Formation of Local Structures in Protein Folding

GAETANO T. MONTELIONE† and HAROLD A. SCHERAGA*,‡

Institute of Science & Technology, Biophysics Research Division, University of Michigan, 2200 Bonisteel Boulevard, Ann Arbor, Michigan 48109-2099, and Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301

Received June 30, 1988 (Revised Manuscript Received October 27, 1988)

I. Introduction

The problem of protein folding has been one of long-standing interest.[1] It has assumed additional importance in light of recent efforts, using recombinant DNA methodology, to express and fold proteins, to design compact folded protein structures de novo,[2] and to design proteins with altered thermostability.[3] Progress in protein design requires an improved understanding of the fundamental principles governing the folding process.

Much experimental and theoretical evidence has accumulated to support the view that thermodynamically stable local structures play an important role in the initial stages of protein folding.[4-9] Since a large part of this evidence has been obtained from studies of the folding/unfolding of bovine pancreatic ribonuclease A (RNase A), we use this system as an example to illustrate the role of local structures in the folding process. This protein has been the subject of extensive physical chemical studies over the last three decades.[1,10-12]

II. Principles of Protein Folding

RNase A contains four disulfide bonds. Its folding has been studied both (i) by keeping the disulfide bonds intact and (ii) by reducing them and then allowing them to re-form by oxidation. The principles described below have been derived from both types of folding experiments. The mechanisms for folding RNase A from the disulfide-intact or disulfide-reduced unfolded ensembles appear to follow the same general principles, although it is not yet clear how similar the detailed mechanisms are.

For both disulfide-intact and disulfide-reduced RNase A, the amino acid sequence contains all the information that is required for the protein to attain its biologically active conformation in aqueous solution under appropriate conditions.[13] During biosynthesis, RNase A is probably produced as a precursor protein containing an additional 25-residue leader sequence on its amino terminus.[14] Folding/unfolding of purified RNase A, however, occurs reversibly, without the presence of this N-terminal segment; in particular, no additional cellular components such as ribosomes or enzymes are needed for folding, although such components could possibly accelerate the *rate* of folding in the cell.

Thermodynamic vs Kinetic Control. An important concept that has emerged from studies of protein folding is that a distinction must be made between thermodynamic and kinetic control of the folding process. These two types of mechanisms can be distinguished by comparing the distribution of products obtained under identical folding conditions, but starting from different distributions of initial states. In the case of thermodynamic control, the final products of refolding are completely insensitive to the distribution of initial states while, if kinetic barriers are present, the different distributions of unfolded states will be reflected as different distributions of products.

There is considerable evidence that the native structure of RNase A is the thermodynamically most stable one. The first demonstration of this was provided by experiments of Anfinsen and co-workers,[13] who showed that, upon oxidation, disulfide-reduced RNase A refolds to the same biologically active structure as thermally unfolded or solvent-unfolded disulfide-intact RNase A. Further, whereas disulfide-reduced and disulfide-intact unfolded forms of RNase A exhibit different average hydrodynamic,[15] antibody-binding,[16] and nonradiative energy transfer[17] properties, these different distributions of initial states give rise to indistinguishable refolded structures. These results have been confirmed and extended in subsequent extensive studies of the oxidative refolding of RNase A.[18-27] Additional

G. T. Montelione was born in the Bronx, NY, in 1957. He completed both his undergraduate and graduate studies at Cornell University, receiving the Ph.D. degree in 1987. He is currently an Assistant Research Scientist and Damon Runyon–Walter Winchell Cancer Research Fellow in the Biophysics Division of the Institute of Science and Technology of the University of Michigan. His research interests are in the areas of NMR studies of protein structure, dynamics, and folding, the molecular basis of oncogenesis, and molecular design of polypeptides.

H. A. Scheraga was born in Brooklyn, NY, in 1921. He attended The City College of New York, where he received his B.S. degree, and went on to graduate work at Duke University, receiving the Ph.D. degree in 1946 and an ScD. degree (Hon) in 1961. Following postdoctoral work at Harvard Medical School, he joined the faculty at Cornell University, where he is now Todd Professor of Chemistry. His research interests are in the physical chemistry of proteins and other macromolecules, chemistry of blood clotting, and structure of water and dilute aqueous solutions.

† University of Michigan.
‡ Cornell University.

(1) Anfinsen, C. B.; Scheraga, H. A. *Adv. Protein Chem.* **1975**, *29*, 205.
(2) Ho, S. P.; DeGrado, W. F. *J. Am. Chem. Soc.* **1987**, *109*, 6751.
(3) Matthews, B. W. *Biochemistry* **1987**, *26*, 6885.
(4) Scheraga, H. A. In *Protein Folding*; Jaenicke, R., Ed.; Elsevier: Amsterdam, 1980; p 261.
(5) Wetlaufer, D. B. *Adv. Protein Chem.* **1981**, *34*, 61.
(6) Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1982**, *51*, 459.
(7) Gō, N. *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 183.
(8) Rose, G. D. *J. Mol. Biol.* **1979**, *134*, 447.
(9) Scheraga, H. A. *Carlsberg Res. Commun.* **1984**, *49*, 1.
(10) Scheraga, H. A.; Rupley, J. A. *Adv. Enzymol.* **1962**, *24*, 161.
(11) Richards, F. M.; Wyckoff, H. W. *Enzymes* **1971**, *4*, 647.
(12) Blackburn, P.; Moore, S. *Enzymes* **1982**, *15*, 317.
(13) White, F. H., Jr.; Anfinsen, C. B. *Ann. N.Y. Acad. Sci.* **1959**, *81*, 515.
(14) Haugen, T. H.; Heath, E. C. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 2689.
(15) Tanford, C. *Adv. Protein Chem.* **1968**, *23*, 121.
(16) Chavez, L. G., Jr.; Scheraga, H. A. *Biochemistry* **1980**, *19*, 1005.
(17) Haas, E.; McWherter, C. A.; Scheraga, H. A. *Biopolymers* **1988**, *27*, 1.
(18) Saxena, V. P.; Wetlaufer, D. B. *Biochemistry* **1970**, *9*, 5015.
(19) Hantgan, R. R.; Hammes, G. G.; Scheraga, H. A. *Biochemistry* **1974**, *13*, 3421.

0001-4842/89/0122-0070$01.50/0

evidence for thermodynamic control of the final protein structure comes from studies of the folding of disulfide-intact RNase A initiated from different solvent-unfolded distributions,[28-30] all of which lead to the same biologically active product.

While the distribution of refolded products of RNase A is thermodynamically controlled, the folding mechanisms themselves do involve kinetic control; i.e., there are multiple folding pathways that are distinguished by kinetic barriers. Guanidine-denatured, urea-denatured, or thermally denatured, disulfide-intact RNase A is an equilibrium mixture of fast- and slow-folding species.[31-39] This kinetic heterogeneity has been attributed, in part, to slow proline cis/trans peptide-bond isomerization,[36] although other kinetic barriers, such as disulfide-bond isomerization,[35] may also be involved. The ratios of fast- and slow-folding species depend on the length of time that the protein has been unfolded.[37,38] Hence, these kinetic barriers are sufficiently large (ca. 20 kcal/mol activation enthalpy) to distinguish different folding pathways of disulfide-intact RNase A. *Kinetic control of the distribution of folding pathways does not, however, imply kinetic control of the folded product(s)* because isomerization can take place in predominantly folded intermediates.[37,39] On the basis of measurements of biological activity, all of these kinetically distinguished folding pathways lead to the same thermodynamically determined native conformation.

Dominance of Short-Range Interactions. In the initial stages of folding in aqueous solution, the nonpolar side chains of the unfolded protein are exposed to water and therefore are driven to associate by short-range hydrophobic interactions.[40] The evidence described below indicates that short-range interactions play an essential role in determining protein structures.[41-44] This is especially true for initial folding intermediates that are generally less compact than fully folded structures and therefore less influenced by long-range interactions. This concept provides the basis for a procedure in which protein structures are built up from successively larger fragments in theoretical calculations of protein structure[45] and for experimental studies of the conformations of *fragments* of proteins described below.

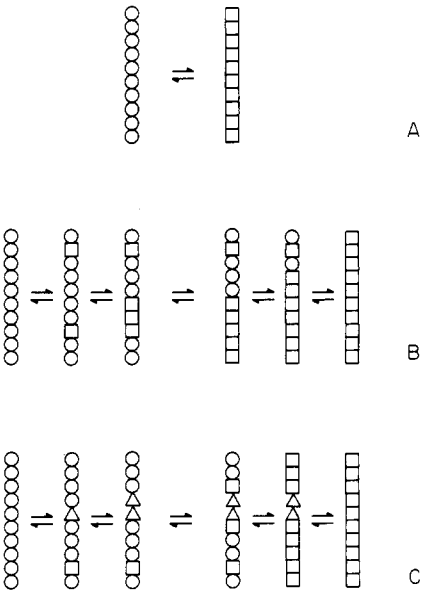

Figure 1. Schematic representations of different classes of protein folding mechanisms. These three classes are distinguished by the types of folding intermediates involved. The symbols ○, □, and △ denote locally disordered, locally ordered native, and locally ordered nonnative protein structures, respectively. (A) Simple two-state folding mechanism. Because of essentially infinite cooperativity, no metastable intermediates can form. (B) G-type multistate folding mechanism. In this case, sequence-specific short-range interactions define native-like local structures in partially folded intermediates; more than one independent local structure can form in a given polypeptide chain. (C) R-type multistate folding mechanism. Some of the local structures that form do not have native-like structures (△). Intermediates with nonnative local structurs must therefore *rearrange* to the native structure in the later stages of the folding pathway.

Support for the concept that short-range interactions dominate comes from conformational energy calculations on oligopeptides,[41-45] and from experimental studies of the thermally induced helix–coil transitions in synthetic host–guest random copolymers of amino acids.[46-48] Such helix–coil transition data, interpreted with the aid of a nearest-neighbor Ising model, lead to nucleation and growth parameters, σ and s, respectively,[49] that correlate with the frequencies of occurrence of α-helical states for residues in protein structures determined by X-ray crystallography.[46] The correlation with experimental data on helical oligopeptides[50-53] is improved by modifying these *intrinsic*

(20) Creighton, T. E. *J. Mol. Biol.* **1979**, *129*, 411.
(21) Galat, A.; Creighton, T. E.; Lord, R. C.; Blout, E. R. *Biochemistry* **1981**, *20*, 594.
(22) Konishi, Y.; Ooi, T.; Scheraga, H. A. *Biochemistry* **1981**, *20*, 3945.
(23) Konishi, Y.; Ooi, T.; Scheraga, H. A. *Biochemistry* **1982**, *21*, 4734.
(24) Konishi, Y.; Ooi, T.; Scheraga, H. A. *Biochemistry* **1982**, *21*, 4741.
(25) Konishi, Y.; Ooi, T.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5734.
(26) Scheraga, H. A.; Konishi, Y.; Ooi, T. *Adv. Biophys.* **1984**, *18*, 21.
(27) Scheraga, H. A.; Konishi, Y.; Rothwarf, D. M.; Mui, P. W. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 5740.
(28) Ahmad, F.; Bigelow, C. C. *J. Mol. Biol.* **1979**, *131*, 607.
(29) Denton, J. B.; Konishi, Y.; Scheraga, H. A. *Biochemistry* **1982**, *21*, 5155.
(30) Lynn, R. M.; Konishi, Y.; Scheraga, H. A. *Biochemistry* **1984**, *23*, 2470.
(31) Tsong, T. Y.; Baldwin, R. L.; Elson, E. L. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 1809.
(32) Garel, J.-R.; Baldwin, R. L. *J. Mol. Biol.* **1975**, *94*, 611.
(33) Schmid, F. X.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 4764.
(34) Lin, L.-N.; Brandts, J. F. *Biochemistry* **1983**, *22*, 564.
(35) Mui, P. W.; Konishi, Y.; Scheraga, H. A. *Biochemistry* **1985**, *24*, 4481.
(36) Brandts, J. F.; Halvorson, H. R.; Brennan, M. *Biochemistry* **1975**, *14*, 4953.
(37) Cook, K. H.; Schmid, F. X.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 6157.
(38) Rehage, A.; Schmid, F. X. *Biochemistry* **1982**, *21*, 1499.
(39) Schmid, F. X.; Blaschek, H. *Eur. J. Biochem.* **1981**, *114*, 111.
(40) Matheson, R. R., Jr.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 819.
(41) Kotelchuck, D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *61*, 1163.
(42) Kotelchuck, D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1969**, *62*, 14.
(43) Kotelchuck, D.; Dygert, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1969**, *63*, 615.
(44) Scheraga, H. A. *Pure Appl. Chem.* **1973**, *36*, 1.
(45) Vásquez, M.; Scheraga, H. A. *Biopolymers* **1985**, *24*, 1437.
(46) Scheraga, H. A. *Pure Appl. Chem.* **1978**, *50*, 315.
(47) Sueki, M.; Lee, S.; Powers, S. P.; Denton, J. B.; Konishi, Y.; Scheraga, H. A. *Macromolecules* **1984**, *17*, 148.
(48) Vásquez, M.; Scheraga, H. A. *Biopolymers* **1988**, *27*, 41.
(49) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526.
(50) Bierzynski, A.; Kim, P. S.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 2470.
(51) Kim, P. S.; Bierzynski, A.; Baldwin, R. L. *J. Mol. Biol.* **1982**, *162*, 187.

values of s (based on short-range interactions) by inclusion of specific medium-range interactions.[48,54,55] Similar statistical mechanical models for β-bend-to-coil and antiparallel β-sheet-to-coil transitions are currently under development[56] and will be used to interpret experimental data[57] on such transitions.

Multistate Folding Pathways. Protein folding mechanisms are distinguished by their degree of cooperativity, i.e., the degree to which the folding of one part of the protein is coupled to folding of the remainder of the protein. For mechanisms involving "infinite" cooperativity, i.e., a two-state model, plots of free energy vs folding reaction coordinate exhibit two local minima, one each for the fully unfolded and the fully folded (native) protein. In this case, both short- and long-range interactions would be involved in the folding mechanism from the beginning. On the other hand, the cooperativity of the folding process is reduced if short-range interactions dominate in the initial stages of folding since short-range interactions would define rapidly forming local structures in partially folded intermediates. The formation of these structures should be independent of the conformations adopted by the rest of the protein. In later stages of folding, these locally ordered structures coalesce under the influence of longer range interactions. The corresponding plot of free energy vs reaction coordinate would then have additional local minima corresponding to partially or incorrectly folded intermediates.

A schematic comparison of two-state and multistate folding mechanisms is shown in Figure 1. Within the class of multistate folding mechanisms, two subclasses are also represented. The subclasses, referred to as R-type and G-type, are distinguished by the presence or absence of local structures with nonnative backbone conformations. G-type and R-type pathways are analogous to (but not identical with) the growth-type and rearrangement-type folding mechanisms, described by Konishi et al.,[25] which are distinguished on the basis of the role of native and nonnative interactions *in the structures that form in the rate-limiting step.* The nonnative structure on an R-type pathway, as defined here, need not be involved in the rate-limiting step.

A two-state mechanism is actually a special case of a multistate mechanism in which the populations of the intermediate states are negligibly small. For disulfide-intact RNase A, the folding mechanism(s) are highly cooperative.[58-60] This led early workers to conclude that the folding is best described as a two-state phenomenon involving an equilibrium between native, folded RNase A and an ensemble of statistically coiled unfolded conformations with little or no formation of partially folded species.[58,60] The inadequacies of the two-state model, however, were already evident in data from early thermodynamic studies of the thermal unfolding transition.[59]

Within the thermal transition region, a multistate folding mechanism involves many species in equilibrium with one another.[61] These partially folded species, in principle, can be characterized by thermodynamic measurements because the equilibrium constants K_i between locally folded and unfolded regions are different for different segments, i, of the polypeptide chain (Figure 1). Hence, in a multistate mechanism, conformational probes located in different regions of the protein measure different folded/unfolded equilibrium constants within the transition region while, for a two-state mechanism, these equilibrium constants are identical for all regions of the protein. One useful probe to distinguish native from nonnative conformations of RNase A is the susceptibility of specific buried peptide bonds to hydrolysis by proteolytic enzymes. The fully folded protein is essentially resistant to hydrolysis by trypsin, chymotrypsin, aminopeptidase, and carboxypeptidase, while all of these proteases cleave at sequence-specific peptide bonds of fully thermally unfolded RNase A. The degree of proteolytic susceptibility of these peptide bonds can therefore be used as a crude measure of the degree of local unfolding. In the absence of partially folded intermediates, all sites should become accessible to an equal degree as the temperature is increased through the transition region. However, in an extensive series of studies,[62-66] it was found that the individual proteolytic cleavage sites are accessible to *different* degrees through the thermal transition. These observations are consistent with the presence of partially folded structures in which some peptide bonds are in environments accessible to proteases while others remain in folded environments inaccessible to proteolytic cleavage. Using this information, partially folded intermediates have been roughly characterized.[66] Further experimental evidence for these *thermodynamic intermediates* is discussed in ref 65, 67, and 68.

NMR spectroscopy is an excellent method for identifying partially folded equilibrium species in the transition region. Both the ^1H and ^{13}C NMR spectra exhibit an equilibrium between folded and unfolded conformational ensembles in the thermal transition region, with a rate of interconversion that is slow compared to the respective chemical shift time scales. With methods now available for making complete sequence-specific resonance assignments,[69] it is possible, in principle, to measure the folding/unfolding equilibrium (K_i) and rate (k_i) constants individually for each residue i of the protein through the transition region. While complete sequence-specific resonance assignments are only now becoming available for RNase A, Benz and Roberts[70] have used the four His $C^\epsilon H$ resonances to

(52) Shoemaker, K. R.; Kim, P. S.; Brems, D. N.; Marqusee, S.; York, E. J.; Chaiken, I. M.; Stewart, J. M.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 2349.
(53) Rico, M.; Santoro, J.; Bermejo, F. J.; Herranz, J.; Nieto, J. L.; Gallego, E.; Jiménez, M. A. *Biopolymers* **1986**, *25*, 1031.
(54) Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5585.
(55) Vásquez, M.; Pincus, M. R.; Scheraga, H. A. *Biopolymers* **1987**, *26*, 351.
(56) Mattice, W. L.; Scheraga, H. A. *Biopolymers* **1985**, *24*, 565.
(57) Milburn, P. J.; Meinwald, Y. C.; Takahashi, S.; Ooi, T.; Scheraga, H. A. *Int. J. Pept. Protein Res.* **1988**, *31*, 311.
(58) Harrington, W. F.; Schellman, J. A. *C. R. Trav. Lab. Carlsberg* **1956**, *30*, 21.
(59) Hermans, J., Jr.; Scheraga, H. A. *J. Am. Chem. Soc.* **1961**, *83*, 3283.
(60) Brandts, J. F.; Hunt, L. *J. Am. Chem. Soc.* **1967**, *89*, 4826.
(61) Poland, D. C.; Scheraga, H. A. *Biopolymers* **1965**, *3*, 401.
(62) Rupley, J. A.; Scheraga, H. A. *Biochemistry* **1963**, *2*, 421.
(63) Ooi, T.; Rupley, J. A.; Scheraga, H. A. *Biochemistry* **1963**, *2*, 432.
(64) Ooi, T.; Scheraga, H. A. *Biochemistry* **1964**, *3*, 641, 648.
(65) Burgess, A. W.; Weinstein, L. I.; Gabel, D.; Scheraga, H. A. *Biochemistry* **1975**, *14*, 197.
(66) Burgess, A. W.; Scheraga, H. A. *J. Theor. Biol.* **1975**, *53*, 403.
(67) Chen, M. C.; Lord, R. C. *Biochemistry* **1976**, *15*, 1889.
(68) Gilbert, W. A.; Lord, R. C.; Petsko, G. A.; Thamann, T. J. *J. Raman Spectrosc.* **1982**, *12*, 173.
(69) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; J. Wiley & Sons: New York, 1986.

compare equilibrium constants (K_i) through the transition region at four different sites. Their data suggest that these sites have slightly different values of K_i through most of the thermal transition and provide preliminary evidence for partially folded thermodynamic intermediates.

Chain-Folding Initiation Structures. Associated with the ideas of (i) the dominance of short-range interactions and (ii) a multistate folding mechanism is the concept that sequence-specific interactions define local structures that can form independently in several different regions of the polypeptide chain in the initial stages of the folding process. Such early-folding structures are an essential feature of hierarchical folding mechanisms[8,40,71-80] and have been referred to as "nuclei",[40,72,81] "locally independently nucleating continuous segments" (LINCS),[75] "kernels",[6] "seeds",[82] and "chain-folding initiation structures" (CFIS).[80] Baldwin has pointed out[6,82] that the term "nucleation" has an ambiguous meaning in the context of protein folding and should be avoided. When there are many local free-energy maxima along a folding pathway, it is no longer clear which local maximum is the analogue of a two-state "nucleation structure".[80] In any case, the structures with which we are concerned here correspond to local free-energy *minima*. For this reason, we have adopted the more descriptive term "chain-folding initiation structure" (CFIS) in referring to thermodynamically metastable local conformations that form in the initial stages of folding.[40,80] These structures limit the conformational space accessible to the protein through specific short- and medium-range interactions, thereby directing subsequent folding events, and provide stable core structures around which the rest of the polypeptide chain can fold. They do not necessarily correspond to the structures that form in the rate-limiting step, nor do they correspond to the intermediates themselves. Early-forming folding intermediates may be regarded as polypeptide chains that contain one or more CFISs.

III. Experimental Approaches to Characterizing Chain-Folding Initiation Structures

Over the last several years, we and others have attempted to develop a more detailed picture of the kinds of structures that RNase A adopts in the initial stages of folding. This work has been directed primarily along three avenues: (i) studies of polypeptide fragments, (ii) studies of disulfide-reduced derivatives, and (iii) studies of kinetic folding intermediates. If short-range interactions do indeed define local structures in the initial stages of folding, it follows that properly chosen peptide fragments of RNase A should adopt some degree of ordered structure by themselves in solution. Accordingly, a large part of our recent work has focused on characterizing the conformations adopted by polypeptide fragments of RNase A under conditions of solvent and temperature at which the disulfide-intact protein is fully folded. During the course of these studies, it became clear that further information could be obtained from structural studies of disulfide-reduced RNase A. Using these model systems, spectroscopic probes have been applied to characterize local structures under equilibrium conditions. Finally, we have recently begun to use these probes to identify locally ordered structures in kinetic folding intermediates.

Studies of Fragments of RNase A. Fragments of RNase A may be prepared by peptide synthesis,[83-85] by chemical[86,87] or enzymatic[88-91] fragmentation, or by genetic engineering. In each case, it is useful to use theoretical methods to identify potential chain-folding initiation sites before isolation and structural study. In a G-type folding mechanism, chain-folding initiation structures form with native or native-like conformations, and candidate initiation sites can be identified from an analysis of the three-dimensional structure of the native protein.[73,76-78] Likely conformations for chain-folding initiation include α-helical and β-bend conformations that can be stabilized by local interactions alone. Such structures are identified either by model building, by interactive molecular graphics, or by more sophisticated theoretical methods. Especially useful in this regard are contact maps derived from the three-dimensional structure of the native protein.[73,76-78] Conformational energy calculations can also be used to characterize the propensity for local interactions to define local structures[40,92,93] and thereby to predict the locations of chain-folding initiation sites.

The contact map for RNase, derived from its X-ray crystal structure, is shown in Figure 2. Assuming that local native-like conformations form before those that involve long-range interactions, this contact map was used to identify six partially overlapping candidate chain-folding initiation sites: residues 4–11 (site A), 25–34 (site B), 51–57 (site C), 53–79 (site D), 71–111 (site E), and 103–124 (site F).[89] Conformational free energy calculations also identify residues 106–118 (site F) as the energetically most favorable local "hydrophobic pocket" structure.[40] While CFISs that form with nonnative local interactions may be overlooked by analysis of the contact map, these predictions are helpful in designing experimental studies of formation of local native structure in fragments.

(70) Benz, F. W.; Roberts, G. C. K. *J. Mol. Biol.* **1975**, *91*, 345.
(71) Lewis, P. N.; Momany, F. A.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 2293.
(72) Wetlaufer, D. B. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 697.
(73) Tanaka, S.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 3802.
(74) Karplus, M.; Weaver, D. L. *Nature* **1976**, *260*, 404.
(75) Rose, G. D.; Winters, R. H.; Wetlaufer, D. B. *FEBS Lett.* **1976**, *63*, 10.
(76) Tanaka, S.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 1320.
(77) Tanaka, S.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 291.
(78) Némethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 6050.
(79) Gō, N.; Abe, H. *Biopolymers* **1981**, *20*, 991.
(80) Montelione, G. T.; Arnold, E.; Meinwald, Y. C.; Stimson, E. R; Denton, J. B.; Huang, S.-G.; Clardy, J.; Scheraga, H. A. *J. Am. Chem. Soc.* **1984**, *106*, 7946.
(81) Rackovsky, S.; Scheraga, H. A. *Acc. Chem. Res.* **1984**, *17*, 209.
(82) Baldwin, R. L. *Trends Biochem. Sci.* **1986**, *11*, 6.
(83) Stimson, E. R.; Montelione, G. T.; Meinwald, Y. C.; Rudolph, R. K. E.; Scheraga, H. A. *Biochemistry* **1982**, *21*, 5252.
(84) Stimson, E. R.; Meinwald, Y. C.; Montelione, G. T.; Scheraga, H. A. *Int. J. Pept. Protein Res.* **1986**, *27*, 569.
(85) Mitchinson, C.; Baldwin, R. L. *Proteins* **1986**, *1*, 23.
(86) Gross, E.; Witkop, B. *J. Biol. Chem.* **1962**, *237*, 1856.
(87) McWherter, C. A.; Thannhauser, T. W.; Fredrickson, R. A.; Zagotta, M. T.; Scheraga, H. A. *Anal. Biochem.* **1984**, *141*, 523.
(88) Richards, F. M. *C. R. Trav. Lab. Carlsberg, Sér. Chim.* **1955**, *29*, 322.
(89) Hirs, C. H. W. *J. Biol. Chem.* **1960**, *235*, 625.
(90) Thannhauser, T. W.; McWherter, C. A.; Scheraga, H. A. *Anal. Biochem.* **1985**, *149*, 322.
(91) Milburn, P. J.; Scheraga, H. A. *J. Protein Chem.* **1988**, *7*, 377.
(92) Oka, M.; Montelione, G. T.; Scheraga, H. A. *J. Am. Chem. Soc.* **1984**, *106*, 7959.
(93) Vásquez, M.; Scheraga, H. A. *J. Biomol. Struct. Dyn.* **1988**, *5*, 705, 757.

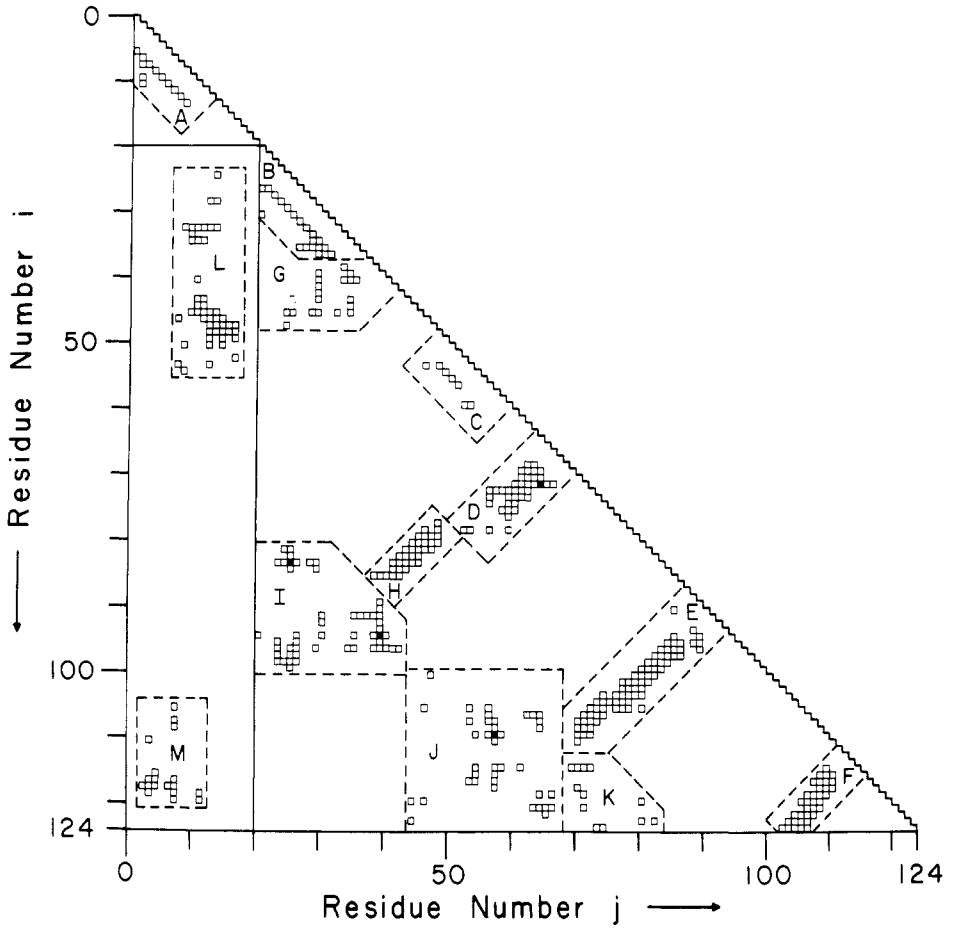

Figure 2. Contact map of RNase S.[78] Each point of the map represents the presence (square) or absence (no marking) of a contact between amino acid residues i and j. A contact is defined in ref 78. Contacts between residues are omitted from the figure whenever $|i - j| < 4$. The pairs of half-cystine residues forming the disulfide bonds are denoted by solid squares. Contact regions A–M are bounded by dashed lines. The six contact regions A–F along the diagonal represent compact structures involving short-range contacts. These are candidates for chain-folding initiation sites.

Site A. N-Terminal α-Helix (Residues 4–11). Extensive studies[16,50–53,85,94–99] of the S- and C-peptide fragments of RNase (residues 1–20 and 1–13, respectively) demonstrate that local interactions within chain-folding initiation site A are sufficient to define a native-like α-helical structure under conditions of solvent, pH, and temperature at which the complete disulfide-intact protein is fully folded. This α-helical conformation has been characterized in solution by circular dichroism (CD) spectroscopy[50,51,85] and by the identification of slowly exchanging amide proteins.[95,97] The local structure(s) present in these fragments are in rapid equilibrium with an unfolded distribution of structures. The folded and unfolded structures interconvert rapidly on the time scale of the ^{1}H NMR experiment, giving rise to resonance chemical shifts that are the weighted average of the folded and unfolded environments.[96,97] This dynamic equilibrium is both temperature and pH dependent and can be monitored by either CD or NMR spectroscopy.

Site D (Residues 53–79). While chain-folding initiation structures are expected to form in the absence of disulfide bonds, the energetics of formation of disulfide bonds and of local structures can potentially be coupled.[80,91] An excellent example of the enthalpic contribution of local structures to formation of disulfide bonds occurs in CFIS D of disulfide-reduced RNase A. CFIS D contains three cysteines, Cys^{58}, Cys^{65}, and Cys^{72}, each separated by six amino acid residues. If entropic considerations alone (i.e., loop entropy) determine the distribution of early-forming disulfide bonds, there would be an approximately equal probability of forming the 58–65 and 65–72 disulfide bonds (the probabilities are not exactly equal because of competition with more distant cysteines). In disulfide-intact RNase A, Cys^{65} is bonded to Cys^{72}, while Cys^{58} is bonded to Cys^{110}. The preference for the 65–72, rather than the 58–65, disulfide may be due to either of two effects: (i) short-range interactions within site D stabilize the 65–72 (or destabilize the 58–65) disulfide bond at an early stage of the folding process, or (ii) the 58–65 and 65–72 disulfide loops may be of equal free energy in the initial stages of folding, but as the protein folds, longer range interactions cause a rearrangement to the 65–72 and 58–110 disulfide pairings. Thermodynamic studies[91] indicate that local interactions do indeed play a role in directing the formation of disulfide bonds in a *fragment* of RNase A corresponding to residues 50–79. In this fragment, the 65–72 is favored over the 58–65 loop by $\Delta G^\circ = -1.1 \pm 0.1$ kcal/mol.[91] Hence, local interactions within CFIS D significantly stabilize the native with

(94) Klee, W. A. *Biochemistry* **1968**, *7*, 2731.
(95) Silverman, D. N.; Kotelchuck, D.; Taylor, G. T.; Scheraga, H. A. *Arch. Biochem. Biophys.* **1972**, *150*, 757.
(96) Bierzynski, A.; Baldwin, R. L. *J. Mol. Biol.* **1982**, *162*, 173.
(97) Kuwajima, K.; Baldwin, R. L. *J. Mol. Biol.* **1983**, *169*, 281, 299.
(98) Rico, M.; Nieto, J. L.; Santoro, J.; Bermejo, F. J.; Herranz, J.; Gallego, E. *FEBS Lett.* **1983**, *162*, 314.
(99) Kim, P. S.; Baldwin, R. L. *Nature* **1984**, *307*, 329.

respect to the nonnative eight-residue loop.

Sites E and F (Residues 71–111 and 103–124). We have also examined the role of local interactions in directing folding within chain-folding initiation sites E and F. In the crystal structure of RNase A,[100,101] the proposed chain-folding initiation site E forms part of an antiparallel β-sheet structure with a type III β-bend at Gly^{88}-Ser^{89} and a type VI β-bend (i.e., with a cis Pro peptide bond) at Tyr^{92}-Pro^{93}. Site F forms part of an independent antiparallel β-sheet structure with a type VI β-bend (again involving a cis Pro peptide bond) at Asn^{113}-Pro^{114}. The symmetry of the amino acid sequence around one of the β-bends of site E, Tyr^{92}-Pro^{93}-Asn^{94}, and in the vicinity of the β-bend of site F, Asn^{113}-Pro^{114}-Tyr^{115}, suggested a possible role for short-range interactions involving Tyr, Pro, and Asn which could direct the folding of RNase A. In order to examine this hypothesis, several experimental[80,83,84,102] and theoretical[92,103] studies of synthetic and proteolytic fragments of RNase A incorporating these sequences were carried out.

For the synthetic peptide Ac-Tyr-Pro-Asn-NHMe (YPN), solution NMR data indicate an equilibrium mixture of cis and trans Tyr-Pro peptide bonds[83] but do not provide evidence for formation of local structure. In the solid state, X-ray crystallography reveals an extended backbone conformation with all-trans peptide bonds.[80] Some insight into the local structural preferences of YPN was obtained by conformational free energy calculations.[92] Most low-energy conformers with trans Tyr-Pro peptide bonds have extended backbone structures like that of the crystal structure, while those with cis Tyr-Pro peptides almost exclusively are type VI β-bends. Hence, the calculations indicate that a native-like chain reversal is preferred in the subpopulation of molecules with cis peptide bonds.

Ac-Asn-Pro-Tyr-NHMe (NPY) is also a mixture of cis and trans Asn-Pro conformers in solution.[80] For this peptide, both NMR and Raman studies provide evidence for a significant fraction of hydrogen-bonded β-bend conformers within the predominant (trans) ensemble.[80] In the crystal, NPY adopts a type I β-bend conformation with two intramolecular hydrogen bonds and a trans Asn-Pro peptide bond. These experimental results are complemented by conformational free energy calculations that indicate that interactions within NPY favor β-bend conformations with *both* Asn-Pro peptide bond conformations. However, according to these calculations,[92] the location of the β-bend migrates from the Pro-Tyr to the Asn-Pro sequence as the peptide bond isomerizes from trans to cis.

Recently, we have extended these studies of local structure to larger fragments of CFIS F. Conformational energy calculations[103] on the peptide fragment corresponding to residues His^{105}-Val^{124} predict that native-like β-sheet conformations are energetically favorable for both cis and trans Asn^{113}-Pro^{114} peptide-bond conformations. Small amounts of native-like structure have also been detected in the polypeptide fragment Ser^{80}-Val^{124} by immunochemical methods.[16] Tyrosine fluorescence lifetime studies of O-T-16 (a tryptic fragment of performic acid oxidized RNase corresponding to residues 105–124) reveal the presence of at least two environments for the single tyrosine chromophore.[102] The heterogeneity of the fluorescence lifetimes in O-T-16 does *not* correlate with X-Pro peptide bond conformational heterogeneity which can be resolved by NMR spectroscopy. Instead, the fluorescence lifetime heterogeneity in O-T-16 arises from the presence of two or more conformations which interconvert rapidly on the ^{1}H NMR time scale ($\tau \ll 1$ ms) but are distinguishable on the fluorescence lifetime time scale ($\tau > 1$ ns). These data provide evidence for a population of nonrandom polypeptide conformations of O-T-16 under conditions of solvent, pH, and temperature at which the complete disulfide-intact RNase molecule is fully folded. However, it is not yet clear whether this transient local structure incorporates the β-bend conformations which have been characterized for trans and cis NPY.

Disulfide-Reduced Derivatives of RNase A. For RNase A, several studies indicate the presence of local structures in thermally, pH-, $LiClO_4$-, and guanidine-denatured RNase A. Structures that are important in the folding mechanism *under folding conditions*, however, may be unstable under solvent conditions at which the native protein structure is destabilized. Nonnative solvent conditions could also favor the formation of local structures that have nothing to do with folding in aqueous solution. Although structural studies of conformations adopted by protein fragments, discussed above, are potentially of great value in identifying local interactions *under folding conditions*, they are limited to the degree that they rely on predictive schemes for deciding which sequences to isolate, clone, or synthesize. In view of the shortcomings of these strategies for identifying CFISs, we have begun to develop systems (applicable to the entire polypeptide chain) in which long-range interactions responsible for the high cooperativity are suppressed while short- and medium-range interactions are relatively unperturbed. For disulfide-containing proteins, one such model system is the reduced protein under folding conditions.

Significant amounts of local structure, identified by CD[21,104,105] and Raman[21,106] spectroscopy, are present in disulfide-reduced RNase and in reduced cysteine S-blocked derivatives. In interpreting these results, it must be kept in mind that thiol-blocking reagents, some of which have formal charges, may perturb the conformational distributions in reduced RNase.[107] Binding of antibodies to disulfide-reduced and reduced S-carboxymethylated RNase A indicate that polypeptide segments 1–13, 31–79, and 80–124 have trace amounts of native structure at 4 °C.[16] The most convincing evidence for formation of local structure in disulfide-reduced RNase A comes from ^{1}H NMR studies.[107] Sequence-specific His C$^\epsilon$H resonance assignments are available for native RNase A from chemical modification and pH titration studies, and corresponding resonances of reduced S-sulfonated RNase could therefore

(100) Wlodawer, A.; Bott, R.; Sjölin, L. *J. Biol. Chem.* **1982**, *257*, 1325.
(101) Borkakoti, N.; Moss, D. S.; Palmer, R. A. *Acta Crystallogr. B* **1982**, *B38*, 2210.
(102) Haas, E.; Montelione, G. T.; McWherter, C. A.; Scheraga, H. A. *Biochemistry* **1987**, *26*, 1672.
(103) Pincus, M. R.; Gerewitz, F.; Wako, H.; Scheraga, H. A. *J. Protein Chem.* **1983**, *2*, 131.
(104) Schaffer, S. W. *Int. J. Pept. Protein Res.* **1975**, *7*, 179.
(105) Takahashi, S.; Kontani, T.; Yoneda, M.; Ooi, T. *J. Biochem.* **1977**, *82*, 1127.
(106) Lord, R. C.; Relyea, N. M. *J. Raman Spectrosc.* **1981**, *11*, 315.
(107) Swadesh, J. K.; Montelione, G. T.; Thannhauser, T. W.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 4606.

be assigned by partial hydrogen–deuterium exchange methods.[107] With these four His resonances as sequence-specific probes of local structure, NMR studies[107] provide evidence for a temperature-dependent equilibrium between ordered and disordered local conformations involving His[12]. By comparison with results for S-peptide and C-peptide,[50,51,53,96-99] it is likely that this locally ordered conformation corresponds to an α-helical structure.

Kinetic Folding Studies. It is important to recognize that local structures identified in equilibrium structural studies may not play an active role in the folding process. The studies described above provide detailed structural information about the kinds of structures that are defined by short-range interactions under folding conditions. However, as has been pointed out elsewhere,[25] the dominant folding mechanism(s) for a particular protein coresond to the isothermal folding pathway(s) of lowest free energy of activation. This implies that, while the folding intermediates that are detected in kinetic measurements provide correct information about the folding mechanism(s), conformers detected in equilibrium measurements may or may not be important in the predominant folding pathway(s).

The evidence from NMR spectroscopy described above supports the view that the N-terminal α-helix (proposed chain-folding initiation site A) can form independently in disulfide-reduced RNase A. The lifetime of the ordered structure in $8SSO_3$-RNase, $\tau \ll 1$ ms, is much shorter than the fastest phase of the folding of the disulfide-intact protein, with a characteristic time of 20–400 ms.[31,34] *Therefore, the degree of local structure involving His[12] will be determined by a rapidly established equilibrium due to short- and medium-range interactions, prior to subsequent folding events.* Both stopped-flow[108] and continuous recycled flow (CRF)[109] ^{1}H-NMR studies demonstrate the rapid formation of local structure involving His[12] in slow-folding kinetic intermediates of disulfide-intact RNase. The temperature dependence of the His[12] C$^\epsilon$H resonance in this intermediate[109] is similar to that of $8SSO_3$-RNase.[107] While the intermediate(s) detected by continuous recycled flow (CRF) ^{1}H NMR is predominantly disordered,[109] the His[12] residue is involved in a dynamic equilibrium between locally ordered and disordered environments similar to those of $8SSO_3$-RNase A. Although several studies[30,110,111] indicate that the rate of folding of the S-peptide α-helix of RNase A is limited by the folding of other parts of the protein, these kinetic data do not exclude very rapid formation of a small equilibrium population of local helix on the submillisecond time scale, which is indicated by the NMR data.[107,109]

IV. Summary

The principles outlined in section II (thermodynamic vs kinetic control, dominance of short-range interactions, multistate folding pathways) provide the basis for the concept of chain-folding initiation structures. Experimental studies of fragments and of disulfide-reduced derivatives of RNase A support the view that short-range interactions are sufficient to define local structures in several *independent* segments of the polypeptide chain. These include fragments corresponding to proposed chain-folding initiation sites A, D, and F. Conformational free energy calculations indicate that local interactions within site E favor the formation of a β-bend with a cis Tyr[92]-Pro[93] peptide bond. The local structure present in fragments of site A, which appears to be an α-helix, is also observed in kinetic folding intermediates. The available data indicate that RNase A utilizes a *multiple initiation site mechanism*[4,80] for folding, involving the formation of several local structures along the polypeptide chain in early-forming intermediates. This conclusion is consistent with a hierarchical folding process in which successively longer range interactions come into play as the protein folds.

Note Added in Proof. A recent NMR study[112] provides additional evidence for local structure in sites D, E, and F in an early-forming kinetic intermediate during the folding of RNase A.

We thank M. Alder, K. H. Altmann, J. M. Beals, M. E. Denton, M. A. Eastman, G. Némethy, F. Ni, D. M. Rothwarf, S. Talluri, T. W. Thannhauser, and M. P. Weiner for helpful discussions.

Registry No. RNase, 9001-99-4.

(108) Blum, A. D.; Smallcombe, S. H.; Baldwin, R. L. *J. Mol. Biol.* **1978**, *118*, 305.
(109) Adler, M.; Scheraga, H. A. *Biochemistry* **1988**, *27*, 2471.
(110) Labhardt, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 7674.
(111) Lin, S. H.; Konishi, Y.; Nall, B. T.; Scheraga, H. A. *Biochemistry* **1985**, *24*, 2680.
(112) Udgaonkar, J. B.; Baldwin, R. L. *Nature (London)* **1988**, *335*, 694.