RESEARCH ARTICLE SUMMARY

PROTEIN DESIGN

De novo design of transmembrane β barrels

Anastassia A. Vorobieva, Paul White*, Binyong Liang*, Jim E. Horne, Asim K. Bera, Cameron M. Chow, Stacey Gerben, Sinduja Marx, Alex Kang, Alyssa Q. Stiving, Sophie R. Harvey, Dagan C. Marx, G. Nasir Khan, Karen G. Fleming, Vicki H. Wysocki, David J. Brockwell, Lukas K. Tamm, Sheena E. Radford, David Baker†

INTRODUCTION: Despite their key biological roles, only a few proteins that fold into lipid membranes have been designed de novo. A class of membrane proteins—transmembrane β barrels (TMBs)—forms a continuous sheet that closes on itself in lipid membranes. In addition to the challenge of designing β-sheet proteins, which are prone to misfolding and aggregation if folding is not properly controlled, the computational design of TMBs is complicated by limited understanding of TMB folding. As a result, no TMB has been designed de novo to date.

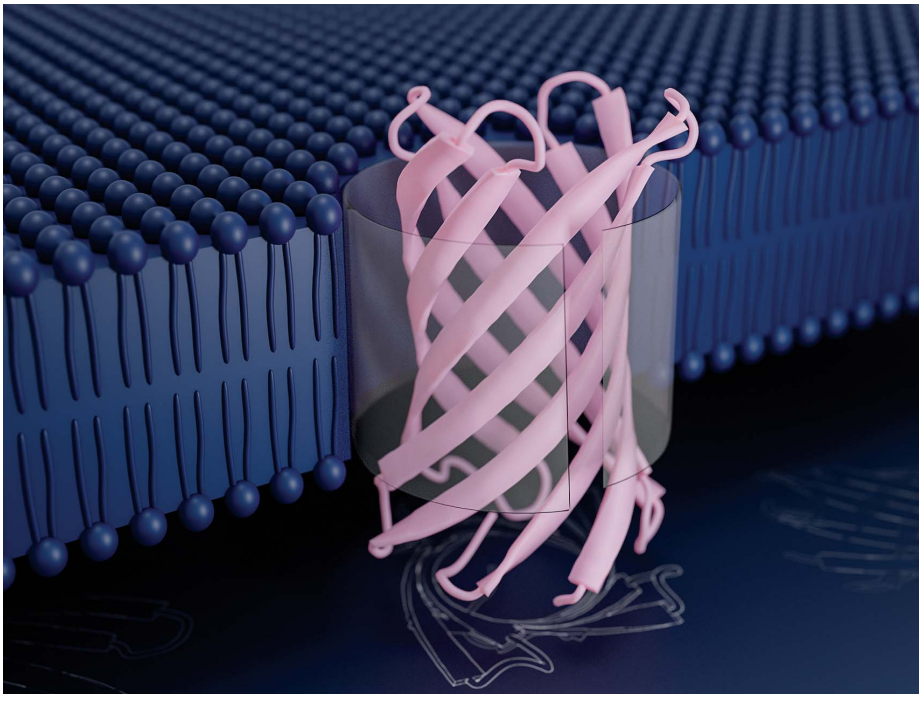
Although the folding of TMBs in vivo is catalyzed by the β-barrel assembly machinery (BAM), many TMBs can also fold spontaneously in synthetic membranes to form stable pores, making them attractive for biotechnology and single-molecule analytical applications. Hence, de novo design of TMBs has potential both for understanding the determinants of TMB folding and membrane insertion and for the custom engineering of TMB nanopores.

RATIONALE: We used de novo protein design to distill key principles of TMB folding through several design-build-test cycles. We iterated between hypothesis formulation, its implementation into computational design methods, and experimental characterization of the resulting proteins. To focus on the fundamental principles of TMB folding in the absence of complications due to interactions with chaperones and BAM in vivo, we focused on the challenge of de novo design of eight-stranded TMBs, which can fold and assemble into synthetic lipid membranes.

RESULTS: We used a combination of purely geometric models and explicit Rosetta protein structure simulations to determine the constraints that β-strand connectivity and membrane embedding place on the TMB architecture. Through a series of design-build-test cycles, we found that, unlike almost all other classes of proteins, locally destabilizing sequences are critical for expression and folding of TMBs, and that the β-turns that translocate through the bilayer during folding have to be destabilized to enable correct assembly in the membrane. Our results suggest that premature formation of β hairpins may result in off-target β-sheet structures that compete with proper membrane insertion and folding, and hence the β hairpins of TMBs must be designed such that they are only transiently formed prior to membrane insertion, when the protein is in an aqueous environment. In the hydrophobic environment of the lipid bilayer, the full TMB can assemble because the membrane-facing nonpolar residues, which would tend to cluster nonspecifically in an aqueous environment, instead make favorable interactions with the lipids. As the TMB assembles, the β hairpins are stabilized by interactions with the neighboring β strands.

Using computational methods that incorporate the above insights, we designed TMB sequences that successfully fold and assemble into both detergent micelles and lipid bilayers. Two of the designs were highly stable and could fold into liposomes more rapidly and reversibly than the transmembrane domain of the model outer membrane protein A (tOmpA) of *Escherichia coli*. A nuclear magnetic resonance solution structure and a high-resolution crystal structure for two different designs closely match the design models, showing that the TMB design method developed here can generate new structures with atomic-level accuracy.

CONCLUSION: This study elucidates key principles for de novo design of transmembrane β barrels, ranging from constraints on β-barrel architecture and β-hairpin design, as well as local and global sequence features. Our designs provide starting points for the bottom-up elucidation of the molecular mechanisms underlying TMB folding and interactions with the cellular outer membrane folding and insertion machinery. More generally, our work demonstrates that TMBs can be designed with atomic-level accuracy and opens the door to custom design of nanopores tailored for applications such as single-molecule sensing and sequencing. ■



CREDIT: IAN HAYDON.

De novo–designed eight-stranded transmembrane β barrels fold spontaneously and reversibly into synthetic lipid membranes. The illustration shows the crystal structure of the protein TMB2.17 designed in this study, which adopts a structure identical to the design model.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: dabaker@uw.edu
Cite this article as A. A. Vorobieva *et al.*, *Science* **371**, eabc8182 (2021). DOI: 10.1126/science.abc8182

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abc8182

RESEARCH ARTICLE

PROTEIN DESIGN

De novo design of transmembrane β barrels

Anastassia A. Vorobieva[1,2]*, Paul White[3]†, Binyong Liang[4]†, Jim E. Horne[3]‡, Asim K. Bera[1,5], Cameron M. Chow[1,5], Stacey Gerben[1], Sinduja Marx[6], Alex Kang[1,5], Alyssa Q. Stiving[7]§, Sophie R. Harvey[7], Dagan C. Marx[8], G. Nasir Khan[3], Karen G. Fleming[8], Vicki H. Wysocki[7], David J. Brockwell[3], Lukas K. Tamm[4], Sheena E. Radford[3], David Baker[1,2,5]¶

Transmembrane β-barrel proteins (TMBs) are of great interest for single-molecule analytical technologies because they can spontaneously fold and insert into membranes and form stable pores, but the range of pore properties that can be achieved by repurposing natural TMBs is limited. We leverage the power of de novo computational design coupled with a "hypothesis, design, and test" approach to determine TMB design principles, notably, the importance of negative design to slow β-sheet assembly. We design new eight-stranded TMBs, with no homology to known TMBs, that insert and fold reversibly into synthetic lipid membranes and have nuclear magnetic resonance and x-ray crystal structures very similar to the computational models. These advances should enable the custom design of pores for a wide range of applications.

Advances in de novo protein design have yielded water-soluble proteins of increasing complexity (*1*–*5*), and several examples of α-helical membrane proteins (*6*, *7*). However, the de novo design of an integral transmembrane β barrel (TMB) has not yet been achieved. The unassisted folding of TMBs into lipid bilayers in vitro likely involves concerted membrane insertion and folding of β hairpins (*8*, *9*), and how this is encoded in the sequences of TMBs is not well understood because of experimental challenges in characterizing their rugged folding pathways (*10*, *11*). To prevent misfolding and aggregation in vivo, an array of chaperones assist TMB folding and assembly in the outer membranes of prokaryotes, mitochondria, and chloroplasts (*12*). The lipid-folding–water-aggregation trade-off places poorly understood constraints on the global sequence properties of TMBs, slowing down the development of de novo design methods. Instead, TMB engineering has proceeded by modification of naturally occurring TMBs, which has yielded nanopores for single-molecule DNA sequencing (*13*), small-molecule sensing (*14*, *15*), or water-filtering bioinspired membranes (*16*).

To shed light on the sequence determinants of folding and stability of TMBs, and to enable the custom design of TMBs for specific applications, we set out to design TMBs de novo. We started by studying the constraints membrane embedding puts on both the backbone geometry and the sequence of transmembrane β barrels.

Geometric constraints on transmembrane β-barrel backbones

TMBs are formed from a single β sheet that twists and bends to close on itself, so that all membrane-embedded backbone polar groups are hydrogen-bonded and shielded from the lipid environment. Insertion of TMBs into the lipid membrane is oriented (*17*), with β strands usually connected with long loops on the translocating (trans) side of the β barrel (extracellular in bacteria) and short β-turns on the nontranslocating (cis) side (Fig. 1A). The β-barrel architecture is characterized by two discrete parameters: the number of strands (n) and the shear number (S)—the sum of residue offsets (register shifts) between the neighbor strands, starting at any strand and tracing around the β barrel (fig. S1A) (*18*). The ideal β-barrel radius r (eq. S1) and angle of the strands with the main barrel axis θ (eq. S2) are functions of S and n (table S1) (*19*). S and n also define the packing arrangement of side chains in the β barrel. There are S continuous strips of side chain Cβ atoms perpendicular to the β strands (fig. S1, B and C); half of these Cβ strips point toward the lumen and the other half toward the β-barrel exterior.

We focused on the simplest and smallest β-barrel architecture of eight β strands. We first considered a shear number of eight ($n = S$). In this configuration, the total register shift is distributed equally among the four β hairpins (two-residue offset between each β hairpin), and the four Cβ strips pointing toward the lumen of the barrel are arranged in fourfold-symmetric rungs with the Cα-Cβ vectors (which indicate the direction of the side chains) pointing at each other (Fig. 1B, left). This symmetric arrangement combined with a small β-barrel radius does not allow tight jigsaw-puzzle–like packing—the side chains clash with each other rather than interdigitating. To enable better packing, we broke the symmetry in the core by increasing the register shift between two β hairpins from two to four residues, resulting in a shear number of 10. In this case, there are five intertwined Cβ strips that spiral around the barrel axis, and the Cα-Cβ vectors point between rather than at each other so that the side chains can pack in a more interdigitated fourfold screw-like pattern (Fig. 1B, right).

The uneven distribution of register shifts between β hairpins complicates interactions with the lipid membrane, which can be approximated as two planes that must be parallel to ensure constant membrane thickness. In natural TMBs, the cis (periplasmic) β-turns are close to the periplasmic lipid–water boundary (fig. S2, A to D). The β-turn residues closely match the sequence preferences observed in water-soluble β barrels (mostly polar residues), but the lipid-exposed residues flanking these β-turns are predominantly hydrophobic (fig. S2, H to K) (*20*) and define the cis boundary of the transmembrane region ("membrane anchor residues"; fig. S2, A to D). The geometric challenge is that differences in the register shifts between β hairpins result in a screw-like arrangement of the four anchor residues with a translation Z along the main β-barrel axis (eqs. S3 to S5); hence, if the β-barrel axis is along the membrane normal, the anchor residues cannot all be in the same plane. The vertical offset of the anchor residues can be made more compatible with the planarity requirement by tilting the β barrel in the membrane by an angle $\alpha = \arctan(Z/C)$, where the denominator is the length of the arc between anchor residues 1 to 4 projected onto the plane perpendicular to the main axis (eq. S6) (Fig. 1C and fig. S3B). In the case of a β barrel with symmetry ($n = 8$, $S = 8$), the vertical offset between each anchor residue, approximated from the geometric model, is close to zero (supplementary text), and no tilt is required. When S is increased to 10 by increasing the register shift between one pair of hairpins to four residues, the barrel must be tilted by ~6.7° to the transmembrane axis (Fig. 1E, top) to bring the anchor residues into the same

[1]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. [2]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA. [3]Astbury Centre for Structural Molecular Biology, School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, USA. [4]Department of Molecular Physiology and Biological Physics and Center for Membrane and Cell Physiology, University of Virginia, Charlottesville, VA 22903, USA. [5]Institute for Protein Design, University of Washington, Seattle, WA 98195, USA. [6]Department of Molecular Engineering and Sciences, University of Washington, Seattle, WA 98195, USA. [7]Department of Chemistry and Biochemistry, Resource for Native Mass Spectrometry Guided Structural Biology, The Ohio State University, Columbus, OH 43210, USA. [8]TC Jenkins Department of Biophysics Johns Hopkins University, Baltimore, MD 21218, USA.
*Present address: VIB-VUB Center for Structural Biology, VIB, Brussels, Belgium.
†These authors contributed equally to this work. ‡Present address: Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, UK. §Present address: Analytical Research and Development, Merck & Co., Inc., 770 Sumneytown Pike, West Point, PA 19846, USA.
¶Corresponding author. Email: dabaker@uw.edu

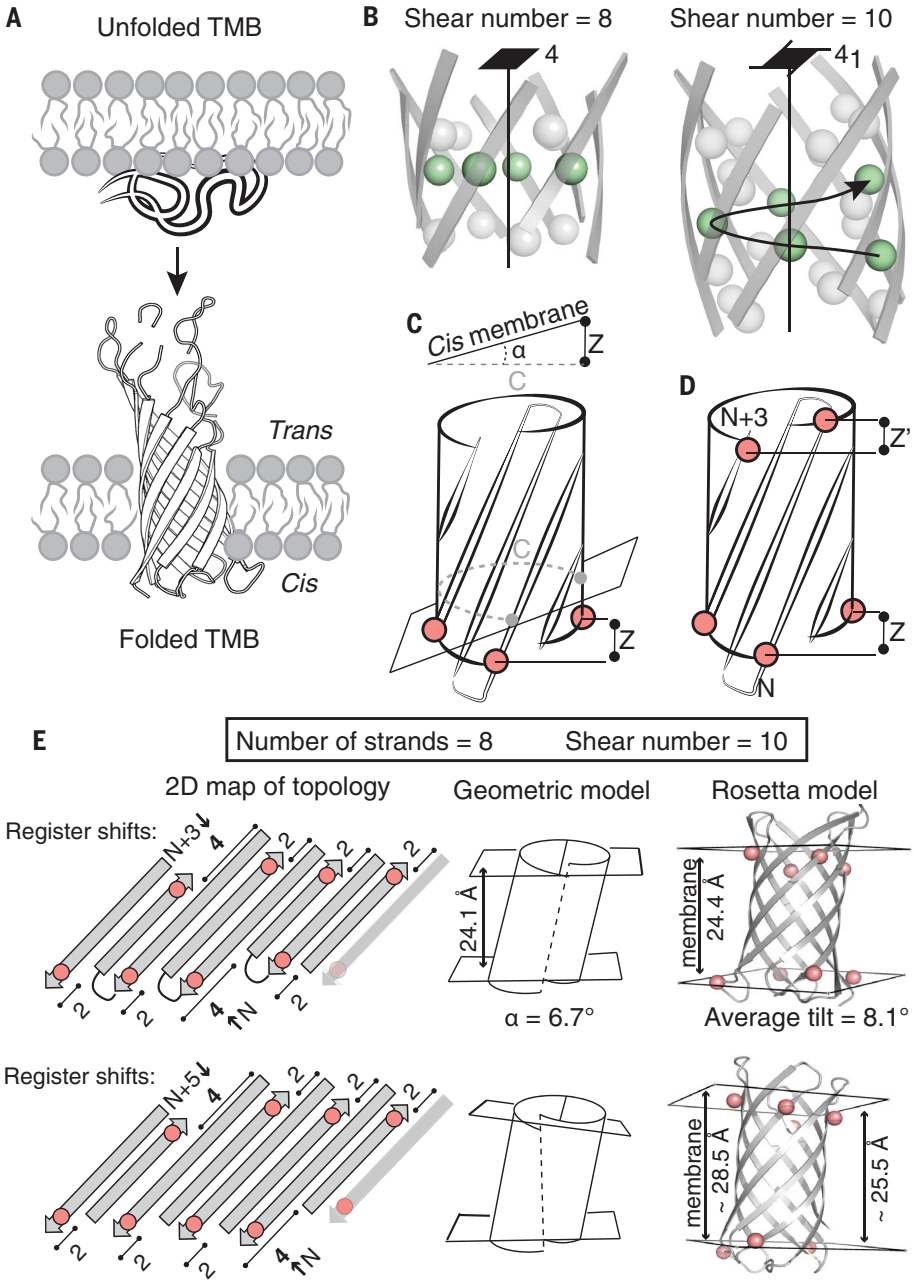


Fig. 1. Geometric principles for TMB backbone design. (**A**) The trans side of the β barrel is the side that translocates through the lipid membrane during TMB folding and insertion. The cis β-turns remain on the initial protein–lipid interaction side of the membrane. (**B**) Comparison of side-chain packing arrangements in eight-strand β barrels with shear numbers of 8 (left, fourfold-symmetric packing) and 10 (right, fourfold screw resulting in a jigsaw-like packing). In (C) to (E), the membrane-anchoring residues are shown as pink spheres. (**C** and **D**) Geometric model of membrane-association constraints on the β-barrel architecture. (C) Asymmetric register shifts between the β hairpin can be accommodated by tilting the β barrel to the transmembrane axis by an angle $\alpha = \arctan(Z/C)$. (D) The change of level Z between two anchor residues on the cis side of the β barrel must be matched by the change of level Z' between the two stacking anchor residues on the trans side. Because of β strands staggering to the main β-barrel axis, an anchor residue on the cis side of a strand N stacks with an anchor residue on the trans side of strand N+3. (**E**) The geometric model (center) and Rosetta model (right) predict similar tilt angles (α) of the β barrel to the membrane axis and constant hydrophobic thickness, for β-strand arrangements with matching Z and Z' (double register shifts located on strand N in cis and on strand N+3 in trans) (top). Both models show inconsistent hydrophobic thickness for β-barrel architectures with double register shifts located on strand N in cis and on strand N+5 in trans (bottom).

plane. To test the validity of this simple geometric model, we explicitly assembled TMB backbones using Rosetta (*21*) and predicted their placement in the membrane [materials and methods (*22*)], which yielded an average tilt angle (8.1°; Fig. 1E, top) close to that of the geometric model. Placing the four-residue register shift after each of the four cis hairpins resulted in tilts with similar amplitude but different directions relative to the membrane axis (supplementary text and fig. S4, B to G); we focused on the placement in which the four-residue register shift is in the middle of the β sheet.

We next investigated the structural consequences of the fact that the planes representing the cis and trans membrane boundaries must be roughly parallel to each other to keep the hydrophobic thickness constant. To achieve this, the offset Z between any two neighbor anchor residues on the cis face must be matched by a similar offset Z' between the anchor residues above it on the trans face. For barrel topologies of (n = 8, S = 10) spanning a membrane of 24 Å, an anchor residue on the cis side of strand N stacks along the main β-barrel axis with the anchor residue on the trans side of strand N+3 because of the staggered β strands (Fig. 1D and supplementary text). Hence, to maintain constant thickness, the register shift between strands N and N+1 on the cis side must be equal to the register shift between strands N+3 and N+4 on the trans side. To confirm this prediction of our geometric model, we set the cis side register shift between strands N and N+1 to four residues and ran Rosetta design simulations and transmembrane plane predictions on backbones with a matching four-residue register shift on the trans side between (i) strands N+3 and N+4 and (ii) strands N+5 and N+6. We averaged planes representing the membrane boundary in cis and trans and found, consistent with the model, parallel planes and constant hydrophobic thickness for the N+3 case (i) but a 3-Å change of thickness in the N+5 case [(ii), Fig. 1E, bottom].

We used this constant hydrophobic thickness constraint to guide the distribution of the register shifts around the β barrel. The cis hairpins were closed with short β-turns associated with a β bulge, which match the local twist of the β strands [these are abundant in water-soluble (*5*) and transmembrane β barrels (fig. S2A)]. On the trans side, the strands were connected with canonical β-turn sequences with strong β-hairpin nucleating properties [3:5 type I β-turns + G1 bulge with canonical Ser-Asp-Gly (SDG) sequence (*21*–*23*)] in place of the long loops found in native TMBs; such turns were previously used to design water-soluble β barrels (fig. S2, E and G) (*5*). To relieve strain from high β-sheet curvature, we placed glycine kinks (*5*)—glycine residues in a

fully extended conformation within β strands—into the blueprint such that (i) every Cβ strip pointing to the core of the barrel contains a glycine and (ii) no Cβ strip contains more than four nonglycine residues in a row (¼ of the average barrel circumference). The glycine kinks in the designed β-barrel blueprint stack along four vertical lines (Fig. 2, A and B) such that the resulting Rosetta models have four regions of strong β-sheet bending that delimit a lumen with a distinctive square-shaped cross-section not observed in naturally occurring β barrels (fig. S2F).

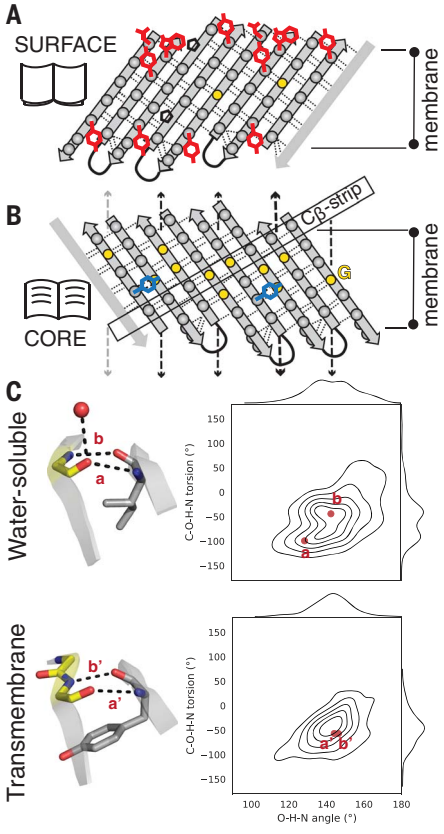


Fig. 2. Sequence features defining de novo TMB fold and shape. (**A** and **B**) Two-dimensional schematic representation of the connectivity (hydrogen bonds as dashed lines) between β strands in the TMB designs. Side chains are shown as gray spheres and glycine residues as yellow dots. Aromatic girdle motifs are shown in red, tyrosines of the mortise–tenon motifs in blue, and prolines as black pentagons. Glycine kinks were arranged to bend the β sheet into four corners (vertical arrows). (**C**) Hydrogen bond geometries between pairs of residues involving a glycine kink. Left: Examples from crystal structures of water-soluble (PDB ID: 6CZH) and transmembrane β barrels (PDB ID: 1BXW). Glycine residues are in yellow, and water molecules are shown as red dots. Right: Distributions of the C-O-H-N and O-H-N angle values describing the hydrogen bond geometry in β-barrel crystal structures.

Sequence design and initial experimental characterization

To delimit the upper and lower membrane boundaries, four tyrosine residues were placed two positions upstream of the anchor residues on the cis side, and alternating tyrosine and tyrosine-tryptophan motifs were placed at the trans boundary [Fig. 2A and supplementary text; such "aromatic girdles" are observed in native TMBs (*24*)]. To design the remainder of the sequence, we first experimented with the approach that we took for helical transmembrane proteins (*6*), using standard Rosetta design methods to design core residues (which results in largely hydrophobic interiors, as in helical transmembrane proteins) and resurfacing the outside with hydrophobic residues. However, this resulted in sequences that had strong amyloid propensity (fig. S5). To reduce the amyloid propensity and the hydrophobicity [native TMBs are usually less hydrophobic than α-helical membrane proteins (*25*)], we experimented with requiring all residues in the interior of the barrel (excluding the glycine kink positions) to be polar and the surface residues to be hydrophobic, resulting in the hydrophobic-polar sequence pattern characteristic of the β-sheet secondary structure but inside-out compared to water-soluble β barrels. To help define the register between β strands we placed tyrosine residues adopting (+60,90) rotamer angles to closely interact with the groove formed by a neighbor glycine kink (*26*) [the "mortise–tenon" motif (*27*, *28*)]. We placed two such motifs in the regions where defining the register shift seemed likely to be particularly important: Tyr^{69} (Y69) on strand 5, where the four-residue register shifts in cis and trans produce a larger vertical offset in the β sheet, and Tyr^{11} (Y11) on strand 1, where the β sheet closes on itself but lacks a register-defining β-turn between the first and last strands (Fig. 2B and fig. S6). Finally, we designed full β-barrel sequences using Rosetta combinatorial sequence design and the ref2015 energy function (*29*) with increased weight on the electrostatics term to favor side-chain–side-chain hydrogen bonds in the core of the β barrels. As expected, the secondary structure of the resulting designs was accurately recapitulated by secondary-structure prediction programs (Fig. 3A).

Folding of TMBs is chaperone-mediated and catalyzed in vivo by the β-barrel assembly machinery (BAM) complex in Gram-negative bacteria, the sorting and assembly machinery (SAM) complex in mitochondria, and the OEP80 insertase in the outer chloroplast membrane (*30*). Because it was unclear whether our TMB designs would be able to interact with the chaperone machinery to fold in the outer membrane of *Escherichia coli*, we expressed the designed sequences in the cytoplasm, anticipating that they would form inclusion bodies that could then be solubilized in urea–guanidinium chloride [both natural and engineered TMBs have been produced in this way (*31*)]. We obtained *E. coli* codon-optimized synthetic genes for nine designs (set TMB0, fig. S7), but no protein of the correct molecular weight was produced upon the induction of protein expression (data S2). Reasoning that the designed sequences may have had too much positive charge, which can impair translation (*32*), in a second round of 16 designs, we reduced the number of charged residues in the core of the protein (set TMB1, fig. S7). Again, none was expressed in *E. coli* (data S2).

Because of the failures at the expression stage, experimental feedback to improve the design methodology could not be obtained. To gain insight, we instead compared our designs to sequences of natural eight-strand TMBs. We noted two differences: First, the natural TMBs often have long and disordered trans loops rather than short β-turns (*20*), and second, the secondary-structure propensity of their transmembrane β strands was lower than that of the designs we had tried to express (Fig. 3A). We hypothesized that the high β-turn and/or β-strand propensities of our designed sequences could result in rapid formation of off-target β-sheet structures when expressed in the cytoplasm, which could be cleared rapidly or hinder growth of expressing cells.

Role of trans β-turns

We first explored the role of the trans loops in TMB folding and expression by redesigning the native TMB of the protein OmpA (tOmpA), replacing its trans loops with the canonical SDG β-turn sequence used in our designs (fig. S8, A and B). The relooped tOmpA construct (OmpSDG) was expressed at high levels in *E. coli* (where it was found in inclusion bodies), but it could not be correctly refolded (Fig. 3C and fig. S10, C and D). To understand this observation, we carried out Rosetta energy landscape calculations on short β-turns at the trans membrane boundary of natural TMB structures and observed that their sequences have relatively low propensity to form β-turn structures (supplementary text) compared to the β-turns of soluble β barrels and the SDG β-turns of OmpSDG. We constructed and tested four variants of tOmpA (Omp*Trans*1-4) that each contain two such 3:5 type I β-turns with suboptimal sequences [these designs are shorter than the shortest variant of tOmpA previously reported, which has trans connections of 5 to 18 residues (*33*)]. The proteins were again expressed at high levels in inclusion bodies (table S2), but this time all four of these sequences showed a heat-modifiable band [analyzed by cold SDS–polyacrylamide gel electrophoresis (PAGE)] when folded into n-dodecyl-β-D-maltoside DDM

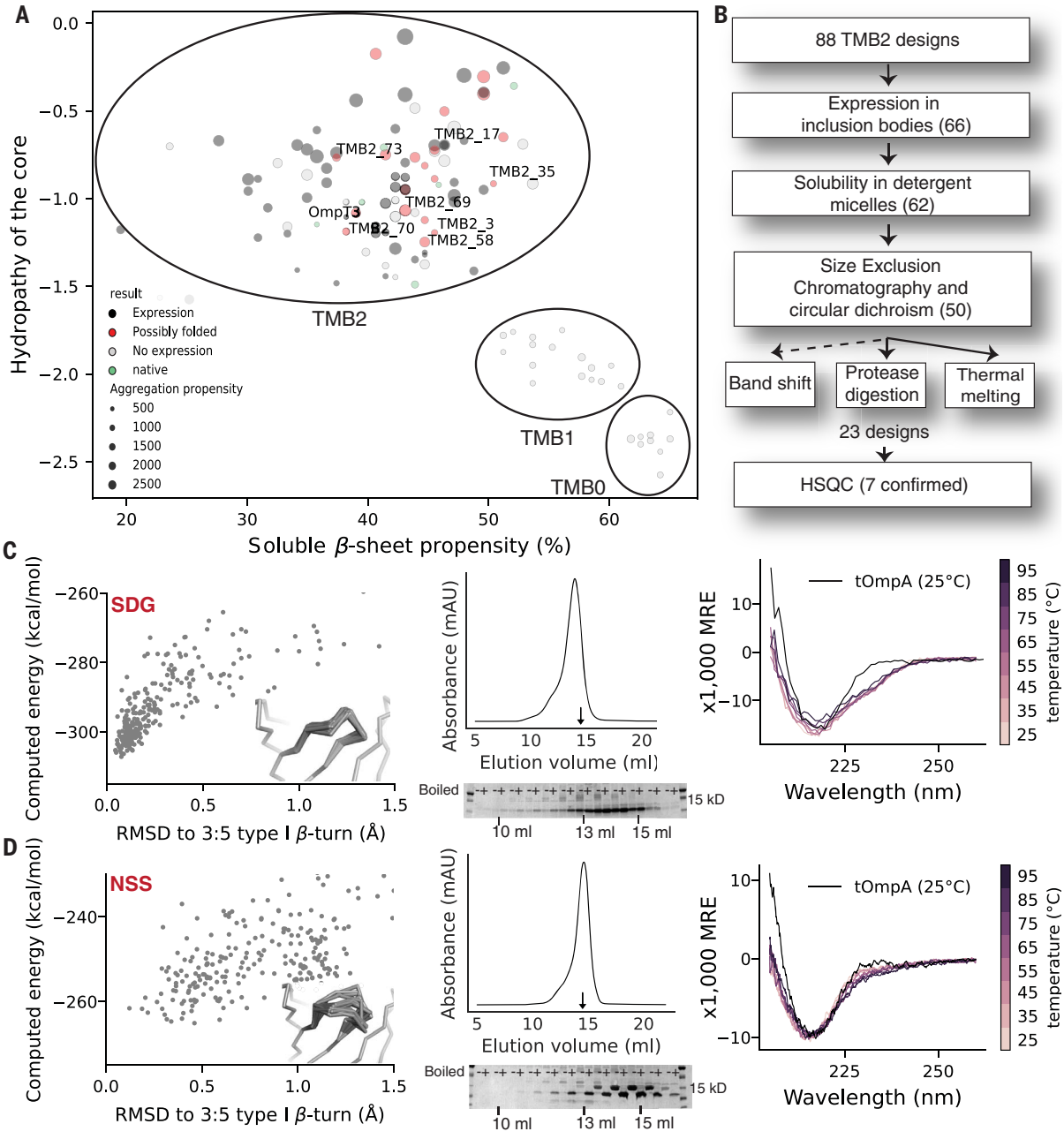


Fig. 3. Negative design is critical for de novo TMB folding. (**A**) Successful design of TMBs requires reducing β-sheet propensity of the transmembrane β strands. *x* axis, β-sheet propensity of the transmembrane region [calculated with RaptroX (*59*)]; *y* axis, hydrophobicity of the core [GRAVY hydropathy index (*60*)]. Gray spheres, nonexpressing TMB designs (repeated twice); black circles, expressing designs that do not fold; red, TMB designs that pass biochemical folding screening (repeated in two different detergents)—labels indicate that the folded species was validated by HSQC; green, naturally occurring TMBs with eight strands. Circle size indicates aggregation propensity of the sequence predicted with TANGO (*61*). (**B**) Experimental workflow. The number of TMB2 designs satisfying each criterion is shown in parentheses. (**C** and **D**) Proper folding of tOmpA requires negative design against strong β-turn nucleating sequences on the trans side. Left: Rosetta energy landscapes of designs with canonical low energy (C) or suboptimal (D) sequences substituted in a 3:5 type I β-turn with a G1 β-bulge. Conformational perturbations were generated with kinematic loop closure (*62*); the inset shows the backbone conformations of the 25 lowest-energy models. Center: After refolding in 2× CMC DDM detergent, Omp*Trans*3 (bottom panel) elutes on SEC similarly to tOmpA (arrow, 14.62 ml for OmpTrans3 and 14.53 ml for tOmpA) and runs as a heat-modifiable species on SDS-PAGE characteristic of folded tOmpA, whereas (top panel) the OmpAAG peak elutes earlier (13.96 ml) and does not show a band shift (band shift assay repeated three times; SEC repeated two times). Right: The far-UV CD spectrum of Omp*Trans*3 (bottom panel), but not OmpAAG (top panel), is similar to that of tOmpA (repeated two times).

detergent micelles and large unilammelar vesicles (LUVs), characteristic of properly folded tOmpA (Fig. 3D and fig. S10, C and D). The best expressed design, Omp*Trans*3, was characterized in more detail; following refolding in detergent micelles, it had a retention time similar to that of native tOmpA on a size exclusion chromatography (SEC) column (Fig. 3D and fig. S12), a similar native mass spectrometry (nMS) profile (fig. S13), well-dispersed resonance peaks by ^{1}H-^{15}N–heteronuclear single-quantum coherence nuclear magnetic

resonance (HSQC NMR) in Fos-choline-12 (DPC) detergent (fig. S10B) and a circular dichroism (CD) spectrum similar to that of tOmpA in both DDM micelles (Fig. 3D) and in 1,2-diundecanoyl-*sn*-glycero-3-phosphocholine (DUPC, *di*$C_{11:0}$PC) LUVs with distinctive peaks at ~220 and 231 nm (fig. S10A) (*34*). These data suggest that either long loops or short suboptimal β-hairpin loops on the trans side are necessary to slow down nucleation of the trans β-hairpins and allow proper folding of TMBs in vitro. However, simply replacing the trans loops on four of the TMB0 designs with the extracellular loops of tOmpA or scrambled versions of these loops did not appreciably increase protein expression (fig. S9), suggesting that there was a further problem with the properties of the transmembrane β strands in the original designs.

Reducing β-sheet propensity

We sought to further use negative design (*35*, *36*) to disfavor off-target states and slow down folding, this time through reduction of the high secondary-structure propensity of the β strands in our designs. We increased the hydrophobicity of the side chains in the β-barrel lumen, thereby disrupting the strict alternation of polar and hydrophobic residues along the β strands. To do so, we experimented with the design of networks of hydrogen bonds surrounded with scattered hydrophobic patches. We extended the mortise–tenon motifs to include a hydrogen bond between the tyrosine and a negatively charged Asp or Glu residue to seed the design of hydrogen bond networks. Possible positions for the Asp or Glu were exhaustively searched using the Rosetta HBNet protocol (*37*) to design β-barrel backbones with preinstalled Tyr-Gly-Asp/Glu (YGD/E) motifs at one or both of the locations identified above. We used Rosetta combinatorial sequence optimization to design the remainder of the positions facing the core of the barrel, allowing all 18 amino acids other than Cys and Pro.

To further lower the β-sheet propensity, we experimented with incorporation of glycine residues [which destabilize β strands (*38*)] on the hydrophobic outer surface of the β barrels. To guide placement of the glycines, we compared crystal structures of natural TMBs with those of water-soluble β barrels, which rarely have water-exposed surface glycine (supplementary text). We found that the extended backbone conformation of core glycine kinks in the water-soluble β barrels results in noncanonical out-of-plane backbone hydrogen bond geometry characteristic of a left-hand twist (O–H–N angle ~130°; C–O–H–N dihedral ~−100°; Fig. 2C, top, and fig. S14A, center), whereas the surface residues preceding the glycine kink have a more pronounced right-hand twist (C–O–H–N dihedral >0°, fig. S14A, right) than canonical in-plane backbone hydrogen bonds (O–H–N angle ~155°; C–O–H–N dihedral ~0°; fig. S14A, left) (*39*). The backbone carbonyls of glycine kinks involved in extremely out-of-plane hydrogen bonds (resulting in strongly bent β strands) in water-soluble β barrels are often exposed to solvent to interact with a water molecule or other hydrogen bond donor in crystal structures (fig. S15, E and F). Such exposed carbonyls are likely disfavored on the lipid-buried surface of TMBs because there are no water molecules in the bilayer to stabilize them, and TMBs indeed have a smaller population of glycine kinks and preglycine hydrogen bonds deviating from in-plane geometry (Fig. 2C, bottom, and fig. S14B, center). We hypothesized that glycines in positions preceding glycine kinks could allow more canonical in-plane hydrogen bonds and hence reduce unfavorable surface exposure of the carbonyls to the apolar lipid environment, and we confirmed this with explicit Rosetta design calculations (supplementary text). In the subsequent sequence design calculations, we identified strongly bent glycine kinks in the β-barrel blueprint and placed glycines in surface-exposed positions directly preceding them.

Negative design of loops and strands enables de novo TMB design

We carried out three iterations of core and surface design according to the above principles using the suboptimal loops from Omp*Trans3* on the trans side, allowing the backbone to relax on the basis of the current sequence by gradient-based energy minimization at each step (glycine placement, in particular, allows local backbone rearrangement). The design calculations converged on 20 distinct core hydrogen bond network architectures with overall amino acid composition similar to that of natural eight-strand TMBs (fig. S7D). Codon-optimized synthetic genes were obtained for several representatives of each network architecture for a total of 88 designs (set TMB2). In sharp contrast with the lack of expression in our previous unsuccessful design rounds, 66 of these designs were well expressed in inclusion bodies, as intended. To test the influence of the trans loops on expression, we expressed variants of 20 of these designs incorporating the extracellular loops of tOmpA. The same designs expressed or did not express with the short suboptimal β-turns or the long tOmpA loops (data S2), indicating that the transmembrane β strands—rather than the β-strand connections—carry the main sequence determinants of cytoplasmic expression (the turn sequence does matter for subsequent assembly and membrane insertion, as exemplified by the failure of OmpSDG to fold properly).

Characterization of folding, stability, and structure

To test the ability of the designs to stably fold to TMB structures in vitro, we followed procedures used to fold tOmpA and other natural TMBs (Fig. 3B) (*40*, *41*). Briefly, the inclusion bodies were dissolved in 8 M urea and rapidly diluted into DDM, DPC, or *n*-octyl-β-D-glucopyranoside (OG) detergents at twice the critical micelle concentration (2× CMC) (data S4). Out of the 66 expressing designs, 62 formed soluble species in such conditions. We purified the protein–detergent complexes by SEC and characterized the 50 designs that had a SEC retention volume expected for a monomeric TMB (similar to the eight-stranded tOmpA monomer and Omp*Trans3*) and a far-ultraviolet (UV) CD spectrum characteristic of a β-sheet protein. The band-shift assay used to monitor native TMB folding (*42*) was uninformative for the identification of folded de novo–designed TMBs. Instead, we found good agreement between the resistance of a design to protease digestion and thermostability up to 95°C of the characteristic β-sheet far-UV CD spectrum. In total, 23 designs satisfied the biochemical screening criteria, suggesting that they fold into a TMB structure. Eleven such designs were randomly selected for analysis by ^{1}H-^{15}N HSQC NMR in DPC detergent micelles, and seven had well-dispersed chemical-shift profiles, characteristic of a folded protein in this detergent (figs. S16 and S17, validated designs; fig. S18, designs that failed biochemical tests; fig. S19, designs that passed the biochemical tests but appear misfolded by NMR).

We selected two de novo designs, TMB2.17 (highest BLAST E-value to the nonredundant protein database: 0.10) and TMB2.3 (BLAST E-value: 0.035) and the Omp*Trans*3 construct for detailed biophysical characterization in a lipid bilayer to determine whether the proteins exhibit the expected properties for a membrane-spanning β barrel (using tOmpA as a control). After refolding into 100 nm DUPC LUVs, all four proteins had far-UV CD spectra characteristic of a β-sheet protein both in 0.24 and 2 M urea, and distinct from the spectra of both the fully unfolded proteins in 8 M urea and the proteins refolded in the absence of lipid (figs. S10A and S20). We next determined the stability of the folded proteins by monitoring their ability to fold into or unfold out of LUVs at increasing urea concentrations, monitored by the change of fluorescence intensity between water-exposed and lipid-embedded surface tryptophans (*43*). The designed TMB proteins are more thermodynamically stable [midpoint urea concentration for folding (Cm^F) 5.7 and 7.2 M for TMB2.3 and TMB2.17, respectively; Fig. 4A) than tOmpA (Cm^F = 4.7 M), and Omp*Trans*3 is the most stable protein as it appears folded even in 9 M urea (fig. S21), in agreement with the far-UV CD data. It has been previously shown that the folding and unfolding transitions of many natural TMBs exhibit hysteresis due to the high kinetic barrier to unfolding and extraction from the membrane environment (*11*, *44*, *45*).

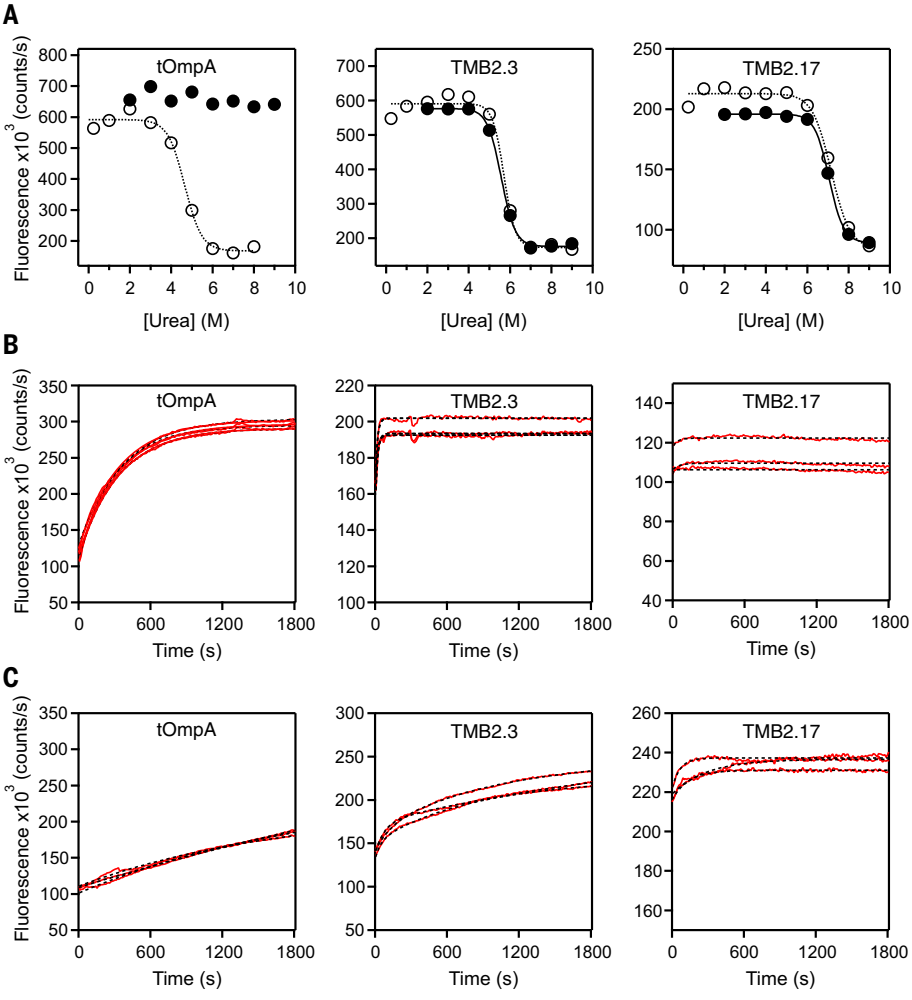


Fig. 4. Folding of de novo–designed TMB2.3 and TMB2.17 compared to tOmpA in synthetic lipid membranes. (**A**) Urea dependence of folding and unfolding in DUPC LUVs. The fluorescence intensity at 335 nm was plotted against urea concentration to determine the midpoint urea concentration for folding (C_m^F) (open circles, dashed line) and unfolding (C_m^{UF}) (filled circles, solid line). (**B** and **C**) Kinetics of folding into (B) DUPC and (C) DMPC LUVs at a lipid to protein ratio of 3200:1 (mol/mol) in 50 mM glycine-NaOH (pH 9.5), 2 M urea at 25°C monitored by tryptophan fluorescence at 335 nm over 30 min (red line). Data were fitted with a single exponential function to determine folding rate constants (black dashed line). Three technical replicates each.

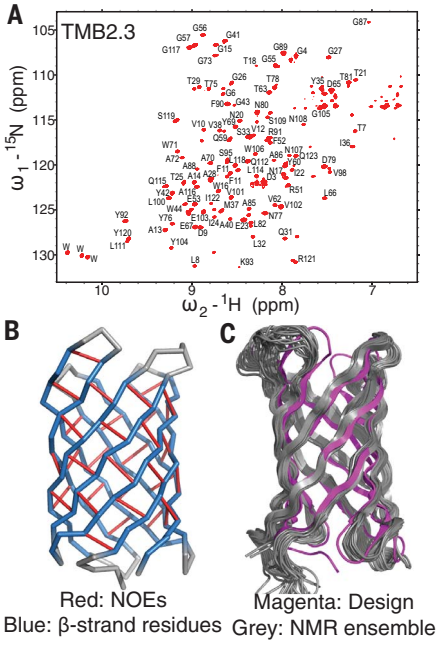


Fig. 5. NMR structure of TMB2.3 in DPC detergent micelles. (**A**) Assigned ^{15}N-^{1}H TROSY (transverse relaxation–optimized spectroscopy) spectrum of TMB2.3. (**B**) NMR constraints mapped on the TMB2.3 design model. Residues predicted to have β-sheet secondary structure are colored in blue. Collected interresidues NOEs are shown as red sticks. (**C**) TMB2.3 Rosetta design model (magenta) aligned to the 20 lowest-energy models generated with NMR constraints (gray).

Under the conditions tested here, this behavior was observed for tOmpA but not for the designs TMB2.3 and TMB2.17, which showed superimposable and reversible unfolding and folding transitions, suggesting reduced kinetic stability relative to tOmpA. These observations likely explain the lack of a band shift in SDS-PAGE: The de novo designs unfold during electrophoresis because of lower kinetic barriers to unfolding (*46*) (fig. S22). The equilibrium unfolding curves for TMB2.3 and TMB2.17 fitted well to a two-state transition (fig. S23) with unfolding free energies (ΔG^0_{UF}) of 38 and 56 kJ mol^{-1}. These ΔG^0_{UF} values fall within the range of natural TMBs [ΔG^0_{UF} 10 to 140 kJ mol^{-1} (*43*, *47*–*49*)].

The designed TMBs fold more than an order of magnitude more rapidly than tOmpA [(*50*); folding rate constant of 3×10^{-3} s^{-1} for tOmpA]—too rapid to allow accurate measurement of the folding rate constant (Fig. 4B). Tryptophan fluorescence emission spectra of the end point of the folding reactions confirmed that the TMBs were indeed fully folded (fig. S24). To confirm that the designs integrate into the lipid bilayer rather than folding on the lipid surface or in the absence of lipid, proteins dissolved in 8 M urea were diluted into 2 M urea without lipid or into LUVs composed of 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC, *di*$C_{14:0}$PC). Consistent with previous results showing that the folding rates of natural TMBs are inversely correlated with lipid chain length (*9*, *51*), the designed TMBs fold more slowly into lipids of longer acyl chain length (Fig. 4C) and do not fold in the absence of lipid (fig. S25B), confirming that they indeed integrate into the lipid bilayer.

To characterize the structure of the designed TMBs in solution, we solved the structure of TMB2.3 folded into DPC detergent micelles using NMR spectroscopy (table S3). Resonance peaks for 107 of the 117 nonproline residues of TMB2.3 were fully assigned; 6 more were partially assigned (Fig. 5A and fig. S26A). Four out of six nonassigned residues are located in the trans β-turn regions—the remaining two are the N- and C-terminal residues of the protein. The secondary-structure TMB2.3 calculated with TALOS-N (*52*) consists of eight β strands that closely match the β-strand boundaries in the designed model (fig. S26C). Nine out of 11 glycine residues pointing toward the core of the β barrel (glycine kink residues) have the designed torsional irregularities based on the positive Cα chemical shifts (*41*) (fig. S27, A and B) and the more extended predicted backbone conformations (ϕ and Ψ closer to 180°; data S5). To validate the residue connectivity between the β strands, we collected a total of 81 unique nuclear Overhauser effects (NOEs) between amide protons; these suggest 72 interstrand backbone hydrogen bonds that are in agreement with the β-strand

connectivity of the design and the shear number of 10 across the β barrel (Fig. 5C and fig. S26D). The NMR structure ensemble generated on the basis of the chemical shifts and NOE information agrees closely with the design model [average backbone root mean square deviation (RMSD) of 2.2 Å, Fig. 5B]. We observed low-intensity additional resonance peaks for a subset of residues, indicating the presence of a (minor) secondary conformation. The secondary signals strong enough for analysis were consistent with the secondary-structure assignment and NOEs of the main conformation, indicating that the secondary conformation does not involve modification of the β-barrel architecture. The residues with double peaks cluster in the cis region of strands 1, 2, and 8 (fig. S26B); these could result from close proximity to the flexible N terminus or transient dimeric interactions identified by native mass spectrometry in detergent micelles (figs. S28 and S29).

To determine structure at the atomic level, we crystallized TMB2.17 and solved the structure at 2.05-Å resolution (table S4). All but two residues located in one trans β-turn were resolved in the electron density map. The crystal structure of TMB2.17 closely matches the design model (1.1-Å backbone RMSD over all residues, Fig. 6A), and the β barrel has a wide lumen delimited by glycines in an extended conformation, which form kinks in the β strands as designed (Fig. 6, B and C). The two YGD/E interactions (Y69, Y11, G27, G89, D39, E103) belonging to the extended mortise–tenon motifs are present in the crystal structure, and the second shell of interactions, involving K71, E53, and Q29, is also properly recapitulated with additional interactions to water molecules (Fig. 6D); these extended side-chain hydrogen bond networks fill the lumen of the β barrel. Overall, the buried amino acid side-chain conformations and interactions in the design model are in very good agreement with the crystal structure (Fig. 6E; compare pink and gray). We compared TMB2.17 to the transmembrane region of tOmpA, the only natural TMB sequence with a known structure returned by a BLAST search for sequences similar to TMB2.17 in the nonredundant sequence database (the E-value of 1.6 is in the range expected from random matches; alignment shown in fig. S30). The shape of the β-barrel lumen is quite different in the two proteins (Fig. 6B), as are the amino acid identities and packing arrangements of the core side chains (Fig. 6F).

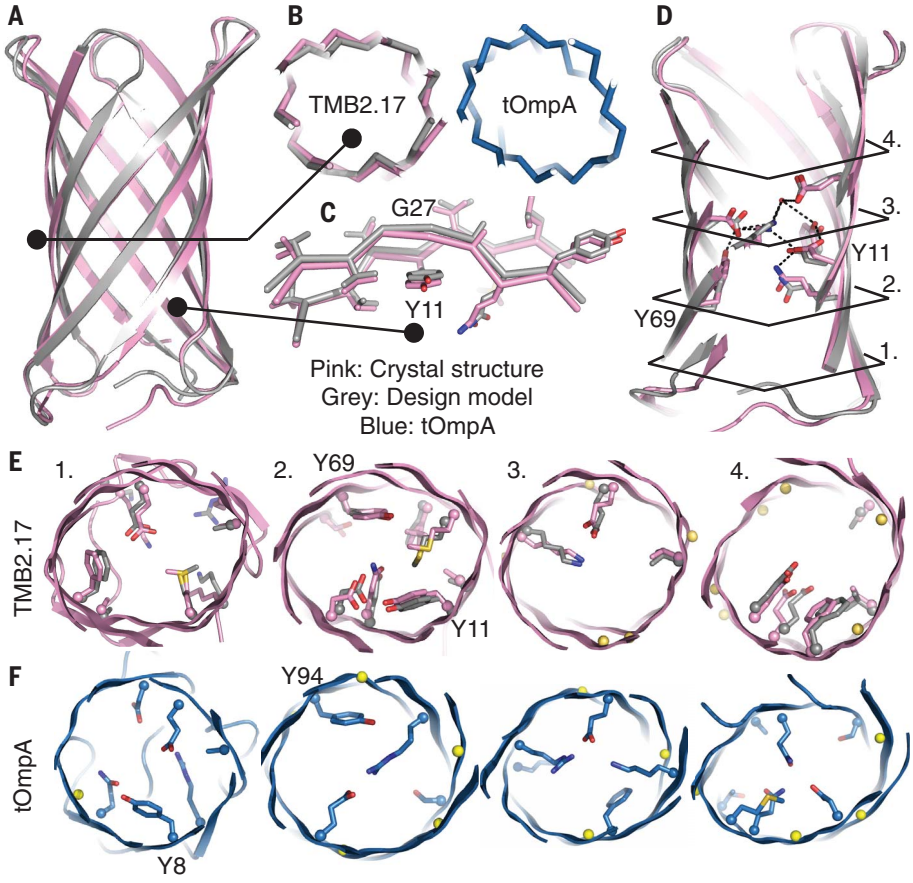


Fig. 6. Crystal structure of TMB2.17 is nearly identical to the design model. The crystal structure (pink) of TMB2.17 determined in DPC detergent, superimposed on the design model (gray), and compared to the crystal structure of the naturally occurring tOmpA (blue; PDB ID, 1BXW). (**A**) Full backbone superposition. (**B**) Comparison of transverse β-barrel cross-section geometries. (**C**) Superposition of the β strands around a mortise–tenon motif, showing the extended backbone conformation of the glycine kink (G27) and the rotamer of the tyrosine involved in the aromatic rescue interaction (Y11), which are nearly identical in crystal structure and design model. (**D**) Superposition of the side chains involved in the core network of polar interactions around the two mortise–tenon motifs. The black lines indicate the locations of the four transverse slices for which core packing is shown in (**E**) for the design model and crystal structure; the two are very similar. Corresponding slices in tOmpA (**F**) are quite different in both shape and amino acid composition and placement. Cα atoms are shown as spheres, and glycine kink residues are colored in yellow; the positions of the tyrosines in the mortise–tenon folding motifs are labeled.

Conclusions

Both the initial failures and the ultimate success of our hypothesize, design, and test approach to de novo TMB design inform our understanding of the sequence determinants of TMB folding and structure. The sequential approach previously used to build helical transmembrane proteins (*6*)—the design of proteins with hydrophobic cores compatible with folding of water-soluble proteins and subsequent hydrophobic residue resurfacing to convert them to membrane proteins—yielded sequences strongly predicted to form amyloid. Designs with more polar cores, which had high β-sheet propensity because of the perfect alternation of hydrophobic and polar residues, systematically failed to express in *E. coli*. Iterative improvement of the design protocol ultimately enabled the generation of a set of sequences, with more than 8% of sequences encoding proteins able to adopt a β-barrel fold (based on ^{1}H-^{15}N-HSQC NMR). The NMR structure and high-resolution crystal structure of two of these designs are very close to the design models. The key to this success was introducing glycine kinks, β bulges, and register-defining side-chain interactions—also critical for the folding of water-soluble β barrels (*5*, *53*) and hence important for defining β-barrel architecture irrespective of the solvent environment—and balancing the hydrophobicity and β-sheet propensities of the sequences.

Our results suggest that, to enable TMB expression and folding, the β hairpins of outer membrane β barrels need to be sufficiently unstable in water that they do not form off-target β-sheet–containing species and become populated at high levels only in the context of the fully folded state in the hydrophobic environment of the membrane. Slowing down the folding and assembly of trans hairpins could also allow more time for passage of the mostly hydrophilic amino acids in these

β-strand connections across the lipid membrane, which likely has a large activation barrier. The overall β-sheet propensity and hydrophobicity of our successful designs are in the range of those of naturally occurring TMB sequences, suggesting that the natural TMBs might be under a similar negative selection pressure against formation of non-native β-sheet structures in an aqueous environment (*54*). Further work is required to determine whether the design principles applied here enable TMB folding into biological membranes [whose properties present a formidable kinetic barrier to folding (*55*)]. In Gram-negative bacteria, the BAM complex is responsible for accelerating the assembly of natural TMB substrates into the outer membrane by lowering the kinetic barrier to folding (*55*). Our design strategy incorporates neither signals for BAM complex association, such as the conserved β signal (*56*), nor evolution-conserved functional motifs and hence represents a "blank slate" for probing the trade-offs between TMB folding, stability, and function, as well as the evolutionary constraints on OMP trafficking and biogenesis.

Larger TMBs share similar sequence properties with the eight-strand TMBs considered in this study (supplementary text), suggesting that the general design principles and methods that we have described here should be applicable to the design of larger pore-containing β barrels, after the generalization of the β-barrel architecture definition rules (glycine kinks, β bulges, etc.) to different combinations of n and S. The extent to which essentially all of the key design features are recapitulated with atomic-level accuracy in the crystal structure of TMB2.17 suggests considerable control over TMB structure, which should enable custom design of transmembrane pores with geometric and chemical properties tailored for specific applications.

Materials and methods summary

The complete material and methods section is available in the supplementary materials.

De novo protein design

The protein backbones were generated from blueprints and constraint descriptors using the Rosetta BluePrintBDR (*57*). The lowest-energy backbone models were used to guide combinatorial sequence optimization. For each of the tested models, the desired amino acid composition was achieved (i) by modifying the reference energy function (*29*) (weights of individual terms and reference weights of amino acids) and (ii) by biasing sampling using HBNet (*37*), TaskOperations, and constraints.

Protein production and screening

The designs were ordered as codon-optimized synthetic genes (Integrated DNA Technologies) and transformed into One Shot BL21 Star (DE3) chemically competent *E. coli* (Invitrogen). Protein expression in inclusion bodies was induced overnight at 37°C in Studier autoinduction medium. The inclusion bodies were washed with 0.1% (w/v) of Brij-35 and at least three cycles of pelleting and resuspension. The proteins were solubilized at 80 μM concentration in 6 M urea or 8 M guanidinium chloride and were refolded by dilution (drop by drop to 4 μM final concentration) into stirred refolding buffer [20 mM Tris, 150 mM NaCl, 0.1 % (w/v) DPC (pH 8.0)]. The solution was stirred overnight at room temperature. The protein-detergent complex was further purified by SEC (Superdex 200 increase 10/300 GL, Cytiva).

Tryptophan fluorescence

Tryptophan residues were incorporated in the designs at the trans lipid–water boundary. The change in intrinsic tryptophan fluorescence was monitored (PTI QuantaMaster, Photon Technology International) as the proteins [originally dissolved in 50 mM glycine-NaOH (pH 9.5), 8 M urea] fold into LUVs. For equilibrium studies, a lipid-to-protein ratio (LPR) of 600:1 (mol/mol) was used, and a final protein concentration of 0.4 μM. TMB folding and unfolding were allowed to proceed overnight at 25°C. Kinetics of TMB folding were measured at a final protein concentration of 0.4 μM and an LPR of 3200:1 (mol/mol). The TMBs were rapidly diluted 20-fold from 8 M urea and mixed into LUVs, and fluorescence emission was monitored at 335 nm over 30 min at 25°C.

NMR

NMR spectra were collected on a Bruker Avance 800 MHz spectrometer equipped with a cold-probe. For backbone assignments of the TMB2.3, TROSY 3D experiments [HNCA, HN(CA)CB, HNCO, HN(CA)CO] were collected on a ^{2}H, ^{13}C, ^{15}N-labeled sample with a nonuniformed sampling (NUS) technique. Two 3D NOE experiments, ^{15}N-^{15}N-^{1}H HSQC-NOESY-HSQC and ^{15}N-^{1}H-^{1}H NOESY-TROSY, were performed with mixing times of 120 ms, also in the NUS mode. In addition, a TROSY-based 2D ^{1}H-^{15}N heteronuclear NOE experiment was collected with a saturation recovery delay of 5 s with an interleaved approach.

Crystallography

Diffraction-quality crystals of TMB2.17 in DPC appeared in classic vapor diffusion method in 0.1 M Tris at pH 8.5 and 10 % (w/v) PEG8000 (MemStart+MemSys HT, Molecular Dimensions). Diffraction data were collected at the Advanced Photon Source (APS), 24ID-E beamline, with a Dectris EIGER 16M detector. Starting phases were obtained by molecular replacement using the designed model. Model bias was reduced using phenix.autobuild (*58*) with rebuild-in-place set to false and with simulated annealing and prime-and-switch phasing.

REFERENCES AND NOTES

1. R. A. Langan *et al.*, De novo design of bioactive protein switches. *Nature* **572**, 205–210 (2019). doi: 10.1038/s41586-019-1432-8; pmid: 31341284
2. A. H. Ng *et al.*, Modular and tunable biological feedback control using a de novo protein switch. *Nature* **572**, 265–269 (2019). doi: 10.1038/s41586-019-1425-7; pmid: 31341280
3. D.-A. Silva *et al.*, De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186–191 (2019). doi: 10.1038/s41586-018-0830-7; pmid: 30626941
4. E. Marcos *et al.*, De novo design of a non-local β-sheet protein with high stability and accuracy. *Nat. Struct. Mol. Biol.* **25**, 1028–1034 (2018). doi: 10.1038/s41594-018-0141-6; pmid: 30374087
5. J. Dou *et al.*, De novo design of a fluorescence-activating β-barrel. *Nature* **561**, 485–491 (2018). doi: 10.1038/s41586-018-0509-0; pmid: 30209393
6. P. Lu *et al.*, Accurate computational design of multipass transmembrane proteins. *Science* **359**, 1042–1046 (2018). doi: 10.1126/science.aaq1739; pmid: 29496880
7. N. H. Joh, G. Grigoryan, Y. Wu, W. F. DeGrado, Design of self-assembling transmembrane helical bundles to elucidate principles required for membrane protein folding and ion transport. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**, 20160214 (2017). doi: 10.1098/rstb.2016.0214; pmid: 28630154
8. J. H. Kleinschmidt, T. den Blaauwen, A. J. Driessen, L. K. Tamm, Outer membrane protein A of Escherichia coli inserts and folds into lipid bilayers by a concerted mechanism. *Biochemistry* **38**, 5006–5016 (1999). doi: 10.1021/bi982465w; pmid: 10213603
9. J. H. Kleinschmidt, L. K. Tamm, Secondary and tertiary structure formation of the β-barrel membrane protein OmpA is synchronized and depends on membrane thickness. *J. Mol. Biol.* **324**, 319–330 (2002). doi: 10.1016/S0022-2836(02)01071-9; pmid: 12441110
10. E. J. Danoff, K. G. Fleming, Novel Kinetic Intermediates Populated along the Folding Pathway of the Transmembrane β-Barrel OmpA. *Biochemistry* **56**, 47–60 (2017). doi: 10.1021/acs.biochem.6b00809; pmid: 28001375
11. C. P. Moon, S. Kwon, K. G. Fleming, Overcoming hysteresis to attain reversible equilibrium folding for outer membrane phospholipase A in phospholipid bilayers. *J. Mol. Biol.* **413**, 484–494 (2011). doi: 10.1016/j.jmb.2011.08.041; pmid: 21888919
12. D. Chaturvedi, R. Mahalakshmi, Transmembrane β-barrels: Evolution, folding and energetics. *Biochim. Biophys. Acta Biomembr.* **1859**, 2467–2482 (2017). doi: 10.1016/j.bbamem.2017.09.020; pmid: 28943271
13. T. Z. Butler, M. Pavlenok, I. M. Derrington, M. Niederweis, J. H. Gundlach, Single-molecule DNA detection with an engineered MspA protein nanopore. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20647–20652 (2008). doi: 10.1073/pnas.0807514106; pmid: 19098105
14. X. Guan, L.-Q. Gu, S. Cheley, O. Braha, H. Bayley, Stochastic sensing of TNT with a genetically engineered pore. *ChemBioChem* **6**, 1875–1881 (2005). doi: 10.1002/cbic.200500064; pmid: 16118820
15. F. Haque, J. Lunn, H. Fang, D. Smithrud, P. Guo, Real-time sensing and discrimination of single chemicals using the channel of phi29 DNA packaging nanomotor. *ACS Nano* **6**, 3251–3261 (2012). doi: 10.1021/nn3001615; pmid: 22458779
16. Y.-M. Tu *et al.*, Rapid fabrication of precise high-throughput filters from membrane protein nanosheets. *Nat. Mater.* **19**, 347–354 (2020). doi: 10.1038/s41563-019-0577-z; pmid: 31988513
17. T. Surrey, F. Jähnig, Refolding and oriented insertion of a membrane protein into a lipid bilayer. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7457–7461 (1992). doi: 10.1073/pnas.89.16.7457; pmid: 1502158
18. A. D. McLachlan, Gene duplications in the structural evolution of chymotrypsin. *J. Mol. Biol.* **128**, 49–79 (1979). doi: 10.1016/0022-2836(79)90308-5; pmid: 430571
19. A. G. Murzin, A. M. Lesk, C. Chothia, Principles determining the structure of beta-sheet barrels in proteins. I. A theoretical analysis. *J. Mol. Biol.* **236**, 1369–1381 (1994). doi: 10.1016/0022-2836(94)90064-7; pmid: 8126726
20. M. W. Franklin, J. S. G. Slusky, Tight Turns of Outer Membrane Proteins: An Analysis of Sequence, Structure, and Hydrogen Bonding. *J. Mol. Biol.* **430** (18 Pt B), 3251–3265 (2018). doi: 10.1016/j.jmb.2018.06.013; pmid: 29944853
21. E. de Alba, M. A. Jiménez, M. Rico, J. L. Nieto, Conformational investigation of designed short linear peptides able to fold into β-hairpin structures in aqueous solution. *Fold. Des.* **1**, 133–144 (1996). doi: 10.1016/S1359-0278(96)00022-3; pmid: 9079373

22. T. Blandl, A. G. Cochran, N. J. Skelton, Turn stability in β-hairpin peptides: Investigation of peptides containing 3:5 type I G1 bulge turns. *Protein Sci.* **12**, 237–247 (2003). doi: 10.1110/ps.0228603; pmid: 12538887
23. J. S. Richardson, E. D. Getzoff, D. C. Richardson, The beta bulge: A common small unit of nonrepetitive protein structure. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2574–2578 (1978). doi: 10.1073/pnas.75.6.2574; pmid: 275827
24. W. C. Wimley, Toward genomic identification of β-barrel membrane proteins: Composition and architecture of known structures. *Protein Sci.* **11**, 301–312 (2002). doi: 10.1110/ps.29402; pmid: 11790840
25. L. K. Tamm, H. Hong, B. Liang, Folding and assembly of β-barrel membrane proteins. *Biochimica et Biophysica Acta Biomembr.* **1666**, 250–263 (2004). doi: 10.1016/j.bbamem.2004.06.011
26. J. S. Merkel, L. Regan, Aromatic rescue of glycine in β sheets. *Fold. Des.* **3**, 449–455 (1998). doi: 10.1016/S1359-0278(98)00062-5; pmid: 9889161
27. D. L. Leyton *et al*., A mortise-tenon joint in the transmembrane domain modulates autotransporter assembly into bacterial outer membranes. *Nat. Commun.* **5**, 4239 (2014). doi: 10.1038/ncomms5239; pmid: 24967730
28. M. Michalik *et al*., An evolutionarily conserved glycine-tyrosine motif forms a folding core in outer membrane proteins. *PLOS ONE* **12**, e0182016 (2017). doi: 10.1371/journal.pone.0182016; pmid: 28771529
29. H. Park *et al*., Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016). doi: 10.1021/acs.jctc.6b00819; pmid: 27766851
30. R. Misra, Assembly of the β-Barrel Outer Membrane Proteins in Gram-Negative Bacteria, Mitochondria, and Chloroplasts. *ISRN Mol. Biol.* **2012**, 708203 (2012). doi: 10.5402/2012/708203; pmid: 27335668
31. M. Fioroni, T. Dworeck, F. Rodriguez-Ropero, *ß-barrel Channel Proteins as Tools in Nanotechnology: Biology, Basic Science and Advanced Applications* (Springer Science & Business Media, 2013).
32. R. D. Requião *et al*., Protein charge distribution in proteomes and its impact on translation. *PLOS Comput. Biol.* **13**, e1005549 (2017). doi: 10.1371/journal.pcbi.1005549; pmid: 28531225
33. R. Koebnik, Structural and functional roles of the surface-exposed loops of the β-barrel membrane protein OmpA from Escherichia coli. *J. Bacteriol.* **181**, 3688–3694 (1999). doi: 10.1128/JB.181.12.3688-3694.1999; pmid: 10368142
34. E. J. Danoff, K. G. Fleming, The soluble, periplasmic domain of OmpA folds as an independent unit and displays chaperone activity by reducing the self-association propensity of the unfolded OmpA transmembrane β-barrel. *Biophys. Chem.* **159**, 194–204 (2011). doi: 10.1016/j.bpc.2011.06.013; pmid: 21782315
35. S. J. Fleishman, D. Baker, Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* **149**, 262–273 (2012). doi: 10.1016/j.cell.2012.03.016; pmid: 22500796
36. J. S. Richardson, D. C. Richardson, Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2754–2759 (2002). doi: 10.1073/pnas.052706099; pmid: 11880627
37. S. E. Boyken *et al*., De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352**, 680–687 (2016). doi: 10.1126/science.aad8865; pmid: 27151862
38. D. L. Minor Jr., P. S. Kim, Measurement of the β-sheet-forming propensities of amino acids. *Nature* **367**, 660–663 (1994). doi: 10.1038/367660a0; pmid: 8107853
39. T. Kortemme, A. V. Morozov, D. Baker, An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J. Mol. Biol.* **326**, 1239–1259 (2003). doi: 10.1016/S0022-2836(03)00021-4; pmid: 12589766
40. A. Ebie Tan, N. K. Burgess, D. S. DeAndrade, J. D. Marold, K. G. Fleming, Self-association of unfolded outer membrane proteins. *Macromol. Biosci.* **10**, 763–767 (2010). doi: 10.1002/mabi.200900479; pmid: 20491126
41. J.-L. Popot, Folding membrane proteins in vitro: A table and some comments. *Arch. Biochem. Biophys.* **564**, 314–326 (2014). doi: 10.1016/j.abb.2014.06.029; pmid: 24997361
42. A. Schüßler, S. Herwig, J. H. Kleinschmidt, Kinetics of Insertion and Folding of Outer Membrane Proteins by Gel Electrophoresis. *Methods Mol. Biol.* **2003**, 145–162 (2019). doi: 10.1007/978-1-4939-9512-7_7; pmid: 31218617
43. H. Hong, L. K. Tamm, Elastic coupling of integral membrane protein stability to lipid bilayer forces. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4065–4070 (2004). doi: 10.1073/pnas.0400358101; pmid: 14990786
44. G. H. M. Huysmans, S. A. Baldwin, D. J. Brockwell, S. E. Radford, The transition state for folding of an outer membrane protein. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4099–4104 (2010). doi: 10.1073/pnas.0911904107; pmid: 20133664
45. C. L. Pocanschi, G. J. Patel, D. Marsh, J. H. Kleinschmidt, Curvature elasticity and refolding of OmpA in large unilamellar vesicles. *Biophys. J.* **91**, L75–L77 (2006). doi: 10.1529/biophysj.106.091439; pmid: 16891370
46. S. Ohnishi, K. Kameyama, Escherichia coli OmpA retains a folded structure in the presence of sodium dodecyl sulfate due to a high kinetic barrier to unfolding. *Biochim. Biophys. Acta* **1515**, 159–166 (2001). doi: 10.1016/S0005-2736(01)00410-2; pmid: 11718671
47. C. P. Moon, N. R. Zaccai, P. J. Fleming, D. Gessmann, K. G. Fleming, Membrane protein thermodynamic stability may serve as the energy sink for sorting in the periplasm. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4285–4290 (2013). doi: 10.1073/pnas.1212527110; pmid: 23440211
48. C. P. Moon, K. G. Fleming, Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10174–10177 (2011). doi: 10.1073/pnas.1103979108; pmid: 21606332
49. H. Hong, S. Park, R. H. Flores Jiménez, D. Rinehart, L. K. Tamm, Role of aromatic side chains in the folding and thermodynamic stability of integral membrane proteins. *J. Am. Chem. Soc.* **129**, 8320–8327 (2007). doi: 10.1021/ja068849o; pmid: 17564441
50. J. H. Kleinschmidt, L. K. Tamm, Folding intermediates of a β-barrel membrane protein. Kinetic evidence for a multi-step membrane insertion mechanism. *Biochemistry* **35**, 12993–13000 (1996). doi: 10.1021/bi961478b; pmid: 8855933
51. N. K. Burgess, T. P. Dao, A. M. Stanley, K. G. Fleming, Beta-barrel proteins that reside in the Escherichia coli outer membrane in vivo demonstrate varied folding behavior in vitro. *J. Biol. Chem.* **283**, 26748–26758 (2008). doi: 10.1074/jbc.M802754200; pmid: 18641391
52. Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009). doi: 10.1007/s10858-009-9333-z; pmid: 19548092
53. J. M. Hemmingsen, K. M. Gernert, J. S. Richardson, D. C. Richardson, The tyrosine corner: A feature of most Greek key β-barrel proteins. *Protein Sci.* **3**, 1927–1937 (1994). doi: 10.1002/pro.5560031104; pmid: 7703839
54. C. M. Bishop, W. F. Walkenhorst, W. C. Wimley, Folding of β-sheets in membranes: Specificity and promiscuity in peptide model systems. *J. Mol. Biol.* **309**, 975–988 (2001). doi: 10.1006/jmbi.2001.4715; pmid: 11399073
55. J. E. Horne, D. J. Brockwell, S. E. Radford, Role of the lipid bilayer in outer membrane protein folding in Gram-negative bacteria. *J. Biol. Chem.* **295**, 10340–10367 (2020). doi: 10.1074/jbc.REV120.011473; pmid: 32499369
56. B. Schiffrin, D. J. Brockwell, S. E. Radford, Outer membrane protein folding from an energy landscape perspective. *BMC Biol.* **15**, 123 (2017). doi: 10.1186/s12915-017-0464-5; pmid: 29268734
57. N. Koga *et al*., Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012). doi: 10.1038/nature11600; pmid: 23135467
58. P. D. Adams *et al*., PHENIX: A comprehensive Python-based system for macromolecular structure solution. *International Tables for Crystallography* (2012), pp. 539–547.
59. M. Källberg, G. Margaryan, S. Wang, J. Ma, J. Xu, RaptorX server: A resource for template-based protein structure modeling. *Methods Mol. Biol.* **1137**, 17–27 (2014). doi: 10.1007/978-1-4939-0366-5_2; pmid: 24573471
60. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982). doi: 10.1016/0022-2836(82)90515-0; pmid: 7108955
61. A.-M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **22**, 1302–1306 (2004). doi: 10.1038/nbt1012; pmid: 15361882
62. A. Stein, T. Kortemme, Improvements to robotics-inspired conformational sampling in rosetta. *PLOS ONE* **8**, e63090 (2013). doi: 10.1371/journal.pone.0063090; pmid: 23704889
63. A. Vorobieva, vorobieva/TransmembraneBBarrels: TMB_manuscript_supplementary_data (version bbarrel_manuscript), Zenodo (2020); http://doi.org/10.5281/zenodo.4068108.

ACKNOWLEDGMENTS

We thank I. Anishanka, Q. Cong, D. Kim, S. Berhanu Lemma, L. Stewart, L. Carter, X. Li, M. Dewitt, and A. Saleem for their help with computational and experimental work, and L. Goldschmit and P. Vecchiato for IT support. We also thank many members of the Baker, IPD, Radford, and Brockwell groups for discussions; the Advanced Photon Source beamline 24-ID-E for data collection; and I. Haydon for summary art. **Funding:** We acknowledge funding from HHMI (D.B. and A.A.V.), Fulbright Belgium and Luxembourg (A.A.V.), BBSRC (BB/M011151/1 to J.E.H.), MRC (MR/P018491/1 to P.W. and G.N.K.), NIH grants R01 GM051329 and P01 GM072694 for the NMR work (L.K.T.), R01 GM079440 (K.G.F.), T32 GM008403 (K.G.F.), and The Open Philanthropy Project Improving Protein Design Fund (D.B. and C.M.C.), Air Force Office of Scientific Research (FA9550-18-1-0297 to D.B. and C.M.C.), The Nordstrom Barrier Institute for Protein Design Directors Fund (D.B. and S.M.), and Eric and Wendy Schmidt by recommendation of the Schmidt Futures program (D.B. and S.G.). The CD instrument in Leeds was funded by the Wellcome Trust (094232/Z/10/Z). The native mass spectrometry studies were supported by a NIH P41 grant (GM128577 to V.H.W.). Northeastern Collaborative Access Team beamline was supported by NIH grants P30GM124165 and S10OD021527, and DOE contract DE-AC02-06CH11357. **Author contributions:** A.A.V. and D.B. designed the research. A.A.V. developed the design methods and expressed the designs with help of S.G., S.M., and C.M.C. A.A.V., S.G., S.M., and D.C.M. cloned, expressed, and characterized the tOmpA variants, supervised by K.G.F. and D.B. J.E.H. and A.A.V. conceived the biochemical screen and screened the designs with help of S.M. and C.M.C. C.M.C. expressed isotopically labeled proteins, and B.L. performed the NMR experiments and solved the NMR structure, supervised by L.K.T. P.W. designed, performed, and analyzed the biophysical characterization in LUVs, with help from G.N.K. and supervised by S.E.R. and D.J.B. A.Q.S. and S.R.H. performed and analyzed the native mass spectrometry experiment, supervised by V.H.W. A.K. set up crystallization trays, and A.K.B. collected data and solved the crystal structure. A.A.V. and D.B. wrote the manuscript with input from all authors. **Competing interests:** A.A.V., D.B., and J.E.H. are inventors on a U.S. provisional patent application submitted by the University of Washington that covers the described sequences. **Data and materials availability:** The Rosetta software suite is available free of charge to academic users and can be downloaded from https://www.rosettacommons.org. The scripts, datasets, design models, and sequence alignments used for this study are available from GitHub (https://github.com/vorobieva/TransmembraneBBarrels) and have been archived in Zenodo (*63*). The NMR structure of the design TMB2.3 and the crystal structure of TMB2.17 have been deposited in the Protein Data Bank (PDB) (6X1K, 6X9Z). Plasmids of the constructs are available upon request to the corresponding author.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6531/eabc8182/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S35
Tables S1 to S4
References (*64*–*97*)
Databases S1 to S6

View/request a protocol for this paper from *Bio-protocol*.

17 May 2020; accepted 7 December 2020
10.1126/science.abc8182

Science

De novo design of transmembrane β barrels

Anastassia A. Vorobieva, Paul White, Binyong Liang, Jim E. Horne, Asim K. Bera, Cameron M. Chow, Stacey Gerben, Sinduja Marx, Alex Kang, Alyssa Q. Stiving, Sophie R. Harvey, Dagan C. Marx, G. Nasir Khan, Karen G. Fleming, Vicki H. Wysocki, David J. Brockwell, Lukas K. Tamm, Sheena E. Radford and David Baker

Science **371** (6531), eabc8182.
DOI: 10.1126/science.abc8182

Building a barrel

Computational design offers the possibility of making proteins with customized structures and functions. The range of accessible protein scaffolds has expanded with the design of increasingly complex cytoplasmic proteins and, recently, helical membrane proteins. Vorobieva *et al.* describe the successful computational design of eight-stranded transmembrane β-barrel proteins (TMBs). Using an iterative approach, they show the importance of negative design to prevent off-target structures and gain insight into the sequence determinants of TMB folding. Twenty-three designs satisfied biochemical screens for a TMB structure, and two structures were experimentally validated by nuclear magnetic resonance spectroscopy or x-ray crystallography. This is a step toward the custom design of pores for applications such as single-molecule sequencing.

Science, this issue p. eabc8182