http://www.jstage.jst.go.jp/browse/biophysics

BIOPHYSICS

Vol. 5, pp. 25–35 (2009)
doi: 10.2142/biophysics.5.25

Stochastic emergence of multiple intermediates detected by single-molecule quasi-static mechanical unfolding of protein

Akihiro Fukagawa[1,2], Michio Hiroshima[1,2], Isao Sakane[1] and Makio Tokunaga[1,2,3,4]

[1]*Structural Biology Center, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan*
[2]*Research Center for Allergy and Immunology, RIKEN, Yokohama, Kanagawa 230-0045, Japan*
[3]*Department of Genetics, School of Life Science, The Graduate University for Advanced Studies (Sokendai), Mishima, Shizuoka 411-8540, Japan*
[4]*Department of Biological Information, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta, Midori, Yokohama, Kanagawa 226-8501, Japan*

Received 15 December, 2008; accepted 25 February, 2009

Experimental probing of a protein-folding energy landscape can be challenging, and energy landscapes comprising multiple intermediates have not yet been defined. Here, we quasi-statically unfolded single molecules of staphylococcal nuclease by constant-rate mechanical stretching with a feedback positioning system. Multiple discrete transition states were detected as force peaks, and only some of the multiple transition states emerged stochastically in each trial. This finding was confirmed by molecular dynamics simulations, and agreed with another result of the simulations which showed that individual trajectories took highly heterogeneous pathways. The presence of Ca^{2+} did not change the location of the transition states, but changed the frequency of the emergence. Transition states emerged more frequently in stabilized domains. The simulations also confirmed this feature, and showed that the stabilized domains had rugged energy surfaces. The mean energy required per residue to disrupt secondary structures was a few times the thermal energy (1–3 k_BT), which agreed with the stochastic feature. Thus, single-molecule quasi-static measurement has achieved notable success in detecting stochastic features of a huge number of possible conformations of a protein.

Corresponding author: Makio Tokunaga, Department of Biological Information, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259-B-35 Nagatsuta, Midori, Yokohama, Kanagawa 226-8501, Japan.
e-mail: mtoku@bio.titech.ac.jp

Key words: single-molecule detection, protein folding, molecular force microscopy, molecular dynamics, energy landscape

Many models have been put forward to explain the high efficiency of protein folding, and to account for variations in protein size and structure, including the framework model[1,2], the hydrophobic-collapse model[3–5] and the nucleation-condensation model[6]. The 'funnel theory'[7–10] provides a general description of folding using energy landscapes, and predicts the presence of many intermediates and multiple pathways; however, ensemble averaging makes it difficult for experiments to accurately probe the energy landscape; therefore, single-molecule force-detection experiments have been carried out using atomic force microscopy (AFM) or optical tweezers[11–13] with the aim of directly detecting mechanical properties. Stretching and unfolding of individual protein molecules has been studied as a reversible pathway, whereas atomic-level information about the mechanical unfolding process has been derived from molecular dynamics (MD) simulations[14–18].

However, experiments and simulations have revealed discrepancies between force profiles. In addition, only one or two mechanical intermediates have been detected by single-molecule experiments[17–23]. By contrast, research using spectroscopy, thermodynamics and nuclear magnetic resonance (NMR) has characterized folding processes by kinetics using schemes with multiple intermediates and

parallel pathways[24–27]. The enumeration of essential intermediates, and identification of their structures and pathways, could be the key to understanding the folding mechanisms; however, these issues remain to be resolved.

Here, we introduce direct control of the end-to-end molecular distance for single-molecule mechanical unfolding using laser-radiation pressure, which allows stretching at conditions close to equilibrium, that is, quasi-statically. Quasi-static force measurement enables the detection of free energy changes directly. Multiple transition states and stochastic pathways were detected in the unfolding of staphylococcal nuclease (SNase), and were confirmed by MD simulations. Further, we found that transition states emerged more frequently in unfolding of a stabilized domain.

Methods

Mutant protein A1C/Q149C

Expression plasmids containing mutations were constructed from pMT7-SN and mutagenic primers using a MutaGene phagemid *in vitro* mutagenesis kit (Bio-Rad)[28] and the Kunkel method[29]. The C-terminus was mutated first followed by the N-terminus, and both termini were mutated to Cys (A1C/Q149C). The expression plasmid was transformed into competent BL21(DE3)/pLysS cells (Novagen Inc.). The expressed protein was purified using a cation-exchange column SP Sepharose (GE Healthcare)[30].

Inter-/intramolecular-force microscopy

The inter-/intramolecular-force microscopy (IFM) system[31–33] (Fig. 1a) was based on an inverted microscope (IX70, Olympus) with an objective (PlanApo 60× oil, numerical aperture (NA) 1.4). A water-immersion objective (LUMPlanFl 40× W, NA 0.8, working distance (WD) 3.3 mm, Olympus) was installed above the specimen. Direct detection of cantilever displacement is essential for the feedback, instead of the detection of the angular change used in AFM; since the feedback position-control system needs detection of the absolute position of the cantilever tip, detection methods of angular change are unusable. The angle of the cantilever tip depends on not the absolute position but the relative position of the cantilever tip to the position of the cantilever root. We detected the displacement by forming a magnified image of the cantilever on a split-photodiode (magnification ~96×) by illumination with a red light-emitting diode (LED; L6112-01, Hamamatsu). The differential output of the photodiode was amplified and used as a feedback system (CF201, Sentech), which modulated the laser intensity (820 nm) for position control with a response frequency of 50 kHz. The maximum intensity of the laser was 30 mW, which exerted a force of ~100 pN on the cantilever. The displacement was calibrated by oscillating the cantilever base with a piezo actuator. As the direction of cantilever displacement was inclined at 45° from the optical axis, the detected displacement by the photosensor was proportional to cos (45°) times the displacement.

Hand-made cantilever

Flexible cantilevers (~0.2 pN/nm; 300 μm length, 20 μm width, 0.2 μm thickness) were made by pulling a heated thin coverslip (0.03 mm thick; Matsunami Glass)[31,32]. A single glass bead (3 μm in diameter; Bangs Lab) was glued to a distal corner of the thin sheet (cantilever) with epoxy adhesive (Araldite Rapid; Huntsman Advanced Materials). The distal end of the cantilever and the bead were coated with gold (~28 nm and ~10 nm thick, respectively) to reflect the laser beam and to chemisorb the cysteinyl residue of the protein, respectively.

Single-molecule force microscopy

A coverslip washed with 0.1 M KOH and rinsed with water was coated with gold (~2 nm thick) and then glued to the bottom of a dish. Next, 10 μg/ml mutant SNase in buffer (50 mM NaCl and 10 mM HEPES, pH 7.4 in the absence of Ca^{2+}, or 1 mM $CaCl_2$ in addition to the buffer in the presence of Ca^{2+}) was incubated for 5 min on the gold-coated coverslip in the dish, and any unabsorbed protein was washed away with the buffer. The dish was placed on the microscope and filled with buffer. The protein extension trial (see Supplementary Information, Fig. S1)[31–33] was repeated. Protein extension was observed in ~1% of the trials. This low ratio of signal-observed trials to all trials indicated that two or more protein molecules seldom adhered to the probe and extended, and confirmed the single-molecule extension. All experiments were carried out at 27°C. Data were recorded on a DAT tape with a sampling rate of 48 kHz, and processed with a Butterworth filter (cutoff width 2 nm) on a personal computer for noise reduction.

MD simulation

MD simulations were carried out using Discovery Studio 1.7 (Accelrys) on Windows XP-based personal computers (Terra) with a CHARMm force field and an implicit solvent model of Generalized Born with simple switching (GBSW) and a time step of 1 fs. The SNase crystallography structure (1STN, Lys6–Ser141) in the absence of Ca^{2+} was heated and equilibrated to 300 K, and used as the initial structure. The starting structure of SNase in the presence of Ca^{2+} was derived from the crystallography structure with thymidine 3′,5′-diphosphate (pdTp)/Ca^{2+} complex (1SNC, Leu7–Ser141) by removing the pdTp molecule. Fifty initial structures were repeatedly generated by heating to 310 K in 0.2 ps and equilibrated at 300 K for 1.2 ps. The end-to-end distance between N- and C-termini was increased from 4.1 to 48.0 nm with 0.1 nm increments per 2-ps step at 300 K (temperature controlled every 5 ps). Force was derived as the derivative of total energy (sum of kinetic energy and potential energy) versus the end-to-end distance, and raw

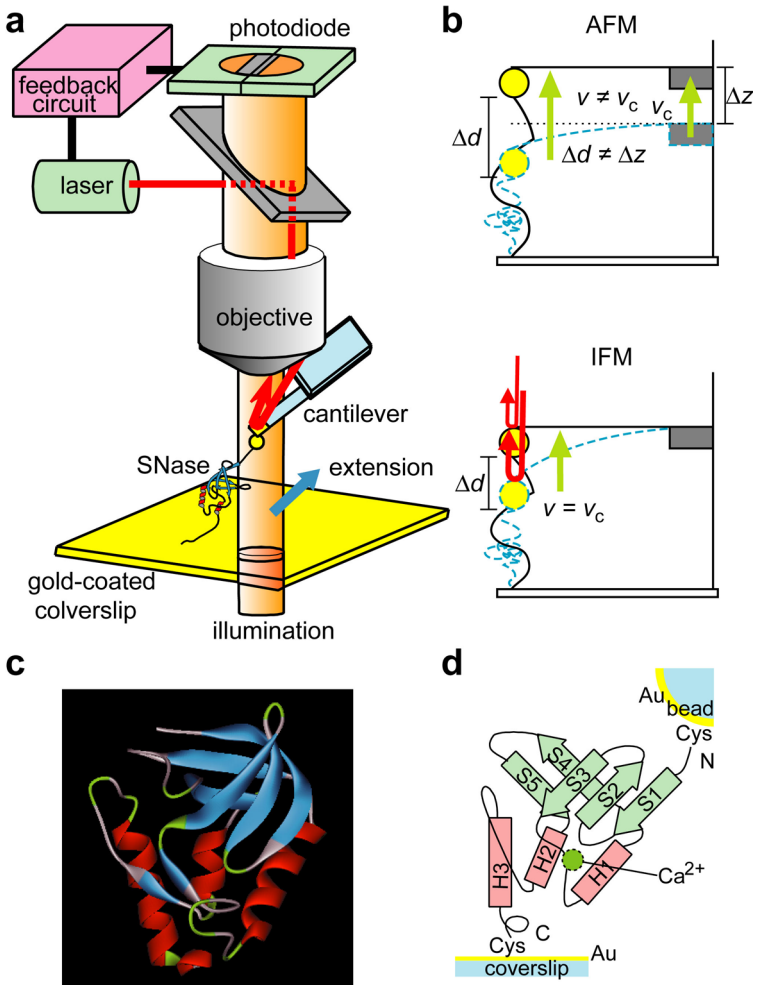

Figure 1 Single-molecule force microscopy and protein structure. (a) IFM[31–33]. The position of the cantilever tip was controlled using laser-radiation pressure. Feedback was applied such that the total forces exerted on the cantilever were equilibrated (see Supplementary Information, Fig. S1). (b) The difference between mechanical stretching using AFM and IFM. Before (broken blue lines) and after (solid black lines) stretching experiments. (top) In an AFM experiment, displacement (Δz) and velocity (Δv_c) of the cantilever were not equal to those of the terminus of the protein molecule (Δd and Δv). (bottom) In an IFM experiment, the terminus of the protein molecule was stretched directly by pulling up the cantilever tip at a constant rate v_c. (c) The SNase crystallographic structure (Protein Data Bank identifier (PDB ID): 1STN) lacked disordered regions (Ala1–Lys5 and Glu142–Gln149)[40]. The red, light-blue and green ribbons represent α-helices, β-structures and turns, respectively. (d) Scheme for SNase unfolding. Secondary-structure notations: S1–S5 (β-structure) and H1–H3 (α-helix). SNase is composed of two domains: the amino (N)-terminal β-barrel domain consisting of S1–S5 and H1, and the C-terminal α-domain consisting of H2 and H3. The Ca^{2+} binding site (dotted circle) consists of two turns between S1 and S2 and between S3 and H1.

traces of force were digitally filtered (Hanning, half width 1.0 nm).

Cluster analysis

The square of the Euclidean distance between clusters, which was composed of single data of a force-extension curve in the first calculation, was calculated by summing the difference between two data (force amplitude) at the same data point (extension) in all combinations. A pair of clusters a and b with the smallest distance was clustered into a new cluster c. The new square of the Euclidean distance between the new cluster c and another cluster x was defined as

$$d_{xc}^{\ 2} = \mathrm{A}d_{xa}^{\ 2} + \mathrm{B}d_{xb}^{\ 2} + \mathrm{C}d_{ab}^{\ 2},$$

where d_{ij} denotes the Euclidean distance between clusters i and j. A, B, and C were defined by the Ward method as

$$\mathrm{A} = (N_x + N_a)/(N_x + N_c)$$
$$\mathrm{B} = (N_x + N_b)/(N_x + N_c)$$
$$\mathrm{C} = -N_x/(N_x + N_c),$$

where N_i denotes the number of components composing cluster i. Iteration of the procedure decreased the number of clusters ($N_{cluster}$) and increased the number of components composing each cluster (N_i). The number of clusters ($N_{cluster}$) was determined as follows. The minimum Euclidean distance d_{min} decreased with the increase in $N_{cluster}$, and was

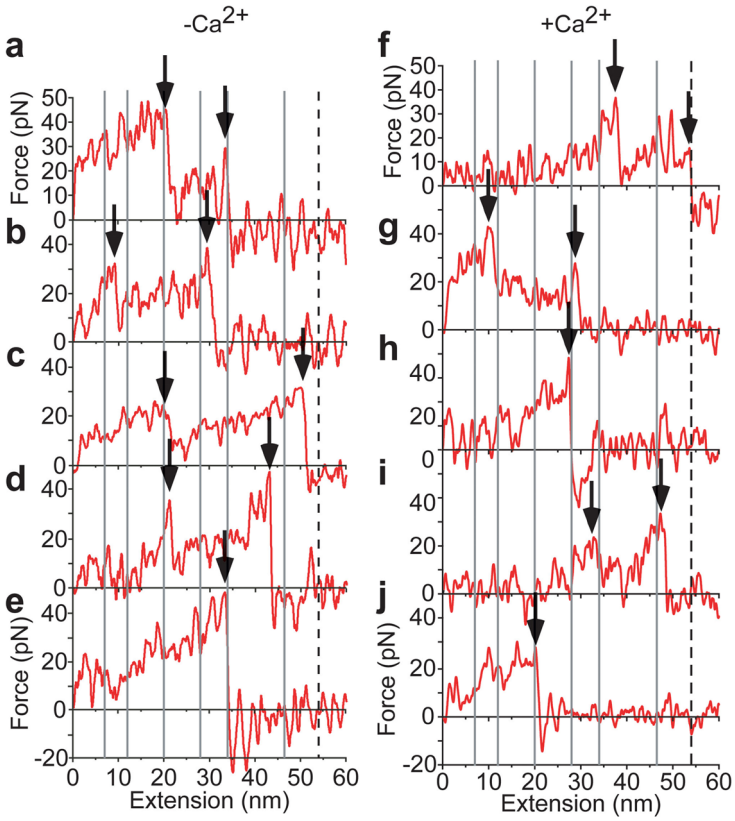

Figure 2 Force-extension relationships for SNase unfolding achieved experimentally. (a–e) Single-molecule force measurements of SNase unfolding, in the absence of Ca^{2+} and (f–k) in the presence of Ca^{2+}. Arrows indicate force peaks. Gray vertical lines represent peak locations in the histogram of the force peaks shown in Fig. 3a and b. Dotted lines represent the estimated maximum length (0.36 nm/residue).

fitted with an exponential decay function, $d_{min} = d_0 \times \exp(-N_{cluster}/N_0)$. Since N_0 was given as 10.2, the number of clusters $N_{cluster}$ was determined as 10.

Force-peak detection

Force peaks accompanied by a decrease within 2 nm in the force amplitude three times larger than the noise amplitude (experiments, Fig. 2; MD simulations, Fig. S3) were used to analyze force-peak distribution in both experiments (Fig. 3a, b) and MD simulations (Fig. 3c, d). Noise amplitude was determined as the standard deviation of the force amplitude in the force-extension relationship after subtracting 3-nm moving averages.

Protein structure

3D structures and secondary structure cartoons were drawn by DS Visualizer (Accelrys).

Trajectory map and energy landscape

The number of residues composing α-helices and β-structures (*Na* and *Nb*, respectively) was counted according to the secondary-structure definition[34] every 1 nm of the end-to-end distance. As the trajectories fluctuated, moving data sets (*Na*, *Nb*) averaged from three were obtained as a function of the end-to-end distance, and were used to depict each trajectory with a B-spline by graph software Origin (OriginLab). The trajectory density map, the energy landscape and deviation of the energy landscape were depicted by Excel (Microsoft). Total energy was used for the energy landscape.

Results

Single-molecule quasi-static microscopy

We used a modified form of inter-/intramolecular-force microscopy (IFM)[31–33], which is a refined version of AFM[35] (Fig. 1a; see Supplementary Information, Fig. S1). Flexible hand-made cantilevers, with ~100-times lower stiffness than conventional cantilevers, were used to achieve sub-piconewton resolutions. Feedback control using laser-radiation pressure was incorporated to control the position of the cantilever tip with sub-nanometer resolutions. Since the root of the cantilever is manipulated in conventional AFM, the displacement (Δd) and velocity (v) of the terminus of the protein molecule were different from those of manipulation (Δz and v_c; Fig. 1b, top). This issue was overcome by manipulating the cantilever tip with the radiation-pressure feedback system. Direct control of the end-to-end molecular distance allowed manipulation to be achieved at conditions close to equilibrium, that is, quasi-statically

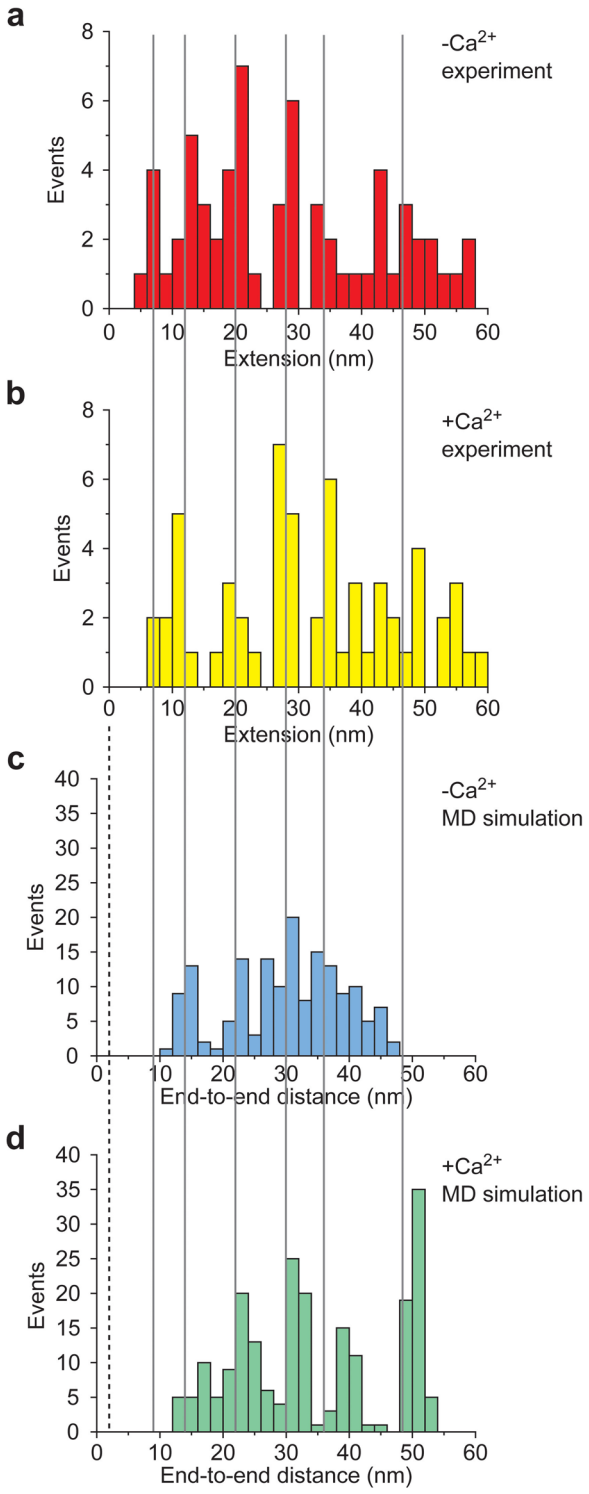

Figure 3 Agreement between experiments and MD simulations in the distribution of force-peak locations. (a and b) Distribution of force-peak locations in force-extension relationships derived experimentally, as shown in Fig. 2a–e and f–j from 42 and 43 force-extension curves in the absence and presence of Ca^{2+}, respectively. (c and d) Distribution derived by MD simulations as shown in Supplementary Information, Fig. S3, a–e and f–j in the absence and presence of Ca^{2+}, respectively, each from 50 force-extension curves. The peaks in (a) and (b) are at extensions of 7, 12, 20, 28, 34 and 46.5 nm.

(Fig. 1b, bottom). SNase is a model protein for folding studies of 149 amino acids (16 kDa)[36–39] with nuclease activity, which is activated by Ca^{2+} (ref. 31). The termini on opposite sides[40] were bridged between a substrate and the tip of the cantilever (Fig. 1c and d) and stretched at a constant rate.

Stochastic appearance of multiple force peaks

Single-molecule force measurements of SNase in the absence and presence of Ca^{2+} produced force-extension curves with one or two peaks at variant extensions (Fig. 2). The force dropped vertically just after its peak, which is a feature of force-extension relations measured using the direct control system of extension. By contrast, force-extension curves from previous single-molecule studies showed force drops with slopes, which decreased the resolution.

Energy-extension curves are obtained as integrals of force-extension curves (Supplementary Information, Fig. S2). The energy increased as the protein ends were pulled, in other words, the end-to-end distance increased; therefore, we calculated energy deviations from trend lines, and compared with the force-extension curves. The force peak corresponded to the local maximum of the energy deviation (see Supplementary Information, Fig. S2). This suggests that force peaks correspond to transition states for unfolding.

A histogram of force-peak locations from all 42 and 43 force-extension curves of SNase in the absence and presence of Ca^{2+}, respectively, showed six peaks (Fig. 3a and b), indicating multiple transition states. That only some of the transition states emerged stochastically in each trial is a distinguishing feature of the single-molecule detection of unfolding processes.

Agreement between experiments and MD simulations

To examine transition states and unfolding pathways, we carried out MD simulations of SNase stretching in the absence and presence of Ca^{2+} (50 simulations for each condition). MD force-extension showed similar profiles to those obtained experimentally (see Supplementary Information, Fig. S3). Force peaks corresponded to breaks in the secondary structure or hydrophobic interaction (Fig. 4). The histogram of force-peak locations from all simulations also showed a distribution similar to the experimental histogram (Fig. 3c and d). It was composed of five peaks, which largely agreed with the experimental results, except for the absence of a peak corresponding to the experimental 7-nm peak. This peak might have been brought about by stretching the terminal region (Ala1–Lys5 and Glu142–Gln149 in the absence of Ca^{2+}; Ala1–Lys6 and Glu142–Gln149 in the presence of Ca^{2+}), which is not included in the simulation, and might correspond to the finding that the carboxy (C)-terminal end region (Trp140–Gln149) is important for structural stability[41]. The best agreement was obtained when the MD axis of the end-to-end distance was shifted to the left by 2.0 nm (broken line in Fig. 3c and d). This might

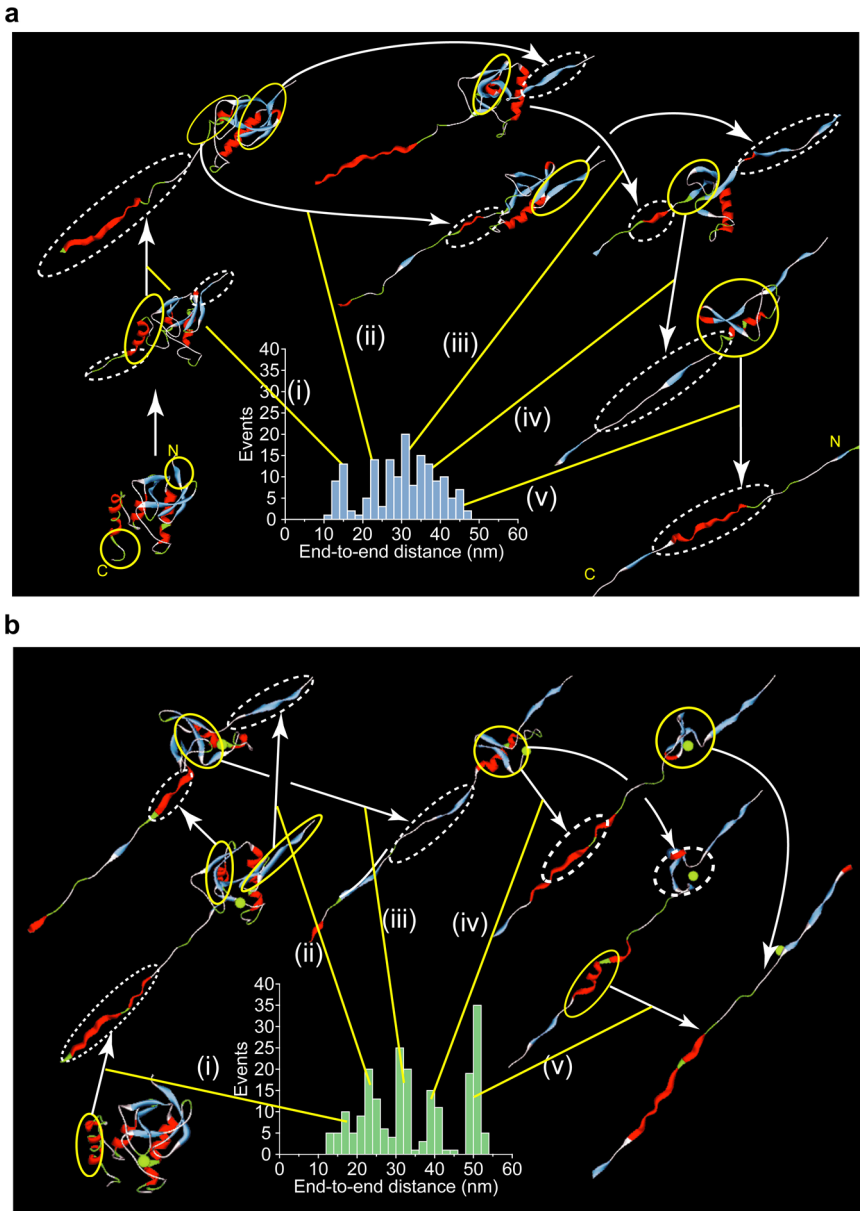

Figure 4 Intermediate structures and structural origins of force peaks obtained by MD simulations. Representative structures of intermediates from MD simulations in the absence (a) and presence (b) of Ca^{2+}. Structures of interest before and after transition are shown as yellow solid and white broken circles, respectively. Histograms (inset) show distributions of force-peak locations in the force-extension relationships (as in Fig. 3c and d). Force peaks corresponded to the breaks of secondary structures or hydrophobic interactions.

have been caused by anchoring the configuration of the protein termini on experimental surfaces. The agreements between experiments and simulations reflected the direct control of the extension and constant-rate stretching.

By contrast, the amplitude of the force and energy differed between experiments and simulations. The force-amplitude distributions were similar, but the experimental and simulation force-peak amplitudes were 26 and 304 pN in the absence of Ca^{2+}, and 34 and 322 pN in the presence of Ca^{2+}, respectively (averaged ratio of 1 : 10.4) (see Supplementary Information, Fig. S4). The energy amplitudes of the mean energy-extension relationship for the experiments and simulations were 442 pNnm (at 34 nm) and 3,645 pN·nm (at 36 nm) in the absence of Ca^{2+}, and 358 pNnm (at 34 nm) and 3,886 pN·nm (at 36 nm) in the presence of Ca^{2+}, respectively (averaged ratio of 1 : 9.4). Force and energy were previously reported to be proportional to the logarithm of the pulling speed[12,42,43]. In the present study, the pulling speeds of the experiments and simulations were 70 nm/s (v_{exp}) and 0.050 nm/ps (v_{sim}), respectively. The logarithm of their ratio, $\log_{10}(v_{sim}/v_{exp})$, was 8.9, which was comparable to the ratio of the force amplitude of 10.4 and the energy amplitude of 9.4. Thus, the difference in amplitude may be explained by the difference

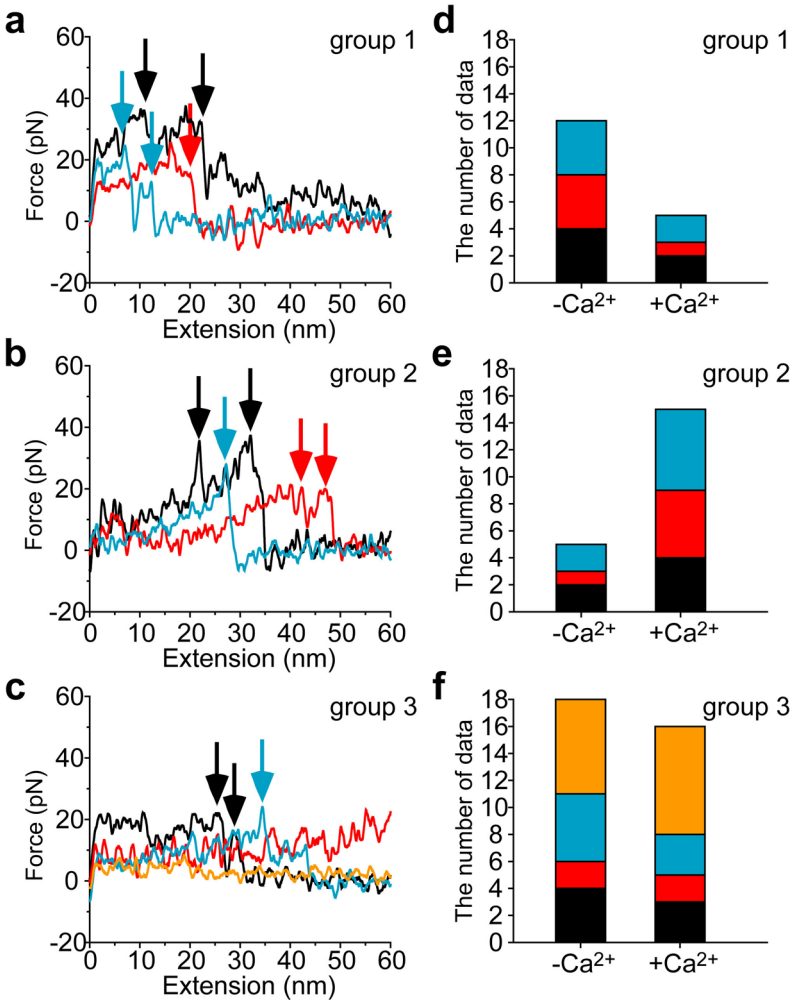

Figure 5 Ca^{2+} effect clarified by cluster analysis. (a–c) The mean force-extension curves of clusters in three groups. All 85 experimental force-extension curves (data) in both the absence and presence of Ca^{2+} were partitioned into 10 clusters by cluster analysis. The data in each cluster were averaged, and 10 mean force-extension curves of the clusters (shown in a–c) were obtained. The mean curves were used for further cluster analysis, and the 10 clusters were partitioned into 3 groups. Arrows indicate force peaks. (d–f) Number of data comprising each cluster in the group 1 (d), 2 (e), and 3 (f), respectively. Data were counted separately in the absence (left) and presence (right) of Ca^{2+}. All 3 clusters in group 1 (black, red, blue) showed data 1.5 times or higher in the absence of Ca^{2+} than in the presence of Ca^{2+}. In contrast, all 3 clusters in group 2 (black, red, blue) showed data 1.5 times or higher in the presence of Ca^{2+} than in the absence of Ca^{2+}.

in the pulling speed, which might be related to the number of possible conformations of the protein per unit time.

Ca^{2+} binding shifted the experimental mean amplitude of the force peak from 26 to 34 pN upon binding of Ca^{2+} (see Supplementary Information, Fig. S4a and b). In MD simulation, the force histogram showed another broad distribution above the major peak only in the presence of Ca^{2+} (see Supplementary Information, Fig. S4c and d). A previous study using circular dichroism (CD) showed that the global protein structure was stabilized in the presence of Ca^{2+} (ref. 44). The increase in force in the present study may reflect the stabilization of the SNase structure by Ca^{2+} binding.

Shift in force-peak locations by Ca^{2+}

To clarify the Ca^{2+} effect, we performed cluster analysis of all 85 experimental force-extension curves in both the absence and presence of Ca^{2+}. The 85 force-extension curves were partitioned into 10 clusters. The mean force-extension curves of the 10 clusters were further divided into 3 groups (Fig. 5). The mean force-extension curves of the clusters in the groups 1 and 2 showed distinct force peaks, while those in group 3 did not. The force-peak amplitude of the mean force-extension curves in groups 1 and 2 was 5 times larger than that of the noise amplitude (Fig. 5a and b; see Methods section).

All force peaks in group 1 were located under 23 nm (Fig. 5a), and all force peaks in group 2 were located at 21.9 nm and over 23 nm (Fig. 5b). The majority of data in group 1 were in the absence of Ca^{2+} (Fig. 5d). In contrast, the majority in group 2 were in the presence of Ca^{2+} (Fig. 5e). Group 3 (Fig. 5c) were composed of data in both the absence and presence of Ca^{2+} comparably (Fig. 5f). This implies that

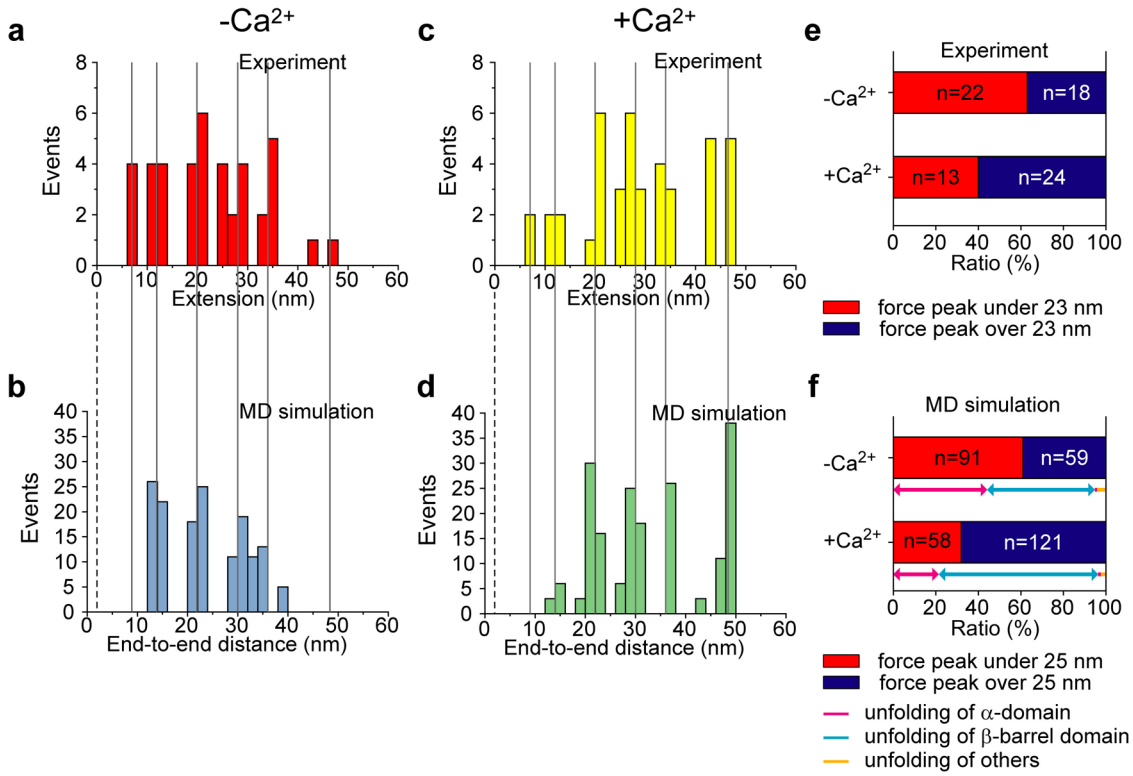

Figure 6 Shift in force-peak distribution by Ca^{2+} binding. (a–d) Distribution of force-peak locations in mean force-extension curves by cluster analysis (Fig. 5a–c) derived experimentally (a and c) and by MD simulation (b and d) in the absence (a and b) and presence (c and d) of Ca^{2+}. The number of force peaks was weighted by the number of data comprising the clusters. Gray vertical lines represent peak locations in the distribution of force peaks without cluster analysis shown in Fig. 3a and b. (e and f) The ratio of the number of force peaks in the former and latter half of extension. The boundary distance was 23 nm for experimental data (see text on Fig. 5 results) and 25 nm for MD data (23+2 nm). Magenta, cyan and orange bars indicate that the force peak of interest emerged during unfolding of the α-domain, β-barrel domain, and other events, respectively in MD simulations.

Ca^{2+} binding shifts the region with a high probability of force-peak emergence from the former to the latter half in extension.

Force peaks emerged more frequently in a stabilized domain

The distribution of experimental force-peak locations of the mean force-extension curves by cluster analysis (Fig. 6a and c) showed several peaks in the same locations as those without cluster analysis (Fig. 3a and b). Cluster analysis was also applied to MD force-extension curves. The histogram of MD force-peak locations by cluster analysis (Fig. 6b and d) again showed a distribution similar to that without cluster analysis (Fig. 3c and d) and to the experimental histogram (Fig. 6a and c). Since force peaks correspond to transition states as mentioned above, this result indicates that the detection of multiple transition states in unfolding was confirmed by MD simulation and cluster analysis.

The majority of force peaks emerged in the former half of the extension (less than 23 nm) in the absence of Ca^{2+}, and in the latter half in the presence of Ca^{2+} (Fig. 6e and f). Their ratios were again similar between the experiments and MD simulations. MD simulations showed that force peaks in the former half mainly occurred during unfolding of the α-domain and that almost all in the latter half occurred during unfolding of the β-barrel domain (Fig. 6f). In other words, Ca^{2+} binding increased force peaks during unfolding of the β-barrel domain. Previous NMR studies showed that ligand binding stabilized the β-barrel domain[45,46]. Taken together, the present findings show that force peaks, i.e., transient states, emerge more frequently in stabilized domains.

Discussion

Stochastic heterogeneous pathways and multiple transition states

In the present study, inter-/intramolecular-force microscopy detected multiple transition states and showed that some, but not all, emerged stochastically (Figs. 2 and 3). This stochastic feature agrees well with predictions from the energy landscape theory, which cannot be detected by ensemble-averaging experiments.

These findings were confirmed by MD simulations, which showed multiple transition states and heterogeneous pathways with stochastic appearance (Figs. 4 and 7; see Supplementary Information, Movies S1 to S3 and Supple-

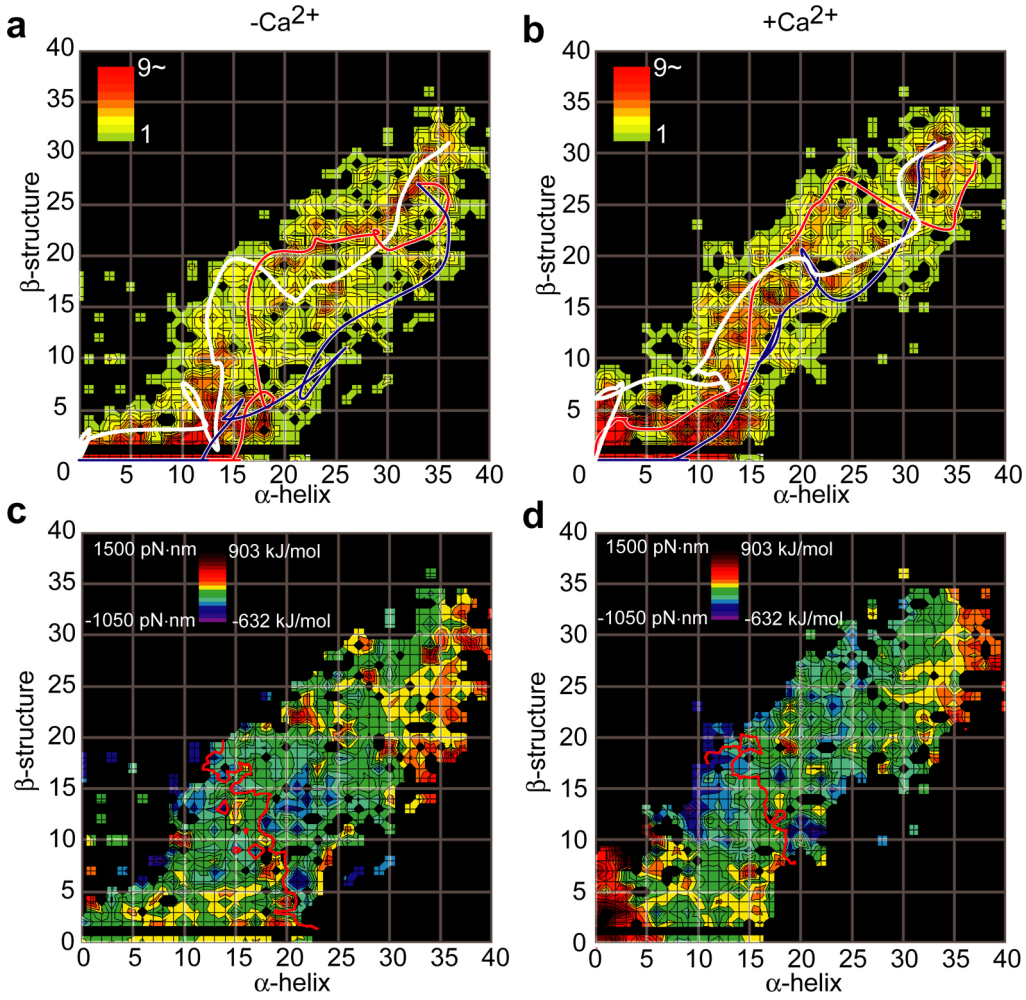

Figure 7 Trajectory density map and 2D 'topographic' energy landscape by MD simulations. (a and b) Trajectory density maps in the absence (a) and presence (b) of Ca^{2+} in αβ number coordinates (see Methods). Three representative trajectories are shown as lines. Trajectories start from the upper right area (N′: 34.6±2.4, 29.5±3.2) in the absence of Ca^{2+}, (N′: 34.5±2.1, 31.6±1.9) in the presence of Ca^{2+}, and end at the coordinate origin (U: 0, 0). (c and d) 2D energy landscapes in the absence (a) and presence (b) of Ca^{2+}. Deviation of the energy landscape from the fitted plane is depicted (see Supplementary Information, Fig. S5). Red curve represents the boundary distance of 25 nm (see Fig. 6f).

mentary Discussion). Individual trajectories underwent vigorous fluctuations, and the 2D trajectory density map clearly showed that individual trajectories were not uniquely determined and took highly heterogeneous pathways (Fig. 7). Meanwhile, kinetic studies analyzed folding processes using schemes with multiple intermediates and parallel pathways[24–27], which differed in number depending on each kinetic study. This characteristic might be comparable to the present finding that individual trajectories stochastically take heterogeneous pathways through multiple intermediates in a complicated energy landscape.

Origin of force peaks in MD simulations

MD simulations showed that force peaks corresponded to breaks in the secondary structure or hydrophobic interaction (Fig. 4). Secondary structures were generally disrupted one at a time (see Supplementary Information, Fig. S6), but with great variation (Fig. 7). The termini of regions where contacts existed between side chains were at loop structures and were often near hydrophobic clusters (see Supplementary Information, Fig. S6), indicating that the intermediates were stabilized by interactions among hydrophobic clusters.

2D energy landscapes by MD simulations were composed of many peaks and troughs (Fig. 7c and d). Trajectories tended to escape energetic peaks and to stochastically search for a route of lower energy. When a trajectory crossed an energetic saddle point or peak, a force peak appeared in the extension curve. As the energies dynamically changed due to a large number of protein conformations, the locations did not precisely correspond with the landscape of the mean energy. Meanwhile, energetic troughs near steep peaks showed high densities on the trajectory map; hence, they corresponded to unfolding intermediates.

The β-barrel domain is stabilized by Ca^{2+} binding

MD simulations showed that unfolding of the α-domain

occurred first followed by the β-barrel domain (Fig. 4; see Supplementary Information, Fig. S6 and Supplementary Discussion), which agrees well with previous results by calorimetric[47] and NMR spectroscopic[48] studies. Cluster analysis showed that groups 1 and 2 were characterized by force peaks in the former half in extension (under 23 nm) and in the latter half (Figs. 5 and 6). The estimated end-to-end distance of the stretched α-domain (Met98-Gln149) and the N-terminal disordered region (Ala1-Lys5) was 20.52 nm (57 residues times 0.36 nm/residue). The boundary length of 23 nm was comparable to the sum of this 20.52 nm and the diameter (3 nm) of the remaining folded β-barrel domain; therefore, force peaks below and above 23 nm are regarded as unfolding of the α-domain and the β-barrel domain, respectively. Indeed, MD simulations showed that force peaks in the former half mainly occurred during unfolding of the α-domain and that almost all in the latter half occurred during unfolding of the β-barrel domain (Fig. 6f).

MD simulations showed that Ca^{2+} binding destabilized the α-domain but stabilized the β-barrel domain (Supplementary Information, Fig. S6). Previous NMR studies showed that ligand binding stabilized the β-barrel domain[45,46]. X-ray crystallography showed that the Ca^{2+} binding site is located in the β-barrel domain (Fig. 1). The flexible loops comprising the Ca^{2+} binding sites were structured and stabilized by Ca^{2+} binding[45].

Frequent force-peak emergence in rugged energy landscape

Cluster analysis showed that in the absence of Ca^{2+}, the majority of force peaks emerged in the former half while in the presence of Ca^{2+} they emerged in the latter half (Fig. 6e and f). Taken together with the results of MD simulations, force peaks emerged more frequently during unfolding of the α-domain in the absence of Ca^{2+}, and during unfolding of the β-barrel domain in the presence of Ca^{2+}. In other words, transient states emerged more frequently in stabilized domains.

2D energy landscapes highlighted the ligand binding effects (Fig. 7c and d). The energy landscape in the former half of extension (upper right area in Fig. 7c) in the absence of Ca^{2+} was more rugged than in the latter half. In the presence of Ca^{2+}, in contrast, the latter half (lower left area in Fig. 7d) was more rugged than the former half. This finding indicates that stabilized domains have a more rugged energy landscape, where transient states emerge with a higher probability.

Disruption energy per residue is comparable to thermal energy

Experimental energy values required per residue to disrupt α-helices and β-structures were estimated using the slope of the fitted plane to the 2D energy landscape (see Supplementary Information, Fig. S5) and the ratio of the energy amplitude of experiments and simulations (1:9.4). The values per residue to disrupt α-helices and β-structures were 11.0 and 3.6 pNnm or 2.7 and 0.9 k_BT (where k_B is the Boltzmann constant and T is the absolute temperature), respectively (corresponding to 104 and 34 pNnm/residue in MD simulations, respectively; see Supplementary Information, Fig. S5). The two-state model fitting a hump-like transition in single-molecule mechanical unfolding of GFP gave an equilibrium free energy of ΔG=22 k_BT for a seven-residue α-helix[21], which agrees well with the present result of 3 k_BT per residue by direct measurement. These figures are comparable to thermal energy (k_BT), which explains the dynamic and stochastic feature of the trajectories.

Conclusion

In the present study, we detected discrete several transition states, only some of which stochastically emerged in each trial. Ligand binding changed the emerging probability. This feature, somewhat analogous to that of quantum phenomena, is a reflection of the huge number of possible conformations of a protein. As demonstrated here, quasi-static force measurement of single molecules, which is performed close to equilibrium, enables researchers to detect free energy changes directly.

Acknowledgments

We thank K. Kuwajima (Okazaki Institute), Y. Hasebe (Sentech), T. Ohmura (Accelrys), Y. Awaji (Terra) and K. Sakata-Sogawa (RIKEN) for support. This work was supported by the Toray Science Foundation (M.T.), by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT; M.T.), and by the Advanced and Innovational Research Program of MEXT (M.T.).

References

1. Kim, P. S. & Baldwin, R. L. Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem. Biophys.* **59**, 631–660 (1990).
2. Karplus, M. & Weaver, D. L. Protein folding dynamics: The diffusion-collision model and experimental data. *Protein Sci.* **3**, 650–668 (1994).
3. Kuwajima, K. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins* **6**, 87–103 (1989).
4. Dill, K. A. Dominant forces in protein folding. *Biochemistry* **29**, 7133–7155 (1990).
5. Ptitsyn, O. B. Molten globule and protein folding. *Adv. Protein Chem.* **47**, 83–229 (1995).
6. Fersht, A. R. Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. USA* **92**, 10869–10873 (1995).
7. Leopold, P. E., Montal, M. & Onuchic, J. N. Protein folding funnels: a kinetic approach to the sequence–structure relationship. *Proc. Natl. Acad. Sci. USA* **89**, 8721–8725 (1992).
8. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. Navigating

the folding routes. *Science* **267**, 1619–1620 (1995).
9. Dobson, C. M. & Karplus, M. The fundamentals of protein folding: Bringing together theory and experiment. *Curr. Opin. Struct. Biol.* **9**, 92–101 (1999).
10. Go, N. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* **12**, 183–210 (1983).
11. Tskhovrebova, L., Trinick, J., Sleep, J. A. & Simmons, R. M. Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature* **387**, 308–312 (1997).
12. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109–1112 (1997).
13. Kellermayer, M. S., Smith, S. B., Granzier, H. L. & Bustamante, C. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science* **276**, 1112–1116 (1997).
14. Lu, H., Isralewitz, B., Krammer, A., Vogel, V. & Schulten, K. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* **75**, 662–671 (1998).
15. Paci, E. & Karplus, M. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *J. Mol. Biol.* **288**, 441–459 (1999).
16. Klimov, D. K. & Thirumalai, D. Stretching single-domain proteins: phase diagram and kinetics of force-induced unfolding. *Proc. Natl. Acad. Sci. USA* **96**, 6166–6170 (1999).
17. Marszalek, P. E., Lu, H., Li, H., Carrion-Vazquez, M., Oberhauser, A. F., Schulten, K. & Fernandez, J. M. Mechanical unfolding intermediates in titin modules. *Nature* **402**, 100–103 (1999).
18. Ohta, S., Alam, M. T., Arakawa, H. & Ikai, A. Origin of mechanical strength of bovine carbonic anhydrase studied by molecular dynamics simulation. *Biophys. J.* **87**, 4007–4020 (2004).
19. Hertadi, R., Gruswitz, F., Silver, L., Koide, A., Koide, S., Arakawa, H. & Ikai, A. Unfolding mechanics of multiple OspA substructures investigated with single molecule force spectroscopy. *J. Mol. Biol*. **333**, 993–1002 (2003).
20. Schwaiger, I., Kardinal, A., Schleicher, M., Noegel, A. A. & Rief, M. A mechanical unfolding intermediate in an actin-crosslinking protein. *Nat. Struct. Mol. Biol*. **11**, 81–85 (2004).
21. Dietz, H. & Rief, M. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc. Natl. Acad. Sci. USA* **101**, 16192–16197 (2004).
22. Cecconi, C., Shank, E. A., Bustamante, C. & Marqusee, S. Direct observation of the three-state folding of a single protein molecule. *Science* **309**, 2057–2060 (2005).
23. Li, L., Huang, H. H., Badilla, C. L. & Fernandez, J. M. Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *J. Mol. Biol.* **345**, 817–826 (2005).
24. Kiefhaber, T. Kinetic traps in lysozyme folding. *Proc. Natl. Acad. Sci. USA* **92**, 9029–9033 (1995).
25. Wallace, L. A. & Matthews, C. R. Sequential vs. parallel protein-folding mechanisms: Experimental tests for complex folding reactions. *Biophys. Chem.* **101–102**, 113–131 (2002).
26. Kamagata, K., Sawano, Y., Tanokura, M. & Kuwajima, K. Multiple parallel-pathway folding of proline-free staphylococcal nuclease. *J. Mol. Biol.* **332**, 1143–1153 (2003).
27. Maki, K., Cheng, H., Dolgikh, D. A., Shastry, M. C. & Roder, H. Early events during folding of wild-type staphylococcal nuclease and a single-tryptophan variant studied by ultrarapid mixing. *J. Mol. Biol.* **338**, 383–400 (2004).
28. Ikura, T., Tsurupa, G. P. & Kuwajima, K. Kinetic folding and cis/trans prolyl isomerization of staphylococcal nuclease. *Biochemistry* **36**, 6529–6538 (1997).
29. Kunkel, T. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**, 488–492 (1985).
30. Maki, K., Ikura, T., Hayano, T., Takahashi, N. & Kuwajima, K. Effects of proline mutations on the folding of staphylococcal nuclease. *Biochemistry* **38**, 2213–2223 (1999).
31. Tokunaga, M., Aoki, T., Hiroshima, M., Kitamura, K. & Yanagida, T. Subpiconewton intermolecular force microscopy. *Biochem. Biophys. Res. Commun.* **231**, 566–569 (1997).
32. Aoki, T., Hiroshima, M., Kitamura, K., Tokunaga, M. & Yanagida, T. Non-contact scanning probe microscopy with sub-piconewton force sensitivity. *Ultramicroscopy* **70**, 45–55 (1997).
33. Fukagawa, A., Hiroshima, M., Sakane, I. & Tokunaga, M. Direct observation of multiple and stochastic transition states by feedback-controlled single-molecule force measurement. *Anal. Sci.* **25**, 5–7 (2009).
34. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
35. Binnig, G., Quate, C. F. & Gerber, C. Atomic force microscope. *Phys. Rev. Lett.* **56**, 930–933 (1986).
36. Tucker, P. W., Hazen, E. E. Jr. & Cotton, F. A. Staphylococcal nuclease reviewed: A prototypic study in contemporary enzymology. IV. The nuclease as a model for protein folding. *Mol. Cell. Biochem.* **23**, 131–141 (1979).
37. Shortle, D. & Meeker, A. K. Residual structure in large fragments of staphylococcal nuclease: Effects of amino acid substitutions. *Biochemistry* **28**, 936–944 (1989).
38. Walkenhorst, W. F., Edwards, J. A., Markley, J. L. & Roder, H. Early formation of a beta hairpin during folding of staphylococcal nuclease H124L as detected by pulsed hydrogen exchange. *Protein Sci.* **11**, 82–91 (2002).
39. Hirano, S., Mihara, K., Yamazaki, Y., Kamikubo, H., Imamoto, Y. & Kataoka, M. Role of C-terminal region of staphylococcal nuclease for foldability, stability, and activity. *Proteins* **49**, 255–265 (2002).
40. Hynes, T. R. & Fox, R. O. The crystal structure of staphylococcal nuclease refined at 1.7 Å resolution. *Proteins* **10**, 92–105 (1991).
41. Yin, J. & Jing, G. Tryptophan 140 is important, but serine 141 is essential for the formation of the integrated conformation of staphylococcal nuclease. *J. Biochem. (Tokyo)* **128**, 113–119 (2000).
42. Evans, E. & Ritchie, K. Strength of a weak bond connecting flexible polymer chains. *Biophys. J.* **76**, 2439–2447 (1999).
43. Klimov, D. K. & Thirumalai, D. Native topology determines force-induced unfolding pathways in globular proteins. *Proc. Natl. Acad. Sci. USA* **97**, 7254–7259 (2000).
44. Sugawara, T., Kuwajima, K. & Sugai, S. Folding of staphylococcal nuclease A studied by equilibrium and kinetic circular dichroism spectra. *Biochemistry* **30**, 2698–2706 (1991).
45. Alexandrescu, A. T., Jahnke, W., Wiltschech, R. & Blommers, M. J. J. Accretion of structure in staphylococcal nuclease: An ^{15}N NMR relaxation study. *J. Mol. Biol.* **260**, 570–587 (1996).
46. Nicholson, L. K., Kay, L. E., Baldisseri, D. M., Arango, J., Young, P. E., Bax, A. & Torchia, D. A. Dynamics of methyl groups in proteins as studied by proton-detected ^{13}C NMR spectroscopy. Application to the leucine residues of staphylococcal nuclease. *Biochemistry* **31**, 5253–5263 (1992).
47. Carra, J. H., Anderson, E. A. & Privalov, P. L. Three-state thermodynamics analysis of the denaturation of staphylococcal nuclease mutants. *Biochemistry* **33**, 10842–10850 (1994).
48. Jacobs, M. D. & Fox, R. O. Staphylococcal nuclease folding intermediate characterized by hydrogen exchange and NMR spectroscopy. *Proc. Natl. Acad. Sci. USA* **91**, 449–453 (1994).