Supplementary Material: Chemical bonds in collagen rupture selectively under tensile stress

James Rowe,[1] Konstantin Röder[1*]

[1]Yusuf Hamied Department of Chemistry, University of Cambridge, UK

*To whom correspondence should be addressed; E-mail: kr366@cam.ac.uk.

S1 Overview of simulations conducted

(a) Energy landscape explorations

System	Force (pN)	Potential
$[CH_3\text{-}CO\text{-}(GPO)_7\text{-}CH_3]_3$	0, 10, 50, 100, 250, 500, 750	AMBER ff14SB
$[CH_3\text{-}CO\text{-}(GPP)_7\text{-}CH_3]_3$	0, 10, 50, 100, 250, 500, 750	AMBER ff14SB
$[CH_3\text{-}CO\text{-}GPO\text{-}GPO\text{-}GP\text{-}CH_3]_3$	3000, 4000	xTB GFN2

(b) Bond breaking simulations

System	Additional alterations	Number of residues per strand	Unrestrained residues per strand	Number of structures
GPO	no initial force	8	5	200
	no initial force	21	9	100
	pre-tensioned at 10 pN	8	5	200
GPP	no initial force	9	9	100
GPO	Gly to Ala mutation in the leading strand only	12	6	100
GPO	deletion of Hyp in the trailing strand only	12	9	100
Cross link	LKNL cross link	8	8	50
Cross link	deH-LNL cross link	8	8	50
Segment of *Homo sapien* tropocollagen	no initial force	21	21	50

Table S1: Overview of all simulations conducted for this study. a) Details of energy landscape explorations with DPS, b) details of the bond breaking simulations. All systems contain three chains, apart from the cross-linked systems, which contain six chains.

S2 *endo/exo* percentages for non-reactive pulling simulations

(a) AMBER databases

Forces/ pN	G**P**O % endo	GP**O** % exo
0	54.05 %	100.00 %
10	63.22 %	100.00 %
50	69.10 %	100.00 %
100	73.05 %	100.00 %
250	97.13 %	99.99 %
500	99.77 %	99.81 %
750	99.96 %	98.74 %

(b) xTB databases

Forces/ pN	G**P**O % endo	G**P**O % exo
3000	65.11 %	2.15 %
4000	77.33 %	0.28 %

Forces/ pN	GP**O** % endo	GP**O** % exo
3000	98.80 %	0.67 %
4000	99.37 %	0.04 %

Table S2: Puckering percentages for the AMBER (a) and xTB databases (b) for the GPO model protein. For AMBER, the endo percentage for the X-PRO and the exo percentage for the Y-HYP are given. Missing values to 100 % are the exo and endo state, respectively. For xTB, both endo and exo percentages are reported for the X-proline and Y-hydroxyproline. Missing values to 100 % are planar proline rings.

Forces/ pN	G**P**P % endo	GP**P** % endo
0	85.68 %	14.47 %
10	88.00 %	18.74 %
50	90.79 %	20.58 %
100	90.43 %	21.61 %
250	95.88 %	47.59 %
500	99.28 %	89.85 %
750	99.16 %	93.81 %

Table S3: Thermal average of puckering percentages for the GPP model protein for both the X- and the Y-position proline residues. Missing values to 100 % are exo-configurations.

S3 Structural properties of the interrupted sequences

The interrupted sequences consist of GPY repeats, where Y is either O or P, with a single Y-position deleted in the trailing strand. The sequence around that interruption is hence GPO-GP-GPO or GPP-GP-GPP. Energy landscapes were explored for both, GPO and GPP repeats, and the structural ensembles surveyed. In both cases we make two main observation.

Firstly, the structures kink around the deletion, as shown in Fig. S1 (left). The second observation is that all changes are localised around the deletion, as the hydrogen bonding pattern characteristic for tropocollagen is conserved away from the interruption, as schematically shown in Fig. S1 (right).

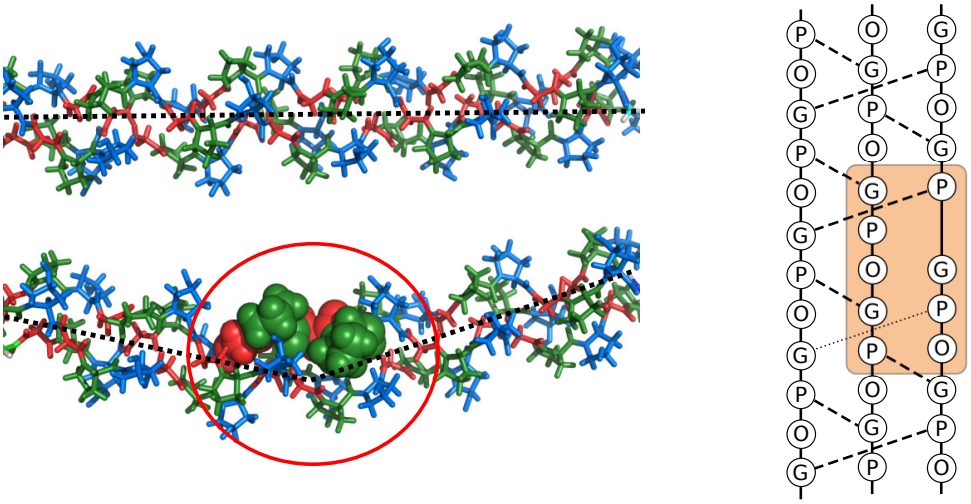

Figure S1: Structural changes observed in GPO repeat sequences when a single Y-position hydroxyproline is deleted in the trailing strand (so called interrupted sequences). Left: Kinking is observed in the interrupted sequence (bottom), around the deletion (spheres), compared to the GPO model peptide straight structure (top). Right: Hydrogen bonding patterns in the collagen with a deletion. The changes are localised around the deletion (box), with the hydrogen bonding between the trailing strand and the other two strands interrupted. The hydrogen bond involving the glycine in the repeat after the deletion is completely absent, and the hydrogen bond involving the proline in the same repeat is partly absent, and only observed in 40 % of structures in GPO and 77 % in GPP repeats.