Direct NMR Measurement of the Folding Kinetics of a Trimeric Peptide†

Xiaoyan Liu,‡ Donald L. Siegel,‡ Pei Fan,‡ Barbara Brodsky,§ and Jean Baum*,‡

Department of Chemistry, Rutgers University, Piscataway, New Jersey 08855-0939, and Department of Biochemistry, UMDNJ-Robert Wood Johnson Medical School, Piscataway, New Jersey 08854

Received September 21, 1995; Revised Manuscript Received December 20, 1995⊗

ABSTRACT: Direct NMR measurements of the folding kinetics are performed on a collagen-like triple helical peptide. The triple helical peptide was designed to model a biologically important region of collagen and has the sequence $(POG)_3ITGARGLAG(POG)_4$. Triple helical peptides were synthesized with specifically labeled ^{15}N amino acid residues in key positions, and the kinetics of folding of the individual residues were monitored directly by measuring the loss of monomer intensity and the increase in trimer intensity. The residues at the terminal ends and central region could be followed independently and quantitated directly. Residues located at the terminal ends have rates and kinetics of folding that are distinct from residues in the central region of the peptide. This allows the monitoring of different steps in the folding mechanism and the postulation of the existence of a kinetic intermediate. The NMR data are consistent with a mechanism of association/nucleation and propagation. Hereditary connective tissue diseases are associated with mutations that result in abnormal folding of collagen, and the NMR folding experiments on a collagen-like peptide provide a basis for characterizing the molecular defect in folding mutations.

The objective of studying protein folding is to determine the mechanism by which a protein proceeds from the unfolded state to the fully folded state. Characterization of the folding processes of proteins is important to understand the determinants of protein structure and function. A number of NMR[1] approaches have been used to understand protein folding and to characterize folding intermediates in structural and kinetic terms. One approach is to characterize the conformation of denatured states of proteins, which can vary from highly unfolded forms in which the protein contains no residual structure, to compact forms in which the protein contains well defined and persistent structure (Kuwajima, 1989; Dill & Shortle, 1991; Shortle, 1993; Dobson, 1994; Wüthrich, 1994; Ptitsyn, 1995). Studies on denatured states and small peptide models (Dyson & Wright, 1993) have helped clarify the factors that are important in early folding events. A second approach is to use NMR pulse labeling techniques where hydrogen exchange kinetics are used to monitor the location of protected amides during the kinetics of protein refolding (Roder et al., 1988; Udgaonkar & Baldwin, 1988; Baldwin, 1993; Dobson et al., 1994; Elöve et al., 1994; Woodward, 1994). Several proteins have been studied successfully using these methods and have given information about the secondary structures of early folding intermediates.

A complementary NMR approach to pulse labeling and denaturation experiments is to apply real time kinetics to the folding of proteins. The ability to detect the loss of intensity of the unfolded peaks or the increase of intensity of the folded peaks would allow direct monitoring of reaction kinetics and would form the basis for postulating folding mechanisms. Direct measurement of kinetics of folding in real time is often difficult to implement with NMR because the folding of globular proteins is typically fast relative to the detection time. However, real time NMR folding studies have been used in certain cases to follow unfolding of proteins (Kiefhaber et al., 1995), to probe slow folding events such as proline isomerization (Blum et al., 1978; Adler & Scheraga, 1988; Akasaka et al., 1991; Koide et al., 1993), and to monitor transient folding intermediates that resemble a molten globule state (Balbach et al., 1995). Folding in these real time experiments has been initiated on a time scale of seconds to minutes either by a temperature jump (Blum et al., 1978; Adler & Scheraga, 1988; Akasaka et al., 1991; Kautz & Fox, 1993) or by stopped-flow approaches involving dilution from denaturant (Frieden et al., 1993; Koide et al., 1993; Balbach et al., 1995). In this paper, we present a triple helical system for which folding is slow and for which real time NMR folding can be monitored.

Folding of the triple helix offers an opportunity to look at multichain molecular assembly. In triple helices this is a process which is very slow compared to the folding of most proteins. The triple helix is a basic structural motif in proteins found in all collagens, as well as in other proteins, such as macrophage scavenger receptor, C1q, and mannose binding protein (Hoppe & Reid, 1994; Brodsky & Shah, 1995; Prockop & Kivirikko, 1995). Two-dimensional NMR studies (Fan et al., 1993; Li et al., 1993) and recent X-ray crystallographic studies (Bella et al., 1994) have confirmed the basic model of the triple helical conformation as being three polyproline II-like chains supercoiled about each other (Rich & Crick, 1961). The three chains are staggered by

† This work was supported by NIH Grants GM45302 (J.B.) and AR19626 (B.B.). J.B. is a Sloan Fellow and a Camille and Henry Dreyfus Teacher-Scholar.

* To whom correspondence should be addressed.

‡ Rutgers University.

§ UMDNJ-Robert Wood Johnson Medical School.

⊗ Abstract published in *Advance ACS Abstracts,* March 1, 1996.

[1] Abbreviations: NMR, nuclear magnetic resonance; CD, circular dichroism; O, one-letter abbreviation for 4-hydroxyproline; HSQC, heteronuclear single-quantum coherence; SE-HSQC, sensitivity-enhanced single-quantum coherence; T3−785, a peptide derived from human type III collagen starting at residue 785 with the sequence $(POG)_3ITGARGLAG(POG)_4$.

0006-2960/96/0435-4306$12.00/0

one residue with respect to each other and are linked through interchain hydrogen bonds. The conformation requires that every third residue be a glycine, generating a repeating (Gly-X-Y)$_n$ pattern, and a high proportion of the residues in each chain are the imino acids proline and hydroxyproline. The folding of the triple helix in collagen has been studied by CD spectroscopy and enzymatic digestion (Engel & Prockop, 1991; Kielty et al., 1993). The initial stages of collagen folding involve a series of enzymatic posttranslational modifications that occur only on the unfolded chain, including the hydroxylation of Pro (in the Y position of the Gly-X-Y sequence) to Hyp. Three C-terminal propeptides of procollagen associate, and this is followed by a zippering of the triple helix from the C- to N-terminus.

Mutations in the collagen triple helix found in hereditary connective tissue diseases, resulting from a single substitution of Gly by another amino acid residue (Gly→X), are associated with abnormal folding (Engel & Prockop, 1991; Kuivaniemi et al., 1991; Byers, 1993). For instance, more than 150 different Gly substitutions in the chains of type I collagen have been associated with lethal and nonlethal cases of osteogenesis imperfecta, a disease where bones are brittle and fragile (Byers, 1993; Prockop & Kivirikko, 1995). Gly substitutions have also been identified in type III collagen and result in Ehlers Danlos syndrome type IV (EDS IV), where large arteries rupture (Prockop & Kivirikko, 1995). Defective folding results from these Gly substitutions and may be implicated in their etiology (Byers, 1993; Raghunath et al., 1994). Mutated collagens, with Gly→X substitutions, show increased posttranslational modifications, suggesting a decreased rate of folding since the modifications occur only on the unfolded chains (Byers, 1993). Direct observations have confirmed this delayed folding (Raghunath et al., 1994).

The defective rate of folding of triple helix formation in some connective tissue diseases makes the triple helix an attractive system for studying folding, and its slow folding makes it amenable to real time NMR studies. Here we describe NMR studies on a triple helical peptide, (POG)$_3$-ITGARGLAG(POG)$_4$, denoted T3−785. It has been known that peptides with Gly as every third residue and a high proportion of imino acids will adopt a triple helical conformation (Sakakibara et al., 1973; Heidemann & Roth, 1982; Long et al., 1992; Fields et al., 1993; Li et al., 1993; Anachi et al., 1995). The peptide T3−785 includes POG triplets at each end to provide stability and a central nine-residue sequence from human type III collagen which lacks imino acids. This sequence contains a Gly (position 790) which is substituted by a Ser residue in a case of EDS IV (Tromp et al., 1989). It also serves to model a recognition region, since it contains the unique trypsin cleavage site and is immediately adjacent to the collagenase cleavage site (Kadler, 1994). By selectively labeling T3−785 at specific positions with ^{15}N-enriched amino acids, and by applying 2D heteronuclear NMR, real time NMR folding studies are performed and the kinetics of folding of individual residues are monitored.

MATERIALS AND METHODS

Chemicals. ^{15}N-Enriched (99%) Gly and Ala were purchased from Cambridge Isotope Laboratory (Woburn, MA). Deuterated water (D_2O) and methanol (CH_3OD) were purchased from Aldrich Chemical Co. (St. Louis, MO).

Peptide Synthesis and Purification. The folding studies were conducted on a 30-mer triple helical peptide (POG)$_3$-ITGARGLAG(POG)$_4$ (Fan et al., 1993; Li et al., 1993). This peptide is denoted as T3−785 since residues ITGARGLAG-POG correspond to residues 785−796 in the human type III chain. Three selectively ^{15}N-labeled T3−785 peptides, namely, T3−785a, T3−785b, and T3−785c, were synthesized on an Applied Biosystems 430A peptide synthesizer using stepwise solid-phase procedures. The T3−785a peptide contains ^{15}N residues at Ala 13 and Gly 24; T3−785b contains ^{15}N residues at Gly 15, Leu 16, and Ala 17; and the T3−785c peptide contains ^{15}N residues at Gly 6 and Ala 13 residues, as shown in Table 1. Both T3−785a and T3−785c were acetylated at the N-terminus, and T3−785b was not acetylated. The completion of each coupling step was monitored by a Kaiser test, in which completed coupling was indicated colorimetrically. The peptides were purified by high-pressure liquid chromatography (HPLC) using a preparative Dynamax C-18 column. The molecular weight of each purified peptide was confirmed by mass spectrometry.

NMR Spectroscopy. NMR experiments were performed on a Varian Unityplus 500 MHz spectrometer. Two-dimensional data sets were processed on a Silicon Graphics workstation using the FELIX 2.05 software package (Biosym, Inc.). Sensitivity-enhanced single-quantum coherence spectroscopy (SE-HSQC) (Cavanagh et al., 1991; Palmer et al., 1991; Kay et al., 1992), heteronuclear single-quantum coherence spectroscopy (HSQC) (Davis et al., 1992), HMQC-NOESY (Fesik & Zuiderweg, 1988), and HMQC-TOCSY (Gronenborn et al., 1989) experiments were recorded at 10 and 45 °C using gradients for coherence pathway selection. The time domain data sets consisted of 2048 complex points in the t_2 dimension and 96 increments in the t_1 dimension. The spectral widths were 3 and 5 kHz in the ^{15}N and ^{1}H dimensions, respectively. The concentrations of the peptide solutions used in all NMR experiments were 4.5−18 mM with 10% D_2O/90% H_2O. The pH of the solutions were adjusted to the range of 2−2.5.

The NMR folding experiment was performed in a manner similar to that described by Kautz and Fox (1993). The sample was denatured outside of the NMR spectrometer by heating the sample at 50 °C for 10 min and then immersed in an ice/water bath for 9 s to reduce the temperature to 10 °C. The temperature of the sample in the ice/water bath was measured directly with a dual channel thermometer which was immersed in the sample. After the sample was cooled to 10 °C, the sample was placed into the NMR probe which was equilibrated to 10 °C. Data acquisition began immediately after the sample was placed in the probe. The total elapsed time from removal of the sample from the 50 °C water bath to the start of data acquisition was 28 s.

The equilibrated temperature of the sample inside the NMR probe was determined by using a 10% methanol/90% D_2O sample which had been subjected to the same protocol as described above. First, a temperature versus chemical shift calibration was performed on the methanol sample in order to have a plot of the relationship between methyl chemical shift and temperature. Then, the methanol sample was heated and cooled as described above, and the chemical shift was used to determine the temperature of the sample after insertion into the NMR probe. Four independent experiments on the methanol sample indicated that, im-

mediately after insertion into the NMR probe, the sample temperature was 10 ± 1 °C and reached 10 ± 0.1 °C within 30 s.

In some experiments where detailed analysis was not carried out, the NMR folding experiment was performed only in two steps. First, the sample was denatured outside of the NMR spectrometer by heating at 50 °C for 5 min. Second, the sample was placed into the NMR probe which had been equilibrated to 10 °C, and the sample was allowed to equilibrate for 4.0 min to reach 10 ± 1 °C based on the methanol sample. In both cases a series of 2D heteronuclear correlation spectra were acquired every 4 min for a period of 1−2 h, and kinetics of folding were monitored by measuring cross-peak heights and/or volumes as a function of time. The intensities of monomer and trimer peaks were normalized to the corresponding 10 °C equilibrium values.

Kinetic Analysis. The evaluation of kinetic order is based on the general equations (Laidler, 1987):

first order: $$\ln\left(\frac{[A]}{[A_0]}\right) = -kt \quad (1)$$

second order: $$\frac{1}{[A]} - \frac{1}{[A_0]} = kt \quad (2)$$

third order: $$\frac{1}{[A]^2} - \frac{1}{[A_0]^2} = kt \quad (3)$$

where $[A_0]$ is the initial concentration of monomer and [A] is the concentration of monomer at any given time. Linear least squares fits of plots of ln [A], $[A]^{-1}$, and $[A]^{-2}$ vs t were performed for all the residues that were monitored by NMR. In addition, the residuals for the first-, second-, and third-order fits were plotted for each residue. The residual is defined as

$$r_i = I_{\mathrm{obsd},i} - I_{\mathrm{calcd},i} \quad (4)$$

where $I_{\mathrm{obsd},i}$ is the observed intensity at time i, and $I_{\mathrm{calcd},i}$ is the corresponding point from the best fit equation.

RESULTS

NMR Assignments of the ^{15}N-Labeled Residues in T3−785 Peptides. To obtain the kinetics of folding of individual residues at different positions in the triple helix sequence, ^{15}N-labeled amino acid residues were incorporated into T3−785. Because of the repetitive nature of the sequence, rather than putting all of the ^{15}N-labeled residues within a single peptide, where individual resonances may not be resolved, three separate peptides were synthesized to look at the kinetics of folding of the termini and central regions. Peptides T3−785a, T3−785b, and T3−785c are listed in Table 1.

To monitor the kinetics of folding of individual residues by NMR, sequence specific assignments must be obtained. The proton resonances of the labeled residues in T3−785b were assigned in earlier NMR studies (Li et al., 1993). Because the peptides T3−785a and T3−785c are identical to T3−785b in all respects other than the position of the labeled residues, the labeled residues in the peptides T3-785a and T3−785c were assigned in a similar manner, using HSQC, HMQC-NOESY and HMQC-TOCSY experiments.

Table 1: Triple Helical Peptide Sequences Showing ^{15}N-Labeled Residues

Peptide	Sequence[a]
T3-785a	POGPOGPOGITG**A**RGLAGPOGPO**G**POGPOG
T3-785b	POGPOGPOGITGAR**GLA**GPOGPOGPOGPOG
T3-785c	POGPO**G**POGITG**A**RGLAGPOGPOGPOGPOG

[a] The ^{15}N labels are indicated with bold and underline.

The amide region of the heteronuclear correlation spectra for T3−785a at 10 °C and 45 °C is shown in panels A and B of Figure 1, respectively. In the heteronuclear correlation spectrum at 10 °C, the ^{15}N Ala 13 residue gives rise to four peaks, and the Gly 24 residue gives rise to two peaks. From our earlier NMR studies, we have seen that a mixture of single-stranded and triple helical molecules are present in solution at 10 °C for T3−785, while only single strands are present at 45 °C (Li et al., 1993). Heteronuclear correlation experiments on T3−785 at high temperature indicate that Ala 13 and Gly 24 each has one cross peak that corresponds to the monomer state of these two residues. For the low temperature spectrum of Ala 13 and Gly 24, one peak in the spectrum corresponds to this monomer form. For Ala 13, three additional peaks, which are derived from the trimer form, are observed. Three peaks arise because the non-equivalent environment surrounding each Ala gives rise to small differences in chemical shift for the Ala residues on the three different chains of the trimer. For Gly 24, there is only a single peak that appears for the trimer form because the repetitive sequence that surrounds the Gly results in a symmetric environment for the Gly on the three chains. On the basis of CD equilibrium melting curves, a larger monomer peak than expected is seen in solution by NMR (Fan et al., 1993; Li et al., 1993). The large monomer peak relative to the trimer peak results from the significantly faster relaxation properties of the trimer form. ^{15}N relaxation experiments performed on T3−785 indicate a factor of 10 difference in the T_2 relaxation of the trimer versus monomer (Fan et al., 1993; Li et al., 1993) which results in broadened lines for the trimer form, as well as weaker intensity for the HSQC cross peaks due to poorer transfer of magnetization.

Kinetics of Folding of T3−785 by NMR. Folding of individual residues in T3−785a was followed by monitoring the decrease in intensity of the ^{15}N-labeled monomer peaks (Figure 2A,B) and by monitoring the increase in intensity of ^{15}N-labeled trimer peaks (Figure 2C,D). For both the Ala and Gly residues, the trimer or folded peaks are resolved from the monomer peaks and can be monitored independently. During the folding process, the monomer intensity of Gly 24 disappears faster than the monomer intensity of Ala 13, and the trimer intensity of Gly 24 appears faster than the trimer intensity of Ala 13. These results can be seen by comparing the difference of the first data points in Figure 2A,B (monomer disappearance) and Figure 2C,D (trimer appearance). For example, at concentrations of 18 mM, and at 30 s after initiation of refolding, Ala 13 still has approximately 70% of the monomer intensity (Figure 2A) whereas Gly 24 only has 30% of the monomer intensity (Figure 2B). The disparity in initial amplitudes of the kinetic phases of Ala 13 versus Gly 24 indicates that the Gly residue is folding into the trimer form more quickly than the Ala residue.

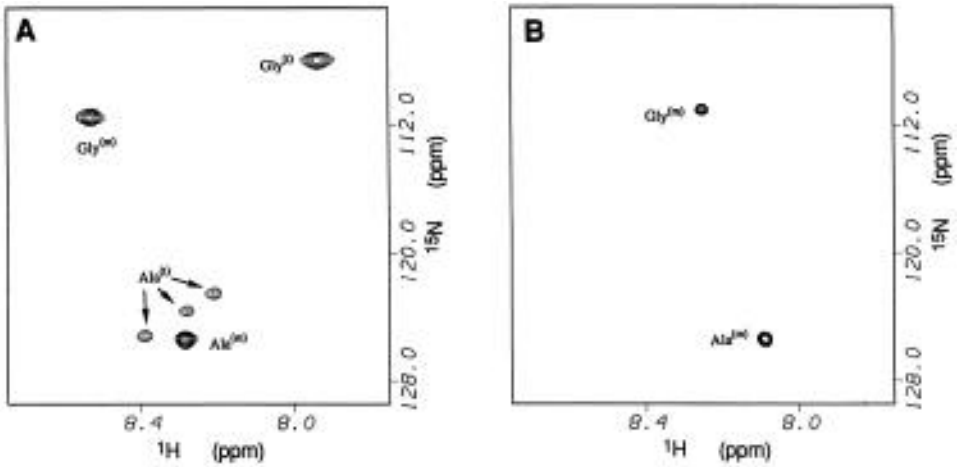

FIGURE 1: SE-HSQC spectra in 90% H_2O/10% D_2O (pH 2.0) at 10 °C (panel A) and 45 °C (panel B) for peptide T3−785a. The peaks corresponding to the monomer state of T3−785a are denoted with a superscript m, while those peaks corresponding to the triple helical state are denoted with a superscript t.

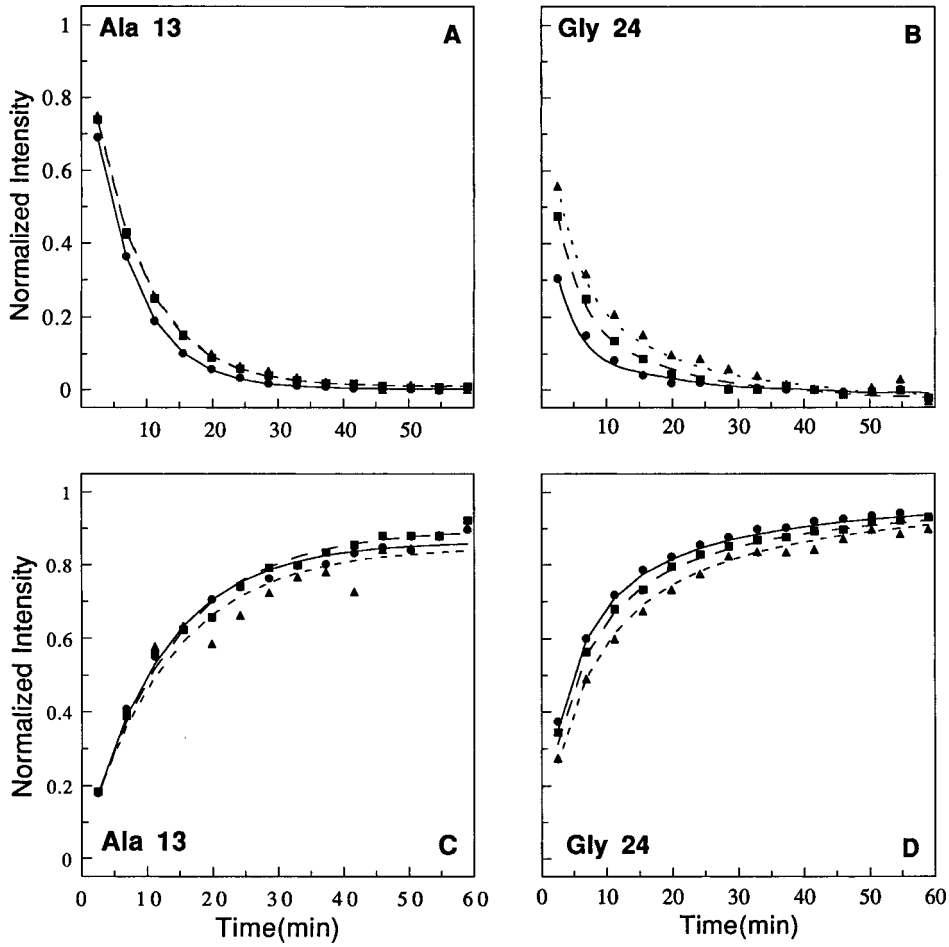

FIGURE 2: NMR folding profiles at three concentrations, 18 mM (●), 9 mM (■), and 4.5 mM (▲), for Ala 13 and Gly 24 of T3−785a (90% H_2O/10% D_2O at pH 2.0). The lines for Ala 13 are first-order fits to the data; the lines for Gly 24 are second-order fits. The disappearance of the monomer peaks for Ala 13 and Gly 24 is shown in panels A and B, respectively, and the appearance of the trimer peaks for these residues is shown in panels C and D. In panel A, the 9 and 4.5 mM concentrations are overlapped. The signal to noise for the appearance of the trimer intensity of Ala 13 (panel C) is inferior to the signal to noise for the appearance of the trimer of Gly 24 (panel D). This arises because Ala 13 is composed of three distinct and weak trimer peaks, whereas Gly 24 is composed of a single trimer peak (see Figure 1A).

To determine the kinetic orders of the Ala 13 and Gly 24 residues, two complementary approaches were used. The first approach is to fit the kinetic data for Ala 13 and Gly 24 to first-, second-, and third-order equations and to plot the residuals to assess quantitatively which order fits the data best. For Ala 13 (Figure 3A−C) the residuals are smallest for the first-order fit to the data. In the case of Gly 24 (Figure 3D−F) residuals are large for first-order kinetics, indicating that unlike Ala 13, the Gly is not following first-order kinetics. The residuals for second- and third-order kinetics are small for Gly 24, with the residuals for second order being slightly smaller than those of third order.

The second approach to determine the kinetic order is to perform the NMR folding experiments at different concen-

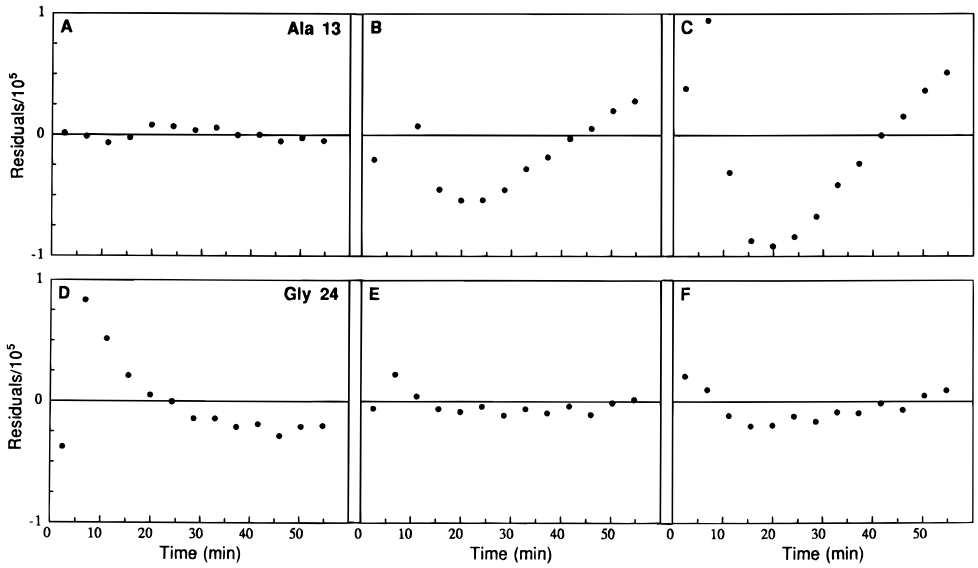

FIGURE 3: Residual plots for T3−785a (18 mM in 90% H_2O/10% D_2O at pH 2.0) indicating the differences between the experimental data and the best fit of the data to first-, second-, and third-order kinetics. The residual for point i is defined as follows: residual_i = observed intensity_i − calculated intensity_i, where the calculated intensity is based on the kinetic fit to the data. The residuals have been scaled by a factor of 10^5 for presentation. The residuals of the first-, second-, and third-order kinetics for the disappearance of the monomer peaks of Ala 13 are shown in panels A−C, respectively, while the residuals of the first-, second-, and third-order kinetics for the disappearance of the monomer peaks of Gly 24 are shown in panels D−F.

Table 2: T3−785a Rate Constants, as a Function of Concentration, Calculated for the Best Fit Kinetic Orders

	Ala 13	Gly 24	
concn (mM)	first order (s^{-1})	second order (M^{-1} s^{-1})	third order (M^{-2} s^{-1})
18	0.0024	0.76	140
9	0.0021	0.69	220
4.5	0.0021	0.99	770

trations and to calculate the rate constants of Ala 13 and Gly 24 at these different concentrations. These data are complementary to the residual plots shown above. Folding experiments on the T3−785a peptide were performed at three different concentrations: 18, 9, and 4.5 mM. Figure 2A,B shows the folding profile of the disappearance of the monomer peaks of Ala 13 and Gly 24 at these three concentrations, while Figure 2C,D shows the appearance of the trimer peaks of the same residues. Table 2 indicates the calculated rate constants, as a function of concentration, for Ala 13 and Gly 24 assuming different kinetic orders. It can be seen that the rate constants for Ala 13, assuming first-order kinetics, are all very similar to one another, confirming the data observed in the residual plots and indicating that Ala 13 fits first-order kinetics. The data for Gly 24 show that the rate constants, as a function of concentration, are more similar to one another for second-order kinetics than for third-order kinetics. This is similar to the observation that the residual plots are slightly smaller for second-order relative to third-order kinetics. Therefore, the data for Gly 24 appear to fit second-order kinetics slightly better than third-order but unlike Ala 13 do not fit first-order kinetics.

To determine whether folding can occur from the N-terminal as well as C-terminal end, the peptide T3−785 with the ^{15}N-labeled Gly, now incorporated at the N-terminal end (T3−785c), was studied. The kinetics of folding were monitored by following the loss of intensity of the monomer peaks as a function of time (Figure 4A) and show that Gly 6 at the N-terminal end folds faster than the labeled Ala 13 residue in the center of the peptide, much as seen for peptide T3−785a for the C-terminal end. Therefore, the peptides appear to fold from either the C- or N-terminal end.

To determine whether the rates of folding are independent of amino acid residue type, a peptide (T3−785b) with three sequentially labeled residues, Gly 15, Leu 16, and Ala 17, near the middle of the peptide, was studied (Figure 4B). The three sequentially labeled residues in the center of the peptide all fold at approximately the same rate, indicating that the rates of folding are independent of amino acid residue type. The rates of folding for the sequential residues in T3−785b are similar to the rate of folding of the central Ala 13 in peptides T3−785a and T3−785c.

DISCUSSION

The folding of collagen triple helices follows a mechanism distinct from those seen in globular proteins. While globular proteins involve the formation of early secondary structure with a hydrophobic core (Udgaonkar & Baldwin, 1995), the triple helix involves the formation of a relatively uniform helical conformation where three chains must associate and where the structure is stabilized by interchain hydrogen bonds (Ramachandran, 1967). The chains also contain a high proportion of imino acids, typically on the order of 20% (Brodsky & Shah, 1995). These features slow down the folding rate, such that the folding of triple helices occurs over a period of minutes rather than the time scale of milliseconds seen for globular proteins or coiled coil α helices.

The direct, real time NMR folding kinetics of residues labeled at specific positions in the triple helical peptide T3−785 allow us to consider a mechanism for folding of triple helical peptides. The kinetics of folding were monitored by measuring the decrease in the intensity of the monomer peak and the appearance of the trimer peaks of ^{15}N-labeled resonances. The residues at the terminal ends and central region could be followed directly and quantitated separately. The kinetics of folding for Ala 13, in the center of the chain,

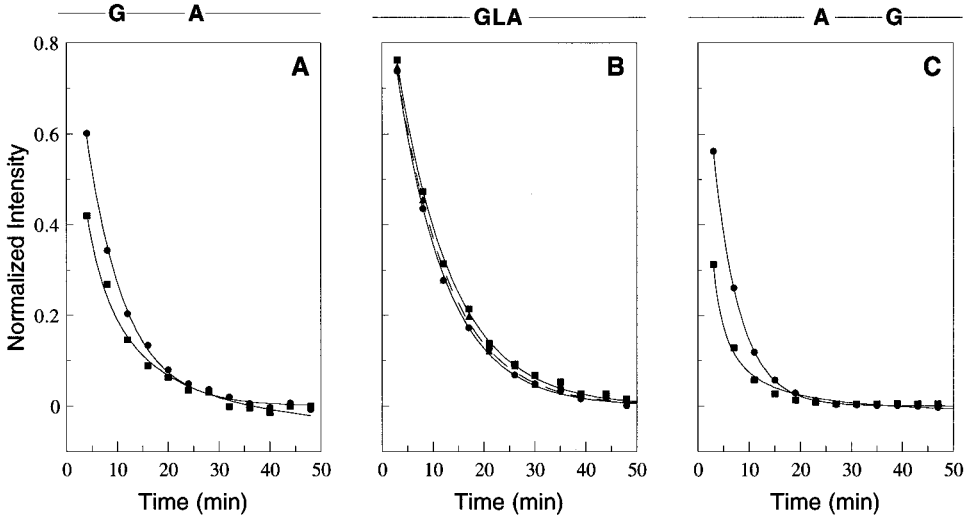

FIGURE 4: NMR folding profiles for three selectively ^{15}N-labeled T3−785 peptides: (A) T3−785c, (B) T3−785b, and (C) T3−785a (sequences are listed in Table 1). The folding data for Ala (●), Gly (■), and Leu (▲) are shown. The lines in panels A and C for the Gly residue are second-order fits to the data, while the rest of the lines are first-order fits to the data. The labeled positions are schematically shown on top of the corresponding panels.

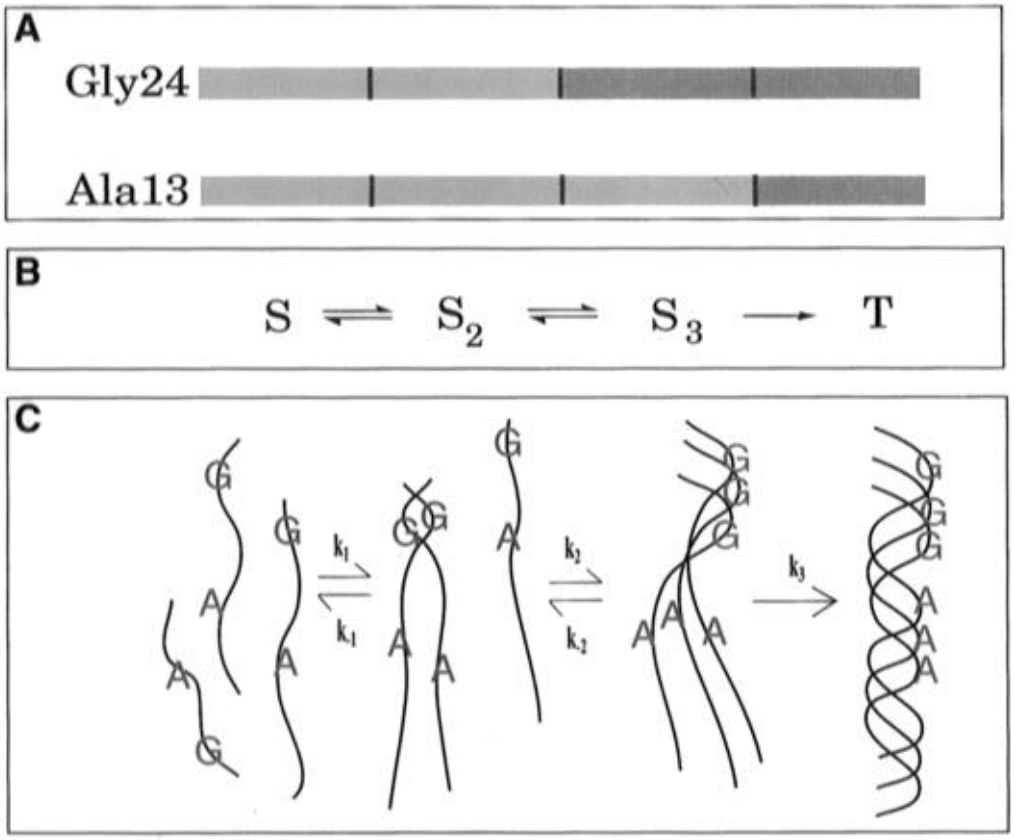

FIGURE 5: Proposed mechanism for the folding of T3−785. Panel A is a schematic representation of the presence of the trimer chemical shift for Ala 13 and Gly 24. The residues whose chemical shift is at the trimer frequency are shown in red, and the residues whose chemical shift is not at the trimer frequency are shown in blue. For example, in the S_3 state, Gly 24 is found at the trimer chemical shift and is indicated in red, but Ala 13 in this state is not in a trimeric environment and is shown in blue. Panel B shows the proposed mechanism, where S represents the monomer state; S_2, the association of two chains; S_3, the nucleated trimer; and T, the fully folded triple helix. Panel C is a cartoon representation of how the changes in chemical shift could arise from different conformational states. The residues being monitored are shown in the color corresponding to their chemical shift. Thus, both Gly and Ala are shown in blue in the S state, indicating their nonhelical nature. In the S_3 state, Gly is drawn in red to indicate its helical environment following nucleation, while Ala is still shown in blue, reflecting the nonhelical local conformation of the central residues. In the T state, both Gly and Ala are shown in red, indicating that both residues are now in the triple helical state.

are both slower and of a different kinetic order (first order) than for Gly 6 or Gly 24 (second order) at the N- and C-terminus, respectively. The different kinetic orders seen for this triple helical peptide suggest an ability to distinguish different steps in the folding mechanism. In addition, the population of protein with a monomer chemical shift decreases more slowly for Ala 13 than for Gly 24, and conversely, the population with trimer chemical shift increases more rapidly for Gly 24 than for Ala 13 as illustrated in Figure 5A. This lag implies that there is a point in the

folding process where the local conformation of Gly 24 is largely helical while the central Ala 13 is still in the unfolded state. This observation requires the postulation of kinetic intermediates for the folding of triple helical peptides.

The NMR data are consistent with a simple mechanism of association/nucleation and propagation as previously proposed for collagen, based on CD and enzyme susceptibility studies (Bächinger et al., 1980; Engel et al., 1980). A schematic three-step mechanism is shown in Figure 5B,C. In the first step, the monomer chains, labeled S, associate into dimers, labeled S_2. The nucleated trimer species with interchain hydrogen bonds, labeled S_3, are formed in the second step. In the schematic drawing, the trimer is nucleated at one end; however, the data indicate that it is likely that the nucleation can occur from either the C-terminal or the N-terminal end. The nucleated trimer species S_3 represents the required intermediate that comes from the disparity in initial amplitudes of the kinetic phases of Ala 13 versus Gly 24. Given the difference in triple helix content of Gly 24 relative to Ala 13, especially in the early folding times, some molecules must already be in the S_3 state, where the Gly has a trimer chemical shift and the Ala has a monomer chemical shift. The last step of the folding mechanism is a slow, first-order propagation step where the nucleated species propagates into a triple helix, labeled T. The rate constant of 0.002 s^{-1} for Ala 13 measured directly from the NMR data at 10 °C is consistent with cis/trans isomerization of imino acids at this temperature (Brandts et al., 1975; Bächinger et al., 1978). However, the folding of T3−785 is much slower than the folding of $(POG)_{10}$ (unpublished results), suggesting that the nine-residue replacement of three (POG) units by ITGARGLAG, which contains no imino acids, is likely to be slowing down the propagation stage. This may be because the conformation of the single-chain form in imino acid poor regions is not as favorable for rapid propagation as the conformation of the single-chain form of POG sequences.

The mechanism proposed in Figure 5 is the simplest possible mechanism that accounts for the data. However, the observation that the sum of the normalized monomer and trimer intensities is not constant throughout the experiment suggests that the actual mechanism is more complicated than proposed. This observation implies that some of the molecules are passing through a state with a chemical shift which is neither monomer nor trimer but that the population of this state, or states, is too low at any one time to be detected directly. This state could arise, for example, from a third chain associating incorrectly with the S_2 state. Alternatively, the three chains are out of register and unable to propagate.

The NMR studies on triple helical peptides illustrate the ability of this method to directly observe different steps in the folding mechanism. Previous spectroscopic and enzymatic digestion studies probing the folding of the collagen triple helix indicate a nucleation event at the C-terminus followed by propagation in a zipper-like fashion to the N-terminus, consistent with data reported here for small triple helical peptides (Engel et al., 1980; Engel & Prockop, 1991). CD monitoring of the folding of collagen-like peptides, such as $(Pro\text{-}Pro\text{-}Gly)_{10}$ (Sutoh & Noda, 1974) and the 36-residue cyanogen bromide fragment α1-CB2 (Piez & Sherman, 1970), shows third-order kinetics. While optical methods observe only the rate-limiting step, or some complex combination of several steps, NMR can be used to monitor the kinetics of folding of individual residues at different locations. In the NMR data, the association/nucleation step fits second-order kinetics best. These studies provide a basis for studying folding mutations in collagen triple helices and give a view into the folding of a uniform helical structure with interstrand hydrogen bonds, which may relate to the folding of β sheets in globular proteins.

ACKNOWLEDGMENT

We thank Professor Alan Goldman for helpful discussions.

REFERENCES

Adler, M., & Scheraga, H. A. (1988) *Biochemistry 27*, 2471−2480.
Akasaka, K., Naito, A., & Nakatani, H. (1991) *J. Biomol. NMR 1*, 65−70.
Anachi, R., Siegel, D. L., Baum, J., & Brodsky, B. (1995) *FEBS Lett. 368*, 551−555.
Bächinger, H. P., Bruckner, P., Timpl, R., & Engel, J. (1978) *Eur. J. Biochem. 90*, 605−613.
Bächinger, H. P., Bruckner, P., Timpl, R., Prockop, D., & Engel, J. (1980) *Eur. J. Biochem. 106*, 619−632.
Balbach, J., Forge, V., vanNuland, N. A. J., Winder, S. L., Hore, P. J., & Dobson, C.M. (1995) *Nat. Struct. Biol. 2*, 865−870.
Baldwin, R. L. (1993) *Curr. Opin. Struct. Biol. 3*, 84−91.
Bella, J., Eaton, M., Brodsky, B., & Berman, H. M. (1994) *Science 266*, 75−81.
Blum, A. D., Smallcombe, S. H., & Baldwin, R. L. (1978) *J. Mol. Biol. 118*, 305−315.
Brandts, J. F., Halvorson, H. R., & Brennan, M. (1975) *Biochemistry 14*, 4953−4963.
Brodsky, B., & Shah, N. (1995) *FASEB J. 9*, 1537−1546.
Byers, P. H. (1993) in *Connective Tissue and Its Heritable Disorders: Molecular, Genetic, and Medical Aspects* (Royce P. M., & Steinmann, B., Eds.) pp 351−407, Wiley-Liss, New York.
Cavanagh, J., Palmer, A. G., Wright, P. E., & Rance, M. (1991) *J. Magn. Reson. 91*, 429.
Davis, A. L., Keeler, J., Laue, E. D., & Moskau, D. (1992) *J. Magn. Reson. 98*, 207−216.
Dill, K. A., & Shortle, D. (1991) *Annu. Rev. Biochem. 60*, 795−825.
Dobson, C. M. (1994) *Curr. Opin. Struct. Biol. 4*, 636−640.
Dobson, C. M., Evans, P. A., & Radford, S. E. (1994) *Trends Biochem. Sci. 19*, 31−37.
Dyson, H. J., & Wright, P. E. (1993) *Curr. Opin. Struct. Biol. 3*, 60−65.
Elöve, G. A., Bhuyan, A. K., & Roder, H. (1994) *Biochemistry 33*, 6925−6935.
Engel, J., & Prockop, D. (1991) *Annu. Rev. Biophys. Biophys Chem. 20*, 137−152.
Engel, J., Bächinger, H. P., & Bruckner, P. (1980) in *Protein Folding* (Jaenicke, R., Ed.) pp 345−368, Elsevier, Amsterdam, The Netherlands.
Fan, P., Li, M.-H., Brodsky, B., & Baum, J. (1993) *Biochemistry 32*, 13299−13309.
Fesik, S. W., & Zuiderweg, E. R. P. (1988) *J. Magn. Reson. 78*, 588−593.
Fields, C. G., Lovdahl, C. M., Miles, A. J., Hagen, V. L. M., & Fields, G. B. (1993) *Biopolymers 33*, 1695−1707.
Frieden, C., Hoeltzli, S. D., & Ropson, I. J. (1993) *Protein Sci. 2*, 2007−2014.
Gronenborn, A. M., Bax, A., Wingfield, P., & Clore, G. M. (1989) *FEBS Lett. 243*, 93−98.
Heidemann, E., & Roth, W. (1982) *Adv. Polym. Sci. 43*, 143−203.
Hoppe, H. J., & Reid, K. B. M. (1994) *Protein Sci. 3*, 1143−1158.
Kadler, K. (1994) *Protein Profile 1*, 519−638.
Kautz, R. A., & Fox, R. O. (1993) *Protein Sci. 2*, 851−858.
Kay, L. E., Kiefer, P., & Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
Kiefhaber, T., Labhardt, A. M., & Baldwin, R. L. (1995) *Nature 375*, 513−515.

Kielty, C. M., Hopkinson, I., & Grant, M. E. (1993) in *Connective Tissue and its Heritable Disorders: Molecular, Genetic, and Medical Aspects* (Royce, P. M., & Steinmann, B., Eds.) pp 103−147, Wiley-Liss, New York.
Koide, S., Dyson, H. J., & Wright, P. E. (1993) *Biochemistry 32*, 12299−12310.
Kuivaniemi, H., Tromp, G., & Prockop, D. J. P. (1991) *FASEB J. 5*, 2052−2060.
Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 6*, 87−103.
Laidler, K. J. (1987) *Chemical Kinetics*, Harper & Row, New York.
Li, M.-H., Fan, P., Brodsky, B., & Baum, J. (1993) *Biochemistry 32*, 7377−7387.
Long, C. G., Li, M.-H., Baum, J., & Brodsky, B. (1992) *J. Mol. Biol. 225*, 1−4.
Palmer, A. G., Cavanagh, J., Wright, P. E., & Rance, M. (1991) *J. Magn. Reson. 93*, 151−170.
Piez, K. A., & Sherman, M. R. (1970) *Biochemistry 9*, 4134−4140.
Prockop, D. J., & Kivirikko, K. I. (1995) *Annu. Rev. Biochem. 64*, 403−434.
Ptitsyn, O. B. (1995) *Curr. Opin. Struct. Biol. 5*, 74−78.
Raghunath, M., Bruckner, P., & Steinmann, B. (1994) *J. Mol. Biol. 236*, 940−949.
Ramachandran, G. N. (1967) in *Treatise on Collagen* (Ramachandran, G. N., Ed.) pp 103−183, Academic Press, New York.
Rich, A., & Crick, F. H. (1961) *Nature 176*, 915−916.
Roder, H., Elöve, G. A., & Englander, S. W. (1988) *Nature 335*, 700−704.
Sakakibara, S., Inouye, K., Shudo, K., Kishida, Y., Kobayashi, Y., & Prockop, D. J. (1973) *Biochim. Biophys. Acta 303*, 198−202.
Shortle, D. (1993) *Curr. Opin. Struct. Biol. 3*, 66−74.
Sutoh, K., & Noda, H. (1974) *Biopolymers 13*, 2477−2488.
Tromp, G., Kuivaniemi, H., Shikata, H., & Prockop, D. J. (1989) *J. Biol. Chem. 264*, 1349−1352.
Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 694−699.
Udgaonkar, J. B., & Baldwin, R. L. (1995) *Biochemistry 34*, 4088−4096.
Woodward, C. K. (1994) *Curr. Opin. Struct. Biol. 4*, 112−116.
Wüthrich, K. (1994) *Curr. Opin. Struct. Biol. 4*, 93−99.

BI952270D