Biophysical Journal
Article

Biophysical Society

# On the Interpretation of Force-Induced Unfolding Studies of Membrane Proteins Using Fast Simulations

Zongan Wang,[1,2] John M. Jumper,[1,2] Karl F. Freed,[1,*] and Tobin R. Sosnick[2,3,*]
[1]Department of Chemistry, James Franck Institute, [2]Department of Biochemistry and Molecular Biology, and [3]Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois

ABSTRACT Single-molecule force spectroscopy has proven extremely beneficial in elucidating folding pathways for membrane proteins. Here, we simulate these measurements, conducting hundreds of unfolding trajectories using our fast *Upside* algorithm for slow enough speeds to reproduce key experimental features that may be missed using all-atom methods. The speed also enables us to determine the logarithmic dependence of pulling velocities on the rupture levels to better compare to experimental values. For simulations of atomic force microscope measurements in which force is applied vertically to the C-terminus of bacteriorhodopsin, we reproduce the major experimental features including even the back-and-forth unfolding of single helical turns. When pulling laterally on GlpG to mimic the experiment, we observe quite different behavior depending on the stiffness of the spring. With a soft spring, as used in the experimental studies with magnetic tweezers, the force remains nearly constant after the initial unfolding event, and a few pathways and a high degree of cooperativity are observed in both the experiment and simulation. With a stiff spring, however, the force drops to near zero after each major unfolding event, and numerous intermediates are observed along a wide variety of pathways. Hence, the mode of force application significantly alters the perception of the folding landscape, including the number of intermediates and the degree of folding cooperativity, important issues that should be considered when designing experiments and interpreting unfolding data.

SIGNIFICANCE Single-molecule force spectroscopy, such as atomic force spectroscopy and magnetic tweezers, has proven extremely beneficial in elucidating folding pathways for membrane and other proteins, especially in detecting sparsely populated intermediates. These measurements are challenging from both experimental and computational standpoints, partly because of the demanding computation resources required. Here, we introduce a fast computational tool that can accurately replicate forced unfolding measurements of membrane proteins in a variety of different experimental modes, with a resolution comparable to the highest resolution measurements. We find that the mode of applying force can greatly alter the perception of the folding landscape, an important issue that should be considered when designing experiments and interpreting unfolding data.

## INTRODUCTION

Single-molecule force spectroscopy (SMFS) is a powerful tool to investigate the dynamics of biomolecules. The ever-expanding repertoire of single-molecule manipulation techniques includes atomic force microscope (AFM), optical tweezers, and magnetic tweezers (MT) (1). These methods have been proven beneficial in detecting sparsely populated intermediates and elucidating unfolding pathways of soluble (2–5) and membrane proteins (6–14). Simulations and theory have aided the experimental SMFS studies by revealing the complexity of the folding process (15–21). These measurements are challenging from the computational standpoint, in part because of the demanding computation resources required to simulate the experimental timescales (17). Coarse-grained models enable more extensive sampling and allow for slower, more realistic pulling velocities and lower forces (20) that better match the experimental studies in providing increased likelihood of observing transient intermediates.

---

*Submitted March 2, 2019, and accepted for publication September 12, 2019.*

*Correspondence: freed@uchicago.edu or trsosnic@uchicago.edu

John M. Jumper's present address is DeepMind Technologies, London, United Kingdom.

Editor: Thomas Perkins.

https://doi.org/10.1016/j.bpj.2019.09.011

A major challenge in coarse graining is to establish the right balance between accelerating the simulations and retaining the critical features of the system. We have addressed this challenge with our new *Upside* model, which can fold de novo proteins shorter than 100 residues in cpu-hours (22,23). To this physics-based model, which uses six atoms per residue and realistic Ramachandran maps, we have incorporated a new knowledge-based membrane burial potential that accounts for the changing level of side-chain exposure to lipids (thereby correcting for the replacement of lipid-protein interactions by protein-protein interactions as helices approach) (24). The membrane burial potential also includes unfavorable energies for unsatisfied H-bond donors and acceptors in the membrane, which allows helices to fold and unfold within the bilayer during the simulations.

Here, we conduct numerous simulations mimicking two high-resolution SMFS unfolding experimental studies. The first is an AFM study of bacteriorhodopsin (bR) (11), a light-driven proton pump with seven transmembrane (TM) helices (25). The second is an MT study of GlpG, a rhomboid intramembrane protease from *Escherichia coli* with six TM helices (9). We perform hundreds of trajectories to test our ability to reproduce the experimental data and then investigate the effects of pulling under various protocols, including different spring stiffness and operating in constant velocity or constant force mode (Fig. 1). The logarithmic dependence of pulling velocities on the observed rupture forces is also examined to compare our values to those obtained in experiment, which are conducted at lower pulling velocities, as well as to obtain the distance to the transition state (6,26,27).

Simulations of bR unfolding provide an example where force is applied normal to the bilayer on the C-terminus. The simulations largely match the experimental AFM data; for example, many of the same unfolding intermediates are observed (11). The trajectories for the unfolding of the monomeric bR are similar, but not identical to those for the unfolding of the trimeric form. In the MT study of GlpG, in which force is applied parallel to the bilayer to match the experimental geometry, a principal component

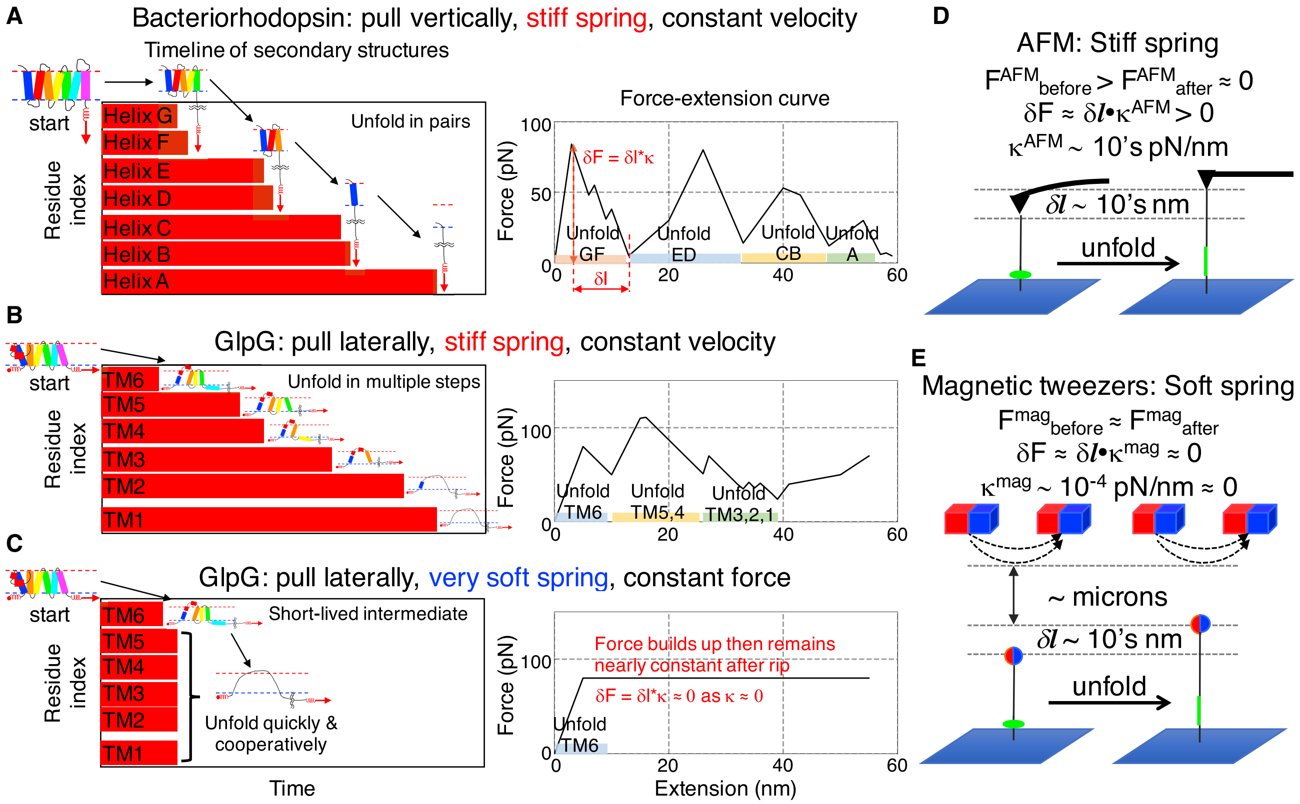

FIGURE 1 Idealized forced unfolding trajectories for bR and GlpG highlighting how different pulling protocols influence the observation of intermediates. The virtual springs (in *red*) exert force by moving perpendicular (bR) or parallel (GlpG) to the bilayer surface at a constant velocity. (*A* and *B*) After a region unfolds in either pulling geometry, a significant drop in force, δF, occurs with a stiff spring as the unfolded chain length increases by an amount δl. This rapid drop increases the probability that another intermediate will be observed. (*C*) However, with a very soft spring, force is maintained as $\delta\text{Force} = -\delta l \times \kappa^{\text{mag}} \approx 0$, and the probability of observing another intermediate is lowered. The horizontal red bars represent the unfolding of helical structures versus time and illustrate the timeline for the secondary structures (TSS). For simplicity, GlpG's two small interfacial helices are not shown. (*D* and *E*) The experimental setup with the AFM and MT is shown. Whereas AFM studies often use a stiff spring ($10 \sim 10^5$ pN/nm), MT studies intrinsically employ a soft spring ($\sim 10^{-4}$ pN/nm) because of the weak spatial dependence of the magnetic field, which varies on the scale of microns (1). Because the force is nearly unchanged after the unfolding event, MT effectively operate in a constant force mode. To see this figure in color, go online.

analysis (PCA) (28) identifies that unfolding occurs stepwise from either end or the middle of the protein when the same spring stiffness is used as in the AFM simulations of bR. This behavior is in apparent disagreement with the experimental study in which few intermediates were found and only the C- to N-terminal unfolding route was observed (9). However, when we better mimic the MT experiment by using a softer spring, the experimental findings are reproduced. This strong dependence on the spring constant highlights that the details of force application can significantly alter the perception of the folding landscape.

## METHODS

### Structure and sequence of bR and GlpG

The bR structure (Protein Data bank [PDB]: 1qhj) and orientation within the lipid bilayer was obtained from the Orientations of Proteins in Membranes (OPM) database (29). The membrane thickness were set to 30.0 nm as identified by OPM. Truncated versions of bR (used in the calibration of the contour length per amino acid) were made from the native structure of bR, and are in the same orientations as in the native bR.

The native structure and orientation within the lipid bilayer of GlpG (PDB: 2xov) were taken from the OPM database (29). The membrane thickness was set to 28.8 nm as predicted by OPM. Two GlpG mutants were made from the native structure using Swiss-Pdbviewer (30): E166A and G261V.

### *Upside* algorithm modified for application of force

*Upside* is a non-Gō, physics-based model where each residue is represented with five atoms (N, C$\alpha$, C, H, O) and a side-chain bead, and with the polypeptide backbone having a geometry given by residue- and neighbor-dependent Ramachandran maps (31). The energy function includes H-bonds, side chain/side chain and side chain/backbone interactions (including helix capping), and a solvation term. The energy function is trained using contrastive divergence. The side chains are represented by multiposition, amino acid-, and directional-dependent beads. Their positional probabilities are given by the probability distribution having the lowest global free energy for all side chains (minimize $G = E - TS$). The use of an instantly equilibrated side-chain probability distribution (i.e., repacking) calculated at every molecular dynamics (MD) step is, to our knowledge, novel and greatly smooths the energy surface and enables *Upside* simulations to be extremely fast. Force is applied to the chosen C$\alpha$ based on the virtual cantilever's spring constant ($\kappa$) and position, which moves at a velocity $v$. The applied force is computed as $-\kappa \times$ (tip position – C$\alpha$ position).

In terms of the *Upside* program, the force function is implemented in the AFMPotential class in *Upside* (in bond.cpp). The tension function used in the force clamp simulation is implemented in the TensionPotential class in bond.cpp. Its implementation is simpler because there is no need to estimate the time and store the tip position and residue position.

### PCA of unfolding trajectories of GlpG

The programs MDTraj (32) and scikit-learn (33) were used to perform the PCA using the C$\alpha$-C$\alpha$ distances below 8 Å in the native state to define contacts. Structures from all trajectories under the same set of simulation conditions were included in the PCA. To derive the principal components (PCs), we used the C$\alpha$-C$\alpha$ distances obtained at $T = 1.0$ ($\sim$300 K), spring constant $\kappa = 0.05\ k_B T/\text{Å}^2$ ($\sim$21 pN/nm at 300 K), and a pulling velocity $v = 0.001$ Å/*Upside* time step ($\sim 10^6$ nm/s). These PCs are used for the projection at the other conditions for comparison purposes.

## RESULTS

### Testing of our implementation of force in MD simulations

We first conducted a variety of tests of the new capability of our *Upside* algorithm (22–24) (Supporting Materials and Methods) to describe force unfolding. The calibration of the force levels is verified by evaluating the thermal fluctuations (in the $z$ direction) of the tip of the virtual cantilever when attached to different masses. These fluctuations are found to be proportional to the spring constant $\kappa$ as required by the equipartition theorem: $k_B T = \kappa \langle z^2 \rangle$ (Fig. S1; (34)). In addition, the unfolded segments are well described by a worm-like chain (WLC) model (Figs. S2 and S3 *E*; (35)). This behavior provides evidence that our pulling rate of 0.001 Å/*Upside* time step is slow enough that the unfolded polypeptide has enough time to sample conformational space and behave as an entropic spring.

To further examine our force calibration and *Upside*'s ability to replicate all-atom simulations, we also compared our simulations to those observed in all-atom, explicit solvent studies of already unfolded ubiquitin molecules (Fig. S3; (18)). *Upside* closely matches the force-extension curve (FEC) and contour length $Lc$ of the all-atom simulations. However, the Ramachandran maps of the ($\phi$, $\psi$) backbone dihedral angles are moderately different over the force range 30–100 pN. Our maps have more $\beta$-angles and less polyproline 2 and helical-like angles. Our dihedral potentials are obtained from a coil library (31), which can be used to predict local chain behavior (e.g., NMR residual dipolar couplings (36)), whereas the maps seen in all-atom simulations are known to depend strongly on the particular force field (37). Overall, these tests provide confidence for us to proceed with the study of forced-induced unfolding of membrane proteins.

### AFM simulations of bR unfolding

In the experimental AFM study, bR molecules are arranged in a lattice of closely packed trimers embedded in purple membranes, and monomers are removed one by one (11). In our simulations, we also pull on single bR molecules, both as monomers and as trimers in a bilayer. Compared to the experiment, our simulations of trimers lack the protein-protein interactions between adjacent trimers found in the purple membranes; these interactions are replaced with protein-lipid interactions in the simulations. The molecules are located within an implicit membrane bilayer modeled using our new membrane burial potential (24) with force being applied with a virtual spring attached at bR's C-terminus. Fig. 1 *A* depicts the experimental setup and introduces the plot of the timeline of secondary structure (TSS), a convenient pictorial representation of the folding simulations that highlights the time evolution of helix unfolding (22). Here, each solid horizontal red bar denotes the presence of a folded helix, with the height reflecting the extent of the folded portion in the helix. Fig. 2 *A* presents an actual trajectory produced by our simulations.

The force is increased by moving the spring vertically at a constant velocity normal to the bilayer (*z* direction) (Figs. 1 *A* and 2 *A*; Video S1 *A*). Force typically accumulates to ~100 pN before becoming an unfolding event occurs, whereupon a newly unfolded portion of the protein extends in the *z* direction. As a result, the spring returns toward its equilibrium position and the force drops. The magnitude of the drop, δF, is proportional both to δl, the length of the newly unfolded segment, and κ, the stiffness of the cantilever: δForce = –δl × κ. Each unfolding event signals the presence of an intermediate or the final release of the entire protein from the bilayer. Sequential unfolding events produces a sawtooth FEC curve with the depth of the valleys reflecting how close the cantilever returns to its neutral position. Our simulated FECs recapitulate key features of the experiments, including the extension of the unfolded segments being well described by the WLC model (6–8,11) (Eq. S1).

Because the conformation of the protein is known at every time point in the simulations, we may identify the

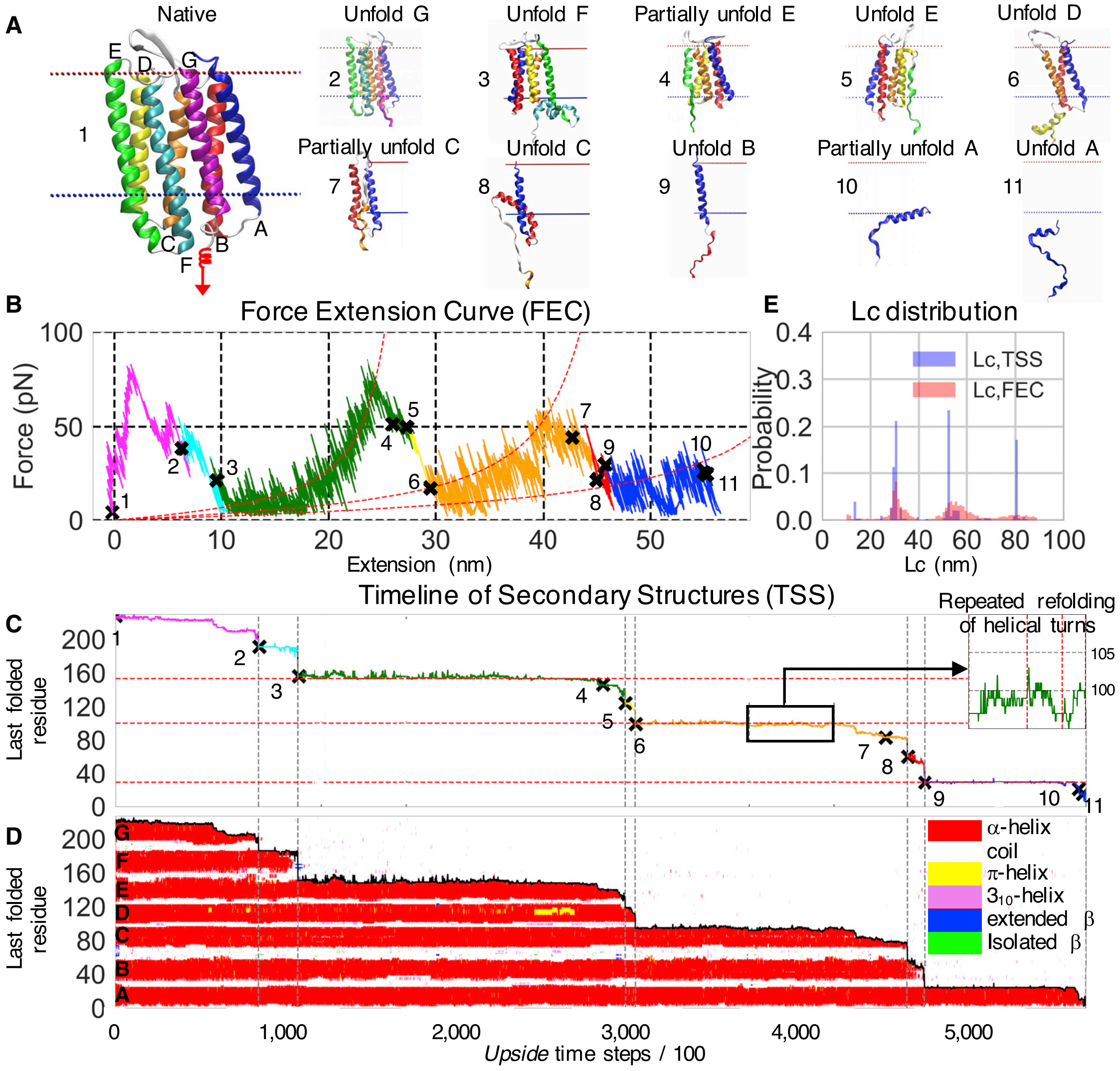

FIGURE 2 A representative unfolding trajectory of bR. (*A*) Typical unfolding intermediates (Video S1 *A*) are shown. (*B*) The force-extension curve (FEC) is shown. The red dashed curves are fit to the WLC model, using the contour lengths (*Lc*) of the most populous states (as seen in *C*). (*C*) Shown is the demarcation of index of the most C-terminal residue that remains folded as determined from the TSS plot (*D*, where each red strip represents a helix and hence identifies the folded regions in the protein). The "X" marks and the associated numbers in (*B*) and (*C*) denote reference points of the trajectory in (*A*) and serve to map the FEC to the TSS. The red horizontal dashed lines in (*C*) identify the most populous intermediates during the unfolding of the ED and CB helix pairs and of helix A. The green line in the inset in (*C*) is an example of repeated fraying/refolding of helical ends in the C helix; other examples of back-and-forth folding transitions are presented in Fig. S4. Secondary structure designations follow conventions in the Dictionary of Secondary Structure of Protein (48) in which coil refers to either H-bonded turn, bend and loop, and irregular elements. Gray vertical dashed lines in (*C*) and (*D*) define the time points at which a given TM helix has completely unfolded. (*E*) Probability distributions of the *Lc* values obtained from the FEC (as in *B*) fit with the WLC model and directly from the TSS (as in *C*). To see this figure in color, go online.

sawtooth pattern as reflecting the sequential unfolding of pairs of TM helices, starting sequentially from the C-terminal helices in the order of GF, ED, and then CB. The pairwise unfolding is a known consequence of the up-down topology of the protein (pulling out a single helix would yield an energetically unfavorable conformation with a single unfolded segment traversing the bilayer) (11). The pairwise unfolding of TM helices is readily observed in the TSS plot.

The first GF helix pair unfolds relatively quickly because the pair is connected directly to the cantilever and the force rapidly accumulates upon movement of the spring. Our simulations lack the capability of including the retinal, which is attached to helix G (38). As the retinal stabilizes the protein (39), our simulations probably underestimate the force that is required to remove helix G. Fortunately, helix G is removed first as the spring is attached to the C-terminus so that the rest of the trajectory should be unaffected and can be reasonably compared to experiments. The FEC for the remainder of the trajectory is dominated by the buildup of force as the unfolded segments are stretched (entropic tension), punctuated with drops in force reflecting the unfolding of the pairs of helices. The final step involves the release of helix A from the bilayer.

The position of the last residue remaining folded at a given time point is readily identified from the simulations. This position is used to construct a plot highlighting the structure of the intermediates, their boundaries, and the lengths of the unfolded segments (Fig. 2 *C*). The helical regions are plotted as a function of time in the TSS plot (Fig. 2 *D*). The contour length (*Lc*) and folded regions also can be inferred from the FEC, as done experimentally. We find that they generally agree with their true values obtained directly from the simulated conformations (as shown in the TSS; Fig. 2 *E*). This agreement supports the standard experimental assumption that the unfolded region is located only at the C-terminus, whereas the remaining TM helices remain stationary within the bilayer (11). However, the agreement is not absolute because our simulations find that partially folded helices can translate vertically in the bilayer (Fig. 2 *A*10) or change from $\alpha$-helix to $3_{10}$ helix or $\pi$-helix conformations (Fig. S5 *A*) and even unfold at the amino terminus (Fig. S5 *B*). These events, although infrequent, are likely to be missed in experiments.

An impressive feature of the experimental AFM study (11) is the ability to observe back-and-forth unfolding and refolding events at the ends of the helices for two to four residues, representing a half to a full helical turn. We likewise observe these back-and-forth transitions between microstates in all three major regions (ED, CB, and A helices) (Figs. 2 *C* and 3; Fig. S4).

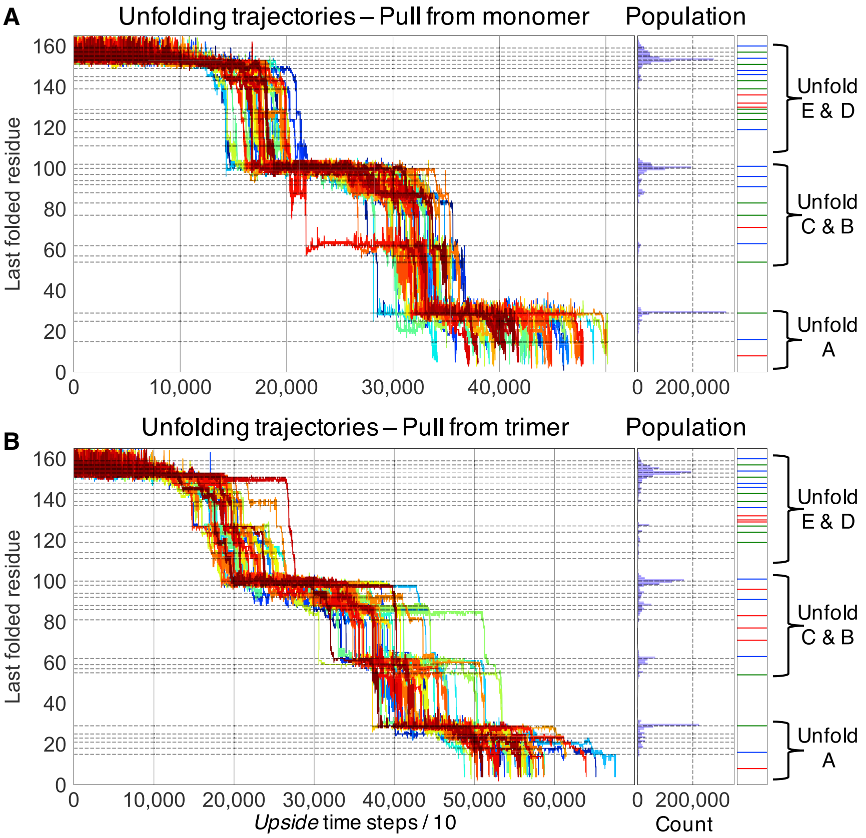

FIGURE 3 Unfolding trajectories and intermediates of bR. Unfolding behavior of (*A*) monomeric and (*B*) trimeric bR is shown. Unfolding trajectories plotted according to the index of the last folded residue (*left*), the corresponding population distribution (*middle*), and the intermediates found in the experiment (11) (*right*) are shown. The time spent at each position is histogrammed (*blue bars*) and fit using multiple Gaussians to identify the population and position of the simulated intermediates (designated with the *black dashed lines* in the *left* and *middle* regions). The blue, green, or red solid lines (*right*) denote intermediates exactly matching the experiment, matching within one residue, or not observed in our simulations, respectively. Of the 45 trajectories conducted on trimeric bR, helix A unfolds by itself in 43 trajectories. Of the 90 trajectories of monomeric bR, helix A unfolds by itself in 48 trajectories. In the rest of the trajectories, more than 2 TM helices are pulled out of the membrane in the last observed unfolding event (these abbreviated trajectories are not shown for simplicity). Helix A is more stable in the trimers because it can interact with the other two bR molecules. Therefore, more helix A intermediates are observed (Table S2), and the mean unfolding force for the major intermediate in helix A is higher (Table S1). Note, the last four residues (I4, Q3, A2, E1) are not in the PDB: 1qhj file and therefore are not included in the simulations and the comparison. To see this figure in color, go online.

## Unfolding intermediates of monomeric and trimeric bR

To further test the agreement between our simulations and the AFM measurements (11), we compare the populations and structures of the intermediates. Following the procedure employed for soluble proteins (3), the population distribution of intermediates obtained from the TSS plot are fit with multiple Gaussian functions, assuming a width of one amino acid (Fig. S6). More and longer-lived intermediates are observed in simulations of trimeric versus monomeric bR (Fig. S8 *G*; Tables S1 and S2). Presumably, this difference is due to the additional protein-protein contacts in the trimer. In the simulations of monomeric (trimeric) bR, we identify 29 (32) intermediates with 15 (15), 11 (11), and 3 (6) having folded-unfolded boundaries in the ED, CB, and A helices, respectively (Fig. 3; Table S2). Among this group of 29 (32) intermediates, 11 (9) exactly correspond to 1 of the 26 experimental intermediates and another 10 (10) are within 1 residue of an experimental intermediate. For both the monomer and trimer, we fail to identify 5 intermediates (3 near the bottom of the E helix, 1 in the middle of the loop connecting the CB helices, and 1 at the bottom of the A helix) while identifying 8 (13) that are not observed experimentally (Table S2).

The disparity in identifying intermediates may reflect real differences, such as errors in our energy function, pulling speed, effective temperature, or the study of isolated trimers rather than the full lattice of trimers. However, the different protocols for identifying intermediates, either the use of the experimental FEC or the simulated TSS, can also affect the determination of the intermediates. For example, we observe that the protein can have unfolded regions between folded regions (Fig. S5 *B*), a possibility that is not considered in the experimental analysis.

## Force levels and pulling velocities in bR simulations

Generally, the use of a fast pulling speed or a soft spring constant can produce an FEC where the applied force has insufficient time to relax back to zero after an unfolding event and the trough-to-peak depth in the sawtooth pattern is reduced. Our study sets the velocities of the cantilever and spring constant to 0.001 Å/*Upside* time step (∼$10^6$ nm/s; see Estimating *Upside* Temperature and Time Scale in Supporting Materials and Methods) and 21 pN/nm, respectively, chosen partly so that the simulations match the experimental sawtooth pattern (11) and to be computationally feasible. The experimental and simulated FEC traces are similar over the range of experimental velocities and spring constants of 30–3000 nm/s and 13–58 pN/nm, respectively. To further test whether our pulling velocities are appropriate for making comparisons to the experimental data, we decreased our speed by a factor of 10. This change has minimal effect on the depth of our sawtooth pattern except at the fastest pulling speeds (Fig. S8; Table S1), supporting the view that we employ sufficiently slow pulling speeds to enable the unfolded regions to equilibrate.

In the simulations, the major intermediate of the helix pair ED exhibits an average simulated unfolding force of 88.7 ± 3.0 and 83.4 ± 2.4 pN for bR in trimeric and monomeric forms, respectively (Table S1). These values are close to the experimental value of 94 ± 1 pN (11). However, our pulling velocities are three to four orders of magnitude faster, which should result in higher rupture forces. According to the Bell-Evans model (26), the most probable rupture force, $F_{mp}$, is partly determined by the pulling velocity $v$ and the width of the potential barrier $\Delta x^{\neq}$ (the distance from initial state to transition state where the separation distance is the reaction coordinate), as well as the natural transition rate $k_0$ according to $F_{mp} = (\beta \times \Delta x^{\neq})^{-1} ln(\beta \times \Delta x^{\neq} \times \kappa \times v/k_0)$ where $\beta = (k_B \mathrm{T})^{-1}$, assuming the potential barrier does not move under force (6,27). By fitting the mean unfolding forces with log(pulling velocities), the barrier distances for helix pairs ED, CB, and helix A (Fig. 4; Table S7) are obtained. We observe the logarithmic relationship for the three unfolding events involving helix pairs ED, CB, and helix A (Fig. 4). This finding indicates that our pulling velocity is slow enough that force is still linear with log(velocity). At the experimental pulling speed, the extrapolated values of the rupture force are 19.1, 9.4, and −32 pN for the ruptures of the ED, CB, and A helices, respectively, whereas the experimental values are 94, 49, and 62 pN. This discrepancy could be due to an error in our conversion of the *Upside* time step to real time (see Supporting Materials and Methods), our membrane potential being too weak, or other

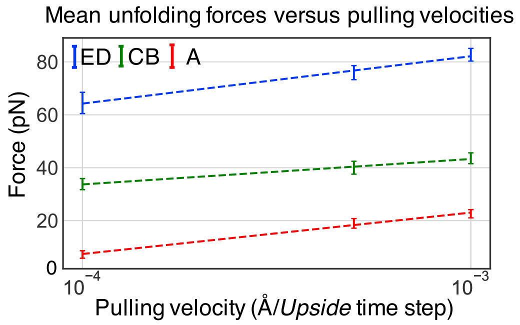

FIGURE 4 Mean unfolding force and pulling velocity. The mean unfolding forces of the major helix pairs ED, CB, and helix A (Table S1) are computed from the trajectories at each pulling velocity. A logarithmic dependence of the unfolding force on the pulling velocity is observed in the range 0.0001–0.001 Å/*Upside* time step (∼$10^5$–$10^6$ nm/s). The distributions of the mean unfolding forces are shown in Fig. S7, and the transition distances are provided in Table S7. The error bars indicate the standard deviation of the rupture forces at each velocity. To see this figure in color, go online.

inaccuracies in the *Upside* energy function. For example, doubling of the membrane potential significantly stabilizes the last helix, helix A, in the membrane bilayer and increases the rupture force of helix A from 22.6 to 38.9 pN (Table S1).

## GlpG can unfold along multiple routes with well-populated intermediates

Pulling simulations on GlpG were performed using the same pulling velocity and spring stiffness as used in the bR simulations. But, to match the experimental pulling geometry used in the MT apparatus, the force was applied horizontally, pulling on both GlpG's N- and C-termini that are located on the same side of the bilayer (Figs. 1 *B* and 5 *A*; Video S1 *B*). Specifically, the C-terminal spring is translated horizontally to the membrane surface to generate force parallel to the surface, whereas the N-terminal spring is held fixed (similar outcomes are produced when the N-terminal spring moves at the same velocity and the C-terminal spring is fixed; Fig. S12 *A*; Table S3).

Figs. 5 *E* and S9 illustrate the diversity of unfolding pathways, including examples of unfolding starting at either terminus or, more rarely, starting with the central helices. Unfolding from the N-terminus ("N→C pathway") typically proceeds sequentially for three helices: TM1→TM2→TM3 followed by the unfolding of TM4, TM5, and TM6 in any order (Figs. 4 *C* and 5 *E*, *left*; 29 of 50 trajectories in Figs. S9 *A* and S10). Unfolding from the C-terminus ("C→N pathway") typically proceeds sequentially: TM6→TM5→TM4→TM3→TM2→TM1 (Fig. 5 *E*, *middle*; 3 of 50 trajectories in Figs. S9 *D* and S11). Other pathways are shown in Fig. S9.

A PCA often is employed to interpret high-variance collective protein motions in simulations such as those observed along folding pathways (28). Internal coordinates, such as interresidue separations, are a poor separation, (e.g., two very different structures can have the same N-to-C separation depending on whether the unfolding initiates at the N- or C-terminus). Hence, we used the native C$\alpha$-C$\alpha$ contacts to derive our PCs (Fig. 5 *F*).

The results of the PCA are plotted as a population heat map for the first two PCs, PC0 and PC1. The interpretation of PCs can be challenging because these mathematical constructs typically do not refer to any specific real-space conformation or motion. For the unfolding of GlpG, however, we empirically find PC0 is related to the lateral expansion of the protein, whereas PC1 is related to whether unfolding occurs from the N-terminus, the center, or the C-terminus. Each of these three scenarios is illustrated with a single unfolding trajectory (Fig. 5 *F*, *thin red line* in the three maps). The unfolding pathway begins from the right, fully folded side and proceeds to the left, fully extended side. For unfolding from the N- or C-terminus, respectively, the path goes along the lower edge (Fig. 5 *F*, *left*; Fig. S9 *A*) or the upper edge (Fig. 5 *F*, *middle*; Fig. S9 *D*). For unfolding beginning in the middle of the protein, the pathway traverses the center of the map (Fig. 5 *F*, *right*; Fig. S9 *C*). Beyond highlighting the pathway heterogeneity, the PCA heat maps also emphasize that unfolding may occur through ~10 intermediates (Figs. S10 and S11), including both the intermediates observed in the MT experiment ($I_1$ and $I_2$ in Fig. 5 *F*, *middle*) (9).

The dependence on spring velocity and constant were also examined, as done with bR. The reduction of the velocity by 10-fold or the spring constant by 5-fold had minimal effect on the PCA plots (Fig. S12 *A*). These results support the robustness of our method. We are unable to readily apply Bell-Evans-style correction to determine the disruption forces extrapolated to the experimental pulling velocity because of the diversity of unfolding barriers in the GlpG landscape. Nevertheless, the correction should be similar to that for bR.

## More stable N-domain of GlpG is more likely to unfold

Unexpectedly, we find that the more stable N-terminal domain (40) has higher tendency to unfold before the C-terminal domain (Fig. S9; Table S3). This counterintuitive observation arises from the differential hydrophobicity of the TM helices (41). TM2, TM5, and TM1 are the most hydrophobic, whereas TM6 is the least. Although hydrophobicity promotes helix insertion into the bilayer (stage I folding (42)), it has the complementary effect of promoting dissociation of helices from the other hydrophobic helices in the bilayer because these helices are "well solvated" by the lipids in the bilayer. In fact, TM5 is completely dissociated from the other TM helices in the crystal structure of GlpG (43). The dissociation of TM1 and TM2 is energetically less costly than the dissociation of TM6 and TM4 because the latter process exposes polar and charged groups (Fig. 6 *A*; and the near-native and N1 intermediates in Fig. S10 *C*).

In addition, the 34-residue segment between TM1 and TM2 is of sufficient length to allow TM1 to dissociate from the other TM helices and remain upright in the bilayer. However, the linker between TM5 and TM6 has only 10 residues, so TM6 must tilt in the bilayer or the GlpG structure must distort for TM6 to dissociate. As a result, the TM6 helix does not readily dissociate on the C→N pathway; rather, the ends of the TM1 and TM6 helices unfold first (Fig. 6 *B*; C2 intermediate in Fig. S11 *C*). These multiple factors explain the preference for unfolding to occur along the N→C pathway.

## Cooperative unfolding of GlpG occurs with a force clamp and a soft spring

The extensive pathway diversity and large number of intermediates found in the simulations discussed in the previous

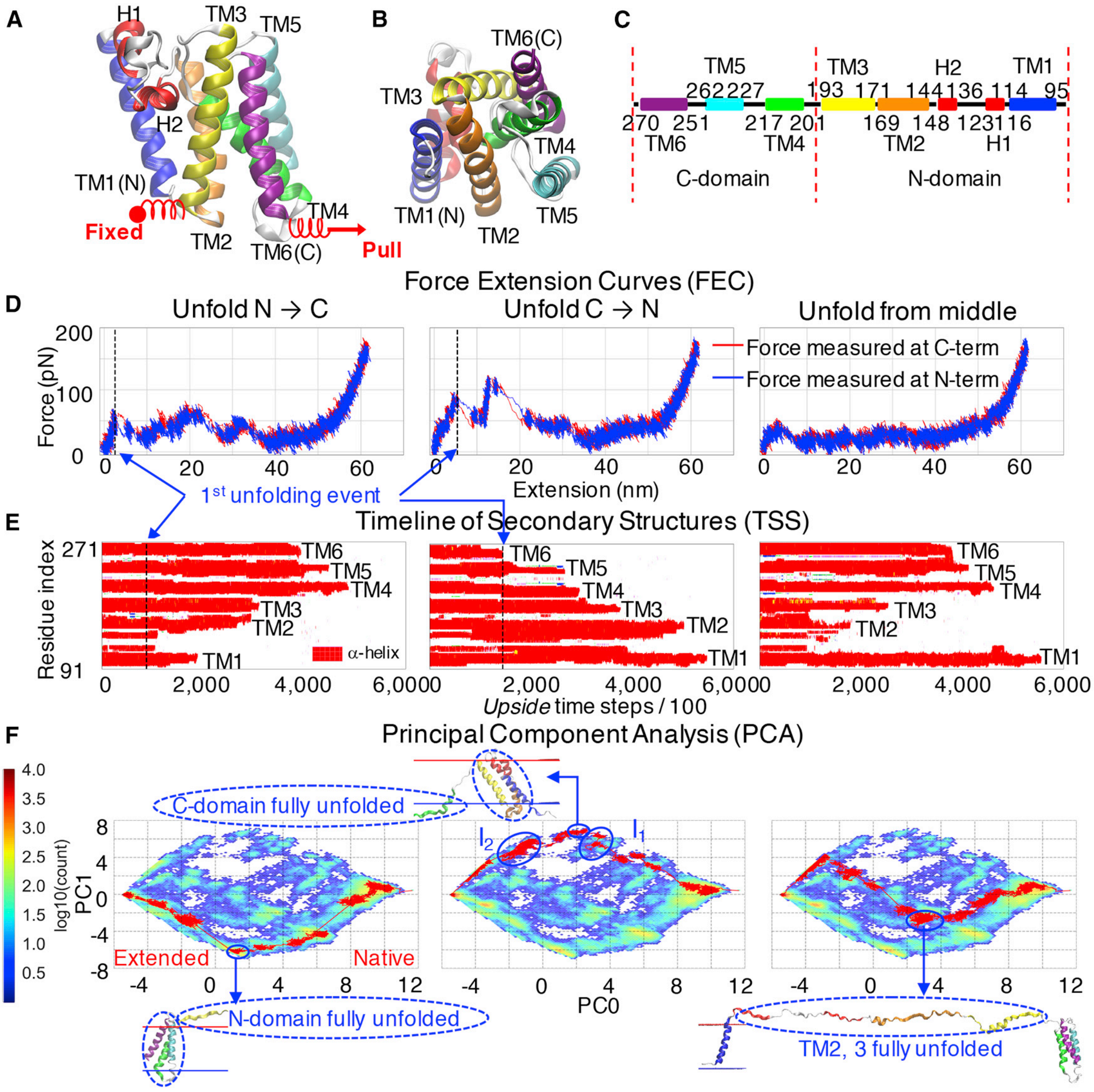

FIGURE 5 Unfolding pathway diversity for GlpG. (*A* and *B*) Side and bottom views of GlpG (PDB: 2xov) and (*C*) secondary structure and definition of N- and C-domains are shown. (*D–F*) FEC, TSS, and PCA plots illustrate unfolding beginning from the N-terminus (*left*; 14th trajectory from Fig. S9 *A*), the C-terminus (*middle*, first trajectory in Fig. S9 *D*), and the middle (*right*, 41st trajectory in Fig. S9 *C*). Videos of the three unfolding trajectories can be found in Video S1 *B*. In the FEC plots (*D*), the forces generated by the two springs (*blue* and *red lines*) are very similar, indicating that the force has sufficient time to equilibrate across the protein, a necessary condition for meaningful comparisons to experiment, except for an occasional small time lag at one end of the protein just when an unfolding event occurs at the other end of the protein. For instance, when the TM1 helix unfolds first, the force measured at the N-terminus drops faster than that at the C-terminus ((*D*) *left*, at extension ∼5 nm), whereas the force at the C-terminus drops faster when TM helices close to that end unfold first ((*D*) *middle*, at extension ∼10 and 20 nm). Unfolding pathways are defined by the sequence order of the unfolding of TM helices. In (*F*), the PCA heat map is evaluated from 50 trajectories. Red curves depict an unfolding pathway from the native to the fully extended state. PC0 relates to the end-to-end distance, whereas PC1 relates to whether unfolding begins from one end of GlpG or the other. The N→C and C→N pathways diverge first (proceeding along the *lower* and *upper* edge of the heat map, respectively) as the protein expands under force. The structures along the two pathways become the most distinct at the bottom and top of the heat map, where either the N- and C-domain are unfolded along the N→C and C→N pathways, respectively. Then, the two pathways converge as more TM helices unfold. The two blue circles in the middle subplot of (*F*) replicate the two experimentally observed unfolding intermediates $I_1$ and $I_2$, formed by the unfolding of TM6 and TM5, followed by TM4 and TM3, respectively, with the final step being the unfolding of TM2 and TM1 (9). To see this figure in color, go online.

section were not observed in the experimental MT study. That study also concluded that unfolding initiates only from the C-terminus and goes through two intermediates (9). We propose that the differences between experiment and the simulations shown in the prior sections are primarily a result of a difference in the mode of force application, possibly compounded by limited experimental time resolution.

The GlpG unfolding simulations presented above mimic a typical AFM measurement in that force builds up as the

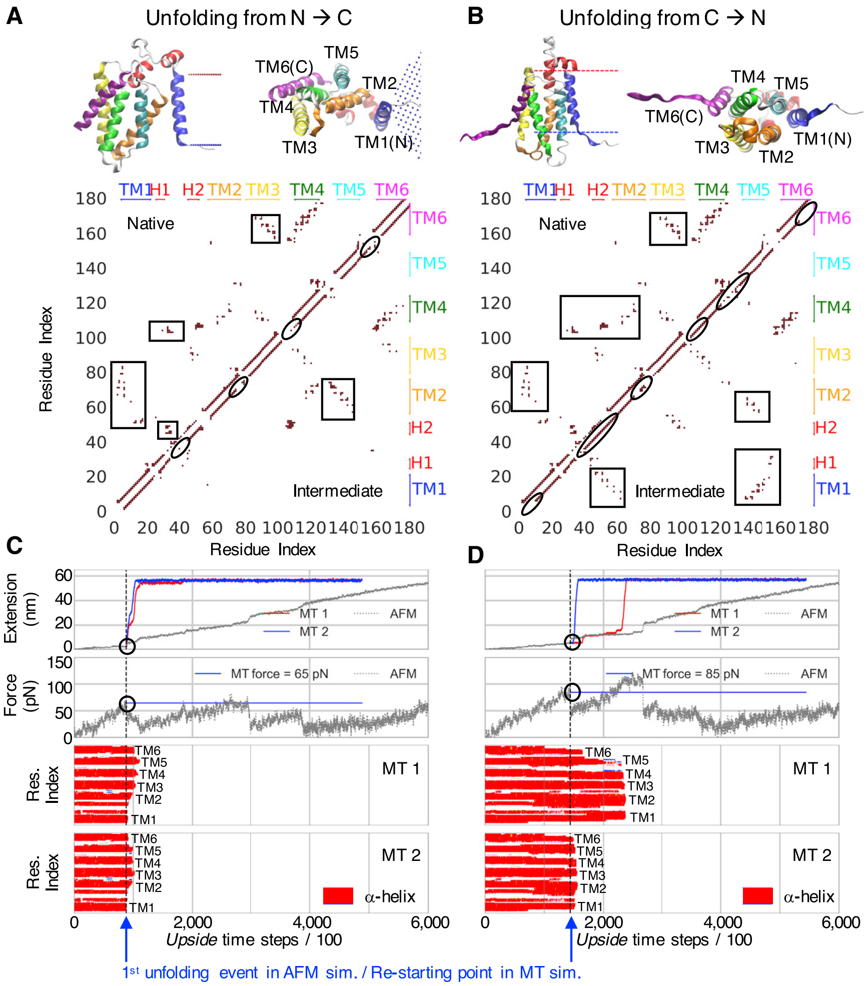

FIGURE 6 GlpG simulations using a soft spring, as in a MT measurement. (*A* and *B*) Structures and contact maps of the major unfolding intermediate on the N→C and the C→N pathways are shown. The intermediates selected for restarting simulations from the N→C and the C→N pathways correspond to the near-native intermediate in Fig. S10 *C* and the C2 intermediate in Fig. S11 *C*, respectively. Differences in contacts between the native state and the intermediates are marked by black rectangles and ellipses (along the main *diagonal*). The length of the maroon lines near the main diagonal identifies the length of the folded portion of the helices (e.g., TM6 in the N→C pathway is present in the first intermediate but is partially unfolded in its counterpart along the C→N pathway). (*C* and *D*) Shown are extension and force profiles over time, along with TSS plots for two examples of unfolding trajectories (denoted MT1 and MT2). After an initial force loading period, the protein begins to unfold (*blue arrows*), and the force is held constant for the rest of the trajectory reflecting the use of a soft spring. For comparison, the extension and force profiles (*gray*) for the N→C and C→N trajectories are shown for a simulation conducted using a stiff spring (denoted AFM) where the force builds up and relaxes after unfolding events (Fig. 5). To see this figure in color, go online.

cantilever is translated, followed by a rapid drop in the force after each unfolding event as the newly unfolded region provides sufficient slack to allow the cantilever to relax back toward its equilibrium position: $\text{Force}_{\text{after}} = \text{Force}_{\text{before}} - \delta l \times \kappa \approx 0$ (Fig. 1, *B* and *D*). The rapid relaxation of force reduces the probability that any other part of the protein unfolds in the same kinetic event. Consequently, unfolding occurs with multiple distinguishable intermediates, and the FEC has multiple peaks and the distinctive sawtooth pattern.

In a typical MT experiment, the pulling force increases as magnets move toward the bead, which is attached to the protein. However, once a portion of the protein unfolds, the force on the bead is largely unchanged because the magnetic field varies on a scale of microns, whereas the bead moves only tens of nanometers as a protein segment unfolds (Fig. 1 *E*). Hence, the bead still resides in nearly the same magnetic field as it did before the unfolding event and it therefore experiences the same force. As $\text{Force}_{\text{after}} = \text{Force}_{\text{before}} - \delta l \times \kappa \approx \text{Force}_{\text{before}}$, or $\delta l \times \kappa \approx 0$, the MT intrinsically function as a very soft spring (Fig. 1, *C* and *E*). Consequently, the force level present at the beginning of the first unfolding event is retained for the remainder of the measurement (as found in Fig. 3 a in (9)).

At this elevated force given by the level at the first rupture event, the protein often is pulled apart in a few or even a single all-or-none process. Hence, few intermediates are observed. Generally, folding appears more cooperative with softer springs. This effect is most pronounced for "brittle" proteins where the first unfolding event requires a higher force than the subsequent unfolding events so that only the first event is kinetically distinguishable from the unfolding of the rest of the protein.

Appreciating this effect of having a soft spring, we mimic the MT experiment by employing a modified force clamp protocol (Fig. 1 *C*). Force is gradually increased until the first unfolding event, whereupon the force is held constant for the remainder of the trajectory (Fig. 6, *C* and *D*). Using this protocol, the N→C and C→N pathways are investigated in detail by restarting 40 total simulations from the structure present right at the point of first rupture on each route (1 structure for 20 N→C trajectories (Fig. 6 *A*) and another structure for 20 C→N trajectories (Fig. 6 *B*). The ruptures occur at a force of 65 and 85 pN, respectively, for the N→C and C→N pathways.

As anticipated, the unfolding of GlpG under this modified protocol of force clamp is more cooperative along both unfolding pathways as compared to those pathways when a stiff spring is used. After the first unfolding event with the softer spring, all the remaining helices are seen to unfold almost immediately and nearly in unison (Fig. 6, *C* and *D*, all red bars disappear at the same time in the TSS, unlike the behavior in Fig. 5 *E*). Additionally, fewer intermediates are seen, and they are more transient especially along the N→C unfolding pathway.

For the 20 trajectories conducted on each of the N→C and C→N pathways, we observe 1 major intermediate (at extension of ~20.5 nm) and 2 major intermediates (at extensions of ~5.5 and ~12.5 nm), respectively (Figs. 6, *C* and *D* and 7 *B*). The difference between the use of a stiff and soft spring is readily apparent in the PCA heat maps in which the soft spring measurement yields only one well-populated intermediate on either the N→C or C→N pathway, and little population appears elsewhere on the PCA heat map (Fig. S12 *C*). In contrast, the measurement using a stiff spring populates dozens of intermediates across the entire map (Figs. 5 *F*, S10 *A*, S11 *A*, and S12 *A*).

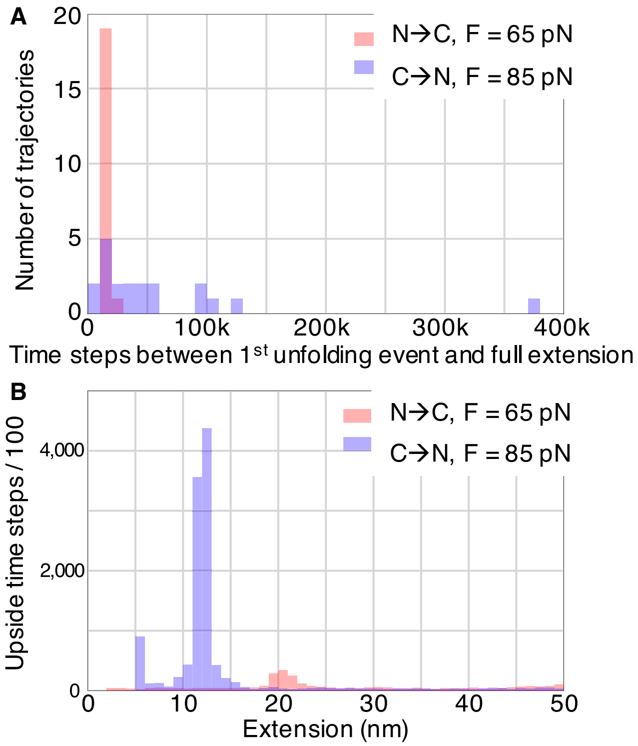

FIGURE 7 Lifetime and position of GlpG intermediates observed when using a soft spring for 40 simulations. (*A*) The number of trajectories as a function of lifetime between the first unfolding event and the fully extended state for the 20 unfolding trajectories originating from either end of GlpG are shown. (*B*) The corresponding aggregate time spent at each extension in an intermediate is shown. Each peak represents an intermediate (one is observed for the N→C pathway and two for the C→N pathway). The fully unfolded state has an extension above 50 nm, which defines the upper limit of the *x* axis. To see this figure in color, go online.

Furthermore, the behaviors of the simulations along the two pathways differ in a manner consistent with experiments. Simulated intermediates formed on the C→N pathway live longer than those on the N→C route (Fig. 7). After the first rupture and formation of an intermediate on the N→C unfolding pathway, the intermediate unfolds in less than 20,000 time steps for 19 of 20 simulations, whereas on the C→N unfolding pathway, the intermediate does not unfold even within 50,000 steps for 7 of 20 simulations despite the elevated level of force (65 vs. 85 pN, respectively; Fig. 7 *A*). Consequently, more intermediates are observed on the C→N pathway despite the fact that more molecules unfold though the N→C route (Fig. S9; Table S3).

The use of a 60 Hz CCD camera in the experiment (9) also could result in the unfolding of GlpG appearing more cooperative with a preference for unfolding beginning from the C-terminus. At this relatively slow frame rate, intermediates populated for less than ~16 ms could elude detection. Because the intermediates are longer lived on the C→N pathway, they are more likely to be detected than those on the N→C pathway. Thus, the use of a slow camera could increase the apparent degree of folding cooperativity and introduce a bias for observing intermediates on the C→N route.

## Other SMFS modes applied to GlpG

To explore other SMFS modes of unfolding, we pulled on GlpG in the same manner as in the AFM simulations of bR with force being applied vertically with a stiff spring attached to either the N- or the C-terminus. As with bR, we observe the characteristic sawtooth pattern (Fig. S14). However, the TM helices of GlpG are not as hydrophobic as bR's (41), so GlpG's TM helices are pulled out at lower force (~60 pN for the first pair of helices versus ~80 pN for bR). TM1 is pulled out in a distinct event in 4 of 20 simulations when the force is applied to the C-terminus, whereas TM6 is pulled out in 3 of 20 trajectories when the force is applied to the N-terminus. In the rest of the trajectories, more than one TM helix is pulled out in the same final unfolding event. Moreover, TM6 sometimes unfolds before TM5 when the N-terminus is pulled (Fig. S14 *B*), a behavior consistent with TM6 being intrinsically unstable within the bilayer (41). This inversion in the order of unfolding would be hard to infer from the FEC because it violates the assumption that the helices unfold according to their sequence order. The difference between this AFM mode and other modes well illustrated in the PCA plots, which display a series of intermediates going along the edges of the heat map (Fig. S12 *B*). Because no corresponding AFM experiment has yet been conducted with GlpG, these simulations provide testable predictions. Studies of other SMFS modes and the effects of mutations and temperature can be found in the Supporting Materials and Methods.

## DISCUSSION

We have modified our fast *Upside* MD algorithm (22,23) to include our new membrane burial potential (24) to study the forced unfolding of bR and GlpG in a variety of different modes of force application. *Upside* models the polypeptide backbone with five atoms and residue-dependent Ramachandran maps. Each type of side chain is represented with a multiposition, orientation-dependent bead. The side chains are repacked with a probability distribution that minimizes their free energy using belief propagation at every MD step. The great acceleration in speed afforded by this method allows us to run hundreds of simulations and conduct PCA to map out the energy surface. Unfolding trajectories of membrane proteins as large as trimeric bR are generated in less than 10 cpu-days with results that agree well with experimental studies (Figs. 2 and 3; (11)). That *Upside* captures many of the experimental features and identifies most intermediates is a positive indication that a reasonably accurate representation of the system has been achieved, including having a membrane burial potential with an appropriate energetic penalty for unfolding helices within the bilayer.

A major finding of our study is that the mode of force application and the strength of the spring strongly affect the observed details and apparent unfolding cooperativity. The use of a stiff spring constant $\kappa$ (and one with a very fast response time) enables the system to rapidly lower the force right after an unfolding event having extension $\delta$l, as $\text{Force}_{\text{after}} = \text{Force}_{\text{before}} - \delta\text{l} \times \kappa < \text{Force}_{\text{before}}$. Because the force on the rest of the protein has rapidly dropped, little additional unfolding occurs until the force is increased again. Hence, many intermediates are likely to be observed, and the FEC will have a distinct sawtooth pattern. With a soft spring, however, the force after an unfolding event remains at the same elevated level even after the extension of the chain because of unfolding, as $\text{Force}_{\text{after}} = \text{Force}_{\text{before}} - \delta\text{l} \times \kappa \approx \text{Force}_{\text{before}}$. Because subsequent species often are less likely to withstand the same high level of force as more folded species can (such as the native protein), unfolding will occur with few distinct steps and high apparent cooperativity (Fig. S12, *C* and *D*).

Conceptually, the applied force also can remain high after a rupture if the length of the already unfolded segments is sufficiently long that the chain (and whatever handles are used to connect the protein to the instrument) can act as weak entropic springs. This possibility will reduce the effective spring constant of the entire system. This effect becomes more pronounced in the later stages of unfolding because the length of the unfolded regions becomes increasingly longer with each unfolding event. This bias may reduce the probability of observing late intermediates.

Membrane proteins can be unfolded by either pulling vertically at one terminus or laterally at both ends. Each mode explores different regions of the energy surface (Fig. S12). Pulling vertically produces a more deterministic unfolding route because TM helices often unfold sequentially and in pairs, whereas the remaining portion of the protein largely remains intact. Pulling laterally tends to break interhelical contacts at a variety of locations, which leads to gross structural rearrangements even before any TM helix unfolds.

Few SMFS simulations have been conducted for membrane proteins, partly because of the computation resources required. One all-atom MD study (17) identified a number of key residues that resist mechanical unfolding in the intermediate states probed by the experiment (11), although its pulling speed (1–50 m/s) was too fast to observe WLC behavior for the unfolded segments. A 2016 CG study used the same pulling rate as ours ($\sim 10^6$ nm/s) and found WLC behavior (20). Although many features are similar between this and our studies, we observed more intermediates (Table S2).

A recent GlpG force-induced unfolding study concluded that the two-stage membrane folding model (44) is overly simplistic because isolated helices can coexist with a folded domain, and all the helices do not have to be in the bilayer before the initiation of folding (45). Although our study reproduces these two items, some technical differences between the two studies are worth noting. We explore different parts of the landscape as we explicitly simulate the experimental pulling process by directly applying force with springs, whereas the other study performs umbrella sampling with an energetic bias determined by the N-to-C separation. Also, we allow helices to unfold and do not employ a Gō model nor stabilize the N-terminus to promote the C→N pathway. The group's earlier study found that folding could occur along either the N→C or C→N pathways (46) as we observe, although we also observe unfolding from the center of the protein.

Although our method has widespread applications, several issues exist. First, we cannot refold the membrane proteins by relaxing the force, whereas refolding to the native is observed in the experimental studies (9,10,13). Second, we assume an infinite flat membrane bilayer, which is valid for bR, but experimentally, GlpG (9) is embedded in bicelles, which may not be large enough to accommodate all the states we generate in our simulations. Third, our estimated force rupture values have some uncertainty. We largely match the values for bR at our typical pulling speed. However, the extrapolation is low for the values appropriate for the experimental pulling speeds using the Bell-Evans relationship between log(pulling velocity) and rupture force. Improvements in our energy function are in progress that may address these issues.

## CONCLUSIONS

We have developed an accurate and fast atomic-level method to conduct hundreds of realistic unfolding simulations to characterize the energy surface for force-induced unfolding. The method reproduces many of the experimental features of SMFS studies for the unfolding of bR (11) and GlpG (9). Our simulations can assist experimental studies by helping convert FECs to structures, pathways, and energies. For example, we identified the more stable amino-terminal domain of GlpG as the more likely terminus to unfold, but it escaped detection in experiment because of the cooperative unfolding behavior along this route. The counterintuitive unfolding of this more stable end (40) arises in part from higher hydrophobicity of the amino-terminal TM helices, a finding that highlights a general folding property for membrane helices: although increased hydrophobicity promotes insertion into the bilayer, it also enhances dissociation because the lipid bilayer effectively behaves as a good solvent for isolated hydrophobic helices. In contrast, TM helices that bury polar or charged groups upon association tend to remain associated because the cost of exposing their nonhydrophobic moieties to the bilayer is high.

Our major finding is that the mode of force application significantly alters the properties of the energy surface. The use of a constant force, either explicitly or by employing a soft spring, reduces the probability of observing intermediates while increasing the apparent unfolding cooperativity as compared to the use of stiff springs that can rapidly relax after an unfolding event. Finally, our simulation tools can be employed to investigate SMFS transition paths (47) as well as conduct complicated "*gedanken*" pulling experiments beyond current experimental capabilities, such as pulling on multiple sites in multiple directions with different strengths of the springs, and with either membrane or soluble proteins.

All data and analysis codes supporting the findings of this study are available from the corresponding authors upon reasonable request. The full simulation package of *Upside* as well as the necessary parameter files are available to public access on GitHub: https://github.com/sosnicklab/upside-md.

## SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2019.09.011.

## AUTHOR CONTRIBUTIONS

Z.W. conceived the study, designed the research, implemented the AFM function into *Upside*, performed the simulations, analyzed the results, and wrote the manuscript. J.M.J. developed *Upside* and reviewed the codes of AFM function. K.F.F. provided theoretical guidance and wrote the manuscript. T.R.S. designed the research, analyzed the results, and wrote the manuscript.

## ACKNOWLEDGMENTS

We thank members in our group, J. Bowie and D. Min for helpful discussions and the reviewers and editor for their useful comments. Simulations were conducted using the Midway resource of the Research Computing Center at the University of Chicago.

This work is supported by National Institutes of Health/National Institute of General Medical Sciences grants GM055694 (to T.R.S. and K.F.F.) and GM087519 (to E. Perozo).

## REFERENCES

1. Neuman, K. C., and A. Nagy. 2008. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods.* 5:491–505.
2. Stigler, J., F. Ziegler, …, M. Rief. 2011. The complex folding network of single calmodulin molecules. *Science.* 334:512–516.
3. Sen Mojumdar, S., N. Scholl Z, …, M. T. Woodside. 2017. Partially native intermediates mediate misfolding of SOD1 in single-molecule folding trajectories. *Nat. Commun.* 8:1881.
4. Marszalek, P. E., H. Lu, …, J. M. Fernandez. 1999. Mechanical unfolding intermediates in titin modules. *Nature.* 402:100–103.
5. Schönfelder, J., R. Perez-Jimenez, and V. Muñoz. 2016. A simple two-state protein unfolds mechanically via multiple heterogeneous pathways at single-molecule resolution. *Nat. Commun.* 7:11777.
6. Janovjak, H., J. Struckmeier, …, D. J. Müller. 2004. Probing the energy landscape of the membrane protein bacteriorhodopsin. *Structure.* 12:871–879.
7. Kessler, M., and H. E. Gaub. 2006. Unfolding barriers in bacteriorhodopsin probed from the cytoplasmic and the extracellular side by AFM. *Structure.* 14:521–527.
8. Oesterhelt, F., D. Oesterhelt, …, D. J. Müller. 2000. Unfolding pathways of individual bacteriorhodopsins. *Science.* 288:143–146.
9. Min, D., R. E. Jefferson, …, T. Y. Yoon. 2015. Mapping the energy landscape for second-stage folding of a single membrane protein. *Nat. Chem. Biol.* 11:981–987.
10. Serdiuk, T., D. Balasubramaniam, …, D. J. Müller. 2016. YidC assists the stepwise and stochastic folding of membrane proteins. *Nat. Chem. Biol.* 12:911–917.
11. Yu, H., M. G. Siewny, …, T. T. Perkins. 2017. Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. *Science.* 355:945–950.
12. Min, D., R. E. Jefferson, …, J. U. Bowie. 2018. Unfolding of a ClC chloride transporter retains memory of its evolutionary history. *Nat. Chem. Biol.* 14:489–496.
13. Kessler, M., K. E. Gottschalk, …, H. E. Gaub. 2006. Bacteriorhodopsin folds into the membrane against an external force. *J. Mol. Biol.* 357:644–654.
14. Thoma, J., B. M. Burmann, …, D. J. Müller. 2015. Impact of holdase chaperones Skp and SurA on the folding of β-barrel outer-membrane proteins. *Nat. Struct. Mol. Biol.* 22:795–802.
15. Pierse, C. A., and O. K. Dudko. 2017. Distinguishing signatures of multipathway conformational transitions. *Phys. Rev. Lett.* 118:088101.
16. Lu, H., B. Isralewitz, …, K. Schulten. 1998. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* 75:662–671.
17. Kappel, C., and H. Grubmüller. 2011. Velocity-dependent mechanical unfolding of bacteriorhodopsin is governed by a dynamic interaction network. *Biophys. J.* 100:1109–1119.
18. Stirnemann, G., D. Giganti, …, B. J. Berne. 2013. Elasticity, structure, and relaxation of extended proteins under force. *Proc. Natl. Acad. Sci. USA.* 110:3847–3852.

19. Zhuravlev, P. I., M. Hinczewski, …, D. Thirumalai. 2016. Force-dependent switch in protein unfolding pathways and transition-state movements. *Proc. Natl. Acad. Sci. USA.* 113:E715–E724.
20. Yamada, T., T. Yamato, and S. Mitaku. 2016. Forced unfolding mechanism of bacteriorhodopsin as revealed by coarse-grained molecular dynamics. *Biophys. J.* 111:2086–2098.
21. Gumbart, J., M. C. Wiener, and E. Tajkhorshid. 2007. Mechanics of force propagation in TonB-dependent outer membrane transport. *Biophys. J.* 93:496–504.
22. Jumper, J. M., N. F. Faruk, …, T. R. Sosnick. 2018. Trajectory-based training enables protein simulations with accurate folding and Boltzmann ensembles in cpu-hours. *PLoS Comput. Biol.* 14:e1006578.
23. Jumper, J. M., N. F. Faruk, …, T. R. Sosnick. 2018. Accurate calculation of side chain packing and free energy with applications to protein molecular dynamics. *PLoS Comput. Biol.* 14:e1006342.
24. Wang, Z., J. M. Jumper, …, T. R. Sosnick. 2018. A membrane burial potential with H-bonds and applications to curved membranes and fast simulations. *Biophys. J.* 115:1872–1884.
25. Subramaniam, S., and R. Henderson. 2000. Molecular mechanism of vectorial proton translocation by bacteriorhodopsin. *Nature.* 406:653–657.
26. Evans, E. 2001. Probing the relation between force–lifetime–and chemistry in single molecular bonds. *Annu. Rev. Biophys. Biomol. Struct.* 30:105–128.
27. Woodside, M. T., and S. M. Block. 2014. Reconstructing folding energy landscapes by single-molecule force spectroscopy. *Annu. Rev. Biophys.* 43:19–39.
28. Ernst, M., F. Sittel, and G. Stock. 2015. Contact- and distance-based principal component analysis of protein dynamics. *J. Chem. Phys.* 143:244114.
29. Lomize, M. A., I. D. Pogozheva, …, A. L. Lomize. 2012. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* 40:D370–D376.
30. Guex, N., and M. C. Peitsch. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis.* 18:2714–2723.
31. Ting, D., G. Wang, …, R. L. Dunbrack, Jr. 2010. Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. *PLoS Comput. Biol.* 6:e1000763.
32. McGibbon, R. T., K. A. Beauchamp, …, V. S. Pande. 2015. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* 109:1528–1532.
33. Pedregosa, F., G. Varoquaux, …, E. Duchesnay. 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12:2825–2830.
34. Levy, R., and M. Maaloum. 2002. Measuring the spring constant of atomic force microscope cantilevers: thermal fluctuations and other methods. *Nanotechnology.* 13:33–37.
35. Bustamante, C., J. F. Marko, …, S. Smith. 1994. Entropic elasticity of lambda-phage DNA. *Science.* 265:1599–1600.
36. Jha, A. K., A. Colubri, …, T. R. Sosnick. 2005. Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc. Natl. Acad. Sci. USA.* 102:13099–13104.
37. Robustelli, P., S. Piana, and D. E. Shaw. 2018. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. USA.* 115:E4758–E4766.
38. Curnow, P., N. D. Di Bartolo, …, P. J. Booth. 2011. Stable folding core in the folding transition state of an alpha-helical integral membrane protein. *Proc. Natl. Acad. Sci. USA.* 108:14133–14138.
39. Cladera, J., J. Torres, and E. Padrós. 1996. Analysis of conformational changes in bacteriorhodopsin upon retinal removal. *Biophys. J.* 70:2882–2887.
40. Guo, R., K. Gaffney, …, H. Hong. 2016. Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat. Chem. Biol.* 12:353–360.
41. Yang, Y., R. Guo, …, H. Hong. 2018. Folding-degradation relationship of a membrane protein mediated by the universally conserved ATP-dependent protease FtsH. *J. Am. Chem. Soc.* 140:4656–4665.
42. Hong, H. 2014. Toward understanding driving forces in membrane protein folding. *Arch. Biochem. Biophys.* 564:297–313.
43. Baker, R. P., and S. Urban. 2012. Architectural and thermodynamic principles underlying intramembrane protease function. *Nat. Chem. Biol.* 8:759–768.
44. Popot, J. L., and D. M. Engelman. 1990. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry.* 29:4031–4037.
45. Lu, W., N. P. Schafer, and P. G. Wolynes. 2018. Energy landscape underlying spontaneous insertion and folding of an alpha-helical transmembrane protein into a bilayer. *Nat. Commun.* 9:4949.
46. Schafer, N. P., H. H. Truong, …, P. G. Wolynes. 2016. Topological constraints and modular structure in the folding and functional motions of GlpG, an intramembrane protease. *Proc. Natl. Acad. Sci. USA.* 113:2098–2103.
47. Hoffer, N. Q., K. Neupane, …, M. T. Woodside. 2019. Measuring the average shape of transition paths during the folding of a single biological molecule. *Proc. Natl. Acad. Sci. USA.* 116:8125–8130.
48. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.

**Biophysical Journal, Volume *117***

## Supplemental Information

# On the Interpretation of Force-Induced Unfolding Studies of Membrane Proteins Using Fast Simulations

**Zongan Wang, John M. Jumper, Karl F. Freed, and Tobin R. Sosnick**

# Supplemental Information for

## On the interpretation of force-induced unfolding studies of membrane proteins using fast simulations

Zongan Wang[1,2], John M. Jumper[1,2,3,5], Karl F. Freed[1,2,*], Tobin R. Sosnick[2,3,4,*]
[1]Department of Chemistry, [2]James Franck Institute, [3]Department of Biochemistry and Molecular Biology, [4]Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL, 60637.
[5]Current address: DeepMind, 5 New Street Square, London EC4A 3TW, UK.
[*]To whom correspondence may be addressed.
Email: trsosnic@uchicago.edu or freed@uchicago.edu

**This PDF file includes:**

- Supplemental information text
- Supplemental information methods
- Figures S1 to S14
- Tables S1 to S7
- Caption for movie S1
- References for SI reference citations

# Supplemental Information Text

**Other SMFS modes applied to GlpG.**

We also performed standard force clamp simulations where the force is rapidly set and held at a constant value throughout the unfolding trajectory. The values generally are set at a force substantially less than the level where the first unfolding event occurs when operating under pulling mode with increasing force. We find that multi-step sequential unfolding from both N-to-C and C-to-N are more likely to be seen under lower force (e.g., 40 versus 60 pN), though the protein tends to unfold more cooperatively and more quickly at either force compared to the pulling with the stiff cantilever (**Figs. S12D and S13**).

**Altering the pathway fluxes using mutation, temperature and spring constant.**

We also performed unfolding simulations on destabilizing GlpG mutants to examine the effects on the unfolding pathways. Of the investigated residues having an H-bonding side chain in the N-domain, the E166A mutation is the most destabilizing (1). This residue is located near the bottom of the TM2 helix and forms two H-bonds to the backbone nitrogens of Val96 and Thr97 on the TM1 helix and two to the side chains of Thr97 on the TM1 and Ser171 on the TM3 helix (1). The G261V mutation on helix TM6 is at the center of the GxxxGxxxA motif that enables the close backbone-backbone association of the TM4 and TM6 helices. This mutation decreases the $T_m$ by 28.1±0.08 °C and increases the probability of unfolding from the C-domain by 50% (10 to 15 events, of a total of 50) (**Fig. S12A and Table S3**). To our surprise, the disruption of the H-bond network at the bottom of the triad of the three TM helices in the N-domain barely changes the probability of initiating the unfolding from this end (40 versus 41 events of a total of 50) (**Fig. S12A and Table S3**).

Unfolding from the N- rather than the C-terminus is 4-fold more probable at 270 K. At 300 K, the ratio is reduced to 1.2. And weakening the spring constant by a factor of 5 (0.01 to 0.05 $k_B\mathrm{T}/\text{Å}^2$) further reduces the branching ratio to 0.7 (**Fig. S9 and Table S3**). Even though there are differences, the fundamental heterogeneous pathway behavior remains (**Fig. S12A**).

# Supplemental Information Methods

**Calibration of virtual cantilever using thermal fluctuations.**
To test whether our spring constant, κ, functions as intended, we compared the observed thermal fluctuations of the cantilever to those expected from the equipartition theorem, $<z^2> = k_B T/\kappa$ (2). We used the first 3, 10, 20, 50 residues of bR and ran simulations with the first residue attached to the virtual cantilever and the rest of the segment restrained as a rigid body. In this case, we can measure the thermal fluctuation of the tip of the cantilever via the fluctuation of the residue (**Fig. S1A**). The square root of the mean fluctuations has a linear relation with the reverse of the square root of the spring constant (**Fig. S1B**) (2).

**Worm-like chain (WLC) model and the analytical solution to contour length**
Unfolded proteins and nucleic acids behavior under force can be described with a worm-like chain (WLC) model for polymer elasticity (3-5). We present an analytic solution for the contour length Lc as a function of force and extension. In the WLC model (3), the force F and the extension x of the unfolded protein has the following relation:

$$F=\frac{k_BT}{L_p}\left[\frac{1}{4}\left(1-\frac{x}{L_c}\right)^{-2}+\frac{x}{L_c}-\frac{1}{4}\right] \qquad (1)$$

where $k_B$ is the Boltzmann constant, T is the temperature, Lp = 0.4 nm is the persistent length of unfolded polypeptide, and Lc is the contour length (total length) of the unfolded polypeptide.
Let $\alpha=\frac{k_BT}{L_p}$, $\lambda=1-\frac{x}{L_c}$, $\omega=\frac{4F}{\alpha}-3$ and substitute them into **Eq. 1**, we have

$$4\lambda^3+\omega\lambda^2-1=0(2)$$

According to Cardano's method (6), any cubic equation can be solved analytically: $ax^3+bx^2+cx+d=0$ $(a\neq 0, a,b,c,d\in\mathbb{R})$.

Let $x=y-\frac{b}{3a}\Rightarrow y^3+\left(-\frac{b^2}{3a^2}+\frac{c}{a}\right)y+\left(\frac{2b^3}{27a^3}-\frac{bc}{3a^2}+\frac{d}{a}\right)=0$

Let $\begin{cases}P=-\frac{b^2}{3a^2}+\frac{c}{a}\\ Q=\frac{2b^3}{27a^3}-\frac{bc}{3a^2}+\frac{d}{a}\end{cases}\Rightarrow y^3+Py+Q=0$

Let $\Delta=\left(\frac{P}{3}\right)^3+\left(\frac{Q}{2}\right)^2$ and $\begin{cases}S=\left(-\frac{Q}{2}+\sqrt{\Delta}\right)^{1/3}\\ T=\left(-\frac{Q}{2}-\sqrt{\Delta}\right)^{1/3}\end{cases}$, we have three roots: $\begin{cases}y_1=S+T\\ y_2=\beta S+\beta^2T\\ y_3=\beta^2S+\beta T\end{cases}$, where

$\beta=\frac{-1+i\sqrt{3}}{2}$ and $\beta^2=\frac{-1-i\sqrt{3}}{2}$ are the two complex cubic roots of -1.

Here, Δ is the discriminant of the cubic equation.

If $\Delta > 0$, there is only one real root, $y_1$ and two complex roots $y_2$ and $y_3$.
If $\Delta = 0$, if $P = Q = 0$ all three roots are equal to 0, otherwise there are three real roots and two of them are equal.
If $\Delta < 0$, there are three unequal real roots with the following relation:

$$\begin{cases} x_1 + x_2 + x_3 = -\dfrac{b}{a} \\ {}^{1}\!/_{x_1} + {}^{1}\!/_{x_2} + {}^{1}\!/_{x_3} = -\dfrac{c}{d}, \text{ where } x_i = y_i - \dfrac{b}{3a}, \text{ i = 1, 2, 3.} \\ x_1 \bullet x_2 \bullet x_3 = -\dfrac{d}{a} \end{cases}$$

Now, back to **Eq. 2**, let $\lambda = y - \frac{\omega}{12}$ (3), we have

$$y^3 + \frac{-\omega^2}{48} y + \left( \frac{\omega^3}{27 * 32} - \frac{1}{4} \right) = 0 \text{ (4), and}$$

$$\Delta = \left( \frac{-\omega^2}{48 * 3} \right)^3 + \left( \frac{\omega^3}{27 * 64} - \frac{1}{8} \right)^2 = \frac{1}{64} \left( 1 - \frac{\omega^3}{27 * 4} \right) \text{ (5).}$$

In a standard SMFS experiment of unfolding bR, the force F is between 0 and 500 pN. At T = 298 K, $k_B$T = 4.114 pN•nm. Only when F < 20 pN is $\Delta > 0$; otherwise $\Delta < 0$. Therefore, **Eq. 2** has only one real root mostly (F ≥ 20 pN, $\Delta > 0$), which is the solution to our problem.

When $\Delta < 0$, we have

$$\begin{cases} \lambda_1 + \lambda_2 + \lambda_3 = -\dfrac{\omega}{4} < 0 \\ {}^{1}\!/_{\lambda_1} + {}^{1}\!/_{\lambda_2} + {}^{1}\!/_{\lambda_3} = 0 \quad \text{(6).} \\ \lambda_1 \bullet \lambda_2 \bullet \lambda_3 = \dfrac{1}{4} > 0 \end{cases}$$

Assuming $\lambda_1 \le \lambda_2 \le \lambda_3$, we have $\lambda_1 < \lambda_2 < 0 < \lambda_3$ and $\lambda_3$ is the root we want. In summary, Lc can be solved analytically given force and extension.

**Unfolding pathway analysis of bR.**
For every frame in trajectory, the Lc of the already unfolded segment can be determined through either FEC or TSS. Assuming intact secondary structure remains unchanged within the bilayer, Lc is uniquely determined given a force and an extension (labeled as Lc,FEC), from which we can infer how many residues have unfolded. Force was measured and recorded into the H5 file during the simulation, while extension was calculated as the distance that the Cα atom of the C-terminus has moved. On the other hand, if the number of unfolded residues is known first, Lc can be determined by mapping the number of unfolded residues to pre-determined Lc value (labeled as Lc,TSS) (**Table S6**). Secondary structures were computed by the `compute_dssp` function

in *MDTraj* (9), which follows the DSSP definition (10). Then Lc,TSS is obtained after identifying the most C-terminal residue which remained folded.

Trajectories plotted according to the last (C-terminal) folded residue were smoothed by a Savitzky-Golay filter (11) in *Scipy* (12), in which the `window_length` was set to 11, `polyorder` 3, `mode` 'nearest'. Then, the population distribution was histogrammed and fitted with multiple Gaussian functions to identify the number position of the simulated intermediates. Amplitudes and positions were fit assuming a width (standard deviation) of one amino acid, i.e. the positional uncertainty is assumed to be ±1 amino acid. Three major unfolding regions, denoted ED, CB, and A, were fit separately. In general, the more intermediates, the smaller the fitting error. To prevent over-fitting, we initially assume that intermediates are evenly distributed within each major unfolding region and obtained the fitting error as a function of the number of intermediates. Later, by adjusting the number of intermediates and their initial positions manually, we acquired fewer intermediates with a relatively low fitting error (**Fig. S6**).

**Calculation of contour length (Lc).**

We measured Lc from simulations of truncated bR molecules. The truncation points were chosen to match the experimental intermediates (13). For example, we simulated "A160", a truncated version having residues 161-232 unfolded to match the intermediate where residues 1-160 are folded while 161-232 are unfolded. For each of the truncated bR species, we fit its FEC with a WLC model (**Eq. 1**) using a fixed persistence length (Lp) of 0.4 nm, estimated by experiment (14) (**Fig. S2A, B**), to determine Lc (**Fig. S2C and Tables S4, S6)**. From this plot of Lc values, we obtained an average slope of 0.390 nm•residue$^{-1}$ in agreement with the experimental Lc estimate of 0.40±0.02 nm•residue$^{-1}$ (15) (the average distance between consecutive $C_\alpha$'s is 0.38 nm for actual proteins). Our Lc value is ~7% larger than 0.364 nm•residue$^{-1}$, a value recently obtained by a high precision measurement (13). Remarkably, for the truncated bR molecules, our Lc values exhibit the same minor deviations from linearity as those observed experimentally. The reproduction of these small deviations implies that they are real. The only reasonable source of the variability is a sequence dependent for Lc, consistent with experimental (15) and simulation (16) findings. Beyond providing support for the accuracy of our simulations, the residue dependence should be useful in identifying the sequence of the segment that is unfolded for a given Lc value.

**Force clamp simulations of ubiquitin.**

We unfolded ubiquitin (1ubq.pdb) to its fully extended state under high force (800 pN) and ran force clamp simulations with a constant force applied to both ends of the protein (procedure described in **Table S5**). We replicated all-atom MD results (16) in a few cpu-hours. Without force, the highly stretched polypeptide contracts considerably but remains extended under force as low as 10 pN (**Fig. S3A**). The distributions of ($\phi$, $\psi$) angles and end-to-end distances at different forces were similar to those of the all-atom MD study (16) (**Fig. S3B, C**). Also, we obtained good fitting of the average end-to-end distances and the applied forces according to the WLC model (**Fig. S3E**). The Lc was determined by minimizing the least-squares fitting error (**Fig. S3D**).

### Clustering analysis of GlpG unfolding trajectories.

We chose the Gaussian mixture algorithm implemented in *scikit-learn* (17) to perform the clustering analysis performed on the structures in each trajectory (**Figs. S10B and 11B**). A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions, which is good for density estimation. The number of clusters was estimated and supplied to the program. The corresponding density for each training point was measured and the point with the maximal density was chosen as the center to represent the cluster. The cluster centers were considered as the intermediate states along the unfolding pathway.

### Estimating *Upside* temperature and time scale

We use Verlet integration with a time unit of 0.009 *Upside* time steps and random number generator to implement the Langevin dynamics with a thermalization time scale of 0.135 time units. The time scale of thermalization (related to Langevin friction) is chosen to maximize the effective diffusion rate of chains while effectively controlling simulation temperature. As Langevin dynamics with any friction coefficient produces the same Boltzmann ensemble, we chose to maximize equilibration of our system rather than attempt to match a solvent viscosity.

The precise time and temperature scale of the *Upside* model is unclear because of the coarse-graining process. As compared to all-atom, explicit solvent simulations on a decapeptide using Charmm36, *Upside* is about 10,000-fold faster in part due to the lack of solvent and the smoothing of side chain interactions. This smoothing is likely to have a disproportionate effect for condensed structures as compared to extended structures. From the transition rates between folded and unfolded states (well-to-well barrier crossing process), we estimate the time unit for barrier crossing events to be ~1 psec. From transition rate between Ramachandran basins in the extended state (chain motions within a thermodynamic well), we estimate the time unit to be ~0.1 psec. As noted above, each time unit is 0.009 time steps. Irrespective of the issues with defining an absolute time scale, the equilibrium population distribution that determines the free energy is expected to be approximately correct, as well as the order of dynamical folding events.

### Simulation configurations of gradual pulling simulations.

In addition to the standard configuration for running *Upside* folding simulation of small soluble proteins (18, 19), we implement our membrane burial potential, which dynamically accounts for the degree of side chain exposure to lipids (20) and a new pulling function in the *Upside* simulations.

(1) Prepare the initial protein structure in a pickle file format the input.

```
python PDB_to_initial_structure.py \
pdbname.pdb pdbname                \
--allow-unexpected-chain-breaks    \
--record-chain-breaks              \
--disable-recentering
```

(2) Prepare the H5 file for the simulation, which includes all simulation parameters.

```
python upside_config.py                                  \
  --output            pdbname.h5                         \
  --fasta             pdbname.fasta                      \
```

```
  --initial-structure  pdbname.initial.pkl                \
  --hbond-energy       $(cat UPSIDE_param_dir/ff_1/hbond) \
  --dynamic-rotamer-1body                                 \
  --rotamer-placement   UPSIDE_param_dir/ff_1/sidechain.h5\
  --rotamer-interaction UPSIDE_param_dir ff_1/sidechain.h5\
  --environment UPSIDE_param_dir/ff_1/environment.h5      \
  --rama-sheet-mixing-energy $(cat UPSIDE_param_dir/ff_1/sheet)\
  --rama-library  UPSIDE_param_dir/common/rama.dat        \
  --reference-state-rama
UPSIDE_param_dir/common/rama_reference.pkl                \
  --membrane-thickness   membrane_thickness               \
  --membrane-potential   membrane_potential_fpath         \
  --ask-before-using-AFM AFM_fpath                        \
  --AFM-time-initial     0
```

AFM_fpath is the path to the file that defines the residue to which force will be applied, tip position, spring constant and pulling velocity. Force can be applied to one or more residues. In *Upside*, the unit of the energy is $k_B$T: 1 $k_B$T ≈ 4.114 pN•nm at T = 1.0 (≈ 298 K). The unit of the spring constant is $k_B$T/Å$^2$: 1 $k_B$T/Å$^2$ ≈ 41.14 pN/Å = 411.4 pN/nm. 1 *Upside* time step ≈ 0.1 ns, so the pulling velocity 0.001 Å/*Upside* time step ≈ $10^6$ nm/s, the same as the extraction velocity in the CG-MD simulations (21).

(3) Run *Upside*.

```
upside pdbname.h5              \
  --seed           random_seed \
  --temperature    temperature \
  --frame-interval frame_intvl \
  --duration       duration    \
  --disable-recentering
```

**Simulation configurations of force clamp simulations.**

The only difference with the configuration above is in the preparation of the H5 file. A tension file is supplied to *Upside* instead of an AFM file, which defines the pulling residue and pulling force. One or more residues can be pulled.

```
python upside_config.py                                   \
  --output               pdbname.h5                       \
  --fasta                pdbname.fasta                    \
  --initial-structure    pdbname.initial.pkl              \
  --hbond-energy    $(cat UPSIDE_param_dir/ff_1/hbond)    \
  --dynamic-rotamer-1body                                 \
  --rotamer-placement UPSIDE_param_dir /ff_1/sidechain.h5 \
  --rotamer-interaction UPSIDE_param_dir ff_1/sidechain.h5\
  --environment       UPSIDE_param_dir/ff_1/environment.h5\
  --rama-sheet-mixing-energy $(cat UPSIDE_param_dir/ff_1/sheet)\
  --rama-library   UPSIDE_param_dir/common/rama.dat       \
  --reference-state-rama
UPSIDE_param_dir/common/rama_reference.pkl                \
```

```
  --membrane-thickness   membrane_thickness               \
  --membrane-potential   membrane_potential_fpath         \
  --tension              tension_fpath
```

**Fasta sequences of the proteins used in the study.**

**1. bR (1qhj.pdb)**

```
QAQITGRPEWIWLALGTALMGLGTLYFLVKGMGVSDPDAKKFYAITTLVPAIAFTMYLSMLLGY
GLTMVPFGGEQNPIYWARYADWLFTTPLLLLDLALLVDADQGTILALVGADGIMIGTGLVGALT
KVYSYRFVWWAISTAAMLYILYVLFFGFTSKAESMRPEVASTFKVLRNVTVVLWSAYPVVWLIG
SEGAGIVPLNIETLLFMVLDVSAKVGFGLILLRSRAIFGEAEAPEPSAGDGAAATS
```

**2. GlpG (2xov.pdb)**

```
ERAGPVTWVMMIACVVVFIAMQILGDQEVMLWLAWPFDPTLKFEFWRYFTHALMHFSLMHILFN
LLWWWYLGGAVEKRLGSGKLIVITLISALLSGYVQQKFSGPWFGGLSGVVYALMGYVWLRGERD
PQSGIYLQRGLIIFALIWIVAGWFDLFGMSMANGAHIAGLAVGLAMAFVDSLN
```

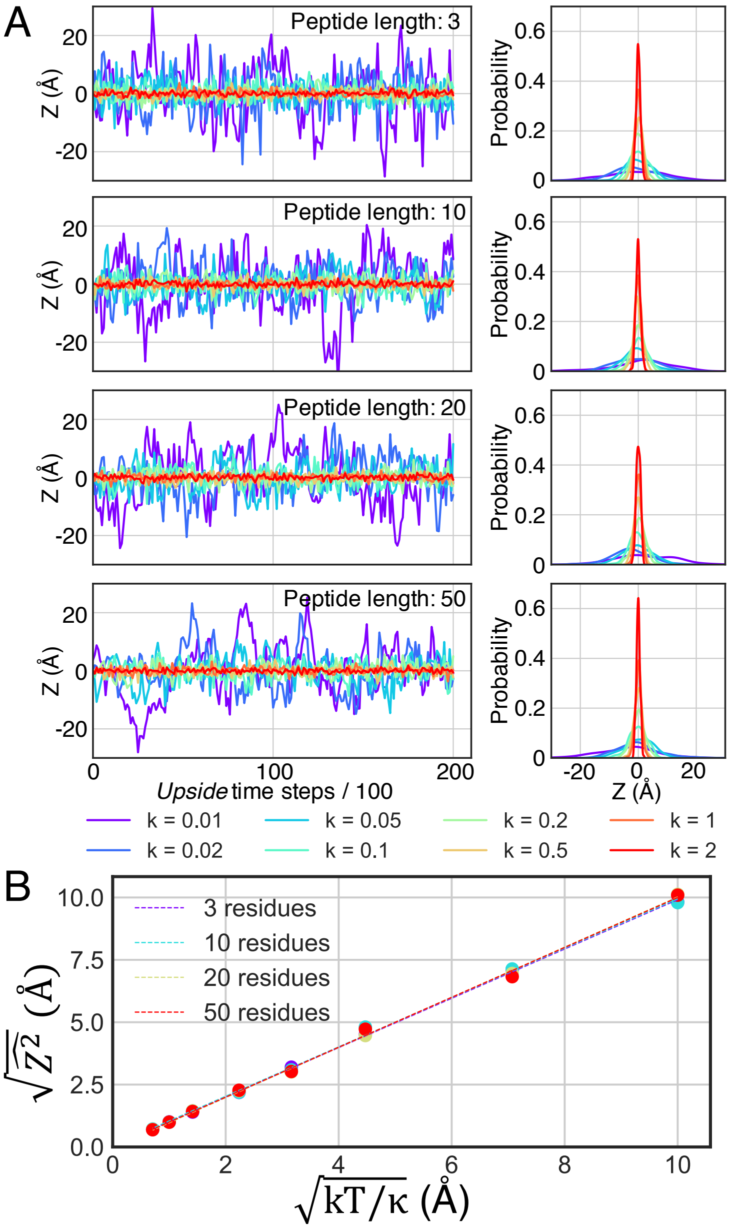

**Fig. S1. Calibrating our virtual cantilever: Stiffness, thermal fluctuations, and the equipartition theorem. A.** Thermal fluctuations and their distributions. A 3-50 residue segment of bR is attached to the tip of the cantilever and the fluctuations of the residue attached at the end to the cantilever are measured (i.e., same location, but with varying mass). **B.** In agreement with the equipartition theorem, $<z^2> = k_BT/\kappa$. The spring constant ($\kappa$) is in *Upside* unit: $k_B$T/Å$^2$.

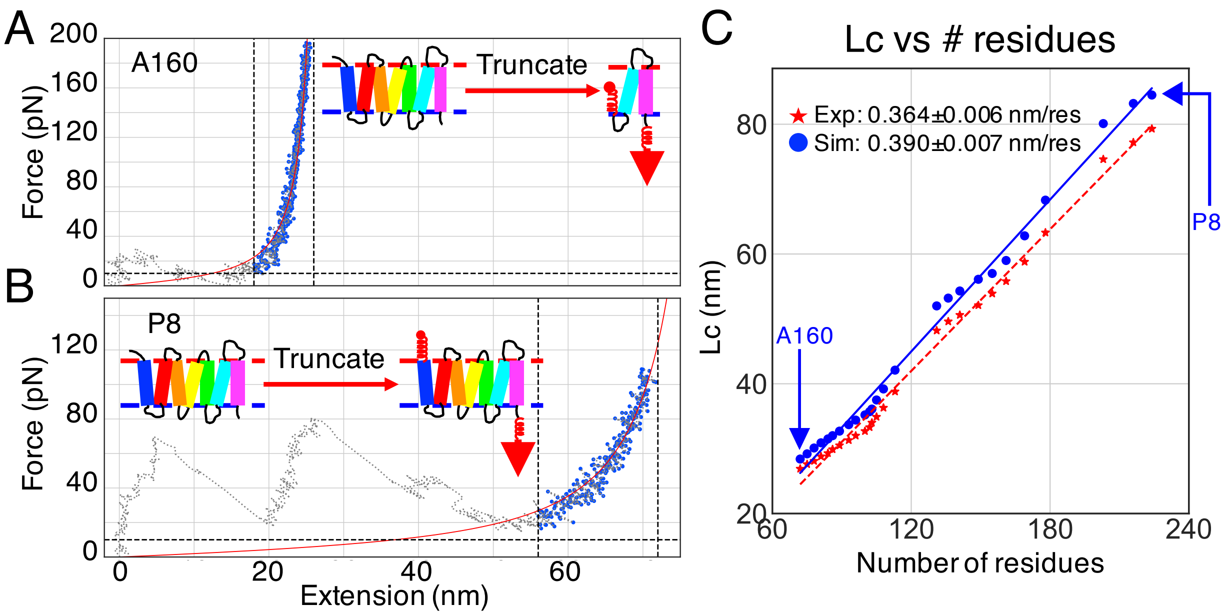

**Fig. S2. Calibration of contour length (Lc) per amino acid. A, B.** WLC fitting of the end-to-end distance (extension) and the force using a fixed Lp of 0.4 nm of truncated bR species A160 and P8, respectively. WLC fitting (curve in red) was performed on the data points in blue between the vertical black dashed lines (≥ 10 pN). The elasticity of the unfolded segment is well described by the WLC model. Note, during the initial (low force) portion, helices are being pulled out so the WLC is not applicable. Specifically, in **panel A**, the grey portion of the force-extension curve depicts the unfolding and pulling out of helix pair GF (similarly as shown in **Figs. 2B and S5**, at extension between 0 ~ 20 nm). After both helices G and F have been pulled out, the spring attached at the N-terminus (residue A160 in this case) starts getting extended, and thereafter we can see the WLC behavior (blue dots). **C.** Lc of unfolded segment as a function of number of residues from simulations compared to experiment (13). In the experiment, the number of unfolded amino acids is calculated based on $n_{aa} = (\Delta L_0 + \Delta d)/L_0^{aa}$, where Δd is the vertical distance of the folded structure along the pulling axis in native bR(22) and $L_0^{aa} = 0.366$ nm is the Lc per amino acid based the distance between the 1st intermediate in the helix pair ED (A160) and the 1st in the helix A (V29). The same deviations from linearity are observed in the simulations and experiment. This similarity implies that there is a similar sequence dependence that both highlights the accuracy of the simulations, and that the residue dependence could be useful in identifying the sequence of the segment that is unfolded for a given Lc value.

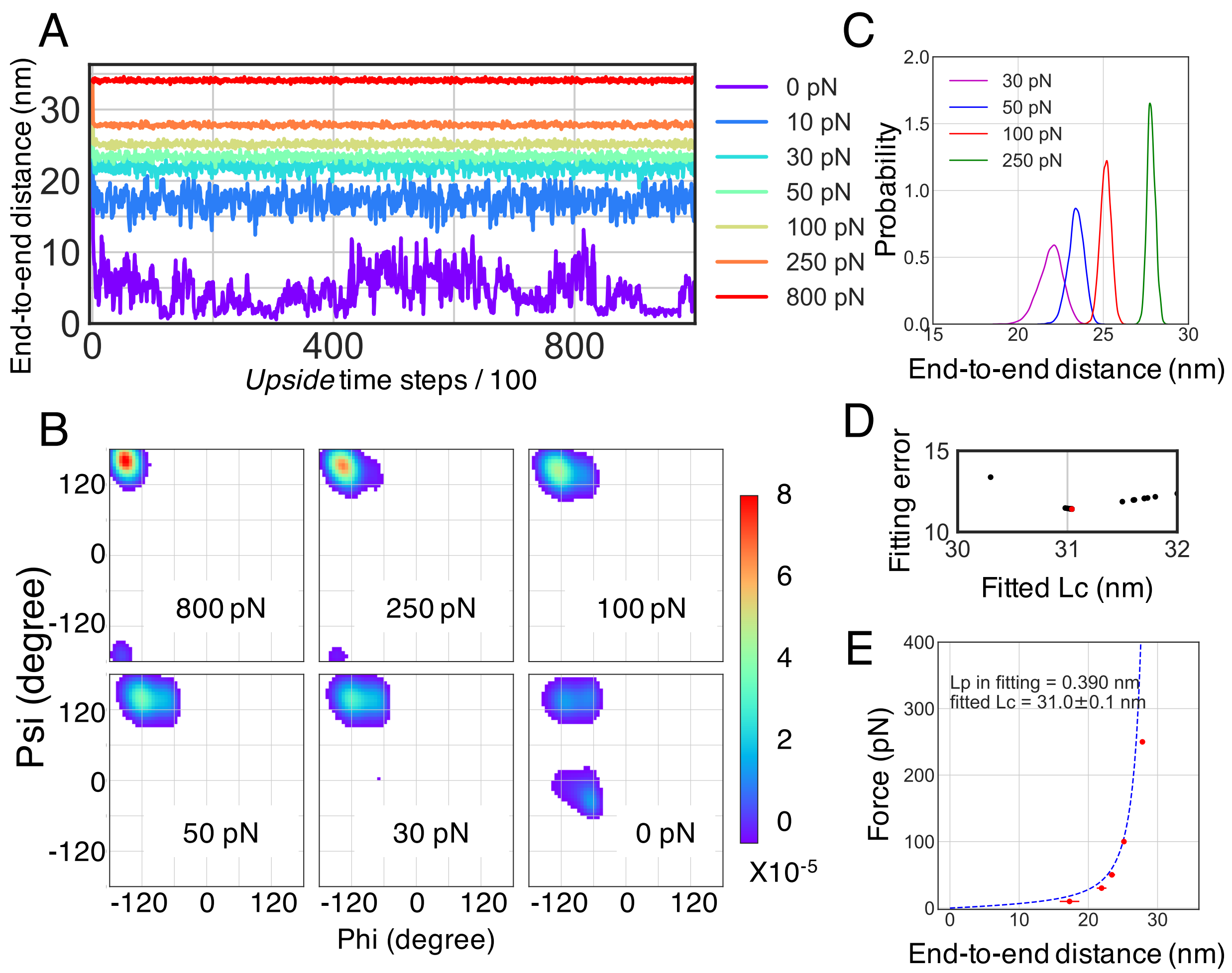

**Fig. S3. Reproduction of the all-atom MD of unfolded ubiquitin. A.** End-to-end distances of the protein under force. **B.** Distribution of ϕ and ψ dihedral angles of all 76 residues of the protein over time (in comparison to Fig. 2A of ref. (16)). Noticeably, there are populations of (ϕ, ψ) in the α-helix and left-handed helix regions under lower forces (100 pN, 50 pN, and 30 pN) in ref. (16). This is because of a specific version of force filed was used in the that study, in which the dihedral potentials are modified (23). In other words, our (ϕ, ψ) distributions would agree with ref. (16) if we had tuned our dihedral potentials. Nonetheless, (not) tuning the dihedral potentials does not affect the non-bonded interactions between side chains, as pointed out by ref. (16) (in the section "modified dihedral potentials" in SI of the paper), hence affect the end-to-end distances and the contour length. **C.** Probability density distribution of end-to-end distances (in comparison to Fig. S2 of ref. (16)). **D.** Fitting error versus fitted Lc. The minimum fitting error was obtained at Lc = 31.0 nm (red dot). **E.** WLC fitting of the average end-to-end distance and the applied force (in comparison to Fig. 1C of ref. (16)). The error bars show the standard deviation of the end-to-end distances. When fitting the data to obtain the Lc value, the value of Lp was fixed at 0.39 nm. Our fitted value for Lc is 31.0 nm, ~ 9% larger than 28.4 nm in ref. (16). The deviation of the force-extension to the WLC fitting is bigger at lower force (< 50pN) due to the interactions of non-neighboring residues when the protein collapses under low force, in which regime the chain behavior cannot be described by ideal chain models.

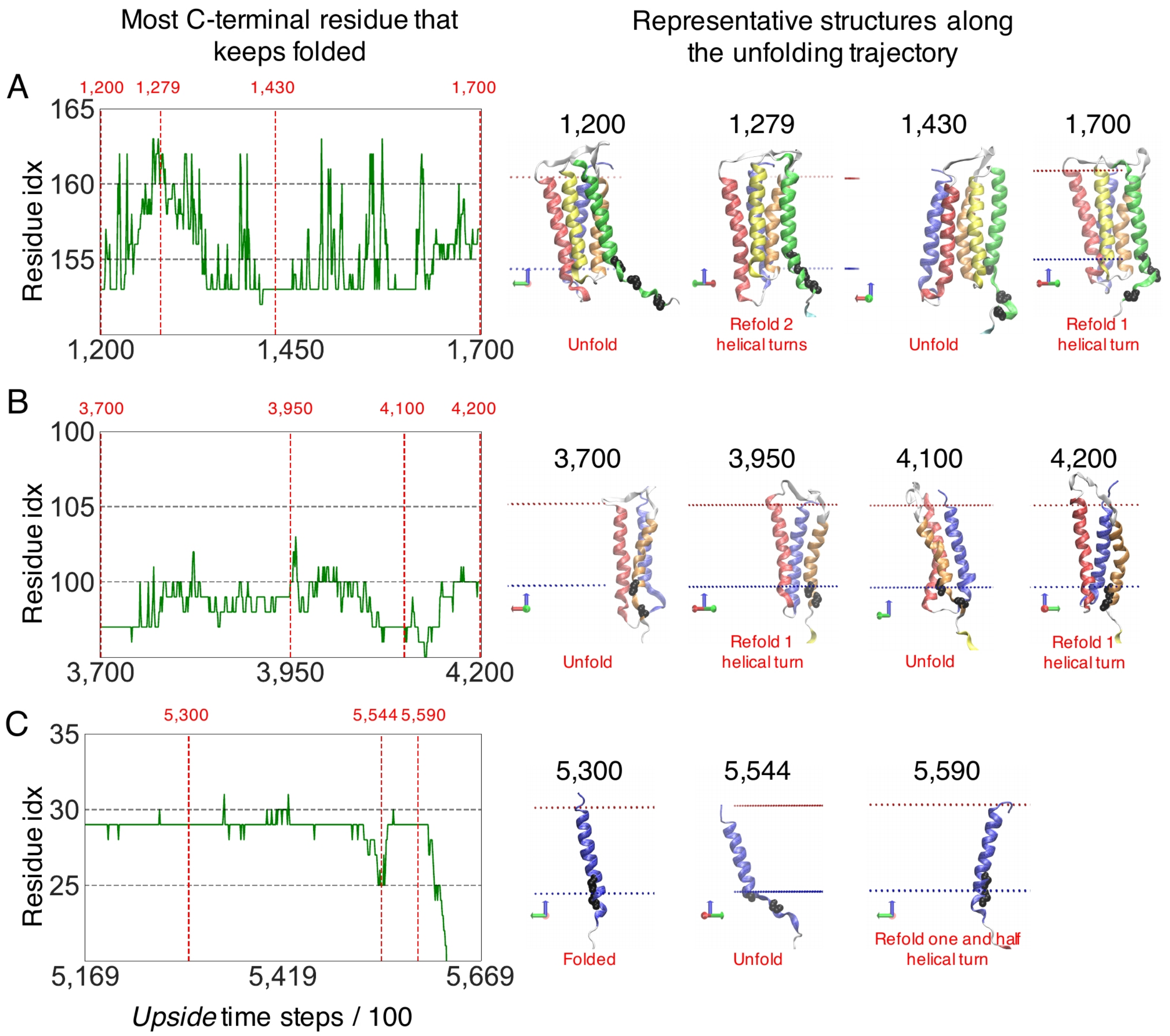

**Fig. S4. Near equilibrium back-and-forth transitions between intermediates in helix pair ED (panel A), helix pair CB (panel B), and helix A (panel C).** The inset in **panel C** of **Fig. 2** is enlarged and shown in the left column. In addition, one segment of TSS in the unfolding regions ED (from frame number 1200 to 1700) and one segment in A (from frame number 5169 to 5669) are shown. Representative structures are selected at the red vertical dashed lines and shown in the right column. The frame indices of those structures are above each plot. The TM helices are colored in the same set of codes as in **Fig. 1A**. Residues 161, 157 and 153 in structures in **panel A**, residues 100 and 97 in **panel B**, and residues 29 and 25 in **panel C** are plotted in the VDW presentation in black. The rapid back-and-forth transitions, which are considered as a hallmark of near-equilibrium measurement (13), can be seen in all three major unfolding regions in our simulations.

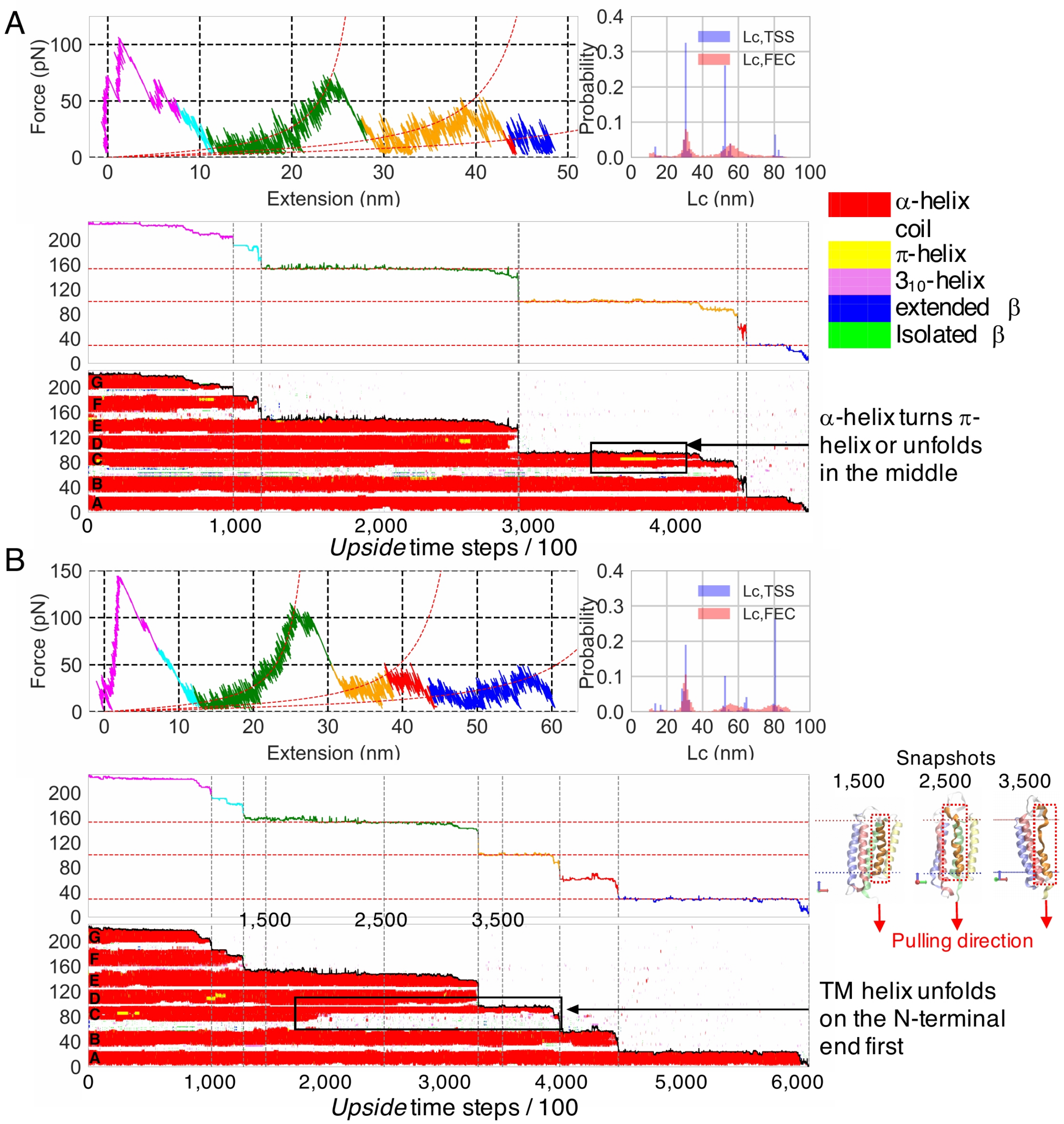

**Fig. S5. Unfolding trajectories of bR.** The trajectories largely support the common assumption that secondary structures remain intact within the membrane bilayer during the unfolding process. However, exceptions can be seen in these two TSS: **A.** Part of a TM helix may turn into π-helix or unfold in the middle (e.g., Helix C, black box in the TSS plot). **B.** A TM helix can unfold from the N-terminal end rather than the C-terminal end (e.g., Helix C, black box in the TSS plot). Snapshots taken at the 1500th, 2500th, are 3500th frames are shown. Helix C in orange (in red dashed box) unfolds at its N-terminal end first.

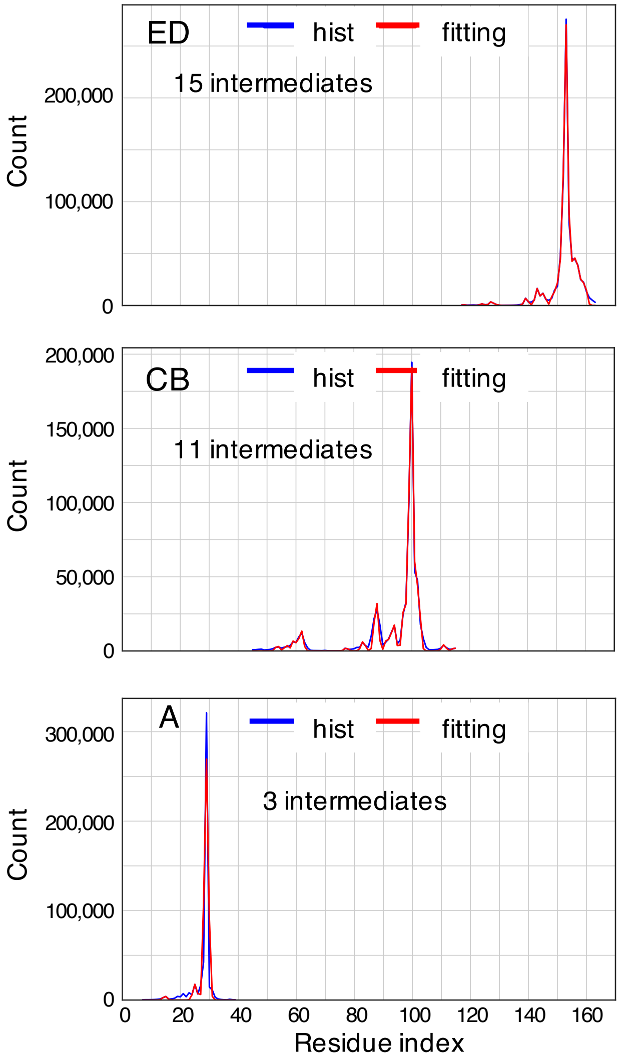

**Fig. S6. Identifying intermediates by fitting with multiple Gaussian functions.** For the 48 trajectories shown in **Fig. 3A**, the time spent at each position is histogrammed (blue lines) and fit using multiple Gaussians (red) with standard deviation ($\sigma$) of one residue to identify the number and position of intermediates. The number and initial position of the Gaussians was manually adjusted to minimize the fitting error; additional Gaussians were added until the error plateaued. The upper, middle and lower panels refer to unfolding occurring within the ED, CB or A helices, respectively. The index refers to the last residue that remains folded, as identified in the TSS, and is listed in **Table S2**.

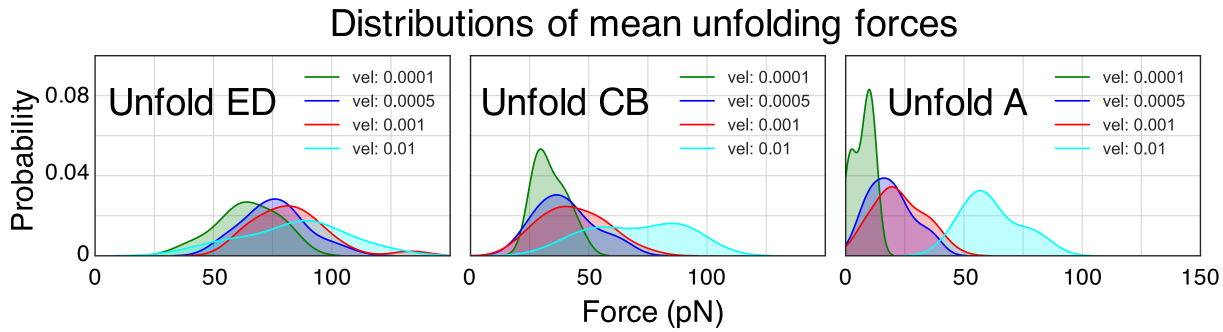

**Fig. S7. Distributions of mean unfolding forces at different pulling velocities.** The mean unfolding forces of the major intermediates in helix pairs ED, CB, and helix A (**Table S1**) are computed, respectively, from the trajectories simulated at each pulling velocity. Only the trajectories in which we observe all 7 helices are unfolded in the order ED, CB, and A, are included in the analysis. The velocities are in unit Å/*Upside* time step.

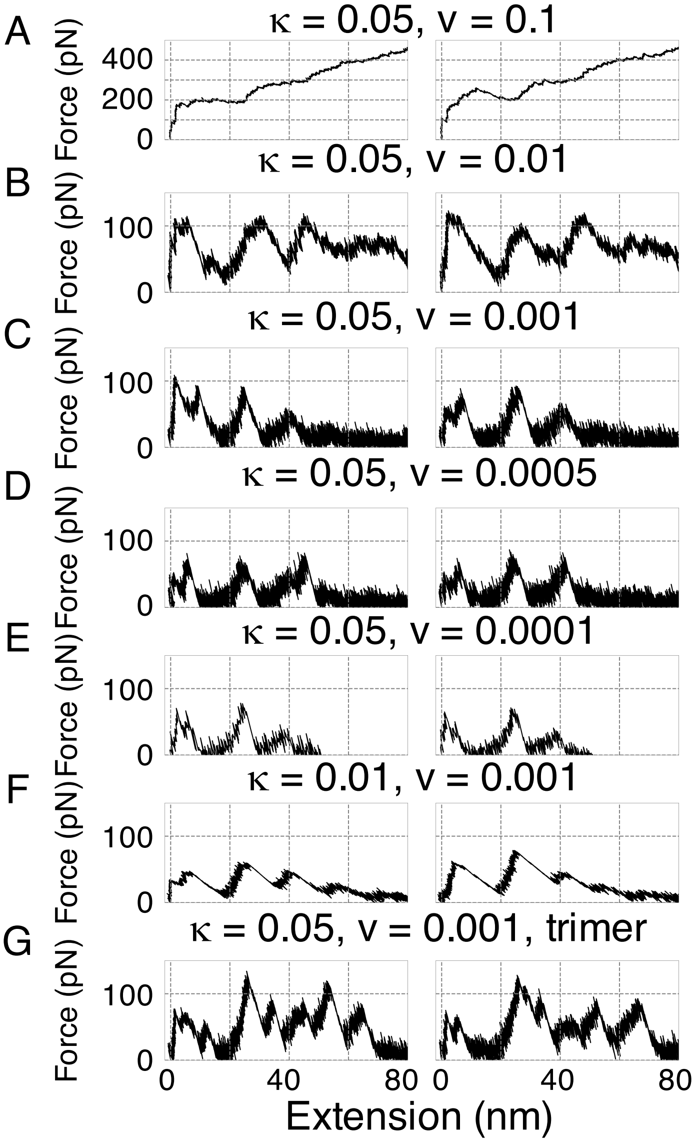

**Fig. S8. Force-extension curves (FECs) of unfolding trajectories of monomeric (A to F) and trimeric bR (G) under different conditions.** Two trajectories are shown for each set of simulation conditions. Details of the simulation setup are listed in **Table S5**. The unit of the spring constant is $k_B$T/Å$^2$: 1 $k_B$T/Å$^2$ ≈ 41.14 pN/Å = 411.4 pN/nm at T ≈ 300K. The pulling velocity 0.001 Å/*Upside* time step ≈ $10^6$ nm/s. WLC behavior is not observed at the highest pulling velocity (**panel A**). At higher pulling velocity (**panels A, B**), force does not restore to zero after the entire protein is pulled out of the membrane bilayer due to the friction of the solvent. As the pulling velocity decreases (from 0.01 to 0.0001, **panels B to E**), the saw-tooth pattern in general has the same depth (the ruptures have similar slopes). When the spring constant decreases (from 0.05 to 0.01), the saw-tooth patterns become shallower, i.e. the slope of the rupture is decreased (**panels C, F**). More intermediates can be observed in the FECs of trimer (**panel G**, **Tables S2**).

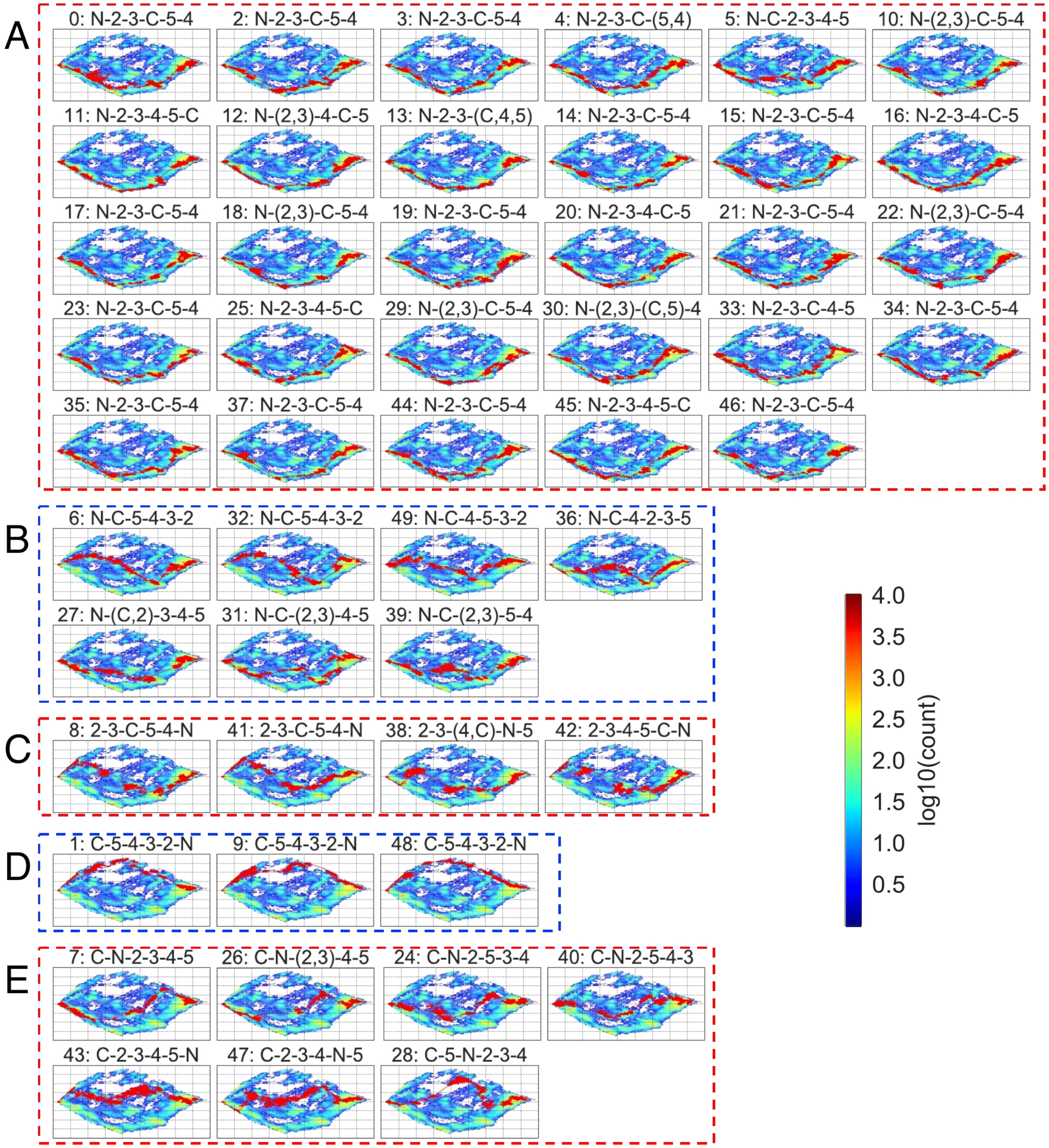

**Fig. S9. Unfolding trajectories of GlpG obtained with a stiff cantilever** (T = 0.9 ≈ 270 K, κ = 0.05 $k_B$T/Å$^2$, v = 0.001 Å/*Upside* time step; pulling the C-terminus and fixing the N-terminus with an equal strength spring). The heat map is obtained from 50 trajectories. The red curve of each subplot is the unfolding pathway from the native state to the fully extended (FE) state for a given trajectory. The trajectories are categorized based on their unfolding pathways. The title of each subplot indicates the index of the trajectory and its unfolding pathway. For example, "4: N-2-3-C-(5,4)" denotes the unfolding pathway of the fourth trajectory where the unfolding order is N→2→C→(5, 4), in which TM5 and TM4 unfold nearly simultaneously and therefore are put in parentheses.

**A.** Unfolding starts from TM1 (N) and proceeds to the C-domain when all the TM helices in the N-terminal domain unfold. The pathways traverse the lower edge of the PCA plots.

**B.** Unfolding starts from TM1, followed by the unfolding of TM6, leading to zigzag pathways across the PCA plots.

**C.** Unfolding starts from the middle of GlpG.

**D.** Unfolding starts from TM6 and proceeds to the N-domain when all the TM helices in the C-domain unfold. In contrast to **panel A**, the pathways flank the upper edge of the PCA plots.

**E.** Similar to **panel B**, with unfolding starting from TM6, followed by the unfolding of a TM helix in the N-domain, which results in zigzag pattern through the middle on the PCA plot.

The ratio of unfolding from the N-domain first to unfolding from the C-domain first is 40:10 (number of trajectories in panels **A**+**B**+**C** : number of trajectories in panels **D**+**E**).

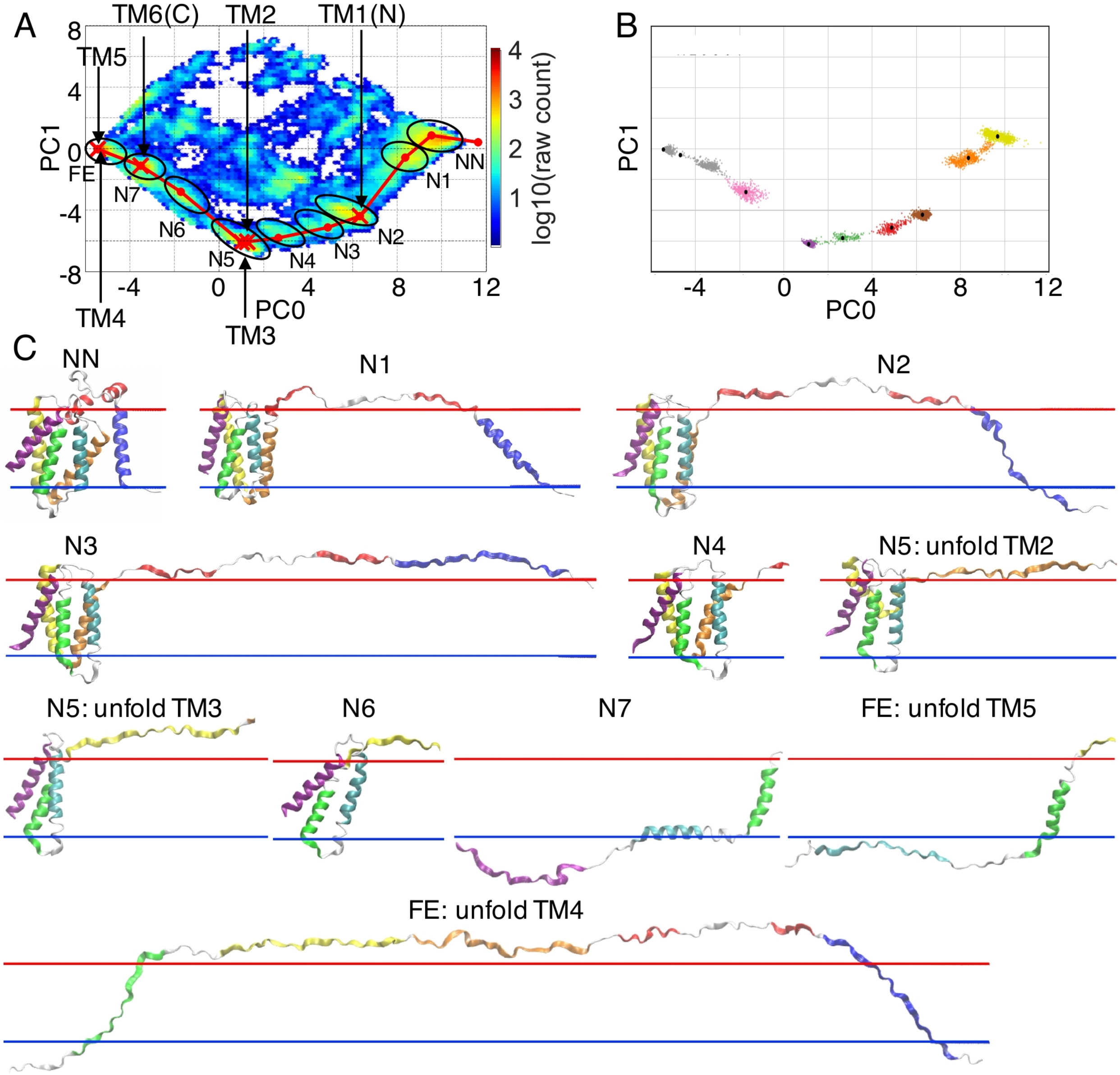

**Fig. S10. Examples of an N→C unfolding pathway of GlpG. A.** Unfolding pathway connected by representative structures on the PCA plot. The representative intermediates are chosen such that they are either the cluster center or the structure when a TM helix unfolds. These structures are considered as the intermediates. We use NN, N1, …, N7, and FE to denote the clusters as well as the intermediates. NN is short for near-native, and FE fully-extended. **B.** Clustering analysis of the trajectory. Nine clusters are identified. **C.** Snapshots of the representative structures. For the illustrative reasons, unfolded segments sometimes are not shown in the snapshots when there is no significant conformational change.

In the NN state, helices rearrange. In N1, the two interfacial helices H1, H2 unfold and separate. In N2, TM1 (N) unfolds. In N3, TM1 flips to the other side of bilayer. In N4, TM4 partially unfolds in its C-terminal. In N5, TM2 and TM3 unfold. In N6, the helices in the C-domain rearrange. In N7, TM6 (C) unfolds, and the C- terminus of TM4 refolds. In FE, TM5 and TM4 unfold, unfolded TM1 may re-enter the bilayer.

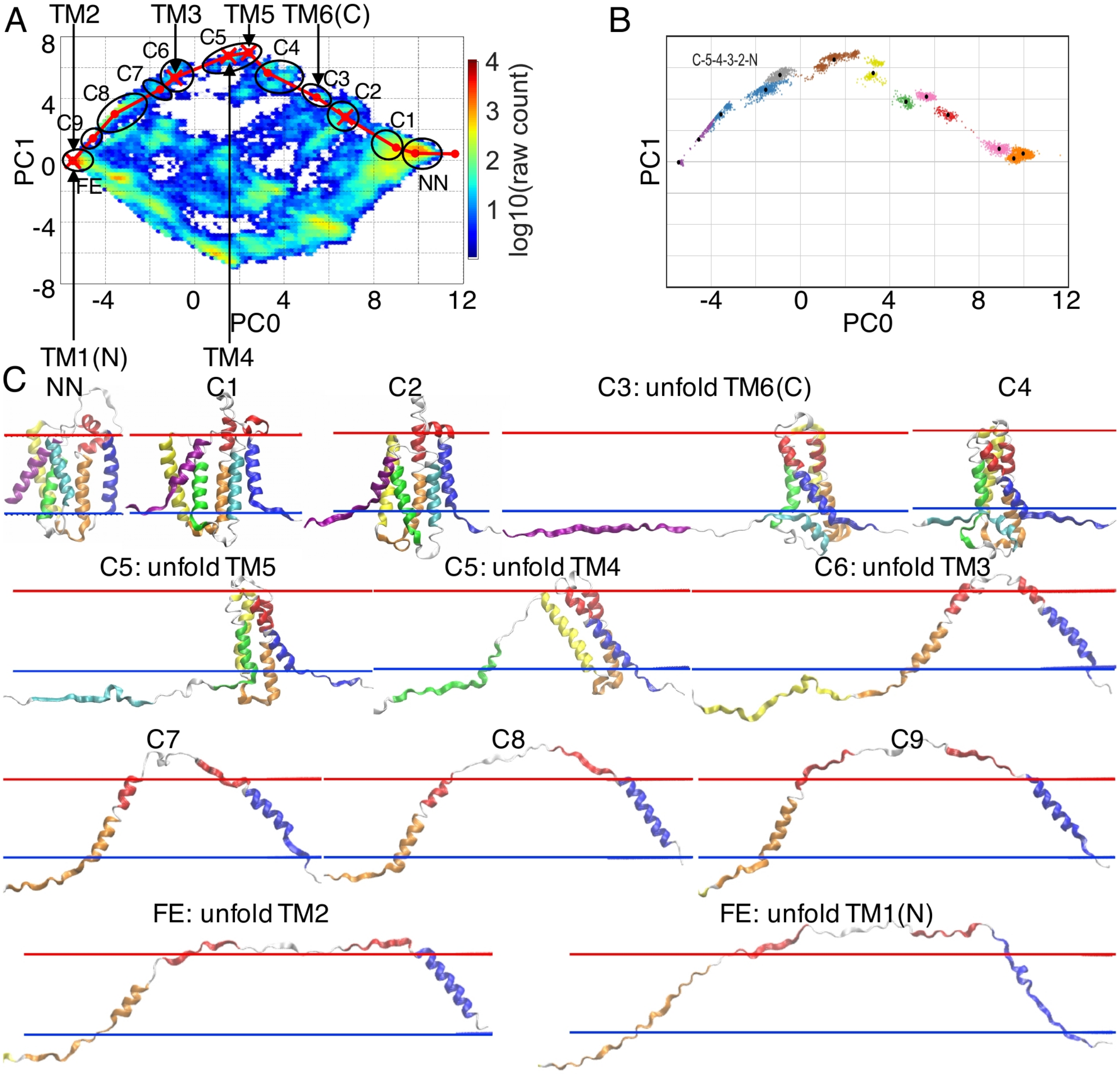

**Fig. S11. Example of a C→N unfolding pathway of GlpG. A.** Unfolding pathway connected by representative structures on the PCA plot**.** We use NN, C1, …, C9, and FE to denote the clusters as well as the intermediates. **B.** Clustering analysis. Thirteen clusters are identified. **c.** Snapshots of the representative structures.

In the NN state, helices rearrange. In C1, the interfacial helix H2 aligns with TM2 and pushes part of TM2 out of the bilayer, TM2 bends, TM1 (N) and TM6 (C) partially unfold, and H1 unfolds. In C2, TM6 unfolds one more helical turn, TM2 partially unfolds in its C-term and H1 refolds. In C3, TM6 unfolds, TM5 comes out of the bilayer, H1 aligns with TM1, and TM1 tilts in order to accommodate the elongation in its C-term due to the alignment of H1. In C4, two more helical turns of TM1 unfold, and TM4 and TM5 partially unfold. In C5, TM5 and TM4 unfold. In C6, TM3 unfolds, TM1 and TM2 come apart. In C7, H1 unfolds. In C8, TM1 and TM2 come further apart. In C9, H2 unfolds. In FE, TM2 and TM1 unfold.

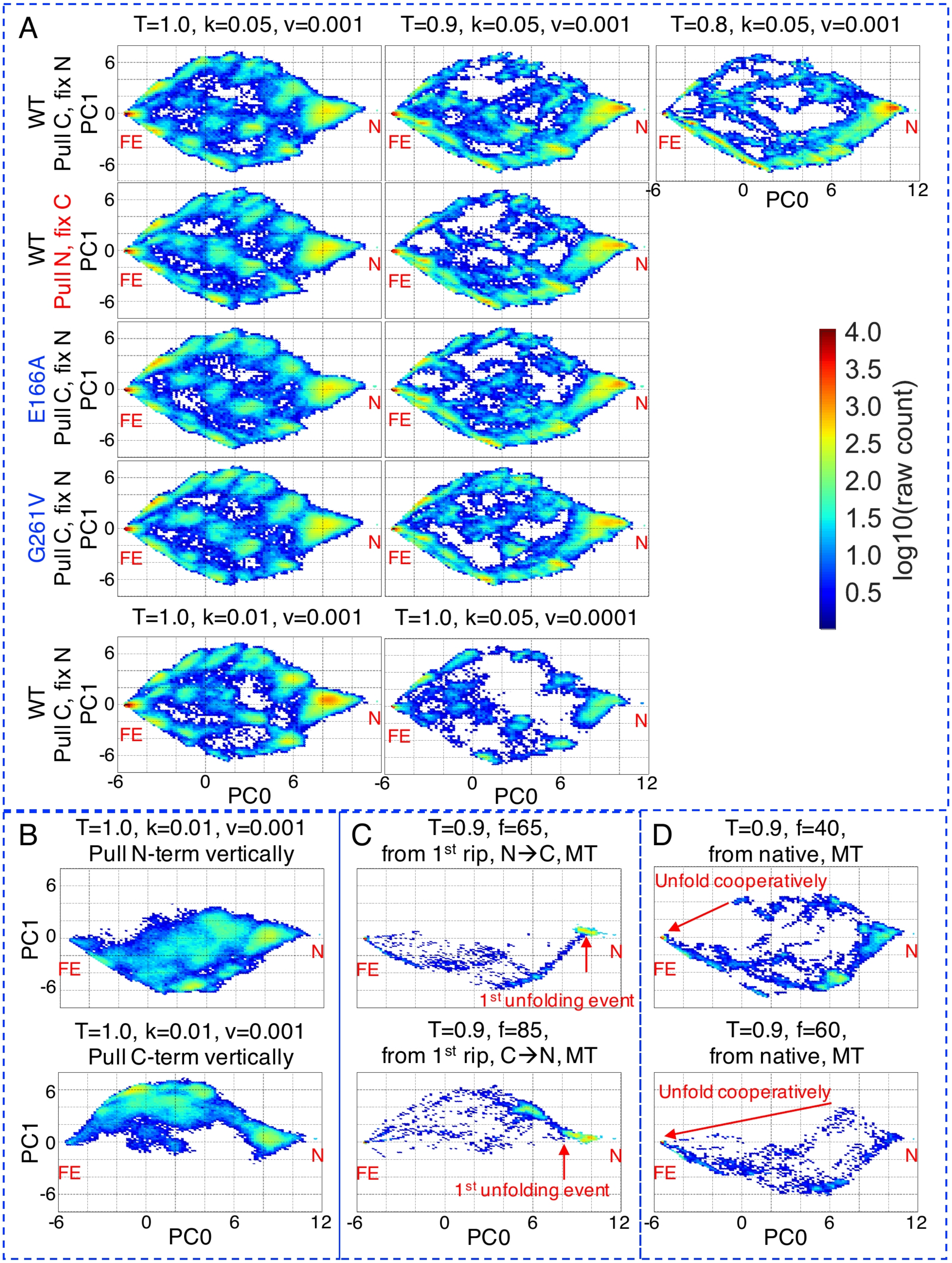
A
T=1.0, k=0.05, v=0.001
T=0.9, k=0.05, v=0.001
T=0.8, k=0.05, v=0.001
WT
Pull C, fix N
WT
Pull N, fix C
E166A
Pull C, fix N
G261V
Pull C, fix N
T=1.0, k=0.01, v=0.001
T=1.0, k=0.05, v=0.0001
WT
Pull C, fix N
PC1
PC0
FE
N
6
0
-6
12
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
log10(raw count)
B
T=1.0, k=0.01, v=0.001
Pull N-term vertically
T=1.0, k=0.01, v=0.001
Pull C-term vertically
C
T=0.9, f=65,
from 1st rip, N→C, MT
1st unfolding event
T=0.9, f=85,
from 1st rip, C→N, MT
1st unfolding event
D
T=0.9, f=40,
from native, MT
Unfold cooperatively
T=0.9, f=60,
from native, MT
Unfold cooperatively

**Fig. S12. Principal component analysis of unfolding trajectories of GlpG under various simulation protocols.**
**A.** Stiff cantilever mode, pulling laterally. Each of the PCA plots is comprised of 50 trajectories, except for the trajectories using a 10x slower pulling velocity (v = 0.0001 Å/*Upside* time step), which contains only 10 trajectories. Despite differences, the fundamental heterogeneous pathway behavior remains.

**B.** Stiff cantilever mode, force is applied to either the N- or C-terminus vertically. The PCA plots for pulling at C- and N-terminus contains 19 or 20 trajectories. Notably, the PCA heat maps obtained in this mode fill in the blanks in the middle of the heat maps obtained in **panel A**. Those may represent structures that largely maintain the tertiary structure for the region embedded in the membrane, which would be difficult to observe in mode **A** because the tertiary structure is disrupted. Besides, the pulling the N-terminus produces "deterministically" N→C pathways as expected, and vice versa.

**C.** Modified MT mode simulations, pulling laterally, were re-started at the 1st unfolding event in the N→C or C→N pathway ($T = 0.9 \approx 270$ K, $\kappa = 0.05\ k_B T/\text{Å}^2$, v = 0.001 Å/*Upside* time step).

**D.** MT mode, pulling laterally, simulations were started from the native structure. The unfolding is more cooperative under higher force and in the C→N pathway than the reverse. 20 trajectories are included in each PCA plot in **panels C and D**. N and FE stand for native and fully-extended in each subplot, respectively.

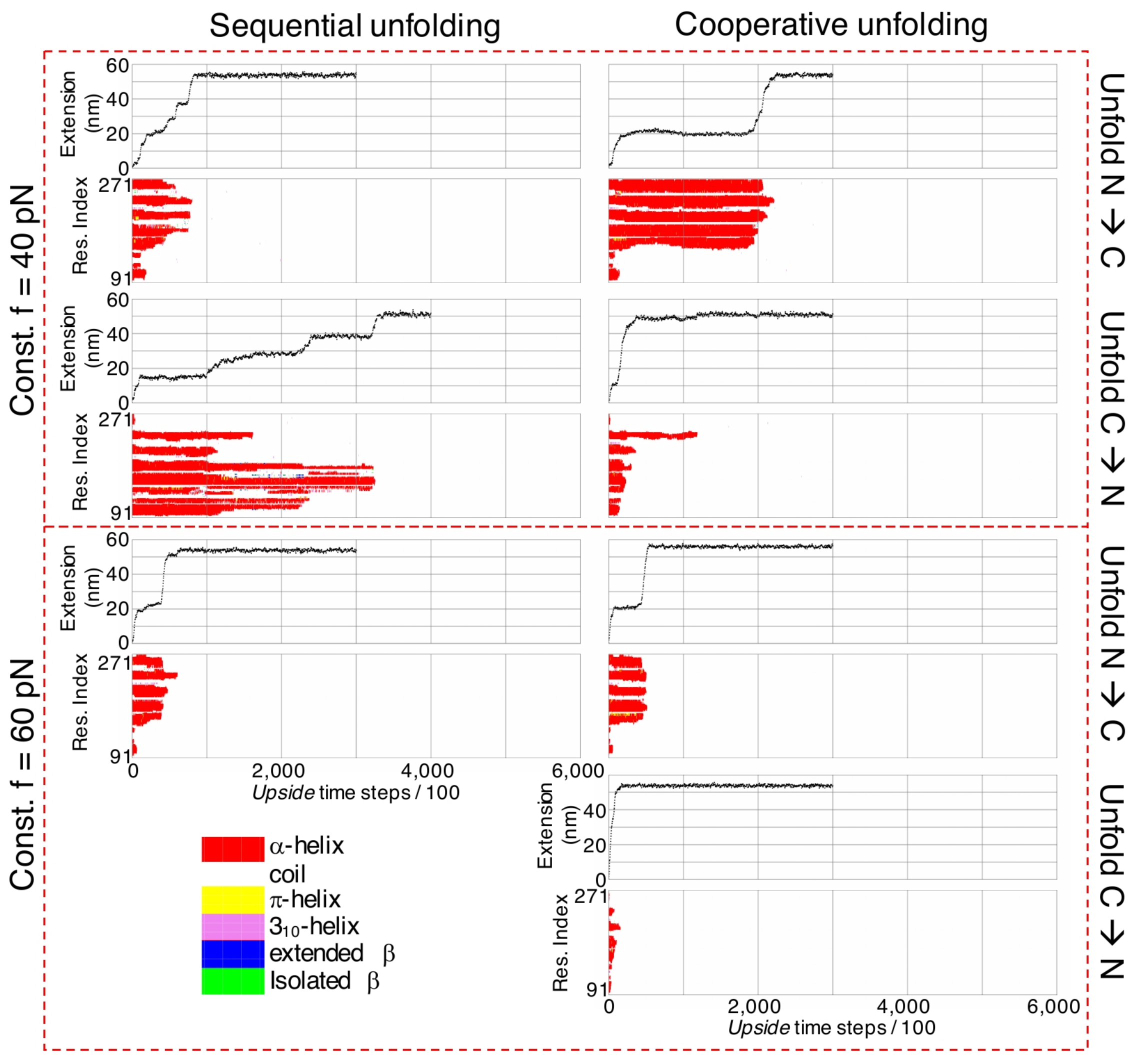

**Fig. S13. Sequential and cooperative unfolding pathways of GlpG in force clamp simulations started from the native structures.** Four trajectories under a constant force of 40 pN (upper panel) and three trajectories under a constant force of 60 pN (lower panel) are shown. Every trajectory is presented by a FEC and a TSS plot. Trajectories in the left column are examples of sequential unfolding pathways (with at least 3 intermediates that can be identified on the extension plot) whereas those in the right column are examples of cooperative unfolding pathways (with no more than 2 intermediates identified from the extension plot). In the 20 simulations under 60 pN, we did not observe any trajectories unfolding from C- to N-domain sequentially.

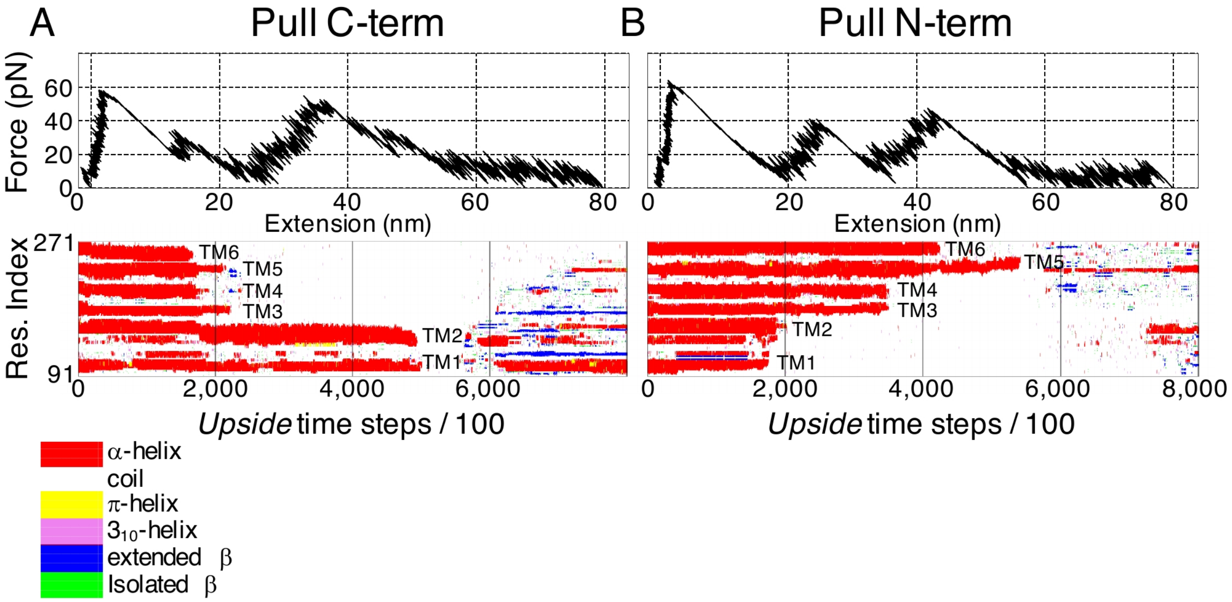

**Fig. S14. Unfolding GlpG by pulling vertically in AFM, stiff cantilever mode. A, B.** FEC and TSS plots of an example trajectory pulling from the C- and N-terminus, respectively. 20 simulations were performed in each case. We observed that all TM helices become completely unfolded in 4 and 3 trajectories when pulling on the C- and N-terminus, respectively. After all the protein is pulled out, the extended chain starts collapsing and forms H-bonds again. Notably, TM6 unfolds before TM5 (**panel B**), implying that TM6 is not very stable by itself in the lipid bilayer.

**Table S1. Comparison of the mean unfolding force (in pN) and the s.e.m. for bR intermediates observed in experiment (13) and simulations.** The spring constant ($\kappa$) is in $k_B$T/Å$^2$; the pulling velocity ($v$) is in Å/*Upside* time step; and the temperature ($T$) is in *Upside* temperature unit ($1 \approx 300$ K). Numbers in parenthesis in the 1st column indicate the number of trajectories in which we observe all 7 helices are unfolded in the order ED, CB, and A versus the total number of trajectories we have simulated. The major intermediate identified in each region is put in the parentheses, as indexed by the last folded residue, in the 2nd to 4th columns. The two rows highlighted in light green are the results presented in the main text.

| | Major intermediate in ED helix pair (residue index) | Major intermediate in CB helix pair (residue index) | Major intermediate in helix A (residue index) |
|---|---|---|---|
| Experiment | 94±1 (A160) | 49±2 (V101) | 62±0.6 (V29) |
| $\kappa = 0.05$, $v = 0.1$ (16 / 20) | 218.5±3.8 (F153) | 311.4±4.2 (L99) | 443.8±5.8 (V29) |
| $\kappa = 0.05$, $v = 0.01$ (11 / 20) | 83.9±6.4 (F153) | 70.2±5.8 (L99) | 61.5±3.5 (V29) |
| $\kappa = 0.05$, $v = 0.001$ (48 / 90) | 82.7±2.4 (F153 [a]) | 43.6±2.1 (L100) | 22.6±1.6 (V29) |
| $\kappa = 0.05$, $v = 0.0005$ (25 / 60) | 75.9±1.1 (F153) | 40.0±1.9 (L100) | 18.9±2.1 (V29) |
| $\kappa = 0.05$, $v = 0.0001$ (10 / 20) | 64.5±4.0 (F153) | 33.8±2.1 (L100) | 7.2±1.4 (V29) |
| $\kappa = 0.01$, $v = 0.001$ (46 / 80) | 69.7±1.4 (F153) | 41.2±1.0 (L100) | 22.5±0.9 (V29) |
| $\kappa = 0.05$, $v = 0.001$, trimer (43 / 45) | 88.7±3.0 (F153 [a]) | 74.5±4.4 (L100) | 57.5±2.7 (V29) |
| $\kappa = 0.05$, $v = 0.001$, MPx2 [b] (20 / 20) | 91.5±3.3 (F153) | 51.4±3.0 (L100) | 38.9±2.2 (V29) |

**a.** The comparison between experiment and simulations is conducted for the most populated intermediate, which is given in the parentheses; however, we also observe a K159 intermediate in the simulations which corresponds to the major experimental intermediate.

**b.** The membrane potential in this set of simulations is doubled, while all the other potentials keep unchanged. The increase of membrane potential significantly stabilizes the last helix, helix A, in the membrane bilayer, as the unfolding of helix A is observed in all 20 of 20 trajectories.

**Table S2. Comparison of bR intermediates identified in the 2017 experiment (13), our monomer and trimer simulations** (spring constant κ = 0.05 $k_B$T/Å$^2$, pulling velocity = 0.001 Å/*Upside* time step, T ~ 300 K) and **a 2016 CG study (21)** (see Fig. 5A in ref. (21), the intermediates are taken from the analysis of force peak groups, which were compared to previous experiments(24-26))**.** The position of an intermediate is indexed by the last folded residue of that intermediate in the protein.

| | 2017 experiment | Monomer simulations | Trimer simulations | 2016 CG study | Description |
|---|---|---|---|---|---|
| Intermediates in helix pair ED | 160 | 159 | 159 | | Top of helix E |
| | 157 | 157 | 157 | 156.8 | |
| | 154 | 155, 153 | 155, 153 | | |
| | 151 | 151 | 151 | 150.8 | |
| | 148 | 149 | 148 | | |
| | 146 | 145 | 145 | | |
| | 143 | 143 | 143 | | |
| | 139 | 139 | 139 | 140.7 | |
| | 136 | | 137 | | |
| | 132 | | | | |
| | 130 | | | | Bottom of helix E |
| | 129 | 129 | | | |
| | 127 | 127 | 127 | | Top of helix D |
| | 124 | 124 | 124 | | |
| | 119 | 118 | 119 | | |
| | | 115 | 114 | | |
| | | 111 | 111 | | |
| Intermediates in helix pair CB | 101 | 102, 100 | 100 | 101.4 | Top of helix C |
| | 96 | 97 | 98 | 95.4 | |
| | | 94 | 94 | | |
| | 91 | 92 | 92 | | |
| | | 88 | 88, 86 | 89.0 | |
| | 83 | 83 | 81 | | Bottom of helix C |
| | 77 | 77 | | 75.7 | |
| | 71 | | | | |
| | 63 | 62 | 62 | | Top of helix B |
| | | 57 | 59, 57 | | |
| | 54 | 54 | 55 | | |
| Intermediates in helix A | | | | 33.8 | |
| | 29 | 29 | 29 | 29.6 | Top of helix A |
| | | 25 | 25 | | |
| | | | 23, 21 | 21.2 | |
| | | | 18 | 19.1 | |
| | 16 | 15 | 15 | | |
| | 8 | | | | Bottom of helix A |

**Table S3. Summary of unfolding pathways of GlpG.**

| Pulling scheme | T | κ or F [a] | WT/mutant | N- → C-domain [b] | C- → N-domain [c] |
|---|---|---|---|---|---|
| Gradual-pulling simulations from the native structure | | | | | |
| Pull C-term, fix N-term | 1 | κ = 0.05 | WT | 27 | 23 |
| Pull C-term, fix N-term | 1 | κ = 0.05 | E166A | 27 | 23 |
| Pull C-term, fix N-term | 1 | κ = 0.05 | G261V | 22 | 28 |
| | | | | | |
| Pull C-term, fix N-term | 1 | κ = 0.01 | WT | 21 | 29 |
| | | | | | |
| **Pull C-term, fix N-term** [d] | **0.9** | κ = **0.05** | **WT** | **40** | **10** |
| Pull C-term, fix N-term [e] | 0.9 | κ = 0.05 | WT | 7 | 3 |
| Pull C-term, fix N-term | 0.9 | κ = 0.05 | E166A | 41 | 9 |
| Pull C-term, fix N-term | 0.9 | κ = 0.05 | G261V | 35 | 15 |
| | | | | | |
| Pull C-term, fix N-term | 0.8 | κ = 0.05 | WT | 42 | 8 |
| | | | | | |
| Pull N-term, fix C-term | 1.0 | κ = 0.05 | WT | 32 | 18 |
| Pull N-term, fix C-term | 0.9 | κ = 0.05 | WT | 43 | 7 |
| Force clamp simulations from the native structure [f] | | | | | |
| Pull N-term, fix C-term | 0.9 | F = 40 | WT | 16 | 4 |
| Pull N-term, fix C-term | 0.9 | F = 60 | WT | 13 | 7 |

**a.** κ or F is listed as relevant to the mode of applying force, gradual pulling or force, respectively.

**b.** The number of trajectories with an unfolding pathway that is initiated from the N-domain. For example, in **Fig. S9**, **A**, **B** and **C** are all counted as N- to C-domain unfolding pathways.

**c.** The number of trajectories with an unfolding pathway that is initiated from the C-domain. For example, in **Fig. S9**, **D** and **E** are both counted as C- to N-domain unfolding pathways.

**d.** The primary data set that is shown in the main text.

**e.** The only data set of gradual pulling that uses a pulling velocity = 0.0001 Å/*Upside* time step, 10x slower than all the other data sets of gradual pulling.

**f.** "From the native structure" refers to starting the simulation from the native structure instead of an intermediate at the first unfolding event.

**Table S4. Contour length (Lc) of bR intermediates.** The truncated bR is named after the structural position (i.e., residue index) of the intermediate, as defined by the last folded residue. Numbers in parentheses indicate the number of residues of the truncated bR molecules. For example, A160 has 72 residues. The Lc of the truncated bR in its fully extended state in simulation, and the Lc of unfolded segment of the corresponding intermediate in experiment are listed.

| Truncated bR | Lc, simulation (nm) (Fitted with Lp = 0.4 nm) | Lc, experiment (nm) | Description |
|---|---|---|---|
| A160 (72) | 28.4 | 26.9 | Top of helix E |
| T157 (75) | 29.2 | 27.6 | |
| F154 (78) | 30.1 | 28.1 | |
| V151 (81) | 30.9 | 28.8 | |
| I148 (84) | 31.5 | 29.3 | |
| L146 (86) | 32.0 | 29.9 | |
| A143 (89) | 32.7 | 30.5 | |
| A139 (93) | 33.7 | 31.3 | |
| V136 (96) | 34.4 | 32.0 | |
| S132 (100) | 35.2 | 32.7 | |
| V130 (102) | 35.7 | 33.3 | Bottom of helix E |
| K129 (103) | 36.1 | 34.0 | |
| L127 (105) | 37.5 | 34.9 | Top of helix D |
| V124 (108) | 39.2 | 36.3 | |
| I119 (113) | 42.1 | 38.8 | |
| V101 (131) | 52.0 | 48.2 | Top of helix C |
| D96 (136) | 53.2 | 49.6 | |
| P91 (141) | 54.3 | 50.6 | |
| Y83 (149) | 56.1 | 52.1 | Bottom of helix C |
| P77 (155) | 57.0 | 53.9 | |
| F71 (161) | 59.0 | 55.8 | |
| G63 (169) | 62.8 | 58.8 | Top of helix B |
| F54 (178) | 68.3 | 63.3 | |
| V29 (203) | 80.1 | 74.6 | Top of helix A |
| G16 (216) | 83.2 | 77.2 | |
| P8 (224) | 84.5 | 79.3 | Bottom of helix A |

**Table S5. Simulation details.** Parameters used in the *Upside* simulation are summarized in the table. Units for κ, v, F, and T are the same as in **Table S2**.

| System | Cantilever[a] | Attachments | κ | v | F | T | Number of simulations |
|---|---|---|---|---|---|---|---|
| Ubiquitin (fully-extended) [b] | Soft | Pull both termini in opposite direction | | | 0, 10, 30, 50, 100, 250, 800 | 1.0 | 1 per F |
| Truncated bR species | Stiff | Pull C-term vertically [c], fix N-term [d] | 0.05 | 0.001 | | 1.0 | 1 per species |
| bR | Stiff | Pull C-term vertically | 0.05 | 0.1 | | 1.0 | 20 |
| bR | Stiff | Pull C-term vertically | 0.05 | 0.01 | | 1.0 | 20 |
| bR | Stiff | Pull C-term vertically | 0.05 | 0.001 | | 1.0 | 90 |
| bR MPx2 [e] | Stiff | Pull C-term vertically | 0.05 | 0.001 | | 1.0 | 20 |
| bR | Stiff | Pull C-term vertically | 0.01 | 0.001 | | 1.0 | 80 |
| bR | Stiff | Pull C-term vertically | 0.05 | 0.0005 | | 1.0 | 60 |
| bR | Stiff | Pull C-term vertically | 0.05 | 0.0001 | | 1.0 | 20 |
| bR trimer [f] | Stiff | Pull C-term vertically | 0.05 | 0.001 | | 1.0 | 45 |

(To be continued in the next page.)

| System | Cantilever[a] | Attachments | κ | v | F | T | Number of simulations |
|---|---|---|---|---|---|---|---|
| GlpG | Stiff | Pull C-term laterally [c], fix N-term | 0.05 | 0.001 | | 0.8, 0.9, 1.0 | 50 per T |
| GlpG | Stiff | Pull N-term laterally, fix C-term | 0.05 | 0.001 | | 0.9, 1.0 | 50 per T |
| GlpG | Stiff | Pull C-term laterally, fix N-term | 0.05 | 0.0001 | | 0.9 | 10 |
| GlpG E166A | Stiff | Pull C-term laterally, fix N-term | 0.05 | 0.001 | | 0.9, 1.0 | 50 per T |
| GlpG G261V | Stiff | Pull C-term laterally, fix N-term | 0.05 | 0.001 | | 0.9, 1.0 | 50 per T |
| GlpG | Stiff | Pull C-term laterally, fix N-term | 0.01 | 0.001 | | 1.0 | 50 |
| GlpG intermediates before 1st rip | Soft | Pull C-term laterally, fix N-term | | | 64.6, 84.5 | 0.9 | 20 per F [g] |
| GlpG | Soft | Pull C-term laterally, fix N-term | | | 40, 60 | 0.9 | 20 per F |
| GlpG | Stiff | Pull C-term vertically | 0.01 | 0.001 | | 1.0 | 20 |
| GlpG | Stiff | Pull N-term vertically | 0.01 | 0.001 | | 1.0 | 20 |

**a.** Soft mode refers to the use of a very soft cantilever to mimic a magnetic tweezers measurement where the force is held essentially constant force after the first unfolding event occurs as the magnetic field varies slowly, on micron length scale, which is longer than the unfolded segments.

**b.** The simulations were started from a fully extended state.

**c.** The direction is relative to the membrane bilayer.

**d.** Held with an equally stiff spring.

**e.** The membrane potential in this set of simulations is doubled, while all the other potentials keep unchanged.

**f.** Chain A of the trimer is pulled, while the other two subunits of the trimer are allowed to undergo conformational changes freely.

**g.** One of the output file is corrupted, so there are only 19 trajectories useful for analysis.

**Table S6. Inferred Lc values (in nm) associated with each residue in helices E to A of bR.** Lc values obtained directly from the simulations of truncated bR (**Fig. S2**) are in red. The Lc value is for the unfolded segment C-terminal to a residue. For example, the contour length for the unfolded segment from the C-terminus to K159 (having residues 160 to 232) is 28.7 nm. Note, the last 4 residues (I4, Q3, A2, E1) are not in 1qhj.pdb, and therefore are not included in the simulations. In this case we don't observe the last intermediate A2 in the experiment (13).

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Helix E | **A160** | K159 | S158 | **T157** | F156 | G155 | **F154** | F153 | L152 | **V151** |
| | **28.4** | 28.7 | 28.9 | **29.2** | 29.5 | 29.8 | **30.1** | 30.4 | 30.6 | **30.9** |
| | Y150 | L149 | **I148** | Y147 | **L146** | M145 | A144 | **A143** | T142 | S141 |
| | 31.1 | 31.3 | **31.5** | 31.8 | **32.0** | 32.2 | 32.5 | **32.7** | 33.0 | 33.2 |
| | I140 | **A139** | W138 | W137 | **V136** | F135 | R134 | Y133 | **S132** | Y131 |
| | 33.4 | **33.7** | 33.9 | 34.2 | **34.4** | 34.6 | 34.8 | 35.0 | **35.2** | 35.4 |
| | **V130** | K129 | T128 | | | | | | | |
| | **35.7** | 36.1 | 36.8 | | | | | | | |
| Helix D | **L127** | A126 | G125 | **V124** | L123 | G122 | T121 | G120 | **I119** | M118 |
| | **37.5** | 38.1 | 38.6 | **39.2** | 39.8 | 40.4 | 40.9 | 41.5 | **42.1** | 42.6 |
| | I117 | G116 | D115 | A114 | G113 | V112 | L111 | A110 | L109 | I108 |
| | 43.2 | 43.8 | 44.3 | 44.8 | 45.4 | 46.0 | 46.5 | 47.0 | 47.6 | 48.2 |
| | T107 | G106 | Q105 | D104 | A103 | D102 | | | | |
| | 48.7 | 49.2 | 49.8 | 50.4 | 50.9 | 51.4 | | | | |
| Helix C | **V101** | L100 | L99 | A98 | L97 | **D96** | L95 | L94 | L93 | L92 |
| | **52.0** | 52.2 | 52.5 | 52.7 | 53.0 | **53.2** | 53.4 | 53.6 | 53.9 | 54.1 |
| | **P91** | T90 | T89 | F88 | L87 | W86 | D85 | A84 | **Y83** | R82 |
| | **54.3** | 54.5 | 54.8 | 55.0 | 55.2 | 55.4 | 55.6 | 55.9 | **56.1** | 56.2 |
| | A81 | W80 | | | | | | | | |
| | 56.4 | 56.6 | | | | | | | | |
| Linker btw C and B | Y79 | I78 | **P77** | N76 | Q75 | E74 | G73 | G72 | **F71** | P70 |
| | 56.7 | 56.8 | **57.0** | 57.3 | 57.7 | 58.0 | 58.3 | 58.7 | **59.0** | 59.5 |
| | V69 | M68 | T67 | L66 | G65 | Y64 | | | | |
| | 60.0 | 60.4 | 60.9 | 61.4 | 61.8 | 62.3 | | | | |
| Helix B | **G63** | L62 | L61 | M60 | S59 | L58 | Y57 | M56 | T55 | **F54** |
| | **62.8** | 63.4 | 64.0 | 64.6 | 65.2 | 65.9 | 66.5 | 67.1 | 67.7 | **68.3** |
| | A53 | I52 | A51 | P50 | V49 | L48 | T47 | T46 | I45 | A44 |
| | 68.8 | 69.2 | 69.7 | 70.2 | 70.7 | 71.1 | 71.6 | 72.1 | 72.5 | 73.0 |
| | Y43 | F42 | K41 | K40 | A39 | D38 | P37 | | | |
| | 73.5 | 74.0 | 74.4 | 74.9 | 75.4 | 75.9 | 76.3 | | | |
| Linker | D36 | S35 | V34 | G33 | M32 | G31 | K30 | | | |
| | 76.8 | 77.3 | 77.7 | 78.2 | 78.7 | 79.2 | 79.6 | | | |
| Helix A | **V29** | L28 | F27 | Y26 | L25 | T24 | G23 | L22 | G21 | M20 |
| | **80.1** | 80.3 | 80.6 | 80.8 | 81.1 | 81.3 | 81.5 | 81.8 | 82.0 | 82.2 |
| | L19 | A18 | T17 | **G16** | L15 | A14 | L13 | W12 | I11 | W10 |
| | 82.5 | 82.7 | 83.0 | **83.2** | 83.4 | 83.5 | 83.7 | 83.8 | 84.0 | 84.2 |
| | E9 | **P8** | R7 | G6 | T5 | I4 | Q3 | **A2** | E1 | |
| | 84.3 | **84.5** | 84.9 | 85.3 | 85.7 | Not in 1qhj.pdb | | | | |

**Table S7. Parameters describing the potential barriers that stabilize secondary structural elements of bR against force.** The values in comparison are the width of potential ($\mathbf{\Delta x^{\neq}}$ [Å]) from a 2004 experimental study (7). The uncertainty of the fitted width of potential is computed from the s.e.m. of the unfolding forces. In specific, mean force ± s.e.m of forces is used in equation $\boldsymbol{F_{mp} = (\beta \cdot \Delta x^{\neq})^{-1} ln \frac{\beta \cdot \Delta x^{\neq} \cdot \kappa \cdot v}{k_0}}$ where $\beta = (k_B\mathrm{T})^{-1}$ to obtain the upper and lower boundaries, $\mathbf{\Delta x}^{\neq}_{\boldsymbol{upper}}$ and $\mathbf{\Delta x}^{\neq}_{\boldsymbol{lower}}$, of the width of potential. The uncertainty is $\boldsymbol{[(\Delta x^{\neq}_{upper} - \Delta x^{\neq}) + (\Delta x^{\neq} - \Delta x^{\neq}_{lower})]/2}$.

| | Monomer simulations | 2004 experiment |
|---|---|---|
| Helix pair ED | 5.3±0.5 Å | 3.2 Å |
| Helix pair CB | 9.9±0.1 Å | 8.6 Å |
| Helix A | 6.1±0.1 Å | 6.8 Å |

**Movie S1. Unfolding trajectories of bR and GlpG.** Simulations are run with a spring constant $\kappa = 0.05\ k_B\mathrm{T}/\text{Å}^2$ and pulling velocity = 0.001 Å/*Upside* time step at $\mathrm{T} \approx 300$ K.

**A.** Force-induced unfolding of bR in the AFM (stiff spring) mode. Force is applied vertically to the membrane surface.

**B.** Force-induced unfolding of GlpG in the AFM (stiff spring) mode, unfolding from N→C, from C→N, and from the middle, respectively. Force is applied laterally to the membrane surface.

## REFERENCES

1. Baker, R. P., and S. Urban. 2012. Architectural and thermodynamic principles underlying intramembrane protease function. Nat Chem Biol 8:759-768.
2. Levy, R., and M. Maaloum. 2002. Measuring the spring constant of atomic force microscope cantilevers: thermal fluctuations and other methods. Nanotechnology 13:33-37.
3. Bustamante, C., J. F. Marko, E. D. Siggia, and S. Smith. 1994. Entropic elasticity of lambda-phage DNA. Science 265:1599-1600.
4. Strick, T., J. F. Allemand, V. Croquette, and D. Bensimon. 2000. Twisting and stretching single DNA molecules. Prog Biophys Mol Bio 74:115-140.
5. Min, D., R. E. Jefferson, J. U. Bowie, and T. Y. Yoon. 2015. Mapping the energy landscape for second-stage folding of a single membrane protein. Nat Chem Biol 11:981-987.
6. Evans, R. A. 1966. Abramowitz M - Handbook of Mathematical Functions with Formulas Graphs and Mathematical Tables Nbs Applied Mathematics Series 55. Ieee Spectrum 3:161-&.
7. Janovjak, H., J. Struckmeier, M. Hubain, A. Kedrov, M. Kessler, and D. J. Muller. 2004. Probing the energy landscape of the membrane protein bacteriorhodopsin. Structure 12:871-879.
8. Woodside, M. T., and S. M. Block. 2014. Reconstructing folding energy landscapes by single-molecule force spectroscopy. Annu Rev Biophys 43:19-39.
9. McGibbon, R. T., K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernandez, C. R. Schwantes, L. P. Wang, T. J. Lane, and V. S. Pande. 2015. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophys J 109:1528-1532.

10. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577-2637.
11. Savitzky, A., and M. J. E. Golay. 1964. Smoothing + Differentiation of Data by Simplified Least Squares Procedures. Anal Chem 36:1627-&.
12. Oliphant, T. E. 2007. Python for scientific computing. Comput Sci Eng 9:10-20.
13. Yu, H., M. G. Siewny, D. T. Edwards, A. W. Sanders, and T. T. Perkins. 2017. Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. Science 355:945-950.
14. Muller, D. J., M. Kessler, F. Oesterhelt, C. Moller, D. Oesterhelt, and H. Gaub. 2002. Stability of bacteriorhodopsin alpha-helices and loops analyzed by single-molecule force spectroscopy. Biophys J 83:3578-3588.
15. Ainavarapu, S. R., J. Brujic, H. H. Huang, A. P. Wiita, H. Lu, L. Li, K. A. Walther, M. Carrion-Vazquez, H. Li, and J. M. Fernandez. 2007. Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. Biophys J 92:225-233.
16. Stirnemann, G., D. Giganti, J. M. Fernandez, and B. J. Berne. 2013. Elasticity, structure, and relaxation of extended proteins under force. Proc Natl Acad Sci U S A 110:3847-3852.
17. Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12:2825--2830.
18. Jumper, J. M., N. F. Faruk, K. F. Freed, and T. R. Sosnick. 2018. Trajectory-based training enables protein simulations with accurate folding and Boltzmann ensembles in cpu-hours. PLoS Comput Biol 14:e1006578.
19. Jumper, J. M., N. F. Faruk, K. F. Freed, and T. R. Sosnick. 2018. Accurate calculation of side chain packing and free energy with applications to protein molecular dynamics. PLoS Comput Biol 14:e1006342.
20. Wang, Z., J. M. Jumper, S. Wang, K. F. Freed, and T. R. Sosnick. 2018. A Membrane Burial Potential with H-Bonds and Applications to Curved Membranes and Fast Simulations. Biophys J.
21. Yamada, T., T. Yamato, and S. Mitaku. 2016. Forced Unfolding Mechanism of Bacteriorhodopsin as Revealed by Coarse-Grained Molecular Dynamics. Biophys J 111:2086-2098.
22. Subramaniam, S., and R. Henderson. 2000. Molecular mechanism of vectorial proton translocation by bacteriorhodopsin. Nature 406:653-657.
23. Mackerell, A. D., Jr., M. Feig, and C. L. Brooks, 3rd. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25:1400-1415.
24. Kessler, M., and H. E. Gaub. 2006. Unfolding barriers in bacteriorhodopsin probed from the cytoplasmic and the extracellular side by AFM. Structure 14:521-527.
25. Sapra, K. T., J. Doehner, V. Renugopalakrishnan, E. Padros, and D. J. Muller. 2008. Role of extracellular glutamic acids in the stability and energy landscape of bacteriorhodopsin. Biophys J 95:3407-3418.
26. Voitchovsky, K., S. A. Contera, and J. F. Ryan. 2007. Electrostatic and steric interactions determine bacteriorhodopsin single-molecule biomechanics. Biophys J 93:2024-2037.