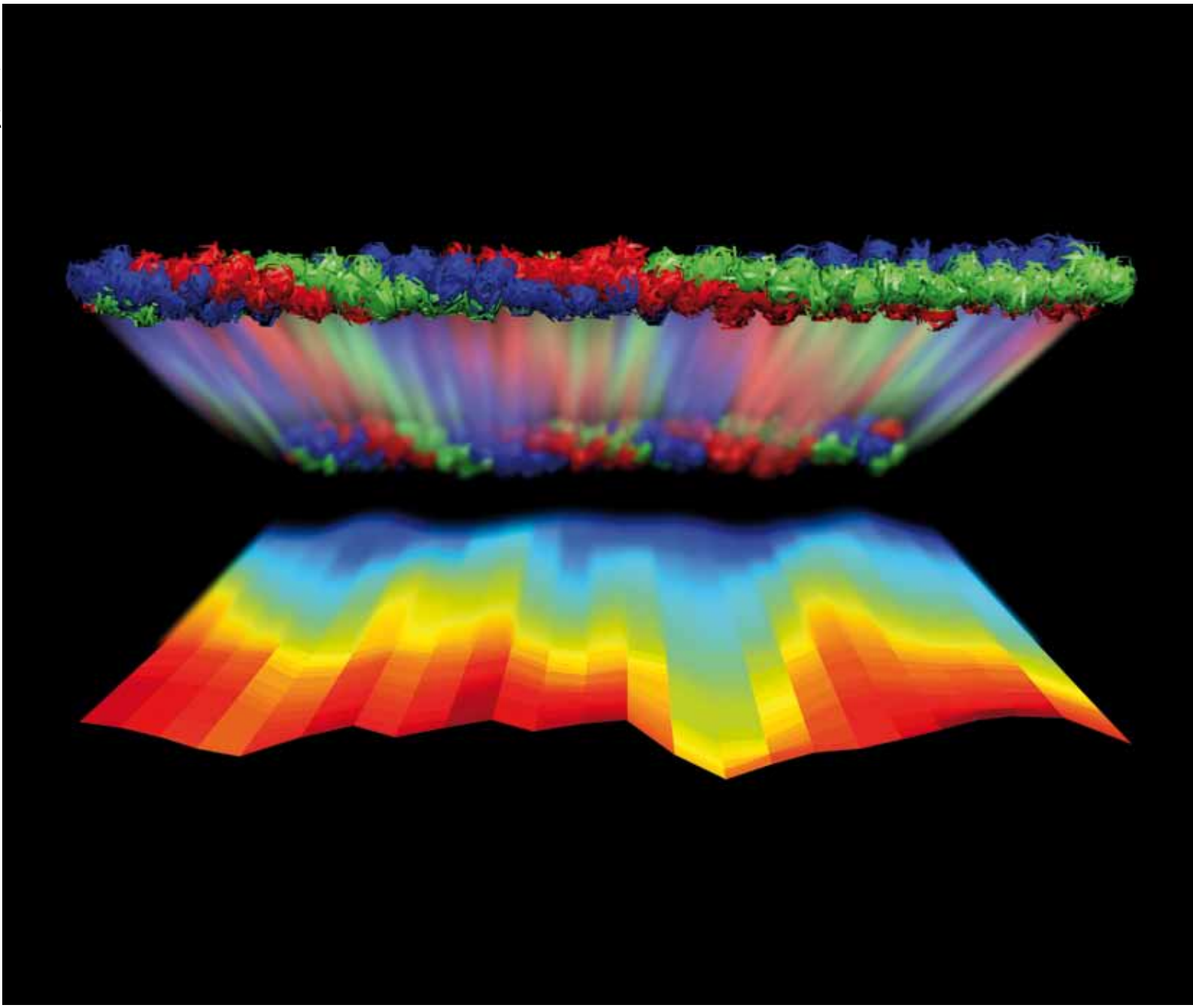
Integrative Biology

Quantitative biosciences from nano to macro

www.rsc.org/ibiology

Volume 1 | Number 7 | July 2009 | Pages 437–488



ISSN 1757-9694

RSC Publishing

1757-9694(2009)1:7;1-N

Nanomechanical sequencing of collagen: tropocollagen features heterogeneous elastic properties at the nanoscale

Sebastien G. M. Uzel[ab] and Markus J. Buehler*[ac]

Received 6th April 2009, Accepted 26th May 2009
First published as an Advance Article on the web 9th June 2009
DOI: 10.1039/b906864c

Collagen is the most important structural protein in biology and is responsible for the strength and integrity of tissues such as bone, teeth, cartilage and tendon. Here we report a systematic computational sequencing of the effect of amino acid motif variations on the mechanical properties of single tropocollagen molecules, with a particular focus on elastic deformation at varying applied strains. By utilizing a bottom-up computational materiomics approach applied to four model sequence motifs found in human type I collagen, we show that variations in the amino acid motif severely influence the elastic behavior of tropocollagen molecules, leading to softening or stiffening behavior. We also show that interpeptide interactions *via* H-bonds vary strongly with the type of motif, which implies that it plays a distinct role in the molecule's stability. The most important implication of our results is that deformation in tropocollagen molecules is highly inhomogeneous, since softer regions deform more than stiffer regions, potentially leading to strain and stress concentrations within collagen fibrils. We confirm the hypothesis of inhomogeneous molecular deformation through direct simulation of stretching of a segment of tropocollagen from human type I collagen that features the physiological amino acid sequence. Our results show that the biomechanical properties of tropocollagen must be understood in the context of the specific amino acid sequence as well as the state of deformation, since the elastic properties depend strongly on the amount of deformation applied to a molecule.

1. Introduction

Collagen, the most abundant protein in vertebrates, is a connective tissue protein that is responsible for the mechanical strength and integrity of many tissues such as bone, teeth, cartilage and tendons.[1–10] Collagenous tissues typically consist of triple helical tropocollagen molecules that have highly conserved lengths of $L \approx 300$ nm with a molecular diameter of roughly 1.5 nm. In fibrous collagens such as type I collagen, self-assembly of fibrils and fibers occurs by entropy-driven alignment of molecules along each other in a well organized, periodic fashion, stabilized by intermolecular adhesion and cross-links between lysine and hydroxylysine. These staggered arrays of tropocollagen molecules form collagen fibrils, that arrange to form collagen fibers and other higher order assembly patterns (Fig. 1(a)). The biological role of collagen is directly related to its mechanical properties, which is evident since it is found in key structural protein materials such as bone, tendon, cartilage and fascia. For example, in bone collagen plays the role of a biological glue that effectively links mineral platelets to form a strong and tough nanocomposite. In tendon, the ability of collagen to store energy

[a] *Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave. Room 1-235A&B, Cambridge, MA, USA. E-mail: mbuehler@mit.edu; Fax: +1-617-324-4014; Tel: +1-617-452-2750*
[b] *Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA, USA*
[c] *Center for Computational Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA, USA*

Insight, innovation, integration

The integration of an engineering analysis based on molecular dynamics simulation used here to identify Young's modulus and other mechanical properties of the fundamental constituents of collagenous tissue brings about a synergistic integration of technology and biology, with great potential for biomedical advances to better understand genetic diseases associated with a breakdown of collagenous tissues. Specifically, an innovative utilization of the bottom-up approach *via* molecular dynamics as put forth in our paper shows for the first time that, due to variations in the amino acid sequence, tropocollagen molecules feature highly heterogeneous mechanical properties at the nanoscale, leading to an inhomogeneous distribution of deformation within molecules. Our study provides a quantitative link between genetic information and the mechanical properties of the most abundant constituent of connective tissues in biology, enabled by a computational biology approach.

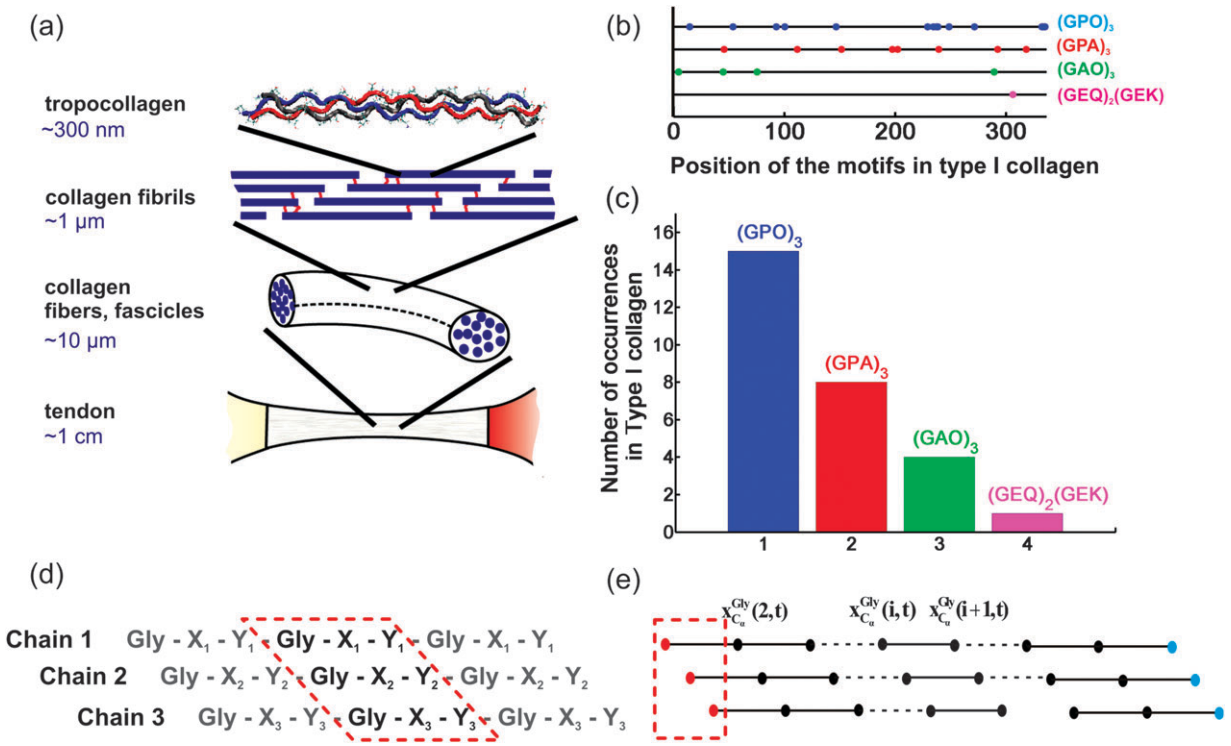


Fig. 1 Collagenous tissue structure and models description. Connectives tissues such as bone or tendon are composed of type I collagen, arranged in several levels of hierarchy. Panel (a) represents four of those hierarchical levels as observed in tendon. Panel (b) shows the locations of the four model collagen segments (each 30 residues long) that are investigated in this paper. Panel (c) displays the number of occurrences of the model collagen segments in human type I collagen. The motifs studied here are $(GPO)_3$, $(GAO)_3$, $(GPA)_3$ and $(GEQ)_2(GEK)$, where G, P, A, O, E, Q and K stand for glysine, proline, alanine, hydroxyproline, glutamate, glutamine and lysine, respectively. Panel (d) shows how a "motif" is defined. Panel (e) illustrates how the boundary conditions are applied and which C_α carbons are designated in the first strand in order to compute the local strain using eqn (5).

at large deformation without breaking is important for locomotion, where collagenous tissue acts as a spring between muscle and bone. Similar strengthening effects due to collagen are found in blood vessels and in the skin. The collagen content in fascia, a soft tissue constituent of the connective tissue system that permeates the human body, is vital to its ability provide structural integrity and to act as a shock absorber. A highly organized, crystalline form of collagen is found in the cornea and in the lens of the eye, where it provides both structural stability as well as optical properties. Several medical disorders have been linked to a breakdown of collagen, including genetic diseases such as *osteogenesis imperfecta* (brittle bone disease), Ehlers-Danlos syndrome (extreme stretchability of skin and fragile blood vessels), as well as Alport syndrome (kidney disease). Specifically in genetic disorders, structural changes to the biochemical make-up of individual collagen molecules are responsible for higher-length scale catastrophic tissue failure. Furthermore, the response of collagen to mechanical shock and large deformation is crucial for the understanding of various tissue injuries.

In this article we focus on the mechanical properties of single tropocollagen molecules, the basic constituent of all collagenous tissues. Since the typical physiological loading condition for collagenous tissues is tensile loading, we emphasize our analysis on the mechanical response of tropocollagen molecules under axial stretch. Earlier studies on tropocollagen molecules and collagen fibrils using optical tweezers[11–14] and AFM/MEMS devices,[15–18] as well as bottom-up molecular dynamics simulations[19–23] have not explicitly considered the effects of the amino acid sequence on the resulting material properties. Larger-scale models of collagen molecules and fibrils based on a series of springs[10] have not yet incorporated molecular-level sequence information into the constitutive description. In particular, most theoretical and simulation studies have been focused on model sequences of tropocollagen sequences, based on synthetic collagen-like peptides.[24] Even though all tropocollagen molecules show a particular sequence pattern universal to all collagenous proteins (*e.g.* glycine is found at approximately every third residue), there exists a significant variation of sequence patterns throughout the molecule. The variation of the sequence patterns may result in significant changes in the mechanical behavior. This is because the mechanisms of energy storage and release and the mechanisms by which side chains move with respect to each other under deformation could be strongly influenced by key biochemical parameters such as the side chain volumes or alterations in charged groups. For example, the volumes of the residue side chains range from 60 $Å^3$ for glycine to 227.8 $Å^3$ for tryptophan, which may control the equilibrium structure of a relaxed molecule without any load being applied. Similarly, the charge content of side chains varies significantly and could

severely influence the molecular properties. However, the effect of such sequence variations on the mechanical properties remains poorly understood thus far.

Here, we resolve this issue by explicitly considering the effect of amino acid sequence on the resulting mechanical properties of tropocollagen. To enable a systematic, quantitative analysis we define four model sequences. The sequences studied here are composed of 10 repeats of the motifs $(GPO)_3$, $(GAO)_3$, $(GPA)_3$ and $(GEQ)_3(GEK)$, where G, P, A, O, E, Q and K stand for glycine, proline, alanine, hydroxyproline, glutamate, glutamine and lysine, respectively. The analysis reported here therefore includes four distinct molecules comprised of 30 residues each (since there are three chains in a tropocollagen molecule, the total number of amino acid residues considered is $3 \times 30 = 90$). Fig. 1(b) shows the locations of the four model collagen segments that are investigated in this paper. Fig. 1(c) displays the frequency of occurrence of the four collagen motifs in physiological human type I collagen molecules. The $(GPO)_3$ motif is the most frequent one, followed by the $(GAO)_3$ and $(GPA)_3$ motifs, respectively. The $(GEQ)_2(GEK)$ motif is the one that appears least frequently. After creating atomistic models of these model collagen segments (including relaxation and equilibration), we measure the force–extension history and carry out a comparative analysis between the four models, focusing on the variation of stiffness as the deformation of the molecule is increased (the derivative of the force–extension history). Following the analysis of the four model collagen segments we also investigate the stretching behavior of a model that features the physiological sequence of human type I collagen. The key steps in the experimental protocol are summarized in Fig. 2, where the tensile loading condition is depicted in Fig. 2(c).

2. Results and discussion

We begin our analysis by carrying out tensile deformation tests of the four tropocollagen model segments. The stress-strain curves corresponding to the four collagen model sequences are

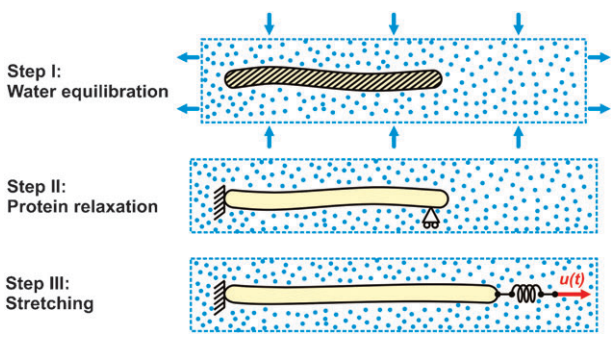


Fig. 2 Illustration of simulation protocol used to equilibrate, relax and stretch tropocollagen molecules. The molecular dynamics simulation protocol consists of three steps. In step I, we equilibrate the water box through adaptation to the controlled pressure. In step II, we perform protein relaxation, where one end is fixed and the other end is kept aligned in the direction orthogonal to the molecular axis. Step III consists of carrying out steered molecular dynamics (SMD) stretching under constant velocity pulling. The dots in the graphs correspond to water molecules (solvent).

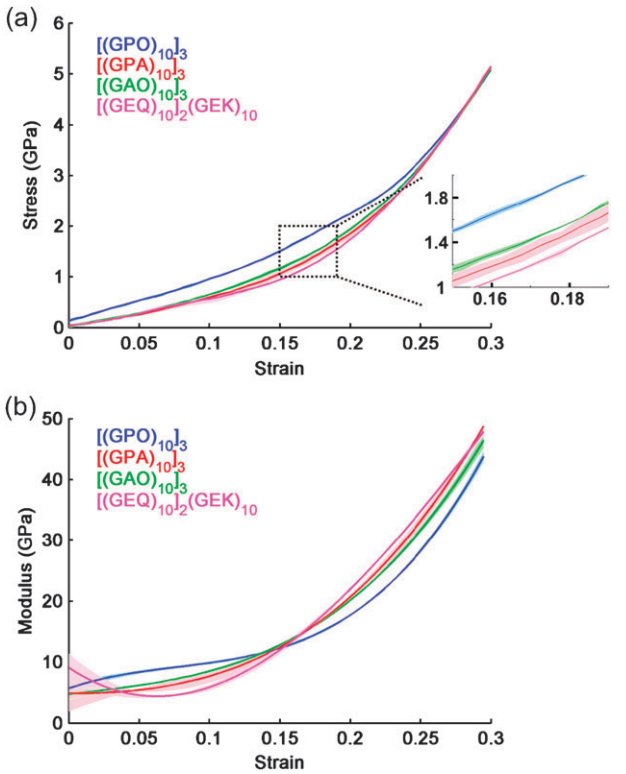


Fig. 3 Nanomechanical response for varying sequences. Panel (a) shows the stress strain behavior for the four model collagen segments studied here. For strains less than 25%, each motif exhibits a distinct behavior (where the results are highly repeatable). For strains larger than 25% all four curves approach each other, corresponding to the beginning of stretching of the protein backbone (which is expected to be independent of the motif of amino acids). Panel (b) shows the Young's modulus as a function of strain, for all four model collagen segments. The results show a significant variation in Young's modulus for the four sequences, as well as a strong dependence of it on the applied strain.

shown in Fig. 3(a), and the associated Young's moduli are depicted in Fig. 3(b). The analysis shows that the Young's moduli range between 4 and 10 GPa for small values of strain. These moduli are consistent with experimental studies[11–14] and earlier molecular dynamics simulation results.[19–25] The results further show that the shape of the Young's modulus dependence on strain and the quantitative numerical values are strongly affected by the amino acid motif. It can either be a strictly increasing function of the strain, such as in the case of $(GPO)_3$ motif, or a decreasing and subsequently increasing function, as it is the case for $(GEQ)_2(GEK)$. In addition, the values of Young's moduli change drastically as the stain increases. The shaded bands in the plots represent the standard deviation over the three runs per motif, providing support for statistical significance of the results put forth here.

We continue with a comparative analysis of the results shown in Fig. 3. Fig. 4(a) shows the relative ratio between the $(GPO)_3$ motif (here taken as a reference case) and the three other motifs, defined as:

$$\eta = \frac{E_X - E_{GPO}}{E_{GPO}}, \qquad (1)$$

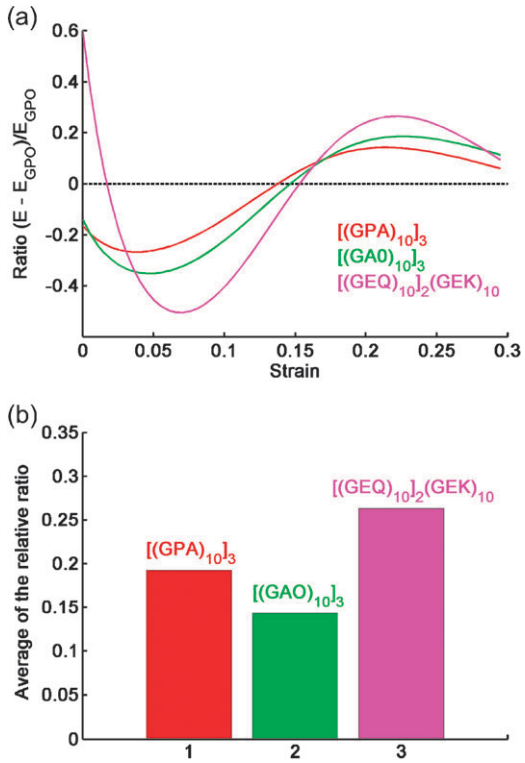


Fig. 4 Young's modulus (stiffness) comparison for varying sequences. Panel (a) shows the relative ratio between the Young's modulus of the reference case (GPO) and the three other cases. We observe that at a certain deformation, this relative ratio can exceed 25%, even reaching 50% at approximately 7% strain for the motif $(GEQ)_2(GEK)$. Panel (b) shows the average of the ratios of the data shown in panel (a).

where X is $(GPA)_3$, $(GAO)_3$ or $(GEQ)_2(GEK)$. For strain less than 15%, those three motifs are softer than the reference one, whereas for strain greater than 15%, they become stiffer. The difference in the Young's modulus can reach up to 50%. Past 25% deformation, the ratio η approaches zero, which is consistent with the fact that all Young's moduli approach the same value as the backbone is being stretched and thus the stiffness is no longer influenced by the type of motif (the structure of the backbone is identical in all amino acids). Interestingly, at $\varepsilon \approx 15\%$, the values for Young's modulus for all cases seems to approach the same value. A more general measure of the difference of the stiffness is given in Fig. 4(b); it corresponds to the average of the ratio defined above (over all strains considered).

Fig. 5 presents the evolution of H-bonds in each of the four sequences during stretching. We find that each sequence exhibits a different pattern. The amount of H-bonding starts decreasing at around 6% strain for $(GPO)_3$ and $(GAO)_3$, whereas it remains roughly constant for $(GPA)_3$ and $(GEQ)_2(GEK)$. At 15% strain, we observe a change in the slope of all the four curves, which may partly explain why all moduli approach the same value at this particular strain level (since the stiffness relates to the rate of H-bonds breaking). At large strain, the ratio between the number of H-bonds is at a maximum between motif $(GPO)_3$ and $(GEQ)_2(GEK)$. The number of H-bonds in the latter is

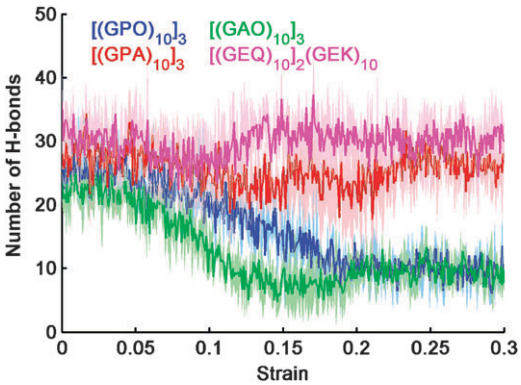


Fig. 5 Evolution of the number of H-bonds during stretching for varying sequences. This figure shows how the motif can influence the way H-bonds break. For motifs $(GPO)_3$ and $(GAO)_3$ the amount of H-bonding reduces by 70% over 30% strain, whereas for cases $(GPA)_3$ and $(GEQ)_2(GEK)$, the number of H-bonds remains roughly the same. This indicates that the stability of the protein is affected by its sequence motifs.

approximately five times greater than in the former. This result shows that the stability of the molecule, which is related to the number of H-bonds, is affected by the type of motif composing the protein.

Our results show a significant effect of the amino acid motif on the resulting mechanical properties. This result has important implications for the deformation of physiological tropocollagen molecules, which are composed of a diverse combination of sequence motifs along the molecular axis. Specifically, since the elastic stiffness varies over the sequence and thus along the molecular axis, we expect that mechanical deformation will be heterogeneous, where some regions of the molecules are stretched more (the softer regions), and others are stretched less (the stiffer regions). To test this hypothesis, we carry out a tensile test of a segment of the tropocollagen molecule that features the exact sequence of human type I collagen.

Fig. 6(a) represents the local deformation along the first α_1 chain of the triple helix, showing a measure of the local strain (a measure for the local deformation) over the length of the tropocollagen molecule, for a segment of a collagen molecule that features the exact sequence of human type I collagen. The graph reveals regions of relatively small deformation, that is, stiffer motifs (bluish, "cold" colors) and regions of large deformation, that is, softer motifs (reddish, "hot" colors) (Fig. 6(b)). The representation depicted in Fig. 6(c) shows the deformation averaged over the last 100 ps of the simulation, when the total strain is approximately 13%. This visualization facilitates a straightforward comparison of the local strain values with the average strain values, characterized by the dotted red line. The plot clearly shows that the distribution of strain is highly inhomogeneous, suggesting a significant heterogeneity of the stiffness along the molecular axis. The largest and smallest strains represent 32% less and 34% more compared with the average strain, respectively. The large local strains pose the question about whether or not covalent bond breaking may be induced. In the results presented here, at an overall strain of 13%, the most

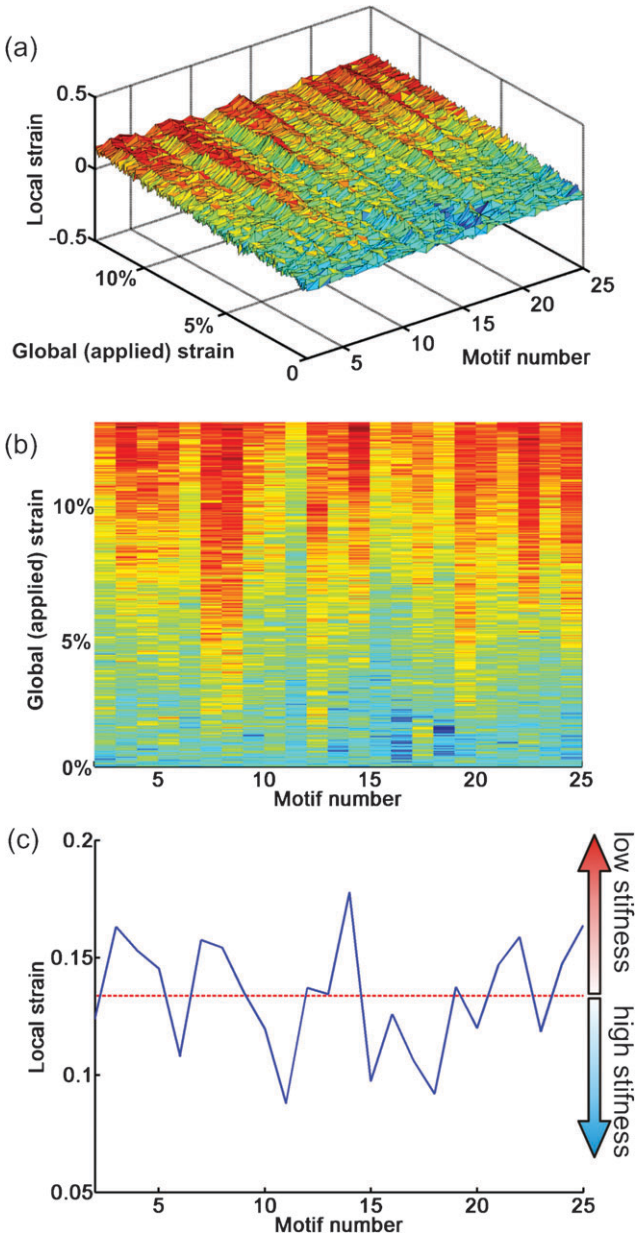


Fig. 6 Local deformation of a segment of human type I collagen molecule. Panel (a) represents the local deformation along the first α_1 chain of the triple helix. The strain is measured by computing the relative displacement of C_α carbons in each of the glycine amino acids (one every three residues). Panel (b) is a top view of the plot in panel (a). Panel (c) exhibits regions of low deformation (stiffer) and regions of high deformation (softer) relative to the average deformation.

stretched section in the molecule reaches approximately 17% strain. Earlier molecular dynamics studies of stretching a collagen-like peptide using the first principles based reactive force field ReaxFF[22] (with a full description of the chemistry of bond breaking and formation) suggest that bond breaking does not occur until approximately 50% molecular strain is reached. These results indicate that at the strain levels considered here no covalent bond rupture is expected to occur.

3. Discussion and conclusion

The key insight reported in this paper is that the type of amino acid motif that defines the tropocollagen molecule has significant effects on its mechanical properties, a phenomenon that has not yet been quantitatively investigated in the literature. Our study thereby makes a link between biochemical parameters (amino acid sequence) and associated biologically relevant functional properties (elasticity, stiffness, energy storage capacity). The findings reported here have important implications in understanding the elasticity and strength properties of a broad range of connective tissues, as they elucidate the fundamental heterogeneous nanomechanical response of collagen's basic constituent. Through molecular dynamics simulation, applied to provide a quantitative assessment of the mechanical properties of single tropocollagen molecules, we showed that changes in the composition lead to large changes in the stiffness, with a relative change of the elastic stiffness reaching 50%. Our results further corroborate the notion that collagen molecules constitute a highly nonlinear material, and that the mechanical properties should not be compared at a single deformed configuration but must be understood in the context of continuously changing strain, as they evolve during the elongation of the molecule. Finally, consistent with our finding that the motif controls the deformation of tropocollagen molecules, the stiffness of a type I tropocollagen molecule model with the physiological sequence varies continuously along the molecule, displaying a distribution of stiff and soft regions. The change of large and small deformation regions within the collagen molecule could potentially lead to strain and stress concentrations within collagen fibrils (where many collagen molecules are assembled in a staggered fashion into larger structures as shown in Fig. 1(a)). Specifically, relatively soft regions within the collagen fibril might behave similarly to small cracks and induce intermolecular sliding and unfolding at defined locations and in a defined sequence. In light of this hypothesis, the characteristic distribution of stiff and soft regions within the molecular packing might perhaps be related to the biomechanical function at varying levels of strain.

More generally, the precise control of unfolding of specific domains at defined levels of strain could provide a means to link biochemical sensing (*e.g. via* protein docking) to mechanical deformation, which could be crucial in understanding mechanotransduction processes. For example, the regions in Fig. 6 that show large levels of strain, or perhaps larger patterns of distributions of strain, may resemble highly specific binding sites for proteins associated with biochemical signaling cascades that are activated when collagen molecules are under severe mechanical stress (*e.g.* to initiate tissue remodeling). This concept of mechanically induced biochemical lock-and-key coupling in the context of signaling cascades might not only apply to collagen, but also to many other extra- and intracellular filamentous proteins found in biology that have been associated with mechanotransduction mechanisms (*e.g.* intermediate filaments).

Our results illustrate the importance of a computational biology approach through molecular dynamics simulations for two main reasons. Firstly, such a method offers great flexibility

in the molecular definition as opposed to experimental approaches that have to cope with existing molecular compositions or that would have to create these molecules through genetic engineering approaches. Secondly, our bottom-up approach provides a rapid, inexpensive computational microscope suitable for the rapid screening of a great variety of motifs. This approach enables us to quantify the nanomechanical response specific to particular amino acid motifs, at resolutions in space and time unmatched to any experimental protocol. These methods are fundamental to facilitate breakthroughs in understanding collagenous tissue behavior through a bottom-up approach that links nano to macro. Future studies could focus on linking the mechanical heterogeneities to the chemical and physical properties of the amino acids that make up the particular sequence motifs. For example, computational inference methods could be used to solve the inverse problem of identifying sequence patterns and associated biochemical properties related to stiff and soft regions, respectively.

The computational biology approach utilized here provides theoretical predictions of the behavior of a tropocollagen molecule at the interface of structural mechanics, structural biology and biochemistry. Specifically, the results put forth in this study show that biochemistry and biomechanics cannot be considered independently. In conventional biomechanics, collagenous tissues such as cartilage or bone are typically treated as bulk materials, characterized by parameters such as Young's modulus, Poisson's ratio and some measure of anisotropy (that is, orientation dependence of properties). Our study shows that there exist significant local structural and property variations at many relevant tissue scales, from nano to macro (amino acid, molecular, fibrillar, *etc*.) that can not be captured by conventional nonlocal continuum theories.

Our finding that specific changes in the tropocollagen sequence motif modulate the mechanical properties at the molecular level is a computationally driven hypothesis, which, if confirmed by experimental studies, can have profound implications on our understanding of genetic components of many connective tissue diseases. In order to validate this hypothesis experimentally, further studies should be carried out. Even though currently there are no experimental results available that consider the deformation of tropocollagen molecules along the molecular axis under applied deformation, in a recent experimental study[26] it was found that (under no stretch applied) the axial pitch per residue is not uniform along the molecular axis. Specifically, the distances between amino acids were found to range between 2.70 Å and 3.04 Å, somewhat close to our predictions ranging between 2.82 Å and 3.28 Å (referring specifically to the first segment investigated here). Future experimental studies might be carried out in a similar setting but with external strain applied to the molecule, thereby enabling a direct comparison with the theoretical predictions reported here. Our results may also be combined with spring models of collagen,[10] where simulation results as reported here could provide quantitative input from a bottom-up perspective towards a multi-scale model of collagenous tissues.

The significance of the findings reported here reaches beyond the specific application to collagen. Similar sequence dependent mechanical properties may be found for many other structural proteins, such as elastin or keratin, or intermediate filaments such as actin or microtubules. A quantitative understanding of fundamental relationships between mechanical properties and sequence patterns could also be critical to enable the design of novel peptide based or genetically engineered biomaterials. Currently, many existing biomaterials show deficiencies with respect to mechanical properties, a limitation that might be overcome by including a better understanding of the relevance of specific sequence patterns in the design process.

The integration of an engineering analysis (*e.g*. to identify Young's modulus and other mechanical properties) applied to a biological material (collagenous tissue) brings about a synergistic integration of technology and biology, with great potential for biomedical advances. For example, a quantitative understanding of the effect of motif on mechanical properties of the basic collagen building block and its relationship to the amino acid sequence may be crucial to advance our understanding of a broad range of genetic disorders associated with defects in collagen genes, including *osteogenesis imperfecta*, Ehlers-Danlos syndrome, Alport syndrome, and dwarfism (*e.g*. in spondyloepiphyseal dysplasia congenital, SED).[27–31] For these medical disorders, an association of the disease phenotype with changes of mechanical properties of connective tissue has been established, without knowing the underlying mechanisms. Our work helps to establish a quantitative link between the level of genetic information and resulting mechanical properties and how the overall tissue behavior changes. To further explore behavior of collagenous tissues at larger scales, additional analysis of intermolecular interaction (*e.g*. adhesion) and the associated mechanical properties of collagen fibrils and tissue should be carried out, perhaps using methods of multi-scale modeling and simulation.[21] Such work could be based on the results reported here and is left to future studies.

4. Materials and methods

4.1 Model structures and geometries

Four model collagen segments (model A) are investigated, as well as a portion of real human type I collagen (model B). For model A, the four model segments are composed of different motifs repeated 10 times, leading to molecules with a total of 90 residues. A motif is defined here as a unit block of tropocollagen, composed of three chains of three residues, that is, by a total of 9 amino acids. These motifs are $(GPO)_3$, $(GAO)_3$, $(GPA)_3$ and $(GEQ)_2(GEK)$, where G, P, A, O, E, Q and K stand for glycine, proline, alanine, hydroxyproline, glutamate, glutamine and lysine, respectively. The first three motifs are relatively frequent in type I collagen, whereas the last one is encountered only once and was chosen randomly among the 245 distinct motifs with similar frequency of occurrence (Fig. 2(b–c)). The segment of real type I collagen corresponds to the first 1/13th of the triple helical region, starting from the N-terminal. The full amino acid composition of this tropocollagen molecule is based on the gene sequences GenBank RefSeq NM_000088.3 and NM_000089.3, and can be found in www.le.ac.uk/ge/collagen/COL1A1_numbering.

pdf and www.le.ac.uk/ge/collagen/COL1A2_numbering.pdf. The sequence corresponding to the segment studied here is listed below. For both α_1 chains: GPMGPSGPRGLPGPP-GAPGPQGFQGPPGEPGEPGASGPMGPRGPPGPPGK-NGDDGEAGKPGRPGERGPPGPQGARGLP. For the α_2 chain: GPMGLMGPRGPPGAAGAPGPQGFQGPAGEP-GEPGQTGPAGARGPAGPPGKAGEDGHPGKPGRPGE-RGVVGPQGARGFP.

The input structures are created using the software THeBuScr,[32,33] which enables us to build a tropocollagen structure based on any specified amino acid sequence. Each segment comprises of 30 amino acids per chain for model A and 78 for model B, which is equivalent to an approximate length of 90 Å and 230 Å, respectively. This assures a reasonable computational time (the simulations take 12 days on several CPUs on a linux cluster; thus longer segments are currently computationally inhibited).

All proteins are solvated in a periodic water box using VMD. The dimensions of the box are 150 Å × 30 Å × 30 Å for model A and 350 Å × 30 Å × 30 Å for model B. The solvent is composed of TIP3P water molecules. Each system includes a total of approximately 16 500 atoms for model A, and 53 000 atoms for model B.

4.2 Computational method and parameterization

Molecular dynamics (MD) simulations are carried out using the NAMD code and the CHARMM force field[34] that includes parameters for hydroxyproline amino acids. The electrostatic interactions are modeled by the particle-mesh Ewald sums (PME) method. Energy minimization is performed using a conjugate gradient (CG) scheme. The equilibration is then carried out at a temperature of 310 K (= 37 °C). Finally, stretching of the proteins is performed *via* the steered molecular dynamics (SMD) method.[35] This method is based on the concept of pulling the center of mass of a collection of chosen atoms *via* a spring along the direction of the molecular axis, while keeping the center of mass of another group of atoms fixed through a stiffer spring. Here, the spring constants are $k_{\text{steered}} = 3 \times 10^5$ kJ mol^{-1} nm^{-2} and $k_{\text{fixed}} = 4 \times 10^3$ kJ mol^{-1} nm^{-2}.

Each simulation is set up in three steps as follows (see also Fig. 2(b)):

Energy minimization (10 ps) *and equilibration* (100 ps) *of the water molecules with fixed protein atoms*. This enables the water box to adapt its shape in order to reach the controlled pressure without disturbing the protein configuration.

Energy minimization (10 ps) *and equilibration* (1000 ps) *of the entire system*. During this period, the molecule is clamped on one side and restrained on the other side to *x*-displacements as depicted in Fig. 2(b). Thus, the molecule can freely expand to a relaxed state, remaining parallel to the pulling direction.

Stretching of the molecule for 5 ns. The C_α carbons at the N-terminal are fixed while the C_α carbons at the C-terminal are displaced at a constant velocity $v = 1$ m s^{-1}. Each of the four cases is repeated three times to facilitate statistical analysis.

All simulations are performed in an *NPT* ensemble (that is, constant number of particles, constant pressure, constant temperature), coupled to a thermostat at 310 K, at a pressure of 1 atm.

4.3 Analysis of nanomechanical properties

The stress σ inside the protein is defined as the ratio between the force exerted to the molecule F and its cross-section area A, where A is taken to be 214.34 Å^2 (following a suggestion put forth in ref. 22). The stress is defined as

$$\sigma = \frac{F}{A}. \tag{2}$$

The strain ε describes the relative elongation of the protein. If L is the length at a given time and L_0 the initial length, the strain ε (engineering strain) is given by

$$\varepsilon = \frac{L - L_0}{L_0}, \tag{3}$$

where L_0 was set to 90 Å, corresponding to the average of all initial molecular lengths of the 12 runs carried out.

Young's modulus E relates the stress and the strain and is generally dependent on the state of deformation (that is, strain ε) and gives a measure of the stiffness of the material. In the elastic approximation, it is defined as the derivative of the stress with respect to strain evaluated at the strain ε (generally, Young's modulus depends on the strain)

$$E(\varepsilon) = \frac{\mathrm{d}\sigma(\varepsilon)}{\mathrm{d}\varepsilon}. \tag{4}$$

The raw force and displacement data are stored every 100 integration steps, where the time step used in our simulation is 1 fs. The force–displacement data are converted into stress–strain curves using the relations and parameters given above (stress = force divided by molecular cross-sectional area, and strain is derived by measuring the elongation divided by the initial length). The results are then averaged in 0.5% increments of strain over intervals of 3% strain, from 0 to 30% strain, which results in 60 bins. Finally, a fourth order polynomial is fitted to the bin averaged results to obtain a smooth relation between the stress and strain that can be used to carry out the differentiation defined in eqn (4) (where the Young's modulus is calculated by computing the first derivative of $\sigma(\varepsilon)$). The results are depicted as the averaged stress–strain and modulus–strain curves over the three simulations carried out for each sequence. The standard deviations are represented by pastel bands. The upper (or lower) edge of each band is defined as the average value plus (or minus) the standard deviation over the three cases. They enable to measure the statistical relevance of the discrepancies between stresses and Young's moduli.

4.4 Physiological type I collagen segment

Local deformation was extracted from SMD simulations in order to point out the influence of the variety of amino acids in the mechanical properties. To do so, the relative displacement is computed in each chain, using a 5% strained state as the initial configuration in order to preclude perturbations due to thermal agitation. Measurements are performed on C_α carbons in each of the glycine amino acids (that is, for one in every three residues). The strain is measured by computing

the relative displacement of C_α carbons in each of the glycine amino acids (one in every three residues) as follows:

$$\varepsilon(i,t) = \frac{x_{C_\alpha}^{\mathrm{Gly}}(i+1,t) - x_{C_\alpha}^{\mathrm{Gly}}(i,t)}{x_{C_\alpha}^{\mathrm{Gly}}(i+1,0) - x_{C_\alpha}^{\mathrm{Gly}}(i,0)}. \qquad (5)$$

Note that in Fig. 5, the first value of the strain computed in the first chain was removed because the first C_α carbons at the N-terminal in the three chains are fixed and slightly staggered. Therefore, the second C_α carbon in chain 1 cannot expand as much as the other and the deformation at that position can not be considered for the evaluation. This is illustrated in Fig. 1(e).

Acknowledgements

Support for this research was provided by a National Science Foundation CAREER Award (grant number CMMI-0642545). SGMU acknowledges support through a Presidential Fellowship at MIT.

References

1 M. A. Meyers, *et al.*, Biological materials: Structure and mechanical properties, *Prog. Mater. Sci.*, 2008, **53**, 1.
2 P. Fratzl and R. Weinkamer, Nature's hierarchical materials, *Prog. Mater. Sci.*, 2007, **52**, 1263.
3 *Collagen Structure and Mechanics*, ed. P. Fratzl, Springer, 2008.
4 M. J. Buehler and Y. C. Yung, Deformation and failure of protein materials in physiologically extreme conditions and disease, *Nat. Mater.*, 2009, **8**(3), 175–188.
5 J. Currey, *Collagen and the Mechanical Properties of Bone and Calcified Cartilage*, in *Collagen: Structure and Mechanics*, ed. P. Fratzl, Springer, New York, 2008.
6 J. D. Currey, *Bones Structure and Mechanics*, Princeton University Press, Princeton, NJ, 2002.
7 J. K. Rainey, C. K. Wen and M. C. Goh, Hierarchical assembly and the onset of banding in fibrous long spacing collagennext term revealed by atomic force microscopy, *Matrix Biol.*, 2002, **21**(8), 647–660.
8 L. Bozec and M. Horton, Topography and mechanical properties of single molecules of type I collagen using atomic force microscopy, *Biophys. J.*, 2005, **88**(6), 4223–4231.
9 J. P. R. O. Orgel, *et al.*, Microfibrillar structure of type I collagen *in situ*, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**(24), 9001–9005.
10 F. H. Silver, *et al.*, *The role of collagen in energy storage and dissipation in extracellular matrix*, in *Bionanotechnology-Proteins to Nanodevices*, ed. R. V. L. V. Renugopalakrishnan, Springer, Netherlands, Dordrecht, The Netherlands, 2006.
11 S. Cusack and A. Miller, Determination of the Elastic-Constants of Collagen by Brillouin Light-Scattering, *J. Mol. Biol.*, 1979, **135**(1), 39–51.
12 N. Sasaki and S. Odajima, Stress-strain curve and Young's modulus of a collagen molecule as determined by the X-ray diffraction technique, *J. Biomech.*, 1996, **29**(5), 655–658.
13 Y. L. Sun, *et al.*, Direct quantification of the flexibility of type I collagen monomer, *Biochem. Biophys. Res. Commun.*, 2002, **295**(2), 382–386.
14 R. Harley, *et al.*, Phonons and Elastic-Moduli of Collagen and Muscle, *Nature*, 1977, **267**(5608), 285–287.
15 S. J. Eppell, *et al.*, Nano measurements with micro-devices: mechanical properties of hydrated collagen fibrils, *J. R. Soc. Interface*, 2006, **3**(6), 117–121.
16 J. A. J. van der Rijt, *et al.*, Micromechanical testing of individual collagen fibrils, *Macromol. Biosci.*, 2006, **6**(9), 697–702.
17 Z. L. Shen, *et al.*, Stress-strain experiments on individual collagen fibrils, *Biophys. J.*, 2008, **95**(8), 3956–3963.
18 Z. L. Shen, *et al.*, Stress-strain experiments on individual collagen fibrils, *Biophys. J.*, 2008, **95**(8), 3956–3963.
19 A. Gautieri, M. J. Buehler and A. Redaelli, Deformation rate controls elasticity and unfolding pathway of single tropocollagen molecules, *J. Mechan. Behav. Biomed. Mater.*, 2009, **2**, 130–137.
20 M. J. Buehler and S. Y. Wong, Entropic elasticity controls nanomechanics of single tropocollagen molecules, *Biophys. J.*, 2007, **93**(1), 37–43.
21 M. J. Buehler, Nature designs tough collagen: Explaining the nanostructure of collagen fibrils, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**(33), 12285–12290.
22 M. J. Buehler, Atomistic and continuum modeling of mechanical properties of collagen: Elasticity, fracture and self-assembly, *J. Mater. Res.*, 2006, **21**(8), 1947–1961.
23 C. M. Stultz, The folding mechanism of collagen-like model peptides explored through detailed molecular simulations, *Protein Sci.*, 2006, **15**(9), 2166.
24 R. Z. Kramer, *et al.*, Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair, *J. Mol. Biol.*, 2000, **301**(5), 1191–1205.
25 S. Vesentini, *et al.*, Molecular assessment of the elastic properties of collagen-like homotrimer sequences, *Biomech. Model. Mechanobiol.*, 2005, **3**(4), 224–234.
26 G. J. Cameron, D. E. Cairns and T. J. Wess, The variability in type I collagen helical pitch is reflected in the D periodic fibrillar structure, *J. Mol. Biol.*, 2007, **372**(4), 1097–1107.
27 D. J. Prockop and K. I. Kivirikko, Collagens: molecular biology, diseases, and potentials for therapy, *Annu. Rev. Biochem.*, 1995, **64**, 403–434.
28 B. G. Hudson, *et al.*, Alport's syndrome Goodpasture's syndrome and type IV collagen, *N. Engl. J. Med.*, 2003, **348**(25), 2543–2556.
29 B. P. Sokolov, *et al.*, Exclusion of COL1A1, COL1A2, and COL3A1 genes as candidate genes for Ehlers-Danlos syndrome type I in one large family, *Hum. Genet.*, 1991, **88**(2), 125–129.
30 A. Gautieri, *et al.*, Molecular and mesoscale mechanisms of osteogenesis imperfecta disease in collagen fibrils, *Biophys. J.*, 2009, **97**(3).
31 A. Gautieri, *et al.*, Single molecule effects of osteogenesis imperfecta mutations in tropocollagen protein domains, *Protein Sci.*, 2009, **18**(1), 161–168.
32 J. Rainey and M. Goh, A statistically derived parameterization for the collagen triple-helix, *Protein Sci.*, 2004, **13**(8), 2276–2276.
33 J. Rainey and M. Goh, An interactive triple-helical collagen builder, *Bioinformatics*, 2004, **20**(15), 2458–2459.
34 T. Lazaridis and M. Karplus, Effective energy function for proteins in solution, *Proteins: Struct., Funct., Bioinf.*, 1999, **35**(2), 133–152.
35 H. Lu, *et al.*, Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation, *Biophys. J.*, 1998, **75**(2), 662–671.