The Severity of Osteogenesis Imperfecta: A Comparison to the Relative Free Energy Differences of Collagen Model Peptides

Kyung-Hoon Lee,[1] Krzysztof Kuczera,[2] Mark M. Banaszak Holl[1]

[1] *Michigan Nanotechnology Institute in Medicine and Biological Sciences, and Department of Chemistry, University of Michigan, Ann Arbor, MI 48109*

[2] *Department of Chemistry and Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66045*

Received 7 June 2010; revised 13 September 2010; accepted 4 October 2010
Published online 13 October 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.21552

ABSTRACT:

Molecular dynamics simulations were carried out to calculate free energy differences between the folded and unfolded states of wild type and mutant collagen model peptides. The calculated stability of the collagen models was compared with the severity of osteogenesis imperfecta. Free energy differences of Gly → Xaa (Xaa: Ser, Cys, Glu, and Asp) mutations between the wild type and the mutants at position 15 of the model peptide were 3.8, 4.2, 5.6, and 8.8 kcal/mol, respectively. The corresponding free energy differences of a second Gly mutation at the same position in different chains were, on average, 1.3, 1.5, 2.9, and 5.4 kcal/mol, respectively. Free energy simulations were also performed to estimate the relative stability between an oxidized form and a reduced form of the mutants containing two Cys residues, which indicated that the mutant of the collagen-like peptide containing an intramolecular disulfide bond was more stable than the mutant containing one Cys residue but less stable than the wild type. The calculated free energy differences between an oxidized and a reduced form of the mutants containing two Cys residues are 0.8 and 2.6 kcal/mol for the disulfide bonds between Chains A and B and between Chains A and C, respectively. © *2010 Wiley Periodicals, Inc. Biopolymers 95: 182–193, 2011.*

Keywords: free energy simulation; thermodynamic integration; stochastic boundary molecular dynamics simulation; collagen-like peptide; disulfide bond

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version. You can request a copy of the preprint by emailing the Biopolymers editorial office at biopolymers@wiley.com

INTRODUCTION

Collagen is a major matrix protein of skin, bone, tendon, and other tissues. It is a heterotrimer composed of two α1(I) chains and one α2(I) chain. Each chain has a pitch of 9.5 Å with nearly three residues per turn and a left-handed twist within a right-handed triple helix for a monomeric collagen. The characteristic structural motif of collagen is a triple helix with 338 repeating units of Gly-X-Y in the helical portion of each chain where Gly is positioned at every third residue, which is believed to be essential for close packing of the chains near the central axis.[1–3] Proline is the most commonly found residue in the repeating units. Proline in the Y position is often post-translationally hydroxylated by prolyl hydroxylase, and the 4-hydroxyproline (Hyp) provides the triple helix with greater thermal stability than proline.[4] This high content of Pro and Hyp is necessary to stabilize individual polyproline II-like helices. The most common triplet in collagen, Gly-Pro-Hyp, is approximately 10% of the total sequence.[5]

Osteogenesis imperfecta (OI),[6] a generalized hereditable disorder of connective tissue leading to fragile bones and sus-

Additional Supporting Information may be found in the online version of this article.

Correspondence to: Mark M. Banaszak Holl; e-mail: mbanasza@umich.edu

Contract grant sponsor: DARPA
Contract grant number: F017952-055766

©2010 Wiley Periodicals, Inc.

ceptibility to fracture,[7–9] is mainly caused by mutations of the Gly residue in the triple helix repeat of $(Gly\text{-}X\text{-}Y)_n$ in human Type I collagen, although deletions, insertions, and splicing defects can also lead to the disease. The majority of deletions, insertions, and splicing anomalies that change the triple helical domain of α1(I) or α2(I) chains causes lethal phenotypes. However, the common Gly substitutions in Gly-X-Y triplets of the helix produce heterogeneous disorder with phenotypes that range from prenatally lethal to mild.[7] One-third of the Gly mutations in α1(I) are lethal when the substituting residues are charged or have a branched side chain.[7] Single point mutations of collagen result in the change of Gly to Ala, Cys, Ser, Val, Glu, Arg, and Asp. These mutations occur all over the range of the α1(I) and α2(I) chains and show varied clinical severity. Severity of OI depends on the identity of the mutated amino acids, the position of the mutation sites, the environment of the residues adjacent to the mutation site, and the level of expression of the mutant allele.[10–12] Statistical analysis indicates that Asp is consistently lethal; Glu, Val, and Arg are lethal unless the mutation sites are near the N-terminus; and Ala, Ser, and Cys are lethal or nonlethal all along the chains.[13]

A composite model was designed for clinical lethality caused by Gly mutations in the α1(I) chain of collagen.[14] The lethality of OI was predicted by a computational method based on physicochemical properties of individual residues surrounding the mutation site.[15] Calculation of relative stability and experimental work with model peptides indicated that mutations within a very stable and robust local environment resulted in a more severe clinical effect.[11,16] The relative free energy of unfolding for collagen-like molecules for mutation from Gly to Ala was previously calculated to investigate the energetic and structural effects of a single point mutation of collagen.[17] Their simulation results indicated that each Gly mutation did not contribute equally to the free energy change difference, and the sum of the three calculated mutations at each chain was in agreement with the experimental value.[17] The simulated results demonstrated that interchain hydrogen bonding was significantly affected by variations of mutated amino acids.

Stabilities of the collagen model peptides previously investigated using a host–guest triple-helical peptide demonstrated that the destabilizing effect caused by the substitution for Gly correlated with the severity of OI mutation in the α1 chain of Type I collagen.[18] However, the Gly substitutions in their studies on the homotrimer peptide systems caused the destabilizing effects to be exaggerated owing to three Gly replacements rather than one or two mutations. With the homotrimeric model peptides, the destabilizing effects on heterozygous or homozygous mutants cannot be investigated. Moreover, because of the difficulty of synthesizing heterotrimeric triple helix, only a few heterotrimer collagen model peptides have been synthesized.[19–21] In natural OI mutations, one or two Gly residues are mutated in the α1(I) chain of Type I collagen because collagen is composed of two α1(I) and one α2(I) chains. Free energy simulations can be designed to overcome these difficulties. Furthermore, the total free energy changes can be separated into contributions from each term in potential energy function and each part of the system, enabling us to analyze in detail the destabilizing effect caused by Gly mutations, which cannot be obtained from the experiments.

To understand the relationship between the thermal stability of collagens with the severity of OI we have carried out free energy simulations mutating Gly to Asp as a severe phenotype, to Glu as a moderate one, and to Cys and Ser as mild cases. Free energy simulations for the first and second Gly → Xaa mutations (Xaa: Asp, Glu, Cys, and Ser) were performed to estimate free energy differences between the wild type and heterozygous and homozygous collagen mutants, respectively. The simulation results were used to compare free energy differences between a wild type and its mutants with severity of the OI from statistical analysis. Furthermore, the free energy difference between an oxidized form and a reduced form of the mutant containing two Cys residues has been calculated to investigate the effect of the disulfide bond and to quantify the relative energetic difference of the mutant with a disulfide bond relative to the mutants containing a Cys or two reduced Cys and the wild type. Our goal was to provide insight into the microscopic mechanism of collagen's mutations by analyzing predicted changes in thermal stability between the wild type and mutants containing Asp, Cys, Glu, and Ser, related to OI. These calculated results provide a better understanding of the details of free energy differences of the collagens caused by a single point mutation and suggest a qualitative correlation between collagen thermal stability and severity of the OI. We believe that the simulation results and methods can be used to help predict the severity of unknown mutations in collagen-related diseases.

METHODS

The molecular dynamics (MD) simulations have been carried out using the molecular simulation program CHARMM[22] Version 35 and the CHARMM22 all-atom topology and parameters[23] except the parameters of 4-hydroxyproline (Hyp, O), which were transferred from Ser and Thr amino acids in the CHARMM protein force field. All hydrogen atoms were explicitly included, and an atom-based 12.0 Å nonbonded cutoff distance was used with a switching function between 10.0 and 12.0 Å for van der Waals terms and a shift function at 12.0 Å for electrostatics in all energy evaluations. The Verlet algorithm was used with a 1-fs time step in the MD simulations.

Relative Free Energy Differences Between the Wild Type and the Mutants (Containing Cys, Ser, Asp, and Glu) by Single Point Mutations

We have carried out MD simulations of the mutations from Gly to Asp, Glu, Cys, and Ser at position 15 of a 27-residue collagen model peptide for the folded state and at position 3 of a five-residue peptide for the unfolded state. The central Gly of the peptide was mutated because the middle part has better structural integrity of triple-helix than the termini of the peptide do. Both peptide systems for the folded and unfolded state were acetylated at N-terminus and amidated at C-terminus. The initial coordinate set for the folded collagen system was obtained from a crystal structure (PDB code: 1CGD) from protein data bank.[24] The structure of the system is shown in Figure 1. Hydrogens for the folded systems were added by using the CHARMM program.[22] After a TIP3P[25] water sphere of 16 Å radius centered at the mutated residue Gly 15 C_α was overlaid and solvent overlapping the solute was deleted, the systems contained the part of the proteins (Chain A: $O^{11}G^{12}P^{13}O^{14}G^{15}P^{16}O^{17}G^{18}P^{19}$, Chain B: $P^{13}O^{14}G^{15}P^{16}O^{17}G^{18}P^{19}O^{20}G^{21}$, and Chain C: $G^{12}P^{13}O^{14}G^{15}P^{16}O^{17}G^{18}P^{19}O^{20}$) and 443 TIP3P waters for the simulations of the first Gly mutations and the same part of the protein and on average 436 TIP3P waters for the second Gly mutations. The initial coordinates of the unfolded system were modeled as a monomeric five-residue peptide obtained by truncating the folded structure in an extended conformation. After overlay of a TIP3P water sphere of 16 Å radius centered at the Gly C_α and deletion of solvent overlapping the solute, the unfolded peptide system consisted of five residues and 545 water molecules.

Each system was energy minimized by 1000 steps of the adopted basis Newton–Raphson optimization algorithm.[22] All MD simulations were carried out by using the stochastic boundary MD (SBMD) method.[26] The spherical dynamics reaction region from the energy-minimized structure with a radius of 14 Å centered on the C^α atom of a mutating residue from the energy-minimized structure was used. In this reaction region, all atoms performed Newtonian dynamics under the CHARMM force field. The dynamics region was surrounded by a 2-Å buffer region in which atoms carried out Langevin dynamics with a friction coefficient of 200 ps^{-1}.[27] Water molecules were restrained to stay within 16 Å of the system center by a spherical mean-field potential.[26] The structure of collagen-like peptide and partially solvated collagen-like peptide structure with water molecules for the simulation are shown in Figure 1.

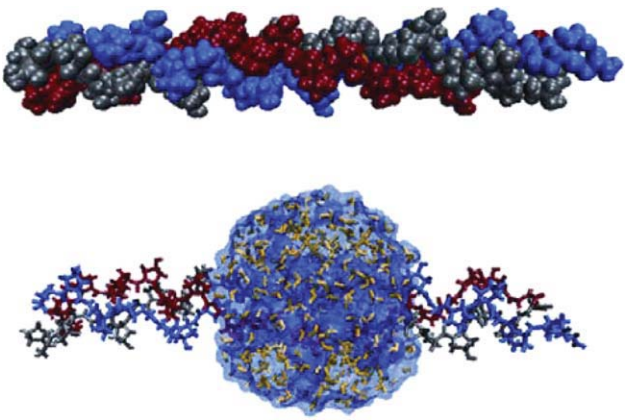

FIGURE 1 The crystal structure of the 27-residue collagen-like triple-helical peptide, $[Ac\text{-}(POG)_9\text{-}NH_2]_3$, for the folded state (PDB code: 1CGD) shown as a space-filling model (top) and the peptide solvated with a water sphere of radius 16 Å centered at C_α atom of Gly 15 shown as a ball and stick model (bottom).[25]

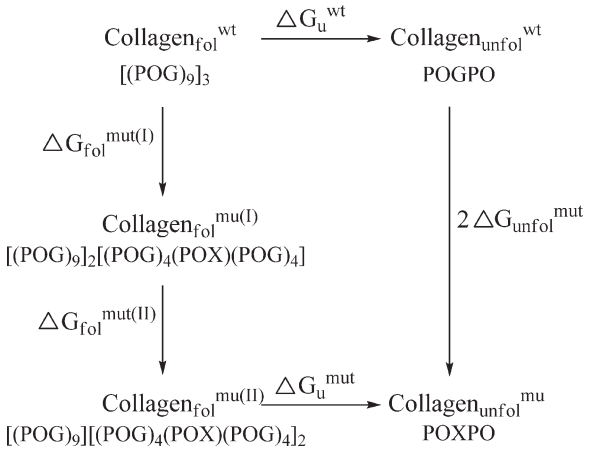

FIGURE 2 Thermodynamic cycle used to calculate free energy differences of the first and second Gly → Xaa mutations at folded and unfolded state. The calculation (ΔG_{unfol}^{mut}) at the unfolded state was determined once and used for each Gly mutation.

$$\begin{aligned}
&\Delta\Delta G_{calc}\ (\text{1st mutation}) = \Delta G_{fol}^{mut(I)} - \Delta G_{unfol}^{mut}\\
&\Delta\Delta G_{calc}\ (\text{2nd mutation}) = \Delta G_{fol}^{mut(II)} - \Delta G_{unfol}^{mut}\\
&\Delta\Delta G_{calc}\ (\text{1st and 2nd mutations}) = \Delta G_{fol}^{mut(I)} + \Delta G_{fol}^{mut(II)} - 2\Delta G_{unfol}^{mut}\\
&\Delta\Delta G_{expt} = \Delta G_u^{wt} - \Delta G_u^{mut}\\
&\Delta(\Delta G) = \Delta G_{fol}^{mut(I)} + \Delta G_{fol}^{mut(II)} - 2\Delta G_{unfol}^{mut} = \Delta G_u^{wt} - \Delta G_u^{mut}
\end{aligned} \tag{1}$$

$$\Delta G = \int_0^1 \langle \Delta U \rangle_\lambda d\lambda \approx \sum_i \langle \Delta U \rangle_{\lambda_i} \Delta_{\lambda_i} \tag{2}$$

$$\Delta U = U_1 - U_0$$

$$\Delta G = \int_0^1 \langle \Delta U_{self} + \Delta U_{int} \rangle_\lambda d\lambda = \Delta G_{self} + \Delta G_{int} \tag{3}$$

The simulation results are related to the experimental measurements in the thermodynamic cycle in Eq. (1) (Figure 2). Because the mutation processes of the collagens at the folded and unfolded state are not chemically balanced, physically meaningful results are obtained by calculating $\Delta(\Delta G)$ values. The thermodynamic windowing method was used for the free energy simulations. The thermodynamic properties were calculated in the coupling parameter approach[27–30] using a hybrid potential energy function given by $U(\lambda) = (1 - \lambda)U_{Gly} + \lambda U_{Xaa} = U_{Gly} + \lambda\Delta U$, where U_{Gly} corresponds to system with a Gly residue, U_{Xaa} corresponds to the same system with a mutated residue (Xaa = Asp, Glu, Cys, and Ser) at

the same position, and $\Delta U = U_{Xaa} - U_{Gly}$. The coupling parameter λ defines the progress of the mutation with U_{Gly} for the initial state ($\lambda = 0$) and U_{Xaa} for the final state ($\lambda = 1$). The simulations were carried out at five values of λ (0.1, 0.3, 0.5, 0.7, and 0.9). After the solvent overlay and deletion of solvents overlapping the solute, solvent only was equilibrated for 100 ps with the peptide model constrained. This was followed by a 200 ps system equilibration and 1 ns MD trajectory generation for each window at $\lambda = 0.1$, 0.3, 0.5, 0.7, and 0.9. Starting at the end of this trajectory, we additionally performed a simulation of the reverse mutation. Based on these two trajectories the free energy differences of the mutations for forward ($\lambda = 0 \rightarrow 1$) and reverse ($\lambda = 1 \rightarrow 0$) direction were calculated, and the average values of the two simulations were reported. In each case, the free energy difference ΔG between the initial ($\lambda = 0$) and final ($\lambda = 1$) states was calculated by the thermodynamic integration method in Eq. (2).[27] A thermodynamic cycle was used to estimate the free energy change because of the first and second Gly → Xaa mutations (Figure 2). Because of linearity of the $U(\lambda)$, the total free energy differences were separated into each contribution from different parts of the system such as the model peptide and the solvent and from the different terms in potential energy function such as covalent or noncovalent (electrostatic or van del Waals) interactions in Eq. (3). The free energy change of the folded-state model for the second Gly mutation was calculated in two different chains because of different chemical environment and geometry of each mutation site. The first mutation was carried out in Chain A and the second mutation both in Chains B and C. Based on the two trajectories generated for each mutation, we estimated the hysteresis, the difference between the magnitudes of free energy change in the forward and reverse directions. The calculations were performed with the computer program CHARMM, which includes a free energy module and the SBMD methodology.[26] The average temperature was 299 K for the forward and reverse simulations at the folded and unfolded states. One nanosecond of SBMD of the systems required about 20 h of central processing unit time on the dual Quad-core Intel Xeon cluster.

Relative Free Energy Difference Between an Oxidized Form and a Reduced Form of Mutants Containing Two Cys Residues

We have also carried out MD simulations to calculate the free energy difference between an oxidized form and a reduced form of the mutants containing two Cys residues. The oxidized form for the mutants contains a disulfide bond between two Cys residues at position 15 of a 27-residue collagen model peptide for the folded state. The initial structure of the unfolded peptide model was composed of two three-residue fragments ($OC_{Oxi}P$) (C_{Oxi}: an oxidized form of a Cys residue) connected by a disulfide bond. The unfolded peptide model was truncated and in an extended conformation. The oxidized form for the unfolded state contains a disulfide bond at position 2 of two three-residue peptides. The oxidized and reduced forms of the mutants are acetylated at N-terminus and amidated at C-terminus. The Cys-S-S-Cys → 2 Cys-SH mutations were introduced in the folded and unfolded systems. After overlay of a TIP3P water sphere of 23 Å radius centered at the Cys 15 C_β and deletion of solvent overlapping the solute, the unfolded peptide system consisted of six residues and 1641 water molecules. For the folded collagen systems, the initial coordinates were obtained from a crystal structure (PDB code: 1CGD).[24] Hydrogen atom positions for the folded systems were built in by using the CHARMM program. After overlay of a TIP3P water sphere of 23 Å radius centered at the mutated residue C_β and deletion of solvent overlapping the solute, the folded system contained 1484 water molecules, with the protein part: Chain A: $O^8G^9P^{10}O^{11}G^{12}P^{13}O^{14}\mathbf{X}^{15}P^{16}O^{17}G^{18}P^{19}O^{20}G^{21}P^{22}$, Chain B: $P^{10}O^{11}G^{12}P^{13}O^{14}\mathbf{X}^{15}P^{16}O^{17}G^{18}P^{19}O^{20}G^{21}P^{22}O^{23}G^{24}$, and Chain C: $G^9P^{10}O^{11}G^{12}P^{13}O^{14}\mathbf{X}^{15}P^{16}O^{17}G^{18}P^{19}O^{20}G^{21}P^{22}O^{23}$.

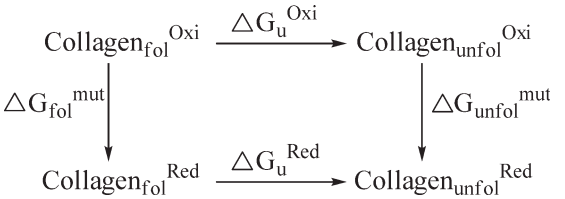

FIGURE 3 Thermodynamic cycle used to calculate free energy differences between an oxidized form and a reduced form of the mutant containing Cys residues at the folded and unfolded state.

Two model peptide systems were prepared. One model peptide has a disulfide bond between Chain A and Chain B, where X's are Cys_{oxi} in Chains A and B and Gly in Chain C (Cys_{oxi} is an oxidized form of a Cys residue). The other model has a disulfide bond between Chain and C, where X's are Cys_{oxi} in Chain A and C and Gly in Chain B.

The system was minimized and simulated with SBMD in the same way as the free energy simulations for a single residue mutation previously mentioned. The thermodynamic cycle (Figure 3) was used to estimate the stability difference between an oxidized and a reduced form of the mutant by calculating free energy differences of chemical mutations at the folded and unfolded states ($\Delta\Delta G = \Delta G_{fol}^{mut} - \Delta G_{unfol}^{mut}$). The thermodynamic properties were also calculated in the coupling parameter approach[27–30] using a hybrid potential energy function given by $U(\lambda) = (1 - \lambda)U_{Oxi} + \lambda U_{Red} = U_{Oxi} + \lambda\Delta U$, where U_{Oxi} corresponds to the mutated system with a pair of oxidized residues forming a disulfide bond, U_{Red} corresponds to the same system with a pair of reduced residues at the same position, and $\Delta U = U_{Red} - U_{Oxi}$ (U_{Oxi} for the initial state ($\lambda = 0$) and U_{Red} for the final state ($\lambda = 1$)). The equilibration and dynamics at each λ were carried out for 200 ps and 1 ns, respectively. The forward and reverse simulations were performed at two different chains containing a disulfide bond. One nanosecond of SBMD of the systems required about 40 h of central processing unit time on the dual Quad-core Intel Xeon cluster.

Error Analysis

The bin-average method was used to estimate the errors of the free energy change differences for all the mutations performed in this study. For each simulation, the data were divided into 20 contiguous bins of equal size within each window for $\lambda = 0.1$, 0.3, 0.5, 0.7, and 0.9. An average within each window and the variance of the sample of the 20 averages were calculated. The errors of the overall free energy difference changes were then evaluated by using standard error analysis, with the reported errors corresponding to 95% confidence intervals of the mean.[31] The resulting statistical errors of the free energy difference changes for all the mutations were about 0.5 kcal/mol. For each mutation, the estimated free energy changes

obtained from the separate forward and reverse trajectories agreed within their statistical errors. This indicates that the free energy simulation results are converged.

Approximations Used

Our goal was to build a simple molecular model, based on the experimental structure, which captures the main interactions responsible for collagen properties, but is computationally tractable. Thus, our simulations involve a number of simplifications and approximations. To limit the number of simulations, we do not consider all possible mutation patterns, but only single mutations in the center of Chain A and second mutations in the centers of Chains B and C. The calculated free energy changes only approximately correspond to the quantities shown in the thermodynamic cycle in Figure 2. In the simulations, we do not use the full 27-residue collagen triple helix to represent the folded state, but only the central fragment centered at the mutation site, as described in Methods section. Analogously, only a five-residue central fragment rather than the full 27-residue peptide was used to represent the unfolded state. We hope that any neglected long-range effects to individual free energy changes ΔG will cancel in the calculations of their differences $\Delta\Delta G$. Further, we use the SBMD method in which only part of the collagen-like peptide close to the mutation site was solvated and allowed to fluctuate, with standard nonbond cutoffs, limited explicit solvation, and without including counterions. This simulation setup does not allow any large structural changes in the collagen-like peptide because of constraints. Our peptide systems were solvated in a water sphere of a radius of 16 Å to expedite the simulations, and we used a relatively short nonbonded cutoff distance of 14 Å. The limited solvation and use of nonbonded cutoffs do not allow for an accurate description of long-range electrostatic effects, introducing a bias in the calculated free energy changes. This is especially true for mutations from neutral to charged side chains, Gly → Glu and Asp, for which there is a change in the total system charge and electrostatic contributions are dominant. The simulations could be improved by introducing more water, including counterions, and use of longer nonbonded cutoffs or more accurate methods such as multipole expansion.[32] In the current setup, the simulation results involving mutations to charged amino acids are quite approximate, and we hope for cancellation of at least some of the neglected long-range effects through the use of thermodynamic cycles. We used only five simulation windows with $\lambda = 0.1$, 0.3, 0.5, 0.7, and 0.9, with 6.2 ns of total MD for each free energy simulation. In our study, we used free energy decompositions, which, unlike the total free energy changes, exhibit path dependence.[33–36] Specifically, we decompose the $\Delta\Delta G$ obtained from alchemical processes in the thermodynamic cycle using the uniform scaling approach, i.e., with all energy terms simultaneously scaled by the same value of the coupling parameter λ. Boresch et al.[34] argued that decomposition of the alchemical $\Delta\Delta G$ is appropriate for obtaining physical insights into the effects of point mutations on protein–ligand binding and protein thermal stability, whereas decomposition of $\Delta\Delta G$ from the chemical steps would be appropriate for analyzing effects on reaction paths for binding or folding. Expanding this analysis to our thermodynamic cycles, we can argue that in global terms decomposing the alchemical $\Delta\Delta G$ for collagen complex formation is the appropriate approach for analyzing microscopic effects on triple-helix stability.[33,34] Our decomposition results still exhibit local path dependence because of our choice of uniform scaling. In terms of theoretical properties, the uniform scaling results have the pleasing property of symmetric distribution of nondiagonal terms.[34] It might be expected that nonuniform scaling, e.g., decoupling the switching of electrostatic and van der Waals terms would lead to somewhat different component values.[34] With these caveats in mind, we use the decompositions to provide helpful qualitative insights into the forces driving changes in stability because of mutation. Finally, any predictions based on empirical force fields are, by their nature, approximate because of limits in parameter accuracy.

RESULTS AND DISCUSSION

The molecular structures of the collagen-like peptide and its partially solvated peptide with a water sphere are represented schematically in Figure 1. We have carried out SBMD simulations on the solvated part of the peptides for the mutations from Gly to Asp, Cys, Glu, and Ser to investigate the stability of the mutated collagen-like peptides relative to the wild type. The mutations of Gly → Asp and Gly → Glu were used to represent the most severe and moderate phenotypes of OI, respectively. The Gly → Ser and Gly → Cys mutations were chosen to represent mild phenotypes of the disease. The MD simulations for the first and second Gly mutations were carried out to estimate the degree of destabilizing effect from each mutation to understand the effect on the stability of homozygous and heterozygous mutants in collagen Type I. Furthermore, we have also performed free energy simulations to quantify the free energy difference between an oxidized form and a reduced form of the mutant containing two Cys residues.

Relative Free Energy Differences Between the Wild Type and the Mutants (Containing Cys, Ser, Asp, and Glu) by Single Point Mutations

Relative total free energy differences indicating the thermal stability of the wild type and mutants containing Cys, Asp, Glu, and Ser at position 15 in Chain A of the homotrimeric collagen-like peptide are summarized in Table I. The calculated results ($\Delta\Delta G_{\text{Total}}$) indicate that the wild type is more stable than the mutants with an Asp, a Glu, a Cys, and a Ser residue by 8.8 ± 0.5, 5.6 ± 0.6, 4.2 ± 0.5, and 3.8 ± 0.3 kcal/mol, respectively. The $\Delta\Delta G_{\text{Total}}$ values follow the same trends as the stability of mutants of collagen-like peptides with respect to the wild type, which is closely associated with the severity of OI from statistical analysis.

The free energy changes of the folded state ($\Delta G^{\text{Fold}}_{\text{Total}}$) for mutations Gly → Ser and Gly → Cys are 12.6 and 11.4 kcal/mol, respectively, and the corresponding free energy changes of the unfolded state ($\Delta G^{\text{Unfold}}_{\text{Total}}$) are 8.8 and 7.2 kcal/mol. These two mutations destabilize both the systems at the

Table I Contributions to Free Energy Differences ($\Delta\Delta G$) for Gly → Xaa Mutations in the Collagen-Like Peptides (Xaa: Cys, Asp, Glu, and Ser) (Unit: kcal/mol)

Xaa	Ser	Cys	Glu	Asp
$\Delta\Delta G_{\text{Total}}$	3.8	4.2	5.6	8.8
$\Delta\Delta G_{\text{Elec}}$	−0.2	0.0	−6.5	−0.5
$\Delta\Delta G_{\text{vdW}}$	2.8	2.3	4.7	4.0
$\Delta\Delta G_{\text{Cov}}$	1.2	1.9	4.1	5.3
Folded state				
$\Delta G^{\text{Fold}}_{\text{Total}}$	12.6	11.4	−76.8	−95.6
$\Delta G^{\text{Fold}}_{\text{Elec}}$	1.5	2.7	−104.3	−118.0
$\Delta G^{\text{Fold}}_{\text{vdW}}$	2.9	−0.5	7.0	8.2
$\Delta G^{\text{Fold}}_{\text{Cov}}$	8.1	9.2	20.5	14.2
Unfolded state				
$\Delta G^{\text{Unfold}}_{\text{Total}}$	8.8	7.2	−82.4	−104.4
$\Delta G^{\text{Unfold}}_{\text{Elec}}$	1.7	2.7	−101.1	−117.5
$\Delta G^{\text{Unfold}}_{\text{vdW}}$	0.1	−2.9	2.3	4.2
$\Delta G^{\text{Unfold}}_{\text{Cov}}$	6.9	7.4	16.4	8.9

Average values of forward and reverse calculations. Statistical errors of $\Delta\Delta G_{\text{Total}}$ for mutants containing a Ser, a Cys, a Glu, or an Asp relative to the wild type are 0.3, 0.5, 0.6, and 0.5 kcal/mol, respectively.

folded and unfolded states, but they destabilize the systems at the folded state more than those at the unfolded state. By way of contrast, the free energy changes for mutations Gly → Glu and Gly → Asp at the folded state ($\Delta G^{\text{Fold}}_{\text{Total}}$) are −76.8 and −95.6 kcal/mol, respectively, and the corresponding energy changes at the unfolded state ($\Delta G^{\text{Unfold}}_{\text{Total}}$) are −82.4 and −104.4 kcal/mol, respectively. The mutations from Gly to Glu and Asp stabilize the system at the folded and unfolded state, but they stabilize the systems at the folded state less than those at the unfolded state. For the Gly → Glu and Gly → Asp mutations, the electrostatic energy term is the most significant contribution to the free energy changes both at the folded and unfolded states. The remarkably different contribution to the free energy change at the folded and unfolded states between Gly → Ser and Cys mutations and Gly → Glu and Asp mutations is due to unfavorable and favorable electrostatic energy term, respectively.

The contributions of different terms in potential energy function and different structural groups are shown in the Supporting Material (Table S1). The self and interaction energy terms based on Eq. (2) have been also evaluated and are given in Table S1. Each contribution is separated into covalent, electrostatic, and van der Waals terms defined in the CHARMM potential energy function. The net effects ($\Delta\Delta G_{\text{Total}}$) for all the systems are to destabilize mutants relative to the wild type. The most dominant unfavorable contributions to the mutants come from van der Waals interaction terms, which mainly determine the stability of all the mutants with respect to the wild type. The unfavorable covalent interaction energy terms for the mutants containing Ser, Cys, Glu, and Asp are 1.2, 1.9, 4.1, and 5.3 kcal/mol, respectively. These values indicate the extent of destabilizing effect of the mutants with respect to the wild type by single mutations. The mutant with Asp is the most unstable, and the mutant bearing Ser is the least unstable among the four variants. For the mutant with Glu, the van der Waals term is an even more unfavorable term than its covalent term, and is the most unfavorable contribution compared with the terms in the other mutants. However, the Glu mutant is more stable than the Asp one because of the favorable electrostatic contribution to the Glu-containing system.

We estimated average backbone dihedral angles from the trajectories of forward and reverse direction for the first single point mutations in Table II. The dihedral angle means and standard deviations surrounding the mutation site (residue 14, 15, and 16) were calculated for all φ and ψ values. The average values of φ at the mutation site for all the mutations were in the range of −72° to −80° and were similar values to the typical $\varphi = -78°$[37] found in polyproline II helices. However, the dihedral angle ψ values at the mutation sites for the mutations from Gly to Asp and Glu were −138.8° and −143.8°, respectively, which are significantly different from the typical ψ value (149°) of polyproline II.[37] The variation of φ and ψ along the peptide chain for Gly → Xaa mutations shows that torsion angles of all three chains are similar except the mutation sites on Chain A in the first mutations of Gly → Glu and Gly → Asp, where the helix of the chain containing the mutated residue unwinds locally (Table II). Only two residues in mutants containing Glu and Asp residue remarkably deviate from the typical angles for φ (−78°) and ψ (149°) in polyproline II, whereas most of the torsion angles remain unchanged.[37] These dihedral angle deviations during the simulation for the Gly → Asp and Gly

Table II Average Values of Peptide Dihedral Angles (φ and ψ) of Residues 14, 15, and 16 in Chain A from the Trajectories with $\lambda = 0.9$ for the First Gly → Xaa Mutations (Xaa: Cys, Asp, Glu, and Ser) (Average ± SD; Unit: Degrees)

Xaa	Cys	Asp	Glu	Ser
		φ		
Hyp 14	−64.8 ± 10.0	−60.1 ± 8.5	−60.6 ± 9.0	−64.7 ± 9.9
Gly 15	−79.7 ± 14.9	−80.3 ± 9.0	−77.8 ± 9.2	−71.8 ± 12.8
Pro 16	−55.1 ± 12.2	−71.8 ± 13.3	−70.8 ± 12.1	−56.3 ± 12.2
		ψ		
Hyp 14	141.9 ± 21.6	129.1 ± 8.6	129.8 ± 8.2	141.9 ± 16.9
Gly 15	140.6 ± 8.1	−138.8 ± 16.2	−143.8 ± 19.6	142.7 ± 8.7
Pro 16	155.7 ± 24.1	150.7 ±10.7	152.1 ± 16.4	157.3 ± 26.3

→ Glu mutations indicate introduction of significant structural deformations from the initial minimized structure used in the MD simulations (Figure 4). In general, the values of the torsion angles do not deviate significantly from those starting structures. In the crystal structure of the Ala-containing mutant of collagen-like peptide, only five residues in the substitution zone of the peptide exhibited significantly deviated torsion angles.[24]

We calculated the root mean square deviation by fitting the solvated backbone atoms of the average structure from the trajectories ($\lambda = 0.9$) of each mutation to a minimized structure during the SBMD simulations. The root mean square deviation of the solvated backbone atoms for Gly → Ser and Gly → Cys mutations is 0.6 Å and that of Gly → Asp and Gly → Glu mutation is 0.8 and 0.7 Å, respectively. The structural changes of the collagen-like peptide generated by the Gly → Xaa mutations are small. The structural deformations estimated from dihedral angle analysis of the Gly → Asp and Gly → Glu mutations seem to be related to decreasing the stability of the collagen-like peptide mutants containing Asp and Glu, leading to OI. The subtle structural change of the Gly substitutions would provide structural effect of Gly mutations in collagen as seen in the crystal structure of the Ala-containing mutant of the collagen model peptide.[24] This small structural change would cause pathological states in fibril collagens.[38] However, although the local structural deviation is small, the untwisting of the triple helix may cause spatial rearrangement between the three helical chains, contributing to the long-range effects observed in Gly substitution.[39,40]

Interaction energies between the side chain of the residue Xaa 15 and the H_α's of Gly 15 on the other two chains were calculated for all the first Gly → Xaa mutations in Table III. The interaction energy differences between H_α's of Gly 15 on Chains B and C relative to the side chain of the mutation site on Chain A are ∼1 kcal/mol for Gly → Ser and Gly → Cys mutations and ∼3 kcal/mol for Gly → Glu and Gly → Asp mutations. These interaction energies from Gly 15 on Chains B and C are due mainly to electrostatic energies rather than van der Waals terms for all the mutations. The interaction energy from the Gly 15 on Chains B and C stabilizes the mutants, although the net effect ($\Delta\Delta G_{\text{Total}}$) destabilizes the mutant relative to the wild type. The different magnitudes of the interaction energies between the Gly 15 on the different chains and the mutation site indicate that the chemical environments of each Gly 15 on Chains B and C are different for all four mutations even if the mutations are at the same position in homotrimeric peptides. Thus, the simulations of the second Gly → Xaa mutations were carried out in two different positions on Chains B and C.

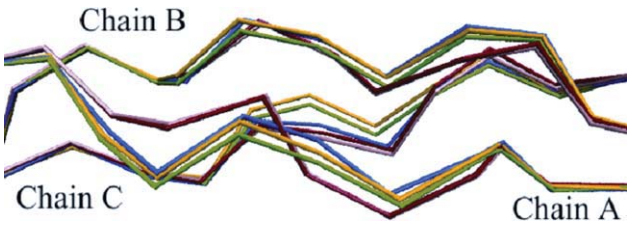

FIGURE 4 Overlay of the backbone atoms between a minimized initial structure (blue) and average structures from the trajectories ($\lambda = 0.9$) of each mutation. Yellow, Gly → Ser mutation; green, Gly → Cys mutation; pink, Gly → Glu mutation; red: Gly → Asp mutation. The figure was prepared using the VMD program.[36]

Table III Contributions of Interaction Energy of H_α's of Gly 15 on Chain B and Chain C with Respect to the Mutation Site to the Free Energy Differences Between Wild Type and Mutants (Gly → Xaa; Xaa: Cys, Asp, Glu, and Ser) (Unit: kcal/mol)

Xaa	Chain B	Chain C
Ser	−0.5	−1.3
Cys	−0.5	−1.2
Glu	−2.3	−4.8
Asp	−2.9	−6.2

Statistical errors for all the mutations were within 0.3 kcal/mol.

The total free energy change differences for the second Gly → Xaa mutations were calculated at position 15 on two different chains (Chains B and C) and are shown in Table IV. The average values of total free energy changes of the mutations on two different sites follow the same trend of destabilizing effects on the stability of the mutants of the collagen-like peptides shown in the first Gly mutations (Figure 5). However, the magnitude of destabilizing effect on the mutated collagen-like peptides by the second Gly mutations is smaller than that by the first Gly mutations. The corresponding free energy differences of a second Gly mutation at position 15 in different chains were, on average, 1.3, 1.5, 2.9, and 5.4 kcal/mol, respectively. The relative stability of the collagen-like peptides with second mutations calculated by MD simulations was qualitatively consistent with severity of OI for the mutants. The statistical percent OI lethalities of Ser, Cys, Glu, and Asp mutations for Gly are 30, 39, 50, and 68%, respectively,[14] showing the same trend as our simulation results of stability of the mutants with respect to the wild type. Our simulation results have shown that, for the first and second Gly mutations, the Gly → Ser and Gly → Cys mutations destabilize the collagen-like peptide systems at the folded and unfolded states, but destabilize the peptides at the folded state more than those at the unfolded state. However, the Gly → Glu and Gly → Asp mutations stabilize the peptide systems at the unfolded and folded states, but stabi-

Table IV Total Free Energy Change Differences ($\Delta\Delta G$) for Second Gly Mutations at Two Different Chains for the Gly → Xaa Mutations in the Collagen-Like Peptide (Xaa: Cys, Asp, Glu, and Ser) (Unit: kcal/mol)

Xaa	Ser	Cys	Glu	Asp
A:B—the first mutation on chain A and the second mutation on chain B				
$\Delta\Delta G_{Total}$	0.9	0.7	4.3	3.4
$\Delta\Delta G_{Elec}$	−1.2	0.5	4.8	4.1
$\Delta\Delta G_{vdW}$	1.9	0.1	−0.7	−1.9
$\Delta\Delta G_{Cov}$	0.2	0.2	0.1	1.3
Folded state				
ΔG^{Fold}_{Total}	9.6	7.8	−78.1	−101.0
ΔG^{Fold}_{Elec}	0.6	3.2	−96.2	−113.4
ΔG^{Fold}_{vdW}	2.0	−2.8	1.6	2.2
ΔG^{Fold}_{Cov}	7.1	7.5	16.6	10.2
Unfolded state				
$\Delta G^{Unfold}_{Total}$	8.8	7.2	−82.4	−104.4
ΔG^{Unfold}_{Elec}	1.7	2.7	−101.1	−117.5
ΔG^{Unfold}_{vdW}	0.1	−2.9	2.3	4.2
ΔG^{Unfold}_{Cov}	6.9	7.4	16.4	8.9
A:C—the first mutation on chain A and the second mutation on chain C				
$\Delta\Delta G_{Total}$	1.5	2.3	1.4	7.3
$\Delta\Delta G_{Elec}$	0.2	0.6	0.5	3.0
$\Delta\Delta G_{vdW}$	1.2	0.1	−0.1	1.4
$\Delta\Delta G_{Cov}$	0.1	1.6	1.1	2.9
Folded state				
ΔG^{Fold}_{Total}	10.3	9.5	−81.0	−97.1
ΔG^{Fold}_{Elec}	1.9	3.3	−100.6	−114.5
ΔG^{Fold}_{vdW}	1.3	−2.8	2.1	5.6
ΔG^{Fold}_{Cov}	7.1	9.0	17.5	11.8
Unfolded state				
$\Delta G^{Unfold}_{Total}$	8.8	7.2	−82.4	−104.4
ΔG^{Unfold}_{Elec}	1.7	2.7	−101.1	−117.5
ΔG^{Unfold}_{vdW}	0.1	−2.9	2.3	4.2
ΔG^{Unfold}_{Cov}	6.9	7.4	16.4	8.9

Average values of forward and reverse calculations. Statistical errors of $\Delta\Delta G_{Total}$ for all second Gly mutations on Chains B and C are approximately 0.5 kcal/mol.

lize the systems at the unfolded state more than those at the folded state (Tables I and IV). The contributions to $\Delta\Delta G$ are decomposed into their individual potential energy terms and are presented in Table S2 (Supporting Information). For the second Gly mutations to Cys, all the energy terms (covalent, electrostatic, and van der Waals) contributing to $\Delta\Delta G_{Total}$ at both mutations in Chains B and C stabilize the wild type. The difference of total free energy changes for the second mutations Gly → Asp and Gly → Glu is mainly from the electrostatic interaction term. The van der Waals term stabilizes the mutant in Chain B but destabilizes it in Chain C. The difference of the overall free energy change for mutation Gly → Ser comes from the electrostatic interaction energy term. Noncovalent interaction energy terms for all the second mutations determine the differences of the total free energy changes in Chains B and C.

Electrostatic terms in all the mutations at the folded and unfolded states make much larger contributions to free energy changes of the mutations than van der Waals and covalent terms. Although these significant contributions to free energy changes from electrostatic terms at the folded state are mostly canceled out by those at the unfolded state, the stabilizing and destabilizing effects of electrostatic terms are the most crucial to determine the relative thermal stability of the mutants with respect to the wild type. The side chains of Glu and Asp can make good hydrogen bonds with solvent, which provides these side chains with favorable electrostatic term at both the folded and unfolded state. Moreover, this favorable electrostatic term is larger at the unfolded state than at the folded state. The bulky side chains of Glu and Asp caused dihedral angle deformations at the mutation site of the peptide backbones at the folded state during the simulation (Table II), which could cause the mutants to become less stabilized relative to the wild type.

MD simulation on the fully solvated collagen for the G → A mutation was previously carried out to estimate stability of the mutant relative to the wild type.[17] Their estimated free energy difference between the mutant containing three Ala residues and the wild type was 10.8 ± 3.0 kcal/mol,[17] which was consistent with the experimental value of 11.9 kcal/mol.[24] The simulated $\Delta\Delta G_{Total}$ values of the first and second

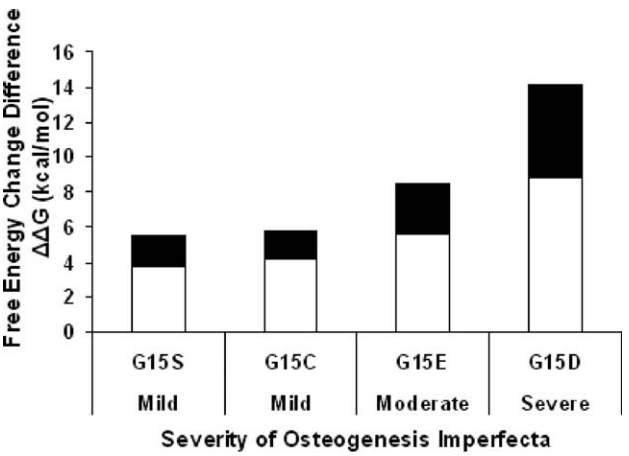

FIGURE 5 Comparison of contributions to the free energy differences (kcal/mol) for the Gly → Xaa mutations (Xaa: Ser, Cys, Glu, and Asp) with the severity of OI. First Gly → Xaa mutations (white) and average values of two second Gly → Xaa mutations (black). A mutant containing two Cys residues is a reduced form.

G → A mutations from the fully solvated system were 6.36 ± 1.54 and 3.84 ± 2.14 kcal/mol, respectively.[17] To validate our partially solvated collagen-like peptide model, we also carried out free energy simulations for the G → A mutation using the same simulation method as previously described. The $\Delta\Delta G_{Total}$ value of the first G → A mutation from our simulations is 5.07 ± 0.60 kcal/mol ($\Delta G^{Fold} - \Delta G^{Unfold} = 10.42 - 5.35$), and the average value of the second mutation on Chains B and C is 1.62 ± 0.50 kcal/mol ($\Delta G^{Fold} - \Delta G^{Unfold} = 6.97 - 5.35$). Our simulation results are consistent with the corresponding simulation values from the previous fully solvated system. Our free energy change differences consistent with the corresponding previous results with fully solvated system indicated that our partially solvated peptide system is valid to estimate stability between the wild type and mutants of the collagen model.

All mutations described so far were performed for residues on Chain A on the triple helix. To estimate the influence of this choice on the calculated free energies, we also carried out free energy simulations for the mutation from Gly to Ser in Chains B and C. The results indicate that the free energy changes for first and second G → S mutations are similar for Gly residues located in all three chains (Table S3, Supporting Information).

Free Energy Difference Between a Reduced Form and an Oxidized Form of a Mutant Containing Two Cys Residues

The mutants of the model peptide containing two Cys residues at position 15 could possibly exist as an oxidized or a reduced form. To estimate the relative free energy difference between the oxidized and reduced form of the mutant, free energy simulations were carried out using the same SBMD approach with the thermodynamic integration method. Furthermore, because of the different chemical environment of Cys at position 15 on Chains B and C shown in Table III, we also calculated the free energy difference for mutants containing a disulfide bond between Chains A and B, as well as between Chains A and C with respect to the corresponding reduced forms. Calculated free energy differences between the oxidized and reduced forms for the two mutants are shown in Table V. The free energy changes for the mutation from the oxidized to the reduced form ($Cys_{Oxi} \rightarrow Cys_{Red}$) are 7.0 and 8.9 kcal/mol for the mutants containing a disulfide bond between Chains A and B (A:B) and Chains A and C (A:C), respectively, for the folded state. The corresponding free energy change is 6.21 kcal/mol for the unfolded state. The mutation $Cys_{Oxi} \rightarrow Cys_{Red}$ destabilizes both the folded and unfolded states, but it destabilizes the former system more than the latter. The difference in the free energy in aqueous solution between the oxidized and reduced forms is calculated to be $\Delta\Delta G = \Delta G^{Fold}_{Total} - \Delta G^{Unfold}_{Total} = 0.8 \pm 0.4$ kcal/mol for A:B and 2.7 ± 0.4 kcal/mol for A:C. As expected, the mutants containing a disulfide bond between Chains A and B and between Chains A and C are more stable than the corresponding reduced forms by 0.8 and 2.7 kcal/mol, respectively. Because the covalent term stabilizes the system at the unfolded state more than both mutants at the folded state, the most dominant favorable contribution to $\Delta\Delta G_{Total}$ for stabilizing the oxidized form is from the covalent term for both the mutants having a disulfide bond (between Chains A and B, 2.8 kcal/mol and between Chains A and C, 3.1 kcal/mol). However, the electrostatic term destabilizes the system at the folded and unfolded state, with a stronger effect in the latter state. The net effect from the electrostatic term stabilizes the reduced mutants relative to the oxidized mutants. The interaction energy terms for both cases contribute to the $\Delta\Delta G_{Total}$ more significantly than self terms (Table S4, Supporting Information). Electrostatic and van der Waals energy terms stabilize the reduced forms, but the covalent term stabilizes the oxidized form. However, because the covalent interaction energy terms are more significant than the noncovalent interaction energy terms, the overall contribution of interaction energy terms to stability is more favorable to the oxidized forms than to the reduced forms. The relative stability of the mutants containing one Cys or two (oxidized or reduced) Cys residues relative to the wild type is shown in

Table V Contributions to ($\Delta\Delta G$) for $X = C_{Oxi} \rightarrow C_{Red}$ in the Collagen-Like Peptides (Unit: kcal/mol)

Xaa	A:B[a]	A:C[b]
$\Delta\Delta G_{Total}$	0.8	2.7
$\Delta\Delta G_{Elec}$	−1.2	−1.0
$\Delta\Delta G_{vdW}$	−0.8	0.6
$\Delta\Delta G_{Cov}$	2.8	3.1
Folded state		
ΔG^{Fold}_{Total}	7.0	8.9
ΔG^{Fold}_{Elec}	8.0	8.2
ΔG^{Fold}_{vdW}	0.8	2.2
ΔG^{Fold}_{Cov}	−1.8	−1.5
Unfolded state		
$\Delta G^{Unfold}_{Total}$	6.2	
ΔG^{Unfold}_{Elec}	9.2	
ΔG^{Unfold}_{vdW}	1.6	
ΔG^{Unfold}_{Cov}	−4.6	

Average values of forward and reverse calculations. Statistical errors of $\Delta\Delta G_{Total}$ for all the mutations from a reduced to an oxidized form are about 0.4 kcal/mol.

[a] A:B means an oxidized mutant containing a disulfide bond between Chains A and B.

[b] A:C means an oxidized mutant containing a disulfide bond between Chains A and C.

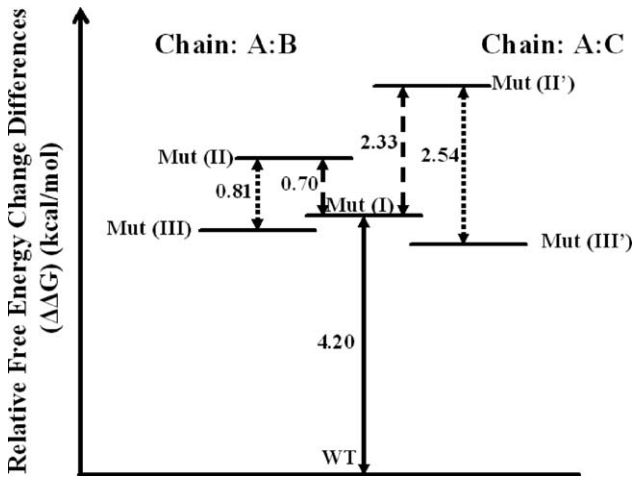

FIGURE 6 Relative free energy differences of the wild type and mutants containing one Cys and two oxidized and reduced Cys residues of the collagen-like peptides. Chain A:B means the oxidized mutant bearing a disulfide bond between Chains A and B (unit: kcal/mol). WT, wild type; Mut(I), mutant containing a Cys residue on Chain A; Mut(II), mutant containing two Cys residues (reduced) on Chains A and B; Mut(II′), mutant containing two Cys residues (reduced) on Chains A and C; Mut(III), mutant containing a disulfide bond between Chains A and B; Mut(III′), mutant containing a disulfide bond between Chains A and C; first mutation, WT → Mut(I) (solid line); second mutation, Mut(I) → Mut(II) and Mut(I′) → Mut(II′) (sparsely dotted line); and disulfide bond formation, Mut(II) → Mut(III) and Mut(II′) → Mut(III′) (dotted line).

Figure 6. The wild type is more stable than the mutants containing a disulfide bond by approximately 4 kcal/mol and than the mutant containing a Cys residue at position 15 on Chains B and C by 4.9 and 6.5 kcal/mol, respectively. The mutant with a disulfide bond is more stable than the mutant containing one Cys residue but less stable than the wild type (Figure 6).

For the mutants containing two Cys residues, both oxidized forms of the mutants were more stable than the reduced forms by 0.8 and 2.7 kcal/mol for the disulfide bond between Chains A and B and Chains A and C, respectively. The main contributions to these free energy change differences ($\Delta\Delta G_{\text{Total}}$) are from electrostatic and covalent terms in both mutants. In the $\text{Cys}_{\text{Oxi}} \rightarrow \text{Cys}_{\text{Red}}$ mutations, the electrostatic term destabilizes the peptides at the folded and unfolded states, but it disfavors the systems at the unfolded state more than those at the folded state. On the contrary, the covalent term stabilizes the mutants at the folded and unfolded states, but it favors the system at the unfolded state more than that at the folded state. The relative free energy differences of $\Delta\Delta G(\text{oxi} \rightarrow \text{red})$ for both mutations are similar to the values experimentally obtained from T4 lysozyme cross-linked by a disulfide bond.[41] The T4 lysozyme was engineered to generate five different mutants containing a disulfide bond, which provided the enzyme with increased stability.[41,42] The five mutated T4 lysozymes bearing a disulfide bond were more stable than the reduced protein by 0.8–2.8 kcal/mol.[41] The different values of $\Delta\Delta G_{\text{Total}}(\text{oxi} \rightarrow \text{red})$ of the mutated T4 lysozymes were influenced by location of the disulfide bond and protein flexibility. These results indicated that stabilization of the protein by the effect of the cross-link on the entropy of the unfolded polypeptide was offset by the strain energy from formation of the disulfide bond in the folded protein.[41] Our calculated values of $\Delta\Delta G_{\text{Total}}(\text{oxi} \rightarrow \text{red})$ of the mutants containing two Cys residues are similar to the experimental values ($\Delta\Delta G_{\text{Total}}(\text{oxi} \rightarrow \text{red})$) from T4 lysozymes cross-linked by a disulfide bond.[41]

Beck et al.[18] studied the relative stability of residues replacing Gly in the host–guest peptide composed of the repeating units of (Gly-Pro-Hyp) to isolate the influence of Gly mutations on the helix disruption in a constant and homogenous environment. Their results demonstrated that the order of disruption in the seven Gly → Xaa peptides (Xaa: Ala, Cys, Ser, Val, Arg, Glu, and Asp) correlated with the identity of the residue replacing Gly in the OI. Their experimental thermodynamic analysis suggested that the decrease in stability of the peptides results from entropic destabilization.[18] The relative clinical severity was consistent with the order of disruption of different residues in their peptide.[13] Relative stability of normal collagen Type I and its mutants containing one or two α1(I) chains that had a Cys residue within the triple-helix has been investigated with skin fibroblasts grown from three patients with OI.[43] This study showed that the collagen mutants with two Cys residues formed intra- and intermolecular disulfide bonds. Although direct thermal measurements on OI collagens may be complicated by heterogeneity of the molecule population and the use of digestive enzymes to measure thermal stability, the thermal stability from all three patients having Cys-containing collagen mutants indicated that the mutant containing disulfide bonds is thermally more stable than the mutant containing one Cys residue but less stable than a wild type. This relative stability of the normal and mutated collagens is consistent with our simulated relative stability of the collagen-like peptides.

The mutated collagens containing two mutated residues are energetically less stable than the mutated collagens bearing one mutated residue for mutations of Gly to Asp, Glu, and Ser but not for mutation of Gly to Cys because of the formation of a disulfide bond in the mutants containing two Cys residues. Thus, comparison of the overall relative free energy difference of the Gly → Xaa mutations studied in this work indicates that the order of destabilizing effect can be represented as $\text{Cys}_{\text{oxi}} < \text{Ser} < \text{Cys}_{\text{red}} < \text{Glu} < \text{Asp}$. This

order of disruptive effect of the residue mutated for Gly follows the order of observed severity of OI from statistical analysis. The free energy change differences provide a basis at the atomic level for estimating other factors such as the amino acid sequence environment adjacent to the mutation site or the location of the mutation site within the collagen chain, in determining the clinical severity of OI phenotypes.

The Buehler group[44,45] has assessed the effect of Gly mutations related to OI on the molecular, intermolecular, and collagen fibril mechanical properties. OI mutations have been known to affect the mechanical properties of collagenous tissues. The authors estimated the Young's modulus (used to characterize the stiffness of the materials) by way of expressing the peptide's stiffness as two functions of the hydropathy and the residue volume. The Young's modulus of the reference peptide and its Gly mutants were determined to be 3.96 $\pm$ 0.21 GPa[46,47] and 3.37 $\pm$ 0.32 $\sim$ 3.78 $\pm$ 0.29 GPa,[44] respectively. They also carried out steered MD simulations to understand the effect of locally softened domains on the structural behavior of the tropocollagen models composed of $[(GPO)_5\text{-}(XPO)\text{-}(GPO)_4)]_3$ (X = Ala, Ser, Cys, Arg, Val, Glu, and Asp) with protonated N-terminals and deprotonated C-terminals in explicit water[44] using the GROMACS program and GROMOS96 43a1 force field.[48,49] They found that mutations showing more severe phenotypes were related to softer tropocollagen mechanical properties, which was confirmed by a linear curve fit to the Young's modulus results over the severity of OI single mutations.[44] The Poisson–Boltzmann method was used to obtain estimates of the intermolecular interaction energies in aqueous solvent.[45] At this intermolecular level, OI mutations caused a weakening of intermolecular adhesion, increase in intermolecular equilibrium spacing, and reduction in likelihood of cross-link formation.[45] Coarse-grained mesoscale MD simulations of collagen fibrils were carried out to investigate reduction of stiffness caused by OI. At this collagen fibril level, the mutations caused reduction in strength of more than 50% and 35% in a cross-linked and a cross-linked free fibril, respectively, compared with a wild type and a formation of localized stress concentration leading to early material failure.[45]

The stability of the collagen helix provided by the repeating sequence of Gly, Pro, and Hyp residues is reduced by single point mutations of Gly by a charged, polar, or hydrophobic residue. Our simulated results indicated that mutants containing an Asp or a Glu are less stable than the mutants containing a Ser or a Cys, which are less stable than the wild type. Replacing Gly by large and charged side chains tended to yield largest decreases in the stability of the collagen. Furthermore, structurally, the triple helix is more distorted by larger and charged side chains of Glu and Asp. The loss of hydrophobic stabilization of the triple helix affects the kinetics of the folding. The folding rate of the mutants is slower than that of the wild type.[50] The reduced folding rate of the triple helix is related to the OI. The extent of the defective folding or altered folding mechanism associated with the decreased stability of the mutated collagens may determine the degree of clinical severity of the OI. Our results showed that depending on the identity of the residue replacing Gly, mutated collagens exhibit a small decrease in stability or show local conformational perturbation at an atomistic level. The slight changes are also associated with local mechanical properties of single tropocllagen molecules.[44] The most unstable collagen mutant containing an Asp has the most soft mechanical property among the four mutants studied. According to previous work, the mutated collagen molecules may be incorporated into fibrils with normal molecules, but these mixed fibrils are not mineralized in bone in a normal way.[51,52]

CONCLUSIONS

Free energy simulations of a collagen-like peptide for the mutations from Gly to Ser, Cys, Glu, and Asp have been carried out to investigate the effect of the single mutated residue on the stability of the collagen, leading to OI. The calculated free energy changes agree with the trend in severity of OI from statistical analysis in collagen Type I. Of the four mutations, the most disruptive residue such as Asp not only is consistent with the most severe clinical symptom, but also has most soft mechanical property. The relative free energy change differences are also consistent with the clinical lethality of OI from collagen Gly mutations. Furthermore, free energy changes between an oxidized and a reduced form of the mutant containing two Cys residues demonstrate that the mutant containing a disulfide bond is more stable than the mutant containing one Cys residue but less stable than the wild type. This calculated result implies that the oxidized form is mostly found in the mutant having two Cys residues rather than the reduced form because of more stable oxidized form of the mutant. This result is consistent with clinical experimental results with three patients demonstrating the existence of intra- or intermolecular disulfide bond and the stability of mutants with a single mutated Cys residue and two reduced and oxidized Cys residues relative to the wild type. The free energy differences between the oxidized and the reduced form of the Cys-containing mutant are 0.8 and 2.7 kcal/mol, respectively, for the mutants containing the disulfide bonds between Chains A and B and between Chains A and C. This work will improve our ability to predict the effect on the stability of mutated collagens relative to a wild

type and to understand and analyze the clinical severity of OI and collagen-related diseases because of a single point mutation or by disulfide bond formation.

The authors thank Prof. Charles L. Brooks III for commenting on the manuscript and allowing us to use his dual Quad-core Intel Xeon cluster and Dr. Jennifer L. Knight for commenting on the manuscript.

REFERENCES

1. Rich, A.; Crick, F. H. C. J Mol Biol 1961, 3, 483–506.
2. Fraser, R. D. B.; MacRae, T. P.; Suzuki, E. J Mol Biol 1979, 129, 463–481.
3. Prockop, D. J.; Kivirikko, K. I. Annu Rev Biochem 1995, 64, 403–434.
4. Engel, J.; Chen, H.-T.; Prockop, D. J. Biopolymers 1977, 16, 601–622.
5. Ramachandran, N. G.; Kartha, G. Nature 1955, 176, 593–594.
6. Ahmed, M.; Liu, Z. J.; Lukyanov, A. N.; Signoretti, S.; Horkan, C.; Monsky, W. L.; Torchilin, V. P.; Goldberg, S. N. Radiology 2005, 235, 469–477.
7. Byers, P. H. Disorders of Collagen Biosynthesis and Structure; McGraw-Hill: New York, 1989.
8. Byers, P. H.; Tsipouras, P.; Bonadio, J. F.; Starman, B.; Schwartz, R. C. Am J Hum Genet 1988, 42, 237–248.
9. Byers, P. H.; Steiner, R. D. Annu Rev Med 1992, 43, 269–282.
10. Byers, P. H. Connective Tissue and Its Heritable Disorders; Wiley-Liss: New York, 1993.
11. Bachiner, H. P.; Morris, N. P.; Davis, J. M. Am J Med Genet 1993, 45, 152–162.
12. Wang, Q.; Orrison, B. M.; Marini, J. C. J Biol Chem 1993, 268, 25162–25167.
13. Dalgleish, R. Nucleic Acids Res 1998, 26, 253–255.
14. Bodian, D. L.; Madhan, B.; Brodsky, B.; Klein, T. E. Biochemistry 2008, 47, 5424–5432.
15. Radmer, R. J.; Klein, T. E. Biochemistry 2004, 43, 5314–5323.
16. Bachiner, H. P.; Davis, J. M. Int J Biol Macromol 1991, 13, 152–156.
17. Mooney, S. D.; Huang, C. C.; Kollman, P. A.; Klein, T. E. Biopolymers 2001, 58, 347–353.
18. Beck, K.; Chan, V. C.; Shenoy, N.; Kirkpatrick, A.; Ramshaw, J. A. M.; Brodsky, B. Proc Natl Acad Sci USA 2000, 97, 4273–4278.
19. Ottl, J.; Battistuta, R.; Pieper, M.; Tschesche, H.; Bode, W.; Kuhn, K.; Moroder, L. FEBS Lett 1996, 398, 31–36.
20. Gauba, V.; Hartgerink, J. D. J Am Chem Soc 2008, 130, 7509–7515.
21. Gauba, V.; Hartgerink, J. D. J Am Chem Soc 2007, 129, 15034–15041.
22. Brooks, B. R.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187–217.
23. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L. J.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; III Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Karplus, M. J Phys Chem B 1998, 102, 3386–3616.
24. Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. Science 1994, 266, 75–81.
25. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. J Chem Phys 1983, 79, 926–935.
26. Brooks, C. L.; Karplus, M. J Mol Biol 1983, 208, 159–181.
27. Beveridge, D.; DiCapua, F. Annu Rev Biophys Chem 1989, 18, 431–492.
28. Straatsma, T. P.; McCammon, J. A. Annu Rev Phys Chem 1992, 43, 407–435.
29. Kollman, P. A. Chem Rev 1993, 93, 2395–2417.
30. Brooks, C. L.; Karplus, M.; Pettitt, B. M. Adv Chem Phys 1988, 71, 1–249.
31. Bevington, P. R. Data Reduction and Error Analysis for the Physical Sciences; McGraw-Hill: New York, 1992.
32. Swart, M.; Van Duijnen, P. T.; Snijders, J. G. J Comput Chem 2001, 22, 79–88.
33. Boresch, S.; Archontis, G.; Karplus, M. Proteins: Structure, Function and Genetics; 1994, 20, 25–33.
34. Boresch, S.; Karplus, M. J Mol Biol 1995, 254, 801–807.
35. Mark, A. E.; van Gunsteren, W. F. J Mol Biol 1994, 240, 167–176.
36. Brady, G. P.; Sharp, K. A. J Mol Biol 1995, 254, 77–85.
37. Bella, J.; Brodsky, B.; Berman, H. M. Structure 1995, 3, 893–906.
38. Bogaert, R.; Tiller, G. E.; Weis, M. A.; Gruber, H. E.; Rimoin, D. L.; Cohn, D. H.; Eyre, D. R. J Biol Chem 1992, 267, 22522–22526.
39. Vogel, B. E.; Doelz, R.; Kadler, K. E.; Hojima, Y.; Engel, J.; Prockop, D. J. J Biol Chem 1988, 263, 19249–19255.
40. Lightfoot, S. J.; Holmes, D. F.; Brass, A.; Grant, M. E.; Byers, P. H.; Kadler, K. E. J Biol Chem 1992, 267, 25521–25528.
41. Matsumura, M.; Becktel, W. J.; Levitt, M.; Matthews, B. W. Proc Natl Acad Sci USA 1989, 86, 6562–6566.
42. Perry, L. J.; Wetzel, R. Science 1984, 226, 555–557.
43. Starman, B. J.; Eyre, D.; Charbonneau, H.; Harrylock, M.; Weis, M. A.; Weiss, L.; Graham, J. M.; Byers, P. H. J Clin Invest 1989, 84, 1206–1214.
44. Gautieri, A.; Vesentini, S.; Redaelli, A.; Buehler, M. J. Protein Sci 2009, 18, 161–168.
45. Gautieri, A.; Uzel, S.; Vesentini, S.; Redaelli, A.; Buehler, M. J. Biophys J 2009, 97, 847–865.
46. Sasaki, N.; Odajima, S. J Biomech 1996, 29, 655–658.
47. Sun, Y. L.; Luo, Z. P.; Fertala, A.; An, K. N. J Biomech 2004, 37, 1665–1669.
48. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. Comp Phys Commun 1995, 91, 43–56.
49. Lindahl, E.; Hess, B.; van der Spoel, D. J Mol Mod 2001, 7, 306–317.
50. Raghunath, M.; Bruckner, P.; Steinmann, B. J Mol Biol 1994, 236, 940–949.
51. Traub, W.; Arad, T.; Vetter, U.; Weiner, S. Matrix Biol 1994, 14, 337–345.
52. Cohen-Solal, L.; Zylberberg, L.; Sangalli, A.; Lira, M. G.; Mottes, M. J Biol Chem 1994, 269, 14751–14758.

Reviewing Editor: David E. Wemmer