

Engineering Gelation Kinetics in Living Silk Hydrogels by Differential Dynamic Microscopy Microrheology and Machine Learning

Rhett L. Martineau, Alexandra V. Bayles, Chia-Suei Hung, Kristofer G. Reyes, Matthew E. Helgeson, and Maneesh K. Gupta*

Microbes embedded in hydrogels comprise one form of living material. Discovering formulations that balance potentially competing for mechanical and biological properties in living hydrogels—for example, gel time of the hydrogel formulation and viability of the embedded organisms—can be challenging. In this study, a pipeline is developed to automate the characterization of the gel time of hydrogel formulations. Using this pipeline, living materials comprised of enzymatically crosslinked silk and embedded *E. coli*—formulated from within a 4D parameter space—are engineered to gel within a pre-selected timeframe. Gelation time is estimated using a novel adaptation of microrheology analysis using differential dynamic microscopy (DDM). In order to expedite the discovery of gelation regime boundaries, Bayesian machine learning models are deployed with optimal decision-making under uncertainty. The rate of learning is observed to vary between artificial intelligence (AI)-assisted planning and human planning, with the fastest rate occurring during AI-assisted planning following a round of human planning. For a subset of formulations gelling within a targeted timeframe of 5–15 min, fluorophore production within the embedded cells is substantially similar across treatments, evidencing that gel time can be tuned independent of other material properties—at least over a finite range—while maintaining biological activity.

1. Introduction

A new frontier in materials development exists in combining living components with traditional material matrices. When living cells, for example, bacteria, are encapsulated within

traditional material matrices, new and potentially useful properties can be conferred. For example, embedding bacteria with inherent sensitivity to environmental signals such as pH or the presence of heavy metals^[1] can impart environmental awareness to a material. Enhancements are not limited to competencies which natively exist in embedded organisms; embedded cells can be genetically reprogrammed to be sensitive to non-native signals, as in the case of *Saccharomyces cerevisiae* engineered to be responsive to the presence of fungal pathogens.^[2] When the embedded cells are also capable of responding to sensed signals, for example, by secreting growth factors,^[3] enzymes,^[4] or functionalized protein fibers,^[4] then sensing and actuation circuits are in place to enable smart, evolvable, and living biomaterials.^[5]

A conceptual bacteria-enabled material is shown in Figure 1. In this illustration, genetically engineered bacteria are embedded into a hydrogel matrix. The cells possess a desired environmental sensitivity and are programmed to produce

and secrete protein upon environmental stimulus. The matrix is designed for specific applications with regard to material stiffness and moldability, optical transparency, permeability to diffusing molecules including water and secreted proteins, and cure time.

R. L. Martineau, C.-S. Hung, M. K. Gupta
Materials and Manufacturing Directorate
Air Force Research Laboratory
2179 12th St. B652/R122, WPAFB, OH 45433-7717, USA
E-mail: maneesh.gupta.2@us.af.mil

A. V. Bayles^[†]
Department of Chemical Engineering
University of California Santa Barbara
3357 Engineering II, Santa Barbara, CA 93106, USA

K. G. Reyes
Department of Materials Design and Innovation
University at Buffalo
Buffalo, NY 14260, USA

M. E. Helgeson
Department of Chemical Engineering
University of California Santa Barbara
Santa Barbara, CA 93106-5080, USA

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adbi.202101070>.

^[†]Present address: Soft Materials, Department of Materials, ETH Zürich, Zürich 8093, Switzerland

DOI: 10.1002/adbi.202101070

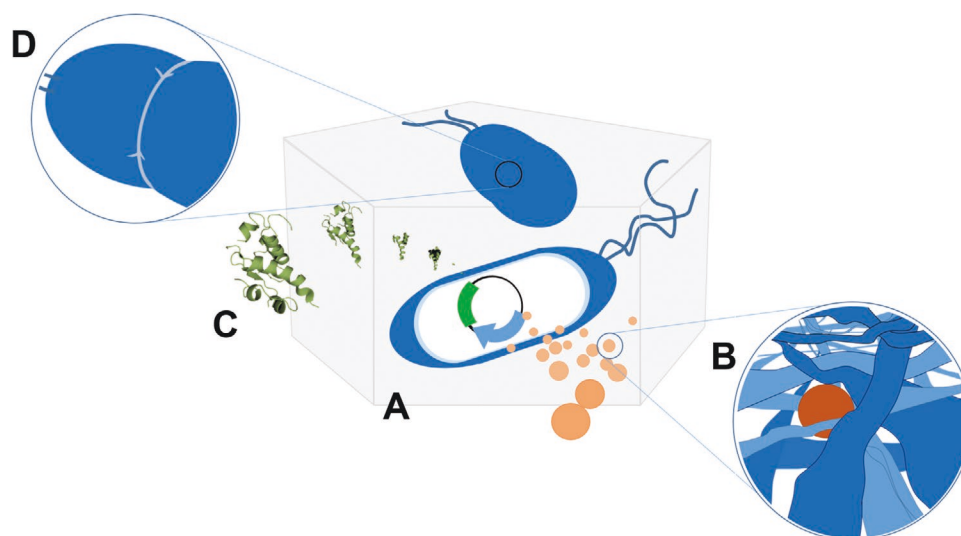


Figure 1. A generic bacteria-embedded living biomaterial. Living biomaterials can be formed by embedding microbes into a material matrix. A) To impart specific biological activity, the microbes may be genetically engineered. B) Signal molecules, (e.g., small molecules ≈ 1 nm in diameter), may be required to diffuse through the embedding medium to be internalized or otherwise sensed by the microbes. C) In response to signal molecules, the microbes may produce biological materials which are transported extracellularly and then diffuse through the embedding medium. D) For some embedding media, microbial mobility is limited and binary fission is difficult, whereas for less stiff or more porous matrices microbial mobility and reproduction are unhindered.

1.1. Living Bacterial Silk Hydrogels

There is a significant body of literature describing efforts to embed living cells into hydrogels. Among hydrogels of recent interest, silk-based hydrogels are prominent.^[6] A number of properties make silk-based hydrogels attractive: in its natural, fibrous forms, silk has outstanding strength and elasticity.^[7] Silk is both biocompatible and biodegradable.^[8] When reconstituted from cocoons produced at industrial scale, silk is abundant. Furthermore, as a genetically encoded, proteinaceous material, a large number of varieties can potentially be produced using recombinant methods.^[9] In addition to sequence modifications at the genetic level, silk as a protein possesses handles for chemical functionalization and bioconjugation, further expanding the palette of material options for engineered applications.^[10] Finally, silk has been successfully processed from aqueous solutions into fibers, hydrogels, foams, and films.^[6e]

Due to its favorable properties, silk materials have featured in numerous applications including: scaffolds for tissue engineering^[6a,8c,11]; injectable matrices for drug delivery^[6d,6g,12]; immobilization matrices for biocompatible, sub-dermal sensors^[6d,13]; optical devices^[8a,14]; and bio-functionalized bulk materials for microfluidic devices.^[15]

Bacteria are an attractive option for adding biological function to materials due to their robustness and genetic tractability. Previous reports include embedding *Bacillus sphaericus* in hydrogels for calcium carbonate precipitation for self-healing concrete,^[16] *E. coli*^[17] and *B. subtilis*^[18] into 3D printing matrices for programmable secretion of biomolecules, *E. coli* into polyacrylamide and agarose for sustained secretion of drugs,^[19] *E. coli* into polyvinyl alcohol for heavy metal toxicity biosensing,^[1] and *Pseudomonas putida* and *A. xylinum* into hyaluronic acid, fumed silica, and κ -carrageenan for bioremediation and in situ production of bacterial nanocellulose.^[20]

1.2. Linking Formulation Conditions with Material Properties

A key requirement for successfully designing living materials is linking formulation conditions with final material properties and—concurrently—biological activity.^[21] Material properties of general interest include those related to material formulation and processing (cure time, polymer fraction), mechanical properties (toughness, stiffness, fatigue strength), and transport properties (mass diffusivity, hydraulic permeability, etc.). Biological activity includes viability and motility of embedded organisms, ability to proliferate, responsivity of embedded organisms to signals of interest, or activity of secreted biologicals within or upon the embedding matrix. Generally speaking, trade-offs exist between idealized mechanical properties and idealized biological activity. A material's suitability for a given application may depend critically upon finding an optimum among competing characteristics.

In developing a novel living material to achieve this balance of properties, a vast space of formula component permutations may need to be characterized. In the case of enzymatically crosslinked silk, hydrogels can be formed from aqueous solutions of purified silk protein, horseradish peroxidase (HRP) as an enzymatic crosslinker, and hydrogen peroxide.^[22] Gelation kinetics and final material stiffness vary widely over this multidimensional, continuous formulation space depending on the concentrations of the individual components.^[6e–h,10–11,22] The inclusion of microbes or other biological components within the embedding medium also impacts the material properties. Conversely, components in the embedding medium potentially impact the biological properties of the living material. As an illustrative example, Hasturk and coworkers embedded human mesenchymal stem cells into hydrogels comprised primarily of silk.^[6c] The materials were desired to rapidly gel once mixed

from precursor solutions and to achieve both high stiffness and high bioactivity in the final hydrogel. Casting gels from precursors dispersed in distilled water resulted in poor survivability of the cells due to osmotic stress, but casting gels from precursors mixed into buffered salt solutions resulted in unacceptably slow gelation. To find formulation conditions that met all performance criteria, it was necessary to navigate a formulation space defined by a large number of parameters. Silk fraction, hydrogen peroxide concentration, crosslinking enzyme concentration, embedded cell concentration, buffer strength, and even the concentrations of potential crosslinking additives were all critical parameters impacting both the material and the biological properties of the final material. In the cited study, only a fraction of treatment combinations could be tested, possibly leaving the most suitable material formulation undiscovered.

1.3. Estimating Gel Times by Automated DDM Microrheology

Rapid screening of the mechanical properties of hydrogel formulations would facilitate living materials developments. For many applications, for example, injectable living hydrogels with biological activity, the gel time of the material (i.e., the time scale over which solid-like mechanical properties develop) is one property that could critically impact a material's successful deployment. Prior to this work, despite the existence of a number of techniques for gel time measurement in hydrogels (see Supporting Information Sections S1–S3), a fully automatable process for gel time measurement in hydrogels had not been reported. Limitations of previous techniques impacting utility for automated implementation include large sample volumes, laborious setup, a need for user interactivity, and ambiguous gelation criteria.

Among existing techniques most promising for automated application is multiple particle tracking (MPT) microrheology.^[23] Indeed, MPT has been used for high-throughput screening.^[24] By tracking the particle trajectories of individual tracer beads embedded in a sample, MPT can be used to produce families of mean-squared displacement (MSD) curves over the course of gelation reactions. Once such MSDs are obtained, a number of robust techniques can be applied to ascertain the gel time.^[24b,d,25] However, the major limitation of MPT precluding automated application is its need for user supervision, as the particle tracking algorithms generally require tuning of analysis parameters from sample to sample. For evolving samples such as gelation reactions, analysis parameters may even require repeated tuning.

A new image analysis technique known as differential dynamic microscopy (DDM) has recently been adapted to conduct microrheological measurements.^[26] Like MPT microrheology, the technique requires commonly available optical equipment. In contrast with MPT microrheology, however, DDM requires no a priori input selection or interactive tuning^[26–27] and is therefore well-suited for automated application. Through Fourier analysis of image intensity fluctuations^[26a,c,28] and the application of theories for thermally-driven motion, DDM quantifies the MSD of embedded tracer beads without the need to track individual particle trajectories.^[26b,29]

1.4. Experimental Approach and Findings

In this work, we present a new experimental methodology and pipeline based on DDM for the automated screening of material formulations for gel time. We present some of DDM's strengths and weaknesses for that task in relation to conventional gel time estimation techniques based on MPT. We also apply novel machine learning algorithms specifically formulated for characterizing a 4D formulation space in terms of gel time. As a demonstrative use case, we develop a living silk hydrogel embedded with viable microorganisms, anticipating applications such as injectable and self-setting biological delivery materials. We asked the following coupled optimization questions: What bacteria-embedded silk hydrogel formulations produce gelation within 5–15 min? Within those formulations, which promote maximum bacterial activity? These coupled criteria—gelation within 5–15 min and maximum bacterial activity—were arbitrarily chosen as examples of criteria that might be targeted when engineering living materials.

In answer to these questions, we demonstrate and validate the capabilities of this new methodology through three experiments: 1) correlation of gel time estimation by a novel adaptation of DDM microrheology—requiring no user inputs and thus fully automatable—to gel time estimation by traditional methods based on MPT which do require user inputs, 2) characterization and prediction of gel time regime boundaries in bacteria-laden silk hydrogels spanning a 4D formulation space by use of novel machine learning algorithms and an automated gel time characterization pipeline, and 3) comparison of bacterial protein production in silk hydrogels of various formulations gelling in 5–15 min by use of confocal microscopy.

In brief, we find that DDM can be used as the basis to objectively and autonomously estimate gel times in silk hydrogels. In conjunction with machine learning, the estimates of gel time by DDM were useful in characterizing regions within a 4D formulation space which were likely to gel faster than 5 min, more slowly than 15 min, or within 5–15 min. Individual bacteria-embedded silk hydrogel formulations exhibited gel times ranging from instantaneous gelation to no gelation at all, with 8 of 63 treatments gelling within the targeted 5–15 min window. Among a subset of samples that gelled within the 5–15 min window, biological productivity of embedded *E. coli*, as measured by fluorescent protein production and colony morphology, was found to be relatively constant despite a significant difference in formula constituency.

Over the course of the 4D formula space characterization campaign, we employed Bayesian machine learning (ML) models and algorithms to perform decision-making under uncertainty to accelerate knowledge acquisition. ML supervision can proceed without operator intervention, however the fastest rate of learning was observed when a round of expert user decision making was interspersed between rounds of artificial intelligence (AI) decision making. Without operator intervention, the ML algorithms were capable of approximating gel regime boundaries in as few as 20 treatments.

The findings of this work collectively illustrate the complex relationship between formulation and properties in living silk hydrogels and show that automated materials characterization techniques such as DDM microrheology can help to navigate

seemingly intractable and vast formulation spaces, especially when paired with ML-informed modeling and decision making algorithms. Importantly, the techniques deployed in this work—in particular gel time estimation by DDM—are easily transferred to other laboratories and can be implemented on equipment which is readily available.

2. Results and Discussion

2.1. Comparison of Gel Time Estimation by DDM and by MPT

Characterization of gelation kinetics requires an estimation of the gel time. A number of methods for gel time estimation exist based on various dynamic linear viscoelastic material functions, and these methods are summarized in detail in the Supporting Information Sections S1–S3. These methods include determination of the point of crossover between the elastic and viscous moduli ($G' \approx G''$) for a given deformation frequency or frequency range, the point at which the loss tangent $\tan \delta$ becomes independent of frequency, or by time-cure superposition (TCS), in which time-domain shifting of the viscoelastic response is used to identify the asymptotic point at which shift factors transition from pre-gel to post-gel scaling. Each of these methods is accessible to MPT microrheology by way of MSD curves.

As mentioned previously, DDM can be used to produce MSDs without user supervision and is thus potentially amenable to automation. When used to calculate MSD curves, DDM compares well with MPT over a wide variety of conditions.^[26,29] As MSD curves can serve as the foundation for rigorous microrheological assessment of gelation, both DDM and MPT can be used to measure gel time. However, automating gel time estimation by way of MSDs can be challenging for a number of reasons: 1) there exists no simple gelation criterion directly accessible from MSDs. That is, a power law exponent threshold of 0.5 (or any other value) cannot be used to accurately indicate gelation in all gelling systems, 2) computing elastic and viscous moduli from MSD data with limited frequency content is difficult, especially in gelling systems, potentially precluding the use of gel time estimations based on G' and G'' , and 3) automating TCS can be difficult and subjective, and requires pre- and post-gelation MSD curves which may not always be available.

During a gelation reaction, probe particles undergo progressively smaller cumulative displacements approaching the gel point and beyond, where probe particle motion is restricted by the elasticity of the gel. DDM fails to replicate MPT-extracted MSD curves in samples exhibiting very small cumulative displacements (see Supporting Information Section S4), where the weak intensity fluctuations associated with particle motion are not easily discernable from other contributors to the image structure function.^[26b] Under these circumstances, the DDM signal generally fails to meet the tolerance criterion for successful MSD estimation. In some cases under these circumstances, the tolerance criterion is met but no physically meaningful MSD curves are produced. An example of the difference in the sensitivity range between MPT and DDM is highlighted in Section S4 of the Supporting Information, where MPT yields physically reasonable MSD curves at time points well beyond those that produce physically reasonable MSD curves by DDM.

The lack of MSD curves resulting from DDM analyses of late-stage gelation limits the amount of data that can be observed during a gelation reaction, potentially precluding gel time estimation by G'/G'' cross-over, TCS, or $\tan(\delta)$ analysis. However, in practice, the point at which DDM fails to extract physically reasonable MSD curves may itself be useful as an estimate of the gel time. In this work, we develop a technique for gel time estimation based on identifying the time during a gelation reaction after which no MSD curves can be extracted from $D(q, \Delta t)$ curves obtained by DDM. We term this method DDM-RL, referring to the resolution limit of DDM for MSD extraction. If DDM-RL does in fact produce useful estimates of the gel time, then the parameter-free nature of the analysis makes DDM microrheology an excellent basis for an automatable pipeline for objective gel time characterization.

The first investigation presented in this work examines DDM-RL as a method for gel time estimation. Estimates obtained by DDM-RL are compared with two conventional methods based on further processing of MSDs obtained by MPT: 1) converting the MSDs into mechanical moduli and then observing the time at which G' exceeds G'' over all measured frequencies (termed here, MPT- G'/G''), 2) applying TCS to the MSDs (MPT-TCS). DDM-RL estimates are also compared with estimates obtained by oscillatory rheology of bulk samples, based on the time at which the loss tangent, $\tan(\delta)$, becomes frequency-invariant.

Passive probe microrheology data from three silk hydrogel formulations with distinct gel times were analyzed by DDM-RL, MPT- G'/G'' , and MPT-TCS. For this comparison, none of the silk formulas contained bacterial biomass. All gels contained 3% silk (w/v). The HRP and H_2O_2 concentrations varied as: 1) 20 U mL⁻¹ HRP with 53 ppm H_2O_2 , 2) 10 U mL⁻¹ with 25 ppm H_2O_2 , and 3) 3.33 U mL⁻¹ HRP with 25 ppm H_2O_2 . Results are shown in Figure 2. Estimates of gel times obtained by DDM-RL are reported as the interval between the time of the last resolved MSD curve and the sum of that time plus the sampling interval. See Section 4 for experimental detail. Estimates by MPT- G'/G'' are reported as the range of times bracketing the presumed switch from a pre-gel to post-gel regime based on the criterion that $G' > G''$ (or $G' \approx G''$) for all measured frequencies. Estimates by MPT-TCS are reported as the range of times bracketing the estimated critical gelation time.

In general, during gelation processes, the MSDs measured by both DDM and MPT are expected to exhibit power law behavior with decreasing power law exponents as gelation proceeds, that is, a decreasing linear slope on a log–log scale. For the cases of DDM here, each sample exhibits power law slopes greater than 0.5 for all time points at which the MSD can be resolved. After converting MSDs obtained by DDM to G'/G'' plots, no time point exhibited G' in excess of G'' for all frequencies. This is due to the resolution limit of DDM in extracting MSDs, with a lower bound on MSD related to the pixel size and the influence of intensity fluctuations away from the focal plane.^[26b] However, DDM does produce MSDs for some time points where $G' > G''$ for some frequencies. In the case of 3% silk, 10 U mL⁻¹ HRP with 25 ppm H_2O_2 (Figure 2B), DDM produced an MSD curve for the 5.3 min time point that corresponds to $G' \approx G''$ for many of the lower frequencies. This demonstrates that DDM can resolve MSD curves at time points very close to the gel time.

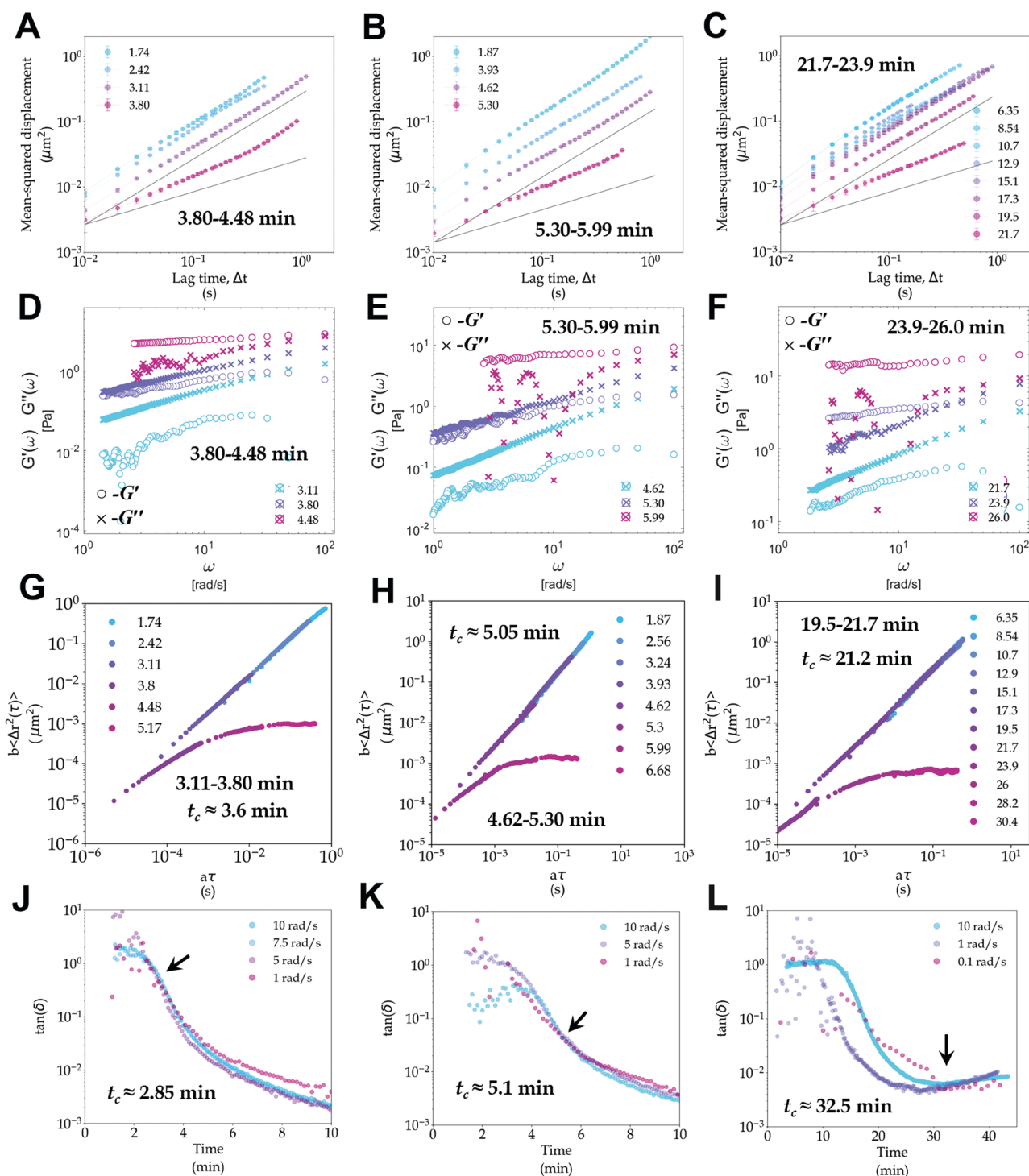


Figure 2. Estimation of gel time by DDM-RL, MPT- G'/G'' , MPT-TCS, and bulk rheology. Gel time was estimated by A–C) DDM-RL, by D–F) MPT- G'/G'' crossover, G–I) MPT-TCS, and J–L) bulk rheology for three gels. Each gel contained 3% silk but varied in the other constituents. For the sample comprised of 20 U mL⁻¹ HRP with 53 ppm H₂O₂, gel times were estimated at 3.80–4.48 min by DDM-RL (A), 3.80–4.48 min by MPT- G'/G'' (D), 3.11–3.80 min by MPT-TCS with a critical gelation time of approximately 3.6 min (G), and at approximately 2.85 min by bulk rheology (J). For the sample comprised of 10 U mL⁻¹ HRP with 25 ppm H₂O₂, estimates were between 5.30–5.99 min by DDM-RL (B), 5.30–5.99 min by MPT- G'/G'' (E), 4.62–5.30 min by MPT-TCS with an estimated critical gelation time of approximately 5.05 min (H), and at approximately 5.1 min by bulk rheology (K). The sample comprised of 3.33 U mL⁻¹ HRP with 25 ppm H₂O₂, was measured to gel at 21.7–23.9 min by DDM-RL (C), 23.9–26.0 min by MPT- G'/G'' (F), 19.5–21.7 min by MPT-TCS (I) with an estimated critical gelation time of 21.2 min, and at approximately 32.5 min by bulk rheology (L). Grey lines (A–C) indicate power law slopes of $\frac{1}{2}$ and 1. Panels (D–F) show G' and G'' plots obtained by local power-law fitting of MSDs obtained by MPT. Panels (G–I) show time-cure superposition results from shifting MSDs obtained by MPT. Panels (J–L) show results of bulk rheology, with critical gel times estimated by the time of $\tan(\delta)$ frequency-invariance.

In the three cases presented in Figure 2, estimates of gel time by DDM-RL and by MPT- G'/G'' approximately agree. Likewise, for an additional three replicates of the 3% silk, 3.33 U mL⁻¹ HRP, and 25 ppm H₂O₂ case, each estimate by DDM-RL overlaps with the estimate by MPT- G'/G'' (data not shown). In most of these cases, the latest gel time estimated by DDM-RL matches the earliest time suggested by MPT- G'/G'' . That is, DDM fails to resolve MSDs when G' begins to be equal to or to exceed G'' for some frequencies. As a result, in comparison with MPT- G'/G'' on these samples, DDM-RL tends to slightly underestimate the gel time. It is also possible that DDM-RL reports gelation when gelation has not (and will not) occur, although this has not been observed in any of the cases reported in this study. On the other hand, when DDM-RL gel times are compared with MPT-TCS for these gelling samples, DDM-RL indicates gelation after gelation has occurred (and therefore may be able to observe the critical gel point if the observation timing was perfectly synchronized with gel point.) In general, whether in comparison with MPT- G'/G'' or with MPT-TCS, gel time estimation by DDM-RL lies within a few observation intervals on these gelling silk samples.

If drift and convective currents exist in the sample (e.g., as might be indicated in Figure 2A at the 3.8 min time point), DDM-RL might tend to overestimate the gel time. Indeed, many of the MSD curves reported in this study exhibit upturns in the MSDs at large lag times. Dedrifting algorithms applied to the MPT data eliminate these upturns (see Supporting Information S5). The drift is not unexpected as the measurements were taken in open well dishes to facilitate robotic handling. Importantly, the drift in these samples does not appear to significantly impact the gel time estimates by DDM-RL.

As applied here, DDM does not identify drift or convective currents in the sample. A potentially useful adaptation to the DDM algorithm could identify and computationally eliminate excessive drift or convective currents, as is currently possible with MPT. Such an adaptation to DDM which compensates for unidirectional flow was recently developed.^[30]

Despite the difficulties of gel time estimation by DDM, DDM still appears to be useful in screening scenarios where automated analysis is desirable. Each of the samples observed to gel by way of MPT- G'/G'' crossover, MPT-TCS, and bulk rheology was also successfully reported to gel by DDM-RL, with estimates of gel time by DDM-RL generally agreeing with either MPT- G'/G'' or MPT-TCS within one measurement interval. Gel time estimates from each of the microrheology techniques are less congruent with estimates from bulk rheology, consistent with other comparisons in the literature.^[25d,31]

Based on these findings, we argue that for objective, parameter-free characterization of gel time in hydrogel formulations, DDM-RL produces approximate gel times that are sufficiently accurate to be useful as an initial screen. Although the estimation of gel time by this technique is not rigorously supported by theory, the MSD curves collected during execution of the technique do capture important rheological information that is quantitatively comparable to more rigorous viscoelastic measures obtained by MPT. As desired, the MSDs obtained by DDM can be processed to quantify the evolution of complex shear modulus or viscosity as a sample gels. Also, since the same raw data is used for both DDM and MPT, the image stacks for a

given sample can be reprocessed by MPT if additional information is required—for example, the explicit particle trajectories or van Hove displacement distributions.

As already mentioned, DDM does not require a user to fine-tune parameters for each analysis and is therefore an excellent option for automated characterizations. This is especially beneficial when the sample is optically changing over the course of an experiment, as was sometimes observed for these gelling samples. As samples gelled, changes in the opacity required careful tuning of the processing parameters when analyzing by MPT. In contrast, analysis by DDM proceeded without regard for changes in optical opacity.

Parameter-free measurement of evolving MSD curves in gelling samples could have potentially been obtained using a recently developed convolutional neural network (CNN).^[32] The CNN was trained using both simulated and real particle imaging videos and can convert image stacks into MSDs without user-supplied parameters. We processed image stacks from a silk gelation (see aitracker.net) and observed that CNN based particle tracking and conversion to MSDs was excellent—and agreed well with DDM-based MSDs—for early time points. However, for later stages in the gelation process, the CNN algorithms failed. See Supporting Information S6 for details which indicate that automated gel time estimation by CNN holds significant promise but is not currently suitable for routine use.

2.2. Automated Screening of Gel Time in Bacteria-Laden Silk Hydrogels in 4D Formulation Space

Having established that DDM-RL is a useful estimator of gel time, we constructed a pipeline for its automated implementation. The pipeline consists of the following operations (Figure 3A): 1) formula specification, 2) mixing and dispensing, 3) video microscopy, 4) DDM microrheology, and 5) gel time estimation by DDM-RL. Once the gel time was determined, the experiment's results were added to a growing ensemble of results, and a new formula specification was generated by an experimental planner and the cycle repeated.

The possible combinations of formula components in a 4D, continuous formulation space are potentially overwhelming. To address this issue, we deployed a novel ML strategy based on iterative Bayesian closed-loop design. In contrast to more traditional design of experiment (DOE) approaches based on collecting batches of observations to fill a model design space, we coupled small sets of exploratory experiments with Bayesian beliefs to more adaptively navigate the formulation space, relying on a sequential experimental design and a decision policy known as the Generalized Knowledge Gradient (GKG). The method was specifically devised to optimally identify gel regime boundaries in the design space. The basic scheme is illustrated in Figure 3B. After an experiment, the gel time landscape over the 4D input space is expressed as a Gaussian process (GP) model and used to estimate the gel regime boundaries. The GKG policy then selects the formulation that is expected to most impactfully change this prior estimate of the boundary. The policy employs a Monte Carlo (MC) approach to calculate the expected change of the boundary. Namely, the GP model is then also used to simulate new data points within the gelation time landscape from which

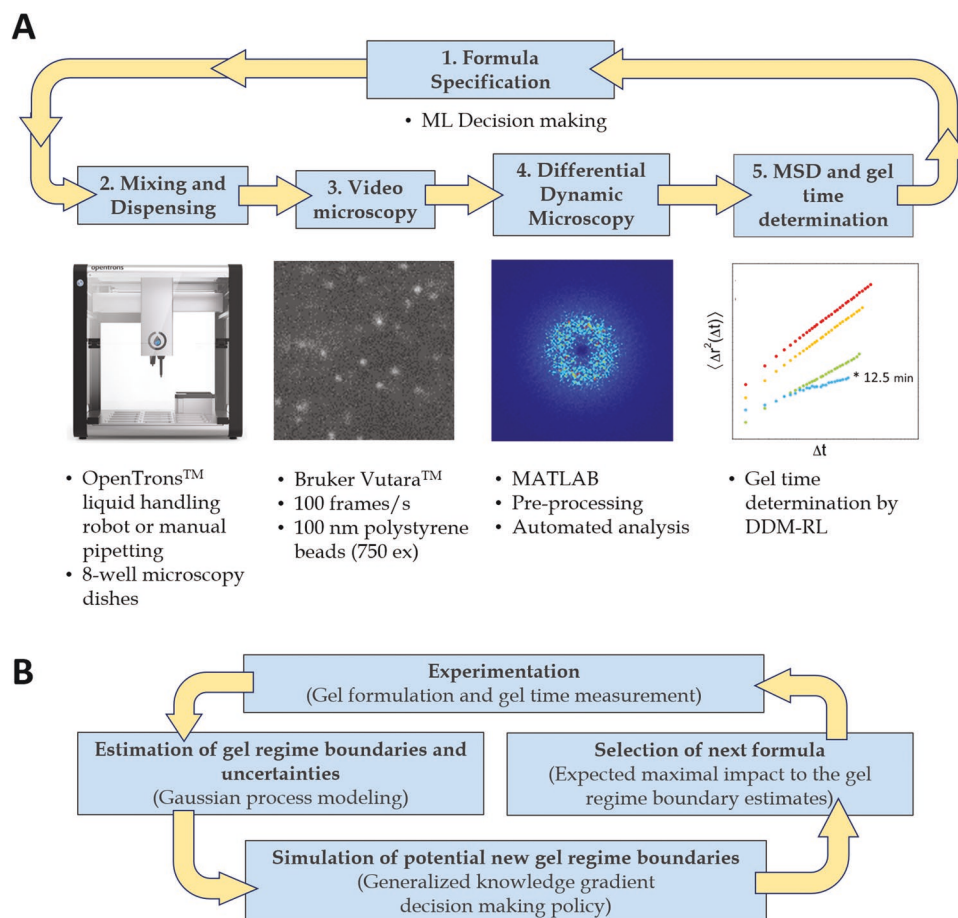


Figure 3. Automated gel time characterization pipeline. A) A closed-loop experimental campaign to characterize a 4D silk hydrogel formulation space was conducted iteratively. B) Each iterate of the campaign began with a formulation specification suggested by a decision-making policy informed by Bayesian ML models of gelation regimes, or by an expert operator. Next, the suggested formulas were mixed using a pipetting robot or manually. Samples were loaded into open wells and manually placed onto the microscope for image acquisition. Images were acquired automatically at set intervals and framerates and then processed using a series of automated MATLAB scripts which implemented DDM. The gel time was estimated from the time course of MSD curves using DDM-RL. Whereas the estimated gel times were earlier specified as an interval, when interacting with the ML algorithms the gel time was estimated as $\frac{1}{2}$ the sampling interval added to the time of the last resolved MSD.

new potential posterior gel regime boundaries are calculated and compared with the prior boundaries, obtaining a measure of difference between the two estimated boundaries. This is repeated several times to obtain an average measure of difference for a given formulation. The formula specification which is expected to maximally impact the gel regime boundary, as measured by this MC-averaged difference, is then selected as the next physical experiment to run, and the cycle repeats.

By focusing on specific desiderata in an adaptive manner, we can drive the exploration to achieve the particular design objectives of an investigation—in this case, finding formulas that gel within 5 to 15 min—without needing to learn the design space landscape in general. As a result, fewer experiments may be required, as has been demonstrated for a number of related sequential approaches in comparison with traditional global or space filling designs.^[33] This efficiency is particularly important in higher-dimensional design spaces as is the case in this application. The specifics of the boundary-learning ML method we developed for this work are presented in Section 4, and in Section S7 of the Supporting Information.

The results of a campaign to characterize gel times by DDM-RL in a 4D formulation space are presented below in Figure 4. The results were obtained from 63 formula characterization experiments. A typical iteration of the campaign loop, testing 4 formulations at a time, required approximately 4 h to complete. See Section 4 for details.

The formula samples were organized into three gel regimes: samples gelling within the selected 5–15 min window, samples gelling faster than 5 min, and samples taking longer than 15 min to gel or not gelling at all. Individual observations are shown in the plots in Figure 4A which also show predicted gel regime boundaries as the campaign progresses.

Selection of formulations for characterization was alternated between our ML algorithm (picking four treatments per iteration) and a domain expert (picking one treatment at a time) attempting to pick treatment combinations that gelled within the 5–15 min window based on domain expertise.

Gelation data and the Bayesian beliefs on gelation time, as presented in Figure 4, have been projected from full 4D input space to two dimensions by plotting as $\text{H}_2\text{O}_2/\text{Cell}$ versus

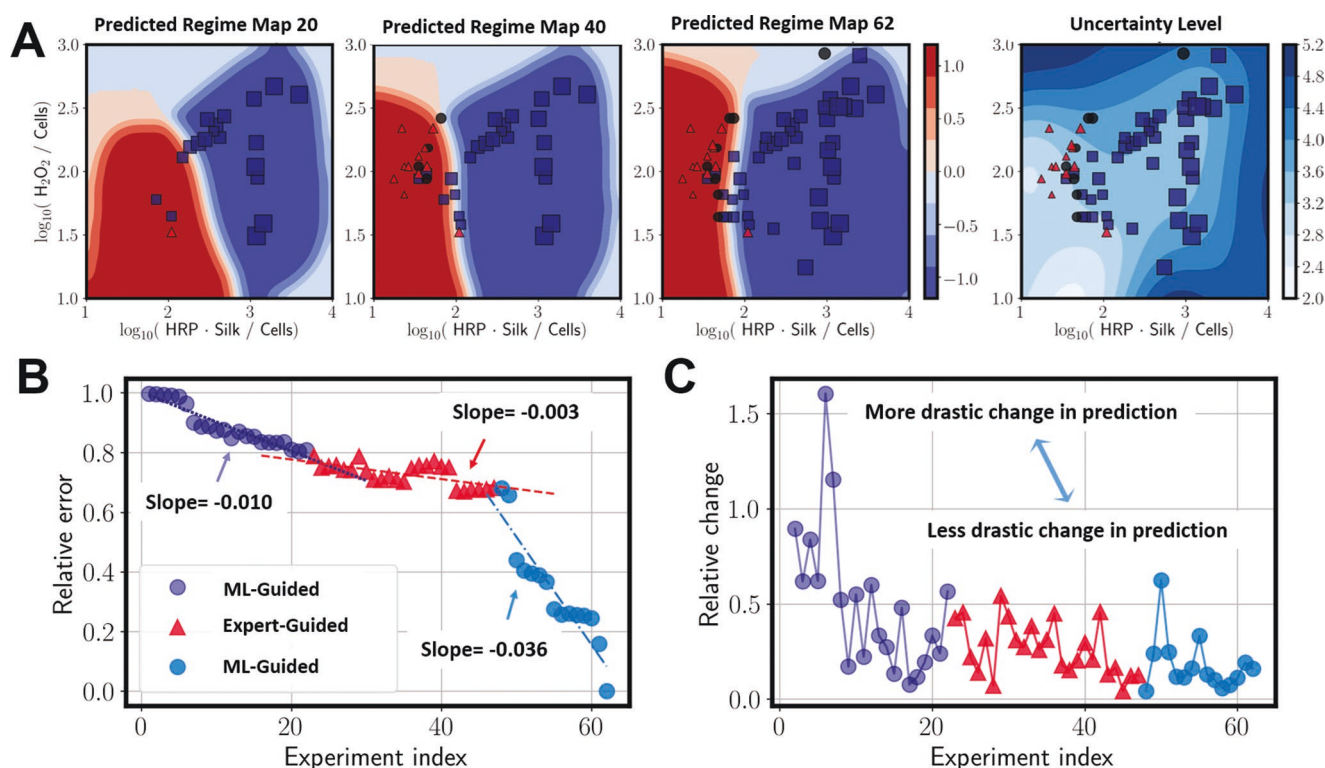


Figure 4. ML and expert-guided experiment planning. A) Low-dimensional visualizations of the mean-estimates of the gelation regimes after 20, 40, and 62 experiments. Observations that gelled within the targeted window of 5–15 min are shown as black circles. Observations that gelled faster than 5 min are shown as blue squares. Observations that took longer than 15 min, or that did not gel at all in the 30 min observation window, are marked as red triangles. The size of the marker reflects the relative concentration of HRP, with the larger markers indicating larger concentrations. Colored regions indicate areas in the formulation space predicted by an SVM classifier to produce samples gelling in three gelation regimes: 1) faster than 5 min (blue region), slower than 15 min or not gelling at all within the observation period of 30 min (red region), or 3) samples gelling within the targeted window of 5–15 min (lighter red and blue). Units are coded gel regime units, with $\ell_i = -1$ representing gel times less than 5 min and $\ell_i = 1$ representing gel times greater than 15. The plot at far right in Panel A presents the uncertainty associated with the predictions after 62 experiments, representing the error bars (in minutes) on the corresponding gel regime predictions. B) The error in the estimates of the full-dimensional gelation time boundary obtained for each iteration in the campaign relative to the estimate obtained after 63 observations. Purple dots show the relative error during an initial ML-guided phase of the campaign. Red triangles show the errors observed during a phase in which a domain-expert selected the formula components with the intent to find samples that gelled within the desired timeframe. Blue circles show errors incurred during a second ML-guided phase of the campaign. Dashed lines are a linear fit to the data, with the slopes of these lines indicating a measure of the rate of learning, with more negative rates signifying faster learning. The expert opinion phase had the slowest rate of 0.003, which is more than 3 times slower than that of the initial ML-guided phase (rate = 0.01) and ten times slower than the final ML-guided phase (rate = 0.036). C) A comparison of successive gel time regime boundary estimates, as measured by relative change. Larger values indicate experiments that more drastically change the prediction. The most drastic changes occur in the beginning of the first ML-guided phase of the campaign.

HRP*Silk/Cells. Exploratory data visualization revealed this representation to be the most differentiating with respect to gelation kinetics.

Not all observations fell within regions delineated by the model predictions. Notably, though, the approximate gel regime boundaries are qualitatively projected in as little as 20 observations, with only minor alterations observed progressing through the full 63 experiments. Interestingly, the learning rate observed at the beginning of the experimental campaign is substantially slower than that observed in the later stages of the campaign. In both ML-guided phases, learning was faster than that achieved when a domain expert attempted to find regions within the targeted 5–15 min window, suggesting the policy designed specifically to quickly learn the gel regime boundaries is successfully doing so. Of additional note, the treatments selected during the expert-guided stage appeared to result in an increased learning rate in the following ML-guided stage. Experiments selected by

the expert were exploitation ones – experiments believed by the domain expert to produce gelation times in the target window, while ML-selected experiments attempted to balance exploitation and parameter space exploration. These results suggest alternative exploration/exploitation stages may be beneficial. Another interesting question raised here is the potential for collaborative teaming between ML algorithms and domain experts. The use of ML allows for reasoning in high dimensions with statistical analysis, while human experts can make suggestions using prior or unmodeled knowledge not built into the ML models. These results point to the beneficial effect that combining both approaches has to accelerating discovery.

The rapid rate of learning at the end of the 63 experiments suggests that more data are necessary to definitively characterize the gel regime boundaries. However, the decreasing error between each iteration's expectations and the expectations in the final iteration suggests that predictive quality in the dataset is converging.

The gelation times of silk hydrogels in this study varied widely from near instantaneous gelation, to gelation over the course of tens of minutes, to not gelling at all even after many hours (as observed incidentally; the observation period for microrheology was 30 min). The formulation transition from samples that gel faster than 5 min to samples that take longer than 15 min to gel is abrupt; the gelation window of 5–15 min is extremely narrow under the conditions explored in this study. This is an example of a situation where discovering the gel regime boundaries may be preferable to fully characterizing the formulation space, as a full characterization is expensive to obtain and a sparsely occupied space-filling design might poorly characterize the boundaries.

Also indicated in Figure 4C is the relative concentration of HRP in each formulation. The size of the marker indicates the relative concentration of HRP, with larger markers indicating higher concentrations of HRP. A clear relationship exists between HRP concentration and gel time, with higher HRP concentrations tending to gel the reactions more quickly. This finding agrees with other literature detailing HRP-mediated silk hydrogelation.^[6c]

Table S1 (see Supporting Information) presents the data in tabular form. Silk was added in the range of 1.71–6.0% w/v, hydrogen peroxide ranged from 15–300 parts per million, HRP from 5–243 U mL⁻¹, and cell concentrations from 0.1–1.4 × 10⁹ cells mL⁻¹. 8 of the 63 samples, 7 obtained during the expert guided phase, gelled within the specified 5–15 min window. Note that the object of the ML algorithm was to discover the gel regime boundaries in the formulation space and not to find specific formulas that gelled within the 5–15 min window.

2.3. Comparison of Protein Production in 5–15 min Silk Hydrogels

A second experiment was designed to determine which formulas that gelled within 5–15 min promoted maximum bacterial activity. Four such formulas were selected for further study. To facilitate comparison, each chosen formula contained a matched bacterial load of 1.14 × 10⁹ cells mL⁻¹. Beyond that, the selected formulas sampled the widest range of constituent concentrations available in the previously obtained 5–15 min samples and included: 1) 5.31% silk, 175 ppm H₂O₂, and 10 U mL⁻¹ HRP, 2) 3% silk, 50 ppm H₂O₂, and 18 U mL⁻¹ HRP, 3) 2% silk, 125 ppm H₂O₂, and 20 U mL⁻¹ HRP, and 4) 3% silk, 300 ppm H₂O₂, and 28 U mL⁻¹ HRP.

For these experiments, fresh cultures were grown to the required OD₆₀₀ and added to the treatments immediately. This is in contrast to the way cells were prepared and added to the treatments in the 4D formula space experiment presented in Figure 4. In that case, all cells for the entire campaign were prepared in one large batch and frozen in aliquots to ensure matched cell fractions throughout the characterization campaign. In these experiments addressing bacterial viability, fresh cells were added to gel formulas. Once fresh cells were added to the formulas, the samples were allowed to gel, overlaid with culture medium, and cultured for 24 h under inducing conditions.

From preliminary growth experiments, it had been observed that both the fluorescence intensity produced by the

embedded cells and evidence of extended bacterial growth (as indicated by the appearance of bacterial colonies) varied significantly based on location within a gel. That is, the further away a cell was from the nutrient overlay within any given sample, the less intense, generally, the fluorescence. In fact, induction of fluorescence in cells located deep within a gel was extremely poor (see Section S8 of the Supporting Information.) In gels seeded with dilute suspensions of cells, where the oxygen and other nutrient levels were not limiting during early stages of growth, a number of distinct colony morphologies (including oblong spherical, filamentous, and “hairy” filamentous) were produced in different hydrogel formulations even in deep sections (see Section S9 in the Supporting Information). In contrast, when cells were added in higher concentrations (such as was done for the replicated treatments), cells grew into colonies only near the surface of the gel–liquid interface perhaps as facilitated by more readily available oxygen or other nutrients.

To ensure similar nutrient and oxygen availability between samples, gels were cast into thin sheets by adding approximately 100 µL of pre-gel material to the open wells and then spreading the material to cover the entire well bottom (≈1 cm²). Casting into thin sheets created areas that could be observed microscopically to contain cells growing in gels near the liquid–gel interface, where a pathway to oxygen and nutrient exchange was not blocked by competing cells. Confocal stacks used for comparisons between treatments were selected to show regions spanning the coverglass, at various z-positions within the hydrogel including near to the gel–liquid interface, and extending into the culture medium overlay. Thus each imaged gel section was no thicker than approximately 180 µm. Figure 5A illustrates the imaged regions. Results from imaging through the thin sections of the four bacteria-laden silk hydrogels cast in formulas that previously gelled within 5–15 min is presented in Figure 5.

Of the four treatments recast with fresh cells, two produced gels that consistently gelled in the expected 5–15 min window. These treatments were: 3% silk, 50 ppm H₂O₂, and 18 U mL⁻¹ HRP; and 2% silk, 125 ppm H₂O₂, and 20 U mL⁻¹ HRP. These growth experiments, with concurrent gel time measurements, were repeated on two, three, and four separate days (see Figure 5B) starting from fresh inoculum. In general, the repeatability of each treatment across days is excellent.

The two treatments that did not gel within the expected window (treatments 1 and 4 from Figure 5B) were observed to gel within 30 min. The inability to repeat gel times in all treatments of the growth experiments compared with the gel times observed for those matched treatments in the 4D characterization campaign may be due to a number of factors including the difference in cell seeding stock (fresh vs frozen), variability in the age or activity of the other hydrogel precursors, or the discrepancy in reaction temperature between the gel time screening campaign and the campaign to study bioactivity in the matched subset of formulas gelling within the targeted 5–15 min window. Specifically, reactions during the 4D characterization campaign were conducted at 30 °C, whereas the bioactivity assays were done without that control. The lower temperature in the bioactivity assays (conducted at room temperature) likely explains the slower gelation observed in those

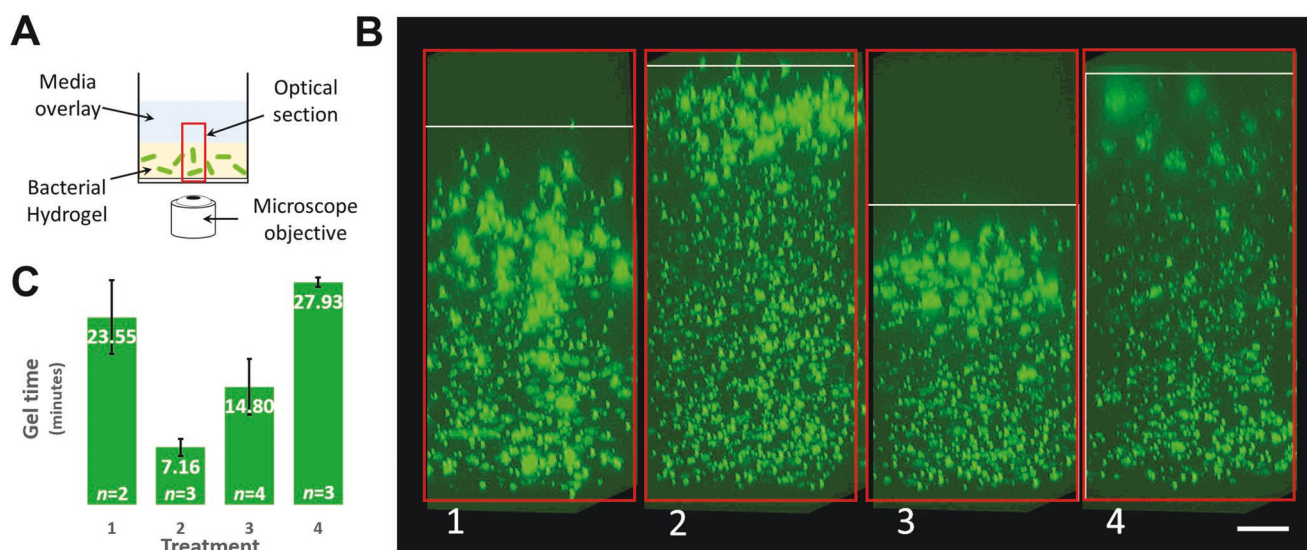


Figure 5. Overnight protein production in 5–15 min silk hydrogel formulas. A) The confocal imaging setup. A relatively thin section of bacteria-laden hydrogel was imaged in an inverted configuration, with images spanning the region beneath the coverglass, through the hydrogel sample, and into the overlying culture medium. B) Representative confocal micrographs of bacterial cells growing near the gel-medium interface for four treatments that were previously observed to gel within 5–15 min. Each confocal image is a side view of the maximum intensity projection of a stack of $80 \times 80 \mu\text{m}^2$ images sampled every $0.5 \mu\text{m}$ over a depth of $200 \mu\text{m}$. Micrograph orientation matches the orientation depicted in (A). Cells positioned toward the top side of the images were progressively closer to the gel-medium interface, and gel sections were chosen such that the top side of each image shows (as a lack of fluorescence) the liquid media overlaying the hydrogel. The white rectangles mark the estimated location of the gelled material, whereas red rectangles illustrate the full stack of images taken through the sample. C) The gel times for each replicated treatment measured by DDM-RL, with gel times reported in minutes and the number of replicates for each treatment indicated. Treatment formulas included 1) 5.31% silk, 175 ppm H_2O_2 , and 10 U mL^{-1} HRP, 2) 3% silk, 50 ppm H_2O_2 , and 18 U mL^{-1} HRP, 3) 2% silk, 125 ppm H_2O_2 , and 20 U mL^{-1} HRP, and 4) 3% silk, 300 ppm H_2O_2 , and 28 U mL^{-1} HRP. Scale bar is $20 \mu\text{m}$.

samples. These discrepancies highlight the potential need to characterize a biomaterial system in a manner that anticipates the intended use as closely as possible. On the other hand, the ability to reproduce gel times in two of the four selection treatments in spite of the experimental differences suggests that certain regions of the bacteria-embedded silk formula space are quite robust.

Biological activity in the treatments was assessed according to three criteria that could be observed microscopically. The first criterion was the relative number of cells, as suggested by the number of *E. coli*-sized fluorescent puncta. The second was the amount of fluorescent protein in the cells, as judged by the brightness of the individual puncta. The third was the amount of growth as measured by the morphology of a colony extending beyond individual *E. coli*-sized puncta.

By visual inspection, cell numbers appeared roughly constant across treatments. As each treatment was seeded with a matched bacterial population, the observation of similar quantities of *E. coli*-sized puncta across treatments suggests no significant differences in reproduction. For each of the four treatments, fluorescence intensity in *E. coli*-sized puncta was high near the glass surface and generally diminished as the optical path length increased. As the availability of oxygen, required for fluorophore maturation, is expected to be higher as the distance from the coverglass increases, the observed attenuation in the fluorescence intensity with increased distance from the coverglass is attributed to optical attenuation in the material. In fact, the clarity of distinct silk hydrogels does differ depending on formula (data not shown), and the

formula with the highest H_2O_2 concentration demonstrated the most significant optical attenuation. The variations in optical attenuations made it impossible to compare the treatments by fluorescence intensity differences alone without complex normalization schemes.

Despite the challenges in comparing treatments based on simple fluorescence intensity, Figure 5 presents evidence that bacterial growth within each of the distinct formulations was qualitatively similar. Near the gel-liquid interface for each formulation, cells grew into sphere-like colonies. The thickness of the region producing the colony formations was similar across treatments. Cells far from the gel-liquid interface did not manifest similar colony morphology. For the two cases which repeatedly gelled within the chosen window of 5–15 min, differences in confocal images are not pronounced. This observation suggests that other material properties, for example stiffness and optical clarity, may be able to be tuned independent of the gel time and biological productivity. Further studies on bacterial growth within hydrogels are needed to study the range in which material properties can be varied while maintaining bacterial activity.

3. Conclusion

Exciting opportunities exist for living biomaterials. Challenges to their development and perfection include the potential vastness of material formulation spaces and potential trade-offs in competing materials and biological properties. Tools to facilitate

rapid screening of large numbers of formula permutations and processing conditions are needed to maximize return on development efforts.

We modified a recently developed tool for microrheology known as DDM to enable a new framework for automated screening of gel times in bacteria-laden silk hydrogels. The last point in the gelation process for which DDM resolves physically meaningful MSD curves was developed as a simple criterion for the onset of gelation. This approach, termed DDM-RL, is shown to correlate approximately with gel time measured by MPT- G'/G'' cross-over and MPT-TCS in hydrogels comprised of enzymatically crosslinked silk. DDM-RL thus provides a supervision-free route to the approximate determination of gelation kinetics which is valuable for automated or high-throughput screens. When greater precision or more comprehensive data is required, TCS, conversion of MSDs from MPT into mechanical moduli, or the determination of individual particle trajectories and displacement distributions can all be performed on the same raw data through MPT. These procedures, however, require an expert user to adjust a number of parameters and to subjectively determine the suitability of the analysis. DDM-RL, though less accurate and yielding less information than MPT-based techniques, can be objectively implemented without expert user interactions and is thus an enabling technique for screening large numbers of samples in automated pipelines.

In hydrogels with added *E. coli* cells, a range of gel times is observed when formula components (silk, HRP, hydrogen peroxide, and bacterial cells) are varied over wide ranges. Within a small subset of treatments that gel within 5–15 min with matched bacterial loads, biological activity as measured by production of fluorescent protein appeared to be roughly the same. This suggests that other material properties, for example stiffness and optical clarity, may be able to be tuned independent of the gel time while maintaining biological activity.

The mix of properties important to the success of biomaterial design efforts, in combination with the potential vastness of formulation spaces, highlights the need for tools to facilitate rapid and intelligent screening. In this regard, we demonstrated that a combination of Bayesian ML models and optimal decision-making under uncertainty was more efficient than an expert user at choosing experimental treatments to characterize a complex formulation space. Interestingly, the data obtained when switching planner responsibility between an ML algorithm and expert user suggests the possibility that cooperative experimentation between ML and experts could result in increased learning in comparison to either agent acting alone.

4. Experimental Section

Silk Hydrogelation: Silk hydrogels were made according to previously published procedures^[22] which were based on di-tyrosine crosslinking by HRP in the presence of hydrogen peroxide. A two-part mixing scheme was employed to control gelation and mixing by separating the HRP from the hydrogen peroxide until the time of mixing. Part A contained

water, silk, cells, and HRP. Part B contained water, silk, cells, and hydrogen peroxide. Tracer beads were added to the silk formulas at 1:700 dilution, resulting in approximately 100–150 beads per $40 \times 40 \mu\text{m}^2$ field of view. Beads were near-IR calibration beads (750 nm excitation) with a diameter of 100 nm, dialyzed against water to remove cytotoxic sodium azide. (F8799, Thermofisher).

Silk was isolated from *Bombyx mori* cocoons according to previously published procedures.^[6a,22] In brief, cocoons were obtained commercially (e.g., Aurora Silk, Portland, Oregon, USA), delaminated and scraped clean of debris, boiled in sodium carbonate in water, soaked in lithium bromide in water, and dialyzed. Final silk concentration was assayed gravimetrically. Typical silk yields were approximately 50 mL of 6.5% w/v silk from 7.5 g of dewormed, delaminated cocoons.

Formulation treatments were prepared from frozen stock aliquots. Silk, HRP, and bacterial cell aliquots were all stored at -80°C at concentrations of roughly 6% silk w/v, 2000 U mL^{-1} HRP, and cells at $10 \times \text{OD}_{600}$ of 1.25 (all in water). Stock hydrogen peroxide (0.65 M, 1 mL 30% w/w H_2O_2 in 14 mL water), was stored at 4°C . Prior to use, a silk aliquot was thawed to room temperature and then lightly spun down. In some cases, a miniscule pellet of precipitated silk would appear on the tube side wall, but precipitated silk was not added to experimental samples.

Throughout this work, cell fractions in the silk formulations are expressed as cells mL^{-1} , based on a conversion of 8×10^8 cells mL^{-1} per OD_{600} unit. Cells were added to silk hydrogels in fractions ranging from ≈ 0.13 – 1.43 OD_{600} equivalents, corresponding to approximately 0.1 – 1.4×10^9 cells mL^{-1} .

Bacterial Culture: *E. coli* BL21 (DE3) was transformed with the pDawn plasmid and used for all experiments involving bacterial viability and induction. pDawn puts expression of dsRed-Express2 under white light illumination control^[34] and carries a resistance marker for kanamycin. Bacterial aliquots for mixing into hydrogels were prepared according to two protocols, depending on the experiment. For experiments characterizing living material gel time in 4D formulations (as shown in Figure 4), a large number of bacterial aliquots were made of overnight bacteria harboring the pDawn plasmid. Bacteria were first shaken in LB at 37°C with kanamycin ($50 \mu\text{g mL}^{-1}$) overnight, and then the overnight culture was spun down and resuspended in PBS (pH 7.4; 8 g NaCl, 0.2 g KCl, 1.42 g Na_2HPO_4 , 0.25 g KH_2PO_4 per liter of water). The suspension was then titrated to an OD_{600} of 1.25 using PBS and then pelleted and resuspended in water to 1/10 volume for a $10 \times$ concentration of OD_{600} of 1.25. Aliquots of 120 μL were placed into a -80°C freezer and stored until use, thawed immediately before use and kept on ice during use. All aliquots were used within one week.

For experiments characterizing biological production of bacteria in silk hydrogels (as shown in Figure 5), cells were first grown overnight in LB with $50 \mu\text{g mL}^{-1}$ kanamycin monosulfate (GoldBio, MO). 400 μL of overnight culture was then added to 20 mL M9 minimal broth (VWR International, LLC; 10 g L^{-1} with 0.24 g MgSO_4 , 0.011 g CaCl_2 , and 2 g glucose), supplemented with $1 \times$ vitamins (100 $\times$ MEM vitamin solution, Thermofisher), $1 \times$ amino acids (50 $\times$ MEM amino acids solution, Thermofisher), and kanamycin again at $50 \mu\text{g mL}^{-1}$. Cells were grown for 4–5 h until an OD_{600} of approximately 1. Cells were then spun down and concentrated in M9 salts (10 g L^{-1} M9 minimal broth base only) to achieve an OD_{600} of 1.25. The cells were then pelleted again and resuspended in 1/10 volume water and placed on ice to achieve a concentration matching bacteria used in the gel time experiments. Cells were kept in the dark during preparation of aliquots in order to keep background fluorescent protein to a minimum. Within 20 min after concentrating into water, cells were added to hydrogel precursor solutions and mixed together. Gels were cast as 100 μL slabs in 8-well chambered coverglass (Lab-Tek, Rochester NY, USA). Approximately 30 min after gel reaction initiation, the samples were overlaid with 600 μL of fresh M9 medium with antibiotics and supplements and placed into an incubator at 37°C . Cells were grown overnight with continuous, incandescent illumination to induce protein expression (dsRed-Express2) from the pDawn plasmid.

DDM Microrheology: Samples ranging in volumes from 100–400 μL were imaged within open (unsealed), 8-well chambered coverglass slides (Lab-Tek, Rochester NY, USA). Note that the volumes required for DDM were the same as those required for MPT but were larger for these experiments for convenience in sample preparation, handling, and imaging in the larger, open-well format. For experiments characterizing the 4D formulation space, the samples were held at 30 $^{\circ}\text{C}$ throughout the course of gelation using a heated stage whereas replicated treatments for the growth and viability study were at room temperature. Images were taken from a focal plane $\approx 100\ \mu\text{m}$ above the coverglass. A typical acquisition framerate was 100 Hz, acquired using a CMOS camera on a Vutara 352 microscope (Bruker, Billerica, MA, USA) with a 60 $\times$ water immersion lens. Tracking bead density was selected to result in approximately 100 beads per frame (per $40 \times 40\ \mu\text{m}^2$ field of view). Beads were carboxylate-functionalized polystyrene of 100 nm diameter, reporting in the near-IR wavelengths (F8799, Thermofisher). Videos of 1000 frames each were captured at several points as each sample gelled. A typical gelation reaction was observed every 2 min over a 30 min timeframe. Intensity fluctuations in each image stack were converted to an image structure function using DDM implemented in MATLAB, as previously reported.^[26b,26c,28] The MSD curves were then extracted from the image structure functions^[26b] using a standard deviation acceptance cutoff of 0.025 a.u. in the plateau regions of the $D(q, \Delta t)$ for each wave vector q . See Section S10 in the Supporting Information for more detail. Additional convenience code was written as part of this work in order to automate the collection and processing of gelation experiments. See Section S11 in the Supporting Information for details.

Gel Time Estimates by MPT and DDM: Gel times were estimated by MPT by two methods: MPT- G'/G'' and MPT-TCS. MPT- G'/G'' consisted of first computing the MSD at each time point,^[26b] using a local power law approximation to compute the G'/G'' as functions of frequency,^[35] and by then identifying the time at which G' exceeds G'' for all frequencies. Gel time estimation by MPT-TCS was done by applying shift factors to the lag times and MSDs, starting from both pre-gel and post-gel time points. Alignment was assumed successful when the MSDs were maximally overlapping and the plots of shift factors versus reaction coordinate were sufficiently linear on a log–log scale. See Section S3, Supporting Information, for detailed discussion.

Neither static nor dynamic MPT errors were corrected for in these data, as static and dynamic error correction methods require sampling closely-matched controls. These controls were not easily obtained in gelling systems, especially when gel opacity was changing as was observed for some of these samples.

For DDM-based gel times as supplied to the ML algorithms, estimates were obtained by adding half of the sampling interval to the time associated with the last MSD curve calculated by DDM. As an example, if a gelation reaction was observed every 2 min and the last physically reasonable MSD curve reported by DDM occurred at min 14, then the gel time supplied to the ML algorithm was 15 min. For the purpose of comparison with other techniques, gel times by DDM-RL were estimated to occur in the time intervening the last resolved MSD and one full observation interval thereafter.

Occasionally, DDM failed to resolve MSDs at some pre-gelation time points but would succeed in later pre-gelation time points. See Panel B of Figure S9, Supporting Information, at time points 2.56 and 3.24 min for an example. The reason for the failure was not apparent but forced the estimation of gel time by the last time point that DDM resolved an MSD rather than the first time point for which it failed to resolve. See S12 in the Supporting Information for more detail.

In some cases, the last or last few MSD curves of a gelling solution would result in physically nonsensical MSD curves. These curves would appear erratic, relatively independent of lag time, and report MSDs several orders of magnitude higher than earlier (temporally adjacent) MSD curves (e.g., see Supporting Information Section S13). These erratic curves likely result from intensity fluctuations due to background

noise or out of plane light which do not correlate with particle motion but nevertheless pass the statistical acceptance criteria.^[26b] Erratic curves were discarded manually by inspection.

A typical iteration of the gel time estimation campaign loop (Figure 3) consisted of four formulations tested in parallel, with each iteration taking approximately 4 h. Each set of four reactions was observed microscopically for 30 min. DDM-RL required approximately 5 min per time point per formula treatment when the images were analyzed using exponential lag time spacing. A typical reaction was observed at 15 time points, with 1000 images taken at each time point. Once gel times were determined, results were supplied to the ML experimental planner which required approximately 10 min to produce four new recommendations for the next iteration.

In contrast, processing by MPT (as was done to produce Figure 2) required approximately 15–20 min per gelling sample, with intensive user interaction for the first 5–10 min. Note that our MPT software also performed dedrifting, calculated G' and G'' , and annotated videos with particle identifications and could be run more quickly without these additional actions. MPT-TCS analysis added more time to the analysis, with manual shifting and plotting requiring approximately 30 min per gelation reaction, for some samples much longer.

Bulk Rheology: Bulk rheology was performed using a parallel plate rheometer with a fixed bottom plate and 25 mm rotating top plate (Discovery HR-2, TA instruments, New Castle, DE, USA). Immediately after mixing, 490 μL of gelling silk solutions were pipetted onto the bottom plate and held in place by surface tension. Temperature was maintained at 30 $^{\circ}\text{C}$ in order to match parallel optical microscopy experiments. Rheology was performed by oscillating at various frequencies ranging from 0.1–10 rad s^{-1} with strain held at 1.25%. Samples were run serially, one gelation reaction per oscillation frequency. Loss angle was plotted over time for each experiment, and the point at which loss angle was invariant with frequency (the point at which each curve intersected) was taken as the gel point.

Identifying Silk Formulas Promoting Maximal Biological Activity: Bacterial colonies were examined with a Bruker Opterra confocal microscope (Bruker, Billerica, MA, USA), imaging directly through the bulk gels. Image stacks were acquired with an $80 \times 80\ \mu\text{m}^2$ field of view, 200 μm thick, sampling every 0.5 μm . 561 nm laser illumination was at 1–5% with an integration time of 100 ms. Embedded bacteria were imaged close to the coverglass and extending 200 μm up through the gel and into the cell culture medium. 3D volumes were projected into 2D using maximum intensity projections.

Machine Learning: After n iterations of the campaign loop, the planner built a GP model $B^n = GP(\mu^n, \Sigma^n)$ of the gelation time $f^*(x)$, where x represented the input parameters of a gelation experiment. Here, $\mu^n(x)$ and $\Sigma^n(x, x')$ represented the time- n mean and covariance functions, respectively, and together, the GP model represented a probability distribution over functions that the authors believed the unknown gelation time response function f^* was sampled from, based on the n experiments performed. These functions could be defined using standard formulas and a prior belief $B^0 = GP(\mu^0, \Sigma^0)$, which was manually specified as

$$\mu^0(x) = c \quad (1)$$

$$\Sigma^n(x, x') = \sigma^0 \exp\left[-\|x - x'\|_2^2 / \ell\right] \quad (2)$$

where $c = 10\ \text{s}$ is a prior estimate of the average gelation time, $\sigma_0 = 5\ \text{s}$ is a prior uncertainty around this initial, constant estimate. The (normalized) length scale $\ell = 0.2$ describes the presumed statistical relationship between the gelation time $f^*(x)$ and $f^*(x')$. See Rasmussen for details.^[36]

With this belief, the GKG decision-making policy was used to select the next set of experiments to run.^[37] Namely, the authors considered the input condition that maximized

$$x_{\text{next}} = \operatorname{argmax}_x V_{\text{CKG}}(x)(\text{AAA}) \quad (3)$$

where

$$V_{\text{GKG}}(x) = \mathbb{E}[\mathcal{L}[f^{n+1}]|x_{\text{next}} = x] - \mathcal{L}[f^n] \quad (4)$$

is the GKG acquisition function. Here:

- $\mathcal{L}[\cdot]$ is a feature operator, that calculates a feature vector from an estimate of the gelation-time response function. Namely, a set of query points $\{x_1, x_2, \dots, x_L\}$ and a bitmap representation predicted optimal region were defined:

$$\mathcal{L}[f^n] = (b_1, b_2, \dots, b_L)^T \quad (5)$$

$$b_i = \begin{cases} 1, & t_{i0} \leq f^n(x_i) \leq t_{hi} \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

- The expected value term $\mathbb{E}[\mathcal{L}[f^{n+1}]|x_{\text{next}} = x]$ is the average value of the feature vector $\mathcal{L}[f^{n+1}]$ that the authors would obtain if they were to a) run the experiment x as the next experiment, b) obtain the output, c) fit the time- $(n+1)$ GP, d) calculate the feature vector from the time- $(n+1)$ GP. It is an average over the result of this *simulation* of what would happen if the authors were to run experiment x . In practice, this quantity was calculated using MC methods. An estimate was obtained by averaging over the results of statistical simulations of what would happen if the authors were to run experiment x . The simulated output response was sampled from the time- n GP, and subsequent calculations were made upon and averaged over 32 sampled responses.

A 4D grid with logarithmic scaling was used to define query points of the GP surrogate model. At each grid point x_i , the surrogate was used to define a label ℓ_i to classify it into one of three discrete classes (gelation regimes) based on whether the surrogate predicted a gelation time of less than 5 min ($\ell_i = -1$), greater than 15 min ($\ell_i = 1$), or between 5 and 15 min ($\ell_i = 0$). The grid point was then transformed to a 2D point $z_i = h(x_i)$ according to the domain-expert defined change in variables. The data set $\{(z_i, \ell_i)\}$ was then used to train a kernel Support Vector Machine (SVM) classifier model using a squared-exponential kernel to obtain a classification of points of the 2D space. The colored regions in the predicted gelation regime map indicate the SVM model's prediction of class at points in the 2D space.

A similar procedure was used to obtain the 2D visualization of model uncertainties shown as the last plot in Figure 4A. At each grid point x_i , the GP model was queried to obtain the standard deviation representing the uncertainty σ_i associated with the model's mean prediction of the gelation time. The data points $\{(z_i, \sigma_i)\}$ were used to train a kernel SVM regressor using a squared-exponential kernel. This SVM model was then used to calculate the projected estimate of the uncertainty in the 2D space.

Error estimates (Figure 4B) were computed from the GP model on the gelation time, $g(x)$. Using the hypervolume as metric, the region corresponding to $5 < g(x) < 15$ min for successive iterates of model fitting (Ω_n) was compared with the region predicted by the final model (Ω_{final}). Relative error was computed as:

$$\text{Relative error}(n) = \frac{|\Omega_{\text{final}} - \Omega_n|}{|\Omega_{\text{final}}|} \quad (7)$$

Likewise, the relative change (Figure 4C) was computed as follows:

$$\text{Relative change}(n) = \frac{|\Omega_{n+1} - \Omega_n|}{|\Omega_n|} \quad (8)$$

where Ω_{n+1} is the predicted hypervolume resulting from several scenarios given a choice of potential experiments.

Note that transformations and the use of SVM models in 2D space were solely for visualization. The decision-making and belief modeling were done in the original 4D space.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was funded by the Synthetic Biology for Military Environments Applied Research for the Advancement of S&T Priorities program of the Office of the Under Secretary of Defense for Research and Engineering. MEH and AVB were supported by the National Science Foundation under Award no. CBET-1729108. The authors thank Jeremy Kemball for assistance with operating the robotic liquid handler.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

biomaterials, differential dynamic microscopy, hydrogels, living materials, microrheology

Received: July 16, 2021

Revised: September 20, 2021

Published online: November 23, 2021

- [1] Y. Zhang, T. Ren, J. He, H. Tian, B. Jin, *Colloids Surf. A* **2019**, 563, 318.
- [2] N. Ostrov, M. Jimenez, S. Billerbeck, J. Brisbois, J. Matragrano, A. Ager, V. W. Cornish, *Sci. Adv.* **2017**, 3, e1603221.
- [3] W. Srifa, N. Kosaric, A. Amorin, O. Jadi, Y. Park, S. Mantri, J. Camarena, G. C. Gurtner, M. Porteus, *Nat. Commun.* **2020**, 11, 2470.
- [4] J. Huang, S. Liu, C. Zhang, X. Wang, J. Pu, F. Ba, S. Xue, H. Ye, T. Zhao, K. Li, Y. Wang, J. Zhang, L. Wang, C. Fan, T. K. Lu, C. Zhong, *Nat. Chem. Biol.* **2019**, 15, 34.
- [5] P. Q. Nguyen, N. D. Courchesne, A. Duraj-Thatte, P. Praveschotinunt, N. S. Joshi, *Adv. Mater.* **2018**, 30, 1704847.
- [6] a) P. H. Chao, S. Yodmuang, X. Wang, L. Sun, D. L. Kaplan, G. Vunjak-Novakovic, *J. Biomed. Mater. Res., Part B* **2010**, 95, 84; b) M. Ciocci, I. Cacciotti, D. Seliktar, S. Melino, *Int. J. Biol. Macromol.* **2018**, 108, 960; c) O. Hasturk, K. E. Jordan, J. Choi, D. L. Kaplan, *Biomaterials* **2020**, 232, 119720; d) W. He, P. Li, Y. Zhu, M. Liu, X. Huang, H. Qi, *New J. Chem.* **2019**, 43, 2213; e) S. Kapoor, S. C. Kundu, *Acta Biomater.* **2016**, 31, 17; f) B. Kundu, N. E. Kurland, S. Bano, C. Patra, F. B. Engel, V. K. Yadavalli, S. C. Kundu, *Prog. Polym. Sci.* **2014**, 39, 251; g) H. Wu, S. Liu, L. Xiao, X. Dong, Q. Lu, D. L. Kaplan, *ACS Appl. Mater. Interfaces* **2016**, 8, 17118; h) Z. Zhu, S. Ling, J. Yeo, S. Zhao, L. Tozzi, M. J. Buehler, F. Omenetto, C. Li, D. L. Kaplan, *Adv. Funct. Mater.* **2018**, 28, 1704757.
- [7] Z. Shao, F. Vollrath, *Nature* **2002**, 418, 741.
- [8] a) M. B. Applegate, G. Perotto, D. L. Kaplan, F. G. Omenetto, *Biomed. Opt. Express* **2015**, 6, 4221; b) N. E. Davis,

- L. N. Beenken-Rothkopf, A. Mirsoian, N. Kojic, D. L. Kaplan, A. E. Barron, M. J. Fontaine, *Biomaterials* **2012**, 33, 6691; c) S. Yodmuang, S. L. McNamara, A. B. Nover, B. B. Mandal, M. Agarwal, T. A. Kelly, P. H. Chao, C. Hung, D. L. Kaplan, G. Vunjak-Novakovic, *Acta Biomater.* **2015**, 11, 27.
- [9] A. Baba, S. Matsushita, K. Kitayama, T. Asakura, H. Sezutsu, A. Tanimoto, T. Kanekura, *J. Biomed. Mater. Res., Part B* **2019**, 107, 97.
- [10] J. Chen, H. Venkatesan, J. Hu, *Adv. Eng. Mater.* **2018**, 20, 1700961.
- [11] C. Vepari, D. L. Kaplan, *Prog. Polym. Sci.* **2007**, 32, 991.
- [12] a) W. Zhang, X. Wang, S. Wang, J. Zhao, L. Xu, C. Zhu, D. Zeng, J. Chen, Z. Zhang, D. L. Kaplan, X. Jiang, *Biomaterials* **2011**, 32, 9415; b) E. Wenk, H. P. Merkle, L. Meinel, *J. Control Release* **2011**, 150, 128; c) D. Yu, C. Sun, Z. Zheng, X. Wang, D. Chen, H. Wu, X. Wang, F. Shi, *Int. J. Pharm.* **2016**, 503, 229.
- [13] a) X. You, J. J. Pak, *Sens. Actuators, B* **2014**, 202, 1357; b) Y. Song, Z. Lin, L. Kong, Y. Xing, N. Lin, Z. Zhang, B.-H. Chen, X.-Y. Liu, *Adv. Funct. Mater.* **2017**, 27, 1700628; c) K. Min, S. Kim, S. Kim, *Proc. Natl. Acad. Sci. U.S.A.* **2017**, 114, 6185.
- [14] R. K. Pal, N. E. Kurland, C. Wang, S. C. Kundu, V. K. Yadavalli, *ACS Appl. Mater. Interfaces* **2015**, 7, 8809.
- [15] a) S. Zhao, Y. Chen, B. P. Partlow, A. S. Golding, P. Tseng, J. Coburn, M. B. Applegate, J. E. Moreau, F. G. Omenetto, D. L. Kaplan, *Biomaterials* **2016**, 93, 60; b) C. J. Bettinger, K. M. Cyr, A. Matsumoto, R. Langer, J. T. Borenstein, D. L. Kaplan, *Adv. Mater.* **2007**, 19, 2847.
- [16] J. Y. Wang, D. Snoeck, S. Van Vlierberghe, W. Verstraete, N. De Belie, *Constr. Build. Mater.* **2014**, 68, 110.
- [17] X. Liu, H. Yuk, S. Lin, G. A. Parada, T. C. Tang, E. Tham, C. de la Fuente-Nunez, T. K. Lu, X. Zhao, *Adv. Mater.* **2018**, 30, 1704821.
- [18] L. M. Gonzalez, N. Mukhitov, C. A. Voigt, *Nat. Chem. Biol.* **2020**, 16, 126.
- [19] S. Sankaran, J. Becker, C. Wittmann, A. Del Campo, *Small* **2019**, 15, 1804717.
- [20] M. Schaffner, P. A. Ruhs, F. Coulter, S. Kilcher, A. R. Studart, *Sci. Adv.* **2017**, 3, eaao6804.
- [21] S. Mondal, P. Das, A. K. Chakraborty, *Mater. Today: Proc.* **2017**, 4, 9833.
- [22] B. P. Partlow, C. W. Hanna, J. Rnjak-Kovacina, J. E. Moreau, M. B. Applegate, K. A. Burke, B. Marelli, A. N. Mitropoulos, F. G. Omenetto, D. L. Kaplan, *Adv. Funct. Mater.* **2014**, 24, 4615.
- [23] a) J. C. Crocker, D. G. Grier, *J. Colloid Interface Sci.* **1996**, 179, 298; b) T. G. Mason, *Rheol. Acta* **2000**, 39, 371; c) Y. Gao, M. L. Kilfoil, *Opt. Express* **2009**, 17, 4685.
- [24] a) V. Breedveld, D. J. Pine, *J. Mater. Sci.* **2003**, 38, 4461; b) E. M. Furst, T. M. Squires, *Microrheology*, Oxford University Press, New York **2017**; c) L. L. Josephson, W. J. Galush, E. M. Furst, *Biomicrofluidics* **2016**, 10, 043503; d) K. M. Schultz, A. V. Bayles, A. D. Baldwin, K. L. Kiick, E. M. Furst, *Biomacromolecules* **2011**, 12, 4178.
- [25] a) W. Hong, G. Xu, X. Ou, W. Sun, T. Wang, Z. Tong, *Soft Matter* **2018**, 14, 3694; b) T. Larsen, K. Schultz, E. M. Furst, *Korea-Aust. Rheol. J.* **2008**, 20, 165; c) T. H. Larsen, E. M. Furst, *Phys. Rev. Lett.* **2008**, 100, 146001; d) T. Moschakis, B. S. Murray, E. Dickinson, *J. Colloid Interface Sci.* **2010**, 345, 278; e) K. M. Schultz, E. M. Furst, *Soft Matter* **2012**, 8, 6198.
- [26] a) A. V. Bayles, T. M. Squires, M. E. Helgeson, *Soft Matter* **2016**, 12, 2440; b) A. V. Bayles, T. M. Squires, M. E. Helgeson, *Rheol. Acta* **2017**, 56, 863; c) R. Cerbino, V. Trappe, *Phys. Rev. Lett.* **2008**, 100, 188102.
- [27] R. Cerbino, D. Piotti, M. Buscaglia, F. Giavazzi, *J. Phys.: Condens. Matter* **2018**, 30, 025901.
- [28] F. Giavazzi, D. Brogioli, V. Trappe, T. Bellini, R. Cerbino, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2009**, 80, 031403.
- [29] P. Edera, D. Bergamini, V. Trappe, F. Giavazzi, R. Cerbino, *Phys. Rev. Mater.* **2017**, 1, 073804.
- [30] J. A. Richards, V. A. Martinez, J. Arlt, *Soft Matter* **2021**, 17, 3945.
- [31] a) M. Caggioni, P. T. Spicer, D. L. Blair, S. E. Lindberg, D. A. Weitz, *J. Rheol.* **2007**, 51, 851; b) A. M. Corrigan, A. M. Donald, *Eur. Phys. J. E: Soft Matter Biol. Phys.* **2009**, 28, 457.
- [32] J. M. Newby, A. M. Schaefer, P. T. Lee, M. G. Forest, S. K. Lai, *Proc. Natl. Acad. Sci. U.S.A.* **2018**, 115, 9026.
- [33] S. Chen, K.-R. G. Reyes, M. K. Gupta, M. C. McAlpine, W. B. Powell, *SIAM/ASA J. Uncertainty Quantif.* **2015**, 3, 320.
- [34] R. Ohlendorf, R. R. Vidavski, A. Eldar, K. Moffat, A. Moglich, *J. Mol. Biol.* **2012**, 416, 534.
- [35] M. A. Escobedo-Sanchez, J. P. Segovia-Gutierrez, A. B. Zuccolotto-Berneiz, J. Hansen, C. C. Marciniak, K. Sachowsky, F. Platten, S. U. Egelhaaf, *Soft Matter* **2018**, 14, 7016.
- [36] C. E. Rasmussen, in *Summer School on Machine Learning*, Springer, Berlin **2003**.
- [37] K. G. Reyes, F. J. Alexander, in *Handbook on Big Data and Machine Learning in the Physical Sciences*, Vol. 2 (Eds: K. Kleese van Dam, K. G. Yager, S. I. Campbell, R. Farnsworth, M. vanDam), **2020**, Ch. 13.