Reconciling Observations of Sequence-Specific Conformational Propensities with the Generic Polymeric Behavior of Denatured Proteins†

Hoang T. Tran, Xiaoling Wang, and Rohit V. Pappu*

Department of Biomedical Engineering and Center for Computational Biology, Washington University in St. Louis, Campus Box 1097, St. Louis, Missouri 63130

Received February 2, 2005; Revised Manuscript Received June 30, 2005

ABSTRACT: Radii of gyration of denatured proteins vary with chain length and are insensitive to details of amino acid sequence. Observations of sequence independence in polymeric properties conflict with results from spectroscopic experiments, which suggest the presence of sequence-specific residual structure in denatured states. Can we reconcile the two apparently conflicting sets of observations? To answer this question, we need knowledge of the ensemble of conformations accessible to proteins in good solvents. The excluded-volume limit provides an ideal mimic of polymers in good solvents. Therefore, we attempt to solve the "reconciliation problem" by simulating conformational ensembles accessible to peptides and proteins in the excluded-volume limit. Analysis of these ensembles for a variety of polypeptide sequences leads to results that are consistent with experimental observations of sequence-specific conformational preferences in short peptides and the scaling behavior of polymeric quantities for denatured proteins. Reconciliation in the excluded-volume limit comes about due to a tug of war between two factors, namely, minimization of steric overlap and the competing effects of conformational entropy. Minimization of steric overlap promotes chain stretching and leads to experimentally observed sequence-dependent preferences for locally extended segments such as polyproline II helices, β-strands, and very short stretches of α-helix. Conformational entropy opposes chain stretching, and the calculated persistence length for sequence-dependent conformational preferences is less than five amino acids. This estimate does not vary with amino acid sequence. The short persistence lengths lead directly to experimental observations of generic sequence-independent behavior of radii of gyration for denatured proteins.

There is growing recognition of the important role played by unfolded proteins in the thermodynamics and kinetics of protein folding (*1*−*17*). Recent evidence indicates that many important biological functions can be ascribed to partially or fully unfolded forms of proteins (*18*−*21*). Furthermore, self-association of unfolded or partially unfolded proteins nucleates misfolding and aggregation of proteins: processes that are associated with a variety of neurodegenerative and systemic diseases (*14*). An improved understanding of unfolded proteins is clearly in order (*22*, *23*).

It is believed that the unfolded state is stabilized by the addition of chemical denaturants such as urea and GdnHCl. Several attempts have been made to characterize the ensemble of conformations accessible to denatured proteins. The use of synchrotron sources in X-ray and neutron scattering studies allows reliable measurement of the sizes of denatured proteins (*24*−*26*). Advances in optical and NMR[1] spectroscopy allow scrutiny of the ensemble of conformations accessible to peptides and proteins under different solution conditions (*27*−*44*). Peptide studies are important because it is not trivial to assign NMR signals or interpret spectroscopic data for denatured proteins (*45*). Peptides cannot have long-range tertiary structure. Therefore, they represent reasonable model systems to study local conformational preferences of the denatured state under folding conditions (*4*, *32*, *33*, *38*).

Recent evidence suggests that left-handed polyproline II (P_{II}) helices are important motifs for short peptides (*29*−*44*). These observations have helped rekindle interest in the hypothesis of Tiffany and Krimm (*47*−*50*), who proposed that denatured proteins are a concatenation of segments that fluctuate between the P_{II} helix motif and a "wide sampling of a standard dipeptide energy map (*50*)". Is this hypothesis valid? This question cannot be answered without first solving the "reconciliation problem" (*54*) identified recently by Plaxco and co-workers (*25*, *26*).

The Reconciliation Problem. How does one reconcile observations of native and non-native site-specific conformational propensities in peptides and residual structure in denatured proteins with unequivocal evidence that polymeric quantities such as average radii of gyration (R_g) are insensitive to the details of amino acid sequence (*25*, *26*)? Using SAXS data, Plaxco and co-workers have shown that for chemically denatured proteins, irrespective of amino acid sequence, the variation of R_g with chain length (N) takes the form $R_g = R_o N^{0.6}$. The value of R_o is 1.93 ± 0.01. R_o

† The work of H.T. is supported by funds from the National Science Foundation, Grant MCB-0416766, awarded to R.V.P. X.W. is supported by a grant from the March of Dimes Foundation awarded to R.P.

* To whom correspondence should be addressed. Telephone: (314) 362-2057. Fax: (314) 362-0234. E-mail: pappu@biomed.wustl.edu.

[1] Abbreviations: NMR, nuclear magnetic resonance; P_{II}, polyproline II; P_{hyp}, polyhydroxyproline; R_g, radius of gyration; R_e, end-to-end distance; SAXS, small-angle X-ray scattering; UV−CD, ultraviolet−circular dichroism.

10.1021/bi050196l CCC: $30.25
Published on Web 08/06/2005

provides a measure of the volume excluded by chain residues to accommodate favorable interactions with the surrounding solvent. In direct contrast to the SAXS data, results from optical and NMR spectroscopy indicate that site-specific and local secondary structure propensities for peptides (*27*−*44*, *52*) and denatured proteins (*5*, *6*, *8*, *10*−*13*, *46*) are sensitive to amino acid sequence and perturbations in solution conditions (*53*).

In this work, we describe a simple solution to the reconciliation problem. Before doing so, we summarize the necessary polymer physics jargon to provide an appropriate context for the development of our solution.

Terminology. Consider a linear, flexible polymer with N residues. Quantities such as the average radius of gyration (R_g) and the average end-to-end distance (R_e) follow universal power law dependence on the chain length, N. Specifically, the variation of R_g with N can be written as $R_g = R_o N^{\nu}$. Here, $R_o \equiv R_o(\upsilon_{ex},b,d)$, and υ_{ex} is a measure of the average volume excluded per residue to accommodate favorable interactions with the surrounding solvent; b is the Kuhn length of the chain defined as the ratio R_e^2/R_L, where R_L is the contour length of the polymer; and d is the hard-sphere diameter of each residue (*55*−*58*).

For rodlike polymers $R_g \sim N$. The symbol "∼" is meant to denote that R_g "scales as" N. For flexible, nonrodlike polymers the energetic cost of bringing a pair of residues into a specified distance of separation r will depend on the balance of chain−chain and chain−solvent interactions. In an ideal (very good) solvent, the residues are chemically identical to the surrounding solvent. Most chain−chain and chain−solvent interactions counterbalance each other, and the effective chain−chain interactions are best described using purely excluded-volume interactions between residues. As a result, $R_g \sim N^{0.6}$ and $\upsilon_{ex} = b^2d$ (*58*).

In a typical good solvent, the chain residues are chemically equivalent although not identical to the surrounding solvent. This yields $R_g \sim N^{0.6}$ and $0 < \upsilon_{ex} < b^2d$. Most importantly, in good solvents υ_{ex} is positive because chain−solvent contacts are preferred to chain−chain contacts and $R_g \sim N^{0.6}$, which implies that the distribution of accessible configurations is still determined primarily by excluded-volume effects (*58*). This last assertion has received unequivocal support from detailed theoretical calculations, which show that real chains in good solvents have the same universal features as self-avoiding walks on lattices (*59*−*63*).

In addition to the solvent regimes discussed above, there are two other solvent types for flexible chains. These are theta and poor solvents (*55*, *56*, *58*). In a theta solvent, the effects of repulsive interactions exactly counterbalance the effects of attractive interactions, and $R_g \sim N^{0.5}$ and $\upsilon_{ex} = 0$. This is the regime of ideal chain models such as freely jointed chains, freely rotating chains, and the rotational isomeric approximation of Flory (*64*). In a poor solvent attractive interactions dominate, and this yields compact configurations such that $R_g \sim N^{0.33}$ and $-b^2d < \upsilon_{ex} < 0$ (*58*). The extreme limit of a poor solvent, also known as a nonsolvent, corresponds to $R_g \sim N^{0.33}$ and $\upsilon_{ex} = -b^2d$ (*58*).

Modeling the Denatured State Ensemble. The data of Plaxco and co-workers (*26*) for the variation of R_g with chain length for denatured proteins are consistent with Tanford's observation that denatured proteins behave like polymers in good solvents (*2*, *3*). Conversely, folded proteins, which are compact globules, behave like chains in a poor solvent (*25*, *56*).

We have developed an atomistic model to predict the ensemble of conformations accessible to denatured proteins. This is accomplished by analyzing conformational ensembles accessible to different polypeptides in the excluded-volume limit. In the following sections, we show that our calculations meet the dual challenges of (a) predicting sequence-specific behavior of peptides, which we find to be consistent with numerous spectroscopic observations, and (b) reproducing good solvent scaling observed in SAXS experiments. We conclude with an explanation of our solution to the reconciliation problem, which rests on the direct connection between excluded-volume ensembles and denatured states.

MATERIALS AND METHODS

Potential Functions. We use purely repulsive inverse power potentials (*67*, *66*) to model interatomic excluded-volume interactions. For a given conformation, specified by a unique set of backbone ϕ, ψ, and side chain χ angles, the inverse power potential energy (U) is a sum of pairwise interactions:

$$U = \sum_i \sum_{j<i} \epsilon_{ij}(\sigma_{ij}/r_{ij})^n \qquad (1)$$

We set $n = 14$, although as shown in the Results and Discussion section our results are relatively insensitive to this choice. Bond lengths and bond angles are fixed at equilibrium values taken from the work of Engh and Huber (*76*). The peptide unit is always *trans* with $\omega = 179.5°$. The summation in eq 1 is over all unique pairs of *nonbonded atoms* (including 1−4 interactions). In the interest of clarity, we reiterate that the only interactions in our model are the nonbonded excluded-volume terms. In eq 1, σ_{ij} is the hard-sphere contact distance (*68*), r_{ij} the interatomic separation, and the dispersion parameters ϵ_{ij} are determined by the static polarizabilities of individual atoms (*69*, *70*). Parameters for σ_{ij} and ϵ_{ij} are available as Supporting Information.

For proline, we supplement the potential in eq 1 with an additional term to model the coupling between puckering of the pyrrolidine ring, C_γ-*endo* or C_γ-*exo*, and rotations around the preferred backbone ϕ-angle (*71*−*74*). In a Monte Carlo simulation, when the selected residue is proline, we first randomly choose a pucker state. For C_γ-*exo* we set $\chi_1 = -28°$ and $\chi_2 = 41°$, and for C_γ-*endo* we set $\chi_1 = 32°$ and $\chi_2 = -40°$. To ensure that only reasonable pucker-state-dependent ϕ-angles are preferred, we add a harmonic restraint potential $U_\phi = k_\phi(\phi - \phi_o)^2$ to the inverse power potential function in eq 1. Here, $k_\phi = 0.006$ kcal mol^{-1} deg^{-2}, $\phi_o = -71°$ if the pucker state is C_γ-*endo*, and $\phi_o = -60.8°$ if the pucker state is C_γ-*exo*. The restraint potentials were parametrized to reproduce the pucker-dependent distribution of ϕ-angles seen for proline in a database of high-resolution crystal structures (*75*).

Monte Carlo Simulations. The degrees of freedom in all of our calculations are the backbone ϕ, ψ, and side chain χ angles. Conformations were generated using torsional space Metropolis Monte Carlo simulations (*77*, *78*). The goal in Monte Carlo sampling is to generate conformations that minimize the overall free energy of the system. In simulations

such as ours, the free energy is $-RT \ln Q_{conf}$, where Q_{conf} is the conformational partition function. Q_{conf} takes the form $Q_{conf} = \int d\phi_1 d\psi_1 d\chi_1^{(i)} \ldots d\phi_N d\psi_N d\chi_N^{(i)} \exp(-U/RT)$, where R is the ideal gas constant, T is the simulation temperature, and U is the potential function from eq 1, and the integral is to be carried out over all possible values of ϕ, ψ, and side chain χ angles for each amino acid. Estimating Q_{conf} is impossible for large multidimensional systems. An alternative approach is to sample conformational space "intelligently" by increasing the likelihood of sampling conformations that make significant contributions to Q_{conf}, namely, those with favorable Boltzmann weights. This class of sampling approaches known as importance sampling allows one to generate ensembles that collectively minimize the overall free energy. Although these approaches do not provide an estimate of Q_{conf}, they allow us to identify conformations that make significant contributions to Q_{conf} at equilibrium. The Metropolis Monte Carlo (MMC) method is one such approach (*78*, *79*). A detailed discussion of this technique is beyond the scope of this work since excellent pedagogic descriptions are available in the standard simulation literature (*77*, *79*).

The major decision to make in designing an MMC simulation is the choice of move sets; i.e., how do we generate truly random, uncorrelated conformations in a large multidimensional space? In the current work, we use move sets known commonly as pivot moves. All trial conformations are generated using single-residue moves. The residue whose conformation is to be altered is selected at random. The ϕ, ψ, and χ angles for the selected residue are assigned random values. To improve sampling efficiency, the (ϕ,ψ) values are chosen from regions devoid of steric clashes. This is done by selecting (ϕ,ψ) values from the allowed regions of a Ramachandran map. Each side chain torsion is set by random selection of angles drawn from the interval [−180°, 180°]. The temperature in all calculations is 298 K. For each peptide, the number of Monte Carlo moves was 1.1×10^8, the first 10^7 of which were discarded. For proteins where the number of amino acids is greater than 50, the number of moves is increased to 10^9. Snapshots were saved every 10^4 moves, for 10^4 snapshots per simulation. Standard block averaging methods were used to estimate statistical errors (*77*).

Calculation of Conformational Propensities Using the Method of Inherent Structures. Conformational ensembles for peptides are determined primarily by the set of backbone (ϕ,ψ) angles accessible to each residue. We calculate the probabilities of all amino acids for different (ϕ,ψ) angles in a series of host−guest peptides. To calculate conformational propensities, we need a reliable method for clustering. We use the method of inherent structures introduced by Stillinger and Weber (*80*, *81*). In this approach, the (ϕ,ψ,χ) space for each peptide is divided into "basins", one surrounding each local minimum, and the individual conformations from the simulation are mapped to their respective basins. A summary of the principles that underlie this clustering technique follows.

Formally, the canonical partition function at temperature T is written as $Z = \int \exp[-\beta U(\mathbf{X})]\, d\mathbf{X}$, where $U(\mathbf{X})$ is the potential energy of conformation $\mathbf{X}$. Here, $\mathbf{X}$ denotes ϕ, ψ, and χ angles and $\beta = 1/RT$. Replacing $U(\mathbf{X})$ by $U_\gamma + [U(\mathbf{X}) - U_\gamma] = U_\gamma + \Delta_\gamma U(\mathbf{X})$, where U_γ is the potential energy of the local minimum γ reached by minimization from conformation $\mathbf{X}$, yields $Z = \sum_\gamma \{\exp(-\beta U_\gamma) \int_{R(\gamma)} \exp(-\beta \Delta U)\, d\mathbf{X}\} = w_\gamma(\beta) \exp(-\beta U_\gamma)$. $R(\gamma)$ is the envelope of conformations that can be mapped via energy minimization to local minimum γ, while $w_\gamma(\beta)$ is the temperature-dependent weight of the basin centered around this minimum. Local minima obtained via conformational mapping are *inherent structures*, and the weights $w_\gamma(\beta)$ provide a measure of the effective volume of each basin in phase space.

For each peptide, we first use torsional space MMC simulations to generate an equilibrium distribution of conformations. From the equilibrium Monte Carlo "trajectory", we choose a set of uncorrelated snapshots. These are snapshots saved once every 10^4 Monte Carlo moves. Starting from each of these snapshots, we carry out torsional space energy minimizations to generate a set of inherent structures. Energy minimizations are carried out using a Polak−Ribiere conjugate gradient energy minimization (*82*). The convergence criterion is set to 10^{-6} kcal mol^{-1} rad^{-1}. For a particular host−guest peptide, an inherent structure is characterized by a set of ϕ, ψ, and χ angles. We are interested in backbone conformational propensities of guest residues. Therefore, for each sequence, we analyze the complete set of inherent structures and compute the frequencies with which residues of interest adopt particular (ϕ,ψ) angles.

RESULTS AND DISCUSSION

Testing the Accuracy of Inverse Power Potentials. We first show that the use of inverse power potentials leads to results for alanine dipeptide (*N*-acetylalanine-*N*′-methylamide) (Figure 1A) that agree with those obtained using more sophisticated potentials. To make this comparison, we systematically varied the backbone (ϕ,ψ) angles in 1° increments between −180° and 180°. The resultant contour map for the potential energy surface is shown in Figure 6A. For each of the sampled conformations we compute the Boltzmann weights at $T = 298$ K. These weights, when correctly normalized, yield probability values that range between 0 and 1 for each point in (ϕ,ψ) space. Conformational statistics for specific regions are obtained by summing over the probabilities associated with (ϕ,ψ) values that lie within the region. Results are tabulated in Table 1. These statistics are compared to those obtained by Hu et al. (*83*), who generated conformational probability distributions for alanine dipeptide in water using hybrid quantum mechanics/molecular mechanics (QM/MM) molecular dynamics simulations. In Table 1, we show that our predictions compare favorably with the results of Hu et al. In Figure 1B, we show that the quality of agreement is not overly sensitive to the choice of repulsive exponent, n in eq 1.

Why Do Results Obtained Using Models of Vastly Differing Complexities Agree with Each Other? Three sets of interactions determine conformational statistics for alanine dipeptide. These are (1) intrapeptide sterics, (2) conformation-dependent peptide solvation, and (3) intrapeptide electrostatic interactions. For conformations where peptide solvation is favorable, the gas-phase intrapeptide electrostatic interactions are unfavorable. Conversely, when peptide solvation is unfavorable, intrapeptide electrostatic interactions are favored. Because conformation-dependent solvation

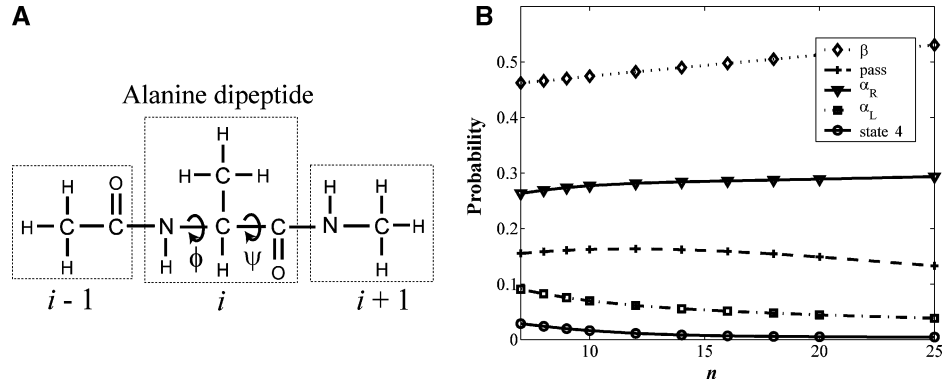

FIGURE 1: (A) Structure of alanine dipeptide, which consists of a single alanine residue flanked by two peptide bonds. (B) Variation of alanine dipeptide conformational statistics with *n*, the exponent of the inverse power potential. Definitions of conformational regions are those of Hu et al. (*83*).

Table 1: Comparison of Distribution of Conformational Populations for Alanine Dipeptide

	conformational propensities	
interval[a]	calcd by Hu et al. using QM/MM[b]	calcd using inverse power potentials[b]
β	0.48	0.49
pass	0.16	0.16
α_R	0.27	0.28
α_L	0.07	0.05
state 4	0.01	0.01

[a] Definitions of conformational intervals and names used for these intervals are identical to those of Hu et al. (*83*). [b] Hu et al. reported conformational propensities as probabilities. We follow their convention here.

balances the effects of intrapeptide gas-phase electrostatics, steric interactions become the main determinants of peptide conformational propensities (*84*).

Calculation of Intrinsic Conformational Propensities and Comparison to Experimental Data. We calculate intrinsic conformational propensities using inherent structures for blocked amino acids of the form *N*-acetyl-Xaa-*N*′-methylamide. Xaa refers to one of the 22 amino acids, which includes the 20 naturally occurring ones plus norvaline (v) and norleucine (l). We compare our propensities to those measured by Eker et al. (*43*). In their experiments, the amide I′ profiles of AXA tripeptides were obtained using a combination of FTIR, polarized Raman, and vibrational CD spectroscopies. The AXA tripeptides are the closest mimics of the blocked peptides (dipeptides) used in our calculations. For each guest residue X, Eker et al. use the measured amide I′ profiles to assign single values of (ϕ,ψ) drawn from the upper left quadrant of (ϕ,ψ) space as the most likely backbone conformation for the guest. On the basis of these assignments, they classified guest residues as preferring P_{II}, β, or both. These experiments cannot provide quantitative information about conformational distributions. Therefore, the best we can do is to compare our detailed quantitative distributions to qualitative trends seen from experiments.

To calculate conformational propensities, we focus on nine $40° \times 40°$ regions in (ϕ,ψ) space and compute the probability of finding inherent structures within each region. Definitions and identities of the nine regions are shown in Table 2. In Figure 2, we show results to justify our choice of conformational regions in (ϕ,ψ) space. This figure shows that (ϕ,ψ) values for dominant inherent structures form well-defined clusters. Tabulation of intrinsic propensities for all 22 amino acids is available as Supporting Information. The data are plotted in Figure 3. We now summarize the main findings.

(1) Conformational propensities vary with amino acid type, and P_{II} is the dominant region for many amino acids. Exceptions are glycine and amino acids with β- and γ-branched hydrocarbons in their side chains. In agreement with Eker et al. (*43*), we find that proline, lysine, aspartic acid, and glutamic acid show marked preference for P_{II}.

(2) Alanine has approximately equivalent propensities for P_{II} and β_P. The findings of Eker et al. (*43*) are similar to ours for alanine but disagree with ours for glycine. While this disagreement might be cause for concern, we note that our calculated propensities for glycine and alanine agree with the estimates of Ferreon and Hilser (*38*). They used isothermal titration calorimetry to measure variations in binding affinities of wild-type and point mutants of the Sem5 peptide to an SH3 domain. Ferreon and Hilser estimate the P_{II} propensities of alanine and glycine to be ~30% and ~10%, respectively.

(3) Leucine is unique in its clear preference for the three bridge regions $P_{II} \leftrightarrow \beta_A$, $P_{hyp} \leftrightarrow \beta_P$, and $\alpha_R \leftrightarrow \alpha_P$. Eker et al. (*43*)-observed this as well and suggest that this preference is a consequence of conformational averaging and reflects an alanine-like behavior where leucine has equal propensity for P_{II} and β. Our results indicate otherwise. The distribution of ϕ angles for inherent structures of leucine is unimodal. This is unlike alanine, which has a bimodal distribution for its ϕ angle.

(4) The β-branched amino acids valine and isoleucine have very low P_{II} propensities. Instead, the preferred regions in the top left quadrants are P_{hyp} (*85*), $P_{II} \leftrightarrow \beta_A$, and, to a lesser extent, β_A. The effect of β-branching becomes clear when one compares differences in propensities of valine to norvaline and leucine or isoleucine to norleucine.

Context-Dependent Conformational Propensities. Conformational propensities for all 22 amino acids (including norvaline and norleucine) were calculated in different host sequences of varying lengths. In the host−guest sequences, *N*-acetyl-$(haa)_L$-Xaa-$(haa)_L$-*N*′-methylamide, the host residues haa are one of glycine, alanine, valine, phenylalanine, or proline. The number of host residues (haa) on either side of the guest, *L*, varies from $L = 0$ to $L = 4$. Here, we summarize the main results from 440 Monte Carlo simulations. Detailed tabulations of context-dependent propensities and distribu-

Table 2: Definitions of Conformational Regions for Calculation of Conformational Propensities[a]

region	(ϕ,ψ) interval	region	(ϕ,ψ) interval	region	(ϕ,ψ) interval
P_{II}	$-100° < \phi \leq -60°$ $140° < \psi \leq 180°$	P_{hyp}	$-100° < \phi \leq -60°$ $100° < \psi \leq 140°$	α_R	$-100° < \phi \leq -60°$ $-80° < \psi \leq -40°$
$P_{II}\leftrightarrow\beta_A$	$-140° < \phi \leq -100°$ $140° < \psi \leq 180°$	$P_{hyp}\leftrightarrow\beta_P$	$-140° < \phi \leq -100°$ $100° < \psi \leq 140°$	$\alpha_R\leftrightarrow\alpha_P$	$-140° < \phi \leq -100°$ $-80° < \psi \leq -40°$
β_A	$-180° \leq \phi \leq -140°$ $140° < \psi \leq 180°$	β_P	$-180° \leq \phi \leq -140°$ $100° < \psi \leq 140°$	α_P	$-180° \leq \phi \leq -140°$ $-80° < \psi \leq -40°$

[a] The significance of names assigned to conformational regions is as follows: P_{II} ≡ polyproline II; $P_{II}\leftrightarrow\beta_A$ ≡ the bridge region between P_{II} and the canonical antiparallel β-strand regions; β_A ≡ encompasses backbone torsion angles for canonical antiparallel β-strands; P_{hyp} ≡ encompasses canonical backbone torsion angles preferred by polyhydroxyproline (*73*); $P_{hyp}\leftrightarrow\beta_P$ ≡ the bridge region between P_{hyp} and the canonical parallel β-strand region; β_P ≡ encompasses backbone torsion angles for canonical parallel β-strands; α_R ≡ encompasses backbone torsion angles for canonical right-handed α-helices; $\alpha_R\leftrightarrow\alpha_P$ ≡ bridge between α_R and region proximal to α_R, i.e., α_P.

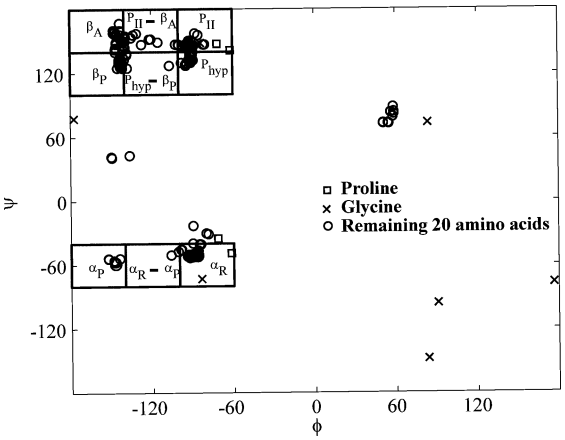

FIGURE 2: (ϕ,ψ) values of dominant inherent structures for all dipeptides. Conformational regions used in inherent structure analysis are shown as boxed regions.

tions of inherent structures are available as Supporting Information.

(1) Context-Dependent Propensities Show an Increased Preference for (ϕ,ψ) Values That Generate Extended Conformations. In Figure 4, we plot the changes in cumulative propensities as a function of peptide length, *L*, for the lower and upper left quadrants of (ϕ,ψ) space, respectively. The lower and upper left quadrants are defined as follows: In both cases ϕ is either less than 0° or lies in the interval 150° $\leq \phi \leq$ 180°. For the lower left quadrant $-150° \leq \psi \leq 0°$; for the upper left quadrant ψ is either greater than 0° or $-180° \leq \psi < -150°$. We plot the context-dependent propensities as a function of peptide length for each of the host types. Specifically, we plot $\Delta_{i,i+1}$ defined as the difference between the propensity for a quadrant in a host peptide of length $L = L_{i+1}$ and the propensity for the same quadrant in a host of length $L = L_i$. *L* refers to the number of host residues on one side of the guest. If $\Delta_{i,i+1} \approx 0$, then the propensities do not change upon adding two flanking host residues. The first row of plots in Figure 4 shows that there is a systematic decrease in the cumulative propensities for the lower left quadrant of (ϕ,ψ) space. This decrease is offset by a concomitant increase in the propensity for the upper left quadrant as shown by the second row of plots in Figure 4. The increase in propensity for the upper left quadrant is consistent with previous calculations (*66*, *86*) and observations from a cross section of experiments on peptides and denatured proteins (*37*, *87*−*89*).

(2) The Effect of Host Context on Conformational Propensities Varies with Host Type. Glycyl hosts have the smallest effect on conformational propensities of the guest. Conversely, guest propensities change the most in prolyl hosts. The effects of alanine, phenylalanine, and valine hosts are between those of glycine and proline. The observation of minimal perturbations due to glycyl hosts is consistent with the recent NMR data of Chen et al. (*41*). They showed that flanking glycines do not have a major effect on the conformational propensities of the central alanine in a GGAGG pentapeptide.

(3) A Steric Segment Is Five Residues Long. Additions of host residues beyond those in blocked pentapeptides do not change the propensities of guest residues. Since our potentials are based purely on excluded volume, we refer to pentapeptide host−guest systems as steric segments. This observation is supported by the fact that $\Delta_{i,i+1} \approx 0$ for all host sequences shown in Figure 4 when $L > 2$ in *N*-acetyl-$(haa)_L$-Xaa-$(haa)_L$-*N*′-methylamide. $\Delta_{i,i+1}$ is the difference in propensity of a guest residue in a host of length $L + 1$ and the corresponding propensity in a host of length *L*. For polyalanine hosts, the dominant context-dependent interactions are $i \rightarrow i + 4$ backbone−side chain contacts between alanine residues in the host sequence. In polyvaline and polyphenylalanine hosts, there are additional ($i \rightarrow i + 2$) side chain−backbone and side chain−side chain contacts between the host residues. The effect of proline on guest residues dominates with just one proline on either side of a guest residue. This is because of side chain−backbone contacts involving the δ-carbon of the proline immediately C-terminal to the guest and the β-carbon of the guest residue (*92*). Results shown in Figure 4, which estimate the length of a steric segment to be five amino acids, are consistent with suggestions based on recent work from other groups (*86*, *90*, *91*).

(4) The Calculated Intrinsic and Context-Dependent P_{II} Propensities Shown in Figure 5 Agree with Trends Measured by UV−CD Spectroscopy. As noted earlier, measurement of P_{II} propensities have taken on special significance in light of renewed interest in the hypothesis of Tiffany and Krimm (*47*−*50*). From measurements in polyproline hosts, Rucker et al. (*40*) conclude that the β-branched amino acids, especially valine and isoleucine, have the lowest P_{II} propensity. Similarly, Rucker et al. found that amino acids with long straight side chains such as lysine, arginine, glutamic acid, and glutamine have high propensities for P_{II}. Our results, shown in Figure 5, are in general agreement with those of Rucker et al. The major disagreements pertain to glycine and leucine. We disagree with their assignment of a high P_{II} propensity to glycine: Since glycine is achiral and not detectable by UV-CD, the measurements of Rucker et al. correspond to two triproline fragments that are in P_{II}.

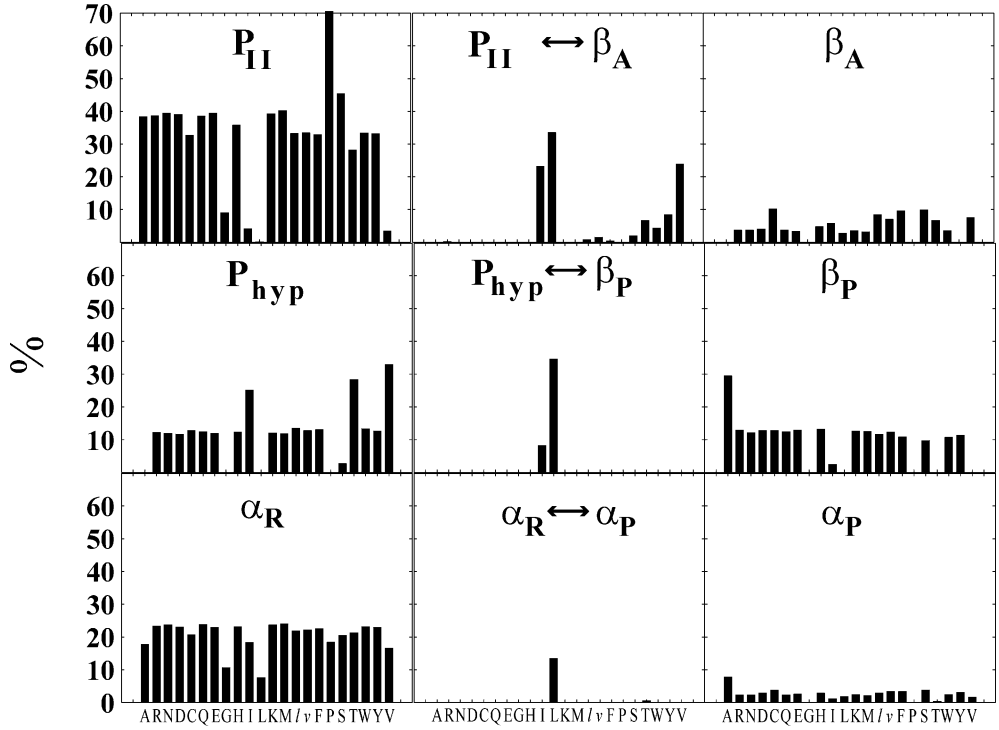

FIGURE 3: Bar plots of intrinsic propensities for all 22 amino acids. Single-letter codes for the nonnatural amino acids norvaline (*v*) and norleucine (*l*) are italicized.

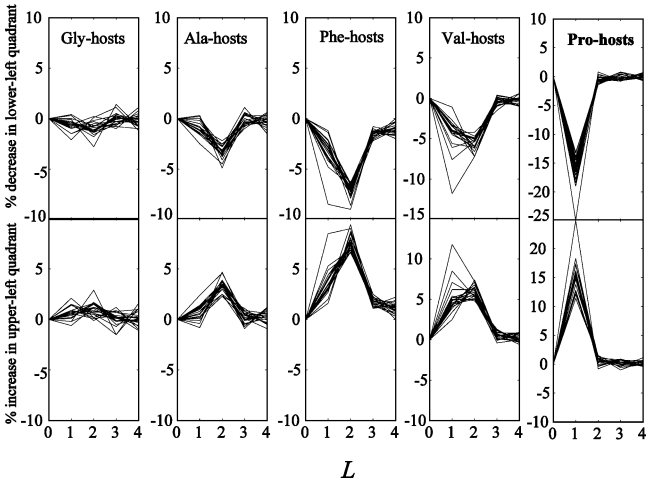

FIGURE 4: Change in cumulative propensities for lower and upper left quadrants as a function of increasing length L of host−guest peptides. Each panel corresponds to a specific host residue. Each curve within a panel represents 1 of 22 guest residues. The host−guest peptides are of the form *N*-acetyl-(haa)$_L$-Xaa-(haa)$_L$-*N*′-methylamide, where haa refers to the host residue and Xaa to the guest. In each panel we plot $\Delta_{i,i+1}(L)$ as a function of L. $\Delta_{i,i+1}$ is the difference in propensity of a guest residue in a host of length $L + 1$ and the corresponding propensity in a host of length L. Definitions of quadrants are provided in the text. The top row shows decreases in cumulative propensities for the lower left quadrant. The bottom row shows concomitant increases in cumulative propensities for the upper left quadrant.

Interestingly, their conclusions for leucine disagree with those of Eker et al. (*43*). We have already established leucine's unique preference for bridge regions, a result that agrees with the findings of Eker et al.

(5) Surprisingly, We Predict That Proline Hosts Diminish the P_{II} Propensities for Many Guest Residues. Prolines immediately C-terminal to guest residues force the latter to adopt positive ψ angles (*92*). Additionally, we find that there is a narrowing in the width of the basin surrounding P_{II}-like inherent structures. This is best illustrated in Figure 6 where we compare the energy landscape for alanine dipeptide to that of alanine in an *N*-acetyl-Pro-Ala-Pro-*N*′-methylamide peptide. The number of low-energy contours surrounding the P_{II} minimum decreases due to the effect of the C-terminal proline (Figure 6B). In addition, the P_{II} conformation ceases to be the global minimum for alanine in the peptide with flanking prolines. This leads to decreases in P_{II} propensities. The degree of decrease depends on the identity of the guest residue. In prolyl hosts, the increase in cumulative propensities for the upper left quadrant, shown in Figure 4, translates to an increase in the propensity for β_P (see Supporting Information).

Summary of Results for Host−Guest Peptides. We have shown that analysis of conformational distributions in the excluded-volume limit yields results that agree with the experimentally observed trends for conformational propensities of short peptides. We identify distinct preferences of short peptides for local conformations such as P_{II}, β, and α_R. In the excluded-volume limit, given the constraints of covalent geometry and the planarity of peptide units, locally extended conformations are the ones that are most effective at minimizing steric overlap. Which of these conformations is preferred will depend primarily on amino acid type and secondarily on the local sequence context. Furthermore, the precise probabilities for these conformations will depend on solution conditions (*52*). The latter determine the degree to which peptide−solvent interactions counterbalance gas-phase, intrapeptide electrostatic effects (*84*). The preceding discussion completes the first half of our solution to the reconciliation problem where the goal was to reproduce experimental observations of conformational specificity.

Solving the Second Half of the Reconciliation Problem for Denatured Proteins. How do polymeric quantities such

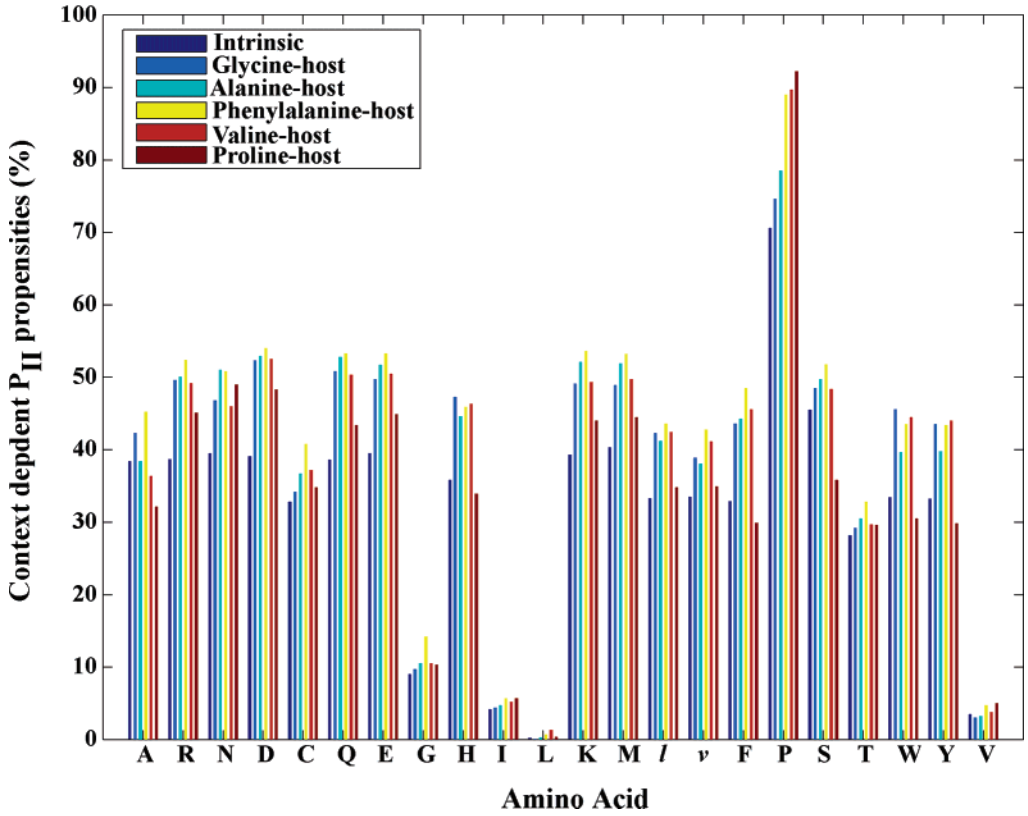

FIGURE 5: Intrinsic and context-dependent P_{II} propensities for all 22 amino acids. Context-dependent propensities are shown for host−guest peptides of the form *N*-acetyl-(haa)$_2$-Xaa-(haa)$_2$-*N*′-methylamide.

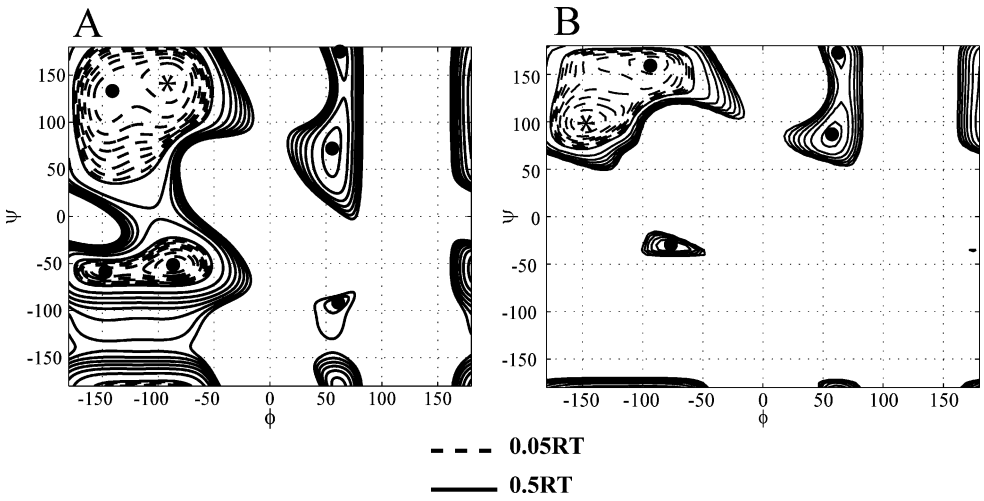

FIGURE 6: Inverse power potential energy surface for alanine dipeptide (A) and alanine in the context of two flanking prolines (B). In both plots, positions of local minima are marked using solid circles. The global minimum is marked using a (+) sign. The dashed contours are in 0.05*RT* intervals from the global minimum. The solid contours are in intervals of 0.5*RT* from the global minimum. Here $RT = 0.59$ kcal/mol. To generate the plot in panel A, proline residues were fixed in the P_{II} conformation, $(\phi,\psi) = (-79°, 145°)$, and the energy surface is generated by systematic conformational enumeration on a 5° × 5° grid.

as R_g scale with chain length for polypeptides in the excluded-volume limit? We find scaling behavior that is consistent with predictions for polymers in good solvents. We are also able to recapitulate the scaling data measured by SAXS by Plaxco and co-workers for chemically denatured proteins.

Scaling of R_g with Chain Length. Ensemble averages for R_g are computed from data generated using MMC simulations. The chains studied are of the form *N*-acetyl-Xaa$_N$-*N*′-methylamide. Here, Xaa is either glycine or alanine. For both classes of homopolymers we performed simulations for chains of length *N* ranging from 50 to 500.

In Figure 7, we show plots of our results for $\ln(R_g)$ versus $\ln(N)$ for polyglycine and polyalanine. Scaling laws for R_g take the form $R_g = R_o N^\nu$ or $\ln(R_g) = \ln(R_o) + \nu \ln(N)$. The parameter R_o is used to quantify the average volume excluded by each residue for favorable interactions with the surrounding solvent and is a measure of the packing density. The precise value for R_o depends on the type of solvent and the details of chemical structure. The upper bound on R_o is reproduced for chains in the excluded-volume limit.

In a good solvent, the exponent ν is universal. Flory's mean-field theory predicts that $\nu = 0.6$ (*55*, *58*). Rigorous renormalization group theories predict that $\nu = 0.5885$ for

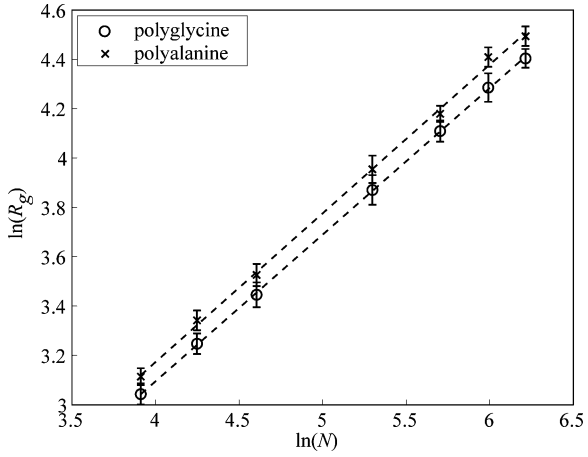

FIGURE 7: Plots of ln(R_g) with ln(N), where N is the chain length. Data shown are for two classes of homopolymers: polyglycine and polyalanine. Parameters for lines of best fit are as follows. For polyglycine, the scaling exponent $\nu = 0.60 \pm 0.01$ and $R_o = 2.0 \pm 0.01$, and for polyalanine, $\nu = 0.61 \pm 0.01$ and $R_o = 2.1 \pm 0.01$.

linear chains such as polyethylene (*59*, *60*). Similar results were obtained from enumeration of self-avoiding walks (*61*). Experimental data for synthetic homopolymers and denatured proteins yield exponents that lie in the interval [0.58, 0.62] (*26*, *57*, *58*, *93*).

We obtain the following parameters from our calculations (Figure 7). For polyglycine, $\nu = 0.60 \pm 0.01$ and $R_o = 2.0 \pm 0.01$. For polyalanine $\nu = 0.61 \pm 0.01$ and $R_o = 2.1 \pm 0.01$. Norms of residuals between the raw data and lines of best fit are 0.075 for polyglycine and 0.071 for polyalanine, respectively.

Testing Good Solvent Scaling for Chemically Denatured Proteins. Plaxco and co-workers (*25*, *26*) have used SAXS data to quantify the scaling behavior for 28 chemically denatured proteins. These proteins do not have prosthetic groups and are free of disulfide bonds, and their glycine and proline contents are small (*26*). They obtained a best-fit scaling exponent of $\nu = 0.598 \pm 0.028$ and an intercept of $R_o = 1.983 \pm 0.1$ (*26*). We generated conformational ensembles in the excluded-volume limit for the proteins studied by Plaxco and co-workers. The list of proteins studied and a tabulation of calculated ensemble average values for R_g are available from Table 3.

Figure 8 shows a log−log plot of our calculated R_g values versus chain length for the proteins listed in Table 3. We obtain a best-fit scaling exponent of 0.62 ± 0.01 and $R_o = 2.08 \pm 0.02$. Our estimates for the scaling exponent and the intercept R_o agree with those from experiment. The small discrepancies in scaling exponent reflect finite size artifacts since, i.e., the field-theoretic exponent of 0.5885 is valid only in the asymptotic limit of infinitely long chains (*60*). Differences in the calculated and measured values for the intercept R_o are due to the effects of preferential interactions with cosolutes that distinguish real systems from the excluded-volume limit (*94*−*97*). It should be stressed that the excluded-volume limit mimics the behavior of a chain in an "ideal solvent". The surprisingly favorable comparison between R_o values from experimental data and calculations suggests that denaturing environments are close facsimiles of ideal solvent conditions. Stated differently, the experimentally observed packing densities for denatured proteins agree with estimates based purely on excluded-volume considerations. The take home message is that under denaturing conditions proteins are loosely packed, swollen coils.

Table 3: List of Sequences Used for Demonstration of Good Solvent Scaling[a]

ID	no. of residues	R_g (Å)	protein/peptide
1	8	7.9 ± 0.04	angiotensin II
2	16	11.9 ± 0.1	AK-16
3	27	16.1 ± 0.2	AK-27
4	32	18.0 ± 0.2	AK-32
5	37	19.6± 0.4	AK-37
6	40	22.4 ± 0.4	N-terminal cytochrome *c*
7	56	27.2 ± 0.3	protein G
8	59	28.0 ± 0.4	SH3 domain of Fyn
9	67	30.0 ± 0.3	cold shock protein
10	76	33.8 ± 0.7	ubiquitin
11	87	36.5 ± 0.7	protein L
12	90	37.5 ± 0.6	barstar
13	92	37.3 ± 0.8	λ repressor
14	98	38.6 ± 0.7	acylphosphatase
15	104	39.6 ± 0.6	horse cytochrome *c*
16	112	42.6 ± 0.8	SH2 domain of PI 3-kinase
17	124	45.3 ± 0.6	ribonuclease
18	129	46.2 ± 0.7	CheY
19	129	45.6 ± 0.7	lysozyme (partially reduced)
20	131	46.5 ± 1.1	intestinal fatty acid binding protein
21	149	50.6 ± 0.7	staphylococcal nuclease
22	154	51.2 ± 0.9	apomyoglobin
23	167	55.6 ± 1.1	dihydrofolate reductase
24	257	70.5 ± 0.9	outer surface protein A
25	260	71.4 ± 1.6	bovine carbonic anhydrase II
26	268	72.8 ± 0.7	tryptophan synthase α chain
27	380	89.1 ± 2.1	creatine kinase
28	416	93.3 ± 2.2	yeast phosphoglycerate kinase
29	549	108.5 ± 1.9	groEL

[a] Data from this table are plotted in Figure 9. The AK peptides are synthetic peptides of Kohn et al. (*26*) with sequence $(AAKAA)_nGY$, where $n = 5$, 6, or 7 (*26*).

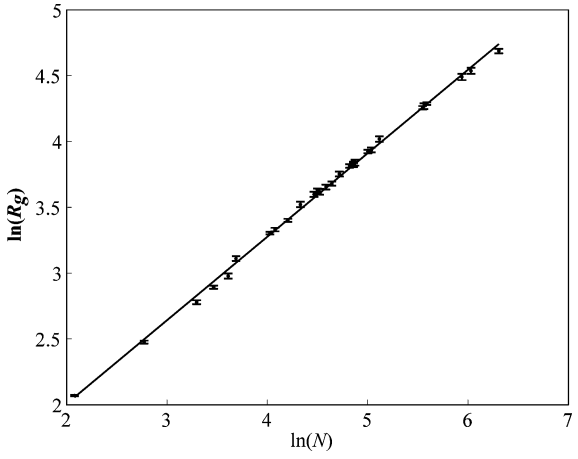

FIGURE 8: Plots of ln(R_g) with ln(N), where N is the chain length for proteins listed in Table 3. Raw data used to make the plots are shown in Table 3. Parameters for lines of best fit are as follows. The scaling exponent $\nu = 0.62 \pm 0.01$ and $R_o = 2.08 \pm 0.02$.

Why Do We Obtain Good Agreement between Theory and Experiment for the Scaling Behavior of Chemically Denatured Proteins? The answer lies in the observation that even for short protein chains ($N \approx 10$) the R_g values scale with chain length like polymers in a good solvent (*26*). The dogma is that scaling behavior is typically valid for chain lengths that are at least an order of magnitude longer than the

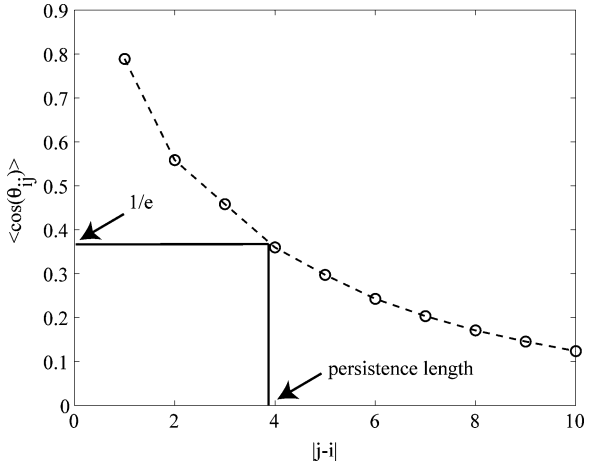

FIGURE 9: Illustration of the method used to calculate $\langle N_p \rangle$. Shown here is the analysis of the data extracted from the conformational ensembles generated via Metropolis Monte Carlo simulations of a 56-residue protein, protein G, studied in the excluded-volume limit. $\langle N_p \rangle$ corresponds to the number of distance in sequence space past which the correlations between bond vectors (see text) decay to 1/e.

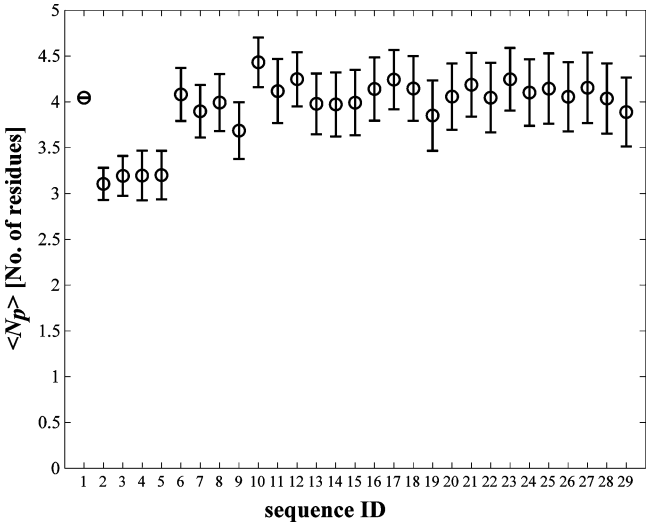

FIGURE 10: Calculated persistence lengths in terms of $\langle N_p \rangle$ for proteins listed in Table 3. The numeric identities are those shown in column 1 of Table 3.

persistence length, l_p (*62*). Favorable comparison to experimental data must therefore arise from the close correspondence between the estimate for l_p in the excluded-volume limit and the true value for l_p in denaturing environments. This quantity has not been measured to date for denatured proteins, so our estimate for l_p should be marked as a prediction from our calculations.

To calculate persistence lengths, we use the definition first given by Porod and described in detail by Öttinger and coworkers (*98*). For a polypeptide chain of length N, the persistence length is $l_p = 1/(N-1)\sum_{i=1}^{N-1}(1/l_i)\sum_{j=1}^{N-1}(\mathbf{l}_i \cdot \mathbf{l}_j)$. Here, $\mathbf{l}_i$ refers to the "bond vector" between α-carbon atoms i and $i+1$. The length of this vector is l_i, and for polypeptides $l = l_i = l_j = 3.8$ Å. We are interested in the ensemble average value of l_p, and this is obtained by computing the average l_p for conformations generated in the equilibrium ensemble. Alternatively, l_p can be estimated by plotting the ensemble average of $\langle |(\mathbf{l}_i \cdot \mathbf{l}_j)/l^2| \rangle = \langle |\cos \theta_{ij}| \rangle$ as a function of $|j-i|$, the distance in sequence space. When the bonds are highly correlated, $\langle |\cos \theta_{ij}| \rangle = 1$. As, $|j-i|$ increases, the correlations decay exponentially. The value of $|j-i|$ for which $\langle |\cos \theta_{ij}| \rangle \approx 1/e$ is the persistence length in terms of the number of residues, $\langle N_p \rangle$. Figure 9 shows a sample plot of $\langle |\cos \theta_{ij}| \rangle$ versus $|j-i|$ for protein G. This plot also illustrates the method used to extract an estimate for $\langle N_p \rangle$.

In Figure 10, we plot the persistence length in terms of $\langle N_p \rangle$ for each of the proteins listed in Table 3. We find that, in the excluded-volume limit, $3 \leq \langle N_p \rangle \leq 5$ for all sequences studied. We obtain similar results for the polyglycine and polyalanine homopolymers (data not shown). The estimate for $\langle N_p \rangle$ is consistent with the estimates for the length of a steric segment from our studies of conformational propensities in host−guest peptides shown in Figure 4.

Solution to the Reconciliation Problem. Resolution of the reconciliation paradox can be clearly stated in light of our predictions for the value of $\langle N_p \rangle$. Conformational ensembles generated in the excluded-volume limit provide an accurate mimic of the ensemble available to proteins under highly denaturing conditions. In the excluded-volume limit, short peptides prefer locally extended conformations, and these preferences are consistent with experimental observations. In particular, we have shown that it is possible to reproduce the segmental preference for conformations such as P_{II}, β-strands, and single turns of α-helices. Local stretching is a consequence of the minimization of steric overlap subject to the constraints of maintaining chain connectivity. As chain lengths increase, conformational entropy opposes the effect of stretching. In the absence of entropic counterbalancing, denatured states would assume rodlike configurations, and R_g would scale linearly with chain length. The tug of war between conformational entropy and local stretching leads to swollen coils and the experimentally observed scaling behavior for polymeric quantities such as R_g with chain length N. We predict that local preferences in denatured proteins extend over segments that are no longer than five amino acids. Simply stated, denatured proteins are not stiff, rodlike chains. Interestingly, when proteins fold by collapsing to form compact globules, l_p and N_p actually decrease (Tran and Pappu, manuscript in preparation). This observation, although underappreciated, is consistent with predictions from polymer theory (*63*) and atomic force microscopy measurements (*99*).

We find preferences for P_{II}-like geometries in short sequence stretches. Within these stretches, the preference for P_{II}-like geometries depends on the specific amino acid sequence. These observations are consistent with the hypothesis of Tiffany and Krimm (*48*, *49*), which has been the subject of renewed interest (*22*, *23*, *27*, *29*, *33*, *36*−*40*, *66*). However, we argue, given previous results from our group (*84*), that the preference for P_{II}-like geometries is really a default characteristic of most amino acid sequences in the excluded-volume limit.

Our work leads to a confounding observation that denaturants (8 M urea or 6 M GdnCl) appear to behave like ideal solvents, not just good solvents. Support for this observation derives from the fact that we are able to mimic ensembles accessible to proteins and polypeptides in harshly denaturing environments by considering the behavior of these polymers in the excluded-volume limit alone. Not only do we obtain close agreement for the scaling exponent, but we also obtain

close agreement for the intercept R_o. As noted earlier, R_o can be used to quantify packing densities under denaturing conditions. The question to pursue is as follows: *How does excluded volume alone become a dominant determinant of conformational preferences in denaturing environments?* Investigation of how the balance of preferential interactions and specificity in chain−chain interactions can lead precisely to the excluded-volume limit ought to lead to an improved understanding of the forces that govern protein stability and specificity.

CONCLUSIONS

We have proposed a solution to the reconciliation problem for denatured proteins. The solution lies in the observation that the excluded-volume limit provides an accurate mimic of ensembles accessible to denatured proteins. In the excluded-volume limit, sequence-specific conformational preferences are exceedingly local. The persistence of sequence-specific conformational preferences is limited to segments that are less than five residues long. Therefore, conformational entropy opposes propagation of local segmental preferences. The balance between local stretching and conformational entropy leads to the observed behavior of polymeric quantities for denatured proteins.

Fitzkee and Rose (*100*) recently showed that their model for denatured proteins also reproduces good solvent scaling. They modeled denatured proteins as self-avoiding walks of rigid elements of native structure. They were able to reproduce good solvent scaling because they only include excluded-volume effects in their model. However, the local structure in the Fitzkee and Rose model is entirely artificial. Conversely, we are able to use a single atomistic model and reproduce both local propensities and scaling behavior of denatured proteins.

NOTE ADDED IN PROOF

The computer code developed by the authors to carry out the calculations discussed in this work is available from the corresponding author upon request.

ACKNOWLEDGMENT

The authors are grateful to Nathan Baker, Trevor Creamer, Ken Dill, Elliot Elson, Alan Grossfield, and Gary Pielak for helpful discussions and critical comments about the manuscript.

SUPPORTING INFORMATION AVAILABLE

Tables of parameters for inverse power potentials and conformational propensities (intrinsic and context-dependent) for all 22 amino acids. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation, *Adv. Protein Chem. 14*, 1−63.
2. Tanford, C. (1968) Protein denaturation, *Adv. Protein Chem. 23*, 121−282.
3. Tanford, C. (1970) Protein denaturation: Part C. Theoretical models for the mechanism of denaturation, *Adv. Protein Chem. 24*, 1−95.
4. Dill, K. A., and Shortle, D. (1991) Denatured states of proteins, *Annu. Rev. Biochem. 60*, 795−825.
5. Shortle, D., and Ackerman, M. S. (2001) Persistence of native-like topology in a denatured protein in 8M urea, *Science 293*, 487−489.
6. Hodsdon, M. E., and Frieden, C. (2001) Intestinal fatty acid binding protein: The folding mechanism as determined by NMR studies, *Biochemistry 40*, 732−742.
7. van Gunsteren, W. F., Burgi, R., Peter, C., and Daura, X. (2001) The key to solving the protein-folding problem lies in an accurate description of the denatured state, *Angew. Chem., Int. Ed. Engl. 40*, 351−355.
8. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-range interactions within a nonnative protein, *Science 295*, 1719−1722.
9. Shortle, D. (2002) The expanded denatured state: An ensemble of conformations trapped in a locally encoded topological space, *Adv. Protein Chem. 62*, 1−23.
10. Dyson, H. J., and Wright, P. E. (2002) Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance, *Adv. Protein Chem. 62*, 311−340.
11. Myers, J. M., and Oas, T. G. (2002) Mechanism of fast protein folding, *Annu. Rev. Biochem. 71*, 783−815.
12. Zagrovic, B., and Pande, V. S. (2003) Structural correspondence between the α-helix and the random-flight chain resolves how unfolded proteins can have nativelike properties, *Nat. Struct. Biol. 10*, 955−961.
13. Mohana-Borges, R., Goto, N. K., Kroon, G. J. A., Dyson, H. J., and Wright, P. E. (2004) Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings, *J. Mol. Biol. 340*, 1131−1142.
14. Uversky, V. N., and Fink, A. L. (2004) Conformational constraints for amyloid formation: the importance of being unfolded, *Biochim. Biophys. Acta 1698*, 131−153.
15. Minton, A. P. (2000) Implications of macromolecular crowding for protein assembly, *Curr. Opin. Struct. Biol. 10*, 13−15.
16. Ghaemmaghami, S., and Oas, T. G. (2001) Quantitative protein stability measurement *in vivo*, *Nat. Struct. Biol. 8*, 879−882.
17. Dedmon, M. M., Patel, C. N., Young, G. B., and Pielak, G. J. (2002) FlgM gains structure in living cells, *Proc. Natl. Acad. Sci. U.S.A. 99*, 12681−12684.
18. Plaxco, K. W., and Gross, M. (1997) The importance of being unfolded, *Nature 386*, 657.
19. Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: Re-assessing the protein structure−function paradigm, *J. Mol. Biol. 293*, 321−331.
20. Dunker, A. K., Brown, C. J., and Obradovic, Z. (2002) Identification and functions of usefully disordered proteins, *Adv. Protein Chem. 62*, 25−49.
21. Uversky, V. N. (2002) Natively unfolded proteins: A point where biology waits for physics, *Protein Sci. 11*, 739−756.
22. Rose, G. D. (2002) Getting to know U, *Adv. Protein Chem. 62*, xv−xxi.
23. Baldwin, R. L. (2002) A new perspective on unfolded proteins, *Adv. Protein Chem. 62*, 361−366.
24. Doniach, S. (2001) Changes in biomolecular conformation seen by small-angle X-ray scattering, *Chem. Rev. 101*, 1763−1778.
25. Millett, I. S., Doniach, S., and Plaxco, K. W. (2002) Toward a taxonomy of the denatured state: Small angle scattering studies of unfolded proteins, *Adv. Protein Chem. 62*, 241−262.
26. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., Sosnick, T. R., Hasan, M. Z., Pande, V. S., Ruczinski, I., Doniach, S., and Plaxco, K. W. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins, *Proc. Natl. Acad. Sci. U.S.A. 101*, 12491−12496.
27. Woody, R. W. (1992) Circular dichroism of unordered polypeptides, *Adv. Biophys. Chem. 2*, 37−79.
28. Spector, S., Rosconi, M., and Raleigh, D. P. (1999) Conformational analysis of peptide fragments derived from the peripheral subunit binding domain from the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophillus*: Evidence for nonrandom structure in the unfolded state, *Biopolymers 49*, 29−40.
29. Woutersen, S., and Hamm, P. (2000) Structure determination of trialanine in water using polarization sensitive two-dimensional vibrational spectroscopy, *J. Phys. Chem. B 104*, 11316−11320.
30. Poon, C.-D., Samulski, E. T., Weise, C. F., and Weishaar, J. C. (2000) Do bridging water molecules dictate the structure of a model dipeptide in aqueous solution?, *J. Am. Chem. Soc. 122*, 5642−5643.

31. Schweitzer-Stenner, R., Eker, F., Huang, Q., and Griebenow, K. (2001) Dihedral angles of trialanine in D_2O determined by combining FTIR and polarized visible Raman spectroscopy, *J. Am. Chem. Soc. 123*, 9628−9633.
32. Shi, Z., Olson, C. A., Rose, G. D., Baldwin, R. L., and Kallenbach, N. R. (2002) Polyproline II structure in a sequence of seven alanine residues, *Proc. Natl. Acad. Sci. U.S.A. 99*, 9190−9195.
33. Shi, Z., Woody, R. W., and Kallenbach, N. R. (2002) Is polyproline II a major backbone conformation in unfolded proteins?, *Adv. Protein Chem. 62*, 163−240.
34. Barron, L. D., Blanch, E. W., and Hecht, L. (2002) Unfolded proteins studied by Raman optical activity, *Adv. Protein Chem. 62*, 51−90.
35. Keiderling, T. A., and Xu, Q. (2002) Unfolded peptides and proteins studied with infrared absorption and vibrational circular dichroism spectra, *Adv. Protein Chem. 62*, 111−161.
36. Eker, F., Griebenow, K., and Schweitzer-Stenner, R. (2003) Stable conformations of peptides in solution studied using UV circular dichroism spectroscopy, *J. Am. Chem. Soc. 125*, 8178−8185.
37. Ding, L., Chen, K., Santini, P. A., Shi, Z. S., and Kallenbach, N. R. (2003) The pentapeptide GGAGG has P_{II} conformation, *J. Am. Chem. Soc. 125*, 8092−8093.
38. Ferreon, J. C., and Hilser, V. J. (2003) The effect of polyproline II structure on denatured state entropy, *Protein Sci. 12*, 447−457.
39. Rucker, A. L., and Creamer, T. P. (2002) Polyproline II helical structure in protein unfolded states: Lysine peptides revisited, *Protein Sci. 11*, 980−985.
40. Rucker, A. L., Pager, C. T., Campbell, M. N., Qualls, J. E., and Creamer, T. P. (2003) Host−guest scale of left-handed polyproline II helix formation, *Proteins: Struct., Funct., Genet. 53*, 68−75.
41. Chen, K, Liu, Z. G., and Kallenbach, N. R. (2004) The polyproline II conformation in short alanine peptides is noncooperative, *Proc. Natl. Acad. Sci. U.S.A. 101*, 15352−15357.
42. Chellgren, B. W., and Creamer, T. P. (2004) Short sequences of non-proline residues can adopt the polyproline II helical conformation, *Biochemistry 43*, 5864−5869.
43. Eker, F., Griebenow, K., Cao, X. L., Nafie L. A., and Schweitzer-Stenner, R. (2004) Preferred peptide backbone conformations in the unfolded state revealed by the structure analysis of alanine-based AXA tripeptides in aqueous solution, *Proc. Natl. Acad. Sci. U.S.A. 101*, 10054−10059.
44. Ferreon, J. C., and Hilser, V. J. (2004) Thermodynamics of binding to SH3 domains: The energetic impact of polyproline II (P_{II}) helix formation, *Biochemistry 43*, 7787−7797.
45. Shortle, D. (1996) Structural analysis of non-native states of proteins by NMR, *Curr. Opin. Struct. Biol. 6*, 24−30.
46. Fieber, W., Kragelund, B. B., Meldal, M., and Poulsen, F. M. (2005) Reversible denaturation of acid-denatured ACBP controlled by helix A4, *Biochemistry 44*, 1375−1384.
47. Tiffany, M. L., and Krimm, S. (1968) Circular dichroism of poly-L-proline in an unordered conformation, *Biopolymers 6*, 1767−1770.
48. Tiffany, M. L., and Krimm, S. (1969) Circular dichroism of the "random" polypeptide chain, *Biopolymers 8*, 347−359.
49. Tiffany, M. L., and Krimm, S. (1973) Extended conformations of polypeptides and proteins in urea and guanidine hydrochloride, *Biopolymers 12*, 575−587.
50. Krimm, S., and Tiffany, M. L. (1974) The circular dichroism spectrum and structure of unordered polypeptides and proteins, *Isr. J. Chem. 12*, 189−200.
51. Miller, W. G., and Goebel, C. V. (1968) Dimensions of protein random coils, *Biochemistry 7*, 3925−3935.
52. Liu, Z., Chen, K., Ng, A., Shi, Z., Woody, R. W., and Kallenbach, N. R. (2004) Solvent dependence of P_{II} conformation in model alanine peptides, *J. Am. Chem. Soc. 126*, 15141−15150.
53. Whittington, S. J., Chellgren, B. W., Hermann, V. M., and Creamer, T. P. (2005) Urea promotes polyproline II helix formation: Implications for protein denatured states, *Biochemistry 44*, 6269−6275.
54. Fleming, P. J., and Rose, G. D. (2005) Conformational properties of unfolded proteins, in *Protein folding handbook* (Buchner, J., and Kiefhaber, T., Eds.) Vol. 2, Part 1, Chapter 20, pp 710−736, Wiley-VCH, Weinheim.
55. Flory, P. J. (1953) *Principles of polymer chemistry*, Cornell University Press, Ithaca, NY.
56. Chan, H. S., and Dill, K. A. (1991) Polymer principles in protein structure and stability, *Annu. Rev. Biophys. Biophys. Chem. 20*, 447−490.
57. Grosberg, A. Y., and Khokhlov, A. R. (1994) Statistical physics of macromolecules, *AIP series in polymers and complex materials*, AIP Press, New York.
58. Rubenstein, M., and Colby, R. H. (2003) *Polymer Physics*, Oxford University Press, Oxford.
59. Le Guillou, J. C., and Zinn-Justin, J. (1980) Critical exponents from field theory, *Phys. Rev. B 21*, 3976−3998.
60. Schäfer, L. (1984) Unified description of temperature and concentration crossover in the excluded volume problem. I. Osmotic pressure and correlation lengths, *Macromolecules 17*, 1357−1370.
61. Li, B., Madras, N., and Sokal, A. D. (1995) Critical exponents, hyperscaling and universal amplitude ratios for two- and three-dimensional self-avoiding walks, *J. Stat. Phys. 80*, 661−754.
62. Schäfer, L. (1999) *Excluded volume effects in polymer solutions as explained by the renormalization group*, Springer, Berlin.
63. deGennes, P.-G. (1979) *Scaling concepts in polymer physics*, Cornell University Press, Ithaca and London.
64. Flory, P. J. (1969) *Statistical mechanics of chain molecules*, Chapter 7, Hanser Publishers, Munich.
65. Richards, F. M. (1977) Areas, volumes, packing and protein structure, *Annu. Rev. Biophys. Bioeng. 6*, 151−176.
66. Pappu, R. V., and Rose, G. D. (2002) A simple model for polyproline II structure in unfolded states of alanine-based peptides, *Protein Sci. 11*, 2437−2455.
67. Hoover, W. G., Gray, S. G., and Johnson, K. W. (1971) Thermodynamic properties of the fluid and solid phases for inverse power potentials, *J. Chem. Phys. 55*, 1128−1136.
68. Hopfinger, A. J. (1973) *Conformational properties of macromolecules*, Academic Press, New York.
69. Pauling, L. (1970) Chapter 11, in *General Chemistry*, 3rd ed., W. H. Freeman Press, San Francisco, CA.
70. Slater, J. C., and Kirkwood, J. G. (1931) The van der Waals forces in gases, *Phys. Rev. 37*, 682−696.
71. Venkatachalam, C. M., Price, B. J., and Krimm, S. (1974) Theoretical analysis of pyrrolidine ring puckering and the conformational energies of proline and 5-methylproline dimers, *Macromolecules 7*, 212−220.
72. Vitagliano, L., Berisio, R., Mastrangelo, A., Mazzarella, L., and Zagari A. (2001) Preferred proline puckerings in *cis* and *trans* peptide groups: implications for collagen stability, *Protein Sci. 10*, 2627−2632.
73. Benzi, C., Improta, R., Scalmani, G., and Barone, V. (2002) Quantum mechanical study of the conformational behavior of proline and 4R-hydroxyproline dipeptide analogues in a vacuum and in aqueous solution, *J. Comput. Chem. 23*, 341−350.
74. Allen, W. D., Czinki, E., and Csaszar, A. G. (2004) Molecular structure of proline, *Chem. Eur. J. 10*, 4512−4517.
75. Lovell, S. C., Davis, I. W., Arendall, W. B., III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation, *Proteins: Struct., Funct., Genet. 50*, 437−450.
76. Engh, R. A., and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement, *Acta Crystallogr. 47*, 392−400.
77. Frenkel, D., and Smit, B. (2002) Chapter 3, in *Understanding molecular simulation*, Academic Press, London, U.K.
78. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953) Equation of state calculations by fast computing machines, *J. Chem. Phys. 21*, 1087−1092.
79. Binder, K. (1995) *In Monte Carlo and Molecular Dynamics Simulations in Polymer Science* (Binder, K., Ed.) Chapter 1, Oxford University Press, Oxford.
80. Stillinger, F. H., and Weber, T. A. (1984) Packing structures and transitions in liquids and solids, *Science* 225, 983−989.
81. Stillinger, F. H., and Weber, T. A. (1985) Inherent structure theory of liquids in the hard-sphere limit, *J. Chem. Phys. 83*, 4767−4775.
82. Luenberger, D. G. (1984) Chapter 9, in *Linear and nonlinear programming*, 2nd ed., Addison-Wesley, Reading, MA.
83. Hu, H., Elstner, M., and Hermans, J. (2003) Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptides" (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution, *Proteins: Struct., Funct., Genet. 50*, 451−463.
84. Drozdov, A. N., Grossfield, A., and Pappu, R. V. (2004) The role of solvent in determining conformational preferences of alanine dipeptide in water, *J. Am. Chem. Soc. 126*, 2574−2581.

85. Brahmachari, S. K., Bansal, M., Ananthanarayanan, V. S., and Sasisekharan, V. (1979) Structural investigation on poly(4-hydroxyl-L-proline). 2. Physicochemical studies, *Macromolecules*, *12*, 23−28.
86. Pappu, R. V., Srinivasan, R., and Rose, G. D. (2000) The Flory isolated pair hypothesis is not valid for polypeptide chains: Implications for protein folding, *Proc. Natl. Acad. Sci. U.S.A. 97*, 12565−12570.
87. Penkett, C. J., Redfield, C., Dodd, I., Hubbard, J., McBay, D. L., Mossakowska, D. E., Smith, R. A., Dobson, C. M., and Smith, L. J. (1997) NMR analysis of main-chain conformational preferences in an unfolded fibronectin-binding protein, *J. Mol. Biol. 274*, 152−159.
88. Petrescu, A.-J., Calmettes, P., Durand, D., Receveur, V., and Smith, J. C. (2000) Change in backbone torsion angle distributions on protein folding, *Protein Sci. 9*, 1129−1136.
89. Yang, W. Y., Larios, E., and Gruebele, M. (2003) On the extended β-conformation propensity of polypeptides at high temperature, *J. Am. Chem. Soc. 125*, 16220−16227.
90. Ohkubo, Y. Z., and Brooks, C. L. (2003) Exploring Flory's isolated-pair hypothesis: statistical mechanics of helix-coil transitions in polyalanine and the C-peptide from RNase A, *Proc. Natl. Acad. Sci. U.S.A. 100*, 13916−13921.
91. Kentsis, A., Mezei, M., Gindin, T., and Osman, R. (2004) Unfolded state of polyalanine is a segmented polyproline II helix, *Proteins 55*, 493−501.
92. MacArthur, M. W., and Thornton, J. M. (1991) Influence of proline residues on protein conformation, *J. Mol. Biol. 218*, 397−412.
93. Pedersen, J. S., and Schurtenberger, P. (2004) Scattering functions of semidilute solutions of polymers in a good solvent, *J. Polym. Sci., Part B: Polym. Phys. 42*, 3081−3094.
94. Timasheff, S. N. (2002) Protein hydration, thermodynamic binding, and preferential hydration, *Biochemistry 41*, 13473−13482.
95. Yang, M., Ferreon, A. C. M., and Bolen, D. W. (2000) Structural thermodynamics of a random coil protein in guanidine hydrochloride, *Proteins: Struct., Funct., Genet., Suppl. 4*, 44−49.
96. Pace, C. N., Grimsley, G. R., and Scholtz, J. M. (2005) *Protein folding handbook* (Buchner, J., and Kiefhaber. T., Eds.) Vol. 1, Part 1, Chapter 3, pp 45−69, Wiley-VCH, Winheim.
97. Felitsky, D. J., and Record, M. T. (2004) Application of the local-bulk partitioning and competitive binding models to interpret preferential interactions of glycine betaine and urea with protein surface, *Biochemistry 43*, 9276−9288.
98. Kröger, M., Ramírez, J., and Öttinger, H. C. (2002) Projection from an atomistic chain contour to its primitive path, *Polymer 43*, 477−487.
99. Thompson, J. B., Hansma, H. G., Hansma, P. K., and Plaxco, K. W. (2002) The backbone conformational entropy of protein folding: Experimental measures from atomic force microscopy, *J. Mol. Biol. 322*, 645−652.
100. Fitzkee, N. C., and Rose, G. D. (2004) Reassessing random-coil statistics in unfolded proteins, *Proc. Natl. Acad. Sci. U.S.A. 101*, 12497−12502.

BI050196L