Change in backbone torsion angle distribution on protein folding

ANDREI-JOSÉ PETRESCU,^{1,2} PATRICK CALMETTES,² DOMINIQUE DURAND,^{2,3}
VERONIQUE RECEVEUR,² AND JEREMY C. SMITH⁴

¹Institute of Biochemistry of the Romanian Academy, 77700 Bucharest 17, Romania

²Laboratoire Leon Brillouin, CEA-Saclay, 91191 Gif sur Yvette, France

³LURE, Université Paris-Sud, 91401 Orsay, France

⁴Lehrstuhl für Biocomputing, IWR, Universität Heidelberg, Im Neuenheimer Feld 368, D-69120, Heidelberg, Germany

(RECEIVED January 3, 2000; FINAL REVISION April 18, 2000; ACCEPTED April 21, 2000)

Abstract

Understanding protein folding requires the determination of the configurational space accessible to the protein at different stages in folding. Here, computer simulation analysis of small angle neutron scattering results is used to probe the change in the distribution of configurations on strong denaturation of a globular protein, phosphoglycerate kinase, in 4 M guanidine hydrochloride solution. To do this atomic-detail ensembles of the unfolded protein chain are modeled and their scattering profiles compared with the experiment. The local conformational statistics are found to strongly influence the experimental intensity at scattering vectors between 0.05 and 0.3 Å⁻¹. Denaturation leads to a reduction in the protein atom-pair distance distribution function over the ~3–15 Å region that is associated with a quantifiable shift in the backbone torsional angle (φ, ψ) distribution toward the β region of the Ramachandran plot.

Keywords: atomic-detail models of unfolded protein ensemble; protein folding; small-angle neutron scattering

A complete understanding of protein folding kinetics and thermodynamics requires the determination of the configurational statistics of proteins at various stages in folding. This involves the development and combination of experimental and theoretical tools for characterizing the large number of possible configurations that can coexist in any given unfolded state (see, for example, Wolynes et al., 1995; Kataoka & Goto, 1997; Shortle, 1996; Dill & Chan, 1997; Dobson & Karplus, 1997; Duan & Kollman, 1998; Shakhnovich & Fersht, 1998; Smith et al., 1998; Radford & Dobson, 1999; Neira & Fersht, 1999; Meekhof & Freund, 1999).

Stable local structures in a denatured polypeptide chain can be experimentally detected with spectroscopies such as NMR, fluorescence, and circular dichroism. In particular, in recent years work using NMR has progressed to the stage where three-dimensional (3D) heteronuclear techniques allow resonance assignment and the measurement of structural data such as coupling constants, NOEs, and relaxation rates to be carried out with a detail comparable to studies of folded proteins (see, for example, Smith et al., 1996; Schwalbe et al., 1997; Hennig et al.,

1999). This has allowed information on the distribution of backbone torsional angles (ϕ, ψ) to be obtained that suggests that the local conformations of the highly denatured states are largely determined by short-range interactions within the polypeptide chain.

There is a need to complement the local information derived from NMR and other spectroscopic techniques with information on nonlocal (longer range) interactions so as to obtain a more complete description of the accessible configurational space of the chain. Moreover, it is also important to examine possible interdependence of the nonlocal and local configurational properties. To do this, techniques giving access to appropriate statistical configuration distribution functions are required. In this regard, solution scattering of neutrons and X-rays is particularly promising, as the scattered intensity is determined by pair distance distribution functions of the system under study.

Small-angle scattering is commonly used to determine the size (radius of gyration) of biological macromolecules. Although this is valuable information for denatured states (Eliezer et al., 1995; Kataoka et al., 1995; Chen et al., 1996; Kataoka & Goto, 1997), there exists the additional possibility that this technique can be extended to provide a more detailed picture, involving the determination of the chain flexibility at higher resolution. With this aim, we have pursued a program to characterize the configurational statistics of a strongly-denatured protein (Calmettes et al., 1993, 1994; Petrescu et al., 1997, 1998; Receveur et al., 1997). The system investigated is phosphoglycerate kinase (PGK) in 4 M

Reprint requests to: Jeremy C. Smith, Lehrstuhl für Biocomputing, IWR, Universität Heidelberg, Im Neuenheimer Feld 368, D-69120 Heidelberg, Germany; e-mail: biocomputing@iwr.uni-heidelberg.de.

Abbreviations: NOE, nuclear Overhauser effect; PDB, Protein Data Bank; PGK, phosphoglycerate kinase; SANS, small-angle neutron scattering; LE, locally extended; EVB, excluded-volume beta; EVN, excluded-volume native-like.

guanidinium chloride, chosen as a model of an unfolded protein without disulfide bonds. Small-angle neutron scattering (SANS) was chosen over its X-ray counterpart as it allows the difference in sign of the scattering lengths of hydrogen and deuterium to be exploited so as to optimize the contrast between the protein and solvent at high concentrations of denaturant.

The radius of gyration of the denatured PGK, extrapolated to zero protein concentration, was found to be 90.0 ± 4.5 Å, largely increased from that of the native state (23.8 ± 0.3 Å). At higher resolution, to examine the configurational statistics it is necessary to construct theoretical models of the protein chain, to calculate their scattering properties, and compare with experiment. In this way, it was shown that the SANS data can be reproduced by the scattering of a random flexible chain of spheres of excess scattering density, with five residues per sphere, subject to excluded volume interactions (Petrescu et al., 1998). This coarse-grained modeling quantifies the increase on unfolding of the long-range (≥ 20 Å) chain flexibility.

Clearly, the ultimate aim of experiments on unfolded states is to obtain an atomic-detail description of the configurational distribution. Small-angle scattering experiments are limited in this endeavor insofar as they yield principally low-resolution, spherically-averaged information. However, the possibility does exist that the effects of statistical local conformational preferences might propagate into the small-angle region. This is examined here, using atomistic models of unfolded PGK. It is found that the effects of denaturation on local chain conformational probability distributions are visible in the experimental profile and involve a quantifiable change in the pair distance distribution function associated with a shift in the φ, ψ distribution toward the β region of the Ramachandran plot.

Results

To describe the chain statistics, three atomic-detail models of the denatured protein were generated. They differ in the distribution of φ, ψ backbone torsion angles sampled in construction of the polypeptide chains, and in the presence or absence of excluded volume interactions, i.e., noninterpenetration of chain segments at long distances. All models were built in two steps: first the low-resolution pathway of the chain was defined with a chain-of-spheres, and then atomic-detail models were built into the spheres. Full descriptions are given in Materials and methods. The first model we call "Locally Extended" (LE), as this describes the resulting peptide chain conformation. The aim of this was to build a random polypeptide chain without imposing excluded volume interactions that is in as close agreement as possible with the SANS experiment. The second and third models are called "Excluded Volume" (EV). In contrast to LE, for these models excluded volume interactions, i.e., repulsion between chain segments distant in the sequence, were included. For the "EV Beta" (EVB) model random, nonnative segments of polypeptide chain were used to build atomic-detail models into the spheres, whereas for "EV Native" (EVN) the atomic model was constructed so as to conserve as much as possible the local structure of the native protein.

A measure of the difference in the local configurational statistics of the three atomic-detail models is the way they sample φ, ψ space. The φ, ψ distributions obtained before and after the constrained energy minimization of the atomic-detail chains and from the native protein crystal structure are shown in Figure 1. The corresponding θ distributions of the energy-minimized chains are

shown in Figure 2 (see Materials and methods for definition of θ). The local structure in EVN was derived from the native protein crystal structure, and these two φ, ψ distributions are correspondingly similar. In the φ, ψ distribution obtained from the LE procedure, most residues are found in the β or extended region (upper left quadrant) of the φ, ψ map, hence the retrospective "LE" acronym, and this remained the case after energy minimization. Energy minimization of the EVB model, in which five-residue segments of the LE chain were oriented in spheres (see Materials and methods), leads to significant flattening of the θ distribution (see Fig. 2). This shifts the φ, ψ distribution somewhat toward that of the native structure, although it still has a larger population in the upper left (β) quadrant of the map than does the native protein. In both the φ, ψ maps and in the θ distributions, then EVB lies between LE and EVN.

The origin of the φ, ψ differences is related to the contour length of a sphere model C , defined as Nl where N is the number of spheres of diameter l . The values for LE are $N = 415$ and $l = 3.8$ Å, giving $C = 1,577$ Å, whereas for the EV models they are $N = 83$ and $l = 14.0$ Å, giving $C = 1,162$ Å. Therefore, C is $\sim 35\%$ longer in LE. Consequently, the LE chains are less locally compact than EV. This reflects the fact that to achieve the same radius of gyration chains without excluded volume require local expansion and hence must adopt principally extended local structure. The difference in local compactness perpendicular to the chain axis is visible in Figure 3, in which typical segments of the three models are shown.

Individual configurations of the whole chain are shown in Figure 4. The 14 Å sphere diameter of the EV models is significantly larger than the space occupied by the associated amino acid side chains. Consequently, few nonlocal contacts (i.e., between residues far away in the sequence) are seen in the EV atomic-detail models, in contrast to the LE model, for which frequent nonlocal chain contacts are observed.

Kratky plots, which are suitable for probing chain statistics, are presented for the three models and the experiment in Figure 5. The theoretical profiles differ significantly from one another above $q \sim 0.03$ Å⁻¹.

The difference between the LE and EVB scattering profiles arises primarily from excluded volume. Using both analytical and sphere-based models, it has been demonstrated that PGK under the present conditions exhibits excluded-volume behavior (Calmettes et al., 1994; Petrescu et al., 1998). Under the present conditions, the long-range interactions between different segments of the chain are generally repulsive—this is seen both in the present excluded-volume chain modeling and in the experimental determination of the second virial coefficient (see Petrescu et al., 1998). In 4 M guanidinium chloride, the protein–protein charge–charge interactions are highly screened. Moreover, our description is consistent with the absence of long-range interactions as seen in most NMR studies of highly-unfolded proteins to date.

In the present work, the deviation of LE from experiment is clear both at low q ($0.03 < q < 0.06$ Å⁻¹) where the intensity is overestimated, and for $0.06 < q < 0.3$ Å⁻¹ where it is underestimated. Therefore, the present work demonstrates that a model without excluded volume is inappropriate also in the case of explicit all-atom modeling.

The progression LE $\rightarrow$ EVB $\rightarrow$ EVN is accompanied by increased intensity of $q^2 I(q)$ in the intermediate q region: $0.05 < q < 0.2$ Å⁻¹. Excluded volume is identically treated in the EV models, and therefore the differences between them arise solely

Fig. 1. ϕ, ψ distributions. “Random Beta” corresponds to the LE distribution before energy minimization. The energy-minimized LE, EVB, and EVN distributions are also shown, together with that of the native crystal structure. Arrows indicate the distributions from which energy minimization was initiated.

Fig. 2. Normalized distribution of the virtual bond angle θ in energy-minimized chain ensembles.

from their ϕ, ψ distributions. Inspection of Figure 5 indicates that model EVB is the closest to experiment and is the only model remaining within experimental error for the whole q range.

The origin of the differences in the scattering profile can be traced to the atom pair distance distribution functions, $p(r)$ from the three models, shown in Figure 6. $p(r)$ is given by the Fourier transform of $I(q)$. Marked differences in $p(r)$ are seen above 3 Å and correlate with the differences in the Kratky plots. In calculations both with or without explicit solvent molecules included, the progression of the number density in the ~3–15 Å region is $\text{EVN} > \text{EVB} > \text{LE}$, consistent with the above-mentioned higher local compactness for the more native-like model.

Discussion

Strongly denatured proteins are often considered to be completely unfolded, i.e., in a “random coil” state with no residual structure.

Fig. 3. Local structure of atomic-detail models demonstrating differences in cross-sectional compactness.

But even in such a state the polypeptide chain will have local conformational preferences. This leads to fundamental problems in protein folding concerning the determination of these preferences, environmental effects on them, their reconciliation with nonlocal configurational statistics and possible deviations from native state

local conformational distributions. Here, we have presented a new approach that promises to be of use in addressing these problems. The results demonstrate how the combination of atomic-detail simulation with small-angle neutron scattering can be used to provide information on the both the global and local configurational sta-

Fig. 4. The overall path for one member of the ensemble of chains of each model. Green, EVN; purple, EVB; yellow, LE.

Fig. 5. Experimental Kratky plot compared with those of the configurationally-averaged atomic-detail models.

tistics of a strongly denatured protein. The scattering profile distinguishes between different apparently physically-realistic ϕ, ψ distributions of the polypeptide chain: for identical medium-resolution (chains-of-spheres) paths, the scattering profile of atomic models containing a native-like ϕ, ψ distribution (EVN) can be clearly distinguished from one containing a higher proportion of β structure (EVB). It is found here that denaturation leads to the polypeptide chain adopting EVB statistics. This involves a reduction in the pair distance density in the $\sim 3\text{--}15\text{ \AA}$ range, associated with a shift in the ϕ, ψ distribution toward the upper left-hand quadrant of the Ramachandran plot (the β region), and a flattening of the corresponding θ distribution. The flat θ distribution ensures that local conformational freedom remains for the chain, in contrast to the LE model. As LE was built assuming the absence of excluded volume interactions, it requires local extension of the polypeptide chain to attain the experimental radius of gyration, restraining the accessible ϕ, ψ space. Therefore, for LE not only does the best possible scattering profile still disagree with experiment, but also the chain configurational entropy might be expected to be unfavorably low.

Fig. 6. Atom-pair radial distribution function from the three models. Upper: of the protein atoms only. Lower: including the explicit solvent atoms.

The present approach provides information complementary to that obtained by NMR, in which coupling constants have been interpreted in terms of averaged residue-dependent backbone angle distributions taken from native proteins from the Protein Data Bank (PDB) (Bolin et al., 1996; Smith et al., 1996; Penkett et al., 1997). In the context of the present work, particularly relevant results were obtained by Smith et al. (1996), who examined $^3J_{HN\alpha}$ coupling constants predicted from the PDB distributions and experimental values for a set of peptides that other evidence suggest are highly unstructured. Two PDB-derived models were examined, one in which the population of ϕ, ψ space for a random coil residue was taken as that for all residues of the same type in the database (including secondary and coil structure) and another where only residues in coils were considered. The latter coil model has a distribution shifted somewhat away from the α to the β region of the Ramachandran plot. Better agreement with experiment was seen for the coil set, a result in a similar spirit to that obtained here. Using a similar procedure, Serrano (1995) found good agreement between the $^3J_{HN\alpha}$ coupling constants calculated using the distribution of ϕ angles for coil residues in the PDB and those measured for some short tetra- and pentapeptides. In later work on a highly unfolded 130-residue fragment of the *Staphylococcus aureus* fibronectin-binding protein, a significant increase was found in the population of the β region of ϕ, ψ space for residues preceded by a β -branched or aromatic side chain (Penkett et al., 1997).

In NMR studies of proteins under highly denaturing conditions, evidence has been found for residual structure (e.g., Fong et al., 1998; Koepf et al., 1999; Luisi et al., 1999). The question arises

whether this might be in conflict with the present modeling. However, residue interactions seen by NMR in highly unfolded states have tended to be local in the sequence, and such interactions are unlikely to change the $p(r)$ functions significantly. Moreover, if $p(r)$ were significantly changed, it would no longer be in agreement with the SANS experiment.

Whereas NMR has the advantage of providing local, residue-specific information, the present approach allows the long- and short-range residue-averaged properties of the chain to be combined, via the radial distribution function. In the “folding funnel” parlance, the width of the funnel is related to the accessible configurational space of the protein chain. As the system examined here, strongly denatured, can be thought of as representing the unfolded end state of protein folding, the present results provide information on the widest rim of the folding funnel, and which regions of conformational space it contains. This information can be directly incorporated in statistical mechanical analyses of protein folding. With further experimental and theoretical work, the present combination of experiment and theory may also prove fruitful in determining the configurational space of different proteins and at other points in the funnel, i.e., at various stages in the folding process.

Materials and methods

Generation of atomic-detail chains

The experimental method and scattering profile are documented in Petrescu et al. (1998). In the theoretical analysis, each atomic-detail model consists of an ensemble of chains constructed in two stages. (1) In the first stage, a low resolution model, consisting of an ensemble of chains of spheres of constant excess scattering length density, was generated by a simple Monte Carlo procedure. The parameters describing the chain ensemble were varied until the calculated scattering profile reproduced the low scattering wave-vector (q) portion of the experimental scattering curve ($q < 0.05 \text{ \AA}^{-1}$). This ensures that the configurationally-averaged shape of the chain and its radius of gyration are reproduced. The methods for this stage and the results of the application of a variety of chain-of-sphere models are described in detail in Petrescu et al. (1997, 1998). (2) In the second stage the fitted chain of spheres model was used to provide bounds for the construction of an all-atom model. The atomic-detail models constructed obey the sphere paths. In all cases, the very low- q profile is unaltered when atomic detail is included—the effects of going to atomic-detail resolution are seen at higher q .

Using the above procedure, three models were constructed:

LE model

The aim of the LE (“Locally Extended”) model was to build a completely random polypeptide chain, without imposing excluded volume interactions, that respects elementary properties of peptide backbone geometry and is in as close agreement as possible with the SANS experiment. “Locally Extended” refers to a result of this model, as described in Results. In the first stage, the scattering object was modeled as an ensemble of chains of spheres with one residue per sphere ($p = 1$), i.e., 415 spheres. The absence of excluded volume means that, although interpenetration between adjacent spheres is not permitted, it is allowed for second neighbour and higher-order spheres. The chain-of-spheres geometry can

Fig. 7. Local geometrical parameters used for chain generation statistics.

be simply described, as shown in Figure 7, in terms of θ , the angle between adjacent vectors connecting three consecutive sphere centers, and the complementary angle ω . In this representation, the chain statistics are only weakly dependent on ω , which can to a first approximation be sampled from a uniform distribution (Flory, 1953). The sphere radius was varied until the experimental scattering profile at low q was reproduced.

An LE model in agreement with experiment at low q cannot be derived using the θ distribution of the native protein—this always resulted in a radius of gyration two to three times smaller than experiment. Furthermore, it is instructive to consider the effect of the peptide chain geometry on the θ distribution required to fit the low- q data. For a simple $p = 1$ chain of spheres in which adjacent spheres do not interpenetrate, θ is constrained to fall within 0 – 120° . It was found that if in the chain construction θ is sampled uniformly over this range then the diameter of the spheres required to fit the low- q scattering profile is $4.70 \text{ \AA}$. This is unrealistically large, as for a $p = 1$ polypeptide chain it should be close to the C_α – C_α distance which is $3.8 \text{ \AA}$. However, consideration of basic peptide geometry itself leads to further restriction of θ , to ~ 37 – 110° , the extreme values of which correspond to $(\varphi, \psi) = (180^\circ, 180^\circ)$ and $(0^\circ, 0^\circ)$, respectively. When θ is uniformly sampled over the 37 – 110° range, the required sphere diameter is lowered to $4.20 \text{ \AA}$. As this is still significantly larger than the C_α – C_α distance, we conclude that θ cannot be uniformly distributed in the range 37 – 110° . An optimal nonuniform distribution was determined, such that the correct low- q scattering is preserved and elementary aspects of local peptide geometry are obeyed (diameter = $3.8 \text{ \AA}$). This has a peak at $\theta \sim 45^\circ$, as compared to a mean value of $\theta \sim 75^\circ$ for the uniform distribution over 37 – 110° .

In the second stage of the LE modeling, the derived θ distribution was used to generate a corresponding distribution of φ, ψ polypeptide backbone angles. The obtained φ, ψ angle distribution was then randomly sampled to obtain φ, ψ sequences, which were used to construct an ensemble of atomic-detail polypeptide chains. The chains thus generated were energy minimized with the CHARMM program (Brooks et al., 1983) and potential function version 25. To conserve the low resolution features of each path, the minimizations were performed with harmonic constraints on every tenth C_α atom along the sequence.

EV models

The second and third models are called “Excluded Volume” (EV). In the first stage of construction of both EV models, the scattering object was again considered to be an ensemble of chains of spheres of excess scattering density. However, in contrast to the LE model, this time excluded volume interactions, i.e., repulsion between chain segments distant in the sequence, were included. The excluded-volume spheres model used here has been described in detail in Petrescu et al. (1998) and is a $p = 5$ chain of spheres of radius $7 \text{ \AA}$ with interpenetration allowed for up to 30% of the sphere diameter. This model is in agreement with the SANS experiment over the whole accessible q -range (Petrescu et al., 1998).

In the second stage of the EV modeling, two different methods were used:

EVB model. For the EVB (“EV Beta”) model random, non-native segments of polypeptide chain were used to build atomic-detail models into the spheres. For this, five-residue-long segments of the LE-generated polypeptide chain were translated and rotated into the spheres, and energy-minimized as described in the LE-modeling section above.

EVN model. For EVN (“EV Native”) the atomic model was constructed so as to conserve as much as possible the local structure of the native protein. To do this, consecutive groups of five amino acids from the native PGK crystal structure (Watson et al., 1982) were translated and rotated into the individual spheres. The protein was then energy minimized with harmonic constraints on every C_α .

Previous work indicated that the cross-sectional diameter of the basic scattering unit of excess scattering density (i.e., the spheres) is about $14 \text{ \AA}$ (Petrescu et al., 1998). This diameter is larger than the cross-sectional diameter of the polypeptide chain. Consequently, it is necessary to include contrast between the solvent region close to the peptide chain (out to a $\sim 14 \text{ \AA}$ diameter) and the bulk.

To do this in a simple way, a box of explicit solvent molecules was constructed and equilibrated with molecular dynamics using CHARMM. The details of construction of the box and the exact composition are unimportant for the present purposes (see below). The equilibrated box was translated along the protein chain and solvent molecules overlapping the protein or at a distance greater than $6 \text{ \AA}$ from any protein atom were deleted.

Calculation of the scattering profiles

SANS profiles of individual atomic-detail protein configurations were computed as described in Petrescu et al. (1997). In this method, each residue is partitioned into one backbone and one or two side-chain groups of atoms, depending on its size. Each group of atoms is considered to be a point scatterer positioned at the scattering-length centroid of the group. The scattering from this simplified representation is about 60 times quicker to calculate than that of all the atoms and has been shown to be accurate to within 1% of the all-atom scattering curve (Petrescu et al., 1997). The scattering profile $I(q)$ was calculated as follows:

$$I(q) = \sum_{i,j}^N b_i^c b_j^c \frac{\sin(qr_{ij})}{qr_{ij}} \quad (1)$$

where N is the number of scattering units in the PGK chain and associated explicit solvent, and r_{ij} the distance between the scat-

tering points i, j . b_i^c takes into account the contrast between the explicit atoms and the bulk solvent (i.e., that outside the explicit solvent region) as follows:

$$b_i^c = \sum_{\nu} b_{i,\nu} - \frac{\theta^{\text{group}}}{\theta^{\text{bulk}}} b^{\text{bulk}} \quad (2)$$

where ν runs over the group of atoms forming the scattering unit and θ^{group} and θ^{bulk} are the molecular volumes of the group and the average bulk solvent molecule (corresponding to 4 M GdmCl in D₂O solution), respectively. b^{bulk} is the corresponding average bulk solvent scattering length.

For each of the three models, 50 random-walk configurations of the protein chain were generated and their scattering profiles computed and averaged, assuming that all solvent is a bulk continuum (50 configurations has been shown to be sufficient to generate a converged SANS profile (Petrescu et al., 1997)). To examine the effect of the explicit solvent, single configurations of the protein chain were explicitly solvated as described above and the scattering curve from the ~25,000 atom system calculated. The chemical composition of the explicit solvent region is unknown and undeterminable from the neutron data. Therefore, to simplify, the centroids of the explicit solvent molecules were considered as scattering points each with the same, single scattering length b_i^c . The effect of the explicit solvent on the scattering profile was found to be similar for each model and configuration examined and is equivalent to multiplying by a function increasing approximately monotonically with q . Therefore, the explicit solvent correction was subsequently applied to the configurationally-averaged scattering profiles of the protein chain. Closest agreement with experiment was found for all models of the protein configurations with the same b_i^c of -1.6 ± 0.4 fm. Variation of b_i^c within physically reasonable limits produced results closely similar to those shown in Results.

References

- Bolin KA, Pitkeathly M, Miranker A, Smith LJ, Dobson CM. 1996. Insight into random coil conformation and an isolated helix: Structural and dynamical characterization of the C-helix peptide from hen lysozyme. *J Mol Biol* 261:443–453.
- Brooks BR, Brucoleri RE, Olafson B, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular minimization and dynamics. *J Comput Chem* 4:187–217.
- Calmettes P, Durand D, Desmadril M, Minard P, Receveur V, Smith JC. 1994. How random is a highly-denatured protein? *Biophys Chem* 53:105–114.
- Calmettes P, Roux B, Durand D, Desmadril M, Smith JC. 1993. Configurational distribution of strongly-denatured phosphoglycerate kinase. *J Mol Biol* 231:840–848.
- Chen L, Hodgson KO, Doniach S. 1996. A lysozyme folding intermediate revealed by solution X-ray scattering. *J Mol Biol* 261:658–671.
- Dill KA, Chan HS. 1997. From Levinthal to pathways and funnels. *Nat Struct Biol* 4:10–18.
- Dobson CM, Karplus M. 1997. Fundamentals of protein folding: Bringing together theory and experiment. *Curr Opin Struct Biol* 9:92–101.
- Duan Y, Kollman PA. 1998. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 282:740–744.
- Eliezer D, Jennings PA, Wright PE, Doniach S, Hodgson KO, Tsuruta H. 1995. The radius of gyration of an apomyoglobin folding intermediate. *Science* 270:487–488.
- Flory PJ. 1953. *Principles of polymer chemistry*. Ithaca, NY: Cornell University Press.
- Fong S, Bycroft M, Clarke J, Freund SMV. 1998. Characterisation of urea-denatured states of an immunoglobulin superfamily domain by heteronuclear NMR. *J Mol Biol* 278:417–429.
- Hennig M, Bermel W, Spencer A, Dobson CM, Smith LJ, Schwalbe H. 1999. Side-chain conformations in an unfolded protein: Chi1 distributions in denatured hen lysozyme determined by heteronuclear ¹³C, ¹⁵N NMR spectroscopy. *J Mol Biol* 288:705–723.
- Kataoka M, Goto Y. 1997. X-ray solution scattering studies of protein folding. *Fold Des* 1:107–114.
- Kataoka M, Nishii I, Fijisawa T, Ukey T, Tokunaga F, Goto Y. 1995. Structural characterization of the molten globule and native states of apomyoglobin. *J Mol Biol* 249:215–228.
- Koepp EK, Petrassi HM, Sudol M, Kelly JW. 1999. WW: An isolated three-stranded antiparallel beta-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(4):841–853.
- Luisi DL, Wu WJ, Raleigh DP. 1999. Conformational analysis of a set of peptides corresponding to the entire primary sequence of the N-terminal domain of the ribosomal protein L9: Evidence for stable native-like secondary structure in the unfolded state. *J Mol Biol* 287:395–407.
- Meekhof AE, Freund SM. 1999. Probing residual structure and backbone dynamics on the milli- to picosecond timescale in a urea-denatured fibronectin type III domain. *J Mol Biol* 286:579–592.
- Neira JL, Fersht AR. 1999. Exploring the folding funnel of a polypeptide by biophysical studies on protein fragments. *J Mol Biol* 285:1309–1333.
- Penkett CJ, Redfield C, Dodd I, Hubbard J, McBay DL, Mossakowska DE, Smith RAG, Dobson CM, Smith LJ. 1997. NMR analysis of main-chain conformational preferences in an unfolded fibronectin-binding protein. *J Mol Biol* 274:152–159.
- Petrescu AJ, Receveur V, Calmettes P, Durand D, Desmadril N, Roux B, Smith JC. 1997. Small angle scattering by a strongly-denatured protein: Analysis by random polymer theory. *Biophys J* 72:335–342.
- Petrescu AJ, Receveur V, Calmettes P, Durand D, Smith JC. 1998. Excluded volume in the configurational distribution of a strongly-denatured protein. *Protein Sci* 7:1396–1403.
- Radford SE, Dobson CM. 1999. From computer simulations to human disease: Emerging themes in protein folding. *Cell* 97:291–298.
- Receveur V, Calmettes P, Smith JC, Desmadril M, Coddens G, Durand D. 1997. Dynamical change on denaturation of phosphoglycerate kinase revealed by quasielastic neutron scattering. *Proteins Struct Funct Genet* 28:380–387.
- Schwalbe H, Fiebig KM, Buck M, Jones JA, Grimshaw SB, Spencer A, Glaser SJ, Smith LJ, Dobson CM. 1997. Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry* 36:8977–8991.
- Serrano L. 1995. Comparison between the ϕ angle distribution of amino acids in the protein data base and NMR data indicates that amino acids have various propensities in the random coil conformation. *J Mol Biol* 254:322–333.
- Shakhnovich E, Fersht AR. 1998. Protein folding: Think globally, interact locally. *Curr Opin Struct Biol* 8:65–67.
- Shortle DR. 1996. Structural analysis of non-native states of proteins by NMR. *Curr Opin Struct Biol* 6:24–30.
- Smith LJ, Bolin KA, Schwalbe H, MacArthur MW, Thornton JM, Dobson CM. 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 LJ, Mark AE, Dobson CM, van Gunsteren WF. 1998. Molecular dynamics simulations of peptide fragments from hen lysozyme: Insight into non-native protein conformations. *J Mol Biol* 280:703–719.
- Watson HC, Walker NPC, Shaw PJ, Bryant TN, Wendell PL, Fothergill LA, Perkins RE, Conroy SC, Dobson MJ, et al. 1982. Sequence and structure of yeast phosphoglycerate kinase. *EMBO J* 1:1635–1640.
- Wolynes PG, Onuchic J, Thirumalai D. 1995. Navigating the folding routes. *Science* 267:1619–1620.