ChemComm

COMMUNICATION

View Article Online
View Journal

Cite this: DOI: 10.1039/d1cc05277b

Received 19th September 2021,
Accepted 19th October 2021

DOI: 10.1039/d1cc05277b

rsc.li/chemcomm

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

Haoyuan Shi, Liming Zhao, Chenxi Zhai and Jingjie Yeo *

The stiffnesses, β-structures, hydrogen bonds, and vibrational modes of wild-type collagen triple helices are compared with osteogenesis imperfecta-related mutants using integrative structural and dynamic analysis *via* molecular dynamics simulations and Markov state models. Differences in these characteristics are strongly related to the unwound structural states in the mutated regions that are specific to each mutation.

Collagen plays a critical mechanical role in connective tissues and bones in mammalian bodies. However, mutations of type I collagen in humans can lead to incurable and debilitating diseases, such as the rare genetic disease, osteogenesis imperfecta (OI).[1,2] Gly substitution mutations in collagen's helical domain is one specific mutation underlying heritable connective tissue diseases for OI,[3] causing susceptibility to bone fractures, bone deformities, and growth deficiencies. Type I collagen is a heterotrimer with two α-1(I) and one α-2(I) chains, exhibiting triple-helical regions with Gly–X–Y repetitive triplets that endow impressive elasticity, toughness, and fracture strength.[2] Gly substitutions disrupt helical folding,[4] change the affinity with other extracellular matrix molecules and cell receptors,[5] and inhibit proper crosslinking.[6,7] These are critical factors for the material behaviours and properties of collagen at the molecular scale, specifically for structural, mechanical and dynamic characteristics. Numerous studies use molecular dynamics (MD) simulations to unravel the mechanical properties[8–10] and structural changes[11,12] in mutated collagen triple helices. However, as these studies generally used non-equilibrium MD techniques, the equilibrium structure–property relationships of wild-type and mutant collagen remain elusive. Systematically understanding the equilibrium mechanical behaviour of OI-related mutant collagen renders deeper molecular insights into the underlying biophysics of the mutation, thus potentially providing a direct path for targeted drug design and molecular therapies for OI and other related diseases. Furthermore, the equilibrium MD methods used herein are directly relevant for rigorously determining the structural traits and dynamics of scleroprotein mutations in general.

We analyse the structure and dynamics by combining equilibrium MD simulations, Markov state models (MSMs), and normal mode analysis (NMA) on a wild-type collagen sequence from human type I collagen as well as four OI-related mutated collagens. The axial, bending, and torsional stiffnesses of every collagen molecule are determined from natural fluctuations using equilibrium MD simulations. Equilibrium structures for further structural analysis is obtained from MSMs by constructing a Markov chain on short-time configuration spaces. Finally, NMA is performed to explore the vibrational modes of the collagen molecules in detail. Interestingly, although all collagens unwind at the mutation sites, these subtle structural changes lead to markedly different mechanical properties and vibrational modes.

Fig. S2, (ESI†) shows the sequence of the wild-type collagen. Triplets of GPP amino acids were added at both ends of the peptides to avoid potential effects at the terminals. The triple helix structure of wild-type collagen was generated by THeBuScr,[13] and the mutated collagens were then generated by replacing Gly at the corresponding sites in α-1(I) chains using QwikMD plugin[14] in VMD[15] (see ESI† for details). To study the natural dynamics of collagen, we performed MD simulations in NAMD[16] with the all-atom CHARMM36m forcefield[17] for each collagen structure in the *NPT* ensemble at 310 K for 100 ns after energy minimisation (see ESI† for details). From the time evolution of each system's potential energy (Fig. S1, ESI†), we considered each system to be well-equilibrated by 60 ns, hence the last 40 ns of each simulation with 2000 conformational ensembles was used for subsequent analyses.

Naturally occurring thermal fluctuations of collagen molecules in equilibrium MD simulations are correlated with the axial, bending, and torsional stiffnesses based on the equipartition theorem:[18]

$$\frac{1}{2}k_{\mathrm{b}}T = \frac{1}{2}k_x\langle(\Delta x)^2\rangle$$

J^2 *Lab for Engineering Living Materials, Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY, 14853, USA.*
E-mail: jingjieyeo@cornell.edu

† Electronic supplementary information (ESI) available: Methods and analysis sections, and additional tables, figures and movies. See DOI: 10.1039/d1cc05277b

T is the temperature, k_b is the Boltzmann constant, and x is a specific geometrical property of the molecule: the length, the end deflection in a specific direction, or the twist angle relative to the reference, thereby corresponding to the axial, single directional bending, and torsional stiffness of k_x, respectively. Δx is the difference between the instantaneous and average values. $(\Delta x)^2$ is the mean square difference that describe the fluctuations at the averaged state, calculated from the dynamic trajectories of equilibrium MD simulations. Fig. S3, (ESI†) shows the scheme of calculating the instantaneous collagen length, end deflection, and twist angle. The collagen length and twist angle are the sum of the length and twist angle of every single GXY unit, respectively, while the end deflection is the distance between the end coordinates of the instantaneous structure and the ensemble-averaged structure after aligning the first GXY unit of all ensembles (Fig. S4, ESI†). The details are provided in ESI.†

To process the data in an unbiased manner, the moving block bootstrap method[19] was used for data analysis before applying the formula for the equipartition theorem. This method provides a confidence interval of three standard deviations (ESI† and Fig. S5). Fig. 1 and Table S1 (ESI†) show the stiffnesses of the wild-type and mutated collagens. The axial, bending, and torsional stiffnesses of wild-type collagen are 504.098 ± 12.726 pN nm^{-1}, 0.528 ± 0.015 pN nm^{-1}, and 7.812 ± 0.084 pN nm^{-1}, respectively. Since the wild-type collagen triple helix maintains a stable structure resembling that of a prismatic rod, the Euler–Bernoulli beam theorem can be coupled with the equipartition theorem to determine the rigidities or elastic moduli (Table S2 and Methods in ESI†).[18,20] This approximation was also applied to other rod-like structures, including actin filaments, DNA strands, or polymeric fibrils with high length-to-diameter ratios.[20,21] The longitudinal Young's modulus of this wild-type collagen is 6.716 ± 0.171 GPa, within a reasonable range from previous simulation and experimental work.[8–10] The in-plane shear modulus of the collagen triple helix is first put forward in our work to be 502.318 ± 5.481 MPa. Compared with wild-type collagen, the mutated collagen structures unwound in the Gly mutation sites and expanded the radius of the backbone of the two GXY units before and after the mutation sites (Fig. S6, ESI†). These highly localised conformational shifts destabilised the collagen structures and led to significant changes in stiffness. The collagen with the 1025th site mutation had approximately twice the average potential energy compared to the wild-type collagen, while other mutated collagens exhibited similar potential energies (Fig. S1, ESI†). This energy difference signified the severe disturbance in the structure introduced by Gly substitution mutations at the 1025th site.

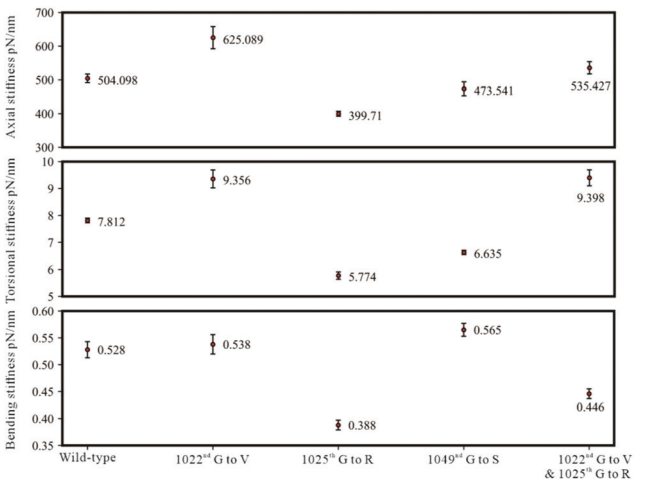

Fig. 1 Axial, bending and torsional stiffness of the wild-type and four mutated collagens.

Gly mutations in the 1025th site reduced all stiffnesses by over 20%, whereas in comparison, the 1049th site mutation had a significantly lower reduction with a slight increase in bending stiffness (Fig. 1 and Table S1, ESI†). However, the 1022nd site mutation led to a 20% increase in axial and torsional stiffness compared to the wild-type collagen. Moreover, the collagen with mutations at both the 1022nd and 1025th sites had stiffnesses that were intermediate compared to those two individual mutations (Fig. 1). The stability of the α-carbon atoms in the collagen molecules was examined using the root mean square fluctuations (RMSFs). Most of the α-carbon atoms in Gly had the lowest RMSFs in its GXY unit, indicating the vital role of Gly in maintaining a stable collagen's triple helical structures (Fig. S7, ESI†). Thus, for the 1025th and 1049th site mutation, unwound triple helical structures had increased RMSFs in that region. However, consistent with the stiffness results, the collagen with 1022nd site mutation had decreased RMSFs in the unwound region, especially for chain B, with relatively low but smooth fluctuations in that area.

To explore the structure–property relationship in collagen, we constructed MSMs for each mutation. MSMs are statistical models for extracting lengthy dynamics information while obtaining the expectation values of stationary and kinetic quantities by integrating independent trajectories.[22] To reduce the error of MSMs resulting from the scarcity of the data in 2000 ensembles, we re-obtained 40 000 ensembles in 4 × 100 ns for each mutation (see ESI† for details). To verify the consistency between the previous 2000 ensembles in 40 ns with the trajectories fed into MSMs, we projected 2000 ensembles using TICA onto the spaces constructed by the MSMs data sources. Demonstrating consistency between the two data sources, almost all configurations existing in the 2000 ensembles were projected onto the basins formed by MSM data, thereby verifying the commonality of the structural changes between the 2000 ensembles and the 40 000 ensembles (Fig. S8, ESI†).

Mutated collagens tended to have larger quantities of β-structures since Gly residues prevent their formation, which can lead to harmful aggregates.[23] Here, the β-structures of typical microstates were quantified for each mutation. The stationary probabilities of every microstate were obtained from the MSMs, from which seven most probable states were selected for β-structure analysis (Fig. S9, ESI†). These seven states were generally in the global minima, with the minor exception of one or two states, implying that the MD-sampled structures reflected the β-structures at equilibrium (Fig. 2a). For each microstate, we sampled 100 frames to compute the extensiveness of β-structures in a specific segment, together with the means

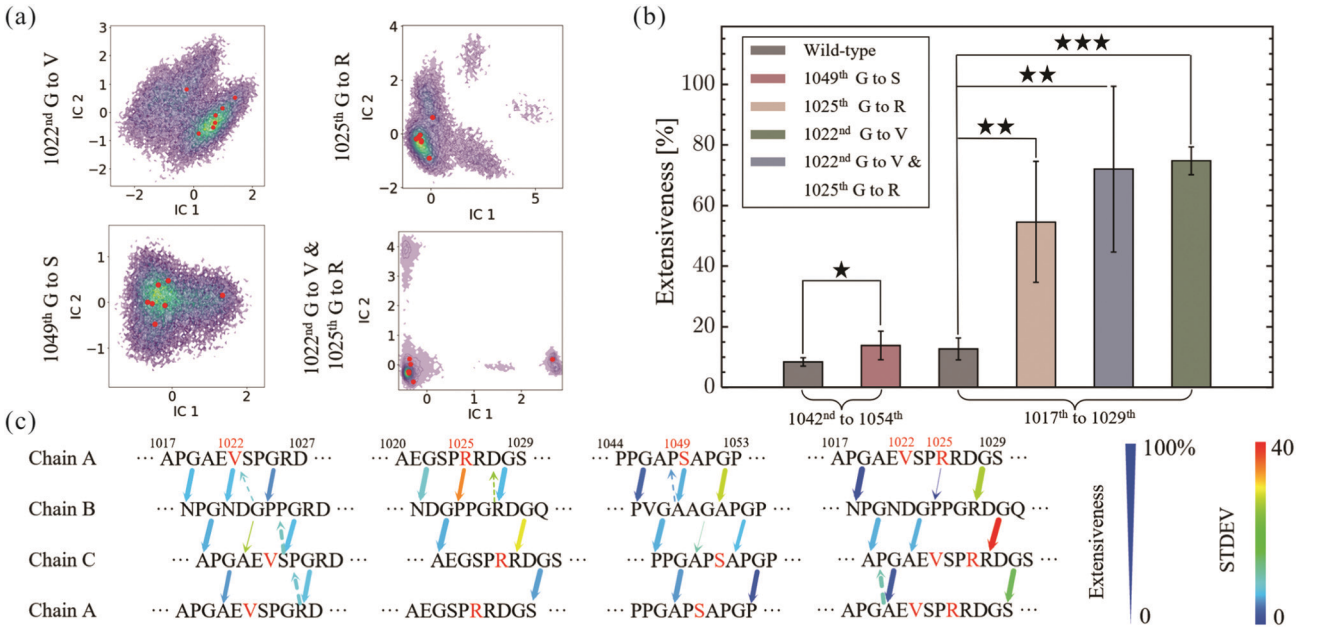

Fig. 2 (a) Distributions of the selected microstates (red dots) in the projected spaces formed by the first two independent components. Coloured areas denote the density of all data. (b) Bar chart of the extensiveness of all mutation cases. Extensiveness is defined as the percentage of frames out of 100 frames that possess β-structures. Stars indicate the *p*-value (One star: 0.05 to 0.01; Two stars: 0.01 to 0.001; Three stars: below 0.001). (c) Interchain H-bonds from donors to acceptors (arrows). Thickness and colour reflect the extensiveness and the standard deviation. The original G-to-X H-bonds and newly formed H-bonds after mutation (Fig. S11, ESI†) are denoted by the solid line and dash line, respectively.

and the standard deviations of the seven extensivenesses for each mutation (Fig. 2b). The significant differences among the extensivenesses, denoted by star symbols, were computed by the two independent samples *t*-test. For the mutation in the 1049th site, the relatively high *p*-value, compared with the other mutation cases, indicated that the extensiveness of the β-structures did not significantly increase compared to the reference segment in the wild-type collagen. For the 1025th site mutation, although considerable β-structures emerged, the high standard deviation illustrated the instability of β-structures among the microstates. On the contrary, the 1022nd site mutation produced greater quantities of β-structures that were more stable. In comparison, mutations on both the 1022nd and 1025th sites led to an intermediate result: there was large extensiveness over 70% and the most severe instability. Therefore, the silenced fluctuations and increased stiffnesses in collagens with 1022nd site mutation was due to the formation of stable β-structures in the unwound region, while unwound structures in other mutated collagens exhibited instabilities.

The collagen triple-helix is stabilised by the interchain H-bonds linking the secondary amine of a glycine residue with a peptide carbonyl group in the previous X residue. Thus, we examined those H-bonds' viability and detected other H-bonds that increased due to mutations (Fig. 2c). For example, the 1025th site mutation broke three original H-bonds and perturbed two original H-bonds, likely causing the significant reduction in stiffnesses. Conversely, in the 1022nd site mutation, an original Gly-to-X H-bond was lost, while the others remained favourably extensive and stable. Furthermore, three new H-bonds were detected in most trajectories. The stability of the original and emergent H-bonds resulted in stable β-structures and the increased stiffnesses. As for the mutations in the 1022nd and 1025th sites, double-site mutations weakened and even broke multiple G-to-X H-bonds. Though having high extensiveness, the remaining H-bonds became precarious. This phenomenon was consistent with the high RMSFs in the unwound region (Fig. S7, ESI†). Finally, for the 1049th site mutation, an original H-bond was lost while a new one emerged, and the other H-bonds remained consistent with the wild-type collagen, implying that the stiffnesses were not expected to change a lot. Overall, the G-to-X H-bonds linking Chain A and C tended to break prior to the breakage in the other H-bonds, which was consistent with the higher RMSFs of Chain A and C in the unwound region compared to Chain B (Fig. S7, ESI†). Similar results of H-bond changes were observed in 2000 ensembles (Fig. S10, ESI†). Furthermore, the formation of most new H-bonds primarily involved Hyp, Arg, Ser, and Asp amino acids, which have polar side chains that serve as the donors and acceptors. Thus, it is reasonable to expect that the chemical structures of the side chains govern the formation of new H-bonds to a large extent.

Mutated collagens' altered stiffnesses and conformations consequently changes vibrational modes, which may further perturb the ability to self-assemble into fibrils and disrupt the mechanical functions. To examine this, elastic network models were constructed with α-carbon atoms interconnected by uniform springs and the normal modes for equilibrated collagen structures obtained by MSMs were analysed (see ESI† for details). Since the energy was equally distributed among the modes, lower-frequency modes exhibited larger oscillation amplitudes, dominant in the accessible vibrations within the molecule.[24] Here, representative

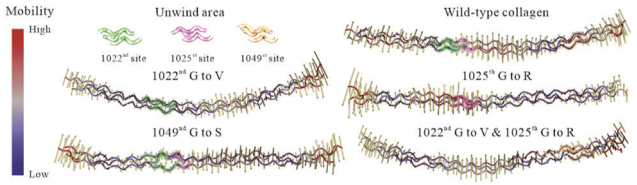

Fig. 3 Dynamic snapshots of collagens show the relative vibration amplitudes and direction of each α-carbon based on the linear combination of the first 20 lowest-frequency modes. Yellow arrows show the amplitude and direction of each α-carbon.

dynamics of the collagen models were obtained as a linear combination of the first 20 lowest frequency modes, scaled by mode amplitude, since higher frequency modes were negligible with minimal effects on dynamics.

The dynamics of wild-type collagen were characterised by a symmetrical vibration with a characteristic twist in the N-terminal region (Fig. 3 and Movie S1, ESI†). However, such dynamics were subverted in all mutated collagens. For the collagen mutated in the 1025th or 1049th sites, the mobility near the mutation sites significantly increased, dominated by a twist vibration (Movie S3 and Movie S5, ESI†). This phenomenon supported our earlier finding of decreased torsional stiffnesses in these two collagen models. In contrast, the 1022nd site mutation silenced the dynamics in the adjacent region (Movie S2, ESI†), consistent with the results in MSMs that formed stable β-structures. The overall dynamics were completely altered in the double site mutations, tending to have more irregular vibrations due to the large unwound region in the structures (Movie S4, ESI†).

In conclusion, our study harnessed the natural fluctuations in equilibrium MD simulations to calculate the stiffnesses of wild-type collagen and several OI-related Gly substitution mutations, identifying remarkable differences in the stiffnesses for various mutated collagen. These differences were attributed to the unwound structural characteristics after mutations. For example, Gly mutations in the 1022nd site led to increased stable short β-structures with new H-bonds, resulting in a more stable local state with overall high stiffness, whereas mutations in the 1025th site disrupted existing H-bonds and decreased the overall number of H-bonds, leading to an erratic unwound structure with more intensive fluctuations. Despite these differences, the vibrational modes of equilibrium structures for mutated collagens changed significantly, having more irregular vibrational modes than the wild-type collagen. Future research will focus on the mutated collagens' behaviours at larger scale in assembled structures. Eventual connections between these molecular characteristics with collagen's macroscale structure–property–dynamic relationship will reveal the mechanisms of genetic diseases on collagen functions from a molecular scale for developing targeted, individualised molecular therapies.

The authors acknowledge support from the US National Science Foundation (Grant No. 2038057), Cornell University's Cornell-China Center, and faculty startup grant. The authors also acknowledge computational resources provided by the XSEDE program under Grant TG-MAT200004 and TG-BIO210063, and computational resources provided by Graphite and G2 cluster from Cornell University.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 H. Kang, A. C. S. Aryal and J. C. Marini, *Transl. Res.*, 2017, **181**, 27–48.
2 A. Forlino and J. C. Marini, *Lancet*, 2016, **387**, 1657–1671.
3 N. Hammami-Hauasli, H. Schumann, M. Raghunath, O. Kilgus, U. Lüthi, T. Luger and L. Bruckner-Tuderman, *J. Biol. Chem.*, 1998, **273**, 19228–19234.
4 J. Engel and D. J. Prockop, *Annu. Rev. Biophys. Biophys. Chem.*, 1991, **20**, 137–152.
5 M. C. Erat, D. A. Slatter, E. D. Lowe, C. J. Millard, R. W. Farndale, I. D. Campbell and I. Vakonakis, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 4195–4200.
6 W. A. Cabral, I. Perdivara, M. Weis, M. Terajima, A. R. Blissett, W. Chang, J. E. Perosky, E. N. Makareeva, E. L. Mertz, S. Leikin, K. B. Tomer, K. M. Kozloff, D. R. Eyre, M. Yamauchi and J. C. Marini, *PLoS Genet.*, 2014, **10**, e1004465.
7 L. Arseni, A. Lombardi and D. Orioli, *Int. J. Mol. Sci.*, 2018, **19**, 1407.
8 S. Varma, J. P. R. O. Orgel and J. D. Schieber, *Biophys. J.*, 2016, **111**, 50–56.
9 A. Ghanaeian and R. Soheilifard, *J. Mech. Behav. Biomed. Mater.*, 2018, **86**, 105–112.
10 J. W. Bourne, L. Shi and P. A. Torzilli, *J. Mech. Behav. Biomed. Mater.*, 2020, **108**, 103835.
11 J. Hartmann and M. Zacharias, *PLoS Comput. Biol.*, 2021, **17**, e1009079.
12 K. A. Clements, A. M. Acevedo-Jake, D. R. Walker and J. D. Hartgerink, *Biomacromolecules*, 2017, **18**, 617–624.
13 J. K. Rainey and M. C. Goh, *Bioinformatics*, 2004, **20**, 2458–2459.
14 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.
15 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
16 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.
17 K. Vanommeslaeghe and A. D. Mackerell, *Biochim. Biophys. Acta, Gen. Subj.*, 1850, **2015**, 861–871.
18 I. Adamovic, S. M. Mijailovich and M. Karplus, *Biophys. J.*, 2008, **94**, 3779–3789.
19 H. R. Kunsch, *Ann. Stat.*, 1989, **17**, 1217–1241.
20 A. Shamloo and B. Mehrafrooz, *Cytoskeleton*, 2018, **75**, 118–130.
21 D. Boal, *Mechanics of the Cell*, Cambridge University Press, Cambridge, 2nd edn, 2012.
22 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.
23 M. D. Shoulders and R. T. Raines, *Annu. Rev. Biochem.*, 2009, **78**, 929–958.
24 I. Bahar, T. R. Lezon, A. Bakan and I. H. Shrivastava, *Chem. Rev.*, 2010, **110**, 1463–1497.

