JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Local vs Global Motions in Protein Folding

Gia G. Maisuradze,[†] Adam Liwo,[‡] Patrick Senet,[†,§] and Harold A. Scheraga*[,†]

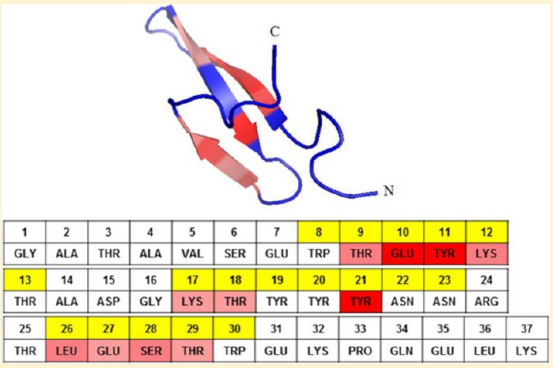
[†]Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, United States

[‡]Faculty of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland

[§]Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS-Université de Bourgogne, 9 Av. A. Savary, BP 47 870, F-21078 Dijon Cedex, France

Supporting Information

ABSTRACT: It is of interest to know whether local fluctuations in a polypeptide chain play any role in the mechanism by which the chain folds to the native structure of a protein. This question is addressed by analyzing folding and nonfolding trajectories of a protein; as an example, the analysis is applied to the 37-residue triple β-strand WW domain from the Formin binding protein 28 (FBP28) (PDB ID: 1E0L). Molecular dynamics (MD) trajectories were generated with the coarse-grained united-residue force field, and one- and two-dimensional free-energy landscapes (FELs) along the backbone virtual-bond angle θ and backbone virtual-bond-dihedral angle γ of each residue and principal components, respectively, were analyzed. The key residues involved in the folding of the FBP28 WW domain are elucidated by this analysis. The correlations between local and global motions are found. It is shown that most of the residues in the folding trajectories of the system studied here move in a concerted fashion, following the dynamics of the whole system. This demonstrates how the choice of a pathway has to involve concerted movements in order for this protein to fold. This finding also sheds light on the effectiveness of principal component analysis (PCA) for the description of the folding dynamics of the system studied. It is demonstrated that the FEL along the PCs, computed by considering only several critically-placed residues, can correctly describe the folding dynamics.



INTRODUCTION

In spite of long, assiduous experimental and theoretical studies, one of the most important problems of biophysics and biochemistry, how proteins reach their biologically active conformations, still remains unsolved. In order to understand the kinetics and thermodynamics of protein folding, it is important to know why proteins fold properly or do not reach their native state and what governs the way that proteins fold. From theoretical and conceptual points of view, it is well-known that a study of free-energy landscapes (FELs) holds the key to understand how proteins fold and function.[1−3] However, the selection of a correct model for protein folding kinetics and the coordinates along which the intrinsic folding pathways can be identified in order to interpret experimental data still remains challenging. Some progress to treat this problem has been made recently based on the analysis of molecular dynamics (MD) simulations of protein folding,[4−12] particularly, in the use of principal component analysis (PCA),[7−10] transition disconnectivity graphs (TRDG),[4−6] and network analysis.[6,12] This involves proper choices for the reaction coordinates and the treatment of progress variables. However, the question as to what governs the way in which proteins fold was not addressed in those studies.

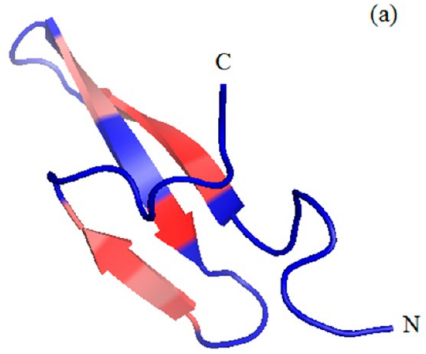
It is well established that such important questions about protein folding can be answered by examining the behavior of small model proteins and peptides. Therefore, in the present study, in order to find the reasons why a protein may or not fold properly to its native state, we analyzed MD trajectories of a particular 37-residue protein as an example, namely, the triple-β-stranded WW domain from the Formin binding protein 28 (FBP28) (PDB ID: 1E0L)[13] (Figure 1), generated with the coarse-grained united-residue (UNRES) force field.[10,14−21]

The FBP28 WW domain is a member of the WW domain family. The WW domains have been the subject of extensive theoretical[7,9,10,22−30] and experimental[31−36] studies because of their small size, biological importance,[37] and interesting fast-folding kinetics. From the folding point of view, the FBP28 WW domain is a very interesting system to study because of its biphasic folding kinetics[22,23,33] [i.e., coexistence of slow (three-state) and fast (two-state) phases]. Also, there is an interesting dilemma offered by β-sheets regarding the mechanism controlling the folding rate. Previous experimental and theoretical studies have shown that either local interactions at the turns[38,39] or the interstrand interactions of hydrophobic residues control the folding,[40,41] although some studies illustrated a close competition between these mechanisms.[42]

It should be noted that, in spite of the biological importance of the WW domain, our interest in this work is in the folding

Received: February 28, 2013

   dx.doi.org/10.1021/ct4001558 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX



(b)

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	LEU	GLU	SER	THR	TRP	GLU	LYS	PRO	GLN	GLU	LEU	LYS

Figure 1. Experimental NMR-derived structure of the triple β-strand WW domain from the Formin binding protein 28 (FBP28) (1E0L) [panel (a)]. Amino acid sequence of 1E0L [panel (b)]. The numbers of residues forming the β-strands are highlighted in yellow color. The residues contributing the most to the mean-square-fluctuations are highlighted in red in both panels. Depending on the percentage of the contribution, the color varies from light red (~5−8%) to dark red (~10% and higher).

problem per se (therefore, a new approach for the analysis of MD trajectories of protein folding is presented) and not specifically in the properties of the WW domain protein. Recently, we studied[30] not only the wild type FBP28 WW domain but also its mutants and truncated mutants at different temperatures and widely discussed the agreements and discrepancies of our results with experiment.[31−34] Here, several important aspects of folding (folding time, folding-transition temperature, folding scenario, the behavior of hairpins, and the role of loops in folding) are compared with experiment (see Materials and Methods and Results and Discussion sections).

We address the protein folding of the FBP28 WW domain by carrying out 10 canonical UNRES MD trajectories at 330 K, which have been grouped in three categories: fast-folding [3 MD trajectories in which this protein folds within the first 40 ns (corresponding to about 40 μs of actual time because the time scale is reduced by about 1000 in UNRES MD simulations[43])], slow-folding (4 MD trajectories in which this protein folds after 40 ns), and nonfolding (3 MD trajectories in which this protein never folds during the entire MD simulation) (Figure S1 in the Supporting Information). All trajectories start with the same initial (extended) structure but with different velocities. From an analysis as to why different initial velocities of the MD simulations lead to proper or improper folding pathways, light can be shed on the mechanisms leading to folding or misfolding of a protein in solution.

There is strong evidence[44−47] that protein folding is initiated and governed by local motions of the residues which, subsequently, result in global conformational changes. Therefore, in the present study, the folding and nonfolding trajectories of the FBP28 WW domain were analyzed in terms of the relation between the local dynamics of each residue and the global dynamics of the entire system. This local and global analysis enabled us to determine which residues play an important role in folding of the FBP28 WW domain. We also answered questions regarding the effectiveness of PCA, and the role of formation of loops and β-strands, for folding and nonfolding trajectories of the FBP28 WW domain.

The paper is organized as follows. The Materials and Methods section contains descriptions of the UNRES model, details of MD simulations, and the PCA method. The Results and Discussion section contains four subsections. In the first two subsections, the free-energy profiles (FEPs) along the internal coordinates of the backbone, for folding and nonfolding MD trajectories, respectively, are presented and discussed to identify the key residues participating in successful folding of the FBP28 WW domain. The distances, as a function of time, and the FEPs along the distances between selected pairs of residues for folding and nonfolding MD trajectories are illustrated and discussed in the third subsection. The fourth subsection answers the question as to how reaction coordinates for folding the FBP28 WW domain can be built from PCA using a relevant subset of internal coordinates of the polypeptide. A summary and conclusions are given in the Conclusions section.

MATERIALS AND METHODS

UNRES Model and Simulation Details. The UNRES model of polypeptide chains[10,17,48] is illustrated in Figure 2. A polypeptide chain is represented as a sequence of α-carbon (C^α) atoms linked by virtual $C^\alpha\cdots C^\alpha$ bonds with united peptide groups halfway between the neighboring C^α's, and united side chains, whose sizes depend on the nature of the amino acid residues, attached to the respective C^α's by virtual $C^\alpha\cdots$SC bonds. The effective energy is expressed by eq 1[21]

$$\begin{aligned} U = {} & w_{SC}\sum_{i<j} U_{SC_iSC_j} + w_{SCp}\sum_{i\neq j} U_{SC_ip_j} \\ & + w_{pp}f_2(T)\sum_{i<j-1} U_{p_ip_j} + w_{tor}f_2(T)\sum_i U_{tor}(\gamma_i) \\ & + w_{tord}f_3(T)\sum_i U_{tord}(\gamma_i, \gamma_{i+1}) + w_b\sum_i U_b(\theta_i) \\ & + w_{rot}\sum_i U_{rot}(\alpha_{SC_i}, \beta_{SC_i}, \theta_i) + w_{bond}\sum_i U_{bond}(d_i) \\ & + \sum_{m=3}^{6} w_{corr}^{(m)} f_m(T) U_{corr}^{(m)} + w_{SS}\sum_i U_{SS;i} \end{aligned} \tag{1}$$

with[21]

$$f_m(T) = \frac{\ln(e + e^{-1})}{\ln\left\{\exp\left[\left(\frac{T}{T_0}\right)^{m-1}\right] + \exp\left[-\left(\frac{T}{T_0}\right)^{m-1}\right]\right\}}; \quad T_0 = 300\ \text{K} \tag{2}$$

where the successive terms represent side chain−side chain, side chain−peptide, peptide−peptide, torsional, double-torsional, bond-angle bending,[16] side-chain local (dependent on the angles α and β of Figure 2), distortion of virtual bonds, multibody (correlation) interactions, and formation of disulfide bonds, respectively. The w's are the relative weights of each term. The correlation terms arise from a cumulant expansion[17,49] of the restricted free energy function of the simplified chain obtained from the all-atom energy surface by integrating out the secondary degrees of freedom. The temperature-dependent factors of eq 2,

dx.doi.org/10.1021/ct4001558 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

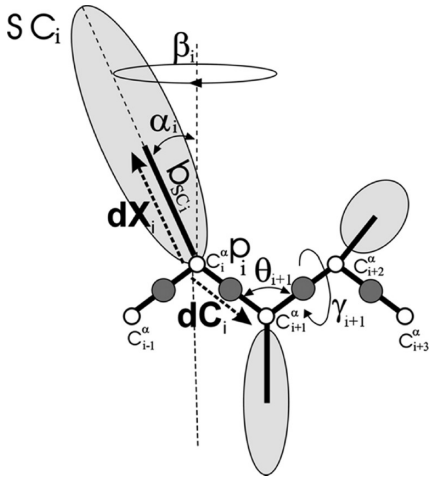


Figure 2. The UNRES model of polypeptide chains. The interaction sites are peptide-bond centers (p) and side-chain ellipsoids of different sizes (SC) attached to the corresponding α-carbons with different "bond lengths", b_{SC}. The α-carbon atoms are represented by small open circles. The equilibrium 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 $\mathbf{dX}_i$ ($C^{\alpha}_i...SC_i$), $i = 2, 3,..., N-1$ (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}.

introduced in our recent work[21] and discussed further in ref 50, reflect the fact that the UNRES effective energy is an approximate cumulant expansion of the restricted free energy. The virtual-bond vectors are the variables used in molecular dynamics.

We carried out canonical MD runs at 330 K (10 trajectories), with the force field parametrized[51] on the β-strand protein 1E0L and the α-helical protein 1ENH. The Berendsen thermostat[52] was used to maintain constant temperature. The trajectories selected for detailed analysis corresponded to the near folding-transition temperature, T_f = 339 K[51] for 1E0L (experimental T_f = 337 K[33]), at which the protein either exhibits quite a stable native state and, once the system folds, it remains in the native state until the end of the simulation or the native state is still stable, but only a few folding/unfolding events are observed. The time step in molecular dynamics simulations was δt = 0.1 mtu (1 mtu = 48.9 fs is the "natural" time unit of molecular dynamics[53]), and the coupling parameter of the Berendsen thermostat was τ = 1 mtu. A total of 120,000,000 steps (about 0.6 μs of MD time) were run for each trajectory. It should be noted that the experimental folding time of the FBP28 WW domain is ~30 μs,[33] while the presented fast-folding trajectory folds in ~37 ns. However, in order to compare the experimental and UNRES time, the latter should be multiplied by ~1000 because the fast degrees of freedom are averaged out.[53] After taking this extension of the time scale into account, it can be seen that the experimental and computational folding times are comparable.

Principal Component Analysis in Internal Coordinates. In the PCA method in internal coordinates,[7−10,54−56] the correlated internal motions of a protein are expressed quantitatively by the covariance matrix C_{ij} [eq 3] of the $2 \times (N-2)$ variables [eq 4a] representing the coarse-grained angles θ and of the $2 \times (N-3)$ variables [eq 4b] representing the coarse-grained angles γ (see Figure 2 for the definitions of the angles)

$$C_{ij} = \langle (u_i - \langle u_i \rangle)(u_j - \langle u_j \rangle) \rangle \tag{3}$$

where <...> denotes the average over all sampled conformations in the trajectory with i and j varying between 2 and $2 \times (2N-5)$ with N being the number of residues ($N > 3$). The variables u in eq 3 are defined by

$$\{u_{4i-7};\ u_{4i-6}\} = \{\cos[\theta_i(t)],\ \sin[\theta_i(t)]\} \tag{4a}$$

for i = 2 to $N-1$ and by

$$\{u_{4i-5};\ u_{4i-4}\} = \{\cos[\gamma_i(t)],\ \sin[\gamma_i(t)]\} \tag{4b}$$

for i = 2 to $N-2$.

The diagonalization of the covariance matrix C_{ij} leads to $2 \times (2N-5)$ eigenvalues λ^k, ordered by decreasing value: $\lambda^1 > \lambda^2 > ... > \lambda^{2X(2N-5)}$, and eigenvectors $\boldsymbol{e}^k = \{\boldsymbol{e_1}^k, \boldsymbol{e_2}^k, ..., \boldsymbol{e_{2N-5}}^k\}^T$, where T is the transpose, with $\boldsymbol{e}_i^k = \{\boldsymbol{e}_i^k(1); \boldsymbol{e}_i^k(2)\}$ where $\boldsymbol{e}_i^k(n)$ is the nth component of the projection of the eigenvector $\boldsymbol{e}^k$ on the ith internal coordinate. The contribution of each angle i to a mode k is the so-called influence:

$$\nu_i^k = [(e_i^k(1))^2 + (e_i^k(2))^2] \tag{5}$$

The total mean-square fluctuation MSF of the coarse-grained internal coordinates can be decomposed into modes:

$$MSF = \sum_i \langle (u_i - \langle u_i \rangle)^2 \rangle = \sum_k \sum_i \lambda^k \nu_i^k \tag{6}$$

The eigenmodes with the largest eigenvalues λ^k correspond to the collective modes contributing the most to the MSF of the protein [see eq 6]. The projection of the trajectory on the eigenvector $\boldsymbol{e}^k$ is named the principal component q_k.

RESULTS AND DISCUSSION

We studied two types of trajectories of the FBP28 WW domain: folding (i.e., in which the protein folds during the trajectory) and nonfolding (i.e., in which the protein never folds during the entire MD simulation). In order to understand the folding mechanism of the FBP28 WW domain protein, we studied the local motions of each residue along the sequence. In particular, we investigated the FEPs along the backbone virtual-bond angle θ and backbone virtual-bond-dihedral angle γ, of each residue (defined in Figure 2). The quantity θ_i for residue i is the angle formed by the vectors (virtual bonds) joining three successive C^{α} atoms ($i-1$, i, $i+1$) along the primary sequence. The first angle along the sequence is θ_2 and the last is θ_{N-1} in which N is the total number of residues. The quantity γ_i for residue n is the dihedral angle formed by the vectors (virtual bonds) joining four successive C^{α} atoms ($i-1$, i, $i+1$, and $i+2$) along the primary sequence. The first dihedral angle along the sequence is γ_2 and the last is γ_{N-2}. This approach along the backbone virtual-bond-dihedral angles γ_i was used previously for describing how a single native protein diffuses on its free-energy landscape.[57−59]

It should be noted that, although the FEPs along the θ_i and γ_i angles of the entire trajectory [$\mu(\theta) = -k_B T \ln P(\theta)$, $\mu(\gamma) = -k_B T \ln P(\gamma)$, where P, T, and k_B are the probability distribution function (pdf), the absolute temperature, and the Boltzmann constant, respectively] are very helpful to identify the key residues in the folding process,[10] the analysis of the FEPs of the entire trajectory does not provide information about the way in which each residue explores its own FEP in the course of time nor to what extent the motion of each residue is coupled to the global motion of the protein as it proceeds toward its native state. To answer these questions, we have selected one folding trajectory

dx.doi.org/10.1021/ct4001558 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

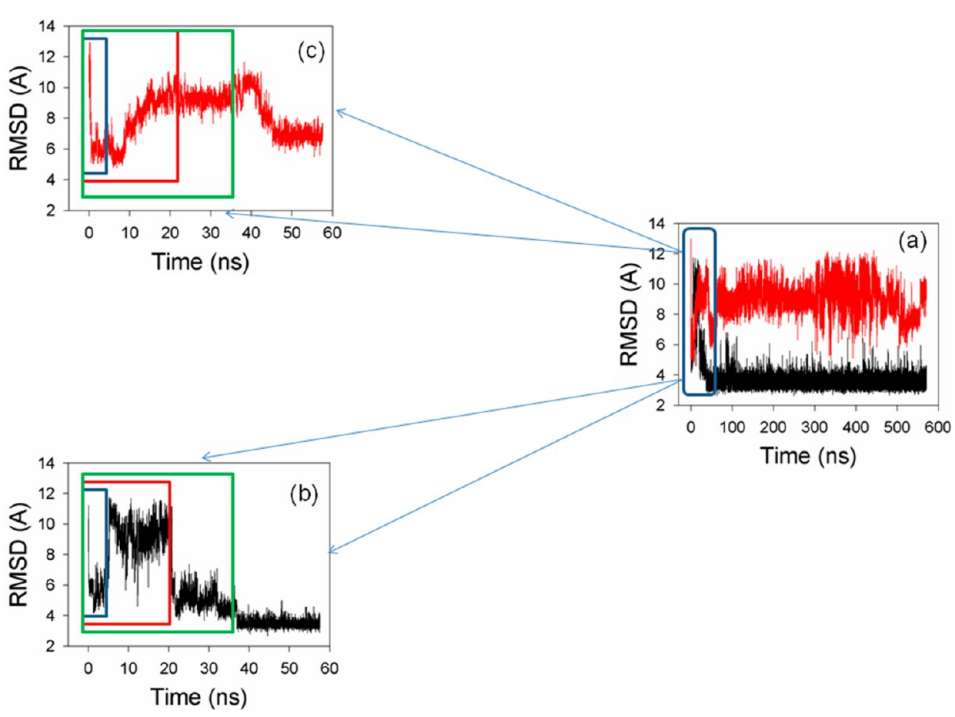


Figure 3. RMSDs vs time for fast-folding (black) and nonfolding (red) MD trajectories (panel a); RMSDs of the first 58 ns period of time of folding [panel (b)] and nonfolding [panel (c)] trajectories. The blue, red, and green rectangles in panels (b) and (c) correspond to 4.5 ns, 21 ns, and 37 ns time periods, respectively.

and calculated FEPs along the θ_i and γ_i angles for certain periods of time in the non-native state, during which significant structural changes occur before the protein reaches its native state. In order to find out which residues play a crucial role in folding, we compared these FEPs to those computed for one nonfolding trajectory in the same time-periods in which the FBP28 WW domain makes noticeable jumps although it never folds. It should be noted that the selection of time periods does not influence the FEPs of the nonfolding trajectory significantly, as shown below in the "FEPs along θ_i and γ_i angles of the nonfolding trajectory" subsection.

It should be noted that the FEPs presented here are effective FEPs because they are computed from a nonequilibrium probability density and depend on the time duration and on the initial conditions of the trajectory. The effective FEP differs from the actual FEP, which is an equilibrium thermodynamic property and should be computed from the entire sets of trajectories (folding and nonfolding). Because of the dependence of the effective FEP on the time duration of the trajectory and on the initial conditions, we use the effective FEP to analyze in detail the MD trajectories and extract the reasons why a protein folds or not in a single MD trajectory.

Figure 3 shows how the periods of time, over which the FEPs were calculated, were selected. Panel (a) illustrates the RMSD as a function of time of folding (black) and nonfolding (red) MD trajectories. The first ∼58 ns of both trajectories were selected and expanded [panels (b) and (c)]. In the first 37 ns, the FBP28 WW domain remains unfolded in the folding trajectory, forming unfolded and intermediate states, and subsequently jumps into the native state ($t > 37$ ns). Based on significant changes in the RMSD of the folding trajectory [panel (b)], three time-periods were selected to calculate the FEPs. These are as follows: ∼4.5 ns (blue rectangle), ∼21 ns (red rectangle), and ∼37 ns (green rectangle).

FEPs along θ_i and γ_i Angles of the Folding Trajectory. Figure 4 shows the FEPs along the θ_i and γ_i angles computed from the folding [Figures 4(a) and 4(b), respectively] and nonfolding [Figures 4(c) and 4(d), respectively] MD trajectories. The blue, red, green, and black curves correspond to the FEPs calculated over 4.5 ns, 21 ns, and 37 ns and over the entire duration of the trajectories, respectively. It should be noted that, in order to avoid some confusion caused by the deep minimum in the neighborhood of 180°, the range of the fluctuations of the γ_i angles, which is between [−180°; 180°], was shifted to [0°; 360°] in Figures 4(b) and 4(d).

Two main (representative) shapes in the FEPs computed over the entire folding trajectory [black lines in Figure 4(a)] can easily be distinguished along the sequence of the θ_i angles. The first shape, with one deep minimum at ∼90°, represents FEPs of loop and end residues (FEPs with black numbers); the second shape, with a broad, deep minimum at ∼120°−130° and a narrow (less deep) local minimum at ∼90°, represents FEPs of β-strand residues (FEPs with red numbers); the shapes of the FEPs along the θ_i angles, which include residues from both loops and β-strands, or ends and β-strands (FEPs with green numbers) either have a "transition" shape between the representative shapes for the loops and β-strands (θ_i, i = 8, 31) or resemble one of them [i.e., the FEPs along θ_i, i = 7, 13, 14, 16, 23−25, 30 resemble the representative shape of loop residues, and the forms of the FEPs along θ_i, i = 17, 26 resemble the representative shape of the β-strand residues]. Only a few FEPs along the θ_i angles (θ_i, i = 22, 32, 33) have forms that are different from their representative shapes.

Based on the FEPs along the θ_i angles of the selected time-periods, during which the whole system remains in the non-native state [blue, red, and green lines in Figure 4(a)], most of the θ angles within the first and middle β-strands (θ_i, i = 9−12, 18−21) completely explore the narrow local minimum at ∼90°. These θ angles jump to their native minimum, as does the whole system simultaneously, and stay in the deep global minimum at ∼120°−130° during the rest of the trajectory [black lines in Figure 4(a)]. It should be noted that, in the non-native state,

dx.doi.org/10.1021/ct4001558 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

these θ angles not only explore the local minimum but also spend some time (less than in the local minimum for θ_i, $i = 9, 10, 18$; and more than in the local minimum for θ_i, $i = 11, 12, 19-21$) in the global minimum [blue, red, and green lines in Figure 4(a)]. The θ angles of the third β-strand (θ_i, $i = 27-29$) partially explore the narrow local minimum at ~90°, while the protein is in the

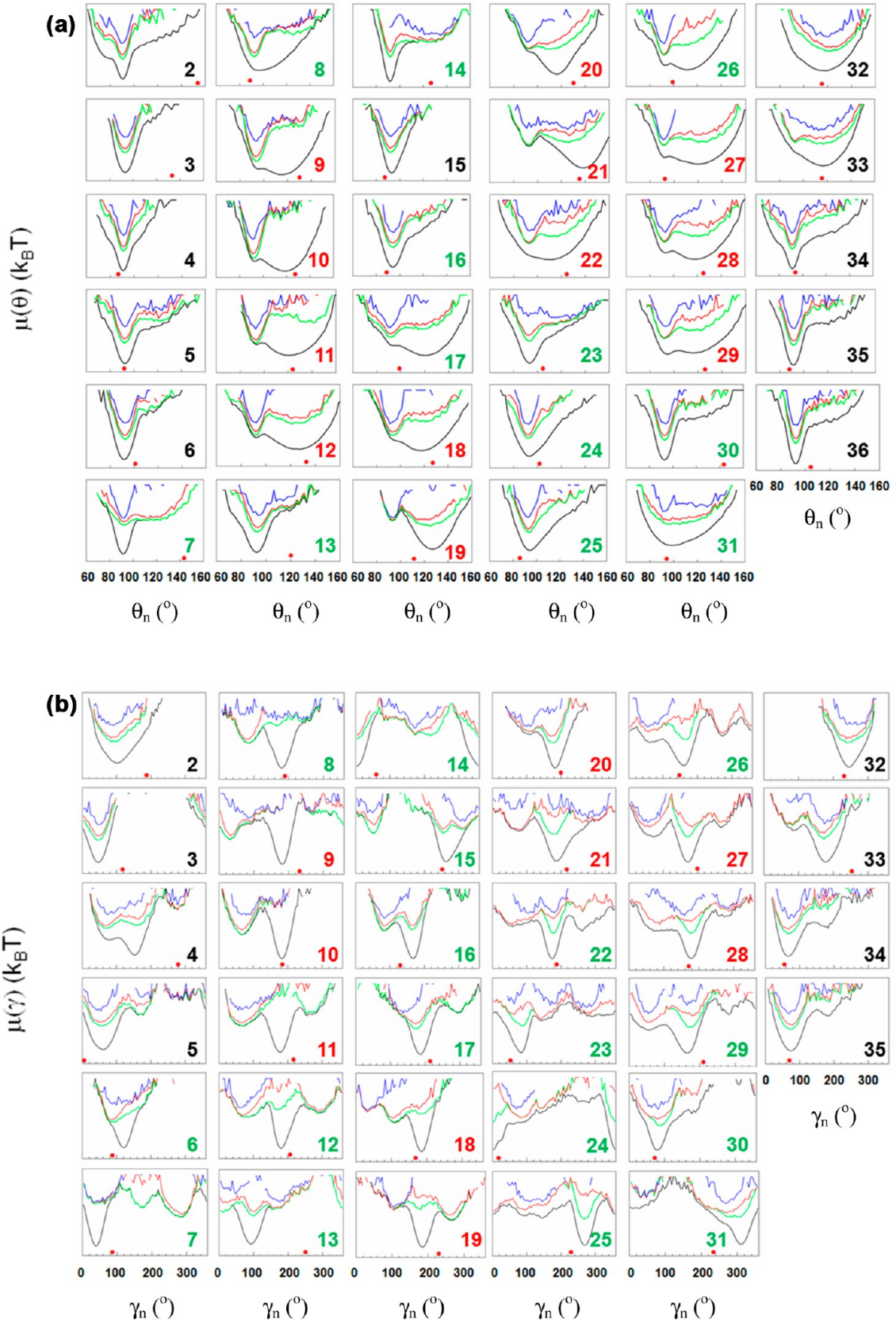


Figure 4. continued

dx.doi.org/10.1021/ct4001558 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

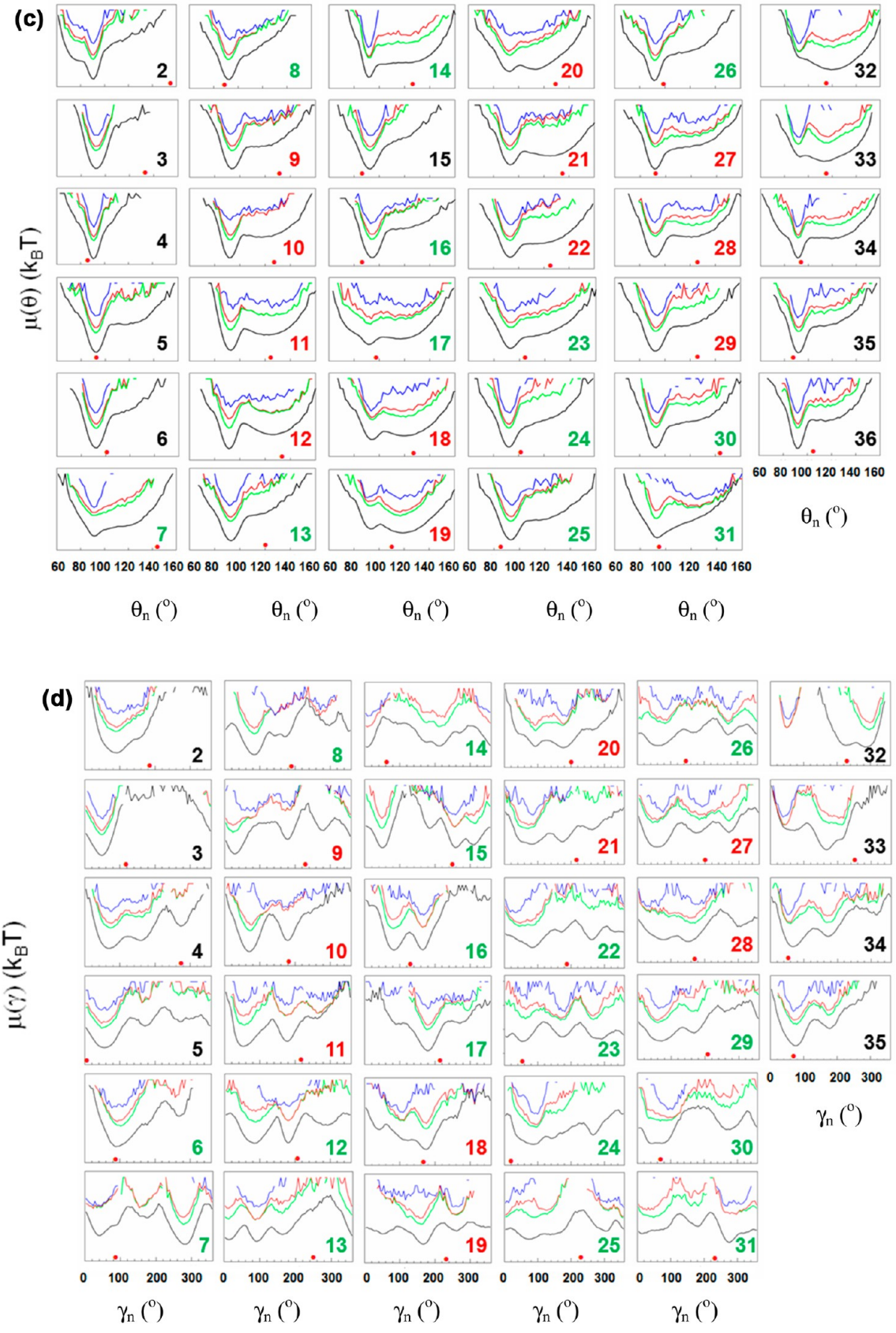


Figure 4. Free-energy profiles (FEPs), $\mu(\theta)$ and $\mu(\gamma)$, along the θ and γ angles for folding [panels (a) and (b), respectively] and nonfolding [panels (c) and (d), respectively] MD trajectories of the FBP28 WW domain. Blue, red, green, and black curves correspond to FEPs computed over 4.5 ns, 21 ns, 37 ns, and over the entire MD trajectories, respectively. The black numbers pertain to FEPs along the θ and γ angles which include only residues of the loops and ends; the red numbers pertain to FEPs along the θ and γ angles which include only residues of β-strands; the green numbers pertain to FEPs along θ and γ angles which include the residues from both loop and β-strands or from both ends and β-strands. The NMR-derived structural data (small red circles at the bottom of each panel) are computed from the first model of the PDB ID code 1E0L.[13]

dx.doi.org/10.1021/ct4001558 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

non-native state. After the protein jumps to the native state [shown by the decrease of the RMSD (Figure 3(b)], the θ angles of the third β-strand start to interconvert between the local (90°) and global (∼120°−130°) minima. Like the θ angles of the first and middle β-strands, the θ angles of the third β-strand also start exploring the native minimum in the non-native state [blue, red, and green lines in Figure 4(a)]. Moreover, the θ angles of the first β-strand edges (θ_i, i = 7, 13, 14), the N-terminus (θ_i, i = 5, 6), and some θ angles in the neighborhood of the middle β-strand (θ_i, i = 17, 23, 26) either completely or partially explore their local minima or "shoulder" before the protein jumps to the native state. The rest of the θ angles (θ_i, i = 2−4, 8, 15, 16, 22, 24, 25, 30−36) gradually explore their own minima during the entire trajectory and, hence, do not follow changes in the dynamics of the whole system.

A more detailed, quantitative representation of the results illustrated in Figure 4(a) is given in Table S1(a) [see the Supporting Information].

Unlike the FEPs along the θ_i angles shown in Figure 4(a), the shapes of the FEPs along the γ_i angles of the full trajectory [black lines in Figure 4(b)] are diverse. However, the FEPs for the β-strand residues (FEPs with red numbers) and their edges (FEPs with green numbers) are similar to each other. In particular, they exhibit one deep minimum at ∼180° (γ_i, i = 8−12, 16−22, 26−29) along with a few shallow minima in the region of 40−100° (γ_i, i = 8−12, 16, 18−22, 26−29) and in the vicinity of 270° (γ_i, i = 9, 11, 12, 17, 19, 22, 26−29). There is no representative shape for the FEPs at the loop residues connecting the β-strands. This is not surprising because there are not enough loop residues to define a γ angle; consequently, every γ angle in this part of the protein is influenced by the neighboring β-strand. Only a few γ_i angles at the N- and C-terminal loops have FEPs similar to each other (γ_i, i = 5, 34, 35). It should be noted that, as in the FEP along the θ_{33} angle, the shape of the FEP for γ_{33} is similar to the representative shape of the FEPs of the β-strand residues.

The FEPs along the γ_i angles for selected time-periods [blue, red, and green lines in Figure 4(b)] clearly illustrate that the residues of the first and middle β-strands and their edges (γ_i, i = 7−21) completely explore all possible shallow local minima before jumping to the native state and then remain in a deep global minimum until the end of the trajectory [black lines in Figure 4(b)]. As in the FEPs along the θ angles [Figure 4(a)], the residues of the third β-strand and its edges (γ_i, i = 24−30) partially explore the local minima in the non-native state [blue, red, and green lines in Figure 4(b)] and, based on the fluctuations of the θ and γ angles in the course of time (not shown), start to jump back and forth between the local and global minima when the whole system is in the native state [black lines in Figure 4(b)].

As for the θ angles, a more detailed, quantitative representation of the results illustrated in Figure 4(b) is given in Table S1(b) [see the Supporting Information].

Based on the results of the FEPs along the θ_i and γ_i angles (Figure 4) and those of Table S1(a,b), it can be concluded that, in the folding trajectory, (i) most of the residues of hairpin I (β-strands 1 and 2) completely follow all the changes in the dynamics of the whole system (these residues are called consistent); (ii) the residues of the third β-strand partially follow the changes in the dynamics of the whole system (these residues are called partially consistent), which indicates that the third β-strand is not as conformationally stable as the first two.

Also, it should be noted that most of the FEPs along the γ angles representing the β-strands and their edges exhibit three minima [one (deepest) corresponds to the native state, and the other two (shallow) correspond to the unfolded and intermediate states, respectively], while the FEPs along the θ angles representing the same parts of the protein have mainly two minima [one (deepest) corresponds to the native state, and the second (shallow) corresponds to the unfolded and intermediate states]; such behavior of θ and γ angles indicates that the γ angles are more sensitive and correlated to the global motions of the protein than are the θ angles.

FEPs along θ_i and γ_i Angles of the Nonfolding Trajectory. The FEPs along the θ_i angles for the loop and end residues (black numbers) computed over the entire nonfolding trajectory [black lines in Figure 4(c)] are similar to the FEPs along the θ_i angles for the loop and end residues computed over the entire folding trajectory [black lines in Figure 4(a)]. However, most of the FEPs along the θ_i angles pertain to β-strands (red numbers), and some FEPs along the θ_i angles for the edges of loops and β-strands, which include the motions of residues from both loops and β-strands of the nonfolding trajectory (green numbers), differ from the corresponding FEPs of the folding trajectory [black lines in Figure 4(a,c)]. Indeed, in the nonfolding trajectory, most of the FEPs along the θ_i angles pertain to β-strands (θ_i, i = 9−12, 18, 20−22, 27−29) and have a deep minimum at ∼90° and a shallow minimum at ∼120°−130°; the reverse is true for these residues in the folding trajectory. It, therefore, turns out that, in the nonfolding trajectory, the β-strand residues are trapped in local minima.

The FEPs along the θ_i angles computed over the selected time-periods and over the entire duration of the nonfolding trajectory [blue, red, green, and black lines in Figure 4(c)] show that almost each angle (except θ_{32}) gradually explores its own FEP.

By comparing the FEPs along the γ_i angles computed over the entire duration of the nonfolding trajectory [black lines in Figure 4(d)] to the FEPs along the γ_i angles computed over the entire duration of the folding trajectory [black lines in Figure 4(b)], differences can be observed not only for the angles pertaining to the β-strands (γ_i, i = 9−11, 19−21, 27, 28) but also for the angles pertaining to the loop and the edges between loops and β-strands (γ_i, i = 4, 6−8, 12, 13, 15, 16, 22−25, 31, 32). The main difference for the angles pertaining to the β-strands is that the deepest minimum of their FEPs is shifted from 180° (in the folding trajectory) to the region of 0°−80° (in the nonfolding trajectory), which is similar to the shift occurring for the FEPs along the θ_i angles pertaining to the β-strands, i.e., from 120°−130° [in the folding trajectory, Figure 4(a)] to 90° [in the nonfolding trajectory, Figure 4(c)]. For the nonfolding trajectory, the molecule also explores wider ranges of the γ_i angles, which is manifested in a small number of regions in which the FEPs are undefined [Figure 4(d)], compared to those in the plots corresponding to the folding trajectory (Figure 4(b)]. This indicates that the protein explores a larger portion of conformational space when it does not fold, as expected. The regions, which are undefined in the FEPs along the γ_i angles in the folding trajectory and explored by the γ_i angles in the nonfolding trajectory, are far from the NMR-derived structural data for most of the angles [red circles in Figure 4(b,d)]; therefore, the exploration of these regions either delays the folding or it can be one of the reasons for nonfolding. It should be noted that, despite the many differences between the FEPs along the γ_i angles computed from the folding and nonfolding trajectories, there are some similarities, as well, at the flexible N- and C-terminal parts (γ_i, i = 2, 3, 33, 35) [Figure 4(b,d)].

In the nonfolding trajectory, almost all θ and γ angles explore their FEPs gradually, and a majority of the angles pertaining to the β-strands and their edges are trapped in their local minima and do not conform to the global minimum [Figure 4(c,d)]. This is in contrast to the folding trajectory in which the local minima

dx.doi.org/10.1021/ct4001558 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(non-native state) were explored before the global minimum (native state) [Figure 4(a,b)].

Based on the analysis of the FEPs described above for the folding and nonfolding trajectories, the following can be concluded: (i) most of the residues of all three β-strands (especially the first and middle ones) and their edges and loops connecting the β-strands play a crucial role in folding the FBP28 WW domain; (ii) the residues of the N- and C-termini do not follow the changes in the dynamics of the whole system; hence, their contribution to the folding event is minimal; (iii) the main reason for nonfolding is that many residues (especially the residues of the β-strands) are trapped in a metastable conformation (i.e., in local minima of the FEPs computed over the entire folding trajectory), and there is no consistency between the FEPs along the angles involving these residues and the whole system, in contrast to the behavior observed for the folding trajectory; (iv) our finding, that almost all residues (except the residues at the N- and C-termini) in the folding trajectories move more or less in a concerted (correlated) fashion and the angles involving these residues are consistent with the changes of the whole system (monitored by the RMSD), reflects the observation[7] that the internal coordinate PCA applied to the θ and γ angles was very efficient for the description of the folding and less efficient for the description of the nonfolding trajectories of the FBP28 WW domain. The point is that the collective motion in a protein is any motion that involves a number of atoms moving in a concerted fashion, and PCA appears to be the best method for extracting collective variables from MD simulations.[60]

In order to justify and strengthen the aforementioned statements concluded from the analysis of only one folding and one nonfolding trajectory, we computed the FEPs along the θ and γ angles for the three fast-folding trajectories [Figure S2(a,b) in the Supporting Information]. We also computed the Pearson average correlation coefficients R[59,61][Figure 5(a)] between the FEPs along the θ and γ angles of the three fast-folding [Figure S1(a−c)] and the four slow-folding [Figure S1(d−g)] trajectories and the correlation coefficients [Figure 5(b,c)] between the FEPs along the θ and γ angles of all folding (7 MD trajectories) and the three nonfolding trajectories [Figure S1(h−j)]. All FEPs of the three fast-folding trajectories are very similar, hardly distinguishable from each other [black lines in Figure S2(a,b)]. Minor differences between the FEPs of the non-native states [red lines in Figure S2(a,b)] are caused by the time-differences between the non-native states explored by the system in the fast-folding trajectories. The average correlation coefficients between the FEPs along both the θ and γ angles of the fast-folding and slow-folding trajectories [Figure 5(a)] are very high ($R > 0.96$ for FEPs along the θ angles, $R > 0.87$ for FEPs along the γ angles), which indicates the similarity between the folding trajectories. The differences between the average correlation coefficients for the FEPs along the θ and γ angles of all folding and nonfolding trajectories are noticeable at β-strands, edges and N- and C-termini [Figure 5(b,c)]. In particular, for the FEPs along the θ angles, R_{avg} = 0.86, 0.76, 0.83 for the first, second, and third β-strands, respectively, whereas R_{avg} = 0.98, 0.95 for the N- and C-termini, respectively [Figure 5(b)]. For the FEPs along the γ angles, R_{avg} = 0.74, 0.66, 0.65 for the first, second, and third β-strands, respectively, whereas R_{avg} = 0.95, 0.90 for the N- and C-termini, respectively [Figure 5(c)]. The average correlation coefficients for the edges are R_{avg} = 0.55, 0.80, 0.12, 0.64 [Figure 5(c)]. These results indicate that the FEPs of the angles representing the C- and N-termini of the protein in the folding and nonfolding trajectories are highly correlated; and the

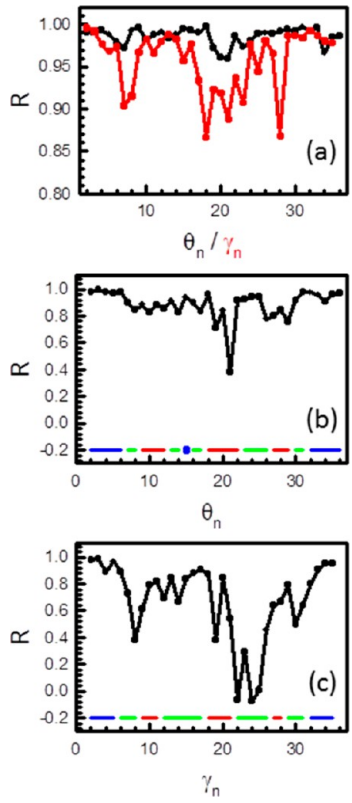


Figure 5. The average correlation coefficients R computed between the FEPs along the θ (black line) and γ (red line) angles of the three fast- and the four slow-folding trajectories [panel (a)] and along θ [panel (b)] and along γ [panel (c)] angles of all seven folding and three nonfolding trajectories. The solid lines on the bottom of panels (b) and (c) correspond to the β-strands (red), β-strands edges (green), and N- and C-termini and loops (blue) regions.

FEPs of the angles representing the β-strands and their edges of the protein in the folding and nonfolding trajectories are less correlated. From the above analysis, the conclusions based on one folding and one nonfolding trajectory can be extrapolated to other folding and nonfolding trajectories of the FBP28 WW domain.

It should be noted that our recent studies[7] of the folding and nonfolding trajectories for FBP28 WW domain by PCA also revealed the strong correlations between the first few PCs and the RMSDs in the folding trajectories and noncorrelations between the PCs and the RMSDs of the folding and nonfolding trajectories. Similar results have been obtained for the trajectories studied here (not shown).

Although the FEPs along the θ_i and γ_i angles identified the residues that participate in the folding of the FBP28 WW domain (see Figure 4), it is still not clear which parts of the protein form first, and which residues play a crucial role in folding. In order to answer these questions, the FEPs along the distances between selected residues for both folding and nonfolding trajectories were examined.

FEPs along the Distances between the Residues in Folding and Nonfolding Trajectories. Figures 6 and 7 illustrate the FEPs along the distances between the selected residues [GLU(10)−TYR(20), THR(13)−LYS(17), ALA(14)−GLY(16), TYR(20)−SER(28), ASN(23)−LEU(26)] for folding and nonfolding trajectories, respectively. The residues in both trajectories [see panels (c), (f), (i), (l), and (o) in both Figures 6 and 7] were selected with the purpose of trying to identify the consequence of the formation of native contacts in all parts of

dx.doi.org/10.1021/ct4001558 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

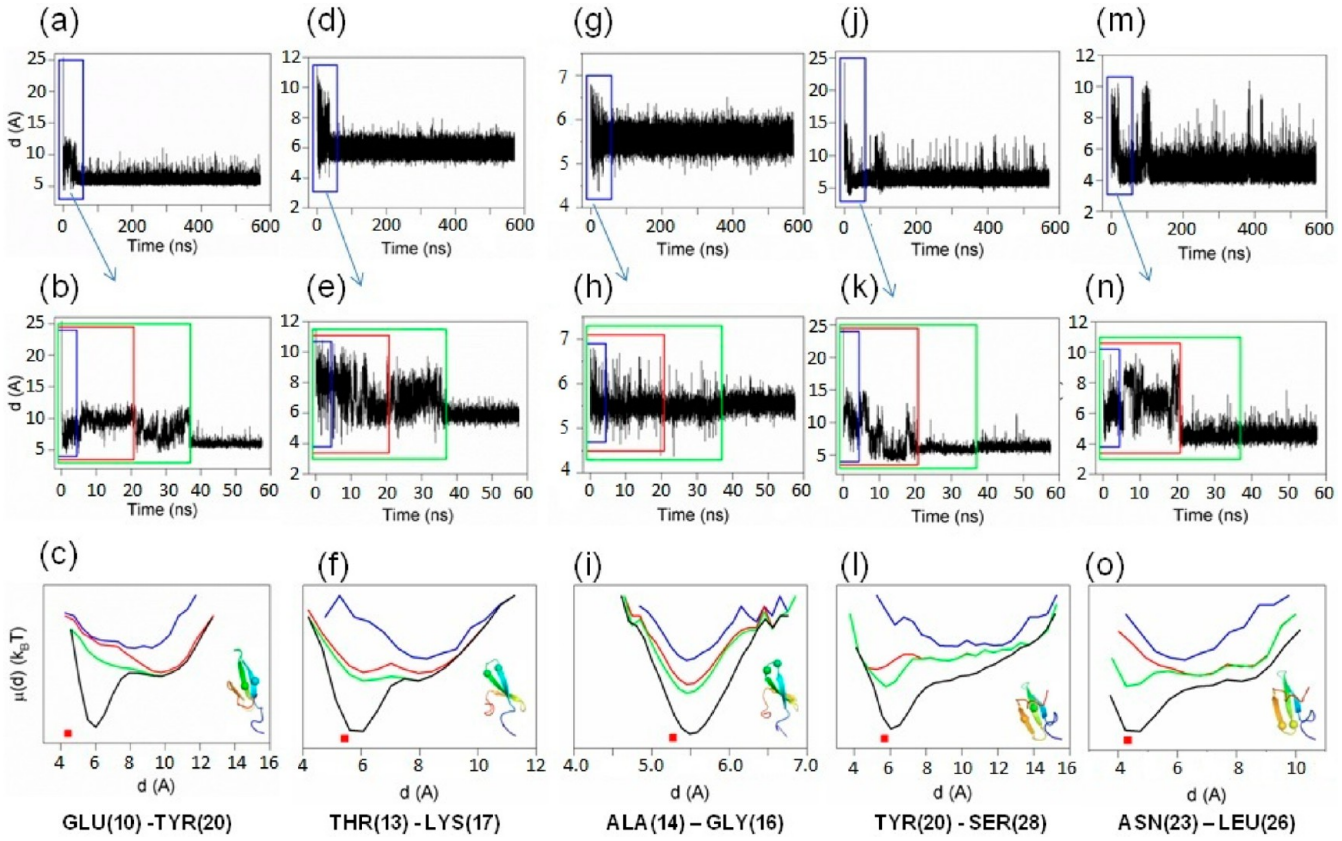


Figure 6. The distances GLU(10)−TYR(20) [panels (a) and (b)], THR(13)−LYS(17) [panels (d) and (e)], ALA(14)−GLY(16) [panels (g) and (h)], TYR(20)−SER(28) [panels (j) and (k)], and ASN(23)−LEU(26) [panels (m) and (n)] as a function of time for the entire folding trajectory and for the 4.5 ns (blue rectangle), 21 ns (red rectangle), and 37 ns (green rectangle) time periods; and the FEPs along these distances [panels (c), (f), (i), (l), (o)] for 4.5 ns, 21 ns, and 37 ns time periods (blue, red, green lines) and the entire trajectory (black line). Small red squares at the bottom of panels (c), (f), (i), (l), (o) correspond to the distances between the selected residues (represented by spheres in the molecular structures shown in the insets) from the NMR experiment.[13] Distances were computed between the C^{α} of the residues.

the protein and those residues that might be key for folding. The FEPs in Figures 6 and 7 were calculated for the same time-periods as the FEPs in Figure 4; hence, the FEPs in Figures 6 and 7 correspond to the large changes in the RMSD of the whole system illustrated in Figure 3.

Hairpin I. The first three pairs of the selected residues represent hairpin I in folding [panels (a)−(i) of Figure 6] and in nonfolding [panels (a)−(i) of Figure 7] trajectories. In particular, (1) panels (a) and (b) of Figures 6 and 7 illustrate the distance between residues GLU(10) and TYR(20) as a function of time for the entire trajectory and for the selected time periods, respectively, and panel (c) of Figures 6 and 7 shows the FEPs along the distances for the selected time-periods (blue, red, green lines) and the entire trajectory (black line); (2) panels (d)−(f) and (g)−(i) of Figures 6 and 7 illustrate the same variables (distances and free energy) for the following pairs of residues: THR(13)−LYS(17), and ALA(14)−GLY(16), respectively.

Residues GLU(10) and TYR(20) are located in the middle of the first and second β-strands, respectively; hence, the distance between them as a function of time describes the behavior of hairpin I. Based on the results shown on panels (a) and (b) of Figures 6 and 7, the changes in the dynamics of the whole system, illustrated by the RMSD vs time in Figure 3, are detected correctly by the distance between residues GLU(10) and TYR(20) for both the folding and nonfolding trajectories, which indicates that, regardless of the type of trajectories (folding or nonfolding), the middle part of hairpin I follows the dynamics of the whole system. The reasons for folding and nonfolding are illustrated in panel (c) of Figures 6 and 7, in which the FEPs computed from the folding and nonfolding trajectories differ from each other. In particular, the FEP computed over the entire folding trajectory [black line in panel (c) of Figure 6] exhibits one shallow minimum (at ∼9.8 Å), which is explored completely before reaching the native state [see blue, red, and green lines in panel (c) of Figure 6], and one narrow deep minimum (at ∼6.0 Å), formed after jumping into the native state. By contrast, the FEP computed over the entire nonfolding trajectory [black line in panel (c) of Figure 7] shows two deep minima [at ∼8.8 Å (deepest) and ∼5.9 Å], and both of them are gradually explored over time [see blue, red, and green lines in panel (c) of Figure 7]. Based on these results, it can be concluded that, in the folding trajectory, hairpin I (at least its middle part) is very stable. In other words, formation of hairpin I coincides with the jump of the protein into the native state, and it remains formed until the end of the trajectory. Unlike the folding trajectory, hairpin I in the nonfolding trajectory is very unstable, and the amplitude of the fluctuations in the distance of its middle part (between two deep minima) is ∼3.0 Å [panels (a)−(c) of Figure 7]. It should be noted that the calculated distance between residues GLU(10) and TYR(20) in the folding trajectory (∼6.0 Å) differs from the experimental value, small red square (∼4.4 Å).[13]

In order to corroborate the conclusions about the stability/instability of hairpin I, a second pair of residues [THR(13) and LYS(17)] located on the edges of the first and middle β-strands and

dx.doi.org/10.1021/ct4001558 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

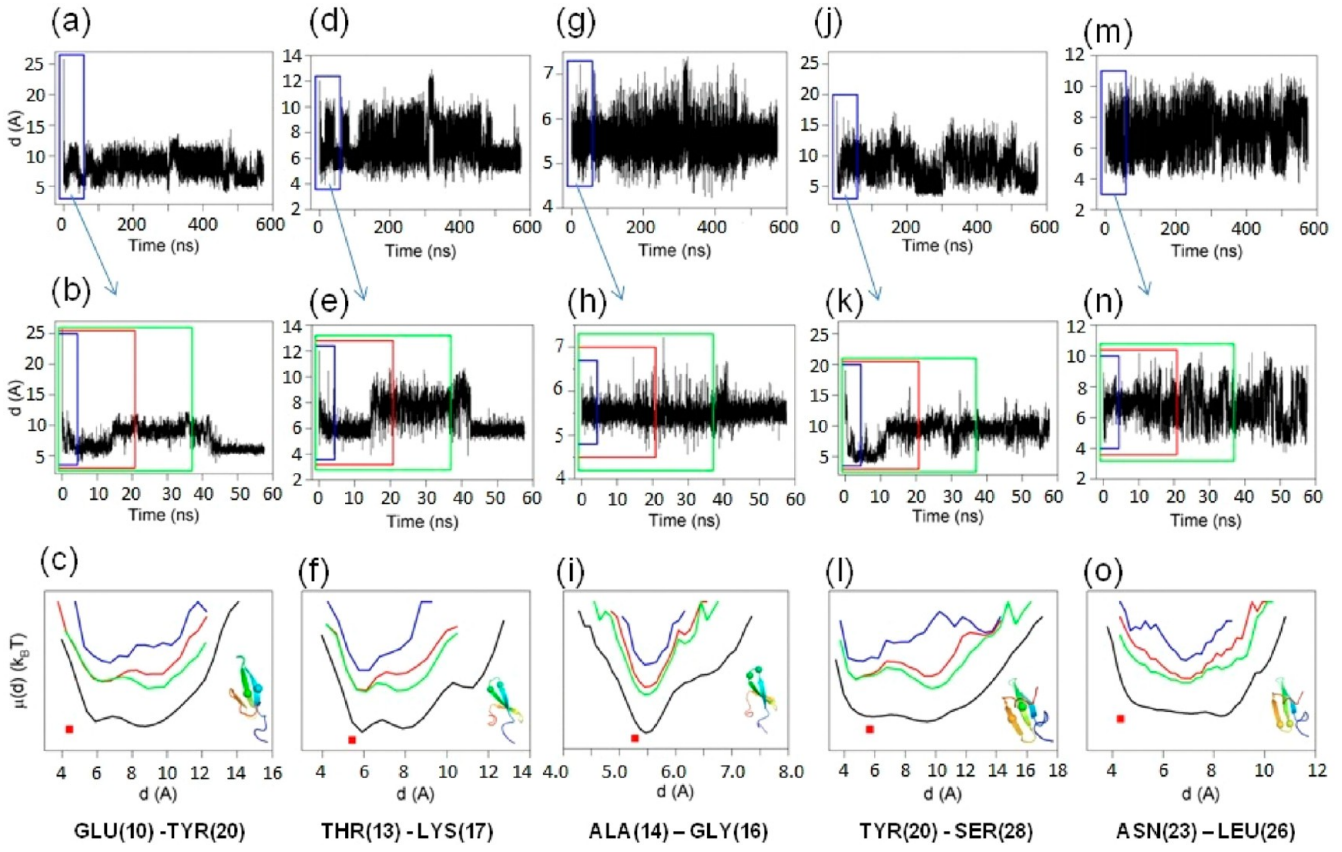


Figure 7. The same as in Figure 6 but for the nonfolding trajectory.

the loop connecting these β-strands was selected [panels (d)−(f) of Figures 6 and 7]. The results shown in panels (d) and (e) of Figures 6 and 7 indicate that the distance between residues THR(13) and LYS(17) is as sensitive to the changes of the whole system (monitored by the RMSD, Figure 3), as was the distance between residues GLU(10) and TYR(20), except for the first large change of the protein RMSD at ∼4.5 ns [see blue rectangle in panel (e)], which was not as clearly detected by residues THR(13)−LYS(17) as was by residues GLU(10)−TYR(20). However, the FEPs along the distances between the both pairs of residues [GLU(10)−TYR(20) and THR(13)−LYS(17)] computed over the entire folding trajectory are similar to each other [panels (c, f) of Figure 6]; in addition, the deep minimum in panel (f) (∼5.8 Å), which corresponds to the native basin, is closer to the experimental value, small red square (∼5.4 Å)[13] than the one for residues GLU(10)−TYR(20) [panel (c)]. The FEPs in panels (c) and (f) (black lines) indicate that hairpin I is very stable and follows the dynamics of the whole protein. The FEP along the distance between residues THR(13)−LYS(17), computed over the entire nonfolding trajectory [panel (f) of Figure 7], reveals even a more complicated shape with three minima (∼5.9 Å, 7.9 Å, and 11.3 Å) than the one for residues GLU(10)−TYR(20) [panel (c) of Figure 7], which is additional evidence for the instability of hairpin I in the nonfolding trajectory.

In order to understand the role played by the loop of hairpin I in folding/nonfolding, the fluctuations and the FEPs along a third distance, selected between residues ALA(14) and GLY(16), located in the loop connecting the first and middle β-strands, were computed [panels (g)−(i) of Figures 6 and 7]. In both the folding and nonfolding trajectories, the fluctuations of the distance ALA(14)−GLY(16) over time [panels (g) and (h) of Figures 6 and 7] do not reflect the large jumps found in the RMSDs (Figure 3) as well as the distances of previous pairs of residues did [panels (a, b) and (d, e) in Figures 6 and 7]. However, the behavior of the distance between residues ALA(14) and GLY(16) over time differs in the folding and in the nonfolding trajectories [compare panels (g) and (h) of Figures 6 and 7]. The loop of hairpin I is more flexible at the beginning of the folding trajectory than in the nonfolding trajectory [see the amplitude of the fluctuations during the first ∼4.5 ns (blue rectangle) in panel (h) of Figures 6 and 7]. However, in the folding trajectory, the amplitude of the fluctuations decreases gradually and becomes stable in the intermediate state (between ∼21 ns and ∼37 ns). On the contrary, the loop of hairpin I in the nonfolding trajectory is quite rigid at the beginning of the trajectory but becomes more flexible over time, and the amplitude of the fluctuations increases almost twice [panels (g) and (h) of Figure 7].

The FEPs along the distance between ALA(14) and GLY(16) for both trajectories, in general, exhibit a similar shape with one minimum at ∼5.5 Å for all time periods; however, the differences observed in panels (g) and (h) regarding the fluctuations are reflected in the shapes of the FEPs [panel (i) of Figures 6 and 7]. Based on the results illustrated in panels (g)−(i) of Figures 6 and 7, it can be concluded that the loop connecting the first and middle β-strands forms in the early stage of the trajectory and remains formed until the end of the trajectory with enhanced stability over time for the folding trajectory and reduced stability over time for the nonfolding trajectory. Therefore, formation of the loop of hairpin I seems to play an important role in the early stage of folding and helps the protein to fold if the loop fluctuates slightly in the vicinity of the most probable distance ALA(14)−GLY(16) (5.5 Å), whereas too large fluctuations may impede

dx.doi.org/10.1021/ct4001558 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

folding. This finding is in agreement with an earlier experimental study,[31] in which the major influence of formation of the loop of hairpin I on the folding rate was observed. It should be noted that, since the angle θ_{15} is formed by residues ALA(14)−ASP(15)−GLY(16), there is a striking similarity between the FEPs along θ_{15} [panels (a, c) of Figure 4] and along the distance ALA(14)−GLY(16) [panel (i) of Figures 6 and 7] for both folding and nonfolding trajectories.

Hairpin II. The two other distances were selected to represent hairpin II in folding [panels (j)−(o) of Figure 6] and nonfolding [panels (j)−(o) of Figure 7] trajectories. In particular, panels (j)−(l) and (m)−(o) of Figures 6 and 7 illustrate the same variables (fluctuations of the distances over time and free energy) as the above [panels (a)−(i)] for the distances: TYR(20)−SER(28) and ASN(23)−LEU(26), respectively. It should be noted that, because of the short length of the loop connecting the second and third β-strands, the pair of residues representing this loop was not selected for this analysis.

Residues TYR(20) and SER(28) are located in the middle of the second and third β-strands, respectively; hence, the time-dependence of the distance between them describes the general behavior of hairpin II. Based on the results shown in panels (j) and (k) of Figures 6 and 7, the changes in the dynamics of the whole system (monitored by the RMSD as shown in Figure 3) are detected correctly in the fluctuations of the distance TYR(20)−SER(28) over time both in the folding and in nonfolding trajectories. This indicates that, regardless of the type of trajectory (folding or nonfolding), the middle part of hairpin II follows the dynamics of the whole system. However, unlike hairpin I (panels (a) and (c) of Figure 6), after reaching the native state, hairpin II exhibits an unstable behavior with large fluctuations [as reflected by the distance TYR(20)-SER(28) in panel (j) of Figure 6]. Hence, the FEP computed over the entire folding trajectory [black line in panel (l) of Figure 6] exhibits only one deep minimum (∼6.0 Å), and a shoulder instead of a shallow minimum, which is not explored completely (see blue, red, and green lines) before jumping into the native state. It should be noted that the distance TYR(20)-SER(28) starts exploring the native basin much earlier (after 4.5 ns) [red line in panel (l) of Figure 6] than the whole system jumps to the native state and, after that, interconverts between the native basin and a local minimum (shoulder) located around 8−10 Å [panel (l) of Figure 6].

The FEP along the distance TYR(20)−SER(28) computed over the entire nonfolding trajectory [black line in panel (l) of Figure 7] exhibits one broad minimum and resembles the FEP illustrated in panel (c) of Figure 7. Like hairpin I, based on these results, hairpin II in the nonfolding trajectory is also very unstable; the distance between the middle parts of the β-strands of hairpin II oscillates between ∼5 Å and ∼10 Å [panel (l) in Figure 7].

The results for the last pair of selected residues, ASN(23) and LEU(26), located at the edges of the middle and third β-strands and the loop connecting these β-strands of the folding trajectory [panels (m)−(o) of Figure 6] also indicate that, despite the consistency with the dynamics of the whole system in the non-native state, after jumping into the native state of the whole system, the distance between residues ASN(23) and LEU(26) still fluctuates with a large amplitude (∼5.0 Å) from time to time [panel (m) of Figure 6] showing the highest flexibility of hairpin II. The corresponding FEP of the distance ASN(23)−LEU(26) computed over the entire folding trajectory [panel (o) of Figure 6], as expected, is quite similar to the FEP for the previous pair [panel (l) of Figure 6]. Unlike the folding trajectory, the distance between residues ASN(23) and LEU(26) in the nonfolding trajectory does not follow the dynamics of the whole system [panels (m) and (n) of Figure 7], and the corresponding FEPs exhibit a broad, "unstructured" shape [panel (o) of Figure 7].

It should be noted that this behavior of hairpin II, which is caused mainly by the high flexibility of the third β-strand, leads to biphasic kinetics in the folding of the FBP28 WW domain.[22,23,33] The conclusions drawn in these studies[22,23,33] regarding the biphasic kinetics coincide with ours. In particular, the misregistered structure in hairpin II is the cause of the biphasic kinetics, which was traced to the mobility of the third β-strand within the context of otherwise folded conformations.

Moreover, the behavior of hairpins I and II and the role of loops in folding, observed in this study, is in harmony with earlier experimental[31,62] and theoretical[26] results.

PCA of the Folding Trajectory. Since most of the residues move in a concerted fashion in the folding trajectory of the FBP28 WW domain, as was observed above [Figure 4(a,b)], it is interesting to examine whether the FEL along the corresponding PCs computed with the involvement of several θ_i and γ_i angles can correctly describe protein folding. Therefore, three different analyses of the folding trajectory were carried out by PCA: (i) PCs were computed by considering the whole protein; (ii) PCs were computed by considering only five sets of θ and γ angles (9−11, 26, 29); (iii) PCs were computed by considering only six sets of θ and γ angles (2−4, 33−35).

The angles in analyses (ii) and (iii) were not selected randomly. Indeed, they correspond to the angles with the largest and smallest contributions, respectively, to the first two PCA modes (defined in Materials and Methods); i.e., the principal modes (with the largest eigenvalues λ^k) contributing the most to the structural fluctuations [mean-square-fluctuations (MSF)] of the protein. Figure 8 illustrates the contributions of the two main principal modes [solid lines with filled (principal mode 1) and empty (principal mode 2) rectangles] to the MSF along the θ [panel (a)] and γ [panel (b)] angles for the folding (black lines) and nonfolding (red lines) trajectories.

Based on the results shown in panel (a), the main contribution to the fluctuations in both trajectories comes from the β-strands; however, some contribution from the loop connecting the first and middle β-strands should also be mentioned. Panel (b) illustrates different results for the folding and nonfolding trajectories. In particular, the major contribution to the fluctuations in the folding trajectory comes from the β-strands, especially from the first β-strand; however, the contributions from the loops, especially from the loop connecting the middle and third β-strands, are also noticeable. The major contribution to the fluctuations in the nonfolding trajectory comes from the loop connecting the first and middle β-strands; however, the contributions from the β-strands are also noticeable. It should also be noted that the magnitude of the contribution of principal mode 1 at some θ and γ angles for the nonfolding trajectory is several times greater than the one for the folding trajectory. This is not surprising because the system in the nonfolding trajectory is more flexible and the amplitude of the fluctuations is much larger than in the folding trajectory; consequently, λ^1 for the nonfolding trajectory is a few times larger than the one for the folding trajectory.

Comparing the results shown in Figures 4 and 6−8, we can see the correlation between local and global motions. The angles, whose fluctuations over time are consistent with the dynamics of the whole system (i.e., the fluctuations of the RMSD), are the main contributors to the MSF; hence, they are the main players

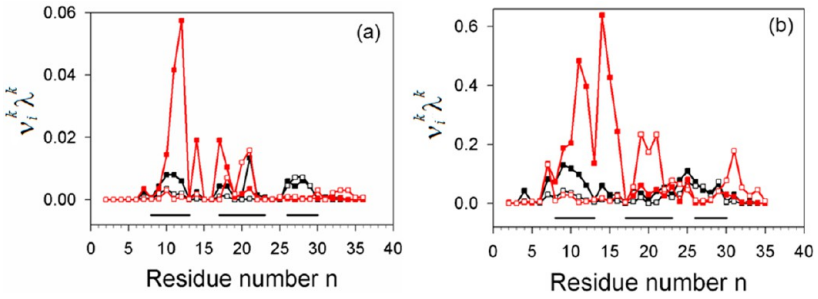


Figure 8. Contributions of the principal modes [$v_i^k\lambda^k$, solid lines with filled (principal mode 1) and empty (principal mode 2) rectangles] to the mean-square-fluctuations for the folding (black lines) and nonfolding (red lines) trajectories along the θ [panel (a)] and γ [panel (b)] angles. The solid black lines on the bottom of the panels correspond to the β-strand regions.

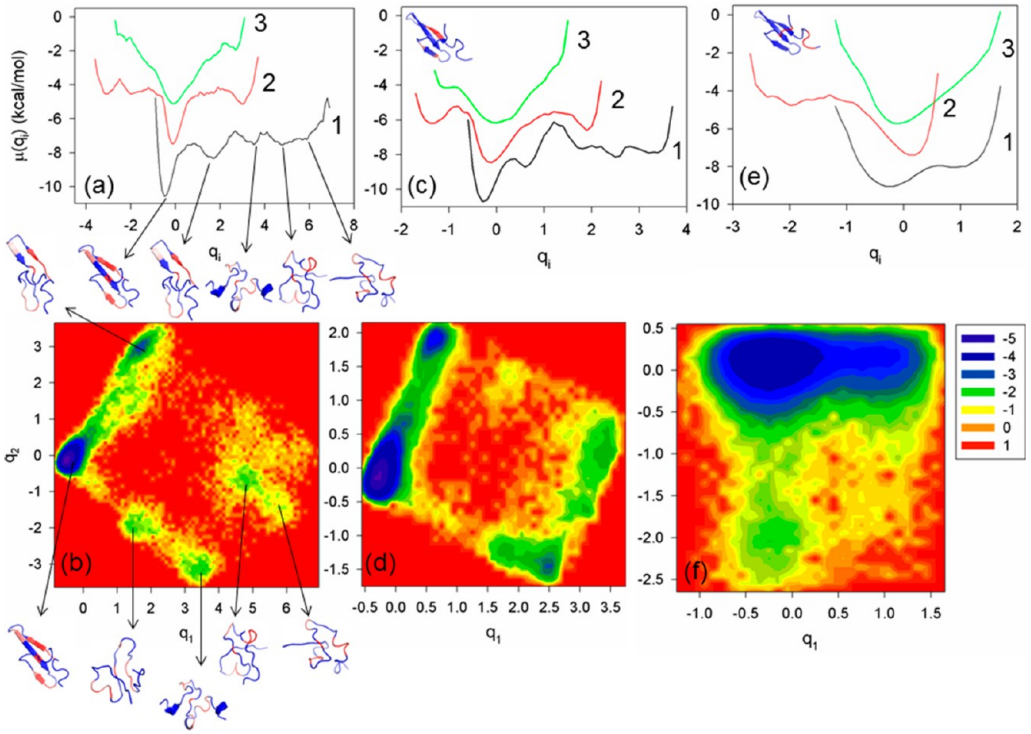


Figure 9. Free-energy profiles, $\mu(q_i)$, for the first three PCs, (q_i), and free-energy landscapes (in kcal/mol) along the first two PCs, for the folding trajectory of the FBP28 WW domain, computed by considering the whole molecule [panels (a) and (b)], five sets of θ and γ angles 9−11, 26, 29 [panels (c) and (d)], and six sets of θ and γ angles 2−4, 33−35 [panels (e) and (f)]. Some residues of the representative structures at the minima of the FEP along the first PC [panel (a)] and the FEL along the first two PCs [panel (b)] are highlighted in red based on their contributions to the mean-square-fluctuations (see Figure 1). The highlighted parts of the native structure, illustrated in panels c and e (in red), correspond to the residues 9−11, 26, 29, and 2−4, 33−35, respectively.

in folding. In other words, they impel the protein to follow their dynamics.

The FEPs, $\mu(q_i) = -k_BT\ln P(q_i)$, constructed along the first three PCs (i = 1−3), and the FELs along the first two PCs, $\mu(q_1,q_2) = -k_BT\ln P(q_i,q_2)$, for the whole protein [panels (a) and (b)], for only five sets of θ and γ angles (9−11, 26, 29) [panels (c) and (d)], and for only six sets of θ and γ angles at the N- and C-termini of protein (2−4, 33−35) [panels (e) and (f)] for the folding trajectory, are shown in Figure 9. The sets of θ and γ angles (9−11, 26, 29, and 2−4, 33−35) were selected by their maximal and minimal contributions in principal modes 1 and 2.

Both FEPs and FELs, computed by considering the whole system [panels (a) and (b)] and five sets of θ and γ angles [panels (c) and (d)], are very similar to each other, which is not surprising since all five angles are the main contributors to the MSF, and their local motions are consistent with the dynamics of the whole system. Both FEPs along the first PC and FELs exhibit three-state folding dynamics, which coincide with the results obtained in earlier theoretical[7,9,10,25,30] and experimental[33] studies. The FEPs and FEL computed by considering six sets of θ and γ angles at the N- and C-termini (panels e and f) do not resemble the FEPs and FELs illustrated in panels a−d and, therefore, do not describe folding dynamics correctly.

CONCLUSIONS

By analyzing the MD trajectories of the FBP28 WW domain, generated with the coarse-grained UNRES force field, in terms of the local motions of each residue and PCA, we have studied the dynamics of the folding and nonfolding trajectories and found the following:

(i) The loop connecting the first and middle β-strands forms in the early stage of simulation[26] and consequently plays an important

dx.doi.org/10.1021/ct4001558 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

role in folding/nonfolding of the protein at the beginning of the trajectory.

(ii) The dynamics of the residues of the first and middle β-strands and their edges with the loop connecting these β-strands, in the folding trajectory, are consistent with the dynamics of the whole system and exhibit very stable behavior; hence, their role in folding is crucial. The residues of the third β-strand in the folding trajectory follow the motions of the whole system on some level; however, after reaching the native state their behavior is not as stable as that of the residues of the first two β-strands. None of the residues in the nonfolding trajectory exhibits complete consistency with the dynamics of the whole system.

(iii) In the folding trajectory, hairpin I forms first and remains formed until the end of the trajectory, whereas hairpin II, because of the behavior of the third β-strand, shows unstable behavior, which is the reason for biphasic kinetics in the folding of the FBP28 WW domain. These findings support earlier experimental[31,33,36,62] and theoretical[22,23,26] results.

(iv) The residues, whose motions are consistent with the dynamics of the whole system in the folding trajectory, are the main contributors to the MSF in principal modes 1 and 2. This indicates the correlations between local and global motions in protein folding, which, to the best of our knowledge, is a new observation.

(v) Unlike the nonfolding trajectory, for the folding trajectory most of the FEPs along the θ_i and γ_i angles for loops and β-strand residues can be identified by their shapes, and most of the residues move in a concerted fashion and follow the dynamics of the whole system. The main differences between the FEPs of the folding and nonfolding trajectories are along the angles belonging to the β-strands.

(vi) The reason why PCA has proved to be an effective tool for the analysis of protein folding trajectories is that this type of analysis involves concerted motions of many residues, which can be captured by a few PCs with the largest eigenvalues.

(vii) The key residues, involved in the folding/nonfolding of the studied trajectories and identified here, are the "main players", in general, in the folding/nonfolding of the FBP28 WW domain. Perhaps, this also applies to other β-proteins (for this reason some additional studies are planned in the future).

(viii) Five sets of θ and γ angles are enough to construct the FEP and FEL thereby correctly describing the folding dynamics of the FBP28 WW domain.

Finally, it should be noted that the questions addressed in the present study, i.e., why the FBP28 WW domain protein folds or does not fold and what governs the FBP28 WW domain to fold, are very important for understanding misfolding and, consequently, the diseases associated with misfolding of the WW domain proteins.

■ ASSOCIATED CONTENT

Supporting Information

Table S1(a),(b) and Figures S1 and S2. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: has5@cornell.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (GM-14312) and the National Science Foundation (MCB10-19767) and conducted by using the resources of (a) our 588-processor Beowulf cluster at the Baker Laboratory of Chemistry and Chemical Biology, Cornell University, (b) the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center, (c) the Beowulf cluster at the Department of Computer Science, Cornell University, (d) the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdańsk, and (e) the Interdisciplinary Center of Mathematical and Computer Modeling (ICM) at the University of Warsaw.

■ REFERENCES

(1) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **1991**, *254*, 1598−1603.

(2) Brooks, C. L., III.; Onuchic, J. N.; Wales, D. J. Taking a walk on a landscape. *Science* **2001**, *293*, 612−613.

(3) Wales, D. J. *Energy landscapes*; Cambridge University Press: Cambridge, U.K., 2003; p 681.

(4) Krivov, S. V.; Karplus, M. Hidden complexity of free energy surfaces for peptide (protein) folding. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14766−14770.

(5) Auer, S.; Miller, M. A.; Krivov, S. V.; Dobson, C. M.; Karplus, M.; Vendruscolo, M. Importance of metastable states in the free energy landscapes of polypeptide chains. *Phys. Rev. Lett.* **2007**, *99* (178104), 1−4.

(6) Krivov, S. V.; Muff, S.; Caflisch, A.; Karplus, M. One-dimensional barrier-preserving free-energy projections of a β-sheet miniprotein: New insights into the folding process. *J. Phys. Chem. B* **2008**, *112*, 8701−8714.

(7) Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. Principal component analysis for protein folding dynamics. *J. Mol. Biol.* **2009**, *385*, 312−329.

(8) 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), 1−4.

(9) 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.

(10) 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.

(11) Maisuradze, G. G.; Liwo, A.; Ołdziej, S.; Scheraga, H. A. Evidence, from simulations, of a single state with residual native structure at the thermal denaturation midpoint of a small globular protein. *J. Am. Chem. Soc.* **2010**, *132*, 9444−9452.

(12) Yin, Y.; Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. Hidden folding pathways in free-energy landscapes uncovered by network analysis. *J. Chem. Theory Comput.* **2012**, *8*, 1176−1189.

(13) 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.

(14) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scherega, H. A. Prediction of protein conformation on the basis of a search for compact structures; test on avian pancreatic polypeptide. *Protein Sci.* **1993**, *2*, 1715−1731.

(15) Liwo, A.; Ołdziej, S.; Pincus, M. R.; Wawak, R. J.; Rackowsky, S.; Scheraga, H. A. A united-residue force field for off-lattice protein-structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J. Comput. Chem.* **1997**, *18*, 849−873.

(16) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Ołdziej, S.; Scheraga, H. A. A united-residue force field for off-lattice protein-structure simulations. II: Parameterization of local interactions and determination of the weights of energy terms by Z-score optimization. *J. Comput. Chem.* **1997**, *18*, 874−887.

(17) 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.
(18) Liwo, A.; Ołdziej, S.; Czaplewski, C.; Kozlowska, U.; Scheraga, H. A. Parametrization of backbone-electrostatic and multibody contributions to the UNRES force field for protein-structure prediction from ab initio energy surfaces of model systems. *J. Phys. Chem. B* **2004**, *108*, 9421−9438.
(19) Ołdziej, S.; Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. Optimization of the UNRES force field by hierarchical design of the potential-energy landscape. 2. Off-lattice tests of the method with single proteins. *J. Phys. Chem. B* **2004**, *108*, 16934−16949.
(20) Ołdziej, S.; Lagiewka, J.; Liwo, A.; Czaplewski, C.; Chinchio, M.; Nanias, M.; Scheraga, H. A. Optimization of the UNRES force field by hierarchical design of the potential-energy landscape. 3. Use of many proteins in optimization. *J. Phys. Chem. B* **2004**, *108*, 16950−16959.
(21) Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Ołdziej, S.; Wachucik, K.; Scheraga, H. A. Modification and optimization of the united-residue (UNRES) potential energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins. *J. Phys. Chem. B* **2007**, *111*, 260−285.
(22) 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.
(23) 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.
(24) 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.
(25) Mu, Y.; Nordenskiold, L.; Tam, J. P. Folding, misfolding, and amyloid protofibril formation of WW domain FBP28. *Biophys. J.* **2006**, *90*, 3983−3992.
(26) 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.
(27) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517−520.
(28) 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.
(29) Beccara, S. A.; Škrbić, T.; Covino, R.; Faccioli, P. Dominant folding pathways of a WW domain. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 2330−2335.
(30) 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.
(31) 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.
(32) 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.
(33) 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.
(34) 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.
(35) 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.
(36) 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.
(37) Serpell, L. C. Alzheimer's amyloid fibrils: structure and assembly. *Biochim. Biophys. Acta* **2000**, *1502*, 16−30.
(38) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. Folding dynamics and mechanism of β-hairpin formation. *Nature* **1997**, *390*, 196−199.
(39) Munoz, V.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. A statistical mechanical model for β-hairpin kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5872−5879.
(40) Pande, V. S.; Rokhsar, D. S. Molecular dynamics simulations of unfolding and refolding of a β-hairpin fragment of protein G. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9062−9067.
(41) Dinner, A. R.; Lazaridis, T.; Karplus, M. Understanding β-hairpin formation. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9068−9073.
(42) Guo, C. L.; Levine, H.; Kessler, D. A. How does a β-hairpin fold/unfold? Competion between topology and heterogeneity in a solvable model. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10775−10779.
(43) 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 α-helical systems. *J. Phys. Chem. B* **2005**, *109*, 13798−13810.
(44) Matheson, R. R.; Scheraga, H. A. A method for predicting nucleation sites for protein folding based on hydrophobic contacts. *Macromolecules* **1978**, *11*, 819−829.
(45) Fersht, A. R. Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10869−10873.
(46) Daggett, V.; Fersht, A. R. The present view of the mechanism of protein folding. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 497−502.
(47) Lewandowska, A.; Ołdziej, S.; Liwo, A.; Scheraga, H. A. β-Hairpin-forming peptides; models of early stages of protein folding. *Biophys. Chem.* **2010**, *151*, 1−9.
(48) Scheraga, H. A.; Liwo, A.; Ołdziej, S.; Czaplewski, C.; Pillardy, J.; Ripoll, D. R.; Vila, J. A.; Kazmierkiewicz, R.; Saunders, J. A.; Arnautova, Y. A.; Jagielska, A.; Chinchio, M.; Ninias, M. The protein folding problem: Global optimization of force fields. *Front. Biosci.* **2004**, *9*, 3296−3323.
(49) Kubo, R. J. Generalized cumulant expansion method. *Phys. Soc. Jpn.* **1962**, *17*, 1100−1120.
(50) Shen, H.; Liwo, A.; Scheraga, H. A. An improved functional form for the temperature scaling factors of the components of the mesoscopic UNRES force field for simulations of protein structure and dynamics. *J. Phys. Chem. B* **2009**, *113*, 8738−8744.
(51) Liwo, A.; Czaplewski, C.; Ołdziej, S.; Kozłowska, U.; Makowski, M.; Kalinowski, S.; Kaźmierkiewicz, R.; Shen, H.; Maisuradze, G.; Scheraga, H. A. Optimization of the physics-based united-reisdue force field (UNRES) for protein folding simulations. In *NIC Symposium, Julich, Germany*, 2008; Munster, G., Wolf, D., Kremer, M., Eds.; NIC-Directors: Julich, Germany, 2008; pp 63−70.
(52) 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.
(53) 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.
(54) Mu, Y.; Nguyen, P. H.; Stock, G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins* **2005**, *58*, 45−52.
(55) Altis, A.; Nguyen, P. H.; Hegger, R.; Stock, G. Dihedral angle principal component analysis of molecular dynamics simulations. *J. Chem. Phys.* **2007**, *126* (244111), 1−10.
(56) Maisuradze, G. G.; Leitner, D. M. Free energy landscape of a biomolecule in dihedral principal component space: sampling convergence and correspondence between structures and minima. *Proteins* **2007**, *67*, 569−578.

(57) Senet, P.; Maisuradze, G. G.; Foulie, C.; Delarue, P.; Scheraga, H. A. How main-chains of proteins explore the free-energy landscape in native states. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 19708−19713.
(58) Cote, Y.; Senet, P.; Delarue, P.; Maisuradze, G. G.; Scheraga, H. A. Nonexponential decay of internal rotational correlation functions of native proteins and self-similar structural fluctuations. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 19844−19849.
(59) Cote, Y.; Senet, P.; Delarue, P.; Maisuradze, G. G.; Scheraga, H. A. Anomalous diffusion and dynamical correlation between the side chains and the main chain of proteins in their native state. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 10346−10351.
(60) Hayward, S.; Gō, N. Collective variable description of native protein dynamics. *Annu. Rev. Phys. Chem.* **1995**, *46*, 223−250.
(61) Pearson, K. Mathematical contributions to the theory of evolution. III Regression, heredity and panmixia. *Philos. Trans. R. Soc., A* **1896**, *187*, 253−318.
(62) Deechongkit, S.; Dawson, P. E.; Kelly, J. W. Toward assessing the position-dependent contributions of backbone hydrogen bonding to β-sheet folding thermodynamics employing amide-to-ester perturbations. *J. Am. Chem. Soc.* **2004**, *126*, 16762−16771.