BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

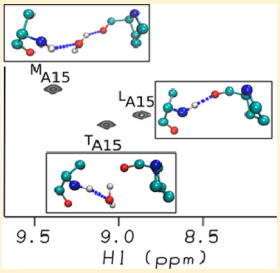
Dynamic Water-Mediated Hydrogen Bonding in a Collagen Model Peptide

Iwen Fu, David A. Case,* and Jean Baum*

Department of Chemistry and Chemical Biology and BioMaPS Institute, Rutgers University, Piscataway, New Jersey 08854, United States

Supporting Information

ABSTRACT: In the canonical $(G\text{-}X\text{-}Y)_n$ sequence of the fibrillar collagen triple helix, stabilizing direct interchain hydrogen bonding connects neighboring chains. Mutations of G can disrupt these interactions and are linked to connective tissue diseases. Here we integrate computational approaches with nuclear magnetic resonance (NMR) to obtain a dynamic view of hydrogen bonding distributions in the $(POG)_4\text{-}(POA)\text{-}(POG)_5$ peptide, showing that the solution conformation, dynamics, and hydrogen bonding deviate from the reported X-ray crystal structure in many aspects. The simulations and NMR data provide clear evidence of inequivalent environments in the three chains. Molecular dynamics (MD) simulations indicate direct interchain hydrogen bonds in the leading chain, water bridges in the middle chain, and nonbridging waters in the trailing chain at the G → A substitution site. Theoretical calculations of NMR chemical shifts using a quantum fragmentation procedure can account for the unusual downfield NMR chemical shifts at the substitution sites and are used to assign the resonances to the individual chains. The NMR and MD data highlight the sensitivity of amide shifts to changes in the acceptor group from peptide carbonyls to water. The results are used to interpret solution NMR data for a variety of glycine substitutions and other sequence triplet interruptions to provide new connections between collagen sequences, their associated structures, dynamical behavior, and their ability to recognize collagen receptors.



Collagen is the most abundant protein in the human body and is remarkable as it plays a dual role as a structural protein that provides tensile strength to skin, bone, cartilage, and blood vessel walls and as a protein that is highly biologically active and involved in many extracellular matrix interactions.[1−3] The linear $(G\text{-}X\text{-}Y)_n$ repetitive triple-helical collagen motif presents a picture very different from that of a typical folded globular protein, and its interactions have different features or specificity. The triple helix is a rodlike folded form in which the glycine residues are all buried near a central axis, while the residues in the X and Y positions are largely exposed to solvent.[4,5] The collagen triple helix has a one-residue stagger, and the hydrogen bonding (H-bond) of the triple helix is between the three chains with the backbone amide protons of the Gly residues forming interchain hydrogen bonds (H-bonds) and the carbonyl oxygen of the X residue in the adjacent chain. All polar residues and hydrophobic side chains are on the exterior, and there is no hydrophobic core. The X and Y positions are frequently occupied by Pro and hydroxyproline (abbreviated by the three-letter code Hyp or one-letter code O), respectively, which are important to triple helix stability, and Gly-Pro-Hyp is the most common and most stabilizing tripeptide sequence.

Defects in the triple helix domain of collagen have been associated with a number of human collagen diseases.[1,6−8] Osteogenesis Imperfecta (OI) or brittle bone disease affects roughly one in 10 000 individuals and results from mutations in type I collagen, a heterotrimer comprised of two $\alpha1$(I) chains and one $\alpha2$(I) chain.[6−8] The most common mutation is a single-base change that leads to the replacement of a single Gly with another residue anywhere along the $(G\text{-}X\text{-}Y)_n$ sequence, breaking the repeating tripeptide pattern. The severity of the disease varies widely, ranging from mild cases with multiple fractures to perinatal lethal cases. The molecular basis of the disease is still not understood, but it has been shown that mutations result in perturbation of the local conformation and dynamics of the triple helix structure around the mutation site.[6,9,10]

Collagen model peptides (CMPs) have been used to model the conformation, dynamics, folding, and stability of the collagen triple helix.[11−15] The simplest model of an OI mutation, a Gly to Ala substitution in the center of the repeating $(POG)_{10}$ sequence (termed G → A peptide hereafter), has been shown to form a very stable triple helix but with a significant decrease in thermal stability relative to that of the peptide with a repeating $(POG)_{10}$ sequence.[16] Studies of a heterotrimer model of a Gly to Ala substitution show that the thermal stability varies as a function of the number of chains that contain an Ala substitution.[17] The X-ray crystal structure[18,19] reveals a highly localized disrupted screw symmetry at the mutation site and shows a loss of direct interchain H-bonds. For each chain, the direct H-bonds are replaced with water-mediated H-bonds in which water creates a bridge between the backbone group of the Ala NH to the O=C group of the Pro in the neighboring chain.

Received: June 5, 2015
Revised: September 4, 2015
Published: September 4, 2015

ACS Publications © 2015 American Chemical Society

DOI: 10.1021/acs.biochem.5b00622
Biochemistry 2015, 54, 6029−6037

The unique rodlike structure of the repeating sequence of the triple helix make structural studies by solution nuclear magnetic resonance (NMR) challenging.[20] However, NMR data on G → A peptides suggest that the structural and dynamic perturbations are very sensitive to the local sequence context surrounding the mutation[21−24] and that perturbations induced by the Gly to Ala mutation in the $(POG)^{10}$ context deviate from those in the crystal environment. The NMR data indicate that the Gly to Ala substitutions lead to nonequivalence of the Ala residues in three strands with one of the strands likely to form a good backbone H-bond while the other two do not, and one strand having a very significant downfield-shifted ^{1}H resonance relative to other two resonances in the HSQC spectra. The NMR[23] data suggest that the nature of the H-bonding in the G → A peptide is more asymmetric and dynamic than that of the crystal structure. However, because of the lack of chain-specific NMR assignments and in light of the complex H-bonding configurations of the triple helix[18,19] that include interchain H-bonding, an extensive hydration network, and in certain instances water-mediated H-bonds, a full description of the H-bonding structure and dynamics is impossible using NMR data alone.

In this study, we integrate NMR and molecular dynamics (MD) simulations with explicit solvent to elucidate the complex H-bonding dynamics of a triple helix that contains a Gly mutation site that models a disease state. The MD simulations capture structural perturbations consistent with the NMR studies, supporting the notion that the solution conformation of CMPs deviates from the X-ray crystal structure in many aspects. MD simulations using explicit waters on the simplest model peptide of OI, the G → A peptide, show that the rodlike molecule has significant bending motions, on a time scale of hundreds of picoseconds, and that the three triple-helical chains display distinctly different H-bonding dynamics. The H-bonding occupancies at the Ala mutation site are different in the three chains and have complex dynamics; water is moving in and out to allow the formation of direct H-bonds or to form dynamic interstitial water-mediated H-bonds. Theoretical calculations of NMR chemical shifts on the MD snapshots that contain interstitial waters can account for the unusual downfield NMR chemical shifts observed at Gly substitution sites and highlight the sensitivity of the ^{1}H chemical shift to the complex dynamics of the H-bonding occupancies. We can explore in this unique system the differences between direct (interchain) H-bonding and water-mediated interactions in ways that are generally not possible with globular proteins. These results, combined with NMR data for other peptides, suggest a general model for the structural and dynamic consequences of glycine interruptions in collagen.

MATERIALS AND METHODS

MD System Description and Setup. The initial coordinates of the collagen-like peptides, $(POG)_4$-(POA)-$(POG)_5$ (termed G → A peptide) and $(POG)_{10}$, were taken from the published crystal structures of Protein Data Bank (PDB) entries 1CAG[18] and 1V7H,[25] respectively. MD simulations were performed with the Amber 99SB[58] protein force field using the AMBER10 programs.[59] In the simulation, the protein is placed in a truncated octahedral periodic box of the explicit SPC water model with approximately 27 700 water molecules. The distance from the surface of the box to the closest atom of the solute is set to 15 Å. The system is heated from 0 to 293 K over 1 ns and followed by a 3 ns simulation with restraints and then a 1 ns simulation without restraints in the *NPT* ensemble. Finally, a MD production simulation is performed in the *NVE* ensemble. In MD simulations, the SHAKE algorithm is used to constrain all bonds involving hydrogen atoms and the time step is 1 fs. The potential energy function represents the contribution from bond, dihedral angle, van der Waals, and electrostatic terms. The particle mesh Ewald method was used to calculate the electrostatic interaction, and a nonbonded cutoff of 8 Å was used during the MD simulation. The simulations were run for 170 ns for G → A peptide and 100 ns for $(POG)_{10}$ peptides. A separate 1 μs simulation of the G → A peptide in solution (see Figure S3 of the Supporting Information) confirmed that the structure is stable on this longer time scale, and that no significant changes were seen compared to the 170 ns simulation analyzed here. All trajectories were saved every 2 ps for future analysis. Postprocessing and analysis were conducted using the cpptraj analysis tool.[60]

Crystal Simulations. Simulations of the G → A peptide used the same force field parameters described above. We created a supercell consisting of six unit cells: one copy along the *a* (triple helix) direction and three and two copies along the *b* and *c* directions, respectively. Acetate ions modeled in PDB entry 1CGD[19] were modeled and neutralized by an equal number of sodium ions. The number of water molecules needed to maintain the observed unit cell dimensions was determined by trial and error to be 3384. The simulation temperature was maintained at 263 K (the temperature of data collection) by a Langevin thermostat with a collision frequency of 5 ps^{-1}. A constant pressure was maintained with a weak-coupling (Berendsen) barostat with a time constant of 3 ps. After equilibration for 280 ns (during which time restraints on the peptide to keep it near its crystallographic position were gradually relaxed), 300 ns of production simulation were performed. The final supercell dimensions were 173.36, 42.15, and 50.58 Å, compared to 173.46, 42.17, and 50.62 Å, respectively, in the experiment. (The β angle of 95.82° was not allowed to vary.) In a typical snapshot, the backbone atoms of the triple helix differed by 1.2 Å (root-mean-square distance) from the deposited model in PDB entry 1CGD.

Theoretical Chemical Shift Calculation. Proton and nitrogen chemical shifts for the NH atoms of the Gly and Ala residues were estimated using the automated fragmentation model originally introduced by He et al.[31] In this model, a quantum region is defined around the residue whose shift is being calculated, which includes that residue and its neighbors out to a cutoff of ∼3.5 Å. Water molecules within a similar distance are also included. Protein residues beyond the cutoff are included as molecular mechanics (MM) atoms, and the effects of waters beyond the cutoff are treated with a continuum dielectric (Poisson) model. Details are given elsewhere.[32,35,61] For these calculations, we used the OLYP density functional with a TZVP basis set (as recommended elsewhere[61]), using the demon-2k (version 3) program. For each calculation, we averaged more than 400 equally spaced snapshots with an interval of 50 ps from the last 20 ns of the MD simulation. The computed shielding was compared to those for amide nitrogen and hydrogen in ubiquitin (using the same computational model), to obtain the reported chemical shifts. The same snapshots were analyzed by the SHIFTX2 program[33] using default parameters, with no homology information being used.

NMR Experiments. Peptide $(POG)_4POA(POG)_5$ (designed G → A peptide) was synthesized by the Tufts University

DOI: 10.1021/acs.biochem.5b00622
Biochemistry 2015, 54, 6029−6037

Core Facility and purified using a Waters XTerra Prep C18 column on an Amersham Biosciences fast protein liquid chromatography system.[23] The G → A peptide was made with selectively ^{15}N-labeled residues at positions Ala15 and Gly24; labeling was limited by the repetitive sequence. NMR samples for the G → A peptide were prepared in a 10% D_2O/90% H_2O solvent at pH 2 with concentrations of 3 mM.

NMR experiments for the G → A peptide were performed on a Varian INOVA 500 MHz spectrometer. The HSQC experiment was performed at 15 °C. The $^3J_{HNHA}$ coupling constants were measured and corrected. For the measurements of amide proton temperature gradients, ^{1}H−^{15}N HSQC spectra were obtained at 0−25 °C with an interval of 5 °C. The sample was equilibrated at each temperature for at least 3 h. Amide proton temperature gradients were obtained by linear regression analysis of the amide proton chemical shifts versus temperature. All data were processed using the FELIX 2004 software package (Felix NMR, Inc., San Diego, CA) and/or NMRPipe[62] and analyzed with FELIX 2004 or NMRView[63] as described previously.[23,56]

RESULTS

MD Simulations Indicate Structural and Dynamic Perturbations at the Gly Substitution Sites. MD simulations offer a complement to X-ray and NMR structures by providing models and structures that can be analyzed for conformational flexibility, as well as global and local motions. CMPs have been typically considered to be relatively rigid rodlike molecules, and the crystal structure of the G → A peptide (Figure 1a) appears as a straight structure. Our simulations show small but noticeable bending motions, on a time scale of hundreds of picoseconds for the $(POG)_{10}$ peptide and the G → A peptide (Figure 1b) . The distribution of bending angles for both peptides suggests that the triple helix cannot be treated as a rigid rod, and notably, the G → A peptide has a distribution of bending angles slightly larger than that of the $(POG)_{10}$ peptide. The bending angle calculated from the crystal structure of either $(POG)_{10}$ or G → A peptide is nearly linear (~176°) and notably different from those observed in the solution MD simulations, where values from 150° to 180° are found. Snapshots corresponding to three different bending angles of G → A peptide are shown in Figure 1c, highlighting the non-negligible conformational fluctuations in solution of the mutation peptide relative to the $(POG)_{10}$ peptide. Such bending implies a nonequivalence in the three chains.

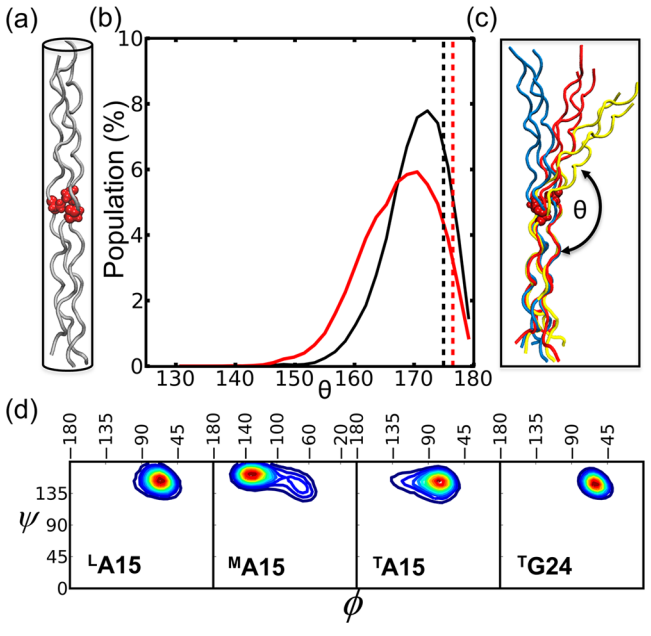


Figure 1. (a) Schematic representation of the G → A crystal structure (PDB entry 1cag[18]) with Ala residues highlighted as red van der Waals spheres. (b) Distributions of bend angles, θ (shown in panel c). The dashed line indicates angles of 175° and 177° calculated from the crystal structure of $(POG)_{10}$ (PDB entry 1v7h[25]) in black and G → A peptide in red, respectively. (c) MD selected snapshots corresponding to three bend angles (~180° in blue, ~160° in red, and ~145° in yellow) for the G → A peptide. (d) Ramachandran plot (ϕ and ψ) from the MD trajectories. Warmer colors represent larger populations.

The local perturbations in conformation induced by the Gly to Ala substitution can be further characterized by comparing the ϕ and ψ backbone angles of the three chains at the Ala at position 15 (termed A15) with those of G24 that acts as an internal control for a region that is not disrupted (Figure 1d). The angles in a repeating (G-X-Y) region are centered at −61° and 148°, characteristic of a PPII conformation. The three A15 and G24 residues have similar ψ angle distributions but adopt different ϕ angles. The Ala residue in the leading chain (^{L}A15) (where "L", "M", and "T" stand for the leading, middle, and trailing strands, respectively) is most similar to the G24 control, whereas ^{T}A15 shows a small population farther from the PPII conformation and ^{M}A15 adopts dihedral angles closer to the β region of the Ramachandran plot, confirming that the disruption creates an inequivalence between the three chains. The X-ray conformation[18,19] also has a more negative ϕ angle for the middle chain, and we show below that these trends are consistent with NMR coupling constants in solution.[23] The presence of a more extended backbone conformation in the middle chain (relative to the leading and trailing chains) is consistent with a greater propensity for bending, allowing the path length of this chain to be somewhat longer than those of the other two chains.

A Gly → Ala Substitution Results in an Asymmetric Water-Mediated H-Bonding Distribution within the Three Chains. The characteristic H-bonds that exist for Gly residues in the $(G\text{-}X\text{-}Y)_n$ repeating environment are significantly altered at the Gly to Ala substitution site and modestly altered at the G18 position one triplet C-terminal to it. Typically, in the $(G\text{-}X\text{-}Y)_n$ repeating unit, there is a direct H-bond between the Gly NH and the O═C of the X residue in the neighboring chain (Figure 2a).[18,19,25] In the G → A peptide (Figure 2b), the terminal G6 and G24 residues and the G12 residue that is just N-terminal to A15 have an N−O distance distribution that is identical for all three chains and is consistent with a typical interchain H-bond. By contrast, for A15, the distribution of the N−O bond distance is different for the three chains, and for each chain, the distribution is bimodal. The peak near 2.8 Å is consistent with a direct H-bond, whereas the second peak, at 4.5−5 Å, indicates the loss of a direct H-bond. The leading chain, ^{L}A15, has an interchain H-bond 72% of the time, whereas ^{T}A15 and ^{M}A15 have interchain H-bond occupancies of 30 and 17%, respectively. H-bonding disruption is also seen at G18, C-terminal to the mutation site, but not at G12, N-terminal to the mutation site. Therefore, in addition to inequivalent H-bonding distributions at the mutation site, the distribution of H-bonds relative to the substitution site is not symmetric.

DOI: 10.1021/acs.biochem.5b00622
Biochemistry 2015, 54, 6029−6037

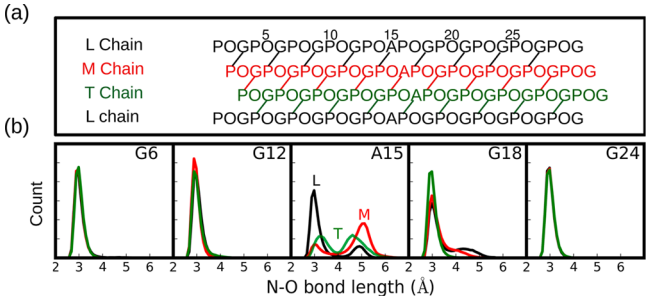


Figure 2. Hydrogen bonding topology and patterns in the G → A peptide. (a) One-residue staggering of the triple helix leads to leading (L), middle (M), and trailing (T) chains. (b) Distribution of N−O bond distances between N of the indicated residue and the carbonyl oxygen of the Pro in the neighboring chain. Chains L, M, and T are colored black, red, and green, respectively.

An analysis of the H-bond dynamics in the G → A peptide is given in Figure 3. ^{L}A15 has direct interchain H-bonds between

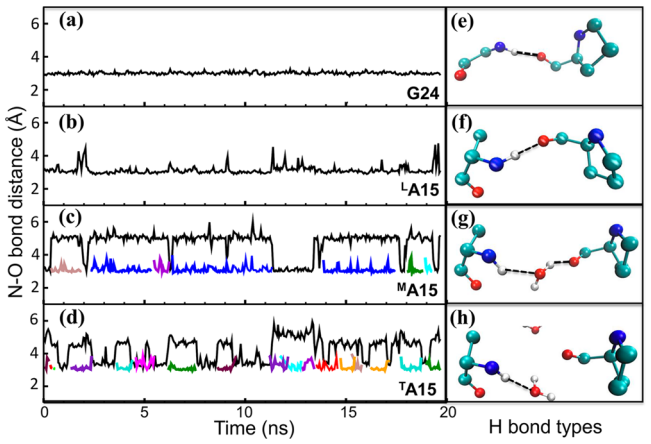


Figure 3. N−O bond distances (left) between the amide atom (N) of G24 and A15 residues and the nearby oxygen atom (O) from either a Pro carbonyl group (interchain distance colored black) or a water molecule (solvent H-bond shown as other colors; water-mediated H-bonds are always colored blue, and H-bonds to dangling water molecules are presented in different colors). For G24 (a) and ^{L}A15 (b), only interchain H-bonding is observed. For ^{M}A15 (c) and ^{T}A15 (d), alternate types of H-bonds are found. The right panel shows the most representative H-bond type corresponding to the residue in the left box, and the H-bond is indicated with a dashed line. Panels e and f show interchain H-bonds corresponding to G24 and ^{L}A15, respectively. Panel g shows an interstitial water-mediated H-bond for ^{M}A15 (colored blue in panel c). Panel h shows a dangling solvent H-bond for ^{T}A15 (colored with many colors indicating different dangling solvent bonds in panel d).

the NH and OC for almost the entire length of the MD simulation (Figure 3b,f), except for brief disruptions, e.g., at 2, 12, and 19 ns; G24 in the repeating $(G\text{-}X\text{-}Y)_n$ region is similar but lacks these brief disruptions (Figure 3a,e). By contrast, ^{M}A15 exhibits interstitial NH−H_2O−OC water bridges with lifetimes of a few nanoseconds (Figure 3c,g), and ^{T}A15 has NH−H_2O dangling water H-bonds in which the waters move in and out on the 100 ps time scale (Figure 3d,h). The loss of interchain H-bonds appears to be compensated by solvent H-bonds at all times during the simulation. The three chains are inequivalent with regard to the type of H-bonds at the mutation sites and show an asymmetry in terms of the residence times for the water H-bonds. Parallel simulations in the crystalline environment show no evidence of direct H-bonding in the L and M chains, and a partial loss of a dangling water in the T chain (Figure S1). This suggests that the differences we see relative to the crystal configuration are not artifacts of the force field but are more strongly related to the change in triple helix environment in going from crystal to solution.

NMR Chemical Shifts in the G → A Peptide Are Accurately Predicted from MD Snapshots That Contain Water-Mediated H-Bonds. NMR studies of the G → A peptide have shown the nonequivalence of the three Ala residues in solution in terms of both their chemical shifts and their H-bonding patterns.[22,23,26] The ^{1}H−^{15}N HSQC spectrum of the G → A peptide (Figure 4a) indicates that A15 gives rise

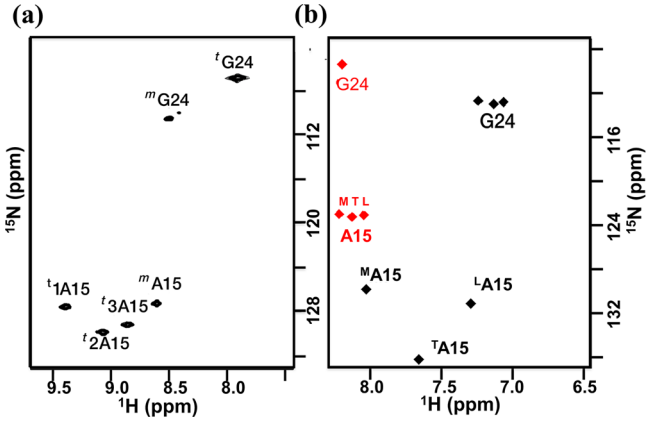


Figure 4. (a) NMR HSQC spectrum[23] of G → A peptide measured at 15 °C, where three A15 trimers (^{t}1A15, ^{t}2A15, and ^{t}3A15) are downfield-shifted in the amide proton dimension relative to the A15 monomer (^{m}A15). (b) AF-QMMM (black) and SHIFTX2 (red) calculations of chemical shifts on the 400 snapshots along 20 ns long MD simulations of the G → A peptide in the trimer state.

to three trimer peaks while G24 gives rise to only one trimer peak because of overlapping resonances arising from the similar $(POG)_4$ environment surrounding the Gly NH. Three amide protons from alanine range from 8.8 to 9.4 ppm, and the most downfield-shifted proton is bonded to the least downfield nitrogen (Figure 4a). At the time these measurements were taken, it was not possible to make assignments to specific chains, and they were simply labeled "1", "2", and "3", as in Figure 4a.[22,23] Amide proton temperature gradients (NH $\Delta\delta/\Delta T$) can be used to indicate the existence of H-bonding; a value greater than −4.6 ppb/°C suggests the existence of an internal (non-water) H-bond.[27−29] The NH $\Delta\delta/\Delta T$ measurements (Table 1) show that the H-bonding of the three chains is inequivalent; two Ala residues have a much more negative temperature dependence than the third, suggesting a lack of interchain H-bonding in two of the three chains.

Automated fragment quantum mechanics/molecular mechanics (AF-QMMM) calculations[31,32] were used to calculate NMR chemical shifts for the G24 and A15 residues using 400 MD snapshots from the last 20 ns of the simulation (Figure 4b). For comparison, the same snapshots were analyzed using the SHIFTX2 empirical model,[33] which has been calibrated against observed shifts in globular proteins (Figure 4b). (Details are given in Materials and Methods.) The calculated chemical shifts from the AF-QMMM method show a pattern similar to that observed in the experimental NMR spec-

DOI: 10.1021/acs.biochem.5b00622
Biochemistry 2015, 54, 6029−6037

Table 1. NMR Amide Proton Temperature Gradient (NH $\Delta\delta/\Delta T$) and Coupling Constants ($^3J_{HNHA}$) for Residues G24 and A15 in the G → A Peptide and the Corresponding Dihedral Angles from Experimental Data (NMR[23] and X-ray[18,19]) and MD Simulations[a]

	NH temperature gradient (ppb/°C)	$^3J_{HNHA}$ (Hz)		ϕ (deg)		
residue	NMR	NMR	MD	NMR	MD	X-ray
G24	−3.5	4.7 ± 0.3	5.6 ± 1.4	−65 ± 2	−71 ± 10	−72
^{L}A15	−4.3	5.6 ± 0.6	5.3 ± 1.6	−72 ± 5	−70 ± 14	−81
^{M}A15	−12.2	8.6 ± 0.5	8.5 ± 1.8	−96 ± 5	−114 ± 27	−104
^{T}A15	−10.7	5.6 ± 0.8	7.0 ± 1.8	−72 ± 6	−85 ± 20	−61

[a]The relation between $^3J_{HNHA}$ and ϕ dihedral angle is based on the parametrized Karplus equation:[30] $^3J_{HNHA} = 6.51\cos^2(\phi - 60) - 1.76\cos(\phi - 60) + 1.6$. The error is the standard deviation from the mean.

trum.[22,23] ^{M}A15 is significantly downfield-shifted relative to ^{T}A15 and ^{L}A15, and G24 in the $(POG)_4$ region is shifted relative to ^{L}A15. In contrast, the SHIFTX2 calculations show that the three Ala residues have a very narrow range of chemical shifts relative to each other and G24 is centered at the same ^{1}H chemical shift as the three Ala residues. The amide proton shifts in A15 in the three chains differ by only 0.2 ppm in the SHIFTX2 results, compared to 0.8 ppm in the AF-QMMM model, and 0.6 ppm in the observed spectrum. These data can be understood by looking at the distribution of amide chemical shifts for the individual snapshots using both the AF-QMMM and SHIFTX2 calculations. Figure 5a shows that individual

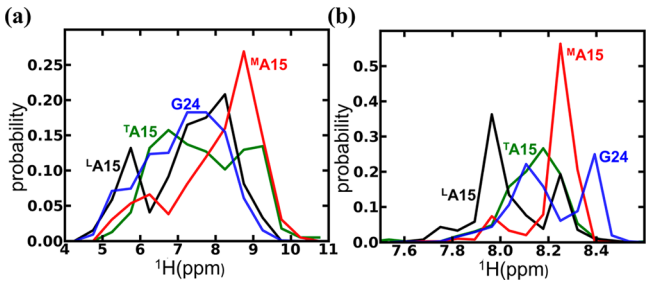


Figure 5. Distribution of amide proton ^{1}H chemical shifts using (a) AF-QMMM and (b) SHIFTX2 calculations on the 400 snapshots along a 20 ns MD simulation. Note that the range of chemical shifts is significantly wider with the AF-QMMM calculations than with SHIFTX2 prediction. See Figure S4 of the Supporting Information for results using the SHIFTS program.

snapshots using AF-QMMM have a wide range (up to 5 ppm) of proton shifts and underscore the well-known difficulties of interpreting amide proton shifts.[32,34−36] One must have both accurate estimates of shifts of individual conformers and of relative populations to obtain a useful estimate of the average resonance position. The QM/MM-calculated ^{1}H chemical shifts are systematically underestimated by ∼1.4 ppm, a pattern that has been seen in studies of folded proteins, as well.[32,37] The empirical SHIFTX2 models are trained to give average shifts from a single conformation and thus have a bias toward understating conformational variability (Figure 4b). The variation among snapshots (Figure 5b) is much smaller, ∼1 ppm for SHIFTX2. In spite of these differences, the pattern of proton shifts (L < T < M chain) is the same in both calculations, as shown in Figure 4b. Calculations using the SHIFTS model[35] also understate the differences among the A15 resonances in the three chains, although the distribution of shifts is wider than that seen in Figure 5b for SHIFTX2 (see Figure S4 of the Supporting Information). This wider distribution is in line with expectations, because SHIFTS was calibrated against quantum calculations whereas SHIFTX2 was fit to experimental data.

To explore the basis for the relationship between the values of the calculated chemical shifts and the asymmetric H-bonding distributions, correlations between the ^{1}H chemical shift and the distance to either a direct H-bond or a water-mediated H-bond were examined. The AF-QMMM simulations indicate that the amide proton chemical shift is sensitive to the type of H-bond acceptor, particularly for short H-bond distances (Figure 6a), and that H-bonding to water results in

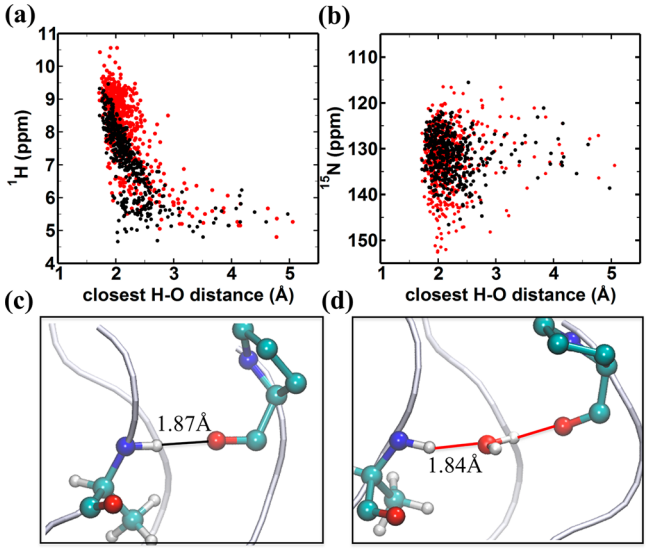


Figure 6. AF-QMMM chemical shifts for the amide hydrogen (a) and nitrogen (b) of A15, demonstrating the dependence of chemical shift on the distance to the closest oxygen atom either from the carbonyl group (black dot) of a neighboring chain or from a water molecule (red dot). (c and d) Representative snapshots with high downfield ^{1}H chemical shifts attributed to H-bonding interactions shown by solid lines for (c) a direct H-bond with a ^{1}H shift of 9.46 ppm for ^{L}A15 and (d) a water-mediated H-bond with a ^{1}H shift of 10.02 ppm for ^{M}A15.

downfield-shifted ^{1}H resonances relative to direct H-bonds of the same length. This trend is not observed for the ^{15}N chemical shift (Figure 6b). For the direct H-bonds between the Ala NH and the Pro O═C, the ^{1}H shift is very sensitive to the distance between the donor and acceptor and decreases linearly with a steep slope for values between 2.0 and 2.5 Å. Representative examples of the amide proton shift resulting from a direct H-bond for ^{L}A15 and a water-mediated H-bond for ^{M}A15 are shown in panels c and d of Figure 6. These results suggest that, at a given distance, H-bonding to water is more electron deshielding than direct H-bonding to a peptide

DOI: 10.1021/acs.biochem.5b00622
Biochemistry 2015, 54, 6029−6037

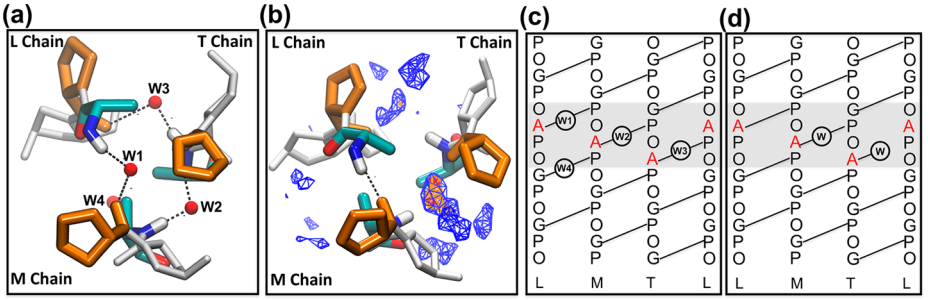


Figure 7. Water distribution around the Gly substitution sites: (a) crystal structure[18] and (b) MD solution from a top view of the N-terminal triple helix. (a) W1 bridges the amide group of ^{L}A15 in the L chain and carbonyl group of Pro in the M chain. W2 and W3 are the water bridges of ^{M}A15 and ^{T}A15, respectively. W4 represents a water bridge for ^{L}G18. (b) The MD simulated water density is shown with contour lines, which are coded by the total occupancy level of the volume. Volumes occupied >75% are colored red, >50% orange, and >25% blue. Crystal interstitial waters (W1−W3) are marked as red spheres. The three Ala residues are shown as licorice. Pro and Hyp are colored orange and gray, respectively. Schematic of the H-bonding topology around the Gly substitution sites (gray area) of a G → A peptide in the (c) crystal form and (d) solution form. Direct interchain H-bonds are indicated as solid lines, and H-bonding to water is represented with the W label in the circle.

carbonyl, and that the enhanced downfield ^{1}H chemical shift is related to the distance of the closest H-bond acceptor. These data are consistent with the fact that the distributions of chemical shifts for the different residues seen in Figure 5a resemble the population distribution for the H-bonding topology seen in Figure 2b. Both distributions of distances (Figure 2b) of the amide proton shifts (Figure 5a) are roughly bimodal for the three Ala residues: the relatively infrequent breaking of H-bonds in the leading chain [noted above (cf. Figure 3b)] is evident in the shift distribution, and the dynamic swapping of direct and water-mediated H-bonds seen in panels c and d of Figure 3 is also present in the shift distributions.

The theoretical HSQC spectra can be used to assign the chain identity and stagger to the three Ala residues in the HSQC spectrum. Residue 1A15 is assigned to ^{M}A15, residue 2A15 to ^{T}A15, and residue 3A15 to ^{L}A15 (Figure 4). The calculations predict that the most-downfield NMR chemical shift observed experimentally corresponds to the ^{M}A15 amide proton H-bond. Having the chain assignments in the NMR spectrum allows us to examine the relationship between the NMR temperature dependence and the nature of the H-bonding observed in the MD simulations. The H-bond analyses of the Ala residues from the MD simulation are in excellent agreement with the NMR amide proton temperature-dependent chemical shifts. The data suggest that the ^{L}A15 is consistent with the formation of direct H-bonds, while two other Ala residues (^{M}A15 and ^{T}A15) are consistent with the lack of dbonding occupancies. H-Bonds at the Ala mutation site and oneirect H-bonds.

A comparison of the experimental NMR[23] and theoretical $^{3}J_{HNHA}$ coupling constants (Table 1) provides further support for the assignments described above. G24 and ^{L}A15 have similar coupling constants that are consistent with a PPII conformation, whereas ^{M}A15 (both experimentally and computationally from the MD simulation) shows a larger coupling constant that is more consistent with conformational sampling of the β region. As noted above, this feature of the middle chain is also present in the G → A peptide crystal structure.[18]

DISCUSSION

Gly substitutions in the repeating $(G\text{-}X\text{-}Y)_n$ sequence of collagen result in connective tissue diseases that can range from lethal to nonlethal phenotypes; characterization of the triple helix structure is complex, and here we are interested in understanding the effect on structure and dynamics of a G to A disruption in the canonical $(Gly\text{-}X\text{-}Y)_n$ sequence. X-ray crystallography[18,19] has provided insight into such sequences via the detailed description of the simplest model peptide of OI (G → A peptide). X-ray data have shown that the central zone that surrounds the mutation site has a small bulge and that the H-bonding network is altered. All three alanine residues have a water-mediated H-bond, rather than a direct H-bond, from the alanine NH to the carbonyl oxygen of the proline on the X residue. As opposed to the X-ray structure, solution NMR studies[22,23] have shown that the disruption at the alanine substitution site leads to an asymmetry of the three chains; the alanine amide proton in one of the chains is significantly downfield-shifted relative to the other two, and the chemical shift temperature dependence suggests that only a single chain contains a direct H-bond. Because of the anisotropic shape of the peptide, the one-residue stagger, and the screw symmetry of the three supercoiled PPII chains, it has been impossible to assign the resonances to the individual chains and interpret these unusual chemical shifts and temperature-dependent data using NMR alone.[22−24]

MD simulations have been performed on CMPs to provide a structural and dynamic view of the collagen triple helix to improve our understanding of recognition regions and the nature of Gly interruptions.[38−43] The G → A peptide was studied in early MD simulations by Klein and Huang.[40] These used an earlier version of the Amber force field, which included a small shell of 267 water molecules, and conducted the runs for 0.5−1 ns. Hydrogen bonding patterns changed very little during the course of the simulation, and almost no bending was seen. Improvements in programs and computers now permit calculations like those presented here, which employ periodic boundary conditions, ∼27 000 water molecules, and simulation times that are two orders of magnitude longer than those available in 1999. In particular, we note that much of the transition to three inequivalent chains took place only after simulation for 40 ns, as illustrated in Figure S2 of the Supporting Information.

Other recent MD simulations of triple helices have focused on hydration effects in a variety of sequence contexts.[44−49] A striking difference between the crystal structure and the MD solution structures is the water density and its mobility around the mutation sites (Figure 7). In the crystal, the amide groups

DOI: 10.1021/acs.biochem.5b00622
Biochemistry 2015, 54, 6029−6037

of all three Ala residues (and one Gly, which is one triplet C-terminal to Ala) do not have direct H-bonds but instead have interstitial water bridges (Figure 7a). In contrast to the equivalent environment in the crystal structure of the three Ala residues, the solution MD simulations indicate an inequivalent distribution of solvent around the three Ala residues such that ^{L}A15 has no water occupancy and ^{M}A15 has the highest-density (>75%) contour of water occupancy (Figure 7b). A schematic of the H-bonding in the crystal and MD simulations indicates the differences in water-mediated H-bonding patterns (Figure 7c,d).

A direct interpretation of the unusual NMR chemical shifts and temperature-dependent data is obtained from the MD results for the G → A peptide reported here. The H-bonding patterns in the NMR temperature-dependent chemical shift experiments are mirrored in the MD simulations and together support the view that different chains display different H-bonding occupancies. H-bonds at the Ala mutation site and one triplet C-terminal to it have complex dynamics in which water is moving in and out to break or form interchain H-bonds. Specifically, one Ala NH is involved in water-mediated H-bonding, while another has a direct H-bond showing that the H-bonding configuration is different at the three axial levels and that one Ala can maintain a direct H-bond in the context of a G → A mutation while the other two cannot.

We note the strong agreement between the calculated chemical shifts from the MD simulations and the experimental NMR chemical shifts. Using AF-QMMM calculations,[31,32] the calculated chemical shifts are able to reproduce both the unusually downfield shifted Ala ^{1}H chemical shift and the wide range of chemical shifts that arise for the three distinct Ala residues. Attempts at predicting the chemical shifts for the triple helix using empirical approaches such as SHIFTX2[33] result in a significantly weaker correlation with experimental data; notably, the range of chemical shifts in the empirical calculations is extremely narrow. The AF-QMMM calculations provide insight into the reasons for the unusually downfield shifted A15 resonance and the range of chemical shifts observed at the A15 position. They suggest that there is a strong dependence of the chemical shift on the population of interstitial water relative to the population of backbone NH amide protons that have direct H-bonds. The better agreement of the AF-QMMM calculations with the experimental data relative to SHIFTX2 suggests that inclusion of the solvent is critical for obtaining more accurate chemical shift calculations[32] and highlights the sensitivity of the ^{1}H chemical shift to the complex dynamics of the H-bonding occupancies.

The NMR conformation and dynamics of the G → A peptide can be compared with those of a G → S peptide [T1-898(G901S)] that models a natural glycine replaced with serine at site 901 associated with a mild case of OI.[23] Peptide T1-898(G901S) [sequence of Ac-$(GPO)_4$-GPV-SPA-GAR-$(GPO)_4$GY-$CONH_2$] contains the collagen sequence GPV-SPA-GAR from position 898 to 906 in the α1 chain, thereby making this a more realistic model of OI. As in the G → A peptide, T1-898(G901S) shows an asymmetry in the solution behavior of the three chains and increased flexibility C-terminal to the mutation site relative to the N-terminal site. Strikingly, the chemical shift and temperature-dependent chemical shift profile for T1-898(G901S) are similar to those of the G → A peptide. One serine resonance (trailing chain) is highly downfield shifted relative to the other two, and another serine (middle chain) shows temperature-dependent chemical shift data suggestive of a direct H-bond. We are investigating whether the downfield-shifted resonance in T1-898(G901S) is due to the presence of interstitial water, but the MD simulations of the G → A peptide and the similarity of the NMR parameters are highly suggestive. In both NMR studies and MD simulations, we observe an asymmetry in the dynamics of the C-terminus relative to the N-terminus. This is consistent with previous suggestions that clinical severity is related to dynamics immediately adjacent to the mutation site[24,50] and to the existence of destabilizing sequences C-terminal to the mutation site.[22,50−52]

OI is a disease that arises from the substitution of a single glycine to another residue in fibrillar collagens, which have an absolute requirement for Gly as every third residue. However, nonfibrillar collagens, such as type IV collagen in basement membranes, also contain breaks in the $(Gly\text{-}X\text{-}Y)_n$ repeating pattern.[53,54] In nonfibrillar collagens, these breaks are not associated with disease but rather may be involved in molecular recognition or flexibility.[54] Interruptions in nonfibrillar collagens may be of variable length and are classified by the number of residues between the repeating Gly residues; for example, Gly residues that are separated by four residues are termed G4G interruptions.[53] Single Gly substitutions can be considered G5G interruptions (for example, GPOAPOG), suggesting that the properties that are observed for the OI substitution peptides may be relevant to peptides that model interruptions.[26,53,55] We have previously shown that peptides[53,56,57] that model natural interruptions in the $(Gly\text{-}X\text{-}Y)_n$ sequence show NMR patterns similar to those of peptides that model G to X mutations. NMR studies[56,57] of the GAAVMG peptide (G4G) show a highly localized structural perturbation with close packing of the Val residue near the central axis, as seen in the G → A and T1-898(G901S) peptides. The intriguing trends in chemical shifts that we have observed in peptides that contain Gly substitutions are also observed in the GAAVM peptide; asymmetry in the behavior of the three chains at the $(GXY)_n$ break is observed along with ^{1}H chemical shifts that are significantly downfield shifted for one chain relative to the other two. On the basis of the MD simulations for the G → A peptide, we suggest that the downfield-shifted resonance arises from water-mediated H-bonding at one of the Val residues in the interruption region. This is consistent with hypotheses by Bella,[26] based on the existing crystal structure for a G1G interruption, Hyp^{-} (···POGPGPOG···), that the H-bonding topology for a G4G peptide (which is commensurate with a G1G peptide) would contain water-mediated H-bonds. The complementary NMR and MD data suggest that disruption of the repeating $(Gly\text{-}X\text{-}Y)_n$ sequence either through Gly to X mutations or through natural interruptions may result in the introduction of interstitial water H-bonds at the noncanonical $(G\text{-}X\text{-}Y)_n$ sites. We can explore in this unique system the differences between direct (interchain) H-bonding and water-mediated interactions in ways that are generally not possible with globular proteins. These results, combined with NMR data for other peptides, suggest a general model for the structural and dynamic consequences of glycine interruptions in collagen.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.5b00622.

DOI: 10.1021/acs.biochem.5b00622
Biochemistry 2015, 54, 6029−6037

Distribution of N−O bond distances for A15 in the crystal simulations of a G → A peptide (Figure S1), tracking the H−O bond distance between Ala NH and the carbonyl oxygen of Pro in the neighbor for each A15 residue over a 170 ns MD solution simulation (Figure S2), backbone fluctuations for a 1 μs solution simulation (Figure S3), and distribution of amide proton shifts using the SHIFTS program (Figure S4) (PDF)

AUTHOR INFORMATION

Corresponding Authors

*Center of Integrative Proteomics Research, Rm. 208b, 174 Frelinghuysen Rd., Piscataway, NJ 08854. Phone: +1-848-445-5885. E-mail: case@biomaps.rutgers.edu.

*Center of Integrative Proteomics Research, 174 Frelinghuysen Rd., Piscataway, NJ 08854. Phone: +1-848-445-5254. E-mail: baum@chem.rutgers.edu.

Funding

This work was supported by National Institutes of Health Grants GM45811 (D.A.C.) and GM45302 (J.B.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Helen Berman for many helpful discussions.

REFERENCES

(1) Myllyharju, J., and Kivirikko, K. I. (2004) Collagens, Modifying enzymes and their mutations in humans, flies and worms. *Trends Genet. 20*, 33−43.

(2) Shoulders, M. D., and Raines, R. T. (2009) Collagen structure and stability. *Annu. Rev. Biochem. 78*, 929−958.

(3) Heino, J. (2007) The collagen family members as cell adhesion proteins. *BioEssays 29*, 1001−1010.

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

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

(6) 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., et al. (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.

(7) Kuivaniemi, H., Tromp, G., and Prockop, D. J. (1997) Mutations in fibrillar collagens (types I, II, III, and XI), fibril-associated collagen (type IX), and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels. *Hum. Mutat. 9*, 300−315.

(8) Byers, P. H., and Cole, W. G. (2002) Osteogenesis Imperfecta. In *Connective Tissue and Its Heritable Disorders: Molecular, Genetic, and Medical Aspects* (Royce, P. M., and Steinmann, B., Eds.) 2nd ed., John Wiley & Sons, Inc., Hoboken, NJ.

(9) Forlino, A., Cabral, W. A., Barnes, A. M., and Marini, J. C. (2011) New perspectives on Osteogenesis Imperfecta. *Nat. Rev. Endocrinol. 7*, 540−557.

(10) Makareeva, E., Cabral, W. A., Marini, J. C., and Leikin, S. (2006) Molecular mechanism of alpha1(I)-Osteogenesis imperfecta/Ehlers-Danlos syndrome: unfolding of an N-anchor domain at the N-terminal end of the type I collagen triple helix. *J. Biol. Chem. 281*, 6463−6470.

(11) Brodsky, B., and Persikov, A. V. (2005) Molecular structure of the collagen triple helix. *Adv. Protein Chem. 70*, 301−339.

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

(13) Fields, G. B. (2010) Synthesis and biological applications of collagen-model triple-helical peptides. *Org. Biomol. Chem. 8*, 1237−1258.

(14) Fallas, J. A., O'Leary, L. E. R., and Hartgerink, J. D. (2010) Synthetic collagen mimics: self-assembly of homotrimers, heterotrimers and higher order structures. *Chem. Soc. Rev. 39*, 3510−3527.

(15) Xu, Y. (2009) Thermal stability of collagen triple helix. *Methods Enzymol. 466*, 211−232.

(16) Long, C. G., Li, M. H., Baum, J., and Brodsky, B. (1992) Nuclear magnetic resonance and circular dichroism studies of a triple-helical peptide with a glycine substitution. *J. Mol. Biol. 225*, 1−4.

(17) Gauba, V., and Hartgerink, J. D. (2008) Synthetic collagen heterotrimers: structural mimics of wild-type and mutant collagen type I. *J. Am. Chem. Soc. 130*, 7509−7515.

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

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

(20) Xiao, J., and Baum, J. (2009) Structural insights from ^{15}N relaxation data for an anisotropic collagen peptide. *J. Am. Chem. Soc. 131*, 18194−18195.

(21) Bhate, M., Wang, X., Baum, J., and Brodsky, B. (2002) Folding and conformational consequences of glycine to alanine replacements at different positions in a collagen model peptide. *Biochemistry 41*, 6539−6547.

(22) Buevich, A., and Baum, J. (2001) Nuclear magnetic resonance characterization of peptide models of collagen-folding diseases. *Philos. Trans. R. Soc., B 356*, 159−168.

(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) 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.

(25) Okuyama, K., Hongo, C., Fukushima, R., Wu, G., Narita, H., Noguchi, K., Tanaka, Y., and Nishino, N. (2004) Crystal structures of collagen model peptides with Pro-Hyp-Gly repeating sequence at 1.26 A resolution: implications for proline ring puckering. *Biopolymers 76*, 367−377.

(26) Bella, J. (2014) A first census of collagen interruptions: collagen's own stutters and stammers. *J. Struct. Biol. 186*, 438−450.

(27) Cierpicki, T., and Otlewski, J. (2001) Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J. Biomol. NMR 21*, 249−261.

(28) Cierpicki, T., Zhukov, I., Byrd, R. A., and Otlewski, J. (2002) Hydrogen bonds in human ubiquitin reflected in temperature coefficients of amide protons. *J. Magn. Reson. 157*, 178−180.

(29) Tomlinson, J. H., and Williamson, M. P. (2012) Amide temperature coefficients in the protein G B1 domain. *J. Biomol. NMR 52*, 57−64.

(30) Vuister, G. W., and Bax, A. (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNHA) coupling constants in 15N-enriched proteins. *J. Am. Chem. Soc. 115*, 7772−7777.

(31) He, X., Wang, B., and Merz, K. M. (2009) Protein NMR chemical shift calculations based on the automated fragmentation QM/MM approach. *J. Phys. Chem. B 113*, 10380−10388.

(32) Zhu, T., Zhang, J. Z. H., and He, X. (2013) Automated fragmentation QM/MM calculation of amide proton chemical shifts in proteins with explicit solvent model. *J. Chem. Theory Comput. 9*, 2104−2114.

(33) Han, B., Liu, Y., Ginzinger, S. W., and Wishart, D. S. (2011) SHIFTX2: Significantly improved protein chemical shift prediction. *J. Biomol. NMR 50*, 43−57.

DOI: 10.1021/acs.biochem.5b00622
Biochemistry 2015, 54, 6029−6037

(34) He, X., Wang, B., and Merz, K. M. (2009) Protein NMR chemical shift calculations based on the automated fragmentation QM/MM approach. *J. Phys. Chem. B 113*, 10380−10388.
(35) Case, D. A. (2013) Chemical shifts in biomolecules. *Curr. Opin. Struct. Biol. 23*, 172−179.
(36) Sitkoff, D., and Case, D. A. (1997) Density functional calculations of proton chemical shifts in model peptides. *J. Am. Chem. Soc. 119*, 12262−12273.
(37) Jiang, X. N., and Wang, C. S. (2009) Rapid prediction of the hydrogen bond cooperativity in N-methylacetamide chains. *ChemPhysChem 10*, 3330−3336.
(38) Stultz, C. M. (2002) Localized unfolding of collagen explains collagenase cleavage near imino-poor sites. *J. Mol. Biol. 319*, 997−1003.
(39) Stultz, C. M. (2006) The folding mechanism of collagen-like model peptides explored through detailed molecular simulations. *Protein Sci. 15*, 2166−2177.
(40) Aliev, A. E., Courtier-Murias, D., Bhandal, S., and Zhou, S. (2010) A combined NMR/MD/QM approach for structure and dynamics elucidation in the solution state: pilot studies using tetrapeptides. *Chem. Commun. 46*, 695−697. Aliev, A. E., Kulke, M., Khaneja, H. S., Chudasama, V., Sheppard, T. D., and Lanigan, R. M. (2014) Motional timescale predictions by molecular dynamics simulations: Case study using proline and hydroxyproline sidechain dynamics. *Proteins: Struct., Funct., Genet. 82*, 195−215.
(41) Mooney, S. D., Huang, C. C., Kollman, P. A., and Klein, T. E. (2001) Computed free energy differences between point mutations in a collagen-like peptide. *Biopolymers 58*, 347−353.
(42) Mooney, S. D., and Klein, T. E. (2002) Structural models of Osteogenesis Imperfecta-associated variants in the COL1A1 Gene. *Mol. Cell. Proteomics 1*, 868−875.
(43) 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.
(44) Vitagliano, L., Berisio, R., and De Simone, A. (2011) Role of hydration in collagen recognition by bacterial adhesins. *Biophys. J. 100*, 2253−2261.
(45) Ravikumar, K. M., and Hwang, W. (2008) Region-specific role of water in collagen unwinding and assembly. *Proteins: Struct., Funct., Genet. 72*, 1320−1332.
(46) Gurry, T., Nerenberg, P. S., and Stultz, C. M. (2010) The contribution of interchain salt bridges to triple-helical stability in collagen. *Biophys. J. 98*, 2634−2643.
(47) Raman, S. S., Gopalakrishnan, R., Wade, R. C., and Subramanian, V. (2011) Structural basis for the varying propensities of different amino acids to adopt the collagen conformation. *J. Phys. Chem. B 115*, 2593−2607.
(48) Singam, E. R. A., Balamurugan, K., Gopalakrishnan, R., Subramanian, S. R., Subramanian, V., and Ramasami, T. (2012) Molecular dynamic simulation studies on the effect of one residue chain staggering on the structure and stability of heterotrimeric collagen-like peptides with interruption. *Biopolymers 97*, 847−863.
(49) Teng, X., and Hwang, W. (2014) Chain registry and load-dependent conformational dynamics of collagen. *Biomacromolecules 15*, 3019−3029.
(50) Bodian, D. L., Madhan, B., Brodsky, B., and Klein, T. E. (2008) Predicting the clinical lethality of Osteogenesis Imperfecta from collagen glycine mutations. *Biochemistry 47*, 5424−5432.
(51) Hyde, T. J., Bryan, M. A., Brodsky, B., and Baum, J. (2006) Sequence dependence of renucleation after a Gly mutation in model collagen peptides. *J. Biol. Chem. 281*, 36937−36943.
(52) 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.
(53) 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.
(54) 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.
(55) Hwang, E. S., Thiagarajan, G., Parmar, A. S., and Brodsky, B. (2010) Interruptions in the collagen repeating tripeptide pattern can promote supramolecular association. *Protein Sci. 19*, 1053−1064.
(56) 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.
(57) 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.
(58) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Genet. 65*, 712−725.
(59) Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowley, M., Walker, R. C., Zhang, W., Merz, K. M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., and Kollman, P. A. (2008) *AMBER 10*, University of California, San Francisco.
(60) Roe, D. R., and Cheatham, T. E. (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput. 9*, 3084−3095.
(61) Tang, S., and Case, D. A. (2011) Calculation of chemical shift anisotropy in proteins. *J. Biomol. NMR 51*, 303−312.
(62) 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.
(63) 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.

DOI: 10.1021/acs.biochem.5b00622
Biochemistry 2015, 54, 6029−6037