Environmentally Controlled Curvature of Single Collagen Proteins

Nagmeh Rezaei,[1] Aaron Lyons,[1] and Nancy R. Forde[1,*]
[1]Department of Physics, Simon Fraser University, Burnaby, Canada

ABSTRACT   The predominant structural protein in vertebrates is collagen, which plays a key role in extracellular matrix and connective tissue mechanics. Despite its prevalence and physical importance in biology, the mechanical properties of molecular collagen are far from established. The flexibility of its triple helix is unresolved, with descriptions from different experimental techniques ranging from flexible to semirigid. Furthermore, it is unknown how collagen type (homo- versus heterotrimeric) and source (tissue derived versus recombinant) influence flexibility. Using SmarTrace, a chain-tracing algorithm we devised, we performed statistical analysis of collagen conformations collected with atomic force microscopy to determine the protein's mechanical properties. Our results show that types I, II, and III collagens—the key fibrillar varieties—exhibit similar molecular flexibilities. However, collagen conformations are strongly modulated by salt, transitioning from compact to extended as KCl concentration increases in both neutral and acidic pH. Although analysis with a standard worm-like chain model suggests that the persistence length of collagen can attain a wide range of values within the literature range, closer inspection reveals that this modulation of collagen's conformational behavior is not due to changes in flexibility but rather arises from the induction of curvature (either intrinsic or induced by interactions with the mica surface). By modifying standard polymer theory to include innate curvature, we show that collagen behaves as an equilibrated curved worm-like chain in two dimensions. Analysis within the curved worm-like chain model shows that collagen's curvature depends strongly on pH and salt, whereas its persistence length does not. Thus, we find that triple-helical collagen is well described as semiflexible irrespective of source, type, pH, and salt environment. These results demonstrate that collagen is more flexible than its conventional description as a rigid rod, which may have implications for its cellular processing and secretion.

INTRODUCTION

Collagen is the predominant structural protein in vertebrates, for whom it represents more than one quarter of the total protein in our bodies (1,2). It is widely used as a biomaterial and plays a vital physiological role in extracellular matrix and connective tissue mechanics. Collagen can assemble into many different higher-order forms, which fulfill distinct structural and mechanical roles.

Over 28 different types of human collagen have been identified (1,2), of which the most prevalent are fibrillar collagens such as types I, II, and III. These collagens assemble to create highly ordered fibrils, which in turn are the building blocks for the extracellular matrix and for fibers, which act as load- and tension-bearing structures in connective tissues. Not surprisingly, changes in collagen's composition are associated with a wide variety of diseases, including os-

teogenesis imperfecta (3) and Ehlers-Danlos syndrome (4). In addition to genetic mutations in collagen, physiological dysfunction can arise from alterations in post-translational modifications and aging via nonenzymatic glycation and cross-link formation (5). Such chemical changes at the protein level correlate with altered tissue structure and mechanics and pathologically affect human health. Because of the hierarchical nature of collagen structure, identifying the mechanisms by which chemical changes modify tissue mechanics requires understanding how they impact mechanics at the molecular level.

Collagen's molecular structure is a right-handed triple helix (Fig. 1 A), ∼300 nm long and 1–2 nm in diameter (2). This triple helix comprises three left-handed polyproline-II-like helices (α-chains) with a characteristic $(Gly-X-Y)_n$ repeat amino acid sequence. Although frequently proline or hydroxyproline, the X and Y amino acids are variable and provide the sequence diversity that defines individual collagen types. Different collagen types are also distinguished as homotrimeric (three identical α-chains) or heterotrimeric. Although the sequences of the different types of
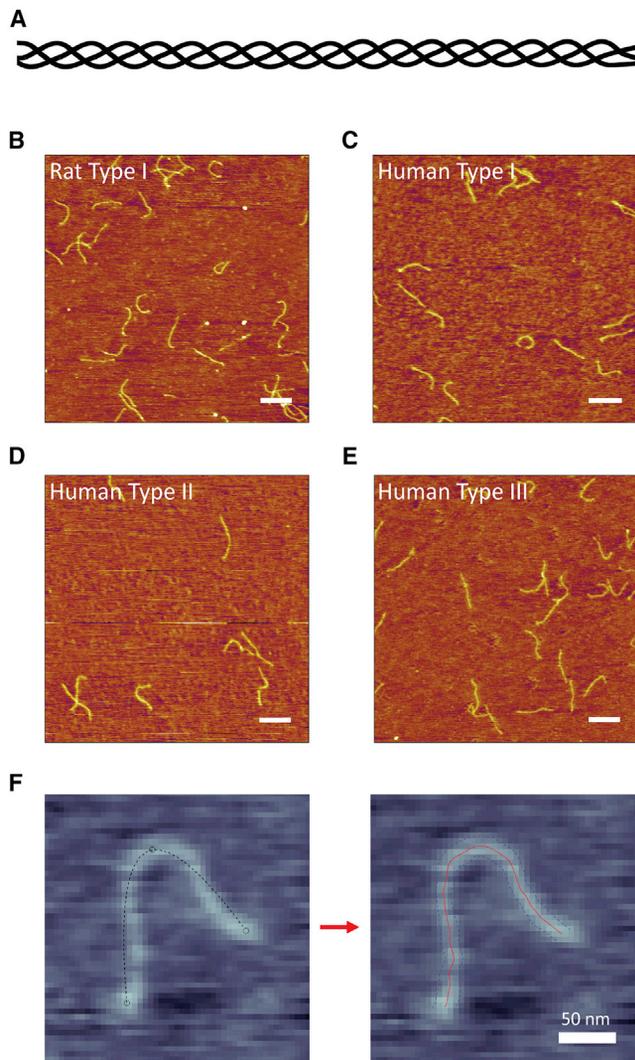
FIGURE 1   Imaging individual collagen proteins. (*A*) A schematic of triple-helical collagen. (*B–E*) Representative AFM images of each of the four collagen samples investigated in this work. For each sample, collagen was deposited from a solution of 100 mM KCl + 1 mM HCl. Scale bars, 250 nm. (*B*) Rat tail-derived type I collagen. (*C*) Recombinant human type I collagen. (*D*) Cartilage-derived human type II collagen. (*E*) Recombinant human type III collagen. (*F*) SmarTrace starts with user input points along a collagen chain (*left image*). Initial splines connecting the three input points (*dashed lines*) do not follow the chain. The SmarTrace algorithm identifies both the center line of the chain (*red*) and its width (*blue*—used for chain tracing but no further analysis), as shown in the right image. To view this figure in color, go online.

collagen are well established, their mechanical properties are not.

One property commonly used to describe the mechanical properties of chain-like biomolecules is the persistence length—the length over which the orientation of the chain remains correlated in the presence of thermal noise. Although the persistence lengths of other biological polymers such as DNA are well established and robust to experimental technique, this is not the case for collagen (Table 1).

Early solution-based studies found that collagen's triple helix should be considered semirigid, exhibiting a persistence length of $p \approx 130$–180 nm (6–8). These values likely led to the conventional description of collagen as a possessing a stiff, rod-like structure (9,10). However, most recent single-molecule experimental and simulation studies contradict this description, finding persistence lengths for collagen as short as $p \approx 10$ nm (11–17). These values imply that collagen is highly flexible, adopting compact, coiled configurations in solution. Other experimental and simulation approaches have determined persistence lengths throughout this range, with another cluster of values around $p \approx 40$–60 nm (14,18–20). It is remarkable that this fundamental parameter of persistence length is so poorly established for a protein of such mechanical importance and ubiquity.

Examination of Table 1 suggests that there is diversity not only in the approaches used to study collagen's flexibility but also in the types of collagen, their sources, and solution conditions. A comparison of different tissue-derived collagen types was performed in (18), which determined the type I heterotrimer to be slightly more rigid than the type III homotrimer. A similar difference between homo- and heterotrimeric collagen sequences was found in recent molecular dynamics simulations (16). Collagen's remodeling by MMP-1, a collagenase that unwinds the triple helix, has been suggested to depend on the source of collagen (e.g., recombinant, from cell culture, or from tissue) (21), implying that post-translational enzymatic and age-related nonenzymatic modifications may alter its mechanics at the molecular level. A recent study found that collagen's flexibility is influenced by its solution environment, with collagen appearing flexible at low salt in acidic conditions and more rigid in high-ionic-strength, neutral pH buffers (13). This result runs counter to the expectations of the behavior of polyelectrolytes like DNA, which are expected to become more flexible as ionic strength is increased (22,23). Understanding collagen's response in different solution conditions is relevant not only to its mechanical role in distinct tissue environments but also to understanding its intracellular compactness as it traverses a pH gradient during cellular processing and secretion (24,25).

In this work, we use atomic force microscopy imaging and characterization to investigate how composition, source, and chemical environment affect the flexibility of collagen's triple helix. By comparing different fibrillar collagens (types I, II, and III; tissue-derived and recombinantly expressed; homo- and heterotrimeric triple helical structures), we find that molecular composition does not significantly impact collagen's overall flexibility. In contrast, we find that both ionic strength and pH independently impact the apparent flexibility of collagen, providing estimates of persistence length that span the range from flexible to semirigid depending on solution environment. Careful consideration of polymer chain statistics shows, however, that treating collagen as

TABLE 1   Literature Estimates of Persistence Length for Fibrillar, Molecular Collagens

Persistence Length, p (nm)	Collagen Type and Source	Solution Conditions	Method	Reference
10	57-residue segment of homotrimeric mouse type I	water, neutral pH	Molecular dynamics	(16)
12	Bovine dermis type I	water, pH ~5	AFM imaging	(13)
13	45-residue segment of rat type I	10 mM NaCl, neutral pH	Molecular dynamics	(17)
11–15	Cell-derived human type I procollagen; recombinant human type II procollagen	>10 mM buffer, neutral pH	Optical tweezers stretching	(11,12)
16	30-residue homotrimeric peptide	water, neutral pH	Steered molecular dynamics	(15)
22	57-residue segment of heterotrimeric mouse type I	water, neutral pH	Molecular dynamics	(16)
15–65	Recombinant human type II procollagen	>10 mM buffer, neutral pH	Optical tweezers stretching	(14)
40	Fetal bovine type III	50 mM acetic acid + equal volume glycerol	Electron microscopy	(18)
51	60-nm segment of human type I collagen	water, neutral pH	Coarse-grained molecular dynamics	(20)
57	Calf dermis type I	50 mM acetic acid + equal volume glycerol	Electron microscopy	(18)
130	Rat skin type I	>10 mM buffer, pH ~5	Viscometry	(6)
135–165	Bovine dermis type I	>10 mM buffer, neutral pH	AFM imaging	(13)
161	Rat skin type I	0.3 M acetate + 6 mM NaCl, pH 4	Rheology	(7,8)
160–165	Bovine dermis type I	1 mM HCl	Dynamic light scattering	(19)
167	Rat skin type I	>10 mM buffer, neutral pH	Rheology	(8)

a standard worm-like chain does not describe its properties in most of these chemical environments. Instead, we show that a curved worm-like chain model that includes inherent molecular curvature is a far more appropriate descriptor of the observed conformations of collagen on mica. With this model, collagen's bending flexibility depends much less on solution conditions; instead, salt concentration and pH modulate its global curvature. We find that curvature does not depend strongly on the type or source of collagen, nor does the persistence length. To the best of our knowledge, these results constitute the first experimental analysis of curved worm-like chains and provide a new explanation for the conflicting reports on the flexibility of triple-helical collagen.

MATERIALS AND METHODS

Collagen sources

Recombinant human type I and recombinant human type III collagen were expressed in yeast and obtained from FibroGen (San Francisco, CA) (generous gifts of Alexander Dunn, Stanford University), rat tail tendon-derived type I collagen (Cultrex 3440-100-01) was purchased from R&D Systems (Oakville, Canada), and human cartilage-derived type II collagen (CC052) was purchased from EMD Millipore (Burlington, MA). All stocks are of pepsin-treated collagen and are have a concentration between 2 and 5 mg/mL in 20 mM acetic acid.

Sample preparation

The desired solution conditions were obtained by solution exchange using Millipore Amicon Ultra-0.5 spin filters (50 kDa, UFC505096), then dilution to ~1 μg/mL collagen before deposition. 50 μL was deposited onto freshly

cleaved mica (Highest Grade V1 AFM Mica Discs, 10 mm; Ted Pella, Redding, CA) for 20 s. After deposition, samples were rinsed five times with 1 mL ultrapure water to remove unbound proteins, and the mica was dried under a flow of filtered compressed air. It is important to note that all collagen molecules were imaged in these dry conditions. Thus, solution conditions quoted refer to the condition under which collagen was deposited onto mica, at room temperature.

Atomic force microscopy imaging

Images of collagen adsorbed to mica were collected with an Asylum Research MFP-3D atomic force microscope (Asylum Research, Santa Barbara, CA) using tapping mode in air. AFM tips with a 325 kHz resonance frequency and 40 N/m force constant (MikroMasch, HQ:NSC15/AL BS) were used for image collection and were changed as necessary to preserve image quality.

Chain tracing

The SmarTrace algorithm used to trace collagen chains from AFM images was developed in MATLAB (26) and uses a graphical user interface adapted from (27). It is available from the authors upon request. A detailed description of the SmarTrace workflow and validation can be found in the Supporting Materials and Methods. Briefly, the algorithm uses cross-correlation analysis of a template cross section with the imaged chain to identify its center line and width. Continuity constraints on the local directionality and width enable the tracing of chains in noisy environments.

Data analysis

Traced chains were sampled using bootstrapping to extract statistics of the chain for flexibility determination. Following the method of Faas et al. (28), each chain was randomly divided into nonoverlapping segments of lengths drawn from a set of input values (here, $s = 10, 20, 30, ..., 200$ nm). The

maximal segment length does not need to be the same as the chain contour length, so it allows the use of partially traced chains. This is particularly useful when ends of a molecule are not clear or chains intersect, as it allows subsections of the chain to be included in the analysis. $\langle R^2(s) \rangle$ and $\langle \cos\theta(s) \rangle$ were determined from all segments of each length s.

Resultant $\langle R^2(s) \rangle$ and $\langle \cos\theta(s) \rangle$ were fitted to equations derived from the inextensible two-dimensional worm-like chain (WLC) model with (Eqs. 3 and 4) or without (Eqs. 1 and 2) intrinsic curvature. Derivations for Eqs. 3 and 4 are presented in the Supporting Materials and Methods.

The kurtosis of each angular distribution was calculated at each segment length s as $n[\sum_i (\theta_i - \langle\theta\rangle)^4]/[\sum_i (\theta_i - \langle\theta\rangle)^2]^2$, for a distribution across n values of θ_i. The standard error of the kurtosis was determined using Eq. S19.

Values for p and κ_o presented in Table 2 are an average of the results from independent fits to $\langle R^2(s) \rangle$ and to $\langle \cos\theta(s) \rangle$. Reported errors, Δ, represent the propagated error of the 95% confidence intervals of the respective fit parameters or half of the difference between the $\langle R^2(s) \rangle$ and $\langle \cos\theta(s) \rangle$ fit parameters, whichever is larger. A full list of fitting parameters from all models and samples is provided in Table S1. The total length of collagen chains traced in each sample condition is also included in Table S1.

Model selection used a likelihood-based approach, the Bayesian Information Criterion (BIC) (29). Details are provided in the Supporting Materials and Methods, with results provided in Table S2.

Validation tests

The performance of the SmarTrace algorithm and data analysis code was evaluated by tests on simulated polymers and on images of DNA. Details are provided in the Supporting Materials and Methods.

RESULTS AND DISCUSSION

Atomic force microscopy (AFM) is a well-established tool for imaging the conformations of flexible biopolymers (30). We have used the technique to image collagens of different types and sources deposited from a range of solution conditions and have characterized how these parameters influence the flexibility of collagen's triple helix.

Collagen type and source

To investigate how collagen's flexibility depends on its source, we compared four different samples of collagen. These represent three types of fibrillar collagens (types I, II, and III) of either rat (type I) or human (all three) genetic origin. Although all of these samples are capable of self-assembly into highly ordered fibrils, they were imaged under sufficiently dilute conditions such that assembly did not occur. Type I collagen is a heterotrimer ($\alpha_1(I)_2\alpha_2(I)_1$), whereas types II and III are each homotrimeric ($\alpha_1(II/III)_3$). Additionally, to explore how collagen's source influences its flexibility, our samples encompass both tissue-derived (rat type I and human type II) and cell-derived (human types I and III) sources. The latter were recombinantly expressed in yeast and possess only prolyl hydroxylation as a post-translational modification (31). This contrasts with mammalian-produced collagens, which have extensive post-translational modifications including hydroxylation of prolines and lysines as well as O-glycosylation of hydroxylysines (14,24).

Representative AFM images of each of the four collagen samples are shown in Fig. 1, *B–E*. Each image was obtained by depositing collagen from a room-temperature solution of 100 mM KCl + 1 mM HCl onto freshly cleaved mica, then rinsing with water and drying before imaging. For all four collagen samples, the chains appear semiflexible, with contour lengths of ~300 nm, as expected (32). There are no obvious qualitative differences in flexibility among these samples.

To quantify the flexibility of each collagen type, the imaged chains were traced to provide backbone contours for conformational analysis. Existing chain-tracing algorithms proved problematic for many of our images, in some cases because of the background levels of noise (28) or in

TABLE 2 Summary of Fitting Parameters Determined in This Work

Collagen Type	Solvent Condition	Standard Model Fits		Curved Model Fits			
		p (nm)	Δp (nm)	p (nm)	Δp (nm)	κ_o (nm^{-1})	$\Delta\kappa_o$ (nm^{-1})
Rat I	water	59	17	102	21	0.015	0.003
Rat I	100 μM KCl	40	8	57	14	0.014	0.003
Rat I	1 mM KCl	70	15	105	21	0.011	0.001
Rat I	10 mM KCl	91	9	89	11	0.000	N/A
Rat I	100 mM KCl	120	22	120	46	0.000	N/A
Rat I	1 mM HCl	39	8	57	11	0.016	0.002
Rat I	10 mM KCl + 1 mM HCl	75	15	117	26	0.010	0.001
Rat I	100 mM KCl + 1 mM HCl	103	10	101	18	0.000	N/A
Human I	100 mM KCl + 1 mM HCl	86	8	88	14	0.000	N/A
Human II	100 mM KCl + 1 mM HCl	98	7	101	10	0.002	N/A
Human III	100 mM KCl + 1 mM HCl	84	7	83	11	0.001	N/A
Rat I	20 mM acetic acid	33	7	49	8	0.018	0.002
Human I	20 mM acetic acid	38	9	59	10	0.017	0.001
Human II	20 mM acetic acid	57	18	97	17	0.017	0.002
Human III	20 mM acetic acid	44	11	69	14	0.017	0.002

Fit parameters are averages of the values from independent $\langle R^2(s) \rangle$ and $\langle \cos\theta(s) \rangle$ fits of the data; errors, Δ, represent the error in these estimates. N/A indicates uncertainties that are unphysically large. The total contour lengths of collagen traced in each condition is provided along with a comprehensive list of fitting parameters and χ^2 values in Table S1.

others from the sensitivity of the algorithm to the locations of user-input starting points (27). Hence, we developed a new chain-tracing program, dubbed SmarTrace. The algorithm incorporates pattern matching to identify the best centerline of a chain and refines an initial guess at this centerline by using direction and width continuity constraints. Required user input is minimal: backbone contours are identified from only a few clicks near the chain (see Fig. 1 F). Details of SmarTrace's methodology and workflow are provided in the Supporting Materials and Methods.

From traced chains, we implemented tools of polymer physics to analyze the statistical properties of chain conformations and determine persistence lengths (33). To perform this analysis, chains were segmented randomly into nonoverlapping pieces of different contour lengths (e.g., $s = 10, 20, \ldots, 200$ nm). This approach allows partially traced chains to be included in the analysis (28). For each segment, we calculated its squared end-to-end distance $R^2(s)$ and the change in orientation between its starting and ending tangent vectors $\hat{t}(s) \cdot \hat{t}(0) = \cos \theta(s)$; these quantities were then averaged over all segments of length s within the population. Length-dependent trends in mean-square end-to-end distance, $\langle R^2(s) \rangle$, and tangent vector correlation, $\langle \cos \theta(s) \rangle$, were compared with the predictions of the WLC model for polymers equilibrated in two dimensions:

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

and

$$\langle \cos \theta(s) \rangle = e^{-\frac{s}{2p}}. \qquad (2)$$

The validity of the SmarTrace chain-tracing algorithm and analysis approaches was established by testing their performance on AFM images of DNA and on simulated chains (see Supporting Materials and Methods; Figs. S1–S3).

We first investigated whether collagen molecules are equilibrated on the mica and are well described by the WLC model. For type I collagen from rat deposited from 100 mM KCl + 1 mM HCl, both mean-square end-to-end distance $\langle R^2(s) \rangle$ and tangent vector correlation $\langle \cos \theta(s) \rangle$ are well described by the WLC model (Fig. 2, A and B). The agreement between persistence lengths from the fits to Eqs. 1 and 2, $p = 106 \pm 6$ nm and $p = 101 \pm 8$ nm, respectively, further suggest that collagen deposited from this solution condition behaves as an equilibrated two-dimensional WLC. This assumption is corroborated by Gaussian distributions of bending angles (Fig. 2 C). The kurtosis of the bending angle distribution is $\sim$3 (the value characteristic of a normal distribution) at all contour lengths (Fig. 2 D). These measures all indicate that the conformations of collagen on mica represent a sample equilibrated in two dimensions and that rat type I collagen has a persistence length of $\sim$100 nm.

The persistence lengths were similarly determined for the other collagen samples (Fig. 2 E). The data and fits are shown in Fig. S4, and persistence lengths are presented in Fig. 2 F. Table 2 provides a summary of the fit parameters, representing an average of the $\langle R^2(s) \rangle$- and $\langle \cos \theta(s) \rangle$-derived results; the complete summary of parameters from each fit is given in Table S1, along with reduced χ^2 values used to assess goodness of fit. All four collagen samples exhibit similar persistence lengths, within the range of 84–103 nm in these solution conditions. We do not observe significant differences in flexibility for homotrimeric (II, III) collagen versus heterotrimeric (I) collagen, in contrast with previous findings (18,34). Tissue-derived collagens (rat I, human II) appear to be slightly more rigid than the yeast-expressed recombinant (human I, III) collagens, which correlates with their increase in post-translational modifications. However, these differences are small in the context of the wide range of values in the literature (Table 1). Thus, differences in type, source, and extent of post-translational modifications do not appear to cause substantial differences in molecular flexibility.

Influences of salt and pH

Given a previous report of significantly different collagen flexibility in buffers with high salt concentration and neutral pH, compared with low-ionic-strength, acidic solutions (13), we next investigated how pH and salt concentration independently impact the flexibility of collagen. Rat type I collagen was deposited onto mica from a range of different solution conditions in a controlled ionic environment. In one set of experiments, solutions contained only varying concentrations of potassium chloride, from 0 mM (water) to 100 mM KCl. In a second set of experiments, solutions again contained varying concentrations of KCl but were acidified to pH $\approx$ 3 by 1 mM HCl.

Images of collagen at different KCl concentrations and pH are shown in Fig. 3 A. It is immediately apparent that salt affects the conformations of collagen: the protein transitions from more compact structures at low ionic strength to much more extended conformations at high ionic strength. A similar trend is seen in the acidic conditions. Conformations were analyzed to obtain mean-square end-to-end distances and tangent correlations as a function of contour length and were fitted to Eqs. 1 and 2, respectively. This analysis finds persistence length to increase significantly with ionic strength at both neutral and acidic pH (Fig. 3 B; Table 2). Our results corroborate and elaborate on the previous findings of increased flexibility at low pH and salt compared with neutral pH and high salt (13).

The dependence of persistence length on salt concentration is striking. In neutral solutions, the determined persistence length increases approximately threefold as the KCl concentration is increased to 100 mM. In acidic solutions, persistence lengths are shorter but still increase, more than doubling as the concentration of KCl increases from 1 to
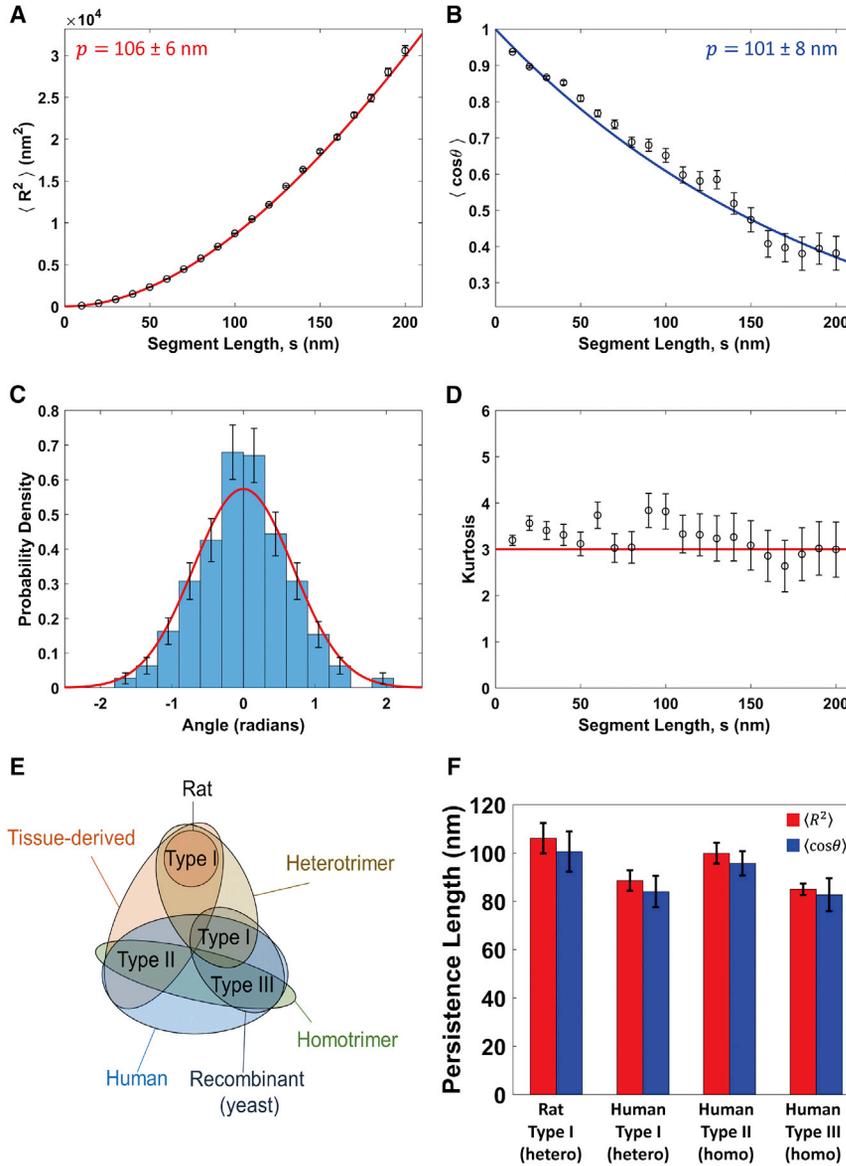
FIGURE 2 Determination of collagen's persistence length using the standard worm-like chain (WLC) model. (A–D) Analysis for rat type I collagen deposited from a solution of 100 mM KCl + 1 mM HCl. (A) Mean-square end-to-end distance as a function of segment length, $\langle R^2(s) \rangle$. Points represent mean values determined at 10 nm segment-length intervals, with error bars representing standard errors of the mean. The red line is a fit with Eq. 1, yielding a persistence length of $p = 106 \pm 6$ nm (error represents 95% confidence interval of the fit). (B) Tangent vector correlation as a function of segment length, $\langle \cos \theta(s) \rangle$. Points and errors are similarly represented as in (A). The blue line is a fit with Eq. 2, yielding $p = 101 \pm 8$ nm. Fits to the WLC model in (A) and (B) capture the trends in the data and yield comparable persistence lengths. (C) An angular histogram for $s = 50$ nm segment lengths (blue bars). Error bars represent $\sqrt{N}$ counting error. This distribution agrees well with that expected for a WLC of persistence length $p = 103$ nm (the average of the results from (A) to (B)), which is a normal distribution with a mean of zero and a variance of s/p (red line). (D) Kurtosis of the angular distributions extracted from the traced collagens at different segment lengths is shown. Error bars represent the standard error in the kurtosis (Eq. S19). The kurtosis of the angular distributions is close to three (red line) for all segment lengths, indicating that the collagens behave as equilibrated Gaussian chains on the mica surface. (E) A Venn diagram illustrating the similarities and differences between the different collagen samples tested, including trimeric identity and source. (F) Persistence length for each collagen sample deposited from a solution of 100 mM KCl + 1 mM HCl, obtained using $\langle R^2(s) \rangle$ and $\langle \cos \theta(s) \rangle$ analyses. The similarity among these samples suggests that collagen type and source have little impact on the mechanical properties of collagen at the molecular level. To view this figure in color, go online.

100 mM. Intriguingly, the range of persistence lengths exhibited here spans a significant portion of values reported and unreconciled in the literature (Table 1).

The trend of increasing stiffness with increasing ionic strength is unexpected, as it is opposite to predictions for polyelectrolytes and to observations for biopolymers such as DNA (22,23). However, examination of the data and fits in some of the solution conditions, particularly at lower salt concentrations, reveals that the experimental data are not well described by the standard WLC model (e.g., Fig. 4; see also Figs. S5 and S6). For $\langle R^2(s) \rangle$, the agreement between model and data is reasonable at shorter segment lengths (e.g., $s < 100$ nm), but longer segments exhibit shorter-than-expected end-to-end distances. The disagreement with the standard WLC model is even more apparent when considering the $\langle \cos \theta(s) \rangle$ behavior. There appears to be an oscillatory modulation of the tangent vector correlation whereby chains maintain directionality at small segment lengths, reorient substantially at intermediate lengths, and in many cases, exhibit anticorrelated tangent vectors at longer contour lengths. These are not features of a well-equilibrated WLC. Instead, the oscillatory character of the tangent vector decorrelation suggests that curvature may play a role in the observed conformations of collagen.

Curved WLC

To describe an intrinsically curved WLC (cWLC), we assume that the lowest-energy conformation of the chain is curved (bent) rather than straight. The extent of angular fluctuations about this curved state is determined by the
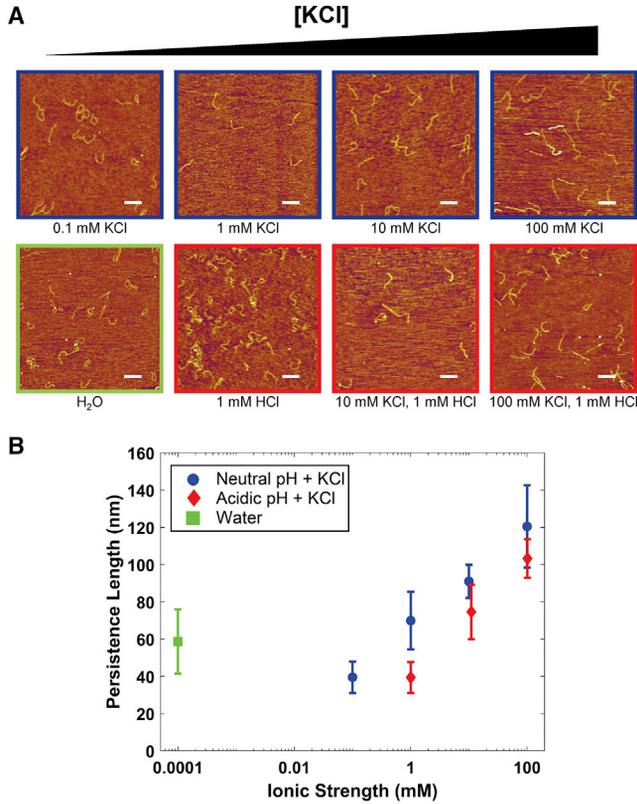
A

[KCl]

0.1 mM KCl   1 mM KCl   10 mM KCl   100 mM KCl

H₂O   1 mM HCl   10 mM KCl, 1 mM HCl   100 mM KCl, 1 mM HCl

B

FIGURE 3 Effects of ionic strength on persistence length. (*A*) AFM images of rat type I collagen deposited from solutions covering a range of KCl concentration, for both neutral and acidic (~3) pH. Scale bars, 250 nm. As salt concentration increases, the collagen chains appear to straighten. (*B*) Persistence lengths in these conditions. The reported values are the average of p extracted from fits to $\langle R^2(s)\rangle$ and to $\langle \cos\theta(s)\rangle$, and errors are the larger of the propagated error in this mean or half the separation between the two values of p. For the purposes of graphical representation, water was assumed to have an ionic strength of 10^{-7} M. Increasing the ionic strength causes a large increase in the apparent persistence length, an effect that is reduced slightly in the acidic condition for the same ionic strengths. To view this figure in color, go online.

angular bending potential and the thermal energy of the system, which relate to persistence length as for the standard WLC. From this cWLC model, equations for the mean-square end-to-end distance and mean tangent vector correlation can be derived (see Supporting Materials and Methods):

$$\langle R^2(s)\rangle_c = \frac{4sp}{\left(1 + 4\kappa_o^2 p^2\right)^2}\left\{1 - \frac{2p}{s}\left(1 - 4\kappa_o^2 p^2\right)\right.$$
$$\times \left[1 - \cos(\kappa_o s)e^{-\frac{s}{2p}}\right] \qquad (3)$$
$$\left. + \frac{4\kappa_o p^2}{s}\left[\kappa_o s - 2\sin(\kappa_o s)e^{-\frac{s}{2p}}\right]\right\}$$

and

$$\langle \cos\theta(s)_c\rangle = \cos(\kappa_o s)e^{-\frac{s}{2p}}. \qquad (4)$$

Here, κ_o represents the inherent curvature of the chain, which is the inverse of its inherent radius of curvature: $\kappa_o = R_o^{-1}$. Eqs. 3 and 4 describe the expected behavior of an intrinsically curved, inextensible WLC equilibrated in two dimensions and are two-dimensional versions of three-dimensional results derived previously (35). Of note, the cWLC model is distinct from approaches that assume a random local curvature (28): in the cWLC model, the lowest-energy conformation of the chain is globally curved, i.e., bends always in the same direction.

Fig. 4, *A* and *B* show fits of the cWLC model to the $\langle R^2(s)\rangle$ and $\langle \cos\theta(s)\rangle$ data from rat type I collagen deposited from 1 mM HCl. The cWLC captures trends in both $\langle R^2(s)\rangle$ and $\langle \cos\theta(s)\rangle$ significantly better than the standard model. Notably, the cWLC model captures the longer-length anticorrelation of the tangent vectors ($\langle \cos\theta < 0\rangle$), which under the standard WLC model is unphysical. Additionally, the cWLC fits provide a distinct interpretation of the compact structures observed at low ionic strength (Fig. 3): rather than being more flexible at lower salt concentration, collagen is instead more curved.

Salt-dependent trends in persistence length and curvature obtained from the cWLC model are shown in Fig. 4, *C* and *D*. In contrast to the monotonic increase in persistence length with increasing KCl concentrations found when using the standard WLC model (Fig. 3), the results from the cWLC fits indicate that persistence length varies only modestly over the range of concentrations used here (Fig. 4 *C*). Instead, the curvature depends strongly on ionic strength, decreasing from $\kappa_o = 0.02$ nm^{-1} to 0 (i.e., straightening) as the concentration of KCl is increased (Fig. 4 *D*). This effect is shown schematically in Fig. 4 *E*.

Statistical considerations demonstrate that the cWLC is a better model than the WLC for collagen in almost all of the tested solution conditions. In lower-concentration KCl conditions, χ^2 is significantly lower when the data are fitted with the cWLC model than with the WLC model (Table S1). In contrast, the WLC model produces a lower χ^2 than the cWLC in only three conditions (10 mM KCl, 100 mM KCl, and 1 mM HCl + 100 mM KCl). In these three solution conditions, we find that fitting with the cWLC model gives the same persistence length and zero curvature, i.e., the cWLC model reduces to the same results as the standard WLC model. We further assessed these two models using the BIC, which uses likelihood analysis to assess the probability that one model is a better descriptor of the data than another. The BIC incorporates a stronger penalty for additional fit parameters than does χ^2 analysis (Eq. S29), thereby challenging more strongly the ability of a model with additional fit parameter(s) to describe the data. Even so, the BIC also indicates a strong preference for the cWLC model in the lower-salt conditions (Table S2).
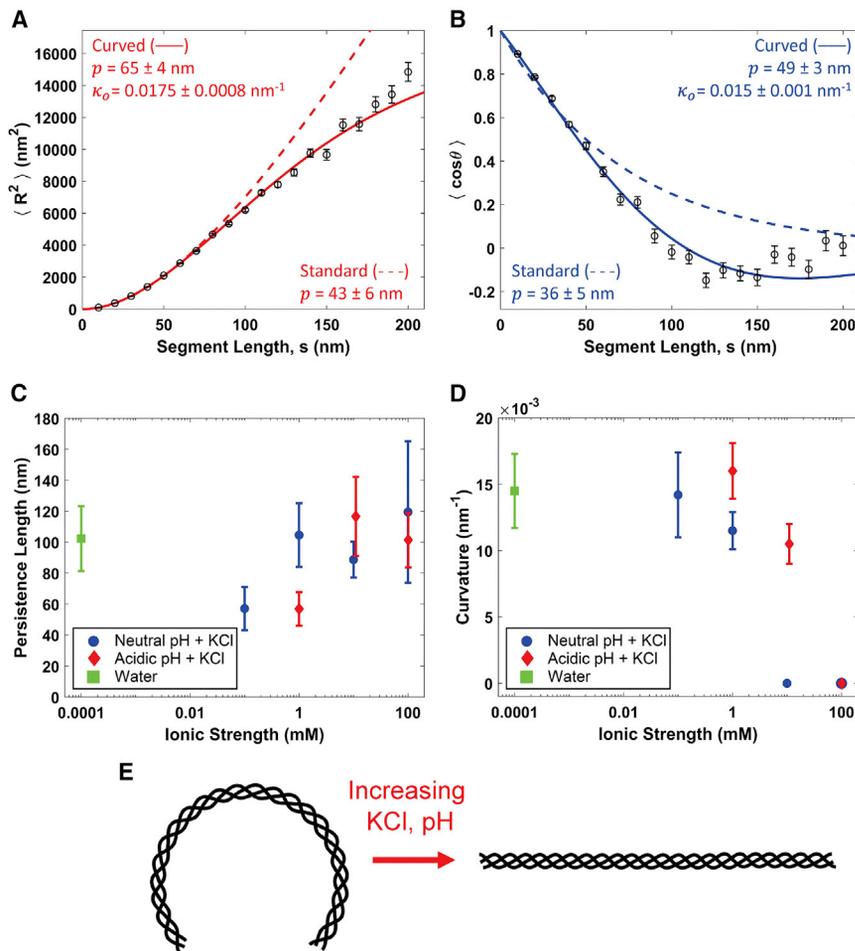
FIGURE 4 At low ionic strength, collagen is better described as a cWLC. (A) Fits of the mean-square end-to-end distance using both the standard (Eq. 1, *dashed red line*) and curved (Eq. 3, *solid red line*) WLC models to data extracted from rat tail type I collagen deposited from 1 mM HCl. (B) Corresponding fits of the standard (Eq. 2, *dashed blue line*) and curved (Eq. 4, *solid blue line*) WLC tangent vector correlation functions. The data are more accurately described by the curved model equations, as shown statistically by a significant improvement in reduced χ^2 values (see text). The standard WLC underestimates the persistence length of the collagens in this condition, presumably interpreting the induced curvature as additional fluctuation. (C) Persistence lengths for the different cosolute conditions presented in Fig. 3 Rather than the trend with ionic strength seen with the standard WLC fits, there is a less significant and less obviously monotonic variation in persistence length as a function of ionic strength. Instead, the observed systematic conformational changes are attributed to a variation in innate curvature, which drops as the ionic strength of the solution is increased (D). For the purposes of graphical representation, water was assumed to have an ionic strength of 10^{-7} M. Error bars on $\langle R^2(s) \rangle$ and $\langle \cos\theta(s) \rangle$ values are as described for Fig. 2, and error bars on p and κ_o are obtained as described for Fig. 3. (E) A schematic illustrating the transition of collagen from a molecule with a curved backbone at low ionic strength and pH ~3 (*left*) to one with a straight backbone at higher ionic strength and neutral pH (*right*). The persistence length characterizes fluctuations about this lowest-energy conformation. To view this figure in color, go online.

Origins of collagen curvature

To determine whether the induction of curvature is ion specific, images of rat type I collagen deposited from 20 mM acetic acid were collected and analyzed with the standard and cWLC models (Fig. 5, *A–D*). Conformations of collagen deposited from acetic acid are better described by the cWLC model, consistent with observations at low ionic strengths of KCl. Again, naïve application of the standard WLC model results in persistence lengths smaller than those obtained from cWLC fits. To test for equilibration of the samples, we considered the distribution of bend angles. Because of chain curvature, these distributions should not be represented by a single Gaussian centered at $\theta(s) = 0$. Rather, the distributions are expected to be described by two Gaussians, centered symmetrically about zero at an angle $\theta(s) = \kappa_o s$ and with standard deviation $\sigma(s) = \sqrt{s/p}$ (Eq. S20). The experimental angular distribution is indeed well described by this bimodal function (Fig. 5 *D*). The agreement between predicted and measured angular distributions is strong evidence of collagen equilibration on the mica surface in this low-salt condition. The fit parameters in acetic acid ($p = 49$ nm, $\kappa_o = 0.018$ nm^{-1})

are similar to those from the 1 mM HCl solution (Table 2), which has comparable ionic strength and pH. Thus, at low ionic strength, there does not seem to be a strong effect of chloride versus acetate anions on collagen's mechanical properties.

It is possible that the observed curvature is an effect specific to heterotrimeric collagen. Because type I collagen has one α-chain which is different than the other two, a larger or smaller propensity for surface interactions with this chain could produce a surface-induced curvature to minimize its energy (36). To address this possibility, types I, II, and III human collagens were also imaged in 20 mM acetic acid and analyzed with the cWLC model (Fig. S7). The results of this analysis are shown in Fig. 5, *E* and *F*. Although the different types vary somewhat in their persistence lengths, their curvatures are nearly identical under these solution conditions, regardless of their trimeric identity. Thus, surface interactions with a unique chain in heterotrimeric collagen are not responsible for the observed curvature.

The curvature observed at low ionic strengths may nonetheless result from molecular interactions with the surface. Intrinsically straight, chiral semiflexible chains may adopt curved conformations at an interface (36). The extent to
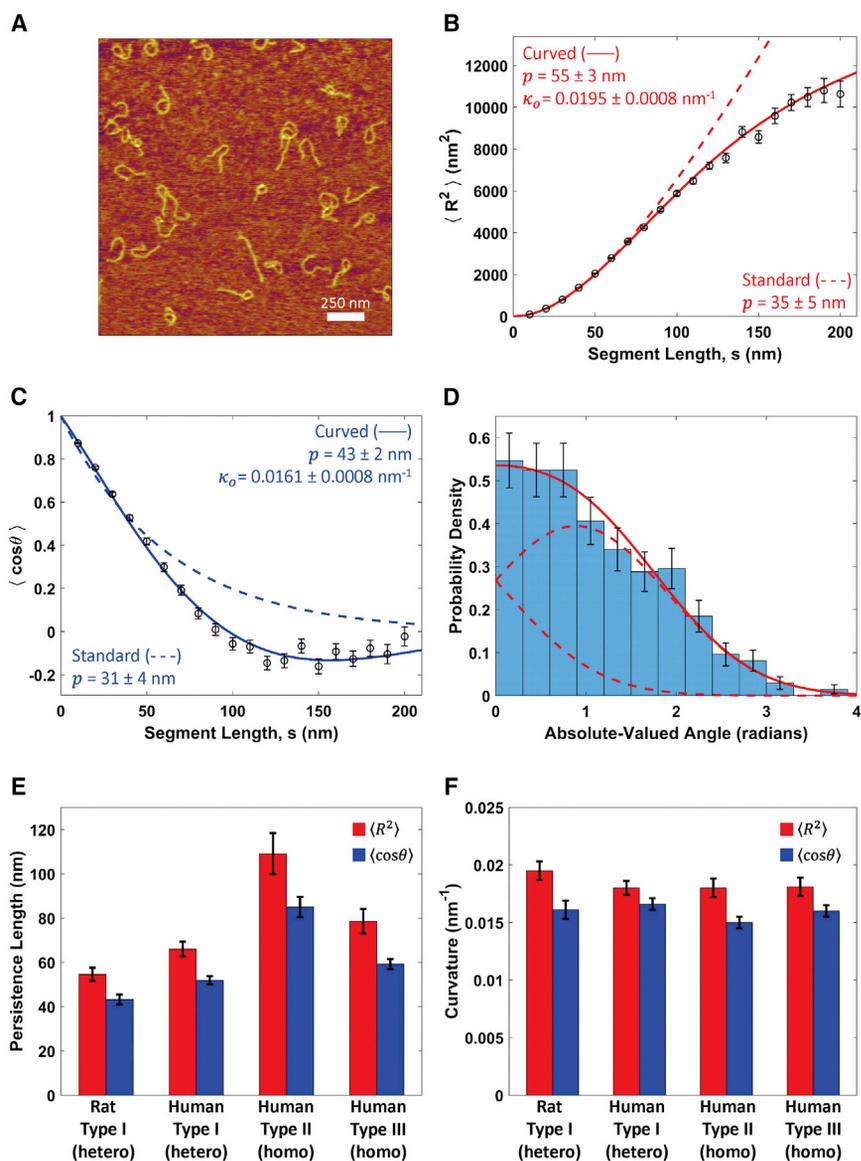
FIGURE 5 Behavior of different collagen types in acetic acid. (*A*) A representative image of rat tail type I collagen deposited from 20 mM acetic acid. (*B*) Standard (Eq. 1, *dashed red line*) and curved (Eq. 3, *solid red line*) WLC model are shown fitted to the mean-square end-to-end distance, $\langle R^2(s) \rangle$. (*C*) Standard (Eq. 2, *dashed blue line*) and curved (Eq. 4, *solid blue line*) WLC models are shown fitted to the tangent vector correlation, $\langle \cos\theta(s) \rangle$. Data in both (*A*) and (*B*) are described more accurately by the cWLC model, as seen also by the reduced χ^2 values (Table S1) and the BIC (Table S2). (*D*) The absolute-valued angular distribution of the same collagen molecules at a segment length of $s = 50$ nm. The red line represents the expected distribution (Eq. S3) for a cWLC using a persistence length $p = 51$ nm and intrinsic curvature $\kappa_o = 0.0182$ nm^{-1}, the average values of each parameter extracted from the fits shown in (*B*) and (*C*). The dashed red lines are the two Gaussian components that make up this distribution, as discussed in the Supporting Materials and Methods. (*E* and *F*) Persistence lengths and curvatures, respectively, of different collagen types deposited from 20 mM acetic acid. As with the different collagen types deposited from 100 mM KCl + 1 mM HCl (Fig. 2 *F*), there is only modest variation among the different samples, demonstrating that, also at lower ionic strength and different cosolute conditions, only minor mechanical differences exist between different collagen types and sources. To view this figure in color, go online.

which the chain curves at an interface will be determined by a competition between the free energy of surface-molecular interactions and the torsional twist energy of the chain. Specifically, curvature in this model requires a chiral organization of sites on the chain that interact differentially at the interface compared with the rest of the chain sites. In the current context, these could be the side chains of the three individual α-chains of collagen, which are arranged in a right-handed superhelical structure about the central axis of the collagen molecule (2). If these sites interact strongly with the mica surface, the free energy gained by adhering to the surface could outweigh the energetic cost of altering the twist (local helical pitch) of the triple helix. In this scenario, the observed trends in curvature with salt concentration (Fig. 4 *F*) imply a salt dependence of collagen's torsional stiffness and/or its interactions with the surface. Salt-depen-

dent interactions with mica are well established for other biopolymers (*e.g* (33)) and could be expected to contribute here (37). In particular, K$^+$ affects the interactions of self-assembling collagen with a mica surface (37–39).

Collagen's stability in solution is also influenced by salt and pH. Collagen has a lower thermal stability at low ionic strength and acidic pH than at high ionic strength and neutral pH (40), and its thermal stability and helicity decrease at lower concentrations of KCl (41). The conditions found to destabilize collagen in solution are the same as those promoting increased curvature in our measurements. It is thus possible that this destabilization of the triple helix reduces the torsional stiffness of collagen, enabling it more easily to adopt curved conformations on mica.

In addition to the ionic-strength dependence of curvature, our images and conformational analysis demonstrate a

dependence on pH: collagen appears more curved in acidic conditions (Fig. 4 F). At a pH of 3, collagen carries a net positive charge. At neutral pH, collagen instead has a roughly equal balance of positively and negatively charged residues. The increased curvature at acidic pH could result from stronger electrostatic affinity to the negatively charged mica surface (37). Alternatively, electrostatic repulsion between α-chains could destabilize the triple helix and reduce its torsional stiffness, particularly at low ionic strength when shielding from cosolute anions is weak (at 1 mM ionic strength, the Debye length $\approx$ 10 nm). At neutral pH, the potential for salt bridge formation could strengthen attractions between α-chains (41,42), thereby increasing the torsional stiffness of the triple helix.

It is also possible that collagen itself is intrinsically bent, even in solution (43). Certainly within the fibrillar superstructure, collagen molecules display conformations that are locally bent (44). Studying how salt influences conformations in solution would be key to discriminating which interactions are surface specific in our study. A change in torsional stiffness and/or alteration of preferred twist angle at lower ionic strength would be expected to alter the average shape of collagen in solution. Light scattering or other solution-based techniques could be used to study how collagen's compactness in solution depends on salt concentration and thereby determine how the triple helix itself responds to changes in ionic strength.

Comparison with other estimates of collagen flexibility

Taken together, our results indicate that collagen should be regarded as a semiflexible polymer at room temperature with a persistence length in the range of 50–120 nm. This is more flexible than solution-based estimates yet more rigid than found in almost all previous single-molecule approaches (Table 1). Inherent biases and assumptions built into data interpretation within each methodology may contribute to the range of results reported in the literature. Here, we describe possible reasons for the discrepancies.

With AFM imaging, one must play particular heed to molecule-surface interactions and to the possibility of kinetically trapping polymers in nonequilibrium configurations on the surface. In our measurements, the observed conformations appear well equilibrated in two dimensions (Figs. 2, C and D and 5 D). Furthermore, although electrostatic interactions with mica may play a role in the apparent curvature of collagen on the surface, we have found that the persistence length is not strongly affected.

In single-molecule stretching experiments, one source of bias is the underestimation of persistence length when the contour length $L < 20p$ (45), as is the case for collagen. Any structural changes in the triple helix that occur during stretching may also modify the bending energy: such changes have been inferred from enzymatic cleavage assays

of collagen under force (21,46–48) and may contribute to softening of collagen's force-extension response (14).

Many of the simulations of collagen have used steered molecular dynamics, which may not investigate its equilibrium flexibility. Additionally, the results of molecular dynamics simulations depend on the force fields used (49), which have been developed predominantly for globular proteins rather than the unique fold of collagen.

Interpretations of light-scattering data require a model (such as an ellipsoidal scatterer (19)), and it is possible to describe decay curves equally well assuming both compact and more extended molecular geometries (50). Furthermore, the effective size of the protein, as deduced from solution-based studies such as light scattering and rheology, is often affected by intermolecular interactions; thus, concentration-dependent studies are essential to determine the spatial extent of a given molecule in isolation (51). Given collagen's propensity to self-associate (52,53), intermolecular interactions may contribute to larger estimates of collagen's size in solution. The rheological studies of (8) found that pH does not have a significant effect on collagen's flexibility; however, to reach this conclusion, the authors had to assume that type I collagen exhibited a nonuniform flexibility at neutral pH, with ends considerably more flexible than the central region of the triple helix.

The two studies most similar to ours, which used AFM imaging (13) and electron microscopy (18) to analyze collagen's flexibility, based their estimates of persistence length on the distance between the endpoints of the collagen chain. As we have shown (Figs. 4 and 5), naïve application of Eq. 1 can lead to significant underestimation of persistence length for curved chains. Consistently, our estimates of collagen's persistence length are greater than theirs obtained at low ionic strength. At high ionic strength and neutral pH, our results and those of the previous AFM imaging study (13) are in reasonable agreement: in this condition, collagen is well described by the standard WLC model and curvature may be disregarded. In general, however, our findings demonstrate the importance of mapping the functional forms of $\langle R^2(s) \rangle$ and $\langle \cos \theta(s) \rangle$ and caution against determining persistence length by using only chain end-point separations.

In our description and analysis of collagen, we have made two major assumptions. First, neither the WLC nor cWLC model takes into account the helical nature of collagen. Thus, flexibility of the chain derives only from bending deformations. If the intrinsic curvature described by our model arises because of a surface-collagen interaction inducing altered twist in the protein, then torsional stiffness of the triple helix (54) is an important parameter, and twist-bend coupling (55) may be required to fully describe collagen's flexibility.

Second, our analysis procedures treat collagen as an apolar, homogenous WLC. We are not able to discriminate between N $\rightarrow$ C and C $\rightarrow$ N directionality in the imaged chains. Furthermore, persistence lengths and curvatures

reported in this work represent average values for the entire collagen sequence. In adopting this approach, we have implicitly assumed that the $(Gly-X-Y)_n$ triple-helix-forming motif dominates collagen's response, with local sequence variations (e.g., fraction of prolines occupying the X and Y positions) viewed as introducing only minor perturbations. The flexibility of fibrillar collagen triple helices appears approximately uniform in electron microscopy images (18), supporting this assumption. However, molecular dynamics simulations (56) and modeling of atomistic collagen structures (54) suggest that collagen possesses a sequence-dependent flexibility commensurate with the expected variations in pitch observed for different triple helical sequences (2,44). It is interesting to speculate how a sequence-dependent mechanical signature may provide additional physical cues to interaction partners (48). Experimentally elucidating the detailed sequence-dependent flexibility of collagen is an exciting future prospect that will require higher-precision algorithms than applied thus far to this mechanically important protein.

CONCLUSIONS

Using AFM, we determined the persistence length of different types and sources of collagen in the presence of different cosolutes (K^+, Cl^-, acetate). Although collagens of different type and source exhibited only minor differences in persistence length, our initial findings implied that ionic strength and to a lesser extent pH modulate collagen's flexibility. Closer inspection of these results, however, revealed this interpretation to be an artifact of the model used for analysis. Rather than behaving as an intrinsically straight WLC, collagen deposited in low salt conditions was found to adopt preferentially curved conformations on the mica surface. Statistical analysis of chain conformations demonstrates that this curvature is a thermodynamic property, with chains appearing equilibrated in two dimensions. The extent of curvature depends strongly on salt concentration and pH, with persistence length remaining roughly constant across all KCl conditions. Thus, collagen possesses an experimentally tunable curvature and is to our knowledge the first such example of a cWLC.

Our findings provide a possible explanation for the range of persistence lengths reported for collagen in the literature: rather than a direct modulation of its flexibility, the presence of certain cosolutes may induce curvature along the collagen backbone. Whether or not the curvature seen in our results is intrinsic to collagen or is caused by cosolute-dependent interactions with the imaging surface remains to be determined. If intrinsic, then the modulation of collagen's conformation by salt and pH has potentially broad implications. The acidification experienced along the secretory pathway (25) may play a heretofore unexplored role in controlling collagen's intracellular compactness. Changes in salt concentration experienced upon secretion may also contribute to extracellular self-assembly of fibrillar collagens; for example, chloride ions are essential for the distinct network assembly of type IV collagen (57). Such changes in pH and cosolute concentrations could therefore influence collagen self-assembly through modification of its conformational properties: collagen does not assemble into fibrillar structures at low pH or at low ionic strength (38,58), precisely those conditions in which we observe its strongest curvature.

The strong agreement between measures of chain flexibility and the predictions of the cWLC model suggest that this formalism may be useful for interpreting the mechanical properties of other biological filaments. The intrinsic curvature of certain DNA sequences (59–61) is important for gene regulation (62), and amyloid fibrils (36) and coiled-coil proteins (63) have been found to exhibit preferentially curved conformations at interfaces. Applying the cWLC model to describe these systems and others is likely to provide further insight into their equilibrium structures and the energies governing their mechanical properties.

The datasets generated and analyzed during the current study are available from the corresponding author on request.

SUPPORTING MATERIAL

Supporting Materials and Methods, seven figures, and two tables are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495 (18)31027-0.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

SUPPORTING CITATIONS

References (64–68) appear in the Supporting Material.

REFERENCES

1. Shoulders, M. D., and R. T. Raines. 2009. Collagen structure and stability. *Annu. Rev. Biochem.* 78:929–958.

2. Bella, J. 2016. Collagen structure: new tricks from a very old dog. *Biochem. J.* 473:1001–1025.

3. Marini, J. C., A. Forlino, …, P. H. Byers. 2007. Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum. Mutat.* 28:209–221.

4. Byers, P. H., and M. L. Murray. 2012. Heritable collagen disorders: the paradigm of the Ehlers—Danlos syndrome. *J. Invest. Dermatol.* 132:E6–E11.

5. Myllyharju, J., and K. I. Kivirikko. 2001. Collagens and collagen-related diseases. *Ann. Med.* 33:7–21.

6. Utiyama, H., K. Sakato, …, M. Kurata. 1973. Flexibility of tropocollagen from sedimentation and viscosity. *Biopolymers.* 12:53–64.

7. Nestler, F. H., S. Hvidt, …, A. Veis. 1983. Flexibility of collagen determined from dilute solution viscoelastic measurements. *Biopolymers.* 22:1747–1758.

8. Amis, E. J., C. J. Carriere, …, A. Veis. 1985. Effect of pH on collagen flexibility determined from dilute solution viscoelastic measurements. *Int. J. Biol. Macromol.* 7:130–134.

9. Alberts, B., A. Johnson, …, P. Walter. 2007. Molecular Biology of the Cell. Garland Science, New York, NY.

10. Raote, I., M. Ortega-Bellido, …, V. Malhotra. 2018. TANGO1 builds a machine for collagen export by recruiting and spatially organizing COPII, tethers and membranes. *eLife.* 7:e32723.

11. Sun, Y. L., Z. P. Luo, …, K. N. An. 2002. Direct quantification of the flexibility of type I collagen monomer. *Biochem. Biophys. Res. Commun.* 295:382–386.

12. Sun, Y. L., Z. P. Luo, …, K. N. An. 2004. Stretching type II collagen with optical tweezers. *J. Biomech.* 37:1665–1669.

13. Lovelady, H. H., S. Shashidhara, and W. G. Matthews. 2014. Solvent specific persistence length of molecular type I collagen. *Biopolymers.* 101:329–335.

14. Wieczorek, A., N. Rezaei, …, N. R. Forde. 2015. Development and characterization of a eukaryotic expression system for human type II procollagen. *BMC Biotechnol.* 15:112.

15. Buehler, M. J., and S. Y. Wong. 2007. Entropic elasticity controls nanomechanics of single tropocollagen molecules. *Biophys. J.* 93:37–43.

16. Chang, S. W., S. J. Shefelbine, and M. J. Buehler. 2012. Structural and mechanical differences between collagen homo- and heterotrimers: relevance for the molecular origin of brittle bone disease. *Biophys. J.* 102:640–648.

17. Varma, S., J. P. Orgel, and J. D. Schieber. 2016. Nanomechanics of type I collagen. *Biophys. J.* 111:50–56.

18. Hofmann, H., T. Voss, …, J. Engel. 1984. Localization of flexible sites in thread-like molecules from electron micrographs. Comparison of interstitial, basement membrane and intima collagens. *J. Mol. Biol.* 172:325–343.

19. Claire, K., and R. Pecora. 1997. Translational and rotational dynamics of collagen in dilute solution. *J. Phys. Chem. B.* 101:746–753.

20. Vesentini, S., A. Redaelli, and A. Gautieri. 2013. Nanomechanics of collagen microfibrils. *Muscles Ligaments Tendons J.* 3:23–34.

21. Adhikari, A. S., E. Glassey, and A. R. Dunn. 2012. Conformational dynamics accompanying the proteolytic degradation of trimeric collagen I by collagenases. *J. Am. Chem. Soc.* 134:13259–13265.

22. Baumann, C. G., S. B. Smith, …, C. Bustamante. 1997. Ionic effects on the elasticity of single DNA molecules. *Proc. Natl. Acad. Sci. USA.* 94:6185–6190.

23. Wenner, J. R., M. C. Williams, …, V. A. Bloomfield. 2002. Salt dependence of the elasticity and overstretching transition of single DNA molecules. *Biophys. J.* 82:3160–3169.

24. Canty, E. G., and K. E. Kadler. 2005. Procollagen trafficking, processing and fibrillogenesis. *J. Cell Sci.* 118:1341–1353.

25. Demaurex, N. 2002. pH Homeostasis of cellular organelles. *News Physiol. Sci.* 17:1–5.

26. MATLAB and Statistics Toolbox Release. 2017. The MathWorks, Inc., Natick, MA.

27. Lamour, G., J. B. Kirkegaard, …, J. Gsponer. 2014. Easyworm: an open-source software tool to determine the mechanical properties of worm-like chains. *Source Code Biol. Med.* 9:16.

28. Faas, F. G., B. Rieger, …, D. I. Cherny. 2009. DNA deformations near charged surfaces: electron and atomic force microscopy views. *Biophys. J.* 97:1148–1157.

29. Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer, Berlin, Germany.

30. Usov, I., and R. Mezzenga. 2015. FiberApp: an open-source software for tracking and analyzing polymers, filaments, biomacromolecules, and fibrous objects. *Macromolecules.* 48:1269–1280.

31. Báez, J., D. Olsen, and J. W. Polarek. 2005. Recombinant microbial systems for the production of human collagen and gelatin. *Appl. Microbiol. Biotechnol.* 69:245–252.

32. Bozec, L., and M. Horton. 2005. Topography and mechanical properties of single molecules of type I collagen using atomic force microscopy. *Biophys. J.* 88:4223–4231.

33. Rivetti, C., M. Guthold, and C. Bustamante. 1996. Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J. Mol. Biol.* 264:919–932.

34. Perret, S., C. Merle, …, F. Ruggiero. 2001. Unhydroxylated triple helical collagen I produced in transgenic plants provides new clues on the role of hydroxyproline in collagen folding and fibril formation. *J. Biol. Chem.* 276:43693–43698.

35. Craig, A., and E. M. Terentjev. 2006. Auxiliary field theory of polymers with intrinsic curvature. *Macromolecules.* 39:4557–4565.

36. Jordens, S., E. E. Riley, …, R. Mezzenga. 2014. Adsorption at liquid interfaces induces amyloid fibril bending and ring formation. *ACS Nano.* 8:11071–11079.

37. Narayanan, B., G. H. Gilmer, …, C. V. Ciobanu. 2014. Self-assembly of collagen on flat surfaces: the interplay of collagen-collagen and collagen-substrate interactions. *Langmuir.* 30:1343–1350.

38. Jiang, F., H. Hörber, …, D. J. Müller. 2004. Assembly of collagen into microribbons: effects of pH and electrolytes. *J. Struct. Biol.* 148:268–278.

39. Leow, W. W., and W. Hwang. 2011. Epitaxially guided assembly of collagen layers on mica surfaces. *Langmuir.* 27:10907–10913.

40. Zanaboni, G., A. Rossi, …, R. Tenni. 2000. Stability and networks of hydrogen bonds of the collagen triple helical structure: influence of pH and chaotropic nature of three anions. *Matrix Biol.* 19:511–520.

41. Freudenberg, U., S. H. Behrens, …, C. Werner. 2007. Electrostatic interactions modulate the conformation of collagen I. *Biophys. J.* 92:2108–2119.

42. Keshwani, N., S. Banerjee, …, G. I. Makhatadze. 2013. The role of cross-chain ionic interactions for the stability of collagen model peptides. *Biophys. J.* 105:1681–1688.

43. Walker, K. T., R. Nan, …, S. J. Perkins. 2017. Non-linearity of the collagen triple helix in solution and implications for collagen function. *Biochem. J.* 474:2203–2217.

44. Orgel, J. P., T. C. Irving, …, T. J. Wess. 2006. Microfibrillar structure of type I collagen in situ. *Proc. Natl. Acad. Sci. USA.* 103:9001–9005.

45. Seol, Y., J. Li, …, M. D. Betterton. 2007. Elasticity of short DNA molecules: theory and experiment for contour lengths of 0.6-7 microm. *Biophys. J.* 93:4360–4373.

46. Adhikari, A. S., J. Chai, and A. R. Dunn. 2011. Mechanical load induces a 100-fold increase in the rate of collagen proteolysis by MMP-1. *J. Am. Chem. Soc.* 133:1686–1689.

47. Camp, R. J., M. Liles, …, J. W. Ruberti. 2011. Molecular mechanochemistry: low force switch slows enzymatic cleavage of human type I collagen monomer. *J. Am. Chem. Soc.* 133:4073–4078.

48. Kirkness, M. W. H., and N. R. Forde. 2018. Single-molecule assay for proteolytic susceptibility: force-induced collagen destabilization. *Biophys. J.* 114:570–576.

49. Varma, S., M. Botlani, …, J. D. Schieber. 2015. Effect of intrinsic and extrinsic factors on the simulated D-band length of type I collagen. *Proteins.* 83:1800–1812.

50. Shayegan, M. 2014. Determining Local Viscoelastic Properties of Collagen Systems using Optical Tweezers. Simon Fraser University, Burnaby, Canada.

51. Muschol, M., and F. Rosenberger. 1995. Interactions in undersaturated and supersaturated lysozyme solutions: static and dynamic light scattering results. *J. Chem. Phys.* 103:10424–10432.

52. Shayegan, M., and N. R. Forde. 2013. Microrheological characterization of collagen systems: from molecular solutions to fibrillar gels. *PLoS One.* 8:e70590.

53. Shayegan, M., T. Altindal, …, N. R. Forde. 2016. Intact telopeptides enhance interactions between collagens. *Biophys. J.* 111:2404–2416.

54. Xu, F., H. Zheng, …, V. Nanda. 2017. Parallels between DNA and collagen - comparing elastic models of the double and triple helix. *Sci. Rep.* 7:12802.

55. Nomidis, S. K., F. Kriegel, …, E. Carlon. 2017. Twist-bend coupling and the torsional response of double-stranded DNA. *Phys. Rev. Lett.* 118:217801.

56. Ravikumar, K. M., and W. Hwang. 2008. Region-specific role of water in collagen unwinding and assembly. *Proteins.* 72:1320–1332.

57. Cummings, C. F., V. Pedchenko, …, B. G. Hudson. 2016. Extracellular chloride signals collagen IV network assembly during basement membrane formation. *J. Cell Biol.* 213:479–494.

58. Harris, J. R., A. Soliakov, and R. J. Lewis. 2013. In vitro fibrillogenesis of collagen type I in varying ionic and pH conditions. *Micron.* 49:60–68.

59. Calladine, C. R., H. R. Drew, and M. J. McCall. 1988. The intrinsic curvature of DNA in solution. *J. Mol. Biol.* 201:127–137.

60. Zuccheri, G., A. Scipioni, …, B. Samorì. 2001. Mapping the intrinsic curvature and flexibility along the DNA chain. *Proc. Natl. Acad. Sci. USA.* 98:3074–3079.

61. Vologodskaia, M., and A. Vologodskii. 2002. Contribution of the intrinsic curvature to measured DNA persistence length. *J. Mol. Biol.* 317:205–213.

62. Goyal, S., T. Lillian, …, N. C. Perkins. 2007. Intrinsic curvature of DNA influences LacR-mediated looping. *Biophys. J.* 93:4342–4359.

63. Murray, D. H., M. Jahnel, …, M. Zerial. 2016. An endosomal tether undergoes an entropic collapse to bring vesicles together. *Nature.* 537:107–111.

64. Landau, L. D., L. P. Pitaevskii, …, E. M. Lifshitz. 1986. Theory of Elasticity. Butterworth-Heinemann, Oxford, UK.

65. Rezaei, N. 2016. Mechanical Studies of Single Collagen Molecules Using Imaging and Force Spectroscopy. Simon Fraser University, Burnaby, Canada.

66. Sonka, M., V. Hlavac, and R. Boyle. 2014. Image Processing, Analysis and Machine Vision. Cengage Learning, Boston, MA.

67. Murugesapillai, D., S. Bouaziz, …, M. C. Williams. 2017. Accurate nanoscale flexibility measurement of DNA and DNA-protein complexes by atomic force microscopy in liquid. *Nanoscale.* 9:11327–11337.

68. Cramer, D. 1997. Basic Statistics for Social Research: Step-by-Step Calculations and Computer Techniques using Minitab. Routledge, New York.

Supplemental Information

Environmentally Controlled Curvature of Single Collagen Proteins

Nagmeh Rezaei, Aaron Lyons, and Nancy R. Forde

Derivation of the curved Worm-Like Chain (cWLC) model

The standard WLC model assumes that there is an energetic cost to bend an intrinsically straight rod. Specifically, to bend a semi-flexible, intrinsically straight rod with length δl into a circular arc with central angle $\delta\theta$, the energy required is (1)

$$E_{\text{bend}} = \frac{\alpha\delta\theta^2}{2\delta l} = \frac{p\delta\theta^2}{2\delta l}k_{\text{B}}T. \tag{S1}$$

α is the bending rigidity of the chain, related to the persistence length p through the relation $p = \alpha/k_{\text{B}}T$, where $k_{\text{B}}T$ is the product of the Boltzmann constant and the absolute temperature. The central angle of this arc, $\delta\theta$, is equivalent to the angle between the tangents at the beginning and end of the segment. The lowest energy conformation of the segment is a straight line, with the energy increasing harmonically about this configuration. The Boltzmann distribution provides the distribution of angles that this segment adopts in the presence of thermal noise, which is given by

$$P(\delta\theta) = \sqrt{\frac{p}{2\pi\delta l}}\exp\left(-\frac{p(\delta\theta)^2}{2\delta l}\right). \tag{S2}$$

This a normal distribution with mean $\langle\delta\theta\rangle = 0$ and variance $\sigma_{\delta\theta}^2 = \frac{\delta l}{p}$.

If the chain is intrinsically curved, the bending energy is modified to reflect deviations from this bent state.

$$E_{\text{bend}} = \frac{p(\delta\theta - \kappa_o\delta l)^2}{2\delta l}k_{\text{B}}T, \tag{S3}$$

where κ_o is the intrinsic curvature of the chain, defined as $\kappa_o = \langle\frac{d\theta}{dl}\rangle$. The distribution of angles that this intrinsically curved segment adopts in the presence of thermal noise is given by

$$P(\delta\theta) = \sqrt{\frac{p}{2\pi\delta l}}\exp\left(-\frac{p(\delta\theta - \kappa_o\delta l)^2}{2\delta l}\right). \tag{S4}$$

As before, this is a normal distribution with variance $\sigma_{\delta\theta}^2 = \frac{\delta l}{p}$, but now centered about a mean $\langle\delta\theta\rangle = \kappa_o\delta l$.

We now consider a longer chain segment, comprised of N segments each of length δl, such that the total length of this chain $l = N\delta l$. This longer chain is not necessarily a circular arc, but is made up of N short circular arc segments, each with angular difference $\delta\theta_i$ distributed about $\kappa_o\delta l$ as per equation (S4). The angular difference θ between the ends of this larger segment is given by the sum of the N angles adopted by the smaller circular arcs:

$$\theta = \sum_{i=1}^{N}\delta\theta_i. \tag{S5}$$

Since the sum of normal random variables is normally distributed, the full distribution of θ is completely described by its mean and variance, given respectively by

$$\langle \theta \rangle = N\langle \delta\theta \rangle = \kappa_o N\delta l = \kappa_o l \tag{S6}$$

and

$$\sigma_\theta^2 = N\sigma_{\delta\theta}^2 = \frac{N\delta l}{p} = \frac{l}{p}. \tag{S7}$$

The angular distribution of this segment of length l is therefore given by

$$P(\theta) = \sqrt{\frac{p}{2\pi l}} e^{-\frac{p(\theta-\kappa_o l)^2}{2l}}. \tag{S8}$$

Note that, because $\langle \theta \rangle \neq 0$, this chain exhibits a net global curvature.

For any arbitrary segment length s, we can now calculate the average tangent vector correlation as

$$\langle \hat{t}(s+s') \cdot \hat{t}(s') \rangle = \langle cos\theta(s) \rangle = \sqrt{\frac{p}{2\pi l}} \int_{-\infty}^{\infty} \exp\left(-\frac{p(\theta-\kappa_o s)^2}{2s}\right) \cos\theta \, d\theta, \tag{S9}$$

where s' defines an arbitrary starting position of the segment along the contour. This simplifies to

$$\langle cos\theta \rangle = \exp\left(-\frac{s}{2p}\right) \cos(\kappa_o s). \tag{S10}$$

However, since $\hat{t}(s+s') \cdot \hat{t}(s') \leq 1$ always, then its average $\langle \hat{t}(s+s') \cdot \hat{t}(s') \rangle \leq 1$. This requires that $s \geq 0$, so we can write

$$\langle cos\theta(s) \rangle = \exp\left(-\frac{|s|}{2p}\right) \cos(\kappa_o |s|) \tag{S11}$$

for all s.

Similarly, we can calculate the mean squared end-to-end distance of the segment as

$$\langle R^2(s) \rangle = \langle \vec{R}(s) \cdot \vec{R}(s) \rangle = \langle \left(\int_0^s \hat{t}(s')ds'\right)\left(\int_0^s \hat{t}(s'')ds''\right) \rangle = \int_0^s \int_0^s \langle \hat{t}(s') \cdot \hat{t}(s'') \rangle ds'ds'', \tag{S12}$$

where the order of operations was interchanged because averaging and integrating are both linear operations. Using $\langle \hat{t}(s') \cdot \hat{t}(s'') \rangle = \exp\left(-\frac{|s''-s'|}{2p}\right) \cos(\kappa_o |s'' - s'|)$ from equation (S11) gives

$$\langle R^2(s) \rangle = \int_0^s \left\{ \int_0^{s''} \exp\left[-\frac{(s''-s')}{2p}\right] \cos[\kappa_o(s'' - s')] \, ds' + \right.$$
$$\left. \int_{s''}^s \exp\left[-\frac{(s'-s'')}{2p}\right] \cos[\kappa_o(s' - s'')] \, ds' \right\} ds''. \tag{S13}$$

Evaluating this expression yields

$$\langle R^2(s) \rangle = \frac{4sp}{(1+4\kappa_0{}^2p^2)^2} \left\{ 1 - \frac{2p}{s}(1 - 4\kappa_0{}^2p^2)\left[1 - \cos(\kappa_0 s)\exp\left(-\frac{s}{2p}\right)\right] + \right.$$
$$\left. \frac{4\kappa_0 p^2}{s}\left[\kappa_0 s - 2\sin(\kappa_0 s)\exp\left(-\frac{s}{2p}\right)\right]\right\}. \tag{S14}$$

Once again, this expression is valid only for $s \geq 0$, as negative values of s can produce values of $\langle R^2(s) \rangle$ that are negative. Thus,

$$\langle R^2(s) \rangle = \frac{4|s|p}{(1+4\kappa_0{}^2p^2)^2} \left\{ 1 - \frac{2p}{|s|}(1 - 4\kappa_0{}^2p^2)\left[1 - \cos(\kappa_0|s|)\,e^{-\frac{|s|}{2p}}\right] + \right.$$
$$\left. \frac{4\kappa_0 p^2}{|s|}\left[\kappa_0|s| - 2\sin(\kappa_0|s|)\,e^{-\frac{|s|}{2p}}\right]\right\}. \tag{S15}$$

In the case where $\kappa_o = 0$, which corresponds to the standard two-dimensional worm-like chain, the expressions for the tangent vector correlation and mean squared end-to-end distance reduce to

$$\langle \cos\theta(s) \rangle = \exp\left(-\frac{|s|}{2p}\right) \tag{S16}$$

and

$$\langle R^2(s) \rangle = 4|s|p\left\{ 1 - \frac{2p}{|s|}\left[1 - \exp\left(-\frac{|s|}{2p}\right)\right]\right\}. \tag{S17}$$

These are identical to previously derived results (2), and to equations (2) and (1) in the main text, respectively.

Because both $\langle R^2(s) \rangle$ (equation S15) and $\langle \cos\theta(s) \rangle$ (equation S11) are even functions of curvature κ_o, we report only the magnitude of the curvature from our fits.

SmarTrace Algorithm

From AFM images of biopolymers, we wished to analyze chain configurations and extract mechanical properties. For this purpose, we developed a robust and efficient chain tracing software in MATLAB, dubbed "SmarTrace", which traces the centerline of each molecule with sub-pixel resolution and provides comprehensive statistical analysis of the chains.

Chain tracing

An overview of the SmarTrace workflow is discussed below. Details are available in reference (3).

1.   In a visual user interface (adapted from the EasyWorm package (4) within MATLAB (5)), the user selects a few points on or near the backbone of the chain to be traced.
2.   SmarTrace fits an initial spline to these user-defined points and extracts points along the spline separated by one nanometer.
3.   The program uses top-hat and median filtering (6) to improve the signal-to-noise ratio of the region surrounding the chain.
4.   To detect the best path describing the chain, a search window is defined for each point on the spline curve along the tangential direction of the initial spline. A search grid with sub-pixel resolution determines (interpolated) intensity values of the image for each grid point.
5.   For each point on the initial spline, a template pattern is matched with each point in the grid within the search window. The template resembles the intensity pattern of a cross-section of the chain with varying widths.
6.   A matching score is calculated for each possible width and location of this pattern. Cross-correlation scores are used to determine the best centerline position and width of the chain.
7.   To ensure stable results, a penalty term is added for sudden changes in width and/or direction of the chain.
8.   After the scores are finalized, the re-weighted maximum scores are used to extract the width and center of the chain at points spaced approximately 1 nanometer apart. A B-spline is fit to these points, resulting in a piecewise smooth polynomial that represents the polymer chain.

This method is not very sensitive to image quality and can successfully detect the chain backbone in noisy images; even if parts of the chain are slightly faded, the code is still able to trace the entire chain. The results also do not depend on where the user selects the points (validated by having different users trace the same set of experimental chains, and finding the same persistence length), and the initially selected points do not need to be located exactly on the chain centerline. These features, along with an efficient computation algorithm, make the code fast and easy to use for tracing and analyzing images of single polymers, as obtained for example by atomic force microscopy, electron microscopy or fluorescence microscopy.

<u>Statistical analysis of chain properties</u>

After the chains have been traced, statistical approaches are used to analyze their flexibility. First, the traced chains are partitioned into segments of varying lengths. Utilizing a method introduced in reference (7), each molecule is divided into multiple segments, with lengths drawn randomly from a pre-defined set of input values (here, 10 nm, 20 nm, 30 nm, ..., 200 nm). Once the lengths of these segments have been determined, their positions on the chain are shuffled to avoid accumulating shorter lengths towards one end of the chain. This process is repeated 50 times for each chain, allowing different regions of the chain to contribute to the statistics of different segment lengths. Within each draw, the sampled chain segments are nonoverlapping. As rare, longer segments are less useful for mathematical fitting and persistence length calculations, choosing a relatively short maximum segment length provides more samples in each bin. This method also allows for the use of partially traced chains, which is particularly helpful when ends of a molecule are not clear or chains intersect. In the current work, end regions (5 pixels $\approx$ 19.5 nm) of the chains were excluded from further analysis.

Several statistics are then calculated from the samples at each segment length s: the mean-squared end-to-end distance, $\langle R^2(s) \rangle$, and the mean cosine of the angle between the start and end of the segment, $\langle \cos \theta(s) \rangle$.

Validation of SmarTrace and analysis procedures

Validation with DNA images

DNA molecules of ~1 µm contour length were prepared by digestion of the pBluescript KS(+) plasmid with SspI, then purifying the longer linear, blunt-ended 2828 bp fragment. For imaging, DNA was deposited from a buffer containing 4 mM HEPES, 10 mM NaCl, 2 mM MgCl$_2$ at pH 7.4, then dried before imaging (Figure S1A). Note that Mg^{2+} is required for DNA deposition due to its negatively charged backbone; in contrast, at the pH values we have studied collagen, the protein has positively charged residues and thus does not require bridging cations for deposition.

DNA images were traced and analyzed with SmarTrace. Analysis of the angular distributions showed that they exhibit a kurtosis of 3 at all but the shortest segment length analysed (Figure S1B), consistent with a Gaussian distribution and equilibration. The mean squared end-to-end distance, $\langle R^2(s) \rangle$, is shown in Figure S1C along with a 2D worm-like chain fit (Equation 1). Figure S1D shows the experimental data for tangent vector correlation, $\langle \cos \theta (s) \rangle$, and the 2D WLC fit (Equation 2). A persistence length of $p = 62 \pm 3$ nm was obtained from these analyses (Table S1), consistent with previous results (8). Statistical analysis demonstrates that DNA deposited and imaged under these conditions is better described by the standard WLC than by the curved WLC (Table S2), and with a fit to the cWLC returning $\kappa_o = 0$ and $p = 62$ nm, *i.e.*, reducing to the standard WLC model.

We also traced and analysed images of DNA molecules with 1 µm contour length (*N*=24), which were provided with the software package Easyworm (4). A persistence length of 52 ± 2 nm was obtained from these analyses, consistent with expectations for DNA (2, 4).

Validation with simulated worm-like chains

To validate the methodology used for chain tracing and analysis within SmarTrace, two-dimensional worm-like chains were simulated and converted into pseudo-AFM images. These images were then traced with SmarTrace to ensure that the algorithm was able to accurately recover the input persistence length and curvature. The workflow for the simulation is as follows:

1.   The desired persistence length p, curvature κ_0, contour length L and width of the chains w – as well as the average number of chains per image – are input by the user.
2.   For the first chain, an angle θ_0 between 0 and 360°, as well as two values x_0 and y_0 between 0 and 2000 nm are sampled randomly from a uniform distribution. These values represent the starting angle and initial xy-coordinates of the chain, respectively.
3.   The curvature of the chain is chosen to be either κ_0 or $-\kappa_0$ with equal chance; this represents the ability of the chain to "lie down" on the mica surface with either a left-handed or right-handed curvature.
4.   Using a step size $\delta s = 0.5$ nm, an angle $\delta\theta$ is sampled from a normal distribution with mean $\pm\kappa_o \delta s$ and variance $\frac{\delta s}{p}$.

5. The next point on the chain is placed at $x_1 = x_0 + \delta s \cos(\theta_1)$ and $y_1 = y_0 + \delta s \sin(\theta_1)$, where $\theta_1 = \theta_0 + \delta\theta$.
6. This process is repeated, choosing a random $\delta\theta$ as above, generating $\theta_{i+1} = \theta_i + \delta\theta$ and placing the next point on the chain at $x_{i+1} = x_i + \delta s \cos(\theta_{i+1})$ and $y_{i+1} + \delta s \sin(\theta_{i+1})$.
7. This process is terminated after n steps, where $n\delta s = L$.
8. Steps 2 through 7 are repeated for every chain in the image.
9. A 512-by-512 array of pixels is overlaid on top of the image, discretizing the image into square pixels with a side length of 3.90625 nm, the same as all experimental images used in this work.
10. The intensity of each pixel is populated by considering every point on each chain as an intensity source, which contributes an intensity

$$I(x,y) = I_o \exp\left(-\frac{(x-x_i)^2+(y-y_i)^2}{2w^2}\right) \tag{S18}$$

to the pixel centered at (x, y). The intensity contributions from every point on every chain are then summed to yield the overall intensity of this pixel.
11. At this point, a 512-by-512 pixel noise matrix is generated and overlaid with the image containing the simulated chains. This noise matrix contains experimentally realistic correlated noise, obtained as described in the following subsection.
12. The matrix containing the intensities from the simulated chains is then scaled and summed with the noise matrix, with the scaling factor being chosen to visually match the signal-to-noise ratio of our experimental AFM images.
13. This noisy image is then converted to an 8-bit grayscale image, allowing it to be traced with SmarTrace.

We generated two sets of images with parameters chosen to emulate experimentally gathered AFM images of collagen: both sets were given a width parameter of 7 nm and a contour length of 300 nm – chosen to replicate AFM images of collagen. For simplicity, background noise was not included in these simulations.

The first set was generated with a persistence length of 85 nm and zero curvature. An example image and standard WLC model fits to the data (Equations 1, 2 from the main text) are shown in Figures S2A-C. Results from fits with both the WLC and cWLC are included in Table S1, in which it is clear that no intrinsic curvature is found in these chains (see also Table S2). The traced data also reproduce the Gaussian properties of the simulated chains, as shown by the kurtosis and normality of the angular distributions at different segment lengths (Figures S2D and S2E). The standard error in the kurtosis (SEK), used to generate the error bars in Figure S2D, is given by (9)

$$\text{SEK} = \sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}}, \tag{S19}$$

where n is the number of observations comprising the distribution. The expected distribution shown in Figure S2E is given by Eq. S8 with $p = 85$ nm, $s = 50$ nm and $\kappa_o = 0$ nm^{-1}.

The second set of chains was generated with $p = 50$ nm and an intrinsic curvature of $\kappa_o = 0.02$ nm^{-1}; an image of these chains is shown in Figure S3A. Due to the presence of chains with both positive and negative curvature (see step 3, above), the angular distributions extracted from these curved chains will be a sum of two Gaussian distributions, with mean $\pm \kappa_o s$ and variance $\frac{s}{p}$. These two distributions may not have equal amplitude (for a small number of sampled chains), making the properties of the expected distribution difficult to calculate. However, the expected distribution of the absolute-values of the angles is given by

$$P(|\theta|) = \sqrt{\frac{p}{2\pi s}}\left[\exp\left(-\frac{p(|\theta|-\kappa_o s)^2}{2s}\right) + \exp\left(-\frac{p(|\theta|+\kappa_o s)^2}{2s}\right)\right] \tag{S20}$$

regardless of the proportion of the two distributions. Figure S3B shows the expected distribution for $p = 50$ nm, $\kappa_o = 0.02$ nm^{-1} and $s = 50$ nm, plotted with a histogram of the angles extracted from the simulated chains at a 50 nm segment length, demonstrating good agreement between the two distributions. Figures S3C and S3D show fits of the curved WLC expressions (Equations 3, 4) to the traced data, both of which yield persistence lengths and curvatures close to the input parameters. Fitting with the standard WLC model results in an underestimate of persistence length (Table S1), and is a poorer model for these chains (Table S2).

Simulation of AFM Image Noise

In order to model to noise present in the background of our AFM images, we characterized the intensity fluctuations in typical images, then used this information to produce simulated images of chains with realistic noise.

We wished to simulate a row of the image by the vector $\vec{X} = (X_1, \ldots, X_n)^T$, where each element X_i of the vector is a normally distributed random variable with mean μ_i and variance σ_i^2. Here, X_1 refers to the intensity of the leftmost pixel in the row, X_2 to the intensity of the pixel second from the left, and so on. We allowed correlation between nearby pixels within a row, *i.e.*, we allowed for $\langle X_i X_j \rangle \neq 0$. This is similar to experimental observations of correlated noise in the scanning direction. We calculated the intensity of each row by

$$\vec{X} = L\vec{Z} + \vec{\mu}, \tag{S21}$$

where the elements μ_i represent the average intensity of the pixels in the ith column; $\vec{Z} = (Z_1, \ldots, Z_n)$, where the Z_i are independently distributed normal random variables with zero mean and unit variance ($\langle Z_i Z_j \rangle = \delta_{ij}$); and L is an n-by-n matrix determined from experimental images that introduces correlations in the simulated noise.

To construct L and $\vec{\mu}$, we used three scarcely populated 2000 nm-by-2000 nm AFM images (containing fewer than 6 chains per image) and assumed the intensities to be uncorrelated between rows. μ_i was given by the average intensity of the n pixels in the ith column. We also determined the correlation matrix C, in which the correlation matrix elements $\langle c_i c_j \rangle$ represent the covariance between the intensities of columns i and j, averaged over all rows, while diagonal

elements comprise the variance of intensities within each column ($c_{ii} = \langle c_i c_i \rangle = \sigma_i^2$). Because $\boldsymbol{C}$ is Hermitian and positive-definite, $\boldsymbol{L}$ can be found from $\boldsymbol{C} = \boldsymbol{L}\boldsymbol{L}^T$. Practically, we determine $\boldsymbol{L}$ by Cholesky decomposition of $\boldsymbol{C}$, implemented using the MATLAB function `chol(C)` (5).

Realizations of $\vec{X}$ are then generated by simulating a vector $\vec{Z}$ of normal variables with zero mean and unit variance, matrix multiplying by $\boldsymbol{L}$, and adding $\vec{\mu}$. These values are then used to populate the first row of pixel intensities in the noise matrix, after which new realizations of $\vec{X}$ are simulated to populate the subsequent rows.

Bayesian Information Criterion (BIC)

The Bayesian Information Criterion (BIC) (10) was used to assess whether the standard or curved WLC better describes the data:

$$BIC = -2\ln\hat{L} + k\ln N. \tag{S22}$$

$\hat{L}$ represents the maximum likelihood, k the number of parameters in the model ($k = 1$ for the WLC and $k = 2$ for the cWLC), and N the number of data points to be fit ($N = 20$ for each data set in our study).

To determine the BIC, we need to obtain the maximum likelihood, by determining the fit parameters θ (e.g. p for the WLC or p and κ_o for the cWLC) that maximize the likelihood function $L(\theta; y_i, \sigma_i^2, f)$ for each model description f, given our data. We assume that our observed data are independent and sampled from normal distribution with means y_i and variances σ_i^2. In this case, the likelihood function is given by

$$L(\theta; y_i, \sigma_i^2, f) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-(y_i - f(s_i,\theta))^2}{2\sigma_i^2}\right). \tag{S23}$$

This can be rewritten as

$$L(\theta; y_i, \sigma_i^2, f) = \frac{1}{(2\pi)^{\frac{n}{2}} \prod_{i=1}^{n} \sigma_i} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \frac{(y_i - f(s_i,\theta))^2}{\sigma_i^2}\right). \tag{S24}$$

The sum within the exponential is the chi-squared statistic

$$\chi^2(\theta) = \sum_{i=1}^{n} \frac{(y_i - f(s_i,\theta))^2}{\sigma_i^2}, \tag{S25}$$

allowing us to further simplify the likelihood function to

$$L(\theta; y_i, \sigma_i^2, f) = \frac{1}{(2\pi)^{\frac{n}{2}} \prod_{i=1}^{n} \sigma_i} \exp\left(-\frac{\chi^2(\theta)}{2}\right). \tag{S26}$$

Since χ^2 is strictly positive, and the variable parameters θ appear only within $\chi^2(\theta)$, we simply need to choose the parameters such that the chi-squared value is minimized, as this will maximize the value of L. This is precisely what has been done in the fitting, where θ have been determined to minimize $\chi^2(\theta)$. The maximum value of the likelihood function is then just

$$\hat{L}(\theta; y_i, \sigma_i^2, f) = \frac{1}{(2\pi)^{\frac{n}{2}} \prod_{i=1}^{n} \sigma_i} \exp\left(-\frac{\chi_{min}^2}{2}\right). \tag{S27}$$

Now, the BIC can be determined from equation (S21):

$$BIC = \chi_{min}^2 + n \ln(2\pi) + 2 \sum_{i=1}^{n} \ln(\sigma_i) + k \ln N. \tag{S28}$$

Since the second and third terms are independent of the model tested, for comparison purposes we need to calculate only

$$BIC = \chi_{min}^2 + k \ln N. \tag{S29}$$

The model with the lower BIC is more likely; the probability that the other model describes the data is given by

$$P_{other} = \exp\left[-\frac{1}{2}\left(\text{BIC}_{\text{larger}} - \text{BIC}_{\text{smaller}}\right)\right]. \tag{S30}$$

The results of this analysis are tabulated in Table S2 for experimental data on collagen and DNA and for simulated chains. Because often the best fit with the cWLC model results in zero curvature, χ_{min}^2 is identical between the WLC and cWLC models. In this case, the values of BIC differ by ~3 due to the difference in number of model parameters k, and thus by equation (S30) have a probability of ~0.2 that the cWLC model better describes the data.

Table S1 - Comprehensive list of fitting parameters obtained for different collagen types and co-solute conditions. Persistence lengths, p, are given in nm, while curvature κ_0 is in units of nm^{-1}. Reported errors, Δ, represent 95% confidence intervals in the fitting parameters. χ^2_{red} values are determined by dividing the minimized χ^2_{min} values by the number of degrees of freedom, $N\text{-}k$, where N is the number of points in each data set ($N = 20$) and k is the number of model parameters ($k = 1$ for the WLC and $k = 2$ for the cWLC).

Collagen Type	Solution	Length Traced (μm)	Standard Model Fits						Curved Model Fits									
			$\langle R^2 \rangle$			$\langle \cos\theta \rangle$			$\langle R^2 \rangle$					$\langle \cos\theta \rangle$				
			p	Δp	χ^2_{red}	p	Δp	χ^2_{red}	p	Δp	κ_o	$\Delta\kappa_o$	χ^2_{red}	p	Δp	κ_o	$\Delta\kappa_o$	χ^2_{red}
Rat I	Water	14.7	64	13	111	53	11	55	117	10	0.0165	0.0007	3.1	87	6	0.0125	0.0005	1.6
Rat I	100 μM KCl	6.3	41	5	19	38	7	19	56	7	0.0150	0.0023	6.1	59	12	0.0134	0.0022	7.2
Rat I	1 mM KCl	12.4	78	12	42	61	10	30	119	9	0.0125	0.0007	2.2	90	6	0.0105	0.0006	1.7
Rat I	10 mM KCl	17.2	97	5	5.6	85	4	1.9	95	10	0.0	0.3	6.3	83	6	0.0	0.4	2.2
Rat I	100 mM KCl	19.9	123	13	27	118	18	23	121	29	0.0	4.2	29	118	36	0.0	100.8	25
Rat I	1 mM HCl	20.4	43	6	94	36	5	38	65	4	0.0175	0.0008	3.8	49	3	0.0145	0.0011	3.6
Rat I	10 mM KCl + 1 mM HCl	33.6	83	11	70	66	10	57	135	10	0.0115	0.0006	3.1	99	4	0.0094	0.0003	1.2
Rat I	100 mM KCl + 1 mM HCl	20.9	106	6	13	101	8	12	102	11	0.0	0.4	16	100	13	0.0	6.5	13
Human I	100 mM KCl + 1 mM HCl	20.7	89	4	7.4	84	6	8.8	87	7	0.0	0.2	8.0	88	12	0.0	0.3	10
Human II	100 mM KCl + 1 mM HCl	12	100	4	3.0	96	5	2.2	106	6	0.0047	0.0016	2.1	95	8	0.0	0.1	2.4
Human III	100 mM KCl + 1 mM HCl	16.7	85	2	2.1	83	7	8.5	85	4	0.0014	0.0054	2.2	81	11	0.0	0.3	9.1
Rat I	20 mM Acetic Acid	24.7	35	5	101	31	4	39	55	3	0.0195	0.0008	3.1	43	2	0.0161	0.0008	2.0
Human I	20 mM Acetic Acid	50.4	42	6	186	34	6	132	66	3	0.0180	0.0006	4.1	52	2	0.0166	0.0005	2.1
Human II	20 mM Acetic Acid	13.2	64	14	120	50	12	78	109	9	0.0180	0.0008	3.9	85	5	0.0150	0.0005	1.3
Human III	20 mM Acetic Acid	40.4	50	9	229	39	7	130	79	6	0.0181	0.0008	8.3	59	2	0.0160	0.0005	1.9
Simulated WLC	p=85 nm	24.5	83	5	17	87	2	1.5	84	9	0.0000	0.3635	18	87	4	0.0010	0.0034	1.5
Simulated cWLC	p=50 nm, κ_o=0.02 nm⁻¹	17.7	41	5	68	36	9	118	58	2	0.0178	0.0006	1.2	65	4	0.0192	0.0008	2.6
2828 bp dsDNA	See supporting text	19.6	61	1	1.7	62	2	1.6	61	3	0.0000	1.6412	1.8	62	4	0.0	0.6	1.7

Table S2. Model selection using the Bayesian Information Criterion (BIC). BIC was calculated for the straight (WLC) and curved (cWLC) models using equation (S29).

| Sample	Condition	$\langle R^2 \rangle$				$\langle \cos\theta \rangle$			
		BIC straight	BIC curved	More probable?	P other?	BIC straight	BIC curved	More probable?	P other?
Rat Type-I	Water	2112	62	curved	< 1E-7	1040	36	curved	< 1E-7
Rat Type-I	100 µM KCl	366	116	curved	< 1E-7	358	136	curved	< 1E-7
Rat Type-I	1 mM KCl	804	46	curved	< 1E-7	579	37	curved	< 1E-7
Rat Type-I	10 mM KCl	109	119	straight	7E-3	39	45	straight	5E-2
Rat Type-I	100 mM KCl	516	523	straight	5E-2	444	447	straight	2E-1
Rat Type-I	1 mM HCl	1798	75	curved	< 1E-7	730	70	curved	< 1E-7
Rat Type-I	10 mM KCl + 1 mM HCl	1334	62	curved	< 1E-7	1082	27	curved	< 1E-7
Rat Type-I	100 mM KCl + 1 mM HCl	259	285	straight	2E-6	237	240	straight	2E-1
Human Type-I	100 mM KCl + 1 mM HCl	144	150	straight	5E-2	171	186	straight	6E-4
Human Type-II	100 mM KCl + 1 mM HCl	60	44	curved	3E-4	45	48	straight	2E-1
Human Type-III	100 mM KCl + 1 mM HCl	43	46	straight	2E-1	165	169	straight	1E-1
Rat Type-I	20 mM Acetic Acid	1916	61	curved	< 1E-7	752	41	curved	< 1E-7
Human Type-I	20 mM Acetic Acid	3533	79	curved	< 1E-7	2519	44	curved	< 1E-7
Human Type-II	20 mM Acetic Acid	2283	76	curved	< 1E-7	1480	29	curved	< 1E-7
Human Type-III	20 mM Acetic Acid	4348	156	curved	< 1E-7	2477	41	curved	< 1E-7
Simulated WLC	p = 85 nm	332	337	straight	8E-2	31	34	straight	2E-1
Simulated cWLC	p = 50 nm κ_0 = 0.02 nm^{-1}	1304	27	curved	< 1E-7	2253	52	curved	< 1E-7
2828 bp dsDNA	4 mM HEPES, 10 mM NaCl, 2 mM MgCl$_2$ (pH 7.4)	36	39	straight	2E-1	33	36	straight	2E-1
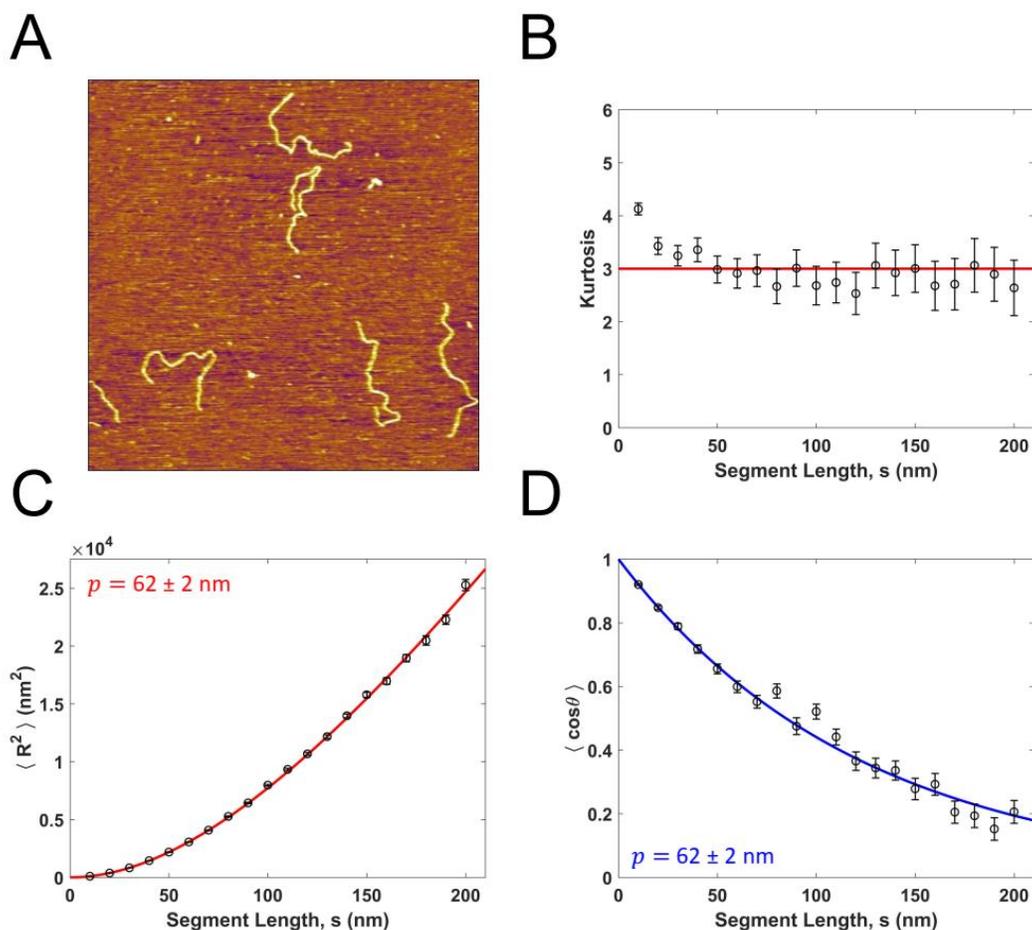
Figure S1. Quantifying the flexibility of dsDNA. (A) Example AFM image of 2828 bp linear dsDNA. (B) Kurtosis of angle distributions is close to 3 for all segment lengths, indicating that angles are normally distributed. (C) Mean squared end-to-end distance versus length along DNA molecules. Symbols represent the mean for each segment length while error bars show the standard error of the mean. The persistence length from the fit (red line; Equation 1) is 62 ± 2 nm. (D) Average tangent-tangent correlation vs. length along DNA molecules. The persistence length from the fit (red line; Equation 2) is 62 ± 2 nm. These values for persistence length are internally consistent and are in agreement with previous measurements of DNA deposited onto mica and imaged in air (8).
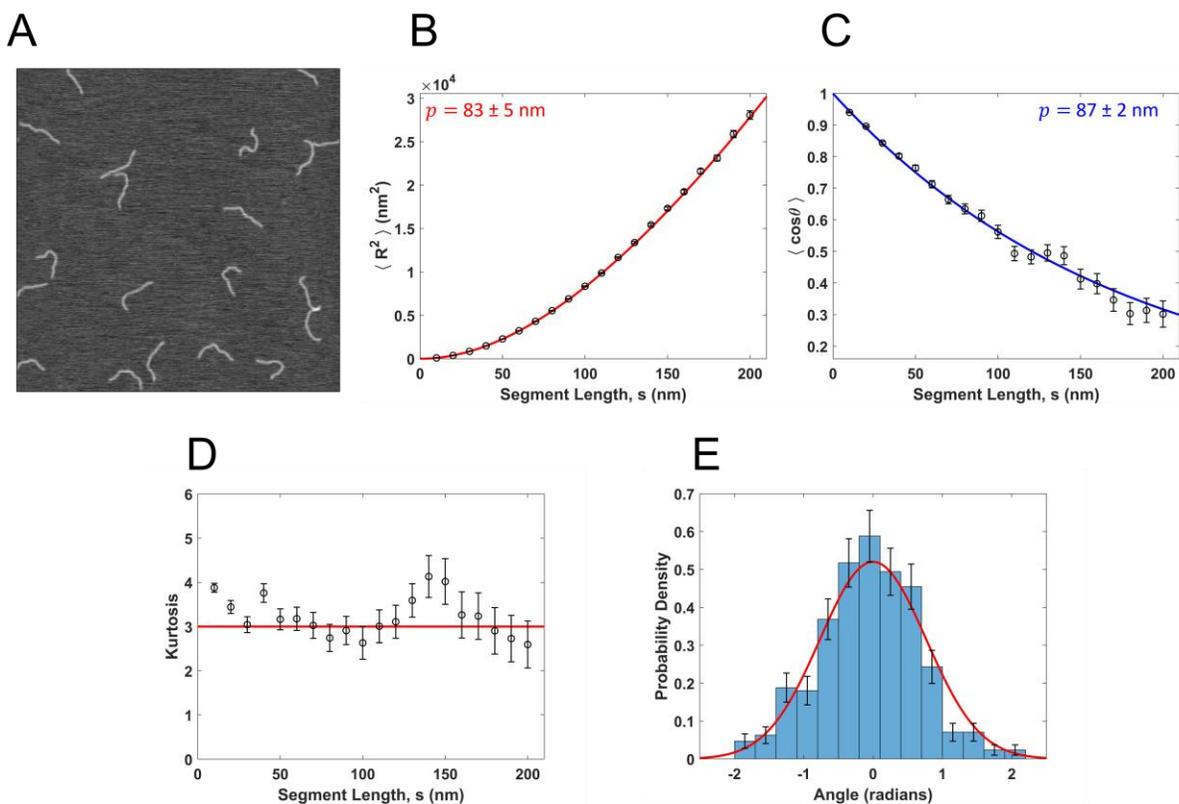
Figure S2. Test of SmarTrace analysis procedures using simulated worm-like chains. (A) Example image of simulated worm-like chains with a persistence length of $p = 85$ nm. After being traced, the data were fit with the standard WLC expressions for mean squared end-to-end distance (Eq. 1) and mean tangent vector correlation (Eq. 2), as shown in (B) and (C), respectively. Both fits yield persistence lengths consistent with the value input into the simulations. (D) Kurtosis of the angular distributions extracted from the simulated chains at different segment lengths. A kurtosis near the expected value of 3 is achieved at all length scales. Errors in the kurtosis are determined using Eq. S19. (E) Individual angular distributions are also well described as Gaussian, as shown by good agreement between the expected and extracted angular distributions at a 50 nm segment length.
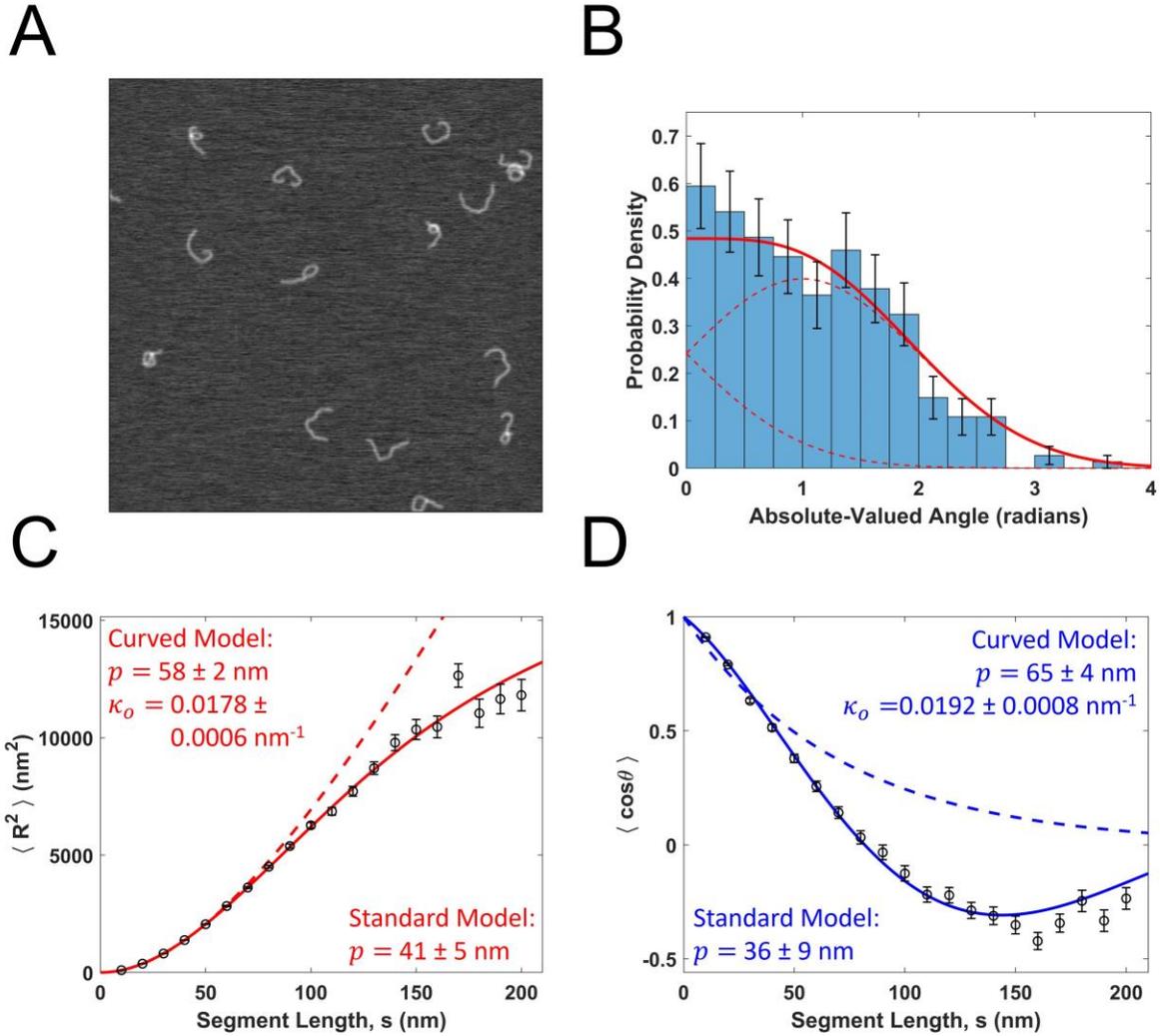
Figure S3. Analysis of simulated curved worm-like chains. (A) Example image of simulated curved worm-like chains with a persistence length of $p = 50$ nm and curvature of $\kappa_o = 0.02$ nm^{-1}. (B) Histogram of the absolute-valued angles extracted from the traces of these simulated chains at a segment length of $s = 50$ nm, as well as the expected probability distribution (solid red line, Eq. S20) and the two Gaussian components that comprise it. (C, D) Fits of data from these simulated chains to the curved worm-like chain expectations for mean squared end-to-end distance (Eq. 3) and mean tangent vector correlation (Eq. 4), respectively. Both fits yield parameters close to those input in the chain simulations.
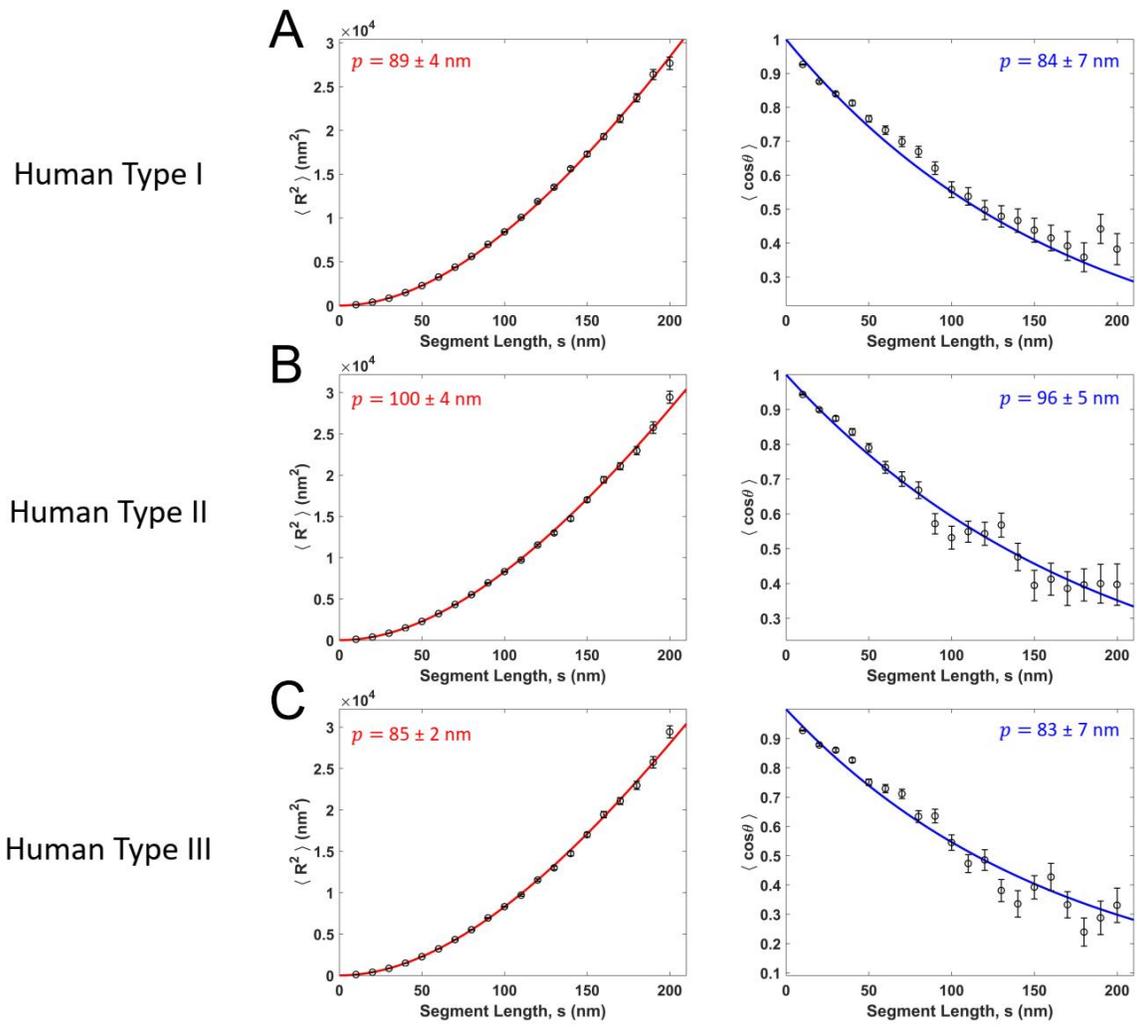
Figure S4. Standard WLC model fits to human collagens (A) type I, (B) type II, and (C) type III, all deposited from 100 mM KCl with 1 mM HCl. $\langle R^2 \rangle$ fits (Eq. 1) are shown in the left column in red, and $\langle \cos \theta \rangle$ fits (Eq. 2) are shown in the right column in blue.
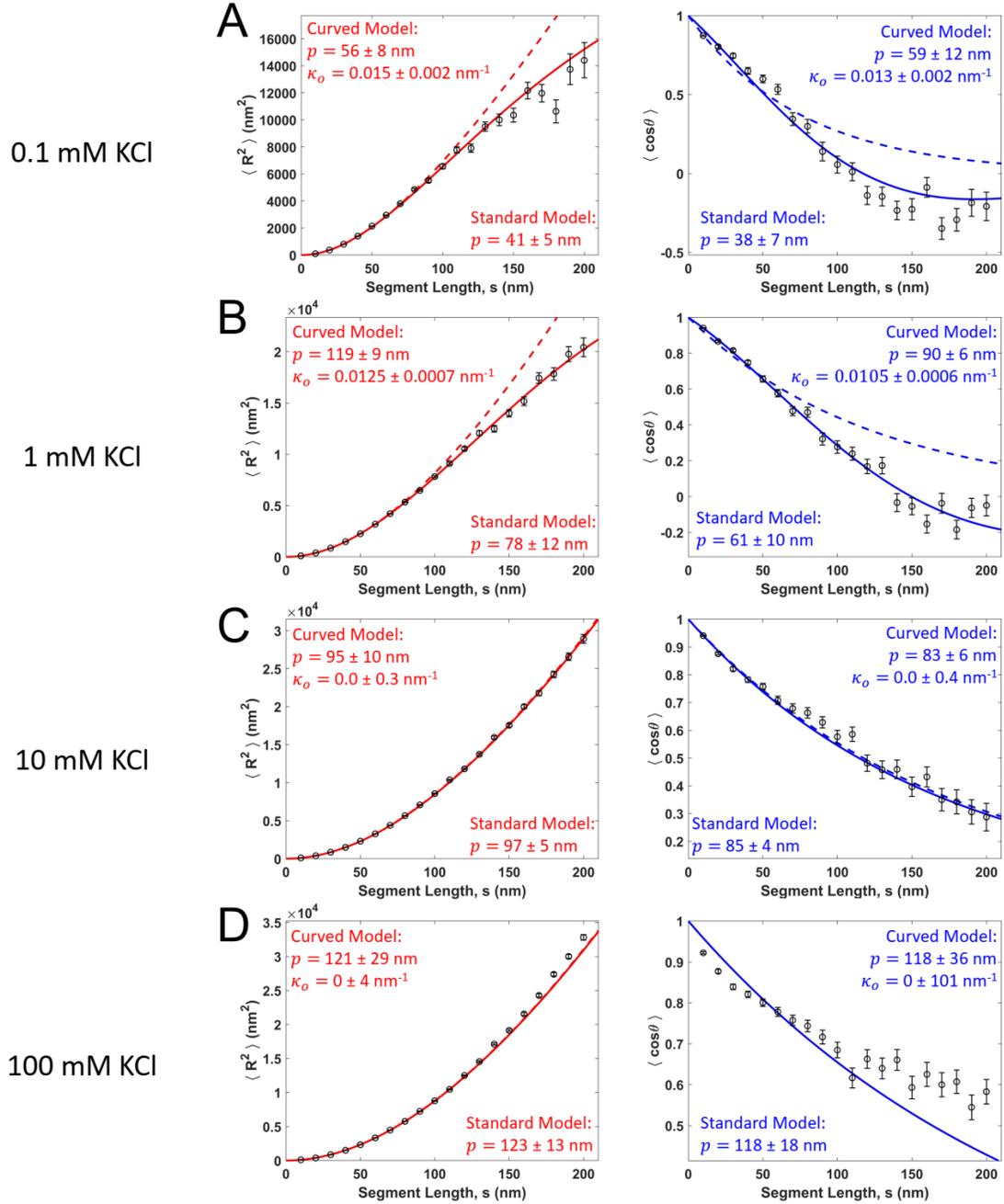
Figure S5. Standard (dashed lines) and curved (solid lines) WLC model fits to rat tail collagen type I data deposited from (A) 0.1 mM KCl, (B) 1 mM KCl, (C) 10 mM KCl and (D) 100 mM KCl. $\langle R^2 \rangle$ fits are shown in the left column in red, and $\langle \cos \theta \rangle$ fits are shown in the right column in blue. As the KCl concentration is increased, the standard and curved fits converge to the same result.
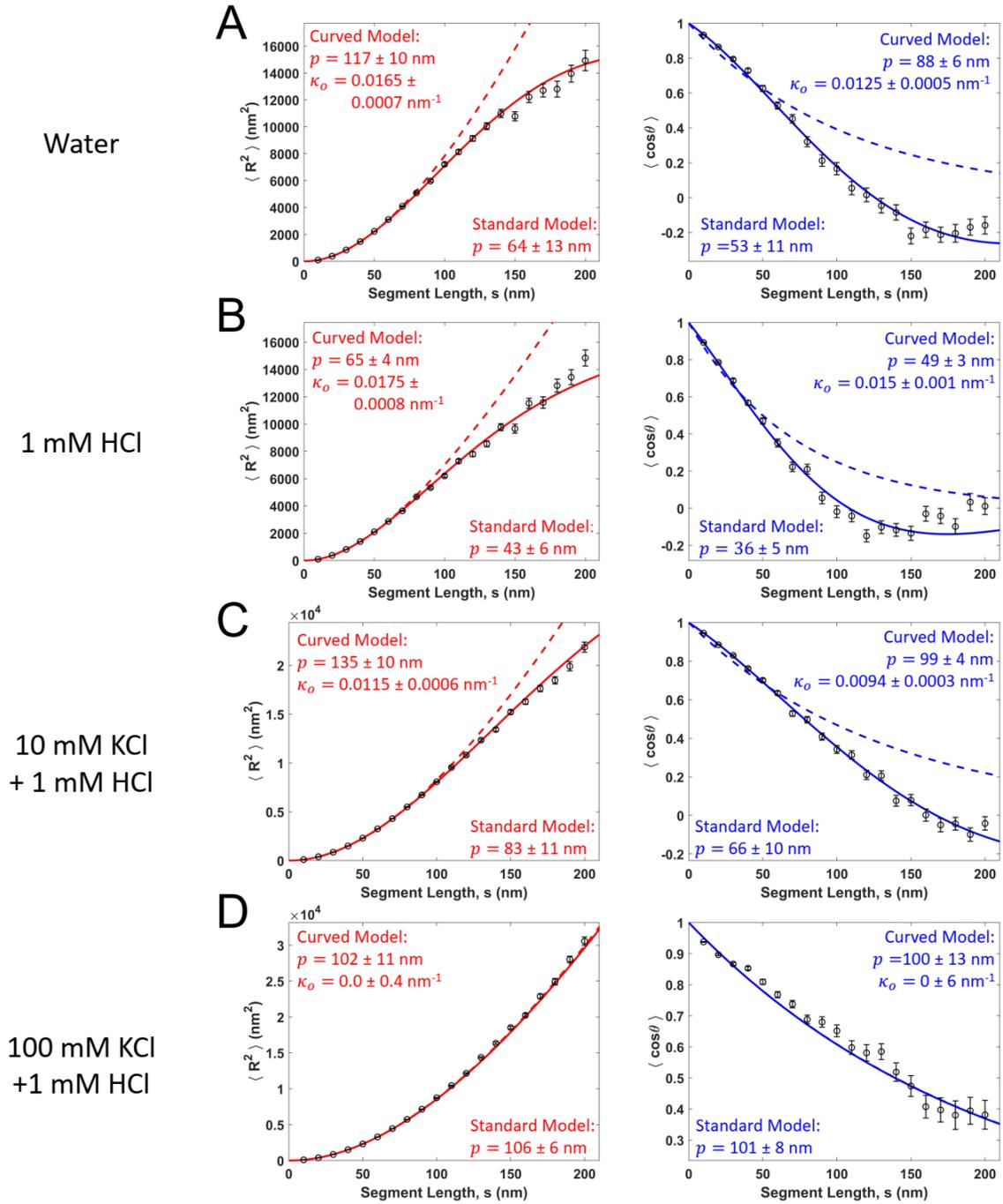
Figure S6. Standard (dashed lines) and curved (solid lines) WLC model fits to rat tail collagen type I data from (A) water, (B) 10 mM KCl + 1 mM HCl and (C) 100 mM KCl + 1 mM HCl. $\langle R^2 \rangle$ fits are shown in the left column in red, and $\langle \cos\theta \rangle$ fits are shown in the right column in blue.
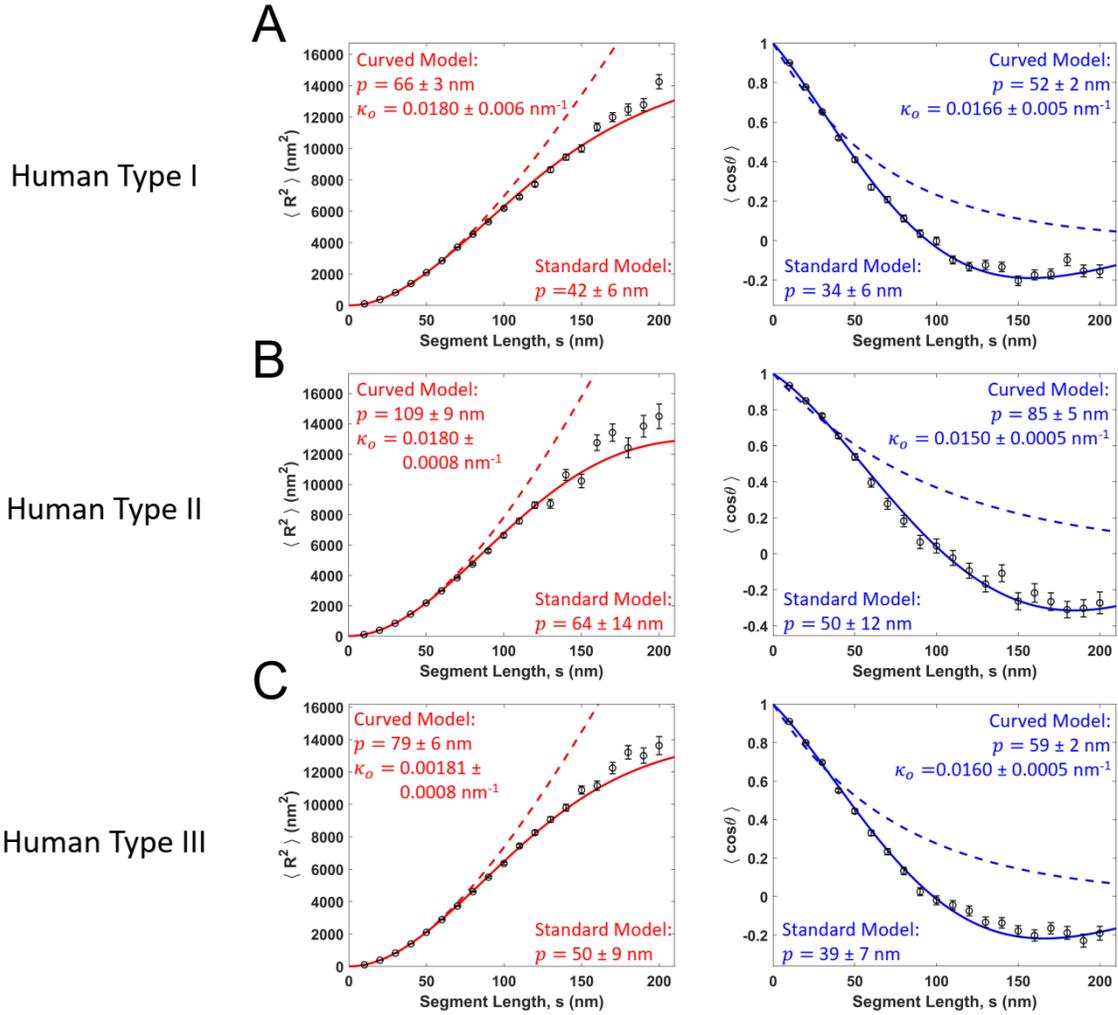
Figure S7. Standard (dashed lines) and curved (solid lines) WLC model fits to (A) type I, (B) type II and (C) type III human collagens deposited from 20 mM acetic acid. $\langle R^2 \rangle$ fits are shown in the left column in red, and $\langle \cos \theta \rangle$ fits are shown in the right column in blue.

Supporting References

1. Landau, L. D., L. P. Pitaevskii, A. M. Kosevich, and E. M. Lifshitz. 1986. Theory of Elasticity. Butterworth-Heinemann.
2. Rivetti, C., M. Guthold, and C. Bustamante. 1996. Scanning Force Microscopy of DNA Deposited onto Mica: Equilibration versus Kinetic Trapping Studied by Statistical Polymer Chain Analysis. Journal of Molecular Biology 264:919-932.
3. Rezaei, N. 2016. Mechanical Studies of Single Collagen Molecules Using Imaging and Force Spectroscopy. Simon Fraser University, Burnaby, Canada.
4. Lamour, G., J. Kirkegaard, H. Li, T. Knowles, and J. Gsponer. 2014. Easyworm: an open-source software tool to determine the mechanical properties of worm-like chains. Source Code for Biology and Medicine 9:16.
5. MATLAB and Statistics Toolbox Release 2017b. The MathWorks, Inc., Natick, Massachusetts, United States.
6. Sonka, M., V. Hlavac, and R. Boyle. 2014. Image Processing, Analysis and Machine Vision. Cengage Learning.
7. Faas, F. G. A., B. Rieger, L. J. van Vliet, and D. I. Cherny. 2009. DNA Deformations near Charged Surfaces: Electron and Atomic Force Microscopy Views. Biophysical Journal 97:1148-1157.
8. Murugesapillai, D., S. Bouaziz, L. J. Maher, N. E. Israeloff, C. E. Cameron, and M. C. Williams. 2017. Accurate nanoscale flexibility measurement of DNA and DNA-protein complexes by atomic force microscopy in liquid. Nanoscale 9:11327-11337.
9. Cramer, D. 1997. Basic statistics for social research : step-by-step calculations and computer techniques using Minitab. Routledge, London; New York.
10. Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer.