Computer-aided design of elastin-like polypeptides with controlled viscoelastic and structural properties

Diego López Barreiro,[1] Abel Folch-Fortuny,[2] Iain Muntz,[3]
Jens C. Thies,[4, *] Cees M. J. Sagt,[1, †] and Gijsje H. Koenderink[3, ‡]

[1] *DSM Bioscience and Process Innovation, DSM,*
Alexander Fleminglaan 1, 2613 AX, Delft, The Netherlands
[2] *DSM Biodata and Translation, DSM, Alexander Fleminglaan 1, 2613 AX, Delft, The Netherlands*
[3] *Department of Bionanoscience, Kavli Institute of Nanoscience Delft,*
Delft University of Technology, van der Maasweg 9, 2629 HZ, Delft, The Netherlands
[4] *DSM Biomedical, DSM, Urmonderbaan 22, 6160 BB, Geleen, The Netherlands*

Abstract. The biofabrication of structural proteins with controllable properties via amino acid sequence design is interesting for biomedicine and biotechnology, yet design rules that link amino acid sequence to material properties remain largely unknown. Molecular dynamics (MD) simulations can aid in unveiling such rules, but the lack of a standardised framework to interpret the outcome of those simulations hinders their predictive value for the design of *de novo* structural proteins. To address this, we developed a model that unambiguously classifies a library of *de novo* elastin-like polypeptides (ELPs) with varying numbers and locations of hydrophobic/hydrophilic and physical/chemical-crosslinking blocks according to their thermoresponsiveness at physiological temperature. Our approach does not require long simulation times or advanced sampling methods. Instead, we apply (un)supervised data analysis methods to a dataset of molecular properties from relatively short MD simulations (150 ns). We also investigate the rheological properties and microstructure of ELP hydrogels, revealing several handles to tune them: chain hydrophilicity/hydrophobicity or block distribution control the viscoelasticity and thermoresponsiveness, whereas ELP concentration defines the network permeability. Our findings provide an avenue to accelerate the design of *de novo* ELPs with bespoke material properties.

INTRODUCTION

Many natural materials like bone, silk cocoons, mussel threads, or insect wings, consist of a self-assembling biopolymeric scaffold formed by structural proteins such as collagen, elastin, silk fibroin, or resilin [1]. These structural proteins display impressive mechanical, structural, and biological traits from a materials science perspective, such as the ability to combine toughness and strength in a lightweight material, to interact with the environment, or to promote cell adhesion [2, 3]. Those traits arise from the collective effects of nanoscale features encoded in the amino acid sequence of structural proteins, such as the frequency and distribution of ordered/disordered, hydrophobic/hydrophilic, charged/uncharged, or chemical/physical crosslinking blocks [4].

Structural proteins have inspired material scientists to manufacture multifunctional materials for fields including medicine [5], energy harvesting [6], and biosensing [7]. However, harvesting structural proteins from nature is inefficient and constrains us to using only biopolymers selected for by evolution [8]. Luckily, developments in bioprocess engineering and molecular and synthetic biology enable the use of microbial cultures for the biofabrication of *de novo* structural proteins [3, 9]. This allows us to rationally design new protein sequences that encode well-defined chain topologies and interactions to control the mechanical, structural, and biological features of these proteins.

Elastin is one of the most studied structural proteins [10]. It is a key component of the extracellular matrix of mammals, conferring elasticity to tissues like blood vessels, skin, or lungs [11]. Its precursor, tropoelastin, is rich in hydrophobic VPGVG blocks and displays a lower critical solution temperature (LCST) behaviour: it is water-soluble below the LCST, but it coacervates and phase-separates above it [12]. This has inspired the development of elastin-like polypeptides (ELPs) [13] for the manufacture of dynamic and thermoresponsive materials in fields like tissue engineering [14], drug delivery [15], microfluidics [16], and actuation [17]. Like tropoelastin, ELP solutions also display LCST behaviour [18, 19]. ELPs are intrinsically disordered and highly dynamic both below and above the LCST [20, 21]. Above the LCST, ELP–ELP interactions become favoured over ELP-solvent interactions, and thus the solution segregates into ELP-rich and solvent-rich phases [21]. The properties of ELP-based materials can be controlled by sequence design (e.g., block arrangement or molecular weight), or by processing conditions (e.g., protein concentration or solution conditions) [22], and numerous studies have shown the ability of ELPs to form a rich landscape of sophisticated nanostructures above the LCST [23–25].

ELPs typically consist of repetitions of the pentapeptide building block VPGXG [18, 19], where X is referred to as the guest amino acid and can be any amino acid except proline [13]. *De novo* ELPs are designed fol-

---

* Jens.Thies@dsm.com
† Cees.Sagt@dsm.com
‡ G.H.Koenderink@tudelft.nl

lowing a modular approach, in which variations of the VPGXG building block are encoded into a polypeptide chain that is produced via microbial fermentation. Early research focused on elucidating the molecular origins of the elasticity and LCST behaviour of ELPs [19, 26–28] and on characterising selected ELP for the manufacture of films [29], fibres [23], nanoparticles [30], or hydrogels [31]. This unveiled the influence of parameters like guest amino acid [32] or polypeptide molecular weight (MW) [33]. Yet, precise heuristics that link sequence design to macroscopic dynamic, mechanical, or structural features for *de novo* ELPs remain largely unknown. It was only recently that some studies started uncovering such sequence–property relationships [34–37]. To that end, molecular dynamics (MD) simulations can accelerate the design of *de novo* ELPs by rapidly and simultaneously screening multiple permutations of building blocks [38]. This holds promise to predict which sequences might display interesting material traits before synthesising them.

MD simulations have been used to study the elasticity and LCST of ELPs [20, 26, 28, 35]. However, the predictive power of MD simulations for the design *de novo* ELPs is yet to be fully harnessed. It is generally stated that simulation times in the microsecond range [20] or advanced sampling methods like replica-exchange [35] are needed to attain statistically converged sampling of ELPs via MD simulations. Furthermore, the lack of a standardised framework to interpret the outcome of MD simulations has hindered their use for the design of *de novo* ELPs. However, several studies have claimed that MD simulations in the 10-100 ns range can capture the LCST of different homopolymeric or multiblock ELPs. This was achieved by monitoring molecular properties like radius of gyration, intra-ELP hydrogen bonds, secondary structure, or water molecules in the hydration layer of ELPs [20, 26, 28, 35, 39–41], all of which could reflect the molecular collapse typically associated with the LCST behaviour of ELPs [35].

Here we propose a different approach to develop a predictive computational framework for the LCST behaviour of *de novo* ELPs. Our work draws inspiration from quantitative structure-activity relationships (QSAR) and quantitative structure-property relationships (QSPR) approaches [42, 43], and uses relatively short (150 ns per replica) fully atomistic MD simulations. We validated the model using a library of eight *de novo* ELPs, which sweeps varying numbers and locations of hydrophobic/hydrophilic and physical/chemical crosslinking blocks. Two clusters were identified among our library based on principal component analysis (PCA) of a set of 27 molecular properties sampled from MD simulations. These clusters, but not the individual molecular properties, coincided with the presence or absence of LCST behaviour at 37 °C, as determined experimentally. Subsequently, we developed a regression model that accurately discriminated *de novo* ELPs according to their LCST behaviour based on data from MD simulations. We also investigated the rheological properties,

secondary structure, and supramolecular microstructure of our ELP library using experiments and theory. This allowed us to bridge sequence-level features with nano-, micro-, and macroscopic structural and mechanical properties, providing an avenue for the rational design of *de novo* ELPs with bespoke material properties.

RESULTS

Predicting the LCST behaviour of an ELP library

A library of eight *de novo* ELPs with size-matched MW (34.5-35.6 kDa) was designed (Table 1). All ELPs contained 48 repeats of the VPGVG building block, while the following molecular modulators were varied: guest amino acid, location of the crosslinking blocks, and identity of the crosslinking block. For a given MW, the LCST can be tuned by changing the amino acid in the X position [19, 32]. We hence compared a nonpolar guest amino acid (isoleucine, I) versus an ionic one (glutamic acid, E). Homopolymeric ELPs are known to form coalescing aggregates above the LCST, which segregate from the aqueous phase [36]. To prevent segregation and promote the formation of a bicontinuous hydrogel network, we used physical (IPAVG, GAGAGS) or chemical (VPGKG) crosslinking blocks. IPAVG blocks in E_{AE}, E_{AI}, $E_{AE,triblock}$, and $E_{AI,triblock}$ can form kinetically arrested gels [44]. GAGAGS blocks in SE_E and SE_I were based on β-sheet-forming blocks from silk fibroin. These blocks can arrest ELP segregation by forming β-sheet structures [31, 37]. Finally, VPGKG blocks were introduced in E_{KE} and E_{KI} to form covalent bonds between ELP chains via lysine (K) residues upon addition of glutaraldehyde [24]. To test how the positioning of crosslinking blocks affects the LCST behaviour and hydrogel functionality, we investigated multiblock (E_{AE} and E_{AI}) and triblock arrangements ($E_{AE,triblock}$ and $E_{AI,triblock}$). The effect of these modulators can be visualised by the Kyte-Doolittle hydropathy scale [45] of these ELPs (Figure 1a).

Previous research suggested that ELPs with as few as 10 VPGXG blocks manifested LCST behaviour in MD simulations [41]. Therefore, to reduce the computational cost of our MD simulations, we studied the ability to undergo LCST at 37 °C of shorter multimers of our library equivalent to ELPs with 35 VPGXG blocks (Table S1). Implicit solvent simulations were first run, starting with the fully extended ELP chain, to achieve a folded structure (Figure 1b). The resulting structure was then solvated in explicit water and, when needed, its electrical charges were adjusted to achieve charge neutrality. The solvated system was then run in triplicate for 150 ns more for each ELP. After 100 ns, ELPs were stable in terms of root mean square deviation (RMSD) of atomic positions (Figure S1-S2), indicating that the system was sufficiently equilibrated to begin data sampling. Thereafter, we sampled 27 different molecular properties every
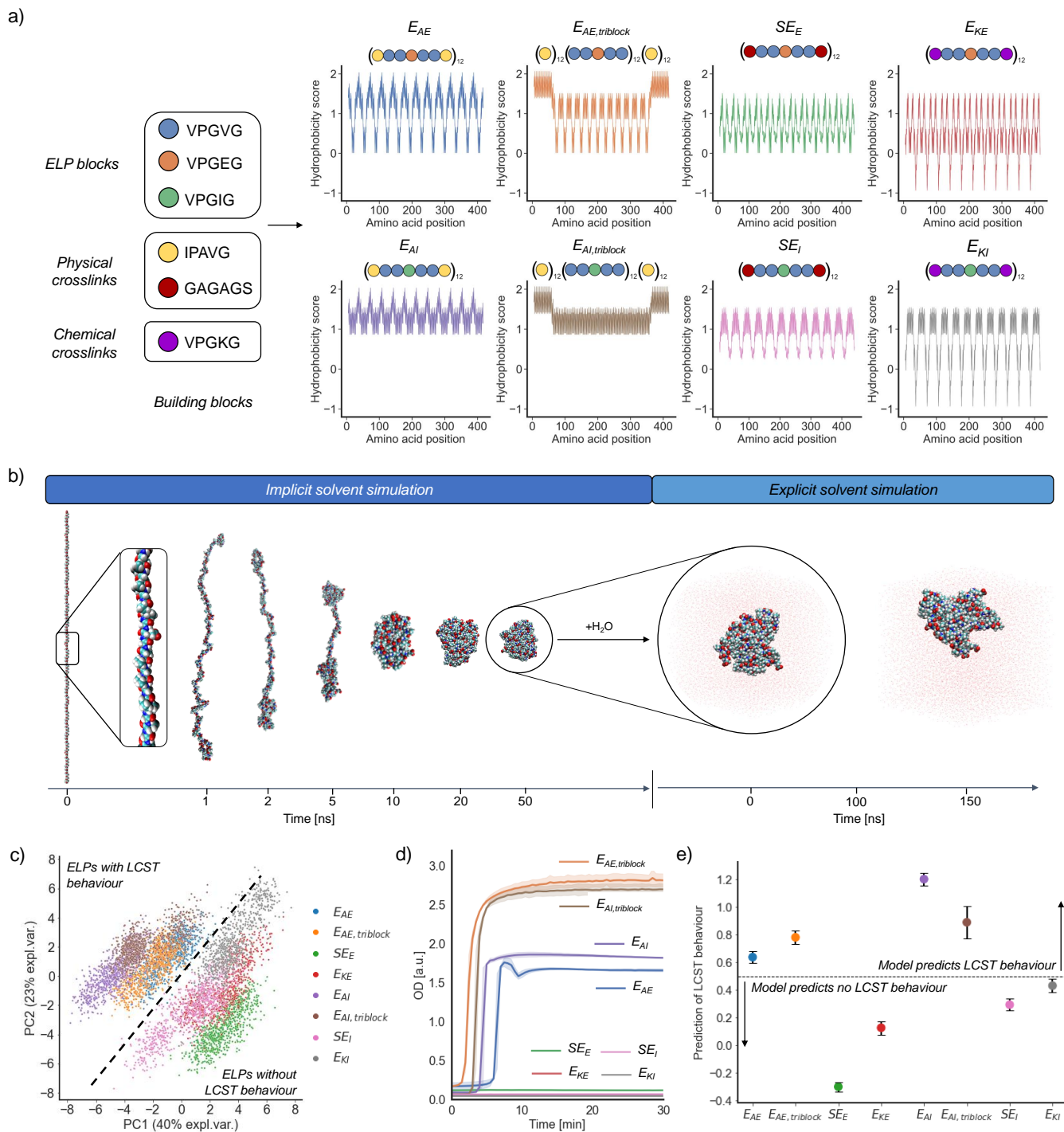
Figure 1. **ELP library design, MD simulations, and data analysis. a.** Schematic representation of the distribution of building blocks along the sequences of our ELP library, together with their hydropathy plots as calculated using the Kyte-Doolittle scale. **b.** Screenshots of the folding process (in implicit solvent) and dynamic run (in explicit solvent) of MD simulations for $E_{AE,triblock}$. **c.** Principal component analysis (PCA) scores plot. Each dot represents a timepoint from the last 50 ns of MD simulations (n=3) on principal component 1 (x-axis) and 2 (y-axis). A dashed line separates the two clusters detected. **d.** Time evolution of the optical density for ELP solutions in milliQ water (15 wt %) at 37 °C (n=2). **e.** Predictions of the LCST behaviour of our ELP library at 37 °C (as observed via optical density) with a partial least-squares regression discriminant analysis (PLS-DA) model that uses molecular properties sampled from MD simulations as input data. The value on the y-axis indicates the prediction of the LCST behaviour for each ELP. When the prediction is above 0.5 (dotted line), the model predicts LCST behaviour. The error bars represent the 99% confidence intervals from the LCST predictions per ELP during the double cross validation procedure.

Table 1. Name, amino acid sequence, and theoretical and real (determined via LC-MS) MW (in kDa) of the ELP library.

Name	Sequence	Theoretical MW	Experimental MW*
E_{AE}	[(IPAVG)(VPGVG)$_2$(VPGEG)(VPGVG)$_2$(IPAVG)]$_{12}$	35.6	35.4
$E_{AE,triblock}$	(IPAVG)$_{12}$[(VPGVG)$_2$(VPGEG)(VPGVG)$_2$]$_{12}$(IPAVG)$_{12}$	35.6	35.4
SE_E	[(GAGAGS)(VPGVG)$_2$(VPGEG)(VPGVG)$_2$(GAGAGS)]$_{12}$	34.7	34.6
E_{KE}	(VPGKG)$_{12}$[(VPGVG)$_2$(VPGEG)(VPGVG)$_2$]$_{12}$(VPGKG)$_{12}$	35.6	35.5
E_{AI}	[(IPAVG)(VPGVG)$_2$(VPGIG)(VPGVG)$_2$(IPAVG)]$_{12}$	35.6	35.3
$E_{AI,triblock}$	(IPAVG)$_{12}$[(VPGVG)$_2$(VPGIG)(VPGVG)$_2$]$_{12}$(IPAVG$_{12}$	35.4	35.3
SE_I	[(GAGAGS)(VPGVG)$_2$(VPGIG)(VPGVG)$_2$(GAGAGS)]$_{12}$	34.5	34.4
E_{KI}	(VPGKG)$_{12}$[(VPGVG)$_2$(VPGIG)(VPGVG)$_2$]$_{12}$(VPGKG)$_{12}$	35.4	35.3

*the error associated with this measurement is 0.5-1.0 Da   .

0.2 ns for the next 50 ns. This yielded 250 data points per molecular property and per replica. These properties include intrapeptide hydrogen bonding, radius of gyration, RMSD, solvent accessible surface area (SASA), secondary structure, or water molecules in the hydration layer, among others. Their selection was based on prior studies showing their connection with the LCST behaviour of specific ELP sequences [20, 35, 39, 41]. A full list of all the molecular properties sampled can be found in Table S2.

Individually, these 27 molecular properties did not reveal significant differences among the ELPs in our library that would allow us to identify an LCST behaviour (Figures S3-S6). Inspired by QSAR and QSPR approaches [42, 43], where a set of molecular descriptors (the molecular properties obtained from MD simulations) are connected with a property (presence or absence of LCST), we decided to analyse these molecular properties collectively, exploring their multidimensional variable space via PCA. By analysing all data points, rather than averaging them, we increased the statistical accuracy of our study, while accounting for temporal fluctuations in the molecular properties during the simulation. PCA summarises the main features of a high-dimensional multivariate dataset (the output of our simulations) into a few artificial variables, called principal components (PCs). We found that the first two PCs accounted for 63% of the variance observed in the 27-dimensional dataset. As shown in Figure 1c, where each dot corresponds to a time point in the MD simulations, the first two PCs divided our ELPs into two distinct clusters: E_{AE}, $E_{AE,triblock}$, E_{AI}, and $E_{AI,triblock}$ (all of which contain IPAVG blocks) on the one hand, and SE_E, E_{KE}, SE_I, and E_{KI} (which contain GAGAGS or VPGKG blocks) on the other hand.

We then explored the connection between the PCA clusters and the experimental LCST behaviour of our library. ELPs were heterologously expressed in *E. coli*. Purification was performed by inverse temperature cycling [46]. The correct transformation of plasmids containing the ELP genes and the purity of the recovered ELPs were confirmed by agarose (Figure S7) and sodium dodecyl sulphate polyacrylamide gel electrophoresis (Figure S8), respectively. The resulting ELPs were within 2% of the expected amino acid composition, as determined by total amino acid analysis (Table S3). The correspon-

dence between the theoretical and experimental MW was confirmed by intact LC-MS (Table 1). LCST is typically accompanied by a turbidity change for ELP solutions [28]. Thus, we prepared cold (4 °C) 15 wt % ELP solutions in milliQ water and monitored their optical density when placed in a microplate reader at 37 °C (Figure 1d). Interestingly, ELPs divided into the same clusters shown by PCA: E_{AE}, E_{AI}, $E_{AE,triblock}$ and $E_{AI,triblock}$ polypeptides showed LCST behaviour, whereas SE_E, E_{KE}, SE_I, and E_{KI} polypeptides did not. This suggested that the collective evaluation via PCA of multiple molecular properties obtained from short MD simulations (150 ns) may be a suitable tool to predict the LCST behaviour of ELPs.

PCA provides an unsupervised exploratory model but does not make any predictions about the significance of the two clusters identified. To speed up the development of *de novo* ELPs, we also need to be able to predict the LCST behaviour of ELPs not used to build the model. To address this, we switched to a supervised model in which we included the experimental data from turbidity measurements. This aimed at identifying the combination of MD molecular properties in our dataset that discriminated ELPs based on their LCST behaviour at 37 °C. We did so by applying a partial least squares regression discriminant analysis (PLS-DA) model [47, 48]. PLS-DA is a suitable approach for the analysis of a high-dimensional dataset (in this case, our 27 molecular properties) with a binary outcome (presence or absence of LCST). The principle behind PLS-DA methods is similar to PCA, but maximising covariance between the dataset of molecular properties and their LCST outcome, instead of maximising the variance observed within the molecular properties only. This generated a model that summarised both the molecular properties obtained via MD simulations and predicted the outcome (presence or absence of LCST).

The presence or absence of LCST behaviour at 37 °C was coded in the PLS-DA model as 1 and 0, respectively. To develop our PLS-DA model, we applied a double cross validation procedure. In total, each simulation was predicted 7 times, using 6 simulations to build the model and 1 to validate it. When an ELP simulation was left out of the PLS-DA model and then predicted, a numeric value was produced (Figure 1e). When that value was above 0.5 [48], the model predicted that the ELP displayed

LCST behaviour at 37 °C. The predictions for each simulation were combined in 99% confidence intervals. When the confidence intervals do not cross the boundaries of the correct class (LCST behaviour yes/no), the experiment is properly predicted. This approach mimics real-world design-build-test-learn cycles [49], where a model is first built with the current knowledge, then a new experiment is designed using that knowledge and predicted through the model, and finally the predictions are validated experimentally. When leaving E_{AE}, $E_{AE,triblock}$, E_{AI}, and $E_{AI,triblock}$ out of the model and then using the model to predict their LCST behaviour, the 99% confidence interval was entirely above 0.5. In turn, the same interval for SE_E, E_{KE}, SE_I, and E_{KI} was entirely below 0.5. The regression coefficients of the PLS-DA model can be found in Figure S9 and Table S4. Thus, the PLS-DA model agreed with the PCA and turbidity data, attaining a 100% prediction accuracy.

To sum up, MD studies of ELPs typically assess molecular properties individually or in pairs, which has helped to establish the links between some molecular properties (e.g., hydrophobic or hydrophilic SASA) and LCST behaviour. However, this usually required long simulation times or computationally expensive advanced sampling methods [20, 35], and the reported results were typically associated with large uncertainties. Instead, here we achieved an unambiguous separation between the ELPs that undergo LCST at 37 °C and those that do not by collectively analysing 27 molecular properties (via PCA and PLS-DA) using relatively short simulation times in explicit solvent (150 ns). This finding hints that MD simulations hold great potential to rapidly screen and predict qualitative trends in the LCST behaviour of ELPs.

Rheological characterisation of the ELP library

The computational and turbidity data suggest that hydrophobic IPAVG blocks are instrumental in triggering the LCST at 37 °C in our ELP library, even in spite of the ionic guest residues in E_{AE} and $E_{AE,triblock}$ that normally hinder gelation [19, 24]. To get further insight into the sequence-property relationships, we performed rheological measurements. The LCST of our ELP library was determined as the temperature at which ELP solutions underwent a sharp increase in their storage (G') and loss (G'') moduli during a temperature sweep from 4 to 37 °C (f=1 Hz, γ=0.3 %) (Figure 2a). Such an increase was only observed for E_{AI}, $E_{AE,triblock}$, and $E_{AI,triblock}$ (which formed stiff physical hydrogels) but not for $E_{AE,triblock}$. This shows that LCST behaviour as measured by optical density does not necessarily indicate the ability of an ELP solution to form a hydrogel network capable of supporting mechanical stress. The formation of chemical hydrogels via interchain covalent bonds was also assessed for E_{KE} and E_{KI} by adding glutaraldehyde (40:1 glutaraldehyde:ELP molar ratio). Without glutaraldehyde, neither of these ELPs gelled. But upon adding glutaraldehyde, both formed a gel (Figure S10), reaching steady state after 2 h (Figure S11). Chemical hydrogels at 37 °C lacked the turbidity typically associated with LCST, indicating that their gelation originated from network formation via covalent bonds, rather than from an LCST phenomenon.

The value of the LCST was affected by the sequence composition and block distribution. $E_{AI,triblock}$ switched at 18 °C from a solution to a stiff hydrogel. A similar transition was observed for $E_{AE,triblock}$, but at a higher temperature (24 °C), due to its ionic guest amino acid (glutamic acid). In contrast, the moduli of E_{AI} increased gradually between 25 and 30 °C. We attribute this to the ability of the non-IPAVG regions in E_{AI} to hydrophobically interact with and hinder the self-interactions of the short IPAVG blocks. Early research had shown that a multiblock distribution of hydrophobic blocks in ELPs with the same MW raised the LCST [50]. Here we demonstrate that a multiblock distribution also weakens the mechanical properties of the network.

Sequences with different physical crosslinking blocks (IPAVG vs GAGAGS) but the same block distribution (i.e., E_{AE} vs SE_E; or E_{AI} vs SE_I) also showed differences in thermoresponsiveness. While sequences with IPAVG blocks changed their turbidity in the 4-37 °C range (E_{AE} and E_{AI}), equivalent sequences with silk-like blocks (SE_E and SE_I) did not. An increase in moduli did not happen for SE_E and SE_I even at extended incubation times (4 h) at 37 °C (Figure S12). This observation indicates that the increased hydrophilicity and lack of thermoresponsiveness of GAGAGS blocks impeded the gelation of SE_E and SE_I. We found the lack of LCST of SE_I surprising, given the hydrophobicity of its guest amino acid isoleucine. The inclusion of silk-like blocks is a common strategy to form durable hydrogels, due to the ability of silk-like blocks to physically associate into thermodynamically stable β-sheets [31, 37]. However, our computational and experimental data showed that SE_E and SE_I do not undergo LCST at 37 °C. This differs from previous work that showed the ability of ELPs with silk-like blocks and a similar MW (though with larger elastin:silk ratios than in our study) to gel at 37 °C [37]. A recent study namely proposed that the aggregation of the ELP regions is a necessary first step to bring silk-like blocks in close proximity to form β-sheets [37]. Thus, we hypothesise that the length of ELP regions (5 VPGXG blocks) between GAGAGS blocks in our library was insufficient to trigger their aggregation.

We also assessed the reversibility of the gelation of these ELP solutions over three temperature cycles between 4 and 37 °C, using a resting time of 30 min between each temperature ramp. The networks disassembled upon cooling for the physical hydrogels (E_{AI}, $E_{AE,triblock}$ and $E_{AI,triblock}$), but not for the chemical ones (Figure 2b). Noteworthy differences were identified for compositionally identical ELPs but with building blocks scrambled in a different manner. For instance, E_{AI} and $E_{AI,triblock}$ formed physical hydrogels
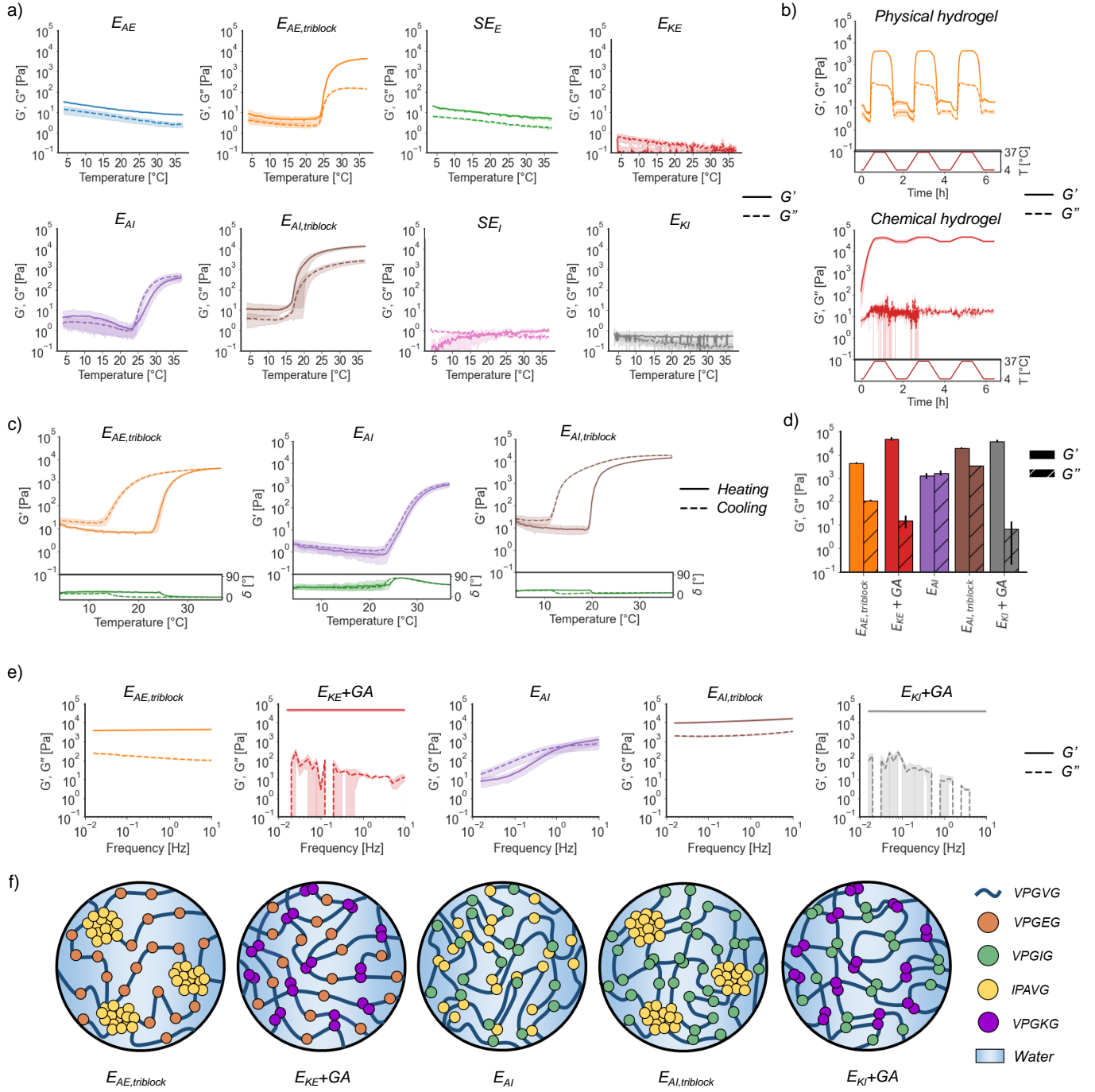
Figure 2. **Rheological properties of the ELP library. a.** Rheological characterisation (storage modulus G' and loss modulus G'') of ELP solutions in milliQ water (15 wt %) during a temperature sweep between 4 and 37 °C (n=4). **b.** Thermal cycling of ELP solutions (15 wt %) at a heating/cooling rate of 1 °C/min (f=1 Hz, γ=0.3%), with 30 min of resting time between temperature ramps: representative reversible physical hydrogel from $E_{AE,triblock}$ (top) and irreversible chemical hydrogel from E_{KE} after glutaraldehyde addition (bottom) (n=2). **c.** Hysteresis of ELP solutions forming physical hydrogels over three temperature cycles (n=2). **d.** Plateau G' and G'' of hydrogels obtained from ELP solutions in milliQ water (15 wt %) at 37 °C (f=1 Hz, γ=0.3%) (n=2). **e.** Frequency sweeps (0.01-10 Hz) of ELP solutions in milliQ water (15 wt %) (T=37 °C, γ=0.3%) (n=2). **f.** Schematic representation of network formation above the LCST for the different ELP designs.

at 37 °C, yet with different hysteresis behaviour (Figure 2c and S13). Hysteresis is usually absent in ELPs with VPGXG blocks [40], but triblock arrangements in our library did display hysteresis. We attribute this behaviour to IPAVG blocks, which are known to have hysteresis between solvation and desolvation [23, 24, 44]. However, the presence of IPAVG blocks was a necessary condition, but not sufficient, to cause hysteresis: $E_{AI,triblock}$ and $E_{AE,triblock}$ showed hysteresis upon cooling, whereas E_{AI} did not. Multiblock arrangements of IPAVG blocks increased the LCST, prevented hysteresis and formed weaker networks: $E_{AI,triblock}$ networks were one order of magnitude stiffer (G'=20000±1100 Pa) than those from E_{AI} (G'=1300±400 Pa), despite having the same number of IPAVG blocks.

The linear viscoelastic (LVE) region of these hydrogels was determined via amplitude sweeps at 37 °C (Figure S14). Triblock designs and chemical hydrogels showed a linear response over the entire range of strain amplitudes (0.01-15%). In contrast, the LVE region for E_{AI} only extended up to strains of $\sim$ 3% and the moduli started decreasing with strain thereafter. We also determined the steady state moduli of these networks after 30 min at 37 °C (f=1 Hz, γ=0.3 %) (Figure 2d). For physical hydrogels, G' was 1300±400 Pa (E_{AI}), 4500±400 Pa ($E_{AE,triblock}$) and 20000±1100 Pa ($E_{AI,triblock}$), whereas G'' was 1700±400 Pa (E_{AI}), 110±10 Pa ($E_{AE,triblock}$) and 3400±200 Pa ($E_{AI,triblock}$). Chemical hydrogels showed little effect of the amino acid sequence, with comparable values for G' (47440±7531 for E_{KE} and 38690±5069 for E_{KI}) or G'' (16±9 for E_{KE} and 7±6 for E_{KI}). This finding highlights that the covalently crosslinked macromolecular network has a larger influence on the mechanical properties than the amino acid sequence. Frequency sweeps (Figure 2e and S15) revealed that the relaxation spectra varied across the ELP library. Triblock designs and chemical hydrogels showed an almost frequency-independent behaviour for G' and G'' (though the determination of G'' was noisy for chemical hydrogels due to their highly elastic character). In turn, E_{AI} showed a crossover indicative of stress relaxation at $f \sim$ 2 Hz.

The effect of the hydrophilic/hydrophobic character of the guest amino acid (E *vs* I) on the viscoelastic properties and LCST was discerned by comparing $E_{AE,triblock}$ and $E_{AI,triblock}$ hydrogels. The steady state G' for $E_{AI,triblock}$ was 4.6 times larger than for $E_{AE,triblock}$. However, the steady state G'' for $E_{AI,triblock}$ was 29.5 times larger than that of $E_{AE,triblock}$. As mentioned before, the nonpolar mid-block of $E_{AI,triblock}$ can likely more easily mix with IPAVG end-blocks. This could promote dissipative interchain hydrophobic contacts. In contrast, the ionic character of VPGEG blocks in $E_{AE,triblock}$ could enhance the aggregation of IPAVG end-blocks, preventing mixing between mid- and end-blocks and reducing the viscous response.

We expected ionic ELPs in our library (E_{AE}, $E_{AE,triblock}$, E_{KE}, and E_{KI}) to form weaker gels or to not gel at all [35], because the self-assembly of ionic polymers is typically hindered by electrical repulsion and generally requires that those repulsions are overcome by pH adjustment or by adding counterions [51]. Indeed, E_{KE} and E_{KI} only gelled after the formation of a macromolecular network upon the addition of glutaraldehyde. In turn, E_{AE} and $E_{AE,triblock}$ did show LCST in the 4-37 °C range (as shown by turbidity data), but with different mechanical outcomes: $E_{AE,triblock}$ gelled and E_{AE} did not. The already discussed differences between multiblock and triblock IPAVG arrangements were likely exacerbated in E_{AE} and $E_{AE,triblock}$ by the presence of an ionic guest amino acid. It has been proposed that high concentrations of ionic ELPs increase the electrostatic repulsion between chains. This can force some hydrophobic blocks (e.g., IPAVG blocks) to become solvent-exposed, triggering the formation of a hydrogel network [51]. According to this mechanism, the higher charge density in the midblock of $E_{AE,triblock}$ could facilitate its gelation. In turn, the more distributed ionic blocks in E_{AE}, together with the reduced length of its IPAVG blocks, would hinder IPAVG-IPAVG associations needed to form the hydrogel network.

Collectively, these observations allowed us to develop an understanding of the different networks that is schematically summarised in Figure 2f. Overall, the highest moduli were obtained for chemical hydrogels. Nonetheless, the viscoelastic properties of physical hydrogels were remarkably high for triblock designs, in the range of tissues like lung, muscle, cartilage, or kidney [52]. This indicates that a rational design of the ELP sequence can deliver materials with mechanical properties in the order of magnitude of chemical hydrogels, but solely relying on physical interactions and without the need of crosslinking agents. This adds to other advantages of physical hydrogels, such an stimuli-responsiveness, self-healing properties, and tailored viscoelasticity [53].

Microstructural characterisation of ELP hydrogels

Our understanding of the nano- and microstructure of ELP hydrogels has been hampered by the intrinsic disorder of ELPs and the high mobility and viscosity of their coacervates [21]. However, such data is critical for potential applications since it determines the ability of cells to infiltrate hydrogels in tissue regeneration applications and the diffusion of small molecules through the hydrogel network in drug delivery applications. Therefore, we investigated the links between amino acid sequence and the nano- and microstructure of the hydrogel-forming ELPs in our library. We combined Fourier-transform infrared (FTIR) spectroscopy and scanning electron microscopy (SEM) with techniques that probed the hydrogels in a hydrated state. The latter included rheology (discussed in the previous section) and mesh size determination by fluorescence recovery after photobleaching (FRAP). FTIR spectra were used to assess the secondary struc-

ture of flash-frozen and freeze-dried ELPs in solution (4 °C) and hydrogel (37 °C) states (15 wt %). To do so, we inspected the amide I region of the spectra (1600-1700 cm^{-1}). This region is commonly used for quantitative determination of the secondary structure in protein materials [54]. We found a modest increase in ordered structures after gelation (from ~26-32% to ~32-40%) (Figure 3a,b), as shown by the increase in the peak associated to β-sheets at 1620 cm^{-1}. This increase was less pronounced in chemical hydrogels (Table S5), likely because covalent crosslinks hindered the reshuffling of the ELP network at 37 °C. The modest increase in order observed in physical hydrogels likely originated from IPAVG blocks, which can lead to more structural order above LCST [55]. Despite the modest shift towards more order above LCST, the hydrogels remained mostly disordered (Figure S16). The high G' of $E_{AE,triblock}$, $E_{AI,triblock}$, or chemical hydrogels -despite their lack of order- indicates that nanoscale order is not a requirement for mechanical reinforcement in ELP materials.

To assess the microstructure of the hydrogels, we first estimated their mesh size from their measured G' using rubber elasticity theory [56] (Equation (4)). The predicted mesh sizes ranged from 4.5 ± 0.7 (E_{KE}) to 15.0 ± 4.2 nm (E_{AI}) (Table 2). We also probed the mesh size by performing FRAP diffusion measurements for fluorescein isothiocyanate labelled dextran tracers of different sizes (MW of 40 and 150 kDa, radii of ca. 4.5 and 8.5 nm, respectively) within ELP hydrogels incubated at 37 °C (Figure 3c,d) for 1 h (physical hydrogels) or 2 h (chemical hydrogels). Glutaraldehyde used to prepare chemical hydrogels did not interfere with the diffusion of dextrans (Figure S17). We found that the diffusion of 40 kDa dextrans was unhindered (i.e., identical to that of a pure dextran solution in milliQ water) for E_{AI}. For the rest of our ELP library, diffusion of 40 kDa dextrans was hindered by the presence of the hydrogel network. The larger 150 kDa dextrans showed an even slower fluorescence recovery (Figure 3e), in this case also within E_{AI} networks. Fitting the fluorescence recovery data to a single exponential (Figure S18) allowed us to estimate the diffusivity of dextrans in the different ELP networks (Table 2). This value was normalised by the diffusivity D_0 in solution calculated using the Stokes-Einstein relation (73.0 and 38.6 μm/s^2 for 40 and 150 kDa dextrans, respectively) (Figure 3f). Generally, chemical hydrogels led to lower diffusivities. Given that the diffusion of 150 kDa dextrans (hydrodynamic radius of 8.5 nm) was hindered in all samples, this suggests that the mesh sizes were below 17 nm. This estimate is consistent with the mesh sizes calculated from G' via rubber elasticity theory. The diffusion of dextrans was similar in all samples, despite the variations in the amino acid sequence. A similar phenomenon was reported in networks formed by ELPs fused to helical partially ordered polypeptides [36], where MW, helical percentage, or helix sequence did not impact their void volume.

We note that the mesh size estimates for hydrated hydrogels obtained using Equation (4) or from probe diffusivity measurements with FRAP differed by three orders of magnitude from the average pore sizes measured via SEM on dried hydrogels (5.2 ± 1.1 µm for $E_{AE,triblock}$, 2.7 ± 0.7 µm for E_{KE}+GA, 1.5 ± 0.4 µm for $E_{AI,triblock}$, or 0.8 ± 0.3 µm for E_{KI}+GA, determination of the average pore size was not possible for E_{AI} due to its highly heterogeneous structure) (Figure S19). We hypothesise that the large pores observed via SEM are artefacts of the drying process needed to image the samples with this technique. Thus, our results show that techniques that probe hydrogels in hydrated, non-perturbed conditions are required to obtain reliable microstructural data for ELP hydrogels.

DISCUSSION

Analysis of the multidimensional dataset of molecular properties obtained via short MD simulations allowed us to develop a predictive model for the LCST behaviour of ELPs at 37 °C. This model unambiguously classified a library of 8 ELPs by their presence or absence of LCST behaviour. It did so without the need for computationally expensive long simulation times or advanced sampling methods. Experimental characterisation indicated that ELPs with physical crosslinking of IPAVG blocks in triblock arrangements formed stiff and reversible hydrogels at physiological temperature. Nonpolar guest amino acids (isoleucine) facilitated the formation of stiffer hydrogel networks, albeit with a higher viscous response. Chemical hydrogels attained the highest elastic response, but without thermoresponsiveness. Triblock sequences exhibited hysteresis upon cooling that was not observed for multiblock arrangements. Silk-like GAGAGS blocks hindered the formation of hydrogel networks, highlighting the need for larger ELP regions (more than the five VPGXG repeats used here) between silk-blocks to enable self-assembly and network formation. Theory and experiments also demonstrated the need to probe ELP hydrogels in a hydrated and unperturbed manner to obtain reliable information about the network structure. The hydrogels displayed mesh sizes below 17 nm at ELP concentrations of 15 wt %, with little dependence on the ELP sequence.

Overall, our results reveal multiple handles to orthogonally control the features of ELP hydrogels via block selection and sequence organisation: chain hydrophilicity/hydrophobicity and/or block distribution can be used to control the viscoelastic properties and LCST, whereas polypeptide concentration defines the network permeability. Furthermore, we showed the potential of MD simulations to mitigate the costs of developing *de novo* ELP designs by predicting their LCST behaviour before synthesising them. This opens the possibility to explore a larger sequence space at a faster pace, where it will be interesting to test the robustness of the model for different MWs or new building blocks.
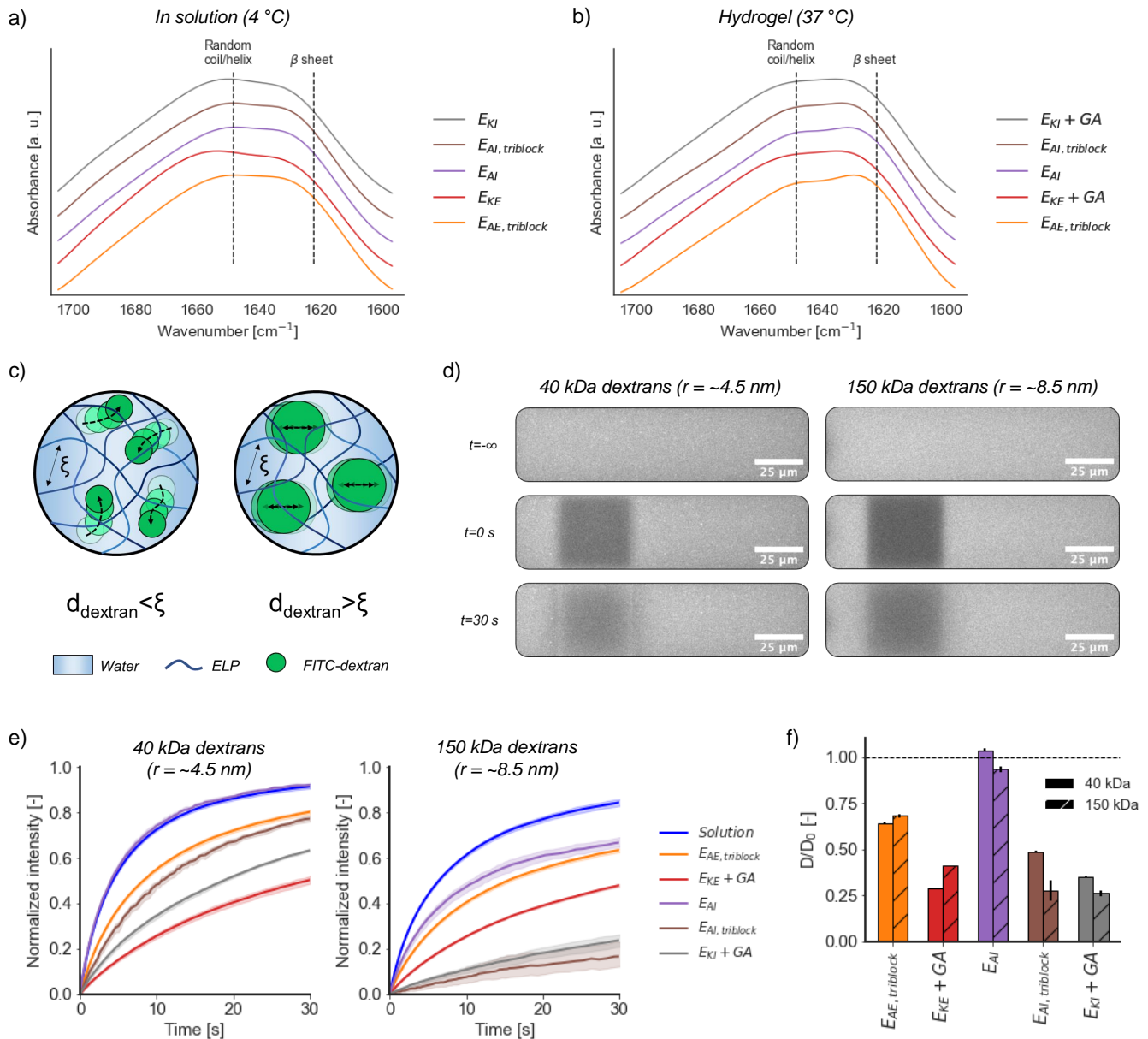
Figure 3. **Microstructural characterisation of ELP hydrogels. a.** Normalised FTIR spectra of the amide I region for freeze-dried and cryo-fractured ELP samples obtained from 15 wt % solutions after overnight incubation at 4 °C (n=2). Spectra are shifted along the y-axis for clarity. **b.** FTIR spectra of the amide I region for freeze-dried and cryo-fractured ELP hydrogels (15 wt %) after incubation at 37 °C for 1 h (physical hydrogels) or 2 h (chemical hydrogels, in the presence of glutaraldehyde) (n=2). **c.** Schematic representation of the diffusion of dextrans of different MW within ELP hydrogels with mesh size ξ. **d.** Snapshots of a representative photobleaching experiment for ELP hydrogels (15 wt %) at 37 °C using FITC-dextrans with MWs of 40 and 150 kDa. **e.** Fluorescence recovery curves for hydrogel samples at 37 °C for 40 kDa dextrans (left, hydrodynamic radius r $\sim$ 4.5 nm) and 150 kDa dextrans (right, r $\sim$ 8.5 nm) (n=5). **f.** Ratio between the diffusivity D of 40 and 150 kDa dextrans in ELP hydrogels (15 wt %, 37 °C) and the predicted diffusivity D_0 from Equation (6) in milliQ water (1 mg/mL).

METHODS

Synthesis and characterization of the ELP Library

Synthetic DNA sequences were designed to encode for the different elastin-like polypeptides (ELPs) and pur-chased from GeneArt (Table 1). The DNA fragments were then transformed into an electrocompetent *E. coli* K12 strain. The correct transformation of plasmids con-taining the ELP genes was confirmed by agarose gel electrophoresis. To that end, plasmids were isolated using the NucleoSpin® plasmid DNA purification kit (Macherey-Nagel) according to the manufacturer's pro-

Table 2. Microstructural characterisation of ELP hydrogels: mesh size ξ as calculated from the gel storage modulus G' following Equation (4); fluorescence recovery rate k_{FRAP} obtained by fitting FRAP data and diffusion constant D.

ELP design	ξ [nm]	Dextrans 40 kDa		Dextrans 150 kDa	
		k_{FRAP} [s^{-1}]	D [μm^2/s]	k_{FRAP} [s^{-1}]	D [μm^2/s]
Solution 1 mg/mL	-	0.179 ± 0.002	73.0 ± 0.8	0.138 ± 0.001	38.6 ± 0.3
$E_{AE,triblock}$	9.9 ± 0.8	0.115 ± 0.001	47.4 ± 0.4	0.094 ± 0.001	26.7 ± 0.3
E_{KE} + GA	4.5 ± 0.7	0.051 ± 0.001	21.0 ± 0.4	0.057 ± 0.000	16.2 ± 0.3
E_{AI}	15.0 ± 4.2	0.186 ± 0.002	76.7 ± 0.4	0.129 ± 0.002	36.7 ± 0.6
$E_{AI,triblock}$	6.0 ± 0.3	0.087 ± 0.001	35.9 ± 0.4	0.038 ± 0.008	10.8 ± 2.3
E_{KI} + GA	4.8 ± 0.6	0.063 ± 0.001	26.0 ± 0.4	0.036 ± 0.002	10.2 ± 0.6

tocol. Transformants were randomly selected and used for bacterial fermentation in 2-L shake flasks containing 500 mL of Terrific Broth medium. Cultivation was performed at 27 °C and ELP expression was induced with L-arabinose when the optical density of the culture at 600 nm reached 0.6. After overnight expression, cells were harvested by centrifugation at 5500 rcf for 20 min at 4 °C. The supernatant was decanted, and the cell pellets were subjected to a freeze-thaw cycle to rupture them. Thereafter, cell pellets were resuspended in PBS and tip sonicated to enhance the release of ELP. The LCST behaviour of ELPs allowed for their purification via inverse temperature cycling [46]. The aggregation of ELPs was triggered by adding 2 M NaCl and incubating the solutions at 42 °C for 1 h. For ELPs containing glutamic acid, the pH was adjusted to 4 to protonate glutamic acid residues and facilitate the coacervation [46]. Purified ELPs were desalted in 3000 MWCO Amicon® ultra-15 centrifugal filter units (MilliporeSigma). The desalted materials were then resuspended in milliQ water, flash-frozen and lyophilised, followed by storage at −20 °C until further use. The purity of ELPs was assessed via sodium dodecyl sulphate−polyacrylamide gel electrophoresis (SDS-PAGE) using NuPAGE 4-12% Bis-tris gels (Invitrogen). Mark12 unstained standard (Thermo Fisher) was used as protein ladder. Gels were stained using SYPRO® Red gel staining agent (Invitrogen) following the manufacturer's protocol. Total amino acid analysis was performed using the Accq Tag method after chemical hydrolysis (Waters). The theoretical hydrophobicity of the ELPs was calculated using the Kyte-Doolittle scale [45]. Their molecular weight was assessed via intact LC-MS.

Computational modelling

Input structures for molecular dynamics (MD) simulation were created based on the ELP sequences from Table 1. To reduce the computational cost of these simulations, we used 5-mer sequences instead of the 12-mer used experimentally (Table S1). Extended conformations of each ELP were built using the software Avogadro (version 1.2.0) [57]. MD simulations were performed using the software NAMD 2.13 developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign [58]. The Chemistry at HARvard Macromolecular Mechanics (CHARMM) force field was used for these simulations [59]. This force field is widely used for studying proteins, including ELPs [20, 35, 39]. The atomic structures were visualised using the Visual Molecular Dynamics (VMD) graphics software [60].

Implicit solvent simulations: Each extended ELP structure was subjected to MD simulations in implicit solvent to obtain a folded structure. The equilibration and folding of ELPs was assessed by the evolution of the root mean squared displacement (RMSD) of their atomic positions. First, the structure was subjected to energy minimization for 20000 timesteps to relax the polypeptide, using the steepest descent algorithm. This was followed by Langevin dynamics in Generalized Born implicit solvent [61]. A simulation step of 2 fs was applied at 310 K for a total simulation time of 50 ns. The short-range electrostatic interactions and Lennard-Jones interactions were evaluated with a cutoff of 18 Å and a switch distance of 16 Å [62].

Explicit solvent simulations: Simulations in explicit water were performed at 310 K. The final structures obtained in implicit solvent simulations were solvated in a TIP3P water box with 3D periodic boundary conditions. The distance between any ELP atom and the edge of the periodic box was at least 12 Å to avoid spurious effects of self-interactions for the ELP chain. Na$^+$ and Cl$^-$ ions were added to neutralise the charge of the system in simulations containing ionic ELPs. The ShakeH algorithm was applied to all the bonds containing hydrogen atoms. The energy of the systems was minimised for 20000 timesteps, followed by Langevin dynamics for 150 ns using the NPT ensemble at 310 K and pressure (1 bar) was exerted through the Nosé-Hoover Langevin piston. The long-range electrostatic Coulombic interactions were calculated using particle mesh Ewald method with a grid spacing of 1 Å. A cutoff distance of 12 Å was applied for electrostatic and van der Waals interactions, with a switch distance of 10 Å to avoid hard cuts [62]. A timestep of 2 fs was applied. The backbone of the ELPs was restrained for the first 1 ns of simulation. Three replicas were performed for each system, and simulation data was sampled every 0.2 ns.

Data analysis: Molecular properties were sampled from the last 50 ns of the MD trajectories. The evolution of the RMSD of the atomic positions during the simulation was analysed for the protein backbone using the RMSD Trajectory Tool from VMD. The count of intraprotein and protein-solvent hydrogen bonds throughout the simulation trajectories was analysed using the Hbonds plugin from VMD, using a distance cutoff of 3.5 Å and a D-H-A angle cutoff of 30° as geometric criteria. Custom TCL scripts were developed to analyse in VMD the solvent accessible surface area (SASA) (using a standard water probe radius of 1.4 Å), the radius of gyration of ELPs, and the evolution of the secondary structure (using the STRIDE algorithm) [63]. Water molecules in the hydration shell of the ELP were defined as those within 3.15 Å from the ELP backbone. The interaction energies were calculated using the molecular mechanics energy function in NAMD 2.13.

For exploratory data analysis purposes, principal component analysis (PCA) was used. PCA searches for the variable subspace summarising the main features of the data [64]. PCA transforms the original (correlated) variables (in this case, the molecular properties sampled from MD simulations) into a lower number of uncorrelated variables or principal components (PCs). The PCA model equation is

$$X = TP' + E \qquad (1)$$

where X is the data matrix (having observations by rows and variables by columns), T is the score matrix containing the PCs, P is the loading matrix containing the linear combination of the original variables to build the PCs (being P' its transposed), and E is the residual matrix.

A predictive model for MD data was developed via a supervised model based on partial least squares regression discriminant analysis (PLS-DA). PLS is a multivariate statistics method commonly used to predict an output variable y from a set of X predictors [65]. PLS, similarly to PCA, reduces the dimensionality of the X variable space by finding the linear combinations of predictors that best summarise the X variable space and best predict the output variable. Therefore, the components found by the PLS algorithm maximise the covariance between X and y. The scores (or PLS components) are obtained as

$$T = XW \qquad (2)$$

where W is the normalised weights matrix. On one hand, these scores reconstruct well the original predictor matrix $X = TP' + E$. On the other hand, y can be predicted using the score matrix [47] as

$$y = Tq + F = XWq + F = Xb + F \qquad (3)$$

where F is the residual matrix and b are the PLS coefficients. In our case, the response variable is not quantitative but qualitative (presence or absence of LCST behaviour), and therefore the discriminant version of PLS (PLS-DA) was used [48]. To perform PLS-DA, the presence or absence of LCST behaviour was coded as 1 or 0, respectively. When using the regression model to predict, a quantitative value is produced, which is later converted to a qualitative 0/1 taking as threshold the mean between those two values as reference [48].

Thorough validation of PLS-DA models is needed when dealing with large datasets. Single cross validation approaches, where the model is built with a subset of samples and validated with an external set, often leads to too optimistic results [66, 67]. That is why a more stringent double cross validation procedure was used here, which comprised the following steps:

1. The 8 MD simulations (one per ELP, including 3 replicas for each ELP) were split in 3 groups: training (including simulations 1 to 7) and test (including simulation 8).

2. The simulations in the training set were further split in 2 groups: model (including simulations 1 to 6) and validation (including simulation 7).

3. The 6 model simulations were used to build a partial least squares regression discriminant analysis model (PLS-DA).

4. The validation simulation was then projected into the PLS-DA model. The number of components of the PLS-DA model was selected at this step, including the minimum number of components needed to obtain perfect classification (LCST behaviour yes/no for each ELP) of the validation set.

5. The PLS-DA model built in step 3 with as many components as defined in step 4 was used to predict the result of the test set, left out in step 1.

Steps 2 to 5 were afterwards repeated 6 times, switching the validation simulation from 1 to 6, and using the remaining ones to build the model. In the same way, steps 1 to 5 were then repeated switching the test simulation from 1 to 7, using the remaining ones to build the model and validate it.

Hydrogel Formation and Characterisation

Experiments were performed using ELPs solutions in milliQ water prepared at 4 °C and with a polypeptide concentration of 15 wt %. This concentration has been reported to yield stiff ELP hydrogels [44]. In all cases (except for chemical hydrogels) solutions were incubated at 4 °C overnight. For the lysine-containing ELPs, chemical hydrogels were formed by dissolving the polypeptides in milliQ water containing an excess of glutaraldehyde

as crosslinking agent (40:1 glutaraldehyde:ELP molar ratio). The dissolution of lysine-containing ELPs was performed on ice for 10 min to facilitate ELP dissolution in milliQ water. Higher temperatures were prevented during this step, as they accelerated the reaction between ELP and glutaraldehyde and rapidly formed a stiff chemical hydrogel that could not be pipetted onto the rheometer.

Turbidimetry: The ability of ELPs to undergo LCST in the 4-37 °C range was analysed via optical density measurements using a microplate reader (Multiskan GO, Thermo Scientific). 100 µL of each ELP solution (15 wt %) were loaded in a 96-well plate at 4 °C. The plate was then introduced in the microplate reader preheated at 37 °C, and the absorbance of the solutions was recorded at 350 nm.

Rheology: Rheological characterisation was performed to assess the ability of ELP solutions to form hydrogels, and to characterise their viscoelastic properties. Experiments were performed using small amplitude oscillatory shear rheology on a stress-controlled rheometer (Anton Paar MCR 501) equipped with a cone-plate geometry with a diameter of 20 mm and a cone angle of 1°. The temperature was controlled by a Peltier system. For each test, 42 µL of the ELP solutions (15 wt %) were loaded at 4 °C onto the bottom plate using a pipette, followed by equilibration for 5 min. Solutions without glutaraldehyde were incubated overnight at 4 °C before loading them onto the rheometer. For experiments with glutaraldehyde, ELPs were dissolved in a milliQ water-glutaraldehyde solution for 10 min on ice, and quickly pipetted onto the Peltier plate. Low viscosity mineral oil (Sigma Aldrich) was applied to the air-sample interface around the measuring geometry to prevent water evaporation. The LCST of ELP solutions was analysed with a temperature sweep between 4 and 37 °C (with a heating rate of 1 °C/min). The reversibility of gelation was assessed with three temperature cycles between 4 and 37 °C (with a heating/cooling rate of 1 °C/min). Holding times of 30 min were applied at 4 and 37 °C, after each temperature ramp. Strain sweeps were performed from 0.01% to 15% at a frequency of f=1 Hz to evaluate the linear viscoelastic region of these hydrogels. Frequency sweeps were carried out at 37 °C from 0.01 to 15 Hz, using a constant strain amplitude γ=0.3%.

Fourier Transform Infrared Spectroscopy (FTIR): Freeze-dried samples of ELP solutions and hydrogels were analysed by FTIR to assess their secondary structure before and after gelation. 80 µL of cold ELP solutions were placed in microcentrifuge tubes. Samples before gelation were directly submerged in liquid N_2 for 1 min. Hydrogel samples were prepared by incubating ELP solutions at 37 °C for 1 h (physical hydrogels) or 2 h (chemical hydrogels), before flash-freezing them in liquid N_2 for 1 min. All samples were then lyophilised and cryo-fractured prior to FTIR analysis. Infrared spectra were measured in a Bruker Vertex 70 Attenuated Total Re-

flectance FTIR device equipped with a Harrick split pea accessory. For each measurement, 64 scans with a resolution of 2 cm^{-1} were recorded in the range of 650 to 4000 cm^{-1}. The secondary structure of the polypeptides is related to the C=O stretching vibration and can be determined by performing peak deconvolution over the amide I region (1595−1705 cm^{-1}). This was done using the lmfit package for curve fitting from Python. The deconvolution was carried out using five primary peaks assigned to different secondary structures: 1620 cm^{-1} (β-sheet), 1645, 1660 and 1670 cm^{-1} (random coil/helix), and 1700 cm^{-1} (β-turn) [68]. The peak positions were allowed to shift 4 cm^{-1} to obtain a reconstituted curve as close as possible to the original spectra. The amide I region from all spectra was normalised to its highest value, to facilitate the comparison between different samples. We used the Levenberg-Marquardt least-squares method, and a Gaussian model was selected for the band shape.

Mesh size determination: We estimated the typical mesh size ξ (distance between physical or chemical crosslinking points) of ELP hydrogels from the measured G', based on rubber elasticity theory [56]:

$$G' = k_B T \xi \tag{4}$$

Equation (4) can be applied to gels and physically crosslinked networks of flexible chains [56], with $k_B T$ representing the thermal energy at the temperature T used for G' determination.

Fluorescent recovery after photobleaching (FRAP): Diffusion coefficients of fluorescent dextran probes in ELP hydrogels (15 wt %) were determined via FRAP experiments. ELPs were dissolved in fluorescein isothiocyanate (FITC)-labelled dextran solutions (40 and 150 kDa, 1 mg/mL, Sigma) at 4 °C. FRAP experiments were performed on a Leica Stellaris 8 Falcon confocal microscope at 37 °C equipped with a white light laser. The FRAP protocol started with 10 images at low laser power and a pixel dwell time of 3.6125 μs to determine the baseline fluorescence. Subsequently, a square region of 37.5x37.5 μm was photobleached using an excitation wavelength of 491 nm at 100% intensity for 7 seconds. A time series was then recorded to follow fluorescence recovery, acquiring a frame (pixel dwell time of 3.6125 μs, image size 512 x 128) every 0.33 s for the first 20 s, and every 1 s thereafter. FRAP data was subjected to full-scale normalisation [69], using the fluorescence of the whole imaging area to normalise and correct for laser fluctuations, photobleaching during image acquisition, and fluorescence loss during photobleaching. The fluorescence recovery was analysed by fitting the normalised data to a single exponential curve (Equation (5)) using the curve_fit function of the SciPy module from Python:

$$I(t) = A(1 - e^{-k_{FRAP}t}) \tag{5}$$

where A corresponds to the plateau intensity after fluorescence recovery, k_{FRAP} is the recovery rate, and t is

the time after photobleaching. We calculated the free diffusivity D_0 of pure dextran solutions using the Stokes-Einstein relationship (Equation (6)):

$$D_0 = \frac{k_B T}{6\pi\eta r} \qquad (6)$$

where $k_B T$ is the thermal energy at the temperature T used in FRAP experiments (37 °C), η is the viscosity of the solvent (0.692 mPa s for milliQ water at 37 °C), and r is the dextran particle radius (4.5 nm for 40 kDa dextrans, 8.5 nm for 150 kDa dextrans). To calculate the diffusivity $D_{hydrogel}$ of dextrans within ELP hydrogels, we assumed that $D \sim k_{FRAP} \cdot L^2$, where L is the characteristic length scale of our FRAP experiments (the size of the bleached area, 37.5 μm). Given that the size of the bleached spots was constant in our FRAP experiments, it follows that

$$D_{hydrogel} = \frac{k_{FRAP}}{k_{FRAP,0}} D_0 \qquad (7)$$

Scanning Electron Microscopy (SEM): Freeze-dried samples of ELP hydrogels were analysed by SEM to investigate their microstructure. 80 µL of each solution (15 wt %) were incubated in microcentrifuge tubes at 37 °C for 1 h (physical hydrogels) or 2 h (chemical hydrogels). The tubes were then submerged in liquid N_2 for 1 min, followed by lyophilisation. The resulting samples were then cryo-fractured and coated with a 2 nm layer of iridium using a Q150T S/E/ES sputter coater (Quorum Technologies). SEM images were obtained using a FEI Teneo LoVa microscope with trinity detector. Analysis of the pore sizes of SEM samples was performed using the software ImageJ [70] by measuring the diameter of 30 randomly selected pores.

Author contributions. DLB designed, purified, and characterised the ELP library, and performed rheology experiments and MD simulations. AFF developed the (un)supervised predictive model from computational data. IM assisted with rheology experiments and performed FRAP analysis. GHK, CMJS, and JCT conceptualised the research and provided guidance and supervision. All authors participated in the design of experiments, data interpretation, and manuscript preparation.

Competing interests. The authors declare no competing interest.

[1] M. J. Harrington and P. Fratzl, Progress in Materials Science , 100767 (2020).

[2] W. Huang, D. Restrepo, J.-Y. Jung, F. Y. Su, Z. Liu, R. O. Ritchie, J. McKittrick, P. Zavattieri, and D. Kisailus, Advanced Materials **31**, 1901561 (2019).

[3] Y. J. Yang, A. L. Holmberg, and B. D. Olsen, Annual Review of Chemical and Biomolecular Engineering **8**, 549 (2017).

[4] A. Pena-Francesch and M. C. Demirel, Frontiers in Chemistry **7**, 1 (2019).

[5] A. Lee, A. R. Hudson, D. J. Shiwarski, J. W. Tashman, T. J. Hinton, S. Yerneni, J. M. Bliley, P. G. Campbell, and A. W. Feinberg, Science **365**, 482 (2019).

[6] X. Jia, C. Wang, V. Ranganathan, B. Napier, C. Yu, Y. Chao, M. Forsyth, F. G. Omenetto, D. R. MacFarlane, and G. G. Wallace, ACS Energy Letters **2**, 831 (2017).

[7] H. Tao, B. Marelli, M. Yang, B. An, M. S. Onses, J. A. Rogers, D. L. Kaplan, and F. G. Omenetto, Advanced Materials **27**, 4273 (2015).

[8] C. Gilbert and T. Ellis, ACS Synthetic Biology **8**, 1 (2019).

[9] J. C. van Hest and D. A. Tirrell, Chemical Communications **19**, 1897 (2001).

[10] S. M. Mithieux and A. S. Weiss, in *Advances in Protein Chemistry*, Vol. 70 (Academic Press, 2005) pp. 437–461.

[11] H. Vindin, S. M. Mithieux, and A. S. Weiss, Matrix Biology **84**, 4 (2019).

[12] J. Ozsvar, C. Yang, S. A. Cain, C. Baldock, A. Tarakanova, and A. S. Weiss, Frontiers in Bioengineering and Biotechnology **9**, 138 (2021).

[13] S. Roberts, M. Dzuricky, and A. Chilkoti, FEBS Letters **589**, 2477 (2015).

[14] A. Ibáñez-Fonseca, T. Flora, S. Acosta, and J. C. Rodríguez-Cabello, Matrix Biology **84**, 111 (2019).

[15] J. C. Rodríguez-Cabello, F. J. Arias, M. A. Rodrigo, and A. Girotti, Advanced Drug Delivery Reviews **97**, 85 (2016).

[16] Y. Wu, B. O. Okesola, J. Xu, I. Korotkin, A. Berardo, I. Corridori, F. L. P. di Brocchetti, J. Kanczler, J. Feng, W. Li, Y. Shi, V. Farafonov, Y. Wang, R. F. Thompson, M. M. Titirici, D. Nerukh, S. Karabasov, R. O. Oreffo, J. Carlos Rodriguez-Cabello, G. Vozzi, H. S. Azevedo, N. M. Pugno, W. Wang, and A. Mata, Nature Communications **11**, 1 (2020).

[17] E. Wang, M. S. Desai, and S. W. Lee, Nano Letters **13**, 2826 (2013).

[18] D. Urry, M. Long, B. Cox, T. Ohnishi, L. Mitchell, and M. Jacobs, Cardiovascular Research **371**, 597 (1974).

[19] D. W. Urry, D. C. Gowda, T. M. Parker, C. Luan, M. C. Reid, C. M. Harris, A. Pattanaik, and R. D. Harris, Biopolymers **32**, 1243 (1992).

[20] S. Rauscher and R. Pomès, eLife **6**, 1 (2017).

[21] S. E. Reichheld, L. D. Muiznieks, F. W. Keeley, and S. Sharpe, Proceedings of the National Academy of Sciences **114**, E4408 (2017).

[22] D. H. Le and A. Sugawara-Narutaki, Molecular Systems Design and Engineering **4**, 545 (2019).

[23] K. Nagapudi, W. T. Brinkman, B. S. Thomas, J. O. Park, M. Srinivasarao, E. Wright, V. P. Conticello, and E. L. Chaikof, Biomaterials **26**, 4695 (2005).

[24] R. E. Sallach, W. Cui, J. Wen, A. Martinez, V. P. Conticello, and E. L. Chaikof, Biomaterials **30**, 409 (2009).

[25] C. González-Obeso, M. González-Pérez, J. F. Mano, M. Alonso, and J. C. Rodríguez-Cabello, Small **16**, 1 (2020).

[26] B. Li, D. O. Alonso, B. J. Bennion, and V. Daggett, Journal of the American Chemical Society **123**, 11991 (2001).

[27] A. Krukau and I. Brovchenko, Biomacromolecules **8**, 2196 (2007).

[28] N. K. Li, F. G. Quiroz, C. K. Hall, A. Chilkoti, and Y. G. Yingling, Biomacromolecules **15**, 3522 (2014).

[29] R. Machado, A. Da Costa, V. Sencadas, A. M. Pereira, T. Collins, J. C. Rodríguez-Cabello, S. Lanceros-Méndez, and M. Casal, Macromolecular Bioscience **15**, 1698 (2015).

[30] Y. Wu, J. A. Mackay, J. R. Mcdaniel, A. Chilkoti, and R. L. Clark, Biomacromolecules **10**, 19 (2009).

[31] A. Fernández-Colino, F. J. Arias, M. Alonso, and J. Carlos Rodríguez-Cabello, Biomacromolecules **15**, 3781 (2014).

[32] J. R. McDaniel, D. C. Radford, and A. Chilkoti, Biomacromolecules **14**, 2866 (2013).

[33] D. E. Meyer and A. Chilkoti, Biomacromolecules **5**, 846 (2004).

[34] F. G. Quiroz and A. Chilkoti, Nature Materials **14**, 1164 (2015).

[35] A. Tarakanova, W. Huang, A. S. Weiss, D. L. Kaplan, and M. J. Buehler, Biomaterials **127**, 49 (2017).

[36] S. Roberts, T. S. Harmon, J. L. Schaal, V. Miao, K. J. Li, A. Hunt, Y. Wen, T. G. Oas, J. H. Collier, R. V. Pappu, and A. Chilkoti, Nature Materials **17**, 1154 (2018).

[37] A. Ibáñez-Fonseca, D. Orbanic, F. J. Arias, M. Alonso, D. I. Zeugolis, and J. C. Rodríguez-Cabello, Small **16**, 2001244 (2020).

[38] C. Zhai, T. Li, H. Shi, and J. Yeo, J. Mater. Chem. B **8**, 6562 (2020).

[39] J. D. Tang, C. E. McAnany, C. Mura, and K. J. Lampe, Biomacromolecules **17**, 3222 (2016).

[40] N. K. Li, S. Roberts, F. G. Quiroz, A. Chilkoti, and Y. G. Yingling, Biomacromolecules **19**, 2496 (2018).

[41] B. Zhao, N. K. Li, Y. G. Yingling, and C. K. Hall, Biomacromolecules **17**, 111 (2016).

[42] S. Yousefinejad and B. Hemmateenejad, Chemometrics and Intelligent Laboratory Systems **149**, 177 (2015).

[43] P. Carracedo-Reboredo, J. Liñares-Blanco, N. Rodríguez-Fernández, F. Cedrón, F. J. Novoa, A. Carballal, V. Maojo, A. Pazos, and C. Fernandez-Lozano, Computational and Structural Biotechnology Journal **19**, 4538 (2021).

[44] M. J. Glassman and B. D. Olsen, Biomacromolecules **16**, 3762 (2015).

[45] J. Kyte and R. F. Doolittle, Journal of Molecular Biology **157**, 105 (1982).

[46] J. C. Rodríguez-Cabello, A. Girotti, A. Ribeiro, and F. J. Arias, Methods in Molecular Biology **811**, 17 (2012).

[47] S. Wold, M. Sjöström, and L. Eriksson, Chemometrics and Intelligent Laboratory Systems **58**, 109 (2001).

[48] M. Barker and W. Rayens, Journal of Chemometrics **17**, 166 (2003).

[49] M. K. Gupta, D. T. Wagner, and M. C. Jewett, MRS Bulletin **45**, 999 (2020).

[50] A. Ribeiro, F. J. Arias, J. Reguera, M. Alonso, and J. C. Rodríguez-Cabello, Biophysical Journal **97**, 312 (2009).

[51] S. Acosta, L. Poocza, L. Quintanilla-Sierra, and J. C. Rodríguez-Cabello, Biomacromolecules **22**, 158 (2021).

[52] J. Irianto, C. R. Pfeifer, Y. Xia, and D. E. Discher, Cell **165**, 1820e1 (2016).

[53] S. Correa, A. K. Grosskopf, H. Lopez Hernandez, D. Chan, A. C. Yu, L. M. Stapleton, and E. A. Appel, Chemical Reviews **121**, 11385 (2021).

[54] X. Hu, X. Wang, J. Rnjak, A. S. Weiss, and D. L. Kaplan, Biomaterials **31**, 8121 (2010).

[55] M. J. Glassman and B. D. Olsen, Macromolecules **48**, 1832 (2015).

[56] P. de Gennes, Macromolecules **9**, 587 (1976).

[57] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, Journal of Cheminformatics **4**, 17 (2012).

[58] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, Journal of computational chemistry **26**, 1781 (2005).

[59] J. Huang and A. D. Mackerell, Journal of Computational Chemistry **34**, 2135 (2013).

[60] W. Humphrey, A. Dalke, and K. Schulten, Journal of Molecular Graphics **14**, 33 (1996).

[61] A. V. Onufriev and D. A. Case, Annual Review of Biophysics **48**, 275 (2019).

[62] D. E. Tanner, K. Y. Chan, J. C. Phillips, and K. Schulten, Journal of Chemical Theory and Computation **7**, 3635 (2011).

[63] D. Frishman and P. Argos, Proteins: Structure, Function, and Bioinformatics **23**, 566 (1995).

[64] J. Jackson, *A user's guide to principal components* (Wiley Series in Probability and Statistics, 2004).

[65] P. Geladi and B. Kowalski, Analytica Chimica Acta **185**, 1 (1986).

[66] E. Szymańska, E. Saccenti, A. K. Smilde, and J. A. Westerhuis, Metabolomics **8**, 3 (2012).

[67] J. A. Westerhuis, H. C. Hoefsloot, S. Smit, D. J. Vis, A. K. Smilde, E. J. Velzen, J. P. Duijnhoven, and F. A. Dorsten, Metabolomics **4**, 81 (2008).

[68] S. Ling, N. Dinjaski, D. Ebrahimi, J. Y. Wong, D. L. Kaplan, and M. J. Buehler, ACS Biomaterials Science & Engineering **2**, 1298 (2016).

[69] M. A. Rapsomaniki, P. Kotsantis, I. E. Symeonidou, N. N. Giakoumakis, S. Taraviras, and Z. Lygerou, Bioinformatics **28**, 1800 (2012).

[70] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, Nature Methods **9**, 671 (2012).