)]; **Cobra_VDAC**—ETE66681.1 isoform 3 [*Ophiophagus hannah*]; **Nematode_VDAC**—CDP92897.1 isoform 1 [*Brugia malayi*]; **Rice_VDAC**—CAB82853.1—isoform 1 [*Oryza sativa* subsp. japonica]. **(B)** Available VDAC structures. The 13th and 14th β-strands containing the motif are highlighted in yellow with the marked mortise-tenon pair in blue and red respectively. The amino acid pair forming the motif is given under the names of the proteins. PDB codes: Mouse VDAC1 [3EMN]; Human VDAC1 [2JK4]; Zebrafish VDAC2 [4BUM].

https://doi.org/10.1371/journal.pone.0182016.g010

Methods

Bioinformatic analysis

To analyze the occurrence of mortise-tenon motifs in known OMP β-barrel structures, we used the SCOPe database, which groups them into 15 families within 7 superfamilies. To obtain multiple sequence alignments (MSA) for these families, we mapped them to their corresponding Pfam entries by performing sequence searches initiated with representatives of each family against the Pfam database[61]. The seed alignments provided for each family by Pfam

were used for detecting conserved mortise-tenon motifs. Representative structures from each SCOPe family were analyzed manually to identify tenon-mortise motifs. Next, the respective multiple sequences alignments were analysed for the conservation of these motifs across the family; only motifs conserved in at least 70% of the sequences in the MSA were deemed important. For multiple structural alignments, the PROMALS3D webserver tool was used [62] with default settings. The NCBI accession numbers for the amino acid sequences used are provided in all relevant figure legends.

Protein structure visualization

The opens source software PyMOL (https://www.pymol.org/) was used for 3D visualization of all structures.

Cloning

WT *E. coli OmpX* without the signal sequence (residues 1–23) in pET3b (pET3bOmpX) was used from a previous study [4].

The tyrosine mutant (OmpX Y80G) was constructed using a two-step PCR mutagenesis protocol [63]. Briefly, the N- and C- terminal parts were amplified separately including a Tyrosine 80 to Glycine mutation in the primers

FP1: 5–GATTATCATATGGCGACTTCTACCGTAACTGGC
RP1: 5–GCCGATGCTTGCCCAGTCGTTAATGC
FP2: 5–GCATTAACGACTGGGCAAGCATCGGCGGTGTAGTGGGTGTGGGTTATG
RP2: 5–AAAAGAGGATCCTTAGAAGCGGTAACCAACACCG

The two resulting PCR products were fused in a second PCR via their complementary sequences using the forward and reverse primers from the first step, and the product was digested and cloned into pET3b using the NdeI and BamHI sites as above (pET3bOmpX_Y80G). All constructs were transformed into chemically competent *E. coli* TOP10 cells (Invitrogen), and confirmed by sequencing.

Protein production and purification

Large-scale protein expression and purification was achieved using previously described methods [4]. In brief, pET3bOmpX and pET3bOmpX_Y80G were transformed into chemically competent *E. coli* C41(DE3) cells. Cells were grown at 37°C in LB medium in shaking flasks, induced with 1mM IPTG (isopropyl-β-d-thiogalactopyranoside) when the OD600 reached 0.6, and harvested after 4h by centrifugation at 4°C.

Cells containing inclusion bodies were suspended in PBS supplemented with 10mM $MgCl_2$, 10 mM $MnCl_2$ and a pinch of DNase, and lysed with a French Press three times. Inclusion bodies were purified from the membrane debris by slow centrifugation for 15 min at 2000g. The resulting pellet was resuspended in 1,5% (v/v) Triton X-100 in PBS, centrifuged again at 2000g, followed by washing with MiliQ water three times to remove the detergent. The crudely purified inclusion bodies were dissolved in urea buffer (6M Urea, 1mM EDTA, 20mM Tris-HCl pH8.5) and purified using a GE Healthcare FPLC system using a 20 ml Mono Q anion-exchange column with a linear gradient of 0–1 M NaCl in urea buffer. The protein purity was checked on an SDS-PAGE gel. Samples were diluted to the desired concentration in urea-containing buffer.

Stopped-flow spectroscopy

Kinetic refolding traces were obtained using a BioLogic stopped-flow apparatus SFM-300/s—MOS-200 with a FC-08 microcuvette accessory. All experiments were done at 25°C. The

excitation light was set to 295 nm and the emission was recorded with a cut-off filter at 320 nm. One 2mL syringe was loaded with unfolded protein at 5 mg/ml concentration in urea buffer (6M Urea, Tris-HCl pH8.5, 1mM EDTA), and the two remaining 10 mL syringes were loaded with detergent-rich buffer (1% C_8POE (w/v) (octyl-polyoxyethylene), Tris-HCl pH8.5, 1mM EDTA). The mixing ratio was set to 1:20 (v/v protein to detergent solution), and a flow of 14 mL/s was applied at a final volume of 422 µl per experiment. The acquisition start time was set to 10ms before the stop of mixing to minimize the dead time. Due to software limitations for the number of time points that could be recorded, we divided the acquisition into 3 segments: timebase 1 with 5001 data points recorded every 0.00005s; timebase 2 with 2749 data points recorded every 0.001s, and timebase 3 with 44 data points recorded every 0.5s. 500 single shots were collected for both wt and Y80G OmpX, and the raw data analyzed by custom-designed kinetic analysis software described in the following section.

Kinetic analysis (software)

Stopped-flow data were analyzed with in-house software developed by us. Our method is based on a maximum likelihood approach that fits multi-exponential decays directly to the kinetic refolding traces. This is based on the following standard equation for multi-exponential fits,

$$f(t) = c + b\,t - \sum_{x=1}^{n} a_x e^{-k_x t}$$

where c is the offset (constant baseline) of the measurement, b is an optional linear component that corrects for instrument drift and that was minimal for our experiments, a is the amplitude and k the rate constant of each exponential component. Thus, each exponential decay has two parameters, an amplitude and a rate constant. Therefore, the total number of parameters that needs to be fitted is twice the number of exponentials plus an additional parameter for the baseline. We account for deviations from the ideal model by assuming a Gaussian noise model whose parameters we also estimate (for each of the three time bases we estimate a separate noise level). The maximum likelihood approach results in a non-linear least-squares fitting problem that we solve by running the Nelder-Mead algorithm starting from randomly varying initial conditions. Additional exponential components are added progressively, starting from the best fit obtained with the previous model and adding one additional exponential component one by one. We fit each refolding trace in a fully automated way starting with a single exponential decay, which is successively refined to a maximum of five exponential decays. By monitoring the chi-squared residual as a function of the number of exponentials, we find that both the wt as well as the mutant data can be modelled with two exponentials; higher order models do not improve the quality of the fit significantly. Our software is implemented in Python 2.7 and depends on the numpy and scipy libraries (Habeck et al. In preparation).

For comparison, the single-, double, and triple-exponential fit parameters are shown in Table 3.

The software is provided in S2 File.

Folding, refolding and heat stability

The folding, refolding and heat stability for both wt and Y80G OmpX was accessed *in vitro* in the detergent SB-12 (N-Dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate. For the folding experiment, the folding reaction was carried out for 2 hours on ice, and at 20, 30, 40 and 50˚C. SDS page loading buffer was added to stop the refolding reaction, and the samples were directly loaded onto SDS page gels without centrifugation or heating. For refolding experiments, wt and

Table 3. Fit parameters for different exponential fits. The chi-square values suggest that a single-exponential fit does not describe the data well, while the standard deviations of the individual rates increase dramatically for the triple-exponential fit, to a level where the errors are as large as the calculated values for rates (and amplitudes) 2 and 3.

	single exponential fit		double exponential fit		triple exponential fit	
	wt	**Y80G**	**wt**	**Y80G**	**wt**	**Y80G**
Avg. rate 1	8,03 ± 0,3	6,45 ± 0,17	8,95 ± 0,40	7,02 ± 0,19	9,07 ± 0,45	7,02 ± 0,18
Avg. amp. 1	100 ± 0	100 ± 0	91,7 ± 0,68	91,7 ± 0,42	85,68 ± 9,14	87,80 ± 10,62
Avg. rate 2	-	-	0,85 ± 0,24	0,51 ± 0,04	2,44 ± 2,83	1,26 ± 2,07
Avg. amp. 2	-	-	8,3 ± 0,68	8,3 ± 0,42	10,34 ± 7,41	10,87 ± 8,27
Avg. rate 3	-	-	-	-	3,61 ± 2,97	3,41 ± 1,96
Avg. amp. 3	-	-	-	-	3,97 ± 3,21	1,33 ± 2,74
Avg. chi^2	2,52 ± 0,59	3,41 ± 0,69	0,51 ± 0,16	1,67 ± 0,27	0,53 ± 0,25	1,19 ± 0,34

https://doi.org/10.1371/journal.pone.0182016.t003

Y80G were denatured at 100˚C followed by 2 hours of incubation at 20˚C prior to loading on an SDS page gel. For the heat stability assays, Denatured protein in 8M Urea was firstly refolded by rapidly diluting it 1:20 into pre-chilled buffer with SB-12, followed by RT incubation for 2 hours. The samples were centrifuged to remove aggregated proteins and transferred to a water bath heated to 85˚C for defined time periods. Samples were removed from the water bath at the indicated time points, and SDS-PAGE loading buffer was added immediately to arrest the reaction [4]. All samples were loaded onto a 15% SDS-PAGE in a water-cooled chamber to keep the temperature low, and silver-staining method was used for detection [64]. Estimation of folding yields were based on band intensities.

Fluorescence measurements of urea-unfolded samples

Fluorescence measurements were essentially performed as described earlier [4]. In brief, samples of folded OmpX were mixed with detergent buffer and urea stock solutions to achieve constant protein and detergent concentrations at different concentrations of chaotrope as indicated in Fig 8. Measurements were taken with a JASCO FP-8000 series spectrophotometer using standard settings, with an excitation wavelength of 280 nm.

Supporting information

S1 File.
(ZIP)

S2 File.
(ZIP)

Acknowledgments

Author Contributions

Conceptualization: Dirk Linke.

Data curation: Marcin Michalik, Marcella Orwick-Rydmark, Vikram Alva, Thomas Arnold.

Formal analysis: Marcella Orwick-Rydmark, Michael Habeck, Vikram Alva, Thomas Arnold, Dirk Linke.

Funding acquisition: Dirk Linke.

Investigation: Marcin Michalik, Thomas Arnold.

Methodology: Marcin Michalik.

Project administration: Dirk Linke.

Software: Michael Habeck.

Validation: Vikram Alva, Dirk Linke.

Visualization: Marcin Michalik, Marcella Orwick-Rydmark, Michael Habeck.

Writing – original draft: Marcin Michalik, Marcella Orwick-Rydmark, Dirk Linke.

Writing – review & editing: Marcin Michalik, Marcella Orwick-Rydmark, Michael Habeck, Vikram Alva, Dirk Linke.

References

1. Benz R. Permeation of hydrophilic solutes through mitochondrial outer membranes: review on mitochondrial porins. Biochim Biophys Acta. 1994; 1197: 167–96. https://doi.org/10.1016/0304-4157(94)90004-3 PMID: 8031826

2. Fischer K, Weber A, Brink S, Arbinger B, Schünemann D, Borchert S, et al. Porins from plants. Molecular cloning and functional characterization of two new members of the porin family. J Biol Chem. New York, New York, USA: ACM Press; 1994; 269: 25754–60. PMID: 7523392

3. Dupont M, Dé E, Chollet R, Chevalier J, Pagès JM. *Enterobacter aerogenes* OmpX, a cation-selective channel *mar-* and osmo-regulated. FEBS Lett. 2004; 569: 27–30. https://doi.org/10.1016/j.febslet.2004.05.047 PMID: 15225603

4. Arnold T, Poynor M, Nussberger S, Lupas AN, Linke D. Gene duplication of the eight-stranded β-barrel OmpX produces a functional pore: a scenario for the evolution of transmembrane β-barrels. J Mol Biol. Elsevier Ltd; 2007; 366: 1174–84. https://doi.org/10.1016/j.jmb.2006.12.029 PMID: 17217961

5. Hong H, Patel DR, Tamm LK, van den Berg B. The outer membrane protein OmpW forms an eight-stranded β-barrel with a hydrophobic channel. J Biol Chem. 2006; 281: 7568–77. https://doi.org/10.1074/jbc.M512365200 PMID: 16414958

6. Wimley WC. The versatile β-barrel membrane protein. Curr Opin Struct Biol. 2003; 13: 404–411. https://doi.org/10.1016/S0959-440X(03)00099-X PMID: 12948769

7. Remmert M, Biegert A, Linke D, Lupas AN, Söding J. Evolution of outer membrane β-barrels from an ancestral β β hairpin. Mol Biol Evol. 2010; 27: 1348–58. https://doi.org/10.1093/molbev/msq017 PMID: 20106904

8. Wimley WC. Toward genomic identification of β-barrel membrane proteins: Composition and architecture of known structures. Protein Sci. Cold Spring Harbor Laboratory Press; 2009; 11: 301–312. https://doi.org/10.1110/ps.29402 PMID: 11790840

9. Ulrich T, Rapaport D. Biogenesis of β-barrel proteins in evolutionary context. Int J Med Microbiol. Elsevier GmbH.; 2015; 305: 259–264. https://doi.org/10.1016/j.ijmm.2014.12.009 PMID: 25596888

10. Kim KH, Aulakh S, Paetzel M. The bacterial outer membrane β-barrel assembly machinery. Protein Sci. Hoboken: Wiley Subscription Services, Inc., A Wiley Company; 2012; 21: 751–768. https://doi.org/10.1002/pro.2069 PMID: 22549918

11. Allen WJ, Phan G, Waksman G. Structural biology of periplasmic chaperones. McPherson A, editor. Adv Protein Chem Struct Biol. Academic Press; 2009; 78: 51–97. https://doi.org/10.1016/S1876-1623(08)78003-9 PMID: 20663484

12. Moon CP, Zaccai NR, Fleming PJ, Gessmann D, Fleming KG. Membrane protein thermodynamic stability may serve as the energy sink for sorting in the periplasm. Proc Natl Acad Sci USA. National Academy of Sciences; 2013; 110: 4285–90. https://doi.org/10.1073/pnas.1212527110 PMID: 23440211

13. Bakelar J, Buchanan SK, Noinaj N. The structure of the β-barrel assembly machinery complex. Science (80-). 2016; 351: 180–186. https://doi.org/10.1126/science.aad3460 PMID: 26744406

14. Rollauer SE, Sooreshjani MA, Noinaj N, Buchanan SK. Outer membrane protein biogenesis in Gram-negative bacteria. Philos Trans R Soc B Biol Sci. 2015; 370: 20150023. https://doi.org/10.1098/rstb.2015.0023 PMID: 26370935

15. Noinaj N, Rollauer SE, Buchanan SK. The β-barrel membrane protein insertase machinery from Gram-negative bacteria. Curr Opin Struct Biol. Elsevier Ltd; 2015; 31: 35–42. https://doi.org/10.1016/j.sbi.2015.02.012 PMID: 25796031

16. Gu Y, Li H, Dong H, Zeng Y, Zhang Z, Paterson NG, et al. Structural basis of outer membrane protein insertion by the BAM complex. Nature. 2016; 531: 64–69. https://doi.org/10.1038/nature17199 PMID: 26901871

17. Calmettes C, Judd A, Moraes TF. Structural Aspects of Bacterial Outer Membrane Protein Assembly. Adv Exp Med Biol. 2015; 883: 255–70. https://doi.org/10.1007/978-3-319-23603-2_14 PMID: 26621472

18. Noinaj N, Kuszak AJ, Balusek C, Gumbart JC, Buchanan SK. Lateral opening and exit pore formation are required for BamA function. Structure. 2014; 22: 1055–62. https://doi.org/10.1016/j.str.2014.05.008 PMID: 24980798

19. Kleinschmidt JH. Folding of β-barrel membrane proteins in lipid bilayers—Unassisted and assisted folding and insertion. Biochim Biophys Acta—Biomembr. Elsevier B.V.; 2015; 1848: 1927–1943. https://doi.org/10.1016/j.bbamem.2015.05.004 PMID: 25983306

20. Paramasivam N, Habeck M, Linke D. Is the C-terminal insertional signal in Gram-negative bacterial outer membrane proteins species-specific or not? BMC Genomics. BioMed Central; 2012; 13: 510. https://doi.org/10.1186/1471-2164-13-510 PMID: 23013516

21. Robert V, Volokhina EB, Senf F, Bos MP, Van Gelder P, Tommassen J. Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. PLoS Biol. Public Library of Science; 2006; 4: e377. https://doi.org/10.1371/journal.pbio.0040377 PMID: 17090219

22. Leyton DL, Johnson MD, Thapa R, Huysmans GHM, Dunstan R a, Celik N, et al. A mortise-tenon joint in the transmembrane domain modulates autotransporter assembly into bacterial outer membranes. Nat Commun. Nature Publishing Group; 2014; 5: 4239. https://doi.org/10.1038/ncomms5239 PMID: 24967730

23. Baldermann C, Lupas A, Lubieniecki J, Engelhardt H. The regulated outer membrane protein Omp21 from *Comamonas acidovorans* is identified as a member of a new family of eight-stranded β-sheet proteins by its sequence and properties. J Bacteriol. 1998; 180: 3741–9. 0021-9193/98/$04.00?0 PMID: 9683466

24. Vogt J, Schulz GE. The structure of the outer membrane protein OmpX from *Escherichia coli* reveals possible mechanisms of virulence. Structure. 1999; 7: 1301–1309. https://doi.org/10.1016/S0969-2126(00)80063-5 PMID: 10545325

25. Walser R, Kleinschmidt JH, Skerra A, Zerbe O. β-Barrel scaffolds for the grafting of extracellular loops from G-protein-coupled receptors. Biol Chem. 2012; 393: 1341–55. https://doi.org/10.1515/hsz-2012-0234 PMID: 23109547

26. Pautsch A, Vogt J, Model K, Siebold C, Schulz GE. Strategy for membrane protein crystallization exemplified with OmpA and OmpX. Proteins. 1999; 34: 167–72. PMID: 10022352

27. Arnold T, Linke D. The use of detergents to purify membrane proteins. Curr Protoc Protein Sci. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2008;Chapter 4: Unit 4.8.1–4.8.30. https://doi.org/10.1002/0471140864.ps0408s53 PMID: 18729050

28. Fox NK, Brenner SE, Chandonia J-M. SCOPe: Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. Nucleic Acids Res. 2014; 42: D304–9. https://doi.org/10.1093/nar/gkt1240 PMID: 24304899

29. Moon CP, Fleming KG. Using tryptophan fluorescence to measure the stability of membrane proteins folded in liposomes. Michael L. Johnson JMH and GKABT-M in E, editor. Biothermodynamics, Part D. Academic Press; 2011; 492: 189–211. http://dx.doi.org/10.1016/B978-0-12-381268-1.00018-5

30. Horne JE, Radford SE. A growing toolbox of techniques for studying β-barrel outer membrane protein folding and biogenesis. Biochem Soc Trans. 2016; 44: 802–9. https://doi.org/10.1042/BST20160020 PMID: 27284045

31. Linke D, Frank J, Pope MS, Soll J, Ilkavets I, Fromme P, et al. Folding kinetics and structure of OEP16. Biophys J. 2004; 86: 1479–87. https://doi.org/10.1016/S0006-3495(04)74216-2 PMID: 14990475

32. Raschle T, Rios Flores P, Opitz C, Müller DJ, Hiller S. Monitoring backbone hydrogen-bond formation in β-Barrel membrane protein folding. Angew Chem Int Ed Engl. 2016; 55: 5952–5. https://doi.org/10.1002/anie.201509910 PMID: 27062600

33. Hiller S, Wider G, Imbach LL, Wüthrich K. Interactions with hydrophobic clusters in the urea-unfolded membrane protein OmpX. Angew Chem Int Ed Engl. 2008; 47: 977–81. https://doi.org/10.1002/anie.200703367 PMID: 18161715

34. Tafer H, Hiller S, Hilty C, Fernández C, Wüthrich K. Nonrandom structure in the urea-unfolded *Escherichia coli* outer membrane protein X (OmpX). Biochemistry. 2004; 43: 860–9. https://doi.org/10.1021/bi0356606 PMID: 14744128

35. Stanley AM, Fleming KG. The process of folding proteins into membranes: challenges and progress. Arch Biochem Biophys. 2008; 469: 46–66. https://doi.org/10.1016/j.abb.2007.09.024 PMID: 17971290

36. Kleinschmidt JH, Tamm LK. Folding intermediates of a β-barrel membrane protein. Kinetic evidence for a multi-step membrane insertion mechanism. Biochemistry. American Chemical Society; 1996; 35: 12993–3000. https://doi.org/10.1021/bi961478b PMID: 8855933

37. Schweizer M, Hindennach I, Garten W, Henning U. Major proteins of the *Escherichia coli* outer cell envelope membrane. Interaction of protein II with lipopolysaccharide. Eur J Biochem. 1978; 82: 211–7. https://doi.org/10.1111/j.1432-1033.1978.tb12013.x PMID: 340230

38. Rosenbusch JP. Characterization of the major envelope protein from Escherichia coli. Regular arrangement on the peptidoglycan and unusual dodecyl sulfate binding. J Biol Chem. 1974; 249: 8019–29. PMID: 4609976

39. He L, Garamus VM, Funari SS, Malfois M, Willumeit R, Niemeyer B. Comparison of small-angle scattering methods for the structural analysis of octyl-β-maltopyranoside micelles. J Phys Chem B. American Chemical Society; 2002; 106: 7596–7604. https://doi.org/10.1021/jp020034o

40. Hammouda B. Temperature effect on the nanostructureof SDS micelles in water. J Res Natl Inst Stand Technol. 2013; 118: 151. https://doi.org/10.6028/jres.118.008 PMID: 26401428

41. Malliaris A. Temperature dependence of the micelle aggregation number and rate of intramicellar exclmer formation in aqueous surfactant solutions. J Phys Chem. 1985; 89: 2709–2713.

42. Gessmann D, Mager F, Naveed H, Arnold T, Weirich S, Linke D, et al. Improving the resistance of a eukaryotic β-barrel protein to thermal and chemical perturbations. J Mol Biol. 2011; 413: 150–61. https://doi.org/10.1016/j.jmb.2011.07.054 PMID: 21835183

43. Chaturvedi D, Mahalakshmi R. Juxtamembrane tryptophans have distinct roles in defining the OmpX barrel-micelle boundary and facilitating protein-micelle association. FEBS Lett. Federation of European Biochemical Societies; 2014; 588: 4464–4471. https://doi.org/10.1016/j.febslet.2014.10.017 PMID: 25448987

44. Burmann BM, Wang C, Hiller S. Conformation and dynamics of the periplasmic membrane-protein–chaperone complexes OmpX–Skp and tOmpA–Skp. Nat Struct Mol Biol. Nature Publishing Group; 2013; 20: 1265–1272. https://doi.org/10.1038/nsmb.2677 PMID: 24077225

45. Pavlova O, Peterson JH, Ieva R, Bernstein HD. Mechanistic link between β barrel assembly and the initiation of autotransporter secretion. Proc Natl Acad Sci USA. 2013; 110: E938–47. https://doi.org/10.1073/pnas.1219076110 PMID: 23431155

46. Struyvé M, Moons M, Tommassen J. Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. J Mol Biol. 1991; 218: 141–8. http://dx.doi.org/10.1016/0022-2836(91)90880-F PMID: 1848301

47. Noskov SY, Rostovtseva TK, Chamberlin AC, Teijido O, Jiang W, Bezrukov SM. Current state of theoretical and experimental studies of the voltage-dependent anion channel (VDAC). Biochim Biophys Acta. 2016; 1858: 1778–90. https://doi.org/10.1016/j.bbamem.2016.02.026 PMID: 26940625

48. Lemasters JJ, Holmuhamedov E. Voltage-dependent anion channel (VDAC) as mitochondrial governator—thinking outside the box. Biochim Biophys Acta. 2006; 1762: 181–90. https://doi.org/10.1016/j.bbadis.2005.10.006 PMID: 16307870

49. Rostovtseva TK, Tan W, Colombini M. On the role of VDAC in apoptosis: fact and fiction. J Bioenerg Biomembr. 2005; 37: 129–42. https://doi.org/10.1007/s10863-005-6566-8 PMID: 16167170

50. Colombini M. VDAC: the channel at the interface between mitochondria and the cytosol. Mol Cell Biochem. 2004; 256–257: 107–15. Available: http://www.ncbi.nlm.nih.gov/pubmed/14977174 PMID: 14977174

51. Hiller S, Garces RG, Malia TJ, Orekhov VY, Colombini M, Wagner G. Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. Science. 2008; 321: 1206–10. https://doi.org/10.1126/science.1161302 PMID: 18755977

52. Bayrhuber M, Meins T, Habeck M, Becker S, Giller K, Villinger S, et al. Structure of the human voltage-dependent anion channel. Proc Natl Acad Sci USA. 2008; 105: 15370–5. https://doi.org/10.1073/pnas.0808115105 PMID: 18832158

53. Ujwal R, Cascio D, Colletier J-P, Faham S, Zhang J, Toro L, et al. The crystal structure of mouse VDAC1 at 2.3 A resolution reveals mechanistic insights into metabolite gating. Proc Natl Acad Sci USA. 2008; 105: 17742–7. https://doi.org/10.1073/pnas.0809634105 PMID: 18988731

54. Schredelseker J, Paz A, López CJ, Altenbach C, Leung CS, Drexler MK, et al. High resolution structure and double electron-electron resonance of the zebrafish voltage-dependent anion channel 2 reveal an oligomeric population. J Biol Chem. 2014; 289: 12566–77. https://doi.org/10.1074/jbc.M113.497438 PMID: 24627492

55. Neuwald AF, Liu JS, Lawrence CE. Gibbs motif sampling: detection of bacterial outer membrane protein repeats. Protein Sci. Cold Spring Harbor Laboratory Press; 1995; 4: 1618–1632. https://doi.org/10.1002/pro.5560040820 PMID: 8520488

56. Alva V, Söding J, Lupas AN. A vocabulary of ancient peptides at the origin of folded proteins. Kuriyan J, editor. eLife. eLife Sciences Publications, Ltd; 2015; 4: e09410. https://doi.org/10.7554/eLife.09410 PMID: 26653858

57. Voigt CA, Martinez C, Wang Z-G, Mayo SL, Arnold FH. Protein building blocks preserved by recombination. Nat Struct Biol. 2002; 9: 553–558. https://doi.org/10.1038/nsb805 PMID: 12042875

58. Miot M, Betton J-M. Protein quality control in the bacterial periplasm. Microb Cell Fact. 2004; 3: 4. https://doi.org/10.1186/1475-2859-3-4 PMID: 15132751

59. Tommassen J. Assembly of outer-membrane proteins in bacteria and mitochondria. Microbiology. 2010; 156: 2587–96. https://doi.org/10.1099/mic.0.042689-0 PMID: 20616105

60. Young JC, Hartl FU. A stress sensor for the bacterial periplasm. Cell. 2003; 113: 1–2. PMID: 12679025

61. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016; 44: D279–D285. https://doi.org/10.1093/nar/gkv1344 PMID: 26673716

62. Pei J, Kim B-H, Grishin N V. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res. 2008; 36: 2295–2300. https://doi.org/10.1093/nar/gkn072 PMID: 18287115

63. Li S, Wilkinson MF. Site-directed mutagenesis: a two-step method using PCR and DpnI. Biotechniques. 1997; 23: 588–90. PMID: 9343667

64. Nesterenko M V, Tilley M, Upton SJ. A simple modification of Blum's silver stain method allows for 30 minute detection of proteins in polyacrylamide gels. J Biochem Biophys Methods. 1994; 28: 239–42. PMID: 8064118