Electronic Supplementary Material (ESI) for Chemical Communications.

Specific osteogenesis imperfecta-related Gly substitutions in type I collagen induce distinct structural, mechanical, and dynamic characteristics

Haoyuan Shi,[a] Liming Zhao,[a] Chenxi Zhai[a] and Jingjie Yeo[*a]

[a]J[2] Lab for Engineering Living Materials, Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY, 14853, United States.

1. Method

1.1. MD simulation

The fully atomistic (FA) model of the chosen wild-type collagen sequence with triplets of GPP amino acids at both ends was generated using the Triple-Helical collagen Building Script (THeBuScr)[1](http://structbio.biochem.dal.ca/jrainey/THeBuScr.html). THeBuScr predicted the three-dimensional structure of the triple-helical collagen molecule with the primary GXY sequence. The mutated collagens were generated by replacing Gly at the corresponding sites in the pro-α 1(I) chain (chain A and C) with other residues using the QwikMD plugin[2] in Visual Molecular Dynamics (VMD).[3] These mutations are related to osteogenesis imperfecta (OI2 or OI3) based on UniProtKB - P02452 (https://www.uniprot.org/uniprot/P02452).[4] We constructed four mutated collagen models, including the 1022nd Gly to Val, the 1025th Gly to Arg, the 1049th Gly to Ser, and the double site mutation at both the 1022nd and 1025th Gly. Together with the wild-type collagen, all five collagen models were used as input structures for MD simulations in NAMD 2.13.[5] In simulations, we used the latest CHARMM36m forcefield[6] with updated parameters for many amino acids, particularly hydroxyproline and hydroxylysine which are common amino acids found in large numbers in collagen. The CHARMM potential function[7] is:

$$U_{CHARMM} = \sum_{bonds} K_b\left(b_{ij} - b_0\right)^2 + \sum_{angles} K_\theta\left(\theta_{ijk} - \theta_0\right)^2$$
$$+ \sum_{dihedrals} K_\varphi\left[1 + \cos\left(n\varphi_{ijkl} - \delta\right)\right]^2 + \sum_{improper} K_\phi\left(\phi_{ijkl} - \phi_0\right)^2$$
$$+ \sum_{Urey-Bradley} K_{UB}(U_{ik} - U_0)^2$$
$$+ \sum_{nonbonded} \left\{\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi D r_{ij}}\right\}$$

where K_b, K_θ, K_φ, K_ϕ, and K_{UB} are the bond, angle, dihedral angle, improper angle and Urey–Bradley force constants, respectively; b_{ij}, θ_{ijk}, φ_{ijkl}, ϕ_{ijkl}, and U_{ik} are the bond length, bond angle, dihedral angle, improper torsion angle, and Urey–Bradley 1,3-distance respectively; b_0, θ_0, φ_0, ϕ_0, and U_0 are the equilibrium terms for such variables; n is the periodicity and δ the phase of a torsion; ε_{ij} is the well depth of the Lennard-Jones potential; σ_{ij} is the distance at the LJ minimum; q is the partial atomic charge; D is the effective dielectric constant; and r_{ij} is the distance between any atoms i and j.

Each collagen structure was explicitly solvated in a TIP3P[8] water box and neutralized by Na^+ or Cl^- ions. These systems were energy minimized with the conjugate gradient algorithm. NPT simulations of 1 ns were performed with harmonic constraints to the α-carbon atoms, followed by 100 ns NPT MD production runs without any constraints. Langevin dynamics[9] and Nosé-Hoover Langevin piston[10] were used for temperature and pressure control at 310 K and 1.013 bar, respectively. Rigid bonds were used with the SHAKE algorithm[11], allowing a timestep of 2 fs. The Particle-mesh Ewald (PME) method was used to calculate long-range electrostatic interactions.[12] The potential energy showed that the mutated structure equilibrated in the first 60 ns simulations (Fig. S1), and thus the last 40 ns simulation with 2,000 conformational ensembles were used for structural and dynamic analysis.

1.2. Stiffness calculations

Axial Stiffness

The axial stiffness can be calculated based on the formula:[13,14]

$$\frac{1}{2}k_bT = \frac{1}{2}k_L\langle(\Delta L)^2\rangle$$

Where T is the temperature, k_b is the Boltzmann constant, L is the length of the collagen and k_L is the axial stiffness. ΔL is the difference between instantaneous length and average length, and $(\Delta L)^2$ is the mean square of ΔL or equilibrium fluctuations, calculated by the dynamic trajectories of equilibrium MD simulations. Note that the axial stiffness is sensitive to the length, which affects the equilibrium fluctuations of collagen. Since collagen is a triple-helical structure containing repeat GXY units, the total collagen length can be calculated by adding all single GXY units together, as shown in Fig. S3a. Here, we defined the length of a single GXY unit as the length between the centroid of three α-carbon atoms of Gly in the current and the next GXY unit. The length of the last GXY unit considered the length between the centroid of three α-carbon atoms of Gly in this GXY unit and the centroid of

three OT2 atoms of Y in this GXY unit. Taking the wild-type collagen as an example, Fig. S4a shows the fluctuation of its length based on the last 40 ns simulation with 2,000 conformational ensembles or frames, which is related to the axial stiffness according to the formula above.

Bending Stiffness

The bending stiffness can be calculated based on the formula:

$$2 \times \frac{1}{2} k_b T = \frac{1}{2} k_\delta \langle (\Delta\delta)^2 \rangle$$

Where δ is the end deflection of the collagen and k_δ is the bending stiffness. The factor 2 in the left is due to the end deflection fluctuations are in two dimensions. In order to calculate the fluctuations of end deflection based on the natural fluctuations in equilibrium MD simulations, we aligned the first GXY unit of the 2,000 conformational ensembles and projected the coordinates of the centroid of three OT2 atoms of Y in the last GXY unit on to x-y plane, as shown in Fig. S3b with result of the wild-type collagen in Fig. S4b. By calculating the average endpoint coordinate of 2,000 frames, we can calculate the $(\Delta\delta)^2$ term based on the mean square of the distance between the instantaneous endpoint coordinate and the average endpoint coordinate.

Torsional Stiffness

The torsional stiffness can be calculated based on the formula:

$$\frac{1}{2} k_b T = \frac{1}{2} k_\theta \langle (\Delta\theta)^2 \rangle$$

Where θ is the twist angle of a single chain around the central axis and k_θ is the torsional stiffness. Still, we calculated the total twist angle by adding the twist angle of all single GXY units together. As shown in Fig. S3c, the twist angle in a single GXY unit was defined as the included angle between two vectors, starting from the centroid of three α-carbon atoms in the current and the next GXY unit, with the direction being along the α-carbon atom in Gly of Chain C of the current and the next GXY unit, respectively. Fig. S4c shows the fluctuations of the twist angle for the wild-type collagen.

1.3. Moving block bootstrap

Moving block bootstrap[15] was used for data analysis to avoid bias in conformational ensembles in stiffness calculations. Since the fluctuations are time-dependent, moving block

bootstrap is suitable for data analysis in this work by creating a sampling window to perform a bootstrap analysis on a time series. The functions of the moving block bootstrap we used are in the tsmoothie library (https://github.com/cerlymarco/tsmoothie) in Python. The steps are as follow and shown in Fig. S5: Assume that there are n data. First, we chose two hyper-parameters, window length w and block length b. Then, we randomly selected n/b blocks to recreate a new data set with n data. We calculated the stiffness based on the new data set. Next, repeat selection process N times, and calculated N times stiffness. Finally, we obtained the average stiffness with standard deviation based on N stiffness. In this work, the data set contained 2,000 data for 2,000 frames, and we selected w to be 10, b to be 500, and N to be 2,000. Besides, we also tried several different hyper-paramteters with the error being no more than 5%. The stiffnesses of the collagens are in Table S1.

1.4. Elastic modulus calculations

Since the wild-type collagen triple helix maintains a stable structure resembling that of a prismatic rod with a circular cross-section (Fig. S6a), the Euler-Bernoulli beam theorem can be applied to solve for rigidities or elastic moduli. Also, stiffness is sensitive to the structure's length and hard to compare with previous reported data. Thus, the obtained Young's moduli in this work can also be compared to previous studies to validate the rationality of the stiffness calculation methods. The radius of the collagen triple helix is defined as the average distance of the farthest non-hydrogen atom on each residue in the cross-section to the central axis (Fig. S6b). Note that the mutated collagens unwound in the mutation sites (Fig. S6c), which destroyed the orderly structures that allow elastic modulus calculations.

Young's modulus

According to Hooke's law, the stress is proportional to the strain:

$$\frac{F}{A} = E\frac{\Delta L}{L}$$

where E is Young's modulus, F is the force along to the rod axis, A is the area of cross-section, L is the rod length, and ΔL is the change in length. The stiffness k_L can also be written as the ratio of force to displacement:

$$k_L = \frac{F}{\Delta L}$$

Therefore, Young's modulus can be calculated by:

$$E = \frac{L}{A} k_L$$

Bending rigidity

The end deflection can be derived from the bending of a cantilevered rod:

$$\Delta\delta = \frac{FL^3}{3EI}$$

Where I is the cross-sectional moment of inertia. Moreover, the bending stiffness satisfy:

$$k_\delta = \frac{F}{\Delta\delta}$$

Therefore, bending rigidity can be calculated by:

$$EI = \frac{k_\delta L^3}{3}$$

Note that we can also calculate Young's modulus based on bending rigidity. In this work, we used those two methods to calculate Young's modulus to verify self-consistency.

Torsional rigidity and shear modulus

The torsional stiffness is defined as:

$$k_\theta = \frac{GJ}{L}$$

where G is in-plane shear modulus, J is the polar moment of inertia, and GJ is torsional rigidity.

All the results of the wild-type collagen are shown in Table S2.

1.5. MSM Construction and Data Analysis

The Markov State Model (MSM) is a stochastic process model used to study molecular dynamics. Based on the Markov chain, MSM is used to construct a transition matrix and obtain the stationary distribution. In this work, MSM was implemented by PyEMMA.[16] Rather than requiring a single long trajectory, MSM integrates some short, independent trajectories and extracts dynamics information from them.[17] Thus, for each mutation, we performed 4×100 ns simulations starting at unique initial configurations. The different initial configurations were obtained via 50 ps heating under 400 K followed by 50 ps annealing

under 310 K. Frames were saved every 0.01 ns. The other simulation set-up was the same as part 1.1. Also, we conducted the same calculation process for those 4 repeated simulations for each case to calculate the stiffnesses, and the results are consistent with Section S1.2 with an error of less than 3%.

In terms of the construction of MSM, all trajectories were parameterized by the backbone dihedral angles, namely φ and ψ, because they uniquely define the conformation of a polypeptide backbone and are key in the stabilization of the triple-helical structure.[18] Time-lagged independent component analysis[19] and the k-means method were used to reduce the feature dimensions and to cluster the samples into microstates. Based on the convergence of VAMP-2 scores,[20] we identified 30 microstates for the wild-type and the 1049th site mutation collagen and 50 microstates for the others. Then, we employed MSM to verify the consistency of the 40 ns ensembles used to compute the stiffness and the 400 ns simulations fed into MSM. As indicated by Fig. S8, the majority of the samples in the 40 ns trajectories lay in the basins formed by the MSM data, which proved the consistency between the two data sources. To guarantee the Markovianity, the MSM of double-site mutation collagen used a lagged time of 1 ns, and the single-site mutation collagens in 1022nd site or 1025th used a lagged time of 2 ns, and the model of wild-type collagen and the 1049th site mutation collagen used a larger lagged time of 8 ns, the values of which were comparable with the previous work.[21] The transition probability network was given by Bayesian estimation. According to Bayes' theorem, the posterior probability has the following property:

$$P(X|C) \propto P(X)P(C|X),$$

where $C = (c_{ij})$ is the observed effective count matrix.[22] Instead of focusing on the transition probability matrix $P = (p_{ij})$, the Bayesian reversible MSM works on a modified matrix $X = (x_{ij})$, which is correlated to P via the stationary distribution π as follows:

$$x_{ij} = \pi_i p_{ij},$$

so that matrix X obeys detailed balance.[22] The prior probability has a form of:

$$P(X) \propto \prod_{i \geq j} x_{ij}^{b_{ij}},$$

where b_{ij} is the prior count.[22] Therefore, the posterior can be expressed as:[22]

$$P(X|C) \propto \prod_{i \geq j} x_{ij}^{b_{ij}} \prod_{i,j} \left(\frac{x_{ij}}{\sum_k x_{ik}} \right)^{c_{ij}}.$$

After the construction of MSM, as denoted by the red dots in Fig. S8, we drew the first seven highest-probability microstates from the stationary distribution for the analysis of β structures and hydrogen bonds. Then, 100 frames for each microstate were randomly extracted to count the extensiveness of the β structures and hydrogen bonds so that we can calculate the mean and the standard deviation among the seven microstates, as shown in Fig. 2b and c. The significant differences provided in Fig. 2b were obtained by the two independent samples t-test. The test statistics is written as:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{{s_x}^2}{n} + \frac{{s_y}^2}{m}}},$$

where the numerator is the difference between the sample means. The symbols s_x (s_y) and n (m) are the standard deviation and the corresponding sample size, respectively. Of note is that the variances were not homogeneous except in the case of 1022nd site mutation. Thus, the DOF of the 1022nd site mutation data was $n + m - 2$, while the DOF of the others were approximated by Welch–Satterthwaite equation:

$$v = \frac{\left(\frac{s_x^2}{n} + \frac{s_y^2}{m}\right)^2}{\left(\frac{s_x^2}{n^2(n-1)} + \frac{s_y^2}{m^2(m-1)}\right)}.$$

1.6. Normal mode analysis with the elastic network model

Normal mode analysis (NMA) can describe the flexible states accessible to a protein around an equilibrium position, based on the physics of small oscillations.[23,24] When protein is near the equilibrium state q^0, the potential energy then can be expanded as a power series in q:

$$V(q) = V(q^0) + \sum_i \left(\frac{\partial V}{\partial q_i}\right)^0 (q_i - q_i^0) + \frac{1}{2}\sum_{i,j} \left(\frac{\partial^2 V}{\partial q_i \partial q_j}\right)^0 (q_i - q_i^0)\left(q_j - q_j^0\right) + \cdots$$

The first term is the potential energy at a minimum, which can be set to zero. The second term is identically zero at any local minimum of the potential. Thus, the potential energy can be written as:

$$V(q) = \frac{1}{2}\sum_{i,j} \left(\frac{\partial^2 V}{\partial q_i \partial q_j}\right)^0 (q_i - q_i^0)\left(q_j - q_j^0\right) = \frac{1}{2}\Delta q^T \mathbf{H} \Delta q$$

where **H** is the Hessian matrix. The eigenvectors of the Hessian matrix are normal mode vectors and are defined as the normal modes of the molecule. The eigenvalues of the Hessian matrix give a squared frequency of vibration, $\lambda_k = \omega_k^2$. Since the energy is equally distributed among the modes, we can use $A = \frac{1}{\sqrt{\lambda_k}}$ to reflect relative amplitude in the different modes, and lower-frequency modes exhibit larger oscillation amplitudes, dominant in the accessible vibrationswithin the molecule.

Although NMA is less computationally expensive than MD simulation, proteins containing N atoms still need to solve a Hessian matrix of $3N \times 3N$ in size. Therefore, the elastic network model (ENM) is widely used to simplify the protein as a network of masses coupled with harmonic potentials.[23,24] In this work, we built the anisotropic network model (ANM) of ENM for all collagen models obtained by MSMs in ProDy 2.0,[25] where every node corresponded to α-carbon atoms and edges were modelled by springs.[26] In this model, we only need to consider the second partial derivative of the potential function or the Hessian matrix that is the potential for each pair α-carbons[26] and then solve for vibrational direction and the relative amplitude in the different modes based on

$$\mathbf{H} = U\Lambda U^T$$

The column of the matrix U is an eigenvector that describes the vibrational direction, and the eigenvalues of the Hessian matrix is the diagonal of Λ that give a squared frequency of vibration, $\lambda_k = \omega_k^2$. We can then use $A = \frac{1}{\sqrt{\lambda_k}}$ to reflect relative amplitude in the different modes. The vibrational direction and amplitude for each α-carbons can be obtained based on the superposition principle for each pair of α-carbons. The representative dynamics were obtained as a linear combination of the first 20 lowest frequency modes, scaled by mode amplitude. The cutoff distance of interactions was 15 Å, and the spring constant was 1. Higher frequency modes were neglected with minimal effects on dynamics.

1.7 Molecular dynamics data analysis tools

The RMSFs of all collagen models were calculated based on the 2,000 conformation ensembles via MDAnalysis tools.[27] The secondary structures were determined using the STRIDE algorithm[28] in VMD. The criteria of hydrogen bonds were set to be within 0.35 nm between a hydrogen atom and a hydrogen bond acceptor, and less than 35° of the bonding angle. All molecules were visualized using VMD and in-house TCL scripts.

1 J. K. Rainey and M. C. Goh, *Bioinformatics*, 2004, **20**, 2458–2459.

2 J. V Ribeiro, R. C. Bernardi, T. Rudack, J. E. Stone, J. C. Phillips, P. L. Freddolino and K. Schulten, *Sci. Rep.*, 2016, **6**, 26536.

3 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.

4 *Nucleic Acids Res.*, 2021, **49**, D480–D489.

5 J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M. C. R. Melo, B. K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V Kalé, K. Schulten, C. Chipot and E. Tajkhorshid, *J. Chem. Phys.*, 2020, **153**, 44130.

6 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller and A. D. J. MacKerell, *Nat. Methods*, 2017, **14**, 71–73.

7 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.

8 W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey and M. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

9 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.

10 S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, *J. Chem. Phys.*, 1995, **103**, 4613–4621.

11 V. Kräutler, W. F. van Gunsteren and P. H. Hünenberger, *J. Comput. Chem.*, 2001, **22**, 501–508.

12 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *jcp*, 1995, **103**, 8577–8593.

13 I. Adamovic, S. M. Mijailovich and M. Karplus, *Biophys. J.*, 2008, **94**, 3779–3789.

14 A. Shamloo and B. Mehrafrooz, *Cytoskeleton*, 2018, **75**, 118–130.

15 H. R. Kunsch, *Ann. Stat.*, 1989, **17**, 1217–1241.

16 M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J. H. Prinz and F. Noé, *J. Chem. Theory Comput.*, 2015, **11**, 5525–5542.

17 J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte and F. Noé, *J. Chem. Phys.*, 2011, **134**, 174105.

18 J. Engel and H. P. Bächinger, *Top. Curr. Chem.*, 2005, **247**, 7–33.

19 S. Schultze and H. Grubmüller, .

20 H. Wu and F. Noé, *J. Nonlinear Sci.*, 2020, **30**, 23–66.

21 S. Park, T. E. Klein and V. S. Pande, *Biophys. J.*, 2007, **93**, 4108–4115.

22 B. Trendelkamp-Schroer, H. Wu, F. Paul and F. Noé, *J. Chem. Phys.*, 2015, **143**, 11B601_1.

23 J. A. Bauer and V. Bauerová-Hlinková, *Homol. Mol. Model. - Perspect. Appl.*, 2021, 1–18.

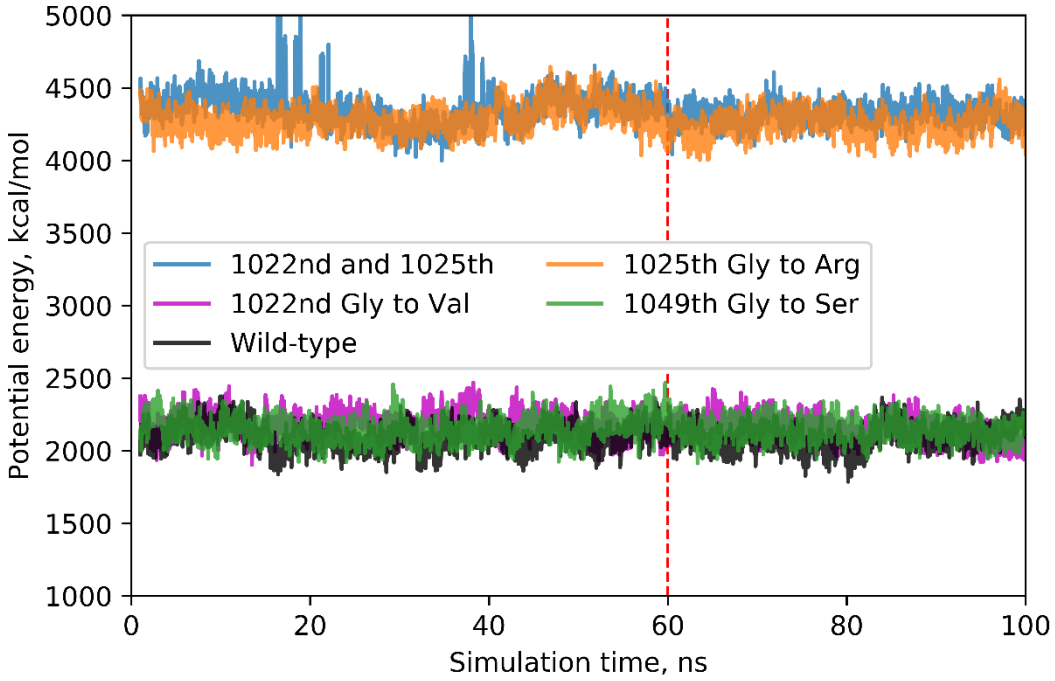
24 I. Bahar, T. R. Lezon, A. Bakan and I. H. Shrivastava, *Chem. Rev.*, 2010, **110**, 1463–1497.

25 S. Zhang, J. M. Krieger, Y. Zhang, C. Kaya, B. Kaynak, K. Mikulska-Ruminska, P. Doruker, H. Li and I. Bahar, *Bioinformatics*, , DOI:10.1093/bioinformatics/btab187.

26 A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin and I. Bahar, *Biophys. J.*, 2001, **80**, 505–515.

27 R. Gowers, M. Linke, J. Barnoud, T. Reddy, M. Melo, S. Seyler, J. Domański, D. Dotson, S. Buchoux, I. Kenney and O. Beckstein, *MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations*, 2016.

28 D. Frishman and P. Argos, *Proteins*, 1995, **23**, 566–579.



275

276 **Fig. S1.** The potential energy vs simulation time for collagen models. The last 40 ns were
277 used for analysis.

278

Fig. S2. The sequence of the wild-type collagen with OI-related Gly substitution mutation sites (red).

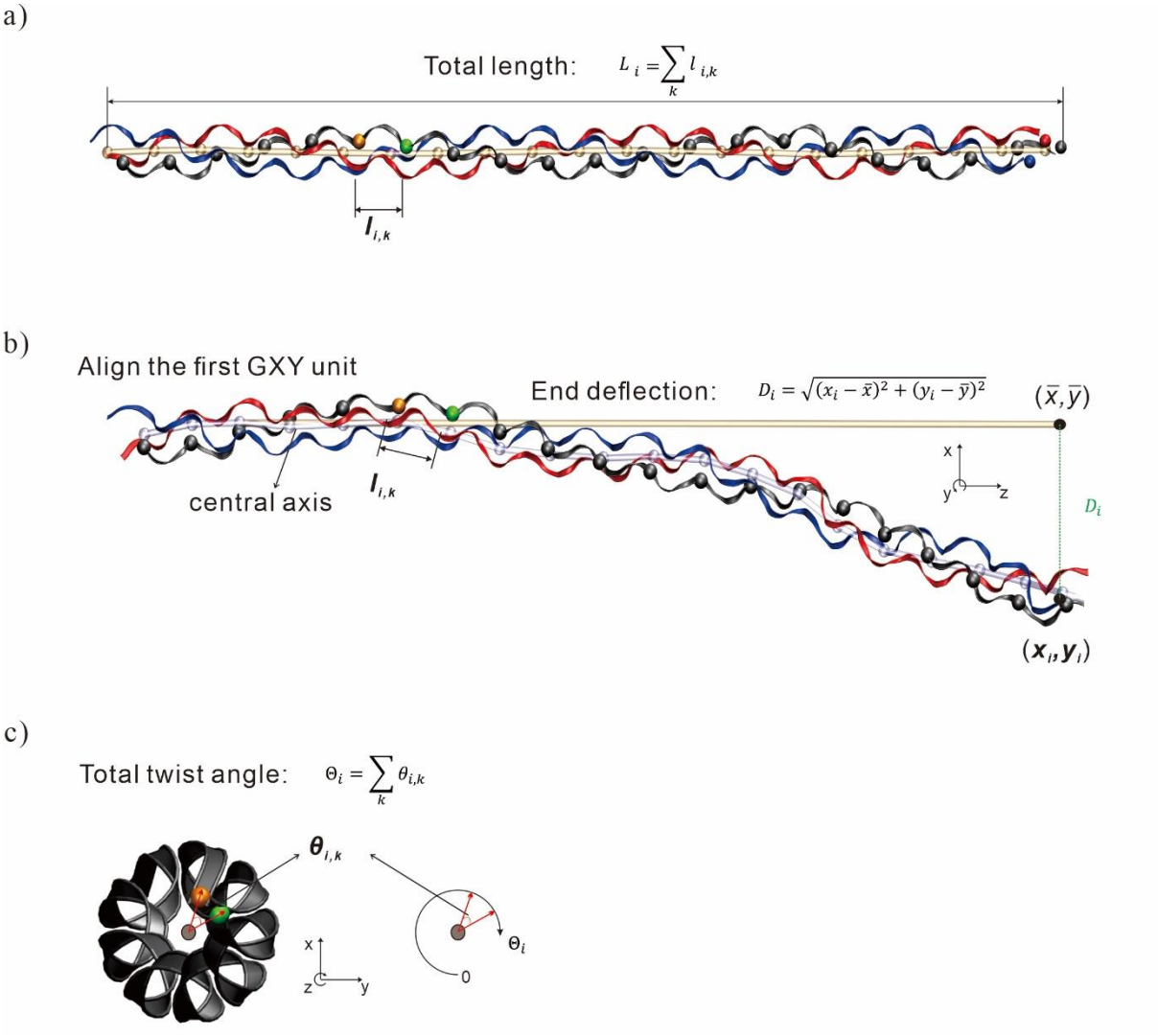
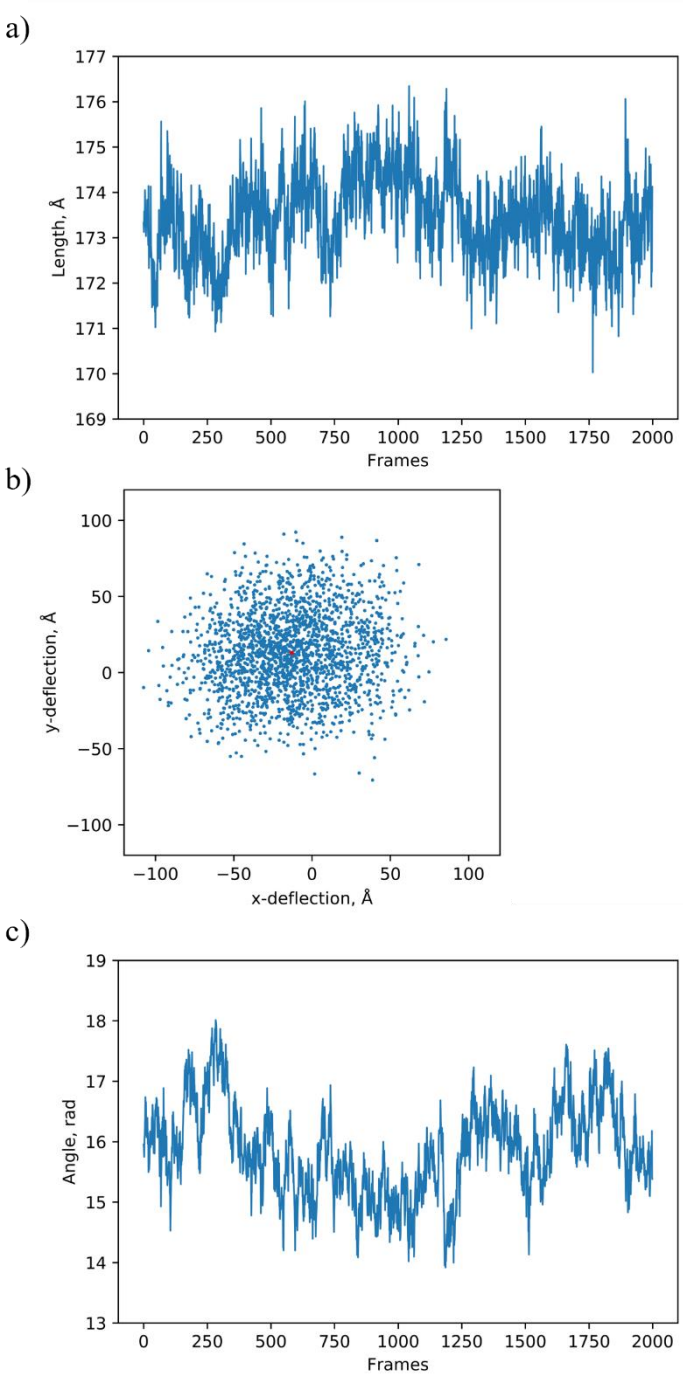


Fig. S3. Scheme of calculating instantaneous collagen a) length, b) end deflection, and c) twist angle.



288

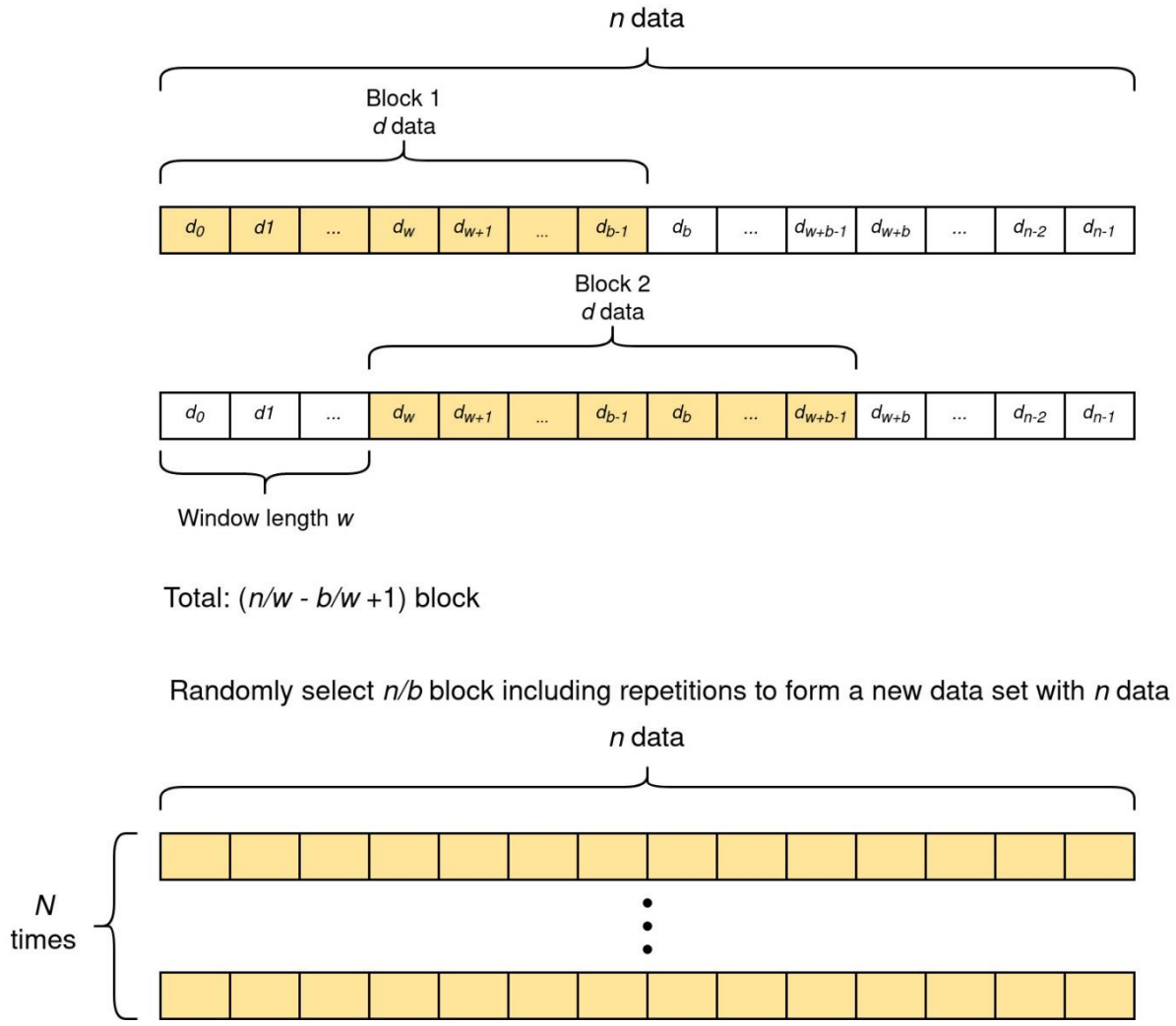
289 **Fig. S4.** Fluctuations of the a) length, b) endpoint coordinates and c) twist angle for the wild-
290 type collagen.

291

292

293

294



295

296 **Fig. S5.** Scheme of moving block bootstrap method to avoid bias in conformational
297 ensembles in stiffness calculations.

298

299

300

301

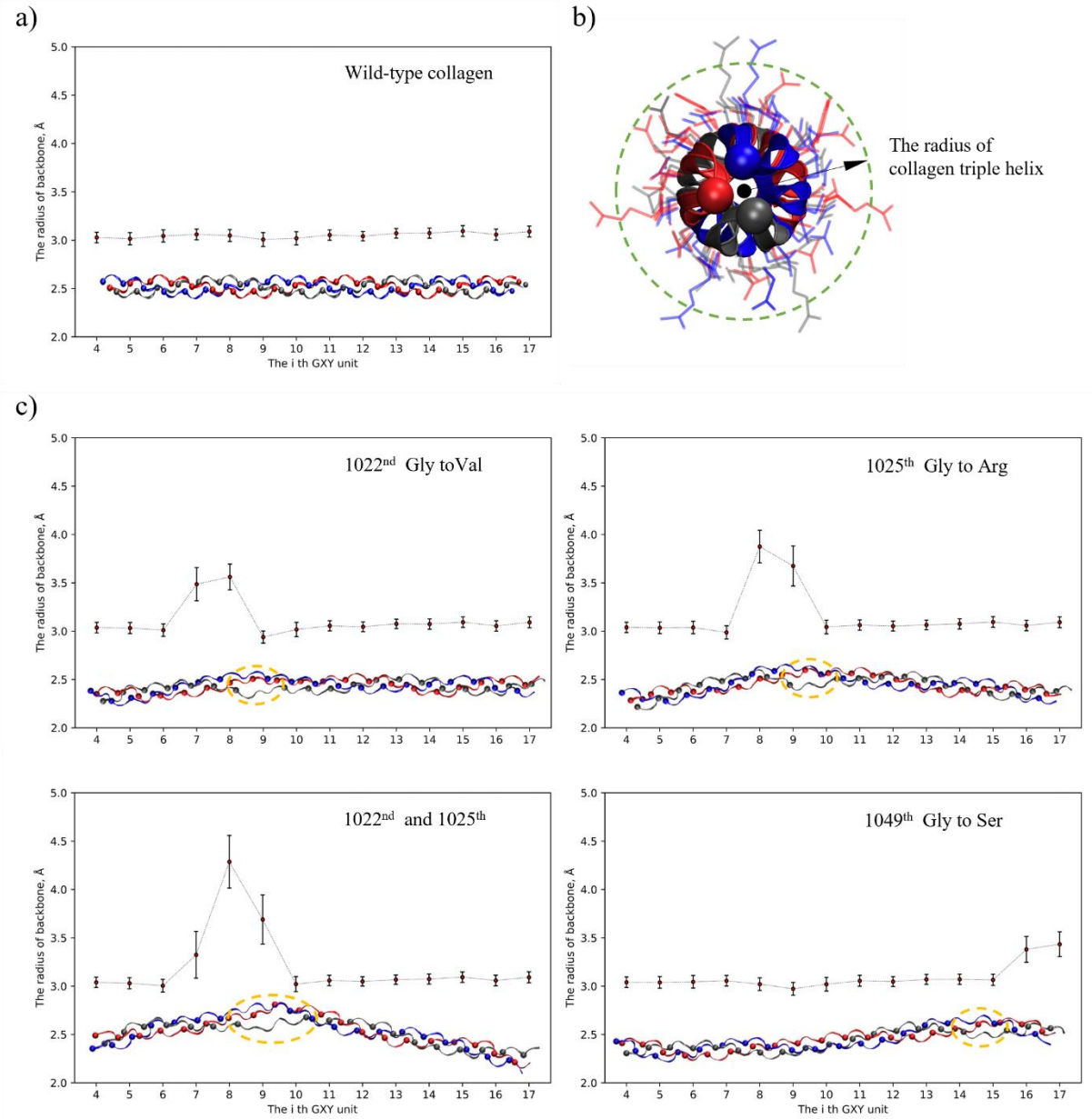


Fig. S6. a) The wild-type collagen maintained a stable structure after MD simulations. b) The equivalent radius of the wild-type collagen triple helix, considered as a circular cross-section. c) The mutated collagens unwound and led to a larger radius of backbone in the mutation sites.

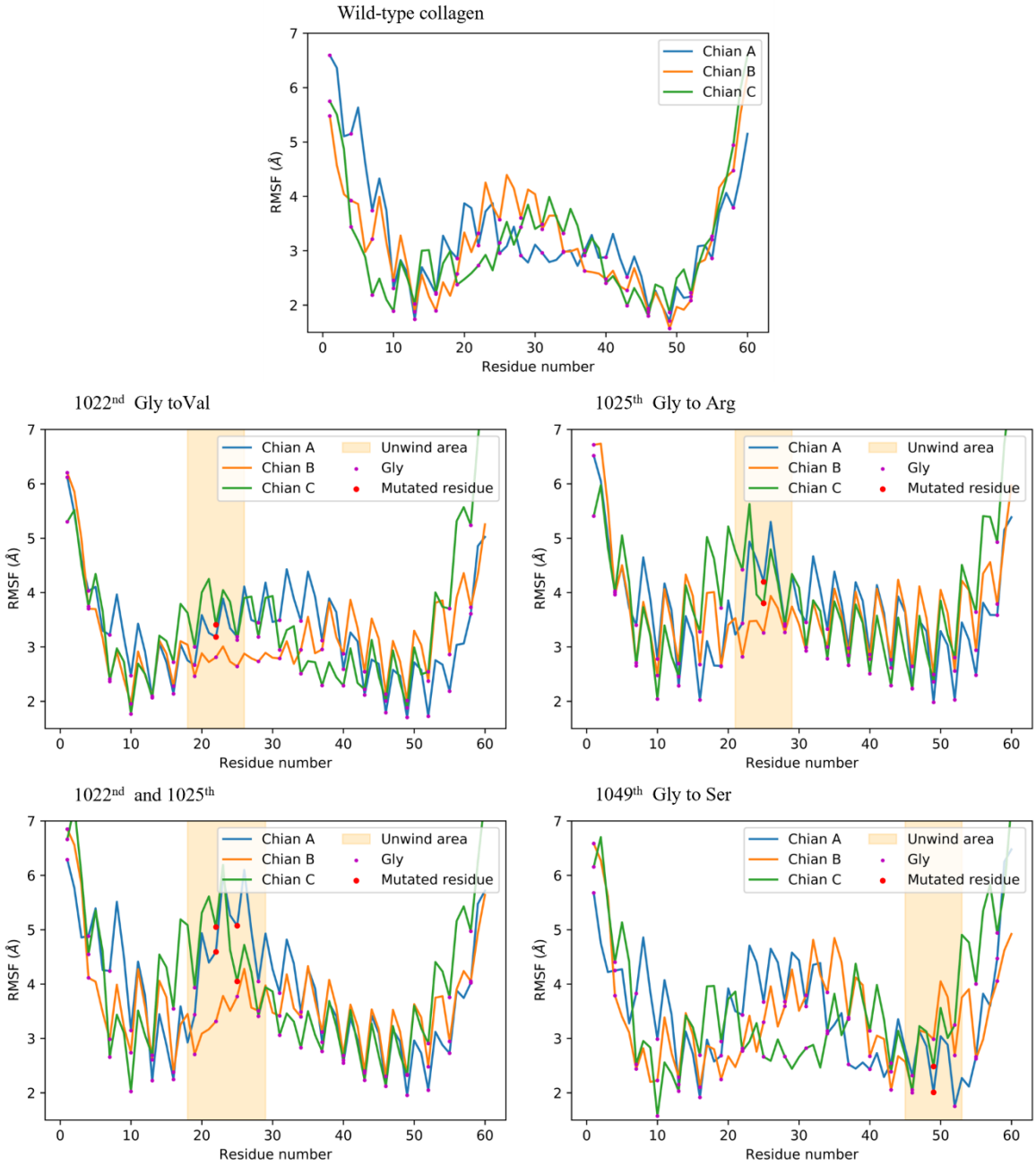


Fig. S7. Root-mean-square fluctuation (RMSF) of α-carbon in collagen models.

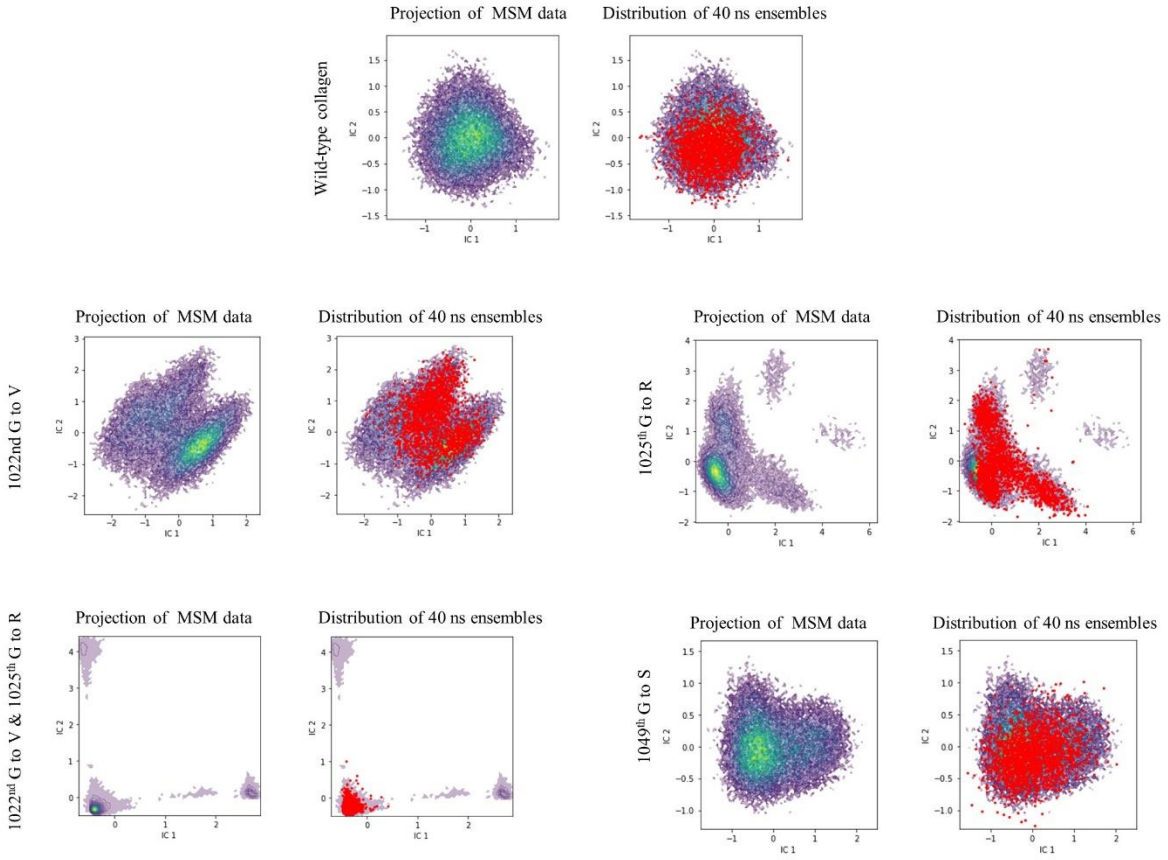


Fig. S8. Verification of the consistency between the MSM data and the 40 ns ensembles. The ensembles were plotted as red dots. The x-axis and y-axis correspond to the first and the second independent components obtained from TICA, respectively.

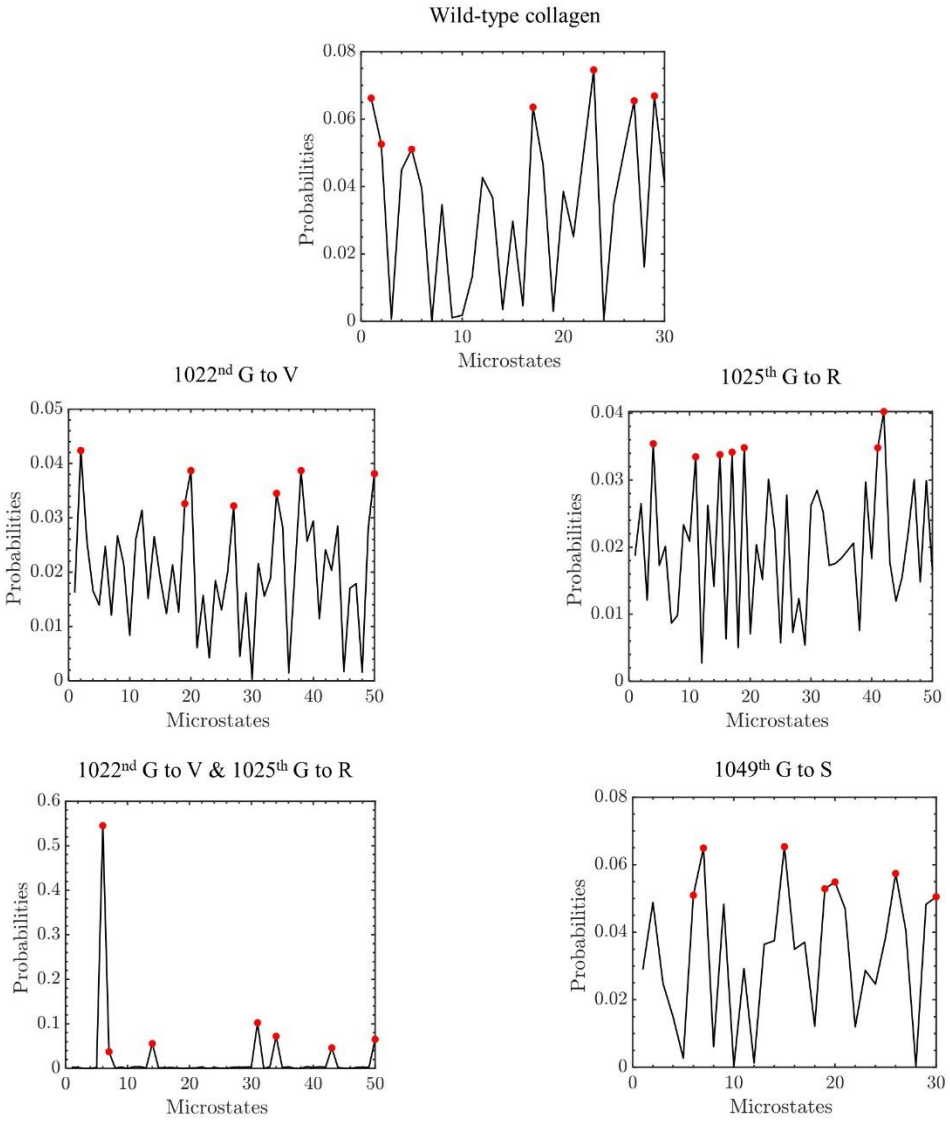


Fig. S9. Stationary probabilities of all microstates. The selected states were marked as red dots.

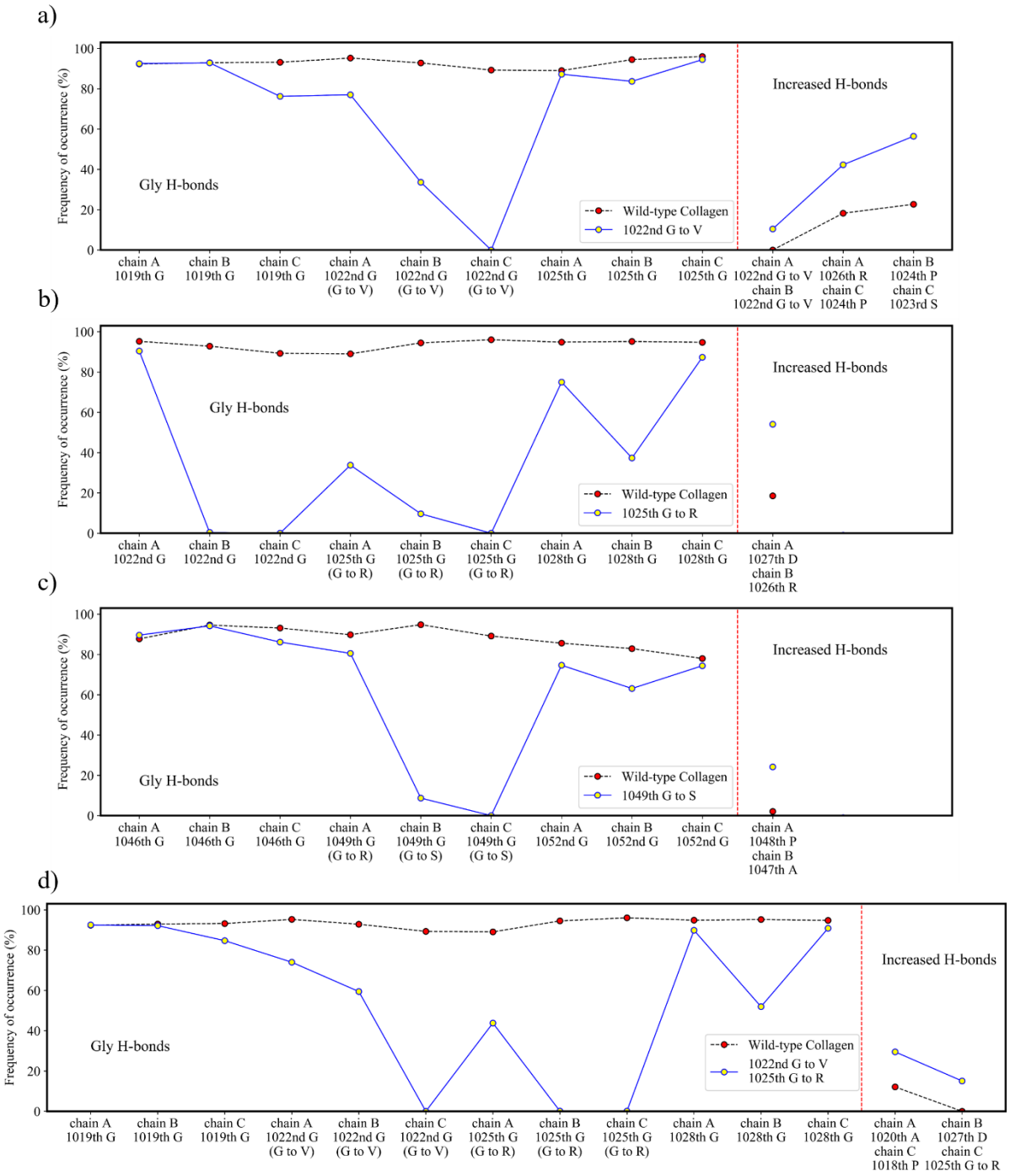
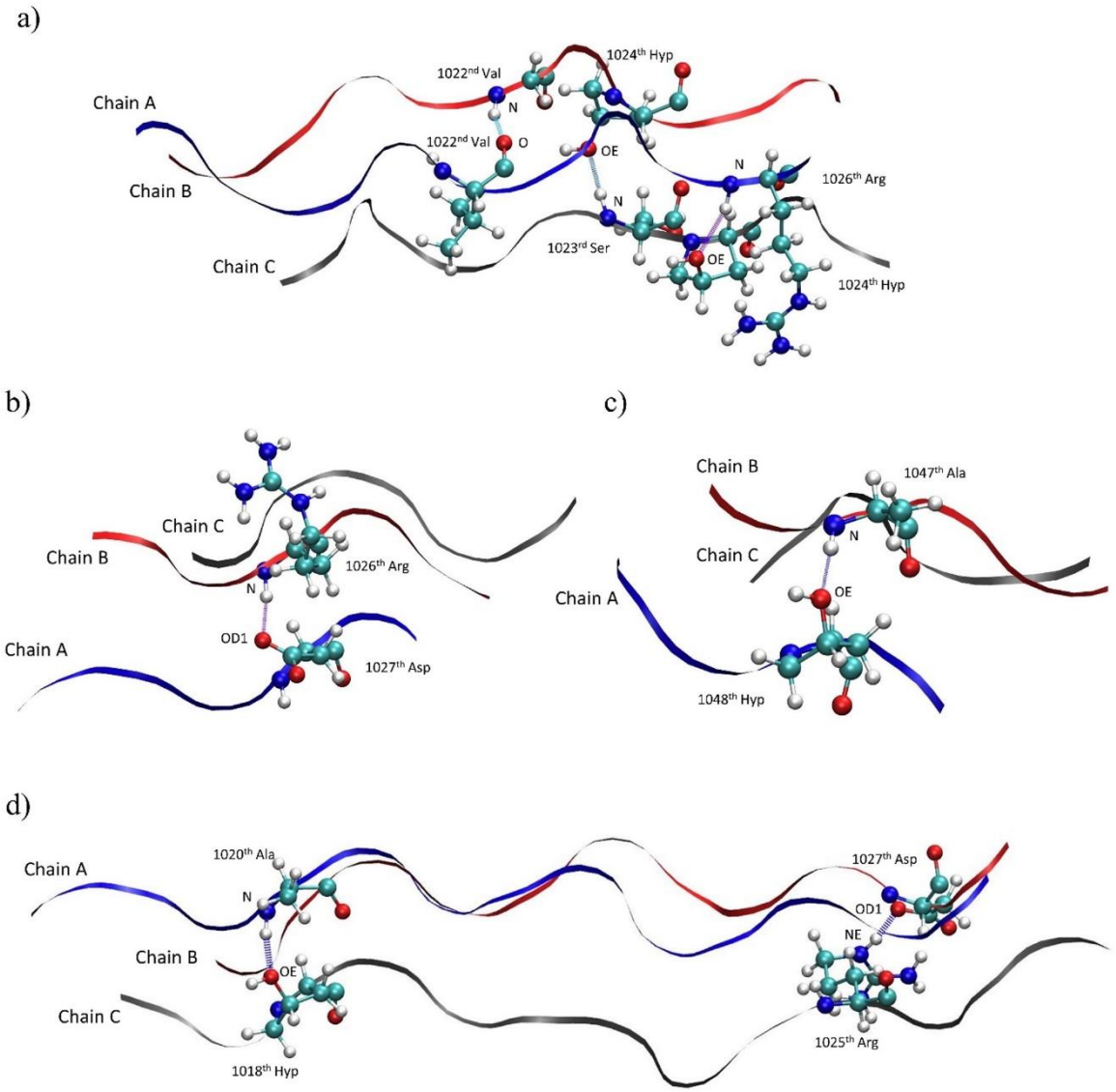


Fig. S10. The frequency of occurrence, based on 2,000 conformational ensembles, of Gly H-bonds in the mutated GXY unit and the GXY units before and after the mutated GXY unit, together with the increased H-bonds due to mutations. a) Wild-type collagen; b) 1022nd G to V; c) 1025th G to R; d) 1049th G to S; e) 1022nd G to V & 1025th G to R.



330

331 **Fig. S11.** Schematic diagram of increased H-bonds due to the mutations. Carbon is cyan,
332 oxygen is red, nitrogen is blue, and hydrogen is white. a) Wild-type collagen; b) 1022nd G to
333 V; c) 1025th G to R; d) 1049th G to S; e) 1022nd G to V & 1025th G to R.

334

335 **Table S1.** Stiffness of the wild-type collagen and several OI-related mutated collagens.

Cases	Length Å	Axial stiffness pN/nm	Bending stiffness pN/nm	Torsional stiffness pN·nm
Wild-type collagen	173.418	504.098 ± 12.726	0.528 ± 0.015	7.812 ± 0.084
1022nd Gly to Val	174.163 (0.430%)	625.089 ± 33.042 (24.001%)	0.538 ± 0.018 (1.894%)	9.356 ± 0.333 (19.76%)
1025th Gly to Arg	173.266 (-0.876%)	399.710 ± 7.152 (-20.708%)	0.388 ± 0.009 (-26.515%)	5.774 ± 0.135 (-26.089%)
1049th Gly to Ser	174.173 (0.435%)	473.541 ± 20.619 (-6.062%)	0.565 ± 0.012 (7.008%)	6.635 ± 0.078 (-15.067%)
1022nd and 1025th	174.019 (0.347%)	535.427 ± 18.195 (6.215%)	0.446 ± 0.009 (-15.530%)	9.398 ± 0.297 (20.302%)

336

Table S2. Rigidity and elastic modulus of the wild-type collagen.

Cases	Radius Å	Young's Modulus I Gpa	Bending rigidity EI $pN \cdot nm^2$	Young's Modulus II Gpa	Torsional rigidity $pN \cdot nm^2$	Shear modulus Mpa
Wild-type collagen	6.437	6.716 ± 0.171	917.647 ± 25.875	6.805 ± 0.192	135.471 ± 1.476	502.318 ± 5.481