Supplementary Information

An intramembrane chaperone catalyzes conformational editing in membrane protein folding

Shi Ho Kim†, W.C. Bhashini Wijesinghe†, Eojin Kim†, Jaehyun Nam, Seoyoon Kim, Minseung Kim, Victor W. Sadongo, and Duyoung Min*

Department of Chemistry, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea

† These authors contributed equally to this work.

* Corresponding author: dymin@unist.ac.kr

Contents

SpyTag003–GlpG–SnoopTag

SpyTag003 GlpG

MRGVPHIVMVDAYKRYKGGSGGSAALRERAGPVTWVMMIAAVVVFIAMQILGDQEVMLWLAWPFDPTLKFEFWRYFTHALMHFSLMHILFNLL
WWWYLGGAVEKRLGSGKLIVITLISALLSGYVQQKFSGPWFGGLSGVVYALMGYVWLRGERDPQSGIYLQRGLIIFALIWIVAGWFDLFGMSMA
NGAHIAGLAVGLAMAFVDSLNARKRKGGSGGSKLGDIEFIKVNKGGSHHHHHH

SnoopTag 6xHisTag

SnoopCatcher–Peptide linker–AviTag

6xHisTag SnoopCatcher

MHHHHHHGGSKPLRGAVFSLQKQHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAFQIVNGEVRDV
TSIVPQDIPATYEFTNGKHYITNEPIPPKGSVVPSTQPVTTPPATTKPPATTIPPSDDPNAVVPGLNDIFEAQKIEWHE

Peptide linker AviTag

SpyCatcher003

6xHisTag TEV protease cleavage site SpyCatcher003 C: Cysteine for DNA attachment

MSYYHHHHHHDYDIPTTENLYFQGGCGTTLSGLSGEQGPSGDMTTEEDSATHIKFSKRDEDGRELAGATMELRDSSGKTISTWISDGHVKDFYL
YPGKYTFVETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAHT

YidC

YidC

MDSQRNLLVIALLFVSFMIWQAWEQDKNPQPQAQQTTQTTTTAAGSAADQGVPASGQGKLISVKTDVLDLTINTRGGDVEQALLPAYPKELNSTQ
PFQLLETSPQFIYQAQSGLTGRDGPDNPANGPRPLYNVEKDAYVLAEGQNELQVPMTYTDAAGNTFTKTFVLKRGDYAVNVNYNVQNAGEKPLE
ISSFGQLKQSITLPPHLDTGSSNFALHTFRGAAYSTPDEKYEKYKFDTIADNENLNISSKGGWVAMLQQYFATAWIPHNDGTNNFYTANLGNGIAAI
GYKSQPVLVQPGQTGAMNSTLWVGPEIQDKMAAVAPHLDLTVDYGWLWFISQPLFKLLKWIHSFVGNWGFSIIIITFIVRGIMYPLTKAQYTSMAKM
RMLQPKIQAMRERLGDDKQRISQEMMALYKAEKVNPLGGCFPLLIQMPIFLALYYMLMGSVELRQAPFALWIHDLSAQDPYYILPILMGVTMFFIQK
MSPTTVTDPMQQKIMTFMPVIFTVFFLWFPSGLVLYYIVSNLVTIIQQQLIYRGLEKRGLHSREKKKSHHHHHH

6xHisTag

SpyCatcher002–spNW50–SpyTag002

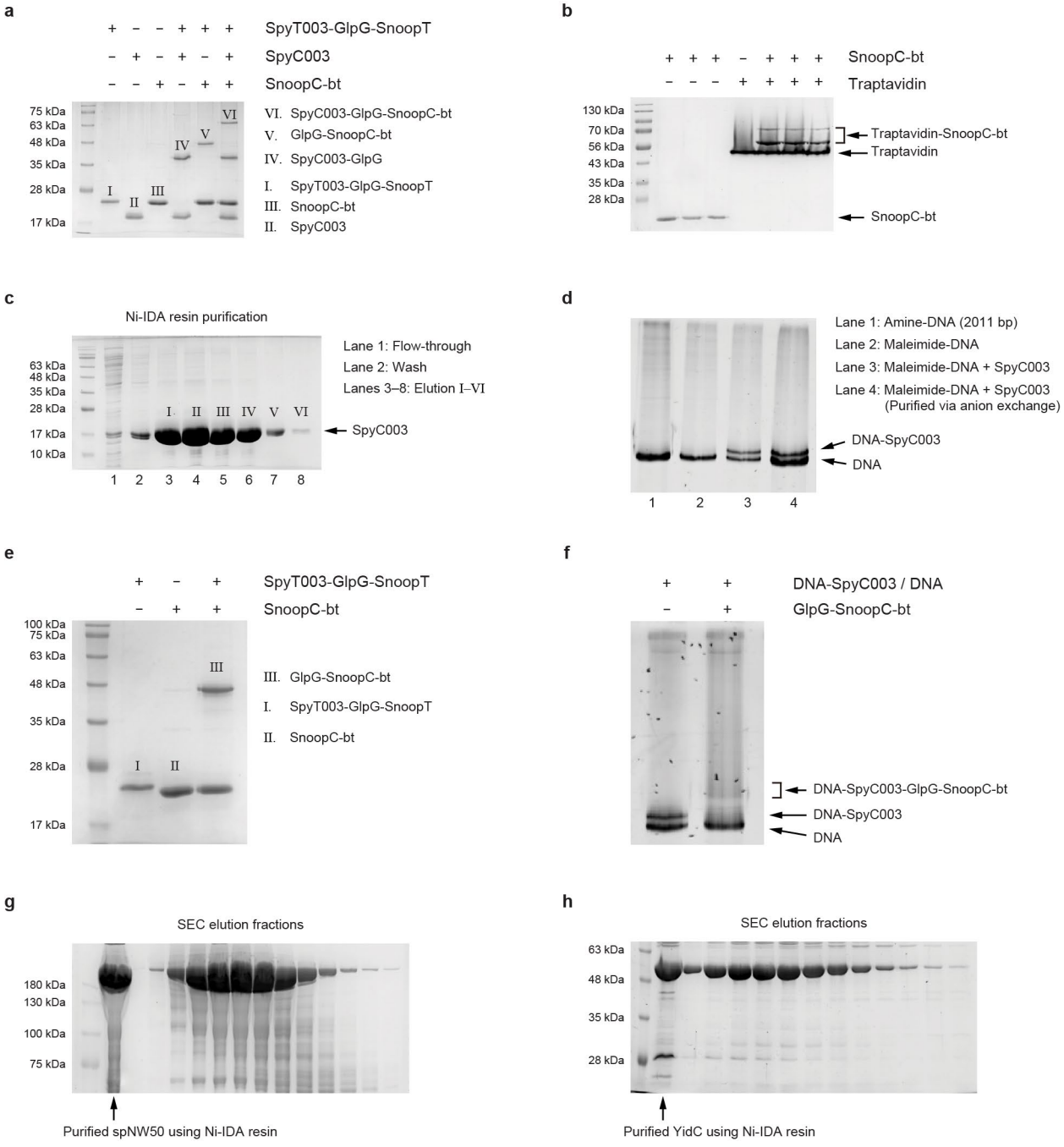
6xHisTag SpyCatcher002

MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSGAMVTTLSGLSGEQGPSGDMTTEEDSATHIKFSKRDEDGRELAGATMELRDSSGKTI
STWISDGHVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGEATKGDAHTSTFSKLREQLGPVTQEFWDNLEKETEGLRQEMSKD
LEEVKAKVQPYLDDFQKKWQEEMELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMRDRARAHVDALRTHLAPYSDELRQRLAARLEALK
ENGGARLAEYHAKATEHLSTLSEKAKPALEDLRQGLLPVLESFKVSFLSALEEYTKKLNTQGTPVTQEFWDNLEKETEGLRQEMSKDLEEVKAKV
QPYLDDFQKKWQEEMELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMRDRARAHVDALRTHLAPYSDELRQRLAARLEALKENGGARLA
EYHAKATEHLSTLSEKAKPALEDLRQGLLPVLESFKVSFLSALEEYTKKLNTQGTPVTQEFWDNLEKETEGLRQEMSKDLEEVKAKVQPYLDDFQ
KKWQEEMELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMRDRARAHVDALRTHLAPYSDELRQRLAARLEALKENGGARLAEYHAKATE
HLSTLSEKAKPALEDLRQGLLPVLESFKVSFLSALEEYTKKLNTQGTPVTQEFWDNLEKETEGLRQEMSKDLEEVKAKVQPYLDDFQKKWQEEM
ELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMRDRARAHVDALRTHLAPYSDELRQRLAARLEALKENGGARLAEYHAKATEHLSTLSEK
AKPALEDLRQGLLPVLESFKVSFLSALEEYTKKLNTQGTPVTQEFWDNLEKETEGLRQEMSKDLEEVKAKVQPYLDDFQKKWQEEMELYRQKVE
PLRAELQEGARQKLHELQEKLSPLGEEMRDRARAHVDALRTHLAPYSDELRQRLAARLEALKENGGARLAEYHAKATEHLSTLSEKAKPALEDLR
QGLLPVLESFKVSFLSALEEYTKKLNTQLPGTGAAALEVPTIVMVDAYKRYK

spNW50 SpyTag002

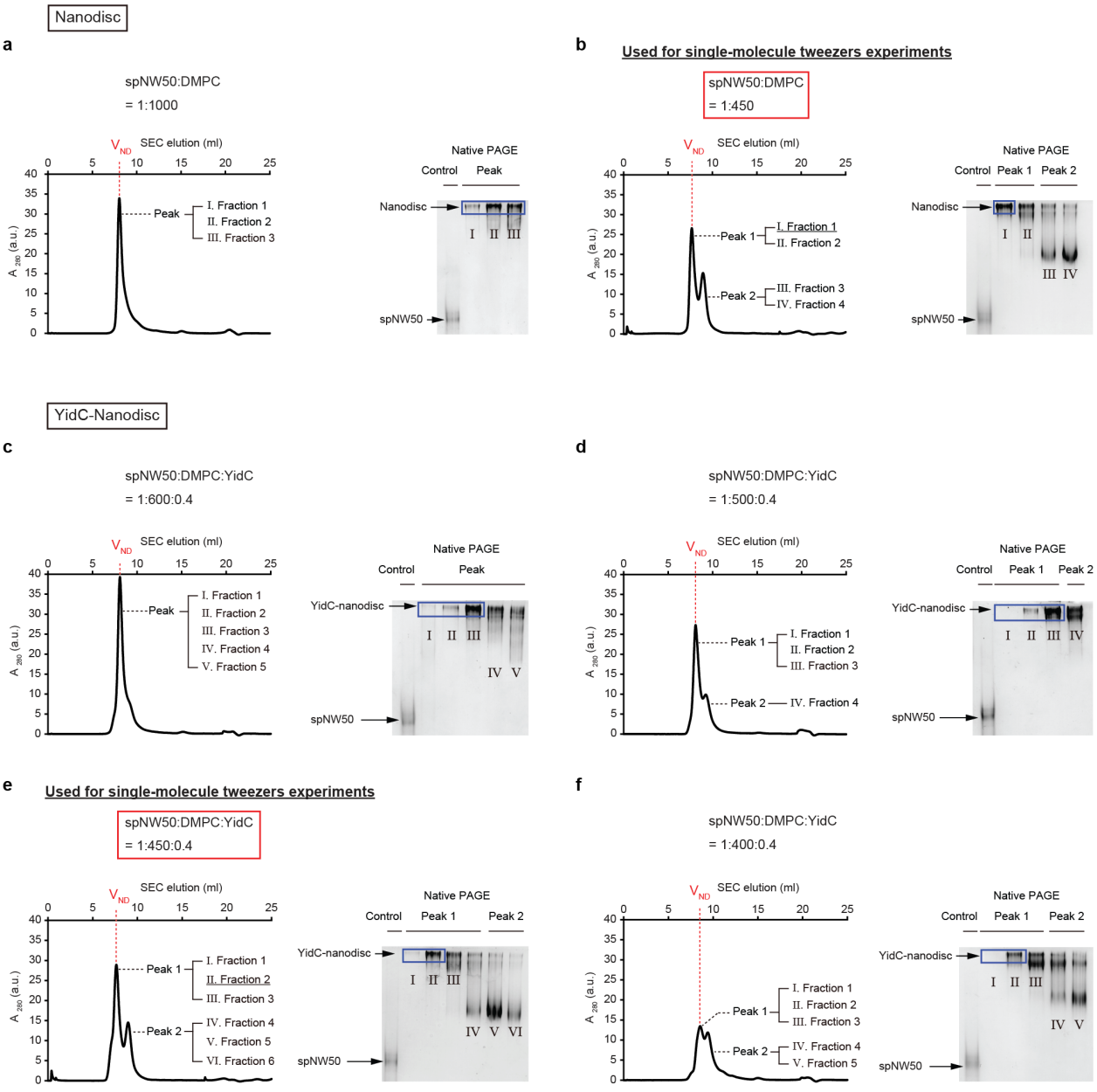
Supplementary Fig. 1 | Amino acid sequences of proteins. SpyTag003 and SnoopTag on GlpG spontaneously bind to SpyCatcher003 and SnoopCatcher, respectively, forming isopeptide bonds. SpyCatcher003 contains a single cysteine used for DNA handle attachment. The peptide linker between SnoopCatcher and AviTag acts as a flexible spacer that offsets the molecular construct from the glass surface. The AviTag enables site-specific biotinylation for anchoring to the glass surface via biotin–traptavidin binding. YidC is incorporated into nanodiscs to construct the YidC-

nanodisc system. The membrane scaffold protein (MSP) spNW50 is spontaneously circularized through isopeptide bond formation between SpyCatcher002 and SpyTag002. The MSP belt encircles lipid molecules to form nanodiscs.



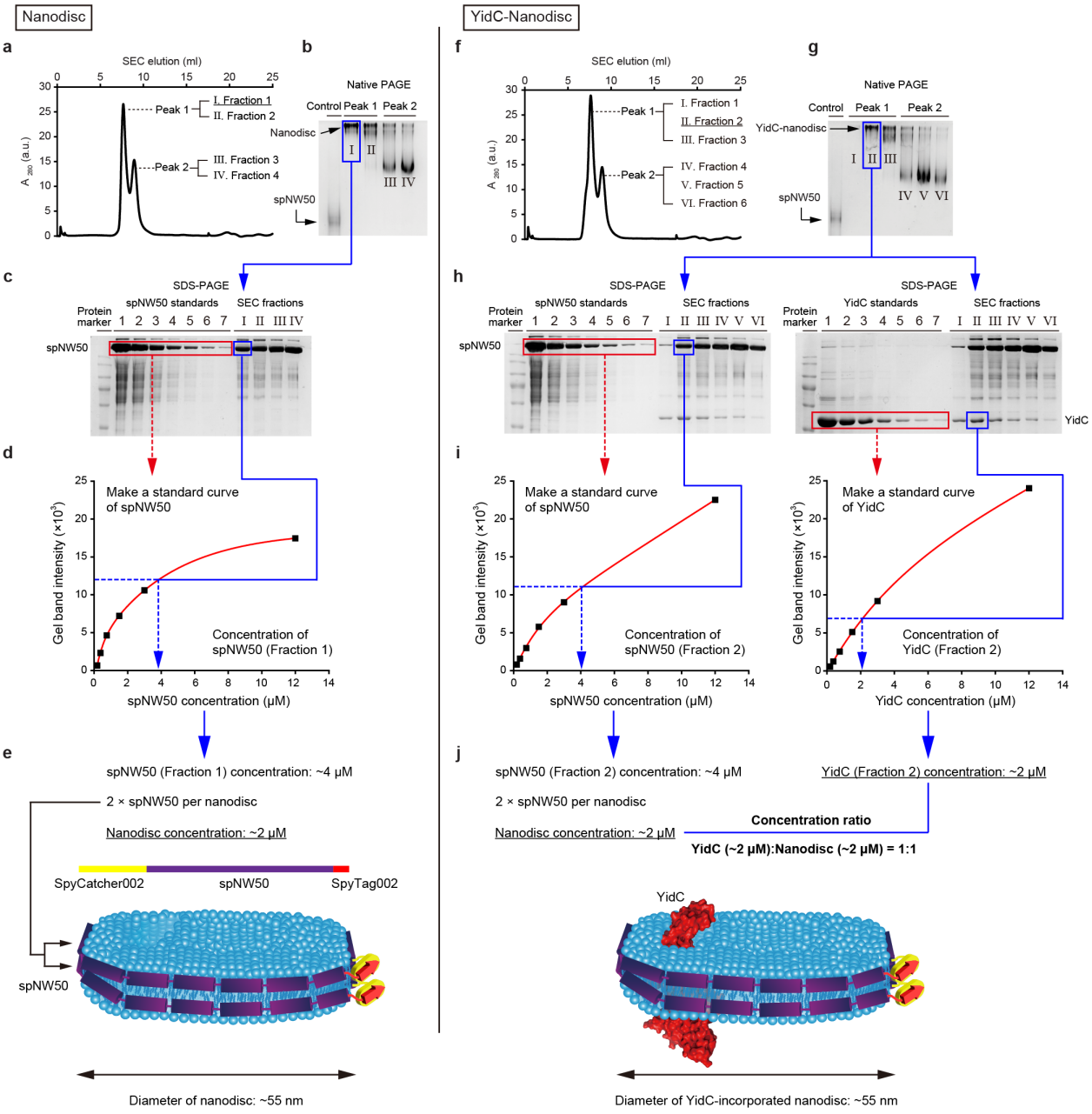
Supplementary Fig. 2 | Protein purification and molecular conjugations. (a) Purification and conjugation of GlpG confirmed by 12% SDS-PAGE. Successful protein conjugations mediated by SpyT003–SpyC003 or SnoopT–SnoopC binding were tested after protein purification. **(b)** Purification of SnoopC and confirmation of biotinylation by a binding test with traptavidin (15% SDS-PAGE). The three gel lanes for SnoopC-bt and for the mixture of SnoopC-bt and traptavidin correspond to different fractions of SnoopC-bt collected after purification using size-exclusion

chromatography (SEC). **(c)** Purification of SpyC003 confirmed by 15% SDS-PAGE. **(d)** Attachment of a 2011-bp DNA handle to SpyC003 confirmed by 6% SDS-PAGE. **(e,f)** Conjugation of GlpG, SnoopC-bt, and the DNA–SpyC003 construct confirmed by 12% SDS-PAGE (e) and 6% SDS-PAGE (f). SnoopC-bt was first attached to GlpG (e), followed by DNA handle attachment (f). Conjugation of GlpG with SnoopC-bt was mediated by SnoopT–SnoopC binding (e), while conjugation with DNA–SpyC003 was mediated by SpyT003–SpyC003 binding (f). **(g,h)** Purification of spNW50 and YidC confirmed by 8% SDS-PAGE (g) and 10% SDS-PAGE (h). Both proteins were purified using Ni-IDA resin and then further purified by SEC. Abbreviations: SnoopC, SnoopCatcher; SnoopC-bt, biotinylated SnoopCatcher; SnoopT, SnoopTag; SpyC003, SpyCatcher003; SpyT003, SpyTag003.



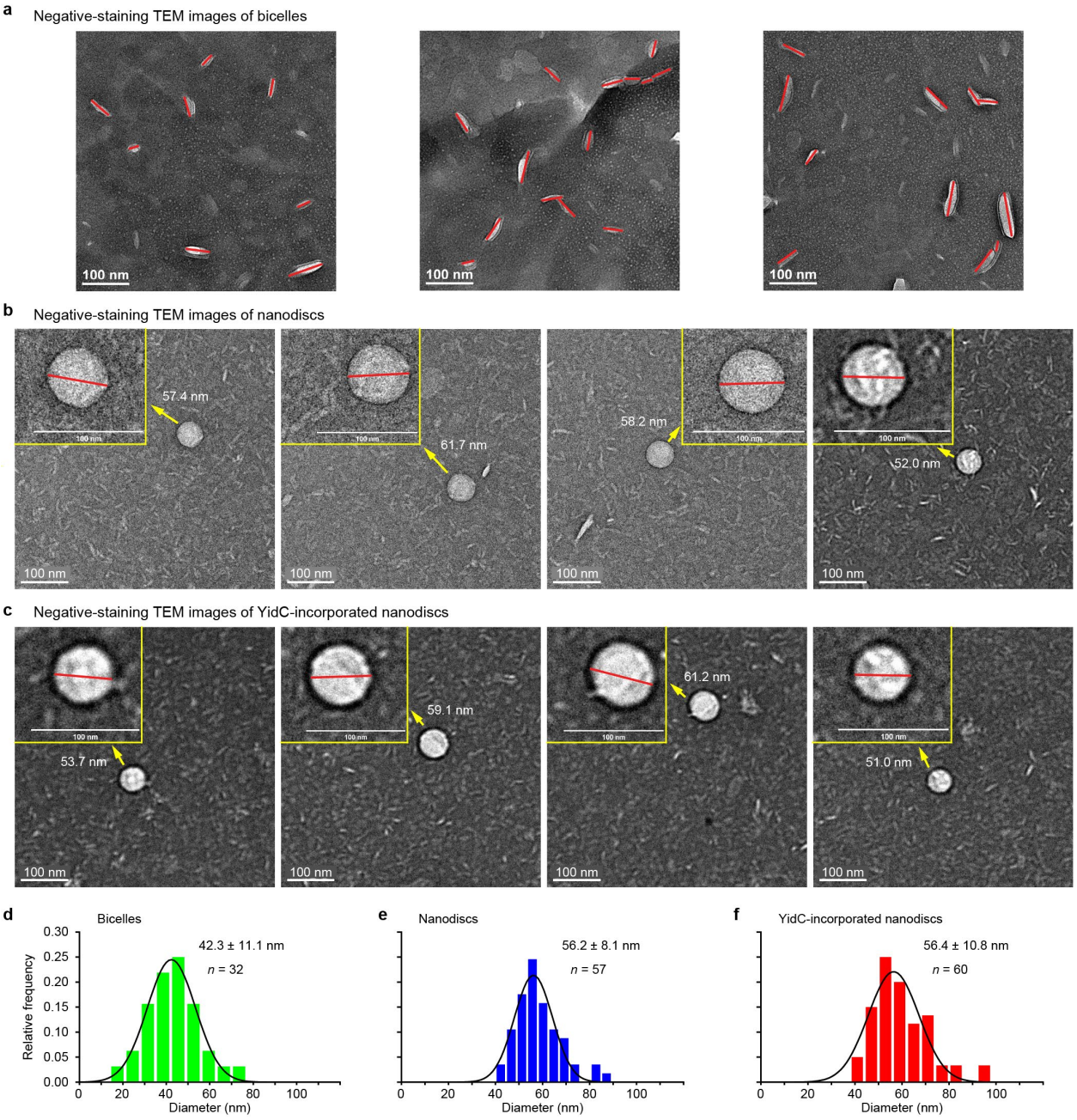
Supplementary Fig. 3 | Purification and optimization of nanodisc samples. (a–f) Purification of nanodiscs and YidC-incorporated nanodiscs prepared with different spNW50:DMPC molar ratios, analyzed by size-exclusion chromatography (SEC; left in each panel) and 6% native PAGE of the SEC elution peak fractions (right in each panel). A_{280} indicates the absorbance at 280 nm. The control nanodisc sample (a), prepared at a 1:1000 spNW50:DMPC molar ratio, showed a single SEC peak, consistent with a previously reported condition yielding monodisperse nanodiscs[1]. The single peak corresponding to properly assembled nanodiscs was eluted at a volume of ~8 ml, denoted V_{ND}, which represents the characteristic elution volume of monodisperse nanodiscs. Nanodisc formation was further confirmed by native PAGE. Fractions from the single

SEC peak showed single bands near the top of the native PAGE gel (outlined in blue); the upward shift of these bands relative to free spNW50 confirms successful nanodisc assembly. Different spNW50:DMPC molar ratios with reduced DMPC content (c–f) were tested to optimize YidC incorporation. Reducing DMPC led to the formation of multiple species, yielding two peaks in the SEC chromatograms. The first peak, containing properly assembled nanodiscs, was eluted at the characteristic elution volume of ~8 ml (V_{ND}), and the corresponding fractions showed single bands (outlined in blue) at positions closely matching those observed for the control nanodisc sample. Among all tested conditions, the nanodisc sample prepared at a 1:450:0.4 spNW50:DMPC:YidC molar ratio (e) yielded preparations containing approximately one YidC molecule per nanodisc (see Methods and Supplementary Fig. 4 for details). The same spNW50:DMPC molar ratio of 1:450 was used to prepare nanodiscs without YidC (b). The underlined fractions of the first elution peak (fraction 1 for nanodiscs and fraction 2 for YidC-incorporated nanodiscs) were further characterized by transmission electron microscopy (TEM) to confirm morphology and size (Supplementary Fig. 5), and were used for single-molecule tweezers experiments.



Supplementary Fig. 4 | Quantification of the YidC:nanodisc stoichiometry. (a) Size-exclusion chromatography (SEC) elution profile of the nanodisc sample. A_{280} indicates the absorbance at 280 nm. **(b)** 6% native PAGE of fractions from the SEC elution peaks confirming successful nanodisc assembly (Supplementary Fig. 3). Elution fraction 1 (outlined in blue), showing a single band near the top of the native gel, was used for single-molecule tweezers experiments. **(c)** 10% SDS-PAGE for quantifying the spNW50 concentration ([spNW50]) in the assembled nanodisc sample. Gel bands outlined in red correspond to spNW50 standards of known concentrations used to generate the standard curve shown in panel d. **(d)** Standard curve for spNW50 used to quantify

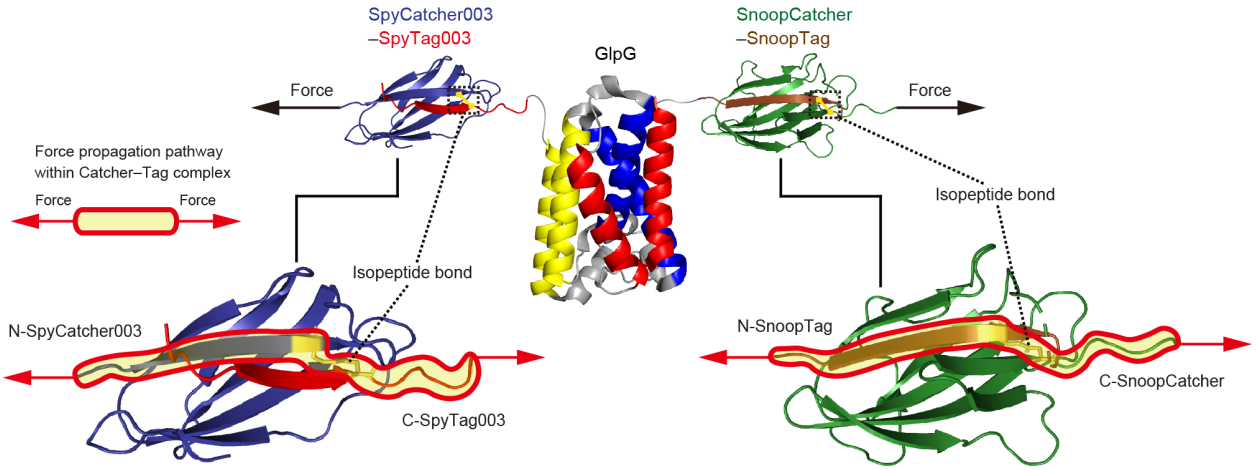
[spNW50] in the assembled nanodisc sample. [spNW50] of elution fraction 1 was interpolated from the standard curve. **(e)** Quantification of the nanodisc concentration. Each nanodisc is formed by two molecules of spNW50 through circularization via SpyTag–SpyCatcher binding. Thus, the nanodisc concentration is estimated as [nanodisc] = [spNW50]/2. **(f)** SEC elution profile of the YidC-incorporated nanodisc sample. **(g)** 6% native PAGE of fractions from the SEC elution peaks confirming successful nanodisc assembly (Supplementary Fig. 3). Elution fraction 2 (outlined in blue), showing a single band near the top of the native gel, was used for single-molecule tweezers experiments. **(h)** 10% SDS-PAGE for quantifying the concentrations of spNW50 (left) and YidC (right). Gel bands outlined in red correspond to spNW50 and YidC standards of known concentrations used to generate the standard curves shown in panel i. **(i)** Standard curves for spNW50 (left) and YidC (right) used to quantify [spNW50] and [YidC] in the assembled nanodisc sample. [spNW50] and [YidC] of elution fraction 2 were interpolated from the standard curves. **(j)** Quantification of the nanodisc concentration and YidC:nanodisc stoichiometry. Each nanodisc is formed by two molecules of spNW50, as described above ([nanodisc] = [spNW50]/2). Based on [nanodisc] obtained here and [YidC] obtained in panel i, the YidC:nanodisc stoichiometry is estimated to be 1:1 for the YidC-incorporated nanodisc sample.



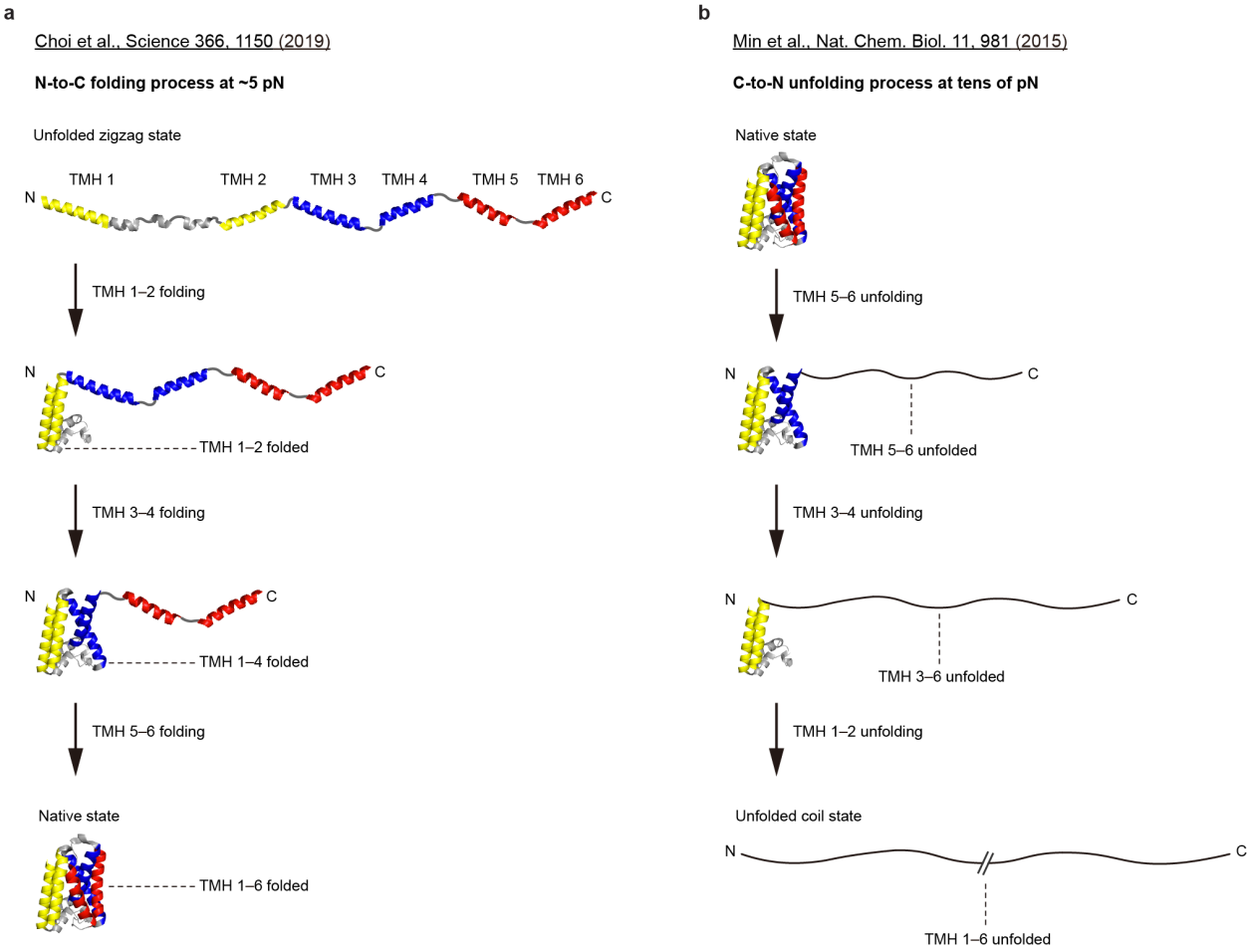
Supplementary Fig. 5 | Size characterization of bicelles and nanodiscs. (**a**–**c**) Negative-staining transmission electron microscopy (TEM) images of 1.3% (w/v) bicelles (a), 500 nM nanodiscs (b), and 500 nM YidC-incorporated nanodiscs (c). Red lines indicate the diameters of the corresponding bicelles and nanodiscs. Insets in panels b and c show magnified views of individual nanodiscs, indicated by yellow arrows. Scale bars in all images indicate 100 nm. (**d**–**f**) Size distributions of bicelles (d), nanodiscs (e), and YidC-incorporated nanodiscs (f) analyzed from the TEM images (n = 32, 57, and 60 data points for bicelles, nanodiscs, and YidC-incorporated

nanodiscs, respectively). The relative frequencies of diameter values were fitted with a Gaussian function, yielding 42.3 ± 11.1 nm (s.d.) for bicelles, 56.2 ± 8.1 nm (s.d.) for nanodiscs, and 56.4 ± 10.8 nm (s.d.) for YidC-incorporated nanodiscs.

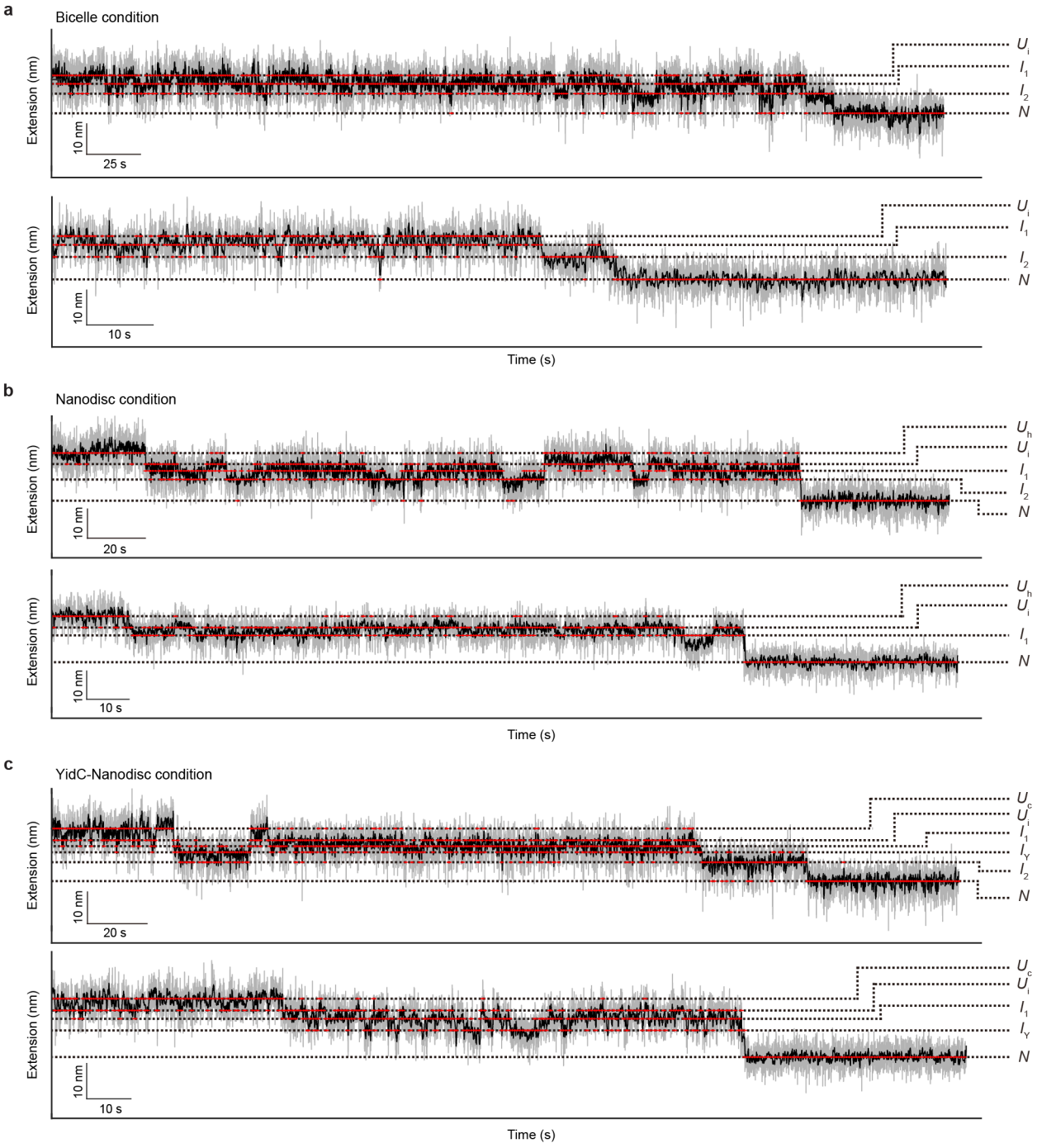
5



Supplementary Fig. 6 | Molecular construct and force propagation. Molecular construct used for single-molecule tweezers experiments. The N- and C-terminal ends of GlpG, genetically tagged with SpyTag003 (red) and SnoopTag (brown), respectively, are covalently linked to SpyCatcher003 (blue) and SnoopCatcher (green) via isopeptide bonds, respectively. Black arrows indicate the force-pulling directions, and dotted lines indicate the internal isopeptide bonds within the Catcher–Tag complexes. Unfolding of the complexes by a pulling force is prevented by these isopeptide bonds, as the force is expected to propagate along the shortest pathway through the covalent isopeptide bonds within the complexes. Force propagation pathways are highlighted by yellow backgrounds with red borderlines and arrows. The N- or C-terminus of the Catcher proteins and peptide Tags is indicated by abbreviations, such as N-SpyCatcher003 and C-SpyTag003.

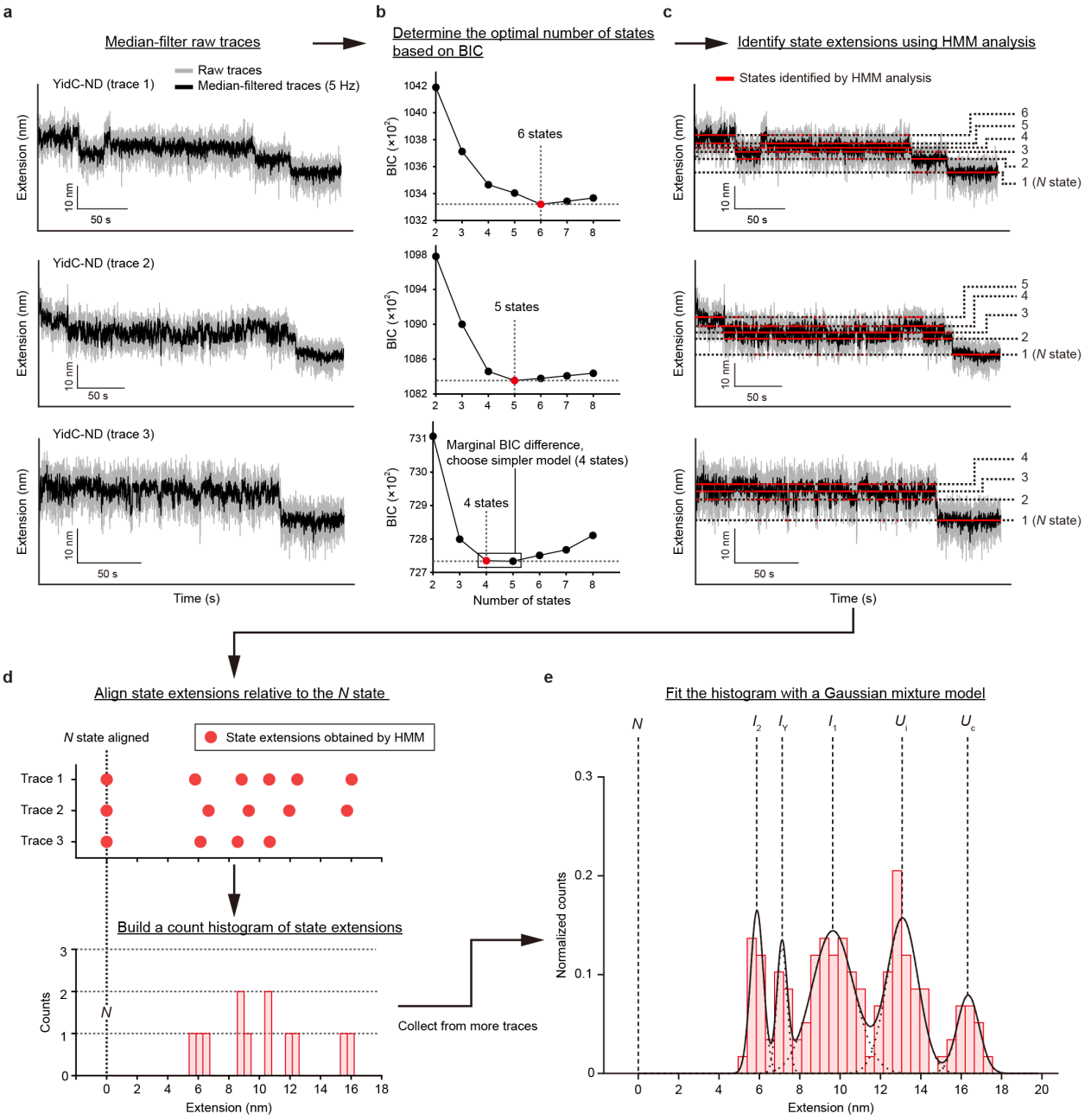


Supplementary Fig. 7 | Folding and unfolding pathways of GlpG. (a) Schematic of the N-to-C folding process of GlpG at low forces such as ~5 pN, as characterized in a previous single-molecule tweezers study[2]. GlpG transitions from the unfolded zigzag state (similar to the unfolded helical state, but with all six transmembrane helices (TMHs) expected to be slightly inserted into the lipid bilayer) to the native state through sequential N-to-C folding of TMH pairs. **(b)** Schematic of the C-to-N unfolding process of GlpG at high forces such as tens of pN, as characterized in a previous single-molecule tweezers study[3]. GlpG transitions from the native state to the unfolded coil state through sequential C-to-N unfolding and unraveling of TMH pairs. Dashed lines indicate the specific TMH segments folded or unfolded at each state. The labels N and C in the figure indicate the N- and C-terminal ends of GlpG, respectively.



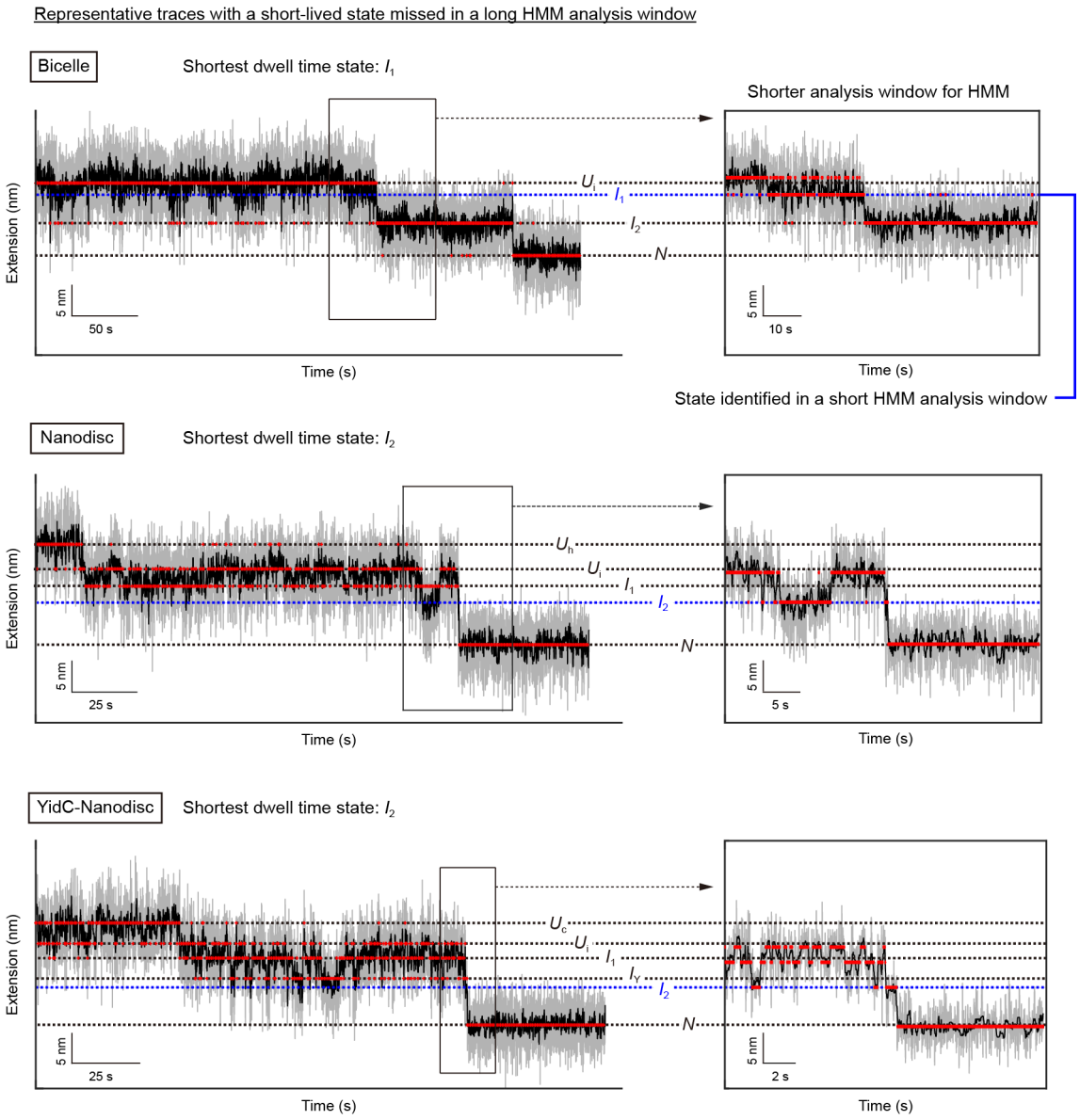
Supplementary Fig. 8 | Representative on-pathway trajectories of GlpG. (**a–c**) Time-resolved extension traces of GlpG at 5 pN under each bilayer condition: bicelles (a), nanodiscs (b), and YidC-incorporated nanodiscs (c). Gray and black traces represent raw data and median-filtered data (5-Hz window), respectively. Red traces represent the states obtained using a hidden Markov model (HMM) analysis with the optimal number of states determined by the Bayesian information

criterion (BIC).

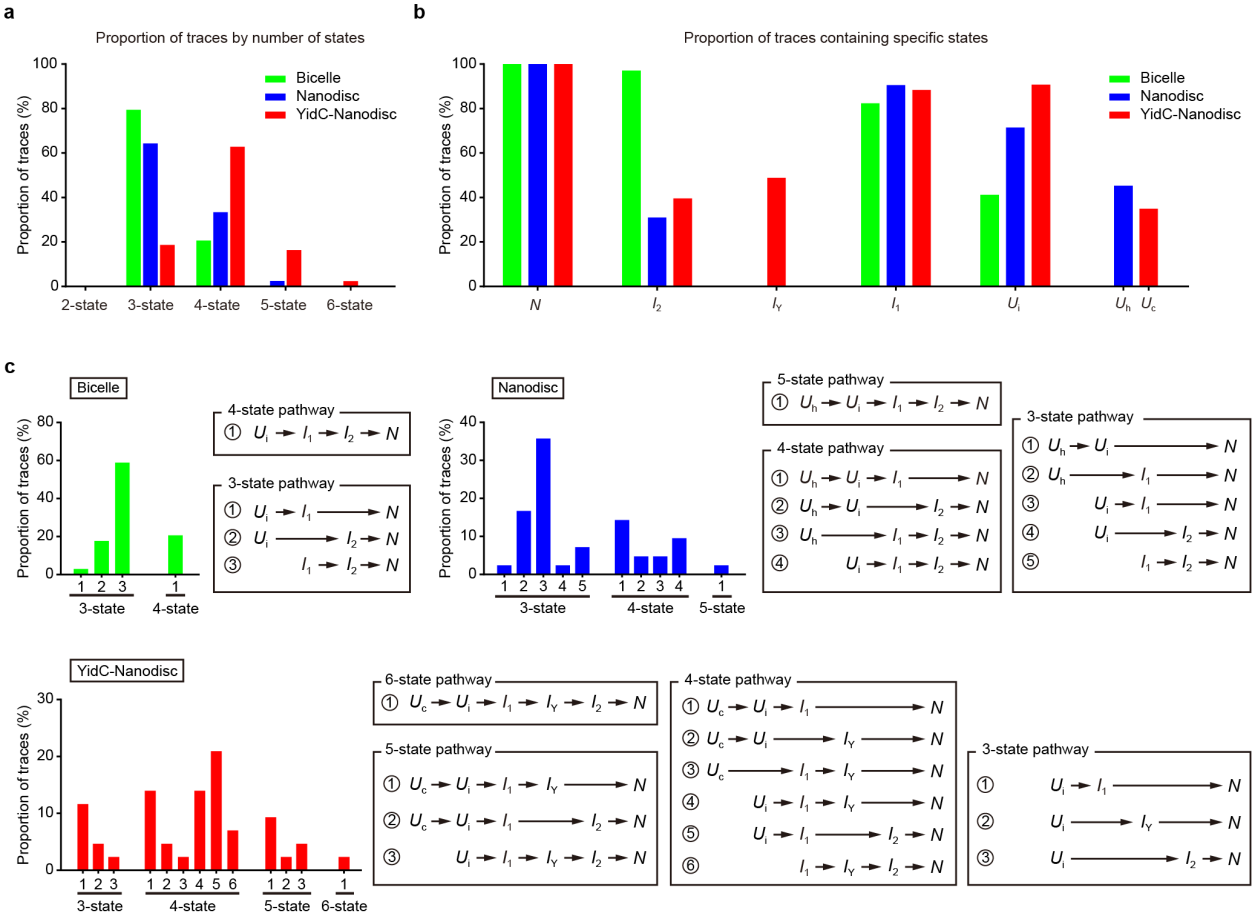


Supplementary Fig. 9 | Analysis workflow for extension state identification. (a) Median filtering of raw time-resolved extension traces at 5 pN with a 5-Hz window. Traces from the YidC-incorporated nanodisc (YidC-ND) system are shown here as an example. **(b)** Determination of the optimal number of states using the Bayesian information criterion (BIC), marked as a red data point. **(c)** Identification of distinct extension states using a hidden Markov model (HMM) analysis with the given number of states. **(d)** Construction of a count histogram of HMM-assigned extension values from multiple traces, relative to the *N* state. **(e)** Fitting of the extension histogram with a Gaussian mixture model (GMM; see Supplementary Fig. 12 for details). Solid and dashed

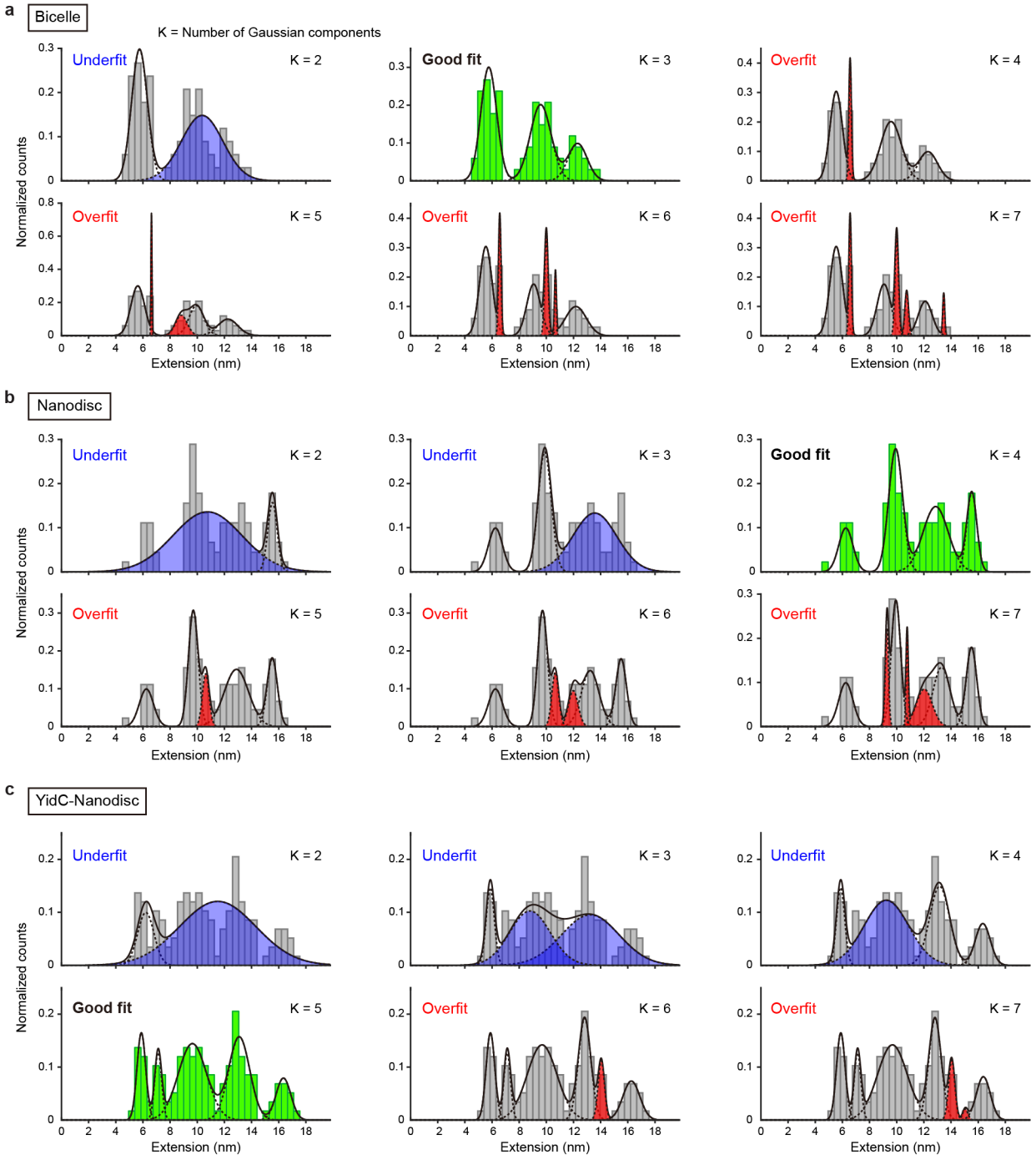
curves represent the GMM fit and the individual Gaussian components, respectively. Each peak of the GMM fit corresponds to a specific state of GlpG during folding in the YidC-ND system, as detailed in the main text.



Supplementary Fig. 10 | Dependence of state identification on the analysis time window. Representative time-resolved extension traces at 5 pN in each bilayer system: bicelles (top), nanodiscs (middle), and YidC-incorporated nanodiscs (bottom). For each system, the left panel displays the trace analyzed using a hidden Markov model (HMM) with a long analysis window, in which a short-lived state is missed. The right panel shows the trace re-analyzed with a shorter HMM analysis window, in which the previously undetected state is identified, indicated by a blue dashed line. Gray, black, and red traces represent raw data, median-filtered data, and states identified using the HMM analysis, respectively.

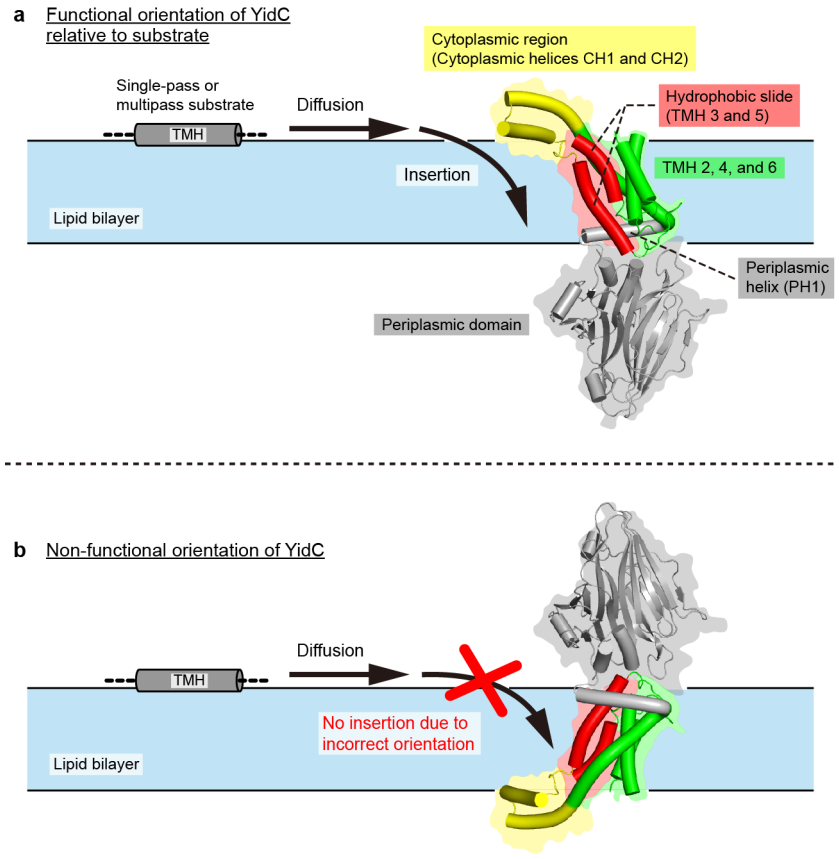


Supplementary Fig. 11 | Statistics for trace populations of GlpG. (**a**) Proportions of time-resolved extension traces (hereafter, traces) at 5 pN, sorted by the number of states in each bilayer system: bicelles (green), nanodiscs (blue), and YidC-incorporated nanodiscs (red). (**b**) Proportions of traces containing a specific state in each bilayer system. (**c**) Proportions of traces showing each folding pathway observed in the analysis in each bilayer system. The variation in observed pathways, listed to the right of each bar plot, reflects the limited detection of relatively short-lived states when a long HMM analysis window is used (see Supplementary Fig. 10 for details). All data are obtained from $n = 34$, 42, and 43 traces across 13, 10, and 13 molecules for bicelles, nanodiscs, and YidC-incorporated nanodiscs, respectively.

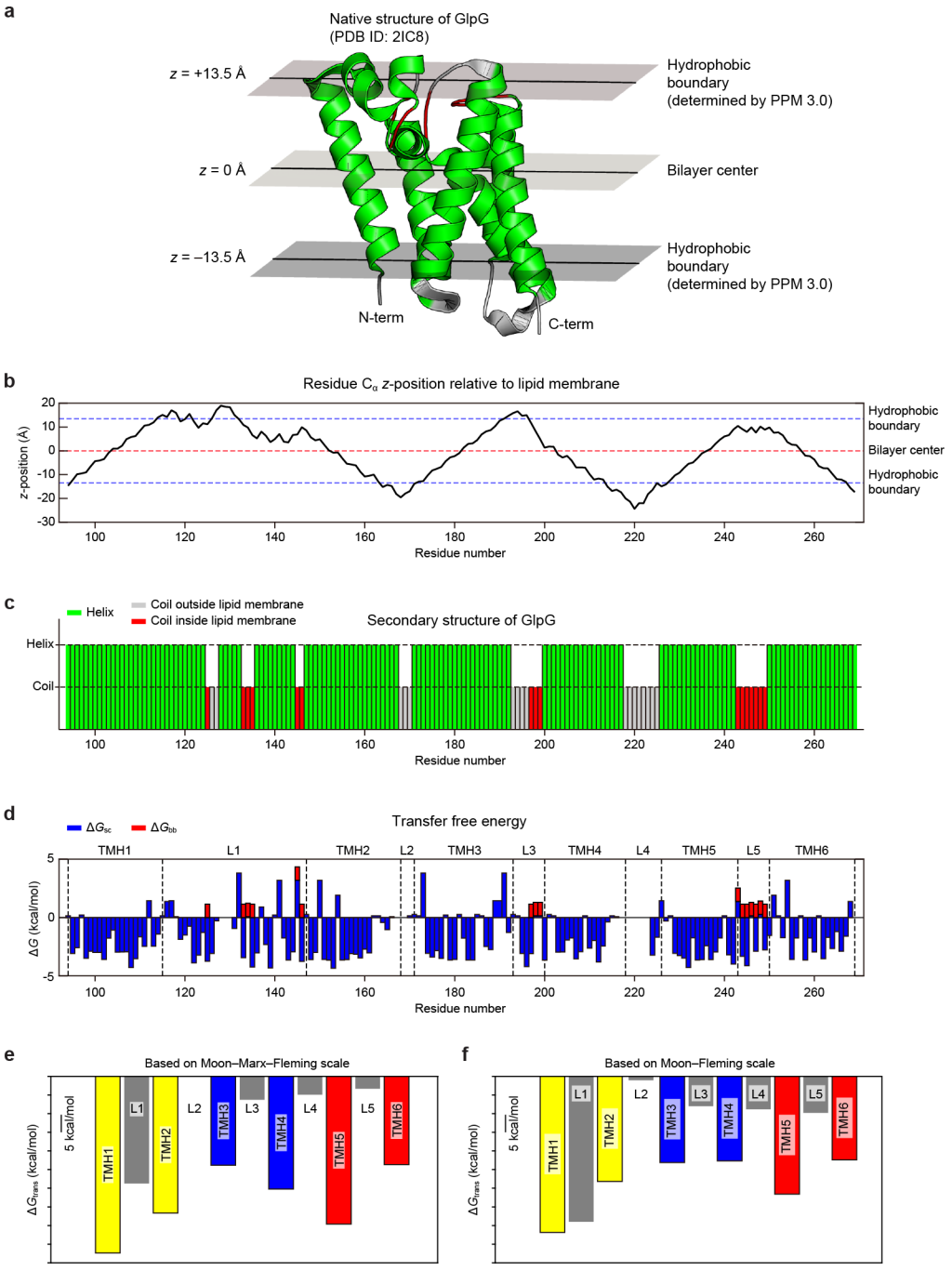


Supplementary Fig. 12 | Identification of the number of states in extension histograms. (**a**–**c**) Normalized count histograms of hidden Markov model (HMM)-assigned extensions from time-resolved traces at 5 pN in each bilayer system: bicelles (a), nanodiscs (b), and YidC-incorporated nanodiscs (c). Each histogram was fitted with a Gaussian mixture model (GMM) using different numbers of Gaussian components (K = 2–7). Black solid and dashed curves indicate the GMM fit

and the individual Gaussian components, respectively. Good GMM fits are shown in green, while underfits and overfits are shown in blue and red, respectively, for the corresponding Gaussian components. The optimal GMMs (i.e., the optimal numbers of states in the extension histograms) correspond to K = 3, 4, and 5 for bicelles, nanodiscs, and YidC-nanodiscs, respectively. All data are obtained from n = 34, 42, and 43 traces across 13, 10, and 13 molecules for bicelles, nanodiscs, and YidC-incorporated nanodiscs, respectively.

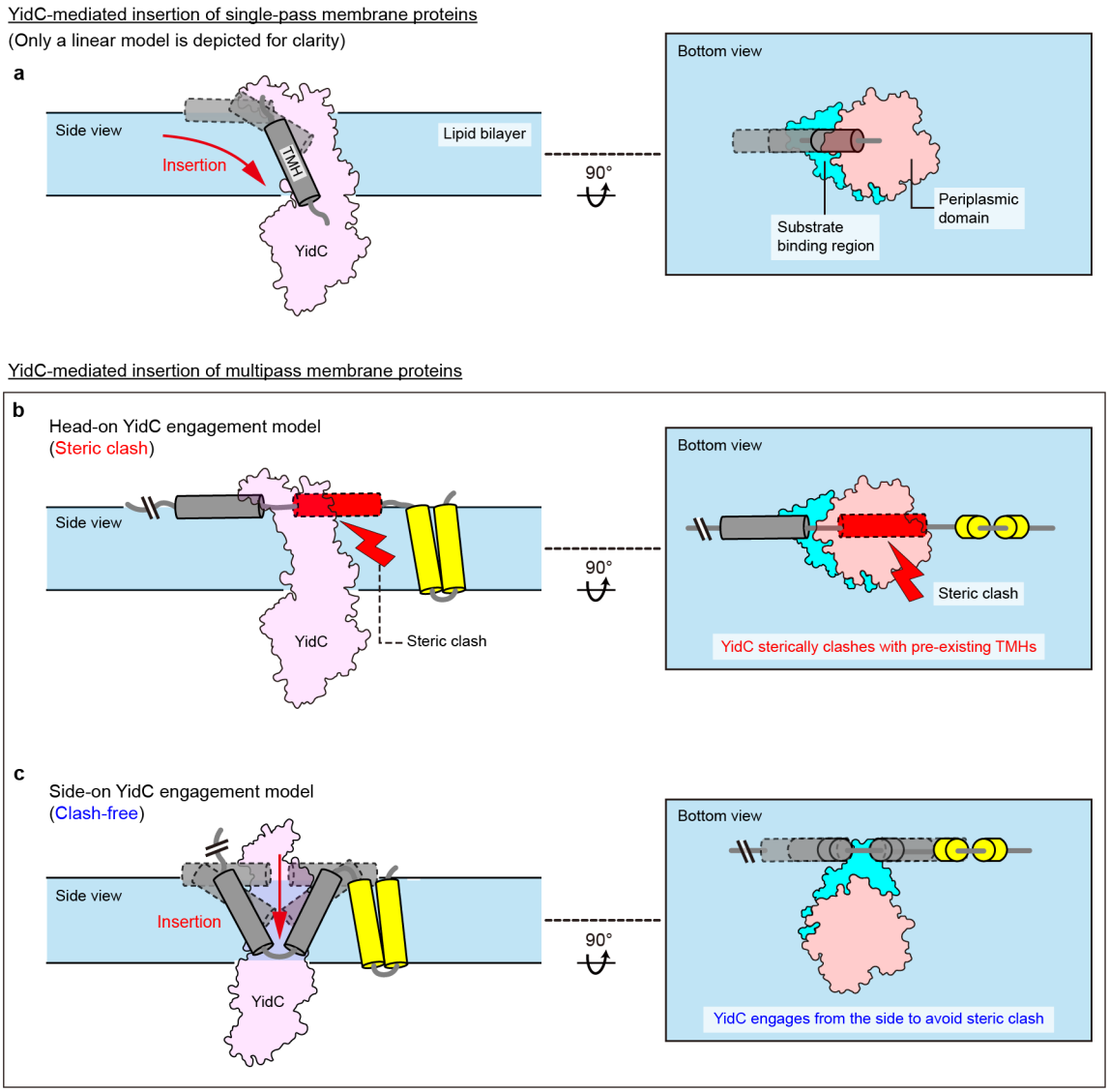


Supplementary Fig. 13 | Functional and non-functional orientations of YidC. (**a**) Functional orientation of YidC relative to the substrate, which facilitates substrate insertion into the lipid bilayer. (**b**) Non-functional, inverted orientation of YidC. Structural elements of YidC are colored as follows: cytoplasmic region CH1/CH2 (yellow); TMHs 3 and 5, which form the hydrophobic slide (red); TMHs 2, 4, and 6 (green); and the periplasmic helix PH1 and periplasmic domain (gray).

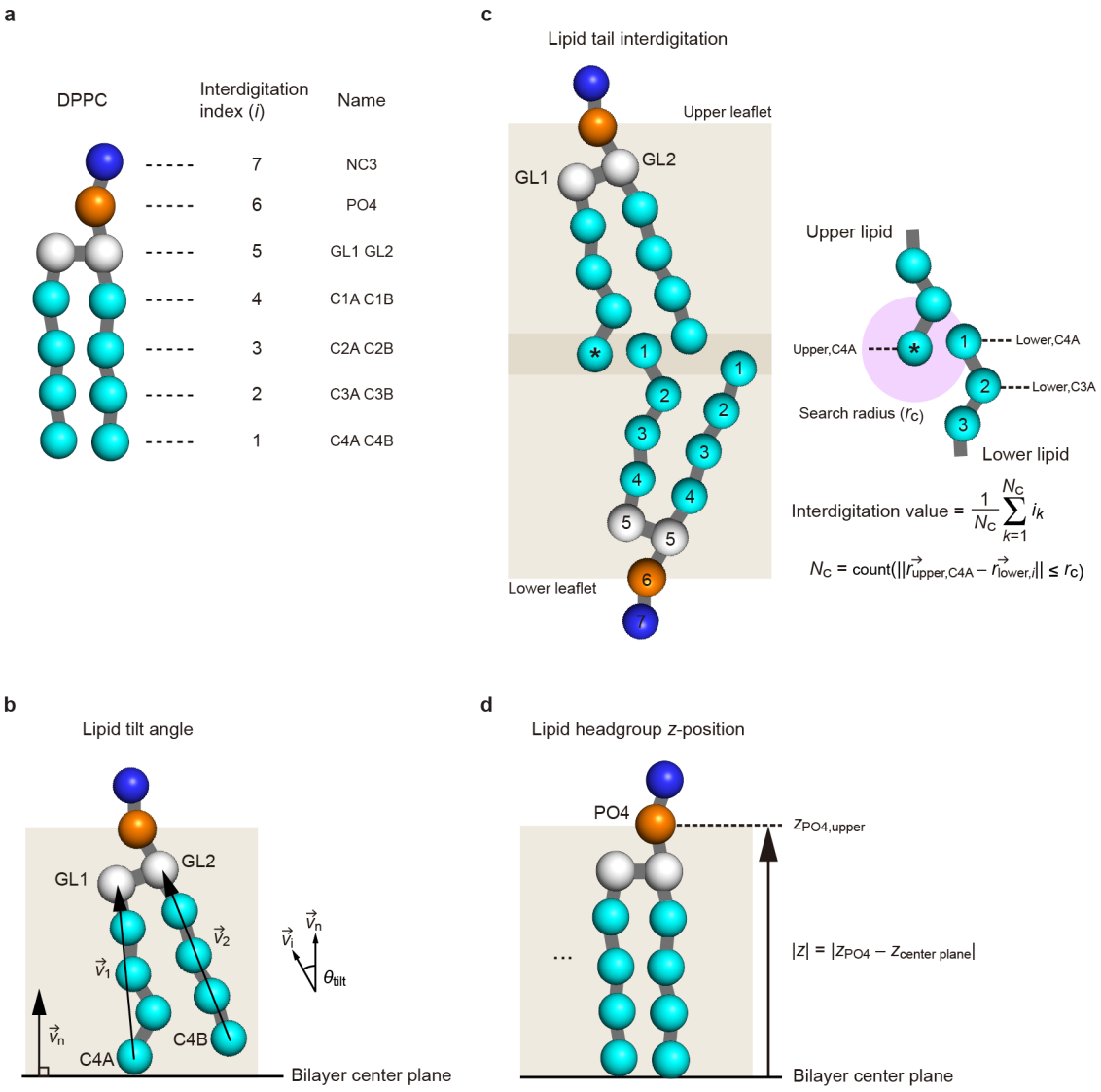


Supplementary Fig. 14 | Water-to-bilayer transfer free energy analysis. (**a**) Structure of GlpG (PDB ID: 2IC8) embedded in a DMPC lipid bilayer, obtained from PPM 3.0. The bilayer center is defined at $z = 0$, and the hydrophobic boundaries are indicated at $z = \pm 13.5$ Å. (**b**) C_α z-positions of residues along the bilayer normal extracted from the PPM-oriented structure. Dashed lines mark the bilayer center and the hydrophobic boundaries. The hydrophobic boundaries correspond to the carbonyl plane of the lipid bilayer. (**c**) Secondary structure assignment across the sequence.

Residues in helix and coil states are shown in green and gray/red, respectively. Coil residues that fall within the hydrophobic core (red) receive a backbone penalty in the water-to-bilayer free energy calculation (see panel d). (**d**) Residue-level water-to-bilayer transfer free energies. The depth-dependent side-chain transfer free energies (ΔG_{sc}) were obtained from the Moon–Marx–Fleming hydrophobicity scale (see Methods for details). To account for the energetic cost of unsatisfied backbone hydrogen bonds within the lipid bilayer, an energetic penalty of 1.15 kcal/mol (ΔG_{bb}, derived from Wimley and White's study; see Methods) was applied to each non-helical residue located within the hydrophobic core. The segments corresponding to transmembrane helices (TMHs) and linker regions (denoted by L) are annotated above the plot. (**e**) Water-to-bilayer transfer free energies for each TMH and linker (ΔG_{trans}) based on the Moon–Marx–Fleming hydrophobicity scale. (**f**) Water-to-bilayer transfer free energies based on the Moon–Fleming hydrophobicity scale without depth dependence.

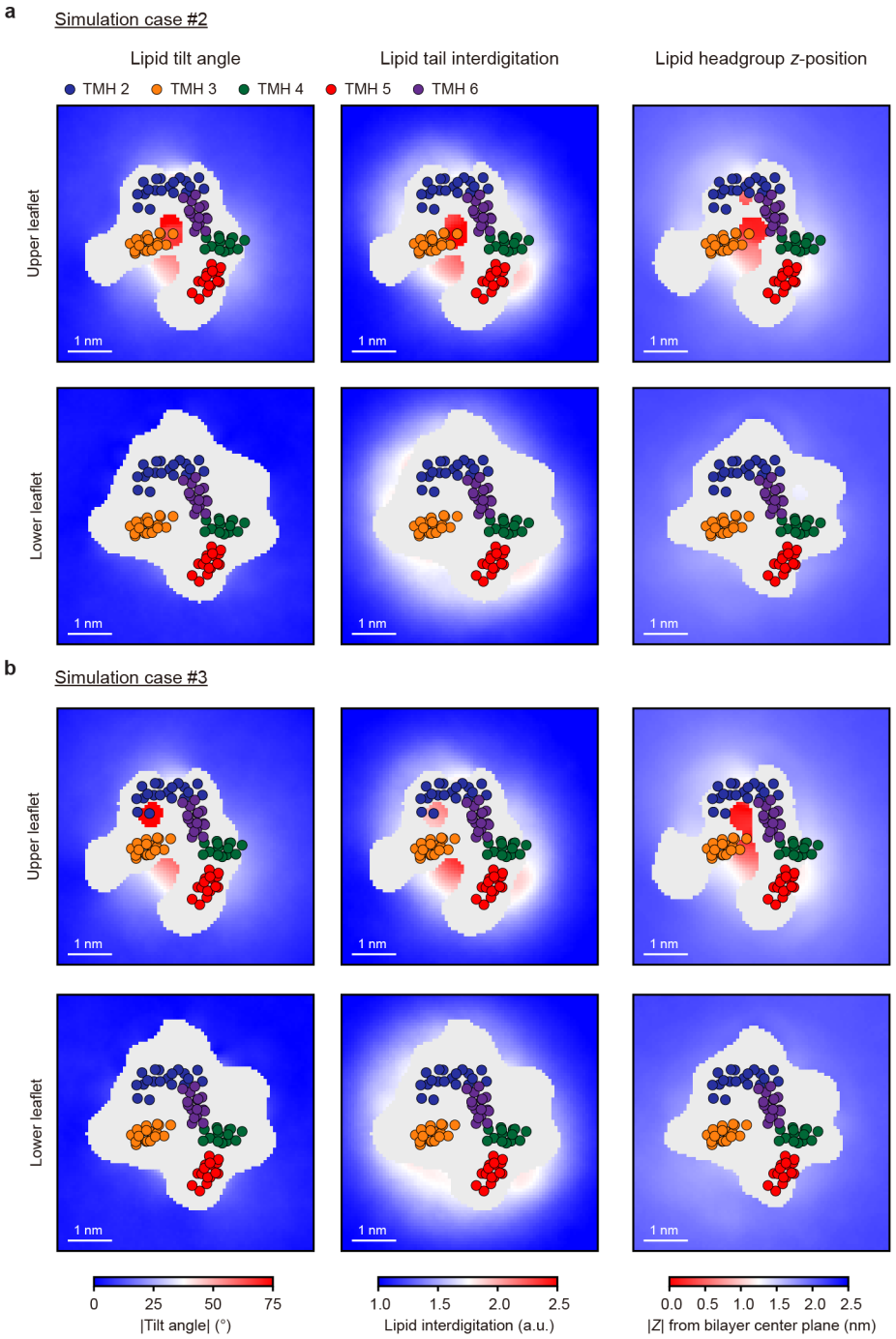


Supplementary Fig. 15 | Models for YidC-mediated membrane protein insertion. (**a**) YidC-mediated insertion model for single-pass membrane proteins (only a linear model is depicted for clarity). In this model, a transmembrane helix (TMH, gray) inserts into the lipid bilayer by sliding head-on along the hydrophilic TMH 3–5 groove of YidC (pink). The left and right panels show the side and bottom views of the YidC-embedded membrane (sky blue). (**b**) Steric clash in head-on YidC engagement model for multipass membrane proteins. If YidC approaches a multipass membrane protein in the same manner as for single-pass membrane proteins, pre-existing TMHs would sterically clash with YidC (indicated by the yellow lightning bolt). (**c**) Proposed side-on YidC engagement model for clash-free insertion. In this model, YidC interacts with a multipass membrane protein from the side and facilitates the insertion of a TMH pair, constrained by the connecting linker. This YidC–substrate orientation bypasses potential steric clashes, making this engagement model highly plausible.



Supplementary Figure 16 | Definitions of physical parameters for membrane deformation. (**a**) Coarse-grained DPPC lipid based on the Martini model. The structure and assigned bead names (e.g., NC3, PO4) are shown on the right. Interdigitation indices (i = 1–7) are assigned as integer values in order of increasing distance from the bilayer center. (**b**) Lipid tilt angle. The tilt angle (θ_{tilt}) for each lipid tail is defined as the angle between the bilayer normal vector ($\boldsymbol{v}_n$) and the lipid vectors ($\boldsymbol{v}_1$ and $\boldsymbol{v}_2$). These lipid vectors are defined from the terminal tail beads (C4A and C4B) to the respective glycerol beads (GL1 and GL2). (**c**) Lipid tail interdigitation. Interdigitating opposing-leaflet beads are identified within a search radius of r_c = 6 Å (purple circle) from a reference terminal tail bead (asterisk; C4A or C4B). The interdigitation value is defined as the mean interdigitation index of the identified opposing-leaflet beads (see the formula in the panel). (**d**) Lipid headgroup vertical position. The vertical distance ($|z|$) is defined as the absolute distance between the phosphate bead (PO4; orange) and the bilayer center plane (black line), i.e., $|z| = |z_{PO4}$

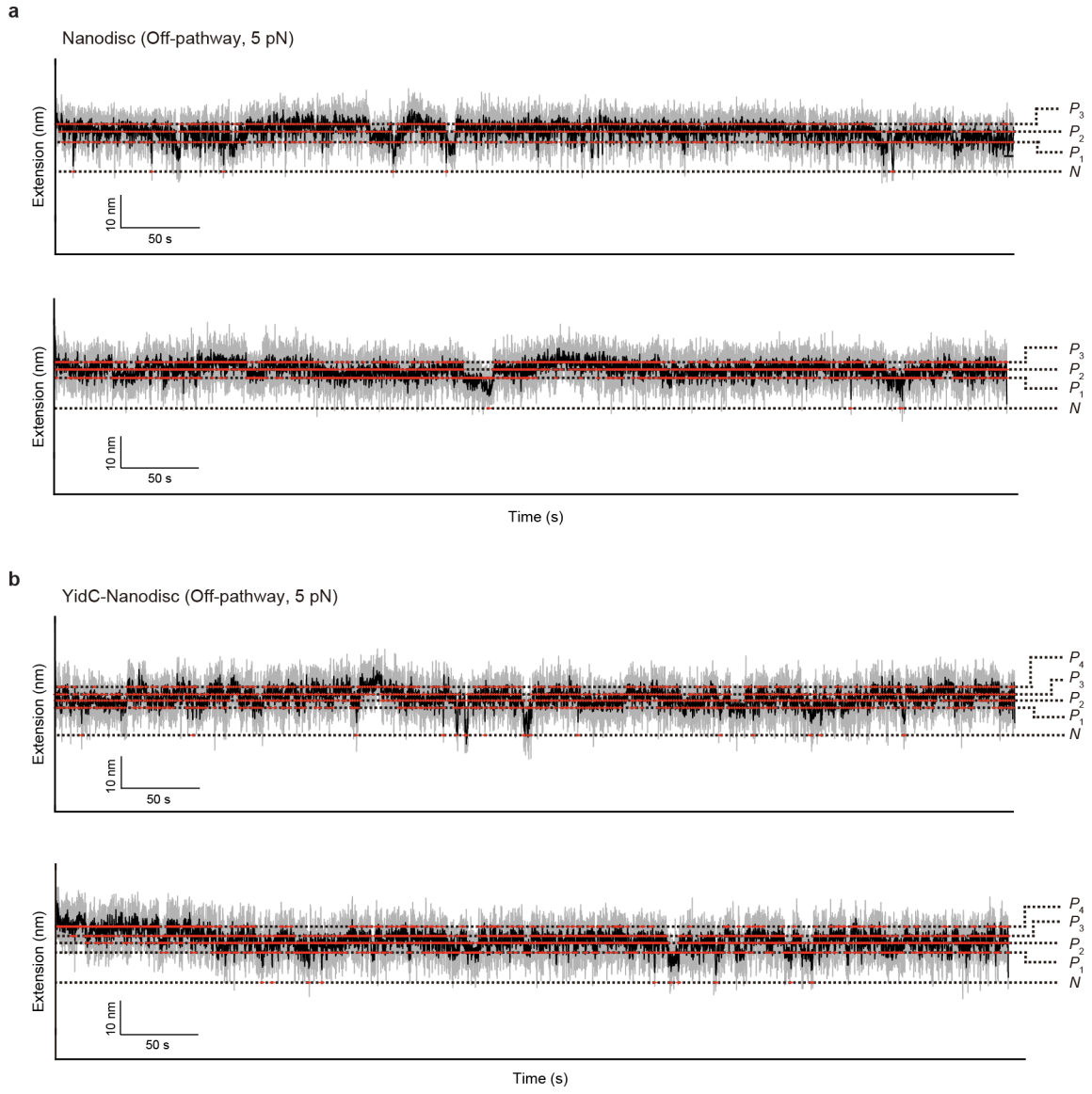
$- z_{\text{center plane}}|$. The bilayer center plane is defined as the averaged z-coordinate of PO4 beads from both leaflets in the bulk membrane region ($r \geq 4$ nm from the protein center).



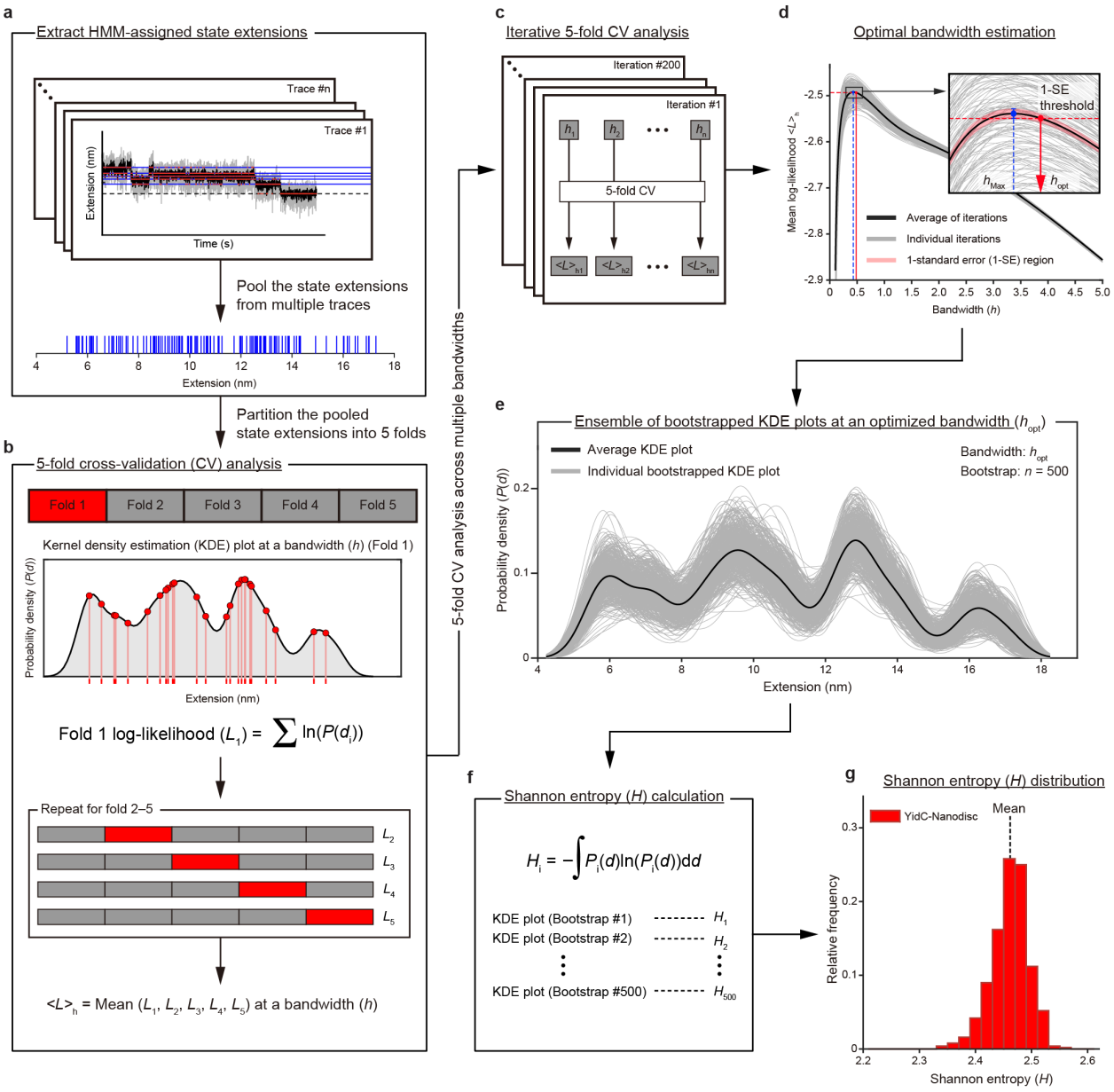
Supplementary Figure 17 | Membrane deformation from independent simulation replicates. (**a**,**b**) 2D maps of local membrane deformation around YidC for simulation case #2 (a) and simulation case #3 (b), analogous to the data for simulation case #1 shown in Fig. 5f–h. Spatial distributions of lipid tilt angle (left), lipid tail interdigitation (middle), and lipid headgroup vertical position (|*z*|; right) are shown for the upper (top) and lower (bottom) leaflets. The TMH 2–6

backbone positions of YidC (circles) are shown as structural reference. YidC-occupied regions (gray) indicate areas where lipid occupancy is absent during the simulations. Scale bars represent 1 nm.

5



Supplementary Fig. 18 | Representative off-pathway trajectories of GlpG. (**a**,**b**) Time-resolved extension traces of GlpG at 5 pN under each bilayer condition: nanodiscs (a) and YidC-incorporated nanodiscs (b). Gray and black traces represent raw data and median-filtered data (5-Hz window), respectively. Red traces represent the states obtained using a hidden Markov model (HMM) with the optimal number of states determined by the Bayesian information criterion (BIC).



Supplementary Fig. 19 | Workflow for Shannon entropy analysis. (**a**) Extraction of hidden Markov model (HMM)-assigned state extensions across multiple traces to construct a pooled state extension dataset. (**b**) 5-fold cross-validation (CV) analysis for optimization of the kernel density bandwidth in constructing the probability density of extension (*P*) using Gaussian kernel density estimation (KDE). The extension dataset was partitioned into five subsets, and the log-likelihood (*L*) for one subset (red) was calculated from the extension probability density at a given bandwidth (*h*). This procedure was repeated across all five subsets to calculate the mean log-likelihood for a given bandwidth. (**c**) 200-iterative rounds of random partitioning of the dataset and 5-fold CV analysis across multiple bandwidths. (**d**) Identification of the optimal bandwidth (h_{opt}) based on the one-standard-error (1-SE) rule, as denoted in the inset. (**e**) Ensemble of bootstrapped KDE

probability density curves (n = 500, gray) at h_{opt}, overlaid with the average KDE curve (black). (**f**) Calculation of the Shannon entropy (H) value for each bootstrap iteration. (**g**) Histogram of Shannon entropy values for the probability density of extension (n = 500 bootstrap iterations). See Methods for details.

Supplementary Table 1 | Molecular dynamics simulation systems. This table summarizes the simulation systems, including their molecular components and system dimensions.

	YidC-bilayer system	bilayer-only system
Protein (YidC)	1 molecule	0 molecule
Upper leaflet (DPPC)	142 molecules	159 molecules
Lower leaflet (DPPC)	148 molecules	159 molecules
Water (PW)	7911 molecules	7702 molecules
Na^+	93 ions	90 ions
Cl^-	95 ions	90 ions
Total number of beads	12,756 beads	11,698 beads
System size (x×y×z)	10.0×10.0×14.9 nm^3	10.0×10.0×14.1 nm^3
Simulation time	10.0 μs	10.0 μs
Independent replicates	3 cases	1 case

References

1. Zhang, S. et al. One-step construction of circularized nanodiscs using SpyCatcher-SpyTag. *Nat Commun* **12**, 5451 (2021).
2. Choi, H.K. et al. Watching helical membrane proteins fold reveals a common N-to-C-terminal folding pathway. *Science* **366**, 1150-1156 (2019).
3. Min, D., Jefferson, R.E., Bowie, J.U. & Yoon, T.Y. Mapping the energy landscape for second-stage folding of a single membrane protein. *Nat Chem Biol* **11**, 981-7 (2015).