Folding Simulations of a De Novo Designed Protein with a $\beta\alpha\beta$ Fold

Yifei Qi,[△] Yongqi Huang,[△] Huanhuan Liang, Zhirong Liu,[*] and Luhua Lai[*]
Beijing National Laboratory for Molecular Sciences, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, College of Chemistry and Molecular Engineering, and Center for Theoretical Biology, Peking University, Beijing, China

ABSTRACT $\beta\alpha\beta$ structural motifs are commonly used building blocks in protein structures containing parallel β-sheets. However, to our knowledge, no stand-alone $\beta\alpha\beta$ structure has been observed in nature to date. Recently, for the first time that we know of, a small protein with an independent $\beta\alpha\beta$ structure (DS119) was successfully designed in our laboratory. To understand the folding mechanism of DS119, in the study described here, we carried out all-atom molecular dynamics and coarse-grained simulations to investigate its folding pathways and energy landscape. From all-atom simulations, we successfully observed the folding event and got a stable folded structure with a minimal root mean-square deviation of 2.6 Å with respect to the NMR structure. The folding process can be described as a fast collapse phase followed by rapid formation of the central helix, and then slow formation of a parallel β-sheet. By using a native-centric Gō-like model, the cooperativity of the system was characterized in terms of the calorimetric criterion, sigmoidal transitions, conformation distribution shifts, and free-energy profiles. DS119 was found to be an incipient downhill folder that folds more cooperatively than a downhill folder, but less cooperatively than a two-state folder. This may reflect the balance between the two structural elements of DS119: the rapidly formed α-helix and the slowly formed parallel β-sheet. Folding times estimated from both the all-atom simulations and the coarse-grained model were at microsecond level, making DS119 another fast folder. Compared to fast folders reported previously, DS119 is, to the best of our knowledge, the first that exhibits a parallel β-sheet.

INTRODUCTION

The protein folding problem has been studied using both experimental and theoretical approaches over the years. Despite great advances during the last decades, it is still not fully clarified. With the rapid increase of computational power, all-atom ab initio simulation of protein folding has become feasible. Since the seminal work of Yong Duan and co-workers (1), many folding simulations have been conducted using all-atom models, and most of these have focused on a limited number of fast-folding proteins (2), such as the villin headpiece subdomain (1,3–6), the albumin binding domain (7), protein A (8–11), and Trp-cage (12–14). These model systems have provided valuable information on the protein folding mechanism.

The $\beta\alpha\beta$ motifs, with two parallel β-strands packed against one α-helix, are the building blocks for proteins with α/β structures. Theoretical investigation reveals that the $\beta\alpha\beta$ structure possesses high designability (H. Chen, C. Zeng, and C. Tang, unpublished result). As hypothesized by Li et al., there is a designability principle underlying nature's selection of protein folds (15), and this principle implies that the $\beta\alpha\beta$ fold may exist as an independent structure. However, to our knowledge, the stand-alone form of the $\beta\alpha\beta$ motif has never been observed in nature until now. Recently, a small protein with an individual $\beta\alpha\beta$ motif (DS119) was successfully designed de novo and its monomeric structure was solved by NMR (16). To our knowledge, it is the first reported example of the novel $\beta\alpha\beta$ fold. The small size and the building block nature of DS119 make it an ideal model system for studying the folding of α/β proteins.

De novo protein design has made significant progress in recent years (17), and several novel proteins have been created by either computational or empirical design (18–20). Although well structured and fairly stable, these novel proteins may exhibit folding characteristics that differ from those of natural proteins. For example, most small natural proteins fold cooperatively in a two-state-like manner, whereas the de novo designed protein Top7 (21) was found to fold in a significantly less cooperative way, with severe chevron rollovers and a very rugged corresponding energy landscape. A recent coarse-grained simulation (22) revealed that the noncooperative behavior of Top7 comes from its native topology, which differs from that of natural proteins. Thus, it is of great interest to investigate the folding mechanism of de novo designed proteins, which will be helpful in understanding protein folding mechanism and further improving protein design methods.

Since systematic kinetic/thermodynamic studies of DS119 or similar proteins with parallel β-sheets have not been available, to our knowledge, in this study we performed all-atom molecular dynamics and coarse-grained simulations to investigate the folding pathway and energy landscape of DS119. In the all-atom simulations, the folding event was observed in one of the four trajectories. The folding process could be described as a fast collapse phase followed by a fast formation of the helix, then slow formation of the parallel β-sheet. The major rate-limiting step of DS119 folding was the formation of the parallel β-sheet. From the coarse-

Submitted August 15, 2009, and accepted for publication October 13, 2009.

△Yifei Qi and Yongqi Huang contributed equally to this work.

*Correspondence: LiuZhiRong@pku.edu.cn or lhlai@pku.edu.cn

Editor: Gregory A. Voth.

0006-3495/10/01/0321/9 $2.00

doi: 10.1016/j.bpj.2009.10.018

grained Gō model simulations, DS119 was shown to be an incipient downhill folder, that is, one that folds more cooperatively than a downhill folder but less cooperatively than a two-state folder. Folding times estimated from all-atom simulations and the coarse-grained model were both at the microsecond level, making DS119 another fast folder with a unique parallel β-sheet that, to the best of our knowledge, has not been observed in other fast folders.

MODELS AND METHODS

All-atom folding simulations

The DS119 protein contains 36 residues in the sequence GSGQVRTIW VGGTPEELKKLKEEAKKANIRVTFWGD, as described in our article about its design (16). Simulations were performed using the AMBER (23) package. The FF03 force field (24) was used. The solvation effects were incorporated with the GB/SA implicit solvent model. Starting from a fully extended structure, 500 steps of steepest descent followed by 500 steps of conjugate gradient minimization were carried out. The minimized structure was further heated and equilibrated at 300 K for 50 ps. For the production run, the temperature was controlled at 300 K with a coupling time constant of 2 ps. Ionic strength was set to 0.2 M. The time step used was 2 fs. Bonds involving hydrogen atoms were constrained with SHAKE (25). The nonbonded interaction and GB pairwise summation cutoff was 12 Å. Four trajectories lasting for 1 μs each, with different initial speed assignments, were generated. Among these, the folding event was observed in the third trajectory only, which was then extended to 2 μs; the folding-pathway analysis was based on this extended trajectory.

Among the 20 NMR structures obtained for DS119, the first structure was used as the reference structure in the root mean-square deviation (RMSD) calculation. The RMSD was calculated over the main-chain atoms (C_α, N, C, O) of the helix and sheet residues, i.e., Val^5-Val^{10}, Thr^{13}-Ala^{27}, and Ile^{29}-Trp^{34}, because other residues have relatively large fluctuations in the NMR structures. We also calculated the RMSD with the C_α atoms of all residues except the two flexible ends (Gly^1-Ser^2 and Gly^{35}-Asp^{36}) as the fitting atoms. The value was ~1 Å larger than the main-chain RMSD. Thus, we reported only the main-chain RMSD here. The solvent-accessible surface area was calculated with an in-house program using Boolean logic and look-up tables (26). The hydrophobic core was defined as I^8, V^{10}, L^{17}, L^{20}, A^{24}, I^{29}, V^{31}, and F^{33}.

Coarse-grained Gō model simulation

As an alternative to all-atom simulations, we carried out coarse-grained simulations on DS119 and compared its folding behavior with those of nine other small proteins to study its folding cooperativity. The structure-based continuum Gō-like model (27,28) was used, where the protein conformation is represented by C_α atoms. The interactions were specified to stabilize the native structure. The potential function and parameters used in the model were identical to those of the ''without-solvation'' model of Liu and Chan (29). The interaction strength was controlled by the parameter ε. A pair of residues, i, j, was defined as being in native contact if 1), the residues were separated in the sequence by at least three other residues, that is, $|i-j|>3$; and 2), two nonhydrogen atoms, one from each residue, were <4.5 Å apart in the Protein Data Bank (PDB) structure. Langevin dynamics was used to perform thermodynamic and kinetic simulations. Histogram techniques with bias potentials were applied to improve the sampling efficiency of equilibrium properties.

The PDB codes and residue truncations of the protein fragments studied in the coarse-grained simulations were summarized in Table S1 in the Supporting Material. For PDB structures determined by NMR, we used the first conformer in the coordinate file to construct the coarse-grained model in all cases except BBL, where only one NMR structure was available. The same choice of conformation was used by Artem et al. in their recent work (30).

RESULTS AND DISCUSSION

Folding simulation by all-atom molecular dynamics using the implicit solvation model

We performed simulations with the AMBER (23) package using the generalized Born model (31) and surface area (GB/SA) to account for solvation effects (see Models and Methods for details). After energy minimization and equilibration of the fully extended starting conformation, four trajectories, each lasting for 1 μs, with different initial speed assignments, at 300 K, were generated. The folding event was observed in the third trajectory only, which was then extended to 2 μs, and the following analysis was based on that trajectory.

Folding event

Starting from a fully extended chain, the protein underwent fast collapse within 50 ps (Fig. 1, *inset*). This phenomenon has been observed in many folding simulation studies at similar timescales (3,8,10,32). In the 300 ns after the collapse, there was a large conformational adjustment during which the RMSD increased from <6 Å to >9 Å (Fig. 1). The RMSD decreased to 4 Å at ~300 ns, indicating that the structure was near the native state. After that, small structural readjustments lasted for ~500 ns. At ~900 ns, the protein reached the native state and stayed in the basin for the rest of the simulation time. The best simulated structure had a minimal main-chain RMSD of 2.6 Å (aligned with the NMR structure in Fig. 2). The simulated structure closely resembled the key features of the NMR structure: the helix was well formed, and the two strands were aligned in the correct direction. Despite the similarity in global fold and secondary structures, small differences in local structure were observed: for the simulated structure, the helix was slightly bent and the N-terminal formed a hairpin due to backbone-backbone hydrogen bonds. These small differences may be caused by the implicit solvation model used in this study and the possible structural fluctuations.

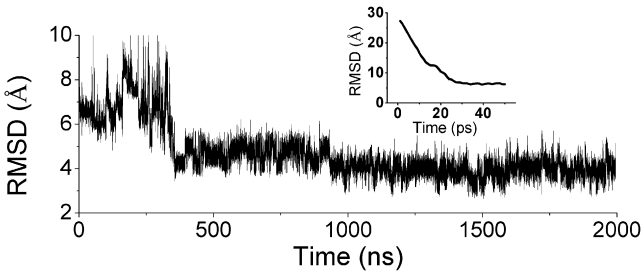

FIGURE 1 RMSD change of the main-chain atoms in the third trajectory of the all-atom simulation of DS119. (*Inset*) RMSD of the first 50 ps, illustrating fast collapse of the extended conformation. The structure comparison and RMSD calculation were done with the regular secondary structures assigned in the NMR structure of DS119.

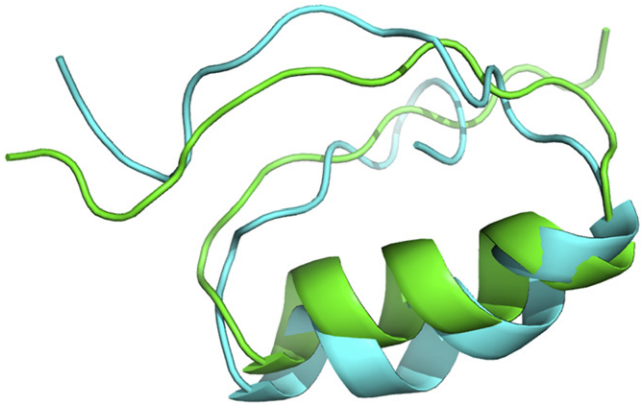

FIGURE 2 Alignment of the best simulated structure (*cyan*) and the NMR structure (*green*, PDB code 2KI0). The main-chain RMSD is 2.6 Å.

Considering that the folding event was observed in one of the four trajectories, the folding time estimated from our simulations was at least at the microsecond level, which is comparable to that of many fast-folding proteins (2). It is noteworthy that most fast-folding proteins are either all-α proteins or proteins containing antiparallel β-sheets only (see Kubelka et al. (2)). A parallel β-sheet generally requires tertiary contacts that have longer sequence distance than an antiparallel β-sheet, which may result in a longer folding time. The fast folding behavior of DS119 observed here may be due to its simple topology, which smoothes the energy landscape and lowers the energy barrier between the folded and unfolded states.

Hydrophobic packing

We also monitored the change of the solvent-accessible surface area (SASA) of the hydrophobic core with time (Fig. S1). After the initial rapid decrease, the SASA increased in the conformation adjustment stage. After the protein had arrived at its near-native state, there were still relatively large fluctuations in SASA, which indicated that the two β-strands were not well formed during the simulation. The minimal SASA was ~400 Å^2, larger than the average SASA of the NMR structures (292 Å^2).

When DS119 was designed, a pair of tryptophan residues was introduced to prevent protein aggregation. The alignment of the two indole rings might be face to face according to the nuclear Overhauser effect distances (16). We also monitored the distance between the centers of the two indole rings during the folding process. In the NMR structure, the distance between the two indole rings is 3.6 Å. However, during the simulation, the distance was ~10 Å after the protein folded (Fig. S2). This result, together with the SASA analysis, indicated that the hydrophobic core of the protein was partially exposed during the simulation, which may stem from the inability of the implicit water model to account for hydrophobic interactions.

Folding pathway

After the initial fast collapse, the protein formed a compact structure that had a size similar to that of the NMR structure. The N-terminal of the helix then formed quickly within 50 ns. The helix further elongated in the subsequent 70 ns. The C-terminal of the helix was formed at 122 ns, accompanied by the formation of a hairpinlike structure of the helical region (Fig. 3). This conformation corresponded to the adjustment phase mentioned above. The helix was fully formed at ~250 ns. Considering the different scales of simulation time and real time, this is comparable to the 1.7-μs folding time of a naturally occurring helical peptide investigated experimentally (33). Although the helix formed quickly, the two strands were highly flexible. The N-terminal strand bent over initially at ~355 ns, which caused the sharp RMSD decrease seen in Fig. 1, and it took another 1.3 μs for the other strand to form a compact packing. The folding pathway can be described as an initial collapse followed by fast formation of the central helix and slow formation of the β-sheet. Compared with β-sheet formation, helix

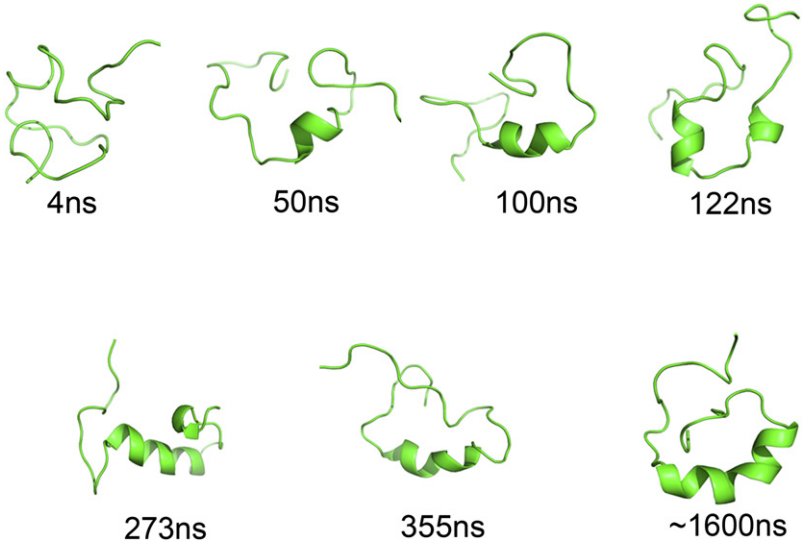

FIGURE 3 Representative snapshots of the DS119 folding trajectory.

formation is easier, because it is determined by local interactions. For the β-sheet, especially the two-stranded parallel β-sheet here, which lacks tight turns between its strands, the two strands are controlled by nonlocal interactions, and constraining the strands together will cause large entropy loss. Thus, the major rate-limiting step of the DS119 folding is the formation of the parallel β-sheet.

Coarse-grained Gō model simulations reveal that DS119 is an incipient downhill folder

Though all-atom simulations provide valuable information in atomic detail for the folding process, it is often difficult to construct the folding energy landscape with parallel pathways (34,35), as they are computationally expensive even with the implicit solvent models. As an alternative, coarse-grained models are more efficient for obtaining the folding energy landscape. The structure-based continuum Gō-like model (27,36) with coarse-grained chain representation has been widely used in protein-folding problems to facilitate understanding of the topology (37), the cooperativity (38–40), the folding-rate diversity in small proteins (41), and, recently, the coupling between folding and binding (42,43). In this section, we utilized the native-centric Gō model to characterize the folding energy landscape and cooperativity of DS119.

Contact order of DS119 and comparison with proteins of similar size

The overall topology is a significant factor influencing the folding kinetics of a protein. It has been pointed out that for proteins of similar chain length, contact order (CO) from residue contact counts and relative contact order (RCO) from atomic contact counts provide useful information on the folding rates (44,45) and cooperativity (38). We constructed the CO and RCO values for DS119 and other proteins of similar size (Table 1). To make the comparison more reasonable, we focused on proteins with similar chain lengths of ~35 residues, i.e., NTL9, DS119, Psbd41, HP36, and BBL. Although CO and RCO gave slightly different rank orders, a clear trend emerged: the CO and RCO values of NTL9 and DS119 were much higher than those of Psbd41, HP36 and BBL, suggesting that NTL9 and DS119 were more cooperative than the latter proteins.

TABLE 1 Thermodynamic properties and contact order of C_α-chain model proteins

Protein	CO*	RCO†	T_m‡	$\Delta H_{vH}/\Delta H_{cal}$
NTL9	0.412	0.213	1.045	0.76
CI2	0.348	0.156	0.98	0.83
E3BD F166W	0.278	0.140	0.925	0.61
BBL H166W	0.278	0.112	0.875	0.55
DS119	0.414	0.206	0.87	0.53
Psbd41	0.270	0.114	0.83	0.45
HP36	0.239	0.112	0.78	0.46
BBL	0.272	0.095	0.75	0.37

CO, contact order; RCO, relative CO.
*CO calculated as in Wallin and Chan (59), using only residue contact counts.
†RCO computed using atomic contact counts according to the definition of Plaxco et al. (45).
‡Transition temperature, T_m, determined from the peak of the heat capacity in Fig. 4.

Calorimetric cooperativity criterion

Thermodynamic simulations of the coarse-grained Gō model allowed us to explore the calorimetric cooperativity of DS119 by comparing it with proteins of similar chain length. The heat capacities of six protein fragments differed remarkably in peak height and half-peak width, from the sharp, narrow curve of NTL9 to the flat, broad curve of BBL. DS119 lay in the middle in terms of both sharpness and width (Fig. 4).

Although a peak in the heat capacity curve was not a sufficient condition for thermodynamic cooperativity and two-state folding (46), we constructed the calorimetric criterion, $\Delta H_{vH}/\Delta H_{cal}$, which has been widely used in protein-folding experiments and simulations. In this expression, the calorimetric enthalpy is defined as $\Delta H_{cal} = \int dT C_p(T)$ and the van 't Hoff enthalpy is defined as $\Delta H_{vH} = 2\sqrt{RT_m^2 C_p(T_m)}$, where T_m is the transition temperature determined from the peak of the heat capacity (Fig. 4) and $C_p(T_m)$ is the corresponding heat capacity value (47). The ratio $\Delta H_{vH}/\Delta H_{cal}$ was calculated for eight protein models (Table 1). Baseline was constructed by following the protocol proposed by Artem et al. (30). We found that CI2 was the most cooperative and BBL the least cooperative among the protein fragments studied. DS119 showed a moderate value of $\Delta H_{vH}/\Delta H_{cal}$, 0.53, indicating that it was less cooperative than CI2 but more cooperative than BBL (Table 1). The rank order of cooperativity determined by $\Delta H_{vH}/\Delta H_{cal}$ was CI2 > NTL9 > E3BD F166W > BBL H166W >

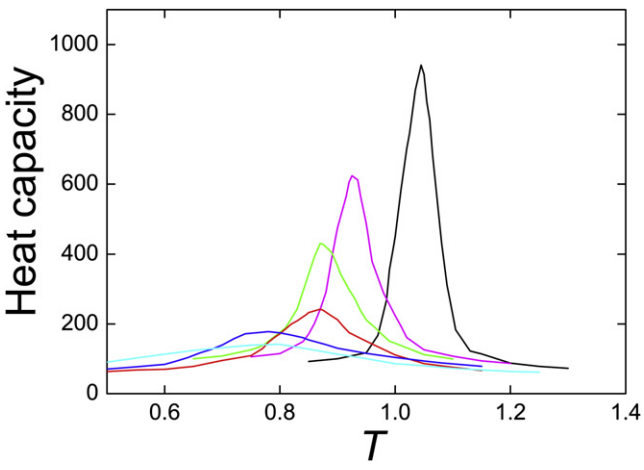

FIGURE 4 Heat capacities of the coarse-grained protein fragments NTL9 (*black*), E3BD F166W (*magenta*), BBL H166W (*green*), DS119 (*red*), HP36 (*blue*), and BBL (*cyan*). The temperature was measured in units of ε/R, where ε is the interaction scale used in the model potential.

DS119 > HP36 > Psbd41 > BBL. It has been pointed out (30) that when desolvation effects and/or the many-body effects are taken into consideration, the folding cooperativity of the models increases and the rank order for models with similar $\Delta H_{vH}/\Delta H_{cal}$ values may change. This rank order of folding cooperativity, especially the rank location of DS119, will be further discussed below.

The sigmoidal transitions

Sigmoidal transition of average energy or folding fraction as a function of temperature is a feature of a cooperative two-state transition. We compared the temperature variations of potential energy $\langle \tilde{E}(T) \rangle$ and the fractional number of native contacts $\langle Q(T) \rangle$ for six proteins: NTL9, E3BD F166W, BBL H166W, DS119, HP36, and BBL (Fig. 5). The normalized potential energy is defined as in Badasyan et al. (30):

$$\langle \tilde{E}(T) \rangle \equiv \frac{\langle E(T) \rangle - \langle E(T) \rangle_N}{\langle E(T) \rangle_D - \langle E(T) \rangle_N}, \quad (1)$$

where $\langle E(T) \rangle$ is the average potential energy at temperature T. $\langle E(T) \rangle_N$ and $\langle E(T) \rangle_D$ are the corresponding average potential energy values for the native state and denatured state, respectively, and are determined from the low- and high-temperature regions far away from the sigmoidal transition. Both $\langle \tilde{E}(T) \rangle$ and $\langle Q(T) \rangle$ showed that the sigmoidal transition of NTL9 was significantly sharper than that of BBL. DS119 exhibited a moderate sigmoidal transition between NTL9 and BBL. The rank order of sharpness of sigmoidal transition, CI2 > NTL9 > E3BD F166W > BBL H166W > DS119 > HP36 > BBL, was consistent with the rank order of folding cooperativity obtained from the calorimetric criterion.

Conformation distribution shifts with temperature

The population distribution of the fractional number of native contacts, Q, distinguished folding cooperativity among NTL9, DS119, and BBL (Fig. 6). For the most cooperative model, NTL9, the population of the native (or denatured) state decreased (or increased) as temperature increased, whereas the relative location of the native and denatured states remained unchanged (Fig. 6 *a*). Conversely, for the least cooperative model, BBL, only one populated state was identified, and the location of this state shifted with temperature, i.e., from the denatured-like state to the nativelike state, when temperature decreased (Fig. 6 *b*), which is consistent with the experimental observation that the folding of BBL is downhill (48). DS119, behaving like NTL9, exhibited two states in the conformation distribution, one identified as the native state ($Q \sim 0.65$) and the other as the denatured state ($Q \sim 0.25$) (Fig. 6 *c*). The location of the native and denatured states shifted slightly with temperature (i.e., when simulation temperature increased from 0.75 to 0.85, Q of the folded state decreased from 0.66 to 0.62; when temperature decreased from 0.95 to 0.85, Q of the denatured state increased from 0.21 to 0.30) (Fig. 6 *c*). Another feature of DS119 was the highly populated states with intermediate Q located between the native and denatured states near the transition temperature. Thus, DS119 was a two-state-like folder with features of a downhill folder.

Free-energy profiles

From a comparison of the free-energy profiles of the CI2, DS119, and BBL models under the transition temperatures, it was clear that these three models showed distinct free-energy barrier heights (Fig. 7). CI2 possessed the highest barrier of 3.5 k_BT and well-separated folded and unfolded states. BBL, on the other hand, showed no energy barrier along the free-energy profile, with only one free-energy minimum. DS119 showed a broad and flat free-energy landscape with an energy barrier of ~0.5 k_BT, and the locations corresponding to the folded and unfolded states were ambiguous. The heights of the overall free-energy barriers followed the rank order of folding cooperativity, CI2 > DS119 > BBL. It has been suggested that lack of significant free-energy barrier (~2 k_BT) along the folding reaction coordinate results in downhill folding (49). More exactly, it is proposed, folding with a low but nonzero free-energy barrier indicates incipient downhill folding (50), such as the folding behavior of DS119 discussed here. The kinetics and thermodynamics of downhill folders, or incipient downhill folders such as BBL (51), are very sensitive to the environment because of the low barriers involved.

Rank ordering of folding cooperativity

From a different perspective, in the above discussion, the rank order of folding cooperativity was given as CI2 > NTL9 > DS119 > BBL. This rank order, together with the free-energy barrier analysis, unambiguously suggested that the designed DS119 protein was an incipient downhill

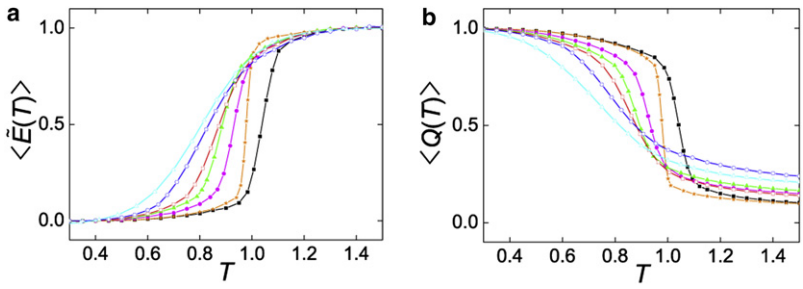

FIGURE 5 (*a*) Average normalized potential energy, $\langle \tilde{E}(T) \rangle$, and (*b*) fractional number of native contacts $\langle Q(T) \rangle$ as functions of temperature, T, for CI2 (*orange*), NTL9 (*black*), E3BD F166W (*magenta*), BBL H166W (*green*), DS119 (*red*), HP36 (*blue*), BBL (*cyan*). Data points were obtained from direct simulations at the given temperature. The average normalized potential energy, $\langle \tilde{E}(T) \rangle$, is defined by Eq. 1 in the text.

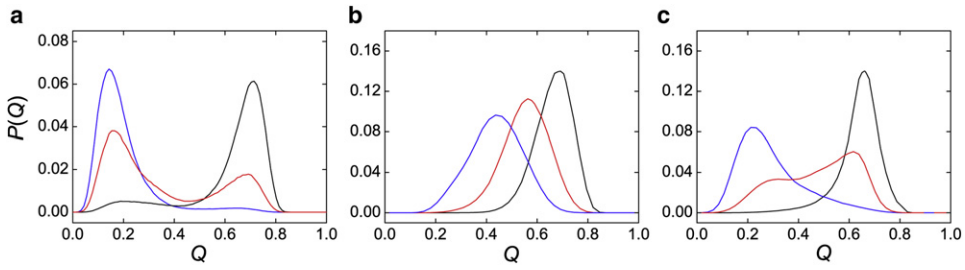

FIGURE 6 Conformation distribution shifts with temperature change. $P(Q)$ was the normalized population distribution obtained in simulations as a function of Q. The highest temperature, moderate temperature, and lowest temperature (*blue*, *red*, and *black*, respectively) were 1.10, 1.05, and 1.00 for NTL9 (*a*), 0.80, 0.70, and 0.60 for BBL (*b*), and 0.95, 0.85, and 0.75 for DS119 (*c*).

folder. The high CO value of DS119 was not enough to lead it to fold in a two-state way, which may originate from its distinct topology, consisting of two parallel β-strands and one α-helix. This suggests that when discussing folding cooperativity using only the contact order criterion, care should be taken when proteins with remarkably different topologies are encountered.

Folding kinetics from coarse-grained simulations

We monitored 400 folding trajectories of DS119 under transition temperature ($T = 0.85$) from 400 randomly chosen unfolded conformations ($Q < 0.05$) with different initial velocities. A typical trajectory is shown in Fig. S3. Consistent with results from the all-atom simulations, the protein collapsed first and the helix formed quickly and was stable during the entire folding process. Though contacts between the two strands and the helix were partially formed during the folding process, the two strands were rather dynamic and were not correctly packed toward each other, since no native contacts were formed between them. Compared with the C-terminal strand, the N-terminal strand was less dynamic, although fewer contacts existed between this strand and the helix. Taking into consideration all 400 folding trajectories, before reaching the folded state (defined as the minimum with the greater Q from the free-energy profile of Fig. 7), the average contacts within the α-helix, between the α-helix and the N-terminal β-strand, between the α-helix and the C-terminal β-strand, and between the two β-strands were 60%, 30%, 23%, and 5% formed, respectively. Thus, correct packing of the two strands, in particular, packing of the C-terminal strand, was the rate-limiting step. The same conclusion was obtained from the all-atom simulations.

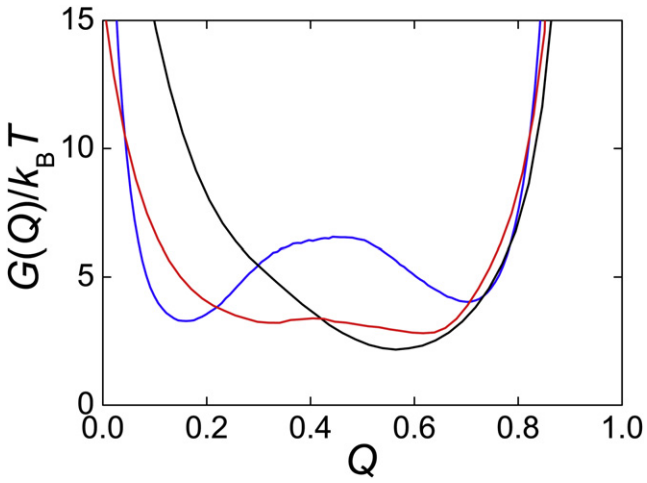

FIGURE 7 Free-energy profiles for CI2 (*blue*), DS119 (*red*), and BBL (*black*) under transition temperatures. Free energy was calculated using $G(Q)/k_BT = -\ln[P(Q)]$, where Q is the fractional number of native contacts and $P(Q)$ is the normalized population distribution obtained in simulations as a function of Q.

The folding rate prediction for DS119

Computer simulations provide valuable information on the kinetics of protein folding, irrespective of whether an all-atom or coarse-grained model is used. The Gō-like model has been widely used in understanding the folding kinetics of small proteins, and the folding rates from simulations correlate remarkably well with experimentally determined folding rates (41,52). On the basis of this correlation, we predicted the folding rate of DS119, whose kinetic data have not yet been reported. We first constructed a data set of seven small proteins that have been well studied both experimentally (53–58) and theoretically (52,59), then simulated their folding rates under transition midpoints. For the seven proteins we studied here, the simulated folding rates correlated well with the experimentally determined folding rates (Fig. 8). From this correlation, we were able to predict the folding rate of DS119 from the fitting line, and the predicted value was ~10^5 s^{-1}. This result was in agreement with the all-atom simulations.

It is notable that DS119, like many designed proteins, exhibit fast folding behavior. As most protein design methods do not take account of folding dynamics explicitly, the reasons for the fast folding property should be examined. One possible reason is that most designed proteins are relatively small in size and have simple topology, thus requiring fewer distant contacts and less structural rearrangement during folding. For example, the 73-residue three-helix bundle protein α_3D folds at a minimal time of 3.2 $\pm$ 1.2 μs (60). FSD-1ss, another 29-residue designed protein with one helix and one antiparallel β-sheet, exhibits microsecond folding kinetics near its T_m (61). In fact, several studies have revealed that folding time is positively correlated with protein size (62–65). Conversely, the 93-residue protein

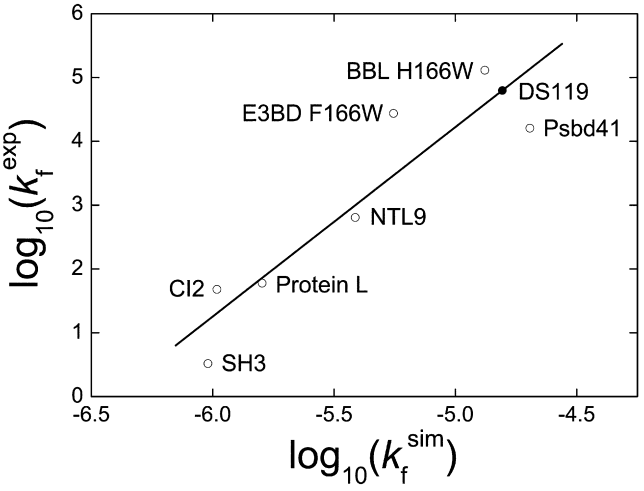

FIGURE 8 Correlation between simulated folding rates and experimentally determined folding rates. The correlation coefficient was 0.92. Simulated folding rates were estimated by calculating the mean first-passage time for folding transition from 400 randomly chosen unfolded states ($Q < 0.05$) using the coarse-grained Gō model under transition temperatures, T_m. The experimental folding rates were collected from the literature (53–58). The predicted folding rate for DS119 from the fitting line is denoted by the solid circle.

Top7 (21), consisting of two helices and five strands, was found to fold in a complex way and its slow folding rate was estimated to approach 1 s^{-1}. The other possible reason might come from currently used design methods, as most computational methods generally result in well-packed hydrophobic cores. As function design is relatively difficult now, most protein design methods focus on designing specific tertiary structures by using protein stability as the objective function. With this stability requirement, the designed proteins—unlike natural proteins, which are also under functional constraints—generally have good hydrophobic packing and stable secondary structures. As the hydrophobic effect is dominant in protein folding, a well-packed core may speed up the folding process. For example, by optimizing the hydrophobic content of the GA module of an albumin-binding domain, Feng Gai and co-workers were able to increase the folding rate up to six folds (66).

CONCLUSION

In this study, we used all-atom molecular dynamics simulations and a coarse-grained model to study the folding pathway and energy landscape of a de novo designed $\beta\alpha\beta$ protein, DS119. From all-atom simulations, we successfully observed the folding event with a minimal RMSD of 2.6 Å relative to the NMR structure. Different cooperativity studies using the coarse-grained model, i.e., calorimetric criterion, sigmoidal transitions, conformation distribution shifts, and free-energy profiles, all indicate that DS119 is an incipient downhill folder, which folds more cooperatively than a downhill folder such as BBL but less cooperatively than a two-state folder such as CI2. This may reflect a balance between the two structural elements of the $\beta\alpha\beta$ motif: the rapidly formed α helix and the slowly formed parallel β-sheet. Folding time estimated from all-atom simulations and the coarse-grained model were all at the microsecond level. This makes DS119 among the fast folders, which, to the best of our knowledge, is the only one with the unique parallel β-sheet. Further experimental studies are expected to verify this prediction.

SUPPORTING MATERIAL

Three figures and a table are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(09)01625-7.

This work was supported in part by grants from the Ministry of Science of Technology of China (2009CB918500 to Z.L. and L.L.) and the National Natural Science Foundation of China (20228306 and 90103029 to L.L., and 10721403 to L.L. and Z.L.).

REFERENCES

1. Duan, Y., and P. A. Kollman. 1998. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*. 282:740–744.
2. Kubelka, J., J. Hofrichter, and W. A. Eaton. 2004. The protein folding "speed limit". *Curr. Opin. Struct. Biol.* 14:76–88.
3. Lei, H., X. Deng, …, Y. Duan. 2008. The fast-folding HP35 double mutant has a substantially reduced primary folding free energy barrier. *J. Chem. Phys.* 129:155104.
4. Lei, H., and Y. Duan. 2007. Two-stage folding of HP-35 from ab initio simulations. *J. Mol. Biol.* 370:196–206.
5. Jayachandran, G., V. Vishal, and V. S. Pande. 2006. Using massively parallel simulation and Markovian models to study protein folding: examining the dynamics of the villin headpiece. *J. Chem. Phys.* 124:164902.
6. Zagrovic, B., C. D. Snow, …, V. S. Pande. 2002. Simulation of folding of a small α-helical protein in atomistic detail using worldwide-distributed computing. *J. Mol. Biol.* 323:927–937.
7. Lei, H., and Y. Duan. 2007. Ab initio folding of albumin binding domain from all-atom molecular dynamics simulation. *J. Phys. Chem. B*. 111:5458–5463.
8. Jang, S., E. Kim, …, Y. Pak. 2003. Ab initio folding of helix bundle proteins using molecular dynamics simulations. *J. Am. Chem. Soc.* 125:14841–14846.
9. Vila, J. A., D. R. Ripoll, and H. A. Scheraga. 2003. Atomically detailed folding simulation of the B domain of staphylococcal protein A from random structures. *Proc. Natl. Acad. Sci. USA*. 100:14812–14816.
10. Lei, H., C. Wu, …, Y. Duan. 2008. Folding processes of the B domain of protein A to the native state observed in all-atom ab initio folding simulations. *J. Chem. Phys.* 128:235105.
11. Jagielska, A., and H. A. Scheraga. 2007. Influence of temperature, friction, and random forces on folding of the B-domain of staphylococcal protein A: all-atom molecular dynamics in implicit solvent. *J. Comput. Chem.* 28:1068–1082.
12. Chowdhury, S., M. C. Lee, …, Y. Duan. 2003. Ab initio folding simulation of the Trp-cage mini-protein approaches NMR resolution. *J. Mol. Biol.* 327:711–717.
13. Paschek, D., H. Nymeyer, and A. E. García. 2007. Replica exchange simulation of reversible folding/unfolding of the Trp-cage miniprotein in explicit solvent: on the structure and possible role of internal water. *J. Struct. Biol.* 157:524–533.

14. Zhou, R. 2003. Trp-cage: folding free energy landscape in explicit water. *Proc. Natl. Acad. Sci. USA.* 100:13280–13285.

15. Li, H., R. Helling, …, N. Wingreen. 1996. Emergence of preferred structures in a simple model of protein folding. *Science.* 273:666–669.

16. Liang, H., H. Chen, …, L. Lai. 2009. De novo design of a $\beta\alpha\beta$ motif. *Angew. Chem. Int. Ed. Engl.* 48:3301–3303.

17. Butterfoss, G. L., and B. Kuhlman. 2006. Computer-based design of novel protein structures. *Annu. Rev. Biophys. Biomol. Struct.* 35:49–65.

18. Kuhlman, B., G. Dantas, …, D. Baker. 2003. Design of a novel globular protein fold with atomic-level accuracy. *Science.* 302: 1364–1368.

19. Dahiyat, B. I., and S. L. Mayo. 1997. De novo protein design: fully automated sequence selection. *Science.* 278:82–87.

20. Summa, C. M., M. M. Rosenblatt, …, W. F. DeGrado. 2002. Computational de novo design, and characterization of an A(2)B(2) diiron protein. *J. Mol. Biol.* 321:923–938.

21. Watters, A. L., P. Deka, …, D. Baker. 2007. The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell.* 128:613–624.

22. Zhang, Z., and H. S. Chan. 2009. Native topology of the designed protein Top7 is not conducive to cooperative folding. *Biophys. J.* 96:L25–L27.

23. Case, D. A., T. E. Cheatham, 3rd, …, R. J. Woods. 2005. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26:1668–1688.

24. Duan, Y., C. Wu, …, P. Kollman. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24: 1999–2012.

25. Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of Cartesian equations of motion of a system with constraints: molecular dynamics of N-alkanes. *J. Comput. Phys.* 23:327–341.

26. Legrand, S. M., and K. M. Merz. 1993. Rapid approximation to molecular surface area via the use of Boolean logic and look-up tables. *J. Comput. Chem.* 14:349–352.

27. Go, N. 1983. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* 12:183–210.

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

29. Liu, Z. R., and H. S. Chan. 2005. Solvation and desolvation effects in protein folding: native flexibility, kinetic cooperativity and enthalpic barriers under isostability conditions. *Phys. Biol.* 2:S75–S85.

30. Badasyan, A., Z. Liu, and H. S. Chan. 2009. Interplaying roles of native topology and chain length in marginally cooperative and noncooperative folding of small protein fragments. *Int. J. Quantum Chem.* 109:3482–3499.

31. Onufriev, A., D. Bashford, and D. A. Case. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins.* 55:383–394.

32. Lee, I. H., S. Y. Kim, and J. Lee. 2009. Dynamic folding pathway models of the villin headpiece subdomain (HP-36) structure. *J. Comput. Chem.* 31:57–65.

33. Mukherjee, S., P. Chowdhury, …, F. Gai. 2008. Folding kinetics of a naturally occurring helical peptide: implication of the folding speed limit of helical proteins. *J. Phys. Chem. B.* 112:9146–9150.

34. Dill, K. A., and H. S. Chan. 1997. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* 4:10–19.

35. Bryngelson, J. D., J. N. Onuchic, …, P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins.* 21:167–195.

36. Ueda, Y., H. Taketomi, and N. Go. 1978. Studies on protein folding, unfolding, and fluctuations by computer simulation. 2. 3-Dimensional lattice model of lysozyme. *Biopolymers.* 17:1531–1548.

37. Koga, N., and S. Takada. 2001. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J. Mol. Biol.* 313:171–180.

38. Badasyan, A., Z. Liu, and H. S. Chan. 2008. Probing possible downhill folding: native contact topology likely places a significant constraint on the folding cooperativity of proteins with approximately 40 residues. *J. Mol. Biol.* 384:512–530.

39. Kaya, H., and H. S. Chan. 2003. Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: how adequate is native-centric topological modeling? *J. Mol. Biol.* 326:911–931.

40. Cho, S. S., P. Weinkam, and P. G. Wolynes. 2008. Origins of barriers and barrierless folding in BBL. *Proc. Natl. Acad. Sci. USA.* 105: 118–123.

41. Ferguson, A., Z. Liu, and H. S. Chan. 2009. Desolvation barrier effects are a likely contributor to the remarkable diversity in the folding rates of small proteins. *J. Mol. Biol.* 389:619–636.

42. Turjanski, A. G., J. S. Gutkind, …, G. Hummer. 2008. Binding-induced folding of a natively unstructured transcription factor. *PLOS Comput. Biol.* 4:e1000060.

43. Huang, Y., and Z. Liu. 2009. Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the ''fly-casting'' mechanism. *J. Mol. Biol.* 393:1143–1159.

44. Baker, D. 2000. A surprising simplicity to protein folding. *Nature.* 405:39–42.

45. Plaxco, K. W., K. T. Simons, and D. Baker. 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* 277:985–994.

46. Knott, M., and H. S. Chan. 2006. 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.* 65:373–391.

47. Kaya, H., and H. S. Chan. 2000. Polymer principles of protein calorimetric two-state cooperativity. *Proteins.* 40:637–661.

48. Sadqi, M., D. Fushman, and V. Muñoz. 2006. Atom-by-atom analysis of global downhill protein folding. *Nature.* 442:317–321.

49. Gruebele, M. 2005. Downhill protein folding: evolution meets physics. *C. R. Biol.* 328:701–712.

50. Liu, F., D. G. Du, …, M. Gruebele. 2008. An experimental survey of the transition between two-state and downhill protein folding scenarios. *Proc. Natl. Acad. Sci. USA.* 105:2369–2374.

51. Garcia-Mira, M. M., M. Sadqi, …, V. Muñoz. 2002. Experimental identification of downhill protein folding. *Science.* 298:2191–2195.

52. Chavez, L. L., J. N. Onuchic, and C. Clementi. 2004. Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc.* 126:8426–8432.

53. Jackson, S. E., and A. R. Fersht. 1991. Folding of chymotrypsin inhibitor-2. 1. Evidence for a 2-state transition. *Biochemistry.* 30:10428–10435.

54. Scalley, M. L., and D. Baker. 1997. Protein folding kinetics exhibit an Arrhenius temperature dependence when corrected for the temperature dependence of protein stability. *Proc. Natl. Acad. Sci. USA.* 94:10636–10640.

55. Spector, S., and D. P. Raleigh. 1999. Submillisecond folding of the peripheral subunit-binding domain. *J. Mol. Biol.* 293:763–768.

56. Cobos, E. S., V. V. Filimonov, …, J. C. Martínez. 2003. A thermodynamic and kinetic analysis of the folding pathway of an SH3 domain entropically stabilised by a redesigned hydrophobic core. *J. Mol. Biol.* 328:221–233.

57. Horng, J. C., V. Moroz, and D. P. Raleigh. 2003. Rapid cooperative two-state folding of a miniature α-β protein and design of a thermostable variant. *J. Mol. Biol.* 326:1261–1270.

58. Ferguson, N., T. D. Sharpe, …, A. R. Fersht. 2005. Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family. *J. Mol. Biol.* 353:427–446.

59. Wallin, S., and H. S. Chan. 2006. Conformational entropic barriers in topology-dependent protein folding: perspectives from a simple native-centric polymer model. *J. Phys. Condens. Matter.* 18:S307–S328.

60. Zhu, Y., D. O. Alonso, …, F. Gai. 2003. Ultrafast folding of α3D: a de novo designed three-helix bundle protein. *Proc. Natl. Acad. Sci. USA.* 100:15486–15491.

61. Sadqi, M., E. de Alba, …, V. Muñoz. 2009. A designed protein as experimental model of primordial folding. *Proc. Natl. Acad. Sci. USA.* 106:4127–4132.

62. Gutin, A. M., V. I. Abkevich, VI, and E. I. Shakhnovich. 1996. Chain length scaling of protein folding time. *Phys. Rev. Lett.* 77:5433–5436.

63. Wolynes, P. G. 1997. Folding funnels and energy landscapes of larger proteins within the capillarity approximation. *Proc. Natl. Acad. Sci. USA.* 94:6170–6175.

64. Li, M. S., D. K. Klimov, and D. Thirumalai. 2004. Thermal denaturation and folding rates of single domain proteins: size matters. *Polymer (Guildf.).* 45:573–579.

65. Naganathan, A. N., and V. Muñoz. 2005. Scaling of folding times with protein size. *J. Am. Chem. Soc.* 127:480–481.

66. Zhu, Y., X. Fu, T. Wang, A. Tamura, S. Takada…, 2004. Guiding the search for a protein's maximum rate of folding. *Chem. Phys.* 307: 99–109.