Electronic Supplementary Material (ESI) for Soft Matter.

Supporting Material: The spatial-temporal characteristics of type I collagen-based extracellular matrix

Chris Jones,[*] Long Liang[†], Daniel Lin[*], Yang Jiao [†,‡], Bo Sun[*1]

[1*]*Department of Physics, Oregon State University, Corvallis, OR, 97331;* [†]*Department of Physics, Arizona State University, Tempe, AZ, 85287;* [‡]*Materials Science and Engineering, Arizona State University, Tempe, AZ 85287*

S1. THE TEMPERATURE FLUCTUATION DURING GELATION

To monitor the exact temperature of the self-assembling collagen matrix, we used a thermo sensor in contact with the microwells next to the gel sample (main text, Materials and Methods). The typical recordings are shown in Fig. S1. After the initial temperature drop due to the storage temperature of reagents, our setup maintained the gelation temperature within $\pm 0.3^{\circ}C$.

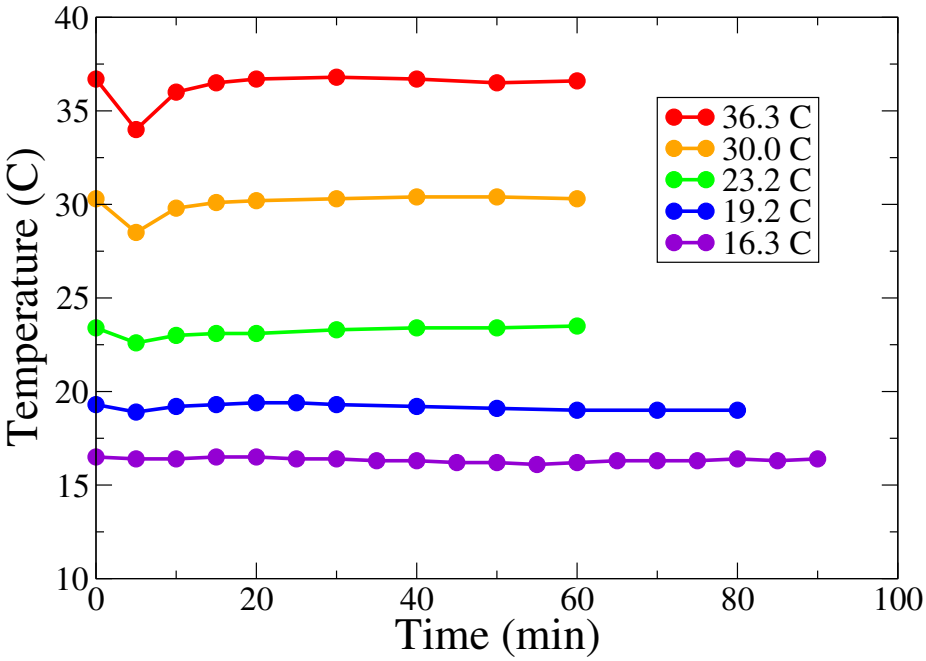

FIG. S1: The typical temperature recordings during gelation.

S2. MORE DETAILED IMAGING OF FIBER CLUSTERS IN COLLAGEN GEL

The spatial heterogeneity of collagen gel is closely related with the existence of fiber clusters – bundles of fibers that are aligned closely. Fig. S2(A1-A2) are typical confocal reflection images of collagen gel formed around room temperature, where fiber clusters can be easily identified visually. Fig. S2(B1-B2) offer better details with higher optical magnification and numerical aperture. To help colocalize the zoom-in and zoom-out images, we have introduced low density particles in the gel as markers.

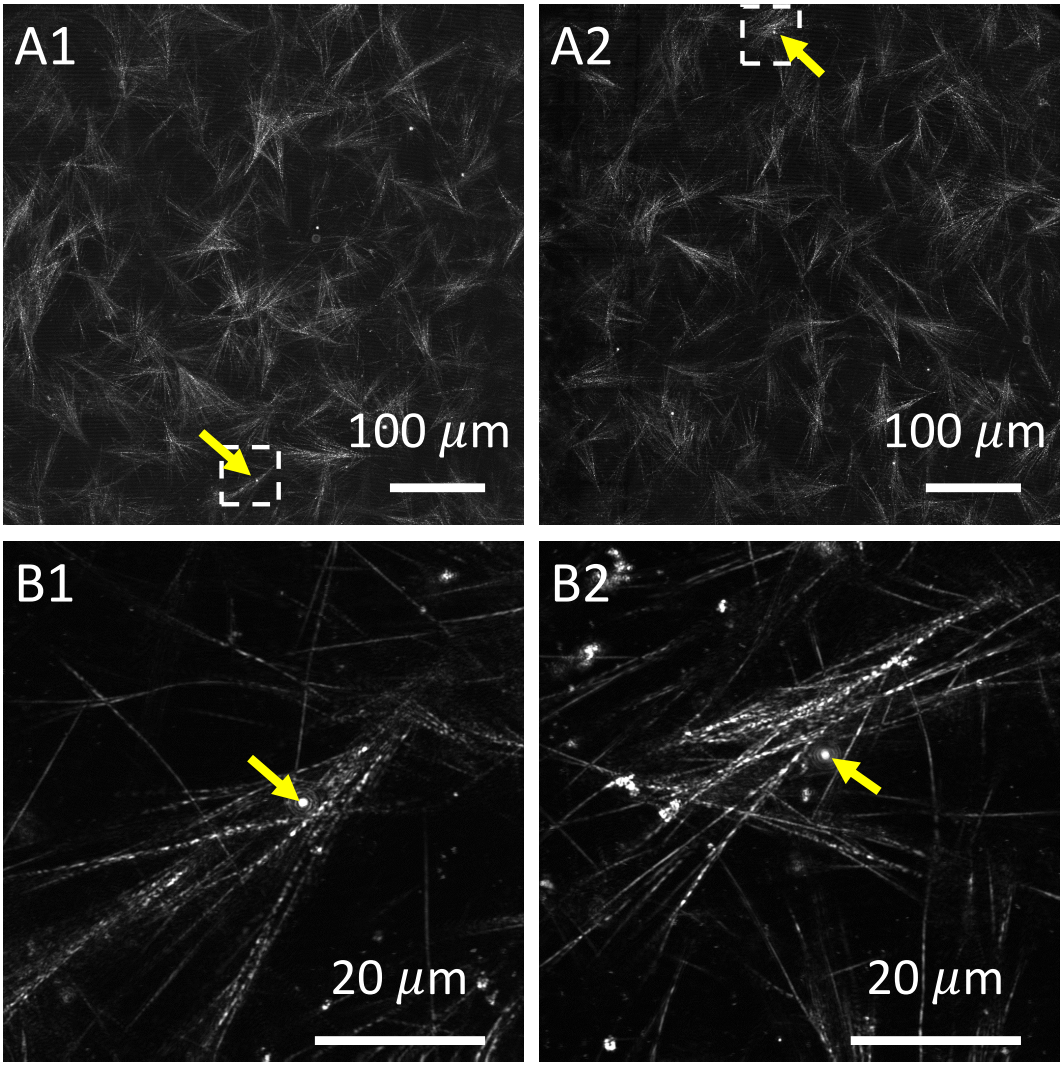

FIG. S2: The fiber clusters contribute to the structural heterogeneity of collagen gel. (A1-B1): A typical confocal reflection image of collagen gel showing many fiber clusters (A1), where one of them (enclosed by white dashed square) is further zoomed in with higher magnification optics. Low density microparticles serve as colocalization markers (yellow arrow). (A2-B2) Another pair of images demonstrate the fiber clusters in collagen gel.

S3. THE TWO POINT DENSITY CORRELATION FUNCTION ENCODES GEOMETRY OF FIBER NETWORKS

The two point correlation functions $g(r)$ and $\Theta(r)$ we calculated from the confocal images of collagen fibers reflect the intrinsic statistics of the collagen microstructure. To further demonstrate the correspondence, we have calculated $g(r)$ for simulated images and evaluate

the curve with double exponential fitting $g(r) = a_1 e^{-l_1 r} + a_2 e^{-l_2 r}$.

The simulated images are generated based on the following protocol:

- Randomly choose nc locations $[x_i, y_i]$ in a 1024 x 1024 bitmap as the centers of clusters. x_i, y_i are drawn from uniform distribution between 1 to $L = 1024$ (chosen to match the experiment image size).

- Draw fc lines centered on each point $[x_i, y_i]$. These lines have Gaussian profile along the transverse direction $e^{-\frac{d^2}{\sigma^2}}$, and each line has a length uniformly chosen from $[ll, \frac{3}{2}ll]$. The orientations of the lines are also uniformly chosen from $[0, 2\pi]$. When two lines intersect, their corresponding intensity add on the bitmap.

Based on extensive examination of the parameter space, we find the following cases most relevant for interpreting the microstructure of collagen networks.

Changing line lengths ll

We have collected the statistics for $nc = 100$, $\sigma = 2$, $fc = 3$ and ll ranging from 20 to 120. For each ll, 50 random sample images are generated. As shown in Fig. S3A, $g(r)$ decays slower at greater ll. The residual of double exponential fitting δg is small and demonstrates damped oscillation. To characterize the shape of δg, we label the first peak-to-peak distance (up to pixel resolution) as w_1. A typical simulation image is shown in the inset of Fig. S3A. Fig. S3B characterizes the fitting results. As can be seen, l_1 is not sensitive to ll, while l_2 grows linearly with ll, suggesting that l_2 measures the size of each cluster ($\propto ll$), while l_1 measures the other length scale in the image: the thickness of the lines σ (see further validation below). This is consistent with the shift of relative weight between a_1 and a_2: for longer lines, statistics for l_1 get more contribution along the line.

Changing line thickness σ

We have collected the statistics for $nc = 100$, $ll = 60$, $fc = 3$ and σ ranging from 1 to 3. For each σ, 50 random sample images are generated. As shown in Fig. S4A, $g(r)$ decays slower at greater σ. A typical simulation image is shown in the inset of Fig. S4A. The double exponential fitting is again a good fit with small residuals δg. Fig. S4B characterizes

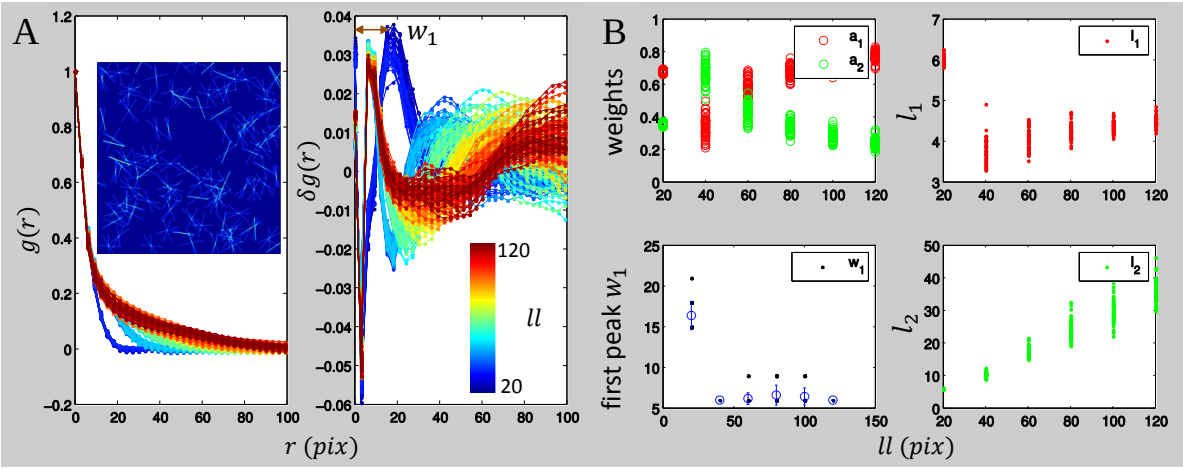

FIG. S3: The density correlation function for simulated ideal line assembly based on the protocol described in the text. $nc = 100$, $\sigma = 2$, $fc = 3$. For each ll, 50 random samples are generated. (A) The two point density correlation function $g(r)$ and the residuals from double exponential fitting $\delta g(r)$. Inset: a typical simulated image. The curves are color coded based on the corresponding line length. (B) The results of nonlinear curve fitting $g(r) = a_1 e^{-l_1 r} + a_2 e^{-l_2 r}$ for each random sample. w_1 represents the first peak-to-peak distance in the residual (arrow in (A)).

the fitting results. As can be seen, l_1 linearly increases with σ, while l_2 is not sensitive to σ, the exact behavior we would expect if l_1 measures the density fluctuation transverse to the line and l_2 measures density fluctuation across each cluster. As σ increases, we would also expect the relative weight of l_1 to decrease following a decrease in the aspect ratio of each line object. This is consistent with the simulation result in Fig. S4B too.

Changing density of clusters $\frac{nc}{L^2}$

We have collected the statistics for $ll = 60$, $\sigma = 2$, $fc = 3$ and nc ranging from 20 to 120 (L is fixed at 1024, the linear size of simulated bitmaps as before). For each nc, 50 random sample images are generated. As shown in Fig. S5, the correlation function is stable as the density of cluster increases by 6 fold. This is consistent with our interpretation of $g(r)$ discussed above, and holds for a wide range of nc until significant overlapping between clusters happen ($nc \cdot ll^2 >> 2L^2$).

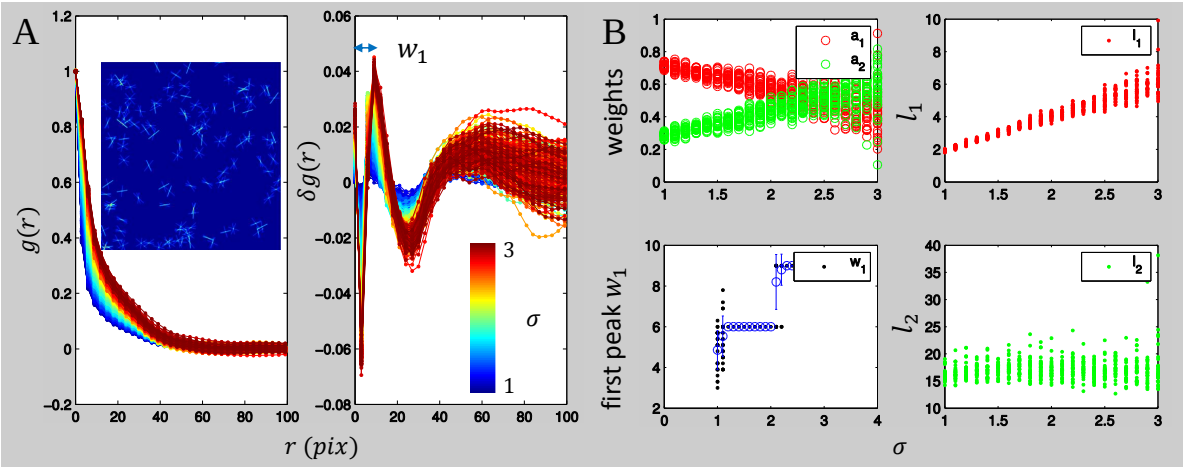

FIG. S4: The density correlation function for simulated ideal line assembly based on the protocol described in the text. $nc = 100$, $ll = 60$, $fc = 3$. For each σ, 50 random samples are generated. (A) The two point density correlation function $g(r)$ and the residuals from double exponential fitting $\delta g(r)$. Inset: a typical simulated image. The curves are color coded based on the corresponding line length. (B) The results of nonlinear curve fitting $g(r) = a_1 e^{-l_1 r} + a_2 e^{-l_2 r}$ for each random sample. w_1 represents the first peak-to-peak distance in the residual (arrow in (A)).

S4. THE INFORMATION CONTENTS OF DIFFERENT STRUCTURAL DESCRIPTORS

In this section, we discuss the relations of three structural descriptors: the density and nematic two-point correlation functions $g(r)$ and $\Theta(r)$, and the average pore size of the network.

First we will show that $g(r)$ is sensitive to the average pore size of the network.

The pore-size distribution $p(\delta)$ widely used in collagen gel characterization can be computed by randomly selecting a point in the pore space and register the radius of largest sphere center at the selected point entirely in the pore space. After a sufficiently large number of sphere radii are registered, one bins the radii to generate a histogram, which represents $p(\delta)$. The average pore size is then defined as:

$$\delta_0 = \int p(\delta)\delta d\delta. \tag{1}$$

Figure S6 shows the plot of the length scale $l2$ in $g(r)$ and the average pore size δ_0 for different collagen network. One can clearly see the positive correlation between these two

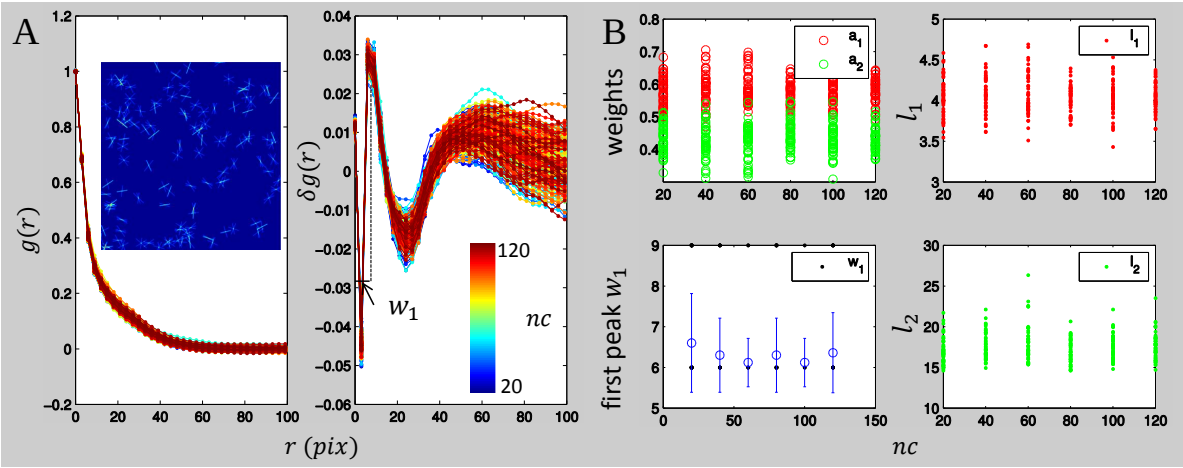

FIG. S5: The density correlation function for simulated ideal line assembly based on the protocol described in the text. $ll = 60$, $\sigma = 2$, $fc = 3$ and nc ranging from 20 to 120. For each nc, 50 random sample images are generated. (A) The two point density correlation function $g(r)$ and the residuals from double exponential fitting $\delta g(r)$. Inset: a typical simulated image. The curves are color coded based on the corresponding line length. (B) The results of nonlinear curve fitting $g(r) = a_1 e^{-l_1 r} + a_2 e^{-l_2 r}$ for each random sample. w_1 represents the first peak-to-peak distance in the residual (arrow in (A)).

quantities. We believe the reason is that the pore size is positively correlated with fiber length and is on the same magnitude of the fiber length, which is manifested as the larger length scale in $g(r)$ characterized by $l2$.

Next, we will show that while both $g(r)$ and $\Theta(r)$ demonstrate consistent trends for collagen gel formed at varying temperatures, in general, they encode complementary information of a network. To demonstrate this, we have generated paired simulation networks with similar $g(r)$ but distinct $\Theta(r)$. In below, we will report both 2D and 3D versions of the comparison.

In the case of 2D networks, we slightly modified the algorithm described in section S2. For a given set of fc, ll, σ and cluster locations, we let the orientational angles of fibers in each cluster to be randomly chosen between $[\beta_1, \beta_2]$. By controlling $\beta_2 - \beta_1$, the nematic order of the simulated networks can be tuned.

We have generated two paired networks that share the same fc, ll, σ and cluster locations, but very different nematic orders as controlled by $\beta_2 - \beta_1$. A typical example is shown in

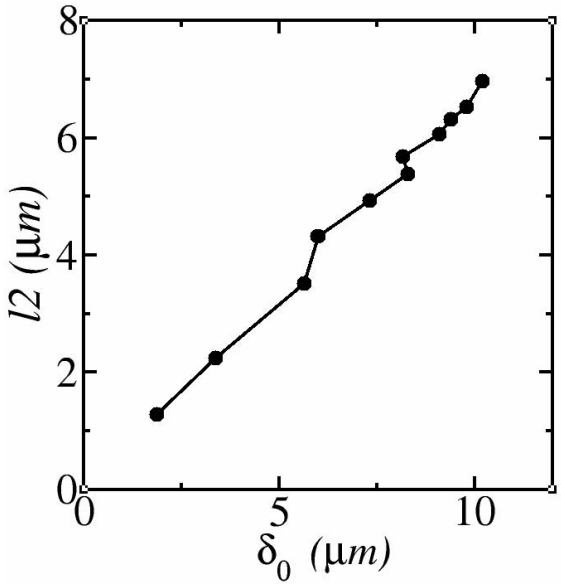

FIG. S6: Positive correlation of the length scale $l2$ and average pore size δ_0 for different collagen network.

Fig. S7. Network on the top left panel is generated with $\beta_1 = 10$, $\beta_2 = 70$. Network on the top right panel is generated with $\beta_1 = 10$, $\beta_2 = 170$. These two networks have almost identical density correlation functions $g(r)$, while the nematic correlation functions $\Theta(r)$ differ significantly.

The same effect can be illustrated in 3D networks. Following the same principle of the 2D case, we used inverse stochastic reconstruction to generate 3D network structures with similar $g(r)$ but distinct $\Theta(r)$ as shown in Fig. S8.

The examples clearly shows that while $g(r)$ is sensitive to the fiber thickness and cluster size, $\Theta(r)$ is also determined by the orientational order of the fibers. To characterize a fibrous network, it is necessary to take into account of both $g(r)$ and $\Theta(r)$.

S5. EFFECTS OF SIMULATED FIBER INTENSITY ON THE TWO POINT STATISTICS

In the virtual collagen network microstructure obtained from the kinetic Monte Carlo growth algorithm, each fiber is represented by a spherocylinder with a definitive radius. In the experimental image, the network structure is represented by a continuous grayscale intensity map, in which the high-intensity regions correspond to the fibers. As indicated in

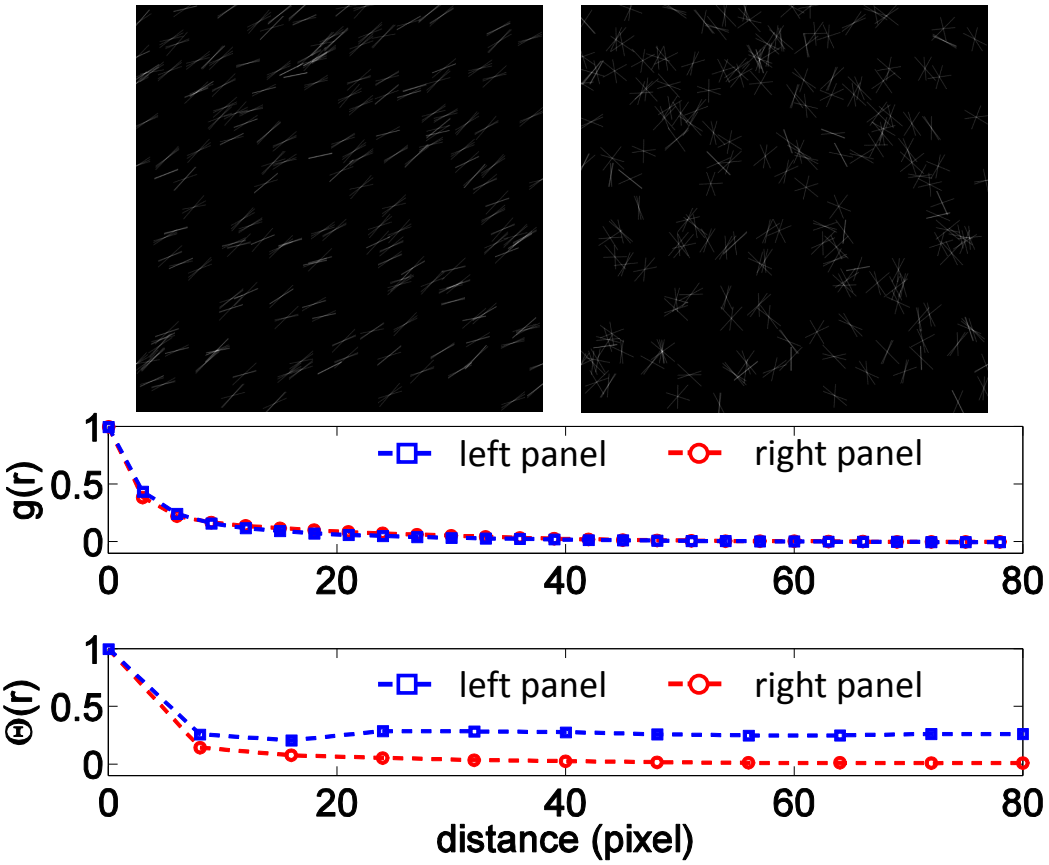

FIG. S7: Top panels: simulated networks with $ll = 60$, $\sigma = 1$, $fc = 3$, $nc = 180$. Network on left panel is set with higher nematic order $[\beta_1, \beta_2] = [10, 70]$. Network on the right panel is set with a lower nematic order $[\beta_1, \beta_2] = [10, 170]$. Bottom panels: the density and nematic correlation functions calculated from the images on the top panels.

the main paper, in order to quantitatively compare the experimental and simulated collagen network structures, the simulated network is also converted to an intensity map as follows:

- (i) The virtual collagen network microstructure (i.e., a "packing" of spherocylinder in a simulation box) is discretized into a 3D binary array. If an entry of the array is associated with a spherocylinder (i.e., a fiber), its value is set to be 1; otherwise, it is set to be 0.

- (ii) The 3D binary array presenting the simulated network structure is then convoluted with a Gaussian kernel, i.e., $f(r) = \exp(-r^2/\sigma^2)$, where σ is the width of the kernel. The convolution operation generates a grayscale image that mimics the experimentally obtained fiber intensity map.

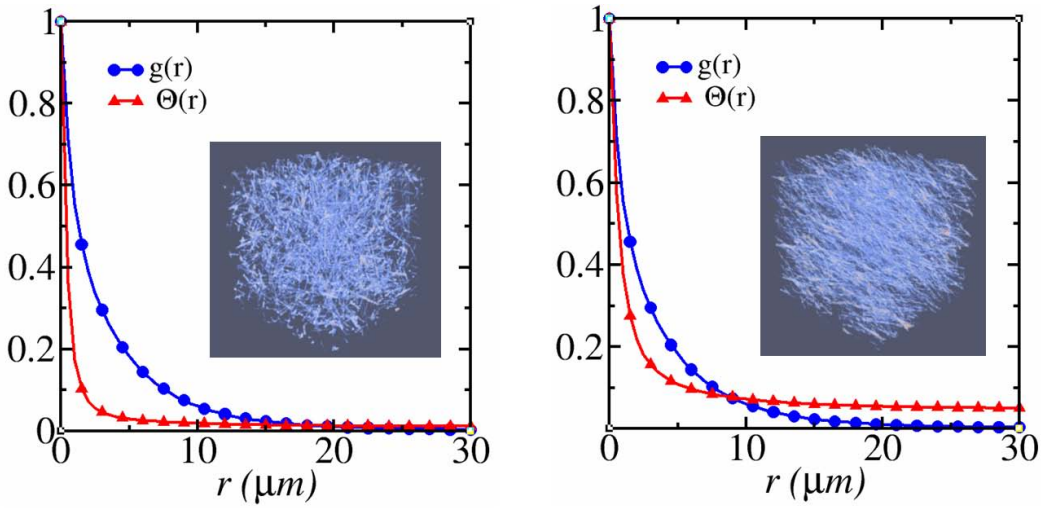

FIG. S8: Constructed collagen structures that possess similar $g(r)$ but distinct $\Theta(r)$, suggesting that these two quantities contain independent structural information.

The two-point statistics of the convoluted image can be then computed and compared to the experimental results. It is clear that the width of the Gaussian kernel σ needs to be carefully chosen such that the statistics associated with the convoluted image can be compared with experimental data on the same footing. Here, we study the effects of the kernel width σ on the two-point intensity correlation function for the simulated network.

In particular, we focus on a simulated network with $n = 1000$ fibers and average fiber length $\bar{L} = 0.43$. The simulated intensity maps for the network with various σ values are shown in the left panel of Fig. S9. It can be seen that increasing kernel width essentially leads to "thicker" fibers. The right panel of the figure shows the two-point intensity correlation functions for the simulated intensity maps with different σ values, which indicates that larger σ values correspond to a longer correlation length. This result is consistent with our study of the effect of fiber thickness on the two-point statistics presented in Sec. S2.

S6. EQUIVALENCE OF THE TWO POINT DENSITY CORRELATION FUNCTION OBTAINED FROM 3D FIBER NETWORKS AND 2D SLICES

In the main paper, we compare the two-point density correlation functions obtained from 2D experimental images and 3D simulated collagen structures. In this section, we provide justification of such comparison.

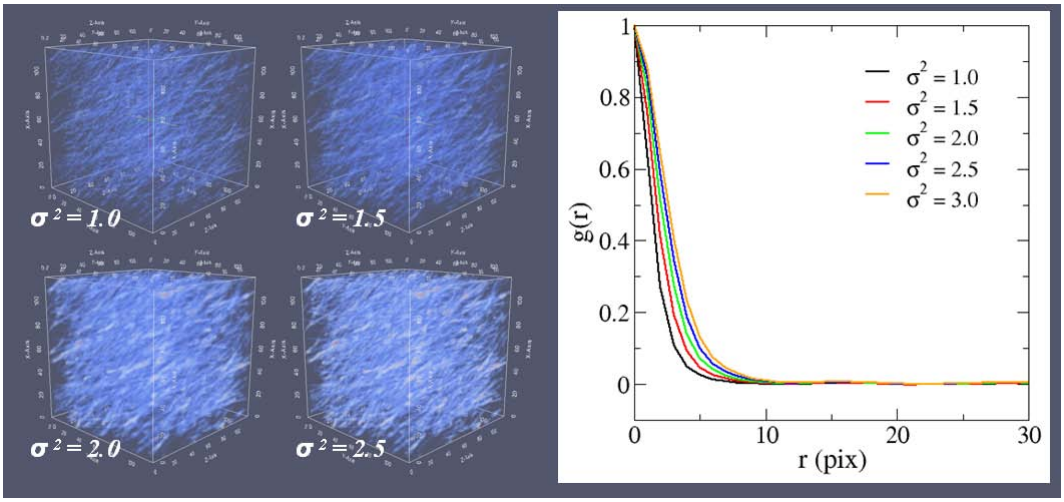

FIG. S9: The effects of Gaussian kernel width on the two-point correlation function $g(r)$ computed from 3D simulated intensity maps. Left panel: Simulated intensity maps for a collagen network (with $n = 1000$ fibers and average fiber length $\bar{L} = 0.43$) associated with various σ values. The unit of the fiber length is the edge length of the cubic simulation box. Right panel: Two-point intensity correlation functions for the simulated intensity maps with different σ values.

It is well established in statistical geometry that if a heterogeneous material (e.g., a collagen network composed of fiber phase and interstitial fluid phase) is statistically homogeneous (i.e., the material microstructure does not contain a preferred center) and isotropic (i.e., the microstructure is rotationally invariant), the two-point correlation function associated with the phase of interest for the 3D material can be obtained from its 2D slices. This is because the correlation function provides statistics of pair distances, which are scalar quantities. Thus, the statistics obtained from a sufficient number of 2D slices or a single 2D slice that is sufficiently large are representative of those obtained from the 3D structure.

We computed the two-point density correlation function $g(r)$ for different simulated collagen networks both from the full 3D microstructure and 2D slices. Specifically, we use $n_s = 10$ successive 2D slices to compute $g(r)$. This is to mimic a 2D experimental image which contains collagen fibers within a thin but finite layer. Figure S10 shows comparison of $g(r)$ obtained from 3D structures and associated 2D slices along different orthogonal planes for different simulated collagen networks. It can be clearly seen that the $g(r)$'s are virtually identical to one another, which indicates that it is equivalent to use the statistics obtained from either 2D or 3D simulated collagen structure for comparison with the 2D experimental

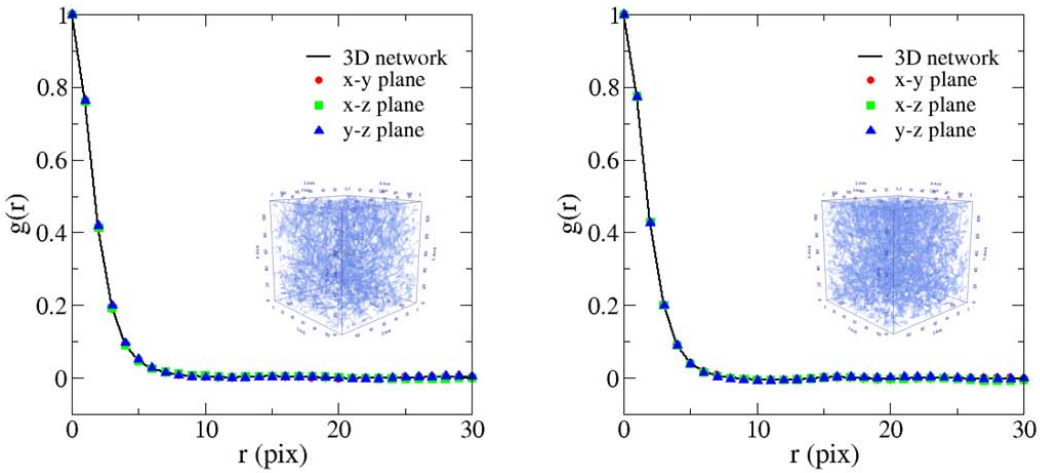

FIG. S10: Comparison of the two-point density correlation function $g(r)$ computed from 3D structure and 2D slices of different simulated collagen networks. Left panel: A simulated collagen network with $n = 1200$ fibers and average fiber length $\bar{L} = 0.32$. Right panel: A simulated collagen network with $n = 3500$ fibers and average fiber length $\bar{L} = 0.13$. The unit of the fiber length is the edge length of the cubic simulation box.

data.

S7. TEMPERATURE DEPENDENCE OF THE FITTING PARAMETERS FOR THE DENSITY CORRELATION FUNCTION AND GLOBAL NEMATIC ORDER FROM SIMULATIONS

In the main paper, we have shown the simulated density correlation functions $g(r)$ and nematic correlation functions $\Theta(r)$ associated with different collagen gels as temperature varies. Here, we provide the fitting parameters a_1, a_2, l_1 and l_2 for $g(r)$ as well as the global nematic order parameter $< s >$ as a function of gelation temperature T.

As shown in Fig. S11(A), the parameter a_1 (a_2) increases (decreases) as T increases, indicating that structures at smaller length scales play a more important role in determining the two-point statistics as the network becomes more homogeneous as the gelation temperature increases, which is consistent with experimental results. Fig. S11(B) and (C) respectively show the two length-scale parameters l_1 and l_2 associated with the network structure as a function of T. Both length scales decay as T increases, indicating a transition of gel structure

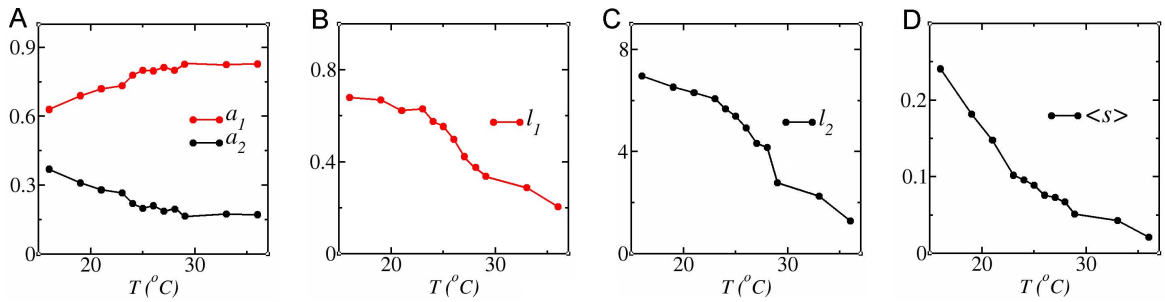

FIG. S11: The effect of gelation temperature T on the fitting parameters a_1, a_2, l_1 and l_2 for $g(r)$ as well as the global nematic order parameter $< s >$.

from a "bundled" network to a homogeneous one. Such a transition is also clearly manifested in the global nematic order parameter, which is monontically decaying as T increases as shown in Fig. S11(D).

S8. EFFECTS OF HYDRODYNAMICAL EFFECTS ON THE SIMULATED COLLAGEN STRUCTURE

The hydrodynamic effects are added to facilitate the formation of fiber bundles (with similar orientation) as observed in the experiments. It is implemented as follows: if a fiber is longer than a critical value, its rotational motion is not random any more, but is biased to align with a nearby long fiber. Specifically, the final orientation of the moved fiber is obtained by perturbing the orientation of the nearby "reference" long fiber. When the fibers are sufficiently long and at high densities, the exclusion volume effects can also impose orientational order. However, we found that without hydrodynamic effects as implemented in the model, the orientational order achieved by exclusion volume effects is not sufficient when comparing to experimental data. Figure S12 shows two simulated collagen networks with the same model parameters except that one of them does not consider the biased moves due to hydrodynamical effects. It can be clearly seen that the network with hydrodynamical effects possesses a higher degree of nematic order, which is consistent with experimental data.

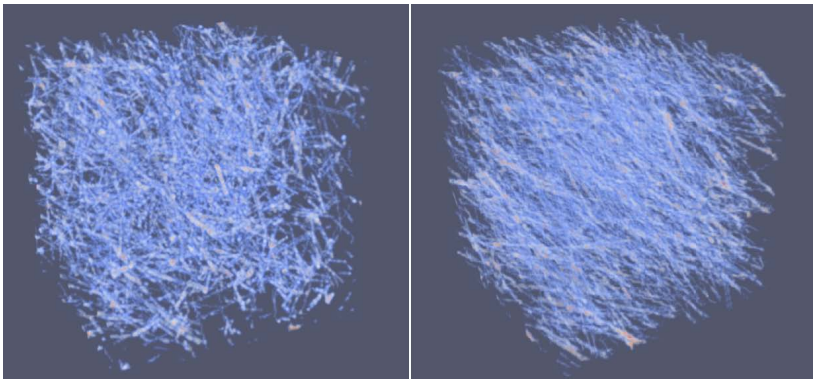

FIG. S12: Simulated collagen networks with same model parameters except that one of them does not consider the biased moves due to hydrodynamical effects (left panel). It can be clearly seen the network with hydrodynamical effects (right panel) possesses a higher degree of nematic order, which is consistent with experimental data.

S9. COMPARISON WITH OTHER MOLECULAR LEVEL MODELS

We note that the temperature dependence of the nucleation rate and growth rate has been investigated in previous works based on molecular level modeling. For example, as reported by Wood (*Biochem. J* (1960) **75**,598), the nucleation rate k_1 and growth rate k_3 were introduced as concentration independent parameters in the governing equation for number of nuclei [Eq. (2)] and fiber mass [Eq.(4)]. The temperature dependence of a new combined parameter $K_1 = k_1(k_3)^3$, characterizing the extent of precipitation, was obtained for selected temperatures. Although a general mathematical form of the dependence was not given in Wood's paper, a dramatic increase of K_1 as T increases can be clearly seen in Table 3, which is presumably exponential in form (the number of data points was not large enough to make definitive conclusions).

The growth rate and nucleation rate defined in the current work distinguish themselves form those in Wood's work in two aspects: (1) they are parameters for discrete particle-growth model instead of those in continuum governing equations; (2) they characterize the "growth" on a coarse-grained scale while the parameters in Wood's work are associated with molecular systems. Therefore, we feel that a quantitative comparison would lead to limited insights. Qualitatively, the parameters from both works appear to exhibit an exponential dependence on T.