PAPER

Check for updates

Chemical bonds in collagen rupture selectively under tensile stress†

James Rowe [ID] and Konstantin Röder [ID] *

Collagen fibres are the main constituent of the extracellular matrix, and fulfil an important role in the structural stability of living multicellular organisms. An open question is how collagen absorbs pulling forces, and if the applied forces are strong enough to break bonds, what mechanisms underlie this process. As experimental studies on this topic are challenging, simulations are an important tool to further our understanding of these mechanisms. Here, we present pulling simulations of collagen triple helices, revealing the molecular mechanisms induced by tensile stress. At lower forces, pulling alters the configuration of proline residues leading to an effective absorption of applied stress. When forces are strong enough to introduce bond ruptures, these are located preferentially in X-position residues. Reduced backbone flexibility, for example through mutations or cross linking, weakens tensile resistance, leading to localised ruptures around these perturbations. In fibre-like segments, a significant overrepresentation of ruptures in proline residues compared to amino acid contents is observed. This study confirms the important role of proline in the structural stability of collagen, and adds detailed insight into the molecular mechanisms underlying this observation.

1 Introduction

Collagen is the main constituent of the extracellular matrix (ECM), a very important and functional biomaterial.[1–3] Collagen is formed by individual collagen fibrils, which in turn are formed by the entanglement of assemblies of three amino acid strands into a triple helix, so called tropocollagen. Additional interactions between molecules and fibrils through chemical crosslinks stabilise the collagen structure.[4] Tropocollagen is characterised by repeating triplets of amino acids, GXY. The glycine residues facilitate the formation of a characteristic triple helix formed by three strands.[5] The other two positions, X and Y, are exhibiting more variation, but are enriched in proline (P) and (2*S*,4*R*)-hydroxyproline (O), respectively. This enrichment has been linked to structural stability of collagen fibrils.[6,7] In Fig. 1, the

structural features of the hierachical assembly of collagen are illustrated.

Collagen provides structural stability to multicellular organisms by absorbing and resisting forces the extracellular matrix experiences. There are two relevant parts to this stability: Firstly, tropocollagen needs to be difficult to deform and break. Secondly, the assembly of the molecules into fibrils and their assembly into collagen needs to resist mechanical forces as well. It has been noted that the stability of tropocollagen in particular is determined by energetic contributions, *i.e.* by the strength of chemical bonds and interactions formed, rather than entropic.[9]

It is difficult to study these processes at high resolution through experimental techniques, and as a result modelling has been used to gain insight into these processes and determine how mechanical forces are absorbed within collagen. These efforts have led to a description of the force response of tropocollagen.[10] After an initial entropic phase (forces of a few pN), unfolding occurs, where the triple helix starts to unwind and lose its helicity. Once this process is completed at around 5000 pN, the backbone is stretched, until the tropocollagen ruptures. The unfolding process and the backbone stretch lead to a bilinear extension behaviour,[10] with the modelled behaviour closely matching experimental observations.[11] Coarse-grained mesoscale modelling allowed similar insight into the force response of cross-linked fibrils.[12] The properties of tropocollagen alone are not sufficient to explain the mechanical behaviour of collagen.[13] Nonetheless, the deformations of tropocollagen are key to the

Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK.
E-mail: kr366@cam.ac.uk

This journal is © the Owner Societies 2023

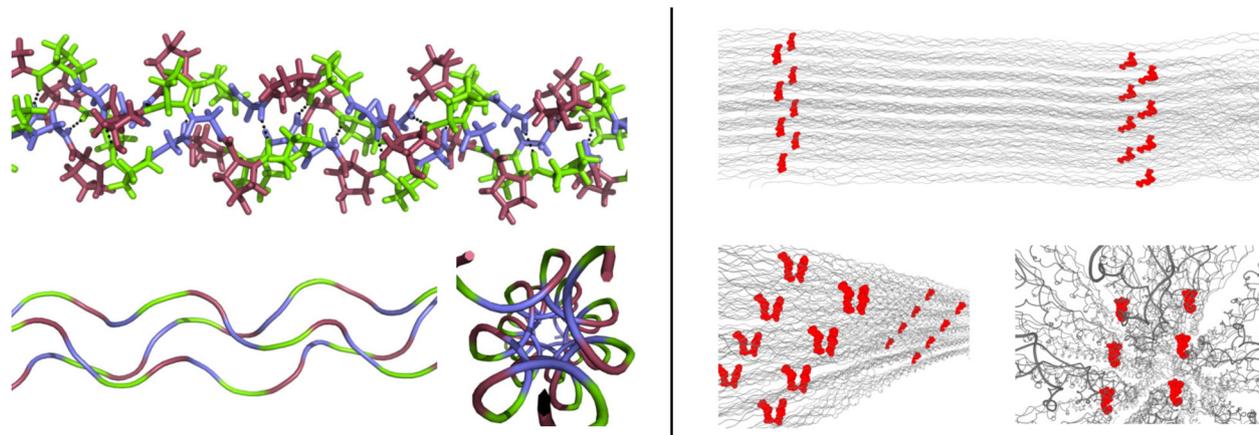*Phys. Chem. Chem. Phys.,* 2023, **25**, 2331–2341 | **2331**

Fig. 1 Model representations of collagen highlighting key structural features. Left: tropocollagen formed by GPO repeats. Glycines are in blue, X-proline in green, and Y-hydroxyproline in red. Hydrogen bonding is indicated with black dotted lines in the stick representation (top). The cartoon representations of the backbones only (bottom) reveal the helical nature of the three strands and the regular pattern they form. The small space inside tropocollagen can be seen. Right: fibre models created with Colbuilder[8] (settings: organism *Homo sapiens*; HLKNL N-terminal crosslinking for 9.C–947.A and 5.B–944.B; DPD C-crosslinking for 1047.C–1047.A–104.C; these choices are purely for illustrative purposes). Crosslinks are shown in red. The alignment of the tropocollagen to form collagen fibres can be seen.

understanding of these properties, as the molecular stretching and uncoiling are required for the hierarchical mechanics displayed by collagen.[13] A detailed insight into these processes can therefore lead towards novel approaches for identifying and treating disease and injury of tissues.[14]

Here, we are concerned with two specific parts of the force response of tropocollagen: (i) how tropocollagen responds to weaker forces, where structural changes are introduced, but no bond breaking occurs. This non-reactive regime corresponds to the molecular uncoiling and backbone stretching described previously.[10] (ii) When bond breaking occurs, where such ruptures are introduced, *i.e.* the reactive regime. The aim of this study is to characterise the molecular mechanisms in detail, adding to the mesoscopic descriptions provided by others.[10,12]

For the non-reactive regime, it has been proposed that changes to the configurations of proline residues are providing flexibility to the collagen fibrils.[15,16] The five-membered ring characteristic for proline can be in an *endo* or *exo* configuration.[5,17] At equilibrium, around 55 to 60% of proline residues at the X position are expected to be *endo*.[18,19] There is a small energy barrier between *endo* and *exo*, which have been calculated to be between 0.0 and 0.5 kcal mol^{-1},[18] around 0.6 kcal mol^{-1},[20] and around 0.3 kcal mol^{-1}.[19] These calculations match experimental observation of a nanosecond scale transition between the two configurations.[21–28] Changes in the relative population between the two configurations are a way to accommodate changes in the backbone configuration, hypothesised to be a mechanism of absorbing forces.[15,16] Indeed, such a change in the *endo*/*exo* populations is found around Gly to Ala mutations.[19] The methyl group in alanine requires more space, pushing the chains apart within the triple helix. As a result a force is exerted on the backbone, and the populations of *endo*/*exo*-configurations shift. As Y-position hydroxyproline is in the *exo*-configuration, stabilising the collagen fibrils,[5,17] in GPO collagen this mechanism

should lead to a shift in *endo*/*exo*-populations for the X-proline. For GPP collagen, the absence of hydroxyproline should lead to more flexibility, but a similar mechanism should be observable, albeit the response to force should be different from GPO collagen.

Much less is known about the reactive regime, apart from the fact that this process requires unravelling of the triple helix, which may lead to breaking of covalent bonds. While it is not known which bonds are likely to break, experimental work suggest that radical species are formed.[29] Radical formation has also been observed in other biomaterials.[30,31] In addition, it has been reported that bond breaking is located in the proximity of cross-linking sites.[12–14,29] Our target is to identify potentially preferred bonds to break, and study the effects of mutations and cross-linking on the location of bond ruptures.

The variation in the constituents of native collagen samples complicate studies of collagen. As a result collagen-like models are often used, in particular GPO and GPP repeats that capture some of the important chemical and physical aspects of collagen fibres. Even in these simplified models, it can be difficult to resolve molecular mechanisms in detail. An additional challenge in the context of mechanical behaviour is the application of forces, which complicate experimental setups. If covalent bonds are broken, an additional challenge is the lifetime of potential products, and whether the process and the resulting products can actually be characterised.

As a result, computational methods can support experimental work in this field. For the non-reactive regime, we used the computational potential energy landscape framework[32,33] to explore the energy landscapes resulting from different forces being applied. The method has previously been used to study the effect of forces on biomolecules,[34] and to study the effects of Gly to Ala mutations in collagen,[19] showing that the method is suited to resolve force-induced changes in collagen model structures. To study the reactive regime, we implemented a

novel simulation scheme to probe the products immediately after the tropocollagen ruptures, employing a semi-empirical potential.

We find evidence for the proposed mechanism of proline ring configuration changes that absorb weaker forces. In the reactive regime, we observe a preference for breaking the C–C$_\alpha$ bond in X-position residues. The introduction of mutations, deletions or cross linking shifts the preference for breaking into the vicinity of these perturbation. In fibre-like segments, the preference for breakages in the C–C$_\alpha$ bond in X-position residues is also observed, with ruptures in proline significantly overrepresented compared to the amino acid content.

2 Methodology

In the non-reactive regime, we used the computational energy landscape framework[32,33] to explore the energy landscapes associated with different forces for GPP and GPO model peptides. In brief, basin-hopping global optimisation[35–37] was used to obtain low-energy starting points for the pulled collagen triple helices. Using discrete pathsampling,[38,39] kinetic transition networks were constructed[40,41] by locating transition state candidates with the doubly-nudged elastic band algorithm.[42–44] Those candidate structures were converged with hybrid eigenvector-following,[45] and the respective local minima found by following approximate steepest-descent paths.

For GPO and GPP, the model proteins consisted of seven repeats per strand (21 residues per strand). The end of the strands were capped with methyl groups connected via peptide bonds (CH$_3$–CO (ACE) and NH–CH$_3$ (NME)). The properly symmetrised[46,47] AMBER ff14SB force field[48] was used with an implicit Generalised Born solvent representation.[49] The forces were applied at the C atom in ACE and the N atom in NME. The applied pulling potential is

$$V_{\text{pull}} = -f(z_1 - z_2),$$

where f is the applied force and z_i is the z coordinate of the atom the pulling force is applied to. Even small forces lead to alignment with the z-axis, meaning this setup results in a constant pulling force applied to the molecule. As an AMBER force field[48] is used, the force will be propagated through the harmonic bonding network. The results presented here are for the same force applied to all three strands. The applied forces were 10, 50, 100, 250, 500 and 750 pN.

For simulations in the reactive regime, we first had to establish the forces required to break bonds in these collagen peptides. The potential chosen for this part of the work is the semi-empirical GFN2 within xTB,[50] which allows for bond breaking. Harmonic constraints are employed in the terminal residues on either end to ensure that the bond breaking, if it occurs, is located away from the ends of the molecules, introducing more realistic breaking points when comparing the model peptides to tropocollagen. First, we established that bond breaking occurs reliably above forces of 6000 pN by running a series of local optimisations for collagen molecules

at various pulling forces. Two energy landscape explorations with xtb were run for forces of 3000 and 4000 pN, as a comparison to the lower force simulations in AMBER. It should be noted that these forces are strong enough to occasionally introduce bond breaking, but not often enough to prohibit the exploration of the energy landscape. For forces at around 6000 pN, bond breaking occurs in roughly 99% of cases, allowing for investigations of the rupturing behaviour. Importantly, this force is close to the point where we see no or very few bond ruptures, meaning we are close to the realistic rupturing threshold, in agreement with observations by others.[10] As for the non-reactive case, forces are applied in parallel to all three strands.

A problem within this reactive framework is posed by the constant force applied. If no fragments are formed, the forces are balanced across the molecules and local minima can be reliably located. However, when fragmentation occurs, the constant force means the fragments will continue drifting apart without convergence to a minimum. One solution is to simply turn of the force at some point, but we considered this as a rather abrupt change in the simulation protocol, which might potentially lead to artefacts in the simulation. Instead, we employed a catching potential. This potential is flat with steep walls, and only yields a significant contribution when the fragments are a specific length l apart. This length is chosen to exceed the length of the molecule significantly, so that fragmentation is possible. Once the fragments are separated by l, the catching potential counteracts the force, allowing convergence of the gradient and hence the location of local minima. The effect is that the potential catches the fragments and leaves them to hover without the necessity to change the applied force. The catching potential has the form

$$V_{\text{catch}} = \begin{cases} 0 & \text{if } d_{\text{e2e}} < l, \\ k\dfrac{\left[d_{\text{e2e}}^2 - l^2\right]^2}{2d_{\text{e2e}}^2} & \text{otherwise;} \end{cases}$$

where d_{e2e} is the end-to-end distance of the molecule, which corresponds to the distance between the atoms to which V_{pull} is applied. With this setup, we simulated bond breaking in a number of different sequences, for sequences that had be stretched with a non-reactive force and linked collagen segments. For each sequence and condition, a number of local minima were taken from energy landscape databases created for this study or that were previously published. Minima were selected at random from these databases.

Our simulation protocol focuses on the instantaneous bond ruptures that occurs under high mechanical forces. Within this protocol, forces are still distributed in the segments that the force is applied to, but we do not equilibrate forces. This approach will therefore give us an indication of bond rupturing mechanisms for sudden mechanical stresses applied to tropocollagen, mimicking rapid onset of forces during for example sudden motions.

We considered bond breaking in GPO and GPP repeats. In addition, we studied the impact of pre-tensioning and of strand length on the rupture behaviour in GPO repeats. Perturbations

This journal is © the Owner Societies 2023

Phys. Chem. Chem. Phys., 2023, 25, 2331–2341 | 2333

to GPO model peptides were introduced in the form of Gly to Ala mutations and Hyp deletions. For the mutation and the deletion, we chose short segments to optimise resource use, as the effects of mutations and deletions are local (see ESI† S3 and previous work[19]). Finally, more realistic fibre-like segments and cross-linked segments were probed, using collagen models obtained from ColBuilder.[8] Due to limits on the available resources, we only studied a single representative fibre-like sequence and two different crosslinks. The main objectives of this extension of the work are (a) to confirm whether or not any observed preference in the GPO models and related systems is reproducible in more realistic fibre-like segments, and (b) whether the experimentally observed location of rupture around crosslinks can also be reproduced. In particular with respect to crosslinks, this study does not aim to provide a comprehensive survey of rupturing in and around them.

Table S1 (ESI†) gives an overview of all simulations conducted for this part of the study.

3 Results

3.1 Changes in the proline puckering

As the puckering of proline has been suggested as a possible mechanism to absorb mechanical forces, the first property to consider when analysing the simulations for the non-reactive regime is the puckering state of the X and Y positions. The exploration of the energy landscape allows an ensemble calculation of structural properties, including higher energy structures. In Fig. 2A–C, the puckering configurations are illustrated for GPO and GPP model proteins. The exact values are provided in the ESI,† Tables S2 and S3. In line with the hypothesis by
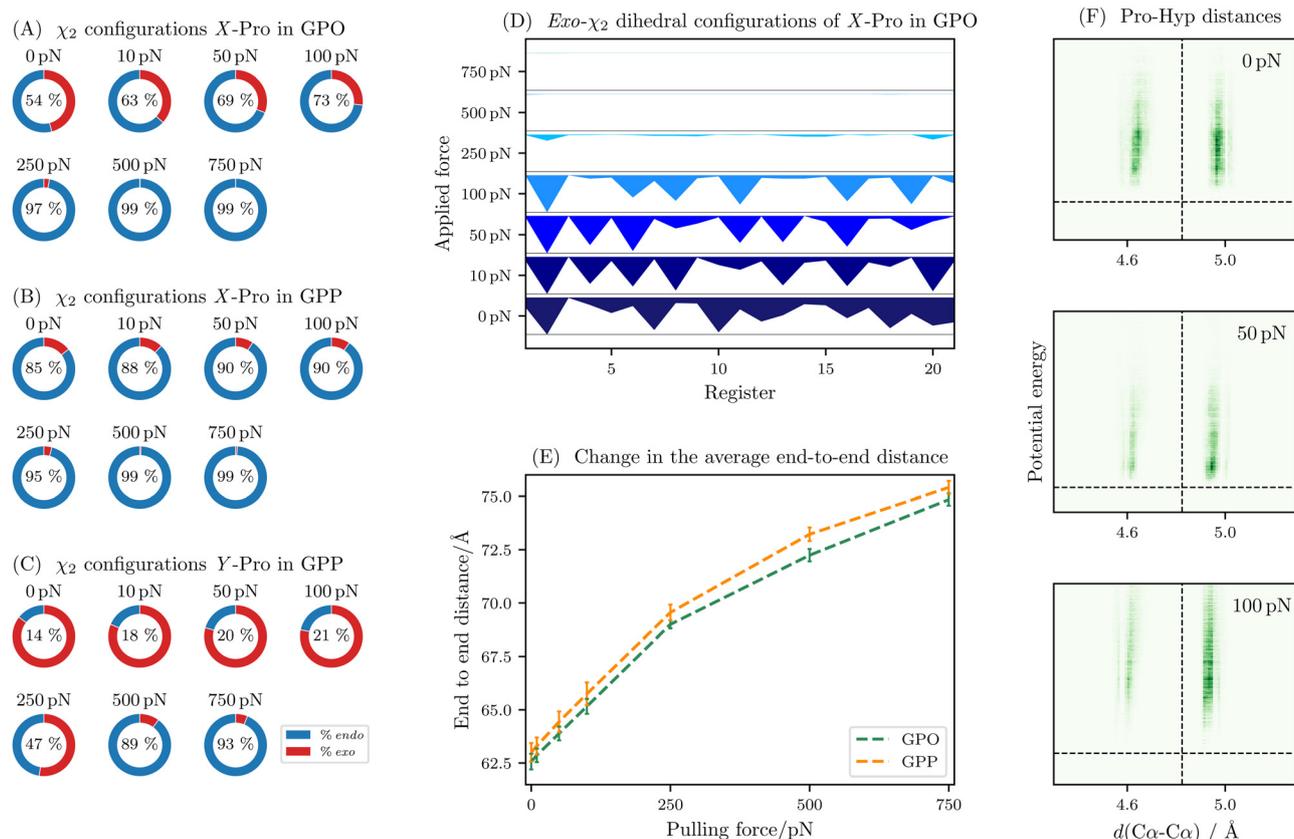

Fig. 2 Overview of structural descriptors for non-reactive pulling simulations. Left: changes to the *endo/exo* distribution of X-proline in GPO (A), X-proline in GPP (B) and the Y-proline in GPP (C) with increasing force. The values are the thermally averaged occupation over the entire molecule (21 registers). (D) Variation of the *endo/exo*-distribution across the registers for increasing pulling forces (going bottom to top) in GPO. For each register and force the distribution is shown as the thermally weighted average of occupation for the proline in the register. The *exo*-content is coloured in shades of blue, while the *endo*-content is in white. As the forces increase, the occupation of *exo*-configurations for the X-proline decreases (less and less blue area). Interestingly, the occupation is not uniformly decaying, but instead some residues are nearly exclusively *endo* at low and medium forces. (E) Change in the end to end distance averaged over the three strands and thermally averaged over all structures for GPO (green) and GPP (orange) with increasing pulling force. (F) Heatmaps of the distribution of the distance between the C_α atoms in Pro and Hyp in each register against the potential energy of the minimum. The average distance[19] (dashed, vertical line) and the global energy minimum (dashed, horizontal line) are shown as references. As the force increases from 0 pN (top) to 50 pN (middle) and 100 pN (bottom), the distribution shifts towards a preference for a longer distance, which is corresponding to the *endo*-configuration of X-proline. The distributions are proxies for the *endo/exo* configurations,[19] and provide additional insight as they show the relative energy of the different configurations.

2334 | *Phys. Chem. Chem. Phys.*, 2023, **25**, 2331–2341

This journal is © the Owner Societies 2023

Chow *et al.*,[15] we observe significant puckering changes with increasing force. For GPO, the X-proline residue first flips to an endo state, which near saturation between 250 and 500 pN applied force. At this point, the Y-hydroxyproline starts to flip towards endo states. At higher pulling forces, the Y-hydroxyproline is nearly exclusively in an endo configuration and the X-proline starts to be planar due to the high forces. For GPP, we

observe a similar behaviour with more and more endo content. Importantly, the transitions occur at lower forces, as expected from the mechanical stability induced by the hydroxylation of the Y-proline.

The transitions are also not uniform across the registers, in line with findings that the puckering angle is not distributed evenly along GPO repeats.[19] Fig. 2D shows that this effect is
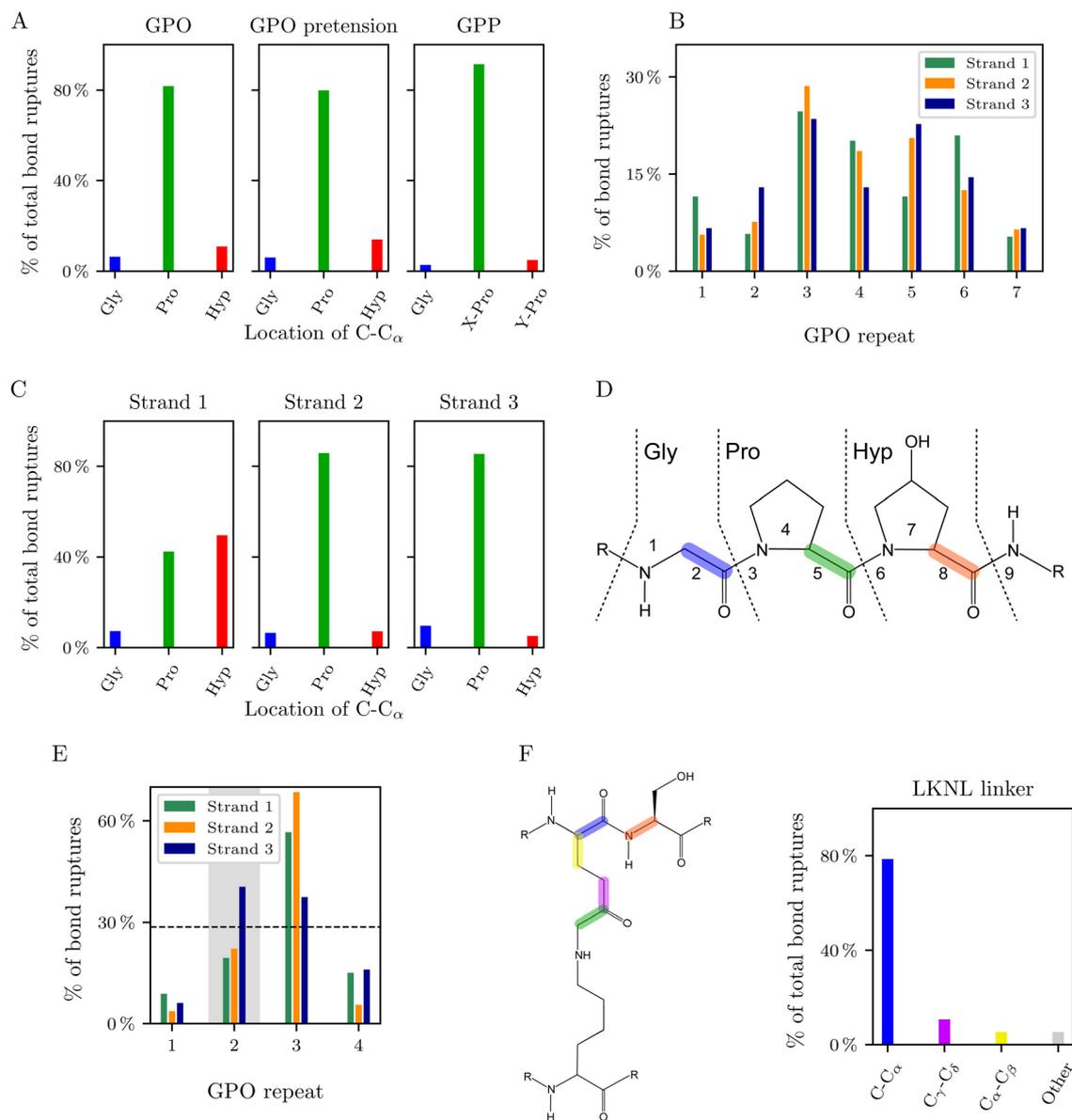

Fig. 3 Overview of the results for bond rupture simulations in collagen models. (A) Distribution of ruptures of the $C-C_\alpha$ bond by residue as percentage of the total number of bond ruptures for GPO model peptides (left), GPO model peptides equilibrated at 10 pN (middle) and for GPP model peptides (right) show a strong preference for bond ruptures in the X-position proline. (B) Distribution of ruptures over the seven GPO repeats, where the central 9 registers, which stretch repeats 3 to 6, are unrestrained. There is no clear locality of the ruptures and they are relatively evenly distributed across the sequence, accounting for the constraints applied to the ends of the collagen strands. (C) Shift in breakage location between the mutated strand 1 (one Gly to Ala mutation) and the unmutated strands when a single Gly to Ala mutation is introduced as percentage of the total bond breakages. The significant shift in the left panel, which is the mutated strand, is due to a high likelihood of breaking in the Hyp preceding the mutation. (D) Schematic of the bonds most likely to break as reference for panels A and C. (E) Bond breakage frequencies in interrupted sequences, *i.e.* where a deletion occurs in repeat 2 in the Y-position in strand 1, leads to ruptures closer to this deletion. (F) Scheme of the bond breaking in the LKNL linker with the relative observed frequency plotted. The highlighted bonds are the only bond for which ruptures were observed, and the percentage of the total ruptures are provided for them. The strong localisation of the bond ruptures around the linker is clear.

This journal is © the Owner Societies 2023

Phys. Chem. Chem. Phys., 2023, **25**, 2331–2341 | **2335**

pronounced at lower forces, where some registers, which are fairly evenly distributed along the length of the molecule, exhibit high percentages of *exo*-configurations. At higher forces, these disappear rapidly.

This behaviour is also reflected in the end to end distance change observed between 0 and 250 pN pulling force and the change observed between 250 and 750 pN, shown in Fig. 2E. In the former regime, which corresponds to the subsequent flipping of more and more X-proline rings, we see an elongation of approximately 0.032 Å per pN, while in the latter saturation regime this rate is reduced to around 0.01 Å per pN. We note that part of the extension will be due to stretching of soft modes in collagen. Given that most ring flipping occurs up to 250 pN, it is likely that this explains the change in behaviour.

The proline puckering can also be monitored by measuring the distance between the C_α atoms of the proline and hydroxyproline in the same register. While the regular structure of the GPO model peptide leads to a sharp distribution for the interchain distances for the distance between proline and glycine, and glycine and hydroxyproline, the puckering states of proline lead to two observed distances between proline and hydroxyproline. As the force increases, this distribution more and more tends to a single value, as shown in Fig. 2F.

3.2 Preferred bond rupture sites in collagen

The first consideration for our study of the bond rupture in collagen models was a survey of different pulling forces to identify the minimum force required to break collagen reliably. For pulling forces around 4000 pN we observed some bond breaking events. For higher pulling forces, we reliably observed such events with a pulling force of 6000 pN being sufficient for consistent rupturing. For even higher forces around 10,000 pN the rupture occurs abruptly, and may introduce artefacts in our simulations. Thus, we used 6000 pN throughout the remaining simulations. While these forces are high in the context of a single bond, it should be noticed that in tropocollagen multiple mechanisms are acting together to absorb and distribute forces. Among those are the proline ring loading described for the lower force regime, the loading of the hydrogen bonding forming the triple helix, and the stretching of a large number of backbone bonds in each chain. As such we would expect that individual bonds only experience a fraction of the force. Our findings here are consistent with the results reported by others,[10] and highlights the extraordinary mechanical stability observed in this biomaterial.

The bond breaking in the GPO model without any pre-stretching showed a clear preference for bond ruptures in proline residues. The bond most likely to break is the $C–C_\alpha$ bond leading to the formation of an aldehyde radical and a pyrrolidine radical, with 80 to 85% of all bond ruptures (see Fig. 3A (left)). We do not observe any significant effects from considering larger GPO strands with regards to the frequency of bond breakages. In the shorter GPO repeats, we observe a preference for a single proline residue in each strand. This observation may be a finite size effect, but we were not able to find a correlation to the puckering state of the structure pulled.

In the larger GPO model peptide, we do not observe such a preference, and the main difference in distribution is related to which residues we restraint in our setup, as shown in Fig. 3B. When a small pre-tension force is applied, no clear preference in which proline the rupture occurs is observed, but otherwise we observe no changes (see Fig. 3A (centre)). The second and third most likely fracture sites in GPO are the same bond, but in hydroxyproline and glycine, accounting for 5 to 6% of all ruptures each.

In GPP models, we also see bond breaking predominantly of the $C–C_\alpha$ bond in the X-proline. In fact, it is even more prevalent in GPP, with close to 90% of all observed bond breakages (see Fig. 3A (right)).

A question that arises in context of these results is their dependence on the simulation protocol and the associated errors. This issue has been partially addressed here, as we tested the influence of the harmonic constraints, the length of the segment we considered, and whether pre-tensioning has any effects on our results. While we observe small changes in the percentages, the overall picture remains very much the same. Further tests, such as much larger sample sizes are unfortunately limited by the computational cost. Regarding the significance of these findings, we can compare it to a random sample, *i.e.* where no preference is found regarding the position of residue. In that case, we expect a third of the ruptures to occur in the X-proline. The observed sample has a p-value that is much smaller than 0.01, showing statistical significance of our findings across these different systems.

3.3 Mutations introduce mechanical weakness

In previous work we demonstrated that Gly to Ala mutations lead to a significant change in the backbone around the mutation site, and an associated shift in the puckering state.[19] When we apply a rupturing force to these mutated structures, we observe a similar rupture pattern in the unmutated strands as we observe for the pure GPO molecules. However, a significant change occurs in the mutated strands. Here, the rupture occurs in the vicinity of the mutated residue. Nearly 80% of the ruptures in the mutated strand occur in the two residues on either site of the mutation and the mutated residue, with over 40% occurring in the preceding residue – in this case hydroxyproline. This results in a significant shift in the rupture location compared to GPO, GPP and the unmutated strands as shown in Fig. 3D.

3.4 Bond breaking localises around deletions

Another possible change observed in experiment is the deletion of residues in individual repeats. Such interruptions have important biological implications.[51–53] Unfortunately, experimental setups are challenging, and generally probe interruptions in all three chain,[54] rather than in one strand. Explorations of the energy landscapes for GPO and GPP model peptide-derived interrupted sequences with one deletion show localised effects of the interruption, where kinking of the tropocollagen is observed at the site of the deletion. This kinking has functional importance,[51] but also may change the structural stability.[53] The hydrogen bonding away from the deletion and between the uninterrupted strands is

not effected, while hydrogen bonding involving the atoms in the repeat after the deletion is lost. More detail on these structural details are provided in the ESI,† Section S3.

The interrupted sequences show a different selectivity for the rupture locations. While the bond breaking still mostly occurs in X-position proline residues in the C–C$_\alpha$ bond, the structural alterations due to the deletion impacts where along the strands the bond rupture events occur. The bond breakages occur in the repeats immediately following the deletion, most pronounced in the trailing and middle chain, with over 60% of ruptures for these strands located in those repeats (see Fig. 3E).

3.5 Cross-links are mechanically weak

Probing of cross-links between strands requires the creation of cross linked models. We derived such model using ColBuilder,[8] and studied two models for human collagen with dehydro-lysino-norleucine (deH-LNL) and lysino-keto-norleucine (LKNL) linkers. In this setup, we have the crosslink, and two

tropocollagen segment on either side. In both cases, we pull the C-terminus of the lysine-bound and the N-terminus of the noleucine-bound end of the linker. Unfortunately, only one of these models produced results. This finding is related to the simulation setup, as we pull the cross-linked strands in opposite directions to load the crosslink. The deH-LNL linker models showed a slight kinking in the strands that are cross-linked. In the simulations this kinking prevented sliding of the chains with respect to each other, and as a result the linker was not loaded. The friction of the two kinked tropocollagen segments was stronger than the force applied. As a result, only the terminal residues in the pulled strands were loaded, and ruptures exclusively occurred in the capping groups. In the LKNL model, the strands were able to slide along each other, allowing for the stress to be distributed along the strands and the linker. This distribution of the force led to breakages in mechanically weak places rather than purely based on the simulation setup. In Fig. 3F, a scheme is shown for the LKNL
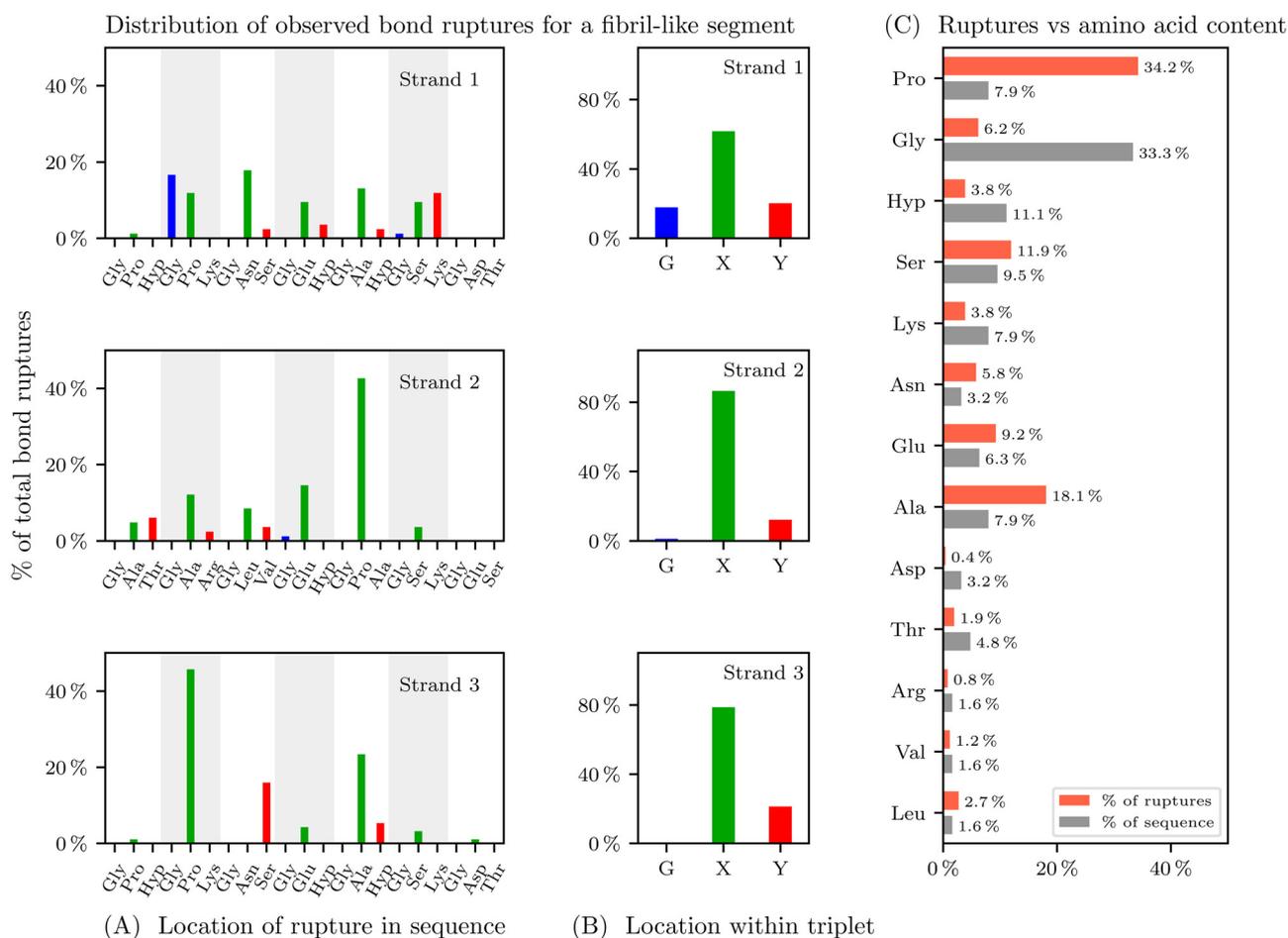

Fig. 4 Overview of the results for bond rupture simulations in a fibre-like segment of 21 residues per strand. (A) Distribution of the observed bond ruptures along the sequence. All bonds breaking are C–C$_\alpha$ bonds. Ruptures in Gly are shown in blue, ruptures in X-position residues are green, and those in Y-position residues in red. The individual GXY repeats are highlighted. A preference for X-position ruptures is observed. (B) Summary of the breakage location by position in the GXY triplets for the individual strands shown in the left panels. Most bond ruptures occur in the X-position (around 78% of all breakages observed). (C) Bond breakages by residue (red) compared to the amino acid content in the fibre-like segment (grey) shows a significant overrepresentation of breakages in Pro compared to its relative amino acid content in the strands. The only other significant overrepresented amino acid is alanine, with glycine and hydroxyproline underrepresented.

This journal is © the Owner Societies 2023

Phys. Chem. Chem. Phys., 2023, **25**, 2331–2341 | **2337**

linker and the observed bond breaking sites. Apart from a few bond breakages in the residue next to the cross-linked residues, most ruptures occur in the linked residues and the linker. Based on the linking and the setup, one C–C$_\alpha$ bond is loaded and nearly 80% of the observed bond breaking events are observed in this bond. While these results show localisation around the linker, this is not sufficient to provide more insight into the nature of this bond breaking. A possible future extension of this work would be an more extensive survey of different linkers, including trivalent ones. The observation of friction as a key mechanism to stop crosslink loading might also be investigated. A final note should be taken of the unequal distribution of rupture sites in the linker. A number of key features might be the cause of these observations, including the geometry of the tropocollagen of the linker, the simulation protocol and the nature of the linker. All of these require further investigation. A recent publication[55] has looked into the mechnical behaviour of crosslinks, and provides a potential explanation for non-symmetric linker breaking, based on the chemical stability of the involved bonds and the need to have well-defined bond breaking for better scavenging of defects. However, this study does not account for friction effects, leading to underestimates of the importance of backbone stability and orientation.

3.6 Fibre-like segments exhibit similar behaviour to collagen model peptides

The final set of simulations aimed to identify whether the described patterns so far are representative for more realistic models for collagen, *i.e.* more complex sequences than GPO and GPP repeats. For this part of the study we tested models for human collagen from ColBuilder.[8] The results for the bond breaking in these sequences are shown in Fig. 4. While some variance is observed in which residues contain the rupture location, which is expected from the variance in the amino acid

sequence (see Fig. 4), the tendency for breaking of C–C$_\alpha$ bonds in the X-position residues is still clear. We exclusively see bond breaking in C–C$_\alpha$ bonds, and on average across all strands 78% of all ruptures are in the X-position residues (see Fig. 4, middle panel), similar to the observations for GPO and GPP repeats (see Fig. 3A).

As we have more variation in the amino acid content, it is possible to compare the frequency of bond ruptures in specific amino acids to their relative content in the sequence. This information is shown in Fig. 4 in the right panel. The outstanding differences between the two percentages are observed for proline, alanine, glycine and hydroxyproline. In proline residues, 34% of all bond breaking events are recorded, while only around 8% of the sequence is made up by proline. A smaller, but still significant increase is observed for alanine, with 18% of all bond ruptures and only 8% of the amino acid content. In contrast, both glycine and hydroxyproline show relatively too few bond breaking events. In glycine residues, which make up 33% of the sequence, only 6% of the bond ruptures are observed. For hydroxyproline, we observe 4% of the bond breaks compared to 11% of the amino acid content.

4 Discussion

4.1 Molecular mechanisms of absorption of forces

The first important point to discuss is the mechanism of force absorption observed in the non-reactive regime. All observations in this study confirm the hypothesis of proline *endo/exo* flipping as the mechanism for non-reactive force absorption.[15] Importantly, this mechanism extend to ring flipping in Y-position hydroxyproline residues in GPO and proline residues in GPP. The proline ring conformation changes are exhibiting the lowest barrier and occur first.

The changes in the *endo/exo* distribution are not gradual, but flipping seems to occur residue by residue. Once the
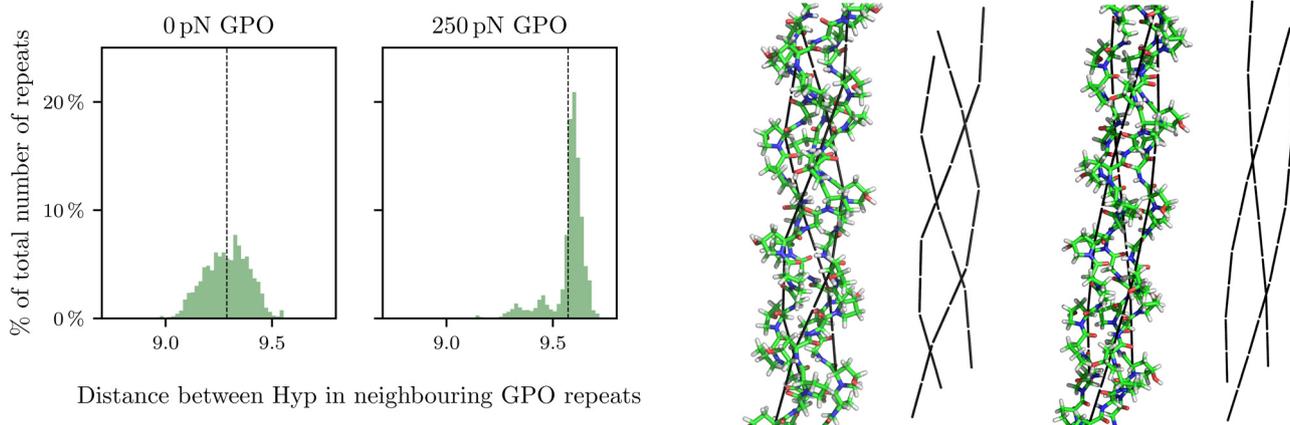

Fig. 5 An illustration of the changes in GPO repeat lengths when pulling forces are applied. Left: Distribution of the distance between the C$_\alpha$ atoms in neighbouring hydroxyprolines along the chains as a proxy for GPO length. The average values for the two forces are given by the dashed lines. As the force is increased and proline ring flips are introduced, the segments extend in length and the distribution becomes sharper. The average distances are 9.29 Å at 0 pN and 9.57 Å for 250 pN applied force. Right: The change in force and proline ring configuration leads to changes in the GPO repeat alignment and length. The lowest energy structures are shown for 0 pN (left) and 250 pN (right) with a representation of the vectors between C$_\alpha$ atoms in neighbouring hydroxyprolines shown in black. The forces lead to a higher symmetry and alignment of the GPO repeat orientations.

2338 | *Phys. Chem. Chem. Phys.*, 2023, **25**, 2331–2341

This journal is © the Owner Societies 2023

proline rings are in the *endo* configuration, the end-to-end distance increase is slowed and larger forces are required for the same additional extension. The ring flipping is related to alternatives in the backbone configuration, allowing the backbone to be in better alignment with the helical axis. As a result the end-to-end length for each GPO repeat increases with the applied force. This extension of repeat length is shown in Fig. 5 (left). As a result, GPO repeats, here represented as the vector from the C_α atom in Hyp in one repeat to the same atom in the next repeat, align much better with the helical axis (see Fig. 5 on the right). The proline residues act as flexible extenders, where at no and low applied forces the backbone is winding somewhat more within the framework of the hydrogen bonds that characterise tropocollagen. Larger pulling forces introduce better alignment, and the ring flip allows for this more extended backbone configuration. Likely this process happens segment by segment, as indicated by the residue by residue flipping observed, rather than in a continuous fashion. Importantly, no changes to the hydrogen bonding network is observed at this stage.

4.2 Bond breaking in collagen

Once forces exceed the threshold for bond breaking, bond ruptures are readily observed. It should be noted here that not only are our finding for this threshold in broad agreement with macrosocpic models,[10] but further that this is significantly larger than the forces required for unfolding observed in non-structural proteins.[34] Clearly, this structural resistance to forces stems from the regular hydrogen-bonding pattern and the in-built molecular mechanisms, like the proline ring flipping, and is faithfully reproduced in our simulations.

Two important preferences are observed in the location of the bonds that break. Firstly, the bond most likely to break are $C–C_\alpha$ bonds. This observation holds true throughout all of the models studied. Two reasons can be identified for this bond as the most likely breakage site. Firstly, both resulting radicals, with the exception for glycine, lead to a radical on a secondary carbon and an amidyl radical (radical on the amide nitrogen). Both are stabilised by hyperconjugation,[56] albeit rearrangements may occur subsequently. Secondly, radicals are more stable on elements with lower electronegativity, and all other backbone options involve C and N rather than two C atoms.

The second key observation about the location of the bond breaking is the preference for X-position residues. Again this is observed across all models with around 80% of all observed ruptures in these residues. This preference might stem from the unique position of X-position residues within the collagen assembly. The X-position residue is hydrogen bonded to glycine, and the resulting network is the fundamental contributor to the tropocollagen structure and help to distribute the applied forces throughout the collagen chains.

As observed in longer repeats, this distribution works well, will fracture sites fairly evenly distributed across the length of the segments considered. In the GPO and GPP repeats the preference for X-position ruptures automatically leads to a preference of ruptures in proline residues. However, these models are only representative to a point. In this case, we need to consider the fibre-like segments to see whether the same distribution of forces through hydrogen bonding still leads to a preference in ruptures located in proline residues. Indeed, such a preference is observed for these fibre-like models, with a clear preference for ruptures to occur in proline. The mechanism behind this observation is likely related to the force absorption described earlier. Not only to we see ring flipping towards exclusively *endo*-configurations, but moreover at even higher forces rings start to become planar. As a result, the proline bonding is destabilised by the absorption of force, likely leading to the observed strong preference in ruptures.

While the rupture preferences for X-residues clearly emerge, and we can also assign an important role to proline in the low force regime, it is harder to quantify these two effects. For the reactive regime, a lot of the preference for proline might be explained by the preference for the X-position. From the results for the fibre-like segment we see more variation between strands, for example between strand 1 and 3, despite them having identical sequences (see Fig. 4). As the leading strand will have a different environment compared to the trailing strand, even if their sequences are identical, it is clear that the local environment impacts bond rupturing. The comparison to more idealised GPO and GPP repeats allows us to identify the overall preferences, but within fibre-like segments we encounter much more noise. As such, our data set for the more realistic fibre segment is not large enough and varied enough to characterise such effects, and further work is required here.

4.3 The effect of mutations, deletions and cross-linking

The above results yield interesting insight into the molecular mechanisms behind the structural stability of collagen. Additional interest stems from biological changes, such as crosslinking and sequence changes. Crosslinking is key to the formation of strong collagen tissues and crosslinks have previously been identified as a likely place for mechanical ruptures.[12–14,29]

Sequence mutations, such as the Gly to Ala mutation, have been associated with hereditary diseases that impact the mechanical properties of collagen tissues.[57] Similarly, while sequence interruptions of the form GX–GXY provide important functional binding motifs,[51] the same motif in non-binding motifs may lead to structural destabilisation.[53]

For all three of these motifs, we observe localisation of bond ruptures in the proximity of these perturbations. In the cases of the mutation and interruption, the local hydrogen bonding pattern and backbone arrangements are changed. These effects can be seen in the alterations of the ring puckering in the Gly to Ala mutated systems[19] and the kinking and hydrogen-bonding interruption in the interrupted sequences (see ESI† S3). As described above, the hydrogen bonding network distributes forces across the three strands and increases the overall stability. The weakest sites, in the unperturbed models proline residues in X-positions, are then the first to rupture. The interruption of the hydrogen bonding and the restriction of the backbone orientations close to the mutation and deletion sites lead to structural weaknesses, and hence start failing. The crosslinking is not

This journal is © the Owner Societies 2023

Phys. Chem. Chem. Phys., 2023, **25**, 2331–2341 | **2339**

supported by such a hydrogen bonding network, and once the linked chains can slide along each other, the crosslink will be loaded. There is no mechanism to provide additional strength to these parts of the structure, leading to bond breaking in and around the linker, as the collagen strands involved are stabilised.

A limitation of our study lies in the choice of systems under investigation. The systems we have selected represent in our opinion a representative cross-section of features of interest found in tropocollagen. A number of additional systems that can be studied in this way are available. Firstly, while we focused on GPY systems for a large part of our study, another common repeat is GAO, with alanine in the place of the important X-proline. As discussed above, there is not only an argument for the important role of proline, but also for the importance of the X-residue in general, due to the hydrogen bonding patterns. We also see alanine overrepresented as a rupture site in our fibre-like simulations. We therefore already observe similar behaviour of alanine to proline in the reactive regime, and further study of these repeats will be insightful. Another potential extension is the simulation of other glycine mutations. Alanine has a smaller effect than other substitutions, as it is still a fairly small amino acid. We would expect other mutations to show similar, but potentially stronger effects to the ones we observed for the Gly to Ala mutation here. Finally, we focused on a single crosslink, mainly to understand whether crosslinking would lead to a change in the bond rupturing pattern, moving the bond ruptures towards the vicinity of the linker. A more detailed on these effects is desirable, and more detail about how crosslinkers are breaking has been published as a preprint, while this study was under review.[55]

5 Conclusions

In this study we investigated the molecular mechanisms of force response in tropocollagen, considering both non-reactive and reactive responses. For the simulations of reactive responses, *i.e.* the rupturing of bonds as a result of applied pulling forces, we implemented a novel simulation setup.

For the non-reactive response, we find evidence to support a hypothesis by Chow *et al.*[15] that proline puckering configurations are key to the absorption of applied forces. Not only do we see changes in the X-proline puckering configurations, but also in Y-hydroxyproline and Y-proline in GPO and GPP models, respectively. The X-proline adapts first, before the Y-residues are impacted. Puckering changes seem to happen segment by segment, and lead to longer GXY repeats, which are aligned symmetrically to the helical axis.

When forces are large enough for bonds to break, these breakages are mostly occurring in X-position residues, independent on whether we consider GPO and GPP model peptides or more realistic fibre-like segments. Almost exclusively, the $C–C_\alpha$ bond ruptures. Rupture sites are fairly evenly distributed in GPO and GPP repeats. When fibre-like segments are considered, we see a strong preference for ruptures in proline residues compared to their relative amino acid content. Likely,

this observation is related to the loading of proline residues as force absorbing residues.

When the tropocollagen is altered *via* mutations or deletions, the interruption of the hydrogen bonding network and backbone orientation changes lead to localisation of bond breakages in the vicinity of these interruptions. Crosslinking similarly leads to localisation of bond ruptures, which matches experimental observations.

Overall, we conclude that bond breaking under tensile stress in collagen is highly selective, and proline residues are central to our molecular understanding of the force response. Mutations and deletions both weaken the collagen chains mechanically, and crosslinks are also weak points within collagen tissues.

Our findings have interesting implications for the design of biomaterials. The mechanism of in-built extensions through the proline ring flipping at low forces might be a useful principle to add mechanical stability to designed materials. A key principle for this mechanism is the close match in energy between the states, leading to a near equal distribution of both states at equilibrium and the small change in length between the states. Another important finding for future design efforts is in the methodology we present here. The simulations of instantaneous bond breaking relies on a semi-empirical method, and therefore can be extended to a large number of systems. It provides insight into likely breaking points, and highlights how modifications affect stability. Finally, the insights into the mechanisms of bond breaking due to mutations and deletions shows that these disease-relevant changes are fundamentally affecting the backbone stability. These changes are due to structural effects, and while they are localised, such changes in tropocollagen are always primed for mechanical failure.

Author contributions

JR ran the simulations. KR conceived the research idea and setup, and supervised JR. Both authors analysed and interpreted the data. KR wrote this manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 M. Aumailley and B. Gayraud, *J. Mol. Med.*, 1998, **76**, 253–265.
2 C. Frantz, K. M. Stewart and V. M. Weaver, *J. Cell Sci.*, 2010, **123**, 4195–4200.

2340 | *Phys. Chem. Chem. Phys.*, 2023, **25**, 2331–2341

This journal is © the Owner Societies 2023

3 C. Bonnans, J. Chou and Z. Werb, *Nat. Rev. Mol. Cell Biol.*, 2014, **15**, 786–801.

4 B. Brodsky and A. V. Persikov, in *Fibrous Proteins: Coiled-Coils, Collagen and Elastomers*, ed. D. Parry and J. Squire, Academic Press, Cambridge (MA), USA, 2005, vol. 70, pp. 301–339.

5 M. D. Shoulders and R. T. Raines, *Annu. Rev. Biochem.*, 2009, **78**, 929–958.

6 T. V. Burjanadze, *Biopolymers*, 1979, **18**, 931–938.

7 S. M. Krane, *Amino Acids*, 2008, **35**, 703–710.

8 A. Obarska-Kosinska, B. Rennekamp, A. Ünal and F. Gräter, *Biophys. J.*, 2021, **120**, 3544–3549.

9 M. J. Buehler, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 12285–12290.

10 M. J. Buehler and S. Y. Wong, *Biophys. J.*, 2007, **93**, 37–43.

11 S. Eppell, B. Smith, H. Kahn and R. Ballarini, *J. R. Soc., Interface*, 2006, **3**, 117–121.

12 B. Depalle, Z. Qin, S. J. Shefelbine and M. J. Buehler, *J. Mech. Behav. Biomed. Mater.*, 2015, **52**, 1–13.

13 A. Gautieri, S. Vesentini, A. Redaelli and M. J. Buehler, *Nano Lett.*, 2011, **11**, 757–766.

14 J. L. Zitnay, Y. Li, Z. Qin, B. H. San, B. Depalle, S. P. Reese, M. J. Buehler, S. M. Yu and J. A. Weiss, *Nat. Commun.*, 2017, **8**, 14913.

15 W. Y. Chow, C. J. Forman, D. Bihan, A. M. Puszkarska, R. Rajan, D. G. Reid, D. A. Slatter, L. J. Colwell, D. J. Wales, R. W. Farndale and M. J. Duer, *Sci. Rep.*, 2018, **8**, 13809.

16 I. Goldberga, R. Li and M. J. Duer, *Acc. Chem. Res.*, 2018, **51**, 1621–1629.

17 M. D. Shoulders, K. A. Satyshur, K. T. Forest and R. T. Raines, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 559–564.

18 W. Y. Chow, D. Bihan, C. J. Forman, D. A. Slatter, D. G. Reid, D. J. Wales, R. W. Farndale and M. J. Duer, *Sci. Rep.*, 2015, **5**, 12556.

19 K. Röder, *Phys. Chem. Chem. Phys.*, 2022, **24**, 1610–1619.

20 M. Cutini, M. Bocus and P. Ugliengo, *J. Phys. Chem. B*, 2019, **123**, 7354–7364.

21 L. S. Batchelder, C. E. Sullivan, L. W. Jelinski and D. A. Torchia, *Proc. Natl. Acad. Sci. U. S. A.*, 1982, **79**, 386–389.

22 L. W. Jelinski, C. E. Sullivan, L. S. Batchelder and D. A. Torchia, *Biophys. J.*, 1980, **32**, 515–529.

23 H. Saitô, R. Tabeta, A. Shoji, T. Ozaki, I. Ando and T. Miyata, *Biopolymers*, 1984, **23**, 2279–2297.

24 H. Saitô and M. Yokoi, *J. Biochem.*, 1992, **111**, 376–382.

25 D. Reichert, O. Pascui, E. R. de Azevedo, T. J. Bonagamba, K. Arnold and D. Huster, *Magn. Reson. Chem.*, 2004, **42**, 276–284.

26 D. Huster, J. Schiller and K. Arnold, *Magn. Reson. Med.*, 2002, **48**, 624–632.

27 S. K. Sarkar, C. E. Sullivan and D. A. Torchia, *J. Biol. Chem.*, 1983, **258**, 9762–9767.

28 S. K. Sarkar, C. E. Sullivan and D. A. Torchia, *Biochemistry*, 1985, **24**, 2348–2354.

29 C. Zapp, A. Obarska-Kosinska, B. Rennekamp, M. Kurth, D. M. Hudson, D. Mercadante, U. Barayeu, T. P. Dick, V. Denysenkov, T. Prisner, M. Bennati, C. Daday, R. Kappl and F. Gräter, *Nat. Commun.*, 2020, **11**, 2315.

30 H. Chandra and M. C. R. Symons, *Nature*, 1987, **328**, 833–834.

31 M. C. Symons, *Free Radical Biol. Med.*, 1996, **20**, 831–835.

32 J. A. Joseph, K. Röder, D. Chakraborty, R. G. Mantell and D. J. Wales, *Chem. Commun.*, 2017, **53**, 6974–6988.

33 K. Röder, J. A. Joseph, B. E. Husic and D. J. Wales, *Adv. Theory Simul.*, 2019, **2**, 1800175.

34 D. J. Wales and T. Head-Gordon, *J. Phys. Chem. B*, 2012, **116**, 8394–8411.

35 Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 6611–6615.

36 Z. Li and H. A. Scheraga, *J. Mol. Struct.*, 1988, **48**, 333–352.

37 D. J. Wales and J. P. K. Doye, *J. Phys. Chem. A*, 1997, **101**, 5111–5116.

38 D. J. Wales, *Mol. Phys.*, 2002, **100**, 3285–3305.

39 D. J. Wales, *Mol. Phys.*, 2004, **102**, 891–908.

40 F. Noé and S. Fischer, *Curr. Opin. Struct. Biol.*, 2008, **18**, 154–162.

41 D. J. Wales, *Curr. Opin. Struct. Biol.*, 2010, **20**, 3–10.

42 G. Henkelman and H. Jónsson, *J. Chem. Phys.*, 1999, **111**, 7010–7022.

43 G. Henkelman and H. Jónsson, *J. Chem. Phys.*, 2000, **113**, 9978–9985.

44 S. A. Trygubenko and D. J. Wales, *J. Chem. Phys.*, 2004, **120**, 2082–2094.

45 L. J. Munro and D. J. Wales, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1999, **59**, 3969–3980.

46 E. Małolepsza, B. Strodel, M. Khalili, S. Trygubenko, S. N. Fejer and D. J. Wales, *J. Comput. Chem.*, 2010, **31**, 1402–1409.

47 E. Małolepsza, B. Strodel, M. Khalili, S. Trygubenko, S. N. Fejer and D. J. Wales, *J. Comput. Chem.*, 2012, **33**, 2209.

48 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.

49 A. Onufriev, D. Bashford and D. A. Case, *Proteins*, 2004, **55**, 383–394.

50 C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.

51 S. Oka, N. Itoh, T. Kawasaki and I. Yamashina, *J. Biochem.*, 1987, **101**, 135–144.

52 N. Yamaguchi, P. D. Benya, M. van der Rest and Y. Ninomiya, *J. Biol. Chem.*, 1989, **264**, 16022–16029.

53 B. G. Hudson, K. Tryggvason, M. Sundaramoorthy and E. G. Neilson, *N. Engl. J. Med.*, 2003, **348**, 2543–2556.

54 J. Bella, J. Liu, R. Kramer, B. Brodsky and H. M. Berman, *J. Mol. Biol.*, 2006, **362**, 298–311.

55 B. Rennekamp, C. Karfusehr, M. Kurth, A. Ünal, K. Riedmiller, G. Gryn'ova, D. M. Hudson and F. Gräter, Collagen breaks at weak sacrificial bonds taming its mechanoradicals, *bioRxiv*, 2022, preprint, DOI: 10.1101/2022.10.17.512491.

56 J. Hioe, D. Šakić, V. Vrček and H. Zipse, *Org. Biomol. Chem.*, 2015, **13**, 157–169.

57 P. M. Royce, B. Steinmann, P. H. Byers and W. G. Cole, *Osteogenesis Imperfecta*, Wiley-Liss, New York, Chichester, 2nd edn, 2002, pp. 385–430.

This journal is © the Owner Societies 2023

Phys. Chem. Chem. Phys., 2023, **25**, 2331–2341 | **2341**