White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/215131/

Version: Accepted Version

Article:
Berhanu, S., Majumder, S., Müntener, T. et al. (2024) Sculpting conducting nanopore size and shape through de novo protein design. Science, 385 (6706). pp. 282-288. ISSN: 0036-8075

https://doi.org/10.1126/science.adn3796

This is an author produced version of an article published in Science, made available under the terms of the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Sculpting conducting nanopore size and shape through *de novo* protein design

Authors: Samuel Berhanu[†,1,2,$], Sagardip Majumder[†,1,2], Thomas Müntener[‡,3], James Whitehouse[‡,4], Carolin Berner[5,6], Asim K. Bera[1,2], Alex Kang[1,2], Binyong Liang[7], G Nasir Khan[4], Banumathi Sankaran[8], Lukas K. Tamm[7], David J. Brockwell[4], Sebastian Hiller[3], Sheena E. Radford[4], David Baker[1,2,9,*], Anastassia A. Vorobieva[5,6,10,*]

Affiliations:
[1]Department of Biochemistry, The University of Washington, Seattle, WA, USA.
[2]Institute for Protein Design, University of Washington, Seattle, WA, USA.
[3]Biozentrum, University of Basel, Basel, Switzerland.
[4]Astbury Centre for Structural Molecular Biology, School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT.
[5]Structural Biology Brussel, Vrije Universiteit Brussel, Brussels, Belgium.
[6]VUB-VIB Center for Structural Biology, Brussels, Belgium.
[7]Department of Molecular Physiology and Biological Physics and Center for Membrane and Cell Physiology, University of Virginia, Charlottesville, VA 22903, USA.
[8]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.
[9]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA.
[10]VIB Center for AI and Computational Biology, Belgium.
[†]Co-first authors.
[‡]Co-second authors.
*Corresponding authors: anastassia.andreevna.vorobieva@vub.be, dabaker@uw.edu
[$]Current affiliation: The Institute for Environmental Health 15300 Bothell Way Lake Forest Park, Washington USA

Abstract: Transmembrane β-barrels (TMBs) have considerable potential for a broad range of sensing applications. Current engineering approaches for nanopore sensors are limited to naturally occurring channels, which provide sub-optimal starting points. In contrast, *de novo* protein design can in principle create an unlimited number of new nanopores with any desired properties. Here we describe a general approach to designing transmembrane β-barrel pores with different diameters and pore geometries. NMR and crystallographic characterization show that the designs are stably folded with structures close to the design models. The designs have distinct conductances that correlate with their pore diameter, ranging from 110 pS (~0.5 nm pore diameter) to 430 pS (~1.1 nm pore diameter). Our approach opens the door to the custom design of TMB nanopores for sensing and sequencing applications.

Transmembrane β-barrel (TMBs) nanopores formed by a circularly-closed single β-sheet provide rigid scaffolds for the transport of molecules across cellular (*1*) and organelle membranes (*2*), (*3*), (*4*)). Engineering of naturally occurring nanopores has enabled single-molecule enzymology (*5*), protein fingerprinting (*6*), the detection of small molecules and biomarkers (*7*), and the sequencing of biological and synthetic polymers (*8*). Of particular note is nanopore-based DNA sequencing (*9*), which has enabled widely-accessible large-scale genomics, epigenomics and microbiological analysis (*10*). Despite this success, the development of nanopore sensors for robust analysis of molecules beyond DNA sequencing has so far been challenging.

The sensing properties of a nanopore for an analyte of interest can be modulated by introducing mutations into the pore lumen that alter nanopore/analyte interactions (*11*). Yet, it remains challenging to identify a channel suitable for each of the many applications of interest, because there is only a limited set of engineerable naturally occurring nanopores, and these have evolved for functions for the most part very different than the desired applications. Going beyond nature, a conducting pore based on a β-hairpin peptide has been designed that transports poly-lysine peptides (*12*). Such self-assembling β-hairpins are however not suitable as a general approach to nanopore design, because it is challenging to control the channel size and to assemble the pore in lipid membranes. Monomeric 8-stranded TMBs have been designed that stably assemble in detergent and in lipid vesicles, but they are too small to contain a central conducting channel (*13*).

Encouraged by the success designing these narrow TMBs, we reasoned that *de novo* protein design should provide a general approach to creating robust β-barrel nanopore scaffolds for a next generation of nanopore sensors. A key challenge in designing such structures is that the polar-hydrophobic pattern characteristic of globular protein folds must be inverted: the exterior must be largely nonpolar for membrane insertion, and the interior must be largely polar to support a solvated conducting channel. Furthermore, unlike globular proteins, the structure of TMBs must be specified in the vast majority by short-range interactions between residues located on adjacent strands since there is no close-packed core. Finally, the amphipathic β-strands are highly aggregation prone prior to β-barrel assembly, and hence the design must strongly favor intra-chain rather than inter-chain interactions during folding. We set out to develop general methods to overcome these challenges and design stable monomeric channels with tunable pore shapes, sizes, and single-channel conductance.

Computational design

We sought to build from scratch TMB backbones accommodating water-accessible pores starting from the principles elucidated during the design of 8-stranded TMBs lacking pores (*13*). To modulate the size of the pore, we increased the number of β-strands (10, 12 and 14 strands) while keeping the transmembrane span and the connectivity between β-strands (the shear number (*14*, *15*)) constant. This resulted in an increase of the average β-barrel diameter from 16.4 Å for the previously designed 8 strand β-barrels (*13*) to 19.4 Å (10 strands), 22.8 Å (12 strands) and 26.4 Å (14 strands) (Figure 1A, Figure S1). By comparison to 8-stranded TMBs, the diameters of the larger β-barrels do not allow long-range side chain contacts across the pores and the structural properties of the pores (β-strand pairing, β-barrel shape) must be locally encoded. Naturally-occurring TMBs

typically feature long, disordered, loops on one side of the barrel (*16*), which can result in noisy electrophysiology recording and challenging data interpretation when the pores are used for sensing applications (*17*, *18*).

To design quiet pores and reduce the noise, we connected the β-strands on both sides of the barrels with 2- and 3-residues (Figure S2) β-hairpins, the shortest loops we have previously found to support TMB folding (*13*). The first-generation backbones corresponding to these designs were assembled with the Rosetta BlueprintBDR (*19*) application and had similar cylindrical shapes. Such cylindrical β-sheet configurations are strained (*20*, *21*) due to repulsion between side-chains packing the barrel lumen (Figure 1C). Glycine kinks (glycine residues in extended positive-phi conformation (*15*)) were introduced into the blueprint to relieve the strain and to bend the β-strands to form corners in the β-barrel cross-section. We generated four blueprints with the same topology but different glycine kink distributions to design 12-stranded β-barrel backbones with square-, triangle-, rectangle-, or oval-shaped cross-sections (Figure S1). A single glycine kink was used in corners of an angle of >= 90° and several adjacent and/or stacked kinks were placed to form corners of < 90° (Figure 1B). Sequence-agnostic TMB backbones incorporating these constraints were assembled *in silico* and had the shapes expected based on the placement of the glycine kinks (Figure 1D).

A challenge for TMB design is to balance the optimization of the folded β-barrel state in the membrane with delayed folding in water to reduce misfolding and aggregation that would prevent successful integration into a membrane bilayer (*13*, *22*, *23*). For the 8-stranded TMBs, this was achieved by incorporating local secondary-structure frustration (*24*) to reduce premature formation of aggregation-prone β-strands prior to full barrel assembly: hydrophobic amino acids were designed into the water-accessible pore to disrupt the hydrophobic-polar amino acid alternation pattern characteristic of amphipathic β-sheets. To test whether such balancing is necessary for larger β-barrel designs that need to have water-accessible (and hence more polar) channels, we first designed “optimal” 10 and 12-strand TMBs with only polar and charged amino acids facing the pore (Table S3). All 16 such designs failed to express in *E. coli* (Figure S6) possibly because they assembled into toxic β-sheet aggregates instead of inclusion bodies, as was previously observed for similarly optimal 8-stranded TMB designs. We therefore set out to design larger TMB nanopores incorporating local secondary structure frustration. In the water-accessible pore, networks of polar residues were designed around the canonical TMB folding motif Tyr-Gly-Asp/Glu (*13*, *25*, *26*) to optimize strong local β-register defining interactions while alternating with patches of hydrophobic and small, disorder promoting, residues (Gly, Ala, Ser, see Methods). On the lipid-exposed surface, design calculations favored Ser and Thr in close proximity with a glycine kink where they could form a hydrogen bond to the β-strand backbone, effectively mimicking the backbone-water hydrogen bonds observed in strongly bent β-strands of water-soluble β-barrels (Figure 1E). While it is perhaps counterintuitive to expose hydroxyl groups to the lipid environment, we included a small number of these amino acids on the lipid-exposed surface instead of hydrophobic β-branched residues (Methods) to further reduce the β-sheet propensity.

During combinatorial design of sequences for β-barrels of different size, we found that the frequency of incorporation of each amino acid type strongly depended on the curvature of the β-sheet. For each of the generated blueprints, we adjusted the Rosetta

solvation and reference energies (*27*) (Methods) to achieve the desired balance of frustrated and energetically-favorable contacts (Figure S3). Following several iterations of combinatorial sequence design and structure relaxation, designs were selected based on hydrogen bond network descriptors, secondary structure (*28*) and aggregation propensities (*29*) (Figure S4). We previously found that AlphaFold2 with multiple recycles (*30*) could accurately predict the structures of designed TMBs from single sequence input without sequence alignments (*31*), and that the confidence assigned to the model (plDDT) was a good discriminator of the sequences with higher probability of experimentally folding (*32*). We selected 4-10 designs per blueprint for which AlphaFold2 predicted high-confidence structures closely matching the design models (Figure S5).

Experimental characterization of TMB folding

We first tested two sets of TMBs with 10 (four designs) or 12 β-strands with a square cross-section (nine designs). Genes were synthesized and the proteins were expressed as inclusion bodies in *E. coli* to avoid the complexity of targeting the outer membrane (*33*) (Figure 2A). Unlike the 16 “optimal” designs which all failed to express, most sequences incorporating secondary structure frustration were expressed at high levels (12/13, Figure S7). Since most naturally-occurring TMBs can fold *in vitro* (*34*), the purified designs were solubilized in guanidine hydrochloride and refolded by slow dilution into a buffer containing either detergent (fos-choline 12 (DPC) at a concentration double the critical micellar concentration (CMC)) or synthetic lipid vesicles (Material & Methods). As previously observed for the 8-stranded TMB designs, the standard band-shift assay on cold SDS-PAGE used to assess folding of natural TMBs (*35*) was not informative to identify properly folded synthetic TMBs (Figure S8). Instead, the designs were characterized by size exclusion chromatography (SEC), far UV circular dichroism (CD) in the presence of DPC detergent, and tryptophan fluorescence in DUPC ($C_{11:0}$PC) large unilamellar vesicles (LUVs). One 10-strand design (TMB10_163) and one 12-strand design (TMB12_3) with predominantly monomeric SEC profiles (Figure 2A), thermostable CD spectra characteristic of β-sheet (Figure 2B,C) and clear shift of tryptophan fluorescence maximum from ~350 nm (unfolded proteins in 8 M urea or in the absence of lipid) to ~330 nm (folded in LUVs) (Figure S9, S10) were selected for further characterization by urea titration. Both designs showed sharp and reversible folding/unfolding transitions in the presence of DUPC LUVs (Figure 2D) (mid-point urea concentrations for folding (Cm^F): 4.5 ± 0.2 M and 5.5 ± 0.2 M, respectively). The equilibrium unfolding curves were fitted to a two-states transition, with the calculated unfolding free energies (ΔG^0_{UF}) of -35.6 ± 2.7 and -63.1 ± 8.0 kJ/mol (for TMB10_163 and TMB12_3, respectively) in the range of natural (ΔG^0_{UF} -10 to -140 kJ/mol (*36*–*39*)) and previously designed 8-stranded TMBs (-38 and -56 kJ/mol (*13*)).

To confirm that the designs folded by integration into the bilayer rather than partial folding on its surface, the kinetics of folding were recorded in DUPC ($C_{11:0}$PC) membranes, as well as in thicker DMPC ($C_{14:0}$PC) membranes. Integral folding is expected to happen slower in thicker than in thinner membranes, whereas folding on the bilayer surface should be relatively insensitive to its thickness. Dramatically decreased folding rates were observed with DMPC compared with DUPC LUVs (Figure S11), consistent with integral membrane folding.

Encouraged by these results, we assessed the nanopore activity of these two designs following spontaneous insertion into planar dipalmitoylphosphatidylcholine (DPhPC) membranes after dilution out of DPC micelles. The 12-strand TMB12_3 inserted successfully into the membrane, producing distinct jumps of current of reproducible intensities (Figure S12) and stable conductance. While the design TMB10_163 did not have detectable nanopore activity, the variant TMB10_165 (obtained by sampling surface residues with Rosetta (40) and a modified energy function; Methods) with seven mutations on the lipid-exposed surface (T72V, T102V, I114V, L124A, V126I, V138I and V144I) inserted into DPhPC membranes and conducted ions (Figure S12). TMB10_165 had higher stability to protease digestion than TMB10_163, and more dispersed NMR ^{1}H-^{15}N HSQC chemical shift in DPC micelles (Figure S13). The TMB10_165 and TMB12_3 pores remained stably inserted over long periods of time with the longest recording acquired being 2 hours for the TMB12_3 design. Recording of the current-to-voltage response showed monotonic increases in observed conductance with increasing positive or negative voltage, indicative of stable transmembrane channels (I/V curves in Figure S12). Overall, results on TMB10_163, TMB10_165, TMB12_3 and other TMB12 designs with less or no detectable nanopore activity (Figure S15) indicate a strong correlation between membrane integration and nanopore conductance with stable TMB folding *in vitro*.

We next sought to solve the structures of the designs to assess the accuracy of the computational design methods. Although the design TMB10_165 did not form crystals in the conditions screened, TMB10_163 formed crystals which diffracted to 2.5-Å resolution (Table S1). The seven surface-exposed mutations between TMB10_165 and TMB10_163 are shown in Figure 3A. The four copies of the TMB10_163 in the asymmetric unit had a structure similar to the original Rosetta design, with an average RMSD of 1.4 Å over all backbone heavy atoms (Figure 3A) and featured the expected β-strand connectivity (shear number of 12). Most of the sidechains lining the pore had similar rotameric states in the crystal structure and the design model, with remarkable similarity at the level of the designed Tyr-Gly-Asp/Glu folding motifs (Figure 3B). Although TMB10_163 nanopore activity was not observed, analysis of its structure using PoreWalker (*41*) and MOLE 2.5 (*42*) indicated the presence of a water-accessible cylindrical pore with an average diameter ranging from 4.2 to 5.3 Å in the four subunits (Figure 3C, Figure S16), matching the diameter of the pore calculated from TMB10_163 design model (4.6 Å).

We determined the structure of TMB12_3 by NMR spectroscopy. Optimization of the *in vitro* folding conditions showed that the protein was structured in aqueous solution in LDAO detergent micelles, as indicated by well-dispersed amide and side chain methyl spectra (Figure S17, Figure S18). Secondary chemical shifts indicated the presence of twelve β-strands as in the design (Figure S19). Amide and side chain methyl NOEs spanned a dense network of experimental connectivities that reached around the barrel circumference and thus confirmed the correct arrangement of the strands into the predicted barrel structure (Figure 3D). TMB12_3 has the designed β-strands connectivity (shear number of 14) with the barrel closed by the canonical antiparallel β1-β12 seam (Figure 3E, Figure S20, Table S2).

The crystal and NMR structures demonstrate that our computational design method can design TMB nanopores with precisely controlled shear, channel width and shape.

Electrophysiology

Encouraged by the success in designing 10- and 12-stranded β-barrels, we set out to design TMBs with different numbers of β-strands and different shapes. We designed 12-stranded β-barrels with a triangular cross-section (eight designs), an oval cross-section (seven designs), or a rectangular cross-section (nine designs), as well as 14 β-stranded β-barrels (nine designs), incorporating the design features described above for the 10- and 12-stranded TMBs. The designs were obtained as synthetic genes and the proteins were again expressed in inclusion bodies. A lower fraction of 12-stranded TMB designs with a rectangular (4/9 designs) and oval (4/7 designs) cross-section showed a prominent expression band SDS-PAGE gel compared to the square-shaped designs (8/9). This difference could be the result of a less homogeneous distribution of β-sheet destabilizing amino acids (which are easier to introduce in bent than in flat β-sheet regions) in these designs, as suggested by a higher density of strong β-sheet islands co-localizing with predicted early folding regions (*43*) (Figure S21). The difficulty of *de novo* β-barrel design thus depends not only on the size of the TMB pore but also on the shape encoded into the blueprint. We then confirmed that the designs formed soluble, monodispersed, species in DPC micelles with expected β-sheet secondary structure (Figure S22) and proceeded to screen them for nanopore activity.

We evaluated the ability of the designs to insert into planar membranes from dilute detergent solution and form conducting pores (Figure 4). We obtained both 12 (three triangular-shaped, three oval-shaped and two rectangle-shaped) and 14 stranded (two) TMBs that exhibited consistent and stable conductances at positive and negative voltage (Fig 4, 3rd and 5th columns), with multiple sequential insertions corresponding to current jumps of small integral multiples of the base pore conductance (Figure 4, 4th column).

Based on the intensities of the current jumps, we estimated the conductances of single-channel events, which increase with pore size as expected: the 10 stranded TMB design described above had a conductance of 108 ± 1.4 pS, which based on the cylindrical pore access resistance model (*44*) corresponds to a nanopore diameter of approximately 3.5 Å. The 12-stranded designs had similar conductances to each other (210-230 pS) despite their different shapes, consistent with a cylindrical nanopore of around 5 Å. The 14-stranded design had a conductance of 427 ± 2.7 pS consistent with a calculated pore diameter of 7 Å. The predicted diameters are close to the average expected diameters of 4.6 ± 0.7 Å, 9.4 ± 0.8 Å and 10.6 ± 1.4 Å (calculated along the pore of TMB10_165, TMB12_3 and TMB14_8 design models, respectively, using MOLE 2.5 (*42*) (Figure S16)). In comparison to naturally-occurring pores used for sensing, such as OmpG which undergoes both transient and complete occlusion events by its solvent-exposed loops over a timescale of 100 ms (*18*, *45*), our TMB designs show remarkably quiet conductances, with no occlusion events detected over 10 sec measurements (Figure S12). Varying the shape of the pore while keeping the size constant (Figure 4, first column) did not have a large effect on monovalent ion conductance, The net flux of ions likely depends on the pore area more than on its shape, given the flexibility of the long polar side chains lining the channel (Figure S23). We anticipate that modulation of the

nanopore shape and chemical lining should allow control over the permeability of the pores to larger and more complex solutes in the future.

Discussion

Our results demonstrate that it is possible to systematically design transmembrane β-barrels with conducting pores spanning a range of sizes and shapes. Despite the inversion of the hydrophobic exterior and polar core compared to globular proteins, and the almost entirely local nature of the side chain interactions, our approach enables TMB design with atomic level precision, as highlighted by the close agreement between the experimentally determined crystal and NMR structures and the corresponding design models. Whereas the shapes of globular proteins are largely determined by the packing of hydrophobic residues in a central core, the TMB shapes can be specified by strategic placement of glycine residues at which bending takes place to reduce strain. As previously observed for 8-stranded TMBs, a delicate balance between the optimization of tertiary structure energy and negative design (introduction of locally frustrated residues) to disfavor premature β-strand formation before membrane insertion was critical for the expression of the larger TMB nanopores in *E coli* inclusion bodies.

In comparison with previously designed oligomeric protein nanopores - built from self-assembling α-helical (*46*–*49*) or β-hairpin peptides (*12*) - the nanopores presented here have the advantage of being built from a single chain which enables controlled assembly of monodisperse nanopores without alternative oligomeric states and with much greater control over the shape of the transmembrane channel (Figure S24), and efficient folding into detergent micelles and lipid membranes. While the β-hairpin based nanopores were soluble only in lipid nanoparticles (*12*), the monomeric TMB design -, like naturally occurring nanopores used for sensing applications - can be solubilized in detergent and spontaneously insert into DPhPC planar lipid membrane following dilution. The most stable nanopores allowed up to 2 hours of quiet recording, thanks to the use of the shortest loops compatible with TMB folding to connect the β-strands. The design principles presented here provide a solution to the long-standing problem of engineering quiet monomeric pores (*17*, *18*, *45*, *5*0) that has limited the use of monomeric integral TMBs such as OmpG as sensors by fusing analyte-recognition motifs (*51*, *52*) or biotin-bound (*53*, *54*) antibodies in the solvent-exposed loops (*7*). As illustrated in an accompanying manuscript (*55*), the designed nanopores can be converted into ligand gated channels with considerably lower noise and more comprehensible signal analysis than previously engineered channels.

Unlike native pores, which are finite in number, there is no limit on the number of distinct designed pores that can be generated. With further optimization of synthetic TMB nanopore insertion into membranes in multi-channel flow cells (e.g. by coupling the height of designed nanopores with that of matched thick synthetic membranes *(56)*), it should be possible to establish fast design-build-test loops to probe the relation between the chemical properties of a nanopore and the detection of an analyte in the pore lumen (*11*, *57*, *58*). Our approach now enables the custom design of pore geometry and chemistry for applications ranging from detection and selective transport of a wide range of molecules of interest to biopolymer sequencing.

References:

1. M. Dal Peraro, F. G. van der Goot, Pore-forming toxins: ancient, but never really out of fashion. *Nat. Rev. Microbiol.* **14**, 77–92 (2016).
2. D. Ranava, A. Caumont-Sarcos, C. Albenne, R. Ieva, Bacterial machineries for the assembly of membrane-embedded β-barrel proteins. *FEMS Microbiol. Lett.* **365**, fny087 (2018).
3. Building Better Barrels – β-barrel Biogenesis and Insertion in Bacteria and Mitochondria. *J. Mol. Biol.* **433**, 166894 (2021).
4. P. M. Day, K. Inoue, S. M. Theg, Chloroplast Outer Membrane β-Barrel Proteins Use Components of the General Import Apparatus. *Plant Cell*. **31**, 1845–1855 (2019).
5. L. Harrington, L. T. Alexander, S. Knapp, H. Bayley, Single-Molecule Protein Phosphorylation and Dephosphorylation by Nanopore Enzymology. *ACS Nano*. **13**, 633–641 (2019).
6. M. Afshar Bakshloo, J. J. Kasianowicz, M. Pastoriza-Gallego, J. Mathé, R. Daniel, F. Piguet, A. Oukhaled, Nanopore-Based Protein Identification. *J. Am. Chem. Soc.* **144**, 2716–2725 (2022).
7. M. A. V. Fahie, B. Yang, C. M. Chisholm, M. Chen, Protein Analyte Sensing with an Outer Membrane Protein G (OmpG) Nanopore. *Methods Mol. Biol.* **2186**, 77–94 (2021).
8. C. Cao, L. F. Krapp, A. Al Ouahabi, N. F. König, N. Cirauqui, A. Radenovic, J.-F. Lutz, M. D. Peraro, Aerolysin nanopores decode digital information stored in tailored macromolecular analytes. *Sci Adv*. **6,** eabc2661 (2020), doi:10.1126/sciadv.abc2661.
9. S. E. Van der Verren, N. Van Gerven, W. Jonckheere, R. Hambley, P. Singh, J. Kilgour, M. Jordan, E. J. Wallace, L. Jayasinghe, H. Remaut, A dual-constriction biological nanopore resolves homonucleotide sequences with high fidelity. *Nat. Biotechnol.* **38**, 1415–1420 (2020).
10. D. Deamer, M. Akeson, D. Branton, Three decades of nanopore sequencing. *Nat. Biotechnol.* **34**, 518–524 (2016).
11. C. Cao, N. Cirauqui, M. J. Marcaida, E. Buglakova, A. Duperrex, A. Radenovic, M. Dal Peraro, Single-molecule sensing of peptides and nucleic acids by engineered aerolysin nanopores. *Nat. Commun.* **10**, 4918 (2019).
12. K. Shimizu, B. Mijiddorj, M. Usami, I. Mizoguchi, S. Yoshida, S. Akayama, Y. Hamada, A. Ohyama, K. Usui, I. Kawamura, R. Kawano, De novo design of a nanopore for single-molecule detection that incorporates a β-hairpin peptide. *Nat. Nanotechnol.* **17**, 67–75 (2022).
13. A. A. Vorobieva, P. White, B. Liang, J. E. Horne, A. K. Bera, C. M. Chow, S. Gerben, S. Marx, A. Kang, A. Q. Stiving, S. R. Harvey, D. C. Marx, G. N. Khan, K. G. Fleming, V. H. Wysocki, D. J. Brockwell, L. K. Tamm, S. E. Radford, D. Baker, De novo design of transmembrane β barrels. *Science*. **371,** eabc8182 (2021), doi:10.1126/science.abc8182.
14. W. M. Liu, Shear numbers of protein beta-barrels: definition refinements and statistics. *J. Mol. Biol.* **275**, 541–545 (1998).
15. Principles determining the structure of β-sheet barrels in proteins I. A theoretical analysis. *J. Mol. Biol.* **236**, 1369–1381 (1994).
16. 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).
17. B. Pham, C. M. Chisholm, J. Foster, E. Friis, M. A. Fahie, M. Chen, A pH-independent quiet OmpG pore with enhanced electrostatic repulsion among the extracellular loops. *Biochim. Biophys. Acta Biomembr.* **1863**, 183485 (2021).
18. M. Chen, S. Khalid, M. S. P. Sansom, H. Bayley, Outer membrane protein G: Engineering a quiet pore for biosensing. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 6272–6277 (2008).
19. 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).
20. 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).
21. F. R. Salemme, Conformational and geometrical properties of beta-sheets in proteins. III. Isotropically stressed configurations. *J. Mol. Biol.* **146**, 143–156 (1981).
22. T. Surrey, F. Jähnig, Kinetics of Folding and Membrane Insertion of a β-Barrel Membrane Protein (*). *J. Biol. Chem.* **270**, 28199–28203 (1995).
23. 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).
24. E. Di Silvio, M. Brunori, S. Gianni, Frustration Sculpts the Early Stages of Protein Folding. *Angew. Chem. Int. Ed Engl.* **54**, 10867–10869 (2015).
25. 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).
26. 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).
27. R. F. Alford, A. Leaver-Fay, J. R. Jeliazkov, M. J. O'Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W. Labonte, M. S. Pacella, R. Bonneau, P. Bradley, R. L. Dunbrack Jr, R. Das, D. Baker, B. Kuhlman, T. Kortemme, J. J. Gray, The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
28. Z. Wang, F. Zhao, J. Peng, J. Xu, Protein 8-class secondary structure prediction using conditional neural fields. *Proteomics*. **11**, 3786–3792 (2011).
29. 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).
30. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K.

Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583–589 (2021).
31. M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: making protein folding accessible to all. *Nat. Methods*. **19**, 679–682 (2022).
32. A. M. Hermosilla, C. Berner, S. Ovchinnikov, A. A. Vorobieva, Validation of de novo designed water-soluble and transmembrane proteins by in silico folding and melting. *bioRxiv* (2023), p. 2023.06.06.543955.
33. A. Konovalova, D. E. Kahne, T. J. Silhavy, Outer Membrane Biogenesis. *Annu. Rev. Microbiol.* **71**, 539–556 (2017).
34. B. Schiffrin, D. J. Brockwell, S. E. Radford, Outer membrane protein folding from an energy landscape perspective. *BMC Biol.* **15**, 123 (2017).
35. 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).
36. 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).
37. 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).
38. 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).
39. 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).
40. J. K. Leman, B. D. Weitzner, S. M. Lewis, J. Adolf-Bryfogle, N. Alam, R. F. Alford, M. Aprahamian, D. Baker, K. A. Barlow, P. Barth, B. Basanta, B. J. Bender, K. Blacklock, J. Bonet, S. E. Boyken, P. Bradley, C. Bystroff, P. Conway, S. Cooper, B. E. Correia, B. Coventry, R. Das, R. M. De Jong, F. DiMaio, L. Dsilva, R. Dunbrack, A. S. Ford, B. Frenz, D. Y. Fu, C. Geniesse, L. Goldschmidt, R. Gowthaman, J. J. Gray, D. Gront, S. Guffy, S. Horowitz, P.-S. Huang, T. Huber, T. M. Jacobs, J. R. Jeliazkov, D. K. Johnson, K. Kappel, J. Karanicolas, H. Khakzad, K. R. Khar, S. D. Khare, F. Khatib, A. Khramushin, I. C. King, R. Kleffner, B. Koepnick, T. Kortemme, G. Kuenze, B. Kuhlman, D. Kuroda, J. W. Labonte, J. K. Lai, G. Lapidoth, A. Leaver-Fay, S. Lindert, T. Linsky, N. London, J. H. Lubin, S. Lyskov, J. Maguire, L. Malmström, E. Marcos, O. Marcu, N. A. Marze, J. Meiler, R. Moretti, V. K. Mulligan, S. Nerli, C. Norn, S. Ó'Conchúir, N. Ollikainen, S. Ovchinnikov, M. S. Pacella, X. Pan, H. Park, R. E. Pavlovicz, M. Pethe, B. G. Pierce, K. B. Pilla, B. Raveh, P. D. Renfrew, S. S. R. Burman, A. Rubenstein, M. F. Sauer, A. Scheck, W. Schief, O. Schueler-Furman, Y. Sedan, A. M. Sevy, N. G. Sgourakis, L. Shi, J. B. Siegel, D.-A. Silva, S. Smith, Y. Song, A. Stein, M. Szegedy, F. D. Teets, S. B. Thyme, R. Y.-R. Wang, A. Watkins, L. Zimmerman, R. Bonneau, Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat. Methods*. **17**, 665–680 (2020).
41. M. Pellegrini-Calace, T. Maiwald, J. M. Thornton, PoreWalker: a novel tool for the

identification and characterization of channels in transmembrane proteins from their three-dimensional structure. *PLoS Comput. Biol.* **5**, e1000440 (2009).
42. L. Pravda, D. Sehnal, D. Toušek, V. Navrátilová, V. Bazgier, K. Berka, R. Svobodová Vareková, J. Koca, M. Otyepka, MOLEonline: a web-based tool for analyzing channels, tunnels and pores (2018 update). *Nucleic Acids Res.* **46**, W368–W373 (2018).
43. D. Raimondi, G. Orlando, R. Pancsa, T. Khan, W. F. Vranken, Exploring the Sequence-based Prediction of Folding Initiation Sites in Proteins. *Sci. Rep.* **7**, 8826 (2017).
44. S. W. Kowalczyk, A. Y. Grosberg, Y. Rabin, C. Dekker, Modeling the conductance and DNA blockade of solid-state nanopores. *Nanotechnology*. **22**, 315101 (2011).
45. R. R. Sanganna Gari, P. Seelheim, B. Liang, L. K. Tamm, Quiet Outer Membrane Protein G (OmpG) Nanopore for Biosensing. *ACS Sens*. **4**, 1230–1235 (2019).
46. C. Xu, P. Lu, T. M. Gamal El-Din, X. Y. Pei, M. C. Johnson, A. Uyeda, M. J. Bick, Q. Xu, D. Jiang, H. Bai, G. Reggiano, Y. Hsia, T. J. Brunette, J. Dou, D. Ma, E. M. Lynch, S. E. Boyken, P.-S. Huang, L. Stewart, F. DiMaio, J. M. Kollman, B. F. Luisi, T. Matsuura, W. A. Catterall, D. Baker, Computational design of transmembrane pores. *Nature*. **585**, 129–134 (2020).
47. H. T. Kratochvil, L. C. Watkins, M. Mravic, J. L. Thomaston, J. M. Nicoludis, N. H. Somberg, L. Liu, M. Hong, G. A. Voth, W. F. DeGrado, Transient water wires mediate selective proton transport in designed channel proteins. *Nat. Chem.* **15**, 1012–1021 (2023).
48. K. R. Mahendran, A. Niitsu, L. Kong, A. R. Thomson, R. B. Sessions, D. N. Woolfson, H. Bayley, A monodisperse transmembrane α-helical peptide barrel. *Nat. Chem.* **9**, 411–419 (2016).
49. A. J. Scott, A. Niitsu, H. T. Kratochvil, E. J. M. Lang, J. T. Sengel, W. M. Dawson, K. R. Mahendran, M. Mravic, A. R. Thomson, R. L. Brady, L. Liu, A. J. Mulholland, H. Bayley, W. F. DeGrado, M. I. Wallace, D. N. Woolfson, Constructing ion channels from water-soluble α-helical barrels. *Nat. Chem.* **13**, 643–650 (2021).
50. M. A. Fahie, J. Candido, G. Andree, M. Chen, Tuning Protein Discrimination Through Altering the Sampling Interface Formed between the Analyte and the OmpG Nanopore. *ACS Sens*. **6**, 1286–1294 (2021).
51. J. C. Foster, B. Pham, R. Pham, M. Kim, M. D. Moore, M. Chen, An Engineered OmpG Nanopore with Displayed Peptide Motifs for Single-Molecule Multiplex Protein Detection. *Angew. Chem. Int. Ed Engl.* **62**, e202214566 (2023).
52. H.-J. Hwang, J.-S. Kim, J. Lee, J. S. Min, K.-B. Jeong, E. Kim, M.-K. Lee, S.-W. Chi, Single-Molecule Sensing of an Anticancer Therapeutic Protein-Protein Interaction Using the Chemically Modified OmpG Nanopore. *Anal. Chem.* **94**, 7449–7454 (2022).
53. M. A. Fahie, B. Yang, M. Mullis, M. A. Holden, M. Chen, Selective Detection of Protein Homologues in Serum Using an OmpG Nanopore. *Anal. Chem.* **87**, 11143–11149 (2015).
54. M. A. Fahie, M. Chen, Electrostatic Interactions between OmpG Nanopore and Analyte Protein Surface Can Distinguish between Glycosylated Isoforms. *J. Phys. Chem. B*. **119**, 10198–10206 (2015).
55. L. An, M. Said, L. Tran, S. Majumder, I. Goreshnik, G. R. Lee, D. Juergens, J.

Dauparas, I. Anishchenko, B. Coventry, A. K. Bera, A. Kang, P. M. Levine, V. Alvarez, A. Pillai, C. Norn, D. Feldman, D. Zorine, D. R. Hicks, X. Li, M. G. Sanchez, D. K. Vafeados, P. J. Salveson, A. A. Vorobieva, D. Baker, De novo design of diverse small molecule binders and sensors using Shape Complementary Pseudocycles, *bioRxiv* (2023)p. 2023.12.20.572602.
56. J. A. Peruzzi, J. Steinkühler, T. Q. Vu, T. F. Gunnels, P. Lu, D. Baker, N. P. Kamat, Hydrophobic mismatch drives self-organization of designer proteins into synthetic membranes, *bioRxiv* (2022)p. 2022.06.01.494374.
57. A. Sauciuc, B. Morozzo Della Rocca, M. J. Tadema, M. Chinappi, G. Maglia, Translocation of linearized full-length proteins through an engineered nanopore under opposing electrophoretic force. *Nat. Biotechnol.* (2023), doi:10.1038/s41587-023-01954-x.
58. J. Liu, A. Aksimentiev, Molecular Determinants of Current Blockade Produced by Peptide Transport Through a Nanopore. *ACS Nanoscience Au* **4**, 21-29 (2024), doi:10.1021/acsnanoscienceau.3c00046.
59. A. Vorobieva, demo_TMB_design: demo_nanopore_design. (2023) doi: 10.5281/zenodo.10935638
60. S. Majumder, MATLAB analysis scripts for denovo Nanopore ion conductance data. (2023) doi: 10.5281/zenodo.10939541
61. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
62. 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).
63. 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).
64. 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. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
65. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
66. 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).
67. A. Favier, B. Brutscher, NMRlib: user-friendly pulse sequence tools for Bruker NMR spectrometers. *J. Biomol. NMR*. **73**, 199–211 (2019).
68. Y. Xia, K. Sze, G. Zhu, Transverse relaxation optimized 3D and 4D 15n/15N separated NOESY experiments of 15N labeled proteins. *J. Biomol. NMR*. **18**, 261–268 (2000).

69. V. Tugarinov, R. Sprangers, L. E. Kay, Line narrowing in methyl-TROSY using zero-quantum 1H-13C NMR spectroscopy. *J. Am. Chem. Soc.* **126**, 4921–4925 (2004).
70. P. Güntert, L. Buchner, Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR*. **62**, 453–471 (2015).
71. 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).
72. S. P. Skinner, R. H. Fogh, W. Boucher, T. J. Ragan, L. G. Mureddu, G. W. Vuister, CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J. Biomol. NMR*. **66**, 111–124 (2016).
73. Y. Shen, A. Bax, Protein Structural Information Derived from NMR Chemical Shift with the Neural Network Program TALOS-N. *Artificial Neural Networks*, 17–32 (2015).
74. Determination and analysis of urea and guanidine hydrochloride denaturation curves" in *Methods in Enzymology* (Academic Press, 1986), vol. 131, pp. 266–280.
75. A. A. Vorobieva, Rationale in Custom Design of Transmembrane β-Barrel Pores. *Methods Mol. Biol.* **2778**, 345–366 (2024).
76. 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).
77. Y. Zhang, J. Skolnick, TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).

Acknowledgments: We thank Jens Gundlach and Andrew Laszlo from the Department of Physics (University of Washington, Seattle) for helpful suggestions regarding conductance measurements, Ines Hertel for protein production, Bob Schiffrin (fitting of CD denaturation data) and many more colleagues at the VUB-VIB Center for Structural Biology, the Institute for Protein Design and the University of Leeds for helpful discussions. We thank the Advanced Light Source (ALS) beamline 8.2.1 at Lawrence Berkeley National Laboratory for X-ray crystallographic data collection. **Funding.** We acknowledge funding from the Howard Hughes Medical Institute (HHMI, to DB), the Flanders Institute for Biotechnology (VIB, to AAV), the Swiss National Science Foundation via the NCCR AntiResist (grant 180541, to SH), the National Institutes of Health (NIH, grant P01 GM072694 to LKT), the EOS Excellence in Research Program of the Research Foundation - Flanders (FWO) and FRS-FNRS (G0G0818N, JW and SER), a Royal Society Professorial Fellowship (RSRP/R1/211057 to SER) and the Air Force Office of Scientific Research (AFOSR, to DB, SB and SM). The Berkeley Center for Structural Biology is supported by the NIH, National Institute of General Medical Sciences, and the HHMI. The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences and US Department of Energy (DOE) (DE-AC02-05CH11231). Computational resources were provided by the VSC (Flemish Supercomputer Center), funded by the FWO and the Flemish Government (projects lt1_2021-32 and lt1_2022-32 to AAV). **Author contributions.** AAV and DB designed the research. SB, SM and AAV designed nanopore proteins. AAV wrote the design and analysis scripts. SB and SM

expressed, purified proteins and ran SEC and CD in detergent. JW performed CD, equilibrium folding/unfolding and folding kinetics in LUVs with help from GNK and supervised by SER and DJB. TM performed initial ^{1}H-^{15}N NMR screening of designs, optimization refolding conditions and solved the NMR structure of TMB12_3, supervised by SH. AB crystallized and solved the X-ray structure of TMB10_163 with help of AK and BS. BL collected initial NMR data on TMB10 designs, supervised by LKT. CB developed the electrophysiology characterization method and collected initial data. SM collected and analyzed electrophysiology data on nanopore designs. AAV and DB wrote the first manuscript draft with support from SH, TM and SM. All authors provided input on the final manuscript. **Competing interests.** AAV and CB are inventors on an E.U. provisional patent application submitted by the Flanders Institute of Biotechnology that covers the sequences of the square-shaped TMB12 designs. **Data and materials availability.** The scripts and the designed protein models are available from GitHub (https://github.com/vorobieva/demo_TMB_design) and will be archived in Zenodo (59). Analysis scripts for processing ion conductance data as presented in this manuscript are also available on Github (https://github.com/sagardipm/denovoPores) and archived in Zenodo (60).The crystal structure of the design TMB10_163 and the NMR structure of TMB12_3 have been deposited in the Protein Data Bank (PDB) (9FDG, 8UZL).

License information: Copyright © 2024 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the Author Accepted Manuscript (AAM) of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

List of Supplementary materials:

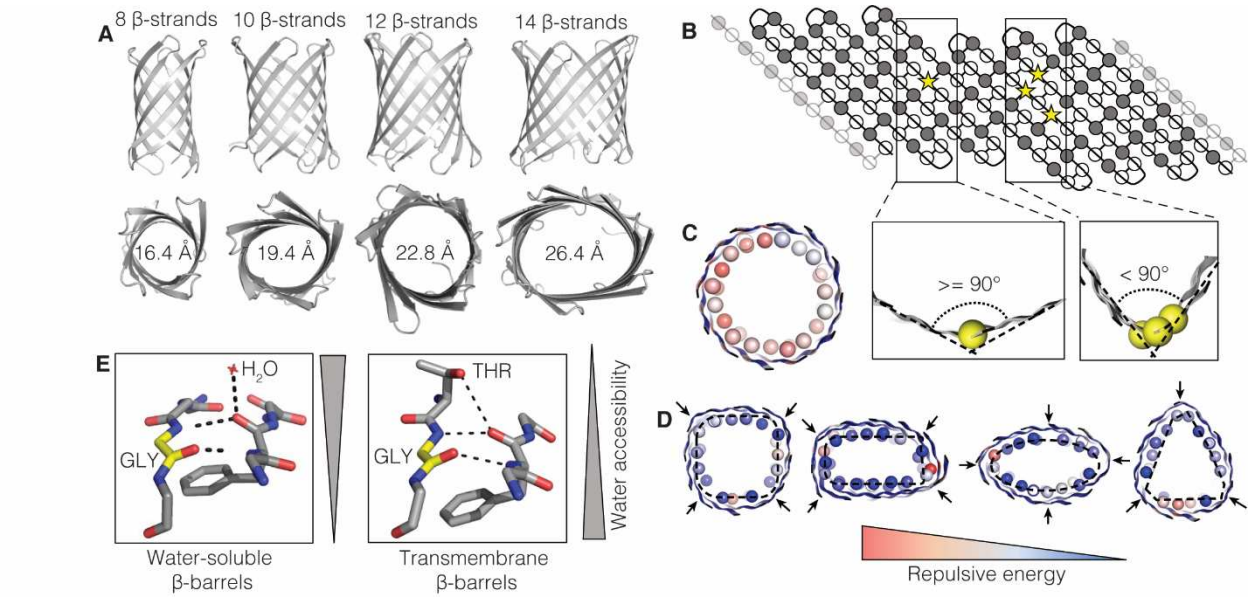

Fig. 1: Sculpting β-barrel geometry. A. Pore diameter can be controlled through the number of β-strands in the β-barrel blueprint. B. β-barrel 2D interaction map. Strong bends in the β-strands (< 90° bend, right) is achieved by stacking several glycine kink residues (yellow spheres) along the β-barrel axis, as opposed to placing one kink (>90° bend, left). C-D. Cross-sections of explicitly assembled β-barrel backbones without (cylinder, C) and with (D) glycine kinks. The Cβ atoms of the residues facing the pore are shown as sphere's and colored based on their respective repulsion energy. Glycine kinks positions are shown with arrows; placement at the corners of the embedded rectangular, oval and triangular shapes (dashed lines in D) generates the desired backbone geometries. E. Polar threonine residues are tolerated on the membrane-exposed surface of TMBs (right) as they can form a hydrogen bond to the backbone, mimicking the interactions with water molecules observed in similarly curved areas of water-exposed β-strands (left).

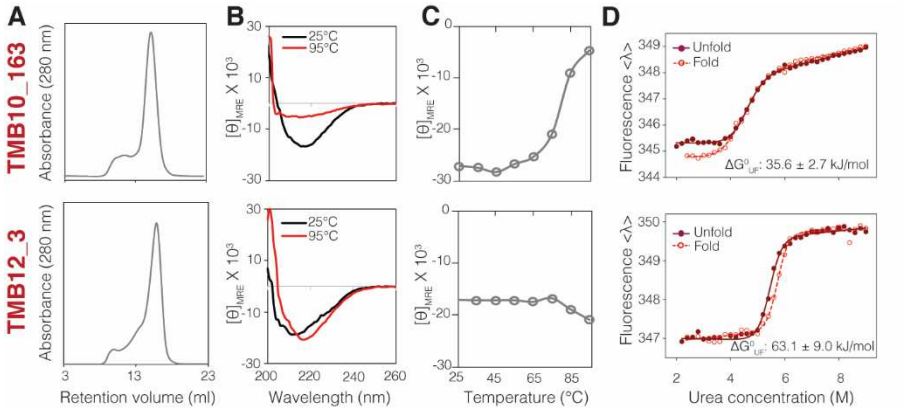

Fig. 2: Biophysical characterization of designed nanopores. Top row: 10-stranded design (TMB10_163); Bottom row: 12-stranded design with a square cross-section (TMB12_3). Both designs elute as one major species with retention time consistent with a monomeric

protein in complex with DPC detergent (A), show distinct negative maxima in far UV CD spectra at 215 nm (B) that remain stable up to >70°C (C), and cooperative and reversible folding/unfolding transitions in DUPC LUVs (where <λ> is the average tryptophan fluorescence emission wavelength (see Methods)) (D)

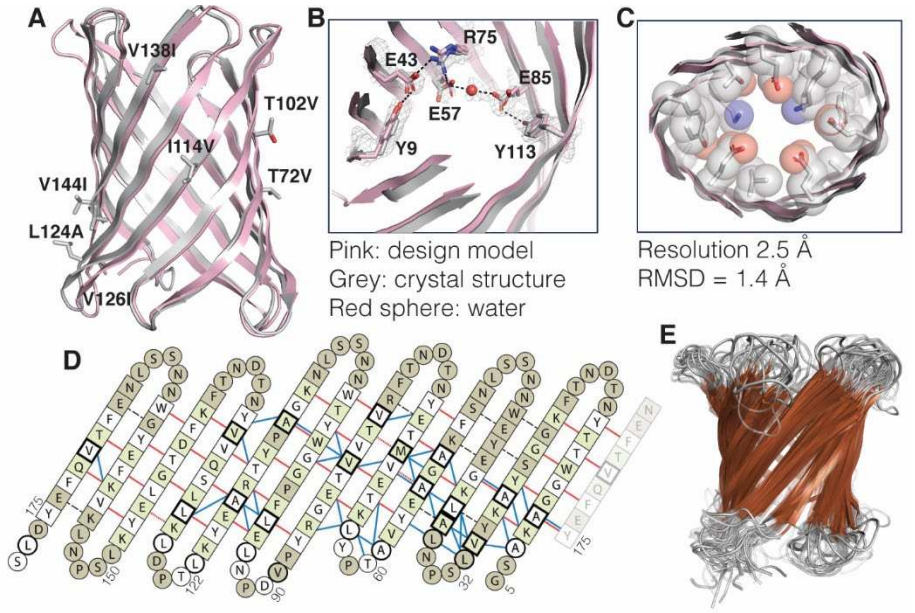

Fig. 3: Experimentally determined nanopore structures closely align with the computational design models. A-C. Crystal structure of TMB10_163 A. backbone superposition. The seven surface residues mutated in TMB10_165 are shown as sticks with the substitution label. B. superposition of side-chains involved in key folding motifs in the lumen, including the $2F_o - F_c$ omit electron density contoured at 1.0 σ. A water-molecule crystallized in the pore is shown as a red sphere. C. cross-sections superposition with residues shown as spheres to highlight the water-accessible pore. D-E. TMB2_13 structure in LDAO micelles. D. Long-range NMR NOE contacts mapped to the expected TMB12_3 hydrogen bonds (dashed black lines). Residues with amide assignment are shown in white and green, unassigned residues are shown in ash gray. Residues with β-sheet secondary structure are shown as squares, all others as circles. Bold outlines indicate available methyl assignments. NOE contacts are shown as red lines (long-range amide-amide, dashes indicating diagonal overlap) and blue lines (contacts involving side chain methyl groups). E. Ensemble of the 20 lowest energy solution NMR structures (β-sheets shown in brown).

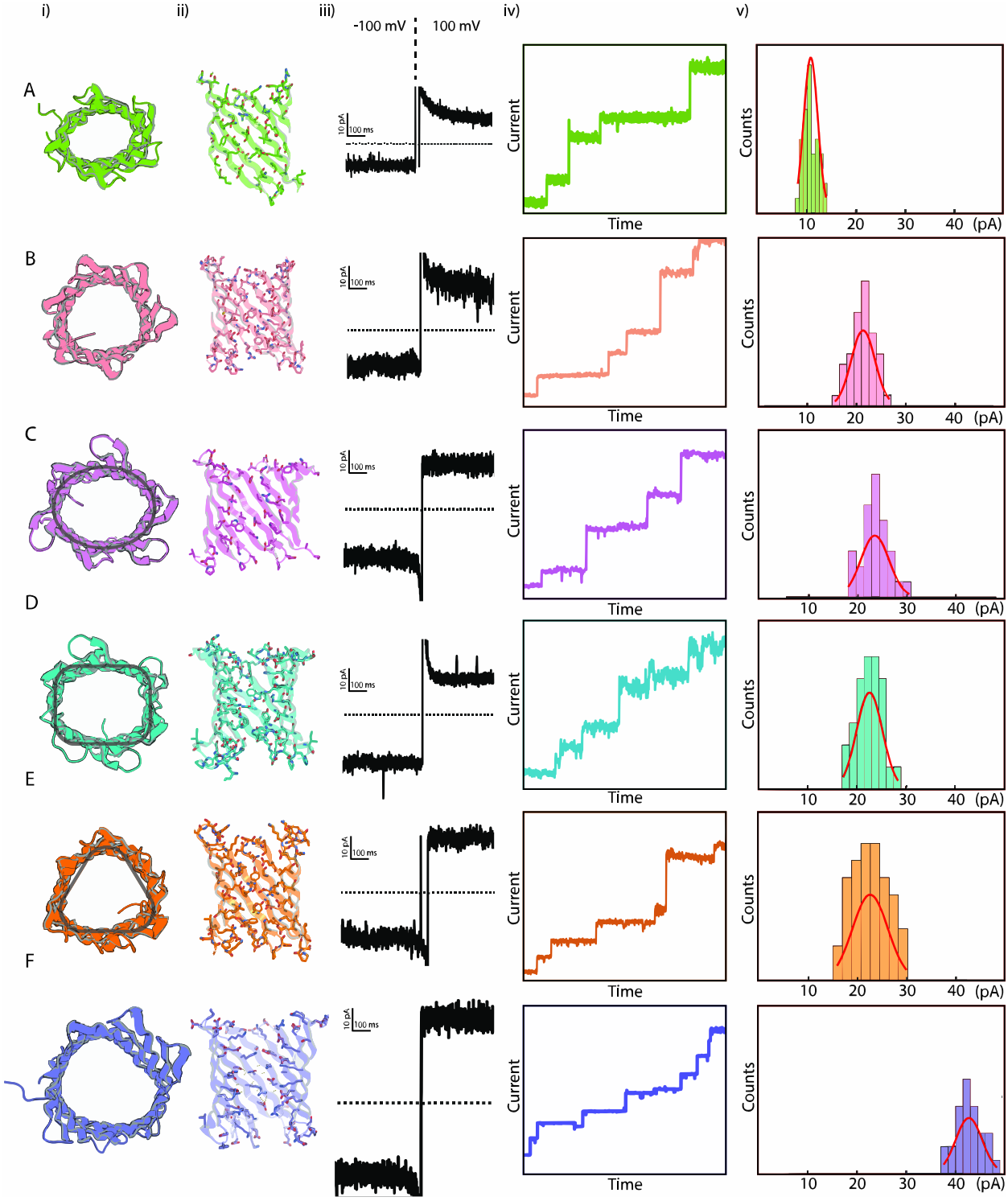

Fig. 4: Conductance of designed nanopores. Designs: A. TMB10_165, B. TMB12_3, C. TMB12_oval_4, D. TMB12_rect_8, E. TMB12_tri_12, F. TMB14_8. i) Top view cartoon representation. ii) Vertical cross sections of the pore. iii) single channel conductance (smallest observed conductance jump). iv) sequential insertions of designed pore in planar lipid bilayer membrane from detergent solubilised sample at low concentrations.

v) histogram of smallest measured current jumps for each design up to 50 pA. The applied voltage across the bilayer was 100 mV and experiments were performed in a buffer containing 500 mM NaCl. A gaussian fit was carried out for the single channel current histograms for each design. For TMB10_165, 38 independent single channel jumps were identified from 3 recordings to plot the histogram shown. Similarly, 44 single channel insertions were identified for TMB12_3 (4 recordings), 29 insertions for TMB12_oval_4 (3 recordings), 30 insertions for TMB12_rect_8 (3 recordings), 45 insertions for TMB12_tri_12 (5 recordings) and 32 insertions for TMB14_8 (3 recordings) to plot the above depicted histograms.

Science
AAAS

Supplementary Materials for

Sculpting conducting nanopore size and shape through de novo protein design

Samuel Berhanu†, Sagardip Majumder†, Thomas Müntener‡, James Whitehouse‡, Carolin Berner, Asim K. Bera, Alex Kang, Binyong Liang, G Nasir Khan, Banumathi Sankaran, Lukas K. Tamm, David J. Brockwell, Sebastian Hiller, Sheena E. Radford, David Baker*, Anastassia A. Vorobieva*

† Equal contribution
‡ Equal contribution

*Correspondence to: anastassia.andreevna.vorobieva@vub.be, dabaker@uw.edu

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:

Material & Methods

General purification of all designs

All designs were purified from *E. Coli* following a similar protocol as described previously in *(13)*. Custom genes in a pET29b vector containing the kanamycin resistance gene were ordered from IDT and chemically transformed into BL21DE3 cells. All proteins were purified from inclusion body fractions following complete denaturation in 6M GuCl (Guanidine Hydrochloride) buffer. Briefly, inclusion pellets were washed several times with buffers containing 1% w/v of Triton X-100 and Brij-35 alternatively. A typically washing step involved resuspension of the insoluble pellet in the appropriate buffer, brief sonication and subsequent incubation for one hour at room temperature or overnight at 4°C. After solubilisation of the pellet in GuCl, the protein was diluted to 80-100 uM and refolded in a buffer containing 25 mM Tris-Cl at pH 8.0, 150 mM NaCl and 0.1% DPC (Dodecyl-Phosphatidyl-Choline) either using a dropwise dilution method or by spontaneous dilution to achieve a final GuCl concentration of 0.3 M. The diluted buffer was concentrated after overnight incubation with shaking at 4°C and run on a S200 Cytiva superdex 200 column. Fractions at expected volume were concentrated with a 10 kDa cutoff filter and used for subsequent analysis.

Conductance measurement in planar lipid bilayers

All ion-conductance measurements were carried out using the Nanion Orbit 16TC instrument (https://www.nanion.de/products/orbit-16-tc/) on MECA chips. Lipid stock solutions were freshly made in dodecane at a final concentration of 5mg/mL. DPhPC (Di-Phytanoyl-Phosphatidyl-Choline) lipids were used for all experiments. Designed proteins were diluted in a buffer containing 0.05% DPC (~ 1 CMC), 25 mM Tris-Cl pH 8.0 and 150 mM NaCl to a final concentration of ~100 nM. Subsequently, 0.5 µL or less of this stock was added to the cis chamber of the chip containing 200 µL of buffer while simultaneously making lipid bilayers using the in-built rotating stir-bar setup. All measurements were carried out at 25°C. Spontaneous insertions were recorded over multiple rounds of bilayer formation. All chips were washed with multiple rounds of ethanol and water and completely dried before testing subsequent designs. A 500 mM NaCl buffer was used on both sides of the membrane for all current recordings. Raw signals were recorded at a sampling frequency of 5 kHz. Only current recordings from bilayers whose capacitances were in the range 15-25 pF were used for subsequent analysis. The raw signals at 5 kHz were downsampled to 100 Hz using an 8-pole bessel filter. Estimation of current jumps were carried out using a custom script with appropriate thresholds. Current jumps larger than 2 times the smallest observed jump were discarded for single channel histogram calculations for each design.

Crystallography and structure determination of TMB10_163

SEC purified sample, at the concentration of 14 mg/ml, was used for crystallization. The crystallization screening was performed using a Mosquito LCP by STP Labtech. Crystals grew successfully in 3.25 M 1,6-hexanediol and 0.01 M HEPES pH 7.5. Crystals were harvested directly from a screening tray, and flash cooled in liquid nitrogen. X-ray diffraction was performed at ALS beamline 8.2.1, data were processed with XDS (*61*), and merged/scaled using Pointless/Aimless in the CCP4 program suite (*62*). The structure was phased by molecular replacement using the designed structure as the search model by Phaser (*63*) and refined with Phenix (*64*). Following molecular replacement, the models were improved, and efforts were made to reduce model bias. Structures were refined in Phenix. Model building was performed using COOT (*65*). The final model was evaluated using MolProbity (*66*). Data collection and refinement statistics are recorded in Table S1. Data deposition, atomic coordinates, and structure factors reported in this paper have been deposited in the Protein Data Bank (PDB), http://www.rcsb.org/ with accession code 8UZL.

TMB12_3 Expression for NMR. BL21(DE3) Lemo cells were transformed with a pET29b-derived expression plasmid for TMB12_3. Cells were grown in M9 minimal medium and expression was induced with 0.5 mM IPTG for 22h at 24°C. For expression of [*U*-99% ^{2}H, ^{15}N, ^{13}C]-labeled samples, M9 was prepared with D_2O, $^{15}NH_4Cl$ and deuterated-^{13}C-glucose. For the expression of the [*U*-99%-^{2}H, ^{15}N] labeled sample, M9 was prepared with D_2O and $^{15}NH_4Cl$. For the selectively labeled [^{15}N-Lys] and [^{15}N-Phe] samples, the desired ^{15}N labeled amino acid was added to the culture 45 min before induction.

Purification and Refolding of TMB12_3. Cells were lysed using a M110L from Microfluidics. Inclusion bodies were isolated and dissolved in denaturing buffer (20 mM Tris/HCl pH 8, 150 mM NaCl, 6M GdnHCl), then dialyzed against H_2O in a 10,000 MWCO dialysis membrane for 2h, followed by centrifugation at 30,000 g to precipitate the protein. Precipitated TMB12_3 was dissolved in 10 mM Tris/HCl pH 8, 7 M urea. Refolding was done at 4°C by dropwise rapid dilution into a stirred refolding buffer (20 mM Tris/HCl, 2 mM EDTA, 0.6 M L-Arg, 15 mM LDAO, pH 10). The dilution ratio was set to 1:20 and after overnight stirring, the refolded protein was dialyzed against 20 mM NaPi pH 6.8, 1 mM EDTA for 2h. The refolded protein was concentrated with MWCO 10,000 and the sample was loaded on an S200 size exclusion column, pre-equilibrated with 20 mM NaPi, 1 mM EDTA, 15 mM LDAO. The fractions containing protein were pooled and concentrated using MWCO 10,000.

Isotope labeling samples. The following samples were made: [U-99%-^{15}N]-TMB12_3 in LDAO, [U-99%-^{2}H, ^{15}N, ^{13}C]-TMB12_3 in LDAO, [U-99%-^{2}H,^{15}N; 99%-$^{1}H^{\beta}$,$^{13}C^{\beta}$-A; 99%-$^{1}H^{\varepsilon}$,$^{13}C^{\varepsilon}$-M; 99%-$^{1}H^{\delta1}$,$^{13}C^{\delta1}$-L; 99%-$^{1}H^{\gamma1}$,$^{13}C^{\gamma1}$-V]-TMB12_3 in [U-99%-^{2}H]-LDAO, [^{15}N-Lys]-TMB12_3 in LDAO, [^{15}N-Phe]-TMB12_3 in LDAO. The final sample conditions were 20 mM Na·PO_4, 1 mM EDTA, pH 6.8, 300–500 mM LDAO, 0.2–1 mM TMB12_3.

NMR experiments. All experiments were carried out at 25°C on Bruker spectrometers operating at field strengths of 700, 800 and 900 MHz. All spectrometers were equipped with a cryogenic triple-resonance probe. The following experiments were recorded: 2D [^{15}N, ^{1}H]-BEST-TROSY (*67*), 3D BEST-TROSY-HNCACB with ^{2}H decoupling (*67*), 3D [^{1}H,^{1}H]-NOESY-^{15}N-TROSY (*68*), 2D ^{13}C-Methyl-SOFAST (*67*), 3D [^{1}H,^{1}H]-NOESY-^{13}C-HMQC (*69*).

Structure calculation. Structure calculation was performed with CYANA 3.98.15 (*70*). All spectra were processed and analyzed with NMRPipe (*71*) and ccpNMR version 3 (*72*). 129 dihedral constraints were derived by TALOS-N (*73*) from the experimentally determined C_α, C_β, N and HN chemical shifts. Only TALOS-N predictions classified as “strong” were used. Tolerances were set to one standard deviation, capped at a maximum of 20°. A total of 205 experimental NOEs were obtained from 3D [^{1}H,^{1}H]-NOESY-^{15}N-TROSY and 3D [^{1}H,^{1}H]-NOESY-^{13}C-HMQC spectra. 97 Hydrogen bond constraints were inferred from the measured NOE data with upper limits of 2.0 / 3.0 Å and lower limits of 1.8 / 2.7 Å for the HN…O and N…O, respectively. In regions with sparse assignment, these constraints were inferred indirectly from the experimentally established β-strand topology (dashed black lines in Figure 3D). A total of 500 structures were calculated and the ensemble of 20 lowest energy structures was selected. Ramachandran statistics of this ensemble showed 81.1 % of residues in most favored regions, 17.6 % in allowed regions, 1.2% in generously allowed regions and 0.1% in disallowed regions.

Calculating the theoretical diameters of nanopores. The diameters of the nanopores were inferred from the observed single-channel conductance (*G* (nS)) using the access resistance model (**eq. 1**), where R is the resistance, S the conductivity of the solution (S/m), d is the diameter of the pore (nm) and L is the pore length (nm). The purely geometric model approximates the

properties of a cylindrical nanopore and assumes homogeneous solution, pore and membrane neutrality and constant potential at the pore mouth.

$$G = \frac{1}{R} = \frac{S\pi d^2}{4L} \quad \text{eq. 1}$$

A pore length of 3.5 nm was used for all calculations based on the total transmembrane span of TMB designs. The conductivity of a solution of 0.5 M NaCl was estimated to 40.5 mS/cm based on the previously reported relationship between NaCl molarity and conductivity of the solution.

Folding kinetics measured by Tryptophan fluorescence. Protein samples were buffer exchanged into unfolding buffer (50 mM Glycine-NaOH pH 9.5, 8 M/10 M urea) using a 0.5 mL ZebaSpin 7K MWCO desalting column. The concentration was determined by nanodrop. For kinetic experiments, 15 µL of unfolded protein was added to 485 µL of pre-warmed (25°C) pre-fold buffer of LUVs in 50 mM glycine-NaOH, pH 9.5, in a QS quartz cuvette to give final concentrations of 0.4 µM OMP, 600-3200 LPR (*mol:mol*), 0.24-4 M urea, 50 mM glycine-NaOH pH 9.5. Immediately after mixing, a time based fluorescence scan was carried out on a PTI QuantaMaster™ spectrofluorometer (Photon Technology International), controlled by FelixGX v4.3 software, with excitation at 280 nm, and emission measured at 335 nm. The slit settings were 0.5 nm for excitation, and 5 nm for emission, to minimize photobleaching. Integration was set at 1 s between time points, and the temperature was maintained at 25°C throughout. Fluorescence emission spectra were measured by exciting tryptophan at 280 nm, and measuring fluorescence emission between 300-400 nm, using the same slit-width settings as above, with samples in urea concentrations between 0.24-9.9 M urea.

Circular dichroism in liposomes. Protein samples were prepared in a similar manner as for the tryptophan fluorescence samples. Samples were made to a 600:1 (*mol:mol*) LPR, with final concentrations of 4 µM TMB, 1.2 mM lipid-LUV, 0.24-8 M urea, 50 mM glycine-NaOH pH 9.5 in a final reaction volume of 300 µL. The reaction was allowed to proceed overnight at 25°C to maximize the fraction of protein folded into the lipid-LUV bilayer. Controls were made where the volume of substrate was replaced with 50 mM glycine-NaOH pH 9.5 and an appropriate volume of urea to match the protein samples. These were used to normalize the data by subtracting their CD signals from the CD signal from the protein containing samples. Measurements were taken using 300 µL of sample in a 1 mm QS quartz cuvette, using a Chirascan plus CD Spectrometer (Applied Photophysics). The bandwidth was set at 2.5 nm, and used adaptive sampling to adjust the integration time for the optimal signal:noise. Four scans were averaged between 260 nm to the lowest useable wavelength for each respective sample, which was the point where the voltage reached its upper limit of 1000 V, after which the data became unusable. During temperature ramp experiments, only single scans were taken as the temperature ranged between 25°C and 87°C.

Equilibrium denaturation analysis. To determine the urea dependence of TMB folding, urea denatured TMBs in 50 mM glycine-NaOH pH 9.5, 10 M urea were diluted into DUPC LUVs at an Lipid-to-Protein ratio (LPR) of 600:1 (mol/mol) to give a final concentration of 0.4 µM TMB in 50 mM glycine-NaOH pH 9.5 containing 2-9.9 M urea, and folding was allowed to proceed overnight at 25°C. For urea dependence of unfolding, TMBs were folded in DUPC LUVs (LPR 600:1 (mol/mol)) in 50 mM glycine-NaOH pH 9.5, 2 M urea overnight at 25°C. Pre-folded TMBs were then unfolded by dilution into 50 mM glycine-NaOH pH 9.5 containing 2-9 M urea to a final TMB concentration of 0.4 µM and incubated overnight at 25°C. Tryptophan fluorescence emission spectra were obtained using a PTI QuantaMaster spectrofluorometer (Photon Technology International) in QS quartz cuvettes with excitation slits set to 1 nm and emission slits set to 5 nm. Fluorescence was excited at 280 nm and emission spectra were acquired between 300-400 nm using a step size of 1 nm and an integration time 0.5 seconds. Average wavelength between 325-

375 nm was calculated using equation 2, where <λ> is the average wavelength, I_λ is the fluorescence intensity at a given wavelength, λ is the wavelength, and ∑I is the sum of the intensity of the entire emission spectra.

$$<\lambda> = \frac{\Sigma I_\lambda * \lambda}{\Sigma I} \qquad \text{eq. 2}$$

The experimental data were fitted to a 2-state transition model (*74*) to extract ΔG_0 (the Gibb's free energy for unfolding in the absence of denaturant), the m-value (m_{UF}, the global dependence of ΔG_0 on the concentration of denaturant and C_m (the transition midpoint) based on Obs_F and Obs_U (the observed <λ> for the folded and unfolded states in the absence of denaturant ([D]=0)), m_f and m_u (the linear dependence of Obs_F and Obs_U to [D]). The observed <λ> was corrected to account for the difference in quantum yield between the folded and unfolded states based on the Q-factor (QF), calculated by taking the ratio of the summed fluorescence intensities at the folded and unfolded states. R is the universal gas constant and T is the absolute temperature.

***De novo* backbones assembly.** The entire computational design pipeline has been described in detail in (*75*). The Rosetta blueprint representations of the beta-barrel backbones were generated based on user input using a custom python script available on GitHub https://github.com/vorobieva/demo_TMB_design/tree/master/generate_blueprint. The script requires the SciPy and BioPython modules and generates a Rosetta blueprint (describing local secondary structure and torsion angle bins per residue) and a Rosetta constraints file (describing backbone-backbone hydrogen bonds in the barrel). Examples of blueprint and constraints files used in this study are available on GitHub https://github.com/vorobieva/demo_TMB_design/tree/master/12_strands_square/assemble_backbones. The backbones were assembled based on such a blueprint and constraints file using the Rosetta BluePrintBDR application (*19*), alternating between sampling of backbone fragments and minimization with hydrogen bonds constraints. The highest-quality protein backbones (250-500 backbones, based on Rosetta `vdw`, `omega` and `rama_prepro` scores) were selected as template for combinatorial sequence design.

Combinatorial TMB sequence design. *De novo* β-barrel backbones assembled using the Rosetta coarse-grained centroid model were subject to one round of fast atomistic refinement of the backbones with the Rosetta full-atom model (`ref2015` energy function with limited sampling depth `-nstruct 1` and limited sequence space). The Tyr-Gly-Asp/Glu TMB folding motifs were then designed into the structures using the HBNet (*76*) Rosetta application, which finds all possible positions of the hydrogen bond acceptor residue in the motif (Asp or Glu) based on defined tyrosine positions. The refined TMB backbones were used as templates for several rounds of combinatorial sequence design, alternating between the water-accessible pore (two rounds of design) and the lipid-exposed surface (three rounds of designs) using re-fitted energy functions (see below). Rosetta resfiles were used to define the set of amino acids sampled at each position in the protein. The surface-exposed residues were constrained to mostly hydrophobic amino acids while all amino acids but PRO and CYS were allowed at pore-lining positions. The β-turn residues were designed using previously identified canonical TMB β-turn sequences (Figure S2). At each iteration of sequence optimization, the whole population of designs was analyzed, population-wide selection metrics were computed (see below) and around 10 % of the designs were selected to be used as input for the next iteration. After an iteration of pore-residue design, the outputs were selected based on backbone quality metrics (Rosetta `omega`, `rama_prepro`, `hbond_lr_bb` scores) and on the computed `total` and `hbond_sc` energies of the hydrogen bond acceptor residues in the Tyr-Gly-Asp/Glu folding motifs. After an iteration of surface-exposed residue design, the outputs were selected based on Rosetta

total_energy and on the retention of the Tyr-Asp/Glu interactions in the designed folding motifs (which are repacked while the surface residues are designed). The resfiles used in this study to generate the different TMB architectures are available on GitHub (https://github.com/vorobieva/demo_TMB_design/tree/master). A complete description of the design pipeline, analysis scripts and example inputs are available on GitHub (https://github.com/vorobieva/demo_TMB_design/tree/master).

Re-fitted TMB-specific energy functions. To fine-tune the amino acid propensities to TMB-specific statistics, the Rosetta reference energy function (ref2015) was modified by testing several variations of weights on one representative TMB backbone. To generate an energy function to design the water-accessible pore-lining residues, the weight of the Rosetta full-atom solvation energy (fasol), electrostatic energy (faelec) and of the reference energies of small disorder-promoting amino acids (ALA and SER) were systematically varied. The scoring function used for subsequent sequence design was selected based on the closest match between the resulting designed sequences and naturally-occurring TMB sequences at the level of the overall hydropathy of the pore and the frequency of ALA and SER amino acids. To generate an energy function to design the lipid-exposed surface residues, the weight of the Rosetta full-atom solvation energy and of the reference energies of large hydrophobic (PHE) and of small disorder-promoting amino acids (GLY and ALA) were systematically varied. A scoring function was selected based on the closest match between the resulting designed sequences and naturally-occurring TMB sequences at the level of the overall hydropathy of the surface and the frequency of ALA and GLY amino acids. The TMB10 variants (TMB10_163-6), differing by only a few surface mutations, were obtained by resampling hydrophobic and Threonine residues on the β-barrel surface using Rosetta FastDesign ('round4' energy function and resfile available in the GitHub repo) and selecting low-energy models.

Comparison of the scoring weights optimized to design the pore residues of TMB12s designs

	TMB12_square	TMB12_oval	TMB12_rectangle	ref2015
fasol	0.9	0.9	0.9	1.0
faelec	1.4	1.45	1.45	1.0
ALA refE	0.4	-0.6	-0.6	1.32468
SER refE	-1.2	-1.8	-1.8	-0.28969

Designs validation and selection. The final designs were filtered based on the desired balance between local secondary structure frustration (β-sheet propensity of the sequence between 30 % and 50 % (predicted with RaptorX (*28*)) and aggregation propensity score predicted with Tango (*29*) smaller than 1500) and sequence-encoded tertiary structure. The sequence/structure compatibility was assessed based on the capacity of AlphaFold2 (*30*) to fold the sequence into the designed TMB structure in single sequence mode (no multiple sequence align input) and using 48 recycles through the network. Selected predictions had AlphaFold2 plDDT scores higher than 0.8 and showed high structure similar to the expected design model when superimposed with TMAlign (*77*), with Root Mean Square Deviations (RMSD) of < 2.1 Å.

Supplementary Table

Table S1. X-ray diffraction data collection and refinement statistics

	TMB10_163 (8UZL)
Resolution range	49.92 - 2.5 (2.58 - 2.5)
Space group	*P* 2_1
Unit cell	55.62, 52.55, 100.05, 90, 93.67, 90
Unique reflections	20045 (1961)
Multiplicity	4.3 (4.5)
Completeness (%)	98.87 (98.54)
Mean I/sigma (I)	4.3 (0.4)
Wilson B-factor	43
R-merge	0.126 (0.923)
R-pim	0.069 (0.493)
$CC_{1/2}$	0.99 (0.83)
Reflections used in refinement	20021 (1960)
R-work	0.227 (0.336)
R-free	0.262 (0.383)
Number of non-hydrogen atoms	4535
macromolecules	4420
ligands	16
solvent	99
Protein residues	568
RMS (bonds)	0.002
RMS (angles)	0.45
Ramachandran favored (%)	96.96
Ramachandran allowed (%)	3.04
Ramachandran outliers (%)	0.00
Average B-factor	52
macromolecules	52
ligands	52
solvent	50

Statistics for the highest-resolution shell are shown in parentheses.

Table S2. NMR and refinement statistics for TMB12_3 in LDAO micelles

	TMB12_3 (9FDG)
NMR distance and dihedral constraints	424
Distance constraints	302
Total NOE	205
Intra-residue	22
Inter-residue	183
Sequential ($\|i - j\| = 1$)	59
Medium-range ($\|i - j\| < 4$)	44
Long-range ($\|i - j\| > 5$)	102
Intermolecular	0
Hydrogen bonds	97
Total dihedral angle restraints	122
f	61
y	61
Structure statistics	
Violations (mean and s.d.)	
Distance constraints (Å)	0.04 +/- 0.01
Dihedral angle constraints (°)	0.9 +/- 0.4
Max. dihedral angle violation (°)	6.3 +/- 3.1
Max. distance constraint violation (Å)	0.71 +/- 0.22
Ramachandran plot statistics	
Residues in most favored regions	81.1 %
Residues in additionally allowed regions	17.6 %
Residues in generously allowed regions	1.2 %
Residues in disallowed regions	0.1 %
Average pairwise r.m.s. deviation* (Å)	
Heavy	2.94 +/- 0.37
Backbone	2.28 +/- 0.41

* Calculated among 20 refined structures

Table S3. “Optimal” designs that failed to express

Designs	Amino acid sequence	BLAST E-value
TMB10_r1_1	MSNSPGARLNIIFRYRSDGTFELTIQFGFKFTVYPEVSIEFQFG VTYKGDSIVEFSDTLGFEFKVYPTLSFYFGFGIQYEDSNRFALT FKFGVKFKLWDPIRLDITVGIRYAGDDSVTVYVGVTLEIQVYPS FEVYFGFGVEYIGESILKFGFEIGFTYW	0.022 (unnamed protein [*Nesidiocoris tenuis*])
TMB10_r1_2	MDRSPGTRIRIEVTYASDGDFQITVTVGIEFTLYPEITFEIAVG FTYRGDSKFAIRFEIGVKFKVYPTILVYFGFGIDWSESNEFEIV FEIGFRVKVYDPFEIELRFGVKYQGLDVFTIYAGITFHFTVYPS VTFYVGFGFEYFGDSIVRFGVTFGFQYN	0.05 (unnamed protein [*Nesidiocoris tenuis*])
TMB10_r1_3	MKPSPGLSIEITVQYNSDGSVDFTVKVGFTIFVYPTVAVTFEVG VTYHGDSRVTFHFTFGFTVTVYPTLTFYAGFGFQWEDSNTFAIQ FDFGFTIKLYDPVRIYFDAGFAYKGDDKVTFYAGFALQITVYPS VEFYVGVGFQYTGDSTSNAGIRFGVQYK	No match
TMB10_r1_4	MKPSPGVRIIIEVQYNSDGSVDFTVKVGFTIFVYPTVAVTFEVG VTYRGDSTVTFHFTFGFSVTVYPTLTFYAGFGFTWDDSNRFAIQ FEAGFTITLYDPVQLYFDFGFTYAGDDKVTFYAGFALQVQVYPS VKFYVGVGFEYTGDSKSNAGIKFGVQYD	No match
TMB10_r1_5	MKPSPGARIVVEVQYNSDGSVDFRVKVGFTIFFYPTLAVTFEVG VTYRGDSRVTLHFTFGFTVTVYPTLTFYAGFGFTWKDSNTFAIQ FDAGFTIKLYDPIRLYFDFGFQYAGDDKLDFYAGFAIEFDVYPS VTFYVGFGAEYTGDSTSNAGVKFGVTYY	0.044 (unnamed protein [*Closterium sp.*])
TMB10_r1_6	MKPSPGVRIIIEVQYNSDGSVDFTVKVGFTIFFYPTVAITFEVG VTYRGDSTVTFHFTFGFSVTVYPTLTFYAGFGFTWDDSNRFAIQ FEAGFTITLYDPVQLYFDFGFTYAGDDKVTFYAGFALQIQVYPS FKFYVGVGFEYTGDSKSNAGIKFGVQYD	No match
TMB10_r1_7	MKPSPGVRIIVEVQYNSDGSVDFTVKVGFTIFFYPTVAVTFEVG VTYRGDSTVTFHFTFGFSVTVYPTLTFYAGFGFTWDDSNRFAIQ FEAGFTITLWDPVQIYFDFGFTYAGDDKVTFYAGFALQIQVYPS FKFYVGVGFEYTGDSKSNAGIKFGVQYD	No match
TMB10_r1_8	MKPSPGVRIIIEFQYNSDGSVDLTIKVGFTIFVYPTVAVTFEVG VTYRGDSTVTFHFTFGFSVTVYPTLTFYAGFGFTWDDSNRFAIQ FEAGFTITLYDPVQIYFDFGFTYAGDDKVTFYAGFALQIQVYPS VKFYVGVGFEYTGDSKSNAGVKFGVQYD	No match
TMB12_r1_11	MSQSPGLKFYITFRWNSDGKLQVEVGFKIELFFYPTVFISLIFG FVWDGDSKLFFQISDGVEFQVWPDLSFYFGAGIRWSSSNKIEIE FKAGFKIRLYDPFEVIFTFGFTWKGDSTLAIQFGITFTFTFYPT FKIQFGFGVQWEGDSKVEFTFGFEIEVKVYPPVKFRFGAGVIWV GDSQLHVGVYVGYEIK	0.038 (unnamed protein [*Nesidiocoris tenuis*])

TMB12_r1_12	MDPSPGVRFYVEFIWRSDGEVTFKAGIQFIFIVYPELTITFEIG FTWEGDSTFAIRVTFGFKVRFWPTLEVYVGFGVQWAEDNKIRFE FKAGIIIKIYDPITLEFRFGVTWFGDSIVKIDFGVAITFRVYPT FEITFGVGATWQGDSKFAINIGVEFRFTLYPPAIITVGFGFTWE GDSTLKFGFYFGYTIE	0.046 (unnamed protein [*Nesidiocoris tenuis*])
TMB12_r1_13	MQDSPGLNFYIKIEWRSDGFLLIFVGFRIQIFLYPNFTIEIEIG VVWQGDSLFSIRITFGFIINLWPTVQFYAGVGVEWSSSNKFTIT FEAGFTIKFYDPLRIVFFFGFQWEGDSKLTIRAGFTVQITFYPS FQVEFGFGFVWFGDSILVFRFGFRIEVQVYPPATVEFGVGVDWR GDSIVTVGVYFGYKIH	0.04 unnamed protein [*Closterium sp.*])
TMB12_r1_14	MDPSPGVRFYVEFIWRSDGEVTFKFGIQFIFIVYPELTITFEIG ATWEGDSTFAIRVTFGFKVRFWPTFEVYVGFGVQWAEDNKIRFE FKAGIIVKLYDPITLEFRFGVTWFGDSIVKIDFGVAITFRVYPT FEITFGVGATWQGDSKFAINIGVEFRFTFYPPAIFTFGFGFTWE GDSTLKFGFYFGYTIE	0.048 unnamed protein [*Nesidiocoris tenuis*])
TMB12_r1_15	MDPSPGVRFYFEVIWRSDGEVTIKFGIQFIFIVYPEFTLTFELG ATWEGDSTFAIRVTFGFKIRIWPTVEFYFGFGLQWAEDNKLRFE FKAGIIIKIYDPITLEFRFGLTWFGDSILKVDFGVAITFRVYPT FEITFGVGATWQGDSKFAINIGVEFRFTVYPPAIITVGFGFTWE GDSTLKFGFYFGYTIE	0.025 unnamed protein [*Nesidiocoris tenuis*])
TMB12_r1_16	MDPSPGVRFYVEFIWRSDGEVTIKFGLQFIIIVYPELTITFEIG ATWEGDSTFAIRVTFGFKVRVWPTIEFYFGFGVQWAEDNKIRFE FKAGIIIKLYDPLTLEFRFGVTWFGDSIVKVDFGVAITFRVYPT FEITFGVGATWQGDSKFAINIGVEFRFTVYPPAIITVGFGFTWE GDSTLKFGFYFGYTIE	0.044 unnamed protein [*Nesidiocoris tenuis*])
TMB12_r1_17	MSDSPGVAFYVKFEWRSDGTFRIEFGVTFKIIVYPDVIFQLTIG FIWEGDSKLKIQVTVGVEIRVWPTLAFYFGTGVRWTSENDVQID IQVGVKIKVYDPFIIDIRFGFTWRGDSKFEFTVGLTLQFRFYPT FTVNFGFGVTWKGDSKVTFEFGVEFRFRVYPPVEIDFGFGAIWF GDSFLQVGFYVGYKVK	0.05 unnamed protein [*Nesidiocoris tenuis*])
TMB12_r1_18	MSQSPGLKFYVTFRWNSDGKLQVEVGFKIEIFVYPTIFISLIFG FVWDGDSKLFIQISDGIEFQVWPDLSFYFGAGIRWSSSNKVEIE FKAGFKIRLYDPFEVIFTFGFTWKGDSTLAIQFGITFTFTFYPT FKIQFGFGVQWEGDSKVEFTFGFEIEVKVYPPVKFRFGAGVIWV GDSQLHVGFYVGYEIK	0.02 unnamed protein [*Nesidiocoris tenuis*])

Table S4. Experimentally tested designs

Design	Amino Acid Sequence	Expression	CD (DPC micelles)	SEC (DPC micelles)	Conductance	BLAST E-value*
TMB10_163	MRTSPGTKPYVKVRW NTDNTVAVAFGAETD YKLAPYLKTGVATET EYNNSSLVKTGTEVK TAYRLGPNAALETVV RYNTDNTFGVEVAIE YRLEPDLSVAPGTRW NNSSLLAPYIKIKYK LGPDLDVVTTIAYNT DNTVGIETKVAYKTD	Yes	Negative maxima CD spectra at 215 nm	Mono-dispersed monomer	No	No match
TMB10_164	MRTSPGTKPYVKVRW NTDNTVAVAVGAETD YKLAPYLKTGVATEV EYNNSSLTKAGTEVK TAYRLGPNAALEVVV RYNTDNTFGVEVAIE YRLGPELSIAPGVRW NNSSLLAPYLKIKYK LGPDLDIVTTIAYNT DNTVGIETKIAYKTD	Yes	Negative maxima CD spectra at 215 nm	Major monomer	No	0.011 (voltage-dependent channel)
TMB10_165	MRTSPGTKPYVKVRW NTDNTVAVAFGAETD YKLAPYLKTGVATET EYNNSSLVKTGTEVK TAYRLGPNAALEVVV RYNTDNTFGVEVAIE YRLEPDLSVAPGVRW NNSSLLAPYVKIKYK LGPNADIVTTIAYNT DNTIGIETKIAYKTD	Yes	Negative maxima CD spectra at 215 nm	Major monomer	**Yes**	No match
TMB10_166	MRTSPGTKPYVKVRW NTDNTVAVAFGAETD YKLAPYLKTGVATET EYNNSSLVKTGTEVK TAYRLGPNAALETVV RYNTDNTFGVEVAIE YRLEPDLSVAPGVRW NNSSLLAPYVKIKYK LGPDLDVVTTIAYNT DNTVGIETKIAYKTD	Yes	Negative maxima CD spectra at 215 nm	Mono-dispersed monomer	**Yes**	No match
TMB12_0	MQEKPGSAEGGTRTL YNTDNTLKSGGYGVY VLSPELVLFAAYFWN NSSLQEFVAGAKYKL	Yes	Negative maxima CD spectra at 215 nm	Major monomer	No	No match

	SPYLETEVHLRYNTD NTFAVDVTTEGEYPV SPELKFRPGATYRWN NSSLNKLRPYLKVEY KLSPDLKGVVEVQYN TDNTVLVWFGATYKL SPDLEVTVMYGWNNS SLNWLLVDVKYKLSE					
TMB12_2	MQDRPGTLKVGGRTV YNTDNTFKSGGYAVY VLTPDLAGKTKYEWN NSSLQSFEFGGQYKL SPYARTEVSVQYNTD NTVKFRVVVEGEYPL SPNLRAFPGAEYVWN NSSLNKLFPYLRFEY DLSPELLGRLEFWYN TDNTFKVKLGAEYKL TPNLSVLVMYGWNNS SLQEFETETRYDLSS	Yes	Negative maxima CD spectra at 215 nm	Mono-dispersed monomer	No	No match
TMB12_3	MQDKPGSAKAGGWTT YNTDNTFKGGSYAKY VLSPNLALKGEYEWN NSSLNSFKAGAEYVA TPYLKTEVMTEYNTD NTFRVTVVTEGRYPV DPNLELFPGGWYTWN NSSLNKGAPYTRAEY KLTPDLKLLSQVVYN TDNTFKFDTGLEYKL SPNLKVKFEYGWNNS SLNEFTVQFEYDLSS	Yes	Negative maxima CD spectra at 215 nm	Mono-dispersed monomer	**Yes**	No match
TMB12_4	MQDKPGTVKGGGKGQ YNTDNTVKGGGYAVY TLSPDLSGKTEYLWN NSSLQELRVGARYRL SPYLETDVEVAYNTD NTFVLRVDTRGEYPL SPELKLFPGAKYAWN NSSLQKGSPYATFVY HADPNLLFKVTFRYN TDNTVEVEVGAEYKL SPNLSTFTAYGWNNS SLNWVRVETRYLFTY	No	N.D.	N.D.	N.D.	No match
TMB12_5	MQKKPGDLEGGTQVK YNTDNTFVFGGYAEY VLSPELSLRGKYDWN NSSLQVFRGGATYKL SPYLKVTFLGEYNTD NTARGAVVFEGEYPL	Yes	Negative maxima CD spectra at 215 nm	Mono-dispersed monomer	No	No match

	SPNLVLFPGAQYDWN NSSLNKGKPYLRVVY DLSPELKLVVEFWYN TDNTFKVFVGTKYKL SPNLELFVWYGWNNS SLNEFVADLRYQLTP					
TMB12_6	MEDKPGTAKGGGETK YNTDNTLDVGGYGEY VLTPELKGFTRYVWN NSSLQSLEVGATYKL SPYARVTVSVEYNTD NTAALKVTVEVFYPL SPNAFTVPGATYVWN NSSLQKGFPYTLFVY QVSPDLVLFVMVVYN TDNTLTLFFGGEYRL SPNLSTRTAYGWNNS SLNWVRTETRYDLSS	Yes	N.D.	Void (aggregate)	N.D.	No match
TMB12_7	METKPGSVWGGGEAK YNTDNTFEAGGYGEY HLTPNLTAFGQYMWN NSSLQKARGGLTYWL SPYARVTVFVEYNTD NTVTLFVVFEVFYPV TPELVLVPGAQYEWN NSSLNSLKPYLVVFY KLSPDLVGKVVFVYN TDNTFQVWFGAVYQL SPNLFVEVMYGWNNS SLNELKTRVEYLLSM	Yes	N.D.	Void (aggregate)	N.D.	No match
TMB12_8	MQDSKGSVRSGTEGE YNTDNTAKGGTYTTY VVTPDLEVTVKYEWN NSSLNRTEVGATYRV SPYARTTVSVKYNTD NTVKFDVTVEGDYPL SPNLKTHPGATYTWN NSSLNKLSPYTTVKY KLSPQLVGKVTFRYN TDNTVEVWFGAEYQL SPELVVEGAYGWNNS SLQKFKGRTEYQLTP	Yes	Negative maxima CD spectra at 215 nm	Mono-dispersed monomer	No	No match
TMB12_9	MQEKPGSLRAGTEFR YNTDNTFSLGFYAWY QLSPNLVGFAKYEWN NSSLNVFEAGAKYKL SPYLETVVSVVYNTD NTAKLKVRTESEYPL SPNLRLRPGADYEWN NSSLQKAFPYLTTEY	Yes	Negative maxima CD spectra at 215 nm	Mono-dispersed monomer, crystals diffracting at 8A.	**Yes**	No match

	KLSPDLVGKTTFVYN TDNTVKVVAGALYKL TPNLEVEFLYGWNNS SLNETVVRFDYVLTP					
TMB12_tri_0	MGSPGRTEVRGEGGY NTDNTFWGVVEFWYE ASPNLSPYVRYKWGK SSKNVGWPGGRYKAS PDAEVDVEVGYNTDN TVVFRVTLRFFYDAS PNAKPYGEFTYEWGK SSKNKSMVYGGAKYK ASPGAEVDFKGGYNT DNTFQGTVVFWYRLS PNAMGYGRYTWGKSS KNEFTGGGKYVGSE	Yes	N.D.	Major monomer	No	No match
TMB12_tri_4	MGSSQSPGRTETRTE GGYNTDNTVKAVTVT FYDLSPNAQPYVVYE WGKSSKNKGEPGGQY RLSPDAWVDVRVGYN TDNTVEFLVVFQFWY DASPDLKPYGEFWYK WGKSSKNKGAVYGGA KYRLSPNAEVEFKGG YNTDNTLEASVVFWY RLSPGAEGYGKYDWG KSSKNAFFGGGKYVG SE	Yes	N.D.	Mostly void (aggregate) but has monomeric peak	**Yes**	No match
TMB12_tri_12	MGSQGSPGRTEVTFE GGYNTDNTAWAATRF KYQASPNAEPYVEYK WGKSSKNVGWPGGKY KLSPDAETDTRTGYN TDNTLRVETETVFWY DASPDAKPYVRFRYV WGKSSKNKGAVYFGG KYKLSPNAEVEVEGG YNTDNTLEGRVEFRY RGSPNLEGYVVYLWG KSSKNEFTGGGTYVG SD	Yes	Negative maxima CD spectra at 215 nm	Mono-dispersed monomer	**Yes**	No match
TMB12_tri_53	MGSPGEGLVRTMVGY NTDNTVEVKTEFWYV LSPDAFPYLRYKWGK SSKNESSPGGTYRLS PDAETTAEAGYNTDN TLKVVVSTTFFYDAS PNLKPYVTFTYKWGK SSKNKFEVYFGGEYK	Yes	N.D.	Mono-dispersed monomer	**Yes**	No match

	LSPNARSVTEVGYNT DNTAWAETKTEYQGS PNVKGVVEYFWGKSS KNQFWFGGEYRGSN					
TMB12_tri_28	MGSGDSPGVSEGRLK FGYNTDNTAWVEVEV WYWASPNLVPYVKYK WGKSSKNDFKPGGRY KLSPDAELDASGGYN TDNTLKADVQLDFWY DASPNLKPYVRFVYK WGKSSKNKFKVYFGG KYKLSPGAESEFEVG YNTDNTAEVRVRFKY RGSPNAWVVVQYDWG KSSKNDGESGVDYKG SE	Yes	N.D.	Monomeric and higher order oligomeric peaks	No	No match
TMB12_tri_14	MGSEDSPGRGETLVM TGYNTDNTAETRGET WYQLSPDFVPYLRYK WGKSSKNEFWPGGTY KLSPDAEVTFEVGYN TDNTLKFRVEFKVFY TATPDLHPYGEVKYE WGKSSKNKGAVYGGG KYRLSPNAEGDVQGG YNTDNTVELTFRFFY VGSPNLVGVVEYTWG KSSKNEFWFGGKYLG SD	Yes	N.D.	Mostly void (aggregate) but has monomeric peak	No	No match
TMB12_tri_36	MGSQPSPGQGESLVK VGYNTDNTADVTVEF WYVASPNLVPYAVYT WGKSSKNKFEPGAKY KLSPDAWVDVKVGYN TDNTAWAESVAEFWY KATPDLEPYVVFVYR WGKSSKNKFEVYFGG RYKLSPNAELDTVTG YNTDNTVVTTSQLDY WLSPNALLVVRYTWG KSSKNRVESGVKYKL SE	Yes	N.D.	Major oligomeric peak	No	No match
TMB12_tri_32	MGSGGSPGQTEVKTE TGYNTDNTVDTVVEF WYVVSPDLVPYFVYF WGKSSKNKFEPGGQY KLSPDAWVDTRGGYN TDNTLEGKTVVEFWY DASPDAKPYVQFHYE	Yes	N.D.	Major monomer	No	No match

	WGKSSKNKFKVYFGG RYRLSPNAELDAATG YNTDNTVVTVVQFDY QGSPGLVGYVRYVWG KSSKNKFVFGGKYRG SE					
TMB12_ova l_1	MGSESRQGSLGAYVR FVYNTDNTAEVGPGF EYEASPHLWVQVDVG WNNSSLVKFSVETSY KSPDGLEVKVGGEYN TDNTARGKVEVVFWY TLSPDLHPYGKVQYF WNNSSLNKGKPGGGF VYKLSPHAEFKFETG YNTDNTVEVHFWTDY KASPELEFSSGGVWN NSSLARTETRAKYKL TP	Weak	N.D.	N.D.	N.D.	0.003 (Signaling mucin HKR1 [*Mizuhopecten yessonsis*])
TMB12_ova l_2	MGSQDKPGTAGGYFR AKYNTDNTAEAGPGG VYVLTPDLKLFVEFG WNNSSLFKLKVFADY KSPDGVDFRSGTEYN TDNTAETQFEVLFWY SVSPEFNPYGKFEYR WNNSSLNKAKPGGGA EYTFSPDLKVWTETG YNTDNTAETTVVVKY RLSPDLEVQTGTTWN NSSLVEFVTEVWYKA SD	Yes	N.D.	Major monomer	**Yes**	No match
TMB12_ova l_3	MGSQPKPGDLGTYLE VEYNTDNTAKTGPGA VYVLSPHLAVETKSG WNNSSLLEVTASLKY ETPDGVVVEVGGTYN TDNTLDLFTDVEFWY TLSPHLHPYGRLRYE WNNSSLNKAQPGGGA LYVLSPHLRTRVDTG YNTDNTVKVKVETDY VVSPHVQVRVGGEWN NSSLFKTVVQFWYKL TE	Yes	N.D.	Major monomer, minor oligomer	**Yes**	No match
TMB12_ova l_4	MGSQDSQGDLGGYLR VEYNTDNTAWVGPGG KYVLSPHAELDVAAG WNNSSLLWTEVAVVY WSPDGVQVKLGTRYN	Yes	Negative maxima CD spectra at 215 nm	Major monomer	**Yes**	No match

	TDNTVSTKLEVLFWY LLSPHLRPYGKTEYQ WNNSSLNKTRPGGGF QYDLSPHLATQFEAG YNTDNTAVAKVKTVY QLSPDAKVEGGTQWN NSSLVKFESQVDYKL SS					
TMB12_ova l_5	MGSQDSQGDAGGYFR VEYNTDNTFWAGPGG KYALSPHLFLDVAAG WNNSSLLKFDFHVKY FSPDGLEVQVGFTYN TDNTLVFFVVVVFWY DLSPHVKPYGRLEYR WNNSSLNKFFPGGGA RYKLSPDLEVQADSG YNTDNTASTRFETVY KVSPEAEFKAGGQWN NSSLFKFQSQFDYKA SS	No	N.D.	N.D.	N.D.	No match
TMB12_ova l_6	MGSQSNPGTLGAYFR AVYNTDNTFEAGPGF EYELTPDLHTQVDVG WNNSSLVKFTVETTY KTPDGVEFTFGGTYN TDNTVVLLVVVVFWY DLTPDLKPYGKTEYW WNNSSLNKVRPGGGF VYRLSPHLATRVDVG YNTDNTLFVKVSTKY DLTPHAEFEAGTEWN NSSLVRFEAVLKYKL SE	No	N.D.	N.D.	N.D.	No match
TMB12_ova l_7	MGSEDKSGSAGSYLR VQYNTDNTVKLEPGS EYTLSPHLDTRVAVG WNNSSLASLKVSTVY RTPDGVEVEAGGEYN TDNTVDGVLRVQTQY QASPHLFPYTRTEYR WNNSSLNRSKPGGGF FYSVSPHLQFEVFVG YNTDNTFVVEVFTRY RVSPDVETKVGAAWN NSSLLSLVAETKYKL SE	Yes	N.D.	Major oligomer	No	No match
TMB12_rec t_1	MGSQDKPGSAGGYAF VGYNNSSLIEFQAGA QYVITPHLKTDVRVG	Weak	N.D.	N.D.	N.D.	No match

	YNTDNTIKVEVKVKY KSPDGIEVEVKGEWN NSSLAATEVWAGAEY SLTPELHPYARAGYR WNNSSLNKPKGEAGA KYVLSPHARADVKTG YNTDNTIETEVTVQY QISPHAGGQVTVKWN NSSLVEVWVGGRYQL SE					
TMB12_rect_2	MGSQEKPGSAGGYAQ VGYNNSSLIKFEAGA QYVITPHLKVDLRVG YNTDNTIEVSLRTSY STPDGIDVSVEVQWN NSSLAAVEVQTGVWY SLTPHIHPYASAGYR WNNSSLNKPFAEAGL LYKLSPHAEAELRTG YNTDNTIRVELTVRY QLSPHLGTWAKTRWN NSSLVETWAGAKYQL SE	No	N.D.	N.D.	N.D.	No match
TMB12_rect_3	MGSQDKPGSLGAYAQ VGYNNSSLIEFQAGL EYVITPHLKTWLDAG YNTDNTAKGAVVVEY KTPDGIETRVRLEWN NSSLFKLDVKTGVMY SLTPHIHPYAAVGYT WNNSSLNKPKVEAGT KYVISPHLQVDVKVG YNTDNTIEFQVDVWY QVSPDAGGWASVKWN NSSLLEVWAGGKYQL SE	Weak	N.D.	N.D.	N.D.	No match
TMB12_rect_4	MGSSEQPGSAGAYAM TGYNNSSLIQTFVGG EYVITPHLKTRVETG YNTDNTARTEVETEY FSPDGAVAKVKAVWN NSSLAELLVVGGAQY FVSPDFFPYAMGGYK WNNSSLNKPVGVAGA KYRLSPHAWLKVEGG YNTDNTAELKVRAEY QISPEIGLLAQVDWN NSSLVEVKAGAKYKL SY	Yes	N.D.	Major monomer, minor oligomer	**Yes**	0.038 (multicopper oxidase domain-containing protein [*Verrucomicrobiaceae bacterium*])
TMB12_rec	MGSSEVPGSAGGYSK	Yes	N.D.	Mostly	No	No match

t_5	TGYNNSSLARTETGA DYVLSPHLVVQVAVG YNTDNTAELVVRVKY ETPDGIEFEVEVVWN NSSLIKFRAMTGAWY SLSPHLFPYAKAGYV WNNSSLNRPQAFAGA QYLLSPDLWLDVFGG YNTDNTIKLFVRVQY WLSPDLGTEAEVVWN NSSLVWVFAGAKYKI SD			void (aggregate), has monomeric peak		
TMB12_rect_6	MGSQPKPGSAGGYAA VGYNNSSLAKVEVGA EYQISPHAHVRLSTG YNTDNTAETVLVVVY ESPDGLRVKVVLRWN NSSLVQVEVMAGGLY VITPHLKPYSMAGYV WNNSSLNKPKGDGGA DYDISPDAQTKVSVG YNTDNTVELVVRVEY RISPDAGGFVEVKWN NSSLVEVKVGGKYKI SW	No	N.D.	N.D.	N.D.	No match
TMB12_rect_7	MGSSSRPGSIGLYAE TGYNNSSLAKTDAGG VYQISPHLKTWFAVG YNTDNTFSGRVQTKY DDPDGFHVSVKTEWN NSSLVSFEVKAGAMY QISPHAQPYAATGYQ WNNSSLNSPFVEVGA RYWISPHAFVDVAVG YNTDNTFVLFVTVFY QVSPDAGGWATGKWN NSSLAETWAGGEYVI DP	Weak	N.D.	N.D.	N.D.	No match
TMB12_rect_8	MGSEKKPGSAGAYAR AGYNNSSLIVGDAGG KYVISPHLETEVRVG YNTDNTAKLVVKVVY ETPDGAFVEVEVVWN NSSLISVVVSAGAWY DLDPHIKPYAQAGYQ WNNSSLNKPFAWAGA RYFISPDLEVDTATG YNTDNTFKLKTAVKY RLSPEAGLRFETVWN NSSLVEVWAGTEYKL SE	Yes	Negative maxima CD spectra at 215 nm	Mostly void (aggregate), has monomeric peak	**Yes**	No match

TMB12_rect_9	MGSQKKPGTAGTYAETGYNNSSLAVTRGGTRYVISPHAELDVSGGYNTDNTAFATVTVDYWTPDGIEFKVTVTWNNSSLAKVVVEAGAWYDLSPDAKPYVKGGYEWNNSSLNKPFGKAGAEYKVSPDLKVQVELGYNTDNTFFLRVQVDYKLSPHLGLKVETNWNNSSLAETWVGAEYQLTP	Yes	N.D.	Mostly void (aggregate), has monomeric peak	No	No match
TMB14_1	MGSESEKDEDVGVEYKKYDNQETEKRVYAKQKKSDKNELGVDYKKKSDNKTEKSVRYKYRYDDKLEVKVEYKKDSDDKDDLKVEVEYKRSDNAKLKAQYDKDKSDKWKVDVGGDYRIKKSEKSSLTPGGKLDVDKSRKVKLEPYARYEYKMSERSKADVELKADSDDKKQVDLKYSYKDSDKSKVDVKASFQKDSKWDAGVEVRYKKS	No	N.D.	N.D.	N.D.	0.0005 (oligogalacturonate-specific porin KdgM family protein [*Kluyvera ascorbeta*])
TMB14_2	MGSKSEKSSSVGVEYSKKDDQDTEKRVYAKQKKDKDNELGVEYKKKSDDKTEKKVDYKYKYSKELEVKVDYKKDSDDKDELYVEVDYKRSSKAYLKAKYRKDDKDDWSVSVGGKYEIKKSEKSYLEPGGELTVDDKNKVDLRPDARYKYKESKDSEADVRLEAQDKDRKSVELEYRYRTDDKSDVRVKADFDKDSQYKAGVEVKYDDD	Yes	N.D.	Mostly void (aggregate), has monomeric peak	**Yes**	No match
TMB14_4	MGSKQEKDTKVGVRYQKKKDDDTELDLEAKQRKDDKLELGVRYKYKSDSSWSLEVDLKYTESDKLDVELSYEYKDDSKKSLDLKVDYKRSKDAKLSASAKAEDDSKWSVSVGGQYRIKQSEKSYLEPGGSVDVDS	No	N.D.	N.D.	N.D.	No match

	SNKVDLKPYVRYRYK RSEKSEAGVEFGKSD KDKKEVKFDYKYEQD SRSYLDLEYRKDKDD DSEWGFKYEYDDS					
TMB14_5	MGSDSSSKEDVSVKY SKDKRNKSRVSVEVK QEQDSSNSLKVRYDK EDDDKSSVELGYEQD DDDKNKLKADVRKDK DDDVELEARYEYKRD QYLYLEAQYKKKKDS KYSLNVGGYYKIPLA DKLEVEPGGDVKYDD DSKKKLSPYVRVKYQ KDSNAEAGAEYKADE DKKSKLKLKLRYKED KDADVELEYEKDSDD KSEVGVRVKYRKD	Yes	N.D.	Void (aggregate)	No	No match
TMB14_6	MGSSDKKDDELEVEY SKDDKSKKSLSVKLK YKQDKDLDLEVEYRK DSDDKNDLRLGLKYR EDKKLELEASVKKDK DNKVELKAELRYKRD SYLKLYASAKARDDS KSELEVGGDYKVPYS DKLYLEPGGSVKYYD DSKSELKPYVKAEYK TDKDAKAEADYEAKE DDKKELKVRLDYKED KDSEVSLKYRADKDN RKEVGVKYKYKKD	No	N.D.	N.D.	N.D.	No match
TMB14_7	MGSESDSKERLRVEY KQDSDNENEVSVELE QKQDSSLDLRVEYKQ KDDDRNSVKLGLDYK QDSRLKLYAEVEYYK DQSVELEAQLEYKED KDLELYAKYKYKDDS KWTLEVGGKYKEPKA KDLEVEPGGSVDYYK DSKSKLNPSVRLEYK EDDKAYADAYLQADD KSQWKLDVKLKYKLD KDAEVDLKYSADDKN KNSVGVELTYKKE	No	N.D.	N.D.	N.D.	No match
TMB14_8	MGSREKKDTKVSVKY SKDDKDDSETSVRLE YKEDDKLALGVQYKK	Yes	Negative maxima CD spectra	Major monomer	**Yes**	0.012 (unnamed protein

	KSDDSSSTKVDLRYD SDDKLKLEAAVEKKD DSRTETEAKLEYKED SYLKLYAKYKKDSDD KYSLEVGGKYRVPKA KDLKVEPGGSVEYYK DSKYKVKPSVRLEYK TDKDARAGVYLEADS DKSWKLTADLEYKLD YNSRLRLEYEADDKN KSSVGVELEYKYD		at 215 nm			product [*Psylliodes chrysocephala*])
TMB14_9	MGSKEKKKTKVEVSY YKDDDDQSRTSVRLE YDLDKDLKLGVEYTK KDDDSSSTQVRLEYK QDKKLDLEASVQKKK DSDTRTSAELKYKRD SYLDLYAKYEKKSDD RYRLEVGGKYSVPLA KDLKVDPGGSVEYYK DSKKKVSPSVKLRYR KDDKADAGVDLRAYE DSSWKLKAYLKYKED KKSSLELEYEADSKN KSQVGVKLRYEED	No	N.D.	N.D.	N.D.	No match

*Protein BLAST search was performed against the non-redundant (nr) protein sequences database. The lowest E-value corresponding to a naturally-occurring protein is given. Many designs were matched with high confidence to previously-published *de novo* designed 8-stranded TMBs with < 50 % sequence identity.

Supplementary Figures

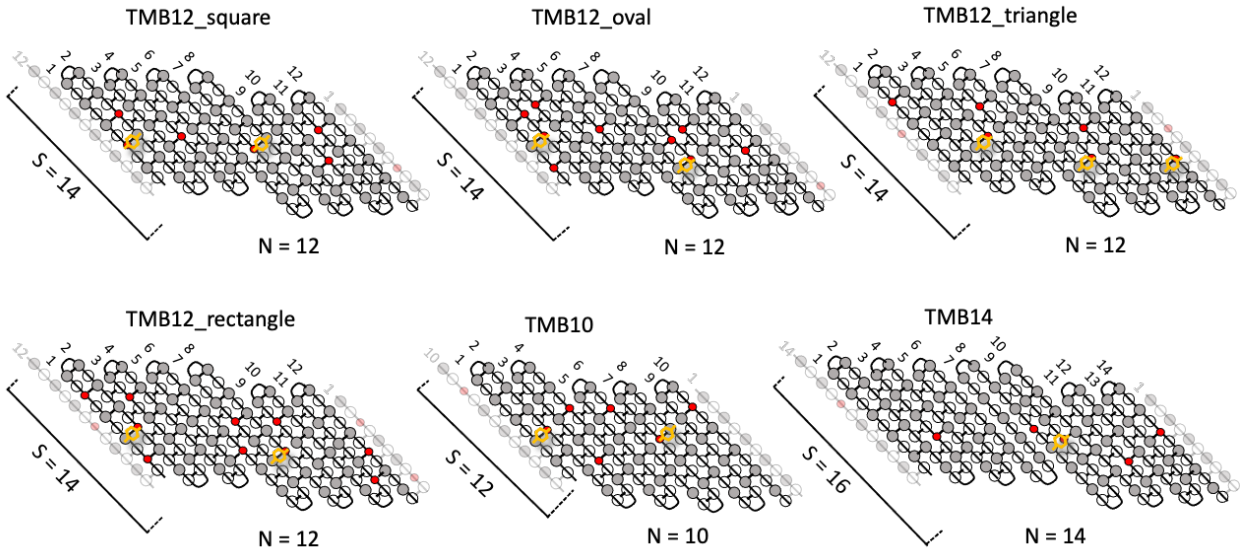

Figure S1: Six new β-barrel blueprints were generated here: one β-barrel of 10 strands (N=10) and a shear number of 12 (S=12), one β-barrel of 14 strands (N=14) and a shear number of 16 (S=16) and three β-barrels of 12 strands (N=12) and a shear number of 14 (S=14). The residues facing the β-barrel lumen and surface are shown as gray and white circles, respectively. Glycine kinks are shown as red circles and are facing the lumen. The tyrosine residues belonging to the Tyr-Gly-Asp/Glu folding motif are shown in orange.

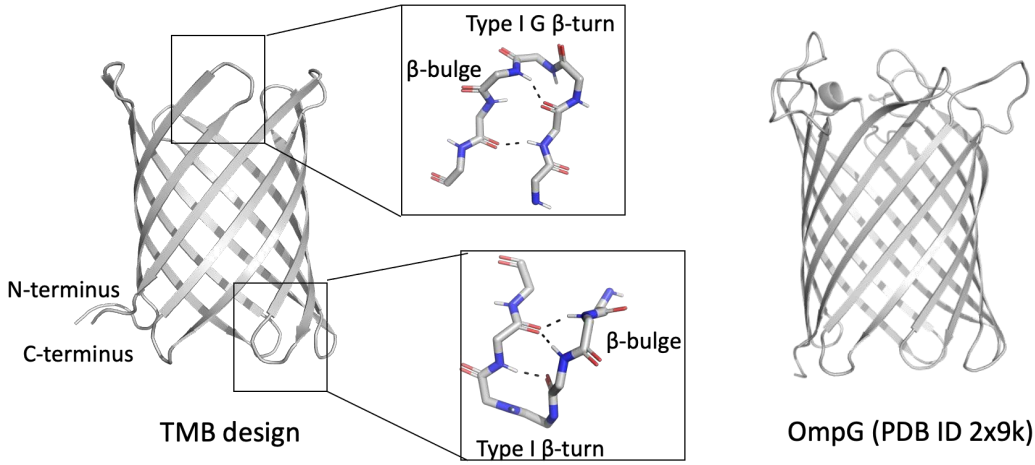

Figure S2: The β-strands of *de novo* designed TMBs are connected with short β-turns on both sides of the barrel: *cis*-hairpins (N- and C-termini side) are connected with canonical type I β-turns preceded by a β-bulge; *trans*-hairpins are connected with type I β-turns directly followed by a G-bulge. By comparison, naturally-occurring TMBs (exemplified here by OmpG, right) feature mostly long, disordered, loops on the *trans* side.

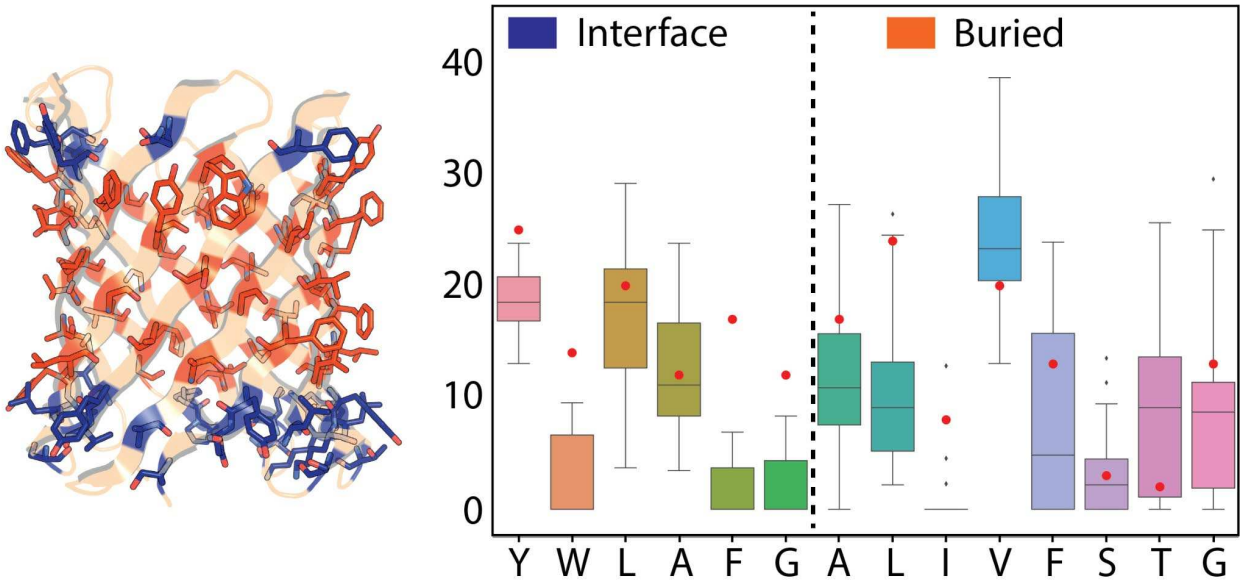

Figure S3: Amino acid composition of the membrane exposed surface of designed beta-barrel pores. Y-axis is calculated composition among all types of amino acids in the interface and buried region for each design respectively. The red dots are averaged amino acid compositions for the indicated amino acids in the respective regions over all transmembrane beta-barrel proteins in the OPM (Outer Membrane Protein) database.

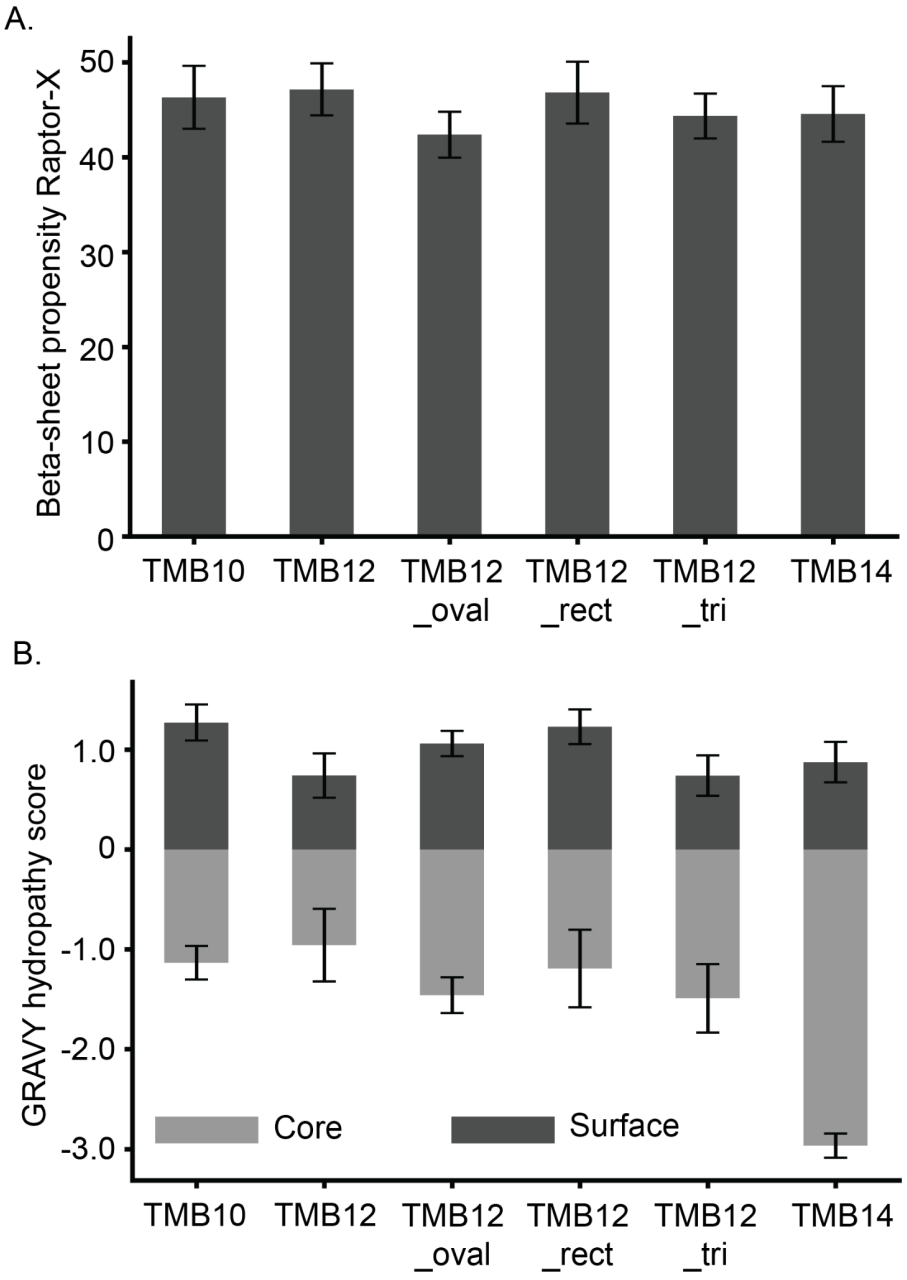

Figure S4: A. Predictions of beta-sheet propensities using RaptorX for different designs. B. GRAVY hydropathy values for the different types of designs and their differences between the pore lining core and surface exposed residues.

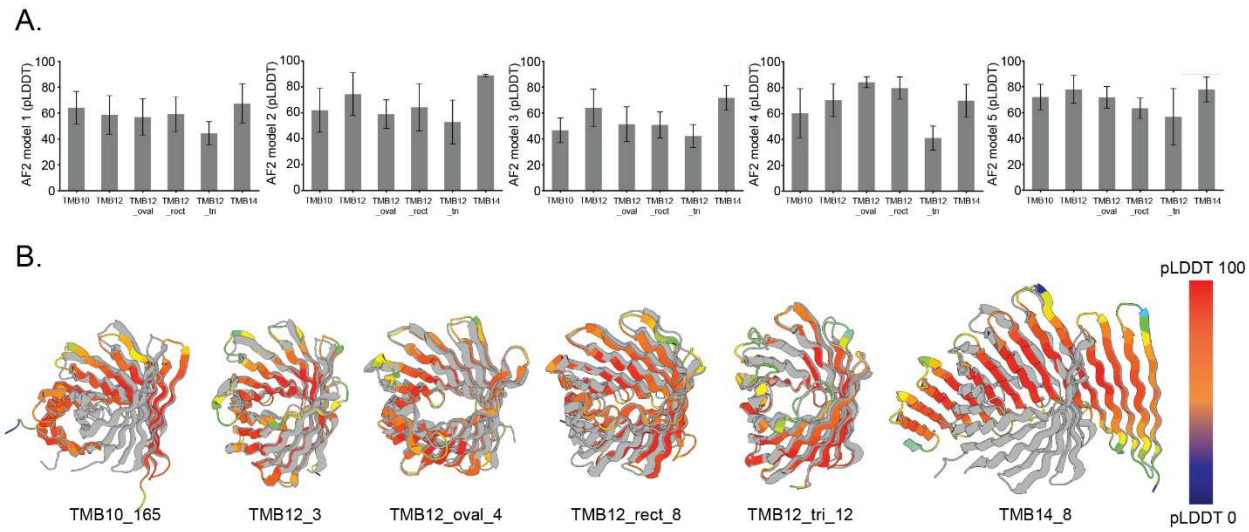

Figure S5: A. Mean pLDDT values for the different types of designs predicted from single sequence without MSA using Alphafold2 and the 5 models. B. Best predicted structures from Alphafold2 (shown as spectrum) are aligned to the design models (shown in gray) for each type of pore. The Alphafold2 structures are colored by per residue pLDDT values.

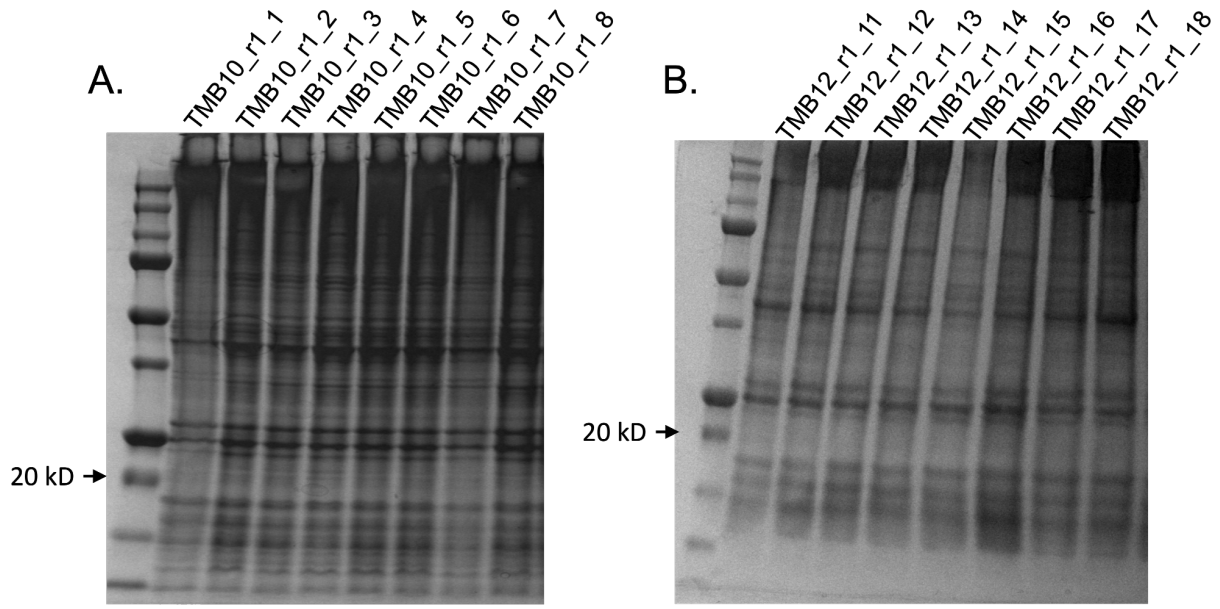

Figure S6: Coomassie stained SDS-PAGE gels showing the first 16 “optimal” 10- and 12-stranded TMB designs which failed to express. The expected molecular weights of the designed proteins were around 22 kDa for TMB12 and 18 kDa for TMB10 designs.

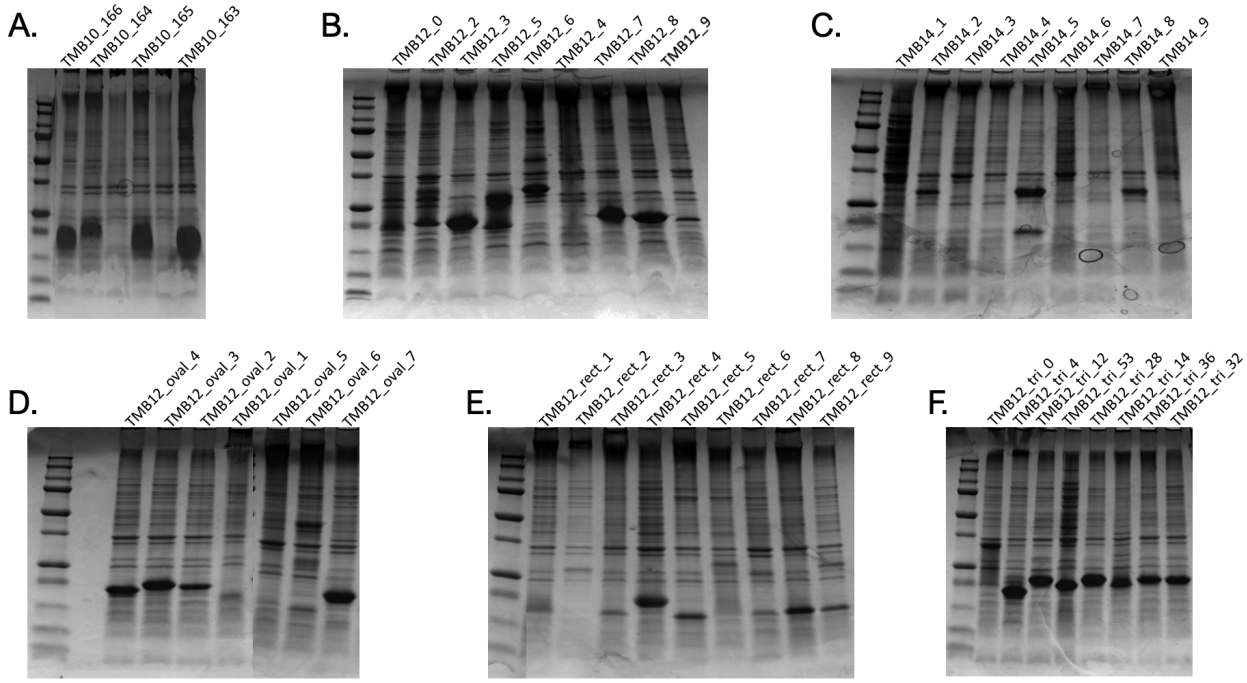

Figure S7: Coomassie stained SDS-PAGE gels showing expression bands for the different designs from insoluble fractions of corresponding lysed cell pellets.

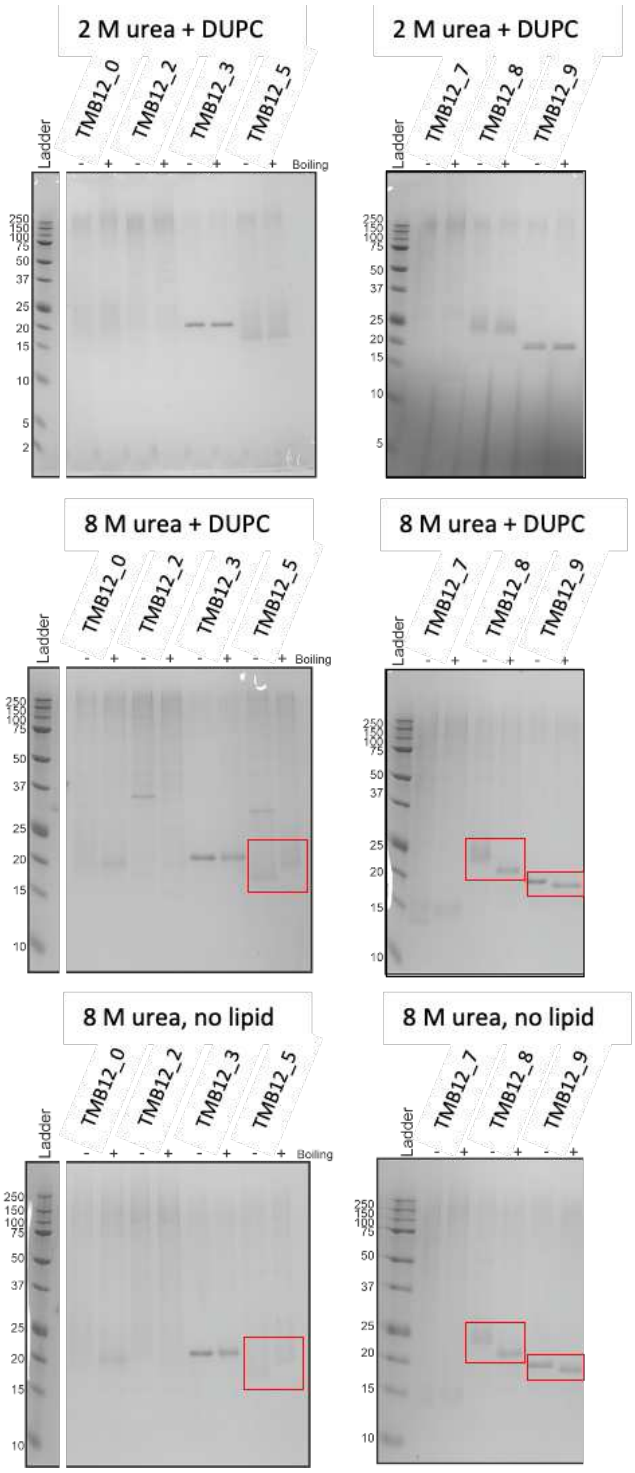

Figure S8: De novo designed TMBs do not exhibit the heat-modifiable behavior on cold SDS-PAGE gel characteristic of folded natural TMBs. No shift in band positions were observed after boiling the samples refolded in DUPC large unilamellar vesicles (LUVs). Band shifts after boiling (red rectangles) were observed only in conditions where the native β-barrel fold can not form (8 M urea with or without lipids).

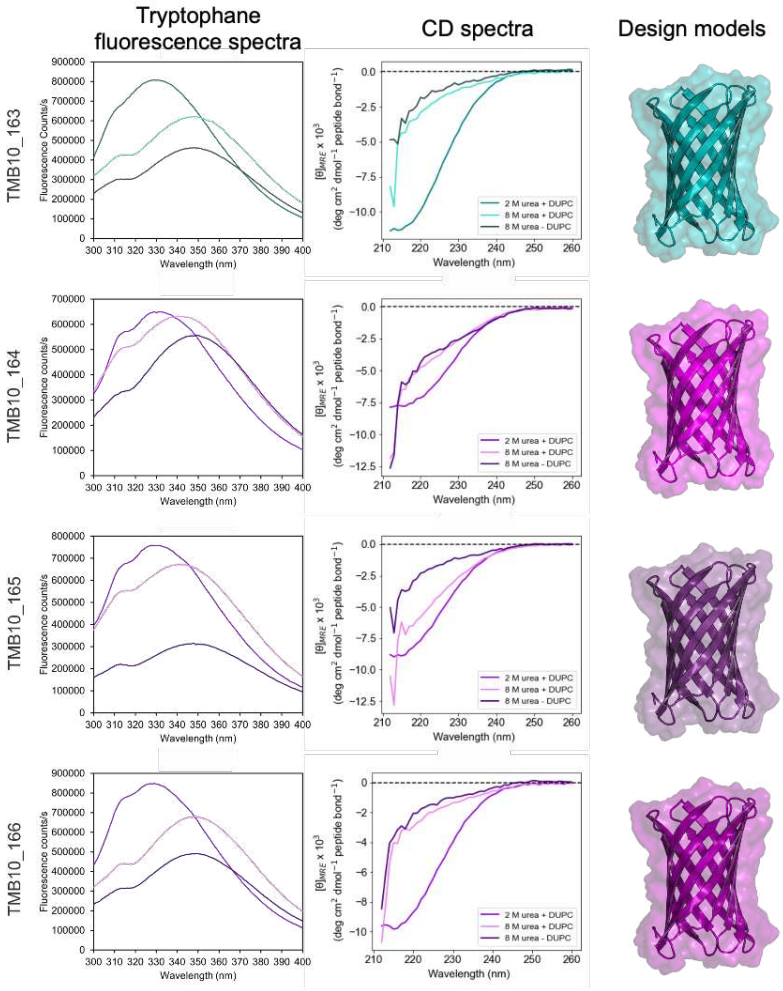

Figure S9: Biophysical characterisation of TMB10 designs (right: design models) for folding in DUPC LUVs. Tryptophan fluorescence spectra (left) and far-UV CD spectra (center) are shown for re-folded proteins in the presence of LUVs and 2M urea (allows TMB folding while reducing aggregation in water), in 8 M urea in the presence of LUVs and in 8 M urea in the absence of lipids. TMB10_163 was selected for further characterization (teal), as it demonstrated a clear fluorescence λ_{max} shift and change in β-sheet structure content between 2 M urea, 8 M urea, and no lipid conditions.

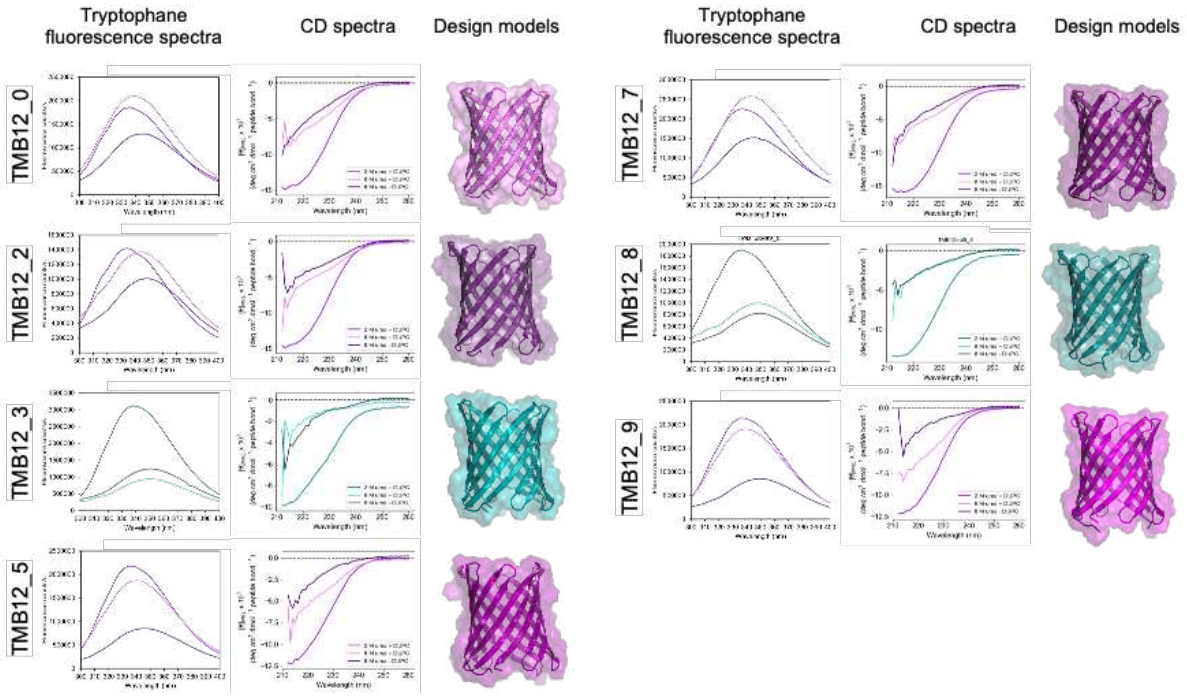

Figure S10: Biophysical characterisation of TMB12 designs with a square shape (right: design models) for folding in DUPC LUVs. Tryptophan fluorescence spectra (left) and far-UV CD spectra (center) are shown for re-folded proteins in the presence of LUVs and 2M urea (allows TMB folding while reducing aggregation in water), in 8 M urea in the presence of LUVs and in 8 M urea in the absence of lipids. TMB12_3 was selected for further characterization (teal), as it demonstrated a clear fluorescence λ_{max} shift and change in β-sheet structure content between 2 M urea, 8 M urea, and no lipid conditions. Similar spectra are observed for the TMB12_9 design, suggesting that the design is also folding into a TMB structure.

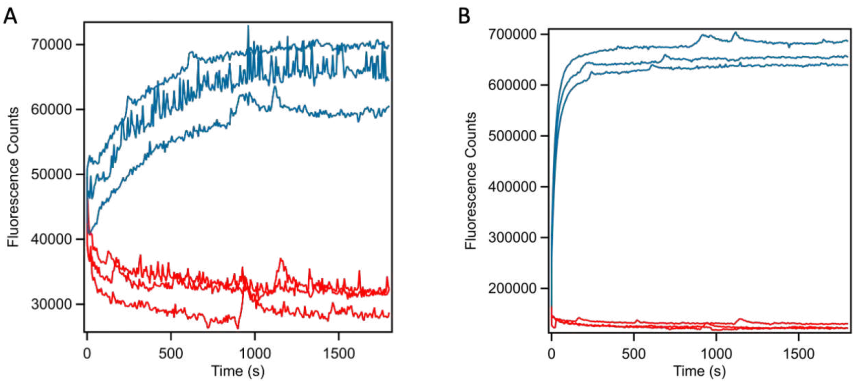

Figure S11: Folding kinetics of designs TMB10_163 (A) and TMB12_3 (B) in DUPC (blue lines) and DMPC (red lines) LUVs at 30°C and monitored by intrinsic tryptophan fluorescence. The difference in folding rates associated with the length of the lipid chain is consistent with intramembrane folding.

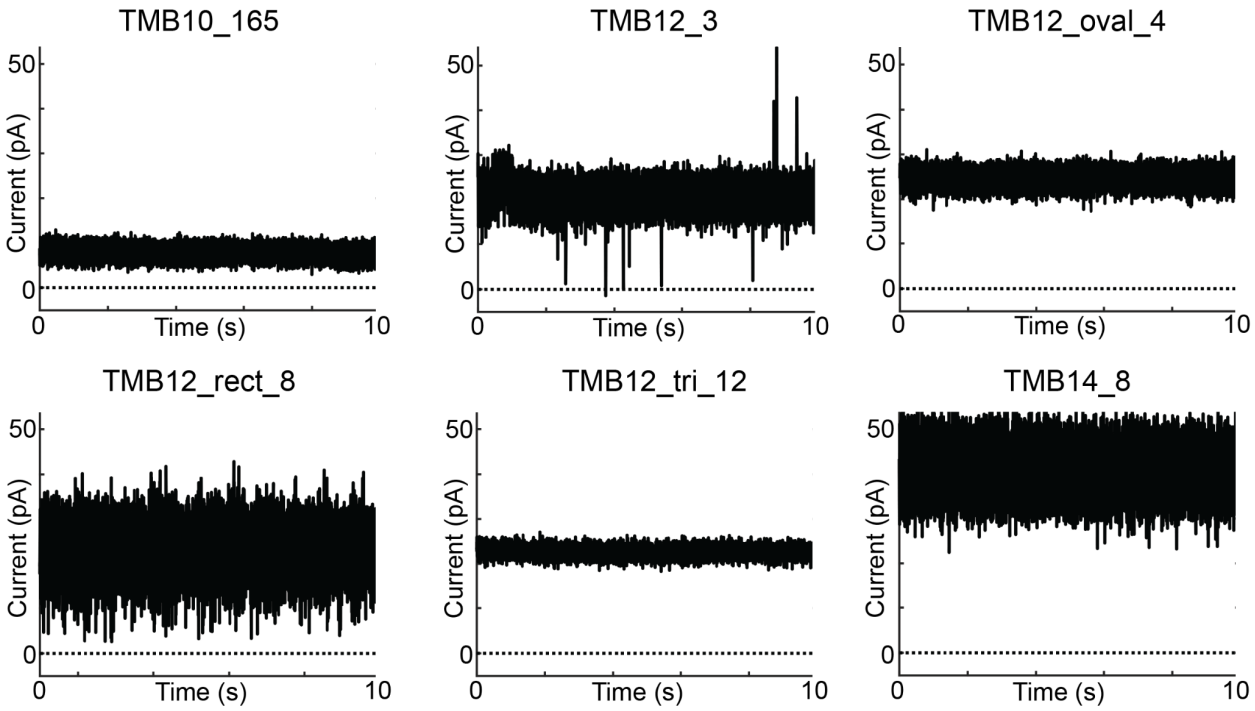

Figure S12: Raw unfiltered current traces of one example for each different type of pores recorded at 5kHz sampling rate. Applied voltage is 100 mV and the cis and trans buffer for all conditions is 500mM NaCl. Different noise levels are a result of the different bilayer capacitances at the time of recording and noise from adjacent cavities in the MECA recording chip from Nanion. 10s reads show characteristics of stable non-gating pores in the membrane.

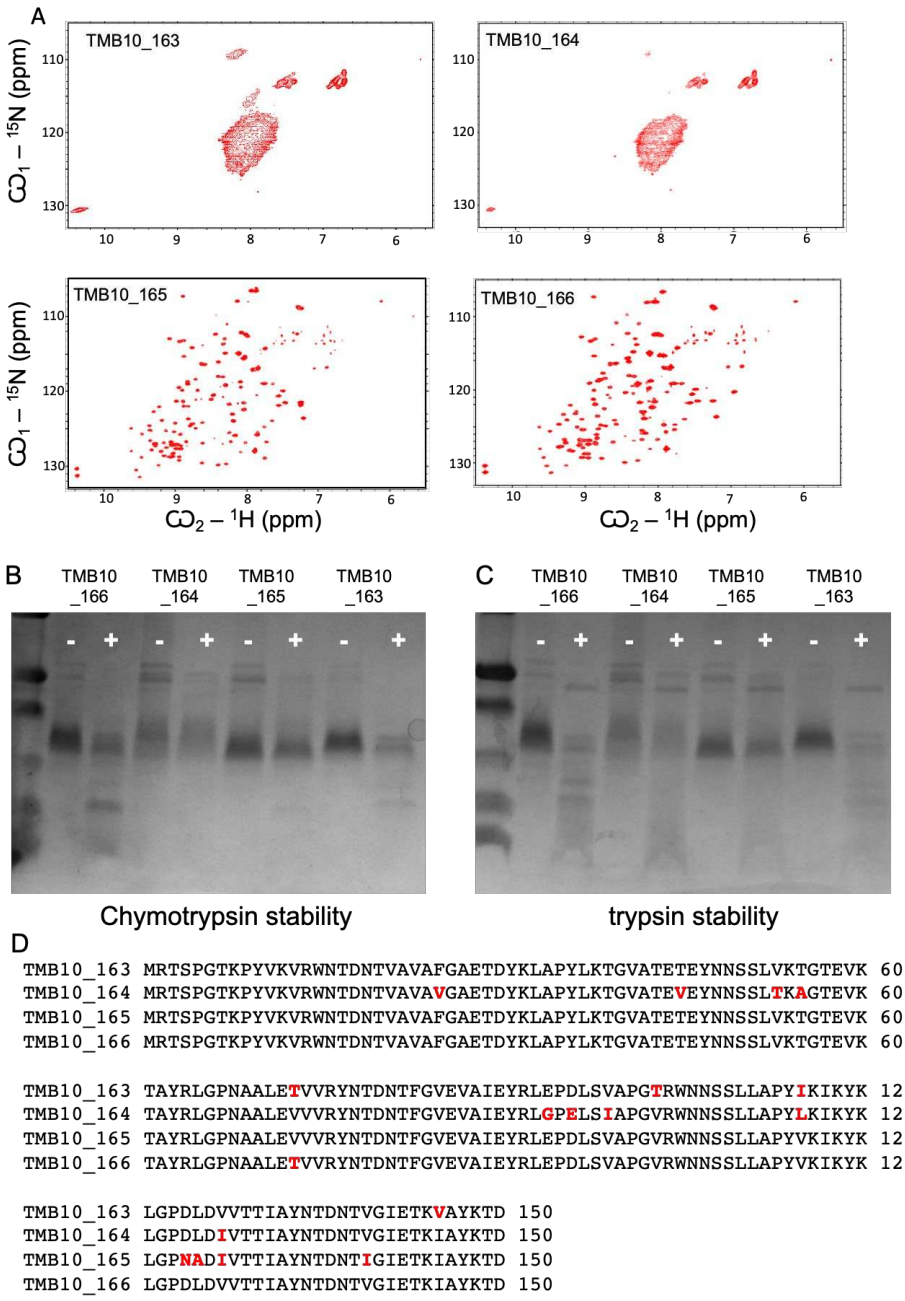

Figure S13: Relative stability of TMB10 designs (163-166). (A) Designs TMB10_165 and TMB10_166 feature well dispersed NMR ^{1}H-^{15}N HSQC spectra. (B-C) Trypsin and chymotrypsin challenge reveals differences in stability between designs, with TMB10_165 showing the highest stability to both proteases. (D) The designs differ only by 2-9 residues on the lipid-exposed surface, introduced using Rosetta.

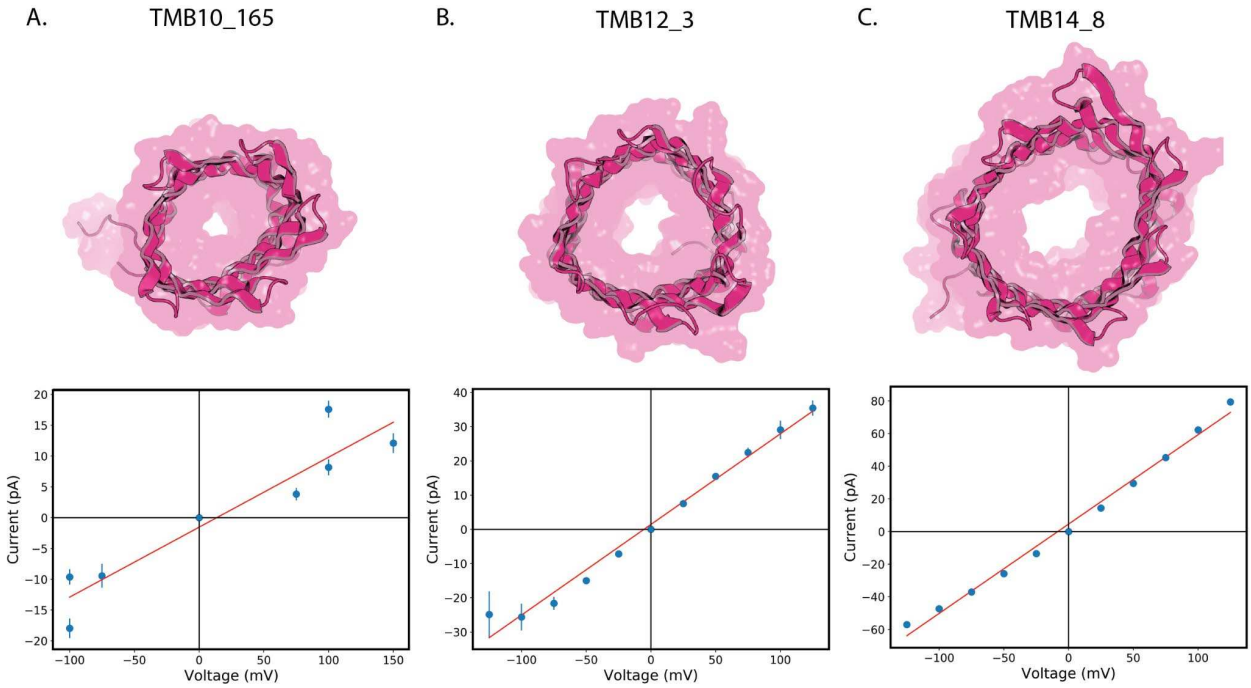

Figure S14: Current vs Voltage plots for three designs pertaining to TMB10_165, TMB12_3 and TMB14_8 pores. All measurements were carried out in 500mM NaCl solution (symmetric across bilayer).

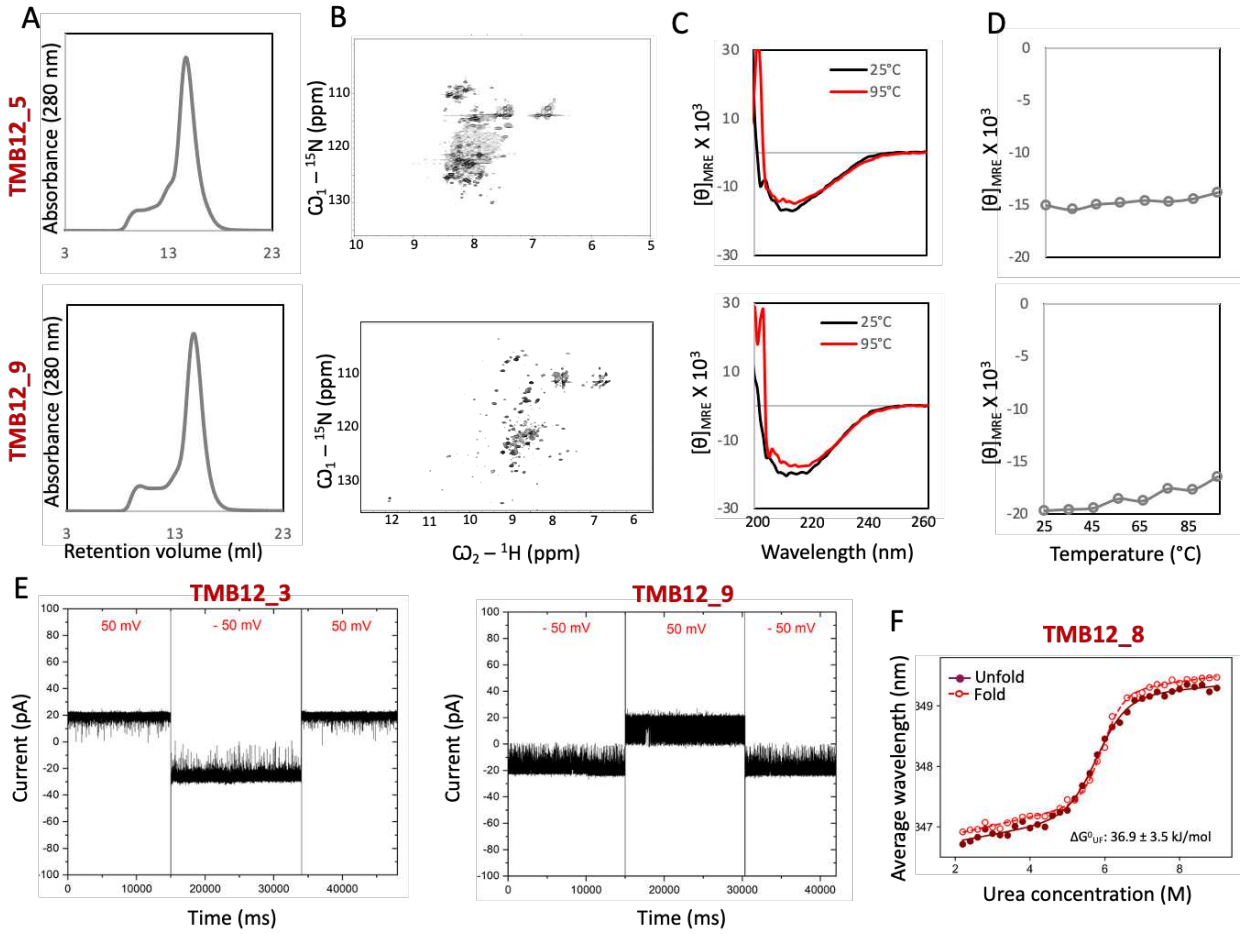

Figure S15: Designs TMB12_5 and TMB12_9 both feature monodispersed SEC elution profiles consistent with a monomeric TMB12 (A) and far-UV CD spectra characteristic of β-sheet proteins in DPC micelles (C) that remain stable up to 95°C (D). However, only TMB12_9 has a dispersed NMR ^{1}H-^{15}N HSQC spectrum indicative of a folded TMB (B). Stable nanopore activity was observed for designs TMB12_3 (stable signal) and TMB12_9 (gated signal suggesting lower stability), but not for TMB12_5 (E). TMB12_8 cooperatively and reversibly folds/unfolds in DUPC LUVs with a similar Cm^{F} to TMB12_3 (5.7 ± 0.3 M) but with a less sharp transition and hence lower unfolding free energy (F).

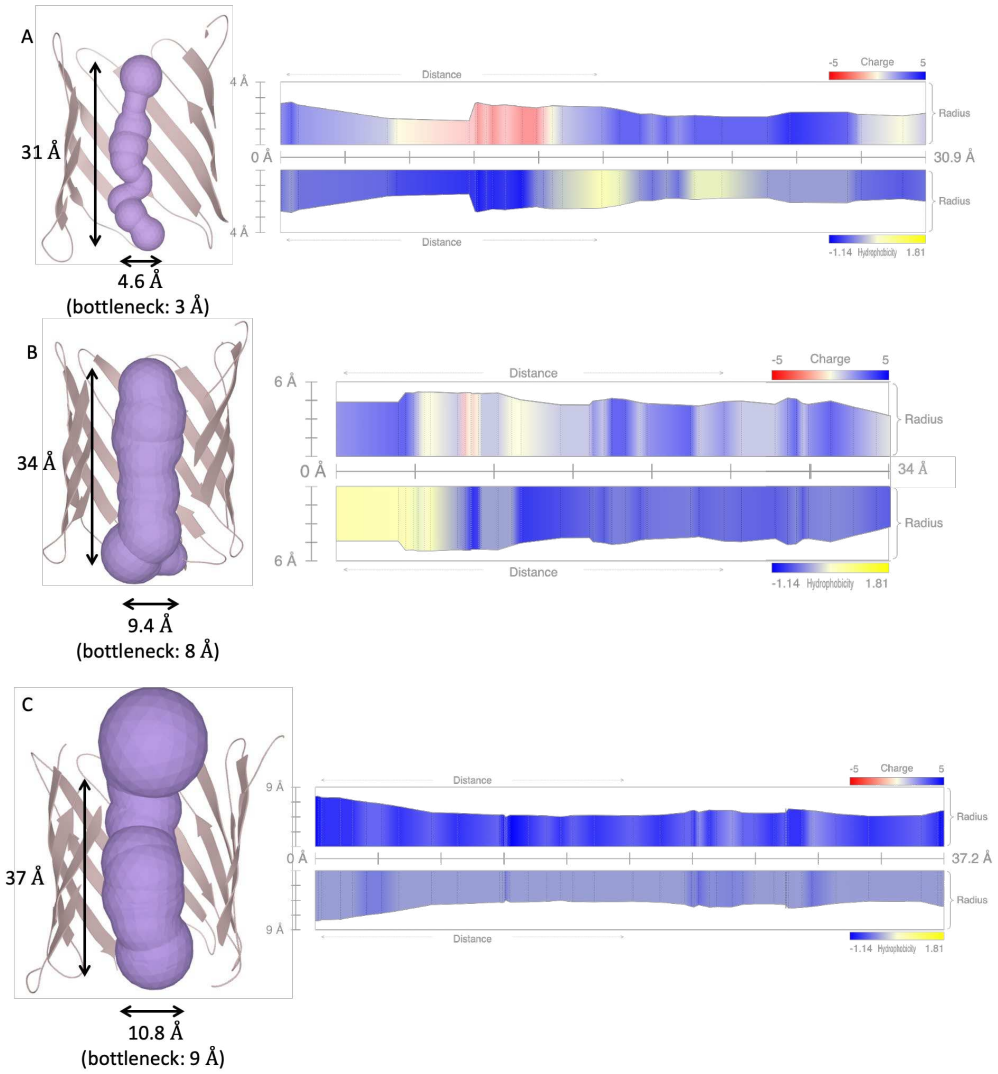

Figure S16: MOLE 2.5 pore size calculations (left), charge and hydrophobicity profiles (right) for designs TMB10_165 (A), TMB12_3 (B) and TMB14 (C).

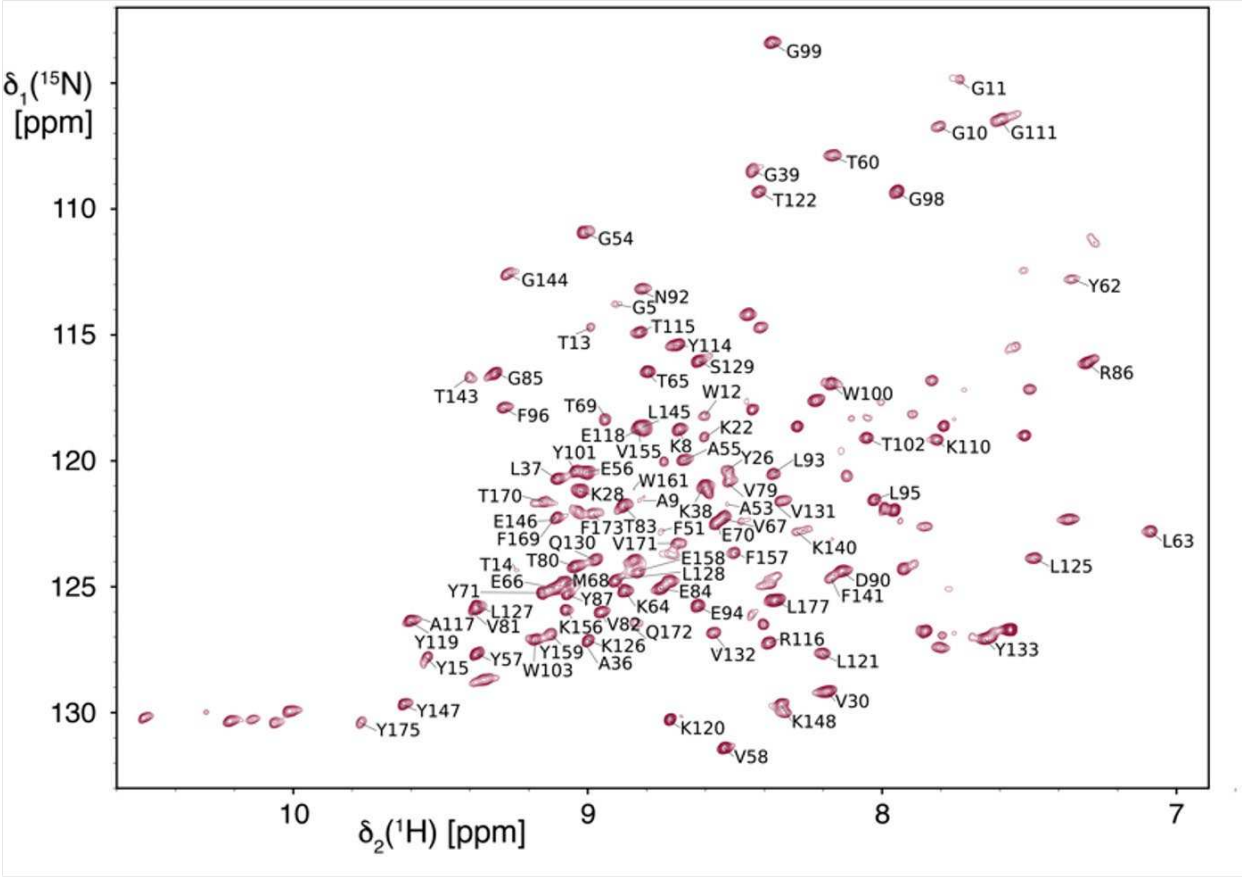

Figure S17: 2D [15N,1H]-TROSY NMR spectrum of [U-2H,15N]-TMB12_3 in LDAO micelles with sequence-specific resonance assignments.

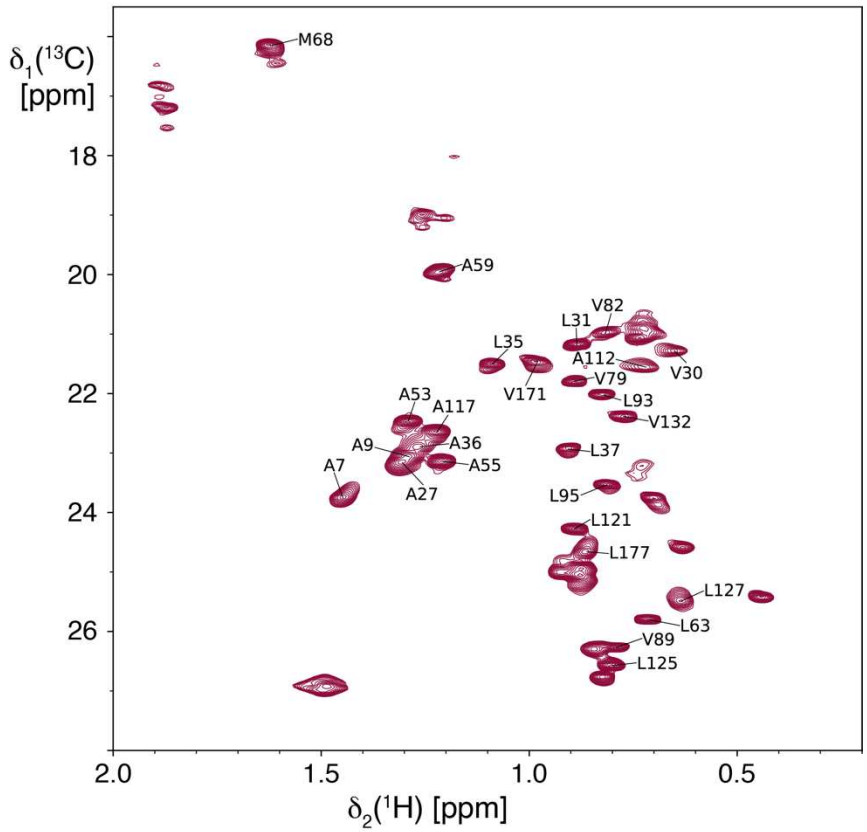

Figure S18: 2D [^{13}C,^{1}H]-HMQC of AMLV-^{1}H^{13}C-methyl-labelled TMB12_3 in LDAO. Sequence-specific resonance assignments are indicated.

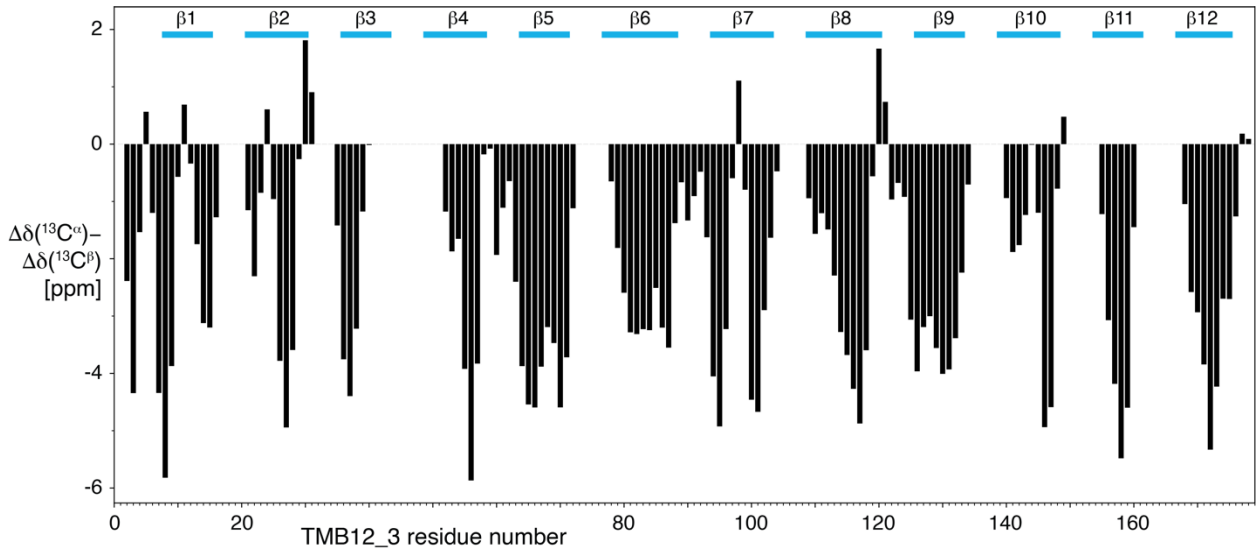

Figure S19: Secondary chemical shifts of TMB12_3 from sequence-specific resonance assignments. Consecutive stretches of large negative values indicate the presence of β-strand secondary structure. The positions of the 12 β-strands are indicated by blue lines.

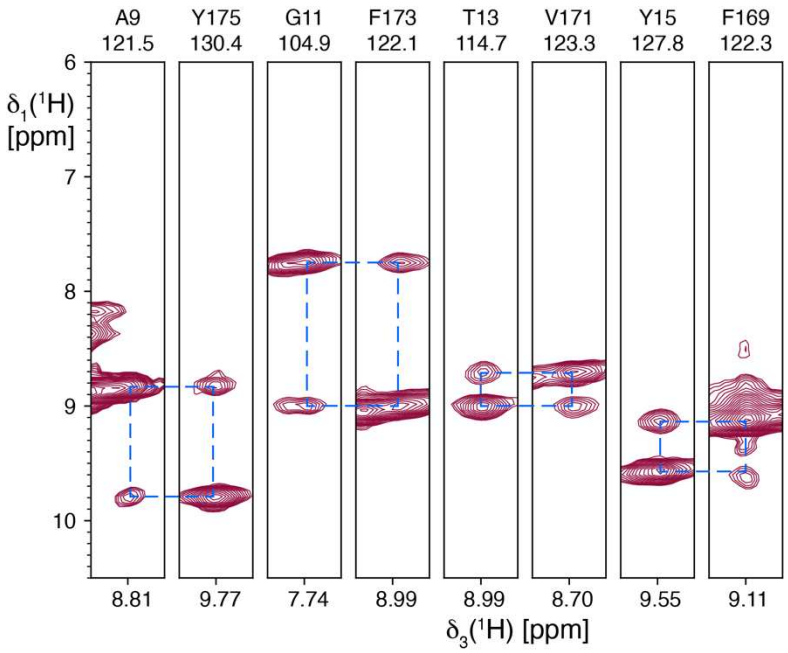

Figure S20: Strips from the 3D [^{1}H,^{1}H]-NOESY-^{15}N-TROSY experiment of TM12_3 in LDAO micelles. Strips were taken for the residue pairs involved in the antiparallel β1–β12 pairing. The NOE cross peaks are connected to the diagonal peaks by blue dashed lines.

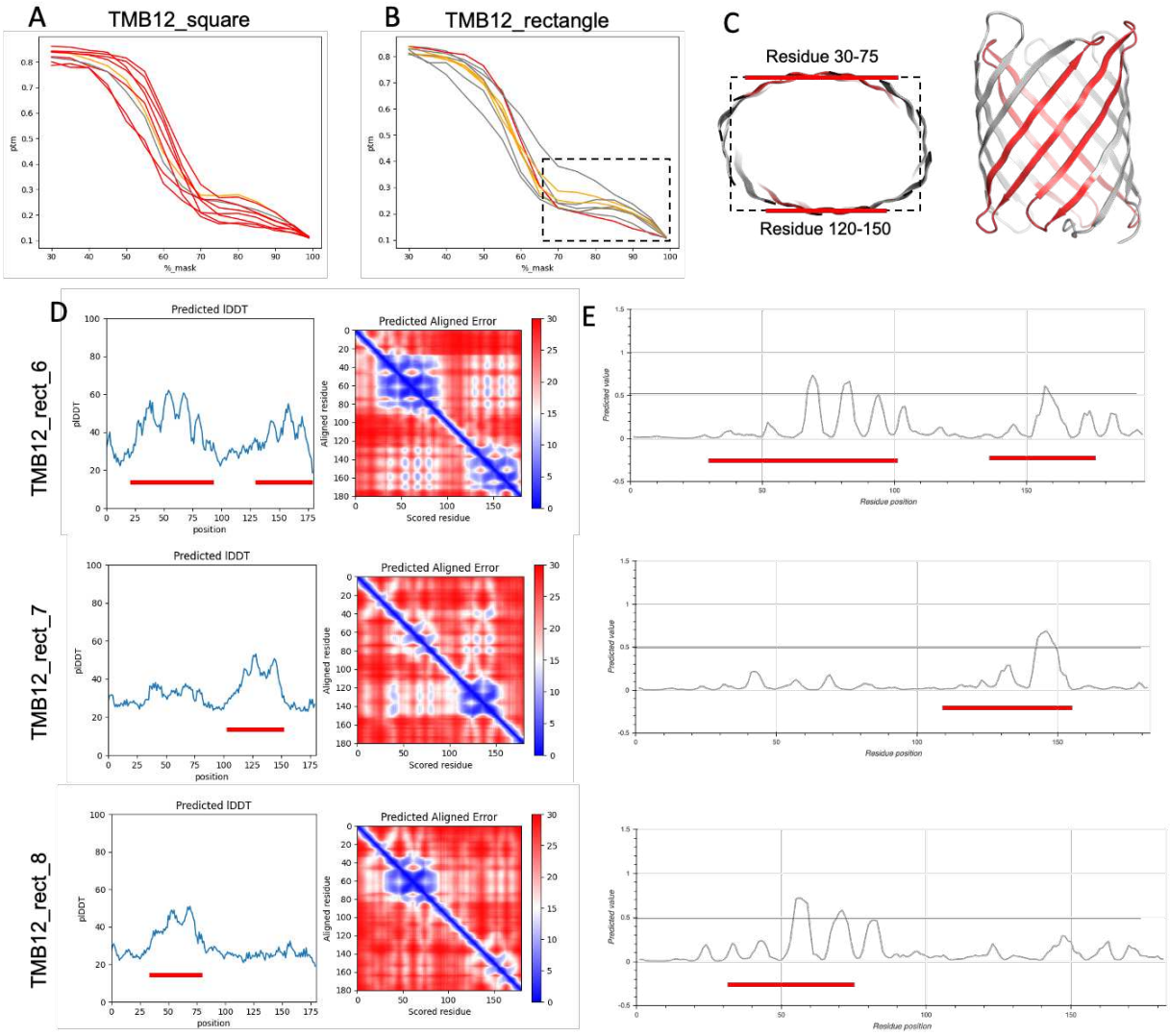

Figure S21: The square-shaped TMB12 designs (A) express at higher levels (red=strongly expressed; orange=weakly expressed; gray=no expression) than rectangle-shaped designs (B) and feature less residual secondary structure content at the end of ESM in silico-melting (*32*) simulations (highlighted by a dashed rectangle in B). Closer analysis of the ESM simulations show that the regions of melting-resistant secondary structure correspond to the long sides of the designed rectangular TMB12 structures (C, D) and co-localize with early-folding regions predicted with EFoldmine (*43*) (E).

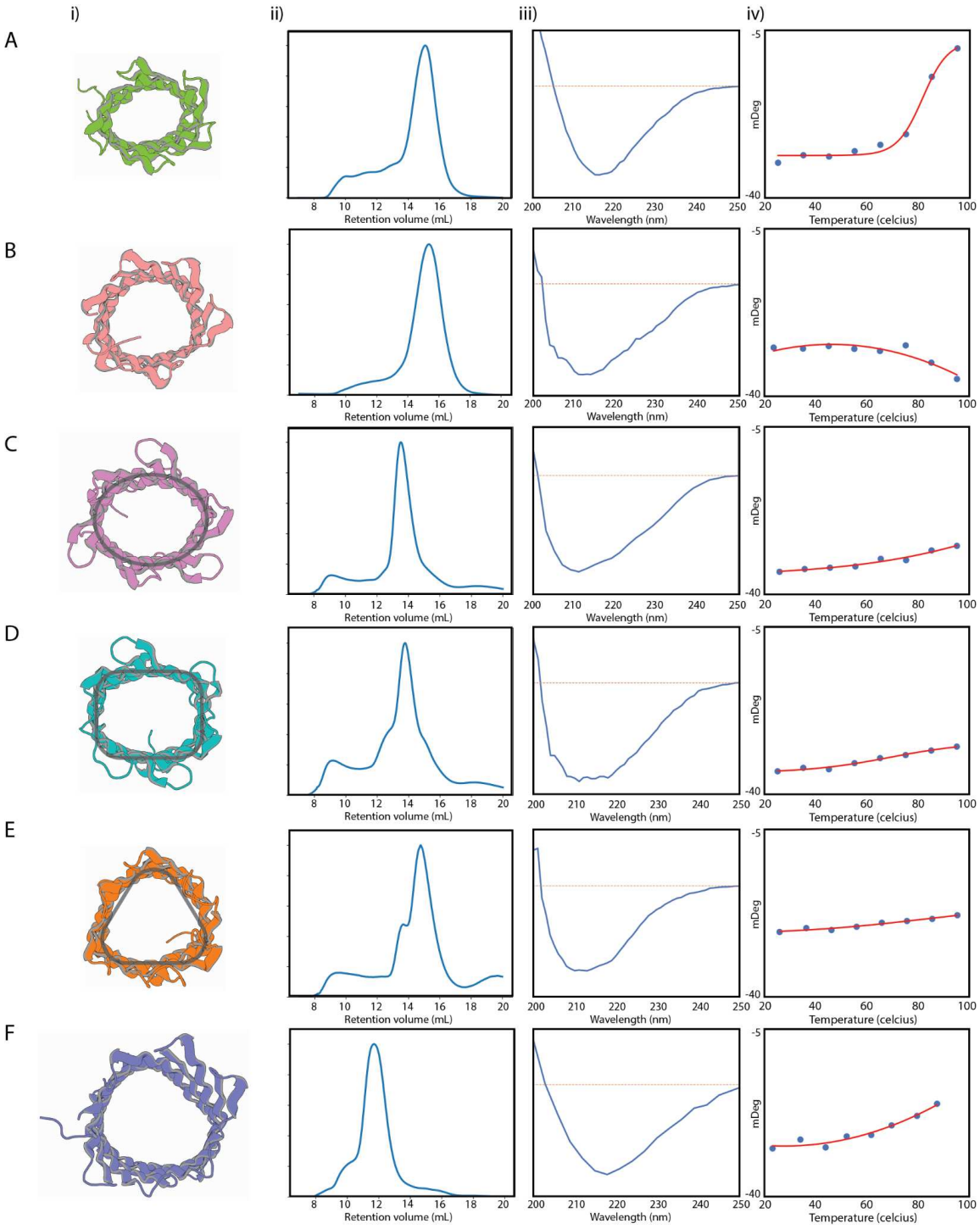

Figure S22: Characterisation of designed TMB pores. A. TMB10_165, B. TMB12_3, C. TMB12_oval_4, D. TMB12_rect_8, E. TMB12_tri_12, F. TMB14_8. i) Respective cartoons indicating top view of the designs. ii) Size Exclusion Chromatography plots for all designs carried out in a buffer containing 0.1% DPC detergent. iii) Corresponding Circular Dichroism (CD) plots in the near-UV range. iv) CD melt plots from 25 C to 95 C.

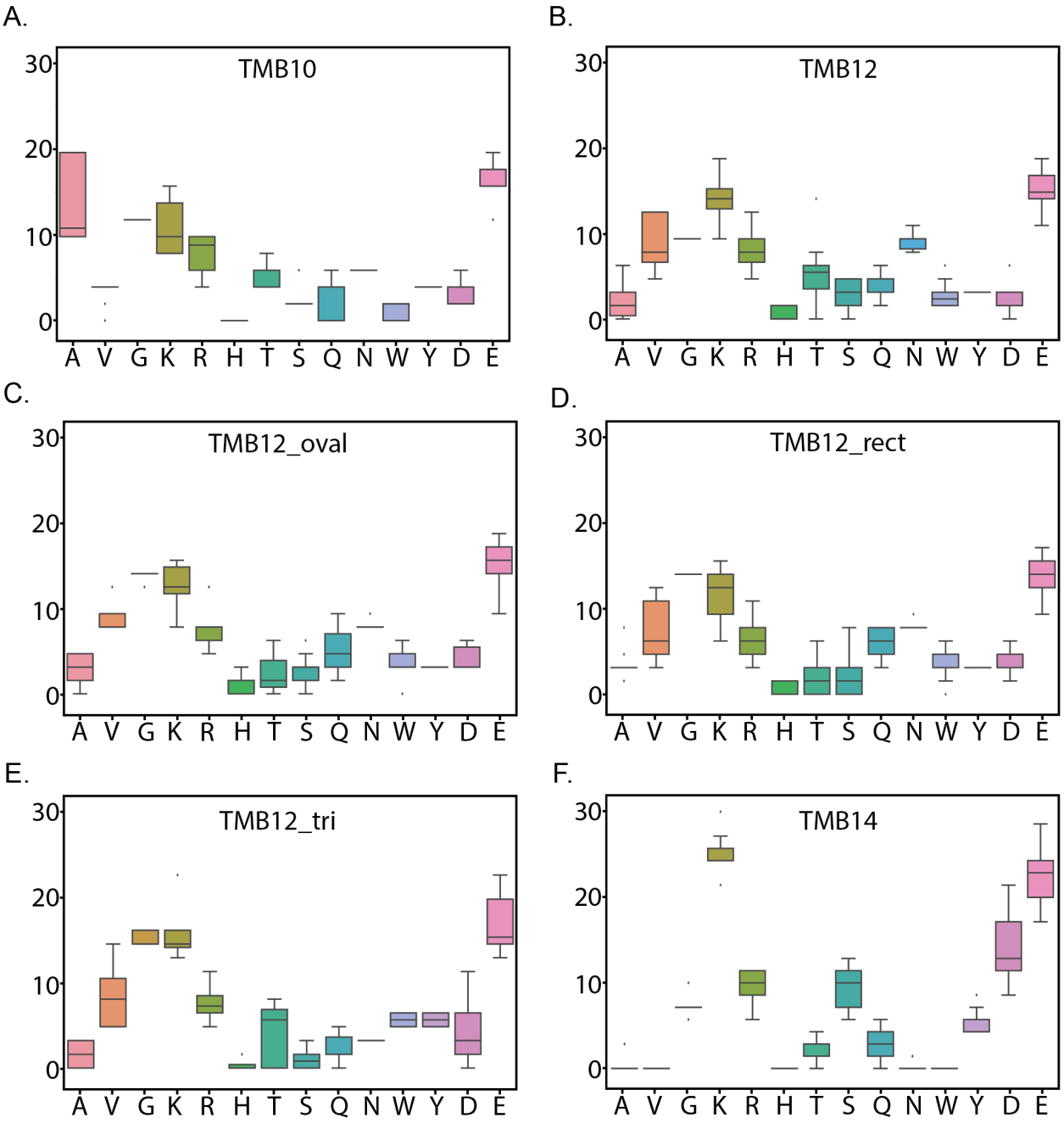

Figure S23: Pore lining amino acid compositions of the different types of designs. Y-axis indicates the percent fraction of the total number of a specific amino-acid within all pore-lining residues.

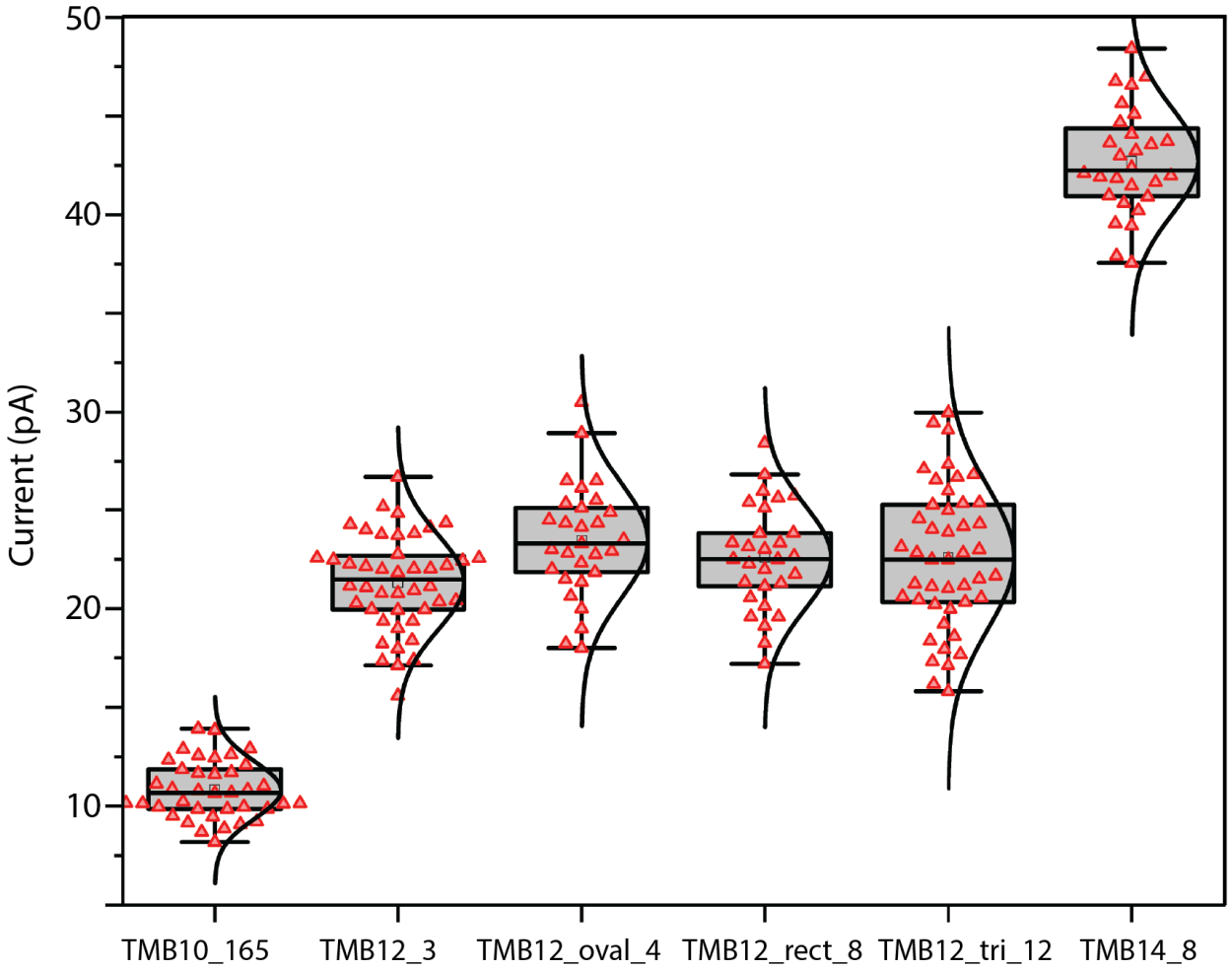

Figure S24: An all points box plot of the minimal step conductances observed across 3 independent recordings for each of the 6 designs shown in Figure 4. The voltage was kept constant at 100 mV throughout the recordings and across all current jumps. Both *cis* and *trans* solutions had 500 mM NaCl as the electrolyte.

References

61. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
62. 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).
63. 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).
64. 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. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
65. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
66. 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).
67. A. Favier, B. Brutscher, NMRlib: user-friendly pulse sequence tools for Bruker NMR spectrometers. *J. Biomol. NMR*. **73**, 199–211 (2019).
68. Y. Xia, K. Sze, G. Zhu, Transverse relaxation optimized 3D and 4D 15n/15N separated NOESY experiments of 15N labeled proteins. *J. Biomol. NMR*. **18**, 261–268 (2000).
69. V. Tugarinov, R. Sprangers, L. E. Kay, Line narrowing in methyl-TROSY using zero-quantum 1H-13C NMR spectroscopy. *J. Am. Chem. Soc.* **126**, 4921–4925 (2004).
70. P. Güntert, L. Buchner, Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR*. **62**, 453–471 (2015).
71. 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).
72. S. P. Skinner, R. H. Fogh, W. Boucher, T. J. Ragan, L. G. Mureddu, G. W. Vuister, CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J. Biomol. NMR*. **66**, 111–124 (2016).
73. Y. Shen, A. Bax, Protein Structural Information Derived from NMR Chemical Shift with the Neural Network Program TALOS-N. *Artificial Neural Networks*, 17–32 (2015).
74. Determination and analysis of urea and guanidine hydrochloride denaturation curves" in *Methods in Enzymology* (Academic Press, 1986), vol. 131, pp. 266–280.
75. A. A. Vorobieva, Rationale in Custom Design of Transmembrane β-Barrel Pores. *Methods Mol. Biol.* **2778**, 345–366 (2024).

76. 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).
77. Y. Zhang, J. Skolnick, TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).