Expanding the molecular language of protein liquid-liquid phase separation

Shiv Rekhi[1#], Cristobal Garcia Garcia[2#], Mayur Barai[3#], Azamat Rizuan[1], Benjamin S. Schuster[3*], Kristi L. Kiick[2,4*], Jeetain Mittal[1,5,6*]

[1]Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843, USA

[2]Department of Materials Science and Engineering, University of Delaware, Newark, DE 19716, USA

[3]Department of Chemical and Biochemical Engineering, Rutgers, the State University of New Jersey, Piscataway, NJ 08854 USA

[4]Department of Biomedical Engineering, University of Delaware, Newark, DE 19176, USA

[5]Department of Chemistry, Texas A&M University, College Station, TX 77843, USA

[6]Interdisciplinary Graduate Program in Genetics and Genomics, Texas A&M University, College Station, TX 77843, USA

[#]These authors contributed equally
[*]benjamin.schuster@rutgers.edu
[*]kiick@udel.edu
[*]jeetain@tamu.edu

Abstract

Understanding the relationship between a polypeptide sequence and its phase separation has important implications for analyzing cellular function, treating disease, and designing novel biomaterials. Several sequence features have been identified as drivers for protein liquid-liquid phase separation (LLPS), schematized as a 'molecular grammar' for LLPS. In this work, we further probed how sequence modulates phase separation and the material properties of the resulting condensates, targeting sequence features previously overlooked in the literature. We generated sequence variants of a repeat polypeptide with either no charged residues, high net charge, no glycine residues, or devoid of aromatic or arginine residues. All but one of twelve variants exhibited LLPS, albeit to different extents, despite significant differences in composition. Further, we found that all condensates formed behaved like viscous fluids, despite large differences in their viscosities. Our results support the model of multiple interactions between diverse residue pairs – not just a handful of residues – working in tandem to drive phase separation and dynamics of condensates.

Introduction

Biomolecular condensates, which often form through liquid-liquid phase separation (LLPS),[1,2] are responsible for diverse cellular functions in the cytoplasm and nucleus.[3-6] Deciphering the role of different amino acids and the associated molecular interactions that drive the thermodynamics and dynamics of LLPS is valuable for underpinning fundamental understanding of these biologically important assemblies,[7-10] unlocking the potential of condensates for applications ranging from materials[11-14] to synthetic biology.[15-17]

Recent studies on natural and artificial sequences have painted a picture of the drivers of protein LLPS.[11,18-26] Many proteins that undergo LLPS are intrinsically disordered or contain intrinsically disordered regions (IDRs), often of low complexity, i.e., enriched in polar, aromatic, and proline and glycine residues.[7,27,28] It has become clear from recent studies that higher content and segregation of aromatic residues, often Tyr, can underlie protein phase separation.[21,23,29,30] Additionally, Arg, but not Lys, has been proposed to drive phase separation through interaction with aromatic residues.[18,19,21,31] It is hypothesized that the overall interactions between the Arg and aromatic sidechains – including hydrogen bonding, sp^2/π, and the electrostatic interaction between the guanidinium group and the aromatic ring – contribute to the strength of this interaction.[12,20,21,24] Many studies have also highlighted the role of charge-charge attraction between oppositely charged residues (Arg/Lys with Asp/Glu) in modulating the phase behavior of different proteins.[19,24,32]

Building upon considerable work that has established important molecular features of LLPS,[11,19-22,24-26,29] here we consider what additional interactions play key contributing roles and how these interactions might be leveraged to tune LLPS and the physical properties of the condensates. To tackle this, we designed and tested sequences that deviate from the existing

heuristics and predictors of phase separation. In conducting these studies, we also addressed another key question: To what extent can protein LLPS be modeled as a process driven by a handful of residues, or is it necessary to account for interactions between many diverse residues?[33]

We began our studies with the sequence $(GRGDSPYS)_{25}$, an artificial intrinsically disordered protein (A-IDP) with upper critical solution temperature (UCST) phase behavior[12,34]. We then carefully designed variants of this polypeptide to dissect the role of different residues individually, and in conjunction with other residues, on *in vitro* phase separation. We examined sequences that are uncharged, have a large net charge, lack aromatic or arginine residues, or are devoid of glycines. Such sequences lack features that are believed to be critical for LLPS based on the current understanding.[20,35-37] Surprisingly, we identified sequences in each category that exhibit LLPS. We obtained mechanistic insights into the atomic interactions between amino acids and quantified changes in the viscosity of the condensates due to specific amino acid substitutions. Collectively, this work highlights the collective contribution of diverse residues and significantly expands our understanding of the molecular language of protein LLPS.

2. Results

Diverse interactions drive A-IDP phase separation

The WT A-IDP $(GRGDSPYS)_{25}$ contains several interactions believed to play a key role in modulating LLPS, including hydrogen-bonding,[24,38] electrostatic attraction/repulsion,[18,19,24] aromatic-cationic interactions,[12,18,19,21,26] sp^2/π interactions,[20,21,26] and van der Waals interactions[10,38] (**Fig. 1a**). Additionally, the A-IDP is similar in length and shows high compositional similarity to proteins from the FET family as well as other naturally occurring prion-like low complexity domains (PLCDs)[27] (**Fig. 1a**), supporting its use to examine the completeness

of the existing molecular grammar originally proposed for FUS-like proteins,[21] later extended to PLCDs.[26]

Turbidimetry experiments at 6 μM protein concentration in phosphate-buffered saline (PBS) show a UCST transition at a temperature (T_t) of 45 °C (**Fig. 1b**). Turbidity measurements at several concentrations (**Supplementary Figure 1**) are used to generate a partial phase diagram, consisting of the low concentration side (left arm), which is also referred to as the saturation concentration (C_{sat}) (**Fig. 1c**). Microscopy images revealed that WT forms spherical droplets that grow with time – a characteristic of LLPS, rather than aggregation (**Fig. 1d**). We monitored droplet fusion events, which confirmed these droplets are contacting, fusing, and relaxing into larger spherical droplets (**Fig. 1e**), verifying their viscous liquid-like nature. Plotting aspect ratio vs. time for a typical fusion event shows that the relaxation time of these fusing droplets is less than 100 ms, comparable to that of similarly sized droplets comprised of the LAF-1 RGG domain[24] (**Supplementary Figure 2**).

As the WT lacked aliphatic hydrophobic residues, which are present in other PLCDs (**Fig. 1a**), we produced recombinant Gly-to-Ala (ARADSPYS)$_{25}$ and Ser-to-Val (GRGDVPYS)$_{25}$ variants. (All sequences described in this study contain an octapeptide repeated 25 times, so for brevity we often drop the subscript 25 hereafter.) Neither variant shows much change in C_{sat} at room temperature, as compared to WT (**Fig. 1g**). However, the Ser-to-Val variant results in a dramatic increase in T_t, from 46.3 ± 1.5 °C to 62.3 ± 2.5 °C, at 6 μM concentration. This indicates that GRGDVPYS condensates are more thermally stable, eventually showing a lower C_{sat} than WT at temperatures above ~40 °C (**Fig. 1g**).

We further find that Gly-to-Ser mutations (SRSDSPYS) lead to comparable LLPS behavior as that of Gly-to-Ala (and WT) at lower concentrations, although reduced T_t are observed

at higher concentrations (**Supplementary Figure 3**), highlighted by a downward shift in the phase diagram (**Fig. 1g**). These results suggest that Ala and Gly impart similar LLPS behavior in A-IDPs, but the polar Ser residue contribution can be similar or different to Ala/Gly, depending on the solution concentration or temperature. Importantly, microscopy experiments show that all these variants form liquid-like droplets, indicating that Gly residues are not strictly required for condensate fluidity (**Fig. 1f**).

Atomistic simulations uncover the role of polar residues

We conducted a microsecond-long all-atom explicit solvent molecular dynamics simulation of a WT protein condensate (**Fig. 2a**) using a previously reported strategy[39] (see Methods) to gain insights into the residue-level molecular interactions stabilizing the protein-rich phase. We find that the condensate remains stable over the 1 μs time with rapid equilibration of solvent density and a relatively slow change in protein concentration profile over this time (**Supplementary Figure 4**). The average protein density after removing the initial 250 ns equilibration is 428 mg/ml (**Fig. 2b**), similar to densities measured experimentally for condensates of the FUS low-complexity (LC) domain (~447 mg/ml)[22] and DDX4 (~400 mg/ml).[19] The polypeptide repeat unit contains a zwitterionic pair (Arg and Asp residues), so the concentrations of Na^+ and Cl^- ions are similar to each other both inside and outside the condensate (**Fig. 2b**), which is confirmed by comparing the simulation results to a previously developed theoretical model[39] for estimating salt concentration within the condensed phase **(Supplementary Figure 5)**. Interestingly, the concentration of ions inside the condensate is half that of outside, reflecting the free energy penalty associated with desolvation of ions.[39]

Next, we calculated the average number of pairwise residue contacts to dissect the prevalence of specific amino acid pairings inside the protein condensate. Consistent with previous

computational work on the FUS LC and LAF-1 RGG domains[39] and NMR experiments of FUS LC,[22,25] we find that effectively all residue pairs form contacts through backbone-backbone (bb-bb), sidechain-sidechain (sc-sc), and backbone-sidechain (bb-sc) atomic interactions (**Fig. 2c top**). Perhaps somewhat surprisingly in the context of the current understanding of LLPS molecular grammar, contacts involving Ser and Gly residues are observed frequently, in addition to contacts between Tyr with Arg, Tyr with itself, and electrostatic interactions between Arg and Asp. In fact, when we sum the number of contacts formed by each residue within the condensate (**Fig. 2c top, inset**), following Tyr and Arg, Ser contacts are prevalent within the condensate, followed by Gly, Asp, and Pro.

The non-normalized contact frequency (**Fig. 2c top**) is the relevant metric for estimating thermodynamic driving forces of LLPS. However, as Ser and Gly appear twice as many times as any other residue in the WT sequence, we normalize the number of contacts by amino acid frequency to assess the contribution of contact pairs on a one-to-one basis. Normalization reveals that in the absence of compositional biases, contacts involving Arg and Tyr are most frequent (**Fig 2c bottom**), in line with the expectation of these residues being the primary drivers of phase separation.[18-21,23,26] Notably, the normalized contacts highlight that Arg and Tyr interact not only with each other and themselves, but also with other residues – e.g., R:P, R:S, S:Y, and G:Y are also prevalent within the condensed phase.

Another important question is whether interactions that drive condensate formation are substantially different than interactions that stabilize the condensate once it has formed. To provide insight into this, we performed a single chain simulation of $(GRGDSPYS)_6$. We observe a strong correlation between single-chain and condensed phase contacts **(Fig. 2d,e)**, implying that the

interactions driving formation of the WT condensate and its stabilization once formed are highly similar.

In most IDPs, Arg and Tyr are much less enriched compared to residues such as Gly, Gln, and Ser (**Fig. 1a**). The simulation results emphasize that residues that are more abundant do in fact contribute to LLPS through favorable contact formation. This aligns with other recent simulation work showing that for an IDP to form liquid-like condensates, a significant fraction of residues must form energetically favorable contacts.[33] Thus, our atomistic simulations strongly support the relevance and contributions of interactions mediated by uncharged, non-aromatic residues in stabilizing protein condensates.

Removal of zwitterionic pair reduces phase separation

To investigate the contribution of non-electrostatic interactions to LLPS, we targeted the zwitterionic pair in WT A-IDP by replacing Arg and Asp with polar residues, Gln and Asn, respectively. Notably, the new variant, GQGNSPYS, also undergoes LLPS at a concentration of 6 μM, albeit with a lower T_t of 25.3 ± 0.6 °C (versus 46.3 ± 1.5 °C for WT) (**Fig. 3a, Supplementary Figure 6**). The reduced LLPS propensity of this variant due to the removal of favorable attractions of Arg with Asp (electrostatic attraction) and Tyr (e.g., cation-π) is also highlighted by a downward shift in the phase diagram and increase of C_{sat} at a fixed temperature (**Fig. 3a**). As with WT, the formation of dense liquid-like droplets and growth in droplet size with time is observed via microscopy (**Fig. 3b, Supplementary Figure 7**). Despite losing favorable interactions upon mutation of Arg and Asp, the polypeptide variant without the zwitterionic pair can still undergo LLPS at relatively low concentrations (**Fig. 3c**), with C_{sat} ~ 4 μM at 20 °C, as compared to FUS LC C_{sat} ~125 μM at physiological salt conditions and hnRNPA1 C_{sat} ~100 μM at 20 °C.[21,26] This

result supports the model of diverse interactions driving LLPS in the WT polypeptide, as identified by our atomistic simulations.

We simulated this new variant by introducing Arg-to-Gln and Asp-to-Asn mutations in the final configuration of the WT system (see Methods) and find that the protein condensate remains stable throughout the 1 μs simulation time (**Supplementary Figure 8**). Like WT, we observe interactions between most amino acid pairs inside the condensed phase (**Supplementary Figure 9**). Comparing the number of contacts between WT and GQGNSPYS (**Fig. 3d**), an overall reduction in contacts is observed upon removing the zwitterionic pair (with a notable exception for contacts involving the Asn residue), consistent with the variant's reduced phase separation in experiments. To reveal changes in the pairwise contact formation in a manner that accounts for differences between sequences in their total number of contacts, we normalize $P_{contact}$ for the variant with $P_{contact}$ of WT, where $P_{contact} = N_{contact} / N_{total}$ and N_{total} is the total number of contacts formed by all residue pairs (**Fig. 3d**). We observe lower contact probability for pairs involving Arg-to-Gln mutations, and higher contact probabilities for Asp-to-Asn pairs, while many other pairs (e.g., G:G, G:Y, S:S, S:Y, P:Y, and Y:Y) show no appreciable change, further highlighting the need to consider uncharged, non-aromatic residues and their contribution to LLPS.

Charged variants exhibit contrasting phase behavior

Next, we replaced only Asp with Asn, yielding repeat unit GRGNSPYS. This allows us to interrogate the role of Arg in LLPS of a polypeptide containing aromatic Tyr but devoid of anionic residues (i.e., in the absence of electrostatic attraction).[11,37,40,41] Remarkably, at 6 μM, we observe formation of droplets with liquid-like material properties for this highly charged polycationic variant (**Fig. 3a, b , Supplementary Figure 6**), though T_t is significantly reduced with respect to WT and is slightly reduced from GQGNSPYS (**Supplementary Figure 6)**. It is worth noting that

the C_{sat} of this polycationic variant is slightly higher than GQGNSPYS at lower temperatures, but lower at higher temperatures (**Fig. 3a and Fig. 3c**), showing the need to consider environmental conditions, such as temperature, as part of the complete molecular language of LLPS. Additionally, T_t of GRGNSPYS increases monotonically with salt concentration (**Extended Data Fig. 1a,b**) possibly due to screening of electrostatic repulsion.

We also tested a polyanionic variant in which we replaced Arg with Gln (GQGDSPYS). The turbidity measurements show no phase separation even up to high concentrations of 80 μM (**Supplementary Figure 6**), which is confirmed by absence of droplet formation in microscopy experiments (**Fig. 3b and Supplementary Figure 7**). The polyanion's behavior is expected given its high net charge (and the absence of elevated salt concentration)[42], but contrasts with the LLPS behavior of the polycationic variant containing Arg.

Atomistic simulations of the condensates formed by the polycationic and polyanionic sequences show a significant expansion of the protein-rich phase within the simulation box (**Supplementary Figure 8)**, and the protein-rich phase incorporates a much higher number of oppositely charged ions to neutralize the excess charge. The number of contacts formed between all residue pairs is lower in these variants than in WT (**Fig. 3e,f, Extended Data Fig. 2a,b**, **Supplementary Figure 9**), but the normalized data again highlights the prevalence of contacts between all other residue pairs (**Fig. 3e, f, Supplementary Figure 9**).

Polycationic Lys variant undergoes phase separation

It is widely believed that attractive contacts between Arg and Tyr are indispensable to the LLPS of proteins that contain these residues, based on extrapolation from the seminal work of Wang et al.[21] on FUS-family proteins. However, a simple empirical theory to predict C_{sat} based on the number of Arg and Tyr residues does not capture the observed behavior for many proteins.[24]

We therefore revisited the role of Arg and Tyr in the context of the A-IDP sequences, and carried out Arg-to-Lys substitutions in the polycationic sequence to test this directly. The new variant GKGNSPYS does not display LLPS at 6 μM (**, Fig. 3b, Supplementary Figure 6**), although upon increasing concentration to 20 μM, this Lys-rich polycationic solution becomes turbid below a T_t of 10.7 ± 2.9°C (**Supplementary Figure 6**). Microscopy confirms that GKGNSPYS assembles into liquid-like droplets at room temperature for 60 μM protein concentration (**Fig. 3b**). In many previous studies, when all Arg residues were mutated to Lys in a sequence with a significant fraction of these residues (13.7% Arg in LAF-1 RGG, 10.2% in DDX4),[19,24] no measurable droplet formation was observed;[19,24,26] this was interpreted by many that Arg is required for LLPS of PLCDs at physiologically relevant conditions. However, based on the saturation curves for the GKGNSPYS and GQGNSPYS variants (**Fig. 3a)**, GKGNSPYS may even phase separate more avidly than GQGNSPYS at higher concentrations, highlighting the significance of mapping LLPS at a range of temperatures and concentrations.

Arg and Lys differ beyond interactions with Tyr

To elucidate the origin of the differences in the interactions of cationic residues (Lys vs. Arg) with Tyr, we computed the probability distribution of Arg/Lys-Tyr sidechain angle and distance from GKGNSPYS and GRGNSPYS condensate simulations (**Fig. 3g**). At shorter distances, a significant fraction of configurations involve stacking of the π-ring above or below the Arg sidechain,[20,43] which is not observed in the case of Lys. Comparing contacts of Arg or Lys with other residues, we observe that all contacts involving Lys are decreased, whereas other contacts remain relatively unchanged compared to the polycationic Arg variant (**Fig. 3h**). In addition to Tyr, the contacts of Lys with polar residues such as Ser and Asn also show a

pronounced reduction[44], implying that these often ignored interactions of Arg with other residues may be essential in LLPS.

Phase separation of the polycationic Arg variant implies that short-range interactions encoded within the polypeptide sequence compensate for long-range electrostatic repulsion due to high net charge. A potential stabilizing force that compensates for electrostatic repulsion is the interaction of the Arg guanidium group with polar atoms through sp^2 and hydrogen bonding interactions, as seen in folded protein structures[20,45,46] and ion channels.[44] To test this, we mutated the Asn and Ser residues at positions 4 and 8 of the GRGNSPYS repeat unit to Ala. LLPS of this new variant, GRGASPYA, is measurably reduced (**Extended Data Fig. 3a,b**). The apparent dependence of the phase behavior of GRGNSPYS on the presence of polar residues appears contrary to previous literature and similar substitutions introduced earlier in this paper, but it is consistent with the simulation results highlighting the role of polar residue interactions in the polycationic sequence. It can further be rationalized as the presence of polar residues increases the polarity of the environment surrounding Arg, minimizing charge repulsion. However, a more complete understanding of the interactions underlying the stabilization of the polycationic condensed phase would involve a non-trivial contribution from other residues, salt ions, and solvent.

Polycationic sequence phase separates without aromatics

The A-IDP variants presented to this point each contain Tyr, which has been shown to be sufficient to induce LLPS in a variety of proteins.[23,29] Our simulation data above identified a diverse array of interactions with Tyr, in addition to the commonly expected self- interactions.

To clearly separate Tyr's contribution, we carried out Tyr-to-Ala substitutions to generate a new variant, GRGNSPAS, which lacks most interactions that are currently understood to be

essential for LLPS. Comparing pairwise contacts between the GRGNSPYS and GRGNSPAS variants highlights the loss of all contacts involving Tyr-to-Ala substitution, and not just Y:Y interactions (**Fig. 4a, Supplementary Figure 10**); this is further evidence that aromatic Tyr contacts with nonaromatic residues are also essential in LLPS.

Remarkably, our experimental data shows that the new GRGNSPAS polypeptide can undergo LLPS and form liquid-like droplets (**Fig. 4b, Fig. 4c)**. At a polypeptide concentration of 40 μM, the T_t was 10.6 ± 4.5°C (**Fig. 4b, Supplementary Figure 11**). For reference, 140 residue-long α-synuclein was found to undergo phase separation at concentrations above 200 μM in the presence of 10% PEG-8000 at physiological salt conditions[47]. This result shows that although Tyr is an important contributor for phase separation, it is not required in the context of our A-IDP sequence.

To further elucidate the contributions of aromatic residues to LLPS, we mutated Tyr-to-Phe or Tyr-to-Trp in the polycationic sequence, generating GRGNSPFS and GRGNSPWS. The trends in phase separation are Trp > Tyr > Phe (**Fig. 4d**), in line with expectations from literature[12,14,24] and contacts from atomistic simulations (**Extended Data Fig. 4a,b**). Microscopy confirms the liquid-like nature of the droplets at concentrations above their respective C_{sat} values (**Fig. 4e)**.

A-IDP variants have tunable material properties

The material properties that emerge upon LLPS are critical to the function, dysfunction, and engineering of biomolecular condensates. To determine how sequence determines not only phase behavior, but also the material properties of A-IDP condensates, we employed passive microrheology using fluorescent tracer beads.[31,48,49] Tracer beads embedded within the condensates, recorded ~1 hr after sample preparation, displayed Brownian motion **(Fig 5a).** The

mean squared displacement (MSD) of the beads increased linearly with lag time for all variants, giving diffusivity exponent close to 1, emphasizing that the condensates behave largely as viscous fluids under the experimental conditions **(Fig 5b)**.

The viscosities of the A-IDP condensates, estimated from the MSD data (see Methods), vary by more than one order of magnitude **(Fig 5c)**. The viscosity of WT was measured as ~ 11 Pa.s, approximately 11,000 times the viscosity of water. Out of nine variants tested, the lowest measured η value was for GRGASPYA, at ~ 2 Pa.s. In contrast, the highest measured η was for the Trp-containing variant, GRGNSPWS, with η ~ 40 Pa.s. This 20-fold range in viscosity highlights the diverse material properties accessible through mutating IDP sequence and altering molecular interactions.

To explore the relationship between thermodynamics and dynamics for the polypeptide sequences, we calculated the correlation between the saturation concentrations for the sequences at 18 $^{\circ}$C and the measured viscosities. Interestingly, we observed only a moderate correlation between η and C_{sat}, with the Pearson correlation coefficient (R) determined to be -0.53. Logarithms of the normalized viscosities and normalized saturation concentrations **(Fig. 5d)** exhibit a power law exponent of -0.44 with an R and R^2 value of -0.86 and 0.74 respectively. Thus, a power-law model reasonably describes the relationship between viscosity and C_{sat} for the sequences used in this study[50] but cannot fully explain our results.

We observe that for GQGNSPYS and most polycationic sequences, η is reduced as compared to the zwitterionic WT sequence, consistent with the increase of C_{sat} for these variants as compared to WT. A notable exception is GRGNSPWS for which a four-fold increase in η is observed despite a small decrease in C_{sat} as compared to WT, suggesting that the bulky indole side chain may slow the dynamics of the condensates. We performed fluorescence recovery after

photobleaching (FRAP) on WT and GRGNSPWS to verify this microrheology result, and indeed, GRGNSPWS exhibits significantly slower FRAP recovery compared to WT (**Extended Data Fig. 5a**).

Several further results stand out. First, we measured only a two-fold reduction in η for the Arg-to-Lys mutation, a smaller decrease than expected based on the 300-fold reduction observed in polyK vs. polyR complex coacervates with UTP[31]. Second, microrheology of the Gly-to-Ser and Gly-to-Ala variants confirms that Gly is not necessary for fluidity of these A-IDP condensates; η of ARADSPYS is comparable to that of WT, and the SRSDSPYS sequence likewise remained liquid-like. However, SRSDSPYS exhibited a two-fold increase in η compared to WT despite similar C_{sat}, which represents another exception to the correlation between C_{sat} and viscosity. This result suggests that replacing Gly with Ser enhances the strength of intermolecular interactions, perhaps due to hydrogen bonding that lowers the mobility and dynamics of the proteins. Third, a 3-fold increase in η for GRGDVPYS (~30 Pa.s) compared to WT suggests that mutating polar Ser to bulkier hydrophobic Val promotes enhanced intermolecular hydrophobic interactions within the condensates, resulting in slowed droplet dynamics and increased viscosity.

Collectively, these observations shed light on the sequence-material properties relationship of the polypeptide condensates. These results suggest that a non-linear relationship exists between phase separation propensity and condensate viscosity[50], but the exceptions we observed point to the feasibility of independently modulating phase behavior and material properties.[51]

3. Discussion

We show that the phase separation of IDPs arises from multiple interactions between diverse residue pairs (**Fig. 6a**), and often overlooked polar and non-polar residues can stabilize the condensed phase.[11,52] The collective presence of a sufficient number of such interacting residues

in an IDP tips the balance towards phase separation,[33] leading to a diverse range of IDP sequences capable of phase separation under biologically relevant conditions. Our proposed model considers the cumulative effect of all amino acids within a sequence towards LLPS, contrary to the Stickers-and-Spacers (SaS) model, which postulates that the driving forces originate only from "sticker" residues.[53,54] It is worth noting that the SaS model is meant to be the simplest possible interpretative theory in the spirit of other successful two-state statistical mechanical models,[55] which can yield general insights evidenced by its widespread adoption in literature.[21,23,26,56] However, while serving as a useful conceptual framework for some IDPs,[21,23,26] the mapping of a real system to a two-state classification is not always straightforward or feasible,[54] and actual molecular mechanisms are best described by physics-based simulation models (in conjunction with laboratory experiments)[22,25,30] which do not a priori classify residues as stickers or spacers.

Solvent-mediated interactions have been shown to play a role in modulating sequence-dependent phase separation of proteins.[57,58] However experimentally determined amino acid solvation free energies[59] cannot predict the changes in C_{sat} for A-IDPs even qualitatively (**Extended Data Fig. 6a**) underlining their inability to serve as predictors for LLPS. We also tested several state-of-the-art sequence-based predictors[20,35-37,60,61] on the A-IDPs and find that the results do not match the experiments (**Extended Data Fig. 6b,c**). The molecular language presented in this work, accounting for all residues, suggests opportunities for the continued development of improved phase separation predictors.

We perform a thermodynamic analysis[26,62] to quantify the changes in C_{sat} upon mutation of different residues (**Fig. 6b, Extended Data Table 1**) and find that that certain mutations have a more pronounced effect on LLPS. However, the relative changes upon mutation are not universal, showing a significant dependence on temperature and the reference sequence (**Extended**

Data Fig. 7a,b, Extended Data Table 2). These results highlight the context-dependent nature of residue contributions to LLPS, including factors not considered here such as sequence patterning.

Considering how our results compare with earlier work, several studies report absence of LLPS when Arg is mutated to Lys,[19,24] frequently attributed to reduced cation-π interactions.[18,19,21,26] Here, the polycationic GKGNSPYS sequence undergoes LLPS at C_{sat} ~25 μM and 20 °C due to the presence of multiple sequence-encoded interactions, such that LLPS is robust to mutation. The differences between Arg and Lys, though notably divergent in their interactions with aromatic groups[63,64] due to hydrogen bonding and cation-π interactions as shown in our prior work on LAF-1 RGG,[24] also extend to their interactions with all other residues. The demonstrated importance of interactions between cationic and polar residues provides a molecular understanding of the lack of phase separation observed in the Lys rich MED1-IDR upon carrying out Ser (polar) to Ala (aliphatic) mutations.[65]

Several experimental studies have suggested that phase separation temperature and viscosity of condensates are correlated.[48,66,67] We observe a power-law dependence of normalized viscosity and C_{sat}, which suggests a coupling between thermodynamics and dynamics of condensates.[50] However, we observe sequences with similar C_{sat} and different viscosities and similar viscosities with different C_{sat}, highlighting that thermodynamics and dynamics of condensates need not always be coupled.[51]

The purely viscous behavior of A-IDPs as opposed to viscoelastic behavior reported in peptide-nucleic acid condensates[48] highlights the importance of considering the sequence-rheology relationship in the context of the experimental system. Similarly, Gly-to-Ala substitutions had minimal effect on rheology of A-IDPs, contrary to FUS, where Gly-to-Ala mutations slow droplet fusion and cause hardening[21]. Additionally, the sequence-rheology relationship may be

time-dependent. Preliminary work shows that over 24 hours all A-IDP condensates in this work maintain spherical droplet morphology with no fibrilization or aggregation (**Extended Data Fig. 8a**). Furthermore, droplet coalescence events and microrheology at 24 hrs for three representative samples (GRGNSPYS, GRGDSPYS, and GQGNSPYS) demonstrate that the condensates retain their liquid-like nature at later timescales **(Extended Data Fig. 8b,c)** motivating future work to explore the sequence dependence of physical aging of condensates.

In conclusion, the insights provided by experiments and atomistic simulations on the relationship between sequence, phase behavior, and material properties are expected to significantly enhance the sequence-based prediction and design of biomolecular condensates.

Acknowledgements

This article is based on the research supported in part by the National Science Foundation (DMR-2004796 to KLK and JM), National Institute of General Medical Science of the National Institute of Health (R01GM136917 to JM and R35GM142903 to BSS), the Welch Foundation (A-2113-202203311 to JM). Use of the Texas A&M High Performance Research Computing is greatly acknowledged for the computational resources utilized in this work. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

SR, CGG, KLK and JM conceived the research. BSS, KLK and JM designed and supervised the research. SR and JM designed the sequences. CGG expressed and purified all polypeptides and performed turbidity experiments. MB performed microscopy, microrheology, and additional turbidity experiments. SR performed the simulations. SR analyzed the simulations aided by AR. SR, CGG, MB, BSS and JM wrote the manuscript with help from other authors.

Competing Interests

The authors declare no competing interests.

Figure Legends

Fig. 1 A diverse range of interactions between residue pairs contributes to phase separation of WT, $(GRGDSPYS)_{25}$. (a) Pie charts comparing the composition of the WT polypeptide to naturally occurring sequences. The composition is segregated into Polar; Aromatic; Anionic; Glycine and Proline; Cationic; and Aliphatic residues. Below the pie charts are cartoons highlighting the different residue pair interactions present in WT. (b) Example turbidimetry experiments on WT to estimate transition temperatures; shown here are triplicate turbidity assays at WT concentration 6 μM. (c) Partial phase diagram of WT obtained through turbidimetry at different WT concentrations in PBS. The dashed line is obtained through a logarithmic fit to the transition temperatures at each polypeptide concentration. Data are presented as mean values +/- SD, n=3 independent experiments. (d) Representative microscopy images of WT at 6 μM concentration at time points 10, 20, 30, and 60 minutes after inducing phase separation (Scale bar, 5 μm). The data presented are representative of multiple images acquired. (e) Images of WT droplets undergoing coalescence over a 100 ms time window (Scale bar, 5 μm). Similar results were obtained from at least two independent analyses (f) Microscopy images of WT, ARADSPYS, SRSDSPYS, and GRGDVPYS variants at 6 μM concentration (Scale bar, 5 μm). (g) Partial phase diagram of WT, ARADSPYS, SRSDSPYS, and GRGDVPYS variants in PBS obtained through turbidimetry at different concentrations. Dashed lines are obtained through a logarithmic fit to the measured transition temperatures at each concentration from turbidity experiments. Data are presented as mean values +/- SD, n=3 independent experiments.

Fig. 2 Atomistic simulations of the WT sequence highlight the diverse interactions encoded within the sequence. (a) Representative snapshot of the atomistic simulation of the condensed phase of WT. Proteins are shown as a semitransparent surface with bonds and atoms shown explicitly in blue; water in red; and Na^+ and Cl^- ions in green and yellow, respectively. (b) Estimated densities of system (protein, water, and ions combined), and of water, protein, sodium ions, and chloride ions from atomistic simulations of the WT condensed phase, consisting of 30 protein chains in the slab geometry. (c) Average residue pair contacts estimated from atomistic simulations of the WT condensed phase, separated into backbone-backbone (bb-bb), backbone-sidechain (bb-sc) and sidechain-sidechain (sc-sc). The inset shows the average contacts formed by each residue during the simulation. The bottom panel shows the average pairwise contacts formed by the residues when normalized by their abundance within the WT sequence, and the inset shows the individual residue contacts with the same normalization applied. (d) Comparison between residue pairwise contacts observed in a atomistic single-chain simulation of a fragment of the WT (6 repeat units), labeled as $N_{contact}^{dilute}$ and the condensed phase, labeled $N_{contact}^{dense}$. Coefficient of determination is shown as a measure of correlation. (e) Snapshot of a representative configuration of the chain in the atomistic single chain simulation. Frequently observed residue pairwise contacts within the simulation are shown surrounding the chain.

Fig. 3 Presence of arginine promotes but is not required for phase separation. (a) Partial phase diagrams for WT (GRGDSPYS), GQGNSPYS, GRGNSPYS, and GKGNSPYS in PBS. GQGDSPYS shows no measurable transition even up to a concentration of 80 μM and is thus omitted from the plot. Dashed lines are obtained through a logarithmic fit to the measured transition temperatures from turbidity experiments at different polypeptide concentrations. Data are presented as mean values +/- SD, n=3 independent experiments. (b) Microscopy images for the WT and the different variants at a concentration of 6 μM, shown in the box bounded by solid lines, and additionally at 60 μM for the GKGNSPYS variant in the box bounded by dashed lines. (Scale bar, 5 μm) (c) Saturation concentrations (C_{sat}) measured at 20 °C for the different variants. GQGDSPYS shows no measurable transition even up to a concentration of 80 μM. (d) Average residue pairwise contacts estimated through atomistic simulations for GQGNSPYS. Residue pairwise contacts are plotted with respect to WT. The diagonal indicates equal number of contacts

in WT and the variant. Residue pairs not involving the mutated residues are shown as gray circles, while residue pairs involving the mutated residues are shown in accordance with the color code for the mutations in the plot. The lower plot shows contact ratio ($P_{contact}^{variant}/P_{contact}^{WT}$) of residue pairs in the variant and the WT. The bars follow the same color code for mutations. (e,f) Contact ratio of residue pairs in the GQGDSPYS (e) and GRGNSPYS (f) variants and the WT. Contacts not involving the mutated residue are shown as gray bars, while contacts involving the mutated residue are shown in color. (g) Angle vs. Distance plots highlighting the frequency of occurrence of different configurations adopted between Tyr and Arg (left) and Lys (right) in atomistic simulations of the condensed phases of the GRGNPYS and GKGNSPYS variants. The snapshot above the respective plots shows the definition of the measured angle, θ, and distance, d. (h) Correlation plot similar to (d), comparing residue pair contacts between GRGNSPYS and GKGNSPYS variants.

Fig. 4 Aromatic residues promote but are not required for phase separation. (a) Average residue pairwise contacts estimated through atomistic simulations for GRGNSPAS with respect to the GRGNSPYS variant. Residue pairs not involving mutated residues are shown as gray circles while residue pairs involving the mutated residues (Y-to-A) are shown as purple triangles. The plot below shows the contact ratio of each of the residue pairwise contacts for the GRGNSPAS variant to that of the GRGNSPYS variant. (b) Partial phase diagrams for WT (GRGDSPYS), GRGNSPYS, and GRGNSPAS in PBS. Dashed lines in the left plot are obtained through a logarithmic fit to the measured transition temperatures at each concentration from turbidity experiments. Data are presented as mean values +/- SD, n=3 independent experiments. (c) Microscopy for WT, GRGNSPYS, and GRGNSPAS variants at a concentration of 6 μM, shown in the box bounded by solid lines, and additionally for 60 μM for GRGNSPAS, shown in the box bounded by dashed lines. (Scale bar, 5μm) (d) Partial phase diagrams for GRGNSPFS and GRGNSPWS variants in PBS, with WT and GRGNSPYS variants shown as a reference. Dashed lines in the left plot are obtained through a logarithmic fit to the measured transition temperatures at each concentration from turbidity experiments. Data are presented as mean values +/- SD, n=3 independent experiments. (e) Microscopy for the WT, GRGNSPYS, GRGNSPFS, and GRGNSPWS variants at a concentration of 6 μM, shown in the box bounded by solid lines, and additionally for 60 μM for GRGNSPFS, shown in the box bounded by dashed lines. (Scale bar, 5μm). Note: Phase diagrams and microscopy data for WT and GRGNSPYS are the same as shown in Fig. 2; they are repeated here for reference.

Fig. 5 Variants result in condensates with diverse material properties. (a) Fluorescence microscopy image of 0.5 μm yellow-green fluorescent polystyrene beads embedded in WT droplet (Scale bar, 5 μm). Inset: Representative trajectory from two-dimensional particle tracking showing Brownian motion of the beads (length of inset box represents 0.02 μm). Similar results were obtained from at least two independent analyses. Representative snapshots from Supplementary Movie 1. (b) Ensemble mean-squared displacement versus lag time for the variants tested in this study. (GQGDSPYS is not shown because it did not phase separate, and GRGNSPAS condensates were too small to analyze using our microrheology method.) (c) Viscosity of the variants, calculated from the particle tracking results after noise correction. Data are presented as mean values +/- SD, n=8 different videos from two independent trials (d) State diagram showing saturation concentration, C_{sat}, and viscosities for the variants tested in this study. C_{sat} is calculated at 18 °C. For viscosity measurements, total concentration differed for different variants. The symbols and colors on the plot are in accordance with the legend of subplot (b). Pearson correlation coefficient (R) and coefficient of determination (R^2) are shown as measures of correlation.

Fig. 6 A multitude of interactions work in tandem to drive LLPS. (a) Snapshot from atomistic simulations of WT, highlighting representative examples of the wide variety of contacts driving phase

separation of WT. Residues are represented by their three letter amino acid codes. (b) Effect of different mutations to C_{sat} at 37 °C, normalized by the number of mutations carried out (denoted by N). C_{sat} denotes the saturation concentration of the variant, whereas C_{ref} denotes the saturation concentration of the reference sequence used to calculate the effect of the mutation. Direct estimates refer to values for which the mutation was carried out in this work, while indirect estimates refer to values where the mutations were not carried out but can be estimated based on combining data from multiple investigated variants. Residues are colored based on the classification used in Fig.1a.

References

1 Feric, M. *et al.* Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **165**, 1686-1697 (2016).

2 Brangwynne, C. P. *et al.* Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729-1732 (2009).

3 Lyon, A. S., Peeples, W. B. & Rosen, M. K. A framework for understanding the functions of biomolecular condensates across scales. *Nature Reviews Molecular Cell Biology* **22**, 215-235 (2021).

4 Alberti, S. & Dormann, D. Liquid–liquid phase separation in disease. *Annual review of genetics* **53**, 171-194 (2019).

5 Alberti, S. & Hyman, A. A. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nature reviews Molecular cell biology* **22**, 196-213 (2021).

6 Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).

7 Martin, E. W. & Holehouse, A. S. Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof. *Emerging topics in life sciences* **4**, 307-329 (2020).

8 Schuster, B. S. *et al.* Biomolecular condensates: Sequence determinants of phase separation, microstructural organization, enzymatic activity, and material properties. *The journal of physical chemistry B* **125**, 3441-3451 (2021).

9 Borcherds, W., Bremer, A., Borgia, M. B. & Mittag, T. How do intrinsically disordered protein regions encode a driving force for liquid–liquid phase separation? *Current opinion in structural biology* **67**, 41-50 (2021).

10 Dignon, G. L., Best, R. B. & Mittal, J. Biomolecular phase separation: from molecular driving forces to macroscopic properties. *Annual review of physical chemistry* **71**, 53-75 (2020).

11 Quiroz, F. G. & Chilkoti, A. Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers. *Nature materials* **14**, 1164-1171 (2015).

12 Dzuricky, M., Rogers, B. A., Shahid, A., Cremer, P. S. & Chilkoti, A. De novo engineering of intracellular condensates using artificial disordered proteins. *Nature chemistry* **12**, 814-825 (2020).

13 Wang, B., Patkar, S. S. & Kiick, K. L. Application of Thermoresponsive Intrinsically Disordered Protein Polymers in Nanostructured and Microstructured Materials. *Macromolecular Bioscience* **21**, 2100129 (2021).

14 Garcia Garcia, C., Patkar, S. S., Jovic, N., Mittal, J. & Kiick, K. L. Alteration of microstructure in biopolymeric hydrogels via compositional modification of resilin-like polypeptides. *ACS Biomaterials Science & Engineering* **7**, 4244-4257 (2021).

15 Good, M. C., Zalatan, J. G. & Lim, W. A. Scaffold proteins: hubs for controlling the flow of cellular information. *Science* **332**, 680-686 (2011).

16 Schuster, B. S. *et al.* Controllable protein phase separation and modular recruitment to form responsive membraneless organelles. *Nature communications* **9**, 2985 (2018).

17 Simon, J. R., Eghtesadi, S. A., Dzuricky, M., You, L. & Chilkoti, A. Engineered ribonucleoprotein granules inhibit translation in protocells. *Molecular cell* **75**, 66-75. e65 (2019).

18 Nott, T. J. *et al.* Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular cell* **57**, 936-947 (2015).
19 Brady, J. P. *et al.* Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings of the National Academy of Sciences* **114**, E8194-E8203 (2017).
20 Vernon, R. M. *et al.* Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *elife* **7**, e31486 (2018).
21 Wang, J. *et al.* A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688-699. e616 (2018).
22 Murthy, A. C. *et al.* Molecular interactions underlying liquid− liquid phase separation of the FUS low-complexity domain. *Nature structural & molecular biology* **26**, 637-648 (2019).
23 Martin, E. W. *et al.* Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694-699 (2020).
24 Schuster, B. S. *et al.* Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. *Proceedings of the National Academy of Sciences* **117**, 11421-11431 (2020).
25 Murthy, A. C. *et al.* 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 (2021).
26 Bremer, A. *et al.* Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nature Chemistry* **14**, 196-207 (2022).
27 Martin, E. W. & Mittag, T. Relationship of sequence and phase separation in protein low-complexity regions. *Biochemistry* **57**, 2478-2487 (2018).
28 Tompa, P., Schad, E., Tantos, A. & Kalmar, L. Intrinsically disordered proteins: emerging interaction specialists. *Current opinion in structural biology* **35**, 49-59 (2015).
29 Lin, Y., Currie, S. L. & Rosen, M. K. Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *Journal of Biological Chemistry* **292**, 19110-19120 (2017).
30 Holehouse, A. S., Ginell, G. M., Griffith, D. & Boke, E. Clustering of Aromatic Residues in Prion-like Domains Can Tune the Formation, State, and Organization of Biomolecular Condensates: Published as part of the Biochemistry virtual special issue “Protein Condensates”. *Biochemistry* **60**, 3566-3581 (2021).
31 Fisher, R. S. & Elbaum-Garfinkle, S. Tunable multiphase dynamics of arginine and lysine liquid condensates. *Nature communications* **11**, 4628 (2020).
32 Devarajan, D. S. *et al.* Effect of Charge Distribution on the Dynamics of Polyampholytic Disordered Proteins. *Macromolecules* **55**, 8987-8997 (2022).
33 Rekhi, S. *et al.* Role of Strong Localized vs Weak Distributed Interactions in Disordered Protein Phase Separation. *The Journal of Physical Chemistry B* **127**, 3829-3838 (2023).
34 Dai, Y. *et al.* Programmable synthetic biomolecular condensates for cellular control. *Nature Chemical Biology* (2023). https://doi.org:10.1038/s41589-022-01252-8
35 Cai, H., Vernon, R. M. & Forman-Kay, J. D. An interpretable machine-learning algorithm to predict disordered protein phase separation based on biophysical interactions. *Biomolecules* **12**, 1131 (2022).

36 Hardenberg, M., Horvath, A., Ambrus, V., Fuxreiter, M. & Vendruscolo, M. Widespread occurrence of the droplet state of proteins in the human proteome. *Proceedings of the National Academy of Sciences* **117**, 33254-33262 (2020).
37 Bolognesi, B. *et al.* A concentration-dependent liquid phase separation can cause toxicity upon increased protein expression. *Cell reports* **16**, 222-231 (2016).
38 Mohanty, P. *et al.* Principles governing the phase separation of multidomain proteins. *Biochemistry* **61**, 2443-2455 (2022).
39 Zheng, W. *et al.* Molecular details of protein condensates probed by microsecond long atomistic simulations. *The Journal of Physical Chemistry B* **124**, 11671-11679 (2020).
40 Kim, S. *et al.* Complexation and coacervation of like-charged polyelectrolytes inspired by mussels. *Proceedings of the National Academy of Sciences* **113**, E847-E853 (2016).
41 Baruch Leshem, A. *et al.* Biomolecular condensates formed by designer minimalistic peptides. *Nature Communications* **14**, 421 (2023).
42 Van Treeck, B. *et al.* RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome. *Proceedings of the National Academy of Sciences* **115**, 2734-2739 (2018).
43 Paloni, M., Bailly, R., Ciandrini, L. & Barducci, A. Unraveling molecular interactions in liquid–liquid phase separation of disordered proteins by atomistic simulations. *The Journal of Physical Chemistry B* **124**, 9009-9016 (2020).
44 Armstrong, C. T., Mason, P. E., Anderson, J. R. & Dempsey, C. E. Arginine side chain interactions and the role of arginine as a gating charge carrier in voltage sensitive ion channels. *Scientific reports* **6**, 21759 (2016).
45 Lee, D., Lee, J. & Seok, C. What stabilizes close arginine pairing in proteins? *Physical Chemistry Chemical Physics* **15**, 5844-5853 (2013).
46 Harms, M. J., Schlessman, J. L., Sue, G. R. & García-Moreno E, B. Arginine residues at internal positions in a protein are always charged. *Proceedings of the National Academy of Sciences* **108**, 18954-18959 (2011).
47 Ray, S. *et al.* α-Synuclein aggregation nucleates through liquid–liquid phase separation. *Nature chemistry* **12**, 705-716 (2020).
48 Alshareedah, I., Moosa, M. M., Pham, M., Potoyan, D. A. & Banerjee, P. R. Programmable viscoelasticity in protein-RNA condensates with disordered sticker-spacer polypeptides. *Nature communications* **12**, 6620 (2021).
49 Jawerth, L. *et al.* Protein condensates as aging Maxwell fluids. *Science* **370**, 1317-1323 (2020).
50 Devarajan, D. S., Wang, J., Nikoubashman, A., Kim, Y. C. & Mittal, J. Sequence-Dependent Material Properties of Biomolecular Condensates. *bioRxiv*, 2023.2005. 2009.540038 (2023).
51 An, Y., Webb, M. A. & Jacobs, W. M. Active learning of the thermodynamics-dynamics trade-off in protein condensates. *Science Advances* **10**, eadj2448 (2024).
52 Urry, D. W. *et al.* Hydrophobicity scale for proteins based on inverse temperature transitions. *Biopolymers: Original Research on Biomolecules* **32**, 1243-1250 (1992).
53 Zeng, X. & Pappu, R. V. Developments in describing equilibrium phase transitions of multivalent associative macromolecules. *Current Opinion in Structural Biology* **79**, 102540 (2023).
54 Pappu, R. V., Cohen, S. R., Dar, F., Farag, M. & Kar, M. Phase transitions of associative biomacromolecules. *Chemical Reviews* (2023).

55 Dill, K. & Bromberg, S. *Molecular driving forces: statistical thermodynamics in biology, chemistry, physics, and nanoscience*. (Garland Science, 2010).
56 Yang, Y., Jones, H. B., Dao, T. P. & Castañeda, C. A. Single amino acid substitutions in stickers, but not spacers, substantially alter UBQLN2 phase transitions and dense phase material properties. *The Journal of Physical Chemistry B* **123**, 3618-3629 (2019).
57 König, B., Pezzotti, S., Ramos, S., Schwaab, G. & Havenith, M. Real-time measure of solvation free energy changes upon liquid-liquid phase separation of α-elastin. *Biophysical Journal*
58 Pezzotti, S., König, B., Ramos, S., Schwaab, G. & Havenith, M. Liquid–Liquid Phase Separation? Ask the Water! *The Journal of Physical Chemistry Letters* **14**, 1556-1563 (2023).
59 Wolfenden, R., Andersson, L., Cullis, P. & Southgate, C. Affinities of amino acid side chains for solvent water. *Biochemistry* **20**, 849-855 (1981).
60 Chu, X. *et al.* Prediction of liquid–liquid phase separating proteins using machine learning. *BMC bioinformatics* **23**, 1-13 (2022).
61 Saar, K. L. *et al.* Learning the molecular grammar of protein condensates from sequence determinants and embeddings. *Proceedings of the National Academy of Sciences* **118**, e2019053118 (2021).
62 Ng, S. C. & Görlich, D. A simple thermodynamic description of phase separation of Nup98 FG domains. *Nature Communications* **13**, 6172 (2022).
63 Gallivan, J. P. & Dougherty, D. A. Cation-π interactions in structural biology. *Proceedings of the National Academy of Sciences* **96**, 9459-9464 (1999).
64 Kumar, K. *et al.* Cation–π interactions in protein–ligand binding: Theory and data-mining reveal different roles for lysine and arginine. *Chemical science* **9**, 2655-2665 (2018).
65 Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
66 Flory, P. J. Thermodynamics of high polymer solutions. *The Journal of chemical physics* **10**, 51-61 (1942).
67 Lin, Y.-H., Song, J., Forman-Kay, J. D. & Chan, H. S. Random-phase-approximation theory for sequence-dependent, biologically functional liquid-liquid phase separation of intrinsically disordered proteins. *Journal of Molecular Liquids* **228**, 176-193 (2017).

Material and methods

Materials

The plasmid DNA encoding the different A-IDP sequences in pQE80L cloning vectors was purchased from Genscript Corporation (Piscataway, NJ). The amino acid sequences and corresponding DNA sequences of all constructs are provided in **Supplementary File 1**. Chemically competent cells of E. coli strain M15-[pREP4] (for transformation of recombinant plasmids) and RNAse (for protein purification) were purchased from Qiagen (Valencia, CA). All other chemicals were obtained from Sigma-Aldrich (St. Louis, MO) or Fisher Scientific (Waltham, MA) and were used as received unless otherwise noted.

Protein Expression and Purification

The E. coli M15-[pREP4] strain was transformed with each DNA plasmid (pQE80L cloning vectors) by heat shock to generate the expression cell stocks employed in protein production. Protein expression and purification were conducted as previously reported by our laboratories[70-73]. The complete protocol can be found in the Supplementary Methods section.

General Characterization of A-IDPs

The purity of samples was assessed by SDS-PAGE using NuPAGE 4-12% Bis-Tris gels (Invitrogen) and stained using a Coomassie stain (GelCode Blue Safe Protein Stain; Invitrogen). For this purpose, the lyophilized samples were dissolved in PBS buffer at a concentration of 1 mg/mL, via sonication for 2 min with a 10 s recovery and diluted with 8M urea buffer (pH 8.0) to a final polypeptide concentration of 0.5 mg/mL (**Supplementary Figure 12**). Amino acid analysis was performed by the Molecular Structure Facility at the University of California, Davis (Davis, CA) using a Hitachi L-800 sodium citrate-based amino acid analyzer (Tokyo, Japan) to determine the composition of each polypeptide (**Supplementary Table 1, Table 2**). The purity and molecular

weight of the dialyzed polypeptides (before lyophilization) was confirmed via UPLC and electrospray ionization mass spectrometry (ESI-MS) (Waters Xevo G2-S Q-TOF MS with Acquity UPLC, Milford, MA) (**Supplementary Figure 13, 14**). The A260/A280 ratios were measured in a UV-Vis spectrophotometer (Cary 60 Bio; Agilent), and the resulting ratios for all sequences range between 0.55 – 0.65, which suggests the absence of any remaining nucleic acid.

Turbidity Assays

a) Kiick Laboratory

Temperature-dependent turbidity assays were conducted in a UV-Vis spectrophotometer (Cary 60 Bio; Agilent) equipped with a Peltier temperature controller. Protein samples were tested in low-volume quartz cuvettes with 1 cm path length (Hellma Analytics). Lyophilized samples were dissolved in PBS (8 mM Na_2HPO_4, 137 mM NaCl, 3mM KCl, 2mM KH_2PO_4, pH 7.4) at double the desired concentration and were sonicated for 2 min with a 10 s recovery time, allowing them to warm up during sonication and exceed the transition temperature. Then, the samples were filtered (0.45 μm, PVDF) at high temperature to eliminate any non-dissolved material, transferred to the quartz cuvette, and incubated at 80 °C in the spectrophotometer. The measured absorbance values (at λ = 280 nm) and the molar extinction coefficients (obtained from the ProtParam tool at web.expasy.org) were used to calculate the polypeptide concentration, and then this information was used to adjust the sample to the desired final concentration and to a volume of 500 μL in PBS. The samples were cooled from 80 °C to 5 °C at a rate of 1 °C min^{-1} and the absorbance was measured at λ = 350 nm every 1 °C throughout the temperature ramp. The transition temperature was defined as the point where absorbance first exceeds 0.03. The average transition temperature (n=3) was used to build the partial phase diagrams corresponding to each sequence.

b) Schuster Laboratory

For verification purposes, temperature-dependent turbidity assays were also conducted in the Schuster Lab in a UV-Vis spectrophotometer (Cary 3500; Agilent) equipped with a multicell Peltier temperature controller. Protein samples were tested in quartz cuvettes with 1 cm path length (ThorLabs) or low-volume quartz cuvettes with 1 cm path length (Hellma Analytics) and sample preparation was conducted as described above in part (a). The samples were filtered using 0.45 μm, PES filters.

Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy (Jasco J-1500 CD spectropolarimeter, Jasco Inc., Easton, MD, USA) was conducted to characterize the secondary structure of the A-IDP sequences. The resulting plots show the average of n=3 samples. A description of the sample preparation and the resulting spectra can be found in **Supplementary Figure 15**. The structural comparison of each sequence can be found on **Supplementary Table 3**. Additionally, the variants used in this work were run through the AmylPred2 consensus aggregation predictor[74], and no aggregation hotspots were identified.

Microscopy: Phase Behavior and Droplet Fusion

Each polypeptide sample was prepared at the desired final concentration by diluting the stock solution (prepared as described in previous section) with PBS (pH 7.4). Prior to imaging, samples were kept at 70 °C in a heat block (Fisher Scientific) in the microscope room, so that there was no droplet assembly before the start of the experiment. The dishes used for microscopy were 16-well glass-bottom dishes (#1.5 glass thickness; Grace Bio-Labs) that were pretreated with 5% Pluronic F-127 (Sigma-Aldrich) for a minimum of 45 minutes. The coated wells were washed with PBS (pH 7.4), and then 100 μL of the protein sample was removed from the heat block and transferred to the imaging well at room temperature to initiate droplet assembly.

Imaging of droplet assembly at different time points after the plating (considered t = 0) was performed on a Zeiss Axio Observer 7 inverted microscope equipped with an Axiocam 702 monochrome sCMOS camera (Zeiss), employing a 63x/1.4 NA plan-apochromatic oil-immersion objective and using differential interference contrast (DIC) transillumination. The same Zeiss microscope and experimental conditions were used for observing droplet fusion, with imaging conducted at a frame rate of approximately 200 Hz. Videos of droplet fusion events were analyzed using MATLAB. All microscopy experiments were conducted at the ambient temperature (17-20 °C).

Video Particle-Tracking Microrheology (VPT)

500 nm diameter yellow-green carboxylate-modified polystyrene beads (FluoSpheres, Invitrogen) were used for VPT microrheology measurements. Each polypeptide sample was prepared at the desired final concentration for microrheology by diluting the stock solution (prepared as described in previous sections) with PBS (pH 7.4). To employ the microrheology technique for all the mutants, which have a wide range of saturation concentrations, we used higher polypeptide concentrations for the variants with elevated saturation concentrations (GKGNSPYS, GRGNSPFS, and GRGASPYA). Since viscosity is an intrinsic material property of the condensate, it should not change with total polypeptide concentration. We verified this experimentally for two polypeptide sequences (**Extended Data Fig. 5b**). This permits us to compare the viscosities of the variants, even though polypeptide concentration varied for the different samples.

Microrheology experiments were prepared by mixing the 200 μL polypeptide sample with the fluorescent tracer beads before initiating droplet assembly in a 96-well plate (#1.5 high-performance cover glass, Cellvis). The samples were incubated at room temperature for 45 min

and then were observed under the microscope to verify that the tracer beads were embedded in the condensates (**see Fig 5a, main text**). Next, the samples in the well plate were centrifuged at 300xg for 1 minute to form a condensate layer or larger-size droplets (>30 μm in diameter); the purpose of this step was to avoid boundary effects and prevent flow of the condensates. To visualize and verify that the beads are embedded in the condensate and away from the condensate interface, we conducted separate verification experiments where we mixed Rhodamine B, a red fluorescent dye, with the protein sample and tracer beads following a similar protocol for microrheology sample preparation. Rhodamine B partitions into the condensate, so the condensate can be visualized as red, particles are green, and the dilute phase and glass appear black. Z-stacks are obtained to visualize the spatial distribution of beads in three dimensions **(Supplementary Figure 16)**. The microscope focus is adjusted to the midsection of the protein sample for particle tracking video acquisition. We note that rhodamine B was only added to the samples for verification experiments and was not present when collecting the microrheology data.

Epifluorescence video imaging was initiated at the 1 hr timepoint using the same microscope and procedure as described in section 4.6, with fluorescence excitation using a 475 nm LED (Colibri 7; Zeiss). Videos of the tracer beads diffusing within the condensate were collected at 200 frames per second for 2000 frames. Imaging was conducted at room temperature (17-20 °C). For each A-IDP variant, two independent samples were made on different days, and 3-5 videos were collected from each sample, with each video containing ~10-50 tracer beads. Viscosity data presented in Fig. 5c and 5d is the average of these two independent trials.

The TrackPy particle tracking code (see Software section) was used to analyze the collected videos, starting with extracting particle trajectories. The mean squared displacement (MSD) was calculated from the trajectories of individual beads, followed by calculating the ensemble-average

MSD (n=4-8 for data shown in Fig. 5b). To remove the static error from the MSD curves for calculating viscosities, we corrected the ensemble-average MSD by subtracting the noise floor from the MSD curves. In general, the ensemble-average MSD often scales as a power law with lag time τ, as given by the following equation:

$$MSD(\tau) = 2dD\tau^{\alpha}$$

where d is the number of dimensions (here d=2, since data collection and analysis were conducted in the x-y plane), D is the diffusion coefficient, and α is the diffusivity exponent. For a purely viscous fluid, the diffusivity exponent α is close to unity. α values for all the condensates tested were in the range of 0.87-1.07. Assuming a purely viscous fluid, with the system at equilibrium, the condensate viscosity η is then calculated using the Stokes-Einstein equation:

$$D = \frac{k_B T}{6\pi\eta R}$$

where k_B is the Boltzmann constant, T is the temperature (Kelvin), and R is the tracer bead radius. Reported viscosities (**Supplementary Table 4**) are averages ± SD across multiple videos (n=8). Pearson's correlation coefficient is calculated using built in Excel function *correl*.

The noise floor of 500 nm beads was measured by adhering the beads to the glass surface of a 96-well plate. We acquired the trajectories of the beads adhered to the glass surface using the same parameters as those used for VPT studies of the polypeptide samples. We find that the particles adhered on the glass surface showed negligible change in MSD over lag time, with MSD measured at approximately 1.02 x 10^{-6} μm^2 (Fig. 5b, n = 4 different videos). To assess the potential impact of particle size and surface chemistry on microrheology results, we performed microrheology on three representative sequences (GRGNSPYS, GQGNSPYS, and GRGDSPYS) with 0.5 μm vs. 1 μm diameter beads and PEGylated vs. carboxylated beads, respectively. The

results show no significant change in the viscosities obtained, validating the suitability of the probes used in this study (**Extended Data Fig. 5c,d,e**).

Software

VPT data analysis was conducted using the open-source particle tracking package TrackPy (v0.5.0)[75] in Python and customized as needed. Custom MATLAB code[24] was used to analyze droplet fusion events. Fiji (version 1.53s) was also used for image processing. Cary WinUV was used for turbidity data collection. Spectra Manager was used to collect Circular Dichroism (CD) data. MS Excel was used for turbidity and CD data analysis.

All-Atom Molecular Dynamics simulations

Generating and running the atomistic WT slab

The initial configuration for the all-atom slab was generated from a Coarse-Grained (CG) simulation using the steps detailed in prior work[39]. The initial CG simulation consisted of 30 chains of the WT peptide using the HPS-Urry model[76], run for 1 microsecond. The final configuration of the CG simulation was backmapped into an atomistic representation using MODELLER[77]. Following this, the atomistic slab was minimized using CAMPARI[78] with the implicit solvent ABSINTH[79] force field. The force field used for production runs was Amber99SBws-STQ[80]. This minimized structure was then solvated with TIP4P/2005[81] water, along with scaled salt interactions[82], in a box of size 10x10x40 nm^3. Following solvation, the system was minimized using GROMACS-2019.4[83] using the steepest descent algorithm. After minimization, 100 mM of NaCl was added to the system in excess of the amount of ions needed to maintain electroneutrality. The solvated polypeptide with ions was then minimized once again using the same parameters as prior minimization and then 100 ns of NVT was run with all bonds constrained using the Nose-Hoover thermostat[84] with a coupling constant of 1 ps for protein, water, and ions and temperature

fixed at 300 K. Following the NVT equilibration, 100 ns of NPT equilibration was run using the Berendsen barostat[85] with isotropic coupling and a constant of 5 ps for pressure control and the Nose-Hoover thermostat with same parameters as for the NVT run. Following the equilibration steps, a 1 microsecond production run in the NPT ensemble using the Langevin Middle Integrator[86] and the Monte Carlo Membrane Barostat in OpenMM-7.7[87] was run. Short-range nonbonded interactions were calculated with a cutoff radius of 0.9 nm, while long-range electrostatics were treated with the PME method[88]. Hydrogen mass was increased by 1.5 times, allowing for a timestep of 4 fs, while hydrogen containing bonds were constrained using the SHAKE algorithm[89]. A friction coefficient of 1 ps^{-1} was used for the Langevin Middle integrator. As the system is periodic in the x and y dimensions, the semi-isotropic membrane barostat was used. The barostat was set to be isotropic in x and y dimensions to ensure both dimensions with periodicity are scaled uniformly, while z dimension was set to be free such that box size can fluctuate independent of x and y dimensions. An advantage of using the membrane barostat is that box size predominantly changes in the z dimension thereby ensuring preservation of the initial slab geometry. Pressure was fixed at 1 bar, temperature at 300K, default surface tension of 200 bar·nm, and a frequency of 1000 time steps for attempted volume changes.

Generation of mutated slabs

A mutant slab was generated from a parent slab using the following steps. The final frame of the parent slab after the production run was separated into two PDB files, one of protein and the other containing water and ions. The protein-only PDB was passed to UCSF Chimera,[90] where the mutations were carried out. The mutated PDB was then saved and combined with the water+ions PDB from the parent slab. Following the combination of the two PDBs, the system was minimized and equilibrated using the same steps and parameters as the WT detailed above, before a

production run of 1 microsecond with the same parameters as WT was run in OpenMM-7.7. In some cases, where the mutation being carried out involves a significant change in the size of the side chain, e.g. the Y to W mutation, a soft-core minimization with lambda set to 0.01 and alpha set to 4 was carried out immediately after combination of the PDBs so as toto remove any clashes resulting from protein and water or ion atoms overlapping in the initial PDB. In cases where net charge of the peptide was non-zero, counterions were added to maintain neutrality prior to minimization and equilibration steps. Following this soft-core minimization, the steps remained the same. A flowchart indicating the sequential mutations carried out to generate the set of variants is available in the Supporting Information (**Supplementary Figure 17**).

Analysis of all-atom simulation trajectories

All analysis shown in the text is calculated after 250 ns of simulation time. The equilibration time of 250ns was estimated using the autocorrelation of the radius of gyration of the chains in the system calculated for the WT (**Supplementary Figure 18**). Density profiles were calculated using Gromacs-2019.4. To verify the partitioning of salt ions in the condensate, we computed radial distribution functions of charged residues and ions. We find that Asp residues show a preference for interaction with Na^+ ions, however Arg shows preferential interactions with other protein residues in addition to Cl^- ions (**Supplementary Figure 5**). Analysis of the salt concentration profiles suggests that these enhanced interactions between Asp and Na^+ are not sufficient to induce a significant difference in the concentration of cations and anions within the condensate. Residue pairwise contacts were calculated using in-house codes written in Python-3.9.12 using the trajectory analysis suite MDAnalysis-2.0.0[91,92]. Any two heavy atoms within 6 Å of each other are considered to be in contact. Residue pairwise contacts were estimated as a sum of number of heavy atoms pairs within the two residues that satisfy the above distance criteria.

Inter and intrachain contacts are calculated together. Details regarding the normalization to generate residue pairwise contacts can be found in work by Zheng et. al.[39] To rule out the possibility that the distance-based contact definition is overcounting contact pairs, we estimate the energetic contribution of different contact pairs using the pairwise option in AmberTools21[93] We find that residue pair contacts apart from those between oppositely charged residues), i.e., involving Gly, Ser, and Tyr, are also energetically favorable (**Supplementary Figure 19**). We further calculated the contact lifetimes of residue pairs within the protein condensate.[94] We find that within the condensed phase, contact association and dissociation events are highly heterogenous, with some residue pairwise contacts forming and breaking dynamically, while others either form and do not undergo dissociation, or rapidly dissociate and do not reform, within the time scale accessible to the simulations performed **(Supplementary Figure 20)**. . All snapshots were generated using VMD-1.9.3[95] and UCSF ChimeraX[96].

Data Availability

Source data are provided with the paper. Starting and ending configurations for all atomistic simulations have been deposited at Zenodo (https://doi.org/10.5281/zenodo.10523201).

Materials are available upon reasonable request to the corresponding authors.

Code Availability

Codes to run and analyze atomistic simulations are available publicly and can be found at https://openmm.org/, https://gromacs.org/ and https://www.mdanalysis.org/.

Codes to reproduce residue pairwise contacts and angle vs. distance distributions have been deposited on Zenodo (https://doi.org/10.5281/zenodo.10523201).

References

68 Li, W. *et al.* Biophysical properties of AKAP95 protein condensates regulate splicing and tumorigenesis. *Nature cell biology* **22**, 960-972 (2020).

69 Conicella, A. E. *et al.* TDP-43 α-helical structure tunes liquid–liquid phase separation and function. *Proceedings of the National Academy of Sciences* **117**, 5883-5894 (2020).

70 Charati, M. B., Ifkovits, J. L., Burdick, J. A., Linhardt, J. G. & Kiick, K. L. Hydrophilic elastomeric biomaterials based on resilin-like polypeptides. *Soft matter* **5**, 3412-3416 (2009).

71 Li, L., Teller, S., Clifton, R. J., Jia, X. & Kiick, K. L. Tunable mechanical stability and deformation response of a resilin-based elastomer. *Biomacromolecules* **12**, 2302-2310 (2011).

72 Li, L., Tong, Z., Jia, X. & Kiick, K. L. Resilin-like polypeptide hydrogels engineered for versatile biological function. *Soft Matter* **9**, 665-673 (2013).

73 Li, L. & Kiick, K. L. Transient dynamic mechanical properties of resilin-based elastomeric hydrogels. *Frontiers in Chemistry* **2**, 21 (2014).

74 Tsolis, A. C., Papandreou, N. C., Iconomidou, V. A. & Hamodrakas, S. J. A consensus method for the prediction of 'aggregation-prone'peptides in globular proteins. *PloS one* **8**, e54175 (2013).

75 Allan, D. B. a. C., Thomas and Keim, Nathan C. and van der Wel, Casper M. and Verweij, Ruben W. *soft-matter/trackpy: Trackpy v0.5.0*, <https://doi.org/10.5281/zenodo.4682814> (2021).

76 Regy, R. M., Thompson, J., Kim, Y. C. & Mittal, J. Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins. *Protein Science* **30**, 1371-1379 (2021).

77 Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Current protocols in bioinformatics* **54**, 5.6. 1-5.6. 37 (2016).

78 Vitalis, A. & Pappu, R. V. Methods for Monte Carlo simulations of biomacromolecules. *Annual reports in computational chemistry* **5**, 49-76 (2009).

79 Vitalis, A. & Pappu, R. V. ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. *Journal of computational chemistry* **30**, 673-699 (2009).

80 Tang, W. S., Fawzi, N. L. & Mittal, J. Refining all-Atom protein force fields for polar-Rich, prion-like, low-Complexity intrinsically disordered proteins. *The Journal of Physical Chemistry B* **124**, 9505-9512 (2020).

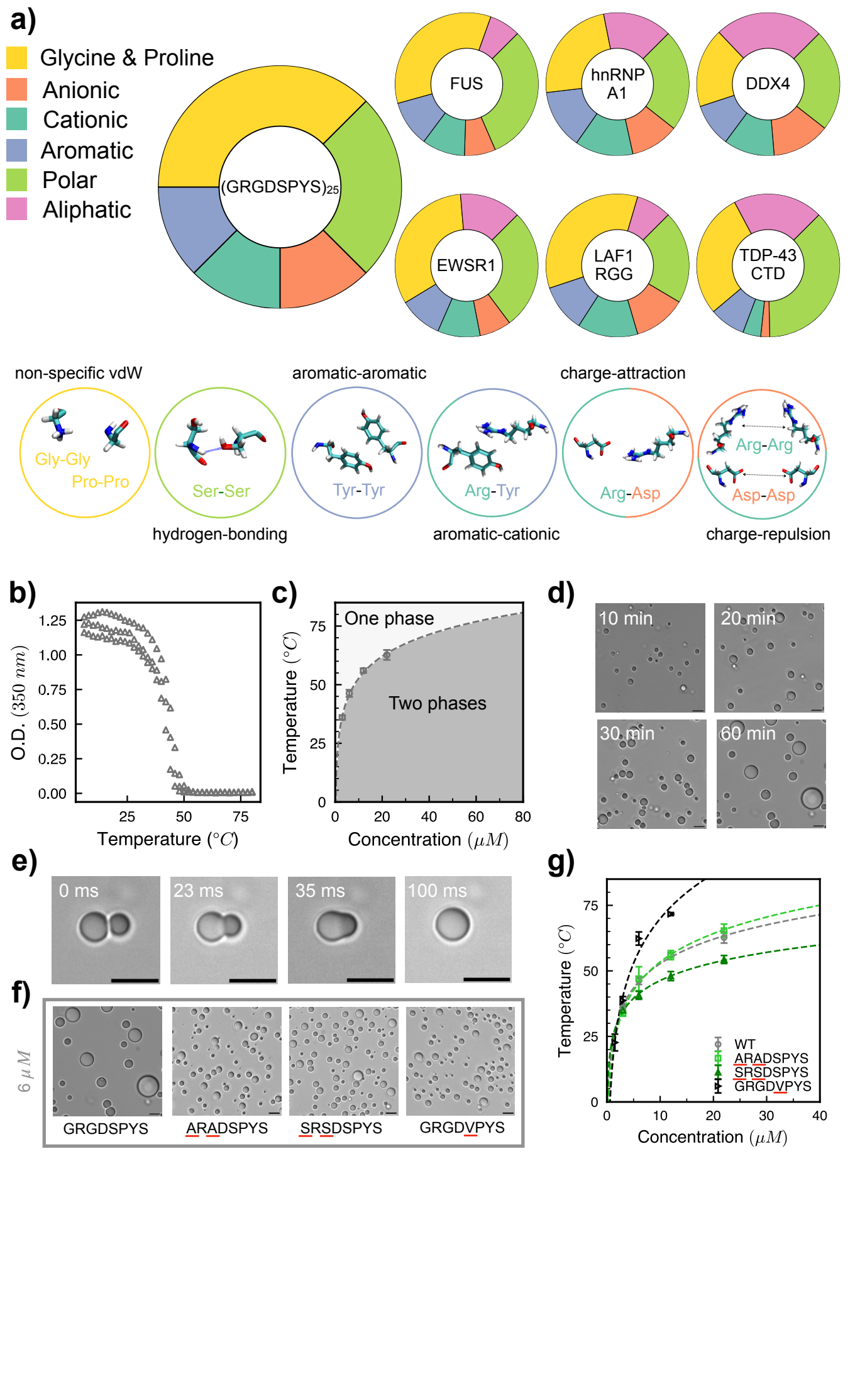
81 Abascal, J. L. & Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *The Journal of chemical physics* **123**, 234505 (2005).

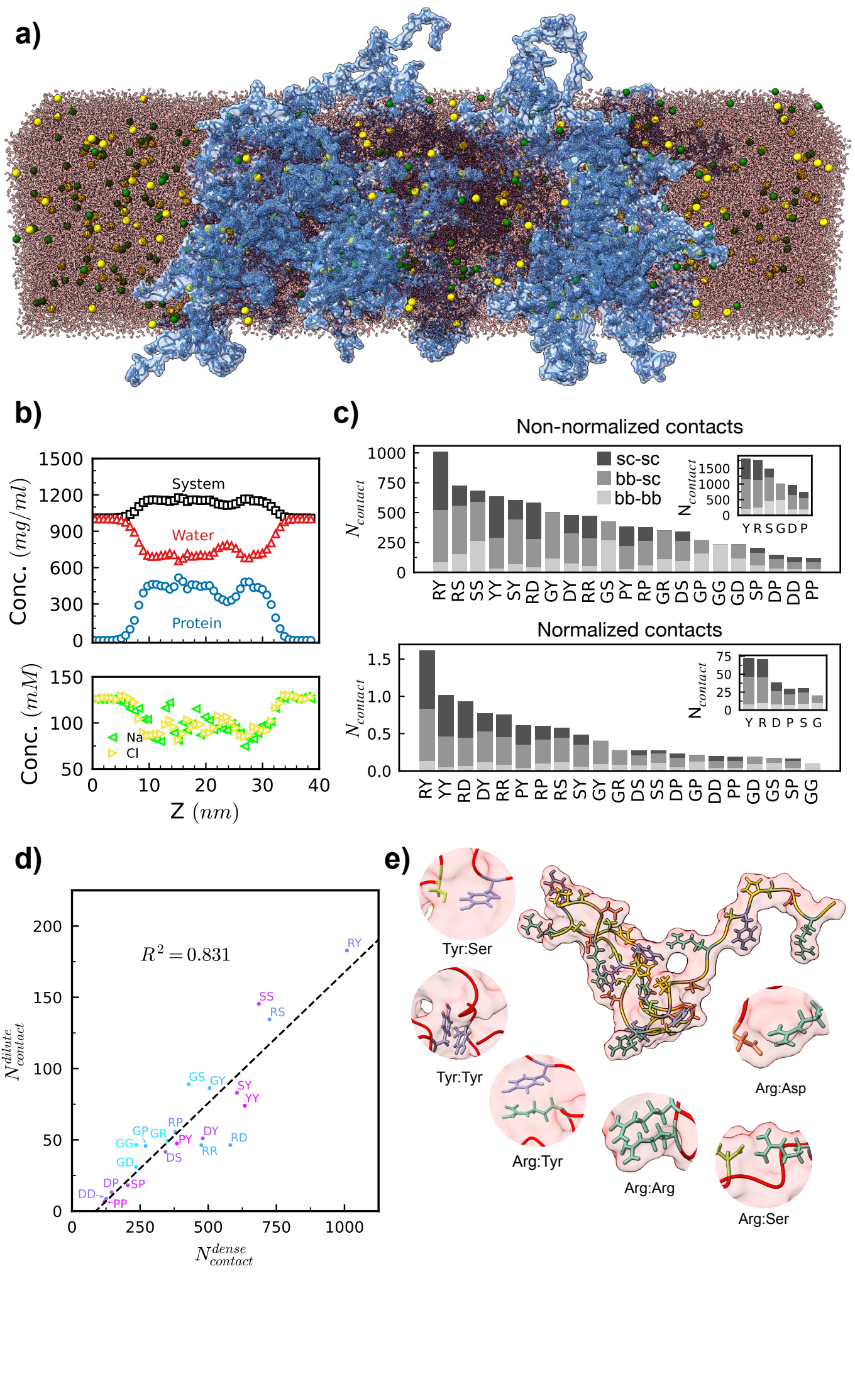
82 Luo, Y. & Roux, B. Simulation of osmotic pressure in concentrated aqueous salt solutions. *The journal of physical chemistry letters* **1**, 183-189 (2010).

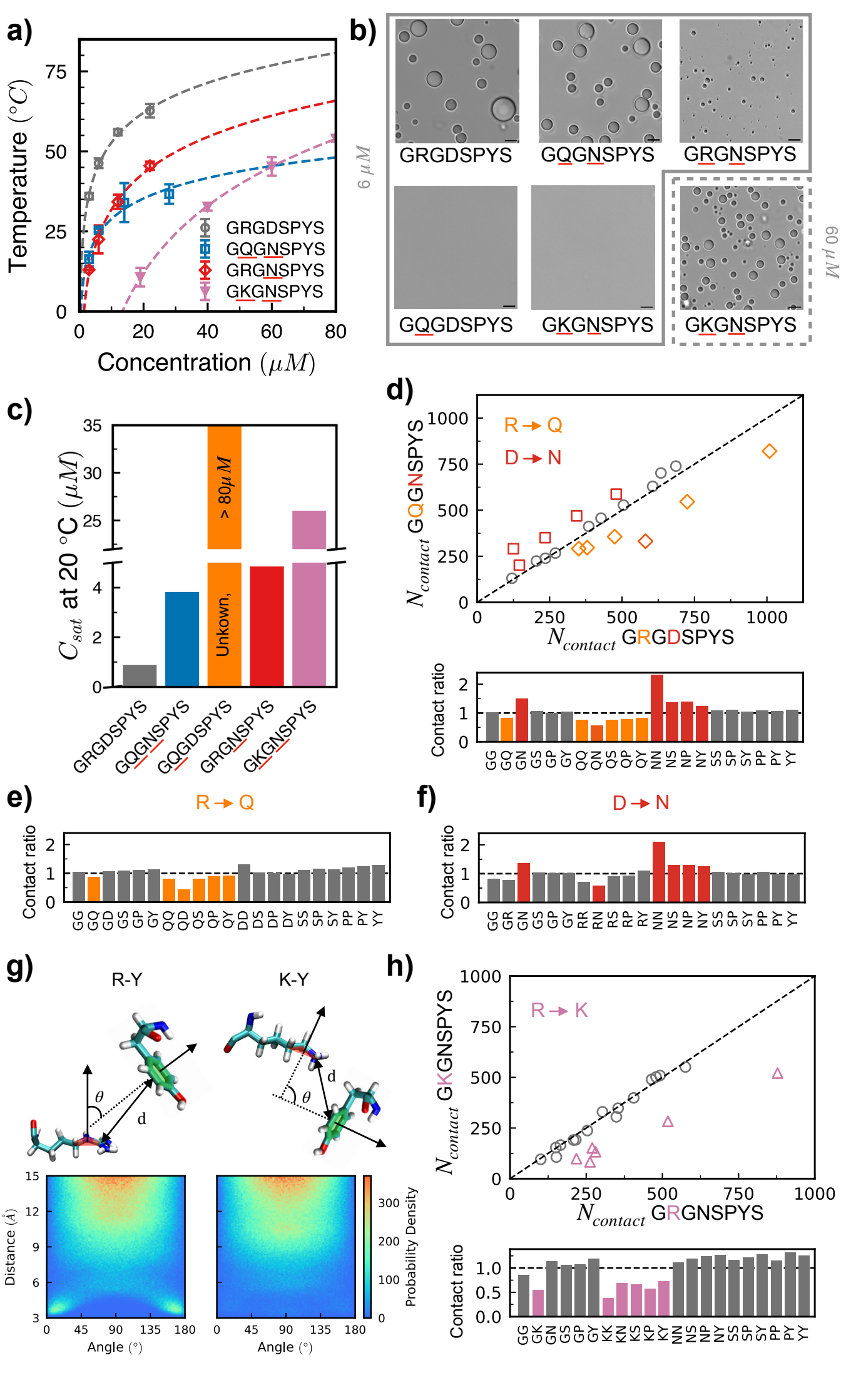
83 Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19-25 (2015).

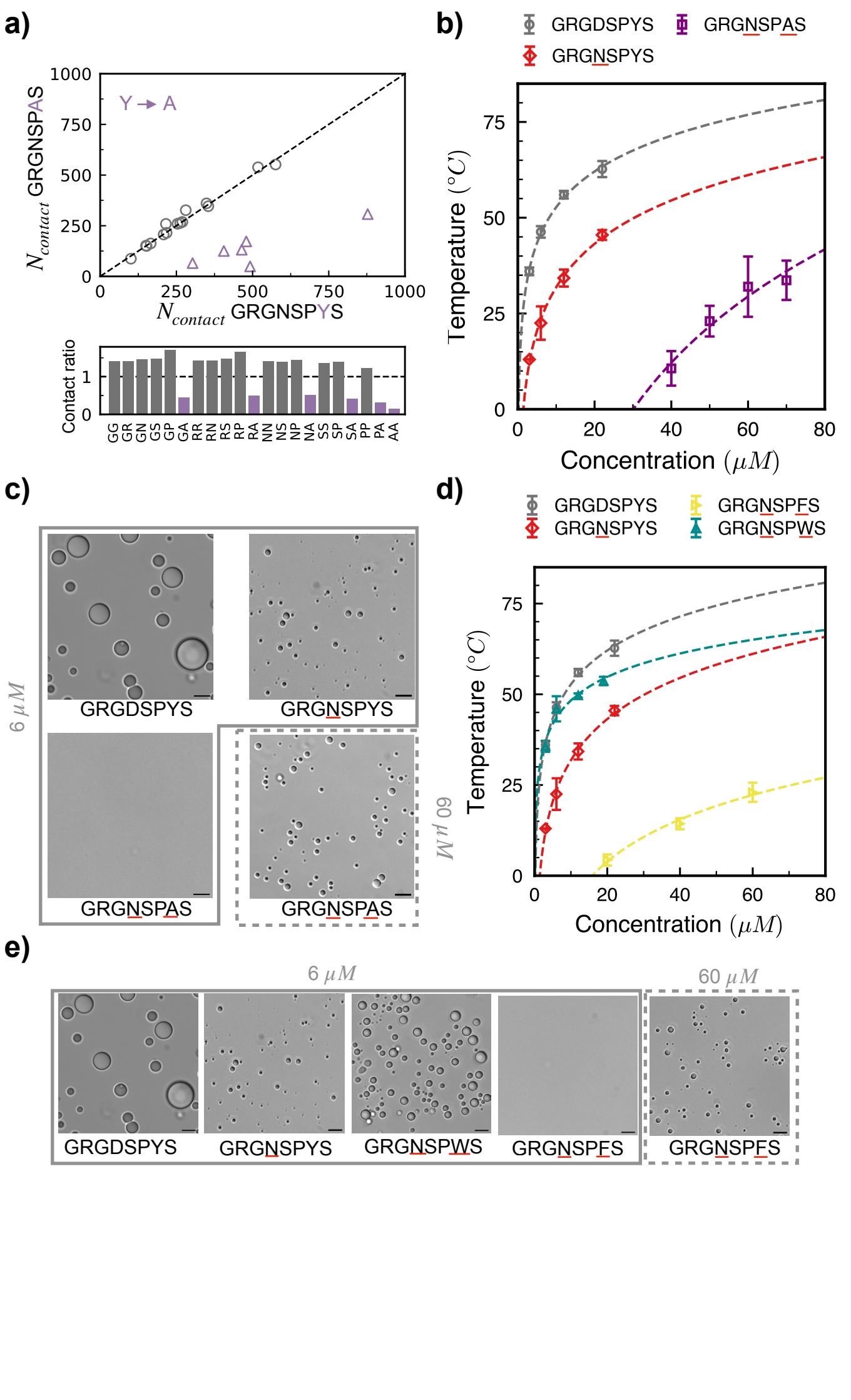
84 Evans, D. J. & Holian, B. L. The nose–hoover thermostat. *The Journal of chemical physics* **83**, 4069-4074 (1985).

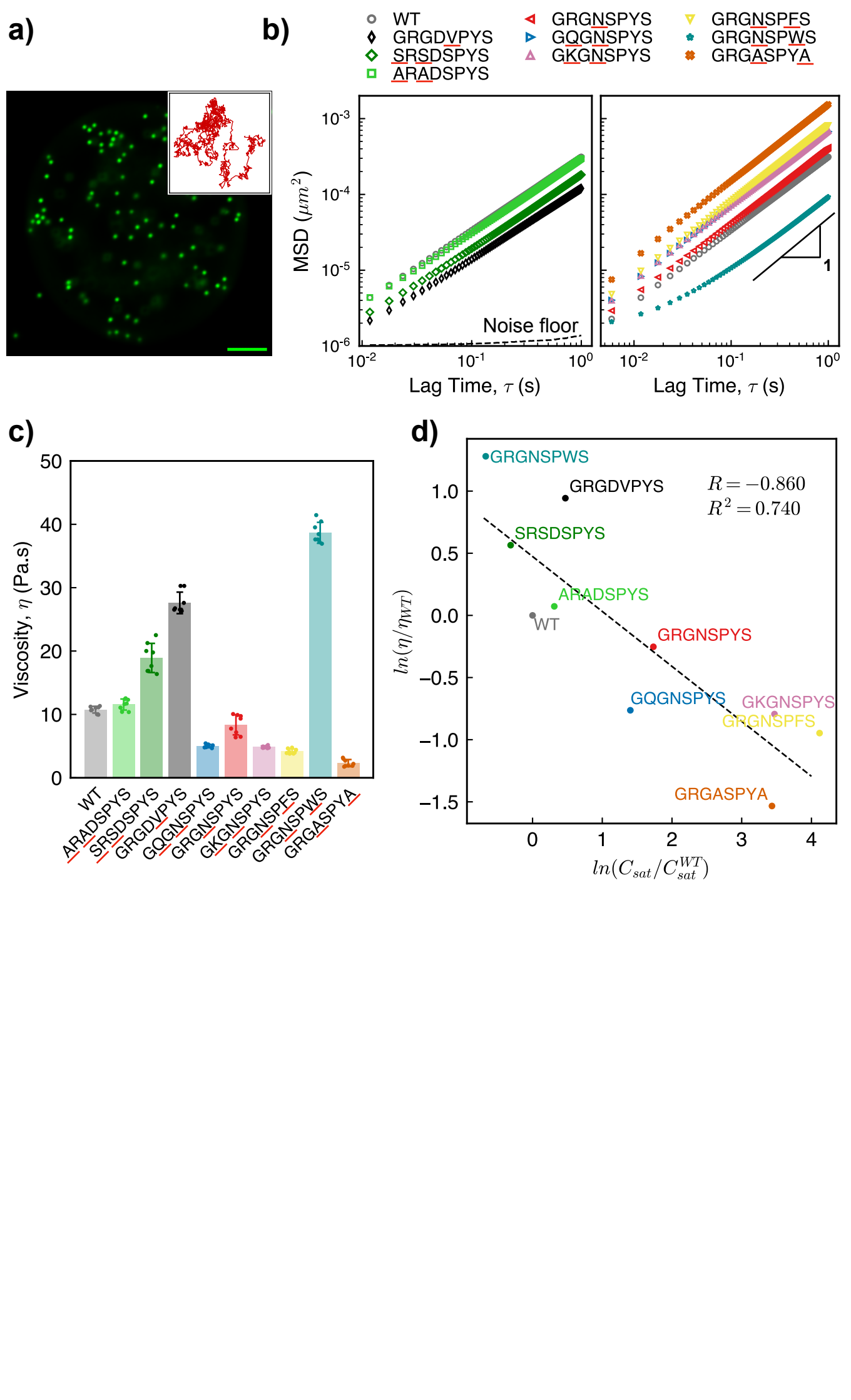
85 Berendsen, H. J., Postma, J. v., Van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **81**, 3684-3690 (1984).
86 Zhang, Z., Liu, X., Yan, K., Tuckerman, M. E. & Liu, J. Unified efficient thermostat scheme for the canonical ensemble with holonomic or isokinetic constraints via molecular dynamics. *The Journal of Physical Chemistry A* **123**, 6056-6079 (2019).
87 Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology* **13**, e1005659 (2017).
88 Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *The Journal of chemical physics* **98**, 10089-10092 (1993).
89 Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics* **23**, 327-341 (1977).
90 Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605-1612 (2004).
91 Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry* **32**, 2319-2327 (2011).
92 Gowers, R. J. *et al.* in *Proceedings of the 15th python in science conference.* 105 (SciPy Austin, TX).
93 D.A. Case, H. M. A., K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. *et al.* Amber 2021. *University of California, San Francisco* (2021).
94 Galvanetto, N. *et al.* Extreme dynamics in a biomolecular condensate. *Nature*, 1-8 (2023).
95 Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *Journal of molecular graphics* **14**, 33-38 (1996).
96 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30**, 70-82 (2021).

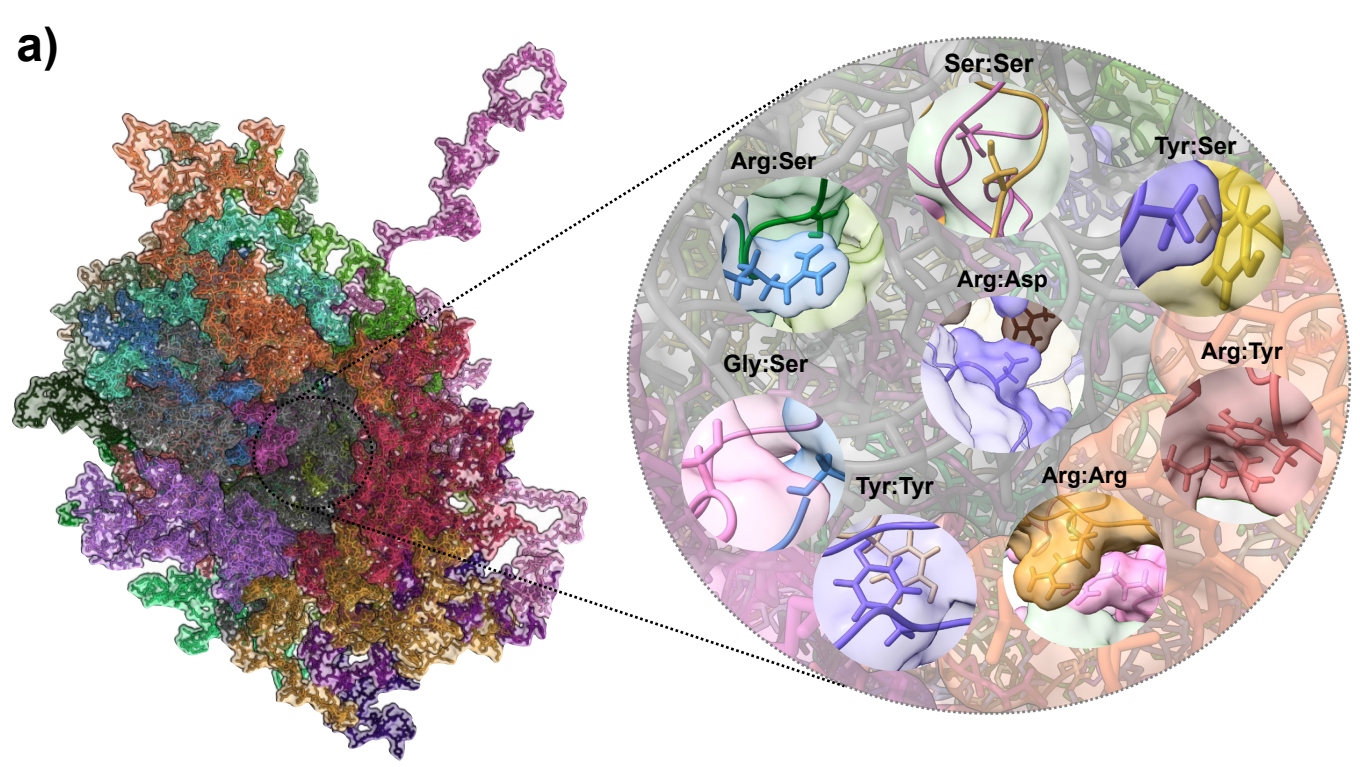

a)
Glycine & Proline
Anionic
Cationic
Aromatic
Polar
Aliphatic
(GRGDSPYS)25
FUS
hnRNP A1
DDX4
EWSR1
LAF1 RGG
TDP-43 CTD
non-specific vdW
Gly-Gly
Pro-Pro
hydrogen-bonding
Ser-Ser
aromatic-aromatic
Tyr-Tyr
aromatic-cationic
Arg-Tyr
charge-attraction
Arg-Asp
charge-repulsion
Arg-Arg
Asp-Asp
b)
O.D. (350 nm)
Temperature (°C)
c)
One phase
Two phases
Temperature (°C)
Concentration (μM)
d)
10 min
20 min
30 min
60 min
e)
0 ms
23 ms
35 ms
100 ms
f)
6 μM
GRGDSPYS
ARADSPYS
SRSDSPYS
GRGDVPYS
g)
Temperature (°C)
Concentration (μM)
WT
ARADSPYS
SRSDSPYS
GRGDVPYS

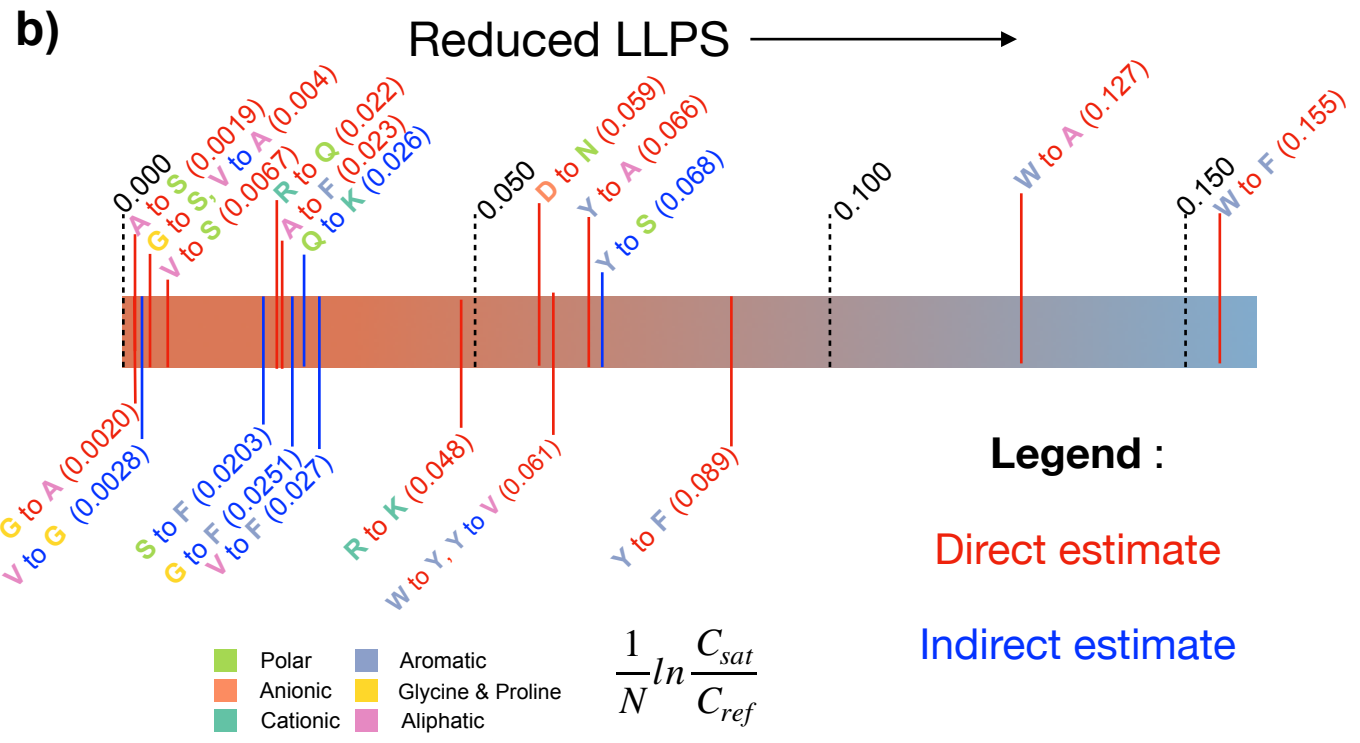

a)
b)
Conc. (mg/ml)
System
Water
Protein
Conc. (mM)
Na
Cl
Z (nm)
c)
Non-normalized contacts
$N_{contact}$
sc-sc
bb-sc
bb-bb
RY RS SS YY SY RD GY DY RR GS PY RP GR DS GP GG GD SP DP DD PP
Y R S G D P
Normalized contacts
RY YY RD DY RR PY RP RS SY GY GR DS SS DP GP DD PP GD GS SP GG
Y R D P S G
d)
$R^2 = 0.831$
$N_{contact}^{dilute}$
$N_{contact}^{dense}$
RY SS RS GS GY SY YY RP GP GR PY DY GG RD RR DS GD DP SP DD PP
e)
Tyr:Ser
Tyr:Tyr
Arg:Tyr
Arg:Arg
Arg:Asp
Arg:Ser


a)
Temperature (°C)
Concentration (μM)
GRGDSPYS
GQGNSPYS
GRGNSPYS
GKGNSPYS
b)
6 μM
60 μM
GRGDSPYS
GQGNSPYS
GRGNSPYS
GQGDSPYS
GKGNSPYS
GKGNSPYS
c)
C_{sat} at 20 °C (μM)
> 80μM
Unkown,
GRGDSPYS
GQGNSPYS
GQGDSPYS
GRGNSPYS
GKGNSPYS
d)
$N_{contact}$ GQGNSPYS
$N_{contact}$ GRGDSPYS
R → Q
D → N
Contact ratio
e)
R → Q
Contact ratio
f)
D → N
Contact ratio
g)
R-Y
K-Y
θ
d
Distance (Å)
Angle (°)
Probability Density
h)
$N_{contact}$ GKGNSPYS
$N_{contact}$ GRGNSPYS
R → K
Contact ratio


a)
1000
750
500
250
0
Y → A
$N_{contact}$ GRGNSPAS
0
250
500
750
1000
$N_{contact}$ GRGNSPYS
Contact ratio
1
0
GG GR GN GS GP GA RR RN RS RP RA NN NS NP NA SS SP SA PP PA AA
b)
GRGDSPYS
GRGNSPAS
GRGNSPYS
Temperature (°C)
75
50
25
0
0
20
40
60
80
Concentration (μM)
c)
6 μM
GRGDSPYS
GRGNSPYS
GRGNSPAS
GRGNSPAS
d)
GRGDSPYS
GRGNSPFS
GRGNSPYS
GRGNSPWS
60 μM
Temperature (°C)
75
50
25
0
0
20
40
60
80
Concentration (μM)
e)
6 μM
60 μM
GRGDSPYS
GRGNSPYS
GRGNSPWS
GRGNSPFS
GRGNSPFS


a)
b)
WT
GRGDVPYS
SRSDSPYS
ARADSPYS
GRGNSPYS
GQGNSPYS
GKGNSPYS
GRGNSPFS
GRGNSPWS
GRGASPYA
MSD (μm^2)
Noise floor
Lag Time, τ (s)
Lag Time, τ (s)
1
c)
Viscosity, η (Pa.s)
WT
ARADSPYS
SRSDSPYS
GRGDVPYS
GQGNSPYS
GRGNSPYS
GKGNSPYS
GRGNSPFS
GRGNSPWS
GRGASPYA
d)
GRGNSPWS
GRGDVPYS
SRSDSPYS
ARADSPYS
WT
GRGNSPYS
GQGNSPYS
GKGNSPYS
GRGNSPFS
GRGASPYA
$R = -0.860$
$R^2 = 0.740$
$ln(\eta/\eta_{WT})$
$ln(C_{sat}/C_{sat}^{WT})$


a)
Ser:Ser
Arg:Ser
Tyr:Ser
Arg:Asp
Arg:Tyr
Gly:Ser
Tyr:Tyr
Arg:Arg


b)
Reduced LLPS
0.000
0.050
0.100
0.150
A to S (0.0019)
G to S, V to A (0.004)
V to S (0.0067)
R to Q (0.022)
A to F (0.023)
Q to K (0.026)
D to N (0.059)
Y to A (0.066)
Y to S (0.068)
W to A (0.127)
W to F (0.155)
G to A (0.0020)
V to G (0.0028)
S to F (0.0203)
G to F (0.0251)
V to F (0.027)
R to K (0.048)
W to Y, Y to V (0.061)
Y to F (0.089)
Legend :
Direct estimate
Indirect estimate
Polar
Anionic
Cationic
Aromatic
Glycine & Proline
Aliphatic
$\frac{1}{N} ln \frac{C_{sat}}{C_{ref}}$

Extended Data Fig. 1 Salt dependence of polycation LLPS.

(**a**)Turbidity curves for the polycationic GRGNSPYS variant at three salt concentrations (50, 137, and 300 mM NaCl). Three independent experiments are shown at each concentration. (**b**) Transition temperatures estimated from the turbidity curves as shown in (**a**). Data are presented as mean values +/− SD, n = 3 independent experiments.

Extended Data Fig. 2 Pairwise contacts of R-to-Q and D-to-N variants.

Average residue pairwise contacts for the (**a**) GRGNSPYS and (**b**) GQGDSPYS variants with respect to WT. Residue pairs not involving mutated residues are shown as dark gray circles while residue pairs involving the mutated residues are shown as red squares (D-to-N) or orange diamonds (R-to-Q).

Extended Data Fig. 3 Turbidity and partial phase diagram of GRGASPYA.

Turbidity (**a**) and partial phase diagram (**b**) of GRGASPYA at different concentrations of protein in PBS. Data are presented as mean values +/− SD, n = 3 independent experiments.

Extended Data Fig. 4 Pairwise contacts of aromatic substitutions in the polycation.

Average residue pairwise contacts for the (**a**) GRGNSPFS and (**b**) GRGNSPWS variants with respect to GRGNSPYS. Upper plots: Residue pairs not involving mutated residues are shown as dark gray circles while residue pairs involving the mutated residues are shown as purple triangles. Lower plots: Contact ratio between residue pairs for the GRGNSPFS and GRGNSPWS variants to that of the GRGNSPYS variant.

Extended Data Fig. 5 Material properties of A-IDP condensates.

(**a**) FRAP of WT and GRGNSPWS with RGG-GFP-RGG as a fluorescent tracer. Data are presented as mean values +/− SD, n = 4 (WT), n = 9 (GRGNSPWS) different condensates from one experiment. (**b**) Effect of total protein concentration on condensate viscosity, as measured by microrheology. Measurements were conducted for GRGNSPFS and GQGNSPYS, at two concentrations each. Data are presented as mean values +/− SD, n = 4 videos from one experiment. (**c**) MSDs measured in polycationic GRGNSPYS condensates using PEGylated and carboxylated beads. (**d**) Viscosity of GRGDSPYS (WT), GQGNSPYS, and GRGNSPYS determined by particle tracking microrheology of 0.5 μm PEGylated vs. carboxylated beads. Two factor with replication ANOVA confirmed difference in viscosities between 0.5 μm PEGylated and carboxylated beads is not statistically significant, with p-value of 0.753. Data are presented as mean values +/− SD, n = 4 videos from one experiment. (**e**) Viscosity of GRGDSPYS (WT), GQGNSPYS, and GRGNSPYS determined by particle tracking microrheology of 0.5 μm vs. 1 μm bead diameters. Two factor with replication ANOVA confirmed difference in viscosities between 0.5 μm and 1 μm beads is not statistically significant, with p-value of 0.268. Data are presented as mean values +/− SD, n = 4 videos from one experiment.

Extended Data Fig. 6 Sequence based predictors of LLPS.

(**a**) Solvation free energy from Wolfenden et al.[1] vs. saturation concentrations measured in this work. Each data point represents a unique variant used in this work. Variants differing by only one residue are connected by lines such that each mutation results in increasing saturation concentration. Dashed lines indicate variants that follow the trend of preferred interaction with solvent leading to lower phase separation propensity, while solid lines show mutations that result in less favorable interaction with solvent and lower phase separation propensity. (**b**) Ratio of phase separation propensity score for each sequence relative to the propensity score for WT, calculated using several online sequence-based

predictors – DeePhase, PScore, PSpredictor, FuzDrop, LLPhysScore and catgranule. Experimental values are shown as black circles. All predictor values are normalized with the WT to account for different scales used by the predictors. In all cases, when the normalized score is above 1, the sequence is predicted to undergo LLPS more avidly than the WT, while values below 1 indicate a lower propensity to undergo LLPS when compared to WT. Experimental values are calculated from the saturation concentration values (C_{sat}) measured at 37 °C. The experimental values are represented as C_{sat} of WT divided by C_{sat} of variant, such that here too, a value above 1 indicates greater phase separation propensity compared to WT, whereas a value below 1 indicates lower phase separation propensity. (**c**) Correlation between experimental values and predictor results. Data for all data sets are normalized from 0 to 1. Symbols are the same as shown in (b) for the predictors.

Extended Data Fig. 7 Temperature dependence of saturation concentrations for A-IDP variants.

(**a**) Ratio of saturation concentrations (C_{sat}) for different sequences with respect to C_{sat} of WT at different temperatures. Lines sloping down indicate that phase separation propensity with respect to the WT is enhanced at higher temperatures, whereas lines sloping upwards indicate reduction in phase separation propensity with respect to WT at higher temperatures. Solid lines indicate the temperatures at which saturation concentration was estimated using turbidimetry experiments, while dashed lines indicate the temperatures at which values were extrapolated from a logarithmic fit to the experimental binodal data. (**b**) Thermodynamic analysis performed for the different variants based on the estimated saturation concentrations at 20 °C. Higher values indicate greater reduction in phase separation propensity upon carrying out the mutation. Direct estimate refers to values that can be calculated from the experimental variants directly, whereas indirect estimate refers to values for which the mutation was not carried out in this work and the values were inferred based on data from multiple related experimental variants.

Extended Data Fig. 8 Droplet morphology and dynamics after 24 hrs.

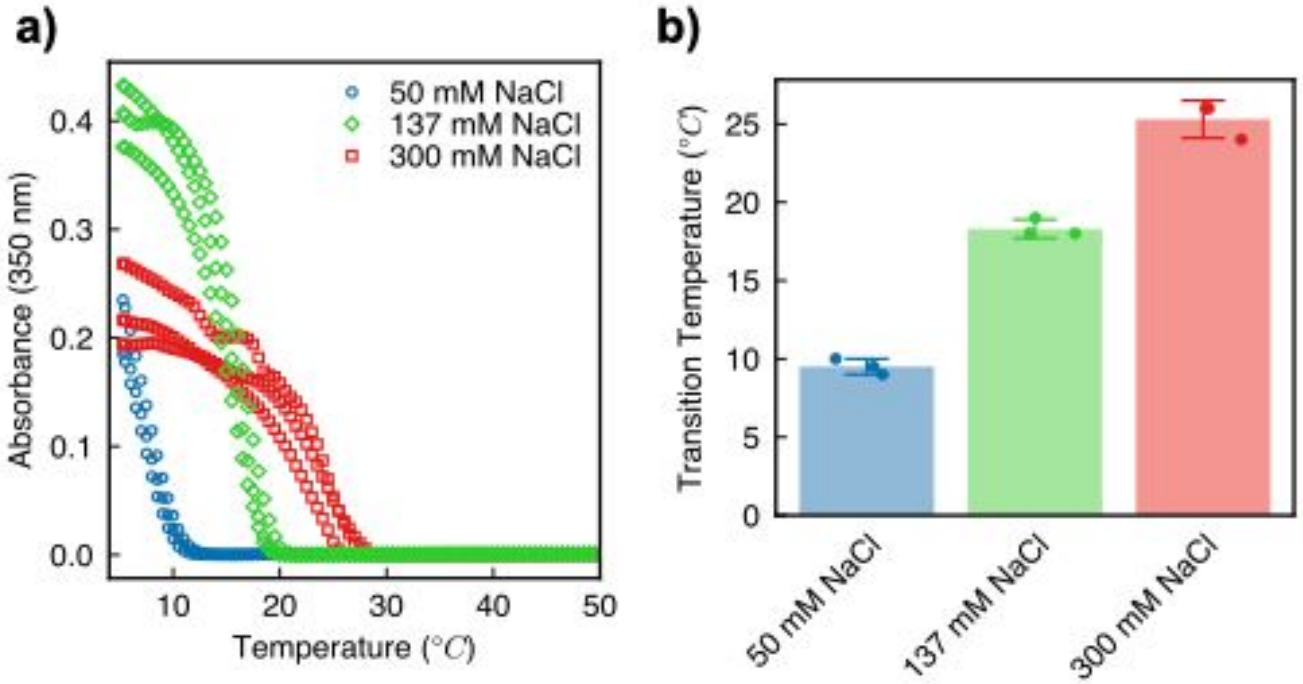
(**a**)Microscopy images for different mutants after 24 hrs of phase separation, showing regular spherical droplets and no signs of fibrillization or aggregation. (Scale bar: 5 μm). Data presented is representative of multiple images acquired for each sample and validated through imaging a second independent sample for 7 out of 12 variants. (**b**) Microscopy images of GRGNSPYS (cationic sequence), GRGDSPYS (WT) and GQGNSPYS (neutral sequence) undergoing droplet fusion and relaxing into a single spherical droplet, showing liquid-like behavior at 24 hrs. (Scale bar: 2 μm). Data presented is representative of results from two independent trials. Representative WT snapshots from Supplementary Movie 2. (**c**) Ensemble mean-squared displacement versus lag time at 24 hrs for the three representative variants (GRGNSPYS, GRGDSPYS, and GQGNSPYS), showing liquid-like behavior even after 24 hrs of phase separation.

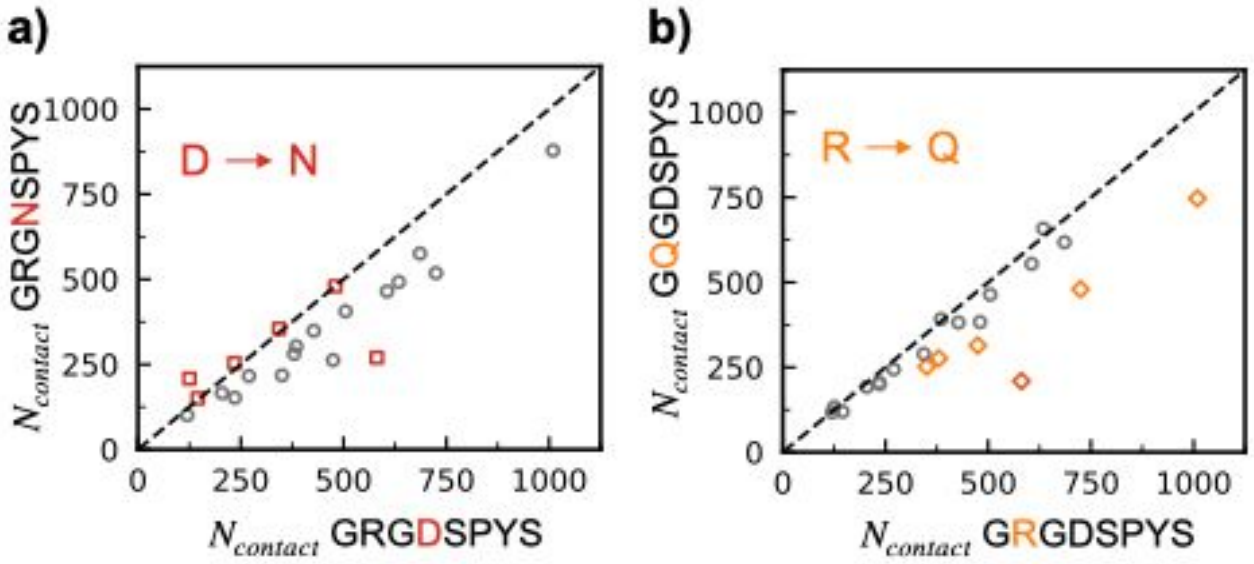
Extended Data Table 1 System specific changes to saturation concentration upon mutation

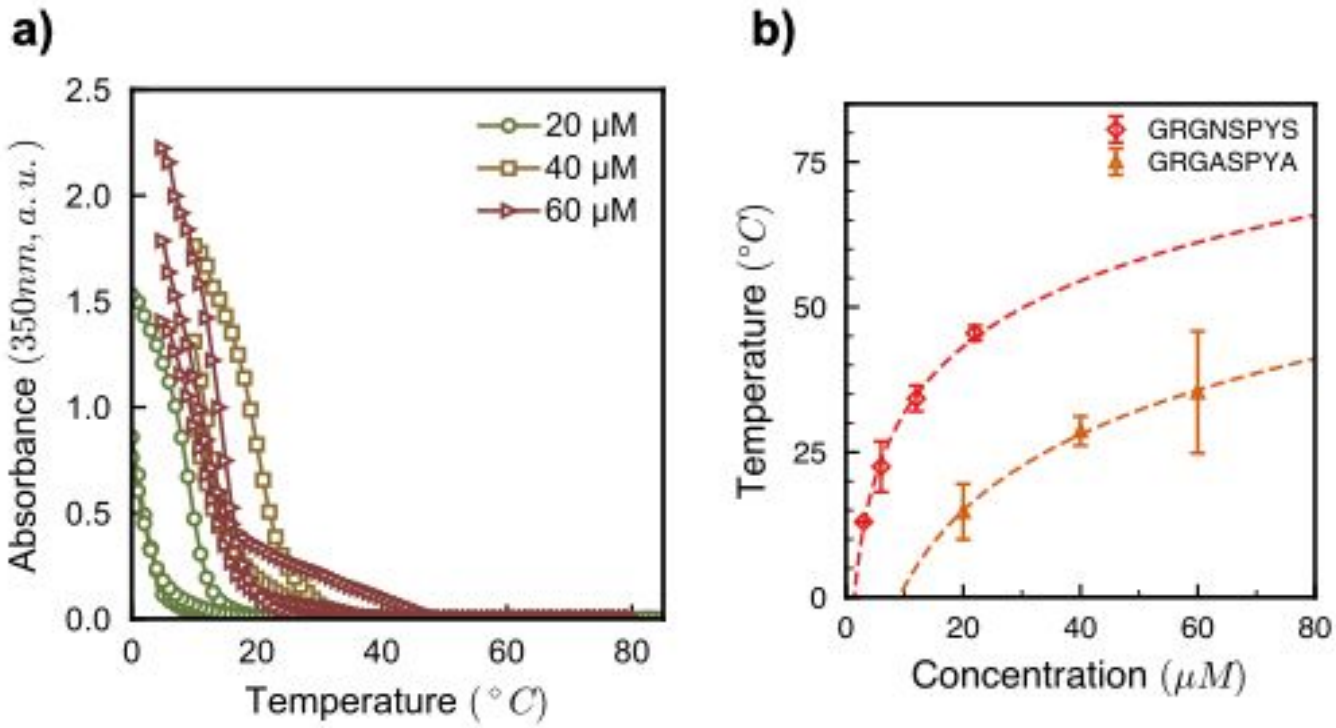
Tabulated values for the effects of different mutations to saturation concentration at 37 °C normalized by number of mutations carried out. The color code follows Fig. 6b (and Figure R8), where blue denotes direct estimates, or values for which the particular mutation was carried out in this work, while red denotes indirect estimates where mutations were not carried out but can be estimated based on combining data from multiple investigated variants.

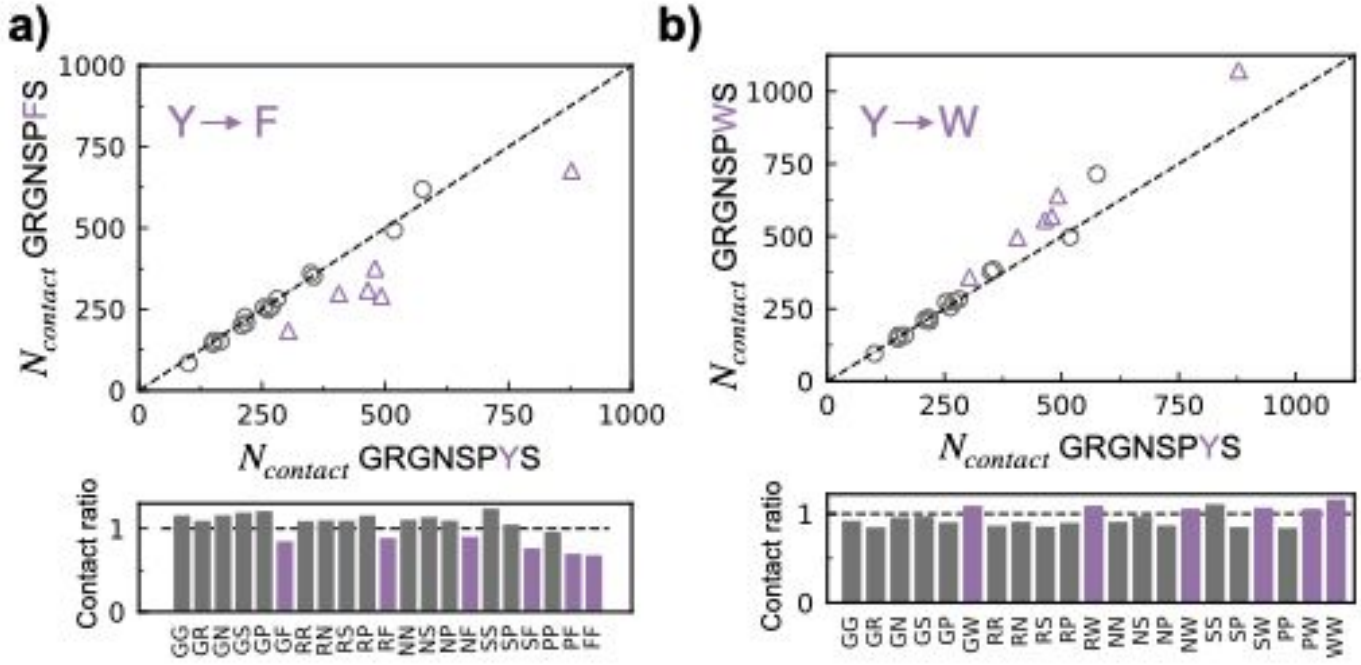
Extended Data Table 2 Sequence dependence of saturation concentration change with mutations

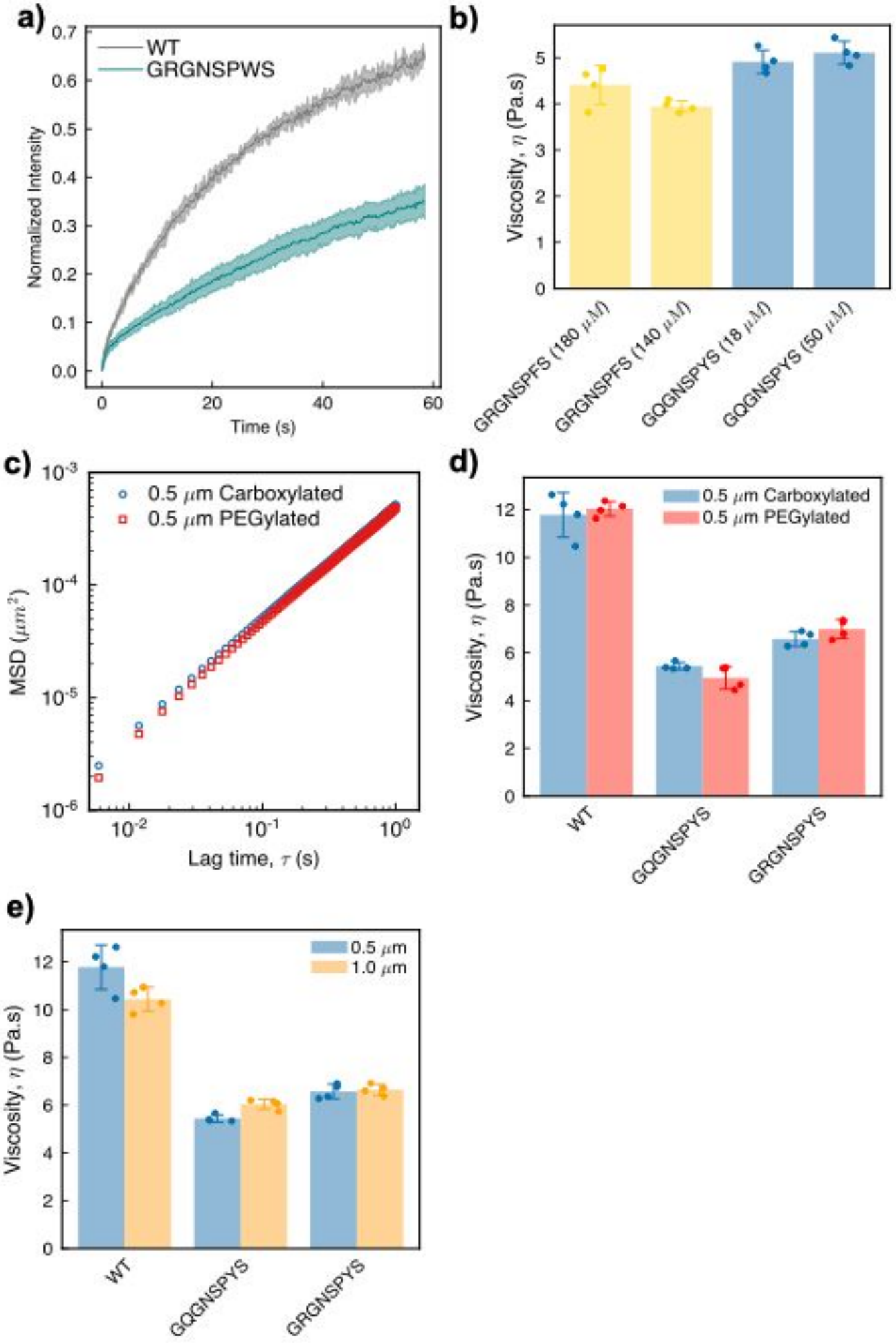
Thermodynamic analysis for frequently carried out mutations based on our work and findings presented in other studies. Mutations are compared based on the changes in saturation concentration with respect to reference saturation concentration upon mutation of a particular residue.


a)
50 mM NaCl
137 mM NaCl
300 mM NaCl
Absorbance (350 nm)
0.0
0.1
0.2
0.3
0.4
10
20
30
40
50
Temperature (°C)
b)
Transition Temperature (°C)
0
5
10
15
20
25
50 mM NaCl
137 mM NaCl
300 mM NaCl


a)
D → N
$N_{contact}$ GRGNSPYS
$N_{contact}$ GRGDSPYS
0
250
500
750
1000
b)
R → Q
$N_{contact}$ GQGDSPYS
$N_{contact}$ GRGDSPYS


a)
b)
Absorbance $(350nm, a.u.)$
Temperature $(^\circ C)$
20 μM
40 μM
60 μM
Temperature $(^\circ C)$
Concentration (μM)
GRGNSPYS
GRGASPYA

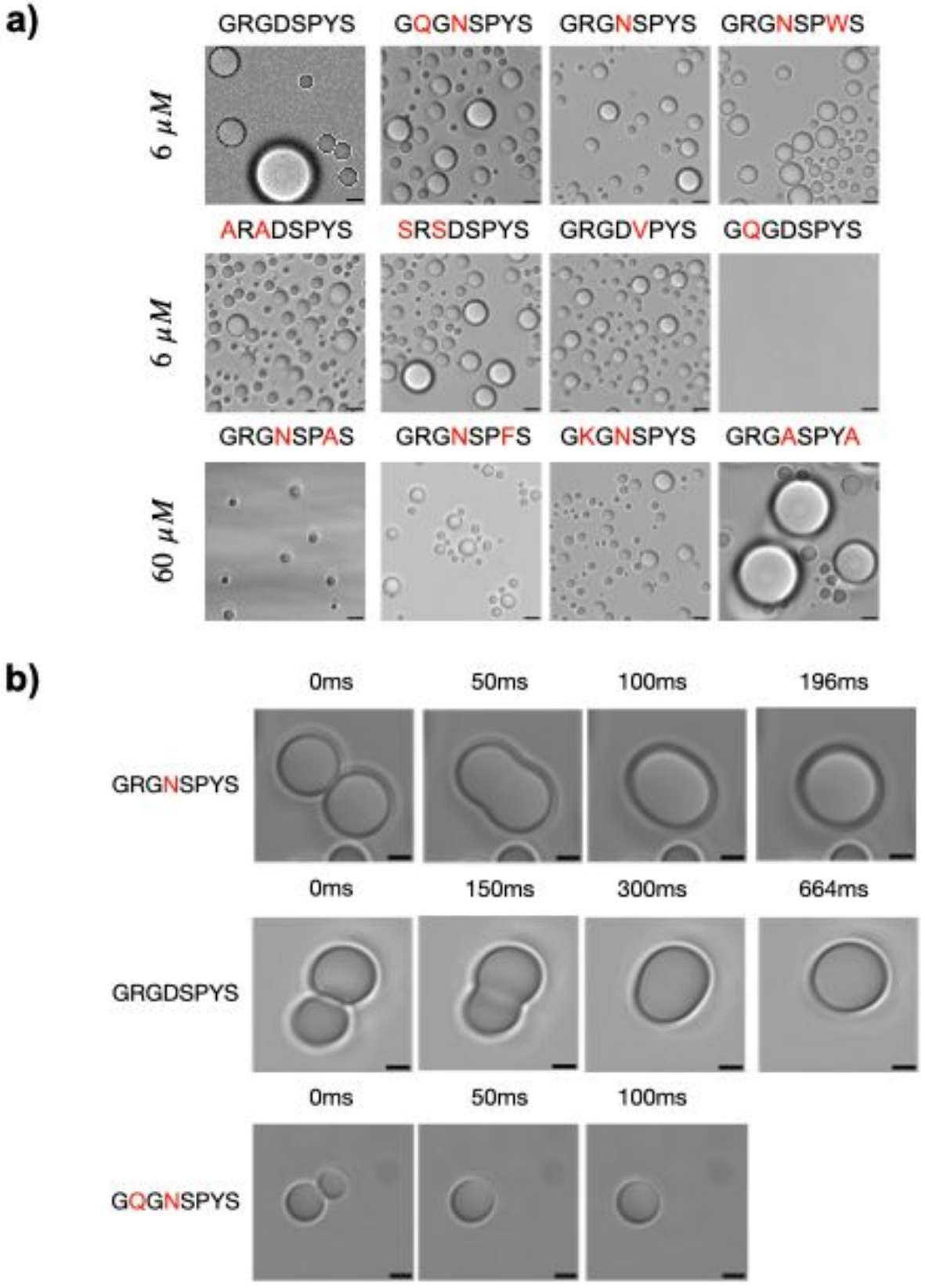

a)
Y→F
$N_{contact}$ GRGNSPFS
$N_{contact}$ GRGNSPYS
Contact ratio
GG GR GN GS GP GF RR RN RS RP RF NN NS NP NF SS SP SF PP PF FF
b)
Y→W
$N_{contact}$ GRGNSPWS
$N_{contact}$ GRGNSPYS
Contact ratio
GG GR GN GS GP GW RR RN RS RP RW NN NS NP NW SS SP SW PP PW WW

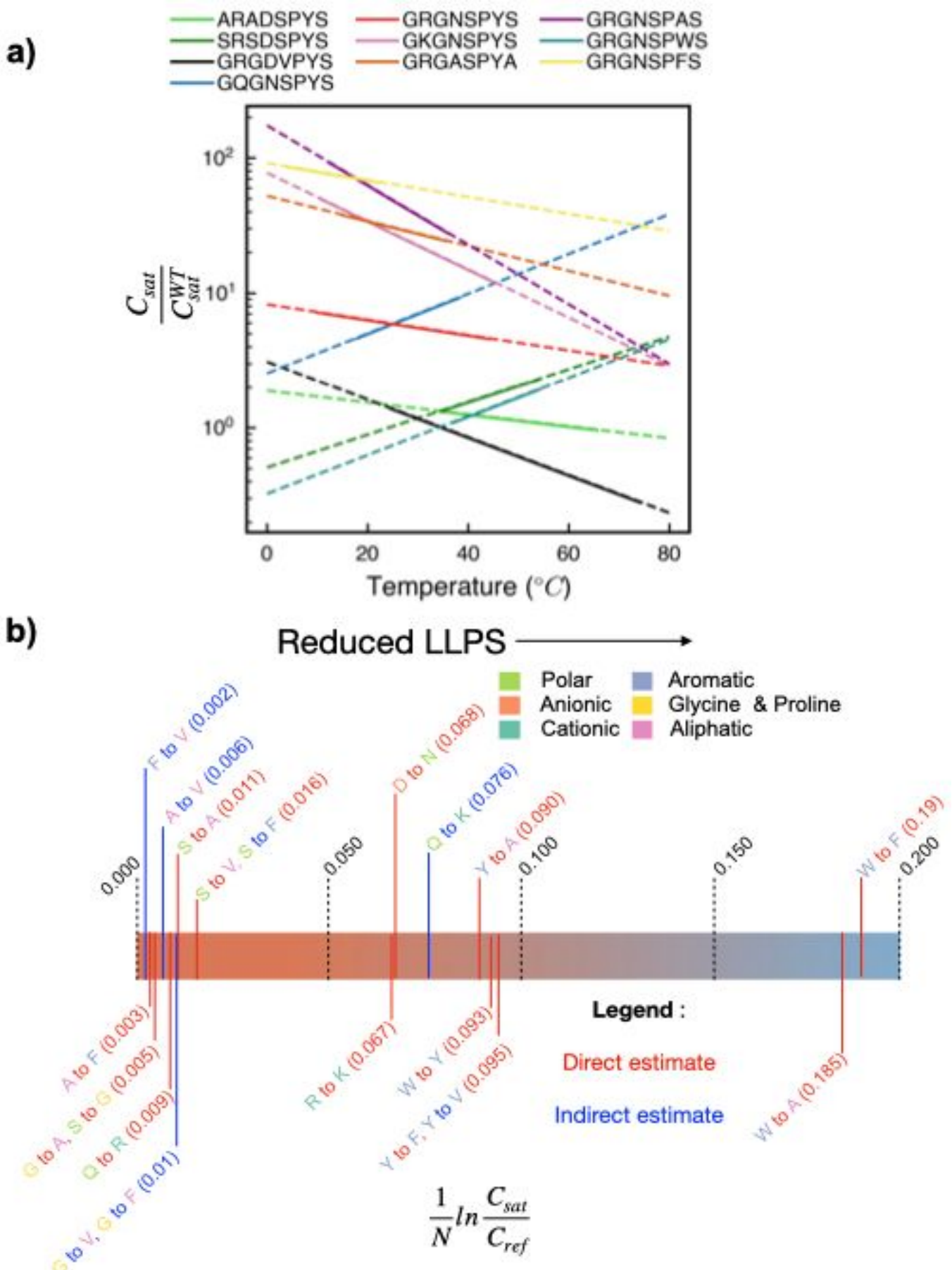

a)
WT
GRGNSPWS
Normalized Intensity
Time (s)
b)
Viscosity, η (Pa.s)
GRGNSPFS (180 μM)
GRGNSPFS (140 μM)
GQGNSPYS (18 μM)
GQGNSPYS (50 μM)
c)
0.5 μm Carboxylated
0.5 μm PEGylated
MSD (μm²)
Lag time, τ (s)
d)
0.5 μm Carboxylated
0.5 μm PEGylated
Viscosity, η (Pa.s)
WT
GQGNSPYS
GRGNSPYS
e)
0.5 μm
1.0 μm
Viscosity, η (Pa.s)
WT
GQGNSPYS
GRGNSPYS

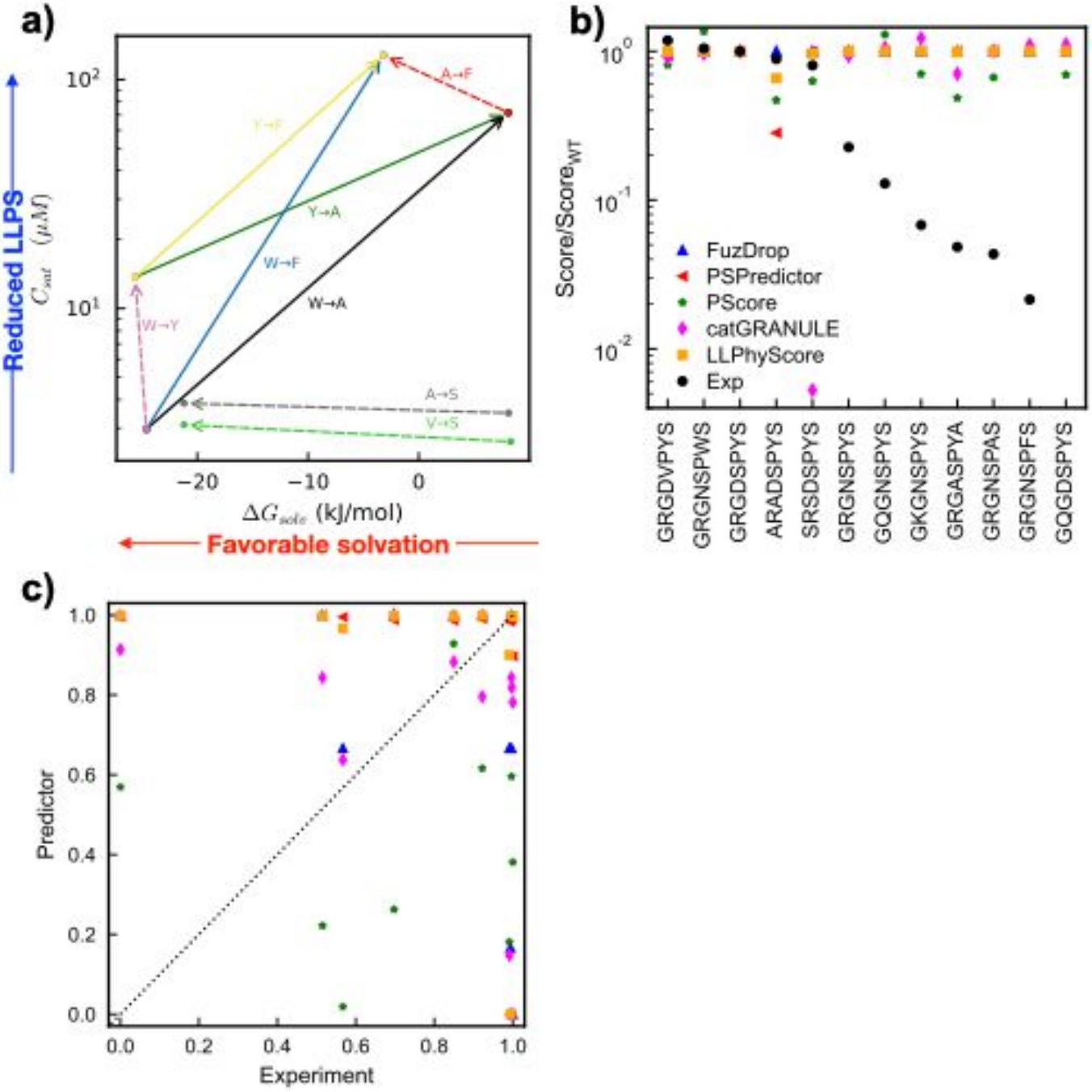
Mutation	Author	Protein	No. of mutations	$\frac{1}{N}\ln\frac{C_{sat}}{C_{sat}^{ref}}$
Y to F	Bremer et. al.[26]	hnRNPA1-LCD	12	0.149
	Bremer et. al.[26]	hnRNPA1-LCD	19	0.157
	Schuster et. al.[24]	LAF1 RGG	10	0.16
	Wang et. al.[21]	FUS	27	0.046
	Li et.al.[68]	AKAP95	6	Between 0 and -0.2
	This work - 20°C	A-IDP	25	0.095
	This work – 37°C	A-IDP	25	0.094
R to K	Dzuricky et. al.[12]	A-IDP	10	0.20
	Dzuricky et. al.[12]	A-IDP	20	0.21
	Bremer et. al.[26]	hnRNPA1	3	0.63
	Bremer et. al.[26]	hnRNPA1	6	0.66
	Brady et. al.[19]	DDX4	24	>0.22
	This work - 20°C	A-IDP	25	0.067
	This work – 37°C	A-IDP	25	0.048
G to A	Wang et. al.[21]	FUS	45	0.01
	Conicella et. al.[69]	TDP43-CTD	1	-1.02
	This work – 20°C	A-IDP	25	0.005
	This work – 37°C	A-IDP	25	0.002

Mutation	Type of mutation	$\frac{1}{N} ln \frac{C_{sat}}{C_{ref}}$
A to S	Aliphatic to Polar	0.0019
G to A	Other to Aliphatic	0.0020
V to G	Aliphatic to Other	0.0028
G to S	Other to Polar	0.0040
V to A	Aliphatic to Aliphatic	0.0040
V to S	Aliphatic to Polar	0.0067
S to F	Polar to Aromatic	0.0203
R to Q	Charged to Polar	0.022
A to F	Aliphatic to Aromatic	0.023
G to F	Other to Aromatic	0.0251
Q to K	Polar to Charged	0.026
V to F	Aliphatic to Aromatic	0.027
R to K	Charged to Charged	0.048
D to N	Charged to Polar	0.059
W to Y	Aromatic to Aromatic	0.061
Y to V	Aromatic to Aliphatic	0.061
Y to A	Aromatic to Aliphatic	0.066
Y to S	Aromatic to Polar	0.068
Y to F	Aromatic to Aromatic	0.089
W to A	Aromatic to Aliphatic	0.127
W to F	Aromatic to Aromatic	0.155

**Red color indicates direct estimates, while blue color indicates indirect estimates*


a)
GRGDSPYS
GQGNSPYS
GRGNSPYS
GRGNSPWS
6 μM
ARADSPYS
SRSDSPYS
GRGDVPYS
GQGDSPYS
6 μM
GRGNSPAS
GRGNSPFS
GKGNSPYS
GRGASPYA
60 μM
b)
0ms
50ms
100ms
196ms
GRGNSPYS
0ms
150ms
300ms
664ms
GRGDSPYS
0ms
50ms
100ms
GQGNSPYS
c)
WT
GQGNSPYS
GRGNSPYS
MSD (μm²)
10⁻³
10⁻⁴
10⁻⁵
10⁻⁶
10⁻²
10⁻¹
10⁰
1
Lag time, τ (s)


a)
ARADSPYS
SRSDSPYS
GRGDVPYS
GQGNSPYS
GRGNSPYS
GKGNSPYS
GRGASPYA
GRGNSPAS
GRGNSPWS
GRGNSPFS
$\frac{C_{sat}}{C_{sat}^{WT}}$
10^2
10^1
10^0
0
20
40
60
80
Temperature (°C)
b)
Reduced LLPS
Polar
Anionic
Cationic
Aromatic
Glycine & Proline
Aliphatic
F to V (0.002)
A to V (0.006)
S to A (0.011)
S to V, S to F (0.016)
D to N (0.068)
Q to K (0.076)
Y to A (0.090)
W to F (0.19)
0.000
0.050
0.100
0.150
0.200
A to F (0.003)
G to A, S to G (0.005)
Q to R (0.009)
G to V, G to F (0.01)
R to K (0.067)
W to Y (0.093)
Y to F, Y to V (0.095)
W to A (0.185)
Legend :
Direct estimate
Indirect estimate
$\frac{1}{N} ln \frac{C_{sat}}{C_{ref}}$


a)
Reduced LLPS
C_{sat} (μM)
10^2
10^1
Y→F
A→F
Y→A
W→F
W→A
W→Y
A→S
V→S
−20
−10
0
ΔG_{sole} (kJ/mol)
Favorable solvation
b)
Score/Score$_{WT}$
10^0
10^{-1}
10^{-2}
FuzDrop
PSPredictor
PScore
catGRANULE
LLPhyScore
Exp
GRGDVPYS
GRGNSPWS
GRGDSPYS
ARADSPYS
SRSDSPYS
GRGNSPYS
GQGNSPYS
GKGNSPYS
GRGASPYA
GRGNSPAS
GRGNSPFS
GQGDSPYS
c)
Predictor
Experiment
0.0
0.2
0.4
0.6
0.8
1.0

Supplementary Information

Expanding the molecular language of protein liquid-liquid phase separation

Shiv Rekhi[1#], Cristobal Garcia Garcia[2#], Mayur Barai[3#], Azamat Rizuan[1], Benjamin S. Schuster[3*], Kristi L. Kiick[2,4*], Jeetain Mittal[1,5,6*]

[1]Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843, USA

[2]Department of Materials Science and Engineering, University of Delaware, Newark, DE 19716, USA

[3]Department of Chemical and Biochemical Engineering, Rutgers, the State University of New Jersey, Piscataway, NJ 08854 USA

[4]Department of Biomedical Engineering, University of Delaware, Newark, DE 19176, USA

[5]Department of Chemistry, Texas A&M University, College Station, TX 77843, USA

[6]Interdisciplinary Graduate Program in Genetics and Genomics, Texas A&M University, College Station, TX 77843, USA

[#]These authors contributed equally
[*]benjamin.schuster@rutgers.edu
[*]kiick@udel.edu
[*]jeetain@tamu.edu

Table of Contents

Supplementary Methods:

Protein Expression and Purification

A single colony of *E. coli* M15[pREP4] containing the desired construct was inoculated in 150 mL of sterile LB media containing 100 μg mL^{-1} antibiotics (ampicillin) and grown overnight. Overnight culture media (150 mL) was used to evenly inoculate 6x750 mL of 2xYT media (yeast extract 10 g L^{-1}, NaCl 5 g L^{-1}, and tryptone 16 g L^{-1}) for protein expression. The 750 mL cultures were grown in a shaker at 37°C until the OD600 reached 0.6−0.8, and then isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM to induce protein expression. After 8 h of culture for protein expression, cells were harvested by centrifugation (5000 rpm for 15 min at 4°C), and the cell pellets were frozen with liquid nitrogen and transferred to a clean plastic bottle. The frozen cell pellets were thawed and resuspended in pH 8.0 native lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, and 10 mM imidazole) with 0.45 g of lysozyme. Lysed cells were further disrupted via sonication on ice, using a Fisher Scientific model 500 Sonic Dismembrator (10 mm tapered horn) for 20 min with a 10-s recovery time and subsequently incubated with RNAse (10 μg mL^{-1}) and DNAse (5 μg mL^{-1}) for 30 minutes. The supernatant from centrifugation (20,000 rpm for 15 min at 4°C) of cell lysate was separated and the cell pellet was resuspended in denaturing lysis buffer B (8M urea, 100 mM NaH_2PO_4, 10 mM Tris·Cl, pH 8.0) via sonication for 1 min with a 10-s recovery time. The supernatant from centrifugation (20,000 rpm for 15 min at 4°C) was collected and the pH was adjusted to 8.0, followed by incubation with Ni-NTA resin for 1 hour at room temperature. The protein-loaded resin was then loaded into a gravitational flow column, washed with denaturing lysis buffer B, denaturing wash buffer C (8M urea, 100 mM NaH_2PO_4, 10 mM Tris·Cl, pH 6.3), denaturing elution buffer D (8M urea, 100 mM NaH_2PO_4, 10 mM Tris·Cl, pH 5.9), and finally eluted with 75 mL denaturing elution buffer E (8M urea, 100 mM NaH_2PO_4, 10 mM Tris·Cl, pH 4.5). Elution E fractions were carefully transferred and dialyzed (MWCO 3.5 kDa) against deionized water (5 L) at room temperature with at least 7 changes of water before lyophilization. The purified protein yield was approximately 30−50 mg per liter of cell culture.

Fluorescence Recovery After Photobleaching (FRAP)

FRAP experiments were performed on a Zeiss Axio Observer 7 inverted microscope equipped with an LSM900 laser scanning confocal module and a 63x/1.4 NA plan-apochromatic oil-immersion objective. WT and $(GRGNSPWS)_{25}$ were mixed with 5% of RGG-GFP-RGG, which partitions into the condensates and serves as a FRAP probe (here, RGG denotes LAF-1 RGG domain)[1]. A region of approximate radius R = 1.5 μm, within droplets whose radii were approximately 2.5R, was bleached with a 405-nm laser. Subsequent fluorescence recovery of the bleached area was recorded with a 488-nm laser for 1 min. Raw FRAP data was normalized and averaged (n = 4 to 9 separate droplets) to obtain the final FRAP recovery curve.

PEGylation Reaction for carboxylated beads

1. Allow PEG and EDC to thaw for 20min
2. Sonicate stock PS beads for 10min prior to aliquoting
3. Aliquot and dilute beads based on above dilutions in UP H2O.
4. Sonicate diluted beads for 7 minutes.
5. Add PEG (PEG5k-NH2, Creative PEGworks) to bead suspension in 1.5mL siliconized tube.
6. Vortex to dissolve PEG.
7. Steps 7-9 should be done quickly.Add Sulfo-NHS (Sigma, stored at 4C) to each tube. Mix to dissolve.
8. Add 200mM borate buffer (pH 8.2).
9. Add EDC to each tube and mix. EDC should be fresh (check open date on container).
10. Place on rotary shaker for 4 hours at room temp.
11. Collection depends on particle size:
    a. Spin 40 nm PS for 12 min at 14,000xg in Amicon Ultra 0.5 mL filter tube, 100k MWCO. Wash 2x in UP H2O
    b. Spin 100 nm PS for 25 min at 21,000 x g. Wash 2x in UP H2O
    c. Spin 200nm for 15 min at 18,000x g. Wash 2x in UP H2O
    d. Spin 500 nm for 12 min at 15,000 x g. Wash 2x in UP H2O
    e. Spin 1 um (or larger) for 10 min at 10,000 x g. Wash 2x in UP H2O.
12. Resuspend collected particles in desired solution back to orginal PS stock volume or 2x diluted from stock

Supplementary Figures:

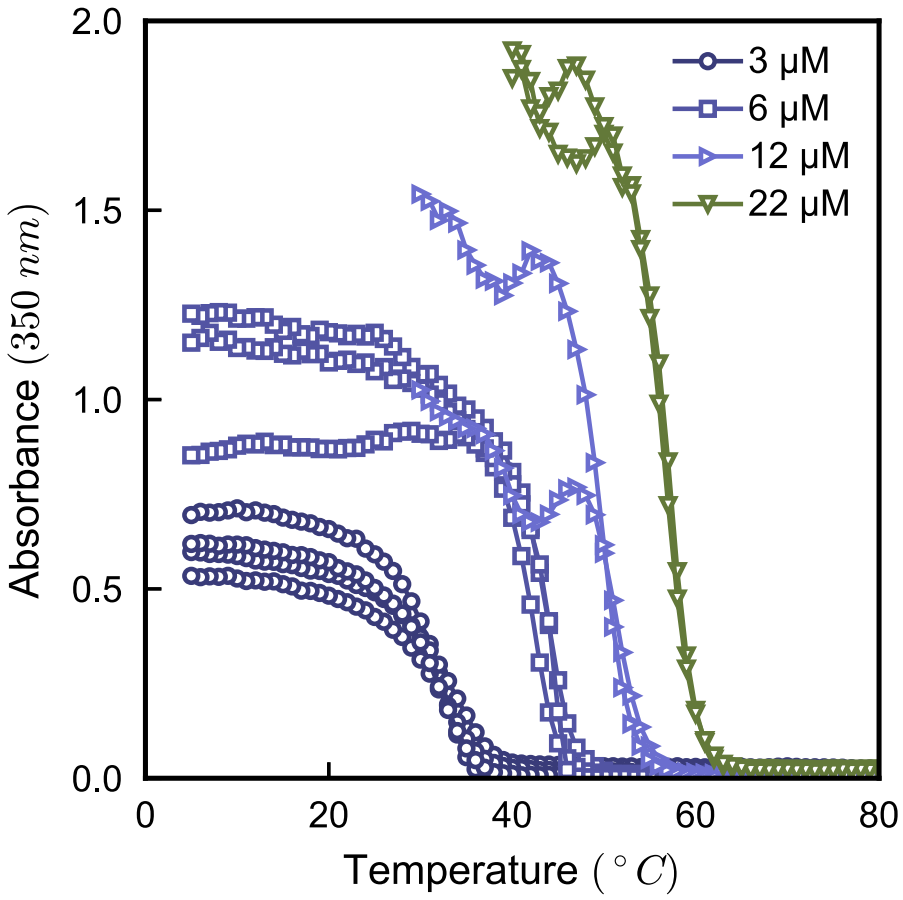


Fig 1. Turbidity measurements for the WT at 4 different concentrations in PBS. The transition temperatures measured are then converted into the partial phase diagram shown in Fig 1.

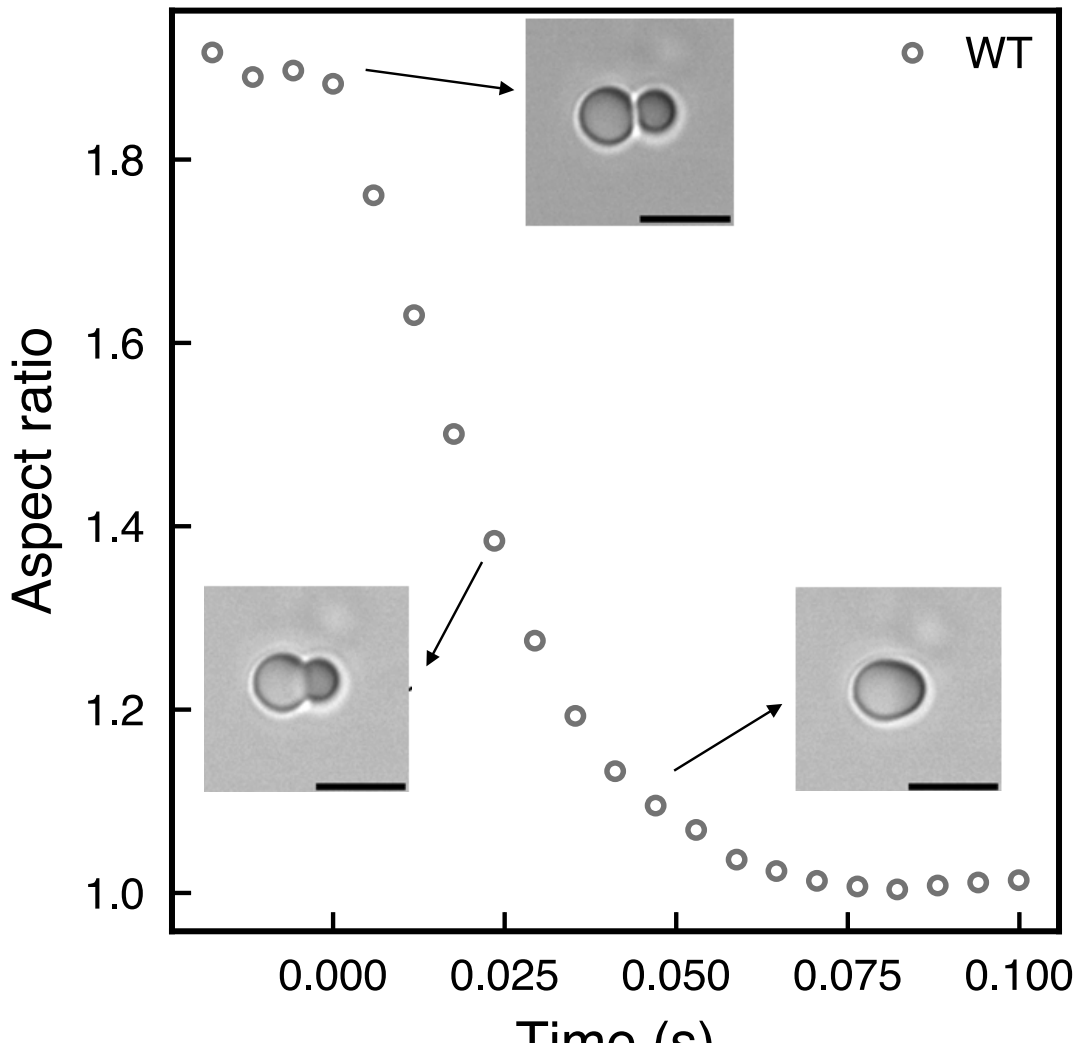


Fig 2. Plot of aspect ratio of fusing droplets relaxing exponentially to a sphere. The droplets are of the WT sequence. Scale bar is 5 μm. Similar results were obtained from at least two independent analyses

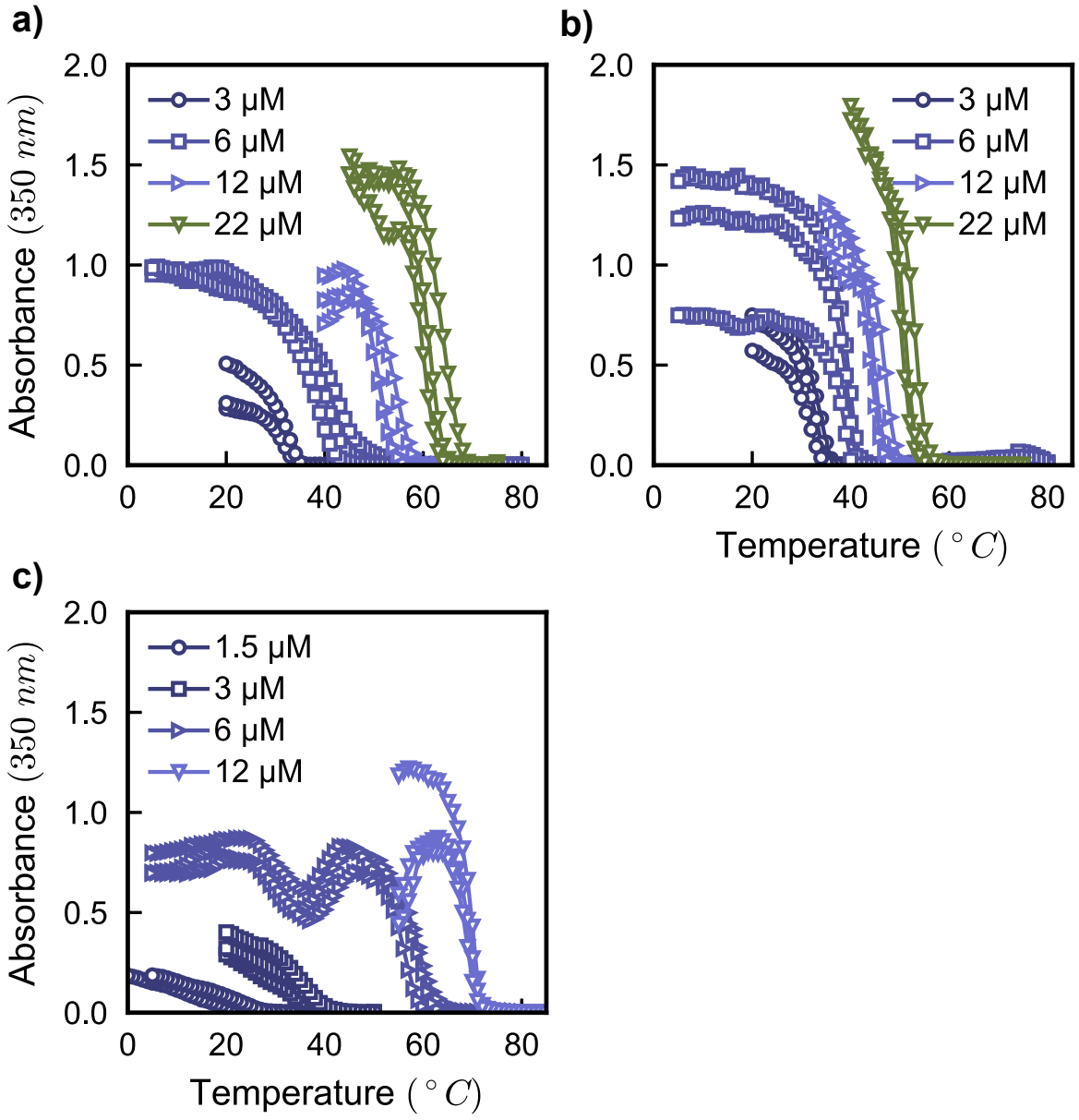


Fig 3. Turbidity experiments for (a) ARADSPYS, (b) SRSDSPYS, and (c) GRGDVPYS conducted at 4 different concentrations in PBS.

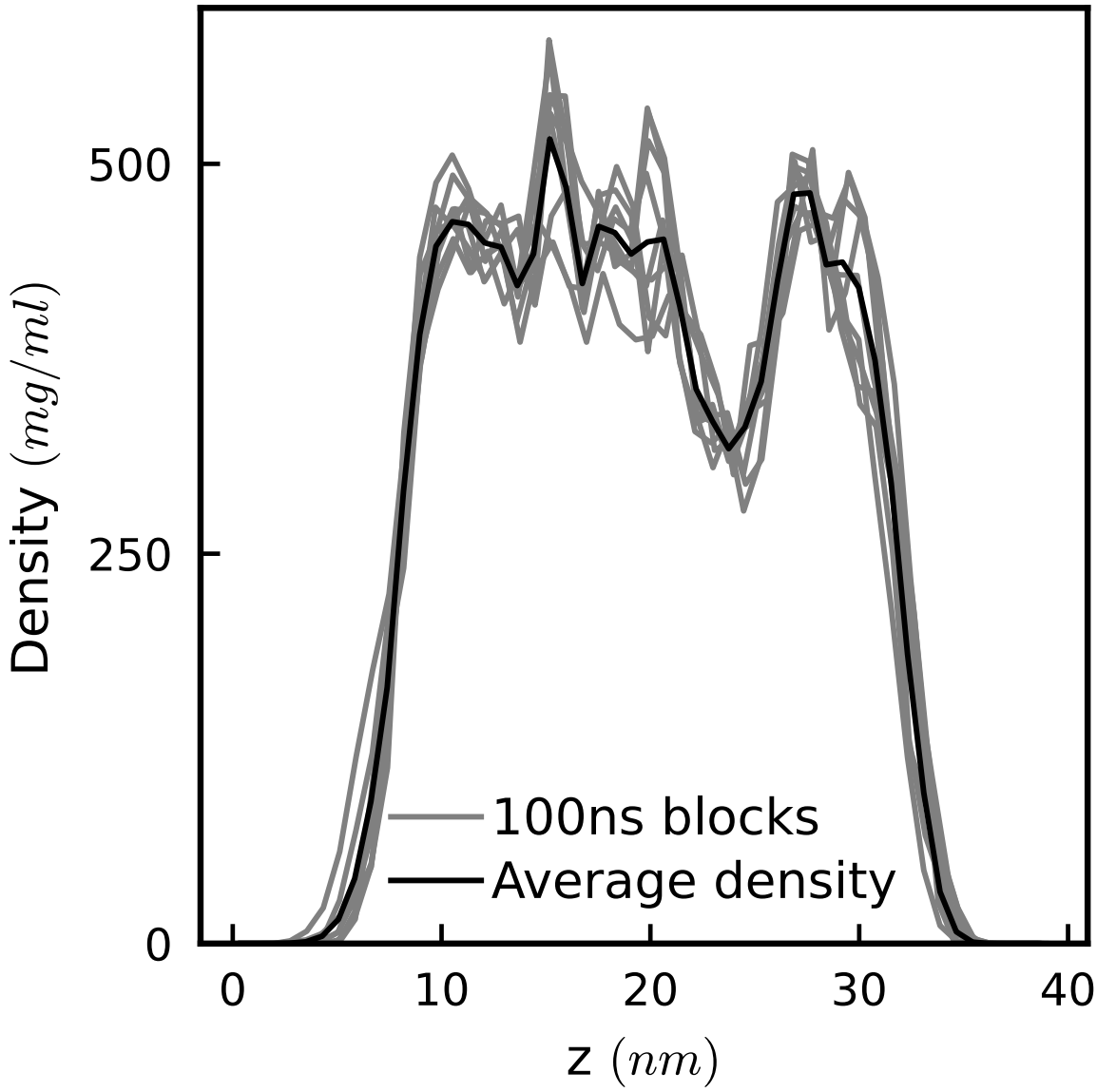


Fig 4. Density profile of the WT A-IDP calculated from atomistic slab simulations. The lines in gray represent calculated densities split into 10 blocks of 100 ns each, e.g., 0-100 ns, 100-200 ns and so on, while the black line represents the averaged density profile over the entire 1 μs trajectory highlighting that the protein density remains stable through the course of the simulation.

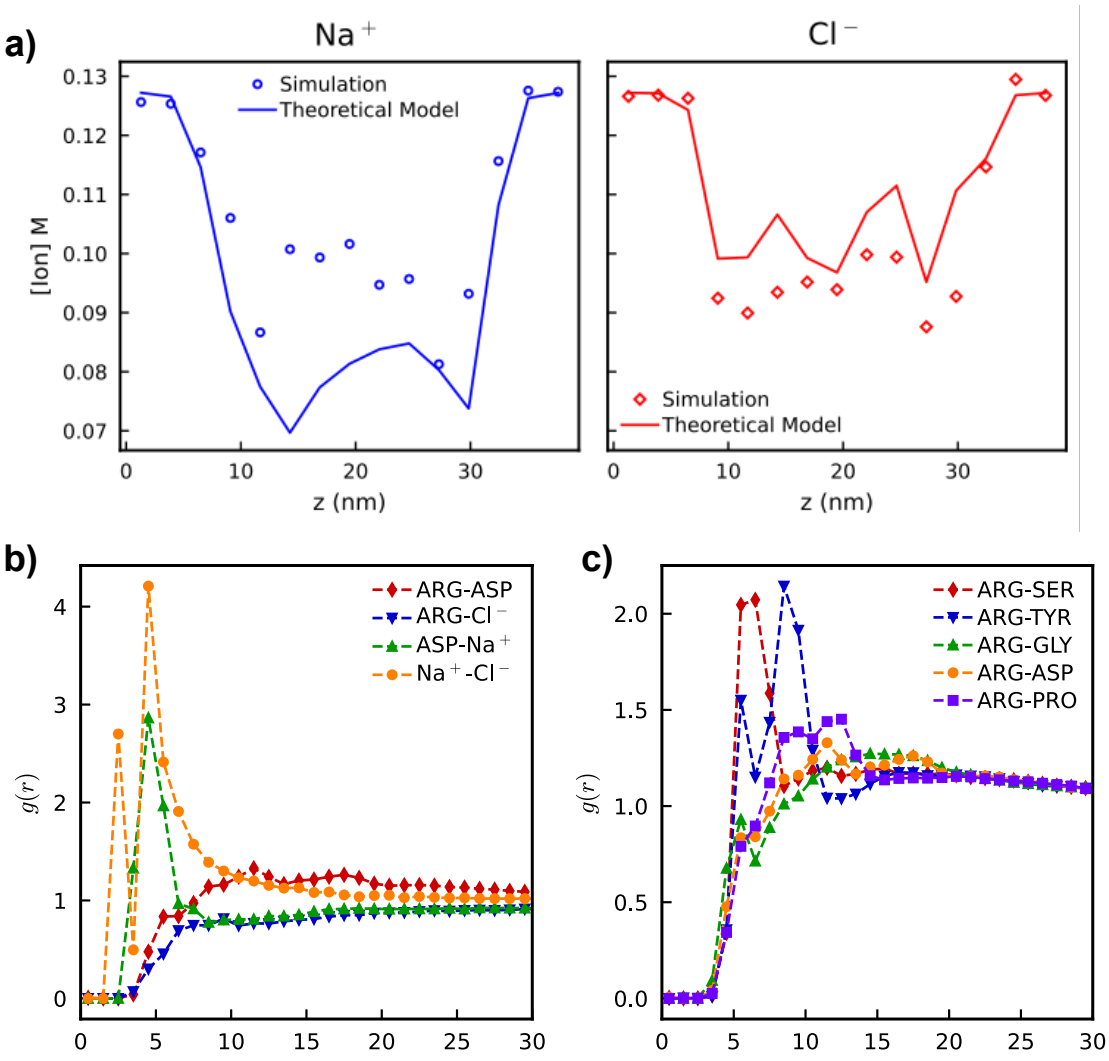


Fig 5. (a) Concentration profiles of Na^+ and Cl^- ions in the atomistic slab simulation as a function of z-coordinate. The symbols represent the estimated concentrations of ions from the atomistic simulation while the solid lines correspond to the predicted concentration profiles on applying the above detailed theoretical expression. Z coordinates of 0-5nm indicate the dilute phase, 5-35nm the protein dense phase and 35-40nm the dilute phase. **(b,c)** Pair correlation functions calculated from atomistic simulations of the WT A-IDP detailing the interactions between **(b)** charged species in the simulation. The specific cases highlighted are, oppositely charged residues (Arg-Asp), positively charged residues and negative counterion (Arg-Cl^-), negatively charged residues and positive counterion (Asp-Na^+) and salt ions (Na^+-Cl^-) and **(c)** Five most frequently occurring pairwise contacts in the atomistic simulations involving Arg residues.

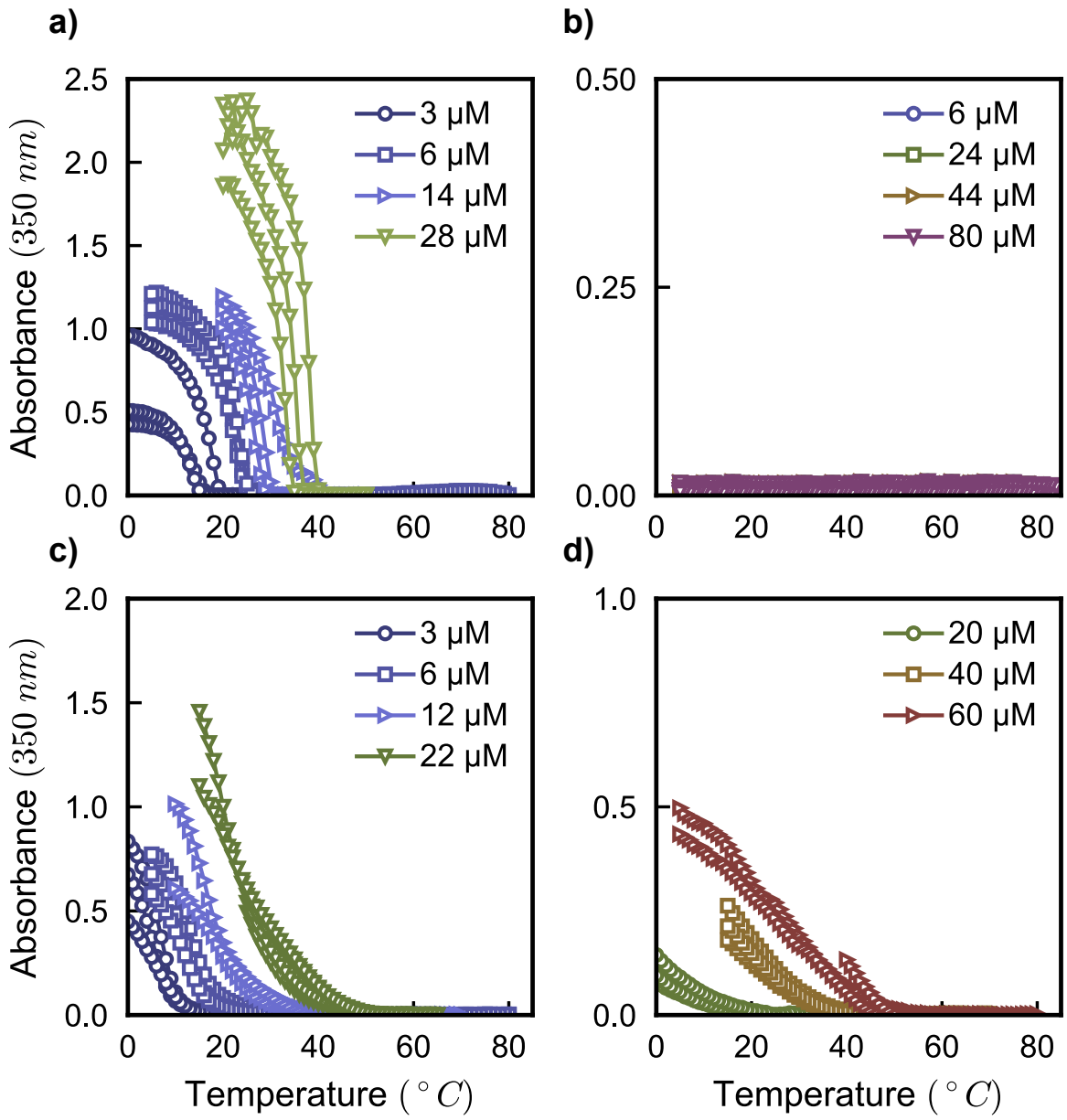


Fig 6. Turbidity experiments for (a) GQGNSPYS, (b) GQGDSPYS, (c) GRGNPSYS, and (d) GKGNSPYS at different concentrations in PBS. Since the GKGNSPYS variant only phase separates at higher concentrations, the measurements are shown from 20 μM onwards. The GQGDSPYS variant shows no measurable transition even up to 80 μM.

.

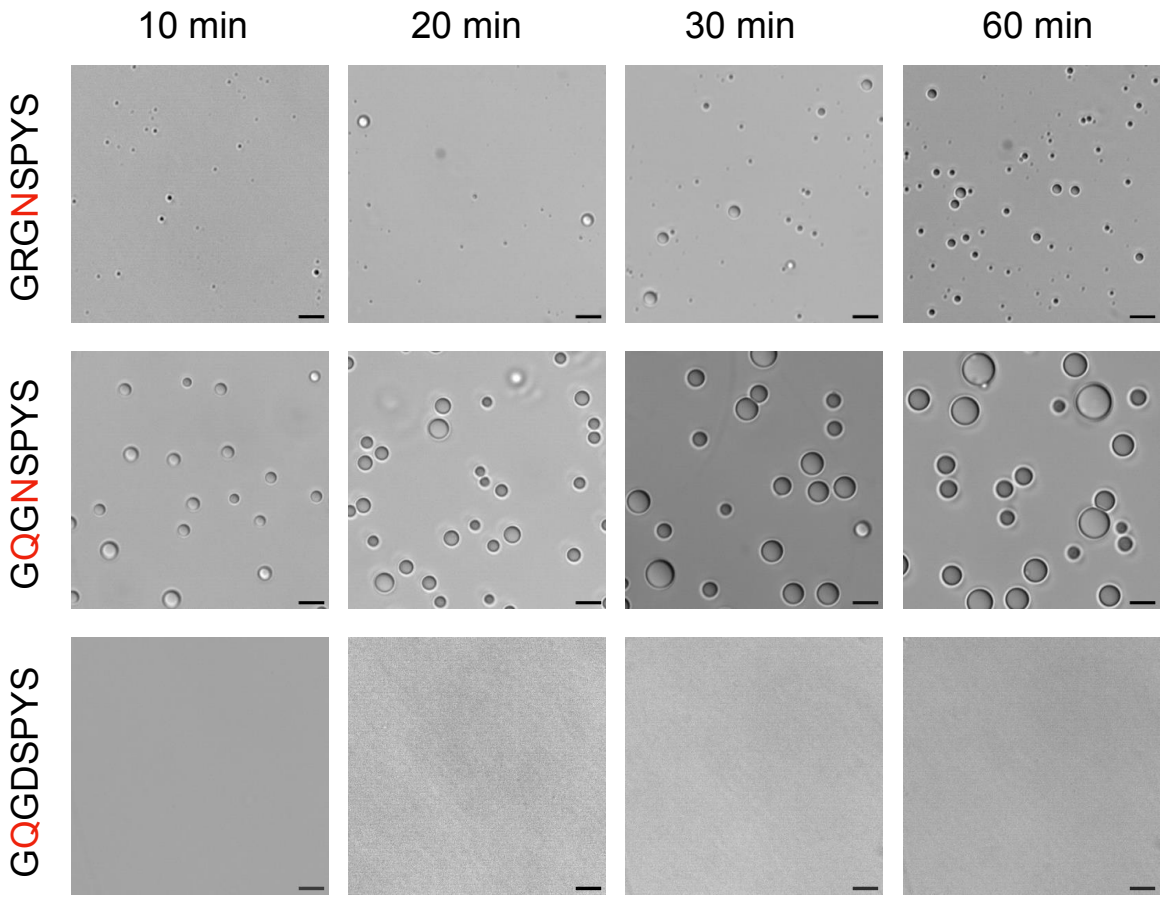


Fig 7. Microscopy images showing comparison of droplet growth over time for cationic sequence (GRGNSPYS), neutral sequence (GQGNSPYS), and anionic sequence (GQGDSPYS). The cationic sequence with tyrosine shows slower growth of droplets as compared to the neutral sequence. (Scale bar: 5 µm). Similar results were obtained from at least two independent analyses

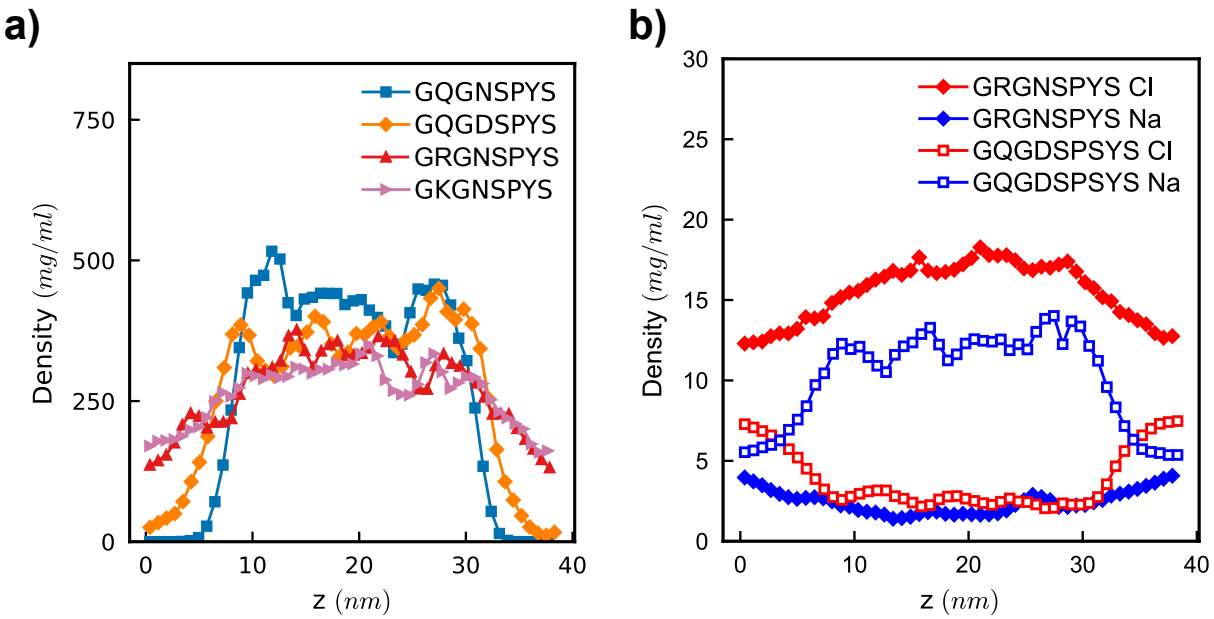


Fig 8. (a) Protein density profile as a function of z dimension for variants presented in Figure 3 (GQGNSPYS, GQGDSPYS, GRGNSPYS, and GKGNSPYS). Densities are averaged over the entire 1 μs atomistic slab trajectories. (b) Ion concentrations as a function of z dimension in the atomistic slab for the two variants with net charge, GQGDSPYS (empty symbols) and GRGNSPYS (filled symbols). Ion concentrations are averaged over the entire 1 μs atomistic slab trajectories.

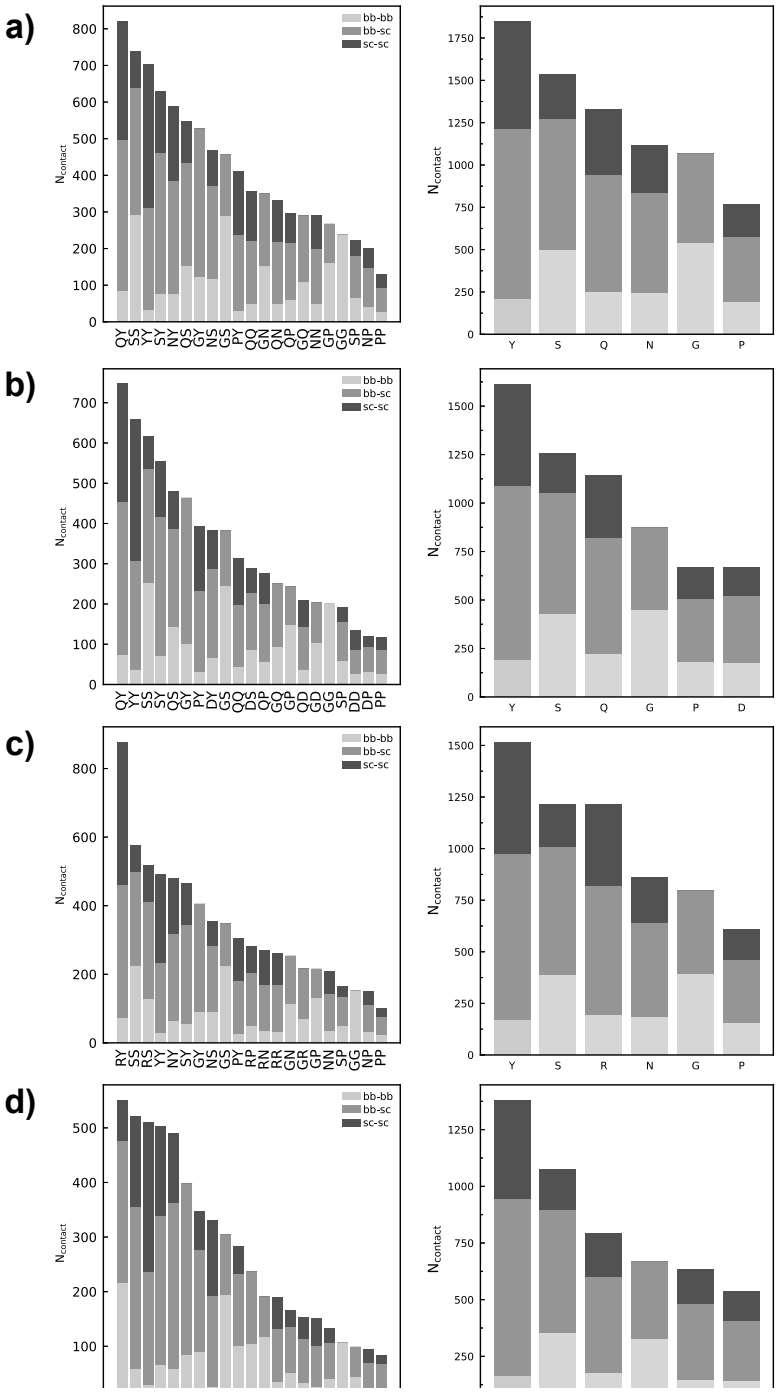


Fig 9. Average residue pairwise contacts (left) and per-residue contacts (right) for the (a) GQGNSPYS, (b) GQGDSPYS, (c) GRGNSPYS, and (d) GKGNSPYS variants, estimated from the atomistic slab trajectories. Both residue pair and per-residue contacts are decomposed into backbone-backbone, backbone-sidechain, and sidechain-sidechain. (Data is not normalized by residue abundance in the sequences.)

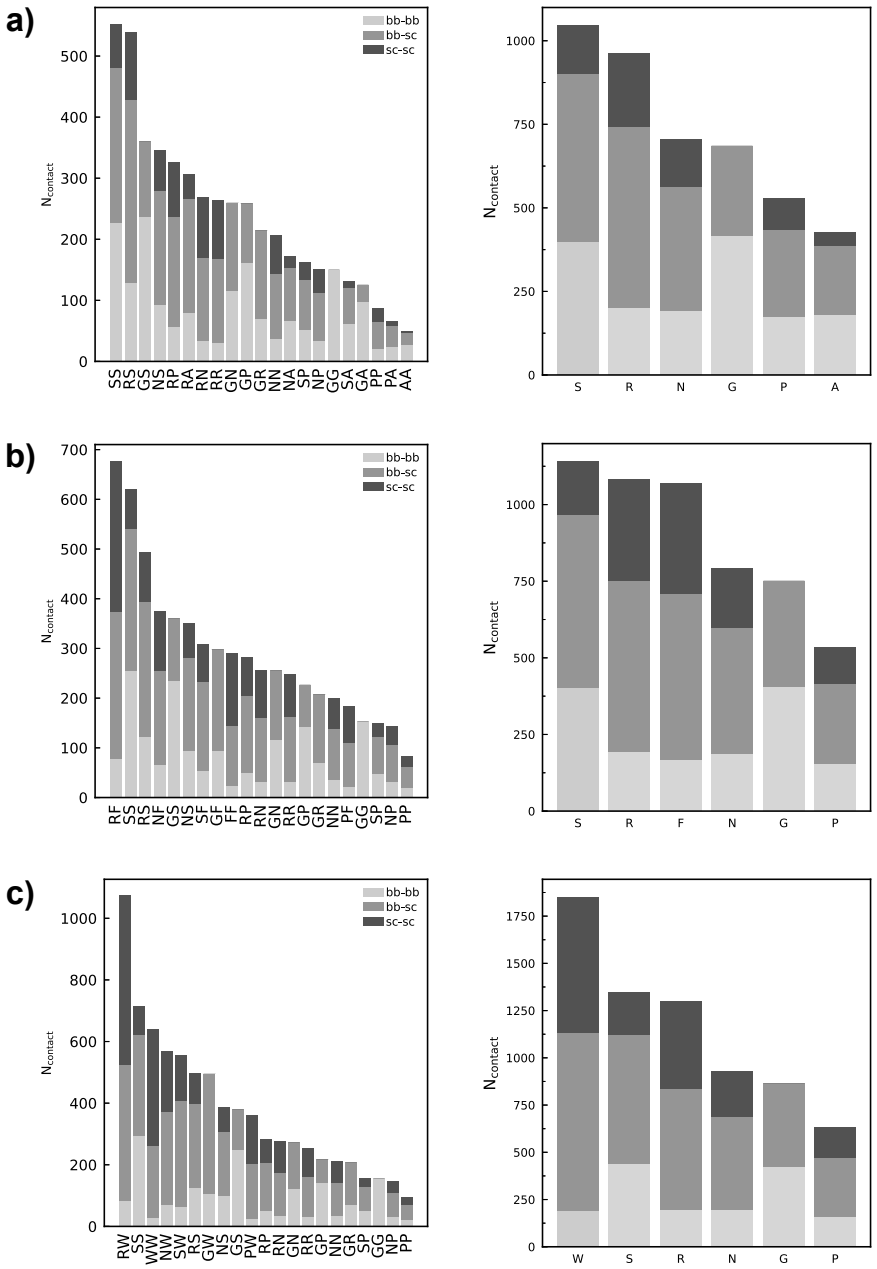


Fig 10. Average residue pair contacts (left) and per-residue contacts (right) for the (a) GRGNSPAS, (b) GRGNSPFS, and (c) GRGNSPWS variants estimated from the atomistic slab trajectories. Both residue pair and per-residue contacts are decomposed into backbone-backbone, backbone-sidechain, and sidechain-sidechain. (Data is not normalized by residue abundance in the sequences.)

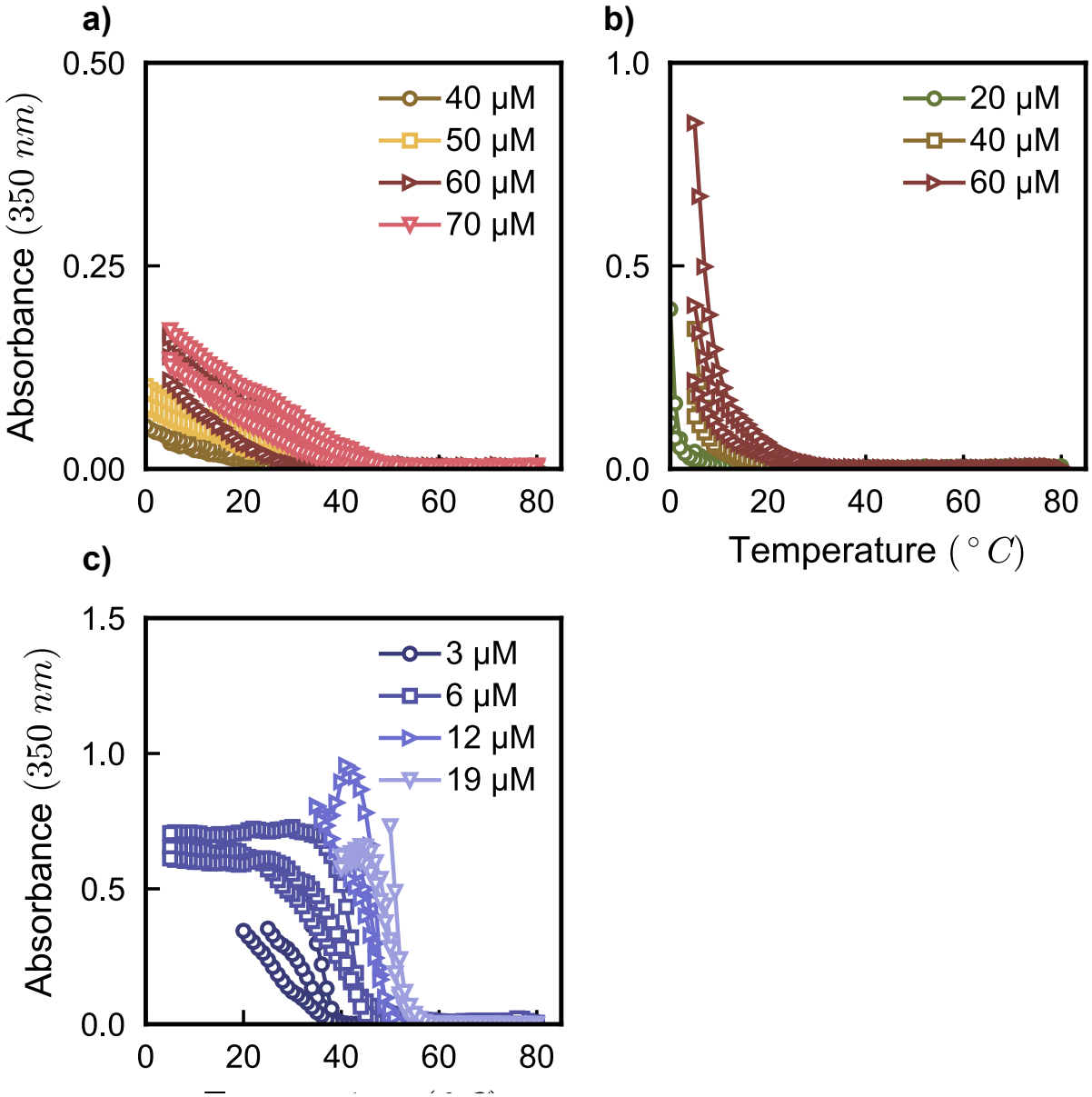
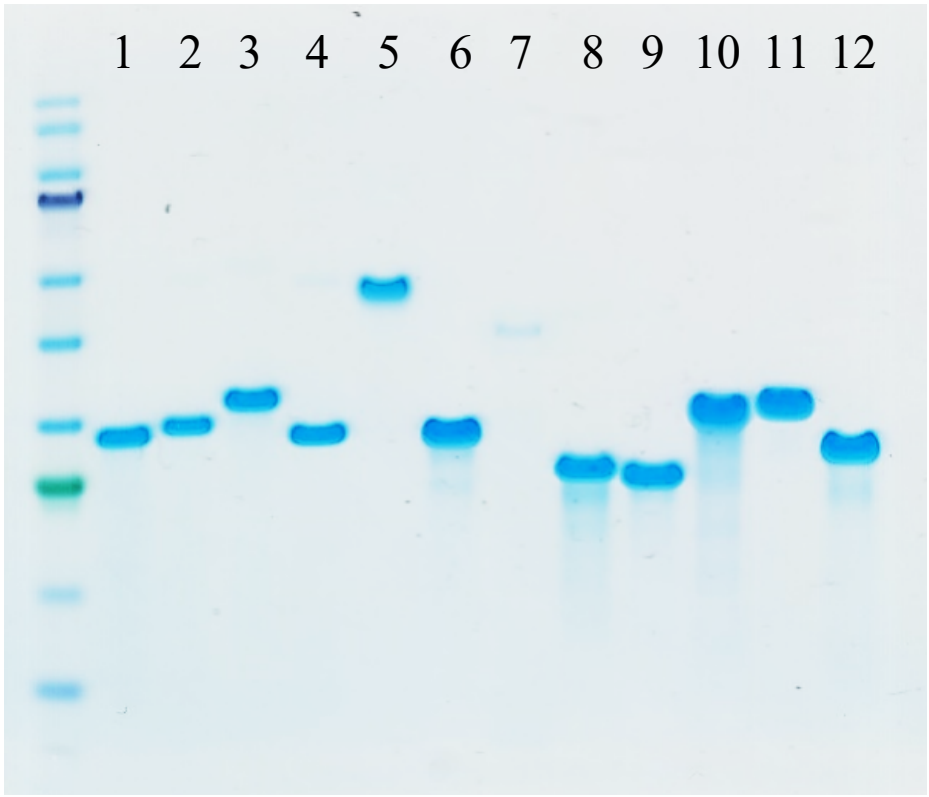


Fig 11. Turbidity measurements at different concentrations in PBS for the (a) GRGNSPAS, (b) GRGNSPFS, and (c) GRGNSPWS variants. GRGNSPAS and GRGNSPFS undergo LLPS at higher concentrations; therefore, for clarity, turbidity assays at lower concentrations where a transition is not observed are omitted.



1. GRGDSPYS (WT)
2. ARADSPYS
3. SRSDSPYS
4. GRGDVPYS
5. GQGNSPYS
6. GRGNSPYS
7. GQGDSPYS
8. GRGNSPWS
9. GRGNSPFS
10. GRGNSPAS
11. GKGNSPYS
12. GRGASPYA

Fig 12. SDS-PAGE of all sequences showing a unique band after purification. All samples were loaded at 0.5 mg/mL and run under standard protocols. Representative image displays results from all 12 samples in a single run. However, SDS-PAGE gel on individual samples were run separately during purification steps. Sequences rich in Gln (Q) have been reported to form oligomers[2] which is likely the origin of the apparent doubled MW in the sequences in columns 5 and 7. The polyanionic sequence GQGDSPYS, in column 7, shows a weaker band compared with the other constructs, owing to weaker staining by the Coomassie blue stain, which interacts with positively charged amino acids. Slight variations in the MW of the constructs arise not only from variations in their theoretical MW but also likely variations in their charge/mass ratios because of their differential adsorption of SDS due to differences in their ionic character.

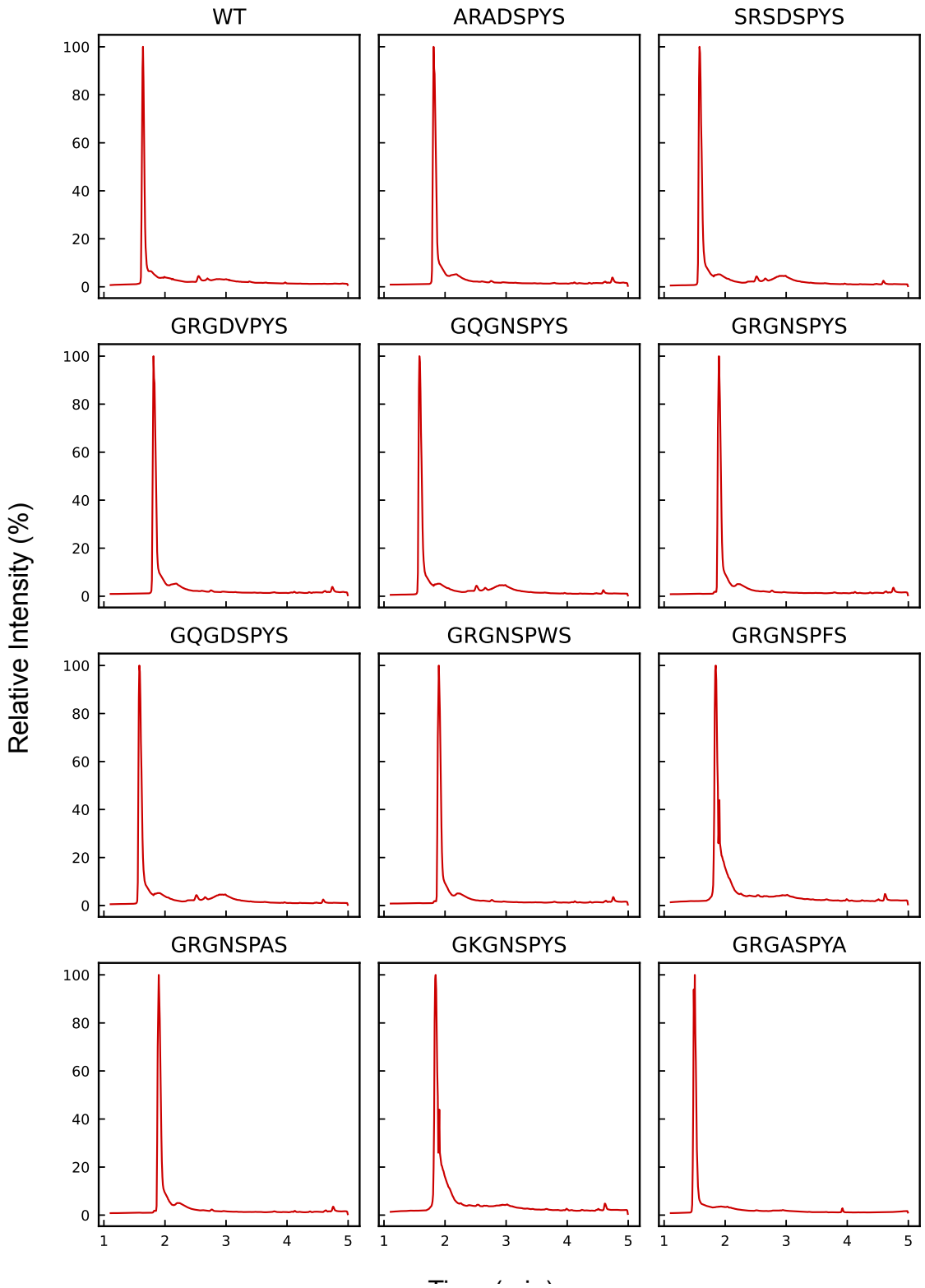


Fig 13. Chromatograms obtained by UPLC (XEVO). The green line in the first plot shows the area analyzed in mass spectrometry for each sequence (Fig S19).

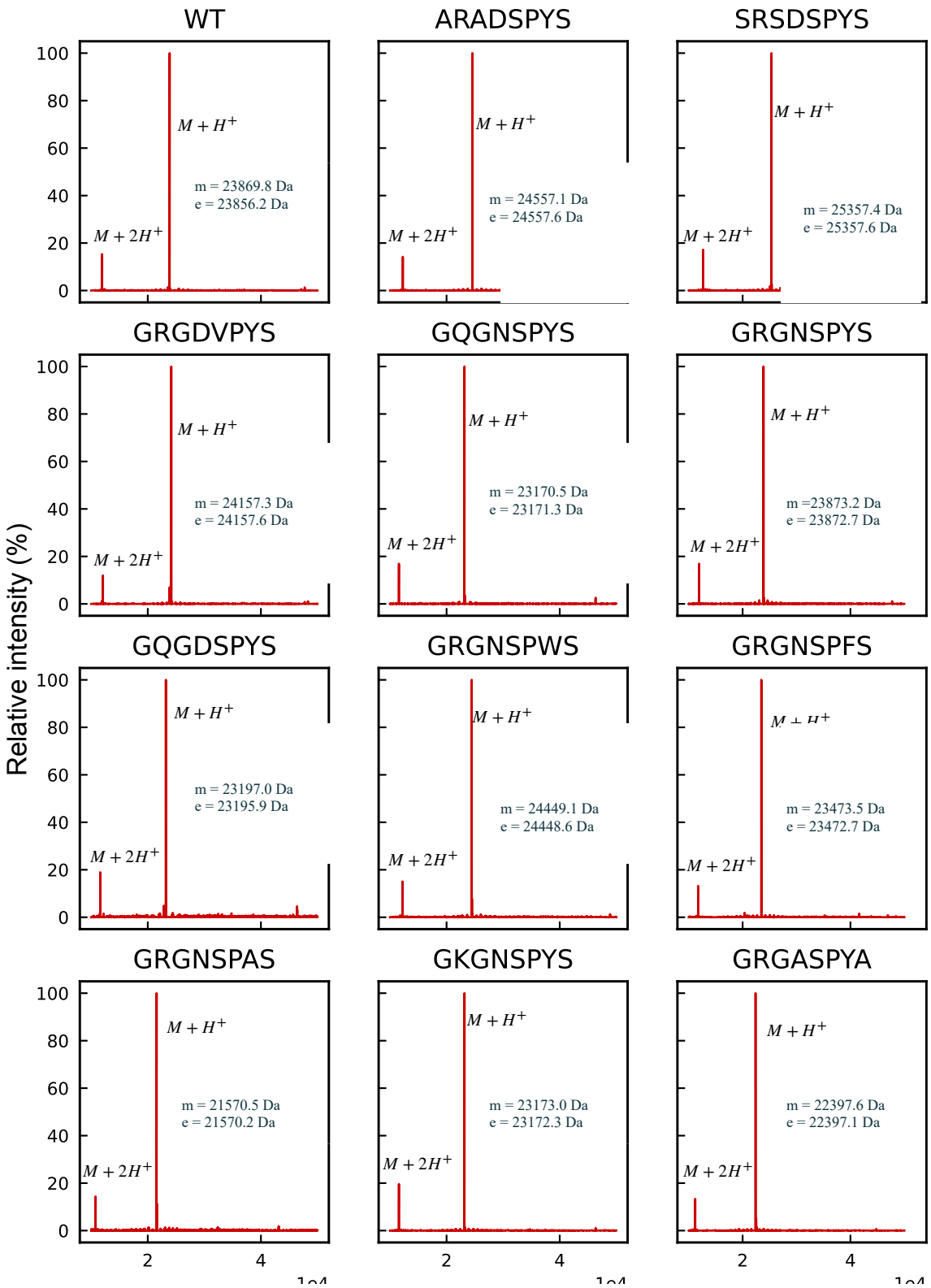


Fig 14. Mass spectra (XEVO) obtained from the UPLC chromatograms from Fig S18. The measured molecular weight is represented by "m" and the theoretical molecular weight is represented by "e".

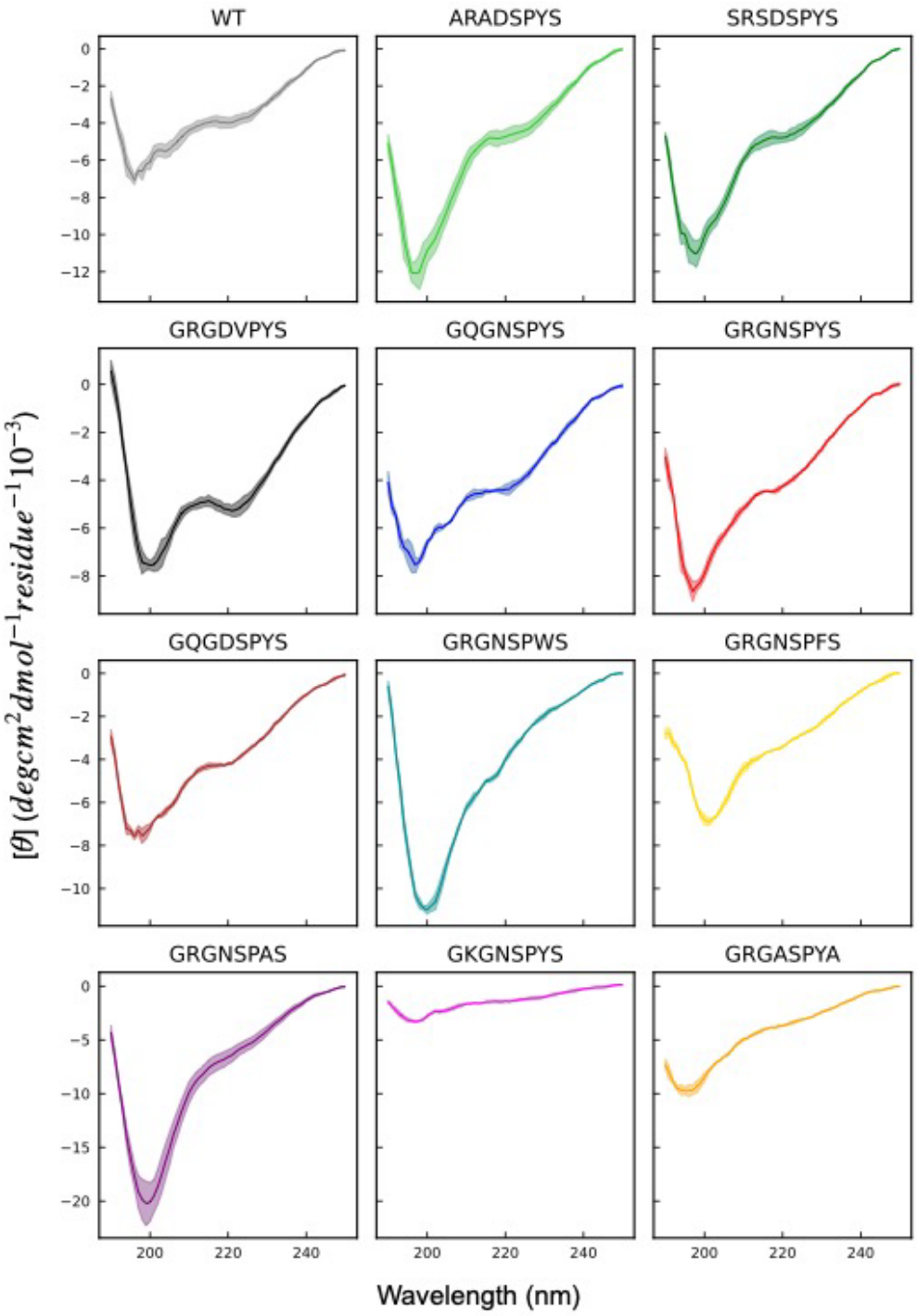


Fig 15. Circular dichroism (CD) spectroscopy (using a Jasco J-1500 CD spectropolarimeter, Jasco Inc., Easton, MD, USA) was conducted to characterize the secondary structure of the RLP sequences. Lyophilized samples were dissolved in DI water at pH 7.4 to a final peptide concentration of 6 µM. The CD spectra were recorded using quartz cells with a 0.2-cm optical path length. The wavelength scans were obtained at 75 °C (above the transition temperature of all the sequences at this concentration) from 190 to 250 nm and were recorded every 1 nm. The spectra show a minimum peak at ~196 nm, characteristic of random coil configurations. Data are presented as mean values +/- SD, n=3 A detailed analysis of the structural contributions to each spectrum is delineated below in Table S4.

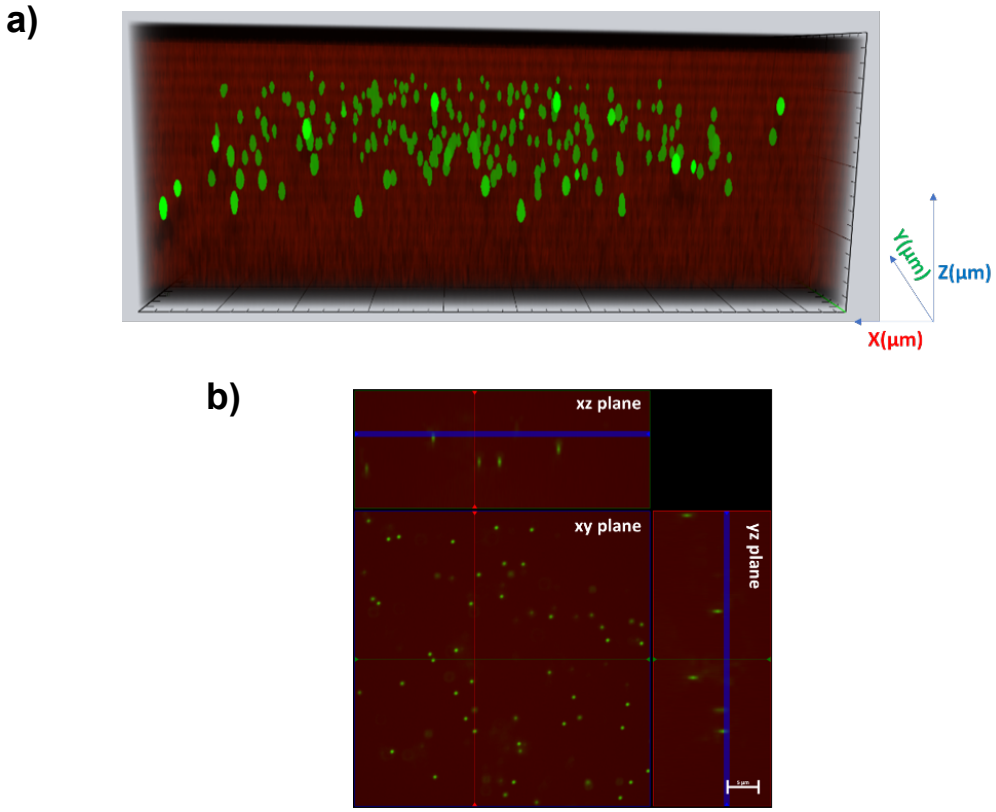


Fig 16. (a) Z-stack of the GRGNSPYS sample with Rhodamine B (red) and 0.5 µm polystyrene beads (green) showing the beads embedded in different z planes of the protein sample. Each major division along the axes corresponds to 5 µm. **(b)** Orthogonal projection of the sample with embedded beads, showing the xy, xz, and yz planes. (Scale bar 5 µm.)

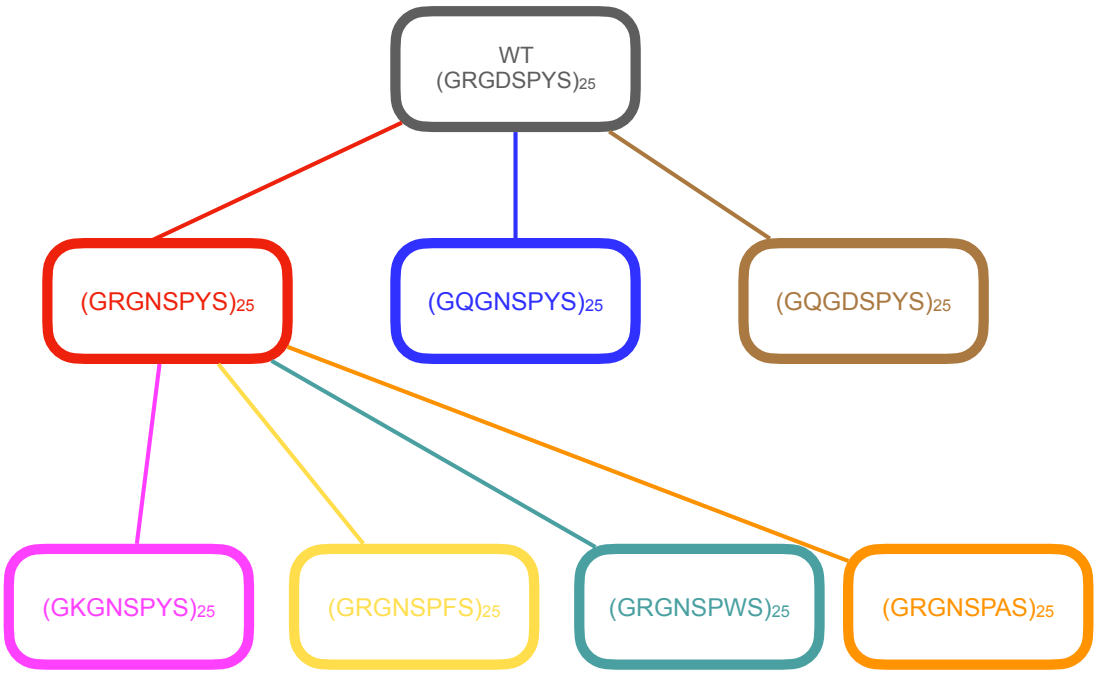


Fig 17. Flowchart representing the different parent and mutant sequences used in the generation of the atomistic slabs for all sequences.

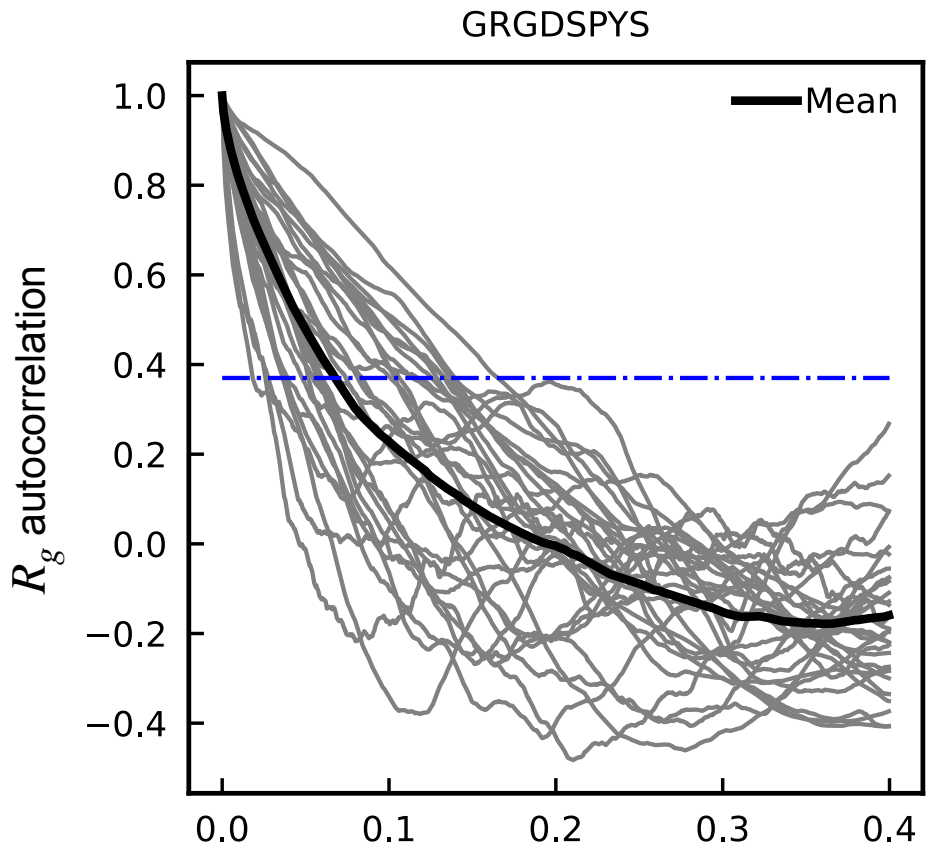


Fig 18. Autocorrelation of the radius of gyration for all chains in the WT atomistic slab. Individual chains are shown in grey; the mean autocorrelation value is shown in black. The horizontal dashed line in blue represents a value of 1/e.

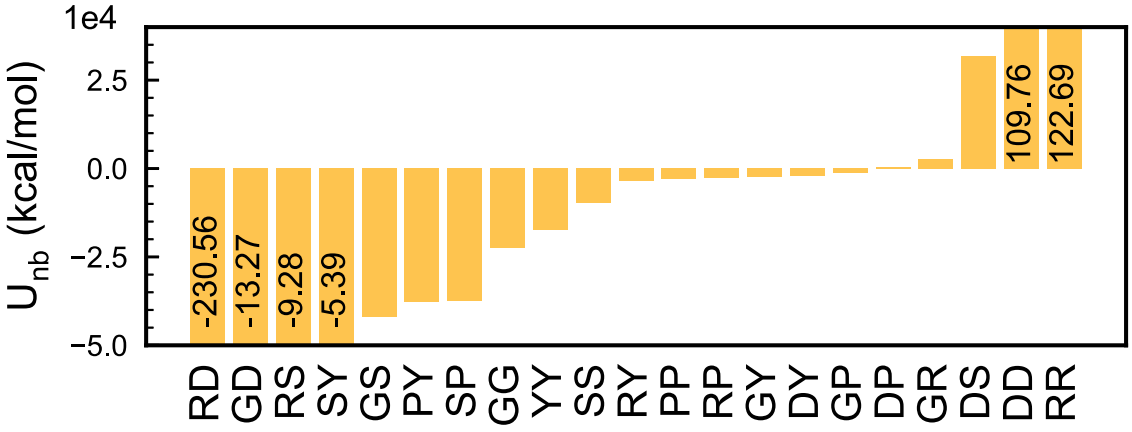


Fig 19. Total pairwise non-bonded potential energy (van der Waals + electrostatic) between residue pairs present in the WT A-IDP estimated from atomistic simulation trajectories. Negative values indicate more favorable interactions between the residue pairs, while positive values indicate repulsion between residue pairs. The numerical values are shown inside the bars for residue pairs with the higher energies.

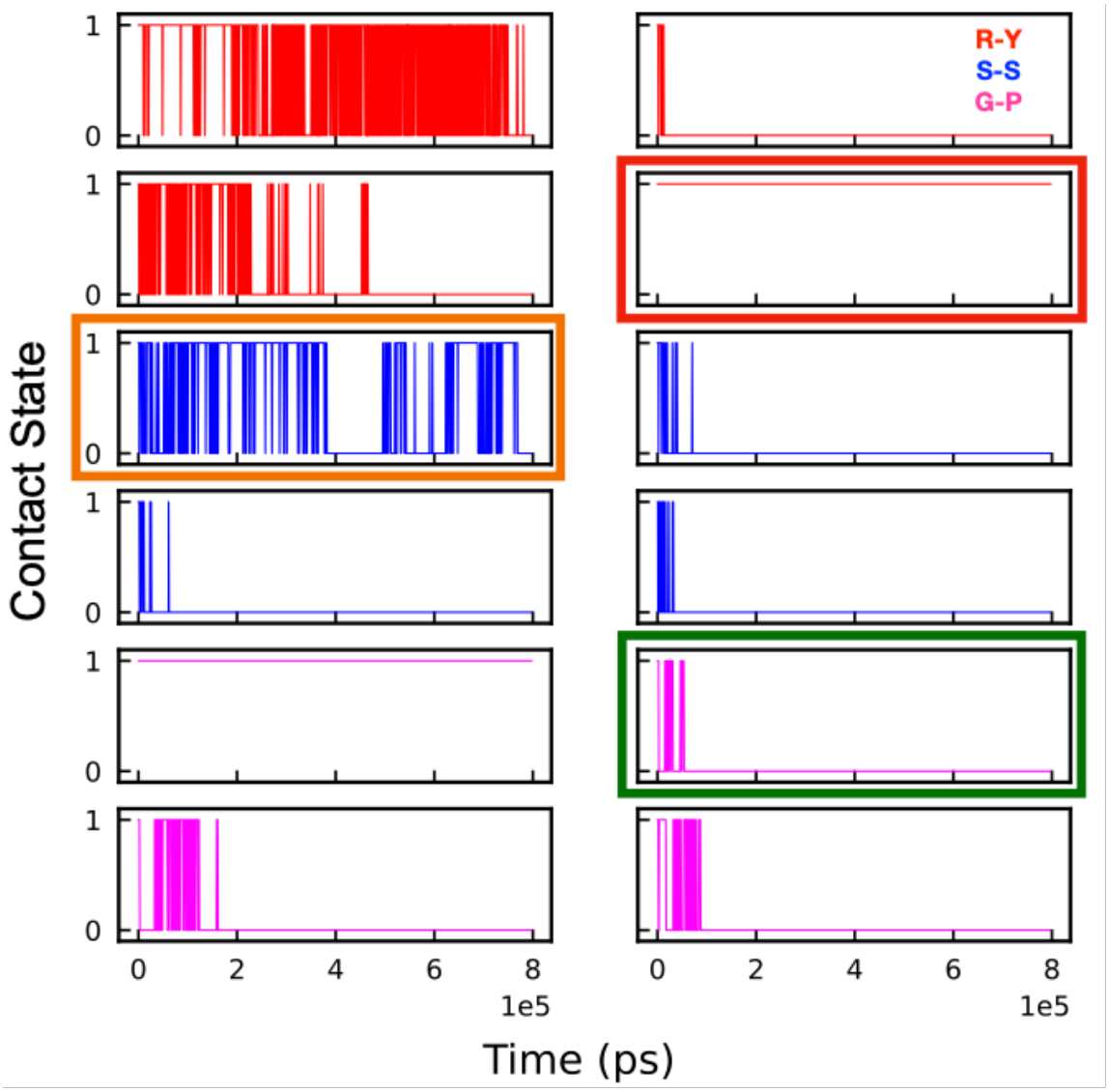


Fig 20. Time evolution of 12 unique intermolecular contacts formed by three residue pairs (R-Y, S-S and G-P) in the atomistic simulation of the WT A-IDP. A value of 1 indicates that the residues are in contact, while a value of 0 indicates that the contact between the residues has broken. The three boxes (red, orange and green) highlight three distinct cases of contact dynamics observed in the simulations. Namely, where contacts dynamically form and break (orange), where a contact forms and breaks transiently and then does not reform again during the simulation (green) and, where the contact forms and then does not break throughout the course of the simulation (red).

Supplementary Tables :

	GRGDSPYS		ARADSPYS		SRSDSPYS		GRGDVPYS		GQGNSPYS		GRGNSPYS	
	Theor	AAA	Theor	AAA	Theor	AAA	Theor	AAA	Theor	AAA	Theor	AAA
Ala	0.0%	1.6%	21.8%	20.7%	0.0%	0.9%	0.0%	1.3%	0.0%	0.4%	0.0%	0.8%
Arg	12.2%	11.2%	12.2%	11.9%	12.2%	12.8%	12.2%	11.7%	1.3%	1.5%	12.2%	12.3%
Asx	11.8%	12.2%	11.8%	12.1%	11.8%	13.0%	11.8%	11.8%	11.8%	12.2%	11.8%	12.1%
Cys	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Glx	1.3%	3.6%	1.3%	2.7%	1.3%	2.5%	1.3%	2.5%	12.2%	11.8%	1.3%	2.1%
Gly	24.0%	21.2%	2.2%	3.1%	2.2%	3.1%	24.0%	21.9%	23.6%	23.8%	23.6%	23.0%
His	2.6%	3.5%	2.6%	3.3%	2.6%	3.1%	2.6%	3.2%	2.6%	3.0%	2.6%	3.0%
Ile	0.0%	1.2%	0.0%	0.9%	0.0%	0.7%	0.0%	0.9%	0.0%	0.3%	0.0%	0.7%
Leu	0.4%	2.0%	0.4%	1.7%	0.4%	1.4%	0.4%	1.6%	0.4%	0.8%	0.4%	1.3%
Lys	0.0%	0.9%	0.0%	0.7%	0.0%	0.5%	0.0%	0.7%	0.0%	0.2%	0.0%	0.4%
Met	0.4%	0.8%	0.4%	0.6%	0.4%	0.7%	0.4%	0.9%	0.4%	0.7%	0.4%	0.8%
Phe	0.9%	1.3%	0.9%	1.3%	0.9%	1.1%	0.9%	1.2%	0.9%	0.9%	0.9%	1.0%
Pro	10.9%	9.0%	10.9%	9.8%	10.9%	10.7%	10.9%	9.5%	11.4%	11.3%	11.4%	10.6%
Ser	23.6%	18.6%	23.6%	18.4%	45.4%	36.3%	12.7%	10.4%	23.6%	20.5%	23.6%	19.3%
Thr	0.4%	1.3%	0.4%	1.1%	0.4%	0.8%	0.4%	1.1%	0.4%	0.6%	0.4%	0.8%
Trp	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Tyr	11.4%	10.0%	11.4%	10.8%	11.4%	11.5%	11.4%	10.5%	11.4%	11.7%	11.4%	11.2%
Val	0.0%	1.7%	0.0%	1.0%	0.0%	0.8%	10.9%	10.9%	0.0%	0.2%	0.0%	0.6%

Table 1. Amino acid composition of the first set of sequences. The blue shading denotes the specific mutation for each sequence. For each sequence, a comparison between the theoretical composition (Theor) and the experimentally determined composition by amino acid analysis (AAA) is shown. Acceptable values range in ±5% of error. The larger variations shown in the table originate mainly from partial amino acid destruction during the hydrolysis procedure (e.g., Ser) and also from small amounts of impurities and quantification error in the limits of the chromatography methodology[3].

	GQGDSPYS		GRGNSPWS		GRGNSPFS		GRGNSPAS		GKGNSPYS		GRGASPYA	
	Theor	AAA	Theor	AAA	Theor	AAA	Theor	AAA	Theor	AAA	Theor	AAA
Ala	0.0%	0.9%	0.0%	0.6%	0.0%	0.7%	10.9%	10.9%	0.0%	0.5%	21.8%	21.7%
Arg	1.3%	1.7%	12.2%	13.0%	12.2%	12.4%	12.2%	12.0%	1.3%	1.6%	12.2%	12.5%
Asx	11.8%	12.3%	11.8%	12.5%	11.8%	12.3%	11.8%	12.2%	11.8%	12.3%	0.9%	1.4%
Cys	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Glx	12.2%	12.4%	1.3%	1.8%	1.3%	2.0%	1.3%	2.8%	1.3%	1.8%	1.3%	1.8%
Gly	23.6%	23.1%	23.6%	24.0%	23.6%	23.4%	23.6%	22.2%	23.6%	24.1%	23.6%	23.5%
His	2.6%	2.6%	2.6%	3.4%	2.6%	3.0%	2.6%	3.8%	2.6%	3.0%	2.6%	3.7%
Ile	0.0%	0.6%	0.0%	0.3%	0.0%	0.5%	0.0%	0.9%	0.0%	0.5%	0.0%	0.3%
Leu	0.4%	1.2%	0.4%	0.8%	0.4%	1.1%	0.4%	1.7%	0.4%	1.0%	0.4%	0.8%
Lys	0.0%	0.5%	0.0%	0.2%	0.0%	0.3%	0.0%	0.7%	10.9%	11.7%	0.0%	0.1%
Met	0.4%	0.7%	0.4%	0.9%	0.4%	0.9%	0.4%	0.8%	0.4%	0.9%	0.4%	0.8%
Phe	0.9%	1.0%	0.9%	0.9%	11.8%	12.0%	0.9%	1.2%	0.9%	1.0%	0.9%	0.9%
Pro	11.4%	10.8%	11.4%	11.0%	11.4%	10.8%	11.4%	10.2%	11.4%	10.6%	11.4%	10.8%
Ser	23.6%	19.7%	23.6%	19.8%	23.6%	19.0%	23.6%	17.8%	23.6%	18.9%	12.7%	10.4%
Thr	0.4%	0.8%	0.4%	0.5%	0.4%	0.6%	0.4%	0.9%	0.4%	0.6%	0.4%	0.5%
Trp	0.0%	0.0%	10.9%	9.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Tyr	11.4%	11.3%	0.4%	0.5%	0.4%	0.6%	0.4%	0.8%	11.4%	11.2%	11.4%	10.4%
Val	0.0%	0.6%	0.0%	0.2%	0.0%	0.5%	0.0%	1.0%	0.0%	0.4%	0.0%	0.3%

Table 2. Amino composition of the second set of sequences. The blue shading denotes the specific mutation for each sequence. For each sequence, a comparison between the theoretical composition (Theor) and the amino acid analysis (AAA) is shown. Acceptable values range in ±5% of error. The larger variations shown in the table originate mainly from partial amino acid destruction during the hydrolysis procedure (e.g., Ser) and also from small amounts of impurities and quantification error in the limits of the chromatography methodology[3].

	Helix	β-sheet		Turn	Disordered
		Antiparallel	Parallel		
GRGDSPYS	3	29.2	6.8	15.5	45.4
ARADSPYS	5	26.8	3.9	17.4	46.9
SRSDSPYS	5.2	26.7	4.6	16.8	46.7
GRGDVPYS	2.6	28.3	7.6	14.9	46.5
GQGNSPYS	4	26.7	6.4	15.1	47.8
GRGNSPYS	3.4	30.8	5.8	15.5	44.5
GQGDSPYS	3.5	28.2	5.9	15	47.4
GKGNSPYS	0	37.4	0	14.6	48
GRGNSPWS	3.5	31.8	2.5	15.3	47
GRGNSPFS	3.8	26.7	4.8	15.5	49.3
GRGNSPAS	8.7	27.3	0	18	45.9
GRGASPYA	4.8	25.6	4.4	16.4	48.8

Table 3. Structural analysis performed in the single spectrum analysis program from BeStSel[4-6]. The method is freely available for academic use at https://bestsel.elte.hu/index.php. The table shows the percentage of each type of structure, confirming that the disordered conformation is the primary contribution to the CD spectra. In addition, the similarities in the contributions of the different secondary conformations indicates that the amino acid substitutions did not lead to significant differences in the intrinsically disordered nature of the variants as compared to the WT sequence, GRGDSPYS.

Sequence	**Viscosity (Pa.s)**
$(GRGDSPYS)_{25}$	10.75 ± 0.53
$(ARADSPYS)_{25}$	11.56 ± 0.87
$(SRSDSPYS)_{25}$	18.91 ± 2.3
$(GRGDVPYS)_{25}$	27.60 ± 1.7
$(GQGNSPYS)_{25}$	5.01 ± 0.26
$(GRGNSPYS)_{25}$	8.35 ± 1.58
$(GRGNSPFS)_{25}$	4.17 ± 0.38
$(GRGNSPWS)_{25}$	38.66 ± 1.65
$(GKGNSPYS)_{25}$	4.86 ± 0.17
$(GRGASPYA)_{25}$	2.32 ± 0.56

Table 4. Noise corrected viscosities measured by passive microrheology for the various polypeptide sequences tested in this study.

References :

1 Schuster, B. S. *et al.* Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. *Proceedings of the National Academy of Sciences* **117**, 11421-11431 (2020).

2 Stott, K., Blackburn, J. M., Butler, P. & Perutz, M. Incorporation of glutamine repeats makes protein oligomerize: implications for neurodegenerative diseases. *Proceedings of the National Academy of Sciences* **92**, 6509-6513 (1995).

3 Rutherfurd, S. M. & Gilani, G. S. Amino Acid Analysis. *Current Protocols in Protein Science* **58**, 11.19.11-11.19.37 (2009).

4 Micsonai, A. *et al.* Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences* **112**, E3095-E3103 (2015).

5 Micsonai, A. *et al.* BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic acids research* **46**, W315-W322 (2018).

6 Micsonai, A., Bulyáki, É. & Kardos, J. BeStSel: from secondary structure analysis to protein fold prediction by circular dichroism spectroscopy. *Structural Genomics: General Applications*, 175-189 (2021).