WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a β-sheet miniprotein

Stefanie Muff and Amedeo Caflisch*

Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

ABSTRACT

The effects of a single-point mutation on folding thermodynamics and kinetics are usually interpreted by focusing on the native structure and the transition state. Here, the entire conformational spaces of a 20-residue three-stranded antiparallel β-sheet peptide (double hairpin) and of its single-point mutant W10V are sampled close to the melting temperature by equilibrium folding–unfolding molecular dynamics simulations for a total of 40 μs. The folded state as well as the most populated free energy basins in the denatured state are isolated by grouping conformations according to fast relaxation at equilibrium. Such kinetic analysis provides more detailed and useful information than a simple projection of the free energy. The W10V mutant has the same native structure as the wild type peptide, and similar folding rate and stability. In the denatured state, the N-terminal hairpin is about 20% more structured in W10V than the wild type mainly because of van der Waals interactions. Notably, the W10V mutation influences also the van der Waals energy at the transition state ensemble causing a shift in the ratio of fluxes between two different transition state regions on parallel folding pathways corresponding to nucleation at either of the two β-hairpins. Previous experimental studies have focused on the effects of denaturant-dependent or temperature-dependent changes in the structure of the denatured state. The atomistic simulations show that a single-point mutation in the central strand of a β-sheet peptide results in remarkable changes in the topography of the denatured state ensemble. These changes modulate the relative accessibility of parallel folding pathways because of kinetic partitioning of the denatured state. Therefore, the observed dependence of the folding process on the starting ensemble raises questions on the biological significance of* in vitro *folding studies under strongly denaturing conditions.

Proteins 2008; 70:1185–1195.

***Key words:** complex network; non-native interactions; transition state; multiple folding pathways; free-energy surface.*

INTRODUCTION

Despite significant advances in the understanding of the protein folding process,[1–8] the link between primary structure and folding thermodynamics and kinetics is not completely clear. In particular, the full elucidation of the mechanism(s) of protein folding requires an in-depth understanding of the denatured state that is the starting ensemble. At physiological (i.e., mainly folding) conditions the denatured state is not only elusive to experimental characterization but also complex and highly heterogeneous even for small proteins and structured peptides.[9–12] Theoretical models and computational approaches have been developed to try to capture the complexity of the free energy surface which governs protein folding.[1,10–13] Evidence from experimental[14] and computational studies[15] indicates that the unfolded population is not featureless and can retain native-like topology, even at high concentration of denaturant according to residual dipolar couplings nuclear magnetic resonance (NMR) measurements.[16] Recently, several biophysical studies have focused on misfolding and the denatured state because of their role in protein-aggregation diseases.[17] As an example, conformers with non-native aromatic clusters have been suggested to play a role in the initiation of amyloidosis from the acid-unfolded state of β_2-microglobulin according to NMR experiments coupled with site-directed mutagenesis.[18]

The main motivation of the present simulation study was to investigate the influence of a single-point mutation on the conformational space and folding pathways of a structured peptide. It was decided to investigate a designed three-stranded antiparallel β-sheet peptide of 20 residues[19] because

Abbreviations: CSN, conformation space netwrok; TSE, transition state ensemble

This Supplementary Material referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat.

Grant sponsor: Swiss National Science Foundation.

*Correspondence to: Amedeo Caflisch, Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.
E-mail: caflisch@bioc.unizh.ch

Received 6 February 2007; Revised 10 April 2007; Accepted 16 April 2007
Published online 10 September 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.21565

it folds reversibly to the correct NMR structure in molecular dynamics simulations[20–22] with a very efficient implicit solvent model.[23] The sequence of the wild type peptide is Thr_1-Trp_2-Ile_3-Gln_4-Asn_5-Gly_6-Ser_7-Thr_8-Lys_9-Trp_{10}-Tyr_{11}-Gln_{12}-Asn_{13}-Gly_{14}-Ser_{15}-Thr_{16}-Lys_{17}-Ile_{18}-Tyr_{19}-Thr_{20} where the Gly-Ser segment at positions 6–7 and 14–15 promote the formation of tight turns resulting in a three-stranded β-sheet consisting of two β-hairpins "sharing" the central strand.[19] The investigated mutant is Trp10Val (abbreviated, using the single-letter code of the amino acids, as W10V hereafter). In long molecular dynamics runs with several folding and unfolding events, the complete information about the native and non-native free energy basins is present in the trajectory. Therefore, peptide conformations can be grouped into free energy minima according to the equilibrium dynamics (kinetic grouping analysis). In the present study, the conformations are sampled close to the melting temperature by simulations totaling 20 μs for each peptide with more than 100 folding and unfolding events. The kinetic analysis reveals a switch of folding pathways, that is, folding begins mainly from the C-terminal and N-terminal hairpin for the wild type and the single-point mutant W10V, respectively, which is consistent with the differences in the denatured state ensemble of the two peptides. This consistency is nontrivial and originates from the very slow equilibration in the denatured state of Beta3s and W10V. In fact, analysis of the molecular dynamics trajectories provides evidence that barriers separating the free-energy basins in the denatured state ensemble are higher than those between individual non-native basins and the folded state, which is a clear indication of kinetic partitioning of the denatured state.

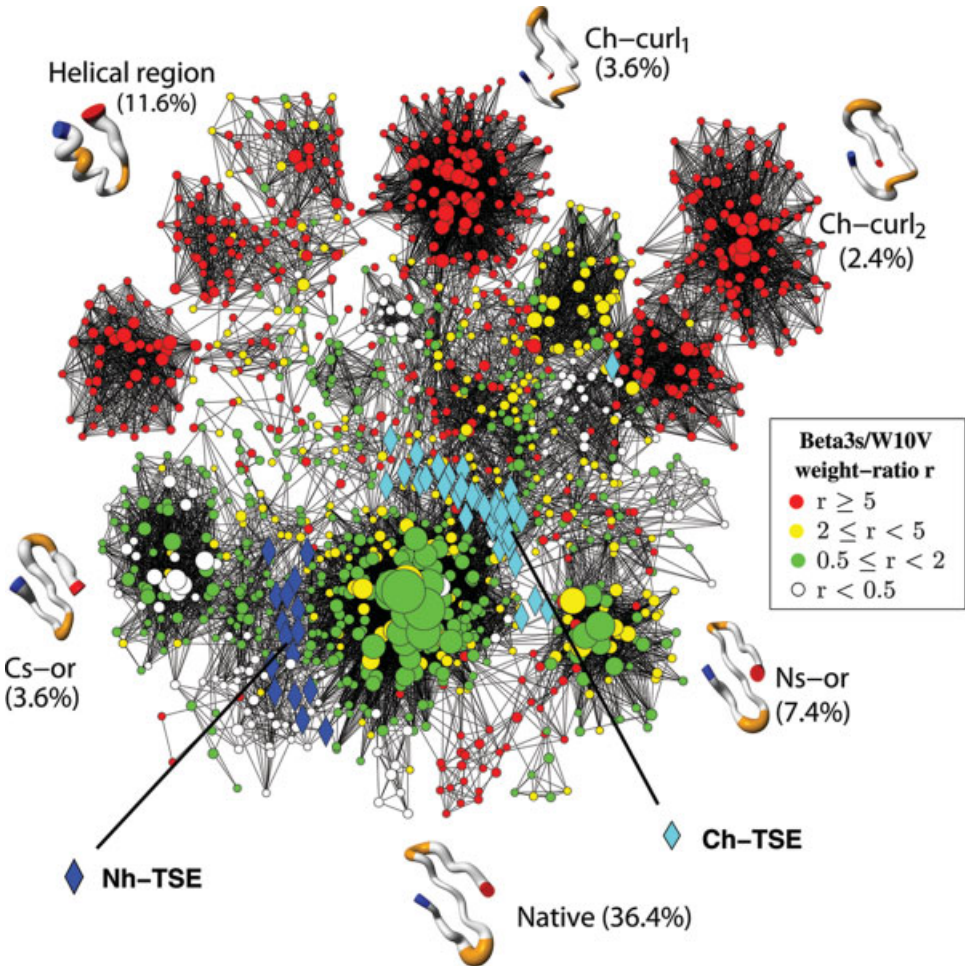

Figure 1

The CSN of Beta3s. Each node (i.e., conformation) of the network represents a secondary structure string. The surface of each node is proportional to its statistical weight and only the 1430 nodes with at least 40 snapshots for Beta3s are shown to avoid overcrowding. Nodes are colored according to the Beta3s/W10V mutant weight-ratio. Conformations in the most populated basins are shown by flexible tubes of variable diameter with N-terminus in blue, C-terminus in red and residues Gly6, Ser7, Gly14, and Ser15, which are at the two turns in the folded structure, in orange. Names of conformations are explained in the legend of Table I. Blue and cyan diamonds emphasize TSE nodes with N-terminal and C-terminal hairpin formed, respectively. This Figure was made using visone (www.visone.de) and MOLMOL.[27]

DOI 10.1002/prot

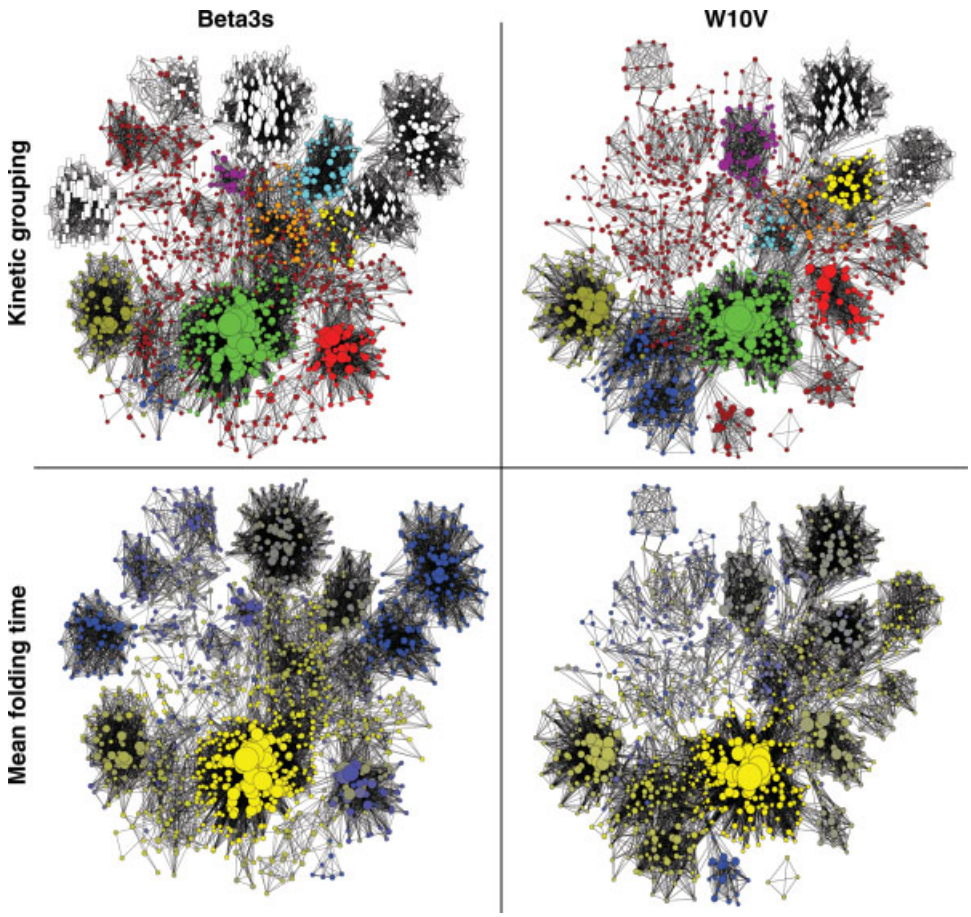

Figure 2

The CSN of Beta3s (left) and W10V (right). (Top) The coloring scheme of the kinetic grouping is chosen such that a large basin with the same most populated node in Beta3s and W10V has the same color in both CSNs (last column in Table 1). White nodes are used for basins populated significantly in only one of the two peptides. Nodes belonging to less populated basins are in brown. (Bottom) The coloring according to the mean folding time (τ_f), which is the kinetic distance from the native state, changes continuously from yellow (folding time of 0 ns) through olive (about 100 ns) to blue (200 ns). Basins identified by the kinetic grouping appear as regions of homogeneous color.

Parallel folding pathways and switches have been already observed[24,25] using protein-engineering Φ-value analysis,[26] but those studies did not report on changes in the denatured state ensemble, which is taken as the reference state in Φ-value analysis. Denaturant-dependent and temperature-dependent changes in the structured denatured state of small helical proteins have been reported.[14] The present simulation results go beyond the available experimental observations by unmasking the consequences of a hydrophobic side chain mutation on both the denatured state ensemble and transition state conformations of a β-sheet peptide.

RESULTS

The following analysis is based on a total simulation time of 20 μs at 330 K for each peptide. The wild type peptide (called Beta3s henceforth) visits 262,433 conformations, that is, nodes of the conformation space network (CSN,[11] see Methods), with a total of 534,383 direct transitions (i.e., CSN links) among nodes. The CSN of the W10V mutant is made up of 245,032 nodes and 476,721 links. A total of 120 and 105 folding events, that is, visits to the native node, were observed for Beta3s and W10V, respectively, with an average folding time of about 0.1 μs for both peptides.

Native state basin

The most populated conformation of Beta3s and its W10V mutant has a statistical weight of 5.6 and 8.8%, respectively. It is the three-stranded antiparallel β-sheet with type II′ turns at residues 6–7 and 14–15 (secondary structure string -EEEESSEEEEEESSEEEE-) in agreement with NMR data.[19] Using this conformation as "attractor," the kinetic grouping (see Methods and Suppl. Mat. for the

illustrative application of kinetic grouping to the alanine dipeptide) yields a native-basin weight of 36 and 41% for Beta3s and W10V, respectively. The CSN of Beta3s is a useful illustration of the underlying free energy surface and its dynamic connectivity. It is colored according to the Beta3s/W10V ratio of the statistical weight of individual nodes (i.e., conformations) in Figure 1, while the corresponding figure for W10V is included in the Suppl. Mat (Fig. S1). Individual nodes in the native state basin have very similar statistical weight for both peptides. As an example, there is only a 1% probability (weighted by the number of snapshots in a node and considering nodes visited at least 10 times during the total simulation time of 40 μs) that the Beta3s/W10V weight-ratio of a native-basin node is larger than 5 or smaller than 1/5, which corresponds to an absolute value of the free energy difference larger than 1 kcal/mol at 330 K. Furthermore, an average value of only 2.1 ± 1.4 Å is obtained for the root mean square deviation of C_α carbon coordinates measured on the 10^8 pairs of structures from subsets of 10^4 native state snapshots for each peptide. The corresponding values within the native basin of Beta3s and W10V are 2.6 ± 1.7 Å and 1.4 ± 0.6 Å, respectively. Hence, the folded state of the two peptides is essentially identical.

Denatured state ensemble

The two peptides show significant differences in the denatured state ensemble (Figs. 1 and 2 top). There is a 33% probability (weighted as explained above) that the Beta3s/W10V weight-ratio of a denatured-state node is larger than 5 (red nodes in Fig. 1) or smaller than 1/5. This result is not affected by statistical error (Suppl. Mat. Section I.). Additional evidence of the large discrepancies in the denatured state ensemble is provided by the CSNs of Beta3s and W10V colored according to the basins identified by the kinetic analysis (Fig. 2 top).

The mean folding time (τ_f) is the kinetic distance from the native state, defined as the time interval Δt between a snapshot and the most populated node, where Δt is averaged over all the snapshots in a given node. By coloring the nodes according to the value of τ_f the basins identified by the kinetic grouping emerge as network "regions" of homogeneous colors (compare top and bottom of Fig. 2). This observation validates the kinetic grouping analysis if one considers that relaxation times to the native node were not used for grouping.

Free energy basins

The changes in the denatured state ensemble due to the W10V mutation are one of the main results of the kinetic grouping (Table I). There is a striking difference between the detailed information provided by the kinetic grouping (and illustrated by the CSN) and the free energy profile along the number of native contacts (thick lines in Fig. 3). The latter is a projection which yields a very similar picture of the denatured state of Beta3s and W10V even if the native state is used as reference for both profiles (as in Fig. 3). The complexity of the denatured state of Beta3s and W10V is captured by kinetic grouping analysis. The enthalpic basins identified by ki-

Table I
Results of the Kinetic Grouping

		Beta3s				W10V				
Conformation	Name	Weight (%) node	basin	τ_f (ns) node	basin	Weight (%) node	basin	τ_f (ns) node	basin	Color
-EEEESSEEEEEESSEEEE-	native	5.6	36.4	—	—	8.8	40.6	—	—	Green
Larger weight in Beta3s										
-EEEESTTEEEEESSEEEE-	Ns-or	1.2	7.4	138	109	0.8	4.0	92	90	Red
---SSGGG---EESSEETT-	Ch-curl$_1$	0.1	3.6	98	90	0	0	—	—	White ovals
---SSGGG-EESSTTTTEE-	Ch-curl$_2$	0.1	2.4	285	257	0	0	—	—	White circles
-----SS--EEEESSEEEE-		0.03	2.2	53	75	0.04	0.9	72	85	Orange
-HHHHHHHHHHHS------	Helix$_{1\text{-}13}$	0.1	2.1	137	122	0.01	0.5	124	151	White squares
--EESSSEEEEESSEEEE-		0.1	1.9	87	84	0.04	0.6	148	134	Cyan
---SSGGG-EEESSSEEEE-		0.1	1.2	200	198	0	0	—	—	White rectangles
---SSSS--EEESTT-EEE-		0.06	0.9	316	263	0	0	—	—	White diamonds
Larger weight in W10V										
-EEEESSEEEEESSEEEE-	Cs-or	0.4	3.6	63	70	0.7	6.0	69	75	Olive
-EEEESSEEEE---TT--B-	Nh-curl	0.04	0.5	59	58	0.1	3.2	69	69	Blue
----STT---EEESSEEEE-		0.1	0.8	139	113	0.3	2.7	108	121	Violet
--BSS-SSSEEE-STTEEE-	Ch-curl$_3$	0	0	—	—	0.1	2.6	104	105	White diamonds
--SSSSS--EEEESSEEEE-		0.03	0.8	103	97	0.1	2.1	111	100	Yellow
-BSSSS---EEEESSEEEE-		0	0	—	—	0.02	0.3	61	53	White circles

Statistical weight of the native basin and the most populated free energy basins in the denatured state as identified by the kinetic grouping. Mean folding times (τ_f) are average values for snapshots in a node or basin. Conformations with names are shown by flexible tubes of variable diameter in Figure 1 for Beta3s and in the Suppl. Mat. (Fig. S1) for W10V: Ns-or, N-terminal strand out of register; Cs-or, C-terminal strand out of register; Nh-curl, curl-like conformation with structured N-terminal hairpin; Ch-curl, curl-like conformation with structured C-terminal hairpin. The colors indicated in the last column are those used in Figure 2 top.

DOI 10.1002/prot

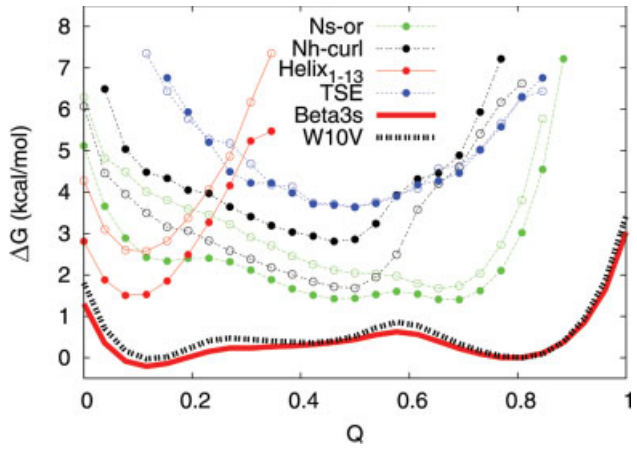

Figure 3

Inadequacy of free energy projection. The free energy profile along the fraction of native contacts Q (see[20] for the list of contacts) is plotted with thick lines without symbols. The contributions of individual free energy basins in the denatured state are shown with thin lines with solid and empty circles for Beta3s and W10V, respectively. The folded state basin is not shown to avoid overcrowding; it overlaps with the thick lines in the range $0.7 \leq Q \leq 1.0$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

netic grouping and the entropically favored helical region (see below and Ref. 11) can be plotted as a supplement to the free energy profile (thin lines with symbols in Fig. 3) to demonstrate the inadequacy of simple one-dimensional projections.

Notably, the basin with helical N-terminal segment and the β-sheet with N-terminal strand out of register (Ns-or) is more populated in Beta3s than W10V, whereas the opposite is observed for the curl-like basin with stable N-terminal hairpin (Nh-curl, see Table I for names of most populated free energy basins). Moreover, the kinetic grouping analysis shows that two of the three most populated free energy basins in the denatured state of W10V have a native-like N-terminal hairpin; they are the β-sheet with C-terminal strand out of register (Cs-or) and the Nh-curl conformation (Table I). Overall, the segment 2–11 has 19% higher content of native secondary structure and 20% more native tertiary contacts (Suppl. Mat. Fig. S5) in the denatured state of the W10V mutant than the wild type peptide.

Analysis of the interaction energy between residue pairs in Cs-or and Nh-curl conformations shows that the loss of favorable interactions of residue 10 with Lys17 and Tyr19 destabilizes Beta3s more than W10V mainly because of the difference in van der Waals interactions (Fig. 4). This loss is compensated, but only partially, in the Cs-or conformations, where the shift of the β-strand register results in non-native van der Waals interactions of residue 10 with Thr16 and Ile18, which are more favorable for the Trp10 side chain of the wild type Beta3s than the smaller Val10 of the mutant. These mainly hydrophobic contacts involving Trp10 provide further evidence for the importance of non-native interactions and are consistent with experimental observations. In fact, residual structure in the denatured state of an SH3 domain and non-native interactions of its Trp36 side chain have been observed by NMR spectroscopy under folding conditions.[28] Interestingly, the native structure of the SH3 domain consists mainly of a hairpin packed against a three-stranded antiparallel β-sheet, and Trp36 is in the central strand of the latter.

Conformations with partial helical content

It was shown in a previous study[11] that the denatured state ensemble of Beta3s is very heterogeneous and

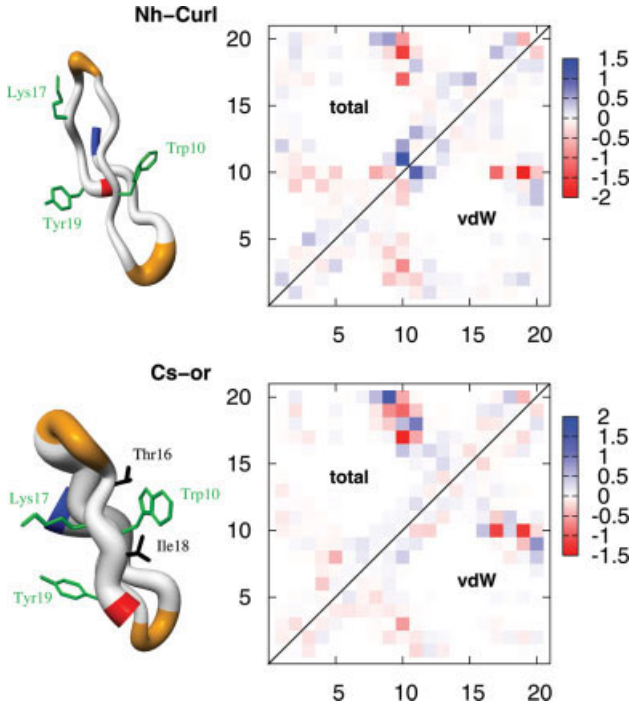

Figure 4

The non-native conformations Nh-curl (top) and Cs-or (bottom) are more populated in W10V than Beta3s. (Left) Flexible tube representation: side chains which interact favorably with residue 10 in the native state but are distant from it in the Nh-curl and Cs-or conformations (see text) are in green. Side chains involved in non-native contacts with Trp10 are in black. (Right) Difference in residue pairwise interaction energy between W10V and Beta3s averaged over the basin. The pairwise energy values in the native state are used as reference and subtracted from all values in the matrix. A red square indicates that the corresponding pair of residues has a more favorable energy in the Nh-curl or Cs-or conformation of W10V than Beta3s. The bars on the right give values in kcal/mol. The upper and lower triangular matrices show the total (sum of van der Waals and electrostatics) and van der Waals energy, respectively, and their similarity indicates that most of the enthalpic stabilization originates from differences in van der Waals energy. In fact, the sum of all pairwise contributions yields a W10V vs. Beta3s difference of −2.0 kcal/mol for the Nh-curl conformation, with a van der Waals contribution of −3.9 kcal/mol. The corresponding values for the Cs-or conformation are −0.2 kcal/mol and −2.2 kcal/mol.

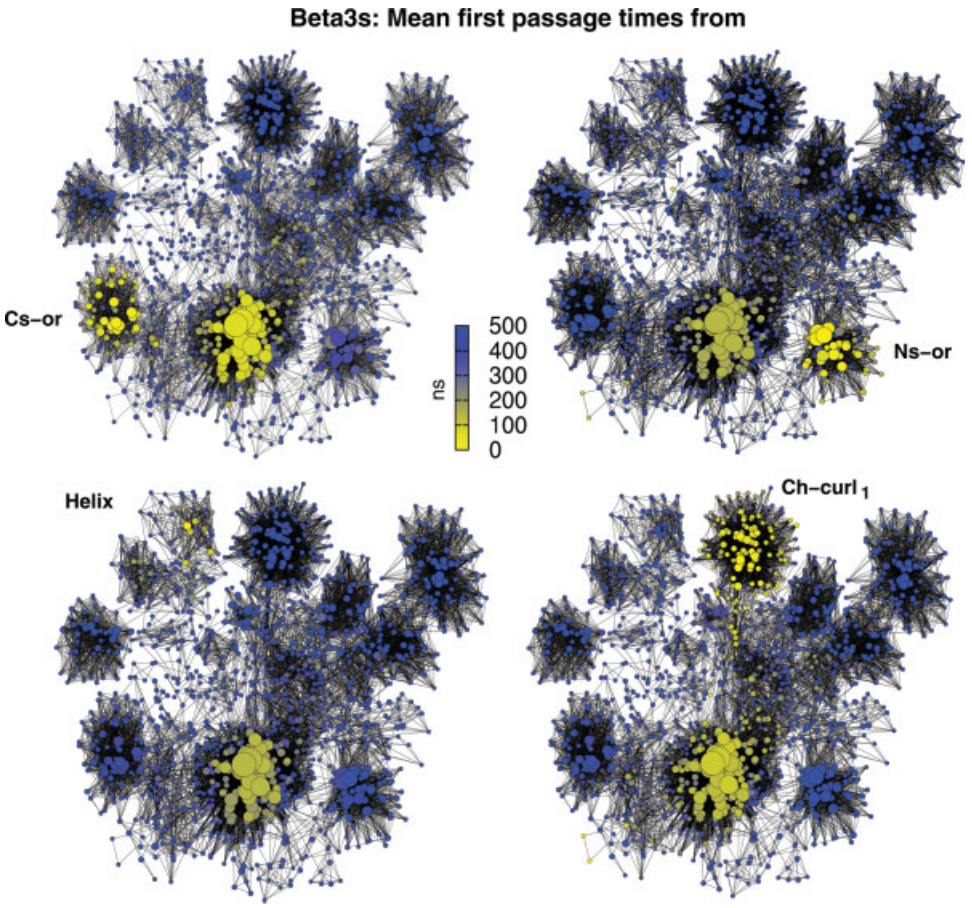

Figure 5

The denatured state is kinetically partitioned. Nodes are colored according to mean first passage times from the most populated node of individual free energy basins to all other nodes of the CSN of Beta3s. Nodes within the basin of the starting node (i.e., the most populated node of the considered basin) are visited relatively fast (yellow), indicating rapid intrabasin transitions and supporting the kinetic grouping analysis (Fig. 2 top). Equilibration between different unfolded basins (blue) is slower than reaching the folded state (olive) which shows that the denatured state is kinetically partitioned, i.e., no fast equilibration takes place between basins in the denatured state. In other words, the native state is a hub.[11] Kinetic partitioning is also observed for the denatured state of W10V (Suppl. Mat. Fig. S3). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

includes conformations with partial helical structure and fluctuating unstructured residues, for example, the C-terminal segment in the helical conformation shown on top left of Figure 1. Here, it is interesting to compare the helical propensity of the two peptides. The conformations of Beta3s and W10V with partial helical content (defined as more than three, four, or five consecutive 3_{10}-, α-, or π-helical residues, respectively) have a statistical weight of 11.6 and 10.7%, respectively. Note that the aforementioned basin with helical N-terminal segment ($Helix_{1–13}$, -HHHHHHHHHHHHS------) is only a fraction (of weight 2.1 and 0.5% for Beta3s and W10V, respectively, Table I) of the conformations with partial helical structure. The latter make up an entropic region in the denatured state.[11] The weight distribution of the nodes in the helical region shows a more pronounced decay than for the native state nodes (Suppl. Mat. Fig. S12), which is consistent with the entropic character of the helical region. In fact, it was demonstrated recently on low-dimensional models, which can be treated analytically, that free energy basins with entropic character have a faster decay of the weight distribution than mainly enthalpic basins.[29]

Interestingly, the percentage of helical content along the sequence shows that Beta3s is more helical than W10V in the central segment, that is, residues 7–13 (Suppl. Mat. Fig. S11). This observation is consistent with the experimental finding that the side chain of valine has a destabilizing effect on the helical structure[30] because of the branching at the C_β carbon.

Kinetic partitioning

The denatured state ensemble of Beta3s and W10V is not in rapid equilibrium. In fact, the mean first passage times (mfpt) from the most populated node of individual

DOI 10.1002/prot

basins in the denatured state to all other nodes, which can be calculated by simply following the trajectories (i.e., the time series of the nodes), reveal that the fastest transitions are those to the native basin. It is useful to color the CSN of Beta3s according to values of mpft from individual free-energy basins (Fig. 5). Folding from four of these basins take in average less than 150 ns (olive nodes in Fig. 5), while the mean transition times to other regions are larger than 500 ns (blue nodes in Fig. 5), indicating that trajectories between non-native basins mostly pass through the folded state. A similar picture is observed for W10V (Suppl. Mat. Fig. S3). The barriers between individual free energy basins in the denatured state and the native state are clearly lower than barriers within the unfolded state. The observation of kinetic partitioning of the denatured state by kinetic grouping analysis provides further evidence to the fact that the native basin of Beta3s and W10V acts as a hub.[11] Importantly, these simulation results are consistent with recent experimental reports on competing folding routes and kinetic partitioning in ubiquitin[31] and the DNA-binding domain of p53.[32] Most notably, the present results provide evidence that folding pathways depend crucially on the denatured state, that is, the starting ensemble. Therefore, *in vitro* folding experiments under strongly unfolding conditions (e.g., high concentration of chemical denaturants) do not necessarily give insights into the folding process under physiological conditions.

Transition state ensemble and switch of folding pathways

The folding probability P_{fold} of a conformation is the likelihood to reach the most populated node before unfolding.[33] It should be close to 0.5 for transition state ensemble (TSE) structures. A total of 55 nodes containing 1703 snapshots for Beta3s and 51 nodes with 1881 snapshots for W10V (diamonds in Fig. 1 and Fig. S1) have been isolated as TSE by the node-P_{fold} analysis.[34] The native secondary structure content in a node can be determined by comparison to the native string. Figure 6 shows the percentage of native secondary structure along the sequence for the TSE of Beta3s and W10V. Notably, there is a TSE switch from predominance of structured C-terminal hairpin in Beta3s to structured N-terminal hairpin in W10V, which is consistent with the redistribution of basin populations in the denatured state ensembles because of the kinetic partitioning, that is, higher barriers between non-native free-energy basins than folding barriers. Furthermore, each TSE string can be unambiguously classified into structured N-terminal (Nh-TSE) or C-terminal (Ch-TSE) hairpin by predominance of native bits in the 2–11 or 10–19 string segment, respectively. In the TSE of Beta3s, 70% of the structures belong to Ch-TSE and only 30% to Nh-TSE, whereas W10V

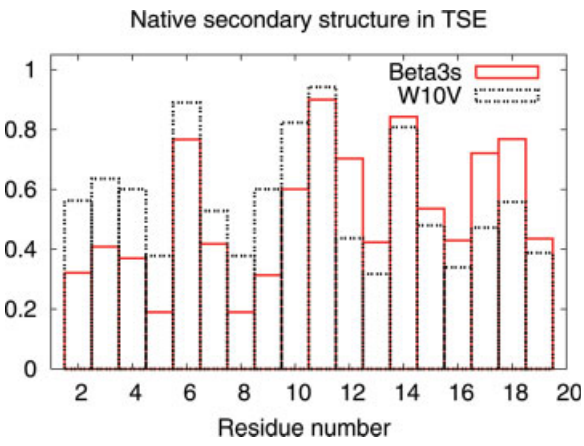

Figure 6

The content of native secondary structure, measured by DSSP,[35] at the TSE illustrates the shift of flux. The pathway through the TSE with formed N-terminal hairpin is more populated in W10V, whereas the pathway through the C-terminal hairpin is more populated in Beta3s. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

switches to 45% Ch-TSE and 55% Nh-TSE. These results are robust with respect to the choice of a τ_{commit} value in the interval from 1.5 to 5 ns (see Suppl. Mat. Fig. S8).

Interestingly, an energetic analysis explains the switch of flux through the two distinct TSEs. The effective energy contribution (sum of intrapeptide and solvation energy terms) to the Nh-TSE activation energy is 0.6 kcal/mol less favorable for W10V than Beta3s. This value is smaller than the effective energy contribution to the Ch-TSE which is 1.6 kcal/mol less favorable for W10V than Beta3s. Therefore, if one assumes similar conformational entropy contributions for both peptides, it follows that the W10V mutation raises more the Ch-TSE than Nh-TSE barrier with respect to the wild type barriers. Most of these energetic effects at the TSE are due to van der Waals interactions as shown in the Suppl. Mat. (Fig. S9).

The free energy projection along the fraction of native contacts shows again its limitations. In fact, using a "naive" criterion of about half of the native contacts formed it is much more likely to select Ns-or conformations than TSE structures (Fig. 3). The former belong to the denatured state ensemble and are clearly pre-critical as indicated by their node-P_{fold} values very close to zero (between 0.001 and 0.005).

CONCLUSIONS

A kinetic analysis based on coarse-graining of molecular dynamics snapshots and grouping according to fast relaxation at equilibrium has been presented for investigating the topography of the conformational space of structured peptides at the melting temperature. Applica-

tion to simulations of reversible folding of a three-stranded antiparallel β-sheet peptide of 20 residues reveals that the W10V mutation in the central strand alters the relative population of the non-native states mainly because of differences in van der Waals energy. In particular, changes are observed in the denatured state ensemble, which is found to be kinetically partitioned, as well as in the relative height of two distinct transition state barriers on parallel folding pathways. As a consequence, the minor folding route in the wild type (first formation of N-terminal hairpin) becomes the major route in the W10V mutant. In other words, the point mutation modulates the flux through each pathway. It is generally accepted that the final stage of folding requires specific side chain packing. The present study provides evidence that a single side chain difference can result in a redistribution of basin populations in a structured denatured state ensemble, that is, at the beginning of the folding process. Significant differences have been observed in the denatured state of a small protein under mildly and strongly denaturing conditions by a joint analysis of fluorescence, differential scanning calorimetry, circular dichroism, and NMR spectroscopy.[36] Therefore, the folding mechanism and pathways postulated on the basis of *in vitro* experiments with high concentrations of guanidinium hydrochloride or urea might not be representative of the folding process under physiological conditions. Further experimental and computational investigations on denatured states under physiological conditions are required because the "starting ensemble" for folding plays a key role and is not just a randomly unfolded polypeptide chain.

Recent network and graph analyses have shown that the surprisingly simple two-state picture of protein folding, often obtained by projecting the free energy onto an arbitrarily chosen progress variable, is not consistent with the complexity of the actual free energy surface.[10–12,37] The simulation results and kinetic analysis provide not only a detailed description of the heterogeneity of the denatured state ensemble, but capture also the changes originating from a single-point mutation, which are hidden in simple projections of the free energy. The kinetic grouping approach presented here is quite general and can be used for investigating the free energy topography of other complex (molecular) systems provided that their conformation space can be sampled by equilibrium simulations.

METHODS

Molecular dynamics simulations

All simulations and part of the analysis of the trajectories were performed with the program CHARMM.[38] The designed 20-residue peptide Beta3s[19] and its W10V mutant were modeled by explicitly considering all heavy atoms and the hydrogen atoms bound to nitrogen or oxygen atoms (PARAM19 force field[38] with the default cutoff of 7.5 Å for the nonbonding interactions). A mean field approximation based on the solvent accessible surface (SAS) was used to describe the main effects of the aqueous solvent on the solute.[23] This choice is justified by two reasons. First, using explicit water simulations it is not possible to sample a statistically significant number of equilibrium folding–unfolding transitions which is a *conditio sine qua non* for the present analysis. Second, it was shown previously using exactly the same SAS-based implicit solvation model that at 330 K Beta3s folds reversibly to its NMR conformation irrespective of the starting structure, and importantly, 23 of the 26 NOE restraints are satisfied.[20] Despite the absence of collisions with water molecules, in the simulations with implicit solvent relative rates are comparable with the values observed experimentally. Helices fold in about 1 ns,[39] β-hairpins in about 10 ns,[39] and triple-stranded β-sheets in about 0.1 μs,[22] while the experimental values are ∼0.1 μs,[40] ∼1 μs[40] and ∼10 μs,[19] respectively. Importantly, the small variations in total solvent accessible surface and radius of gyration during folding of Beta3s at 330 K[21] provide evidence that the lack of solute/solvent friction does not have a significant effect on the main results of the kinetic analysis.

For each peptide 10 molecular dynamics runs of 2 μs each with different initial distribution of velocities were performed with the Berendsen thermostat (coupling constant of 5 ps) at 330K, which is close to the melting temperature of wild-type Beta3s.[21] A time step of 2 fs was used and the coordinates were saved every 20 ps for a total of 10^6 snapshots for each system. The significant amount of sampling is supported by the small variation in the native population measured on two disjoint 10-μs subsets of the trajectories that have a weight of 38 and 35% for Beta3s (38 and 43% for two subsets of the W10V trajectories). Moreover, a small variation is observed for the population of individual free energy basins in the denatured state. As an example, the Cs-or basin has a weight of 3.9 and 3.2% for the two 10-μs subsets of Beta3s (6.2 and 5.9% for W10V), and the population of the Nh-curl basin is 0.35 and 0.65% for the two 10-μs subsets of Beta3s (3.4 and 3.1% for W10V).

Coarse-graining

There are several ways for assigning snapshots (i.e., coordinate sets) to coarse-grained conformations (nodes and strings are used as synonyms in this article) and different types of analysis might require different coarse-graining approaches.[11,12] The coarse-graining used in this work is based on secondary structure strings.[35] A conformation is a single string of secondary structure, for example, the most populated conformation of Beta3s is -EEEESSEEEEEESSEEEE-.[11] There are 8 possible

DOI 10.1002/prot

"letters" in the secondary structure "alphabet": "H", "G", "I", "E", "B", "T", "S", and "−", standing for α helix, 3_{10} helix, π helix, extended, isolated β-bridge, hydrogen bonded turn, bend, and unstructured, respectively. Since the N- and C-terminal residues are always assigned an "−",[35] a 20-residue peptide can in principle assume $8^{18} \simeq 10^{16}$ conformations. The secondary structure-based coarse-graining is appropriate for structured peptides without loops and has three advantages with respect to approaches based on root mean square deviation (RMSD) of atomic coordinates. First, it does not require the use of an arbitrarily chosen threshold value. Second, each node is uniquely defined by its secondary structure string which is a useful conformational "label". Third, the same type of secondary structure classification[35] is used for coarse-graining and analysis of structural properties (e.g., native content in TSE, Fig. 6).

Conformation space network

Conformations (i.e., secondary structure strings) are nodes of the conformation space network (CSN) and the direct transitions between them are links.[11] The number of snapshots with a given secondary structure string is abbreviated as $\tilde{w}$. The statistical weight w of a node is equal to $w = \tilde{w}/N$, where $N = 10^6$ is the total number of snapshots for each of the two peptides. The CSN in Figure 1 and in the Suppl. Mat. (Fig. S1) show only nodes with $\tilde{w} \geq 40$ to avoid overcrowding. It is important to emphasize that the CSN of "heavy" nodes is used solely for illustrative purposes, whereas the quantitative analysis was performed using the kinetic grouping which takes into account the complete trajectory, that is, all nodes (see later).

Kinetic grouping into free energy basins

An important methodological aspect of the present work is that conformations are grouped into disjoint free energy basins not according to arbitrarily chosen geometric characteristics, but rather according to the statistics of the transitions at equilibrium as explained in the two following subsections.

The native state basin

The log-binned distribution of first passage times from any snapshot saved along the trajectory to the native node -EEEESSEEEEEESSEEEE- separates the conformational space into two regions: the fast relaxation within the native basin takes place in less than 1.6 ns, while folding from outside is about two orders of magnitudes slower (Fig. 7). Nodes are assigned to the native basin if they have a node-$P_{fold} \geq 0.5$ (see Ref. 34) using a commitment time τ_{commit} of 1.6 ns, meaning that at least 50% of the structures in a node must visit the native node within 1.6 ns along the trajectory. Note that this is a purely kinetic argument for grouping nodes to a basin. The only condition is that the underlying coarse-graining yields nodes containing kinetically homogeneous structures.

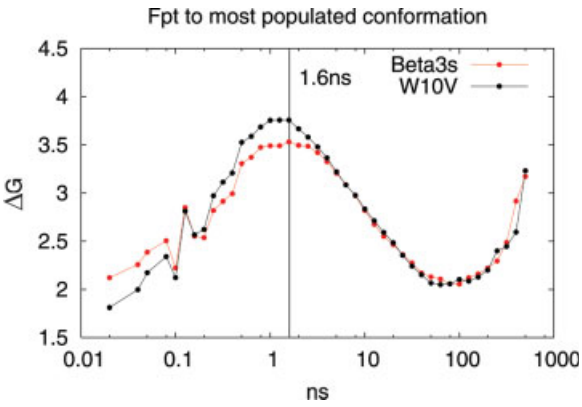

Figure 7

Distribution of first passage times P(fpt) to the most populated node. Values are calculated as $\Delta G = -k_B T \ ln(P(fpt))$ and given in kcal/mol. The fpt can be considered as a geometrically unbiased reaction coordinate. This projection is very useful to determine the transition between intra-basin and inter-basin relaxation, which is emphasized by the logarithmic binning (10 bins/decade) without normalization of the binsize. The peak at about 0.1 ns is an artifact of the binning but does not have any effect on the results. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Since the accuracy of node-P_{fold} depends on the population of the node, a statistical argument to reduce the number of false negatives is used. If a node has $P_{fold} \geq 0.5$ in the limit of infinite sampling, it should be detected with a probability of at least 80% (called "*80% criterion*" hereafter). As an example, a node with 20 snapshots is grouped to the native basin if at least 8 snapshots are separated from the native node by less than 1.6 ns (node-$P_{fold} \geq 0.4$), while a node with 200 snapshots must have at least 92 of them folding within 1.6 ns (node-$P_{fold} \geq 0.46$).

Basins in the denatured state

The kinetic grouping can be extended to automatically detect the attractor regions which are stabilized mainly enthalpically. Instead of calculating node-P_{fold}, which per definition follows the trajectory to the native state, an analogous quantity can be determined between any two nodes. Hence, $P_{commit}(i \to j)$ is defined as the probability to observe a transition from node i to node j within a given commitment time τ_{commit}. The value of $P_{commit}(i \to j)$ is an asymmetric, directed measure of the kinetic similarity of nodes i and j. Once the P_{commit}-matrix has been calculated for nodes with at least 40 snapshots, pairs of nodes (i,j) are grouped together if $P_{commit}(i \to j) \geq 0.5$, where the statistical 80%-criterion is applied. This way of grouping leads to a partitioning into disjoint sets, that is, a disconnected CSN, whose subgraphs correspond to different attractor regions. A value of $\tau_{commit} = 1$ ns is used

because it is in the range of the relaxation times for most of the important basins, but still two to three orders of magnitudes shorter than typical transition times from outside (see Suppl. Mat. Fig. S6). Note that the kinetic grouping is robust with respect to the choice of the commitment time in the interval $0.5 \leq \tau_{commit} \leq 2.0$ ns (Suppl. Mat. Fig. S7). Moreover, different values of τ_{commit} allow one to analyze different levels of ruggedness of the free energy surface as illustrated by the application to the alanine dipeptide (Suppl. Mat.). Since the all-against-all P_{commit} calculations are expensive (growth is quadratic) and to reduce noise caused by nodes lying in high-energy regions, only nodes with at least 40 snapshots were used in a first step (which required about 10 h of CPU time of a Xeon 2 GHz), while the small nodes were grouped to the attractor regions in a post-processing step as detailed in the Suppl. Mat.

Transition state ensemble

Putative transition state ensemble (TSE) nodes are identified by the node-P_{fold} analysis[34] using the aforementioned 80%-criterion and with a τ_{commit} of 1.6 ns to be consistent with the native state basin isolation. Only nodes with weight above a threshold ($\tilde{w} \geq 20$) are considered to guarantee a minimal amount of statistics. Validation with classical P_{fold} analysis[33] is presented in the Suppl. Mat. (Table S-III).

ACKNOWLEDGMENTS

We thank Enrico Guarnera, Francesco Rao, and Gianni Settanni for very interesting and helpful discussions. We also thank Paolo De Los Rios, David Gfeller, and Ben Schuler for a critical reading of the manuscript. We are indebted to Michele Seeber for the program Wordom[41] used to analyze the trajectories. The simulations were performed on the Matterhorn cluster of the University of Zurich and we gratefully acknowledge the support of C. Bolliger and A. Godknecht.

REFERENCES

1. Frauenfelder H, Sligar SG, Wolynes PG. The energy landscape and motions of proteins. Science 1991;254:1598–1603.
2. Dill K, Chan H. From levinthal to pathways to funnels. Nat Struct Biol 1997;4:10–19.
3. Leopold PE, Montal M, Onuchic JN. Protein folding funnels: a kinetic approach to the sequence-structure relationship. Proc Natl Acad Sci USA 1992;89:8721–8725.
4. Dinner AR, Šali A, Smith LJ, Dobson CM, Karplus M. Understanding protein folding via free-energy surfaces from theory and experiment. Trends Biochem Sci 2000;25:331–339.
5. Daggett V, Fersht AR. The present view of the mechanism of protein folding. Nat Rev Mol Cell Biol 2003;4:497–502.
6. Wolynes PG. Energy landscapes and solved protein-folding problems. Phil Trans R Soc A 2005;363:453–467.
7. Thirumalai D, Hyeon C. RNA and protein folding: common themes and variations. Biochemistry 2005;44:4957–4970.
8. Shakhnovich EI. Protein folding thermodynamics and dynamics: where physics, chemistry, and biology meet. Chem Rev 2006;106:1559–1588.
9. Dill KA, Shortle D. Denatured states of proteins. Annu Rev Biochem 1991;60:795–825.
10. Krivov SV, Karplus M. Hidden complexity of free energy surfaces for peptide (protein) folding. Proc Natl Acad Sci USA 2004;101:14766–14770.
11. Rao F, Caflisch A. The protein folding network. J Mol Biol 2004;342:299–306.
12. Hubner IA, Deeds EJ, Shakhnovich EI. Understanding ensemble protein folding at atomic detail. Proc Natl Acad Sci USA 2006;103:17747–17752.
13. Boczko EM, Brooks CL III. First-principles calculation of the folding free energy of a three-helix bundle protein. Science 1995;269:393–396.
14. Ferguson N, Sharpe TD, Schartau PJ, Sato S, Allen MD, Johnson CM, Rutherford TJ, Fersht AR. Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family. J Mol Biol 2005;353:427–446.
15. Zagrovic B, Snow CD, Khaliq S, Shirts MR, Pande VS. Native-like mean structure in the unfolded ensemble of small proteins. J Mol Biol 2002;323:153–164.
16. Shortle D, Ackerman MS. Persistence of native-like topology in a denatured protein in 8 M urea. Science 2001;293:487–489.
17. Dobson CM. Protein folding and misfolding. Nature 2003;426:884–890.
18. Platt G, McParland V, Kalverda AP, Homans SW, Radford SE. Dynamics in the unfolded state of β_2-microglobulin studied by NMR. J Mol Biol 2005;346:279–294.
19. De Alba E, Santoro J, Rico M, Jiménez MA. De novo design of a monomeric three-stranded antiparallel β-sheet. Protein Sci 1999;8:854–865.
20. Ferrara P, Caflisch A. Folding simulations of a three-stranded antiparallel β-sheet peptide. Proc Natl Acad Sci USA 2000;97:10780–10785.
21. Cavalli A, Ferrara P, Caflisch A. Weak temperature dependence of the free energy surface and folding pathways of structured peptides. Proteins: Struct Funct Bioinformatics 2002;47:305–314.
22. Settanni G, Rao F, Caflisch A. Φ-Value analysis by molecular dynamics simulations of reversible folding. Proc Natl Acad Sci USA 2005;102:628–633.
23. Ferrara P, Apostolakis J, Caflisch A. Evaluation of a fast implicit solvent model for molecular dynamics simulations. Proteins: Struct Funct Bioinformatics 2002;46:24–33.
24. Nauli S, Kuhlman B, Baker D. Computer-based redesign of a protein folding pathway. Nat Struct Biol 2001;8:602–605.
25. Wright CF, Lindorff-Larsen K, Randles LG, Clarke J. Parallel protein-unfolding pathways revealed and mapped. Nat Struct Biol 2003;10:658–662.
26. Fersht AR, Matouschek A, Serrano L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol 1992;224:771–782.
27. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph Modell 1996;14(1):51–55.
28. Crowhurst KA, Tollinger M, Forman-Kay JD. Cooperative interactions and a non-native buried Trp in the unfolded state of an SH3 domain. J Mol Biol 2002;322:163–178.
29. Gfeller D, De Los Rios P, Caflisch A, Rao F. Complex network analysis of free-energy landscapes. Proc Natl Acad Sci USA 2007;104:1817–1822.
30. Hecht MH, Zweifel BO, Scheraga HA. Helix-coil stability constants for the naturally occurring amino acids in water: XVII threonine parameters from poly(hydroxylbutyl-glutamine-co-L-threonine). Macromolecules 1978;11:545–551.

 DOI 10.1002/prot

31. Crespo MD, Simpson ER, Searle MS. Population of on-pathway intermediates in the folding of ubiquitin. J Mol Biol 2006;360: 1053–1066.
32. Butler JS, Loh SN. Kinetic partitioning during folding of the p53 DNA binding domain. J Mol Biol 2005;350:906–918.
33. Du R, Pande VS, Grosberg AY, Tanaka T, Shakhnovich EI. On the transition coordinate for protein folding. J Chem Phys 1998;108:334–350.
34. Rao F, Settanni G, Guarnera E, Caflisch A. Estimation of protein folding probability from equilibrium simulations. J Chem Phys 2005;122:184901.
35. Andersen CAF, Palmer AG, Brunak S, Rost B. Continuum secondary structure captures protein flexibility. Structure 2002;10:174–184.
36. Mayor U, Grossmann JG, Foster NW, Freund SMV, Fersht AR. The denatured state of engrailed homeodomain under denaturing and native conditions. J Mol Biol 2003;333:977–991.
37. Caflisch A. Network and graph analyses of folding free energy surfaces. Curr Opin Struct Biol 2006;16:71–78.
38. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983; 4:187–217.
39. Ferrara P, Apostolakis J, Caflisch A. Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. J Phys Chem B 2000;104:5000–5010.
40. Eaton WA, Munoz V, Hagen J, Jas SGS, Lapidus LJ, Henry ER, Hofrichter J. Fast kinetics and mechanisms in protein folding. Ann Rev Biophys Biomolec Struc 2000;29:327–359.
41. Seeber M, Cecchini M, Rao F, Settanni G, Caflisch A. Wordom: a program for efficient analysis of molecular dynamics simulations. Bioinformatics, in press.