

20 Conformational Properties of Unfolded Proteins

Patrick J. Fleming and George D. Rose

20.1 Introduction

The protein folding reaction, U(nfolded) \rightleftharpoons N(ative), is a reversible disorder \rightleftharpoons order transition. Proteins are disordered (U) at high temperature, high pressure, extremes of pH, or in the presence of denaturing solvents; but they fold to uniquely ordered, biologically relevant conformers (N) under physiological conditions. This folding transition is highly cooperative such that individual molecules within the population are predominantly fully folded or fully unfolded; partially folded chains are transitory and rare. Notably, no covalent bonds are made or broken during folding/unfolding; in effect, the transition is simply a re-equilibration in response to changes in temperature, pressure, pH, or solvent conditions. Currently, there are more than 20 000 examples of native proteins in the protein databank. In contrast, the unfolded population, by its very nature, resists ready structural characterization. In this sense, the folding reaction might be more appropriately denoted as $\blacksquare \rightleftharpoons N$.

This chapter traces thinking about the unfolded state from Pauling's and Wu's early suggestions in the 1930s, through the work of Tanford and Flory in the 1960s to the present moment. Early work gave rise to the random coil model for the unfolded state, as described below. Confirmatory findings established this model as the conceptual anchor point for thinking about unfolded proteins – until recently, perhaps. In the past few years, results from both theory and experiment indicate the existence of conformational bias in the unfolded state, a condition that is not addressed by the random coil model. If unfolded conformers are biased toward their native conformation sufficiently, then the random coil model is likely to be superseded by newer, more specific models. Though controversial, such a conceptual shift appears to be underway, as we attempt to describe.

20.1.1 Unfolded vs. Denatured Proteins

The term unfolded protein is generic and inclusive, and it can range from protein solutions in harsh denaturants to protein subdomains that undergo transitory ex-

Protein Folding Handbook. Part I. Edited by J. Buchner and T. Kiefhaber
Copyright © 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30784-2

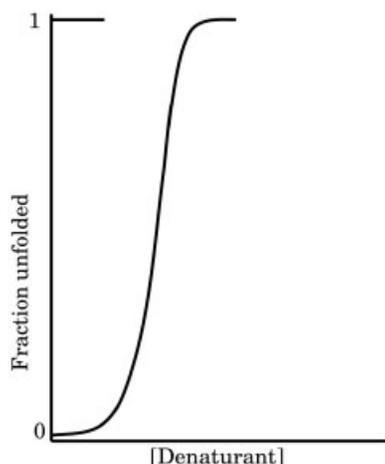


Fig. 20.1. The folding transition. The folding reaction of a typical, small biophysical protein is a highly cooperative, all-or-none transition. At the transition midpoint, half the ensemble is folded and half is unfolded; the population of partially folded/unfolded molecules is negligible. In this idealized plot of an actual

experiment (Ginsburg and Carroll 1965), the population is followed by a conformational probe (e.g., circular dichroism or fluorescence) as a function of denaturant concentration. Upon addition of sufficient denaturant, the probe signal reaches a plateau, indicating that the transition is complete.

cursions from their native format via spontaneous fluctuations. While conceptually complete, this range is too diverse to be practically useful, and it requires further specification. Accordingly, the field has focused more specifically on denatured proteins, the population of unfolded conformers that can be studied at equilibrium under high concentrations of denaturing solvents, high temperature, high pressure, high/low pH, etc. Early experiments of Ginsburg and Carroll (Ginsburg and Carroll 1965) and Tanford (Tanford et al. 1966) demonstrate that such conditions can give rise to a defined equilibrium population in which the unfolding transition is complete (Figure 20.1). In this chapter, we use both terms and rely on the context for specificity.

20.2 Early History

The fact that protein molecules can undergo a reversible disorder \rightleftharpoons order transition originated early in the last century, in ideas proposed by Wu (Wu 1931; Edsall 1995) and Mirsky and Pauling (Mirsky and Pauling 1936). Both papers propose that a theory of protein structure is tantamount to a theory of protein denaturation. In particular, these authors recognized that many disparate physical and chemical properties of proteins are abolished coordinately upon heating. This was unlikely to be mere coincidence. Both Wu and Mirsky and Pauling hypothesized that such properties are a consequence of the protein's structure and are abolished when that

structure is melted. Their hypothesis was later confirmed by Kauzmann and Simpson (Simpson and Kauzmann 1953), at which point the need for an apt characterization of the melt became clear, and protein denaturation emerged as a research discipline.

A widely accepted view assumes that unfolded polypeptide chains can explore conformational space freely, with constraints arising only from short-range local restrictions and longer range excluded volume effects. To a good first approximation, short-range local restrictions refer to repulsive van der Waals interactions between sequentially adjacent residues (i.e., steric effects) captured by the well-known Ramachandran map for a dipeptide (Ramachandran and Sasisekharan 1968). Longer range excluded volume effects also refer to repulsive van der Waals interactions, in this case those between nonbonded atoms that are distant in sequence but juxtaposed in space as the chain wanders at random along a Brownian walk in three dimensions (Flory 1953; Tanford 1968). This random coil model has conditioned most of the thinking in the field.

It is important to realize that the random coil model need not imply an absence of residual structure in the unfolded population. Kauzmann's famous review raised the central question about structure in the unfolded state explicitly (Kauzmann 1959):

■ *For instance, one would like to know the types of structures actually present in the native and denatured proteins. . . . The denatured protein in a good solvent such as urea is probably somewhat like a randomly coiled polymer, though the large optical rotation of denatured proteins in urea indicates that much local rigidity must be present in the chain (p. 4).*

20.3 The Random Coil

A chain molecule is a freely jointed random coil if it traces a random walk in three-dimensional space, in incremental steps of fixed length. The random coil model has enjoyed a long and successful history in describing unfolded proteins. By definition, a random coil polymer has no strongly preferred backbone conformations because energy differences among its sterically accessible backbone conformations are of order $\sim kT$. Accordingly, the energy landscape for such a polymer can be visualized as an "egg crate" of high dimensionality, and a Boltzmann-weighted ensemble of such polymers populates this landscape uniformly.

More than others, this elegant theory was developed by Flory (Flory 1969, pp. 30–31; Cantor and Schimmel 1980, pp. 991–996) and advanced by Tanford (Tanford et al. 1966; Tanford 1968, 1970), who demonstrated that proteins denatured in 6 M guanidinium chloride (a strong denaturant) appear to be structureless, random chains. Tanford's pioneering studies established a compelling framework for interpreting experimental protein denaturation that would survive largely unchallenged for the next 30 years.

Often, the term random coil is used synonymously with the freely jointed chain model (described below), in which there is no correlation between the orientation of two chain monomers at any length scale. That is, configurational properties of a freely jointed chain, such as its end-to-end distance, are Gaussian distributed at all chain lengths. In practice, no actual polymer chain is freely jointed. More realistic models, such as Flory's rotational isomeric-state model, have Gaussian-distributed chain configurations only in the infinite chain limit (Flory 1969, pp. 30–31; Cantor and Schimmel 1980, pp. 991–996). These distinctions notwithstanding, the main characteristic of the random coil holds in all cases, both ideal and real: the unfolded state is structurally featureless because the number of available conformers is large and the energy differences among them are small.

20.3.1

The Random Coil – Theory

Statistical descriptions are the natural way to characterize a large heterogeneous population, such as an ensemble of unfolded proteins. A few key ideas are mentioned here, but they are no substitute for the many excellent treatments of this subject (Flory 1953, 1969; Cantor and Schimmel 1980; Chan and Dill 1991; Dill and Shortle 1991).

The fundamental model is the freely jointed chain (or freely jointed random coil or random flight), a linear polymer of n adjoining links, each of fixed length, with complete freedom of rotation at every junction (Figure 20.2). From this definition, it follows that the angles at link junctions (i.e., bond angles) are completely uncorrelated. This model is completely general because it neglects chemical constraints, and therefore its scope is not restricted to any particular type of polymer chain. However, additional constraints such as chain thickness or hindered bond rotation can be added as appropriate, resulting in more specific models. What can be said about a polymer chain that is devoid of chemistry?

The freely jointed chain is equivalent to Brownian motion with a mean free path of fixed length, as described by Einstein-Smoluchowski theory (Einstein 1956). The

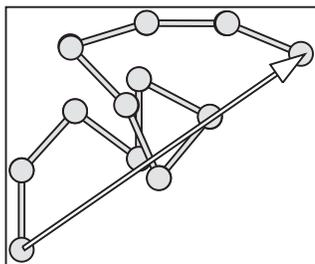


Fig. 20.2. A freely jointed chain. The chain is comprised of links, each of fixed length, l , with freedom of rotation at every junction. For a chain of n links, the vector from the beginning to the end, \vec{r}_n , (shown as the long arrow) is given by summing the links, \vec{r}_i : $\vec{r}_n = \sum_{i=1}^n \vec{r}_i$ and $|\vec{r}_i| = l$.

basic relationship governing both freely jointed chains and Brownian particles is:

$$\sqrt{\langle r^2 \rangle} = l\sqrt{n} \quad (1)$$

where $\sqrt{\langle r^2 \rangle}$ is the root-mean-square end-to-end distance (see Figure 20.2), l is the link length, and n is the number of links in the polymer. In other words, the distance between termini increases as the square root of the number of chain links: doubling the distance requires four times as many links.

The end-to-end distance measures the size of a polymer coil. Another such measure is the radius of gyration, R_G , the root-mean-square distance of link termini from their common center of gravity:

$$R_G^2 = \left[\frac{1}{n+1} \right] \sum_{i=0}^n R_{Gi}^2 \quad (2)$$

where R_{Gi} is the distance of link i from the center of gravity and n is the number of links in the polymer. According to a theorem of Lagrange in 1783, R_G can be rewritten in terms of the individual link vectors, r_{ij} , without explicit reference to the center of gravity (Flory 1969, appendix A).

$$R_G^2 = \frac{1}{2n^2} \sum_{i=1}^N \sum_{j=1}^N r_{ij}^2 \quad (3)$$

The two measures are related:

$$\langle R_G^2 \rangle = \frac{\langle r^2 \rangle}{6} \quad \text{as } n \rightarrow \infty \quad (4)$$

For a freely jointed chain, the values of such configurational measures are Gaussian distributed.

Of course, no real chain is freely jointed. The chemical bonds in real chains restrict motion; bond rotations are never random. Also, each link of a real chain occupies a finite volume, thereby reducing the free volume accessible to remaining links. Accordingly, ideal chains descriptions must be modified if they are to accommodate such real-world constraints.

A strategy for accommodating restricted bond motion is to depart from physical chain links and instead re-represent the chain as though it were comprised of longer, uncorrelated virtual links. The idea underlying this strategy is as follows: a short chain segment (e.g., a dipeptide) is somewhat rigid (Ramachandran and Sasisekharan 1968), but a sufficiently long segment is flexible. Therefore, the chain becomes flexible at some length between the dipeptide and the longer peptide. This leads to the idea of an effective segment, $l_{\text{effective}}$, also called a Kuhn segment, the length scale at which chain segments approach independent behavior and cor-

related orientations between them dwindle away. A chain of length L contains $L/l_{\text{effective}}$ Kuhn segments and can be approximated as a freely jointed chain of Kuhn segments:

$$\langle r^2 \rangle = l_{\text{effective}}^2 \frac{L}{l_{\text{effective}}} = l_{\text{effective}} L \quad (5)$$

A closely related idea is defined in terms of the chain's persistence length, the length scale over which correlations between bond angles "persist". In effect, the chain retains a "memory" of its direction at distances less than the persistence length. Stated less anthropomorphically, the energy needed to bend the chain through a 90° angle diminishes to $\sim kT/2$ at its persistence length, and thus ambient-temperature fluctuations can randomize the chain direction beyond this length. The size of a Kuhn segment is approximately two persistence lengths (i.e., directional correlations die away in either direction).

Current models strive to capture the properties of real chains with more detail than idealized, freely jointed chains can provide. For example, no actual chemical bond is a freely swiveling joint. To treat bond restrictions more realistically, Flory devised the rotational isomeric state approximation (Flory 1969, p. 55), in which bond angles are restricted to discrete values, chosen to correspond to known potential minima (e.g., gauche⁺, gauche⁻, and trans).

A real polymer chain cannot evade itself. Inescapably, the volume occupied by a chain element is excluded from occupancy by other chain elements. Otherwise, a steric clash would ensue: nonbonded atoms cannot occupy the same space at the same time. This excluded volume effect is substantial for proteins and results in a major departure from ideal chain dimensions (Eqs (1)–(4)).

As real polymers fluctuate, contracted coils have more opportunities to experience excluded volume steric clashes than expanded coils, perturbing the chain toward larger dimensions than those expected for ideal polymers.

Chain dimensions are also perturbed by the nature of the solvent. A good solvent promotes chain expansion by favoring chain:solvent interactions over chain:chain interactions. Conversely, a poor solvent promotes chain contraction by favoring chain:chain interactions over chain:solvent interactions. Flory introduced the idea of a Θ -solvent in which, on average, chain:chain interactions exactly counterbalance chain:solvent interactions, leading to unperturbed chain behavior. He pointed out that the notion of a Θ -solvent for a liquid is analogous to the Boyle point for a real gas, the temperature at which a pair of gas molecules follow an ideal isotherm because repulsion arising from volume exclusion is compensated exactly by mutual attraction (Flory 1969, p. 34).

Flory provided a simple relationship that relates the coil dimensions to solvent quality (Flory 1969). For a random coil polymer with excluded volume, the radius of gyration, R_G , is given by:

$$R_G = R_0 n^\nu \quad (6)$$

where R_0 is a constant that is a function of the chain's persistence length, n is the number of links, and ν is the scaling factor of interest that depends on solvent quality. Values of ν range from 0.33 for a collapsed, spherical molecule in poor solvent through 0.5 at the Θ -point (Eq. (1)) to 0.6 in good solvent.

Protein molecules are amphipathic, and their interactions with solvent are complex. However, on balance, denaturing agents such as urea and guanidinium chloride can be considered good solvents. Using Eq. (6), the degree to which unfolded proteins are random coil polymers in denaturing solvents can be assessed by measuring ν , the main topic of Section 20.3.2.

20.3.1.1 The Random Coil Model Prompts Three Questions

The random coil model set the stage for much of the contemporary theoretical work on unfolded proteins. A key question was brought into sharp focus by Levinthal (Levinthal 1969): if the random coil model holds, how can an unfolded protein discover its native conformation in biological real time? In particular, if unfolded protein molecules wander freely in a vast and featureless energy landscape, with barriers of order $\sim kT$, then three related questions arise:

1. *The kinetic question:* How does a protein discover its native conformation in biological real time? If restricted solely to the two most populated regions for a dipeptide, a 100-residue backbone would have $2^{100} \cong 10^{30}$ conformers. With bond rotations of order 10^{-13} s, the mean waiting time en route to the native conformation would be prohibitive just for the backbone. In actuality, experimentally determined folding times range from milliseconds to seconds.
2. *The thermodynamic question:* How does a protein compensate for the large conformational entropy loss on folding? With $2^{100} \cong 10^{30}$ conformers, the entropic price required to populate a single conformation $\cong 30 \times R \ln(10) \cong 40$ kcal mol⁻¹ at room temperature, a conservative estimate.
3. *The dynamic question:* How does a protein avoid meta-stable traps en route to its native conformation? An equivalent way of asking this question is: why do proteins have a unique native conformation instead of a Boltzmann-distributed ensemble of native conformations?

Many investigators have sought to provide answers to these questions. Two notable examples are mentioned here, though only in bare outline.

20.3.1.2 The Folding Funnel

Following earlier work of Frauenfelder et al. (Ansari et al. 1985), who suggested an analogy between proteins and spin glasses, Wolynes and coworkers introduced the notion of a folding funnel (Bryngelson et al. 1995) to describe the progress of a protein population that traverses its energy landscape en route to the folded state. The favorable-high-entropy, unfavorable-high-energy unfolded state is conceptualized as a wide funnel mouth, while the unfavorable-low-entropy, favorable-low-energy native state corresponds to a narrow funnel spout. According to this conception, sloping funnel walls guide the population downhill toward the folded state from

any starting point, answering question 1. During this downhill trajectory, lost entropy is progressively compensated by favorable pairwise interactions, answering question 2. Finally, meta-stable traps can be avoided if the funnel walls are sufficiently smooth (Dill and Chan 1997), answering question 3. As a corollary, it is postulated that evolutionary pressures screen protein sequences, selecting those which can fold successfully in a funnel-like manner (Tiana et al. 2000). The folding funnel evokes a graphic portrait of folding dynamics and thermodynamics but is not intended to address specific structural questions, such as whether a region of interest will be helix or sheet.

20.3.1.3 Transition State Theory

Fersht and coworkers imported transition state theory from small molecule chemical reactions into protein folding (Itzhaki et al. 1995). Akin to the folding funnel, transition state kinetics focus on the entire population, with the transition state species pictured at the top of an energy barrier which separates U from N. But, unlike the folding funnel, only a few key residues comprise the organizational “tipping point”, viz., those that participate in the transition state.

Questions 1–3 are not at issue for small molecule chemical reactions: (1) the mean waiting time for a reaction to occur depends upon a bond vibration, (2) after barrier crossing, the process is steeply downhill, and (3) intermediates between reactant and product are unstable because bond making/breaking energies are large. To the degree that the transition state approximation holds for protein folding (Baldwin and Rose 1999b), similar answers will obtain.

Transition state theory, expressed in the Eyring rate equation, transforms time-dependent kinetics into time-independent thermodynamics via an internal ticking clock: the rate of product formation depends upon the frequency of vibration of a critical bond. In contrast, no covalent bonds are made or broken in a folding reaction, and structure accretion is incremental and hierarchic en route from U to N (Baldwin and Rose 1999a,b). Not surprisingly then, the application of transition state theory to protein folding is complex (Cieplak and Hoang 2003).

Confidence in the application of the transition state approximation to protein folding comes from its success in describing simplified folding reactions (Doyle et al. 1997) and the thermal unfolding of a β -hairpin (Munoz et al. 1997). However, recent work also illustrates the complexities. The transition state can be shifted dramatically without changing a protein's amino acid sequence (Shastry and Udgaonkar 1995; Viguera et al. 1996). In simulations, the folding reaction can produce a broad ensemble of transition states instead of a single, well-defined species (Lazaridis and Karplus 1997). This blurring of the lines is further compounded by other work showing that the transition state resembles nearby folding intermediates (Kazmirski et al. 2001) or is simply a distorted form of the native state (Sanchez and Kiefhaber 2003).

20.3.1.4 Other Examples

The preceding examples illustrate how the random coil model has informed current thinking about unfolded proteins and the folding transition. The search for

answers to the three questions has motivated other studies as well. In yet another example that focuses on question 3, Sali et al. (1994a,b) analyzed the density of states in lattice simulations of folding and found a large energy gap – the e-gap – separating the native state (i.e., the ground state) from the nearest nonnative state (i.e., the 1st excited state). This finding rationalizes the predominance of the native state.

20.3.1.5 Implicit Assumptions from the Random Coil Model

Unfolded state models utilized in computer simulations often incorporate random coil assumptions implicitly. Four such assumptions are mentioned here.

1. The unfolded landscape is smooth. If the energy differences among sterically accessible backbone conformations are of order $\sim kT$, the landscape will be devoid of kinetic traps and conformational bias. This assumption simplifies strategies for exploring the unfolded state in simulations.
2. The isolated-pair hypothesis is valid. Lattice models provide a way to count conformational alternatives explicitly, and they have been used extensively (Chan and Dill 1991). Most often, residues are represented as single lattice points (i.e., all residues are sterically equivalent on a lattice). Consequently, residue-specific steric restrictions beyond the dipeptide are either underweighted or ignored. This practice is valid to the degree that local steric repulsion does not extend beyond nearest chain neighbors, an assumption made explicit in Flory's isolated-pair hypothesis (Flory 1969), which posits that each ϕ, ψ pair is sterically independent.
3. The Go approximation holds. A simplifying idea, introduced by Go (Go 1984), computes the energy of a conformation by rewarding native-like contacts while ignoring nonnative contacts, i.e., fortuitous nonnative contacts are not allowed to develop into kinetic traps. For simulations, this is a useful artifice that can be rationalized in a featureless landscape, where nonnative contacts dissolve as easily as they form.
4. Peptide backbone solvation is uniform. In other words, solvent water does not induce conformational bias in the unfolded state. If the interaction with water were energetically favored by some particular backbone conformation(s), then the unfolded landscape would be preferentially populated by these favored conformers, in violation of the featureless, random coil model.

These four assumptions are examined in Section 20.4.

20.3.2

The Random Coil – Experiment

Is a denatured protein aptly described as a random coil? It was Charles Tanford's experimental work that convinced the field. In numerous studies, Tanford demonstrated that proteins denatured in 6 M guanidinium chloride (GdmCl) have coil dimensions that obey simple scaling laws, consistent with random coil behavior. His

masterful review of protein denaturation in *Advances in Protein Chemistry* (Tanford 1968, 1970) is required reading for anyone interested in this topic.

In essence, the experimental strategy involves measuring coil dimensions for unfolded proteins in solution, fitting them to Eq. (6), and determining whether the scaling exponent, ν , is consistent with a random coil polymer with excluded volume in good solvent. The excluded volume can be obtained directly from any practical technique that depends upon the colligative properties of the polymer solution, such as osmotic pressure. Using such techniques, concentration-dependent deviations from ideality arising from solute:solvent interactions are measured. To extract the excluded volume, the chemical potential of the polymer solution is expanded as a power series in solute concentration – the virial expansion. For purely repulsive interactions, the molar excluded volume is given by the second virial coefficient (Schellman 2002). As mentioned above, excluded volume increases chain dimensions, with ν ranging from 0.33 for a collapsed, spherical molecule in poor solvent through 0.5 at the Θ -point to 0.6 in good solvent.

Tanford documented many experimental pitfalls (Tanford 1968). His investigations emphasized the importance of eliminating any potential residual structure in the unfolded protein by showing that the unfolding transition is complete. In fact, some residual structure is evident in heat-denatured proteins (Aune et al. 1967), but it can be eliminated in 6 M GdmCl. He also cautioned that the radius of gyration alone is an insufficient criterion for assessing random coil behavior, pointing out that a helical rod and a random polypeptide chain have similar values of R_G at chain lengths approximating those of ribonuclease and lysozyme.

20.3.2.1 Intrinsic Viscosity

In classic studies, Tanford used the intrinsic viscosity to determine coil dimensions. The intrinsic viscosity of a protein solution measures its effective hydrodynamic volume per gram in terms of the specific viscosity (van Holde 1971, chapter 7). In particular, if η is the viscosity of the solution and η_0 is the viscosity of solvent alone, the specific viscosity, $\eta_{sp} = (\eta - \eta_0)/\eta_0$, is the fractional change in viscosity produced by adding solute. While η_{sp} is the quantity of interest, it is expressed in an experimentally inconvenient volume fraction concentration scale. This is remedied by converting to the intrinsic viscosity, $[\eta]$, which is η_{sp} normalized by the protein concentration, c , at infinite dilution: $[\eta] = \lim_{c \rightarrow 0} \left(\frac{\eta_{sp}}{c} \right)$. Specific viscosity is a pure number, so intrinsic viscosity has units of reciprocal concentration, milliliters per gram.

The intrinsic viscosity is not a viscosity per se but a viscosity increment owing to the volume fraction of solution occupied by the protein, like η_{sp} . It measures the hydrated protein volume, which will be much larger for randomly coiled molecules than for compactly folded ones; $[\eta]$ scales with chain length, n . The dependence of $[\eta]$ on n is conformation dependent, and Tanford took advantage of this fact. The relevant equation is:

$$[\eta] = Kn^x \quad (7)$$

where K is a constant that depends upon the molecular weight, but only slightly. Intrinsic viscosity is closely related to R_G , and Eqs (6) and (7) have a similar form. If unfolded proteins retain residual structure, each in their own way, the relation between $[\eta]$ and n is expected to be idiosyncratic. Conversely, a series of proteins that conform to Eq. (7) is indicative of random coil behavior.

In fact, for a series of 15 proteins denatured in GdmCl, a plot of $\log[\eta]$ vs. $\log n$ describes a straight line with slope 0.666, the exponent in Eq. (7) (Tanford 1968, figure 6). The linearity of this series and the value of the exponent are strong evidence in favor of random coil behavior.

20.3.2.2 SAXS and SANS

Small angle X-ray scattering (SAXS) can be used to measure the radius of gyration, R_G , directly (Millett et al. 2002). Molecules in a protein solution scatter radiation like tiny antennae (van Holde et al. 1998, chapter 7). In idealized situations, particles scatter independently (Rayleigh scattering), but significant interference occurs when intramolecular distances are of the same order as the wavelength of incident radiation, λ . This is the situation that obtains when proteins are irradiated with X-rays, and it is the basis for all experimental scattering techniques that yield R_G . In this case, the quantity of interest is $P(\theta)$, the ratio of measured intensity to the intensity expected for independent scattering by particles much smaller than λ , as a function of the scattering angle, θ . For a solution of scatterers,

$$P(\theta) = \frac{1}{n^2} \sum_{i=1}^N \sum_{j=1}^N \frac{\sinh r_{ij}}{hr_{ij}} \quad (8)$$

where n is the number of scattering centers, r_{ij} is the distance between any pair of centers i and j , and h is a function of the wavelength and scattering angle:

$$h = \frac{4\pi}{\lambda} \sin \frac{\theta}{2} \quad (9)$$

The double sum over all scattering centers is immediately reminiscent of Eq. (3), in which R_G is rewritten in terms of individual vectors, without explicit reference to their center of gravity. Van Holde et al. (1998, p. 321) show that

$$P(\theta) = \frac{1}{n^2} \sum_{i=1}^N \sum_{j=1}^N (1) - \frac{h^2}{6n^2} \sum_{i=1}^N \sum_{j=1}^N r_{ij}^2 \quad (10)$$

where the first term is unity and R_G is directly related to the double sum in the second term, as in Eq. (3).

Millett et al. used SAXS to determine R_G for a series of proteins under both denaturing and native conditions (Millett et al. 2002, table I and fig. 4). Disulfide cross-links, if any, were reduced in the denatured species. Their experimentally de-

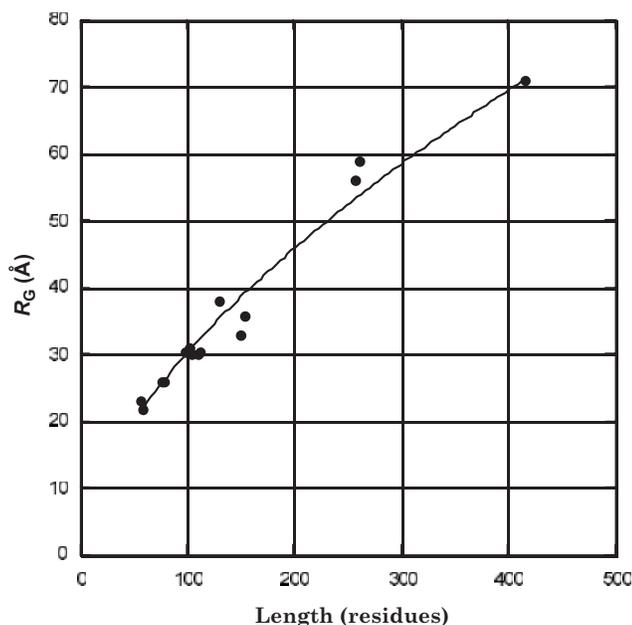


Fig. 20.3. The relationship between chain length and the radius of gyration, R_G , for a series of denatured proteins is well described by Eq. (6). Data points were taken from table 1 in (Millett et al. 2002), obtained using SAXS.

The fitted curve has a value of $\nu = 0.61 \pm 0.03$, in close agreement with theory. This figure reproduces the one in Millett et al. (2002, fig. 4), but with omission of outliers.

terminated values of R_G were fit to Eq. (6), giving values of $\nu = 0.61 \pm 0.03$ for the denatured proteins and $\nu = 0.38 \pm 0.05$ for their native counterparts (Figure 20.3). These values are remarkably close to those expected from theory, viz., $\nu = 0.6$ for a random coil with excluded volume in good solvent and $\nu = 0.33$ for a collapsed, spherical molecule in poor solvent.

The SAXS data provide the most compelling evidence to date in favor of the random coil model for denatured proteins.

20.4 Questions about the Random Coil Model

The random coil model would seem to be on firm ground at this point. However, recent work from both theory and experiment has raised new questions about the validity of the model – questions that provoke considerable controversy. Are they mere quibbles, or are they the prelude to a deeper understanding of the unfolded state?

Familiarity conditions intuition. At this point, the random coil model has conditioned our expectations for several decades. Should we be surprised that the di-

mensions of unfolded proteins are well described by a single exponent? Size matters here. As Al Holtzer once remarked, a steel I-beam behaves as a Gaussian coil if you make it long enough. But, at relevant length scales, the fact that proteins and polyvinyl behave similarly is quite unanticipated. After all, proteins adopt a unique folded state, whereas nonbiological polymers do not.

Flory emphasized this difference (Flory 1969): “Synthetic analogs of globular proteins are unknown. The capability of adopting a dense globular configuration stabilized by self-interactions and of transforming reversibly to the random coil are characteristics peculiar to the chain molecules of globular proteins alone” (p. 301).

The new questions center around the possibility of conformational bias and/or residual structure in unfolded proteins (Baldwin and Zimm 2000), even those unfolded in strong denaturing solvents like 6 M GdmCl (Plaxco and Gross 2001). We turn now to this discussion.

20.4.1

Questions from Theory

Superficially, the question of whether polypeptide chains are true random coils seems amenable to straightforward analysis by computer simulation. In principle, chains of n residues could be constructed, one at a time, using some plausible model (e.g., the Flory rotational isomeric model) to pick backbone dihedral angles. The coil dimensions and other characteristics of interest could then be analyzed by generating a suitable population of such chains. In practice, the excluded volume problem precludes this approach for chains longer than ~ 20 residues, where the likelihood of encountering a steric clash increases sharply, killing off nascent chains before they can elongate. Naively, one might think that the problem can be solved by randomly adjusting offending residues until the clash is relieved, but this tactic biases the overall outcome. In fact, the only unbiased tactic is to rebuild the chain from scratch, resulting almost invariably in other clashes at new sites for chain lengths of interest. Such problems have thwarted attempts to analyze the unfolded population via simulation and modeling.

20.4.1.1 The Flory Isolated-pair Hypothesis

Nearly all theoretical treatments of the unfolded state assume that local steric repulsion does not extend beyond nearest chain neighbors. This simplifying assumption was made explicit in Flory’s isolated-pair hypothesis (Flory 1969, p. 252), which posits that each ϕ, ψ pair is sterically independent of its neighbors.

Recently, the isolated-pair hypothesis was tested by exhaustive enumeration of all sterically accessible conformations of short polyalanyl chains (Pappu et al. 2000). To count, ϕ, ψ space was subdivided into a uniform grid. Every grid square, called a mesostate, encloses a $60^\circ \times 60^\circ$ range of ϕ, ψ values, with 36 mesostates in all. Each such mesostate was sampled extensively and at random to determine whether alanyl dipeptides with ϕ, ψ values in this range are sterically allowed. Only 14 mesostates are populated; the remaining 22 mesostates experience ubiquitous steric

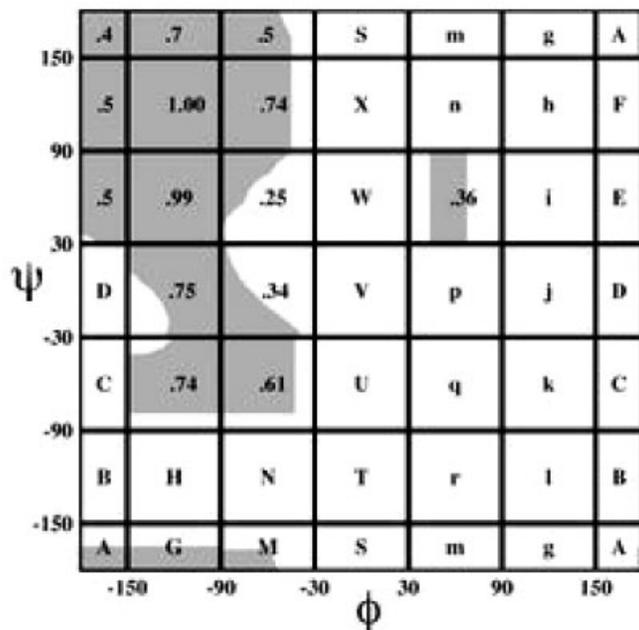


Fig. 20.4. Testing the Flory isolated-pair hypothesis. ϕ, ψ space for a dipeptide was subdivided into 36 alphabetically labeled coarse-grain grid squares, called mesostates. Treating the atoms as hard spheres, a Ramachandran plot (shown in gray) was computed by generating 150 000 randomly chosen ϕ, ψ conformations within each

mesostate and testing for steric collisions (Pappu and Rose 2002). Twenty-two mesostates have no allowed population; in each of these cases, every ϕ, ψ value results in a steric collision. In the 14 remaining mesostates, the fraction of sterically allowed samples, $0 < \Lambda_i \leq 1$, was determined, as shown.

clashes throughout their entire range. Reconstruction of allowed ϕ, ψ space from mesostate sampling recreates the dipeptide map (Ramachandran and Sasisekharan 1968) and provides an acceptance ratio for each mesostate (Figure 20.4). The acceptance ratio, Λ_i , is the fraction of sterically allowed samples for mesostate i , ranging from 0 to 1.

These Λ s were then used to test the isolated-pair hypothesis. Specifically, short polypeptide chains of length $n = 3 \dots 7$ were tested by enumerating all possible strings over the 14 allowed mesostates and sampling them extensively. If the isolated-pair hypothesis holds, then ϕ, ψ angles in each mesostate are independent, and the fraction of sterically allowed conformers for each string is given by the product of individual acceptance ratios, $\prod_i^n \Lambda_i$. But, if there are steric clashes between nonnearest neighbors in the string, then $\Lambda_{\text{string}} < \prod_i^n \Lambda_i$, invalidating the hypothesis.

From this analysis, the isolated-pair hypothesis was found to be valid in the upper left quadrant of ϕ, ψ space but invalid in all other allowed regions. This finding

makes sense physically: upon adopting ϕ, ψ values from the upper left quadrant, the chain is extended, like a β -strand, and nonnearest neighbors are separated. However, with ϕ, ψ values from any of the remaining five allowed mesostates (see Figure 20.4), the chain is contracted, like a helix or turn, and nonnearest neighbors are juxtaposed, enhancing opportunities for steric interference.

The failure of the isolated-pair hypothesis for short peptides ($n = 7$) challenges the random coil model, possibly in a major way. Steric restrictions obtain in the folded and unfolded states alike. The failure of the hypothesis for contracted chains implies that such conformers will be reduced selectively in the unfolded state, resulting in a population that is more extended than random coil expectations. Structurally, this shift in the population will result in a more homogeneous ensemble of unfolded conformers, and thermodynamically, it will reduce the entropy loss accompanying folding. But is it significant?

Studies of van Gunsteren et al. (van Gunsteren et al. 2001) and Sosnick and his colleagues (Zaman et al. 2003) concur that the size of conformational space that can be accessed by unfolded molecules is restricted in peptides. However, Ohkubo and Brooks (Ohkubo and Brooks 2003) argue that restrictions become rapidly insignificant as chain lengths grow beyond $n \cong 7$, with negligible consequences for the random coil model.

In an inventive approach to the problem, Goldenberg simulated populations of protein-sized chains (Goldenberg 2003) by adapting a standard software package that generates three-dimensional models from NMR-derived distance constraints. He analyzed the resultant unfolded state population using several measures, including coil dimensions, and found them to be well described as random coils. A note of caution is in order, however, because a substantial fraction of the conformers generated by this method fall within sterically restricted regions of ϕ, ψ space (Goldenberg 2003, table 1).

20.4.1.2 Structure vs. Energy Duality

Often, the complex interplay between structure and energy has confounded simulations. Small changes in structure can give rise to large changes in energy, and conversely. From a structural point of view, two conformers are distinguishable when their ϕ, ψ angles differ. From a thermodynamic point of view, two conformers are indistinguishable when they can interconvert via a spontaneous fluctuation.

This structure–energy duality has contributed confusion to the Levinthal paradox (Section 20.3.1.1) and many related size estimates of the unfolded population because a single energy basin can span multiple conformers. For example, most sterically accessible conformers of short polyalanyl chains in good solvent (Pappu et al. 2000) are quite extended, as expected in the absence of stabilizing intramolecular interactions. The ϕ, ψ values for these conformers are densely distributed over a broad region in the upper left quadrant of the ϕ, ψ map, as shown in Figure 20.5. When energy differences among these structures are calculated using a simple soft-sphere potential, the population partitions largely into two distinct energy basins, one that includes β -strands and another that includes polyproline II helices

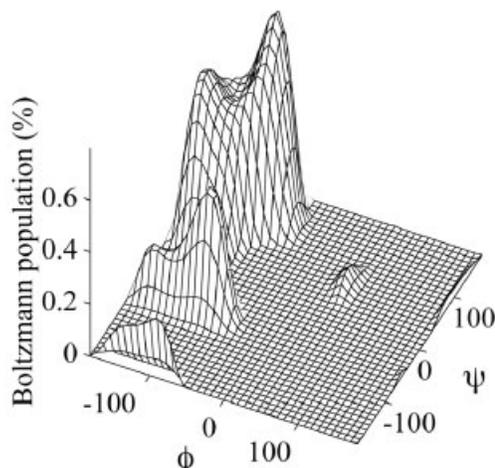


Fig. 20.5. A single energy basin can span multiple conformers. Most sterically accessible conformers of short polyalanyl chains in good solvent are extended. Using a soft-sphere potential, the Boltzmann-weighted population for the alanyl dipeptide is predominantly in the upper left quadrant of ϕ, ψ space and partitions into two distinct energy basins, one that includes polyproline II helices (larger) and another that includes β -strands (smaller) (Pappu and Rose 2002). At 300 K, conformers within each basin interconvert spontaneously.

(Pappu and Rose 2002). All conformers within each basin can interconvert spontaneously at room temperature (i.e., $\Delta A_{i,j} \leq RT$ at 300 K, where $\Delta A_{i,j}$ is the Helmholtz free energy difference between any two conformers, i and j , R is the universal gas constant and T the temperature in Kelvin). Thus, apparent structural diversity is reduced to two thermodynamically homogeneous populations.

20.4.1.3 The “Rediscovery” of Polyproline II Conformation

More than three decades ago, Tiffany and Krimm proposed that disordered peptides are comprised of left-handed polyproline II (P_{II}) helical segments interspersed with bends (Tiffany and Krimm 1968a,b). They were led to this prescient proposal by the similarity between the optical spectra of P_{II} helices and nonprolyl homopolymers. Even earlier, Schellman and Schellman had already argued that the spectrum of unfolded proteins was unlikely to be that of a true random coil (Schellman and Schellman 1964). Following these early studies, the ensuing literature disclosed a noticeable similarity between the spectra of P_{II} and unfolded proteins, but such suggestive hints failed to provoke widespread interest – until recently. See Shi et al. for a thorough review (Shi et al. 2002b).

The designation “polyproline” can be misleading. The circular dichroism (CD) spectrum, characteristic of actual polyproline or collagen peptides, has a pronounced negative band near 200 nm and a positive band near 220 nm. However, similar spectra can be obtained from peptides that are neither “poly” (Madison and Schellman 1970) nor proline-containing (Tiffany and Krimm 1968b).

The P_{II} conformation is a left-handed helix with three residues per turn ($\phi, \psi \cong -75^\circ, +145^\circ$), resulting in three parallel columns spaced uniformly around the long axis of the helix. This helix has no intrasegment hydrogen bonds, and, in solution, significant fluctuations from the idealized structure are to be expected. The P_{II} conformation is forced by sterics in a polyproline sequence, but it is adopted readily by proline-free sequences as well (Creamer 1998).

Only three repetitive backbone structures are sterically accessible in proteins: α -helix, β -strand and P_{II} -helix (Ramachandran and Sasisekharan 1968). In the folded population, α -helices and β -strands are abundant, whereas P_{II} -helices are rare. More specifically, isolated residues with P_{II} ϕ, ψ values are common in the non- α , non- β regions, accounting for approximately one-third of the remaining residues, but longer runs of consecutive residues with P_{II} ϕ, ψ -values are infrequent (Stapley and Creamer 1999).

This finding can be rationalized by the fact that P_{II} -helices cannot participate in hydrogen bonds in globular proteins. Hydrogen bonds are eliminated because the spatial orientation of backbone donors and acceptors is incompatible with both intrasegment hydrogen bonding within P_{II} -helices and regular extra-segment hydrogen bonding between P_{II} helices and the three repetitive backbone structures. Upon folding, those backbone polar groups deprived of hydrogen-bonded solvent access can make compensatory hydrogen bonds in α -helices and strands of β -sheet, but not in P_{II} -helices.

Recent work by Creamer and coworkers focused renewed attention on P_{II} (Creamer 1998; Stapley and Creamer 1999; Rucker et al. 2003), raising the question of whether fluctuating P_{II} conformation might contribute substantially to the unfolded population in proteins (Tiffany and Krimm 1968b). Studies performed during the past few years lend support to this idea, as described next.

20.4.1.4 P_{II} in Unfolded Peptides and Proteins

The blocked peptide, *N*-acetylalanine-*N'*-methylamide, is a popular backbone model. Many groups have found P_{II} to be an energetically preferred conformation for this peptide in water (Anderson and Hermans 1988; Grant et al. 1990; Jalkanen and Suhai 1996; Han et al. 1998; Poon and Samulski 2000; Drozdov et al. 2004). Does this finding hold for longer chains?

Again using alanine as a model, Pappu and Rose analyzed the conformational preferences of longer blocked polyalanyl chains, *N*-acetyl-Ala_{*n*}-*N'*-methylamide ($n \leq 7$) (Pappu and Rose 2002). To capture nonspecific solvent effects, they minimized chain:chain interactions, mimicking the chain's expected behavior in good solvent. At physiological temperature, only three energy basins were needed to span $\sim 75\%$ of the population, and within each basin, the population of structures was homogeneous. Notably, the basin corresponding to P_{II} structure was dominant.

Pappu and Rose (Pappu and Rose 2002) used soft-sphere repulsion (the repulsive term in a Lennard-Jones potential) to calculate energy. More extensive testing using detailed force fields was performed by Sosnick and coworkers (Zaman et al. 2003).

It is often assumed that the backbone is solvated uniformly in the unfolded state and that the energy of solvent stripping upon folding is not a significant consideration. This assumption follows directly from the random coil model, in which unfolded conformers are readily interconvertible. However, if unfolded state conformers exhibit conformational biases, it becomes important to question this assumption. Is solvation free energy conformation dependent?

A series of papers by Avbelj and Baldwin (Avbelj et al. 2000; Avbelj and Baldwin 2002, 2003) offered a fresh perspective on this issue, motivated by an inconsistency between the measured energy of peptide hydrogen bond formation (Scholtz et al. 1991) and the corresponding energy derived from a simple thermodynamic cycle (Baldwin 2003). Specifically, their analysis uncovered a large enthalpy deficit (-7.6 kcal mol⁻¹) upon helix formation that could not be reconciled with data from typical model compounds, such as acetamide derivatives (Avbelj et al. 2000). One or more terms had to be missing.

Avbelj and Baldwin's work prompted a re-examination of peptide solvation in proteins by a number of groups, including themselves (Avbelj et al. 2000; Avbelj and Baldwin 2002, 2003). Of particular interest are a series of unrelated simulations (Avbelj et al. 2000; Avbelj and Baldwin 2002, 2003; Zaman et al. 2003; Drozdov et al. 2004; Garcia 2004; Kentsis et al. 2004; Mezei et al. 2004), all of which reach a common conclusion: water interacts preferentially with P_{II} peptides, imparting a previously unsuspected conformational bias.

In sum, both peptide:solvent interactions and peptide:peptide interactions (Pappu and Rose 2002) favor P_{II} conformers. In the former case, water is simply a better solvent for P_{II} than for other conformers, e.g., β -strands and α -helices. In the latter case, P_{II} affords the chain greater entropic freedom (i.e., more "wiggle room").

20.4.2

Questions from Experiment

Early NMR studies provided evidence for residual structure in the denatured state of both proteins (Garvey et al. 1989; Neri et al. 1992) and peptides excised from proteins (Dyson et al. 1992). However, the structured regions seen in proteins were not extensive. Furthermore, most isolated peptides lacked structure, and the few exceptions did not always retain the conformation adopted in the native protein (Dyson et al. 1988).

Peptide studies tell a similar story. A prime example involves the assessment of autonomous stability in the α -helix. Early evidence indicated that the cooperative unit for stable helix formation is ~ 100 residues (Zimm and Bragg 1959), a length that exceeds the average protein helix (~ 12 residues) by almost an order of magnitude. Consequently, the prevailing view in the 1970s was that protein-sized helical peptides would be random coils in isolation. This view was reversed in the 1980s, after Bierzynski et al. (Bierzynski et al. 1982), expanding upon earlier work by Brown and Klee (Brown and Klee 1971), demonstrated helix formation in water at near-physiological temperature for residues 1–13 of ribonuclease, a cyanogen

bromide cleavage product. This finding prompted a re-evaluation of helix propensities in peptides (Lyu et al. 1990; Merutka et al. 1990; O'Neil and DeGrado 1990; Padmanabhan et al. 1990) and motivated numerous biophysical studies of peptides (Scholtz and Baldwin 1992). Summarizing this large body of work, there is evidence for structure in some short peptides in aqueous solvent at physiological temperature, but it is marginal at best and, more often, undetectable altogether.

20.4.2.1 Residual Structure in Denatured Proteins and Peptides

The limited success of these early attempts to detect residual structure strengthened the conviction that denaturation abolishes structure and reinforced the notion that the unfolded state is a random coil. Consequently, the field was stunned when Shortle and Ackerman (Shortle and Ackerman 2001) demonstrated the persistence of native-like structure in staphylococcal nuclease under strongly denaturing conditions (8 M urea). Shortle and Ackerman's finding was based on evidence from residual dipolar couplings in oriented gels. However, their interpretation that these data provide evidence of global organization was questioned recently by Annala and coworkers (Louhivuori et al. 2003). The ultimate conclusions from such work are still unclear, but the perspective has definitely changed and many recent experiments now find evidence for substantial residual structure in the denatured state (e.g., Daggett 1996; Yi et al. 2000; Garcia 2001; Kazmirski et al. 2001; Sanchez 2003; Sridevi 2004).

In a similar vein, Shi et al. reanalyzed a blocked peptide containing seven consecutive alanine residues for the presence of residual structure (Shi et al. 2002a). This peptide is too short to form a stable α -helix and should therefore be a random coil. Contrary to this expectation, the peptide is largely in P_{II} conformation, in agreement with predictions from theory (Pappu and Rose 2002). While not all residues are expected to favor the P_{II} conformation (Rucker et al. 2003), this result shows that the unfolded state is predominantly a single conformer, at least in the case of polyalanine.

20.4.3

The Reconciliation Problem

The measured radii of gyration, R_G , of denatured proteins have values (Millett et al. 2002) that are consistent with those expected for a random coil with excluded volume in good solvent (Section 20.3.2.2). Yet, experimental evidence in both proteins (Shortle and Ackerman 2001) and peptides (Shi et al. 2002a) suggests the presence of residual structure in the unfolded population. How are these seemingly contradictory findings to be reconciled? Millett et al. refer to this as the reconciliation problem (Millett et al. 2002, see their discussion, p. 257).

Paradox is often a prelude to perception. Equation (6), in its generality, necessarily neglects the chemical details of any particular polymer type. Accordingly, the resultant chain description is insensitive to short-range order, apart from the proportionality constant R_0 , which is a function of the persistence length (Section 20.3.1). Sterically induced local order, encapsulated in R_0 , is surely present in unfolded

proteins (Pappu et al. 2000; Fitzkee and Rose 2004), but can it rationalize the apparent contradiction between random coil R_G values and global residual structure (Shortle and Ackerman 2001)? One possible explanation is that multiple regions of local structure dominate the ensemble average to such an extent that they are interpreted as global organization (Zagrovic 2002; Louhivuori et al. 2003).

The coil library may provide a useful clue to the resolution of this puzzle. The coil library is the collection of all nonrepetitive elements in proteins of known structure, that fraction of native structure which remains after α -helix and β -sheet are removed. Given that the library is composed of fragments extracted from solved structures, it is surely not “coil” in the polymer sense. However, the term “coil library” is intended to convey the hypothesis that such fragments do, in fact, represent the full collection of accessible chain conformers in unfolded proteins (Smith et al. 1996). Taken to its logical conclusion, this hypothesis posits that the coil library is a collection of structured fragments in folded proteins and, at the same time, a collection of unstructured fragments in unfolded proteins. If so, this library, together with α -helix, β -strand, and P_{II} helix, represents an explicit enumeration of accessible conformers from which the unfolded ensemble might be reconstructed (Avbelj and Baldwin 2003).

At this writing, the reconciliation problem remains an ongoing question. Regardless of the eventual outcome, this paradox appears to be moving the field in an informative direction.

20.4.4

Organization in the Unfolded State – the Entropic Conjecture

Are there general principles that lead to organization in the unfolded state? If accessible conformational space is vast and undifferentiated, the entropic cost of populating the native basin exclusively will be large. However, if the unfolded state is largely restricted to a few basins, with nonuniform, sequence-dependent basin preferences, then entropy can function as a chain organizer.

Consider two thermodynamic basins, i and j . The Boltzmann-weighted ratio of their populations, n_i/n_j , is given by $(w_i/w_j)e^{-\beta\Delta U}$, where n_i and n_j designate the number of conformers in i and j , w_i and w_j are the degeneracies of state, β is the Boltzmann factor, and ΔU is the energy difference between the two basins. Both entropy and enthalpy contribute to this ratio. If w_i and w_j are conformational biases (i.e., the number of isoenergetic ways the chain can adopt conformations i and j), and w_i/w_j is the dominant term in the Boltzmann ratio, the entropy difference, $\Delta S_{\text{conf}} = R \ln(w_i/w_j)$, would promote organization in the unfolded population.

In particular, if P_{II} is a dominant conformation in polyalanyl peptides, then it is also likely to be favored in unfolded proteins, in which case the unfolded state is not as heterogeneous as previously believed. The usual estimate of about five accessible states per residue in an unfolded protein is based on a familiar argument: the free energy difference between the folded and unfolded populations, ΔG_{conf} , is a small difference between large value of ΔH_{conf} and $T\Delta S_{\text{conf}}$ (Brandts 1964a,b). If

$\Delta G_{\text{conf}} \cong -10 \text{ kcal mol}^{-1}$ (a typical value) and $\Delta H_{\text{conf}} \cong 100 \text{ kcal mol}^{-1}$, then the counterbalancing $T\Delta S_{\text{conf}}$ is also $\sim 100 \text{ kcal mol}^{-1}$. Then $\Delta S_{\text{conf}} \cong 3.33$ entropy units per residue for a 100-residue protein at 300 K. Assuming $\Delta S_{\text{conf}} = R \ln W$, the number of states per residue, W , is 5.34.

However, instead of a reduction in the number of distinct states, this entropy loss on folding could result from a reduction in the degeneracy of a single state, providing the ϕ, ψ space of occupied regions in the unfolded population is further constricted upon folding. For example, a residue in P_{II} is within a room-temperature fluctuation of any sterically allowed ϕ, ψ value in the upper left quadrant of the dipeptide map (Mezei et al. 2004, table 1). Consequently, different ϕ, ψ values from these regions would be thermodynamically indistinguishable and therefore not distinct states at all. As a back-of-the-envelope approximation, consider a residue that can visit any allowed region of the upper left quadrant in the unfolded state. Upon folding, let this residue be constrained to lie within $\pm 30^\circ$ of ideal β -sheet ϕ, ψ values. The reduction in ϕ, ψ space would be a factor of 5.58, approximating the value attributed to distinct states. Similar, but less approximate, estimates can be obtained when the unfolded populations are Boltzmann weighted.

What physical factors might underwrite such entropy effects?

20.4.4.1 Steric Restrictions beyond the Dipeptide

It has long been believed that local steric restrictions do not extend beyond the dipeptide boundary (Ramachandran and Sasisekharan 1968), but, on re-analysis, this conviction requires revision (Srinivasan and Rose 1999; Pappu et al. 2000; Baldwin and Zimm 2000) (see Section 20.4.1.1). In fact, systematic steric restrictions operate over chain regions of several adjacent residues, and they serve to promote organization in unfolded protein chains. Two recent lines of investigation focus on identifying the physical basis for longer range, sterically induced ordering.

In a series of remarkable papers, Banavar, Maritan and their colleagues show that chain thickness alone imposes stringent, previously unrecognized restrictions on conformational space (Maritan et al. 2000; Banavar 2002; Hoang et al. 2004). All the familiar secondary structure motifs emerge automatically when the protein is represented as a self-avoiding tube, coaxial with the main chain, and a single inequality is imposed on all triples of $C\alpha$ atoms (Maritan et al. 2000; Banavar 2002). The further addition of simple hydrogen bond and hydrophobic terms is sufficient to generate the common super-secondary structures (Hoang et al. 2004). These straightforward geometric considerations demonstrate that sequence-independent steric constraints predispose proteins toward their native repertoire of secondary and super-secondary structural motifs.

Investigating the atomic basis for longer range steric restrictions, Fitzkee and Rose found that a direct transition from an α -helix to a β -strand causes an unavoidable steric collision between backbone atoms (Fitzkee and Rose 2004). Specifically, a nonnearest neighbor collision occurs between the carbonyl oxygens of an α -residue at position i (O^{α}_i) and a β -residue at position $i + 3$ (O^{β}_{i+3}). This restriction also holds for the transition from α -helix to P_{II} . These simple steric constraints

have pervasive organizational consequences for unfolded proteins because they eliminate all structural hybrids of the form $\dots\alpha\alpha\beta\dots$ and $\dots\alpha\alpha\alpha P_{II}\dots$, pushing the unfolded population toward pure segments of α , β , and P_{II} interconnected by irregular regions such as those found in the coil library.

20.5 Future Directions

The early analysis of steric restrictions in the alanyl dipeptide (more precisely, the compound $C\alpha\text{-CO-NH-C}\alpha\text{HR-CO-NH-C}\alpha$, which has two degrees of backbone freedom like a dipeptide) by Ramachandran et al. (Ramachandran et al. 1963) has become one of those rare times in biochemistry where theory is deemed sufficient to validate experiment (Laskowski et al. 1993). The fact that the dipeptide map is based only on “hard sphere” repulsion alone led some to underestimate the generality of this work, but not Richards, who commented (Richards 1977):

For chemically bonded atoms the distribution is not spherically symmetric nor are the properties of such atoms isotropic. In spite of all this, the use of the hard sphere model has a venerable history and an enviable record in explaining a variety of different observable properties. As applied specifically to proteins, the work of G. N. Ramachandran and his colleagues has provided much of our present thinking about permissible peptide chain conformations.

The notion that repulsive interactions promote macromolecular organization is not limited to the alanyl dipeptide. Space-filling models (Koltun 1965), which represent each atom literally as a hard sphere, were central to Pauling’s successful model of the α -helix (Pauling et al. 1951) and have widespread application throughout chemistry. Much of the theory of liquids is based on the organizing influence of repulsion interactions (Chandler et al. 1983).

Despite such successes, the existence of sterically induced chain organization has had little influence on models of the unfolded state owing to the strongly held conviction that local steric restrictions extend no further than adjacent chain neighbors. Of course, long-range excluded volume effects do affect the population (Flory 1953; Chan and Dill 1991), as reflected in the exponent of Eq. (6), but they are not thought to play any role in biasing unfolded proteins toward specific conformations. Given the finding of local steric restrictions beyond the dipeptide (Section 20.4.4.1), it is time to re-analyze the problem.

Re-analysis will involve at least three steps: (1) analysis of local steric restrictions beyond the dipeptide, (2) characterization of elements in the coil library, and (3) combination of the results from these two steps. To the extent that useful insights emerge from this prescription, the folding problem may not be as intractable as previously thought.

Acknowledgments

We thank Buzz Baldwin, Nicholas Fitzkee, Haipeng Gong, Nicholas Panasik, Kevin Plaxco, and Timothy Street for stimulating discussion and The Mathers Foundation for support.

References

- 1 ANDERSON, A. G., and HERMANS, J. 1988. Microfolding: conformational probability map for the alanine dipeptide in water from molecular dynamics simulations. *Proteins* **3**, 262–265.
- 2 ANSARI, A., BERENDZEN, J., BOWNE, S. F., FRAUENFELDER, H., IBEN, I. E., SAUKE, T. B., SHYAMSUNDER, E., and YOUNG, R. D. 1985. Protein states and proteinquakes. *Proc. Natl Acad. Sci. USA* **82**, 5000–5004.
- 3 AUNE, K. C., SALAHUDDIN, A., ZARLENGO, M. H., and TANFORD, C. 1967. *J. Biol. Chem.* **242**, 4486–4489.
- 4 AVBELJ, F. and BALDWIN, R. L. 2002. Role of backbone solvation in determining thermodynamic beta propensities of the amino acids. *Proc. Natl Acad. Sci. USA* **99**, 1309–1313.
- 5 AVBELJ, F. and BALDWIN, R. L. 2003. Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: Distributions of phi. *Proc. Natl Acad. Sci. USA* **100**, 5742–5747.
- 6 AVBELJ, F., LUO, P., and BALDWIN, R. L. 2000. Energetics of the interaction between water and the helical peptide group and its role in determining helix propensities. *Proc. Natl Acad. Sci. USA* **97**, 10786–10791.
- 7 BANAVAR 2002. ■
- 8 BALDWIN, R. L. 2003. In search of the energetic role of peptide hydrogen bonds. *J. Biol. Chem.* **278**, 17581–17588.
- 9 BALDWIN, R. L. and ROSE, G. D. 1999a. Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem. Sci.* **24**, 26–33.
- 10 BALDWIN, R. L. and ROSE, G. D. 1999b. Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem. Sci.* **24**, 77–83.
- 11 BALDWIN, R. L. and ZIMM, B. H. 2000. Are denatured proteins ever random coils? *Proc. Natl Acad. Sci. USA* **97**, 12391–12392.
- 12 BIERZYNSKI, A., KIM, P. S., and BALDWIN, R. L. 1982. A salt-bridge stabilizes the helix formed by isolated C-peptide of RNase A. *Proc. Natl Acad. Sci. USA* **79**, 2470–2474.
- 13 BRANDTS, J. F. 1964a. The thermodynamics of protein denaturation. I. The denaturation of chymotrypsinogen. *J. Am. Chem. Soc.* **86**, 4291–4301.
- 14 BRANDTS, J. F. 1964b. The thermodynamics of protein denaturation. II. A model of reversible denaturation and interpretations regarding the stability of chymotrypsinogen. *J. Am. Chem. Soc.* **86**, 4302–4314.
- 15 BROWN, J. E. and KLEE, W. A. 1971. Helix-coil transition of the isolated amino terminus of ribonuclease. *Biochemistry* **10**, 470–476.
- 16 BRYNGELSON, J. D., ONUCHIC, J. N., SOCCI, N. D., and WOLYNES, P. G. 1995. Funnels, pathways and the energy landscape of protein folding-a synthesis. *Proteins Struct. Funct. Genet.* **21**, 167–195.
- 17 CANTOR, C. R. and SCHIMMEL, P. R. 1980. *Biophysical Chemistry*. Part III: *The Behavior of Biological Macromolecules*. Freeman, New York.
- 18 CHAN, H. S. and DILL, K. A. 1991. Polymer principles in protein structure and stability. *Annu. Rev. Biophys. Chem.* **20**, 447–490.
- 19 CHANDLER, D., WEEKS, J. D., and ANDERSEN, H. C. 1983. The van der Waals picture of liquids, solids and

- phase transformations. *Science* **220**, 787–794.
- 20 CIEPLAK, M. and HOANG, T. X. 2003. Universality classes in folding times of proteins. *Biophys. J.* **84**, 475–488.
 - 21 CREAMER, T. P. 1998. Left-handed polyproline II helix formation is (very) locally driven. *Proteins* **33**, 218–226.
 - 22 DAGGETT 1996. ■
 - 23 DILL, K. A. and CHAN, H. S. 1997. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19.
 - 24 DILL, K. A. and SHORTLE, D. 1991. Denatured states of proteins. *Annu. Rev. Biochem.* **60**, 795–825.
 - 25 DOYLE, R., SIMONS, K., QIAN, H., and BAKER, D. 1997. *Proteins Struct. Funct. Genet.* **29**, 282–291.
 - 26 DROZDOV, A. N., GROSSFIELD, A., and PAPPU, R. V. 2004. The role of solvent in determining conformational preferences of alanine dipeptide in water. *J. Am. Chem. Soc.* **126**, 2574–2581.
 - 27 DYSON, H. J., RANCE, M., HOUGHTEN, R. A., LERNER, R. A., and WRIGHT, P. E. 1988. Folding of immunogenic peptide fragments of proteins in water solution. I. Sequence requirements for the formation of a reverse turn. *J. Mol. Biol.* **201**, 161–200.
 - 28 DYSON, H. J., SAYRE, J. R., MERUTKA, G., SHIN, H. C., LERNER, R. A., and WRIGHT, P. E. 1992. Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding II. Plastocyanin. *J. Mol. Biol.* **226**, 819–835.
 - 29 EDSALL, J. T. 1995. Hsien Wu and the first theory of protein denaturation (1931). In *Advances in Protein Chemistry* (eds D. S. EISENBERG, and F. M. RICHARDS), pp. 1–26. Academic Press, San Diego.
 - 30 EINSTEIN, A. 1956. *Investigations on the Theory of Brownian Movement*. Dover Publications, New York.
 - 31 FITZKEE, N. C. and ROSE, G. D. 2004. Steric restrictions in protein folding: an α -helix cannot be followed by a contiguous β -strand. *Protein Sci.* **13**, 633–639.
 - 32 FLORY, P. J. 1953. *Principles of Polymer Chemistry*. Cornell University Press, New York.
 - 33 FLORY, P. J. 1969. *Statistical Mechanics of Chain Molecules*. Wiley, New York.
 - 34 GARCIA 2001. ■
 - 35 GARCIA, A. E. 2004. Characterization of non-alpha helical conformations in Ala peptides. *Polymer* **45**, 669–676.
 - 36 GARVEY, E. P., SWANK, J., and MATTHEWS, C. R. 1989. A hydrophobic cluster forms early in the folding of dihydrofolate reductase. *Proteins Struct. Funct. Genet.* **6**, 259–266.
 - 37 GINSBURG, A. and CARROLL, W. R. 1965. Some specific ion effects on the conformation and thermal stability of ribonuclease. *Biochemistry* **4**, 2159–2174.
 - 38 GO, N. 1984. The consistency principle in protein structure and pathways of folding. *Adv. Biophys.* **18**, 149–164.
 - 39 GOLDENBERG, D. P. 2003. Computational simulation of the statistical properties of unfolded proteins. *J. Mol. Biol.* **326**, 1615–1633.
 - 40 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.
 - 41 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.
 - 42 HOANG, T. X., TROVATO, A., SENO, F., BANAVAR, J., and MARITAN, A. 2004. What determines the native state folds of proteins? submitted.
 - 43 ITZHAKI, L. S., OTZEN, D. E., and FERSHT, A. R. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* **254**, 260–288.
 - 44 JALKANEN, K. J. and SUHAI, S. 1996. N-Acetyl-L-Alanine N'-methylamide: a

- density functional analysis of the vibrational absorption and birational circular dichroism spectra. *Chem. Phys.* **208**, 81–116.
- 45 KAUZMANN, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1–63.
- 46 KAZMIRSKI, S. L., WONG, K. B., FREUND, S. M. V., TAN, Y. J., FERSHT, A. R., and DAGGETT, V. 2001. Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl Acad. Sci. USA* **98**, 4349–4354.
- 47 KENTISIS, A., GINDIN, T., MEZEI, M., and OSMAN, R. 2004. Unfolded state of polyalanine is a segmented polyproline II helix. *Proteins* **55**, in press.
- 48 KOLTUN, W. L. 1965. Precision space-filling atomic models. *Biopolymers* **3**, 665–679.
- 49 LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S., and THORNTON, J. M. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291.
- 50 LAZARIDIS, T. and KARPLUS, M. 1997. *Science* **278**, 1928–1931.
- 51 LEVINTHAL, C. 1969. How to fold gracefully. Mossbauer spectroscopy in Biological Systems, Proceedings. *University of Illinois Bull.* **41**, 22–24.
- 52 LOUHIVUORI, M., PAAKKONEN, K., FREDRIKSSON, K., PERMI, P., LOUNILA, J., and ANNILA, A. 2003. On the origin of residual dipolar couplings from denatured proteins. *J. Am. Chem. Soc.* **125**, 15647–15650.
- 53 LYU, P. C., LIFF, M. I., MARKY, L. A., and KALLENBACH, N. R. 1990. Side chain contributions to the stability of alpha-helical structure in peptides. *Science* **250**, 669–673.
- 54 MADISON, V. and SCHELLMAN, J. A. 1970. Diamide model for the optical activity of collagen and polyproline I and II. *Biopolymers* **9**, 65–94.
- 55 MARITAN, A., MICHELETTI, C., TROVATO, A., and BANAVAR, J. 2000. Optimal shapes of compact strings. *Nature* **406**, 287–290.
- 56 MERUTKA, G., LIPTON, W., SHALONGO, W., PARK, S. H., and STELLWAGEN, E. 1990. Effect of central-residue replacements on the helical stability of a monomeric peptide. *Biochemistry* **29**, 7511–7515.
- 57 MEZEI, M., FLEMING, P. J., SRINIVASAN, R., and ROSE, G. D. 2004. Polyproline II helix is the preferred conformation for unfolded polyalanine in water. *Proteins* **55**, ■.
- 58 MILLETT, I. S., DONIACH, S., and PLAXCO, K. W. 2002. Toward a taxonomy of the denatured state: small angle scattering studies of unfolded proteins. *Adv. Protein Chem.* **62**, 241–262.
- 59 MIRSKY, A. E. and PAULING, L. 1936. On the structure of native, denatured, and coagulated proteins. *Proc. Natl Acad. Sci. USA* **22**, 439–447.
- 60 MUNOZ, V., THOMPSON, P. A., HOFRICHTER, J., and EATON, W. A. 1997. *Nature* **390**, 196–199.
- 61 NERI, D., BILLETER, M., WIDER, G., and WUTHRICH, K. 1992. NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science* **257**, 1559–1563.
- 62 O'NEIL, K. T. and DEGRADO, W. F. 1990. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* **250**, 646–651.
- 63 OHKUBO, Y. Z. and BROOKS, C. L. 2003. Exploring Flory's isolated-pair hypothesis: statistical mechanics of helix-coil transitions in polyalanine and C-peptide from RNase A. *Proc. Natl Acad. Sci. USA* **100**, 13916–13921.
- 64 PADMANABHAN, S., MARQUSEE, S., RIDGEWAY, T., LAUE, T. M., and BALDWIN, R. L. 1990. Relative helix-forming tendencies of nonpolar amino acids. *Nature* **344**, 268–270.
- 65 PAPPU, R. V. and ROSE, G. D. 2002. A simple model for polyproline II structure in unfolded states of alanine-based peptides. *Protein Sci.* **11**, 2437–2455.
- 66 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. USA* **97**, 12565–12570.
- 67 PAULING, L., COREY, R. B., and BRANSON, H. R. 1951. The structures of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl Acad. Sci. USA* **37**, 205–210.
- 68 PLAXCO, K. W. and GROSS, M. 2001. Unfolded, yes, but random? Never! *Nat. Struct. Biol.* **8**, 659–660.
- 69 POON, C.-D. and SAMULSKI, E. T. 2000. Do bridging water molecules dictate the structure of a model dipeptide in aqueous solution? *J. Am. Chem. Soc.* **122**, 5642–5643.
- 70 RAMACHANDRAN, G. N. and SASISEKHARAN, V. 1968. Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283–438.
- 71 RAMACHANDRAN, G. N., RAMAKRISHNAN, C., and SASISEKHARAN, V. 1963. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99.
- 72 RICHARDS, F. M. 1977. Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151–176.
- 73 RUCKER, A. L., PAGER, C. T., CAMPBELL, M. N., QUALLS, J. E., and CREAMER, T. P. 2003. Host-guest scale of left-handed polyproline II helix formation. *Proteins* **53**, 68–75.
- 74 SALI, A., SHAKHNOVICH, E., and KARPLUS, M. 1994a. Kinetics of protein folding – a lattice model study of the requirements for folding to the native state. *J. Mol. Biol.* **5**, 1614–1636.
- 75 SALI, A., SHAKHNOVICH, E. I., and KARPLUS, M. 1994b. How does a protein fold? *Nature* **477**, 248–251.
- 76 SANCHEZ 2003. ■
- 77 SANCHEZ, I. E. and KIEFHABER, T. 2003. Origin of unusual ϕ -values in protein folding: evidence against specific nucleation sites. *J. Mol. Biol.* **334**, 1077–1085.
- 78 SCHELLMAN, J. A. 2002. Fifty years of solvent denaturation. *Biophys. Chem.* **96**, 91–101.
- 79 SCHELLMAN, J. A. and SCHELLMAN, C. G. 1964. In *The Proteins: Composition, Structure and Function*, 2nd edn, Vol 2, pp. 1–37. Academic Press, New York.
- 80 SCHOLTZ, J. M. and BALDWIN, R. L. 1992. The mechanism of alpha-helix formation by peptides. *Annu. Rev. Biophys. Biomol. Struct.* **21**, 95–118.
- 81 SCHOLTZ, J. M., MARQUSEE, S., BALDWIN, R. L., YORK, E. J., STEWART, J. M., SANTORO, M., and BOLEN, D. W. 1991. Calorimetric determination of the enthalpy change for the alpha-helix to coil transition of an alanine peptide in water. *Proc. Natl Acad. Sci. USA* **88**, 2854–2858.
- 82 SHASTRY, M. C. and UDGAONKAR, J. B. 1995. The folding mechanism of barstar: evidence for multiple pathways and multiple intermediates. *J. Mol. Biol.* **247**, 1013–1027.
- 83 SHI, Z., OLSON, C. A., ROSE, G. D., BALDWIN, R. L., and KALLENBACH, N. R. 2002a. Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl Acad. Sci. USA* **2002**, 9190–9195.
- 84 SHI, Z., WOODY, R. W., and KALLENBACH, N. R. 2002b. Is polyproline II a major backbone conformation in unfolded proteins? *Adv. Protein Chem.* **62**, 163–240.
- 85 SHORTLE, D. and ACKERMAN, M. 2001. Persistence of native-like topology in a denatured protein in 8 M urea. *Science* **293**, 487–489.
- 86 SIMPSON, R. B. and KAUZMANN, W. 1953. The kinetics of protein denaturation. *J. Am. Chem. Soc.* **75**, 5139–5192.
- 87 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.
- 88 SRIDEVI 2004. ■
- 89 SRINIVASAN, R. and ROSE, G. D. 1999. A physical basis for protein secondary structure. *Proc. Natl Acad. Sci. USA* **96**, 14258–14263.
- 90 STAPLEY, B. J. and CREAMER, T. P. 1999. A survey of left-handed

- polyproline II helices. *Protein Sci.* **8**, 587–595.
- 91 TANFORD, C. 1968. Protein denaturation. *Adv. Protein Chem.* **23**, 121–282.
- 92 TANFORD, C. 1970. Protein denaturation. Part C. Theoretical models for the mechanism of denaturation. *Adv. Protein Chem.* **24**, 1–95.
- 93 TANFORD, C., PAIN, R. H., and OTCHIN, N. S. 1966. Equilibrium and kinetics of the unfolding of lysozyme (muramidase) by guanidine hydrochloride. *J. Mol. Biol.* **15**, 489–504.
- 94 TIANA, G., BROGLIA, R. A., and SHAKHNOVICH, E. I. 2000. Hiking in the energy landscape in sequence space: a bumpy road to good folders. *Proteins* **39**, 244–251.
- 95 TIFFANY, M. L. and KRIMM, S. 1968a. *Biopolymers* **6**, 1767–1770.
- 96 TIFFANY, M. L. and KRIMM, S. 1968b. New chain conformations of poly(glutamic acid) and polylysine. *Biopolymers* **6**, 1379–1382.
- 97 VAN GUNSTEREN, W. F., BÜRGI, R., PETER, C., and DAURA, X. 2001. The key to solving the protein-folding problem lies in an accurate description of the denatured state. *Angew. Chem. Int. Ed.* **40**, 351–355.
- 98 VAN HOLDE, K. E. 1971. *Physical Biochemistry*. Prentice-Hall, Englewood Cliffs, NJ.
- 99 VAN HOLDE, K. E., JOHNSON, W. C., and HO, P. S. 1998. *Physical Biochemistry*. Prentice-Hall, Upper Saddle River, NJ.
- 100 VIGUERA, A. R., SERRANO, L., and WILMANN, M. 1996. Different folding transition states may result in the same native structure. *Nat. Struct. Biol.* **3**, 874–880.
- 101 WU, H. 1931. Studies on denaturation of proteins. XIII. A theory of denaturation. *Chin. J. Physiol.* **5**, 321–344.
- 102 YI, Q., SCALLEY-KIM, M. L., ALM, E. J., and BAKER, D. 2000. NMR characterization of residual structure in the denatured state of protein L. *J. Mol. Biol.* **299**, 1341–1351.
- 103 ZAGROVIC 2002. ■
- 104 ZAMAN, M. H., SHEN, M. Y., BERRY, R. S., FREED, K. F., and SOSNICK, T. R. 2003. Investigations into sequence and conformational dependence of backbone entropy, inter-basin dynamics and the Flory isolated-pair hypothesis for peptides. *J. Mol. Biol.* **331**, 693–711.
- 105 ZIMM, B. H. and BRAGG, J. K. 1959. Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* **31**, 526–535.