BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

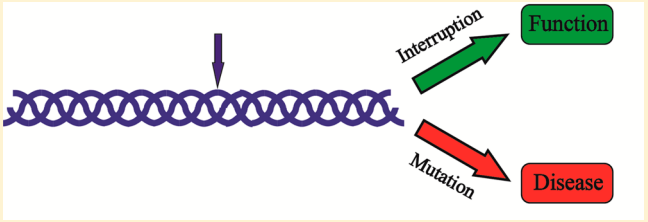
A Natural Interruption Displays Higher Global Stability and Local Conformational Flexibility than a Similar Gly Mutation Sequence in Collagen Mimic Peptides

Xiuxia Sun,[†,#] Yalin Chai,[†,#] Qianqian Wang,[‡] Huanxiang Liu,[‡] Shaoru Wang,[∥] and Jianxi Xiao*[,†]

[†]Key Laboratory of Nonferrous Metal Chemistry and Resources Utilization of Gansu Province, State Key Laboratory of Applied Organic Chemistry, College of Chemistry and Chemical Engineering and [‡]School of Pharmacy, Lanzhou University, Lanzhou 730000, P. R. China

[∥]College of Chemistry and Molecular Sciences, Wuhan University, Wuhan 430072, P. R. China

ABSTRACT: Natural interruptions in the repeating $(Gly\text{-}X\text{-}Y)_n$ amino acid sequence pattern are found normally in triple helix domains of all nonfibrillar collagens, while any Gly substitution in fibrillar collagens leads to pathological conditions. As revealed by our sequence analysis, two peptides, one modeling a natural G5G interruption (POALO) and the other one mimicking a pathological Gly-to-Ala substitution (LOAPO), are designed. Circular dichroism (CD), NMR, and computational simulation studies have discovered significant differences in stability, conformation, and folding between the two peptides. Compared with the Gly substitution sequence, the natural interruption maintains higher stability, higher triple helix content, and a higher folding rate while introducing more alterations in local triple helical conformation in terms of dihedral angles and hydrogen bonding. The conserved hydrophobic residues at the specific sites of interruptions may provide functional constraints for higher-order assembly as well as biomolecular interactions. These results suggest a molecular basis of different biological roles of natural interruptions and Gly substitutions and may guide the design of collagen mimic peptides containing functional natural interruptions.



Natural interruptions in the repeating $(Gly\text{-}X\text{-}Y)_n$ amino acid sequence pattern are found normally in the triple helix domain of all nonfibrillar collagens, while any Gly substitution in fibrillar collagens leads to pathological conditions.[1,2] Collagens represent a large family of triple helical extracellular matrix molecules utilized by cells for structural integrity and other biological functions. Twenty-eight distinct types of collagens have been thus far identified, including the most abundant fibril-forming collagens (Type I, II, III, V, and XI) and all other types designated as nonfibrillar collagens.[3,4] All collagens contain a characteristic repetitive $(Gly\text{-}X\text{-}Y)_n$ sequence pattern, which results in the formation of unique triple helical structure.[5,6]

The conservation of Gly as every third residue is absolutely required throughout the ~1000 residue long triple helix domain in fibrillar collagens. Even the replacement of a single Gly by a bulkier amino acid in the triple helix results in diseases.[7−9] More than 700 different Gly to X mutations have been reported in type I collagen to be associated with connective tissue disorder osteogenesis imperfecta (OI).[7,10] In contrast, such strict requirement of Gly is not retained in the triple helix domain of nonfibrillar collagens, where more than 350 interruptions in the $(Gly\text{-}X\text{-}Y)_n$ pattern have been observed.[2] These naturally occurring interruptions are considered to play critical roles in molecular structures as well as collagen binding and degradation.[11−16] It is not well understood why Gly substitutions and natural interruptions, both disrupting the $(Gly\text{-}X\text{-}Y)_n$ sequence pattern, result in totally contrary outcomes.

Triple-helical peptides have provided a versatile platform to investigate the sequence-specific conformational features of collagen.[17−21] The X-ray structure of peptide $(Pro\text{-}Hyp\text{-}Gly)_{10}[Gly15Ala]$ modeling a Gly-Ala mutation showed a highly localized angular distortion and hydrogen-bonding disruption at the Ala mutation site.[18,22] NMR studies of peptides modeling an Ala mutation in the highly charged Lys-Gly-Asp sequence environment indicated that only one of the three substituting Ala residues is capable of forming a good hydrogen bond.[23] In addition, biophysical studies on peptide models of natural interruptions have shown loss of helix registration and localized perturbations in hydrogen bonding and dihedral angles.[24,25] These studies have greatly enhanced our understanding of the structural features of natural interruptions and Gly substitutions; however, the peptide models contain quite different amino acid sequences, not allowing a direct comparison between these two types of sequential disruptions in collagen.

Peptides herein are designed to model natural interruptions and pathological Gly substitutions in a quite similar sequence environment to facilitate the comparison. The interruptions can

Received: July 3, 2015
Revised: September 3, 2015
Published: September 9, 2015

DOI: 10.1021/acs.biochem.5b00747
Biochemistry 2015, 54, 6106−6113

be classified by the length of amino acids between the repeating Gly-X-Y tripeptides. For instance, interruptions in the form of Gly-X-Y-Gly-$(AA)_n$-Gly-X-Y can be categorized as GnG type interruptions.[2] The G5G interruptions (Gly-AA_1-AA_2-AA_3-AA_4-AA_5-Gly) are particularly interesting, as they resemble the sequence of a Gly mutation in the collagen triple helix domain, where AA_3 represents the mutant residue (Gly-AA_1-AA_2-Mut-AA_4-AA_5-Gly). With the assistance of sequence analysis, two peptides, one modeling a natural G5G interruption (G-POALO-G) and the other one mimicking a pathological Gly-to-Ala substitution (G-LOAPO-G), are designed. Circular dichroism (CD), NMR, and computational simulation studies have discovered significant differences in stability, conformation, and folding between the two peptides. Compared with the Gly substitution sequence, the natural interruption may provide higher stability, higher composition of trimer species, and stronger folding capability while introducing more alterations in local triple helical conformation for collagen mimic peptides. These results suggest a molecular basis of different biological roles of natural interruptions and Gly substitutions and may guide the design of collagen mimic peptides containing functional natural interruptions.

MATERIALS AND METHODS

Sequence Analyses. All Gly-to-Ala mutations in the α1(I) chain of Type I collagen were collected from the Database of Collagen Mutations (www.le.ac.uk/genetics/collagen) at the University of Leicester.[26] Sequences of all nonfibrillar collagens were obtained from Genbank.[27] The amino acids were grouped into hydrophobic (Val, Ile, Leu, Met, Phe, Tyr, and Trp), charged (His, Lys, Arg, Glu, and Asp), small (Gly, Ala, Cys, Ser, and Thr), polar (Asn and Gln), and pro (Pro and Hyp). Fisher's exact test was used to test whether the types of residues (using the above groupings) at each position (AA_1, AA_2, AA_4, AA_5) in the sequence of Gly-AA_1-AA_2-Ala-AA_4-AA_5-Gly of G5G[Ala] interruptions (G5G type interruptions with a central Ala at the AA_3 site) were different from those found adjacent to Gly-to-Ala mutations of the α1(I) chain of Type I collagen. The test was also used to decide whether the types of residues were different at each position (AA_1, AA_2, AA_4, AA_5) in the sequence of Gly-AA_1-AA_2-Ala-AA_4-AA_5-Gly for G5G[Ala] interruptions. Fisher's exact test was utilized to test the statistical significance considering the small sample size. All the tests were two-tailed, and a p-value of <0.05 was considered significant. All the analyses were performed by the software SPSS (Statistical Product and Service Solutions, IBM).

Sample Preparation. Peptides LOAPO and POALO were synthesized by Chinese Peptide Company (Hangzhou, China). Both peptides were selectively ^{15}N labeled at position A19 and G28. NMR samples of both peptides were prepared in 10% D_2O/90% PBS (10 mM, pH 7.0) with a concentration of 5.3 mM.

Circular Dichroism Spectroscopy. CD spectra were recorded on a Chirascan CD spectrometer (Applied Photophysics, UK). Cells with a path length of 10 mm were used, and the temperature of the cells was controlled using a Peltier temperature controller. The CD samples were prepared at a concentration of 1 mg/mL in 10 mM PBS at PH 7. Wavelength scans were conducted from 195 to 260 nm with a 0.5 nm increment per step and a 5 s averaging time at 4 °C, and each scan was repeated three times. CD was applied to determine the thermal stability by monitoring the amplitude of the peak at 225 nm as a function of increasing temperature with an average heating rate of 0.1 °C/min.[28] The peptides were equilibrated for at least 24 h at 4 °C prior to the melting experiments. The melting temperature (T_m) is defined as the temperature at which the fraction folded is equal to 0.5 in the curve fitted to the trimer to monomer transition.

NMR Spectroscopy. All NMR experiments were performed on a Varian INOVA 600 MHz spectrometer equipped with High-Filed Indirect Detection NMR Probe Installation (Varian). $^1H-^{15}N$ heteronuclear single-quantum coherence (HSQC) spectra[29] and three-dimensional (3D) ^{15}N edited NOESY-HSQC experiments[30−32] with a mixing time of 50 ms were carried out for assignments of NMR resonances at 25 °C. 3D HNHA experiments were performed with an H−H coupling period of 25 ms at 27 °C to measure $^3J_{HNH\alpha}$ coupling constants.[33] The correction factor for the $^3J_{HNH\alpha}$ coupling constants was obtained as described.[24] Relaxation R_1, R_2, and heteronuclear NOE measurements[34−36] were done at 25 °C. For measurements of amide proton temperature gradients, a series of $^1H-^{15}N$ HSQC spectra were recorded from 0 to 40 °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 shift versus temperature. All data were processed using NMRPipe[37] and analyzed with NMRView.[38]

The NMR folding experiments were performed as previously described.[39,40] The samples were denatured by heating to 90 °C for 15 min and quickly transferred to the spectrometer that was equilibrated at 25 °C. A series of HSQC spectra were acquired every 5.6 min immediately after the sample was placed in the probe. The kinetics of folding was monitored by measuring cross-peak intensities as a function of time. The intensities of the monomer peaks were normalized following this equation: $I(t) = (I_t - I_f)/(I_0 - I_f)$, where I_t, I_0, and I_f are the intensities at time t, at time 0 and at final time point when momomer:trimer equilibrium has been reached. The rate constant k is obtained by exponential fitting as described.[39,40]

Molecular Simulation. The initial coordinates of peptides LOAPO and POALO were derived from RCSB pdb protein data bank (PDB ID: 1CAG) by mutating the 13th and 16th Pro to Leu of the peptide (1CAG), respectively.[18] The Tleap module of Amber 10.0 package was first used for the preliminary treatment of peptides. Amberff99SB force field was used for the peptides and water molecules. The parameters of 4-hydroxyproline were used as previously described.[41] An appropriate number of Cl- was added to maintain the electroneutrality of the simulation systems. Each system was immersed into a cubic periodic box of TIP3P waters with at least 10 Å distance around the solute. All the molecular dynamics (MD) simulations were performed using Amber 10.0 package. Prior to MD simulation, energy minimization was carried out by steepest descent and conjugated gradient methods to remove the bad contacts between the solute and solvent. The systems were heated from 0 to 310.0 K over a period of 50 ps in the NVT ensemble. One nanosecond equilibration was followed by 100 ns MD simulations without any restraint in the NPT ensemble at a temperature of 310.0 K and a pressure of 1 atm.

RESULTS

Comparison of Sequence Environments of Gly-to-Ala Substitutions and G5G[Ala] Interruptions. Sequence environments of Gly-to-Ala substitutions and G5G[Ala] interruptions (G5G type interruptions with a central Ala at

DOI: 10.1021/acs.biochem.5b00747
Biochemistry 2015, 54, 6106−6113

the AA_3 site) were compared to understand the totally different functional outcomes of substitutions versus interruptions. Gly-to-Ala substitutions and G5G[Ala] interruptions shared the same sequence pattern (Gly-AA_1-AA_2-Ala-AA_4-AA_5-Gly), thus easing the comparison. A total of 33 Gly-to-Ala OI mutations in the $\alpha(1)$ chain of Type I collagen and 19 G5G[Ala] interruptions were collected from the current database. The distribution of amino acid types was compared between Gly-to-Ala mutations and G5G[Ala] interruptions at each position (AA_1, AA_2, AA_4, AA_5) (Figure 1). The difference at positions

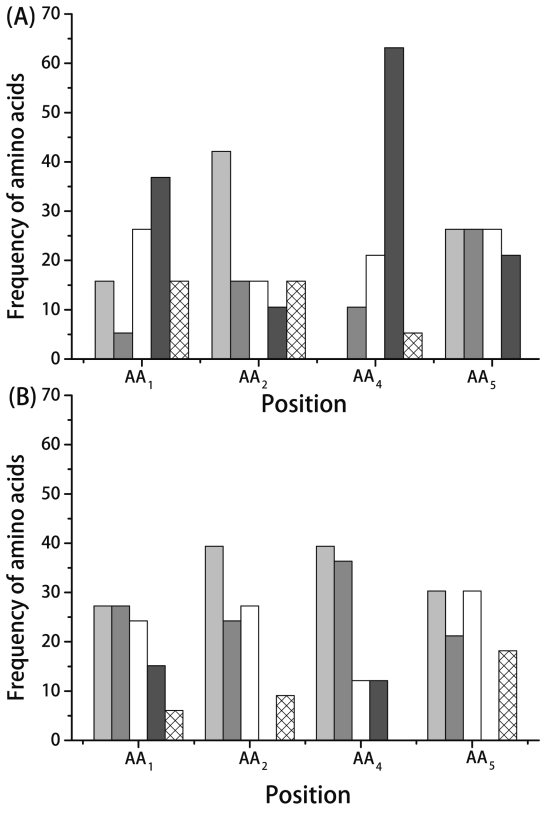


Figure 1. Distribution of amino acid types in the sequence of Gly-AA_1-AA_2-Ala-AA_4-AA_5-Gly. The frequency of amino acid types in the positions (AA_1, AA_2, AA_4, AA_5) of G5G[Ala] interruptions (G5G type interruptions with a central Ala at the AA_3 site) in nonfibrillar collagens (A) and Gly-to-Ala mutations in the α1(I) chain of Type I collagen (B). The amino acids are counted in five types: Pro (light gray), small (gray), charged (white), hydrophobic (dark gray), and polar (sparse white). For G5G[Ala] interruptions, it has a high percentage of hydrophobic residues at the position AA_4 (63%), though Gly-to-Ala mutations at the same position only have 12% hydrophobic amino acids. The distribution of amino acid types at the position AA_4 is significantly different between G5G[Ala] interruptions and Gly-Ala mutations ($p < 0.001$).

AA_1 and AA_2 was tested as insignificant, while position AA_5 showed statistically significant difference ($p = 0.03$). At position AA_5, both Gly-to-Ala mutations and G5G[Ala] interruptions had most Pro, while G5G[Ala] interruptions had much more hydrophobic and less polar amino acids than Gly-to-Ala mutations (Figure 1). The biggest difference was observed for position AA_4 with a p-value <0.001. For G5G[Ala] interruptions, the most distinct feature was that it had a very high percentage of hydrophobic residues (63%) and 0% Pro at the position AA_4. In contrast, at the same position, Gly-to-Ala substitutions have most Pro (39%), but very few hydrophobic amino acids (12%) (Figure 1).

Motif Selection. Interruption and mutation motifs were extracted from natural collagen sequences and inserted into a host model peptide $(GPO)_9$. Peptide POALO (with amino acid sequence: $(GPO)_5$-GPOALO-$(GPO)_4$, interruption site underlined) was used to model a G5G[Ala] interruption at site 386−390 in the $\alpha(1)$ chain of Type XIX collagen, a fibril-associated collagen with interrupted helices. Peptide POALO modeled the most distinct feature of G5G[Ala] interruptions, which contained most hydrophobic amino acids (58% of which is Leu) at position AA_4. LO-A-PO was a sequence environment for Gly-to-Ala OI mutations that occurred more than once, specifically, at site 13 and 304 in the $\alpha(1)$ chain of Type I collagen. Peptide LOAPO (with amino acid sequence: $(GPO)_5$-GLOAPO-$(GPO)_4$, mutation site underlined) was therefore designed to model these two Gly-to-Ala substitutions. Peptides POALO and LOAPO shared very similar amino acid content, in which PO and LO at each side of the central Ala simply exchanged their positions with each other.

Thermal Stability. Although POALO and LOAPO were similar in their sequences, the triple helices composed of POALO were much more stable than those of LOAPO. Peptide POALO displayed a high molecular residue ellipticity signal at 225 nm (MRE_{225} = 4588 deg·cm^2·dmol^{-1}) (Figure 2A). Monitoring the ellipticity at 225 nm with increasing temperature showed a sharp thermal transition with T_m = 55 °C (Figure 2B). In contrast, Peptide LOAPO displayed a significantly reduced MRE value (MRE_{225} = 2370 deg.cm^2.dmol$^{-1)}$, and it even did not display a well-resolved melting curve. This suggests that changing the relative position of PO and LO around the Ala led to loss of a substantial amount of triple-helix structure.

HSQC Spectra. Consistent with the observation of circular dichroism, NMR HSQC showed peptide POALO tended to form more triple helices than peptide LOAPO. In the HSQC spectrum of peptide POALO, only one trimer resonance for Gly28 was observed due to the repetitive GPO environment. Three trimer resonances were assigned for Ala19, since the three different chains had nonequivalent environments resulting from the one-residue staggering within the triple-helix (Figure 3A). Trimer and monomer resonances could be distinguished from each other by the ^{1}H−^{15}N heteronuclear NOE measurements. Other minor resonances in the HSQC spectrum aroused due to cis−trans isomerization of Gly−Pro and Pro−Hyp bonds in the peptide in the unfolded states.[42,43] For peptide LOAPO, the HSQC spectrum showed trimer resonances as well as monomer resonances for Ala19 and Gly28. The unique trimer resonance of Gly28 indicated that it adopted a typical $(Gly\text{-}Pro\text{-}Hyp)_n$ triple-helix. The three weak trimer resonances of Ala19 indicated some ordered triple helix structure at the Ala mutation site (Figure 3B). In contrast to peptide POALO, the much weaker trimer resonances of peptide LOAPO indicated that the peptide formed a much less amount of triple helix species at the Ala mutation site.

Backbone Local Conformation. The backbone conformation was characterized with $^3J_{HNH\alpha}$ coupling constants derived from the trimer resonances by HNHA experiments (Figure 4). Residues in the triple helix structure typically contained phi angles from −55 to −75 deg, giving a corresponding J coupling value of 4−6 Hz.[2,23,44] The labeled residue Gly28 in both peptides showed J coupling values between 4 and 6 Hz, supporting a standard triple-helix

DOI: 10.1021/acs.biochem.5b00747
Biochemistry 2015, 54, 6106−6113

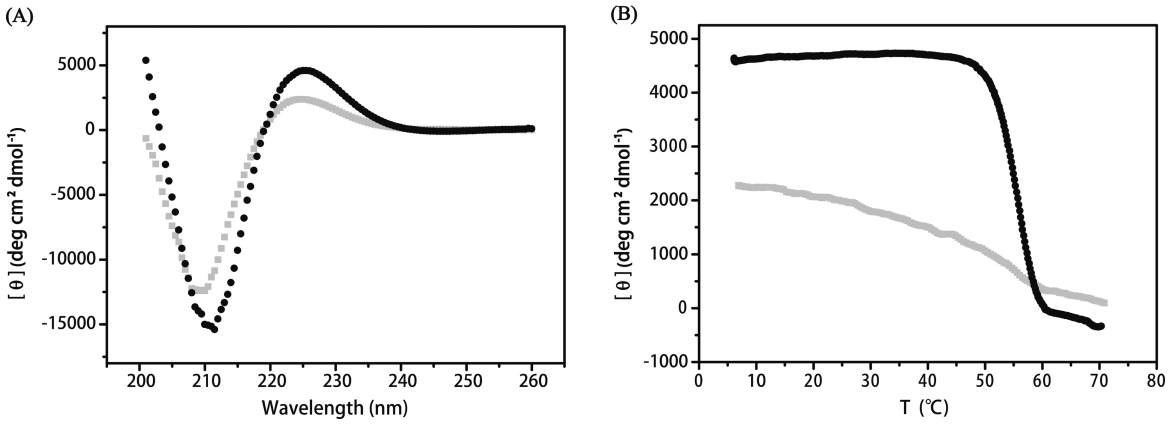


Figure 2. CD wavelength scans (A) and thermal transition (B) of peptides POALO (black) and LOAPO (gray). Both peptides were prepared at a concentration of 1 mg/mL in 10 mM PBS at PH 7. Wavelength scans were conducted at 4 °C. Melting transitions were monitored by CD spectroscopy at 225 nm.

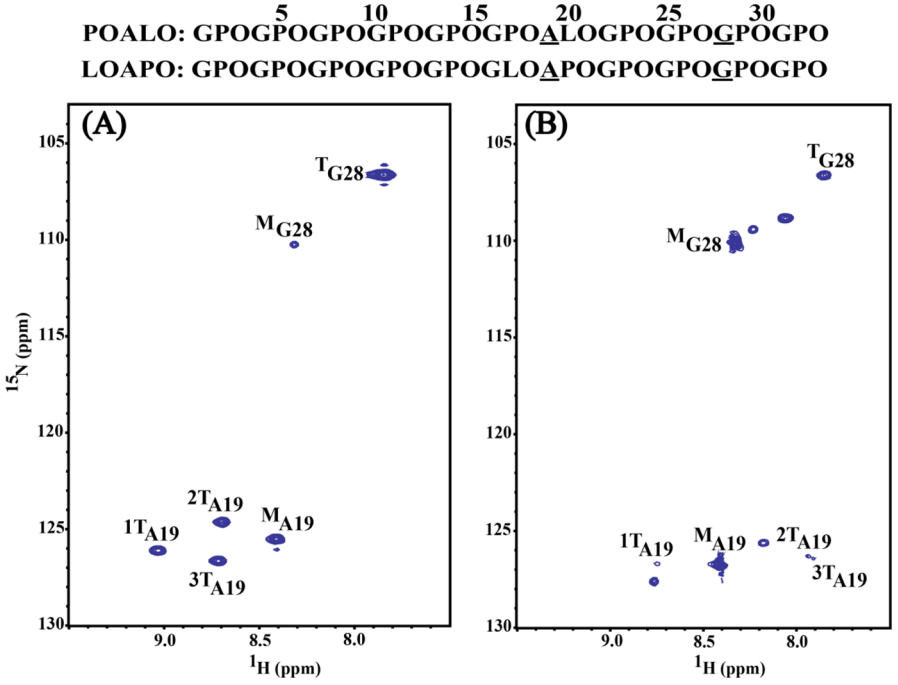


Figure 3. ^{1}H−^{15}N HSQC spectra of peptides POALO (A) and LOAPO (B) at pH 7 in 10 mM PBS at 35 °C. Peptide sequences are shown at the top with ^{15}N-labeled residues colored red. The peaks corresponding to the monomer and trimer states are denoted with a superscript M or T, respectively. The superscripted number 1, 2, and 3 corresponds to the three different chains, respectively.

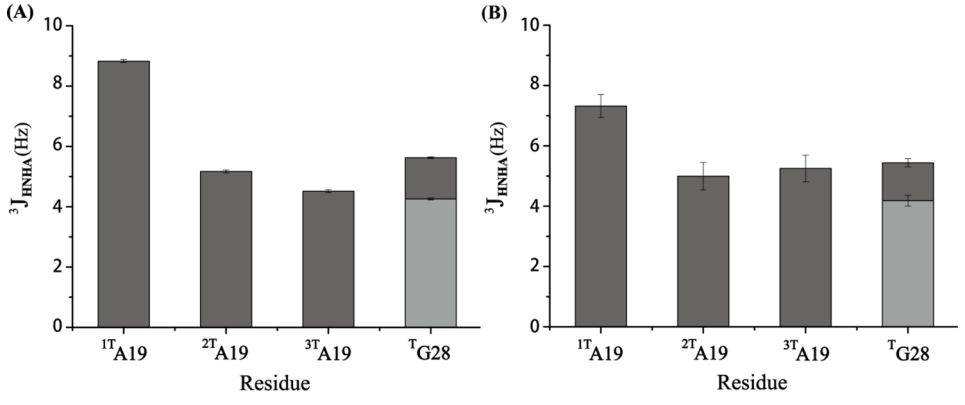


Figure 4. $^{3}J_{HNH\alpha}$ coupling values of peptides POALO (A) and LOAPO (B). Residues in the triple helical conformation typically contain phi angles from −55 to −75 deg and have a corresponding J coupling value of 4−6 Hz.

DOI: 10.1021/acs.biochem.5b00747
Biochemistry 2015, 54, 6106−6113

conformation at the C-terminal end for both peptides (Figure 4). For both peptides, two out of three Ala's (^{2}TA19 and ^{3}TA19) showed J coupling values between 4 and 6 Hz. However, a single Ala (^{1}TA19) had the J coupling value larger than 6 Hz for both peptides, especially for peptide POALO. This higher value may be indicative of a local distorted triple helix conformation of Ala.

Hydrogen Bonding. The amide proton temperature gradients are indicative of hydrogen bonding, with a value higher than −4.6 ppb/°C indicating the existence of a hydrogen bond.[23,24] The monomer resonances of Ala19 and Gly28 in both peptides showed much more negative values than this cutoff value, confirming the absence of any hydrogen bonding (Figure 5). For both peptides, the NH gradient values for Gly28 in the trimer state were much less than −4.6 ppb/°C, supporting the presence of typical hydrogen bonds for the triple helix structure. However, the amino gradient values for the three trimer resonances of Ala19 were very heterogeneous for the two peptides. For peptide POALO, all the three Ala19 in the trimer states showed NH gradient values more negative than −4.6 ppb/°C, which indicated that the Ala's may not participate in any hydrogen bonds. In contrast, for peptide LOAPO, two out three Ala's, ^{2}TA19 and ^{3}TA19, showed NH gradient values less negative than −4.6 ppb/°C, supporting the formation of hydrogen bonds. It suggested that peptide POALO may contain no hydrogen bonds at the interruption site, but peptide LOAPO likely forms two hydrogen bonds at the mutation site.

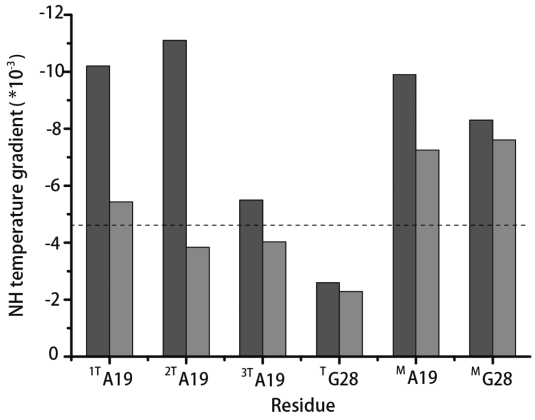


Figure 5. Amide proton NH temperature gradients of peptide POALO (left) and LOAPO (right). The black dashed horizontal line corresponds to a cutoff value for hydrogen bonding, with less negative values than −4.6 ppb/°C indicative of hydrogen bonding.

Real-Time NMR Folding. The folding dynamics of these two motifs were compared by residue-specific real-time NMR folding experiments (Figure 6). The intensities of monomer resonances of Ala19 and Gly28 in both peptides were monitored under the same conditions for ∼250 min (Figure 6). For Ala19, it refolded comparatively much faster in peptide POALO (black, k = 0.083 min^{-1}) than that in peptide LOAPO (red, k = 0.052 min^{-1}), while Gly28 refolded at similar rates (∼0.053 min^{-1}) in both peptides. It suggested that the interruption sequence POALO may have less disturbing effects on the triple helix folding than the mutation sequence LOAPO.

Molecular Simulation. Atomic-level structural details of peptides POALO and LOAPO were obtained with molecular dynamics simulation (Figure 7). The effects of the breaks in the sequence on the conformation were very local. Both peptides maintained the overall triple helix structure, while the hydrogen bonding pattern around the breaks were altered differently between the two peptides (Figure 7A,B). For peptide POALO, only one hydrogen bond was observed for Ala19 at the mutation site; and one Gly22, the Gly just one triplet following the mutation site, got seriously affected in the middle chain and could not form hydrogen bonds (Figure 7C). In contrast, for peptide LOAPO, two of the three Ala's at the mutation site form hydrogen bonds, which were consistent with the amide temperature gradient results (Figure 7D). For Gly16 and Gly22, which were just nearby the Ala19, all maintained the expected hydrogen bonds for the reserved Gly in the triple helix structure (Figure 7D). Besides hydrogen bonding, the dihedral angles were also measured for both peptides (data not shown). Most amino acids contained proper dihedral angles for polyproline II conformation, while major perturbation was observed for the residues near the breaks. For peptide POALO, five residues at the interruption site were affected, while peptide LOAPO only contained three residues out of polyproline II conformation at the mutation site. These results suggested that peptide LOAPO may have a less perturbation effect on the local conformation at the breaks than peptide POALO.

DISCUSSION

Although the repetitive (Gly-X-Y)$_n$ sequence pattern and uniform triple helix structure may be considered as well-defined characteristics of collagen, interruptions and mutations in collagen amino acid sequences present a seemingly

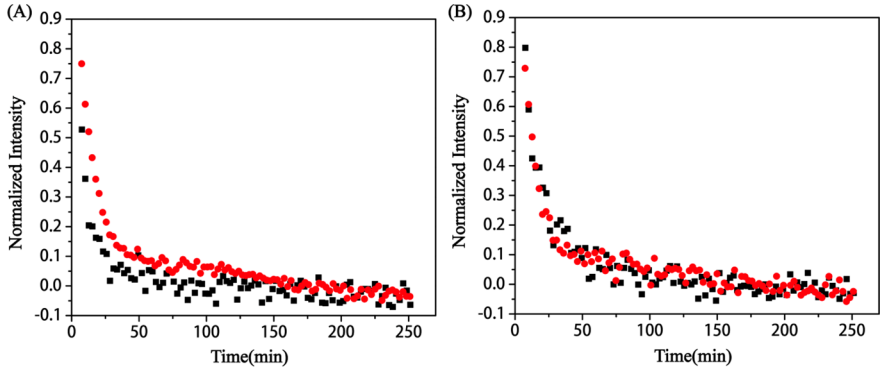


Figure 6. Residue-specific real-time NMR folding of A19 (A) and G28 (B) by monomer decay as a function of time. The intensities of monomer resonances of A19 and G28 in the peptides POALO (black) and LOAPO (red) were monitored for ∼250 min.

DOI: 10.1021/acs.biochem.5b00747
Biochemistry 2015, 54, 6106−6113

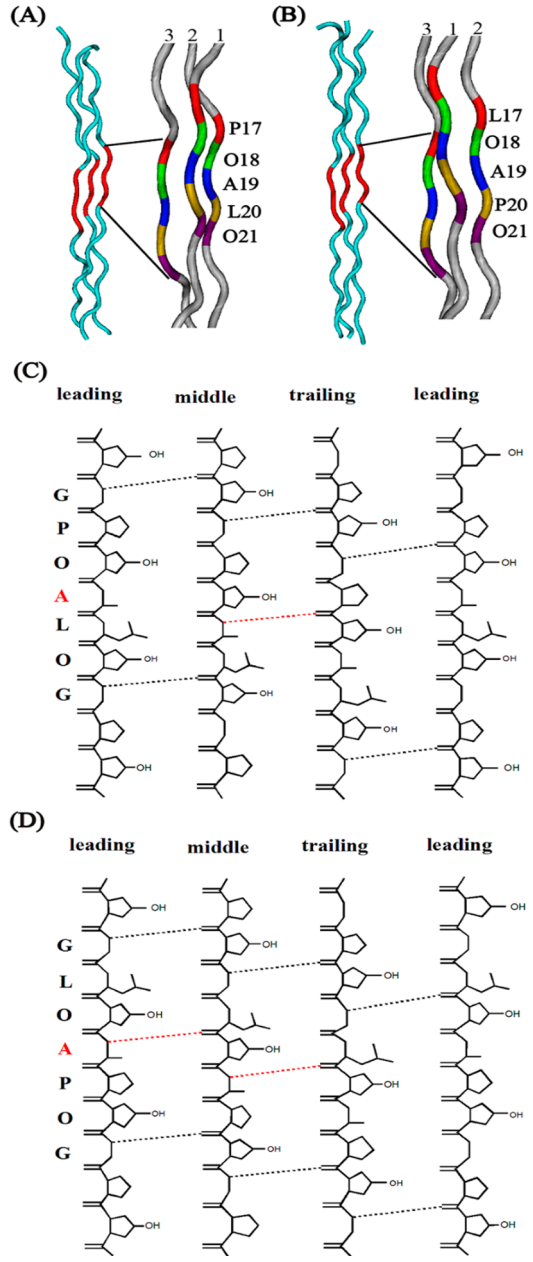


Figure 7. Average structures of peptides POALO (A and C) and LOAPO (B and D) by molecular simulation. The overall ribbon and zoomed-in structure of the breaks of peptides POALO (A) and LOAPO (B). Hydrogen bonding patterns near the breaks in peptides POALO (C) and LOAPO (D) were calculated based on their model structures.

conflicting challenge. Natural interruptions in nonfibrillar collagens play normal functional roles, while Gly substitutions in fibrillar collagens always result in diseases.[2] Our sequence analysis has identified two exemplified motifs for G5G interruptions and Gly-to-Ala substitutions with similar amino acid sequence but different biological roles, which provides a good opportunity to ease the comparison for elucidating their underlying molecular mechanisms.

The triple helix stability of peptides POALO and LOAPO are well correlated with their biological roles. Peptide POALO shows much higher thermal stability and stronger trimer resonances than peptide LOAPO. Stability bias has been correlated with the severities of osteogenesis imperfecta, suggesting that the Gly substitutions leading to larger destabilization cause more severe OI phenotypes.[45] The large reduction of stability and triple helix content of peptide LOAPO indicates that Gly substitutions may have a much larger destabilizing effects than natural interruptions, therefore resulting in pathological conditions. Folding has also been considered as a key factor for osteogenesis imperfecta phenotypes.[46] Peptide POALO has a significantly higher folding rate at the Ala site than peptide LOAPO, suggesting that the natural interruptions may cause less interference in folding than mutations, therefore not leading to diseases. The high stability, high triple helix content, and strong folding capability may ensure the triple helices with interruptions to play normal functions.

In contrast to the global stability profile, peptide POALO shows more local alterations in the triple helix structure at the sites of breaks than peptide LOAPO. Compared with peptide LOAPO, peptide POALO displays a larger J coupling value for ^{1}TA19, suggesting that it may cause a larger deviation from canonical triple helix structure at the interruption site than that at the mutation site. In addition, peptide POALO contains more local residues out of polyproline II conformation than peptide LOAPO based on model structures obtained from dynamics simulations. Furthermore, peptide POALO maintains less hydrogen bonds at the interruption site than peptide LOAPO at the mutation site as revealed by NMR as well as molecular simulation studies. To conclude, peptide POALO with a G5G interruption displays a locally more altered triple helical conformation in terms of dihedral angles and hydrogen bonding compared with peptide LOAPO with a Ala mutation. A number of interruptions have been observed to be close to the binding sites of collagen with other biomolecules,[12,15,16] and these structural deviations may provide recognition signal. In contrast, Gly mutations near the binding sites of collagen tend to result in severe OI phenotypes.[7,47] The global triple helix stability and local conformational flexibility may be complexed with higher-order interactions, and play a balanced role in collagen function.

Although the appearance of more hydrogen bonds but much less stability for peptide LOAPO is striking, there may well be contributions from other factors such as hydration networks and hydrophobic interactions in stabilizing the triple helix structure. A number of factors such as the presence of Gly as every third residue, a high content of imino acids, hydrogen bonding, hydration networks, hydrophobic interactions, and salt bridges have been proposed to contribute to the triple helix stability.[2,18,24,48−53] The backbone amine of the conserved Gly and the backbone carbonyl oxygen of an Xaa residue on the subsequent chain forms direct interchain hydrogen bonds, strongly stabilizing the triple helix structure.[53] When a Gly-Ala mutation is introduced in triple helical peptide (Pro-Hyp-Gly)$_{10}$[Gly15Ala], the direct hydrogen bonding at the mutation site is disrupted, while water-mediated indirect hydrogen bonding can be formed.[18] More importantly, the peptide displays an extensive hydration network with an abundance of hydrogen bonding between water molecules and peptide acceptor groups, thus stabilizing the triple helix structure.[18] X-ray structures of different peptides have shown that the hydration pattern is sequence dependent.[49] In addition, favorable hydrophobic interactions are also considered important, and peptide studies of a G4G interruption show the presence of hydrophobic packing at the interruption site.[24,48,52] Interestingly, recent studies of a heterotrimer

DOI: 10.1021/acs.biochem.5b00747
Biochemistry 2015, 54, 6106−6113

peptide show more hydrogen bonds but slightly less stability than a similar homotrimer peptide.[54] To conclude, triple helix stability is complicated with multiple factors, and favorable hydrophobic interactions, hydration networks as well as water-mediated hydrogen bonds may compensate for the loss of direct hydrogen bonds, and stabilize the triple helix structure.

Our studies here may serve as a prototype G5G interruption in nonfibrillar homotrimer collagen such as type X collagen in hypertrophic cartilage. The most distinct feature of G5G[Ala] interruptions is the critical hydrophobic residue at the AA_4 position. Changing the hydrophobic Leu to Pro in the model peptides at the AA_4 position significantly alters the global stability and local conformation. It is not well understood how hydrophobicity is correlated with triple helix stability, while imino acids nearby a mutation have been shown to destabilize the triple helix structure.[43,50] It is likely that the Pro leads to a significant loss of stability, while the Leu, as a non-imino acid, displays much weaker disruptive effects. Furthermore, previous studies of G1G and G4G interruptions have identified the dominance of hydrophobic residues at the AA_1 and the AA_3 site, repectively.[2,24] The studies suggested that the hydrophobic residues in G1G and G4G interruptions may play a physical basis of their different susceptibility to matrix metalloproteinase.[2,24] Recent studies on collagen peptide models have suggested that hydrophobic residues can mediate peptide self-assembly into various higher order structures.[55] The dominance of hydrophobic groups at the interruption sites may provide functional constraints for higher-order assembly as well as biomolecular interactions.[2,55] Our studies shed light on the molecular basis of the biological role of interruptions as well as the design of collagen mimic peptides containing functional natural interruptions.

AUTHOR INFORMATION

Corresponding Author

*E-mail: xiaojx@lzu.edu.cn.

Author Contributions

#X.S. and Y.C. contributed equally.

Funding

This work was supported by grants from the National Natural Science Foundation of China (Grant No. 21305056) and the Fundamental Research Funds for the Central Universities (Grant No. lzujbky-2014-73).

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

HSQC, heteronuclear single quantum coherence; Hyp, hydroxyproline (three-letter code); O, hydroxyproline (single-letter code)

REFERENCES

(1) Kielty, C. M., and Grant, M. E. (2002) The collagen family: strucutre, assembly, and organization in the extracellular matrix, In *Connective Tissue and Its Heritable Disorders, Molecular, Genetic and Medical Aspects* (Royes, P. M., and Steinmann, B. U., Eds.) pp 159–222, Wiley Liss, New York.

(2) Thiagarajan, G., Li, Y., Mohs, A., Strafaci, C., Popiel, M., Baum, J., and Brodsky, B. (2008) Common interruptions in the repeating tripeptide sequence of non-fibrillar collagens: sequence analysis and structural studies on triple-helix peptide models. *J. Mol. Biol. 376*, 736–748.

(3) van der Rest, M., and Garrone, R. (1991) Collagen family of proteins. *FASEB J. 5*, 2814–2823.

(4) Gordon, M. K., and Hahn, R. A. (2010) Collagens. *Cell Tissue Res. 339*, 247–257.

(5) Ramachandran, G. N., and Kartha, G. (1955) Structure of collagen. *Nature 176*, 593–595.

(6) Rich, A., and Crick, F. H. (1955) The structure of collagen. *Nature 176*, 915–916.

(7) Marini, J. C., Forlino, A., Cabral, W. A., Barnes, A. M., San Antonio, J. D., Milgrom, S., Hyland, J. C., Korkko, J., Prockop, D. J., De Paepe, A., Coucke, P., Symoens, S., Glorieux, F. H., Roughley, P. J., Lund, A. M., Kuurila-Svahn, K., Hartikka, H., Cohn, D. H., Krakow, D., Mottes, M., Schwarze, U., Chen, D., Yang, K., Kuslich, C., Troendle, J., Dalgleish, R., and Byers, P. H. (2007) 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. 28*, 209–221.

(8) Myllyharju, J., and Kivirikko, K. I. (2001) Collagens and collagen-related diseases. *Ann. Med. 33*, 7–21.

(9) Byers, P. H., and Cole, W. G. (2002) Osteogenesis Imperfecta, in *Connective Tissue and Its Hereditable Disorders* (Royce, P. M., and Steinmann, B., Eds.) pp 385–430, Wiley-Liss, New York.

(10) Dalgleish, R. Osteogenesis imperfecta & Ehlers-Danlos syndrome variant databases, In http://www.le.ac.uk/genetics/collagen/, Department of Genetics, University of Leicester.

(11) Brazel, D., Oberbaumer, I., Dieringer, H., Babel, W., Glanville, R. W., Deutzmann, R., and Kuhn, K. (1987) Completion of the amino acid sequence of the alpha 1 chain of human basement membrane collagen (type IV) reveals 21 non-triplet interruptions located within the collagenous domain. *Eur. J. Biochem. 168*, 529–536.

(12) Miles, A. J., Knutson, J. R., Skubitz, A. P., Furcht, L. T., McCarthy, J. B., and Fields, G. B. (1995) A peptide model of basement membrane collagen alpha 1 (IV) 531–543 binds the alpha 3 beta 1 integrin. *J. Biol. Chem. 270*, 29047–29050.

(13) Mohs, A., Popiel, M., Li, Y., Baum, J., and Brodsky, B. (2006) Conformational features of a natural break in the type IV collagen Gly-X-Y repeat. *J. Biol. Chem. 281*, 17197–17202.

(14) Hofmann, H., Voss, T., Kuhn, K., and Engel, J. (1984) Localization of flexible sites in thread-like molecules from electron micrographs. Comparison of interstitial, basement membrane and intima collagens. *J. Mol. Biol. 172*, 325–343.

(15) Miles, A. J., Skubitz, A. P., Furcht, L. T., and Fields, G. B. (1994) Promotion of cell adhesion by single-stranded and triple-helical peptide models of basement membrane collagen alpha 1(IV)531–543. Evidence for conformationally dependent and conformationally independent type IV collagen cell adhesion sites. *J. Biol. Chem. 269*, 30939–30945.

(16) Furcht, L. T., Skubitz, A. P., and Fields, G. B. (1994) Tumor cell invasion, matrix metalloproteinases, and the dogma. *Lab. Invest. 70*, 781–783.

(17) Brodsky, B., Thiagarajan, G., Madhan, B., and Kar, K. (2008) Triple-helical peptides: an approach to collagen conformation, stability, and self-association. *Biopolymers 89*, 345–353.

(18) Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994) Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science 266*, 75–81.

(19) Baum, J., and Brodsky, B. (1999) Folding of peptide models of collagen and misfolding in disease. *Curr. Opin. Struct. Biol. 9*, 122–128.

(20) Xiao, J., Addabbo, R. M., Lauer, J. L., Fields, G. B., and Baum, J. (2010) Local conformation and dynamics of isoleucine in the collagenase cleavage site provide a recognition signal for matrix metalloproteinases. *J. Biol. Chem. 285*, 34181–34190.

(21) Xiao, J., Cheng, H., Silva, T., Baum, J., and Brodsky, B. (2011) Osteogenesis imperfecta missense mutations in collagen: structural consequences of a glycine to alanine replacement at a highly charged site. *Biochemistry 50*, 10771–10780.

(22) Bella, J., Brodsky, B., and Berman, H. M. (1995) Hydration structure of a collagen peptide. *Structure 3*, 893–906.

DOI: 10.1021/acs.biochem.5b00747
Biochemistry 2015, 54, 6106–6113

(23) Li, Y., Brodsky, B., and Baum, J. (2009) NMR conformational and dynamic consequences of a gly to ser substitution in an osteogenesis imperfecta collagen model Peptide. *J. Biol. Chem. 284*, 20660−20667.
(24) Li, Y., Brodsky, B., and Baum, J. (2007) 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. 282*, 22699−22706.
(25) Bella, J., Liu, J., Kramer, R., Brodsky, B., and Berman, H. M. (2006) Conformational effects of Gly-X-Gly interruptions in the collagen triple helix. *J. Mol. Biol. 362*, 298−311.
(26) Dalgleish, R. (1997) The human type I collagen mutation database. *Nucleic Acids Res. 25*, 181−187.
(27) Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2013) GenBank. *Nucleic Acids Res. 41*, D36−42.
(28) Persikov, A. V., Pillitteri, R. J., Amin, P., Schwarze, U., Byers, P. H., and Brodsky, B. (2004) Stability related bias in residues replacing glycines within the collagen triple helix (Gly-Xaa-Yaa) in inherited connective tissue disorders. *Hum. Mutat. 24*, 330−337.
(29) Kay, L. E., Keifer, P., and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc. 114*, 10663−10665.
(30) Messerle, B. A., Wider, G., Otting, G., Weber, C., and Wüthrich, K. (1989) Solvent suppression using a spin-lock in 2D and 3D NMR spectroscopy with H2O solutions. *J. Magn. Reson. 85*, 608−613.
(31) Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) Three-dimensional heteronuclear NMR of 15N labeled proteins. *J. Am. Chem. Soc. 111*, 1515−1517.
(32) Fesik, S. W., and Zuiderweg, E. R. (1988) Heteronuclear three-dimensional nmr spectroscopy. A strategy for the simplification of homonuclear two-dimensional NMR spectra. *J. Magn. Reson. 78*, 588−593.
(33) Vuister, G. W., and Bax, A. (1993) 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. 115*, 7772−7777.
(34) Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry 33*, 5984−6003.
(35) Palmer, A. G., 3rd (1993) Dynamic properties of proteins from NMR spectroscopy. *Curr. Opin. Biotechnol. 4*, 385−391.
(36) Fan, P., Li, M. H., Brodsky, B., and Baum, J. (1993) Backbone dynamics of (Pro-Hyp-Gly)10 and a designed collagen-like triple-helical peptide by 15N NMR relaxation and hydrogen-exchange measurements. *Biochemistry 32*, 13299−13309.
(37) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol NMR 6*, 277−293.
(38) Johnson, B. A., and Blevins, R. A. (1994) Nmr View - a Computer-Program for the Visualization and Analysis of Nmr Data. *J. Biomol. NMR 4*, 603−614.
(39) Liu, X., Siegel, D. L., Fan, P., Brodsky, B., and Baum, J. (1996) Direct NMR measurement of folding kinetics of a trimeric peptide. *Biochemistry 35*, 4306−4313.
(40) Buevich, A. V., Silva, T., Brodsky, B., and Baum, J. (2004) Transformation of the mechanism of triple-helix peptide folding in the absence of a C-terminal nucleation domain and its implications for mutations in collagen disorders. *J. Biol. Chem. 279*, 46890−46895.
(41) Park, S., Radmer, R. J., Klein, T. E., and Pande, V. S. (2005) A new set of molecular mechanics parameters for hydroxyproline and its use in molecular dynamics simulations of collagen-like peptides. *J. Comput. Chem. 26*, 1612−1616.
(42) Buevich, A. V., Dai, Q. H., Liu, X., Brodsky, B., and Baum, J. (2000) Site-specific NMR monitoring of cis-trans isomerization in the folding of the proline-rich collagen triple helix. *Biochemistry 39*, 4299−4308.
(43) Xiao, J., Madhan, B., Li, Y., Brodsky, B., and Baum, J. (2011) Osteogenesis imperfecta model peptides: incorporation of residues replacing Gly within a triple helix achieved by renucleation and local flexibility. *Biophys. J. 101*, 449−458.
(44) Kramer, R. Z., Bella, J., Mayville, P., Brodsky, B., and Berman, H. M. (1999) Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol. 6*, 454−457.
(45) Beck, K., Chan, V. C., Shenoy, N., Kirkpatrick, A., Ramshaw, J. A., and Brodsky, B. (2000) Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *Proc. Natl. Acad. Sci. U. S. A. 97*, 4273−4278.
(46) Baum, J., and Brodsky, B. (1998) Triple-helix folding and its role in collagen diseases. *FASEB Journal 12*, A1329−A1329.
(47) Xiao, J., Yang, Z., Sun, X., Addabbo, R., and Baum, J. (2015) Local amino acid sequence patterns dominate the heterogeneous phenotype for the collagen connective tissue disease Osteogenesis Imperfecta resulting from Gly mutations. *J. Struct. Biol.*, DOI: 10.1016/j.jsb.2015.05.002.
(48) Persikov, A. V., Ramshaw, J. A., Kirkpatrick, A., and Brodsky, B. (2000) Amino acid propensities for the collagen triple-helix. *Biochemistry 39*, 14960−14967.
(49) Kramer, R. Z., Bella, J., Brodsky, B., and Berman, H. M. (2001) The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence. *J. Mol. Biol. 311*, 131−147.
(50) Radmer, R. J., and Klein, T. E. (2004) Severity of osteogenesis imperfecta and structure of a collagen-like peptide modeling a lethal mutation site. *Biochemistry 43*, 5314−5323.
(51) Xu, K., Nowak, I., Kirchner, M., and Xu, Y. (2008) 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. 283*, 34337−34344.
(52) Persikov, A. V., Ramshaw, J. A., Kirkpatrick, A., and Brodsky, B. (2005) Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. *Biochemistry 44*, 1414−1422.
(53) Shoulders, M. D., and Raines, R. T. (2009) Collagen structure and stability. *Annu. Rev. Biochem. 78*, 929−958.
(54) Xiao, J., Sun, X., Balaram, M., Brodsky, B., and Baum, J. (2015) NMR studies demonstrate a unique AAB composition and chain register for a heterotrimeric type IV collagen model peptide containing a natural interruption site. *J. Biol. Chem.*, jbc.M115.654871.
(55) McGuinness, K., Khan, I. J., and Nanda, V. (2014) Morphological diversity and polymorphism of self-assembling collagen peptides controlled by length of hydrophobic domains. *ACS Nano 8*, 12514−12523.

DOI: 10.1021/acs.biochem.5b00747
Biochemistry 2015, 54, 6106−6113