Fast pressure-jump all-atom simulations and experiments reveal site-specific protein dehydration-folding dynamics

Maxim B. Prigozhin[a,1,2], Yi Zhang[b,1], Klaus Schulten[b,c], Martin Gruebele[a,b,c,3], and Taras V. Pogorelov[a,b,d,e,3]

[a]Department of Chemistry, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [b]Beckman Institute for Advanced Science and Technology, Center for Biophysics and Quantitative Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [c]Department of Physics, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [d]National Center for Supercomputing Applications, University of Illinois at Urbana–Champaign, Urbana, IL 61801; and [e]School of Chemical Sciences, University of Illinois at Urbana–Champaign, Urbana, IL 61801

Edited by José N. Onuchic, Rice University, Houston, TX, and approved February 8, 2019 (received for review August 30, 2018)

As theory and experiment have shown, protein dehydration is a major contributor to protein folding. Dehydration upon folding can be characterized directly by all-atom simulations of fast pressure drops, which create desolvated pockets inside the nascent hydrophobic core. Here, we study pressure-drop refolding of three λ-repressor fragment (λ_{6-85}) mutants computationally and experimentally. The three mutants report on tertiary structure formation via different fluorescent helix–helix contact pairs. All-atom simulations of pressure drops capture refolding and unfolding of all three mutants by a similar mechanism, thus validating the nonperturbative nature of the fluorescent contact probes. Analysis of simulated interprobe distances shows that the α-helix 1–3 pair distance displays a slower characteristic time scale than the 1–2 or 3–2 pair distance. To see whether slow packing of α-helices 1 and 3 is reflected in the rate-limiting folding step, fast pressure-drop relaxation experiments captured refolding on a millisecond time scale. These experiments reveal that refolding monitored by 1–3 contact formation indeed is much slower than when monitored by 1–2 or 3–2 contact formation. Unlike the case of the two-state folder [three–α-helix bundle (α_3D)], whose drying and core formation proceed in concert, λ_{6-85} repeatedly dries and rewets different local tertiary contacts before finally forming a solvent-excluded core, explaining the non–two-state behavior observed during refolding in molecular dynamics simulations. This work demonstrates that proteins can explore desolvated pockets and dry globular states numerous times before reaching the native conformation.

protein solvation dynamics | molecular dynamics simulation | pressure jump

Water is a critical participant in protein folding (1). Roughly half of the folding free energy of proteins is related to reorganization of water molecules (2), and the hydrophobic effect plays an important role (3, 4). Understanding folding mechanisms in the context of a protein's hydration is one of the outstanding challenges in molecular biophysics (5, 6). To this end, previous studies using hydrogen exchange combined with mass spectrometry have shown a concerted desolvation of secondary structural units during the process of folding (7). Both dry and wet globules have been observed on the path to the native state using fluorescence and circular dichroism (8, 9) and NMR (10). In addition, molecular dynamics (MD) simulations have captured the process of protein folding at atomic resolution, including explicit simulation of water (11, 12). However, the validation of MD with experimental data that monitor desolvation has been challenging (13).

Pressure denaturation can be used to study protein–water interactions during folding (14). Hydrostatic pressure causes penetration of water molecules into the hydrophobic but imperfectly packed protein core, thereby reducing the protein's partial molar volume (15). Recently, large pressure jumps with microsecond time resolution have been demonstrated experimentally (16, 17) and simulated using MD (18). Desolvated cavities in the protein core (15) are restored by downward pressure jumps, which facilitate refolding without a change in solvent or thermal motion. In addition, pressure perturbation reduces the effects of protein aggregation, common in temperature-jump (T-jump) experiments performed near the denaturation temperature (16, 17). Finally, pressure-jump (P-jump) experiments can utilize intrinsic fluorophores and fluorescence contact quenchers, such as tryptophan and tyrosine (17).

Desolvation and contact formation can occur in many parts of a protein, yet fluorescence experiments that measure protein dynamics usually rely on a single probe or donor–acceptor pair that reports on a single-order parameter. Thus, deviations from two-state behavior are difficult to capture (19). Previous work has used multiprobe tryptophan fluorescence (20, 21), NMR spectroscopy (22, 23), and IR spectroscopy (24, 25), as well as a combination of multiple approaches (26–28), to address this challenge. To probe three sites, we previously engineered the λ_{6-85} mutants λ_{12}, λ_{32}, and λ_{13}, which contained tryptophan–tyrosine pairs as contact probes between α-helices 1–2, 3–2, and 1–3 (29, 30) (sequences are shown in *SI Appendix*, Fig. S1). Tyrosine quenches tryptophan via Dexter energy transfer when the two α-helices containing these residues contact each other, such that

Significance

When proteins fold, water is squeezed out of the hydrophobic core, leaving small desolvated pockets. Such voids are critical for protein flexibility and function, but not much is known about how and when they form during protein folding. We combine long atomic-level simulations with rapid pressure-drop experiments to see how water gets out of a protein as it folds. For one small protein, we find that dry pockets can form repeatedly between various pairs of protein α-helices before a successful folding event finally occurs. For another small protein, drying and folding are much more concerted. Thus, the formation of void pockets that enhance protein flexibility can occur in a concerted fashion or by coincidence of several local events.

Author contributions: K.S., M.G., and T.V.P. designed research; M.B.P., Y.Z., and T.V.P. performed research; M.B.P., Y.Z., K.S., M.G., and T.V.P. analyzed data; and M.B.P., Y.Z., M.G., and T.V.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]M.B.P. and Y.Z. contributed equally to this work.

[2]Present address: Department of Physics, Stanford University, Stanford, CA 94305.

[3]To whom correspondence may be addressed. Email: mgruebel@illinois.edu or pogorelo@illinois.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1814927116/-/DCSupplemental.

Published online March 5, 2019.

the side chains are less than ~5 Å apart. Thus, we can monitor the formation of specific tertiary contacts between secondary structure elements.

As shown in this work, the enhanced computational capacity of MD now can simulate each multiprobe protein mutant individually. We performed all-atom, explicit-solvent MD on a long time scale (110 μs of total simulation time) to simulate P-jump refolding of λ_{12}, λ_{32}, and λ_{13}. A transient refolding event is captured for each mutant to validate that the fluorescent contact probes used in our experiments do not perturb the folding mechanism. In addition, we performed fast pressure-jump relaxation experiments on the mutants λ_{12}, λ_{32}, and λ_{13}, and compared these experimental results with P-jump MD simulations. The computed time scales of local helix–helix contact formation correlate with the much slower experimental global folding time scales monitored by local helix–helix contacts. Experimental and computational analysis suggests that the rate-limiting step in λ_{6-85} folding is the formation of interhelical contacts between noncontiguous α-helices 1 and 3. In addition, MD simulations reveal that multiple local dehydration and rehydration events of the local core contacts precede the critical folding step. The kinetic barrier that limits the rate of hydrophobic core formation must be even greater in experiments than in the simulations, as evidenced by the 60- to 70-fold slower observed experimental refolding time scales.

Results

The three multiprobe λ_{6-85} constructs have been studied previously using 100-ns-long MD simulations, protein expression level analysis, tryptophan fluorescence spectroscopy, T-jump, and circular dichroism spectroscopy (29). The previous MD trajectories were too short to test whether the dynamics of λ_{6-85} are significantly altered from mutant to mutant. Instead, an MD trajectory of a single mutant (11) was used by us for the analysis of λ_{6-85} folding. Thus, Prigozhin et al. (29) assumed that the folding mechanism of one mutant was representative of all.

Here, an independent MD simulation was performed for each of the three Trp/Tyr contact-pair mutants on the Anton 2 supercomputer (31) (*Methods* and *SI Appendix*, section 2). The initial state of each mutant was prepared by mutating the protein conformation of the last frame of a 1.45-μs high-pressure simulation from a previous study (18) (summarized in *SI Appendix*, section 2). Next, the mutants were subjected to an additional 100 ns of simulation at high pressure (5,000 bar), which was flanked by a 10-ns pressure jump and a 10-ns pressure drop, for total of 120 ns (Fig. 1*B*). Finally, λ_{12} was simulated at 1 bar and 350 K for 25 μs, λ_{32} at 1 bar and 340 K for 25 μs, and λ_{13} at 1 bar and 340 K for 60 μs. The simulation temperatures were chosen to be ~40 K higher than the experimental melting temperatures of each mutant because the thermal stability of λ_{6-85} is significantly higher with the Charmm22* simulation force field (32).

In each simulation, the corresponding mutant transiently visited a conformation close to the crystallographic folded state (33). The folded state was observed at 21.0 μs (λ_{12}), 15.0 μs (λ_{32}), and 51.5 μs (λ_{13}) of simulation time, as highlighted with red vertical lines in Fig. 1*A*. The similarity of the simulated proteins to the folded state was characterized by the fraction of native contacts (Q; the equation for the calculation of Q is provided in *SI Appendix*, section 12) and the root-mean-square deviation (rmsd) of α-carbon atoms calculated with respect to the crystallographic folded state of the three mutants [Protein Data Bank (PDB) ID code 3KZ3 (33)], as shown in Fig. 1*A*. Fig. 1*C* zooms in on the trajectories when the folded state ($Q_{max} \geq 0.8$) was reached by each mutant. In addition, Fig. 1*C* shows overlays of the crystallographic state with the near-folded MD-simulated states for each λ_{6-85} mutant to verify that the folded state was achieved.

MD simulations were analyzed to extract the dynamics captured by the fluorescent probes (Fig. 2). For each mutant, all three distances between the side chains of residues located at the positions used for fluorescent probes and quenchers were

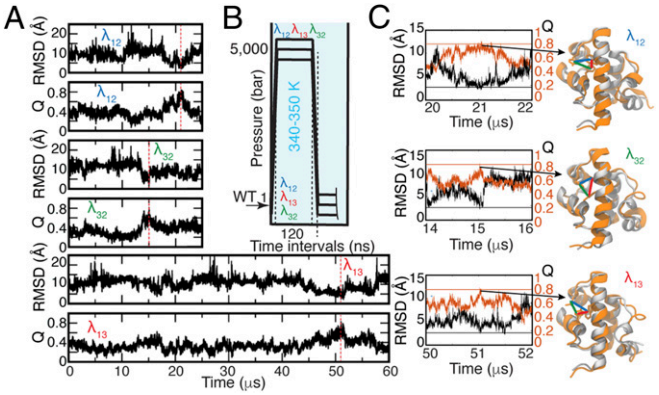

Fig. 1. MD simulations of the three λ_{6-85} multiprobe mutants. (*A*) Time trajectories of Q and rmsd based on α-carbon atoms of the λ_{6-85} crystallographic state [PDB ID code 3KZ3 (33)]. Red dashed vertical lines indicate the time points when the proteins visited the folded state. (*B*) Simulation protocol (details are provided in *SI Appendix*, section 2 and Table S1). The unfolded WT from our previous study (18) was mutated and simulated at 5,000 bar and 340 or 350 K before jumping the pressure down to 1 bar. (*C*) Close-up view of time dynamics of Q (orange) and rmsd (black) near the folded states. Horizontal lines correspond to $Q_{max} = 0.8$ (orange) and rmsd = 2.5 Å (black). The crystallographic state (gray) overlapped with the simulated structures of each mutant at the time when these mutants visited the folded state with $Q_{max} = \sim 0.8$ (21.0 μs for λ_{12}, 15.0 μs for λ_{32}, and 51.5 μs for λ_{13}). Distances between residues used in fluorescence measurements are shown with colored lines (contact 1–2 in blue, contact 3–2 in green, and contact 1–3 in red).

calculated, measuring contacts between helix pairs 1–2, 3–2, and 1–3. Thus, a total of nine time trajectories were generated: three for each of the three mutants (*SI Appendix*, Figs. S2–S4). In the MD simulations, each of the three helix–helix distances was able to reach folded-like values multiple times, while the whole protein was not in the folded conformation. Fig. 2*C* shows structural examples for λ_{12}, with arrows pointing to the corresponding time points on the trajectory (blue = 1–2, green = 3–2, and red = 1–3, while orange is Q, the fraction of total native contacts).

Autocovariance functions of these distances report on the characteristic time scale for the local contact formation/dissolution dynamics between the specified residues, which ultimately lead to helix–helix contacts that result in folding of λ_{6-85} mutants. The nine autocovariance functions shown in Fig. 2*A* suggest that the α-helix 1–2 contact (blue) forms and dissolve fastest and the α-helix contact 1–3 (red) forms and dissolves slowest. All three mutants feature a very similar time scale for each contact pair. Thus, the contact pair time scale is not affected by the mutations. Dynamic fingerprint analysis of the autocovariances (*Methods* and *SI Appendix*, Fig. S5), which uses an inverse Laplace transform to numerically predict time scales that contribute to a given autocovariance function, reveals the following slowest characteristic time scales: 0.35 ± 0.05 μs (contact 1–2), 0.9 ± 0.1 μs (contact 3–2), and 2.5 ± 0.5 μs (contact 1–3). Exponential fits of the autocovariances in Fig. 2*A* produce the same ranking as fingerprinting (0.24–1.6 μs; *SI Appendix*, Fig. S6 and Table S2), with 1–2 being fastest and 1–3 being slowest. Simulated P-jump refolding of λ_{13} is also slowest (Fig. 1 *A* and *C*), hinting that helix contact time scales and overall folding time may be correlated, although a single simulated refolding event is insufficient to draw that conclusion.

Thus, we performed P-jump experiments at 22 °C on the three λ_{6-85} mutants to see if the simulated ordering of helix contact time scales correlates with actual folding times (Fig. 2*B*). The proteins were expressed as described previously (18), pressurized to 1,200 bar (λ_{12}) or 1,500 bar (λ_{32}, λ_{13}), and dropped down to atmospheric pressure within 3 μs (*Methods*) to monitor refolding. The longer dead time than in the study by Dumont et al. (16) precluded resolving the fast "downhill phase." During the P-jump, tryptophan fluorescence lifetime decays were collected

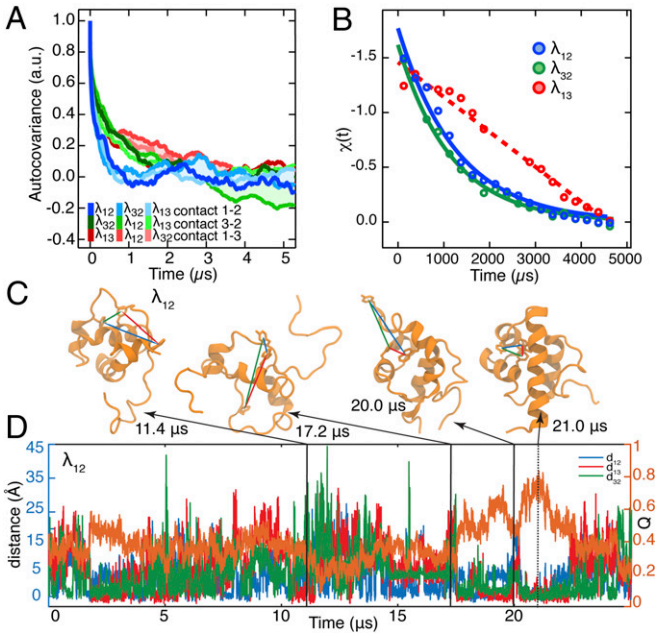

Fig. 2. Dynamics of λ_{6-85} captured by three fluorescent contact pairs. (*A*) Autocovariance analysis of the MD time trajectories of distances between residues used as contact probes in λ_{6-85}. In each simulation, the three residue–residue distances were extracted regardless of whether the mutants had the fluorescent probes in those positions or not. Autocovariance analysis was performed on the resulting time trajectories and shows fast exploration of contacts before folding. (*B*) Experimental fluorescence lifetime relaxation traces collected following a P-jump of $\lambda_{12}(\tau = 1.4 \pm 0.5)$ ms, $\lambda_{32}(\tau = 1.2 \pm 0.3)$ ms, and λ_{13}. The time constant could not be extracted for the slow λ_{13} decay [linear fit (dashed line) is shown instead]. All P-jump time traces are normalized from 0 to 1. (*C*) Representative protein structures with at least one folded-like distance for λ_{12}. Interresidue distances are shown with lines color-coded by the contact pair (λ_{12}, blue; λ_{32}, green; λ_{13}, red). The simulation time points when these structures were observed are indicated. (*D*) Time trajectories of the λ_{12} fraction of native contacts (*Right* axis; orange trace) and distances between the side chains of residues located at the positions used for fluorescent probes (*Left* axis; blue, green, and red traces).

with a photomultiplier tube every 12.5 ns and analyzed as described in *Methods*: $\chi(t)$ measures normalized average lifetime ($\chi = 0$ is shortest, $\chi = 1$ is longest).

P-jump data for λ_{13} in Fig. 2*B* appeared linear and did not fully decay to equilibrium within the 5-ms measurement window because the relaxation time of this mutant is longer than 3 ms. The other two mutants, λ_{12} and λ_{32}, showed P-jump relaxation decays of $\tau_{12} = 1.4 \pm 0.5$ ms and $\tau_{32} = 1.2 \pm 0.3$ ms, respectively (Fig. 2*B* and *SI Appendix*, section 6). Thus, the 1–3 contact yields the slowest experimental folding time, the slowest folding time by MD (single event), and the slowest time scale for local helix contact formation as determined by MD. The 1–2 and 3–2 contacts are always faster than the 1–3 contact, but their time scales are too close to be ranked.

The experimental fluorescence spectrum of tryptophan is sensitive to solvent exposure of the side chain upon pressure denaturation when the dry core is solvated (34) (Fig. 3*A*). The mean fluorescence wavelength of a tryptophan derivative, *N*-acetyl-L-tryptophanamide (NATA), is 371 nm at 1 bar (Fig. 3*B*), but this value is 342 nm for the tryptophan in position 22 of λ_{12} at 1 bar and in the absence of the denaturant guanidine hydrochloride (GuHCl) (Fig. 3*B*). Pressure-induced unfolding of all three λ_{6-85} mutants in the presence of GuHCl increases their mean fluorescence wavelength (Fig. 3 *A* and *B*). (An analysis of pressure-induced equilibrium unfolding of λ_{6-85} mutants is provided in *SI Appendix*, section 7, with *SI Appendix*, Figs. S18–S21 and Table S3 *A* and *B*).

The interdependence of the protein conformational dynamics and the solvation of its hydrophobic residues by water can be visualized in the MD simulations using the solvent-accessible surface area (SASA) of hydrophobic side chains and Q (fraction of total native contacts) calculated for each protein mutant. In Fig. 3 *C–E*, the horizontal blue lines indicate the SASA values of the folded state of each mutant and the horizontal orange lines indicate the Q values of the folded state of each mutant. Although SASA and Q are roughly anticorrelated, it is evident that solvent exclusion of hydrophobic side chains does not necessarily lead to a native-like state. For example, λ_{12} expelled water at the beginning of the simulation (~3 μs; Fig. 3 *C, 1*), but then failed to fold. In this particular case, a transient "dry" globule (8, 35) did not directly lead to the formation of the native state (Fig. 3 *C, 4*). Instead, the hydrophobic core of λ_{12} was first unpacked again and hydrated several more times (Fig. 3 *C, 2*) before water was eventually removed and the protein core folded in a concerted

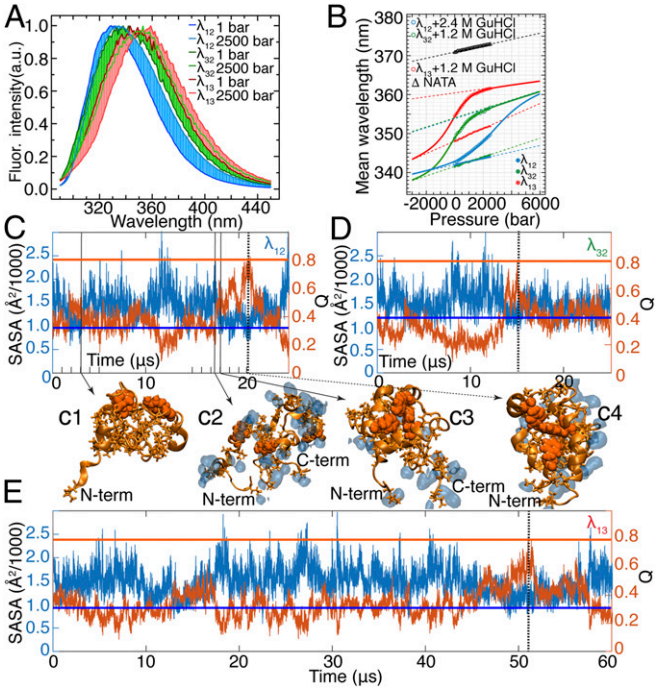

Fig. 3. Water-coupled dynamics in the hydrophobic core of λ_{6-85}. (*A*) Fluorescence (Fluor.) spectra of λ_{6-85} mutants at 1 bar and 2,500 bar (the full datasets are provided in *SI Appendix*, Fig. S21). The point of maximal intensity is normalized to unity to highlight the pressure-induced spectral shift. λ_{12} is in 2.4 M GuHCl; λ_{13} and λ_{32} are in 1.2 M GuHCl. (*B*) Change of mean wavelength of the spectra in *A* as a function of pressure. Connected solid circles are the data collected for each mutant in the absence of GuHCl. Connected open circles represent the data collected at 2.4 M GuHCl for λ_{12} and 1.2 M GuHCl for λ_{13} and λ_{32}. The data for NATA are also shown (△). Solid lines are two-state thermodynamic fits, and dashed lines are the corresponding baselines of folded and unfolded states (the detailed fitting procedure is described in *SI Appendix*, section 7). (*C–E*) SASA of hydrophobic side chains of λ_{6-85} mutants is anticorrelated with Q. Horizontal lines mark SASA (blue) and Q (orange) values in the folded state of each protein mutant. SASA values are 984 Å^2 for λ_{12}, 1,180 Å^2 for λ_{32}, and 908 Å^2 for λ_{13}. Vertical dashed lines (black) mark time points when each protein visited the folded-like conformation (21.0 μs for λ_{12}, 15.0 μs for λ_{32}, and 51.5 μs for λ_{13}). Representative structures of λ_{12} (tryptophan/tyrosine contact pairs are shown as orange van der Waals spheres, hydrophobic side chains are shown as orange sticks) demonstrate that expulsion of water (shown as blue shapes within 2.5 Å from hydrophobic side chains) from the protein core can form a nonnative overpacked (low Q and SASA below folded value) conformation (C1), while the protein core can be hydrated (C2) en route to the native conformation (C3 and C4). C-term, C terminus; N-term, N terminus.

fashion (Fig. 3 *C, 3* and *C, 4*). Thus, local drying is a necessary, but not sufficient, condition for forming a folded structure.

Figs. 1*A* and 3 *C–E* show that the dynamics of λ_{6-85} have low cooperativity, as evidenced by the gradual conformational transitions characterized by such conventional order parameters as Q or rmsd (5) (Fig. 1). This lack of cooperativity has been attributed to downhill folding or short-lived intermediates (36, 37). In Figs. 1 and 3, no clear evidence for a small number of intermediates was observed. Although multiple intermediates with small barriers cannot be ruled out, downhill folding on a rough free energy landscape is the most parsimonious explanation of that limiting case (35).

A seemingly similar model protein, the three–α-helix bundle (α_3D), experiences much more cooperative behavior by the same metrics (11) (Fig. 4). α_3D contains 73 amino acids compared with 80 in λ_{6-85}, and both proteins are α-helical proteins. MD simulations (11) suggest that these proteins have similar folding times of 27 μs and 49 μs, respectively. SASA of only hydrophobic side chains (Fig. 4*B*) calculated using the MD simulations of Shaw and coworkers (11) revealed two-state–like dynamics in the case of α_3D (Fig. 4 *D* and *E*), but less so for λ_{6-85} (Fig. 4 *A* and *B*; details are provided in *SI Appendix*, Fig. S22 and Table S4). Notably, when all side chains are included in the SASA calculation, the folded and unfolded ensembles are indistinguishable for either protein.

We also compared pairwise helix distances for both proteins, using the side chains of Leu10, Ile35, and Ala60 (located roughly in the middle of helices 1, 2, and 3) for α_3D. Autocovariance functions show very different folding dynamics for α_3D and λ_{6-85} (Fig. 4 *C* and *F*): α_3D shows dynamics that are similar between the contact probes, while the interresidue distances of λ_{6-85} show varying time scales. Again, the dynamic fingerprinting technique (38) (*Methods*) revealed that while both proteins have fast relaxations, on the longer time scale, the proteins display radically different behavior. All of the fingerprints of α_3D are clustered near 20 μs. A similar cluster is also observed for λ_{6-85} dynamics at shorter time scales (~4 μs), but contact 1–3 of λ_{6-85} explores the structural space on multiple time scales (red dashed lines, Fig. 4*C*, *Inset*).

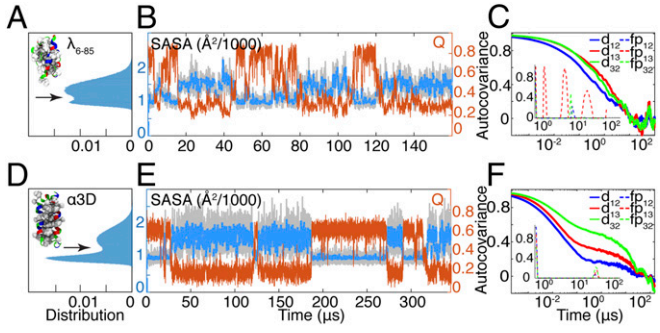

Fig. 4. Comparison of solvation dynamics of λ_{6-85} (*A–C*) and α_3D (*D–F*). Native structures in *A* and *D* are color-coded by residue type (hydrophobic, white; polar, green; basic, blue; acidic, red). Hydrophobic residues used in SASA calculations are shown using the van der Waals representation. (*A* and *D*) SASA serves as a more effective reaction coordinate for the two-state folder, α_3D, compared with λ_{6-85}. α_3D SASA shows a clear saddle point (arrow) separating native and unfolded ensembles. (*B* and *E*) Smoothed (2-μs Gaussian filter) time traces of SASA of the hydrophobic side chains (blue; raw time traces shown in gray) follow the folding dynamics of both the two-state system, α_3D, and the less cooperative λ_{6-85}. Q (orange) is anticorrelated with SASA in the case of α_3D, while there are instances of low SASA and low Q for λ_{6-85}. (*C* and *F*) Autocovariance of distances between side chains of the residues that can be used as probes for interhelical distances (solid lines). Fingerprints (*Inset*, dashed lines) calculated using software developed by Noé et al. (38) report on the characteristic time scales contributing to each autocorrelation function (*SI Appendix*, Fig. S24 and Table S5).

Discussion

MD Simulations Validate Nonperturbative Contact Pairs. The MD simulations presented here suggest that the folding mechanism of λ_{6-85} is not influenced by the mutations that introduced the fluorescent probes for two reasons. First, all of the simulated proteins visited the near-native crystallographic-like state (within 2.0 Å, 2.2 Å, and 2.2 Å rmsd from the crystal structure for λ_{12}, λ_{32}, and λ_{13}, respectively) within their simulation times. The fact that each mutant folded into the same crystallographic-like state during the simulation suggests that the native state is unchanged regardless of the mutations. Second, the autocovariance analysis reported in Fig. 2*A* shows that the characteristic time scales for 1–2, 3–2, and 1–3 contact formation are conserved across the mutants, with 1–2 forming/dissolving most frequently and 1–3 forming/dissolving least frequently. Thus, the collective effect of residues other than the ones in the probe sites dominates the dynamics of the λ_{6-85} conformational search. These results, in combination with the previously reported experimental analysis (29, 30), suggest that the fluorescent contact probe mutations are conservative, thus validating λ_{12}, λ_{32}, and λ_{13} as experimental proxies to study λ_{6-85} folding at the level of the tertiary contact formation.

Such minimal perturbation is by no means universal. Even within λ_{6-85}, we have previously shown that α-helix 4 was not amenable to the incorporation of fluorescent contact probes at the positions that were tested (29). In such a case, it may be necessary to make size-complementary mutations to conserve packing volume, as has been accomplished already for a λ-repressor fragment (39). Validation by MD will enable the use of short-range donor–quencher contact pairs; IR and NMR experiments are, of course, essentially label-free, although they require larger sample concentrations where aggregation may become an issue.

Water Tracks Motions of the Helices in λ_{6-85}. Dynamic fingerprint analysis of the SASA autocovariance for the hydrophobic side chains of α_3D folding simulations (*SI Appendix*, Fig. S23) reveals a time scale comparable to the folding time of the protein, ~27 μs, indicating that water is expelled as global folding proceeds in a two-state manner. In addition, all α_3D distances have similar dynamic content in Fig. 4*F*. In contrast, the fingerprints of λ_{6-85} SASA contain a wide range of fast time-scale components, consistent with the time scale of repeated interhelical contact formation that precedes overall folding. Water can track λ_{6-85} helix packing, both drying and hydrating λ_{6-85} core residues multiple times, before the core is finally dewetted and compacted into the correct local topology characteristic of the native conformation (Fig. 3 *C, 3*). Thus, a local drying transition alone does not necessarily guarantee folding. A number of studies have pointed out the importance of drying the protein core during folding (5, 6, 13).

Similar dynamic patterns are observed in SASA and in the number of water molecules located near the fluorescent probes. In particular, hydration of Trp-51 in our folding simulations of λ_{6-85} mutants reveals this pattern (*SI Appendix*, Figs. S2–S4). Mechanistically, our short-lived folded conformations (Fig. 1*C*) are followed by rapid influx of the water as hinted by the SASA of hydrophobic residues (Fig. 3 *C–E*). Future fast P-jumps with spectral resolution could shed more light on dewetting of at least Trp residues via Stokes shift.

Local and Global Dynamics of λ_{6-85} Are Correlated over Different Time Scales. In the absence of multiple folding/unfolding events, the autocovariance decay curves in Fig. 2*A* are dominated by fast dynamics of the nonnative ensemble. It is this high frequency of local attempts that ultimately leads to global folding, making it critical for the folding mechanism of λ_{6-85}. Local contact formation occurs on a much faster time scale (0.35–2.5 μs from fingerprinting analysis, 0.24–1.6 μs from exponential fits in *SI Appendix*, Table S2) than global folding. As expected, local contact formation occurs near the "folding speed limit" of *ca.*

1 μs (40), which is much faster than activation barrier-limited global refolding.

Nonetheless, the computed speed of local transient contact formation in the nonnative ensemble is correlated with the computed and experimental rates of overall folding. The computed (Fig. 1*A*) and experimental (Fig. 2*B*) overall refolding time scale of λ_{13} relative to λ_{12} and λ_{32} is in agreement with the slower formation/dissolution of the 1–3 contact pair relative to the 1–2 and 3–2 contact pairs (*SI Appendix*, Fig. S5). Trapped state analysis of our folding MD simulations demonstrates the presence of a low-energy trap: a high-Q ensemble with a large 1–3 distance (*SI Appendix*, Fig. S25) in all mutants. Thus, the observed trends suggest that a highly dynamic sampling of the 1–2 helix contact in the nonnative ensemble precedes folding, while the slower contact formation between noncontiguous helices 1 and 3 is part of the rate-limiting step during which at least both of these pieces must fall into place simultaneously.

Comparison with Previous T-Jump Experiments. The refolding dynamics from the pressure-denatured state at 22 °C occur in ≈1 ms for λ_{12} and λ_{32}, and even more slowly for λ_{13}. This observation is in contrast to the ≈50-μs relaxation times observed for these mutants at elevated temperature (40–50 °C higher) by T-jump experiments (29). Using $k(T_1)/k(T_2) = \eta(T_2)/\eta(T_1)\exp[\Delta G^{\dagger}(T_2^{-1} - T_1^{-1})/R]$ to relate rate coefficients k, viscosities η, and activation barriers $\Delta G^{\dagger}$ at two different temperatures, the folding barrier would have to be ≈40 kJ/mol or ≈16 k_BT to account for the reduced rate [$\eta(T_2)/\eta(T_1) \approx 0.45$ for 65 °C vs. 20 °C]. However, natural logarithm of the folding rate [$\ln(k_f)$] of λ_{6-85} is known to have a nonlinear temperature dependence (41), and the pressure- and temperature-denatured states may be structurally different (42), accounting for some of the difference.

Discrepancy Between Experimental and Computational Folding Times. The simulated refolding times of λ_{6-85} after the pressure jump (15–51.5 μs) were 60- to 80-fold faster than the experimental refolding times (1.2 to >3 ms). Some of the possible reasons for this discrepancy are discussed below.

Temperature. The higher simulation temperature (e.g., 340 K in MD of λ_{13} vs. ∼295 K used in experiments) could contribute to faster activated dynamics observed in simulations. Faster activation at a 45 K higher simulation temperature could contribute a factor of ∼10 for a folding barrier of ∼40 kJ/mol. This is also consistent with faster T-jump results at higher temperature, as noted above.

Viscosity. Reduced viscosity of the simulated water is a known artifact of MD simulations that could be partially responsible for fast diffusion and a large prefactor for contact formation and global folding (43). The present simulations were run well above room temperature, and the low viscosity of the TIP3P water model (43) and higher temperature could account for the up to ∼10-fold faster helix contact formation (as low as 0.24 μs for λ_{12}) compared with the "speed limit" set by the P-jump burst phase (2.1 μs in ref. 16).

Sampling. Since only a single transient folding event was observed in each simulation run, the observed folding times are subject to a significant uncertainty caused by insufficient sampling. However, the simulations of the three mutants are consistent (15–50.5 μs), so it is highly unlikely that undersampling can account for the full factor of 60- to 80-fold. Bayesian two-state analysis suggests that it could plausibly account for a factor of 2.

Force field. Some of the difference in computed vs. observed rates may be due to the force field allowing parts of the protein to stick to each other too easily. Such sticking could effectively lower the activation barrier in the simulations relative to experiments. There is evidence for such sticking, forced filed corrections have been implemented to reduce it (44). If sticking is a significant contributor, we predict a somewhat paradoxical outcome: While folding of a single protein in a water box can be sped up in simulations, simulated folding in the presence of other proteins (e.g., a simulated cytoplasm) can be slowed down.

In conclusion, our simulations of three mutants of λ_{6-85} show consistent trends in the interhelix contact formation time scales regardless of mutant identity, validating the minimally perturbative nature of mutants used here to study the mechanism of folding by local contact formation. MD simulations, in agreement with experiments, revealed that fast-folding proteins, such as λ_{6-85}, may show local drying dynamics that are asynchronous with hydrophobic core formation, while bona fide two-state folders, such as α_3D, are highly cooperative and demonstrate close anticorrelation between the hydrophobic core solvation and protein folding. Such asynchronicity does not necessarily imply slower folding and is not related to protein size, as λ_{6-85} and α_3D have very similar size and folding time scales.

Methods

Protein Mutagenesis, Expression, and Purification. Proteins were expressed in *Escherichia coli*, strain BL21 (Stratagene) as described previously (29). The gene of interest was cloned into a pET-15b vector, which contained an N-terminal sequence for six consecutive histidine residues for purification. The cells were transformed with the plasmid containing the gene of interest, and induction was done with 1 mM isopropyl β-D-1-thiogalactopyranoside overnight. The cells were lysed using sonication in the presence of DNase and 1 mM PMSF to reduce protease activity. The protein was then purified using a nickel-nitrilotriacetic acid column (Qiagen), followed by dialysis in 50 mM K_3PO_4 buffer. Mutagenesis was done using a Stratagene site-directed mutagenesis protocol.

Equilibrium Experiments. A Varian Eclipse spectrophotometer was used to acquire fluorescence spectra. A pressurization cell from ISS and a hydrostatic pump (HiP) were used for pressure thermodynamic experiments. A Jasco spectrometer was used to acquire circular dichroism spectra. GuHCl was purchased from Sigma–Aldrich.

Kinetic Experiments. P-jump experiments were performed using a custom-built system, which was described in detail previously (16, 17). Briefly, the sample (∼8 μL at 200 μM concentration) was placed in a dimple machined in a sapphire cube. Pressurization was achieved using an HiP. The pressure was released by rupturing a stainless-steel burst membrane using ∼10-kA current that was generated using a custom capacitor bank. The sample was probed using a mode-locked, frequency-tripled titanium-sapphire laser (center wavelength of ∼285 nm) at an 80-MHz repetition rate. Fluorescence decays were recorded in real time with an oscilloscope using a sampling period of 100 ps. With the 1.2- to 1.5-kbar P-jumps used here, the dead time of the instrument was 3 μs as determined by NATA jumps, about sixfold longer than for 2.5-kbar jumps in a study by Dumont et al. (16).

Data Analysis. Data were analyzed using Igor Pro (Wavemetrics) and MATLAB (MathWorks). Two-state thermodynamic fits were done as described by Prigozhin et al. (29). Fluorescence lifetime data from the P-jump experiments were recorded in real time using a sampling period of 100 ps. Fluorescence decays, $f(t_i)$ at $t_i = 0, 12.5, 25, \cdots$ ns were then parametrized by decomposition into a before-jump profile f_1 and a postjump (5 ms) profile f_2 as $f(t_i) = a_1(t_i)\, f_1 + a_2(t_i)\, f_2$. Signal was reported as $\chi(t) = a_1(t_i)/[a_1(t_i) + a_2(t_i)]$, where $\chi(t)$ is a dimensionless parameter that reports on the change of fluorescence lifetime as a function of time. This two-state assumption was sufficient to fit the data, although it does not imply true two-state folding. The parameter $\chi(t)$ was fitted to a single-exponential function of the form $A_*\exp[-t/\tau]$.

Atomic Models. To make direct comparisons with experiments, three λ-repressor (λ_{6-85}) mutants were constructed at the same pressure-denatured state with the exact amino acid sequences as in the experiments (*SI Appendix*, Fig. S1). The initial structures of the λ_{6-85} mutants were taken from a previous computational model (18) [PDB ID code 3KZ3 (33)], and the following mutations were made to match the exact sequence as shown in *SI Appendix*, Fig. S1*A*: F22W, Q33Y for λ_{12}, F22W, F51Y for λ_{13}, and Q33Y, F51W for λ_{32}. The mutated structures were then solvated using the TIP3P water model (45) and 55 mM of NaCl. Each prepared system contained ∼69,000 atoms.

MD Simulations. MD simulations were performed with the CHARMM22* force field (46) for protein and ions in a $9 \times 9 \times 9$-nm^3 water box with appropriate ionization and periodic boundary conditions in the isothermal-isobaric ensemble at the temperature chosen for each λ_{6-85} mutant. To mimic the P-jump

experiments, the steps summarized in Fig. 1*B* and *SI Appendix*, section 2 were then taken in the simulations.

MD Simulations on NAMD2. Step 2 in the above protocol and the equilibration was performed with NAMD2 (47). The systems to be simulated were minimized for 6,000 steps, followed by 12 ns of equilibration with harmonic constraints on heavy atoms [force constant = 1 kcal/(mol Å^2)]. The constant temperature was controlled by Langevin dynamics, and the constant pressure was regulated by the Nosé–Hoover–Langevin piston method (48). The particle mesh-Ewald method was employed to calculate long-range electrostatic forces (49). The time step was set to 2 fs. Data analysis and figure rendering were done using the VMD computer program (50).

MD Simulations on Anton 2. Refolding simulations for all three λ-repressor mutants were carried out on the Anton 2 platform (31) for a total time of 110 μs ($\lambda_{12} = \lambda_{32} = 25$ μs, $\lambda_{13} = 60$ μs). The multigrator integration method was applied (51). Short-range forces were evaluated at every time step, and long-range electrostatics were calculated every three time steps using the Gaussian split Ewald method (52). The time step was set to 2.5 fs.

Dynamics Fingerprint Analysis. Dynamic fingerprint analysis of autocovariance of MD distance and SASA were performed using software developed by Noé et al. (38). In brief, the method performs a multiexponential fit, which reports on characteristic time scales present in the autocovariance time traces. Noise was assumed to be constant with intensity (SD) of $0.5/\sqrt{time}$. The profiles were optimized (38) for 2×10^7 steps (default) or 5×10^7 steps for d_{13}. Time scales of the fingerprint peaks were estimated using single-Gaussian fits (*SI Appendix*, Table S5). Convergence was checked with likelihood, which is constant after a relatively small number of iterations (*SI Appendix*, Fig. S24).

ACKNOWLEDGMENTS. We thank D. E. Shaw Research (DESR) for providing MD data for λ_{6-85} and α_3D, and Dr. Anna Jean Wirth for assistance with setting up the P-jump. Anton 2 computer time was provided by the Pittsburgh Supercomputing Center (PSC) through Grant R01GM116961 from the NIH. The Anton 2 machine at the PSC was generously made available by DESR. We also thank DESR for providing access to the simulation data. This work was supported by NIH Grants R01GM093318 (to M.G.) and 9P41GM104601 (to K.S.). M.B.P. was supported by the Howard Hughes Medical Institute International Student Research Fellowship and the Helen Hay Whitney Foundation Postdoctoral Fellowship.

1. Bellissent-Funel MC, et al. (2016) Water determines the structure and dynamics of proteins. *Chem Rev* 116:7673–7697.
2. Ebbinghaus S, et al. (2007) An extended dynamical hydration shell around proteins. *Proc Natl Acad Sci USA* 104:20749–20752.
3. Chandler D (2005) Interfaces and the driving force of hydrophobic assembly. *Nature* 437:640–647.
4. Berne BJ, Weeks JD, Zhou R (2009) Dewetting and hydrophobic interaction in physical and biological systems. *Annu Rev Phys Chem* 60:85–103.
5. Cheung MS, García AE, Onuchic JN (2002) Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc Natl Acad Sci USA* 99:685–690.
6. Fernández-Escamilla AM, et al. (2004) Solvation in protein folding analysis: Combination of theoretical and experimental approaches. *Proc Natl Acad Sci USA* 101:2834–2839.
7. Hu W, et al. (2013) Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc Natl Acad Sci USA* 110:7684–7689.
8. Jha SK, Udgaonkar JB (2009) Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc Natl Acad Sci USA* 106:12289–12294.
9. Sarkar SS, Udgaonkar JB, Krishnamoorthy G (2013) Unfolding of a small protein proceeds via dry and wet globules and a solvated transition state. *Biophys J* 105:2392–2402.
10. Fu Y, et al. (2012) Coupled motion in proteins revealed by pressure perturbation. *J Am Chem Soc* 134:8543–8550.
11. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. *Science* 334:517–520.
12. Snow CD, Sorin EJ, Rhee YM, Pande VS (2005) How well can simulation predict protein folding kinetics and thermodynamics? *Annu Rev Biophys Biomol Struct* 34:43–69.
13. Das P, Kapoor D, Halloran KT, Zhou R, Matthews CR (2013) Interplay between drying and stability of a TIM barrel protein: A combined simulation-experimental study. *J Am Chem Soc* 135:1882–1890.
14. Silva JL, Foguel D, Royer CA (2001) Pressure provides new insights into protein folding, dynamics and structure. *Trends Biochem Sci* 26:612–618.
15. Roche J, et al. (2012) Cavities determine the pressure unfolding of proteins. *Proc Natl Acad Sci USA* 109:6945–6950.
16. Dumont C, Emilsson T, Gruebele M (2009) Reaching the protein folding speed limit with large, sub-microsecond pressure jumps. *Nat Methods* 6:515–519.
17. Prigozhin MB, et al. (2013) Misplaced helix slows down ultrafast pressure-jump protein folding. *Proc Natl Acad Sci USA* 110:8087–8092.
18. Liu Y, Prigozhin MB, Schulten K, Gruebele M (2014) Observation of complete pressure-jump protein refolding in molecular dynamics simulation and experiment. *J Am Chem Soc* 136:4265–4272.
19. Liu F, et al. (2009) A one-dimensional free energy surface does not account for two-probe folding kinetics of protein α3D. *J Chem Phys* 130:061191.
20. Sato S, Religa TL, Fersht AR (2006) Phi-analysis of the folding of the B domain of protein A using multiple optical probes. *J Mol Biol* 360:850–864.
21. Royer CA, Mann CJ, Matthews CR (1993) Resolution of the fluorescence equilibrium unfolding profile of trp aporepressor using single tryptophan mutants. *Protein Sci* 2:1844–1852.
22. Alderson TR, Charlier C, Torchia DA, Anfinrud P, Bax A (2017) Monitoring hydrogen exchange during protein folding by fast pressure jump NMR spectroscopy. *J Am Chem Soc* 139:11036–11039.
23. Roche J, et al. (2013) Effect of internal cavities on folding rates and routes revealed by real-time pressure-jump NMR spectroscopy. *J Am Chem Soc* 135:14610–14618.
24. Davis CM, Cooper AK, Dyer RB (2015) Fast helix formation in the B domain of protein A revealed by site-specific infrared probes. *Biochemistry* 54:1758–1766.
25. Ma J, Pazos IM, Zhang W, Culik RM, Gai F (2015) Site-specific infrared probes of proteins. *Annu Rev Phys Chem* 66:357–377.
26. Li P, Oliva FY, Naganathan AN, Muñoz V (2009) Dynamics of one-state downhill protein folding. *Proc Natl Acad Sci USA* 106:103–108.
27. Davis CM, Reddish MJ, Dyer RB (2017) Dual time-resolved temperature-jump fluorescence and infrared spectroscopy for the study of fast protein dynamics. *Spectrochim Acta A Mol Biomol Spectrosc* 178:185–191.
28. Panick G, et al. (1998) Structural characterization of the pressure-denatured state and unfolding/refolding kinetics of staphylococcal nuclease by synchrotron small-angle X-ray scattering and Fourier-transform infrared spectroscopy. *J Mol Biol* 275:389–402.
29. Prigozhin MB, Chao S-H, Sukenik S, Pogorelov TV, Gruebele M (2015) Mapping fast protein folding with multiple-site fluorescent probes. *Proc Natl Acad Sci USA* 112:7966–7971.
30. Sukenik S, Pogorelov TV, Gruebele M (2016) Can local probes go global? A joint experiment-simulation analysis of λ(6-85) folding. *J Phys Chem Lett* 7:1960–1965.
31. Shaw DE, et al. (2014) Anton 2: Raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. *SC14: International Conference for High Performance Computing, Networking, Storage and Analysis* (IEEE Computer Society, Los Alatimos, CA), pp 41–53.
32. Piana S, Klepeis JL, Shaw DE (2014) Assessing the accuracy of physical models used in protein-folding simulations: Quantitative evidence from long molecular dynamics simulations. *Curr Opin Struct Biol* 24:98–105.
33. Liu F, Gao YG, Gruebele M (2010) A survey of lambda repressor fragments from two-state to downhill folding. *J Mol Biol* 397:789–798.
34. Vivian JT, Callis PR (2001) Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys J* 80:2093–2109.
35. Acharya N, Mishra P, Jha SK (2016) Evidence for dry molten globule-like domains in the pH-induced equilibrium folding intermediate of a multidomain protein. *J Phys Chem Lett* 7:173–179.
36. Prigozhin MB, Gruebele M (2013) Microsecond folding experiments and simulations: A match is made. *Phys Chem Chem Phys* 15:3372–3388.
37. Gruebele M (2005) Downhill protein folding: Evolution meets physics. *C R Biol* 328:701–712.
38. Noé F, et al. (2011) Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments. *Proc Natl Acad Sci USA* 108:4822–4827.
39. Lim WA, Sauer RT (1989) Alternative packing arrangements in the hydrophobic core of lambda repressor. *Nature* 339:31–36.
40. Muñoz V, Eaton WA (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc Natl Acad Sci USA* 96:11311–11316.
41. Yang WY, Gruebele M (2004) Rate-temperature relationships in lambda-repressor fragment lambda 6-85 folding. *Biochemistry* 43:13018–13025.
42. Wirth AJ, Liu Y, Prigozhin MB, Schulten K, Gruebele M (2015) Comparing fast pressure jump and temperature jump protein folding experiments and simulations. *J Am Chem Soc* 137:7152–7159.
43. Feller SE, Pastor RW, Rojnuckarin A, Bogusz S, Brooks BR (1996) Effect of electrostatic force truncation on interfacial and transport properties of water. *J Phys Chem* 100:17011–17020.
44. Yoo J, Aksimentiev A (2018) New tricks for old dogs: Improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Phys Chem Chem Phys* 20:8432–8449.
45. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
46. Piana S, Lindorff-Larsen K, Shaw DE (2011) How robust are protein folding simulations with respect to force field parameterization? *Biophys J* 100:L47–L49.
47. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802.
48. Martyna GJ, Tobias DJ, Klein ML (1994) Constant pressure molecular dynamics algorithms. *J Chem Phys* 101:4177–4189.
49. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092.
50. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14:33–38, 27–28.
51. Lippert RA, et al. (2013) Accurate and efficient integration for molecular dynamics simulations at constant temperature and pressure. *J Chem Phys* 139:164106.
52. Shan Y, Klepeis JL, Eastwood MP, Dror RO, Shaw DE (2005) Gaussian split Ewald: A fast Ewald mesh method for molecular simulation. *J Chem Phys* 122:54101.