NMR and CD Spectroscopy Show that Imino Acid Restriction of the Unfolded State Leads to Efficient Folding†

Yujia Xu,‡,# Timothy Hyde,§,# Xin Wang,§ Manjiri Bhate,‡ Barbara Brodsky,*,‡ and Jean Baum*,§

Department of Biochemistry, Robert-Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854, and Department of Chemistry, Rutgers University, Piscataway, New Jersey 08854

Received January 6, 2003; Revised Manuscript Received May 15, 2003

ABSTRACT: Protein folding is determined by molecular features in the unfolded state, as well as the native folded structure. In the unfolded state, imino acids both restrict conformational space and present cis−trans isomerization barriers to folding. Because of its high proline and hydroxyproline content, the collagen triple-helix offers an opportunity to characterize the impact of imino acids on the unfolded state and folding kinetics. Here, NMR and CD spectroscopy are used to characterize the role of imino acids in a triple-helical peptide, T1−892, which contains an 18-residue sequence from type I collagen and a C-terminal $(Gly-Pro-Hyp)_4$ domain. The replacement of Pro or Hyp by an Ala in the $(Gly-Pro-Hyp)_4$ region significantly decreases the folding rate at low but not high concentrations, consistent with less efficient nucleation. To understand the molecular basis of the decreased folding rate, changes in the unfolded as well as the folded states of the peptides were characterized. While the trimer states of the peptides are all similar, NMR dynamics studies show monomers with all trans $(Gly-Pro-Hyp)_4$ are less flexible than monomers containing Pro → Ala or Hyp → Ala substitutions. Nucleation requires all trans bonds in the $(Gly-Pro-Hyp)_4$ domain and the constrained monomer state of the all trans nucleation domain in T1−892 increases its competency to initiate triple-helix formation and illustrates the impact of the unfolded state on folding kinetics.

In recent years, attempts have been made to relate the mechanism of protein folding to structural features in the unfolded state (*1*−*4*). For example, the retention of nativelike long-range contacts or persistent local interactions in the unfolded chains of globular proteins is thought to affect folding through the limitation of conformational space, creating a biased population distribution of unfolded conformers (*5*−*8*). Imino acids are unique in their impact on the unfolded state of proteins. The ring structure dramatically limits the Φ,Ψ conformational space and an unusually large population of cis isomers is present. The equilibrium between cis and trans isomers for X−Pro[1] bonds in the unfolded state mandates a slow cis−trans isomerization step for folding to the native state, which is often the rate-limiting step in protein folding (*9*, *10*).

The consequences of imino acids on structure and folding are accentuated in the collagen motif, where an imino acid content of >20% is required to stabilize the triple-helical structure (*11*). The triple-helix consists of three extended, left-handed polyproline II-like chains, supercoiled around each other (*12*, *13*). The ring structure of imino acids fixes the Φ angle to that of the polyproline II conformation, while dramatically restricting Ψ angles to two minima, one of which is that of the polyproline II conformation (*14*, *15*). As a result, the high imino acid content stabilizes the extended individual chains. The close packing of the three chains requires that every third residue, located near the central axis, be Gly. This generates a $(Gly-X-Y)_n$ repeating sequence pattern, where X is frequently occupied by proline (Pro) and Y is often occupied by hydroxyproline (Hyp), formed as a posttranslational modification of Pro. Hydroxyproline confers a greater stabilization on the triple-helix than Pro, and it has been suggested that this occurs through water-mediated hydrogen bonding of the Hyp (*13*, *16*−*18*) or through a stereoinductive effect (*19*−*21*).

One approach to clarifying the role of imino acids in the collagen triple helix is through the study of model peptides. Peptides of sufficient length with Gly as every third residue and a high imino acid content will form a stable triple-helical structure and serve as models for collagen folding (*22*−*24*). Fibril forming collagens nucleate folding at the C-terminus, followed by C- to N-terminal zipper like propagation, limited by cis−trans isomerization of imino acids (*11*, *25*). A similar directionality of folding has been observed for a triple-helical peptide, T1−892, which was designed to contain the $(Gly-Pro-Hyp)_4$ nucleation domain found at the C-terminus of type I collagen (*26*, *27*). To investigate the impact of imino acids on the triple-helix, peptide homologues of T1−892 were characterized in which a single Pro or Hyp residue in the $(Gly-Pro-Hyp)_4$ domain is replaced by an Ala. CD and NMR studies show the replacement of an imino acid by Ala

† This work was supported by grants from NIH (GM45302 to J.B. and GM60048 for B.B.) and the Children's Brittle Bone Foundation (to B.B.).

* To whom correspondence should be addressed. E-mail: baum@rutchem.rutgers.edu. Phone: (732) 445−5666; Fax: (732) 445−5312; brodsky@umdnj.edu, Phone: (732) 235−4048.

‡ University of Medicine and Dentistry of New Jersey.

§ Rutgers University.

# Authors are co-equal contributors.

[1] Abbreviations: CD, circular dichroism; both three letter codes and one letter codes have been used to denote common amino acids; hydroxyproline is denoted by Hyp or O.

10.1021/bi034006n CCC: $25.00
Published on Web 07/01/2003

does not affect the folded state. Imino acid replacement does affect the rate of folding at low concentrations, indicating that Pro and Hyp play a role in efficient nucleation. A restricted mobility seen for (Gly−Pro−Hyp)$_4$ but lost upon any Ala substitution is likely to relate to conformational restrictions in this imino acid rich domain which may facilitate the initiation of triple-helix folding.

MATERIALS AND METHODS

Sample Preparation. All peptides were synthesized by Synpep Corporation (Dublin, CA), and were purified and characterized as previously described (*28*).

Circular Dichroism Spectroscopy. Circular dichroism spectra were recorded on an Aviv Model 62DS spectrometer. Cells of 0.1-cm path length were used, and the temperature in the cell was controlled using a Hewlett-Packard Peltier thermoelectric temperature controller. Samples were prepared at varying concentrations, with peptides dried in vacuo over P_2O_5 for at least 48 h prior to weighing. The data were collected with 1-nm bandwidth and 2 s time averaging unless otherwise noted. The concentrations for CD experiments were determined using the absorbance at 214 nm with an extinction coefficient of 2×10^3 cm^{-1} M^{-1} per residue. Peptide T1−892 was also synthesized with a C-terminal Tyr for concentration determination; for this peptide an extinction coefficient of 1.2×10^3 M^{-1} cm^{-1} at 280 nm was utilized.

CD Equilibrium Experiments. The equilibrium melting curves were obtained by monitoring the change in mean residual ellipticity at 225 nm while the temperature is either increased from 0 to 60 °C, or decreased from 60 to 0 °C, with a temperature step of 1 to 4 °C to ensure a minimum of 10 data points in the melting transition zone. The samples were equilibrated for 4−9 h at each temperature until the signal reached a steady state, and the agreement of the thermal profiles obtained by heating and by cooling were used to ensure reversibility. The reversible melt experiment was carried out for each peptide at varying concentrations ranging from 0.14 to 2.1 mM. Due to the increased dynode voltage at 225 nm of samples at high concentrations, the temperature melt experiments of the three peptides at 2.1 mM were monitored at a wavelength of 227 nm, where the dynode voltage is relatively low (≤ 350 V).

The equilibrium melt curves are plotted as fraction folded F vs temperature, as defined in the previous work (*27*, *29*). The melting temperature, T_m, was defined as the temperature at which F reaches 0.5 during an equilibrium melt experiment, which also marks the midpoint of the transition. The transition of monomer to triple helix can be described by the reversible two-state model:

$$\mathrm{TH} \rightleftarrows 3\mathrm{M}$$

where TH is the triple-helical trimer state (the folded state) and M is the monomer state (the unfolded state). The equilibrium constant of such a transition at temperature T is defined as

$$K(T) = \frac{[\mathrm{M}]^3}{[\mathrm{TH}]} = \frac{3\mathrm{C_o}^2(1 - F(T))^3}{F(T)} \qquad (1)$$

where [M] and [TH] are the molar concentrations of the monomer and trimer, respectively, and C_o is the total molar concentration of the peptide (moles of monomer chain).

The melting curve expressed in fraction folded can be fit directly to the following equation to obtain the enthalpy ($\Delta H°$) and entropy ($\Delta S°$) of the transition, provided that the change of heat capacity during the transition is negligible (*30*):

$$\Delta H° - T\Delta S° = -RT \ln[3C_0^2(1 - F(T))^3/F(T)] \qquad (2)$$

The concentration dependence of the T_m for the trimerization process of triple helix formation can be expressed as

$$\frac{1}{T_m} = \frac{\Delta S^0 + R \ln 0.75}{\Delta H^0} + \left(\frac{R}{\Delta H^0}\right) \ln C_0^2 \qquad (3)$$

Thus, the change of enthalpy (ΔH^0) and entropy (ΔS^0) of the transition can be estimated from a linear plot of ($1/T_m$) vs ln C_0^2. This treatment provides an alternative method for analyzing the thermodynamic behavior of the peptides using the melting data at different concentrations, independent from the approach of eq 2. The linearity of the plot is also an indication of the two-state behavior of the monomer-to-trimer transition.

Curve fitting of CD data was conducted using the software Origin by Microcal except for the fit to eq 2, which was done using the software Scientist by microMath.

Folding Experiments. The CD folding experiments were carried out by monitoring the recovery of mean residual ellipticity after cooling from 70 to 5 °C at 225 nm (0.14−0.36 mM) or 227 nm (2.1 mM) as described previously (*27*). An automated cell changer was used to simultaneously monitor the folding of five samples at relatively low concentrations (0.14−0.36 mM). The dead time with the use of the cell changer is typically in the 40−100 s range depending on the position of the sample in the changer. The half time of refolding, $t_{1/2}$, was determined as the time for the fraction folded to reach 0.5. The third order folding rate and the percentage of competent monomer were estimated by fitting the folding data to a two-step folding model as described previously (*27*).

NMR Spectroscopy. Peptides were prepared as previously described (*26*). The pH of each sample was adjusted to 2.5 by addition of HCl. The peptides are very soluble even at high concentration and at acidic pH. CD experiments performed at acid and neutral pH indicate that the properties of the peptides are the same. A Varian INOVA spectrometer equipped with a triple resonance probe and pulsed field gradients operating at a proton frequency of 499.938 MHz was utilized. Two-dimensional data were processed on a Silicon Graphics workstation using the FELIX 97 software package (MSI, San Diego, CA). {^{1}H−^{15}N} heteronuclear single quantum coherence (HSQC) spectra were used for all experiments. {^{1}H−^{15}N} NOE relaxation experiments were performed at 0, 10, 20, 30, and 40 °C. Temperatures were verified by calibration with methanol (± 0.1 °C). For measurement of {^{1}H−^{15}N} NOE relaxation, spectra were collected in the presence and absence of ^{1}H saturation (*31*). Enhanced sensitivity pulsed field gradient versions of all experiments were used (*32*). The proton carrier frequency was set to the water resonance and the nitrogen carrier frequency was set to the center of the nitrogen chemical shift

Scheme 1[a]

```
                  1    5        10        15        20        25        30
T1-892(O24A): Ac-GPAGPAGPVGPAGARGPAGPOGPAGPOGPOGV-CONH2
T1-892:       Ac-GPAGPAGPVGPAGARGPAGPOGPOGPOGPOGY-CONH2
T1-892(P26A): Ac-GPAGPAGPVGPAGARGPAGPOGPOGAOGPOGV-CONH2
```

[a] The residues shown in bold italic mark the positions of the ^{15}N-labeled residues. The substitution sites, O24 → A and P26 → A, are boxed.

range. Spectral widths were 6000.2 Hz in the proton dimension and 1600 Hz in the nitrogen dimension for T1−892 and 1100 Hz in the nitrogen dimension for T1−892-(O24A) and T1−892(P26A). All collected data had 2048 complex data points in the t_2 dimension and 64 increments in the t_1 dimension.

The monomer resonances of Gly25 for all three peptides were assigned based on agreement of observed intensities with those calculated for populations of cis/trans isomers, as previously described (*26*). The percentage of cis isomers was taken from oligopeptide studies, assuming that X−Hyp is equivalent to X−Pro: 6% cis for Pro−Hyp, 13.7% cis for Gly−Pro, and 7.7% cis for Ala−Hyp bonds (*33*). Assuming isomerization events are independent, the resonances corresponding to Gly25 appeared to be influenced by the isomerization state of five surrounding X−imino acid bonds (*26*).

RESULTS

Peptide Design. An 18-residue sequence from the α1 chain of type I collagen (residues 892−909) forms the N-terminus of peptide T1−892. Our previous studies show that the imino acid rich (Gly−Pro−Hyp)$_4$ domain at the C-terminus of peptide T1−892 acts as a nucleation site to facilitate the association of three monomer chains into a triple-helix (*26*, *27*). To understand the role of Pro and Hyp in the initiation of triple-helix formation, two variants of T1−892 have been designed with Pro → Ala or Hyp → Ala substitutions in the (Gly−Pro−Hyp)$_4$ region, designated peptide T1−892(P26A) and T1−892(O24A), respectively (Scheme 1).

CD Spectroscopy: Spectra, Stability, and Thermodynamics. The two substituted peptides with imino acid substitutions Hyp24 → Ala and Pro26 → Ala assume the same triple helical structure as that of T1−892. The CD spectra of the three peptides at 5 °C are almost identical, showing the 225 nm maximum characteristic of the triple-helical conformation, with a mean residual ellipticity close to +4600 deg cm^2 dmol^{-1} (Figure 1).

All three peptides undergo a trimer−monomer transition as the temperature is increased, which can be monitored by the maximum at 225 nm. The equilibrium nature of the transition is indicated by the good agreement of the data obtained by cooling and by heating (Figure 1A). This equilibrium is reached only with prolonged equilibration time at each temperature. The midpoint of the thermal transition, T_m, for peptides T1−892(P26A) and T1−892(O24A) is at slightly lower temperatures (2 and 4 °C lower, respectively), compared to that of T1−892 at the same concentration. The equilibrium melting temperature of the three peptides is clearly dependent on peptide concentration (Figure 1B), as expected for a system undergoing a reversible unfolding from a trimeric folded state to a monomeric denatured state. Plots

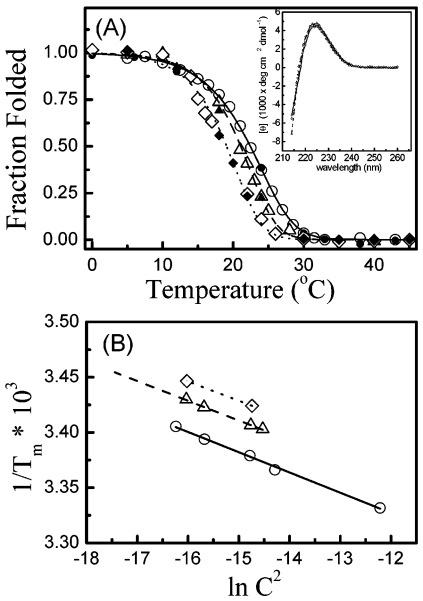


FIGURE 1: Equilibrium thermal unfolding of peptides T1−892 (circle, solid line), T1−892(P26A) (triangle, dashed line), and T1−892(O24A) (diamond, dotted line) monitored by CD. (A) Reversible temperature melt of samples at 0.62 mM in PBS buffer. The lines are the fit of experimental data collected by heating (open symbols) or cooling (solid symbols) using eq 2. Inset: Circular dichroism spectra of the three peptides (0.4 mM) at 5 °C. (B) Concentration dependence of the T_m obtained from the reversible melt of samples ranging from 0.14 to 2.1 mM of the three peptides. The lines represent the linear fit using eq 3. Because of the exceptionally long equilibration time necessary to reach thermal equilibrium for peptide T1−892(O24A), the reversible thermal melt profiles were obtained only at 0.36 and 0.62 mM.

Table 1: Equilibrium Parameters from Reversible Temperature Melt Experiments

peptide	T1−892	T1−892(P26A)	T1−892 (O24A)
thermal stability (T_m, °C)[a]	22.9	20.4	18.9
ΔH (kcal/mol)[b]	−104.1	−122.6	−125.4
ΔS (cal/mol)[b]	−322	−387	−399
ΔG (kcal/mol)[c], 5 °C	−14.6	−15.0	−14.4
$k_3 \times 10^4$(s^{-1} M^{-2})[d]	13.3	3.49	1.65
{M_{tr}}%[d]	41	49	47

[a] Extrapolated to 0.63 mM for comparison. [b] Results of fitting data in Figure 2a to eq 2. [c] Calculated from the ΔH and ΔS at the specified temperature, assuming the ΔC_p of thermal transition is negligible. [d] The folding data of the three peptides at 4 °C (0.32 mM) were fit to the unbranched, two-step folding model (Scheme 1 in ref *27*) to estimate the third-order rate constant (k_3) and the percentage of the competent monomer ({M_{tr}}%).

of $1/T_m$ vs ln C^2 for the three peptides show the linear nature expected for an equilibrium two-state trimer−monomer transition.

Both the thermal profiles and the linear $1/T_m$ vs ln C^2 plots are fit to a two-state model to obtain thermodynamic parameters for the trimer−monomer transition (Table 1). The good agreement of the estimations (data not shown for $1/T_m$ vs ln C^2) obtained from the two independent methods demonstrates the adequacy of the fitting model used. Thermodynamic parameters of the folding of T1−892(P26A) and T1−892(O24A) peptides show an unfavorable entropy loss that is largely compensated for by the more favorable enthalpy change, resulting in relatively small differences in free energy. The favorable enthalpy may result from ad-

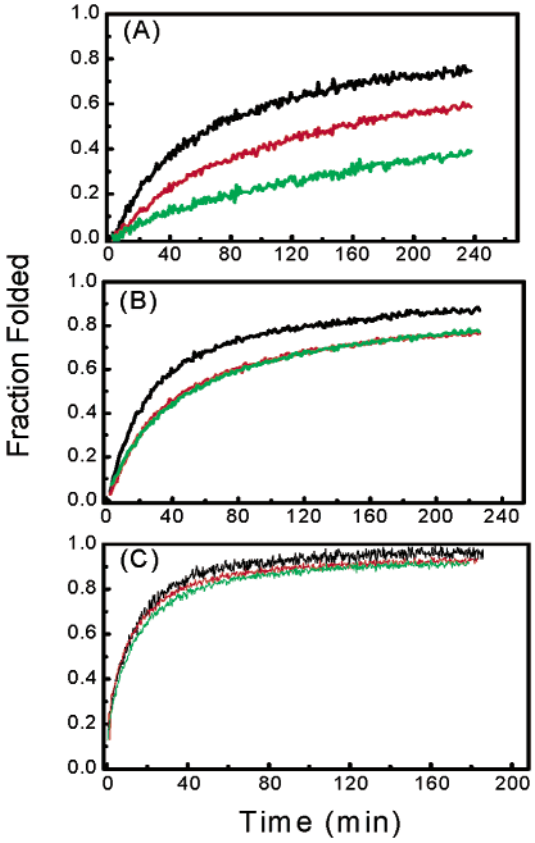
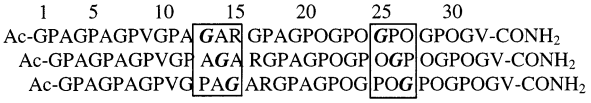


FIGURE 2: The folding profiles monitored by CD of T1−892 (black), T1−892(P26A) (red), and T1−892(O24A) (green) with concentrations of 0.14 (A), 0.32 (B), and 2.1 mM (C). The data were collected with a time constant of 10 s for the 0.32 and 0.14 mM samples, and with a time constant of 2 s for the 2.1 mM sample due to the faster folding rate at this concentration.

ditional hydrogen bonding capability to solvent gained from the introduction of an amide N−H group in Ala (*13*, *18*, *24*).

CD Spectroscopy: Folding. The folding kinetics of the three peptides at concentrations 0.14, 0.32, and 2.1 mM were monitored by CD at 5 °C, where the CD spectra and the ΔG values for the three peptides are very similar (Figure 1 and Table 1). Relative folding rates of different peptides at different concentrations can be compared qualitatively based on the empirical parameter $t_{1/2}$, the time at which the fraction folded equals 0.5. At low concentrations (0.14 mm), the $t_{1/2}$ of T1−892 is about 68 min, while that of peptide T1−892-(P26A) nearly doubled to 150 min. Under the same conditions, the folding of peptide T1−892(O24A) is even slower, with the fraction folded barely reaching 40% at the end of the data collection (240 min) (Figure 2). Thus, the replacement of a Hyp by Ala has a much more dramatic effect on folding at low concentrations than the replacement of a Pro. The difference in folding rates between the three peptides becomes less as the concentration is increased. At 0.32 mM, the rates of T1−892(P26A) and T1−892(O24A) are very similar ($t_{1/2}$ = 52 min for T1−892(O24A), $t_{1/2}$ = 48 min for T1−892(P26A)), and both are still slower than T1−892 ($t_{1/2}$ = 27 min). At the highest concentration (2.1 mM), the folding rates for the three peptides are identical within experimental error ($t_{1/2} \sim 8-10$ min; Figure 2). Thus, the folding rates for the Pro26 → Ala substituted peptide and the Hyp24 → Ala substituted peptide differ significantly at low concentration but not at higher concentrations.

Scheme 2



A more quantitative kinetic analysis was attempted, although the folding data cannot be fit to simple first, second, or third order kinetics. The folding reaction of T1−892 has been characterized by a two-step folding model (*27*). This model includes concentration-independent cis−trans isomerization of imino acids followed by a concentration-dependent third-order nucleation reaction. The differences in the folding rate observed for the three peptides at low, but not high, concentrations supports a difference in the third-order nucleation step which is likely to be rate limiting at low concentration. Application of this model to folding data at intermediate concentration (0.32 mM) indicates that the third-order rate constants of T1−892(P26A) and T1−892(O24A) peptides are decreased by nearly 4- and 10-fold, respectively, compared to that of T1−892 (Table 1). Estimations of the percentage of the competent monomer at the equilibrium unfolded state obtained from this fitting (41, 49, and 47% for T1−892, T1−892(P26A), and T1−892(O24A) peptides, respectively) are in good agreement with the values of 42, 49, and 46% calculated, assuming the requirement of all trans residues in the C-terminal nucleation domain (*26*, *27*). This good agreement indicates that the nucleation mechanism did not change significantly with Pro or Hyp to Ala substitution, even though the rate is reduced by these replacements.

NMR Spectroscopy: Assignments of Trimer and Monomer Resonances. NMR studies were carried out on the three peptides containing ^{15}N-labeled amino acids at specific positions along the peptides, as shown in Scheme 1. Resonance assignments were made for the HSQC spectrum of each peptide based on previous studies (*26*, *35*). In each spectrum, monomer and trimer resonances for Gly13 and Gly25 can be observed. The monomer and the trimer resonances of G13 were identified based on the nearly identical chemical shift of the peaks to those of T1−892. The remaining resonances were assigned to G25 residues, since this is the only other labeled residue in these two substituted peptides. The trimer and monomer resonance assignments were confirmed based on R_2 relaxation measurements (data not shown) which reflect the differential relaxation properties of the monomers and trimers.

(a) Trimer. The trimer resonances are characteristically at higher nitrogen field strength than the corresponding monomers and have lower intensity per mole due to different relaxation properties of the rigid rodlike trimer compared to the unfolded monomer. There is more than one Gly13 trimer resonance, since the staggering of chains by one residue in the GPAGAR sequence produces a different chemical environment for each of the three chains in the triple-helix (Scheme 2). In the repetitive sequence environment of (Gly−Pro−Hyp)$_4$, the resonances from the Gly25 residues of the T1−892 peptide are degenerate, giving only one broad trimer resonance (Scheme 2; Figure 3). The Pro26 → Ala and Hyp24 → Ala substitutions in T1−892(P26A) and T1−892-(O24A) peptides altered the chemical environment around

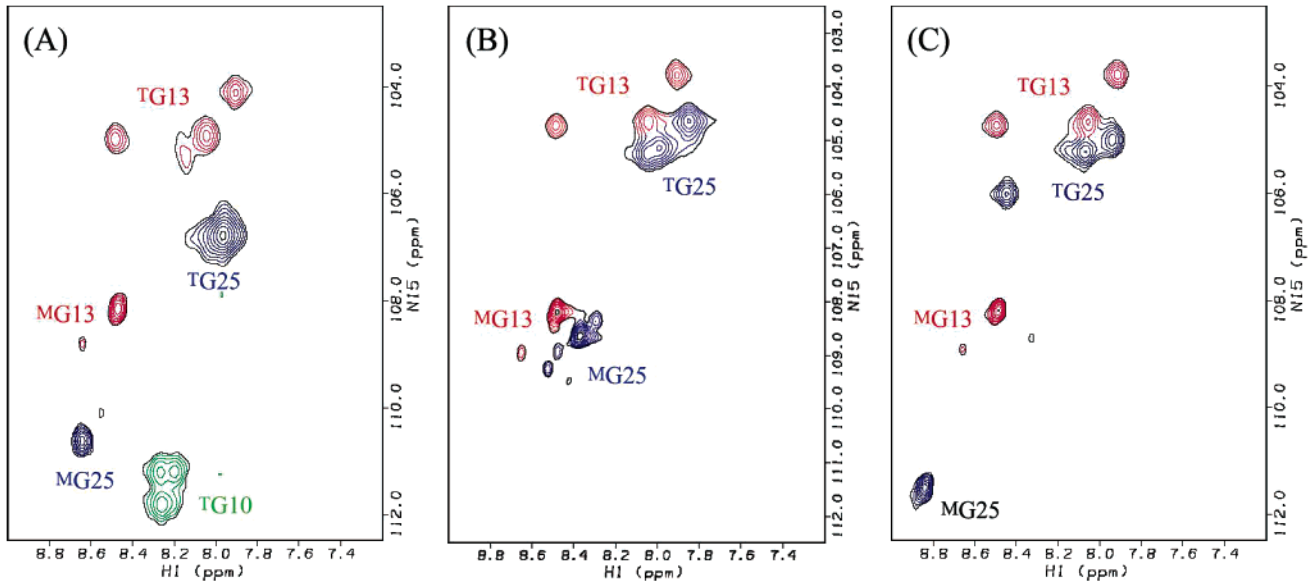


FIGURE 3: {^{1}H−^{15}N} HSQC spectra obtained at 0 °C of T1−892 (A), T1−892(O24A) (B), and T1−892(P26A) (C). The abbreviations ^{T}G13, ^{T}G25, ^{T}G10, ^{M}G13, and ^{M}G25 are for trimer Gly13, trimer Gly25, trimer Gly10, monomer Gly13, and monomer Gly25, respectively. T1−892 contains ^{15}N-labeled residues at positions Ala6, Gly10, Gly13, and Gly25, while peptides T1−892(O24A) and T1−892(P26A) are only labeled at positions Gly13 and Gly25.

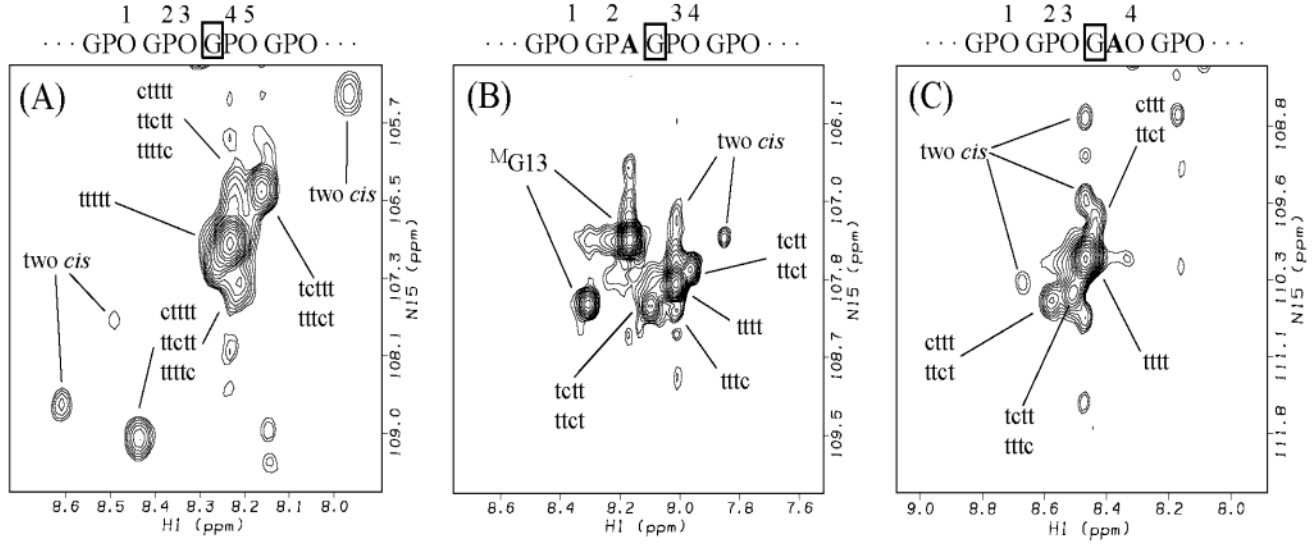


FIGURE 4: An expanded region of the {^{1}H−^{15}N} HSQC spectra obtained at 40 °C of T1−892 (A), T1−892(O24A) (B), and T1−892-(P26A) (C) showing only the monomer form of Gly25. The trans bonds are represented by "t" and cis bonds are represented by "c". The most intense resonance is the all trans conformer, and the less intense resonances indicate the existence of one or more cis bonds.

the Gly25 residues, resulting in multiple trimer resonances for Gly25 in these peptides (Figure 3). The trimer resonances for Gly13 in the substituted peptides have similar chemical shifts to those seen for the T1−892 peptide (Figure 3). In contrast, significant chemical shift changes are seen for the trimer resonances of Gly25, the residue adjacent to the Hyp24 → Ala and Pro26 → Ala substitutions.

(b) Monomer. The monomer forms of Gly13 and Gly25 show multiple resonances as a result of cis−trans isomerization of imino acids near the labeled glycine (*26*). The resonances seen for Gly13 in T1−892 (major peak ∼86%; minor peak ∼13%) reflect two interconverting conformers undergoing cis−trans isomerization about one Gly−Pro bond (*26*) and is unchanged in peptides T1−892(P26A) and T1−892(O24A). In contrast, the large number of cis−trans isomer resonances seen for Gly25 in T1−892 is different in the HSQC spectra of T1−892(O24A) and T1−892(P26A), showing the effects of the local amino acid sequence on the population of cis−trans conformational isomers (Figure 4).

In all peptides, the most intense resonance can be unambiguously assigned to the all trans isomer. The weaker monomer resonances of Gly25 for all three peptides were assigned based on agreement of observed intensities with those calculated for populations of cis−trans isomers (*26*). The resonance at Gly25 appears to be influenced by the isomerization state of the five surrounding X-imino acid bonds at Pro20−Hyp21, Gly22−Pro23, Pro23−Hyp24, Gly25−Pro26, and Pro26−Hyp27 (Table 2). Substitution of Pro26 by Ala in peptide T1−892(P26A) changes two isomerization bonds: Gly25−Pro26 → Gly25−Ala26, which eliminates the 14% cis at this bond, and Pro26−Hyp27 → Ala26−Hyp27, which increases slightly the population of

Table 2: Comparison of Experimental and Theoretical Intensities for Conformational Isomers of T1−892, T1−892(O24A), and T1−892(P26A) at 40 °C

peptide	exptl[a] value (%)	theor[b] values (%)	conformational state
T1−892	63.4	61.9	all trans (ttttt)
	11.9	9.8	one GP cis (ctttt or tttct)
	9.7	9.8	one GP cis (ctttt or tttct)
	4.6	3.9	one PO cis (tcttt, ttctt or ttttc)
	2.5	3.9	one PO cis (tcttt, ttctt or ttttc)
T1−892(O24A)	63.2	65.8	all trans (tttt)
	11.8	10.4	one GP cis (tctt or ttct)
	11.0	10.4	one GP cis (tctt or ttct)
	4.2	4.2	one PO cis (cttt or tttc)
T1−892(P26A)	74.9	70.4	all trans (tttt)
	12.4	11.2	one GP cis (tctt)
	5.0	5.9	one AO cis (tttc)
	4.8	4.5	one PO cis (cttt or ttct)

[a] Experimental values for individual monomer resonances of G25 were calculated as percentages of the total observed intensity. Conformations with more than one cis bond have not been monitored due to low signal-to-noise. [b] Relative amounts of cis−trans isomers were calculated as described in ref *26*, where the isomerization of each isolated imino bond is assumed to be independent of its neighbors and the populations of cis−trans isomers for each bond were obtained from ref *33*.

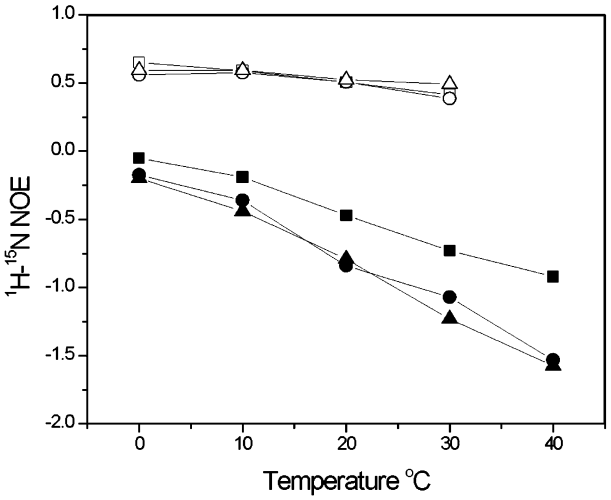


FIGURE 5: {^{1}H−^{15}N} NOE of T1−892 (squares), T1−892(P26A) (circles), and T1−892(O24A) (triangles) as a function of temperature. The trimer Gly25 NOE is shown with open symbols and the all trans Gly25 monomer NOE is shown with filled symbols.

cis-isomers (*33*). Substitution of Hyp24 by Ala in peptide T1−892(O24A) changes one bond involved in isomerization, replacing Pro23−Hyp24 by Pro23−Ala24, eliminating the 6% cis that was at this bond. The assignments are given in Table 2, together with the calculated and observed intensities. The good agreement between observed and calculated intensity indicates the five bonds whose isomerization state impacts Gly25 resonances have cis−trans populations that act independently of each other and have similar equilibrium properties as in oligopeptides.

NMR Spectroscopy: Dynamics. {^{1}H−^{15}N} NOE relaxation experiments can be used to assess the changes in dynamics resulting from the Hyp → Ala and Pro → Ala replacements in the (Gly−Pro−Hyp)$_4$ domain (Figure 5). Extrapolation of the CD results to the high NMR concentrations leads to the prediction of trimer-to-monomer transitions in the ∼30 °C range, allowing the characterization of dynamics in

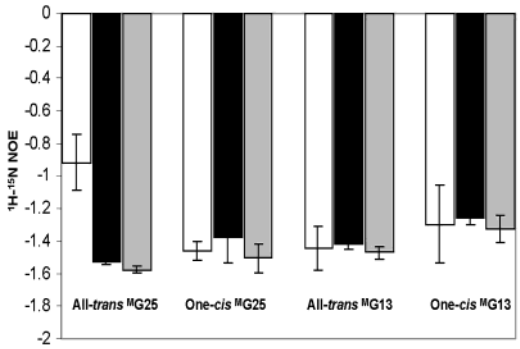


FIGURE 6: The {^{1}H−^{15}N} NOE at 40 °C of various cis−trans isomers of Gly25 and Gly13. The white bars represent T1−892, the dark gray bars represent T1−892(P26A), and the light gray bars represent T1−892(O24A).

different states by varying temperature. In the trimer at low temperature, the positive Gly25 NOEs indicate that the backbone dynamics of the three peptides are highly rigid and similar to one another. The small amount of trimer population that remains at 30 °C still has NOEs characteristic of a rigid molecule, showing little temperature dependence.

Analysis of the dynamics of monomer isomers was also carried out at varying temperatures. The monomer chains are more flexible than the trimers, as seen by the negative {^{1}H−^{15}N} NOE values for Gly13 and Gly25 in all isomers of the three peptides (Figures 5 and 6). Replacement of Hyp24 by Ala and Pro26 by Ala does not affect the NOE values of the all trans isomer of Gly13, which is distant from the replacement sites and which maintains a similar flexibility in all peptides (Figure 6). In contrast, the NOE values of the all trans isomer of Gly25, which is adjacent to the replacement sites, are affected by these amino acid substitutions. A high mobility is seen for the all trans isomer of Gly25 in peptides T1−892(O24A) and T1−892(P26A), while less flexibility is seen in peptide T1−892 (Figure 6). This distinction is more significant at higher temperatures (Figure 5). These results indicate that the repeating Gly−Pro−Hyp imino acid sequence environment of Gly25 in T1−892 constrains the mobility in the all trans monomer in the nucleation domain.

The dynamics of the well-defined resonances assigned to conformers with 1-cis bond were examined and showed a similar high degree of flexibility at all sites in the three peptides (Figure 6). For example, the NOE values seen for 1-cis Gly25 in all three peptides are close to the NOE values observed for all trans Gly25 in T1−892(O24A) and T1−892(P26A), and resemble the values seen for 1-cis Gly13. The dynamics of the 1-cis form was similar to that of the all trans form, except for the case of Gly25 in peptide T1−892, where the introduction of a cis bond resulted in a more negative NOE, indicating a release of the constraints conferred by the all trans (Gly−Pro−Hyp)$_4$ environment.

DISCUSSION

In the collagen model peptides studied here, the loss of a single imino acid decreases the folding rate of the triple-helix, and contrasts with globular proteins, where elimination of one Pro typically accelerates folding. To understand the basis of this decreased triple-helix folding rate, complementary NMR and CD studies were used to define changes in

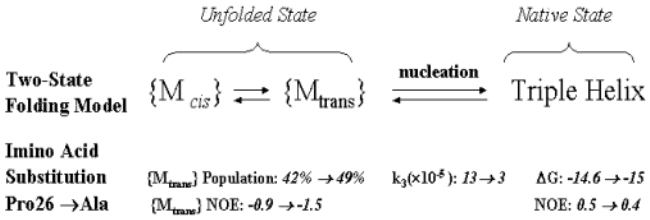


FIGURE 7: Schematic illustration of the effects of the loss of proline on the folding of the triple helix. The populations of $\{M_{cis}\}$ (represents monomers with at least one cis bond in the (Gly−Pro−Hyp)$_4$ nucleation region) and $\{M_{trans}\}$ (represents the population of monomers with all trans bonds in the (Gly−Pro−Hyp)$_4$) are calculated as previously described (*27*). The NOE values are taken from Figure 6. The third-order rate constant of nucleation reaction, k_3 (s^{-1} M^{-2}), and the free energy, ΔG (kcal/mol) at 5 °C, are taken from Table 1. The slower folding rate (k_3) in T1−892 with a Pro26 → Ala substitution arises from a balance of forces. The increased all trans conformation of the unfolded chains is favorable for folding, while the increased mobility of the unfolded state is unfavorable for folding.

the unfolded as well as native states of the peptide set. The triple helical peptides studied here allow direct observation of the unfolded as well as the folded state by NMR spectroscopy.

The folding mechanism of peptide T1−892 includes a concentration-dependent nucleation at the C-terminal (Gly−Pro−Hyp)$_4$ domain followed by a concentration-independent propagation to the N-terminus (*26*, *27*). A similar directionality of folding for the T1−892(O24A) and T1−892(P26A) peptides is suggested by real time NMR folding experiments (see Supporting Information). The observation that the peptides with Pro or Hyp replaced by Ala in the (Gly−Pro−Hyp)$_4$ domain slows down folding at low but not high concentrations indicates the nucleation step is affected by the imino acid replacements.

Folding models for T1−892 based on CD and NMR data include a heterogeneous population of monomers undergoing cis−trans isomerization followed by a third-order folding reaction from monomer to the triple-helix (Figure 7). Fitting results support a nucleation domain composed of all or most of the (Gly−Pro−Hyp)$_4$ sequence and indicates it must be in the all trans form before the monomer is competent to initiate triple-helix formation (*27*). Monomers that contain one or more cis bonds in this domain are not competent to form triple-helices. Given the equilibrium population of interconverting cis−trans isomers, only 42% of the monomer population has all trans bonds in the (Gly−Pro−Hyp)$_4$ domain of T1−892, and is therefore competent to nucleate. The T1−892(O24A) and T1−892(P26A) peptides, having lost one imino acid bond in this sequence, will have a greater population of all trans conformers in the nucleation domain. If the folding rate were only dependent on the percentage of all trans in the monomer (Gly−Pro−Hyp)$_4$ sequence, then T1−892(O24A) and T1−892(P26A) would fold faster than T1−892. The observation that the T1−892(O24A) and T1−892(P26A) fold slower at low concentrations suggests that the folding changes are not due to the population of competent monomers, but rather reflect changes in the third-order rate of association of competent monomers (Table 1).

Changes in the third order rate constant of the nucleation step could be due to perturbations in the native triple-helical structure. However, little change is observed in the triple-helix by CD and NMR spectroscopy as a result of the replacement of Hyp24 → Ala or Pro26 → Ala. All three peptides formed stable and rigid triple-helices. An increased nucleation rate could also be due to changes in the unfolded monomer state. For example, the constrained dihedral angle of the Pro and Hyp residues in the monomer are close to those found in native collagen, and the loss of even one Pro/Hyp could affect monomer dynamics or conformation. Such changes in the monomer but not the trimer state are supported by NMR data.

NMR studies showed significant differences between the dynamics of monomers with an all trans (Gly−Pro−Hyp)$_4$ domain, which are competent to proceed to triple-helix formation, and monomers containing one or more cis bonds, which require isomerization before folding. These results indicate that when Gly25 is surrounded by Gly−Pro−Hyp tripeptide units that are in the all trans conformation, the Gly25 residue is more constrained in the monomer form. The replacement of one imino acid by an Ala residue in the nucleation domain loosens the constraints in the competent monomer. The restricted mobility of Gly25 in the all trans (Gly−Pro−Hyp)$_4$ sequence may make it more effective at nucleation, and may be one of the factors that would explain the altered folding kinetics seen as a result of imino acid replacements.

Thermodynamic analysis is consistent with this hypothesis. The more constrained competent monomer state is consistent with the smaller and more favorable conformational entropy change accompanying the monomer−trimer transition in T1−892 compared with T1−892(O24A) and T1−892(P26A). In contrast, the favorable enthalpy of T1−892(O24A) and T1−892 (P26A) is likely to be due to additional H bonding capacity to solvent.

The peptide design makes it possible to consider the effect of Hyp versus Pro on nucleation. Hyp is known to accelerate triple helix folding, and at low concentrations the data show that the loss of Hyp (T1−892(O24A)) results in a folding rate slower than the loss of Pro (T1−892(P26A)). In this repetitive, helical system, it was possible to detect a change in the unfolded state arising from constraints imposed by multiple imino acids. These monomer constraints influence the initial nucleation step in triple-helix folding and thus affect the overall folding rate. In summary, the decrease in folding rate when a Pro or Hyp is replaced by Ala appears to be a balance of opposing events in the unfolded state: the increased percentage of unfolded monomers with an all trans nucleation domain is counteracted by decreased rigidity of the nucleation domain in the monomer form. An understanding of collagen triple-helix folding and its dependence on imino acid content may shed light on the abnormal folding found as a consequence of mutations in collagen in various connective tissue diseases.

ACKNOWLEDGMENT

We would like to thank Dr. Alexei Buevich for many helpful discussions.

SUPPORTING INFORMATION AVAILABLE

NMR folding curves (0 °C) monitored by the disappearance of monomer for T1-892 (A) and T1892P26A (B) peptide. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Shortle, D. R. (1996) *Curr. Opin. Struct. Biol. 6*, 24−30.
2. Dill, K. A., and Chan, H. S. (1997) *Nat. Struct. Biol. 4*, 10−19.
3. Dobson, C. M., Sali, A., and Karplus, M. (1998) *Angew. Chem., Int. Ed. 37*, 868−893.
4. Myers, J. K., and Oas, T. G. (2002) *Annu. Rev. Biochem. 71*, 783−815.
5. Zhang, O., Kay, L. E., Shortle, D., and Forman-Kay, J. D. (1997) *J. Mol. Biol. 272*, 9−20.
6. Yi, Q., Scalley-Kim, M. L., Alm, E. J., and Baker, D. (2000) *J. Mol. Biol. 299*, 1342−1351.
7. Shortle, D., and Ackerman, M. S. (2001) *Science 293*, 487−489.
8. Li, R., Battiste, J. L., and Woodward, C. (2002) *Biochemistry 41*, 2246−2253.
9. Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975) *Biochemistry 14*, 4953−4963.
10. Creighton, T. E. (1978) *J. Mol. Biol. 125*, 401−406.
11. Engel, J., and Prockop, D. J. (1991) *Annu. Rev. Biophys. Biophys. Chem. 20*, 137−152.
12. Rich, A., and Crick, F. H. C. (1961) *J. Mol. Biol. 3*, 483−506.
13. Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994) *Science 266*, 75−81.
14. Madison, V., and Schellman, J. (1970) *Biopolymmers 9*, 511−567.
15. Némethy, G., and Scheraga, H. A. (1982) *Biopolymers 21*, 1535−1555.
16. Privalov, P. L., Tictopulo, E. I., and Tischenko, V. M. (1979) *J. Mol. Biol. 127*, 203-216.
17. Privalov, P. L. (1982) *Adv. Protein Chem. 35*, 1−104.
18. Kramer, R. Z., Bella, J., Brodsky, B., and Berman, H. M. (2001) *J. Mol. Biol. 311*, 131−147.
19. Holmegren, S. K., Taylor, K. M., Bretscher, L. E., and Raines, R. T. (1998) *Nature 392*, 666−667.
20. Holmgren, S. K., Bretscher, L. E., Taylor, K. M., and Raines, R. T. (1999) *Chem. Biol. 6*, 63−70.
21. Bretscher, L. E., Jenkins, C. L., Taylor, K. M., DeRider, M. L., and Raines, R. T. (2001) *J. Am. Chem. Soc. 123*, 777−778.
22. Brodsky, B., and Ramshaw, J. A. M. (1997) *Matrix Biol. 15*, 545.
23. Baum, J., and Brodsky, B. (1999) *Curr. Opin. Struct. Biol. 9*, 122−128.
24. Baum, J., and Brodsky, B. (2000) *Mechanisms of Protein Folding*, 2nd ed. (Pain, R. H., Ed.) pp 330−351, Oxford University Press.
25. Bulleid, N. J., Wilson, R., and Lees, J. F. (1996) *Biochem. J. 317*, 195−202.
26. Buevich, A. V., Dai, Q.-H., Liu, X., Brodsky, B., and Baum, J. (2000) *Biochemistry 39*, 4299−4308.
27. Xu, Y., Bhate, M., and Brodsky, B. (2002) *Biochemistry 41*, 8143−8151.
28. Yang, W., Battineni, M. L., and Brodsky, B. (1997) *Biochemistry 36*, 6930−6935.
29. Persikov, A. V., Ramshaw, J. A. M., Kirkpatrick, A., and Brodsky, B. (2000) *Biochemistry 39*, 14960−14967.
30. Privalov, P. L., and Makhatadze, G. I. (1990) *J. Mol. Biol. 213*, 385−391.
31. Farrow, N. A., Muhandiram, R., Shinger, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.
32. Kay, L. E., Keifer, P., and Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
33. Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M., and Fischer, G. (1998) *J. Mol. Biol. 279*, 449−460.
34. Fan, P., Li, M.-H., Brodsky, B., and Baum, J. (1993) *Biochemistry 32*, 13299−13309.
35. Liu, X., Kim, S., Dai, Q.-H., Brodsky, B., and Baum, J. (1998) *Biochemistry 37*, 15528−15533.

BI034006N