Part I

1 Principles of Protein Stability and Design

1
Early Days of Studying the Mechanism of Protein Folding

Robert L. Baldwin

1.1
Introduction

Modern work on the mechanism of protein folding began with Chris Anfinsen. He recognized the folding problem, and he asked: How does the amino acid sequence of a protein determine its three-dimensional structure? From this basic question came various research problems, including (1) What is the mechanism of the folding process? (2) How can the three-dimensional structure be predicted from the amino acid sequence? and (3) What is the relation between the folding process in vivo and in vitro? Only the early history of the first problem will be considered here.

The basic facts needed to state the folding problem were already in place before Anfinsen's work. He knew this, well before 1973 when he received the Nobel Prize, and he was somewhat embarrassed about it. In his Nobel address [1], he says in the opening paragraph "Many others, including Anson and Mirsky in the '30s and Lumry and Eyring in the '50s, had observed and discussed the reversibility of the denaturation of proteins." Anfinsen's statement of the folding problem may be dated to 1961 [2], when his laboratory found that the amino acid sequence of ribonuclease A (RNase A) contains the information needed to make the correct four disulfide bonds of the native protein. There are eight –SH groups in the unfolded RNase A chain which could make 105 different S–S bonds. Although in 1961 the reversibility of protein denaturation was recognized by protein chemists, the knowledge that protein denaturation equals protein unfolding had been gained only a few years earlier, in a series of papers from Walter Kauzmann's laboratory, beginning with Ref. [3]. The first protein structure, that of sperm whale myoglobin (2 Å resolution), determined by John Kendrew and his coworkers [4], became available only in 1960. The myoglobin structure confirmed the proposal that proteins possess 3D structures held together by weak, noncovalent bonds, and consequently they might unfold in denaturing conditions. Wu suggested this explanation of protein denaturation as early as 1931 [5].

The other thread in Anfinsen's proposal was, of course, the recognition that protein folding is part of the coding problem. The basic dogma of molecular biology,

Protein Folding Handbook. Part I. Edited by J. Buchner and T. Kiefhaber
Copyright © 2005 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-30784-2

"DNA makes RNA makes protein," was already in place in 1958, and Anfinsen knew that the newly synthesized product of RNA translation is an inactive, unfolded polypeptide chain. How does it fold up to become active? In the 1960s and 1970s there was speculation about a code for folding, and some workers even proposed a three-letter code (i.e., three amino acid residues) [6]. Anfinsen proposed a thermodynamic hypothesis [1, 7]: the newly synthesized polypeptide chain folds up under the driving force of a free energy gradient and the protein reaches its thermodynamically most stable conformation. For protein chemists familiar with reversible denaturation, this appeared obvious: what else should drive a reversible chemical reaction besides the free energy difference between reactant and product? But for molecular biologists interested in knowing how an unfolded, newly synthesized protein is able to fold, Anfinsen's hypothesis represented a considerable leap of faith. In fact, biology does introduce subtle complexities, and more is said below about the thermodynamic hypothesis (see Section 1.3). Michael Levitt and Arieh Warshel made a farseeing proposal in 1975: they argued that an unfolded protein folds under the influence of a molecular force field, and someday the folding process will be simulated with the use of a force field [8].

Starting from Anfinsen's insights, this review examines what happened in the 1960s, 1970s, and 1980s to lay the groundwork for the modern study of how proteins fold in vitro. My review ends when the literature balloons out at the end of the 1980s with the study of new problems, such as: (1) experimental study of the transition state, (2) whether hydrophobic collapse precedes secondary structure, (3) the nature of the conformational reactions that allow hydrogen exchange in proteins, (4) using molecular dynamics to simulate the folding process, (5) the speed limit for folding, (6) helix and β-strand propensities, and (7) the mechanism of forming amyloid fibrils. This review is not comprehensive. The aim is to follow the threads that led to the prevalent view at the end of the 1980s. A few references are added after the 1980s in order to complete the picture for topics studied earlier.

1.2 Two-state Folding

An important achievement of early work on the mechanism of protein folding was the recognition that small proteins commonly show two-state equilibrium denaturation reactions and there are no observable intermediates. After the initial observation by Harrington and Schellman in 1956 that RNase A undergoes reversible thermal denaturation [9], at least five laboratories then examined the nature of the unfolding transition curve (see Ref. [10] for references). In the 1950s and 1960s RNase A was widely studied because of its small size (124 residues), purity and availability. The shape of its unfolding transition curve puzzled almost everyone, because the van't Hoff plot of ln K versus $(1/T)$ (K = equilibrium constant for unfolding) is unmistakably curved, which gives a small but observable asymmetry to the unfolding curve. The slope of the van't Hoff plot, $\Delta H/R$ (ΔH = enthalpy of unfolding, R = gas constant) should be constant if ΔH is constant. The typical explanation then was that stable intermediates are present during protein unfolding

and they explain the shape of the unfolding curve. In 1965 John Brandts [10] recognized that the correct explanation for the peculiar shape of the unfolding curve lies in the unusual thermodynamics of protein folding. Brandts argued that the thermodynamics of unfolding is dominated by hydrophobic free energy, as proposed by Walter Kauzmann [11] in 1959. Then there should be a large positive value of ΔC_p for unfolding, which explains the curvature of the van't Hoff plot [10]. ΔC_p is the difference between the heat capacities of the native and denatured forms of a protein, and a large value for ΔC_p causes a strong dependence of ΔH on temperature.

Brandts' proposal that thermal denaturation of RNase A is a two-state reaction without intermediates [10] was strongly supported, also in 1965, by Ginsburg and Carroll [12], who introduced the superposition test for intermediates and found no populated intermediates in the unfolding reaction of RNase A. In the superposition test, two or more normalized unfolding curves are superimposed and tested for coincidence. They are monitored by at least two probes that report on fundamentally different molecular properties, such as specific viscosity that reports on molecular volume and optical rotation that reports on secondary structure.

In 1966 Lumry, Biltonen and Brandts [13] introduced the calorimetric ratio test for intermediates. In this test, the ratio of the calorimetric and van't Hoff values of ΔH should equal 1 if there are no populated intermediates. In the 1970s, after the development of differential scanning calorimetry by Peter Privalov [14], the calorimetric ratio test became widely used as a criterion for two-state folding.

Charles Tanford and his laboratory undertook a wide-ranging study of whether the denaturation reactions of small proteins are truly two-state reactions. In 1968 Tanford summarized the results in a long and widely quoted review [15]. He recognized that conditions must be found in which denatured proteins are completely unfolded, without residual structure, before one can confidently determine if denaturation is a two-state reaction. His laboratory found that 6 M GdmCl (guanidinium chloride) is a denaturant that eliminates residual structure in water-soluble proteins, whereas thermally denatured proteins retain significant residual structure [15]. (Whether or not the residual structure is related to the structure of the native protein was left for future study.) A key finding was that the reversible denaturation reactions of several small proteins are indeed two-state reactions when 6 M GdmCl is the denaturant [15]. This work opened the way to later study of two basic questions: (1) Are there intermediates in protein folding reactions and, if so, how can they be detected? and (2) How can the energetics of protein folding be measured experimentally? In discussing these problems, I use the terms "unfolded" and "denatured" interchangeably, and a "denatured" protein typically has some residual structure that depends on solvent conditions and temperature.

1.3 Levinthal's Paradox

In 1968 Cyrus Levinthal released a bombshell that became known as Levinthal's paradox [16, 17]. He had begun work on prediction of the 3D structures of proteins

from their amino acid sequences. He observed that, if protein folding is truly a two-state reaction without intermediates, then the time needed to fold can be estimated from the time needed to search randomly all possible backbone conformations. Levinthal estimated that the time needed for folding by a random search is far longer than the life of the universe. A plausible conclusion from his calculation is that there must be folding intermediates and pathways [16, 17]. When Levinthal's calculation was repeated in 1992, with the addition of a small free energy bias as the driving force for folding, the time needed to search all conformations by a random search process was reduced to a few seconds [18]. Note, however, that a free energy bias in favor of the native structure is likely to produce intermediates in the folding process, although not necessarily ones that are populated.

In 1969 Levinthal pointed out [17] that, if it has a choice, a protein folds to the structure dictated by the fastest folding pathway and not to the most stable structure, in contrast to Anfinsen's thermodynamic hypothesis. His proposal was confirmed experimentally in 1996 for a protein from the serpin family, whose members form two different stable structures. The folding pathway of the serpin, plasminogen activator inhibitor 1, was found to be under kinetic control [19]. In 1998 Agard and coworkers [20] found that the stability of the folded structure of α-lytic protease appears to be under kinetic control; i.e., the denatured protein is not only kinetically unable to refold but also thermodynamically more stable than the native form [20]. This surprising deduction does not contradict Anfinsen's thermodynamic hypothesis. The enzymatically active form of this protein is formed after complete folding of a much longer polypeptide whose long Pro sequence is cleaved off after folding is complete.

1.4 The Domain as a Unit of Folding

Knowledge that polypeptide chains are synthesized starting from the N-terminus led to speculation that folding begins from the N-terminus of the chain. In 1970, Taniuchi [21] observed that the correct four S–S bonds of RNase A are not formed and the protein remains unfolded if the four C-terminal residues are deleted from the 124-residue polypeptide chain. Consequently, almost the entire polypeptide chain (maybe all of it) is required for stable folding (but see, however, Section 1.10). In 1969 Goldberg [22] found that intracistronic complementation occurs in β-galactosidase via the presence of at least two independently folding units ("globules") in each of the four identical polypeptide chains. In 1973 Wetlaufer [23] found contiguous folded regions in the X-ray structures of several proteins, in some cases apparently connected by flexible linkers. These three observations taken together gave rise to the concept that the domain ($\sim$100 amino acids) is the unit of stable folding.

Wetlaufer [23] pointed out that a contiguous folded region of the polypeptide chain is likely to arise from a structural folding nucleus. He noted that a structural nucleus might also serve as a kinetic nucleus for the folding process, with the con-

sequence that successive folding events occur rapidly after the nucleus is formed, so that folding intermediates are never populated. His suggestion was often used in the early 1970s to explain why folding intermediates could not be found. In 1981 Lesk and Rose [24] pointed out that each protein domain can typically be divided into two subdomains, each of which is also folded from a contiguous segment of polypeptide chain – although the subdomains are not separated by flexible linkers. Their observation favors a hierarchic mechanism of folding [24].

In 1974 Goldberg and coworkers [25] proposed domain swapping as the explanation for the concentration-dependent formation of large aggregates of refolding tryptophanase, formed at a critical urea concentration. Their work has been taken as a model for understanding the formation of inclusion bodies and the need for chaperones to improve the yield in many folding reactions. Also in the 1970s Jaenicke and coworkers [26] began a systematic investigation of folding coupled to subunit association in the concentration-dependent folding reactions of oligomeric proteins. These subjects are discussed elsewhere in this book.

1.5 Detection of Folding Intermediates and Initial Work on the Kinetic Mechanism of Folding

Demonstration of two-state equilibrium denaturation by Brandts [10] and by Tanford [15] made clear the difficulty of detecting any folding intermediates that might exist. In 1971, reports of complex kinetics of unfolding/refolding by Ikai and Tanford for cytochrome *c* (cyt *c*) [27] and by Tsong, Baldwin, and Elson for RNase A [28] raised hope that fast-reaction methods would succeed in detecting and characterizing kinetic folding intermediates. Complexity in the refolding kinetics of staphylococcal nuclease (SNase) had already been reported in 1970 [29] from Anfinsen's laboratory (see also Ref. [30]).

Ikai and Tanford used a stopped-flow apparatus to measure the folding kinetics of cyt *c* [27, 31] and hen lysozyme [32]. These studies laid the groundwork for systematic investigation of folding reactions that can be represented by a simple sequential model, $U \leftrightarrow I_1 \leftrightarrow I_2 \leftrightarrow \bullet\bullet\bullet N$, with one unfolded form U, one native form, N, and intervening intermediates. Off-pathway intermediates are described by a branched pathway. A mathematical framework for such studies is given in Ref. [33]. Ikai and Tanford concluded initially that they had evidence for incorrectly folded, off-pathway intermediates in the refolding of cyt *c* [27, 31], but they assumed there was only one unfolded form. Concurrent studies of RNase A refolding showed that two or more unfolded forms are present, which suggested that the same is true of other denatured proteins (see Section 1.6). Ikai and Tanford made a further important contribution by developing tests for stopped-flow artifacts that result from mixing concentrated GdmCl with buffer; this problem was a serious issue at the time. In 1973, when Ikai finished his PhD work, Charles Tanford left the field of folding mechanisms to take up the new study of membrane proteins.

Earlier, in 1968/69, Fritz Pohl began using a slow temperature-jump ("T-jump")

method, capable of observing kinetic changes only in the time range of seconds and longer, to study the kinetics of protein folding. He reported apparent two-state kinetics for the unfolding/refolding reactions of chymotrypsin [34], trypsin [35], chymotrypsinogen-A [36], and RNase A [36]. He found that the entire kinetic progress curve for unfolding or refolding could be represented by a single exponential time course and the kinetic amplitude agreed with the value expected from the equilibrium unfolding curve. He interpreted his results as showing that folding is highly cooperative, which strengthened the general view that folding intermediates would be very difficult to detect, perhaps impossible. Ikai and Tanford undertook their stopped-flow study of the unfolding/refolding kinetics of cyt *c*, despite Pohl's evidence that the kinetic folding reactions of small proteins are two-state, because Wayne Fish in Tanford's laboratory had evidence suggesting an equilibrium folding intermediate (A. Ikai, personal communication, 2003).

Tsong, Baldwin, and Elson [28, 37] undertook their fast T-jump study (dead time, 10 μs) of the unfolding kinetics of RNase A because they believed that fast kinetics might reveal intermediates in apparent two-state unfolding reactions. Pörschke and Eigen had observed [38] that short RNA helices show unfolding intermediates in the msec time range even though both the major folding and unfolding reactions follow a single exponential time course in the seconds time range. The unfolding kinetics of RNase A could be fitted to a similar type of nucleation-dependent folding model [28, 37, 40]. The model predicts kinetics in which the relative amplitude of the fast unfolding phase increases rapidly with increasing temperature, in agreement with experiment [28, 37]. A fast T-jump study of the unfolding kinetics of chymotrypsinogen-A [40] gave results like those of RNase A, suggesting that the nucleation model might be generally applicable. For chymotrypsinogen, as for RNase A, there is a fast (milliseconds) phase in unfolding, in addition to the slow unfolding reaction (seconds) studied earlier by Pohl [36]. The fast phases in the unfolding of RNase A [28, 37] and of chymotrypsinogen [40] were missed by Pohl [36] because they account for only a few percent of the total kinetic amplitude in his conditions, and because he measured only the slow unfolding reactions. The relative amplitude of the fast unfolding phase approaches 100% at temperatures near the upper end of the thermal unfolding transition of RNase A (see Section 1.6), but of course the total amplitude becomes small.

The T-jump and stopped-flow unfolding studies of RNase A were monitored by tyrosine absorbance [28, 37] and RNase A has six tyrosine residues. Consequently, the fast and slow unfolding reactions of RNase A might be detecting unfolding reactions in different parts of the molecule that occur in different time ranges. This hypothesis was tested and ruled out by studying a chemically reacted derivative of RNase A containing a single, partly buried, dinitrophenyl (DNP) group [41]. The unfolding kinetics monitored by the single DNP group are biphasic, exactly like the kinetics observed by the six tyrosyl groups. Later work by Paul Hagerman (see below) showed that the fast phase in the unfolding of RNase A arises from an unfolding intermediate that is a minor species in the ensemble of unfolded forms.

Thus, in 1973 the stage was set for kinetic studies of the mechanism of protein folding/unfolding. However, three basic questions required answers before folding

mechanisms could be analyzed. (1) Are the observed folding intermediates on-pathway? (2) Are the intermediates partly folded forms or are they completely unfolded, while the complex unfolding/refolding kinetics result from the interconversion of different denatured forms? (3) How can the structures be determined of folding intermediates whose lifetimes are as short as milliseconds?

1.6 Two Unfolded Forms of RNase A and Explanation by Proline Isomerization

In 1973 Garel and Baldwin found that unfolded RNase A contains two different major denatured forms, a fast-folding species (U_F, ~20%) and a slow-folding species (U_S, ~80%) [42] that refolds 50 times more slowly than U_F in some conditions. The two different denatured forms were discovered when refolding was monitored with a probe specific for the enzymatically active protein, namely binding of the specific inhibitor 2′-CMP. Refolding was studied initially after a pH jump (pH 2.0 → pH 5.8) at high temperatures (to obtain complete unfolding at pH 2.0) [42] and later after dilution from 6 M GdmCl [43]. When the unfolding transition is complete in the initial conditions, either at low pH and high temperature [42] or in 6 M GdmCl [43], both the fast and slow kinetic phases of refolding yield native RNase A as product. In 6 M GdmCl, the denatured protein should be completely unfolded and therefore the fast-folding and slow-folding forms correspond to two different denatured species.

In 1976 Hagerman and Baldwin [44] studied the kinetic mechanism of RNase A unfolding by using a stopped-flow apparatus and pH jumps to analyze the unfolding/refolding kinetics as a function of temperature throughout the thermal unfolding zone at pH 3.0. Their analysis is based on a four-species mechanism, $U_S \leftrightarrow U_F \leftrightarrow I \leftrightarrow N$, in which U_S and U_F are the slow-folding and fast-folding forms of the denatured protein discussed above and I is a new unfolding intermediate observed above T_m. Because I is completely unfolded, as judged either by tyrosine absorbance or by enthalpy content [44], I may be labeled instead as U_3 and the unfolding/refolding mechanism written as $U_S \leftrightarrow U_F \leftrightarrow U_3 \leftrightarrow N$. The proportions of U_S:U_F:U_3 in denatured RNase A were predicted in 1976 to be 0.78:0.20:0.02 [44]. Although U_3 was studied initially only as an unfolding intermediate [44], U_3 should be the immediate precursor of N in refolding experiments according to the sequential unfolding/refolding mechanism. This prediction was confirmed next with experiments using a sequential mixing apparatus [45]. U_3 and U_F are populated transiently by unfolding N with a first mixing step and then U_3 and U_F are allowed to refold after a second mixing step. U_3 forms N much more rapidly than U_F does [45], and in 1994 U_3 was detected in the equilibrium population of denatured RNase A species [46].

The 1976 analysis predicts correctly the equilibrium curve for thermal denaturation from the kinetic data [44], and also shows that refolding of U_F and U_S to N must occur by the sequential mechanism $U_S \leftrightarrow U_F \leftrightarrow N$ and not by the split mechanism $U_S \leftrightarrow N \leftrightarrow U_F$. The latter conclusion follows from the behavior of

the kinetic amplitudes when the relative rates of the fast and slow kinetic refolding reactions are varied by changing the temperature [44]. The issue of a sequential versus a split mechanism was tested in a different manner by Brandts and coworkers [47], who were aware of Hagerman's work (see their discussion). They introduced the interrupted unfolding (or "double-jump") experiment in which the species present at each time of unfolding are assayed by refolding measurements.

Brandts and coworkers [47] proposed a proline isomerization model as an explanation for the two different forms of RNase A. Their model includes two separate hypotheses. The first is that the fast-folding and slow-folding forms of a denatured protein are produced by slow cis–trans isomerization of proline peptide bonds after unfolding occurs. The second is that the fast-folding and slow-folding denatured species account entirely for the complex unfolding/refolding kinetics and no structural folding intermediates are populated. Thus, their unfolding/refolding mechanism for a protein with only one proline residue is $N \leftrightarrow U_F \leftrightarrow U_S$. The fast-folding species U_F has the same prolyl isomer (cis or trans) as N and the slow-folding species U_S contains the other prolyl isomer.

Nuclear magnetic resonance (NMR) studies of proline-containing peptides show that the cis:trans ratio of a prolyl peptide bond commonly lies between 30:70 and 10:90, depending on neighboring residues. Because the proline ring sometimes clashes sterically with neighboring side chains, proline peptide bonds (X-Pro) are quite different from ordinary peptide bonds, for which the % cis is only ~0.1–1%. RNase A contains two cis proline residues, as well as as two trans proline residues, and therefore denatured RNase A should have an unusually high fraction of $[U_S]$ (as observed), because the two cis residues isomerize to trans after unfolding and produce U_S species. A later NMR study of the trans → cis isomerization rate of Gly-Pro in water [48] places it in the same time range as the $U_S \leftrightarrow U_F$ reaction of RNase A.

The proline isomerization model of Brandts and coworkers was very persuasive but it proved difficult to test, particularly because one of its two hypotheses turned out to be wrong, namely that no structural intermediates are populated during the kinetics of unfolding or refolding. In 1978 Schmid and Baldwin [49] found that the $U_F \leftrightarrow U_S$ reaction in unfolded RNase A is acid-catalyzed, although very strong acid, >5 M $HClO_4$, is required for catalysis. Very strong acid is needed for the cis → trans isomerization of both prolyl peptide bonds [50] and ordinary peptide bonds, and the slow rate of the $U_F \leftrightarrow U_S$ reaction of RNase A implies that the critical bonds are prolyl peptide bonds rather than ordinary peptide bonds. The high activation enthalpy (~85 kJ mol^{-1}) expected for isomerization of prolyl peptide bonds was found for the $U_F \leftrightarrow U_S$ reaction of denatured RNase A, both in 3.3 M $HClO_4$ and in 5 M GdmCl [49]. The acid catalysis results were widely accepted as evidence that the $U_F \leftrightarrow U_S$ reaction of RNase A is proline isomerization, and the later discovery of prolyl isomerases (see below) ended any doubts. In further work, the role of proline isomerization in protein folding kinetics has been thoroughly analyzed, especially for RNase T1 [51, 52], by combining mutagenesis of specific proline residues with sequential mixing experiments and with accurate measurement and analysis of kinetic amplitudes and relaxation times.

The $U_F \leftrightarrow U_S$ reaction of unfolded RNase A is quite slow ($\sim$1000 s) at 0 °C and it is straightforward to ask whether partial folding precedes proline isomerization at 0 °C. In 1979, Cook, Schmid, and Baldwin [53] tested this issue. They found that that partial folding does precede proline isomerization and they obtained two quite surprising results. (1) The major partly folded form (I_N) has properties closely resembling those of native RNase A (thus, I_N refers to a native-like intermediate). I_N even has RNase A catalytic activity [54]! (2) Not only does proline isomerization occur within the folded structure of I_N, but also the isomerization rate is speeded up in I_N by as much as 40-fold, compared with the rate of the $U_S \leftrightarrow U_F$ reaction in denatured RNase A [53]. These experiments gave the first clear indication that partly folded, noncovalent intermediates are sometimes populated during the kinetic process of protein folding.

Because U_S species fold slowly in physiological conditions, which is likely to make them susceptible to proteolytic cleavage, prolyl isomerases seemed needed to speed up proline isomerization in vivo. Gunter Fischer and coworkers found the first prolyl isomerase and in 1985 they showed that it catalyzes the $U_S \leftrightarrow U_F$ reaction of RNase A [55]. At least three classes of prolyl isomerases are known today. The role of prolyl isomerases in folding in vivo is discussed elsewhere in this book (see Chapter 25).

1.7
Covalent Intermediates in the Coupled Processes of Disulfide Bond Formation and Folding

Disulfide bonds stabilize the folded structures of proteins that contain S–S bonds and, when they are reduced, the protein typically unfolds. This observation was the starting point of Anfinsen's work [2] when he showed that the folding process directs the formation of the four unique S–S bonds of native RNase A. Tom Creighton had the basic insight that S–S intermediates can be covalently trapped, purified by chromatography, and structurally characterized. Because formation of S–S bonds is linked to the folding process, these S–S intermediates should also be folding intermediates. Beginning in 1974, Creighton reported the isolation and general properties of both the one-disulfide [56] and two-disulfide [57] intermediates of the small protein BPTI (bovine pancreatic trypsin inhibitor, 58 residues, three S–S bonds). He later measured the equilibrium constants for forming each of the three S–S bonds in BPTI [58, 59]. For example, he found that the effective concentration of the two –SH groups that form the S–S bond between cysteine residues 5 and 55 is $\sim 10^7$ M [58]. This value is three orders of magnitude higher than the effective concentration of the two adjacent –SH groups in dithiothreitol [58], and it illustrates the rigid alignment of these two –SH groups by the folded structure of BPTI. Later, after the development of two-dimensional NMR made it possible to determine the structures of protein species that are difficult to crystallize, Creighton and coworkers determined the structures of various S–S intermediates of BPTI (see, for example, Ref. [60]). Stabilization of the BPTI structure by a given

S–S bond depends on the effective concentration of the two –SH groups before the bond is formed, and the increase in effective concentration as successive S–S bonds are formed illustrates strikingly how the cooperativity of protein folding operates [58, 59]. The same principle has been used at an early stage in the folding of BPTI to examine the interplay between S–S bond and reverse turn formation [61]. The pathway of disulfide bond formation in BPTI is complex and has been the subject of considerable discussion [62, 63].

Early work [2] from Anfinsen's laboratory showed that nonnative S–S bonds are often formed during the kinetic process in which unfolded RNase A folds and eventually makes the correct S–S bonds. Because the S–S bond is covalent, some mechanism is needed to break nonnative S–S bonds during folding and allow formation of new S–S bonds. Anfinsen and coworkers reported in 1963 that enzymes such as protein disulfide isomerase [7] have a major role in ensuring that correct S–S bonds are formed. The role of disulfide isomerases in folding in vivo is discussed elsewhere in this book (see Chapter 26).

1.8 Early Stages of Folding Detected by Antibodies and by Hydrogen Exchange

Initial studies of the folding of peptide fragments, and also of proteins made to unfold by removing a stabilizing linkage or cofactor, gave the following generalization: the tertiary structures of proteins are easily unfolded and little residual structure remains afterwards. In 1956 Harrington and Schellman [9] found that breaking the four disulfide bonds of RNase A causes general unfolding of the tertiary structure and also destroys the helical structure, which should be local structure that could in principle survive loss of the tertiary structure. In 1968 Epand and Scheraga [64] tested by circular dichroism (CD) whether peptides from helix-containing segments of myoglobin still form helices in aqueous solution. They studied two long peptides and found they have very low helix contents; they did not pursue the problem. In 1969 Taniuchi and Anfinsen [65] made a similar experiment with SNase. They cleaved the polypeptide chain between residues 126 and 127, which causes the protein to unfold. Both fragments 1–126 and 127–149 were found to lack detectable native-like structure by various physical methods, including circular dichroism. For SNase as for RNase A, circular dichroism indicates that the native protein has some helical structure and at that time (1968) the X-ray structure of myoglobin was known [4], which gives the detailed structures of its eight helices. Thus, the overall conclusion from these experiments was that the helical secondary structures of the three proteins are stable only when the tertiary structures are present. In 1971 Lewis, Momany and Scheraga [66] proposed a hierarchic mechanism of folding in which β-turns play a directing role at early stages of folding by increasing the effective concentrations of locally formed structures, such as helices, that later interact in the native structure.

Anfinsen considered it likely that proteins fold by a hierarchic mechanism [1], and he developed an antibody method for sensitively detecting any native-like

structure still present in a denatured protein [67]. His method is simple in principle. He and his coworkers took fragments 1–126 and 99–149 of SNase, which were devoid of detectable structure by physical criteria, and bound them covalently to individual sepharose columns. Antisera were prepared against both native SNase and the two polypeptide fragments, and polyclonal antibodies were purified by immunoabsorption against each homologous antigen. The antibodies developed against native SNase were tested for their ability to cross-react with the two denatured fragments. The results indicate that a weak cross-reaction occurs and a denatured fragment reacts with antibodies made against native SNase as if the denatured fragment exists in the ''native format'' a small fraction ($\sim$0.02%) of the time. These experiments are the forerunner of modern ones in which monoclonal antibodies are made against short peptides, and some of the monoclonal antibodies cross-react significantly with the native protein from which the peptide is derived. In 1975 Anfinsen and coworkers made the converse experiment [68]. They found that antibodies directed against denatured fragments of SNase are able to cross-react with native SNase. They conclude that antibodies made against denatured fragments detect unfolded SNase in equilibrium with native SNase, even though only a tiny fraction of the native protein, less than 0.01%, is found to be unfolded.

In 1979 Schmid and Baldwin [69] developed a competition method for detecting H-bonded secondary structure formed at an early stage in the refolding of denatured RNase A. Native proteins were known at that time, from Linderstrøm-Lang's development of the hydrogen exchange method [70], to contain large numbers of highly protected peptide NH protons. Shortly afterwards, in 1982, NMR hydrogen exchange experiments by Wagner and Wüthrich [71] demonstrated that, as expected, the highly protected peptide NH protons of BPTI are ones involved in H-bonded secondary structure. In this period the exchange rates of freely exchanging, unprotected peptide NH protons were already known from earlier studies of dipeptides by Englander and coworkers [72]. The peptide NH exchange rates are base-catalyzed and the rates become faster above pH 7 than the measured folding rates of the two major U_S species of denatured RNase A.

Thus, the principle of the competition experiment is straightforward [69]. Refolding of denatured, ^{3}H-labeled RNase A is performed at pH values where the rate of exchange-out of the ^{3}H label from denatured RNase A is either faster or slower, depending on pH, than the observed rate of refolding to form native RNase A. (The exchange rates of peptide NH protons in denatured RNase A can be computed from the peptide data [72].) When exchange-out is slower than folding, the folding process traps many ^{3}H-labeled protons. When exchange-out is faster than the formation of native RNase A, the observed folding rate can be used to predict the number of ^{3}H-labeled protons that should be trapped by folding. However, the observed number of ^{3}H-labeled protons trapped by folding is always much larger than the number predicted in this way. Control experiments, made at the same pH values but in the presence of modest GdmCl concentrations added to destabilize folding intermediates, show no trapped ^{3}H label. The first conclusion is that one or more folding intermediates are formed rapidly and they give protection against exchange-out of ^{3}H label. The second conclusion is that some form of early struc-

ture, probably H-bonded secondary structure, is stable before the tertiary structure is formed. In 1980, a more convenient and informative pulse-labeling version of the competition experiment was tested [73] and found superior to the competition method. Methods of resolving and assigning the proton spectra of native proteins were being developed rapidly in the early 1980s and it was evident that the secondary structures of early folding intermediates would be determined by this approach, probably within a few years. A stopped-flow apparatus could be used to trap protected peptide NH protons by means of $^{2}H-^{1}H$ exchange, and 2D ^{1}H-NMR could then be used to determine the structural locations of the protected $N^{1}H$ protons.

1.9 Molten Globule Folding Intermediates

In 1981, Oleg Ptitsyn and coworkers released a bombshell [74] that was comparable in its impact to Levinthal's paradox. They proposed that the folding intermediates everyone had been searching for were sitting under our noses in plain sight, in the form of partly folded structures formed when certain native proteins are exposed to mildly destabilizing conditions. A few proteins were known to form these curious, partly folded, structures, particularly at acid pH. Ptitsyn and coworkers proposed that the partly folded forms, or ''acid forms,'' were structurally related to authentic folding intermediates. The acid forms were supposed to differ from true folding intermediates essentially only by protonation reactions resulting from pH titration to acid pH. The acid forms were found to have surprising properties which suggested that their secondary structures are stable and native-like and their conformations are compact, even though the acid forms lack fixed tertiary structures (for reviews, see Refs [75, 76]). Until then, most workers had taken it for granted that folding intermediates should simply be ''partial replicas'' of native proteins. They should contain some unfolded segments plus some other folded segments whose tertiary and secondary structures are native-like. Later work has verified essential features of Ptitsyn's proposal, although argument about the details continues.

Ptitsyn's background was in polymer physics, and he was accustomed to analyzing problems involving the conformations of polymers. In 1973, he gave his forecast of a plausible model for the kinetic process of protein folding [77], which resembles the 1971 hierarchic mechanism of Scheraga and coworkers [66] but includes also later stages in the folding process. Ptitsyn later used the term ''framework model,'' coined in a 1982 review by Kim and Baldwin [78], to describe his model. In making his 1981 proposal [74], Ptitsyn was impressed by the resemblance between the physical properties of acid forms and the properties he had hypothesized for early folding intermediates. He must also have been impressed by Kuwajima's results for the acid form of α-LA, which revealed some striking properties of acid forms (see below).

The name ''molten globule'' was given to these acid forms by Ohgushi and Wada [79] in 1983: ''globule'' meaning compact, ''molten'' meaning no fixed tertiary

structure. Ohgushi and Wada were studying two acid forms of cyt *c*, which are converted from one form to the other by varying the salt concentration [79, 80].

In 1981 the best studied of these partly folded acid forms was bovine α-lactalbumin (α-LA), which was chosen by Ptitsyn and coworkers [74] for their initial study of an acid form. In 1976 and earlier, Kuwajima and coworkers had analyzed the pH-dependent interconversion between the acid form (or "A-state") and native α-LA [81]. They found a very unusual equilibrium folding intermediate in GdmCl-induced denaturation at neutral pH [81], which by continuity – as the pH is varied – is the same species as the acid form of α-LA. The explanation for this unusual folding intermediate is that native α-LA is a calcium metalloprotein, a property discovered only in 1980 [82]. The 1976 [81] and earlier studies of α-LA were made with the apoprotein in the absence of Ca^{2+}, and the apoprotein is much less stable than the holoprotein. When GdmCl-induced unfolding of the more stable holoprotein is studied in the presence of Ca^{2+}, no stable folding intermediate is observed [76]. The near-UV and far-UV CD spectra of the 1976 folding intermediate [81] reveal some basic properties of molten globule intermediates. The α-LA folding intermediate has no fixed tertiary structure, as judged by its near-UV CD spectrum, but its secondary structure resembles that of native α-LA, as judged by its far-UV CD spectrum. In 1990, when 2D ^{1}H-NMR was used together with ^{2}H–^{1}H exchange to determine the locations and stability of the helices in a few acid forms, the helices were found at the same locations as in the native structures. Particularly clear results were found for the helices in the acid forms of cyt *c* [83] and apomyoglobin [84].

1.10
Structures of Peptide Models for Folding Intermediates

By 1979 a paradox was evident concerning the stability of helices in folding intermediates. The experiments of Epand and Scheraga and of Taniuchi and Anfinsen indicated that the helices of myoglobin [64] and SNase [65] were unstable when the intact polypeptide chains of these proteins were cut into smaller fragments. On the other hand, the ^{3}H-labeling experiments of Schmid and Baldwin [69] indicated that an early folding intermediate – probably a H-bonded intermediate – of RNase A is stable before the tertiary structure is formed. In 1971 Brown and Klee [85] had found partial helix formation by CD at 0 °C in the "C-peptide" of RNase A. C-peptide is formed by cyanogen bromide cleavage at Met13 and contains residues 1–13, while residues 3–12 form a helix in native RNase A.

In 1978, Blum, Smallcombe and Baldwin [86] used the four His residues of RNase A as probes for structure in an NMR study in real time of the kinetics of RNase A folding at pH 2, 10 °C. These are conditions in which the folding rate is sufficiently slow to take 1D NMR spectra during folding. The carbon-bound protons of the imidazole side chains could be resolved by 1D NMR in that period, provided the peptide N^{1}H protons are first exchanged for ^{2}H. By chemical shift, His12 appears to be part of a rapidly formed, folded structure at 10 °C, although it is unfolded at 45 °C, pH 2 [86]. This result suggests that the N-terminal helix of RNase

A is partly folded at low temperatures, in agreement with the C-peptide study of Brown and Klee [85]. An ensuing study by Bierzynski et al. [87] confirmed that temperature-dependent helix formation does occur. Interestingly, C-peptide helix formation was found to be strongly pH dependent with apparent pK values, indicating that the ionized forms of His12 and either Glu9 or Glu2 are needed for helix formation [87]. Many peptides later, two specific side-chain interactions were found to contribute substantially to C-peptide helix stability [88]: an amino–aromatic interaction between Phe8 and $His12^+$, and a salt bridge between $Glu2^-$ and $Arg10^+$. Both interactions could be seen in the 1970 X-ray structure of RNase S [89], although the Phe 8•••His 12^+ interaction was not recognized as such in 1970.

In 1989, Marqusee et al. [90] found that alanine-based peptides form stable helices in water without the help of any specific interactions. Because alanine has only a $-CH_3$ side chain, this result indicates that the helix backbone itself is stable in water, although the helix has only marginal stability. Side chains longer than that of Ala detract from, rather than increasing, helix stability. In 1985 Dyson and coworkers [91] found that even reverse turns can be detected in short peptides by ^{1}H-NMR, and 10 years later Serrano and coworkers [92] observed formation of a stable β-hairpin in water. Thus, it is possible for all classes of secondary structure to be present, and to aid in directing the folding process, at very early stages of the folding process.

In 1988 a landmark experiment by Oas and Kim [93] showed that peptides are able to model more advanced folding intermediates. Oas and Kim took two peptides from BPTI: P_α, a 16-residue peptide from residues 43–58 which includes the C-terminal helix, and P_β, a 14-residue peptide from residues 20–33 which includes part of the central β-hairpin. Separately, each peptide appears structureless by physical criteria. When the two peptides are joined by forming the 30–51 S–S bond, both helical and β structures appear. NMR characterization indicates that the 3D structure of the peptide complex resembles that of BPTI [93]. The central question about folding intermediates in that period was [78]: are they formed according to the framework model (secondary structure forms before tertiary structure) or the subdomain model (secondary and tertiary structures form simultaneously in local subdomains)? Before the 1988 experiment of Oas and Kim [93], all evidence seemed to support the framework model. After their experiment, the subdomain model became the focus of much further work.

Acknowledgments

I thank Atsushi Ikai for information and Thomas Kiefhaber for discussion.

References

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

2 Anfinsen, C. B., Haber, E., Sela, M., & White, Jr., F. H. (1961). The kinetics of formation of native

ribonuclease, during oxidation of the reduced polypeptide chain. *Proc. NatL Acad. Sci. USA* **47**, 1309–1314.

3 Simpson, R. B. & Kauzmann, W. (1953). The kinetics of protein denaturation. I. The behavior of the optical rotation of ovalbumin in urea solutions. *J. Am. Chem. Soc.* **75**, 5139–5152.

4 Kendrew, J. C., Dickerson, R. E., Strandberg, B. E. et al. (1960). Structure of myoglobin. A three-dimensional Fourier synthesis at 2 Å resolution. *Nature (London)* **185**, 422–427.

5 Wu, H. (1931). Studies on denaturation of proteins. XIII. A theory of denaturation. *Chin. J. Physiol.* **5**, 321–344.

6 Kabat, E. A. & Wu, T. T. (1972). Construction of a three-dimensional model of the polypeptide backbone of the variable region of kappa immunoglobulin light chains. *Proc. Natl Acad. Sci. USA* **69**, 960–964.

7 Epstein, C. J., Goldberger, R. F., & Anfinsen, C. B. (1963). The genetic control of tertiary protein structure: studies with model systems. *Cold Spring Harbor Symp. Quant. Biol.* **28**, 439–444.

8 Levitt, M. & Warshel, A. (1975). Computer simulation of protein folding. *Nature (London)* **253**, 694–698.

9 Harrington, W. F. & Schellman, J. A. (1956). Evidence for the instability of hydrogen-bonded peptide structures in water based on studies of ribonuclease and oxidized ribonuclease. *C.R. Trav. Lab. Carlsberg, Sér. Chim.* **30**, 21–43.

10 Brandts, J. F. (1965). The nature of the complexities in the ribonuclease conformational transition and the implications regarding clathrating. *J. Am. Chem. Soc.* **87**, 2759–2760.

11 Kauzmann, W. (1959). Factors in interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1–63.

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

13 Lumry, R., Biltonen, R., & Brandts, J. F. (1966). Validity of the "two-state" hypothesis for conformational transitions of proteins. *Biopolymers* **4**, 917–944.

14 Privalov, P. L. (1979). Stability of proteins. Small globular proteins. *Adv. Protein Chem.* **33**, 167–241.

15 Tanford, C. (1968). Protein denaturation. A. Characterization of the denatured state. B. The transition from native to denatured state. *Adv. Protein Chem.* **23**, 121–282.

16 Levinthal, C. (1968). Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44–45.

17 Levinthal, C. (1969). How to fold graciously. In *Mössbauer Spectroscopy in Biological Systems* (Frauenfelder, H., Gunsalus, I. C., Tsibris, J. C. M., Debrunner, P. G., & Münck, E., eds), pp. 22–24, University of Illinois Press, Urbana.

18 Zwanzig, R., Szabo, A., & Bagchi, B. (1992). Levinthal's paradox. *Proc. Natl Acad. Sci. USA* **89**, 20–22.

19 Wang, Z., Mottonen, J., & Goldsmith, E. J. (1996). Kinetically controlled folding of the serpin plasminogen activator inhibitor 1. *Biochemistry* **35**, 16443–16448.

20 Sohl, J. L., Jaswal, S. S., & Agard, D. A. (1998). Unfolded conformations of α-lytic protease are more stable than its native state. *Nature (London)* **395**, 817–819.

21 Taniuchi, H. (1970). Formation of randomly paired disulfide bonds in des-(121–124)-ribonuclease after reduction and reoxidation. *J. Biol. Chem.* **245**, 5459–5468.

22 Goldberg, M. E. (1969). Tertiary structure of *Escherichia coli* β-D-galactosidase. *J. Mol. Biol.* **46**, 441–446.

23 Wetlaufer, D. B. (1973). Nucleation, rapid folding and globular intrachain regions in proteins. *Proc. Natl Acad. Sci. USA* **70**, 697–701.

24 Lesk, A. M. & Rose, G. D. (1981). Folding units in globular proteins. *Proc. Natl Acad. Sci. USA* **78**, 4304–4308.

25 London, J., Skrzynia, C., & Goldberg, M. E. (1974). Renaturation

of *Escherichia coli* tryptophanase after exposure to 8 M urea: evidence for the existence of nucleation centers. *Eur. J. Biochem.* **47**, 409–415.

26 Jaenicke, R. & Rudolph, R. (1980). Folding and association of oligomeric enzymes. In *Protein Folding* (Jaenicke, R., ed.), pp. 525–546, Elsevier/North Holland, Amsterdam.

27 Ikai, A. & Tanford, C. (1971). Kinetic evidence for incorrectly folded intermediate states in the refolding of denatured proteins. *Nature (London)* **230**, 100–102.

28 Tsong, T. Y., Baldwin, R. L., & Elson, E. L. (1971). The sequential unfolding of ribonuclease A: detection of a fast initial phase in the kinetics of unfolding. *Proc. Natl Acad. Sci. USA* **68**, 2712–2715.

29 Schechter, A. N., Chen, R. F., & Anfinsen, C. B. (1970). Kinetics of folding of staphylococcal nuclease. *Science* **167**, 886–887.

30 Epstein, H. F., Schechter, A. N., Chen, R. F., & Anfinsen, C. B. (1971). Folding of staphylococcal nuclease: kinetic studies of vtwo processes in acid denaturation. *J. Mol. Biol.* **60**, 499–508.

31 Ikai, A., Fish, W. W., & Tanford, C. (1973). Kinetics of unfolding and refolding of proteins. II. Results for cytochrome *c*. *J. Mol. Biol.* **73**, 165–184.

32 Tanford, C., Aune, K. C., & Ikai, A. (1973). Kinetics of unfolding and refolding of proteins. III. Results for lysozyme. *J. Mol. Biol.* **73**, 185–197.

33 Ikai, A. & Tanford, C. (1973). Kinetics of unfolding and refolding of proteins. I. Mathematical analysis. *J. Mol. Biol.* **73**, 145–163.

34 Pohl, F. M. (1968). Einfache Temperatursprung-Methode im Sekunden-bis Stundenbereich und die reversible Denaturierung von Chymotrypsin. *Eur. J. Biochem.* **4**, 373–377.

35 Pohl, F. M. (1968). Kinetics of reversible denaturation of trypsin in water and water-ethanol mixtures. *Eur. J. Biochem.* **7**, 146–152.

36 Pohl, F. M. (1969). On the kinetics of structural transition I of some pancreatic proteins. *FEBS Lett.* **3**, 60–64.

37 Tsong, T. Y., Baldwin, R. L., & Elson, E. L. (1972). Properties of the unfolding and refolding reactions of ribonuclease A. *Proc. Natl Acad. Sci. USA* **69**, 1809–1812.

38 Pörschke, D. & Eigen, M. (1971). Cooperative non-enzymic base recognition. III. Kinetics of the helix-coil transition of the oligoribouridylic-oligoriboadenylic acid system and of oligoriboadenylic acid alone at acid pH. *J. Mol. Biol.* **62**, 361–381.

39 Tsong, T. Y., Baldwin, R. L., & McPhie, P. (1972). A sequential model of nucleation-dependent protein folding: kinetic studies of ribonuclease A. *J. Mol. Biol.* **63**, 453–475.

40 Tsong, T. Y. & Baldwin, R. L. (1972). Kinetic evidence for intermediate states in the unfolding of chymotrypsinogen A. *J. Mol. Biol.* **69**, 145–148.

41 Tsong, T. Y. & Baldwin, R. L. (1972). Kinetic evidence for intermediate states in the unfolding of ribonuclease A. II. Kinetics of exposure to solvent of a specific dinitrophenyl group. *J. Mol. Biol.* **69**, 149–153.

42 Garel, J.-R. & Baldwin, R. L. (1973). Both the fast and slow refolding reactions of ribonuclease A yield native enzyme. *Proc. Natl Acad. Sci. USA* **70**, 3347–3351.

43 Garel, J.-R., Nall, B. T., & Baldwin, R. L. (1976). Guanidine-unfolded state of ribonuclease A contains both fast- and slow-refolding species. *Proc. Natl Acad. Sci. USA* **73**, 1853–1857.

44 Hagerman, P. J. & Baldwin, R. L. (1976). A quantitative treatment of the kinetics of the folding transition of ribonuclease A. *Biochemistry*, **15**, 1462–1473.

45 Hagerman, P. J., Schmid, F. X., & Baldwin, R. L. (1979). Refolding behavior of a kinetic intermediate observed in the low pH unfolding of ribonuclease A. *Biochemistry* **18**, 293–297.

46 Houry, W. A., Rothwarf, D. M., & Scheraga, H. A. (1994). A very fast phase in the refolding of disulfide-intact ribonuclease A: implications for the refolding and unfolding pathways. *Biochemistry* **33**, 2516–2530.

47 Brandts, J. F., Halvorson, H. R., & Brennan, M. (1975). Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* **14**, 4953–4963.

48 Cheng, H. N. & Bovey, F. A. (1977). Cis-trans equilibrium and kinetic studies of acetyl-L-proline and glycyl-L-proline. *Biopolymers* **16**, 1465–1472.

49 Schmid, F. X. & Baldwin, R. L. (1978). Acid catlysis of the formation of the slow-folding species of RNase A: evidence that the reaction is proline isomerization. *Proc. Natl Acad. Sci. USA* **75**, 4764–4768.

50 Steinberg, I. Z., Harrington, W. F., Berger, A., Sela, M., & Katchalski, E. (1960). The configurational changes of poly-L-proline in solution. *J. Am. Chem. Soc.* **82**, 5263–5279.

51 Kiefhaber, T. & Schmid, F. X. (1992). Kinetic coupling between protein folding and prolyl isomerization. II. Folding of ribonuclease A and ribonuclease T_1. *J. Mol. Biol.* **224**, 231–240.

52 Mayr, L. M., Odefey, C., Schutkowski, Mike, & Schmid, F. X. (1996). Kinetic analysis of the unfolding and refolding reactions of ribonuclease T1 by a stopped-flow double-mixing technique. *Biochemistry* **35**, 5550–5561.

53 Cook, K. H., Schmid, F. X., & Baldwin, R. L. (1979). Role of proline isomerization in folding of ribonuclease A at low temperatures. *Proc. Natl Acad. Sci. USA* **76**, 6157–6161.

54 Schmid, F. X. & Blaschek, H. (1981). A native-like intermediate on the ribonuclease A folding pathway. 2. Comparison of its properties to native ribonuclease A. *Eur. J. Biochem.* **114**, 111–117.

55 Fischer, G. & Bang, H. (1985). The refolding of urea-denatured ribonuclease A is catalyzed by peptidyl-prolyl cis-trans isomerase. *Biochim. Biophys. Acta* **828**, 39–42.

56 Creighton, T. E. (1974). The single-disulfide intermediates in the refolding of reduced pancreatic trypsin inhibitor. *J. Mol. Biol.* **87**, 603–624.

57 Creighton, T. E. (1974). The two-disulfide intermediates and the folding pathway of reduced pancreatic trypsin inhibitor. *J. Mol. Biol.* **95**, 167–199.

58 Creighton, T. E. (1983). An empirical approach to protein conformation stability and flexibility. *Biopolymers* **22**, 49–58.

59 Creighton, T. E. (1988). Disulfide bonds and protein stability. *BioEssays* **8**, 57–64.

60 van Mierlo, C. P. M., Darby, N. J., Neuhaus, D., & Creighton, T. E. (1991). Two-dimensional ^{1}H nuclear magnetic resonance study of the (5–55) single-disulfide folding intermediate of pancreatic trypsin inhibitor. *J. Mol. Biol.* **222**, 373–390.

61 Zdanowski, K. & Dadlez, M. (1999). Stability of the residual structure in unfolded BPTI in different conditions of temperature and solvent composition measured by disulfide kinetics and double mutant cycle analysis. *J. Mol. Biol.* **287**, 433–445.

62 Creighton, T. E. (1992). The disulfide folding pathway of BPTI. *Science* **256**, 111–112.

63 Weissman, J. S. & Kim, P. S. (1992). Response. *Science* **256**, 112–114.

64 Epand, R. M. & Scheraga, H. A. (1968). The influence of long-range interactions on the structure of myoglobin. *Biochemistry* **7**, 2864–2872.

65 Taniuchi, H. & Anfinsen, C. B. (1969). An experimental approach to the study of the folding of staphylococcal nuclease. *J. Biol. Chem.* **244**, 3864–3875.

66 Lewis, P. N., Momany, F. A., & Scheraga, H. A. (1971). Folding of polypeptide chains in proteins: a proposed mechanism for folding. *Proc. Natl Acad. Sci. USA* **68**, 2293–2297.

67 Sachs, D. H., Schechter, A. N., Eastlake, A., & Anfinsen, C. B. (1972). An immunologic approach to

the conformational equilibria of polypeptides. *Proc. Natl Acad. Sci. USA* **69**, 3790–3794.

68 Furie, B., Schechter, A. N., Sachs, D. H., & Anfinsen, C. B. (1975). An immunological approach to the conformational equilibrium of staphylococcal nuclease. *J. Mol. Biol.* **92**, 497–506.

69 Schmid, F. X. & Baldwin, R. L. (1979). Detection of an early intermediate in the folding of ribonuclease A by protection of amide protons against exchange. *J. Mol. Biol.* **135**, 199–215.

70 Linderstrøm-Lang, K. (1958). Deuterium exchange and protein structure. In *Symposium on Protein Structure* (Neuberger, A., ed.) Methuen, London.

71 Wagner, G. A. & Wüthrich, K. (1982). Amide proton exchange and surface conformation of BPTI in solution. *J. Mol. Biol.* **160**, 343–361.

72 Molday, R. S., Englander, S. W., & Kallen, R. G. (1972). Primary structure effects on peptide group hydrogen exchange. *Biochemistry* **11**, 150–158.

73 Kim, P. S. & Baldwin, R. L. (1980). Structural intermediates trapped during the folding of ribonuclease A by amide proton exchange. *Biochemistry* **19**, 6124–6129.

74 Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V. et al. (1981). α-Lalbumin: Compact state with fluctuating tertiary structure? *FEBS Lett.* **136**, 311–315.

75 Ptitsyn, O. B. (1995). Molten globule and protein folding. *Adv. Protein Chem.* **47**, 83–229.

76 Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins Struct. Funct. Genet.* **6**, 87–103.

77 Ptitsyn, O. B. (1973). The stepwise mechanism of protein self-organization. *Dokl. Nauk SSSR* **210**, 1213–1215.

78 Kim, P. S. & Baldwin, R. L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* **51**, 459–489.

79 Ohgushi, M. & Wada, A. (1983). 'Molten-globule state': a compact form of globular proteins with mobile side chains. *FEBS Lett.* **164**, 21–24.

80 Kuroda, Y., Kidokoro, S., & Wada, A. (1992). Thermodynamic characterization of cytochrome *c* at low pH. Observation of the molten globule state and of the cold denaturation process. *J. Mol. Biol.* **223**, 1139–1153.

81 Kuwajima, K., Nitta, K., Yoneyama, M., & Sugai, S. (1976). Three-state denaturation of α-lactalbumin by guanidine hydrochloride. *J. Mol. Biol.* **106**, 359–373.

82 Hiraoka, Y., Segawa, T., Kuwajima, K., Sugai, S., & Murai, N. (1980). α-Lactalbumin: a calcium metalloprotein. *Biochem. Biophys. Res. Commun.* **95**, 1098–1104.

83 Jeng, M.-F., Englander, S. W., Elöve, G., Wand, A. J., & Roder, H. (1990). Structural description of acid-denatured cytochrome *c* by hydrogen exchange and 2D NMR. *Biochemistry* **29**, 10433–10437.

84 Hughson, F. M., Wright, P. E., & Baldwin, R. L. (1990). Structural characterization of a partly folded apomyoglobin intermediate. *Science* **249**, 1544–1548.

85 Brown, J. E. & Klee, W. A. (1971). Helix-coil transition of the isolated amino terminus of ribonuclease. *Biochemistry* **10**, 470–476.

86 Blum, A. D., Smallcombe, S. H., & Baldwin, R. L. (1978). Nuclear magnetic resonance evidence for a structural intermediate at an early stage in the refolding of ribonuclease A. *J. Mol. Biol.* **118**, 305–316.

87 Bierzynski, A., Kim, P. S., & 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.

88 Shoemaker, K. R., Fairman, R., Kim, P. S., York, E. J., Stewart, J. M., & Baldwin, R. L. (1987). The C-peptide helix from ribonuclease A considered

as an autonomous folding unit. *Cold Spring Harbor Symp. Quant. Biol.* **52**, 391–398.

89 Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B., & Richards, F. M. (1970). The 3-dimensional structure of ribonuclease S: interpretation of an electron density map at a nominal resolution of 2 Å. *J. Biol. Chem.* **245**, 305–328.

90 Marqusee, S., Robbins, V. H., & Baldwin, R. L. (1989). Unusually stable helix formation in short alanine-based peptides. *Proc. Natl Acad. Sci. USA* **86**, 5286–5290.

91 Dyson, H. J., Cross, K. J., Houghten, R. A., Wilson, I. A., Wright, P. E., & Lerner, R. A. (1985). The immunodominant site of a synthetic immunogen has a conformational preference in water for a type-II reverse turn. *Nature (London)* **318**, 480–483.

92 Blanco, F. J., Rivas, G., & Serrano, L. (1995). A short linear peptide that folds into a stable *β*-hairpin in aqueous solution. *Nature Struct. Biol.* **1**, 584–590.

93 Oas, T. G. & Kim, P. S. (1988). A peptide model of a protein folding intermediate. *Nature (London)* **336**, 42–48.