A simple model for polyproline II structure in unfolded states of alanine-based peptides

ROHIT V. PAPPU[1] AND GEORGE D. ROSE[2]

[1]Department of Biomedical Engineering and Center for Computational Biology, Washington University in St. Louis, St. Louis, Missouri 63130, USA
[2]The Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, USA

(RECEIVED May 30, 2002; FINAL REVISION July 16, 2002; ACCEPTED July 17, 2002)

Abstract

The striking similarity between observed circular dichroism spectra of nonprolyl homopolymers and that of regular left-handed polyproline II (P_{II}) helices prompted Tiffany and Krimm to propose in 1968 that unordered peptides and unfolded proteins are built of P_{II} segments linked by sharp bends. A large body of experimental evidence, accumulated over the past three decades, provides compelling evidence in support of the original hypothesis of Tiffany and Krimm. Of particular interest are the recent experiments of Shi et al. who find significant P_{II} structure in a short unfolded alanine-based peptide. What is the physical basis for P_{II} helices in peptide and protein unfolded states? The widely accepted view is that favorable chain-solvent hydrogen bonds lead to a preference for dynamical fluctuations about noncooperative P_{II} helices in water. Is this preference simply a consequence of hydrogen bonding or is it a manifestation of a more general trend for unfolded states which are appropriately viewed as chains in a good solvent? The prevalence of closely packed interiors in folded proteins suggests that under conditions that favor folding, water—which is a better solvent for itself than for any polypeptide chain—expels the chain from its midst, thereby maximizing chain packing. Implicit in this view is a complementary idea: under conditions that favor unfolding, chain-solvent interactions are preferred and in a so-called good solvent, chain packing density is minimized. In this work we show that minimization of chain packing density leads to preferred fluctuations for short polyalanyl chains around canonical, noncooperative P_{II}-like conformations. Minimization of chain packing is modeled using a purely repulsive soft-core potential between polypeptide atoms. Details of chain-solvent interactions are ignored. Remarkably, the simple model captures the essential physics behind the preference of short unfolded alanine-based peptides for P_{II} helices. Our results are based on a detailed analysis of the potential energy landscape which determines the system's structural and thermodynamic preferences. We use the inherent structure formalism of Stillinger and Weber, according to which the energy landscape is partitioned into basins of attraction around local minima. We find that the landscape for the experimentally studied seven-residue alanine-based peptide is dominated by fluctuations about two noncooperative structures: the left-handed polyproline II helix and its symmetry mate.

Keywords: Configurational mapping; energy landscape; polyproline II helices; inherent structures; random-coil; packing density

Reprint requests to: Rohit V. Pappu, Department of Biomedical Engineering and Center for Computational Biology, Washington University in St. Louis, One Brookings Drive, Campus Box 1097, St. Louis, Missouri 63130, USA; e-mail: pappu@biomed.wustl.edu.

Abbreviations: CD, circular dichroism; P_{II}, left-handed polyproline II helix; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; sP_{II}, symmetry mate of left-handed polyproline II helix; T, temperature; U_x, potential energy of conformation x; Z, configurational sum.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.0217402.

Folding reactions for two-state globular proteins are well described within the framework of equilibrium thermodynamics (Ginsburg and Carroll 1965; Anfinsen 1973). Under appropriate physiological conditions, the native state (N) emerges spontaneously and reversibly from the unfolded population (U) (Anfinsen 1973). A comprehensive understanding of the folding reaction requires knowledge of conformers populated in U and N under different conditions

(Kauzmann 1959; Tanford 1968; Dill and Shortle 1991; Lattman and Rose 1993; Serrano 1995; Shortle 1996; O'Connell et al. 1999; Shortle and Ackerman 2001; van Günsteren et al. 2001; Zhou and Dill 2001; Klein-Seetharaman et al. 2002). In addition to their obvious importance in protein folding, unfolded states also appear to be physiologically important (Plaxco and Gross 1997; Dobson 1999; Wright and Dyson 1999; Dunker et al. 2001; Tsai et al. 2001; Uversky 2002).

Twenty-five years ago, Richards (1977) analyzed protein crystal structures and concluded that mean packing densities for the interior of folded proteins are approximately 0.75. The packing density, a dimensionless ratio, measures the degree to which compounds are either liquid-like or solid-like, and the observed mean value of 0.75 resembles that of close-packed spheres of uniform size. Although further refined in later work (Richards and Lim 1994), these early conclusions remain valid descriptions of protein native states. In the Richards paradigm, the central question in protein folding becomes: how do attractive intrachain interactions, subject to excluded volume restrictions and the constraints of chain connectivity, conspire to maximize the packing density?

Here, we explore the complementary idea that polypeptides chains *minimize* their packing density on unfolding. For *two-state* proteins (Jackson 1998) and peptides, we propose that upon unfolding, packing density is minimized, while upon folding, packing density is maximized, subject in both cases to constraints imposed by chain connectivity and excluded volume (Ramachandran and Sasisekharan 1968; Richards 1977).

The term "unfolded state" is used to describe the collection of conformers populated under extreme nonnative conditions, including high temperature, high pressure, extremes of pH, and high concentrations of denaturant (Shortle 1996). An unfolded polypeptide chain is equivalent to a chain in a "good solvent" (Flory 1953; Chan and Dill 1991), and its behavior can be modeled using a purely repulsive potential function (Flory 1953; deGennes 1976; Binder and Heerman 1997). In other words, under good-solvent conditions, chain-solvent interactions are maximized; as a consequence the packing density of the chain around itself is minimized. To capture this effect, we use an inverse power potential (Brant et al. 1967; Hoover et al. 1971; Stillinger and Weber 1985) that treats interatomic interactions as soft repulsions. Conformers favored by this potential are free of steric clashes, and pairwise interatomic distances are maximized with respect to hard-sphere contact distances. The chosen inverse power potential prefers conformers that minimize the covolume (Flory 1969) or shared volume between pairs of chain monomers, subject of course to the constraints imposed by chain connectivity. This behavior satisfies our requirement that chain packing density be minimized.

According to a widely accepted model, unfolded polypeptide chains can explore conformational space freely subject to the constraints imposed by local interactions between nearest neighbor residues (Smith et al. 1996). *To a first approximation*, local restrictions refer to steric effects captured in the well-known Ramachandran map for a dipeptide (Ramachandran et al. 1963; Ramachandran and Sasisekharan 1968). However, in recent work (Pappu et al. 2000), we showed that steric interactions between nonnearest neighbors along a backbone should lead unfolded peptides toward extended conformations in which each (ϕ,ψ)-pair preferentially populates the top left quadrant of a Ramachandran map. All extended conformers are free of steric clashes. However, as shown in this work, only a small fraction of allowed conformers satisfy the additional restriction that the packing density of chain around itself be minimized.

We have applied the idea of minimizing chain packing density to two systems studied in recent experiments, an alanine dipeptide or blocked alanine (Poon et al. 2000; Woutersen and Hamm 2000; Schweitzer-Stenner et al. 2001) and a seven-residue alanine peptide (Shi et al. 2002). In each system, we find that preferred conformers fluctuate about the (ϕ,ψ)-values of canonical left-handed polyproline II (P_{II}) helices and their "symmetry mates" (sP_{II}).

Longer polyalanine peptides "fold" to form stable α-helices in water (Scholtz et al. 1991). In comparison to the *noncooperative* P_{II}-helices (Adzhubei et al. 1987) obtained by minimization of chain packing density, the α-helix is a compact structure that results from *maximizing* chain packing density due to the formation of *cooperative* attractive intrachain hydrogen bonds (Pauling et al. 1951) and van der Waals contacts.

Our results are in general agreement with recent experimental data (Poon et al. 2000; Woutersen and Hamm 2000; Schweitzer-Stenner et al. 2001; Shi et al. 2002). They are also consistent with work that suggests a dominant role for P_{II}-helices in unfolded populations for a variety of systems (Woody 1992) including homopolymers (Holzwarth and Doty 1965; Tiffany and Krimm 1968), peptide fragments excised from glycoproteins (Matsumoto et al. 1983), oligomers that are based on residues other than alanine (Rucker and Creamer 2002), and proteins (Adzhubei et al. 1987; Smyth et al. 2001). Our results for the dipeptide agree with many of the general trends seen in numerous molecular dynamics simulations that include all the details of intrachain and chain-solvent interactions (Rossky et al. 1979; Anderson and Hermans 1988; Pettitt and Karplus 1988; Roterman et al. 1989; Schmidt and Fine 1994; O' Connell et al. 1999; Smith 1999; Tobias and Brooks 1999), although our method of analysis with a simple model uncovers P_{II} as the lowest energy conformation for the dipeptide and the 7-mer. We also find good agreement with quantum mechanical calculations that include the effect of hydrogen bonded water molecules (Grant et al. 1990; Jalkanen and Suhai 1996;

Han et al. 1998). It is clear that our good-solvent model, based on the proposal that peptides unfold by minimizing chain packing, unequivocally captures the *specificity* of short unfolded alanine-based peptides to fluctuate about noncooperative P_{II}-like conformers, in agreement with the recent experiments of Kallenbach and coworkers (Shi et al. 2002).

Results

Conceptual framework

Our goal is to calculate the energies associated with each conceivable conformer in a polyalanine peptide *N*-Acetyl-Ala_ν-*N′*-Methyl-amide, where ν denotes the number of alanine residues. Energies are computed using a purely repulsive inverse power potential (Hoover et al. 1971; Stillinger and Weber 1985) that is akin to the repulsive term of a classical 6–12 Lennard-Jones potential function (see Methods). The only degrees of conformational freedom are the peptide-backbone (ϕ,ψ)-angles. Unique chain conformations are derived by specifying (ϕ,ψ)-values for each pair. The Boltzmann-weighted distribution of energies as a function of temperature, that is, distribution of relative free energies, is analyzed to determine whether fluctuations occur about preferred regions in conformational space.

It is impossible to visualize multidimensional energy landscapes for long chains, but an elegant approach introduced by Stillinger and Weber (1984,1985) overcomes this obstacle. Conformational basins are identified by a mapping that connects all conformers to their corresponding local minima. Briefly, the Boltzmann-weighted sum over all conceivable conformers (Z) for a system in a canonical ensemble may be written as:

$$Z = \sum_i \exp(-U_i/k_BT);$$

U_i is the potential energy for conformation i and k_B is the Boltzmann constant. Z, referred to as a configurational sum, acts as an estimate of the canonical partition function (Chandler 1987). The energy U_i of a single conformation can be rewritten as $U_i = \{U_{\eta o} + (U_i - U_{\eta o})\} = U_{\eta o} + \Delta U_{\eta i}$, where $U_{\eta o}$ is the energy of the local minimum obtained by an energy minimization with i as the starting conformation. The sum over all conformations is rewritten as a sum over basins:

$$Z = \sum_\eta \{\exp(-\beta U_{\eta o}) \sum_{\{i\}} \exp(-\beta \Delta U_{\eta i})\},$$

where {i} refers to the set of unique conformations that can be mapped to the local minimum η_o and $\beta = 1/k_BT$. The likelihood P_η of finding the system in basin η is:

$$P_\eta = \frac{\exp(-\beta U_{\eta o}) \sum_{\{i\}} \exp(-\beta \Delta U_{\eta i})}{Z}.$$

In the descriptive terminology of Stillinger and Weber (1984), mapping by energy minimization is *configurational mapping*; local minima are *inherent structures*, and the set of conformations that map to an inherent structure constitute a *basin*. Thermodynamic preferences are expressed as basin occupation probabilities, and structural preferences are identified by analyzing the conformations associated with inherent structures or local minima. Configurational mapping for a simple one-dimensional function is illustrated in Figure 1.

The alanine dipeptide

Energetic preferences for the inverse-power potential function can be visualized from the energy landscape for the alanine dipeptide, which has just two degrees of freedom, ϕ and ψ. The energy landscape as a function of ϕ and ψ obtained from a coarse sampling is shown in Figure 2. To obtain a rigorous mapping of the energy landscape, an elaborate enumeration of conformations was carried out. A total of 4×10^8 uniformly distributed random conformers were generated, and the inverse-power potential energy was calculated for each.

Configurational mapping was used to identify basins of attraction and quantify energetic preferences. A two-step energy minimization (see Methods) was performed for each of the 4×10^8 distinct dipeptide conformations; results are summarized in Table 1. Ten unique minima were found on the dipeptide energy-surface. The locations of these minima and their respective catchment regions are shown in Figure 3, where all points that map to a local minimum are the same color.

The global minimum for the dipeptide corresponds to the (ϕ,ψ)-values for a P_{II}-helix

The global minimum is situated at (ϕ,ψ) = (−81.81°, 146.74°). If all (ϕ,ψ)-pairs for an oligomer assume similar values, the resultant conformation is a helix about some long axis in space (Ramachandran and Sasisekharan 1968). In this case the molecule would be a regular left-handed polyproline II (P_{II}) helix characterized by three residues per turn. Ideal (ϕ,ψ) values in a canonical P_{II} helix vary in the literature. Ramachandran and Sasisekharan (1968) showed that for standard peptide geometries, $\omega = 180°$ and the

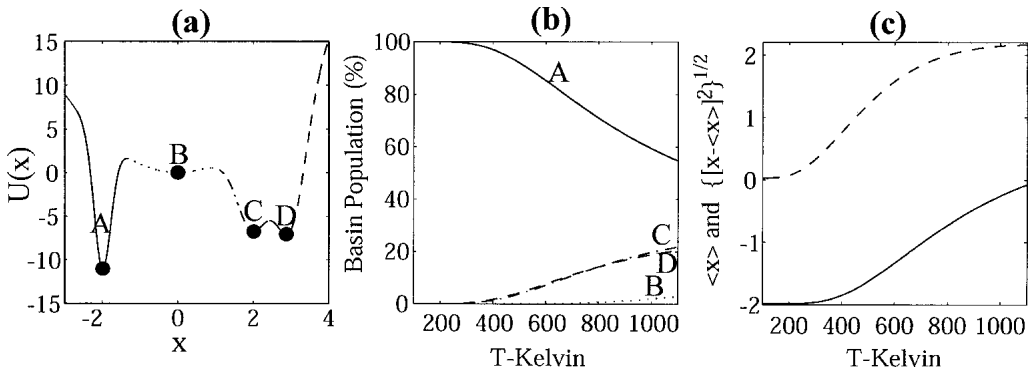

Fig. 1. (*a*) Plot of a one-dimensional function: $U(x) = \varepsilon\{x^2 - a\exp[-b(x + 2)^2] - c\exp[-d(x - 2)^2] - f\exp[-g(x - 3)^2]\}$ as a function of x for $a = f = 15, b = c = 10, d = g = 3$, and $\varepsilon = 1$kcal/mole. The function value $U(x_i)$ is calculated at each point in the interval x_i ([−3, 4] in increments of 0.0001. From each of these points, steepest descent minimizations were performed to identify local minima. Four unique minima exist and are shown as solid circles. Results of configurational mapping are indicated using different curves. As an example, all minimizations starting at points lying on the solid curve (–) converge to the global minimum. (*b*) Plot of the basin populations as a function of temperature. (*c*) Plot of $\langle x\rangle$ (–) and $\langle \Delta x^2\rangle^{1/2}$ (—), the Boltzmann weighted averages as a function of temperature.

N-C_α-C′ bond-angle set at 110° the $(\phi,\psi) = -77.2°$, 145.9°) generates a left-handed three-residue per turn helix with a pitch of 9.12Å. For the rigid peptide geometry chosen in this work, the (ϕ,ψ) values at the global minimum for the dipeptide generate a regular three-residue per turn P_{II} helix with a pitch of 9.07 Å. The second lowest energy conformer for the dipeptide is at $(\phi,\psi) = (-147.42°, 80.94°)$, and the corresponding helix is a three-residue per turn "symmetry mate" of the P_{II} helix referred to hereafter as sP_{II}.

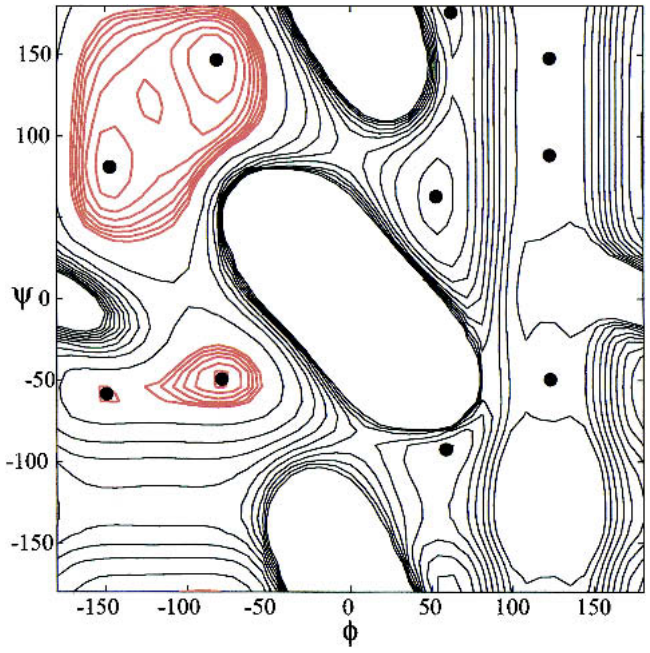

Fig. 2. Energy contour map for the alanine dipeptide. The contour map was generated by sampling (ϕ,ψ)-values on a uniform 9°×9° grid. Contours in black are drawn in 2.0 kcal/mole intervals from the global minimum. Additional contours, shown in red, are drawn in 0.1 kcal/mole intervals from the global minimum to highlight the low-energy regions. Locations of the ten energy minima are shown as filled circles. The (ϕ,ψ)-values and energy values of minima are summarized in Table 1.

Ramachandran and Sasisekharan (1968) used the Cahn-Infold-Prelog (Cahn et al. 1966) designation for the different conceivable helices. According to this convention, the sP_{II} helix would be referred to as a $P3_1L$ helix, where P is the Cahn et al. notation for right-handed axial chirality, 3_1 denotes a three-residue per turn helix, and L indicates that the (ϕ,ψ)-value is in the allowed region for L amino acids (that are not proline). According to this convention, the P_{II} helix would be referred to as the $M3_1L$ helix, where M is the Cahn et al. designation for left-handed chirality. For simplicity we use the notations P_{II} and sP_{II}. The two regular helices, P_{II} and sP_{II}, for a blocked alanine 7-mer are shown in Figure 4.

Table 1. *Summary of alanine dipeptide minima calculated using the inverse power potential*

Minimum No., assigned according to energy rank	(ϕ, ψ)-value at minimum	Inverse power potential energy and $(\Delta U = U_i - U_l)$ in kcal/mole
1	(−81.81°, 146.74°)	1.007 (0.000)
2	(−147.42°, 80.94°)	1.143 (0.136)
3	(−78.69°, −49.40°)	1.264 (0.257)
4	(−149.27°, −58.48°)	1.727 (0.721)
5	(52.86°, 62.70°)	2.425 (1.418)
6	(62.12°, 175.99°)	4.154 (3.147)
7	(58.89°, −92.40°)	8.873 (7.866)
8	(122.45°, 147.68°)	19.230 (18.282)
9	(122.48°, 88.21°)	19.346 (18.339)
10	(123.07°, −49.58°)	19.598 (18.591)

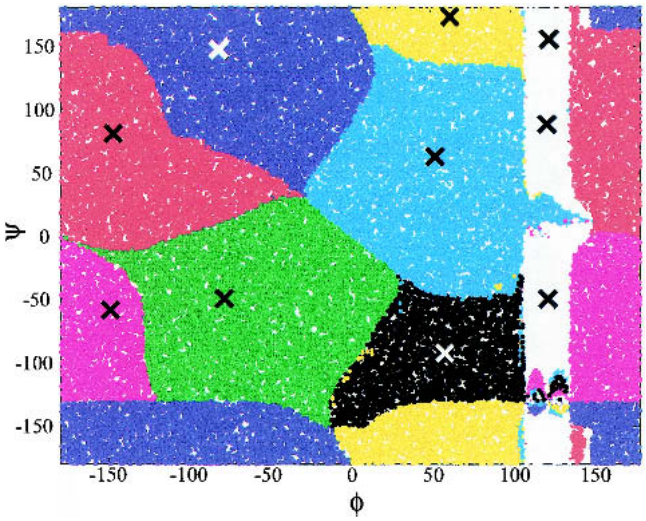

Fig. 3. The ten inherent structure basins from configurational mapping. Within each basin, the local minimum on the potential energy surface is marked with an "x". Same-color points map to the same local minimum. Points within the three highest energy basins are shown in white, and the boundaries for these basins are clear in Figure 2. Only 1/1000th of the sampled points were used to create the image; blank spots within a basin reflect this coarse-graining.

For purely repulsive inverse power potentials, dipeptide-conformers that maximize their pairwise interatomic distances with respect to the hard-sphere contact distances are preferred. In detail, for all atom pairs i and j, as the ratio of interatomic distances to hard-sphere contact distances (r_{ij}/σ_{ij}) increases (subject to covalent constraints), the dipeptide energy decreases. When dipeptide conformers are drawn from the allowed regions of a hard-sphere Ramachandran map (Ramachandran et al. 1963; Pappu et al. 2000), all interatomic distances satisfy the condition ($r_{ij}/\sigma_{ij} \geq 1$). At the global minimum (P_{II}), there are no interatomic contacts with (r_{ij}/σ_{ij}) <1.1 and only two contacts with (r_{ij}/σ_{ij}) <1.25. In comparison, at the penultimate minimum (sP_{II}), the side-chain methyl group is closer to the adjacent carbonyl oxygen atom, resulting in a small increase in energy with respect to the global minimum (Table 1).

Conformational basins

As shown in Figure 3, the basin catchment regions differ in size. The set of 4×10^8 conformers was used to estimate the configurational sum (Z) at a specified temperature, T, which in turn was used to calculate temperature-dependent occupation probabilities for the ten basins. The temperatures used in this work are not directly related to physical temperatures measured in experiments because we made no attempt to parameterize the energies to give agreement with small molecule experimental data. Our temperature scale is therefore a nonphysical temperature scale. Basin occupation probabilities change with temperature, and basins that encompass a large number of low-energy conformers are preferentially populated (Fig. 5).

The fraction of 4×10^8 sampled points required to account for 99.9% of the configurational sum (Z) at two different "temperatures", T = 100K and T = 300K is shown in Figure 6. Estimating the fraction of thermodynamically relevant conformers is a novel approach for assaying the results of conformational sampling (Pappu et al. 2000). Conformational energies are sorted in ascending order, and the number of conformers required to account for some stipulated fraction of the sampled configurational sum is calculated. This approach bears close resemblance to the method of most probable distributions in equilibrium statistical mechanics (Chandler 1987).

In Figure 6, envelopes of points are color-coded by their basin memberships. These envelopes measure the contributions due to conformational entropy. As temperature increases the number of thermodynamically relevant basins will increase; additionally the number of relevant conformers within a basin will also increase. Conversely, as temperature decreases, the population of thermodynamically relevant conformers approaches the immediate vicinity of the global minimum. At T = 300K, only five of the ten basins have significant occupation probabilities.

Fig. 4. Backbone drawings of left-handed polyproline II (P_{II}), in red, and the "symmetry mate" (sP_{II}) in orange for a blocked 7-mer N-Acetyl-Ala_7-N′-Methylamide. In the P_{II} structure, the (ϕ,ψ)-angles for residue are (−81.81°, 146.74°) and for sP_{II}, the (ϕ,ψ)-angles are (−147.42°, 80.94°).

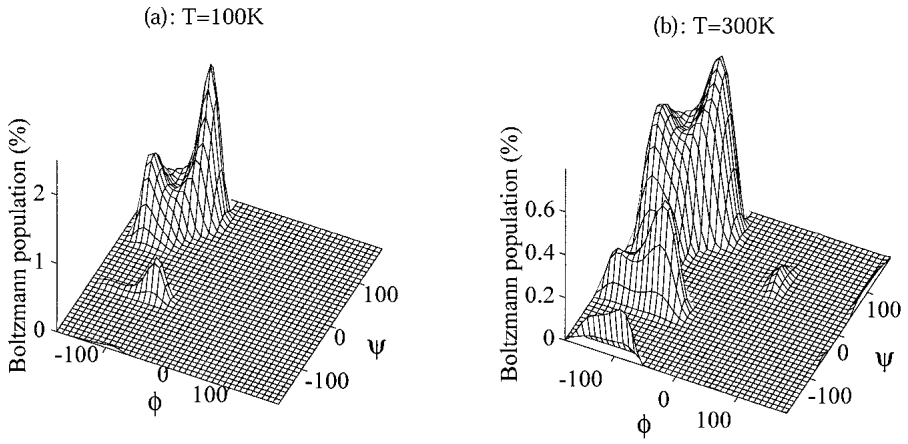

Fig. 5. Topographic map of Boltzmann populations on the (ϕ,ψ)-surface of the alanine dipeptide at two different temperatures. For lower temperatures (*a*), the population is largely distributed around the two lowest-energy conformers. The valley between the two peaks in the upper left quadrant is a saddle region, ~(−120°, 120°).

How do our results for the dipeptide compare with theory and experiment?

Poon et al. (2000) studied the conformations of blocked alanine—the alanine dipeptide—in an aqueous liquid crystal which mimics the behavior of the dipeptide in water, a good solvent for oligopeptides. They interpreted measured residual dipolar couplings from a proton NMR spectrum in terms of one dominant conformation (ϕ = −85°, ψ = 160°). Although the ψ-angle is larger than canonical values, Poon et al. refer to the dominant conformer as the (ϕ,ψ)-value for a P_{II}-helix. The inherent plasticity of P_{II}-helices (Adzhubei et al. 1987) justifies their interpretation.

Woutersen and Hamm (2000) used polarization-sensitive two-dimensional vibrational spectroscopy to study the backbone conformations for trialanine in aqueous solution. Trialanine is similar to the alanine dipeptide in that it has a similar number of degrees of freedom if one assumes that the terminal ϕ and ψ-angles are in a *trans* conformation. They conclude that the dominant conformer for trialanine has (ϕ,ψ)-values of (−60°, 140°).

The experimental work of Poon et al was prompted by earlier theoretical analysis (Grant et al. 1990; Han et al. 1998) for the alanine dipeptide. Starting from eight gas-phase energy minima found by Jalkanen and Suhai (1996), Han et al. (1998) studied the effect of hydration using a density functional approach. They found the dipeptide energies with four coordinated water molecules to be quite different from the gas-phase energies. The global minimum

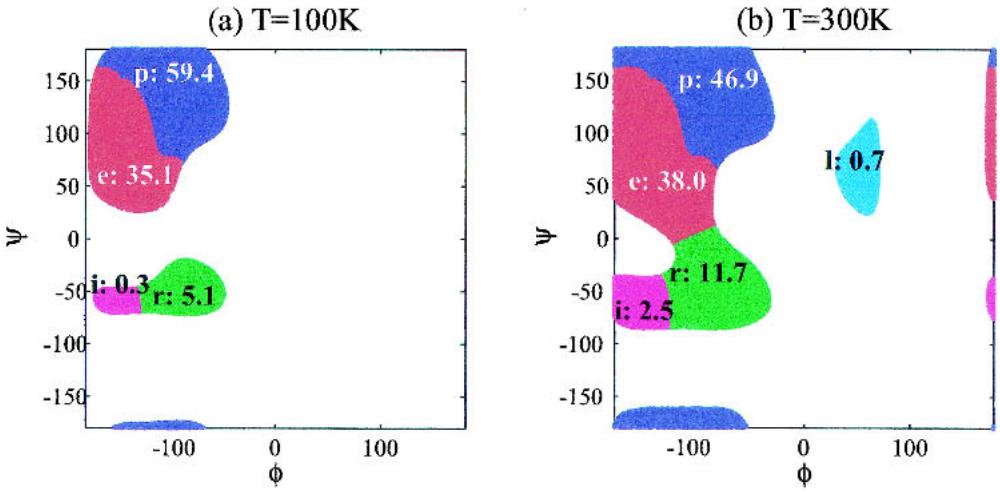

Fig. 6. Temperature-dependent envelopes for alanine dipeptide basins from configurational mapping. The points required to account for 99.9% of the configurational sum calculated using all 4×10^8 conformers is shown at two temperatures: (*a*) T = 300K and (*b*) T = 100K. In order of thermodynamic preference, the five basins are: blue (p), red (e), green (r), magenta (i), and cyan (l). Populations associated with each basin are shown.

with hydration is in the vicinity of the P_{II} conformation. (ϕ,ψ)-values and relative density functional energies for the seven low energy structures found by Han et al. (1998) are shown in Table 2. Also shown are energies calculated using the inverse power potential for these seven structures (column 4, Table 2). The rank ordering of conformations by energy is almost identical, although quantitative agreement of relative energies is poor, as would be expected for a model that ignores the details of peptide-solvent interactions.

What is the source of the good qualitative agreement between our results and those of Han et al.?

This is an especially important question because the results of Han et al. (1998) were obtained using four water molecules coordinated around the two peptide groups of the alanine dipeptide, whereas we ignored all details of intrachain and dipeptide-solvent interactions. The process of solvation may be split into two steps (Pieroti 1965): (1) the creation of a cavity of appropriate *size* and *shape* to accommodate the solute, and (2) introduction of the solute into the cavity to interact favorably with solvent. Our calculation addresses step (1) of the solvation process. It must be true that P_{II}-like conformations possess the appropriate distribution of intramolecular voids to accommodate solvent molecules. If so, the success of our model lies in the general validity of the good-solvent inverse-power potential which strives to minimize intrachain packing density and promote the creation of intra-molecular voids.

Comparison with results from other simulations

Free-energy surfaces for the alanine dipeptide have been the subject of numerous independent investigations based on all atom force-fields, both in vacuo and in the presence of solvent (Rossky et al. 1979; Anderson and Hermans 1988; Pettitt and Karplus 1988; Roterman et al. 1989; Schmidt and Fine 1994; Apostolakis et al. 1999; O' Connell et al. 1999; Smith 1999; Tobias and Brooks 1999). With the exception of Caflisch and coworkers (Apostolakis et al. 1999), who find the global free-energy minimum to be at (ϕ,ψ) ~ (−75°, 136°) for the dipeptide in water, almost all simulations converge on the global minimum being either in the β-region (Anderson and Hermans 1988; O'Connell et al. 1999;) or in the vicinity of the right-handed α-helix (ϕ,ψ) ~(−72°, −56°; Smith 1999). In the work of Tobias and Brooks (1992), the global minimum is in the vicinity of P_{II}, (ϕ,ψ) ~(−80°, 120°). Smith (1999) finds closer agreement with the results of Apostolakis et al. (1999) and those of Tobias and Brooks (1992) using either a continuum dielectric or a Poisson-Boltzmann approach to embed the alanine dipeptide solute in a solvent continuum. In summary, many of the force-field approaches capture the general trend toward P_{II}, although none are unequivocal about this preference, and numerous disagreements exist between the different force-field calculations. Smith's work (Smith 1999) provides an excellent summary of results from different calculations.

Mu and Stock (2002) performed two 20ns simulations at T = 300K and T = 350K for trialanine in water. This is the molecule studied by Woutersen and Hamm (2000) using vibrational spectroscopy. Mu and Stock found that the trialanine molecule spends roughly 80% of the time in the vicinity of β-strand (ϕ ~−122° and ψ~130°) and P_{II} (ϕ~−67° and ψ~132°) conformers.

Calculating the free-energy surface based on the inherent-structure formalism is relatively straightforward. The free-energy difference between a basin surrounding minimum η and the global minimum P_{II} basin at temperature T is: $\Delta A_{\eta o} = -RT\{\ln(P_\eta) - \ln(P_o)\}$, where $P_\eta \equiv P_\eta(T)$ is the occupation probability of basin η, $P_o \equiv P_o(T)$ is the occupation probability of the global minimum basin, and $\Delta A_{\eta o}$ is the difference in Helmholtz free energy. The basin occupation probabilities for the five low-energy basins are plotted as a function of temperature in Figure 7a. The results in this figure were used to calculate the Helmholtz free-energies of the different basins relative to the P_{II} basin, as shown in Figure 7b. At low temperatures, the dipeptide prefers to fluctuate in the immediate vicinity of the P_{II} conformation.

Table 2. *Summary of dipeptide minima from Han et al. (1998)*

Label and (rank) of minimum from Han et al	(ϕ, ψ)-value at minimum	Energies relative to the global minimum $\Delta U = U_x - U_{P_{II}}$ from Table 1 of Han et al. (1998)[a]	Energy in kcal/mole and (rank) calculated using inverse power potential[a]
P_{II} (1)	(−93.55°, 127.62°)	0.000 (1)	0.000 (1)
β′ (2)	(−150.88°, 116.47°)	1.886 (2)	0.128 (2)
α_R' (3)	(−82.08°, −44.11°)	2.465 (3)	0.179 (3)
α_L' (4)	(60.73°, 52.44°)	2.754 (4)	1.690 (4)
α_D' (5)	(66.68°, −111.27°)	3.715 (5)	3.307 (5)
$C_1^{ax\prime}$ (6)	(58.88°, −121.92°)	4.134 (6)	9.583 (7)
α_P' (7)	(−152.94°, −92.01°)	15.140 (7)	4.662 (6)

[a] Numbers in parentheses indicate the rank ordering of conformers based on the energies relative to the P_{II} conformer (ϕ = −93.55°, ψ = 127.62°) found by Han et al. (1998) and not the P_{II} conformer of Table 1.

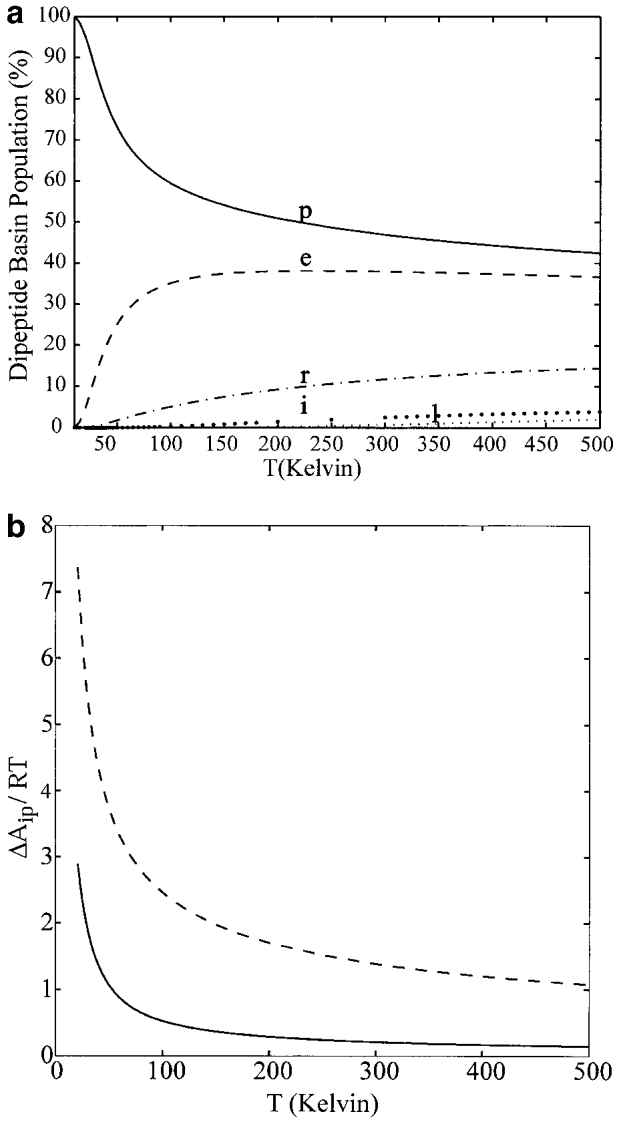

Fig. 7. (*a*) Populations for p (—), e (– –), r (–.), i (:), and l (.) dipeptide basins as a function of temperature. (*b*) Scaled Helmholtz free energy ($\Delta A_{ip}/RT$) of the e-basin (–) and the r-basin (– –) relative to the P_{II} (p-basin) as a function of temperature.

From Figure 7b it is clear that the dipeptide freely fluctuates around the P_{II} and sP_{II} minima, and the two basins are essentially isoenergetic for all temperatures greater than 100K. General agreement with detailed free energy calculations is encouraging, and we believe our model provides a simple interpretation of results seen using detailed all atom force-fields.

There are only two adjustable parameters in our model: the exponent for inverse power potentials (n) and values for the hard sphere radii (σ; see Methods). The latter are dictated by stereochemical considerations (Ramachandran and Sasisekharan 1968). Consequently n is the only adjustable parameter. As the exponent n in $(\sigma_{ij}/r_{ij})^n$ is lowered, inverse-power potentials mimic interactions between softer spheres. Increase of n beyond 12 results in interactions between harder spheres. Table 3 summarizes the location of local minima, the number of local minima and the energies relative to the global minimum for n = 9 and 14. As n decreases, a pair of minima connected by a saddle point shift toward each other, and the relative energies and the energy at the barrier decrease until a single minimum will result from coalescence of basins due to the elimination of the barrier. The relative energy differences between existing minima increases. As n increases beyond 12, all allowed conformers become isoenergetic and their total inverse-power potential energy approaches zero. Conversely, the high-energy conformers also become isoenergetic and the inverse-power potential energy approaches infinity, consistent with the fact that for a hard sphere potential, conformers with steric overlap are disallowed.

The inverse power potentials used here are similar to the modified hindrance potentials of Flory and coworkers (see eqs. 4 and 7 in Brant et al. 1967). The low-energy contours and the location of the global minimum shown in Figure 2 and those from the hindrance potentials of Brant et al. (1967) are essentially identical. The asymmetric distribution of distances in Cartesian space leads to two local minima and a well defined saddle point in the upper left quadrant for the inverse power potential.

Results for longer chains

A polyalanine 7-mer was studied at two temperatures, 100K and 300K. The 7-mer was chosen to facilitate direct comparison with the recent experimental work of Shi et al. (2002). The 7-mer has 14 independently rotatable bonds, so it is impossible to map the multidimensional energy landscape exhaustively. However, as noted earlier, only five of the ten conformational basins are required to account for 99.9% of the configurational sum for the alanine dipeptide over a wide range of temperatures. To make the enumeration of 7-mer conformers tractable, sampling was restricted to just the five basins shown in Figure 6a, that is, 5^7 distinct pockets in the 14-dimensional space of interest.

The restrictions imposed on the sampling are robust and do not lead to a neglect of thermodynamically significant conformers. This is a direct consequence of choosing purely repulsive potentials: energies of longer chains are either a strict sum of dipeptide energies, or increase drastically if the combination of (ϕ,ψ)-values leads to longer-range steric overlaps (Pappu et al. 2000). The latter is true of conformers that are a combination of (ϕ,ψ)-values from all but the two dipeptide basins shown in blue and in red in Figure 6. A stratified Monte Carlo importance sampling (see Methods) was used to generate a total of 5×10^9 independent conformers for the two temperatures.

Table 3. *Location and energies of alanine-dipeptide local minima for different values of the inverse-power potential exponent (n). These values are to be compared to those shown in Table 1 for n = 12*

n = 9			n = 14		
ϕ	ψ	Inverse power potential energy and ($\Delta U = U_i - U_l$) in kcal/mole	ϕ	ψ	Inverse power potential energy and ($\Delta U = U_i - U_l$) in kcal/mole
−86.23°	143.52°	2.156 (0.000)	−80.31	148.32	0.640 (0.000)
−142.17°	86.83°	2.287 (0.131)	−149.62	78.91	0.762 (0.122)
−81.52°	−49.20°	2.626 (0.469)	−77.67	−49.69	0.821 (0.181)
52.44°	65.31°	3.954 (1.797)	−151.77	−58.20	1.209 (0.569)
63.23°	171.66°	5.634 (3.477)	53.24	61.81	1.866 (1.226)
55.67°	−97.49°	9.474 (7.318)	61.64	177.63	3.547 (2.907)
122.90°	141.91°	13.412 (11.256)	60.23	−90.99	8.949 (8.309)
123.04°	98.56°	13.463 (11.307)	122.42	149.83	26.284 (25.644)
125.72°	−48.94°	13.951 (11.795)	122.44	84.97	26.336 (25.695)
–	–	–	122.71	−49.97	26.505 (25.865)

Configurational mapping allows us to analyze conformational preferences on a 14-dimensional energy landscape. For each of the 5×10^9 sampled conformers, the two-step energy minimization described for the dipeptide was used to identify its inherent structure (local minimum). The set of conformers that map to a given inherent structure comprise its basin. Once all inherent structures and basins were identified, the likelihood (P_η) of finding the system within a basin (η) is calculated.

Results from configurational mapping

We identified 16,317 unique inherent structures at 100K and 78,103 unique inherent structures at 300K. The distribution of inherent structure energies at 300K, a superset of the 100K case, is shown in Figure 8. There are only four significant dipeptide basins at T = 100K, which explains the smaller number of basins for the 7-mer at this temperature. The global minimum is a regular P_{II} helix, with ϕ,ψ-angles: $\{\phi_1,\psi_1 = -81.8°, 146.4°; \phi_i,\psi_i = -82.0°, 146.0°; \phi_7,\psi_7 = -82.0°, 146.3°\}$, i = 2, . . . ,6.

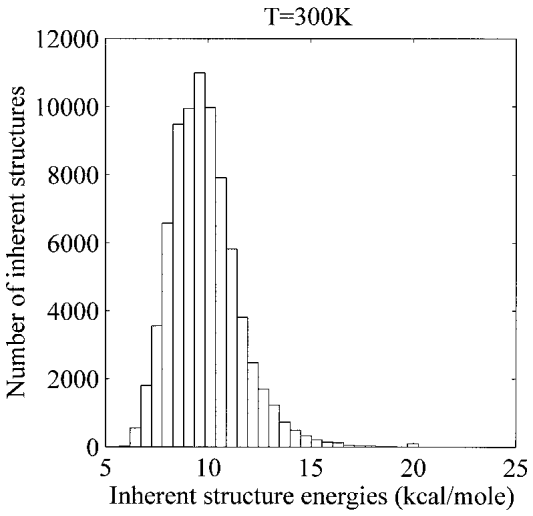

Fig. 8. Distribution of 78,103 inherent structure energies for a polyalanine 7-mer at T = 300K.

As noted earlier, the five relevant alanine dipeptide minima are p ($\phi\approx-81°$, $\psi\approx147°$), e ($\phi\approx-147°$, $\psi\approx81°$), r ($\phi\approx-79°$, $\psi\approx-49°$), i ($\phi\approx-149°$, $\psi\approx-58°$), and l ($\phi\approx53°$, $\psi\approx63°$). On the 14-dimensional landscape, the inverse-power potential energy is the sum of dipeptide energies when (ϕ,ψ)-pairs sample values from the p or e dipeptide basins (Fig. 6), or when no two consecutive (ϕ,ψ)-pairs are from high-energy dipeptide basins, r, i, and l (Fig. 6). Such conformers are isoenergetic, and consequently the energies of several inherent structures (U_η) and associated basin weights (P_η) are identical. The free-energy difference between a basin η and the global minimum P_{II} basin at temperature T is: $\Delta A_{\eta o} = -RT\{\ln(P_\eta) - \ln(P_o)\}$, where $P_\eta \equiv P_\eta(T)$ is the occupation probability of basin η, $P_o \equiv P_o(T)$ is the occupation probability of the global minimum basin, and $\Delta A_{\eta o}$ is the difference in Helmholtz free energy.

As an example, consider two inherent structures, η and γ, with ϕ,ψ-angles: ($\phi_i,\psi_i = -82.0°, 146.0°$, $\phi_7,\psi_7 = -147.5°, 81.6°$, i = 1, . . ,6) and ($\phi_i,\psi_i = -82.0°$, 146.0°, i = 1,2,3,5,6,7 and $\phi_4,\psi_4 = -147.5°, 81.6°$), respectively. The two inherent structures η and γ are isoenergetic, with $U_{\eta o} = U_{\gamma o} = 6.1$ kcal/mole. Additionally, their basin occupation probabilities, P_η and P_γ, are also similar, which leads to similar values for $\Delta A_{\eta o}$ and $\Delta A_{\gamma o}$. The set of inherent-structures realized by permutations of six (ϕ,ψ)-pairs $\in$ p and one (ϕ,ψ)-pair $\in$ e is labeled p6e1, and the label identifies the corresponding free-energy level.

T = 100K

Data for T = 100K are summarized in Table 4, which lists the inherent structure basins that account for 90% of the

Table 4. *Energy landscape for the 7-mer at T = 100K*[a]

Free-energy level code[b]	No. of isoenergetic basins per free energy-level	Free-energy level $\Delta A_{\alpha o}/RT$	Population (%) P_j over all basins i on level j $Pj = \sum_{i=1}^{nbasin} pi$	$\langle C_\alpha \rangle$ distance (Å) between isoenergetic inherent structures[c]	$\langle C_\alpha \rangle$ distance (Å) between isoenergetic inherent structures and the global minimum P_{II} helix[c]	$\langle C_\alpha \rangle$ distance (Å) between isoenergetic inherent structures and sP_{II} helix[c]
p7	1	0.00	1.47	0.00	0.00	1.72
p6e1	7	0.27	8.05	0.12 (0.004)	0.74 (0.20)	1.01 (0.27)
p5e2	21	0.57	18.44	0.22 (0.006)	1.08 (0.15)	1.18 (0.28)
p4e3	35	0.82	23.83	0.34 (0.006)	1.24 (0.19)	1.30 (0.30)
p3e4	35	1.05	18.44	0.45 (0.006)	1.45 (0.14)	1.27 (0.31)
p2e5	21	1.32	8.50	0.57 (0.005)	1.49 (0.16)	1.24 (0.31)
p1e6	7	1.56	2.20	0.69 (0.002)	1.58 (0.18)	1.08 (0.29)
e7	1	1.81	0.24	0.00	1.72	0.00
p6e0rl	7	2.93	0.64	0.89 (0.033)	1.49 (0.61)	2.19 (0.86)
p5e1rl	42	3.63	2.70	0.99 (0.003)	1.68 (0.65)	2.06 (0.68)
p4e2rl	108	4.09	5.08	1.11 (0.046)	1.91 (0.56)	2.09 (0.58)

[a]RT = 0.2 kcal/mole
[b]The free-energy levels listed here account for 90% of the total population at 100K.
[c]$\langle C_\alpha \rangle$ is the root-mean-squared-distance. Parenthetic values are standard deviations about the mean.

sampled population. Free-energy levels are labeled by the number of occurrences of p, e, or r ϕ,ψ-values in the inherent structure. Column 2 lists the number of degenerate basins at each free-energy level. All basins at a given free-energy level have an equal likelihood of being populated; column 4 gives the population at each of these levels.

There is a dominant cluster of basins, within $\Delta A_{\alpha o}/RT \approx 1.8$ (Fig. 9a). This distribution is bracketed between p7 (P_{II}) and e7 (sP_{II}). All inherent structures between p7 and e7 are permutations of ϕ,ψ-values from these two types of helices. Beyond the dominant cluster ($\Delta A_{\alpha o}/RT > 1.8$ in Fig. 9a), the population is defined by conformers with one (ϕ,ψ)-value in the r-basin (centered around $\phi \sim -81°$, $\psi \sim -49°$), primarily p6e0r1, and p5e1r1.

The system is found in the p7 basin in only ~1.5% of the ensemble, but thermal fluctuations to higher free-energy levels access similar structures. To illustrate the structural nature of thermal fluctuations, we use the root-mean-squared-distance (RMSD) between superimposed inherent-structures. Data for the 11 lowest free-energy levels are shown in Table 4, columns 5–7. Inherent structures from the same free-energy level are structurally similar to each other (column 5), but structurally dissimilar both to inherent structures from other levels and to random conformations. For example, the average RMSD of all unique pairwise combinations of the 21 inherent structures at level 3 (p5e2) is 0.22 ± 0.006 Å. In contrast, the average RMSD is 4.7 ± 1.9 Å between any of these inherent structures and 100 conformers generated from random combinations of (ϕ,ψ)-values from p, e, r, i, and l dipeptide basins. The RMSD was also computed between each inherent structure and the global minimum P_{II} conformation (Table 4, column 7). Fluctuations about inherent structures that are structurally similar to the P_{II} helix are lower in free energy and account for a major fraction (~90%) of the equilibrium population. The data from Table 4 (column 7), plotted in Figure 10, reveal this strong correlation between free energy relative to the P_{II} basin and structural similarity to a P_{II} helix. Clearly the landscape is structured, and is dominated by uncorrelated fluctuations of individual residues around P_{II} -helix (ϕ,ψ)-values.

T = 300K

An equivalent analysis was performed at a higher temperature, T = 300K. The results are shown in Figure 9c,d. Comparison of the scales on the abscissa of Figure 9a and c indicate that the free-energy difference between the global minimum and higher-energy basins decreases with increasing temperature. There is a roughly threefold increase in the number of relevant basins, which is consistent with a temperature-dependent increase in energy fluctuations. Two new clusters of inherent-structure basins emerge (Fig. 9c). In the second cluster ($1.9 \leq \Delta A_{\alpha o} \leq 2.8$), one ϕ,ψ-value is in the r-basin ($\phi \sim -81°$, $\psi \sim -49°$), whereas in the third cluster, one ϕ,ψ-value is in the i-basin ($\phi \sim -149°$, $\psi \sim -58°$); remaining ϕ,ψ-values are from either p or e. Overall, the diminished preference for low-energy P_{II} (p7) basins is offset by the entropic preference for higher-energy basins.

The degeneracy for each free-energy level is shown in Figure 9b,d. The basins on each free-energy level are isoenergetic, and this promotes uncorrelated fluctuations of individual residues about preferred (ϕ,ψ)-values. The degeneracy (Fig. 9b,d) emphasizes that conformational entropy is maximized upon unfolding. Dynamic disorder refers to hopping between inherent-structure basins. Equilibrium populations for the 7-mer show clear evidence for dynamic disorder in that the 7-mer is floppy, although individual resi-

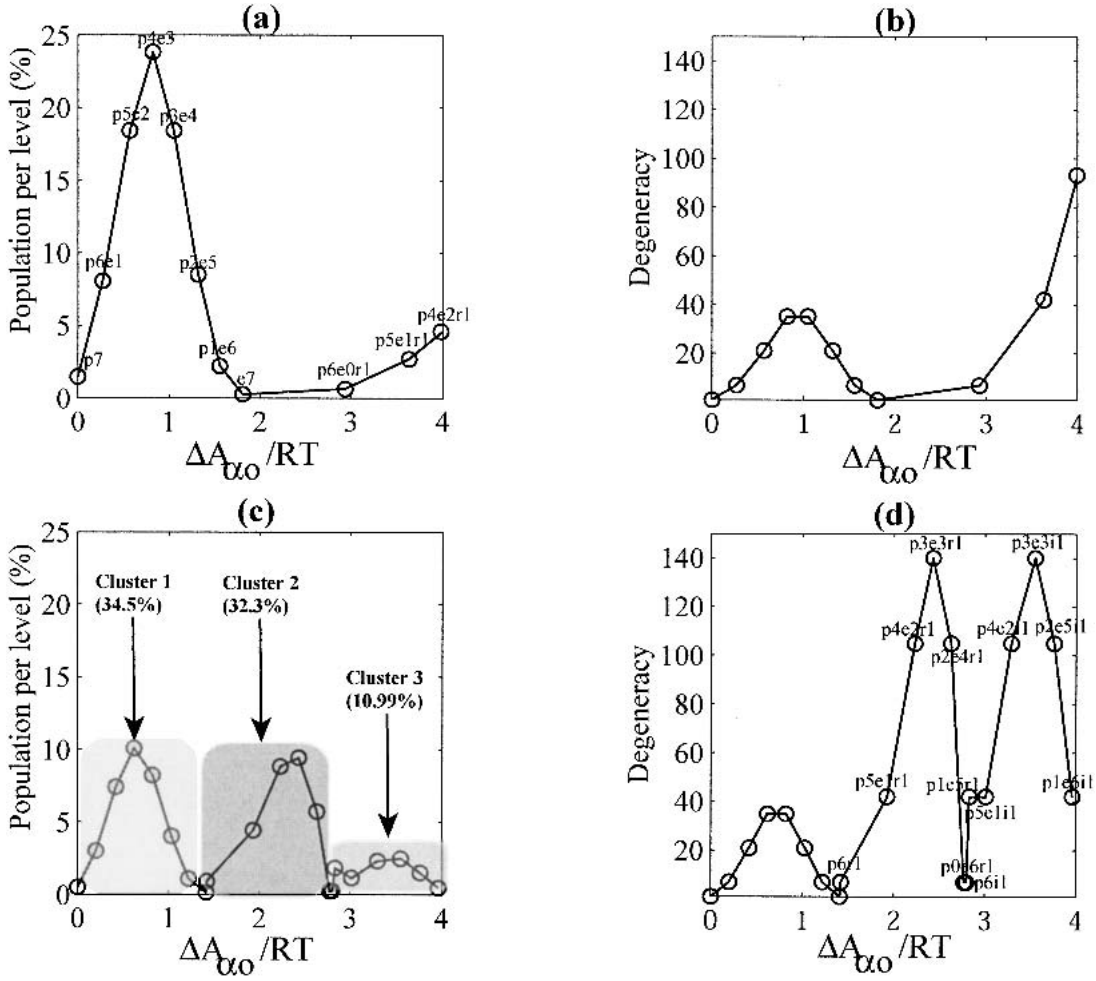

Fig. 9. Population map as a function of the scaled Helmholtz free energy ($\Delta A_{\alpha o}/RT$) relative to the ground state, at T = 100K (*a*,*b*) and T = 300K (*c*,*d*). Raw data used to generate the plots in (*a*) and (*b*) are from Table 4. (*a*) Sum of populations over all basins on each discrete free-energy level at T = 100K. Open circles coincide with observed values. (*b*) The degeneracy (i.e., the number of isoenergetic dihedral angle permutations) for each discrete free energy level. (*c*) and (*d*) are the T = 300K versions of (*a*) and (*b*). In all plots, the solid curves serve only as a visual aid.

dues fluctuate about preferred low-energy P_{II} and sP_{II} (ϕ,ψ)-values. Local preferences give the chain a distinctly nonrandom character.

Making contact with experiment

Our analysis of the 7-mer energy landscape supports the view that individual residues show uncorrelated fluctuations about two preferred dipeptide basins, P_{II} (p) and sP_{II} (e). The quantitative analysis of conformational preferences on the energy landscape goes well beyond the capabilities of typical measurements that report on ensemble-averaged properties. The distribution of inherent-structure weights can be used to compute different properties accessible to experimental measurements. To illustrate this, we used our knowledge of the distribution of Boltzmann weights for conformational basins to calculate scalar coupling constants accessible to a typical NMR experiment.

NMR scalar coupling constants characterize interactions between nuclei separated by a small number of covalent bonds (Wüthrich 1986). The Karplus relation (Karplus 1959) relates values for coupling constants between third nearest neighbors (^{3}J- measured in Hertz) to the dihedral angle about a single rotatable bond. Parameters to extract the relationship between the (ϕ,ψ)-angles and different ^{3}J coupling constants have been worked out in previous studies (Pardi et al. 1984; Wüthrich 1986; Vuister and Bax 1993; Wang and Bax 1995).

We used a prescription outlined in the Methods section to calculate ^{3}J coupling constants as weighted thermodynamic averages. Two types of averages were computed. The first, referred to as an annealed average, uses (ϕ,ψ)-values for

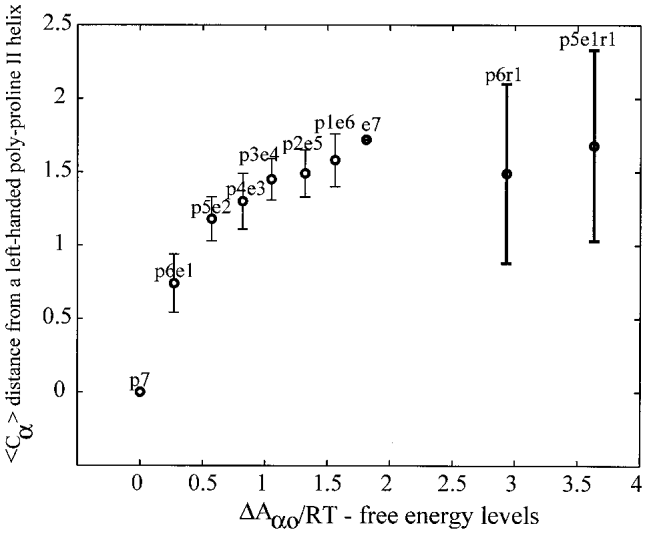

Fig. 10. Structural annotation of the free-energy surface at T = 100K, showing the average distance from the P_{II} helix as a function of free-energy relative to the ground state. Distances were calculated after superposition of α-carbon atoms. Data used to create the plot are from column 5 of Table 4. Error bars denote the standard deviation from the average distance. When the degeneracy (i.e., the number of isoenergetic basins at a given free-energy level) is high, the deviation from the mean is large.

each residue of the 7-mer from the sampled distribution of 5×10^9 conformers. In addition, we computed quenched averages that use the (ϕ,ψ)-values of inherent structures to compute the ^{3}J coupling constants. The two sets of results are summarized in Table 5a for T = 100K and Table 5b for T = 300K. The coupling constants $^3J_{HN\alpha}$, $^3J_{HNCO}$, $^3J_{HN\beta}$, and $^3J_{C(i-1)H\alpha(i)}$ report on the ensemble-averaged value of the ϕ-angle for a particular residue and the measured $^3J_{NiH\alpha i-1}$ values report on the ensemble average for ψ. In the next few paragraphs, we discuss our calculations in the context of the recent measurements of Shi et al. (2002) for the 7-mer.

The temperature dependence of measured coupling constants and those calculated from our sampled equilibrium ensemble cannot be directly compared because our temperature scale is nonphysical. Our model is incomplete as it ignores the details of chain-solvent interactions. Therefore 100K in our calculations does not correspond to 100K in vitro. However, general trends from theory and experiment can be compared as is shown in Tables 5a and 5b.

The calculated $3J_{HN\alpha}$ coupling constant is roughly 7.4 Hz for all residues in the 7-mer. As shown in Figure 11, this value is consistent with ensemble average angles of either $\phi \approx -85°$ or $\phi \approx -155°$. This degeneracy would be referred to as conformational averaging (Smith et al. 1996), and the common assumption would be that all values between $\phi \approx -85°$ and $\phi \approx -155°$ are equally likely. Our analysis of basin occupation probabilities (Fig. 9a,c) allows us to correctly interpret the calculated coupling constants as being consistent with thermal fluctuations of each residue about distinct basins centered about P_{II} and sP_{II}.

The calculated $3J_{HN\alpha}$ coupling constants shown in Tables 5a and 5b are insensitive to changes in temperature, although the populations for each of the 5^7 basins change with temperature as shown in Figure 9a,c. Results for the dipeptide (Figs. 6, 7) help clarify the origins of the thermal insensitivity of calculated coupling constants. As temperature increases, increased disorder leads to increased population in basins around the sP_{II} minimum (e-basin) and the α_R minimum (r-basin). In both cases, the fluctuations lead to values for the $3J_{HN\alpha}$ coupling constant that are similar to values computed for fluctuations in the P_{II} basin. Therefore the model is not athermal. Instead, the calculated coupling constants are insensitive to changes in temperature because increased disorder leads to fluctuations about basins that result in similar values for the coupling constant. The values for the remaining coupling constants $^3J_{HNCO}$, $^3J_{HN\beta}$, and $^3J_{C(i-1)H\alpha(i)}$ that report on the ϕ-angle support the conclusion that the ensemble average is tilted toward P_{II}-like conformers over sP_{II} conformers. Although both basins are prevalent, the P_{II} basin is preferred, and the preference is weak (Table 4).

Shi et al. (2002) used NMR to study the conformers available to an alanine-based peptide, Acetyl-XXAla$_7$OO-Amide (XAO), in water. Here X is diaminobutyric acid and

Table 5a. *Calculated scalar coupling constants for polyalanine at T = 100 K*

Res #	$^1J_{HN\alpha}$ Hz	$^3J_{HNCO}$ Hz	$^3J_{HN\beta}$ Hz	$^3J_{C(i-1)H\alpha(i)}$ Hz	$^3J_{NiH\alpha i-1}$ Hz	$^3J_{HN\alpha}$ Hz	$^3J_{HNCO}$ Hz	$^3J_{HN\beta}$ Hz	$^3J_{C(i-1)H\alpha(i)}$ Hz	$^3J_{NiH\alpha i-1}$ Hz
1	7.63	0.78	1.03	1.01	–0.45	7.47	1.02	1.31	0.82	–0.43
2	7.84	0.90	0.98	1.02	–0.46	7.49	1.02	1.30	0.83	–0.42
3	7.69	0.84	0.97	0.98	–0.47	7.48	1.02	1.30	0.83	–0.42
4	7.58	0.83	0.99	0.98	–0.44	7.48	1.02	1.30	0.83	–0.42
5	7.89	0.91	0.95	0.99	–0.45	7.48	1.02	1.30	0.83	–0.42
6	7.49	0.79	1.02	0.96	–0.48	7.48	1.01	1.31	0.83	–0.43
7	7.83	0.89	1.02	1.08	–0.49	7.46	1.01	1.34	0.82	–0.44

Values in the shaded region are quenched averages obtained using (ϕ,ψ)-values from inherent structures

Table 5b. *Calculated scalar coupling constants for polyalanine at T = 300 K*

Res #	$^{1}J_{HN\alpha}$ Hz	$^{3}J_{HNCO}$ Hz	$^{3}J_{HN\beta}$ Hz	$^{3}J_{C(i-1)H\alpha(i)}$ Hz	$^{3}J_{NiH\alpha i-1}$ Hz	$^{3}J_{HN\alpha}$ Hz	$^{3}J_{HNCO}$ Hz	$^{3}J_{HN\beta}$ Hz	$^{3}J_{C(i-1)H\alpha(i)}$ Hz	$^{3}J_{NiH\alpha i-1}$ Hz
1	7.77	0.90	0.86	1.00	−0.51	7.44	1.03	1.33	0.84	−0.57
2	7.72	1.04	0.87	1.02	−0.44	7.47	1.03	1.30	0.85	−0.54
3	7.34	1.01	1.00	0.91	−0.45	7.44	1.03	1.32	0.84	−0.54
4	7.61	0.85	0.99	1.06	−0.47	7.44	1.03	1.33	0.84	−0.54
5	7.41	0.91	0.92	1.10	−0.49	7.44	1.03	1.33	0.84	−0.54
6	7.56	0.90	0.85	1.10	−0.46	7.43	1.02	1.35	0.83	−0.57
7	7.61	1.02	0.90	0.92	−0.50	7.39	1.02	1.37	0.81	−0.59

Values in the shaded region are quenched averages obtained using (ϕ,ψ)-values from inherent structures

O is ornithine. XAO may be thought of as a soluble version of a 7-residue polyalanine. Well resolved peaks in the amide region permit the measurement of $^{3}J_{HN\alpha}$ scalar coupling constants. Shi et al. (2002) calculated the average value of the backbone ϕ-angle at 2°C to be −70° ± 10° for each of the alanine residues. They used the optimized coefficients of Vuister and Bax (1993) to calculate the ϕ-angle from the measured coupling constants. The coefficients of Vuister and Bax are only slightly different from those of Pardi et al. (1984) used in our calculations. We have used both sets of coefficients and find that our results are essentially insensitive to the choice of Karplus-equation coefficients. Shi et al. found that the measured $^{3}J_{HN\alpha}$ coupling constants increase monotonically with temperature between ~5.5 Hz at 2°C and ~6.1 Hz at 56°C. The coupling constants measured by Shi et al. are consistent with ϕ-angles of ≈−70° at 2°C and −75° at 55°C. Shi et al. did not measure the other coupling constants. Based on the ratio of NOEs between nearest neighbor β-protons, Shi et al. suggest that the ψ-angles for all alanine residues are likely to be: +145° ± 20°. The average ϕ-angle for P_{II} consistent with the calculated coupling constant is ~10° off from the ϕ-angle calculated using experimentally measured coupling constants. However, we have provided ample evidence that the free-energy surface for the inverse-power potentials is dictated by fluctuations about P_{II} and sP_{II}-helices.

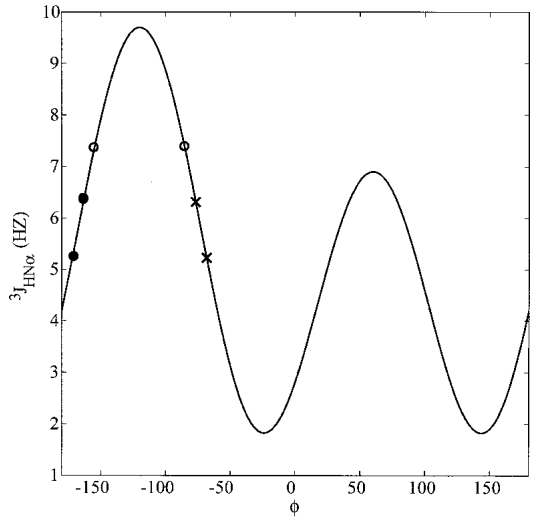

Fig. 11. Plot of the $^{3}J_{HN\alpha}$ coupling constant (in Hertz), calculated using the Karplus (1959) equation, as a function of the angle ϕ. The calculated coupling constants (○) are consistent with two different values for the ϕ-angles. The inherent-structure analysis helps resolve this degeneracy. The coupling constants are a consequence of fluctuations centered about P_{II} and sP_{II} conformers. The coupling constants measured by Shi et al. (x) are also shown. Solid circles (●) indicate alternate values for the ϕ-angle that cannot be ruled out in the coupling constant analysis of Shi et al. These values are consistent with ϕ-angles for putative sP_{II} helices in water. As discussed in the text, it is entirely possible that fluctuations about the sP_{II} set of conformers are higher in energy and therefore inaccessible in water.

Shi et al. interpret the temperature-dependent behavior of measured coupling constants as evidence for an admixture between β-strand and P_{II} helical structures. It should be noted that their coupling constants are also consistent with ϕ-angles between −165° and −170°, as shown in Figure 11. These values correspond to the sP_{II} basin (Fig. 6). Shi et al. did not consider this possibility in interpreting their measurements. It would appear that the sP_{II} conformers can be ruled out because the margin of error in the deduced values for the ψ-angles is inconsistent with the possibility of fluctuations around sP_{II}.

There are two possible interpretations for the temperature dependence of measured scalar coupling constants. The sP_{II} basin may be unfavorable in water, and therefore increased thermal agitation leads to increased sampling within the basin around P_{II}, which in our model subsumes (ϕ,ψ)-values consistent with β-strands that would be part of parallel β-sheets. The interpretation of Shi et al. would be valid if for our model fluctuations for all residues in the 7-mer are confined to the blue regions shown in Figure 6a,b. Alternatively, fluctuations at lower temperatures may be centered about P_{II}, whereas at higher temperatures the system fluctuates about both P_{II} and sP_{II} basins. If the latter were true, the measured coupling constants would have to be relatively insensitive to changes in temperature, as was found with calculated coupling constants. The measured coupling constants change by ~0.6 Hz, and it is not clear whether this is a significant change given that the ensemble averaged

ϕ-angle changes by 5° over this temperature range. Therefore neither alternative can be unequivocally ruled out, although the monotonic increase of measured $3J_{HN\alpha}$ values may preclude fluctuations about sP_{II}.

Applying the configurational mapping formalism to a system that includes details of chain-solvent interactions should clarify the interpretations of measured coupling constants. It is also conceivable that softening the potential (Table 3) by choosing smaller values for the exponent of the inverse-power potentials will lead to improved agreement with measured coupling constants. Upon softening the potential, the ϕ-angle shifts away from the canonical values toward the experimentally measured values (Table 3). The selectivity for extended chain conformers increases with softer potentials and diminishes with hard repulsive potentials.

More importantly, the configurational mapping procedure should allow us to make contact with results from a variety of other measurements including UV circular dichroism (Woody 1992), one-dimensional vibrational absorption spectroscopy (Woutersen and Hamm 2000), two-dimensional infrared polarization spectroscopy (Woutersen and Hamm 2000; Schweitzer-Stenner 2001), and NOE data from two-dimensional NMR experiments (Neuhaus and Williamson 2000). Additional investigations with labeled XAO peptides might permit the measurement of other coupling constants that do not vary steeply for small changes in the ϕ-angle and that have different values for P_{II}, sP_{II}, and α_R ϕ-angles. Differences in coupling constants such as $3J_{Ci-1H\alpha i}$ may be useful to determine the energy difference between different conformers. Labeled peptides may also provide information regarding the ψ-angle through the measurement of the $^3J_{NiH\alpha i-1}$ coupling constant. It is clear that one cannot comment authoritatively on conformational preferences in unfolded states based on the value reported by one number, especially if that number happens to be the scalar coupling constant $^3J_{HN\alpha}$ that gives degenerate values for different ϕ-angles. Only by making contact with data from diverse experimental probes can we fully validate the quantitative predictions of our simple model.

Connection to recent theoretical work

Banavar et al. (2002) recently showed that considerations of chain connectivity, compactness, and excluded volume impose geometrical constraints on the conformations accessible to coarse-grained protein-like polymers. The global radius of a chain computed by considering the smallest radii for circles that circumscribe all triplet combinations of C_α atoms measures the thickness of the chain in a particular conformation. Thickness is related to the "wiggle room" or "free volume" around a chain, and is a direct measure of conformational flexibility. Banavar et al. (2002) found that secondary structure motifs emerge as a consequence of maximizing the thickness upon chain compaction.

We calculated the chain thickness for each of the enumerated inherent structures of the 7-mer at T = 300K. The distribution of the chain thickness values (Fig. 12) is bimodal. The linear, uniformly weighted average thickness is 2.73Å. This value is in close agreement with the chain thickness for a 7-mer that populates mixtures of compact and extended conformations, which would be expected for a statistical random-coil. Conversely, the weighted average, calculated using the Boltzmann weights associated with each inherent structure, is 3.48 Å, a value consistent with preferential population of extended conformers for the 7-mer. The ratio of the weighted average to the linear average, a measure of nonrandomness realized by preferring extended conformers, is ~0.73. This value is in agreement with that computed by Banavar et al. (2002) for the ratio of the chain thickness of β-strands or extended conformers to compact α-helices. Maximizing chain thickness under good solvent conditions, akin to minimizing chain packing density, promotes unhindered conformational fluctuations about P_{II} and sP_{II} helices. Conversely, one would expect maximization of chain thickness under poor solvent conditions, akin to maximizing chain packing density, to lead to α-helix or β-sheet formation. This expectation is consistent with the findings of Banavar et al. (2002).

Discussion

There is long-standing evidence for P_{II}-helices in a variety of peptide and protein systems. Homopolymers of polyglu-

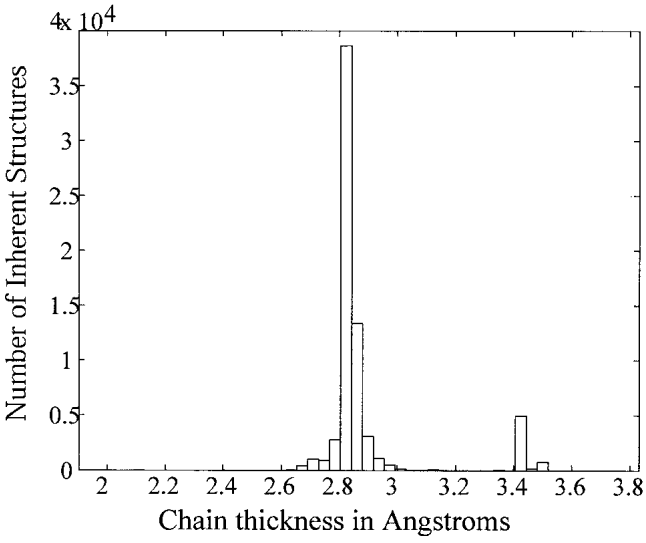

Fig. 12. Distribution of chain thickness values computed for enumerated inherent structures at T = 300K.

tamate and polylysine do not form stable α-helices in water at neutral pH. Consequently, these systems were considered appropriate model systems for unfolded polypeptides. Despite the fact that these chains were expected to mimic random-coil polypeptides, measurements from numerous experiments based on UV-circular dichroism and its predecessor, optical rotatory dispersion, suggested that the putative unordered peptides deviate significantly from the canonical statistical random coil (Kauzmann 1959; Schellman and Schellman 1964; Tiffany and Krimm 1968; Woody 1992). Based on the similarity of the observed CD spectrum for polyglutamate and polylysine at neutral pH to that observed for regular left-handed polyproline II (P_{II}) helices, Tiffany and Krimm (1968) proposed that purportedly unordered peptides or proteins are built of P_{II} segments linked by sharp bends.

Experimental evidence accumulated over the past three decades largely supports the hypothesis of Tiffany and Krimm (1968). Evidence comes from a variety of spectroscopic probes including vibrational spectroscopy (Dukor et al. 1991; Silva et al. 2000), Raman optical activity (ROA; Barron et al. 1997; Smyth et al. 2001), and polarized visible Raman and Fourier Transform IR spectroscopy (Woutersen and Hamm 2000; Schweitzer-Stenner et al. 2001; Asher et al. 2002).

Smyth et al. (2001) showed, using ROA, that a major conformational element in states of static and dynamically disordered peptides and proteins is the relatively *unconstrained* P_{II} helix. Rucker and Creamer (2002) recently revisited the issue of P_{II} helical CD spectra for lysine-based peptides. They studied a seven-residue lysine peptide, as opposed to the homopolymeric polylysine peptides studied in the 1960s, and found that this short peptide retains P_{II} helical CD signals over a range of salt concentrations where the charges are effectively neutralized or screened. They conclude that P_{II} helices must be preferred conformers for polypeptide backbones, a hypothesis that is in agreement with the experiments of Shi et al. (2002) and our calculations.

What is the physical basis for P_{II} helices in peptide unfolded states? The consensus is that P_{II} helices are preferred in water, the good solvent for short peptides, because hydrogen bonding promotes favorable chain-solvent contacts (Anderson et al. 1988; Woody 1992; Case and Brooks 1993; Han et al. 1998; Smith 1999; Sreerama and Woody 1999; Poon et al. 2000; Creamer and Rucker 2002; Shi et al. 2002).

We propose a simple "MiniMax" model to explain folding-unfolding transitions of simple peptides that appear to behave in a two-state fashion. According to our model, chain packing density, subject to constraints of chain connectivity and restrictions of excluded volume, is minimized (Mini) upon unfolding and maximized (Max) upon folding (Richards 1977). We implement this idea to assay the population of conformers preferred for unfolded polyalanine, using a purely repulsive inverse power potential.

We have shown that Flory's isolated-pair hypothesis (Flory 1969) is valid for all oligomer conformers generated from (ϕ,ψ)-values in the top left quadrant of a Ramachandran map (Pappu et al. 2000). Satisfying the isolated-pair hypothesis is a direct way to increase conformational entropy, for if uncorrelated fluctuations of chain monomers are feasible, then the number of accessible conformers is maximized. The question remains if these considerations continue to be valid in solvent. We hypothesize that the only conformers for which the isolated-pair hypothesis remains valid are those that are truly *noncooperative*, that is, those conformers for which individual (ϕ,ψ)-pairs fluctuate about P_{II} and sP_{II} values. In these conformers, solvent must be able to access each peptide group individually. Consequently, in the vicinity of P_{II}, both chain and solvent entropy are maximized. Preliminary evidence supporting this hypothesis comes from the experiments of Barron et al. (1997), although more experiments and detailed calculations of solvent relaxation around chain conformers—an extension of the work initiated by Sreerama and Woody (1999)—are needed.

According to our model, it appears that local structure in peptide unfolded states is dictated by constraints of chain connectivity, restrictions of excluded volume, and the drive to minimize the packing density of a chain around itself. Although the latter is expected to be a direct consequence of the chain's behavior in a good solvent, details of chain-solvent interactions are not necessary to direct the *specificity* of alanine-based peptides for P_{II} helices. However, these details are clearly important for *stabilizing* the chain in the vicinity of P_{II} helices. The distinction we make between specificity and stability resembles the Weeks-Chandler-Andersen picture for the arrangement of molecules in a simple liquid (Chandler et al. 1984) and mirrors previous suggestions along these lines for proteins and peptides (Lattman and Rose 1993; Richards and Lim 1994).

Our calculations are similar to that of Brant et al. (1967) with two crucial distinctions: first, we highlight the conceptual underpinning of inverse power potentials that allows us to envisage chain unfolding as a generalization of Richards' packing paradigm; secondly, our findings that polyalanine preferentially fluctuates about P_{II} and sP_{II} helices, in general agreement with recent experiments, supports the view that unfolded oligopeptides clearly deviate from random-coil behavior. Brant et al. did not find the marked preference for P_{II} in longer chains because they assumed the tenets of the rotational isomeric approximation (Mattice and Suter 1994), which ignores long-range interactions between chain monomers.

The emphasis on packing density is also manifest in recent ideas proposed by Petrescu et al. (2000), Banavar et al. (2002), and Rader et al. (2002). Rader et al. (2002) modeled

the unfolding of several two-state proteins as a spontaneous loss of chain rigidity. Unfolding is characterized in terms of a mean coordination number, a quantity analogous to chain packing density. All proteins studied in their work appear to have similar values for the mean coordination number at the midpoint of the folding-unfolding transition. This observation is consistent with the Tiffany and Krimm (1968) hypothesis that pieces of the unfolded polypeptide chain are likely to flicker in and out of P_{II} -like conformations irrespective of sequence.

Banavar et al. (2002) showed that geometrical considerations lead to a natural selection of secondary structure motifs for segments of protein backbones. In their description, compact secondary structures emerge as a consequence of maximizing chain thickness or conformational entropy measured as a local radius of curvature. We have established a connection to the ideas of Banavar et al. (2002) by showing that P_{II}-helices and related inherent structures are natural geometries when chain thickness is maximized subject to the constraint that packing density be minimized.

In recent work, Petrescu et al. (2000) analyzed computer-generated ensembles of highly denatured phosphoglycerate kinase that reproduced small-angle neutron-scattering profiles. Experimental scattering intensities were found to be strongly dependent on local (ϕ,ψ)-statistics of individual residues. Petrescu et al. also found that denaturation leads to a reduction in the atomic pair distribution function. This is consistent with uncorrelated fluctuations of individual residues which would result by minimization of chain packing. Petrescu et al. (2000) associated the observed reduction of the atomic pair-distribution functions with a discernible shift of backbone (ϕ,ψ)-angles toward the top left quadrant of the Ramachandran map, the region populated by fluctuations about P_{II} and sP_{II} (ϕ,ψ)-values.

Unlike the α-helix, the preferred conformer of short alanine-based peptides in water (Scholtz et al. 1991), the P_{II}-conformation lacks intrasegment hydrogen bonds; moreover individual residues fluctuate freely about P_{II} and sP_{II} (ϕ,ψ)-values. The dynamical disorder seen in experimental measurements for "random-coil" peptides (Merutka et al. 1995; Plaxco et al. 1996) may well be a consequence of the non-cooperative nature of P_{II} geometries (Adzhubei et al. 1987).

Experimental probes of increased sensitivity have been used to characterize conformations populated in the unfolded ensemble (Shortle 1996; Schwalbe et al. 1997; Petrescu et al. 2000; Hodsdon and Frieden 2001; Shortle and Ackerman 2001; Klein-Seetharaman et al. 2002). The extent of residual structure (Shortle and Ackerman 2001) and the nature of favorable nonnative contacts that accumulate under denaturing conditions (Koepf et al. 1999; Hodsdon and Frieden 2001; Klein-Seetharaman et al. 2002) are of particular interest. It remains to be seen whether the idea of minimizing chain packing density helps to explain new data for unfolded proteins that exhibit two-state behavior.

Methods

Values for bond lengths, bond angles, and hard-sphere contact distances are identical to the ones used in our previous work (Pappu et al. 2000). The peptide unit was held fixed in the *trans* configuration ($\omega = 179.5°$) and the N-C_α-C′ bond angle was fixed at 110°. Only interactions between nonbonded pairs of atoms were considered; all other interactions, including those between third-nearest neighbors, were ignored.

The inverse power potential energy (U) is calculated as a sum of pairwise interactions:

$$U = \sum_i \sum_{j<i} \left(\frac{\sigma_{ij}}{r_{ij}} \right)^n .$$

The summation runs over all unique pairs of nonbonded atoms; σ_{ij} is the hard-sphere contact distance and r_{ij} the interatomic separation. The inverse power potential exponent was set to $n = 12$ in all calculations. The energy landscape for the alanine dipeptide (Fig. 2) is qualitatively similar to the $n = 12$ case for a range of exponents between $n = 9$ and $n = 14$. Hard-sphere contact distances (σ_{ij}) and interatomic separations (r_{ij}) are measured in Å, so (σ_{ij}/r_{ij}) is dimensionless. Each pairwise interaction contributes ε_{ij} (σ_{ij}/r_{ij})^{12} to the total energy. For a pair of atoms, the ε_{ij} parameters can be calculated using the Slater-Kirkwood (Slater and Kirkwood 1931) formula, in which ε_{ij} varies as

$$\frac{\alpha_i \alpha_j (\sigma_{ij})^{-6}}{\sqrt{\frac{\alpha_i}{N_i}} + \sqrt{\frac{\alpha_j}{N_j}}};$$

α_i and α_j are the static polarizabilities (in units of Å^3), and N_i and N_j are the number of outer-shell electrons for atoms i and j, respectively. Polarizability increases with atomic radius, and $N_{i(j)}$ increases with atomic number. Nonpolar atoms such as carbon and hydrogen are more polarizable than nitrogen and oxygen atoms. Using standard values for atomic polarizabilities (Lide 2001), our hard-sphere contact distances (σ_{ij}), and values for $N_{i(j)}$ from Pitzer (1959), the calculated ε_{ij}-values result in energy landscapes that are qualitatively similar to the ones obtained by simply setting ε_{ij} = 1kcal/mole for all atom pairs. Units are needed to calibrate Boltzmann weights in terms of RT (R = 1.987 cal/mole/K; T = temperature in Kelvin). The decision to set ε_{ij} to 1kcal/mole for all interatomic pairs means that the two temperatures (100K and 300K) used here are not physical temperatures.

Two-step minimizations

Each starting conformation was initially subjected to steepest descent minimization (Luenberger 1984), perturbing the conformer toward the "trust region" of a nearby local minimum. Within the trust region, an optimally conditioned variable metric minimization method without line searches (Shanno and Phua 1978) was used to accelerate convergence. The convergence criterion was 1.0kcal /mole/degree-radian for the steepest-descent step and 10^{-5} kcal/mole/degree-radian for the final step. The two-step minimization is required to ensure that minimizations which start from conformers with extremely high energies converge to the structurally similar inherent structures. This cannot be guaranteed with nonsteepest-descent optimizers. Routines for the two-step procedure are available from Jay Ponder's TINKER molecular design suite at http://dasher.wustl.edu/~tinker.

Conformational enumeration for the dipeptide

The protocol of generating a large number of conformers by randomly sampling (ϕ,ψ)-values between $-180°$ and $180°$ is termed simple sampling (Binder and Heerman 1997). Unlike the widely used Markov Chain Metropolis Monte Carlo sampling (Metropolis et al. 1953), which generates a sequence of interdependent conformers by sampling from the Boltzmann distribution, in simple sampling one relies on the generation of a large number of *statistically independent* conformers to cover the expanse of conformational space. It is well known that simple sampling is inefficient for larger problems (Binder and Heerman 1997). Even for the dipeptide, simple sampling cannot be used to compute the canonical partition function. Instead we compute a configurational sum that is simply the sum of Boltzmann-weighted energies of all sampled conformers.

We calculated the weight for each inherent structure basin as a function of the number of sampled conformers. The standard error for simple sampling is inversely proportional to the number of enumerated conformers. We found that upward of 10^8 conformers was required to get the individual basin weights (Fig. 5) to converge, where convergence implies that the weights do not change appreciably if fewer conformers were generated.

Stratified importance sampling for the 7-mer

The goal is to generate a total of M_c statistically independent conformers that represent the equilibrium thermodynamic population. Sampling is restricted to pockets in multidimensional space generated by combinations of the five dipeptide basins shown in Figure 6. For the 7-mer, there are 5^7 basins that span all the relevant space. Within each of the 5^7 basins, *statistically independent conformers* are generated according to the following protocol: The number density of points, d_η (the fraction of points within a dipeptide basin shown in Fig. 3) is different for different dipeptide basins. Furthermore, the number density of thermodynamically relevant points within each basin (the fraction of points required to account for 99.9% of the dipeptide configurational sum at temperature T) varies with temperature (Fig. 6). Therefore, $d_{\alpha\eta}(T)$, the fraction of points required to account for the populations shown in Figure 6 within each dipeptide basin α at temperature T is first computed. Next, 5^7 different sampling boundaries (η) were delineated in the 14-dimensional space of the 7-mer using the envelopes of temperature-dependent dipeptide basins (α). M_η-conformers are generated within the sampling boundary η using:

$$M_\eta = M_c \prod_{\alpha=1}^{7} d_{a\eta}(T);$$

$d_{a\eta}(T)$ is the number density of points within each dipeptide basin α that makes up the sampling boundary η. The numbers M_η satisfy the constraint:

$$M_c = \sum_{\eta=1}^{5^7} M_\eta.$$

Each of the sampled conformers was subjected to the two-step minimization procedure. Basin weights were constructed using the results of the configurational mapping and energies of the sampled conformers. After repeated trials we found $M_c = 5\times10^9$ to be adequate to generate converged weights for the inherent-structure basins of the 7-mer at both T = 100K and T = 300K.

The sampling technique is based upon the stratified Monte Carlo method of Lepage (1978). The basic idea is to estimate the configurational sum by subdividing the conformational space of interest into smaller pockets, and concentrating the sampling to regions where the configurational sum is both large and rapidly varying. This method of stratified importance sampling relies on the generation of a large number of statistically independent conformers, yet alleviates the severe convergence problems that plague simple sampling. However, it does require sufficient a priori knowledge of the conformational space to be sampled. This knowledge is available as a combination of exhaustive sampling of the dipeptide and the observation that interactions inaccessible to the dipeptide lead to high-energy conformers for the longer chain.

Calculation of thermodynamic averages

Two types of thermodynamic averages may be computed using the sampled distribution of conformers. For a specific quantity of interest A, the annealed average is:

$$\langle A\rangle = \frac{\sum_{i=1}^{M_c} A_i \exp(-\beta U_i)}{\sum_{i=1}^{M_c} \exp(-\beta U_i)},$$

where $M_c = 5\times10^9$ denotes the total number of sampled conformers, U_i is the energy associated with conformer i, $\beta = 1/RT$, T is the desired temperature, and R = 0.00199 kcal/mole/K is the gas constant. Quenched averages use conformers of the inherent structures alone. Accordingly for the same quantity A, the quenched average is calculated as:

$$\langle A\rangle_Q = \frac{\sum_{i=1}^{M_{IS}} \{A_i^{IS} \exp(-\beta U_i^{IS}) \sum_{j\in i\}} \exp(-\beta \Delta U_{ij}^{IS})\}}{\sum_{i=1}^{M_c} \exp(-\beta U_i)},$$

where $U^{IS}{}_i$ refers to the energy of the inherent structure i, M_{IS} is the total number of inherent structures found from configurational mapping, and $\Delta U^{IS}{}_{ij}$ is the difference between the energy of the sampled conformer j and its inherent structure. The inherent structure formalism becomes very useful for calculations that require numerous pairwise comparisons of large numbers of sampled conformers.

Calculation of scalar coupling constants

Scalar coupling constants were calculated using the Karplus equation $^3J = A\cos^2\theta + B\cos\theta + C$ and parameters available from the literature. The method and parameters A, B, and C we used are identical to those of Smith et al. (1996). The thermodynamic average for a particular coupling constant is calculated as outlined above. Annealed ^3J values for a given angle $\theta\equiv\phi$ or $\theta\equiv\psi$ by calculating ^3J values for each of the 5×10^9 sampled conformers and then computing the average and variance as discussed above. Alternatively, the quenched average is calculated using (ϕ,ψ)-angles from inherent structures and the weights associated with these inherent structures.

Acknowledgments

We thank Robert Baldwin, Jayanth Banavar, Trevor Creamer, Elliot Elson, Carl Frieden, Alan Grossfield, Neville Kallenbach, Michael Paulaitis, Kevin Plaxco, and Christopher Roberts for many helpful discussions. We are especially grateful to Jayanth Banavar, Alan Grossfield, and Michael Paulaitis for careful reading of the manuscript. G.D.R. acknowledges support from the NIH (GM29458) and the Mathers Foundation. R.V.P. is supported by a start-up grant from the Whitaker foundation to the Department of Biomedical Engineering at Washington University in St.Louis.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Adzhubei, A.A., Eisenmenger, F., Tumanyan, V.G., Zinke, M., Brodzinski, S., and Esipova, N.G. 1987. Third type of secondary structure: Non-cooperative mobile conformation. Protein data bank analysis. *Biochem. Biophys. Res. Commun.* **146:** 934–938.

Anderson, A.G. and Hermans, J. 1988. Microfolding—Conformational probability map for the alanine dipeptide in water from molecular-dynamics simulations. *Proteins: Struc. Func. Genet.* **3:** 262–265.

Anfinsen, C.B. 1973. Principles that govern the folding of protein chains. *Science* **181:** 223–230.

Apostolakis, J., Ferrara, P., and Caflisch. A. 1999. Calculation of conformational transitions and barriers in solvated systems: Application to the alanine dipeptide in water. *J. Chem. Phys.* **110:** 2099–2108.

Asher, S.A., Ianoul, A., Mix, G., Karnoup, A., Diem, M., and Schweitzer-Stenner, R. 2001. Dihedral psi angle dependence of the amide III vibration: A uniquely sensitive UV resonance Raman secondary structural probe. *J. Am. Chem. Soc.* **123:** 11775–11781.

Banavar, J.R., Maritan, A., Micheletti, C., and Trovato, A. 2002. Geometry and physics of proteins. *Proteins: Struct. Func. Genet* **47:** 315–322.

Barron, L.D., Hecht, L., and Wilson, G. 1997. The lubricant of life: A proposal that solvent water promotes extremely fast conformational fluctuations in mobile heteropolypeptide structure. *Biochemistry* **43:** 13143–13147.

Binder, K. and Heerman, D.W. 1997. Monte Carlo simulations in statistical physics. In *Springer Series in Solid State Sciences*, 3rd ed., **80:**

Brant, D.A., Miller, W.G., and Flory, P.J. 1967. Conformational energy estimates for statistically coiling polypeptide chains. *J. Mol. Biol.* **23:** 47–65.

Brooks, C.L. and Case, D. 1993. Simulations of peptide conformational dynamics and thermodynamics. *Chem. Rev.* **93:** 2487–2502.

Cahn, R.S., Ingold, C., and Prelog, V. 1966. Specification of molecular chirality. *Angew. Chem. Int. Ed. Engl.* **5:** 385–415.

Chan, H.S. and Dill, K.A. 1991. Polymer principles in protein structure and stability. *Annu. Rev. Biophys. Biophys. Chem.* **20:** 447–490.

Chandler, D. 1987. *Introduction to modern statistical mechanics.* Oxford University Press, New York.

Chandler, D., Weeks, J.D., and Andersen, H.C. 1983. Van der Waals picture of liquids, solids, and phase transformations. *Science* **220:** 787–794.

deGennes, P.G. 1976. *Scaling concepts in polymer physics*, Chapter III, Cornell University Press, Ithaca.

Dill, K.A. and Shortle, D. 1991. Denatured states of proteins. *Annu. Rev. Biochem.* **60:** 795–825.

Dobson, C.M. 1999. Protein misfolding, evolution and disease. *TIBS* **24:** 329–332.

Dunker, A.K., Lawson, J.D., Brown, C.J., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C-H., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., and Orabdovic, Z. 2001. Intrinsically disordered proteins. *J. Mol. Graph. Model.* **19:** 26–59.

Dukor, R.K., Keiderling, T.A., and Gut, V. 1991. Vibrational circular-dichroism spectra of unblocked proline oligomers. *Int. J. Pept. Protein Res.* **38:** 198–203.

Flory, P.J. 1953. *Principles of polymer chemistry*, Chapter X. Cornell University Press, Ithaca.

Flory, P.J. 1969. *Statistical mechanics of chain molecules*, Chapter 7, Hanser Publishers, Munich.

Ginsburg, A. and Carroll, W.R. 1965. Some specific ion effects on the conformation and thermal stability of ribonuclease. *Biochemistry* **4:** 2159–2174.

Grant, J.A., Williams, R.L., and Scheraga, H.A. 1990. Ab initio self-consistent field and potential-dependent partial equalization of orbital electronegativity calculations of hydration properties of N-acetyl-N′-methyl-alanineamide. *Biopolymers* **30:** 929–949.

Han, W-G., Jalkanen, K.J., Elstner, M., and Suhai, S. 1998. Theoretical study of aqueous *N*-Acetyl-L-alanine *N*′-Methylamide: Structures and Raman, VCD, and ROA spectra. *J. Phys. Chem. B* **102:** 2587–2602.

Hodsdon, M.E. and Frieden, C. 2001. Intestinal fatty acid binding protein: The folding mechanism as determined by NMR studies. *Biochemistry* **40:** 732–742.

Holzwarth, G. and Doty, P. 1965. The ultraviolet circular dichroism of polypeptides. *J. Am. Chem. Soc.* **87:** 218–228.

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.

Jackson, S.E. 1998. How do small single domain proteins fold? *Fold. Des.* **3:** R81–R91.

Jalkanen, K.J. and Suhai, S. *N*-Acetyl-L-alanine *N*′-methylamide: A density functional analysis of the vibrational absorption and vibrational circular dichroism spectra. *Chem. Phys.* **208:** 81–116.

Karplus, M. 1959. Contact electron-spin coupling of nuclear magnetic moments. *J. Chem. Phys.* **30:** 11–15

Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14:** 1–63.

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.

Koepf, E.K., Petrassi, H.M., Sudol, M., and Kelly, J.W. 1999. WW: An isolated three-stranded antiparallel β-sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHcl and a disordered thermal unfolded state. *Protein Sci.* **8:** 841–853.

Lattman, E.E. and Rose, G.D. 1993. Protein folding: What's the question? *Proc. Natl. Acad. Sci.* **90:** 439–441.

Lepage, G.P. 1978. A new algorithm for adaptive multidimensional integration. *J. Comput. Phys.* **27:** 192–203.

Lide, D.R. 2001. *CRC Handbook of Chemistry and Physics*, Chap. 5, CRC Press.

Luenberger, D.G. 1984. *Linear and nonlinear programming*, 2nd ed., Chap. 9, Addison-Wesley, Reading, MA.

Matsumoto, I., Jimbo, A., Mizuno, Y., Seno, N., and Jeanloz, R.W. 1983. Purification and characterization of potato lectin. *J. Biol. Chem.* **258:** 2886–2891.

Mattice, W.L. and Suter, U.W. 1994. *Conformational theory of large molecules. The rotational isomeric state model in macromolecular systems*. John Wiley & Sons, New York.

Merutka, G., Dyson, H.J., and Wright, P.E. 1995. Random coil H-1 chemical-shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. *J. Biomol. NMR.* **5:** 14–24.

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.

Mu, Y. and Stock, G. 2002. Conformational dynamics of trialanine in water: A molecular dynamics study. *J. Phys. Chem. B.* **106:** 5294–5301.

Neuhaus, D. and Williamson, M.P. 2000. *The Nuclear overhauser effect in structural and conformational analysis*, Chapter 8, Wiley-VCH, New York.

O'Connell, T.M., Wang, L., Tropsha, A., and Hermans, J. 1999. The "Random-Coil" state of proteins: Comparisons of database statistics and molecular simulations. *Proteins: Struct. Func. Genet.* **36:** 407–418.

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.* **97:** 12565–12570.

Pardi, A., Billiter, M., and Wüthrich, K. 1984. Calibration of the angular dependence of the amide protein-C^{α} proton coupling constants, $^{3}JHN_{\alpha}$, in a globular protein. *J. Mol. Biol.* **180:** 741–751.

Pauling, L., Corey, R.B., and Branson, H.R. 1951. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci.* **37:** 205–211.

Petrescu, A-J., Calmettes, P., Durand, D., Receveur, V., and Smith, J.C. 2000. Change in backbone torsion angle distribution on protein folding. *Protein Sci.* **9:** 1129–1136.

Pettitt, B.M. and Karplus, M. 1988. Conformational free-energy of hydration for the alanine dipeptide—Thermodynamic analysis. J. Phys. Chem. **92:** 3994–3997.

Pieroti, R.A. 1965. Aqueous solutions of nonpolar gases. *J. Phys. Chem.* **69:** 281–288.
Pitzer, K.S. 1959. Inter- and intramolecular forces and molecular polarizability. *Adv. Chem. Phys.* **2:** 59–83.
Plaxco, K.W. and Gross, M. 1997. The importance of being unfolded. *Nature* **386:** 657–659.
Plaxco, K.W., Morton, C.J., Grimshaw, S.B., Jones, J.A., Pitkeathly, M., Campbell, I.D., and Dobson C.M. 1996. The effects of guanidine hydrochloride on the "random coil" conformations and NMR chemical shifts of the peptide series GGXGG. *J. Biomol. NMR* **10:** 221–230.
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.
Rader A.J., Hespenheide B.M., Kuhn L.A., and Thorpe, M.F. 2002. Protein unfolding: Rigidity lost. *Proc. Natl. Acad. Sci.* **99:** 3540–3545.
Ramachandran, G.N. and Sasisekharan, V. 1968. Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23:** 283–437.
Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. 1963. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7:** 95–99.
Richards, F.M. 1977. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **6:** 151–176.
Richards, F.M. and Lim, W.A. 1994. An analysis of packing in the protein folding problem. *Q. Rev. Biophys.* **26:** 423–498.
Roterman, I.K., Gibson, K.D., and Scheraga, H.A. 1989. A comparison of the CHARMM, AMBER and ECEPP potentials for peptides. 2. ϕ-ψ maps for the A-Acetyl-Alanine N′-Methyl Amide. Comparisons, contrasts and simple experimental tests. *J. Biomol. Struct. Dyn.* **3:** 421–453.
Rossky, P.J., Karplus, M., and Rahman, A. 1979. A model for the simulation of an aqueous dipeptide solution. *Biopolymers* **18:** 825–854.
Rucker, A.L. and Creamer, T.P. 2002. Polyproline II helical structure in protein unfolded states: Lysine peptides revisited. *Protein Sci.* **11:** 980–985.
Schellman, J.A., and Schellman, C.G. 1964. In *The poteins*, 2nd ed. (ed. H. Neurath), Vol. 2, pp. 1–137. Academic Press, New York.
Schmidt A.B. and Fine, R.M. 1994. A CFF91-based continuum solvation model: Solvation free energies of small organic molecules and conformations of the alanine dipeptide in solution. *Molec. Simulat.* **6:** 347–365.
Scholtz, J.M., Qian, H., York, E.J., Stewart, J.M., and Baldwin, R.L. 1991. Parameters of helix-coil transition theory for alanine based peptides of varying chain lengths in water. *Biopolymers* **31:** 1463–1470.
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.
Serrano, L. 1995. Comparison between the ϕ distribution of the amino acids in the protein database and NMR data indicates that amino acids have various ϕ propensities in the random coil conformation. *J. Mol. Biol.* **254:** 322–333.
Shanno, D.F. and Phua, K.H. 1978. Numerical comparison of several variable-metric algorithms. *J. Optim. Theor. Appl.* **25:** 507–518.
Shi, Z., Olson, C. A., Rose, G.D., Baldwin, R.L., and Kallenbach, N.R. 2002. Polyproline II structure in a sequence of 7-residue alanine residues. *Proc. Natl. Acad. Sci.* (in press).
Shortle, D. 1996. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* **10:** 27–34.
Shortle, D. and Ackerman, M.S. 2001. Persistence of native-like topology in a denatured protein in 8M urea. *Science* **293:** 487–489.
Silva, R. A. G. D., Sherman, S. A., Perini, F., Bedows, E., and Keiderling, T. A. 2000. Folding studies on the human chorionic gonadotropin β-subunit using optical spectroscopy of peptide fragments. *J. Am. Chem. Soc.* **122:** 8623–8630.
Slater, J.C. and Kirkwood, J.G. 1931. The van der Waals forces in gases. *Phys. Rev.* **37:** 682–696.
Smith, L.J., Bolin, K.A., Schwalbe, H., MacArthur, M.W., Thornton, J.M., and Dobson, C.M. 1996. Analysis of main chain torsion angles in proteins: Prediction of NMR coupling constants for native and random coil conformations. *J. Mol. Biol.* **255:** 494–506.
Smith, P.E. 1999. The alanine dipeptide free energy surface in solution. *J. Chem. Phys.* **111:** 5568–5579.
Smyth, E., Syme, C.D., Blanch, E.W., Hecht, L., Vašák, M., and Barron, L.D. 2001. Solution structure of native proteins with irregular folds from Raman Optical Activity. *Biopolymers* **58:** 138–151.
Sreerama, N. and Woody, R.W. 1999. Molecular dynamics simulations of polypeptide conformations in water: A comparison of α, β and poly(Pro)II conformations. *Proteins: Struct. Func. Genet.* **36:** 400–406.
Stillinger, F.H. and Weber, T.A. 1984. Packing structures and transitions in liquids and solids. *Science* **225:** 983–989.
Stillinger, F.H. and Weber, T.A. 1985. Inherent structure theory of liquids in the hard-sphere limit. *J. Chem. Phys.* **83:** 4767–4775.
Tanford, C. 1968. Protein denaturation. *Adv. Prot. Chem.* **23:** 121–282.
Tiffany, M.L. and Krimm, S. 1968. New chain conformations of poly(glutamic acid) and polylysine. *Biopolymers* **6:** 1379–1382.
Tobias, D.J. and Brooks, C.L. 1992. Conformational equilibrium in the alanine dipeptide in the gas phase and aqueous solution: A comparison of theoretical results. *J. Phys. Chem.* **96:** 3864–3870.
Tsai, C-J., Ma, B., Sham, Y.Y., Kumar, S., and Nussinov, R. 2001. Structured disorder and conformational selection. *Proteins: Struct. Func. Genet.* **44:** 418–427.
Uversky, V.N. 2002. Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.* **11:** 739–756.
van Günsteren, W.F., Burgi, P., Peter, C., and Daura, X. 2001. The key to solving the protein-folding problem lies in an accurate description of the denatured state. *Angew. Chemie Int. Ed.* **40:** 351–355.
Vuister, G.W. and Bax, A. 1993. Quantitative *J* correlation. A new approach for measuring homonuclear $J(H^NH^\alpha)$ coupling constants in the ^{15}N-enriched proteins. *J. Am. Chem. Soc.* **115:** 7772–7777.
Wang, C., and Bax, A. 1995. Reparameterization of the Karplus relation for $^3J(H^\alpha\text{-}N)$ and $^3J(H^N\text{-}C')$ in peptides from uniformly 13C/15N-enriched human ubiquitin. *J. Am. Chem. Soc.* **117:** 1810–1813.
Wilson, G., Hecht, L., and Barron, L.D. 1996. Residual structure in unfolded proteins revealed by Raman Optical Activity. *Biochemistry* **35:** 12518–12525.
Woody, R.W. 1992. Circular dichroism of unordered polypeptides. *Adv. Biophys. Chem.* **2:** 37–79.
Wright, P.E., and Dyson, H.J. 1999. Intrinsically unstructured proteins: Reassessing the structure-function paradigm. *J. Mol. Biol.* **41:** 415–427.
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.
Wüthrich, K. 1986. *NMR of proteins and nucleic acids*, Chapter 9, John Wiley & Sons, New York.
Zhou, H-X. and Dill, K.A. 2001. Stabilization of proteins in confined spaces. *Biochemistry* **40:** 11290–11293.