Sequence Environment of Mutation Affects Stability and Folding in Collagen Model Peptides of Osteogenesis Imperfecta

ARTICLE *in* BIOPOLYMERS · JANUARY 2011

Impact Factor: 2.39 · DOI: 10.1002/bip.21432 · Source: PubMed

3 AUTHORS, INCLUDING:

Barbara Brodsky
Tufts University
136 PUBLICATIONS **6,995** CITATIONS

Published in final edited form as:
Biopolymers. 2011 ; 96(1): 4–13. doi:10.1002/bip.21432.

Sequence Environment of Mutation Affects Stability and Folding in Collagen Model Peptides of Osteogenesis Imperfecta

Michael A. Bryan, **Haiming Cheng**, and **Barbara Brodsky***
Department of Biochemistry, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, Piscataway, New Jersey 08854

Abstract

Osteogenesis Imperfecta (OI), a disorder characterized by fragile bones, is often a consequence of missense mutations in type I collagen which change one Gly in the repeating (Gly-Xaaa-Yaa)n sequence to a larger amino acid. The impact of local environment and the identity of the residue replacing Gly was investigated using two sets of triple-helical peptides. Gly mutations in the highly stable $(Pro\text{-}Hyp\text{-}Gly)_{10}$ system are compared with mutations in T1-865 peptides where the mutation is located within a less stable natural collagen sequence. Replacement of a Gly residue by Ala, Ser, or Arg leads to significant triple-helical destabilization in both peptide systems. The loss of stability (ΔT_m) due to a Gly to Ala or Gly to Ser change was greater in the more rigid $(Pro\text{-}Hyp\text{-}Gly)_{10}$ peptides than in the T1-865 set, as expected. But the final T_m values, which may be the more biologically meaningful parameters, were higher for the $(Pro\text{-}Hyp\text{-}Gly)_{10}$ mutation peptides than for the corresponding T1-865 mutation peptides. In both peptide environments, a Gly to Arg replacement prevented the formation of a fully folded triple-helix. Monitoring of folding by differential scanning calorimetry showed a lower stability species as well as the fully folded triple-helical molecules for T1-865 peptides with Gly to Ala or Ser replacements, and this lower stability species disappears as a function of time. The difficulty in propagation through a mutation site in T1-865 peptides may relate to the delayed folding seen in OI collagens, and indicates a dependence of folding mechanism on the local sequence environment.

Keywords

collagen; osteogenesis imperfecta; folding; triple helix; peptides

INTRODUCTION

Mutations in different types of collagens lead to a range of hereditary diseases. The clinical phenotype resulting from a mutation depends on the tissue location and function of the type of collagen involved [1]. There are twenty-eight distinct types of collagens [2–3], defined as extracellular matrix molecules with a triple-helix domain. The most abundant are fibrillar collagens which provide the structural framework for the extracellular matrix of bone, skin, tendons, ligaments, and blood vessels. Mutations in the fibrillar type I collagen, which is the major protein in bone, lead to Osteogenesis Imperfecta (OI) [4] and missense mutations which change one Gly residue to a larger residue are common. The mechanism by which a Gly missense mutation leads to the development of OI and the factors that lead to varying clinical severity of OI are not well understood.

*Correspondence to: Barbara Brodsky, Department of Biochemistry, University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, NJ 08854. brodsky@umdnj.edu.

The replacement of Gly by a larger residue as a result of a missense mutation alters the regularity of the triple-helix structure. The collagen triple-helix is a rod like structure consisting of three polyproline II-like polypeptide chains supercoiled about a common axis [5–8]. Each chain has a $(Gly\text{-}Xaa\text{-}Yaa)_n$ repeating sequence motif, where the Xaa and Yaa positions are often occupied by Pro and hydroxyproline (Hyp) respectively. Residues in the X and Y positions are largely exposed to the solvent while the Gly residues are buried in the central region of the molecule. Gly is required as every third residue for ideal close packing of the three strands and for formation of interchain peptide hydrogen bonds. The replacement of Gly by any other residue must result in distortion of the core of the triple helix which may play a role in the development of collagen diseases such as OI.

OI is an autosomal dominant disorder characterized by varying levels of bone fragility, ranging from a mild phenotype with occasional fractures to perinatal lethal cases [4,9–10]. The majority of OI mutations are single base changes in the α1 or α2 chains of type I collagen which replace one Gly in the repeating $(Gly\text{-}Xaa\text{-}Yaa)_n$ sequence of collagen. A single base change can transform a Gly codon to a codon for one of eight residues, Ala, Ser, Val, Cys, Arg, Asp, Glu, and Trp, or to a stop codon. In the more than 600 missense mutations in type I collagen reported to result in OI[4], Gly residues are most frequently replaced by Ser or Cys, while Arg, Val, Asp, Glu and Ala are also well represented.

A number of ideas have been proposed to explain how a Gly mutation within the type I collagen triple-helix affects mineralization and causes OI [4]. Proposals include delayed folding which can lead to endoplasmic reticulum associated degradation and impaired collagen secretion [11], or altered fibril structure which may affect interactions with proteoglycans or other matrix components [4]. Further, it is difficult to explain why a Gly to Ser mutation at one location leads to a mild or moderate case of OI, while a nearby Gly to Ser mutation leads to a lethal phenotype. The local sequence environment of a mutation has been considered as a possible factor in the varying clinical severity of OI, including features such as local stability, the proximity to a salt bridge, and the stability of a triplet adjacent to the mutation triplet [4,12–14]. There is recent evidence that more large scale parameters, e.g. the domain location of the mutation may be important [15].

The use of model triple-helical peptides complements recent studies on OI collagens expressed in fibroblasts, offering the potential to focus on the effect of a Gly substitution in a well defined simple environment, where the structural consequences can be defined by a range of biophysical techniques. A high resolution structure, obtained by x-ray crystallography, for a peptide with a Gly to Ala replacement in the very stable peptide $(Pro\text{-}Hyp\text{-}Gly)_{10}$, denoted here as $(POG)_{10}$, shows a local perturbation of the triple-helix conformation and loss of hydrogen bonding, with a disruption of the superhelix register [7]. Thermal stability studies on a host-guest triple-helical peptide set where the central Gly in a host guest peptide system, $(Pro\text{-}Hyp\text{-}Gly)_8$, is replaced by varying residues indicates the degree of destabilization, ΔT_m, is dependent on the identity of the residue replacing Gly[16]. Some peptides with a Gly mutation within a collagen sequence form only a partially folded species [17]. A more recent peptide design, T1-865, includes a strong natural N-terminal renucleating sequence $Gly\text{-}Pro\text{-}Hyp\text{-}(Gly\text{-}Ala\text{-}Hyp)_3$ as well as a C-terminal $(Gly\text{-}Pro\text{-}Hyp)_4$ cap surrounding a natural α1(I) chain sequence, and NMR studies on peptides with labeled residues at specific sites indicate a fully triple-helical structure for both the native peptide and the peptide containing a Gly to Ala substitution [18].

The structural and folding consequences of replacing Gly by different residues are compared here in two peptide systems containing different local environments. The peptide $(POG)_{10}$ has the highest possible imino acid content and represents the most rigid environment, while the peptide, T1-865 contains a less rigid natural type I collagen sequence. Although a

replacement of Gly by Ala or Ser leads to a greater loss of stability in the more rigid $(POG)_{10}$ set, the final stability of the substituted peptide is higher than the same mutation in T1-865, because of the greater original thermal stability of the parent $(POG)_{10}$ peptide. Folding of the T1-865 peptides with Gly to Ala and Gly to Ser mutations leads to an accumulation of a less stable species in addition to the fully folded molecule, which is not seen in the corresponding $(POG)_{10}$ peptides. The results suggest that propagation is paused at the Gly substitution site and complete folding is eventually achieved by a non-linear mechanism.

MATERIALS AND METHODS

Peptides

The peptide $(Pro\text{-}Hyp\text{-}Gly)_{10}$ denoted $(POG)_{10}$ was obtained from Peptides International (Louisville, Kentucky). Peptides containing Gly substitutions in the $(Gly\text{-}Pro\text{-}Hyp)_{10}$ set were obtained as follows: POG(G→S), $(Pro\text{-}Hyp\text{-}Gly)_4$-Pro-Hyp-Ser-$(Pro\text{-}Hyp\text{-}Gly)_5$, synthesized by Alta Bioscience (Edgbaston, Birmingham, UK); POG(G→A), $(Pro\text{-}Hyp\text{-}Gly)_4$-Pro-Hyp-Ala-$(Pro\text{-}Hyp\text{-}Gly)_5$ and POG(G→R), $(Pro\text{-}Hyp\text{-}Gly)_4$-Pro-Hyp-Arg-$(Pro\text{-}Hyp\text{-}Gly)_5$ were synthesized by Tuffs University Core Facility (Boston, MA). All peptides in the T1-865 peptide set were synthesized by the Tufts University Core Facility (Boston, MA): T1-865, $(Gly\text{-}Pro\text{-}Hyp)$-$(Gly\text{-}Ala\text{-}Hyp)_3$-Gly-Pro-Val-Gly-Pro-Ala-Gly-Ala-Arg-$(Gly\text{-}Pro\text{-}Hyp)_4$; T1-865(G→A), $(Gly\text{-}Pro\text{-}Hyp)$-$(Gly\text{-}Ala\text{-}Hyp)_3$-Gly-Pro-Val-Ala-Pro-Ala-Gly-Ala-Arg-$(Gly\text{-}Pro\text{-}Hyp)_4$; T1-865(G→S), $(Gly\text{-}Pro\text{-}Hyp)$-$(Gly\text{-}Ala\text{-}Hyp)_3$-Gly-Pro-Val-Ser-Pro-Ala-Gly-Ala-Arg-$(Gly\text{-}Pro\text{-}Hyp)_4$; T1-865(G→R), $(Gly\text{-}Pro\text{-}Hyp)$-$(Gly\text{-}Ala\text{-}Hyp)_3$-Gly-Pro-Val-Arg-Pro-Ala-Gly-Ala-Arg-$(Gly\text{-}Pro\text{-}Hyp)_4$; and T1-865(G→D), $(Gly\text{-}Pro\text{-}Hyp)$-$(Gly\text{-}Ala\text{-}Hyp)_3$-Gly-Pro-Val-Asp-Pro-Ala-Gly-Ala-Arg-$(Gly\text{-}Pro\text{-}Hyp)_4$. All peptides were purified using a reverse-phase HPLC system (Shimadzu) on a C-18 column and eluted in 0.1% trifluoroacetic acid with a binary gradient of 0–40% (v/v) water/acetonitrile. The identity of the peptides and their purity were confirmed by laser desorption mass spectrometry.

Circular Dichroism

Circular dichroism spectra were recorded on an Aviv model 62DS spectrophotometer (Aviv Biomedical, Inc.). Cells of path length 0.1 cm were used, and the temperature of the cells was regulated using a Hewlett-Packard Peltier thermoelectric temperature controller. Samples were dissolved in phosphate buffered saline (PBS), at pH 7.0, c=1mg/ml. Peptide concentrations were determined by tyrosine absorbance at 275 nm using $\varepsilon^{275} = 1400$ $M^{-1}{\cdot}cm^{-1}$ for all peptides with the exception of $(Pro\text{-}Hyp\text{-}Gly)_{10}$, which was determined by weight since it doesn't include a C-terminal tyrosine residue.

CD was applied to follow thermal stability and unfolding by monitoring the amplitude of the peak at 225nm as a function of increasing temperature. Standard procedures of a temperature step of 0.3°C with 2-minute equilibration time at each temperature and a 10-sec data collection time were employed [19]. This procedure gives an average heating rate of 0.1°C/min. Prior to all melting experiments collagen model peptides were pre-equilibrated at 0°C for 48hr. The melting temperature (T_m) is defined as the temperature at which the fraction folded $F(T_m)$ is equal to 0.5 as described previously [19].

CD was also used to monitor folding. Samples were denatured at 80°C for 15 minutes and then rapidly quenched in an ice bath for 14 seconds and immediately transferred to the CD cell holders pre-equilibrated at 0°C. The dead time was on the order of 40 seconds. The ellipticity at 225 nm was monitored with a time constant of 2 seconds and time interval of 10

seconds. The half time of refolding $t_{1/2}$ was determined as the time for the fraction folded (F) to reach 0.5.

Differential Scanning Calorimetry

Differential scanning calorimetry (DSC) measurements were performed on a Nano-DSC II, Model 6100 scanning calorimeter from Calorimetry Sciences Corp. All DSC profiles were obtained at a scan rate of 1oC/min and each curve was baseline subtracted prior to data analysis. The thermal transition values obtained by DSC are typically ~8°C degrees higher than those obtained by CD because of the faster heating rate under non-equilibrium conditions [19]. All peptides were dialyzed in PBS buffer for 24–48 hrs prior to DSC studies, and concentrations were determined after dialysis. Calorimetric enthalpy values (ΔH_{cal}) were obtained by integrating the excess heat capacity curve.

DSC was also used to monitor the products formed at different folding times. Samples were heated at 80°C for 20 minutes to ensure that all peptides were denatured. After heat denaturation, peptides were immediately incubated at 0°C and allowed to refold for 0.25, 0.3, 1, 2, 3.5, 5, 25, or 50 hrs. Following incubation, peptides were then transferred to the DSC and run at a scan rate of 1°C/min over a temperature range of 0–40°C or 0–85°C (depending on the thermal stability of the peptide studied). These experiments monitor the unfolding transition of the structure formed by the peptide after folding at low temperature for a specific length of time, as previously described by Miles and Bailey [20]. DSC scans take about 60 minutes to complete, which is on the same time scale as the average folding time of the Gly substitution mutant peptides.

When two peaks are seen in the DSC scan, a deconvolution program "C_pCal" provided with DSC instrument is applied, which uses a second order polynomial function to deconvolute the calorimetric thermograms. This method has been used to obtain the area of the individual peaks of the experimental plots of T1-865 (G→A) and T1-865 (G→S) at 0.3, 1, 2, 5, 25 and 50 hrs incubation periods for 1 and 10 mg/ml. The second order polynomial function gives an approximate fit to the two DSC peaks and provides an estimate of calorimetric enthalpies for each peak as a function of temperature. A second approximation method utilized the $C_p$max (Y-axis) of each peak, plotting the maximum value along the Y-axis for each of the two peaks as a function of time.

RESULTS

Peptide design

The effects of a Gly substitution within the rigid and well established $(POG)_{10}$ peptide set are compared with its effect within the natural and locally less rigid GPVGPAGAR sequence, α1(I) chain residues 898–906, within peptide T1-865. Clinical studies show that a Gly to Ser mutation at position 901 is associated with a mild case of OI (Gly901Ser)[4] and Gly replacements at this site are modeled in the T1-865 system. This central region is capped at the C-terminus by a $(Gly\text{-}Pro\text{-}Hyp)_4$ sequence and at the N-terminus by the sequence Hyp rich sequence $GPO(GAO)_3$ which is found at residues 865–879 of the α1(I) chain [18]. The Gly residue corresponding to position 901 is replaced by Ala, Ser, Arg, or Asp in peptide T1-865, while the central Gly in $(POG)_{10}$ is replaced by Ala, Ser, or Arg.

Impact of a Gly substitution in a $(Pro\text{-}Hyp\text{-}Gly)_{10}$ sequence

The peptide $(POG)_{10}$ has a well characterized collagen CD spectrum with a maximum at 225 nm (MRE_{225} = 4340 deg·cm^2·dmol^{-1}) and a thermal stability of T_m=57°C [19], while the same peptide with a Gly to Ala replacement, designated as POG(G→A) has been shown to form a less stable triple-helix, T_m = 26.5°C, with a localized perturbation defined by x-ray

crystallography [7]. Replacement of the central Gly by a Ser resulted in a new peptide, designated as POG(G→S), with a CD spectrum (MRE_{225}= 4150 deg·cm^2·dmol^{-1}) similar to POG(G→A) and $(POG)_{10}$, but with a lower stability, T_m = 20°C (Figure 1A; Table 1). A Gly to Arg substitution, peptide POG(G→R), results in a CD spectrum with a decrease in the MRE_{225} = 3100 deg·cm^2·dmol^{-1} and only the beginning of a thermal transition at low temperature. Calculations estimate a thermal stability of T_m = 5.5°C, but this is not an accurate value since the low MRE and the shape of the transition suggest a partially folded trimer.

Impact of a Gly substitution in the collagen-like peptide sequence T1-865

As reported previously, peptide T1-865 has a thermal stability of 35°C, with an MRE_{225} = 4330 deg·cm^2·dmol^{-1}, while the same peptide with a Gly to Ala replacement [T1-865(G→A)] had a fully folded triple-helix structure, with a similar MRE_{225} = 4140 deg·cm^2·dmol^{-1} and a decreased Tm of 15.5°C [18]. This peptide design was extended by replacing the Gly with Ser, Arg and Asp (Figure 1B; Table 1). The similarities of peptide T1-865(G→S) to the T1-865(G→A) peptide, with MRE_{225} = 4120 deg·cm^2·dmol^{-1} and a T_m=14.5°C, suggest this peptide has again folded completely around the Gly to Ser substitution. The presence of a Gly to Arg or a Gly to Asp substitution causes the molar ellipticity to decrease to 2920 deg·cm^2·dmol^{-1} for T1-865(G→R) and 1690 deg cm^2 dmol^{-1} for T1-865(G→D) and results in low T_m values (T_m~7.0°C for T1-865(G→R) and T_m~ 5°C for T1-865(G→D)). The low ellipticity and Tm values indicate a substantial loss in helical content in both Gly to Arg and Asp substituted peptides, consistent with the notion that these two peptides are partially folded.

Calorimetric enthalpy of Gly substituted peptides

Enthalpic measurements were carried out by Differential Scanning Calorimetry (DSC) to investigate factors responsible for the decreased thermal stability due to a Gly substitution. A substitution from Gly to Ala or Ser in the $(POG)_{10}$ peptide set results in a decrease in ΔH_{cal} from 390 kJ/mol in $(POG)_{10}$ to 289 and 285 kJ/mol for POG(G→A) and POG(G→S) respectively (Figure 1C; Table 1). Replacement of a Gly to Ala or Ser in the peptide T1-865 set results in a decrease in ΔH_{cal} from 434 kJ/mol for T1-865 to 301 and 299 kJ/mol for T1-865(G→A) and T1-865(G→S) respectively (Figure 1D; Table 1). The replacement of a Gly by Ala or Ser resulted in a similar decreased enthalpy (285–299kJ/mol) in both peptide environments, and this decrease could be due to loss of direct hydrogen bonding within the triple-helix or to effects on the hydration shell. The very low calorimetric enthalpy values observed for Gly to Arg [POG(G→R)] as well as Gly to Arg [T1-865(G→R)] or Gly to Asp [T1-865(G→D)] reflect the partially folded nature of these peptides at low temperature. All transition temperatures observed by DSC are about 8–10°C higher than seen by CD which can be attributed to a faster heating rate [19].

Folding of peptides with Gly substitutions

CD spectroscopy was used to investigate the effect of Gly substitutions on peptide folding (c=1mg/ml, 0°C). None of the folding profiles fit simple 1st, 2nd, or 3rd order kinetics, so empirical fraction folded half times ($t_{1/2}$) were used to compare folding (Table 1). The introduction of a Gly to Ala or Gly to Ser into the $(POG)_{10}$ or T1-865 set both resulted in a significant delay in folding time. The folding half life increased from $t_{1/2}$ = 20 min for $(POG)_{10}$ to $t_{1/2}$ = 68 min for POG(G→A) and 67 minutes for POG(G→S). In the T1-865 peptide system, the values increased from $t_{1/2}$ = 27 min for T1-865 to $t_{1/2}$ = 60 min for T1-865(G→A), and $t_{1/2}$ = 64 min for T1-865(G→S) (Figure 2). Folding cannot be analyzed for peptides POG(G→R), T1-865(G→R) and T1-865(G→D) since they are unable to form completely folded native trimers under these conditions.

The folding was also investigated using DSC, which provides a measure of thermal stability and calorimetric enthalpy for molecular species formed at different folding times [20]. Following denaturation and refolding for different lengths of time, the DSC profiles of the peptide $(POG)_{10}$ and its G→A and G→S homologues only show the final molecular species, with the intensity of the fully triple-helical species increasing with folding time. For instance, DSC measurements of the peptide POG(G→S) collected after 15 minutes of folding show a single species with a transition temperature of 31.0°C and ΔH_{cal} = 110 kJ/mol, while the DSC profile after 50 hours folding time shows a single transition at the same temperature, 31.0°C, but with a significantly higher enthalpy (ΔH_{cal} = 285 kJ/mol) (Figure 3A). These results suggest that both POG(G→A) and POG(G→S), as well as $(POG)_{10}$, form only the fully folded molecule, which continues to populate with time between 15 minutes and 50 hours.

The effect of mutations on folding in the $(POG)_{10}$ system contrasts with that for the T1-865 system, where a less stable species in addition to the fully folded species is seen at early folding times, The parent peptide T1-865 shows a single transition temperature at 42°C with ΔH_{cal} = 394 kJ/mol after 2 hours of folding and a larger peak (ΔH_{cal} = 434 kJ/mol) at 42°C after a 50 hours incubation period (Data not shown). These results show that the amount of completely folded T1-865 species is increasing as a function of time. Both T1-865(G→A) and T1-865(G→S) show a more complicated course of events. At short incubation times a second transition appears at a lower temperature in addition to transition of the fully triple-helical state. At folding times up to 5 hours, both peptides show a biphasic DSC, with a clear low temperature transition at 14.5°C in addition to the transition of the fully folded form at higher temperatures. As shown for peptide T1-865(G→S) (Figure 3B), the disappearance of intensity of the peak at 14.5°C coincides with an increase in intensity in the peak at 19.5°C After 50 hours of incubation, there is just one thermal transition at 20.5°C and 19.5°C respectively, representing the fully folded triple-helical molecules of T1-865(G→A) and T1-865(G→S). This pattern suggests some non-native species with lower stability is slowly being converted to the final native triple-helical form. This species could be an on-pathway or off-pathway intermediate with a relatively long lifetime.

To further investigate the origin of the two transitions observed in the early folding stages, DSC profiles of peptide T1-865(G→S) were compared at concentrations of 1mg/ml and 10 mg/ml after different folding times (Figure 3C). Two methods were used to estimate the kinetics of the two DSC peaks as a function of folding time: (1) monitoring areas of the deconvoluted peaks and (2) plotting the change in maximum value of Cp at the transition temperature for each peak as a function of time. Both methods gave similar results. The population of the lower stability species rises to a maximum followed by a progressive decrease at long folding times, with similar kinetics for both 1mg/ml and 10mg/ml concentrations (Figure 3D). The most stable peak, representing the native fully triple-helical population, shows growth followed by a plateau with increasing folding time. The growth rate of the fully folded species is significantly faster at 10 mg/ml than at 1 mg/ml.

DISCUSSION

The effect of Gly missense mutations on triple-helix properties was studied within two different peptide environments, which differ in imino acid content, stability, and rigidity. These two environments are also likely to differ in the parameters of the superhelix and the tightness of the triple-helix [21–23]. Results reported here indicate the local sequence environment of a Gly replacement affects both the degree of destabilization and the net stability. It was hypothesized that a Gly missense mutation in a highly stable region of the collagen molecule will lead to greater disruption and more severe clinical consequences than the same missense mutation in a less stable region [13]. As reported here, the same mutation

caused a greater loss of thermal stability (ΔT_m) in the more rigid $(POG)_{10}$ context than in T1-865, supporting the proposed relationship between a higher original local stability T_m^{orig} and a greater ΔT_m value. However, the net final stability $T_m^{mut} = T_m^{orig} - \Delta T_m$ at a mutation site was observed to still be greater in the more stable sequence with a mutation than in a less stable region with the same mutation. For example, the Gly to Ser replacement causes a larger relative decrease in T_m in the more stable $(POG)_{10}$ peptide (ΔT_m = 36.5°C) compared with T1-865 (ΔT_m =20.5°C), but the POG(G→S) peptide stability (T_m=20.5°C) is still greater than that of T1-865 (G→S) (T_m=14.5°C) (Fig. 4A). Destabilizing residues such as an Arg mutation appear to result in a similar low stability in both a very stable and a rather unstable sequence, even though there is a much greater decrease in the more stable region (Fig 4B). It is possible that the final net stability at the mutation T_m^{mut} site is a more biologically relevant parameter than the local stability of the original sequence T_m^{orig} or the decrease in stability ΔT_m alone, and could help explain why no correlation has been observed between local stability and clinical severity for the large OI mutation database [4]. It is interesting to note that Gly to Ser mutations, which have T_m^{mut} values sensitive to their local environment in peptides, show differing clinical severity at different locations along the α1(I) chain, while Gly to Arg mutations, which lead to a similar low net stability regardless of the local peptide context, are more severe, except for the N-terminal anchor region [24].

Host guest peptides of the form $(POG)_8$ with a single Gly replaced by all residues available by a single base change indicated that the degree of destabilization depends on the identity of the residue replacing Gly, with the order Gly<Ala<~Ser<Cys<Arg<Val<Glu<Asp, where the chemical chaperone TMAO was used to promote stability [16]. Replacement of Gly by Ala, Ser, and Arg in the $(POG)_{10}$ peptide set and by Ala, Ser, Arg and Asp in the T1-865 peptide set indicate an order of destabilization (Ala<~Ser<Arg<Asp) similar to that reported for the host-guest peptides [16]. These results suggest that the relative order of destabilization for Gly mutations on the triple helix is independent of the peptide environment. When a Gly residue at a specific location is replaced by different amino acids in different OI patients, Val, Asp and Arg are associated with the more severe phenotype [4], an effect that could be related to their greater destabilization. However, the ordering of triple-helix destabilization due to the identity of the residue replacing Gly does not seem to hold for full length collagen chains with Gly mutations, where the degree of destabilization caused by a mutation in collagen appears related to the domain location rather than the identity of the residue replacing Gly [15]. The effect of different residues replacing Gly are exaggerated in short peptides compared to the situation in long collagen molecules, and the effects on a multidomain collagen molecule may be different than on a single domain peptide triple-helix. In addition, all peptides studied here are homotrimers with mutations in all three chains, while dominant OI mutations give rise to collagen molecules where only one or two of the three chains contain a Gly mutation [15]. Recent studies from the Hartgerink laboratory reported formation of heterotrimeric collagen peptides containing 1 or 2 chains with a Gly to Ser replacement, and these results suggested that the first Gly replacement causes the greatest destabilization, with lesser ΔT values seen for subsequent mutations [25,26].

Differences between the two peptide systems studied here highlight the importance of local sequence environment on triple-helix folding in the presence of a Gly mutation. Although OI collagens can form full length triple-helical molecules, it has been reported that folding of collagen molecules is slowed down in fibroblast cultures when there is an OI Gly substitution in the triple-helix [27]. A pause in folding at the mutation site is also suggested by the increased levels of post translational modification of lysine observed N-terminal to the Gly mutation site, since these modifications can only occur on unfolded collagen chains[9]. Both $(POG)_{10}$ and T1-865 peptides can accommodate Gly to Ser and Gly to Ala

substitutions within a full length triple-helix but showed delayed overall folding. For the $(Pro\text{-}Hyp\text{-}Gly)_{10}$ peptides with Gly replacements, only the unfolded peptide or the fully folded molecule are observed; no other molecular species are seen. In the T1-865 peptides, the mutations lead to the appearance of an intermediate species at early folding times which is less stable than the fully folded triple-helix. This suggests the replacement of Gly by a larger residue in T1-865 mutation peptides creates a barrier to propagation which leads to an accumulation of long lived partially folded species. The absence of any observable partially folded species in Gly substituted $(POG)_{10}$ peptide may relate to its optimized sequence for promoting renucleation faster than the detection time of our instruments.

The simplest possible scheme to explain the presence of an intermediate of lower stability for T1-865(G→A) and T1-865(G→S) would be a 3 state model where nucleation at the C-terminus (3^{rd} order) is followed by 1^{st} order propagation up to the Gly substitution site (Figure 5A). The accumulation of the partially folded triple-helix would reflect a slow rate-limiting step necessary to proceed and produce the fully folded triple-helical molecule. Only the first nucleation step (k3) in this simple model is concentration dependent, predicting that increased concentration will lead to a faster loss in monomer species and faster growth of the intermediate species (3^{rd} order reaction), but no change in the 1^{st} order rate of formation of the native state from the intermediate. However, experimental observations show increased concentration leads to no change in the kinetics of monomer to intermediate species but does result in a faster rate of formation of the final native state from the intermediate (Fig. 3D). Previous studies have shown that the $GPO(GAO)_3$ sequence found at the N-terminal of peptide T1-865 can act as a nucleation domain but is less efficient compared with $(GPO)_4$ at the same site[18]. At higher concentrations (e.g. 10mg/ml), it is likely that nucleation more readily occurs at the N-terminal $GPO(GAO)_3$ sequence as well as at the more favorable $(GPO)_4$ C-terminus. The potential for nucleation from either end creates a more complex “diamond” model where the presence of a mutation could lead to either an N-terminal or a C-terminal partially folded species (Figure 5B). At low concentrations, C-terminal nucleation is likely to be favored, but the accumulation of a lower stability species suggests difficulty in completion of the fully folded molecule from this partially folded complex. At higher concentrations, a significant amount of N-terminal nucleation could allow an alternate route which more easily folds through the mutation site to the native form. The presence of a more stable and nucleation favorable $(GPO)_4$ sequence vs. a less favorable $GPO(GAO)_3$ sequence on the other side of mutation site could be a factor in the ease of folding of an N-terminally partially folded species vs. a C-terminally partially folded species. The lack of observable low stability non-native species in the $(POG)_{10}$ peptides with mutations supports the hypothesis that the presence of Pro-Hyp-Gly sequences immediately adjacent to the mutation site facilitates refolding and decreases the potential for accumulating partially folded collagen species.

The studies on mutations in two environments reported here have highlighted the factors that determine the degree of destabilization in homotrimer peptide models, indicating that ΔT will be greater in a more stable local sequence and that ΔT will be greater for Gly to Arg or Asp mutations compared with Gly to Ala or Ser mutations regardless of the local environment. The observation of a folding barrier in the T1-865 peptide system may provide insight into the delay and pause reported for OI collagen folding [9,27]. It is not known whether OI collagens still follow a zipper-like unidirectional folding to a complete triple-helix or have a more complicated pathway. The appearance of a significant amount of long lived partially folded species in the T1-865 mutant peptides, but not the more stable $(POG)_{10}$ mutant peptides, indicates the importance of local environment in the folding mechanism. OI collagens with different mutation sites may have distinctive folding features that influence clinical phenotype.

Acknowledgments

This work was supported by NIH grant GM60048 (BB), an NIH fellowship 5F31 GM072149 (MAB) and a China Scholarship Council (CSC) fellowship (HC). We acknowledge the helpful discussions of Drs. Jean Baum, Angela Mohs, Karunakar Kar, and Anton Persikov.

Abbreviations

(Pro-Hyp-Gly)$_{10}$	hydroxyproline is denoted by Hyp in the 3-letter code and by O in the 1-letter code
OI	Osteogenesis Imperfecta
Tm	melting temperature
Hyp	Hydroxyproline (three-letter code)
O	Hydroxyproline (single-letter code)
CD	Circular dichroism
DSC	Differential Scanning Calorimetry
MRE	mean residue ellipticity
$t_{1/2}$	fraction folded half time

References

1. Myllyharju J, Kivirikko KI. Trends Genet. 2004; 20:33–43. [PubMed: 14698617]
2. Kielty, CM.; Grant, ME. The collagen Family: Structure Assembly and Organization in the Extracellular Matrix. In: Royce, PM.; Steinmann, B., editors. Connective Tissue and its Hereditable Disorders: Molecular, Genetic and Medical Aspects. Wiley-Liss; New York: 2002. p. 159-222.
3. Veit G, Kobbe B, Keene RD, Paulsson M, Koch M, Wagener RJ. Biol Chem. 2006; 281:3494–3504.
4. Marini JC, Forlino A, Cabral WA, Barnes AM, San Antonio JD, Milgrom S, Hyland JC, Korkko J, Prockop DJ, De Paepe A, Coucke P, Symoens S, Glorieux FH, Roughley PJ, Lund AM, Kuurila-Svahn K, Hartikka H, Cohn DH, Krakow D, Mottes M, Schwarze U, Chen D, Yang K, Kuslich C, Troendle J, Dalgleish R, Byers PH. Hum Mutat. 2007; 28:209–221. [PubMed: 17078022]
5. Ramachandran GN, Kartha G. Nature. 1954; 174:269–70. [PubMed: 13185286]
6. Rich A, Crick FHJ. Mol Biol. 1961; 3:483–506.
7. Bella J, Eaton M, Brodsky B, Berman HM. Science. 1994; 266:75–81. [PubMed: 7695699]
8. Shoulders MD, Raines RT. Ann Rev Biochem. 2009; 78:929–958. [PubMed: 19344236]
9. Byers PH. Philos Trans R Soc Lond B Biol Sci. 2001; 356:151–158. [PubMed: 11260795]
10. Byers, PH.; Cole, WG. Osteogenesis Imperfecta. In: Royce, PM.; Steinmann, B., editors. Connective Tissue and its Hereditable Disorders: Molecular, Genetic and Medical Aspects. Wiley-Liss; New York: 2002. p. 385-430.
11. Forlino A, Kuznetsova NV, Marini JC, Leikin S. Matrix Biol. 2007; 26:604–614. [PubMed: 17662583]
12. Xu K, Nowak I, Kirchner M, Xu YJ. Biol Chem. 2008; 283:34337–34344.
13. Bachinger HP, Morris NP, Davis JM. American journal of medical genetics. 1993; 45:152–162. [PubMed: 8456797]
14. Bodian DL, Madhan B, Brodsky B, Klein TE. Biochemistry. 2008; 47:5424–5432. [PubMed: 18412368]
15. Makareeva E, Mertz EL, Kuznetsova NV, Sutter MB, Deridder AM, Cabral WA, Barnes AM, McBride DJ, Marini JC, Leikin SJ. Biol Chem. 2007; 283:4787–4798.
16. Beck K, Chan VC, Shenoy N, Kirkpatrick A, Ramshaw JA, Brodsky B. Proc Natl Acad Sci USA. 2000; 97:4273–4278. [PubMed: 10725403]

17. Liu X, Kim S, Dai QH, Brodsky B, Baum J. Biochemsitry. 1998; 37:15528–15533.
18. Hyde TJ, Bryan MA, Brodsky B, Baum JJ. Biol Chem. 2006; 281:36937–36943.
19. Persikov AV, Xu Y, Brodsky B. Protein Sci. 2004; 13:893–902. [PubMed: 15010541]
20. Miles CA, Bailey AJ. J Mol Biol. 2004; 337:917–931. [PubMed: 15033361]
21. Okuyama K, Wu G, Jiravanichanun N, Hongo C, Noguchi K. Biopolymers. 2006; 84:421–432. [PubMed: 16518844]
22. Kramer RZ, Bella J, Mayville P, Brodsky B, Berman HM. Nat Struct Biol. 1999; 6:454–457. [PubMed: 10331873]
23. Brodsky B, Persikov AV. Adv Protein Chem. 2005; 70:301–339. [PubMed: 15837519]
24. Makareeva E, Cabral WA, Marini JC, Leikin SJ. Biol Chem. 2006; 281:2463–2470.
25. Gauba V, Hartgerink JDJ. Am Chem Soc. 2008; 130:7509–7515.
26. Brodsky B, Baum J. Nature. 2008; 453:998–999. [PubMed: 18563144]
27. Raghunath M, Bruckner P, Steinmann B. J Mol Biol. 1994; 236:940–949. [PubMed: 8114103]

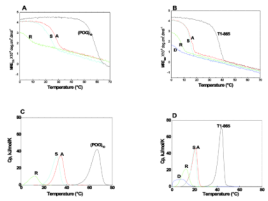

Figure 1.

[A] CD thermal unfolding profiles for $(POG)_{10}$ (black); the POG(G→A) peptide (A, red); the POG(G→S) peptide (S, cyan); and the POG(G→R) peptide (R, green). [B] CD thermal unfolding profiles for T1-865 (black); the T1-865(G→A) peptide (A, red); the T1-865(G→S) peptide (S, cyan); the T1-865(G→R) peptide (R, green); and the T1-865(G→D) (D, blue); [C] DSC transitions of $(POG)_{10}$, (POG)G→A, (POG)G→S, and (POG)G→R, using the same colors and notation described above. [D] DSC transitions of T1-865, T 1-865(G→A), T1-865(G→S), T1-865(G→R), and T1-865(G→D).

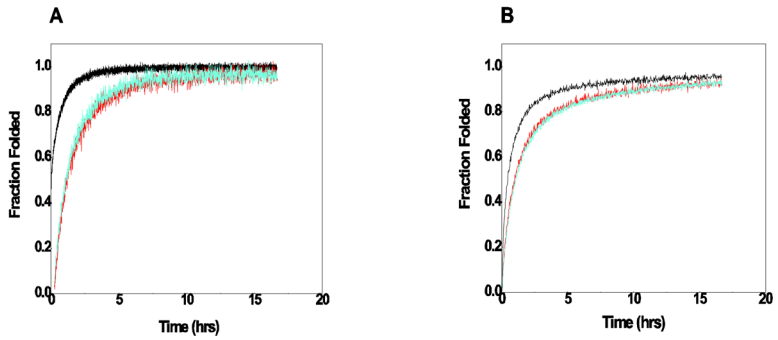

Figure 2.
[A] CD folding curve of $(POG)_{10}$ (black), POG(G→A) (red), and POG(G→S) (cyan). [B] CD folding for T1-865 (black), T1-865(G→A) (red) and T1-865(G→S) (cyan).

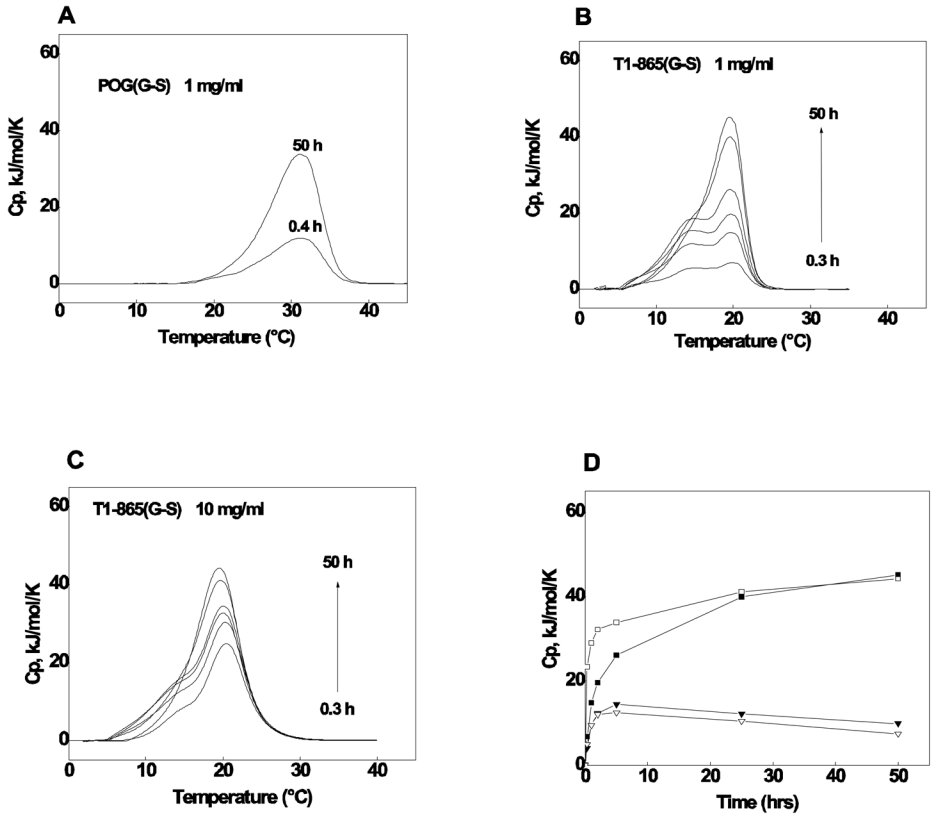

Figure 3.

[A] DSC profiles of POG(G→S) after refolding for 0.4hrs and 50 hrs; [B] DSC profiles of T1-865(G→S) 1mg/ml after folding for different lengths of time, 0.3, 1, 2, 5, 25 and 50 hrs (arrow represents increasing time); [C] DSC profiles of T1-865(G→S) 10 mg/ml after folding for periods of 0.3, 1, 2, 5, 25, and 50 hrs (arrow represents increasing time points); [D] Plot of Cp_{max} data for T1-865(G→S) taken from panels B and C as a function of time for the fully folded native state of 1mg/ml (solid squares) and 10mg/ml (open squares) and for the lower stability species transition at 1mg/ml (solid triangles) and 10mg/ml (open triangles).

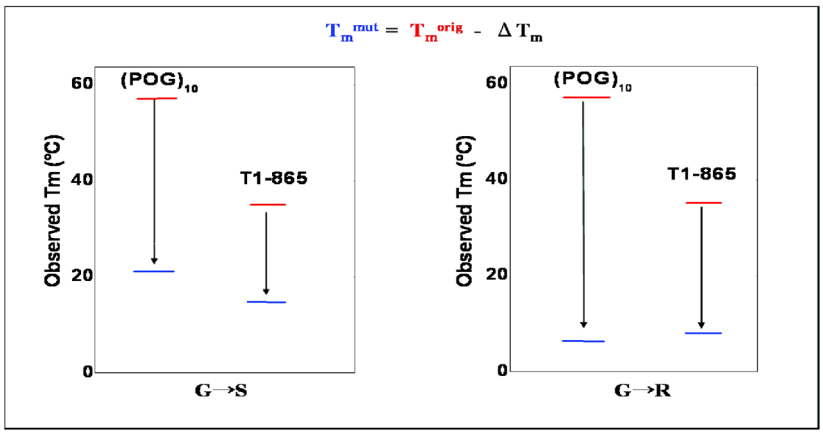

Figure 4.

Comparison of the impact of Gly substitutions on thermal stability (T_m) of the triple helix in different environments. On the left is illustrated the effect of a Gly to Ser replacement on the T_m value of $(POG)_{10}$ and T1-865, showing that the degree of destabilization ΔT_m is greater in the $(POG)_{10}$ context, yet the final stability T_m^{mut} is still higher for (POG)(G→S) than T1-865(G→S). On the right is illustrated the effect of a Gly $_{mut}$ to Arg replacement, show a greater destabilization for $(POG)_{10}$, but very similar T_m final thermal stability for both peptide environments.

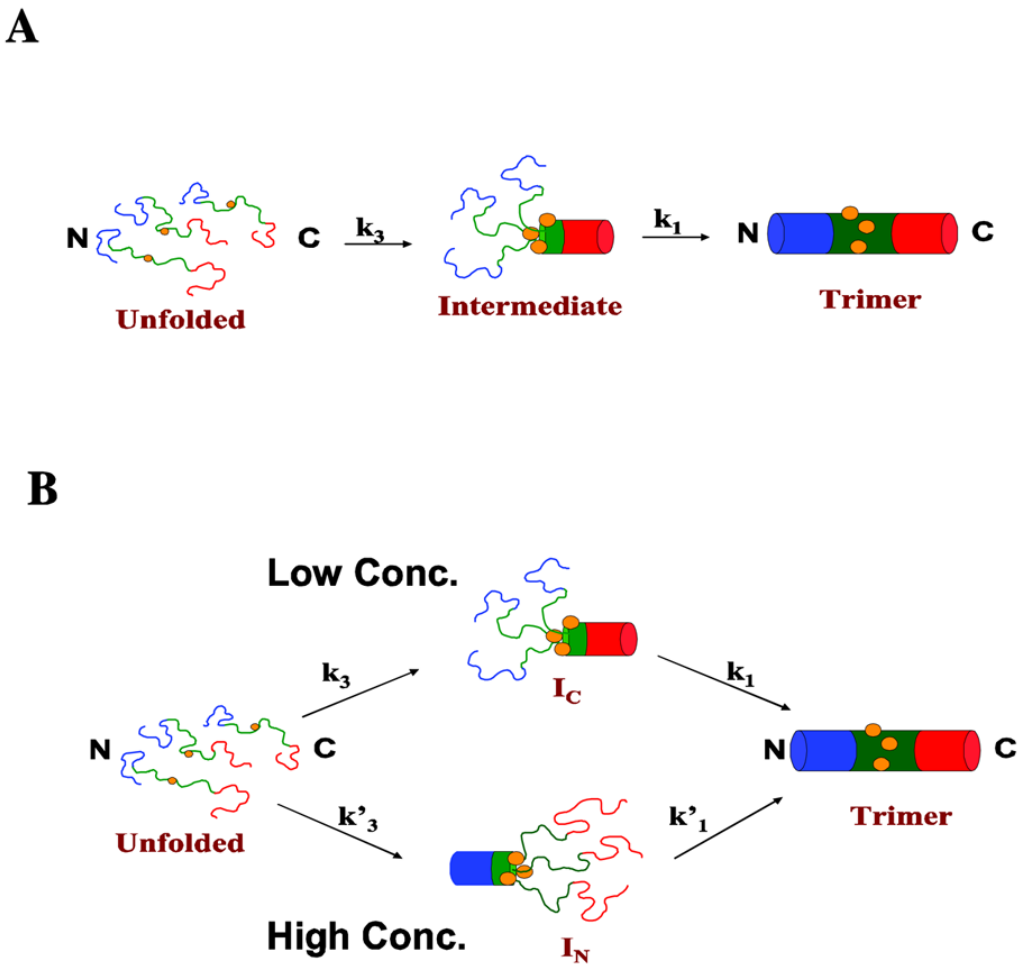

Figure 5.

A) Scheme showing a linear model for folding of T1-865(G→S), nucleated at the C-terminal $(POG)_4$ sequence (red) and folding from the C- to N-terminus. A pause at the site of a Gly substitution could lead to a molecular species which is less stable than the fully folded molecule, which eventually leads to all fully folded species. (B) A more complex scheme where nucleation can be initiated largely or exclusively at the $(POG)_4$ C-terminus (red) at lower concentration, but which could be nucleated significantly at the N-terminus $GPO(GAO)_3$ sequence (blue) at higher concentrations. At high concentrations, it is possible that N-terminal nucleation would occur at significant amounts and lead to fully folded molecules at a faster rate than the C-terminal nucleation.

Table 1

Effects of different Gly substitutions on various collagen like environments, with thermal transition temperature (T_m), mean residue ellipticityat 225nm (MRE_{225}), calorimetric enthalpy (ΔH_{cal}) and half time for folding ($t_{½}$).

Peptide	Sequence	T_m °C	MRE_{225} deg·cm^2·dmol^{-1}	ΔH_{cal} kJ/mol	$t_{½}$ min
$(POG)_{10}$	$(POG)_4PO\underline{G}(POG)_5$	57.0	4340	390	20
POG(G→A)	$(POG)_4PO\underline{A}(POG)_5GY$	26.5	4130	289	68
POG(G→S)	$(POG)_4PO\underline{S}(POG)_5Y$	20.5	4150	285	67
POG(G→R)	$(POG)_4PO\underline{R}(POG)_5GY$	5.5[a]	3100	---[a]	---[a]
T1-865	$GPO(GAO)_3G\underline{P}VG\underline{P}AGAR\ (GPO)_4GY$	35.0	4330	434	27
T1-865(G→A)	$GPO(GAO)_3G\underline{P}V\underline{A}PAGAR\ (GPO)_4GY$	15.5	4140	301	60
T1-865(G→S)	$GPO(GAO)_3GPV\underline{S}PAGAR\ (GPO)_4GY$	14.5	4110	299	64
T1-865(G→R)	$GPO(GAO)_3GPV\underline{R}PAGAR\ (GPO)_4GY$	7.0[a]	2890	---[a]	---[a]
T1-865(G→D)	$GPO(GAO)_3GPV\underline{D}PAGAR\ (GPO)_4GY$	<5.0[a]	1690	---[a]	---[a]

[a]The peptide did not adopt a fully triple-helical state, even at low temperature, so that a meaningful Tm, calorimetric enthalpy, and half time of folding could not be determined.