

# On the Origins of the Weak Folding Cooperativity of a Designed $\beta\beta\alpha$ Ultrafast Protein FSD-1

Chun Wu<sup>1</sup>, Joan-Emma Shea<sup>1,2\*</sup>

**1** Department of Chemistry and Biochemistry, University of California, Santa Barbara, Santa Barbara, California, United States of America, **2** Department of Physics, University of California, Santa Barbara, Santa Barbara, California, United States of America

## Abstract

FSD-1, a designed small ultrafast folder with a  $\beta\beta\alpha$  fold, has been actively studied in the last few years as a model system for studying protein folding mechanisms and for testing of the accuracy of computational models. The suitability of this protein to describe the folding of naturally occurring  $\alpha/\beta$  proteins has recently been challenged based on the observation that the melting transition is very broad, with ill-resolved baselines. Using molecular dynamics simulations with the AMBER protein force field (ff96) coupled with the implicit solvent model (IGB = 5), we shed new light into the nature of this transition and resolve the experimental controversies. We show that the melting transition corresponds to the melting of the protein as a whole, and not solely to the helix-coil transition. The breadth of the folding transition arises from the spread in the melting temperatures (from  $\sim 325$  K to  $\sim 302$  K) of the individual transitions: formation of the hydrophobic core,  $\beta$ -hairpin and tertiary fold, with the helix formed earlier. Our simulations initiated from an extended chain accurately predict the native structure, provide a reasonable estimate of the transition barrier height, and explicitly demonstrate the existence of multiple pathways and multiple transition states for folding. Our exhaustive sampling enables us to assess the quality of the Amber ff96/igb5 combination and reveals that while this force field can predict the correct native fold, it nonetheless overstabilizes the  $\alpha$ -helix portion of the protein ( $T_m = \sim 387$  K) as well as the denatured structures.

**Citation:** Wu C, Shea J-E (2010) On the Origins of the Weak Folding Cooperativity of a Designed  $\beta\beta\alpha$  Ultrafast Protein FSD-1. PLoS Comput Biol 6(11): e1000998. doi:10.1371/journal.pcbi.1000998

**Editor:** Victor Muñoz, Centro de Investigaciones Biológicas, Spain

**Received:** June 11, 2010; **Accepted:** October 14, 2010; **Published:** November 18, 2010

**Copyright:** © 2010 Wu, Shea. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The David and Lucile Packard Foundation (JES), National Science Foundation grants MCB-0642088 (JES) are gratefully acknowledged. Simulations were performed on the Lonestar cluster at Texas Advanced Computing Center (LRAC MCA055027 to JES) and the Dell cluster of CNSI. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: shea@chem.ucsb.edu

## Introduction

Small “ultrafast” folders (proteins that fold on the order of microseconds), both naturally occurring and designed, have received considerable attention in the last few years. These proteins have the singular advantage of being computationally tractable, thus bridging the gap between experimental and *in silico* studies. They permit not only a testing of the accuracy of computational force fields, but also (in the event that the force field proves to be adequate) an assessment of protein folding theories. Folding mechanisms and the possibility of multiple folding pathways, both of which can be difficult to determine from standard bulk measurements, can be resolved through an analysis of molecular dynamics simulations.

FSD-1 is a 28 residue designed ultrafast folder with a  $\beta\beta\alpha$  (hairpin/helix) fold [1]. The protein has a well-defined hydrophobic core, and unlike the more commonly studied  $\beta\beta\alpha$  BBA5 protein (which has a D-proline at the  $\beta$ -turn position), only contains naturally occurring residues. The folding time of FSD-1 has not been reported, but the folding kinetics of a close analog, FSD-1ss (involving substitution of two non-natural aromatic residues at positions 6 and 26) have been monitored using laser-induced temperature-jump spectroscopy (Table 1) [2]. This modified protein displayed two folding phases ( $\tau_1 \sim 150$  ns and  $\tau_2 \sim 4.5$   $\mu$ s) at 322 K, placing FSD-1ss at the top range of known ultrafast folders. Although the N-terminal region FSD-1ss (residues

16–23) adopts a loose U-shape rather than the tight  $\beta$ -hairpin seen in FSD-1, the overall tertiary structure of FSD-1ss is similar to FSD-1 with C $\alpha$  Root Mean Square Deviation (C $\alpha$ -RMSD) of only 2.2 Å. One can therefore expect that FSD-1 folds on similar microseconds timescales as FSD-1ss.

The thermal unfolding of FSD-1, as determined from Circular Dichroism (CD) [1] and Differential Scanning Calorimetry (DSC) [3], is reversible, but weakly cooperative, with a relatively low melting temperature ( $T_m = 315$  K). Mayo and co-workers, who designed this protein, attributed the observed transition to the melting of the entire protein. Feng et al. [3] have however recently challenged this interpretation and have proposed that the broad transition, which lacks clearly defined folding or unfolding baselines, in fact reflects only the melting of the  $\alpha$ -helical segment (residues 14–26) of the protein. Should this interpretation be correct, then one would have to reconsider FSD-1 as a model system for studying the folding of  $\alpha/\beta$  proteins.

Prior simulations of the FSD-1 protein have met with mixed (and sometimes conflicting) results, and have not provided a clear picture of the nature of the melting transition. For instance, a replica exchange molecular dynamics (REMD) simulation of FSD-1 in explicit solvent, using the Amber protein force field (ff03), TIP3P water and the NVT ensemble, predicted a melting temperature of 411.59 K [4],  $\sim 100$  K higher than the experimental value of 315 K. Simulations performed using a different force field, water model and simulation protocol (OPLS-AA/L

## Author Summary

The protein folding process, in which a linear chain of amino acids reaches its biologically active three-dimensional shape, is fundamental to life. Small “ultrafast” folders, proteins that fold in microseconds, have received considerable attention, because these proteins serve as model systems for the folding of larger proteins, and thus permit a testing of the accuracy of computational models as well as an assessment of protein folding theories. FSD-1, a designed small ultrafast folder with a  $\beta\beta\alpha$  fold, has been actively studied in the last few years as a model system for mixed  $\alpha/\beta$  fold proteins. The suitability of this protein to describe the folding of naturally occurring proteins has however recently been challenged based on the observation that the melting transition is very broad, with ill-resolved baselines. Prior simulations have not been successful in providing an interpretation of this broad melting transition. In the present study, our extensive molecular dynamics simulations using the AMBER protein force field (ff96) coupled with the implicit solvent model (IGB=5) shed new light on the nature of the folding transition of this protein, as well as reveal the strengths and weaknesses of the force field in predicting the thermodynamics and kinetics of folding.

2001 force field, TIP4P water model and NPT ensembles with REMD simulation) lead to a melting temperature that is 84K higher than what is observed experimentally [3]. These unsatisfactory results may be the result of inadequate force fields, or/and due to insufficient sampling. In order to overcome possible sampling issues related to the use of explicit solvents, a number of groups have turned to coarse-grained protein models [5] or to implicit water models. Pak and co-workers, using the CHARMM 19 force field in conjunction with a GB solvation model [6] witnessed the folding of FSD-1 to a structure 2.56 Å C $\alpha$ -RMSD from the NMR structure at 440 K in 15 ns, i.e. at a temperature well above the experimental  $T_m$  and with a folding time off by orders of magnitude. The authors had better success using replica exchange molecular dynamics (REMD) simulations and a newly modified version (param99MOD5) of AMBER 99 with GBSA implicit solvent [7], obtaining a computational melting temperature of  $\sim$ 309 K. However, the predictive power of these simulations is uncertain given that FSD-1 was used as a training peptide in the optimization of the force field. Finally, using a newly optimized force field in combination with the recently developed implicit solvent (IGB=5) [8], Lei and co-workers [9] were able to fold the double mutant of FSD-1 (FSDEY) [10], into its native state with high population ( $>$ 64.2%) and high fidelity (1.29 Å). However, the computationally generated heat capacities failed to produce a melting transition at 315K, and the melting of the helix

was observed at 360K which is higher than that of stable helical protein [11].

In the present paper, we investigate the folding of FSD-1 using the Amber ff96 protein force field combined with the implicit water solvent IGB=5. This combination has been shown by Dill and co-workers to have a good balance of  $\alpha/\beta$  propensity in the case of small peptides [12]. We have recently been able to use this combination to investigate the conformations adopted by natively disordered amyloid peptides (e.g. prion fragment [13] and amylin [14]) which can sample a variety of conformations including  $\alpha$ ,  $\beta$  and  $\alpha/\beta$ . Recent successes of this force field/implicit solvent combination include the successful folding of the 39-residue NTL9 protein [15] by the Pande group and the unfolding of the 64-residue protein L [16]. The majority of studies involving the ff96/igb5 model have focused on assessing the model’s reliability in generating accurate structural properties of proteins. Here, we present a thorough thermodynamic and kinetic analysis that enables, through a direct comparison with experimental data, an in-depth evaluation of the strength and weaknesses of this force field/implicit water model combination. Our simulations offer the first comprehensive interpretation of the broad transition at 315 K and resolve the issue of whether it corresponds to the melting of the protein or to an only helix-coil transition. Furthermore, our extensive simulations enable a thorough exploration of the energy landscape for folding, a structural characterization of transition state ensemble, and explicitly demonstrate the existence of multiple folding routes. Our simulations indicate that while ff96/igb5 can be reliably used to identify the folded state of the protein, as well as predict the thermodynamic order of formation of the structural elements in folding, there remain areas in which the protein force field/implicit water model needs to be improved. In particular, we find that ff96/igb5 overstabilizes both the denatured state and the helical portion of the protein.

## Results

### REMD simulation

**REMD simulations predict the correct native fold.** REMD simulations started from an extended chain conformation were performed as detailed in Methods section. 16 replicas ranging from 271 K to 475 K were used, each of a 1.25  $\mu$ s duration, leading to a cumulative simulation time of 20.0  $\mu$ s. The convergence of the REMD simulations was rigorously verified by a block analysis: the total sampling time of 1.25  $\mu$ s for the replica at 280 K were equally divided into five blocks, the population of the folded structure ( $<$ 3.0 Å to the NMR structure) were calculated for each block, and a good convergence was found during the 500 ns of the simulations (Text S1). All analyses were done over only the last 500 ns of the simulations.

To ascertain whether our simulations can predict the correct structure of FSD-1, we follow the typical two step protocol used in the structural prediction community: 1.) identification of the most populated structure (i.e. the lowest free energy structure) by a clustering method, 2.) comparison of this structure with the experimental structure. We used a pair-wise clustering algorithm (see Methods section) to pick up the centroid structure of the most populated structural family from the ensemble at 280 K. Fig. 1 shows the comparison between our predicted structure at 280 K and the NMR structure measured at 280 K. An excellent match is found only not between the over-all backbone but also between the side chains in the hydrophobic core. The important ramification of these simulations is that the Amber 96/IGB-5 combination is capable of identifying the correct native fold as the lowest free energy structure. This force field/ implicit water

**Table 1.** Folding times for three  $\beta\beta\alpha$  proteins FSD-1, FSD-1ss and BBA5 of similar size.

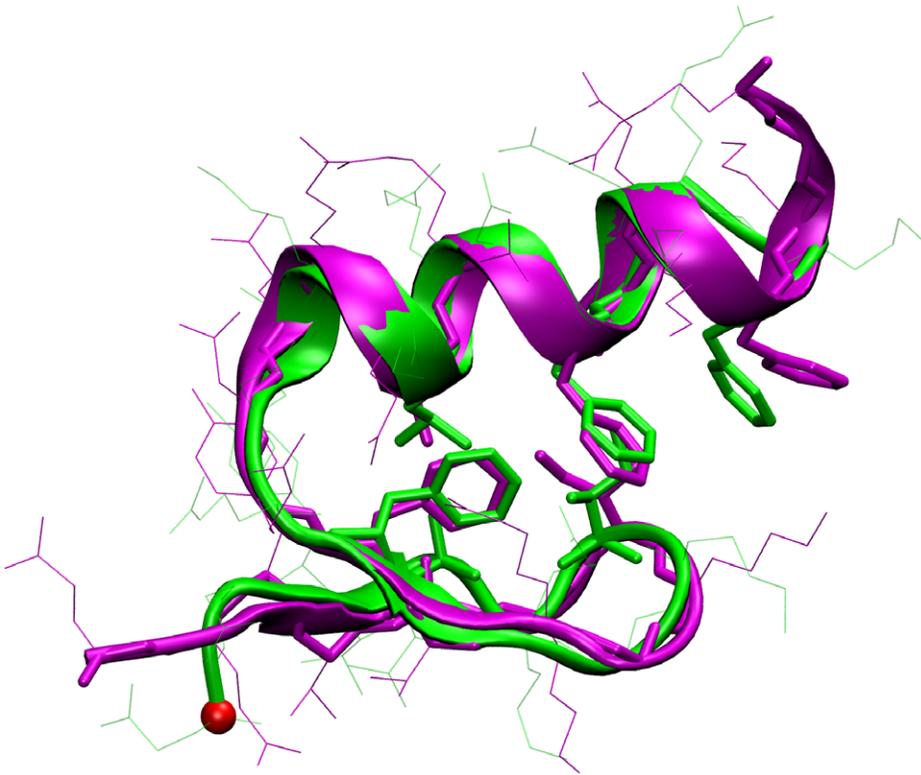
	FSD-1 <sup>a</sup>	FSD-1ss <sup>b</sup>	BBA5 <sup>c</sup>
This work	2.2 $\pm$ 0.7 $\mu$ s, 323 K	-	-
Experiments	-	4.5 $\mu$ s, 322 K	7.5 $\mu$ s, 298 K

<sup>a</sup>Based on the free energy barrier at 323 K.

<sup>b</sup>Ref. 2.

<sup>c</sup>Ref. 17.

doi:10.1371/journal.pcbi.1000998.t001



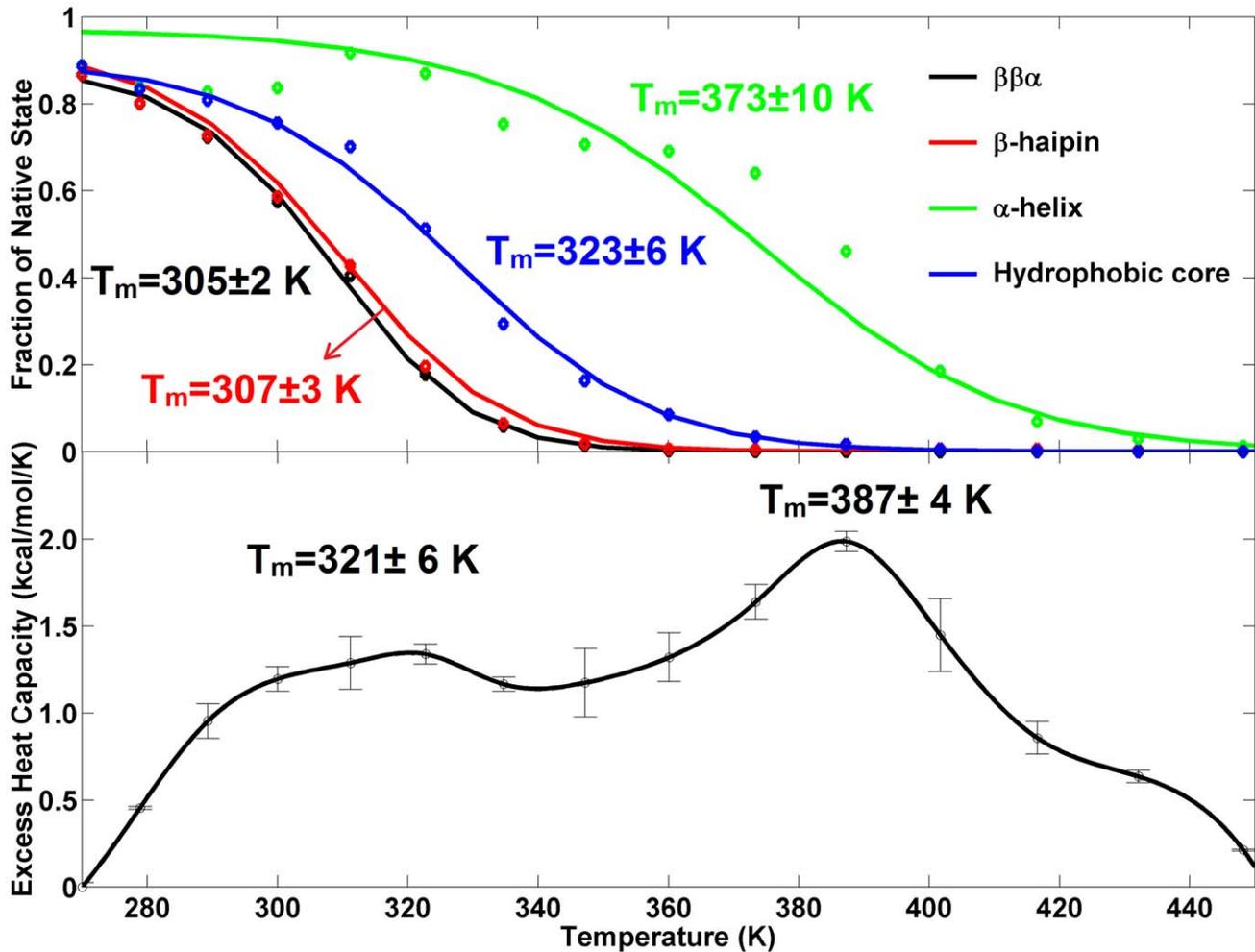
**Figure 1. Predicted structure of FSD-1 (Green) at 280 K from REMD initiated from an extended conformation.** It is 1.7 Å of C $\alpha$ -RMSD compared with the NMR structure at 280 K (purple, Protein Data Bank ID code: 1FSD).  
doi:10.1371/journal.pcbi.1000998.g001

combination could be very useful in structure prediction studies of proteins whose folds have not yet been experimentally characterized.

**Thermodynamics of folding.** The melting transition can be objectively analyzed from the heat capacity profile. The sharp slope at the lowest temperatures makes it difficult to determine the native state baseline necessary for getting the excess heat capacity profile. In our analysis, we set the absolute heat capacity value at the lowest temperature as the native state baseline. The resulting excess heat capacity profile (Fig. 2, bottom) has two broad peaks at temperature  $321 \pm 6$  K and  $387 \pm 4$  K indicative of two distinct melting transitions. The denaturation enthalpies for the two transitions are 70.0 kcal/mol and 122.0 kcal/mol, respectively.

The structural nature of the two transitions can be identified from the van't Hoff analysis (see Methods section) on the melting of the various structural components of FSD-1 (the C-terminal  $\alpha$ -helix, N-terminal  $\beta$ -hairpin, the hydrophobic core and the whole protein). The upper panel of Fig. 2 shows the fitted curves and the thermodynamic parameters are listed in Table 2. The curve fitting is good for all structural components except for the  $\alpha$ -helix. The lack of sharpness of the transitions in all curves indicates a weak cooperativity. The melting temperatures are  $305 \pm 2$  K,  $307 \pm 3$  K,  $323 \pm 6$  K and  $373 \pm 10$  K for the tertiary fold, the N-terminal  $\beta$ -hairpin, the hydrophobic core and the C-terminal  $\alpha$ -helix. Given the correspondence between the transitions temperatures from the excess heat capacity profile and the melting curves, we conclude that: (1) the first transition at lowest temperatures ( $321 \pm 6$  K) corresponds to the formation of the hydrophobic core ( $323 \pm 6$  K), the N-terminal  $\beta$ -hairpin ( $307 \pm 3$  K) and the tertiary fold ( $305 \pm 2$  K). (2) the spread ( $\sim 323$  K to  $\sim 305$  K) in the melting temperatures of the hydrophobic core,  $\beta$ -hairpin and tertiary fold

accounts for the broadness of the first transition; (3) the second transition at higher temperatures ( $387 \pm 4$  K) involves the formation of the  $\alpha$ -helix ( $373 \pm 10$  K). Since the melting temperature of even a very stable helix is less than 343 K [11], the melting temperature of the second transition appears to be overestimated, likely a result of an imbalance in the ff96/igb5 force field parameters. Nonetheless, given that the experimental CD spectrum at elevated temperatures (353 K) shows the presence of helical structure [3] in addition to random coils, it appears that the helix is more stable than the other structural elements (although certainly not to the extent seen in our simulations). Hence the relative stability order of the four structural components seen in our simulations should still be correct (i.e. the helix is still slightly more stable than the hydrophobic core and the helix-to-coil is the last step of the denaturation). In other words, the order of the melting temperatures for the four components ( $373 \pm 10$  K  $>$   $323 \pm 6$  K  $>$   $307 \pm 3$  K  $>$   $305 \pm 2$  K) suggests the following thermodynamic folding sequence: Initial formation of the C-terminal  $\alpha$ -helix and followed by the hydrophobic core and thereafter the simultaneous formation of the N-terminal  $\beta$ -hairpin and the tertiary fold. (Again, since the force field likely overstabilizes the  $\alpha$ -helix, the initial formation of the C-terminal  $\alpha$ -helix and the hydrophobic core should be more concurrent than is seen in simulation, with a more realistic temperature gap being much smaller ( $\ll 50$  K) than seen here. In terms of the van't Hoff thermodynamic parameters (i.e. the changes of enthalpy  $\Delta H_m$ , and heat capacity  $\Delta C_p$  at  $T_m$ ) for each structural component (see Table 2), the values are generally in good agreement with experiment ( $T_m = 314$  K,  $\Delta H_m = 10.4$  kcal/mol,  $\Delta C_p = 120.7$  cal/mol/K). The agreement is less satisfactory with regard to the melting temperature of the  $\alpha$ -helix ( $373 \pm 10$  K), consistent with our earlier observations, and the



**Figure 2. Thermodynamic features obtained from REMD simulations.** Top: The fitted melting curves of different structural elements. The original data points are indicated by diamonds and the obtained melting temperatures are noted. Bottom: The excess heat capacity as a function of temperature. Standard error is obtained by dividing data into two blocks for analysis. doi:10.1371/journal.pcbi.1000998.g002

heat capacity changes for the tertiary fold, ( $\Delta C_p = 336.0 \text{ cal/mol/K}$ ) and for the N-terminal  $\beta$ -hairpin ( $\Delta C_p = 229.7 \text{ cal/mol/K}$ ). Overall, the qualitative agreement between the simulations and the CD experiment supports the interpretation that the melting of these four structural components is involved in the denaturation transition.

Despite the qualitative agreement with the experimental CD data, comparison to experimental thermodynamic data shows a number of differences with our simulations that point to some deficiencies in the simulation protocol or/and the force field/implicit water model used. The primary discrepancy lies in the fact that we see two peaks in our theoretical heat capacity plots, while the DSC experiment [3] only shows one broad peak at a temperature 314 K with a denaturation enthalpy of 12–15 kcal/mol (Table 2). The experimental peak corresponds well with the first peak in our simulations in terms of location. The presence of a second high temperature peak (not seen in experiment) can be attributed to: 1) the high temperatures (up to 465K) used in the REMD simulations (in contrast, the temperature range in the DSC experiment was 283 K–353 K); 2) the force field/implicit water model related over-stabilization of the  $\alpha$ -helix which would lead to an artificially high melting temperature for this structural element. Tuning of the force field to better balance the relative

stabilities of the helical and sheet components might shift the second peak down in temperature so as to overlap with the first peak. The second difference to account for is the elevated denaturation enthalpy in the first peak (70.0 kcal/mol. vs. the experimental value of 14 kcal/mol). This could be due once again to deficiencies in the ff96/igb5 model, or to the fact that we used a Langevin thermostat with a low friction coefficient. To examine the latter, we performed the same REMD simulations, but this time with a Berendsen thermostat with a coupling constant of 2.0 ps (unpublished data). In this case, we obtain a similar heat capacity profile with two peaks (Text S1), but the native state baseline is well resolved and the denaturation enthalpy of the first peak is now 12.8 kcal/mol, in good agreement with experiment. Hence the thermostat appears to be the primary reason for the difficulty in defining the native state baseline and the overestimation of the denaturation enthalpy for the first peak. A detailed comparison between different thermostats will be reported in future work.

**Folding free energy plot and transition state.** The folding free energy surface at 323K (close to the melting temperature of  $321 \pm 6$  K identified from the heat capacity profile) is plotted as a function of the  $C\alpha$ -RMSD of FSD-1 in Fig. 3. The plot shows two basins, a narrow one centered at  $C\alpha$ -RMSD = 2.0 Å and a

**Table 2.** Calculated and experimental thermodynamic parameters.

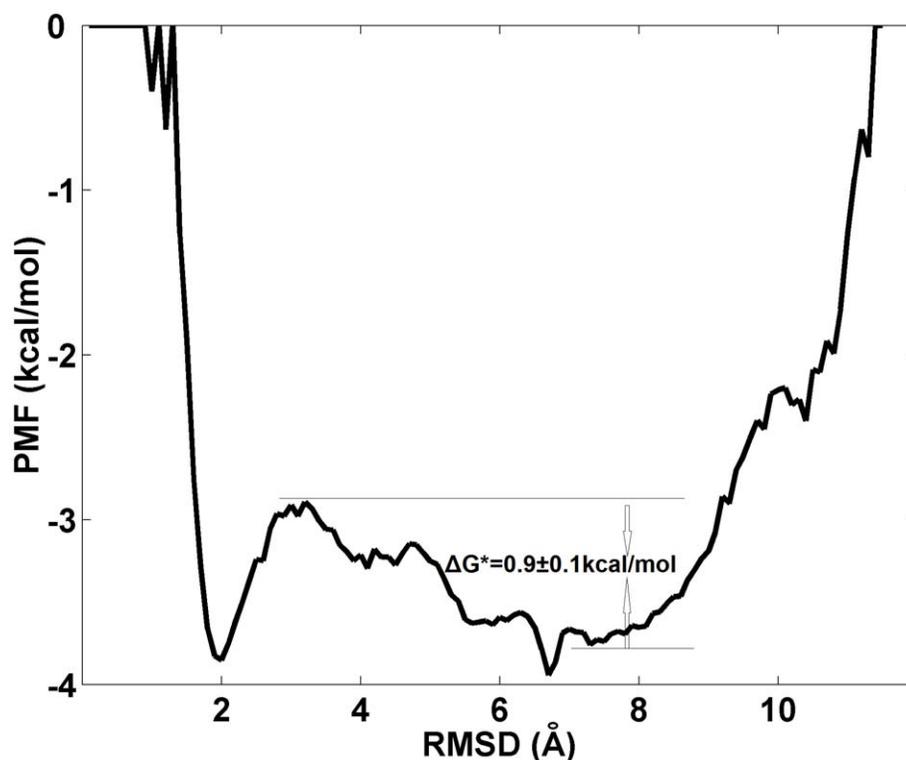
Measurement	Cutoff (Å)	T <sub>m</sub> (K)	ΔH <sub>m</sub> (kcal/mol)	ΔC <sub>p</sub> (cal/mol/K)
CD experiment <sup>a</sup>	-	315	10.4	120.7
melting of ββ <sup>b</sup>	Cα-RMSD = 3.0	305 ± 2	14.4 ± 3.7	336.0 ± 78.0
melting of β-hairpin <sup>b</sup>	Cα-RMSD = 3.0	307 ± 3	13.6 ± 3.0	229.7 ± 66.0
melting of hydrophobic core <sup>b</sup>	R <sub>g</sub> = 7.7	323 ± 6	11.6 ± 3.2	188.0 ± 60.0
melting of α-helix <sup>b</sup>	Cα-RMSD = 3.7	373 ± 10	13.6 ± 4.0	125.1 ± 50.0
DSC experiment <sup>c</sup>	-	314	12 to 15	-
excess heat capacity peak 1 <sup>b</sup>	-	321 ± 6	70.0	-
excess heat capacity peak 2 <sup>b</sup>	-	387 ± 4	122.0	-

<sup>a</sup>Ref. 7.<sup>b</sup>This work (see Fig. 2).<sup>c</sup>Ref. 3.

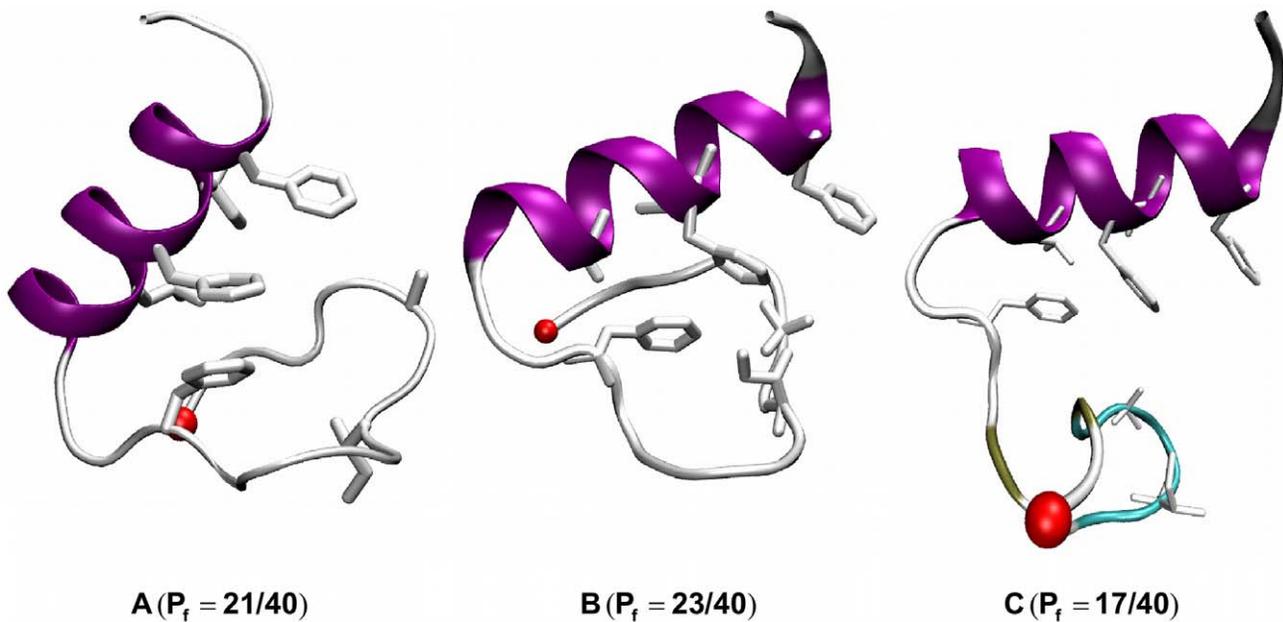
doi:10.1371/journal.pcbi.1000998.t002

broader one centered at Cα-RMSD = 7.0 Å, separated by a barrier at Cα-RMSD = 3.0 Å. The first basin represents the folded state, whereas the second represents the unfolded state. A common feature of the structures residing at the barrier is full formation of the C-terminal helix and partial formation of the hydrophobic core and the N-terminal β-hairpin. If the order parameter chosen (Cα-RMSD) is close to the true reaction coordinate, then the barrier region should contain transition state structures. We identify a representative/centroid structure (Fig. 4A) from our clustering of the conformations present in the barrier region and use it for a  $P_{\text{fold}}$  test. In the  $P_{\text{fold}}$  test, 40 constant temperature molecular dynamics simulations (200 ns of each) are initiated from this structure and the number of folding and unfolding trajectories

are counted. For this structure, about half of trajectories folded and other half unfolded ( $P_{\text{fold}} \sim 0.5$ ), intimating that this structure belongs to the transition state ensemble and that the Cα-RMSD might be a satisfactory reaction coordinate. (While we have identified a transition state structure from this analysis, it is difficult to comment on “how good” a reaction coordinate the Cα-RMSD really is. A “true” reaction coordinate would contain *only* transition state structures at the barrier region, this constituting the necessary and sufficient condition for being a reaction coordinate. An analysis of all the structures residing at the top of the barrier to assess the quality of the reaction coordinate would simply be computationally prohibitive). Next, we use the barrier height to estimate the folding time. Following Kramer’s theory (see the

**Figure 3.** Folding free-energy of FSD-1 against Cα-RMSD of the whole protein at 323K.

doi:10.1371/journal.pcbi.1000998.g003



**Figure 4. Transition structures identified from MD simulations.** A: Identified from the 323 K replica of the REMD simulation by clustering analysis. B–C: Identified from the two successful CMD folding trajectories at 300 K.  $P_{\text{fold}}$  test results for TS states are noted. doi:10.1371/journal.pcbi.1000998.g004

method section for details), the folding time is estimated to be  $\tau_{\text{folding}} = 2.2 \pm 0.7 \mu\text{s}$  from equation 7 by input of the free energy barrier ( $\Delta G^* = 0.9 \pm 0.1 \text{ kcal/mol}$ ) and pre-exponential factor ( $A = 0.56 \mu\text{s}$ ). The order of magnitude is comparable to those of two similar proteins [2,17] (e.g. similar size and the same fold) listed in Table 1. In particular, this folding time is of the same order as the one ( $4.2 \mu\text{s}$ ) measured at 322 K for the fluorophore mutant (FSD-1ss) of FSD-1. (We note that because our reaction coordinate is not the “true” reaction coordinate, our computational value is a lower bound estimate of the true folding time).

#### Conventional Molecular Dynamics (CMD) simulations

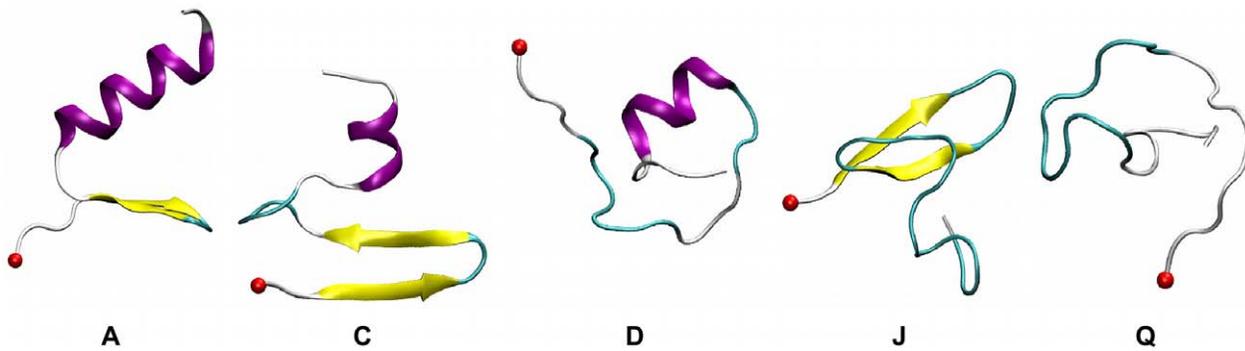
To directly probe folding routes, we performed 20 CMD simulations ( $1 \mu\text{s}$  each) at 300K starting from an extended conformation. The last snapshot of each trajectory (Text S1) offers a quick structural assessment: a successful  $\alpha/\beta$  fold was formed in 2 trajectories (A–B), an unpacked  $\alpha/\beta$  fold was observed in 1 trajectory (C), a partial  $\alpha$ -helix was formed in 6 trajectories (D–I), a partial  $\beta$ -sheet was formed in 7 trajectories (J–P) and a compact coil-turn fold was formed in 4 trajectories (Q–T). A representative structure for each structural family is shown in Fig. 5. Altogether, 10% of the trajectories folded to the correct  $\alpha/\beta$  fold, yielding an estimated folding time of  $10 \pm 7 \mu\text{s}$  (see Methods section). This time needs to be corrected taking into account the fact that our simulations used a friction coefficient  $1/60$  of the one in water. The corrected folding time is at least  $600 \mu\text{s}$ , far larger than our estimate of  $2.2 \mu\text{s}$  from the barrier height from REMD (or the  $4.5 \mu\text{s}$  folding time of the FSD1-ss mutant). The implication is that the CMD simulations lead to a much rugged folding landscape than “reality” and that the ff96/igb5 combination may overstabilize unfolded structures (e.g. C–Q in Fig. 5).

The  $C\alpha$ -RMSD of each trajectory (Text S1) was calculated (defined as  $C\alpha\text{-RMSD} < 3.0 \text{ \AA}$  for at least consecutive 30 ns) for each trajectory. The two successful folding trajectories are shown in Fig. 6 and Fig. 7. By applying the clustering analysis described in the Methods section to each trajectory, we identified structural

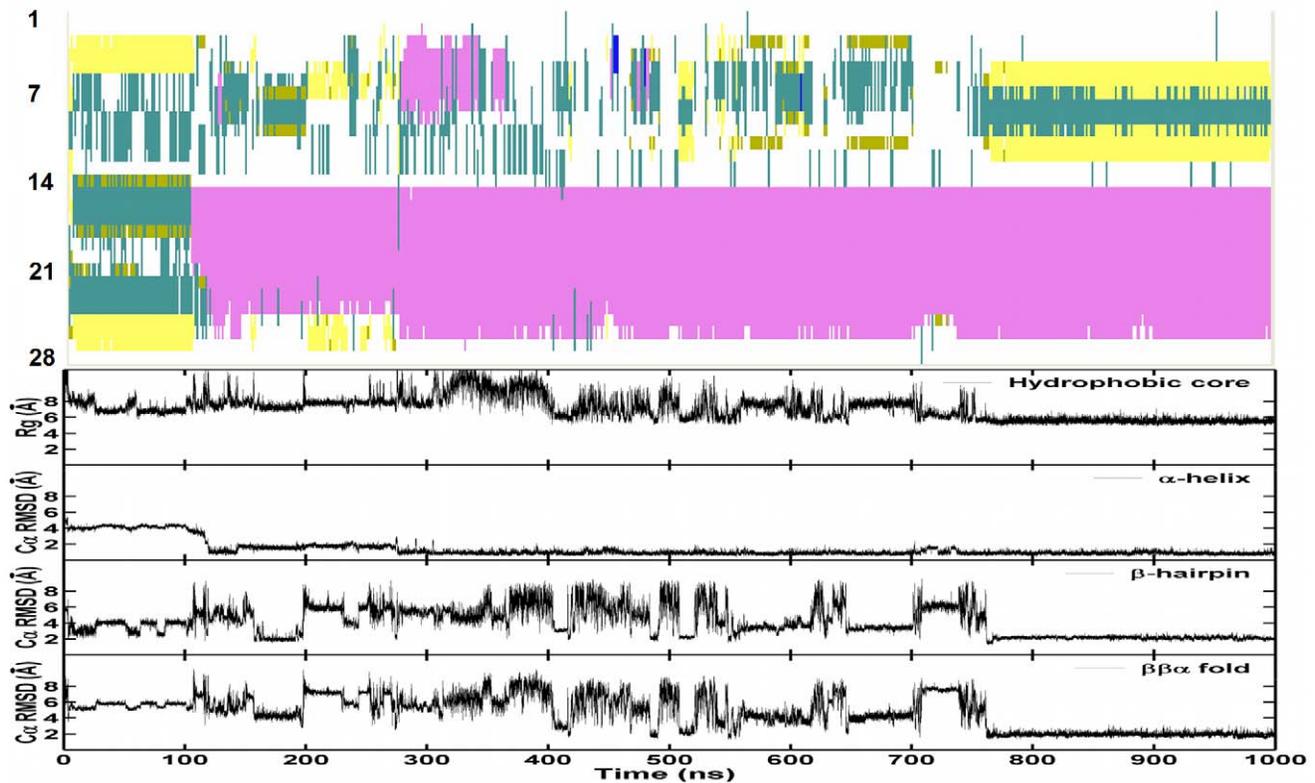
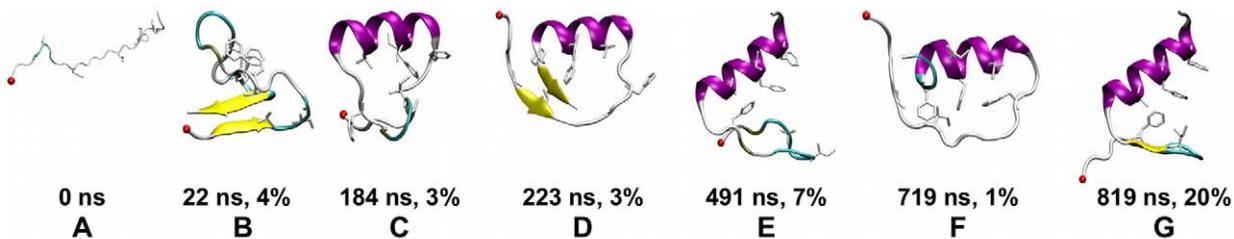
families whose population is greater than 1% of the total snapshots. The representative structures (the centroid of each structural family), with their respective time of occurrence and abundance, are shown in Fig. 6 and Fig. 7. Strikingly, two different folding routes are observed in these two trajectories.

**Trajectory 1.** Following hydrophobic collapse, a non-native main chain fold (A) emerges, consisting of a short two-strand antiparallel  $\beta$ -sheet formed at the two ends, a large disordered loop and a non-native hydrophobic core. The non-native main chain fold and the non-native hydrophobic core then unfold and the correct C-terminal native  $\alpha$ -helix (B) forms, initiating from the middle of the protein chain and rapidly extending to the C-terminus (within 20 ns). The N-terminal fragment sampled various conformations (non-native ones in structures D and F, and native-like in structures C and E) before reaching the native  $\beta$ -hairpin (G). Some features are noted from the time development of the 4 order parameters: 1) the formation of C-terminal  $\alpha$ -helix (within 120 ns) preceded the formation of the N-terminal  $\beta$ -hairpin, hydrophobic core and global  $\beta\beta\alpha$  fold (within 780 ns); 2) the formation of N-terminal  $\beta$ -hairpin was highly correlated to the formation of the tertiary fold with a correlation coefficient of 0.88, indicating a concurrent assembly of the  $\beta$ -hairpin and tertiary fold; 3) the formation of the hydrophobic core was moderately correlated with the folding of the tertiary fold with correlation coefficients of 0.69, suggesting a spread between the formation of the hydrophobic core and the tertiary fold.

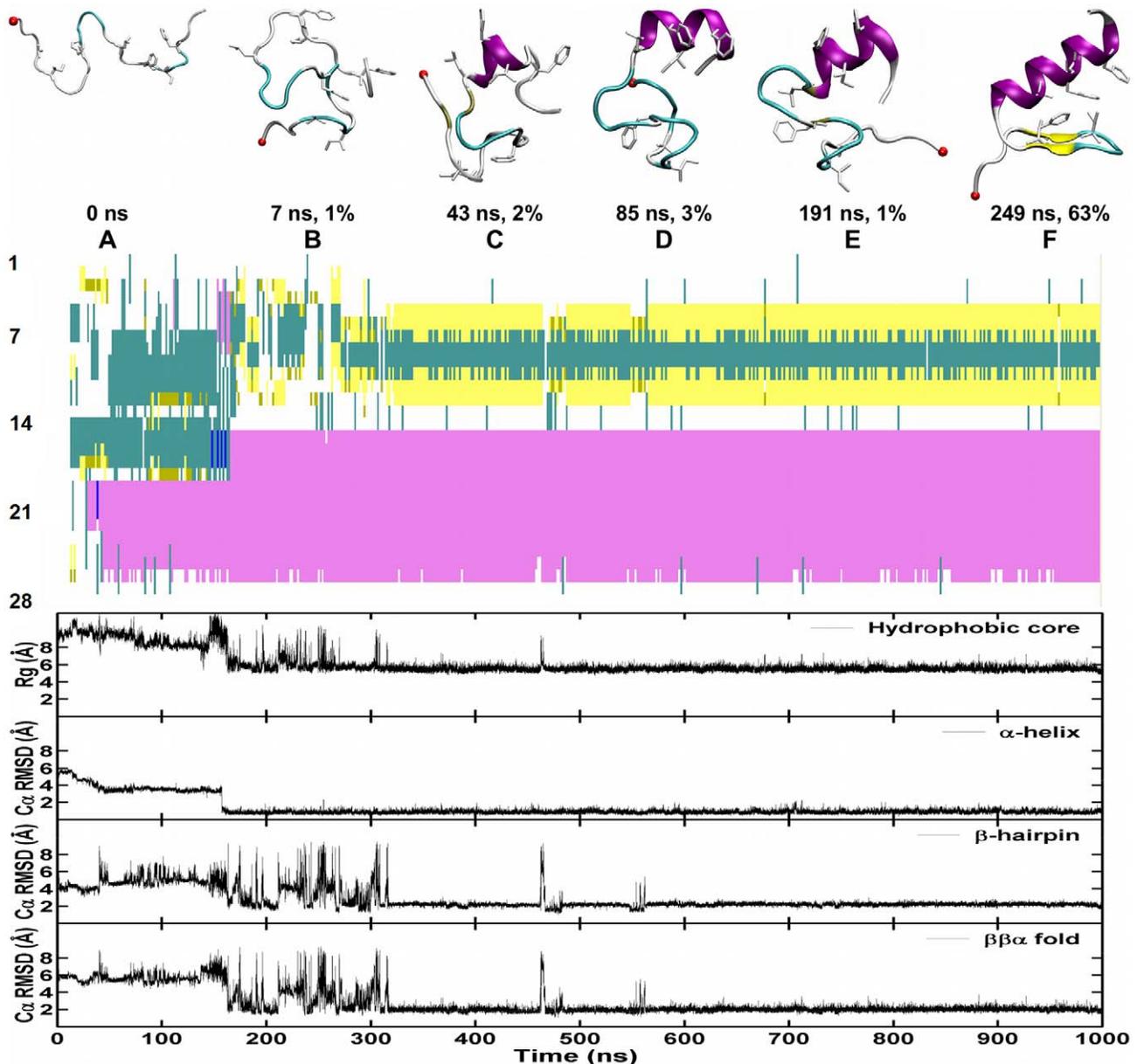
**Trajectory 2.** After collapse, a compact disordered structure without any hydrophobic core (B) was formed. The C-terminal native  $\alpha$ -helix (B) was initiated from the lower half of the protein chain and extended toward both the C-terminus and the N-terminus (within 140 ns). Again, the N-terminal fragment sampled various conformations (the native-like one in C, and non-native ones in D and E) before reaching the native  $\beta$ -hairpin (G). Trajectory 2 shares the main three features observed in Trajectory 1 (and listed in the previous paragraphs), yet the details of the folding pathway differ, and folding occurs faster (320 ns).



**Figure 5. Representative structures from the last snapshots of 20 CMD trajectories.** The backbone is shown in cartoon and the secondary structure is coded by color: coil in silver,  $\alpha$ -helix in purple,  $\beta$ -sheet in yellow, isolated  $\beta$ -bridge in tan and turn in cyan.  
doi:10.1371/journal.pcbi.1000998.g005



**Figure 6. Successful folding trajectory 1 from CMD simulations of FSD-1 at 300 K.** Top: The representative snapshots of top structural families are presented together with their abundance and time. The backbone is shown in cartoon. The secondary structure is coded by color: coil in silver,  $\alpha$ -helix in purple,  $\beta$ -sheet in yellow, isolated  $\beta$ -bridge in tan and turn in cyan. N-terminal is shown by a red VDW ball. Middle: The development of secondary structure. Bottom: The development of four order parameters: the radius of gyration of the hydrophobic core, the  $C\alpha$ -RMSD of the N-terminal hairpin (residues 3–13), the C-terminal  $\alpha$ -helix (14–26) and the whole protein (residues 3–26) against the NMR structure.  
doi:10.1371/journal.pcbi.1000998.g006



**Figure 7. Successful folding trajectory 2 from CMD simulations of FSD-1 at 300 K.** Top: The representative snapshot of top structural families is presented together with its abundance and time. The backbone is shown in cartoon. The secondary structure is coded by color: coil in silver,  $\alpha$ -helix in purple,  $\beta$ -sheet in yellow, isolated  $\beta$ -bridge in tan and turn in cyan. N-terminal is shown by a red VDW ball. Middle: The development of secondary structure. Bottom: The development of four order parameters: the radius of gyration of the hydrophobic core, the  $C_{\alpha}$ -RMSD of the N-terminal hairpin (residues 3–13), the C-terminal  $\alpha$ -helix (14–26) and the whole protein (residues 3–26) against the NMR structure.  
doi:10.1371/journal.pcbi.1000998.g007

In summary, both trajectories share the following common features: the early formation of the  $\alpha$ -helix within 150 ns, and the concurrent formation of  $\beta$ -hairpin and tertiary structure within 320–780 ns. These features are consistent with the thermodynamic data obtained from the replica exchange simulations. Despite the overall similarity in folding mechanism, the individual trajectories show clear differences in the details of how the individual components form: 1) the C-terminal helix can initiate from different sites (residues 15–20 in Trajectory 1 vs. residues 19–22 in Trajectory 2); 2.) A significant heterogeneity exists across the two successful folding trajectories (for instance, trajectory 1 shows early transient population of a structure with two  $\beta$ -strands). Our direct observation of multiple folding routes

is in direct line with the folding funnel energy landscape perspective [18].

**Transition state structures from the two folding trajectories.** Putative TSS were identified from both trajectories as the structures that will fold into the native structure in the next time step [19]. 40 independent simulations starting from the putative structures (200 ns of each) were performed to calculate a  $P_{\text{fold}}$  value and determine whether or not these conformations belong to the transition state ensemble. For the two structures shown in Fig. 4 (B–C), (each coming from one of the two folding trajectories), we observed about a half of 40 trajectories (23 and 17) reaching the native state. Clearly, these two “kinetic” transition state structures share similar structural

features to the “thermodynamic” transition state structures obtained from the REMD data (Fig. 4A): full formation of the C-terminal helix and partial formation of the hydrophobic core and the N-terminal  $\beta$ -hairpin. The major difference between these structures, which reflects the heterogeneity of the transition state ensemble, is located at the  $\beta$ -hairpin part of FSD-1.

## Discussion

FSD-1, a designed small  $\beta\beta\alpha$  ultrafast folder, has received considerable experimental and computational attention as a model system for studying protein folding mechanisms and for testing computational models. However, the validity of using FSD-1 as a prototypical folder has recently been challenged by Feng et al [3]. They argue that the broad melting transition at 315 K observed by CD and DSC is mainly due to the melting of the helical portion of the protein, rather than to the melting of the entire protein. They point out that an overall protein transition should exhibit a better-defined baseline and a higher melting temperature (for comparison, HP35 folds at 342 K) and the  $\beta$ -hairpin part appears to be flexible and lack of stability based on their REMD simulations using OPLS-AA/L 2001 force field, TIP4P water model and NPT ensemble. Their simulations and others [4], thus far, have not been able to address this controversy and provide an explanation for the broad melting transition of FSD-1, with computational folding transitions all lying at much higher temperatures than the experimental ones. Reliable folding of mixed  $\alpha/\beta$  fold proteins like FSD-1 is notoriously difficult, mostly because most force fields are either heliophilic [20,21] or  $\beta$ -centric [12].

In the present paper, we investigate the folding of FSD-1 using the Amber ff96 force field combined with the implicit solvent IGB=5, which appears to offer a reasonable balance of helical and beta propensities. Our REMD simulations show two broad peaks in the excess heat capacity at  $T_m = 321 \pm 6$  K and  $T_m = 387 \pm 4$  K. These peaks correspond to the two structural transitions identified from the computational thermal-denaturation curves: 1) formation of the hydrophobic core ( $T_m = 323 \pm 6$  K),  $\beta$ -hairpin ( $T_m = 307 \pm 3$  K) and tertiary fold ( $T_m = 305 \pm 2$  K) and 2) formation of the  $\alpha$ -helix ( $T_m = 373 \pm 10$  K). By comparing the simulations results with the DSC and CD experiments, we find that the first transition qualitatively agrees with the experiments, whereas the second transition is an artifact of the simulations resulting from the ff96/igb5 induced overstabilization of the  $\alpha$ -helix. This suggests that the improved igb5 solvation model does not fully resolve the secondary balance problem inherent to the heliophilic Amber ff96 force field and that further refinement of the protein force field is necessary.

While the force field clearly overstabilizes the helix, the fact that the helix is more stable than the other secondary structural elements is confirmed by the CD spectra. A re-examination of the experimental CD spectra of FSD-1 at 353 K and 277 K presented in Fig. 2 of ref 3 shows that the spectrum above the experimental  $T_m = 315$  K possesses helical features rather than coil-only features and the CD spectrum below the experimental  $T_m = 315$  K possesses both  $\beta$ -hairpin-rich and helix-rich features rather than helix-only features. Indeed, the spectrum at 353 K lacks random-coil features (i.e. very low ellipticity above 210 nm and negative bands near 195 nm); instead it shows helical negative bands at 203 nm and 222 nm, indicating the presence of helical structure. In addition, the CD spectrum of FSD-1 at 277 K not only shows two helix bands at 207 and 220 nm but also contains a sheet band at 218 nm, intimating that the protein shows signs of a folded  $\alpha/\beta$  protein at this temperature. Furthermore, the CD

spectrum of FSD-1 is significantly different from the CD spectra of the helix-only protein [11], suggesting it contains both  $\alpha$  and  $\beta$  secondary structure.

Put together, our prediction is as follows: although the  $\alpha$ -helix is the most stable structural component (i.e. the helix-to-coil transition is dominant in the denaturation process), the broad denaturation transition seen experimentally also has a significant contribution arising from the spread in melting temperatures of the hydrophobic core, the  $\beta$ -hairpin. In other words, the breadth does not arise solely from the helix-coil transition as had been proposed by Feng and coworkers [3]. Our simulations explain the experimentally observed lack of well-defined baselines as due to the marginal stability of this protein, coupled with the spread in melting temperatures of the individual secondary and tertiary components. We suggest that mutations that would enhance the stability of the  $\beta$ -hairpin could improve the overall stability and the cooperativity of this designed protein, making it an even better model for natural  $\alpha/\beta$  proteins.

The folding free energy landscape at temperatures near the folding temperature ( $T_m = 321 \pm 6$  K) shows two well-defined folded and unfolded basins. Transition state structures identified from the structures at the top of the barrier satisfying the  $P_{fold}$  analysis revealed common features: full formation of the C-terminal helix and partial formation of the hydrophobic core and the N-terminal  $\beta$ -hairpin. The free energy barrier height enables us to calculate the folding time of FSD-1. Following Kramer's theory, the folding time is estimated to be  $\tau_{folding} = 2.2 \pm 0.7$   $\mu$ s, which is of the same order of magnitude as that of FSD-1ss [2], which shows two folding phases ( $\tau_1 \sim 150$  ns and  $\tau_2 \sim 4.5$   $\mu$ s). Should FSD-1 present the same two phases, we could assign the first fast phase to the formation of hydrophobic core, and assign the concurrent formation of N-terminal  $\beta$ -hairpin and the tertiary fold to the slower phase ( $\sim 4.5$   $\mu$ s).

Furthermore, our CMD trajectories provide detailed atomic information of possible folding routes starting from a straight chain. Although the general folding features gleaned from the REMD thermodynamic analysis (e.g. early formation of the C-terminal helix and thereafter folding N-terminal hairpin and the tertiary fold) are seen in the two successful folding trajectories, a significant heterogeneity is present across the two folding trajectories: 1) the C-terminal helix can initiate from different sites (residues 15–20 in Trajectory 1 vs. residues 19–22 in Trajectory 2); 2) different structures were visited along the folding routes; 3) the transition state structures found in each trajectory, validated by  $P_{fold}$  analysis, differ, supporting the notion of different transition structures for multiple routes [22]; 4) significant non-native topologies and secondary structures are also sampled along the two folding routes. The presence of these structures underlines the importance of considering non-native interactions, in addition to the natively favored interactions used in Go-like models [23]. Non-native conformations play a role in modulating folding times and mechanisms. Put together, our all-atom CMD data provide direct evidences to a funneled energy landscape [18] with multiple folding pathways and a diverse transition state ensemble. Nonetheless, the low number of successful folding runs (only 2 out of 20 trajectories) and the resulting lengthy folding time ( $\sim 600$   $\mu$ s vs. 2.2  $\mu$ s) indicates that the CMD simulations generate a much more rugged folding landscape than reality. A better modeling by the implicit protein force field for the unfolded state might resolve the problem.

It is interesting to compare the folding mechanism of FSD-1 to a similar protein (BBA5) with  $\beta\beta\alpha$  fold. In the case of BBA5, Pande and coworkers showed that this protein follows a diffusion-collision model [17,24]: docking of the preformed  $\alpha$ -helix and  $\beta$ -hairpin.

Thus, for BBA5, the formation of the  $\beta$ -hairpin precedes the formation of the tertiary fold, whereas in the case of FSD-1, we have shown that the  $\beta$ -hairpin and the tertiary fold form concurrently. Clearly, the D-proline present at the turn region of the  $\beta$ -hairpin of BBA facilitates the formation of this structural element, but, as a result, it may reduce the overall folding cooperativity of the protein. It is possible that the introduction of non-natural amino-acids makes this protein less “funnel-like”. Nonetheless, BBA5 and FSD-1 fold in microsecond despite these differences in folding mechanism, consistent with contact order theory that suggests that native topology is a major factor in determining the folding time [25].

## Methods

The AMBER 8 simulation package [26] is used in both molecular dynamics simulations and data processing. The protein is modeled using the AMBER all-atom point-charge force field, ff96 [27]. Solvation effects are represented by the recent implicit solvent model (IGB=5) [8] plus the surface term (gbsa=1, 0.005 kcal/Å<sup>2</sup>/mol) with an effective salt concentration of 0.2 M.

### REMD and CMD folding simulations started from an extended chain conformation

An initial energy minimization was performed on an extended chain conformation and the minimized structure was used as the input for both replica exchange molecular dynamics (REMD) and conventional molecular dynamics simulations (CMD) simulations. In REMD simulation [28,29,30,31], 16 replicas were set up with initial temperatures exponentially spaced from 271 to 465 K for solution phase calculations (i.e 271.0 280.0 289.3 300.0 311.2 322.7 334.7 347.2 360.1 373.4 387.3 401.7 416.6 432.1 448.2 465.0, see reference [32] for the algorithm used to optimize them). 20 CMD simulations were conducted at 300 K. Initial velocities for each trajectory were generated according to the Maxwell-Boltzmann distribution for its target temperature. The first 1.0 ns of REMD simulation was performed to equilibrate the system at its target temperatures. After equilibrium, exchanges between neighboring replicas were attempted every 1000 MD steps (2.0 ps) and the exchange rate among replicas was ~20%. SHAKE [33] was applied to constrain all bonds connecting hydrogen atoms and thus a time step of 2.0 fs. In order to reduce computation time, non-bonded forces were calculated using a two-stage RESPA (reference system propagator algorithm approach) [34] where the forces within a 12 Å radius were updated every step and those beyond 12 Å were updated every two steps. The Langevin dynamics was used to control the temperature 300K using a collision frequency of 1.0 ps<sup>-1</sup>. The lower collision frequency than a typical value (~60 ps<sup>-1</sup>) for water solvent was used for a better conformational sampling. The center of mass translation and rotation were removed every 250 MD steps (0.5 ps). Each trajectory was run for 1.25  $\mu$ s and 1.0  $\mu$ s respectively in REMD and CMD simulations, giving a cumulative simulation time of 20.0  $\mu$ s. The trajectories were saved at 2.0 ps intervals for further analysis.

### Secondary structure and tertiary structure analysis

For analysis of secondary structure, the STRIDE program of Frishman and Argos [35] is used. For analysis of tertiary structural families, the snapshots are clustered by the GROMACS protocol [36], in which the structure similarity score is based on pair wise C $\alpha$ -RMSD of 2.0 Å. This is done in order to reduce a large number of the sampled structures into a few structural families. The structure with the largest number of neighboring structures

within the cutoff, was selected as the representative structure of the structural family/cluster.

### Four order parameters

The formation of hydrophobic core, secondary structures and tertiary structure are important events in the protein folding process. To monitor the formation of the hydrophobic core of FSD-1, we calculate the radius of gyration for the hydrophobic core formed by the residues Ala5, Ile7, Phe12, Leu18, Phe21, Ile22, and Phe25. To characterize the secondary and tertiary structure formation, the C $\alpha$ -RMSDs of the N-terminal hairpin (residues 3–13), the C-terminal  $\alpha$ -helix (residues 14–26) and the whole protein (residues 3–26), are calculated against the NMR structure (pdb id: 1FSD). The terminal residues are flexible and thus are not included the calculation.

### Thermodynamic properties

The fractions of the folded state  $P_f$  at various temperatures are directly calculated from the REMD trajectories based on each of the four order parameters. Here, the folded state is defined by setting a cutoff of C $\alpha$ -RMSD or the radius of gyration (Rg) for each order parameter. The cutoff is the value that separates the two populations in the distribution of each order parameter at 323 K (Text S1). The values are listed in Table 2.

The melting transition temperature  $T_m$  is obtained at the midpoint where  $P_f = 0.5$ . Assuming a two-state thermodynamic model, the thermodynamic parameters in the direction of folding to unfolding can be obtained from the following van't Hoff analysis [37]:

$$\Delta H(T) = \Delta H_m + \Delta C_p(T - T_m) \quad (1)$$

$$\Delta S(T) = \Delta S_m + \Delta C_p * \ln(T/T_m) \quad (2)$$

$$\Delta G(T) = -RT \ln\left(\frac{P_f}{1-P_f}\right) = \Delta H_m\left(1 - \frac{T}{T_m}\right) + \Delta C_p\left(T - T_m - T \ln\left(\frac{T}{T_m}\right)\right), \quad (3)$$

where  $\Delta H_m$ ,  $\Delta S_m$  and  $\Delta C_p$  (assumed to be constant across temperature) are the changes in the van't Hoff enthalpy, entropy at  $T_m$  and heat capacity at constant pressure.

Also,  $T_m$  (i.e. peak temperature values) and the denaturation enthalpy,

$$\Delta H_{cal} = \int_{T_0}^{T_f} \Delta C_p dT, \quad (4)$$

can be obtained from the excess heat capacity profile, which is calculated by subtracting the heat capacity of the native state (i.e. the lowest temperature) from the absolute heat capacity profile [38]:

$$\Delta C_p = C_p - C_{p,o}. \quad (5)$$

The absolute heat capacity ( $\Delta C_p$ ) is estimated from the potential energy distribution at each temperature from the REMD simulation by

$$C_p \approx C_v = (\langle E^2 \rangle - \langle E \rangle^2) / RT^2 \quad (6)$$

where  $E$  is the potential energy,  $R$  is the gas constant, and  $T$  is the temperature.

**Estimation of the folding time:** The folding time  $\tau_{folding}$  can be obtained from Kramers' theory of unimolecular reaction rates in solution [39,40,41]:

$$\tau_{folding} = A * \exp\left(\frac{\Delta G^*}{RT}\right), \quad (7)$$

where  $\Delta G^*$  is the height of the free energy barrier,  $R$  is the gas constant and  $T$  is the absolute temperature,  $A$  is the pre-exponential factor. As to the 40-residue protein BBL,  $A$  is about  $0.8 \mu s$  [42] estimated by the measured relaxation time [43] of the Förster resonance energy transfer (FRET) efficiency for the acid-denatured state of the BBL with donor and acceptor fluorophores attached to the N and C termini at 305 K. Using the linear length scaling suggested by the homopolymer collapse theory [44], the pre-exponent factor  $A$  for a 28 residue FSD-1 is about  $0.56 \mu s$  ( $0.80 \mu s * 28/40$ ).

Alternatively, the folding time  $\tau_{folding}$  based on two-state folding model can be estimated from a large number of CMD simulations ( $N_{total}$ ) of a short duration ( $t \ll \tau_{folding}$ ) [45]:

$$\tau_{folding} = \frac{N_{total}}{N_{folded}} * t \pm \frac{N_{total}}{(N_{folded})^{3/2}} * t, \quad (8)$$

where  $N_{folded}$  is the number of the folded trajectories.

## Supporting Information

**Text S1** Analysis of folding simulation data. Block analysis of the REMD, distribution of the four order parameters at 323K, heat capacity profile using Berendsen thermostat, RMSDs of 40 CMD trajectories from TS1, Final snapshots and RMSDs of 20 CMD trajectories from an extended conformation and RMSDs of 80 CMD trajectories from TS2 and TS3 are included.

Found at: doi:10.1371/journal.pcbi.1000998.s001 (4.42 MB DOC)

## Author Contributions

Conceived and designed the experiments: CW JES. Performed the experiments: CW. Analyzed the data: CW. Wrote the paper: CW JES.

## References

- Dahiyat BI, Mayo SL (1997) De novo protein design: Fully automated sequence selection. *Science* 278: 82–87.
- Sadqi M, de Alba E, Perez-Jimenez R, Sanchez-Ruiz JM, Munoz V (2009) A designed protein as experimental model of primordial folding. *Proc Natl Acad Sci U S A* 106: 4127–4132.
- Feng JWA, Kao J, Marshall GR (2009) A Second Look at Mini-Protein Stability: Analysis of FSD-1 Using Circular Dichroism, Differential Scanning Calorimetry, and Simulations. *Biophys J* 97: 2803–2810.
- Li WF, Zhang J, Wang W (2007) Understanding the folding and stability of a zinc finger-based full sequence design protein with replica exchange molecular dynamics simulations. *Proteins* 67: 338–349.
- Derreumaux P (1999) From polypeptide sequences to structures using Monte Carlo simulations and an optimized potential. *J Chem Phys* 111: 2301–2310.
- Jang S, Shin S, Pak Y (2002) Molecular dynamics study of peptides in implicit water: Ab initio folding of  $\beta$ -hairpin,  $\beta$ -sheet, and  $\beta\beta\alpha$ -motif. *J Am Chem Soc* 124: 4976–4977.
- Kim E, Jang S, Pak Y (2007) Consistent free energy landscapes and thermodynamic properties of small proteins based on a single all-atom force field employing an implicit solvation. *J Chem Phys* 127: 145104.
- Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* 55: 383–394.
- Lei HX, Wang ZX, Wu C, Duan Y (2009) Dual folding pathways of an  $\alpha/\beta$  protein from all-atom ab initio folding simulations. *J Chem Phys* 131: 165105.
- Sarisky CA, Mayo SL (2001) The  $\beta\beta\alpha$  fold: Explorations in sequence space. *J Mol Biol* 307: 1411–1418.
- Scholtz JM, Marqusee S, Baldwin RL, York EJ, Stewart JM, et al. (1991) Calorimetric Determination of the Enthalpy Change for the  $\alpha$ -Helix to Coil Transition of an Alanine Peptide in Water. *Proc Natl Acad Sci U S A* 88: 2854–2858.
- Shell MS, Ritterson R, Dill KA (2008) A test on peptide stability of AMBER force fields with implicit solvation. *J Phys Chem B* 112: 6878–6886.
- Grabenaus M, Wu C, Soto P, Shea JE, Bowers MT (2010) Oligomers of the prion protein fragment 106–126 are likely assembled from ordered  $\beta$ -hairpins: A combined experimental and computational study. *J Am Chem Soc* 132: 532–539.
- Dupuis NF, Wu C, Shea JE, Bowers MT (2010) Human Islet Amyloid Polypeptide forms ordered  $\beta$ -hairpin monomers: a possible amyloidogenic conformation. *J Am Chem Soc* 131: 18283–18292.
- Voelz VA, Bowman GR, Beauchamp K, Pande VS (2010) Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9(1–39). *J Am Chem Soc* 132: 1526–1528.
- Voelz VA, Singh VR, Wedemeyer WJ, Lapidus LJ, Pande VS (2010) Unfolded-State Dynamics and Structure of Protein L Characterized by Simulation and Experiment. *J Am Chem Soc* 132: 4702–4709.
- Snow CD, Nguyen N, Pande VS, Gruebele M (2002) Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* 420: 102–106.
- Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, Pathways, and the Energy Landscape of Protein-Folding - a Synthesis. *Proteins* 21: 167–195.
- Li AJ, Daggett V (1996) Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations. *J Mol Biol* 257: 412–429.
- Best RB, Buchete NV, Hummer G (2008) Are current molecular dynamics force fields too helical? *Biophys J* 95: L7–L9.
- Freddolino PL, Liu F, Gruebele M, Schulten K (2008) Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys J* 94: L75–L77.
- Lei HX, Dastidar SG, Duan Y (2006) Folding transition-state and denatured-state ensembles of FSD-1 from folding and unfolding simulations. *J Phys Chem B* 110: 22001–22008.
- Go N, Taketomi H (1978) Respective Roles of Short-Range and Long-Range Interactions in Protein Folding. *Proc Natl Acad Sci U S A* 75: 559–563.
- Jang SM, Kim E, Pak YS (2008) All-atom level direct folding simulation of a  $\beta\beta\alpha$  miniprotein. *J Chem Phys* 128: 105102.
- Plaxco KW, Simons KT, Baker D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 277: 985–994.
- Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, et al. (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26: 1668–1688.
- Kollman P, Dixon R, Cornell W, Fox T, Chipot C, et al. (1997) The development/application of “minimalist” organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. In: Gunsteren W, Weiner P, eds. *Computer Simulation of Biomolecular Systems*. Dordrecht: Escom. pp 83–96.
- Swendsen RH, Wang JS (1986) Replica Monte-Carlo Simulation of Spin-Glasses. *Phys Rev Lett* 57: 2607–2609.
- Geyer CJ (1991) Markov chain Monte Carlo maximum likelihood. In: Keramidas EM, ed. *Computing science and statistics: Proceedings of the 23rd Symposium on the Interface*. Fairfax Station VA: Interface Foundation. pp 156–163.
- Hukushima K, Nemoto K (1996) Exchange Monte Carlo method and application to spin glass simulations. *J Physical Soc Japan* 65: 1604–1608.
- Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314: 141–151.
- Wu C, Murray MM, Bernstein SL, Condoron MM, Bitan G, et al. (2009) Experimental and theoretical study of the structure of Amyloid- $\beta$  C-terminal fragments. *J Mol Biol* 387: 592–501.
- Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J Comput Phys* 23: 327–341.
- Procacci P, Berne BJ (1994) Multiple Time-Scale Methods for Constant-Pressure Molecular-Dynamics Simulations of Molecular-Systems. *Mol Phys* 83: 255–272.
- Frishman D, Argos P (1995) Knowledge-based protein secondary structure assignment. *Proteins* 23: 566–579.
- Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, et al. (1999) Peptide folding: When simulation meets experiment. *Angew Chem Int Ed* 38: 236–240.
- Becktel WJ, Schellman JA (1987) Protein Stability Curves. *Biopolymers* 26: 1859–1877.

38. Freire E, Biltonen RL (1978) Statistical Mechanical Deconvolution of Thermal Transitions in Macromolecules .1. Theory and Application to Homogeneous Systems. *Biopolymers* 17: 463–479.
39. Kramers HA (1940) Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* 7: 284–304.
40. Berne BJ, Borkovec M, Straub JE (1988) Classical and Modern Methods in Reaction-Rate Theory. *J Phys Chem* 92: 3711–3725.
41. Socci ND, Onuchic JN, Wolynes PG (1996) Diffusive dynamics of the reaction coordinate for protein folding funnels. *J Chem Phys* 104: 5860–5868.
42. Kubelka J, Hofrichter J, Eaton WA (2004) The protein folding ‘speed limit’. *Curr Opin Struct Biol* 14: 76–88.
43. Sadqi M, Lapidus LJ, Munoz V (2003) How fast is protein hydrophobic collapse? *Proc Natl Acad Sci U S A* 100: 12117–12122.
44. Pitard E, Orland H (1998) Dynamics of the swelling or collapse of a homopolymer. *Europhys Lett* 41: 467–472.
45. Pande VS, Baker I, Chapman J, Elmer SP, Khaliq S, et al. (2003) Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers* 68: 91–109.