

# Microphase separation produces interfacial environment within diblock biomolecular condensates

Andrew P Latham<sup>1</sup>, Longchen Zhu<sup>2</sup>, Dina A Sharon<sup>1</sup>, Songtao Ye<sup>2,3</sup>, Adam P Willard<sup>1</sup>, Xin Zhang<sup>2\*</sup>, Bin Zhang<sup>1\*</sup>

<sup>1</sup>Department of Chemistry, Massachusetts Institute of Technology, Cambridge, United States; <sup>2</sup>Department of Chemistry, School of Science and Research Center for Industries of the Future, Westlake University, Hangzhou, China; <sup>3</sup>Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, China

## eLife assessment

This **important** study investigates the structural organization of a series of diblock elastin-like polypeptide condensates. The methodology is highly **compelling** as it combines multiscale simulations and fluorescence lifetime imaging microscopy experiments. The results increase our understanding of model biomolecular condensates.

### \*For correspondence:

zhangxin@westlake.edu.cn (XZ);  
binz@mit.edu (BZ)

**Competing interest:** See page 13

**Funding:** See page 13

**Preprint posted**

02 April 2023

**Sent for Review**

24 July 2023

**Reviewed preprint posted**

21 November 2023

**Reviewed preprint revised**

10 April 2024

**Reviewed preprint revised**

07 May 2024

**Version of Record published**

26 March 2025

**Reviewing Editor:** Rosana

Collepardo, University of  
Cambridge, United Kingdom

© Copyright Latham et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

**Abstract** The phase separation of intrinsically disordered proteins is emerging as an important mechanism for cellular organization. However, efforts to connect protein sequences to the physical properties of condensates, that is, the molecular grammar, are hampered by a lack of effective approaches for probing high-resolution structural details. Using a combination of multiscale simulations and fluorescence lifetime imaging microscopy experiments, we systematically explored a series of systems consisting of diblock elastin-like polypeptides (ELPs). The simulations succeeded in reproducing the variation of condensate stability upon amino acid substitution and revealed different microenvironments within a single condensate, which we verified with environmentally sensitive fluorophores. The interspersed hydrophilic and hydrophobic residues and a lack of secondary structure formation result in an interfacial environment, which explains both the strong correlation between ELP condensate stability and interfacial hydrophobicity scales, as well as the prevalence of protein-water hydrogen bonds. Our study uncovers new mechanisms for condensate stability and organization that may be broadly applicable.

## Introduction

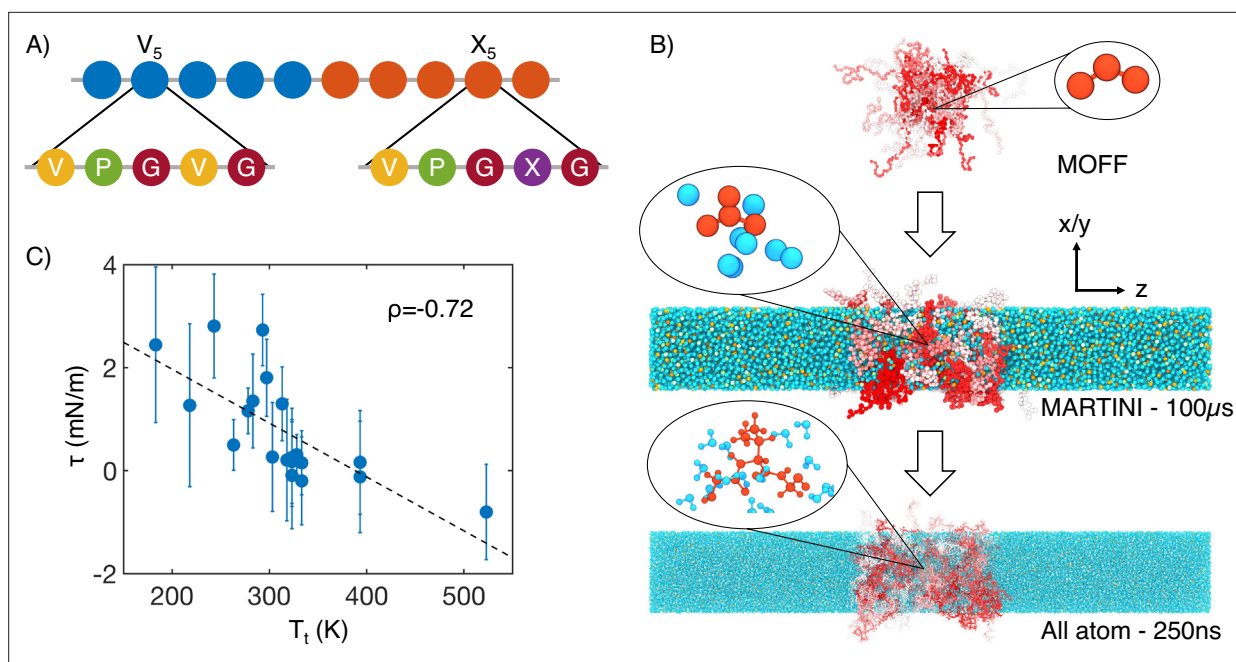
Biological condensates are found in both the cytosol (Brangwynne et al., 2009; Folkmann et al., 2021; Patel et al., 2015; Ma and Mayr, 2018) and nucleus (Feric et al., 2016; Larson et al., 2017; Strom et al., 2017; Sabari et al., 2018; Leicher et al., 2022; Latham and Zhang, 2022b), playing essential roles in a variety of cellular processes (Banani et al., 2017; Uversky, 2017) from stress response (Riback et al., 2017) to genome organization (Lin et al., 2021; Sabari et al., 2020). Similar to membrane-bound organelles, they assemble a collection of molecules to raise the efficiency of sophisticated tasks. The lack of a membrane barrier allows fast material exchange between condensates and the cellular environment, rendering the molecular composition and stability of condensates more prone to regulations by external signals (Hnisz et al., 2017; Riback et al., 2020; Klein et al., 2020).

Intrinsically disordered proteins (IDPs) that promote multivalent, promiscuous interactions are key drivers of condensate formation (Hyman et al., 2014; Boeynaems et al., 2018; Abyzov et al., 2022). Multiple mechanisms, including electrostatic, cation- $\pi$ ,  $\pi$ - $\pi$ , hydrogen bonding, and hydrophobic interactions, contribute to the affinity among various chemical groups (Dignon et al., 2020; Das et al., 2020; Murthy et al., 2021). Above a threshold concentration, as predicted by the Flory-Huggins theory (Flory, 1942), interactions among IDPs can drive liquid-liquid phase separation to produce a highly concentrated phase that nevertheless remains dynamic. The simplicity of theory, while insightful, may prove insufficient for a comprehensive understanding of biological condensates.

Much remains to be learned regarding the connection between amino acid sequences and protein phase behaviors, or the so-called ‘molecular grammar’ of protein condensates (Schuster et al., 2021; Choi et al., 2020; Wang et al., 2018; Schneider et al., 2019; Elbaum-Garfinkle et al., 2015; Kilgore and Young, 2022). These systems often exhibit complex viscoelastic behaviors and substructures with a layered organization (Latham and Zhang, 2022a; Mittag and Pappu, 2022; Kar et al., 2022; Wu and King, 2023), defying the mean-field assumption in the Flory-Huggins theory. Several more advanced theories have been introduced to better account for long-lived structural features that might persist in the polymer network. The sticker and spacer model (Tanaka, 1989; Semenov and Rubinstein, 1998) has been adopted by Pappu and coworkers to account for strong and specific interactions that might form physical cross-links among protein molecules (Harmon et al., 2017). In the meantime, the block copolymer theory may explain the microphase separation that can lead to layered structures (Leibler, 1980; Bates and Fredrickson, 1990; Matsen and Schick, 1994; Matsen and Bates, 1996; Grason, 2006; Swann and Topham, 2010; Shi, 2021). Because of their inherent assumptions, different theories are more appropriate for some systems but not others. High-resolution structural characterizations of condensates could offer further insight into their organizational principles and the applicable theories.

Further decoding the condensate grammar could also benefit from studies of simpler systems. Natural proteins often utilize highly complex amino acid sequences, rendering the attribution of an individual residue’s contribution to collective phenomena challenging. On the other hand, elastin-like polypeptides (ELPs), which are composed of pentapeptide repeats of valine-proline-glycine-X-glycine, serve as excellent models for studying protein phase behaviors. ELPs undergo phase separation upon heating with a lower critical solution temperature (LCST) (Urry, 1997; Rauscher and Pomès, 2017; Baul et al., 2020; Cinar et al., 2019; Dignon et al., 2019). Importantly, the guest residue X can be modified to any amino acid except proline with modern engineering approaches (McDaniel et al., 2010), enabling systematic characterization of the impact on condensate stability and organization upon introducing specific residues.

We combine multiscale simulations with fluorescence lifetime imaging microscopy (FLIM) to probe condensates formed by diblock ELPs and investigate the contribution of amino acid composition to condensate stability. The simulation approach allows the sampling of large-scale conformational rearrangements while providing atomistic resolution to quantify the solvation environment of individual residues. It succeeds in reproducing the stability of condensates and reveals the formation of tubular network phases similar to gyroids. Such structures deviate from weak micelles that have been proposed for diblock ELPs in solution and result in heterogeneous microenvironments within the condensate, which we verify in vitro with FLIM. In addition, we find that the condensate stability exhibits a striking correlation with a hydrophobicity scale derived from interfacial transfer free energy, supporting an interfacial microenvironment of the condensate interior. The chemical specificity of the microenvironment is dictated by the peptide sequence that prevents a complete microphase separation between hydrophobic and hydrophilic groups. As a result, all condensates remain highly solvated after phase separation, producing water molecules that maintain hydrogen bonding with the exposed peptide backbone that persists even in residues with hydrophobic side chains. Our study takes a significant stride at connecting amino acid sequences with the structural and physical properties of biological condensates.



**Figure 1.** Multiscale simulations enable thermodynamic and structural characterization of elastin-like polypeptide (ELP) condensates. **(A)** Illustration of the sequence for the simulated diblock ELPs that consist of five-amino acid repeats, where X is substituted with a guest amino acid. **(B)** Overview of the three-step multiscale simulation approach that gradually increases the model resolution. Simulations of V5L5 were used to produce the example configurations at each step. Peptides are shaded red-white, while water molecules, chlorine, and sodium are colored in blue, green, and orange, respectively. Inserts of a G-V-G repeat and surrounding water molecules are shown to indicate the resolution of each model. **(C)** Correlation between the simulated surface tension ( $\tau$ ) of 20 ELP condensates and the transition temperatures ( $T_t$ ) of related systems (Urry, 1997).  $\rho$  is the Pearson correlation coefficient between the two data sets, and the dashed line is the best fit between simulation and experimental data. Error bars represent the SD of estimates from five independent time windows.

## Results

### Multiscale simulations of diblock ELP condensates

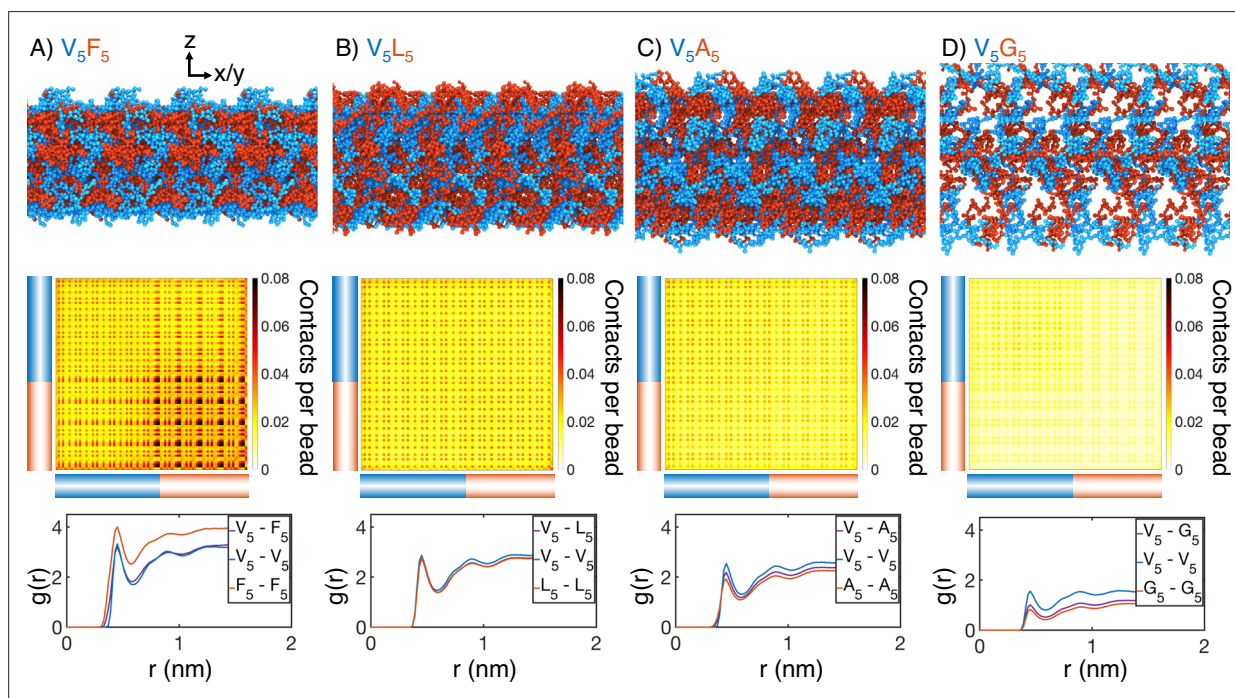
Decoding the condensate grammar necessitates connecting protein sequences with collective physical properties. ELPs stand out because of their sequence simplicity and amenability for biological engineering, allowing systematic exploration and precise attribution of amino acid contributions. We focus on diblock ELPs of sequence  $(V-P-G-V-G)_n(V-P-G-X-G)_n$ , which we abbreviate as  $V_nX_n$  (Figure 1A). The guest residue in the first block is set as valine to promote phase separation, and we explore 20 systems in which every natural amino acid is substituted into the X position in the second block.

We adopt a multiscale approach to balance accuracy and efficiency for simulating ELP condensates (Figure 1B). The first stage of this strategy is to simulate condensate formation using a coarse-grained force field, MOFF (Latham and Zhang, 2021; Latham and Zhang, 2019; Latham and Zhang, 2020; Latham and Zhang, 2022c), with one bead per amino acid and implicit solvation. These simulations are highly efficient for system relaxation and promoting large-scale conformational rearrangements that occur over slow timescales. We further converted these  $\alpha$ -carbon-based models to approximately four heavy atoms per bead resolution and introduced explicit water and ions for simulations with the MARTINI force field (Marrink et al., 2007). The latest version of this force field (MARTINI3) provides a more balanced set of parameters for protein-protein interactions and has been applied to study biological condensates (Souza et al., 2021; Thomasen et al., 2022; Benayad et al., 2021; Tsanai et al., 2021). These simulations further relax the protein configurations and the partition of water and counterions in condensed and dilute phases. Finally, we carried out explicit solvent all-atom simulations for 250 ns to produce an accurate characterization of the chemical environment of the condensates.

Utilizing the simulated condensate conformations, we computed various quantities to benchmark against experimental measurements. While the critical temperature has been widely used as a measure for condensate stability, determining it computationally is expensive. As an alternative, we

computed the surface tension,  $\tau$ , using 100- $\mu$ s-long MARTINI simulations performed with the NP<sub>N</sub>AT ensemble (Zhang *et al.*, 1995). As detailed in the ‘Supplemental theory’ in Appendix 1, an inverse relationship is expected between  $\tau$  and the critical temperature,  $T_c$ , for systems exhibiting LCSTs. We further approximate  $T_c$  with the transition temperatures ( $T_t$ ) of ELP sequences (Urry, 1997), which are the temperatures at which ELPs undergo an LCST transition at a specified solution condition.  $T_t$  was shown to be linearly proportional to  $T_c$  (McDaniel *et al.*, 2013; Meyer and Chilkoti, 2004). As expected, a negative correlation can be readily seen between computed surface tension and experimental  $T_t$  (Figure 1C). This observed negative correlation between  $T_t$  and  $\tau$  supports the simulation approach’s accuracy in reproducing the sequence-dependent changes in ELP phase behavior.

We note that the experimental values were determined using different ELP sequences from those simulated here, and the relation between the surface tension and  $T_c$  is likely nonlinear (Roe, 1975), both contributing to the imperfect agreement. In another study (Ye *et al.*, 2024), we adopted the same multiscale approach to determine the dielectric constant for several ELP condensates. The simulated values agree well with those determined using FLIM experiments.



**Figure 2.** Internal organization of elastin-like polypeptide (ELP) condensates for (A) V<sub>5</sub>F<sub>5</sub>, (B) V<sub>5</sub>L<sub>5</sub>, (C) V<sub>5</sub>A<sub>5</sub>, and (D) V<sub>5</sub>G<sub>5</sub>. The uppermost panels present representative configurations from MARTINI simulations for each system. Periodic images along the x and y dimensions are shown for clarity, and the condensate-water interface is perpendicular to the z-axis. Only proteins are shown, with the X- and V-substituted halves of the peptides shown in red and blue, respectively. The central panels show contact maps between amino acids from different peptides. The blue and red bars indicate the V- and X-substituted half of the peptides. The lower panels plot the radial distribution functions,  $g(r)$ , for amino acids only from the V-substituted half of the peptides (V<sub>5</sub>-V<sub>5</sub>), only from the X-substituted half of the peptides (X<sub>5</sub>-X<sub>5</sub>), and between the two halves (V<sub>5</sub>-X<sub>5</sub>). We limited the calculations to amino acid pairs from different peptides.

The online version of this article includes the following figure supplement(s) for figure 2:

**Figure supplement 1.** Mass density profiles of elastin-like polypeptide (ELP) condensates from MARTINI simulations.

**Figure supplement 2.** Mass density profiles of elastin-like polypeptide (ELP) condensates from all-atom simulations.

**Figure supplement 3.** Theoretical phase diagram (Matsen and Bates, 1996) and corresponding morphologies for diblock copolymers.

**Figure supplement 4.** Representative configurations of (A) V<sub>5</sub>F<sub>5</sub> and (B) V<sub>5</sub>L<sub>5</sub> condensates from MARTINI simulations.

**Figure supplement 5.** Finite size effect checks of our simulation methodology by (A) doubling the number of chains in the simulation box (V<sub>5</sub>F<sub>5</sub>; 80 chains) or by (B) doubling the length of the peptide sequence (V<sub>10</sub>F<sub>10</sub>).

**Figure supplement 6.** Intermolecular contact maps for elastin-like polypeptide (ELP) peptides from MARTINI simulations.

**Figure supplement 7.** Radial distribution functions  $g(r)$  from MARTINI simulations support the microphase separation of different elastin-like polypeptide (ELP) blocks.

## Microphase separation of ELP condensates

Upon validating the accuracy of the simulation protocol for reproducing collective properties of ELP condensates, we examined their structural organization. Hassouneh et al. argued that, in dilute solutions, ELP diblocks self-assemble into the so-called weak spherical micelles with dense cores and almost unstretched coronas (Hassouneh et al., 2015). At phase separation conditions, whether similar structural features are preserved in condensates remains unclear.

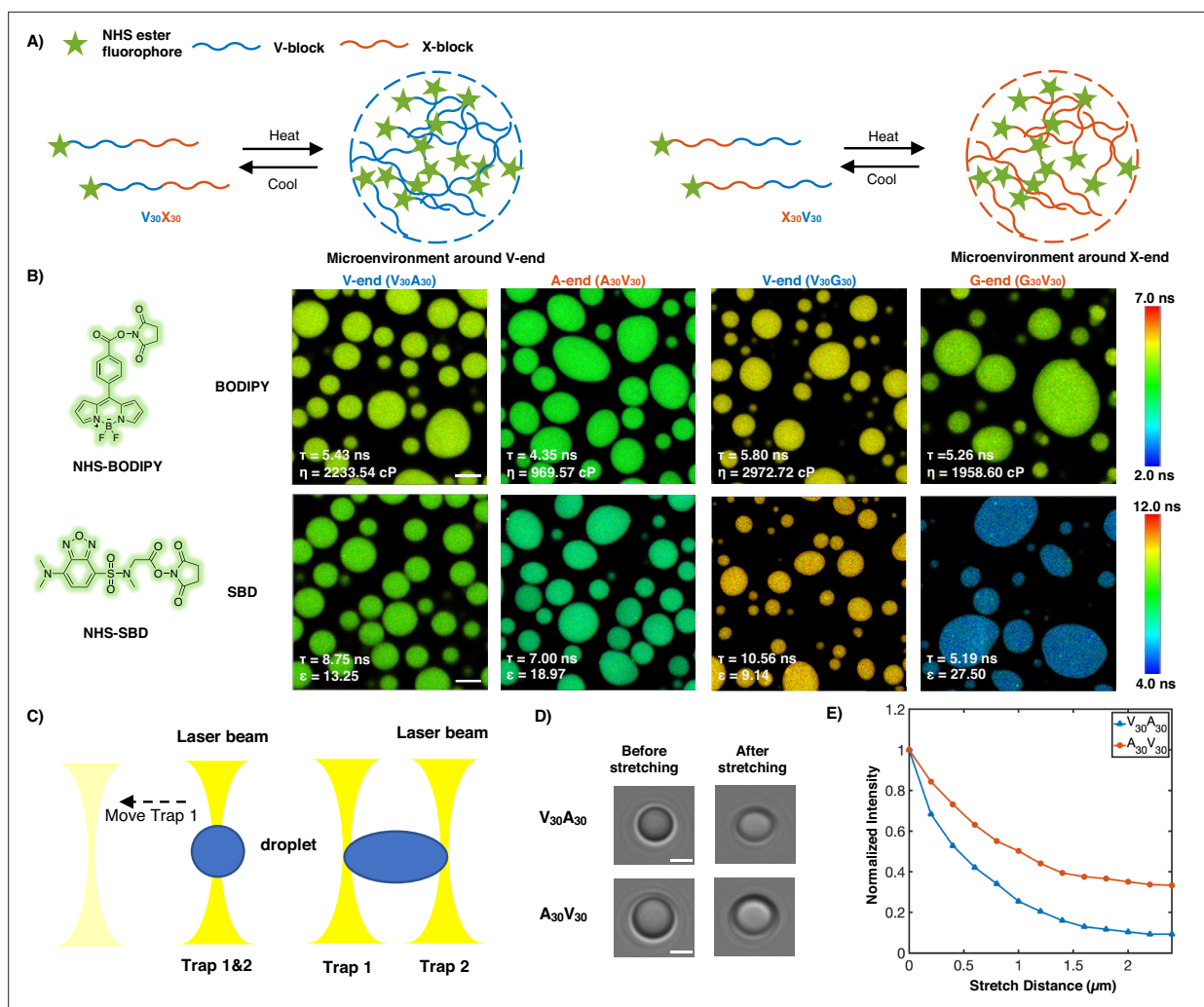
We found that most ELP condensates undergo a microphase separation. Representative configurations for four typical systems from MARTINI simulations are shown in the top panels of **Figure 2**, with the V blocks in blue and X in red. Overall, the microphase separation manages to orient the more hydrophobic blocks toward the condensate interior as opposed to the interface. The relative density of guest blocks near the center of the condensate positively correlates with their hydrophobicity, as can be seen in both MARTINI (**Figure 2—figure supplement 1**) and all-atom (**Figure 2—figure supplement 2**) simulations.

Surprisingly, microphase separation did not produce lamellar morphology as expected for block copolymers with equal volume fraction of the two blocks (**Figure 2—figure supplement 3**; Bates and Fredrickson, 1999; Mai and Eisenberg, 2012; Zhulina et al., 2005; Leibler, 1980; Bates and Fredrickson, 1990; Matsen and Schick, 1994; Matsen and Bates, 1996; Grason, 2006; Swann and Topham, 2010; Shi, 2021). In particular, the condensates appear to form gyroid-like structures, in which the V and X blocks form two interpenetrating networks (**Figure 2—figure supplement 4**). This morphology also differs from micelle-like structures seen in simplified hydrophobic-polar (HP) polymers (Statt et al., 2020; Wessén et al., 2022). It promotes interfacial contacts while maintaining substantial self-interactions as well. Weak interfacial tension between different ELP blocks has also been noted by Hassouneh et al., 2015. These qualitative observations are insensitive to the polymer length and system size in our simulations (**Figure 2—figure supplement 5**). While longer diblock copolymers drive more prominent microphase separation, similar gyroid structures can also be observed. Notably, for more hydrophilic X blocks, the condensates begin to dissolve into a collection of micelle-like structures, consistent with the predictions by Hassouneh et al. in dilute solutions (Hassouneh et al., 2015).

Quantitative characterization of the condensate interior supports the presence of interpenetrating networks with substantial interfacial contacts as well. For example, significant contacts between V and X blocks can be readily seen in the inter-chain contact maps (**Figure 2, Figure 2—figure supplement 6**). We further computed the radial distribution functions for amino acids in various blocks (**Figure 2, Figure 2—figure supplement 7**). For all systems, we found that the probability of finding cross-block contacts V-X is often comparable to the intra-block contacts between X-X or V-V, depending on the relative hydrophobicity of V and X. For example, the more hydrophobic F blocks exhibit strong self-clustering, much more prominent than V-V blocks that are comparable to V-F blocks. On the other hand, clustering among V-V blocks and V-A blocks is more substantial than among A-A blocks. These results are again robust with respect to system setups (**Figure 2—figure supplement 5**).

To experimentally test the microphase separation behavior uncovered in simulations, we studied the micro-physicochemical properties of the V-end and X-end of the peptides. We constructed diblock peptides with the combination of 30 pentameric repeats of V block and X (A or G) block, namely  $V_{30}A_{30}$  and  $V_{30}G_{30}$  ('Experimental sequences' in Appendix 1). The amino-termini of  $V_{30}A_{30}$  and  $V_{30}G_{30}$  sequences were subsequently labeled with environmentally sensitive BODIPY or SBD fluorophores (Liu et al., 2017; Shen et al., 2023), whose lifetime could be measured to quantify the viscosity or polarity of the V-end (**Figure 3A**, left panel) (Chambers et al., 2018). These probes have been reported to be only sensitive to single physicochemical properties (Ye et al., 2024; Shen et al., 2023). To avoid artifacts induced by fluorophore labeling, we usually used ELPs labeled with a low fraction of dyes. We also constructed  $A_{30}V_{30}$  and  $G_{30}V_{30}$  diblock peptides, wherein the viscosity or polarity of the A-end or the G-end could be measured by fluorophores that are attached at the amino-terminus (**Figure 3A**, right panel).

Using FLIM, we found that the lifetime of BODIPY for the V-end (5.43 ns) was longer than that for the A-end (4.35 ns), suggesting that the V-end indeed has a higher microviscosity than the A-end ( $\eta_V = 2233.54$  cp vs  $\eta_A = 969.57$  cp). Accordingly, the lifetime of SBD was longer for the V-end (8.75 ns) than the A-end (7.00 ns), indicating that the micropolarity of the V-end was lower than the A-end ( $\epsilon_V = 13.25$  vs  $\epsilon_A = 18.97$ ). These observations could be largely attributed to the greater extent of dehydration at the V-end due to its higher local peptide density. We further showed that the observed



**Figure 3.** Experimental support of the microphase separation of elastin-like polypeptide (ELP) condensates. **(A)** Reversible ELP condensates formation via changing temperature. NHS ester fluorophores are attached at the amino-termini of  $V_{30}X_{30}$  and  $X_{30}V_{30}$  to detect the different micro-physicochemical properties between the V-end and the X-end. **(B)** Structures of NHS-BODIPY and NHS-SBD, and FLIM images of  $V_{30}A_{30}$ ,  $A_{30}V_{30}$ ,  $V_{30}G_{30}$ , and  $G_{30}V_{30}$  labeled with respective fluorophores. The fluorescence lifetime of each image is the average acquired from three independent experiments. Scale bar: 5  $\mu$ m. **(C)** Schematic diagram of optical tweezers stretching experiment. **(D)** Bright-field images of  $V_{30}A_{30}$  and  $A_{30}V_{30}$  in the stretching experiment using the optical tweezer. Scale bar: 2  $\mu$ m. **(E)** Normalized droplets fluorescence intensity changes while stretching (red curve:  $V_{30}A_{30}$ ; blue curve:  $A_{30}V_{30}$ ). Fifteen droplets (at size  $\sim 4$   $\mu$ m) were imaged and used for statistical analysis.

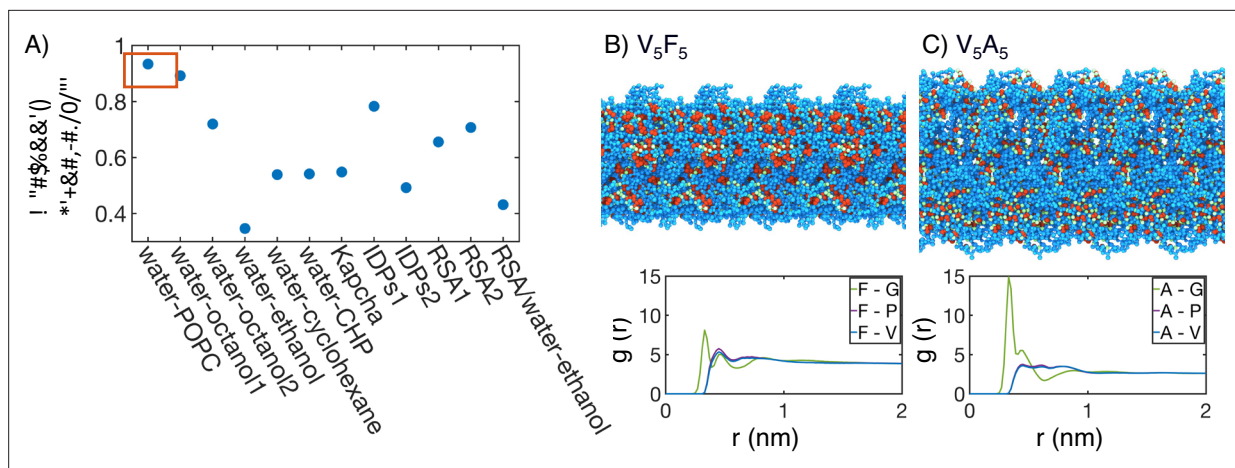
The online version of this article includes the following figure supplement(s) for figure 3:

**Figure supplement 1.** Experimental  $T_f$  of elastin-like polypeptide (ELP) sequences.

**Figure supplement 2.** Fluorescence lifetime imaging microscopy (FLIM) image of unlabeled elastin-like polypeptide (ELP) condensates.

differences are not results of possible artifacts arising from any subtle distinctions between the two sequences  $V_{30}A_{30}$  and  $A_{30}V_{30}$  ('Experimental characterization of ELP condensates' in Appendix 1, **Figure 3—figure supplements 1 and 2**). Similar results were observed using the V-G sequences. FLIM experiments revealed that the V-end was more viscous than the G-end ( $\eta_V = 2972.72$  cp vs  $\eta_G = 1958.60$  cp) and the V-end was less polar than the G-end ( $\epsilon_V = 9.14$  vs  $\epsilon_G = 27.50$ ). These experimental observations provided the first line of evidence to support the microphase separation, as suggested by the simulation results. It is worth noting that the reported values, although related, may not quantitatively represent the steady-state viscosity. This discrepancy arises from the slow relaxation timescale inherent in ELP condensates with viscoelastic properties.

To further investigate these phenomena in real time in motion, optical tweezers were used to stretch the droplet to observe the viscosity changes of different blocks. A single diblock condensate



**Figure 4.** Interior of elastin-like polypeptide (ELP) condensates exhibits interfacial properties as a result of microphase separation. **(A)** Pearson correlation coefficient,  $\rho$ , between the stated measure of hydrophobicity and the condensate transition temperature,  $T_t$ , measured by [Urry, 1997](#). To remove discrepancies in which end of the scale is hydrophobic and which end is hydrophilic, all hydrophobicity scales, including the Urry scale, are first normalized such that 1 corresponds to the most hydrophobic residue and 0 corresponds to the least hydrophobic residue. Hydrophobicity measures considered include experimental measures of water-solvent transfer free energies water-1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine [POPC]-interface ([White, 1996](#)), water-octanol1 ([Wimley et al., 1996](#)), water-octanol2 ([Wilce et al., 1995](#)), water-ethanol ([Zimmerman et al., 1968](#); [Wilce et al., 1995](#)), water-cyclohexane ([Radzicka and Wolfenden, 1988](#)), and water-N-cyclohexyl-2-pyrrolidone [CHP] ([Lawson et al., 1984](#)), atomic level analysis of moieties within amino acids ([Kapcha and Rossky, 2014](#)), computational approaches to estimate hydrophobicity within disordered proteins (IDPs1 [Tesei et al., 2021](#) and IDPs2 [Dannenhoffer-Lafage and Best, 2021](#)), bioinformatics techniques that approximate the burial propensity, or relative solvent accessibility (RSA) of amino acids (RSA1 [Rose et al., 1985](#) and RSA2 [Tien et al., 2013](#)), and a method that mixed protein burial fraction with water-vapor transfer free energy (RSA/water-vapor [Kyte and Doolittle, 1982](#)). The red box highlights the high correlation between  $T_t$  and water-octanol or water-POPC interface transfer free energies. **(B, C)** Condensate organization from MARTINI simulations for  $V_5F_5$  **(B)** and  $V_5A_5$  **(C)**. The top panels provide protein-only views for simulated condensates, with the guest residue X, glycine adjacent to the guest residue, and the remaining residues colored red, green, and blue, respectively. Condensate images are repeated periodically in the x/y plane. The bottom panel shows the overall radial distribution function from the guest amino acid to those amino acids native to the ELP sequence.

The online version of this article includes the following figure supplement(s) for figure 4:

**Figure supplement 1.** Residue-specific radial distribution function  $g(r)$  from MARTINI simulations.

was captured by two optical traps (traps 1 and 2) at the same position, and it was stretched as trap 1 moved away from the original position (**Figure 3C and D**). We used the peptides  $V_{30}A_{30}$  and  $A_{30}V_{30}$ , with BODIPY labeled at their amino-terminus, to study the microviscosity changes where the fluorescence intensity decreases as the viscosity decrease. We found that the normalized fluorescence intensity of the V-end decreases greater than the A-end, indicating that the viscosity change of the V-end is larger (**Figure 3E**). We anticipate that during stretching condensates deform and molecules may deviate from the preferred chemical environment established with microphase separation. Such deviations impact the hydrophobic segments more, leading to more significant reduction in their interactions with surrounding molecules and correspondingly their viscosity.

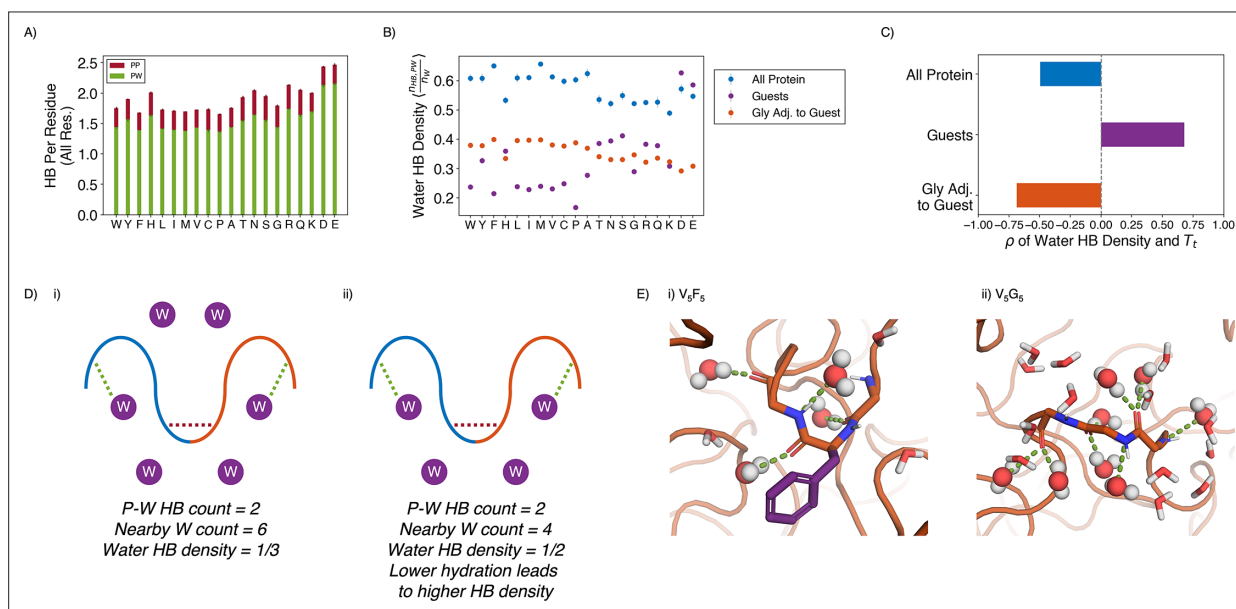
### Frustration produces condensate interior with interfacial properties

The strong dependence of molecular organization on amino acid hydrophobicity suggests that the solvation environment of individual residues might be a determining factor for condensate stability. Indeed, as shown in the ‘Supplemental theory’ of Appendix 1, the critical temperature is closely related to the free energy cost of transferring polymer beads from a solution state to a polymer-only environment. This transfer free energy is often used to quantify the hydrophobicity of amino acids ([Regy et al., 2021](#); [White, 1996](#); [Wimley et al., 1996](#); [Wilce et al., 1995](#); [Zimmerman et al., 1968](#); [Radzicka and Wolfenden, 1988](#); [Lawson et al., 1984](#)). To explore their relationship more quantitatively, we compared the transition temperature for ELP condensates measured by [Urry, 1997](#) to several hydrophobicity scales. **Figure 4A** shows that water-octanol and water-POPC interface transfer free energies best correlate with the transition temperature. Meanwhile, other hydrophobicity measures, even including those derived specifically for disordered proteins, do not match as well. We note that our conclusion here is not dependent on the limited number of hydrophobicity measures

examined in this study, and can be seen in a previous clustering of 98 different hydrophobicity scales (Simm *et al.*, 2016).

The correlation analysis of the Urry transition temperature with various hydrophobicity scales supports a unique chemical environment emerging from the collective behaviors of condensates. To better understand the microenvironment surrounding guest residues in X blocks, we computed the radial distribution functions using MARTINI simulations. As shown in **Figure 4B and C**, **Figure 4—figure supplement 1**, the guest residues are immediately surrounded with hydrophilic glycine residues as the nearest neighbors. In the second shell, more hydrophobic valine and proline residues appear. These profiles reveal the interfacial nature of the chemical environment surrounding the guest residues consisting of both hydrophobic and hydrophilic amino acids. This interfacial solvation environment is consistent with the strong correlation of the Urry scale with water-POPC interface transfer free energy. Therefore, ELP condensates do not simply feature a hydrophobic interior that separates from hydrophilic residues as seen for folded proteins, and hence the poor correlation shown in **Figure 4A** with the burial propensity of amino acids (RSA1, RSA2).

The ELP peptide sequences partly dictate the interfacial environment. For example, the guest residues are juxtaposed between two glycine residues in the pentamer (**Figure 1A**), creating the immediate shell of hydrophilic residues. However, the presence of valine in the pentamer prevents the movement of GXG motifs to the interface due to the clustering of hydrophobic residues that prefer the condensate interior. Therefore, while microphase separation occurs in ELP condensates, frustration remains in the system. Hydrophilic residues cannot completely separate from hydrophobic ones due to constraints imposed by the amino acid sequence, creating unique microenvironments. Such



**Figure 5.** Water hydrogen bonding environment of condensates from all-atom simulations. **(A)** Bar chart depicting the average number of protein-water (PW, in green) and protein-protein (PP, in red) hydrogen bond (HB) per residue for each condensate system. The x-axis presents the guest residue of the system and is ordered by the Urry scale. Error bars represent SDs of four independent estimates. **(B)** Water HB density for each of the 20 condensates, for three different residue selections. The metric is shown for all residues in the system (blue), guest residues (purple), and glycine residues adjacent to the guest residue in the sequence (orange). The x-axis presents the guest residue of the system and is ordered by the Urry scale. Error bars represent SDs of four independent estimates. **(C)** Correlation coefficients of the water HB density shown in **(B)** with Urry  $T_t$ . **(D)** Schematics depicting two systems with low (i) and high (ii) water hydrogen bond density. The protein is depicted as blue and orange lines, and the water molecules are depicted as purple circles. Protein-protein and protein-water hydrogen bonds are drawn as red and green dashed lines respectively. **(E)** Visualizing hydrogen bond density with illustrative snapshots for  $V_5F_5$  (i) and  $V_5G_5$  (ii) condensate. Side chain carbon atoms are shown in purple. Water molecules near the protein are explicitly depicted, with oxygen and hydrogen atoms colored in red and white. Water molecules hydrogen bonded to the protein are depicted as spheres (with hydrogen bonds shown in green lines).

The online version of this article includes the following figure supplement(s) for figure 5:

**Figure supplement 1.** Secondary structure of simulated protein condensates, splitted into four panels **(A-D)**, with protein names indicated in the legends.

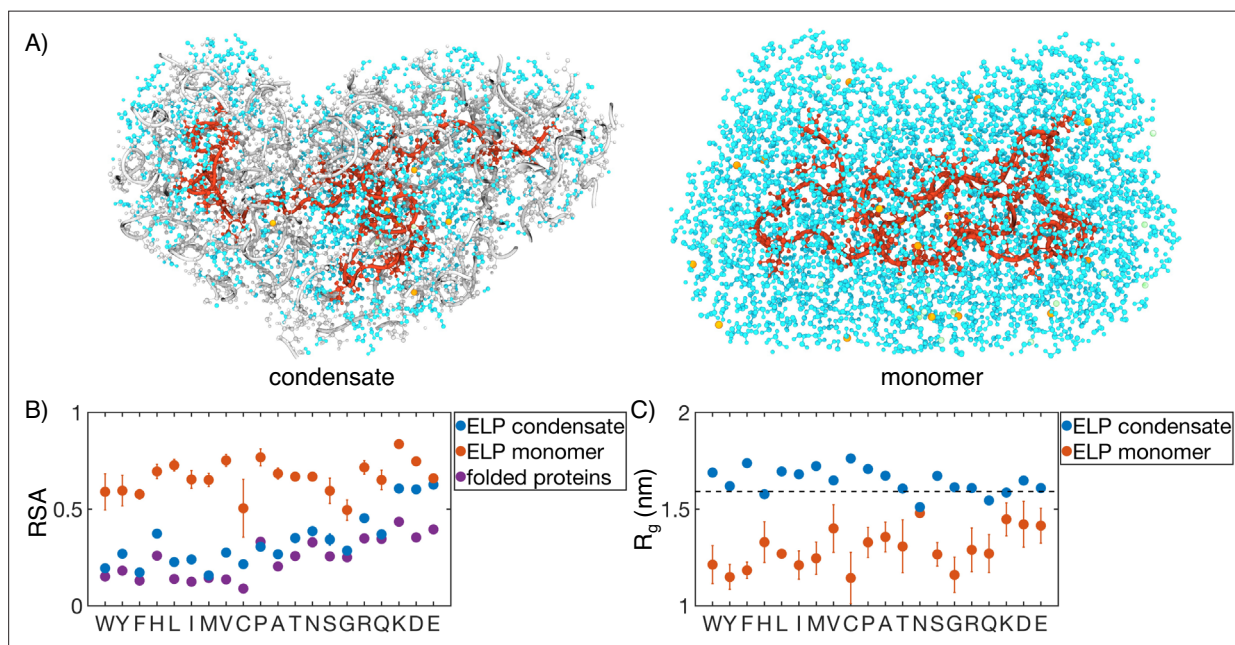
a microenvironment arises from the collective behavior of many proteins and can deviate from individual chains (Wei *et al.*, 2017), explaining the less-than-ideal correlation between the Urry scale and the hydrophobicity scales optimized to reproduce the conformational ensemble of IDP monomers.

## Solvation environment from atomistic simulations support interfacial properties

The interfacial environment uncovered in the previous section suggests the presence of significant solvation in the condensate interior. For condensates with lower surface tension, such as  $V_5G_5$ , the protein density near the center of the condensate is only around ~30%. Even for the stable condensates with high surface tension, such as  $V_5F_5$ , the protein density near the center only reaches ~70%. The relative density of protein, water, and ions in different systems generally follows the trend observed for the transition temperature (Appendix 1—figures 1 and 2).

The large water mass fractions in the condensates prompt the question of the extent to which water is engaged in meeting the protein's hydrogen bonding requirements. We determined the average number of hydrogen bonds per residue with ELP residues or water from extensive atomistic simulations. As shown in Figure 5A, over 75% of the hydrogen bonds are between protein and water in all of the condensates. Prior work on a different ELP condensate (Rauscher and Pomès, 2017) similarly indicated that water-protein hydrogen bonds constitute most of the hydrogen bonds found in the system. This finding highlights that ELPs, lacking noticeable secondary structures (Figure 5—figure supplement 1), are less capable of meeting its own hydrogen bonding needs, compared to folded proteins (Sticke *et al.*, 1992). Therefore, high water concentration within the condensate is crucial to fulfilling the protein's hydrogen bonding requirements.

Strikingly, given the importance of water molecules, the average number of hydrogen bonds per residue remained relatively constant across condensates, even when the amount of water inside the



**Figure 6.** Comparison of protein conformation and solvation between elastin-like polypeptide (ELP) condensates and monomers. **(A)** Zoom-in view of a single peptide chain (red) from atomistic simulations of  $V_5A_5$  condensate (left) and monomer (right). Atoms within 1 nm of the peptide are shown, including water (cyan), chlorine (green), sodium (orange), and protein atoms from other chains (gray). **(B)** Relative solvent accessibility (RSA) of guest residues estimated from atomistic simulations of condensates (blue) and monomers (orange). For comparison, the corresponding values estimated using folded proteins are included as purple dots (Tien *et al.*, 2013). Error bars represent the SD of four independent time estimates. **(C)** Comparison of the radius of gyration ( $R_g$ ) for peptides estimated from atomistic simulations of condensates (blue) and monomers (orange). The dashed line represents the expected  $R_g$  of an ideal chain, utilizing values that have been previously suggested for intrinsically disordered proteins (IDPs) (Hofmann *et al.*, 2012; Dignon *et al.*, 2018a). Error bars represent the SD of four independent time estimates.

The online version of this article includes the following figure supplement(s) for figure 6:

**Figure supplement 1.**  $R_g$  of individual peptide blocks from MARTINI **(A)** and atomistic **(B)** simulations.

condensate varied by almost two folds. Therefore, we hypothesize that each water molecule in more hydrophobic condensates must engage more in hydrogen bonding than each water in hydrophilic condensates. To test this hypothesis, we determined the water hydrogen bond density, defined as the ratio of the number of protein-water hydrogen bonds divided by the total number of water molecules near the condensate (**Figure 5D and E**). As shown in **Figure 5B and C**, the water hydrogen bond density indeed negatively correlates with the condensate transition temperature,  $T_i$ , supporting our hypothesis that more hydrophobic condensates would have a higher hydrogen bond density.

The water hydrogen bond density also highlights an interfacial environment of blended hydrophobic and hydrophilic regions. For example, when we limit the analysis to water molecules near the guest residues, X, the hydrogen bond density now becomes positively correlated with  $T_i$ . Therefore, locally, protein-water interactions are strongly influenced by the chemistry of side chains, and water molecules have fewer hydrogen bonding opportunities with more hydrophobic residues. However, the opposite trend is observed for the hydrogen bond density near glycine residues that immediately surround the guest residues. Therefore, the exposed backbones retain water molecules even when hydrophobic side chains do not provide hydrogen bonding opportunities, and the inseparability between the two further contributes to the observed interfacial property of condensates.

As a complementary measure of the solvation environment of individual residues, we computed the relative solvent accessibility (RSA) for the guest amino acids using atomistic simulations (**Figure 6A and B**). The RSA measures the solvent-accessible surface area (SASA) normalized by the maximum possible solvent exposure. RSA values in ELP condensates are much lower than those computed from all-atom simulations of ELP monomers, consistent with the increase of polymer density upon phase separation. However, they are almost always higher than the values estimated for folded proteins due to many water molecules roaming inside the condensates (**Tien et al., 2013**). Therefore, while ELP condensates provide a mechanism to shield hydrophobic amino acids from solvent, they are less effective in doing so than folded proteins.

While the amino acid RSA values increase upon phase separation, unlike protein folding, the increase does not result from polymer collapse. We observe a significant expansion for almost all the studied peptides upon condensation, supporting the notion that water is a poor solvent for ELP (**Figure 6C**). The expansion is most evident for more hydrophobic ELPs, the radii of gyration of which are now comparable to that of an ideal Gaussian chain. Notably, the sizes of the two separate peptide blocks are similar in most systems, supporting an interfacial environment that can accommodate both hydrophobic and hydrophilic residues (**Figure 6—figure supplement 1**).

## Discussion

We carried out multiscale simulations to elucidate the connection between protein sequences and emergent properties of ELP condensates. The multiscale approach overcomes challenges in sampling slow conformational rearrangements to produce equilibrium atomistic condensate configurations. The simulations successfully reproduced condensate stability variation upon amino acid substitution. While our study is performed at set salt concentration and temperature to isolate the contributions of amino acid hydrophobicity to condensate organization, future studies may consider implementing temperature (**Dignon et al., 2019**) or salt (**Wohl et al., 2021**)-dependent models to explore how solution conditions affect the organization of ELP condensates.

We found that diblock ELPs undergo microphase separation, with the more hydrophobic blocks localizing toward the condensate interior. The more hydrophobic blocks increase the local protein density, resulting in higher viscosity and lower dielectric constants, as determined in FLIM experiments. In contrast to the interior of folded proteins, these blocks are significantly solvated with water occupying over 30% of mass density. Unlike the weak micelles formed in solution (**Hassounah et al., 2015**), ELP condensates exhibit gyroid-like morphologies that promotes contacts between blocks. Further studies on the thermodynamic stability of these morphologies and comparing with predictions from the self-consistent field theory shall provide more insights into the driving forces for their emergence.

Our study uncovered a remarkably simple grammar for determining the stability of ELP condensates. A single quantity specific to the guest residue suffices to predict the emergent properties of the condensate, as evidenced by the strong correlation of the transition temperatures with the transfer free energies. The lack of higher-order effects is striking, supporting a mean field approximation to

account for the contribution of individual residues in the system. However, the hydrophobicity scale is itself sensitive to the local microenvironment that is sequence specific. Such a microenvironment arises from the collective behavior of many proteins, can deviate from that of individual chains, and is likely sensitive to the solution conditions (Welsh et al., 2022), which are held constant in our study. Future work on systems with double amino acid substitutions or changes to salt concentration or temperature could elucidate the generality of the mean field interpretation and the additivity of individual contributions.

While the transferability of the extracted rule for condensate stability to other systems remains to be shown, there are reasons to be hopeful. For instance, we anticipate that the microphase separation that produces the interfacial properties is a general principle for condensate organization. For most IDPs, the hydrophobic and hydrophilic residues are often interspersed throughout the chain to prevent the collapse of individual proteins (Martin et al., 2020; Zheng et al., 2020) and, correspondingly, the complete phase separation as in the lamellar phase. In addition, the hydrophilic backbones are expected to be exposed due to a lack of secondary structure formation, contributing to the interfacial property in the presence of hydrophobic side chains. As such, we anticipate that future research will discover the broad presence of interfacial environments within microphase separated motifs, as observed here.

## Methods

### Molecular dynamics simulation details

To better understand the properties of biological condensates, we performed simulations on 40 diblock ELPs with the sequence (V-P-G-V-G)<sub>n</sub>-(V-P-G-X-G)<sub>n</sub>. We set  $n = 5$  to balance computational efficiency and polymer topological effects. As shown in **Figure 2—figure supplement 5**, the results presented in the main text are qualitatively robust with respect to the system size and peptide length.

The multiscale simulations of ELP condensates started with an implicit solvent model, MOFF (Latham and Zhang, 2021), and a similar homopolymer model, to generate initial configurations for each system. Next, we converted the  $\alpha$ -carbon-only configurations into those consistent with the MARTINI force field (Souza et al., 2021) for coarse-grained, explicit solvent simulations. These simulations measured the coexistence between the condensate and solvent, lasted for 100  $\mu$ s, and were conducted in the NP<sub>N</sub>AT ensemble. Finally, we converted the MARTINI configurations into atomistic structures for simulations with CHARMM36m (Huang et al., 2017) force field, with protein-water interactions shifted to better characterize IDPs. Atomistic simulations were conducted in the NP<sub>N</sub>AT ensemble and lasted 250 ns. All simulations were conducted using the software GROMACS (Berendsen et al., 1995) and are described in detail in ‘Condensate simulation details’ in Appendix 1.

In addition to the condensates, we performed atomistic simulations of ELP monomers. These simulations began from RoseTTAFold predictions of the peptide structure (Baek et al., 2021; Bonneau et al., 2001) and lasted for 1  $\mu$ s in the NPT ensemble. Full details are available in ‘Monomer simulation details’ in Appendix 1.

### Analysis of condensate simulations

#### Surface tension

In explicit solvent simulations, two coexisting phases are present, in which proteins form a dense slab in the x-y plane, and the solvent creates a dilute phase in the z-dimension (Dignon et al., 2018b; Ladd and Woodcock, 1977). We controlled the pressure normal to the protein-solvent interface by performing simulations in the NP<sub>N</sub>AT ensemble (Zhang et al., 1995). The surface tension ( $\tau$ ) was calculated from these simulations as

$$\tau = \frac{L_z}{2N} * (P_{ZZ} - \frac{1}{2}(P_{XX} + P_{YY})), \quad (1)$$

where  $L_z$  is the length of the Z-dimension and  $N$  is the number of protein droplets.  $P_{XX}$ ,  $P_{YY}$ , and  $P_{ZZ}$  are the components of the pressure tensor along the three axes.

#### Clustering

Since the condensate can diffuse along the z-axis, that is, the direction perpendicular to the interface, we aligned the simulated configurations using the position of the largest cluster of peptide molecules.

To find the largest cluster, we first computed a center of mass contact matrix between peptide molecules with a distance cutoff of 4 nm. Then, using a depth first search algorithm (*Tribello et al., 2017*), we identified the size and location of the largest protein cluster. The system was then shifted such that the z coordinate of the center of mass of the largest cluster is at zero. We implemented the alignment using the software MDAnalysis (*Gowers et al., 2016; Michaud-Agrawal et al., 2011*).

### Contact map

Protein contact maps were calculated using a cutoff distance of 0.6 nm. Specifically, two residues were defined as in contact if any pair of their coarse-grained beads were within 0.6 nm.

### Radial distribution function

Radial distribution functions were calculated using the InterRDF function of MDAnalysis between 0.00001 nm and 2 nm. For analyzing the microphase separation of each condensate, we divided each protein into the V-substituted blocks and X-substituted blocks. This radial distribution function excluded intra-chain contacts to better capture how inter-chain contacts determine condensate organization. To analyze the origin of chain frustration within diblock phase separation, we computed the radial distribution function of each guest amino acid to the amino acids native to the sequence. Intra-chain contacts were included to fully capture the local environment experienced by the guest amino acid.

### Relative solvent accessibility

The relative solvent accessibility (RSA) was measured based on the SASA of guest amino acids determined from all-atom simulations. We first calculated the SASA for each residue using GROMACS tools *sasa*, with a probe radius of 0.14 nm (*Eisenhaber et al., 1995*). We then averaged across all repeats of guest amino acids and normalized the SASA by the maximum possible value for a given amino acid to compute RSA (*Tien et al., 2013*; see **Appendix 1—table 1**).

### Radius of gyration

The estimated  $R_g$  from atomistic simulations were compared to the following analytical expression (*Hofmann et al., 2012*):

$$R_g = \sqrt{\frac{2l_p b}{(2\nu + 1)(2\nu + 2)}} N^\nu, \quad (2)$$

where  $N$  is the length of the polymer,  $b = 0.38$  nm is the bond length, and  $l_p = 0.40$  nm is the persistence length.

### Secondary structure

Secondary structure for protein condensates was calculated from atomistic simulations using the GROMACS *do\_dssp* command (*Touw et al., 2015; Kabsch and Sander, 1983*). **Figure 5—figure supplement 1** displays the ordered fraction of secondary structures, which includes  $\alpha$ -helices,  $\beta$ -sheets,  $\beta$ -bridges, and turns.

### Hydrogen bonding

Protein-protein and protein-water hydrogen bonds were determined using the hydrogenbonds package within MDAnalysis (*Gowers et al., 2016; Michaud-Agrawal et al., 2011; Smith et al., 2019*). A hydrogen bond was defined as having a donor heavy atom within 3.30 Å of an acceptor heavy atom and a donor heavy atom-hydrogen atom-acceptor heavy atom angle of 135.0° or larger. The nearby water count for each analysis selection was determined by counting the water molecules whose oxygen atoms were within 4.0 Å of the heavy atoms of this selection. Water hydrogen bond density for each residue selection for each frame was computed as the ratio of protein-water hydrogen bonds to the number of nearby water molecules (**Appendix 1—figure 3**).

## Experimental methods

### Fluorescence lifetime imaging microscopy (FLIM)

The proteins of interest were labeled with NHS ester fluorophore. We used ELPs with 1% BODIPY labels or 2.5% SBD labels to form condensates, which avoid the artifacts induced by fluorophores. Droplets were formed with the final concentration of 70  $\mu$ M ELP in 2 M NaCl for V-A and 1.5 M  $\text{NH}_4\text{SO}_4$  for V-G diblock, respectively. A drop of droplets containing solution was placed on a 0.17 mm coverslip with a 500  $\mu$ m spacer. Images were acquired using Leica Falcon Fluorescence Microscope equipped with Wipac pulse laser and  $\times 63/0.12$  oil-immersion objective. The BODIPY was excited at 488 nm and the SBD was excited at 448 nm. The fluorescence lifetime fitting and image analysis were performed in LAS X and ImageJ.

### Optical tweezer experiments

The protocol describes the preparation and manipulation of a sample containing a BODIPY-labeled ELP protein using an optical tweezer, specifically the C-Trap from LULICKS. The goal is to form a single large condensate from small droplets and then stretch it using the optical tweezers while monitoring its fluorescence intensity. The first step is to mix the BODIPY-labeled ELP protein with NaCl solution to obtain a final concentration of 20  $\mu$ M protein in 2 M NaCl. The sample is then loaded into the flow channel of the optical tweezer. The dual trap mode of the C-Trap is then used to move the two traps to the same position. The sample is flushed until a single large condensate of about 4  $\mu$ m is caught by the dual trap via the fusion of small droplets. The traps are then moved along with the droplet to another flow channel containing 2 M NaCl buffer. The channel and pressure are turned off to avoid any additional force on the droplet. The next step is to stretch the droplet by moving trap 1 0.2  $\mu$ m each time. Fluorescent images of each stretch are taken using a 488 nm excitation laser. The mean fluorescence intensity of the whole droplet is then quantified using ImageJ.

## Acknowledgements

This work was supported by the National Institutes of Health (grant # R35GM133580, BZ), the Research Center for Industries of the Future (RCIF) at Westlake University (XZ), PEW Biomedical Scholars Program 00033066 (XZ), and the National Science Foundation under CHE-1654415 (APW). APL acknowledges support from the National Science Foundation Graduate Research Fellowship Program (grant # 1745302). DAS acknowledges support from the Hertz Foundation and the National Science Foundation Graduate Research Fellowship Program (grant # 2141064).

## Additional information

### Competing interests

Bin Zhang: Reviewing editor, *eLife*. The other authors declare that no competing interests exist.

### Funding

| Funder                                         | Grant reference number | Author          |
|------------------------------------------------|------------------------|-----------------|
| National Institute of General Medical Sciences |                        | Bin Zhang       |
| National Science Foundation                    | CHE-1654415            | Adam P Willard  |
| National Institutes of Health                  | R35GM133580            | Bin Zhang       |
| Pew Charitable Trusts                          | 00033066               | Xin Zhang       |
| National Science Foundation                    | 1745302                | Andrew P Latham |
| Hertz Foundation                               |                        | Dina A Sharon   |

| Funder                      | Grant reference number | Author        |
|-----------------------------|------------------------|---------------|
| National Science Foundation | 2141064                | Dina A Sharon |
| Westlake University         |                        | Xin Zhang     |

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Andrew P Latham, Conceptualization, Formal analysis, Funding acquisition, Validation, Investigation, Visualization, Methodology, Writing – original draft, Writing – review and editing; Longchen Zhu, Dina A Sharon, Songtao Ye, Data curation, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing – original draft, Writing – review and editing; Adam P Willard, Supervision, Funding acquisition, Investigation, Methodology, Writing – original draft, Project administration, Writing – review and editing; Xin Zhang, Bin Zhang, Conceptualization, Formal analysis, Supervision, Funding acquisition, Investigation, Methodology, Writing – original draft, Project administration, Writing – review and editing

### Author ORCIDs

Bin Zhang  <https://orcid.org/0000-0002-3685-7503>

### Peer review material

Reviewer #1 (Public Review): <https://doi.org/10.7554/eLife.90750.4.sa1>

Reviewer #2 (Public Review): <https://doi.org/10.7554/eLife.90750.4.sa2>

Author response <https://doi.org/10.7554/eLife.90750.4.sa3>

## Additional files

### Supplementary files

MDAR checklist

### Data availability

The current manuscript is a computational study, so no data have been generated for this manuscript.

## References

- Abyzov A**, Blackledge M, Zweckstetter M. 2022. Conformational dynamics of intrinsically disordered proteins regulate biomolecular condensate chemistry. *Chemical Reviews* **122**:6719–6748. DOI: <https://doi.org/10.1021/acs.chemrev.1c00774>, PMID: 35179885
- Baek M**, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, Millán C, Park H, Adams C, Glassman CR, DeGiovanni A, Pereira JH, Rodrigues AV, van Dijk AA, Ebrecht AC, Opperman DJ, et al. 2021. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**:871–876. DOI: <https://doi.org/10.1126/science.abj8754>, PMID: 34282049
- Banani SF**, Lee HO, Hyman AA, Rosen MK. 2017. Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews. Molecular Cell Biology* **18**:285–298. DOI: <https://doi.org/10.1038/nrm.2017.7>, PMID: 28225081
- Bates FS**, Fredrickson GH. 1990. Block copolymer thermodynamics: theory and experiment. *Annual Review of Physical Chemistry* **41**:525–557. DOI: <https://doi.org/10.1146/annurev.pc.41.100190.002521>, PMID: 20462355
- Bates FS**, Fredrickson GH. 1999. Block Copolymers—Designer Soft Materials. *Physics Today* **52**:32–38. DOI: <https://doi.org/10.1063/1.882522>
- Baul U**, Bley M, Dzubiella J. 2020. Thermal compaction of disordered and elastin-like polypeptides: a temperature-dependent, sequence-specific coarse-grained simulation model. *Biomacromolecules* **21**:3523–3538. DOI: <https://doi.org/10.1021/acs.biomac.0c00546>
- Benayad Z**, von Bülow S, Stelzl LS, Hummer G. 2021. Simulation of FUS protein condensates with an adapted coarse-grained model. *Journal of Chemical Theory and Computation* **17**:525–537. DOI: <https://doi.org/10.1021/acs.jctc.0c01064>, PMID: 33307683
- Berendsen HJC**, van der Spoel D, van Drunen R. 1995. GROMACS: a message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **91**:43–56. DOI: [https://doi.org/10.1016/0010-4655\(95\)00042-E](https://doi.org/10.1016/0010-4655(95)00042-E)

- Boeynaems S**, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, Schymkowitz J, Shorter J, Wolozin B, Van Den Bosch L, Tompa P, Fuxreiter M. 2018. Protein phase separation: a new phase in cell biology. *Trends in Cell Biology* **28**:420–435. DOI: <https://doi.org/10.1016/j.tcb.2018.02.004>, PMID: 29602697
- Bonneau R**, Tsai J, Ruczinski I, Chivian D, Rohl C, Strauss CEM, Baker D. 2001. Rosetta in CASP4: progress in ab initio protein structure prediction. *Proteins* **5**:119–126. DOI: <https://doi.org/10.1002/prot.1170>, PMID: 11835488
- Brangwynne CP**, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, Jülicher F, Hyman AA. 2009. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**:1729–1732. DOI: <https://doi.org/10.1126/science.1172046>, PMID: 19460965
- Case D**. 2020. Amber. San Francisco: University of California.
- Chambers JE**, Kubánková M, Huber RG, López-Duarte I, Avezov E, Bond PJ, Marciniak SJ, Kuimova MK. 2018. An optical technique for mapping microviscosity dynamics in cellular organelles. *ACS Nano* **12**:4398–4407. DOI: <https://doi.org/10.1021/acsnano.8b00177>
- Choi J-M**, Holehouse AS, Pappu RV. 2020. Physical principles underlying the complex biology of intracellular phase transitions. *Annual Review of Biophysics* **49**:107–133. DOI: <https://doi.org/10.1146/annurev-biophys-121219-081629>, PMID: 32004090
- Cinar H**, Fetahaj Z, Cinar S, Vernon RM, Chan HS, Winter RHA. 2019. Temperature, hydrostatic pressure, and osmolyte effects on liquid-liquid phase separation in protein condensates: physical chemistry and biological implications. *Chemistry* **25**:13049–13069. DOI: <https://doi.org/10.1002/chem.201902210>, PMID: 31237369
- Dannenhoffer-Lafage T**, Best RB. 2021. A data-driven hydrophobicity scale for predicting liquid-liquid phase separation of proteins. *The Journal of Physical Chemistry B* **125**:4046–4056. DOI: <https://doi.org/10.1021/acs.jpcc.0c11479>
- Das S**, Lin Y-H, Vernon RM, Forman-Kay JD, Chan HS. 2020. Comparative roles of charge,  $\pi$ , and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins. *PNAS* **117**:28795–28805. DOI: <https://doi.org/10.1073/pnas.2008122117>, PMID: 33139563
- Dignon GL**, Zheng W, Best RB, Kim YC, Mittal J. 2018a. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *PNAS* **115**:9929–9934. DOI: <https://doi.org/10.1073/pnas.1804177115>, PMID: 30217894
- Dignon GL**, Zheng W, Kim YC, Best RB, Mittal J. 2018b. Sequence determinants of protein phase behavior from a coarse-grained model. *PLOS Computational Biology* **14**:e1005941. DOI: <https://doi.org/10.1371/journal.pcbi.1005941>, PMID: 29364893
- Dignon GL**, Zheng W, Kim YC, Mittal J. 2019. Temperature-controlled liquid-liquid phase separation of disordered proteins. *ACS Central Science* **5**:821–830. DOI: <https://doi.org/10.1021/acscentsci.9b00102>, PMID: 31139718
- Dignon GL**, Best RB, Mittal J. 2020. Biomolecular phase separation: from molecular driving forces to macroscopic properties. *Annual Review of Physical Chemistry* **71**:53–75. DOI: <https://doi.org/10.1146/annurev-physchem-071819-113553>, PMID: 32312191
- Eisenhaber F**, Lijnzaad P, Argos P, Sander C, Scharf M. 1995. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry* **16**:273–284. DOI: <https://doi.org/10.1002/jcc.540160303>
- Elbaum-Garfinkle S**, Kim Y, Szczepaniak K, Chen CC-H, Eckmann CR, Myong S, Brangwynne CP. 2015. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *PNAS* **112**:7189–7194. DOI: <https://doi.org/10.1073/pnas.1504822112>
- Feric M**, Vaidya N, Harmon TS, Mitrea DM, Zhu L, Richardson TM, Kriwacki RW, Pappu RV, Brangwynne CP. 2016. Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **165**:1686–1697. DOI: <https://doi.org/10.1016/j.cell.2016.04.047>, PMID: 27212236
- Flory PJ**. 1942. Thermodynamics of high polymer solutions. *The Journal of Chemical Physics* **10**:51–61. DOI: <https://doi.org/10.1063/1.1723621>
- Folkman AW**, Putnam A, Lee CF, Seydoux G. 2021. Regulation of biomolecular condensates by interfacial protein clusters. *Science* **373**:1218–1224. DOI: <https://doi.org/10.1126/science.abg7071>, PMID: 34516789
- Gowers R**, Linke M, Barnoud J, Reddy T, Melo M, Seyler S, Domański J, Dotson D, Buchoux S, Kenney I, Beckstein O. 2016. MDAnalysis: a python package for the rapid analysis of molecular dynamics simulations. Python in Science Conference Austin, Texas. DOI: <https://doi.org/10.25080/Majora-629e541a-00e>
- Grason G**. 2006. The packing of soft materials: molecular asymmetry, geometric frustration and optimal lattices in block copolymer melts. *Physics Reports* **433**:1–64. DOI: <https://doi.org/10.1016/j.physrep.2006.08.001>
- Harmon TS**, Holehouse AS, Rosen MK, Pappu RV. 2017. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6**:30294. DOI: <https://doi.org/10.7554/eLife.30294>, PMID: 29091028
- Hassouneh W**, Zhulina EB, Chilkoti A, Rubinstein M. 2015. Elastin-like polypeptide diblock copolymers self-assemble into weak micelles. *Macromolecules* **48**:4183–4195. DOI: <https://doi.org/10.1021/acs.macromol.5b00431>, PMID: 27065492
- Helfand E**, Tagami Y. 1972. Theory of the interface between immiscible polymers. II. *The Journal of Chemical Physics* **56**:3592–3601. DOI: <https://doi.org/10.1063/1.1677735>
- Hnisz D**, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. 2017. A phase separation model predicts key features of transcriptional control. *Cell* **169**:13–23. DOI: <https://doi.org/10.1016/j.cell.2017.02.007>

- Hofmann H, Soranno A, Borgia A, Gast K, Nettels D, Schuler B. 2012. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *PNAS* **109**:16155–16160. DOI: <https://doi.org/10.1073/pnas.1207719109>, PMID: 22984159
- Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell AD Jr. 2017. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **14**:71–73. DOI: <https://doi.org/10.1038/nmeth.4067>, PMID: 27819658
- Hyman AA, Weber CA, Jülicher F. 2014. Liquid-liquid phase separation in biology. *Annual Review of Cell and Developmental Biology* **30**:39–58. DOI: <https://doi.org/10.1146/annurev-cellbio-100913-013325>, PMID: 25288112
- Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**:2577–2637. DOI: <https://doi.org/10.1002/bip.360221211>, PMID: 6667333
- Kapcha LH, Rossky PJ. 2014. A simple atomic-level hydrophobicity scale reveals protein interfacial structure. *Journal of Molecular Biology* **426**:484–498. DOI: <https://doi.org/10.1016/j.jmb.2013.09.039>
- Kar M, Dar F, Welsh TJ, Vogel LT, Kühnemuth R, Majumdar A, Krainer G, Franzmann TM, Alberti S, Seidel CAM, Knowles TPJ, Hyman AA, Pappu RV. 2022. Phase-separating RNA-binding proteins form heterogeneous distributions of clusters in subsaturated solutions. *PNAS* **119**:e2202222119. DOI: <https://doi.org/10.1073/pnas.2202222119>, PMID: 35787038
- Kilgore HR, Young RA. 2022. Learning the chemical grammar of biomolecular condensates. *Nature Chemical Biology* **18**:1298–1306. DOI: <https://doi.org/10.1038/s41589-022-01046-y>, PMID: 35761089
- Klein IA, Boija A, Afeyan LK, Hawken SW, Fan M, Dall'Agnese A, Oksuz O, Henninger JE, Shrinivas K, Sabari BR, Sagi I, Clark VE, Platt JM, Kar M, McCall PM, Zamudio AV, Manteiga JC, Coffey EL, Li CH, Hannett NM, et al. 2020. Partitioning of cancer therapeutics in nuclear condensates. *Science* **368**:1386–1392. DOI: <https://doi.org/10.1126/science.aaz4427>, PMID: 32554597
- Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**:105–132. DOI: [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0)
- Ladd AJC, Woodcock LV. 1977. Triple-point coexistence properties of the Lennard-Jones system. *Chemical Physics Letters* **51**:155–159. DOI: [https://doi.org/10.1016/0009-2614\(77\)85375-X](https://doi.org/10.1016/0009-2614(77)85375-X)
- Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, Agard DA, Redding S, Narlikar GJ. 2017. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**:236–240. DOI: <https://doi.org/10.1038/nature22822>, PMID: 28636604
- Latham AP, Zhang B. 2019. Improving coarse-grained protein force fields with small-angle X-ray scattering data. *The Journal of Physical Chemistry B* **123**:1026–1034. DOI: <https://doi.org/10.1021/acs.jpcb.8b10336>
- Latham AP, Zhang B. 2020. Maximum entropy optimized force field for intrinsically disordered proteins. *Journal of Chemical Theory and Computation* **16**:773–781. DOI: <https://doi.org/10.1021/acs.jctc.9b00932>, PMID: 31756104
- Latham AP, Zhang B. 2021. Consistent force field captures homologue-resolved HP1 phase separation. *Journal of Chemical Theory and Computation* **17**:3134–3144. DOI: <https://doi.org/10.1021/acs.jctc.0c01220>, PMID: 33826337
- Latham A, Zhang B. 2022a. Molecular determinants for the layering and coarsening of biological condensates. *Authorea*. DOI: <https://doi.org/10.22541/au.166792086.60577556/v1>
- Latham AP, Zhang B. 2022b. On the stability and layered organization of protein-DNA condensates. *Biophysical Journal* **121**:1727–1737. DOI: <https://doi.org/10.1016/j.bpj.2022.03.028>, PMID: 35364104
- Latham AP, Zhang B. 2022c. Unifying coarse-grained force fields for folded and disordered proteins. *Current Opinion in Structural Biology* **72**:63–70. DOI: <https://doi.org/10.1016/j.sbi.2021.08.006>, PMID: 34536913
- Lawson EQ, Sadler AJ, Harmatz D, Brandau DT, Micanovic R, MacElroy RD, Middaugh CR. 1984. A simple experimental model for hydrophobic interactions in proteins. *The Journal of Biological Chemistry* **259**:2910–2912. DOI: [https://doi.org/10.1016/S0021-9258\(17\)43235-2](https://doi.org/10.1016/S0021-9258(17)43235-2), PMID: 6699000
- Leibler L. 1980. Theory of microphase separation in block copolymers. *Macromolecules* **13**:1602–1617. DOI: <https://doi.org/10.1021/ma60078a047>
- Leicher R, Osunsade A, Chua GNL, Faulkner SC, Latham AP, Watters JW, Nguyen T, Beckwitt EC, Christodoulou-Rubalcava S, Young PG, Zhang B, David Y, Liu S. 2022. Single-stranded nucleic acid binding and coacervation by linker histone H1. *Nature Structural & Molecular Biology* **29**:463–471. DOI: <https://doi.org/10.1038/s41594-022-00760-4>, PMID: 35484234
- Lin X, Qi Y, Latham AP, Zhang B. 2021. Multiscale modeling of genome organization with maximum entropy optimization. *The Journal of Chemical Physics* **155**:010901. DOI: <https://doi.org/10.1063/5.0044150>, PMID: 34241389
- Liu Y, Fares M, Dunham NP, Gao Z, Miao K, Jiang X, Bollinger SS, Boal AK, Zhang X. 2017. AgHalo: a facile fluorogenic sensor to detect drug-induced proteome stress. *Angewandte Chemie International Edition* **56**:8672–8676. DOI: <https://doi.org/10.1002/anie.201702417>
- Ma W, Mayr C. 2018. A membraneless organelle associated with the endoplasmic reticulum enables 3'UTR-Mediated Protein-Protein Interactions. *Cell* **175**:1492–1506. DOI: <https://doi.org/10.1016/j.cell.2018.10.007>, PMID: 30449617
- Mai Y, Eisenberg A. 2012. Self-assembly of block copolymers. *Chemical Society Reviews* **41**:5969–5985. DOI: <https://doi.org/10.1039/c2cs35115c>, PMID: 22776960

- Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH. 2007. The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B* **111**:7812–7824. DOI: <https://doi.org/10.1021/jp071097f>
- Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, Grace CR, Soranno A, Pappu RV, Mittag T. 2020. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**:694–699. DOI: <https://doi.org/10.1126/science.aaw8653>, PMID: 32029630
- Matsen MW, Schick M. 1994. Stable and unstable phases of a diblock copolymer melt. *Physical Review Letters* **72**:2660–2663. DOI: <https://doi.org/10.1103/PhysRevLett.72.2660>, PMID: 10055940
- Matsen MW, Bates FS. 1996. Unifying weak- and strong-segregation block copolymer theories. *Macromolecules* **29**:1091–1098. DOI: <https://doi.org/10.1021/ma951138i>
- Mazur J, Jernigan RL. 1991. Distance-dependent dielectric constants and their application to double-helical DNA. *Biopolymers* **31**:1615–1629. DOI: <https://doi.org/10.1002/bip.360311316>, PMID: 1814508
- McDaniel JR, Mackay JA, Quiroz FG, Chilkoti A. 2010. Recursive directional ligation by plasmid reconstruction allows rapid and seamless cloning of oligomeric genes. *Biomacromolecules* **11**:944–952. DOI: <https://doi.org/10.1021/bm901387t>, PMID: 20184309
- McDaniel JR, Radford DC, Chilkoti A. 2013. A unified model for de novo design of elastin-like polypeptides with tunable inverse transition temperatures. *Biomacromolecules* **14**:2866–2872. DOI: <https://doi.org/10.1021/bm4007166>, PMID: 23808597
- Mehler EL, Solmajer T. 1991. Electrostatic effects in proteins: comparison of dielectric and charge models. *Protein Engineering, Design and Selection* **4**:903–910. DOI: <https://doi.org/10.1093/protein/4.8.903>
- Meyer DE, Chilkoti A. 1999. Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nature Biotechnology* **17**:1112–1115. DOI: <https://doi.org/10.1038/15100>
- Meyer DE, Chilkoti A. 2004. Quantification of the effects of chain length and concentration on the thermal behavior of elastin-like polypeptides. *Biomacromolecules* **5**:846–851. DOI: <https://doi.org/10.1021/bm034215n>
- Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. 2011. MDAAnalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **32**:2319–2327. DOI: <https://doi.org/10.1002/jcc.21787>, PMID: 21500218
- Mittag T, Pappu RV. 2022. A conceptual framework for understanding phase separation and addressing open questions and challenges. *Molecular Cell* **82**:2201–2214. DOI: <https://doi.org/10.1016/j.molcel.2022.05.018>, PMID: 35675815
- Murthy AC, Tang WS, Jovic N, Janke AM, Seo DH, Perdikari TM, Mittal J, Fawzi NL. 2021. Molecular interactions contributing to FUS SYGQ LC-RGG phase separation and co-partitioning with RNA polymerase II heptads. *Nature Structural & Molecular Biology* **28**:923–935. DOI: <https://doi.org/10.1038/s41594-021-00677-4>, PMID: 34759379
- Patel A, Lee HO, Jawerth L, Maharana S, Jahnel M, Hein MY, Stoyanov S, Mahamid J, Saha S, Franzmann TM, Pozniakovski A, Poser I, Maghelli N, Royer LA, Weigert M, Myers EW, Grill S, Drechsel D, Hyman AA, Alberti S. 2015. A liquid-to-solid phase transition of the ALS protein fus accelerated by disease mutation. *Cell* **162**:1066–1077. DOI: <https://doi.org/10.1016/j.cell.2015.07.047>
- Radzicka A, Wolfenden R. 1988. Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* **27**:1664–1670. DOI: <https://doi.org/10.1021/bi00405a042>
- Rauscher S, Pomès R. 2017. The liquid structure of elastin. *eLife* **6**:e26526. DOI: <https://doi.org/10.7554/eLife.26526>, PMID: 29120326
- Regy RM, Thompson J, Kim YC, Mittal J. 2021. Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins. *Protein Science* **30**:1371–1379. DOI: <https://doi.org/10.1002/pro.4094>, PMID: 33934416
- Reichheld SE, Muiznieks LD, Keeley FW, Sharpe S. 2017. Direct observation of structure and dynamics during phase separation of an elastomeric protein. *PNAS* **114**:E4408–E4415. DOI: <https://doi.org/10.1073/pnas.1701877114>, PMID: 28507126
- Riback JA, Katanski CD, Kear-Scott JL, Pilipenko EV, Rojek AE, Sosnick TR, Drummond DA. 2017. Stress-triggered phase separation is an adaptive, evolutionarily tuned response. *Cell* **168**:1028–1040. DOI: <https://doi.org/10.1016/j.cell.2017.02.027>, PMID: 28283059
- Riback JA, Zhu L, Ferrolino MC, Tolbert M, Mitrea DM, Sanders DW, Wei M-T, Kriwacki RW, Brangwynne CP. 2020. Composition-dependent thermodynamics of intracellular phase separation. *Nature* **581**:209–214. DOI: <https://doi.org/10.1038/s41586-020-2256-2>, PMID: 32405004
- Roe RJ. 1975. Theory of the interface between polymers or polymer solutions. I. Two components system. *The Journal of Chemical Physics* **62**:490–499. DOI: <https://doi.org/10.1063/1.430502>
- Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH. 1985. Hydrophobicity of amino acid residues in globular proteins. *Science* **229**:834–838. DOI: <https://doi.org/10.1126/science.4023714>, PMID: 4023714
- Rubinstein M, Colby RH. 2003. *Polymer Physics*. Oxford University Press. DOI: <https://doi.org/10.1093/oso/9780198520597.001.0001>
- Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, Abraham BJ, Hannett NM, Zamudio AV, Manteiga JC, Li CH, Guo YE, Day DS, Schuijers J, Vasile E, Malik S, Hnisz D, Lee TI, Cisse II, Roeder RG, et al. 2018. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**:eaar3958. DOI: <https://doi.org/10.1126/science.aar3958>, PMID: 29930091
- Sabari BR, Dall'Agnese A, Young RA. 2020. Biomolecular condensates in the nucleus. *Trends in Biochemical Sciences* **45**:961–977. DOI: <https://doi.org/10.1016/j.tibs.2020.06.007>, PMID: 32684431

- Schauperl M, Podewitz M, Waldner BJ, Liedl KR. 2016. Enthalpic and entropic contributions to hydrophobicity. *Journal of Chemical Theory and Computation* **12**:4600–4610. DOI: <https://doi.org/10.1021/acs.jctc.6b00422>, PMID: 27442443
- Schneider R, Blackledge M, Jensen MR. 2019. Elucidating binding mechanisms and dynamics of intrinsically disordered protein complexes using NMR spectroscopy. *Current Opinion in Structural Biology* **54**:10–18. DOI: <https://doi.org/10.1016/j.sbi.2018.09.007>
- Schuster BS, Regy RM, Dolan EM, Kanchi Ranganath A. 2021. Biomolecular condensates: sequence determinants of phase separation. *Microstructural Organization, Enzymatic Activity, and Material Properties*. *J. Phys. Chem. B* **125**:3441–3451. DOI: <https://doi.org/10.1021/acs.jpcc.0c11606>
- Segur JB, Oberstar HE. 1951. Viscosity of glycerol and its aqueous solutions. *Industrial & Engineering Chemistry* **43**:2117–2120. DOI: <https://doi.org/10.1021/ie50501a040>
- Semenov AN, Rubinstein M. 1998. Thermoreversible gelation in solutions of associative polymers. 1. statics. *Macromolecules* **31**:1373–1385. DOI: <https://doi.org/10.1021/ma970616h>
- Shen B, Jung KH, Ye S, Hoelzel CA, Wolstenholme CH, Huang H, Liu Y, Zhang X. 2023. A dual-functional BODIPY-based molecular rotor probe reveals different viscosity of protein aggregates in live cells. *Aggregate* **4**:2–8. DOI: <https://doi.org/10.1002/agt2.301>
- Shi AC. 2021. Frustration in block copolymer assemblies. *Journal of Physics. Condensed Matter* **33**:253001. DOI: <https://doi.org/10.1088/1361-648X/abf8d0>, PMID: 33862614
- Simm S, Einloft J, Mirus O, Schleiff E. 2016. 50 years of amino acid hydrophobicity scales: revisiting the capacity for peptide classification. *Biological Research* **49**:31. DOI: <https://doi.org/10.1186/s40659-016-0092-5>, PMID: 27378087
- Smith P, Ziolek RM, Gazzarrini E, Owen DM, Lorenz CD. 2019. On the interaction of hyaluronic acid with synovial fluid lipid membranes. *Physical Chemistry Chemical Physics* **21**:9845–9857. DOI: <https://doi.org/10.1039/C9CP01532A>
- Souza PCT, Alessandri R, Barnoud J, Thallmair S, Faustino I, Grünewald F, Patmanidis I, Abdizadeh H, Bruininks BMH, Wassenaar TA, Kroon PC, Melcr J, Nieto V, Corradi V, Khan HM, Domański J, Javanainen M, Martinez-Seara H, Reuter N, Best RB, et al. 2021. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature Methods* **18**:382–388. DOI: <https://doi.org/10.1038/s41592-021-01098-3>
- Statt A, Casademunt H, Brangwynne CP, Panagiotopoulos AZ. 2020. Model for disordered proteins with strongly sequence-dependent liquid phase behavior. *The Journal of Chemical Physics* **152**:152. DOI: <https://doi.org/10.1063/1.5141095>
- Sticke DF, Presta LG, Dill KA, Rose GD. 1992. Hydrogen bonding in globular proteins. *Journal of Molecular Biology* **226**:1143–1159. DOI: [https://doi.org/10.1016/0022-2836\(92\)91058-W](https://doi.org/10.1016/0022-2836(92)91058-W)
- Strom AR, Emelyanov AV, Mir M, Fyodorov DV, Darzacq X, Karpen GH. 2017. Phase separation drives heterochromatin domain formation. *Nature* **547**:241–245. DOI: <https://doi.org/10.1038/nature22989>, PMID: 28636597
- Swann JMG, Topham PD. 2010. Design and application of nanoscale actuators using block-copolymers. *Polymers* **2**:454–469. DOI: <https://doi.org/10.3390/polym2040454>
- Tanaka F. 1989. Theory of thermoreversible gelation. *Macromolecules* **22**:1988–1994. DOI: <https://doi.org/10.1021/ma00194a077>
- Taylor N, Elbaum-Garfinkle S, Vaidya N, Zhang H, Stone HA, Brangwynne CP. 2016. Biophysical characterization of organelle-based RNA/protein liquid phases using microfluidics. *Soft Matter* **12**:9142–9150. DOI: <https://doi.org/10.1039/C6SM01087C>
- Tesei G, Schulze TK, Crehuet R, Lindorff-Larsen K. 2021. Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *PNAS* **118**:e2111696118. DOI: <https://doi.org/10.1073/pnas.2111696118>, PMID: 34716273
- Thomasen FE, Pesce F, Roesgaard MA, Tesei G, Lindorff-Larsen K. 2022. Improving martini 3 for disordered and multidomain proteins. *Journal of Chemical Theory and Computation* **18**:2033–2041. DOI: <https://doi.org/10.1021/acs.jctc.1c01042>
- Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, Wilke CO. 2013. Maximum allowed solvent accessibilities of residues in proteins. *PLOS ONE* **8**:e80635. DOI: <https://doi.org/10.1371/journal.pone.0080635>, PMID: 24278298
- Touw WG, Baakman C, Black J, te Beek TAH, Krieger E, Joosten RP, Vriend G. 2015. A series of PDB-related databanks for everyday needs. *Nucleic Acids Research* **43**:D364–D368. DOI: <https://doi.org/10.1093/nar/gku1028>, PMID: 25352545
- Tribello GA, Giberti F, Sossa GC, Salvalaglio M, Parrinello M. 2017. Analyzing and driving cluster formation in atomistic simulations. *Journal of Chemical Theory and Computation* **13**:1317–1327. DOI: <https://doi.org/10.1021/acs.jctc.6b01073>, PMID: 28121147
- Tsanai M, Frederix PWJM, Schroer CFE, Souza PCT, Marrink SJ. 2021. Coacervate formation studied by explicit solvent coarse-grain molecular dynamics with the martini model. *Chemical Science* **12**:8521–8530. DOI: <https://doi.org/10.1039/d1sc00374g>, PMID: 34221333
- Urry DW. 1997. Physical chemistry of biological free energy transduction as demonstrated by elastic protein-based polymers. *The Journal of Physical Chemistry B* **101**:11007–11028. DOI: <https://doi.org/10.1021/jp972167t>
- Uversky VN. 2017. Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Current Opinion in Structural Biology* **44**:18–30. DOI: <https://doi.org/10.1016/j.sbi.2016.10.015>, PMID: 27838525

- Wang J, Choi J-M, Holehouse AS, Lee HO, Zhang X, Jahnel M, Maharana S, Lemaitre R, Pozniakovsky A, Drechsel D, Poser I, Pappu RV, Alberti S, Hyman AA. 2018. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**:688–699. DOI: <https://doi.org/10.1016/j.cell.2018.06.006>, PMID: 29961577
- Wassenaar TA, Pluhackova K, Böckmann RA, Marrink SJ, Tieleman DP. 2014. Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. *Journal of Chemical Theory and Computation* **10**:676–690. DOI: <https://doi.org/10.1021/ct400617g>
- Wei M-T, Elbaum-Garfinkle S, Holehouse AS, Chen CC-H, Feric M, Arnold CB, Priestley RD, Pappu RV, Brangwynne CP. 2017. Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nature Chemistry* **9**:1118–1125. DOI: <https://doi.org/10.1038/nchem.2803>, PMID: 29064502
- Welsh TJ, Krainer G, Espinosa JR, Joseph JA, Sridhar A, Jahnel M, Arter WE, Saar KL, Alberti S, Collepardo-Guevara R, Knowles TPJ. 2022. Surface electrostatics govern the emulsion stability of biomolecular condensates. *Nano Letters* **22**:612–621. DOI: <https://doi.org/10.1021/acs.nanolett.1c03138>
- Wessén J, Das S, Pal T, Chan HS. 2022. Analytical formulation and field-theoretic simulation of sequence-specific phase separation of protein-like heteropolymers with short- and long-spatial-range interactions. *The Journal of Physical Chemistry. B* **126**:9222–9245. DOI: <https://doi.org/10.1021/acs.jpcc.2c06181>, PMID: 36343363
- White SH. 1996. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Structural & Molecular Biology* **3**:842–848. DOI: <https://doi.org/10.1038/nsb1096-842>
- Wilce MCJ, Aguilar MI, Hearn MTW. 1995. Physicochemical basis of amino acid hydrophobicity scales: evaluation of four new scales of amino acid hydrophobicity coefficients derived from RP-HPLC of peptides. *Analytical Chemistry* **67**:1210–1219. DOI: <https://doi.org/10.1021/ac00103a012>
- Wimley WC, Creamer TP, White SH. 1996. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35**:5109–5124. DOI: <https://doi.org/10.1021/bi9600153>, PMID: 8611495
- Wohl S, Jakubowski M, Zheng W. 2021. Salt-dependent conformational changes of intrinsically disordered proteins. *The Journal of Physical Chemistry Letters* **12**:6684–6691. DOI: <https://doi.org/10.1021/acs.jpclett.1c01607>, PMID: 34259536
- Wu T, King MR. 2023. Single fluorogen imaging reveals spatial inhomogeneities within biomolecular condensates. *bioRxiv*. DOI: <https://doi.org/10.1101/2023.01.26.525727>
- Ye S, Latham AP, Tang Y, Hsiung CH, Chen J, Luo F, Liu Y, Zhang B, Zhang X. 2024. Micropolarity governs the structural organization of biomolecular condensates. *Nature Chemical Biology* **20**:443–451. DOI: <https://doi.org/10.1038/s41589-023-01477-1>, PMID: 37973891
- Zhang Y, Feller SE, Brooks BR, Pastor RW. 1995. Computer simulation of liquid/liquid interfaces. I. Theory and application to octane/water. *The Journal of Chemical Physics* **103**:10252–10266. DOI: <https://doi.org/10.1063/1.469927>
- Zheng W, Dignon G, Brown M, Kim YC, Mittal J. 2020. Hydropathy patterning complements charge patterning to describe conformational preferences of disordered proteins. *The Journal of Physical Chemistry Letters* **11**:3408–3415. DOI: <https://doi.org/10.1021/acs.jpclett.0c00288>, PMID: 32227994
- Zhulina EB, Adam M, LaRue I, Sheiko SS, Rubinstein M. 2005. Diblock copolymer micelles in a dilute solution. *Macromolecules* **38**:5330–5351. DOI: <https://doi.org/10.1021/ma048102n>
- Zimmerman JM, Eliezer N, Simha R. 1968. The characterization of amino acid sequences in proteins by statistical methods. *Journal of Theoretical Biology* **21**:170–201. DOI: [https://doi.org/10.1016/0022-5193\(68\)90069-6](https://doi.org/10.1016/0022-5193(68)90069-6), PMID: 5700434

## Appendix 1

### Condensate simulation details

We employed a multiscale approach to study ELP condensates. Our simulations consisted of four steps with increasing resolution. We start with homopolymer and MOFF (Latham and Zhang, 2021) simulations that help generate initial configurations for the peptides within the condensate. Then we use the results from the explicit solvent coarse-grained model, MARTINI (Souza et al., 2021), and all-atom explicit solvent force field, CHARMM36m (Huang et al., 2017), for further examination and quantitative analysis. Here, we describe the force field and simulation parameters at each step. Simulations were conducted using GROMACS (Berendsen et al., 1995). Homopolymer and MOFF configurations were performed on GROMACS 2018.4, which allowed for tabulated interactions not allowed in current versions of GROMACS. MARTINI and all-atom simulations were performed using GROMACS 2021.

### Homopolymer simulations

We used an identical homopolymer simulation to help generate starting configurations for elastin-like polypeptides. By using an identical configuration, all future simulations begin with the same molar density of protein, despite different sequences of ELPs. Our homopolymer force field was based on the MOFF force field. It assumed random coil secondary structure and non-bonded interactions identical to MOFF's valine-valine interaction. More specifically,

$$U_{\text{hom}}(\mathbf{r}) = U_{\text{bond}} + U_{\text{angle}} + U_{\text{contact}}, \quad (3)$$

where  $U_{\text{bond}}$ ,  $U_{\text{angle}}$ , and  $U_{\text{contact}}$  account for the bonds between neighboring amino acids, angles between three neighboring amino acids, and nonbonded interactions between amino acids that are separated by more than three bonds, respectively.

The bonding potential  $U_{\text{bond}} = \sum_i V_b(r_{i,i+1})$ , where  $i$  is the bead and

$$V_b(r_{i,i+1}) = \frac{k_b}{2}(r_{i,i+1} - r_0)^2. \quad (4)$$

We used  $k_b = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  and  $r_0 = 0.38 \text{ nm}$ . The angular potential  $U_{\text{angle}} = \sum_i V_a(\theta_i)$  with

$$V_a(\theta_i) = \frac{k_a}{2}(\theta_i - \theta_0)^2. \quad (5)$$

$\theta_i$  is the angle formed between the three consecutive beads  $i$ ,  $i + 1$ , and  $i + 2$   $\text{kJ mol}^{-1} \text{ deg}^{-2}$ , and  $\theta_0 = 127^\circ$ . The contact potential,  $U_{\text{contact}} = \sum_{ij} V_{\text{nb}}(r_{ij}, \epsilon_{IJ})$ . The pairwise potential is a combination of excluded volume and contact terms given by

$$V_{\text{nb}}(r_{ij}) = \frac{|\epsilon|\sigma^{12}}{r_{ij}^{12}} + \epsilon C(r_{ij}). \quad (6)$$

$i$  and  $j$  correspond to interacting beads,  $\epsilon = -2.327339 \text{ kJ mol}^{-1}$ , and  $\sigma = 0.586 \text{ nm}$ .

In this force field, we placed 40 polymers in a  $75 \text{ nm} \times 75 \text{ nm} \times 75 \text{ nm}$  simulation box. After steepest decent energy minimization, we performed an NPT simulation for 0.1 s at 150 K and 1 bar, using a Parrinello–Rahman isotropic barostat and time coupling constant of 1 ps. This resulted in a configuration of forty homopolymers compressed into a  $6.3782 \text{ nm} \times 6.3782 \text{ nm} \times 6.3782 \text{ nm}$  simulation box.

### MOFF simulations

From the starting configuration produced by the homopolymer simulation, we wanted to relax the contacts to be representative of those with ELPs before we began our explicit solvent simulations. For this equilibration step, we turned to the MOFF force field. ELPs are known to have minimal secondary structure in both the dense and dilute state (Rauscher and Pomès, 2017; Reichheld et al., 2017), so we assumed the secondary structure to be similar to that of a random coil. Under these simplifications, MOFF can be specified as

$$U_{\text{hom}}(\mathbf{r}) = U_{\text{bond}} + U_{\text{angle}} + U_{\text{electrostatics}} + U_{\text{contact}}, \quad (7)$$

where  $U_{\text{bond}}$  is described by **Equation 4**,  $U_{\text{angle}}$  is described by **Equation 5**,  $U_{\text{electrostatics}}$  accounts for electrostatic interactions between charged amino acids, and  $U_{\text{contact}}$  accounts for non-bonded interactions.

$U_{\text{electrostatics}}$  accounts for electrostatic interactions between charged residues. Based on the Debye–Hückel theory, it can be approximated as

$$U_{\text{electrostatics}} = \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij} \epsilon(r_{ij})} \exp(-r_{ij}/\lambda_D), \quad (8)$$

where  $\epsilon_0$  is the permittivity of free space,  $\lambda_D$  is the Debye screening length,  $r_{ij}$  is the distance between particles  $i$  and  $j$ , and  $q_i$  and  $q_j$  are charges of particles  $i$  and  $j$ . Charges of individual residues can be found in Table S1 of **Latham and Zhang, 2021**. We used a distance-dependent dielectric constant,  $\epsilon(r_{ij})$ , to capture the change in the solvation environment upon protein folding (**Mazur and Jernigan, 1991; Mehler and Solmajer, 1991**). This dielectric takes the form

$$\epsilon(r_{ij}) = A + \frac{B}{1 + ke^{-\lambda B r_{ij}}}, \quad (9)$$

where  $A = -8.5525$ ,  $k = 7.7839$ ,  $\lambda = 0.03627 \text{ nm}^{-1}$ , and  $B = \epsilon_w - A$ , where  $\epsilon_w = 78.4$  is the dielectric constant of water.

The contact potential,  $U_{\text{contact}} = \sum_{ij} V_{\text{nb}}(r_{ij}, \epsilon_{IJ})$ . The pairwise potential is a combination of excluded volume and contact terms given by

$$V_{\text{nb}}(r_{ij}, \epsilon_{IJ}) = \frac{|\epsilon_{IJ}| \sigma_{IJ}^{12}}{r_{ij}^{12}} + \epsilon_{IJ} C(r_{ij}). \quad (10)$$

$I$  and  $J$  correspond to the amino acid type of bead  $i$  and  $j$  is defined using the individual size of each amino acid type, and full parameters for the implementation of this potential can be found in Table S7 and Table S8 of **Latham and Zhang, 2021**.

In this force field, we conducted energy minimization, and then a 0.1  $\mu\text{s}$  NVT simulation, with a timestep of 10 fs. During this simulation, we utilize the GROMACS simulated annealing protocol to linearly raise the temperature from 150 K to 300 K, with a time coupling constant of 100 ps.

## MARTINI simulations

The final implicit solvent, coarse-grained MOFF configuration was used as a starting point for our explicit solvent coarse-grained simulations with the MARTINI force field. To convert our  $\alpha$ -carbon only configuration to a MARTINI representation, we converted the  $\alpha$ -carbon snapshot to an all-atom representation using tleap, which is part of AMBER tools 2020 (**Case, 2020**). Next, we converted the all-atom configuration to a MARTINI3 configuration using martinize2 (**Souza et al., 2021**). We used random coil secondary structure throughout the entire peptide and enforced neutral termini for each peptide chain. We then expanded the Z-dimension of the simulation box to 40 nm and centered the protein in the simulation box. All simulations thus started with box dimensions of 6.3782 nm  $\times$  6.3782 nm  $\times$  40.0000 nm. This change in box dimensions was preparation for slab simulations, which enabled us to probe the physical properties of the condensate (**Dignon et al., 2018b**). In this setup, the high-density protein phase became a dense slab that was periodic in  $x$  and  $y$ , but did not interact with its periodic image in  $z$  due to the size of the box. The lack of periodic interactions in the  $z$  direction allowed the formation of two interfaces between a dense biopolymer phase and the solvent phase. We then solvated the peptide with MARTINI water and 1 M NaCl.

We started with steepest decent energy minimization. Then, we performed NVT equilibration with a timestep of 10 fs for a total simulation time of 5 ns. The protein and solvent were held at 313.15 K by separately coupling them to a velocity-rescaling thermostat with a time coupling constant of 1 ps. We then equilibrated our simulations in an NPT ensemble for 10 ns with a timestep of 20 fs. We separately coupled protein and solvent to a temperature of 313.15 K with a velocity-rescaling thermostat with a time coupling constant of 1 ps. The pressure was fixed using a Berendsen barostat with a reference pressure of 1 bar, compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$ , and a time coupling constant of 5 ps. We then performed our production simulation in an NPT ensemble with semi-isotropic pressure coupling, which is sometimes called an NP<sub>N</sub>AT ensemble (**Zhang et al., 1995**). This technique fixes the pressure normal to the protein surface, which allows us to calculate the surface tension of the protein-solvent interface. Temperature was coupled separately to 313.15 K for the

protein and solvent using a velocity-rescaling thermostat with a time coupling constant of 1 ps. A Parrinello–Rahman semi-isotropic barostat was used to couple pressure at a time coupling constant of 10 ps. A compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$  and pressure of 1 bar was used in the Z-dimension, but the compressibility was set to 0 in the X–Y dimensions. The production simulations lasted for 100  $\mu\text{s}$  and used a timestep of 20 fs. The first 25  $\mu\text{s}$  was discarded for equilibration, and configurations were recorded every 1000 ps for analysis.

Finally, not every  $\alpha$ -carbon, implicit solvent representation from MOFF could be stably converted into a higher resolution, explicit solvent MARTINI representation. If energy minimization was unable to resolve clashes within the MARTINI configuration we converted from the final step in the MOFF simulation, we iteratively moved back one snapshot and repeated the procedure until a stable configuration was found.

We note that many studies of IDPs with MARTINI have scaled the protein-water interactions to improve agreement between simulation and experiment (Thomassen *et al.*, 2022; Benayad *et al.*, 2021). However, in the case of ELPs studied here, we find the surface tensions given by the default MARTINI model are reasonable based on experimentally informed theories of ELP condensates (Hassouneh *et al.*, 2015) and experimental observations of surface tension in other biological condensates (Taylor *et al.*, 2016), which suggest biological condensates may have interfacial tensions as large as  $\sim 1 \text{ mN/m}$ .

## Atomistic simulations

Proteins were described using the CHARMM36m force field (Huang *et al.*, 2017). Proteins were capped using acetylated N-terminus and amidated C-terminus. TIP3P water was used, and the water-hydrogen Lennard–Jones well depth was scaled to  $-0.10 \text{ kcal/mol}$ . This scaling has been shown to better describe IDPs in previous work (Huang *et al.*, 2017).

All-atom simulations were started based on the final timestep of MARTINI simulations. MARTINI configurations were converted to CHARMM36m configurations using the backward script (Hofmann *et al.*, 2012; Wassenaar *et al.*, 2014). This code converts MARTINI configurations into all-atom representations and then ensures configuration stability through a series of short energy minimization and restrained molecular dynamics simulations. The final output for these configurations was utilized as the starting point for all-atom simulations. If the backward script did not result in as stable all-atom representation, we iteratively moved back one snapshot in the MARTINI and repeated the backward procedure until a stable all-atom initial configuration was found. We performed an additional round of energy minimization, and then equilibrated the protein in the NVT ensemble for 100 ps with a timestep of 1 fs. We used a Nose–Hoover thermostat to separately couple the protein and solvent to 313.15 K with a time coupling constant of 1.0 ps. Position restraints were applied with a force constant of  $400 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  for the backbone and  $40 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  for the side chains. We then performed NPT equilibration for 5 ns with a timestep of 2 fs. We used a Nose–Hoover thermostat to separately couple the protein and solvent to 313.15 K with a time coupling constant of 1.0 ps, and coupled the pressure using an isotropic Parrinello–Rahman barostat with a compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$ , a reference pressure of 1 bar, and a time coupling constant of 2 ps. We then performed an additional equilibration in the  $\text{NP}_{\text{N}}\text{AT}$  ensemble by using a semiisotropic barostat for 5 ns with a timestep of 2 fs. We used a Nose–Hoover thermostat to separately couple the protein and solvent to 313.15 K with a time coupling constant of 1.0 ps. We coupled the pressure using an semiisotropic Parrinello–Rahman barostat with a compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$  and a reference pressure of 1 bar in the Z-dimension, but set the compressibility to 0 in the X–Y dimensions. The pressure-time coupling constant was 2 ps. We then performed production runs in the  $\text{NP}_{\text{N}}\text{AT}$  ensemble. Like the previous simulation, we used a Nose–Hoover thermostat to separately couple the protein and solvent to 313.15 K with a time coupling constant of 1.0 ps, and coupled the pressure using an semiisotropic Parrinello–Rahman barostat with a compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$  and a reference pressure of 1 bar in the Z-dimension, but set the compressibility to 0 in the X–Y dimensions. The pressure-time coupling constant was 2 ps. Simulations lasted for 250 ns, and the first 50 ns was discarded for equilibration. We recorded timesteps every 10 ps for analysis.

## Monomer simulation details

In addition to the condensates, we carried out atomistic simulations of ELP monomers for comparison. These simulations were conducted on all 20 sequences and began from RosettaFold predictions of the

peptide structure (Baek *et al.*, 2021; Bonneau *et al.*, 2001). RoseTTAFold was used in all cases except for V<sub>5</sub>M<sub>5</sub>, where not enough homologs were available for a prediction (Baek *et al.*, 2021). In this case, Rosetta ab initio predictions were used instead (Bonneau *et al.*, 2001). The force field was identical to the condensate simulations, and thus used the CHARMM36m force field, with shifted water potentials and capped termini (Huang *et al.*, 2017). Each peptide was placed in a simulation box of 8 nm × 8 nm × 8 nm, which gives a >99% chance of the side of the simulation box being longer than the end-to-end distance of an ideal chain with a Kuhn length of 0.55 nm, as has been suggested for IDPs (Dignon *et al.*, 2018a). Peptides were solvated in water with an ionic strength of 1 M.

After solvating each protein, we conducted energy minimization. Then, we performed 100 ps of equilibration with a timestep of 1 fs in the NVT ensemble. We used a Nose–Hoover thermostat to separately couple the protein and solvent to 313.15 K with a time coupling constant of 1.0 ps. Position restraints were applied with a force constant of 400 kJ mol<sup>−1</sup> nm<sup>−2</sup> for the backbone and 40 kJ mol<sup>−1</sup> nm<sup>−2</sup> for the side chains. Next, we performed NPT equilibration for 5 ns with a timestep of 2 fs. Simulations were conducted at 313.15 K and 1.0 bar. Finally, we conducted a production simulation in the NPT ensemble for 1 μs with a timestep of 2 fs, at 313.15 K and 1.0 bar. We excluded the first 100 ns of this production run for additional equilibration, which left 900 ns for analysis.

## Experimental characterization of ELP condensates

### Plasmid synthesis

Cloning vector pET-28a+ was modified to contain endonuclease recognition sites of BseRI, AclI, and BglI for inserting the designed ELP sequence, similar to McDaniel *et al.*, 2010. Modified pET-28a+ was linearized using BseRI (NEB, China) and dephosphorylated with rSAP (NEB) for ELP assembly. ssDNA encoding for five pentapeptides (5-mer) sequences was ordered through GeneWiz, China. Then the ssDNA chains were annealed and phosphorylated to form double-stranded duplexes with sticky overhangs. Linearized pET-28a+ vectors and duplexes were ligated by the T4 DNA ligase (NEB). After obtaining 5-mer ELP plasmids, we performed the iterative cloning steps to increase the repeats through recursive directional ligation by the plasmid reconstruction method (McDaniel *et al.*, 2010) until achieving the target length of ELP. The sequences of all plasmids created at each step were confirmed by DNA sequencing.

### Protein expression and purification

The targeted ELP plasmids were transformed into *Escherichia coli* BL21 (DE3) cells. Cells were incubated overnight on the LB plate with kanamycin. On the second day, a single colony was inoculated into 5 mL Luria Broth (LB) and serially diluted (1:100) into four different starter flasks (A–D). Starter cultures were shaken for 12–14 hours at 37 °C, and the culture with the OD<sub>600</sub> 0.6–0.8 was inoculated into 1 L Terrific Broth (TB) containing 50 μg/mL kanamycin at 1:100 dilution. This culture was grown at 37 °C and 220 rpm until the media OD<sub>600</sub> reached 0.8, and the culture was then induced to express the protein with 0.5 mM isopropyl-β-D-thiogalactoside (IPTG). The cultures were then kept shaking for 20–24 hours at 37°C and 200 rpm before harvesting. The cells were collected by centrifuging at 6000 rpm at 4 °C for 10 min and resuspended in 1× PBS buffer (pH 7.4). The proteins were released from the cells by sonicating on ice for 6 min, with 2 s of pulsing followed by 8 s of resting on ice. The lysate was centrifuge at 14,000 rpm at 4°C for 40 min. The supernatant was then collected and mixed with polyethyleneimine to a final concentration of 0.5 w/v% to precipitate nucleic acids. The mixture was then centrifuged at 14,000 rpm and 4°C for 40 min. The supernatant was collected and added with NaCl(s) to get the final concentration of 4.5 M NaCl. The mixture was then shaken at 37°C to precipitate ELP and centrifuged at 10,000 rpm at 37°C for 40 min. The pellet was saved and resuspended in double-distilled H<sub>2</sub>O. This mixture was then shaken on ice to dissolve the ELP. The inverse transition cycling was repeated 1–2 more times (Meyer and Chilkoti, 1999). The protein solution was dialyzed in a 10 kDa membrane (SnakeSkin, Thermo Fischer Scientific) against 2 L double-distilled H<sub>2</sub>O at 4°C, changing the dialysis water one time. The protein purity was determined by 12% SDS-PAGE gel electrophoresis and stained with 0.5 M copper chloride. ELP proteins were then concentrated and lyophilized to remove the water.

### Sequence dependence of microviscosity and polarity

All ELP sequences were expressed in bacteria *E. coli* and purified to more than 95% purity, as analyzed by SDS-PAGE (Appendix 1—figure 4). The purified ELPs were subjected to mass spectroscopy measurement to confirm their identity (Appendix 1—table 2 and Appendix 1—figure 5). Using turbidity as a signal to report on the formation of liquid droplets as a function of temperature change,

we found that  $V_{30}A_{30}$  and  $A_{30}V_{30}$  exhibited the same  $T_i$  (transition temperature for phase separation) values (**Figure 3—figure supplement 1A and B**), suggesting that both sequences underwent almost identical intermolecular interactions to transit from the one-liquid diffusive phase to the two-liquid phase separated phase. The same is true for  $V_{30}G_{30}$  and  $G_{30}V_{30}$  (**Figure 3—figure supplement 1C and D**). To measure the viscosity and polarity of different micro-organizations of blocks, we used the established microenvironment-sensitive fluorogenic probes: boron dipyrromethene (BODIPY) and 7-sulfonamide benzoxadiazole (SBD). Based on previous reports, the fluorescence lifetime ( $\tau$ ) of BODIPY would increase with elevating microviscosity, whereas the  $\tau$  of SBD would rise with decreasing micro-polarity. The correlation of their  $\tau$  values can be quantitatively determined using standard solvents, thus providing equations to quantify viscosity ( $\eta$ ) and polarity (as  $\epsilon$ , dielectric constants) via the measurement of  $\tau$  values (**Appendix 1—figure 6**).

**Appendix 1—table 1.** Maximum allowed solvent-accessible surface area (SASA) for each amino acid, taken from *Tien et al., 2013*.

| Amino acid | Maximum SASA (Å <sup>2</sup> ) |
|------------|--------------------------------|
| ALA        | 129                            |
| ARG        | 274                            |
| ASN        | 195                            |
| ASP        | 193                            |
| CYS        | 167                            |
| GLN        | 225                            |
| GLU        | 223                            |
| GLY        | 104                            |
| HIS        | 224                            |
| ILE        | 197                            |
| LEU        | 201                            |
| LYS        | 236                            |
| MET        | 224                            |
| PHE        | 240                            |
| PRO        | 159                            |
| SER        | 155                            |
| THR        | 172                            |
| TRP        | 285                            |
| TYR        | 263                            |
| VAL        | 174                            |

**Appendix 1—table 2.** Mass spectrum of purified ELP.

|                | Mass (Da)   |          |
|----------------|-------------|----------|
|                | Theoretical | MALDI-MS |
| $V_{30}A_{30}$ | 24097.04    | 23898.62 |
| $A_{30}V_{30}$ | 24097.04    | 23916.19 |
| $V_{30}G_{30}$ | 23676.23    | 23495.55 |
| $G_{30}V_{30}$ | 23676.23    | 23498.98 |

Since we used  $V_{30}X_{30}$  and  $X_{30}V_{30}$  to quantify the V- and X-end of the V-X blocks, it is possible that the observed differences arose from the innate property of the  $V_{30}X_{30}$  and  $X_{30}V_{30}$  sequences. To rule

out this artifact, we formed the ELP condensates with sequences of  $V_{30}X_{30}$ ,  $X_{30}V_{30}$ , or the  $V_{30}X_{30}$  and  $X_{30}V_{30}$  mixture. The condensates were subsequently treated with the aldehyde-BODIPY and methyl-ester SBD fluorophores without the NHS ester reactive warhead (**Figure 3—figure supplement 2**). After brief incubation, aldehyde-BODIPY and methyl-ester SBD fluorophores were recruited into and homogeneously distributed in the ELP condensates. The fluorescence lifetime of aldehyde-BODIPY was the same for  $V_{30}A_{30}$  (4.96 ns),  $A_{30}V_{30}$  (4.99 ns), and their mixture (4.98 ns) (**Figure 3—figure supplement 2B**, upper panel). Interestingly, this value is around the average (4.89 ns) of the A-end (4.35 ns) and the V-end (5.43 ns)-labeled NHS-BODIPY. For the SBD measurement, methyl-ester SBD resulted in almost identical lifetime values of  $V_{30}A_{30}$  (8.25 ns),  $A_{30}V_{30}$  (8.27 ns), and their mixture (8.28 ns) (**Figure 3—figure supplement 2B**, lower panel), again around the average values (7.88 ns) of the A-end (7.00 ns) and the V-end (8.75 ns)-labeled NHS-SBD. In addition to the V-A blocks, similar observations were made for the V-G blocks as  $V_{30}G_{30}$  and  $G_{30}V_{30}$  sequences (**Figure 3—figure supplement 2C**). The slight difference between the results is attributed to the experiment errors. Because the fluorophores did not covalently label the amino-terminus of the ELP peptides, their lifetime reports closer to the averaged property of the condensates instead of the microscopic property of the V-end or the X-end when the number of molecules is sufficient and the molecular distribution has no preference. Our results reveal that the  $V_{30}X_{30}$  and  $X_{30}V_{30}$  condensates exhibited similar macroscopic viscosity or polarity, suggesting that the previously observed different viscosity or polarity of  $V_{30}X_{30}$  and  $X_{30}V_{30}$  could be attributed to the microscopic property of the V-end or X-end.

## Supplemental theory

Relationship between the transition temperature  $T_i$  and the critical temperature  $T_c$

The temperature measured by Urry et al. and presented in the main text corresponds to the transition temperature,  $T_i$  (Urry, 1997). This temperature is defined as the value above which ELP forms an insoluble coacervate phase. The value of  $T_i$  depends on the concentration of the polymer solution.

As shown by Chilkoti and coworkers (Meyer and Chilkoti, 1999; Meyer and Chilkoti, 2004), the critical temperature  $T_c$  is indeed linearly related to  $T_i$  with the following relationship:

$$T_i = T_c + \frac{k}{\text{length}} \ln \frac{C_c}{\text{conc}}. \quad (11)$$

The above equation highlights the dependence of  $T_i$  on the chain length (length) and polymer concentration (conc). The parameter  $C_c$  is the corresponding theoretical polypeptide concentration that would be required to achieve  $T_c$  and  $k$  is the proportionality constant.

Relationship between the critical temperature  $T_c$ , the Flory–Huggins parameter  $\chi$ , and the surface tension,  $\tau$

The Flory–Huggins parameter,  $\chi$ , is defined as

$$\chi = \frac{z}{2k_B T} [\epsilon_{pp} + \epsilon_{ss} - 2\epsilon_{ps}] = \frac{z\Delta\epsilon}{k_B T}, \quad (12)$$

where  $\Delta\epsilon = \frac{1}{2}[\epsilon_{pp} + \epsilon_{ss} - 2\epsilon_{ps}]$  is the coordination number,  $T$  is the temperature,  $k_B$  is the Boltzmann constant, and  $\epsilon_{pp}$  are the strength of polymer-polymer, solvent-solvent, and polymer-solvent interactions respectively (Flory, 1942).

From the original derivation of Flory–Huggins theory, it can be shown that phase separation occurs when  $\chi$  is greater than its critical value, or  $\chi > \chi_c \approx \frac{1}{2}$ . From  $\chi_c$ , we can derive the critical temperature as

$$T_c = \frac{2z\Delta\epsilon}{k_B}. \quad (13)$$

Therefore,  $\chi = \frac{T_c}{2T}$

Relationships between the Flory–Huggins parameter,  $\chi$ , and interfacial tension ( $\tau$ ) have been investigated, and the relationship can be approximated as

$$\tau \propto \chi^\alpha, \quad (14)$$

where  $\alpha$  is a positive constant, whose exact value depends on the proximity of  $\chi$  to the critical value of  $\chi$  necessary for phase separation ( $\chi_c$ ) (Roe, 1975; Helfand and Tagami, 1972). Correspondingly, we have  $\tau \propto T_c^\alpha$ .

We note that the expression for  $\chi$  (Equation 12) is simplified and assumes that  $\Delta\epsilon$  is independent of temperature. However, changes in solvent packing upon phase separation can result in entropic contributions Rubinstein and Colby, 2003. In general,  $\Delta\epsilon$  should be expressed as  $\Delta\epsilon = \Delta h + T\Delta s$ . The entropic contributions are essential for giving rise to phase separations with lower critical temperatures. Therefore, in the generalized case, we have

$$\chi = \frac{z\Delta\epsilon}{k_B T} = \frac{z(\Delta h + T\Delta s)}{k_B T}. \quad (15)$$

At the critical temperature, we have

$$\frac{1}{2} = \frac{z(\Delta h + T_C \Delta s)}{k_B T_C}, \quad (16)$$

and

$$T_C = \frac{z\Delta h}{k_B/2 - z\Delta s}. \quad (17)$$

For systems exhibiting lower critical solution temperatures (LCST),  $\Delta h < 0$  and  $\Delta s > 0$ . For  $T_c > 0$ , we have  $k_B/2 < z\Delta s$ .

Plugging the above expression back to Equation 15, we have

$$\chi = \frac{k_B T_C/2 - z\Delta s T_C + zT\Delta s}{k_B T} = \frac{T_C}{T} \left( \frac{1}{2} - \frac{z\Delta s}{k_B} \right) + \frac{z\Delta s}{k_B}. \quad (18)$$

Furthermore, as shown by Schauperl et al., 2016,  $\Delta s$ , while significant, remains relatively constant across different amino acids. Correspondingly, we have  $\tau \propto \left[ \frac{T_c}{T} \left( \frac{1}{2} - \frac{z\Delta s}{k_B} \right) + \frac{z\Delta s}{k_B} \right]^\alpha$  holds true in the more general case as well. Since  $\frac{z\Delta s}{k_B} > \frac{1}{2}$ , the slope between  $\tau$  and  $T_c$  is negative.

## Relationship between the critical temperature, $T_c$ and hydrophobicity scales

As shown in Equation 13, in the simple case  $T_c$  is linearly related to  $\Delta\epsilon$ . For the more general case, from Equation 17, we have

$$T_C = \frac{z\Delta h}{k_B/2 - z\Delta s} = \frac{z\Delta\epsilon}{k_B/2 - z\Delta s} - \frac{zT\Delta s}{k_B/2 - z\Delta s}. \quad (19)$$

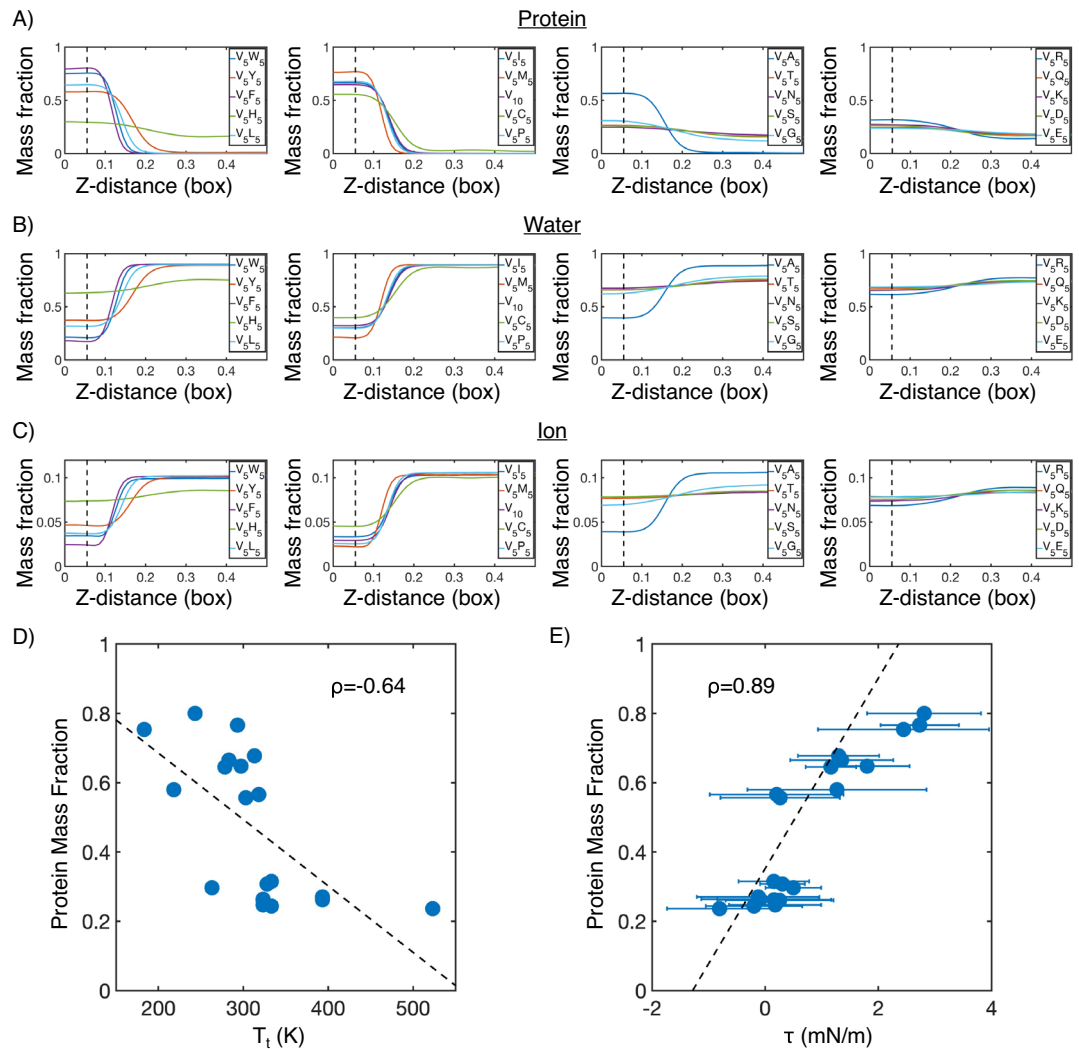
Assuming that  $\Delta s$  is relatively independent of amino acids, as shown by Schauperl et al., 2016,  $T_c$  is again linearly proportional to  $\Delta\epsilon$ .

$\Delta\epsilon = \frac{1}{2}[\epsilon_{pp} + \epsilon_{ss} - 2\epsilon_{ps}]$  can indeed be interpreted as the free energy cost of transferring a polymer bead from a solution phase to a polymer phase. It corresponds to the change of energy from a mixed state, with contacts between polymer and solvent ( $\epsilon_{ps}$ ), to the demixed state with only polymer-polymer ( $\epsilon_{pp}$ ) and solvent-solvent ( $\epsilon_{ss}$ ) contacts.

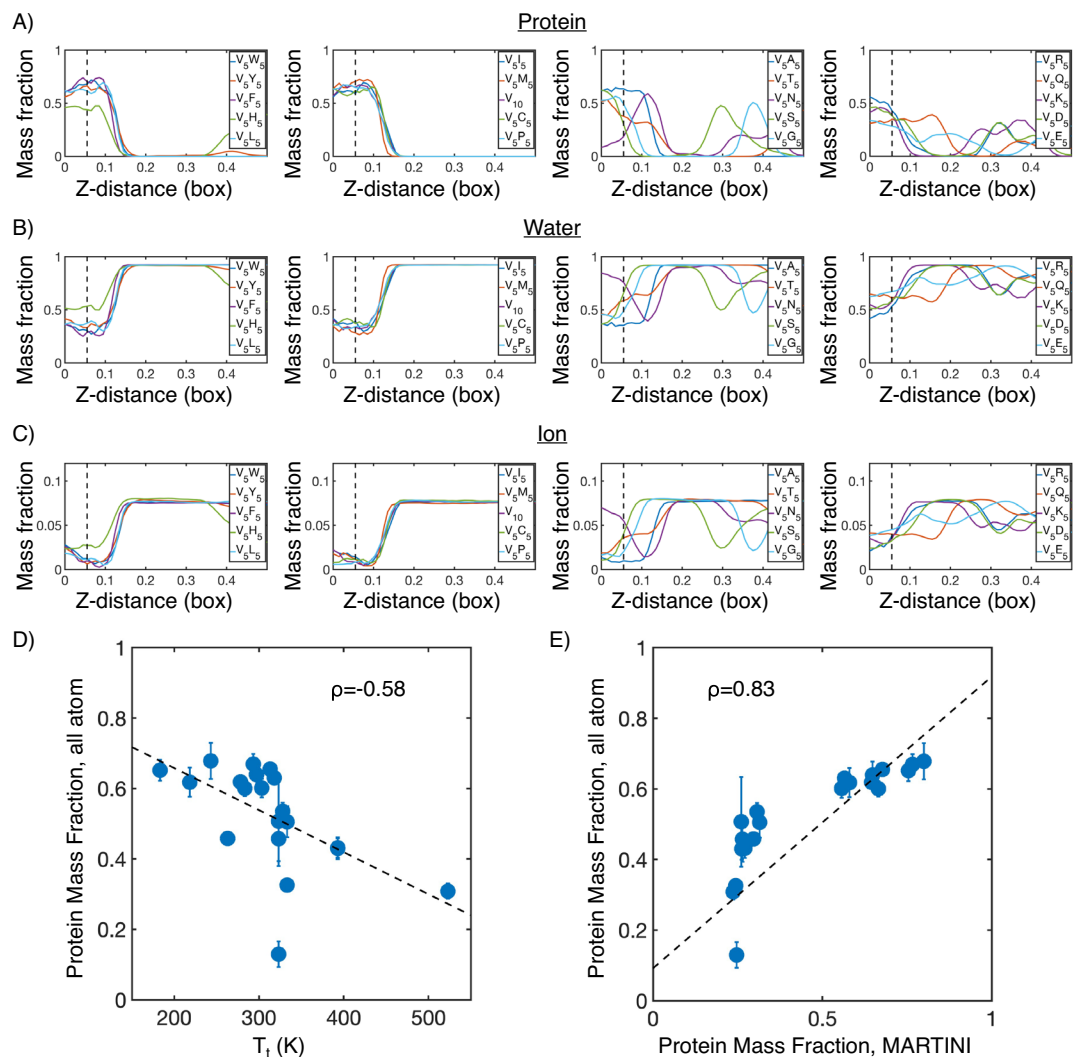
Therefore, the transfer free energy, and the interactions among amino acids of ELPs, are expected to correlate with the critical temperature.

## Experimental sequences

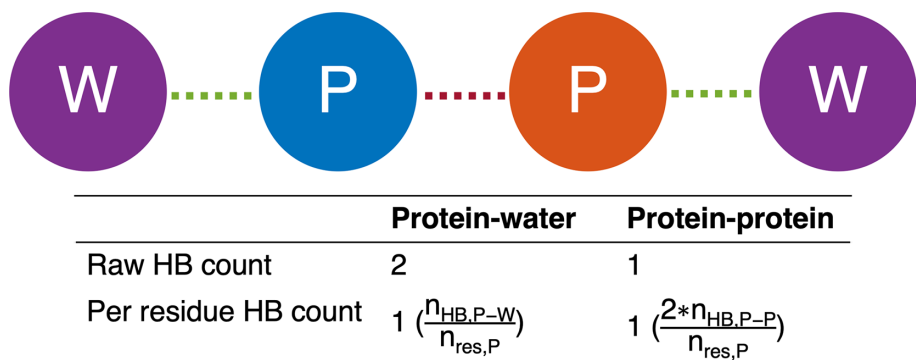
V<sub>30</sub>A<sub>30</sub>  
M-(GVGVP)<sub>30</sub>-(GAGVP)<sub>30</sub>-GY  
A<sub>30</sub>V<sub>30</sub>  
M-(GAGVP)<sub>30</sub>-(GVGVP)<sub>30</sub>-GY  
V<sub>30</sub>G<sub>30</sub>  
M-(GVGVP)<sub>30</sub>-(GGGVP)<sub>30</sub>-GY  
G<sub>30</sub>V<sub>30</sub>

M-(GGGVP)<sub>30</sub>-(GVGVP)<sub>30</sub>-GY


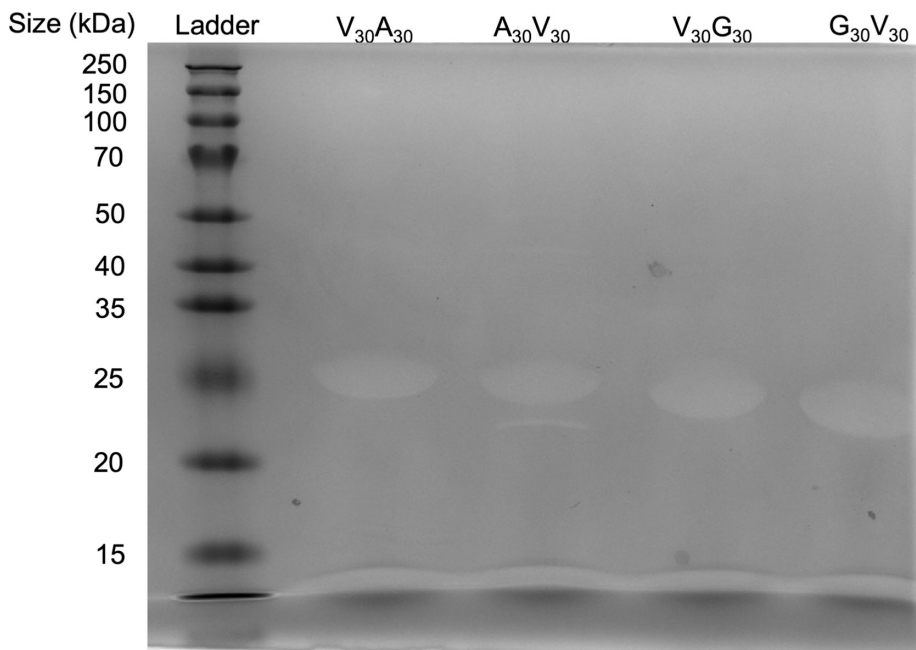
**Appendix 1—figure 1.** Mass fraction from MARTINI condensate simulations. (A) Protein mass fraction along the Z-distance from the condensate center defined using the largest cluster. The Z-axis is defined as the direction perpendicular to the condensate-water interface. The dashed line represents a Z-distance of 0.06 box lengths. Average values below this threshold are used for correlation analysis in (D) and (E). (B) Water mass fraction along the Z-distance from the condensate center. (C) Ion mass fraction along the Z-distance from the condensate center. (D) Correlation between the average protein mass fraction and transition temperature ( $T_t$ ) from Urry, 1997.  $\rho$  is the Pearson correlation coefficient between the two data sets, and the dashed diagonal line is the best fit line. Error bars represent SDs of the mean and are smaller than the symbols. (E) Correlation between the average protein mass fraction and the simulated surface tension ( $\tau$ ).  $\rho$  is the Pearson correlation coefficient between the two data sets, and the dashed diagonal line is the best fit line. Error bars represent SDs of the mean taken over six equally spaced box length intervals.



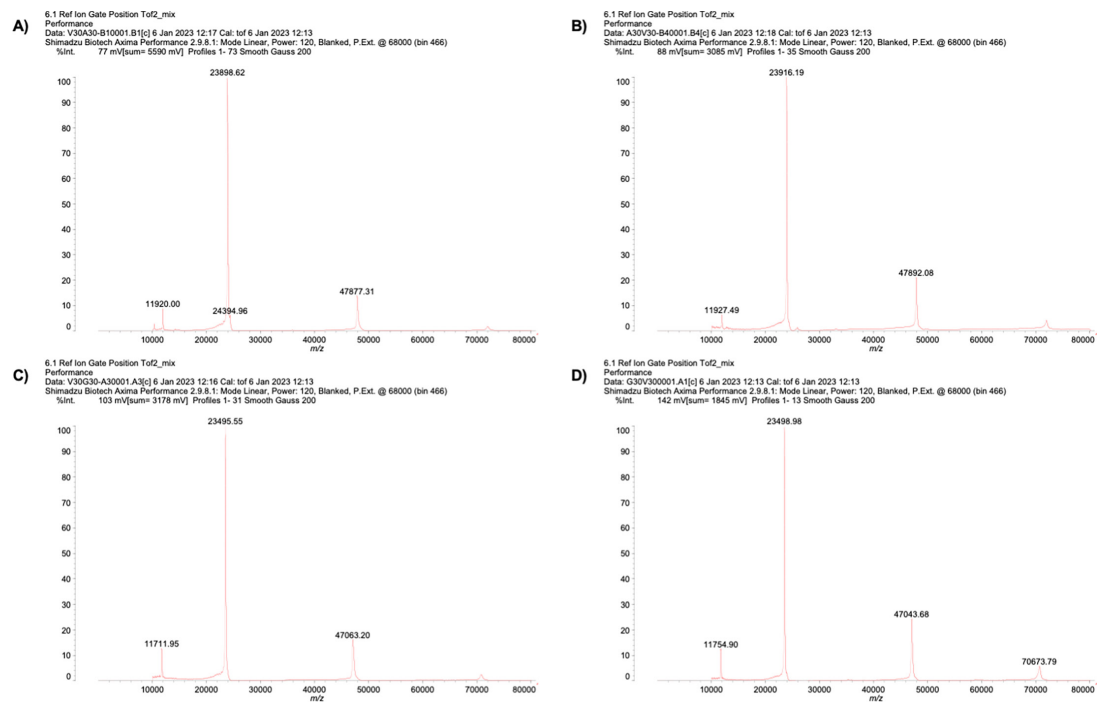
**Appendix 1—figure 2.** Mass fraction from atomistic condensate simulations. (A) Protein mass fraction along the Z-distance from the condensate center defined using the largest cluster. The Z-axis is defined as the direction perpendicular to the condensate-water interface. The dashed line represents a Z-distance of 0.06 box lengths. Average values below this threshold are used for correlation analysis in (D) and (E). (B) Water mass fraction along the Z-distance from the condensate center. (C) Ion mass fraction along the Z-distance from the condensate center. (D) Correlation between the average protein mass fraction in all-atom simulations and transition temperature ( $T_t$ ) from Urry, 1997.  $\rho$  is the Pearson correlation coefficient between the two data sets, and the dashed diagonal line is the best fit line. Error bars represent SDs of the mean taken over six equally spaced box length intervals. (E) Correlation between the average protein mass fraction from all-atom simulations and the average protein mass fraction from MARTINI simulations.  $\rho$  is the Pearson correlation coefficient between the two data sets, and the dashed diagonal line is the best fit line. Error bars represent SDs of the mean taken over six equally spaced box length intervals.



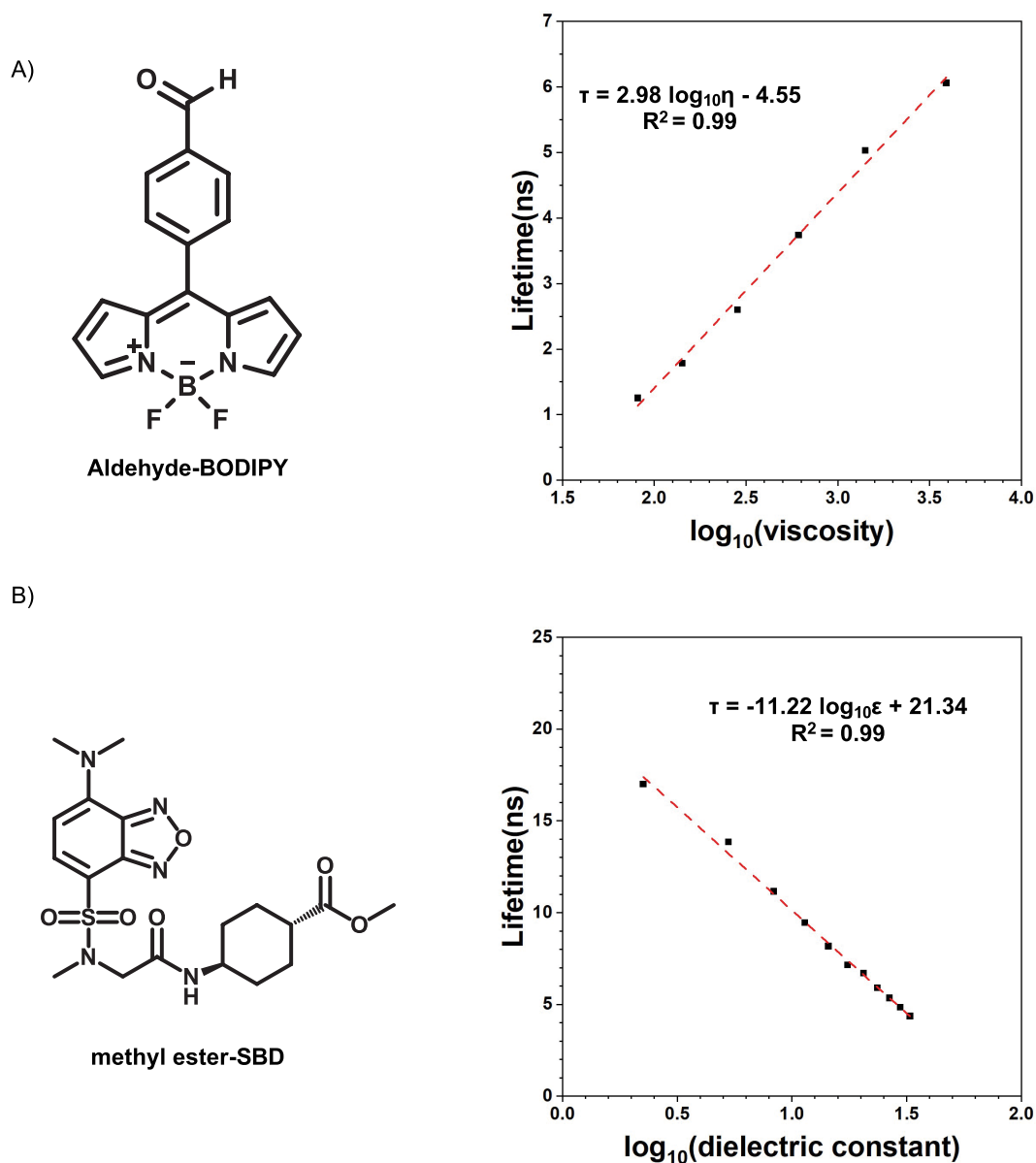
**Appendix 1—figure 3.** Graphical illustration of number of hydrogen bonds per protein residue. The protein is depicted as blue and orange circles, and the water molecules are depicted as purple circles. Protein-protein and protein-water hydrogen bonds are drawn as red and green dashed lines, respectively. The number of protein-water hydrogen bonds per protein residue is calculated by taking the raw number of protein-water hydrogen bonds ( $n_{HB,P-W}$ ) and dividing by the number of residues in the protein ( $n_{res,P}$ ). Meanwhile, the number of protein-protein hydrogen bonds per protein residue is calculated by doubling the number of protein-protein hydrogen bonds ( $n_{HB,P-P}$ ) and dividing by the number of residues in the protein ( $n_{res,P}$ ). The extra factor of 2 accounts for the fact that, in the case of protein-protein hydrogen bonds, both the donor and the acceptor must reside within the protein.



**Appendix 1—figure 4.** SDS-PAGE gel of elastin-like polypeptide (ELP). The purity of the protein is >95%.  
The online version of this article includes the following source data for appendix 1—figure 4:  
**Appendix 1—figure 4—source data 1.** Original files for western blot analysis displayed in **Appendix 1—figure 4**.  
**Appendix 1—figure 4—source data 2.** Original western blots for **Appendix 1—figure 4**, indicating the relevant bands and treatments.



**Appendix 1—figure 5.** Mass spectrum of (A)  $V_{30}A_{30}$ , (B)  $A_{30}V_{30}$ , (C)  $V_{30}G_{30}$ , and (D)  $G_{30}V_{30}$ .



**Appendix 1—figure 6.** Calibration curves of fluorophores. (A) Lifetime-viscosity calibration curve is quantified in the glycerol at different temperatures (10°C  $\eta = 3900$  cP; 20°C  $\eta = 1,412$  cP; 30°C  $\eta = 612$  cP; 40°C  $\eta = 284$  cP; 50°C  $\eta = 142$  cP; 60°C  $\eta = 81.3$  cP) *Segur and Oberstar, 1951*?. The fluorophore concentration is 5  $\mu$ M. (B) Lifetime-dielectric constant calibration curve is quantified in the MeOH-1,4-dioxane mixture.  $\epsilon_{mix} = \phi_1 \epsilon_1 + \phi_2 \epsilon_2 = 32.7$ ,  $\epsilon_{1,4-dioxane} = 2.25$ .