

Article

Probing the Nanosecond Dynamics of a Designed Three-Stranded Beta-Sheet with a Massively Parallel Molecular Dynamics Simulation

Vincent A. Voelz ¹, Edgar Luttmann ¹, Gregory R. Bowman ² and Vijay S. Pande ^{2,*}

¹ Department of Chemistry / Stanford University, Stanford, California 94305, USA;
E-Mails: vvoelz@stanford.edu (V.V.); luttmann@stanford.edu (E.L.)

² Biophysics Program / Stanford University, Stanford, California 94305, USA;
E-Mail: gbowman@stanford.edu

* Author to whom correspondence should be addressed; E-Mail: pande@stanford.edu;
Tel. +1-650-723-3660; Fax: +1-650-725-0259

Received: 15 January 2009; in revised form: 4 March 2009 / Accepted: 9 March 2009 /

Published: 10 March 2009

Abstract: Recently a temperature-jump FTIR study of a designed three-stranded sheet showing a fast relaxation time of $\sim 140 \pm 20$ ns was published. We performed massively parallel molecular dynamics simulations in explicit solvent to probe the structural events involved in this relaxation. While our simulations produce similar relaxation rates, the structural ensemble is broad. We observe the formation of turn structure, but only very weak interaction in the strand regions, which is consistent with the lack of strong backbone-backbone NOEs in previous structural NMR studies. These results suggest that either ^DP^DP-II folds at time scales longer than 240 ns, or that ^DP^DP-II is not a well-defined three-stranded β -sheet. This work also provides an opportunity to compare the performance of several popular forcefield models against one another.

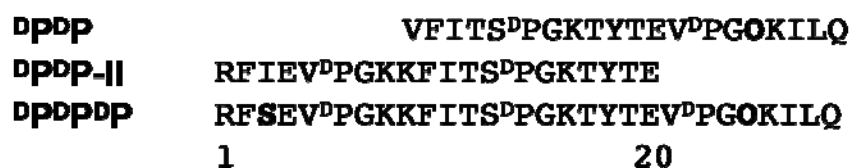
Keywords: Ultrafast folding; downhill folding; DPDP; DPDP-II; designed beta-sheet proteins.

1. Introduction

Xu *et al.* have recently studied the nanosecond time scale folding dynamics of a designed three-stranded sheet mini-protein [1]. This peptide, called ^DP^DP-II, is one of many peptide sequences originally designed by the Gellman group for the purposes of elucidating the sources of thermodynamic stability and folding cooperativity of beta-hairpin and beta-sheet structures [2]. The most stable of these beta-sheet designs have a scaffold that incorporates successive D-proline and glycine residues (^DPG) in the turn regions, a motif shown to form a stable Type-II' turn [3].

Of great interest as model systems have been several three- and four-stranded beta-sheet designs from the Gellman group. The first of these, ^DP^DP, was studied by NMR and shown to have cross-strand NOEs and chemical shifts indicative of beta-sheet populations [4]. Syud *et al.* built upon ^DP^DP, producing (among others): ^DP^DP-II, a three-stranded sheet whose C-terminal hairpin is identical to the N-terminal hairpin of ^DP^DP (Figure 1); and ^DP^DP^DP, a four-stranded composite of ^DP^DP and ^DP^DP-II [5]. The stability of these designs was assessed in a similar fashion by NMR.

Figure 1. Designed beta-sheet peptides designed by the Gellman group: ^DP^DP [4], ^DP^DP-II [5], and ^DP^DP^DP [5]. ^DP denotes D-Proline, and O denotes Ornithine.



Recently, designed ^DPG-turn beta-sheet peptides have become interesting candidates for ultrafast folding beta-sheet systems. Many proteins have been engineered to fold quickly [6-8], close to the “speed limit” of folding [9,10]. Upper limits on protein folding rates are thought ultimately to be controlled by the conformational search rate for forming intermolecular contacts [11,12]. For beta-sheet proteins, the entropic barriers of turn formation are rate limiting [13]. Indeed, a designed variant of human Pin1 WW domain, with a ^DPG substitution in the turn region, shows a 10-fold increase in folding rate compared to native sequence, up to $(\sim 10 \mu\text{s})^{-1}$, becoming one of the fastest folding beta-sheet proteins to date [13].

It had been hypothesized that because of the reduced conformational entropy of the ^DPG turn regions, the folding landscape of ^DP^DP-II might not have activation barriers to folding, and shown to be that of a “downhill” folder [8], with kinetics shaped mainly by landscape roughness. Using temperature-jump FTIR, Xu *et al.* showed that ^DP^DP-II has “the fastest T-jump relaxation rate observed for a beta-sheet system so far” of $(\sim 140 \pm 20 \text{ ns})^{-1}$, with single-exponential relaxation kinetics [1]. More recently, T-jump FTIR studies of the related four-stranded peptide, ^DP^DP^DP, show similar single-exponential kinetics, but with a folding time of $\sim 440 \text{ ns}$ [14].

While the single-exponential kinetics of ^DP^DP-II can be fit to a two-state Arrhenius-type model, Xu *et al.* showed that one-dimensional Langevin models of dynamics over a rough free-energy surface [15,16] explain the data equally well, which is their preferred interpretation. In the case of the four-

stranded $^{\text{D}}\text{P}^{\text{D}}\text{P}^{\text{D}}\text{P}$, Xu *et al.* suspect that many parallel but degenerate refolding pathways may be present [14].

One reason to prefer the “downhill” interpretation is the lack of features typical of activated folding kinetics [15,16]. The fast rate of the relaxation (on the time scale of helix-coil transitions) imply that unfolded and folded ensembles must have a similar degree of compactness, and end-to-end distances as measured by FRET show little sensitivity to temperature. Xu *et al.* suggest a reason for this is the reduced accessible conformational space imposed by the rigid $^{\text{D}}\text{PG}$ turns. Smith and Tokmakoff used time-resolved infrared spectroscopy along with site-specific isotopic labeling techniques to show that the $^{\text{D}}\text{PG}$ turn region of de novo hairpin peptide PG12 does not undergo significant rearrangement upon a temperature-jump, where as the mid-strand regions rearrange on a ~ 130 -ns time scale [17]. Their results support a model where the unfolded state is an expanded but native-like ensemble.

Simulation studies have shed some light onto the thermodynamics of $^{\text{D}}\text{PG}$ -turn beta-sheet proteins. While there have been no previous simulations of $^{\text{D}}\text{P}^{\text{D}}\text{P}$ -II, several groups have simulated the related $^{\text{D}}\text{P}^{\text{D}}\text{P}$ peptide, which shares an 11-residue stretch of hairpin residues (Figure 1). Wang and Sung simulated a 100 ns molecular dynamics trajectory of $^{\text{D}}\text{P}^{\text{D}}\text{P}$ using an implicit solvent model, starting from an extended conformation [18]. Their results show $^{\text{D}}\text{P}^{\text{D}}\text{P}$ folding to beta-sheet structures, and agree with experimental findings that the $^{\text{D}}\text{PG}$ turn is more stable than designed three-stranded peptides with NG or GS turns. Roe *et al.* used replica exchange molecular dynamics in a modified AMBER99 forcefield with an implicit solvation model to sample the thermodynamics of $^{\text{D}}\text{P}^{\text{D}}\text{P}$ [19]. REMD (12 replica trajectories each of ~ 130 ns) dramatically enhanced the convergence of the free energy landscape compared to single-replica MD. The two hairpins of $^{\text{D}}\text{P}^{\text{D}}\text{P}$ show simulated populations of $\sim 50\%$ and $\sim 75\%$, respectively, consistent with NMR and CD studies [4,20], and estimates of thermodynamic cooperativity of -1 to -3 kcal/mol. The less stable of the two $^{\text{D}}\text{P}^{\text{D}}\text{P}$ hairpins comprises the C-terminal sequence of $^{\text{D}}\text{P}^{\text{D}}\text{P}$ -II.

We have been interested in $^{\text{D}}\text{P}^{\text{D}}\text{P}$ -II as a target for molecular simulation for several reasons. It appears to be the fastest-folding beta-sheet protein so far, with relaxation kinetics within the time scale range that can be effectively addressed with all-atom molecular simulation. Moreover, the experimental kinetics remain ambiguous as to whether activation barriers exist for this peptide. To investigate the underlying conformational dynamics, we perform massively parallel molecular dynamics simulations of $^{\text{D}}\text{P}^{\text{D}}\text{P}$ -II in explicit solvent.

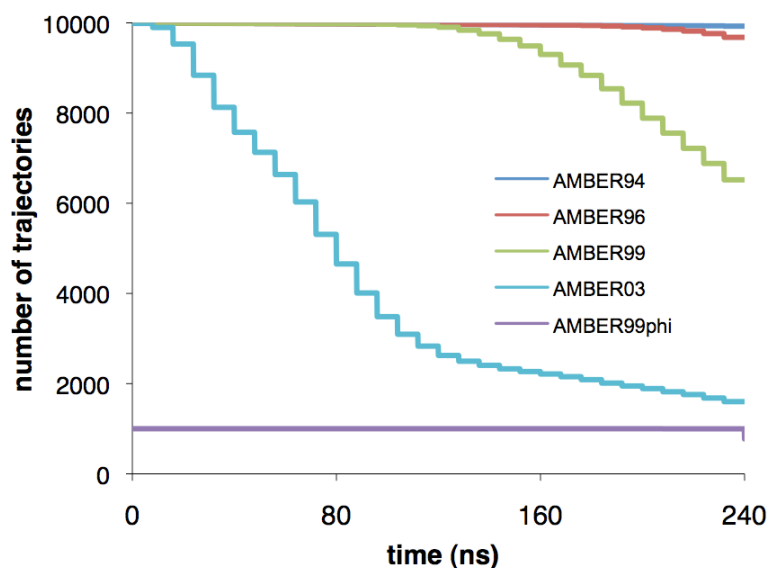
As we report below, the reaction coordinates of average radius of gyration and solvent-accessible surface area of backbone C=O over time show good agreement with the experimentally measured relaxation rates, but we observe very few three-stranded sheet structures folded within 240 ns, regardless of the forcefield model used. These results suggest that either $^{\text{D}}\text{P}^{\text{D}}\text{P}$ -II folds at time scales longer than 240 ns, or that $^{\text{D}}\text{P}^{\text{D}}\text{P}$ -II is not a stable well-defined β -sheet, which is consistent with previous NMR spectroscopic data [5].

2. Results and Discussion

The Folding@Home distributed computing platform [21] was used to simulate molecular dynamics (MD) trajectories, each up to 240 ns in length, for five different forcefields, for a total of ~ 8.2 ms of simulation. Simulations were performed using the GROMACS simulation package [22], with

AMBER94 [23], AMBER96 [24], AMBER99 [25], AMBER99 ϕ [26], and AMBER03 [27] forcefields (see Methods section). 1000 total trajectories were generated for the AMBER99 ϕ simulations, and 10,000 trajectories each were generated for the other forcefield simulations. Figure 2 shows the distribution of trajectory lengths for each forcefield tested.

Figure 2. The distribution of trajectories achieving a given trajectory length, shown for the forcefields tested in this study.



2.1. Simulated relaxation kinetics

Time-resolved FTIR measurements cannot directly determine whether a protein is folded, but instead report the status of backbone amide groups, which may be closely related. To best connect with the relaxation rates experimentally measured using FTIR [1], we therefore analyzed the ensemble time course of the total solvent-accessible surface area of backbone C=O groups, as well as the average radius of gyration of the entire molecule. In general, reaction coordinates must be carefully chosen because a poor choice of can yield projection-dependent results [28,29]. The C=O solvent-accessible surface area is a measure that closely connects with the measured amide I band, which is known to be sensitive to hydration status [30]. The radius of gyration is a global quantity that does a good job of characterizing the structural distribution and compactness of a conformational ensemble.

We simulated 1,000 trajectories each (100 each for AMBER99 ϕ) from 10 different starting configurations (Figure 3) taken randomly from a high-temperature equilibration trajectory of ^DP^DP-II started from a semi-extended state (see Methods). Figure 4a shows a typical trace of the ensemble-average radius of gyration over time (for a particular combination of forcefield and starting conformation), which fits well to a bi-exponential curve. Figure 6a shows a typical trace of the ensemble-average C=O solvent-accessible surface area over time. The kinetics also fit well to a bi-exponential curve. In both cases, the kinetics show a fast equilibration phase (usually $\tau_1 \sim 1-10$ ns) and a slower relaxation phase ($\tau_2 \sim 100$ ns). Similar kinetics were computed across all forcefields and starting conformations (Figure 4b). In most cases, the fast phase corresponds to fast equilibration of the starting conformation. Alternatively, in some cases, the fitted values of τ_1 were extremely short (~ 0.1

ns), less than the snapshot frequency, indicating that the kinetics may be better described as a single-exponential process with rate constant τ_2 . Numerical values for all fitted kinetic parameters are shown in the Supplementary Material (Tables S1 and S2).

Figure 3. Ten different starting conformations taken from a high-temperature trajectory were used to seed the simulations.

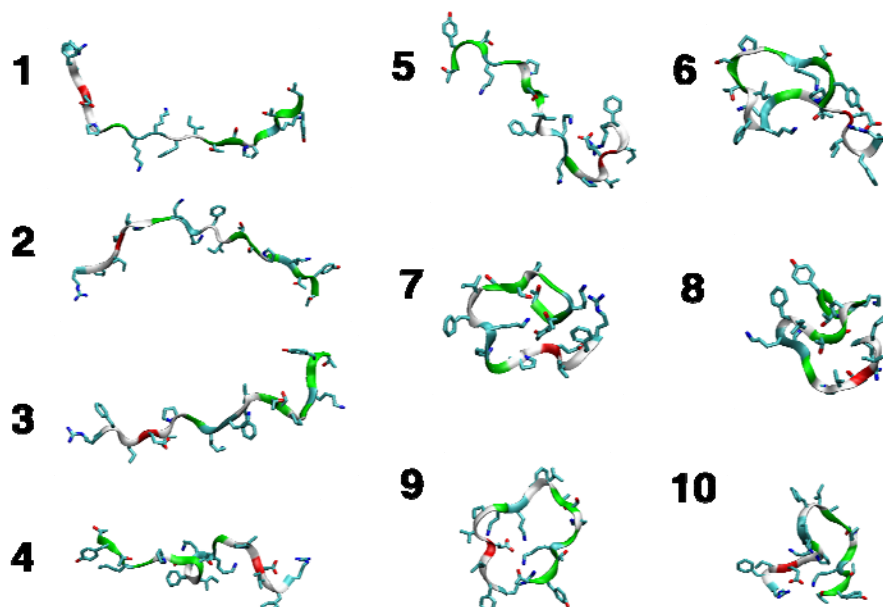
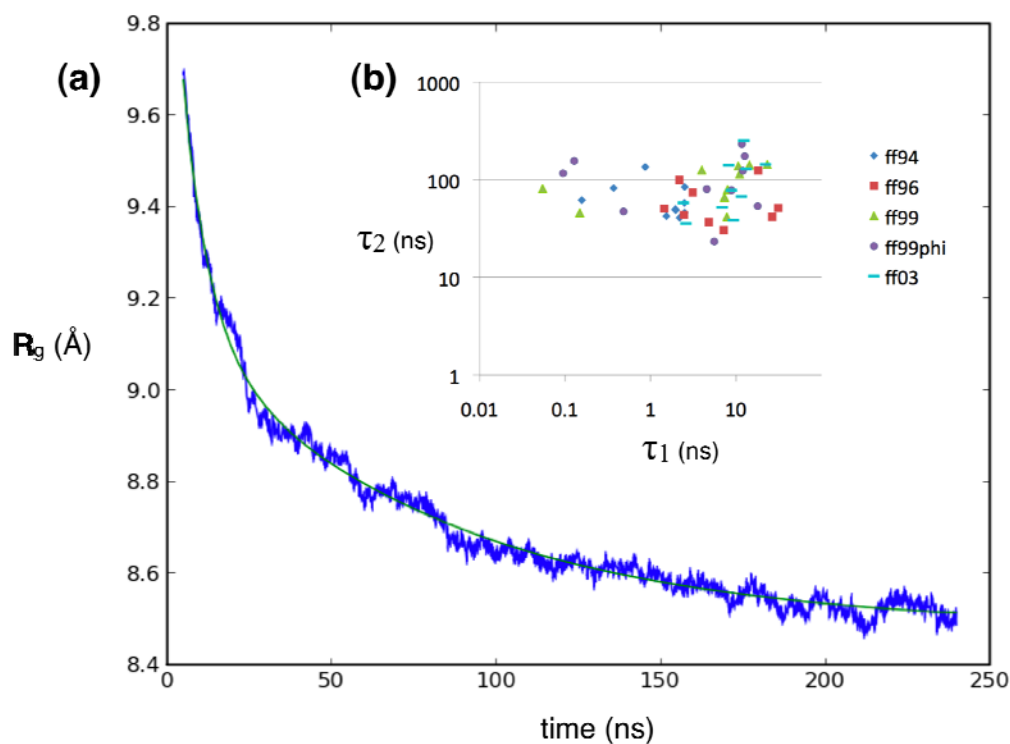


Figure 4. Simulated relaxation kinetics for ${}^{\text{D}}\text{P}^{\text{D}}\text{P-II}$, as characterized by the average radius of gyration. (a) An example trace of the average radius of gyration over time (blue) with the best-fit bi-exponential curve (green). (b) Fitted bi-exponential time constants τ_1 and τ_2 across all forcefields and starting conformations.



Regardless of forcefield choice, the slow relaxation times estimated from our simulations are consistent, ranging from about ~60 to ~100 ns for the radius of gyration reaction coordinate (Figures 4b, 5), and ~80 to ~150 ns (Figures 6b, 7) for the solvent-accessible surface area. Both compare very favorably to the experimentally measured relaxation time of $\sim 140 \pm 20$ ns obtained by T-jump infrared spectroscopy [1].

Figure 5. Average simulated relaxation times for $D^D P^D P-II$, for each forcefield, as characterized by average radius of gyration. Error estimates are computed from the standard deviation across the 10 starting conformations.

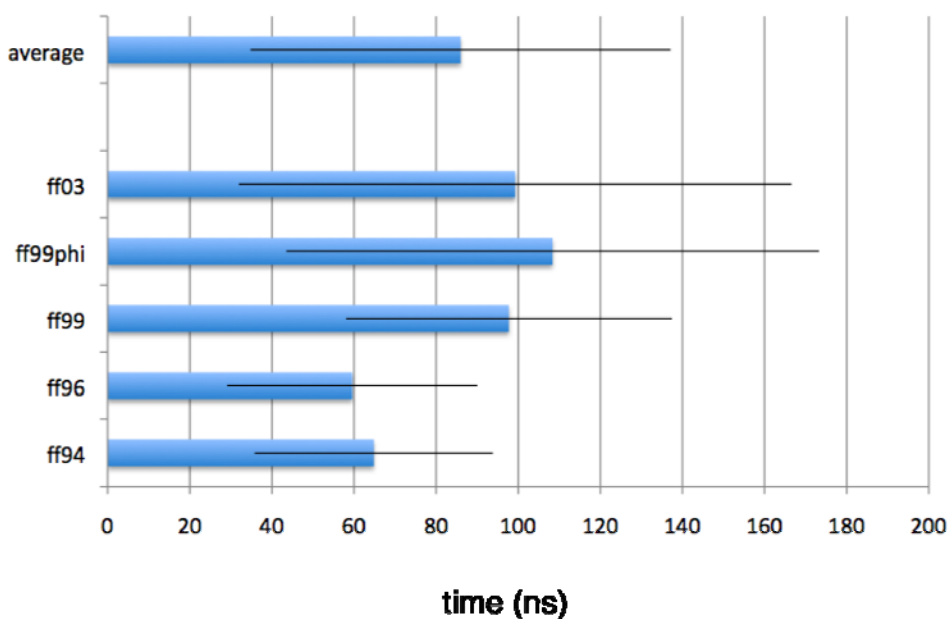


Figure 6. Simulated relaxation kinetics for $D^D P^D P-II$, as characterized by average C=O solvent-accessible surface area. Description is as Figure 4.

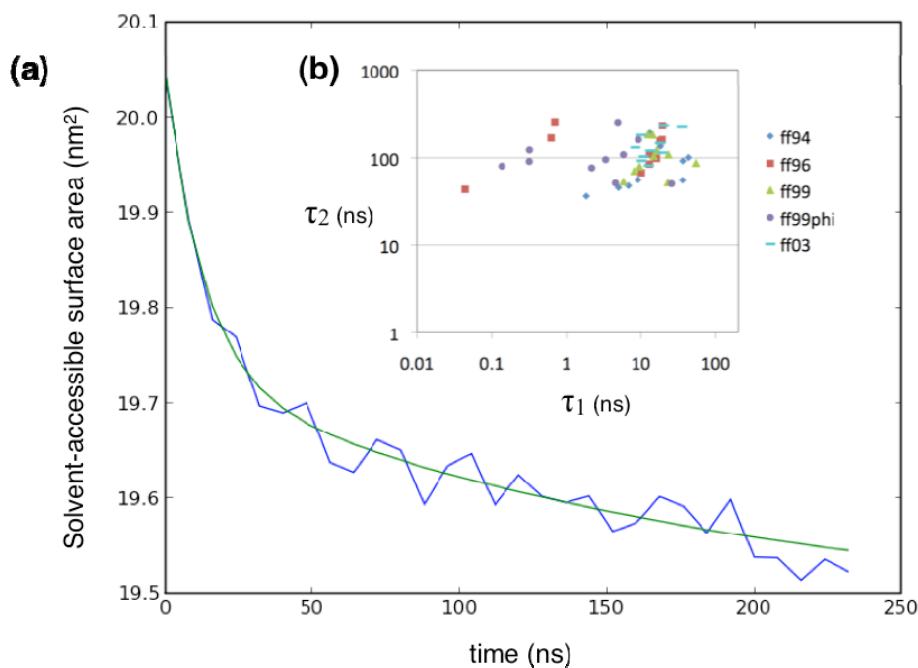
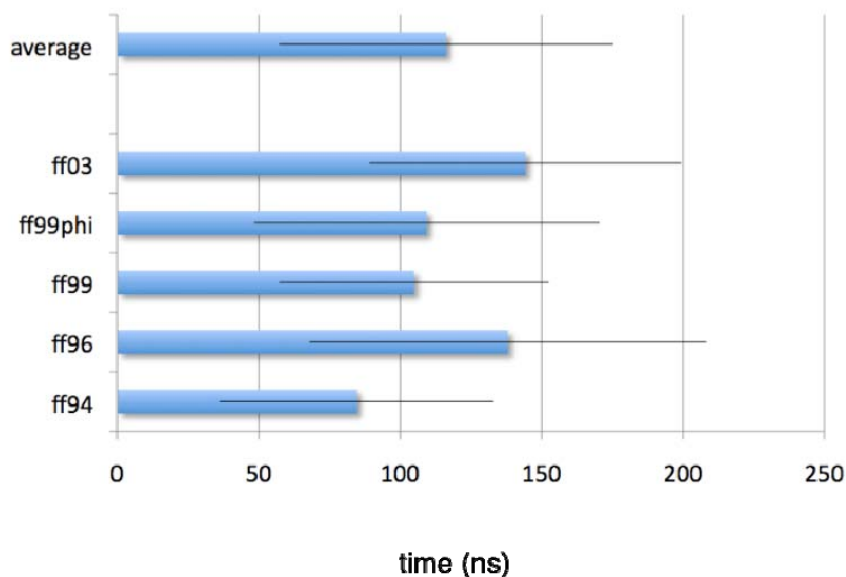


Figure 7. Average simulated relaxation times for $^D P^D P$ -II, for each forcefield, as characterized by average C=O solvent-accessible surface area. Error estimates are computed from the standard deviation across the 10 starting conformations.



The agreement between simulated and experimental rates is comparable to other contemporary examples of physical kinetics simulations [31]. The slightly faster relaxation rates observed in the simulations may in part reflect the anomalously high diffusion constant of the TIP3P water model [32].

The average radius of gyration at 240 ns across all forcefields is $8.32\text{\AA} \pm 0.47\text{\AA}$, and the average value of the exponential baseline, C , is $8.28\text{\AA} \pm 0.47\text{\AA}$. This reflects a more conformationally expanded ensemble than seen in simulations of $^D P^D P$, which showed radius of gyration of $\sim 7\text{\AA}$ for unfolded states, $\sim 6.5\text{\AA}$ for partially unfolded states, and $\sim 5.5\text{\AA}$ for a fully strand-paired native-state conformation [19].

2.2. Secondary structures over time

The per-residue secondary structure over time for each forcefield was calculated using the DSSP algorithm [33]. The general features observed across the different forcefields include fast formation of the $^D P G$ turn regions, and negligible amounts of sheet formation as quantified by the amount of backbone hydrogen-bonded strand content (see Supplementary Material). It should be noted that strand content may be somewhat underestimated due to the stringent definition required by DSSP.

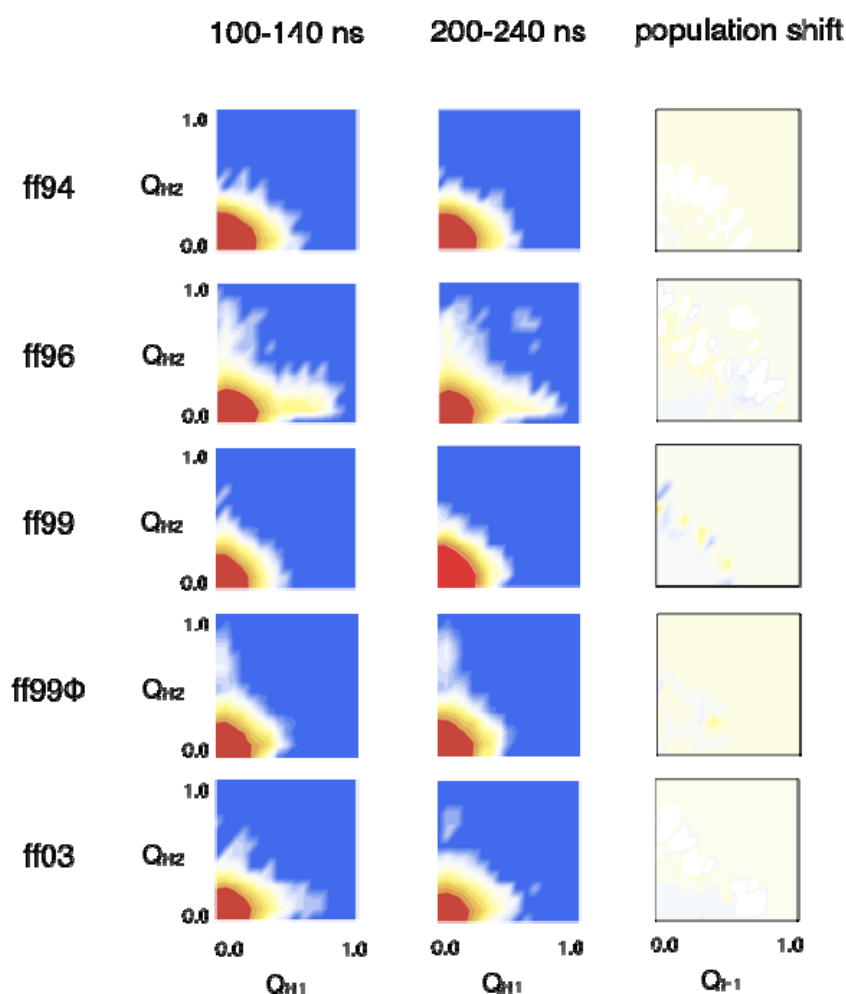
The amounts of secondary structure across different forcefields reproduce previously noted secondary structural biases [26]. For example, AMBER94 is slightly biased toward more helical conformations and has more populated turn regions (as defined by DSSP), while AMBER96 biased toward beta-sheet conformations, which detectable populations of strand (see Supplementary Material). The more modern forcefields of AMBER99, AMBER99phi, and AMBER03 all show comparable amounts of secondary structural propensities intermediate between AMBER94 and AMBER96. In all cases the DSSP populations are relatively static after ~ 100 ns.

2.3. Hairpin formation over time

2.3.1. QH1 and QH2 at 100-150 ns and 200-230 ns over time

To examine hairpin formation, we computed two quantities, Q_{H1} and Q_{H2} , reporting the fraction of “native” contacts in (N-terminal) hairpin 1 and (C-terminal) hairpin 2, respectively (see Methods). The quantities Q_{H1} and Q_{H2} were used as reaction coordinates to compute the landscape of sampled conformations at two time slices: 100-140 ns and 200-240 ns (Figure 8).

Figure 8. Conformational landscapes for ${}^D P^D P$ -II. Histograms of sampled populations were constructed in reaction coordinates Q_{H1} and Q_{H2} , which monitor the fraction of hairpin 1 and hairpin 2 contacts, respectively. Populations at times 100-140 ns and 200-240 ns are shown in the first two columns. The third column shows a difference map of the population shift over this time. Distributions are plotted on a log-scale, with each color gradation representing one unit $k_B T$ of free energy at room temperature.

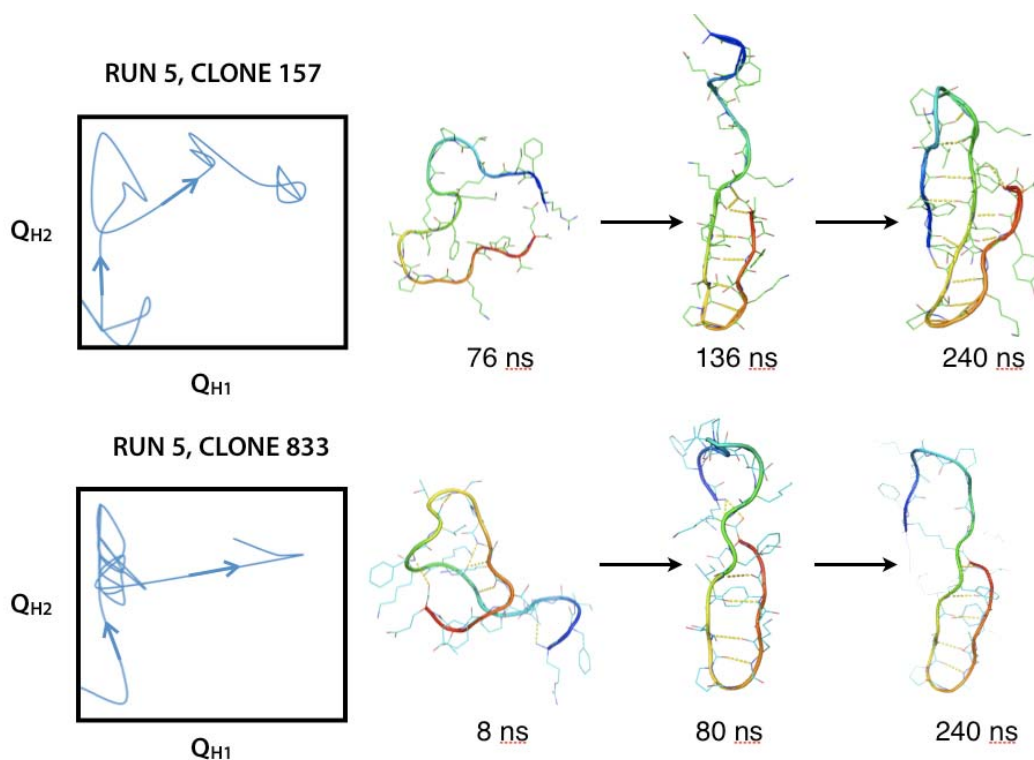


Regardless of the choice of forcefield, the conformational landscape mostly disfavors the formation hairpin. Recall that the less stable of the two ${}^D P^D P$ hairpins comprises the C-terminal sequence of ${}^D P^D P$ -II, corresponding to hairpin 2. With the exception of the AMBER96 simulations, only the formation of hairpin 2 is mostly observed, and only then with a population of 3% or less at 200 ns. For

the AMBER96 trajectories, formation of both hairpin 1 and hairpin 2 is observed. For all the simulations, comparisons of the conformational landscape at 100 ns and 200 ns shows very little change in hairpin populations on the ~100 ns time scale (Figure 8).

Folding to a three-stranded sheet is observed for only two out of a total of 10,000 AMBER96 trajectories (Figure 9). One of these two trajectories shows a fully hydrogen-bonded three-stranded sheet structure, while the other shows only hairpin 2 with defined hydrogen bonds, but is otherwise “native” according to inter-residue contacts defined by the Q_{H1} and Q_{H2} reaction coordinates.

Figure 9. Only two of 10,000 AMBER96 trajectories show folding events for $^D P^D P$ -II within 240 ns. Shown is the time course of reaction coordinates Q_{H1} and Q_{H2} , which monitor the fraction of hairpin 1 and hairpin 2, with conformational snapshots. The second of the two trajectories is “native” by our reaction-coordinate definition, although hairpin 1 does not have a fully hydrogen-bonded structure.



Across all of the forcefields we studied, most all of our simulations do not produce stable three-stranded hairpin conformations. We think that this result is very unlikely to be due to poor sampling. With as many as 10,000 simulation replicas per forcefield, there should be a strong likelihood of observing at least some trajectories reaching the folded state [34]. It is possible that forcefield deficiencies may be at work here, but we tested a wide range forcefields, and consistently found negligible amounts of three-stranded. Parallel simulation techniques to accelerate kinetic sampling also has its limits on short timescales where first-passage times are short compared to the folding time [35], but that is not the case here. If the experimentally observed relaxation does indeed correspond to folding, then the overlap in simulated and experimental relaxation time scales should be very favorable for observing transitions to native conformational ensembles.

Is $^{\text{D}}\text{P}^{\text{D}}\text{P-II}$ a stable folded three-stranded sheet? While Syud *et al.* reported qualitative NOE data for $^{\text{D}}\text{P}^{\text{D}}\text{P-II}$, this peptide was the least well-folded compared to the other designed sequences in this paper [5]. The measured NMR resonances were weak, and aggravated by poor dispersion, so only key inter-residue contacts hinting at the designed structure were reported (Syud and Gellman, personal communication). Combined with our simulation results, this suggests that perhaps $^{\text{D}}\text{P}^{\text{D}}\text{P-II}$ is unstable as a three-stranded sheet, and may not be a very relevant model system for studying beta sheet peptides.

Similar plasticity has been observed in another designed three-stranded sheet, the betanova peptide [36]. Both betanova and the $^{\text{D}}\text{PG}$ -turn peptides of Gellman *et al.* were designed with stable turns and hydrogen-bonded strand regions, to be used as model systems to study beta-sheet cooperativity. WW domains, by contrast, are three-stranded beta-sheet proteins found in nature, whose structures are well-defined [37]. Unlike designed beta-sheet peptides, WW domains additionally possess a conserved network of hydrophobic interactions between their termini. Thus, in general, beta-sheet model systems such as $^{\text{D}}\text{P}^{\text{D}}\text{P-II}$ may not have the necessary amount of long-range cooperative interactions needed to fully stabilize their structure.

2.4. Conformational clustering and Markov State Model (MSM) analysis

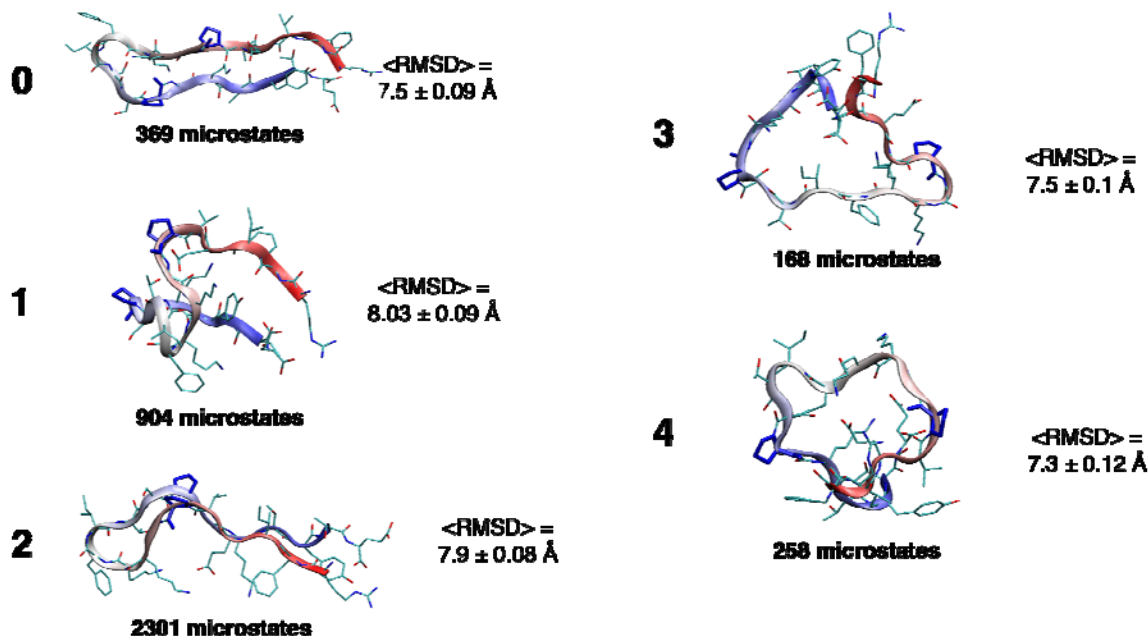
Kinetics-based conformational clustering was performed for all snapshots from the AMBER96 trajectories. The AMBER96 trajectories were chosen as they contained the greatest extent of beta-sheet structure, and the only observed folding events. Our clustering procedure was used to identify five macrostate clusters calculated to be the most metastable, which were used to construct Markov State Models [38-40] of the dynamics (see Methods).

We constructed a series of MSMs from matrices of macrostate transition counts, using different lag times ranging from 8 to 240 ns. The performance of these models reveals much about the underlying folding landscape.

The most striking result of the MSM-building procedure was our failure to identify well-separated metastable states that would indicate large activation barriers on the folding landscape. The first indication of this comes from our clustering algorithm, designed to identify the most kinetically metastable states. Only five metastable states were identified, and each contained a broad ensemble of microstate conformations, with average RMSD between any two microstates ranging from 7.3-8.0 Å (Figure 10).

Regardless of lag time, the spectrum of relaxation rates predicted by the MSM is broad, without a large gap that would indicate a pronounced separation of time scales (Figure S2, Supplementary Material). These results, at least within the time scale of our simulations, are not inconsistent with either multistate folding or the “downhill” folding interpretation of Xu *et al.* Moreover, as we increase the lag time used to build the models, the longest implied timescale also increases. If clear activation barriers were present, such that metastable dynamics on the > 10 ns time scale resulted, the implied timescales should level off as the lag time increases. This result is not simply a consequence of poor state definitions, because the kinetic clustering procedure we use should insure that the macrostates are the most metastable basins.

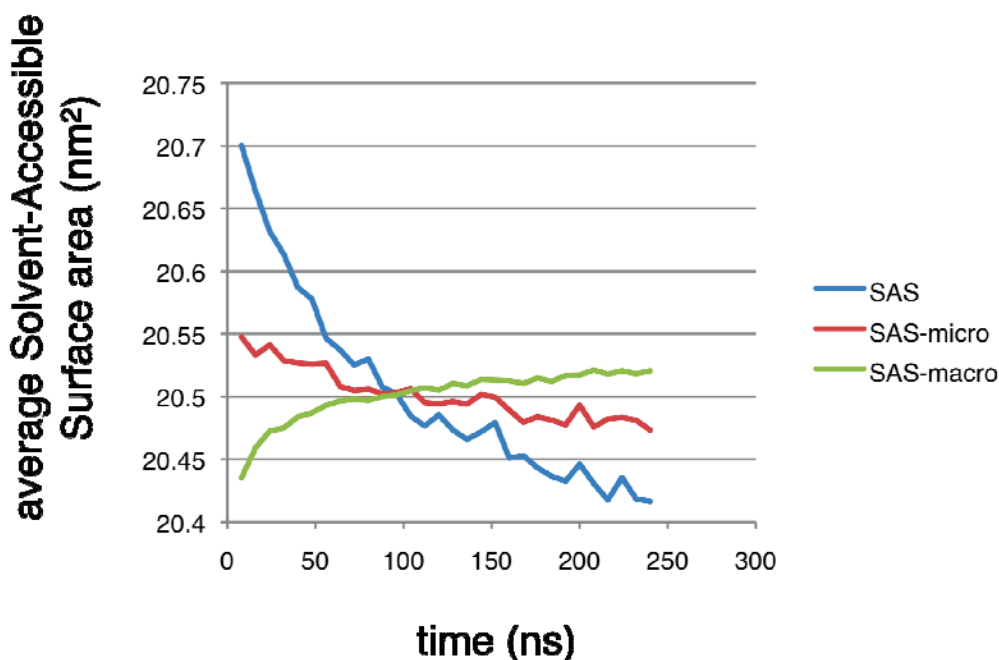
Figure 10. Kinetics-based clustering was used to find five maximally metastable macrostates (see Methods). The representative conformations shown for each state are the most probable conformations in that state. Shown next to each representative conformation is the average RMSD between microstates in that cluster, a measure of the compactness of the conformational ensemble, and the number of microstates (of 4000 total) comprising each macrostate.



The variability of simulated relaxations across the ten starting conformations offer an additional indication of the absence of large barriers. This is not only evident from the bi-exponential fits of average radius of gyration and average solvent accessible surface area over time, but also from individual MSMs we built using trajectory data generated from each conformation (Figures S3 and S4, Supplementary Material). Similar kinds of heterogeneity in relaxation dynamics for different starting conformations have been observed in previous parallel simulations of ultrafast folders [9].

The other striking result of our MSM-building procedure is the unexpected sensitivity of the average C=O solvent-accessible surface area (SAS) to expanded states (Figure 11). When we use the average SAS of each macrostate to compute a projection of the time evolution of the SAS observable, the effects of averaging over each macrostate is severe enough to produce a signal that increases over time instead of decreasing. When the average SAS is projected onto each microstate, this effect is less severe, yet is still present. The simulation data suggest that the average SAS is more sensitively dependent on expanded conformations that quickly collapse, as compared to more compact conformations. Given our good overall results in recapitulating experimentally observed relaxation rates, we remain confident that our representative set of starting conformations is a useful ensemble to compare with FTIR T-jump experiments. However, the SAS projections underscore the importance of choosing experimental observables that overlap well with the reaction coordinate of interest (in this case, the folding reaction) as to best report the underlying dynamics.

Figure 11. Average C=O solvent-accessible surface area (SAS) over time computed from simulation snapshot data (blue), compared to the average SAS of each microstate projected onto the 4000 microstate populations over time (red), and average SAS of each macrostate projected onto 5 macrostate populations over time (green). The differential effects produced by averaging over microstates and macrostates indicate a sensitive dependence of the SAS observable on short-lived expanded conformations.



One question partially addressed by our work is how experiment and simulation might be used to distinguish between two-state vs. “downhill” folding. As we have shown, one potential indication of so-called “downhill” folding from simulations might be a failure to build a Markovian kinetic model able to describe dynamics as transitions between well-defined metastable states. However, our work suggests that perhaps $^D P^D P-II$ is not a well-defined beta-sheet structure, which brings into question what is meant by “folding” in this case.

Can simulations help suggest experiments that could discriminate downhill vs. activated folding? This is a challenging task, as the observed experimental kinetics for “downhill” folders may depend on many factors. Liu and Gruebele, using one-dimensional Langevin models, present an excellent elucidation of the possible experimental outcomes that can arise from slight differences in folding landscapes (such as native-biases, roughness, and barrier heights) and the reaction coordinate-dependence of reporter probes [16]. Using simulations to identify observables that connect well with folding reaction-coordinates may be particularly useful. For example, our simulations of $^D P^D P-II$ suggest that the C=O solvent-accessible surface area (SAS) is more sensitive to expanded versus compact states. The insensitivity may be in part because the SAS is an aggregate measure across all peptide residues. To the extent that the SAS correlates with the amide I band spectroscopic observable in FTIR T-jump experiments, we suggest that multiple time-resolved FTIR experiments using isotopic labeling of specific residues, combined with microscopic information about peptide conformations from simulation, would help to better resolve folding landscapes for ultrafast folding proteins.

3. Experimental Section

3.1. System preparation and simulation protocol

Ten initial starting conformations were selected iteratively from a 1 ns stochastic dynamics (SD) simulation at 3000K, with 9 Å cutoffs for Coulomb and vdW interactions, integration time step of 1 fs, neighbor searching on a grid every 10 steps, at solvent (shear) viscosity of 10 ps⁻¹. Ten conformations were picked iteratively from a collection of snapshots saved every 1 ps. After picking the first conformation, the most diverse structure (as measured by RMSD) was picked as the next. This procedure was repeated to create a structurally diverse starting set. Each chosen (nearly random) structure was then minimized and equilibrated for the production runs.

Production runs were performed using the TIP3P water model [32] for explicit solvation. A rhombic dodecahedral box of largest dimension 58.7Å was used with periodic boundary conditions. The box contained a ^DP^DP-II molecule with uncapped termini, approximately 4,650 water molecules (this number varied slightly with starting conformation) and two chloride counterions to achieve a net neutral charge. Molecular dynamics (MD) simulations were ran at 308 K in the NVT ensemble with a 2 fs integration time step. The same cut-off and neighbor-list settings above were employed, along with a reaction-field electrostatics model, Berendsen temperature coupling, and constrained bonds with the LINCS algorithm. Trajectory snapshots were recorded every 100 ps. Total C=O solvent-accessible surface area was calculated for each snapshot from the set of all carbon and oxygen atoms in the backbone carbonyl groups, using a solvent probe radius of 1.4Å.

3.2. Exponential curve fitting

Best-fit parameters $\beta^*=(A, B, C, \tau_1, \tau_2)$ for bi-exponential curves of the form $f(t) = A\exp(-t/\tau_1) + B\exp(-t/\tau_2) + C$ were calculated for time series of the average radius of gyration and C=O solvent-accessible surface area, by using a simulated annealing protocol to minimize the sum of squared errors. The first 5 ns of the time series were omitted from the fitting procedure. Variances σ_i^2 in average radius of gyration at each time point i were calculated by non-parametric bootstrap of 100 samples. Errors in parameter estimates for each β_j were calculated as diagonal elements of the covariance matrix $\mathbf{C}(\beta^*) = (\mathbf{F}^T \mathbf{W} \mathbf{F})^{-1}$, where \mathbf{F} is the (N x 5) Jacobian matrix

$$F_{ij} = \left. \frac{\partial f(t_i, \beta)}{\partial \beta_j} \right|_{\beta^*} \quad (2)$$

and \mathbf{W} is an N x N diagonal matrix of inverse variances: $W_{ij} = 1/\sigma_i^2$ for $i=j$, $W_{ij} = 0$ for $i \neq j$ [41].

3.3. Secondary structure and “native” hairpin contacts

The DSSP algorithm was used to assess the extent of helix, strand (sheet), turn, and loop secondary structures [33]. DSSP recognizes eight types of secondary structures based on hydrogen bonding patterns: G (3₁₀ helix), H (alpha helix), I (pi helix), B (beta bridge), E (extended sheet), T (turn), S (loop). We monitor helix content as the total of G, H, I, the strand content as the total of B and E.

Q_{H1} and Q_{H2} report the fraction of native contacts present for (N-terminal) hairpin 1 and (C-terminal) hairpin 2, respectively. We use the same criteria derived by Roe *et al.*, who used a model of the native conformation to define “native” contacts in each of the two possible hairpins [19]. For hairpin 1, the set of native sidechain contacts (C_α for glycine) is defined as residue pairs (R1,I3), (R1,T12), (F2,I1), (I3,V5), (I3,T12), (E4,G7), (E4,K9), (V5,F10) and native backbone hydrogen bonds (R1-H, T12-O), (R1-O, T12-H), (I3-H, F10-O), (I3-O, F10-H). For hairpin 2, the set of native sidechain contacts is defined as residue pairs (K8,F10), (K8,20), (K9,20), (F10,T17), (F10,T19), (I11,S13), (I11,Y18), (I11,E20), (T12, DP-14), (T12,G15), (T12,T17), (S13,Y8) and native backbone hydrogen bonds (K9-H, E20-O), (I11-H, Y18-O), (I11-O, Y18-H), (S13-H, K16-O). A contact between sidechains is defined when centroid distances $< 6.5\text{\AA}$ and a backbone contact is defined when hydrogen donor-acceptor distance $< 2.5\text{\AA}$.

3.3. Kinetics-based clustering for building Markov State Models (MSM)

Representative conformations were extracted from the simulation data using a procedure previously described [42], though constant temperature simulations were used. This method uses Markov State Models (MSMs) to identify kinetically related regions of phase space. Thus, two conformations will be found in the same state if a simulation can move between them quickly but will be grouped into different states if transitioning between them is slow. The definitions of fast and slow are based on the timescales observed in the simulations [39,42].

The first step in building such an MSM is to group conformations with a high degree of structural similarity into small sets called microstates. In this study 4,000 microstates were generated based on their all-atom RMSD using a k -centers clustering algorithm [43]. A desirable feature of this algorithm is that the resulting microstates have approximately equal volumes so their populations are directly related to their densities, or free energies. If each microstate is sufficiently small then it is assumed that structural similarity is equivalent to kinetic similarity since it should take a very short time to transition between very similar conformations. Kinetically related microstates, as judged by the number of transitions between them observed in the data, are then grouped together using the PCCA algorithm and this lumping is refined using a simulated annealing scheme [38,44,45]. The center of the most populated microstate from each macrostate is then selected as the representative conformation for that macro state as it is the most probable.

3.4. Markov State Model (MSM) construction

The matrix of transition probabilities \mathbf{T} between the five macrostates was computed from the trajectory data. The entries of this matrix T_{ij} contain the probability of transitioning from state i to state j in time τ , which ranged from 8 ns to 240 ns. Diagonalization of $(\mathbf{T}^T - \mathbf{1})$ produces a set of eigenvalues μ_k and corresponding eigenvectors \mathbf{e}_k which describe the dynamics of state populations $\mathbf{p}(t)$ as a linear combination of relaxation processes:

$$\mathbf{p}(t) = \sum_k \alpha_i \mathbf{e}_k e^{-\lambda_k t} \quad (1)$$

where $\lambda_k = [-\ln \mu_k]/\tau$, and the α_i are determined by the initial state populations $\mathbf{p}(0)$ [38,39]. Thus λ_k^{-1} are the set of implied timescales involved in the relaxation dynamics.

4. Conclusions

We performed massively parallel folding simulations of ^DP^DP-II to investigate the conformational dynamics underlying its nanosecond refolding dynamics. The simulated relaxation rates, as monitored by average radius of gyration and average C=O solvent-accessible surface area, agree well with the single-exponential relaxation rates experimentally measured by T-jump FTIR. Furthermore, Markov state models built from the trajectory data do not show a separation of metastable timescales consistent with large activation barriers. These results, at least within the time scale of our simulations, are not inconsistent with either multistate folding or the “downhill” folding interpretation of Xu *et al.* However, despite the agreement with experimental kinetics, we observe very few trajectories that fold to stable three-stranded beta-sheet structures. These results suggest that either ^DP^DP-II folds at time scales longer than 240 ns, or that ^DP^DP-II is not a well-defined three-stranded β -sheet. The latter interpretation is consistent with previous NMR spectroscopic data [5].

Supplementary Materials

Supplementary materials are available online at <http://www.mdpi.com/1422-0067/10/3/1013/s1>.

Acknowledgements

The authors would like to thank Faisal Syud and Samuel Gellman for their insightful comments about the ^DP^DP-II design, and Frank Noe for his comments on Markov State Models. We acknowledge support from NSF FIBR (NSF EF-0623664) and NIH (NIH U54 GM072970), and thank the valued contributors of the Folding@Home distributed computing project, without whom this work would not have been possible.

References and Notes

1. Xu, Y.; Purkayastha, P.; Gai, F. Nanosecond folding dynamics of a three-stranded beta-sheet. *J. Am. Chem. Soc.* **2006**, *128*, 15836-15842.
2. Espinosa, J.F.; Syud, F.; Gellman, S. Analysis of the factors that stabilize a designed two-stranded antiparallel β -sheet. *Protein Sci.* **2002**, *11*, 1492-1505.
3. Stanger, H.; Gellman, S.H. Rules for antiparallel β -sheet design: D-Pro-Gly is superior to L-Asn-Gly for β -hairpin nucleation. *J. Am. Chem. Soc.* **1998**, *120*, 4236-4237.
4. Schenck, H.; Gellman, S.H. Use of a designed triple-stranded antiparallel β -sheet to probe β -sheet cooperativity in aqueous solution. *J. Am. Chem. Soc.* **1998**, *120*, 4869-4870.
5. Syud, F.; Stanger, H.; Mortell H.S.; Espinosa, J.F.; Fisk, J.D.; Fry, C.G.; Gellman, S.H. Influence of strand number on antiparallel beta-sheet stability in designed three- and four-stranded beta-sheets. *J. Mol. Biol.* **2003**, *326*, 553-568.

6. Arora, P.; Oas, T.; Myers J. Fast and faster: A designed variant of the B-domain of protein A folds in 3 microsec. *Protein Sci.* **2004**, *13*, 847-853.
7. Kubelka, J.; Chiu, T.; Davies, D.; Eaton, W.; Hofrichter, J. Sub-microsecond protein folding. *J. Mol. Biol.* **2006**, *359*, 546-553.
8. Yang, W.; Gruebele, M. Folding λ -Repressor at Its Speed Limit. *Biophys. J.* **2004**, *87*, 596-608.
9. Ensign, D.; Kasson, P; Pande, V. Heterogeneity Even at the Speed Limit of Folding: Large-scale Molecular Dynamics Study of a Fast-folding Variant of the Villin Headpiece. *J. Mol. Biol.* **2007**, *374*, 806-816.
10. Kubelka, J.; Hofrichter, J.; and Eaton, W.A. The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76-88.
11. Eaton, W.A.; Muñoz, V.; Thompson, P.A.; Henry, E.R. Kinetics and Dynamics of Loops, Alpha-Helices, Beta-Hairpins, and Fast-Folding Proteins. *Acc. Chem. Res.* **1998**, *31*, 745-753.
12. Ghosh, K.; Ozkan, S.B.; Dill K.A. The ultimate speed limit to protein folding is conformational searching. *J. Am. Chem. Soc.* **2007**, *129*, 11920-11927.
13. Deechongkit, S.; Nguyen, H.; Jager, M.; Powers, E.; Gruebele, M; Kelly, J.W. β -Sheet folding mechanisms from perturbation energetics. *Curr. Opin. Struct. Biol.* **2006**, *16*, 94-101.
14. Xu, Y.; Bunagan, M.R.; Tang, J.; Gai, F. Probing the Kinetic Cooperativity of β -Sheet Folding Perpendicular to the Strand Direction. *Biochemistry* **2008**, *47*, 2064-2070.
15. Gruebele, M. Comment on probe-dependent and nonexponential relaxation kinetics: Unreliable signatures of downhill protein folding. *Proteins* **2008**, *70*, 1099-1102.
16. Liu, F; Gruebele, M. Downhill dynamics and the molecular rate of protein folding. *Chem. Phys. Lett.* **2008**, *461*, 1-8.
17. Smith, A; Tokmakoff, A. Probing local structural events in β -hairpin unfolding with transient nonlinear infrared spectroscopy. *Angew. Chem. Int. Ed.* **2007**, *46*, 7984-7987.
18. Wang, H.; Sung, S. Molecular dynamics simulations of three-strand β -sheet folding. *J. Am. Chem. Soc.* **2000**, *122*, 1999-2009.
19. Roe, D.R.; Hornak, V.; Simmerling, C. Folding cooperativity in a three-stranded β -sheet model. *J. Mol. Biol.* **2005**, *352*, 370-381.
20. Kuznetsov, S.V.; Hilario, J.; Keiderling, T.A.; Ansari, A. Spectroscopic studies of structural changes in two-sheet-forming peptides show an ensemble of structures that unfold noncooperatively. *Biochemistry* **2003**, *42*, 4321-4332.
21. Shirts, M.R.; Pande, V.S. Screen savers of the world, unite! *Science* **2000**, *290*, 1903-1904.
22. Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306-317.
23. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, Jr., K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W; Kollman, P.A. A second generation force field for the simulation of proteins nucleic acids and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
24. Kollman, P.; Dixon, R.; Cornell, W; Fox, T.; Chipot, C.; Pohorille, A. The development/application of a "minimalist" organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. In *Computer Simulations of*

Biomolecular Systems: Theoretical and Experimental Applications, van Gunsteren, W.F., Wiener, P.K., Eds.; Escom: Dordrecht, The Netherlands, 1997; pp. 83-96.

25. Wang, J.; Cieplak, P.; Kollman, P.A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049-1074.
26. Sorin, E.J.; Pande, V.S. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.* **2005**, *88*, 2472-2493.
27. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M.L.; Xiong, G. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comp. Chem.* **2003**, *24*, 1999-2012.
28. Best, R.B.; Hummer, G. Reaction coordinates and rates from transition paths. *Proc. Natl. Acad. Sci.* **2005**, *102*, 6732-6737.
29. Juraszek, J.; Bolhuis, P. Rate constant and reaction coordinate of Trp-Cage folding in explicit water. *Biophys. J.* **2008**, *95*, 4246-4257.
30. Walsh, S.T.R.; Cheng, R.P.; Wright, W.W.; Alonso, D.O.V.; Daggett, V.; Vanderkooi, J.M.; DeGrado, W.F. The hydration of amides in helices; a comprehensive picture from molecular dynamics, IR, and NMR. *Protein. Sci.* **2003**, *12*, 520-531.
31. Snow, C.; Sorin, E.; Rhee, Y.; Pande, V. How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 43-69.
32. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.
33. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577-2637.
34. Shirts, M.R.; Pande, V.S. Mathematical analysis of coupled parallel simulations. *Phys. Rev. Lett.* **2001**, *86*, 4983-4987.
35. Marianayagam, N.; Fawzi, N.L.; Head-Gordon, T. Protein folding by distributed computing and the denatured state ensemble. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16684-16689.
36. Colombo, G.; Roccatano, D.; Mark, A.E. Folding and stability of the three-stranded beta-sheet peptide betanova: Insights from molecular dynamics simulations. *Prot. Struct. Func. Genet.* **2002**, *46*, 380-392.
37. Macias, M.J.; Gervais, V.; Civera, C.; Oschkinat, H. Structural analysis of WW domains and design of a WW prototype. *Nat. Struct. Biol.* **2000**, *7*, 375-379.
38. Chodera, J.D.; Singhal, N.; Pande, V.S.; Dill, K.A.; Swope, W.C. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* **2007**, *126*, 155101.
39. Noe, F.; Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.* **2008**, *8*, 154-162.
40. Chodera, J.D.; Swope, W.C.; Pitera, J.W.; Dill, K.A. Long-time protein folding dynamics from short-time molecular dynamics simulations. *Multiscale Model. Sim.* **2006**, *5*, 1214-1226.
41. Bates, D.M.; Watts, D.G. *Nonlinear Regression Analysis and Its Applications*. Wiley: New York, 1988.

42. Bowman, G.R.; Huang, X.; Pande, V.S. Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* **2009**, in press.
43. Dasgupta, S.; Long, P.M. Performance guarantees for hierarchical clustering. *J. Comp. Sys. Sci.* **2005**, *70*, 555-569.
44. Deuffhard, P. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Lin. Alg. Appl.* **2000**, *315*, 39-59.
45. Deuffhard, P.; Weber, M. Robust Perron cluster analysis in conformation dynamics. *Lin. Alg. Appl.* **2005**, *398*, 161-184.

© 2009 by the authors; licensee Molecular Diversity Preservation International, Basel, Switzerland. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).