

# Free Energy Simulation to Investigate the Effect of Amino Acid Sequence Environment on the Severity of Osteogenesis Imperfecta by Glycine Mutations in Collagen

Kyung-Hoon Lee, Mark M. Banaszak Holl

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

Received 7 October 2010; revised 16 December 2010; accepted 4 January 2011

Published online 28 January 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.21593

## ABSTRACT:

Molecular dynamics simulations were carried out to calculate the free energy change difference of two collagen-like peptide models for Gly→Ser mutations causing two different osteogenesis imperfecta phenotypes. These simulations were performed to investigate the impact of local amino acid sequence environment adjacent to a mutation site on the stability of the collagen. The average free energy differences for a Gly→Ser mutant relative to a wild type are 3.4 kcal/mol and 8.2 kcal/mol for a nonlethal site and a lethal site, respectively. The free energy change differences of mutant containing two Ser residues relative to the wild type at the nonlethal and lethal mutation sites are 4.6 and 9.8 kcal/mol, respectively. Although electrostatic interactions stabilize mutants containing one or two Ser residues at both mutation sites, van der Waals interactions are of sufficient magnitude to cause a net destabilization. The presence of Gln and Arg near the mutation site, which contain large and polar side chains, provide more destabilization than amino acids containing small and nonpolar side chains. © 2011 Wiley Periodicals, Inc. *Biopolymers* 95: 401–409, 2011.

*Keywords:* collagen; osteogenesis imperfecta; free energy simulation; stochastic boundary molecular dynamics simulation; thermodynamic integration

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](mailto:biopolymers@wiley.com)

## INTRODUCTION

Collagen, a major matrix protein of skin, bone, tendon, and other tissues, is the most abundant protein in all mammals. It is composed of a heterotrimer containing two  $\alpha 1(I)$  and one  $\alpha 2(I)$  chains with over 1000 residues each. Type I collagen has 338 triplet repeating units of Gly-X-Y within each chain where Gly is placed at every third position as required for the formation of the tight helix and the close packing of the three chains. In vertebrate collagens, proline and 4-hydroxyproline (Hyp) are most frequently found at positions X and Y, respectively. A high content of Pro and Hyp is necessary for stabilizing the extended polyproline II-like helices. The helical domain is capable of self-association to form fibrils and is used for binding to integrins, proteoglycans, matrix metalloproteinases, and so on.<sup>1,2</sup> In this helical domain there are sequence variations to specify binding sites and regions for local microunfoldings and forming cooperative folding units.<sup>3,4</sup>

Mutation of collagen leads to heritable connective tissue disorders, such as osteogenesis imperfecta (OI),<sup>5</sup> which is characterized by a defective skeletal matrix and susceptibility to fractures.<sup>6–8</sup> The most common cause of OI is single point mutation from Gly to another residue such as Cys, Ser, Ala,

Correspondence to: Mark M. Banaszak Holl; e-mail: [mбанасза@umich.edu](mailto:mбанасза@umich.edu)

Contract grant sponsor: DARPA

Contract grant number: F017952-055766

© 2011 Wiley Periodicals, Inc.

Arg, Val, Glu and Asp. These mutations occur over the full range of the  $\alpha 1(I)$  and  $\alpha 2(I)$  chains of collagen and result in clinical phenotypes ranging from perinatal lethal to mild.<sup>9</sup> Important factors for the severity of OI include the type of  $\alpha 1(I)$  and  $\alpha 2(I)$  chains, the nature of amino acids substituted for Gly, amino acid sequence environment adjacent to a mutation site, position of the mutation with respect to C-terminus, and level of expression of mutant allele.<sup>9–11</sup> The effect of amino acid sequence is the subject of this study.

According to statistical analysis of the severity of OI, substitution of Asp for Gly is consistently lethal. Arg, Val, and Glu are lethal unless the mutation sites are near the N-terminus.<sup>12</sup> This observation indicates the C- to N-terminal folding direction of the collagen triple-helix;<sup>13</sup> however, the severity of phenotype for the mutation from Gly to Ala, Cys, and Ser does not follow this trend in  $\alpha 1(I)$  and  $\alpha 2(I)$  chains.<sup>12</sup> These amino acids show various phenotypes all along the chains and the clinical phenotypes of Gly→Cys and Gly→Ser mutations are determined by amino acid sequence environment adjacent to a mutation site.<sup>11,14,15</sup> Our previous stochastic boundary molecular dynamics (SBMD) simulations of collagen model peptides of [Ac-(Gly-Pro-Hyp)<sub>9</sub>-NH<sub>2</sub>]<sub>3</sub> were performed to calculate free energy change differences of mutants containing an Asp, a Cys, a Glu or a Ser residue relative to the wild type, which correlates with the clinical severity of OI from statistical analysis.<sup>16</sup> As expected based on the statistical analysis of OI, the free energy change differences of mutants bearing a Cys or a Ser residue relative to the wild type are smaller than those of the mutants containing a Glu and an Asp residue; however, the clinical phenotypes of the Gly→Cys and Gly→Ser mutations can be changed by the amino acid sequence environment adjacent to a mutation site.<sup>11,14,15</sup> One of the remarkable instances showing different phenotypes for the Gly→Ser mutation is found in human collagen at positions 901 and 913. In this case, nonlethal (901) and lethal (913) phenotypes are observed even when the type of mutation is the same and the two mutation sites are in close proximity.<sup>17,18</sup>

Homotrimeric triple helical collagen model peptides containing the nonlethal (901) and lethal (913) mutation sites for Gly→Ser mutation were synthesized to study the stability and folding of the peptides.<sup>19</sup> It was found that the Gly→Ser mutation at site 913 caused a significant decrease in thermal stability of the mutant relative to the wild type at site 901.<sup>19</sup> Although of interest, this study is limited by the nature of the homotrimeric peptide and the non-native addition of the four repeating units of Gly-Pro-Hyp at the C-terminus of the peptides. Furthermore, one amino acid mutation of the model peptide leads to mutation of three amino acids of the triple helix model because the peptides are homotrimeric.

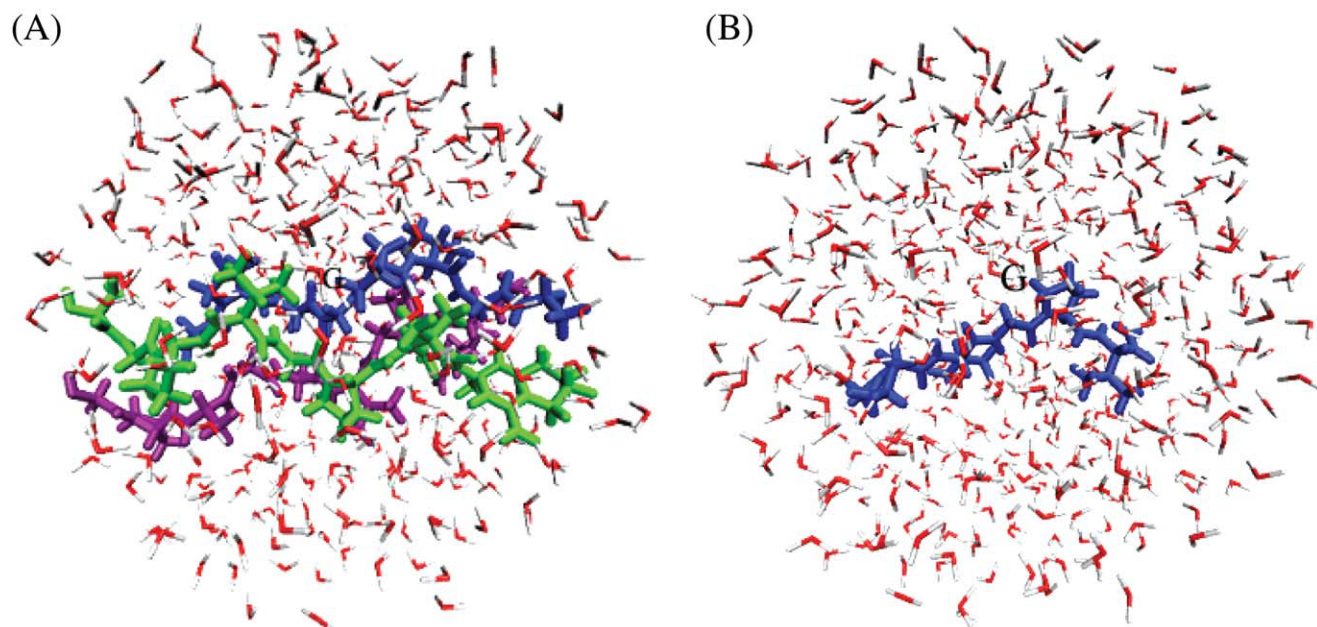
These experimental limitations can be overcome by a computational study on the triplet helical system because the heterotrimeric peptide systems can be easily built in silico. Mutations of one or two residues can be carried out to study the stability of the heterozygous and homozygous mutants relative to the wild type. Moreover, decomposition of the total free energy into each contribution from the potential energy function provides detailed analysis of the effect of mutations on the thermal stability of the system.

Herein we have performed SBMD simulations to calculate free energy change differences caused by the amino acid sequence environment surrounding the Gly→Ser mutation (see Figure 1). Two 30-residue heterotrimeric peptides containing a mutation site were built based on a human collagen sequence and a previously reported crystal structure of collagen-like peptide.<sup>20</sup> The peptides were prepared to include local regions of the  $\alpha 1(I)$  chain bearing mutation sites of the nonlethal Gly→Ser mutation at position 901<sup>17</sup> and the lethal Gly→Ser mutation at position 913.<sup>18</sup> We utilized standard free-energy simulation with thermodynamic integration and a coupling parameter approach to estimate free energy changes for each peptide system at the folded and unfolded states.

Our goal is to provide a microscopic understanding of the effect of amino acid sequence environment adjacent to a mutation site on their thermal stability of heterozygous and homozygous mutant collagens. These calculated results provide a better understanding of the details of free energy differences of the collagens caused by the single point mutation. The calculated free energy change differences of nonlethal and lethal mutants relative to the wild type are compared to the severity of OI mutants obtained from statistical analysis. These free energy change differences are also expected to agree with the trend in the previous free energy change differences from the simulations of the collagen models of [(Gly-Pro-Hyp)<sub>10</sub>]<sub>3</sub> for the mutations from Gly to Cys, Ser, Glu, and Asp.<sup>16</sup> The simulation method can be used to predict the severity of unknown mutations in OI and collagen-related diseases caused by amino acid sequence environment adjacent to a mutation site.

## METHODS

We carried out MD simulations for the mutation Gly → Ser at position 15 of two 30-residue collagen-like peptides for the folded state and for position 3 of two 5-residue monomeric extended peptides for the unfolded state. Amino acid sequences of folded and unfolded states from human collagen were employed. The 30-residue collagen sequences containing a mutation site at positions 901 and 913 (in bold) were selected for two collagen model peptide systems (a sequence for a peptide containing a mutation site at position 901: blue lines and a sequence for the other



**FIGURE 1** Collagen model peptides (wild type) containing a mutation site at position 901 solvated with a water sphere at the (A) folded and (B) unfolded states. The mutation site is indicated by G (for glycine).

peptide containing a mutation site at position 913: red lines in Figure 2). A mutation site is positioned at the center of each peptide (position 15) for the folded state because of the structural integrity of the central part of the peptides. Both peptide systems for the folded and unfolded states were acetylated at N-terminus and amidated at C-terminus. The peptide model for the folded state was built based on the sequence of human collagen and a crystal structure of a collagen-like peptide of [(Gly-Pro-Hyp)<sub>10</sub>]<sub>3</sub>

from the protein data bank (PDB code: 1CGD).<sup>20</sup> The structure and sequence of the collagen model peptides are shown in Figures 1 and 2, respectively.

A TIP3P<sup>21</sup> water sphere of 16 Å radius centered at the mutated residue Gly 15 H<sub>α</sub> was overlaid and water molecules overlapping the collagen-like peptide were deleted. The solvated wild type peptide systems for each simulation contained the following part of the peptide and water molecules.

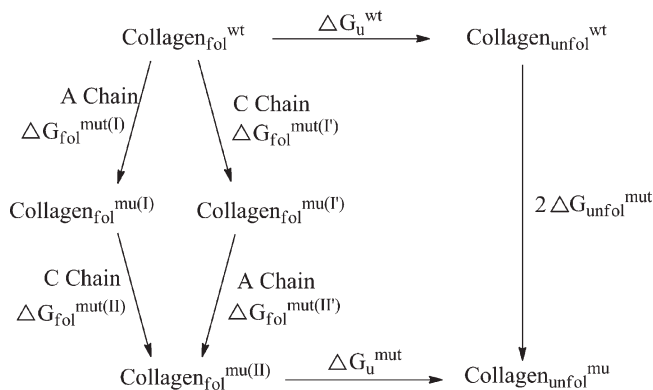
	Gly→Ser mutation at position 901	Gly→Ser mutation at position 913
Folded state		
Chain A: α1(I)	A <sup>11</sup> G <sup>12</sup> P <sup>13</sup> V <sup>14</sup> G <sup>15</sup> P <sup>16</sup> V <sup>17</sup> G <sup>18</sup> A <sup>19</sup>	A <sup>11</sup> G <sup>12</sup> P <sup>13</sup> Q <sup>14</sup> G <sup>15</sup> P <sup>16</sup> R <sup>17</sup> G <sup>18</sup> D <sup>19</sup>
Chain B: α2(I)	P <sup>13</sup> A <sup>14</sup> G <sup>15</sup> A <sup>16</sup> V <sup>17</sup> G <sup>18</sup> P <sup>19</sup> R <sup>20</sup> G <sup>21</sup>	P <sup>13</sup> Q <sup>14</sup> G <sup>15</sup> I <sup>16</sup> R <sup>17</sup> G <sup>18</sup> D <sup>19</sup> K <sup>20</sup> G <sup>21</sup>
Chain C: α1(I)	G <sup>12</sup> P <sup>13</sup> V <sup>14</sup> G <sup>15</sup> P <sup>16</sup> V <sup>17</sup> G <sup>18</sup> A <sup>19</sup> R <sup>20</sup>	G <sup>12</sup> P <sup>13</sup> Q <sup>14</sup> G <sup>15</sup> P <sup>16</sup> R <sup>17</sup> G <sup>18</sup> D <sup>19</sup> K <sup>20</sup>
Unfolded state	PVGPV	PQGPR

Each peptide system was minimized using the adopted basis Newton-Raphson<sup>22</sup> optimization algorithm (1000 steps). The MD simulations were carried out using the molecular simulation program CHARMM Version 35<sup>22</sup> and the CHARMM22<sup>23</sup> all-atom topology and parameters. 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. All the simulations on the minimized peptides were carried out by using the SBMD method.<sup>24</sup> The spherical dynamics reaction region contains the part of the peptide structure solvated with a water sphere of a radius of 14 Å centered on the H<sup>α</sup>

atom of a mutating residue. In this dynamics region all atoms performed Newtonian dynamics under the CHARMM force field. The dynamics region was surrounded by a 2-Å buffer region where the atoms carried out Langevin dynamics with a friction coefficient of 200 ps<sup>-1</sup>.<sup>25</sup> Water molecules were restrained to stay within 16 Å of the system center by a spherical mean-field potential.<sup>24</sup>

The calculated results ( $\Delta\Delta G_{\text{cal}}$ ) can be related to the experimental measurements (Eq. 2) in the thermodynamic cycle in Figure 3. The thermodynamic cycle for double mutations was setup in the same way as previously published.<sup>16,26</sup> The thermodynamic windowing method was used to calculate free energy changes of the system. Thermodynamic properties were estimated using the coupling





**FIGURE 3** Thermodynamic cycle used to calculate free energy differences of the first and second Gly→Ser mutations at the folded and unfolded states. The calculation ( $\Delta G_{\text{unfol}}^{\text{mut}}$ ) at the unfolded state was carried out once and used for each Gly mutation.  $\alpha'1(I)$ : Gly at the central position is replaced by Ser. Collagen<sub>fol</sub><sup>mu(I)</sup>:  $\alpha'1(I)\alpha1(I)\alpha2(I)$  (mutant containing a Ser residue on Chain A). Collagen<sub>fol</sub><sup>mu(I')</sup>:  $\alpha1(I)\alpha'1(I)\alpha2(I)$  (mutant containing a Ser residue on Chain C). Collagen<sub>fol</sub><sup>mu(II)</sup>:  $\alpha1(I)\alpha2(I)$  (mutant containing two Ser residues). Collagen<sub>unfol</sub><sup>wt</sup>: Ac-PVGPV-NH<sub>2</sub> at 901; Ac-PQGPR-NH<sub>2</sub> at 913. Collagen<sub>unfol</sub><sup>mu</sup>: Ac-PVSPV-NH<sub>2</sub> at 901; Ac-PQSPR-NH<sub>2</sub> at 913.

than the mutants containing two Ser residues at positions 901 and 913 by 4.6 and 9.8 kcal/mol, respectively. The free energy change differences ( $\Delta\Delta G_{\text{Total}}$ ) also demonstrate that the wild type is more stable than the mutant containing one Ser at positions 901 and 913 by an average 3.4 and 8.4 kcal/mol, respectively. The free energy change differences of the homozygous and heterozygous mutants at position 901 relative to the wild type are smaller than those of the corresponding mutants at position 913 by 4.8 and 5.0 kcal/mol, respectively. The average value of the free energy difference of the heterozygous mutants with a Ser at position 901 relative to the wild type is smaller than that of the heterozygous mutant at position 913 by 4.8 kcal/mol. These energy differences indicate that both homozygous and heterozygous mutants at position 901 are more stable than those at position 913.

The first and second Gly→Ser mutations destabilize both the folded and unfolded states but destabilization of the folded state is greater. This destabilizing trend is the same for both peptide systems. The destabilizing effect is larger for the 1st Gly mutation (by 9.7 kcal/mol on chain A and 6.7 kcal/mol on chain C) than the 2nd mutation (by 0.1 kcal/mol on chain C and 3.1 kcal/mol on chain A) for the mutant containing a Ser at position 913. The destabilizing effect is smaller for the 1st Gly mutation (by 2.2 kcal/mol on chain A and 4.5 kcal/mol on chain C) than the 2nd mutation (by 2.4 kcal/mol on chain C and 0.1 kcal/mol on chain A) for the mutant containing a Ser at position 901. Contributions to the free energy change difference from electrostatic terms

( $\Delta\Delta G_{\text{Elec}}$ ) stabilize the mutants containing one or two Ser residues relative to the wild type for the Gly→Ser mutations at both positions. However, van der Waals terms ( $\Delta\Delta G_{\text{vdW}}$ ) destabilize the mutants having one or two Ser residues with respect to the wild type at both mutation sites. For the folded and unfolded states of both collagen model peptides almost all of the covalent and noncovalent contributions disfavor the mutants containing one or two Ser residues relative to

**Table I** Contributions to Free Energy Change Differences ( $\Delta\Delta G$ ) for 1st and 2nd Gly→Ser Mutations at Positions 901 and 913 in the Collagen-like Peptides (kcal/mol)

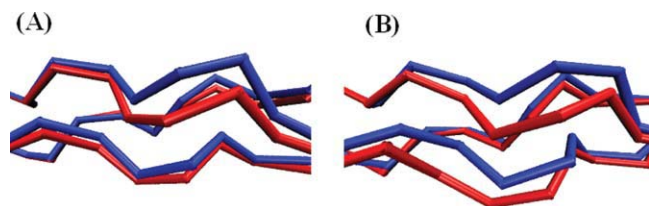
	Chain A	Chain C	A:C	C:A
<b>At position 901</b>				
Mutation			1st	2nd
$\Delta\Delta G_{\text{Total}}$	2.2	4.5	2.4	0.1
$\Delta\Delta G_{\text{Elec}}$	-3.6	-2.5	-2.4	-3.5
$\Delta\Delta G_{\text{vdW}}$	5.2	6.4	5.2	4.0
$\Delta\Delta G_{\text{Cov}}$	-0.3	-0.5	-0.3	-0.4
<b>Folded state</b>				
$\Delta G_{\text{Total}}^{\text{Fold}}$	11.5	13.8	11.7	9.4
$\Delta G_{\text{Elec}}^{\text{Fold}}$	-0.1	1.0	0.2	-0.9
$\Delta G_{\text{vdW}}^{\text{Fold}}$	5.0	6.2	5.0	3.8
$\Delta G_{\text{Cov}}^{\text{Fold}}$	6.6	6.4	6.3	6.5
<b>Unfolded state</b>				
$\Delta G_{\text{Total}}^{\text{Unfold}}$		9.3		
$\Delta G_{\text{Elec}}^{\text{Unfold}}$		2.6		
$\Delta G_{\text{vdW}}^{\text{Unfold}}$		-0.2		
$\Delta G_{\text{Cov}}^{\text{Unfold}}$		6.9		
<b>At position 913</b>				
Mutation			1st	2nd
$\Delta\Delta G_{\text{Total}}$	9.7	6.7	0.1	3.1
$\Delta\Delta G_{\text{Elec}}$	-1.8	-2.0	-2.0	-1.8
$\Delta\Delta G_{\text{vdW}}$	8.6	8.0	2.1	3.0
$\Delta\Delta G_{\text{Cov}}$	2.9	0.9	0.0	2.0
<b>Folded state</b>				
$\Delta G_{\text{Total}}^{\text{Fold}}$	19.6	16.6	10.0	13.0
$\Delta G_{\text{Elec}}^{\text{Fold}}$	0.6	0.4	0.4	0.6
$\Delta G_{\text{vdW}}^{\text{Fold}}$	9.4	8.5	2.9	3.8
$\Delta G_{\text{Cov}}^{\text{Fold}}$	9.6	7.6	6.7	8.7
<b>Unfolded state</b>				
$\Delta G_{\text{Total}}^{\text{Unfold}}$		9.9		
$\Delta G_{\text{Elec}}^{\text{Unfold}}$		2.4		
$\Delta G_{\text{vdW}}^{\text{Unfold}}$		0.8		
$\Delta G_{\text{Cov}}^{\text{Unfold}}$		6.7		

A:C—the first mutation on chain A and the second mutation on chain C.

C:A—the first mutation on chain C and the second mutation on chain A.

Average values of forward and reverse calculations from three independent simulations for each mutation are reported.

Statistical errors of  $\Delta\Delta G_{\text{Total}}$  for 1st and 2nd Gly mutations at positions 901 and 913 are  $\sim 0.5$  and  $0.4$  kcal/mol, respectively. Statistical errors of  $\Delta\Delta G_{\text{Elec}}$  for 1st and 2nd mutations at positions 901 and 913 are  $0.5$  and  $0.2$  kcal/mol, respectively. The errors of  $\Delta\Delta G_{\text{Cov}}$  for the corresponding positions are  $0.5$  and  $0.3$  kcal/mol, respectively. The errors of  $\Delta\Delta G_{\text{vdW}}$  for both positions are  $\sim 0.6$  kcal/mol.



**FIGURE 4** RMS deviations of backbone atoms of average structures (red) from the trajectories ( $\lambda = 0.9$ ) for the 1st Gly→Ser mutation (A) at position 901 in chain A and (B) at position 913 in chain A with respect to each corresponding minimized initial structure (blue).

the wild type; however, all the Gly→Ser mutations disfavor the systems at the folded states more than those at the unfolded states. The relative free energy change differences of heterozygous and homozygous Ser-containing mutants are greater for the peptide system at lethal mutation site than that at the nonlethal mutation site. The significant free energy differences between the systems are mainly from the 1st mutations rather than the 2nd mutations.

The root-mean-square deviations (RMSDs) of all the solvated backbone atoms (chain A: residue 11–19, chain B: residue 13–21, chain C: residue 12–20) of the average structures with respect to minimized initial structures were estimated from three independent MD trajectories ( $\lambda = 0.9$ ). The RMSDs for the Gly→Ser mutations at positions 901 and 913 are in the range of 0.6–0.8 Å and 0.9–1.4 Å, respectively. The RMSDs of the backbone atoms in each chain from the 1st Gly mutations at position 901 are 0.4–0.7 Å for chain A, 0.5–0.8 Å for chain B, and 0.6–1.0 Å for chain C; however, the corresponding RMSDs from the 1st mutations at position 913 are 1.0–1.2 Å for chain A, 0.9–1.7 Å for chain B, and 0.9–1.1 Å for chain C (see Figure 4). The RMSD of the backbone atoms of the system at the mutation position 901 are smaller than those of the system at the position 913.

Average dihedral angles ( $\phi$ ,  $\psi$ ) of residues near the mutation site, residues 14, 15, and 16 of the model peptide (residue 900, 901, and 902 in human collagen sequence), from the trajectories with  $\lambda = 0.9$  for the first and second Gly→Ser mutations exhibit  $\phi$  from  $-50^\circ$  to  $-71^\circ$  and  $\psi$  from  $132^\circ$  to  $161^\circ$ , respectively. These dihedral angles ( $\phi$ ,  $\psi$ ) are similar to the typical ( $\phi$ ,  $\psi$ ) angles found in polyproline II, ( $-78^\circ$ ,  $149^\circ$ ).<sup>30</sup> The average dihedral angles ( $\phi$ ,  $\psi$ ) of the corresponding residues from the trajectories with  $\lambda = 0.9$  for the first and second Gly→Ser mutations at position 913  $\phi$  are from  $-55^\circ$  to  $-80^\circ$  and  $\psi$  from  $134^\circ$  to  $163^\circ$ . However, the  $\psi$  values of the residue at positions 14 and 15 (residues 912 and 913 in human collagen sequence) for the 1st mutation on chain A are  $119^\circ$  and  $-159^\circ$ , respectively. These values are significantly deviated from the typical dihedral angle values

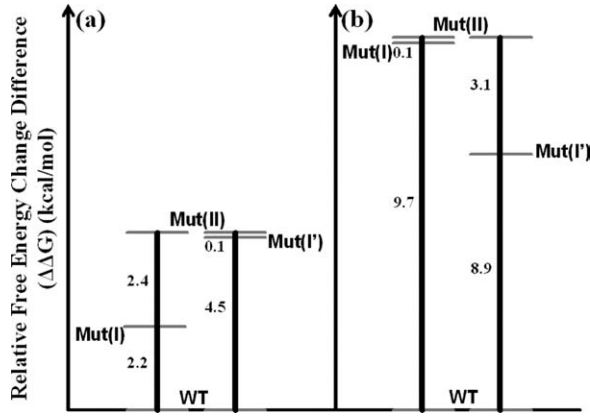
of polyproline II. The larger RMSDs of the backbone atoms for 1st Gly mutations at position 913 and the significant deviation of dihedral  $\psi$  angles of the 14th and 15th residues from the typical dihedral angles of polyproline II are caused by strong steric interactions between neighboring side chains at the mutation site.

## DISCUSSION

The common single point OI mutations in Type I collagen show various phenotypes ranging from mild to prenatally lethal ones. The single base substitutions lead to mutation of Gly by another amino acid, breaking the repeating sequence pattern of Gly-X-Y in the collagen triple helix. Statistical analysis of clinical phenotypes of OI has demonstrated that all Gly→Asp mutations reported in the  $\alpha 1(I)$  chain are lethal; Gly→Arg, Val and Glu are lethal unless the mutation sites are near the N-terminus; Gly→Ala, Cys and Ser are lethal or nonlethal all along the chains.<sup>12</sup> The gradient model suggests that mutations near the C-terminus are more severe than those near the N-terminus. Although this model supports the correlation for Gly→Arg and Val mutations in the  $\alpha 1(I)$  chain, the regional model emphasizes that local amino acid sequence environment and stability or binding sites are important in determining the severity of OI. This model has been proposed for the most common Gly→Cys and Ser mutations.<sup>31,32</sup>

To investigate the effect of amino acid sequence environment on the severity of OI, an amino acid sequence of human collagen from residue 887 to residue 928 was employed. Two collagen model peptides for Gly→Ser mutations were prepared: (1) a mild phenotype at position 901 and (2) a lethal phenotype at position 913. Our simulation results show that free energy change differences of the peptide systems for 1st Gly→Ser mutations at positions 901 and 913 are on average 3.4 and 8.2 kcal/mol, respectively, and that free energy differences of the mutants containing two Ser residues relative to the wild types are 4.6 and 9.8 kcal/mol at the corresponding positions (see Figure 5). The free energy change differences of heterozygous and homozygous mutants at position 901 are smaller than those at position 913 by  $\sim 5$  kcal/mol. The severe clinical phenotype of the mutant at position 913 is mainly caused by the 1st mutation. In the mutation at position 913 on chain A, the thermal stability of the heterozygous mutant relative to the wild type is similar to that of homozygous mutant.

The free energy change differences of first and second Gly→Ser mutations at position 901 on chain A are 2.2 and 2.4 kcal/mol, respectively. The relative destabilizing effects by the 1st and 2nd Gly mutations are almost identical. This



**FIGURE 5** Relative free energy change differences of the wild type and mutants containing one or two Ser residues of the collagen-like peptides. (Unit: kcal/mol). (a) at the nonlethal mutation site (at position 901), (b) at the lethal mutation site (at position 913). WT: Wild type, Mut(I) and Mut(I'): Mutant containing a Ser residue on chains A and C, respectively, and Mut(II): Mutant containing two Ser residues.

1st Mutation: WT→Mut(I) on chain A, WT→Mut(I') on chain C  
 2nd Mutation: Mut(I)→Mut(II) on chain A, Mut(I')→Mut(II) on chain C.

$$\begin{aligned}\Delta\Delta G_{\text{calc}} \text{ (1st mutation on chain A)} &= \Delta G_{\text{fol}}^{\text{mut(I)}} - \Delta G_{\text{unfol}}^{\text{mut}} \\ \Delta\Delta G_{\text{calc}} \text{ (1st mutation on chain C)} &= \Delta G_{\text{fol}}^{\text{mut(I')}} - \Delta G_{\text{unfol}}^{\text{mut}} \\ \Delta\Delta G_{\text{calc}} \text{ (2nd mutation on chain A)} &= \Delta G_{\text{fol}}^{\text{mut(II)}} - \Delta G_{\text{unfol}}^{\text{mut(I)}} \\ \Delta\Delta G_{\text{calc}} \text{ (2nd mutation on chain C)} &= \Delta G_{\text{fol}}^{\text{mut(II')}} - \Delta G_{\text{unfol}}^{\text{mut(I')}} \\ \Delta\Delta G_{\text{calc}} \text{ (1st and 2nd mutations)} &= \Delta G_{\text{fol}}^{\text{mut(I)}} + \Delta G_{\text{fol}}^{\text{mut(II)}} \\ &\quad - 2\Delta G_{\text{unfol}}^{\text{mut}} \\ &= \Delta G_{\text{fol}}^{\text{mut(I')}} + \Delta G_{\text{fol}}^{\text{mut(II')}} \\ &\quad - 2\Delta G_{\text{unfol}}^{\text{mut}}\end{aligned}$$

$$\begin{aligned}\text{At the folded state : } \Delta G_{\text{fol}}^{\text{mut(I)}} + \Delta G_{\text{fol}}^{\text{mut(II)}} &= \Delta G_{\text{fol}}^{\text{mut(I')}} \\ &\quad + \Delta G_{\text{fol}}^{\text{mut(II')}} \\ \Delta\Delta G_{\text{expt}} &= \Delta G_{\text{u}}^{\text{wt}} - \Delta G_{\text{u}}^{\text{mut}}\end{aligned} \quad (1)$$

$$\begin{aligned}\Delta(\Delta G) &= \Delta G_{\text{fol}}^{\text{mut(I)}} + \Delta G_{\text{fol}}^{\text{mut(II)}} - 2\Delta G_{\text{unfol}}^{\text{mut}} \\ &= \Delta G_{\text{fol}}^{\text{mut(I')}} + \Delta G_{\text{fol}}^{\text{mut(II')}} - 2\Delta G_{\text{unfol}}^{\text{mut}} \\ &= \Delta G_{\text{u}}^{\text{wt}} - \Delta G_{\text{u}}^{\text{mut}}\end{aligned} \quad (2)$$

$$\begin{aligned}\Delta G &= \int_0^1 \langle \Delta U \rangle_{\lambda} d\lambda \approx \sum_i \langle \Delta U \rangle_{\lambda_i} \Delta \lambda_i \\ \Delta U &= U_1 - U_0\end{aligned} \quad (3)$$

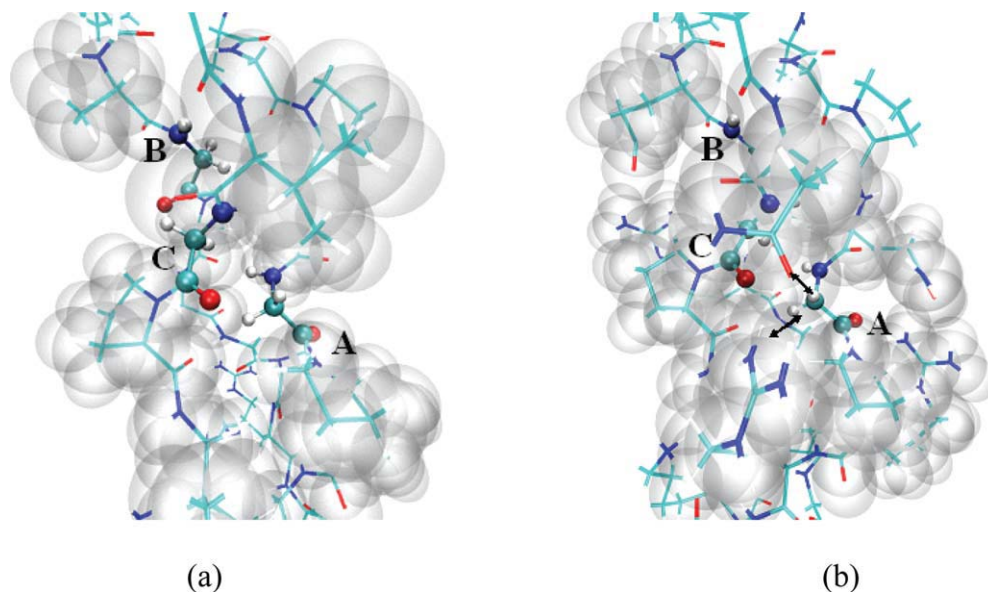
degree of destabilization caused by 1st and 2nd Gly mutations is different from that of previous simulations on the fully and partially solvated helical triple peptide systems composed of repeating units of (Gly-Pro-Hyp).<sup>26,16</sup> With this homotrimeric peptide model, the 1st mutated residues cause larger destabilizations of all the helical triple collagen-like peptides than the 2nd mutated residues. The heterotrimeric

collagen-like peptide model used in this study is more structurally flexible and is composed of more varied residue sequences than the homotrimeric model. The different degree of destabilizing effects by the 1st and 2nd mutations with respect to the wild type might be caused by a different type of heterotrimeric and homotrimeric triple helical model, structural robustness of the chains or amino acid sequence in the triple-helical peptides.

The different chemical and geometrical environments of the nonpolar or polar side chains of the residues near the mutation sites affect the nonlethal or lethal phenotypes of the Gly→Ser mutations. The polar and bulky side chains of the Gln and Arg residues at positions 14 and 17 on chain C are placed closely across the mutating residue at position 15 on chain A (position site in the model peptides) for the lethal mutation. By way of contrast, the nonpolar Val residues are placed at the corresponding positions on chain C for nonlethal mutation (see Figure 6). During the Gly→Ser mutation in these peptide systems, the mutated residue could interact with polar or bulky side chains of the residues more unfavorably than with side chains of small or nonpolar residues. These unfavorable van der Waals interactions between side chain of the mutating residue and the side chains of the amino acids, such as Gln, Pro, and Arg, near the mutation site more significantly decrease the stability of the lethal mutant for the Gly→Ser mutation.

The RMSD of the backbone atoms of the average structures for the 1st Gly mutation at position 913 is  $1.4 \pm 0.1 \text{ \AA}$  (see Figure 4), which is twice as large as that for the mutation at position 901 ( $0.7 \pm 0.1 \text{ \AA}$ ). The mutant containing a Ser at position 913 is less stable than that at position 901 by 7.2 kcal/mol. The dihedral angle deviation near the mutation site at position 913 also seems to be related to unfavorable contributions to relative free energy change difference of the mutant containing a Ser at position 913 relative to the wild type. The more significant structural deviations estimated from RMSD of the backbone atoms and dihedral angles near the mutation sites are found in the 1st Gly mutation at position 913.

We previously calculated relative free energy change differences of collagen-like peptides between the wild type and mutants containing Ser, Cys, Glu and Asp residues.<sup>16</sup> The relative free energy changes for 1st mutations from Gly to Ser, Cys, Glu and Asp were 3.8, 4.2, 5.6, and 8.8 kcal/mol, respectively. The corresponding relative free energy change differences of mutants containing two mutated residues with respect to the wild type were 5.1, 5.7, 8.5, and 14.2 kcal/mol, respectively.<sup>16</sup> The statistical analysis of OI mutations and relative free energy change differences demonstrated that the Gly→Ser and Cys mutations showed mild phenotypes and Gly→Asp exhibited a severe case.<sup>12,16</sup> In this study the rela-

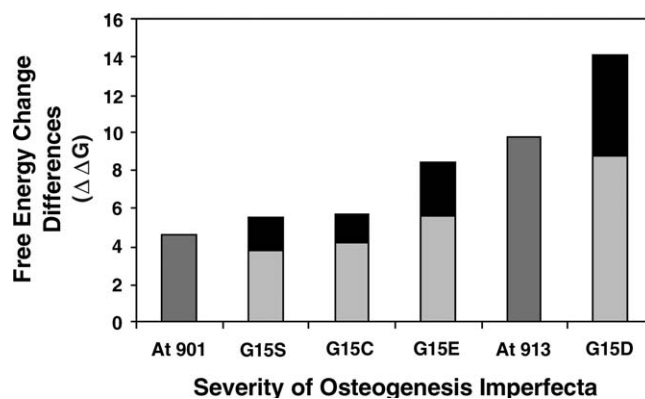


**FIGURE 6** The minimized collagen peptide models ( $\lambda = 0$ ) of (a) nonlethal and (b) lethal mutation sites. (Color: Red, blue, white and light blue indicate oxygen, nitrogen, hydrogen and carbon, respectively.) A, B and C indicate the chains of the model peptides. Three Gly's at position 15 in both peptide models are shown using CPK model and the residues at positions 14, 16, and 17 are shown with VDW model. The rest of the residues are shown with lines. The arrows in (b) indicate the unfavorable interactions between the side chains of the mutating residue on chain A and the Gln and Arg on chain C.

tive free energy change differences of the first and second Gly→Ser mutations at positions 901 and 913 correlate with mild and severe clinical phenotypes, respectively.<sup>17,18</sup> The mutants containing two Ser residues are less stable than their wild types by 4.6 kcal/mol at a nonlethal position and 9.8 kcal/mol at a lethal position (see Figure 5). The different clinical phenotypes of these mutations at different sites are consistent with the correlation of our calculated free energy change difference of the mutants relative to the wild type with the severity of OI from statistical analysis (see Figure 7).<sup>16</sup>

Mooney et al. previously carried out free energy simulations on a fully solvated collagen-like peptide for the mutation Gly→Ala to estimate thermal stability of the mutant relative to the wild type.<sup>26</sup> The folded and unfolded peptide systems were [Ac-(Gly-Pro-Hyp)<sub>10</sub>-NH<sub>2</sub>]<sub>3</sub>, and Ac-Pro-Hyp-Gly-Pro-Hyp-NH<sub>2</sub>, respectively where the central Gly was mutated to Ala. The simulation result was consistent with the experimental value.<sup>20</sup> We also carried out MD simulations on the same collagen-like peptide systems for the mutation Gly→Ala, but in a partially solvated water sphere using the SBMD method.<sup>16</sup> The simulation results with the partially solvated peptide models were consistent with the fully solvated systems, which validated the partially solvated collagen peptide model.<sup>16</sup> In this article we employ the same simulation

method with the partially solvated model to expedite the free energy simulations on our triple helical peptides. To achieve a more accurate free energy changes and realistic pictures of the peptide systems, we would better simulate the peptide systems in more extensive solvent systems without truncation or



**FIGURE 7** Correlations of the free energy change differences (kcal/mol) for the Gly→Ser mutations at positions 901 and 913 with the severity of OI. For comparison, previously estimated free energy change differences of Gly→Xaa mutations (Xaa: Ser, Cys, Glu and Asp) with the severity of OI (Lee et al.). A mutant containing two Cys residues is a reduced form. (Color: light grey, first Gly→Xaa mutations; black, average values of two second Gly→Xaa mutations; and grey, first and second Gly→Ser mutations at positions 901 and 913).



constraints. Longer simulations would improve statistical sampling of the system. High temperature simulation could improve the model for unfolded structure of the peptide model systems.<sup>33</sup>

## CONCLUSIONS

We carried out MD simulations of collagen-like peptides for the Gly→Ser mutations to investigate the effect of amino acid sequence environment on the severity of OI. The peptide sequences of the model peptides are from human collagen containing a mutation site at positions 901 and 913 for the Gly→Ser mutation, which show nonlethal and lethal phenotypes, respectively.<sup>17,18</sup> The Ser-containing nonlethal heterozygous and homozygous mutants are predicted to be more stable than the corresponding lethal mutants by 4.8 and 5.2 kcal/mol, respectively. The different clinical severity of OI for Gly→Ser mutations is consistent with the relative free energy differences caused by amino acid sequence environment adjacent to a mutation site. The polar and bulkier amino acids near the mutation site provide more destabilizing effect on the lethal mutant for the Gly→Ser mutations. The free energy simulation results for the Gly→Ser mutation of collagen-like peptide show that the lethality of the clinical phenotype of OI is in agreement with the severity of OI obtained from statistical analysis. The free energy change differences from the Gly→Ser mutations also agree with the trend in the severity of OI obtained on the collagen model system, [(Pro-Hyp-Gly)<sub>9</sub>]<sub>3</sub> for the Gly→Xaa mutations (Xaa: Ser, Cys, Glu, Asp).<sup>16</sup> Our MD simulation results indicate that the simulation method to calculate the relative free energy change differences of the homozygous or heterozygous mutants relative to the wild type with various clinical phenotypes can be used to estimate the degree of severity of OI and other collagen-related diseases.

The authors thank Dr. Charles L. Brooks III for allowing us to use his dual Quad-core Intel Xeon cluster and for commenting on this manuscript. The figures were prepared by using the VMD program.<sup>34</sup>

## REFERENCES

- Kadler, K. E. *Prot Profile* 1994, 1, 519–638.
- Fields, G. B. *Connect Tissue Res* 1995, 31, 235–243.
- Ryhänen, L.; Zaragoza, E.; Uitto, J. *Arch Biochem Biophys* 1983, 223, 562–571.
- Privalov, P. L. *Adv Protein Chem* 1982, 35, 1–104.
- 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.
- Byers, P. H. *Disorders of Collagen Biosynthesis and Structure*; McGraw-Hill Inc.: New York, 1989.
- Byers, P. H.; Tsipouras, P.; Bonadio, J. F.; Starman, B.; Schwartz, R. C. *Am J Hum Genet* 1988, 42, 237–248.
- Byers, P. H.; Steiner, R. D. *Annu Rev Med* 1992, 43, 269–282.
- Byers, P. H. *Connective Tissue and its Heritable Disorders*; Wiley-Liss: New York, 1993.
- Bachiner, H. P.; Morris, N. P.; Davis, J. M. *Am J Med Genet* 2005, 45, 152–162.
- Wang, Q.; Orrison, B. M.; Marini, J. C. *J Biol Chem* 1993, 268, 25162–25167.
- Dalgleish, R. *Nucleic Acids Res* 1998, 26, 253–255.
- Bonadio, J. F.; Byers, P. H. *Nature* 1985, 316.
- Marini, J. C.; Lewis, M. B.; Wang, Q.; Chen, K. J.; Orrison, B. M. *J Biol Chem* 1993, 268, 2667–2673.
- Marini, J. C.; Grange, D. K.; Gottesman, G. S.; Lewis, M. B.; Koeplin, D. A. *J Biol Chem* 1989, 264, 11893–11900.
- Lee, K.-H.; Kuczera, K.; Banaszak Holl, M. M. *Biopolymers* 2011, 95, 182–193.
- Mottes, M.; Sangalli, A.; Valli, M.; Lira, M. G.; Tenni, R.; Buttitta, P.; Pignatti, P. F.; Cetta, G. *Hum Genet* 1992, 89, 480–484.
- Cohn, D. H.; Wallis, G. A.; Zhang, X.; Byers, P. H. *Matrix* 1990, 10, 236–236.
- Yang, W.; Battineni, M. L.; Brodsky, B. *Biochemistry* 1997, 36, 6930–6935.
- Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. *Science* 1994, 266, 75–81.
- Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J Chem Phys* 1983, 79, 926–935.
- Brooks, B. R.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. *J Comp Chem* 1983, 4, 187–217.
- 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.; Wiorcikiewicz-Kuczera, J.; Karplus, M. *J Phys Chem B* 1998, 102, 3386–3616.
- Brooks, C. L., III; Karplus, M. *J Mol Biol* 1983, 208, 159–181.
- Beveridge, D.; DiCapua, F. *Annu Rev Biophys Chem* 1989, 18, 431–492.
- Mooney, S. D.; Huang, C. C.; Kollman, P. A.; Klein, T. E. *Biopolymers* 2001, 58, 347–353.
- Straatsma, T. P.; McCammon, J. A. *Annu Rev Phys Chem* 1992, 43, 407–435.
- Kollman, P. A. *Chem Rev* 1993, 93, 2395–2417.
- Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Adv Chem Phys* 1988, 71, 1–249.
- Bella, J.; Brodsky, B.; Berman, H. M. *Structure* 1995, 3, 893–906.
- Wenstrup, R.; Shrago-Howe, A.; Lever, L.; Phillips, C.; Byers, P.; Cohn, D. *J Biol Chem* 1991, 266, 2590–2594.
- Baum, J.; Brodsky, B. *Curr Opin Struct Biol* 1999, 9, 122–128.
- Brooks, C. L., III. *Curr Opin Struct Biol* 1998, 8, 222–226.
- Humphrey, W.; Dalke, A.; Schulten, K. *J Mol Graph* 1996, 14, 33–38.

Reviewing Editor: J. Andrew McCammon