PROTEINS: Structure, Function, and Bioinformatics 55:493–501 (2004)

SHORT COMMUNICATION

Unfolded State Of Polyalanine Is a Segmented Polyproline II Helix

Alex Kentsis, Mihaly Mezei, Tatyana Gindin, and Roman Osman*
Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York University, New York, New York

ABSTRACT **Definition of the unfolded state of proteins is essential for understanding their stability and folding on biological timescales. Here, we find that under near physiological conditions the configurational ensemble of the unfolded state of the simplest protein structure, polyalanine α-helix, cannot be described by the commonly used Flory random coil model, in which configurational probabilities are derived from conformational preferences of individual residues. We utilize novel effectively ergodic sampling algorithms in the presence of explicit aqueous solvation, and observe water-mediated formation of polyproline II helical (P_{II}) structure in the natively unfolded state of polyalanine, and its facilitation of α-helix formation in longer peptides. The segmented P_{II} helical coil preorganizes the unfolded state ensemble for folding pathway entry by reducing the conformational space available to the diffusive search. Thus, as much as half of the folding search in polyalanine is accomplished by preorganization of the unfolded state. Proteins 2004;55:493–501.** © 2004 Wiley-Liss, Inc.

Key words: protein folding; hierarchical organization; search nucleation; unfolded state preorganization

INTRODUCTION

The importance of accurately defining the unfolded state of proteins was recognized early by Levinthal, who concluded that folding of a random coil by way of a diffusive search of its combinatorially vast conformational space is incompatible with biological energies and timescales of protein folding.[1] Consequently, either the conformational space of unfolded polypeptides deviates significantly from that of a random coil, or the conformational search is not entirely diffusive, being guided along folding pathway(s). The latter is a foundation of the framework folding models,[2] including the experimentally supported search-nucleation mechanism,[3] while the former is an implication of the hierarchical organization of proteins that are composed of nested structural motifs,[4] the most basic of which, such as helices and hairpin turns, can form on timescales several orders of magnitudes faster than those of protein folding.[5]

In support of the nonrandom nature of the unfolded state, original studies by Tanford and colleagues suggested that unfolded polypeptides can possess residual structure under nondenaturing conditions,[6] and using CD spectroscopy Tiffany and Krimm[7] observed P_{II} conformations in natively unfolded polyglutamyl peptides and hypothesized their origin in P_{II} helices 4–7 residues in length. Recently, advanced spectroscopic studies indicated that a substantial fraction of residues in unfolded polyalanine peptides exists in P_{II} conformation.[8–11] Polypeptides in canonical P_{II} configuration are characterized by backbone dihedral angles of $(\phi, \psi) = (-75°, +145°)$, which result in 3_{10}-helices that do not contain intramolecular hydrogen bonds and are more extended and solvated as compared to α-helices. It is unknown whether the macroscopic experimentally observed P_{II} structural features correspond to conformational preferences of individual residues, or to an average of microscopic configurational states that are composed of contiguous stretches of multiple residues in P_{II} conformation. The former is consistent with Flory's isolated pair hypothesis and the random, albeit conformationally biased (statistical), coil model of the unfolded state.[12] The latter cannot be obtained from Flory's model and implies that the unfolded state configurational space of proteins is preorganized in the form of P_{II} structural segments. A quantitative demonstration of such preorganization may provide insight into its contribution to the folding search, and is thus of major importance for understanding, predicting, and designing protein structures.

Numerous simulations of short peptides have been carried out in the past, including those of polyalanine in

The Supplementary Materials Referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/index.html

Grant sponsor: National Institutes of Health Medical Scientist Training Program (A. Kentsis and T. Gindin). Grant sponsor: National Institutes of Health; Grant numbers: DK 43036, CA 63317, and CA 80728.

*Correspondence to: Roman Osman, Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York University, New York, NY 10029. E-mail: roman.osman@mssm.edu

Received 23 August 2003; Accepted 21 October 2003

Published online 1 April 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20051

aqueous environment,[13–17] and a few have focused on the role of unfolded state structure.[18,19] Recent work[20] explored a blocked A_{21} peptide in explicit water using replica exchange molecular dynamics (MD). The authors observed a substantial fraction of α-helix, but the rest of the conformations, labeled as coil, remained undifferentiated. Recently, Mu et al.[8] and Zaman et al.[21] examined conformational preferences of alanine tripeptides and noted that the P_{II} conformation has a major population in the unfolded state.

In this article, we present results of statistically converged and apparently accurate calculations of the unfolded state ensemble of 7 (A_7) and 14 (A_{14}) residue polyalanine peptides flanked by pairs of diaminobutyrate and ornithine residues (see Methods section), recently examined by Kallenbach and colleagues using NMR and circular dichroism (CD) spectroscopies.[11,22] These peptides are of particular interest because A_7 is natively unfolded, allowing direct characterization of the unfolded state, without the use of chemical or thermal denaturation, and A_{14} is about 30% helical, enabling us to evaluate the contribution of unfolded state structure to α-helical (protein) stability. Moreover, the physics of polyalanine reflects inherent and universal properties of the polypeptide backbone, which are preserved in natural proteins and modified in a sequence specific manner.

METHODS

Molecular Systems

Simulations were performed with all-atom CHARMM27[23] and AMBER98 (http://amber.scripps.edu) force fields and the TIP3P water model.[24] CHARMM27 and AMBER98 force fields were produced by setting force constants for ϕ and ψ torsion potentials to zero. CHARMM27a98 and AMBER98c27 were produced by replacing the ϕ and ψ torsion potentials in CHARMM27 with those of AMBER98, and those of AMBER98 with those of CHARMM27, respectively.

Calculations were performed on the blocked polyalanine peptides Ac-$X_2A_7O_2$-NH_2 (A_7) and Ac-$X_2A_{14}O_2$-NH_2 (A_{14}), where X and O represent diaminobutyrate and ornithine, respectively, as studied experimentally by Kallenbach and colleagues.[11,22] Force field parameters for both diaminobutyrate and ornithine were derived from lysine, since their side-chains contain two and one fewer methylene groups, respectively. Systems with periodic boundary conditions (PBC) were constructed by randomly placing waters into a face-centered cubic cell until appropriate density was reached, using partial specific molecular volume of water of 30 $Å^3$, with the total number of water molecules adjusted based on the partial molar volume of component amino acids of the solute.[25] Initial solute configurations were generated manually in vacuum, solvated, energy minimized, and equilibrated in PBC. Primary hydration shells (PHS) were constructed from the PBC systems by deleting water molecules outside the PHS radius as measured from the nearest solute heavy atom. The PHS systems of A_7 and A_{14} contained 414 and 554 water molecules, respectively, corresponding to hydration shells of 6–8 Å and 2–3 molecular layers in thickness on average.

Monte Carlo Simulations

Our implementation of the PHS for Monte Carlo (MC) simulations replaces bulk molecular solvation with a restraining potential at the surface of a molecular solvation layer that maintains proper water structure and energetics by mimicking steric exclusion by the outlying bulk water, and by accurately representing interfacial water structure for waters that are outside of the PHS but within the restraining potential.[26] Usage of a size-independent reference shell energy allows the primary hydration shell to adopt arbitrary shape and size as dictated by conformational sampling of the solute, while capturing principal solvation effects such as density, solvation energy, and molecular water structure.[26] MC-PHS appears to reproduce solvation structure and energetics of both highly polar and nonpolar amino acid solutes, as well as to correctly recapitulate the experimentally observed aqueous stability of a β-hairpin peptide, making it suitable for studies of structure and thermodynamics of flexible polypeptides in the context of explicit aqueous solvation.[26] Current work characterizes its performance using larger solutes and extends its capabilities as described.

MC simulations were performed using a force-biased Metropolis procedure as implemented in the program MMC (http://inka.mssm.edu/~mezei/mmc). Systems were thermalized for 10,000 sweeps, as judged from energy equilibration, and evolved for 50 million sweeps, saving configurations every 100th sweep, where one sweep represents one step of all the degrees of freedom, including those of the solute and the solvent. Bond lengths and angles were kept constant. We utilized the shuffled cyclic procedure for single-solute torsional moves,[27] and the reverse proximity criterion for local torsional moves of several torsional angles,[28] which were performed half of the time. Appendix 1 in the Supplementary Material describes the iterative algorithm used to obtain the proximal solution to the loop-closing problem for a peptide backbone. Although not necessary for convergence of simulations of pentapeptides,[26] usage of local torsional moves was required for efficient sampling of the larger solutes studied here (Supplementary Tables I–III). Both solute and solvent step sizes were tuned to yield mean acceptance rates of 20–40%. The radius of the PHS was updated every third sweep, using the normalized reference shell energy of 0.15 kcal/mol/molecule and the restraining force constant of 3.0 kcal/mol/$Å^2$,[29] as described previously.[26] All nonbonded interactions were included.

Convergence and Accuracy

To evaluate sampling efficiency, we calculate the evolution of the apparent self-diffusion coefficient, $\Omega_{a-b}(x) = [\langle f_a(x)\rangle - \langle f_b(x)\rangle]^2$, as a function of simulation length x, where f_a and f_b are phase space variables, such as protein backbone dihedral angles, of two independent simulations starting from different initial conditions a and b. If the sampling of phase space is ergodic, $\Omega_{a-b}(x)/\Omega_{a-b}(0)$

decays to zero at long x. This is a necessary but insufficient condition of ergodicity, since it depends on the choice of initial conditions a and b. To evaluate convergence and ergodicity directly, we carried out 4 independent simulations using different random number seeds for the initial MC moves and their sizes, and 4 different initial solute configurations, chosen to span the configurations of interest: P_{II} (ϕ, ψ) = ($-75°$, $+145°$), P_I (ϕ, ψ) = ($+145°$, $-75°$), α-helix (ϕ, ψ) = ($-57°$, $-47°$), and β-strand (ϕ, ψ) = ($-139°$, $+135°$).

Analysis

Although canonical secondary structure elements have well-defined regular geometries, conformations and configurations in solution at ambient temperature exhibit considerable plasticity, which in computational studies are additionally dependent on force field parameters. Moreover, torsional geometries and their associated basins may depend on the aggregate nature of polypeptide interactions.

We use the term *conformation* to refer to backbone geometries of individual peptide residues, and *configuration* to refer to molecular geometries of groups of residues. Consequently, we utilize a self-consistent method for defining conformational basins using a stepwise optimal clustering algorithm based on a self-organizing neural net, as implemented in ART-2 by Brooks and coworkers.[15] Briefly, the cluster assignment of the dihedral angles extracted from simulation trajectories are optimized subject to a constraint on the cluster radius, such that no member of a cluster is farther than a specified distance from the cluster center. Because the convergence of such minimizations is sensitive to initial conditions, we test the robustness of assignments to conformational basins by recalculating the cluster assignments using reshuffled simulation trajectories (data not shown). Probabilities of sampling of the configurational basins based on conformational assignments as defined in this manner were calculated using a set of home-built programs, available upon request. Additionally, robustness of self-consistent clustering was evaluated by assigning secondary structure elements using the definitions implemented in PROSS (http://roselab.jhu.edu/utils/pross.html). In this manner, we use P_{II} to refer to backbone geometries in the P_{II} conformational basin, α to refer to helical geometries, including those of α-helical residues, and S to refer to residues having strand geometries, including residues in β-strands [Fig. 1(a)].

Scalar $^3J_{NH\alpha}$ coupling constants were calculated using the Karplus relation, as calibrated by Vuister and Bax.[30] Interaction energies, radial and orientational distribution functions, and coordination numbers were calculated according to standard methods,[31] referenced to the center of mass of the solute, as implemented in MMC using proximity analysis.[32] Sensitivity analysis of the decomposed interaction energies was done according to the method of Zhang et al.[33]

RESULTS AND DISCUSSION

In order to ensure computational accuracy, we have employed two different solute potential energy functions, an effectively ergodic MC algorithm to sample configurational space, and an explicit MC-PHS water to represent molecular aqueous solvation. MC-PHS maintains principal solvation effects such as density, solvation energy, and fine water structure,[26] while efficiently coupling structural rearrangements of the solvent to those of the solute. As a result, simulation lengths of 5×10^7 sweeps exceed the apparent time for self-diffusion of the backbone dihedral angles by more than 3 orders of magnitude (Supplementary Figs. 1 and 2, Supplementary Tables I–III), a degree of sampling currently unattainable by MD and MC simulations of canonical ensembles in condensed phase. The simulations appear to be converged, as judged from the results obtained from 4 independent simulations using initial configurations maximally distributed in torsional phase space (Fig. 1).

Unfolded state ensembles calculated using CHARMM27 and AMBER98 force fields exhibit several differences (Supplementary Fig. 3), most notably in the geometry of calculated P_{II} conformations [Fig. 1(a and b)], but are statistically indistinguishable in the mean probabilities of individual alanine residues sampling the respective P_{II} conformational basins, in spite of significant differences in their parameterisations of the polypeptide backbone [Fig. 1(c and d)]. More importantly, mean P_{II} conformational probabilities of 0.42 ± 0.089 and 0.47 ± 0.076 calculated using CHARMM27 and AMBER98, respectively, are in excellent agreement with the experimentally observed value of 0.4 ± 0.08.[11] Additionally, calculated $^3J_{HN\alpha}$ scalar coupling constants (Supplementary Table IV) are in agreement with those observed experimentally.[11]

To characterize the *configurational* preferences of the unfolded state ensemble, we enumerate the occurrences of varying lengths of contiguous stretches of alanines in P_{II} conformation, and compare the calculated probability distributions with those expected for a Flory polymer with identical but independent conformational preferences [Fig. 2(a and b)]. Probabilities of sampling P_{II} configurations are significantly greater for both CHARMM27 and AMBER98 simulations as compared to the rapidly decaying power law distributions of their respective Flory polymers [Fig. 2(a and b)], indicating that the unfolded state ensemble of polyalanine deviates significantly from that of a statistical coil. Specifically, P_{II} segments 4–6 residues in length, corresponding to about 1–2 helical turns, appear to be roughly isoenergetic with stretches of residues in the strand conformation, as indicated by their sampling probability of 0.4–0.5 [Fig. 2(a)]. Biphasic nature of the apparent configurational P_{II} probabilities relative to those of a Flory-like statistical coil with a breakpoint at 4–6 residues is consistent with such a segmental structure [Fig. 2(b)]. Snapshots of the calculated ensemble are shown in Figure 2(c). Consequently, the unfolded state ensemble of polyalanine is a segmented P_{II} helical coil, whose P_{II} segments are roughly isoenergetic with intervening strand configurations [Fig. 2(a)].

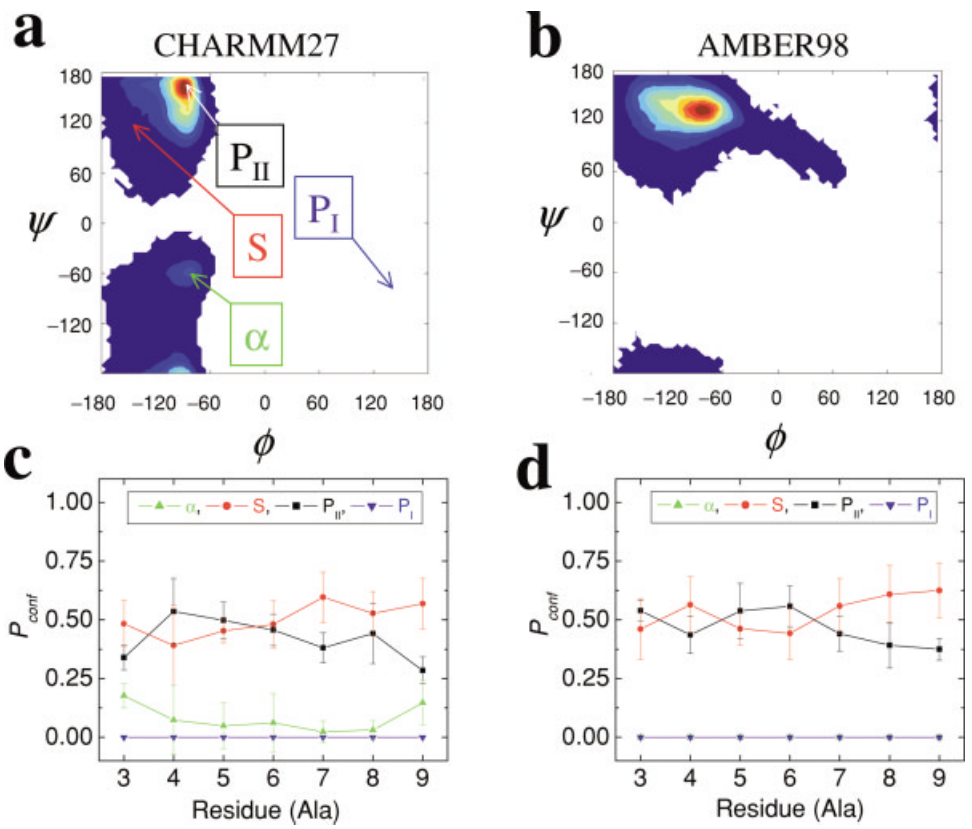
a
CHARMM27
b
AMBER98
180
120
60
0
−60
−120
ψ
P_{II}
P_I
S
α
−180 −120 −60 0 60 120 180
ϕ
c
d
1.00
0.75
0.50
0.25
0.00
P_{conf}
α, S, P_{II}, P_I
3 4 5 6 7 8 9
Residue (Ala)

Fig. 1.

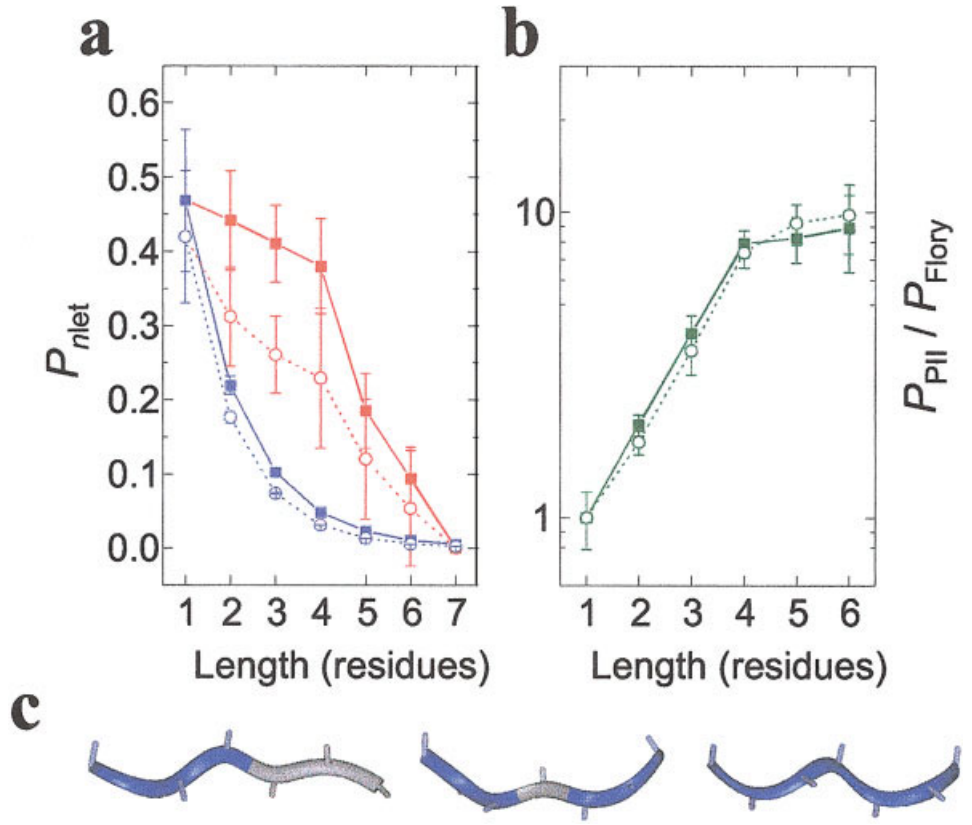
a
b
c
0.6
0.5
0.4
0.3
0.2
0.1
0.0
P_{nlet}
1 2 3 4 5 6 7
Length (residues)
10
1
P_{PII} / P_{Flory}
1 2 3 4 5 6
Length (residues)

Fig. 2.

The corresponding difference in conformational entropy between the observed P_{II} segmentally helical and expected Flory statistical coils is about 1 cal/mol/K per residue and 3–4 cal/mol/K per P_{II} helical turn (Table I). The apparent linear additivity of conformational entropies of P_{II} helix formation in the unfolded state of A_7 polyalanine indicates that P_{II} formation is largely noncooperative (Table I), consistent with predominantly diffusive nature of unfolded state configurational sampling. Considering that the calorimetrically determined loss of conformational entropy for α-helix formation in polyalanine is about 2 cal/mol/K/residue,[11,34] formation of P_{II} helical structures in unfolded polypeptides significantly reduces the contribution of conformational entropy to protein stability, conventionally thought to be the major opposing force for folding, as dictated by random and statistical coil models of the unfolded state.

To examine the contribution of solvating water to the configurational preferences of polyalanine, the accompanying article in this issue by Mezei et al.[35] presents results of thermodynamic integration calculations of conformational free energy of polyalanine. Remarkably, they observe that regular P_{II} helices produce a less disruptive effect on surrounding water organization as compared to β-strands, the solvation of the latter requiring the formation of entropically unfavorable peptide:water:peptide bridges and general reorganization of the surrounding bulk, much akin to a hydrophobic moiety.[35] We take a complementary approach and decompose the interaction energies of the conformationally segregated A_7 unfolded state ensembles using proximity analysis (Supplementary Fig. 4). In agreement with Mezei et al., our analysis suggests that the stability of P_{II} conformations relative to strand is entropic in origin, being a property of the surrounding water solvent (Supplementary Fig. 4). Such entropic, water-mediated origin of stability of P_{II} helical structure in the unfolded state of polyalanine, coupled with the noncooperative nature of P_{II} itself constitute a fundamental revision in our understanding of protein folding. In this manner, the hydrophobic effect appears to be not only a principal determinant of protein stability and folding pathways,[36] but also to act early in the folding process to preorganize the unfolded state structure.

To directly examine the contribution of unfolded state structure to α-helical (protein) stability, we calculated the configurational partition functions of A_{14} polyalanine, which due to its increased length is able to sample α-helical configurations at sufficient probabilities for analysis. These calculations also appear to be converged as judged from the equivalency of the results of 4 independent simulations using CHARMM27, and accurate, as estimated from the mean α-helical content of 0.25 ± 0.035, as compared with the experimental value of about 30%.[22] Furthermore, calculated length-dependent configurational probabilities of α-helix formation are in general agreement with experimental values,[37,38] whereby formation of short α-helices (nucleation) is unfavorable, with an apparent mean $\sigma = 0.00084 \pm 0.00027$, and formation of long α-helices is favorable (propagation), with an apparent mean $s = 1.1 \pm 0.053$ [Fig. 3(a)]. Configurational P_{II}

TABLE I. Differences in Conformational Entropies for the A_7 Peptide Between Observed Segmentally Helical P_{II} coil and Statistical Coil With Identical Conformational Preferences

	CHARMM27		AMBER98	
N	Δ*S*/seg[a]	Δ*S*/res[b]	Δ*S*/seg[a]	Δ*S*/res[b]
1	—	—	—	—
2	1.1 ± 0.59	0.55 ± 0.42	1.4 ± 0.47	0.70 ± 0.33
3	2.5 ± 0.64	0.83 ± 0.37	2.8 ± 0.89	0.93 ± 0.51
4	4.0 ± 1.1	1.0 ± 0.55	4.1 ± 1.5	1.0 ± 0.75
5	4.4 ± 1.6	0.88 ± 0.72	4.2 ± 2.1	0.84 ± 0.94
6	4.5 ± 2.1	0.75 ± 0.86	4.3 ± 2.5	0.72 ± 1.0

[a]Calculated using $S = R\,P \ln P$, where R is the gas constant, and expressed in cal/mol/K ± 1σ.
[b]Calculated by dividing Δ*S*/segment by the number of contained residues. Note that for segments greater than 3 residues in length, the calculated Δ*S* values report on the configurational entropy of forming a P_{II} helical turn. Since the differences in conformational entropies per residue calculated using different segment lengths are equal to each other, this linear additivity indicates that P_{II} helix formation is noncooperative, consistent with the largely diffusive nature of the conformational search in the unfolded state.

Fig. 1. Contour maps of all individual A_7 alanine (ϕ, ψ) pairs evolved using the (**a**) CHARMM27 and (**b**) AMBER98 parameter sets starting from the P_I configuration (ϕ, ψ) = (+145°, −75°), with the red to blue gradient corresponding to decreasing densities of backbone conformations. Note that the initial P_I configuration is not significantly sampled, consistent with its aqueous instability. There is an absence of α-helical conformations in ensembles calculated using AMBER98, consistent with the improved backbone torsional parameterization. Presence of a small population of left-handed α-helices with (ϕ, ψ) ≈ (+60°, +60°) has been observed to be the potential energy minimum in vacuo.[52] Geometries observed in the centers of P_{II} clusters have (ϕ, ψ) values of (−86°, +161°) using CHARMM27 and (−83°, +133°) using AMBER98, being somewhat more extended and compacted, respectively, as compared to the canonical P_{II} geometry of (−75°, +145°), highlighting the importance of self-consistent microscopic state clustering. Relative probabilities of sampling conformational basins are plotted as a function of alanine position in A_7 using (**c**) CHARMM27 and (**d**) AMBER98 force fields. Error bars represent ±1σ, and are a heuristic measure of convergence of the calculations. Mean P_{II} conformational probabilities calculated using modified CHARMM27 and AMBER98 potential energy functions which lack their ϕ and ψ torsional energy terms (Supplementary Fig. 3), producing the CHARMM27 and AMBER98 force fields similar to Garcia and Sanbonmatsu's modification of AMBER94,[20] are in agreement with those calculated using intact force fields, consistent with the entropic, water-mediated stability of P_{II} that is largely decoupled from the particular form of the potential energy function (Supplementary Fig. 4).

Fig. 2. Apparent P_{II} configurational basin probabilities (**a**) as a function of length for the simulations using CHARMM27 (red dashed circles) and AMBER98 (red solid squares) forcefields, as well as those expected for the respective Flory polymers (blue), generated by enumerating the occurrences of varying lengths of residues in P_{II} conformation (*n*lets), as defined using the self-consistent clustering approach, and normalized by the total possible number of nlets. Configurational basin probabilities expected for Flory polymers were calculated using $P_{\text{nlet}} = \prod_{i=1}^{n} P_i$, since the conformational sampling of adjacent residues in a random coil is independent of each other; setting P_1 of the expected Flory polymer to the value observed in the calculated ensemble creates a statistical coil with identical conformational but different configurational preferences. Ratio (**b**) of the apparent configurational P_{II} probabilities observed for A_7 and those of a Flory polymer with identical (but independent) backbone conformational preferences (green) calculated using CHARMM27 (dashed circles) and AMBER98 (solid squares). Snapshots (**c**) of the calculated unfolded state ensembles of A_7 containing alanine segments in P_{II} helical (blue) and strand configurations (gray). Polypeptide backbone and methyl side-chains are rendered as ribbons and cylinders, respectively.

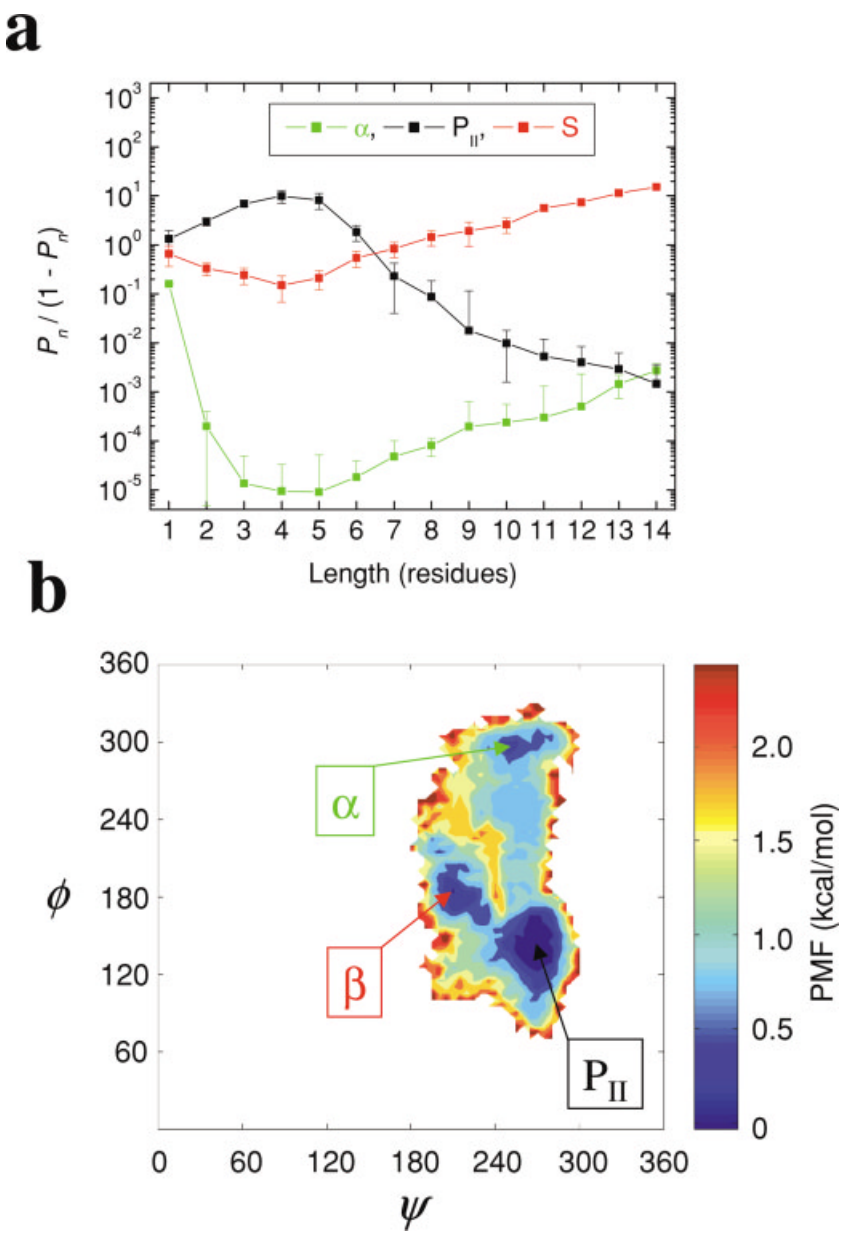

Fig. 3. Configurational probability distribution functions (**a**) of A_{14} showing the internally referenced relative stabilities of α, P_{II}, and strand configurations as a function of length. Note that maximum of P_{II} and minimum of α-helical stabilities coincide at length of 4–5 residues, suggestive of nucleation of α-helices by P_{II} structure in the polyalanine unfolded state, and that depletion of P_{II} configurations is concomitant with the appearance of α-helical structure. Moreover, stabilities of P_{II} and strand configurations have a crossover point at length of 6–7 residues, consistent with the segmentally helical P_{II} coil nature of the polypeptide unfolded state determined from A_7 calculations. Helix–coil parameters were determined by a linear fit of the calculated stabilities ln $[P_n/(1 - P_n)]$ as a function of length n, with the slopes corresponding to free energy differences for the addition of one residue to short (σs; $n = \{1–3\}$) and long (s; $n = \{4–14\}$) α-helices. This corresponds to nucleation (σ) and propagation (s) of α-helices. Potential of mean force (**b**) for backbone dihedral angles of alanine residues in the field of 3 α-helical residues, corresponding to the Gibbs free energies of nucleating the first α-helical turn in the mean field of 3 α-helical residues, referenced to the P_{II} conformational basin, as determined by using $-RT \ln (P/PP_{II})$, where values of P are the probabilities of sampling 5° bins of the (ϕ, ψ) phase space. P_{II} and β conformational basins are separated by a barrier of about 0.5 kcal/mol, less than the thermally available k_BT at 298 K, consistent with the largely diffusive and noncooperative origin of P_{II} structure in the unfolded state of polyalanine. On the other hand, cooperative α-helix formations from the P_{II} and β basins involve barriers of about 1.1 and 1.9 kcal/mol, respectively. Note that the (ϕ, ψ) range has been converted to (0°, 360°) to maintain the continuity of the PMF representation.

stability appears to have a maximum at length of 4–5 residues [Fig. 3(a)], consistent with the segmentally P_{II} helical nature of the A_7 unfolded state [Fig. 2(a)]. As for the A_7 peptide, A_{14} also exhibits significant differences from Flory behavior for lengths up to 7–8 residues (Table II). Most importantly, length-dependent stabilities of P_{II} and

TABLE II. Differences in Conformational Entropies for the A_{14} Peptide Between Observed Segmentally Helical P_{II} Coil and Statistical Coil With Identical Conformational Preferences

N	ΔS/seg[a]
1	—
2	1.2 ± 0.21
3	2.3 ± 0.35
4	3.7 ± 0.52
5	4.6 ± 0.96
6	5.1 ± 1.3
7	4.97 ± 2.0
8	0.13 ± 1.3
9	0.044 ± 0.52
10	0.026 ± 0.42
11	0.015 ± 0.58
12	0.010 ± 0.72
13	0.0070 ± 0.41
14	0.0039 ± 0.24

[a]Calculated using $S = R\,P \ln P$, where R is the gas constant, and expressed in cal/mol K ± 1σ.

α-helical configurations are inversely related, whereby melting of P_{II} (in the unfolded state) is concomitant with formation of α-helices [Fig. 3(a)], suggesting that P_{II} formation in the unfolded state may facilitate length-dependent α-helix formation.

In order to directly evaluate the contribution of unfolded state P_{II} structure to the kinetics of α-helix formation, we calculated the potential of mean force (PMF) for the formation of an α-helical turn in A_{14} by enumerating the backbone torsions of alanine residues when they are in the field of 3 α-helical residues. Such a PMF represents the Gibbs free energy surface forming a 4-residue α-helical turn in the mean field of 3 α-helical residues, as referenced to the most stable conformation of the fourth residue [Fig. 3(b)]. The conformational basins in the PMF surface identified from the analysis correspond to distorted P_{II}, β, and α, with population-weighted average (ϕ, ψ) = ($-81°$, $+137°$), ($-157°$, $+173°$), and ($-98°$, $-69°$), respectively [Fig. 3(b)]. Most importantly, the barrier for α-helix formation from the P_{II} conformation is about 0.8 kcal/mol lower than that from β [Fig. 3(b)], suggesting that P_{II} structure in the unfolded state kinetically facilitates α-helix formation. The energetic preference for such a pathway of α-helix nucleation from P_{II} may be due to the requirement of moving only the ψ angle, ϕ already having an approximate α-helical geometry in P_{II}. On the other hand, formation of α from β requires the motion of both ϕ and ψ angles [Fig. 3(b)]. Moreover, there is an absence of peptide:water: peptide bridges in P_{II},[35] structures that may need to be disrupted in order to complete an α-helical turn from β conformation. In a sense that the transition state for the conversion of P_{II} to α resembles the configurational basin from which it originates, P_{II} structures in the unfolded state of polyalanine act to preorganize as well as guide the ensemble for folding pathway entry. In this way, the segmentally helical P_{II} coil acts to preorganize the unfolded state ensemble of polyalanine and to kinetically guide its α-helix formation.

Cooperativity that defines α-helix formation is characteristic of protein folding in general, with both phenomena exhibiting a nonlinear, sigmoidal dependence on temperature. Both the largely all-or-none behavior and the associated two-state transition mean that the microscopic ensemble is either largely folded or unfolded. However, statistical mechanical models of the helix–coil transition that assume full cooperativity among residues[39] overestimate helical content, as well as its dependence on temperature. Consequently, Zimm and Bragg,[40] and later, Lifson and Roig[41] reduced the overall cooperativity of the transition by employing an Ising-like model, limiting it to nearest neighbor cooperativity. Thus, helix formation is partitioned into fully cooperative nucleation and largely noncooperative extensions steps, thereby reducing the overall cooperativity of the reaction, and more accurately reproducing the observed helical content. The Ising model and its variants are the only way to reduce two-state reaction cooperativity in the presence of a random reference state.[42] However, nearest neighbor cooperativity is not required if the reference state is organized, as is the case for the coil state of polyalanine (Fig. 2). Because the loss of conformational entropy in the formation of a P_{II} helix is about half of α-helix formation, the entropic penalty associated with nucleating the first α-helical turn in the Zimm–Bragg model is significantly reduced in the segmentally helical P_{II} coil (Table I). The Ising-like reduction of overall cooperativity in the helix–(random) coil model of Zimm and Bragg is thus effectively equivalent to a model based on the (preorganized) segmentally helical P_{II} coil state, in which α-helix formation itself is fully cooperative, but the overall cooperativity is reduced by the noncooperative P_{II} formation. Such a model is supported by the calculated configurational thermodynamics and kinetics of A_{14}, in which depletion of P_{II} is concomitant with appearance of α-helical structure (Fig. 3). Modification of the Zimm–Bragg formalism to include a third state leads to a reduction of overall cooperativity, as was demonstrated by inclusion of the 3_{10}-helix.[43] A more pronounced effect is expected from the introduction of a P_{II}-helix, since its formation is less cooperative as compared to that of a 3_{10}-helix (Table I). Finally, the Zimm–Bragg random coil model is incompatible with the observed kinetics of α-helix formation, whereby in addition to the exponential phase corresponding to cooperative barrier crossing, several stretched and nonexponential relaxations are observed,[44,45] indicative of noncooperative processes; these could be attributed to processes such as P_{II} formation. A mechanism of α-helix formation that emerges from these considerations involves diffusive, noncooperative P_{II} formation in the segmentally helical P_{II} coil state that facilitates nucleation and propagation of α-helices.

Although recently Flory's independent pair hypothesis was observed to be sterically invalid in the α-helical region of the Ramachandran map, no steric interference was observed for extended conformations such as β and P_{II}.[46] Here, we demonstrate that Flory's hypothesis and the associated assumption of randomness of the unfolded state fail for these conformations as well, albeit as a result of an

entropic effect of the surrounding water, whereby polyalanine segments up to 7 residues in length form P_{II} helices [Figs. 2(a) and 3(a)]. Thus, the conformational space actually sampled by unfolded polypeptides is considerably smaller than the one assumed by random and statistical coil models of the unfolded state. The extent of cooperation (κ) in achieving observed biological efficiencies of protein folding between such (thermodynamic) preorganization of the unfolded state and (kinetic) facilitation by folding pathways, regardless of their nature, can be quantified by $\kappa = \Delta S_U/\Delta S_F$, where ΔS_U is the loss of conformational entropy in the unfolded state, and ΔS_F is the change in conformational entropy upon folding. In the case of α-helix formation by polyalanine, calorimetrically determined $\Delta S_F \approx 2$ cal/mol/K/residue,[11] and ΔS_U, calculated as the difference in Shannon entropies of the observed segmentally helical P_{II} coil and the expected Flory statistical coil with identical conformational biases, is about 1 cal/mol/K/residue (Table I), yielding $\kappa \approx 0.5$. Thus, as much as half of the folding search of α-helix formation in polyalanine is accomplished by water-mediated, entropic preorganization of the unfolded state into segmentally helical P_{II} coil. Values of κ and the underlying unfolded state structure propensities are expected to depend both on polymer nature and sequence (unpublished observations), and promise to be useful both for polymer structure prediction and design.

Although the cooperative nature of protein stability and folding requires the existence of folding pathways, diffusive search in the unfolded state prior to the crossing of the rate-limiting barrier can nevertheless be productive, particularly if the configurational space available to the search is reduced relative to conventional random coil models.[4] Such a process is in fact a corollary of the statistical mechanics of polymer folding of Wolynes and colleagues[47] and the associated funnel-like topologies of several experimentally and theoretically determined folding free energy landscapes.[48,49] The segmentally helical P_{II} coil observed in the unfolded state of polyalanine in water, reflecting the inherent physics of the polypeptide backbone, accomplishes just that, capturing more than half of the unfavorable conformational entropy of α-helix formation. Such reduction of the otherwise biologically intractable conformational space acts to preorganize the unfolded state ensemble for folding pathway(s) entry. Indeed, although local structural motifs such as helices and turns can form on ultrafast nanosecond timescales in isolation, several orders of magnitude faster than those of protein folding,[5] their formation in proteins is often,[50] but not always,[51] coupled to the (cooperative) traversal of the rate-limiting barrier, suggesting that preorganization of the unfolded state both locally in the vicinity of P_{II} helices, and globally as a result of chain compaction, is directly related to the nucleation of folding pathways. In this way, the requirements for efficient folding are satisfied both by the thermodynamic organization of the conformational search in the unfolded state and its kinetic guidance by the resulting folding pathways.

ACKNOWLEDGMENTS

We are grateful to Neville Kallenbach, George Rose, William Eaton, Tobin Sosnick, Angel Garcia, and Alex Proekt for helpful and critical discussions, to Allan Capili and Kathy Borden for reading the manuscript, and to Benjamin Goldsteen for technical support. We thank George Rose, Tobin Sosnick, and Angel Garcia for sharing unpublished manuscripts.

REFERENCES

1. Levinthal C. Are there pathways for protein folding? Extrait J Chim Phys 1968;65:44–45.
2. Daggett V, Fersht AR. Is there a unifying mechanism for protein folding? Trends Biochem Sci 2003;28:18–25.
3. Sosnick TR, Mayne L, Englander SW. Molecular collapse: the rate-limiting step in two-state cytochrome c folding. Proteins 1996;24:413–426.
4. Baldwin RL, Rose GD. Is protein folding hierarchic?: I. Local structure and peptide folding. Trends Biochem Sci 1999;24:26–33.
5. Eaton WA, Munoz V, Hagen SJ, Jas GS, Lapidus LJ, Henry ER, Hofrichter J. Fast kinetics and mechanisms in protein folding. Annu Rev Biophys Biomol Struct 2000;29:327–359.
6. Aune KC, Salahuddin A, Zarlengo MH, Tanford C. Evidence for residual structure in acid- and heat-denatured proteins. J Biol Chem 1967;242:4486–4489.
7. Tiffany ML, Krimm S. New chain conformations of poly(glutamic acid) and polylysine. Biopolymers 1968;6:1379–1382.
8. Mu Y, Kosov DS, Stock G. Conformational dynamics of trialanine in water: 2. Comparison of AMBER, CHARMM, GROMOS, and OPLS force fields to NMR and infrared experiments. J Phys Chem B 2003;107:5064–5073.
9. Woutersen S, Hamm P. Isotope-edited two-dimensional vibrational spectroscopy of a short alpha-helix in water. J Chem Phys 2001;114:2727–2737.
10. Eker F, Cao X, Nafie L, Schweitzer-Stenner R. Tripeptides adopt stable structures in water: a combined polarized visible Raman, FTIR, and VCD spectroscopy study. J Am Chem Soc 2002;124:14330–14341.
11. Shi Z, Olson CA, Rose GD, Baldwin RL, Kallenbach NR. Polyproline II structure in a sequence of seven alanine residues. Proc Natl Acad Sci USA 2002;99:9190–9195.
12. Flory PJ. Statistical mechanics of chain molecules. New York: Wiley; 1969.
13. Soman KV, Karimi A, Case DA. Unfolding of an alpha-helix in water. Biopolymers 1991;31:1351–1361.
14. Daggett V, Levitt M. Molecular dynamics simulations of helix denaturation. J. Mol Biol 1992;223:1121–1138.
15. Karpen ME, Tobias DJ, Brooks CL III. Statistical clustering techniques for the analysis of long molecular dynamics trajectories: analysis of 2.2-ns trajectories of YPGDV. Biochemistry 1993;32:412–420.
16. Hummer G, Garcia AE, Garde S. Conformational diffusion and helix formation kinetics. Phys Rev Lett 2000;85:2637–2640.
17. Hummer G, Garcia AE, Garde S. Helix nucleation kinetics from molecular simulations in explicit solvent. Proteins 2001;42:77–84.
18. Vila JA, Ripoll DR, Villegas ME, Vorobjev YN, Scheraga HA. Role of hydrophobicity and solvent-mediated charge–charge interactions in stabilizing alpha-helices. Biophys J 1998;75:2637–2646.
19. Garcia AE. Characterization of non-alpha helical conformations in Ala peptides. 2003. Submitted for publication.
20. Garcia AE, Sanbonmatsu KY. Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. Proc Natl Acad Sci USA 2002;99:2782–2787.
21. Zaman MH, Shen MY, Berry RS, Freed KF, Sosnick TR. Investigations into sequence and conformational dependence of backbone entropy, inter-basin dynamics and the Flory isolated-pair hypothesis for peptides. J Mol Biol 2003;331:693–711.
22. Spek EJ, Olson AC, Shi Z, Kallenbach NR. Alanine is an intrinsic α-helix stabilizing amino acid. J Am Chem Soc 1999;121:5571–5572.
23. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T,

Nguyen DT, Prod'hom B, Reiher IWE, Roux B III, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
24. Jorgensen W, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
25. Perkins SJ. Protein volumes and hydration effects: the calculations of partial specific volumes, neutron scattering matchpoints and 280-nm absorption coefficients for proteins and glycoproteins from amino acid sequences. Eur J Biochem 1986;157:169–180.
26. Kentsis A, Mezei M, Osman R. MC-PHS: a Monte Carlo implementation of the primary hydration shell for protein folding and design. Biophys J 2003;84:805–815.
27. Mezei M. One the selection of the particle to be perturbed by the Monte Carlo method. J Comp Phys 1981;39:128–136.
28. Mezei M. Efficient Monte Carlo sampling for long molecular chains using local moves, tested on a solvated lipid bilayer. J Chem Phys 2003;118:3874–3880.
29. Rosenhouse-Dantsker A, Osman R. Application of the primary hydration shell approach to locally enhanced sampling simulated annealing: computer simulation of thyrotropin-releasing hormone in water. Biophys J 2000;79:66–79.
30. Vuister GW, Bax A. Quantitative J correlation: a new approach for measuring homonuclear three-bound j(HN-H) coupling constant in 15N-enriched proteins. J Am Chem Soc 1993;115:7772–7777.
31. Ben-Naim A. Statistical thermodynamics for chemists and biochemists. New York: Plenum Press; 1992.
32. Mezei M. Modified proximity criteria for the analysis of solvation of a polyfunctional solute. Mol Simulat 1988;1:327–332.
33. Zhang H, Wong CF, Thacher T, Rabitz H. Parametric sensitivity analysis of avian pancreatic polypeptide (APP). Proteins 1995;23: 218–232.
34. Lopez MM, Chin DH, Baldwin RL, Makhatadze GI. The enthalpy of the alanine peptide helix measured by isothermal titration calorimetry using metal-binding to induce helix formation. Proc Natl Acad Sci USA 2002;99:1298–1302.
35. Mezei M, Fleming PJ, Srinivasan R, Rose GD. The solvation free energy of the peptide backbone is strongly conformation-dependent. Proteins 2003. In press.
36. Dill KA. Polymer principles and protein folding. Protein Sci 1999;8:1166–1180.
37. Platzer KEB, Ananthanarayanan VS, Andreatta RH, Scheraga HA. Helix-coil stability constants for the naturally occurring amino acids in water: IV. Alanine parameters from random poly(hydroxypropyl-glutamine-co-L-alanine). Macromolecules 1972;5:177–187.
38. Rohl CA, Scholtz JM, York EJ, Stewart JM, Baldwin RL. Kinetics of amide proton exchange in helical peptides of varying chain lengths: interpretation by the Lifson–Roig equation. Biochemistry 1992;31:1263–1269.
39. Schellman JA. The factors affecting the stability of hydrogen-bonded polypeptide structures in solution. J Phys Chem 1958;62: 1485–1494.
40. Zimm BH, Bragg JK. Theory of phase transition between helix and random coil in polypeptide chains. J Chem Phys 1959;31:526–535.
41. Lifson S, Roig A. On the theory of helix-coil transitions in polypeptides. J Chem Phys 1961;34:1963–1974.
42. Ising E. Beitrag zur theorie des ferromagnetismus. [Contribution to the theory of ferromagnetism] Zeitschrift fur Physik 1925;31: 253–258.
43. Sheinerman FB, Brooks CL III. Helices in peptides and proteins as studied by modified Zimm–Bragg theory. J Am Chem Soc 1995;117:10098–10103.
44. Huang CY, Getahun Z, Zhu Y, Klemke JW, DeGrado WF, Gai F. Helix formation via conformation diffusion search. Proc Natl Acad Sci USA 2002;99:2788–2973.
45. Salbelko J, Ervin J, Gruebele M. Observation of strange kinetics in protein folding. Proc Natl Acad Sci USA 1999;96:6031–6036.
46. Pappu RV, Srinivasan R, Rose GD. The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. Proc Natl Acad Sci USA 2000;97:12565–12570.
47. Onuchic JN, Suthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 1997;48:545–600.
48. Shea JE, Brooks CL III. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. Annu Rev Phys Chem 2001;52:499–535.
49. Schuler B, Lipman EA, Eaton WA. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. Nature 2002;419:743–747.
50. Krantz BA, Mayne L, Rumbley J, Englander SW, Sosnick TR. Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. J Mol Biol 2002;324:359–371.
51. Kubelka J, Eaton WA, Hofrichter J. Experimental tests of villin subdomain folding simulations. J Mol Biol 2003;329:625–630.
52. Roterman IK, Gibson KD, Scheraga HA. A comparison of the CHARMM, AMBER and ECEPP potentials for peptides: I. Conformational predictions for the tandemly repeated peptide (Asn-Ala Asn-Pro)9. J Biomol Struct Dyn 1989;7:391–419.