Supplementary Information for "Optimizing gelation time for cell encapsulation through active learning"

Yuxin Luo, Juan Chen, Mengyang Gu, Yimin Luo

Note 1: Procedure to remove the convective contribution

Towards the later stage of gelation when probe displacements are extremely small, varying rates of crosslinking within the sample may lead to locally uneven stresses and a convective tail at large Δt for the MSD. Hence, these MSDs were modeled and corrected using the Ornstein–Uhlenbeck (O-U) function with a convective term [3]:

$$\langle \Delta r^2(\Delta t) \rangle = 4\sigma_s^2 - 4\sigma_s^2 \rho^{\Delta t} + \mu_D^2 \Delta t^2 = 4\sigma_s^2(1 - \rho^{\Delta t}) + \mu_D^2 \Delta t^2 \tag{1}$$

This function describes the MSD as a combination of a constant term, a decaying exponential, and a squared convective term. μ_D represents the velocity. To remove the convective contribution, we fit the gel branch of the MSDs to the parametrized form of MSD in Eq. 1, then subtract the part representing convective term from the MSD, an example of the processed MSD is shown in Fig. S1. Figure S1 shows a comparison of the MSD curves before and after removing the convective term, and the subsequent shifting procedure. By modeling the MSD with an O-U process and performing this procedure, the upward tail of the MSD at relatively large lag times is flattened, more accurately reflecting the scenario where tracer particles are fully immobilized after gelation.

Note 2: Ab initio uncertainty quantification (AIUQ) in scattering analysis of microscopy

Here, we briefly reviewed the model and procedure of AIUQ [2] for scattering analysis of microscopy videos. Consider a vectorized microscopy image of $N = N_1 \times N_2$ pixels at time t, where the image intensity vector is denoted by $\mathbf{y}(t) = [y(\mathbf{x}_1, t), ..., y(\mathbf{x}_N, t)]^T$ at time t, where $\mathbf{x}_i = (x_{i,1}, x_{i,2})^T$ is the two dimensional (2D) location of at pixel i in the Cartesian space, for $i = 1, ..., N$. AIUQ imposes a probabilistic model of the image intensity in the original Cartesian space

$$\mathbf{y}(t) = \frac{1}{\sqrt{N}} \mathbf{W}^* \mathbf{z}(t) + \boldsymbol{\epsilon}(t), \tag{2}$$

where the $N \times N$ matrix $\mathbf{W}^*$ is a 2D complex conjugate of the Fourier basis, which relates the N observations of an image at time t from Cartesian space $\mathbf{x} = (x_1, x_2)^T$ to a set of random factor processes $\mathbf{z}(t)$ in the reciprocal space $\mathbf{q} = (q_1, q_2)^T$ and $\boldsymbol{\epsilon}(t) \sim \mathcal{MN}(\mathbf{0}, \frac{\bar{B}}{2}\mathbf{I}_N)$ is an N-dimensional Gaussian white noise vector with variance $\frac{\bar{B}}{2}$ and $\mathbf{I}_N$ being the identity matrix of N dimensions. The N-dimensional complex-valued latent factor $\mathbf{z}(t)$ is split into the real and imaginary parts: $\mathbf{z}(t) = \mathbf{z}_{re}(t) + i\mathbf{z}_{im}(t)$, where $\mathbf{z}_{re}(t) = (z_{1,re}(t), ..., z_{N,re}(t))^T$ and $\mathbf{z}_{im}(t) = (z_{1,im}(t), ..., z_{N,im}(t))^T$ are two N-dimensional vectors.

For isotropic processes, the j'th row of latent factor is assumed to have the same intermediate scattering function (ISF) corresponding to the jth ring of the Fourier transformed image intensity. For any j', the random factor vectors over n time points $\mathbf{z}_{\mathbf{j'},re} = (z_{\mathbf{j'},re}(t_1), ..., z_{\mathbf{j'},re}(t_n))^T$ and $\mathbf{z}_{\mathbf{j'},im} = (z_{\mathbf{j'},im}(t_1), ..., z_{\mathbf{j'},im}(t_n))^T$ are assumed to independently follow multivariate normal distributions $\mathbf{z}_{\mathbf{j'},re} \sim \mathcal{MN}(\mathbf{0}, \frac{A_j}{4}\mathbf{R}_j)$ and $\mathbf{z}_{\mathbf{j'},im} \sim \mathcal{MN}(\mathbf{0}, \frac{A_j}{4}\mathbf{R}_j)$, with $\frac{A_j}{4}\mathbf{R}_j$ being the covariance matrix for $j = 1, ..., J$. The (k_1, k_2)th entry of $\mathbf{R}_j$ is characterized by ISF: $R_j(k_1, k_2) = f_{\boldsymbol{\theta}}(q_j, \Delta t_k)$ with $\Delta t_k = |k_2 - k_1|\Delta t_{min}$ with Δt_{min} being the interval between two consecutive time frames and $f_{\boldsymbol{\theta}}(q_j, \Delta t_k)$ is the ISF with

parameters θ at a scalar Fourier magnitude q_j and lag time point Δt_k, for $j = 1, ..., J$ and Δt_k. Further denote the index set $\mathcal{S}_j$ that satisfies $\mathcal{S}_j = \{(j'_1, j'_2) : q^2_{j'_1,1} + q^2_{j'_2,2} = q^2_j\}$ for $j = 1, ..., J$. For the isotropic Ornstein–Uhlenbeck process assumed herein, the ISF is

$$f_{\boldsymbol{\theta}}(q_j, \Delta t_k) = \exp\left\{-\frac{q_j^2 \sigma_s^2 (1 - \rho^{\Delta t_k})}{4}\right\},$$ (3)

which consists of two parameters $\boldsymbol{\theta} = (\sigma_s, \rho)$.

The parameters of AIUQ include $(\boldsymbol{\theta}, \boldsymbol{A}_{1:J}, \bar{B})$. First for any given $\bar{B}$, we use an unbiased estimator of J amplitude parameters,

$$A_{est,j} = \frac{2}{S_j n} \sum_{\mathbf{j}' \in \mathcal{S}_j} \sum_{k=1}^n |\hat{y}_{\mathbf{j}'}(t_k)|^2 - \bar{B}$$ (4)

where $\hat{y}_{\mathbf{j}'}(t_k)$ is the j'th row of the Fourier transformed intensity $\hat{\mathbf{y}}(t) = \frac{\mathbf{W}\mathbf{y}(t)}{\sqrt{N}} = \hat{\mathbf{y}}_{re}(t) + i\hat{\mathbf{y}}_{im}(t)$ and $S_j = \#\mathcal{S}_j$ is the number of items in the set $\mathcal{S}_j$. Then we estimate the parameters $\{\boldsymbol{\theta}, \bar{B}\}$ by maximizing the logarithm of the marginal likelihood function after integrating out the random factor $\mathbf{Z}$ and with a plug-in estimator of $\mathbf{A}_{est,1:J}$:

$$\log\{\mathcal{L}(\boldsymbol{\theta}, \bar{B})\} = \log\left\{\prod_{j=1}^J \prod_{\mathbf{j}' \in \mathcal{S}_j} p_{MN}(\hat{\mathbf{y}}_{re,\mathbf{j}'}; \mathbf{0}, \boldsymbol{\Sigma}_j) p_{MN}(\hat{\mathbf{y}}_{im,\mathbf{j}'}; \mathbf{0}, \boldsymbol{\Sigma}_j)\right\},$$

$$= C - \sum_{j=1}^J \left\{-S_j \log|\boldsymbol{\Sigma}_j| - \frac{\sum_{\mathbf{j}' \in \mathcal{S}_j} \hat{\mathbf{y}}^T_{j,re} \boldsymbol{\Sigma}_j^{-1} \hat{\mathbf{y}}_{j,re}}{2} - \frac{\sum_{\mathbf{j}' \in \mathcal{S}_j} \hat{\mathbf{y}}^T_{j,im} \boldsymbol{\Sigma}_j^{-1} \hat{\mathbf{y}}_{j,im}}{2}\right\},$$ (5)

where $\boldsymbol{\Sigma}_j = \frac{A_{est,j}}{4} \mathbf{R}_j + \frac{\bar{B}}{4}$ and C is a normalizing constant not relevant to the parameters $(\boldsymbol{\theta}, \bar{B})$.

Then we estimate the model parameter $\boldsymbol{\theta}$ and noise parameter $\bar{B}$ by maximizing the marginal likelihood from Eq. (5):

$$(\boldsymbol{\theta}_{est}, \bar{B}_{est}) = \text{argmax}_{\boldsymbol{\theta}, \bar{B}} \log \mathcal{L}(\boldsymbol{\theta}, \bar{B}).$$ (6)

The AIUQ package [4] implements the generalized Schur algorithm [1, 5] that substantially accelerates the computation in estimation without approximation.

Note 3: Procedure to compute Pearson correlation coefficient r

The covariance between data x and y accounting for the weights w is computed by:

$$\text{cov}(x, y; w) = \frac{\sum_{i=1}^n w_i (x_i - m(x; w))(y_i - m(y; w))}{\sum_{i=1}^n w_i},$$ (7)

where m is the weighted mean:

$$m(x; w) = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}.$$

The weighted correlation coefficient is

$$r = \frac{\text{cov}(x, y; w)}{\sqrt{\text{cov}(x, x; w)\text{cov}(y, y; w)}}.$$ (8)

When computing the unweighted correlation coefficient, all data points are weighed equally and $w_i = 1$.

Additional tables

Table S1 compares the gelation time using manual shifting of the MPT tracking data or using the automated AIUQ analysis. The critical power exponent at the gelation point is also included in this analysis at different concentration and temperature conditions.

Table S2 compares the gelation time determined by microrheology, using time-cure superposition, versus bulk rheology, which uses the G', G" crossover. The gelation time is determined from Fig. S7.

Table S1: Gelation time t_{gel} using MPT and manual shifting versus using automated AIUQ analysis

Concentration (mg/mL)	Temperature (°C)	pH	n	t_{gel} (min) by MPT	t_{gel} (min) by AIUQ
20	25	7	0.74	41.5	43.4
20	35	7	0.73	34.5	33.9
23	28	7	0.74	40.5	44.6
30	27	7	0.72	19.3	21.1
32	30	7	0.77	13.8	15.7
35	25	7.4	0.76	8.5	14.1
30	33	7.4	0.75	28.3	36.4

Table S2: Gelation time t_{gel} using MPT versus bulk rheometry, using the G', G" crossover point.

Concentration (mg/mL)	Temperature (°C)	pH	t_{gel} (min) by MPT	t_{gel} (min) by bulk rheology
20	25	7.4	80	~ 5
30	25	7.4	38	~ 5
35	25	7.4	9.25	< 2
25	30	7.4	65	< 2
30	30	7.4	45	< 2
40	30	7.4	10	< 2

Additional figures

Figure S1 provides an example of de-drift procedure following **Note 1**.

Figure S2-S3 demonstrate the procedure for active learning. Fig. S2 illustrates the procedure of sequential minimization of error by placing the next testing points where the error is the biggest in the previous error assessment. The color denotes the 95% confidence interval length. Fig. S3 illustrates the gelation response surface learned using linear versus Gaussian process regression. The average interval length and root-mean-square errors are also shown for both methods.

Figure S4 shows UV-vis spectroscopy to track the kinetics of TPEG-SG hydrolysis in pH = 7.4 and pH = 7 solvent. TPEG-SG was dissolved in PBS or DI water (at 30 mg/mL) immediately before experiments and placed inside a quartz cuvette with a 2 mm path length, and measured by an ultraviolet-vis Cary 3E Spectrophotometer. The varying rates of hydrolysis lead to differences in learning efficiency between these two conditions.

Figures S5-S7 show bulk rheology plots. Fig. S5 shows the frequency-dependent G', G" for collagen-TPEG double networks after they have been fully crosslinked. Fig. S6 shows the transient G', G" from mixing to about 3 hours after crosslinking, for collagen-TPEG double networks. Finally, Fig. S7 shows the time-dependent G', G" from mixing to when the crossover occurs.

Figure S8-S12 illustrate the procedures for shape analysis. Fig. S8 demonstrates the processing steps for extracting the shape using the best-fit ellipses. Fig. S9 plots the histograms of the aspect ratio distributions across all temperature concentration conditions for the encapsulated cells. Fig. S10 illustrates the procedure of determining the aspect ratio by the gyration tensor or extracting the area and shows that those metrics are highly correlated to the aspect ratio determined by the best-fit ellipse. Fig. S11 compares cell shapes on 2D substrates versus 3D encapsulations. Finally, Fig. S12 illustrates the variability within the cell shape data.
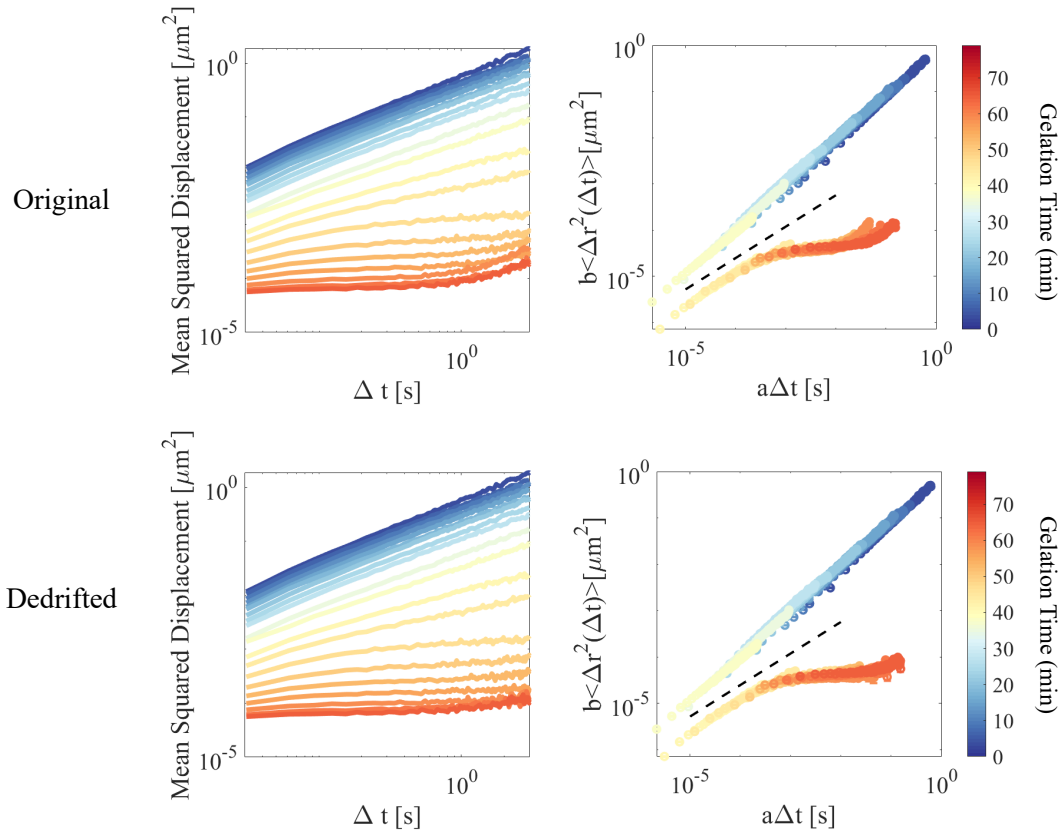

Figure S1: Comparison between MSDs for TPEG formulation at 23 mg/mL, crosslinked at 28°C with and without removing the convective contribution.
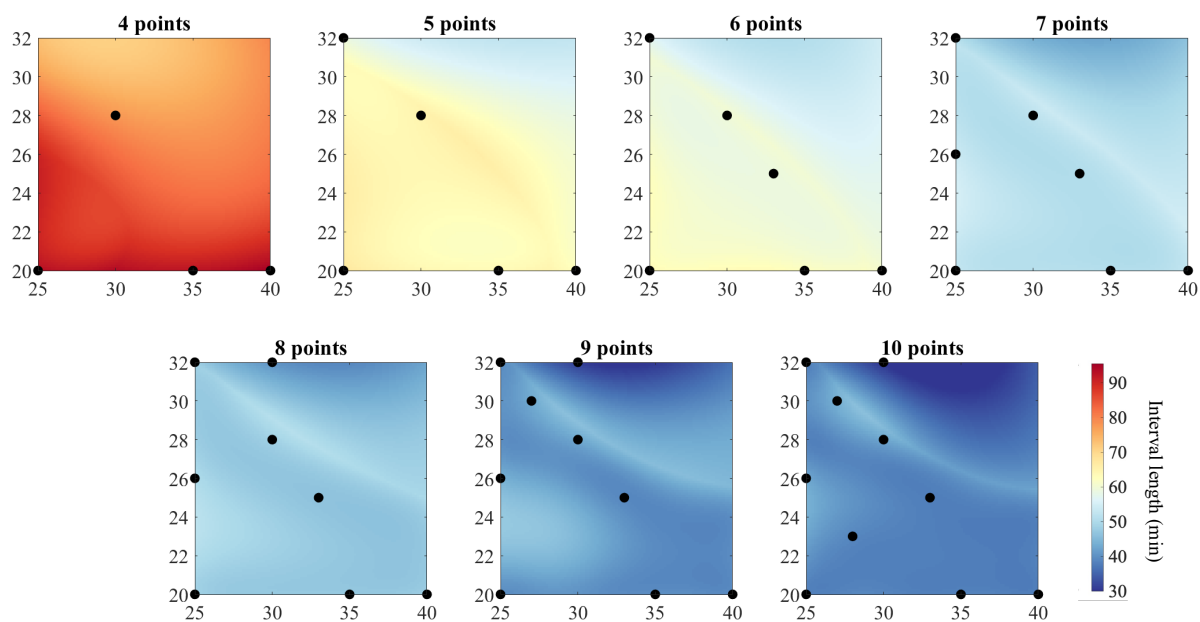
Figure S2: Sequential minimization of error in predicting the response surface: Visual representation of the interval after sequential selection of 4-10 points in the pH = 7 case.
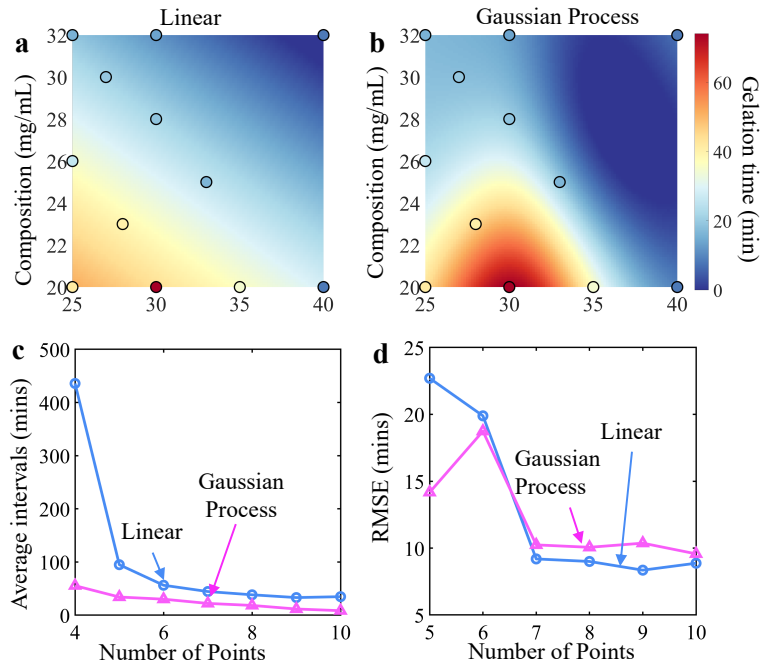

Figure S3: (a-b) Predicted gelation time surface at neutral pH = 7 based on compositions and temperatures using 10 observed points using (a) linear regression and (b) Gaussian Process Regression fit. Filled circles represent experimentally measured gelation time, and their colors denote observed values. (c-d) Comparison of (c) average interval length across the entire surface and (d) out-of-sample RMSE between linear (blue circles) and Gaussian Process (magenta triangles) regressions.
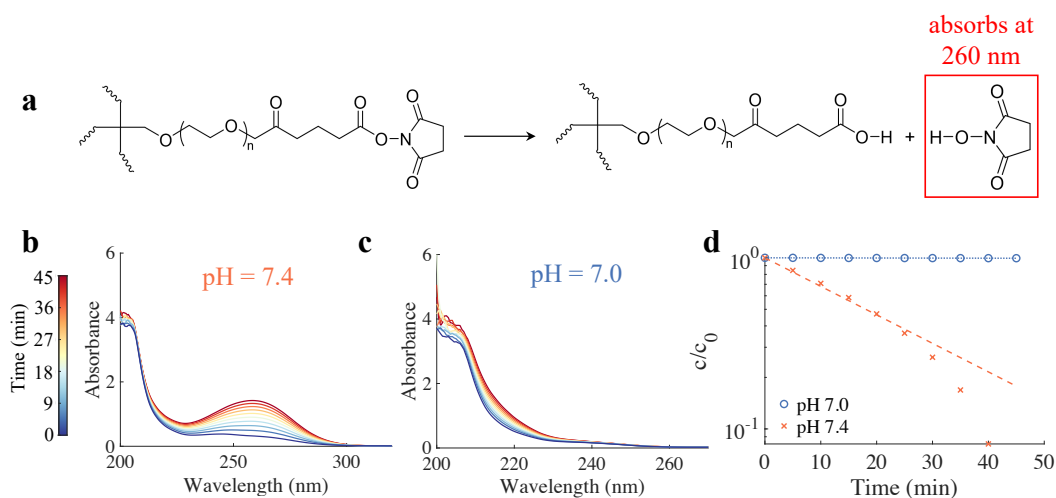
Figure S4: (a) The hydrolysis reaction produces an increasing concentration of the hydrolyzed free glutamate groups (boxed in red). A spectrum was measured every 5 minutes at (b) PBS (pH = 7.4) and (c) DI water (pH = 7). (d) The peak absorbance intensity (260 nm), which is proportional to the concentration of the TPEG-SG, is normalized with respect to the initial value and plotted over time. We found the half-life of hydrolysis at pH = 7.4 to be 26 minutes, whereas the original amount of TPEG-SG barely breaks down at pH = 7 in comparison.
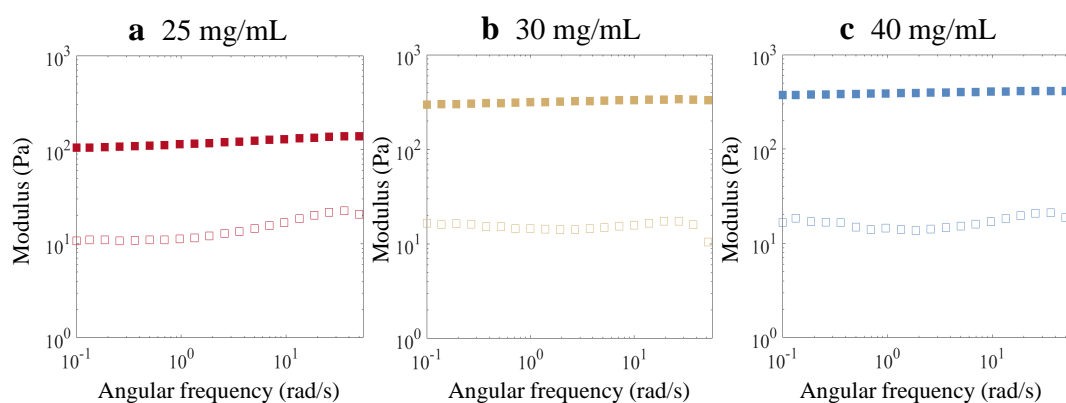
Figure S5: Frequency-dependent storage (G') and loss (G") moduli measured for the collagen-TPEG double network fully crosslinked at 30 °C, after 24 hours.
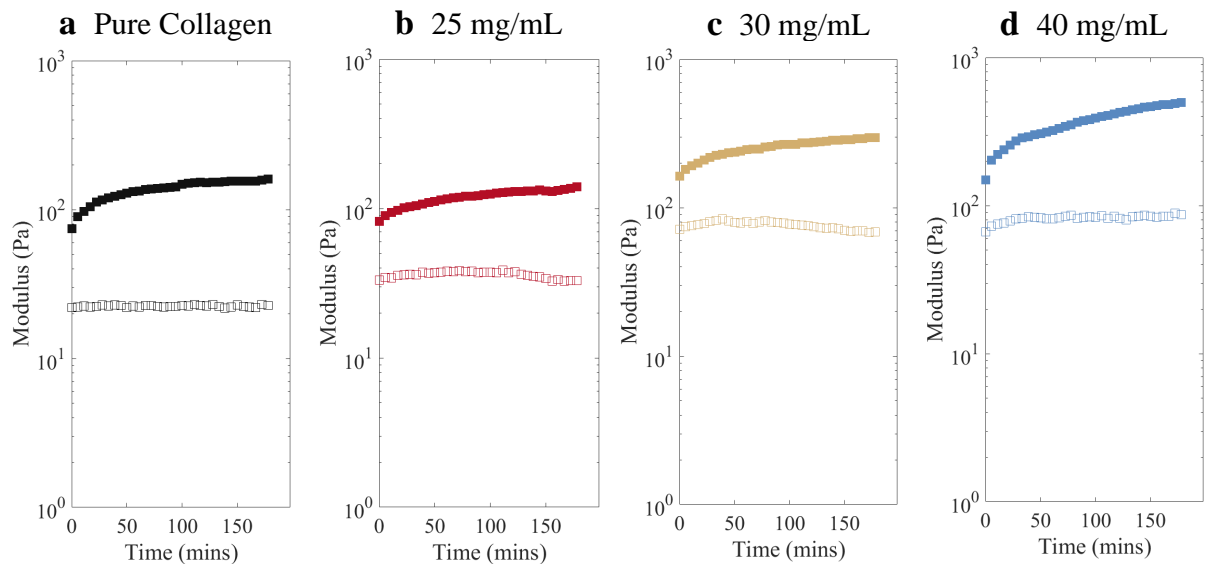
Figure S6: Time-dependent storage (G') and loss (G") moduli measured for the collagen-TPEG double network crosslinking at 30 °C.
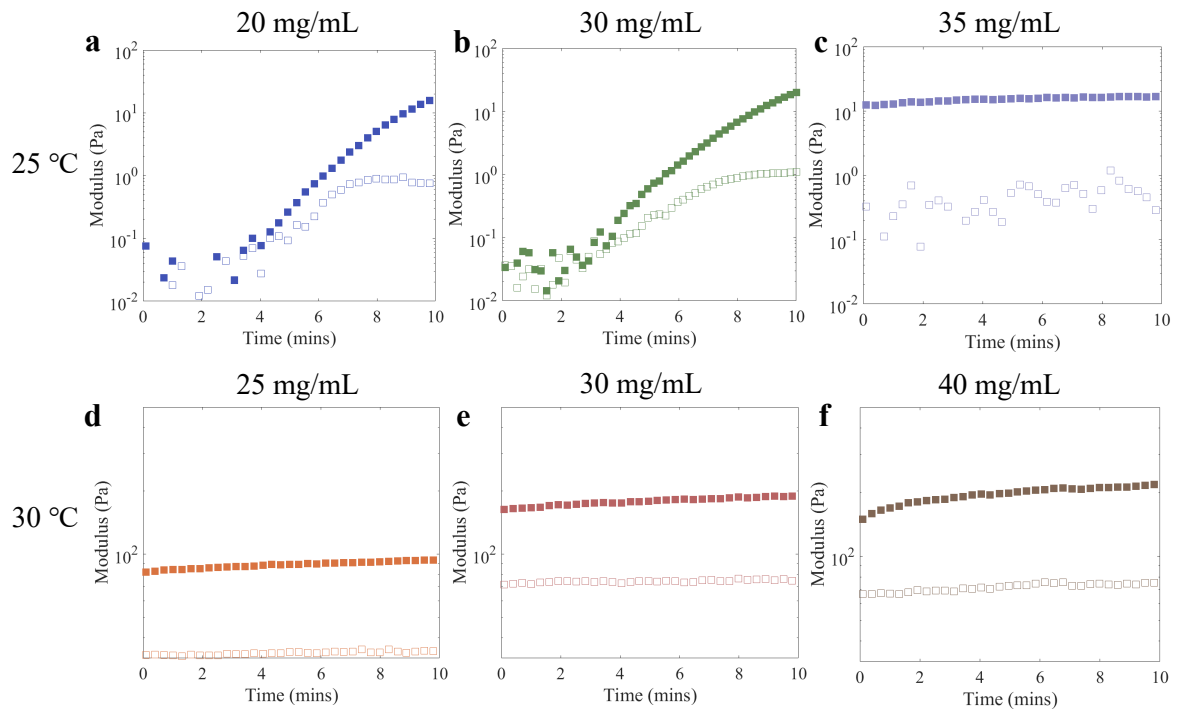

Figure S7: Time-dependent storage (G') and loss (G") moduli measured for the TPEG-only (no collagen) network crosslinking at (a-c) 25 °C and (d-f) 30 °C. The rheology is measured at 1 Hz, 1 rad/s using a 20 mm parallel plate geometry with a gap size of 650 μm.
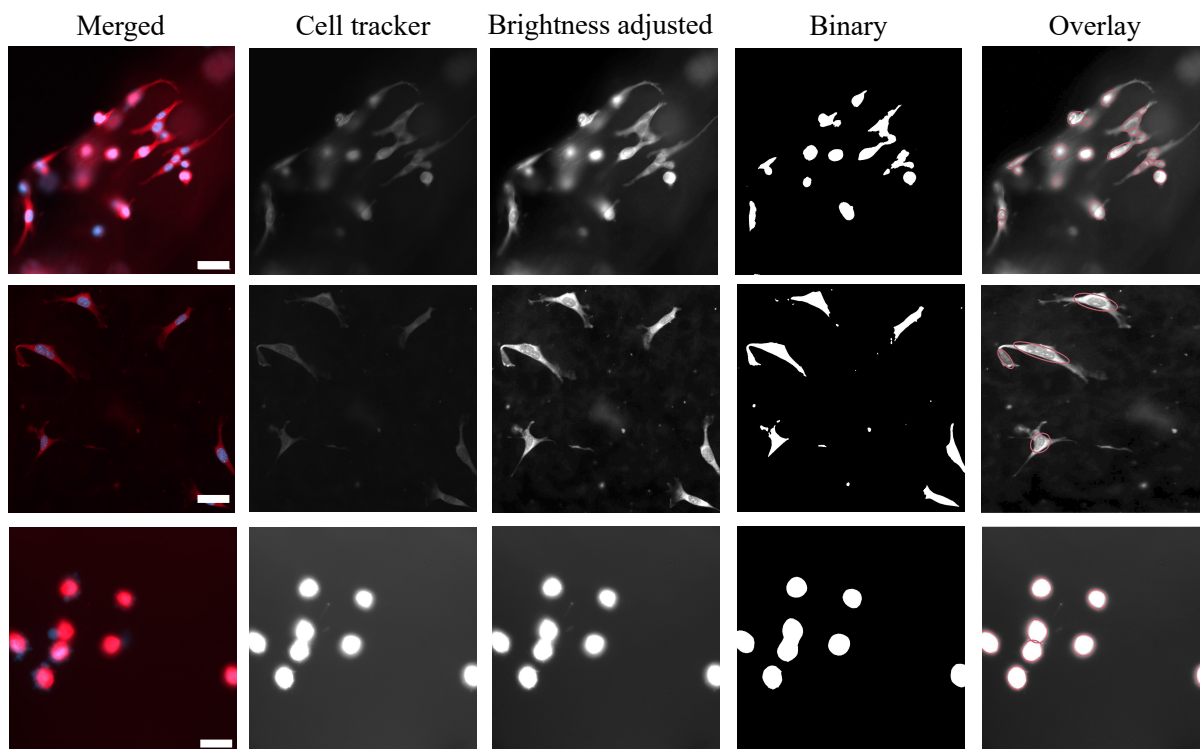
Figure S8: Filtering and fitting ellipse to microscopy images to extract the aspect ratio. The first two rows show the processing of images of cells with extended morphologies at high and low density, and the last row shows those with rounded morphology. The red ellipses denote the best fit by automated filtering on the image. The scale bar is 100 μm.
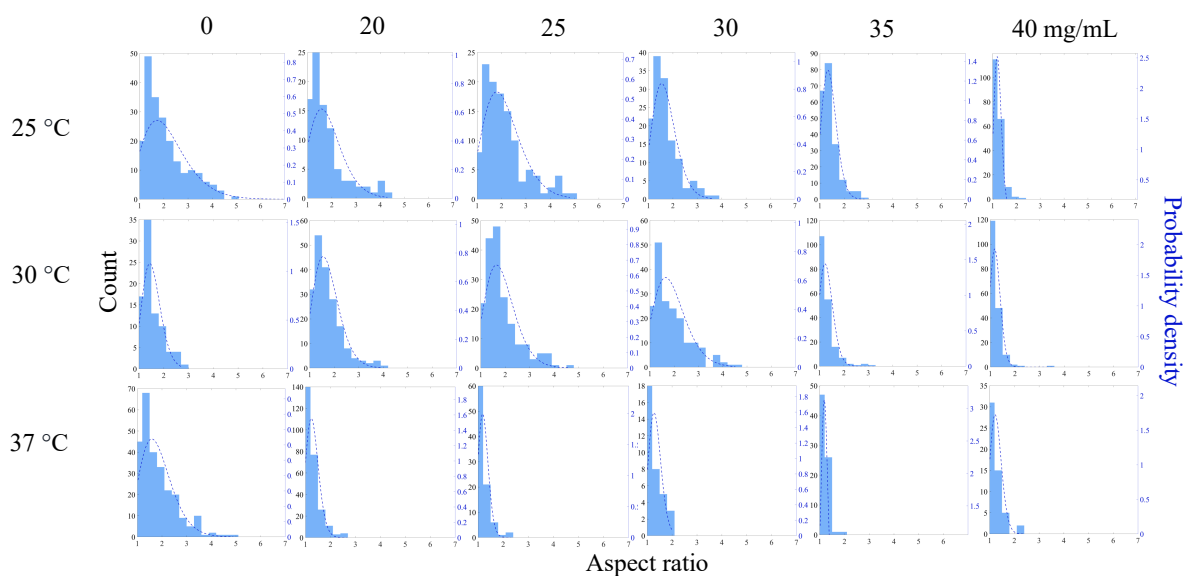

Figure S9: Distribution of cell aspect ratio for different gelation conditions. Each row represents a certain temperature and each column represents a given TPEG concentration. The histograms represent counts of the observed aspect ratio. The dotted lines represent the best-fit gamma distribution probability density.
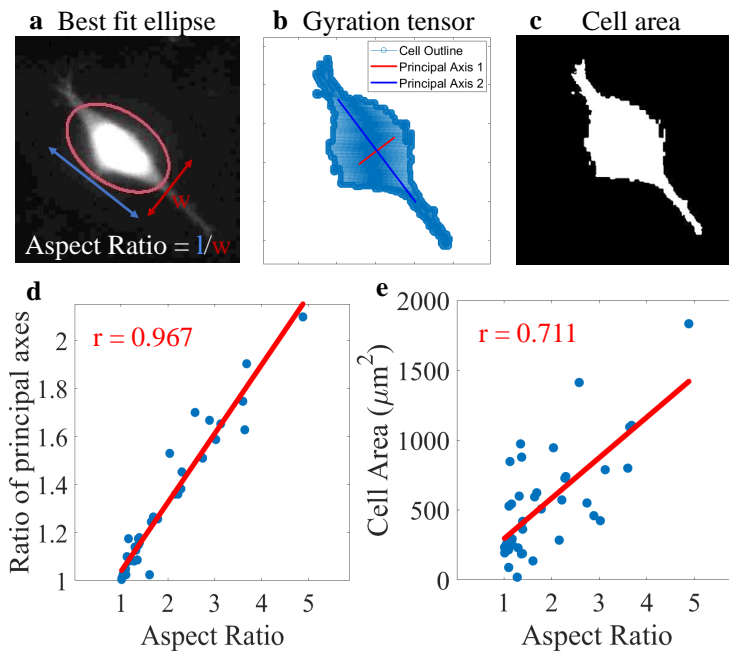
Figure S10: (a-c) The cell shape is quantified by three different metrics. (a) The aspect ratio (AR) is defined as the ratio of the major axis to the minor axis of the best-fit ellipse. (b) The ratio of the principal axes is obtained from the gyration tensor. (c) The cell area is calculated by binarizing the image. The scale bars are 50 μm. (d-e) Scattering plots and correlation coefficient of the ratio of the principle axes and the cell area versus the aspect ratio.
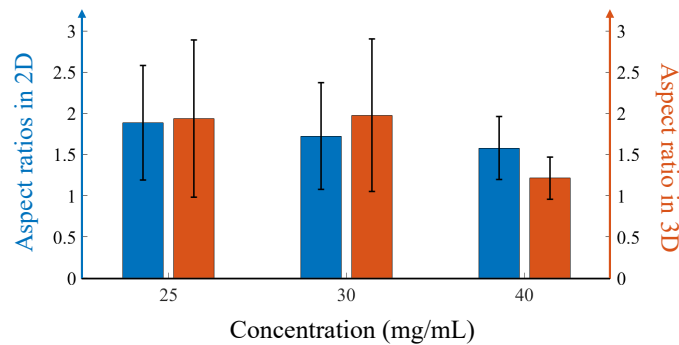

Figure S11: Histograms of aspect ratios for cells on 2D substrates (left aixs) and in 3D hydrogels (right axis). The averages are computed from ~200 cells.
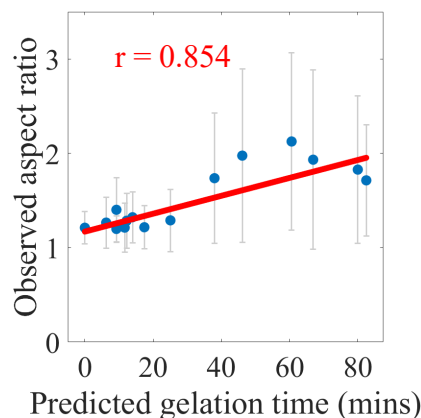

Figure S12: Correlation analysis between gelation time and aspect ratio, where the error bar is quantified by the standard deviation across 200 cells. Cell shapes exhibit a degree of variability due to heterogeneity within the cell population and their sensitivity to the local matrix environment.

References

[1] Gregory S Ammar and William B Gragg. Superfast solution of real positive definite toeplitz systems. *SIAM Journal on Matrix Analysis and Applications*, 9(1):61–76, 1988.

[2] Mengyang Gu, Yue He, Xubo Liu, and Yimin Luo. Ab initio uncertainty quantification in scattering analysis of microscopy. *Phys. Rev. E*, 110(3):034601, 2024.

[3] Mengyang Gu, Yimin Luo, Yue He, Matthew E Helgeson, and Megan T Valentine. Uncertainty quantification and estimation in differential dynamic microscopy. *Phys. Rev. E*, 104(3):034610, 2021.

[4] Yue He, Xubo Liu, and Mengyang Gu. *AIUQ: Ab Initio Uncertainty Quantification*, 2024. R package version 0.5.2.

[5] Yun Ling. *Superfast Inference for Stationary Gaussian Processes in Particle Tracking Microrheology*. PhD thesis, University of Waterloo, 2019.