Biochemistry

 Volume 27, Number 19 September 20, 1988

Perspectives in Biochemistry

Conformation of Peptide Fragments of Proteins in Aqueous Solution: Implications for Initiation of Protein Folding†

Peter E. Wright,* H. Jane Dyson, and Richard A. Lerner
Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, California 92037
Received July 12, 1988

ABSTRACT: Applications of sensitive new technologies, in particular, two-dimensional NMR spectroscopy, have allowed detection of folded structures in short peptide fragments of proteins in aqueous solution under conditions where native proteins fold. These structures are in rapid dynamic exchange with unfolded states. These observations provide evidence in support of models for protein folding which postulate localized regions of folded structure as initiation sites for the folding process. Since these initiation processes are expected to be rapid, such models are consistent with kinetic evidence that the rate-determining steps of protein folding occur late in the process and probably involve rearrangement of incorrectly folded intermediates.

The mechanism by which proteins fold into their native three-dimensional structures remains one of the central unsolved problems of molecular biology. Since the early experiments of Anfinsen and co-workers [reviewed in Anfinsen (1973)], it has been recognized that protein folding is a spontaneous event and that all the information required for correct folding is contained within the amino acid sequence. The available experimental data indicate that many proteins fold in vitro into their native conformations on a time scale ranging from less than a second to a few minutes. It is clear that folding cannot occur by a random search of all conformations, which, for a protein of only 100 amino acid residues, would take on the order of 10^{50} years or longer (Levinthal, 1968; Wetlaufer, 1973; Karplus & Weaver, 1976). It is now generally accepted that protein folding does not occur by a random conformational search but proceeds via local folded intermediates that function as sites for cooperative growth (Anfinsen, 1972; Wetlaufer, 1973; Ptitsyn & Rashin, 1975; Richards, 1977; Ptitsyn & Finkelstein, 1980; Scheraga, 1980; Jaenicke, 1980; Lesk & Rose, 1981; Richardson, 1981; Ghelis & Yon, 1982; Kim & Baldwin, 1982; Karplus & Weaver, 1976; Weaver, 1984). A large number of mechanistic models of protein folding have been proposed. These can be reduced to two generalized working models (Kim & Baldwin, 1982), the framework model, in which elements of secondary structure are formed early in folding, and modular assembly, in which subdomains of the protein fold completely without the need for prior formation of secondary structure.

Much is now known about the later, slow stages of protein folding that frequently involve proline cis–trans isomerization (Brandts et al., 1975) or rearrangement of incorrectly formed disulfide bridges (Creighton, 1978). Experimental evidence for formation of kinetically observable intermediates during the later stages of protein folding is accumulating, and techniques such as NMR[1] spectroscopy are beginning to provide direct information about the nature of these species (Roder & Wüthrich, 1986; States et al., 1987). However, identification of the transient structures that are likely to initiate protein folding and direct the folding pathway(s) is very much more difficult because the polypeptide chain folds rapidly and cooperatively into a compact folded structure. In principle, it should be possible to identify folding initiation sites by studying peptide fragments of proteins for evidence of folded structures. However, early attempts to find structure in peptide fragments of proteins were generally unsuccessful or gave equivocal results (Epand & Scheraga, 1968; Taniuchi & Anfinsen, 1969; Hermans & Puett, 1971; Howard et al., 1975)

† Supported by Grants GM38794 and CA27489 from the National Institutes of Health.

[1] Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser effect spectroscopy; ROESY, rotating frame NOESY spectroscopy; $d_{\alpha N}(i,j)$, $d_{NN}(i,j)$, etc., intramolecular distance between the protons $C^{\alpha}H$ and NH, NH and NH, etc. on residues i and j; $^3J_{HN\alpha}$, NH–$C^{\alpha}H$ coupling constant; CD, circular dichroism; FTIR, Fourier transform infrared spectroscopy.

0006-2960/88/0427-7167$01.50/0

and led to a general belief that short linear peptides do not contain secondary structure in water solution. For many years, the only exception appeared to be the C-peptide (residues 1–13) of ribonuclease A, which was shown to adopt helical structure in the monomeric state in aqueous solution at temperatures near 0 °C (Brown & Klee, 1971).

Most of the early attempts to identify secondary structures (i.e., putative folding initiation sites) in peptide fragments of proteins were probably unsuccessful for two main reasons. The first problem lies in the selection of appropriate peptides for study. A priori prediction of peptides that are likely to contain protein folding initiation sites on the basis of the three-dimensional structure of the folded protein (Tanaka & Scheraga, 1977; Matheson & Scheraga, 1978; Zehfus & Rose, 1986) is difficult at best since it is by no means obvious that structures involved in initiation of folding will be retained in the final folded protein. Secondly, the populations of folded conformations are expected to be rather small, and sensitive techniques are required for their detection. Most early attempts to detect structured regions of peptide fragments of proteins were performed with relatively insensitive methods. Advances in antipeptide antibody technology and two-dimensional NMR spectroscopy now provide a means both to identify peptides that have a high folding propensity in aqueous solution and to obtain direct experimental information on their preferred conformations.

It is now well established that short synthetic peptides are able to induce antibodies that recognize, with high frequency, the cognate sequence in the folded protein (Lerner, 1984; Niman et al., 1983). This phenomenon is difficult to explain unless the peptide is able to adopt a highly preferred conformation, either free in solution or on the B-cell receptor, approaching that which it adopts in the native protein or else that the protein itself approaches disorder through conformational fluctuations (Tainer et al., 1984, 1985; Dyson et al., 1988c). We have recently demonstrated that several short peptide sequences, each capable of inducing antibodies that bind to the cognate sequence in the folded protein, exhibit conformational preferences for secondary structure (helix, turns) in aqueous solution (Dyson et al., 1985, 1988b). These experiments, which are described in more detail below, have led us to propose that highly immunogenic peptides frequently have a high propensity to adopt folded structures (Dyson et al., 1988c). Immunological presentation of peptide immunogens has a problem in common with protein folding in that, if an antibody that cross-reacts with the folded protein is to be successfully induced, the conformational space available to the peptide must be reduced.

The potential of NMR for conformational studies of flexible linear peptides has long been recognized. Until recently however, this potential has not been realized because of problems arising from averaging of NMR parameters over all populated conformations. Short linear peptides in water typically gave "random coil" NMR spectra, and evidence for nonrandom conformations could only be obtained by indirect methods (Wüthrich & Grathwohl, 1974; Bundi & Wüthrich, 1979). In contrast, relatively stable, well-defined structures have often been found for linear peptides in organic solvents. However, these are largely irrelevant for identification of folding initiation structures, which must perforce be formed in an aqueous medium during protein folding.

With improvements in experimental techniques, considerable evidence has accumulated in recent years which shows that short linear peptide fragments of proteins are not always the disordered, structureless entities in aqueous solution that the early experiments suggested. This evidence will be reviewed in the present paper, along with the profound implications that observation of peptide structures has for mechanisms of protein folding. We exclude from this review peptide fragments that are large enough to form stable, independently folded domains (Wetlaufer, 1981).

Identification of Peptide Structures by NMR Spectroscopy. Two-dimensional NMR spectroscopy provides a highly sensitive approach for detection and identification of folded structures in aqueous solutions of peptides. In contrast to many other forms of spectroscopy, such as circular dichroism, infrared, and Raman spectroscopy, NMR spectroscopy provides structural information, once resonances have been assigned, at specific sites throughout the peptide. Several NMR parameters provide information about molecular structure. Information about dihedral angles can be obtained from spin–spin coupling constants, the existence of hydrogen bonding can be inferred from the temperature coefficients of the amide proton chemical shifts and from amide proton exchange data, and most importantly, direct information on the distances between protons in the peptide can be inferred from the nuclear Overhauser effect (NOE). The mere observation of a direct NOE between two protons implies the presence of at least a threshold population of conformers in which these protons are close together. In folded forms of a peptide, protons located in distant parts of the peptide chain are brought into close proximity. Since the NOE is weighted according to the sixth power of the interproton distance, NOEs can be observed and used as a diagnostic of structure for even quite small populations of folded conformers.

Depending on the size of the peptide, NOE data are obtained from conventional NOESY spectra (Jeener et al., 1979) or from rotating-frame NOESY (ROESY) spectra (Bothner-By et al., 1984). ROESY spectra are employed for short peptides for which $\omega_0\tau_c \leq 1$. For such peptides NOESY cross peaks either are nonexistent or are too weak to be observed reliably (Ernst et al., 1987). In practice, it appears difficult to observe NOEs between protons that are more than approximately 3.5 Å apart in small peptides, even when long mixing time NOESY or ROESY experiments are used (Dyson et al., 1988a,b). This arises in part because of the short correlation times of monomeric linear peptides of fewer than 20–30 residues in aqueous solution. The populations of the folded forms are frequently rather small, which also limits the interproton distance for which an NOE can be observed.

Interpretation of NOE data for a small linear peptide in water solution is complicated by conformational averaging. Barriers between low-energy minima in ϕ,ψ space have been estimated to be only a few kilocalories per mole (Zimmerman et al., 1977), and hence peptides are expected to sample a number of backbone conformations on a nanosecond time scale. NMR parameters such as the chemical shift and coupling constant are thus a population-weighted average over all conformers, and the NOESY spectrum will contain cross peaks representative of all conformations that have sufficiently high populations. Fortunately, steric constraints impose restrictions on the allowed backbone conformations (Ramachandran et al., 1963) so that only a relatively limited range of backbone ϕ and ψ dihedral angles is heavily populated (Figure 1). This makes it possible to interpret, at a qualitative level, patterns of NOEs involving backbone protons. It should be noted, however, that it is not possible to use NOE-derived distance constraints and current distance geometry or molecular dynamics methods to determine meaningful structures for flexible peptides that adopt multiple conformations.

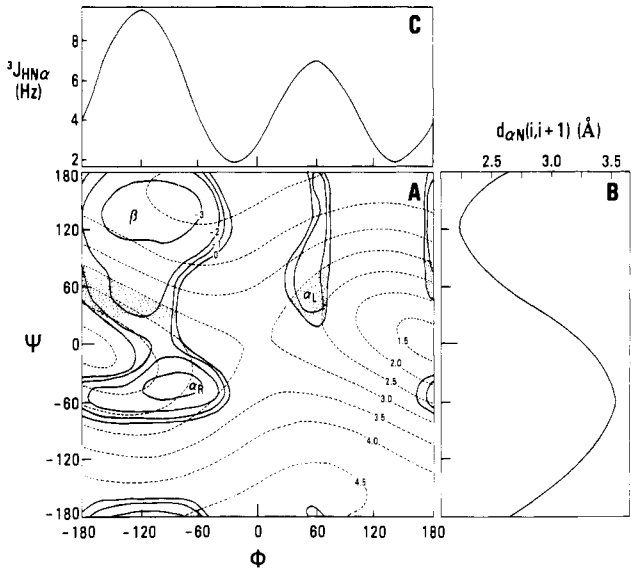

FIGURE 1: (A) Conformational energy diagram for the alanine dipeptide [adapted from Ramachandran et al. (1966)]. Energy contours (solid lines) are drawn at intervals of 1 kcal·mol^{-1} decreasing from 0 kcal·mol^{-1}. The potential energy minima for the extended structure (β), right-handed α-helix (α_R), and left-handed α-helix (α_L) are labeled. The dashed lines show the dependence of the sequential $d_{NN}(i,i + 1)$ distance (in angstroms) on the ϕ and φ dihedral angles (Billeter et al., 1982). Hatched regions show allowed regions of ϕ,φ space where both $d_{NN}(i,i + 1)$ and $d_{\alpha}N(i,i + 1)$ are less than 3.0 Å. (B) Plot of the $d_{\alpha N}(i,i + 1)$ distance as a function of the dihedral angle φ [$d_{\alpha N}(i,i + 1)$ does not depend on ϕ] (Billeter et al., 1982). (C) Variation of the coupling constant $^3J_{HN\alpha}$ between the NH and C$^{\alpha}$H protons of the same amino acid residue as a function of the dihedral angle ϕ (Pardi et al., 1984).

Patterns of NOE Connectivities and Backbone Conformations. The distance between backbone protons on adjacent amino acid residues of a peptide depends only on the dihedral angles ϕ and ψ (Figure 1). Patterns of sequential NOE connectivities between protons on adjacent residues, the $d_{\alpha N}(i,i + 1)$ and $d_{NN}(i,i + 1)$ NOE connectivities in the terminology of Wüthrich et al. (1984), provide information about ϕ and ψ and hence about backbone conformation. The potential energy diagram shown in Figure 1A can be used as a guide for interpretation of NMR data, and particularly the NOE, for linear peptides. While there are still disagreements about the detailed features of conformational energy surfaces, it is evident that broad minima occur for non-glycine peptides in the β (extended chain), α_R, and α_L regions (Zimmerman et al., 1977; Lau & Pettitt, 1987; Anderson & Hermans, 1988). These regions represent the ranges of allowed dihedral angles that are to be expected in peptides; the relative depths of the potential wells for each region will depend on the particular amino acid residues in the sequence. For non-glycine peptides, the α_L region is of relatively high energy and is only weakly populated.

Different patterns of NOE connectivities are observed for each of the major minima in ϕ,ψ space (Figure 1). Strong $d_{\alpha N}(i,i + 1)$ NOE connectivities in the absence of $d_{NN}(i,i + 1)$ NOEs indicate that backbone dihedral angles are predominantly in the β (extended chain) region of ϕ,ψ conformational space. The peptide does not, of course, adopt a unique conformation but fluctuates over an ensemble of extended-chain conformations with a somewhat restricted range of ϕ and ψ angles. The presence of $d_{NN}(i,i + 1)$ NOE connectivities is taken as an indication that the conformational ensemble includes local structures with dihedral angles in the α_R (or α_L) region of conformational space.

It is often found that sequential NOEs characteristic of both β and α backbone conformations are present for a given peptide. For the more than 50 peptides examined in our laboratory by ROESY and NOESY spectroscopy, $d_{\alpha N}(i,i + 1)$ NOE connectivities are invariably observed between every pair of amino acids in the sequence. For only a few peptides, extended sequences of $d_{NN}(i,i + 1)$ NOE connectivities are also observed. Figure 1 shows that there are two regions of ϕ,ψ space where both $d_{\alpha N}(i,i + 1)$ and $d_{NN}(i,i + 1)$ distances are shorter than 3 Å. These are the α_L-helical region and the bridging region between the β (extended-chain) region and the α_R-helix minimum. Similar observations have been made by Balaram and co-workers (Ramakrishnan et al., 1987). Observation of extended sequences of both $d_{\alpha N}(i,i + 1)$ and $d_{NN}(i,i + 1)$ NOE connectivities in the spectrum of a given peptide could thus indicate conformations in one of these regions. A more likely explanation is conformational averaging; i.e., the peptide samples both the α_R and β (extended-chain) regions of ϕ,ψ space. On the other hand, observation of both $d_{\alpha N}(i,i + 1)$ and $d_{NN}(i,i + 1)$ connectivities at a single site in a peptide is often indication of a highly preferred local structure, e.g., a type II β-turn for which the ϕ and ψ angles for residue 3 lie in the α_L region (see below).

It is important to emphasize that the observation of $d_{\alpha N}(i,i + 1)$ or $d_{NN}(i,i + 1)$ NOE connectivities provides information about the *local* dihedral angle populations at each residue. Observation of sequential NOE connectivities is of itself insufficient evidence for *folded* peptide structures. Additional information is needed, either from other NMR parameters or from other forms of spectroscopy. The best diagnostic is the observation of medium range $i,i + 2$, $i,i + 3$, or $i,i + 4$ NOEs or even longer range NOEs that immediately identify a significant population of nonrandom folded structures in the conformational ensemble. Additional information may come from unusual $^3J_{NH\alpha}$ coupling constants or evidence of hydrogen-bonding interactions from lowered values of the amide proton temperature coefficients or reduced amide proton exchange rates. In identifying folded conformations of flexible linear peptides, observation of characteristic patterns of NOE connectivities representative of secondary structure elements in proteins (Wüthrich et al., 1984) is important.

In summary, modern two-dimensional NMR experiments provide a powerful approach, when interpreted with due caution, to determination of the conformational properties of flexible peptides. We turn now to the nature of the folded conformations adopted by linear peptides in aqueous solution.

Conformations of Peptides in Water: Hydrophobic Clusters. Linear peptides have been shown by NMR methods to form intramolecular hydrophobic clusters in aqueous solution. Such a structure was first identified in a synthetic fragment of parathyroid hormone (Bundi et al., 1976, 1978). A local nonrandom structure stabilized by hydrophobic interactions between the side chains of valine and tryptophan in the sequence VQWL was identified in peptides as short as four or five residues. A similar hydrophobic cluster has also been identified involving residues Ile, Val, and Phe in the peptide KIVFKNNA corresponding to residues 26–33 of French bean plastocyanin (J. Sayre, H. J. Dyson, R. A. Lerner, and P. E. Wright, unpublished observations).

Conformations of Peptides in Water: Helical Structures. Highly detailed studies of the conformation of the ribonuclease S-peptide and C-peptide and sequences related to them have been reported (Bierzynski et al., 1982; Kim & Baldwin, 1984; Shoemaker et al., 1985, 1987a,b; Rico et al., 1984, 1986). These studies have relied heavily on CD spectroscopy and one-dimensional NMR methods. With the negative ellipticity

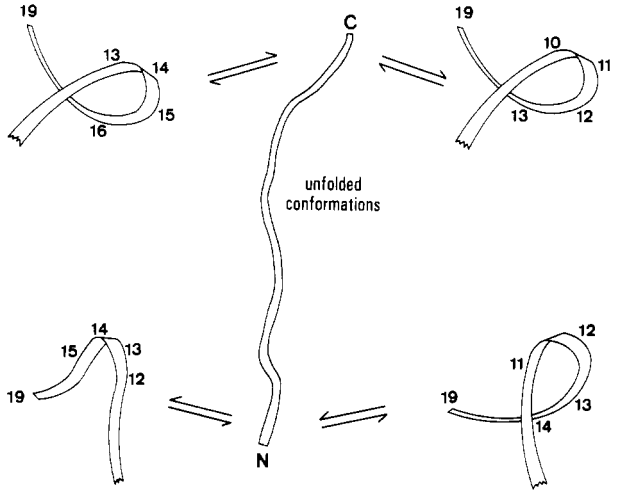

FIGURE 2: Schematic representation of the "nascent helix" observed for the myohemerythrin C-helix peptide in water solution. The structural ensemble contains a series of turn-like conformations in equilibrium with unfolded conformations (Dyson et al., 1988b).

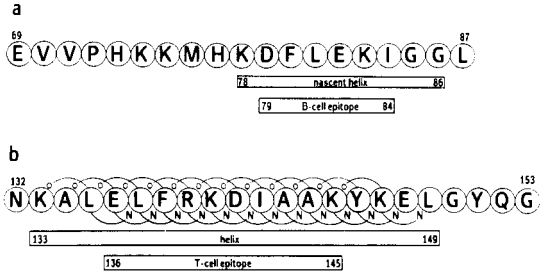

FIGURE 3: (a) The amino acid sequence of the myohemerythrin C-helix peptide showing the location of the B-cell antigenic determinant (Fieser et al., 1987) and of the nascent helix in water solution (Dyson et al., 1988b). (b) The amino acid sequence of the C-terminal cyanogen bromide cleavage fragment of sperm whale myoglobin showing the location of the helical region (J. P. Waltho, R. A. Lerner, and P. E. Wright, submitted for publication) and of the T-cell epitope (Berkower et al., 1987).

at 222 nm as a diagnostic of helix formation, it was estimated that the C-peptide (residues 1–13 of ribonuclease) contains about 30% helix at 1.7 °C, pH 5.22 (Bierzynski et al., 1982; Kim & Baldwin, 1984). Comparison of NMR chemical shifts for the C-peptide lactone with those for the S-peptide suggests the presence of a "helix-stop" signal in the S-peptide sequence (Kim & Baldwin, 1984). Subsequently, sequence variants of the C-peptide have been synthesized and have been found to contain both higher and lower populations of helix, according to the ellipticity at 222 nm (Baldwin, 1986; Shoemaker et al., 1987a).

The stability of the α-helix formed by the ribonuclease S- and C-peptides and their analogues is much higher than is predicted from Zimm and Bragg σ and s parameters derived from amino acid copolymers (Zimm & Bragg, 1959; Sueki et al., 1984). Recent experiments strongly suggest that the helix formed by the ribonuclease C-peptide is stabilized by favorable interactions between charged side chains and the helix dipole and by specific interactions between side chains. The proposed side-chain interactions include a hydrogen-bonded Glu 2–Arg 10 salt bridge and possible contributions from an interaction between the aromatic rings of Phe 8 and His 12 (Shoemaker et al., 1987a,b; Rico et al., 1986). Scheraga and co-workers (Vásquez et al., 1987) have now extended the Zimm–Bragg theory to account for the stabilizing effect of peptide charge–dipole interactions. The role of side-chain salt bridges and interactions of charged side chains with the helix dipole in stabilization of the helix have subsequently been tested in several peptides of de novo design (Marqusee & Baldwin, 1987).

A much more elementary and less stable structure than the helix formed by the ribonuclease S- and C-peptides has been observed by NMR in a synthetic immunogenic peptide corresponding to the C-helix, residues 69–87, of myohemerythrin (Dyson et al., 1988b). The conformational ensemble contains a set of turn-like structures [as shown by $d_{NN}(i,i+1)$ and $d_{\alpha N}(i,i+2)$ NOEs], distributed over the C-terminal half of the peptide and rapidly interconverting by way of unfolded states (Figure 2). These structures, termed nascent helix (Dyson et al., 1988b), are stabilized into helical structures with long-range order in water/trifluoroethanol mixtures. The transient helical turns in water solution that were identified by NMR could not be detected by CD experiments. The N-terminal half of the peptide adopts predominantly β (extended-chain) backbone conformations, even in the presence of trifluoroethanol, but is stabilized in the helix by long-range interactions in the folded protein.

The nascent helical region of the myohemerythrin C-helix peptide encompasses the epitope for a number of peptide-induced monoclonal antibodies that recognize the intact protein (Fieser et al., 1987) (Figure 3). This same region constitutes an immunodominant region for antibodies to native myohemerythrin (Getzoff et al., 1987). Another striking example of the coincidence of immunological recognition and the presence of NMR-detectable secondary structure in water solution is the recent observation that a T-cell stimulating peptide from sperm whale myoglobin forms a well-defined amphipathic helix in water solution (J. P. Waltho, R. A. Lerner, and P. E. Wright, submitted for publication). The helical structure persists under a wide variety of conditions and is characterized by an extensive network of $d_{\alpha\beta}(i,i+3)$ and $d_{\alpha N}(i,i+3)$ NOE connectivities. The T-cell epitope (Berkower et al., 1986) is contained within the helical region (Figure 3).

Several other NMR studies have been reported recently that indicate at least partial helix formation by peptides in aqueous solution. Jiménez et al. (1987) presented evidence that a peptide corresponding to residues 50–61 of ribonuclease A contains a small population of helix in water solution. Both NMR and CD measurements indicate formation of an α-helix, stabilized by intermolecular association, in a fragment of bovine growth hormone (Brems et al., 1987; Gooley & MacKenzie, 1988). Recent claims that a 45-residue peptide of adenylate kinase contains an α-helix (Fry et al., 1988) are not well substantiated since they are based solely on sequential NOE connectivities and there is serious discrepancy between the NMR results and both FTIR and CD measurements of helix content.

Conformations of Peptides in Water: Reverse Turns. Another type of regular secondary structure that has been the subject of much recent experimental and theoretical interest is the reverse turn. It is now clear that turns in short linear peptides in water solution can be sufficiently stable to be detected by NMR methods. Evidence has been presented for a detectable population of turn conformations in water solutions of a terminally blocked tripeptide derived from a proposed folding initiation site of ribonuclease (Montelione et al., 1984). Turns have also been identified in an antigenic peptide derived from herpes simplex virus glycoprotein (Williamson et al., 1986). The most extensive investigations have centered around the peptide YPYDVPDYA, corresponding to residues 98–106 of the influenza virus hemagglutinin HA1 chain and containing the immunodominant region of a longer (36-residue) immunogenic peptide. NMR studies of YPYDVPDYA revealed

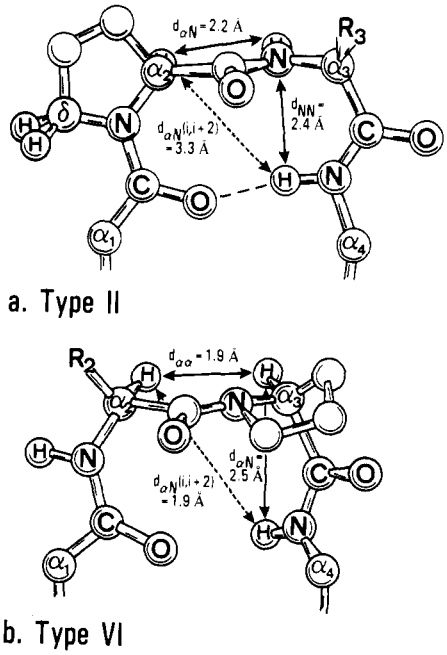

FIGURE 4: Diagrammatic representation of (a) a type II β-turn with *trans*-proline at position 2 and (b) a type VI β-turn with *cis*-proline at position 3. Sequential interproton distances are indicated with solid lines and the $d_{\alpha N}(i,i+2)$ distances by dotted lines. The dashed line in (a) represents a 1→4 hydrogen bond. The configuration of the type VI turn in (b) is consistent with NMR observations (Dyson et al., 1988a).

a conformational preference in water solution for a β-turn involving residues YPYD in the trans but not the cis isomeric form (Dyson et al., 1985). Subsequent experiments verified that the β-turn is retained with somewhat reduced population in the trans form of the peptide YPYDV and even in the four-residue peptide YPYD (Dyson et al., 1988a).

The observation of significantly populated reverse-turn conformations in such small peptides is surprising. The peptide YPYDV provides an ideal system for evaluation of the factors that stabilize β-turns in aqueous solution. A series of peptides was synthesized in which the residues at positions 1, 3, and 4 of the turn (Figure 4) were systematically varied; proline was retained at position 2 (Dyson et al., 1988a). For peptides of sequence YPXDV, where X represents all L amino acids except Trp and Pro, the highest turn population was found in YPGDV; approximately 50% of the trans isomer was estimated to be in a β-turn conformation at 5 °C. The presence of a type II β-turn was established unambiguously by observation of an extensive network of NOE connectivities, including the medium-range $d_{\alpha N}(i,i+2)$ NOE, characteristic of a β-turn (Figures 4 and 5). The low temperature coefficient of the Asp 4 amide proton resonance suggests the presence of an intramolecular hydrogen bond.

The cis–trans isomerism of the Tyr-Pro sequence in these peptides gives rise to two distinct sets of proton resonances, since isomerization is slow on the NMR time scale. In the majority of the peptides studied, the cis isomer adopts no detectable folded conformation and is predominantly extended chain. However, for several peptides containing an additional amino acid residue at the N-terminus, e.g., AYPYDV, there is a strong preference for folded conformations in both the cis and trans isomers (Dyson et al., 1988a). The predominant folded structure in the cis form of the peptides appears to be a type VI β-turn with *cis*-proline at position 3 of the turn (Figure 4b). The type VI β-turn is present in high population, as much as 70% of the cis form of the peptide (Dyson et al., 1988a).

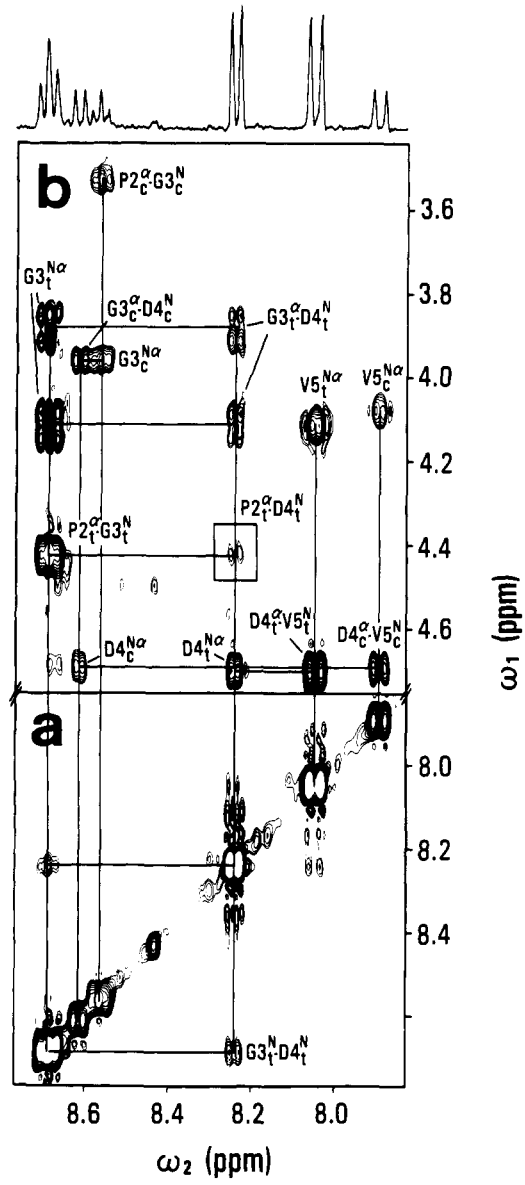

FIGURE 5: Portion of a 300-MHz ROESY spectrum of the peptide YPGDV in 90% H_2O/10% ^2H_2O at pH 4.1 and 5 °C. A spin-locking mixing period of 400 ms was used, with a radio-frequency field of 4.1 kHz. (a) Amide diagonal region; (b) $C^{\alpha}H$/NH cross-peak region (Dyson et al., 1988a).

Considerable progress has now been made in elucidating the factors that contribute to the stability of the β-turn in *trans*-YPGDV in water solution (Dyson et al., 1988a). The nature of the amino acid at position 4 influences the β-turn stability, and there is a preference for a deprotonated Asp 4 side chain. Electrostatic interactions with the charged N-terminus appear to contribute only slightly to stability. Substitution of glycine at position 3 by other residues destabilizes the turn. It appears that the nature of the residue at position 3 is the primary determinant of β-turn stability. For peptides of sequence YPXDV the temperature coefficient of the hydrogen-bonded Asp 4 amide proton resonance provides a measure of the β-turn population. The β-turn populations in water solution determined in this way were found to correlate with statistical β-turn probabilities derived from protein crystal structures (Chou & Fasman, 1978) (Figure 6). This indicates that it is frequently short-range interactions specified by the local amino acid sequence, rather than medium- to long-range interactions in the folded protein, that determine the β-turn conformation in the folded state. Such sequences

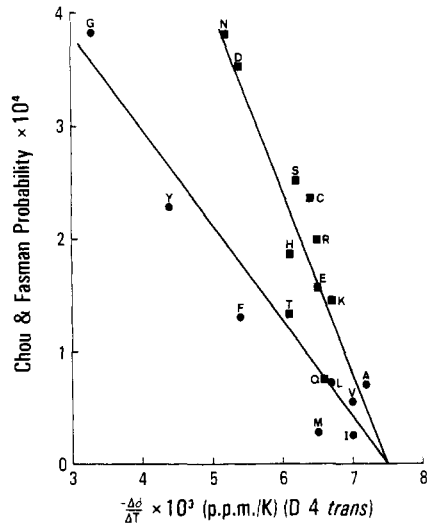

FIGURE 6: Plot of temperature coefficients for the Asp 4 amide proton resonances for the trans peptides of sequence YPXDV against turn probability calculated for each sequence by the method of Chou and Fasman (1978). Peptides for which residue X may be classified as hydrophobic (plus Gly) are represented by circular symbols; those for which residue X may be classified as hydrophilic are represented by square symbols. In each case the peptide is identified by the single-letter code for residue X. The straight lines represent lines of best fit calculated by the method of least squares (Dyson et al., 1988a).

are therefore ideal candidates for protein folding initiation sites.

Initiation of Protein Folding. A plausible first step in protein folding is the formation of local secondary structure or hydrophobic clusters in one or more regions of the polypeptide chain. The principal objection to such models has, in the past, been the lack of direct experimental evidence for local structures in unfolded proteins and in short peptide fragments of proteins. Clearly this objection is no longer a valid one. Folded structures have now been observed in many linear peptides. Although these structures are only marginally stable, they might be expected to play a fundamental role in initiation of protein folding by restricting the conformational space that must be searched and by directing subsequent folding pathways. Formation of such structures depends only on short-range interactions between amino acid residues in short, contiguous segments of the polypeptide chain. The folded structures that have been identified in peptide fragments are relatively unstable and are in rapid dynamic exchange with unfolded states. Their role in protein folding must clearly be restricted to the very early events, prior to formation of kinetically observable intermediates and much before the rate-determining step.

Throughout this paper, we deliberately refer to *initiation* of protein folding rather than use the more common term *nucleation.* Scheraga and co-workers (Montelione et al., 1984) have also advocated use of the term initiation of chain folding. Initiation of protein folding is a quite different process from a typical nucleation-dependent process such as crystallization. Nucleation of crystal formation from a supersaturated solution is rate-determining, whereas initiation or *seeding* (Baldwin, 1986) of protein folding is a rapid process that precedes the rate-determining step. This important distinction has not always been appreciated and has led to confusion (Creighton, 1987, 1988). Numerous kinetic studies of protein folding have indicated that the rate-determining step is toward the end of the folding pathway, close to the fully folded state (Creighton, 1985).

A role for local secondary structures in initiation of folding has been proposed in several folding models. The α-helix is a favorite candidate for folding initiation sites (Anfinsen, 1972; Finkelstein & Ptitsyn, 1976; Lim, 1978; Presta & Rose, 1988). This role clearly gains support from the observation that formation of relatively stable α-helices in linear peptides is not such a rare event as was originally supposed (Sueki et al., 1984) and several examples are now known. It should be noted, however, that the stability of initiation structures is not really an issue; even nascent helical structures (Dyson et al., 1988b) substantially restrict conformational space and hence can influence early folding events prior to the formation of stable secondary structures.

Reverse turns have also been suggested to play an important role in initiation of protein folding (Lewis et al., 1971; Zimmerman & Scheraga, 1977). The surprisingly high stability of β-turns in several small linear peptides suggests that turns are to be expected in polypeptide chains under folding conditions, where they may have a profound influence on protein folding pathways. In common with the α-helix, formation of reverse turns significantly restricts the conformational space available to the folding polypeptide chain. In addition, reversal of the chain brings distant parts of the polypeptide into proximity and thus promotes medium- and long-range interactions with consequent further stabilization of structure.

Of course, the experimental observation of β-turns (or α-helix) in linear peptides in water does not in itself establish a role for these structures in initiation of protein folding. Some recent studies do, however, lend support to the notion that reverse turns play an important role in protein folding. In examining temperature-sensitive mutants of the bacteriophage P22 tailspike protein, Yu and King (1988) have identified single-site mutations that interfere with polypeptide chain folding and assembly at the restrictive temperature. At the permissive temperature the protein is folded and assembled correctly. Many of the mutations that affect folding are at sites that are predicted to be solvent accessible and have high β-turn probabilities. Interestingly, in many cases the mutations do not substantially decrease the probability of β-turn formation based on statistics for folded proteins (Chou & Fasman, 1978), e.g., a Pro-Gly to Pro-Arg mutation (Yu & King, 1988). However, on the basis of the turn populations in short peptides in aqueous solution (Dyson et al., 1988a), this same substitution might be expected to greatly decrease the β-turn population in the "unfolded" state (see Figure 6). Additional insights into the likely role of β-turns in protein folding come from a series of Monte Carlo calculations of highly simplified model proteins (Kolinski et al., 1987; Skolnick et al., 1988). In simulations of the folding of a four-strand β-barrel a *unique* native structure is formed only when certain regions of the molecule are, at the very least, neutral toward bend formation. Introduction of strong bend forming regions serves to increase the stability of the native state (Skolnick et al., 1988). Simulations of the folding of a four-helix bundle lead to similar conclusions (J. Skolnick, personal communication). In both simulations the presence of some residual secondary structure in the unfolded state is required for folding to occur. Extrapolating to real proteins, these simulations suggest that regions of the amino acid sequence that have a statistical preference for formation of loops or bends are likely to be important for folding to a unique and stable globular state.

In addition to secondary structure elements, local hydrophobic clusters have been suggested to play a role in initiation of protein folding (Matheson & Scheraga, 1978; Rose & Roy, 1980; Lesk & Rose, 1981). NMR studies lend some support to these ideas. As discussed above, local clusters of amino acids, apparently stabilized by hydrophobic interactions, have

been observed in short linear peptides.

How stable must local folded structures be in order to be effective in initiating protein folding? As noted above, even transient, metastable structures in rapid equilibrium with unfolded states would have the effect of reducing the conformational space available to the nascent polypeptide chain and hence could initiate early folding events. The secondary structure elements (β-turn, nascent helix, α-helix) and hydrophobic clusters that have been identified to date appear to be, to a greater or lesser extent, relatively unstable structures in rapid exchange with unfolded states. This is probably an important criterion of primitive chain folding initiation structures since it helps to prevent the locking in of incorrectly folded regions at early stages of folding. It should be emphasized that the role of initiation structures is to direct the pathway or pathways of protein folding. The interactions important for secondary structure formation in the earliest stages of protein folding may not persist in the final folded state. The local structures that initiate folding may be retained and further stabilized or rearranged into different structures as the protein folds into its tertiary conformation. The secondary structure of a folded protein is generally stabilized by interactions between residues distant in the sequence, interactions that would not be expected to be present in the earliest intermediates or, obviously, in small peptide fragments. This means that the key amino acid residues required for folding initiation structures to form may be different from those involved in the secondary structures observed in folded proteins. As a consequence, secondary structure prediction algorithms, which are based on statistics of occurrence of the structures in folded proteins, may not always be helpful in predicting early folding intermediates. The hemagglutinin peptide provides an example of a structure that does not persist in the folded protein, since the YPYD reverse turn found in solution (Dyson et al., 1985, 1988a) is not observed in the X-ray structure (Wilson et al., 1981). On the other hand, the α-helix observed in the ribonuclease S-peptide, for example, is retained in the folded protein (Kim & Baldwin, 1984).

It is likely that the stability of some of the secondary structure elements that have been identified experimentally is unusually high, e.g., the α-helix in the ribonuclease C-peptide or the highly populated reverse turn in YPGDV. Other peptides, such as the C-helix peptide of myohemerythrin, adopt primitive structures of only marginal stability in aqueous solution. Even peptide structures of extremely low stability, which might be difficult to identify by present physical techniques (but might still induce protein-reactive antipeptide antibodies), could play a role in the earliest folding events. Harrison and Durbin (1985) have presented a compelling argument that protein sequences have evolved to favor a multiplicity of folding pathways; the protein is then much less susceptible to disruption of folding by random mutations. For similar reasons, folding via an obligatory, highly stable initiation structure would be evolutionarily undesirable. For example, initiation of folding by an essential and relatively stable Pro-Gly turn would render the protein highly susceptible to mutational disruption of folding. On the other hand, if all that is required for proper initiation of protein folding is a statistical preference for, say, a β-turn, then any one of a large number of amino acid sequences will suffice, although substitutions may well affect the stability of the final folded state. Likewise, it appears probable that folding will be initiated through local structure formation in many regions of the polypeptide chain and that, at least in the very early stages, the folding pathway will have built-in redundancy. This again will diminish the susceptibility of protein folding to random mutational errors.

The Nature of the Unfolded Polypeptide Chain. An important question is whether or not there is residual structure in the nascent ("unfolded") polypeptide chain under conditions that favor folding, i.e., in the absence of denaturants. The frequency with which structures have now been observed in peptide fragments of proteins in water strongly implies that local, transient structures involving both hydrophobic clusters and elements of secondary structure should be present in the nascent polypeptide chain, in rapid dynamic equilibrium with fully unfolded states. For certain small proteins such as ribonuclease, there is a growing body of evidence that suggests the retention of residual structure in the "unfolded" state (Takahashi et al., 1977; Chavez & Scheraga, 1980; Haas et al., 1988).

A Model for Protein Folding. The framework model of protein folding gains strong support from the observation of transient folded structures in many short linear peptide fragments of proteins and from the evidence for residual structure in unfolded proteins that is now beginning to accumulate. Numerous mechanistic models of protein folding have been proposed and have been widely reviewed (Jaenicke, 1980; Kim & Baldwin, 1982; Ghelis & Yon, 1982). The most attractive models, on the basis of present experimental data, are the diffusion–collision-type models (Karplus & Weaver, 1976, 1979; Weaver, 1982) in which local, structured microdomains form rapidly, followed by diffusion, collision, and coalescence into the folded state. These models must be distinguished carefully from the original nucleation models (Wetlaufer, 1973), in which the earliest structures are formed by a slow random search mechanism followed by rapid growth and coalescence into the native folded structure. This latter model is inconsistent with experimental evidence that the rate-determining step occurs near the end of the folding pathway (Creighton, 1985).

The following model for folding of a single domain protein, which is really an extension of the diffusion–collision model, appears to be consistent with current experimental knowlege of in vitro protein folding. The model assumes that the earliest events in folding involve formation of elements of secondary structure or local hydrophobic clusters in several regions of the polypeptide chain. These are only marginally stable, transient structures in rapid dynamic equilibrium with fully unfolded states and are presumed to be very similar to the folded structures observed in short peptide fragments. They are "stabilized" by short-range and medium-range (3–4 residue) interactions determined by the local amino acid sequence. These local elements of structure diffuse together by random Brownian processes and coalesce, resulting in growth and further stabilization. The first structures that result from coalescence are likely to be highly localized within the protein sequence and are probably still relatively unstable and in rapid exchange with unfolded states. As folding proceeds, the polypeptide chain collapses rapidly to a globular state that is likely to contain a hydrophobic core and extensive secondary and supersecondary structure and probably corresponds to the first kinetically observable folding intermediates. Rearrangement into a compact, folded, but nonnative structure follows. The final stage of protein folding, which is frequently the rate-determining step, is the conversion of the nonnative structure into the fully folded native protein. This often requires rearrangement of disulfide bridges or isomerization of *cis*- and *trans*-prolines.

With the new technologies that are now available, we can look forward to rapid progress on many aspects of the protein folding problem in the next few years. The combination of antipeptide antibody technology with two-dimensional NMR methods provides an approach for identification and structural characterization of peptide fragments of proteins that have a high propensity to fold in aqueous solution. These techniques, coupled with the ability to synthesize large numbers of peptides with specific changes in the amino acid sequence, promise to provide new insights into the earliest events in protein folding and may eventually reveal the fundamental amino acid sequence code for protein folding. Looking toward the later stages of folding, site-specific mutagenesis and two-dimensional NMR techniques hold great promise for structural characterization of metastable intermediates, and hydrogen-exchange techniques (Schmid & Baldwin, 1979; Roder & Wüthrich, 1986) should provide detailed information on the pathways involved. The greatest challenge remains the folding of the nascent polypeptide chain in vivo. Progress in this area has been reviewed recently (Tsou, 1988).

ACKNOWLEDGMENTS

We thank Drs. Jeffrey Skolnick and Jan Hermans for making available unpublished manuscripts, Drs. David Case, John Osterhout, and Jonathan Waltho for helpful discussions, and Lonnie Harvey for assistance in preparation of the manuscript.

REFERENCES

Anderson, A. G., & Hermans, J. (1988) *Proteins* (in press).
Anfinsen, C. B. (1972) *Biochem. J. 128*, 737–749.
Anfinsen, C. B. (1973) *Science (Washington, D.C.) 181*, 223–230.
Baldwin, R. L. (1986) *Trends Biochem. Sci. (Pers. Ed.) 11*, 6–9.
Berkower, I., Buckenmeyer, G. K., & Berzofsky, J. A. (1986) *J. Immunol. 136*, 2498–2503.
Bierzynski, A., Kim, P. S., & Baldwin, R. L. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 2470–2474.
Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321–346.
Bothner-By, A. A., Stephens, R. L., Lee, J., Warren, C. D., & Jeanloz, R. W. (1984) *J. Am. Chem. Soc. 106*, 811–813.
Brandts, J. F., Halvorson, H. R., & Brennan, M. (1975) *Biochemistry 14*, 4953–4963.
Brems, D. N., Plaisted, S. M., Kauffman, E. W., Lund, M., & Lehrman, S. R. (1987) *Biochemistry 26*, 7774–7778.
Brown, J. E., & Klee, W. A. (1971) *Biochemistry 10*, 470–476.
Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 299–311.
Bundi, A., Andreatta, R., Rittel, W., & Wüthrich, K. (1976) *FEBS Lett. 64*, 126–129.
Bundi, A., Andreatta, R. H., & Wüthrich, K. (1978) *Eur. J. Biochem. 91*, 201–208.
Chavez, L. G., & Scheraga, H. A. (1980) *Biochemistry 19*, 1005–1012.
Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.
Creighton, T. (1978) *Prog. Biophys. Mol. Biol. 33*, 231–297.
Creighton, T. E. (1985) *J. Phys. Chem. 89*, 2452–2459.
Creighton, T. E. (1987) *Nature (London) 326*, 547–548.
Creighton, T. E. (1988) *Science (Washington, D.C.) 240*, 267.
Dyson, H. J., Cross, K. J., Houghten, R. A., Wilson, I. A., Wright, P. E., & Lerner, R. A. (1985) *Nature (London) 318*, 480–483.
Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A., & Wright, P. E. (1988a) *J. Mol. Biol. 201*, 161–200.
Dyson, H. J., Rance, M., Houghten, R. A., Wright, P. E., & Lerner, R. A. (1988b) *J. Mol. Biol. 201*, 201–217.
Dyson, H. J., Lerner, R. A., & Wright, P. E. (1988c) *Annu. Rev. Biophys. Biophys. Chem. 17*, 305–324.
Epand, R. M., & Scheraga, H. A. (1968) *Biochemistry 7*, 2864–2872.
Ernst, R. R., Bodenhausen, G., & Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford.
Fieser, T. M., Tainer, J. A., Geysen, H. M., Houghten, R. A., & Lerner, R. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8568–8572.
Finkelstein, A. V., & Ptitsyn, O. B. (1976) *J. Mol. Biol. 103*, 15–24.
Fry, D. C., Byler, D. M., Susi, H., Brown, E. M., Kuby, S. A., & Mildvan, A. S. (1988) *Biochemistry 27*, 3588–3598.
Getzoff, E. D., Geysen, H. M., Rodda, S. J., Alexander, H., Tainer, J. A., & Lerner, R. A. (1987) *Science (Washington, D.C.) 235*, 1191–1196.
Ghelis, C., & Yon, J. (1982) *Protein Folding*, Academic, New York.
Gooley, P. R., & Mackenzie, N. E. (1988) *Biochemistry 27*, 4032–4040.
Haas, E., McWherter, C. A., & Scheraga, H. A. (1988) *Biopolymers 27*, 1–21.
Harrison, S. C., & Durbin, R. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4028–4030.
Hermans, J., & Puett, D. (1971) *Biopolymers 10*, 895–914.
Howard, J. C., Ali, A., Scheraga, H. A., & Momany, F. A. (1975) *Macromolecules 8*, 607–622.
Jaenicke, R., Ed. (1980) *Protein Folding*, Elsevier/North-Holland, Amsterdam.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.
Jiménez, M. A., Nieto, J. L., Herranz, J., Rico, M., & Santoro, J. (1987) *FEBS Lett. 221*, 320–324.
Karplus, M., & Weaver, D. L. (1976) *Nature (London) 160*, 404–406.
Karplus, M., & Weaver, D. L. (1979) *Biopolymers 18*, 1421–1437.
Kim, P. S., & Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459–489.
Kim, P. S., & Baldwin, R. L. (1984) *Nature (London) 307*, 329–334.
Kolinski, A., Skolnick, J., & Yaris, R. (1987) *Biopolymers 26*, 937–962.
Lau, W. F., & Pettitt, B. M. (1987) *Biopolymers 26*, 1817–1831.
Lerner, R. A. (1984) *Adv. Immunol. 36*, 1–44.
Lesk, A. M., & Rose, G. D. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 4304–4308.
Levinthal, C. (1968) *J. Chim. Phys. Phys.-Chim. Biol. 65*, 44–45.
Lewis, P. N., Momany, F. A., & Scheraga, H. A. (1971) *Proc. Natl. Acad. Sci. U.S.A. 68*, 2293–2297.
Lim, V. I. (1978) *FEBS Lett. 89*, 10–14.
Marqusee, S., & Baldwin, R. L. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8898–8902.
Matheson, R. R., & Scheraga, H. A. (1978) *Macromolecules 11*, 819–829.
Montelione, G. T., Arnold, E., Meinwald, Y. C., Stimson, E. R., Denton, J. B., Huang, S.-G., Clardy, J., & Scheraga, H. A. (1984) *J. Am. Chem. Soc. 106*, 7946–7958.
Niman, H. L., Houghten, R. A., Walker, L. E., Reisfeld, R. A., Wilson, I. A., Hogle, J. M., & Lerner, R. A. (1983)

Proc. Natl. Acad. Sci. U.S.A. 80, 4949–4953.
Pardi, A., Billeter, M., & Wüthrich, K. (1984) *J. Mol. Biol. 180*, 741–751.
Presta, L. G., & Rose, G. D. (1988) *Science (Washington, D.C.) 240*, 1632–1641.
Ptitsyn, O. B., & Rashin, A. A. (1975) *Biophys. Chem. 3*, 1–20.
Ptitsyn, O. B., & Finkelstein, A. V. (1980) *Q. Rev. Biophys. 13*, 339–386.
Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963) *J. Mol. Biol. 7*, 95–99.
Ramachandran, G. N., Venkatachalam, C. M., & Krimm, S. (1966) *Biophys. J. 6*, 849–872.
Ramakrishnan, C., Sukumar, M., & Balaram, P. (1987) *Biochem. Biophys. Res. Commun. 149*, 953–959.
Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151–176.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167–339.
Rico, M., Gallego, E., Santoro, J., & Bermejo, F. J. (1984) *Biochem. Biophys. Res. Commun. 123*, 757–763.
Rico, M., Santoro, J., Bermejo, F. J., Herranz, J., Meto, J. L., Gallego, E., & Jiménez, M. A. (1986) *Biopolymers 25*, 1031–1053.
Roder, H., & Wüthrich, K. (1986) *Proteins 1*, 34–42.
Rose, G. D., & Roy, S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 4643–4647.
Scheraga, H. A. (1980) *Protein Folding* (Jaenicke, R., Ed.) pp 261–288, Elsevier/North-Holland, Amsterdam.
Schmid, F. X., & Baldwin, R. L. (1979) *J. Mol. Biol. 135*, 199–215.
Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaikin, I. M., Stewart, J. M., & Baldwin, R. L. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2349–2353.
Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M., & Baldwin, R. L. (1987a) *Nature (London) 326*, 563–567.
Shoemaker, K. R., Fairman, R., Kim, P. S., York, E. J., Stewart, J. M., & Baldwin, R. L. (1987b) *Cold Spring Harbor Symp. Quant. Biol. 52*, 391–398.
Skolnick, J., Kolinski, A., & Yaris, R. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5057–5061.
States, D. J., Creighton, T. E., Dobson, C. M., & Karplus, M. (1987) *J. Mol. Biol. 195*, 731–739.
Sueki, M., Lee, S., Powers, S. P., Denton, J. B., Konishi, Y., & Scheraga, H. A. (1984) *Macromolecules 17*, 148–155.
Tainer, J. A., Getzoff, E. D., Alexander, H., Houghten, R. A., Olson, A. J., Lerner, R. A., & Hendrickson, W. A. (1984) *Nature (London) 312*, 127–133.
Tainer, J. A., Getzoff, E. D., Paterson, Y., Olson, A. J., & Lerner, R. A. (1985) *Annu. Rev. Immunol. 3*, 501–535.
Takahashi, S., Kontani, T., Yoneda, M., & Ooi, T. (1977) *J. Biochem. (Tokyo) 82*, 1127–1133.
Tanaka, S., & Scheraga, H. A. (1977) *Macromolecules 10*, 291–304.
Taniuchi, H., & Anfinsen, C. B. (1969) *J. Biol. Chem. 244*, 3864–3875.
Tsou, C. L. (1988) *Biochemistry 27*, 1809–1812.
Vásquez, M., Pincus, M. R., & Scheraga, H. A. (1987) *Biopolymers 26*, 373–386.
Weaver, D. L. (1982) *Biopolymers 21*, 1275–1300.
Weaver, D. L. (1984) *Biopolymers 23*, 675–694.
Wetlaufer, D. B. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 697–701.
Wetlaufer, D. B. (1981) *Adv. Protein Chem. 34*, 61–92.
Williamson, M. P., Hall, M. J., & Handa, B. K. (1986) *Eur. J. Biochem. 158*, 527–536.
Wilson, I. A., Wiley, D. C., & Skehel, J. J. (1981) *Nature (London) 289*, 366–373.
Wüthrich, K., & Grathwohl, C. (1974) *FEBS Lett. 43*, 337–340.
Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.
Yu, M. H., & King, J. (1988) *J. Biol. Chem. 263*, 1424–1431.
Zehfus, M. H., & Rose, G. D. (1986) *Biochemistry 25*, 5759–5765.
Zimm, B. H., & Bragg, K. J. (1959) *J. Chem. Phys. 31*, 526–535.
Zimmerman, S. S., & Scheraga, H. A. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4126–4129.
Zimmerman, S. S., Pottle, M. S., Nemêthy, G., & Scheraga, H. A. (1977) *Macromolecules 10*, 1–9.