
Article:
Vorobieva, AA, White, P, Liang, B et al. (2021) De novo design of transmembrane β-barrels. Science, 371 (6531). eabc8182. ISSN: 0036-8075

https://doi.org/10.1126/science.abc8182

Title: *De novo* design of transmembrane β-barrels

Authors: 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,*]

Affiliations:

[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

[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

[8]TC Jenkins Department of Biophysics Johns Hopkins University, Baltimore, MD 21218

[†]These authors contributed equally

[‡]Current address: Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland

[§]Current address: Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, UK

[||]Current address: Analytical Research & Development, Merck & Co., Inc. 770 Sumneytown Pike, West Point, PA 19846.

*Correspondence to: dabaker@uw.edu

Abstract: Transmembrane β-barrel proteins (TMBs) are of great interest for single-molecule analytical technologies since 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 NMR and X-ray crystal structures very close to the computational models. These advances should enable the custom design of pores for a wide range of applications.

One sentence summary: *De novo* design reveals principles of transmembrane β-barrel folding.

Main text:

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 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 non translocating (*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 neighbour 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. S1B,C), half of the 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 8 β-strands. We first considered a shear number of 8 (n=S). In this configuration, the total register shift is distributed equally among the four β-hairpins (2-residue offset between each β-hairpin) and the four Cβ-strips pointing toward the lumen of the barrel are arranged in 4-fold symmetric rungs with the Cα -Cβ vectors (which indicate the direction of the sidechains) 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 2 to 4 residues resulting in a shear number of 10. In this case, there are five intertwined Cβ-strips which 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. S2A-D). While the β-turn residues closely match the sequence preferences observed in water-soluble β-barrels (mostly polar residues), the lipid-exposed residues flanking these β-turns are predominantly hydrophobic (Fig. S2H-K) (*20*) and define the *cis* boundary of the transmembrane region ("membrane anchor residues", Fig. S2A-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 (eq. S3-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, 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 4 residues, the barrel must be tilted by approximately 6.7° to the transmembrane axis (Fig. 1E, top) to bring the anchor residues into the same 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 (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, Fig. S4B-G); we focused on the placement in which the 4-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 neighbour 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, 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 4-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 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. S2E,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 a) every Cβ-strip pointing to the core of the barrel contains a glycine and b) no Cβ-strip contains more than 4 non-glycine 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. 2A,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).

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, supplementary text; such "aromatic girdles" are observed in native TMBs (*24*)). To design the remainder of the sequence, we first experimented with the approach 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 re-surfacing 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 neighbour 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: Y69 on strand 5 where the 4-residue register shifts in *cis* and *trans* produce a larger vertical offset in the β-sheet, and 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, 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 sidechain-sidechain 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 foldase/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 *E. 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 9 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 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 8-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 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/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. S8A,B). The re-looped tOmpA construct (OmpSDG) expressed at high levels in *E. coli* (where it was found in inclusion bodies), but it could not be correctly re-folded (Fig. 3C, Fig. S10C,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-PAGE) when folded into DDM detergent micelles and LUVs, characteristic of properly folded tOmpA (Fig. 3D, Fig.

S10C,D). The best expressed design Omp*Trans*3 was characterized in more detail; following refolding in detergent micelles it had a similar retention time to native tOmpA on a Size Exclusion Chromatography (SEC) column (Fig. 3D, Fig. S12), a similar native mass spectrometry (nMS) profile (Fig. S13), well-dispersed resonance peaks by ^{1}H-^{15}N-HSQC NMR in Fos-choline-12 (DPC) detergent (Fig. S10B) and a similar CD spectrum to tOmpA in both DDM micelles (Fig. 3D) and in 1,2-diundecanoyl-sn-glycero-3-phosphocholine (DUPC, diC11:0PC) LUVs with distinctive peaks at ~220 nm 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 significantly increase protein expression (Fig. S9), suggesting 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 and to seed the design of the hydrogen bonds networks. Possible positions for the Asp or Glu were exhaustively searched using the Rosetta HBNet protocol (*37*) to design β-barrel backbones with pre-installed 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 result in non-canonical out-of-plane backbone hydrogen bond geometries characteristic of a left-hand twist (O--H--N angle ~ 130°; C--O--H--N dihedral ~ -100°, Fig. 2C, top, Fig. S14A, center), while the surface residues preceding the glycine kink have 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 is often exposed to solvent to interact with a water molecule or other hydrogen bond donor in crystal structures (Fig. S15E,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 pre-glycine hydrogen bonds deviating from in-plane geometry (Fig. 2C, bottom, 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 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 based on 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 8-strand TMBs (Fig. S7D). Codon optimized synthetic genes were obtained for several representatives of each network architecture for a total of 90 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/did not express with the short sub-optimal 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 2X CMC (Data S4). Out of the sixty-six expressing designs, sixty-two formed soluble

species in such conditions. We purified the protein/detergent complexes by SEC and characterized the fifty designs which had a SEC retention volume expected for a monomeric TMB (similar to the 8-stranded tOmpA monomer and Omp*Trans*3) and a far-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 a 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, twenty-three 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 shifts profiles, characteristic of a folded protein in this detergent (Fig. S16 and Fig. 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 non-redundant 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 M and 2 M urea, and distinct from the spectra of the fully unfolded proteins in 8 M urea and from the proteins refolded in the absence of lipid (Fig. S10A, Fig. S20). We next determined the stability of the folded proteins by monitoring their ability to fold into/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 M and 7.2

M for TMB2.3 and TMB2.17, respectively, Fig. 4A) than tOmpA (Cm^F = 4.7 M), while 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/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*). 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/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 due to 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 kJ.mol^{-1} and 56 kJ.mol^{-1}. These ΔG^0_{UF} values fall within the range of natural TMBs (ΔG^0_{UF} 10-140 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 x 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 non-proline residues of TMB2.3 were fully assigned; 6 more were partially assigned (Fig. 5A, Fig. S26A). Four out of six non-assigned 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 using TALOS-N (*52*) consists of eight β-strands that closely match the β-strand boundaries in the designed model (Fig. S26C). Nine out of eleven 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. S27A,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 inter-strand 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, Fig. S26D). The NMR structure ensemble generated based on the chemical shifts and NOE information agrees closely with the design model (average backbone 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 (Fig. S28, Fig. 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. 6B,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 non-redundant 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 sidechains (Fig. 6F).

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 re-surfacing 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 model. The key to this success was introducing glycine kinks, β-bulges and register-defining sidechain 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 TMBs 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 tradeoffs 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 8-strand TMBs considered in this study (supplemental text), suggesting that the general design principles and methods 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.

References and notes:

1. R. A. Langan, S. E. Boyken, A. H. Ng, J. A. Samson, G. Dods, A. M. Westbrook, T. H.

Nguyen, M. J. Lajoie, Z. Chen, S. Berger, V. K. Mulligan, J. E. Dueber, W. R. P. Novak, H. El-Samad, D. Baker, De novo design of bioactive protein switches. *Nature*. **572**, 205–210 (2019).

2. A. H. Ng, T. H. Nguyen, M. Gómez-Schiavon, G. Dods, R. A. Langan, S. E. Boyken, J. A. Samson, L. M. Waldburger, J. E. Dueber, D. Baker, H. El-Samad, Modular and tunable biological feedback control using a de novo protein switch. *Nature*. **572**, 265–269 (2019).

3. D.-A. Silva, S. Yu, U. Y. Ulge, J. B. Spangler, K. M. Jude, C. Labão-Almeida, L. R. Ali, A. Quijano-Rubio, M. Ruterbusch, I. Leung, T. Biary, S. J. Crowley, E. Marcos, C. D. Walkey, B. D. Weitzner, F. Pardo-Avila, J. Castellanos, L. Carter, L. Stewart, S. R. Riddell, M. Pepper, G. J. L. Bernardes, M. Dougan, K. C. Garcia, D. Baker, De novo design of potent and selective mimics of IL-2 and IL-15. *Nature*. **565**, 186–191 (2019).

4. E. Marcos, T. M. Chidyausiku, A. C. McShan, T. Evangelidis, S. Nerli, L. Carter, L. G. Nivón, A. Davis, G. Oberdorfer, K. Tripsianes, N. G. Sgourakis, D. Baker, De novo design of a non-local β-sheet protein with high stability and accuracy. *Nat. Struct. Mol. Biol.* **25**, 1028–1034 (2018).

5. J. Dou, A. A. Vorobieva, W. Sheffler, L. A. Doyle, H. Park, M. J. Bick, B. Mao, G. W. Foight, M. Y. Lee, L. A. Gagnon, L. Carter, B. Sankaran, S. Ovchinnikov, E. Marcos, P.-S. Huang, J. C. Vaughan, B. L. Stoddard, D. Baker, De novo design of a fluorescence-activating β-barrel. *Nature*. **561**, 485–491 (2018).

6. P. Lu, D. Min, F. DiMaio, K. Y. Wei, M. D. Vahey, S. E. Boyken, Z. Chen, J. A. Fallas, G. Ueda, W. Sheffler, V. K. Mulligan, W. Xu, J. U. Bowie, D. Baker, Accurate computational design of multipass transmembrane proteins. *Science*. **359**, 1042–1046 (2018).

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** (2017), doi:10.1098/rstb.2016.0214.

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).

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. *Journal of Molecular Biology*. **324** (2002), pp. 319–330.

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).

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).

12. D. Chaturvedi, R. Mahalakshmi, Transmembrane β-barrels: Evolution, folding and

energetics. *Biochim. Biophys. Acta Biomembr.* **1859**, 2467–2482 (2017).

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).

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).

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).

16. Y.-M. Tu, W. Song, T. Ren, Y.-X. Shen, R. Chowdhury, P. Rajapaksha, T. E. Culp, L. Samineni, C. Lang, A. Thokkadam, D. Carson, Y. Dai, A. Mukthar, M. Zhang, A. Parshin, J. N. Sloand, S. H. Medina, M. Grzelakowski, D. Bhattacharya, W. A. Phillip, E. D. Gomez, R. J. Hickey, Y. Wei, M. Kumar, Rapid fabrication of precise high-throughput filters from membrane protein nanosheets. *Nat. Mater.* (2020), doi:10.1038/s41563-019-0577-z.

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).

18. A. D. McLachlan, Gene duplications in the structural evolution of chymotrypsin. *J. Mol. Biol.* **128**, 49–79 (1979).

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).

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**, 3251–3265 (2018).

21. E. de Alba, E. de Alba, M. Angeles Jiménez, M. Rico, J. L. Nieto, Conformational investigation of designed short linear peptides able to fold into β-hairpin structures in aqueous solution. *Folding and Design*. **1** (1996), pp. 133–144.

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 Science*. **12** (2003), pp. 237–247.

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).

24. W. C. Wimley, Toward genomic identification of β-barrel membrane proteins: Composition and architecture of known structures. *Protein Science*. **11** (2009), pp. 301–312.

25. L. K. Tamm, H. Hong, B. Liang, Folding and assembly of β-barrel membrane proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. **1666** (2004), pp. 250–263.

26. J. S. Merkel, L. Regan, Aromatic rescue of glycine in β sheets. *Folding and Design*. **3**

(1998), pp. 449–456.

27. D. L. Leyton, M. D. Johnson, R. Thapa, G. H. M. Huysmans, R. A. Dunstan, N. Celik, H.-H. Shen, D. Loo, M. J. Belousoff, A. W. Purcell, I. R. Henderson, T. Beddoe, J. Rossjohn, L. L. Martin, R. A. Strugnell, T. Lithgow, A mortise-tenon joint in the transmembrane domain modulates autotransporter assembly into bacterial outer membranes. *Nat. Commun.* **5**, 4239 (2014).

28. M. Michalik, M. Orwick-Rydmark, M. Habeck, V. Alva, T. Arnold, D. Linke, An evolutionarily conserved glycine-tyrosine motif forms a folding core in outer membrane proteins. *PLoS One*. **12**, e0182016 (2017).

29. H. Park, P. Bradley, P. Greisen Jr, Y. Liu, V. K. Mulligan, D. E. Kim, D. Baker, F. DiMaio, Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016).

30. R. Misra, Assembly of the β-Barrel Outer Membrane Proteins in Gram-Negative Bacteria, Mitochondria, and Chloroplasts. *ISRN Mol Biol*. **2012**, 708203 (2012).

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, L. Fernandes, H. J. A. de Souza, S. Rossetto, T. Domitrovic, F. L. Palhano, Protein charge distribution in proteomes and its impact on translation. *PLoS Comput. Biol.* **13**, e1005549 (2017).

33. R. Koebnik, Structural and Functional Roles of the Surface-Exposed Loops of the β-Barrel Membrane Protein OmpA fromEscherichia coli. *Journal of Bacteriology*. **181** (1999), pp. 3688–3694.

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).

35. S. J. Fleishman, D. Baker, Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell*. **149**, 262–273 (2012).

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).

37. S. E. Boyken, Z. Chen, B. Groves, R. A. Langan, G. Oberdorfer, A. Ford, J. M. Gilmore, C. Xu, F. DiMaio, J. H. Pereira, B. Sankaran, G. Seelig, P. H. Zwart, D. Baker, De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science*. **352**, 680–687 (2016).

38. D. L. Minor, P. S. Kim, Measurement of the β-sheet-forming propensities of amino acids. *Nature*. **367** (1994), pp. 660–663.

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. *Journal of Molecular Biology*. **326** (2003), pp. 1239–1259.

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).

41. J.-L. Popot, Folding membrane proteins in vitro: a table and some comments. *Arch. Biochem. Biophys.* **564**, 314–326 (2014).

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).

43. H. Hong, L. K. Tamm, Elastic coupling of integral membrane protein stability to lipid bilayer forces. *Proceedings of the National Academy of Sciences*. **101** (2004), pp. 4065–4070.

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).

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**, L75–7 (2006).

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).

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).

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).

49. H. Hong, S. Park, R. H. F. 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).

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** (1996), pp. 12993–13000.

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).

52. Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS : a hybrid method for predicting

protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR*. **44** (2009), pp. 213–223.

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 Science*. **3** (1994), pp. 1927–1937.

54. C. M. Bishop, W. F. Walkenhorst, W. C. Wimley, Folding of β-sheets in membranes: specificity and promiscuity in peptide model systems. *Journal of Molecular Biology*. **309** (2001), pp. 975–988.

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).

56. B. Schiffrin, D. J. Brockwell, S. E. Radford, Outer membrane protein folding from an energy landscape perspective. *BMC Biol.* **15**, 123 (2017).

57. M. Källberg, G. Margaryan, S. Wang, J. Ma, J. Xu, RaptorX server: A Resource for Template-Based Protein Structure Modeling. *Methods in Molecular Biology* (2014), pp. 17–27.

58. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).

59. 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).

60. A. Stein, T. Kortemme, Improvements to robotics-inspired conformational sampling in rosetta. *PLoS One*. **8**, e63090 (2013).

61. E. Marcos, B. Basanta, T. M. Chidyausiku, Y. Tang, G. Oberdorfer, G. Liu, G. V. T. Swapna, R. Guan, D.-A. Silva, J. Dou, J. H. Pereira, R. Xiao, B. Sankaran, P. H. Zwart, G. T. Montelione, D. Baker, Principles for designing proteins with cavities formed by curved β sheets. *Science*. **355**, 201–206 (2017).

62. Y.-R. Lin, N. Koga, R. Tatsumi-Koga, G. Liu, A. F. Clouser, G. T. Montelione, D. Baker, Control over overall shape and size in de novo designed proteins. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5478–85 (2015).

63. N. Koga, R. Tatsumi-Koga, G. Liu, R. Xiao, T. B. Acton, G. T. Montelione, D. Baker, Principles for designing ideal protein structures. *Nature*. **491**, 222–227 (2012).

64. H. Kamisetty, S. Ovchinnikov, D. Baker, Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. *Proceedings of the National Academy of Sciences*. **110** (2013), pp. 15674–15679.

65. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. **28** (2012), pp. 3150–3152.

66. M. B. Ulmschneider, M. S. P. Sansom, Amino acid distributions in integral membrane protein structures. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. **1512** (2001), pp. 1–14.

67. D. Gront, D. W. Kulp, R. M. Vernon, C. E. M. Strauss, D. Baker, Generalized fragment picking in Rosetta: design, protocols and applications. *PLoS One*. **6**, e23294 (2011).

68. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*. **6**, 277–293 (1995).

69. Website, (available at Goddard TD, Kneller DG SPARKY 3. University of California, San Francisco. Available at http://www.cgl.ucsf.edu/home/sparky/).

70. S. G. Hyberts, A. G. Milbradt, A. B. Wagner, H. Arthanari, G. Wagner, Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *J. Biomol. NMR*. **52**, 315–327 (2012).

71. Y. Shen, A. Bax, Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR*. **56**, 227–241 (2013).

72. C. D. Schwieters, J. J. Kuszewski, G. Marius Clore, Using Xplor-NIH for NMR Molecular Structure Determination. *ChemInform*. **37** (2006), , doi:10.1002/chin.200644278.

73. M. T. Marty, A. J. Baldwin, E. G. Marklund, G. K. A. Hochberg, J. L. P. Benesch, C. V. Robinson, Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376 (2015).

74. W. Kabsch, XDS. *Acta Crystallographica Section D Biological Crystallography*. **66** (2010), pp. 125–132.

75. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).

76. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

77. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *International Tables for Crystallography* (2012), pp. 539–547.

78. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta*

Crystallogr. D Biol. Crystallogr. **60**, 2126–2132 (2004).

79. C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. Arendall 3rd, J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson, D. C. Richardson, MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

80. D. R. Flower, The lipocalin protein family: structure and function. *Biochemical Journal*. **318** (1996), pp. 1–14.

81. L. H. Greene, E. D. Chrysina, L. I. Irons, A. C. Papageorgiou, K. Ravi Acharya, K. Brew, Role of conserved residues in structure and stability: Tryptophans of human serum retinol-binding protein, a model for the lipocalin superfamily. *Protein Science*. **10** (2009), pp. 2301–2316.

82. J. H. Kleinschmidt, Folding of β-barrel membrane proteins in lipid bilayers - Unassisted and assisted folding and insertion. *Biochim. Biophys. Acta*. **1848**, 1927–1943 (2015).

83. M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg, A. L. Lomize, OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–6 (2012).

84. R. Jackups, S. Cheng, J. Liang, Sequence Motifs and Antimotifs in β-Barrel Membrane Proteins from a Genome-Wide Analysis: The Ala-Tyr Dichotomy and Chaperone Binding Motifs. *Journal of Molecular Biology*. **363** (2006), pp. 611–623.

85. R. Jackups, J. Liang, Interstrand Pairing Patterns in β-Barrel Membrane Proteins: The Positive-outside Rule, Aromatic Rescue, and Strand Registration Prediction. *Journal of Molecular Biology*. **354** (2005), pp. 979–993.

86. J. A. Stapleton, T. A. Whitehead, V. Nanda, Computational redesign of the lipid-facing surface of the outer membrane protein OmpA. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9632–9637 (2015).

87. D. L. Leyton, M. D. Johnson, R. Thapa, G. H. M. Huysmans, R. A. Dunstan, N. Celik, H.-H. Shen, D. Loo, M. J. Belousoff, A. W. Purcell, I. R. Henderson, T. Beddoe, J. Rossjohn, L. L. Martin, R. A. Strugnell, T. Lithgow, A mortise-tenon joint in the transmembrane domain modulates autotransporter assembly into bacterial outer membranes. *Nat. Commun.* **5**, 4239 (2014).

88. P. Craveur, A. P. Joseph, J. Rebehmed, A. G. de Brevern, β-Bulges: extensive structural analyses of β-sheets irregularities. *Protein Sci.* **22**, 1366–1378 (2013).

89. M. A. Jiménez, Design of monomeric water-soluble β-hairpin and β-sheet peptides. *Methods Mol. Biol.* **1216**, 15–52 (2014).

90. P.-Y. Chen, C.-K. Lin, C.-T. Lee, H. Jan, S. I. Chan, Effects of turn residues in

directing the formation of the β-sheet and in the stability of the β-sheet. *Protein Science*. **10** (2001), pp. 1794–1800.

91. K. Fujiwara, S. Ebisawa, Y. Watanabe, H. Toda, M. Ikeguchi, Local sequence of protein β-strands influences twist and bend angles. *Proteins: Structure, Function, and Bioinformatics*. **82** (2014), pp. 1484–1493.

92. I. Walsh, F. Seno, S. C. E. Tosatto, A. Trovato, PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Res.* **42**, W301–7 (2014).

93. 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).

94. O. Conchillo-Solé, N. S. de Groot, F. X. Avilés, J. Vendrell, X. Daura, S. Ventura, AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics*. **8**, 65 (2007).

95. G. E. Crooks, WebLogo: A Sequence Logo Generator. *Genome Research*. **14** (2004), pp. 1188–1190.

96. H. Wang, K. K. Andersen, B. S. Vad, D. E. Otzen, OmpA can form folded and unfolded oligomers. *Biochim. Biophys. Acta*. **1834**, 127–136 (2013).

97. R. A. Laskowski, J. Jabłońska, L. Pravda, R. S. Vařeková, J. M. Thornton, PDBsum: Structural summaries of PDB entries. *Protein Sci.* **27**, 129–134 (2018).

98. A. G. Murzin, A. M. Lesk, C. Chothia, Principles determining the structure of beta-sheet barrels in proteins. II. The observed structures. *J. Mol. Biol.* **236**, 1382–1400 (1994).

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 for 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 Ian Haydon for summary art. **Funding:** We acknowledge funding from HHMI (DB and AAV), Fulbright Belgium and Luxembourg (AAV), BBSRC (JEH (BB/M011151/1)), MRC (PW and NK (MR/P018491/1)), NIH grants R01 GM051329 and P01 GM072694 for the NMR work (LKT), R01 GM079440 (KGF), T32 GM008403 (KGF) and The Open Philanthropy Project Improving Protein Design Fund (DB and CC), Air Force Office of Scientific Research (DB and CC (FA9550-18-1-0297)), The Nordstrom Barrier Institute for Protein Design Directors Fund (DB and SM), Eric and Wendy Schmidt by recommendation of the Schmidt Futures program (DB and

SG). The CD instrument in Leeds was funded by the Wellcome Trust (094232/Z/10/Z). The native MS studies were supported by a NIH P41 grant (GM128577) (to VHW). Northeastern Collaborative Access Team beamline supported by NIH grants P30GM124165 and S10OD021527, and DOE contract DE-AC02-06CH11357. **Author contributions:** AAV and DB designed the research. AAV developed the design methods and expressed the designs with help of SRG, SM and CMC. AAV, SRG, SM and DCM cloned, expressed and characterized the tOmpA variants, supervised by KGF and DB. JEH and AAV conceived the biochemical screen and screened the designs with help of SM and CMC. CMC expressed isotopically labeled proteins and BL performed the NMR experiments and solved the NMR structure, supervised by LKT. PW designed, performed and analyzed the biophysical characterization in LUVs, with help from GNK and supervised by SER and DJB. AQS and SRH performed and analyzed native mass spectrometry experiment, supervised by VHW. ASK set up crystallization trays and AKB collected data and solved crystal structure. AAV and DB wrote the manuscript with input from all authors. **Competing interests:** AAV, DB and JEH 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 http://www.rosettacommons.org. The scripts, datasets, design models and MSAs used for this study have been deposited in Zenodo (DOI: 10.5281/zenodo.4068108). The NMR structure of the design TMB2.3 and the crystal structure of TMB2.17 have been deposited into PDB (6X1K, 6X9Z). Plasmids of the constructs are available upon request to the corresponding author.

Supplementary materials:

Materials and Methods

Supplementary text

Figures S1-S35

Tables S1-S4

External Databases S1-S6

References (*61-98*)

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 8-strand β-barrels with shear numbers of 8 (left, 4-fold symmetric packing) and 10 (right, 4-fold screw resulting in a jigsaw-like packing). In panels C-E, the membrane anchoring residues are shown as salmon spheres. (C, D) Geometric model of membrane-association constraints on the β-barrel architecture. (C) Asymmetric register shifts between the β-hairpins 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 match 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).

Fig. 2. Sequence features defining *de novo* TMB fold and shape. (A and B) 2D schematic representation of the connectivity (hydrogen bonds as dashed lines) between β-strands in the TMB designs. Side-chains are shown as grey 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 are yellow and water molecules are red dots. Right: distributions of the C-O-H-N and O-H-N angle values describing the hydrogen bond geometry in crystal structures.

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 (*57*)). Y axis: hydrophobicity of the core (GRAVY hydropathy index (*58*)). Grey spheres, non-expressing 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 folded species was validated by HSQC; Green, naturally occurring TMBs with 8 strands. Circle size - aggregation propensity of the sequence predicted with TANGO (*59*). (B) Experimental workflow. The number of unique designs (excluding loop doublons) satisfying each criteria is shown in brackets. (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 sub-optimal (D) sequences substituted in a 3:5 type I β-turn with a G1 β-bulge. Conformational perturbations were generated using Kinematic Loop Closure (*60*); the inset shows the backbone conformations of the twenty-five lowest-energy models. Center: After refolding in 2X CMC DDM detergent, Omp*Trans*3 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,

while 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, but not OmpAAG, is similar to that of tOmpA (repeated two times).

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 ($C\text{-}_m^{UF}$) (filled circles, solid line). Kinetics of folding into (B) DUPC and (C) DMPC LUVs at a Lipid to Protein Ratio (LPR) 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 minutes (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 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 inter-residues NOEs are shown as red sticks. (C) TMB2.3 Rosetta design model (magenta) aligned to the 20 lowest energy models generated with NMR constraints (grey).

Fig. 6. Crystal structure of TMB2.17 is nearly identical to the design model. The crystal structure (pink) was determined in DPC detergent, superimposed on the design model (grey) and compared

to the crystal structure of the naturally occurring tOmpA (blue, PDB ID: 1BXW). (A) Full backbone superposition. (B) Comparison of the 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 panel 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.

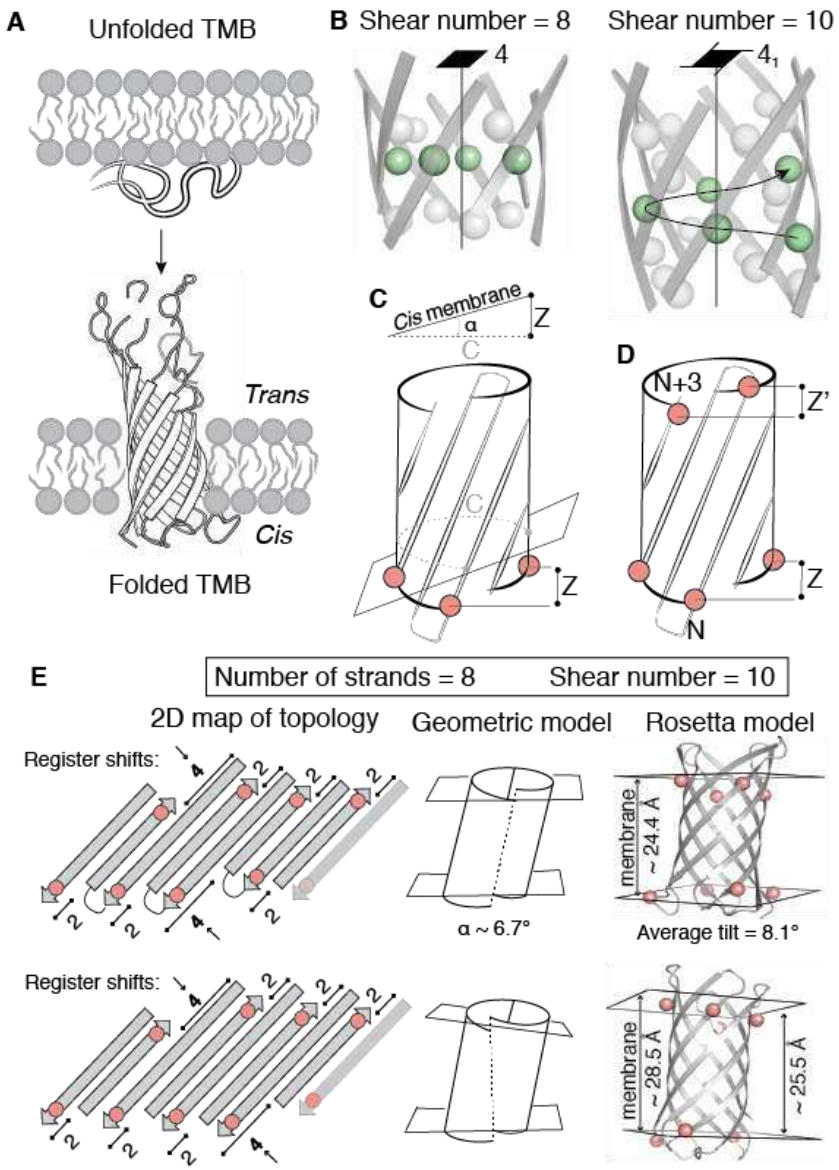

Fig. 1. Geometric principles for TMB backbone design. (A) The *trans* side of the β-barrel is the side that translocates through the lipid membrane during TMB folding and insertion. The *cis* β-turns remain on initial protein/lipid interaction side of the membrane. (B) Comparison of side-chain packing arrangements in 8-strand β-barrels with shear numbers of 8 (left, 4-fold symmetric packing) and 10 (right, 4-fold screw resulting in a jigsaw-like packing). In panels C-E, the membrane anchoring residues are shown as salmon spheres. (C, D) Geometric model of membrane-association constraints on the β-barrel architecture. (C) Asymmetric register shifts between the β-hairpins can be accommodated by tilting the β-barrel to the transmembrane axis by an angle = arctan(Z/C). (D) Because of the staggered β-strands, an anchor residue on the *cis* side

of a strand N stacks with an anchor residue on the *trans* side of strand N+3 along the main β-barrel axis; the change of level Z between two anchor residues on the *cis* side of the β-barrel must be match by the change of level Z' between the two stacking anchor residues on the *trans* side. (E) The geometric model (center) and Rosetta modelling followed by hydrophobic thickness prediction with PPM (right) predict similar tilt angles (□) of the β-barrel to the membrane axis for a given β-strand arrangement: both models show inconsistent hydrophobic thickness for β-barrel architectures with double register shifts are located on strand 1 in *cis* (strand N) (top) and on strand 6 in *trans* (strand N+5) (bottom).

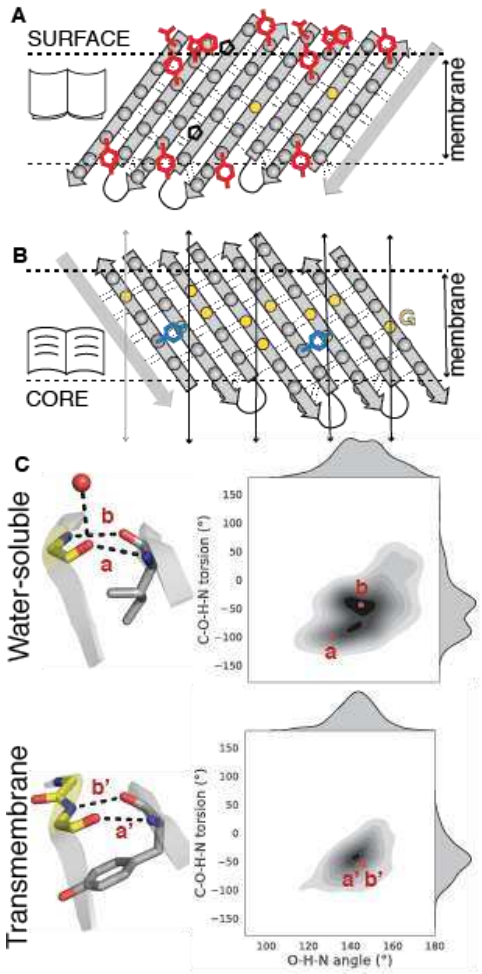

Fig. 2. Sequence features defining *de novo* TMB fold and shape. (A and B) 2D schematic representation of the connectivity (hydrogen bonds as dashed lines) between β-strands in the TMB designs. Side-chains are shown as grey spheres and glycine residues as yellow dots. Aromatic girdle motifs are shown 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 are yellow and water molecules are red dots. Right: distributions of the C-O-H-N and O-H-N angle values describing the hydrogen bond geometry in crystal structures.

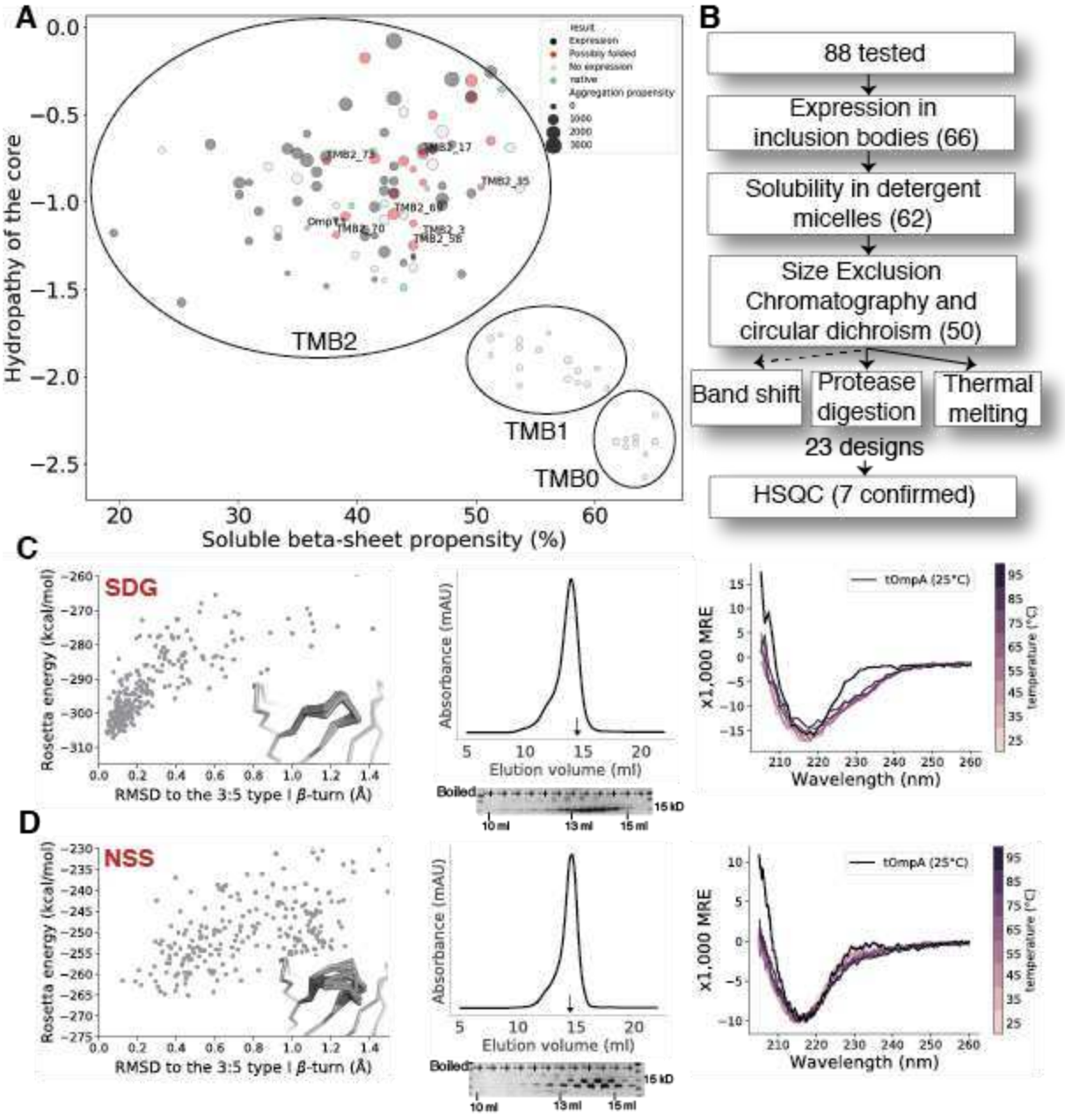

Fig. 3. Negative design is critical for *de novo* TMB folding. (A) Successful design of TMBs requires reducing β-sheet propensity of the transmembrane β-strands. X axis: β-sheet propensity of the transmembrane region (calculated with RaptroX (*54*)). Y axis: hydrophobicity of the core (GRAVY hydropathy index (*55*)). Grey spheres, Non-expressing TMB designs; black circles, expressing designs that do not fold; red, TMB designs that pass biochemical folding screening -- labels indicate folded species was validated by HSQC; Green, naturally occurring TMBs with 8

strands. Circle size, aggregation propensity of the sequence predicted with TANGO (*56*). (B) Experimental workflow. The number of unique designs (excluding loop doublons) satisfying each criteria is shown in brackets. (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 sub-optimal (D) sequences substituted in a 3:5 type I β-turn with a G1 β-bulge. Conformational perturbations were generated using Kinematic Loop Closure (*57*); the inset shows the backbone conformations of the twenty-five lowest-energy models. Center: After refolding in 2X CMC DDM detergent, OmpTrans3 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, while the OmpAAG peak elutes earlier (13.96 ml) and does not show a band shift. Right: The far-UV CD spectrum of OmpTrans3, but not OmpAAG, is similar to that of tOmpA.

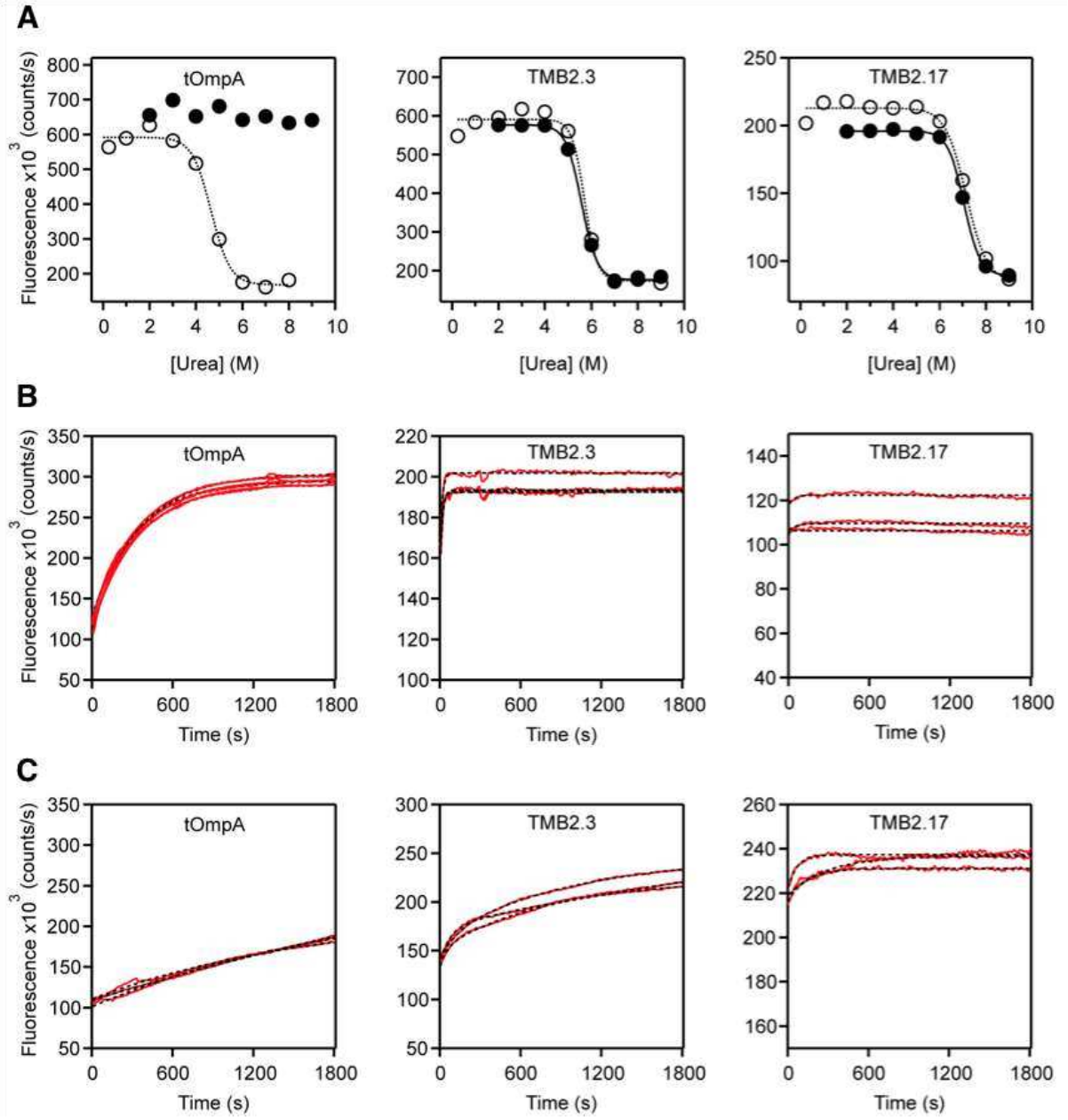

Fig. 4. Folding of *de novo* designed TMB2.3 and TMB2.17 compared to tOmpA in synthetic lipid membranes. (A) Urea dependence of folding and unfolding in DUPC LUVs. The fluorescence intensity at 335 nm was plotted against urea concentration to determine the midpoint urea concentration for folding (C_m^F) (open circles, dashed line) and unfolding ($C\text{-}_m^{UF}$) (filled circles, solid line). Kinetics of folding into (B) DUPC and (C) DMPC LUVs at a lipid to protein ratio (LPR) 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 minutes (red line). Data were fitted with a single exponential function to determine folding rate constants (black dashed line). Three replicates are shown for each.

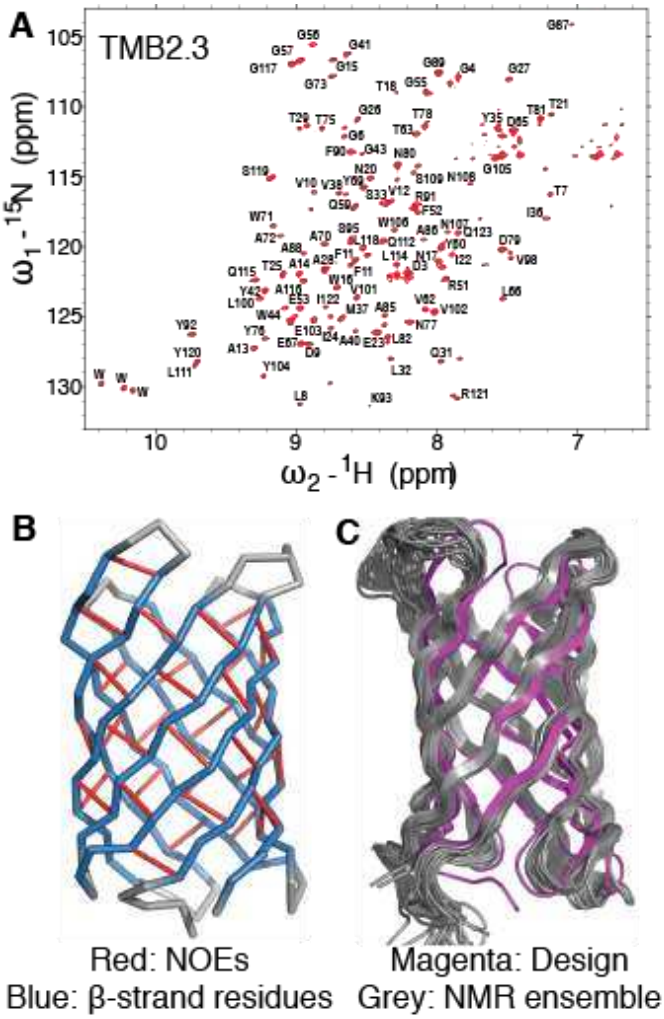

Fig. 5. NMR structure of TMB2.3 in DPC detergent micelles. (A) Assigned ^{15}N-^{1}H TROSY spectrum of TMB2.3. (B) NMR constraints mapped on the TMB2.3 design model. Residues predicted to have β-sheet secondary structure are colored in blue. Collected inter-residues NOEs

are shown as red sticks. (C) TMB2.3 Rosetta design model (magenta) aligned to the 20 lowest energy models generated with NMR constraints (grey).

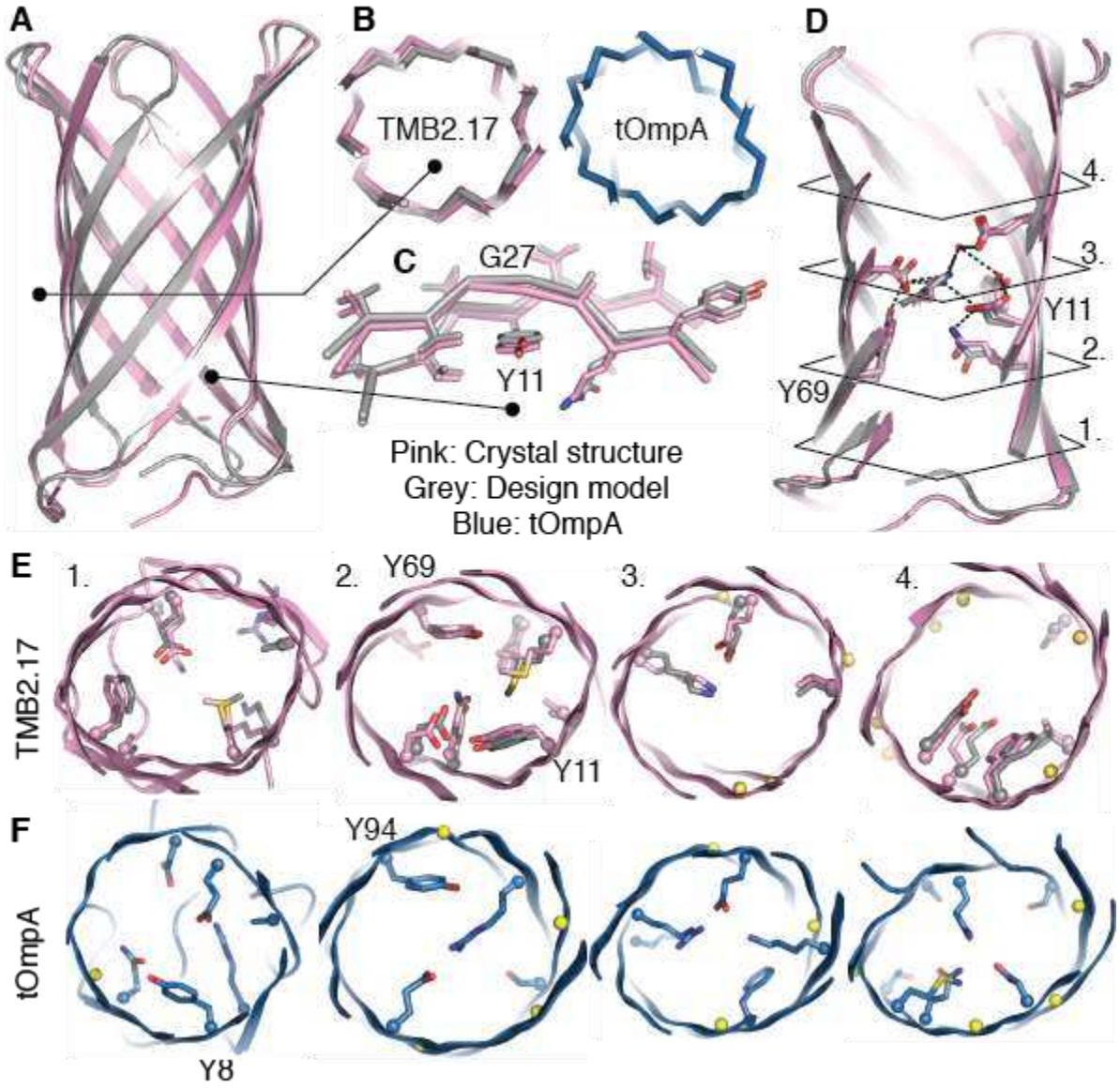

Fig. 6. Crystal structure of TMB2.17 in DPC detergent (pink) superimposed to the design model (grey) and compared to the crystal structure of the naturally occurring tOmpA (green, PDB ID: 1BXW). (A) Full backbone superposition. (B) Comparison of the transverse β-barrel cross-section geometries. (C) Superposition 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 for the design model and crystal structure (E; the two are very similar) and compared to core packing in tOmpA (F) which is quite different. 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.