Folding and Conformational Consequences of Glycine to Alanine Replacements at Different Positions in a Collagen Model Peptide†

Manjiri Bhate,‡,§ Xin Wang,§,|| Jean Baum,*,|| and Barbara Brodsky*,‡

Department of Biochemistry, UMDNJ−Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, New Jersey 08854, and Department of Chemistry, Rutgers University, Piscataway, New Jersey 08854

Received January 23, 2002; Revised Manuscript Received March 22, 2002

ABSTRACT: The collagen model peptide T1−892 includes a C-terminal nucleation domain, (Gly-Pro-Hyp)$_4$, and an N-terminal (Gly-X-Y)$_6$ sequence taken from type I collagen. In osteogenesis imperfecta (OI) and other collagen diseases, single base mutations often convert one Gly to a larger residue, and T1−892 homologues modeling such mutations were synthesized with Gly to Ala substitutions in either the (Gly-Pro-Hyp)$_4$ domain, Gly25Ala, or the (Gly-X-Y)$_6$ domain, Gly10Ala. CD and NMR studies show the Gly10Ala peptide forms a normal triple-helix at the C-terminal end and propagates from the C- to the N-terminus until the Gly → Ala substitution is encountered. At this point, triple-helix folding is terminated and cannot be reinitiated, leaving a nonhelical N-terminus. A decreased thermal stability is observed as a result of the shorter length of the triple-helix. In contrast, introduction of the Gly to Ala replacement at position 25, in the nucleation domain, shifts the monomer/trimer equilibrium toward the monomer form. The increased monomer and lower trimer populations are reflected in the dramatic decrease in triple-helix content and stability. Unlike the Ala replacement at position 10, the Ala substitution in the (Gly-Pro-Hyp)$_4$ region can still be incorporated into a triple-helix, but at a greatly decreased rate of folding, since the original efficient nucleation site is no longer operative. The specific consequences of Gly to Ala replacements in two distinctive sequences in this triple-helical peptide may help clarify the variability in OI clinical severity resulting from mutations at different sites along type I collagen chains.

Mutations in members of the collagen family have been implicated in an increasing number of hereditary connective tissue diseases (*1*−*8*). Osteogenesis imperfecta (OI)[1] results from mutations in type I collagen, the major collagen in bone, while some chondrodysplasias are caused by mutations in type II collagen, the major cartilage collagen. Other collagen diseases include Ehlers-Danlos Syndrome type IV (mutations in type III collagen), Alport syndrome (type IV collagen), Bethlem myopathy (type VI collagen), and dystrophic form of epidermolysis bullosa (type VII collagen). Observed defects include splicing errors, deletions, and nonsense codons, but the most frequent mutations are single base changes leading to the replacement of a required Gly residue in the collagen triple-helix. There is evidence that such collagen Gly substitutions fall in the class of human folding diseases (*6*, *9*−*13*). A better understanding of disease etiology may come from characterization of the effect of Gly replacements on the folding and conformation of the collagen triple-helix.

The collagen triple-helix constitutes the major motif in fibril-forming collagens (types I, II, III, V, XI), and also occurs as a domain in at least 15 types of nonfibrillar collagens and in noncollagenous proteins involved in host-defense (*14*, *15*). The molecular conformation of the collagen helix consists of three extended polyproline II-like chains supercoiled around a common axis (*16*−*18*). As a result of its unique supercoiled conformation, there are strict constraints on the amino acid sequence. The close packing of the three chains requires that every third residue in each chain be a Gly, producing the characteristic (Gly-X-Y)$_n$ amino acid sequence pattern. The three chains are staggered by one residue with respect to each other and have interchain peptide hydrogen bonds of the form NH (Gly)···OC (X). The extended polyproline II-like nature of the individual chains is stabilized by a high content of sterically restricted imino acids. Pro residues are incorporated into both X and Y positions in the growing polypeptide chain, but only those located in the Y position become posttranslationally modified to hydroxyproline (Hyp). The triplet Gly-Pro-Hyp is the most stabilizing tripeptide sequence and also the most frequent, constituting 10% of all triplets. Both Pro and Hyp promote triple-helix stability, but Hyp confers a greater stabilizing effect (*19*−*21*). The mechanism of the additional stabilization

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

* To whom correspondence should be addressed. B.B.: Department of Biochemistry, UMDNJ−Robert Wood Johnson Medical School, Piscataway, NJ 08854; (tel) 732-235-4048, (fax) 732-235-4783, (e-mail) brodsky@umdnj.edu. J.B.: Department of Chemistry, Rutgers University, Piscataway, NJ 08854; (tel) 732-445-5666, (fax) 732-445-5312, (e-mail) baum@rutchem.rutgers.edu.

‡ UMDNJ−Robert Wood Johnson Medical School.

§ These authors made equal contributions to this work.

|| Rutgers University.

[1] Abbreviations: CD, circular dichroism; Hyp, three-letter code for hydroxyproline; OI, osteogenesis imperfecta; MRE and [θ], mean residue ellipticity.

10.1021/bi020070d CCC: $22.00 © 2002 American Chemical Society
Published on Web 04/26/2002

by Hyp may relate to hydrogen bonding through a water network, its preferential ring conformation, or inductive effects (*18, 22*−*24*).

The folding of collagen is a multistep process, complicated by its multimer nature, the need for correct register of the repetitive Gly-X-Y sequence, and its high imino acid content. Type I collagen purified from tendon or skin can refold to a native triple-helix in vitro, but extensive time and critical temperature control are needed to promote the correct registration of the three chains (*25, 26*). In vivo, the molecules are synthesized in a procollagen form, with N- and C-terminal globular propeptides flanking the $(Gly\text{-}X\text{-}Y)_n$ central domain (*6, 27, 28*). Three globular C-propeptides self-associate and are responsible for chain selection and trimerization. Tethering at the C-terminal end facilitates the correct register of the three chains so that triple-helix nucleation can proceed (*29*). A distinctive $(Gly\text{-}Pro\text{-}Hyp)_4$ -Gly-Pro-Pro sequence found at the C-terminus of the triple-helix domain of the α1 and α2 chains of type I collagen and a related Hyp-rich terminal domain in type III collagen are likely to be involved in the formation of a stable triple-helix nucleus (*6, 29*). This repetitive sequence is unique in that no more than two contiguous Gly-Pro-Hyp units are found at any other site in the collagen chain. Following nucleation, the triple-helix conformation is propagated in a zipper-like mechanism from the C- to the N-terminus (*11, 30, 31*). In the unfolded equilibrium state, a significant proportion of imino acids (∼10%) have cis peptide bonds, and the propagation rate is limited by cis−trans isomerization to form the native collagen triple-helix with all-trans peptide bonds (*30, 31*).

Peptides with Gly as every third residue and a high content of Pro and Hyp adopt stable triple-helices and can serve as models for collagen (*12, 13, 32, 33*). Although the C-propeptide-mediated mechanism of trimerization found in collagen is absent, peptides of the form $(Gly\text{-}X\text{-}Y)_n$ can self-associate to form triple-helices when the length and imino acid content are sufficient. Peptide T1−892, a 30-mer, contains a C-terminal $(Gly\text{-}Pro\text{-}Hyp)_4$ sequence, modeling the nucleation domain of type I collagen, and an N-terminal sequence of residues 892−909 from the α1(I) chain of type I collagen (GPAGPAGPVGPAGARGPA). Previous NMR and CD studies have shown that peptide T1−892 adopts a stable collagen-like triple-helical conformation which is rigid along its full length (*34, 35*). Real time NMR studies have shown that the Gly-Pro-Hyp-rich C-terminal domain initiates folding in this peptide (*36*).

Single amino acid replacements in peptide T1−892 can be used to model mutations found in type I collagen in OI (*34, 35*). OI is characterized by fragile bones, with a wide range of clinical phenotypes (*1*−*3*). The disease results from mutations in either the α1(I) or the α2(I) chain of type I collagen, the major collagen in bone (*1, 2, 3, 37*). In different cases of OI, mutations of single base changes leading to a substitution for one required Gly residue are found at sites all along the triple-helix domain, including the $(Gly\text{-}Pro\text{-}Hyp)_4$ nucleation region at the C-terminus (*38*). There is direct evidence of delayed folding for a number of OI collagens (*10*), and indirect evidence of slower folding for all OI mutations (*3, 9, 11*).

Here, model peptides are used to compare the effects of Gly substitutions in the $(Gly\text{-}Pro\text{-}Hyp)_4$ nucleation domain with those in the propagation domain on stability, conformation, and folding. One peptide homologous to T1−892 was synthesized with a Gly to Ala replacement at a site corresponding to residue 901 in the α1(I) sequence, where a Gly substitution was observed to result in a mild case of OI (*39*). A second peptide homologous to T1−892 was synthesized with a Gly to Ala replacement in the middle of the $(Gly\text{-}Pro\text{-}Hyp)_4$ nucleation domain. The Gly → Ala replacement in the $(Gly\text{-}Pro\text{-}Hyp)_4$ domain generates a much greater decrease in stability and folding rate than the same replacement in the propagation domain. A progressive disruption of the triple-helix is seen for the peptide with the propagation mutation, while a shift in the equilibrium of trimer/monomer occurs for the nucleation mutation.

MATERIALS AND METHODS

Peptides. Peptides were synthesized by SynPep Corp (Dublin, CA) and purified on a C-18 column by HPLC, and their identity was confirmed by mass spectroscopy. The sequences of the three peptides studied are shown below, with positions of residues designated as 1−30:

```
                    5        10        15        20        25        30
T1-892          G P A G P A G P V G P A G A R G P A G P O G P O G P O G P O G V
T1-892[G10A]    G P A G P A G P V A P A G A R G P A G P O G P O G P O G P O G V
T1-892[G25A]    G P A G P A G P V G P A G A R G P A G P O G P O A P O G P O G V
```

For synthetic reasons, a Gly-Val sequence is included at the C-terminus of each peptide, and the N- and C-termini ends are blocked by acetylation and amidation, respectively, to enhance stability.

Circular Dichroism Spectroscopy. Circular dichroism spectra were recorded on an Aviv model 62DS spectrophotometer, using a Peltier thermoelectric temperature controller. All peptides were dried in vacuo over P_2O_5 for at least 48 h prior to weighing. Samples were dissolved in PBS (0.15 M NaCl, 0.01 M sodium phosphate, pH 7.1) and studied in cells of path length 0.1 cm. For wavelength scans, the signal was collected from 260 to 210 nm at 0.5 nm intervals at 0 °C.

For equilibrium melting experiments, samples were pre-equilibrated in the cold for more than 48 h prior to analysis, unless otherwise noted. The ellipticity at 225 nm was monitored as a function of temperature over an 18 h period, with an average heating rate of 0.15 °C/min and 2 min equilibration time. The fraction folded (F) was calculated as

$$F = (\theta_{\text{observed}} - \theta_{\text{monomer}})/(\theta_{\text{trimer}} - \theta_{\text{monomer}})$$

In this equation, θ_{observed} is the observed ellipticity, θ_{trimer} is calculated from the best-fit equation for the trimer part of the melting curve, and θ_{monomer} is calculated from the best-fit linear equation for the monomer part of the melting curve. Based on previous sedimentation equilibrium studies on similar peptides (*40*), the melting curves were fit to a two-state trimer to monomer transition, and good fits were obtained.

For CD folding experiments, samples were heated at 70 °C for 15 min, and then rapidly cooled in an ice/water bath kept at 0 to −2 °C, prior to placement in the CD machine equilibrated at the folding temperature. The ellipticity at 225 nm was monitored, with a time constant of 2 s and time

interval of 30 s. The time prior to recording of the first folding point is estimated to be 25 s. The folding data were fit to first-, second-, and third-order rate equations:

$$\ln([A]/[A_o]) = -kt \text{ (first order)}$$

$$1/[A] - 1/[A_o] = 2kt \text{ (second order)}$$

$$1/[A]^2 - 1/[A_o]^2 = 6kt \text{ (third order)}$$

where $[A_o]$ is the initial monomeric peptide concentration and [A], the concentration of monomer at any given time t, is calculated as $[A] = (1 - F)[A_o]$. The correlation coefficient and the residual plot were used to evaluate which rate equation best fit the data. The refolding half-time, $t_{1/2}$, is the time it takes for the fraction folded, F, to reach 50%, and is used as an empirical measure of folding rate.

NMR Spectroscopy. (A) NMR Experiments. All NMR experiments were carried out on a Varian INOVA-600 MHz spectrometer, using solutions of 9 mM in 10% D_2O/90% H_2O. The pH was adjusted to a value of 2.5 by titration with 5 mM HCl. The three peptides described above were synthesized with ^{15}N Gly and ^{15}N Ala residues at specific positions (see Figure 2). Heteronuclear single quantum coherence spectroscopy (HSQC) spectra were recorded at 0 °C. All the pulse sequences employed enhanced sensitivity pulsed field gradient techniques. The ^{15}N-{^{1}H} NOE relaxation measurements were performed in the presence and absence of ^{1}H saturation (*41*). Two-dimensional data sets were processed on a Silicon Graphics workstation using FELIX 97 (Molecular Simulation Inc.).

(B) Folding Experiments. The NMR folding experiments were performed as previously described (*36*). The sample was denatured outside of the NMR spectrometer by heating for 15 min at 50 °C and then immersed in an ice/water bath at 0 °C for 15 s to quickly reduce the temperature to 0 °C. A series of 2D ^{1}H-^{15}N HSQC spectra were acquired every 4 min at 0 °C. For each 2D ^{1}H-^{15}N HSQC spectrum, 32 t_1-increments were acquired with 4 scans per increment, and spectral widths were 6000 Hz (^{1}H) and 2000 Hz (^{15}N). The dead time of the first spectrum was 60 s. Kinetics of folding were monitored by measuring cross-peak volumes as a function of time. The folding experiment was repeated 3 times, and average intensities were used in the analysis. Normalization of monomer and trimer peaks was done as described elsewhere (*36*). Scaled data were normalized assuming that the initial intensity of monomer was equal to 1.

(C) Curve Fitting Procedure. The program *Excel* (Microsoft Inc.) was used to fit the decay of the monomer intensity to first-order rate constants.

RESULTS

Studies on the 30-mer peptide T1−892 have been previously reported (*34*−*36*). Here, two homologues of T1−892 with Gly → Ala substitutions at different locations are compared with the native peptide. The positions of residues in this peptide are designated as 1−30 (see Materials and Methods). One peptide, designated T1−892[G10A], has the Gly → Ala change at the 10th position in the peptide, corresponding to site 901 in the α1(I) sequence. In a folding context, this site is part of the propagation domain. The second peptide, T1−892[G25A], has the Gly → Ala change at the 25th position in the peptide, in the $(GPO)_4$ nucleation domain. Each peptide has ^{15}N residues incorporated at specific positions for heteronuclear NMR studies.

Circular Dichroism Spectroscopy Studies of Peptide Conformation, Stability, and Folding. Peptide T1−892 gives a characteristic triple-helical CD maximum at 225 nm, with mean residue ellipticity (MRE) = 3540 deg·cm^2·dmol^{-1}. The thermal profile of T1−892 at 1 mg/mL (0.36 mM) shows a sharp trimer to monomer transition, with $T_m = 26.5$ °C. For peptide T1−892[G10A], with a replacement of Gly at position 10 by Ala, the mean residue ellipticity at 225 nm is decreased to 2040 deg·cm^2·dmol^{-1}(Figure 1a; Table 1), and the thermal stability is reduced to 16 °C (Figure 1b). The replacement of the Gly by Ala at position 25 leads to more dramatic changes. After 48 h equilibration at 0 °C, the CD signal of peptide T1−892[G25A] at 225 nm is very low (MRE ∼ 800 deg·cm^2·dmol^{-1}) compared with the two homologous peptides. The melting temperature for T1−892[G25A] is 6 °C, considerably lower than T_m values for the other peptides.

The folding profiles of peptides with Gly → Ala replacements were compared with the parent peptide (Table 1, Figure 1c). Previously reported studies on T1−892 indicated slow folding rates, taking hours to reach completion (*34*, *36*). The CD folding rate depends strongly on concentration (Table 1), but a pure first-, second-, or third-order reaction is not consistent with the kinetics at all concentrations (data not shown). The folding of T1−892 [G10A] is comparable to that of T1−892 over the concentration range studied (1−5 mg/mL) (Table 1; Figure 1c). This contrasts with the much slower folding seen for peptide T1−892[G25A]. Using the time it takes to reach 0.5 fraction folded ($t_{1/2}$) as an empirical measure of the folding rate, $t_{1/2}$ = 221min for T1−892[G25A] at 1 mg/mL compared with $t_{1/2} = 27$ and 38 min for the other two peptides (Table 1; Figure 1c). The very slow folding rate of T1−892[G25A] fits third-order kinetics, suggesting the concentration-dependent initial reaction involving three chains is rate-limiting.

To investigate whether the low mean residue ellipticity observed for T1−892[G25A] was a result of incomplete folding, samples with higher concentrations were incubated at 0 °C for up to 2 months. Long incubation time at 0 °C for the 1 mg/mL solution led to an increased MRE at 225 nm, going from 800 to 1300 deg·cm^2·dmol^{-1} (data not shown). Higher concentrations also increased the $[\theta_{225}]$ value, going to 1740 deg·cm^2·dmol^{-1} at 2.5 mg/mL and 2120 deg·cm^2·dmol^{-1} at 5 mg/mL (Table 1). A trimer−monomer equilibrium is expected to show increases in MRE such as those observed, since trimer formation will be driven by higher concentration (*42*). Smaller increases in MRE were seen for higher concentrations of the more stable T1−892 and T1−892[G10A] peptides (Table 1).

Nuclear Magnetic Resonance Spectroscopy Studies of Peptide Conformation, Stability, and Folding. Peptide T1−892 was previously characterized by heteronuclear NMR, with ^{15}N-labeled residues at positions Ala6, Gly10, Gly13, and Gly25 (c = 9 mM; 25 mg/mL) (*35*, *36*). Trimer peaks, which disappear at temperatures higher than T_m and have high NOE values, as well as monomer peaks, which have low NOE values and are present at high and low tempera-

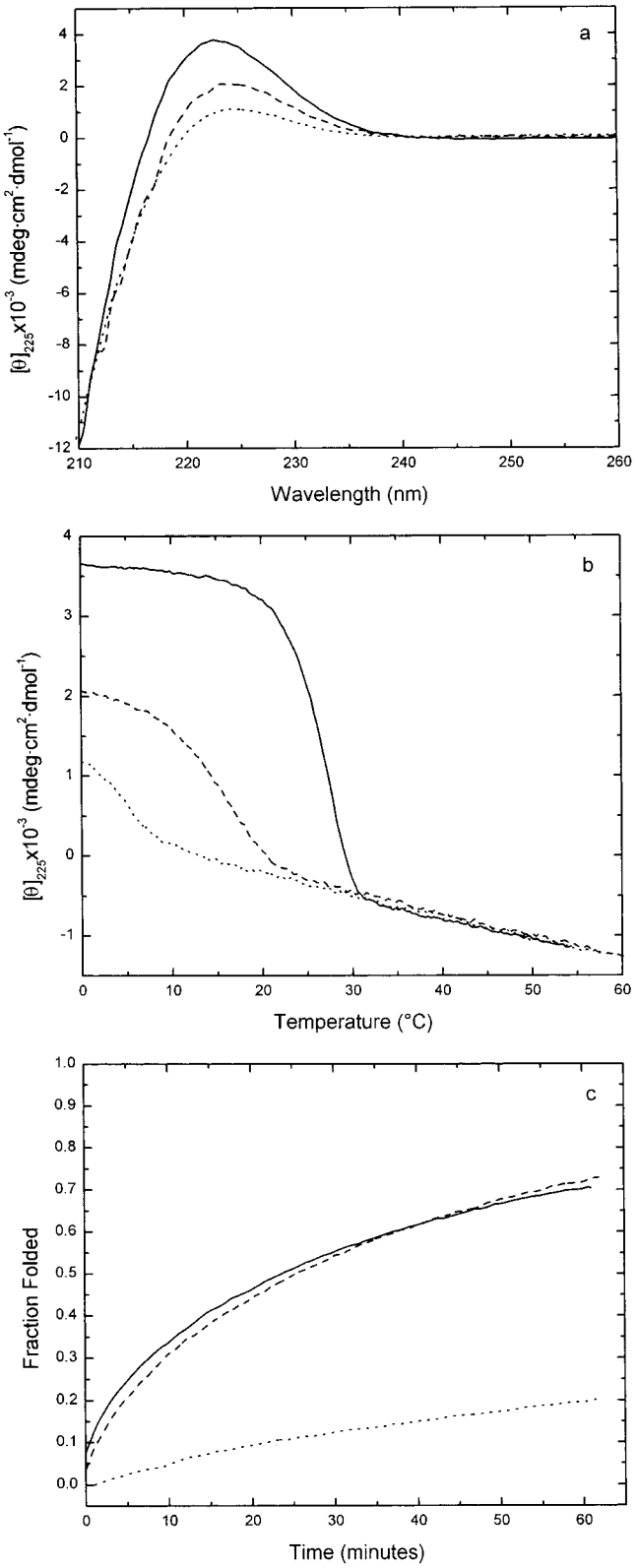

FIGURE 1: (a) Circular dichroism spectra of T1−892 (solid line), T1−892[G10A] (dashed line), and T1−892[G25A] (dotted line) at 0 °C, showing the characteristic triple-helix maximum near 225 nm. (b) Thermal transitions of T1−892 (solid line), T1−892[G10A] (dashed line), and T1−892[G25A] (dotted line), monitoring Θ_{225} as a function of temperature. (c) CD monitoring of folding at 0 °C (after quenching from 70 °C) for T1−892 (solid line), T1−892[G10A] (dashed line), and T1−892[G25A] (dotted line), showing fraction folded (calculated from Θ_{225} values) as a function of time. All samples were 1 mg/mL in PBS.

Table 1: Concentration Dependence of the Mean Residue Ellipticity at 225 nm and the Folding Rate, As Measured by $t_{1/2}$, for Peptide T1−892 and Its Homologues with Gly → Ala Substitutions at Positions 10 and 25

peptide	concn (mg/mL)	MRE_{225} (mdeg·cm^2·dmol^{-1})	$t_{1/2}$ (min)
T1−892	1.0	3540	27
	2.5	3610	18
	5.0	3750	15
T1−892[G10A]	1.0	2040	38
	2.5	2220	17
	5.0	2350	10
T1−892[G25A]	1.0	800	221
	2.5	1740	89
	5.0	2120	42

tures, could be assigned for each residue. These results support a rigid triple-helical conformation along the length of trimeric peptide T1−892, which is in equilibrium with some disordered monomer.

Introduction of a Gly to Ala replacement at position 10 results in a peptide whose trimer form is now only partially triple-helical. In the HSQC spectrum of T1−892 [G10A], the Gly25 residue shows trimer peaks with NOE values expected for high rigidity, while no trimer peaks are observed for Ala6 and Ala 10 (Figure 2). The Gly 13 residue has trimer peaks, but with less dispersion and lower NOE values than seen in T1−892. Thus, the NMR studies suggest that the Gly to Ala replacement at position 10 results in a peptide that is triple-helical C-terminal to the replacement site and not helical at and N-terminal to the replacement site.

No partially triple-helical forms are observed by NMR for peptide T1−892[G25A] (Figure 2). The Gly13 residue has trimer peaks with NOE values similar to those seen in T1−892, indicating its presence in triple-helical conformation. The Ala25 residue also shows trimer, as well as monomer, peaks. The presence of triple-helical peaks with large positive NOE values for both Ala25 and Gly13 suggests that a certain population of T1−892[G25A] is forming a rigid triple-helix even at the Gly → Ala replacement site. However, in contrast to the well-separated monomer and trimer peaks seen for Gly25 in T1−892 and T1−892[G10A], little chemical shift dispersion is seen for Ala25 monomer and trimer resonances. A number of Ala25 peaks disappear at high temperature and have the rigidity expected for a triple-helix, indicating they arose from the triple-helical conformation. But a number of peaks with similar chemical shifts remain at high temperature and show smaller NOE values, suggesting some monomer-like character.

NMR monitoring of specific residues at high concentrations shows the effect of a Gly substitution at different positions on the folding behavior (Figure 3). As reported previously, NMR studies on peptide T1−892 show a fast initial phase, where the monomer intensity of Gly25 drops to 50−60% of its original value, followed by a slower first-order cis−trans isomerization (*36*). The greater magnitude of the fast phase seen for Gly25 compared with Gly13 and Gly10 suggests nucleation at the C-terminus, followed by propagation along the molecule. For T1−892 [G10A], the Gly25 folding closely matches that observed for T1−892, supporting the similarity of their nucleation step, while Gly13 is folding more slowly both in the initial fast phase and in

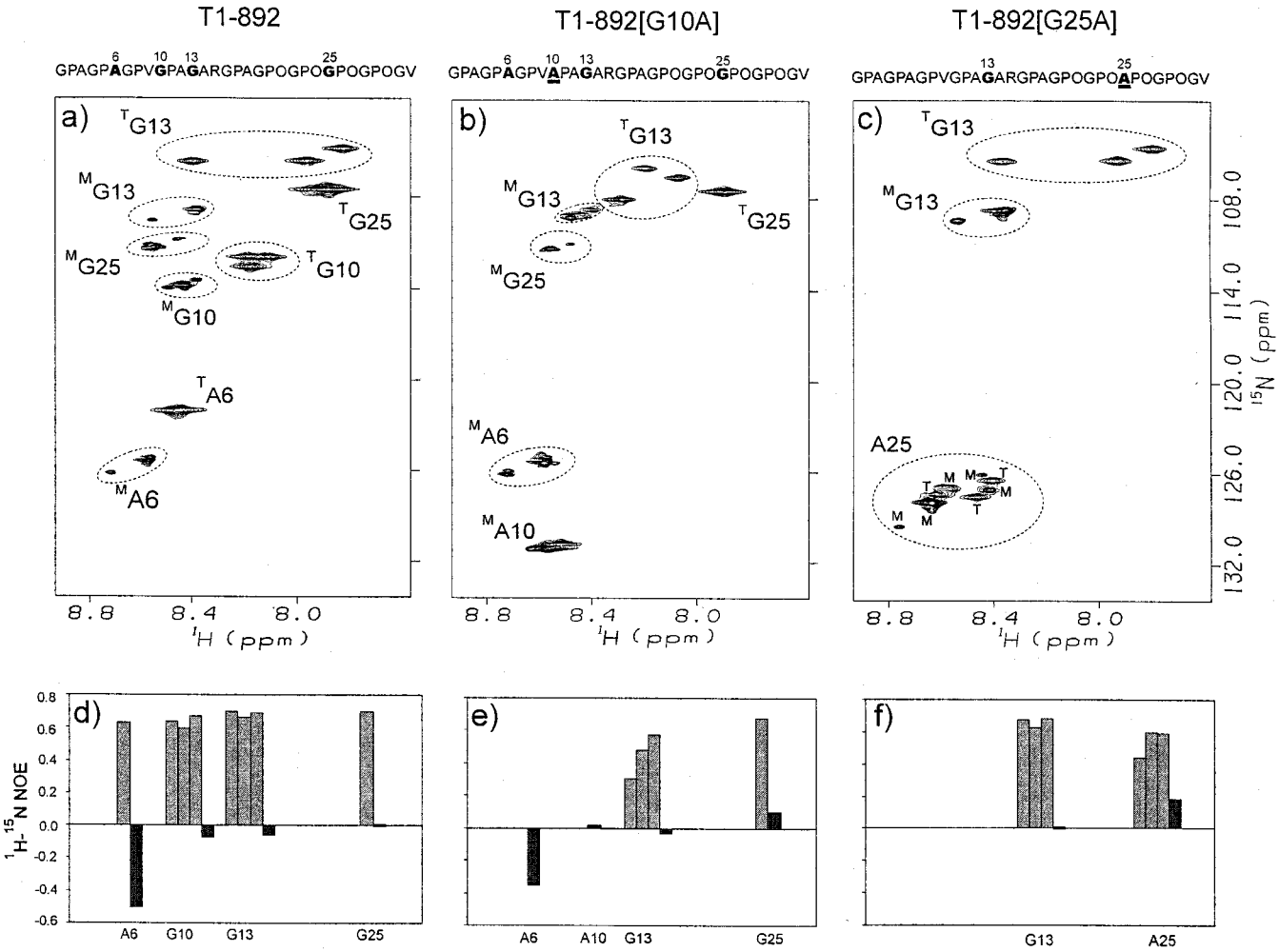

FIGURE 2: ^{1}H-^{15}N heteronuclear single quantum correlation spectrum (HSQC) of (a) T1−892, (b) T1−892 [G10A], and (c) T1−892[G25A]. All samples were $c = 9$ mM in $H_2O/D_2O = 9{:}1$, pH 2.5, 0 °C. The sequences for the peptides are shown at the top with the ^{15}N-labeled positions in boldface and the positions of the Gly-Ala substitution underlined. In the spectra, monomer and trimer peaks are denoted as M and T, respectively, at each position. Heteronuclear ^{15}N-{^{1}H} NOE experiments are shown below for labeled residues in peptides (d) T1−892, (e) T1−892[G10A], and (f) T1−892[G25A], for individual trimer peaks (gray) and monomer peaks (black).

the slower step. Ala6 and Ala10 do not fold at all. NMR monitoring of the folding of T1−892[G25A] shows slower folding of all residues. The fast phase of Ala25 only constitutes 15% of the total intensity and is similar to the value seen for the fast phase of Gly13. The failure of the residue at position 25 to fold before more N-terminal residues contrasts with the situation seen in T1−892 and T1−892[G10A] and suggests a different folding mechanism.

The equilibrium states of the peptides can be inferred from the end stage of the NMR folding experiments (Figure 3). In T1−892, folding reaches an equilibrium with 90%:10% trimer:monomer ratio for Gly25 and Gly13. The same 90%:10% trimer:monomer ratio is seen for residue Gly25 in peptide T1−892[G10A]. However, residue Gly13 in T1−892[G10A] gives a different equilibrium value of 75%:25% trimer:monomer, suggesting some opening and closing of the trimer just C-terminal to the substitution site. Ala10 and Ala6, located in the disrupted region at and N-terminal to the substitution site, remain 100% monomer. For T1−892[G25A], the equilibrium trimer:monomer ratio is ∼50:50 for Ala25 and ∼60:40 for Gly13. The presence of an Ala at position 25 has shifted the equilibrium toward a greater amount of monomer than seen in the other peptides.

DISCUSSION

Conformation: Imino Acid Sequence Context Determines Consequences of Gly → Ala Replacements. The complementarity of CD and NMR techniques allows definition of the different consequences for a Gly to Ala substitution at position 25 compared with position 10 in peptide T1−892. CD studies show the G10A replacement leads to a loss of about 50% of triple-helix content. NMR spectra of individual residues indicate this decreased triple-helical content results from a loss of helical conformation at and N-terminal to the Ala10 replacement site. The nine-residue sequence N-terminal to Ala10 is either insufficiently long or of a sequence such that reinitiation of the helix does not occur in these experiments after the C- to N-propagation is interrupted by the break in the Gly-X-Y repeating pattern. In addition, the residues just C-terminal to the substitution site, although still triple-helical, show signs of increased mobility and unwinding. The T1−892[G10A] molecules are also less stable than T1−892 (T_m = 16 °C compared with 26 °C). Peptide T1−892[G10A] has only 18−20 residues in a triple-helix compared with 30 residues for T1−892, and previous studies have shown stability is dependent on triple-helix length (*43*, *44*). Thus, this loss of helix N-terminal to the Ala10 is

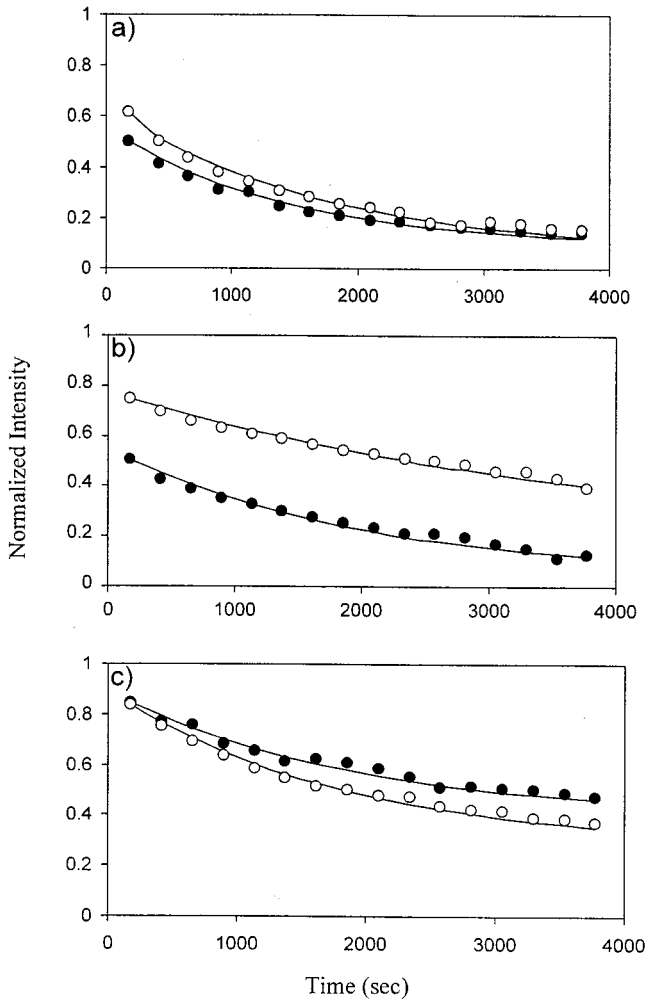

FIGURE 3: Real time folding kinetics of (a) T1−892, (b) T1−892[G10A], and (c) T1−892[G25A], depicted by the time-dependent change in intensity of the monomer peaks for Gly13 (○) and Gly25 (●) in (a) and (b) and for Ala25 (●) in (c). Resonance intensities were measured as volume integrals of the signals in the ^{1}H-^{15}N HSQC spectra recorded every 4 min after initiation of refolding by a rapid temperature quench from 50 to 0 °C. All curves represent the best fit of the second slow phase of the biphasic folding kinetics, starting from the second time interval. Folding is shown for the first 4000 s, and values obtained after overnight at 0 °C are considered to represent equilibrium values.

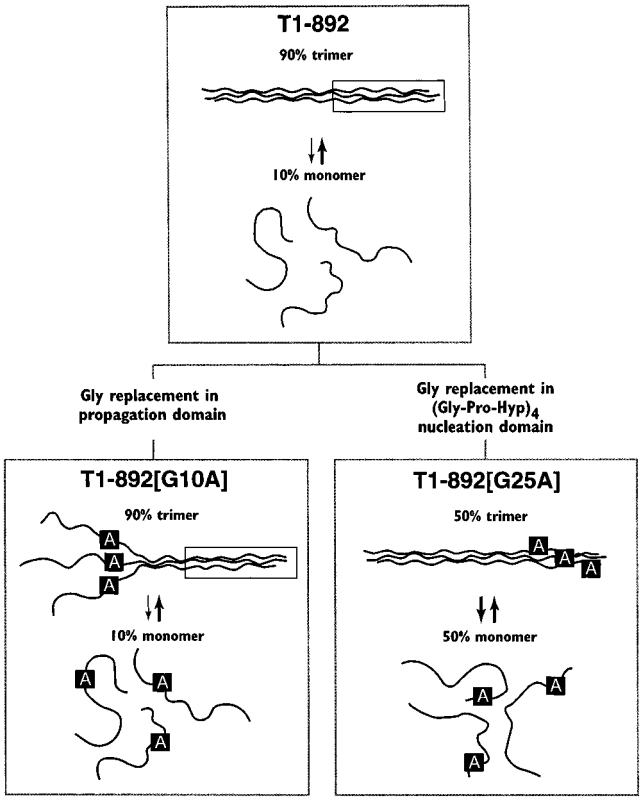

FIGURE 4: Schematic representation of the consequences of a Gly to Ala substitution in peptide T1−892 at position 10 and position 25. The boxed region of T1−892 and T1−892[G10A] molecules represents the $(Gly\text{-}Pro\text{-}Hyp)_4$ nucleation domain, which is no longer present in T1−892[G25A]. A major loss of triple-helix in each molecule can be seen as a result of the Gly10Ala substitution compared with the equilibrium shift seen with a Gly25Ala substitution.

consistent with the observed lower stability as well as its reduced 225 nm ellipticity (Figure 4).

The presence of a Gly to Ala substitution at position 25 in the $(Gly\text{-}Pro\text{-}Hyp)_4$ domain has consequences for the triple-helix very different from the substitution at position 10. CD studies show only a small fraction of the T1−892[G25A] peptide is in triple-helical form under standard conditions (1 mg/mL, 48 h equilibration time), rising to ∼50% at high concentrations and long incubation times at 0 °C. NMR studies indicate that the low triple-helix content in this case is not due to the formation of partially helical molecules, but rather to the presence of more monomers in equilibrium with less trimers under a given set of conditions. The shift in the trimer−monomer equilibrium constant is reflected in the low T_m value (∼6 °C) as well as the low ellipticity of this peptide.

Surprisingly, the trimer population of T1−892[G25A] maintains an ordered triple-helical conformation at the Ala25 site, despite the loss of Gly as every third residue and despite its proximity to the C-terminus of the peptide. The ability of repeating Gly-Pro-Hyp sequences to promote triple-helix formation strongly enough to fold around an interruption has been reported previously in the context of a $(Gly\text{-}Pro\text{-}Hyp)_8$ host peptide (*45*) and a $(Pro\text{-}Hyp\text{-}Gly)_{10}$ peptide (*18*, *40*). A high-resolution structure of the peptide $(Pro\text{-}Hyp\text{-}Gly)_4$-Pro-Hyp-Ala-$(Pro\text{-}Hyp\text{-}Gly)_5$ shows a triple-helix with local disruption of hydrogen bonding and untwisting at the Gly → Ala replacement site (*18*), while NMR studies on this peptide suggest the three Ala residues are nonequivalent and are undergoing hydrogen exchange faster than seen for Gly (*46*). Computational modeling of the effects of a Gly to Ala substitution using molecular dynamics showed perturbations consistent with those observed experimentally (*47*). Some related disarrangement of the triple-helix may be present in T1−892[G25A], but the lack of dispersion between the trimer and monomer peaks for Ala25 contrasts with the resolved Ala resonances in $(Pro\text{-}Hyp\text{-}Gly)_4$-Pro-Hyp-Ala-$(Pro\text{-}Hyp\text{-}Gly)_5$ (*47*) and suggests some additional effect, such as dynamic unfolding/folding from the nearby C-terminus.

Folding: Gly → Ala Replacement in Nucleation vs Propagation Domain. Peptide T1−892 was designed with $(Gly\text{-}Pro\text{-}Hyp)_4$ at its C-terminus to explore the special role of such sequences in folding. The unique $(Gly\text{-}Pro\text{-}Hyp)_4$ sites found at the C-terminus of α1 and α2 chains of type I collagen have been suggested to act as an initiation site for triple-helix folding because of the highly restricted nature

of adjacent imino acids (*48*). The magnitude and time course of the fast folding phase seen in NMR studies of different residues in T1−892 indicate folding starts at the C-terminus, with helix formation of the entire $(Gly\text{-}Pro\text{-}Hyp)_4$ domain (*36*). Relative to T1−892, introduction of a Gly → Ala substitution at position 10 "downstream" of the $(Gly\text{-}Pro\text{-}Hyp)_4$ domain has no influence on either the CD overall folding rate or the NMR folding rate of Gly25, implying nucleation is unaffected. However, the slower NMR folding rate seen for Gly13 and the absence of any triple-helix formation at Ala10 and Ala6 indicate the Gly10 → Ala substitution has an impact on propagation: C- to N-terminal propagation is slowed and then arrested when Ala10 is encountered. This is consistent with a two-step nucleation−propagation mechanism, where $(GPO)_4$ constitutes an isolated nucleation domain that is unaffected by perturbations to the subsequent propagation step.

CD studies show a Gly → Ala replacement at position 25 dramatically decreases the folding rate, while NMR gives insights about the nature of the defective folding. In peptide T1−892[G25A], the first point in the NMR folding curve, which indicates the fast phase, is the same for Ala25 and Gly13, in contrast to a greater magnitude of the fast phase for Gly25 compared with Gly13 for peptides T1−892 and T1−892[G10A]. This suggests that as a result of the Gly25Ala substitution in the nucleation domain, helix initiation is no longer occurring predominantly at the C-terminal (Gly-Pro-Hyp)-rich site. The smaller magnitude of the fast phase for both Ala25 and Gly13 (only ∼15%) supports a different and far less efficient mode of initiating folding. The existence of alternate folding initiation pathways not involving $(Gly\text{-}Pro\text{-}Hyp)_4$ is also supported by the ability of $(Pro\text{-}Pro\text{-}Gly)_{10}$ and the cyanogen bromide peptide α1CB-2 to form triple-helices, albeit with slow folding rates (*43*, *49*).

Comparison of Gly → Ala and Gly → Ser Replacements at the Same Site. The clinical severity of OI depends on the identity of the residues substituting for Gly at a given substitution site in the collagen chain (*3*). For example, replacement of Gly at position 883 of the α1(I) chain by Asp results in a lethal case of OI, while replacement of the same residue by Ser leads to a mild case of OI (*38*, *39*). The nature of the replacement amino acid was also shown to affect stability in a series of simple model peptides of the form $(Gly\text{-}Pro\text{-}Hyp)_3$- Z-Pro-Hyp-$(Gly\text{-}Pro\text{-}Hyp)_4$, where Z = Gly, Ala, Ser, Cys, Arg, Val, Asp, and Glu (*45*). All Gly substitutions were highly destabilizing, and, in the rank ordering, Ala and Ser were the least destabilizing residues.

Comparison of peptide T1−892[G10A] with the previously reported T1-892[G10S] (*34*−*36*) provides an opportunity to investigate the precise effect of the identity of the residue replacing Gly in the context of a known collagen mutation site. The effects of Ala and Ser replacements of Gly10 were very similar in terms of decreased stability (both T_m = 16−17 °C), folding kinetics, and termination of the triple-helix at and N-terminal to position 10, with comparable NMR chemical shifts and dynamics of homologous residues along the chain. Thus, although Ala is smaller than Ser, similar perturbations to conformation and folding were observed in this collagen context.

Of the several hundred individual OI Gly substitutions defined in the $(Gly\text{-}X\text{-}Y)_{338}$ triple-helix domain of the α1 and α2 chains of type I collagen, Gly → Ser changes are the most common and are found as lethal and nonlethal sites all along both α1 and α2 chains (*38*). In contrast, Gly → Ala mutations are rarely found and are often lethal. The peptide comparisons suggest differences in occurrence, distribution, and clinical severity of Ala vs Ser OI mutations are unlikely to be a result of differential effects on the triple-helix features, but more probably reflect the frequency of specific mutations at the DNA level, ascertainment problems in diagnosing mild forms of OI, or different consequences of Ser and Ala changes at the fibrillar level (*3*, *6*).

Peptides as Models for Mutant Collagens. Different mutations along the αl or α2 chains of type I collagen result in variable OI clinical phenotypes, ranging from perinatal lethal to a rather mild form with a high frequency of bone fractures (*3*). Factors that have been suggested to affect clinical phenotype include the chain in which the mutation occurs; the identity of the residue replacing Gly; the location of the mutation along the chain with respect to the C-terminus; the sequence environment of the mutation site; and the availability of potential renucleation sites immediately N-terminal to the site (*3*, *6*, *11*, *13*, *37*, *45*, *50*). It has been previously noted that mutations in imino acid rich domains tend to have the most severe clinical phenotype (*11*). The peptide models in this study illustrate the differing consequences of having a Gly → Ala replacement in the extreme imino acid rich $(Gly\text{-}Pro\text{-}Hyp)_4$ sequence compared with an α1(I) sequence having a high Gly-Pro-Ala content and no Hyp. In peptide T1−892[G10A], the triple-helix could not propagate through the substitution site, and may model the temporary cessation in propagation that results in overmodification of all OI collagens. However, when the Gly substitution occurs in the $(Gly\text{-}Pro\text{-}Hyp)_4$ domain, the triple-helix is present at and around the substitution site, but normal efficient nucleation is prevented and the trimer formation is highly disfavored. The diverse effects seen for the replacement of a Gly by an Ala in two distinctive regions of a collagen-like peptide indicate the sequences surrounding the substitution site affect multiple factors, including folding, propensity for triple-helix formation, and the ability to accommodate the larger replacement residue. It is likely that these factors all contribute to the variability in clinical severity seen for different OI mutation sites.

ACKNOWLEDGMENT

We thank Alexei Buevich for assistance with data acquisition and helpful discussions, Anton Persikov for assistance with figure preparation, and Yujia Xu for discussions.

REFERENCES

1. Myllyharju, J., and Kivirikko, K. I. (2001) Collagen and collagen-related diseases. *Ann. Med. 33*, 7−21.
2. Kuivaniemi, H., Tromp, G., Prockop, D. J., and Liams, C. (1997) Mutations in fibrillar collagens (types I, II, III and XI), fibril-associated collagen (type IX) and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels. *Hum. Mutat. 9*, 300−315.
3. Byers, P. H. (1993) Osteogenesis Imperfecta. in *Connective Tissue and its Hereditable Disorders: Molecular, Genetic and Medical Aspects* (Royce, P. M., and Steinmann, B., Eds.) pp 351−407, Wiley Liss, New York.

4. Bruckner-Tuderman, L. (1999) Hereditary skin diseases of anchoring fibrils. *J. Dermatol. Sci. 20*, 122−133.
5. Camacho Vanegas, O., Bertini, E., Zhang, R. Z., Petrini, S., Minosse, C., Sabatelli, P., Giusti, B., Chu, M. I., and Pepe, G. (2001) Ullrich scleroatonic muscular dystrophy is caused by recessive mutations in collagen type VI. *Proc. Natl. Acad. Sci. U.S.A. 98*, 7516−7521.
6. Byers, P. H. (2001) Folding defects in fibrillar collagens. *Philos. Trans. R. Soc. London, Ser. B 356*, 151−158.
7. Tiller, G. E., Polumbo, P. A., Weis, M. A., Bogaert, R., Lachman, R. S., Cohn, D. H., Rimoin, D. L., and Eyre, D. R. (1995) Dominant mutations in the type II collagen gene, COL2A1, produce spondyloepimetaphyseal dysplasia, Strudwick type. *Nat. Genet. 11*, 87−98.
8. Lemmink, H. H., Schroder, C. H., Monnens, L. A., and Smeets, H. J. (1997) The clinical spectrum of type IV collagen mutations. *Hum. Mutat. 9*, 477−499.
9. Bonadio, J., and Byers, P. H. (1985) Subtle structural alterations in the chains of type I procollagen produce osteogenesis imperfecta type II. *Nature 316*, 363−366.
10. Raghunath, M., Bruckner, P., and Steinmann, B. (1994) Delayed triple helix formation of mutant collagen from patients with osteogenesis imperfecta. *J. Mol. Biol. 236*, 940−949.
11. Bachinger, H. P., Morris, N. P., and Davis, J. M. (1993) Thermal stability and folding of the collagen triple-helix and the effects of mutations in osteogenesis imperfecta on the triple-helix of type I collagen. *Am. J. Med. Genet. 45*, 152−162.
12. Baum, J., and Brodsky, B. (1999) Folding of peptide models of collagen and misfolding in disease. *Curr. Opin. Struct. Biol. 9*, 122−128.
13. Baum, J., and Brodsky, B. (2000) Case study 2. Folding of the collagen triple-helix and its naturally occurring mutants. in *Mechanisms of Protein Folding* (Pain, R. H., Ed.) 2nd ed., pp 330−351, Oxford University Press, Oxford.
14. Kielty, C. M., Hopkinson, I., and Grant, M. E. (1993) The collagen family: structure, assembly and organization in the extracellular matrix. in *Connective Tissue and its Hereditable Disorders: Molecular, Genetic, and Medical Apsects* (Royce, P. M., and Steinmann, B., Eds.) pp 103−148, Wiley Liss, New York.
15. Hakansson, K., and Reid, K. B. (2000) Collectin structure: a review. *Protein Sci. 9*, 1607−1617.
16. Ramachandran, G. N. (1964) Structure of collagen at the molecular level. in *Treatise on Collagen* (Ramachandran, G. N., Ed.) Vol. 1, pp 103−183, Academic Press, New York.
17. Rich, A., and Crick, F. H. C. (1961) The molecular structure of collagen. *J. Mol. Biol. 3*, 483−506.
18. Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994) Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science 266*, 75−81.
19. Sakikabara, S., Inouye, K., Shudo, K., Kishida, Y., Kobayashi, Y., and Prockop, D. J. (1973) Synthesis of (Pro-Hyp-Gly)n of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxyproline. *Biochim. Biophys. Acta 303*, 198−202.
20. Berg, R. A., and Prockop, D. J. (1973) The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. *Biochem. Biophys. Res. Commun. 52*, 115−120.
21. Burjanadze, T. V. (2000) New analysis of the phylogenetic change of collagen thermostability. *Biopolymers 53*, 523−528.
22. Privalov, P. (1982) Stability of proteins. *Adv. Protein Chem. 35*, 1−104.
23. Bretscher, L. E., Jenkins, C. L., Taylor, K. M., DeRider, M. L., and Raines, R. T. (2001) Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc. 123*, 777−778.
24. Vitagliano, L., Berisio, R., Mastrangelo, A., Mazzarella, L., and Zagari, A. (2001) Preferred proline puckerings in cis and trans peptide groups: Implications for collagen stability. *Protein Sci. 10*, 2627−2632.
25. Beier, G., and Engel, J. (1966) The renaturation of soluble collagen. Products formed at different temperatures. *Biochemistry 5*, 2744−2755.
26. Harrington, W. F., and Karr, G. M. (1970) Collagen structure in solution. II. Analysis of refolding kinetics in terms of nucleation and growth processes. *Biochemistry 9*, 3725−3733.
27. McLaughlin, S. H., and Bulleid, N. J. (1998) Molecular recognition in procollagen chain assembly. *Matrix Biol. 16*, 369−377.
28. Lamande, S. R., and Bateman, J. F. (1999) Procollagen folding and assembly: The role of endoplasmic reticulum enzymes and molecular chaperones. *Cell Dev. Biol. 10*, 455−464.
29. Bulleid, N. J., Dalley, J. A., and Lees, J. F. (1997) The C-propeptide domain of procollagen can be replaced with a transmembrane domain without affecting trimer formation or collagen triple-helix folding during biosynthesis. *EMBO J. 16*, 6694−6701.
30. Bachinger, H. P., Bruckner, P., Timpl, R., and Engel, J. (1978) The role of cis−trans isomerization of peptide bonds in the coil-triple helix conversion of collagen. *Eur. J. Biochem. 90*, 605−613.
31. Bachinger, H. P., Bruckner, P., Timpl, R., Prockop, D. J., and Engel, J. (1980) Folding mechanism of the triple-helix in type III collagen and type III pN-collagen. Role of disulfide bridges and peptide bond isomerization. *Eur. J. Biochem. 106*, 619−632.
32. Fields, G. B., and Prockop, D. J. (1996) Perspectives on synthesis and applications of triple-helical collagen model peptides. *Biopolymers 40*, 345−357.
33. Goodman, M., Bhumralkar, M., Jefferson, E. A., Kwak, J., and Locardi, E. (1998) Collagen mimetics. *Biopolymers 47*, 127−142.
34. Yang, W., Battineni, M. L., and Brodsky, B. (1997) Amino acid sequence modulates the disruption by osteogenesis imperfecta glycine substitutions in collagen-like peptides. *Biochemistry 36*, 6930−6935.
35. Liu, X., Kim, S., Dai, Q.-H., Brodsky, B., and Baum, J. (1998) NMR shows asymmetric loss of triple-helix in peptides modeling a collagen mutation in brittle bone disease. *Biochemistry 37*, 15528−15533.
36. Buevich, A. V., Dai, Q.-H., Liu, X., Brodsky, B., and Baum, J. (2000) Site specific NMR monitoring of cis−trans isomerization in the folding of the proline-rich collagen triple-helix. *Biochemistry 39*, 4299−4308.
37. Marini, J. C., Lewis, M. B., Wang, Q., Chen, K. J., and Orrison, B. M. (1993) Serine for glycine substitutions in type I collagen in two cases of type IV osteogenesis imperfecta (OI). Additional evidence for a regional model of OI pathophysiology. *J. Biol. Chem. 268*, 2667−2673.
38. Dagleish, R. (2201) Database of human type I and type III collagen mutations. *http:///www.le.ac.uk/genetics/collagen/col1a1.html.*
39. Mottes, M., Sangalli, A., Valli, M., Gomez Lira, M., Tenni, R., Buttitta, P., Pignatti, P. F., and Detta, G. (1992) Mild dominant osteogenesis imperfecta with intrafamilial variability: the cause is a serine for glycine alpha 1(I) 901 substitution in a type-I collagen gene. *Hum. Genet. 89*, 480−484.
40. Long, C. G., Braswell, E., Zhu, D., Apigo, J., Baum, J., and Brodsky, B. (1993) Characterization of collagen-like peptides containing interruptions in the repeating Gly-X-Y sequence. *Biochemistry 32*, 11688−11695.
41. Farrow, N. A., Muhandiram, R., Singer, A., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry 33*, 5984−6003.
42. Boice, J. A., Dieckmann, G. R., DeGrado, W. F., and Fairman, R. (1996) Thermodynamic Analysis of a designed three-stranded coiled coil. *Biochemistry 35*, 14480−14485.
43. Sutoh, K., and Noda, H. (1974) Conformational Change of the Triple-Helical Structure. III. Stabilizing forces in the triple-helix. *Biopolymers 13*, 2461−2475.

44. Li, M.-H. (1993) NMR Studies of Triple-helical Peptides. Ph.D. Thesis, Rutgers University.
45. Beck, K., Chan, V. C., Shenoy, N., Kirkpatrick, A., Ramshaw, J. A. M., and Brodsky, B. (2000) Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *Proc. Natl. Acad. Sci. U.S.A. 93*, 4273−4278.
46. Buevich, A., and Baum, J. (2001) Nuclear magnetic resonance characterization of peptide models of collagen-folding diseases. *Philos. Trans. R. Soc. London, Ser. B 356*, 159−168.
47. Klein, T. E., and Huang, C. C. (1999) Computational investigations of structural changes resulting from point mutations in a collagen-like peptide. *Biopolymers 49*, 167−183.
48. Josse, J., and Harrington, W. F. (1964) Role of pyrrolidine residues in the structure and stabilization of collagen. *J. Mol. Biol. 9*, 269−287.
49. Piez, K. A., and Sherman, M. R. (1970) Equilibrium and kinetic studies of the helix-coil transition in α1-CB2, a small peptide from collagen. *Biochemistry 9*, 4134−4140.
50. Wenstrup, R. J., Shrago-Howe, A. W., Lever, L. W., Phillips, C. L., Byers, P. H., and Cohn, D. H. (1991) The effects of different cysteine for glycine substitutions within alpha 2(I) chains. Evidence of distinct structural domains within the type I collagen triple helix. *J. Biol. Chem. 266*, 2590−2594.

BI020070D