Sequence Dependence of Renucleation after a Gly Mutation in Model Collagen Peptides

Received for publication, May 30, 2006, and in revised form, September 20, 2006. Published, JBC Papers in Press, September 22, 2006, DOI 10.1074/jbc.M605135200

Timothy J. Hyde1,2, Michael A. Bryan3,4,5, Barbara Brodsky5, and Jean Baum4,4

From the 1Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854 and the 2Department of Biochemistry, University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School, Piscataway, New Jersey 08854

Missense mutations in the collagen triple helix that replace one Gly residue in the (Gly-X-Y)n repeating pattern by a larger amino acid have been shown to delay triple helix folding. One hypothesis is that such mutations interfere with the C- to N-terminal directional propagation and that the identity of the residues immediately N-terminal to the mutation site may determine the delay time and the degree of clinical severity. Model peptides are designed to clarify the role of tripeptide sequences N-terminal to the mutation site, with respect to length, stability, and nucleation propensity, to complete triple helix folding. Two sets of peptides with different N-terminal sequences, one with the natural sequence α(I) 886–900, which is just adjacent to the Gly301 mutation, and one with a GPO(GAO)3 sequence, which occurs at α(I) 865–879, are studied by CD and NMR. Placement of the five tripeptides of the natural α(I) collagen sequence N-terminal to the Gly to Ala mutation site results in a peptide that is folded only C-terminal to the mutation site. In contrast, the presence of the Hyp-rich sequence GPO(GAO)3 N-terminal to the mutation allows complete refolding in the presence of the mutation. The completely folded peptide contains an ordered central region with unusual hydrogen bonding while maintaining standard triple helix structure at the N- and C-terminal ends. These peptide results suggest that the location and sequences of downstream regions favorable for renucleation could be the key factor in the completion of a triple helix N-terminal to a mutation.

Abnormalities in protein folding are known to play a role in many diseases, including those arising from mutations in the collagen triple helix (1). The best characterized collagen disease is osteogenesis imperfecta (OI),5 or brittle bone disease, in which there is defective mineralization of bones in type I collagen (2, 3). Missense mutations that change one Gly in the repetitive (Gly-X-Y)n sequence are the most common mutations (4). Such Gly mutations are found all along the collagen chain, suggesting that the loss of a Gly at any site in the triple helix has pathological consequences. The phenotype of the disease varies widely, depending on the type of amino acid substitution and the site of the mutation (5, 6). There is evidence of abnormal folding of collagen in OI and other collagen diseases, which may relate to the pathology (7, 8).

Folding of the triple helix is a complex, multistep process that includes association of three chains to form the supercoiled polyproline II triple helical structure (8). Collagen is synthesized in a procollagen form, with N- and C-terminal globular propeptides flanking the (Gly-X-Y)n central domain (9, 10). Posttranslational hydroxylation of Pro and Lys residues in the Y positions and further glycosylation of Hyl occur while the chains are unfolded (11, 12). Trimerization occurs through the association of three C-terminal propeptides, and then nucleation of the triple helix takes place at the (Gly-Pro-Hyp)n rich sequence found at the C terminus of the (Gly-X-Y)n region. After nucleation of the three chains, the triple helix propagates in a zipper-like fashion from the C to the N terminus, with a rate-limiting step of cis-trans isomerization of Pro and Hyp (13, 14).

OI has been characterized as a human folding disease (2). Enzymatic digestion studies show that OI mutations result in slower collagen folding rates (2, 7, 15). All OI collagens show increased levels of lysyl hydroxylation and glycosylation N-terminal to the mutation site (4, 16). This has been suggested to result from delayed folding, since posttranslational enzymatic modifications can occur only on the unfolded chain. Thus, the appearance of increased amounts N-terminal to the Gly substitution site suggests that the mutation delays triple helix formation at the site, extending the time during which collagens can be modified by enzymes that act only on the unfolded state.

In order to study details of the folding mechanism, biophysical studies, including NMR and CD, have been applied to collagen model peptides (17). Peptides that satisfy the stringent (Gly-X-Y)n sequence constraint and have a high content of imino acids will form stable triple helices in solution and have been used to model folding of the triple helix (17–19). One system that has been well characterized is peptide T1-892, which contains a C-terminal (Gly-Pro-Hyp)4 sequence modeling the nucleation domain of the Type I sequence and an N-terminal sequence of residues 892–909 from the α(I) chain of
**Renucleation of Collagen Model Peptides**

Type I collagen (GPAGPGPVGPAGGPAGP) (20–22). Peptide T1-892 includes the site of a Gly to Ser mutation at position 901, leading to a mild form of OI. NMR and CD studies have indicated that T1-892 forms a rigid, uniform triple helix along the entire length of the peptide. Substitution of Gly to Ala or Gly to Ser to model the OI mutation results in the formation of a peptide that is folded at the C-terminal end and unfolded at the mutation site and N-terminal to it (23). Based on these data, it has been suggested that that the mutation interrupts the C- to N-terminal folding of the triple helix in the peptide.

In OI collagens, however, full-length triple helical molecules are formed, indicating that triple helix can be present on both sides of the substitution site. To better model OI collagen, peptide studies were initiated to develop models that could fold on both sides of the mutation. The sequences N-terminal to the mutation were varied with respect to length, stability, and nucleation propensity to better understand the requirements for completion of a triple helix. The presence of five tripeptides N-terminal to the mutation with a Gly-Ala-Hyp-rich sequence was necessary and sufficient to lead to a fully trimeric conformation, with triple helix formed both N- and C-terminal to the substitution site. This GAO-rich sequence promotes stability and was shown to act as an effective nucleus for triple helix formation. In contrast, the presence of five tripeptide units of the natural α(I) collagen sequence N-terminal to the α(901) OI mutation does not lead to complete peptide folding, suggesting that the triple helix may not continue past a mutation with the expected zipper-like C to N folding.

**Experimental Procedures**

**Peptides**—Peptides were purchased from SynPep (Dublin, CA) and Alta Bioscience (Birmingham, UK). Peptides were purified on a C-18 column by high pressure liquid chromatography, and their identity was confirmed by mass spectrometry.

**NMR Spectroscopy**—Samples T1-886, T1-886(G16A), T1-865, and T1-865(G16A) were prepared at a pH of 2.5 and a concentration of 7 mM. Concentration was confirmed by UV spectroscopy. Samples T1-892 and T1-892(GAO) were prepared at a concentration of 9 mM. All NMR experiments were carried out on a Varian INOVA 500-MHz spectrometer. Heteronuclear single quantum coherence spectroscopy (HSQC) spectra were recorded at 0 °C with a sweep width of 6000 Hz in the 1H dimension and 1800 Hz in the 15N dimension. Temperatures were calibrated with methanol (+0.1 °C). All of the pulse sequences employed enhanced sensitivity pulsed field gradient techniques (24). The 1H-15N NOE relaxation measurements were performed in the presence and absence of 1H saturation at 0, 10, 20, 30, and 40 °C (25).

The NMR folding experiments were performed as previously described (20). The sample was denatured outside of the NMR spectrometer by heating for 10 min at 50 °C and then immersed in an ice water bath at 0 °C for 30 s to quickly reduce the temperature to 0 °C. The sample was then transferred to the spectrometer, which was precooled at 0 °C. The total dead time for sample transfer was 1 min. A series of two-dimensional 1H-15N HSQC spectra was acquired every 3.5 min at 0 °C. For each two-dimensional 1H-15N HSQC spectrum, 32 t1 increments were acquired with two scans per increment. Each folding experiment was performed three times per peptide, and the intensities were averaged. The kinetics of folding were monitored by measuring cross-peak volumes as a function of time. All two-dimensional data were processed on a Silicon Graphics work station with the Felix 97 software package (MSI, San Diego, CA).

**Circular Dichroism Spectroscopy**—Circular dichroism spectra were recorded on an Aviv model 62DS spectrophotometer. Cuvettes of 1-mm path length were used, and the temperature of the cells was controlled using a Peltier thermoelectric temperature controller. Samples were dissolved in phosphate-buffered saline buffer, pH 7.0, and pre-equilibrated at 4 °C for 2–4 days prior to recording spectra. Wavelength scans were collected with a 2-s averaging time, from 260 to 210 nm at 0.5-nm steps at 5 °C. For temperature-induced denaturation, the ellipticity was monitored at 225 nm. Peptides were equilibrated for 2 min with steps of 0.3 °C, giving an average heating rate 0.1 °C/min (26). The melting temperature, Tm, was defined as the temperature at which the fraction folded (F) is equal to 0.5. For this monomer to trimer system, F was calculated as follows,

\[
F = \frac{(\theta_{\text{observed}} - \theta_{\text{monomer}})}{(\theta_{\text{trimer}} - \theta_{\text{monomer}})}
\]

where \(\theta_{\text{observed}}\) represents the observed ellipticity, \(\theta_{\text{trimer}}\) is the trimer signal, and \(\theta_{\text{monomer}}\) is the monomer signal. The CD folding experiments were carried out by heating the sample to 70 °C for 15 min and then rapidly quenching in an ice water bath and placing in a pre-equilibrated CD cell at 5 °C (21). The dead time was on the order of 25 s. The ellipticity at 225 nm was monitored, with a time constant of 2 s and time interval of 10 s. The half time of refolding, \(t_{1/2}\), was determined as the time for the fraction folded to reach 0.5.

**Results**

Peptides were designed to consider the effects of length, imino acid content, and nucleation propensity on the completion of folding in the presence of a Gly mutation (Fig. 1, a and b).

**Study of Peptides with Natural Sequences N-terminal to a Gly Mutation**—Peptide T1-886 contains a (GPO)4 C-terminal nucleation sequence and seven tripeptide units taken from the α1(I) sequence (residues 886–906) of collagen. A homologous peptide T1-886(G16A) was synthesized with a Gly to Ala mutation at position 16, with five tripeptide units N-terminal to the mutation site (Fig. 1b).

CD spectroscopy was used to characterize the conformation and thermal stability of T1-886 and T1-886(G16A) (Fig. 2). Peptide T1-886 gave a characteristic triple helix CD spectrum at low temperature with a maximum near 224 nm and a minimum at 198 nm (Fig. 2, inset). A sharp thermal transition with a Tm of 25 °C is observed as the temperature is increased. Peptide T1-886(G16A), which contains a Gly to Ala substitution at position 16, shows a significant drop in mean residue ellipticity (MRE) at 225 nm from 5194 degrees cm² dmol⁻¹ to 1988 degrees cm² dmol⁻¹ and a substantial decrease in melting temperature from 25 to 7 °C, indicating a large destabilization of the peptide and loss of triple helix content. The reduced ellipticity and stability could arise from an alternate conformation in the central region of the peptide or could be due to interruption of
Renucleation of Collagen Model Peptides

FIGURE 1. a, C-terminal region of the α1(I) collagen sequence. Triplet sequences used to form peptides in b and c are shown in different colors. b, peptides T1-886 and T1-886(G16A) contain five N-terminal triplets that come directly from residues adjacent to the α1(I) 901 mutation site in collagen (residues 886–909 shown in green) and a C-terminal (GPO)₄ nucleation domain from the C-terminal end of the α1(I) collagen chain (residues shown in magenta). Peptides T1-865 and T1-865(G16A) contain five N-terminal triplets that are GAO-rich (residues 865–879 shown in blue) and downstream of the α1(I) 901 mutation site and a C-terminal (GPO)₄ nucleation domain (residues shown in magenta). c, peptides T1-892 and T1-892(GAO)₃ both contain N-terminal residues 892–909 from the α1(I) sequence (shown in green), and T1-892 contains a C-terminal (GPO)₄ nucleation domain (shown in magenta), whereas T1-892(GAO)₃ contains a C-terminal GPO(GAO)₃ nucleation domain (shown in blue).

FIGURE 2. Thermal unfolding of peptides T1-886 (solid line) and T1-886(G16A) (dashed line) monitored by CD spectroscopy at 225 nm. Inset, CD scan of T1-886 (solid line) and T1-886(G16A) (dashed line) at 0 °C.

the folding at the mutation site and the formation of a partially folded triple helix.

NMR studies were carried out on ¹⁵N-labeled peptides T1-886 and T1-886(G16A) in order to ascertain the conformation of individual residues. The NMR spectrum of peptide T1-886 contains both monomer and trimer resonances for all labeled positions, Gly²⁸, Ala¹⁸, and Gly⁷, consistent with the trimer form being in equilibrium with the monomer form of the peptide (Fig. 3). The appearance of three trimer peaks and positive [¹H-¹⁵N] NOEs (Table 1) for Gly⁷ and Ala¹⁸ is consistent with the nonrepetitive environment surrounding Gly⁷ and Ala¹⁸ and the formation of a rigid triple helix along the peptide chain. As seen previously, in addition to the monomer resonance, a single trimer resonance with positive [¹H-¹⁵N] NOE is seen for the rigid C-terminal repetitive (GPO)₄ triple helical environment at Gly²⁸ (27). In contrast, the HSQC spectrum of T1-886(G16A) shows distinct monomer and trimer peaks only for resonances that are C-terminal to the substitution site, suggesting that the Gly to Ala mutation disrupts the triple helical conformation. The C-terminal Gly²⁸ has a trimer resonance with a positive NOE value, indicating that the C-terminal end forms a rigid triple helix. Residues Gly⁷ at the N-terminal end and Ala¹⁶ at the mutation site show only monomer resonances with characteristic negative [¹H-¹⁵N] NOEs, indicative of very flexible residues. T1-886(G16A) is only partially folded with no triple helix N-terminal to the mutation site.

CD and NMR studies on T1-886(G16A) show that using five triplets of native sequences adjacent to the mutation site to increase the length N-terminal to the substitution site is not sufficient to reinitiate folding after the mutation site. Computational analysis suggests that further increasing the length alone will not promote N-terminal refolding because of the inherent instability of the N-terminal triplets adjacent to 886, GPVGPAKS (28) (data not shown). These sequences do not have high stability or propensity to nucleate. However, at 865–879 there is a sequence of triplets, GPO(GAO)₃GPV, with high imino acid content and high nucleation propensity (Fig. 1). It is possible that this GPO(GAO)₃ sequence could serve as a good renucleation sequence N-terminal to the Gly to Ala mutation site, and peptides were designed to test this hypothesis.

Sequences with High Imino Acid Content at the N Term

Renucleation of Collagen Model Peptides
peaks with positive $[^{1}H-^{15}N]$ NOEs are observed at Ala$^{16}$, indicating that the environment surrounding the mutation site is different for the three chains but rigid on the nanosecond-picosecond time scale. A large downfield chemical shift perturbation is seen for one of the Ala resonances (9.5 ppm) relative to the other two trimer resonances. Based on the NMR data, placement of the GPO(GAO)$_3$ sequence at the N-terminal end of the peptide allows formation of a rigid triple helix at both the C- and N-terminal ends of the peptide as well as the formation of an ordered trimer in the central mutation region.

The effect of the N-terminal GPO(GAO)$_3$ nucleation sequence on the directionality of triple helix folding can be seen in the real time NMR folding experiments of peptides T1-886, T1-886(G16A), T1-865, and T1-865(G16A) (Fig. 6, a–d). Comparison of the folding of native peptides T1-886 and T1-865 shows that T1-886 folds from the C- to N-terminal direction, whereas T1-865 appears to fold as a cooperative unit. Substitution of Gly to Ala in T1-886 results in a peptide that is only partially folded at the C-terminal end but has folding rates at the C-terminal (GPO)$_4$ end that are similar to those seen in the native peptide. Substitution of Gly to Ala in T1-886 results in a peptide that is partially folded at the C-terminal end but shows folding around the mutation site. However, relative to the C-terminal folding, folding at the mutation site and N-terminal to it appears to be slowed down.

**Evidence That GPO(GAO)$_3$ Is a Nucleation Domain**

A peptide T1-892(GAO)$_3$ was designed to investigate whether the sequence GPO(GAO)$_3$ can function as a nucleation domain when located at the C-terminus (Fig. 1c). CD data indicate that T1-892(GAO)$_3$ has an MRE of 4750 degrees cm$^2$ dmol$^{-1}$, similar to T1-892, indicating the formation of triple helix (22). The $T_m$ of T1-892(GAO)$_3$ is 14 °C, lower than the $T_m$ of 20 °C seen for T1-892. HSQC spectra of T1-892(GAO)$_3$ are typical of spectra obtained for folded trimers, and $[^{1}H-^{15}N]$ NOE data indicate that the peptide forms a rigid triple helix along the entire peptide chain (Table 1). Real time NMR folding experiments indicate that the C-terminal Gly$_{25}$ residue folds faster than the N-terminal Gly$_{7}$ residue, indicating that T1-892(GAO)$_3$ folds in a directional manner from the C- to N-terminal direction (data not shown), similar to folding of T1-892 (20, 27). The ability of T1-892(GAO)$_3$ to form a stable triple helix that folds from the C to N-terminal direction indicates that GPO(GAO)$_3$ can function as a C-terminal nucleation domain.

To further compare the nucleation ability of GPO(GAO)$_3$ with (GPO)$_4$, the folding kinetics of T1-892 and T1-892(GAO)$_3$ were compared at two concentrations by CD (Fig. 7). At 1.5 mg/ml, T1-892(GAO)$_3$ folds significantly more slowly ($t_{1/2} =$...
81.8 min) than T1-892 (t½ = 19.2 min), indicating that GPO(GAO)₃ is a less efficient nucleation domain than (GPO)₄. The difference between the folding rates of the two peptides becomes less significant at higher concentration (Fig. 7).

To investigate the molecular basis of the more efficient nucleation capacity of (GPO)₄ compared with GPO(GAO)₃, NMR dynamics experiments were used to assess the flexibility of residue Gly¹⁵ in the monomer and trimer forms of T1-892 and T1-892(GAO)₃ as a function of temperature (Fig. 8). The all-trans monomer Gly¹⁵ residues are more flexible than the trimer Gly¹⁵ residues as seen by the negative [¹H,¹⁵N] NOE values for the monomer and the positive values for the trimer. Substitution of GPO(GAO)₃ in the C-terminal end results in more negative NOEs at position Gly¹⁵ with a more distinct effect at higher temperatures. The more negative NOEs in T1-892(GAO)₃ indicate that the lower imino acid content of GPO(GAO)₃ relative to (GPO)₄ in T1-892 results in more flexibility and mobility in the nucleation domain.

**DISCUSSION**

Redesign of sequences has allowed refolding of a triple helix on the N-terminal as well as the C-terminal side of a Gly substitution in the context of an OI mutation site. Peptides that contain the natural collagen sequence found directly N-terminal to the α(I) 901 mutation site did not generate refolding around the Gly to Ala mutation. Substitution of the N-terminal imino acid-poor sequence with a more stable imino acid-rich GPO(GAO)₃ sequence found further downstream of the mutation results in a stable triple helix that folds around the Gly to Ala mutation site, with rigid triple helical N- and C-terminal regions. Further studies will be performed to understand whether the naturally occurring heterotrimers have similar features to the homotrimer models studied here.

Although the presence of a Gly substitution prevents a continuous triple helix, the NMR data indicate that the central region around the substitution site is ordered. The NMR data on T1-865(G16A) indicate the presence of three distinct trimer resonances at the mutation site, indicating that the three chains are in different environments. The high values of the NOEs indicate that each residue adopts a specific ordered conformation at the mutation site. The ¹H chemical shift for one of the Ala NH resonances at the mutation site is highly downfield shifted (~9.5 ppm). This unusual downfield shift is also seen in the NMR spectra of a different peptide model of OI, (POG)₁₀ with a Gly to Ala substitution (31). Details of the structural perturbations at the mutation site of the (POG)₁₀ peptide with a Gly to Ala substitution have been described by an x-ray crystal structure (32–34), and it has been shown that the breaking of the repeating (Gly-X-Y)ₙ sequence by the Gly to Ala substitution results in an alteration of the conformation with a local untwisting of the triple helix. At the substitution site, direct interchain hydrogen bonds are replaced with interstitial water bridges between the peptide chains. The significant downfield NMR chemical shifts observed for peptide T1-895(G16A) and for peptide (POG)₁₀ with the Gly to Ala substitution may reflect similar unusual hydrogen bonding patterns to the interstitial water bridges seen in the x-ray structure of the (POG)₁₀ Gly to Ala peptide.
The success in complete folding around a mutation site in our peptide systems depended on having the sequence GPO(GAO)_3 adjacent to the mutation. Further peptide designs show that the GPO(GAO)_3 sequence is a good C-terminal nucleation domain, although less efficient than (GPO)_4, particularly at low concentrations. The nature of a renucleation sequence has not been defined in collagen, but it is proposed that a good renucleation sequence may be similar to a good nucleation sequence that allows folding to be initiated. Factors that promote nucleation have been shown to relate to high imino acid content and high stability (9, 35). Experiments performed on a set of host-guest peptides with different tripeptide sequences have shown that GPO sequences fold faster than GAO sequences, but both are good nucleation sequences (35). Further studies have shown that nucleation propensity may also be related to the dynamics of the monomer state with a highly constrained monomer state related to the efficiency of folding (36). The CD and NMR results on peptides T1-892 and T1-892(GAO)_3 indicate that imino acid-rich sequences of the form (GPO)_4 or GPO(GAO)_3 can act as nucleation domains when placed at the C-terminal end of the peptide.

The observation that the GPO(GAO)_3 sequence, which occurs at α1(I) 865–876 but not the natural sequence α1(I) 886–900 placed adjacent to the Gly^901 mutation site allows complete folding on both sides of the mutation has implications for the folding of OI collagens with a Gly substitution. It was previously suggested that the sequence of the triplet located N-terminal to the mutation site is important for renucleation and completion of collagen triple helix formation (2, 23, 37). The implication is that if renucleation occurs, it will take place adjacent to the mutation site and continue in a unidirectional propagation (Fig. 9a). The fact that the sequence that was directly adjacent to the mutation did not result in refolding suggests that renucleation sequences further downstream from the mutation rather than sequences immediately adjacent to it may be required for folding to resume in collagen. The stable and imino acid-rich GPO(GAO)_3 sequence found further downstream in the collagen sequence may act as a second nucleation domain, allowing folding to restart in collagen.

The discontinuity in folding at the mutation site suggests that alternate folding mechanisms may be required to explain OI collagen folding. Assuming that a renucleation site must be encountered in order to restart folding downstream of the mutation, then reverse folding will be required as well in order to form a complete triple helix around the mutation site (Fig. 9b). To investigate the nature of reverse folding, a set of peptides with N-terminal (GPO)_4 nucleation domains was previously designed (27). One peptide, T1^1-892n, is similar in design to peptide T1-892(GAO)_3, and contains a (GPO)_4 nucleation domain N-terminal to the imino acid-poor 892 sequence. NMR and CD studies have shown that the peptide forms a stable rigid triple helix along the entire chain but that the directional nature of the folding is lost when nucleation takes place at the N terminus. To further investigate the details of folding at the N terminus, a second peptide, T1^1-892-G10S, was designed to model the effect of an OI mutation N-terminal to a nucleation domain and contains a Gly to Ser replacement at position 10. This peptide folds up to the Gly to Ser mutation site from the N-terminal end, indicating that formation of a triple helix N-terminal to a mutation site is possible in the presence of a strong (GPO)_4 nucleating sequence at the N-terminal end (27). Reverse N- to C-terminal folding has also been observed in

**FIGURE 7.** CD folding profiles of T1-892 and T1-892(GAO)_3 monitored at various concentrations at 225 nm. T1-892 at 3 mg/ml (black line) and 1.5 mg/ml (dotted black line), compared with T1-892(GAO)_3 at 3 mg/ml (red line) and 1.5 mg/ml (red dotted line).

**FIGURE 8.** (1H-15N) NOE for T1-892 (●) and T1-892(GAO)_3 (▲) as a function of temperature for Gly^865. Open symbols represent the trimer Gly^865, and closed symbols represent the all-trans Gly^865 monomer residue from each peptide.

**FIGURE 9.** Schematic diagram of the folding of OI collagen with a continuous folding model (a) and a discontinuous folding model (b). In both models, nucleation of the three chains occurs at the C-terminal end (depicted by the circle), and folding is interrupted at the α1(I) 901 mutation site (depicted with a red line). In the continuous folding model (a), renucleation occurs adjacent to the 901 mutation site (depicted by the circle), followed by propagation to the N-terminal end (depicted with an arrow), whereas in the discontinuous model (b), renucleation (depicted by the circle) occurs further downstream of the 901 mutation site at the GAO-rich site (depicted by a blue line), requiring reverse and forward folding to form full-length collagen (depicted with the arrows).
peptides that contain disulfide bonds or foldons at the N-terminal ends (38, 39).

The existence of internal renucleation sites and the possibility of folding up to the mutation site from either the N- or C-terminal end supports the potential for a bidirectional mechanism to complete the folding of collagens with OI mutations. Crystal structure and NMR data of OI peptides support the existence of water-mediated hydrogen bonds at the interruption site, and the recruitment of these waters during the bi-directional folding may contribute to the folding delay that is observed in OI collagens. Examination of the collagen sequence shows that there are a number of regions that contain two sequential (GXO-GXO) triplets that may serve as internal renucleation sites. With the discontinuity in the folding and the lack of a unidirectional zipper-like folding model, the distance from the mutation site to the next renucleation site may also play a role in the observed folding delay in OI collagens and the degree of clinical severity.

REFERENCES
1. Myllyharju, J., and Kivirikko, K. I. (2001) Ann. Med. 33, 7–21
2. Bachinger, H. P., Morris, N. P., and Davis, J. M. (1993) Am. J. Med. Genet. 45, 152–162
3. Byers, P. H., and Cole, W. G. (2002) Connective Tissue and Its Hereditable Disorders, Wiley-Liss, New York
4. Byers, P. H. (1990) Trends Genet. 6, 293–300
5. Kuivaniemi, H., Tromp, G., and Prockop, D. (1997) Hum. Mutat. 9, 300–315
6. Byers, P. H. (2001) Philos. Trans. R. Soc. Lond. B Biol. Sci. 356, 151–158
7. Raghunath, M., Bruckner, P., and Steinmann, B. (1994) J. Mol. Biol. 236, 940–949
8. Engel, J., and Prockop, D. J. (1988) Annu. Rev. Biophys. Biophys. Chem. 17, 137–152
9. McLaughlin, S. H., and Bulleid, N. J. (1998) Matrix Biol. 16, 369–377
10. Lamande, S. R., and Bateman, J. F. (1999) Semin. Cell Dev. Biol. 10, 455–464
11. Kiely, C. M., Hopkinson, I., and Grant, M. E. (1993) in The Collagen Family: Structure, Assembly, and Organization in the Extracellular Matrix (Royes, P. M., and Steinmann, B. U., eds) Wiley Liss, New York, pp. 159–221
12. Kivirikko, K. I., and Myllyla, R. (1982) Methods Enzymol. 82, 245–304
13. Bachinger, H. P., Bruckner, P., Timpl, R., and Engel, J. (1978) Eur. J. Biochem. 90, 605–613
14. Bachinger, H. P., Bruckner, P., Timpl, R., Prockop, D. J., and Engel, J. (1980) Eur. J. Biochem. 106, 619–632
15. Bonadio, J., and Byers, P. (1985) Nature 316, 363–366
16. Prockop, D. J. (1988) Trans. Am. Clin. Climatol. Assoc. 100, 70–80
17. Baum, J., and Brodsky, B. (1999) Curr. Opin. Struct. Biol. 9, 122–128
18. Baum, J., and Brodsky, B. (2000) Mechanisms of Protein Folding, 2nd Ed., pp. 330–351, Oxford University Press, Oxford
19. Brodsky, B., and Ramshaw, J. A. (1997) Matrix Biol. 15, 545–554
20. Buevich, A. V., Dai, Q. H., Liu, X., Brodsky, B., and Baum, J. (2000) Biochemistry 39, 4299–4308
21. Xu, Y., Bhat, M., and Brodsky, B. (2002) Biochemistry 41, 8143–8151
22. Yang, W., Battinieni, M. L., and Brodsky, B. (1997) Biochemistry 36, 6930–6935
23. Liu, X. Y., Kim, S., Dai, Q. H., Brodsky, B., and Baum, J. (1998) Biochemistry 37, 15528–15533
24. Kay, L. E., Keifer, P., and Saarinen, T. (1992) J. Am. Chem. Soc. 114, 10663–10665
25. Farrow, N. A., Muhandiram, R., Singer, 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
26. Persikov, A. V., Xu, Y., and Brodsky, B. (2004) Protein Sci. 13, 893–902
27. Buevich, A. V., Silva, T., Brodsky, B., and Baum, J. (2004) J. Biol. Chem. 279, 46890–46895
28. Persikov, A. V., Ramshaw, J. A., and Brodsky, B. (2005) J. Biol. Chem. 280, 19343–19349
29. Long, C. G., Braswell, E., Zhu, D., Apigo, J., Baum, J., and Brodsky, B. (1993) Biochemistry 32, 11688–11695
30. Vogel, B. E., Doelz, R., Kadler, K. E., Hojima, Y., Engel, J., and Prockop, D. J. (1988) J. Biol. Chem. 263, 19249–19255
31. Buevich, A., and Baum, J. (2001) Philos. Trans. R. Soc. Lond. B Biol. Sci. 356, 159–168
32. Buevich, A., Brodsky, B., and Berman, H. M. (1995) Structure 3, 893–906
33. Buevich, A., Brodsky, B., and Berman, H. M. (1996) Connect. Tissue Res. 35, 401–406
34. Buevich, A., Eaton, M., Brodsky, B., and Kim, J. (1996) Science 276, 75–81
35. Ackerman, M. S., Bhat, M., Shenoy, N., Beck, K., Ramshaw, J. A., and Brodsky, B. (1999) J. Biol. Chem. 274, 7668–7673
36. Xu, Y. J., Hyde, T., Wang, X., Bhat, M., Brodsky, B., and Baum, J. (2003) Biochemistry 42, 8696–8703
37. Wenstrup, R. J., Shrager-Howe, A. W., Lever, L. W., Phillips, C. L., Byers, P. H., and Cohn, D. H. (1991) J. Biol. Chem. 266, 2590–2594
38. Frank, S., Boudko, S., Mizuno, K., Schulthess, T., Engel, J., and Bachinger, H. P. (2003) J. Biol. Chem. 278, 7747–7750
39. Barth, D., Kyrielis, O., Frank, S., Renner, C., and Moroder, L. (2003) Chemistry 9, 3703–3714