Characterizing the Flexibility of the Body's Building Block: Collagen

by Alaa Al-Shaer[1], Aaron Lyons, and Nancy R. Forde

Alaa Al-Shaer[a] <aalshaer@sfu.ca>,
Aaron Lyons[b,c] <alyons@ualberta.ca>,
Nancy Forde[b] <nforde@sfu.ca>

[a]Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC V5A 1S6

[b]Department of Physics, Simon Fraser University, Burnaby, BC V5A 1S6

and

[c]Department of Physics, University of Alberta, Edmonton, AB T6G 2E1

WHAT IS COLLAGEN?

Collagen is the fundamental building block of animals and is the most abundant protein in our bodies [1,2]. Its evolutionary origins trace back to the dawn of metazoan organisms, where it supported the coalescence of multicellular life forms [3,4].

Different types of collagen are localized to different tissues and can form various higher-order structures. The most abundant class are the fibril-forming collagens, which align in a staggered parallel fashion and give rise to long fibrillar nanostructures (Fig. 1A). The basic structural unit of this fibril is the 300-nm-long triple-helical collagen molecule. Assembled from three polypeptide chains in the inner compartments of a cell, the molecule is trafficked to the outside of our cells where it assembles into higher-order structures, such as the fibril. However, collagen molecules can also assemble into networks, as shown in Fig. 1B. These networks play an essential role in providing a filtration barrier to macromolecules around organs such as kidneys [5]. Whereas fibrils consist primarily of collagen type I molecules, collagen network structures are typically formed by collagen type IV molecules. In this short article, we report on the mechanical properties of these collagen molecules, comparing the fibril-forming collagen type I to the network-forming collagen type IV.

We have characterized the different collagen types by analyzing them at the molecular level. Collagen's distinct triple-helical structural feature is shared amongst all twenty-nine collagen types, where the triple helix arises due to the polypeptide chains having a unique amino acid composition that is a repeating unit of Gly-X-Y. Having this glycine every third residue is a structural requirement that confers compactness and hydrogen-bonding abilities to stabilize the triple helix [1]. A key feature that differentiates collagen type IV from fibrillar collagens lies in this repeating unit, where collagen type IV has missing glycine residues which lead to triple-helical interruptions within the molecule. A central question is how these interruptions contribute to the flexibility of collagen, and this is what we aimed to address in our studies.

Summary

This article investigates the structural impact of local sequence variations along the contour of collagen, by sampling AFM-based images and performing statistical analyses.

ATOMIC FORCE MICROSCOPY

We use atomic force microscopy (AFM) as our experimental tool to study the mechanical properties of collagen molecules. AFM is a type of scanning probe microscope that utilizes a sharp tip to image the nanoscale topological features of a surface. By using a laser detection system, the instrument can record the deflection of the tip as it scans across the surface to yield a height map of any features present. We start our sample preparation by depositing collagen molecules onto mica – an atomically flat surface commonly used for imaging biological samples. Once deposited, the molecules adhere to the surface [6]. When imaged with AFM, collagen molecules appear as elevated string-like features on the flat mica background (Fig. 1). The AFM images provide us with an ensemble of conformations adopted by the collagen molecules, and we can use these conformations to extract information related to their mechanical properties. But before this is possible, we need a quantitative way to gain access to their conformations.

A custom-built MATLAB code, SmarTrace, was developed to analyze AFM images of individual collagen molecules [6]. The code first extracts the backbone coordinates of each chain in an image. To do so, SmarTrace requires minimal user intervention – the selection of only a few points near each chain – and uses a correlation-based algorithm to best match the backbone of the chain. From these backbone contours, we implement polymer physics tools to determine molecular flexibility.

POLYMER THEORY

Once a set of collagen chains has been traced, the coordinates along the backbone and the tangent vectors along

1. Alaa Al-Shaer received 3rd place in the CAP Best Student Oral Presentation competition at the 2019 CAP Congress at Simon Fraser University in Burnaby, BC.

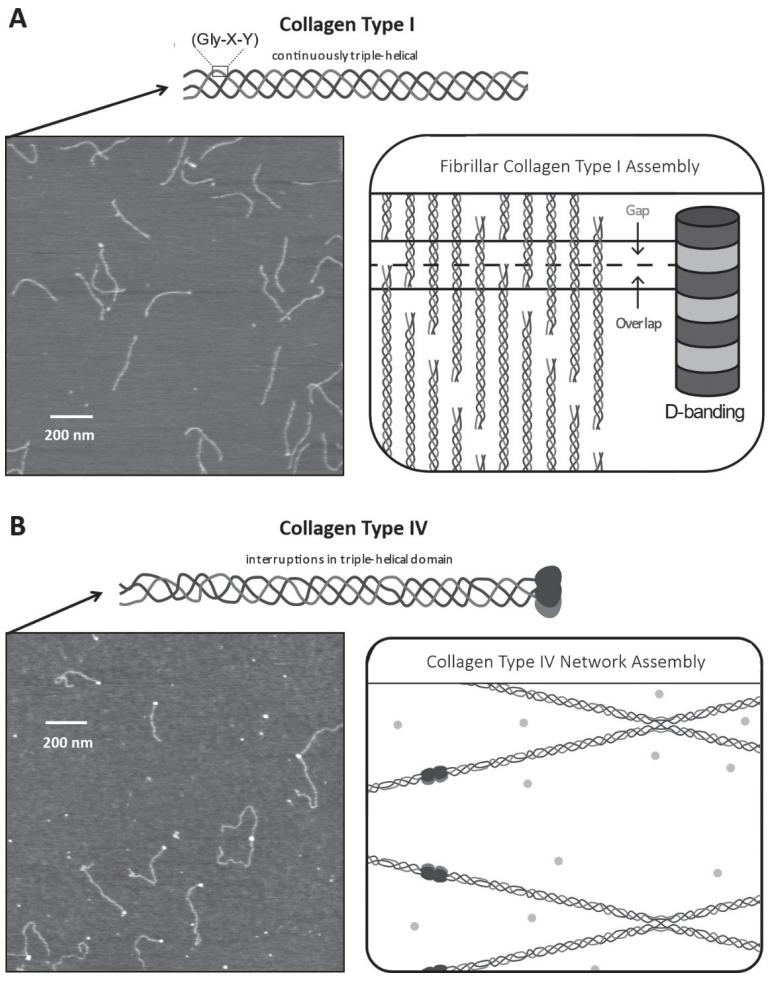

Fig. 1 **Atomic Force Microscopy (AFM) and supramolecular organizations of fibrillar and network-forming collagens.** AFM images of **(A)** collagen type I and **(B)** collagen type IV molecules deposited from 100 mM KCl, 1 mM HCl (pH ~ 3) with schematics of their corresponding molecular structures and supramolecular assembles placed above and to the right of the AFM images, respectively.

the chains are collected by the program. To relate these conformations to collagen's mechanical properties, we use the worm-like chain model, depicted schematically in Fig. 2A. The model describes a semi-flexible polymer, where short segments adopt straight conformations (the lowest-energy state) at zero temperature. In the presence of thermal fluctuations of energy, $k_B T$, the segments adopt smoothly curved conformations. The energy required to bend a segment of length Δs into an arc with angle θ between its ends is related to the polymer's persistence length, which we denote with the symbol p:

$$E_{bend} = \frac{p\theta^2}{2\Delta s} k_B T. \qquad (1)$$

The persistence length is best described. as the length along the contour over which the correlation of chain tangent vectors drops by a factor of e. Thus, persistence length characterizes the chain's flexibility, and is the mechanical property on which we focus.

We calculated the mean squared end-to-end distance $\langle R^2(\Delta s)\rangle$ and the mean dot product of the starting and ending unit tangent vectors $\langle \hat{t}(s).\hat{t}(s+\Delta s)\rangle$ for segments of length ΔS within the pool of traced chains. The mean squared end-to-end distances and tangent vector correlation data are fit with the predictions of the worm-like chain (WLC) model, described in Eqs. 2 and 3, to estimate persistence length (Fig. 2B).

$$\langle R^2(\Delta s)\rangle = 4p\Delta s\left[1-\frac{2p}{\Delta s}\left(1-e^{-\frac{\Delta s}{2p}}\right)\right] \qquad (2)$$

$$\langle \hat{t}(s)\cdot\hat{t}(s+\Delta s)\rangle = \langle \cos\theta(\Delta s)\rangle = e^{-\frac{\Delta s}{2p}} \qquad (3)$$

The plots in Fig. 2B show that collagen can be well-described by the WLC model.

The persistence length obtained for collagen type IV ($p \approx 41$ nm), reveals that it is significantly more flexible than fibrillar collagen type I ($p \approx 104$ nm). Could this enhanced flexibility be attributed to the presence of triple-helical interruptions? We decided to take a closer look at the amino acid sequence of collagen type IV and map out where these interruptions are along the contour of the molecule.

SEQUENCE-DEPENDENT ANALYSIS

To perform a sequence-dependent analysis, we need a sense of directionality to the chains so that we can align them at the same end. The collagen type IV molecules have a discernable "knob" at one end, which marks its C-terminal globular domain, and so we use it as a marker to trace chains starting at the edge of the knob. A schematic structure of collagen type IV is represented in Fig. 3B, where the boxes correspond to tripe-helical segments, and gaps to overlapping interruptions where all three chains possess a non-triple-helix-forming sequence. One can see that the chains have more interruptions at the end further from the knob. In addition, by inspecting the AFM images, we can see that the collagen IV molecules appear to be more flexible towards that end as well. To quantitatively show this, we devised an algorithm that performs a

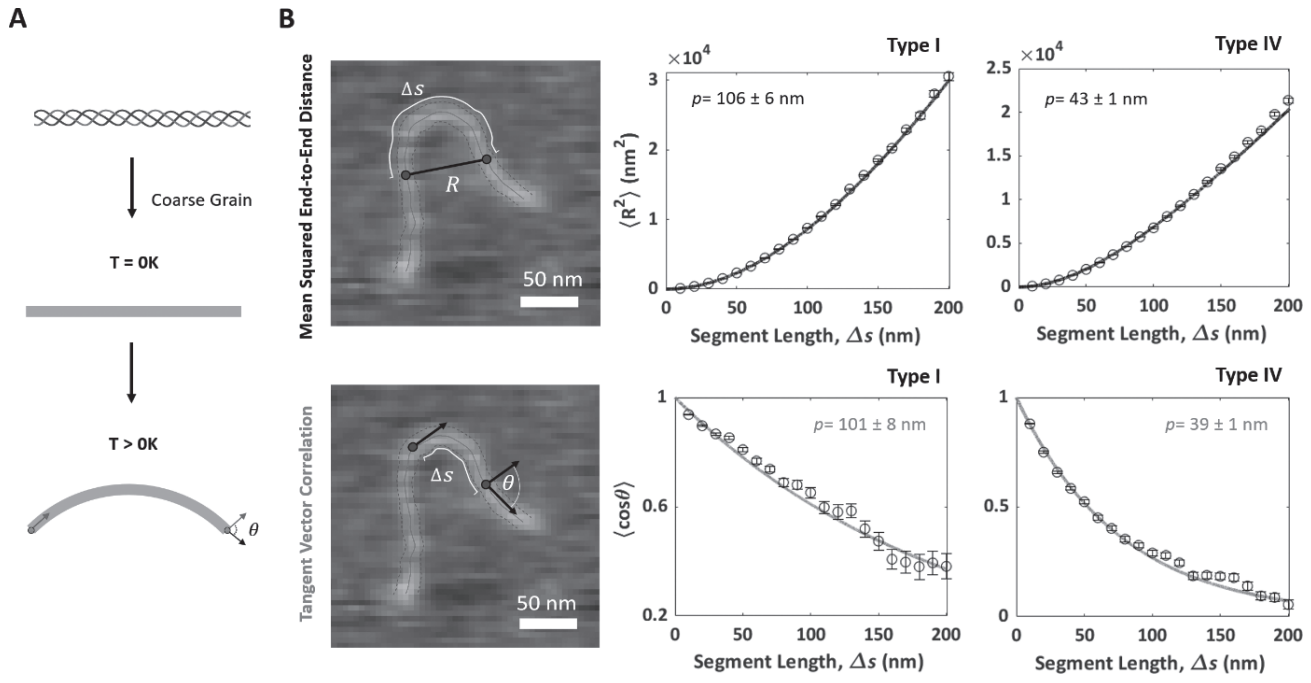

Fig. 2 **Homogeneous analysis of collagen to extract persistence length. A.** The Worm-Like Chain (WLC) model. **B.** The persistence lengths were obtained using $\langle R^2(\Delta s)\rangle$ (black; Eq. 2)) and $\langle \cos\theta(\Delta s)\rangle$ (gray; Eq. 3) analyses, as shown for the collagen data here. These data are fit well by the predictions of the worm-like chain model. The fibrillar collagen type I possess a persistence length of $p \approx 104$ nm [6] ($N = 100$ chains), while collagen type IV exhibits a substantially lower overall persistence length of $p \approx 41$ nm ($N = 280$ chains).

sequence-dependent analysis to determine how these local sequence variations affect flexibility. The algorithm works by calculating an effective persistence length at 1 nm increments along the contour of the chain, using a chain segment of fixed length $\Delta s = 30$ nm centered at each of these positions along the contour, s. Since the angular distribution for a worm-like chain segment is expected to be Gaussian with zero mean and variance equal to the ratio of its segment length Δs and persistence length p, an effective persistence length for the segment centred at position s can be calculated as

$$p^*(s) = \frac{\Delta s}{\sigma_\theta^2(s)}. \qquad (4)$$

As seen in Fig. 3A, the distributions of angles $\theta(\Delta s)$ extracted from our AFM images are well described by a Gaussian distribution with zero mean. Our collagen type IV effective persistence length profile is shown in Fig. 3B, where the profile has been aligned with the expected structure of collagen type IV. We can see that at the end far from the knob, the persistence length significantly drops, in agreement with our visual argument made earlier. This shows that an increase in interruptions is indeed correlated with an enhanced flexibility of the molecule. Also, where there are overlapping interruptions, boxed in black, the persistence length exhibits minima, further supporting the concept of interruptions providing enhanced flexibility to the structure. Lastly, we note that where there are longer stretches of triple-helical segments, we capture the persistence length found for fibrillar collagen type I, which are continuously triple-helical molecules [6].

CONCLUSION

Our results shed light onto the heterogeneity of the collagen type IV molecule and the effects of triple-helical interruptions on its flexibility. By performing statistical analysis, we can gain insight onto the structural consequences of interruptions – a feature that is incompletely understood and lacking a well-defined biological role. The differences seen in the flexibility of fibrillar and network-forming collagen may be manifest by the type of higher-order structures they assemble into, where collagen type I assembles into fibrils which are more rigid structures than the networks formed by collagen type IV.

ACKNOWLEDGEMENTS

This work was supported by a Discovery Grant from NSERC. We are grateful for the assistance provided by our collaborators, Sergey Budko and Billy G. Hudson from Vanderbilt University. We would like to give a special thanks to ChangMin Kim for assistance with AFM instrumentation.

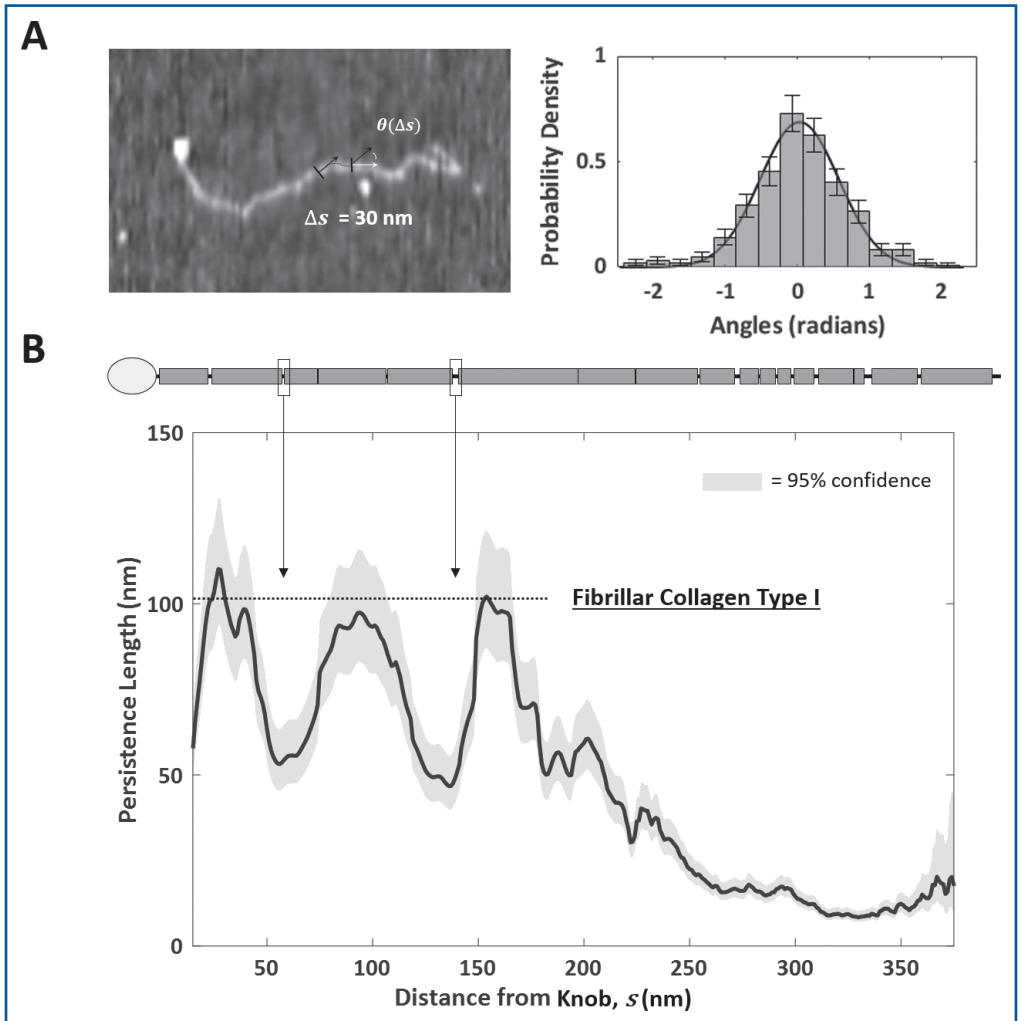

Fig. 3 **Position-dependent persistence length profile of collagen type IV. A.** For segments of length Δs each centered at a distance s along the backbone from the "knob" of each chain, the effective persistence length is determined by the variance of angular changes $\theta(\Delta s)$ among these segments (Eq. 4). **B.** Collagen type IV structural representation (box = triple-helical, gap = overlapping interruption) aligned with the persistence length profile, where the dashed line indicates the persistence length found for the continuously triple-helical fibrillar collagen type I [6], and the arrows correspond to key overlapping interruptions that locally increase flexibility.

REFERENCES

1. A. Sorushanova, *et al*., "The Collagen Suprafamily: From Biosynthesis to Advanced Biomaterial Development", *Adv. Mater.* **31**, 1-39 (2019).
2. M.W.H. Kirkness, K. Lehmann, and N.R. Forde, "Mechanics and Structural Stability of the Collagen Triple Helix", *Curr. Opin. Chem. Biol.* **53**, 98-105 (2019).
3. A.L. Fidler, *et al*., "Collagen IV and basement membrane at the evolutionary dawn of metazoan tissues", *eLife.* **6**, e24176 (2017).
4. A.L. Fidler, *et al*., "The triple helix of collagens – an ancient protein structure that enabled animal multicellularity and tissue evolution", *J. Cell Sci.* **131**, jcs203950 (2018).
5. B.G. Hudson, "The molecular basis of Goodpasture and Alport syndromes: Beacons for the discovery of the collagen IV family", *J. Am. Soc. Nephrol.* **15**, 2514-2527 (2004).
6. N. Rezaei, A. Lyons, and N.R. Forde, "Environmentally controlled curvature of single collagen proteins", *Biophys. J.* **115**, 1457-1469 (2018).