Supplementary Materials

Supplementary Materials for

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]*

correspondence to: dabaker@u.washington.edu

This PDF file includes:

Materials and Methods

Computational Modeling

Transmembrane region design

Orientation, RK ring and YW ring

The orientations of natural transmembrane proteins across the membrane follow the positive-inside rule (*27*)— that is, the side which is more positively charged, probably containing more Arg and Lys residues, would be in the cytoplasm. For transmembrane proteins with even numbers of TMs, the N- and C-termini is prefered to localize in the cytoplasmic side. Most of transmembrane helical hairpins have this orientation on *E. coli* membrane proteomes (*11*). The N- and C-termini of the designs made in this study are designed facing the cytoplasmic side, through adding a ring of Arg and Lys residues, named "RK ring", close to the N- and C-termini end of the helical bundle and designing the Arg and Lys to other polar residues on the other end. Only the changes would not clash are accepted during the design. Amphipathic aromatic residues (i.e., Trp and Tyr) prefers to locate at lipid-water boundary, forming a "YW ring". Trp and Tyr residues may interact with the lipid headgroups and water molecules in the boundary region and also pack with the lipid aliphatic chains, locking the transmembrane protein with the right register in membrane. The YW ring is designed on the other end of the RK ring, without steric clash.

Definition of hydrophobic transmembrane span

The hydrophobic transmembrane span could be defined as the region between the YW and KR rings. As all the designs have central symmetry, the central symmetry axis of designs should be perpendicular to the membrane plane; otherwise more hydrophobic and hydrophilic residues will be exposed to water solvent and buried in lipid membrane, respectively, which is energetically unfavorable. The center symmetry axis is aligned to the z axis, thus, the length of hydrophobic transmembrane region could be expressed as the distance between the mean z-coordinate values of the Cα atoms of YW and KR rings. We tested the lengths ranging from 21 to 35 Å.

Rosetta calculation

Together with the power of Rosetta algorithms, the recently developed RosettaMP (*28*) framework, provides a general modeling and design tool that combines scoring, conformational sampling, and mutation routines in the membrane environment.

Briefly, RosettaMP use a "span" object to store the start and end residue numbers of a single transmembrane span. An updated score function, which is derived from the original RosettaMembrane (*29*) score functions, is implemented in RosettaMP. RosettaMP uses the membrane position to score per-residue and residue pair interactions within the hydrophobic layers. The restructured membrane score function was verified using continuous regression testing and showed good scientific integrity (*28*).

Between the YW and KR rings, diverse hydrophobic residues are designed to replace all the polar residues those with polar atoms not involved in any hydrogen bond network, based on amino acid propensity in the membrane (*29*). The diversity could be achieved by application of an amino acid composition based energy term ("aa_composition") (*30*) in the design energy

function that penalized sequences possessing too many similar nonpolar amino acids . Sometimes Phe could be designed at positions roughly in the middle of the TM region, again, without causing any clash. The sequence was optimized using Rosetta Monte Carlo (*10*) design calculations to obtain low energy sequences.

Helices extension

We used the Crick coiled-coil parameters of 2L4HC2_23 but with lengths up to 14 more residues per helix, which form two additional full helical turns. The same hydrogen bond networks were introduced by specifying the residues at corresponding positions, and the remainder of the sequence was designed using Rosetta Monte Carlo calculations.

The helices were connected into a single chain by adding loops using look-ups to a structural database and Rosetta design, as in (*8*). Briefly, we generated an exhaustive database of loop backbones, spanning two helical regions with five or less residues. Candidate loops were identified via the alignment of the terminal residues of the elongated helical bundle to the database. Candidates within 0.35 Å root-mean-square deviation (RMSD) were then designed using Rosetta Monte Carlo design calculations and the lowest-scoring candidate is selected as the final loop design.

Junction design

RosettaRemodel protocol (*20*) was used to find the α-helical junction that can connect the helical bundle domain and helical repeat protein domain of TMHC4_R. We set up sampling runs for the junction lengths from 0 to 10 residues under four-fold symmetry. Distance constraints between the subunits of the tetrameric helical repeat protein and total energy are used for selection of the optimal helix length, which was found to be 9—other lengths either changed the helical register shifts or caused clashes. The models chosen from the fragment sampling stage for final sequence refinement are subjected to Rosetta Monte Carlo design calculations based on layer design protocol (*30*) to obtain low energy sequences, the sequences are converged quickly and the design with the lowest score are selected for experimental test.

Structural figures

All structural images for figures were generated using PyMOL (*31*).

Experimental Materials and Methods

Reagents

Chemicals used were of the highest grade commercially available and were purchased from Sigma-Aldrich (St. Louis, MO, USA), Invitrogen (Carlsbad, CA, USA), or Qiagen (Hilden, Germany). Detergents were from Anatrace (Maumee, OH, USA) and crystallization reagents were from Hampton (Aliso Viejo, CA, USA).

Cloning and expression

Synthetic genes were obtained from IDT (Coralville, Iowa, USA), Genscript Inc. (Piscataway, N.J., USA) and Gen9 Inc. (Cambridge, MA, USA) and either delivered in pET29b expression vector or as linear dsDNA and sub-cloned into pET-29b in-house via NdeI/XhoI restriction sites. The genes were designed without a stop codon, which allows expression of the protein with a C-terminal hexa-histidine tag. TMHC2 is cloned into pET-28b via NdeI/XhoI restriction sites, and with a N-terminal hexa-histidine tag followed by a thrombin cutting site. The assembled plasmids were transformed into chemically competent *E. coli* BL21(DE3)pLysS cells (Invitrogen). Gene expression was facilitated by growing pre-cultures in Luria-Bertani (LB) medium with a final concentration of 50 µg/ml kanamycin overnight at 37°C. 10 ml pre-cultures were used to inoculate 1L of LB medium, again containing 50 µg/ml kanamycin for plasmid selection. The cultures were grown at 37°C until an OD600 of 0.8-1.0 was reached and expression was induced by addition of isopropyl thio-β-D-galactoside (IPTG) to a final concentration of 0.2 mM. Protein was expressed at 18°C overnight and cells were harvested by centrifugation.

Cell lysis and purification

Cells were resuspended and homogenized in lysis buffer containing 25 mM Tris-HCl pH 8.0 and 150 mM NaCl. After further disruption with a French press, cell debris was removed by low-speed centrifugation for 10 min. The supernatant was collected and ultracentrifuged for 1 h at 150,000g. The membrane fraction was collected and homogenized with buffer containing 25 mM Tris-HCl pH 8.0 and 150 mM NaCl. n-Decyl-β-D-Maltopyranoside (DM; Anatrace) was added to the membrane suspension to a final concentration of 1.5% (w/v) and then incubated for 2 h at 4 °C. After another ultracentrifugation step at 150,000g for 30 min, the supernatant was collected and loaded on Ni^{2+}-nitrilotriacetate affinity resin (Ni-NTA; Qiagen), followed by a wash with 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 30 mM imidazole and 0.2% DM. Proteins were eluted with buffer containing 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 30 mM imidazole and 0.2% DM. After concentration to 10–15 mg ml^{-1}, proteins were further purified by gel filtration (Superdex-200 10/30; GE Healthcare). The buffer for gel filtration contained 25 mM Tris-HCl pH 8.0, 150 mM NaCl and various detergents. The purified proteins are separated on 16.5% Mini-PROTEAN®Tris-Tricine Gel (Bio-Rad) and visualized by Coomassie Blue staining. For TMHC2, the hexa-histidine tag is removed by cleavage of thrombin. After full cleavage, the reaction is stopped by addition of phenylmethanesulfonyl fluoride (PMSF), followed by another round of gel filtration purification. DM buffer is used for general purpose. For AUC experiments, the proteins were buffer exchanged in 20 mM sodium phosphate, pH 7.0, containing 200 mM NaCl supplemented with 0.5% Pentaethylene Glycol Monooctyl Ether (C8E5). For crystallization, different detergents are screened on gel filtration. The peak fractions were collected, concentrated to 10-15 mg ml^{-1}, aliquoted and flash frozen by liquid nitrogen.

Crystallization

The hanging-drop vapour-diffusion method was performed at 20 °C during crystallization. For TMHC2_E, crystals belonging to the space group C2 were obtained with protein purified in the presence of 0.2% n-nonyl-β-D-glucopyranoside (β-NG; Anatrace). The crystallization buffer was 0.05 M magnesium acetate tetrahydrate, 0.05 M sodium acetate 5.5 and 24 % v/v polyethylene glycol (PEG) 400. Rod cluster-shaped crystals appeared in 2-3 days and typically grew to full size in about 1 week. Single crystals could be obtained from one branch of the rod cluster. Crystals were dehydrated by exposing the drops to air for 5 min. For TMHC4_R, crystals in P4 space group were obtained in a detergents mixture of 0.2% β-NG and 0.1% DM. The crystallization buffer was 30 % v/v PEG 400, 100 mM 3-(N-morpholino)propanesulfonic acid (MOPS) pH 7.0, 100 mM NaCl. 10 mM N,N-Dimethyldecylamine-N-oxide (DDAO) was identified in detergent additive screen, which would improve the crystal quality. Plate-shaped crystals appeared in 1 week and typically grew to full size in about 4 weeks.

Data collection and structure determination

Crystal diffraction data for TMHC2_E and TMHC4_R, were collected at ALS beamline BL8.2.1 and BL5.0.1, respectively, and processed with the package HKL-2000 (*32*) with routine procedures. The scaled data were then used for structural determination and refinement. Further processing was carried out with programs from the CCP4 suites (*33*). Data collection statistics are summarized in Supplementary Table 1. For TMHC2_E and TMHC4_R, the best diffraction reached 2.95 Å and 3.9 Å, respectively.

Structure determination of TMHC2_E

From the data, the apparent space group was I212121, and an MR solution was found by Phaser (*34*) with TFZ=9.7, but refinement was unable to improve the structure. We then tried molecular replacement using Rosetta *ab initio* models and in lower symmetry groups. In doing so, we found a solution in C2 with four copies in the asymmetric unit: in two copies the designed dimer was part of the crystal symmetry, and the other two copies formed a dimer. Using Rosetta-Phenix refinement (*35*), the system refined to R/R_{free}=0.258/0.276.

Structure determination of TMHC4_R

Using the design model as well as ~25 models perturbed with RosettaCM (*36*), we were unable to find a solution in the apparent space group, P4212. After trying molecular replacement with lower symmetry, one of the perturbed models was able to place 4 copies in P4 (two pairs each related by tNCS). The original design model was inappropriate for MR as the angle between the transmembrane helices and repeat protein was different in the crystal lattice, however, several of the perturbed models accurately modeled this flexing, giving TFZ values of ~11 once all four copies were placed. This solution in P4 was then straightforward to refine in Phenix-Rosetta, giving a final R/R_{free} of 0.291/0.322.

Circular dichroism (CD) measurements

CD wavelength scan measurements were made on an AVIV CD spectrometer model 420. Protein concentrations ranged from 0.1-0.2 mg/ml in PBS (pH 7.4) buffer plus 0.2% DM. Wavelength scan spectra from 260 to 190 nm were recorded in triplets and averaged. The scanning increment for full wavelength scans was 1 nm. Temperature melts were conducted in 2

°C steps (heating rate of 2 °C/min) and recorded by following the absorption signal at a wavelength of 220 nm. Three sets of wavelength scan spectra were recorded at 25 °C, 95 °C and after cooled down to 25 °C.

TOXCAT-β−lactamase (TβL) assays

TβL assay is a genetic screen based on insertion of membrane-spanning segment to the N-terminus ToxR and C-terminus β-lactamase. ToxR is an oligomerization-dependent transcriptional activator, which could activate a chloramphenicol-resistance gene in this system. Bacterial survival on ampicillin monitors periplasmic localization of the C-terminus, and survival on chloramphenicol correlates with self-association of the membrane span and cytoplasmic localization of the N-terminus. The genes encoding TM designs were cloned into p-Mal vector using *XhoI* and *SpeI* restriction sites, and selected by spectinomycin. The TMs of the human erythrocyte sialoglycoprotein Glycophorin A (GpA) (*13*) is used as a positive control. The resulting plasmids were transformed into *E. coli* XL-1 blue (Agilent), plated on agar plates containing 50 μg/ml spectinomycin, and used to inoculate 10 ml of Luria Broth medium (LB) with 50 μg/ml spectinomycin and grown in a shaker at 200 rpm and 37°C overnight. The cultures were then inoculated into fresh medium, and until the density reached OD_{600} =1. 1 μl of the resulting cultures were plated at different dilutions on large 12-cm petri dishes containing spectinomycin, ampicillin alone or chloramphenicol.

Cell localization

Synthetic genes (codon optimized for human expression) were obtained from IDT and subcloned into pCAGGS vector via NheI and XhoI along with a fluorescent c-terminal protein tag (i.e., mTagBFP, eGFP, or mCherry). HEK293T cells were transiently transfected using TransIT-293T transfection reagent (Mirus Bio) along with constructs encoding the synthetic transmembrane proteins fused to a fluorescent tag. After 12-24 hours, cells were detached by incubation in PBS + 2 mM EDTA (Thermo Fisher Scientific, Sigma-Aldrich) for 4 minutes at room temperature. Cells were then transferred into Opti-MEM + 10% FBS (Thermo Fisher Scientific), seeded in 8 chambered coverglass wells (In Vitro Scientific) pre-coated with 1 mg/ml fibronectin (Thermo Fisher Scientific), and incubated for >4 hours to overnight at 37ºC. Wells were imaged on a spinning-disk confocal microscope (Nikon) at 60x. A line-scan through a region of the plasma membrane was performed using FIJI to determine if the protein of interest localized to the membrane.

Analytical ultracentrifugation

Analytical ultracentrifugation (sedimentation velocity and sedimentation equilibrium) experiments were carried out using a Beckman XL-I analytical ultracentrifuge (Beckman Coulter) equipped with an eight-cell An-50 Ti rotor. The proteins were run in 20 mM sodium phosphate, pH 7.0, containing 200 mM NaCl supplemented with 0.5% C8E5, no density matching was necessary and the solvent density was calculated as 1.0075 g mL^{-1}. The partial specific volume of the protein was calculated by the program Sednterp (*37*). For sedimentation velocity, absorbance at 230 nm versus radial location was recorded during centrifugation at 50,000 rpm at 20 °C. For sedimentation equilibrium, data were collected by UV detector at 20 °C for at least two protein concentrations at three rotor speeds. The data of sedimentation velocity and sedimentation equilibrium were analyzed using Sedfit and Sedphat (*38*).

Fig. S1. Design sequences. Hydrophobic TMs are indicated above the sequences. The invariant amino acids in the alignment are colored red. (A) Sequence alignment of TMHC2 with water-soluble version 2L4HC2_23. (B) Sequence alignment of designed transmembrane dimers with different TMs lengths. (C) Sequence alignment of TMHC2 with TMHC2_E. (D) Sequence of scTMHC2. (E and F) Sequence alignment of TMHC3 with 5L6HC3_1 and TMHC4_R TMs with 5L8HC4_6.

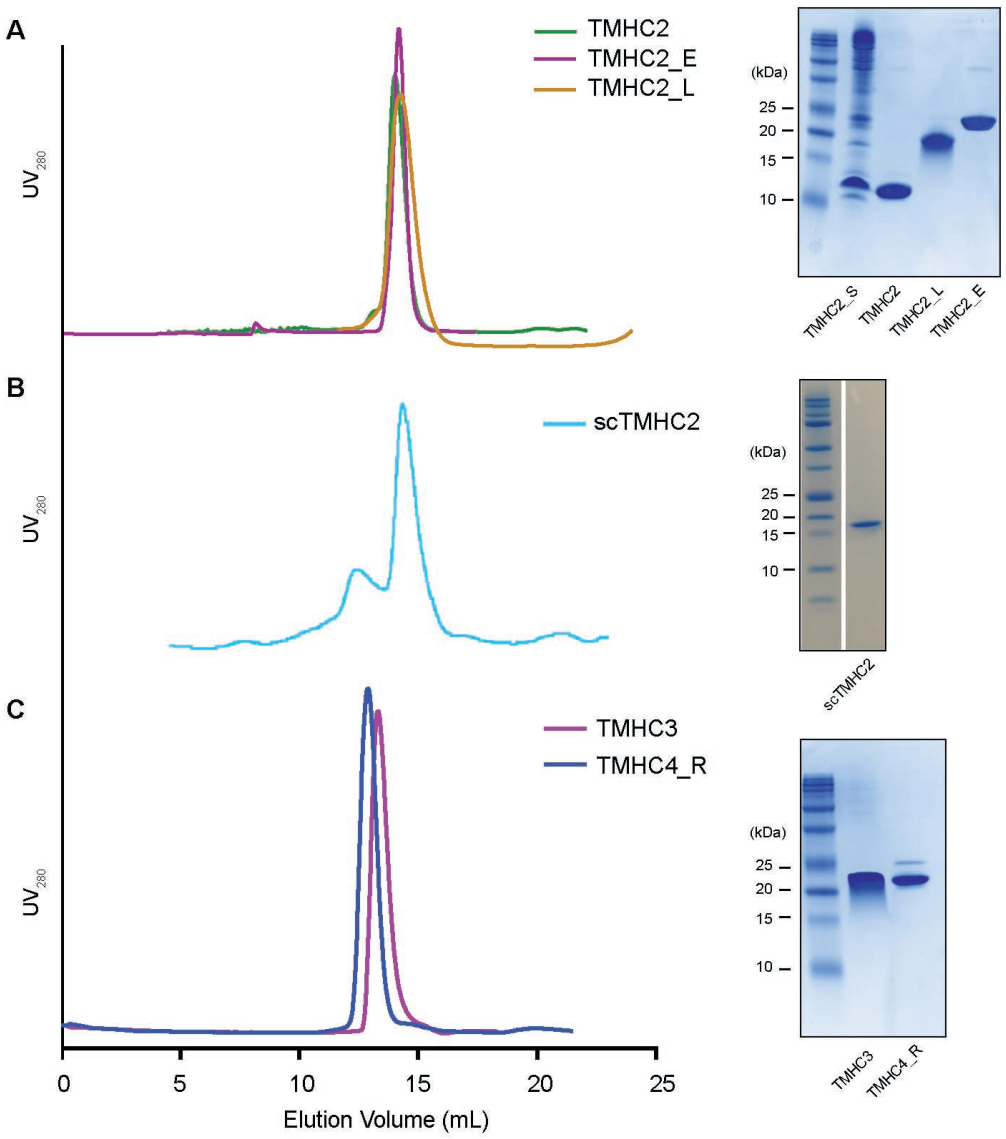

Fig. S2. Purification of designed multipass transmembrane proteins. (A) Representative gel filtration chromatography and SDS-PAGE of TMHC2, TMHC2_L and TMHC2_E. These dimeric designs elute at similar elution volume in gel filtration. TMHC2_L and TMHC2_E run at roughly dimer positions in SDS-PAGE. Only SDS-PAGE is shown for TMHC2_S, which expressed and behaved poorly. (B) Purification of scTMHC2. The elution volume of the major peak is comparable to the dimers. The small peak which elutes earlier is also from scTMHC2, probably due to intermolecular oligomers. Full separation of the two peaks is achieved after single chromatography. (C) Purification of TMHC3 trimer and TMHC4_R tetramer. TMHC3 runs at dimer position in SDS-PAGE, which may be an artifact due to incomplete denaturation.

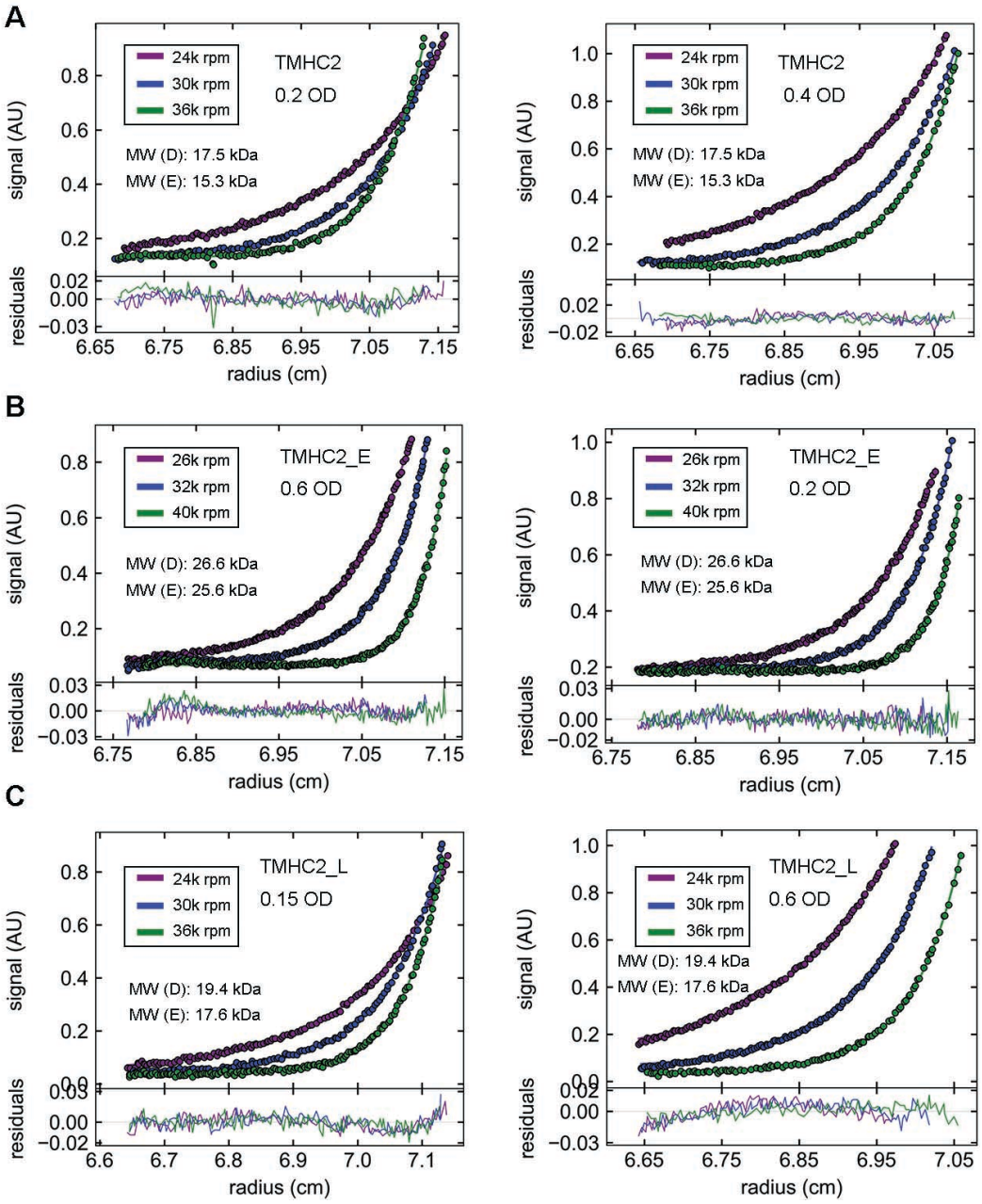

Fig. S3. Analytical ultracentrifugation sedimentation-equilibrium curves of dimeric designs. AUC sedimentation-equilibrium curves are shown at three different rotor speeds for two different concentrations for (A) TMHC2, (B) TMHC2_E, and (C) TMHC2_L. Together with the data in Fig. 1, three protein concentrations are tested. Each data set is well fit globally as a single ideal species in solution corresponding to the dimer molecular weight. 'MW (D)' refers to the molecular weight of the oligomer design and 'MW (E)' refers to the experimentally determined molecular weight. TMHC2 is about 2 kDa smaller than TMHC2_L due to the cleavage of the hexa-histidine tag.

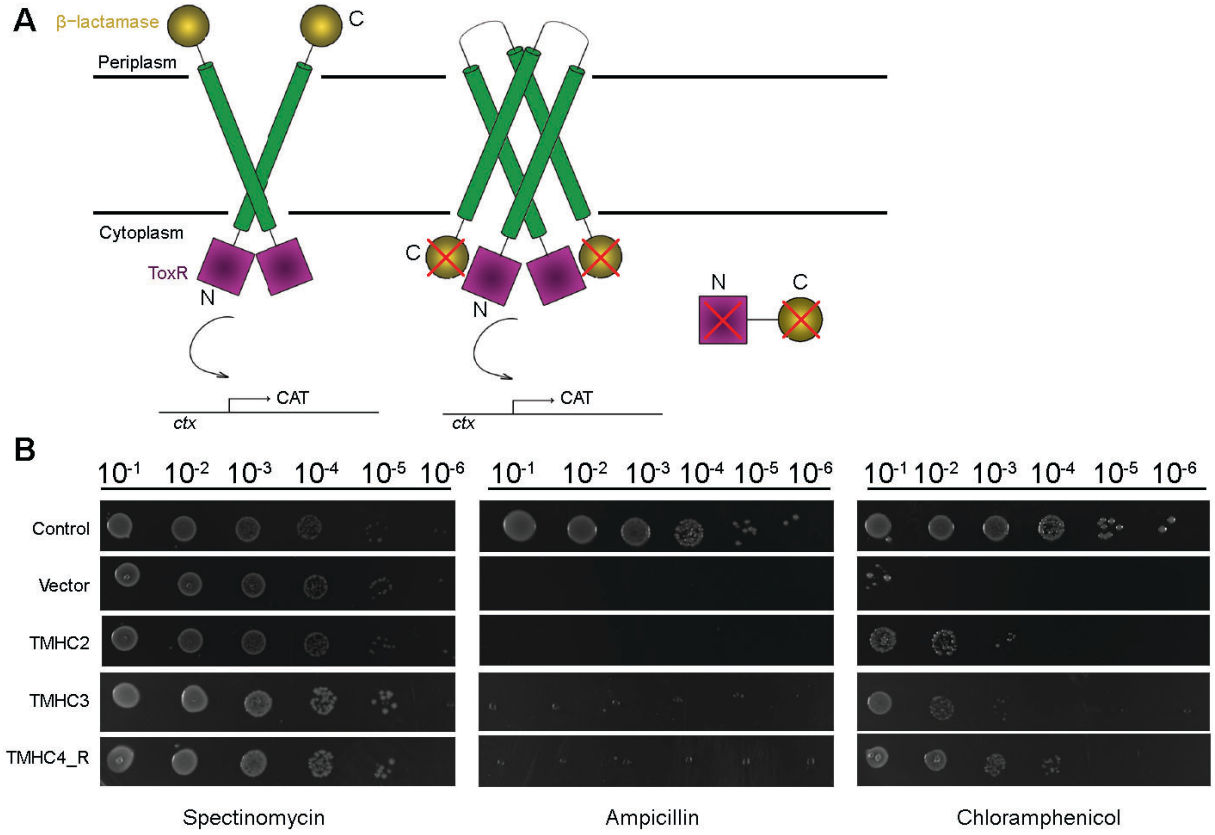

Fig. S4. TOXCAT-β−lactamase (TβL) assays. (A) Schematic diagram of the TβL assay. The TβL genetic construct fuses a transmembrane protein to two antibiotic selection markers: ToxR and β-lactamase, which locate at N- and C- terminal of the transmembrane protein, respectively. ToxR reports on self-association when oligomerize in the cytoplasm, activating the chloramphenicol resistance gene, while β-lactamase function only in the periplasm. Positive control, designed dimer and empty vector control are shown from left to right. (B) Viability of designs in TβL assay. Cells are normalized to the same density, and plated in serial dilutions on LB agar containing different antibiotics. TMHC2, TMHC3 and TMHC4_R designs could survive on chloramphenicol plates ~100 fold better than the empty vector control, but not ampicillin plates, suggesting the designs form oligomers with both termini in the cytoplasm.

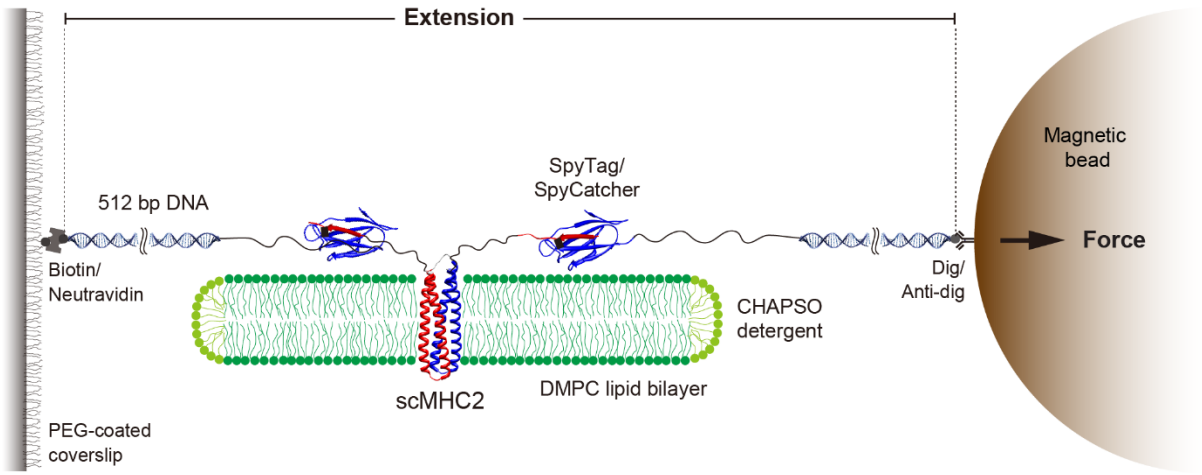

Fig. S5. Schematic diagram of single-molecule forced unfolding experiment. The single chain homodimer scTMHC2 was covalently linked to 512 bp DNA handles through SpyTag-SpyCatcher complex by an established protocol (*15*). The protein was incorporated into a bicelle composed of DMPC lipids and CHAPSO detergents. The protein-DNA hybrid was then attached to the glass surface via a biotin-neutravidin complex and a magnetic bead via a digoxigenin-antidigoxigenin complex. The glass coverslip was passivated by polyethylene glycol to prevent proteins and beads from nonspecific bindings to the surface. Mechanical tension was applied with a pair of magnets to a single scTMHC2, increasing at ~0.5 pN/s from a few pN to tens of pN. The bead height (extension) was measured by monitoring changes of the bead diffraction pattern using a 60 Hz CCD camera. The custom built single-molecule magnetic tweezer apparatus and a detailed experimental procedure was described previously (*14*, *15*).

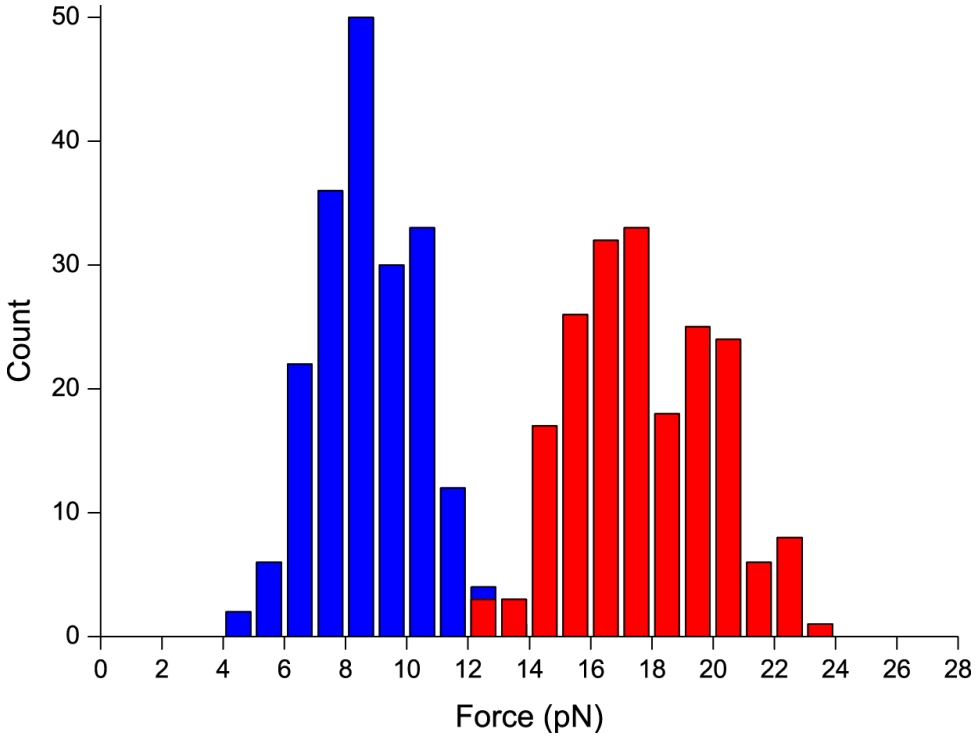

Fig. S6. Unfolding and refolding force distributions. The unfolding and refolding forces were measured from force-extension curves like those shown in Fig. 2C. The unfolding force distribution (red) was constructed from the unfolding events from the native state (N=196). The refolding force distribution (blue) was constructed from the refolding events from the unfolded state (N=196). The forces correspond to the force at the initiation of the event.

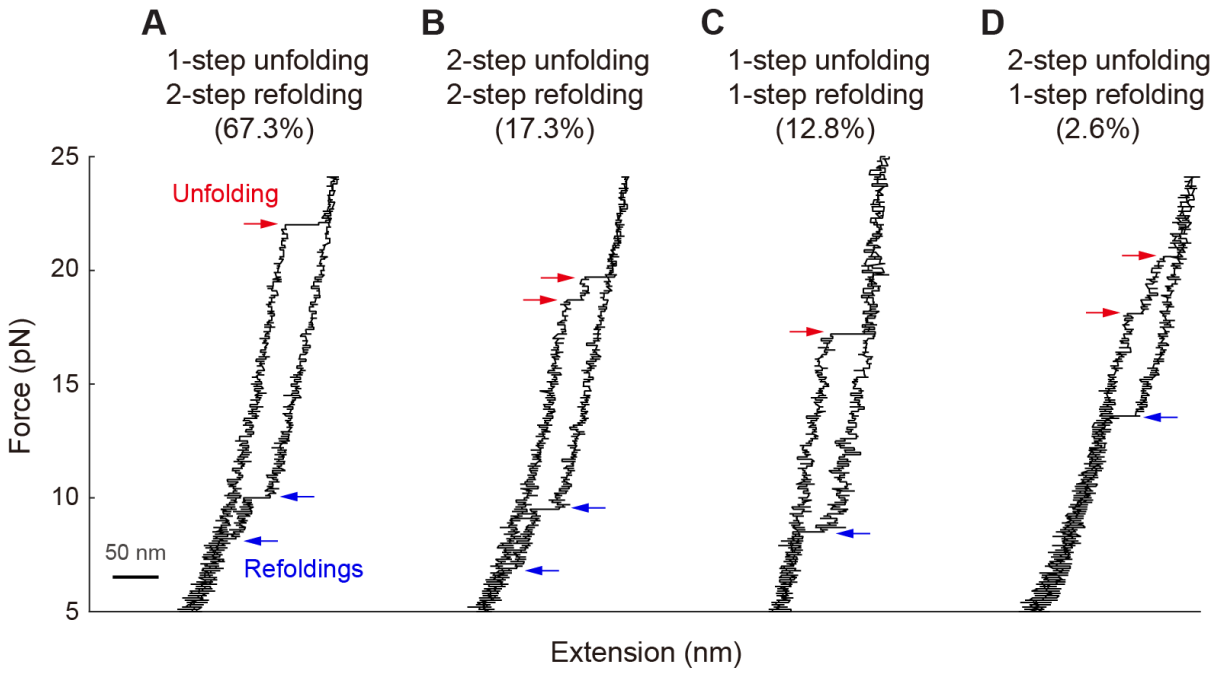

Fig. S7. Single-molecule trace classification. Representative force-extension traces for each type are shown. (A) 1-step unfolding and 2-step refolding. (B) 2-step unfolding and 2-step refolding. (C) 1-step unfolding and 1-step refolding. (D) 2-step unfolding and 1-step refolding. The occurrence probability for each type in all traces (total N=196) was 67.3% (A), 17.3% (B), 12.8% (C), and 2.6% (D).

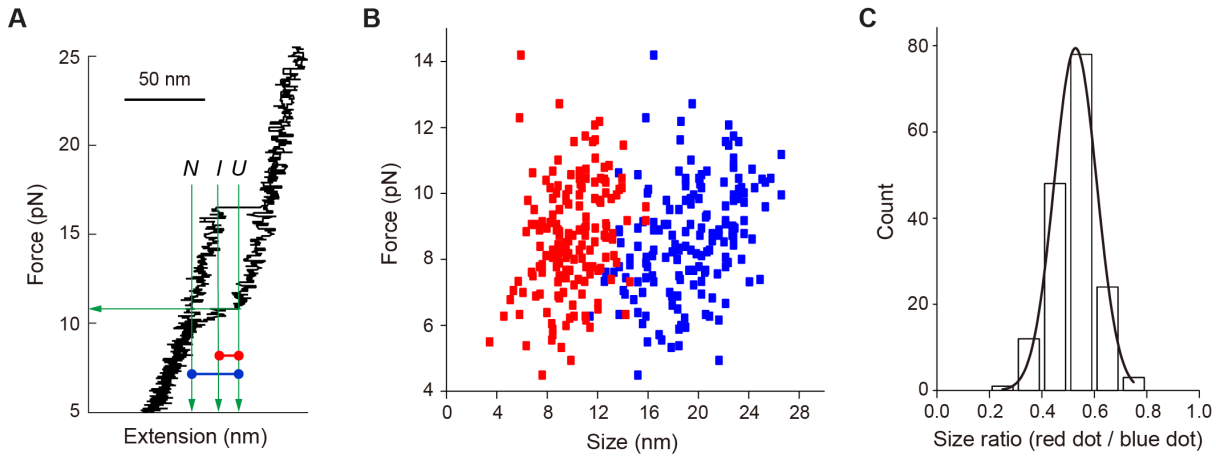

Fig. S8. Refolding size analysis. (A) Example force-extension trace for refolding size analysis. The refolding step size to the intermediate state was measured at the point of a refolding event (red line). For comparison, the total refolding size was measured at the same force by measuring the extension difference between the fully unfolded and the full folded states (blue line). Notations *N*, *I*, and *U* in the panel indicate the native, intermediate, and unfolded states respectively. (B) Scatter plot of extension size *vs* force. The values for intermediate refolding size (*U* to *I*) and the total refolding size (between *N* and *U*) are denoted with red and blue dots respectively (each *N*=166). (C) Count histogram for size ratio. The size ratio was calculated as the intermediate refolding size divided by the total refolding size. The histogram was fitted with Gaussian function (peak: 0.53, standard deviation: 0.08), indicating that half the protein is refolded in the intermediate state.

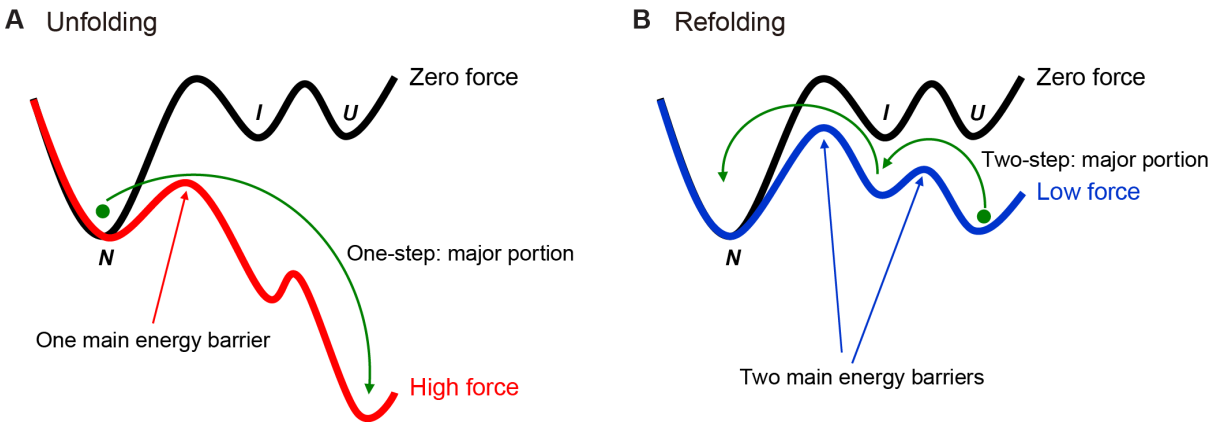

Fig. S9. Conceptual three-state energy landscape. (A) Energy landscape during unfolding at high force. The high force tilts the zero-force landscape toward the unfolded state so that during unfolding the main energy barrier is effectively reduced to the one between the native and intermediate states. (B) Energy landscape during refolding at low force. The landscape is slightly tilted at lower forces and the both energy barriers become prominent during refolding. Notations *N*, *I*, and *U* in the panels indicate the native, intermediate, and unfolded state respectively.

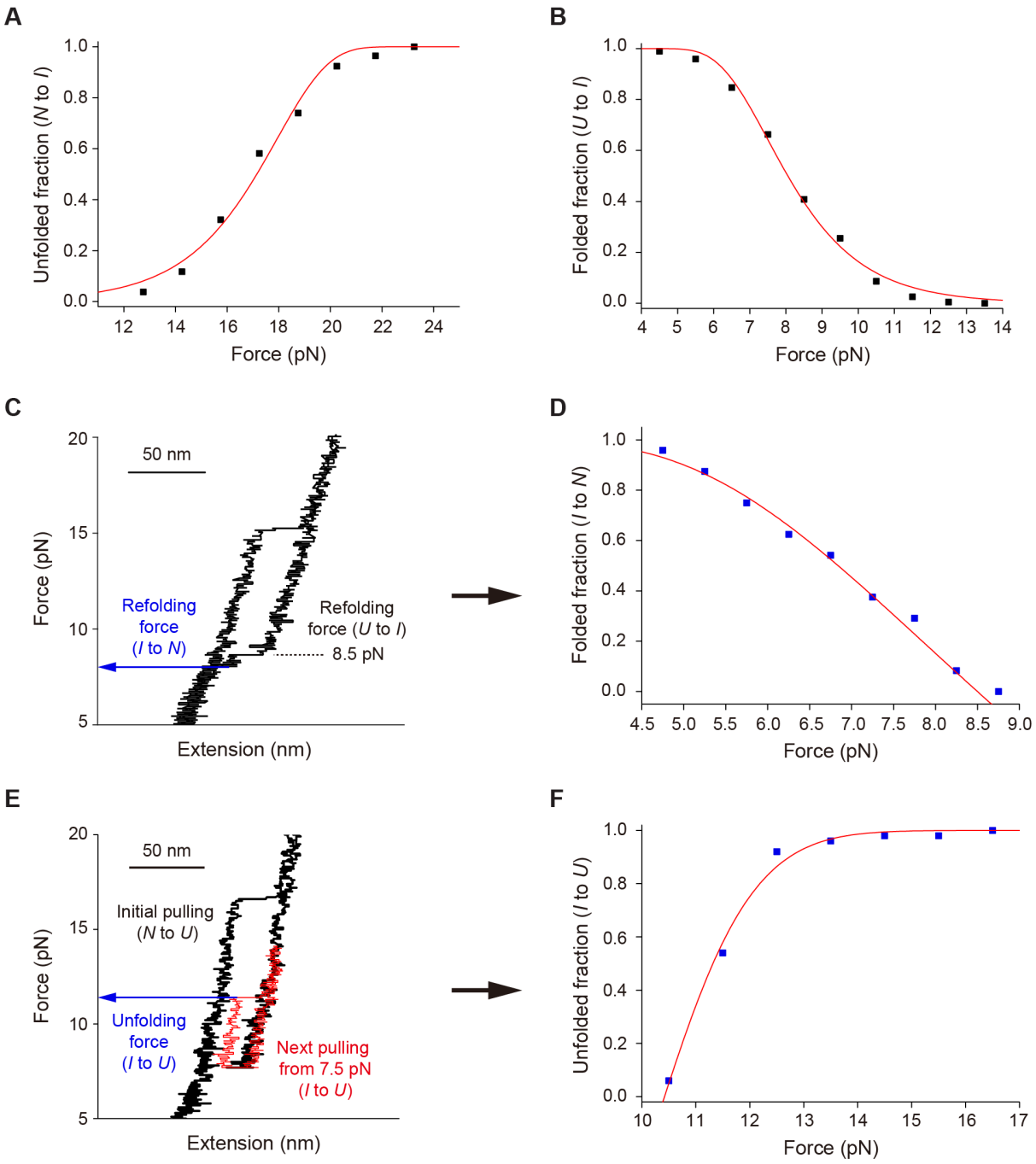

Fig. S10. Unfolding and refolding kinetics analysis. (A) *N*-to-*I* unfolded fraction as a function of force. The unfolded fraction at each force was calculated by integrating the unfolding force distribution up to the force in fig. S6. The data was fitted with a combined equation $P_{NI} = 1 - \exp(-\frac{k_{NI} k_B T}{A x_{NI}^{\dagger}} \frac{e^{f_{NI} x_{NI}^{\dagger} / k_B T}}{f_{NI}})$ from rate equation and Bell equation (*14*, *16*), where P_{NI} is the *N*-to-*I* unfolded fraction, f_{NI} is the *N*-to-*I* unfolding force, $k_B T$ is the thermal energy constant, A is the proportionality constant of $df/dt = Af$ determined from force calibration data, k_{NI} is the *N*-to-*I* rate constant at zero force, and $x_{NI}^{\dagger}$ is the distance between the *N* state and the *N*-to-*I* transition

state. The k_{NI} and $x_{NI}^{\dagger}$ were obtained by the fitting as 2.32(±1.70)·10^{-5} s^{-1} and 2.22(±0.19) nm respectively. The measured rate constant was converted to free energy difference between the *N* state and the *N*-to-*I* transition state using the Kramer equation (*14*, *39*). The energy landscape is summarized in Fig. 2D. (B) *U*-to-*I* folded fraction as a function of force. The folded fraction data was similarly analyzed as above for the unfolded fraction data. The fitting function is $P_{UI} = 1 - \exp(\frac{k_{UI} k_B T}{A x_{UI}^{\dagger}} \frac{e^{-f_{UI} x_{UI}^{\dagger} / k_B T}}{f_{UI}})$ where P_{UI} is the *U*-to-*I* folded fraction, f_{UI} is the *U*-to-*I* refolding force, k_{UI} is the *U*-to-*I* rate constant at zero force, and $x_{UI}^{\dagger}$ is the distance between the *U* state and the *U*-to-*I* transition state. The k_{UI} and $x_{UI}^{\dagger}$ were obtained as 15.63(±7.43) s^{-1} and 2.40(±0.2) nm respectively. (C) Example force-extension trace for extracting *I*-to-*N* kinetics. Refolding events from *I* to *N* were counted as a function of force. (D) *I*-to-*N* folded fraction as a function of force. The data was analyzed by a modified equation $P_{IN} = 1 - \exp[\frac{k_{IN} k_B T}{A x_{IN}^{\dagger}} (\frac{e^{-f_{IN} x_{IN}^{\dagger} / k_B T}}{f_{IN}} - \frac{e^{-f_{UI} x_{IN}^{\dagger} / k_B T}}{f_{UI}})]$ where P_{IN} is the *I*-to-*N* folded fraction, f_{UI} is the *U*-to-*I* refolding force (set at 8.5 pN), f_{IN} is the *I*-to-*N* refolding force, k_{IN} is the *I*-to-*N* rate constant at zero force, and $x_{IN}^{\dagger}$ is the distance between the *I* state and the *I*-to-*N* transition state. The k_{IN} and $x_{IN}^{\dagger}$ were measured as 0.40(±0.48) s^{-1} and 0.84(±0.54) nm respectively. (E) Example force-extension trace for extracting *I*-to-*U* kinetics. Unfolding events from *I* to *U* were counted as a function of force. (F) *I*-to-*U* unfolded fraction as a function of force. The unfolded fraction data was similarly analyzed as the folded fraction data in panel D. The fitting function is $P_{IU} = 1 - \exp[-\frac{k_{IU} k_B T}{A x_{IU}^{\dagger}} (\frac{e^{f_{IU} x_{IU}^{\dagger} / k_B T}}{f_{IU}} - \frac{e^{f_{UI} x_{IU}^{\dagger} / k_B T}}{f_{UI}})]$ where P_{IU} is the *I*-to-*U* unfolded fraction, f_{UI} is the *U*-to-*I* refolding force (set at 7.5 pN), f_{IU} is the *I*-to-*U* unfolding force, k_{IU} is the *I*-to-*U* kinetic rate at zero force, and $x_{IU}^{\dagger}$ is the distance between the *I* state and the *I*-to-*U* transition state. The k_{IU} and $x_{IU}^{\dagger}$ were measured as 0.58(±0.05) s^{-1} and 0.46(±0.001) nm respectively.

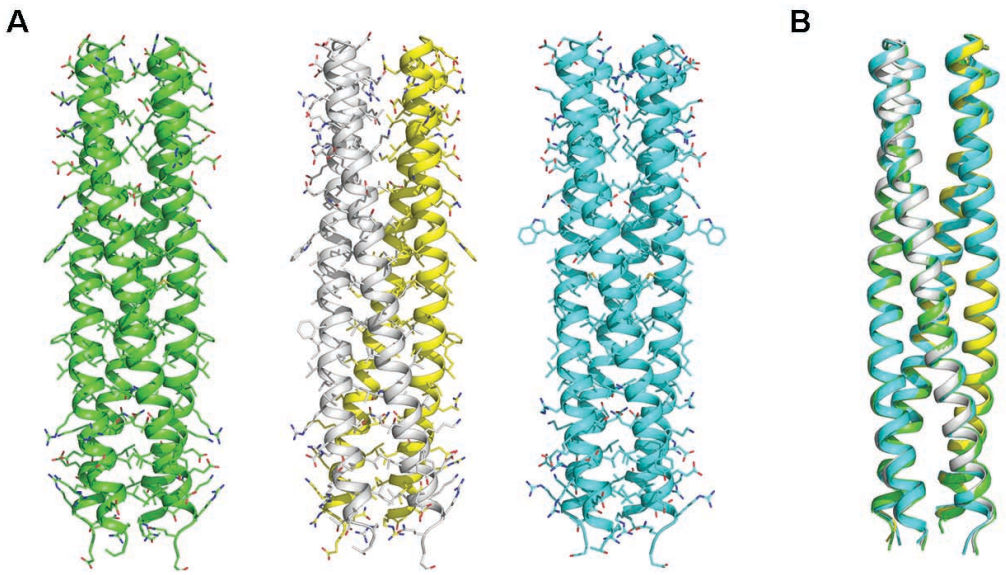

Fig. S11. Nearly identical structures for the three dimers in the crystal of TMHC2_E. (A) Structures for the three TMHC2_E dimers. Monomers are colored the same as Fig. 3B. (B) Structure alignment for the three dimers with Cα RMSDs between 0.60 and 0.84 Å.

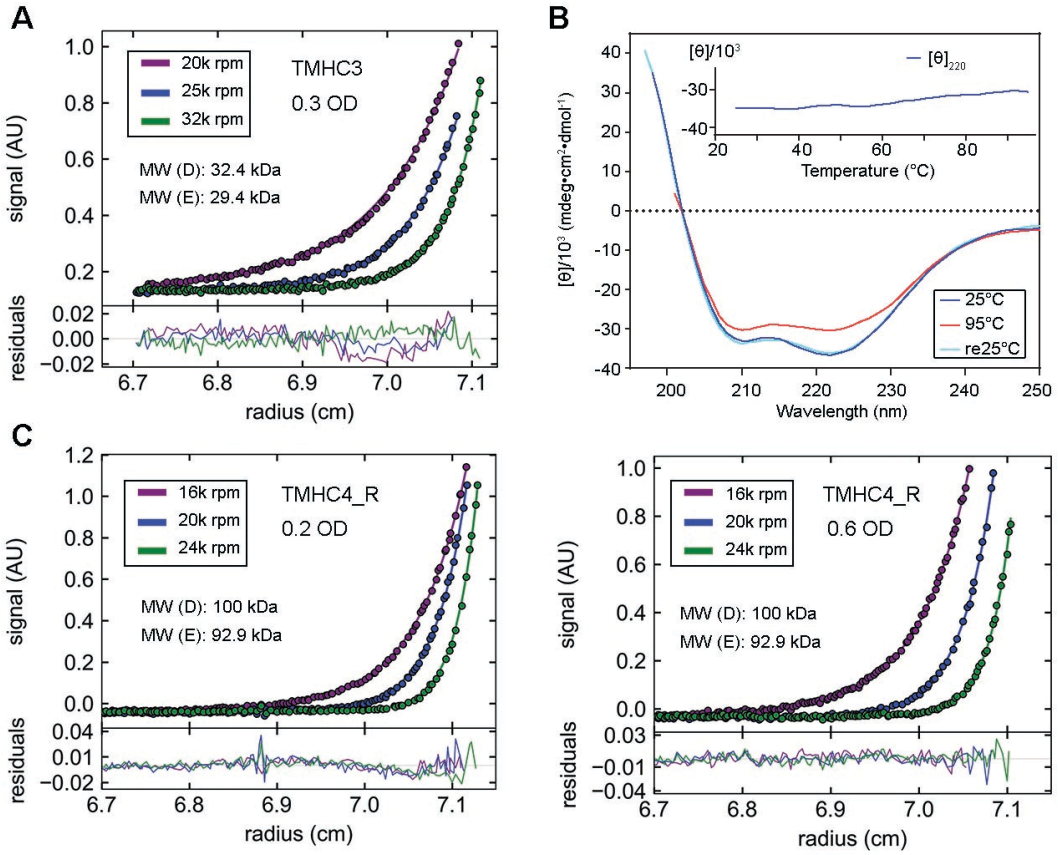

Fig. S12. Structural analysis of trimer and tetramer. (A) AUC sedimentation-equilibrium curves at three different rotor speeds are shown for TMHC3. The data, together with that in Fig. 4C, globally fit to trimer molecular weight. (B) Circular dichroism characterization of TMHC4_R. the design is alpha helical and does not thermally unfold. (C) AUC sedimentation-equilibrium curves for TMHC4_R. The curves for two concentrations, together with that in Fig. 4E, could globally fit to tetramer molecular weight.

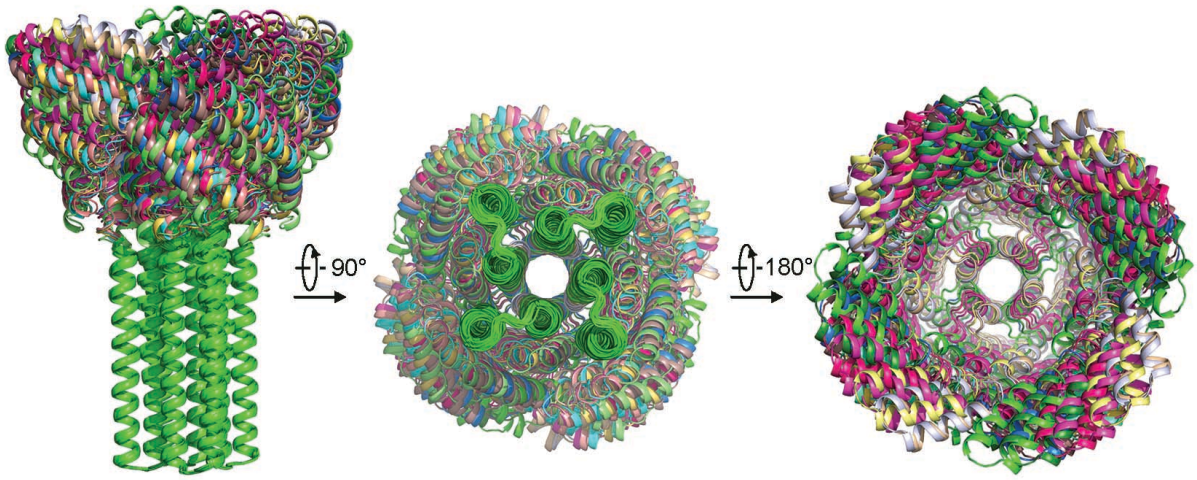

Fig. S13. Sampling the helical junction between helical bundle 5L8HC4_6 and helical repeat homo-tetramer tpr1C4_2. Three successive views of junction assemblies. The ensemble of inserted helical linker and helical repeat domain is shown moving relative to the helical bundle as a result of sampling the helical linker. The tetramer structure of the helical repeat domain kept intact with defined tetrameric distance constraints.

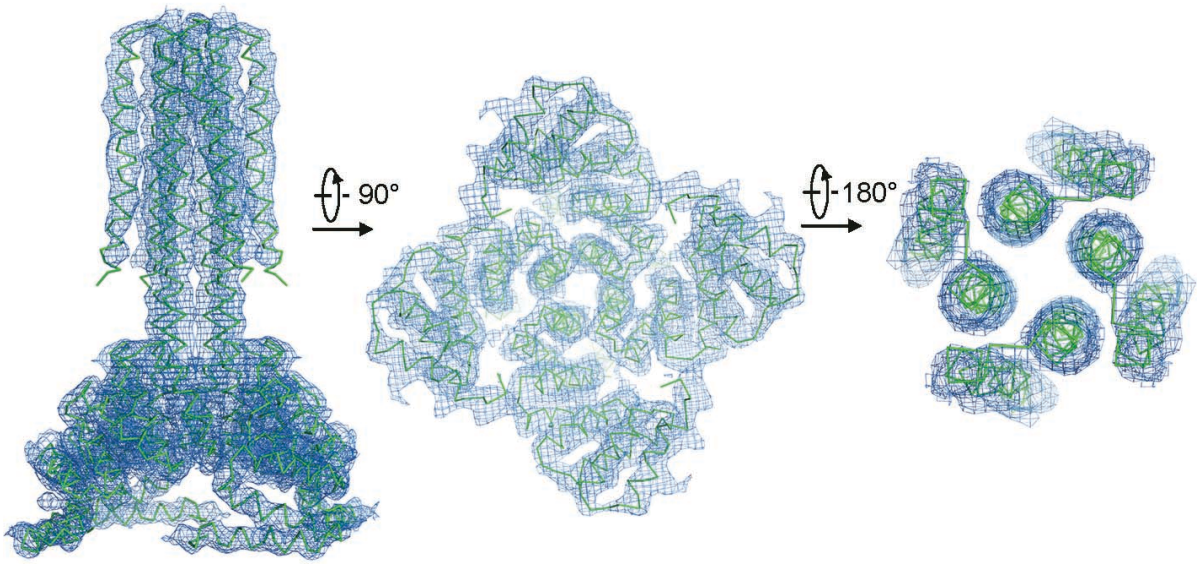

Fig. S14. Representative electron density maps for TMHC4_R. Three successive views of the 2Fo-Fc electron density map for a tetramer of TMHC4_R in the crystal are shown.It is contoured at 1.0σ. A Cα trace of the model is shown (green).

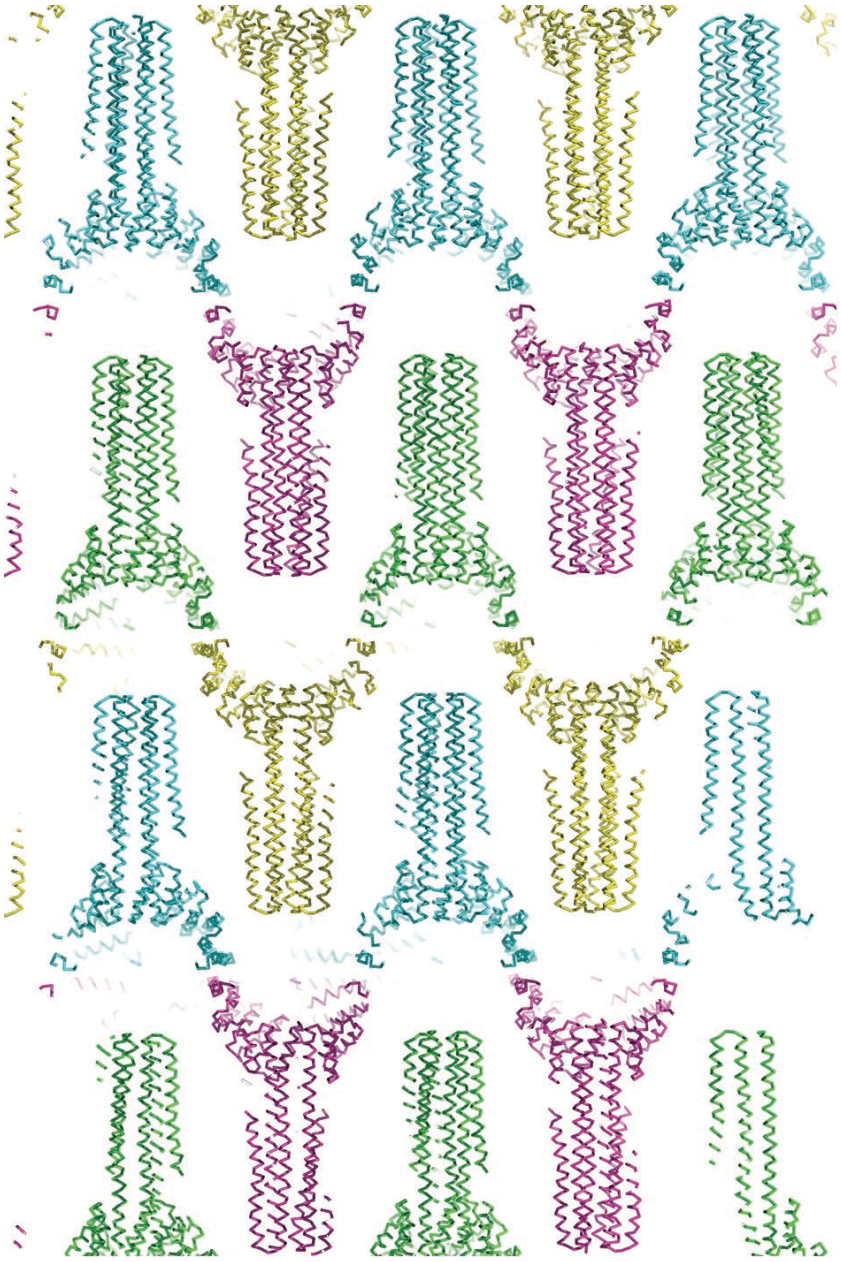

Fig. S15. Crystal lattice packing for TMHC4_R. The helical repeat domain mediates a major portion of the crystal lattice packing. The four tetramers in the crystal are colored green, magenta, yellow and cyan, respectively. There is no direct crystal contact from transmembrane helical bundle, however, detergents may mediate some contacts between helical bundle and helical repeat domains.

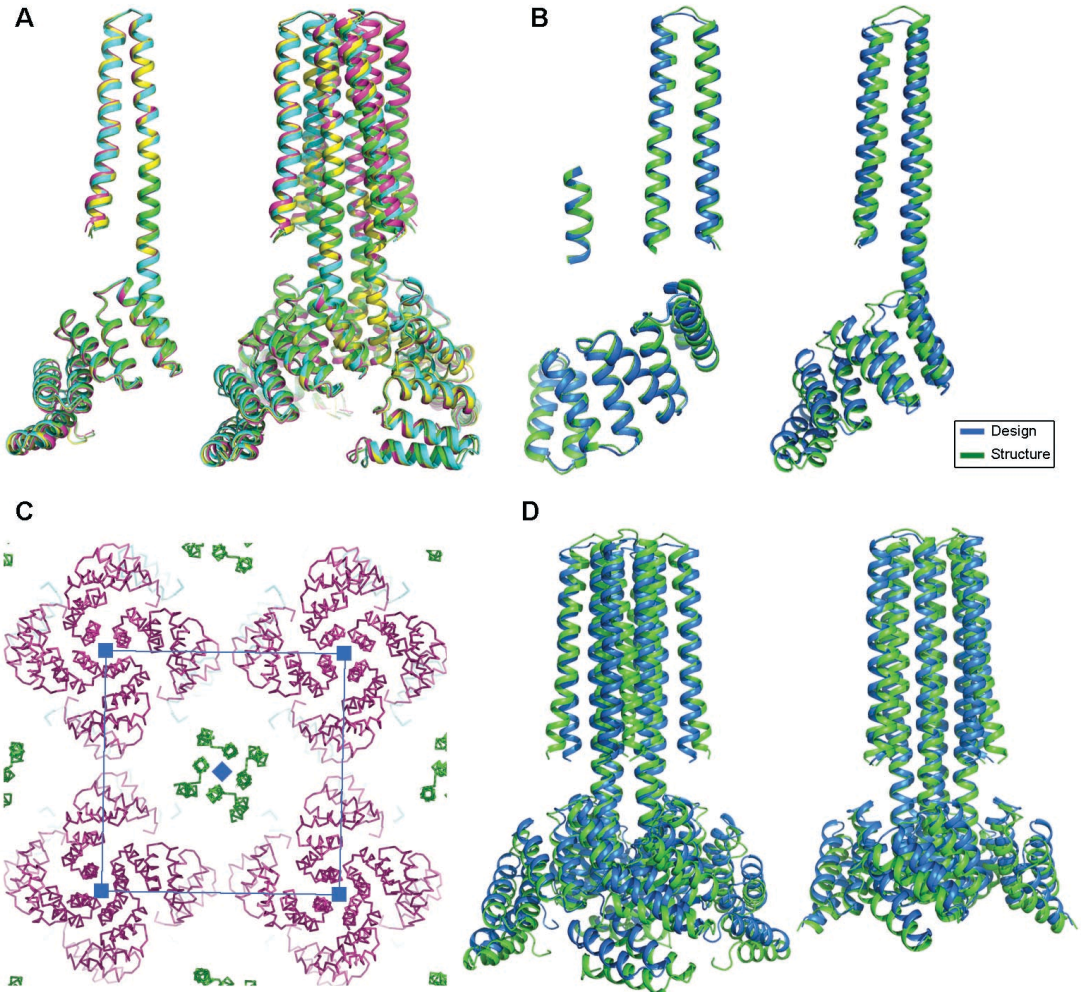

Fig. S16. Structural analysis for TMHC4_R. (A) Structure alignments for the four monomers (left) and tetramers (right). The four monomers and tetramers could be aligned with Cα RMSDs from 0.2 to 0.6 Å and 0.2 to 1.0 Å, respectively. (B) Superpositions of crystal structure and design model for the TMHC4_R monomer. Structure alignments of the transmembrane, linker and HR domains are shown on the left, while the overall structure superposition is on the right. (C) The crystallographic four fold aligns with the C4 axis of the design. The space group diagram (P4) is shown in the background. (D) Structure alignments of crystal structure and design model for the TMHC4_R tetramer. The overall tetramer structure aligns to the design with Cα RMSDs of 3.3-3.8 Å (left). The first 162 residues of the tetramer in crystal structure align to the design with Cα RMSDs of 2.2-2.3 Å (right).

Table S1. Statistics of data collection and refinement

Data	TMHC2_E	TMHC4_R
Integration Package	HKL2000	HKL2000
Space Group	C2	P4
Content per ASU	4 monomers	4 monomers
Unit Cell (Å)	103.5, 121.6, 52.0	80.2, 80.2, 251.6
Unit Cell (°)	90, 119.9, 90	90, 90, 90
Resolution (Å)	50~2.95 (3.03~2.95)	50~3.9 (4.01~3.90)
Outer shell (Å)		
R_{merge}	0.097 (0.635)	0.133 (2.065)
I/sigma	9.6 (1.4)	18.2 (1.0)
$CC_{1/2}$	0.688	0.398
Completeness (%)	92.5 (60.7)	99.6 (97.4)
Number of unique reflections	10,899	14,545
Redundancy	3.5 (2.6)	12.0 (8.7)
R_{work} /R_{free}	0.258/0.276	0.291/0.322
No. atoms		
Overall	6508	6764
Protein	6508	6764
Water	0	0
Other entities	0	0
Average B value (Å^2)		
Protein	84.8	172.5
Water	N/A	N/A
Other entities	N/A	N/A
R.m.s. deviations		
Bonds (Å)	0.011	0.021
Angle (°)	1.257	1.558
Ramachandran plot statistics (%)		
Most favourable	100	99.4
Additionally allowed	0.0	5.9
Generously allowed	0.0	0.0
Disallowed	0.0	0.0

Every diffraction dataset was collected from a single crystal. Values in parentheses are for the highest resolution shell. $R_{merge}=\Sigma_h\Sigma_i|I_{h,i}-I_h|/\Sigma_h\Sigma_i I_{h,i}$, where I_h is the mean intensity of the i observations of symmetry related reflections of h. $R=\Sigma|F_{obs}-F_{calc}|/\Sigma F_{obs}$, where F_{calc} is the calculated protein structure factor from the atomic model (R_{free} was calculated with 5% of the reflections selected).