Forced unfolding mechanism of bacteriorhodopsin as revealed by coarse-grained molecular dynamics

Tatsuya Yamada,[1] Takahisa Yamato,[1,*] and Shigeki Mitaku[2]

[1]Graduate School of Science, Nagoya University, Nagoya, Aichi, Japan

[2]Toyota Physical & Chemical Research Institute, Nagakute, Aichi, Japan

[*]Correspondence: yamato@nagoya-u.jp

Running head: Forced Unfolding of Bacteriorhodopsin

ABSTRACT

Developments in atomic force microscopy (AFM) have opened up a new path toward single-molecular phenomena, in particular, during the process of pulling a membrane protein out of a lipid bilayer. However, the characteristic features of the force–distance (F–D) curve of a bacteriorhodopsin (bR) in purple membrane, for instance, have not yet been fully elucidated in terms of physicochemical principles.

To address the issue, we performed a computer simulation of bR with a novel coarse-grained (C-G) model. Peptide planes are represented as rigid spheres, while the surrounding environment consisting of water solvents and lipid bilayers is represented as an implicit continuum. Force-field parameters were determined on the basis of auxiliary simulations and experimental values of transfer free energy of each amino acid from water to membrane. According to Popot's two-stage model, we separated molecular interactions involving membrane proteins into two parts: (I) affinity of each amino acid to the membrane, and intra-helical hydrogen bonding between main chain peptide bonds; (II) inter-helix interactions. Then, only part (I) was incorporated into the C-G model because we assumed that the part plays a dominant role in the forced unfolding process. As a result, the C-G simulation has successfully reproduced the key features, including peak positions, of the experimental F–D curves in the literature, indicating that the peak positions are essentially determined by the residue-lipid and intra-helix interactions. Furthermore, we investigated the relationships between the energy barrier formation on the forced unfolding pathways and the force peaks of the F–D curves.

INTRODUCTION

Membrane proteins constitute essential parts of living organisms for signal transduction, molecular transport of cells, energy conversion (1) and others. A significant fraction (20–30%) of sequenced genes is accounted for by membrane proteins (2, 3). Developments in atomic force microscopy (AFM) have opened a new path toward single-molecular phenomena, in particular, during the process of pulling a membrane protein out of a lipid bilayer (4–24). Among them, the most studied and analyzed protein is bacteriorhodopsin (bR) (25, 26), which is composed of seven transmembrane (TM) helices. When the stylus tip of the cantilever of the AFM apparatus is attached to either the N- or C-terminal loop of bR and continuously extracted from the purple membrane (PM), the force (F) acting on the stylus tip is recorded as a function of the distance (D) from the membrane surface. Thereby, we are allowed access to direct information concerning the mechanical properties of a single bR chain by analyzing characteristic sawtooth-like features of the F–D curves (Fig. 1).

To elucidate the forced unfolding mechanism of membrane proteins, computational studies have been conducted (27–30). However, there is ongoing debate regarding the origins of the sawtooth-like features. For instance, Kappel and Grubmüller conducted all-atom simulations of the forced unfolding of bR and highlighted the importance of specific anchor points (29). Overall, their F–D curves exhibited similar pattern to those obtained by AFM. They proposed, however, velocity-dependent unfolding mechanism based on the kinetics of intramolecular interaction network. Oesterhelt et al. selectively cleaved the E–F loop and extracted the truncated bR chain from the C-terminal end of helix E. The truncated chain exhibited similar F–D curves to those of the wild type except for the lack of the single peak that corresponds to helices F–G (4). Likewise, Kedrov et al. performed a similar experiment of sodium proton antiporter A (NhaA), another transmembrane α-helical protein (8). These studies indicate that the force peak originates from intrinsic properties of each TM helix rather than the inter-helix interactions. Furthermore, experimental studies (13, 16, 23, 24) suggested that the magnitude and the appearance frequency of the force peaks of the F–D curves were affected by the inter-helix interactions, whereas the peak positions were not. Kessler et al. performed unfolding and refolding experiments of bR by lifting the AFM tip up and down repeatedly, and observed that the peak positions of the F–D curves for the helix pairs ED and CB were always reproduced whereas the magnitudes of these peaks were

sometimes reproduced and sometimes were not (31). According to these experimental observations, they suggested that the refolding of bR occurs based on the two-stage model (32–35) proposed by Popot and Engelman, and the partial recovery of a force peak was due to imperfect formation of native contacts between TM helices.

To address the issue, we developed a computer simulation system for the forced unfolding of bR, on the basis of the two-stage model. According to the two-stage model, we can expect that a helical membrane protein folds in a stepwise manner via the two energetically distinct stages (35). In the first stage, a polypeptide chain forms individually stable TM helices in the membrane, and these TM helices assemble into the complete tertiary structure in the second stage. From energetic point of view, they assumed that the membrane affinities of amino acid residues and main-chain hydrogen bonding in each TM helix, hereafter denoted as class-I interactions, play central roles in the first stage. In the second stage, they assumed that the inter-helix interactions, hereafter denoted as class-II interactions, are responsible for the complete tertiary structure formation. The two-stage model has been supported by experimental observations (18), (31).

For the past two decades or so, computational studies (36–39) have been devoted to the TM helix formation mechanism. We developed a coarse-grained (C-G) model because the unfolding occurs on the time scales much slower than those covered by typical all-atom simulations, making unfolding simulations of membrane proteins unfeasible with the finite computational resources available. The surrounding water and membrane environment were represented as implicit medium, and the Brownian dynamics (BD) method (40–42) was employed for the simulation. Before calculations, we verified that our model system is overdamped and suitable for the BD simulation. The class-I interactions were incorporated in the force-field functions, while the class-II interactions were not. Force-field parameters were determined on the basis of the auxiliary simulations and experimental values of transfer free energy of each amino acid from water to membrane. In this study, we are particularly interested in the force peak positions in the F–D curves and focused on the role of the class-I interactions in forced unfolding of bR. The F–D curves were obtained by using the C-G BD simulation. As a result, we found that the simulated peak positions agree well with those reported in the experimental studies in the literatures. Furthermore, we discussed possible unfolding mechanisms of bR on the basis of the C-G BD calculations.

MATERIALS AND METHODS

C-G model

Here, the polypeptide chain was modeled as a string of beads, each of which represents a $C^{\alpha C}$–CONH–$C^{\alpha N}$ segment hereafter referred to as "peptide bond particle (PBP)" or simply "particle", and a local coordinate system (LCS) was assigned for each PBP. The octet, {$C^{\alpha C}$, C, O, N, H, $C^{\alpha N}$ hbo, hbh} (Figs. 2 and S1(A) in the Supporting Material, Table S1), was introduced for each PBP, where $C^{\alpha C}$ ($C^{\alpha N}$) represents the C^{α} atom in the previous (next) residue attached to C (N), –CO–NH– forms a peptide bond, and hbo (hbh) represents a virtual "bond point" near O (H) on which the H-bonding counterpart is attached. The sequential number of the N- (C-) terminal PBP was set to 0 (N), and $C^{\alpha C}$, C, O, (N, H, $C^{\alpha N}$) in the i-th PBP belong to the i-th (i + 1-st) residue. Thus, a polypeptide is represented as a chain of N + 1 PBPs. For the i-th PBP, the 1×3 column vectors, $\hat{\mathbf{r}}^{\alpha C}, \hat{\mathbf{r}}^{C}, \hat{\mathbf{r}}^{O}, \hat{\mathbf{r}}^{N}, \hat{\mathbf{r}}^{H}, \hat{\mathbf{r}}^{\alpha N}, \hat{\mathbf{r}}^{hbo}, \hat{\mathbf{r}}^{hbh}$, represent the octet arrangement in LCS. Note that henceforth, we use the hat (non-hat) notation of the vectors in LCS (GCS, or global coordinate system) unless otherwise noted. Coordinate transformation from LCS to GCS is done by $\mathbf{r}^{P} = \mathbf{A}_i \hat{\mathbf{r}}^{P} + \mathbf{r}_i$ for the point P, where $\mathbf{r}_i$ (3 × 3 rotation matrix $\mathbf{A}_i$) represent the origin position (orientation) of the i-th PBP. The state of the system was represented as: $\mathbf{q} = \{\mathbf{r}_0, \mathbf{r}_1,\ldots, \mathbf{r}_N, \mathbf{A}_0, \mathbf{A}_1,\ldots, \mathbf{A}_N, \mathbf{r}^{cl}_{base}\}$, where $\mathbf{r}^{cl}_{base}$ indicates the cantilever base position.

Force-field functions of the C-G model

The total potential energy is expressed as $U^{sys} = U^{pep} + U^{mem} + U^{hb} + U^{cl} + U^{wall}$, where U^{pep} is the energy term responsible for holding the PBPs together as a polypeptide chain. U^{mem} and U^{hb} represent the affinity of amino acid residues to the membrane and mainchain H-bonding interactions, while U^{cl} and U^{wall} correspond to the deformation energy of the cantilever and the repulsive particle-wall interactions, respectively.

The first term, U^{pep}, is separated further into $U^{pep} = U^{b} + U^{a} + U^{da} + U^{rep}$, where U^{b} and U^{rep} represent the attractive and repulsive inter-particle interactions, respectively. The i- and (i − 1)-th PBPs are bound together via the harmonic restraint potential, U^{b}_i, working between $C^{\alpha C}$ of the former and $C^{\alpha N}$ of the latter, and the C^{α} atom of the i-th residue is placed at the midpoint between $C^{\alpha C}$ and $C^{\alpha N}$. The constraint energy for bond angles is represented by U^{a}, and U^{da} provides special constraints for proline dihedrals. Each term is expressed as follows

$$\begin{pmatrix} U^{\mathrm{b}} \\ U^{\mathrm{a}} \\ U^{\mathrm{da}} \end{pmatrix} = \sum_{i=1}^{N} \begin{pmatrix} U_i^{\mathrm{b}} \\ U_i^{\mathrm{a}} \\ U_i^{\mathrm{da}} \end{pmatrix}, \quad \begin{pmatrix} U_i^{\mathrm{b}} \\ U_i^{\mathrm{a}} \end{pmatrix} = \begin{pmatrix} \dfrac{k^{\mathrm{b}}}{2}(\mathbf{r}_i^{\alpha\mathrm{C}} - \mathbf{r}_i^{\alpha\mathrm{N}})^2 \\ \dfrac{k^{\mathrm{a}}}{2}(\theta_i - 110.1^\circ)^2 \end{pmatrix},$$

$$U_i^{\mathrm{da}} = \begin{cases} \dfrac{k^{\mathrm{da}}}{2}(\phi_i + 65.4^\circ)^2, & \text{if } i\text{-th residue is proline,} \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

$$U^{\mathrm{rep}} = \sum_{i=0}^{N-3}\left(\sum_{j=i+3}^{N} U_{i,j}^{\mathrm{rep}} \right),$$

$$U_{i,j}^{\mathrm{rep}} = \begin{cases} \dfrac{k^{\mathrm{rep}}}{2}(2r^{\mathrm{rep}} - |\mathbf{r}_j - \mathbf{r}_i|)^2, & \text{for } |\mathbf{r}_j - \mathbf{r}_i| < 2r^{\mathrm{rep}} \\ 0, & \text{ot herwise} \end{cases},$$

where θ_i is the bond angle $\angle\mathrm{NC}^\alpha\mathrm{C}$ in the i-th residue and ϕ_i is the main chain ϕ-angle of

the i-th residue. The values of parameters k^{b}, k^{a}, k^{da} k^{rep}, and r^{rep} were set to 125.02 [J/m^2], 4.618×10^{-22} [J/deg^2], 3.281×10^{-23} [J/deg^2], 100 [N/m], and 0.2366 [nm], respectively. To determine the value of k^{a}, k^{b}, we performed a preliminary calculations. We confirmed that the magnitude of $\left| \mathbf{r}_i^{\alpha\mathrm{C}} - \mathbf{r}_i^{\alpha\mathrm{N}} \right|$ was always ~0.1Å during the simulations.

The next term, U^{mem}, represents the affinity of amino-acid residues to the membrane that is regarded as continuum (Fig. 3) in which a coordinate frame is introduced such that the xy-plane became the middle plane of the membrane system and the z-axis was set perpendicular to it. The contributions from the PBPs to U^{mem} are divided into main chain and side chain contributions, expressed as

$$U^{\mathrm{mem}} = \sum_{i=0}^{N}(\Delta U_i^{\mathrm{mc}} \cdot g(z_i)) + \sum_{i=1}^{N}(\Delta U_i^{\mathrm{sc}} \cdot g(z_i^{\alpha})),$$

$$g(z) = \begin{cases} 0, & \text{for } s+d < |z| \\ \dfrac{1}{2} - \tilde{z} - \dfrac{1}{2\pi}\sin(2\pi\tilde{z}), & \text{for } s \le |z| \le s+d, \\ 1, & \text{for } |z| < s \end{cases} \tag{2}$$

$$\tilde{z} = (1/d)(|z| - s - (d/2)),$$

where z_i (z_i^{α}) represents the z-coordinate of the origin (C$^\alpha$ atom) of the i-th PBP (residue). s (half-thickness of the lipid core region) and d (thickness of the interface region) were set to 1.4 and 0.7 nm respectively. ΔU_i^{mc} and ΔU_i^{sc} were derived from the octanol–water partition coefficients (43, 44) and partially modified in this study (Table

1). The smoothing function, $g(z)$, decreases from 1 (membrane region) to 0 (water region) (Fig. 3). Note that $\Delta U^{mc}_0 = \Delta U^{N\text{-term}}$ (= 9.8 kJ/mol) and $\Delta U^{mc}_N = \Delta U^{C\text{-term}}$ (= 27.1 kJ/mol) in the standard polypeptide chain.

The next term, U^{hb}, provides mainchain H-bonding interactions responsible for TM helix formation.

$$U^{hb} = \sum_{i=1}^{N-4} U_i^{hb}, \ U_i^{hb} = \begin{cases} 0 \ \ \text{if } (i+4)\text{th residue is proline} \\ \{(\Delta U^{hb,m} - \Delta U^{hb,w}) \cdot g(z_i^{hb}) + \Delta U^{hb,w}\} \cdot h(r_i^{hb}), \ \text{otherwise} \end{cases}$$

$$h(r) = \begin{cases} -3(r/r_0^{hb})^4 + 8(r/r_0^{hb})^3 - 6(r/r_0^{hb})^2 + 1, \ \text{for } r < r_0^{hb} \\ 0, \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \text{otherwise} \end{cases}$$

(3)

where r_0^{hb} was set 0.25nm, z_i^{hb} and r_i^{hb} represent the z-component of $(\mathbf{r}^{hbo}_i + \mathbf{r}^{hbh}_{i+4})/2$ and $|\mathbf{r}^{hbo}_i - \mathbf{r}^{hbh}_{i+4}|$, respectively. The same function $g(z)$ was used as that in U^{mem}. If $\mathbf{r}^{hbo}_i = \mathbf{r}^{hbh}_{i+4} = (\mathbf{r}_i + \mathbf{r}_{i+3})/2$, the i-th residue takes the ideal right-handed α-helical structure in which dihedral angles $(\phi_i, \psi_i) = (-60°, -45°)$ (45) (Fig. S1(B)). $\Delta U^{hb,w}$(= −8.726 kJ/mol) and $\Delta U^{hb,m}$(= −24.93 kJ/mol) represent the depth of the H-bonding interaction potential in water and lipid core, respectively. The former was derived from the free energy change associated with the N-methylacetamide dimer formation in water, i.e., +13.0 kJ/mol (46). For this purpose, we constructed a simple model system of a pair of isolated PBPs in water and estimated the strength of the pairwise attraction so that the value of $\Delta F° = -RT \ln \dfrac{[\text{C=O} \cdots \text{H}-\text{N}]}{[\text{C=O}_{free}][\text{H}-\text{N}_{free}]}$ became the experimental value (46). Note that the N-methylacetamide molecule has a similar structure to the polypeptide mainchain unit. The value of $\Delta U^{hb,m}$ was determined by the preliminary simulations for 100 μs with no extraction force ($U^{cl} = 0$) at different values of $\Delta U^{hb,m}$ at −9.0, −9.5, −10.0, −10.5, −11.0$k_B T$, where k_B is Boltzmann's constant and $T = 300$K. Then, we found that each of the seven TM helices forms a stable helix structure when $\Delta U^{hb,m}$ was set to $-10.0 k_B T = -24.93$ kJ/mol, and this value was adopted in the present study (Fig. S2). The stability of each helix was examined using the STRIDE program (47).

The next term, U^{cl}, represents the potential energy of the cantilever whose base and head positions are given by $\mathbf{r}^{cl}_{base}$ and $\mathbf{r}^{cl}_{head}$, respectively (Fig. 3). The expression of the potential is given by

$$U^{cl} = \frac{1}{2}\{k_x^{cl}(x_{b\text{-}h}^{cl})^2 + k_y^{cl}(y_{b\text{-}h}^{cl})^2 + k_z^{cl}(z_{b\text{-}h}^{cl})^2\},$$

$$(x_{b\text{-}h}^{cl}, y_{b\text{-}h}^{cl}, z_{b\text{-}h}^{cl})^{\text{T}} = \mathbf{r}^{cl}_{head} - \mathbf{r}^{cl}_{base},$$

(4)

6

where the value of k^{cl}_z is set as a typical cantilever stiffness used in experiments ([4]), i.e., $k^{\text{cl}}_z = 0.1$ [N/m], while $k^{\text{cl}}_x = k^{\text{cl}}_y$ were set to 50 [N/m]. In the extraction simulation of bR from the N- (C-) terminal side, we set $\mathbf{r}^{\text{cl}}_{\text{head}} = \mathbf{r}^{\text{cl}}_{\text{base}} = \mathbf{r}_0$ ($\mathbf{r}_N$) at the initial state, and then the condition, $\mathbf{r}^{\text{cl}}_{\text{head}} = \mathbf{r}_0$ ($\mathbf{r}_N$), was maintained till the end of the simulation. Note that the force acting on the cantilever along the z-direction is given by $F^{\text{cl}}_z = k^{\text{cl}}_z z^{\text{cl}}_{\text{b–h}}$.

U^{wall} is the particle-wall potential energy due to which the PBPs are expelled from the gray region in Fig. 3.

$$U^{\text{wall}} = \sum_{i=0}^{N} U_i^{\text{wall}}, \ U_i^{\text{wall}} = \begin{cases} \dfrac{k^{\text{wall}}}{2}(r^{\text{wall}} - |\mathbf{r}_i - \mathbf{r}_i^{\text{d}}|)^2, & \text{for } |\mathbf{r}_i - \mathbf{r}_i^{\text{d}}| < r^{\text{wall}} \\ 0, & \text{ot herwise,} \end{cases} \tag{5}$$

where $k^{\text{wall}} = 100.0$ [N/m] and $r^{\text{wall}} = 0.2366$ [nm]. The nearest point to $\mathbf{r}_i$ in the gray region is represented by $\mathbf{r}_i^{\text{d}}$.

BD simulation

It is known that the BD method ([40–42, 48]) is suitable for the protein simulation with the implicit solvent model for the surrounding environment. The following is the calculation procedure:

(i) For $i = 0, 1, \ldots, N$, we calculate the global coordinates of the octet of the i-th PBP using $\mathbf{r}_i(t)$ and $\mathbf{A}_i(t)$ and $\mathbf{r}^{\text{cl}}_{\text{head}}(t)$ at time t.

(ii) Based on the coordinates obtained in the previous step, we calculate the external force and torque, $\mathbf{F}_i(t)$ and $\mathbf{T}_i(t)$, respectively, acting on the i-th PBP at time t.

(iii) Based on the Brownian dynamics method, we calculate $\mathbf{r}_i(t + \Delta t)$ and $\mathbf{A}_i(t + \Delta t)$ from $\mathbf{r}_i(t)$ and $\mathbf{A}_i(t)$, respectively.

(iv) The distance between the cantilever head and the membrane surface, and the force acting on the polypeptide chain by the cantilever are recorded for drawing F–D curves.

(v) The cantilever base position is shifted toward a positive direction along the z-axis by the distance obtained by $v^{\text{cl}} \times \Delta t$, where v^{cl} is the extraction velocity (=1.0 mm/s in this study). In the preliminary calculations, we examined the v^{cl}-dependence of the viscous resistance acting on the polypeptide chain. If the magnitude of the viscous resistance is too large then the calculation results of the F–D curve are significantly disturbed. We found that if we set $v^{\text{cl}} = 1.0$ mm/s, the magnitude of the viscous resistance was suppressed at most less than 3.0 pN, ensuring that the calculations are reliable.

In this study, BD simulations were performed at $T = 298$ K with the Stokes radius of

each PBP was set to 0.4 nm, using the time step $\Delta t = 10$ fs. The viscosity coefficient was set to 8.9×10^{-4} Pa·s for both water and membrane phases. The validity of the simulation condition will be discussed later. In an extraction simulation, we repeated the abovementioned cycle for 10^{10} steps (0.1 ms), and we saved the molecular structure every 10^6 steps together with the time-averaged extraction force, which was averaged over the period of the 10^6 steps. Two types of the simulations of bR extraction from the N- and C-terminus, henceforth denoted as simulations **N** and **C**, respectively, were performed 128 times for each type with different random number sequences (Movies S1 and S2). The computational time for a single simulation run took approximately 202 h using one core of 2.8GHz quad core Intel Xeon E5462 processors.

Initial model

The x-ray crystallographic coordinates (PDB id; 1BRR, chain A) (49) of bR were used to construct the starting structure. Note that the x-ray model is different from the wild type in that G-241 is deleted, and its N-terminal glutamic acid is pyroglutamylated, leading to a chargeless N-terminus. Therefore, we deleted G-241, E-1 was replaced by glutamate in our model, put $\Delta U^{mc}_0 = 0$, and the total amino acid residue length became 247. Then, we placed the bR model so that the helical axes of the TM helices were aligned parallel to the z-axis, with the N- or C-terminus on the upside of the membrane (see Fig. 3) for simulation **N** or **C**, respectively. The protonated Schiff base of retinal (PSBR) was not included in the system because the removal of PSBR does not affect the force-peak positions (16).

Stretched polypeptide chain

The polypeptide segment between the cantilever head to the amino acid residue lying on the top surface of the membrane is called a stretched polypeptide chain (SPC) (16). Here, we assume that the lower end of the chain terminates in the hypothetical surface residue i_s whose C^α atom is exactly on the surface, i.e., $z = d + s$, and the value of i_s can be either integer or non-integer. During simulation **C** (**N**), the length of SPC, $L_{SPC} = 247 - i_s$ ($L_{SPC} = i_s - 1$), and

$$i_s = j + \frac{(d+s) - z^\alpha_j}{z^\alpha_{j+1} - z^\alpha_j},\tag{6}$$

where j is the largest (smallest) integer that satisfies $z^\alpha_j \leq (d+s) < z^\alpha_{j+1}$ ($z^\alpha_{j+1} \leq (d+s) < z^\alpha_j$)

and d and s are 0.7 and 1.4 nm, respectively. Throughout this paper the force peak positions of the simulated F–D curves are defined in terms of L_{SPC} unless otherwise specified.

RESULTS AND DISCUSSION

F–D curve analysis

A superimposed plot of F–D curves of simulation C (N) is shown in Fig. 4A (B), in which we identified 918 (1,420) force peaks by visual inspection. We performed the force peak analysis (Fig. S3): (i) the force magnitude, (ii) peak distance, i.e., the distance between the cantilever head and the membrane surface, and (iii) the SPC length, L_{SPC}, when the cantilever experiences the peak force. As a result of the analysis for the two types of simulations C and N, 13 and 18 distinct groups of peak positions were identified, respectively (Figs. 4C, D). These characteristic features of the F–D curve should be intimately related to the energy landscape of bR along the forced unfolding pathways.

Frequency distribution of the SPC length

During the simulations we recorded the value of L_{SPC} as a function of step number. For instance, if a surface residue experiences a large energy barrier on the membrane surface during extraction and stays there for a long time, the frequency probability of finding L_{SPC} in the trajectories becomes large in proportion to the residence time on the surface. In such a case, the magnitude of the force peak should be large. During the 128 simulation runs for each simulation C (N), the frequency distribution of L_{SPC} was quantitatively evaluated as follows. First, we calculated the occurrence number of L_{SPC}, $N(k)$, for integer k ($0 \leq k \leq 246$), where $L_{SPC} - 0.5 < k \leq L_{SPC} + 0.5$. Then, the appearance frequency, $F(k)$, was defined as $N(k)/N_{max}$, where N_{max} is the maximum number of the $N(k)$ for $0 \leq k \leq 246$. In parallel, we calculated the frequency peak positions $k_p = \sum_{i=k'_p-2}^{k'_p+2} F(i) \cdot i \left/ \sum_{i=k'_p-2}^{k'_p+2} F(i) \right.$, where $F(k'_p-2), F(k'_p-1), F(k'_p+1)$ and $F(k'_p+2)$ are smaller than $F(k'_p)(>0.0003)$. Note that the force peaks were identified by visual inspection in the previous section. However, Fig. 5 shows a strong correlation between the force peaks and the appearance frequencies of L_{SPC}, indicating the validity of the identification of the peak positions in this study.

Comparison with the experimental F–D curves

The F–D curves of simulation C (N) were compared with those obtained experimentally in the literature (20, 21, 24) (Fig. 5). Note that the definitions of the force-peak positions are different among those studies. To compare different studies i.e., L_{SPC} (our study), L_S(24), L_K(20), and L_V(21), we introduced a concept of the standard force-peak positions, L_{std}. For the forced unfolding of bR from (1) the N- and (2) the C-terminal side, the conversion rules are as follows: (1) $L_{std} = L_{SPC}$, $L_{std} = L_K - d$, $L_{std} = L_V + l$, ($l = 2$), and (2) $L_{std} = L_{SPC} + 1$, $L_{std} = L_S$, $L_{std} = L_K - d$, $L_{std} = L_V + l$, ($l = 19$), where $d = (2.1 \text{ nm} - z_{anchor})/(0.36 \text{ nm})$. The values of z_{anchor} were derived from Figure 3 of the literature (20) and 0.36 nm represents the effective length of one residue in a fully stretched polypeptide chain. The values of l for (1) and (2) were determined based on the literature (21). Here, we compared our computational force-peak positions with the experimental measurements with ±3 amino-acid-residue long being the allowance limit. As a result, the concordance rate defined as (the match number, n_{sim})/(the number of experimental peak positions) for simulation C (N) was 8/11(24), 8/13(20), and 7/17(21) (7/13(20) and 6/14(21)). Importantly, the locations of frequently appeared force peaks, indicated by large closed circles in Fig. 5, were almost perfectly matched with the experimental observations.

To appreciate whether the degree of matches between the computational and the experimental (20, 21, 24) peak positions are meaningfully more frequent than accidental coincidence, we further examined and verified our computational results. For a pair of each experiment and simulation C (N), we created a histogram of random matching probability as follows: (1) Choose 13 (18) random positions in the interval between [0, 247]. (2) Count the number of matches, n_{rand}, between these random positions and the experimental peak positions. (3) Processes (1) and (2) were repeated over 10^{10} times, and we finally calculated the average matching probability, p_{match}, as a function of n_{rand}. For simulation C (N), the summation of the matching probability, $\sum_{n_{sim} \leq n_{rand}} p_{match}(n_{rand})$, was

0.00043(24), 0.0043(20), 0.087(21) (0.15(20), 0.37(21)), indicating that the matches of the peak positions between the experiments and simulation C were not regarded as accidental coincidence, while the concordance rates were less satisfactory for simulation N. This indicates that our simulation condition may be closer to the experimental

conditions of Sapra and Kessler than to that of Voïtchovsky. Note that the peak positions were not recorded in the interval [0, 60] in the extraction from the N-terminus by Kessler et al.[20], and we could not obtain any match in this portion.

   With regard to the magnitudes of the force peaks of the F–D curves, we observed that the peak magnitudes fell in the range of 50–120 pN. In the forced unfolding of bR, thermodynamic forces acting on each component of the polypeptide chain are particularly important in the lipid–water interface regions. The force components toward the $+z$ direction acting on typical hydrophobic, e.g., leucine; hydrophilic, e.g., asparagine; and ionizable, e.g., aspartic acid, side chain fragments at the interface region at $z = 1.75$ nm (or -1.75 nm) were -55 (+55), $+46$ (-46), and $+113$ (-113) pN, respectively. An isolated main chain fragment feels strong hydrophilic driving force (59 pN) at the interface toward the outside of the membrane, whereas those of main chain fragment participating in TM helices are significantly decreased to 11 pN because bare –CO and –NH groups are no longer available as a result of hydrogen-bond formation.

Similarity rule of the BD simulation

   In the BD simulation, the translational (orientational) degrees of freedom are updated each step and the changes in these values are proportional to $D^T\Delta t$ ($D^R\Delta t$) for the translational (rotational) motions, where D^T (D^R) represents the translational (rotational) diffusion coefficient. In the forced unfolding simulation, the state of the system also depends on the z-coordinate of the cantilever stylus, and the change in its coordinate is equal to $v^{cl}\times\Delta t$ at every step. Here, we can deduce a useful property, i.e., a similarity rule, of the BD simulation. A pair of simulations beginning from the same initial conditions with a different set of parameters, (η_1, Δt_1, v^{cl}_1) and (η_2, Δt_2, v^{cl}_2), are equivalent under the condition that $\Delta t_1/\eta_1 = \Delta t_2/\eta_2$ and $v^{cl}_1\times\Delta t_1 = v^{cl}_2\times\Delta t_2$ because the translational and rotational diffusion coefficients are inversely proportional to the viscosity coefficient (see **Supporting material S7**).

Validity of the simulation conditions

   As described before, we performed simulations with the following set of parameters: time step of the simulation (Δt_1) = 10 fs, viscosity coefficient (η_1) = 8.9 $\times 10^{-4}$ Pa·s, which is equivalent to the bulk water viscosity, and extraction speed (v^{cl}_1) = 1.0 mm/s. Using the scaling relation as mentioned previously, it is possible to scale this set of

parameters to $(\Delta t_2, \eta_2, v^{cl}_2) = (k\Delta t_1, k\eta_1, v^{cl}_1/k)$, where k is an arbitrary constant. In the following paragraphs, we will see that the parameters Δt_2, η_2, and v^{cl}_2 satisfy both experimental conditions and the requirement for the BD simulation if we choose the scaling constant k appropriately, e.g., if we set $k = 1,000$, each of the parameters, i.e., Δt_2, η_2, and v^{cl}_2 becomes 10 ps, 0.89 Pa·s, and 1 μm/s, respectively.

- Extraction speed

With regard to the extraction speed, the value of v^{cl}_1 is remarkably larger than that adopted in previous experiments (20, 21, 24). Janovjak et al. examined the v^{cl}-dependence of the peak positions of the F–D curves, and observed no remarkable difference between those measured by an experiment with $v^{cl} = 10$ nm/s and those with $v^{cl} = 5,230$ nm/s (18). With faster extraction velocities, i.e., 1–50 m/s, Kappel and Grubmüller performed all-atom simulations of the forced unfolding of bR (29). Overall, their F–D curves exhibited similar pattern to those obtained by AFM. They proposed, however, velocity-dependent unfolding mechanism based on the kinetics of intramolecular interaction network.

Making use of the similarity rule as mentioned above, our simulation can be interpreted as that with another extraction speed, $v^{cl}_2 = v^{cl}_1/k$. The value of the k is chosen so that it represents the ratio of the PM viscosity to that of bulk water. Although the direct estimation of the PM viscosity is not straightforward as described in the next section, v^{cl}_2 almost falls within the range of experimental extraction velocities if k takes value from 180 to 4,800. We would like to emphasize, here, that $k = 1000$ is a reasonable choice of k because the v^{cl}_2 is close to the experimental condition (24), the value of k_u, i.e., the unfolding rate under no applied force, agrees well with the experiment (24), and η_2 has a realistic value for a polypeptide chain in a lipid bilayer (Supporting material S5).

- Viscosity

With regard to the viscosity of the system, we set the viscosity η_1 to that of water η_W ($= 8.9 \times 10^{-4}$ Pa·s) in the simulations. Here, we should take care that the polypeptide chain experiences two different environments during the simulation, i.e., water environment felt by SPC and a lipid environment felt by the remaining portion of bR. We assumed that the latter environment is of central importance in this study for the following reason: We can expect that SPC behaves just like an elastic string connecting the cantilever stylus and the edge of the remaining polypeptide chain that resides in the

membrane. It is likely that this part only mediates mechanical forces from the cantilever, and the viscosity of the environment around SPC does not affect the property of the F–D curves. According to the literature (50), the viscosity value of the aliphatic hydrocarbon with C-14 chain length, i.e., *n*-tetradecane, is 0.002Pa·s, which is two times that of water. Although it is difficult to evaluate the magnitude of the viscosity of PM (η_m), we expect that the value is at least several times larger than η_W, ranging from $180\eta_W$ (51–53) to $4800\eta_W$ (54–55) at *T*=298K.

- Time step

With regard to the time step of the simulation, the following conditions should be satisfied in the BD simulation (40):

$$\exp\left(-\frac{\Delta t}{\tau^{\mathrm{T}}}\right) \ll 1 \Leftrightarrow \Delta t \gg \tau^{\mathrm{T}} = mD^{\mathrm{T}}/k_b T = \frac{m}{6\pi\eta a},$$

$$\exp\left(-\frac{\Delta t}{\tau^{\mathrm{R}}}\right) \ll 1 \Leftrightarrow \Delta t \gg \tau^{\mathrm{R}} = ID^{\mathrm{R}}/k_b T = \frac{m}{20\pi\eta a}, (D^{\mathrm{R}} = k_b T/8\pi\eta a^3, I = (2/5)ma^2)$$

, (7)

where τ^{T} (τ^{R}) and D^{T} (D^{R}) are the translational (rotational) relaxation time and the diffusion coefficient, respectively. *m*, *I*, and *a*, are effective mass, moment of inertia and the radius of a PBP, respectively. If we set $a = 4.0 \times 10^{-10}$ [m], $m = (55 \times 10^{-3}$ [kg] / $6.02 \times 10^{23})$, $\Delta t = 2\Delta t_1$, and $\eta = 2\eta_w$, then $\exp\left(-\Delta t/\tau^{\mathrm{T}}\right) = 0.053$ and $\exp\left(-\Delta t/\tau^{\mathrm{R}}\right) = 5.6 \times 10^{-5}$, indicating that the time step Δt ($\geq 2\Delta t_1$) satisfies the requirements of the BD simulation when the viscosity is ≥two times larger than that of water.

Unfolding process

In simulation **C**, we observed, without exception, that the forced unfolding of bR proceeded in four phases, i.e., $1(GF)^{\mathrm{C}}$, $2(ED)^{\mathrm{C}}$, $3(CB)^{\mathrm{C}}$, and $4(A)^{\mathrm{C}}$ phases (Fig. 6), which is in accordance with the experimental observations in the literature (17–20, 23, 24), whereas 121 out of 128 simulations showed four-phase unfolding, i.e., $1(AB)^{\mathrm{N}}$, $2(CD)^{\mathrm{N}}$ $3(EF)^{\mathrm{N}}$, and $4(G)^{\mathrm{N}}$ phases, in simulation **N** (Fig. S4). In the remaining seven runs of simulation **N**, helices F and G unfolded spontaneously and protruded from the membrane prior to the forced unfolding of helix E. We examined the relationships between the peak positions of the simulated F–D curves and the structures of the partially unfolded bR and found that there is one-to-one correspondence between the peak positions and the folding patterns of bR in the membrane, which is established

13

almost perfectly.

Forced unfolding scenario

In some cases a pair of helices participates in the energy barrier formation (Fig. 7(A)) on the force-unfolding pathways obtained by the simulations.

I: Strong forces toward the $-z$ direction originate from part "a" and "c" against the extraction force. On average, ~60pN per residue is generated from each of the hydrophobic side chains in part "a" on the upper side of the membrane, while contribution from the main chain is only 11 pN toward the opposite direction per PBP because the main chain units in part "a" participate in hydrogen bonding. Here, the magnitude of the forces were evaluated by the equation, $F = -\dfrac{dg}{dz}\Big|_{z=+1.75nm} \times \dfrac{\Delta U}{6.02 \times 10^{23}}$,

where ΔU represents the transfer free energy from water to lipid environment and the function g is provided in Eq. 2. For hydrophobic side chains, ΔU ranges from –11.7 (Leu) to –14.3 (Phe) kJ/mol (Table 1), while that of H-bonded mainchain unit is estimated to be +2.3 kJ/mol (Supporting material S6). Part "b" forms stable α-helix and transmit forces from part "a" to "c". Part "c" consists of two subparts: (1) hydrophobic residues forming helix structure and (2) hydrophilic residues connected to the loop outside the membrane. Among these hydrophilic residues, the side chains of Glu, Arg, and Asp cause a strong force of ~100pN per side chain toward the $-z$ direction.

II: Part "i" plays key role in the resistance against the extraction force, while part "h" forms a stable α-helix in the middle of the membrane. The contribution to the extraction force from part "f" is much smaller than expected, even though this part consists of hydrophobic residues. This is because that part "f" is extended without forming α-helical conformation, and main chain amino and carbonyl groups are exposed. Therefore, the effects of hydrophobic and hydrophilic groups on the free energy of transfer from the lipid phase to water cancel each other out.

III: Parts "l" and "m" are extended and contain hydrophilic residues, i.e., Glu, Arg, and Asp, giving rise to significant resistance against the extraction force, ~160pN in total, because the forces acting on both side and main chains are directed toward the $-z$ direction. On the other hand, part "j" behaves like part "f" as mentioned above, and has little effect on the resistance against the extraction force. Thus, most extraction forces are transmitted to part "l" through part "k".

IV, V, and VI: Although there are some hydrophobic residues in parts "n", "o" and "r", the magnitudes of the forces in the simulated F–D curves were small in these cases. According to the literature (56, 57), some amino acid residues are preferentially absorbed in the membrane-water interface. We may need to refine our force-field functions to evaluate the extraction forces for these types, IV–VI.

In some cases a single helix participates in the energy barrier formation (Fig. 7(B)). As far as the roles of fragments in the interface regions are concerned, there is strong similarity in these cases (Fig. 7(B)) and the previous cases where a pair of helices is involved (Fig. 7(A)). Type VII, VIII, and IX correspond to type I, II, and III, respectively, while type X corresponds to type V and VI.

As we have seen so far, we classified the unfolding intermediate states into ten types, focusing our attention especially on the interface region of the bilayer. Using single α-helical peptides as a model of TM parts of helical membrane proteins, Ganchev et al. performed forced unfolding experiments, and concluded that the interface region plays an important role in stability anchoring TM α-helices into membranes (58). Furthermore, they examined the F–D curves of the peptides in the ordered peptide-rich bilayers with those in the unordered peptide-poor bilayers, and demonstrated that the forced unfolding of the peptides occurred by the same mechanism in the both cases, indicating that the unfolding processes of the helical membrane proteins are controlled by the first-stage interaction of the two-stage model.

It is likely that the interface region plays central role in membrane protein folding as well (56). To address the issue, computer simulation techniques have been employed, and they demonstrated that (i) the polypeptide chain was absorbed in the interface region at the beginning, then (ii) the formation of α-helix took place there, and (iii) the integration of the transmembrane helix completed (36–39). Interestingly, we observed unfolding intermediates similar to the folding intermediate of the second step (ii), i.e., peak positions at 54.0aa, 118.9aa, and 174~188aa (Fig. 6) and 45.7aa, 53.9aa, and 105.7aa (Fig. S4), indicating that folding and unfolding pathways of membrane proteins may share some common intermediates.

Height and width of energy barrier

Janovjak et al. estimated the heights of the energy barriers that exist on the forced unfolding pathways based on the F–D curve measurements, and these values were

19–33 $k_b T$ (22).

Here we estimated the heights of the energy barriers of types I and III. The thickness of the interface region between the membrane and water is ~0.7nm and this region accommodate approximately four hydrophobic residues for the type I case. The average transfer energy of the hydrophobic side chains of (I, L, F, W, Y, V) is ~12 kJ/mol. Thus, the magnitude of the energy barrier for type I is estimated to be ~12 × 4 = 48 kJ/mol (~19 $k_b T$). For type III, the average transfer energy of the charged side chains (R, E, D) is ~ 20 kJ/mol and that of the main chain unit is ~12.4 kJ/mol. Therefore, the height of the energy barrier for type III is approximately 20 + 12.4 ×2 = 44.8 kJ/mol (~18 $k_b T$) as insertion of an ionic side chain and two main-chain units of part "l" and/or "m" into the membrane.

For type I and III, it is possible to assume that the widths of the energy barriers correspond to the thickness of a membrane interface region (~0.7nm) in agreement with the estimation of 0.32~0.86nm based on experimental F-D curves in the literature (18). Note that it is possible to estimate the widths and heights of the energy barriers by using the loading-rate dependence of the magnitude of peak forces (59–61). Focusing especially on the peaks at 90.2-aa and 158.0-aa, we characterize the energy barriers in detail (Supporting material S5).

Branch of forced unfolding pathways

It seems that there is almost one-to-one correspondence between the force peak positions and the unfolding intermediate states with a few exceptions.

In simulation *C*, we observed two different states, i.e., 35.0aa-A and 35.0aa-B, shared the same peak position with extracting forces of 15~35pN and 50~75pN, respectively. Figure 4(C) demonstrates more frequent occurrence of 35.0aa-A than 35.0aa-B, indicating that the pathway via 35.0aa-A is more favorable than that via 35.0aa-B.

Roles of class-I and class-II interactions

So far, we have focused on the peak positions in the F–D curves and the roles of the class-I interaction in forced unfolding of bR. Here, we would like to emphasize that the peak position and the length of the stretched polypeptide chain (SPC) are primarily determined by the boundary residue, which lies on the upper membrane surface, between SPC and the TM helix that is directly connected to SPC. We hereafter call such

a helix "primary helix". Note that a primary helix can be either fully folded, partially folded, or fully extended (Fig. 6), depending on which unfolding intermediate it belongs. To predict each peak position, therefore, we do not need to predict the complete 3D-structure of the entire polypeptide chain at each unfolding intermediate state. Rather, each peak position, which depends on the position of the boundary residue, mostly depends on the structure of the primary helix in each unfolding intermediate.

In what follows, we compare the structure of each TM helix and that of the X-ray structure (49) (Fig. S2B). Seeber et al. conducted atomistic MD simulations (27) of bacterioopsin with the implicit membrane/water IMM1 model. They observed that the Cα-RMSD value of each TM helix ranged from 0.7 to 1.7 Å. Although our simulations in this study are based on the C-G model, we reproduced the structure of each individual TM helix accurately, except for the B-helix.

Furthermore, we compared each helix pair between the X-ray structure (49) and the CG-MD structures (Fig. S2C) because some unfolding intermediates are characterized by the helix pairs (Fig. 6). It turned out that the fold of each helix pair was qualitatively reproduced, except for those including the B-helix. Despite the lack of inter-helix interactions, the positional restraints by the inter-helical loop, probably, worked well to hold the pair of helices together.

As we described, our simulations accurately reproduced the structure of each TM helix, which has to do with the reason why our force field works for reproducing the peak positions despite the relative positions of TM helices are not accurate due to the lack of inter-helical interactions.

Next, we discuss the roles of the class-II interaction. According to the experimental studies (13, 16, 23, 24), the magnitude and the appearance frequency of the force peaks of the F–D curves were affected by the inter-helix interactions, whereas the peak positions were not. In what follows, we discuss possible roles of the class-II interaction.

In the initial state of bR before extraction, we recognize interactions between the Schiff-base retinal and D^{85}, D^{212} and R^{82}, and E^{194} and E^{204}, which should increase the barrier heights of some force peaks. However, Sapra et al. reported that the triple mutation E9Q/E194Q/E204Q did not affected the peak positions in the F-D curves of bR (24). In our opinion, these interactions play important roles in biological functions more than in the structure formation.

With regard to the inter-helical hydrogen bonding, Kappel et al. performed forced

unfolding simulation of bR at the atomic level and analyzed the structure of each anchor point (29), which was stabilized by hydrogen bonds and hydrophobic contacts. During forced unfolding of bR, they observed that these interactions form a highly dynamic network. It is possible that the relative positions of TM helices are quickly optimized and stabilized during extraction processes. They reported the list of anchor points whose structure are stabilized by inter-helical hydrogen bonds. Among them, the position of Thr17 is close to the experimental peak position of Gly21 (20). It is possible that inter-helical hydrogen bonds increase some energy barrier heights of force peaks. Further investigation is needed to study the role of inter-helical hydrogen bonding in forced unfolding of membrane proteins.

Sapra et al. performed forced unfolding experiments on P50A, M56A, Y57A, P91A, and P186A mutants of bR (23). As a result, they observed no change in the peak positions in the F-D curves, although changes in inter-helical packing were expected in these mutants. Incorporating these inter-helix interactions in the force-field functions may improve the accuracy of the potential energy surface, leading to better results of F–D curve simulations.

SUMMARY AND OUTLOOK

We have performed forced unfolding simulations of bR with a novel coarse-grained model and examined the roles of the interactions that stabilize each individual TM helix in the membrane environment in the forced unfolding processes. We are particularly interested in the peak positions in the F–D curves and the agreement between the force peak positions obtained in the simulations and those by the AFM experiments was good, indicating that the interactions considered play important roles in the characterization of the F–D curves. Furthermore, we investigated the relationships between the energy barrier formation on the forced unfolding pathways and the force peaks of the F–D curves.

We plan to distribute the computer program via the website (http://www.comp-biophys.com/yamato-lab/resources.html) in the near future. The application of the current method to other membrane proteins, i.e., halorhodopsin, WALP peptide, and sodium-proton antiporter, are currently in progress and will be reported elsewhere.

AUTHOR CONTRIBUTIONS

T. Yamada: performed research, analyzed data, wrote the paper. T. Yamato: examination of computer programs, interpretation of data, wrote the paper. S. M.: study conception and design, interpretation of data, wrote the paper.

ACKNOWLEDGEMENTS

FIGURE CAPTIONS

Figure 1. A forced unfolding scenario

During the forced unfolding (4), the polypeptide chain of bR is extracted. We denote the polypeptide chain spanning from the stylus to the PM surface and its residue length as SPC (16) and L_{SPC}, respectively.

(A) The stylus of the cantilever is bound to the C-terminus of bR. While the cantilever is raised, SPC goes slack immediately after the F and G-helices are extracted (left), and the cantilever continues moving upward with monotonically increasing tension force until the SPC is fully stretched. Then, we observe a force peak when the helices E and D are unfolded and extracted. After helices E and D are completely extracted, the tension force is again relaxed (right).

(B) Several distinct peaks are observed on a typical F–D curve.

(C) The relationships between the force peaks and energy barrier on the unfolding pathways. It is likely there are multiple unfolding pathways (23) because the patterns of F–D curves are not completely identical in all trials of the forced unfolding experiments.

Figure 2. Peptide bond particle

The orthonormal vectors $\hat{x}, \hat{y}, \hat{z}$ represent LCS associated with the PBP with the octet, $\{C^{\alpha C},\ C,\ O,\ N,\ H,\ C^{\alpha N}\ hbo,\ hbh\}$, where $C^{\alpha C}$ ($C^{\alpha N}$) represents the C^{α} atom in the previous (next) residue attached to C (N), and hbo (hbh) represents "bonding point" near O (H) on which the H-bonding counterparts are attached (Fig. S1(B)). The contact radius $r^{rep} = 0.2366$ nm of the PBP is shown as a gray sphere.

Figure 3. Model system

GCS was introduced so that the xy-plane became the middle plane of the membrane system, and the z-axis was set perpendicular to it. The polypeptide chain is excluded from the gray regions D_{I} and D_{II}.

$$\begin{cases} D_{\mathrm{I}}: -2.5nm \leq z \leq 2.5nm \ \text{ and }\ \left(x^2 + y^2\right)^{1/2} > 2.0nm, \\ D_{\mathrm{II}}: z \leq -3.5nm. \end{cases}$$

The shape of the smoothing function $g(z)$ is shown on the right.

Figure 4. F–D curves and force peak analysis

(A, B) Superimposed plots of 32 F–D curves obtained in simulations C and N are shown in Fig. 4A and B, respectively. The vertical axis represents the force exerted on the cantilever stylus through SPC, while the horizontal axis represents the distance between the upper surface of the membrane and the cantilever stylus. A single trajectory consists of 10,000 data points, and these values were averaged every 10 points. Thus, we obtained 1,000 representative averaged data points for each trajectory. Figure S3A and B illustrate the force peak analysis for simulations C and N, respectively.

(C, D) A plot of the force peaks obtained from the 128 F–D curves for extraction from C-terminus (C) and N-terminus (D) as a function of L_{SPC}. We see distinct clusters of 13 (18) force peaks in Fig. 4C (D). For each cluster, the average L_{SPC} is indicated in the figure.

Figure 5. Comparison of force peak positions between simulation and experiments

Force peak positions and appearance frequencies of L_{SPC} are shown for simulation C (N) in the upper two lines of Fig. 5A (B), where the horizontal x-axis represents $L_{\mathrm{SPC}} + 1$ (L_{SPC}). Because we used the G-241 deletion mutant of bR, the value of L_{SPC} is smaller by 1 than that would be if we used the wild type for simulation C. Note that there was

no peak near the C-terminal end, i.e., amino acids 241–247. Force peak positions (frequency peak positions of L_{SPC}) are indicated by large closed circle (square), large open circle (square), or small open circle (square), depending on whether its appearance probability x (frequency $F(k_p)$) satisfies, $x \geq 0.9$ ($F(k_p) \geq 0.1$), $0.9 > x \geq 0.3$ ($0.1 > F(k_p) \geq 0.01$), or $0.3 > x$ ($0.01 > F(k_p)$), respectively.

For comparison, the experimental results of peak positions are shown for the case of extraction from the C (20, 21, 24) (N (20, 21)) terminus in the lower three (two) lines of Fig. 5A (B). Kessler et al. measured the force peaks with the polypeptide chain length between the terminal loop directly attached to the cantilever and the anchoring point. They assumed that the anchoring point is several angstroms below the membrane surface and the z-coordinates of each anchoring point are shown in Fig. 3 in a previous report (20).

Figure 6. The forced unfolding process
Each phase proceeded from the left to the right as shown in different rows representing $1(GF)^C$, $2(ED)^C$, $3(CB)^C$, and $4(A)^C$ phases from top to bottom in this order. For each peak position of the simulated F–D curves, we observed each individual intermediate state (visible intermediate). Furthermore, we observed additional intermediate states (hidden intermediates) only detected in the frequency distribution analysis. Snapshot images of each intermediate structure are shown with L_{SPC}. If there are multiple conformers for the same L_{SPC}, each conformer is distinguished by an additional suffix, -A or -B. Each snapshot of the hidden intermediates was enclosed in a dashed-line box. The gray arrow on the top of each row represents the unfolding pathway of each phase where the occurrence number was indicated at each branching point.

Figure 7. Ten patterns (I–X) of energy barrier formation
Pair of helices (A) and single helix (B). The hydrophobic (hydrophilic) residues that remain near the membrane surface are represented by white (gray) circles. Black arrows show the transitions from one pattern to another. The correspondence between each pattern and the force-peak position is as follows. I: 24.4aa, 145.6aa; II: 90.2aa, 151.6aa; III: 35.0aa-B, 158.0aa; IV: 54.0aa, 118.9aa, 174aa~188aa; V: 43.1aa, 96.2aa, 106.3aa, 165.8aa, 171.3aa; VI: 35.0aa-A; VII: 213.2aa; VIII: 217.4aa; IX: 225.8aa, 227.9aa, 231.0aa-B; X: 231.0aa-A. In simulation N, all structures at force-peak positions were

classified as types I–X, except for that at 39.9aa.

Table 1. Free energy of transfer from the bulk water to pure lipid environment

Side Chain			
Fragment	ΔU^{sc} [kJ/mol]	Fragment	ΔU^{sc} [kJ/mol]
Ala	−3.5	Leu	−11.7
Arg	+21.0	Lys	+11.8
Asn	+9.6	Met	−6.8
Asp	+23.9	Phe	−14.3
Cys	−2.7	Pro	−7.0
Gln	+6.2	Ser	+5.8
Glu	+19.3	Thr	+4.5
Gly	−1.3	Trp	−13.4
His	+2.2	Tyr	−9.7
Ile	−11.7	Val	−8.7
Main Chain			
Fragment	ΔU^{mc} [kJ/mol]	Fragment	ΔU^{mc} [kJ/mol]
N-terminus	+9.8	>CH-CONH-	+12.4
C-terminus	+27.1		

Free energy of transfer of various fragments, i.e., side- and main chain, N-, and C-termini, from bulk water to pure lipid environment at pH = 7.8. These values were derived from the octanol–water partition coefficients (43) and partially modified in this study (Tables S2 and S3). The value of ΔU^{mc} for the N(C)-terminus, $\Delta U^{\text{N-term}}$ ($\Delta U^{\text{C-term}}$), was evaluated for the system consisting of $-NH_3^+$ and $-NH_2$ (-COOH and $-COO^-$) in chemical equilibrium at pH = 7.8. Similarly, the values of ΔU^{sc}s for the ionizable side chains were evaluated for the equilibrium system consisting of the protonated and deprotonated side chains.

Supporting Citations

References (62–66) appear in the Supporting Material.

References

1. Oesterhelt, D., and W. Stoeckenius. 1973. Functions of a new photoreceptor membrane. Proc. Natl. Acad. Sci. U. S. A. 70: 2853–7.

2. Krogh, A., B. Larsson, G. von Heijne, and E.L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305: 567–80.

3. Sawada, R., and S. Mitaku. 2012. Biological meaning of DNA compositional biases evaluated by ratio of membrane proteins. J. Biochem. 151: 189–96.

4. Oesterhelt, F., D. Oesterhelt, M. Pfeiffer, A. Engel, H.E. Gaub, and D.J. Müller. 2000. Unfolding pathways of individual bacteriorhodopsins. Science. 288: 143–6.

5. Janovjak, H., K.T. Sapra, A. Kedrov, and D.J. Müller. 2008. From valleys to ridges: exploring the dynamic energy landscape of single membrane proteins. Chemphyschem. 9: 954–66.

6. Bippes, C.A., and D.J. Muller. 2011. High-resolution atomic force microscopy and spectroscopy of native membrane proteins. Reports Prog. Phys. 74: 086601.

7. Möller, C., D. Fotiadis, K. Suda, A. Engel, M. Kessler, and D.J. Müller. 2003. Determining molecular forces that stabilize human aquaporin-1. J. Struct. Biol. 142: 369–78.

8. Kedrov, A., C. Ziegler, H. Janovjak, W. Kühlbrandt, and D.J. Müller. 2004. Controlled unfolding and refolding of a single sodium-proton antiporter using atomic force microscopy. J. Mol. Biol. 340: 1143–52.

9. Kedrov, A., M. Krieg, C. Ziegler, W. Kuhlbrandt, and D.J. Muller. 2005. Locating ligand binding and activation of a single antiporter. EMBO Rep. 6: 668–74.

10. Cisneros, D.A., D. Oesterhelt, and D.J. Müller. 2005. Probing origins of molecular interactions stabilizing the membrane proteins halorhodopsin and bacteriorhodopsin. Structure. 13: 235–42.

11. Tanuj Sapra, K., P.S.-H. Park, S. Filipek, A. Engel, D.J. Müller, and K. Palczewski. 2006. Detecting molecular interactions that stabilize native bovine rhodopsin. J. Mol. Biol. 358: 255–69.

12. Kedrov, A., S. Wegmann, S.H.J. Smits, P. Goswami, H. Baumann, and D.J. Muller. 2007. Detecting molecular interactions that stabilize, activate and guide ligand-binding of the sodium/proton antiporter MjNhaP1 from Methanococcus jannaschii. J. Struct. Biol. 159: 290–301.

13. Cisneros, D.A., L. Oberbarnscheidt, A. Pannier, J.P. Klare, J. Helenius, M. Engelhard, F. Oesterhelt, and D.J. Muller. 2008. Transducer binding establishes localized interactions to tune sensory rhodopsin II. Structure. 16: 1206–13.

14. Kawamura, S., A.T. Colozo, D.J. Müller, and P.S.-H. Park. 2010. Conservation of molecular interactions stabilizing bovine and mouse rhodopsin. Biochemistry. 49: 10412–20.

15. Serdiuk, T., M.G. Madej, J. Sugihara, S. Kawamura, S.A. Mari, H.R. Kaback, and D.J. Müller. 2014. Substrate-induced changes in the structural properties of LacY. Proc. Natl. Acad. Sci. U. S. A. 111: E1571–80.

16. Müller, D.J., M. Kessler, F. Oesterhelt, C. Möller, D. Oesterhelt, and H. Gaub. 2002. Stability of bacteriorhodopsin alpha-helices and loops analyzed by single-molecule force spectroscopy. Biophys. J. 83: 3578–88.

17. Janovjak, H., M. Kessler, D. Oesterhelt, H. Gaub, and D.J. Müller. 2003. Unfolding pathways of native bacteriorhodopsin depend on temperature. EMBO J. 22: 5220–9.

18. Janovjak, H., J. Struckmeier, M. Hubain, A. Kedrov, M. Kessler, and D.J. Müller. 2004. Probing the energy landscape of the membrane protein bacteriorhodopsin. Structure. 12: 871–9.

19. Sapra, K.T., H. Besir, D. Oesterhelt, and D.J. Muller. 2006. Characterizing molecular interactions in different bacteriorhodopsin assemblies by single-molecule force spectroscopy. J. Mol. Biol. 355: 640–50.

20. 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.

21. Voïtchovsky, K., S.A. Contera, and J.F. Ryan. 2007. Electrostatic and steric interactions determine bacteriorhodopsin single-molecule biomechanics. Biophys. J. 93: 2024–37.

22. Janovjak, H., H. Knaus, and D.J. Muller. 2007. Transmembrane helices have rough energy surfaces. J. Am. Chem. Soc. 129: 246–7.

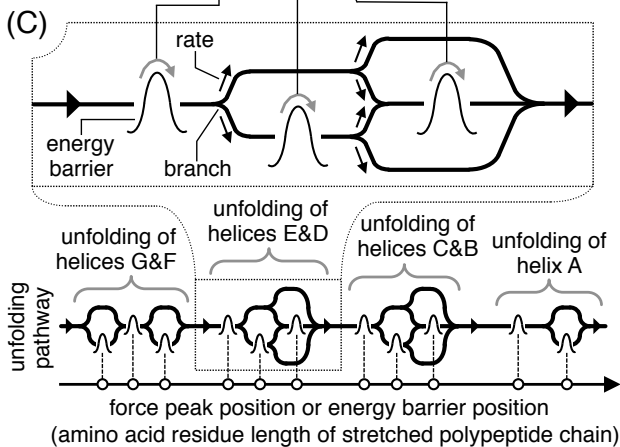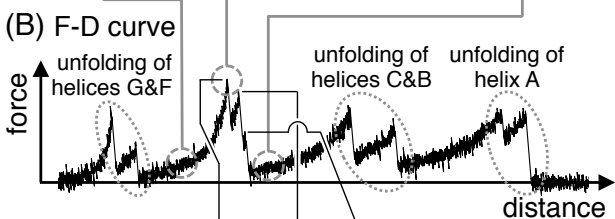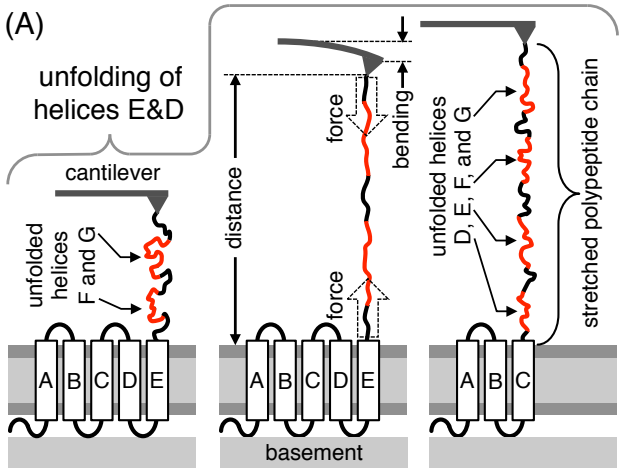23. Sapra, K.T., G.P. Balasubramanian, D. Labudde, J.U. Bowie, and D.J. Muller.

2008. Point mutations in membrane proteins reshape energy landscape and populate different unfolding pathways. J. Mol. Biol. 376: 1076–90.

24.　Sapra, K.T., J. Doehner, V. Renugopalakrishnan, E. Padrós, and D.J. Muller. 2008. Role of extracellular glutamic acids in the stability and energy landscape of bacteriorhodopsin. Biophys. J. 95: 3407–18.

25.　Stoeckenius, W., and R.A. Bogomolni. 1982. Bacteriorhodopsin and related pigments of halobacteria. Annu. Rev. Biochem. 51: 587–616.

26.　Haupts, U., J. Tittor, and D. Oesterhelt. 1999. Closing in on bacteriorhodopsin: progress in understanding the molecule. Annu. Rev. Biophys. Biomol. Struct. 28: 367–99.

27.　Seeber, M., F. Fanelli, E. Paci, and A. Caflisch. 2006. Sequential unfolding of individual helices of bacterioopsin observed in molecular dynamics simulations of extraction from the purple membrane. Biophys. J. 91: 3276–84.

28.　Fanelli, F., and M. Seeber. 2010. Structural insights into retinitis pigmentosa from unfolding simulations of rhodopsin mutants. FASEB J. 24: 3196–209.

29.　Kappel, C., and H. Grubmüller. 2011. Velocity-dependent mechanical unfolding of bacteriorhodopsin is governed by a dynamic interaction network. Biophys. J. 100: 1109–19.

30.　Cieplak, M., S. Filipek, H. Janovjak, and K.A. Krzyśko. 2006. Pulling single bacteriorhodopsin out of a membrane: Comparison of simulation and experiment. Biochim. Biophys. Acta. 1758: 537–44.

31.　Kessler, M., K.E. Gottschalk, H. Janovjak, D.J. Muller, and H.E. Gaub. 2006. Bacteriorhodopsin folds into the membrane against an external force. J. Mol. Biol. 357: 644–54.

32.　Popot, J.L., S.E. Gerchman, and D.M. Engelman. 1987. Refolding of bacteriorhodopsin in lipid bilayers. A thermodynamically controlled two-stage process. J. Mol. Biol. 198: 655–76.

33.　Popot, J.L., and D.M. Engelman. 1990. Membrane protein folding and oligomerization: the two-stage model. Biochemistry. 29: 4031–7.

34.　Hunt, J.F., P. Rath, K.J. Rothschild, and D.M. Engelman. 1997. Spontaneous, pH-dependent membrane insertion of a transbilayer alpha-helix. Biochemistry. 36: 15177–92.

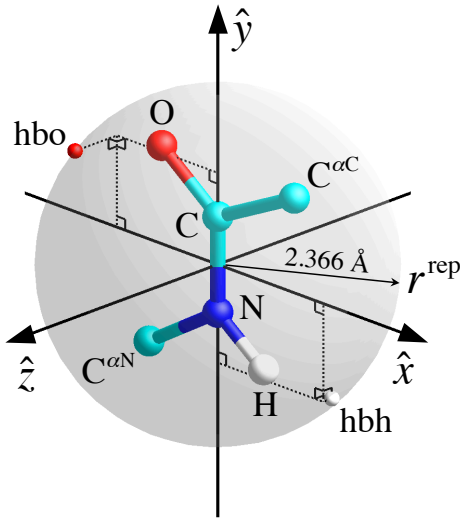35.　Popot, J.L., and D.M. Engelman. 2000. Helical membrane protein folding,

stability, and evolution. Annu. Rev. Biochem. 69: 881–922.

36.   Milik, M., and J. Skolnick. 1993. Insertion of peptide chains into lipid membranes: an off-lattice Monte Carlo dynamics model. Proteins. 15: 10–25.

37.   Baumgärtner, A. 1996. Insertion and hairpin formation of membrane proteins: a Monte Carlo study. Biophys. J. 71: 1248–55.

38.   Maddox, M.W., and M.L. Longo. 2002. A Monte Carlo study of peptide insertion into lipid bilayers: equilibrium conformations and insertion mechanisms. Biophys. J. 82: 244–63.

39.   Kessel, A., D. Shental-Bechor, T. Haliloglu, and N. Ben-Tal. 2003. Interactions of hydrophobic peptides with lipid bilayers: Monte Carlo simulations with M2delta. Biophys. J. 85: 3431–44.

40.   Ermak, D.L., and J.A. McCammon. 1978. Brownian dynamics with hydrodynamic interactions. J. Chem. Phys. 69: 1352.

41.   Fernandes, M.X., and J.G. de la Torre. 2002. Brownian dynamics simulation of rigid particles of arbitrary shape in external fields. Biophys. J. 83: 3039–48.

42.   Chang, C.-E., T. Shen, J. Trylska, V. Tozzini, and J.A. McCammon. 2006. Gated binding of ligands to HIV-1 protease: Brownian dynamics simulations in a coarse-grained model. Biophys. J. 90: 3880–5.

43.   Roseman, M.A. 1988. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. J. Mol. Biol. 200: 513–22.

44.   Leo, A., C. Hansch, and D. Elkins. 1971. Partition coefficients and their uses. Chem. Rev. 71: 525–616.

45.   Mills, F.D., V.C. Antharam, O.K. Ganesh, D.W. Elliott, S.A. McNeill, and J.R. Long. 2008. The helical structure of surfactant peptide KL4 when bound to POPC: POPG lipid vesicles. Biochemistry. 47: 8292–300.

46.   Klotz, I.M., and J.S. Franzen. 1962. Hydrogen Bonds between Model Peptide Groups in Solution. J. Am. Chem. Soc. 84: 3461–3466.

47.   Frishman, D., and P. Argos. 1995. Knowledge-based protein secondary structure assignment. Proteins. 23: 566–79.

48.   Takada, S., Z. Luthey-Schulten, and P.G. Wolynes. 1999. Folding dynamics with nonadditive forces: A simulation study of a designed helical protein and a random heteropolymer. J. Chem. Phys. 110: 11616.

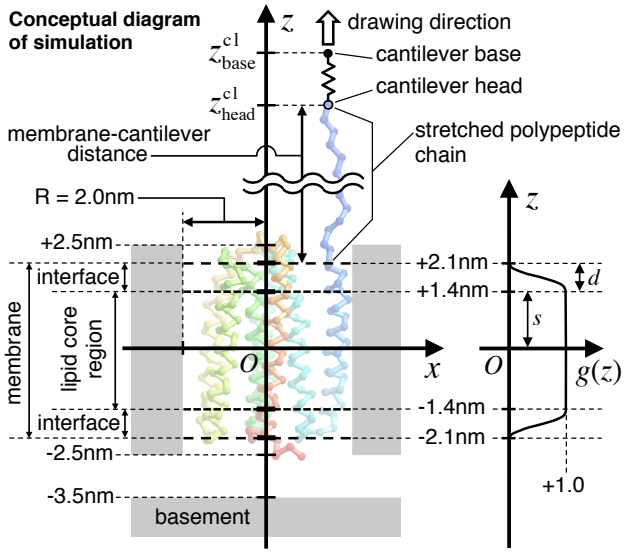49.   Essen, L.O., R. Siegert, W.D. Lehmann, and D. Oesterhelt. 1998. Lipid patches in

membrane protein oligomers: Crystal structure of the bacteriorhodopsin-lipid complex. Proc Natl Acad Sci U S A. 95: 11673–11678.

50. Aralaguppi, M.I., T.M. Aminabhavi, R.H. Balundgi, and S.S. Joshi. 1991. Thermodynamic interactions in mixtures of bromoform with hydrocarbons. J. Phys. Chem. 95: 5299–5308.

51. Vaz, W.L., D. Hallmann, R.M. Clegg, A. Gambacorta, and M. De Rosa. 1985. A comparison of the translational diffusion of a normal and a membrane-spanning lipid in L alpha phase 1-palmitoyl-2-oleoylphosphatidylcholine bilayers. Eur. Biophys. J. 12: 19–24.

52. Vaz, W. L., R. M. Clegg, and D. Hallmann. 1985. Translational diffusion of lipids in liquid crystalline phase phosphatidylcholine multibilayers. A comparison of experiment with theory. Biochemistry. 24: 781–786.

53. Shinitzky, M., and M. Inbar. 1976. Microviscosity parameters and protein mobility in biological membranes. BBA-Biomembranes. 433: 133–149.

54. Vaz, W.L.C., F. Goodsaid-Zalduondo, and K. Jacobson. 1984. Lateral diffusion of lipids and proteins in bilayer membranes. FEBS Lett. 174: 199–207.

55. Peters, R., and R. J. Cherry. 1982. Lateral and rotational diffusion of bacteriorhodopsin in lipid bilayers: experimental test of the Saffman-Delbrück equations. Proc Natl Acad Sci U S A. 79: 4317-4321.

56. Jacobs, R.E., and S.H. White. 1989. The nature of the hydrophobic binding of small peptides at the bilayer interface: implications for the insertion of transbilayer helices. Biochemistry. 28: 3421–37.

57. Wimley, W.C., and S.H. White. 1996. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat. Struct. Biol. 3: 842–8.

58. Ganchev, D.N., D.T.S. Rijkers, M.M.E. Snel, J.A. Killian, and B. de Kruijff. 2004. Strength of integration of transmembrane alpha-helical peptides in lipid bilayers as determined by atomic force spectroscopy. Biochemistry. 43: 14987–93.

59. Bell, G. 1978. Models for the specific adhesion of cells to cells. Science. 200: 618–627.

60. Evans, E., and K. Ritchie. 1997. Dynamic strength of molecular adhesion bonds. Biophys. J. 72: 1541–55.

61. Evans, E. 1999. Looking inside molecular bonds at biological interfaces with dynamic force spectroscopy. Biophys. Chem. 82: 83–97.

62. Pauling, L. 1960. The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry. Cornell university press, Ithaca, NY.

63. Maillard, M.C., M.E. Perlman, O. Amitay, D. Baxter, D. Berlove, S. Connaughton, J.B. Fischer, J.Q. Guo, L.Y. Hu, R.N. McBurney, P.I. Nagy, K. Subbarao, E.A. Yost, L. Zhang, and G.J. Durant. 1998. Design, synthesis, and pharmacological evaluation of conformationally constrained analogues of N,N'-diaryl- and N-aryl-N-aralkylguanidines as potent inhibitors of neuronal Na+ channels. J. Med. Chem. 41: 3048–61.

64. Nozaki, Y., and C. Tanford. 1967. Examination of titration behavior. Methods Enzymol. 11: 715–734.

65. Roseman, M. A. 1988. Hydrophobicity of the peptide $C=O \cdots H\text{-}N$ hydrogen-bonded group. J. Mol. Biol. 201: 621-623.

66. Gilson, M. K., J. A. Given, B. L. Bush, and J. A. McCammon. 1997. The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophys. J. 72: 1047–1069.

(A) unfolding of helices E&D

cantilever

unfolded helices F and G

force

unfolded helices D, E, F, and G

stretched polypeptide chain

distance

force

bending

A B C D E   A B C D E   A B C

basement

(B) F-D curve

force

unfolding of helices G&F

unfolding of helices C&B

unfolding of helix A

distance

(C)

rate

energy barrier

branch

unfolding pathway

unfolding of helices G&F

unfolding of helices E&D

unfolding of helices C&B

unfolding of helix A

force peak position or energy barrier position
(amino acid residue length of stretched polypeptide chain)

Conceptual diagram of simulation

drawing direction

z_{base}^{cl} — cantilever base

z_{head}^{cl} — cantilever head

membrane-cantilever distance

stretched polypeptide chain

R = 2.0nm

+2.5nm

interface

+2.1nm

d

+1.4nm

s

lipid core region

membrane

O

x

O

$g(z)$

-1.4nm

interface

-2.1nm

-2.5nm

-3.5nm

basement

+1.0

Extraction from C-terminal side

Extraction from N-terminal side

Extraction from C-terminal side

1(GF)^C

128

24.4aa

104
3
104
21
3

z
+2.1nm
+1.4nm
0nm
-1.4nm
-2.1nm

Helix F Helix G

F219-GL-^{222}
A^{196}-G-^{198}

35.0aa-A

L^{207}-FMV-L^{211}

35.0aa-B

L206-LF-M^{209}
E^{194}-GA-G ^{197}

43.1aa

V199-PLN-^{203}

54.0aa

S^{183}-AYPV-
VWLIG-S^{193}

Helices
G & F
had been
extracted

2(ED)^C

128

90.2aa

3
10
115
2
11

Helix D Helix E

F153-FG-F156
S^{132}-YR-F^{135}

96.2aa

Y^{147}-ILY-V^{151}

106.3aa

W^{137}-WAI-S^{141}

118.9aa

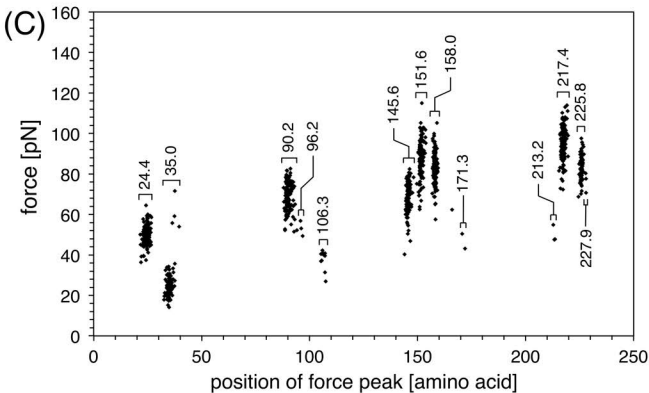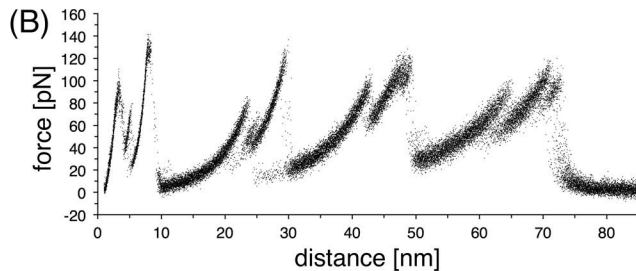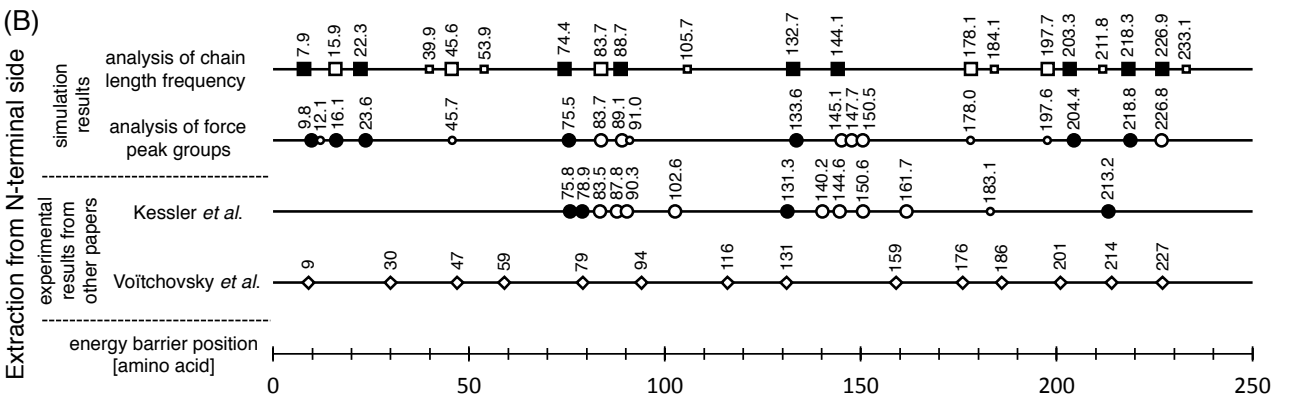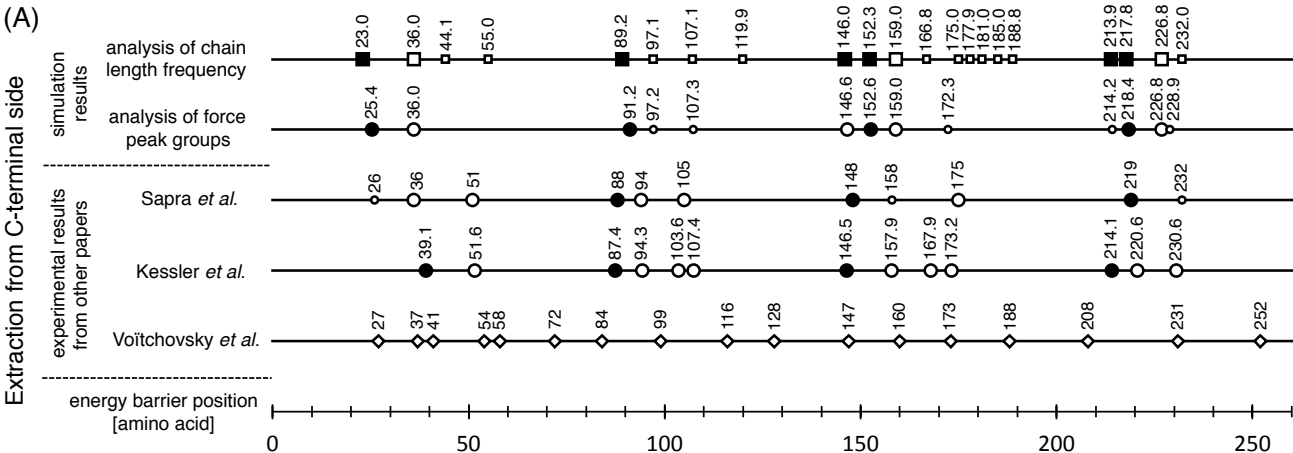L^{109}-ALVGA-
DGIMIGTG-
LVGAL-T^{128}

Helices
E & D
had been
extracted

3(CB)^C

100
27
1
98
1
104
1
1
20
106
2

145.6aa

151.6aa

158.0aa

Helix B Helix C

A^{98}-LLV-D^{102}
Q^{75}-NP-I^{78}

L^{93}-L-L^{95}
E^{74}-QNP-I^{78}

L$^-$87-F-T^{89}
Q^{75}-N-P^{77}

165.8aa

P77-IYW-A^{81}

171.3aa

P^{70}-FGGE-Q^{75}

174aa
~188aa

T^{46}-TLVPAIA-
FTMYLSMLL-
GYGLTMV-P^{70}

Helices
C & B
had been
extracted

4(A)^C

3
3
125
64
3
64
61
3

213.2aa

217.4aa

225.8aa

Helix A

G^{31}-MG-V^{34}
R^7-PEW-I^{11}

L^{28}-V^{29}
R^7-PEW-I^{11}

L^{19}-M-G^{21}
P^8-E-W^{10}

227.9aa

T^{17}-A-L^{19}
G^6-R-P^8

231.0aa-A

W^{12}-LAL-G^{16}

231.0aa-B

A^{14}-L-G^{16}
A^2-Q-I^4

Helices A &
N-terminal
had been
extracted

(A)

I
d — a
b
e — c

II
f
g
h
i

III
j
k
l
m

VI
t — r
s

V
q — o
p

IV
n

Ext.

(B)

VII

VIII

IX

X

Ext.