

Effects of terminal tripeptide units on mechanical properties of collagen triple helices

Milad Masrouri^{1,2} Zhao Qin^{1,2,3 *}

1 Department of Civil and Environmental Engineering, Syracuse University, Syracuse, NY 13244, USA

2 Laboratory for Multiscale Material Modelling, Syracuse University, 151L Link Hall, Syracuse University, Syracuse, NY 13244, USA

3 The BioInspired Institute, Syracuse University, NY 13244, USA

*Materials & Correspondence should be addressed to Z.Q. (zqin02@syr.edu)

KEYWORDS: Molecular dynamics simulations, Collagen triple helix, sequence-mechanics relationship, Hydrogen-Bonds

ABSTRACT: Collagen, a vital protein that provides strength to various body tissues, has a triple helix structure containing three polypeptide chains. The chains are composed mostly of a tripeptide of glycine (G), proline (P), and hydroxyproline (O). Using molecular dynamics simulations and theoretical analysis, the study examines the mechanical response of collagen triple helix structures, made up of three different tripeptide units, when subjected to different fracture loading modes. The results show that collagen with GPO tripeptide units at their C-terminal are mechanically stronger than the POG and OGP units with a single amino-acid frame shift. Our work shows that the N-terminal has less effect on collagen fracture than the C-terminal. The differences in mechanical response are explained by the heterogenous rigidity of the amino acid backbone and the resulting shear lag effect near the terminal. The findings have potential applications in developing tough synthetic collagen for building materials and may stimulate further studies on the connection between terminal repeats and the mechanical-thermal behavior of other structural proteins such as silk, elastin, fibrin, and keratin.

Submitted to: *Extreme Mechanics Letters*

Introduction

Understanding the sequence-structure relationship and the mechanical properties of proteins is important for a wide range of applications, including drug development, biotechnology, and materials science[1]. As the most abundant structural protein for mammals, collagen is the main structural component of skin, bone, tendons, ligaments, and cartilage[2]. It provides strength and support to various tissues in the body and helps to keep them elastic and resilient[3][4][5]. The crucial importance of collagen as a scaffold for animals demands a diverse of essential characteristics, including thermal stability, mechanical strength, and the ability to engage in specific interactions with other biomolecules[6]. Understanding how such properties are derived from the fundamental structural unit of collagen, the triple helix, requires a thorough knowledge of the mechanisms underlying triple-helix structure and stability[7]. Terminal caps, the external amino acids at the N- and C-ends, are often thoroughly studied and engineered for the proteins' overall solubility and stability[7]. As opposed to common α -helix and β -sheet secondary structures[8][9][10][11], there has been limited theoretical study of terminal capping for repeat proteins not conforming α -helix and β -sheet folding structures, such as collagen triple-helix, preventing from understanding how the modification of the N- and C-terminals affect the stability and localization of the protein. The molecular structure of collagen is a right-handed triple helix with a length of ~ 300 nm and a diameter of 1–2 nm[7]. The sequence of amino acids in collagen is given by the repeating tripeptide unit, forming a long repetitive sequence of Glycine-X-Y. X and Y positions are often occupied by proline (Pro, P) and hydroxyproline (Hyp, O), respectively[7]. Glycine (Gly, G) is essential in every third position to allow for a compact packing of the triple-helix[12]. The underlying structure of collagen, triple helix, is formed by three long interwinding polypeptide chains. The triple-helix is stabilized by interstrand hydrogen-bonding,

N-H(Gly)...O=C(Pro), formed between the amide of Gly and the carbonyl of Pro of the neighboring strand[13] (Figure 1a).

Studying native collagen is challenging due to its heterogeneous structure and the difficulty of isolating it[14][15]. Therefore, collagen model peptides (CMP), short sequence of amino acids with 6-10 repeating tripeptides that can bind specifically to unfolded collagen chains, have been employed as synthetic substitutes[16][17][18]. CMPs are used in various biomedical applications, such as tissue engineering and drug delivery, because of their ability to target collagen-rich tissues[19][20]. They can also be designed to understand the structures and functions of natural collagen. By creating CMPs that mimic the structure of collagen, researchers can learn more about how natural collagen works, how it contributes to the support and structure of different tissues, and how it might be influenced by different factors, such as age, disease, and environmental factors[21][22]. They have provided profound understanding of the structure–property relationship of collagen triple helices at the molecular level[23][24][25][26]. The repetitive nature of collagen sequence allows for the synthesis of CMPs in three different tripeptide frames. A frame is defined as a tripeptide unit that can be made of GPO, POG, and OGP[27], which is repeated to form a CPM structure (Figure 1b). Traditionally, most studies on collagen peptides have favored Gly as the C-terminal residue[24][3] since couplings with an activated Gly residue proceed smoothly, and the achiral Gly cannot experience epimerization[28][29]. Besides, resins preloaded with Gly are readily available for the solid phase peptide synthesis of CMPs [30][31]. For many years, (POG)_n and (GPO)_n were thought to be compatible[32][33][34][35]. These collagen mimetic peptides share the same sequence at the middle but differ in the terminal amino acid sequences. Unlike the OGP frame whose examples are rare[14], CMPs with the GPO and POG frames have been used and studied commonly[24][36][17]. However, despite the substantial

endeavors made to comprehend the relationship between structural modifications and the stability of the collagen triple helix, the effect of collagen frame and terminal groups on thermal stability of CMPs had not been extensively explored until recent years[37][38][39]. It was shown that frame-shifted CMPs develop triple helices with significantly different melting temperatures (as high as 10 °C in difference)[40]. It means that CMPs with variations in terminal repeats exhibit distinct end structures and stabilities. It was also demonstrated that the maintenance of the triple-helix's end structure in collagen is primarily facilitated by cross-chain hydrogen bonding mediated by the terminal repeat[40]. These findings raise the question whether the mechanical properties of collagen triple helix are affected by the frame shift, which leads to different sequence of amino acids at the terminals.

Mechanical damage of collagen fibrils and fibers in connective tissue has been extensively studied by both experiments and computational modeling[41][42][43][44][45]. However, there has been little definitive proof of permanent damage to collagen at the molecular level so far, partly due to the limitations of collagen imaging modalities and the unavailability of markers that can recognize unfolded collagen molecules[15]. Using collagen hybridizing peptides, Zitnay et al. have found directly for the first time that irreversible unfolding of collagen molecules is a key feature of the mechanical failure process of tendon fascicles[15], suggesting the sequence near the crosslinking site can be crucial to the collagen mechanics in loading. The terminal-sequence-mechanics relationship of collagen triple helices, nevertheless, is still poorly understood[46]. To address this point, we need to understand how a frame shift and terminal amino acids sequence contributes to mechanical stability of collagen. Up to now, to the best of our knowledge, studies examining mechanical damage to collagen fibrils have not investigated how mechanical properties of a collagen change if their terminals sequence is altered.

Here, we aim to explore the effect of the terminal residues on the mechanical properties of capped frame-shifted collagen triple helices. Using molecular dynamics (MD) simulations, we give insights into the impact of the frame on mechanical behavior of collagen triple helices at the molecular level. Among all the various ways that can unfold collagen molecules, we consider two likely loading mechanisms for irreversible triple helix unfolding (Figure 1c). The loading force and pin apply to the alpha-carbon atoms of the protein backbone leads to mode I and mode II fracture, representing the force that is conducted to the collagen molecule through covalent cross links from its neighboring molecules within a fibril in tearing and tension, respectively. We found out that the interstrand H-bonding mediated by the terminal repeats as well as the difference in stiffness of Gly, Pro, and Hyp residues lead CMPs with (GPO) frame at their C-terminal to possess the highest mechanical properties. Not only can our results play a part in designing future collagen model peptides, assemblies, and materials, but give further insights into the sequence-structure and sequence-mechanics relationship of several intricate proteins.

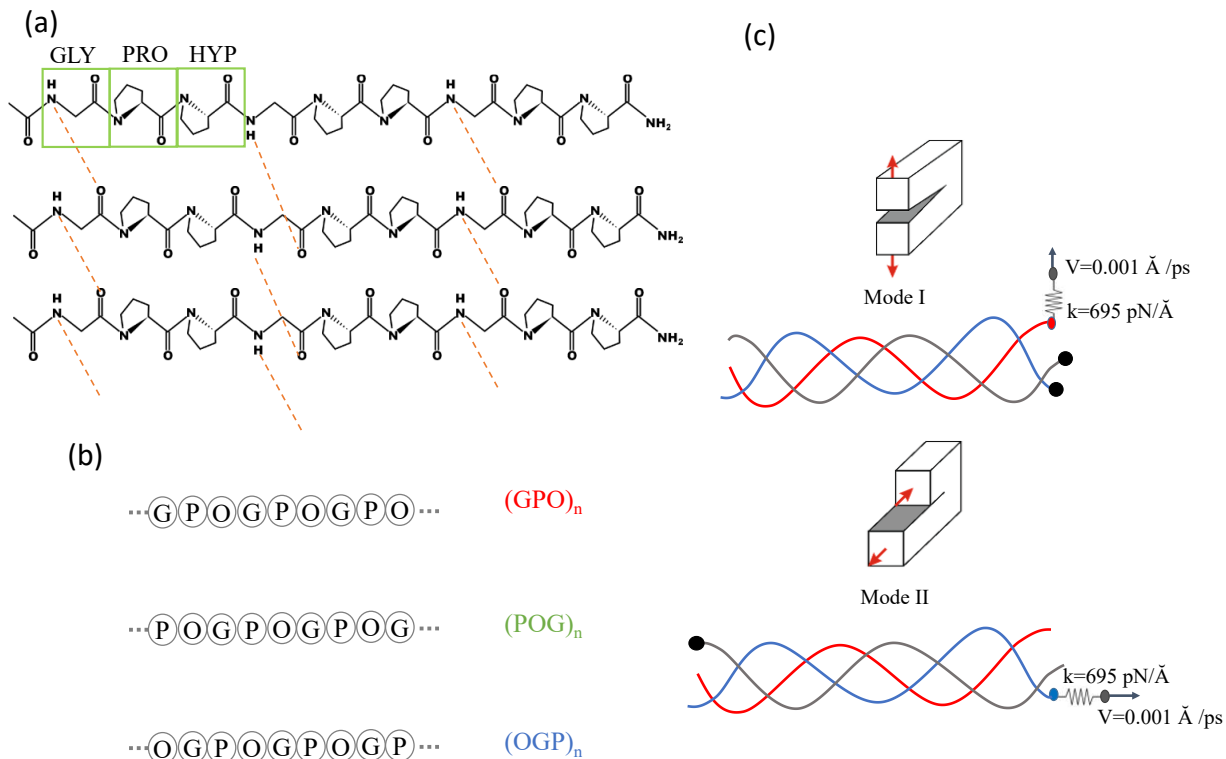


Figure 1. (a) The molecular structure of the collagen triple-helix: Both Pro and Hyp lack an N-hydrogen atom, so the interchain hydrogen bond (the orange dotted lines) is formed only between the amide of Gly and the carbonyl of Pro. (b) The three possible repeating tripeptide units of a collagen model peptide sequence. (c) Two possible damage mechanisms of collagen as a result of mediated stress transferred by cross links.

Results and Discussion

Different CMP strength governed by a frame shift

We investigate whether and how the frame shift affects the mechanics of capped triple helices in light of the recent finding that capped frame-shifted collagen triple helices exhibit noticeably different thermal stability[14][40]. We use fully atomistic molecular dynamics to relax the

collagen model peptides made up of (GPO), (POG), and (OGP) tripeptide units with different number of amino acids (see the method section for details) for a total of 20 ns. To understand how the different tripeptide units affect the mechanical response, we directly unfold the structures by using steered molecular dynamics (SMD) to apply loading forces that simulate mode I and mode II fractures and record force-deformation data. Mode II of failure or shearing is expected to happen when a single strand of the triple helix is pulled out because of a longitudinal or uniaxial loading, which results from molecular sliding between neighboring collagen molecules, mediated by stress transfer through crosslinks that connect collagen strands of neighboring triple helices. On the other hand, when the loading is applied in perpendicular direction, the stress transfer by crosslinks leads to a failure of mode I or tearing failure.

To quantify the effect of the sequence on mechanical properties of the models, we compute the maximum mechanical strength and the energy dissipation. The maximum strength refers to the peak force experienced by the structure during deformation, representing its highest resistance capacity. Energy dissipation is determined by integrating the area beneath the force-deformation curve. To ensure comparability, we normalized the energy dissipation by dividing it by the surface area of the deformed region. Besides, we use an elastic network model to show that the heterogeneous rigidity of amino acids backbone impacts the mechanical properties of collagen triple helices.

To avoid the effects of the length and number of amino acids on mechanical properties, we run the simulations to apply the force on three CMPs with the same number of amino acids but different repeating sequences. With the intention of doing so, we built three collagen triple helices established by every single potential combination of CMP frame models, *i.e.*, (GPO)₇, (POG)₇, and (OGP)₇. Again, these CMPs have the same chain length and share the same sequence at the middle but differ in the terminal amino acid sequences. Figure 2 compares force-deformation curve, energy dissipation, and mechanical strength for these polypeptides. We visualized the

structure at the point of minimum and maximum forces as well as at the ending point for (GPO)₇. At the point of minimum force (point i), the structure appeared folded, and interstrand hydrogen bonds were observed to play a crucial role in holding the chains together. At the point of maximum force (point ii), the structure experienced irreversible failure, indicating a critical point where the applied force reached its peak. This point represented a significant deformation or breaking of intermolecular interactions within the structure. Following the maximum force point, the force gradually decreased until reaching the ending point (point iii).

Although the models have the identical chain length and amino acid quantity and composition, the force-deformation curves illustrate that GPO frame has the higher energy dissipation as well as mechanical strength in comparison with OGP and POG counterparts, which is in good agreement with the reported experimental results for melting temperature of (GPO)₇, (POG)₇, and (OGP)₇[38][40]. Figure 2a indicates that the GPO frame leads to irreversible failure at 1.6 nN, while POG and OGP frames can withstand 1.4 and 1.3 nN respectively. GPO-repeating sequences can establish an additional set of inter-helix H-bonds, thereby increasing stability and mechanical properties of collagen triple helix. Our data suggests that POG is unable to form stable H-bonds at the C-terminal since Gly is flexible, interfering the interaction between H-bond donor and acceptor. As a result, mechanical properties and stability of CMPs holding POG tripeptide unit is reduced. The results for Mode I are shown in Figure 1S, where we observe the same trend, meaning that GPO frame possesses the highest mechanical properties while being unfolded by Mode I of fracture. Our observations are confirmed by repeating the same procedure for (GPO)₅, (POG)₅, and (OGP)₅ for Mode I and II fracture tests. These results are provided as Figure 2S and 3S, respectively.

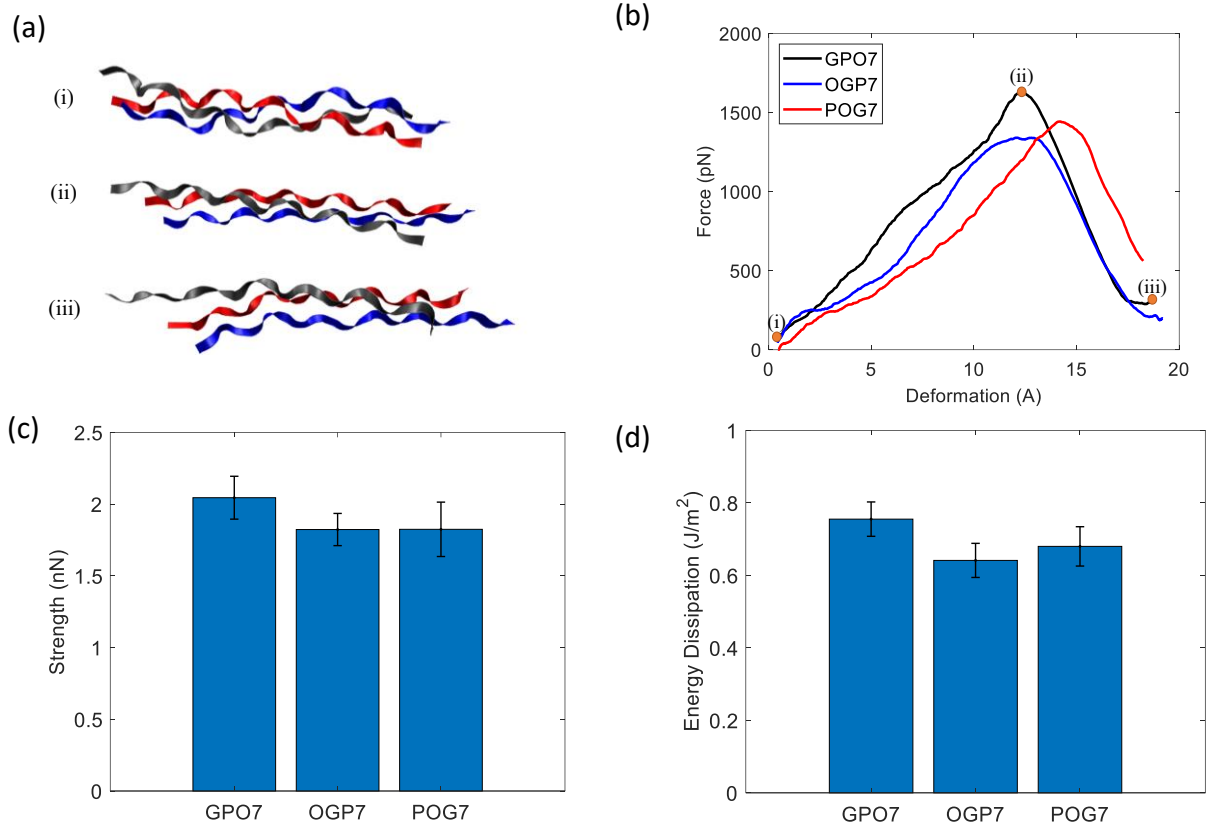


Figure 2. (a) Simulation snapshots at minimum force (i), maximum force (ii) and the end of the deformation (iii). (b) The force-deformation curve for triple helices derived from GPO, POG, and OGP frames for mode II of fracture, and their (c) mechanical strength, and (d) energy dissipation per unit area. The mean value and standard deviation in panel (c) and (d) are obtained from four individual tests.

C-Terminal repeating sequence

It is surprising that the three CMPs of the exact same length, highly similar sequences but different terminal residues lead to very different mechanical strength and energy dissipation in loading. Previous research has shown that the terminal residues tend to have greater mobility and B-factor (the displacement of the atomic positions from its mean position) due to a shortage of

interstrand hydrogen bonds and their extension from the helical core [47][48][49]. Starting with (GPO)₇, we sequentially delete O, P, and G residues from C-terminal to build (GPO)_n, (GPO)_nG, and (GPO)_nGP sequences, where n takes values between 2 and 6. The sequence difference between the three models lies at the C-terminal caps repeats, where (GPO)_nGP has an additional Gly and Pro residues and (GPO)_nG contains an extra Gly residue at their C-terminals as compared to (GPO)_n. Figure 3a shows the simulation snapshots and Figure 3b shows the force-deformation curves for different collagen sequences under a mode II molecular fracture (slipping) with loading force along the molecular axis. It is shown that both the ultimate strength and energy dissipation increase with the chain length (number of amino acids), as summarized in Figure 3c and Figure 3d, respectively. Besides the overall trend, it is also clearly shown that (GPO)_n always leads to a higher strength than (GPO)_nG and (GPO)_nGP for $n \geq 4$ (Figure 3c). Moreover, Figure 3d illustrates how (GPO)_n structures dissipate more energy than (GPO)_nG and (GPO)_nGP structures, despite the first one having smaller number of residues than the latter two in the sequence. We repeat the simulation for each sequence eight times and give the mean value and standard deviation in these plots. It is noted that the mechanical strength only exhibits the consistent pattern for $n \geq 4$ (Figure 3c), which may be attributed to the instability of collagen caused by its insufficient length. To support this, it was found that peptides with repeating sequence of (GPO)_n with $n \leq 5$ cannot self-assemble into a stable triple helix without the need of crosslinks [50][51]. We observe that (GPO)_n is mechanically stronger than (POG)_n and (OGP)_n for collagens with sufficient length ($n \geq 4$, Fig. 2, 1S-3S), which qualitatively agree with the recent work where the melting temperature of (GPO)₇ sequence is higher than (GPO)₇G and (GPO)₇GP sequences in experiments[40].

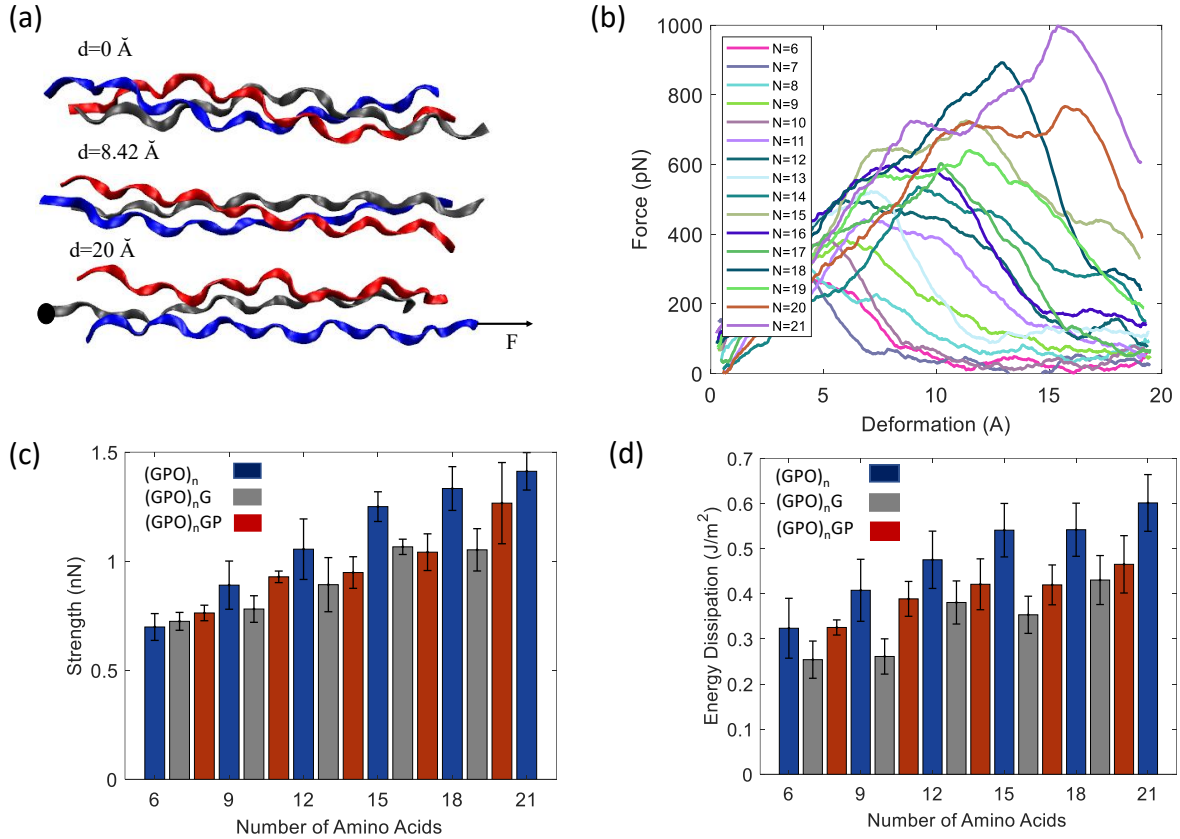


Figure 3. (a) Simulation snapshots at different deformation levels. (b) The Force-deformation curves for CMPs undergoing fracture mode II loading mechanism, N is the length of a collagen chain, where $N=3n$ for $(GPO)_n$, $N=3n+1$ for $(GPO)_nG$, and $N=3n+2$ for $(GPO)_nGP$. (c) Mechanical strength for CMPs with different number of amino acids, which maintain different C-terminal repeating sequence. (d) Effect of number of amino acids on energy dissipation per unit area for collagen model peptides with different C-terminal repeating sequence.

We notice that it is observed the (GPO) frame exhibits a significant lower thermal fluctuation and noticeable smaller radius of gyration compared to (POG) and (OGP) at the C-terminus in our MD simulations, indicating that (GPO) at C-terminus moves less during the thermal oscillations and keeps its size more compact[40], agreeing to our observation for the trend in strength and

energy dissipation. As the structure undergoes deformation, it is generally expected that the radius of gyration will increase. A higher radius of gyration indicates a more extended or unfolded structure. The lower radius of gyration for GPO frame indicates a more compact and less unfolded compared to POG and OGP structure. It also suggests the difference in mechanical strength and energy dissipation can be attributed to its more stable structure at C-terminus. The reason for its higher stability is that the distances and angles between the C-terminal Hyp-NH₂ and Pro-C=O are suitable for forming interstrand H-bonds in GPO structures ending with Hyp-amide. Previous study by Qi et al[40] suggests that the C-terminal Hyp-amide in (GPO)_n is probably responsible for new H-bonding formation, resulting in a more stable and stronger structure, explaining the trend we observe where (GPO)_n has the highest strength and energy dissipation. However, it is still not clear how the extra glycine in (GPO)_nG suddenly destabilizes the triple-helix and decreases the mechanical strength and energy dissipation. To elucidate, at the C-terminal, where conformational constraints are reduced, the backbone of glycine exhibits higher flexibility than proline and hydroxyproline, which has additional bonds between the sidechain and alpha carbon, affecting its flexibility at C-terminal. This flexible nature of glycine appears to impede the Pro...Gly H-bond formation, decreasing the strength and energy dissipation for (GPO)_nG as it is demonstrated in Figure 3c-d. Adding a proline residue increases the strength and energy dissipation since proline is less flexible and leads the (GPO)_nGP structure to be stronger and more stable in comparison with (GPO)_nG. However, (GPO)_nGP lacks the Gly H-bond donor at the C-terminal, leading to preventing the structure from creating the last Pro...Gly H-bonds. This is probably the reason why (GPO)_nGP has less mechanical strength and energy dissipation compared to (GPO)_n.

We observe the same trend when collagen model peptides become unfolded by mode I of fracture. Figure 4S exhibits that for different values of n , $(\text{GPO})_n$ can withstand higher forces than $(\text{GPO})_n\text{G}$ and $(\text{GPO})_n\text{GP}$. When it comes to energy dissipation, $(\text{GPO})_n$ also shows the highest values which is nicely aligned with the findings for mode II fracture mechanism. However, it is noticed that the normalized energy dissipation drops when the number of amino acids increases. This might be caused by the fact that only a limited number of H-bonds carry the load whereas we normalized the energy dissipation by the entire length of the CMPs. In other words, in mode I of fracture, as opposed to mode II, not all the H-bonds are involved while subjecting to the load. For mode II, since the force is applied along the longitudinal axis of the collagen triple helices, all the H-bonds get involved and bear the load.

N-terminal repeating sequence

We serially eliminate G, P, and O residues from N-terminal in order to investigate the impact of frame shift at N-terminal on mechanical properties of collagen triple helices. We build $(\text{GPO})_n$, $\text{PO}(\text{GPO})_{n-1}$, and $\text{O}(\text{GPO})_{n-1}$ polypeptide structures and apply pin and loading force in such a way that the alpha carbon atom of one chain is pulled from the N-terminal while another chain is fixed at the C-terminal (Figure 4a). We notice that changing tripeptide units at N-terminal does not show a considerable influence on mechanical strength and energy dissipation of collagen triple helices. (Figure 4c-d). Moreover, as opposed to what we observe for different frames at C-terminal, the mechanical strength and energy dissipation values do not follow a consistent trend. For example, $(\text{GPO})_n$ has a lower energy dissipation than that of $\text{PO}(\text{GPO})_{n-1}$ and $\text{O}(\text{GPO})_{n-1}$ when $n=7,4,3$, while it shows a higher energy dissipation when $n=6,5$. The same is true for mechanical strength, meaning that $(\text{GPO})_n$ is mechanically stronger than $\text{PO}(\text{GPO})_{n-1}$ and $\text{O}(\text{GPO})_{n-1}$ only if $n=7$. $\text{O}(\text{GPO})_{n-1}$ and $\text{PO}(\text{GPO})_{n-1}$ sequences exhibit higher strength when $n=6,5$ and $n=4,3$, respectively.

The results indicate that the total chain length is dominating the mechanical strength and energy dissipation instead of N-terminal residue sequence, leading to much weaker effect on collagen mechanical stability than the C-terminal sequence. We believe this observation may be attributed to the fact that the N-terminal is acrylated, while the C-terminal is capped with an amide group. In our system, the acrylation of the N-terminal potentially hinders the intermolecular interactions and could lead to less influence on the mechanical strength. On the other hand, the C-terminal cap with an amide group may help stabilize interstrand H-bonds, thereby contributing to the observed mechanical properties.

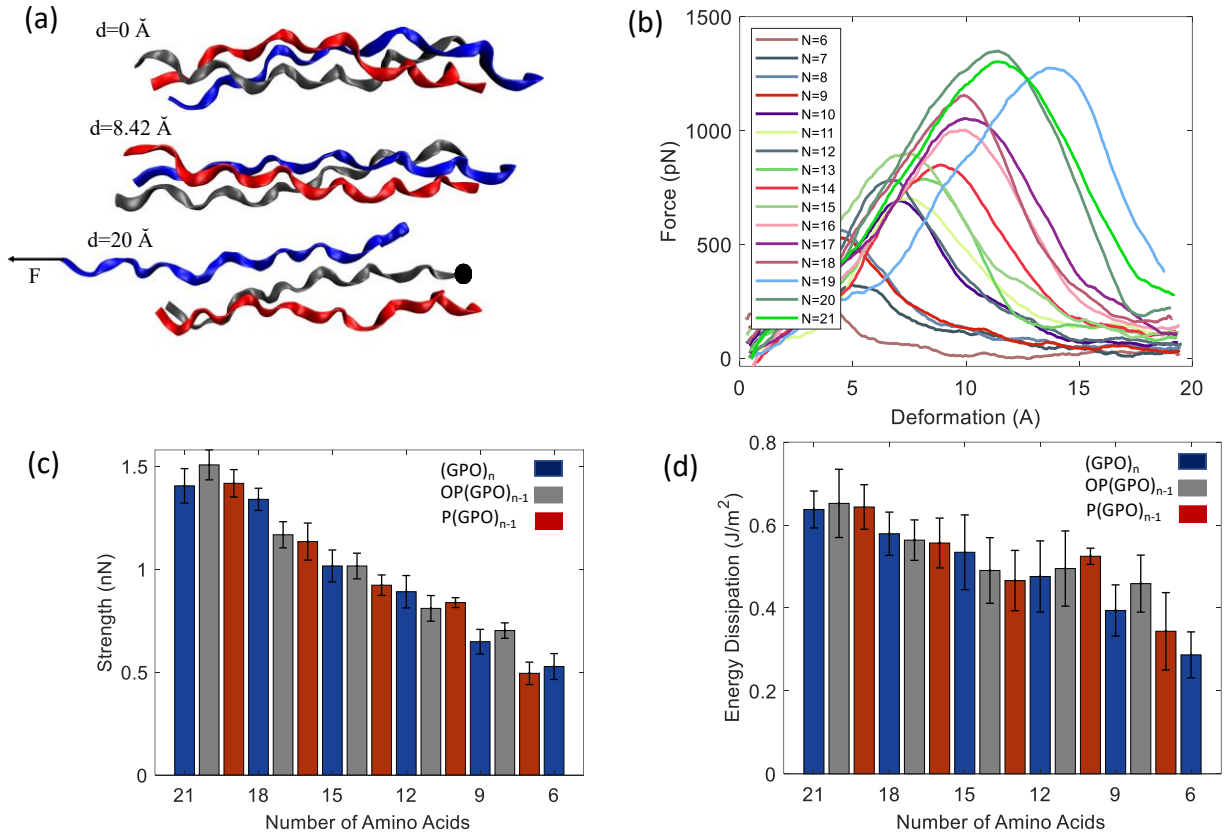


Figure 4. (a) Simulation snapshots at different deformation levels. (b) The force-deformation curves for CMPs undergoing mode II of fracture from N-terminal. (c) Mechanical strength for CMPs with different number of amino acids, which maintain different N-terminal repeating

sequences. **(d)** Effect of number of amino acids on energy dissipation per unit area for collagen model peptides with different N-terminal repeating sequence.

Molecular mechanism by an elastic network

Molecular models and analytical approaches can be used to capture the response of protein structures upon mechanical loading and to illustrate the mechanisms behind their collapse [52][53][54]. In this study, to demonstrate the difference in mechanical response of collagen molecules with different tripeptide units, we use an elastic structural network [53] (details in supplementary information) to model the ideal peptides of different sequence arrangements (chain 1) that bond to a rigid substrate (effect of chain 2 and 3) via H-bonds and subject to lateral loading force (Figure 5a). Unlike Gly, Pro and Hyp residues have rigid rings in their structures (Figure 5b), which increase the stiffness of the backbone of GPO and OGP sequences near the C-terminal in comparison with POG. According to a discretized shear-lag model, the cooperative length, as well as the interfacial strength, is proportional to $\sqrt{K_1/K_2}$ [55] in which K_1 represents the stiffness of the backbone and K_2 is the stiffness of the H-bond (Figure 5S), and therefore the strength of GPO and OGP at the C-terminal can be much higher than POG. It is observed that with the same loading force, the POG sequence yields much larger displacement than GPO and OGP (Figure 5c), which agrees well with the atomistic simulation outcome shown in Figure 2. We also obtain the deformation of each H-bond via the elastic model and equations 1S-4S and observe that the H-bond nearest to the applied force (C-terminal) shows the highest displacement for all three frames while the deformation decays quickly as the amino acid is located further away from the C-terminal. Figure 5d shows that GPO frame is displaced less than two other frames in the vicinity of the C-terminal, leading to later rupture of H-bonds and higher strength in comparison with OGP

and POG counterparts, which is in good agreement with the result of MD simulations shown in Figure 2.

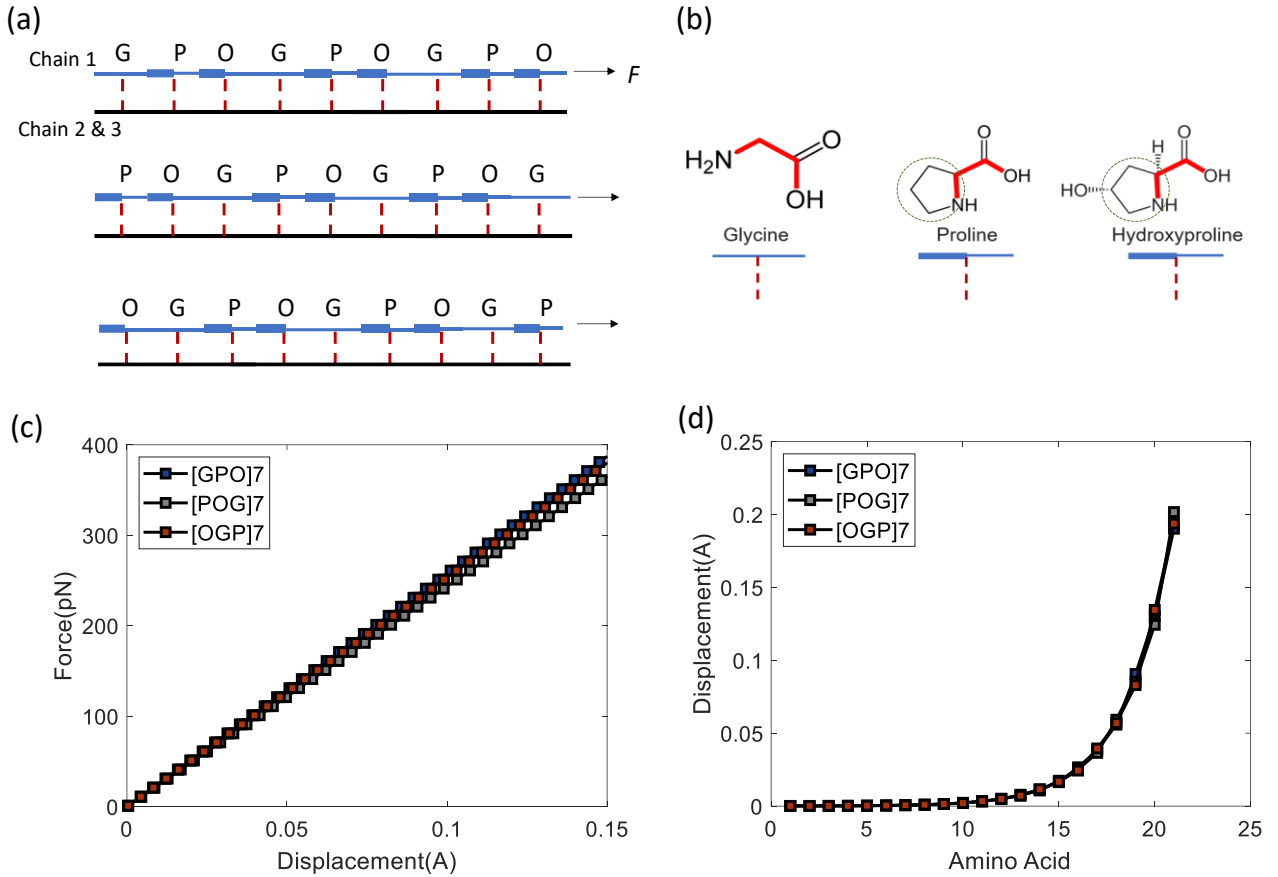


Figure 5. The schematic image depicts the elastic network model used to model different peptide sequences for this study: the backbone by blue solid line and the H-bonds by red dash line. (a) The schematic of different amino acids of different backbone stiffness by an elastic network model. (b) The schematic of different collagen sequences in shear loading by the elastic network model. (c) The force-displacement curve given by the elastic network model for different collagen sequences in shear loading. It is shown that both the effective stiffness, as the ratio between force and displacement, and the area below the force-displacement of POG sequence is lower than that of

GPO and OGP. (d) Displacement of each amino acid for GPO, POG, and OGP frames subjected to a lateral force of 200 pN.

Conclusion

This work analyzes the effects of terminal residues on the mechanical properties of frame-shifted collagen triple helices. The simulations conducted in this study offer valuable insights into the diverse mechanical properties exhibited by collagen-mimetic peptides (CMPs) with distinct terminal tripeptides. Our findings elucidated the influence of hydrogen bonding within the C-terminal region, along with the inherent variations in stiffness associated with Glycine, Proline, and Hydroxyproline, on the mechanical behavior of CMPs. We discovered that the mechanical properties of collagen triple helices partly depend on structural stability. The more stable the structure, the higher the mechanical strength. The collagen model peptides with GPO frame at their C-terminal are mechanically stronger than the POG and OGP equivalents. We also noticed that the N-terminal frame is less critical than the C-terminal frame. Frame shifts at N-terminal do not greatly impact the mechanical properties of CMPs, and there is no discernible pattern observed. Furthermore, the frame rigidity plays an important part as well, and based on our findings GPO frame is more rigid than POG and OGP, leading collagen triple helices with GPO tripeptide units mechanically stronger than the ones holding POG and OGP tripeptide units. However, our simulations cannot explain the effect of collagen-related diseases on the mechanical behavior of the models. For instance, Osteogenesis imperfecta caused by mutation in type 1 collagen results in lower level of collagen and structural defects in collagen. Future investigations may delve into modeling disease-specific scenarios to gain insights into the altered mechanical responses associated with collagen-related disorders.

The value of these results is expected for the development of effective collagen-targeting probes[56][57], and as the building blocks to synthesize collagen materials[58][59]. The mechanical properties and stability of these CMPs is dependent on the repetitive sequence from which they are made. Our work provided a comprehensive direction to choose the appropriate frame exhibiting both high thermal stability and mechanical properties. Moreover, our study may

be an inspiration for new investigation on terminal repeats-mechanics relationship of structural proteins, such as silk, elastin, fibrin, and keratin.

Bones experience daily competition between formation and resorption processes, that may be altered from pathology and aging. Understanding the mechanical rupture of collagen, as the essential building block of bones, from the fundamental length scale and up to the multi-scale structure-mechanics relationship can provide a comprehensive understanding of bone remodeling and are crucial to treatments of bone diseases. Having a molecular model that focuses on the initiation of material failure is also the key to perceive the intimate cross-talks between pathological and physiological agents and the ambitious objectives will be pursued by integrating cutting-edge computational and experimental strategies, that explore modeling, validation and prediction of *in vivo* and *in silico* data from atomistic scale up to the continuum scale.

Methods

Molecular dynamic simulations were carried using the NAMD code with a CHARMM27 all atom energy force field to gain insights into the effects of collagen repeating sequence on mechanical properties.

Modeling and relaxation of the collagen model peptides in MD simulation

Fully equilibrated tropocollagen with a (GPO)₇ sequence was employed as the template. Using a tool of NAMD package, psfgen, a certain number of residues were removed from the C-terminals. The N-terminal was acetylated (Ac-) and C-terminal was capped with amide (-NH₂) group to obtain the structures of interest with Ac-(GPO)_n-NH₂, Ac-(GPO)_nG-NH₂, and Ac-(GPO)_nGP-NH₂ sequence, where n varies between 2 to 6. We repeat the same procedure to sequentially delete residues from N-terminal to build Ac-(GPO)_n-NH₂, Ac-PO(GPO)_{n-1}-NH₂, and Ac-O(GPO)_{n-1}-NH₂ collagen model peptides for n changing between 7 to 3. We use the TIP3P

(Transferable Intermolecular Potential with 3 Points) water model. In the TIP3P model, each water molecule is represented by three interaction sites: two hydrogen atoms and one oxygen atom. The oxygen atom carries a partial negative charge, while the hydrogen atoms carry partial positive charges. The water molecules are hidden for clarity in simulation snapshots representations. Figure 6S demonstrates the collagen model peptide surrounded by water molecules. The molecules are equilibrated in a water solution of 0.1 mol l^{-1} NaCl whose net charge is made zero by adjusting the ratio between cations and anions. The simulation cell (MD box) is maintained at $10 \times 6 \times 5 \text{ nm}^3$ and periodic boundary conditions are implemented to all directions. We employ particle mesh Ewald (PME) algorithm with grid width $< 1 \text{ \AA}$ as a computationally efficient method to precisely calculate and include all the long-distance electrostatics interactions. Each molecular dynamic simulation is performed in the isothermal-isobaric ensemble (NPT) controlled by the Langevin dynamics to reach a constant pressure of 1 atm and temperature of 310 K (100 fs oscillation period and 50 fs damping time scale for pressure control and 5 ps^{-1} for temperature control) with shrinkable volume in the course of the simulation. Simulation time step is 2 fs and we use a rigid bonds model by constraining all the covalent bonds between hydrogen atoms and other heavy atoms. For minimization and equilibration, to begin with, while the side chains and solvent atoms are released, we constrain all backbone atoms in space by an elastic spring with the stiffness of $5 \text{ kcal/mol/\AA}^2$ in all directions. We run 10,000 steps of energy minimization by using conjugate gradient and line search algorithm followed by 20 ns dynamics run for initial temperature. Then, all the applied constraints on the backbone are removed and the simulation is restarted using NPT ensemble in order to fully equilibrate the collagen structures.

Steered molecular dynamics (SMD)

Although there are a variety of ways causing collagen molecules to get unfolded, we consider two different initial boundary conditions to investigate two possible damage mechanisms, fracture

mode I (opening) and fracture mode II (in-plane sliding), as a result of mediated stress transferred by cross links. These cases are scrutinized using steered molecular dynamics (SMD) with a constant loading speed of 0.001 Å/ps and a pulling spring constant of 10 kcal/molÅ². For mode II, the molecule is deformed by fixing one alpha carbon atom at the N-terminal of one chain and pulling an alpha carbon atom along the longitudinal axis at the C-terminal of another chain. To apply load based on fracture mode I, one alpha carbon atom at the C-terminal of two chains are fixed and an alpha carbon atom of the other chain is pulled perpendicular to the longitudinal axis. The force is recorded versus the position of SMD atom to generate the force-deformation curves and extract mechanical strength and energy dissipation of collagen molecules.

Visualization

The utilization of Visual molecular dynamics (VMD) as a post-processing tool for molecular dynamics (MD) simulation results is employed in this study. The software is employed to analyze structural information by providing visual representation of the system under study. This is achieved through the process of iterating through each frame of the MD simulation animation and rendering each atom and chemical bond, thus providing a comprehensive visual representation of the system.

Associated content.

Supplementary materials, including additional figures, are available online.

Author Contributions.

Z.Q. proposed, designed and supervised the research. M.M. prepared the computational model, performed the MD simulations and analyzed the results. M.M. led the manuscript writing and all authors revised and approved the manuscript.

Acknowledgment

The authors acknowledge the National Science Foundation CAREER Grant (Award #: 2145392), the startup funding and Collaboration for Unprecedented Success and Excellence (CUSE) Grant at Syracuse University for supporting the research work.

REFERENCES

- [1] B. Kuhlman, P. Bradley, Advances in protein structure prediction and design, *Nat. Rev. Mol. Cell Biol.* 20 (2019) 681–697. <https://doi.org/10.1038/s41580-019-0163-x>.
- [2] B.J. Bielajew, J.C. Hu, K.A. Athanasiou, Collagen: quantification, biomechanics and role of minor subtypes in cartilage, *Nat. Rev. Mater.* 5 (2020) 730–747. <https://doi.org/10.1038/s41578-020-0213-1>.
- [3] M.D. Shoulders, R.T. Raines, Collagen structure and stability, *Annu. Rev. Biochem.* 78 (2009) 929–958. <https://doi.org/10.1146/annurev.biochem.77.032207.120833>.
- [4] W. Yang, V.R. Sherman, B. Gludovatz, E. Schaible, P. Stewart, R.O. Ritchie, M.A. Meyers, On the tear resistance of skin, *Nat. Commun.* 6 (2015) 1–10. <https://doi.org/10.1038/ncomms7649>.
- [5] D.M. Reilly, J. Lozano, Skin collagen through the lifestages: importance for skin health and beauty, *Plast. Aesthetic Res.* 2021 (2021). <https://doi.org/10.20517/2347-9264.2020.153>.
- [6] C. Dong, Y. Lv, Application of collagen scaffold in tissue engineering: Recent advances

- and new perspectives, *Polymers* (Basel). 8 (2016) 1–20.
<https://doi.org/10.3390/polym8020042>.
- [7] B. Brodsky, A. V Persikov, Molecular structure of the collagen triple helix, *Adv. Protein Chem.* 70 (2005) 301–339. [https://doi.org/10.1016/S0065-3233\(04\)70009-1](https://doi.org/10.1016/S0065-3233(04)70009-1).
- [8] L.S. A.R. Viguera, Stable proline box motif at the N-terminal end of alpha-helices, *Protein Sci.* 8 (1999) 1733–1742. 10.1110/ps.8.9.1733.
- [9] F. FarzadFard, N. Gharaei, H. Pezeshk, S.A. Marashi, β -Sheet capping: Signals that initiate and terminate β -sheet formation, *J. Struct. Biol.* 161 (2008) 101–110.
<https://doi.org/10.1016/j.jsb.2007.09.024>.
- [10] C.J. Pike, M.J. Overman, C.W. Cotman, Amino-terminal deletions enhance aggregation of β -amyloid peptides in vitro, *J. Biol. Chem.* 270 (1995) 23895–23898.
<https://doi.org/10.1074/jbc.270.41.23895>.
- [11] and S.H.G. Flex Freire, Aaron M.Almeida, John D. Fisk, Jay D. Steinkruger, Impact of Strand Length on the Stability of Parallel- Beta-Sheet Secondary Structure, *Angew. Chemie - Int. Ed.* 50 (2011) 8735–8738.
- [12] S.D. Melton, E.A.E. Brackhahn, S.J. Orlin, P. Jin, D.M. Chenoweth, Rules for the design of aza-glycine stabilized triple-helical collagen peptides, *Chem. Sci.* 11 (2020) 10638–10646. <https://doi.org/10.1039/d0sc03003a>.
- [13] N.B. Hentzen, L.E.J. Smeenk, J. Witek, S. Riniker, H. Wennemers, Cross-Linked Collagen Triple Helices by Oxime Ligation, *J. Am. Chem. Soc.* 139 (2017) 12815–12820.
<https://doi.org/10.1021/jacs.7b07498>.

- [14] T. Fiala, E.P. Barros, R. Heeb, S. Riniker, H. Wennemers, Predicting Collagen Triple Helix Stability through Additive Effects of Terminal Residues and Caps, *Angew. Chemie - Int. Ed.* (2022). <https://doi.org/10.1002/anie.202214728>.
- [15] J.L. Zitnay, Y. Li, Z. Qin, B.H. San, B. Depalle, S.P. Reese, M.J. Buchler, S.M. Yu, J.A. Weiss, Molecular level detection and localization of mechanical damage in collagen enabled by collagen hybridizing peptides, *Nat. Commun.* 8 (2017). <https://doi.org/10.1038/ncomms14913>.
- [16] J. Hwang, B.H. San, N.J. Turner, L.J. White, D.M. Faulk, S.F. Badylak, Y. Li, S.M. Yu, Molecular assessment of collagen denaturation in decellularized tissues using a collagen hybridizing peptide, *Acta Biomater.* 53 (2017) 268–278. <https://doi.org/10.1016/j.actbio.2017.01.079>.
- [17] J.A. Fallas, L.E.R.O. Leary, J.D. Hartgerink, Synthetic collagen mimics: Self-assembly of homotrimers, heterotrimers and higher order structures, *Chem. Soc. Rev.* 39 (2010) 3510–3527. <https://doi.org/10.1039/b919455j>.
- [18] S.M. Yu, Y. Li, D. Kim, Collagen mimetic peptides: Progress towards functional applications, *Soft Matter.* 7 (2011) 7927–7938. <https://doi.org/10.1039/c1sm05329a>.
- [19] and D.K. S.Michael Yu, Yang Li, Collagen Mimetic Peptides: Progress Towards Functional Applications, *HSS Public Access.* 7 (2011) 7927–7938. <https://doi.org/10.1039/C1SM05329A.Collagen>.
- [20] M.A. Cejas, W.A. Kinney, C. Chen, G.C. Leo, B.A. Tounge, J.G. Vinter, P.P. Joshi, B.E. Maryanoff, V. Ad, V. Uni, C. Gables, Collagen-Related Peptides : Self-Assembly of Short

- , Single Strands into a Functional Biomaterial of Micrometer Scale, *JACS Commun.* (2007) 2202–2203.
- [21] B.B. Anton V. Persikov, John A.M. Ramshaw, Collagen model peptides Sequence dependence of triple-helix stability, *Biopolym. - Pept. Sci. Sect.* 55 (2000) 436–450.
- [22] L. He, P. Theato, Collagen and collagen mimetic peptide conjugates in polymer science, *Eur. Polym. J.* 49 (2013) 2986–2997. <https://doi.org/10.1016/j.eurpolymj.2013.05.033>.
- [23] W. Wang, C. Yu, J. Chang, T. Chen, Y. Chen, Y. Sun, S. Wang, S. Jao, Biomolecular Chemistry stability but attenuates the specific assembly of, *Org. Biomol. Chem.* 19 (2021) 1860–1866. <https://doi.org/10.1039/d0ob02190c>.
- [24] S.A.H. Hulgán, J.D. Hartgerink, Recent Advances in Collagen Mimetic Peptide Structure and Design, *Biomacromolecules.* 23 (2022) 1475–1489. <https://doi.org/10.1021/acs.biomac.2c00028>.
- [25] H.P. Ba, K. Mizuno, S.P. Boudko, Kinetic Hysteresis in Collagen Folding, *Biophys. J.* 98 (2010) 3004–3014. <https://doi.org/10.1016/j.bpj.2010.03.019>.
- [26] S. Frank, S. Boudko, K. Mizuno, T. Schulthess, H.P. Ba, Collagen Triple Helix Formation Can Be Nucleated at Either End *, *J. Biol. Chem.* 278 (2003) 7747–7750. <https://doi.org/10.1074/jbc.C200698200>.
- [27] W. Roth, E.R. Heidemann, Triple helix–coil transition of covalently bridged collagenlike peptides, *Biopolymers.* 19 (1980).
- [28] W. Henkel, T. Vogl, H. Echner, W. Voelter, C. Urbanke, D. Schleuder, J. Rauterberg, Synthesis and folding of native collagen III model peptides, *Biochemistry.* 38 (1999)

- 13610–13622. <https://doi.org/10.1021/bi9905157>.
- [29] R.S. Erdmann, H. Wennemers, Synthesis of fmoc-pro-hyp(TBDPS)-gly-OH and its application as a versatile building block for the preparation of collagen model peptides, *Synthesis (Stuttg)*. (2009) 143–147. <https://doi.org/10.1055/s-0028-1083281>.
- [30] I.C. Tanrikulu, W.M. Westler, A.J. Ellison, J.L. Markley, R.T. Raines, Templated Collagen “double Helices” Maintain Their Structure, *J. Am. Chem. Soc.* 142 (2020) 1137–1141. <https://doi.org/10.1021/jacs.9b07583>.
- [31] J.M. Dones, I.C. Tanrikulu, J. V. Chacko, A.B. Schroeder, T.T. Hoang, A.L.F. Gibson, K.W. Eliceiri, R.T. Raines, Optimization of interstrand interactions enables burn detection with a collagen-mimetic peptide, *Org. Biomol. Chem.* 17 (2019) 9906–9912. <https://doi.org/10.1039/c9ob01839e>.
- [32] R. Berisio, V. Granata, L. Vitagliano, A. Zagari, Imino acids and collagen triple helix stability: Characterization of collagen-like polypeptides containing Hyp-Hyp-Gly sequence repeats, *J. Am. Chem. Soc.* 126 (2004) 11402–11403. <https://doi.org/10.1021/ja047069h>.
- [33] J. Egli, C. Esposito, M. Muri, S. Riniker, H. Wennemers, Influence of Lipidation on the Folding and Stability of Collagen Triple Helices - An Experimental and Theoretical Study, *J. Am. Chem. Soc.* 143 (2021) 5937–5942. <https://doi.org/10.1021/jacs.1c01512>.
- [34] J.A.M. Ramshaw, N.K. Shah, B. Brodsky, Gly-X-Y Tripeptide Frequencies in Collagen: A Context for Host-Guest Triple-Helical Peptides, 1998.
- [35] C. Byrne, P.A. McEwan, J. Emsley, P.M. Fischer, W.C. Cahn, End-stapled homo and hetero collagen triple helices: A click chemistry approach, *Chem. Commun.* 47 (2011) 2589–2591.

<https://doi.org/10.1039/c0cc04795c>.

- [36] A. V. Persikov, J.A.M. Ramshaw, B. Brodsky, Prediction of collagen stability from amino acid sequence, *J. Biol. Chem.* 280 (2005) 19343–19349. <https://doi.org/10.1074/jbc.M501657200>.
- [37] H.S. Chiu, J.C. Horng, Modulating the Stability of Collagen Triple Helices by Terminal Charged Residues, *J. Phys. Chem. B.* 125 (2021) 7351–7358. <https://doi.org/10.1021/acs.jpcc.1c01631>.
- [38] T. Fiala, E.P. Barros, M.O. Ebert, E. Ruijsenaars, S. Riniker, H. Wennemers, Frame Shifts Affect the Stability of Collagen Triple Helices, *J. Am. Chem. Soc.* 144 (2022) 18642–18649. <https://doi.org/10.1021/jacs.2c08727>.
- [39] N. Rezaei, A. Lyons, N.R. Forde, Environmentally Controlled Curvature of Single Collagen Proteins, *Biophys. J.* 115 (2018) 1457–1469. <https://doi.org/10.1016/j.bpj.2018.09.003>.
- [40] Y. Qi, D. Zhou, J.L. Kessler, R. Qiu, S.M. Yu, G. Li, Z. Qin, Y. Li, Terminal repeats impact collagen triple-helix stability through hydrogen bonding, *Chem. Sci.* 13 (2022) 12567–12576. <https://doi.org/10.1039/d2sc03666e>.
- [41] P.P. Provenzano, D. Heisey, K. Hayashi, R. Lakesand, R. Vanderby, Subfailure damage in ligament: A structural and cellular evaluation, *J. Appl. Physiol.* 92 (2002) 362–371. <https://doi.org/10.1152/jappl.2002.92.1.362>.
- [42] S.P. Veres, J.M. Lee, Designed to fail: A novel mode of collagen fibril disruption and its relevance to tissue toughness, *Biophys. J.* 102 (2012) 2876–2884. <https://doi.org/10.1016/j.bpj.2012.05.022>.

- [43] K.P. Quinn, J.A. Bauman, N.D. Crosby, B.A. Winkelstein, Anomalous fiber realignment during tensile loading of the rat facet capsular ligament identifies mechanically induced damage and physiological dysfunction, *J. Biomech.* 43 (2010) 1870–1875. <https://doi.org/10.1016/j.jbiomech.2010.03.032>.
- [44] S.P. Veres, J.M. Harrison, J.M. Lee, Repeated subrupture overload causes progression of nanoscaled discrete plasticity damage in tendon collagen fibrils, *J. Orthop. Res.* 31 (2013) 731–737. <https://doi.org/10.1002/jor.22292>.
- [45] B. Depalle, Z. Qin, S.J. Shefelbine, M.J. Buehler, Influence of cross-link structure, density and mechanical properties in the mesoscale deformation mechanisms of collagen fibrils, *J. Mech. Behav. Biomed. Mater.* 52 (2015) 1–13. <https://doi.org/10.1016/j.jmbbm.2014.07.008>.
- [46] M.H. Li, P. Fan, B. Brodsky, J. Baum, Two-Dimensional NMR Assignments and Conformation of (Pro-Hyp-Gly)₁₀ and a Designed Collagen Triple-Helical Peptide, *Biochemistry.* 32 (1993) 7377–7387. <https://doi.org/10.1021/bi00080a007>.
- [47] S. Shantanam, MUELLER, Multicomponent Peptide Assemblies, *Physiol. Behav.* 176 (2018) 139–148. <https://doi.org/10.1039/c8cs00115d.Multicomponent>.
- [48] J. Bella, M. Eaton, B. Brodsky, H.M. Berman, Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution., *Science.* 266 (1994) 75–81. <https://doi.org/10.1126/science.7695699>.
- [49] N.K. Shah, J.A.M. Ramshaw, A. Kirkpatrick, C. Shah, B. Brodsky, A host-guest set of triple-helical peptides: Stability of Gly-X-Y triplets containing common nonpolar residues,

- Biochemistry. 35 (1996) 10262–10268. <https://doi.org/10.1021/bi960046y>.
- [50] Y. Xu, M. Kirchner, Collagen mimetic peptides, *Bioengineering*. 8 (2021) 1–24. <https://doi.org/10.3390/bioengineering8010005>.
- [51] S. Sakakibara, K. Inouye, K. Shudo, Y. Kishida, Y. Kobayashi, D.J. Prockop, Synthesis of (Pro-Hyp-Gly)_n of defined molecular weights Evidence for the stabilization of collagen triple helix by hydroxypyroline, *BBA - Protein Struct.* 303 (1973) 198–202. [https://doi.org/10.1016/0005-2795\(73\)90164-5](https://doi.org/10.1016/0005-2795(73)90164-5).
- [52] N. Cohen, C.D. Eisenbach, Molecular Mechanics of Beta-Sheets, *ACS Biomater. Sci. Eng.* 6 (2020) 1940–1949. <https://doi.org/10.1021/acsbiomaterials.9b01983>.
- [53] Z. Qin, M.J. Buehler, Cooperative deformation of hydrogen bonds in beta-strands and beta-sheet nanocrystals, *Phys. Rev.* 82 (2010) 1–9. <https://doi.org/10.1103/PhysRevE.82.061906>.
- [54] A. Nova, S. Keten, N.M. Pugno, A. Redaelli, M.J. Buehler, Molecular and nanostructural mechanisms of deformation, strength and toughness of spider silk fibrils, *Nano Lett.* 10 (2010) 2626–2634. <https://doi.org/10.1021/nl101341w>.
- [55] Z. Qin, M.J. Buehler, Cooperativity governs the size and structure of biological interfaces, *J. Biomech.* 45 (2012) 2778–2783. <https://doi.org/10.1016/j.jbiomech.2012.08.043>.
- [56] L.L. Bennink, Y. Li, B. Kim, I.J. Shin, B.H. San, M. Zangari, D. Yoon, S.M. Yu, Visualizing collagen proteolysis by peptide hybridization: From 3D cell culture to in vivo imaging, *Biomaterials*. 183 (2018) 67–76. <https://doi.org/10.1016/j.biomaterials.2018.08.039>.
- [57] J. Hwang, Y. Huang, T.J. Burwell, N.C. Peterson, J. Connor, S.J. Weiss, S.M. Yu, Y. Li, In

- Situ Imaging of Tissue Remodeling with Collagen Hybridizing Peptides, *ACS Nano*. 11 (2017) 9825–9835. <https://doi.org/10.1021/acsnano.7b03150>.
- [58] L.E.R. O’Leary, J.A. Fallas, E.L. Bakota, M.K. Kang, J.D. Hartgerink, Multi-hierarchical self-assembly of a collagen mimetic peptide from triple helix to nanofibre and hydrogel, *Nat. Chem.* 3 (2011) 821–828. <https://doi.org/10.1038/nchem.1123>.
- [59] I.C. Tanrikulu, A. Forticaux, S. Jin, R.T. Raines, Peptide tessellation yields micrometre-scale collagen triple helices, *Nat. Chem.* 8 (2016) 1008–1014. <https://doi.org/10.1038/nchem.2556>.