Severity of Osteogenesis Imperfecta and Structure of a Collagen-like Peptide Modeling a Lethal Mutation Site†

Randall J. Radmer and Teri E. Klein*

Department of Genetics, School of Medicine, Stanford University, Stanford, California 94305

Received September 17, 2003; Revised Manuscript Received March 6, 2004

ABSTRACT: We show that there are correlations between the severities of osteogenesis imperfecta (OI) phenotypes and changes in the residues near the mutation site. Our results show the correlations between the severity of various forms of the inherited disease OI and alteration of residues near the site of OI causing mutations. Among our many observed correlations are particularly striking ones between the presence of nearby proline residues and lethal mutations, and the presence of nearby alanines residues and nonlethal mutations. We investigated the possibility that these correlations have a structural basis using molecular dynamics simulations of collagen-like molecules designed to mimic the site of a lethal OI mutation in collagen type I. Our significant finding is that interchain hydrogen bonding is greatly affected by variations in residue type. We found that the strength of hydrogen bond networks between backbone atoms on different chains depends on the local residue sequence and is weaker in proline-rich regions of the molecule. We also found that an alanine at a site near an OI mutation causes less structural disruption than a proline, and that residue side chains also form interchain hydrogen bonds with frequencies that are dependent on residue type. For example, arginine side chains form strong hydrogen bonds with the backbone of the subsequent peptide chain, while lysine and glutamine less frequently form similar hydrogen bonds. This decrease in the observed hydrogen bond frequency correlates with a decrease in the experimentally determined thermal stability. We contrasted general structural properties of model collagen peptides with and without the mutation to examine the effect of the single-point mutation on the surrounding residues.

Collagen is the most abundant protein in higher animals, accounting for approximately one-quarter of the total protein mass of vertebrates. It adopts a ropelike structure that is strong with respect to tension, with a tensile strength by weight approaching that of steel. Its primary function is structural, and it is the primary protein constituent of higher-animal frameworks, including bones, tendons, ligaments, skin, blood vessels, and supporting membranous tissue. Collagen's obvious structural role and relatively homogeneous structure have led to the common view that it plays a passive supporting role.

However, examination of typical extracellular matrices (ECMs)[1] reveals that collagen's role is more complex than simply providing tensile strength; it also self-assembles into complex frameworks that support the ECM and provide specific binding sites for a wide range of other constituents of the ECMs (*1*). A truncated list of ECM constituents that are known to bind to collagen (*2*, *3*) includes calcium ions, which bind to collagen scaffolding to form bones; proteoglycans, which provide compressive strength to the ECM; fibronectins, which link ECM molecules; and integrins, which are transmembrane proteins that control cellular interaction with and binding to the ECM.

When collagen's essential role in the development and maintenance of the ECM is considered, it is not surprising that mutations that disrupt its function can result in a variety of pathologies. Examples include inherited diseases such as osteogenesis imperfecta (OI) and Ehlers Danlos syndrome (EDS), which are characterized by fragile bones, weak tendons, and thin skin (*4*). The clinical severity of OI can range from mild (characterized by blue sclera and diminutive stature) to prenatally lethal (where the skeletal system fails to develop).

Although the existence of a connection between certain collagen mutations and the diseases described above has been well established, the mechanism by which mutations cause the disease is still poorly understood. Possible explanations include the loss of molecular stability, the loss of the ability to correctly associate with other collagen molecules, and the loss of critical binding sites. In general, most explanations assume a change in structure that results in the loss or change of some function at the structural level or downstream effect in the collagen biosynthesis pathway.

The question of how the severity of various forms of OI relates to the type and location of their causal mutations has caused extensive research of the structure and stability of

† This work was funded by National Institutes of Health Grant AR47720-01 (T.E.K., principal investigator).

* To whom correspondence should be addressed. E-mail: teri.klein@stanford.edu.

[1] Abbreviations: ECM, extracellular matrix; OI, osteogenesis imperfecta; EDS, Ehlers Danlos syndrome; Hyp, hydroxyproline; Gly−Xaa hydrogen bonds, bonds between the backbone amine of a glycine residue and the backbone carbonyl of a Yaa residue on the subsequent chain; Yaa−Yaa hydrogen bonds, bonds between a hydrogen bond donor on the side chain of a Yaa residue and the backbone carbonyl of a Yaa residue on the subsequent chain.

10.1021/bi035676w CCC: $27.50
Published on Web 04/14/2004

Table 1: Frequency of OI Severity, Based on Nearby Residue Types[a]

residue	offset	frequency (%) (no. of occurrences) lethal[b]	frequency (%) (no. of occurrences) nonlethal[c]	significance (%)[d]	residue	offset	frequency (%) (no. of occurrences) lethal[b]	frequency (%) (no. of occurrences) nonlethal[c]	significance (%)[d]
alanine	−5	31 (8)	69 (18)	6.5	histidine	+4	33 (1)	67 (2)	53.9
	−4	39 (9)	61 (14)	28.0	hydroxyproline	−4	56 (19)	44 (15)	20.7
	−2	19 (5)	82 (22)	0.2		−1	43 (16)	57 (21)	36.6
	−1	48 (12)	52 (13)	55.5		+2	50 (19)	50 (19)	43.7
	+1	39 (5)	62 (8)	35.8		+5	41 (14)	59 (20)	29.0
	+2	33 (6)	67 (12)	16.9	leucine	−5	60 (3)	40 (2)	45.2
	+4	37 (7)	63 (12)	24.5		−2	29 (2)	71 (5)	27.1
	+5	64 (9)	36 (5)	16.0		+1	50 (4)	50 (4)	57.9
arginine	−4	46 (6)	54 (7)	57.5		+4	43 (3)	57 (4)	55.6
	−2	33 (1)	67 (2)	53.9	lysine	−4	17 (1)	83 (5)	13.5
	−1	33 (3)	67 (6)	30.7		−1	60 (3)	40 (2)	45.2
	+1	33 (1)	67 (2)	53.9		+1	43 (3)	57 (4)	55.6
	+2	100 (6)	0 (0)	1.1		+2	33 (2)	67 (4)	39.3
	+4	43 (3)	57 (4)	55.6		+4	33 (1)	67 (2)	53.9
	+5	40 (6)	60 (9)	37.8		+5	50 (5)	50 (5)	55.8
asparagine	−4	0 (0)	100 (3)	14.5	methionine	+1	33 (1)	67 (2)	53.9
	−2	67 (2)	33 (1)	46.1	phenylalanine	+4	33 (1)	67 (2)	53.9
	+2	67 (2)	33 (1)	46.1	proline	−5	54 (22)	46 (19)	26.0
	+4	33 (1)	67 (2)	53.9		−2	67 (26)	33 (13)	1.2
	+5	33 (1)	67 (2)	53.9		+1	58 (21)	42 (15)	12.6
aspartic acid	−1	67 (2)	33 (1)	46.1		+4	56 (15)	44 (12)	25.6
	+1	75 (3)	25 (1)	27.5	serine	−5	33 (1)	67 (2)	53.9
	+2	36 (4)	64 (7)	33.5		−1	67 (2)	33 (1)	46.1
	+4	86 (6)	14 (1)	4.7		+1	43 (3)	57 (4)	55.6
	+5	50 (2)	50 (2)	64.8		+2	33 (1)	67 (2)	53.9
glutamic acid	−5	57 (4)	43 (3)	44.4	threonine	−1	67 (2)	33 (1)	46.1
	−2	44 (4)	56 (5)	56.3		+5	50 (2)	50 (2)	64.8
	+1	43 (3)	57 (4)	55.6	valine	−4	67 (4)	33 (2)	29.7
	+4	42 (5)	58 (7)	45.8		−1	25 (1)	75 (3)	35.2
glutamine	−5	33 (1)	67 (2)	53.9		+5	67 (2)	33 (1)	46.1
	−4	67 (2)	33 (1)	46.1					
	−1	100 (3)	0 (0)	10.7					
	+1	25 (1)	75 (3)	35.2					
	+2	40 (2)	60 (3)	54.8					
	+5	67 (4)	33 (2)	29.7					

[a] The observed relative frequency of lethal, nonlethal, and mild OI based on the existence of nearby residues for chain 1 of collagen type I. The position indicates the location of the residue relative to the mutation site. The OI mutations are from the Database of Human Collagen Mutations (*26*, *27*) (www.le.ac.uk/genetics/collagen). [b] Observed frequencies of OI type II based on a total of 46 lethal mutations. [c] Observed frequencies of OI type I, type III, and type IV based on a total of 51 nonlethal mutations. [d] Probability that the observed distribution of lethal vs nonlethal mutations would occur by chance assuming a binomial distribution.

normal and mutant collagen. Full-sized collagen molecules are difficult to study; thus, model collagen peptides of ~30 residues per chain are often used. For example, Brodsky and co-workers have published a number of experimental examinations of the structure and stability of collagen-like peptides designed to mimic regions of native collagen with known OI mutation sites (*5*−*7*). Our group (*8*−*10*) and others (*11*−*13*) have used molecular dynamics simulations to explore the differing details of collagen structure.

In this study, we address the question of how residues near mutation sites influence the degree of OI severity. After presenting statistical correlations between OI severity and the types of residues near OI mutation sites (see Table 1), we use molecular dynamics simulations to examine the structure and dynamics of collagen-like peptides that were designed to model the site of a lethal mutation (G913S) in collagen type I. The simulation results provide insight into the relationship between molecular stability and residue type. Our results indicate a structural connection for some of the correlations between OI severity and nearby residues.

Ultimately, this work will help explain the structural basis underlying the mutations that lead to the resulting disease, as well as a better understanding of why there is a large variation in the observed severity of diseases resulting from seemingly similar mutations. We also believe that this research will help make it possible in the future to predict the severity of unknown mutations.

MATERIALS AND METHODS

Molecular dynamics simulations were performed using the Sander module of the AMBER 7 (*14*) molecular simulation package and the force field of Cornell *et al.* (*15*). For simulations requiring neutral aspartic and glutamic acid residues, standard AMBER ASH and GLH residues were used. The ornithine residue was built by removing the γ-carbon and attached hydrogens from the lysine residue of Cornell *et al.* (*15*). The total charge of the eliminated atoms ($0.0393q$) was divided in half and added to ornithine's β- and γ-carbons to give a net residue charge of 1.

The generalized Born solvation model developed by Onufriev *et al.* (*16*, *17*) was used with a solution ion concentration of 0.2 M. Every 5.0 ps, velocities of each atom were randomly reassigned from a Maxwellian distribution. The SHAKE algorithm was used to fix all bond lengths involving H-atoms to their equilibrium distances. A 2.0 fs time step and 12.0 Å cutoff were used in all cases.

Table 2: Simulated Peptides, Sequences, and Melting Temperatures

Label	Peptide[a]	Sequence[b]	T_m/°C
I	T1-904[c]	[Ac-GARGPAGPQGPRGDKGETGPOGPOGPOGPOGV-Nme]$_3$	30.8
II	T1-904[G11S][c]	[Ac-GARGPAGPQSPRGDKGETGPOGPOGPOGPOGV-NMe]$_3$	8.9
III	T1-904[G11S-P9A]	[Ac-GARGPAGAQSPRGDKGETGPOGPOGPOGPOGV-NMe]$_3$	-
IV	T1-904[pH 2][cd]	[Ac-GARGPAGPQGPRGDKGETGPOGPOGPOGPOGV-NMe]$_3$	22.1
	P1-GPO[e]	[Ac-GPOGPOGPOGPOGPOGPOGPOGPOGG-NH_2]$_3$	47.3
	P1-GPO[O13R][e]	[Ac-GPOGPOGPOGPRGPOGPOGPOGPOGG-NH_2]$_3$	47.2
	P1-GPO[O13Q][e]	[Ac-GPOGPOGPOGPQGPOGPOGPOGPOGG-NH_2]$_3$	41.3
	P1-GPO[O13K][e]	[Ac-GPOGPOGPOGPKGPOGPOGPOGPOGG-NH_2]$_3$	36.8
	P1-GPO[O13N][e]	[Ac-GPOGPOGPOGPNGPOGPOGPOGPOGG-NH_2]$_3$	30.3

[a] Residue numbering starts with the acetyl-capping group (e.g., Ace1, Gly2, etc.) and continues to the *N*-methylamine cap (Nme32). [b] O represents a hydroxyproline residue. [c] From ref *5*. [d] All acidic residues protonated to give the neutral form. [e] From ref *22*.

For each simulated molecule, all three chains were identical (see Table 2). Homotrimer models were used for comparisons between the computational models with published experimental data of homotrimer molecules. Atomic level experimental data on heterotrimer collagen-like peptides are limited. The T1−904 (**I**) peptide was built using gencollagen (*18*) and capped with acetyl and *N*-methylamine end groups. It was minimized for 100 cycles and equilibrated at 10 K for 10 ps, 100 K for 30 ps, 200 K for 40 ps, and 275 K for 120 ps. A 20 ns production simulation was carried out at 275 K [consistent with the experimental conditions of Yang *et al*. (*5*)]. Atomic coordinates were written to disk every 1.0 ps.

Initial structures of the mutant and low-pH peptides (**II**−**IV**) were built using the structure of partially equilibrated **I**, saved at the end of the 200 K equilibration described above. Each of these structures was edited to change residues where appropriate, then minimized, equilibrated, and simulated as described above.

Each simulation was performed using eight nodes of an SGI Origin 3800 class supercomputer, requiring ∼12 h for each nanosecond of simulation time.

The P1-GPO peptides (see Table 2) were also built using gencollagen (*18*) and were capped with acetyl and NH_2 end groups. Equilibration and simulation conditions were as described above, except the total production run was 5 ns in length.

The hydrogen bonding frequencies and correlations in addition to average Φ and Ψ angles were found using the saved coordinates and a program developed by the authors. For the hydrogen bond statistics, it was assumed that all donors formed a maximum of one interaction at any given time, but the number of interactions per acceptors was unlimited. For the results reported here, hydrogen-bonding groups had to be within 2.2 Å of each other to be identified as being involved in hydrogen bonding. For each peptide, up to 10 000 structures were saved and used in the analysis.

RESULTS AND DISCUSSION

Collagen Structure and OI Mutations. Fibrillar collagen molecules, including types I−III, V, and XI, consist of three polypeptide chains twisted into a ropelike, right-handed triple-helical structure. The chains are long, generally consisting of ∼1000 residues each. The three chains can be homogeneous (collagen types II and III each consist of only one chain type) or heterogeneous (collagen type I consists of two identical chains and a third similar but nonidentical chain). The structure of each chain is a left-handed helix with a helical pitch of approximately three residues per turn, called type II polyproline. Typical fibrillar collagen sequences consist of the repeating residue triplet Gly-Xaa-Yaa, where Xaa and Yaa are often the imino acids proline and hydroxyproline, respectively.

The glycine in the residue triplet is essential in fibrillar collagens because it is positioned at the point of contact between the three chains that form the triple helix. A mutation of one of these critical glycines to a larger residue is often responsible for OI and EDS. The larger side chain of any other residue at this location hits the other chains of the triple helix and prevents the formation of a stable complex. Other factors beyond the physical replacement of the core glycine with any other residue that seem to influence severity include residue size, the ability to interact with other residues (by hydrogen bonding or, in the case of cystine, covalent bonding), backbone flexibility, and local structure. We discuss these factors in depth below.

Imino Acid Effect. Imino acids proline and hydroxyproline are critical for maintaining the stability of the triple-helical region of native collagen molecules. Their rigid structure prevents them from adopting many of the conformations that are accessible to other residues, but does permit the adoption of the polyproline type II structure seen in triple helices (*19*). This results in a reduced entropic cost of molecular formation and increased helical stability, relative to the other amino acids.

Hydrogen Bonding Effects. In addition to the stabilizing effects described above, interchain hydrogen bonding also contributes to collagen stability (*20*). The type of hydrogen bond most often found in collagen triple helices (see Figure 1) forms between the backbone amine of the invariant glycines and the backbone carbonyl oxygen of an Xaa residue on the subsequent chain (Gly-Xaa hydrogen bonds). In the case of an idealized collagen peptide, (Gly-Pro-Hyp)$_n$, these are the only possible intramolecular hydrogen bonds. This is because proline and hydroxyproline have no hydrogen bonding groups that can interact directly with other groups on the molecule (unlike amino acids, imino acid's peptide bonds lack a hydrogen that can act as a hydrogen bond

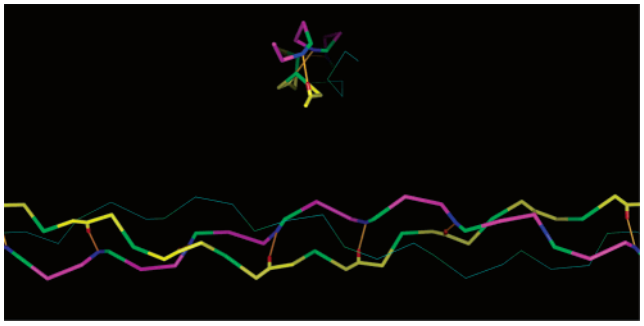

FIGURE 1: α-Carbons, representative interchain hydrogen bonds, and associated atoms of the collagen-like peptide. Chain A is colored magenta, chain B yellow, and chain C cyan, and glycine α-carbons are colored green. The top structure is a truncated end view showing α-carbons only (N-terminal residues are at the front). Glycine nitrogen atoms are colored blue, and Xaa residue oxygen atoms are colored red. Because of the complexity of the full hydrogen bonding network, chains A and B are emphasized, and only the hydrogen bonds linking these two chains are shown (orange). Note that all these hydrogen bonds have the same orientation; all donors are glycine nitrogen atoms on chain A, and all acceptors are the backbone carbonyl groups of Xaa residues on chain B. The same pattern occurs between chain B and chain C and between chain C and chain A (in each case, the first chain provides the hydrogen bond donor and the second the hydrogen bond acceptor). This structure is Protein Data Bank entry 1K6F.

donor). The triple helices have no *intrachain* backbone−backbone hydrogen bonds.

In naturally occurring collagen, many of the residues have side chains capable of forming hydrogen bonds with one of the other chains. A particularly important type of hydrogen bond involves Yaa residues that have side chains with a hydrogen bond donor group (e.g., arginine, lysine, and glutamine), permitting formation of hydrogen bonds to the carbonyl oxygen of equivalent Yaa residues on the subsequent chain (Yaa−Yaa hydrogen bonds). An example of this can be seen in Figure 2, which shows arginine side chains forming Yaa−Yaa hydrogen bonds to carbonyl groups on the subsequent chains. Changes in Yaa residues have been shown to have a significant impact on the stability of model collagen-like peptides, and interchain hydrogen bonding is used to explain some of this variation (*20*−*22*).

Variations in residue sequence lead to variations in structure. However, the variations in the side chain behavior and changes in backbone structure are generally small (*23*), even when it is bound to other molecules. One of the best characterized changes in backbone structure is seen in the crystal structure of a collagen-like peptide determined by Kramer *et al*. (*24*), which shows the helical pitch in imino-poor regions is slightly smaller than in imino-rich regions [see also Rainey *et al*. (*25*)].

Structural Variations. This study was motivated by the question how OI severity is influenced by residues near the mutation sites. To determine how changes in residues and mutation severity correlate with each other, we calculated the observed frequency of lethal and nonlethal mutations versus residue type at positions near each mutation site [mutations from the Database of Human Collagen Mutations (*26*, *27*) at www.le.ac.uk/genetics/collagen]. The results are shown in Table 1, where the statistical significance is the probability that the observed distribution of lethal versus nonlethal mutations would occur by chance alone assuming a binomial distribution. We see the same general trends

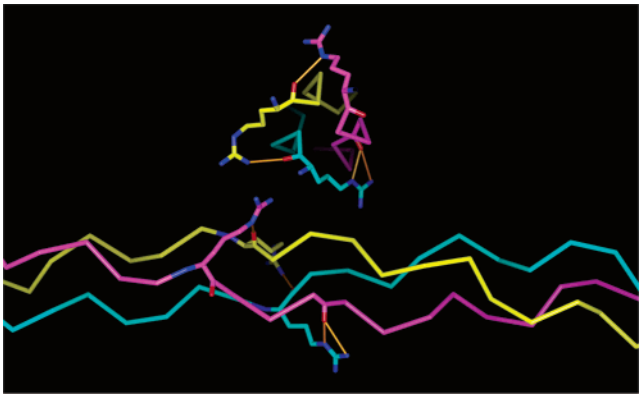

FIGURE 2: Arginine residues and their side chain−backbone hydrogen bonds in the collagen-like peptide. Chain A is colored magenta, chain B yellow, and chain C cyan. Arginine nitrogen atoms are colored blue, and Yaa residue oxygen atoms are colored red. The following three sets of hydrogen bonds are shown in orange: the bond between the ϵ-nitrogen of the arginine on chain A and the backbone carbonyl group of the arginine on chain B, the bond between a terminal nitrogen of the arginine on chain B and the carbonyl group of the arginine on chain C, and the bond between the ϵ-nitrogen and a terminal nitrogen of the arginine on chain C and the backbone carbonyl group of the Yaa residue following the arginine on chain A. For clarity, only the arginine residues, α-carbons of subsequent residues, and the carbonyl group of the Yaa residue following the arginine on chain A are shown. This structure is Protein Data Bank entry 1BKV.

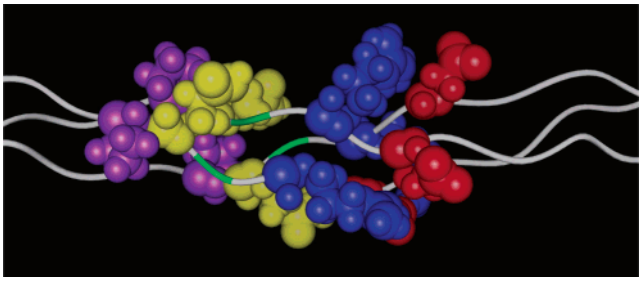

FIGURE 3: Residues near the mutation site. The three backbone chains (white), mutation site glycine (green), and prolines two residues from the mutation site, in the N-terminal direction (purple), are displayed. The glutamines (yellow) between the N-terminus and the mutation site and arginine (blue) and aspartic acid (red) residues two and four positions from the mutation site, in the C-terminal direction, are also displayed.

observed previously [see also Hunter and Klein (*28*)]. For example, the frequency of lethal OI is always greater than 50% when a proline is nearby, and 67% (26 of 39 known cases) if there is a proline two positions from the mutation site, in the N-terminal direction (see Figure 3). It is interesting to contrast these statistics with the cases of nearby hydroxyprolines, which despite their apparent similarity to prolines do not seem to correlate with lethal or nonlethal mutations, and, with alanines, which nearly always correlate with nonlethal mutations. As with proline, the presence of an alanine two residues from the mutation site, in the N-terminal direction, appears to have a particularly pronounced effect, but with the current data set, lethal OI is seen only 19% of the time.

To determine if underlying structural effects may cause these correlations, molecular dynamics simulations were carried out on two sets of related peptides (see Table 2). Preliminary studies showed that the two structures are relatively constant and changes are subtle. The first set of

four peptides [T1−904 (**I**), T1−904[G11S] (**II**), T1−904-[G11S/P9A] (**III**), and T1−904[pH 2] (**IV**)] models the site of the G913S lethal mutation in the α1 chain of collage type I, using the sequence corresponding to type I collagen residues 904−921. Peptides **I**, **II**, and **IV** have been studied by Yang *et al*. (*5*), yielding experimental stability and structure data.

The G913S mutation site has an arginine two positions from the mutation site, in the C-terminal direction, that correlates with 100% lethal mutations (from Table 1, six of six known mutations). It is also notable that the G913S site has an aspartic acid four positions from the mutation site, in the C-terminal direction, that is lethal for 86% of known mutations (six of seven mutations). There is also a glutamine in the position immediately preceding the mutation site that also correlates with 100% lethality (three of three mutations). However, the small number of observed mutations makes the statistical significance of this correlation with lethality suspect (∼15%). The diversity of side chain physiochemical properties surrounding the G913S mutation provides us with an opportunity to answer multiple questions with a single simulation.

*Native Peptide (**I**).* The T1−904 peptide (**I**) models the native collagen structure. It remains helical for the majority of its length (residues 5−29), for the entire 20 ns simulation. The average Φ and Ψ angles from this simulation are shown in Figure 4 (blue lines). Note that the angles are relatively constant, especially in the imino-rich C-terminal region. These results are consistent with the study by Yang *et al*. (*5*) which reports that this peptide has a high helical content on the basis of CD experiments. NMR structural data do not exist for this peptide (B. Brodsky, unpublished results from 2003).

*Single-Mutation Peptide (**II**).* The T1−904[G11S] peptide (**II**) models the G913S mutant collagen by replacing the glycines corresponding to Gly-913 in native collagen type I with a serine (residue 11 in this simulation). The major structural changes observed during the MD simulation occur near the mutation site. They include the disruption of the triple helix (see Table 3), the formation of hydrogen bonds between the hydroxyl groups of various pairs of serines (see Figure 5), and the addition of extensive molecular interactions between serines and glutamines on adjacent chains. Figure 4 shows the Φ and Ψ angles from this simulation (cyan lines). Note that these are similar to the native simulation for most residues, with the primary exception being the region C-terminal of the mutation site, particularly residues 13 and 14 on chain A. The imino acid-rich C-terminal region of the peptide remains helical and stable for the entire simulation. There is a clear reduction in the helical content of the N-terminal region of the peptide (see Table 3), consistent with observed CD data (*5*).[2]

*Double-Mutation Peptide (**III**).* The T1−904[G11S/P9A] peptide (**III**) has a double mutation, and was simulated for comparison of the structural effects of alanine versus proline residues located two positions from the glycine to serine

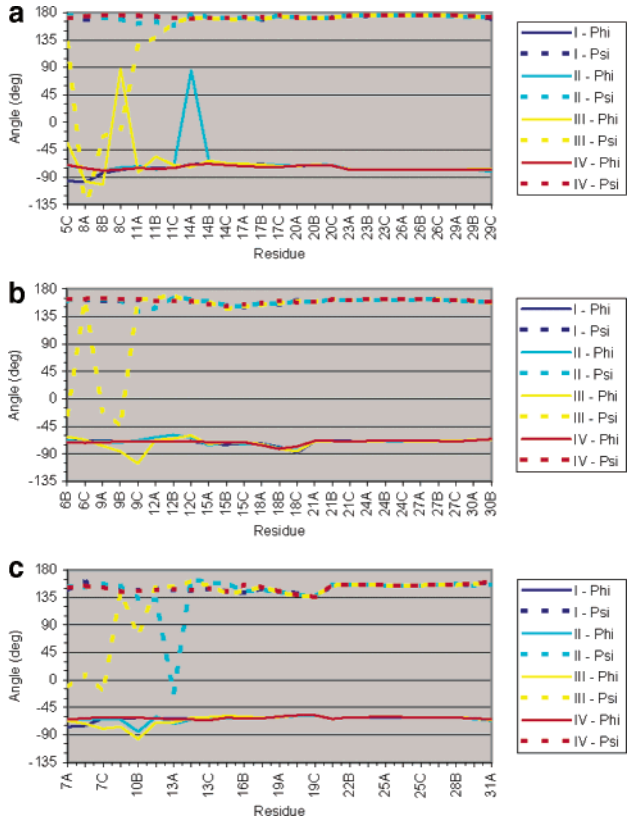

FIGURE 4: Average Φ and Ψ angles for each glycine residue, Xaa residue, and Yaa residue. The average Φ and Ψ angles are displayed. Solid lines represent Φ angles by residue number, and dotted lines represent Ψ angles. (a) Average Φ and Ψ angles for each glycine residue, (b) average Φ and Ψ angles for each Xaa residue, and (c) the average Φ and Ψ angles for each Yaa residue in the native peptide (**I**) (blue), the single mutant (**II**) (cyan), the double mutant (**III**) (yellow), and the native peptide at pH 2 (**IV**) (red) are shown.

mutation, in the N-terminal direction. The molecular structure during the MD simulation is similar to the structure of the single-mutant peptide (**II**). The two primary differences occur on opposite sides of the mutation site. The first difference is in the N-terminal region of the peptide, which is destabilized relative to **II**. This can be seen in Table 4 and in the Φ and Ψ angles plotted in Figure 4 (yellow lines). The disruption is most significant in the region of the proline to alanine mutation site at residue 9. The second difference occurs on the C-terminal side of the mutation site, which is significantly stabilized relative to **II**. In fact, the structure of this region of the peptide more closely resembles the native peptide (**I**) than it does **II**.

*pH 2 Native Structure (**IV**).* The T1−904[pH 2] peptide (**IV**) models the native collagen structure at pH 2. Yang *et al*. tested this collagen-like peptide at pH 2 to determine how the removal of side chain−side chain ionic interactions affected the stability of the peptide. **IV** has the same sequence as **I** except that all acidic residues are in their neutral form (Asp14 and Glu17 carboxylate groups are protonated). The backbone structure is similar to that of the native peptide, but the number of observed hydrogen bonding interactions between the acidic side chains and the basic side chains is greatly reduced relative to the native structure (**I**). This is consistent with experimental data (*5*) since this peptide exhibits a reduced thermal stability at this pH. The Φ and Ψ angles shown in Figure 4 (red lines) are very similar to

[2] We recently learned of unpublished NMR studies of peptide **II** which indicated that the peptide formed a triple helix between the C-terminus and the mutation site, but no triple helix was observed between the N-terminus and the mutation site (Q. H. Dai, J. Baum, and B. Brodsky, unpublished results from 2004).

Table 3: Backbone−Backbone (Gly−Xaa) Hydrogen Bond Statistics for T1−904 Simulations

H-bond pair		H-bond frequency (%)[a]				correlation with the subsequent H-bond			
donor	acceptor	I	II	III	IV	I	II	III	IV
Gly5A H	Ala3B O	8	2	4	1	–	–	–	–
Gly5B H	Ala3C O	24	22	19	69	0.02	0.12	0.32	0.06
Gly5C H	Pro6A O	51	39	24	52	0.04	−0.07	0.18	−0.08
Gly8A H	Pro6B O	33	37	18	38	−0.09	−0.09	0.13	−0.10
Gly8B H	Pro6C O	30	35	17	31	−0.09	−0.08	0.00	−0.10
Gly8C H	Pro/Ala9A O	35	17	3	32	−0.07	−0.09	0.00	−0.06
Gly/Ser11A H	Pro/Ala9B O	45	32	39	41	−0.04	0.00	0.00	−0.07
Gly/Ser11B H	Pro/Ala9C O	41	4	0	35	−0.08	0.00	0.00	−0.07
Gly/Ser11C H	Pro12A O	37	23	37	38	−0.10	0.00	0.00	−0.08
Gly14A H	Pro12B O	51	2	2	52	−0.07	0.00	0.00	−0.10
Gly14B H	Pro12C O	67	29	43	58	−0.11	−0.13	−0.16	−0.17
Gly14C H	Asp15A O	66	44	61	39	0.05	−0.14	0.04	0.00
Gly17A H	Asp15B O	68	57	74	60	0.04	0.03	0.04	0.00
Gly17B H	Asp15C O	76	69	76	57	0.02	−0.02	−0.01	0.00
Gly17C H	Glu18A O	69	68	67	53	0.01	−0.01	0.02	0.00
Gly20A H	Glu18B O	62	58	63	48	−0.01	−0.04	−0.03	−0.02
Gly20B H	Glu18C O	61	60	64	54	0.02	−0.01	−0.05	0.02
Gly20C H	Pro21A O	46	48	46	48	−0.14	−0.12	−0.13	−0.11
Gly23A H	Pro21B O	30	29	31	27	−0.10	−0.10	−0.13	−0.12
Gly23B H	Pro21C O	29	27	27	27	−0.10	−0.14	−0.10	−0.10
Gly 23C H	Pro24A O	32	32	30	31	−0.13	−0.12	−0.13	−0.13
Gly26A H	Pro24B O	28	29	29	29	−0.11	−0.11	−0.12	−0.12
Gly26B H	Pro24C O	29	29	29	30	−0.14	−0.14	−0.15	−0.14
Gly26C H	Pro27A O	32	32	30	30	−0.12	−0.13	−0.14	−0.13
Gly29A H	Pro27B O	28	28	29	29	−0.16	−0.15	−0.17	−0.16
Gly29B H	Pro27C O	33	34	33	34	−0.15	−0.17	−0.14	−0.16
Gly29C H	Pro30A O	35	33	35	31	−0.10	−0.08	−0.09	−0.04
Gly32A H	Pro30B O	26	32	33	33	0.06	0.00	−0.06	−0.06
Gly32B H	Pro30C O	39	40	32	36	–	–	–	–
average		42	34	34	39	−0.06	−0.07	−0.03	−0.08

[a] Frequency of hydrogen bonds observed between backbone nitrogen atoms and backbone carbonyls.

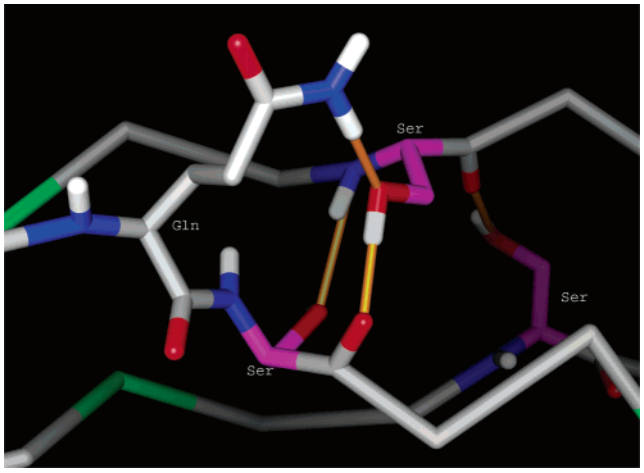

FIGURE 5: Typical serine mutant interactions. Mutant serine residues (magenta), carbon atoms (gray), hydrogen atoms (white), nitrogen atoms (blue), oxygen atoms (red), glycine α-carbon atoms (green), and hydrogen bonds (orange bars) are displayed. Details of the interaction between a glutamine and the backbone carbonyl of an adjacent mutant serine on one chain and the side chain of a mutant serine on another chain are very frequently observed during both mutant simulations (**II** and **III**). This image is taken from the single mutant simulation (**II**).

those of the native structure. The only significant variations are the Φ angle of the neutral glutamic acid at position 18C and the Ψ angle of the lysine at position 16B. As described in detail below, these two residues are observed to interact differently at different pHs.

Interchain Hydrogen Bonding Effects. Interchain hydrogen bonding statistics for each molecular simulation were calculated and are summarized in Tables 3−6. Hydrogen bonds are defined as interactions between hydrogen bond donor and acceptor groups that are within 2.2 Å of each other. Donor groups are assumed to bond with only the nearest acceptor group, setting a maximum of one hydrogen bond for each donor. No limit is placed on acceptor groups; thus, they can participate in any number of hydrogen bonds.

Table 3 shows frequencies of Gly−Xaa hydrogen bonds and the time correlation coefficient between each pair of adjacent hydrogen bonds. In general, the four peptides have

Table 4: Side Chain−Backbone (Yaa−Yaa) Hydrogen Bond Frequencies for T1−904 Simulations

H-bond pair		H-bond frequency (%)[a]			
donor	acceptor	I	II	III	IV
Gln10A N^{ϵ}	Gln10B O	23	0	3	28
Gln10B N^{ϵ}	Gln10C O	28	3	0	14
Gln10C N^{ϵ}	Arg13A O	6	3	20	9
Arg13A N^{ϵ}	Arg13B O	49	47	60	52
Arg13A $N^{\eta 1}$	Arg13B O	8	59	46	9
Arg13A $N^{\eta 2}$	Arg13B O	12	0	0	8
Arg13B N^{ϵ}	Arg13C O	57	64	65	56
Arg13B $N^{\eta 1}$	Arg13C O	9	61	35	11
Arg13B $N^{\eta 2}$	Arg13C O	2	0	3	5
Arg13C N^{ϵ}	Lys16A O	39	39	46	54
Arg13C $N^{\eta 1}$	Lys16A O	21	34	34	38
Arg13C $N^{\eta 2}$	Lys16A O	2	4	2	0
Lys16A N^{ζ}	Lys16B O	41	7	43	5
Lys16B N^{ζ}	Lys16C O	35	27	35	9
Lys16C N^{ζ}	Thr19A O	11	13	4	5
average		23	24	26	20

[a] Frequency of hydrogen bonds observed between side chain nitrogens (including amides and guanadino groups) and backbone carbonyls.

Table 5: Hydrogen Bond Correlation Coefficients for T1−904 Simulations

H-bond pair 1		H-bond pair 2		correlation coefficient[a]			
donor	acceptor	donor	acceptor	I	II	III	IV
Arg13A N$^{\epsilon}$	Arg13B O	Gly14A H	Pro12B O	0.06	0.02	−0.01	0.03
Arg13A N$^{\eta}$	Arg13B O	Gly14A H	Pro12B O	−0.17	0.00	0.00	−0.21
Arg13B N$^{\epsilon}$	Arg13C O	Gly14B H	Pro12C O	0.01	−0.06	0.09	0.07
Arg13B N$^{\eta}$	Arg13C O	Gly14B H	Pro12C O	−0.26	−0.02	−0.06	−0.32
Arg13C N$^{\epsilon}$	Lys16A O	Gly14C H	Asp15A O	0.00	−0.09	−0.01	−0.03
Arg13C N$^{\eta}$	Lys16A O	Gly14C H	Asp15A O	−0.25	0.14	−0.23	−0.29
average				−0.10	0.00	−0.04	−0.13

[a] Calculated correlation coefficients for selected hydrogen bond pairs from the T1−904 simulations.

Table 6: Side Chain−Backbone (Yaa−Yaa) Hydrogen Bond Frequencies for P1 Simulations

	Yaa−Yaa H-bond frequency (%)[a]				
peptide	chain A	chain B	chain C	average	T_m (°C)
P1-GPO[O13R][b]	74	77	81	77	47.2
P1-GPO[O13Q][b]	24	26	29	27	41.3
P1-GPO[O13K][b]	4	9	10	8	36.8
P1-GPO[O13N][b]	8	12	0	7	30.3
average	27	31	30	30	

[a] Frequency of hydrogen bonds observed between side chain nitrogens (including amides and guanadino groups) and backbone carbonyls for the P1-GPO simulations. [b] From ref *22*.

similar hydrogen bonding statistics over most of their length. Differences between the model peptides are seen in the region near the acidic residues in the low-pH simulations (**IV**, residues 15−18), and the region near the mutation sites in the single mutant simulation (**II**, residues 11−17) and the double mutant simulation (**III**, residues 8−14). This is consistent with the difference in the Φ and Ψ angles shown in Figure 4.

For all peptides, two interesting structural differences can be seen. First, Gly−Xaa hydrogen bonds in the imino-poor region (residues 14−20) are twice as frequent as in the peptide's imino-rich region (residues 23−32). Second, there is a small but consistent negative correlation between adjacent hydrogen bonds in imino-rich regions, while there is no such correlation in imino-poor regions.

For each peptide in Table 3, the Gly−Xaa hydrogen bond networks in imino-rich regions are weaker than we expected, since our simulations showed hydrogen bonds in the imino-rich regions forming approximately half as frequently as in imino-poor regions. This reduced frequency is consistent with the study by Rainey and Goh (*25*) which compared crystal structures of imino-rich and -poor regions and estimated average interchain hydrogen bonding energies of −1.9 to −2.0 kcal/mol and −2.8 to −2.9 kcal/mol, respectively. This is also consistent with the host−guest studies done by Persikov *et al*. (*22*) which reported that the enthalpic contribution to association of the imino acid-rich peptides (Gly-Pro-Hyp)$_n$ is among the smallest of 39 peptides examined despite the fact that this peptide has the highest melting temperature.

Time correlation coefficients between pairs of hydrogen bonds shown in Table 3 provide two structural insights. First, the negative correlations between adjacent hydrogen bonds in imino-rich regions indicate that hydrogen bonds are more likely to form when adjacent hydrogen bonds do not exist, suggesting that the local structure is not ideal for the formation of adjacent hydrogen bonds. A reasonable explanation for this is that the rigidity of these residues prevents the formation of an optimal hydrogen bonding network. Second, imino-poor regions are more flexible and have been shown to adopt a different helical pitch than imino-rich regions (*24*), permitting better hydrogen bond interactions.

For the single-mutant peptide (**II**), Table 3 shows that the N-terminal region is destabilized relative to the nonmutant peptides, but the helical structure and hydrogen bonding network seem to be essentially normal from the C-terminal end to approximately nine residues from the mutation site in the C-terminal direction. A closer examination of the hydrogen bonding behavior of the serine residues shows that they frequently form hydrogen bonds with backbone carbonyls on adjacent chains. Serines also interact with backbone amines, and with the side chains of nearby glutamines (see Figure 5), which no longer form the Yaa−Yaa hydrogen bonds seen in the other simulations (shown in Table 4). The result is a rather complex hydrogen bonding network that appears to be relatively rigid with little in common with a native triple helix. Inspection of the Φ and Ψ angles plotted in Figure 4 shows specific disruptions in the region of the mutation site.

For the double mutant simulations (**III**), Table 3 shows hydrogen bonding patterns similar to those seen in the single mutant simulations (**II**). Changing the proline to an alanine two residues from the mutation site, in the N-terminal direction, permits the already destabilized N-terminal end to become more disordered and increases the strength of the hydrogen bonding network between the C-terminus and the mutation site. The structure of the C-terminal end of the double-mutant peptide is more native-like than that of the single-mutant peptide. This can also be seen in the Φ and Ψ angles plotted in Figure 4.

The correlation between lethality and nearby residue type suggests the details of side chain behavior are important. Table 4 gives Yaa−Yaa hydrogen bond frequencies for the central glutamine, arginine, and lysine residues. Arginine is the most interesting residue for three reasons. First, all known mutations with an arginine two positions from the mutation site, in the C-terminal direction, are lethal (see Table 1). Second, arginines in the Yaa position have been show to significantly stabilize collagen model structures in host−guest studies done by Persikov *et al*. (*22*). Third, arginines in the Yaa position have been shown to be part of the dominant binding sites for molecular chaperones such as HSP47 (*29*).

In the simulations reported here, the arginine side chains interact strongly with a Yaa residue on the subsequent chain. For the native simulation (**I**), the hydrogen bond is formed primarily using the arginine's ϵ-nitrogen, although arginine's terminal nitrogens can also form bonds (Table 4). The most

frequently seen arginine structure in this simulation has a fully extended side chain with its aliphatic region fitting in a hydrophobic region of the peptide and its ϵ-nitrogen forming a Yaa−Yaa hydrogen bond with a backbone carbonyl on the subsequent chain.

The presence of the arginine Yaa−Yaa hydrogen bond varies significantly among the four peptides. In almost every case, this interaction is seen more frequently in the mutant peptides (**II** and **III**) and is most frequent in the double-mutant peptide (**III**). In contrast, the Gly−Xaa hydrogen bonds in the same region are substantially weakened relative to those in the native (**I**) simulation (see Gly14 in Table 3).

Less frequently, the arginine residues in the native simulation form Yaa−Yaa hydrogen bonds using their terminal nitrogens (N^{η}). These structures can be characterized by examining time correlation data for these hydrogen bonds and the Gly−Xaa hydrogen bonds connecting the same regions (see Table 5). In each case, the negative correlation indicates that Gly−Xaa hydrogen bonds are seen less frequently when the terminal nitrogen hydrogen bond exists. This means that the terminal nitrogen hydrogen bonds are more frequent when the Gly−Xaa bonds are broken and the peptide chains are partially separated. Since this hydrogen bond presumably helps prevent further chain separation, the ability of arginine to form these "back-up" bonds using their terminal nitrogens is probably another factor explaining the large stabilization seen for arginine (*22*). Examination of Table 5 shows that this is also true for the low-pH simulation (**IV**), but not true for the single mutant (**II**) simulation. The double mutant (**III**) simulation falls between the two extremes with residues Arg13A and Arg13B behaving in a mutant-like manner and Arg13C behaving in a more native-like manner.

The mutant collagen-like peptide simulations each show a significant increase in the frequency of hydrogen bonds involving one of the terminal nitrogens in each arginine (Table 4). This suggests that the triple helix is a bit more open, which is consistent with the serine pushing the chains apart, yet the arginines are still able to form strong bonds using their terminal nitrogens. The observed large increase in the hydrogen bond frequency can be explained by an observed increase in the frequency of interactions between the Asp15 residues and the Arg13 residues in the mutant simulations. This interaction seems to hold the arginines in the appropriate position to form the Yaa−Yaa hydrogen bonds, making them more frequent than in the native simulations.

The Lys16A and Lys16B residues also form Yaa−Yaa hydrogen bonds, and these are seen more frequently in the neutral simulation than in the low-pH simulation (Table 4). The structural explanation for this result is that lysine needs to interact with the negatively charged acid residues to form strong Yaa−Yaa hydrogen bonds because lysine's fully extended side chain is too long to form a hydrogen bond to the Yaa carbonyl on the subsequent chain. Thus, the lysine must fold back on itself, adopting a less stable conformation to bring its amino group close enough to form a hydrogen bond. When this happens, the charged acidic residues are ideally positioned to form strong hydrogen bonds to the lysine amino group and hold it in a conformation that stabilizes the Yaa−Yaa hydrogen bonds, increasing its observed frequency.

In the low-pH simulations, the acidic residues are in their neutral form such that the lysines are far less likely to interact with them. Rather, the lysine residues adopt an extended conformation and interact primarily with the solvent. Examination of the Gly−Xaa hydrogen bond frequencies in this region (Table 3) shows that they are less frequent than in the native simulation, confirming a weaker interaction. Lys16C does not form stable Yaa−Yaa hydrogen bonds in any of the simulations since the side chain is too distant from the acidic residues to interact with and to be stabilized by them.

Lysine residues in the double mutant (**III**) simulation exhibit behavior similar to that seen in the native peptide simulation, where side chain−side chain hydrogen bonds with acidic residues stabilize the lysine Yaa−Yaa hydrogen bonds. However, in the single mutant simulation, they stabilize the arginine Yaa−Yaa hydrogen bonds, but not the lysine Yaa−Yaa hydrogen bonds. The underlying reasons for this difference are not clear; however, it is likely that this affects the peptides' stability, and it clearly affects the structure and dynamics of the exposed surface of the peptides.

The distorted lysine structure seen in the native simulation suggested that a shorter side chain might result in a stronger Yaa−Yaa hydrogen bond. To test this hypothesis, we performed a simulation with each lysine replaced with an ornithine residue (ornithine is similar to lysine, but with one fewer methylene in its side chain). However, the simulation did not support this hypothesis (data not shown). The Yaa−Yaa hydrogen bonding frequency was very low, and inspection of the simulation structures reveals that the ornithine's N^{ξ} often seems to be in good position to form a hydrogen bond, but cannot because its hydrogens straddle and do not make a direct interaction with the appropriate carbonyl oxygen. Surprisingly, the ornithine side chain is unable to adjust adequately to make a good hydrogen bond, presumably because it is already fully extended and is constrained on either side by nearby groups. This is not the case for glutamine and arginine because unlike those of ornithine, their nitrogens are sp^2 hybridized, allowing their hydrogens to point directly at the carbonyl oxygen.

Residues in the Xaa position are more solvent accessible than residues in the Yaa position and do not seem to interact with the backbone of adjacent chains. This suggests that these residues do not affect the molecular stability as much as residues in the Yaa positions [supported by experimental thermal stability data (*20*)]. They do, however, interact with the side chains of residues in the more critical Yaa position, affecting their interchain interactions as described above for arginine and lysine. These residues also appear to be ideally positioned to interact with other molecules, so they may play a critical role in intermolecular recognition and binding.

Throughout this study, we assumed that an increased Yaa−Yaa hydrogen bond frequency implies an increase in molecular stability. To test this assumption, we performed five additional simulations of proline-rich peptides that were taken from Persikov *et al*. (*22*) and are able to form Yaa−Yaa hydrogen bonds. Table 6 shows the hydrogen bond frequency and the melting temperature for each. Notably, the rank orderings of hydrogen bond frequency and melting temperature are the same, supporting our hypothesis that the frequency of Yaa−Yaa bonds does affect molecular stability.

CONCLUDING REMARKS

To date, the question of whether local molecular structure helps to determine the observed severity of various OI mutations remains unanswered. Glycine mutations disrupt the formation of the triple-helical structure normally seen in collagen molecules. One factor determining the resulting OI severity is the ability of the local structure to adapt to the disruption during collagen folding and re-form the normal triple-helical structure on the N-terminal side of the mutation site. This suggests that flexible residues such as alanine on the N-terminal side of the mutation site would be most likely to permit re-formation of the triple helix and would correlate with nonlethal OI. This argument also suggests that proline, the least flexible of the naturally occurring residues, would inhibit re-formation of the triple helix and correlate with lethal mutations.

The strong alanine and proline correlations (Table 1) two residues from the mutation site, in the N-terminal direction, deserve special comments. These correlations may simply be a result of this position's proximity to the mutation site; however, it is interesting to note that the residues in this position on the B and C chains form Gly−Xaa hydrogen bonds with the mutant residues on chains A and B, respectively. On the basis of this, it seemed reasonable to assume that a flexible residue in this position might be better able to adapt to the distorted structure and form stronger hydrogen bonds with a mutant residue, stabilizing the resulting peptide. We tested this idea on two mutant simulations (**II** and **III**), since the critical site in the single-mutant peptide (**II**) is occupied by a proline and is occupied by an alanine in the double-mutant peptide (**III**).

Interestingly, the observed hydrogen bond frequencies between the mutant residues and the residues two positions from the mutation site, in the N-terminal direction, are essentially the same in both **II** and **III** peptide simulations (see Table 3), suggesting that the argument presented above is not complete. Comparison of the N-terminal regions of both peptides (see Table 3 and Figure 4) also seems to be inconsistent with the idea that alanine residues permit the re-formation of the triple helix since the alanine-containing double-mutant peptide exhibits less helical structure than the proline-containing single-mutant peptide.

Inspection of the hydrogen bonding statistics on the C-terminal side of the mutation site suggests a more promising explanation for the differences in lethality. For all the hydrogen bonding statistics that were collected, the structure of the alanine-containing peptide (**III**) becomes native-like closer to the mutation site than does that of the single-mutant peptide (**II**). For **III**, the hydrogen bonding networks (Gly−Xaa and Yaa−Yaa) seem to be disrupted up to residue Arg13C or Gly14C, while the hydrogen bonding networks for **II** are disrupted up to Gly17C (see Tables 3−5). The difference in the details of side chain behavior for all four peptides (**I**−**IV**) is particularly dramatic.

Interestingly, there is little difference in the hydrogen bonding structure at the mutation site, but there are substantial differences more distant from the mutation site. The alanine-containing peptide is more native-like on the C-terminal side of the mutation, arguing that this peptide may model a less lethal, mutant form of collagen. This difference could affect the molecule's stability or its ability to interact with other molecules, affecting its ability to interact with other molecules required for fibril formation.

Despite the apparent homogeneity of the triple-helical structure (in terms of hydrogen bonding partners, helicity, and overall structure), side chain dynamics and the frequency of the hydrogen bonds are very sensitive to residue sequence and/or location. Accurate predictions of the triple-helical structures will need to include these effects.

A better understanding of the details of collagen structure, dynamics, and hydrogen bond networks will improve our ability to predict the physicochemical properties that contribute to the stability of collagen molecules, or lack thereof, and the severity of a single-point mutation.

ACKNOWLEDGMENT

We thank Drs. Russ Altman of Stanford University and Sean Mooney of Indiana University (Indianapolis, IN) for helpful discussions and comments on the manuscript and the Biocomputation Core Facility at Stanford University for the use of the SGI 3800 Origin computer.

REFERENCES

1. Alberts, B. (2002) *Molecular biology of the cell*, 4th ed., Garland Science, New York.
2. Kielty, C. M., and Grant, M. E. (2002) in *Connective Tissue and Its Heritable Disorders* (Royce, P. M., and Steinmann, B., Eds.) pp 159−221, Wiley-Liss, New York.
3. Kreis, T., and Vale, R. (1999) *Guidebook to the extracellular matrix, anchor, and adhesion proteins*, 2nd ed., Oxford University Press, Oxford, U.K.
4. Byers, P. H. (2001) Folding defects in fibrillar collagens, *Philos. Trans. R. Soc. London, Ser. B 356*, 151−157.
5. Yang, W., Battineni, M. L., and Brodsky, B. (1997) Amino acid sequence environment modulates the disruption by osteogenesis imperfecta glycine substitutions in collagen-like peptides, *Biochemistry 36*, 6930−6935.
6. Liu, X., Kim, S., Dai, Q. H., Brodsky, B., and Baum, J. (1998) Nuclear magnetic resonance shows asymmetric loss of triple helix in peptides modeling a collagen mutation in brittle bone disease, *Biochemistry 37*, 15528−15533.
7. 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.
8. 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.
9. Mooney, S. D., and Klein, T. E. (2002) The functional importance of disease-associated mutation, *BMC Bioinf. 3*, 24.
10. Mooney, S. D., Kollman, P. A., and Klein, T. E. (2002) Conformational preferences of substituted prolines in the collagen triple helix, *Biopolymers 64*, 63−71.
11. Vitagliano, L., Nemethy, G., Zagari, A., and Scheraga, H. A. (1995) Structure of the type I collagen molecule based on conformational energy computations: the triple-stranded helix and the N-terminal telopeptide, *J. Mol. Biol. 247*, 69−80.
12. Vitagliano, L., Nemethy, G., Zagari, A., and Scheraga, H. A. (1993) Stabilization of the triple-helical structure of natural collagen by side-chain interactions, *Biochemistry 32*, 7354−7359.
13. Stultz, C. M. (2002) Localized unfolding of collagen explains collagenase cleavage near imino-poor sites, *J. Mol. Biol. 319*, 997−1003.
14. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E. I., Wang, J., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (2002) *AMBER*, version 7, University of California, San Francisco.
15. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A Second Generation Force Field

for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *J. Am. Chem. Soc. 117*, 5179−5197.
16. Onufriev, A., Case, D. A., and Bashford, D. (2002) Effective Born radii in the generalized Born approximation: the importance of being perfect, *J. Comput. Chem. 23*, 1297−1304.
17. Onufriev, A., Bashford, D., and Case, D. A. (2000) Modification of the Generalized Born Model Suitable for Macromolecules, *J. Phys. Chem. B 104*, 3712−3720.
18. Huang, C. C., Couch, G. S., Pettersen, E. F., Ferrin, T. E., Howard, A. E., and Klein, T. E. (1998) The object technology framework: an object-oriented interface to molecular data and its application to collagen, *Pac. Symp. Biocomput. '98*, 349−361.
19. Jenkins, C. L., and Raines, R. T. (2002) Insights on the conformational stability of collagen, *Nat. Prod. Rep. 19*, 49−59.
20. Persikov, A. V., Ramshaw, J. A., Kirkpatrick, A., and Brodsky, B. (2002) Peptide investigations of pairwise interactions in the collagen triple-helix, *J. Mol. Biol. 316*, 385−394.
21. Persikov, A. V., Ramshaw, J. A., and Brodsky, B. (2000) Collagen model peptides: Sequence dependence of triple-helix stability, *Biopolymers 55*, 436−450.
22. Persikov, A. V., Ramshaw, J. A., Kirkpatrick, A., and Brodsky, B. (2000) Amino acid propensities for the collagen triple-helix, *Biochemistry 39*, 14960−14967.
23. Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J., and Liddington, R. C. (2000) Structural basis of collagen recognition by integrin $\alpha_2\beta_1$, *Cell 101*, 47−56.
24. 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.
25. Rainey, J. K., and Goh, M. C. (2002) A statistically derived parametrization for the collagen triple-helix, *Protein Sci. 11*, 2748−2754.
26. Dalgleish, R. (1998) The Human Collagen Mutation Database 1998, *Nucleic Acids Res. 26*, 253−255.
27. Dalgleish, R. (1997) The human type I collagen mutation database, *Nucleic Acids Res. 25*, 181−187.
28. Hunter, L., and Klein, T. (1993) in *The Proceedings of the 1993 Conference on Intelligent Systems and Molecular Biology*, pp 190−197, AAAI Press: Menlo Park, CA.
29. Koide, T., Takahara, Y., Asada, S., and Nagata, K. (2002) Xaa-Arg-Gly triplets in the collagen triple helix are dominant binding sites for the molecular chaperone HSP47, *J. Biol. Chem. 277*, 6178−6182.

BI035676W