Published in final edited form as:
Biochemistry. 2011 December 20; 50(50): 10771–10780. doi:10.1021/bi201476a.

Osteogenesis Imperfecta Missense Mutations in Collagen: Structural consequences of a glycine to alanine replacement at a highly charged site

Jianxi Xiao[†,#], **Haiming Cheng**[‡,§,#], **Teresita Silva**[‡], **Jean Baum**[†,*], and **Barbara Brodsky**[‡,¥,*]

[†]Department of Chemistry and Chemical Biology, BIOMAPS Institute, Rutgers University, 610 Taylor Road, Piscataway, NJ 08854, USA

[‡]Department of Biochemistry, UMDNJ-Robert Wood Johnson Medical School, Piscataway, NJ 08854, USA

[¥]Department of Biomedical Engineering, Tufts University, 4 Colby Street, Medford, MA 02446

Abstract

Glycine is required as every third residue in the collagen triple-helix, and a missense mutation leading to the replacement of even one Gly in the repeating $(Gly\text{-}Xaa\text{-}Yaa)_n$ sequence by a larger residue leads to a pathological condition. Gly to Ala missense mutations are highly underrepresented in osteogenesis imperfecta (OI) and other collagen diseases, suggesting that the smallest replacement residue Ala might cause the least structural perturbation and mildest clinical consequences. The relatively small number of Gly to Ala mutation sites that do lead to OI must have some unusual features, such as greater structural disruption due to local sequence environment or location at a biologically important site. Here, peptides are used to model a severe OI case where a Gly to Ala mutation is found within a highly stabilizing Lys-Gly-Asp sequence environment. NMR, CD and DSC studies indicate this Gly to Ala replacement leads to a substantial loss in triple-helix stability and non-equivalence of the Ala residues in the three chains such that only one of the three Ala residues is capable of form a good backbone hydrogen bond. Examination of reported OI Gly to Ala mutations suggests preferential location at known collagen binding sites, and we propose that structural defects due to Ala replacements may lead to pathology when interfering with interactions.

Keywords

CD; NMR; collagen; mutations; Osteogenesis Imperfecta; triple helix

Collagen, the major extracellular matrix structural protein in multicellular animals, is defined by the presence of a triple-helix conformation, formed by supercoiling of 3 polyproline II like helical chains. A striking feature of the triple-helix conformation is the strict amino acid sequence pattern it generates, requiring Gly as every third residue [1-4]. Only Gly, which has no side chain, can pack into the center of the triple-helix structure without distortion. A missense mutation leading to the replacement of even one Gly in the repeating $(Gly\text{-}Xaa\text{-}Yaa)_n$ sequence by a larger residue leads to a pathological condition. For

[*]Correspondence: brodsky@umdnj.edu (B.B.), Tel. 617-627-0447; jean.baum@rutgers.edu (J.B.), Tel 732-445-5666..
[#]Equal Contribution.
[§]Present address: Department of Biomass and Leather Engineering, Sichuan University, China

Supporting Information Available
Figure S1, Figure S2, Table S1 and Table S2. This material is available free of charge via the internet at http://pubs.acs.org.

instance, a Gly missense mutation in the $(Gly\text{-}Xaa\text{-}Yaa)_{338}$ triple helix domain of the α1(I) or α2(I) chains of type I collagen, the major collagen in bone, leads to a hereditary bone disorder, Osteogenesis imperfecta (OI), characterized by fragile bones [5]. Eight possible amino acids (Ala, Ser, Cys, Asp, Glu, Arg, Val, or Trp), as well as a nonsense codon, can be generated by a single base change in the 4 available Gly codons, and all of these residues have been observed as Gly replacements in collagen diseases [6-8]. There is evidence that the identity of the residue replacing Gly influences the extent of triple-helix disruption and the OI clinical phenotype [9, 8, 10].

Triple-helical peptide studies have shown a single Gly substitution in a $(Gly\text{-}Pro\text{-}Hyp)_8$ environment is highly destabilizing, and the identity of the residue replacing Gly was observed to affect the degree of peptide destabilization. The order of stability loss, from least to greatest was Ala ≤ Ser < Cys < Arg < Val < Glu ≤ Asp [11], and this ranking has been confirmed by peptides of other design [12]. In all cases, Ala and Ser were seen to have the least effect on conformation and stability, and in several cases, Ala was less destabilizing than Ser [12]. The peptide data contrast with results on OI collagens expressed in fibroblasts, where the small degree of destabilization observed (1.5-5°C) depends on the location of the mutation site but not the residue replacing Gly [13]. Tthe relatively short length of collagen model peptides (~10 tripeptide units) results in substantial destabilization due to local perturbations at a Gly substitution site strongly, in contrast to type I collagen (338 tripeptide units), where global T_m values show only a small decrease.

The identity of the residue replacing Gly also appears to influence the clinical severity of Osteogenesis Imperfecta. Ser and Cys, the most frequent replacement residues, are found in lethal OI cases less often than expected, while Val, Asp, Glu and Arg occur more often than expected in lethal cases of OI [8, 14]. Most strikingly, Ala is highly underrepresented in reported OI cases, as well as in all other collagen diseases, suggesting that in many cases Gly to Ala changes may result in a mild phenotype that escapes clinical classification [8, 14]. In the current OI database [15, 6, 7], there are 33 Gly to Ala replacements reported (5.7%) out of a total of 581 Gly missense mutations in the α1(I) chain and only 4 of these Gly to Ala mutations lead to the most severe OI clinical phenotype, perinatal lethal type II. With the extensive number of OI mutations reported, there are now a significant number of multiple OI cases due to different mutations occurring at a single Gly residue site in the α1(I) or α2(I) collagen chain [15, 7, 8]; in the large majority of cases, an Ala mutation leads to a less severe OI type than seen for other replacement residues at the same site (Supplement, Table S1). For example, a Gly to Ala at Gly637 leads to type I OI (mildest form) while a Gly to Val mutation at Gly637 leads to type II OI (perinatal lethal) [15].

Previous studies on model peptides indicate that a Gly to Ala replacement results in a major destabilization of the triple-helix, even though the degree of destabilization is less than observed for replacements by larger residues [11, 16]. The high resolution x-ray structure of a peptide with a Gly to Ala substitution within a highly stabilizing and rigid $(Pro\text{-}Hyp\text{-}Gly)_{10}$ context shows a highly localized distortion, with replacement of the 3 standard direct interchain (Gly)NH…CO(Pro) hydrogen bonds by water mediated hydrogen bonds (Ala)NH…H2O…CO(Pro) [17, 1]. In addition, the angular distortion puts the two helical ends out-of- register. NMR studies on this peptide indicate that the Ala residues at the mutation site are non-equivalent, and that the amide of only one of the three Ala residues is capable of form a good hydrogen bond [18].

It is not surprising that replacing Gly by the next smallest residue, Ala, might lead to less structural distortion and to milder clinical consequences than replacements by larger residues. The relatively small number of Gly to Ala mutation sites that do lead to OI must have some features that cause them to generate a clinical phenotype. The drop in thermal

stability due to a Gly to Ala substitution has been shown to depend on the local sequence environment in the peptide, with a larger μT_m seen in a more rigid, Gly-Pro-Hyp sequence than in a more flexible environment [12]. Thus, one possibility is that observed Gly to Ala replacements that lead to OI are located at sites with unusual amino acid sequence environments that lead to a more disruptive outcome. To investigate this hypothesis, peptides were studied containing the sequence of a Gly to Ala mutation at position 658 in the α1(I) chain, which was observed in a severe case of OI. This mutation is located within a highly charged Lys-Gly-Asp (KGD) local sequence environment. The sequences KGD and KGE were previously shown to contribute an unusually high degree of stability to the triple-helix, and computational analysis showed such sequences are present in fibrillar collagens at a frequency much greater than expected [19]. The effects of the Gly to Ala substitution on overall conformational stability of the peptide were characterized by CD and DSC, while the conformation, dihedral angles and hydrogen bonding were determined at specific sites by NMR. Complete folding around the mutant sequence was achieved when a strong stabilizing $(Gly\text{-}Pro\text{-}Hyp)_n$ sequence was included at the N-terminus, as well as the C-terminus. The Ala replacement led to a substantial loss in triple-helix stability and significant perturbations to triple-helix structure and hydrogen bonding. Examination of the locations of known Gly to Ala OI mutation sites indicate they are not localized to regions of high or low triple-helix stability, but a significant number are situated within known collagen binding sites. We suggest that the distinct but relatively minor local structural perturbations at Gly to Ala sites lead to observable clinical outcomes when they disrupt biologically important interactions.

MATERIALS AND METHODS

Peptides

Peptides were purchased from Tufts University Core Facility (Boston, MA) or Biomer (Pleasanton, CA), and was purified on a Shimadzu high pressure liquid chromatography C-18 column (Shimadzu, Columbia, MD). The identity was confirmed by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (Applied Biosystems Voyager DE Pro).

Circular Dichroism Spectroscopy

Circular dichroism (CD) spectra were recorded on an Aviv model 62DS CD spectrophotometer (Aviv Biomedical Inc., Lakewood, NJ). Cells of path length 10 mm were used, and the temperature of the cells was controlled using a Hewlett-Packard Peltier thermoelectric temperature controller. Samples were prepared at a concentration of 1 mg/mL in 20 mM PBS (10 mM NaH_2PO_4, 10 mM Na_2HPO_4, 150 mM NaCl) at pH 7 or in 0.1M acetic acid at pH 2.9. Peptide concentrations were determined by tyrosine absorbance at 275 nm using $\varepsilon^{275} = 1400\ M^{-1}{\cdot}cm^{-1}$. Wavelength scans were collected from 190 nm to 260 nm with a 0.5nm increment per step and a 5s averaging time. CD was applied to determine the thermal stability by monitoring the amplitude of the peak at 224 nm as a function of increasing temperature with an average heating rate of 0.1°C/min [20]. The melting curve is fit to a trimer to monomer transition, and the melting temperature (T_m) is defined as the temperature at which the fraction folded is equal to 0.5.

Differential Scanning Calorimetry

Differential scanning calorimetry (DSC) measurements were performed on a Nano-DSC II, Model 6100 scanning calorimeter (Calorimetry Sciences Corp., Lindon, UT). All the peptides were dialyzed in buffers for 24h before DSC measurements, and their concentrations were determined after dialysis. All DSC profiles were obtained at a scan rate of 1°C/min and each curve was baseline subtracted prior to data analysis.

NMR Spectroscopy

Peptide sets were synthesized with ^{15}N labeled amino acids at selective positions for NMR characterization (Table 1). Peptide T1-645 was selectively ^{15}N-labeled at position G16, and peptide T1-655 was selectively ^{15}N-labeled at positions G7 and G16. Peptides T1-645[G16A] and T1-655[G16A] were labeled at positions G7 and A16. The samples were prepared in 2mM PBS buffer (15mM NaCl) in 10% D_2O/90% H_2O with concentrations of 2.3-4.5 mM at pH 7 or adjusted to pH 3 by addition of hydrochloric acid.

NMR experiments were performed on a Varian INOVA 600 MHz spectrometer equipped with a cold probe (Agilent Technologies, Santa Clara, CA) or a Bruker 600 or 700 MHz spectrometer (Bruker BioSpin Corporation, Billerica, MA). ^1H-^{15}N heteronuclear single quantum coherence (HSQC) spectra [21] were obtained at 0 °C. The 3D HNHA experiments were performed to measure homonuclear $^3J_{HNHa}$ coupling constants at 15 °C, with a H-H coupling period of 25 ms [22]. The correction factor for the $^3J_{HNHa}$ coupling constants was obtained as described [23]. ^{15}N R_1, R_2 and heteronuclear NOE experiments [24-27] were performed at 10 °C. For measurements of amide proton temperature gradients, ^1H-^{15}N HSQC spectra were obtained from 0 °C to 25 °C or 30°C at intervals of 5°C. The samples were equilibrated at each temperature for at least 2 h. Amide proton temperature gradients were obtained by linear regression analysis of the amide proton chemical shifts *versus* temperature.

All hydrogen exchange experiments were performed at 10°C at pD 3, since amide proton exchange rates at pH 7 are too fast to be measured. The pD is corrected for the glass electrode solvent isotope artifact [28]. The peptide was equilibrated at 10°C for 48 h to ensure that the monomer:trimer equilibrium is reached. The sample was then lyophilized, re-dissolved in 100% D_2O and quickly transferred to the spectrometer which was equilibrated at 10°C. A series of HSQC spectra were acquired as described before [29]. The hydrogen exchange rates k_{ex} were determined by a non-linear least squares fit and the protection factor P is calculated as described before [30, 24, 29]. All data were processed using NMRPipe [31] and analyzed with Sparky [32].

RESULTS

Structural characterization of peptide T1-645 with a Gly to Ala mutation

To clarify the effect of Gly to Ala substitutions on triple-helix structure, a peptide was designed to model a Gly to Ala missense mutation at residue 658 in the α1(I) chain, which is reported to lead to a severe OI III case [15]. The immediate surrounding sequence of Gly658 is GAKGDA, and 3 triplets N-terminal to the mutation site is a KGE sequence (Table 1). Host guest peptide studies have shown that such KGD and KGE sequences contribute a very high degree of stabilization to the triple-helix [33, 19]. Initially, as much of the original collagen sequence as possible N-terminal to the mutation site was included in the control peptide, T1-645, with 13 α1(I) residues N-terminal to Gly658, 5 α1(I) residues C-terminal to it, and a stabilizing C-terminal $(GPO)_4$ sequence: Ac-GPO-*GAK-G*EO-GDA-GAK-G*(A*)DA-GPO*-GPO-GPO-GPO-GPO-GY-NH_2. The native collagen sequence is underlined, and the Gly corresponding to residue 658 (G16) was N^{15}-labeled for site specific NMR studies (Table 1). The homologous peptide T1-645[G16A] containing an Ala at position 16 was also synthesized, where the Ala16 residue at the mutation site and Gly7 N-terminal to the mutation are ^{15}N labeled, as denoted by asterisks (Table 1).

The control peptide T1-645 in PBS (pH 7) adopts a triple helix structure as shown by the characteristic CD spectrum with a maximum at 224 nm (MRE_{224}=3950 deg.cm^2.dmol^{-1}) (Table 1). Monitoring the MRE_{224nm} with increasing temperature gives a thermal transition with T_m=31.5°C (Figure 1A), a value close to the T_m=34°C predicted using the collagen

stability calculator [33]. Differential scanning calorimetry (DSC) shows a sharp thermal transition with a value 4 °C higher than seen by CD, due to the faster heating rate [20] (Figure 1C), and indicates a calorimetric enthalpy of 280KJ/M. The homologous peptide with the Gly to Ala replacement, T1-645[G16A], had a significantly reduced MRE_{224} =2310 $deg.cm^2.dmol^{-1}$, suggesting that the introduction of Ala led to loss of a substantial amount of triple-helix structure (Figure 1A). CD studies showed the thermal stability was dramatically reduced from T_m=31.5°C to ~12°C (Table 1). The DSC confirmed a similar drop in stability and indicated a 50% drop in calorimetric enthalpy compared with the parent peptide (Figure 1C).

NMR studies were carried out on both the parent and the mutation peptides of T1-645. The HSQC spectrum of peptide T1-645 at pH 7 contains typical features for a triple helical conformation, with the ^{15}N labeled residue Gly16 showing three trimer resonances in addition to one monomer resonance, suggesting a monomer:trimer equilibrium in the solution (Figure 2). The triple-helical resonances disappear at 45°C, a temperature higher than melting temperatures of the peptides (data not shown). For peptide T1-645[G16A], the HSQC spectrum shows that Gly7 N-terminal to the mutation site has a monomer but no trimer resonances and that Ala16 at the mutation site has only a weak trimer resonance, together with the monomer resonance (Figure 2). These results suggest the peptide adopts a partially folded conformation with a triple-helical C-terminus, a looser or less stable middle region around the mutation site, and an unfolded N-terminal region. Such a partially folded molecule is consistent the observed low ellipiticity, the low stability, and the decreased enthalpy values.

In 0.1M acetic acid (pH 2.9), peptide T1-645 still forms a complete triple helix, MRE_{224nm}=3800 $deg.cm^2.dmol^{-1}$, but has a decreased T_m=22.5°C, 9°C lower than seen for this peptide at pH 7 (Figure 1B). This is consistent with significant electrostatic contributions to stability from the KGD and KGE sequences. At acidic pH, peptide T1-645[G16A] shows a substantial loss in MRE_{224} (1650 $deg.cm^2.dmol^{-1}$) compared with the pH 7 value (2310 $deg.cm^2.dmol^{-1}$), indicating a further decrease in triple-helix content. The residual triple-helix structure showed a thermal transition near T_m=12°C (Figure 1B), a value similar to the thermal transition for T1-645[G16A] at neutral pH. The HSQC spectrum of T1-645 at pH 3 shows peaks still indicative of a triple-helix structure, while the T1-645[G16A] indicates a partially folded structure (data not shown). The similarity of mutant peptide Tm values at neutral and acidic pH would be expected if the charged KGE and KAD sequences are in unfolded regions, consistent with the NMR results.

Structural consequences of a Gly to Ala mutation in peptide T1-655

The failure of peptide T1-645[G16A] to form a completely triple-helical structure could reflect a folding problem, where the peptide had difficulty incorporating an Ala in a KAD context within a triple-helix. To promote folding and stabilization, the parent peptide and its mutant homologue were modified to replace natural α1(I) residues by a $(GPO)_4$ sequence on the N-terminus as well as the C-terminus of the central α1(I) 655-663 sequence, GAKGDAGPO, to form peptide T1-655. In this peptide, ^{15}N-labelled residues were incorporated at positions Gly7 and Gly16. A homologous peptide was made with the Gly to Ala replacement at position 16 to form T1-655[G16A], and ^{15}N-labelled residues were incorporated at positions Gly 7 and Ala16: Ac-GPOGPOG*POGPOGAKA*DAGPO$(GPO)_4$GY-NH_2, (Table 1).

The control T1-655 peptide formed a stable triple-helix with an MRE_{224}=3800 $deg.cm^2.dmol^{-1}$ and a high thermal stability of T_m=55°C (Figure 3A). The excellent agreement of predicted (T_m=55.2°C) and observed T_m values again confirms the high (+17 °C) stabilizing contribution from KGD in this peptide [19]. The homologous peptide

containing the Gly to Ala mutation T1-655[G16A] now formed a fully triple-helical molecule (MRE_{224nm}=3670 deg.cm^2.dmol^{-1}), but with a large reduction in stability, T_m=30 °C (Figure 3A). There was also a drop in calorimetric enthalpy (Table 1).

At acidic pH, the T1-655 peptide still appears fully helical (MRE_{224nm} =3750 deg.cm^2.dmol^{-1}), but its T_m is decreased by 5 °C compared with its value at neutral pH, suggesting that the KGD made an electrostatic contribution of about this amount to the stability (Figure 3B). In 0.1M acetic acid pH 2.9, T1-655[G16A] showed a small decrease in MRE and T_m=~25°C. Thus, at acidic pH, the mutant as well as the parent peptide showed a 5°C decrease in thermal stability compared to the neutral PBS buffer value (Table 1).

NMR characterization of a Gly to Ala mutation in peptide T1-655

NMR studies were carried out the on the T1-655 peptide set (Figure 4). The HSQC spectra of peptide T1-655 at pH 7 (PBS) show only one trimer resonance for Gly7, since it is in a repetitive GPO environment, and three trimer resonances for Gly16, where the 3 different chains have non-equivalent environments due to the one-residue staggering within the triple-helix (Figure 4A). The ^{1}H- ^{15}N heteronuclear NOE values can distinguish resonances due to the ordered triple-helix (positive values) from resonances due to disordered monomers (negative values) (data not shown). Minor resonances observed in the HSQC spectrum arise due to cis-trans isomerization of Gly-Pro and Pro-Hyp bonds in the peptide in the unfolded states [34, 10]. For peptide T1-655[G16A], the HSQC spectrum at pH 7 shows trimer resonances as well as monomer resonances for Gly7 and Ala16. The unique chemical shift of the trimer resonance for Gly7 indicates it adopts a typical $(Gly\text{-}Pro\text{-}Hyp)_n$ triple-helix, and the presence of three Ala16 trimer residues indicates some ordered triple helix structure at the mutation site (Figure 4B). In contrast to peptide T1-645[G16A], the $(Gly\text{-}Pro\text{-}Hyp)_4$ sequence at the N-terminus of T1-655[G16A] allowed the central region with the mutation to be included within neighboring triple-helix domains.

The NMR spectra were also examined at acidic pH, to see the effect of removing the Glu charge within the KGD sequence. In peptide T1-655, at pH 3, the trimer resonances for Gly7 and Gly16 are still observed, but they are weaker with respect to the monomer resonances (Figure 4A). Comparison of the HSQC spectra show that the trimer resonances of Gly7 are completely overlapped at both pHs, indicating no charge effects on the chemical environment at the N-terminus. However, the three trimer resonances of Gly16 all move upfield in the ^{15}N dimension by ~1ppm when the pH changes from neutral to acidic, suggesting a charge effect on the KGD sequence in the center of the triple helix. Overlaying the HSQC spectra of the T1-655[G16A] peptide at pH 7 and pH 3 shows that the trimer and monomer resonances of Gly7 are the same at both pH values, while Ala16 shows pH-dependent trimer chemical shifts (Figure 4B). In particular, ^{2T}A16 shifts ~ 0.5ppm in the proton dimension, which could result from changes in conformation and/or dynamics. In addition, the three trimer resonances of A16 are much less dispersed at pH 3 than pH 7.

The amide proton temperature gradients can indicate the existence of hydrogen bonding for triple helical peptides. A value greater than −4.6 ppb/°C supports the presence of a hydrogen bond [23, 18]. The monomer resonances of Gly7 and Gly16 in peptide T1-655 and T1-655[G16A] at pH 7 show much more negative values than this cut-off value, indicating the absence of hydrogen-bonding (Figure 5B). For T1-655, the NH gradient values for Gly7 and Gly16 in the trimer state (~−3.6 ppb/°C) supports the presence of typical triple-helix hydrogen bonding (Figure 5A). In the mutant peptide, the Gly7 trimer resonance has a value of −3.4 ppb/°C, which is typical for hydrogen bonded Gly residues in the GPO region [23]. However, the amide gradient values for the three trimer resonances of Ala16 in the mutant peptide are very heterogeneous. One Ala, ^{2T}A16, shows an NH gradient value of −2.3 ppb/°C, a value significantly less negative than −4.6 ppb/°C which indicates the formation of a

typical triple-helix amide hydrogen bond. The NH gradient value for ^{3T}A16 (−10.5 ppb/°C) clearly indicates it is not participating in a hydrogen bond. For ^{1T}A16, the value for the NH gradient (−5.1 ppb/°C) is borderline, consistent with a weaker, transient or water mediated hydrogen bond.

At pH 3, the amide gradient studies of peptide T1-655 support hydrogen bond formation for Gly7 and Gly16 trimer resonances, and lack of hydrogen bonding for the monomer resonances, as seen for pH 7 (Figure 5A). The amide gradient values for the mutant peptide again show a Gly7 trimer resonance with a value of −3.2 ppb/°C (pH 3), typical of hydrogen bonding residues (Figure 5B). Similar to the pH 7 results, only one Ala, ^{2T}A16, shows an NH gradient value typical of hydrogen bonding, but at pH 3, this ^{2T}A16 shows a more negative value, changing from −2.3 ppb/°C to −4.4 ppb/°C, indicating that the hydrogen bond is weakened. The value for the NH gradient for ^{1T}A16 became more negative at pH 3 (−6.1 ppb/°C), suggesting this borderline hydrogen bond may now be disrupted. These results suggest that lowering the pH from 7 to 3, leads to conformational changes that weaken the remaining hydrogen bonds involving Ala at the mutation site, but do not affect the hydrogen bonds of Gly in the N-terminal GPO region.

$^{3}J_{HNH\alpha}$ coupling constants that can be related to the dihedral angle φ were obtained for the trimer peaks of peptides T1-655 and T1-655[G16A] from HNHA experiments at pH 7 (Supplement, Figure S1). Residues in the triple helical conformation typically contain phi angles from −55 to −75 degrees and have a corresponding J coupling value of 4 to 6 Hz [29]. Labeled residues Gly7 and Gly16 in peptide T1-655 show J coupling values around 4 Hz, supporting a standard triple-helix conformation throughout the chain at both pH values. Similarly, Gly7 in peptide T1-655[G16A], shows a J-coupling value of ~4 Hz. In addition, the Ala16 residues in all three chains show J-coupling values between 5 to 6 Hz at neutral pH, suggesting that all three Ala at the mutation site have dihedral angles similar to a typical triple-helix structure when in the current KAD context. Similar results were observed at acid pH, except that for the mutant peptide, the Ala16 trimer resonances in two chains (^{1T}A16 and ^{2T}A16) show small increased values of J-couplings going from ~5.4 to ~6 Hz, suggesting the possibility of some alteration in the triple helix structure (Supplement, Figure S1).

Hydrogen bonding can also be measured using protection factors derived from hydrogen exchange experiments [24, 29]. All hydrogen exchange experiments were performed on peptide T1-655[G16A] at pH 3 at 10°C, since amide proton exchange rates at pH 7 are too fast to be measured. The high protection factor observed for Gly7 at pH 3 indicates that this residue forms hydrogen bonds within the triple-helix and is well protected from solvents (Supplement, Figure S2). The Ala16 residues show much less protection than G7, with varying degrees of protection among the Ala in the 3 chains (Supplement, Figure S2). ^{2T}A16 shows more protection than the other two, consistent with the amide proton chemical shift coefficients suggesting that ^{2T}A16 is the only Ala capable of forming a reasonable hydrogen bond at pH 3.

^{15}N Relaxation experiments R_1, R_2 and NOE are performed for both peptides to evaluate if the mutation site experiences dynamic changes at the fast, pico-second timescale at different pH values (data not shown). The labeled residues G7 and Gly16/A16 in the trimer states for both peptides have high NOE values of >~0.6 at pH 7 as well as pH 3, indicating that they are rigid on the pico-second timescale.

DISCUSSION

The significant underrepresentation of Gly to Ala mutations in OI suggests that some of these missense mutations in type I collagen chains may not come to clinical attention because of a mild phenotype or late clinical onset [14]. Such an argument is favored by the observation of single Gly to Ala substitutions within otherwise perfect $(Gly\text{-}Xaa\text{-}Yaa)_n$ repeats in several invertebrate fibril-forming collagens [35-37]. If a subset of Gly to Ala missense mutations in the α1(I) or α2(I) chain result in an OI clinical phenotype, there may be definable features that predispose this subset to disruption or pathology. The environment of the mutation could be a predisposing factor, since peptide studies showed the degree of destabilization depends on the local sequence around a mutation [12]. Previously, the presence of a KGE sequence near an OI missense mutation was implicated in structural disruption and clinical severity [38], suggesting that the proximity of the highly stabilizing KGE and KGD charged sequences may be important. Structural studies on a peptide containing a Gly to Ala substitution at a KGD site are presented here.

A Gly to Ala mutation within a highly stabilizing charge sequence KGD presents an energy barrier to complete triple-helix folding of a peptide containing a natural α1(I) chain N-terminal sequence. This barrier was overcome by placing a strong folding/nucleation sequence $(Gly\text{-}Pro\text{-}Hyp)_4$ at the N-terminus as well as the C-terminus of the peptide. In type I collagen, triple-helix folding proceeds from the C-terminus to the N-terminus, and there is a long $(Gly\text{-}Xaa\text{-}Yaa)_n$ sequence N-terminal to the mutation at residue 658 in the triple-helix of the α1(I) chain. The presence of a $(Gly\text{-}Pro\text{-}Hyp)_4$ sequence in peptide T1-655 may compensate for the very small number of triplets N-terminal to the mutation in these model peptides. The good agreement between predicted and experimentally observed stability for both the T1-645 and T1-655 peptides indicate that KGD and KGE sequences are making the same large contribution to stability in these α1(I) sequence environments as found when they are isolated sequences in a $(Gly\text{-}Pro\text{-}Hyp)_8$ environment in host-guest peptides, which form the basis for stability predictions [33, 19]. Since the mutation is located within the charged KGD sequence, features were characterized at acidic and neutral pH values, to clarify electrostatic contributions to stability. The difference in T_m observed for acidic vs. neutral pH is much smaller than the total stabilizing contribution of KGD or KGE (~5°C vs. 15-17°C), indicating that stabilizing factors in addition to electrostatic interactions are also arising from the Lys/Asp residues. Other stabilizing interactions involving the Lys and Asp side chains could include hydrogen bonding between side chains, backbone carbonyls or water, which may be less sensitive to pH. When the T_m at pH 7 is compared with the value for pH 3, a $\square T_m$ of ~5 °C is seen for the parent peptide ($\square T_m$= 55°C − 50 °C) as well as for the peptide with the mutation ($\square T_m$= 30°C − 25 °C). This suggests that the electrostatic contribution of the KAD sequence is similar to that of the KGD sequence, even though it may have less favorable hydrogen bonding or hydration effects. Low pH was seen to affect structural features in the mutant peptide: the three trimer resonances of A16 are less dispersed and the one hydrogen bond involving ^{2T}A16 appears weakened. Even though all dihedral angles fall within the normal triple-helix range for the mutant peptide, the Ala16 trimer resonances in two chains (^{1T}A16 and ^{2T}A16) show small increased values of J-couplings at low pH, going from ~5.4 to ~6 Hz, consistent with some changes in conformation or flexibility near the mutation site.

NMR studies on specifically labeled residues within T1-655[G16A] peptide showed the Gly7 near the N-terminus is in a standard triple-helix but that the Ala16 residues at the mutation site are perturbed. The amide displacement, supported by hydrogen exchange studies, indicates that only one of the 3 Ala residues forms a strong hydrogen bond while a second Ala may form a weaker or water mediated hydrogen bond. Despite the perturbations in hydrogen bonding, J-coupling values indicate the dihedral angles for the Ala16 residues

did not vary significantly from those expected for a standard triple-helix. A previous study on a Gly to Ala substitution within a $(Pro\text{-}Hyp\text{-}Gly)_{10}$ peptide showed one of the three Ala residues had a J-coupling of 8.6 Hz, which is not compatible with a triple-helix structure [18], while all three Ala seem to be able to participate in a triple-helix structure in the current KAD context. Both Ala and Gly were found to be rigid on picoseconds time scale, suggesting any perturbations are occurring on a much longer time scale. The changes due to the Gly to Ala replacement in the KGD context are similar to those observed previously in the context of other amino acid sequences, including the most rigid Gly-Pro-Hyp context and an uncharged α1(I) chain sequence (Supplement, Table S2).

Analysis of known Gly to Ala mutations shows they are distributed all along the α1(I) chain and the small number of mutations in the α2(I) chains do not appear clustered (Figure 6; See Supplement Table S1). Examination of the relative stability along the type I triple-helix, using the collagen stability calculator [33], indicates that Gly to Ala missense mutations are found in regions of low, middle, and high stability (Figure 6). Therefore, stability considerations alone cannot predict which Gly to Ala substitutions will lead to a clinical phenotype. Different window sizes were used for the relative stability calculation to see whether immediate tripeptide residues or longer range factors could be important, but no correlation was found between the severity of OI and the local stability of the Gly to Ala mutations.

It has been suggested that Gly missense mutations that disrupt type I collagen binding sites may lead to particularly severe phenotypes or even embryonic lethality [8, 39], and it is plausible that the structural perturbations caused by a Gly to Ala mutation in model peptides could interfere with a collagen interaction site. A number of the Gly to Ala mutations occur in proximity to known interaction sites (Supplement, Table S1). Gly220Ala, which is observed in two independently reported cases, is part of a DDR2 binding site as determined by Toolkit peptide analysis [40, 41]; Gly928Ala is near the known Lys/Hyl cross-linking site [42]; Gly409Ala is C-terminal to a site which binds DDR2, vWF, and SPARC [42, 40, 43, 41]; and Gly304Ala is at an α1β1 integrin binding site [40, 44, 39]. The number of Gly to Ala missense mutations is also highly underrepresented in type III collagen, where there are only 2 out of 104 cases of Gly missense mutations leading to EDS-IV, at positions Gly805Ala and Gly937Ala [15]. Interestingly, the Gly805 site was in a peptide identified as a binding site for LAIR-1 and LAIR-2 [45], while the Gly937 is adjacent to the known cross-linking site KGHR [42].

The number of recognizable interaction sites located at Gly to Ala mutations suggests that the relatively small distortion caused by Ala replacements may lead to disease when they interfere with binding and biological activity. The site of a Gly to Ala mutation would represent a distorted triple-helix with disruption of the screw symmetry of the 3 chains and significant loss of hydrogen bonding. These features are likely to affect the ability of the triple-helix to interact with cell receptors, enzymes or other matrix components. These mutations would thus represent the result of a natural Ala scanning mutagenesis experiment, where clinical phenotype tags those Gly residues necessary for essential interactions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grants from the National Institute of Health (GM45302 to J.B. and GM60048 to B.B.), and the National Science Foundation (DBI-0403062 and DBI-0320746 to J.B.).

ABBREVIATIONS

Hydroxyproline	Hyp in the 3-letter code and by O in the 1-letter code
OI	Osteogenesis Imperfecta
Tm	melting temperature
CD	Circular dichroism
DSC	Differential Scanning Calorimetry
MRE	mean residue ellipticity

REFERENCES

1. Bella J, Eaton M, Brodsky B, Berman HM. Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. Science. 1994; 266:75–81. [PubMed: 7695699]
2. Kielty, CM.; Grant, ME. The collagen family: strucutre, assembly, and organization in the extracellular matrix. In: Royes, PM.; Steinmann, BU., editors. Connective tissue and its heritable disorders, molecular, genetic and medical aspects Edited by. Wiley Liss; New York: 2002. p. 159-222.
3. Ramachandran, GN. Structure of Collagen at the Molecular Level. In: Ramachandran, GN., editor. Treatise on Collagen Edited by. Academic Press; New York: 1967. p. 103-184.
4. Rich A, Crick FH. The molecular structure of collagen. J Mol Biol. 1961; 3:483–506. [PubMed: 14491907]
5. Byers, PH.; Cole, WG. Osteogenesis Imperfecta. In: Royce, PM.; Steinmann, B., editors. Connective tissue and its hereditable disorders Edited by. Wiley-Liss; New York: 2002. p. 385-430.
6. Dalgleish R. The human type I collagen mutation database. Nucleic Acids Res. 1997; 25:181–187. [PubMed: 9016532]
7. Dalgleish R. The Human Collagen Mutation Database 1998. Nucleic Acids Res. 1998; 26:253–255. [PubMed: 9399846]
8. Marini JC, Forlino A, Cabral WA, Barnes AM, San Antonio JD, Milgrom S, Hyland JC, Korkko J, Prockop DJ, De Paepe A, et al. Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. Hum Mutat. 2007; 28:209–221. [PubMed: 17078022]
9. Bodian DL, Madhan B, Brodsky B, Klein TE. Predicting the clinical lethality of osteogenesis imperfecta from collagen glycine mutations. Biochemistry. 2008; 47:5424–5432. [PubMed: 18412368]
10. Xiao J, Madhan B, Li Y, Brodsky B, Baum J. Osteogenesis imperfecta model peptides: incorporation of residues replacing Gly within a triple helix achieved by renucleation and local flexibility. Biophysical Journal. 2011; 101:449–458. [PubMed: 21767498]
11. Beck K, Chan VC, Shenoy N, Kirkpatrick A, Ramshaw JA, Brodsky B. Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. Proc Natl Acad Sci U S A. 2000; 97:4273–4278. [PubMed: 10725403]
12. Bryan MA, Cheng H, Brodsky B. Sequence environment of mutation affects stability and folding in collagen model peptides of osteogenesis imperfecta. Biopolymers. 2011; 96:4–13. [PubMed: 20235194]
13. Makareeva E, Mertz EL, Kuznetsova NV, Sutter MB, Deridder AM, Cabral WA, Barnes AM, Mcbride DJ, Marini JC, Leikin S. Structural heterogeneity of type I collagen triple helix and its role in osteogenesis imperfecta. J Biol Chem. 2008; 283:4787–4798. [PubMed: 18073209]
14. Persikov AV, Pillitteri RJ, Amin P, Schwarze U, Byers PH, Brodsky B. Stability related bias in residues replacing glycines within the collagen triple helix (Gly-Xaa-Yaa) in inherited connective tissue disorders. Hum Mutat. 2004; 24:330–337. [PubMed: 15365990]
15. Dalgleish, R. Osteogenesis imperfecta & Ehlers-Danlos syndrome variant databases. Leiden University Medical Center; http://www.le.ac.uk/genetics/collagen/

16. Long CG, Braswell E, Zhu D, Apigo J, Baum J, Brodsky B. Characterization of collagen-like peptides containing interruptions in the repeating Gly-X-Y sequence. Biochemistry. 1993; 32:11688–11695. [PubMed: 8218237]

17. Bella J, Brodsky B, Berman HM. Hydration structure of a collagen peptide. Structure. 1995; 3:893–906. [PubMed: 8535783]

18. Li Y, Brodsky B, Baum J. NMR conformational and dynamic consequences of a gly to ser substitution in an osteogenesis imperfecta collagen model Peptide. J Biol Chem. 2009; 284:20660–20667. [PubMed: 19451653]

19. Persikov AV, Ramshaw JA, Kirkpatrick A, Brodsky B. Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. Biochemistry. 2005; 44:1414–1422. [PubMed: 15683226]

20. Persikov AV, Xu Y, Brodsky B. Equilibrium thermal transitions of collagen model peptides. Protein Sci. 2004; 13:893–902. [PubMed: 15010541]

21. Kay LE, Keifer P, Saarinen T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J. Am. Chem. Soc. 1992; 114:10663–10665.

22. Vuister GW, Bax A. Quantitative J correlation: a new approach for measuring homonuclear three-bond J(H^NH^A) coupling constants in ^{15}N-enriched proteins. J Am Chem Soc. 1993; 115:7772–7777.

23. Li Y, Brodsky B, Baum J. NMR shows hydrophobic interactions replace glycine packing in the triple helix at a natural break in the (Gly-X-Y)n repeat. J Biol Chem. 2007; 282:22699–22706. [PubMed: 17550894]

24. Fan P, Li MH, Brodsky B, Baum J. Backbone dynamics of (Pro-Hyp-Gly)10 and a designed collagen-like triple-helical peptide by 15N NMR relaxation and hydrogen-exchange measurements. Biochemistry. 1993; 32:13299–13309. [PubMed: 8241186]

25. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. Biochemistry. 1994; 33:5984–6003. [PubMed: 7514039]

26. Palmer AG 3rd. Dynamic properties of proteins from NMR spectroscopy. Curr Opin Biotechnol. 1993; 4:385–391. [PubMed: 7763967]

27. Xiao J, Baum J. Structural insights from (15)N relaxation data for an anisotropic collagen peptide. J Am Chem Soc. 2009; 131:18194–18195. [PubMed: 19954183]

28. Glasoe PK, Long FA. Use of glass electrodes to measure acidities in deuterium oxide. J. Phys. Chem. 1960; 64:188–189.

29. Xiao J, Addabbo RM, Lauer JL, Fields GB, Baum J. Local conformation and dynamics of isoleucine in the collagenase cleavage site provides recognition signal for matrix metalloproteinases. J Biol Chem. 2010; 285:34181–34190. [PubMed: 20679339]

30. Bai Y, Milne JS, Mayne L, Englander SW. Primary structure effects on peptide group hydrogen exchange. Proteins. 1993; 17:75–86. [PubMed: 8234246]

31. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995; 6:277–293. [PubMed: 8520220]

32. Goddard, TD.; Kneller, DG. Sparky - NMR Assignment and Integration Software. University of California; San Francisco: http://www.cgl.ucsf.edu/home/sparky/

33. Persikov AV, Ramshaw JA, Brodsky B. Prediction of collagen stability from amino acid sequence. Journal of Biological Chemistry. 2005; 280:19343–19349. [PubMed: 15753081]

34. Buevich AV, Dai QH, Liu X, Brodsky B, Baum J. Site-specific NMR monitoring of cis-trans isomerization in the folding of the proline-rich collagen triple helix. Biochemistry. 2000; 39:4299–4308. [PubMed: 10757978]

35. Mann K, Gaill F, Timpl R. Amino-acid sequence and cell-adhesion activity of a fibril-forming collagen from the tube worm Riftia pachyptila living at deep sea hydrothermal vents. Eur J Biochem. 1992; 210:839–847. [PubMed: 1483468]

36. Sicot FX, Mesnage M, Masselot M, Exposito JY, Garrone R, Deutsch J, Gaill F. Molecular adaptation to an extreme environment: origin of the thermal stability of the pompeii worm collagen. J Mol Biol. 2000; 302:811–820. [PubMed: 10993725]
37. Yoneda C, Hirayama Y, Nakaya M, Matsubara Y, Irie S, Hatae K, Watabe S. The occurrence of two types of collagen proalpha-chain in the abalone Haliotis discus muscle. Eur J Biochem. 1999; 261:714–721. [PubMed: 10215888]
38. Xu K, Nowak I, Kirchner M, Xu Y. Recombinant collagen studies link the severe conformational changes induced by osteogenesis imperfecta mutations to the disruption of a set of interchain salt bridges. J Biol Chem. 2008; 283:34337–34344. [PubMed: 18845533]
39. Sweeney SM, Orgel JP, Fertala A, Mcauliffe JD, Turner KR, Di Lullo GA, Chen S, Antipova O, Perumal S, Ala-Kokko L, et al. Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. J Biol Chem. 2008; 283:21187–21197. [PubMed: 18487200]
40. Farndale RW, Lisman T, Bihan D, Hamaia S, Smerling CS, Pugh N, Konitsiotis A, Leitinger B, De Groot PG, Jarvis GE, et al. Cell-collagen interactions: the use of peptide Toolkits to investigate collagen-receptor interactions. Biochem Soc Trans. 2008; 36:241–250. [PubMed: 18363567]
41. Konitsiotis AD, Raynal N, Bihan D, Hohenester E, Farndale RW, Leitinger B. Characterization of high affinity binding motifs for the discoidin domain receptor DDR2 in collagen. J Biol Chem. 2008; 283:6861–6868. [PubMed: 18201965]
42. Di Lullo GA, Sweeney SM, Korkko J, Ala-Kokko L, San Antonio JD. Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. Journal of Biological Chemistry. 2002; 277:4223–4231. [PubMed: 11704682]
43. Giudici C, Raynal N, Wiedemann H, Cabral WA, Marini JC, Timpl R, Bachinger HP, Farndale RW, Sasaki T, Tenni R. Mapping of SPARC/BM-40/osteonectin-binding sites on fibrillar collagens. Journal of Biological Chemistry. 2008; 283:19551–19560. [PubMed: 18487610]
44. Raynal N, Hamaia SW, Siljander PR, Maddox B, Peachey AR, Fernandez R, Foley LJ, Slatter DA, Jarvis GE, Farndale RW. Use of synthetic peptides to locate novel integrin alpha2beta1-binding motifs in human collagen III. J Biol Chem. 2006; 281:3821–3831. [PubMed: 16326707]
45. Lebbink RJ, Raynal N, De Ruiter T, Bihan DG, Farndale RW, Meyaard L. Identification of multiple potent binding sites for human leukocyte associated Ig-like receptor LAIR on collagens II and III. Matrix Biol. 2009; 28:202–210. [PubMed: 19345263]

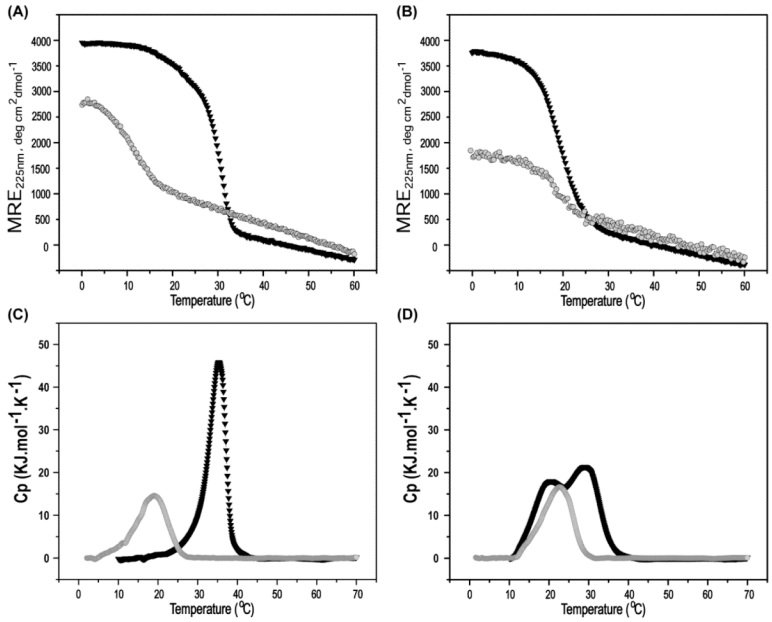

Figure 1.

Thermal transitions of peptide set T1-645 with a Gly to Ala mutation. (A) CD thermal transition of T1-645 (black) and T1-645[G16A] (gray) in 20mM PBS, pH 7.0; (B) CD thermal transition of T1-645 (black) and T1-645[G16A] (gray) in 0.1M acetic acid pH 2.9; (C) DSC profile of T1-645 (black) and T1-645[G16A] (gray) in 20mM PBS, pH 7.0; (D) DSC profile of T1-645 (black) and T1-645(G16A) (gray) in 0.1M acetic acid, pH 2.9.

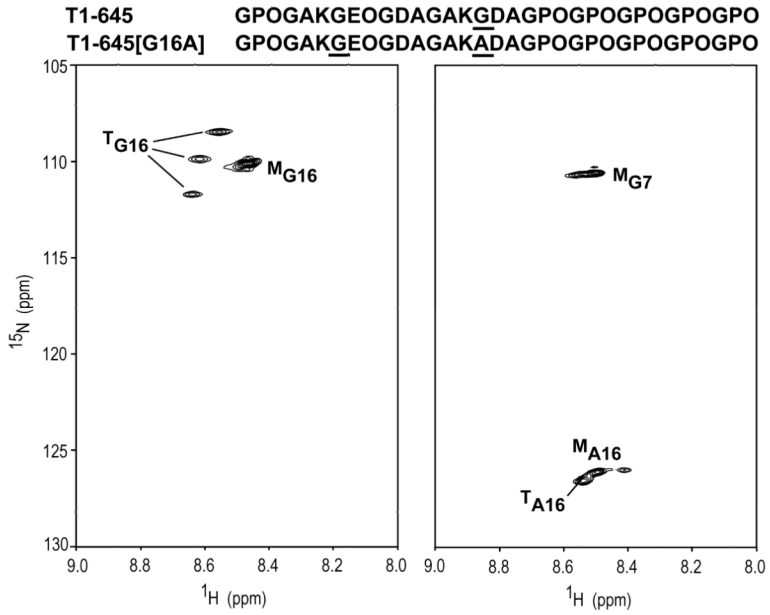

Figure 2.

HSQC spectra of peptides T1-645 (left) and T1-645[G16A] (right) at pH 7, in 2mM PBS and 15mM NaCl at 0°C. Peptide sequences are shown at the top with ^{15}N labeled residues underlined. The peaks corresponding to the monomer and trimer states are denoted with a superscript M or T, respectively.

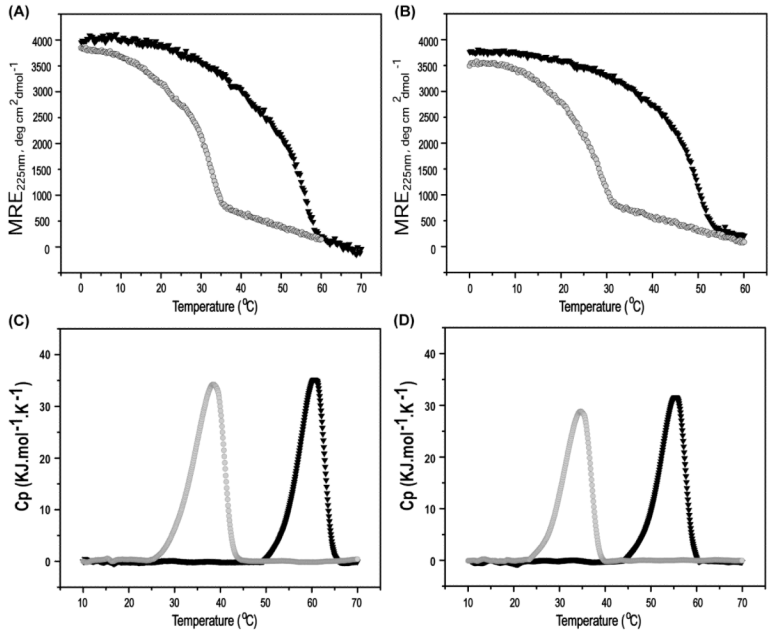

Figure 3.

Thermal stability of the Gly to Ala mutation peptide T1-655 set. **(A)** CD thermal transition of T1-655 (black) and T1-655[G16A] (gray) in PBS pH 7.0; (B) CD thermal transition of T1-655 (black) and T1-655[G16A] (gray) in 0.1M acetic acid pH 2.9; (C) DSC profile of T1-655 (black) and T1-655[G16A] (gray) in PBS pH 7.0; and (D) DSC profile of of T1-655 (black) and T1-655[G16A] (gray) in 0.1M acetic acid pH 2.9.

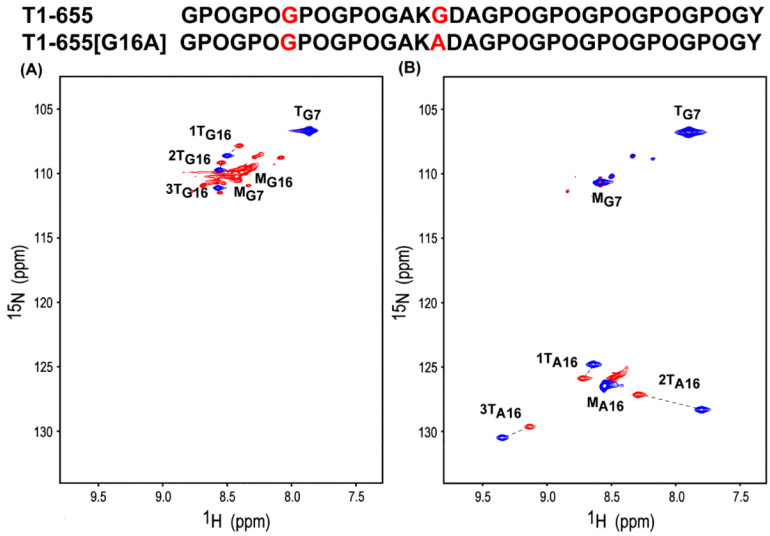

Figure 4.

NMR studies of peptide set T1-655. Peptide sequences are shown at the top with ^{15}N labeled residues in red color. (A) Overlapped ^{1}H-^{15}N HSQC spectra of peptide T1-655 at pH 7 (blue) and pH 3 (red) at 20°C; (B) Overlapped^{1}H-^{15}N HSQC spectra of peptide T1-655[G16A] at pH 7 (blue) and pH 3 (red) at 0°C. The peaks corresponding to the monomer and trimer states are denoted with a superscript M or T, respectively. Minor monomer resonances arise due to cis-trans isomerization in the unfolded state of the Pro/Hyp-rich chains [34].

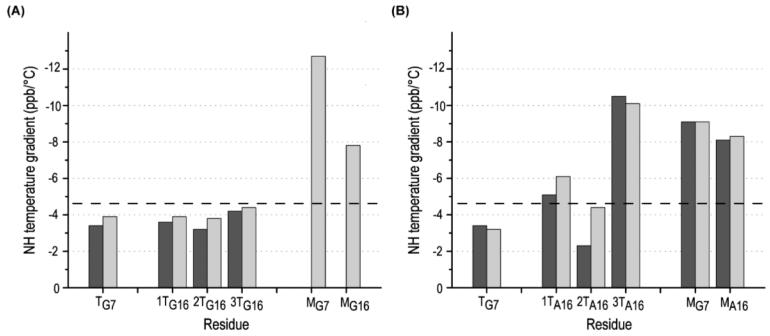

Figure 5.

Amide proton NH temperature gradients of peptide set T1-655. (A) Amide proton NH temperature gradient (ppb/°C) of peptide T1-655 at pH 7 (black) and pH 3 (gray); (B) Amide proton NH temperature gradient (ppb/°C) of peptide T1-655[G16A] at pH 7 (black) and pH 3 (gray); The black dashed horizontal line corresponds to a value of -4.6 ppb/°C, a cut-off for hydrogen bonding, with less negative values indicative of hydrogen bonding.

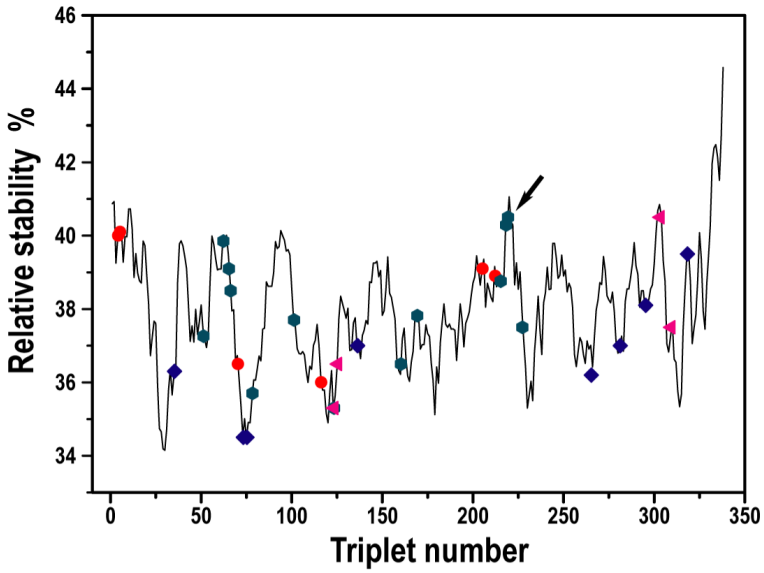

Figure 6. Stability calculation plots with Gly to Ala mutation in COL1A1

The horizontal axis is triplet number from N-terminal to C-terminal (from 0 to 338). The Silence classification of OI clinical phenotype is color coded: Red for mild OI I; blue diamonds for mild/moderate OI IV; green hexagons for severe OI III and pink triangles for lethal OI II. The site of the mutation modeled in peptides is indicated by an arrow.

Table 1

Peptides modeling the Gly to Ala substitution at residue 658 in the α1(I) chain at pH 7.0 and pH 2.9, showing Thermal transition temperature (T_m), Mean residue ellipticity at 224 nm (MRE_{224}), and Calorimetric enthalpy (ΔH_{cal}).

Peptide name	Peptide sequence*	pH	Tm (°C)	MRE_{224}(deg.cm^2·dmol^{-1})	Enthalpy (KJ/mol)
T1-645	Ac-GPO-GAK-GEO-GDA-GAK-GDA-GPOGPO-GPO-GPO-GPO-GY-$CONH_2$	7.0	31.5	3950	280
		2.9	22.5	3800	-†
T1-645[G16A]	Ac-GPO-GAK-GEO-GDA-GAK-ADA-GPOGPO-GPO-GPO-GPO-GY-$CONH_2$	7.0	12	2310	140
		2.9	12	1650	150
T1-655	Ac-GPO-GPO-GPO-GPO-GAK-GDA-GPOGPO-GPO-GPO-GPO-GY-$CONH_2$	7.0	55	3800	320
		2.9	50	3750	270
T1-655[G16A]	Ac-GPO-GPO-GPO-GPO-GAK-ADA-GPOGPO-GPO-GPO-GPO-GY-$CONH_2$	7.0	30	3670	285
		2.9	25	3530	202

*^{15}N labeled residues are underlined and mutated residue is boxed.

†Two transitions were seen for T1-645 at pH 2.9