# Molecular and Mesoscale Mechanisms of Osteogenesis Imperfecta Disease in Collagen Fibrils

Alfonso Gautieri,[†¶] Sebastien Uzel,[†‡] Simone Vesentini,[¶] Alberto Redaelli,[¶] and Markus J. Buehler[†§*]
[†]Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, [‡]Department of Mechanical Engineering, and [§]Center for Computational Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts; and [¶]Cellular and Molecular Biomechanics Research Group, Department of Bioengineering, Politecnico di Milano, 20133 Milan, Italy

ABSTRACT Osteogenesis imperfecta (OI) is a genetic disorder in collagen characterized by mechanically weakened tendon, fragile bones, skeletal deformities, and in severe cases, prenatal death. Although many studies have attempted to associate specific mutation types with phenotypic severity, the molecular and mesoscale mechanisms by which a single point mutation influences the mechanical behavior of tissues at multiple length scales remain unknown. We show by a hierarchy of full atomistic and mesoscale simulation that OI mutations severely compromise the mechanical properties of collagenous tissues at multiple scales, from single molecules to collagen fibrils. Mutations that lead to the most severe OI phenotype correlate with the strongest effects, leading to weakened intermolecular adhesion, increased intermolecular spacing, reduced stiffness, as well as a reduced failure strength of collagen fibrils. We find that these molecular-level changes lead to an alteration of the stress distribution in mutated collagen fibrils, causing the formation of stress concentrations that induce material failure via intermolecular slip. We believe that our findings provide insight into the microscopic mechanisms of this disease and lead to explanations of characteristic OI tissue features such as reduced mechanical strength and a lower cross-link density. Our study explains how single point mutations can control the breakdown of tissue at much larger length scales, a question of great relevance for a broad class of genetic diseases.

## INTRODUCTION

Collagen is a crucial structural protein material, formed through a hierarchical assembly of tropocollagen molecules, arranged in collagen fibrils that constitute the basis for larger-scale fibrils and fibers (Fig. 1 *a*). Diseases associated with collagen abnormalities have severe consequences, because collagens are major structural proteins within most human body tissues (e.g., tendon, skin, bone) that are crucial for the mechanical integrity of organisms. Among the first reported mutations in collagen were those associated with the so-called ''brittle bone disease'' or osteogenesis imperfecta (OI), which involves a catastrophic breakdown of tissue (1). Patients affected by OI exhibit an array of symptoms, including short stature, loose joints, blue sclerae, dentinogenesis imperfecta, hearing loss, and neurological and pulmonary complications (2,3). We specifically point out that in addition to the effects of OI mutations on bone tissue, there are also severe changes to mechanical properties of other connective tissues such as tendon.

The classification of OI is based commonly on clinical features that leads to four major groups, from mild (OI type I) to severe (OI type III and IV) to perinatal lethal (OI type II) (4). The genetic basis for ~90% of all known forms of this disease lies in mutations of type I collagen genes (5), as tabulated in the database of human collagen mutations (http://www.le.ac.uk/genetics/collagen) (6). Missense mutations that alter a glycine (Gly) codon in the genes encoding the characteristic collagen triple helix are the most common causes of OI (5). The replacement of either guanine (G) residue in the glycine codon GGC can theoretically result in the replacement of eight different amino acids: serine (Ser), cysteine (Cys), alanine (Ala), valine (Val), aspartic acid (Asp), glutamic acid (Glu), arginine (Arg), and tryptophan (Trp). All possibilities have been described in association with OI, although the frequency by which the different mutations occur varies considerably, with tryptophan replacements being exceedingly rare (7). Although a general correspondence between the specific mutations and the severity of OI has been reported, the molecular mechanisms of how a single point mutation can alter the mechanical properties at larger tissue scales remain unknown.

Earlier experimental studies have attempted to correlate glycine mutation types and locations with phenotypic severity. Some trends are apparent, such as OI severity increasing with an amino to carboxyl terminal orientation and with substitution by large and charged amino acids (8–10). At present, however, genotype-phenotype correlations are too weak to accurately predict the phenotypic effect of a particular glycine mutation. The answer to the question how a single point mutation in the tropocollagen molecule can cause failure of collagen based protein materials remains unknown. In particular, it remains unclear at what level in the tissue structure and how a single point mutation influences the behavior. To address these points, investigations at all relevant hierarchical levels must be carried out, beginning at the molecular level.

In this study, we report a series of systematic molecular scale based bottom-up computational experiments focused

*Submitted February 11, 2009, and accepted for publication April 28, 2009.*

*Correspondence: mbuehler@mit.edu

Editor: Elliot L. Elson.



0006-3495/09/08/0857/9 $2.00

doi: 10.1016/j.bpj.2009.04.059

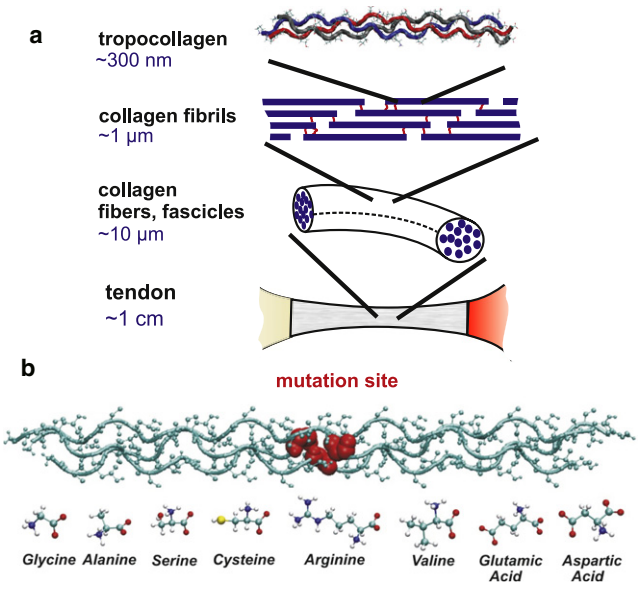


FIGURE 1 Hierarchical collagen structure from tropocollagen molecule to fiber level, and location of OI point mutations in the tropocollagen molecule. (*a*) Hierarchical structure of collagen and associated length scales; including: single tropocollagen molecule, collagen fibrils (staggered arrays of tropocollagen molecules), and fibers/fascicles. (*b*) Molecular geometry of the tropocollagen molecule, indicating the residue that is replaced in the mutation (*thick visualization*). All tropocollagen-like peptides considered in this study share the same structure, consisting of three identical chains made of Gly-Pro-Hyp triplets: $[(GPO)_5\text{-}(\mathbf{X}PO)\text{-}(GPO)_4]_3$. The **X** position of each chain (highlighted in *red* in the *upper part*) is one of seven replacing residues related to OI (*lower part*).

on pure collagenous tissue, carried out using atomistic-level molecular dynamics (MD), adaptive Poisson-Boltzmann solver (APBS) calculations, and a mesoscale molecular model of collagen fibrils. The basis for our study is a reference tropocollagen peptide and seven modified versions with seven types of glycine mutations, as shown in Fig. 1 *b*. The major goal of our analysis is to investigate the effect of these OI mutations on single molecule properties, changes to the lateral intermolecular interaction, as well as changes to mechanical properties of collagen fibrils. We also discuss possible effects on larger-scale properties and mineralization during bone formation. The major question to be addressed in this study is to elucidate how it is possible that a single point mutation at the level of a single tropocollagen molecule can lead to such dramatic, catastrophic effects at much larger tissue scales. We emphasize that the scope of this study is focused solely on pure collagenous tissues, and limited to studies of individual collagen fibrils.

## METHODS

We use a computational multiscale approach to investigate the effects of OI mutations at different levels. A variety of computational techniques are used including: full atomistic simulations for individual and pairs of molecules (in explicit water, using the GROMOS force field), Poisson-Boltzmann simulations, and MD for intermolecular interactions, as well as a coarse-grained bead model to describe the collagen fibril behavior. The results from single molecule and intermolecular properties are fed into the coarse-grained bead model in the spirit of a multiscale simulation approach.

### Molecular model setup

Triple helical tropocollagen molecule models with various sequences are created using the triple-helical collagen building script tool (THeBuScr) (11–13). All simulations are carried out with explicit solvent and the GROMOS96 43a1 force field as reported in earlier studies (13). We choose the simplest model of tropocollagen, with only Gly-Pro-Hyp (GPO) triplets on each of the three chains as the reference system (where Hyp and O are the three letter code and single letter code for the amino acid hydroxyproline, respectively). The central triplet is used to introduce the Gly replacement. The peptide structure is $[(GPO)_5\text{-}(\mathbf{X}PO)\text{-}(GPO)_4]_3$, where the **X** position is occupied by alanine, serine, cysteine, arginine, valine, glutamic acid, or aspartic acid (Fig. 1 *b*). For all peptides, the N-terminals and C-terminals are assumed neutral.

The tropocollagen models used in this study are truncated at 30 amino acids per chain due to computational limitations. This leads to short length tropocollagen segments with a length of ~8 nm. Peptides of comparable length have been used in earlier studies (8).

Further details in regard to the modeling approach, in particular the mechanical analysis and identification of Young's modulus, are included in the Methods section of the Supporting Material.

#### *Model equilibration*

MD simulations are carried out using the GROMACS code (14,15) and the GROMOS96 43a1 force field as used in earlier studies (13), which also includes parameters for the hydroxyproline (HYP) residue. The protein molecules are entirely solvated in a 13 nm × 3 nm × 3 nm periodic water box (that ensure a minimum distance of 0.8 nm between the protein and the box edge). Single point charge water molecules are used for the solvent, leading to a total of ~12,800 atoms for each system. SETTLE (for water) and LINCS algorithms are used to constrain covalent bond lengths involving hydrogen atoms, thus allowing a time step of 2 fs. Nonbonding interactions are computed using a cutoff for neighbor list at 1 nm, with a switching function between 0.8 and 0.9 nm for van der Waals interactions, whereas the particle-mesh Ewald sums method is applied to describe electrostatic interactions. In the case of charged peptides, counterions ($Cl^-$ or $Na^+$) are added to keep the system neutral. The preliminary system energy minimization is carried out by using a steepest descent algorithm until convergence. The system is then equilibrated at a temperature of 310 K (37°C) for 1200 ps of MD. The entire protein is held fixed for the first 200 ps by restraining the atomic positions, and thereafter only the first and the last $C_\alpha$ atoms of each chain are restrained for the following 1,000 ps. We observe that, even in the presence of large mutant amino acids, the RMSD of the proteins reaches a stable value within the 1 ns simulation time, suggesting that the collagen peptides are equilibrated properly.

### Modeling of intermolecular interactions

The lateral intermolecular interactions are calculated between one reference peptide and another peptide that can either be the reference or a mutated one. The combination of the reference peptide and the mutated peptide is chosen because in vivo mutations are not aligned in two neighboring collagen molecules due to the staggered geometry.

We move the second peptide in six different directions from a separation distance of 1.25 nm to a distance of 3 nm, where the distance is calculated as the distance between the center of mass of the proteins. The six directions have been chosen based on the quasi-hexagonal packing of collagen fibrils (16,17) (for the geometry, see Fig. 3 *a*).

The contributions to the total intermolecular interaction energy, the electrostatic and the van der Waals energies, are evaluated by two different

methods. The electrostatic forces are long ranged and thus the finite cut-off necessary in MD simulations would introduce large errors. Although the particle mesh Ewald method provides a reliable approach to treat the electrostatic energies during MD simulations, we choose the APBS (18) to obtain reliable estimates of the electrostatic interaction energies in aqueous solvent. The APBS method has been specifically developed and validated to accurately evaluate the electrostatic contributions to protein-protein binding (19). Therefore we use the APBS method that allow us to determine the electrostatic interaction profile as a function of the distance between two tropocollagen molecules. On the other hand, because the van der Waals interactions are short-ranged they can be accurately calculated through MD simulations in explicit solvent.

### *Electrostatic intermolecular interactions*

The APBS calculation involve, for each pair of peptides ($p$), for each orientation ($o$), and for each relative distance ($d$), the evaluation of the electrostatic energy of three systems: both single peptides ($C_a$ and $C_b$) as separate entities and the complex ($C_c$). The electrostatic interaction energy ($C_{int}$) is calculated as follows:

$$C_{int}(p,o,d) = C_c - (C_a + C_b). \quad (1)$$

Plotting $C_{int}$ obtained for the same peptide pair and direction but different distances yields the electrostatic energy profile as a function of the separation distance between the molecules. We use the standard CHARMM charge distribution at pH 7 to provide atomic-level charges. The equilibrated collagen models are converted from the PDB format to PQR format using the PDB2PQR code (20). We note that APBS does not take temperature dependent vibrations into consideration, and therefore we do not carry out an average over different configurations for the same distance, orientation, and peptide pairing (this is standard practice in APBS calculations of similar problems; see Vallejo et al. (19)). For all APBS calculations we use a coarse grid of 16.0 nm × 13.6 nm × 12.1 nm, a fine grid of 11.3 nm × 10.0 nm × 9.1 nm, and a number of grid point of 225 × 193 × 161. The grid dimensions are given by the PDB2PQR code, whereas the number of grid point is chosen to have a grid resolution of 0.5 Å in all directions.

### *van der Waals intermolecular interactions*

To quantify the van der Waals calculations we use the equilibrated peptides and follow a procedure similar to that described for the electrostatic interaction calculations above. For each peptide pair ($p$), orientation ($o$) and peptide-peptide distance ($d$), three systems are generated, the two single molecules and their complex. Each of these systems is entirely solvated in a 13.0 nm × 8.7 nm × 8.7 nm periodic water box (to ensure a minimum distance of 0.8 nm between the largest protein complex and the box edge).

To avoid "bad" contacts between proteins and solvent, all systems are subject to energy minimization by using a steepest descent algorithm until convergence is achieved. The systems are then equilibrated at a temperature of 310 K (37°C) for 200 ps of MD. The average value during the last 100 ps of the van der Waals energy of both the single peptides ($W_a$ and $W_b$) as separate entities and the value obtained for the complex ($W_c$) are then used to calculate the van der Waals interaction energy ($W_{int}$) as follows:

$$W_{int}(p,o,d) = W_c - (W_a + W_b). \quad (2)$$

### *Total intermolecular interaction energy*

The total lateral intermolecular interaction energy ($E_{int}$) between each peptide pair is then obtained as the sum of the electrostatic interaction energy ($C_{int}$) and the van der Waals interaction energy ($W_{int}$):

$$E_{int}(p,o,d) = C_{int}(p,o,d) + W_{int}(p,o,d), \quad (3)$$

where $p$, $o$, and $d$ are the different peptide pairs, the orientation in which the second peptide is displaced and the relative distance between the two proteins, respectively. The total interaction energies are then averaged over the six different directions to obtain, for each pair of peptides, the interaction energy profile as a function of the distance. These profiles are then fitted to a 12:6 Lennard-Jones (LJ) potential and the values of $\varepsilon$ (energy well depth), $\sigma$ (distance at which the potential is zero), and $d_{min}$ (distance between the particles at the energy minimum) are extracted for each peptide pair. The LJ potential is defined as $E(r) = 4\varepsilon((\sigma/r)^{12} - (\sigma/r)^{6})$. It is noted that given a certain peptide pair, the energy profile shows no significant difference between the energy profiles along different directions (where the different directions correspond to those sketched in Fig. 3 *a*, *top right*; following the hexagonal three-dimensional packing of molecules typically found in collagen fibrils).

## Intermolecular distance and probability of intermolecular cross-linking

In bone and tendon, the tropocollagen molecules form fibrils that are connected through lysine-based intermolecular cross-links. The model we use in this study is based on the discussion reported originally in Miles et al. (21). They suggested that the likelihood of cross-linking is related to the overlap volume of two adjoining tropocollagen molecules and that the increase distances between the molecules (in their case due to the lack of the $\alpha 2$ chain) would lead to a reduced cross-linking. In our case we have the opportunity to advance that model because we calculated the exact intermolecular distance from atomistic simulation (that is, the parameter $d$) for each mutation type. Therefore, if a mutation is close to a lysine residue that would be involved in formation of a cross-link, the increased spacing will affect the likelihood of cross-linking. It is noted that it is not the mutation itself that makes the cross-link, nor has the mutation to be exactly next to the lysine residue, but just close enough (~5–10 residue positions) to affect the local intermolecular distance. The core of the collagen molecule plus the lysine side chains can be considered as a sphere with radius $r$ (where the radius of the tropocollagen molecule is not affected by the mutations as shown in Table S4). For the cross-link reaction to occur, the end groups of two lysine residues must approach one another in space and be at the correct orientation with one another for the reaction to take place. This means that the bigger the overlap between two collagen spheres (that is, the closer the tropocollagen molecules are packed in the fibril), the greater is the probability of the cross-link forming (21). The overlap volume between two collagen spheres as a function of the distance is given by

$$V = 2\pi\left(\frac{2r-d}{2}\right)^2\left(r - \frac{2r-d}{6}\right), \quad (4)$$

where $r$ is the radius of the spheres and $d$ is the distance between the spheres.

## Coarse-grained mesoscale model of collagen fibrils

The mesoscale model of collagen fibrils is identical to that reported in Buehler (22,23), with modifications to reflect changes in the segment's stiffness and intermolecular interactions due to OI mutations. Numerical values of all model parameters are given in Table S3.

To model the effect of mutations, we alter the molecular and intermolecular properties in a segment of six beads (corresponding to the size of the structure investigated at the molecular scale, ≈8 nm), according to the findings reported in Figs. 2 and 3. We consider only the most severe case of OI mutations. To account for the reduction in molecular stiffness (Fig. 2), the spring constant that describes bead-bead stretching is reduced by 15%. The alteration of the intermolecular interaction leads to corresponding changes of $\varepsilon_{LJ}$ and $\sigma_{LJ}$ in Eq. S7. The value of $\sigma_{LJ}$ is chosen to be 20% larger than the reference value (to reflect increased molecular spacing) and $\varepsilon_{LJ}$ is chosen to be 10% of the reference value (to reflect a significantly reduced intermolecular adhesion by ≈90%; see Fig. 3).

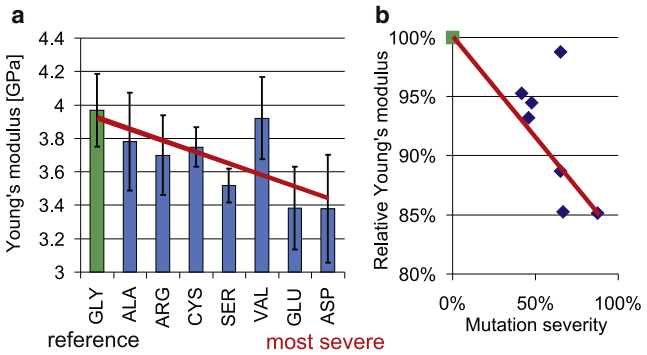


FIGURE 2 Reduction of Young's modulus of a single tropocollagen molecule, as a function of the (*a*) glycine replacement and as a function of (*b*) OI severity. In *a*, the elastic properties of a tropocollagen molecule are depicted as a function of the replacing amino acid residues, which are ordered based on the resulting disease severity (from physiological glycine [GLY] reference case to the most severe OI mutation [ASP]). (*b*) Young's modulus normalized by the reference glycine case, as a function of the mutation severity. More severe mutations tend to lead to a greater reduction of the elastic modulus, with greatest reduction of 15%. Data replotted from Gautieri et al. (13) and shown here for completeness.

## RESULTS AND DISCUSSION

### Single molecule effects

The Young's modulus of a single reference tropocollagen molecule is determined to be 3.96 ± 0.21 GPa, in agreement with corresponding experimental results (24,25). Molecules with glycine mutations display softer mechanical properties than the reference peptide, with values ranging from 3.37 ± 0.32 GPa to 3.78 ± 0.29 GPa. A decrease in Young's modulus up to 15% is observed, depending on the specific type of the replacing residue (Fig. 2 *a*). We note that these results have been reported in an earlier study (13) and are included in this study for completeness.

We further find that mutations related to more severe phenotypes are associated with softer tropocollagen mechanical properties, as illustrated in Fig. 2 *b* (see also Table S1). The softening behavior is confirmed by a linear fit to the Young's modulus results over the severity of the disease for the particular mutations, showing a clear trend.

### Intermolecular interaction effects

We proceed with a systematic investigation of intermolecular interactions, for all seven mutations as well as the reference system, by studying the adhesion energy profile as pairs of molecules are separated (see Fig. 3 *a*, *top left*; where the resulting curves are obtained as averages over different orientations as shown in Fig. 3 *a*, *top right*). The energy profiles presented in Fig. 3 *a* show that for short distances (<1.7 nm) there is generally a strong repulsion between the molecules, whereas for larger distances (>2 nm) the energy goes smoothly to zero. For intermediate distances (roughly between 1.9 nm and 2.2 nm) the energy curves display a minimum corresponding to the equilibrium intermolecular spacing. This energy minimum is in good agreement with the experimentally observed intermolecular distance between tropocollagen molecules in collagen fibrils (21,26).

We find that the intermolecular adhesion is strongly influenced by the type of mutation in the interacting tropocollagen molecules. When the reference tropocollagen molecule is coupled to an OI mutated peptide (Ala, Arg, Asp, Cys, Ser, Glu, or Val), the adhesion energies are reduced compared to the case when the reference peptide is interacting with another reference peptide without OI mutation. This behavior is reflected in the more shallow adhesion profile. In addition to the more shallow adhesion profile, the energy minimum is shifted to larger distances, leading to an increased intermolecular spacing.

A quantitative analysis of OI effects on intermolecular adhesion is carried out by plotting the depth of the energy minimum $\varepsilon$ (equivalent to the adhesion energy; Fig. 3 *b*) as well as the equilibrium distance $d_{min}$ between molecules (Fig. 3 *c*) as a function of the severity of each mutation type (see Table S2 for a list of numerical values of $\varepsilon$ and $d_{min}$). The adhesion energy $\varepsilon$ decreases from 2.8 kcal/mol (reference case) to 0.1 kcal/mol (most severe mutation), reflecting a significant reduction in intermolecular adhesion by >90%. As the severity increases, the value of $d_{min}$ increases by ~30% from ~1.95 nm to 2.5 nm, reflecting a larger intermolecular spacing. The finding that the intermolecular spacing in collagen fibrils increases due to OI mutations is in excellent agreement with the results found in osteogenesis imperfecta murine models, where an increase of ~0.1 nm in the lateral distance between tropocollagen molecules was observed in experimental in vivo studies (21,27). We note that these correlations do not necessarily imply causality. Nonetheless, our observations corroborate earlier results that OI mutations affect the structural properties of collagenous tissues.

The increased distance between tropocollagen molecules may affect the rate of intermolecular cross-linking. For the cross-link reaction to occur, the end groups of two lysine residues must come sufficiently close. Considering the collagen molecules as spheres (including the lysine side chain), then the larger the overlap zone (the closer the collagen molecules are packed in the fibril), the greater the probability is for a cross-link to form (21). Because OI mutations lead to an increase of the separation of the collagen molecules (Fig. 3 *c*), the volume of the overlap is reduced and thus the probability of the cross-link formation is severely compromised, as shown in Fig. 4. This mechanism could explain the lower level of cross-links observed in OI (28,29).

### Collagen fibril-level effects

To investigate the effect of molecular-level mutations on larger-scale mechanical behavior of collagen fibrils, we use

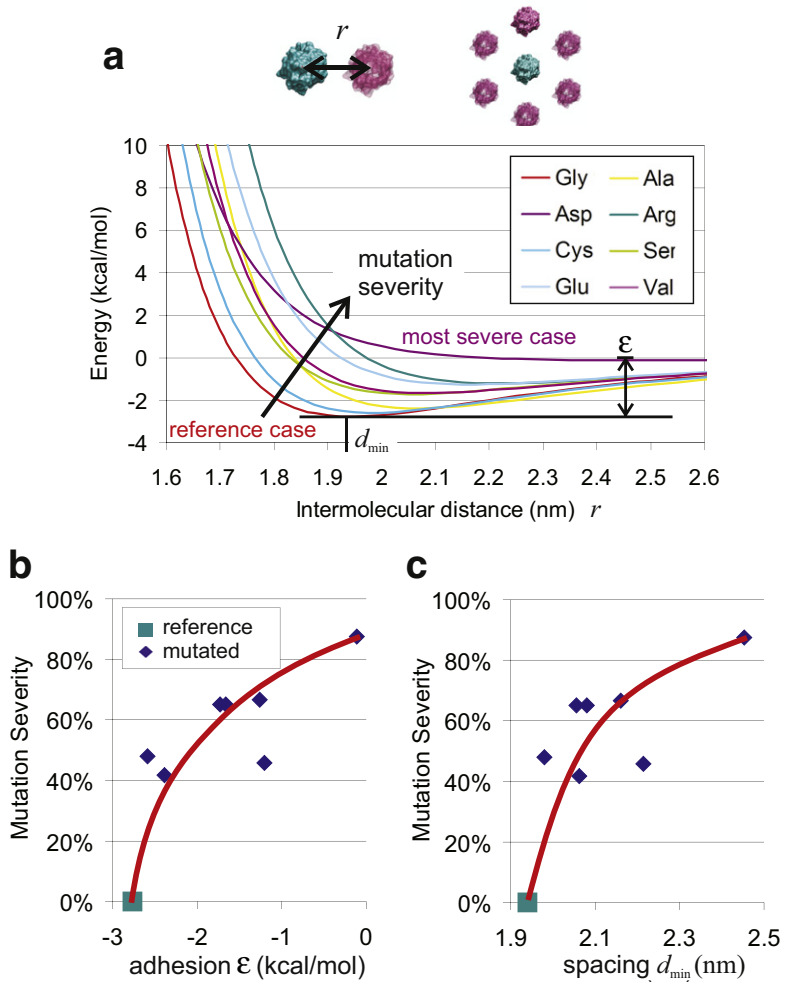


FIGURE 3 Change of intermolecular adhesion profile, adhesion energy $\varepsilon$ and intermolecular spacing $d_{min}$ as a function of OI severity. (*a*) Displays effective adhesion profiles for various mutations, including the reference case (the effective adhesion profile between pairs of molecules [*top left*] is obtained by averaging over all six directions in the hexagonal packing [*top right*]). As the mutations become more severe, the adhesion profiles display a more shallow profile with an equilibrium spacing shifted to larger values. (*b*) The intermolecular adhesion energy $\varepsilon$ plotted over OI severity, reflecting the reduction of adhesion as the mutations become more severe. (*c*) The intermolecular equilibrium spacing plotted over OI severity, reflecting the increase of intermolecular distance as the mutations become more severe. In each panel, the reference case (=absence of mutations) is depicted with a square.

a simple mesoscale model formulation that describes the mechanical behavior of a single collagen fibril, where the nanomechanical characterization of molecular-scale behavior serves as input parameter into the model formulation. This method follows a modeling approach reported earlier in Buehler (22,23) with relevant modifications to reflect the effect of tropocollagen mutations on molecular and intermolecular properties. The mesoscale model provides us with the ability to directly study the properties of collagen fibrils (at scales of larger than micrometers) due to molecular-level, single point mutations (at scales of subnanometers).

We find that the presence of mutations leads to a significant change in the mechanical response of collagen fibrils. Fig. 5 *a* shows the relative change of strength, revealing a loss in strength of more than 35% in a cross-link free fibril. The loss of strength is even greater in a cross-linked fibril (one cross-link at each tropocollagen molecule end), leading to a reduction of almost 50% compared to its reference value. We also find that the effect of the mutation is strongest if the mutation is placed at either end of the molecule, and least strong if it is placed in the center. We note that the strength of cross-link free fibril is significantly reduced compared with those that feature cross-links; nevertheless, it is finite (23). The reason for the finite strength of crosslink free collagen fibrils is that there exist intermolecular adhesion forces, which provide adhesion strength in the molecular assembly. We point out, however, that the strength is many times greater once cross-links are introduced, which qualitatively agrees with experimental studies (30).

A direct quantitative comparison with experimental results of reference and mutated collagen fibrils is not yet feasible, because a systematic comparison of the mechanical properties of reference and mutated collagen fibrils has not yet been carried out. However, a qualitative comparison with experimental results (31) in osteogenesis imperfecta murine tendon shows some agreement for the loss of strength in mutated collagen tissue; as shown in Fig. 5 *a* (albeit we emphasize the experimental measurements were carried out at a much larger tissue scale).

Fig. 5 *b* shows two sample stress-strain curves, illustrating that the presence of OI mutations strongly influences the mechanical response. In addition to the changes of critical strain for the onset of permanent deformation through intermolecular slip and the strength, the elastic modulus is

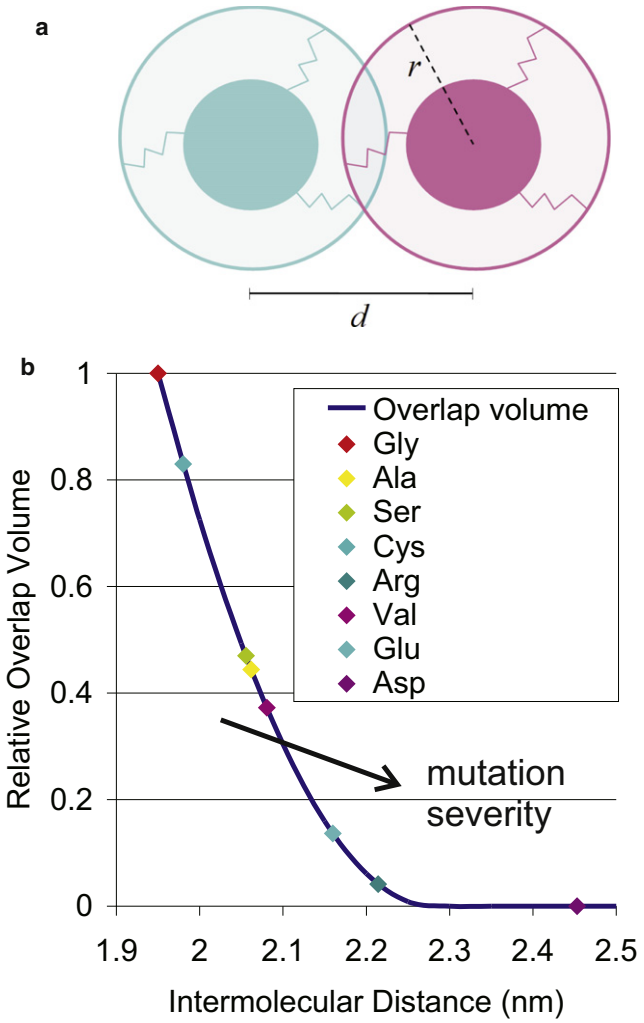


FIGURE 4 Effect of increased intermolecular distance reduces the probability of intermolecular cross-linking in collagen fibrils. (*a*) The geometry of cross-linking tropocollagen molecules. The core of the collagen molecule plus the lysine side chains can be considered as a sphere with radius $r$. For the cross-link reaction to occur, the end groups of the two lysine residues must approach one another in space and be at the right orientation with one another for the reaction to take place. This means that the bigger the overlap between two collagen spheres (that is, the closer the tropocollagen molecules are packed in the fibril), the greater is the probability of the cross-link forming (21). (*b*) The overlap volume (relative to the reference case, Gly) between two collagen spheres as a function of the distance between the spheres (*dark line*). The overlap volume is given by Eq. S3. We use a value of $r = 1.14$ nm, which is given by the sum of the radius of the collagen core (0.5 nm; see Fratzl and Weinkamer (26)) plus the length of the lysine side chain (0.64 nm; see Miles et al. (21)). The plot is then populated with the relative overlap volumes calculated in the presence of each specific OI mutation. In this case we use as intermolecular equilibrium distance $d_{min}$ (see Table S2). The presence of OI mutations, by increasing the intermolecular distance between tropocollagen molecules, considerably reduces the volume of the overlap zone, and thus the probability of the cross-link formation. This mechanism could explain the lower level of cross-links observed in OI (28,29).

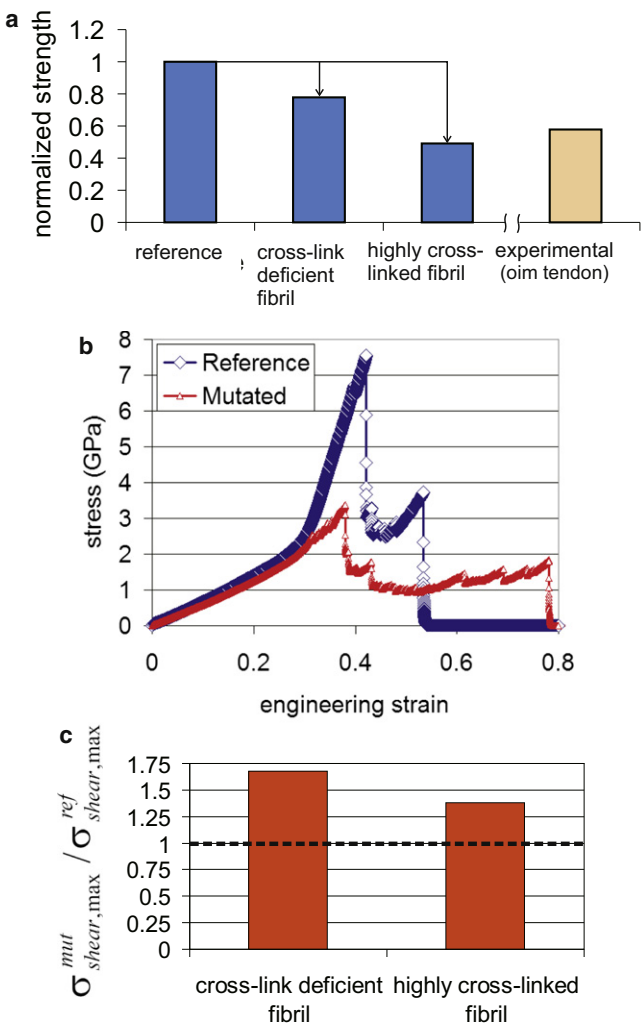


FIGURE 5 Influence of OI mutations on the mechanical properties of a collagen fibril, leading to a significant reduction of mechanical strength and yield strain (for most severe mutation, located at the end of the molecule). (*a*) The decrease of the strength for a mutation (in a cross-link deficient fibril and highly cross-linked fibril), including a qualitative comparison with experimental results (31) in OI tendon (note the difference in scale; our studies are focused on single collagen fibrils; the experiments have been carried out at much larger tendon tissue scales). (*b*) The stress-strain curve for the reference case and the mutated case, showing that the presence of mutations can severely influence the overall mechanical signature (highly cross-linked fibril). Intermolecular sliding sets in at 30% vs. 42% in the mutated fibril, and the maximum stress is significantly reduced (to ~50% of its reference value). The small-deformation elastic modulus is reduced by ~15% under the presence of the OI mutation. (*c*) The ratio of the maximum shear stress in the mutated versus the nonmutated system, based on the average shear stress along tropocollagen molecules, $\sigma^{mut}_{shear,max}/\sigma^{ref}_{shear,max}$ for both the cross-link deficient fibril and highly cross-linked fibril (the stress is evaluated at identical strain levels immediately before the onset of failure). The average maximum shear stress is 68% and 38% higher for the cross-link deficient fibril and highly cross-linked fibril, respectively, causing fibril failure via intermolecular slip.

reduced by ~15% for small deformation elasticity due to the presence of the mutation. Through an analysis of the stress distribution within the collagen fibril we have observed that mutated fibrils include large local peaks of stresses; suggesting that the presence of mutations leads to an increase of the local stress that is several times larger than the applied

stress. Fig. 5 *c* shows the ratio of the maximum shear stress in the mutated versus the nonmutated system, based on the average maximum shear stress along the tropocollagen molecule, $\sigma^{mut}_{shear,max}/\sigma^{ref}_{shear,max}$ for both the cross-link deficient fibril and highly cross-linked fibril. The maximum shear stress is 68% and 38% higher for the cross-link deficient fibril and the highly cross-linked fibril, respectively. The occurrence of large local stresses might explain the reduced failure strength, as intermolecular bonds are broken more easily at locations of stress concentrations. The concept of local stress concentrations is similar to the role of flaws and cracks in materials; where the corners of these defects lead to a magnification of the stress that may induce catastrophic failure at the macroscale (even though the actual size of the defect may be very small) (32,33). The situation in collagen fibrils is similar because the mutated segments feature a reduced adhesion and larger spacing, resembling a crack-like nanoscale inclusion in the material.

Even though our molecular simulation results provide molecular-level insight into possible mechanism of OI related reduction in strength of collagen fibrils, the agreement with experiments at the tendon level is only qualitative in terms of the actual fracture strength and strain. A quantitative comparison of failure stress shows that the mesoscale model predicts much higher stress levels (5–10 times higher) in collagen fibrils than the experimentally observed values in tendon. This discrepancy could be due to the limitations and approximations of the mesoscale model, which represents perfect single fibrils, whereas the experimental studies have been carried out at much larger tissues scales in tendon where more complex structural arrangements with additional voids and imperfections are observed. In fact, recent experimental studies of stretching of single collagen fibrils (34) showed similar stress levels as found in our simulation model, along with similar levels of yield strain.

To obtain a quantitative prediction at the tendon level, further improvement in the mesoscale model will have to be made, and larger structural levels of the tendon structure should be modeled (e.g., larger fibril (super-)structures, details of the hexagonal packing, microfibril coiling, defects, etc.). These studies, however, are beyond the scope of this study where our aim is to focus solely on the effects of OI on the properties of collagen fibrils. We note that the limitations of our mesoscale model apply to both the reference and to the mutated models, so they should not affect the main result, that is, the relative difference in strength between reference and mutated fibril and the underlying mechanisms.

In light of the remaining challenge to investigate OI effects from an experimental perspective, the predictions put forth in Fig. 5 could be regarded as a challenge to experimentalists. Carrying out such experimental studies might be feasible because recent advances in tensile testing now enable the quantification of the stress-strain response of single collagen fibrils (34).

## CONCLUSION

The main focus of this study was to assess the quantitative effect of OI-related glycine replacement mutations on the molecular, intermolecular and collagen fibril mechanical properties. We find that OI mutations have a strong effect on collagenous tissues at several levels in the collagen hierarchy, suggesting that OI must be understood as a multiscale materials phenomenon.

We have identified three major effects of OI mutations on the mechanical properties of collagenous tissues at ultrasmall scales. First, at the single molecule level, where molecular softening occurs as the disease severity increases (Fig. 2). Second, at the intermolecular level, where mutations lead to a weakening of intermolecular adhesion and increase of intermolecular equilibrium spacing as disease severity increases (shown in Fig. 3), leading to a reduction in likelihood of cross-link formation (Fig. 4). Third, at the collagen fibril level, where OI mutations lead to reduction in strength (Fig. 5) through a change in the stress distribution within fibrils, and specifically the formation of localized stress concentrations that lead to early material failure (Fig. 5 *c*).

For collagen fibril properties, the location of the mutations is a crucial factor in determining their impact on material properties, with mutations close to the ends of molecules featuring the strongest effects. Similar effects of the glycine replacement location along the triple helix on the severity of the resulting type of OI have been previously reported in experiment, showing that mutations located closer to the terminals (in particular the C-terminus) of the tropocollagen molecule lead to more severe phenotypes (8,35). Our simulations agree with this general trend. The observed changes in the intermolecular interaction and in the mechanical properties of collagen fibrils could be a major cause of structural and property changes of collagenous tissues, leading to a mechanical deterioration of fibrils and mineralized fibrils. Altogether, the effects described in this study are important to better understand the molecular-scale based multiscale mechanism of this disease. Specifically, our work shows that microscopic events involving only tens to hundreds of atoms can trigger tissue failure with potentially system wide catastrophic consequences.

Specifically, the results of this study can explain several key features found in OI tissues:

- Poor fibril packing. The alternation of collagen strands with normal interactions and strands with increased repulsion due to the presence of glycine replacements may hinder normal fibril packing, leading to a partial loss of the lateral crystallinity, a feature often observed in OI tissues (28,36).
- Decrease of cross-links density. At a larger length scale, the molecular effects of OI mutations observed in our study could help to explain the poor mechanical properties of tendons: the poorer fibril organization and the larger distance between tropocollagen molecules (that

is, less resistant material per unit area) in conjunction with the lower number of cross-links may account for the reduced mechanical strength of tendons typical of OI (31). Eventually, a mechanically inferior collagen matrix plus an increased and less organized mineral content might lead to the phenomenon of brittle bones, an important and physiologically evident feature of OI (28). Further studies will be required to provide more mechanistic insight into these specific mechanisms; this present study was focused solely on the pure collagen phase. Such studies should entail an explicit consideration of the interface between collagen and hydroxyapatite mineral crystals and it changes under the presence of mutations.

Changes in mineral crystals size and shape. A poorer lateral packing of collagen molecules may affect the growth of hydroxyapatite crystals. In an environment characterized by poor fibril packing, mineral crystals fail to grow highly organized in the fibril axis direction. Rather, they might be less organized and round-shaped instead of platelet-like. This feature would be consistent with the observed mineral crystals size and shape found in OI (37–39).

To the best of our knowledge, this is the first bottom-up nanomechanical mechanistic analysis of the effects of OI on the material properties at larger scales from a materials failure perspective. The results reported in this study contribute to a better understanding of the molecular mechanisms underlying OI and could eventually enable us to establish a direct link from the scale of genetic mutations to the macroscale phenomenon of brittle bones, albeit we emphasize that additional work is necessary to study the specific role of the mineral phase in this disease in particular for bone tissues.

We believe this study shows clearly that the severity of the OI mutation types and their effect on intermolecular interaction are correlated. Our findings suggest that the changes in the molecular structure due to the OI mutations change properties already at the molecular level and provide likely explanations of relevant phenomena connected with OI. The development of understanding of how structural modifications lead to the changes in the mechanical properties of higher hierarchical levels is critically important on the path of forming a more holistic picture of OI.

Our results of changes of collagen mechanics at the fibril level might be directly accessible to experimental studies using MEMS devices, as reported in a recent study (34). Such types of investigations might help to shed further light on mechanical effects of OI mutations by comparing to the results shown in Fig. 5. Other future studies could focus on the effect the position of the mutated protein, and what the effect of other triples surrounding the mutated glycine (that is, instead of the stabilizing Gly-Pro-Hyp sequence) would be.

OI is an autosomal dominant disease, and thus mutated structures arise even if only one of the two copies of the collagen gene encoding is mutated. Depending on the assembly of the chains into tropocollagen trimers, we may find molecules with zero, one, two, or three mutations at the same position. Our analysis has thus far considered only the case of three mutations at the same position (Fig. 1). It remains an open question what the effect of different numbers of mutations is. Future studies could be focused on the question whether or not mutation effects are additive (that is, two mutations lead to more severe effects than one mutation alone), or if it is an on/off process (that is, the effects do not change even if there is more than one mutation in the tropocollagen molecule). The modeling approach reported in this study could be adapted to further investigate these issues.

The development of links between failure of protein materials in the context of genetic disease is currently a little explored aspect with great potential to greatly enhance our understanding of the role of materials phenomena in biological systems (40), an effort referred to as materiomics. Thereby, the consideration of how material properties change in diseases could lead to a new paradigm that may expand beyond the focus on biochemical readings alone. This could eventually be important for both disease diagnosis and treatment. The work reported in this study is a beginning to initiate the development of constitutive material models for protein materials (such as in collagenous tissues) in the context of genetic disease. Further work must be done to validate the model through quantitative comparison with experiments, to link the predictions to larger tissue levels, and to incorporate additional structural features into the model. A similar approach as used in this study could be used for many other diseases in which materials failure due to a drastic change of the constitutive behavior of a material component plays a crucial role in disease initiation and progression.

## SUPPORTING MATERIAL

Additional text, equations, references, figures, and tables are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(09)01026-1.

The authors declare no conflict of interest of any sort.

This research was partially supported by the Progetto-Rocca fund and the MIT-Italy program. Support was provided by a National Science Foundation CAREER Award (grant CMMI-0642545) and by the Army Research Office (grant W911NF-06-1-0291).

## REFERENCES

1. Peltonen, L., A. Palotie, and D. J. Prockop. 1980. A defect in the structure of type I procollagen in a patient who had osteogenesis imperfecta: excess mannose in the COOH-terminal propeptide. *Proc. Natl. Acad. Sci. USA*. 77:6179–6183.
2. Rauch, F., and F. H. Glorieux. 2004. Osteogenesis imperfecta. *Lancet*. 363:1377–1385.

3. Primorac, D., D. W. Rowe, M. Mottes, I. Barisic, D. Anticevic, et al. 2001. Osteogenesis imperfecta at the beginning of bone and joint decade. *Croat. Med. J.* 42:393–415.
4. Sillence, D. O., A. Senn, and D. M. Danks. 1979. Genetic heterogeneity in osteogenesis imperfecta. *J. Med. Genet.* 16:101–116.
5. Kozloff, K. M., A. Carden, C. Bergwitz, A. Forlino, T. E. Uveges, et al. 2004. Brittle IV mouse model for osteogenesis imperfecta IV demonstrates postpubertal adaptations to improve whole bone strength. *J. Bone Miner. Res.* 19:614–622.
6. Dalgleish, R. 1997. The human type I collagen mutation database. *Nucleic Acids Res.* 25:181–187.
7. Roughley, P. J., F. Rauch, and F. H. Glorieux. 2003. Osteogenesis imperfecta—clinical and molecular diversity. *Eur. Cell. Mater.* 5:41–47.
8. Beck, K., V. C. Chan, N. Shenoy, A. Kirkpatrick, J. A. M. Ramshaw, et al. 2000. Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *Proc. Natl. Acad. Sci. USA.* 97:4273–4278.
9. Byers, P. H. 2001. Folding defects in fibrillar collagens. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356:151–157.
10. Yang, W., M. L. Battineni, and B. Brodsky. 1997. Amino acid environment modulates the disruption by osteogenesis imperfecta glycine substitutions in collagen-like peptides. *Biochemistry.* 36:6930–6935.
11. Rainey, J., and M. Goh. 2002. A statistically derived parameterization for the collagen triple-helix. *Protein Sci.* 13:2748–2754 (Erratum in *Protein Sci.* 2004. 13:2276).
12. Rainey, J., and M. Goh. 2004. An interactive triple-helical collagen builder. *Bioinformatics.* 20:2458–2459.
13. Gautieri, A., S. Vesentini, A. Redaelli, and M. J. Buehler. 2009. Single molecule effects of osteogenesis imperfecta mutations in tropocollagen protein domains. *Protein Sci.* 18:161–168.
14. Berendsen, H. J. C., D. Vanderspoel, and R. Vandrunen. 1995. GROMACS—a message-passing parallel molecular-dynamics implementation. *Comput. Phys. Commun.* 91:43–56.
15. Vandrunen, R., D. Vanderspoel, and H. J. C. Berendsen. 1995. GROMACS—a software package and a parallel computer for molecular-dynamics. *Abstr. Paper. Am. Chem. Soc. Natl. Meet.* 209:49.
16. Wess, T. J., A. P. Hammersley, L. Wess, and A. Miller. 1998. Molecular packing of type I collagen in tendon. *J. Mol. Biol.* 275:255–267.
17. Orgel, J. P. R. O., A. Miller, T. C. Irving, R. F. Fischetti, A. P. Hammersley, et al. 2001. The In Situ Supramolecular Structure of Type I Collagen. *Structure.* 9:1061–1069.
18. Baker, N. A., D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA.* 98:10037–10041.
19. Vallejo, D. F., F. Zamarreño, D. M. A. Guérin, J. R. Grigera, and M. D. Costabel. 2009. Prediction of the most favorable configuration in the ACBP-membrane interaction based on electrostatic calculations. *Biochim. Biophys. Acta.* 1788:696–700.
20. Dolinsky, T. J., J. E. Nielsen, J. A. McCammon, and N. A. Baker. 2004. PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* 32:W665–W667.
21. Miles, C. A., T. J. Sims, N. P. Camacho, and A. J. Bailey. 2002. The role of alpha2 chain in the stabilization of the collagen type I heterotrimer: a study of the type I homotrimer in oim mouse tissues. *J. Mol. Biol.* 321:797–805.
22. Buehler, M. J. 2006. Nature designs tough collagen: Explaining the nanostructure of collagen fibrils. *Proc. Natl. Acad. Sci. USA.* 103:12285–12290.
23. Buehler, M. J. 2008. Nanomechanics of collagen fibrils under varying cross-link densities: atomistic and continuum studies. *J. Mech. Behav. Biomed. Mat.* 1:59–67.
24. Sun, Y. L., Z. P. Luo, A. Fertala, and K. N. An. 2004. Stretching type II collagen with optical tweezers. *J. Biomech.* 37:1665–1669.
25. Sasaki, N., and S. Odajima. 1996. Stress-strain curve and Young's modulus of a collagen molecule as determined by the x-ray diffraction technique. *J. Biomech.* 29:655–658.
26. Fratzl, P., and R. Weinkamer. 2007. Nature's hierarchical materials. *Prog. Mater. Sci.* 52:1263–1334.
27. Kuznetsova, N., D. J. McBride, and S. Leikin. 2001. Osteogenesis imperfecta murine: interaction between type I collagen homotrimers. *J. Mol. Biol.* 309:807–815.
28. Miller, A., D. Delos, T. Baldini, T. M. Wright, and N. P. Camacho. 2007. Abnormal mineral-matrix interactions are a significant contributor to fragility in *oim*/*oim* bone. *Calcif. Tissue Int.* 81:206–214.
29. Sims, T. J., C. A. Miles, A. J. Bailey, and N. P. Camacho. 2003. Properties of collagen in OIM mouse tissues. *Connect. Tissue Res.* 44: 202–205.
30. Bailey, A. J. 2001. Molecular mechanisms of ageing in connective tissues. *Mech. Ageing Dev.* 122:735–755.
31. Misof, K., W. J. Landis, K. Klaushofer, and P. Fratzl. 1997. Collagen from the osteogenesis imperfecta mouse model (oim) shows reduced resistance against tensile stress. *J. Clin. Invest.* 100:40–45.
32. Broberg, K. B. 1990. Cracks and Fracture. Academic Press, San Diego.
33. Buehler, M. J. 2008. Atomistic Modeling of Materials Failure. Springer, New York.
34. Shen, Z. L., M. R. Dodge, H. Kahn, R. Ballarini, and S. J. Eppell. 2008. Stress-strain experiments on individual collagen fibrils. *Biophys. J.* 95:3956–3963.
35. Liu, X. Y., S. Kim, Q. H. Dai, B. Brodsky, and J. Baum. 1998. Nuclear magnetic resonance shows asymmetric loss of triple helix in peptides modeling a collagen mutation in brittle bone disease. *Biochemistry.* 37:15528–15533.
36. McBride, D. J., V. Choe, J. R. Shapiro, and B. Brodsky. 1997. Altered collagen structure in mouse tail tendon lacking the alpha2(I) chain. *J. Mol. Biol.* 270:275–284.
37. Grabner, B., W. J. Landis, P. Roschger, S. Rinnerthaler, H. Peterlik, et al. 2001. Age- and genotype-dependence of bone material properties in the osteogenesis imperfecta murine model (oim). *Bone.* 29:453–457.
38. Camacho, N. P., L. Hou, T. R. Toledano, W. A. Ilg, C. F. Brayton, et al. 1999. The material basis for reduced mechanical properties in oim bones. *J. Bone Miner. Res.* 14:264–272.
39. Fratzl, P., O. Paris, K. Klaushofer, and W. J. Landis. 1996. Bone mineralization in an osteogenesis imperfecta mouse model studied by small-angle x-ray scattering. *J. Clin. Invest.* 97:396–402.
40. Buehler, M. J., and Y. C. Yung. 2009. Deformation and failure of protein materials in physiologically extreme conditions and disease. *Nat. Mater.* 8:175–188.