Biophysical Journal
Article

Biophysical Society

# Piecewise All-Atom SMD Simulations Reveal Key Secondary Structures in Luciferase Unfolding Pathway

Pan Zhang,[1] David Wang,[2] Weitao Yang,[1,*] and Piotr E. Marszalek[2,*]
[1]Department of Chemistry and [2]Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina

ABSTRACT Although the folding of single-domain proteins is well characterized theoretically and experimentally, the folding of large multidomain proteins is less well known. Firefly luciferase, a 550 residue three-domain protein, has been commonly used as a substrate to study chaperone reactions and as a model system for the study of folding of long polypeptide chains, including related phenomena such as cotranslational folding. Despite being characterized by various experimental techniques, the atomic-level contributions of various secondary structures of luciferase to its fold's mechanical stability remain unknown. Here, we developed a piecewise approach for all-atom steered molecular dynamics simulations to examine specific secondary structures that resist mechanical unfolding while minimizing the amount of computational resources required by the large water box of standard all-atom steered molecular dynamics simulations. We validated the robustness of this approach with a small NI3C protein and used our approach to elucidate the specific secondary structures that provide the largest contributions to luciferase mechanostability. In doing so, we show that piecewise all-atom steered molecular dynamics simulations can provide novel atomic resolution details regarding mechanostability and can serve as a platform for novel mutagenesis studies as well as a point for comparison with high-resolution force spectroscopy experiments.

SIGNIFICANCE Firefly luciferase is a 550 residue multidomain protein commonly used to examine protein folding in large systems. Although firefly luciferase has been analyzed using force spectroscopy, the atomistic details that underlie luciferase mechanostability remain unknown. Here, we modified a previously established method for conducting piecewise all-atom steered molecular dynamics simulations to allow for the creation of force extension plots that can be directly compared with experiment. We ensured the robustness of this method by first testing on the small protein NI3C, then using the method to identify specific secondary structures in the firefly luciferase mechanical unfolding pathway. The identification of these secondary structures provides possible avenues to test firefly luciferase mechanostability and understand how chaperones might mechanically unfold firefly luciferase.

## INTRODUCTION

Understanding the folding of proteins provides potential avenues for therapeutic development in the treatment of various folding-related diseases, such as Parkinson's (1,2), and in potential methods to predict protein structure from amino acid sequences (3,4). Although the folding of small single-domain proteins has been well characterized theoretically (5–10), computationally (11–19), and experimentally (20–26), the folding of multidomain proteins is less well characterized, although single-molecule force spectroscopy approaches have started to fill this void (27–31). Furthermore, various processes, such as cotranslational folding (29,30,32) and the interaction of protein substrates with chaperones (33,34), suggest complications to the picture of protein folding provided by single-domain protein folding studies (29,32,35).

One model system commonly used to study multidomain protein folding, as well as cotranslational folding and chaperone interactions, is the 550 residue multidomain protein firefly luciferase (10,33,36–39). It is known that firefly luciferase folds cotranslationally (37), displays difficulty refolding after denaturation (38,40), and can experience accelerated refolding because of interactions with chaperones (34,36). Although the mechanisms of firefly luciferase

*Submitted June 11, 2020, and accepted for publication October 14, 2020.*
*Correspondence: weitao.yang@duke.edu or pemar@duke.edu
Pan Zhang and David Wang contributed equally to this work.
Editor: Thomas Perkins.
https://doi.org/10.1016/j.bpj.2020.10.023

folding and refolding have been examined via thermodynamic experiments (36,37) and single-molecule studies using force spectroscopy (27,28,41), the precise atomic-level details of how various substructures in firefly luciferase interact to produce properties such as mechanostability remain unknown because of difficulties in achieving atomic-level resolution using experimental means (27,28). Obtaining these atomic-level details, however, is crucial to understanding various phenomena, such as whether chaperone unfolding of firefly luciferase by force in the "unfoldase" model is a plausible mechanism for how chaperones aid firefly luciferase refolding after denaturation (42,43). To alleviate this problem, we used piecewise all-atom steered molecular dynamics simulations to examine firefly luciferase unfolding under force in all-atom resolution.

All-atom molecular dynamics simulations provide a computational approach to examine the details of protein dynamics in all-atom resolution. Such a description is difficult to obtain via experimental means (12,44–52). To examine whether all-atom details would affect the force extension plots of firefly luciferase, we utilized a piecewise approach for unfolding luciferase and conducted all-atom steered molecular dynamics simulations on firefly luciferase piece by piece in a smaller water box (with around 460,000 atoms) based on a modification of a previously established methodology (11). Normally, the complete mechanical unfolding of large proteins such as luciferase using all-atom steered molecular dynamics simulations would involve setting up a very large water box. The large size of the water box during steered molecular dynamics simulation is a result of the length of the linear firefly luciferase polypeptide after forced unfolding, which drastically increases the amount of water molecules needed to solvate the linear chain. Because the length of an unraveled and stretched luciferase polypeptide is ~200 nm (53), the minimal length of the water box for luciferase SMD calculations is around 200 nm, totaling 1,700,000 atoms. This amount is doubled in the case of GROMACS 2018.2, which requires the distance between the two pull points on the polypeptide chain to not exceed one-half the length of the water box. We note that much larger systems (exceeding 60 million atoms) have been simulated using all-atom representations; however, such simulations typically require significant computational resources (such as supercomputers) (54,55). Our approach reduces the demand on computational resources and can be executed on a simple desktop with a single NVIDIA GTX 980 GPU (NVIDIA, Santa Clara, CA) within a reasonable amount of time (2 months for luciferase). Thus, using a small water box and unfolding a large protein in a piecewise fashion not only reduces the computational cost (see Supporting Materials and Methods) but also can improve sampling statistics by increasing the number of possible replicates for simulations. Furthermore, our modifications to this methodology allowed us to produce force extension plots, thereby making possible the direct comparison of our piecewise steered molecular dynamics simulations with both coarse-grained steered molecular dynamics simulations and experimental force spectroscopy recordings of firefly luciferase unfolding.

We conducted four separate piecewise all-atom steered molecular dynamics simulations of luciferase in a small water box and analyzed the consistency of these simulations with themselves and with respect to coarse-grained simulations that were previously used to elucidate luciferase unfolding under force. In doing so, we reveal that piecewise all-atom steered molecular dynamics simulations provide the same overview of unfolding as coarse-grained simulations based on a native-contact model but also provide self-consistent extra details that cannot be properly modeled by coarse-grained simulations. Furthermore, we identified distinct portions of luciferase secondary structure that appear to provide the greatest contribution to luciferase mechanostability. We then analyzed each of these portions to examine why these specific secondary structures are likely to be important to luciferase unfolding under force.

## METHODS

### Piecewise all-atom simulation system

To validate the piecewise pulling method, a relatively small protein NI3C (Protein Data Bank, PDB: 2QYJ (56)) was studied. Because all-atom steered molecular dynamics simulations that unfold the entire protein are still computationally expensive, we performed piecewise pulling simulations with two different-sized water boxes to examine whether water box size would affect the structural details produced by our all-atom steered molecular dynamics simulations. For one piecewise all-atom steered molecular dynamics simulation, the size of the water box was set as $8.0 \times 8.0 \times 25.0$ nm, and for another piecewise all-atom steered molecular dynamics simulation, the water box was set as $8.0 \times 8.0 \times 50.0$ nm. The simulation parameters were set to be identical to those of the following luciferase steered molecular dynamics simulations.

For steered molecular dynamics simulations of luciferase (PDB: 1BA3 (57)), the size of the water box was set as $10.4 \times 9.2 \times 50.0$ nm, which remains unchanged for all piecewise luciferase steered molecular dynamics simulations. The water box was chosen based on coarse-grained steered molecular dynamics simulations and was chosen because it was large enough to accommodate the first unfolding event (peak) identified by these simulations. A homology model was generated using SWISS-MODEL (58), with the 1BA3 structure as a template, to fill in the two missing amino acids (residues 197–198). All proteins were reoriented such that the N-terminus and C-terminus are aligned with the $z$ axis to pull along one of the cardinal axes and to restrict the size of the water box. All simulations were performed in GROMACS 2018.2 (59), with the CHARMM36 protein force field (60) and the TIP3P water model (61). The systems were minimized until the maximal force was smaller than 1000 kcal/mol and equilibrated for 100 ps before pulling simulations were conducted. An integration time step of 2 fs was applied for both equilibration and pulling simulations. The lengths of the bonds were constrained by the LINCS algorithm (62). The ion concentration was set to be 0.15 M. The NPT ensemble was applied for all pulling simulations. We used a Nosé-Hoover thermostat, with a relaxation time constant of 0.5 ps and a reference temperature of 300 K. To control pressure, we used a Parrinello-Rahman barostat, with a relaxation time constant of 1.0 ps and a compressibility of $4.5 \times 10^{-5}$ $bar^{-1}$. Isotropic pressure coupling was used with a reference pressure of 1.0 bar for coupling. The cutoffs for short-range electrostatic and van der Waals

interaction were set to be 1.4 nm. Particle mesh Ewald (63,64) was employed to calculate long-range electrostatic interactions, and periodic boundary conditions were applied. The Fourier grid dimension was set as 0.12 nm, and cubic interpolation was used for PME. The spring constant and the pulling velocity were set as 6 pN/nm and 1 nm/ns, respectively, based on previous system setups (27,65). Constant velocity pulling was used with a dummy atom anchored by a spring to the $C_\alpha$ atom of the C-terminal residue of luciferase.

Because the performance of all-atom SMD simulations is adversely affected by the size of water box, piecewise pulling was applied to address this issue (see Fig. 1). At the beginning, the $C_\alpha$ atom of the N-terminal residue was chosen as the reference atom, and the $C_\alpha$ atom of the C-terminal residue was chosen as the SMD pulling atom, on which a harmonic pulling potential was applied. For each piecewise pulling simulation, after a portion of luciferase was unfolded, the new SMD pulling atom was changed to one of the $C_\alpha$ atoms close to the folded portion of luciferase. Similarly, we moved the reference atom whenever substantial unfolding occurred near the N-terminus of luciferase. It is worth noting that if a protein's extension becomes larger than half the size of the water box in the pulling direction, the SMD simulation will be automatically terminated in GROMACS 2018.2. We changed the SMD pulling atom upon automatic termination of the simulation by GROMACS 2018.2. After beginning the next piecewise pulling simulation, the previously mechanically unfolded portion of luciferase collapses while the remaining folded portion of luciferase is being mechanically unfolded. This ensures the unfolded luciferase in every piecewise simulation occupies less than half of the water box. Therefore, we believe this approach presents significant precautions such that it is reasonable to suggest that the unfolded portion of luciferase will never interact with its images. For each piecewise simulation, the simulation time was set as 100 ns, but the simulation usually terminated early. After changing the SMD pulling atom, the initial time of the new piecewise simulation was set as the final time from the previous piecewise simulation. This

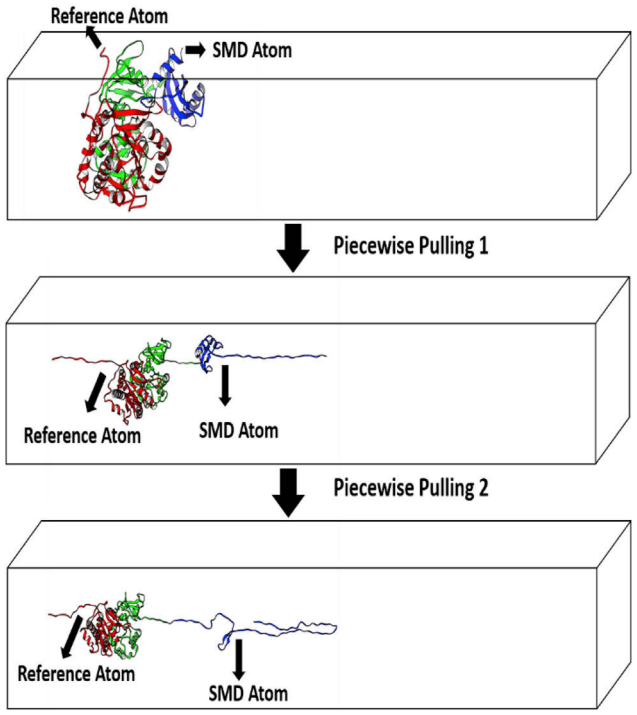


FIGURE 1 Change of SMD dummy atom and reference atom during piecewise pulling simulations (the *black* frame is the simplified sketch of the water box, which is twice as long as the unfolding extension of protein in GROMACS 2018.2). To see this figure in color, go online.

is used to produce the force extension plot, as discussed later. The potential of the $i$-th pulling simulation is represented as

$$U = \frac{1}{2}k[vt - (s_i(t) - d_i)]^2, \tag{1}$$

where $k$ is spring constant, $v$ is pulling velocity, $t$ is simulation time, $s_i(t)$ is the extension between the SMD pulling atom and the reference atom with respect to simulation time, and $d_i$ is the offset to retain force. In the first pulling simulation, $d_1$ was the distance between the initial positions of the SMD pulling atom and the reference atom, which ensured the starting pulling force is zero. Afterwards, $d_i$ was set according to the change of the reference atom and the SMD pulling atom, which is

$$d_i = d_{i-1} - \left|s_i^{\mathrm{i}} - s_{i-1}^{\mathrm{f}}\right| - \left|r_i^{\mathrm{i}} - r_{i-1}^{\mathrm{f}}\right| \,(for\ i > 1), \tag{2}$$

where $s_i^{\mathrm{i}}$ and $r_i^{\mathrm{i}}$ are the initial positions of the SMD pulling atom and the reference atom in the $i$-th pulling simulation, respectively, and $s_{i-1}^{\mathrm{f}}$ and $r_{i-1}^{\mathrm{f}}$ are the final positions of the SMD pulling atom and the reference atom in the $(i-1)$-th pulling simulation, respectively. The simulation time $t$ remained between piecewise pulling simulations so that the force was retained and continuous through all piecewise pulling simulations.

Although this piecewise pulling scheme resembles previous work (11), we did not truncate the protein after each piecewise simulation. This allowed us to perform a continuation simulation after changing the location of the SMD atom because all-atom locations and velocities could be kept the same. In doing so, we eliminated the need for re-equilibration steps and did not add any new solvent molecules to the system. Furthermore, the distance that we set our SMD pulling atom in this piecewise pulling scheme by Eq. 2 allowed us to generate force extension plots. This is significant because it allows for direct comparison with force extension plots generated by continuous coarse-grained simulations and approximates the force extension plots generated by experiment. Because the force is already continuous through the piecewise simulations, the extension is the only quantity that needs to be recovered, which can be obtained by

$$L_i(t) = s_i(t) - r_i(t) - d_i + d_1, \tag{3}$$

where $L_i(t)$ and $r_i(t)$ are the extension and the position of the reference atom with respect to simulation time in the $i$-th pulling simulation, respectively.

## Coarse-grained model generation and simulation

Coarse-grained simulations were performed using GROMACS 2018.2 (59) with a $C_\alpha$ native structure-based force field (66) generated from the SMOG server (67). The coarse-grained model was built after equilibration using the all-atom model. We found the best temperature for coarse-grained simulations using previously established methods (27). A coarse-grained simulation was first performed at 300 K to denature luciferase, then simulations with the denatured coarse-grained model of luciferase were computed for five different temperatures: 130, 140, 150, 160, and 170 K. By analyzing the fraction of native contacts, 150 K was selected as the temperature for coarse-grained steered molecular dynamics simulations because the temperature is near the melting temperature of the coarse-grained luciferase model (i.e., there is a similar probability for the folded and unfolded states (see Supporting Materials and Methods)). An integration time step of 0.5 fs was applied for coarse-grained simulations, and the simulation time for each coarse-grained simulation was set as 150 ns. The cutoff for van der Waals interactions was set as 3.0 nm for the SMOG coarse-grained force field. The spring constant and the pulling velocity were set as 6 pN/nm and 1 nm/ns, respectively, based on previous system setups (11,27). Simulations were repeated 100 times to identify the average behavior of the force extension plots generated by the coarse-grained model of luciferase. A set

of simulations was chosen to represent the variation that we found while conducting coarse-grained simulations.

## RESULTS AND DISCUSSION

### Structural details during piecewise unfolding of NI3C do not vary with water box size

We conducted multiple simulations of the piecewise unfolding of NI3C, a 154-residue synthetic, consensus ankyrin repeats protein (56), with box sizes of 25 and 50 nm in the pulling direction. From our force extension plots (see Fig. 2 *a*) we find that the simulations are very similar to each other, with water box size displaying only small effects on the location of various force peaks. The extension of peak 1 is almost the same between simulations that have different box sizes. This is likely because our different chosen water box sizes can both accommodate the first unfolding event of NI3C. However, the extension of peak 2 for the 25-nm water box size is slightly smaller than that for the 50-nm water box. We attribute this change in extension lengths to a difference in water box size. The 25-nm water box requires a change in pulling site before the chain length of NI3C could be fully extended. Thus, the next force peak occurs slightly before it does in simulations using a larger water box. Extra simulations using a slightly larger 30-nm water box support this explanation (see Supporting Materials and Methods). We expect that this phenomenon may affect all piecewise pulling simulations, unless the water box is large enough to extend the unfolded polypeptide chain to its full length. Thus, peaks may be slightly heterogeneous in their extension lengths, which may be attributed to the shift in pull sites.

Although the extension lengths of the various peaks differ slightly between our 25- and 50-nm box sizes, we found that the structural information produced by the piecewise unfolding of NI3C using different box lengths remains the same. For example, helix 9 (from Thr148 to Asn156) and helix 10 (from Asn158 to Gln166; please note that residue numbering in 2QYJ starts with number 13 and ends with 166) unfold at peak 1 for both simulations (see Fig. 2, *b* and *c*). Similarly, the unfolding event for peak 2, consisting of the unfolding of helix 7 (from Thr115 to Gly124) and helix 8 (from His125 to Ala135), is also the same in simulations with different length water boxes (see Fig. 2, *b* and *c*). After the first two peaks, the unfolding process appears to be more variable (see Fig. 2 *a*). We believe that the variable nature of the later peaks is a result of the more stochastic nature of the unfolding of the N-terminal helix pairs rather than a reflection of the effect of different water box sizes. This is evident in the heterogeneity found within multiple simulations using the same water box size (see Fig. 2 *a*).

The similar locations of force extension peaks and the identical structural details for the force extension peaks across different water box lengths lead us to conclude that piecewise pulling is a reliable method to uncover the force unfolding behavior of large proteins in all-atom resolution.

### Coarse-grained simulations provide a standard three peaks view of atomic force spectroscopy data

From our 100 coarse-grained SMD simulations of luciferase, we found that overall coarse-grained simulations followed previously established single-molecule atomic force spectroscopy (AFM) data (27). Like previously established AFM data, our coarse-grained SMD force extension plots of luciferase also separate into three distinct peaks (see Fig. 3 *a*). After examining the coarse-grained simulations of luciferase, we concluded that these peaks correspond to the sequential unfolding of the three domains of luciferase (27). On a closer inspection, the first unfolding peak of coarse-grained luciferase does display multiple small peaks. We believe, however, that this is a simulation artifact and demonstrate that the number of small peaks in the first peak varies according to simulation parameters (see Supporting Materials and Methods) Although the second peak, after averaging over all 100 curves, displays essentially one broad force peak (Fig. 3 *a*), the details of each individual coarse-grained simulation around the extension

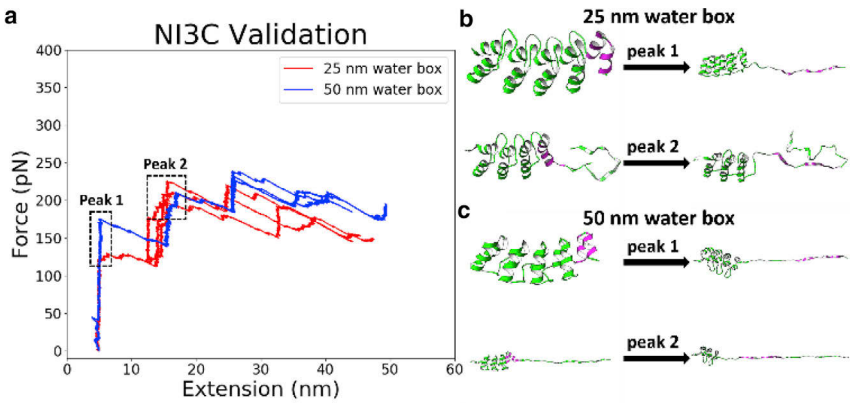


FIGURE 2 NI3C validation using different box sizes shows that structural details remain unchanged by box size. (*a*) Force extension curves from multiple piecewise pulling simulations using two different box sizes are shown. (*b*) Structural details of NI3C unfolding using a 25-nm water box are shown. (*c*) Structural details of NI3C unfolding using a 50-nm water box are shown. To see this figure in color, go online.

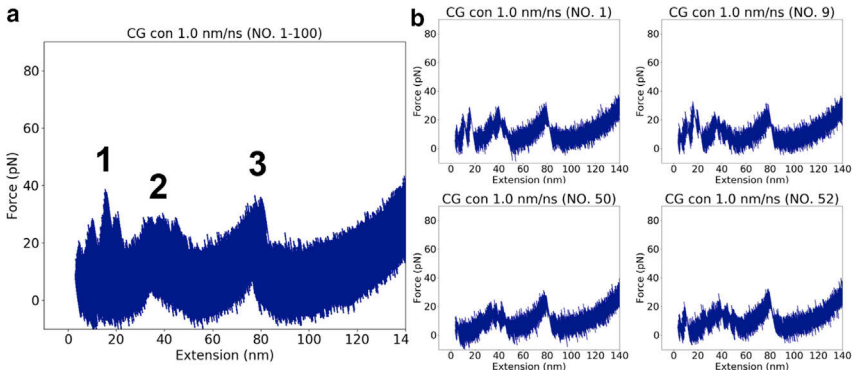


FIGURE 3 Force extension plots from coarse-grained simulations. (*a*) The superpositions of 100 simulations and (*b*) four randomly chosen simulations are shown. To see this figure in color, go online.

corresponding to the second peak differ significantly (see Fig. 3 *b*). In some simulations, the second peak of coarse-grained luciferase only displays a single peak. However, in other coarse-grained simulations, the second peak separates into multiple peaks, suggestive of multiple unfolding intermediate states captured during the unfolding of the middle domain of luciferase.

Thus, we conclude that coarse-grained simulations provide force extension curves that are generally consistent with AFM data. Given the reduced nature of our coarse-grained model and the variety of features found in our coarse-grained simulations, however, we further suggest that these coarse-grained simulations hint at more subtle unfolding events that have not yet been resolved by AFM but may be characterized using all-atom SMD simulations.

## Piecewise all-atom SMD simulations are reproducible and provide more detail than coarse-grained simulations

We note that it is possible to execute hundreds of all-atom SMD simulations while also varying the stretching speed to determine, for example, rupture force distributions and their dependence on loading rate (dynamic force spectra) (49,68) using supercomputing resources. However, in the absence of such resources, we attempted to determine the minimal number of independent all-atom simulations of luciferase that would qualitatively produce representative force extension data. As shown in the Supporting Materials and Methods, we randomly selected 50 sets of four coarse-grained SMD simulations that when aligned appear to be representative of the total 100 replicas overlaid in Fig. 3 *a*. This suggests that utilizing four all-atom simulation replicates should be enough to observe details that are representative of the overall mechanical unfolding process. Therefore, four separate piecewise all-atom SMD simulations were performed in this work, with each simulation being composed of 10–15 piecewise stretching episodes. The total simulation time for each simulation was between 200 and 300 ns. The total simulation time was calculated by summing the simulation times from all piecewise unfolding episodes. The piecewise all-atom SMD trajectories display the same ordering of domain unfolding as coarse-grained SMD simulations. In both sets of simulations, the C-terminal domain of luciferase unfolds first, followed by the middle domain and then the N-terminal domain. Furthermore, the major three peaks of the resulting force extension curves correspond to the unfolding of residues in those three domains. The consistency between the large peaks in the force extension curves of sequential all-atom SMD simulations and coarse-grained SMD simulations suggests that both systems agree with experimental force extension curves and can reproduce that level of detail (see Figs. 2 *c* and 4).

The combined force extension plots of all four simulations is displayed in Fig. 4. Altogether, six overall peaks are shared between the four simulations. In general, six to eight peaks were found in each individual simulation (see individual force extension plots in the Supporting Materials and Methods). For the first six peaks shared between the

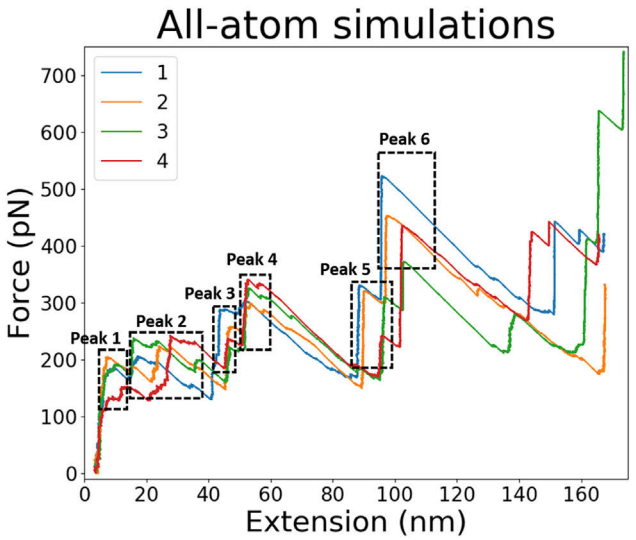


FIGURE 4 The superposition of force extension plots from four all-atom piecewise simulations. To see this figure in color, go online.

four simulations, the extension distance for each peak is similar. The difference in the magnitude of the force peaks between the four independent simulations is expected because it captures the stochastic nature of the mechanically driven unfolding process (68). Furthermore, the force extension curve after peak 6 is somewhat more variable across various simulations. This is likely due to the loss of structural integrity within the remaining luciferase fold when most of the protein has been already unraveled. The increased stochasticity of the last peaks may also be attributed in part to the different SMD pulling atoms across various simulations and warrants further study.

To further validate this method, more piecewise simulations on luciferase and other large proteins are warranted. An additional layer of verification could be provided by comparing the results obtained with the piecewise steered molecular dynamics trajectories with results obtained from a single trajectory steered molecular dynamics simulation performed on a supercomputer.

## All-atom simulation details reveal the unfolding of specific secondary structures in luciferase

We examined the unfolding order of secondary structures using the trajectories provided by our piecewise all-atom steered molecular dynamics simulations (see Fig. 5). The four piecewise simulations display a very similar unfolding order in terms of luciferase secondary structure, with small differences in some peaks. Helix 15 (Arg435 to Lys437) was unfolded at the beginning for all simulations, even though it is in the middle domain. This suggests that helix 15 is the weakest secondary structure element in luciferase. After the unfolding of helix 15, the C-terminal domain was unfolded. The unfolding of the C-terminal domain is represented by peak 1 and peak 2. The unfolded secondary structure in peak 1 is the same across all simulations, but the unfolding order in peak 2 varies between each simulation. Peak 3 corresponds to the unfolding of helix 1 (Ala20 to Ala32) in the N-terminal domain. This small peak occurs in all four of our simulations.

Peak 4 and peak 5 represent the unfolding of the middle domain. In the unfolding of the middle domain of luciferase, a large number of secondary structure elements unfolded in peak 4 in random order, but sheet 1 and helix 9 consistently unfold in peak 5 in all simulations. Unlike the unfolding of the C-terminal and middle domains, the unfolding of the N-terminal domain could be separated into one or three peaks and thus appears more stochastic than the unfolding of the other two domains. The order of unfolded secondary structure elements is very different in peak 6, peak 7, and

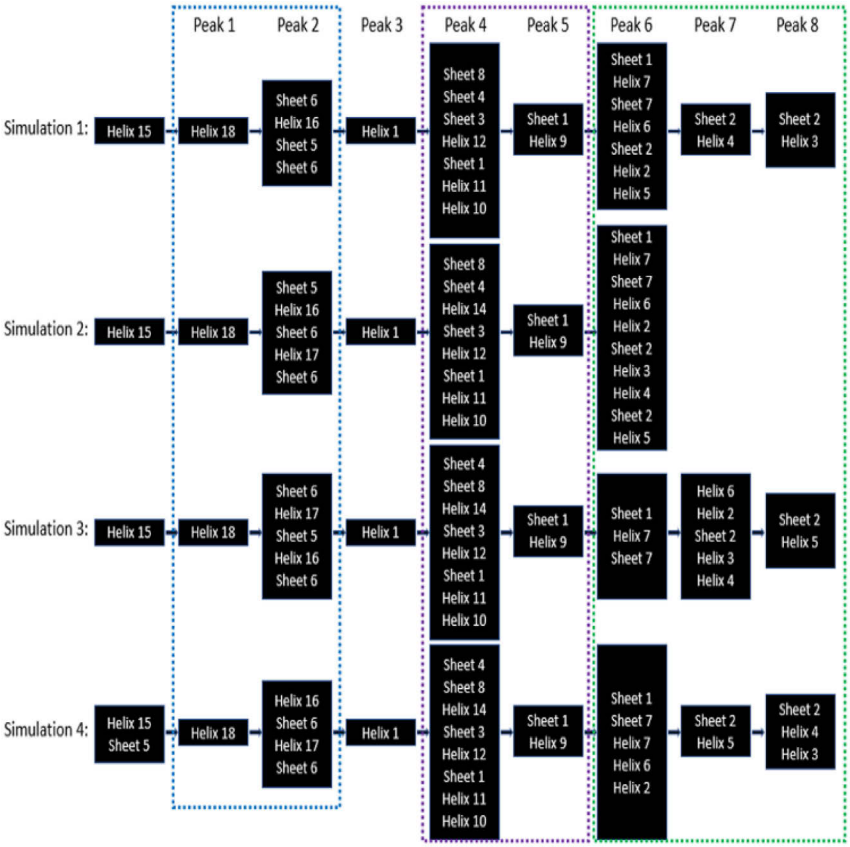


FIGURE 5 Unfolded secondary structures in four all-atom simulations (*blue*: C-terminal domain; *purple*: middle domain; and *green*: N-terminal domain). To see this figure in color, go online.

peak 8 across all four simulations (see Fig. 5). This suggests that the unfolding of structures in the N-terminal domain remains rather variable and that their unfolding order is stochastic. Furthermore, the consistent ordering of secondary structure elements in the unfolding of luciferase with all-atom steered molecular dynamics simulations suggests that various secondary structure elements in luciferase possess different degrees of mechanostability.

The complete unfolding of luciferase in a large water box (~3.4 million atoms) with a single SMD simulation is not performed in this work. Thus, no direct comparison between the complete unfolding of luciferase and piecewise unfolding of luciferase can be drawn. Nevertheless, we believe our results concerning the relative mechanical strengths of secondary structures determined by utilizing piecewise pulling simulations will be informative and similar to the information determined from a complete unfolding of luciferase through a single SMD simulation in a large water box. We believe that this would be the case based on our validation simulations using NI3C with water boxes of different sizes.

## Smaller peaks are dominated by the unfolding of peripheral helices

From the force extension plots generated by our piecewise all-atom steered molecular dynamics simulations (see Fig. 4), we observe that some small peaks occur in a consistent manner for all simulations, namely peak 1, peak 3, and peak 5. We investigated for commonalities in the small peaks of our simulations and found that they all involved the unraveling of small helices (see Fig. 6). Helix 18 (Ala 530 to Ala 540) was the only secondary structure unfolded in peak 1 (see Fig. 6 *a*). It is the secondary structure closest to the C-terminus and is consistently unfolded in the first peak of all our force extension plots. The second small peak is peak 3, where helix 1 (Ala20 to Ala32) was unfolded (see Fig. 6 *b*). This is also a peripheral helix and is located closest to the N-terminus. Helix 1 begins unfolding after the unfolding of the C-terminal domain is finished. Thus, peak 1 and peak 3 suggest that the helices near the end of protein are easily unfolded, correspond to relatively small peaks, and may be precursors to the unfolding of a large domain.

The unfolding of the last peak consists of sheet 1 (Ala234 to Ser237 and Ser282 to Leu284) and helix 9 (from Glu268 to Asp277), which were unfolded sequentially (see Fig. 6 *c*). Because sheet 1 contains 6 strands, the unfolding of sheet 1 found in the small peak is the strand from Ser282 to Leu284, which is separated from the strand at Ala234 to Ser237. In this small peak, not only the helix structure but also the sheet structure is unfolded. However, the removed $\beta$-strand has a relatively weaker interaction compared with the other strands in the $\beta$-sheet because it only consists of three residues compared with four residues in other parts of the $\beta$-sheet. Even though sheet 1 and helix 9 were found in the middle domain, they are the precursor for the unfolding of the N-terminal domain (see Fig. 6 *c*) and become peripheral portions of the overall structure after the middle and C-terminal domains of firefly luciferase are unfolded. Furthermore, although the small peaks usually consist of the unfolding of helices and small $\beta$-strands, we noticed that helices must be larger than a certain size threshold in order for

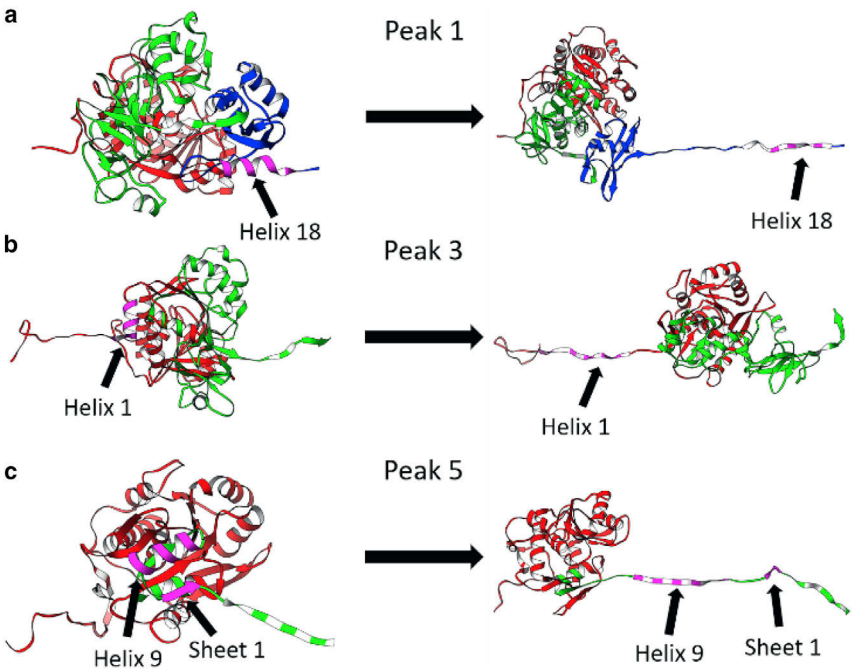


FIGURE 6 Unfolding process of secondary structures for three small peaks. (*a*) Peak 1 is shown; (*b*) peak 3 is shown; and (*c*) peak 5 is shown. The unfolded portions of firefly luciferase are not shown to improve figure clarity. To see this figure in color, go online.

a force peak to be registered. For example, helix 15 was unfolded first in all our piecewise steered molecular dynamics simulations (see Supporting Materials and Methods), but no force peak is displayed. We hypothesize that this is because helix 15 is a relatively short helix with only three residues and the force required for the unfolding of this secondary structure is not large enough to produce a distinct feature preceding force peak 1 and this event is buried in the simulation noise.

However, the presence of these small peaks generated by peripheral helices provides a potential observable for identification with improved spatiotemporal resolution atomic force spectroscopy measurements (20,69). Recent improvements in atomic force microscopy have allowed the observation of multiple intermediates, whereas previous atomic force microscopy measurements could only detect one intermediate (20,69,70). In one specific case, the force unfolding of bacteriorhodopsin using high-spatiotemporal resolution atomic force microscopy detected far more intermediates than previous force spectroscopy unfolding recordings, including the unfolding and refolding of turns in a helix. Furthermore, the improved spatiotemporal resolution force spectroscopy recordings observed intermediates found in previous all-atom steered molecular dynamics simulations (11,20), thereby suggesting that the extra detail provided by all-atom steered molecular dynamics simulations can be observed if the spatiotemporal resolution of force spectroscopy measurements is increased. Thus, we believe that our piecewise steered molecular dynamics simulations could be combined with improved resolution atomic force microscopy measurements to help analyze which structural units correspond to the experimental force extension curves and to compare the accuracy of all-atom molecular dynamics simulations with experiments.

## Unfolding of large sheet structures corresponds to large force peaks for luciferase

We also analyzed the large peaks found in our force extension plots to see whether these peaks provided information on key secondary structures in the luciferase mechanical unfolding pathway. In peak 2, the mechanical unfolding of secondary structures appears to be random (see Fig. 5; Supporting Materials and Methods). This suggests that the secondary structures in the C-terminal domain have equal unfolding propensities and may possess relatively similar mechanical stability. Thus, we believe no obvious anchor point is found in the C-terminal domain.

In contrast to peak 2, which involves the unraveling of the C-terminal domain (see Fig. 5), two secondary structures consistently unfold first in peak 4 across all four piecewise all-atom steered molecular dynamics simulations (see Fig. 5). These structures consist of $\beta$-strands in sheet 4 (Ala422 to Trp424 and Phe430 to Ile432) and $\beta$-strands in sheet 8 (Gly337 to Gly339 and Ala346 to Ile349), which are located in the middle domain of luciferase (see Fig. 7). The strands in both sheets unfold almost simultaneously in peak 4, even though they are far apart from each other, and some secondary structures are found between the two sheets (see Fig. 7). Furthermore, we found that the unfolding of the $\beta$-strands in sheet 4 and sheet 8 correspond to the highest forces produced in peak 4. Thus, we suspect that the $\beta$-strands in sheet 4 and sheet 8 provide major contributions to the mechanical stability of the middle

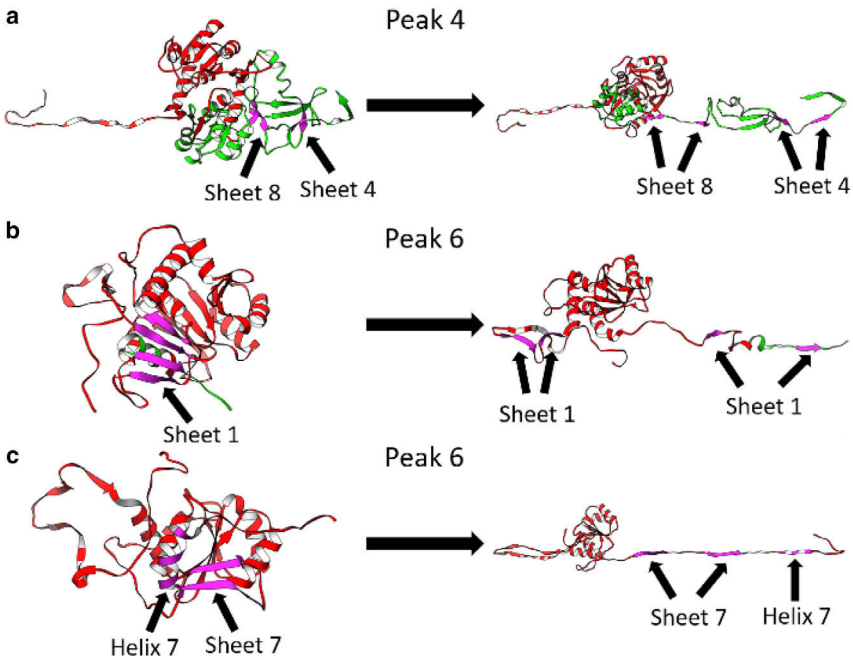


FIGURE 7 Unfolding process of secondary structures for two large peaks. (*a*) The first two secondary structures unfolded in peak 4 are shown. (*b*) The first secondary structures unfolded in peak 6 are shown. (*c*) The second and the third secondary structures unfolded in peak 6 are shown. To see this figure in color, go online.

domain of firefly luciferase. Furthermore, after the unfolding of the $\beta$-strands in sheet 4 and sheet 8, the remaining secondary structures in the middle domain appear to unfold in a random order, thereby furthering the notion that sheet 4 and sheet 8 provide large contributions to the mechanical stability of the middle domain (see Supporting Materials and Methods).

Peak 6 is generally the force peak with the highest forces of all peaks in our force extension curves (see Fig. 4). We found that the first unfolded secondary structure at the beginning of peak 6 is sheet 1, and it consistently unfolds in all our piecewise steered molecular dynamics simulations. The unfolding of sheet 1 comprises the breakage of four $\beta$-strands: one from Arg259 to Leu262, one from Ala234 to Ser237, one from Asn48 to Thr50, and one from Ile38 to Asp42. Sheet 1 is a large sheet structure that is formed by the residues from the N-terminal domain and the middle domain. Although sheet 1 is composed of six strands, one strand (from Glu309 to Ala311) is unfolded in peak 4, and one strand (from Ser282 to Leu284) is unfolded in peak 5. The four strands unfolded in peak 6 strongly interact and simultaneously unfold while producing the highest forces in the force extension plot. Consequently, we believe that sheet 1 is the most consequential structure for the mechanical unfolding pathway of firefly luciferase. After sheet 1, sheet 7 (Ile190 to Asn195 and Gly205 to Pro209) and helix 7 (His210 to Ala220) are the two secondary structures that consistently unfold in peak 6 for all our piecewise all-atom steered molecular dynamics simulations (see Fig. 7 *c*). These two secondary structures belong to the N-terminal domain, and the consistent unfolding of these two structures suggests that they comprise an important portion of the firefly luciferase mechanical unfolding pathway. After the unfolding of sheet 7 and helix 7, the remaining secondary structures in the N-terminal domain unfolded randomly (see Fig. 5; Supporting Materials and Methods) and resulted in multiple peaks that were not entirely consistent between simulations. In general, we determined that sheet structures are found at the initial unfolding regions of the highest force peaks, such as sheet 4 and sheet 8 for peak 4 and sheet 1 for peak 6. We believe that this supports the notion that sheet secondary structures are more stable and require larger force to unfold.

The identification of large sheet structures that consistently unfold at the large force unfolding peaks for firefly luciferase suggests potential mutagenesis experiments that could alter the mechanical unfolding pathway of firefly luciferase. A mutation of the identified $\beta$-strands that correspond to high forces in large force peaks (see Fig. 7) provides a potential experimental observable because destabilizing these sheet structures will likely decrease the maximal forces produced by the various force peaks in the firefly luciferase force extension plots. Furthermore, it would be interesting to examine whether the key sheet structures that were identified as mechanically stable would also be important during heat denaturation experiments that typically are used to produce luciferase-based substrates for the heat-shock-protein-70-assisted refolding. It is possible that destabilization of these sheet structures will lead to the greatest drop in the free energy of folding if the mechanically stable portions of firefly luciferase also correspond to key structures in the thermostability of firefly luciferase.

## CONCLUSIONS

We showed that piecewise all-atom steered molecular dynamics simulations provide a consistent and reproducible picture of firefly luciferase unfolding. These simulations generate more details about specific unfolding events than native-centric coarse-grained models. By examining the force peaks produced by piecewise all-atom steered molecular dynamics simulations, we have identified the secondary structures that play a key role in luciferase mechanostability. We found that the small force peaks before the unfolding of the main domains of firefly luciferase usually involve peripheral helices. Additionally, we found several key sheet structures that consistently correspond to the highest forces in the major force peaks of our force extension curves.

We believe that the identification of these key secondary structures will prove useful in examining the nature of mechanostabilities in firefly luciferase. The identification of these secondary structures may help with the analysis of higher temporal resolution force spectroscopy experiments involving firefly luciferase unfolding. Recent improvements in the temporal resolution of atomic force microscopy techniques allow for the observation of new intermediates that could correspond with the added detail provided by our piecewise all-atom steered molecular dynamics simulations (20,69). Furthermore, the key secondary structures we identified suggest new mutagenesis experiments that could probe how these secondary structures contribute to firefly luciferase mechanostability. The consistency of these piecewise all-atom steered molecular dynamics simulations combined with the added detail they provide vis-à-vis native structure-based coarse-grained models suggest the utility of using piecewise all-atom steered molecular dynamics simulations alongside force spectroscopy experiments to probe the dynamics of proteins.

## SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2020.10.023.

## AUTHOR CONTRIBUTIONS

All authors designed the research and wrote the article. P.Z. carried out the simulations and analyzed the trajectories.

## ACKNOWLEDGMENTS

The authors thank the following funding sources: National Science Foundation (NSF) Grant MCB-1517245, NSF Grant MCB-1817556, and National Institutes of Health Grant R01 GM061870-17. Molecular graphics and analyses were performed with University of California San Francisco Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California San Francisco with support from National Institutes of Health Grant P41-GM103311.

## REFERENCES

1. Dobson, C. M. 2003. Protein folding and misfolding. *Nature.* 426:884–890.
2. Dill, K. A., and J. L. MacCallum. 2012. The protein-folding problem, 50 years on. *Science.* 338:1042–1046.
3. Anfinsen, C. B. 1973. Principles that govern the folding of protein chains. *Science.* 181:223–230.
4. Dill, K. A., S. B. Ozkan, …, T. R. Weikl. 2008. The protein folding problem. *Annu. Rev. Biophys.* 37:289–316.
5. Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600.
6. Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 14:1–63.
7. Bryngelson, J. D., and P. G. Wolynes. 1987. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA.* 84:7524–7528.
8. Plaxco, K. W., K. T. Simons, and D. Baker. 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* 277:985–994.
9. Makarov, D. E., C. A. Keller, …, H. Metiu. 2002. How the folding rate constant of simple, single-domain proteins depends on the number of native contacts. *Proc. Natl. Acad. Sci. USA.* 99:3535–3539.
10. Levy, R. M., W. Dai, …, D. E. Makarov. 2013. How long does it take to equilibrate the unfolded state of a protein? *Protein Sci.* 22:1459–1465.
11. Kappel, C., and H. Grubmüller. 2011. Velocity-dependent mechanical unfolding of bacteriorhodopsin is governed by a dynamic interaction network. *Biophys. J.* 100:1109–1119.
12. Lindorff-Larsen, K., S. Piana, …, D. E. Shaw. 2011. How fast-folding proteins fold. *Science.* 334:517–520.
13. Best, R. B., G. Hummer, and W. A. Eaton. 2013. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. USA.* 110:17874–17879.
14. Piana, S., K. Lindorff-Larsen, and D. E. Shaw. 2013. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. USA.* 110:5915–5920.
15. Voelz, V. A., G. R. Bowman, …, V. S. Pande. 2010. Molecular simulation of *ab initio* protein folding for a millisecond folder NTL9(1-39). *J. Am. Chem. Soc.* 132:1526–1528.
16. Baxa, M. C., K. F. Freed, and T. R. Sosnick. 2009. ψ-constrained simulations of protein folding transition states: implications for calculating. *J. Mol. Biol.* 386:920–928.
17. Stumpe, M. C., and H. Grubmüller. 2009. Urea impedes the hydrophobic collapse of partially unfolded proteins. *Biophys. J.* 96:3744–3752.
18. Echeverria, I., D. E. Makarov, and G. A. Papoian. 2014. Concerted dihedral rotations give rise to internal friction in unfolded proteins. *J. Am. Chem. Soc.* 136:8708–8713.
19. Satija, R., A. Das, and D. E. Makarov. 2017. Transition path times reveal memory effects and anomalous diffusion in the dynamics of protein folding. *J. Chem. Phys.* 147:152707.
20. Yu, H., M. G. Siewny, …, T. T. Perkins. 2017. Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. *Science.* 355:945–950.
21. Marszalek, P. E., H. Lu, …, J. M. Fernandez. 1999. Mechanical unfolding intermediates in titin modules. *Nature.* 402:100–103.
22. Carrion-Vazquez, M., A. F. Oberhauser, …, J. M. Fernandez. 1999. Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl. Acad. Sci. USA.* 96:3694–3699.
23. Gates, Z. P., M. C. Baxa, …, T. R. Sosnick. 2017. Perplexing cooperative folding and stability of a low-sequence complexity, polyproline 2 protein lacking a hydrophobic core. *Proc. Natl. Acad. Sci. USA.* 114:2241–2246.
24. Religa, T. L., J. S. Markson, …, A. R. Fersht. 2005. Solution structure of a protein denatured state and folding intermediate. *Nature.* 437:1053–1056.
25. Englander, S. W., and L. Mayne. 1992. Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. *Annu. Rev. Biophys. Biomol. Struct.* 21:243–265.
26. He, C., C. Hu, …, H. Li. 2015. Direct observation of the reversible two-state unfolding and refolding of an α/β protein by single-molecule atomic force microscopy. *Angew. Chem. Int.Engl.* 54:9921–9925.
27. Scholl, Z. N., W. Yang, and P. E. Marszalek. 2014. Chaperones rescue luciferase folding by separating its domains. *J. Biol. Chem.* 289:28607–28618.
28. Scholl, Z. N., W. Yang, and P. E. Marszalek. 2017. Competing pathways and multiple folding nuclei in a large multidomain protein, luciferase. *Biophys. J.* 112:1829–1840.
29. Bitran, A., W. M. Jacobs, …, E. Shakhnovich. 2020. Cotranslational folding allows misfolding-prone proteins to circumvent deep kinetic traps. *Proc. Natl. Acad. Sci. USA.* 117:1485–1495.
30. Kaiser, C. M., D. H. Goldman, …, C. Bustamante. 2011. The ribosome modulates nascent protein folding. *Science.* 334:1723–1727.
31. Jahn, M., J. Buchner, …, M. Rief. 2016. Folding and assembly of the large molecular machine Hsp90 studied in single-molecule experiments. *Proc. Natl. Acad. Sci. USA.* 113:1232–1237.
32. Liu, K., K. Maciuba, and C. M. Kaiser. 2019. The ribosome cooperates with a chaperone to guide multi-domain protein folding. *Mol. Cell.* 74:310–319.e7.
33. Balchin, D., M. Hayer-Hartl, and F. U. Hartl. 2016. In vivo aspects of protein folding and quality control. *Science.* 353:aac4354.
34. Schröder, H., T. Langer, …, B. Bukau. 1993. DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage. *EMBO J.* 12:4137–4144.
35. Liu, K., X. Chen, and C. M. Kaiser. 2019. Energetic dependencies dictate folding mechanism in a complex protein. *Proc. Natl. Acad. Sci. USA.* 116:25641–25648.
36. Frydman, J., E. Nimmesgern, …, F. U. Hartl. 1994. Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. *Nature.* 370:111–117.
37. Frydman, J., H. Erdjument-Bromage, …, F. U. Hartl. 1999. Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat. Struct. Biol.* 6:697–705.
38. Herbst, R., U. Schäfer, and R. Seckler. 1997. Equilibrium intermediates in the reversible unfolding of firefly (*Photinus pyralis*) luciferase. *J. Biol. Chem.* 272:7099–7105.
39. Buchberger, A., H. Schröder, …, B. Bukau. 1996. Substrate shuttling between the DnaK and GroEL systems indicates a chaperone network promoting protein folding. *J. Mol. Biol.* 261:328–333.
40. Nimmesgern, E., and F. U. Hartl. 1993. ATP-dependent protein refolding activity in reticulocyte lysate. Evidence for the participation of different chaperone components. *FEBS Lett.* 331:25–30.
41. Mashaghi, A., S. Mashaghi, and S. J. Tans. 2014. Misfolding of luciferase at the single-molecule level. *Angew. Chem. Int.Engl.* 53:10390–10393.
42. Clerico, E. M., J. M. Tilitsky, …, L. M. Gierasch. 2015. How hsp70 molecular machines interact with their substrates to mediate diverse physiological functions. *J. Mol. Biol.* 427:1575–1588.

43. Mattoo, R. U. H., and P. Goloubinoff. 2014. Molecular chaperones are nanomachines that catalytically unfold misfolded and alternatively folded proteins. *Cell. Mol. Life Sci.* 71:3311–3325.

44. Zhang, Y., M. Hashemi, …, Y. L. Lyubchenko. 2016. Self-assembly of the full-length amyloid Aβ42 protein in dimers. *Nanoscale.* 8:18928–18937.

45. Best, R. B., and J. Mittal. 2011. Free-energy landscape of the GB1 hairpin in all-atom explicit solvent simulations with different force fields: similarities and differences. *Proteins.* 79:1318–1328.

46. Lovas, S., Y. Zhang, …, Y. L. Lyubchenko. 2013. Molecular mechanism of misfolding and aggregation of Aβ(13-23). *J. Phys. Chem. B.* 117:6175–6186.

47. Bock, L. V., M. H. Kolář, and H. Grubmüller. 2018. Molecular simulations of the ribosome and associated translation factors. *Curr. Opin. Struct. Biol.* 49:27–35.

48. White, K. H., M. Orzechowski, …, K. Visscher. 2011. Mechanical unfolding of the beet western yellow virus -1 frameshift signal. *J. Am. Chem. Soc.* 133:9775–9782.

49. Milles, L. F., K. Schulten, …, R. C. Bernardi. 2018. Molecular mechanism of extreme mechanostability in a pathogen adhesin. *Science.* 359:1527–1533.

50. Williams, M. R., J. C. Tardiff, and S. D. Schwartz. 2018. Mechanism of cardiac tropomyosin transitions on filamentous actin as revealed by all-atom steered molecular dynamics simulations. *J. Phys. Chem. Lett.* 9:3301–3306.

51. Zheng, W. 2014. All-atom and coarse-grained simulations of the forced unfolding pathways of the SNARE complex. *Proteins.* 82:1376–1386.

52. Schoeler, C., K. H. Malinowska, …, M. A. Nash. 2014. Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.* 5:5635.

53. Dietz, H., and M. Rief. 2006. Protein structure by mechanical triangulation. *Proc. Natl. Acad. Sci. USA.* 103:1244–1247.

54. Zhao, G., J. R. Perilla, …, P. Zhang. 2013. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature.* 497:643–646.

55. Perilla, J. R., and K. Schulten. 2017. Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. *Nat. Commun.* 8:15959.

56. Merz, T., S. K. Wetzel, …, P. R. E. Mittl. 2008. Stabilizing ionic interactions in a full-consensus ankyrin repeat protein. *J. Mol. Biol.* 376:232–240.

57. Franks, N. P., A. Jenkins, …, P. Brick. 1998. Structural basis for the inhibition of firefly luciferase by a general anesthetic. *Biophys. J.* 75:2205–2211.

58. Waterhouse, A., M. Bertoni, …, T. Schwede. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46:W296–W303.

59. Abraham, M. J., T. Murtola, …, E. Lindahl. 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* 1–2:19–25.

60. Best, R. B., X. Zhu, …, A. D. Mackerell, Jr. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain $\chi(_1)$ and $\chi(_2)$ dihedral angles. *J. Chem. Theory Comput.* 8:3257–3273.

61. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

62. Hess, B., H. Bekker, …, J. G. E. M. Fraaije. 1997. Lincs: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.

63. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an N•log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

64. Essmann, U., L. Perera, …, L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.

65. Lu, H., and K. Schulten. 2000. The key event in force-induced unfolding of Titin's immunoglobulin domains. *Biophys. J.* 79:51–65.

66. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.

67. Noel, J. K., M. Levi, …, P. C. Whitford. 2016. SMOG 2: a versatile software package for generating structure-based models. *PLoS Comput. Biol.* 12:e1004794.

68. Sedlak, S. M., L. C. Schendel, …, R. C. Bernardi. 2020. Streptavidin/biotin: tethering geometry defines unbinding mechanics. *Sci. Adv.* 6:eaay5999.

69. Edwards, D. T., J. K. Faulk, …, T. T. Perkins. 2017. Force spectroscopy with 9-μs resolution and sub-pN stability by tailoring AFM cantilever geometry. *Biophys. J.* 113:2595–2600.

70. Walder, R., M.-A. LeBlanc, …, T. T. Perkins. 2017. Rapid characterization of a mechanically labile α-helical protein enabled by efficient site-specific bioconjugation. *J. Am. Chem. Soc.* 139:9867–9875.