Biophysical Chemistry 156 (2011) 146–152

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Effect of osteogenesis imperfecta mutations on free energy of collagen model peptides: A molecular dynamics simulation

Kyung-Hoon Lee [a,b], Krzysztof Kuczera [c], Mark M. Banaszak Holl [a,b,*]

[a] Michigan Nanotechnology Institute in Medicine and Biological Sciences, USA
[b] Department of Chemistry, University of Michigan, Ann Arbor, Michigan 49109, USA
[c] Department of Chemistry and Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66045, USA

ARTICLE INFO

Article history:
Received 19 February 2011
Received in revised form 22 March 2011
Accepted 31 March 2011
Available online 8 April 2011

Keywords:
Osteogenesis imperfecta
Collagen
Free energy simulation
Stochastic boundary molecular dynamics simulation
Thermodynamic integration

ABSTRACT

We have carried out stochastic boundary molecular dynamics simulations to estimate free energy changes for substitutions of Gly with Val, Arg and Trp residues in a collagen-like peptide. The relative free energy change differences of mutants containing a Val, an Arg and a Trp relative to the wild type are 5.7, 8.1 and 9.5 kcal/mol, respectively. The corresponding free energy change differences of mutants containing two mutated residues are on average 7.6, 10.5 and 14.7 kcal/mol, respectively. We show that the free energy change differences are correlated with the severity of OI from statistical analysis and mechanical properties of the individual tropocollagen molecules. This simulation result indicates an atomistic-level mechanistic understanding of the effect of OI mutations in terms of stability of the mutants relative to the wild type, which could eventually provide a new strategy for diagnosis and treatment of the disease.

1. Introduction

Collagen is the most abundant protein in mammals and provides the structural framework for blood vessels and most organs. Collagen type I is a heterotrimer composed of two α1(I) and one α2(I) chains. Each chain has a left-handed twist within a right-handed helical triplex for a monomer and consists of 338 repeating units of (Gly-X-X′), where Gly is placed at every third position of helical domain and is essential for close packing of triple helical structure [1–3]. Residues in positions X and X′ are related to the stabilization of the collagen fibrils. Post-translational modification of Pro and Lys residues can occur at residues in position X′ [4]. According to the structure of a collagen-like peptide determined by 'X-ray crystallography, the triple helical peptide exhibits supercoiling of polyproline II helices and interchain hydrogen bonding. The model peptide is a very rigid and robust system because two thirds of peptide sequence is composed of Pro and hydroxyproline (Hyp). Computational and experimental work of relative stability with model peptides indicate that mutations within a very stable local environment of the collagen result in more significant destabilizing effect on the structure, which may cause a more severe clinical phenotype of OI [5,6]. Single point mutations in the Gly position of a collagen helical triplex cause the majority of osteogenesis imperfecta (OI), a brittle-bone disease.

The severity of OI has been thought to be related to the structure and stability of the type I collagen caused by the mutations. But, it is not clear which factors are responsible for patient symptoms and the mechanism by which these changes cause patient phenotype. A change in the melting temperature (T_m) of collagen triple helix has been used as a measure of the destabilizing effect of mutations on the triple helix. Although the change of T_m could be expected to have direct physiological consequences due to decreased thermal stability of the mutated collagens, there is no simple correlation of clinical OI symptoms with the change of T_m caused by the mutations.

Single nucleotide replacements within Gly codons are the most common type of mutation. The nature of the amino acids substituted for Gly and type of chains and position of the mutation mainly determine phenotypes of the mutants caused by the single mutations in the triple helical coding regions of the type I collagen genes [7,8]. In the proα1(I) chain, substitutions of Val, Glu Arg and Asp produce moderate and severe phenotypes, respectively, near the carboxy-terminus of the triple helical peptides owing to the effect of large side chains [7]. The severity of similar substitutions within the proα2(I) chain appears to be milder because collagen is composed of two proα1(I) chains and one proα2(I) chain [7]. Substitutions of tryptophan for glycine are very rare, probably due to the rarity of GGG glycine codons in both *COL1A1* and *COL1A2* genes. The glycine substitutions change the triple helical structure, lower thermal stability and produce delay in triple-helix formation [9–11]. The

* Corresponding author at: 930 N. University Avenue, Department of Chemistry, University of Michigan, Ann Arbor MI 48109, USA. Tel.: +1 734 763 2283; fax: +1 734 763 3790.
E-mail address: mbanasza@umich.edu (M.M.B. Holl).

0301-4622/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2011.03.012

effect on collagen structure has been investigated by modeling studies and by experimental measurements using peptides with substitutions [9–11]. It was suggested that in the region of the mutation a bubble was formed, which could cause a different chain order [9,10] and there could be a change in the pitch of the helix [11].

The effect of OI mutations on the mechanical properties of a single tropocollagen molecule was investigated under a physiologically relevant mechanical loading condition [12]. The study showed that the severity of OI was correlated with the loss of the mechanical stiffness of individual tropocollagen molecules, which was controlled by side-chain volume and hydropathy index of the mutated residues [12]. Mutations exhibiting more severe phenotypes were associated with softer tropocollagen mechanical properties [12]. This indicated that the structural changes caused by OI mutations altered the properties at the single molecule level. Furthermore, mesoscale simulation showed that mutations leading to severe phenotypes exhibited weakened intermolecular adhesion, increased intermolecular spacing, and a reduced failure strength of collagen fibrils [13].

The previous MD simulations were carried out with a fully solvated collagen-like peptide composed of $[(\text{Pro-Hyp-Gly})_{10}]_3$ (Hyp: 4-hydroxyproline) for Gly→Ala mutation at the central position of the triple helical peptide. The free energy change difference between the mutant and the wild type was consistent with the experimental measurement [14]. We also previously performed stochastic boundary MD (SBMD) simulations on the collagen-like peptides to estimate free energy change differences between the wild type and mutants containing an X residue (X: Cys, Ser, Glu or Asp) as well as between the oxidized and reduced forms of the mutants containing two Cys residues. In that study we found that the calculated folding free energy differences were correlated with the severity of OI from statistical analysis [15]. Furthermore, the free energy differences of the oxidized and reduced form of the mutants containing two Cys residues showed that the oxidized form of the mutant is more stable than the reduced form by 0.8 to 2.7 kcal/mol [15], which was similar to the experimental values of free energy change differences (ΔΔG) between the oxidized and reduced form of cross-linked T4 lysozymes by a disulfide bond [16]. These simulation results also show that the oxidized form is more stable than the mutant containing one Cys residue but less stable than the wild type, which is in good agreement with experimental results obtained from three patients bearing collagen mutants containing Cys residues substituted for Gly.

In our previous simulations, we investigated several OI-related collagen Gly mutations, selected to represent varying levels of OI severity: severe (Asp), moderate (Glu) and mild (Ser and Cys). These preliminary studies showed that our free energy simulation protocol could correctly reproduce the observed OI severity effects. In this work we perform calculations for another set of mutations related to OI, involving mutations from Gly to Val, Arg and Trp in the collagen triple helix. These studies expand the range of previously modeled Gly substitutions to more observed residue types and extend the range of predicted changes in collagen stability. Together with the previous studies, the chosen mutations improve our coverage of the range of observed OI severities, covering multiple examples of mild, moderate and severe phenotypes. This provides a more stringent test of our computational approach and strengthens its predictive power for mutations not yet considered. Our results include calculations for the very rare Gly to Trp mutant, aimed at understanding the relation between its stability and OI severity. The Brodsky group experimentally estimated thermodynamic values to study stability of the mutants using the model peptides but did not study for the mutant containing a Trp [17] because Gly to Trp mutation maybe is extremely rare.

Herein we have carried out free energy simulations on partially solvated collagen model peptides to estimate stabilities of the mutants containing X (X = Val, Arg and Trp) residues relative to the wild type. The 1st and 2nd mutations were carried out in our free energy simulations to understand stabilities of the mutants containing one or two mutated residues, respectively. Our goal is to investigate the effect of the OI mutations on the stabilities of the mutants relative to the wild type and the correlation with the severity of OI and mechanical properties of the collagen-like molecules. We expect that these calculated results provide insight into the microscopic understanding of collagen's mutations on the thermal stability and a better understanding of the details of free energy differences of the collagens caused by a single point mutation.

2. Methods

We performed SBMD simulations on homotrimeric collagen model peptides with the simulation program CHARMM [18,19] version 35 and the CHARMM22 all-atom topology and parameters [20]. An atom-based 12.0 Å nonbonded cutoff distance was used with 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. For the folded peptide the initial structure is obtained from a crystal structure of the collagen-like model peptide (PDB code: 1CGD) [9]. The sequence of the folded peptide is $[\text{Ac-}(\text{POG})_9\text{-NH}_2]_3$ (O: 4-hydroxyproline) and the mutation site of the peptide is at position 15. The unfolded peptide model consists of an extended five-residue monomeric peptide (Ac-POGPO-NH_2). N-terminus and C-terminus are acetylated and amidated, respectively, for both folded and unfolded model peptides. The system is composed of spherical regions of 16 Å radius centered at Gly H_α. Hydrogen atoms were built in using the CHARMM program [18].

Each system was minimized by performing 5000 steps of the adopted basis Newton–Raphson (ABNR) optimization algorithm [18]. All simulations were carried out by using the SBMD simulations [21]. In this method, the system is divided into three regions: dynamics, buffer and reservoir regions. In the interior 14 Å sphere, called the dynamics region, all atoms perform normal Newtonian dynamics under the CHARMM force field [18]. In the 2 Å thick external shell, surrounding the dynamics region called a buffer region, stochastic dynamics based upon Langevin equation is used with a friction coefficient of 200 ps^{-1} [22]. All atoms except hydrogen atoms within the buffer region are placed under isotropic harmonic constraints with a force constant of 15 (kcal/mol $Å^2$). TIP3P water molecules [23] were restrained to stay within 16 Å of the system center by a spherical mean-field potential [21]. After deletion of the water molecules overlapped with the peptide, we carried out equilibration of solvent for 100 ps with the solute constrained. The reservoir region is the rest of the system, which is kept fixed all along the simulation.

The Gly→Val, Arg and Trp mutations were introduced in the folded and unfolded systems. The thermodynamic properties were calculated using the coupling parameter approach [22,24–26] with a hybrid potential energy function. The potential energy of the peptide system containing Gly at position 15 is defined as U_0 and that of the mutated system containing Val, Arg or Trp at the same position as U_1. MD simulations were performed for five intermediate systems with $U = (1-\lambda)\ U_0 + U_1$, and λ values of 0.1, 0.3, 0.5, 0.7 and 0.9. We carried out equilibration for 200 ps followed by a dynamics run for 1 ns for each window. The free energy between the initial and final states is evaluated by thermodynamic integration [22]. Each mutation simulation involved 6.3 ns of molecular dynamics for forward reaction and 1-ns of SBMD of the system required about 12 h of CPU time on the dual Quad-core Intel Xeon cluster. The simulation for each mutation was carried out in forward and reverse direction to estimate the degree of thermodynamic reversibility.

We have carried out MD simulations on collagen-like peptide for the mutation of Gly→Xaa (Xaa: Val, Arg, and Trp) at the folded and unfolded states. The simulation results produced the free-energy changes for the Gly→Xaa mutations at positions 15 of the folded state and 3 of the unfolded state. The free energy differences of chemical

mutation at the folded and unfolded states are ΔG_{fol}^{mut} and ΔG_{unfol}^{mut}, respectively. Experiment can measure the free energies of unfolding, ΔG_u^{wt} and ΔG_u^{mut}, of the wild type and mutant, respectively. The simulation results can be related to the experimental measurements via the thermodynamic cycle shown in Fig. 1.

$$\begin{aligned}\Delta(\Delta G)(\text{for first Gly}\rightarrow\text{Xaa mutations}) &= \Delta G_{fol}^{mut(I)} - \Delta G_{unfol}^{mut}\\ \Delta(\Delta G)(\text{for second Gly}\rightarrow\text{Xaa mutations}) &= \Delta G_{fol}^{mut(II)} - \Delta G_{unfol}^{mut}\\ \Delta(\Delta G)(\text{for first and second mutations}) &= \Delta G_u^{wt} - \Delta G_u^{mut}\\ &= \Delta G_{fol}^{mut(I)} + \Delta G_{fol}^{mut(II)} - 2\Delta G_{unfol}^{mut}.\end{aligned}$$

2.1. Error analysis and convergence

The errors of the free energy change differences in this simulation were estimated by the bin-average method [28]. For each simulation the data were divided into 20 contiguous bins of equal size within each window. The errors of the overall free energy difference changes were evaluated by standard error analysis with the reported errors corresponding 95% confidence intervals of the mean [28]. The statistical errors of the free energy difference changes for all the mutations from Gly to Val, Arg and Trp in this study were 0.7, 0.9 and 0.8 kcal/mol, respectively. For each mutation, the free energy changes from the separate forward and reverse trajectories agreed within their statistical errors. This indicates that the free energy simulation results are converged.

2.2. Approximations made in these MD simulations

Our simulations involve a number of simplifications and approximations since we employ a simple molecular model for folded and unfolded states. To limit the number of simulations we do not carry out all possible mutation patterns, only single mutations in the center of chain A and then second mutations in the centers of chains B and C. In the simulations we use only the solvated portion of the triple helix to represent the folded and unfolded states. Furthermore, in this SBMD simulation only part of the peptide model is solvated, allowed to fluctuate and interact with its surroundings with standard nonbonded cutoffs and without including counterions. This simulation setup does not allow any large structural changes in the triple helical peptide due to constraints. The limited solvation and use of nonbonded cutoffs do not allow for an accurate description of long-range electrostatic effects, which could introduce a bias in the calculated free energy changes. This effect could be more significant in the mutation Gly→Arg where electrostatic contributions are dominant. In this setup the simulation results involving a mutant containing a charged residue could be approximate. The simulation could be improved by using more water molecules, counterions, longer nonbonded cutoffs and a more accurate method for estimating electrostatistics such as multipole expansion [29]. We used five simulation windows with $\lambda = 0.1$, 0.3, 0.5, 0.7 and 0.9, with 6.2 ns of total MD for forward reaction in each energy simulation. In our study we decompose ΔΔG from alchemical processes in the thermodynamic cycle with all energy terms simultaneously scaled by the same value of the coupling parameter λ. Decomposition of the alchemical ΔΔG is appropriate for obtaining physical insights into the effects of point mutations on protein thermal stability [30]. In this analysis with our thermodynamic cycle, decomposing the alchemical ΔΔG for collagen complex formation is the appropriate approach for analyzing microscopic effects on triple-helix stability [30,31]. Our decomposition results exhibit local path dependence due to choice of uniform scaling. The uniform scaling results have the property of symmetric distribution of nondiagonal terms [30]. Thus, we employ the decompositions to provide helpful qualitative insights into the forces driving changes in stability caused by mutations. Lastly any predictions based upon empirical force fields are approximate due to limits in parameter accuracy.

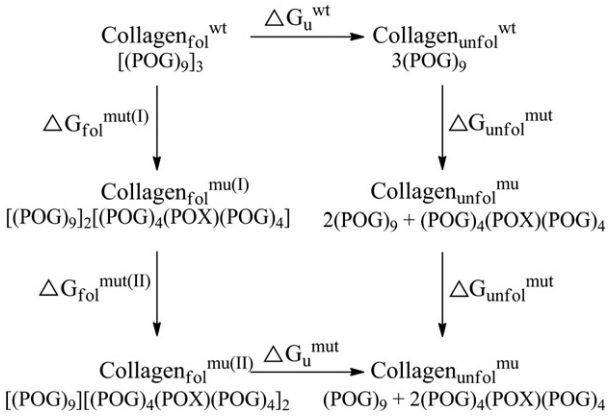

Fig. 1. Thermodynamic cycle used to calculate free energy differences of the first and second Gly→Xaa (Val, Arg and Trp) mutations at folded and unfolded states. The calculation at the unfolded state was determined once and used for each Gly mutation.

3. Results

The molecular structures of the homotrimeric collagen model peptides partially solvated in a water sphere at the folded and unfolded states are shown in Fig. 2. We have performed SBMD simulations on the solvated part of the peptides for the mutations from Gly to Val, Arg and Trp to understand the stability of the mutated collagen model peptides with respect to the wild type. The MD simulation results for the first and second Gly mutations to estimate the degree of destabilizing effect for the mutants in collagen type I are shown in Tables 1 and 2, respectively. The calculated results ($\Delta\Delta G_{Total}$) indicate that the wild type is more stable than the Val-, Arg- and Trp-containing mutants with two mutated residues by 7.6 ± 0.7, 10.5 ± 0.9 and 14.7 ± 0.8 kcal/mol respectively. The $\Delta\Delta G_{Total}$ values calculated in this work follow the same trends as the ones obtained using the same peptide system for the mutations from Gly to Cys, Ser, Glu and Asp, which were found to be closely related to the severity of OI from statistical analysis [15].

The free energy changes of the folded ($\Delta\Delta G_{Total}^{Fold}$) states for the first mutations Gly→Val, Gly→Arg and Gly→Trp are 22.9, −203.6 and 48.6 kcal/mol, respectively. The free energy changes of the unfolded ($\Delta\Delta G_{Total}^{Unfold}$) states for the corresponding first mutations are 17.5, −212.3 and 39.1 kcal/mol, respectively. The Gly→Val and Gly→Trp mutations destabilize both the systems at the folded and unfolded states, but they destabilize the system at the folded state more than those at the unfolded state. However, the mutation Gly→Arg stabilizes the system at the folded and unfolded states but it stabilizes the system at the unfolded state more than that at the folded state. Total free energy change difference is decomposed into covalent, electrostatic and van der Waals terms defined in the CHARMM potential energy function. The net effects ($\Delta\Delta G_{Total}$) for all of the mutations are to destabilize mutants relative to the wild type. The electrostatic, van der Waals and covalent terms all provide unfavorable contribution to 1st Gly→Val and Gly→Trp mutations. But, for 1st Gly→Arg mutation, electrostatic energy term contributes favorably to the total free energy change. The remarkably different contribution to the free energy change at the folded and unfolded states between Gly→Val & Trp and Gly→Arg mutations is from unfavorable and favorable electrostatic energy terms, respectively.

Due to different chemical and geometrical environment of the 2nd mutation sites, the 2nd mutations were carried out at two different positions. The total free energy change differences between the two chains are less than ~1 kcal/mol for all the mutations. The

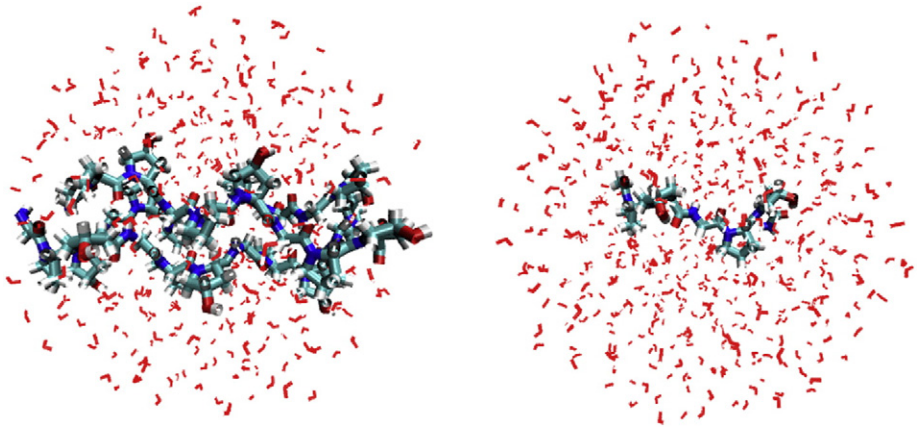

Fig. 2. The partially solvated triple helical peptide (wild type) and extended monomeric peptide (wild type) systems used as models for the folded (left) and unfolded (right) states, respectively in a water sphere (red). (Color: blue: nitrogen, red: oxygen, light blue: carbon, and white: hydrogen) The figures were prepared using the VMD program [27].

destabilizing effects of 2nd mutations for all of the mutations Gly→Val, Gly→Arg and Gly→Trp have shown the same trends as those of the 1st mutations but the magnitudes of the destabilizing effects for the 2nd mutations are smaller than those of 1st mutations. The Val-, Arg- and Trp-containing mutants from 2nd Gly mutations are less stable than the wild type by on average 2.0, 2.4 and 5.8 kcal/mol, respectively.

Statistical errors of $\Delta\Delta G_{Total}$ for the 1st and 2nd mutations Gly→Val, Arg and Trp are within 0.7, 0.9 and 0.8 kcal/mol, respectively.

The root mean square deviations (RMSDs) of solvated backbone atoms of average structures from the MD trajectories ($\lambda = 0.9$) for the mutations Gly→Val, Gly→Arg and Gly→Trp with respect to the minimized initial structures are shown in Table 3. The RMSD of the solvated backbone atoms for 1st mutation Gly→Trp is 1.6 Å, which is larger than those of the 1st mutations Gly→Val and Gly→Arg. Larger structural deviation of the mutants containing a Trp residue may be from its larger side chain. This larger RMSD of the peptide backbone of the Trp-containing mutant is also shown in Fig. 3. Structural deviations of the average structures with respect to the initial structures from the trajectories ($\lambda = 0.9$) for the 1st and 2nd Gly→Arg and Gly→Val mutations are identical.

We calculated average backbone dihedral angles from the MD trajectories ($\lambda = 0.9$) for the 1st mutations of Gly→Val, Gly→Arg and Gly→Trp. The means and standard deviations of the (φ and ψ) dihedral angle values near the mutation sites (residues 14, 15 and 16) are shown in Table 3. The average dihedral angle φ values near the mutation sites for the mutations from Gly to Val and Arg are −86° to −53° and −88°to −56°, respectively, and are similar to the typical value of $\varphi = -78°$ observed in polypropine II helices. The average dihedral angle ψ values near the mutation sites for the mutation from Gly to Trp are −63° to −37° with the dihedral angles at position 15 deviating the most from typical polyproline II values. The dihedral angle ψ values at positions 14, 15 and 16 for the mutations Gly→Val, Gly→Arg and Gly→Trp are 138°–155°, 143°–155° and 136°–151°, respectively, and are similar to the typical ψ value (149°) of polyproline II. The ψ value at position 15 for the mutation Gly→Trp is 135°, which deviates more from the ideal values than results for the mutations Gly→Val and Gly→Arg. The deviations of the dihedral angles for the mutations indicate introduction of structural deformations from the minimized initial structure during the MD simulations. The structural deformation observed in the value of RMSD of backbone atoms for the 1st mutation Gly→Trp may be related to more distorted backbone dihedral angle values at the mutation site of the Trp-containing mutant from the typical values of polyproline II, which may be also associated with more significant

Table 1
Contributions to free energy differences (ΔΔG) for 1st Gly→Xaa mutations at position 15 of the collagen-like peptides (Xaa: Val, Arg and Trp) (Unit: kcal/mol).

1st mutations	Gly→Val	Gly→Arg	Gly→Trp
$\Delta\Delta G_{Total}$	5.7	8.0	9.5
$\Delta\Delta G_{Elec}$	0.8	−1.3	0.3
$\Delta\Delta G_{vdW}$	3.4	4.6	0.2
$\Delta\Delta G_{Cov}$	2.0	4.7	8.9
Folded state			
ΔG^{Fold}_{Total}	22.9	−203.6	48.6
ΔG^{Fold}_{Elec}	−2.8	−257.2	−1.1
ΔG^{Fold}_{vdW}	1.5	−0.8	−0.7
ΔG^{Fold}_{Cov}	24.7	54.4	50.3
Unfolded state			
$\Delta G^{Unfold}_{Total}$	17.2	−211.6	39.1
ΔG^{Unfold}_{Elec}	−3.6	−255.9	−1.4
ΔG^{Unfold}_{vdW}	−1.9	−5.4	−0.9
ΔG^{Unfold}_{Cov}	22.7	49.7	41.4

Average values of forward and reverse calculations are reported.
Statistical errors of $\Delta\Delta G_{Total}$ for the 1st and 2nd mutations Gly→Val, Arg and Trp are within 0.7, 0.9 and 0.8 kcal/mol, respectively.

Table 2
Contributions to free energy differences (ΔΔG) for 2nd Gly→Xaa mutations (A and B) on two different chains in the collagen-like peptides (Xaa: Val, Arg and Trp) (unit: kcal/mol).

2nd mutations	Gly→Val		Gly→Arg		Gly→Trp	
	A	B	A	B	A	B
$\Delta\Delta G_{Total}$	1.8	2.1	1.5	3.2	5.4	6.1
$\Delta\Delta G_{Elec}$	0.7	1.2	1.1	0.1	0.2	0.6
$\Delta\Delta G_{vdW}$	0.1	−0.7	−0.2	1.5	−1.6	−1.1
$\Delta\Delta G_{Cov}$	1.0	1.6	0.6	1.7	6.9	6.6
Folded state						
ΔG^{Fold}_{Total}	19.0	19.3	−210.2	−208.5	43.9	44.6
ΔG^{Fold}_{Elec}	−2.9	−2.4	−254.9	−255.9	−1.4	−0.9
ΔG^{Fold}_{vdW}	−1.9	−2.6	−5.6	−4.0	−3.0	−2.5
ΔG^{Fold}_{Cov}	23.7	24.3	50.3	51.4	48.3	48.1
Unfolded state						
$\Delta G^{Unfold}_{Total}$	17.2		−211.6		38.5	
ΔG^{Unfold}_{Elec}	−3.6		−255.9		−1.5	
ΔG^{Unfold}_{vdW}	−1.9		−5.4		−1.4	
ΔG^{Unfold}_{Cov}	22.7		49.7		41.5	

Average values of forward and reverse calculations are reported.
2nd mutations are the average values at two different mutation sites.
Statistical errors of $\Delta\Delta G_{Total}$ for the 1st and 2nd mutations Gly→Val, Arg and Trp are within 0.7, 0.9 and 0.8 kcal/mol, respectively.

Table 3
RMSDs (root mean square deviations) of all of the solvated backbone atoms with respect to the minimized initial structures in each mutation from the MD trajectories with λ = 0.9. (Unit: Å).

	1st mutations	2nd mutations
Gly → Val	1.2	1.5
Gly → Arg	1.1	1.5
Gly → Trp	1.6	1.4

Fluctuations of the RMSDs for all the 1st and 2nd Gly mutations are 0.1 kcal/mol. The average values of 2nd Gly mutations on two chains are reported.

Table 4
Average dihedral angles (ψ, φ) near the mutation site of the folded peptide system in each mutation from the MD trajectories with λ = 0.9. (Unit: degree).

	1st mutations	Gly → Val	Gly → Arg	Gly → Trp
Residue 14	φ	138 ± 24	146 ± 42	151 ± 64
	ψ	−67 ± 10	−66 ± 9	−62 ± 11
Residue 15	φ	147 ± 36	143 ± 9	135 ± 10
	ψ	−86 ± 17	−88 ± 17	−37 ± 15
Residue 16	φ	155 ± 20	155 ± 41	136 ± 20
	ψ	−53 ± 12	−56 ± 12	−63 ± 14

decreasing effect on the stability of the Trp-containing mutant relative to the wild type (Table 4).

4. Discussion

The most common mutations found in type I collagen leading to OI are caused by single point mutations. The resulting phenotypes range from mild to prenatally lethal ones. The single point mutations of Gly by another amino acid break repeating sequence pattern of Gly-X-X′ in collagen triple helix. Statistical analysis of clinical phenotypes of OI has demonstrated that all Gly → Asp mutations observed in the α1(I) chain are lethal; Gly → Trp mutation is also severe but extremely rare; Gly → Arg, Val and Glu are lethal if the mutation sites are not found near the N-terminus; Phenotypes of Gly → Ala, Cys and Ser mutations are not dependent on the position of mutations. These mutations are lethal or unlethal all along the chains [32]. The gradient model emphasizes that mutations near the C-terminus are more severe than those near the N-terminus and that the folding process of the triple helical chains occurs from C- to N-terminus. This model supports lethality of single point mutations of Gly by Arg, Val and Glu in the α1(I) chain. However, the regional model suggests that local environment of the amino acids at the mutation site and stability and binding sites are crucial in the severity of OI. This regional model is proposed to explain the severity of single point mutations of Gly by Ala, Cys and Ser [10,33].

The free energy change differences of mutants with Xaa residues (Xaa: Val, Arg and Trp) relative to the wild type are in the trend of the correlations of the severity of OI from statistical analysis with the relative free energy change differences. This result is also consistent with our previous free energy change differences for the mutations from Gly to Xaa residues using the same simulation method but different mutants (Xaa: Cys, Ser, Glu and Asp) (Fig. 4) [15]. The free energy change differences of Gly → Xaa (Xaa: Ser, Cys, Glu and Asp) mutations between the wild type and the mutants containing one mutated residue at position 15 were 3.8, 4.2, 5.6 and 8.8 kcal/mol, respectively [15]. The corresponding free energy change differences between the wild type and the mutants bearing two mutated residues were 5.1, 5.7, 8.5 and 14.2 kcal/mol, respectively. The order of destabilizing effect of the residue replaced for Gly follows the order of severity of OI from statistical analysis. The relationship of the free energy change differences with the severity of OI provides insights into microscopic understanding of important factors determining the clinical severity of OI phenotypes.

Combining the results of this work and our previous study [15], the mutants containing one or two mutated Ala, Cys or Ser residue, exhibiting a mild observed phenotype, are less stable than the wild type by ~4 or ~6 kcal/mol, respectively. The mutants bearing one or two mutated Val, Arg or Glu residue, exhibiting a moderate observed phenotype, are less stable than the wild type by 6–8 or 8–11 kcal/mol, respectively. The mutants containing one or two mutated Asp or Trp residue, exhibiting a severe phenotype, are less stable than the wild type by 9–10 kcal/mol or 14–15 kcal/mol, respectively. Thus, these relative free energy changes for all of the mutants would help to predict the clinical severity of OI of unknown mutants from calculated estimates of free energy change differences.

Free energy simulations on the collagen-like peptides using SBMD method were previously employed to estimate free energy change differences of human collagen for the Gly → Ser mutations at positions 901 and 913 showing nonlethal and lethal phenotypes, respectively [34]. The free energy change differences of the two mutations in close proximity are in the trend of correlation of the simulated free energy change differences of single point Gly mutations with the severity of OI from statistical analysis. The relationship of our calculated free energy change differences with the clinical severity of OI could be used to determine the clinical phenotype of mutated collagen by calculating free energy change differences of the collagen mutants relative to the wild type using the SBMD method.

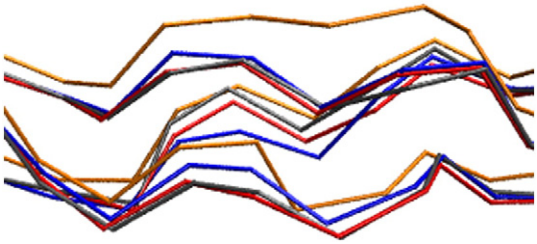

Fig. 3. Overlay of the backbone atoms of average structures from the trajectories (λ = 0.9) of each mutant containing one mutated residue. (Blue: wild type; red, gray and orange: mutants containing an Arg, a Val and a Trp residue, respectively) The figures were prepared using the VMD program [27].

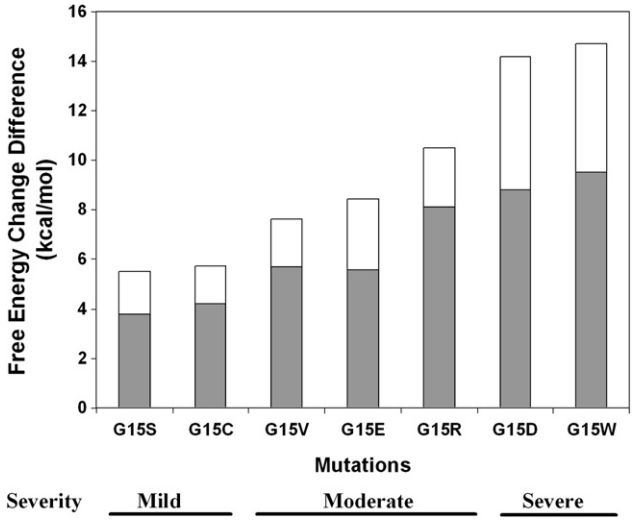

Fig. 4. Relationship of free energy change differences of the mutations and the severity of OI. (Color; gray: 1st mutations, white: 2nd mutations) (G15S, G15C, G15E and G15D mutations from Ref. [15]) The mutant containing two Cys residues is a reduced form. (S: Ser, C: Cys, V: Val, E: Glu, R: Arg, D: Asp, and W: Trp) The severity of OI is based upon the statistical occurrences and observations of OI. The statistical data are published [32] and collected in the collagen mutation database (http://www.le.ac.uk/genetics/collagen/).

The relative stability of host–guest collagen-like peptide mutants relative to the wild type was investigated to understand the influence of Gly mutations on the helix disruption in a constant and homogeneous environment [17]. In this experimental study the peptide mutants were prepared by substitutions of Gly by Xaa (Xaa: Ala, Cys, Ser, Val, Arg, Glu and Asp) residues in the peptide sequence of $[(\text{Gly-Hyp-Pro})_5\text{-}(\text{Xaa-Hyp-Pro})\text{-}(\text{Gly-Hyp-Pro})_4]_3$. Their experimental results showed that the order of disruption of the peptide correlated with the identity of the residues replacing Gly in the OI [17]. The thermodynamic analysis suggested that the decrease of stability of the peptides was due to entropic destabilization [17]. The relative clinical severity was consistent with the order of disruption of different residues in their collagen-like peptide. These experimental relative stabilities between the wild type and mutants of the collagen-like peptide were in approximate agreement with our calculated free energy change differences of the collagen model peptides.

The destabilization of the mutants is dependent upon side-chain size and charge; however, as both experimental data [17] and our simulations demonstrate, the correlation does not simply follow a side-chain volume series, even for residues of identical charge. The exception to the side-chain volume trend is aspartic acid for which the mutant is more unstable than the mutant having a glutamic acid even though the Glu side chain is larger than that of Asp. According to the simulation results, the total free energy change from electrostatic term in the glutamic acid-containing mutant contributes more favorably than that in the aspartic acid-containing mutant by 6 kcal/mol [15]. We expect this is because the glutamic acid side chain is more exposed to solvent than the aspartic acid side chain, providing the glutamic acid-containing mutant with more favorable electrostatic contribution to the total free energy change [15].

Gautieri et al. investigated the effect of Gly mutations in collagen leading to OI on the molecular, intermolecular and fibril mechanical properties [12,13]. The Gly mutations in OI were known to affect the mechanical properties of collagenous tissues [12,13]. The authors used atomistic steered MD simulations in explicit solvent to show that the severity of OI was correlated with the loss of the mechanical stiffness of individual tropocollagen molecules [12]. The values of Young's modulus were estimated to understand the peptides' stiffness as two functions of the hydropathy and the residue volume. The Young's modulus of the reference peptide was determined as 3.96 ± 0.21 GPa, which was consistent with the experimental measurement [35,36] and the values of mutated peptides for Gly were estimated to be 3.37 ± 0.32–3.78 ± 0.29 GPa [12]. This suggests an atomistic-level mechanistic understanding of OI mutations, which could lead to new strategies for diagnosis and treatment of the disease [12]. They also performed MD simulations to understand the effect of locally softened domains on the structural behavior of the tropocollagen models [13]. They found that mutations showing more severe phenotypes were related to softer tropocollagen mechanical properties [13]. They also found that at the intermolecular level, OI mutations caused a weakening of intermolecular adhesion, increase of intermolecular equilibrium spacing and reduction of likelihood of cross-link formation [13]. Through the mesoscale MD simulations, they suggested that the OI mutations caused reduction in strength of more than 50% and 35% in a cross-linked and a cross-linked free fibril, respectively, compared to a wild type [13].

The stability of the collagen model peptides tends to be reduced by single point mutations of Gly by other residues especially having charged and large side chains. The decrease of hydrophobic stabilization of the triple helix caused by Gly mutations affects the kinetics of the folding. This leads to slower folding rate of the mutants compared to the wild type [37]. This reduced folding rate of the mutants is associated with the severity of OI. The extent of the defective folding or altered folding mechanism related to decreased folding rate of the mutants may determine the degree of clinical severity of OI. Our results demonstrated that mutated collagen peptides exhibited a small decrease in stability or showed local conformational perturbation at an atomistic level. The small energetic and structural changes could be associated with local mechanical properties of single tropocollagen molecules [12].

5. Conclusion

Free energy simulations of collagen model peptides for the mutations from Gly to Val, Arg and Trp have been performed to investigate the effect of the single point mutations on the structure and stability of the collagen mutants related to OI. The free energy change differences of mutants containing a Val, an Arg and a Trp residues relative to the wild type are 5.7, 8.1 and 9.5 kcal/mol, respectively. The corresponding free energy change differences on the collagen mutants containing two mutated residues with respect to the wild type are 7.6, 10.5 and 14.7 kcal/mol, respectively. These free energy change differences for the mutations Gly → Val, Gly → Arg and Gly → Trp are caused by introduction of larger and/or charged side chains in place of small and neutral Gly. The Val-, Arg- and Trp-containing collagen mutants show relatively severe clinical phenotypes of OI. These free energy change differences of the mutants relative to the wild type are consistent with correlation of the free energy change differences with the severity of the other mutations obtained in our previous study using the SBMD simulation method. The larger structural deformation observed in the RMSD value of backbone atoms for 1st mutation Gly → Trp and more distorted backbone dihedral angles at the mutation site of the Trp-containing mutant may be associated with more significant destabilizing effect on the Trp-containing mutant.

The mutants bearing one or two residues (Asp or Trp) showing severe phenotypes are less stable than the wild type by 9–10 or 14–15 kcal/mol, respectively, and the mutants containing one or two residues (Val, Arg or Glu) showing moderate phenotypes are less stable than the wild type by 6–8 or 8–11 kcal/mol, respectively. According to our previous simulation study, the mutants containing one or two residues (Ala, Cys or Ser) showing mild phenotypes are less stable than the wild type by ~4 or ~6 kcal/mol, respectively. When the free energy change difference of unknown collagen mutants relative to the wild type is estimated to be greater than 13–14 kcal/mol through SBMD simulation, the mutant should show a severe phenotype. On the other hand, when the free energy difference of the unknown mutant is lower than 4–5 kcal/mol using the same simulation protocol, the phenotype of the mutant is predicted to be mild. These calculated relative energy differences are in agreement with the trend in the clinical severity of OI from statistical analysis in collagen type I and mechanical properties of tropocollagen molecules. This relationship of the free energy changes of the mutants and the clinical severity of OI would be helpful to predict the severity of unknown mutations. This work will improve our ability to predict the destabilizing effect on the stability of mutated collagens with respect to a wild type and to understand the clinical severity of OI and collagen-related diseases due to a single point mutation.

Acknowledgment

We thank Dr. Charles L. Brooks III for allowing us to use his dual Quad-core Intel Xeon cluster.

References

[1] A. Rich, F.H.C. Crick, The molecular structure of collagen, J. Mol. Biol. 3 (1961) 483–506.
[2] R.D.B. Fraser, T.P. MacRae, E. Suzuki, Chain conformation in the collagen molecule, J. Mol. Biol. 129 (1979) 463–481.
[3] D.J. Prockop, K.I. Kivirikko, Collagens: molecular biology diseases and potentials for therapy, Annu. Rev. Biochem. 64 (1995) 403–434.
[4] J. Engel, H.-T. Chen, D.J. Prockop, The triple helix coil conversion of collagen-like polytripeptides in aqueous and nonaqueous solvents. Comparison of the

thermodynamic parameters and the binding of water to (L-Pro-L-Pro-Gly)n and (L-Pro-L-Hyp-Gly)n, Biopolymers 16 (1977) 601–622.
[5] H.P. Bachiner, N.P. Morris, J.M. Davis, Thermal stability and folding of the collagen triple helix and the effects of mutations in osteogenesis imperfecta on the triple helix of type I collagen, Am. J. Med. Genet. 45 (1993) 152–162.
[6] H.P. Bachiner, J.M. Davis, Sequence specific thermal stability of the collagen triple helix, Int. J. Biol. Macromol. 13 (1991) 152–156.
[7] P.H. Byers, Osteogenesis imperfecta, in: P.M. Royce, B. Steinmann (Eds.), Connective Tissue and Its Heritable Disorders. Molecular, Genetic and Medical Aspects, Wileyp-Liss, New York, 1993, pp. 317–350.
[8] J.C. Marini, M.B. Lewis, Q. Wang, K.J. Chem, B.M. Orrison, Serine for glycine substitutions in type I collagen in two cases of type IV osteogenesis imperfecta (OI). Additional evidence for a regional model of OI pathophysiology, J. Biol. Chem. 268 (1993) 2667–2673.
[9] J. Bella, M. Eaton, B. Brodsky, H.M. Berman, Crystal and molecular structure of a collagen-like peptide at 1.9A resolution, Science 266 (1994) 75–81.
[10] J. Baum, B. Brodsky, Folding of peptide models of collagen and misfolding in disease, Curr. Opin. Struct. Biol. 9 (1999) 122–128.
[11] T.E. Klein, C.C. Huang, Computational investigations of structural changes resulting from point mutations in a collagen-like peptide, Biopolymers 49 (1999) 167–183.
[12] A. Gautieri, S. Vesentini, A. Redaelli, M.J. Buehler, Single molecule effects of osteogenesis imperfecta mutations in tropocollagen protein domains, Protein Sci. 18 (2009) 161–168.
[13] A. Gautieri, S. Uzel, S. Vesentini, A. Redaelli, M.J. Buehler, Molecular and mesoscale mechanisms of osteogenesis imperfecta disease in collagen fibrils, Biophys. J. 97 (2009) 847–865.
[14] S.D. Mooney, C.C. Huang, P.A. Kollman, T.E. Klein, Computed free energy differences between point mutations in a collagen-like peptide, Biopolymers 58 (2001) 347–353.
[15] K.-H. Lee, K. Kuczera, M.M. Banaszak Holl, The severity of osteogenesis imperfecta: a comparison to the relative free energy differences of collagen model peptides, Biopolymers 95 (2011) 182–193.
[16] M. Matsumura, W.J. Becktel, M. Levitt, B.W. Matthews, Stabilization of phage T4 lysozyme by engineered disulfide bonds, Proc. Natl. Acad. Sci. U. S. A. 86 (1989) 6562–6566.
[17] K. Beck, V.C. Chan, N. Shenoy, A. Kirkpatrick, J.A.M. Ramshaw, B. Brodsky, Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 4273–4278.
[18] B.R. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, M. Karplus, CHARMM: A program for macromolecular energy, minimization and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.
[19] B.R. Brooks, C.L. Brooks, A.D. Mackerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus, CHARMM: the biomolecular simulation program, J. Comput. Chem. 30 (2009) 1545–1614.
[20] A.D. MacKerell, D. Bashford, M. Bellott, R.L.J. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B 102 (1998) 3386–3616.
[21] C.L. Brooks III, M. Karplus, Solvent effects on protein motion and protein effects on solvent motion: dynamics of the active site region of lysozyme, J. Mol. Biol. 208 (1983) 159–181.
[22] D. Beveridge, F. DiCapua, Free energy via molecular simulation: applications to chemical and biomolecular systems, Annu. Rev. Biophys. Chem. 18 (1989) 431–492.
[23] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.
[24] T.P. Straatsma, J.A. McCammon, Computational alchemy, Annu. Rev. Phys. Chem. 43 (1992) 407–435.
[25] P.A. Kollman, Free energy calculations: applications to chemical and biochemical phenomena, Chem. Rev. 93 (1993) 2395–2417.
[26] C.L. Brooks III, M. Karplus, B.M. Pettitt, Proteins: a theoretical perspective of dynamics, structures and thermodynamics, Adv. Chem. Phys. 71 (1988) 1–249.
[27] W. Humphrey, A. Dalke, K. Schulten, VMD — visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.
[28] P.R. Bevington, Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill, New York, 1992.
[29] M. Swart, P.T. Van Duijnen, J.G. Snijders, A charge analyssis derived from an atomic multipole expansion, J. Comput. Chem. 22 (2001) 79–88.
[30] S. Boresch, M. Karplus, The meaning of component analysis: decomposition of the free energy in terms of specific interactions, J. Mol. Biol. 254 (1995) 801–807.
[31] S. Boresch, G. Archontis, M. Karplus, Free energy simulations: the meaning of the individual contributions from a component analysis, Proteins Struct. Funct. Genet. 20 (1994) 25–33.
[32] R. Dalgleish, The human collagen mutation database 1998, Nucleic Acids Res. 26 (1998) 253–255.
[33] R. Wenstrup, A. Shrago-Howe, L. Lever, C. Phillips, P. Byers, D. Cohn, The effects of different cysteine for glycine substitutions within alpha 2(I) chains. Evidence of distinct structural domains within the type I collagen triple helix, J. Biol. Chem. 266 (1991) 2590–2594.
[34] K.-H. Lee, M.M. Banaszak Holl, Free energy simulation to investigate the effect of amino acid sequence environment on the severity of osteogenesis imperfecta mutations in collagen. Biopolymers accepted for publication.
[35] N. Sasaki, S. Odajima, Stress–strain curve and Young's modulus of a collagen molecule as determined by the X-ray diffraction technique, J. Biomech. 29 (1996) 655–658.
[36] Y.L. Sun, Z.P. Luo, A. Fertala, K.N. An, Stretching type II collagen with optical tweezers, J. Biomech. 37 (2004) 1665–1669.
[37] M. Raghunath, P. Bruckner, B. Steinmann, Delayer triple helix formation of mutant collagen from patients with osteogenesis imperfecta, J. Mol. Biol. 236 (1994) 940–949.