

PROTEIN DESIGN

Accurate computational design of multipass transmembrane proteins

Peilong Lu,^{1,2} Duyoung Min,³ Frank DiMaio,^{1,2} Kathy Y. Wei,^{1,2*} Michael D. Vahey,⁴ Scott E. Boyken,^{1,2} Zibo Chen,^{1,2} Jorge A. Fallas,^{1,2} George Ueda,^{1,2} William Sheffler,^{1,2} Vikram Khipple Mulligan,^{1,2} Wenqing Xu,⁵ James U. Bowie,³ David Baker^{1,2,6†}

The computational design of transmembrane proteins with more than one membrane-spanning region remains a major challenge. We report the design of transmembrane monomers, homodimers, trimers, and tetramers with 76 to 215 residue subunits containing two to four membrane-spanning regions and up to 860 total residues that adopt the target oligomerization state in detergent solution. The designed proteins localize to the plasma membrane in bacteria and in mammalian cells, and magnetic tweezer unfolding experiments in the membrane indicate that they are very stable. Crystal structures of the designed dimer and tetramer—a rocket-shaped structure with a wide cytoplasmic base that funnels into eight transmembrane helices—are very close to the design models. Our results pave the way for the design of multispans membrane proteins with new functions.

In recent years, it has become possible to de novo design, with high accuracy, soluble protein structures ranging from short constrained peptides to megadalton protein cages (1). There have also been advances in membrane protein design, as illustrated by an elegant zinc-transporting transmembrane peptide tetramer named Rocker (2) and an engineered ion-conducting oligomer based on the C-terminal transmembrane segment (TMs) of the *Escherichia coli* polysaccharide transporter Wza (3). Both are single membrane-spanning synthesized peptides with fewer than 36 residues. It has also been possible to design and confirm the transmembrane topology of multipass membrane proteins by using simple sequence hydrophobicity and charge-based models (4), but the extent to which the transmembrane helices pack with each other is not clear. Design of structurally defined multipass membrane proteins has remained a major challenge because of the difficulty in specifying structure within the membrane and in experimentally determining membrane protein structures generally; crystal structures of the full designed oligomeric states of Rocker- and the Wza-derived channel have not yet been determined, and to date, there are no crystal structures of de novo-designed multipass membrane proteins.

A major challenge for membrane protein design stems from the similarity of the membrane environment to protein hydrophobic cores. In the design of soluble proteins, the secondary structure and overall topology can be specified by the pattern of hydrophobic and hydrophilic residues, with the former inside the protein and the latter outside, facing solvent. This core design principle cannot be used for membrane proteins because the apolar environment of the hydrocarbon core of the lipid bilayer requires that outward-facing residues in the membrane also be nonpolar. Buried hydrogen bonds between polar side chains have been demonstrated to play an important role in the association of helical peptides within the membrane, overcoming the degeneracy in the nonpolar interactions (5–7).

We reasoned that a recently developed method for designing buried hydrogen bond networks (8) could allow specification of the packing interactions of transmembrane helices in multipass transmembrane proteins. We first explored the design of helical transmembrane proteins with four TMs—dimers of 76- to 104-residue hairpins or a single chain design of 156 residues—with hydrophobic spanning regions ranging from 21 to 35 Å (Figs. 1A and 2A), repurposing the Ser- and Gln-containing hydrogen bond networks in a designed soluble four-helix dimer with C2 symmetry [2L4HC2_23; Protein Data Bank (PDB) ID 5JOK] (8) to provide structural specificity. Four-helix bundles of different lengths with backbone geometries capable of hosting these networks were produced by using parametric generating equations (9), residues comprising the hydrogen bond networks and neighboring packing residues were introduced, and the remainder of the sequence was optimized by using Rosetta Monte Carlo (10) design calculations to obtain low-energy sequences. Connecting loops between the helices were built

with Rosetta. To specify the orientation of the designs (11) in the membrane when expressed in cells, at the designed lipid-water boundary on the extracellular/periplasmic side, we incorporated a ring of amphipathic aromatic residues and, at the lipid-water boundary on the cytoplasmic side, a ring of positively charged residues (Figs. 1A and 2A). Between these two rings, the surface residues are exposed to the hydrophobic membrane environment; these positions in Rosetta sequence design calculations were restricted to hydrophobic amino acids (supplementary materials). Consistent with the design, TMHMM predicts that the dimer designs contain 2 TMs and the single-chain design (scTMHC2) contains 4 TMs (fig. S1). On average, for each residue ~68% of the side-chain surface area is buried in the design models, which could provide substantial van der Waals stabilization (12).

Synthetic genes encoding the designs were obtained and the proteins expressed in *E. coli* and mammalian cells. The dimer design with the shortest hydrophobic span (15 residues; TMHC2_S) was poorly behaved in both *E. coli* and mammalian cells, but the dimer designs with longer spans—TMHC2, TMHC2_E, and TMHC2_L—localized to the cell membrane when expressed in human embryonic kidney (HEK) 293T cells (Fig. 1B) and in *E. coli*. The designed proteins were purified by extracting the *E. coli* membrane fraction with detergent, followed by nickel-nitrilotriacetic acid (NTA) chromatography and size exclusion chromatography (SEC) with a yield of ~2 mg/L (fig. S2, A and B). The designed proteins TMHC2, TMHC2_E, and TMHC2_L eluted as single peaks in SEC, and in analytical ultracentrifugation (AUC) experiments in detergent solution, the proteins sedimented as dimers, which is consistent with the design models (Fig. 1C and fig. S3). For the single-chain scTMHC2, the major species in SEC was the monomer, with a small side peak that was readily removed by purification (fig. S2B). Circular dichroism (CD) measurements showed that the designs were α -helical and highly thermal stable; the CD spectra at 95°C were similar to those at 25°C (Figs. 1D and 2B). TOXCAT- β -lactamase (T β L) assays (13), which couple *E. coli* survival to oligomerization and proper orientation of fused antibiotic resistance markers on the N and C termini, suggest that the N and C termini of TMHC2 are in the cytoplasm, as in the design models (fig. S4).

We more quantitatively characterized the folding stability of scTMHC2 using single-molecule forced unfolding experiments (Fig. 2) (14, 15). The designed protein reconstituted in a bicelle was covalently attached to a magnetic bead and a glass surface through its N and C termini (Fig. 2A and fig. S5). The distance between the bead and the surface was determined as a function of the applied mechanical tension. In unfolding experiments with the force slowly increasing (~0.5 pN/s), unfolding transitions were observed at ~18 pN and, upon force de-ramping, refolding transitions were observed

¹Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. ²Institute for Protein Design, University of Washington, Seattle, WA 98195, USA.

³Department of Chemistry and Biochemistry, UCLA-DOE Institute, Molecular Biology Institute, University of California, Los Angeles (UCLA), CA 90095, USA. ⁴Department of Bioengineering and Biophysics Group, University of California, Berkeley, Berkeley, CA 94720, USA. ⁵Department of Biological Structure, University of Washington, Seattle, WA 98195, USA. ⁶Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA.

*Present address: Department of Bioengineering, University of California, Berkeley, CA 94720, USA.

†Corresponding author. Email: dabaker@u.washington.edu

at ~ 9 pN (80.1% of the recorded unfolding traces had one-step unfolding transitions, and 84.6% of the refolding transitions had two steps) (Fig. 2C and figs. S6 and S7). Consistent with the internal symmetry of the single-chain design (Fig. 2A

and fig. S5), the two refolding step sizes were very similar (fig. S8). This unfolding and refolding asymmetry is consistent with a three-state free-energy landscape: the native state (N), an intermediate state containing only one hairpin

(I), and an unfolded state (U) (fig. S9). During unfolding at high force, only the barrier between the native and intermediate states is observed, whereas at the lower forces at which refolding occurs, both energy barriers become

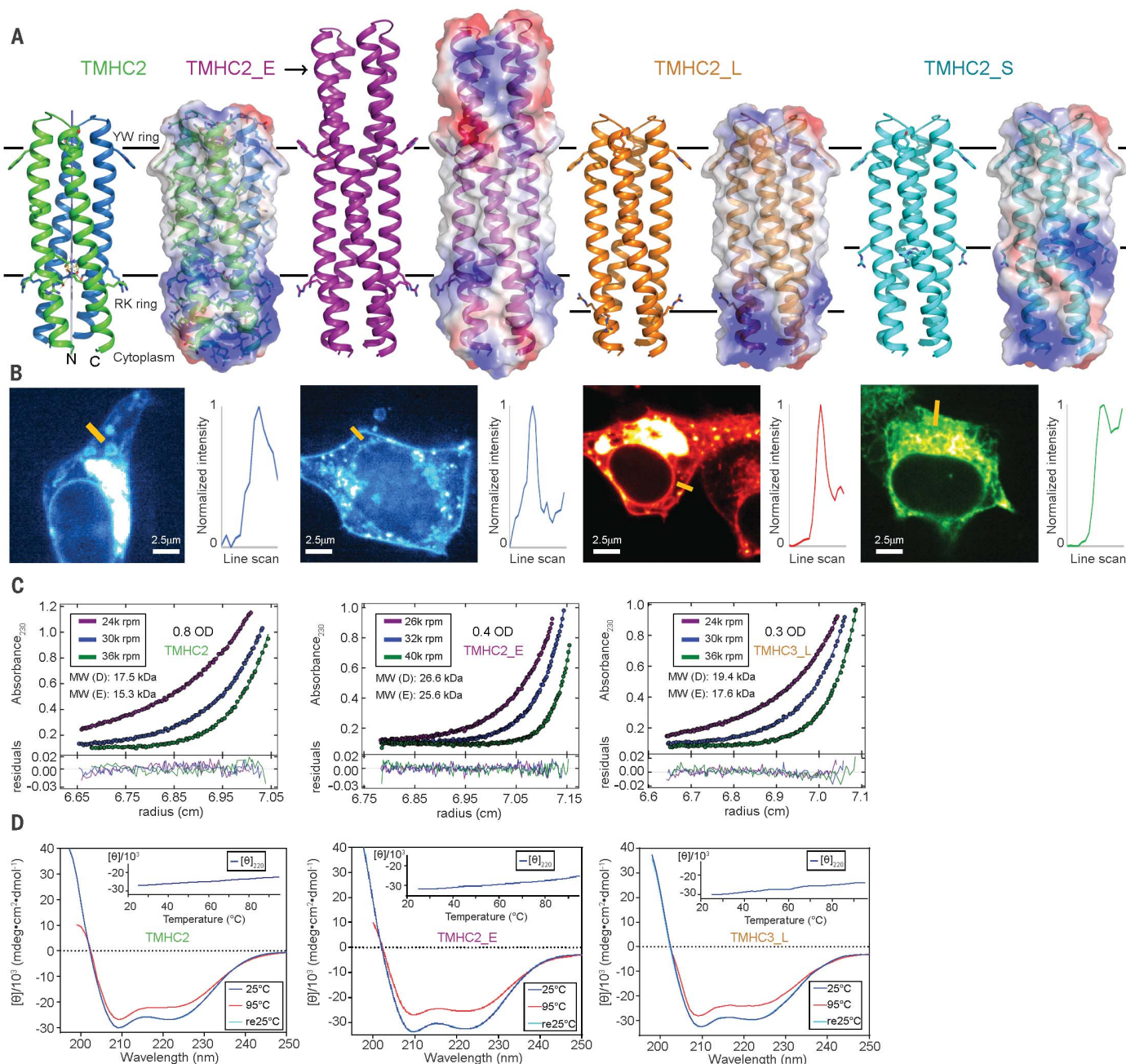


Fig. 1. Design and characterization of proteins with four transmembrane helices.

(**A** and **B**) From left to right, designs and data for TMHC2 (transmembrane hairpin C2), TMHC2_E (elongated), TMHC2_L (long span), and TMHC2_S (short span). (**A**) Design models with intra- and extra-membrane regions with different lengths. Horizontal lines demarcate the hydrophobic membrane regions. Ribbon diagrams are at left, electrostatic surfaces are at right, and the neutral transmembrane regions are in gray. (**B**) Confocal microscopy images for HEK293T cells transfected with TMHC2 fused to mTagBFP, TMHC2_E fused to mTagBFP, TMHC2_L fused to mCherry, and TMHC2_S fused to enhanced green fluorescent

protein. Line scans (yellow lines) across the membranes show substantial increase in fluorescence across the plasma membranes for TMHC2, TMHC2_E, and TMHC2_L, but less substantial increase for TMHC2_S. (**C**) Representative AUC sedimentation-equilibrium curves at three different rotor speeds. Each data set is globally well fit as a single ideal species in solution corresponding to the dimer molecular weight. “MW (D)” and “MW (E)” indicate the molecular weight of the oligomer design and that determined from experiment, respectively. (**D**) CD spectra and (inset) temperature melt. No apparent unfolding transitions are observed up to 95°C .

Fig. 2. Folding stability of the 156-residue single-chain TMHC2 (scTMHC2) design with four transmembrane helices.

(A) Design model (left) and electrostatic surface (right) of scTMHC2. N- and C-terminal helical hairpins are colored green and blue, respectively. Numbers indicate the order of the four TMs in the sequence. The linker connecting the two hairpins is colored magenta. Single-molecule forced unfolding experiments were conducted by applying mechanical tension to the N and C termini of a single scTMHC2 (fig. S5). (B) CD spectra of scTMHC2 at different temperatures. No unfolding transition is observed up to 95°C. (C) Single-molecule force-extension traces of scTMHC2. The unfolding and refolding transitions are denoted with red and blue arrows. (D) Folding energy landscape obtained from the single-molecule experiments. *N*, *I*, and *U* indicate the native, intermediate, and unfolded state, respectively.

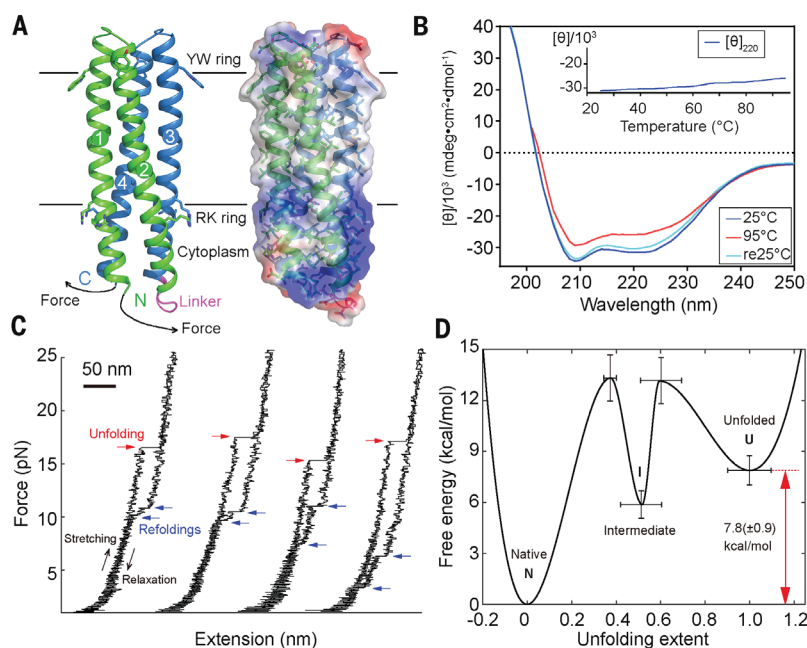
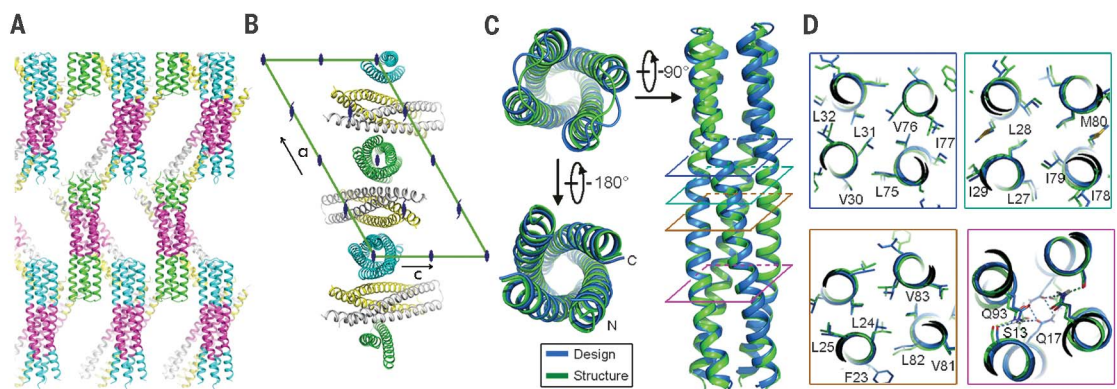


Fig. 3. Crystal structure of the designed transmembrane dimer TMHC2_E. (A and B)

Crystal lattice packing. (A) The extended soluble region mediates a large portion of the crystal lattice packing. The four helical hairpins in the asymmetric unit are colored green, gray, yellow, and blue, respectively. The TMs, in magenta, forms layers in the crystal separating the soluble regions. (B) The C2 axis of the design aligns with the crystallographic twofold. Two monomers (gray and yellow) are paired in a dimer, whereas the other two (green and blue) form two C2 dimers with two crystallographic adjacent monomers. The space group diagram (C121) is shown in the background. (C) Superposition of the TMHC2_E crystal structure and design model (RMSD = 0.7 Å over the core Cα atoms). (D) The side-chain packing arrangements at layers [(C), colored squares] at different depths in the membrane are almost identical to the design model.



prominent (fig. S9). The transition rates between the folded, intermediate, and unfolded states were determined by using the Bell model (16), yielding the relative free energies of the states and the associated barrier heights (Fig. 2D and fig. S10) (14). The overall thermodynamic stability of scTMHC2 is $7.8(\pm 0.9)$ kcal/mol on a per transmembrane helix basis, which is more stable than the naturally occurring helical membrane proteins studied thus far [folding free energy per helix for scTMHC2 is $2.0(\pm 0.2)$ kcal/(mol-helix) compared with 0.7 to 0.9 kcal/(mol-helix) for GlpG (14, 17) and 1.6 to 1.8 kcal/(mol-helix) for bacteriorhodopsin (18); error estimates in parentheses are propagated from the standard errors of the kinetics measurements].

We carried out crystal screens in different detergents for each of the designs and obtained crystals of the design with the most extensive

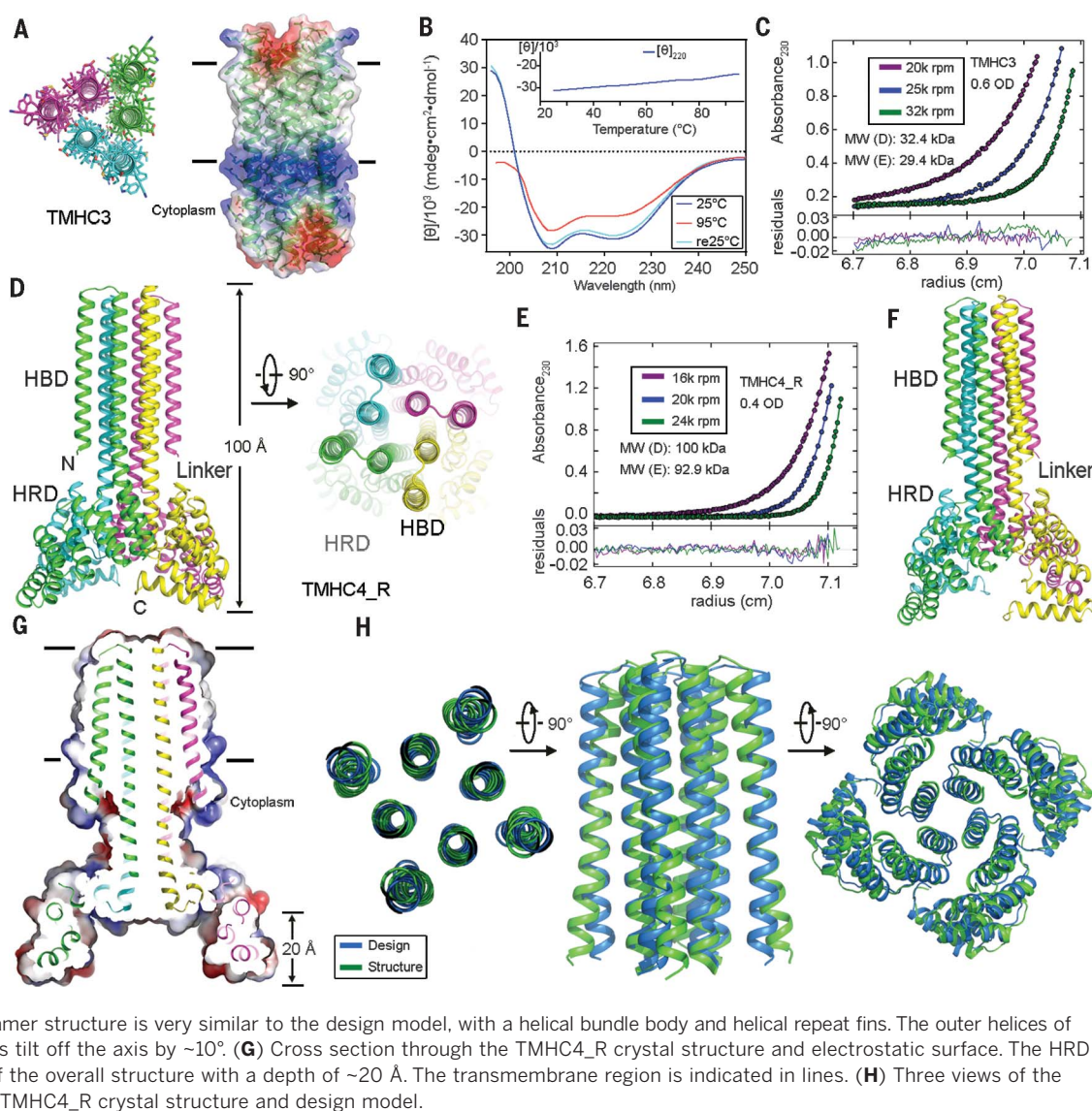
cytoplasmic region, TMHC2_E, in *n*-nonyl- β -D-glucopyranoside (NG). The crystals diffracted to 2.95-Å resolution, and we solved the structure by means of molecular replacement with the design model. As anticipated, the extended soluble region mediates the crystal lattice packing; there are large solvent channels around the designed TMs likely because of the surrounding disordered detergent molecules (Fig. 3A). Each asymmetric unit contains four helical hairpins: Two are paired in a dimer, whereas the other two form two C2 dimers through crystallographic symmetry with two monomers in adjacent asymmetric units. The C2 axis in the design is perfectly aligned with the crystallographic twofold (Fig. 3B). The conformations of the dimers in the three biological units are nearly identical, with very small differences due to crystal packing [$C\alpha$ root-mean-square devia-

tions (RMSDs), 0.60 to 0.84 Å] (fig. S11). Both the overall structure and the core side-chain packing are almost identical in the crystal structure and the design model, with a $C\alpha$ RMSD of 0.7 Å over the core residues (Fig. 3C). Two of the three buried hydrogen bonding residues within the membrane have conformations that almost exactly match the design model (S13 and Q93), but Q17 adopts a different rotamer, with the side-chain nitrogen donating a hydrogen bond to the main-chain carbonyl oxygen (Fig. 3D).

We used a similar approach to design a transmembrane trimer with six membrane-spanning helices (TMHC3) based on the 5L6HC3_1 scaffold (PDB ID 5LZS) (8). Guided by the results with the C2 designs, we chose a hydrophobic span of ~ 30 Å (20 residues) (Fig. 4A). The design was expressed in *E. coli* and purified to

Fig. 4. Stability and structural characterization of designs with six and eight membrane-spanning helices.

(A) Model of designed transmembrane trimer TMHC3 with six transmembrane helices. Stick representation from periplasmic side (left) and lateral surface view (right) are shown. **(B)** CD characterization of TMHC3. The design is stable up to 95°C. **(C)** Representative AUC sedimentation-equilibrium curves at three different rotor speeds for TMHC3. The data fit to a single ideal species in solution with molecular weight close to that of the designed trimer. **(D)** Model of designed transmembrane tetramer TMHC4_R with eight transmembrane helices. The four protomers are colored green, yellow, magenta, and blue, respectively. **(E)** AUC sedimentation-equilibrium curves at three different rotor speeds for TMHC4_R fit well to a single species, with a measured molecular weight of ~94 kDa. **(F)** Crystal structure of TMHC4_R. The overall tetramer structure is very similar to the design model, with a helical bundle body and helical repeat fins. The outer helices of the transmembrane hairpins tilt off the axis by ~10°. **(G)** Cross section through the TMHC4_R crystal structure and electrostatic surface. The HRD forms a bowl at the base of the overall structure with a depth of ~20 Å. The transmembrane region is indicated in lines. **(H)** Three views of the backbone superposition of TMHC4_R crystal structure and design model.



homogeneity, eluting on a gel filtration column as a single homogeneous species (fig. S2C). CD measurements showed that TMHC3 was highly thermostable, with the α -helical structure preserved at 95°C (Fig. 4B). AUC experiments showed that TMHC3 is a trimer in detergent solution, which is consistent with the design model (Fig. 4C and fig. S12A).

To explore our capability to design membrane proteins with more complex topologies, we designed a C4 tetramer with a two-ring, helical membrane-spanning region composed of eight TMs and an extended bowl-shaped cytoplasmic domain formed by repeating structures emanating away from the symmetry axis (Fig. 4D). The design has an overall rocket shape, with a height of ~100 Å, and can be divided into three regions: the helical bundle domain (HBD), the helical repeat domain (HRD), and the helical linker between the two. The central HBD was derived from the soluble design

5L8HC4_6 (8), and the bowl was derived from a designed helical repeat protein homo-oligomer (tpr1C4_2) (19). Helical linkers were built by using RosettaRemodel (20); a nine-residue junction was found to yield the correct helical register (fig. S13). After Rosetta sequence design calculations, a gene encoding the lowest energy design, TMHC4_R, was synthesized. The protein was expressed in *E. coli* and purified by using nickel affinity and gel filtration chromatography; the final yield was ~3 mg/L, and the purified protein chromatographed as a monodisperse peak in SEC (fig. S2C). CD experiments showed that the design was α -helical and thermostable up to 95°C (fig. S12B). AUC measurements showed that TMHC4_R is a tetramer in detergent solution, which is consistent with the design model (Fig. 4E and fig. S12C). After a systematic effort to screen detergents for crystallization, we obtained crystals in a combination of *n*-decyl- β -D-maltopyranoside (DM) and

NG in the P4 space group that diffracted to 3.9-Å resolution. We solved the crystal structure by means of molecular replacement using the design model ($R_{\text{work}}/R_{\text{free}} = 0.29/0.32$, with unambiguous electron density) (table S1 and fig. S14). The crystal lattice packing is primarily between the extended cytoplasmic domains; there may be minor detergent-mediated interactions between the transmembrane and helical repeat (HR) domains as well (fig. S15).

Although the resolution is insufficient for evaluating the details of the side-chain packing, it does allow backbone-level comparisons. There are four TMHC4_R monomers in one asymmetric unit, with nearly identical structures ($C\alpha$ RMSDs between 0.2 and 0.6 Å) (fig. S16A). The $C\alpha$ RMSDs between the structure and design model are 1.2 to 1.8 Å for the monomer transmembrane helices, 0.3 to 0.4 Å for the linkers, 1.1 to 1.5 Å for the HR domains, and 3.3 to 3.6 Å for the overall structure (fig.

S16B). As in the case of the C2 design, the C4 symmetry axis of the design coincides with the crystallographic axes of the crystal lattice (fig. S16C). The four tetramer structures on the crystal C4 axes have overall structures very similar to each other and to the design model (Fig. 4, F and G, and fig. S16A); the tetrameric transmembrane domain, HR domain, and overall tetramer structure have C α RMSDs to the design model of 1.3 to 1.5 Å, 3.3 to 3.8 Å, and 3.3 to 3.8 Å, respectively (Fig. 4H and fig. S16D, left). The deviation in the HR domain may result from crystal packing interactions between the termini; the C α RMSDs over the first 162 residues are 2.2 to 2.3 Å (fig. S16D, right). The main deviation from the design model is a tilting of the outer helices of transmembrane hairpins from the axis by $\sim 10^\circ$ (Fig. 4, F and G).

The agreement between the crystal structures of TMHC2_E and TMHC4_R with the design models demonstrates that transmembrane homo-oligomers containing multiple membrane-spanning regions and extensive extracellular domains can now be accurately designed. For future work, the general approach of first designing and characterizing hydrogen bond network-containing soluble versions of the desired transmembrane structures, and then converting to integral membrane proteins by redesigning the membrane-exposed residues, could be quite robust. Single-molecule forced unfolding and thermal denaturation experiments show that the designed proteins are highly stable. Although the designs lack the classic small residue packing in the core that is thought to be an important driver of membrane protein folding (21–24), like natural membrane proteins they bury more surface area than do typical soluble proteins, maximizing van der Waals packing contributions (12). The range of the design features—variable transmembrane and extracellular helix lengths and superhelical twists, extensive soluble domains,

and diverse oligomeric states—are substantial steps toward the complexity of natural transmembrane proteins with multiple membrane-spanning regions and extra-membrane domains that play important roles in ligand/substrate recognition and structure stabilization, such as in the adenosine 5'-triphosphate-binding cassette transporters, ion channels, ryanodine receptor, and γ -secretase (25, 26). The capability to accurately design complex multipass transmembrane proteins that can be expressed in cells opens the door to the design of a new generation of multipass membrane protein structures and functions.

REFERENCES AND NOTES

- P.-S. Huang, S. E. Boyken, D. Baker, *Nature* **537**, 320–327 (2016).
- N. H. Joh *et al.*, *Science* **346**, 1520–1524 (2014).
- K. R. Mahendran *et al.*, *Nat. Chem.* **9**, 411–419 (2017).
- P. Whitley, I. Nilsson, G. von Heijne, *Nat. Struct. Biol.* **1**, 858–862 (1994).
- C. Choma, H. Gratkowski, J. D. Lear, W. F. DeGrado, *Nat. Struct. Biol.* **7**, 161–166 (2000).
- F. X. Zhou, M. J. Cocco, W. P. Russ, A. T. Brunger, D. M. Engelman, *Nat. Struct. Biol.* **7**, 154–160 (2000).
- C. D. Tatko, V. Nanda, J. D. Lear, W. F. DeGrado, *J. Am. Chem. Soc.* **128**, 4170–4171 (2006).
- S. E. Boyken *et al.*, *Science* **352**, 680–687 (2016).
- P.-S. Huang *et al.*, *Science* **346**, 481–485 (2014).
- C. A. Rohl, C. E. M. Strauss, K. M. S. Misura, D. Baker, in *Methods in Enzymology* (2004), pp. 66–93.
- G. von Heijne, *Nat. Rev. Mol. Cell Biol.* **7**, 909–918 (2006).
- A. Oberai, N. H. Joh, F. K. Pettit, J. U. Bowie, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17747–17750 (2009).
- A. Elazar *et al.*, *eLife* **5**, e12125 (2016).
- D. Min, R. E. Jefferson, J. U. Bowie, T.-Y. Yoon, *Nat. Chem. Biol.* **11**, 981–987 (2015).
- D. Min, M. A. Arbing, R. E. Jefferson, J. U. Bowie, *Protein Sci.* **25**, 1535–1544 (2016).
- G. I. Bell, *Science* **200**, 618–627 (1978).
- R. Guo *et al.*, *Nat. Chem. Biol.* **12**, 353–360 (2016).
- Y.-C. Chang, J. U. Bowie, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 219–224 (2014).
- J. A. Fallas *et al.*, *Nat. Chem.* **9**, 353–360 (2017).
- P.-S. Huang *et al.*, *PLOS ONE* **6**, e24109 (2011).
- S.-Q. Zhang *et al.*, *Structure* **23**, 527–541 (2015).
- R. F. S. Walters, W. F. DeGrado, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13658–13663 (2006).
- X. Feng, P. Barth, *Nat. Chem. Biol.* **12**, 167–173 (2016).
- A. D. Meruelo, S. K. Han, S. Kim, J. U. Bowie, *Protein Sci.* **21**, 1746–1753 (2012).
- Y. Shi, *Cell* **159**, 995–1014 (2014).
- R. Fernandez-Leiro, S. H. W. Scheres, *Nature* **537**, 339–346 (2016).

ACKNOWLEDGMENTS

We thank J. Sumida for AUC support; A. Kang for crystallization support; D. Ma and Z. Wang for crystallography support; and the staff at the Advanced Light Source and P. Huang, Y. Hsia, A. Ford, L. Stewart, C. Xu, and many other members of the Baker laboratory for helpful discussions. This work was facilitated by the Hyak supercomputer at the University of Washington. **Funding:** This work was supported by the Howard Hughes Medical Institute (D.B.) and the National Institutes of Health (grant R01GM063919 to J.U.B.). P.L. was supported by the Raymond and Beverly Sackler fellowship. D.M. was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (grant NRF-2016R1A6A3A03007871). **Author contributions:** P.L. and D.B. designed the research, and P.L., D.M., F.D., K.Y.W., J.U.B., and D.B. wrote the manuscript. P.L. and D.B. carried out design calculations and developed the membrane protein design method. P.L. purified and characterized the designed proteins. D.M. and J.U.B. designed, performed, and analyzed single-molecule forced unfolding experiments. K.Y.W. and M.D.V. performed mammalian cell localization experiment. P.L. crystallized the designed proteins. P.L. collected and analyzed crystallographic data with help from W.X.; F.D. solved structures with help from P.L.; and S.E.B., C.Z., J.A.F., G.U., and W.S. contributed the soluble scaffolds. V.K.M. wrote the amino acid composition-based energy term. All authors discussed results and commented on the manuscript. **Competing interests:** D.B., P.L., S.E.B., C.Z., J.A.F., G.U., and W.S. are inventors on a U.S. provisional patent application submitted by the University of Washington that covers computational design of multipass transmembrane proteins described in this paper. **Data and materials availability:** Coordinates and structure files have been deposited to PDB with accession codes 6B87 (TMHC2_E) and 6B85 (TMHC4_R). All other data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6379/1042/suppl/DC1
Materials and Methods
Figs. S1 to S16
Table S1
References (27–39)

10 October 2017; accepted 11 January 2018
10.1126/science.aag1739