THE JOURNAL OF PHYSICAL CHEMISTRY B

pubs.acs.org/JPCB Article

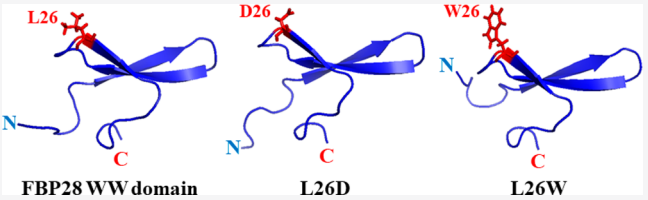
New Insights into Folding, Misfolding, and Nonfolding Dynamics of a WW Domain

Khatuna Kachlishvili,[▽] Anatolii Korneev,[▽] Luka Maisuradze,[▽] Jiaojiao Liu, Harold A. Scheraga, Alexander Molochkov, Patrick Senet, Antti J. Niemi, and Gia G. Maisuradze*[,▽]

Cite This: *J. Phys. Chem. B* 2020, 124, 3855−3872 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Intermediate states in protein folding are associated with formation of amyloid fibrils, which are responsible for a number of neurodegenerative diseases. Therefore, prevention of the aggregation of folding intermediates is one of the most important problems to overcome. Recently, we studied the origins and prevention of formation of intermediate states with the example of the Formin binding protein 28 (FBP28) WW domain. We demonstrated that the replacement of Leu26 by Asp26 or Trp26 (in ∼15% of the folding trajectories) can alter the folding scenario from three-state folding, a major folding scenario for the FBP28 WW domain (WT) and its mutants, toward two-state or downhill folding at temperatures below the melting point. Here, for a better understanding of the physics of the formation/elimination of intermediates, (i) the dynamics and energetics of formation of β-strands in folding, misfolding, and nonfolding trajectories of these mutants (L26D and L26W) is investigated; (ii) the experimental structures of WT, L26D, and L26W are analyzed in terms of a kink (heteroclinic standing wave solution) of a generalized discrete nonlinear Schrödinger equation. We show that the formation of each β-strand in folding trajectories is accompanied by the emergence of kinks in internal coordinate space as well as a decrease in *local* free energy. In particular, the decrease in downhill folding trajectory is ∼7 kcal/mol, while it varies between 31 and 48 kcal/mol for the three-state folding trajectory. The kink analyses of the experimental structures give new insights into formation of intermediates, which may become a useful tool for preventing aggregation.



INTRODUCTION

Protein folding is a rapid and complex process that is difficult to characterize, because folding does not refer to the progressive pathway of a single conformation. Instead, it pertains to interconversions among ensembles of conformations in a back-and-forth progression from the non-native to the native state. In addition, the non-native and native states themselves may consist of a large ensemble of conformations, interconverting at a rapid rate, and characterized by basins with many minima in each state. The two-state model, deduced from calorimetric experiments,[1] is one of the widely used models for the description of folding of single-domain proteins consisting of the non-native and the native state separated by the energetically unfavorable transition state. In this scenario, at the folding-transition temperature (T_f), the non-native (fully unfolded) and native (fully folded) conformations are equally populated; that is, the equilibrium ensemble contains their 50−50% mixture. This means that, for a folding scenario characterized by single-exponential kinetics, a protein goes directly from the unfolded to the native state without detectable intermediates on the way and, hence, makes it difficult to extract mechanistic information. The origins and heights of the barriers separating the unfolded and native states have been studied both computationally and by experiment. For example, Thirumalai[2] showed that, at the atomic level, the free-energy barrier height scales as $N^{1/2}$, where N is the number of residues; while Akmal and Munoz,[3] by investigating six single-domain proteins ranging from 56 to 107 residues, found that the barrier height, determined by the relative contribution from local and nonlocal interactions to protein stability, is only a few kilocalories per mole; and it arises because, in the folding reaction of two-state proteins, the conformational entropy changes at a faster rate than the free energy of stabilization.

Many experimental studies[4−8] have shown that a folding pathway is not always defined in terms of a two-state model. Proteins can fold through intermediate states or undergo one-state (downhill) folding. This scenario of folding was pointed out by Poland and Scheraga 55 years ago.[9] On the basis of the classical Landau theory for critical transitions, Munoz and Sanchez-Ruiz[6] showed that the appearance of a peak in heat-

Received: January 22, 2020
Revised: April 5, 2020
Published: April 9, 2020

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872

capacity curves can very well be explained in terms of a continuous order parameter (enthalpy) at various temperatures and not the competition between the well-defined folded and unfolded state. An analysis of experimental heat-capacity curves showed that the density of states (as a function of enthalpy) has either one maximum or two maxima, indicating either barrierless folding (e.g., for 1BBL) or the existence of a barrier between the folded and unfolded states (e.g., for thioredoxin).[6] By using simplified atomistic and coarse-grained models, Knott and Chan[10] demonstrated that sharp heat-capacity peaks and steeply increasing sigmoidal curves of the temperature dependence of the radius of gyration generally indicate the presence of a barrier (cooperative folding), while low and broad peaks and less steeply increasing curves of the radius of gyration (such as obtained for 1BBL) indicate barrierless (noncooperative) folding. A marginal cooperativity exhibited in downhill folding causes non-concerted structural disassembly. Characterizing the structural properties of the ensemble of configurations at different temperatures with multidimensional NMR, Munoz[11] was able to reconstruct the folding process of the protein, interpreting all the spectroscopic data with statistical mechanical models that included the most relevant partially folded conformations, and determined that downhill folding is characterized by decoupling between structural elements. Moreover, note that, during downhill folding, all the intermediate structures between the denatured and native states are potentially detectable by experiment, which has led to methods that distill mechanistic information from conventional ensemble experiments, such as monitoring how thermal denaturation depends on the structural probe,[12] analyzing heat-capacity thermograms in terms of low-dimensional free-energy surfaces,[13] etc.

Apart from the two-state and downhill models, the observation of two phases in the kinetic progress curves required consideration of a third state; therefore, a three-state model was introduced. The conventional three-state folding model is one in which there is an intermediate state on the pathway from the unfolded to native state, although in some studies[14] a three-state model was defined as a triangular folding mechanism with an energetically trapped intermediate. Kubelka et al.[15] proposed a three-state model in which the interconversion between the intermediate state and native state is much faster than that between the intermediate state and unfolded state. Therefore, the intermediate state lies on the folded side of the major free-energy barrier. Also, the free-energy barrier calculated in this study ranged from 1.6 to 2.0 kcal/mol. Note that, on the basis of recent studies,[16,17] the widely accepted notion that single- and double-exponential kinetics may emerge during two- and three-state (or multistate) folding, respectively, may not always be the case. We have recently shown that the origins of single- and double-exponential kinetics and their correlations with two- and three-state folding scenarios are related to the relative barrier heights between the various states. In particular, single-exponential kinetics can emerge even in three-state (or multistate) folding when one of the free-energy barriers is much higher than the other.[16] Moreover, it was recently demonstrated that the downhill folding can produce double-exponential kinetics.[17]

Protein-folding intermediates on or off the main folding pathway are a common route to the formation of oligomers and amyloid fibrils, which are linked to protein-misfolding diseases[18−22] such as Alzheimer's,[23] Parkinson's,[24] Huntington's,[25] and Creutzfeldt-Jakob's[26] diseases. Therefore, prevention of the aggregation of folding intermediates is one of the most important problems to surmount. The triple-β-stranded WW domain from the Formin binding protein 28 (FBP28) (PDB ID: 1E0L)[27] is a member of the WW domain family. The WW domains, in general, have been the subject of extensive theoretical[16,21,28−39] and experimental[5,17,20,27,40−49] studies because of their small size, biological importance, and fast-folding kinetics. From the folding point of view, the FBP28 WW domain (WT) is a very interesting system to study because of its biphasic folding kinetics.[5,28,29]

Note that a folding mechanism of the FBP28 was disputable for a long time because of its complexity. There are not only differences between experimental and theoretical results but also different experiments that reveal different folding scenarios. In particular, closer to its physiological melting temperature and in the absence of a denaturant, experiments by Nguyen et al.[5] using tryptophan-fluorescence detection revealed slow concentration-independent biphasic kinetics attributed to a folding intermediate. The conclusion regarding three-state folding was challenged by Ferguson et al.,[20] who observed that the FBP28 WW domain readily forms fibrils under similar experimental conditions; hence, the biphasic kinetics has been attributed to an off-pathway intermediate that is a gateway for oligomer formation. Later, our theoretical studies[16,31,32,38] confirmed experimental findings[5] that three-state folding is a main folding scenario for the FBP28 WW domain and that strand-crossing hydrophobic cluster of residues Tyr11, Tyr19, and Trp30 is not associated with the formation of intermediates.

Recently, on the basis of the findings[28,29] that biphasic folding kinetics of the FBP28 WW domain can be caused by slower formation of turn 2 (Arg24 and Thr25) contacts relative to the remainder of the protein, and surface-exposed hydrophobic contact (Tyr21 with Leu26) enforcing the correct registry of hairpin 2, we determined the structures of six new mutants [L26D (PDB ID: 2N4R), L26E (PDB ID: 2N4S), L26W (PDB ID: 2N4T), E27Y (PDB ID: 2N4U), T29D (PDB ID: 2N4V), and T29Y (PDB ID: 2N4W)] by high-resolution NMR spectroscopy and performed extensive studies of folding dynamics of these systems[50] with the coarse-grained united-residue (UNRES) force field.[30,33,51,52] It was found that three-state folding is a major folding scenario for all six mutants; however, two-state and downhill folding scenarios were also identified in ∼15% of the folding trajectories for L26D and L26W. We also discovered that, for formation of intermediates, it is crucial how each hairpin, especially hairpin 1, folds. If both hairpins are formed by the Matheson and Scheraga mechanism[53] (also known as a zipper model[54]), then the system may fold through a downhill or two-state folding scenario. If hairpin 1 is formed by the hydrophobic collapse mechanism,[55] then an intermediate state emerges, and the protein folds through a three-state folding scenario.

Moreover, we tried to elucidate the structural basis for a three-state, downhill, and two-state folding by scrutinizing the structures of all mutants and the FBP28 WW domain determined by high-resolution NMR spectroscopy.[50] For example, for L26D we found that the D26 side chain is consistently oriented toward the Y21 hydroxyl, which suggests the presence of a water-mediated hydrogen bond that stabilizes that specific orientation allowing some "flexibility" during the correct registry of turn 2. In other words, it may either speed up (downhill folding) or slow down (three-state folding) the

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872

1	2	3	4	5	6	7	8	9	10	11	12
GLY	ALA	THR	ALA	VAL	SER	GLU	TRP	THR	GLU	TYR	LYS

13	14	15	16	17	18	19	20	21	22	23	24
THR	ALA	ASP	GLY	LYS	THR	TYR	TYR	TYR	ASN	ASN	ARG

25	26	27	28	29	30	31	32	33	34	35	36	37
THR	ASP/TRP	GLU	SER	THR	TRP	GLU	LYS	PRO	GLN	GLU	LEU	LYS

A

B

Figure 1. Amino acid sequence of the triple β-strand WW domain from the Formin binding protein 28 (FBP28) (1E0L) (A) and experimental NMR-derived structure of 1E0L (B). The mutated residue is highlighted in red color (A) and represented by sphere (B). Residues forming the β-strands are highlighted in yellow color.

correct registry of turn 2 in contrast to the FBP28 WW domain. In the L26W mutant, the W26 substitution presented contacts more similar to the WT, with the orientation of their side chain in the calculated structures resembling that of L26. We observed contacts between the indole of W26 and the aromatic ring of Y21, but the calculated structures result in several orientations of the W26 ring that are compatible with the experimental restraints and do not affect the turn structure; hence, their correlation with different folding scenarios is not straightforward.

To corroborate these findings (especially for L26W), we performed molecular dynamics (MD) simulations with an all-atom force field along with T-jump experiments for L26D and L26W.[56] In that study,[56] our findings from earlier work[50] were validated both theoretically and experimentally. Therefore, it is important to find out what governs the L26W mutant to alter the folding scenario from three-state folding toward downhill folding.

To answer this question and provide new insights into biphasic kinetics, in this study, we (i) examine three-state and downhill folding, as well as nonfolding of L26D and L26W (Figure 1) in terms of internal coordinates (local motions) and the *local* free energies along the entire sequence, as well as with principal components (global motions)[31,32,57−59] and (ii) investigate the experimental structures of the FBP28 WW domain, L26D, and L26W from another, complementary point of view. In particular, instead of analyzing individual interactions that contribute to the formation of folded structure, model-independent principles that are based on symmetry are looked for. We suggest that all the physical forces, no matter how strong or weak they are, combine together to give rise to a particular type of protein dynamics, described by a generalized version of the discrete nonlinear Schrödinger (DNLS) equation, and propose that various parametrizations of a kink, or heteroclinic standing wave solution, of a generalized version of the DNLS equation to be utilized as the basic modular building blocks of folded proteins.[60−69] Note that, in most of the refs [60−66, 68, 69] the epithet topological (dark) soliton was used, but here kink is preferred. This choice highlights that the potential in the DNLS equation displays spontaneous breakdown of a discrete symmetry, and the kink describes the ensuing domain wall;[70−74] in the case of a folded protein an individual kink corresponds to a supersecondary structure such as a helix−loop−helix or strand-loop-strand motif. Also, the kink considered here has no direct relationship with the concept of Davydov's soliton.[75] The kink in a protein is a purely topological (geometric) structure, and the Davydov's soliton, in turn, is associated with collective excitations traveling along a chain (perturbations of electron density). The kinks discussed here can be formally related to the spontaneous formation of intrinsic localized modes (ILMs) in proteins.[76] Indeed, ILMs are large excitations localized on four successive residues and can be viewed as *a pair* of kink and *anti-kink*.[77] However, ILMs occur transiently with a lifetime on a picosecond time scale and are characteristic of the unfolded protein state. In the present work, the kinks do not occur in pairs, are stable structures localized on four to five residues, and are characteristic of the native-state.

One hundred and twenty (for L26D and WT) and ninety-six (for L26W) canonical MD trajectories, generated with the coarse-grained UNRES force field,[30,33,51,52] were run at five and four different temperatures, respectively (24 MD trajectories, with ∼1.4 μs UNRES time and corresponding effectively to ∼1.4 ms actual time of each, at each temperature), which were below, very close to, and above the melting temperatures. The time scale of UNRES is extended by ∼3 orders of magnitude because of averaging out the fast motions of the secondary degrees of freedom[78] and scaling down water friction in MD simulations by a factor of 1000. Hence, the statistics generated by UNRES is several orders of magnitude larger than those obtained by all-atom MD simulations. The trajectories, in which the mutants fold with three-state and downhill scenarios and do not fold, were selected and analyzed in terms of the backbone virtual-bond angle θ and the backbone virtual-bond-dihedral angle γ of each residue (see the UNRES model of polypeptide chains in Figure 2), the *local* free energies,[30,33,50−52] and principal component analysis.[31,32,57−59] Experimental structures of FBP28 WW domain, L26D, and L26W examined in terms of kinks are determined in refs 27 and 50.

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872

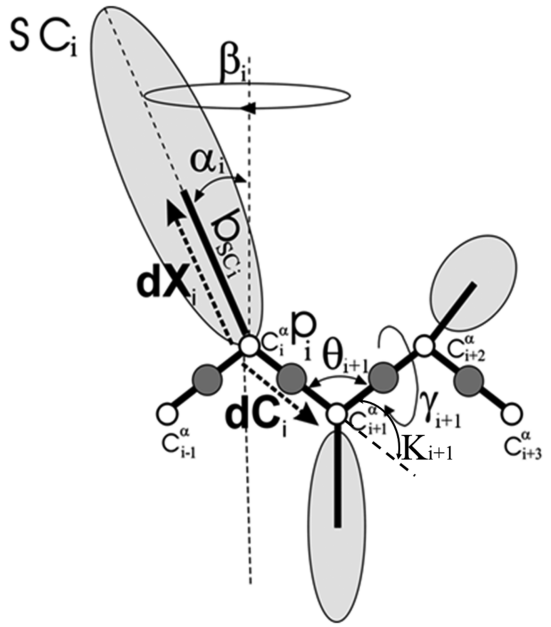


Figure 2. UNRES model of polypeptide chains. The interaction sites are peptide-bond centers (p) and side-chain ellipsoids of different sizes (SC) attached to the corresponding α-carbons with different "bond lengths", b_{SC}. The α-carbon atoms are represented by small open circles. The equilibrium distance of the $C^{\alpha}\cdots C^{\alpha}$ virtual bonds is taken as 3.8 Å, which corresponds to planar *trans* peptide groups. The geometry of the chain can be described either by the virtual-bond vectors $\mathbf{dC}_i$ ($C^{\alpha}_i\cdots C^{\alpha}_{i+1}$), $i = 1, 2, ..., N - 1$, and $\mathbf{dX}_i$ ($C^{\alpha}_i\cdots SC_i$), $i = 2, 3, ..., N - 1$ (represented by thick dashed arrows), where N is the number of residues, or in terms of virtual-bond lengths, backbone virtual-bond angles θ_i, $i = 2, 3, ..., N - 1$, backbone virtual-bond-dihedral angles γ_i, $i = 2, 3, ..., N - 2$, and the angles α_i and β_i, $i = 2, 3, ..., N - 1$ that describe the location of a side chain with respect to the coordinate frame defined by C^{α}_{i-1}, C^{α}_i, and C^{α}_{i+1}. The angles κ_i used here are complements of the θ_i angles, that is, $\pi - \theta_i$.

METHODS

UNRES Force Field. Detailed descriptions of the UNRES force field and its parametrization are available in ref 33, 51, and 52 and at http://www.unres.pl. Therefore, it will be only briefly outlined here. In the UNRES force field,[33,51,52] a polypeptide chain is represented as a sequence of α-carbon (C^{α}) atoms with united peptide groups (p) located halfway between the consecutive C^{α} atoms and united side chains (SC) attached to the C^{α} atoms (Figure 2). The force field has been derived as the potential of mean force (PMF), also termed the restricted free energy function (RFE), of a system of polypeptide chain(s) in solvent, where all degrees of freedom except the coordinates of the C^{α} atoms and SC centers have been averaged out. The effective energy function contains local and site−site interactions, as well as multibody terms, which have been obtained by decomposing the PMF into factors corresponding to clusters of interactions within and between coarse-grained sites.[51] The SC-SC interaction potentials implicitly include the contribution from solvation.[33,51,52] The force field was calibrated to reproduce the structure and thermodynamics of small model proteins. The force field used in this study was calibrated with the α-helical protein 1ENH and the β-strand protein 1E0L.

Kink and Protein Backbone Geometry. Since detailed descriptions of kinks of the DNLS equation are available in refs 60−69, the main aspects of it will be only outlined here. Protein geometry is described in terms of the C^{α} backbone. The C^{α} atoms coincide with the vertices that are denoted by $\mathbf{r}_i$ with $i = 1, ..., N$. At each vertex, there is an orthonormal triplet ($\mathbf{n}_i$, $\mathbf{b}_i$, $\mathbf{t}_i$) with the unit tangent vector

$$\mathbf{t}_i = \frac{\mathbf{r}_{i+1} - \mathbf{r}_i}{|\mathbf{r}_{i+1} - \mathbf{r}_i|} \tag{1}$$

the unit binomial vector

$$\mathbf{b}_i = \frac{\mathbf{t}_{i-1} \times \mathbf{t}_i}{|\mathbf{t}_{i-1} \times \mathbf{t}_i|} \tag{2}$$

and the unit normal vector

$$\mathbf{n}_i = \mathbf{b}_i \times \mathbf{t}_i \tag{3}$$

This defines the discrete Frenet frame[63,67] at the vertex $\mathbf{r}_i$. The C^{α} backbone bond angles κ_i and torsion angles γ_i shown in Figure 2 can be defined in terms of the frame vectors

$$\kappa_{i+1,i} \equiv \kappa_i = \arccos(\mathbf{t}_{i+1}\cdot\mathbf{t}_i) \tag{4}$$

$$\gamma_{i+1,i} \equiv \gamma_i = \mathrm{sgn}[(\mathbf{b}_{i-1} \times \mathbf{b}_i)\cdot\mathbf{t}_i]\cdot\arccos(\mathbf{b}_{i+1}\cdot\mathbf{b}_i) \tag{5}$$

Conversely, when the values of the bond and torsion angles are all known, the discrete Frenet equation

$$\begin{pmatrix} \mathbf{n}_{i+1} \\ \mathbf{b}_{i+1} \\ \mathbf{t}_{i+1} \end{pmatrix} = \begin{pmatrix} \cos\kappa\cos\gamma & \cos\kappa\sin\gamma & -\sin\kappa \\ -\sin\gamma & \cos\gamma & 0 \\ \sin\kappa\cos\gamma & \sin\kappa\sin\gamma & \cos\kappa \end{pmatrix}_{i+1,i} \begin{pmatrix} \mathbf{n}_i \\ \mathbf{b}_i \\ \mathbf{t}_i \end{pmatrix} \tag{6}$$

computes the frame at vertex $i + 1$ from the frame at vertex i. Once all the frames are known, the entire chain is obtained by

$$\mathbf{r}_k = \sum_{i=0}^{k-1} |\mathbf{r}_{i+1} - \mathbf{r}_i|\cdot\mathbf{t}_i \tag{7}$$

Unlike the tangent vector $\mathbf{t}_i$, the normal and binormal vectors ($\mathbf{n}_i$, $\mathbf{b}_i$) do not appear in eq 7. Therefore, if these vectors are simultaneously rotated around the vector $\mathbf{t}$, the C^{α} geometry remains intact, and only the way how it is framed changes. In particular, rotation by π constitutes the discrete $\mathbb{Z}_2$ gauge transformation

$$\gamma_i \to \begin{cases} \gamma_i - \pi, & \text{if } \gamma_i > 0 \\ \gamma_i + \pi, & \text{if } \gamma_i < 0 \end{cases}$$

$$\kappa_k \to -\kappa_k \text{ for all } k \geq i \tag{8}$$

which was proved to be very convenient in analyzing the protein loop structure.[60−69]

Note that regular protein secondary structures correspond to definite values of (κ_i, γ_i). For example, for the standard right-handed α-helix

$$\begin{cases} \kappa \approx \dfrac{\pi}{2} \\ \gamma \approx 1 \end{cases} \tag{9}$$

and for the standard β-strand

$$\begin{cases} \kappa \approx 1 \\ \gamma \approx \pm\pi \end{cases} \tag{10}$$

with the angles in radians.

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872

The variables κ_i and γ_i are mutually connected by the equations of motion, determined by the atomic-level interactions along the protein chain. The following is the Landau−Ginzburg−Wilson free energy F of the protein backbone in terms of the discrete virtual bond and torsion angles[60−69]

$$F = -\sum_{i=1}^{N-1} 2\kappa_{i+1}\kappa_i + \sum_{i=1}^{N}\left\{2\kappa_i^2 + \lambda(\kappa_i^2 - m^2)^2 + \frac{q}{2}\kappa_i^2\gamma_i^2 - p\gamma_i + \frac{r}{2}\gamma_i^2\right\} \tag{11}$$

where λ, q, p, r, and m are parameters. The detailed derivation of eq 11 can be found in refs 60, 61, 66, and 67; here, it suffices to state that this free energy can be shown to relate to the long-distance limit that describes the full microscopic energy of a folded protein in the universal sense of refs 79−82. As such, it does not explain the details of the (sub)atomic-level mechanisms that give rise to protein folding.

To determine the kink content of the backbones, at first, we should define positions of the inflection points (kink centers). We can find the inflection points by analyzing the three-dimensional structure (protein topology) and the experimental values of the angles (κ_i, γ_i). Inflection points correspond to the centers of kinks. Then, the angular spectrum should be reconstructed using the $\mathbb{Z}_2$ transformation at the kink centers (note that the three-dimensional (3D) shape of the protein is invariant under $\mathbb{Z}_2$ transformation). After that, for a given kink structure, we look for a minimum of global energy using a combination of simulated annealing and gradient descent methods. For this proposal a special toolkit for protein structure visualization and analysis was developed [https://protoin.ru/propro/index.php].[68]

In short, the search for the minimum energy occurs in the following order. The virtual-bond angles κ are first extended to negative values, using the $\mathbb{Z}_2$ gauge symmetry (eq 8). The virtual-bond-dihedral angles γ are then expressed as functions of the virtual-bond angles κ

$$\gamma_i[\kappa] = \frac{p}{r + q\kappa_i^2} \equiv \frac{u}{1 + v\kappa_i^2} \tag{12}$$

with $u = p/r$ and $v = q/r$. By inserting eq 12 into eq 11, the virtual-bond-dihedral angles γ are eliminated, and the following system of equations for the motion of the virtual-bond angles κ is obtained

$$\kappa_{i+1} = 2\kappa_i - \kappa_{i-1} + \frac{dV[\kappa]}{d\kappa_i^2}\kappa_i (i = 1, ..., N) \tag{13}$$

where $\kappa_0 = \kappa_{N+1} = 0$ and

$$V[\kappa] = \frac{p}{r + q\kappa^2} + 2(1 - \lambda m^2)\kappa^2 + \lambda\kappa^4 \tag{14}$$

where the familiar structure of the generalized DNLS equation is recognized.[60−69] The kink solution to eq 13 can be constructed numerically by following the iterative procedure of ref 61. But its explicit form, until now, has not been found in terms of elementary functions. However, an excellent approximation is obtained by naively discretizing the heteroclinic standing wave solution to the continuum nonlinear Schrödinger equation[60−69]

$$\kappa_i = \frac{b\ \exp[\sigma_1(i - s)] + a\ \exp[-\sigma_2(i - s)]}{\exp[\sigma_1(i - s)] + \exp[-\sigma_2(i - s)]} \tag{15}$$

and with $\gamma_i[\kappa]$ evaluated from eq 12. Here, s is a parameter that determines the center of the kink. The a, $b \in [0,\pi]$ mod(2π) are parameters that determine the amplitude of the variation of κ and the asymmetry of the inflection regions; they correspond to the minima of the potential energy contribution $V[\kappa]$ in eq 14. The parameters σ_1 and σ_2 are related to the inverse of the range of the kink. Note that, in the case of proteins, the values of a and b are determined entirely by the adjacent helices and strands. Far away from the center of the kink

$$\kappa_i \to \begin{cases} b\ \mathrm{mod}(2\pi)\ i > s \\ a\ \mathrm{mod}(2\pi)\ i < s \end{cases} \tag{16}$$

and according to eqs 9 and 10, the asymptotic values

$$\kappa_i \approx \pi/2 \text{ or } -\pi/2 \text{ and } \kappa_i \approx 1 \text{ or } -1 \tag{17}$$

correspond to the α-helix or β-strand, respectively. To satisfy the monotonic character of the profile of eq 15, the experimentally measured values of κ_i have to vary monotonically along the amino-acid sequence. Otherwise, a multiple of 2π is added to the experimental values. This does not affect the backbone geometry, because κ_i values are defined mod (2π). The σ_1 and σ_2 are intrinsically specific parameters for a given loop. But they specify only the length of the loop, not its shape, which is defined by the functional form of eq 15, and in the case of a and b, they are combinations of the parameters in eq 14.

In eq 12 for the virtual-torsion angles, γ_i, there are only two independent parameters u and v. As a consequence, the profile of γ_i is determined entirely by the profile of κ_i and by the structure of the adjacent regular secondary structures.

Finally, we introduce the concept of folding index of a protein backbone. The formation, evolution, and structure of a loop along a folding protein can be monitored in terms of topologically determined folding indices. Folding index is defined by the following equation

$$Ind_f = \frac{\Gamma}{\pi} \tag{18}$$

where

$$\Gamma = \sum_{i=n_1+2}^{n_2-2} \begin{cases} \gamma_{i,i+1} - \gamma_{i-1,i} - 2\pi \text{ if } \gamma_{i,i+1} - \gamma_{i-1,i} > \pi \\ \gamma_{i,i+1} - \gamma_{i-1,i} - 2\pi \text{ if } \gamma_{i,i+1} - \gamma_{i-1,i} < -\pi \\ \gamma_{i,i+1} - \gamma_{i-1,i} \text{ otherwise} \end{cases} \tag{19}$$

Here, Γ is the total rotation angle (in radians) that the projections of the C^α atoms of the consecutive loop residues make around the north pole. The folding index is a positive integer when the rotation is counterclockwise and a negative integer when the rotation is clockwise. The folding index classifies loop structures and entire folded proteins in terms of its values.[67]

Molecular Dynamics Simulations. For each mutant, a total of 120 canonical MD simulations were performed with the UNRES force field. The Berendsen thermostat[83] was used to maintain constant temperature. The time step in molecular dynamics simulations was $\delta t = 0.1$ mtu (1 mtu = 48.9 fs is the "natural" time unit of molecular dynamics[84]), and the coupling

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872

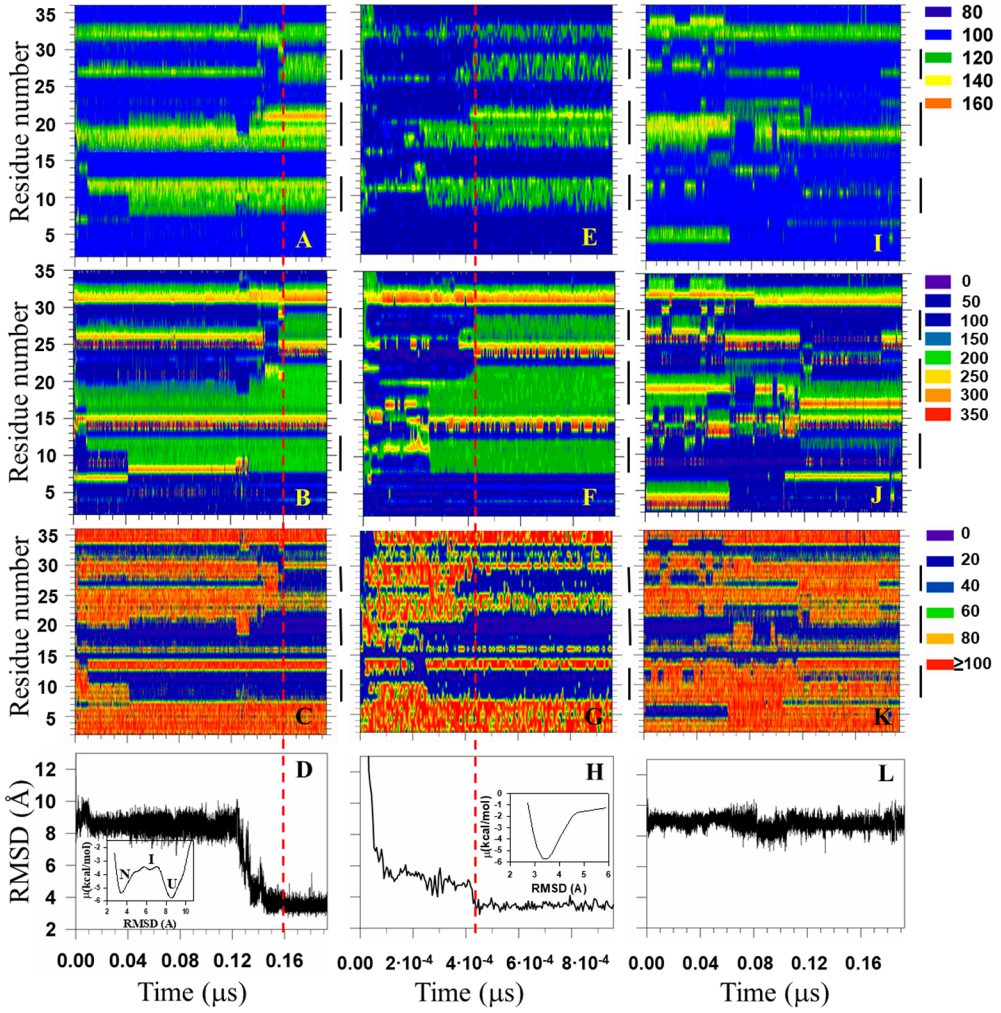


Figure 3. Filled contour plots of the backbone virtual-bond angles θ (deg) (A, E, I), backbone virtual-bond-dihedral angles γ (deg) (B, F, J), and *local* free energies (kcal/mol) (C, G, K) vs time for three-state folding trajectory of L26D mutant (A−C), downhill folding trajectory of L26W mutant (E−G), and nonfolding trajectory of L26D mutant (I−K). The vertical black lines on each panel correspond to the β-strand regions. The RMSDs vs time for three-state folding, downhill folding, and nonfolding trajectory are presented in panels D, H, and L, respectively. The insets in panels D and H represent the free-energy profiles (kcal/mol) as functions of RMSD. Vertical red dashed lines indicate the folding time of L26D and L26W.

parameter of the Berendsen thermostat was τ = 1 mtu. A total of ∼3 × 10^8 MD steps were run for each trajectory, starting from the fully extended structure.

RESULTS AND DISCUSSION

Protein Folding, Misfolding, and Nonfolding in Internal Coordinate Space. Three types of trajectories of the L26D and L26W mutants of the FBP28 WW domain were examined here, in which (i) the protein folds through the intermediate state (i.e., exhibits misfolding during the folding), (ii) the protein exhibits downhill folding (i.e., folds without encountering any significant free energy barrier), and (iii) the protein does not fold. Because the systems fold at the beginning of the trajectories (especially in downhill folding), to better illustrate the results, ∼90% of the trajectories in which the protein remains in the native state were truncated.

One of the main aims of this study was to illustrate how protein folding, misfolding, and nonfolding are reflected on internal coordinates and *local* free energy. As in our previous studies, the protein main chain is described by the positions of the C^α atoms. Assuming constant average virtual bond distances between successive C^α atoms, the protein main chain is entirely described by the backbone virtual-bond angles θ_i between two successive virtual bonds [$(C^\alpha_{i-1},C^\alpha_i)$, and $(C^\alpha_i,C^\alpha_{i+1})$] and the backbone virtual-bond-dihedral angles γ_i built from three successive virtual bonds [$(C^\alpha_{i-1},C^\alpha_i)$, $(C^\alpha_i,C^\alpha_{i+1})$, and $(C^\alpha_{i+1},C^\alpha_{i+2})$]. The contour plots depicted in Figure 3 show how the angles θ_i (panels A, E, I), the angles γ_i (panels B, F, J), and corresponding *local* free energies of the backbone angles θ_i (panels C, G, K) are changing in the course of time when the protein undergoes three-state folding (panels A−D), downhill folding (panels E−H), and no folding (panels I−L). The root-mean-square deviations (RMSDs) versus time illustrated in panels D, H, and L describe how the system folds or does not fold. Note that γ angles explore the neighborhood of 180° in the native state; therefore, to avoid confusion, the range of fluctuations of the γ angles, which is between −180° and 180°, was shifted to [0°; 360°] in panels B, F, and J.

The results show a clear correlation between θ and γ angles and *local* free energies. These variables describe, in detail, how protein folds and why protein does not fold. In particular:

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872

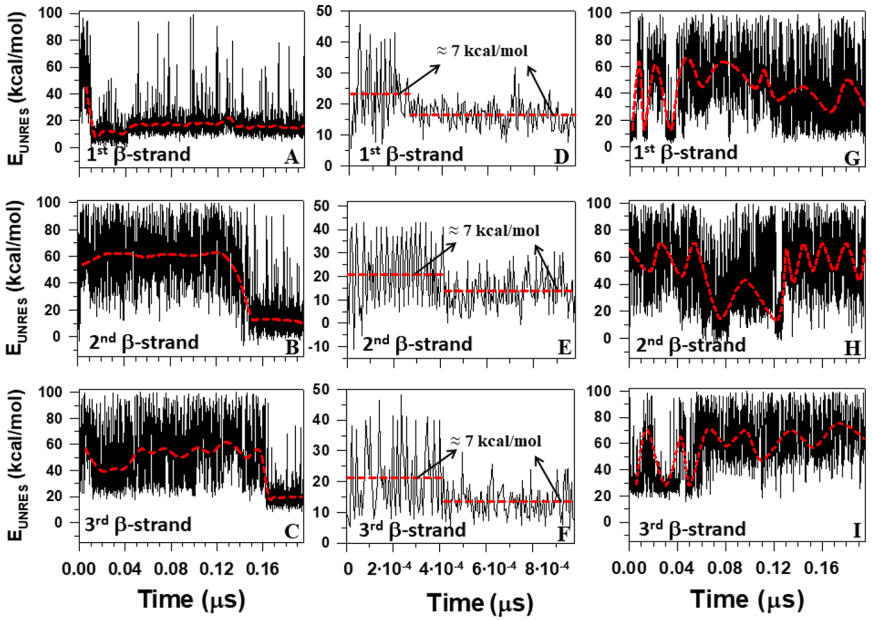


Figure 4. UNRES energy change during the formation of the first, second, and third β-strands for three-state folding trajectory of L26D mutant (A−C), downhill folding trajectory of L26W mutant (D−F), and nonfolding trajectory of L26D mutant (G−I).

i. Three-state Folding. On the basis of the RMSD versus time plot (panel D), the protein remains unfolded during the first 0.12 μs (8Å < RMSD < 10Å), then starts folding (0.12 μs < Time < 0.16 μs), forming an intermediate state, and reaches native state at 0.16 μs (RMSD $\leq$ 4 Å). The contour plots of θ and γ angles (panels A and B) indicate that the formation of all three β-strands starts much earlier in the beginning of trajectory, while the protein is still in an unfolded state. However, they reach complete formation at different times (values of θ and γ angles in β-strands fluctuate around 120° and 180°, respectively; these regions are represented by green color in the contour plots). In particular, the β-strand 1 forms completely at 0.04 μs; ∼90% of the β-strand 2 also forms at 0.04 μs, although it reaches complete formation at ∼0.14 μs; the β-strand 3 forms completely last at 0.16 μs. Corresponding *local* free energies adequately change with the formation of β-strands (panel C). From these results, we can conclude that one of the reasons for the formation of an intermediate state and misfolding is a slow formation of the third β-strand. These findings are in harmony with results obtained in earlier studies.[5,16,28,29,31,32,38,50,56] Moreover, for a clear demonstration of three-state folding, the free-energy profile as a function of RMSD was calculated (see the inset in panel D). Note that, because of averaging out the fast motions of the secondary degrees of freedom, at the course-grained level, the free-energy barriers, illustrated in the inset of panel D, are lower than those at the atomic level.

ii. Downhill Folding. The RMSD versus time plot, presented in panel H, indicates that the protein only needed $\sim 4.3 \times 10^{-4}$ μs to fold (400 times less time than in the three-state folding case). As in three-state folding, in downhill folding the β-strand 1 forms completely first ($\sim 2.5 \times 10^{-4}$ μs), and the β-strands 2 and 3 form completely together at $\sim 4.3 \times 10^{-4}$ μs. Although all three β-strands do not form together, the protein still undergoes downhill folding. The point is that the time difference between the formation of β-strand 1 and β-strands 2 and 3 is very small ($\sim 1.8 \times 10^{-4}$ μs) and not enough for the formation of an intermediate state, which can be corroborated by the free-energy profile as a function of RMSD calculated for the 0.19 μs portion of the trajectory (see the inset in panel H). The changes in *local* free energies in the course of time are in agreement with the changes of θ, γ angles.

iii. Protein Nonfolding. The RMSD of the MD trajectory, illustrated in panel L, fluctuates between ∼8 and 10 Å, which indicates that the protein does not fold in the given time. The contour plots of the θ and γ angles, and *local* free energies, as expected, show very "chaotic behavior"; in particular, β-strand 1 practically does not form at all; instead, transient incomplete α-helical structures can be observed, the formation of β-strand 3 is transient and incomplete, and only β-strand 2 forms almost completely. Moreover, β-strand structures appear in the regions of N and C termini.

These results indicate that the formation of β-strand 1 might be crucial for formation of β-strands 2 and 3, and for folding entirely.

Note that *local* free energies of the backbone dihedral angles γ_i versus time are not shown here, because they behave very similarly to ones illustrated in Figure 3.

Apart from studying, in detail, the representative trajectories of the L26D and L26W mutants, it is of interest to know "a general picture" of the folding dynamics of these mutants. Therefore, we joined folding trajectories at different temperatures and calculated the free-energy profiles as functions of RMSD for both mutants. In addition, for comparison, we did the same calculations for the FBP28 WW domain. The results are shown in Figure S1 (Supporting Information). First, note that, because both mutants fold with the three-state folding scenario in ∼85% of trajectories, none of the free-energy profiles exhibit downhill folding. Clearly, three-state folding in WT is more pronounced than in L26D and L26W, especially at low temperatures. With the increase of temperature intermediate states in all systems gradually disappear, which is in agreement with experimental results.[5,56] These results indicate that, although mutations did not change the folding scenario in all trajectories, they lowered the barriers, which

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872

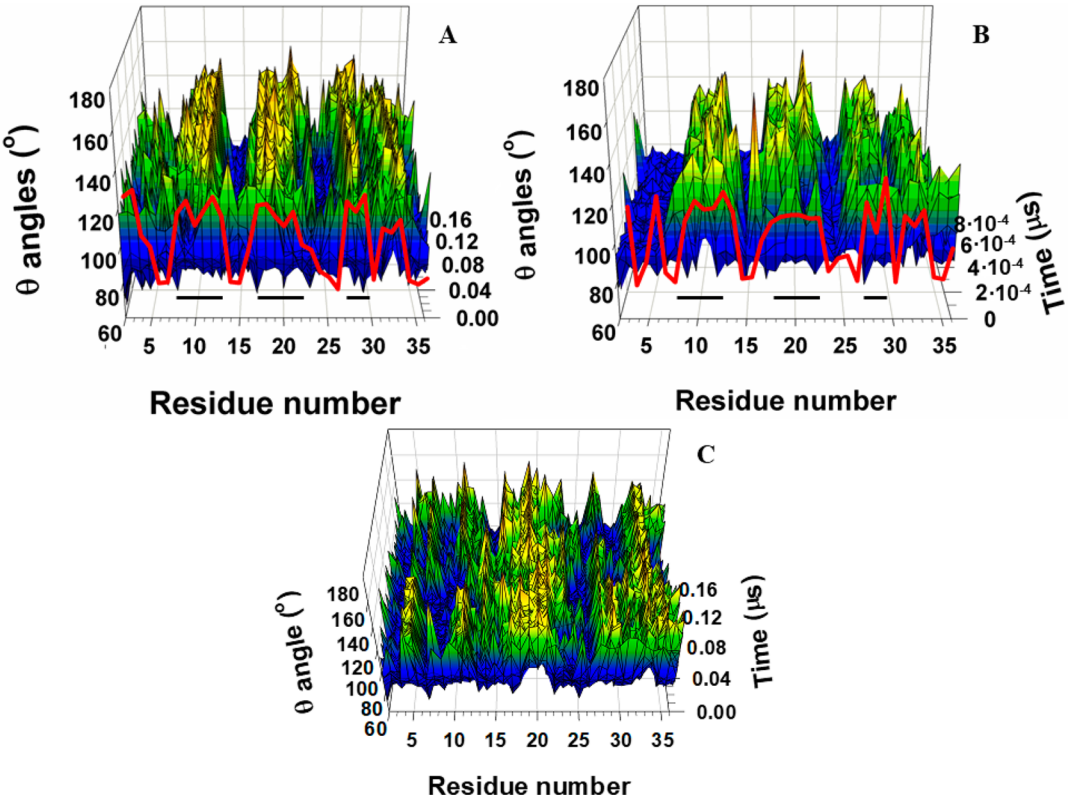


Figure 5. 3D representation of the backbone virtual-bond angles θ vs time for three-state folding trajectory of L26D mutant (A), downhill folding trajectory of L26W mutant (B), and nonfolding trajectory of L26D mutant (C). The NMR-derived structural data (red curves on panels A and B) are computed from the first model of the PDB ID codes 2N4R (L26D) and 2N4T (L26W). The horizontal black lines on panels A and B correspond to the β-strand regions.

enable both systems to fold faster and in some trajectories (~15%) without barriers.

Local Free Energies. As was mentioned in the Introduction, the mechanisms of three-state and downhill folding scenarios were elucidated in our recent study.[50] In particular, in three-state folding, hairpin 1 forms by a hydrophobic collapse mechanism,[55] which causes significant delay of formation of both hairpins, especially hairpin 2, and facilitates the emergence of an intermediate state, while in downhill folding both hairpins are formed by the Matheson–Scheraga mechanism[53] much faster than in three-state folding. It is of interest to know how *local* free energies change during the formation of β-strands and hairpins for both mechanisms, as the purpose of the present work is to identify why the mutations favor one mechanism relative to another.

Figure 4 illustrates how the free energies of the TRP8-THR13 (panels A, D, G), LYS17-ASN23 (panels B, E, H), ASP26-TRP30 (panels C, I), and TRP26-TRP30 (panel F) segments of L26D and L26W, computed by using the UNRES effective energy function, which has the sense of a free energy,[30,33,51,52] change when the formation of β-strands takes place in three-state folding (panels A–C) and downhill folding (panels D–F), as well as when β-strands do not form fully in nonfolding trajectory (panels G–I). The red horizontal dashed lines in panels D–F correspond to the average UNRES energy averaged over "before" and "after" energy jumps in the downhill folding trajectory, whereas the red dashed curves in panels A–C and G–I correspond to the average UNRES energy averaged over the three-state folding and nonfolding trajectories (note that the UNRES energy was not averaged over the downhill trajectory due to its shortness).

It can be seen that, in the downhill folding trajectory, the *local* free energy decreases by ~7 kcal/mol when β-strands form [Figure 4D–F], while the average values of the free-energy decrease during the full formation of β-strands in the three-state folding trajectory are much larger and vary between 31 and 48 kcal/mol [Figure 4A–C]. These findings agree with the well-known notion that, for many proteins, only a modest free-energy gain (generally only −3 to −7 kcal/mol) is associated with the correct folding of a protein compared with its innumerable potential misfolded states.[85] Similar results were reported in a recent study,[67] in which the dynamics and energetics of the formation of loops in the 46-residue N-terminal fragment of the B-domain of staphylococcal protein A were studied. In particular, it was shown that, during the formation of the loops and, consequently, the kinks, the free energy over the putative kink region increased by ~7 kcal/mol.[67] Interestingly, 7 kcal/mol is very close to the free energy of dissociation of one phosphate residue from adenosine triphosphate (ATP).[86] The difference between the previous and presented work is that the UNRES simulations in the previous study[67] were started from the full right-handed α-helix to obtain a clear picture of the loop (consequently kink) formation, whereas in the presented study the MD simulations are started from the fully extended structure, and formation of β-strands are investigated. In the three-state folding trajectory, a hydrophobic collapse is associated with large changes (31–48 kcal/mol) in free energy, which indicates an irreversibility of the process. In other words, if this free energy jump occurs,

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855–3872

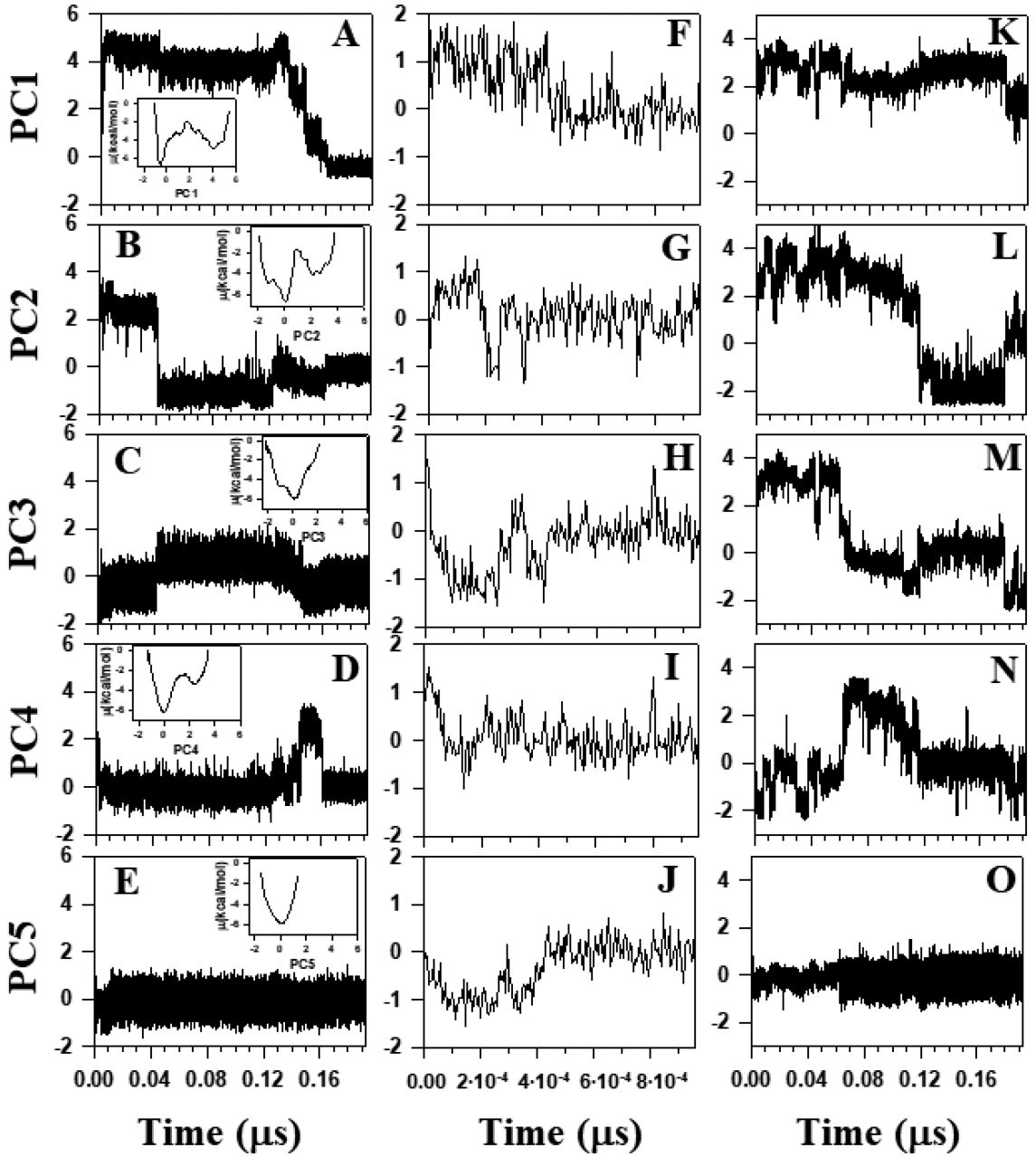


Figure 6. First five principal components of three-state folding trajectory of L26D mutant (A−E), downhill folding trajectory of L26W mutant (F−J), and nonfolding trajectory of L26D mutant (K−O). The insets in panels A−E represent free-energy profiles of the corresponding principal components for the entire three-state folding trajectory (∼1.4 μs).

the protein misfolds as a consequence. It is important to note that the formation of β-strands is accompanied by the emergence of kinks in the internal coarse-grained coordinate space (θ,γ) [as an example, see Figure 5, which is a 3D representation of Figure 3A,E,I].

Moreover, Figure 4 shows that the free energies of TRP8-THR13, LYS17-ASN23, and ASP26-TRP30 segments in the three-state folding (panels A−C) and nonfolding (panels G−I) trajectories oscillate in the course of time. The amplitude of oscillations in the nonfolding trajectory is much larger than in the three-state folding trajectory, which is not surprising given that the β-strands in the nonfolding trajectory are either formed partially or not formed at all [Figure 3I,J]. In addition, the most "oscillating" segment in the three-state folding trajectory of L26D mutant is the third β-strand [Figure 4C], which is also logical, given its flexibility and lowest stability. Another explanation for the *local* energy oscillations was introduced in a recent study.[67] In particular, it was shown that the *local* energy change is caused by oscillation in the distance between the two kinks. Here, even very rough monitoring of the formation of β-strands [Figures 3A,B,I,J and 5A,C] and free-energy oscillations [Figure 4A−C,G−I] of β-strands indicates correlations between the free-energy oscillations and kinks' oscillations. Note that the ability of kinks to oscillate back and forth along the backbone (i.e., appear and disappear in the course of time), when the protein is in an unfolding state, is a very important feature, one that might be used not only for a better understanding of protein folding but also to open a new door in understanding intrinsically disordered proteins, which have been implicated in a number of human

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872

diseases, for example, cancer, cardiovascular disease, neurodegenerative diseases, diabetes, etc.[18,19,22−26,87] Therefore, detailed investigation of intrinsically disordered proteins in terms of kinks is planned in the nearest future.

Principal Component Analysis. Principal component analysis (PCA), a covariance-matrix-based mathematical technique, is an effective method for extracting important motions from molecular dynamics trajectories.[31,32,57−59] PCA rotates the Cartesian or internal coordinate space to a new space with new coordinates, principal components (PCs), a few of which are sufficient to describe a large part of the fluctuations of a protein. Structural fluctuations of θ and γ angles [mean-square fluctuations (MSF)] can be decomposed into collective modes by PCA. The modes have "frequencies" and directions corresponding to the eigenvalues and eigenvectors of the covariance matrix. The modes with the largest eigenvalues correspond to the modes that contribute the most to the structural fluctuations of the protein. It has been shown that the principal modes in folding trajectories are correlated with kink formation.[67] Therefore, it is of interest to know whether and how the changes of free energies and internal coordinates in the course of time in three-state, downhill, and nonfolding trajectories are reflected on PCs, the projections of the trajectories on the eigenvectors.

Figure 6 illustrates the first five PCs of three-state (panels A−E), downhill (panels F−J), and nonfolding (panels K−O) MD trajectories. As was expected, there is a clear correlation between PC1 (Figure 6A) and the RMSD (Figure 3D) for the three-state folding trajectory. Although PC2 (Figure 6B) does not mimic RMSD (Figure 3D), it identifies the changes in RMSD in the course of time. The free-energy profiles along PC1 and PC2 [see insets in Figure 6A,B] are highly rugged, that is, anharmonic, and many local minima appear in a multiple number (two) of coarse-grained minima. Therefore, the free-energy profiles along these two PCs can be characterized as multiply hierarchical.[31,88] The multiply hierarchical PCs are main contributors to the total fluctuations and are associated with global collective motions.[31,88] Some correlation between PC3 and PC4 [Figure 6C,D] and RMSD (Figure 3D) is noticeable in some parts of the trajectory; these PCs mainly identify the transition from the unfolded state to the native state. The probability distributions along PC3 and PC4 are Gaussian-like with a single peak, and the free-energy profiles are characterized by a number of local minima arranged within a single coarse-grained minimum [see insets in Figure 6C,D]. These two PCs belong to a singly hierarchical category.[31,88] The PC5 is not correlated to RMSD. It belongs to a harmonic category (see inset in Figure 6E).[31,88] Since it involves low-amplitude local minima and corresponds to local motions, PC5 does not contribute significantly to the total fluctuations.[31,88] Similar correlations between PCs and RMSD can be observed for downhill folding trajectory; however, because of very fast folding, free-energy profiles along the PCs do not exhibit multiply hierarchical and singly hierarchical categories. The free-energy profiles along all PCs are harmonic. For this reason, they are not shown in Figure 6.

Moreover, comparison of the results depicted in Figures 3A−C,E−G and 6A−C,F−H illustrate that the first two PCs of three-state and downhill folding trajectories capture the changes in internal coordinates (θ, γ) and *local* free energies in the course of time. In particular, the formation of all three β-strands is reflected in PCs. This indicates the correlations between local and global motions in folding trajectories. This is not the case for the nonfolding trajectory, in which none of the PCs is correlated to RMSD. Also, because of the "disorderness" and lack of formation of β-strands, it is very difficult (if not impossible) to determine the correlation between the internal coordinates and *local* free energies and PCs. However, note that the first four PCs of the nonfolding trajectory [Figure 6K−N] oscillate, which indicates that oscillatory behavior of kinks is mirrored on PCs and that they are correlated.

It is important to note that the percentages of total fluctuations captured by the first PCs in the three-state and downhill folding trajectories are ∼40% and 9%, respectively. PCA has proven to be an effective tool for the analysis of protein folding trajectories involving concerted motions of many residues, which can be captured by a few PCs with the largest eigenvalues.[31] These results indicate that PCA can correctly identify the folding scenario. The point is that, as was mentioned earlier, downhill folding proteins exhibit marginal cooperativity causing nonconcerted structural disassembly,[11] which consequently may induce nonconcerted motions. The percentage of captured fluctuations for downhill folding trajectory is quite low (∼9%), which based on the PCA definition is an indication of nonconcerted motions. Therefore, the folding scenario can be determined by the percentage of captured fluctuations.

In the end, note that both *local* free energy analysis and PCA clearly indicate that the kink analysis can be very effective and useful in uncovering the mechanisms of different folding scenarios. Therefore, kinks and their applications for three different systems are discussed in the following section.

Examination of the Experimental Structures of the FBP28 WW Domain and its Mutants in Terms of Kinks. The study of kinks is more easily understood from a differential representation of the protein main chain. Indeed, the angle θ_i is directly related to the local curvature κ_i ($\kappa = \pi - \theta$) of the protein main chain, whereas the angle γ_i is the local torsion of the main chain (see Figure 2 and Methods). Assuming constant virtual bond distances between the C^α atoms, the free energy of the polymer can be written in terms of (κ,γ) internal coordinates using a Landau−Ginzburg−Wilson free-energy model[60−69] [eq 11]. Minimizing this free-energy relative to κ and γ leads to a nonlinear equation of motion for the curvature similar to a generalized DNLS that admits kinks as particular solutions.

In earlier studies,[60−69] the modular building blocks of folded proteins, that is, supersecondary structure such as a helix−loop−helix or strand-loop-strand motif, have been described in terms of the kink of a generalized version of the DNLS equation. It was found that the formation of a kink is initiated by an abrupt change in the orientation of a pair of consecutive side chains in the loop region.[67] It has been shown that, in order to obtain a clear picture of kink formation, it is necessary to start MD simulations from the full α-helical or β-strand structure and investigate the dynamics of the formation of loops; otherwise, it would be blurred by α-helix or β-strand formation.[67] Moreover, kink analysis enabled us to realize the importance of local interactions, specifically the bimodal character of the potential of mean force in virtual-bond angles κ, as the driving force of folding.[67] It is notable that the only long-range interaction present in the Landau−Ginzburg−Wilson Hamiltonian[60−69] is a stepwise Pauli exclusion that introduces self-avoidance and prevents chain crossing. The effects of the long-range Coulomb and van der Waals interactions are accounted for by the global multikink profile

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872

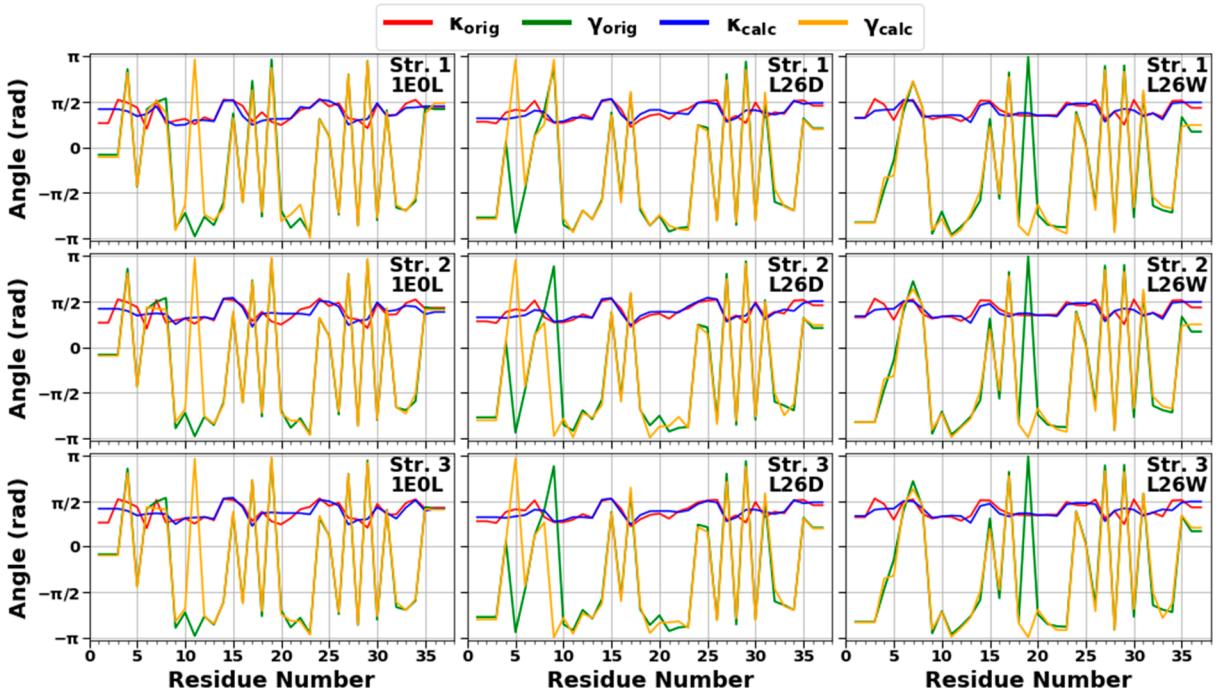


Figure 7. Comparison of experimental and calculated angle spectra of three selected structures for the FBP28 WW domain (left), L26D (middle), and L26W (right) in terms of virtual-bond κ_i (experimental-red, calculated-blue) and torsion γ_i (experimental-green, calculated-yellow) values.

resulting from the minimization of free energy.[89] The multikink structure describes both local and collective motions. Previously, it has been shown that, using kink parameters that are derived from the experimental folded protein structure, its folding can be simulated.[65,66,68,69,90,91] Moreover, a phase diagram of protein with temperature (T) and acidity (pH) as the thermodynamic variables can be constructed as well.[69] Therefore, it is of interest to investigate whether a kink analysis of the experimental structures[27,50] can uncover the mechanisms, which govern L26D and especially L26W to fold through downhill folding.

Kink Structures of the FBP28 WW Domain, L26D and L26W. In this work, we performed a detailed analysis of experimental structures of the FBP28 WW domain (three-state folder) and its mutants, L26D and L26W, which may fold through the three-state, downhill, and two-state folding scenarios, in terms of kinks. Using experimental values of the κ and γ angles (in other words—tertiary structure of protein), we started resolving the kink structures of the FBP28 WW domain, L26D and L26W. The multikink is a configuration that minimizes free energy [eq 11], with an acceptable small deviation, RMSD, from the three-dimensional experimental structure. Minimum energy, as well as RMSD fitting, is achieved using a combination of simulated annealing and gradient descent methods. Multikink structures that are combinations of eight individual kinks, differing in the arrangement along the protein chain, were obtained for each system. With RMSD < 0.8 Å, a total of 20 structures were initially obtained for the FBP28 WW domain and L26D, and 60 structures were obtained for L26W. For each system, three representative structures were then selected for a detailed analysis. Figure 7 illustrates comparison of the κ and γ angles, experimental and calculated, along the chains of three selected structures of the FBP28 WW domain, L26D and L26W. In addition, for each protein, Figures 8 shows experimental values of the κ and γ angles in the first panels and most representative kink structures in the three remaining panels. Data on this figure allow us to scrutinize the correlations between the location of kinks and folding paths. [The experimental and three calculated representative structures for the FBP28 WW domain, L26D, and L26W are illustrated in Figure S2 (Supporting Information)].

The kink locations along almost the entire sequence of the FBP28 WW domain in all three structures are the same, only the centers of the seventh and eighth kinks are moving between the 30th and 31st and between the 33rd and 34th residues, respectively, which indicates the instability of the C-terminal part of the protein (Figure 8A). These findings are in harmony with earlier results obtained by us and other groups.[16,21,28,29,31,32,38] Since the FBP28 WW domain is a three-state folder, we consider these kink locations as a reference for three-state folding.

In L26D, the center of the eighth kink does not change; instead, we observe a movement of the fifth kink center between the 22nd and 23rd residues along with the center of the seventh kink between the 30th and 31st residues (Figure 8B). These results indicate that the C-terminal becomes more stable and that turn 2 has more flexibility, which may either speed up (downhill folding) or slow down (three-state folding) the correct registry of turn 2. The resemblance of the locations of kink centers in the third structure of the FBP28 WW domain (Figure 8A) and the first structure of L26D (Figure 8B) suggests the presence of three-state folding. The differences observed in the fifth kink (the center is shifted from the 23rd residue to 22nd residue) in the second and third structures indicate that L26D can fold through two-state or downhill folding. These findings are in agreement with our recent results.[50]

In our recent studies,[50,56] we found that L26W is the fastest folder among all studied mutants; however, examination of the structural basis of L26W did not reveal the reasons for fast folding. The kink structures of L26W (Figure 8C) exhibit

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872

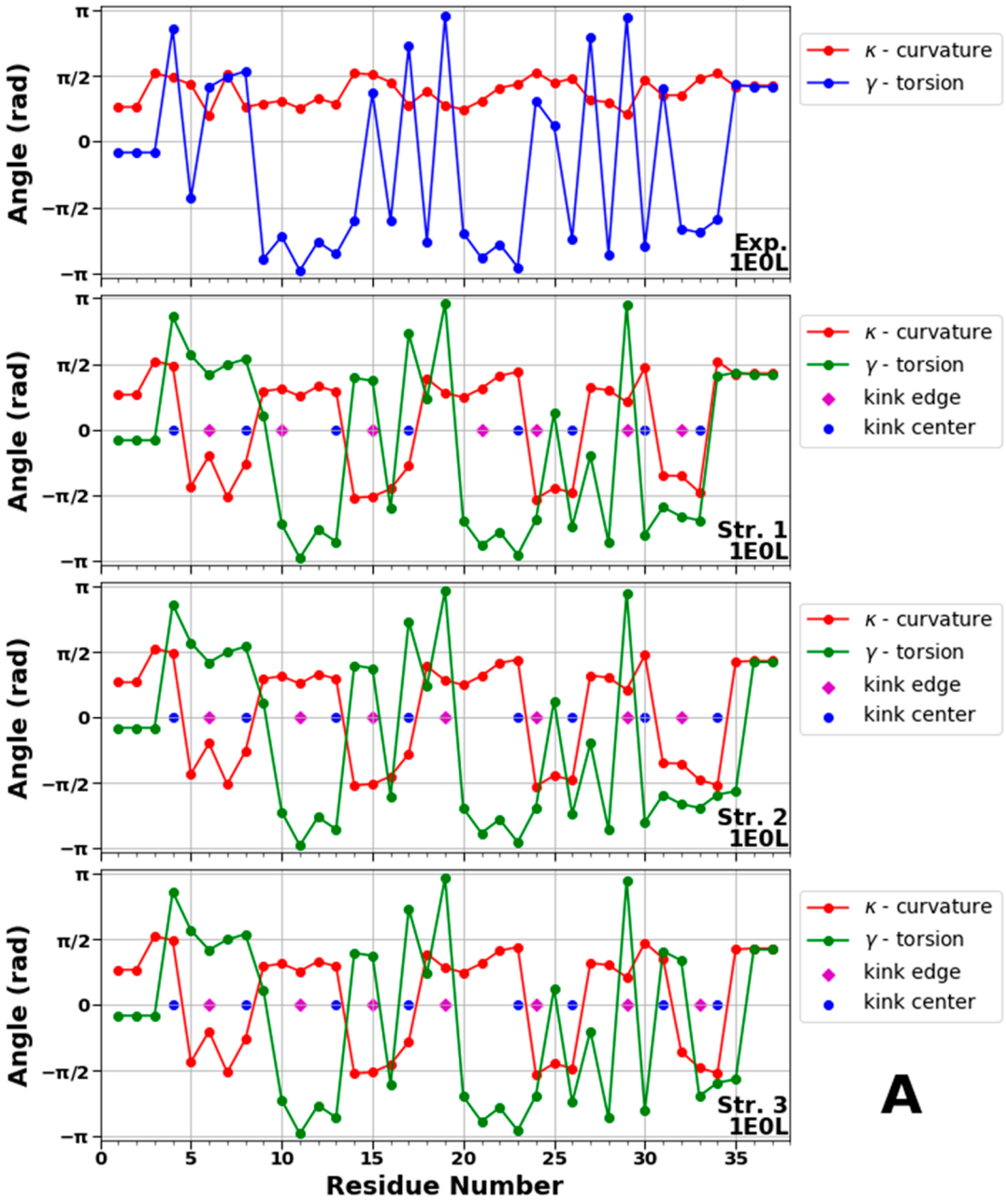


Figure 8. continued

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872

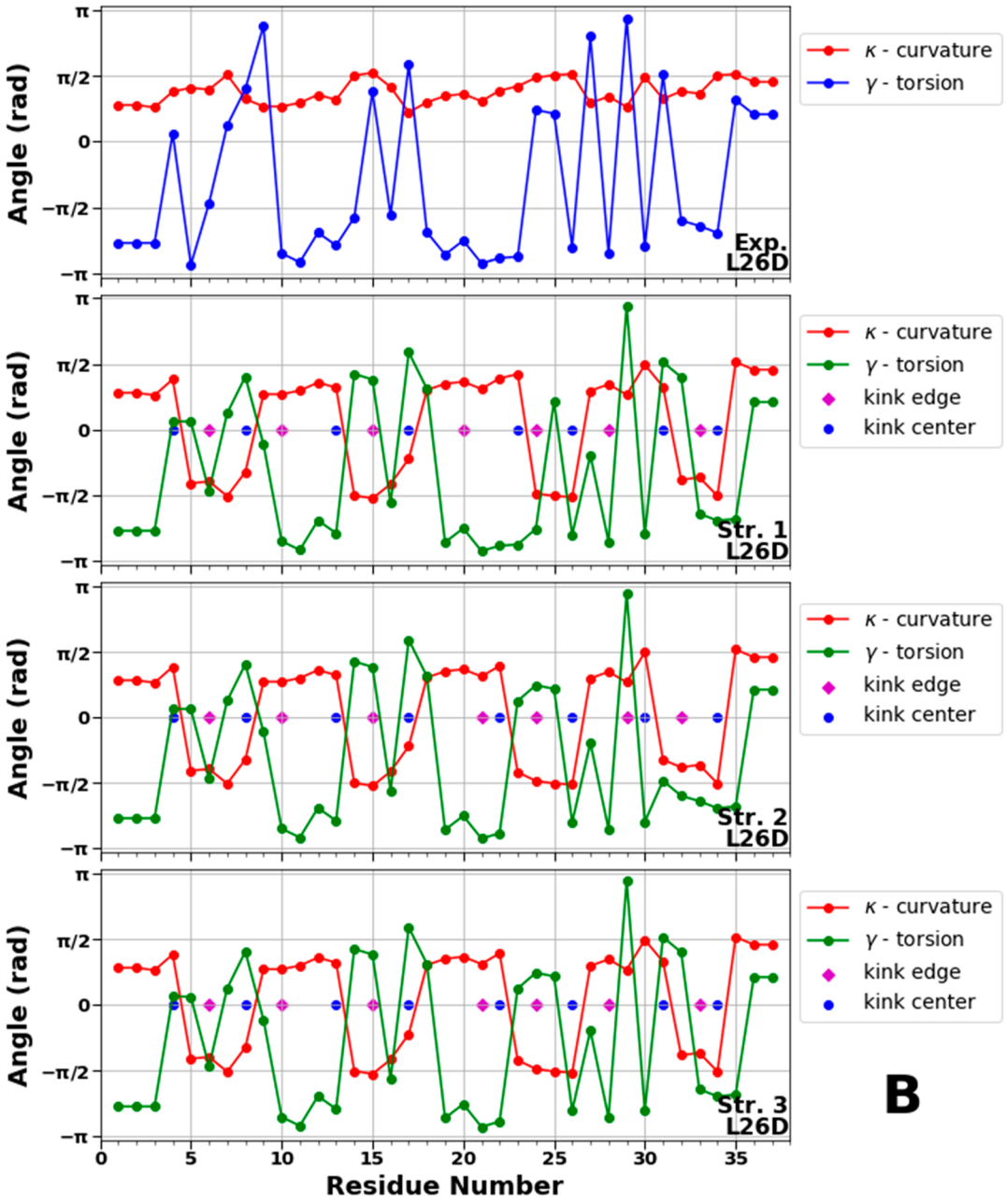


Figure 8. continued

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872

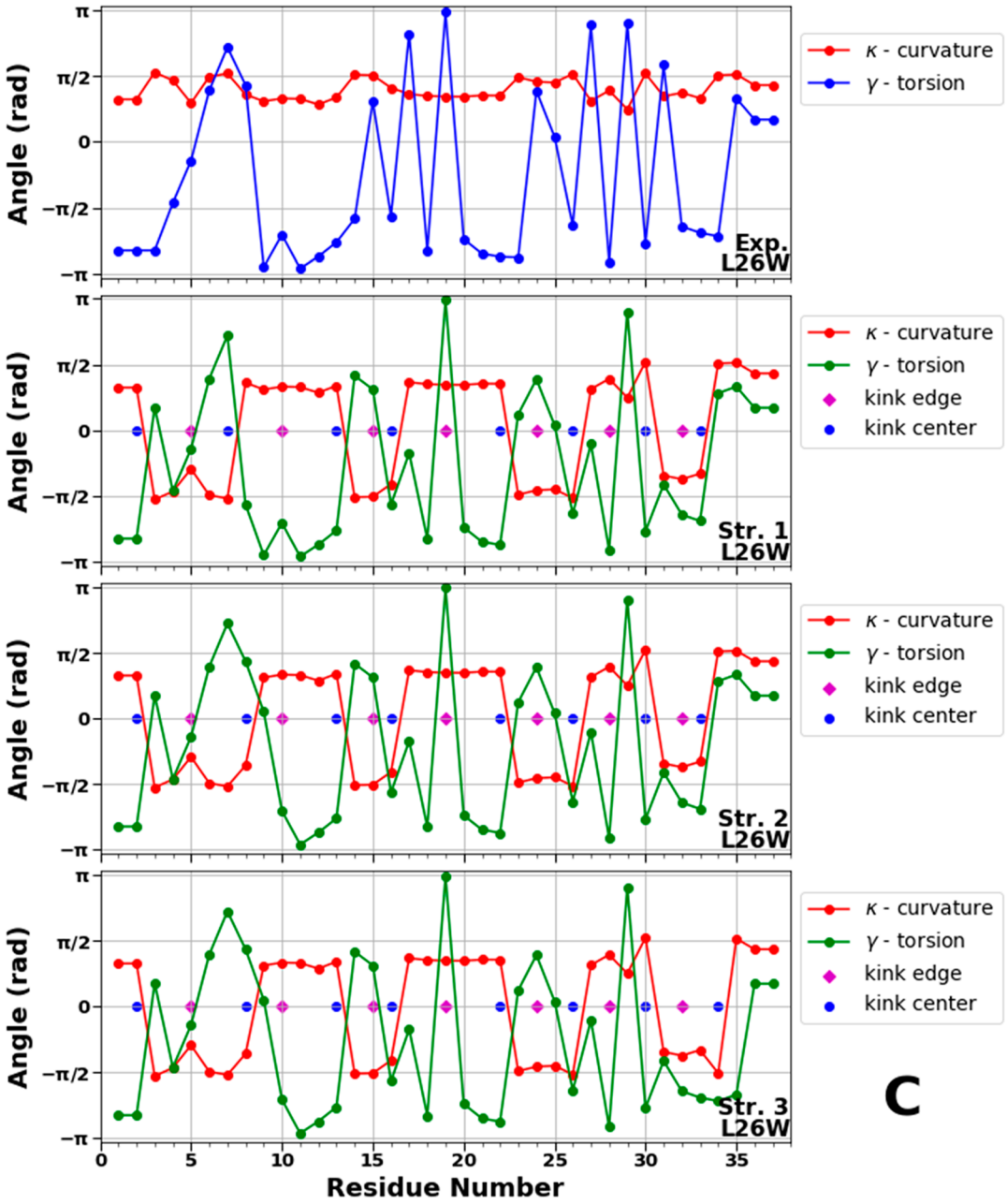


Figure 8. (A−C) Top panels represent experimental values of the virtual-bond κ_i (red) and torsion γ_i (blue) angle spectra of 1E0L (A), 2N4R (B), and 2N4T (C). Remaining three panels in A, B, and C illustrate kink structures, i.e., the virtual-bond κ_i (red) and torsion γ_i (green) angle spectra for three selected structures of 1E0L (A), 2N4R (B), and 2N4T (C), after $\mathbb{Z}_2$ gauge transformation (eq 8) was made. Blue circles represent centers of kinks; purple rhombuses represent the right edges of left kinks.

different locations of centers in the N-terminal region and turn 1 compared to the FBP28 WW domain and L26D. First, note that the center of the fifth kink in L26W is located at the 22nd residue, as it was in the fast-folding structures of L26D, which indicates the fast formation of hairpin 2. Second, the center of the second kink in the first structure of L26W is shifted from the eighth to the seventh residue, which might be an indication of possible destabilization of the delocalized hydrophobic core (Trp8, Tyr20, and Pro33)—"one of the contributors" in the formation of the intermediate state.[5] Third, the shift of the fourth kink center from the 17th to the 16th residue may speed up the formation of turn 1, which along with the destabilized delocalized hydrophobic core speeds up the folding of the entire system. Finally, the center of the first kink in L26W is shifted by two residues (from the fourth residue to the second residue). Since this shift occurs at the N-terminal end, and neither Ala4 nor Ala2 is associated with any hydrophobic core stabilizing the intermediate state, its influence, in our opinion, on the folding scenario is probably not significant. However, it is worth investigating in the future. It is important to mention that these results are correlated with the results obtained in our recent work[50] for the first two principal modes. In particular, the main contributions of the first two principal modes to the mean-square fluctuations along the θ and γ angles in downhill folding trajectories come from the N terminus, the first β-strand, and the first turn [see Figure 1H,I,K,L of ref 50].

Folding Index for the FBP28 WW Domain, L26D, and L26W. Figure 9 shows the accumulation of the folding index[48]

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872

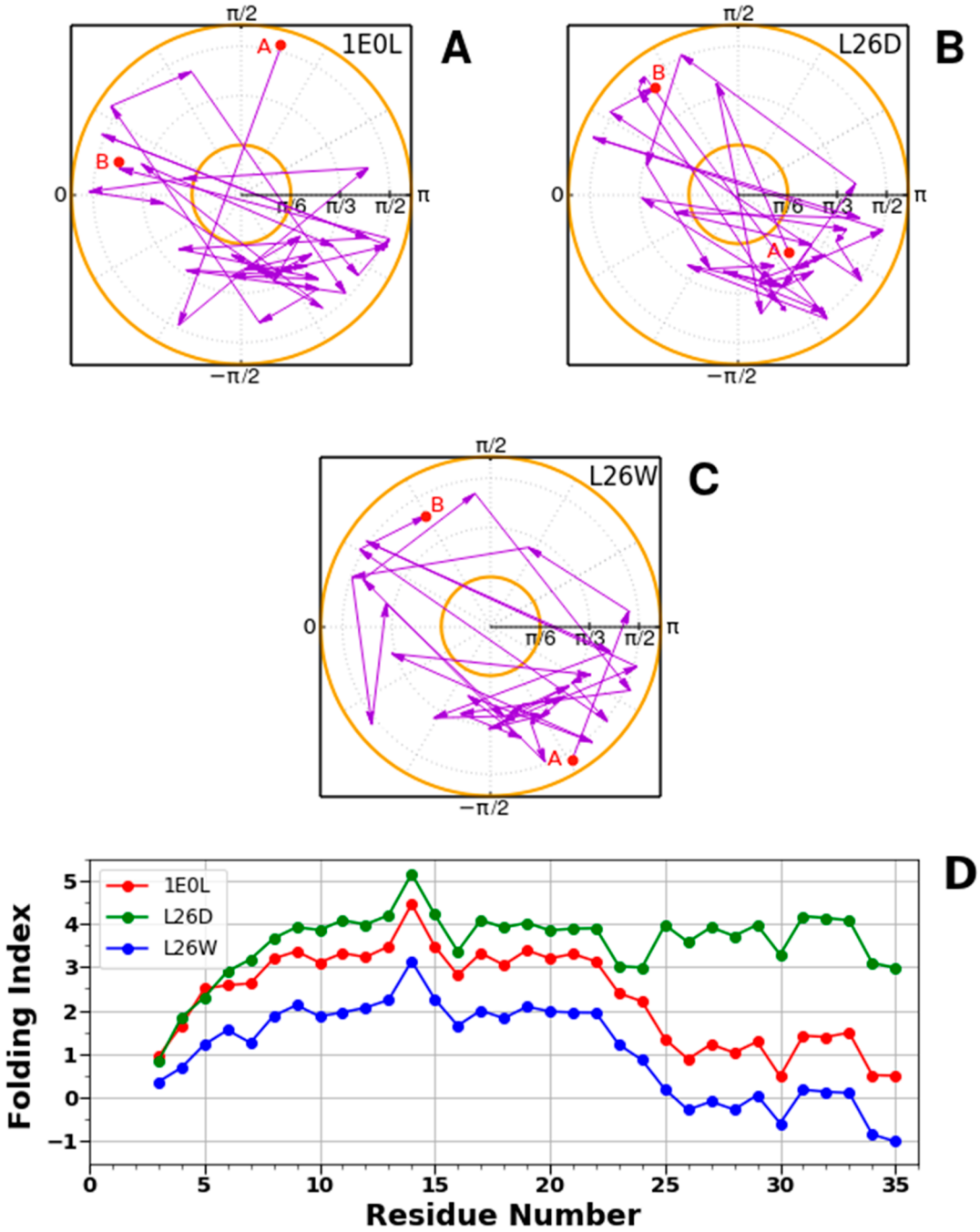


Figure 9. Folding index trajectories for the FBP28 WW domain (A), L26D (B), and L26W (C) and the folding index for the FBP28 WW domain, L26D and L26W (D).

[eqs 18 and 19] along the trajectories for the FBP28 WW domain (A), L26D (B) and L26W (C), together with the folding index dependence on length of proteins (D).

We can see that the folding index does not capture all the differences between the FBP28 WW domain, L26D, and L26W, but it indicates the most crucial sites that are involved in alteration of folding scenarios. In particular, (i) the main difference between the folding indices of L26D and the FBP28 WW domain and L26W is observed in the region of the 22nd and 23rd residues, which is associated with the speed-up of formation of turn 2; (ii) the folding index of L26W behaves differently in the region of the sixth and seventh residues, which is related to possible destabilization of the delocalized hydrophobic core and, also, obtains negative values in the region of the third β-strand.

CONCLUSIONS

The folding, misfolding, nonfolding, and energetics of two mutants (L26D and L26W) of the FBP28 WW domain were examined in terms of the backbone virtual-bond angle θ and backbone virtual-bond-dihedral angle γ of each residue, the *local* free energies, as well as with principal components. It was shown that the "behavior" of θ and γ angles in the course of time provides full information not only about the formation of β-strands but also about folding and reasons for misfolding and nonfolding. Strong correlations between the changes in θ and γ

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872

angles and *local* free energies over the trajectories were observed. The formation of β-strands was accompanied by the emergence of kinks in internal coordinate space and by the decrease of *local* free energy. In the downhill folding trajectory, the *local* free-energy decrease was ∼7 kcal/mol, which falls in the range of free energy that is enough for the correct folding of a protein. In the three-state folding trajectory, the *local* free-energy decrease was much higher because of the folding mechanism (hydrophobic collapse mechanism), which caused misfolding of the protein first. It was illustrated that the oscillations of the free energies of the β-strand segments observed in the unfolded state of the three-state folding trajectory (third β-strand) and in the entire nonfolding trajectory might be caused by instability of kinks along the backbone prompting a protein to appear "disordered". It was, also, shown that the local (internal coordinates) and global (principal components) motions in folding trajectories, as well as the kinks and principal components in nonfolding trajectory, are correlated.

Moreover, experimental structures of the FBP28 WW domain, L26D and L26W, were analyzed in terms of kinks. It was found that, without MD simulations, the kinks are able to capture the sites of protein that govern L26D and L26W to change the folding scenario. This feature makes kinks a very effective method not only for better understanding protein folding but also that would open a new door to the understanding of intrinsically disordered proteins, which have been implicated in a number of human diseases.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.0c00628.

The free-energy profiles as functions of RMSD at different temperatures for L26D, L26W, and FBP28 WW domain. The experimental and three calculated representative structures for the FBP28 WW domain, L26D and L26W (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Gia G. Maisuradze − *Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca 14853-1301, New York, United States;* orcid.org/0000-0002-1901-8433; Phone: 1-607-255-4399; Email: gm56@cornell.edu

Authors

Khatuna Kachlishvili − *Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca 14853-1301, New York, United States*

Anatolii Korneev − *Laboratory of Physics of Living Matter, Far Eastern Federal University, Vladivostok 690950, Russia*

Luka Maisuradze − *Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca 14853-1301, New York, United States; Biochemistry, Quantitative Biology, Biophysics, and Structural Biology (BQBS) Track, Yale University, New Haven 06520-8084, Connecticut, United States*

Jiaojiao Liu − *School of Physics, Beijing Institute of Technology, Beijing 100081, P. R. China*

Harold A. Scheraga − *Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca 14853-1301, New York, United States;* orcid.org/0000-0002-6314-5376

Alexander Molochkov − *Laboratory of Physics of Living Matter, Far Eastern Federal University, Vladivostok 690950, Russia*

Patrick Senet − *Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca 14853-1301, New York, United States; Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS, Univ. de Bourgogne Franche-Comté, Dijon Cedex F-21078, France;* orcid.org/0000-0002-2339-0019

Antti J. Niemi − *Laboratory of Physics of Living Matter, Far Eastern Federal University, Vladivostok 690950, Russia; School of Physics, Beijing Institute of Technology, Beijing 100081, P. R. China; Laboratoire de Mathematiques et Physique Theorique CNRS UMR 6083, Fédération Denis Poisson, Université de Tours, Parc de Grandmont, Tours F37200, France; Nordita, Stockholm University, Stockholm SE-106 91, Sweden*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.0c00628

Author Contributions

G.G.M. designed research; K.K., A.K., L.M., and G.G.M. performed research; K.K., A.K., L.M., J.L., A.M., P.S., A.J.N., and G.G.M. analyzed data; and K.K., A.K., L.M., H.A.S., A.M., P.S., A.J.N., and G.G.M. wrote the paper.

Author Contributions

$^{\nabla}$K.K., A.K., L.M., and G.G.M. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was (i) supported by a grant from the National Institutes of Health (GM-14312) and was (ii) carried out within the state assignment of the Ministry of Science and Higher Education of Russia (Grant No. 0657-2020-0015); (iii) the work by A.J.N. has been supported by the Carl Trygger Foundation, by the Swedish Research Council under Contract No. 2018-04411, and by COST Action CA17139; (iv) P.S. acknowledges support from the Bourgogne Franche-Comté Graduate School EUR-EIPHI (17-EURE-0002). This research was conducted by using the resources of (i) our 588-processor Beowulf cluster at the Baker Laboratory of Chemistry and Chemical Biology, Cornell University, (ii) the Informatics Center of the Metropolitan Academic Network in Gdansk, and (iii) the computing cluster Vostok-1 of Far Eastern Federal University.

■ REFERENCES

(1) Privalov, P. L.; Khechinashvili, N. N. A thermodynamic approach to the problem of stabilization of globular protein structure: A calorimetric study. *J. Mol. Biol.* **1974**, *86*, 665−684.

(2) Thirumalai, D. From minimal models to real proteins: Time scales for protein folding kinetics. *J. Phys. I* **1995**, *5*, 1457−1467.

(3) Akmal, A.; Munoz, V. The nature of the free energy barriers to two-state folding. *Proteins: Struct., Funct., Genet.* **2004**, *57*, 142−152.

(4) Garcia-Mira, M. M.; Sadqi, M.; Fischer, N.; Sanchez-Ruiz, J. M.; Munoz, V. Experimental identification of downhill protein folding. *Science* **2002**, *298*, 2191−2195.

(5) Nguyen, H.; Jäger, M.; Moretto, A.; Gruebele, M.; Kelly, J. W. Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3948−3953.

(6) Munoz, V.; Sanchez-Ruiz, J. M. Exploring protein-folding ensembles: A variable-barrier model for the analysis of equilibrium unfolding experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 17646−17651.
(7) Li, P.; Oliva, F. Y.; Naganathan, A.; Munoz, V. Dynamics of one-state downhill protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 103−108.
(8) Jha, S. K.; Dhar, D.; Krishnamoorthy, G.; Udgaonkar, J. B. Continuous dissolution of structure during the unfolding of a small protein. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11113−11118.
(9) Poland, D. C.; Scheraga, H. A. Statistical mechanics of non covalent bonds in polyamino acids. IX. The two state theory of protein denaturation. *Biopolymers* **1965**, *3*, 401−419.
(10) Knott, M.; Chan, H. S. Criteria for downhill protein folding: calorimetry, chevron plot, kinetic relaxation, and single-molecule radius of gyration in chain models with subdued degrees of cooperativity. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 373−391.
(11) Munoz, V. Thermodynamics and kinetics of downhill protein folding investigated with a simple statistical mechanical model. *Int. J. Quantum Chem.* **2002**, *90*, 1522−1528.
(12) Sadqi, M.; Fushman, D.; Muñoz, V. Atom-by-atom analysis of global downhill protein folding. *Nature* **2006**, *442*, 317−321.
(13) Naganathan, A. N.; Doshi, U.; Fung, A.; Sadqi, M.; Munoz, V. Dynamics, energetics, and structure in protein folding. *Biochemistry* **2006**, *45*, 8466−8475.
(14) Wildegger, G.; Kiefhaber, T. Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *J. Mol. Biol.* **1997**, *270*, 294−304.
(15) Kubelka, J.; Henry, E. R.; Cellmer, T.; Hofrichter, J.; Eaton, W. A. Chemical, physical, and theoretical kinetics of an ultrafast folding protein. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 18655−18662.
(16) Zhou, R.; Maisuradze, G. G.; Suñol, D.; Todorovski, T.; Macias, M. J.; Xiao, Y.; Scheraga, H. A.; Czaplewski, C.; Liwo, A. Folding kinetics of WW domains with the united residue force field for bridging microscopic motions and experimental measurements. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 18243−18248.
(17) Szczepaniak, M.; Iglesias-Bexiga, M.; Cerminara, M.; Sadqi, M.; Sanchez de Medina, C.; Martinez, J. C.; Luque, I.; Muñoz, V. Ultrafast folding kinetics of WW domains reveal how the amino acid sequence determines the speed limit to protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 8137−8142.
(18) Guijarro, J. I.; Sunde, M.; Jones, J. A.; Campbell, I. D.; Dobson, C. M. Amyloid fibril formation by an SH3 domain. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 4224−4228.
(19) Ramirez-Alvarado, M.; Merkel, J. S.; Regan, L. A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8979−8984.
(20) Ferguson, N.; Berriman, J.; Petrovich, M.; Sharpe, T. D.; Finch, J. T.; Fersht, A. R. Rapid amyloid fiber formation from the fast-folding WW domain FBP28. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 9814−9819.
(21) Mu, Y.; Nordenskiold, L.; Tam, J. P. Folding, misfolding, and amyloid protofibril formation of WW domain FBP28. *Biophys. J.* **2006**, *90*, 3983−3992.
(22) Neudecker, P.; Robustelli, P.; Cavalli, A.; Walsh, P.; Lundström, P.; Zarrine-Afsar, A.; Sharpe, S.; Vendruscolo, M.; Kay, L. E. Structure of an intermediate state in protein folding and aggregation. *Science* **2012**, *336*, 362−366.
(23) Alzheimer's Association. 2016 Alzheimer's disease facts and figures. *Alzheimer's Dementia* **2016**, *12*, 459−509.
(24) Kalia, L. V.; Lang, A. E. Parkinson's disease. *Lancet* **2015**, *386*, 896−912.
(25) Roos, R. A. Huntington's disease: a clinical review. *Orphanet J. Rare Dis.* **2010**, *5*, 40.
(26) Niimi, Y.; Iwasaki, Y.; Umemura, T.; Tanaka, F.; Yoshida, M.; Hashizume, Y.; Kitamoto, T.; Hirayama, M.; Sobue, G. MM2-cortical-type sporadic Creutzfeldt-Jakob disease with early stage cerebral cortical pathology presenting with a rapidly progressive clinical course. *Neuropathology* **2008**, *0*, 645−651.
(27) Macias, M. J.; Gervais, V.; Civera, C.; Oschkinat, H. Structural analysis of WW domains and design of a WW prototype. *Nat. Struct. Biol.* **2000**, *7*, 375−379.
(28) Karanicolas, J.; Brooks, C. L., III The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: Lessons for protein design? *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3954−3959.
(29) Karanicolas, J.; Brooks, C. L., III Integrating folding kinetics and protein function: Biphasic kinetics and dual binding specificity in a WW domain. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 3432−3437.
(30) Liwo, A.; Khalili, M.; Scheraga, H. A. Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 2362−2367.
(31) Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. Principal component analysis for protein folding dynamics. *J. Mol. Biol.* **2009**, *385*, 312−329.
(32) Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. Relation between free energy landscapes of proteins and dynamics. *J. Chem. Theory Comput.* **2010**, *6*, 583−595.
(33) Maisuradze, G. G.; Senet, P.; Czaplewski, C.; Liwo, A.; Scheraga, H. A. Investigation of protein folding by coarse-grained molecular dynamics with the UNRES force field. *J. Phys. Chem. A* **2010**, *114*, 4471−4485.
(34) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. B.; Wriggers, W. Atomic-level characterization of the structural dynamics of proteins. *Science* **2010**, *330*, 341−346.
(35) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517−520.
(36) Piana, S.; Sarkar, K.; Lindorff-Larsen, K.; Guo, M.; Gruebele, M.; Shaw, D. E. Computational design and experimental testing of the fastest-folding β-sheet protein. *J. Mol. Biol.* **2011**, *405*, 43−48.
(37) a Beccara, S.; Škrbić, T.; Covino, R.; Faccioli, P. Dominant folding pathways of a WW domain. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 2330−2335.
(38) Maisuradze, G. G.; Zhou, R.; Liwo, A.; Xiao, Y.; Scheraga, H. A. Effects of mutation, truncation and temperature on the folding kinetics of a WW domain. *J. Mol. Biol.* **2012**, *420*, 350−365.
(39) Zhang, M.; Case, D. A.; Peng, J. W. Propagated perturbations from a peripheral mutation show interactions supporting WW domain thermostability. *Structure* **2018**, *26*, 1474−1485.
(40) Jäger, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M. The folding mechanism of a β-sheet: WW domain. *J. Mol. Biol.* **2001**, *311*, 373−393.
(41) Ferguson, N.; Johnson, C. M.; Macias, M.; Oschkinat, H.; Fersht, A. Ultrafast folding of WW domains without structured aromatic clusters in the denatured state. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 13002−13007.
(42) Petrovich, M.; Jonsson, A. L.; Ferguson, N.; Daggett, V.; Fersht, A. R. Φ-analysis at the experimental limits: mechanism of β-hairpin formation. *J. Mol. Biol.* **2006**, *360*, 865−881.
(43) Jager, M.; Zhang, Y.; Bieschke, J.; Nguyen, H.; Dendle, M.; Bowman, M. E.; Noel, J. P.; Gruebele, M.; Kelly, J. W. Structure-function-folding relationship in a WW domain. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 10648−10653.
(44) Dave, K.; Jäger, M.; Nguyen, H.; Kelly, J. W.; Gruebele, M. High-resolution mapping of the folding transition state of a WW domain. *J. Mol. Biol.* **2016**, *428*, 1617−1636.
(45) Hanazono, Y.; Takeda, K.; Miki, K. Structural studies of the N-terminal fragments of the WW domain: Insights into co-translational folding of a beta-sheet protein. *Sci. Rep.* **2016**, *6*, 34654.
(46) Zanetti-Polzi, L.; Davis, C. M.; Gruebele, M.; Dyer, R. B.; Amadei, A.; Daidone, I. Parallel folding pathways of Fip35 WW domain explained by infrared spectra and their computer simulation. *FEBS Lett.* **2017**, *591*, 3265−3275.
(47) Iglesias-Bexiga, M.; Szczepaniak, M.; Sańchez de Medina, C.; Cobos, E. S.; Godoy-Ruiz, R.; Martinez, J. C.; Muñoz, V.; Luque, I. Protein folding cooperativity and thermodynamic barriers of the

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872

simplest β-sheet fold: A survey of WW domains. *J. Phys. Chem. B* **2018**, *122*, 11058−11071.
(48) Davis, C. M.; Zanetti-Polzi, L.; Gruebele, M.; Amadei, A.; Dyer, R. B.; Daidone, I. A quantitative connection of experimental and simulated folding landscapes by vibrational spectroscopy. *Chem. Sci.* **2018**, *9*, 9002−9011.
(49) Dave, K.; Gasic, A. G.; Cheung, M. S.; Gruebele, M. Competition of individual domain folding with inter-domain interaction in WW domain engineered repeat proteins. *Phys. Chem. Chem. Phys.* **2019**, *21*, 24393−24405.
(50) Maisuradze, G. G.; Medina, J.; Kachlishvili, K.; Krupa, P.; Mozolewska, M. A.; Martin-Malpartida, P.; Maisuradze, L.; Macias, M. J.; Scheraga, H. A. Preventing fibril formation of a protein by selective mutation. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 13549−13554.
(51) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. *J. Chem. Phys.* **2001**, *115*, 2323−2347.
(52) Liwo, A.; Czaplewski, C.; Ołdziej, S.; Kozłowska, U.; Makowski, M.; Kałinowski, S.; Kazmierkiewicz, R.; Shen, H.; Maisuradze, G.; Scheraga, H. A. Optimization of the physics-based united-residue force field (UNRES) for protein folding simulations. In *NIC Series, NIC Symposium 2008;* Munster, G., Wolf, D., Kremer, M., Eds.; John von Neumann Institute for Computing (NIC): Jülich, Germany, 2008; Vol. *39*, pp 63−70.
(53) Matheson, R. R.; Scheraga, H. A. A method for predicting nucleation sites for protein folding based on hydrophobic contacts. *Macromolecules* **1978**, *11*, 819−829.
(54) Dill, K. A.; Fiebig, K. M.; Chan, H. S. Cooperativity in protein-folding kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 1942−1946.
(55) Dinner, A. R.; Lazaridis, T.; Karplus, M. Understanding beta-hairpin formation. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9068−9073.
(56) Kachlishvili, K.; Dave, K.; Gruebele, M.; Scheraga, H. A.; Maisuradze, G. G. Eliminating a protein folding intermediate by tuning a local hydrophobic contact. *J. Phys. Chem. B* **2017**, *121*, 3276−3284.
(57) Jolliffe, I. T. *Principal Component Analysis;* Springer: New York, NY, 2002.
(58) Mu, Y.; Nguyen, P. H.; Stock, G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins: Struct., Funct., Genet.* **2005**, *58*, 45−52.
(59) Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. How adequate are one- and two-dimensional free energy landscapes for protein folding dynamics? *Phys. Rev. Lett.* **2009**, *102*, 238102.
(60) Chernodub, M.; Hu, S.; Niemi, A. J. Topological solitons and folded proteins. *Phys. Rev. E* **2010**, *82*, 011916.
(61) Molkenthin, N.; Hu, S.; Niemi, A. J. Discrete nonlinear Schrödinger equation and polygonal solitons with applications to collapsed proteins. *Phys. Rev. Lett.* **2011**, *106*, 078102.
(62) Hu, S.; Krokhotin, A.; Niemi, A. J.; Peng, X. Towards quantitative classification of folded proteins in terms of elementary functions. *Phys. Rev. E* **2011**, *83*, 041907.
(63) Hu, S.; Lundgren, M.; Niemi, A. J. Discrete Frenet frame, inflection point solitons, and curve visualization with applications to folded proteins. *Phys. Rev. E* **2011**, *83*, 061908.
(64) Krokhotin, A.; Niemi, A. J.; Peng, X. Soliton concepts and protein structure. *Phys. Rev. E* **2012**, *85*, 031906.
(65) Krokhotin, A.; Liwo, A.; Niemi, A. J.; Scheraga, H. A. Coexistence of phases in a protein heterodimer. *J. Chem. Phys.* **2012**, *137*, 035101.
(66) Krokhotin, A.; Lundgren, M.; Niemi, A. J. Solitons and collapse in the λ-repressor protein. *Phys. Rev. E* **2012**, *86*, 021923.
(67) Krokhotin, A.; Liwo, A.; Maisuradze, G. G.; Niemi, A. J.; Scheraga, H. A. Kinks, loops, and protein folding with protein A as an example. *J. Chem. Phys.* **2014**, *140*, 025101.
(68) Molochkov, A.; Begun, A.; Niemi, A. J. Gauge symmetries and structure of proteins. *EPJ Web Conf.* **2017**, *137*, 04004.
(69) Begun, A.; Molochkov, A.; Niemi, A. J. Protein tertiary structure and the myoglobin phase diagram. *Sci. Rep.* **2019**, *9*, 10819.
(70) Faddeev, L. D.; Takhtajan, L. *Hamiltonian methods in the theory of solitons;* Springer-Verlag: Berlin, Germany, 1987.
(71) Ablowitz, M. J.; Prinardi, B.; Trubatch, A. *Discrete and continuous nonlinear Schrödinger systems;* Cambridge University Press: Cambridge, UK, 2004.
(72) Kevrekidis, P. *The Discrete nonlinear Schrödinger equation: mathematical analysis, numerical computations and physical perspectives;* Springer-Verlag: Berlin, Germany, 2009.
(73) Manton, N.; Sutcliffe, P. *Topological solitons;* Cambridge University Press: Cambridge, UK, 2004.
(74) Weinberg, S. *The quantum theory of fields;* Cambridge University Press: Cambridge, UK, 1995; Vol. *2*.
(75) Davydov, A. S. Solitons and energy transfer along protein molecules. *J. Theor. Biol.* **1977**, *66*, 379−387.
(76) Nicolaï, A.; Delarue, P.; Senet, P. Intrinsic localized modes in proteins. *Sci. Rep.* **2015**, *5*, 18128.
(77) Dauxois, T.; Peyrard, M. *Physics of Solitons;* Cambridge University Press: Cambridge, UK, 2006.
(78) Khalili, M.; Liwo, A.; Jagielska, A.; Scheraga, H. A. Molecular dynamics with the united-residue model of polypeptide chains. II. Langevin and Berendsen-bath dynamics and tests on model o-helical systems. *J. Phys. Chem. B* **2005**, *109*, 13798−13810.
(79) Widom, B. Surface tension and molecular correlations near the critical point. *J. Chem. Phys.* **1965**, *43*, 3892−3897.
(80) Kadanoff, L. P. Scaling laws for Ising models near T_c. *Physics* **1966**, *2*, 263−272.
(81) Wilson, K. Renormalization group and critical phenomena. I. Renormalization group and the Kadanoff scaling picture. *Phys. Rev. B* **1971**, *4*, 3174−3183.
(82) Fisher, M. E. The renormalization group in the theory of critical behavior. *Rev. Mod. Phys.* **1974**, *46*, 597−616.
(83) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(84) Khalili, M.; Liwo, A.; Rakowski, F.; Grochowski, P.; Scheraga, H. A. Molecular dynamics with the united-residue model of polypeptide chains. I. Lagrange equations of motion and tests of numerical stability in the microcanonical mode. *J. Phys. Chem. B* **2005**, *109*, 13785−13797.
(85) Lindquist, S. L.; Kelly, J. W. Chemical and biological approaches for adapting proteostasis to ameliorate protein misfolding and aggregation diseases - progress and prognosis. *Cold Spring Harbor Perspect. Biol.* **2011**, *3*, No. a004507.
(86) Berg, J.; Tymoczko, J. L.; Stryer, L. *Biochemistry,* 6th ed.; Freeman, W. H.: New York, 2007.
(87) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Intrinsically disordered proteins in human diseases: Introducing the D2 concept. *Annu. Rev. Biophys.* **2008**, *37*, 215−246.
(88) Kitao, A.; Hayward, S.; Gō, N. Energy landscape of a native protein: jumping-among-minima model. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 496−517.
(89) Niemi, A. J. WHAT IS LIFE - Sub-cellular physics of live matter. In *Topological Aspects of Condensed Matter Physics, Les Houches Summer School 103;* Oxford University Press, 2017.
(90) Krokhotin, A.; Lundgren, M.; Niemi, A. J.; Peng, X. Soliton driven relaxation dynamics and protein collapse in the villin headpiece. *J. Phys.: Condens. Matter* **2013**, *25*, 325103.
(91) Peng, X.; Sieradzan, A. K.; Niemi, A. J. Thermal unfolding of myoglobin in the Landau-Ginzburg-Wilson approach. *Phys. Rev. E* **2016**, *94*, 062405.

https://dx.doi.org/10.1021/acs.jpcb.0c00628
J. Phys. Chem. B 2020, 124, 3855−3872