Current Topics

Proline Cis−Trans Isomerization and Protein Folding†

William J. Wedemeyer, Ervin Welker, and Harold A. Scheraga*

Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301

Received September 5, 2002

ABSTRACT: Proline cis−trans isomerization plays a key role in the rate-determining steps of protein folding. The energetic origin of this isomerization process is summarized, and the folding and unfolding of disulfide-intact bovine pancreatic ribonuclease A is used as an example to illustrate the kinetics and structural features of conformational changes from the heterogeneous unfolded state (consisting of cis and trans isomers of X-Pro peptide groups) to the native structure in which only one set of proline isomers is present.

Cis−trans isomerization in X-Pro peptide groups is one of the rate-determining steps in protein folding (*1*). This review is concerned with the effect of such isomerism on this folding process. X-Pro dipeptides exist essentially completely in either the trans or cis forms (Figure 1) in native proteins but as an equilibrium mixture of these two forms in unfolded proteins (*2*). Conformational energy calculations provide an understanding for the overwhelming dominance of the trans form for 19 of the naturally occurring amino acids and for the population of both cis and trans forms in X-Pro peptide groups (*3*).

The relative instability of the cis form compared to the trans form in non-proline dipeptides arises from interactions of the C_1^α and H_1^α atoms with the C_2^α and H_2^α atoms (Figure 1), from stronger favorable electrostatic interactions between O_i and C'_{i+1} in the trans form, and from conformational entropy (*3*). The instability of the cis form of X-Pro peptide groups, on the other hand, is reduced because (i) unfavorable interactions between atoms attached directly to the peptide unit are present in both the cis and trans forms, primarily the steric similarity between the C^α and C^δ atoms bonded to the amide nitrogen of proline, (ii) there is a change in the $O_i\cdots C'_{i+1}$ electrostatic interactions, and (iii) there is a smaller entropy loss in the trans → cis conversion in the X-Pro peptide groups than in non-prolyl peptide groups (*3*).

Since the cis and trans forms of X-Pro peptide groups are almost isoenergetic, with the trans form being slightly more favorable (*3*), small peptides exhibit a mixture of cis and trans forms (*2*), and (because of the electrostatic component of the cis−trans transformation energy) the cis form increases in population in going from a nonpolar to a polar solvent (*2*, *3*). On the other hand, either the completely cis or the completely trans form of X-Pro peptide groups is observed in native proteins because the almost-equal energy between the cis and trans forms is overwhelmed by favorable interactions with neighboring groups in native proteins. Both experimental (*4*) and theoretical (*5*, *6*) observations indicate that there is a high energy barrier (∼20 kcal/mol), limiting the rate of the interconversion between the two forms because of the partial double-bond character of the peptide bond (*7*); hence, the kinetics of cis−trans isomerization are modeled well by a simple two-state reaction, with ω (the peptide bond dihedral angle) as a reaction coordinate. The rate of interconversion is also solvent-dependent, being higher in nonpolar solvents (*6*, *8*).

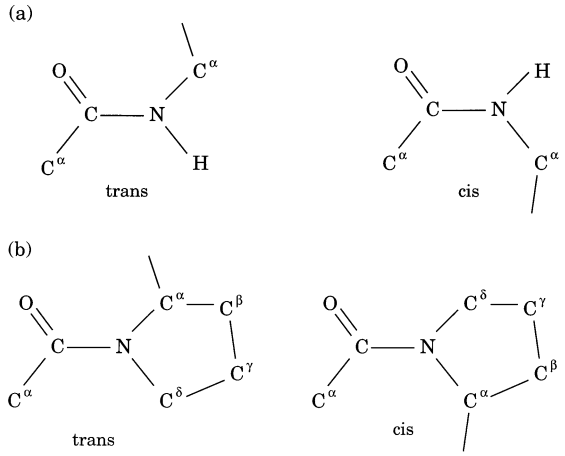

FIGURE 1: Non-proline (a) and proline (b) peptide fragments in trans and cis conformations.

The properties of X-Pro peptide groups appear not only in oligopeptides and globular proteins but also in homopolymers of proline which exist in two extreme forms. Polyproline I is all cis in a right-handed helical conformation, and polyproline II is all trans in a left-handed helical conformation (*4*). The two forms can easily interconvert by a change of solvent (*4*, *9*). For example, conformational energy calculations (*10*) account for the experimental transition

† This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health (Grant GM-24893). Support was also received from the National Foundation for Cancer Research.

* To whom correspondence should be addressed. Telephone: (607) 255-4034. Fax: (607) 254-4700. E-mail: has5@cornell.edu.

10.1021/bi020574b CCC: $22.00
Published on Web 11/05/2002

curves from polyproline II in benzyl alcohol to polyproline I in *n*-butyl alcohol in terms of the slight difference in the binding free energies of both alcohols to the carbonyl groups of form II because of the more open and extended conformation of form II.

The cis−trans interconversion of X-Pro peptide groups can be catalyzed by disruption of the partial double-bond character of the peptide bond. Thus, for example, strong acid increases the interconversion rate by protonation of the carbonyl oxygen (*4*). Enzymes such as peptidylprolyl isomerases (PPIases) likewise catalyze cis−trans isomerization by mechanisms that apparently involve disruption of the partial double-bond character of the peptide bond (*11*, *12*).

The structural and kinetic clarity of proline isomerization can be exploited in two ways. First, nonnative isomers of X-Pro peptide groups can be treated as transient site-directed mutants. At ordinary temperatures (0−35 °C), proline isomerization in general is much slower than conformational folding. When this kinetic decoupling condition holds, the isomeric state of the protein is essentially fixed on the time scale of folding, and the conformational folding of distinct isomeric species can be measured. The effects of these localized disruptions (the non-native proline isomers) on the folding kinetics indicate the role of the interactions within nearby chain segments during conformational folding.

Second, cis−trans isomerization of an X-Pro peptide group can be used to probe its local environment. Changes in the cis−trans *equilibrium* ratio may signal the development of local structure in the protein backbone. Such structure may be missed by less localized probes such as circular dichroism or tyrosine absorbance, particularly in a partially folded intermediate in which only a small fraction of the protein had assumed a stable structure. Similarly, the *kinetics* of isomerization and its susceptibility to catalysts such as strong acid or PPIases may also report on the local environment of the isomerizing proline, e.g., its degree of burial.

In this review, we focus on the influence of such properties of X-Pro peptide groups on the conformational changes involved in protein folding. We will draw primarily on the extensive experience gained from experimental studies of the folding of disulfide-intact bovine pancreatic ribonuclease (RNase A). The first section illustrates the study of conformational folding through proline isomerization in RNase A. The second section presents an overview of analogous studies carried out with other proteins.

Proline Isomerization and Conformational Folding in RNase A

Ribonuclease A is a good model protein for studying the synergism between proline isomerization and conformational folding (*13*, *14*). RNase A has four prolines (Pro93, Pro114, Pro42, and Pro117), of which the first two are cis in the native state and the latter two are trans (*15*). When the protein unfolds, these four prolines isomerize to form a heterogeneous mixture of 16 (2^4) cis−trans isomeric species. The goal has been to determine how each of these 16 species folds conformationally (if at all), and to correlate their folding behavior with the structural disruption caused by the nonnative isomers.

Unfolding of disulfide-intact RNase A is achieved by denaturation with guanidine hydrochloride (GdnHCl). Subsequent dilution leads to conformational folding. Because the reactions are very rapid, stopped-flow unfolding and folding, single- and double-jump, or pulsed-labeling, kinetic experiments are carried out, with the progress of the reaction monitored by the absorption or fluorescence properties of tyrosyl residues, circular dichroism, or deuterium−hydrogen exchange.

The conformational refolding of RNase A has long been known to be heterogeneous (*16*, *17*). Early kinetic studies of tyrosine fluorescence and absorbance indicated at least three distinct refolding phases (*18*): a fast-folding phase (U_F), a major slow-folding phase (U_S^{II}), and a minor very-slow-folding phase (U_S^{I}), accounting for 20, 65, and 15% of the refolding amplitude, respectively. These phases do not result from the presence of intermediates in the conformational refolding, but rather from heterogeneity in the unfolded state (*19*).

It was soon accepted that these distinct phases correspond to the refolding of different isomeric species (*1*), based on two lines of evidence. First, the relative amplitudes of the refolding phases are insensitive to mild refolding conditions and approximate the cis−trans ratio of X-Pro peptide groups in unstructured peptides. Second, when a native protein is unfolded for various times, and then allowed to refold (double-jump unfolding experiments), the variation of the final refolding amplitudes with the unfolding time exhibits several hallmarks of proline isomerization, e.g., a slow reaction (τ on a scale of seconds) with an activation energy of 20 kcal/mol (*1*), catalysis by strong acids (*4*, *20*), and an indifference to the unfolding conditions such as pH and denaturant concentrations (*1*, *21*).

Recent folding studies have discerned several additional phases by varying the final refolding conditions. At low pH, U_F resolves into a very-fast-folding species denoted U_{vf} and a fast-folding species denoted U_f, the equilibrium amplitudes of which (5 and 13%, respectively) are likewise independent of refolding conditions, suggesting the involvement of cis−trans isomers (*22*). U_{vf} is a unique species in which all four prolines have retained the conformation of the native protein, i.e., have not yet had time to isomerize in this unfolded state (see Figure 2). Careful analysis also revealed a small medium-refolding phase U_m, comprising roughly 7−10% of the refolding amplitude (*23*, *24*) which may have been observed in earlier studies (*25*).

Once we had discerned these five refolding phases (U_{vf}, U_f, U_m, U_S^{II}, and U_S^{I}), the central task became the identification of the isomeric species to which they correspond, i.e., to determine the conformational folding of each isomeric species. This could be accomplished by combining the following different approaches. First, it was helpful to determine which prolines are nonessential, reducing the number of isomeric species for consideration (*23*). (Nonessential prolines have no discernible effect on the conformational folding; essential prolines slow conformational folding, and critical prolines completely disrupt conformational folding.) Second, a kinetic analysis of the double-jump refolding kinetics of wild-type RNase A allowed the relative refolding amplitudes to be associated with the relative concentrations of specific isomeric species (*22*, *23*). The prolines of the kinetic model could be identified by carrying out analogous double-jump kinetic experiments on site-

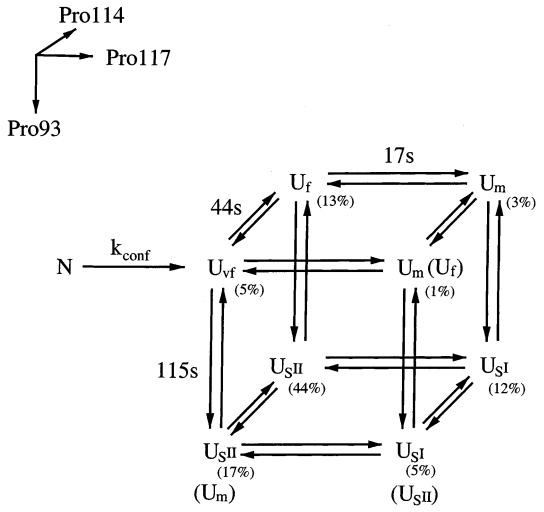

FIGURE 2: Best-fitting box model for the isomerization of Pro93, Pro114, and Pro117 under unfolding conditions (*23*, *32*). The parentheses indicate assignments in the old box model that differ from the new model.

directed mutants that eliminated specific proline isomerizations, typically proline to alanine (*23*). Third, the kinetic model could be cross-checked by directly assessing the isomerizations of specific prolines under unfolding conditions (*26*). Fourth, cis- and trans-locked proline analogues could be synthesized (*27*) and introduced into proteins (*28*, *29*); thus, stable, pure isomeric species can be prepared and their conformational folding studied.

Identifying the Essential Prolines. In a set of single-jump refolding experiments, the essential prolines of RNase A were determined with all possible single proline-to-alanine mutants, i.e., P42A, P93A, P114A, and P117A (*23*). It was concluded that prolines 93, 114, and 117 are essential, whereas Pro42 appears to be nonessential. This is somewhat surprising, since Pro42 occurs in a central, highly conserved β-strand. However, the proline is not buried, and presumably, the protein can rearrange to compensate for the cis isomer (*5*). Alternatively, the relative population of cis Lys41-Pro42 peptide groups may be too small to have a discernible effect on the conformational folding, because Lys-Pro peptide groups are predominantly trans under unfolding conditions (*30*, *31*).

Box Model of Proline Isomerization. In another approach, we investigated proline isomerization under *unfolding* conditions using double-jump refolding experiments on wild-type RNase A and its proline-to-alanine mutants (*22*, *23*). Unfolding conditions eliminated the complications arising from conformational and steric interactions; thus, the isomerization rates could be approximated by analogous data found in short peptides. In the first jumps to high GdnHC1 concentrations, the protein was denatured, allowing each proline to isomerize, although at different rates depending on the sequence environment. The second jumps, which diluted the GdnHC1, were carried out after various delay times, providing different mixtures of the isomeric species and different amplitudes of the refolding phases. Thus, in our double-jump refolding experiments, the amplitudes of the five refolding phases (U_{vf}, U_f, U_m, U_S^{II}, and U_S^{I}) were measured as a function of unfolding time. The data on the wild-type protein suggested that at least three independent proline isomerizations are responsible for the observed refolding phases (*22*, *24*). These processes are represented by a model in which the specific isomeric species are at the eight corners of a box (a so-called "box" model).

The prolines of the kinetic box model were associated with the three essential prolines of RNase A (Pro93, Pro114, and Pro117) as follows. The fitted equilibrium constants of the box model indicate that one proline is trans in the native state, while the other two are cis. Since Pro42 is nonessential, the trans proline of the box model was uniquely identified as Pro117. The two cis prolines were identified separately as Pro114 and Pro93 based on kinetic fitting of double-jump data on the proline-to-alanine mutants (*23*), in which their respective isomerization had been eliminated. The P114A data allowed an unambiguous identification of two cis prolines of the box model (*23*, *32*), which were then verified by an independent assay of these proline isomerizations (see Isomerization Assays of Specific Prolines).

The P93A double-jump data suggested that the Tyr92-Ala93 peptide group is cis in the native state (*23*), a finding subsequently confirmed by NMR (*33*) and X-ray crystallography (*34*). This is remarkable, since the cis content of non-N-substituted amino acids is typically less than 1% (*3*). The persistence of the cis isomer in Ala93 indicates strong conformational interactions maintaining the structure of the major β-hairpin near Pro93, and supporting its identification as a critical structure in conformational folding (*32*, *35*, *36*).

To refine the model, the double jumps for the wild-type and mutant proteins were fit to all possible kinetic models of three independently isomerizing prolines. In the best-fitting box model (*32*), three isomeric species were assigned to a different refolding phase than in the old box model (*23*, *24*) (the old model being indicated in parentheses in Figure 2). However, both the old and new models agree that the nonnative isomers of Pro114, Pro117, and Pro93 cause disruption of conformational folding, with Pro93 causing the most disruption (*22*, *23*, *32*). Although the differences between the two models are only in three positions of the box model, the new box model provides a better fit to the experimental data. Some uncertainties with the assignment of the specific isomeric species, however, still remain. The fitted isomerization rate constant for the Val116-Pro117 peptide is nearly 10 times faster than that observed in short peptides containing the Val-Pro sequence (*37*), in contrast to those of the Tyr92-Pro93 and Asn113-Pro114 peptides which agree roughly with the values of their small-peptide analogues. Second, the box model suggests that the U_m refolding phase should be eliminated in the P93A and P117A mutants; however, no change in the amplitude or rate constant could be observed in these mutants (*23*), which is consistent with the old model. The likely origin of these uncertainties is the fact that eight isomeric species could give rise to eight refolding phases; however, the data allow only five refolding phases to be discerned. Thus, some isomeric species may contribute to more than one experimentally observed refolding phase. This problem can be addressed by preparing stable pure isomeric species.

Isomerization Assays of Specific Prolines. Fluorescence assays for the isomerization of Pro93 and Pro114 in RNase A under unfolding conditions were developed (*26*). The unfolding of RNase A is monophasic when observed by tyrosine absorbance, but biphasic when observed by tyrosine fluorescence (*38*). The fast fluorescence phase coincides with

the absorbance phase, which occurs on the millisecond time scale and corresponds to conformational unfolding, specifically to the exposure of the buried tyrosines (Tyr25, Tyr92, and Tyr97) (*39*) to solvent (*32*). However, the *slow* fluorescence unfolding phase occurs after conformational unfolding, and exhibits the classic hallmarks of proline isomerization.

RNase A has four prolines and six tyrosines, and consequently, we had to determine which prolines contribute to this slow fluorescence unfolding phase, and which tyrosines report on their isomerizations. Accordingly, the slow fluorescence unfolding phase was examined in all possible proline-to-alanine and tyrosine-to-phenylalanine mutants; the former mutants eliminate the slow isomerization, and the latter mutants eliminate its reporting. These knockout mutations demonstrated that Tyr92 reports on the isomerization of Pro93 and Tyr115 reports on that of Pro114 (*26*), with roughly equal amplitudes. To check these assignments, two double mutants were made: Y92F/Y115F (which eliminates both reporting groups) and Y92F/P114A (which eliminates the one reporting group and the other isomerizing proline); in both cases, the slow fluorescence unfolding phase was eliminated entirely. The latter double mutant Y92F/P114A indicates that the reporting is very local, since Y115F does not report on the isomerization of Pro117 (*26*).

Thus, the slow fluorescence unfolding phase of wild-type RNase A corresponds to the isomerizations of Pro93 and Pro114. These isomerizations may be observed *individually* in the Y92F and Y115F mutants, which appear to be very similar to the wild-type protein (*26*) in their enzymatic activities and conformational folding and unfolding. The isomerization time constants (41 and 115 s) for these two prolines measured in the four mutants agree nearly perfectly with the box model isomerization rate constants for these prolines in the wild-type protein, confirming the kinetic fit for these two prolines (*26*, *32*). This fluorescence assay for proline isomerization appears to be general, since similar slow fluorescence unfolding phases have been observed in other proteins where tyrosine is adjacent to proline, e.g., the MerP protein (*40*).

The isomerization rate constant for the Tyr92-Pro93 peptide group was measured independently using the Y92W mutant (*41*). Wild-type RNase A has no tryptophan residues, and thus, the slow Trp fluorescence phase observed under unfolding should correspond to the isomerization of Pro93. A key difference should be noted; in the tyrosine-to-phenylalanine and proline-to-alanine mutants, the wild-type residues are observed and the mutated residue is not observed (except by omission). In contrast, the mutated residue is observed directly in the tyrosine-to-tryptophan mutants, and the Trp-Pro peptide group may not isomerize at the same rate as the Tyr-Pro peptide group. Indeed, from model peptide studies, the rate of isomerization of Trp-Pro peptide groups is 2-fold slower than that of Tyr-Pro and Phe-Pro peptide groups (*37*). However, after correction by this factor of 2, the Tyr92-Pro93 isomerization time constant estimated from the Y92W mutant agrees with that measured in the Y115F and P114A mutants.

An earlier study (*42*) of the homologous guinea pig ribonuclease A, which has no slow fluorescence unfolding phase, misidentified Tyr92 as being solely responsible for the slow fluorescence unfolding phase, which was detected in the bovine form. Their conclusion was based on the sequence differences: Tyr92 and the Asn113-Pro114 peptide group in the bovine form are replaced with Phe92 and the Lys113-Pro114 peptide group, respectively, in the guinea pig. A plausible explanation for the absence of the slow phase is that the Lys-Pro peptide bond is trans in the folded guinea pig form and remains overwhelmingly trans in the unfolded state (*30*, *31*); thus, the contribution of Tyr115 may not have been observable if the Lys113-Pro114 peptide group had undergone negligible isomerization.

This misidentification illustrates a serious problem, however. As shown above, the slow fluorescence unfolding phase in wild-type RNase A is biphasic, corresponding to the sum of two independently isomerizing prolines. However, this biphasic character could be detected only by site-directed mutagenesis; all prior studies [spanning nearly 20 years (*38*)] reported only a single exponential phase. The errors in the fluorescence measurements are not sufficiently small to discern the two phases statistically, although they differ 3-fold in their rate constants. For illustration, the slow fluorescence unfolding of a 50%/50% mixture of Y92F and Y115F also appears to be a single exponential phase (identical to the wild-type phase), although it is unquestionably composed of two exponential phases (*26*). Thus, the conclusions of all such kinetic fittings should be regarded as provisional, until they can be confirmed by site-directed mutagenesis.

It has been proposed that the tyrosine fluorescence changes under proline isomerization reflect interactions of the tyrosine ring with the C40−C95 disulfide bond (*38*). However, the slow unfolding fluorescence phase of the C40A/C95A mutant (in which this disulfide bond is eliminated) is identical to that of wild-type RNase A (*43*). It is more likely that the fluorescence changes reflect interactions of the tyrosine ring with the proline ring (*44*−*46*) or, possibly, with the peptide group (*47*).

Structural Interpretations of Isomeric Folding. The box model indicates that non-native isomers of Pro114, Pro117, and Pro93 cause increasing levels of disruption of conformational folding (*32*). However, kinetic studies have also shown that only the refolding of U_S^I is rate-limited by proline isomerization; the other species fold conformationally before they isomerize, albeit more slowly than in the native isomeric species (*18*, *24*). The native isomer of Pro93 seems to be particularly critical for conformational folding, since a non-native isomer of Pro93 by itself slows conformational folding almost 1000-fold, and conformational folding is blocked entirely in species with a non-native Pro93 isomer and additional destabilizing mutations such as Y25F and Y97F (*26*). Somewhat surprisingly, conformational folding is still relatively rapid in the non-native Pro117 species (*23*), although Pro117 is completely buried in the hydrophobic core of the native protein. This suggests that the acquisition of the major β-hairpin secondary structure near Pro93 is more critical for folding than a structurally correct hydrophobic core. This hypothesis is supported by early antigenic studies (*48*) as well as by large disruptions of conformational folding (*32*) resulting from conservative mutations of tyrosines to phenylalanine in this region, which eliminate a single hydrogen bond.

This structural interpretation of the isomerization data is consistent with our chain-folding initiation site (CFIS) folding

mechanism (*35*, *49*−*52*). Briefly, this mechanism predicts that flickering nuclei of secondary structure motifs such as α-helices and β-hairpins form in regions with high secondary structure propensities and high local concentrations (along the protein backbone) of nonpolar residues. These nuclei associate through medium- and long-range interactions to form a stable core onto which the remainder of the protein folds. In this mechanism, the rate of conformational folding is determined largely by the loop entropy cost of making the medium- and long-range interactions to form the core (i.e., effectively by conformational searching), although electrostatic interactions between the nuclei could play a role as well.

Given this folding scenario, it is plausible to hypothesize that the incorrect isomer at Pro93 inhibits the formation of the adjacent β-hairpin motif by preventing the formation of the nucleating type VI turn at Pro93. Formation of the native hairpin would have to be nucleated by a longer-range interaction. Moreover, formation of the native hairpin may also compete with the formation of non-native local structure, e.g., in the β-bulge region (residues 88−91), which would further slow conformational folding.

Refolding of U_{vf}. In the box model of Figure 2, the concentration of U_{vf} *at equilibrium* is very low. However, this species could be prepared in an almost 100% pure state by a double-jump procedure (*22*, *24*, *41*, *53*−*55*). Starting in the native state (which is isomerically pure), the RNase A was jumped to strongly unfolding conditions (viz., 4.2 M GdnHCl at pH 2 and 15 °C), in which the conformational unfolding proceeds much more rapidly than the proline isomerization. The protein was held under these conditions long enough to allow complete unfolding of the protein to U_{vf} but not long enough to populate any non-native isomeric state significantly (roughly 0.8 s at 15 °C). The protein was then returned to folding conditions, and the conformational refolding of U_{vf} was monitored by various probes such as absorbance, fluorescence, circular dichroism, and H−D exchange. Thus, conformational folding could be studied directly without cis−trans isomerization.

The pH and GdnHCl dependence of the refolding kinetics indicated that a conformational intermediate (denoted $I_ø$) is populated in refolding (*53*). The intermediate appears to be favored under stabilizing conditions, e.g., at low denaturant conditions or neutral pH (*53*). H−D exchange experiments have suggested that the amide protons of the minor hydrophobic core (near the C26−C84 disulfide bond) become strongly protected; on the other hand, those of the major hydrophobic core (near the C58−C110 disulfide bond) receive medium protection, and others receive little or no protection (*55*).

Site-directed mutants also indicate the importance of the minor hydrophobic core in the refolding of U_{vf}. The Y25F and Y97F mutants slow the folding drastically (*32*), suggesting that these tyrosine residues are required for this folding. The Y92W mutant (*41*) exhibits a burst phase by fluorescence, a relatively small burst phase in the refolding of U_{vf}, and a larger burst phase in single-jump refolding experiments from the equilibrium unfolded state. Subsequent experiments (unpublished results) have indicated that this fluorescence burst phase is composed of two well-resolved exponentials on the sub-millisecond time scale, with a still-unresolved burst phase.

Conformational Folding of U_S^{II}. Kinetic studies (*26*, *32*) suggested that the U_S^{II} phase comprises the refolding of two distinct isomeric states. The conformational folding of U_S^{II} has been studied by pulsed H−D exchange methods and correlated with optical probes (*56*, *57*). The results suggest that the C-terminal, predominately β-sheet half of RNase A is almost fully formed in an intermediate of conformational folding, and that the rate-determining step consists of the ordering of the N-terminal, predominately α-helical half onto the β-sheet. Fluorescence resonance energy transfer (FRET) experiments of the fully reduced form under folding conditions (*58*) support this interpretation.

The conformational refolding of U_S^{II} has also been studied using a cross-linking protocol (*59*). Lysines 7 and 41 were cross-linked through their N^{ϵ} atoms using 1,5-difluoro-2,4-dinitrobenzene. The cross-linked protein exhibits the same native structure as wild-type RNase A (*60*), and similar refolding phases (*61*). The cross-linked protein is thermodynamically much more stable than wild-type RNase A (*59*), and this change in stability appears to result largely from entropic (not enthalpic) factors, consistent with the high conformational lability observed in the loop region (residues 16−23). Therefore, the cross-linked protein could be used to assess the association of lysines 7 and 41 in the folding transition state of the wild-type protein (*61*, *62*). The results suggested that the N-terminal α-helix is not associated with the C-terminal β-sheet in the transition state for conformational folding (*61*, *62*) of U_S^{II}. The FRET measures (*58*) also support this suggestion.

Conformational Folding of the Other Isomeric Species. The conformational folding of U_f is generally similar to that of U_{vf} for pH values above 5. This is consistent with experimental and theoretical studies indicating that the cis−trans isomerization of the Asn113−Pro114 peptide bond makes relatively minor, localized changes that do not affect the structure of the adjacent hydrophobic core of residues 58−110 (*5*, *23*, *36*, *63*−*67*). However, at low pH, the conformational folding of U_f differs significantly from that of U_{vf}.

The refolding of U_S^I appears to be rate-limited by proline isomerization (*24*). Hence, the conformational folding of the corresponding isomeric states cannot be studied, although the results indicate that non-native isomers of Pro93 and Pro117 suffice to make the native state thermodynamically unstable relative to the unfolded state.

Preparation of Stable Isomeric Species. The goal of these studies of RNase A was to characterize the conformational folding of individual isomeric species to investigate the role of backbone interactions in pathways of conformational folding. However, except for the native species, it is difficult to prepare a pure isomeric species; the unfolded state is generally a mixture of species due to the isomeric equilibration that occurs under unfolding conditions. The RNase A experiments were able to discern the folding of the individual isomeric species only through careful kinetic analysis of the single-jump and double-jump folding and unfolding experiments carried out on wild-type and mutant proteins. It would be preferable to prepare stable, pure isomeric species so that their conformational folding and unfolding could be assessed in isolation.

Unfortunately, there is no naturally occurring amino acid locked in the cis isomer. Non-prolyl amino acids have a strong preference for the trans isomer, and thus, it is theoretically possible to prepare a stable, pure all-trans isomeric species by mutating every proline to alanine (*23*). Alternatively, 2,4-methanoproline which is locked in the trans conformation (*68*, *69*) and L-5,5-dimethylproline (dmP) which is locked in the cis conformation (*27*) can be used to obtain a stable protein with specific X-Pro peptide groups locked in given conformations. In particular, the cis-locked proline analogue dmP has been prepared and studied by NMR in two tripeptides, Tyr-dmP-Asn and Asn-dmP-Tyr, corresponding to the tripeptides containing the cis prolines of RNase A (residues 92−94 and 113−115) (*27*). In addition, a general synthetic method is being developed for introducing such unnatural proline analogues into proteins (*28*, *29*). Thus, it is possible to prepare stable, isomerically pure species of RNase A and to study their conformational folding directly. This work is underway in our laboratory.

Proline Isomerization in the Folding of Other Proteins

Slow refolding and unfolding phases have been detected in many proteins (*70*−*72*). Most of these slow phases have been correlated with cis−trans isomerization, displaying many of the classic hallmarks, e.g., a slow reaction with an activation energy of 20 kcal/mol, an indifference to solution conditions, or catalysis by PPIases. Unfortunately, these slow phases have been correlated with specific isomeric species for only relatively few proteins, e.g., RNase A (described above), RNase T1 (*73*, *74*), staphylococcal nuclease (SNase) (*75*), bovine pancreatic trypsin inhibitor (BPTI) (*76*), thioredoxin (*77*), iso-2-cytochrome *c* (*78*), $p13^{suc1}$ (*79*), cellular retinoic acid binding protein I [CRABPI (*80*)], the mercury binding protein MerP (*40*), the C_L fragment of the immunoglobulin light chain (*81*), and the SH3 domain of the p85α subunit of bovine phosphatidylinositol 3′-kinase (P13 kinase) (*82*).

Some prolines appear to be completely buried in the hydrophobic core, e.g., Pro117 in RNase, the buried prolines (Pro11, Pro31, and Pro56) of staphylococcal nuclease (*75*), Pro76 of thioredoxin (*77*), Pro39 of RNase T1 (*74*), and Pro85 in CRABPI (*80*). However, some buried prolines do not appear to give rise to slow folding phases, e.g., Pro134 in the Fyn-SH3 domain (*83*), although this may reflect an unusually low non-native cis content in the unfolded state (<4%). Buried prolines are relatively rare, presumably because the proline residue may disrupt the hydrogen bonding of local residues.

Essential prolines appear to be associated with the cis conformation that forms tight turns, e.g., Pro93 in RNase A, Pro143 in the C_L fragment of immunoglobulin (*81*), Pro55 in RNase T1 (*73*), Pro76 in iso-2-cytochrome *c* (*78*), Pro67 in MerP (*40*), Pro90 in $p13^{suc1}$ (*79*), and Pro70 in the SH3 domain of PI3 kinase (*82*). Such hairpin turns usually occur on the surface of a protein, and thus, such prolines are not buried and presumably do not stabilize the folding pathways by direct hydrophobic or hydrogen bonding interactions. Instead, native isomers of such prolines probably stabilize the folding pathways indirectly by promoting the association of residues which in turn stabilize the folding pathways through hydrophobic and hydrogen bonding interactions (*50*). Conversely, non-native isomers likely retard folding by inhibiting such association and also by permitting non-native nuclei to compete with the native nuclei.

In contrast, nonessential prolines are generally in positions that are not critical for rapid folding, or they are associated with isomer-neutral conformations, i.e., conformations which tolerate both isomers. This often occurs in irregular and/or labile structures near the surface of the protein, e.g., prolines 15 and 17 in roe deer RNase (*84*). This type of proline appears to be the most common; examples of such prolines are found in most proteins. There are several examples of proteins with more than 10 proline residues which have no slow-folding phase whatsoever (*85*); presumably, all of the prolines are nonessential.

An unresolved question is how closely the cis−trans ratio and isomerization kinetics of X-Pro peptide groups in proteins under *unfolding* conditions are approximated by their behavior in model peptides. Negligible differences in the rate constants would make it easier to interpret double-jump refolding studies (N → U → N) since the kinetics of the first step (N → U) could be determined in advance from studies of model peptide analogues. This would help in measuring the effects of non-native isomers on the conformational *refolding* kinetics. In general, the evidence indicates that model peptide kinetics correspond roughly to the kinetics in the unfolded protein (*37*), but this correspondence is not robust enough to predict the populations of isomeric species.

Summary

Non-native proline isomers are an effective probe for studying the conformational folding of proteins. Unlike site-directed mutagenesis or changes in disulfide bonding, proline isomerization is a *conformational* change in a single dihedral angle.

Proline isomerization can be exploited in two ways. First, non-native proline isomers can be treated as transient site-directed mutants, since the isomerization rate is typically much slower than the conformational folding rate. The conformational folding of specific isomeric species indicates which chain segments are critical for conformational folding. Second, the isomerization rates and equilibrium cis−trans ratios of specific prolines can be used to detect local structure and to monitor the local environment of the prolines. It should be stressed that the goal of these studies is not to characterize proline isomerization per se, but rather to exploit the kinetic and structural clarity of the isomerization reaction as an assay of conformational structure and of the pathways of conformational folding.

Non-native proline isomers can have widely varying effects on conformational folding. Some prolines have no discernible effect on conformational folding (nonessential prolines); some prolines slow conformational folding (essential prolines), often as much as a 1000-fold, and some prolines completely disrupt conformational folding (critical prolines).

It is remarkable that a change in a single dihedral angle (the peptide bond dihedral angle ω) should be capable of making such dramatic changes in the conformational folding kinetics. A structurally plausible explanation is that the non-native isomer disrupts conformational folding by altering a critical pattern of peptide−peptide hydrogen bonds (i.e., local

secondary structure) that is essential for the subsequent folding. For example, the native cis isomer of Pro93 in RNase A naturally forms a β-turn, which may then act as a chain-folding initiation site (CFIS), catalyzing the formation of the two-stranded β-hairpin (residues 79−104) (*50*). The non-native trans isomer of Pro93 disrupts the native turn (*67*), which requires that a larger loop be closed to form the correct β-hairpin, slowing its formation. Moreover, other chain segments (e.g., the β-bulge from residue 88 to 91) may form incorrect loops which compete with the native loop.

This hypothesis is supported by a survey of the essential prolines of various proteins. Nonessential prolines tend to be found in loosely ordered loop segments of the native fold (e.g., Pro19 in bovine seminal RNase A), or in secondary structure elements which can accommodate both isomers. Essential prolines occur in regions of critical secondary structure, either completely buried in the structure (e.g., Pro117 of RNase A) or in tight turns (e.g., Pro93 of RNase A).

Thus, it seems that specific secondary structure is important at an early stage of conformational folding. This is consistent with the older CFIS model for folding (*35*, *49*−*51*) and disagrees with several "new view" models of protein folding, which postulate that specific secondary and tertiary structure plays a role only in the final stages of protein folding (*86*).

The methodology for assaying proline isomerization has been steadily improving in recent years. Several fluorescence and isomer-specific proteolytic methods have been developed for proteins and peptides, as well as the combination of site-directed mutagenesis and single- and double-jump kinetic experiments.

ACKNOWLEDGMENT

We thank Mahesh Narayan for helpful comments on the manuscript and Walid Houry for helpful comments on an earlier draft of the manuscript.

REFERENCES

1. Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975) *Biochemistry 14*, 4953−4963.
2. Wüthrich, K., Grathwohl, C., and Schwyzer, R. (1974) in *Peptides, Polypeptides and Proteins*" (Blout, E. R., Bovey, F. A., Goodman, M., and Lotan, N., Eds.) pp 300−307, Wiley, New York.
3. Zimmerman, S. S., and Scheraga, H. A. (1976) *Macromolecules 9*, 408−416.
4. Steinberg, I. Z., Harrington, W. F., Berger, A., Sela, M., and Katchalski, E. (1960) *J. Am. Chem. Soc. 82*, 5263−5279.
5. Pincus, M. R., Gerewitz, F., Wako, H., and Scheraga, H. A. (1983) *J. Protein Chem. 2*, 131−146.
6. Kang, Y. K. (2002) *J. Phys. Chem. B 106*, 2074−2082.
7. Corey, R. B., and Pauling, L. (1953) *Proc. R. Soc. London, Ser. B 141*, 10−20.
8. Drakenberg, T., Dahlqvist, K.-I., and Forsén, S. (1972) *J. Phys. Chem. 76*, 2178−2183.
9. Mandelkern, L. (1967) in *Poly-α-Amino Acids* (Fasman, G. D., Ed.) pp 675−724, Marcel Dekker, New York.
10. Tanaka, S., and Scheraga, H. A. (1975) *Macromolecules 8*, 516−521.
11. Stein, R. L. (1993) *Adv. Protein Chem. 44*, 1−24.
12. Schmid, F. X., Mayr, L. M., Mücke, M., and Schönbrunner, E. R. (1993) *Adv. Protein Chem. 44*, 25−66.
13. Neira, J. L., and Rico, M. (1997) *Folding Des. 2*, R1−R11.
14. Raines, R. T. (1998) *Chem. Rev. 98*, 1045−1065.
15. Wlodawer, A., Svensson, L. A., Sjölin, L., and Gilliland, G. L. (1988) *Biochemistry 27*, 2705−2717.
16. Tsong, T. Y., Baldwin, R. L., and Elson, E. L. (1971) *Proc. Natl. Acad. Sci. U.S.A. 68*, 2712−2715.
17. Tsong, T. Y., Baldwin, R. L., and Elson, E. L. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1809−1812.
18. Cook, K. H., Schmid, F. X., and Baldwin, R. L. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 6157−6161.
19. Garel, J.-R., and Baldwin, R. L. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 3347−3351.
20. Schmid, F. X., and Baldwin, R. L. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4764−4768.
21. Schmid, F. X., and Baldwin, R. L. (1979) *J. Mol. Biol. 133*, 285−287.
22. Houry, W. A., Rothwarf, D. M., and Scheraga, H. A. (1994) *Biochemistry 33*, 2516−2530.
23. Dodge, R. W., and Scheraga, H. A. (1996) *Biochemistry 35*, 1548−1559.
24. Houry, W. A., and Scheraga, H. A. (1996) *Biochemistry 35*, 11719−11733.
25. Lin, L.-N., and Brandts, J. F. (1983) *Biochemistry 22*, 564−573.
26. Juminaga, D., Wedemeyer, W. J., and Scheraga, H. A. (1998) *Biochemistry 37*, 11614−11620.
27. An, S. S. A., Lester, C. C., Peng, J.-L., Li, Y.-J., Rothwarf, D. M., Welker, E., Thannhauser, T. W., Zhang, L. S., Tam, J. P., and Scheraga, H. A. (1999) *J. Am. Chem. Soc. 121*, 11558−11566.
28. Welker, E., and Scheraga, H. A. (1999) *Biochem. Biophys. Res. Commun. 254*, 147−151.
29. Cerovsky, V., Welker, E., and Scheraga, H. A. (2002) *J. Pept. Res.* (in press).
30. Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A., and Wright, P. E. (1988) *J. Mol. Biol. 201*, 161−200.
31. Raleigh, D. P., Evans, P. A., Pitkeathly, M., and Dobson, C. M. (1992) *J. Mol. Biol. 228*, 338−342.
32. Juminaga, D., Wedemeyer, W. J., Garduño-Júarez, R., McDonald, M. A., and Scheraga, H. A. (1997) *Biochemistry 36*, 10131−10145.
33. Xiong, Y., Juminaga, D., Swapna, G. V. T., Wedemeyer, W. J., Scheraga, H. A., and Montelione, G. T. (2000) *Protein Sci. 9*, 421−426.
34. Pearson, M. A., Karplus, P. A., Dodge, R. W., Laity, J. H., and Scheraga, H. A. (1998) *Protein Sci. 7*, 1255−1258.
35. Matheson, R. R., Jr., and Scheraga, H. A. (1978) *Macromolecules 11*, 819−829.
36. Oka, M., Montelione, G. T., and Scheraga, H. A. (1984) *J. Am. Chem. Soc. 106*, 7959−7969.
37. Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M., and Fischer, G. (1998) *J. Mol. Biol. 279*, 449−460.
38. Rehage, A., and Schmid, F. X. (1982) *Biochemistry 21*, 1499−1505.
39. Scheraga, H. A. (1967) *Fed. Proc. 26*, 1380−1387.
40. Aronsson, G., Brorsson, A.-C., Sahlman, L., and Jonsson, B.-H. (1997) *FEBS Lett. 411*, 359−364.
41. Sendak, R. A., Rothwarf, D. M., Wedemeyer, W. J., Houry, W. A., and Scheraga, H. A. (1996) *Biochemistry 35*, 12978−12992.
42. Schmid, F. X., Grafl, R., Wrba, A., and Beintema, J. J. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 872−876.
43. Iwaoka, M., Wedemeyer, W. J., and Scheraga, H. A. (1999) *Biochemistry 38*, 2805−2815.
44. Yao, J., Dyson, H. J., and Wright, P. E. (1994) *J. Mol. Biol. 243*, 754−766.
45. Kemmink, J., and Creighton, T. E. (1995) *J. Mol. Biol. 245*, 251−260.
46. Wu, W.-J., and Raleigh, D. P. (1998) *Biopolymers 45*, 381−394.
47. Cowgill, R. W. (1976) in *Biochemical Fluorescence: Concepts*, Vol. 2, pp 441−486.
48. Chavez, L. G., Jr., and Scheraga, H. A. (1980) *Biochemistry 19*, 996−1004.
49. Lewis, P. N., Gō, N., Gō, M., Kotelchuck, D., and Scheraga, H. A. (1970) *Proc. Natl. Acad. Sci. U.S.A. 65*, 810−815.
50. Lewis, P. N., Momany, F. A., and Scheraga, H. A. (1971) *Proc. Natl. Acad. Sci. U.S.A. 68*, 2293−2297.
51. Lewis, P. N., Momany, F. A., and Scheraga, H. A. (1973) *Biochim. Bioophys. Acta 303*, 211−229.
52. Tanaka, S., and Scheraga, H. A. (1977) *Macromolecules 10*, 291−304.
53. Houry, W. A., Rothwarf, D. M., and Scheraga, H. A. (1995) *Nat. Struct. Biol. 2*, 495−503.
54. Houry, W. A., Rothwarf, D. M., and Scheraga, H. A (1996) *Biochemistry 35*, 10125−10133.

55. Houry, W. A., and Scheraga, H. A. (1996) *Biochemistry 35*, 11734−11746.
56. Udgaonkar, J. B., and Baldwin, R. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8197−8201.
57. Udgaonkar, J. B., and Baldwin, R. L. (1995) *Biochemistry 34*, 4088−4096.
58. Navon, A., Ittah, V., Landsman, P., Scheraga, H. A., and Haas, E. (2001) *Biochemistry 40*, 105−118.
59. Lin, S. H., Konishi, Y., Denton, M. E., and Scheraga, H. A. (1984) *Biochemistry 23*, 5504−5512.
60. Weber, P. C., Salemme, F. R., Lin, S. H., Konishi, Y., and Scheraga, H. A. (1985) *J. Mol. Biol. 181*, 453.
61. Lin, S. H., Konishi, Y., Nall, B. T., and Scheraga, H. A. (1985) *Biochemistry 24*, 2680−2686.
62. Wedemeyer, W. J., Welker, E., Narayan, M., and Scheraga, H. A. (2000) *Biochemistry 39*, 4207−4216.
63. Stimson, E. R., Montelione, G. T., Meinwald, Y. C., Rudolph, R. K. E., and Scheraga, H. A. (1982) *Biochemistry 21*, 5252−5262.
64. Ihara, S., and Ooi, T. (1985) *Biochim. Biophys. Acta 830*, 109−112.
65. Schultz, D. A., Schmid, F. X., and Baldwin, R. L. (1992) *Protein Sci. 1*, 917−924.
66. Schultz, D. A., Friedman, A., and Fox, R. O. (1993) *Protein Sci. 2* (Suppl. 1), 67, Abstract 66-M.
67. Montelione, G. T., Arnold, E., Meinwald, Y. C., Stimson, E. R., Denton, J. B., Huang, S.-G., Clardy, J., and Scheraga, H. A. (1984) *J. Am. Chem. Soc. 106*, 7946−7958.
68. Talluri, S., Montelione, G. T., van Duyne, G., Piela, L., Clardy, J., and Scheraga, H. A. (1987) *J. Am. Chem. Soc. 109*, 4473−4477.
69. Piela, L., Némethy, G., and Scheraga, H. A. (1987) *J. Am. Chem. Soc. 109*, 4477−4485.
70. Kim, P. S., and Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459−489.
71. Schmid, F. X. (1993) *Annu. Rev. Biophys. Biomol. Struct. 22*, 123−143.
72. Nall, B. T. (1994) in *Mechanism of Protein Folding* (Pain, R. H., Ed.) pp 80−103, IRL Press, Oxford, U.K.
73. Odefey, C., Mayr, L. M., and Schmid, F. X. (1995) *J. Mol. Biol. 245*, 69−78.
74. Mayr, L. M., Odefey, C., Schutkowski, M., and Schmid, F. X. (1996) *Biochemistry 35*, 5550−5561.
75. Maki, K., Ikura, T., Hayano, T., Takahashi, N., and Kuwajima, K. (1999) *Biochemistry 38*, 2213−2223.
76. Hurle, M. R., Anderson, S., and Kuntz, I. D. (1991) *Protein Eng. 4*, 451−455.
77. Kelley, R. F., and Richards, F. M. (1987) *Biochemistry 26*, 6765−6774.
78. Wood, L. C., White, T. B., Ramdas, L., and Nall, B. T. (1988) *Biochemistry 27*, 8562−8568.
79. Rousseau, F., Schymkowitz, J. W. H., Sánchez delPino, M., and Itzhaki, L. S. (1998) *J. Mol. Biol. 284*, 503−519.
80. Clark, P. L., Weston, B. F., and Gierasch, L. M. (1998) *Folding Des. 3*, 401−412.
81. Goto, Y., and Hamaguchi, K. (1982) *J. Mol. Biol. 156*, 891−910.
82. Guijarro, J. I., Morton, C. J., Plaxco, K. W., Campbell, I. D., and Dobson, C. M. (1998) *J. Mol. Biol. 276*, 657−667.
83. Plaxco, K. W., Guijarro, J. I., Morton, C. J., Pitkeathly, M., Campbell, I. D., and Dobson, C. M. (1998) *Biochemistry 37*, 2529−2537.
84. Krebs, H., Schmid, F. X., and Jaenicke, R. (1983) *J. Mol. Biol. 169*, 619−635.
85. Creighton, T. E. (1978) *J. Mol. Biol. 125*, 401−406.
86. Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. (1997) *Annu. Rev. Phys. Chem. 48*, 545−600.

BI020574B