Summer 7-15-2020

Fusion of Elastin Like Polypeptides to RNA Binding Motifs for Programmable Binding of Nucleic Acids

Dominic Medina
University of New Mexico

Dominic Medina

_____
Candidate

Biomedical Engineering

_____
Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Dr. Nick Carroll, Chairperson

_____

Dr. Gabriel Lopez

_____

Dr. Matthew Lakin

_____

_____

_____

_____

_____

_____

Fusion of Elastin Like Polypeptides to RNA Binding Motifs for Programmable Binding of Nucleic Acids

by

Dominic Medina

Biochemistry, Bachelor's of Science 2017

THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

MASTER'S OF SCIENCE BIOMEDICAL ENGINEERING

The University of New Mexico Albuquerque, New Mexico

JULY, 2020

Fusion of Elastin Like Polypeptides to RNA Binding Motifs for Programmable Binding of Nucleic Acids

By

Dominic Medina

B.S., Biochemistry, University of New Mexico, 2017
M.S., Biomedical Engineering, University of New Mexico, 2020

Abstract

Intrinsically disordered proteins (IDPs) with RNA binding capability have recently been found to play important roles in several cellular processes, including the formation of phase separated membraneless organelles to regulate gene expression. An important class of membraneless organelles is stress granules, which enable cells to survive periods of environmental stress by regulating gene expression in the cytoplasm. Cells use a variety of ribonucleoprotein granules, primarily comprised of RNA and chaperone protein components, and can regulate cellular metabolism by sequestering and processing regulatory components, such as messenger RNAs, within these phases separated 'organelles'. To recapitulate this cellular behavior, the genetically engineered model IDPs was used known as elastin-like polypeptides (ELPs) fused to RNA-binding peptide blocks derived from cellular RNA binding motifs. These modular IDPs are engineered to undergo a reversible phase transition. Below their transition temperature (Tt), ELP fusions are soluble in water, but at temperatures above their Tt, a phase transition occurs which leads to aggregation of RNA-ELP complexes. This programmable behavior demonstrates the utility of ELPs for the triggered formation of synthetic ribonucleoprotein membraneless organelles. The goal of this work is to provide a research basis for the overall engineering goal, which is to engineer a cleavable linker domain within the NA-binding blocks of the ELP fusions, where protease cleavage of previously concatenated blocks will trigger release of the RNA from the granules and the dissolution of the granule itself.

Table of Contents

List of Figures

List of Tables

1. Background

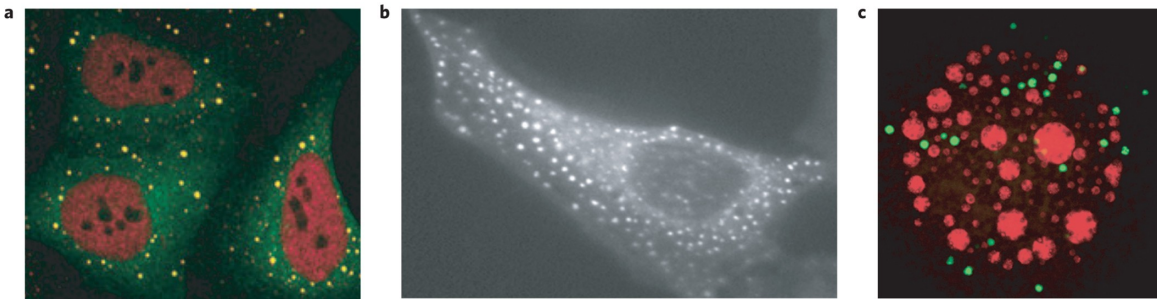

Figure 1. Examples of membrane-less bodies in cells. a. P bodies (yellow) in tissue culture cells. b. Purinosomes. c. Nucleoli (red) and histone locus bodies (green) in the nucleus of a large X. laevis oocyte. This image is reprinted from Brangwynne et al. [43].

1.1 Intrinsically Disordered Proteins

Intrinsically disordered proteins (IDPs) play a key role in cellular signaling and gene regulation machinery, with bioactivity directed via formation of ribonucleic protein assemblies through targeted nucleic acid (NA) binding and liquid phase separation behavior. A variety of transient aqueous protein phases comprised of diverse, unstructured polypeptides enriched with low complexity domains[1–3] have been reported, including P bodies and granules, purinosomes, nuclear bodies, and Cajal bodies[4–8] (Figure 1). Cellular IDPs are characterized by low sequence complexity, by lack of bulky hydrophobic amino acids and are enriched with proline and glycine residues, flanked by charged, hydrophilic amino acids[9]. IDP chains do not fold into stable, well defined, three-dimensional structures due to the structure breaking property of proline residues. These polypeptide structures are disordered and fluctuate via tethered random walk conformations, and undergo liquid-liquid phase separation when soluble, extended coils collapse to form thermal globules that combine to form phase separated granules[9,10].

Intrinsically disordered proteins are conserved across biology and it appears that IDP occurrence scales with organism complexity. For example, an average of 2.0% of archaean and 4.2% of eubacteria proteins are predicted to have long regions of disorder, while 33.0% of eukaryotic proteins are predicted to contain long regions of disorder[11–13]. Eukaryotic cells attain diverse IDP functionality with a relatively small number of translated proteins. This is due, in part, to the fact that the functions of eukaryotic cell IDPs are controlled through a large network of post-translational modifications (PTMs)[14], and studies have shown that PTMs occur most often in disordered regions[15,16]. One key role of cellular IDPs is exemplified through transcriptional regulation, where IDPs provide distinct advantages to transcription factor (TF) function enabled by conformational plasticity, robust nucleic acid binding, and regulation by environmentally triggered PTMs. Long Disordered regions in TFs allow regulatory proteins to form flexible conformations[15,17] when binding with various peptide and NA interaction partners[17], thus enhancing cellular transcriptional and translational activity.

Another key role of NA-binding IDPs are their environmentally triggered assembly as a mechanism for maintaining cellular fitness. Organisms are constantly subjected to environmental changes, which include diverse stress conditions such as fluctuations in heat, oxygen levels, water, and other nutrients and any changes in these conditions can lead to protein unfolding, denaturation, aggregation and degradation that may affect cellular fitness[16,18]. It has been well established that intrinsically disordered proteins are activated to organize biological matter in the cell as a response to environmental stress. Archetypal

examples include cytoplasmic ribonucleoprotein structures known as stress granules and processing bodies (P bodies), membraneless organelles that can sequester and protect both RNAs and proteins from stress-induced damage[19,20] (Figure 2). The sequestration of signaling and regulatory proteins can regulate signaling pathways during stress[21]. Stress granules comprise a complex mixture of protein and nucleic acid components, and they are believed to regulate cellular metabolism by physically sequestering those components within the stress granule structure, including messenger RNAs. Thus, understanding the structure and function of granular IDPs is crucial to understanding a range of behaviors in biological systems, regulated by phase separated membraneless organelles. Importantly, this understanding of intracellular material science will provide opportunities for engineering biomolecular systems in synthetic cells, inspired by the biological functions of membraneless organelles.

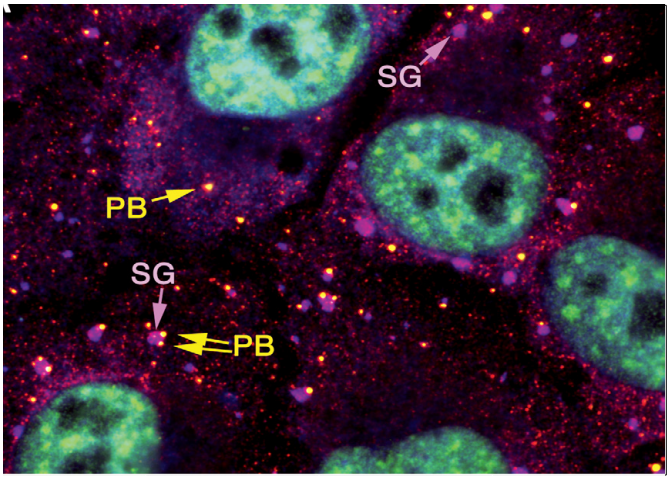

Figure 2. Stress granules (SG, purple arrows) are As stress granules also contain RNA helicase (stained in red), the merged colors appear purple. Similarly, P-bodies (PB, yellow arrows) are visualized by green staining but appear yellow due to RNA co-localization. A single cell can contain isolated P-bodies and stress granules, as well as interacting pairs of stress granules and P-bodies. This figure is reprinted from Anderson & Kedersha, 2009[19].

1.2 Elastin Like Polypeptides

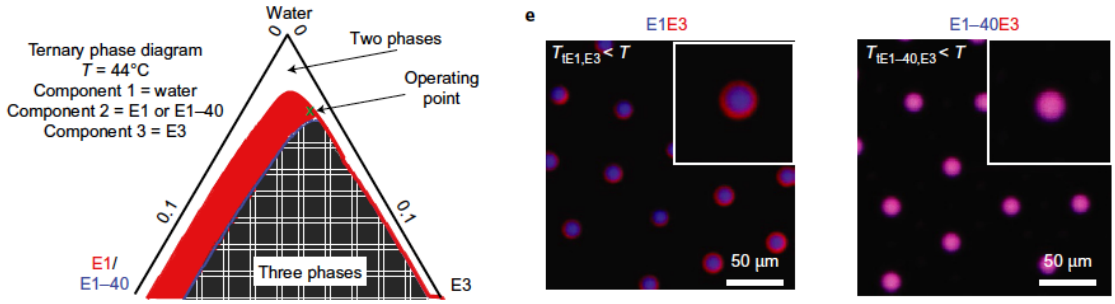

Figure 3. ELP phase diagrams can be tuned to assembled immiscible or fully mixed coacervates by modulation of guest residue X (VPGXG), chain length, and protein volume fraction. This figure is reprinted from Simon et al.[22].

Genetically engineered model IDPs include a class of minimal repeat proteins known as elastin like polypeptides (ELPs), and are composed of Val–Pro–Gly–Xaa– Gly (VPGXG) pentapeptide repeats, where 'Xaa' is a guest residue that can be substituted by any amino acid except proline. The guest residue within the pentameric repeats allow for the modulation of the polypeptide hydropathy and prediction of the associated aqueous phase behavior[23,] thus providing a simple platform for programming assemblies of ELP liquid phases (Figure 3). Below their transition temperature (Tt) ELPs are water-soluble, biocompatible, non-toxic biopolymers that, when heated above the Tt, undergo a reversible temperature phase transition (Figure 4), which is dependent on the chemical properties of the guest residue, ionic state of the ELP, salt concentration, and molecular weight of the ELP[24]. These polypeptides display lower critical solution temperature (LCST) behavior in aqueous solution, meaning they phase separate upon heating above their cloud point or Tt. Upon undergoing a thermal phase transition, they phase separate to form dense, protein-rich liquid aggregates referred to as coacervates[23,25]. The phase separation is sharp (~2°C range) and reversible, as the coacervate formed by heating above the cloud point

temperature dissolves upon cooling the solution below the cloud point[26,27].

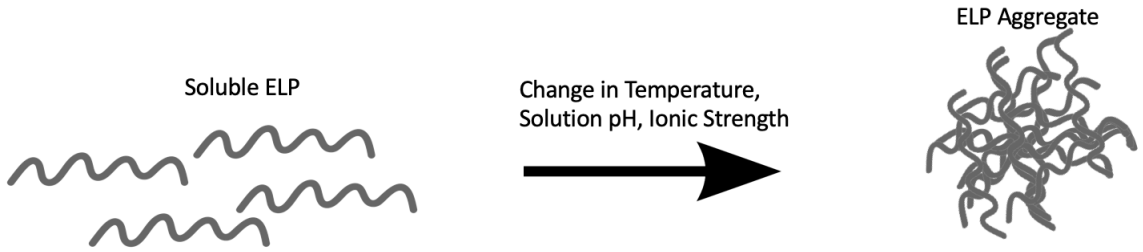

Figure 4. Illustration of stimulus responsive mechanisms-including stimulus triggered aggregation. ELP sequences are shown as $(VPGXG)_n$ motifs, where X is the guest residues mole ratios are described for each ELP with subscript ''n'' indicating the number of pentapeptide repeats in each ELP. This figure was adapted from Hassouneh et al., figure 1 [27].

ELPs were selected as a model IDP material system because they are biocompatible, and the composition, including chain length, and coupled thermal responsive properties can be fine-tuned at the genetic level. Furthermore, ELPs can be expressed at rather high levels in Escherichia coli. This is because the thermally responsive phase behavior of ELPs allows simple and rapid purification of both the polypeptide and its peptide or protein fusions after it's expression in E. coli[23,28,29]. This high level of protein purification efficacy is due to separation of ELPs from cell lysate, exploiting the inverse temperature transition behavior of these polypeptides that causes them to undergo the reversible phase transition from soluble to insoluble upon increasing the temperature or upon lowering the transition temperature from above to below the transition temperature[23]. Specifically, the expressed recombinant VPGXG polymers are soluble in the cellular lysate at low temperatures at which point the normally insoluble cellular lysate components can be removed by centrifugation. Then, upon raising the temperature to the transition temperature, the protein-based polymer comes out of solution to form a visible aggregate which can then be removed from the remaining soluble solution by centrifugation. The repetition of this process is called Inverse Transition Cycling (ITC)[30].

1.3 Elastin Like Polypeptide Fusion Proteins

Another motivating factor for selection of ELPs as our model biopolymer is their amenability to fabrication of fusion proteins with high yield. This modular nature provides an opportunity to create ribonucleoprotein granules by including NA-binding peptide blocks inspired by those found in nature. This level of sophistication in programmed hydropathy and modular material design may be necessary to create synthetic granules capable of regulation and signaling, because commonly used simple dipeptide cationic repeats (e.g. poly-GR, poly-PR), while strong RNA binders with robust phase behavior, actually inhibit protein expression because they also nonspecifically bind to translational complexes and ribosomal assemblies[31]. By contrast, formation of cellular 'membraneless' organelles[32] is governed by more complex multivalent interactions and for this thesis work, key bioinspired concepts of programmable self-assembly, regulation of gene expression, and RNA selectivity in recruitment[33–35] were implemented. There is now widespread scientific recognition that protein granules are highly dynamic organizers of mRNA translation and associated metabolic pathways, and this thesis work aims to provide an experimental framework for the genetic engineering of NA-binding ELP fusions.

Importantly, ELPs offer an immense range of possible designs at the genetic level; they can be genetically fused to peptide and proteins while maintaining the biological activity of the protein and the physical properties of the ELP[36]. Therefore, the fusion protein can be expressed in large quantities and purified rapidly through ITC. This provides a relatively inexpensive and efficient method of purification relative to current column chromatography and Ni-NTA column standards. Fusion protein purification does see some loss relative to unfused ELP, but these losses in ITC can be attributed to a decreased

efficiency in the ELPs phase transition property due to the relative hydrophobicity of the additional protein and the strength of its cysteine bonds[37].

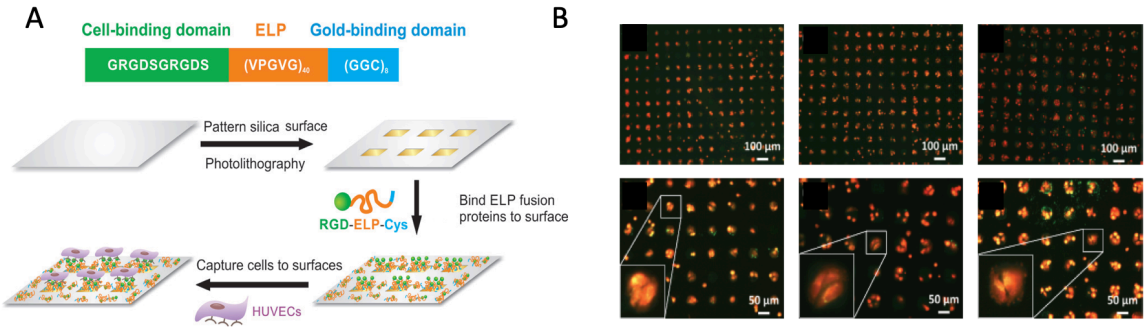

Figure 5. (a) Schematic of cell- and gold-binding ELP fusion protein sequence and process for creating cell patterns on peptide-modified gold surfaces. (b) Fluorescence images of cell (HUVECs) adhering to ELP fusion protein modified patterns. This figure is reprinted from Li et al., Figure 1[38].

Previous studies have demonstrated the ease of synthesizing ELP fusions, with high yields empowered by ITC protein purification. A primary advantage of ELPs is they are inert and do not interact appreciably with cells or other biomolecules[39], thereby providing a 'blank canvas' for engineered bio interactions by mitigating non-specific interactions. For example, cell- and material-binding domains can be inserted at precise locations and spacing within ELP chains, offering direct control over the cell-adhesive properties of the surface. Specifically, Carroll et. al. engineered a recombinantly synthesized ELP with both cell-binding (RGD) and gold-binding (cysteine-rich) domains (Figure 5). Cell-binding ELPs attached to 2D gold microstructures on a silicon oxide substrate enabled the creation of cellular patterns. This work highlights the ability to genetically engineer diverse functions into ELPs, as well as the biocompatibility ELPs can provide to biological systems.

Chilkoti and coworkers engineered an RNA binding ELP peptide by using a recursive directional ligation (RDL) technique[40] for concatenation of a cellular NA binding peptide known as RGG[41]. This cloning process involves RDL cycles of two halves of a

parent plasmid, each containing a copy of an insert, which are ligated together, thereby dimerizing the inserts and regenerating a functional expression plasmid. This process is carried out recursively to assemble an oligomeric gene encoding both ELP repeats and non-ELP blocks. The NA-binding ELP fusions were used to bind and recruit RNA within ELP coacervates for sequester/release of the RNA to control gene expression of a fluorescent protein in a transcription/translation system. The RGG domain has a large degree of disorder and thus, these ELP-RGG fusions are primarily disordered chains.

In this thesis, the aim is to build on previous work by creating fusion ELPs using Golden Gate cloning to engineer ELP-encoding DNA plasmids with a type IIS restriction enzyme landing site, corresponding to a location at the C-terminal end of a model ELP. Our hypothesis is that Golden Gate cloning will enable DNA oligomers encoding non-ELP blocks to easily and quickly be inserted into ELP plasmids for rapid synthesis and prototyping of NA-binding (and other types of) ELP fusions. The goal is to synthesize ELP fusions by concatenation of ordered and disordered NA-binding blocks derived from the human FUS protein. The impact of this work is that design and synthesis of these fusion proteins will be a next step for the overarching goals of (i) synthetic biology via programmable liquid-liquid phase separation of IDPs to control gene regulation in artificial cells and (ii) a move toward engineering synthetic NA-binding IDPs for regulation of bioactivity in living cells that can bind and release NA material (Figure 6) with potential utility in therapeutics and metabolic regulation.

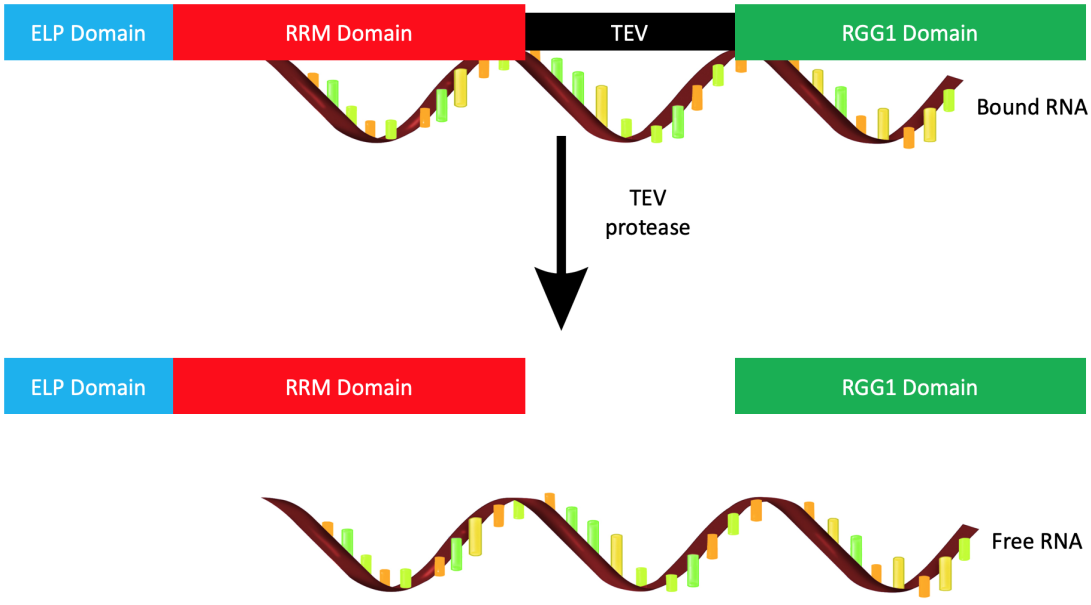

Figure 6. Illustration of proposed RNA release mechanism. While TEV cleavable linker is intact, bound RNA is held tightly to the construct. Upon cleavage of the TEV protease cleavable linker, the construct becomes a weaker RNA binder, which would enable RNA to disassociate more easily.

2. Experimental Theory

2.1 Thermodynamics of ELP Phase Separation

The thermodynamics of ELP phase separation ELPs undergo a phase transition that leads to the proteins forming a coacervate. This reaction is driven by a preference for self-interactions over protein–solvent interactions. There are several factors that influence the characteristics of the coacervate including sequence composition, molecular weight, salt concentration and type, and concentration (volume fraction). This allows for the establishment of comprehensive rules for the programmability of the coacervates[22]. It is likely that cells utilize subtle changes in the amino acid composition and concentration of IDPs to control their phase behavior and compartmentalization locally[22,42].

Using an archetypical set of IDPs based on ELP sequences, Carroll, Lopez et al. recently published a simple in-vitro model system for programming the intermolecular self-

assembly of multicomponent protein-rich liquid phases spanning multiple length scales and architectures. They showed ELP guest residues could be used to modulate protein phase behavior. ELPs exhibit lower critical solution temperature (LCST) phase transition behavior in water; above their cloud point transition temperature (Tt), they phase separate to form dense, protein-rich liquid coacervates. Using analyses based on Flory Huggins mean field theory, they interpreted and successfully predicted the phase behavior of ELP sequences and demonstrated rational design of a diverse library of multicomponent protein-rich structures, ranging from uniform nano-, and micro-scale protein-rich particles to multilayered granular structures (see Figure 7).

To understand LCST phase transitions of ELPs in microdrops, Carroll developed methods based on static and dynamic light scattering to obtain a full LCST phase diagram for a model ELP termed E3 (nominal sequence: $[(VPGXG)_{10}\text{-}GKG]_8$; X= $V_{80\%}, A_{20\%}$) comprising the binodal curve (green plot in Fig. 7) which comprises the transition temperature and its dependence on ELP volume fraction, and the spinodal curve (black plot in Fig. 7) that separates unstable and metastable regions within the binodal phase envelope. The spinodal is thermodynamically defined as the boundary of instability above which instantaneous and rapid phase separation occurs. This is exemplified in Figure 8, where the dynamics of spinodal decomposition of an ELP

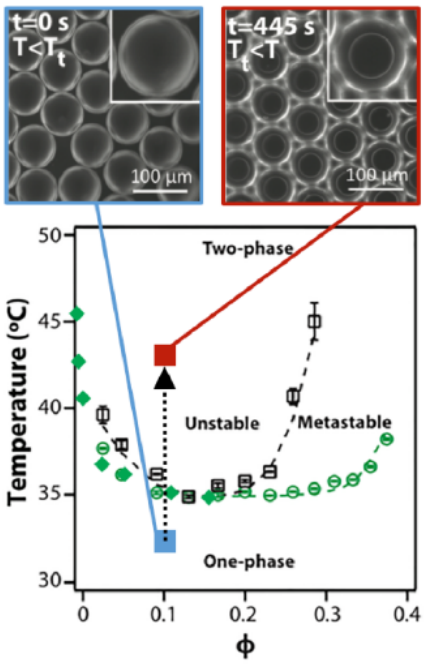

Figure 7. LCST phase diagram for E3 in water (f = vol. fraction E3). Open symbols obtained through light scattering; closed through turbidity measurements. Top: Micrographs before and after liquid/liquid phase separation of E3 in aqueous microdroplets in oil. This figure is adapted Simon et al. figure 2 [22].

solution are readily observable in aqueous microdroplets, where the incipient stages of an ELP phase transition occur homogeneously throughout the droplets; this rapid and simultaneous formation of protein-rich mesoscopic domains corroborates de-mixing through spinodal decomposition. To our knowledge, this model ELP (E3) is the first ELP to have a fully mapped phase diagram with both binodal and spinodal curves. Thus, E3 was chosen as our model ELP to create NA-binding fusions because phase behavior, including concentration-dependent transition temperatures and critical point, can be contrasted between the unmodified E3 polymer and E3 fusions comprising NA-binding blocks that are ordered, disordered, and combinations thereof.

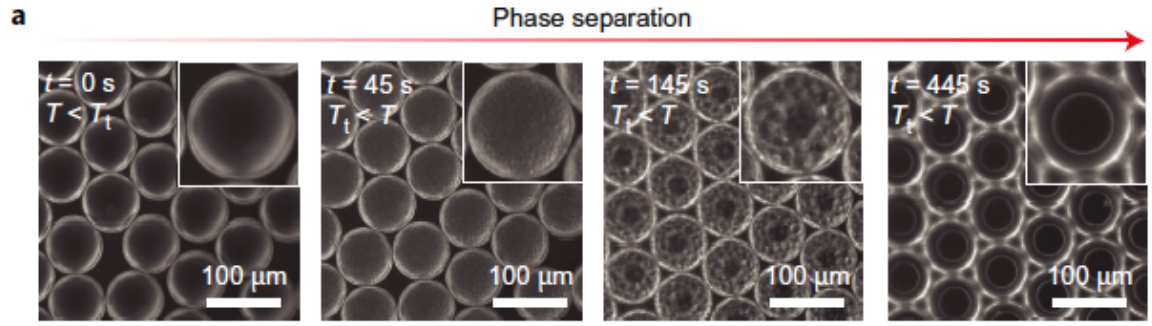

Figure 8. Time-lapse dark-field microscopy images of thermally induced spinodal decomposition of E3 within water microdrops; the phase-separated mesoscopic domains coarsen over time to form a single concentrated coacervate droplet (dark spherical centre) surrounded by a dilute protein solution[22]. This figure is adapted Simon et al. figure 2 [22].

The essential physics of polymer solution phase separation for a polymer is described by the Flory Huggins free energy:

$$\frac{\Delta F}{kT} = \frac{\phi}{N} ln(\phi) + (1-\phi) ln(1-\phi) + \chi\phi(1-\phi),$$

where N is the degree of polymerization (the number of effective monomers), and ϕ is the polymer volume fraction, hence the quantity $(1-\phi)$ is the water volume fraction. The first two terms represent the mean field entropy of mixing per lattice site and scales as the inverse of polymer chain length (N). The last term in F represents the mixing energy per

lattice site. The χ interaction parameter quantifies the balance of energy for chain and solvent interactions and can be thought of as an analogous energy 'penalty'–a large χ value indicates a large penalty for solvent-chain interactions. Above a critical value of χ, the energetic term will overcome the entropy that promotes mixing, such that the free energy of mixing comprises an unstable region and hence favors phase separation into coexisting polymer rich and polymer poor phases (Figure 9).

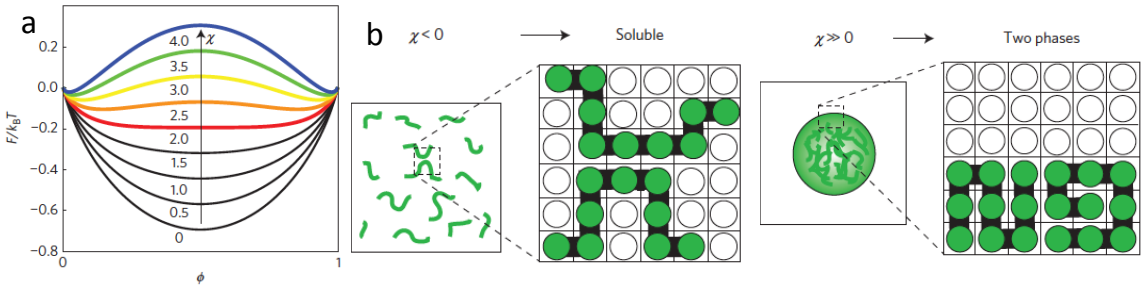

Figure 9. a. Normalized free energy of mixing (F/kBT) as a function of polymer volume fraction (ϕ) for increasing values of χ. Above a critical value of χ, unstable regions exist in F, and phase separation is favored (colored curves). b. Lattice representation depicting polymer solubility for a good solvent (soluble) and a poor solvent (separation into coexisting polymer rich and polymer poor phases. This image is reprinted from figure 2 of Brangwynne et al. [43].

For simple homopolymers comprising low overall charge, such as ELPs, solution phase separation can be described by χ, which quantifies the competition among three types of dipolar interactions, peptide-solvent, peptide-peptide, and solvent-solvent. However, the basic Flory Huggins framework cannot describe the phase behavior for more complex NA-binding polypeptides inspired by cellular proteins: (i) sequences enriched with charged/or aromatic side chains and (ii) proteins comprising ordered, folded blocks concatenated with intrinsically disordered regions. Importantly, the extent to which NA-binding ELP fusions of ordered folded domains affect the phase behavior of IDPs remains largely unexplored.[43] The hypothesis is that addition of ordered folded NA-binding domains to an ELP

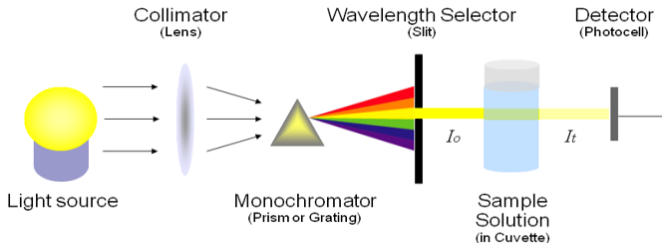

Figure 10. Illustration of the optical arrangement around a sample cuvette is given. White light exits the first optical fiber at the focus of a collimating lens. The collimated white beam crosses the turbid solution and is collected using another lens. This image was reprinted from Kevin Vo's online article "Spectrophotometry" [44].

homopolymer will result in dissimilar phase behavior (transition temperatures as a function of concentration, critical point, etc.) between the unmodified ELP polymer and the ELP fusion.

2.2 Turbidity Measurements

Turbidity measurements (optical density at 600 nm) were used as a primary tool to delineate the liquid-liquid phase boundary of ELP fusions. Temperature-controlled absorption spectrophotometry is an optical technique to measure the cloud point temperature of phase separating polymers.[44] This is achieved by detecting the onset of polymer phase separation, when the solution transitions to a purely scattering media due to formation of phase separated polymer drops comprising assemblies of collapsed thermal globules.

The polymer (e.g. protein) solution of interest is contained in a cuvette, where a collimated beam is crossing it along its width L (Figure 10). The light transmission through the sample is then measured at a single wavelength, typically selected to be in the visible spectral region for most standard protein solutions. First, the transmitted light I_0 is recorded using a "blank" reference cuvette containing only the solvent. Then the transmitted light I_t is recorded, where both the solvent and the polymer are mixed. This corresponds to the transmittance T, which is a measure of the light attenuation along the distance L. According

to the Beer–Lambert–Bouguer law and assuming a collimated beam of a single wavelength crossing a uniform medium, the transmittance T is given by[45]

$$T = {I_t}/{I_0} = e^{-OD},$$

where OD is the so-called optical density. OD corresponds to the average number of scattering and/or absorption events along the distance L. The quantity of importance for measuring cloud point temperatures by detecting the incipient stages of phase separation is the absorbance A, which is related to the optical density as

$$A = \frac{OD}{\ln(10)} = -log_{10}T.$$

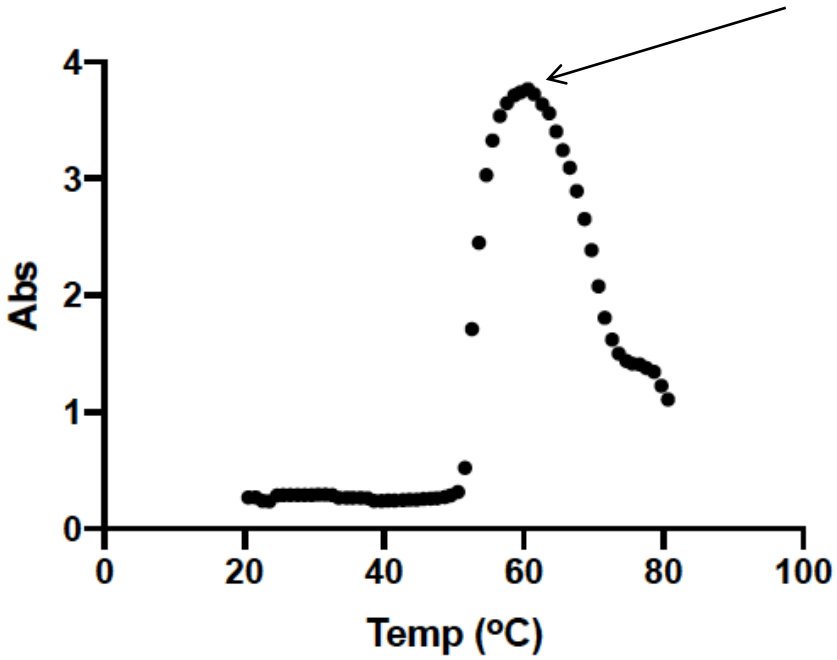

Figure 11. Absorbance as a function of temperature for an ELP solution (wavelength = 380 nm). At T > 50 °C, the ELP phase separates, resulting in a dramatic increase in measured absorbance. Eventually, the ELP droplets coalesce and begin to sediment out of the beam path, resulting in a decrease in absorbance (black arrow).

An absorbance of one corresponds to 10% of light transmittance, while an absorbance of two corresponds to only 0.1% transmittance. Therefore, A gives a quick indication of how much of the light has crossed the sample, while the optical density gives an indication of how many scattering or absorption events has occurred within the probed sample. A typical temperature-controlled absorbance measurement curve is shown in Figure 10. At temperature > 50 °C, the ELP phase separates and an increase in absorbance is measured as collapsed ELP globules scatter light. Eventually, phase separated ELP droplets coalesce and begin to sediment, falling out of the beam path, leading to a decrease in the absorbance (black arrow, Figure 11).

2.3 Gel electrophoresis

Biomolecules often comprise positive and negative electrical charges. Charged biomolecules in a hydrogel under electric field will migrate towards an electrode of opposite charge due to the phenomenon of electrophoresis. The relative mobility of biomolecules depends on factors such as net charge, charge/mass ratio, molecular shape and the porosity and viscosity of the matrix through which the molecule migrates. Protein molecules will migrate with dissimilar velocities, and this mobility depends on both the chemical characteristics of the molecule and on the physical properties of the hydrogel matrix. The velocity of movement, ν, of a charged molecule in an electric field E is described by:

$$v = \frac{E * q}{f},$$

where f is the frictional coefficient and q is the net charge on the molecule[46]. The frictional coefficient describes frictional resistance to mobility and depends on a number of factors

such as mass of the molecule, its degree of compactness, buffer viscosity and the porosity of the matrix in which the experiment is performed. The number of positive and negative charges in the molecule determines the net charge.

Folded proteins comprise a host of different geometries and their surfaces are mosaics with respect to the distribution of amino acid groups with different chemical properties. Because of this diversity (i.e. surface charges and geometries) the molecular weights of folded proteins cannot be simply determined by v in an electric field. For example, positively and negatively charged proteins would migrate to different electrodes under applied electric field. Hence, protein populations must be converted to a uniform geometry to impart a uniform charge/mass ratio to the proteins.[47] To resolve protein samples according to their molecular weight, SDS PAGE is used, where proteins are boiled in the presence of an anionic surfactant, sodium dodecyl sulfate (SDS) (Figure 12). The combination of heat and detergent is sufficient to break the many noncovalent bonds that stabilize protein folds. The amphipathic molecule SDS, consisting of a hydrophobic 12-carbon chain and an anionic sulfate group, permeates the protein interior and binds to hydrophobic groups, transforming the chain to a random coil with negative charges along the chain length.

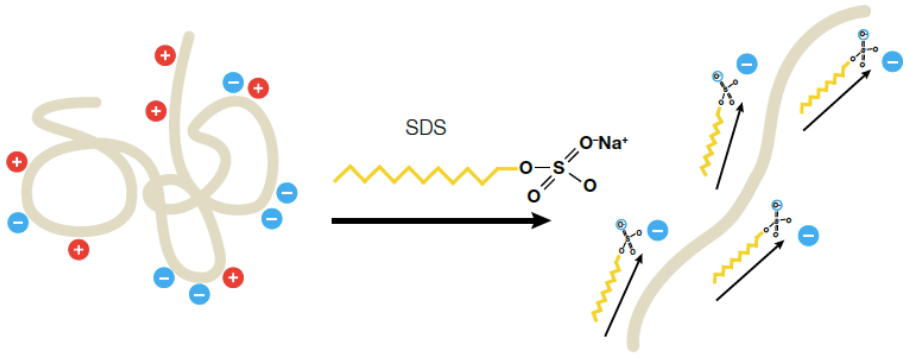

Figure 12. Effect of SDS on protein conformation (ordered to random coil) and charge. This figure is adapted from Manns et al. figure 2 [47].

To control the frictional coefficient f, a polyacrylamide electrophoreses gel was used. Polyacrylamide gels are fabricated using free radical polymerization of acrylamide and Sa crossliner such as bis-acrylamide in the presence of a radical initiator. The pore size of the gel is modulated by the ratio of total monomer concentration (%T in g/100 mL) and weight percentage of crosslinker (%C in g crosslinker/total weight). Higher %T results in a larger polymer-to-water ratio and smaller pore sizes. Gels can be cast with a gradient to enhance separation of proteins with a range of molecular weights. For the protein migrations expected for this thesis project, gel gradient of 4%–20% was selected.

3. Materials and Methods

3.1 Materials

For cloning, the following materials were purchased from New England Biolabs (NEB): Novagen KOD Hot Start DNA polymerase, (Merck KGaA, Darmstadt), supplied with 10X buffer, 25 mM MgSO4, and 2 mM dNTPs. Custom-made primers, oligonucleotides, and gene blocks were purchased from a commercial vendor (Invitrogen, Integrated DNA Technologies). To purify the PCR products, NucleoSpin Extract II kit was purchased from Takarobio (Macherey Nagel, Duren).

The Golden Gate restriction endonucleases XbaI (10 U/µL) and BsaI (10 U/µL), were purchased from New England Biolabs (NEB) supplied with 10 NEBuffer 4, (200 mM Tris–acetate pH 7.5, 100 mM magnesium acetate, 500 mM potassium acetate, 10 mM dithiothreitol), and T4 DNA Ligase 3 U/µL supplied with 10 ligation buffer (300 mM Tris–HCl pH 7.8, 100 mM MgCl2, 100 mM DTT, 10 mM ATP).

To measure DNA concentration, the NanoDrop ND2000 spectrophotometer

(Peqlab, Erlangen) was used. Bacteria was suspended in Luria-Bertani (LB) Medium: 1 % bacto-tryptone, 0.5 % yeast extract, 1 % NaCl in deionized water, adjusted to pH 7.0 with 5 N NaOH. For plasmid selection, antibiotics ampicillin and kanamycin were used: filter-sterilized stocks of 50 mg/mL in H_2O (stored in aliquots at -20C) are diluted 1:1,000 (final concentration: 50 μg/mL) in the final volume medium after the medium has been autoclaved and cooled down. For preparation of bacterial growth plates, the stock was diluted 1:500 (final concentration: 40 μg/mL) in an appropriate amount of LB agar after autoclaving/melting and cooling down.

BL21 (DE3) chemically competent (C2527) protein expression *E. Coli* cells were purchased from NEB. NanoDrop ND2000 (Peqlab, Erlangen) was used to measure optical density of sample measured at wavelength 600nm, OD_{600}. OD_{600} is used to estimate the concentration of bacterial cells in solution.

DNA constructs were sent to an external contractor, GeneWiz (South Plainfield, NJ), for sequencing. Sequence data were analyzed using the SnapGene Viewer software and Lalign online software. Primers T7 (TAATACGACTCACTATAGGG) and U3Rev (TGTAAAACGACGGCCAGT) were used for sequencing of inserts cloned in pET-24a+ derived vectors.

E. coli cell cultures were grown in Terrific Broth (TB) media purchased from MO BIO laboratories Inc. (Carlsbad, CA). DNA Miniprep extraction kit, DNA gel purification kit, and polymerase chain reaction (PCR) purification kits were purchased from Qiagen Inc. (Germantown, MD). *E. coli* cell cultures were collected by centrifugation in a Sorvall RC3B Centrifuge Plus from ThermoFisher Scientific (Waltham, MA).

To verify size and purity of the fusion proteins, the following SDS-PAGE solution components were purchased from Thermofisher Scientific: 40% Acrylamide Ammonium Persulfate, N,N'-Methylene bisacrylamide, DDI H2O, 1.5 M Tris• Cl, 125 mM Tris HCl, pH 6.8, 1M Tris base, 20% Glycerol, Glycine 4%, SDS 20%, ß-Mercaptoethanol, and Bromophenol Blue.

The optical density at 350 nm (OD350) of the fusion protein constructs was measured as a function of temperature on a UV−vis spectrophotometer equipped with a multicell thermoelectric temperature controller, Cary 300 by Agilent Technologies (Santa Clara, CA).

Fluorescent microscopy images were obtained using an Olympus IX83 inverted microscope series (Shinjuku City, Tokyo) equipped with ORCA-Flash4.0 V3 Digital CMOS camera (Hamamatsu Photonics, Hertfordshire, UK) and Olympus cellSens software.

3.2 Cloning

Golden Gate cloning method was used to create ELP fusions with synergistic arginine-rich (RGG) domains and folded RRM blocks that only bind RNA strongly when they are both present on the chain. Golden Gate cloning is a molecular cloning method that allows a for simultaneous and directional assembly of multiple DNA fragments into a single piece using Type IIs restriction enzymes and T4 DNA ligase (Figure 13). The principle of this method is based on the ability of type IIS enzymes to cleave outside of their recognition site, allowing two DNA fragments flanked by compatible restriction sites to be digested and ligated seamlessly. Since the ligated product of interest does not contain the original type IIS recognition site, it will not be subject to re-digestion in a

restriction-ligation reaction. However, all other non-viable products that reconstitute the original site will be re-digested, allowing their landing sites to be made available for further ligation, leading to formation of an increasing amount of the desired product with increasing time of incubation. Since the sequence of the overhangs at the ends of the digested fragments can be chosen to be any 4-nucleotide sequence of choice multiple compatible DNA fragments can be assembled in a defined linear order in a single restriction-ligation step[48].

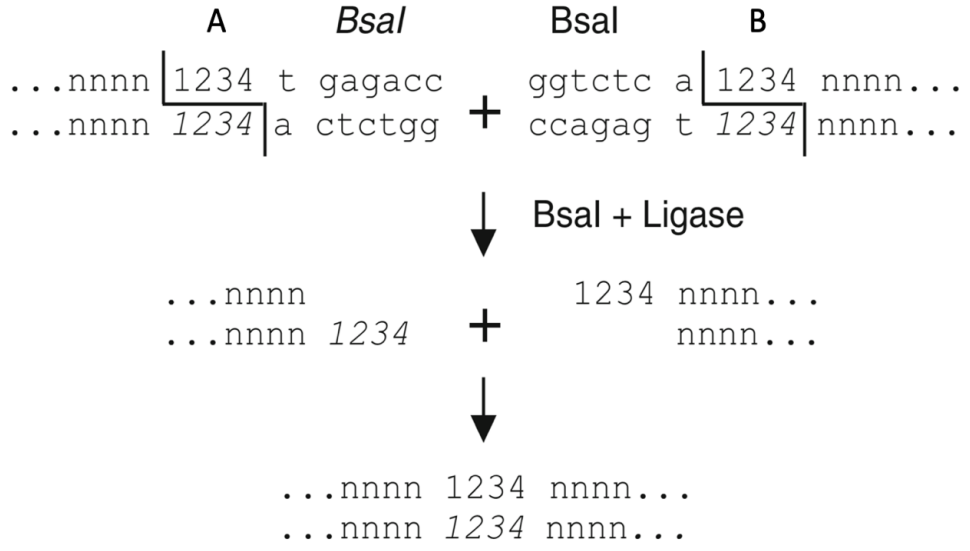

Figure 13. Golden Gate Cloning Method. Two DNA ends terminated by the same 4 nucleotides (sequence A and sequence B, composed of nucleotides 1234, complementary nucleotides noted in italics) flanked by a BsaI recognition sequence, GGTCTC, form two complementary DNA overhangs after digestion with BsaI, allowing for seamless concatenation of two different DNA strands. This figure has been adapted from figure 1 of Engler and Marillonet, 2013 [48].

Fusion sites consist of 4-nucleotide sequences that were used for restriction enzyme digestion and ligation for the DNA assembly step. An example of this is shown in Figure 13. Fusion sites of 4 nucleotides were selected corresponding to type IIS restriction enzymes BsaI and XbaI. Next designed fragments were amplified the designed fragments by PCR, cloning them in intermediate vectors, and sequencing them. Alternatively, these

steps can also be replaced by simply ordering the fragments of interest from a gene synthesis company. This was done by ordered gene blocks with the BsaI landing site on the C-terminal end of the model ELP from IDT. Primers were designed so that the overhangs created by digestion of the amplified products with BsaI and XbaI correspond to the sequence of the chosen fusion sites. Therefore, the sequence ttggtctca was added to the primer sequence. A main requirement for Golden Gate assembly is to not have any internal BsaI sites present within any of the DNA fragments to assemble, as the presence of a BsaI site within one of the modules would lead to redigestion of the shuffled DNA sequences containing such fragment. These linear molecules would not transform into E. coli and therefore will not present false positives. Cloning vectors for generating entry clones for shuffling need to fulfill two requirements: (i) they should not contain any restriction site for the type IIS enzyme chosen for shuffling, and (ii) the antibiotic resistance gene of the entry vector should be different from the one in the destination vector. The vector, pET-24a+ was selected as it meets these requirements. Once entry constructs and the recipient vector were made and sequenced, performing DNA assembly requires pipetting all components into a reaction mixture, incubating the mixture in a thermocycler, and transforming the ligation mix into chemically competent cells. An important factor is to use an equimolar amount of DNA for each of the module sets and the destination vector. The entire ligation product is transformed into chemically competent BL21 (DE3) cells and grown overnight. The cloning vectors were then extracted and sequenced.

3.3 Purification

ELPs undergo a sharp and reversible phase transition at the LCST, this temperature is also commonly referred to as the inverse transition temperature (Tt). Below its LCST,

an ELP adopts a close to random coil conformation, is well solvated and is hence highly soluble in aqueous solution. When the solution is heated and the LCST is reached, ELPs become insoluble and form large micron-size aggregates that are visible to the naked eye[44,45]. This transition is completely reversible, so that the aggregated polypeptide completely dissolves when the temperature is lowered below the LCST of the ELP. The LCST decreases with increasing ionic strength, polymer concentration, and polymer length. The nature of the guest residue also affects the transition temperature, as more hydrophobic amino acid residues lower the transition temperature and *vice versa*[46].

The ELP tag provides a convenient stimulus responsive handle to separate the target protein from other cell lysate contaminants by cycling the ELP fusion protein through its phase transition in cell lysate. After expression, the cells are lysed, and the cell lysate debris is removed by centrifugation. The ELP fusion protein is then separated from soluble contaminants by triggering the phase transition of the ELP fusion protein. Increasing the solution temperature above the inverse transition temperature of the ELP fusion protein or increasing the salt concentration to decrease the transition temperature below solution temperature are two convenient methods to trigger the phase transition of an ELP fusion protein. As increasing temperature might denature certain proteins, it is preferable trigger the phase transition at room temperature by increasing the salt concentration. Triggering the phase transition causes the ELP fusion protein to form aggregates. These aggregates are typically highly concentrated with the ELP fusion protein and are separated from the soluble fraction of the cell lysate by centrifugation at the same temperature. The supernatant is decanted and then discarded. This centrifugation step is called a “hot spin”, though this step is carried out at around ambient temperature in the centrifuge (roughly

24°C) and to distinguish it from the next step of the process that involves centrifugation at low temperature. In this subsequent step, the "cold spin", the pellet of the aggregated ELP fusion protein is dissolved in cold PBS buffer, and then centrifuged at 4°C to remove any insoluble contaminants that may have been trapped in the mass of the aggregated ELP fusion protein. The soluble protein is decanted and retained, and this completes one cycle of inverse temperature cycling (ITC)[22] (Figure 14). Multiple rounds of ITC are typically required to purify ELP fusion protein.

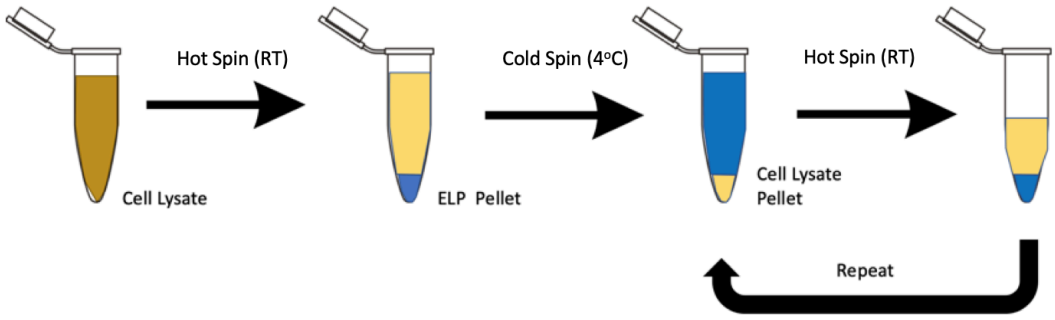

Figure 14. Schematic diagram Protein Purification Cycle. Triggering the phase transition causes the ELP fusion protein to form an ELP pellet of highly concentrated aggregates. The supernatant is decanted and then discarded. This centrifugation step is the "hot spin" (room temperature). In the subsequent step, the "cold spin", the pellet of the aggregated ELP fusion protein is dissolved in cold PBS buffer, and then centrifuged at 4°C to remove any insoluble contaminants that may have been trapped in the mass of the aggregated ELP fusion protein. Cycle is repeated until no contaminate pellet forms. This figure is adapted from Chilkoti et al. protocol [23].

3.4 Fusion Protein Synthesis Verification and Protein Characterization

DNA sequence was verified by sanger sequencing via GeneWiz sequencing, those sequences were then aligned to our designated sequence. Alignment was done by inserting the sequence of the designed protein column one of the Lalign software and inserting the GeneWiz sequencing data into the second column.

Protein purity and molecular weight distribution was characterized by 4%–20% gradient tris-glycerol sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-

PAGE). Purified protein was re-suspended in 20 µl of 2x sample buffer. Tubes were incubated in boiling water for 5 min, then centrifuged at 12,000 x g for 30s. 20 µl of protein solution was then placed into a micro-centrifuge tube and an equal volume of 2x sample buffer. Tubes were incubated in boiling water for 5 minutes then centrifuged at 12,000 x g for 30 seconds. To run the gel, the comb was removed and placed into pre-cast gel into Mini-Protean II electrophoresis apparatus. Freshly prepared 1x running buffer (300 ml) was added to both chambers of the apparatus. The prepared samples were loaded, via pipetting, into the wells of the gel. The gel was run at 100 V until the dye front migrated into the running gel (~15 min), and then increased to 200 V until the dye front reached the bottom of the gel (~45 min).

To stain and de-stain the gel, the gel was removed from the apparatus and the spacers and glass plates. Then the gel was placed into a small tray with ~20 ml staining solution and was left to stain for 1 hour with gentle shaking. Then the stain was poured off and saved. 5 ml DDI H2O was added to the tray to de-stained for 5 min with gentle shaking. A final wash was done with DDI H2O by suspending gel in ~30 ml while shaking gently for 1 hour. Lanes with only one protein band, aligned to the corresponding band in the ladder, indicated a pure sample of the correct molecular weight. Protein concentration was determined by weighing and suspending lyophilized protein in buffered solution and was verified by nanodrop spectrophotometer.

Concentration dependent transition temperatures of the fusion ELP was measured as a function of as a function of temperature via turbidity measurements using a UV−vis spectrophotometer equipped with a multi-cell thermoelectric temperature controller (Cary 300). The ELP fusion was suspended in PBS at concentrations of 5, 50, 100, 200 and

250μM. Added at least 100μL of each solution to heat resistant quartz cuvettes. Samples were heated at a rate of 1 °C/min in a range from 20 °C to 80 °C. The Tt was defined as the temperature corresponding to the inflection point of the turbidity plot. The data was analyzed using Prism8 Graphpad software by plotting turbidity against the temperature.

4. Results

4.1 ELPs as a model biopolymer

To approach the goal of developing a protein that can bind and release nucleic acid material our model protein was selected. ELPs were selected as the model biopolymer because their modular nature allows us to include nucleic acid binding motifs and allows us to tune protein-RNA interactions. ELP peptide repeats are not strong RNA binders but have potent phase behavior and they inhibit protein expression because they have the potential to, non-specifically, bind to translational elements and ribosomal assemblies[2] making ELPs ideal for fusion to nucleic acid binding elements. Membraneless organelle formation, however, is governed by more complex multivalent interactions. Key concepts of programmable self-assembly, regulation of gene expression, and RNA selectivity in recruitment[2,36] were used to design our biomaterial. Protein granules are highly dynamic organizers of mRNA translation and associated metabolic pathways and our biomaterials model is built on this regulatory pathway[40].

The ELP fusion complexes are designed to be responsive to environmental cues, particularly temperature and protease activity. Here the temperature sensitivity takes advantage of the ELP's LCST Tt so upon increasing the temperature above the Tt, the ELP fusion should collapse and undergo liquid-liquid phase separation to form ribonucleoprotein coacervates enriched in DNA, effectively sequestering the DNA. These

complexes are also strong binders of both RNA and DNA and upon exposure to the TEV protease, the following cleavage of the synergistic nucleic acid binding blocks RRM-TEV-RGG at the TEV cleavage site ELP-RRM fusion will release captured RNA. In this work, the goal is to show that such ELP fusions can be successfully developed, for the eventual application of ELP fusion proteins with folded and intrinsically disordered blocks that work in unison to bind nucleic acid[40].

4.2 ELP Golden Gate Fusion Proteins

Controllable RNA binding and release is meant to be achieved by linking two RNA-binding peptide blocks into ELP chains that, individually, are weak RNA binders, but when concatenated together, they achieve strong RNA binding when on a polypeptide chain. In this work, the goal is to establish that these ELP fusion proteins can be successfully genetically cloned, synthesized, and purified. Three different fusion-proteins were designed and made: one by concatenating the weak RNA binding protein FUS RRM to the N-terminal end of the ELP, another by concatenating the weak binding proteins FUS RGG2 with FUS RRM (with dissociation constants KD = 61 μM and KD = 90 μM, respectively), and a cleavable fusion by concatenating a TEV protease cleavable domain between the FUS RRM and FUS RGG. When RGG2 and RRM are adjacent on a polypeptide chain, they work together to form a strong RNA binding protein with K_D = 2.5 μM, which approaches that of full length FUS protein (K_D = 0.7 μM)[40,42]. Model DNA targets with G-quadruplex and double stranded structural motifs are optimal binding targets because the proposed RNA-binding motifs are known to bind strongly these nucleic acid geometries[42].

The two RNA binding domains will be linked using a short tobacco etch virus (TEV) protease cleavable linker (ENLYFQ) and flanked by ELP domains. In future work, E3-12 should be exposed to TEV protease in vitro to cleave the linker and demonstrate the controlled release of RNA from granules, thereby controlling RNA translation. This linker was selected because the TEV domain is a short polypeptide sequence and therefore will minimally interfere the with the fusion protein's nucleic acid binding and phase behaviors. It is expected that the phase behavior of proteins bound to nucleic acids will be altered compared to the phase behavior of the ELP in the absence of RNA.

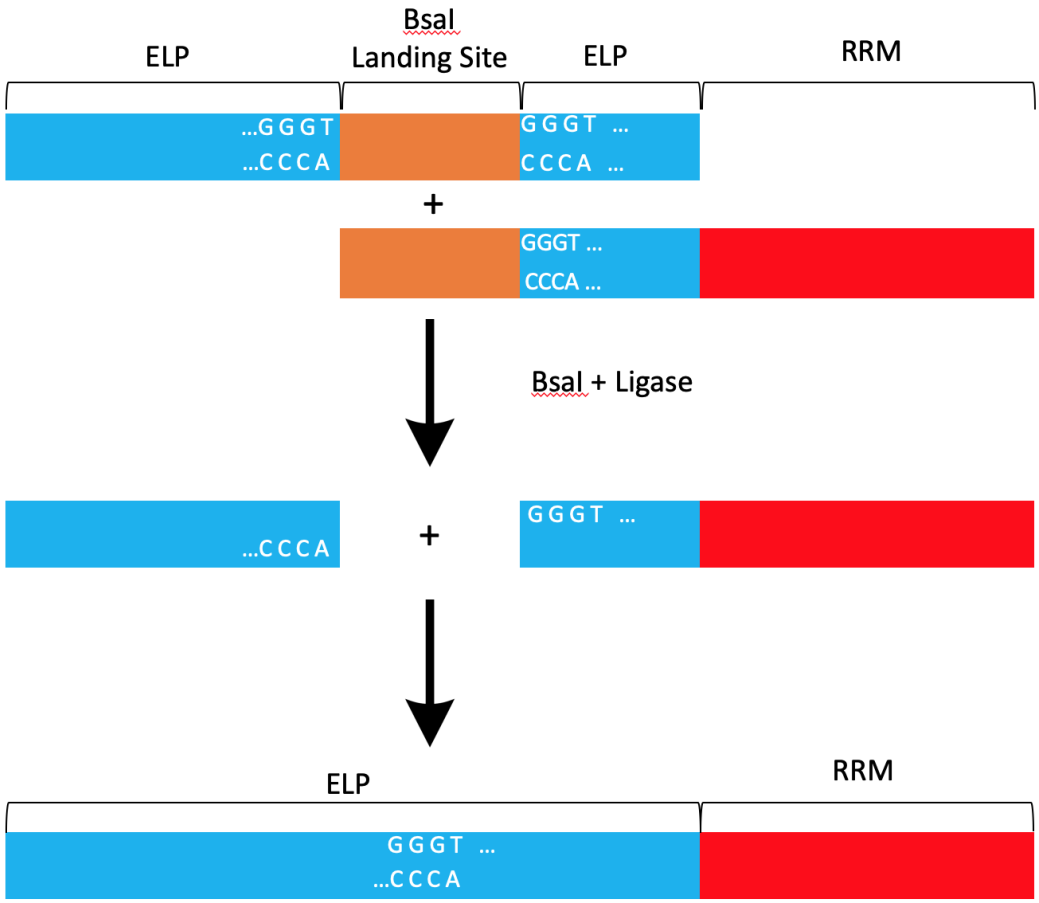

Figure 15. Schematic illustration depicting Golden Gate assembly of DNA sequences encoding a fusion protein comprised of an ELP concatenated with an RNA-binding peptide block.

4.3 Sequencing

Golden Gate cloning was used to create genes encoding protein fusions comprising NA binding blocks fused to E3 ELP. First, a BsaI restriction enzyme landing site (gggttgagccaaatttggtctca) was added to the terminal end of the gene encoding E3 (Figure 15). Golden Gate cloning was then used to fuse the C-terminal end of the ELP to the N-terminal end of the nucleic acid binding block. Protein sequence of the E3-BSA landing site and three new constructs were verified by sanger sequencing (Figure 16). By this sequencing, it was verified that three new constructs were developed that seamlessly fused the E3 ELP to the FUS RRM RGG, FUS RRM, and FUS RRM RGG with the TEV protease cleavable linker, which were designated E3-10, E3-11, and E3-12, respectively (Figure 17). The protein sequence for E3-11 has the E3 protein sequence fused to the RRM binding domain (Table 1). This confirmation of the E3-11 sequencing suggests that these fusion ELP proteins can be accurately cloned using Golden Gate cloning.

E3	MGVGVPGVGVPGAGVPGVGVPGVGVPGVGVPGVGVPGAGVPGVGVPGVG VPGGKGVGVPGVGVGAGVPGVGVPGVGVPGVGVPGVGVPGAGVPGVGVP GVGVPGGKGVGVPGVGVPGAGVPGVGVPGVGVPGVGVPGVGVPGAGVPG VGVPGVGVPGGKGVGVPGVGVPGAGVPGVGVPGVGVPGVGVPGVGVPGA GVPGVGVPGVGVPGGKGVGVPGVGVPGAGVPGVGVPGVGVPGVGVPGVG VPGAGVPGVGVPGVGVPGGKGVGVPGVGVPGAGVPGVGVPGVGVPGVGVP GVGVPGAGVPGVGVPGVGVPGGKGVGVPGVGVPGAGVPGVGVPGVGVPG VGVPGVGVPGAGVPGVGVPGVGVPGGKGVGVPGVGVPGAGVPGVGVPGV GVPGVGVPGVGVPGAGVPGVGVPGVGVPG
RRM	PRDQGSRHDSEQDNSDNNTIFVQGLGENVTIESVADYFKQIGIIKTNKKTGQP MINLYTDRETGKLKGEATVSFDDPPSAKAAIDWFDGKEFSGNPIKVSFATR

RGG	RADFNRGGGNGRGGRGRGGPMGRGGYGGGGSGGGGRGGFPSGGGGGGGQQR
TEV	ENLYFQG

Table 1. Table listing the protein sequences of the ELP E3, FUS RRM, FUS RGG, and TEV protease cleavable domain.

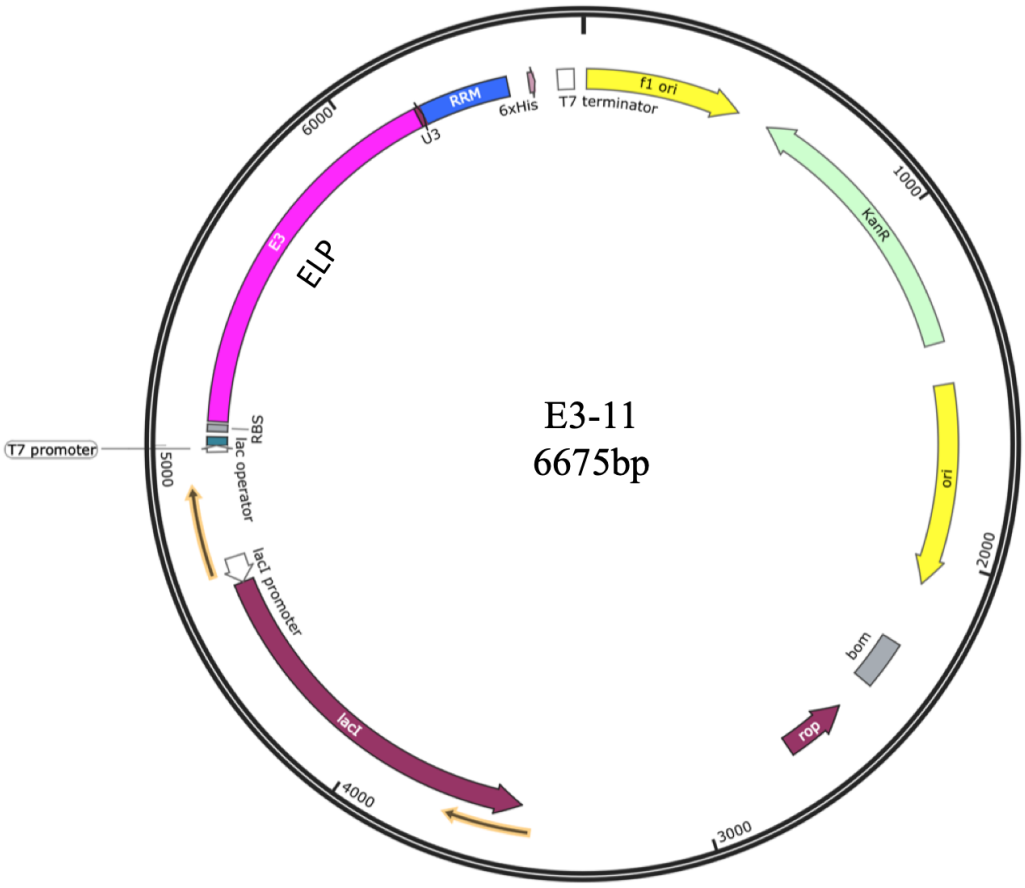

Figure 16. SnapGene plasmid map of E3-11 protein construct. ELP E3 (pink) is seamlessly concatenated with the FUS RRM2 (blue) inserted into pET-24a+ cloning vector.

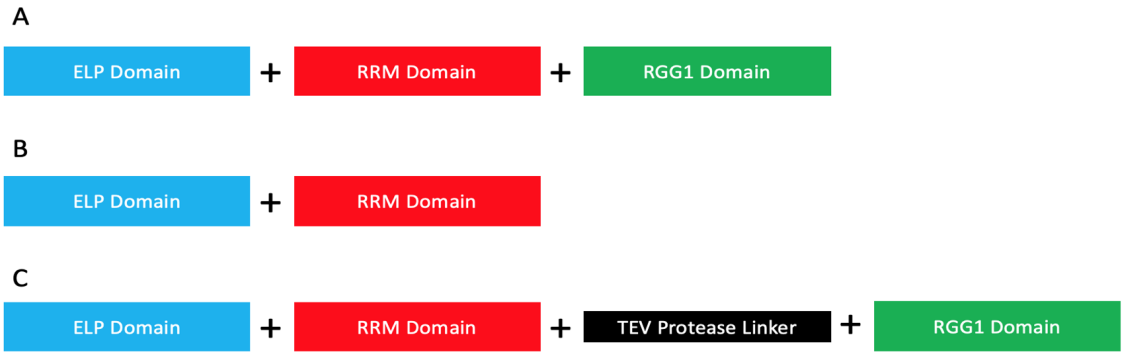

Figure 17. Schematic Illustration of the designed ELP-RNA binding constructs. A) E3-10. ELP concatenated with FUS RRM and RGG1, should bind nucleic acids strongly. B) E3-11. ELP concatenated with FUS RRM, should be a relatively weak nucleic acid binder. C) E3-12. ELP concatenated with FUS RRM, TEV protease cleavable linker, and RGG1, should bind nucleic acid strongly until cleavage of protease cleavable linker, then construct should be a weak binder of nucleic acids.

4.4 Protein Synthesis and Purification

After the overnight expression of the fusion protein in E. Coli, the protein constructs were purified by inverse transition cycling. Construct E3-11 required seven rounds of ITC, while E3-10 required ten rounds before there was no visible pellet after the hot spin, but a visible pellet remained after a cold spin. This is an indication that the protein is pure because below the ELP Tt, the ELP is water soluble so if no pellet forms after a cold centrifugation, it is assumed that there are no other contaminants in the sample. E3- 12 was not synthesized because of lack of time. Protein purity and molecular weight was also verified by SDS-PAGE. The SDS–PAGE analysis used precast Mini-PROTEAN 10–20% gradient gels (Bio-Rad, Hercules, CA) with a discontinuous buffer system, stained with Coomassie brilliant blue. The proteins were considered pure when there was only a single band per lane, corresponding to the construct's molecular weight (Figure 18). E3-10 is 48.5kD and E3-11 is 45.9kD, both of while align to the corresponding size. This result strongly suggests that the proteins purified were E3-11 and E3-10. Protein concentration was determined by nanodrop 3000. After purification, the highest concentration of E3-11

was 0.258μM and highest concentration of E3-10 was 0.144μM. The successful purification of fusion ELP proteins suggests that not only can these fusion proteins be synthesized, but they can be purified by inverse transition cycling, an important result for the overall engineering goal of creating fusion proteins for the eventual encapsulation of NA- material.

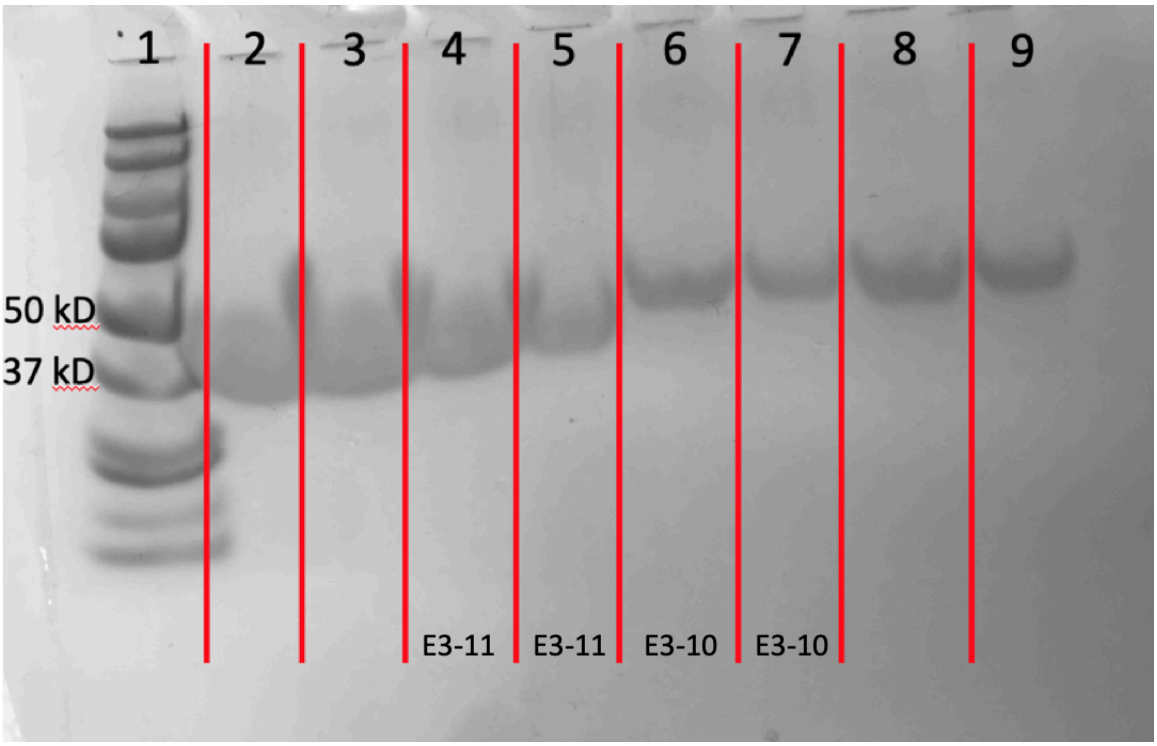

Figure 18. SDS-Page size exclusion of E3-10 (lanes 6 and 7) and E3-11 (lanes 4 and 5) showing that E3-10 aligns to correct size at 49kD and E3-11 aligns at correct size at 46kD. Proteins are also pure as a single band appears per lane. Red bars are to distinguish individual lanes. Lanes 2 and 3 contained very high concentrations of protein.

4.5 ELP Fusion Phase Behavior

The partial phase diagram of E3-11 and the transition temperature as a function of dilute protein concentrations, was determined using turbidity measurements taken on a Cary 300 ultraviolet-visible spectrophotometer equipped with a multi-cell thermo-electric temperature controller (Varian Instruments, Walnut Creek, CA). E3-11 was the only

protein that could be produced in high enough yields to determine the phase behavior. The optical absorbance at 350 nm of ELP fusion solutions were measured in the 20–80°C range (Figure 19). The Tt was determined from the inflection point of the transition-induced change at a heating rate of 1°C min. Increase in temperature beyond Tt results in a sharp increase in turbidity over an approximately 2°C range to a maximum value, because of aggregation of the ELP. At 50uM and 100uM, E3-11 (Figure 20) and E3 (Figure 21) follow a very similar trend. At 50uM and 100uM, the Tt of each of the proteins are very similar. As protein concentration is increased, E3 Tt steadily decreases while E3-11 Tt takes a sharp increase. These E3 measurements are not a perfect comparison however, as E3 was suspended in 100mM sodium phosphate while E3-11 was suspended in 1X PBS. Turbidity measurements and microcopy images of E3 protein were collected by graduate student, Telmo Diez Perez[49].

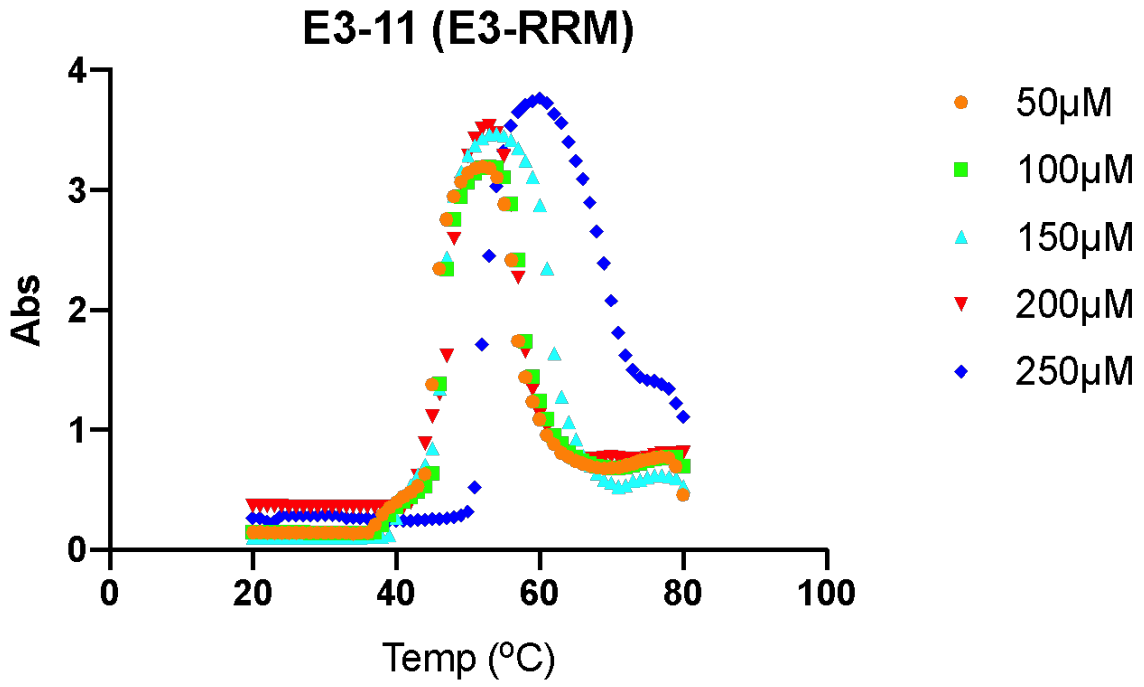

Figure 19. Temperature controlled turbidity measurements for E3-11 to find Tt as a function of E3-11 concentration.

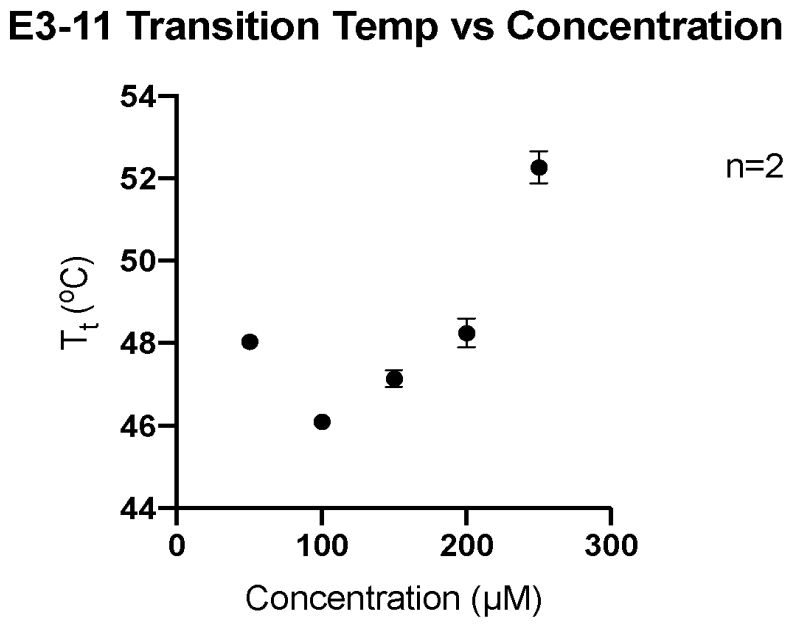

Figure 19. Plot of transition temperature of E3-11 as a function of protein concentration.

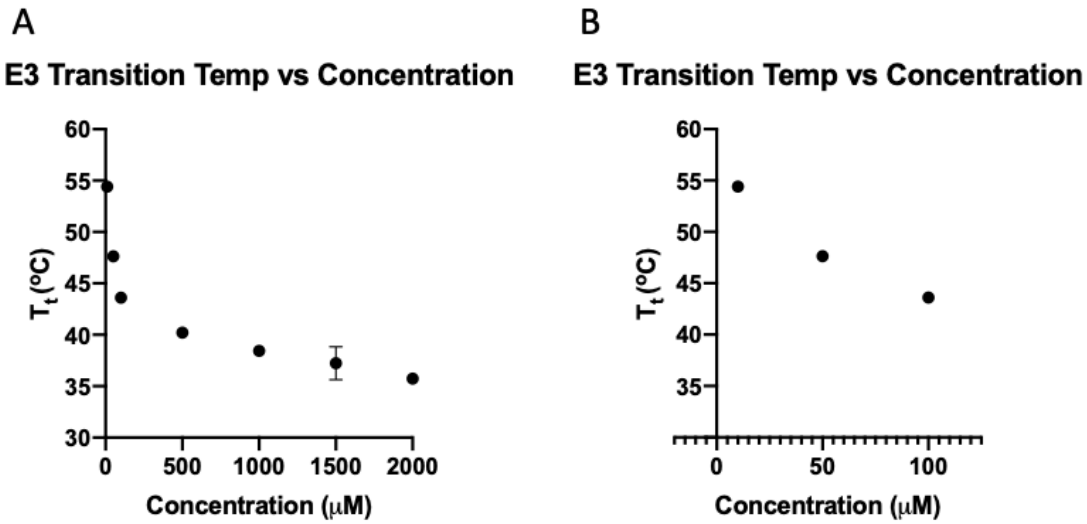

Figure 20. A) Plot of transition temperature of E3 as a function of protein concentration. B) First three data points of plot A.[49]

4.6 Ribonucleoprotein Assemblies

To further explore the potential of the ELP fusion E3-11 to bind and capture nucleic acid material. Their ability to form synthetic stress granules was investigated using confocal

fluorescence microscopy. To prepare the protein for imaging, the protein is suspended in PBS then mixed into an oil phase which generates microdrops that are rich with the suspended protein. E3-11 fusion protein was at a final concentration of 200μM and the fluorescently labeled DNA had a final concentration of 500nM. Fluorescently labeled DNA was incubated with the E3-11 fusion and allowed to associate to labeled DNA. DNA bound to E3-11 fusion because FUS binds nucleic acid material and DNA is easy to handle and fluorescently label. Under a confocal fluorescent microscope, the aggregates at 20°C and 50°C were monitored. At 200μM, the Tt of E3-11 is estimated to be about 48°C. As the above Tt was reached, the E3-11 fusion appears to capture DNA within the formed coacervate (Figure 22). The transition temperature could be altered in the presence of nucleic acid material.

Aggregation of the ELP fusion protein E3-11 does not seem to be reversible however, as the aggregates were not re-solubilized completely when the temperature was lowered below the Tt as only E3 does (Figure 23). E3 is suspended in 100mM sodium phosphate, so it is not a perfect comparison. This interesting result in the shows a potential downshift in Tt during cooling when compared to the Tt while the protein was heating. This could be happening because while in the aggregated state, this partially ordered polymer becomes entangled in the higher order components (Figure 24) of the protein, as ELPs have been shown to do when introduced to ordered proteins[50,51], interfering with the reverse temperature transition. The presence of nucleic acid material could also be contributing to this behavior.

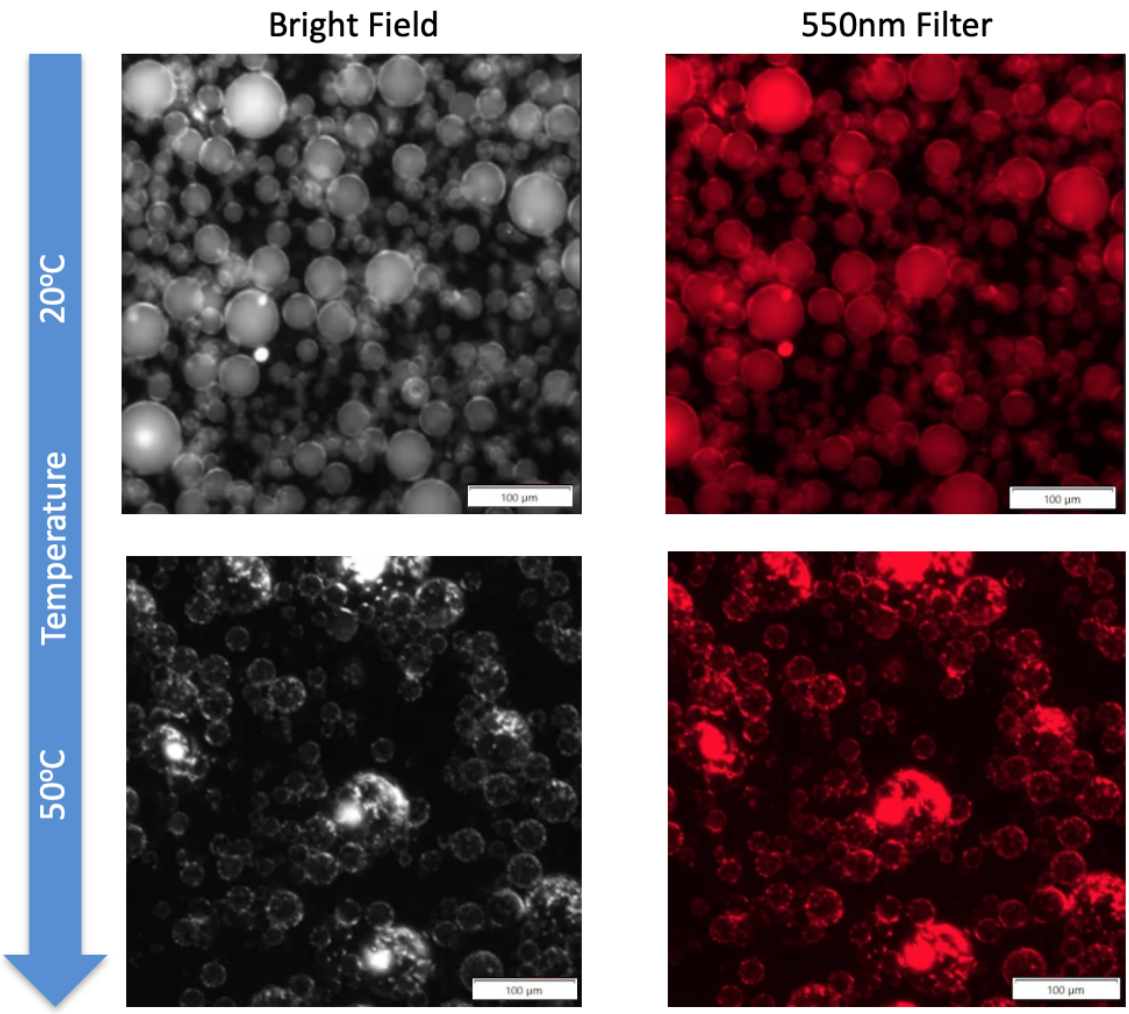

Figure 21. Confocal fluorescence and bright field microscopy images of E3-11 with CY3 labeled DNA. Formation of aggregates enriched with captured DNA as temperature is increased.

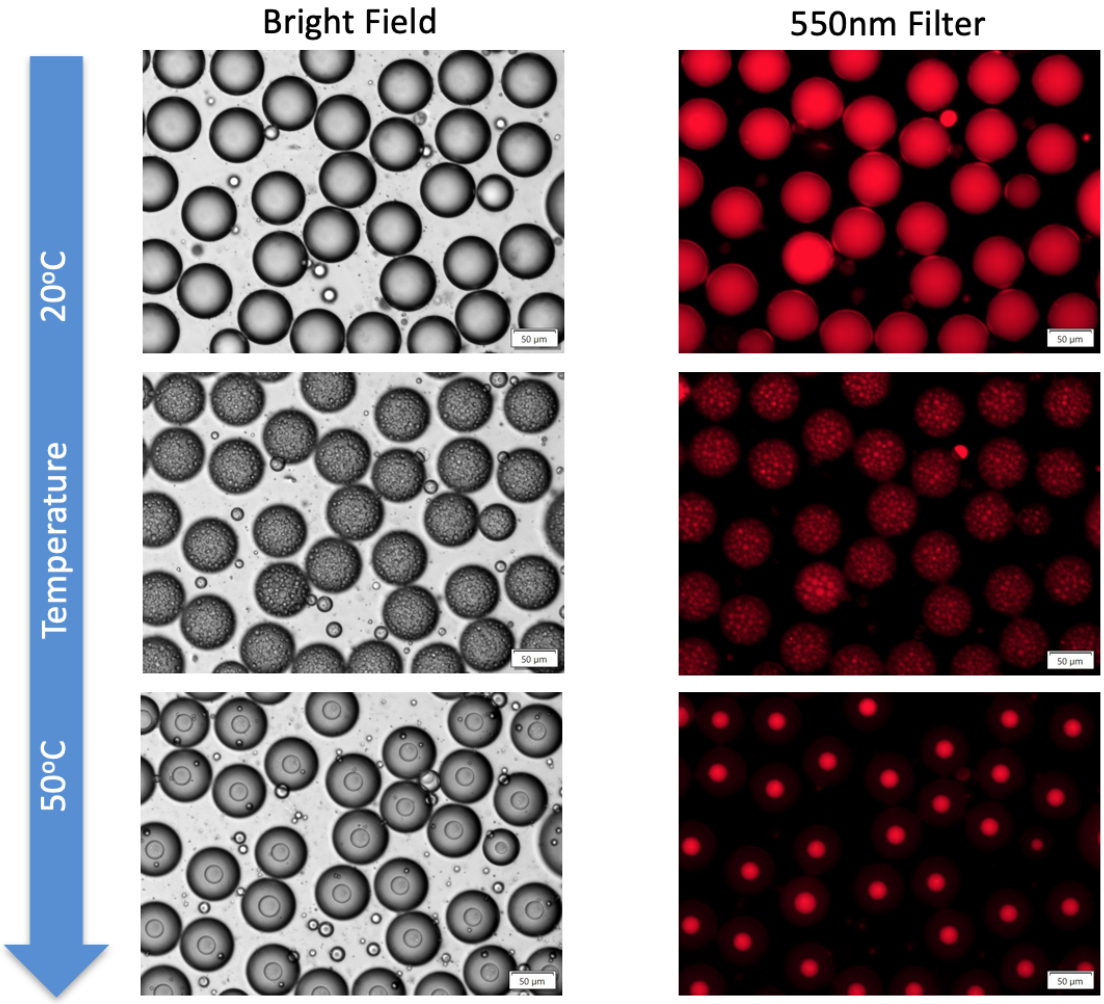

Figure 22. Confocal fluorescence and bright field microscopy images of E3 with CY3 labeled DNA. Formation of aggregates at temperature above Tt[49].

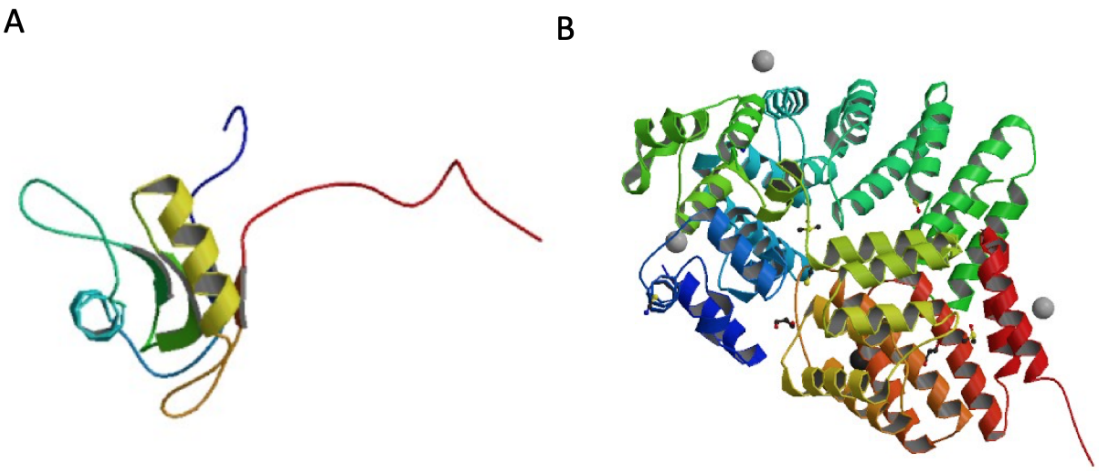

Figure 23. Three-dimensional structures of the fusion elements, A) FUS RRM and B) FUS RGG2

4.7 Acknowledgments

Dr. Peter Davenport, a post-doctoral fellow in the UNM Computer Science department, genetically designed the ELP fusions including the ELP with the Golden Gate cloning landing site. Dr. Davenport instructed me on cloning procedures as well as bacterial proliferation procedures.

Dr. TyAnna Lovato, a research scientist in the UNM Biology Department was also heavily involved in the designing of the ELP fusion fragments. Dr. Lovato also developed the experimental design of the cloning and proliferation of the ELP fusions.

Telmo Diez Perez, a fellow graduate student in Dr. Nick Carroll's lab, collected the data for the E3 ELP. Specifically, Mr. Diez Perez collected the turbidity measurements for E3 and the microscopy images of E3 aggregates.

5. Conclusion

The genetic design, synthesis and potential utility of phase separating ELP fusions that bind NAs is reported herein, working towards programmable gene regulation via environmentally responsive ribonucleoprotein assemblies. By incorporating nucleic acid binding motifs, derived from cell biology, to the C terminus end of a high-yield ELP using Golden Gate cloning, this work has demonstrated the feasibility of creating fusion proteins comprised of synergistic IDP and NA-binding peptide blocks. This work shows that the addition of RNA binding motifs alters the phase diagram of an ELP, but the NA-binding fusion maintains LCST behavior to enable (i) purification by ITC, (ii) substantiation of phase behavior via a measured partial phase diagram and, (iii) formation of ribonucleoprotein coacervates, enriched with DNA.

Future work will be built on the observation that salt concentration has a direct effect on the protein-DNA electrostatic interactions of ELPs. To observe the effect salt has on E3-11 fusion constructs aggregation, turbidity experiments should be carried out where E3-11 is suspended in varying concentrations of salt. This will elucidate the effect of salt on the phase behavior of the fusion construct and, more importantly, the contribution to DNA binding through electrostatic interactions.

It will be necessary to map the phase diagram of E3 fusions in the presence of NAs at variable concentrations, and these data could enable extensions of current theoretical approaches to produce phase diagrams describing the ternary systems of IDP, NA, and buffered water solutions. ELPs are gaining popularity as drug delivery vectors due to their ability to be genetically encodable and to undergo phase transition. Using these properties, it has been shown that chimeric fusion proteins comprising of biologically active RNA binding motifs and ELPs can be encoded and synthesized. This would make ELPs a perfect candidate for the delivery of gene therapeutics if appropriate concentration regimes for nucleic acid binding and release can be established. ELP fusions allow for multiple different interactions based on the type of stimulus or type of signaling input. In future work, ELP fusions could regulate the controlled release or sequestration of genetic material, allowing for even greater control over gene regulation in biological systems.

6. References

1. Han, T. W. *et al.* Cell-free formation of RNA granules: Bound RNAs identify features and components of cellular assemblies. *Cell* **149**, 768–779 (2012).
2. Kato, M. *et al.* Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
3. Weber, S. C. & Brangwynne, C. P. Getting RNA and protein in phase. *Cell* **149**, 1188–1191 (2012).
4. Parker, R. & Sheth, U. P Bodies and the Control of mRNA Translation and Degradation. *Mol. Cell* **25**, 635–646 (2007).
5. Bernardi, R. & Pandolfi, P. P. Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat. Rev. Mol. Cell Biol.* **8**, 1006–1016 (2007).
6. An, S., Kumar, R., Sheets, E. D. & Benkovic, S. J. Reversible compartmentalization of de novo purine biosynthetic complexes in living cells. *Chemtracts* **20**, 501–502 (2008).
7. Elbaum-Garfinkle, S. *et al.* The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7189–7194 (2015).
8. Nott, T. J., Craggs, T. D. & Baldwin, A. J. Membraneless organelles can melt nucleic acid duplexes and act as biomolecular filters. *Nat. Chem.* **8**, 569–575 (2016).
9. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).
10. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208 (2005).
11. Kim, P. M., Sboner, A., Xia, Y. & Gerstein, M. The role of disorder in interaction networks: A structural analysis. *Mol. Syst. Biol.* **4**, 179 (2008).
12. Phillips, A. H. & Kriwacki, R. W. Intrinsic protein disorder and protein modifications in the processing of biological signals. *Curr. Opin. Struct. Biol.* **60**, 1–6 (2020).
13. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* (2004)
14. Bah, A. & Forman-Kay, J. D. Modulation of intrinsically disordered protein function by post-translational modifications. *J. Biol. Chem.* **291**, 6696–6705 (2016).
15. Tsafou, K., Tiwari, P. B., Forman-Kay, J. D., Metallo, S. J. & Toretsky, J. A. Targeting Intrinsically Disordered Transcription Factors: Changing the Paradigm. *J. Mol. Biol.* **430**, 2321–2341 (2018).
16. JianJiong Gao* and Dong Xu. Correlation Between posttranslational Modification and intrinsic Disorder in Protein. *Biocomputing* 94–103 (2012).
17. Pejaver, V. *et al.* The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Sci.* **23**, 1077–1093 (2014).
18. Chavali, S., Gunnarsson, A. & Babu, M. M. Intrinsically Disordered Proteins

Adaptively Reorganize Cellular Matter During Stress. *Trends Biochem. Sci.* **42**, 410–412 (2017).
19. Anderson, P. & Kedersha, N. Stress granules. *Curr. Biol.* **19**, 397–398 (2009).
20. Riback, J. a *et al.* Stress-triggered phase separation is an adaptive, evolutionarily tuned response. *Cell* **168**, 1028–1040 (2018).
21. Wallace, E. W. J. *et al.* Reversible, Specific, Active Aggregates of Endogenous Proteins Assemble upon Heat Stress. *Cell* **162**, 1286–1298 (2015).
22. Simon, J. R., Carroll, N. J., Rubinstein, M., Chilkoti, A. & López, G. P. Programming molecular self-assembly of intrinsically disordered proteins containing sequences of low complexity. *Nat. Chem.* **9**, 509–515 (2017).
23. Chilkoti, A. Purification of recombinant proteins by fusion with thermally-responsive polypeptides. (1999) doi:10.1038/15100.
24. Rincon, A. C. *et al.* Biocompatibility of elastin-like polymer poly(VPAVG) microparticles: in vitro and in vivo studies. *J. Biomed. Mater. Res. Part A* **78**, 343–351 (2006).
25. McDaniel, J. R., Radford, D. C. & Chilkoti, A. A unified model for de novo design of elastin-like polypeptides with tunable inverse transition temperatures. *Biomacromolecules* **14**, 2866–2872 (2013).
26. Meyer, D. E. & Chilkoti, A. Genetically Encoded Synthesis of Protein-Based Polymers with Precisely Specified Molecular Weight and Sequence by Recursive Directional Ligation: Examples from the Elastin-like Polypeptide System. (2002)
27. Hassouneh, W., Zhulina, E. B., Chilkoti, A. & Rubinstein, M. Elastin-like Polypeptide Diblock Copolymers Self-Assemble into Weak Micelles. *Macromolecules* **48**, 4183–4195 (2015).
28. Christensen, T., Hassouneh, W., Trabbic-Carlson, K. & Chilkoti, A. Predicting transition temperatures of elastin-like polypeptide fusion proteins. *Biomacromolecules* **14**, 1514–1519 (2013).
29. Yi, A., Sim, D., Lee, Y. J., Sarangthem, V. & Park, R. W. Development of elastin-like polypeptide for targeted specific gene delivery in vivo. *J. Nanobiotechnology* **18**, 1–14 (2020).
30. McPherson, D. T., Xu, J. & Urry, D. W. Product purification by reversible phase transition following Escherichia coli expression of genes encoding up to 251 repeats of the elastomeric pentapeptide GVGVP. *Protein Expr. Purif.* **7**, 51–57 (1996).
31. Kanekura, K. *et al.* Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. *Hum. Mol. Genet.* **25**, 1803–1813 (2016).
32. Feric, M. *et al.* Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165**, 1686–1697 (2016).
33. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol. Cell* **60**, 208–219 (2015).
34. Kedersha, N., Ivanov, P. & Anderson, P. Stress granules and cell signaling: More than just a passing phase? *Trends Biochem. Sci.* **38**, 494–506 (2013).
35. Protter, D. S. W. & Parker, R. Principles and Properties of Stress Granules. *Trends*

Cell Biol. **26**, 668–679 (2016).
36. Gonzalez, M. A. *et al.* Strong, Tough, Stretchable, and Self-Adhesive Hydrogels from Intrinsically Unstructured Proteins. *Adv. Mater.* **29**, 1–8 (2017).
37. McCarthy, B., Yuan, Y. & Koria, P. Elastin-like-polypeptide based fusion proteins for osteogenic factor delivery in bone healing. *Biotechnol. Prog.* **32**, 1029–1037 (2016).
38. Li, L., Mo, C.-K., Chilkoti, A., Lopez, G. P. & Carroll, N. J. Creating cellular patterns using genetically engineered, gold- and cell-binding polypeptides. *Biointerphases* **11**, 021009 (2016).
39. Macewan, S. R. & Chilkoti, A. Applications of elastin-like polypeptides in drug delivery. *J. Control. Release* **190**, 314–330 (2014).
40. McDaniel, J. R., MacKay, J. A., Quiroz, F. G. & Chilkoti, A. Recursive directional ligation by plasmid reconstruction allows rapid and seamless cloning of oligomeric genes. *Biomacromolecules* **11**, 944–952 (2010).
41. Simon, J. R., Eghtesadi, S. A., Dzuricky, M., You, L. & Chilkoti, A. Engineered Ribonucleoprotein Granules Inhibit Translation in Protocells. *Mol. Cell* **75**, 66-75.e5 (2019).
42. Gall, J. G., Bellini, M., Wu, Z. & Murphy, C. Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes. *Mol. Biol. Cell* **10**, 4385–4402 (1999).
43. Brangwynne, C. P., Tompa, P. & Pappu, R. V. Polymer physics of intracellular phase transitions. *Nat. Phys.* **11**, 899–904 (2015).
44. Vo, K. & Shim, H. Spectrophotometry. *Chem. Libr.* (2020).
45. Regnima, G.-O. *et al.* Quantitative measurements of turbid liquids via structured laser illumination planar imaging where absorption spectrophotometry fails. *Appl. Opt.* **56**, 3929 (2017).
46. Adamson, N. J. & Reynolds, E. C. Rules relating electrophoretic mobility, charge and molecular size of peptides and proteins. *J. Chromatogr. B. Biomed. Sci. Appl.* **699**, 133–147 (1997).
47. Manns, J. M. SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) of Proteins. *Curr. Proticols* **22**, A.3M.1-A.3M.13 (2011).
48. Engler, C. & Marillonnet, S. Combinatorial DNA assembly using golden gate cloning. *Methods Mol. Biol.* **1073**, 141–156 (2013).
49. Diez Perez, T., López, G. P. & Carroll, N. J. DNA-binding by an elastin-like polypeptide for assembly of deoxyribonucleoprotein coacervates. *Publ. Prog.* (2020).
50. Roberts, S. *et al.* Injectable tissue integrating networks from recombinant polypeptides with tunable order. *Nat. Mater.* **17**, 1154–1163 (2018).
51. Glassman, M. J., Avery, R. K., Khademhosseini, A. & Olsen, B. D. Toughening of Thermoresponsive Arrested Networks of Elastin- Like Polypeptides To Engineer Cytocompatible Tissue Scaffolds. *Biomacromolecules* **17**, 415–426 (2016).