

Predicting disordered regions driving phase separation of proteins under variable salt concentration

Esteban Meca¹, Anatol W. Fritsch^{2,3}, Juan M. Iglesias–Artola², Simone Reber^{4,5},

Barbara Wagner⁶

submitted: September 21, 2021

¹ Departamento de Física Aplicada
Radiología y Medicina Física
Universidad de Córdoba
14071 Córdoba, Spain
E-Mail: esteban.meca@uco.es

² Max Planck Institute
of Molecular Cell Biology and Genetics
Pfortenhauerstr. 108
01307 Dresden, Germany
E-Mail: fritsch@mpi-cbg.de
jjglesia@mpi-cbg.de

³ Center for Systems Biology Dresden
01307 Dresden, Germany

⁴ IRI Life Sciences
Humboldt-Universität zu Berlin
10115 Berlin, Germany
E-Mail: simone.reber@iri-lifesciences.de

⁵ University of Applied Sciences Berlin
13353 Berlin, Germany

⁶ Weierstrass Institute
Mohrenstr. 39
10117 Berlin, Germany
E-Mail: barbara.wagner@wias-berlin.de

No. 2875

Berlin 2021



2020 *Mathematics Subject Classification.* 82B26, 65H10.

2010 *Physics and Astronomy Classification Scheme.* 87.15.km, 87.15.kr, 87.15.Zg.

Key words and phrases. Liquid-liquid phase separation, random phase approximation, intrinsically disordered proteins, phase-separation assays.

A.W.F. was supported by the ELBE postdoctoral fellows program and the Max Planck Research Network for Synthetic Biology (MaxSynBio) consortium, jointly funded by the Federal Ministry of Education and Research of Germany and the Max Planck Society.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: <http://www.wias-berlin.de/>

Predicting disordered regions driving phase separation of proteins under variable salt concentration

Esteban Meca, Anatol W. Fritsch, Juan M. Iglesias–Artola, Simone Reber, Barbara Wagner

Abstract

We determine the intrinsically disordered regions (IDRs) of phase separating proteins and investigate their impact on liquid-liquid phase separation (LLPS) with a random-phase approximation (RPA) that accounts for variable salt concentration. We focus on two proteins, PGL-3 and FUS, known to undergo LLPS. For PGL-3 we predict that an IDR near the C-terminus promotes LLPS, which we validate through direct comparison with *in vitro* experimental results. For the structurally more complex protein FUS the role of the low complexity (LC) domain in LLPS is not as well understood. Apart from the LC domain we here identify two IDRs, one near the N-terminus and another near the C-terminus. Our RPA analysis of these domains predict that, surprisingly, the IDR at the N-terminus (aa 1-285) and not the LC domain promotes LLPS of FUS by comparison to *in vitro* experiments under physiological temperature and salt conditions.

1 Introduction

Protein condensation driven by liquid-liquid phase separation (LLPS) is a powerful concept to understand the mesoscale organization of cells. It provides a simple mechanism to form non-membrane bound organelles that separate from the nucleo- and cytosol. Examples of such biomolecular condensates include nucleoli, P-granules, stress granules and centrosomes as reviewed in Alberti and Hyman [2], Banani et al. [4], Shin and Brangwynne [43]. These condensates correspond to a protein-rich phase coexisting with a protein-poor bulk phase.

One of the drivers of cellular LLPS are intrinsically disordered regions (IDRs), these can be either part of a protein or constitute the entire protein (intrinsically disordered proteins, IDPs). IDRs are highly dynamic regions within protein sequences that lack stable secondary or tertiary structure. Yet, they facilitate weak multivalent interactions. On the sequence level, driving forces include electrostatic interactions between charged motifs that promote long- and short-range interactions. Short-range interactions are characterized by directional interactions of dipoles or positive charges with aromatic groups (Brangwynne et al. [10]). Thus, the phase behaviour of a given protein is sequence encoded. Condensation, however, does not only depend on protein structure but also on the environmental conditions. These include temperature, ionic strength of the constituents or their concentration. For example, condensates can dissolve upon raising temperature or salt concentration and can reform when conditions are reverted.

Despite increasing experimental evidence, lending insight into the physical chemistry that is driving LLPS, it remains a challenge to directly predict the phase behaviour of a protein based on its primary sequence and solvent environment. This limits our ability to predict how changes in the amino acid sequence of a protein influence its phase behaviour. Therefore, theoretical predictions for phase diagrams are needed to guide experimental research and provide insights into the molecular basis of physiological and pathological processes related to diseases and ageing.

While molecular dynamics simulations provide detailed biophysical information on a single protein level Rauscher and Pomès [38], they quickly become computationally expensive when applied to large ensembles of phase-separating proteins in a solvent. A coarse-grained lattice-based approach is the classical Flory-Huggins theory (Flory [18], Huggins [22]) and its extension to the Voorn-Overbeek theory (Overbeek and Voorn [34]), which incorporates electrostatic interactions via the Debye-Hückel theory. In the derivation of these theories variations in charge patterns that are responsible for phase-separation are averaged out. They thus cannot capture the structure- or even sequence-specific phase behaviour, which is a signature of IDPs. On an intermediate coarse-grained scale, Field Theoretic Simulations (FTS), rooted in statistical mechanics, are able to incorporate structural information of proteins and has recently been used to predict LLPS of tau, see e.g. Zhang et al. [47]. However, since FTS relies on the full partition function of the free energy it is still numerically very demanding. Asymptotic approximations of the full partition function, such as the Random-Phase Approximation (RPA), reliably account for the structural features of proteins. Indeed, Lin et al. [27, 28] have pioneered the application of RPA to LLPS of phase separating proteins to predict the sequence-specific phase behaviour of the RNA helicase Ddx4, see also the recent review by Dinic et al. [15]. Here, we show that in order to test the phase behaviour under physiological conditions and at the same time obtain a thermodynamically consistent theoretical model, it is necessary to introduce salt concentration as an additional variable into the RPA ansatz. We thus derive phase diagrams for the temperature-protein as well as for the salt-protein concentrations as projections of a 3D (*pointy*) phase diagram to analyse the phase behaviour of two proteins, PGL-3 and FUS, known to undergo LLPS *in vitro* and *in vivo* (Brangwynne et al. [9], Wang et al. [44]). By matching to the exact conditions of *in vitro* experiments, we determine the IDRs that drive LLPS of these proteins.

We first investigate PGL-3, since its phase behaviour is well understood (Brangwynne et al. [9], Saha et al. [40]) and validate our predictions with experimental data of the dilute and condensed phase concentration from *in vitro* studies under physiological salt conditions. Our results confirm that the IDR at the C-terminus drives LLPS of PGL-3. We then focus our analysis on FUS, where the driving forces and sequence domains responsible for LLPS are still under debate (Patel et al. [35], Wang et al. [44]). We identify and analyse three regions as possible candidates to impact the phase behaviour of FUS, the LC region, an IDR at the N-terminus and an IDR at the C-terminus.

Our analysis reveal that the domain at the N-terminus, from amino acids 1 to 285, to be responsible for LLPS when comparing to *in vitro* experimental data of FUS.

2 Results

The phase behavior of a given protein is encoded in its free energy function f ,

$$f = -S + f_{el}.$$

As indicated in Fig. 1, the free energy has an entropic part S , representing the Flory-Huggins interactions

$$-S = \frac{\phi_{aa}}{N} \log \phi_{aa} + \phi_c \log \phi_c + 2\phi_s \log \phi_s + \phi_w \log \phi_w,$$

where N denotes the number of amino acids. Apart from the volume fraction of the amino acids of the protein ϕ_{aa} , the volume fraction of counterions ϕ_c and the volume fraction ϕ_w of water, we also include in our analysis the volume fraction of salt ions KCl (ϕ_s) as an additional variable. The second term f_{el} represents the electrostatic multivalent interactions of protein chains with each other and the surrounding salt solution, and drives LLPS here. Thus, it is this part of the free energy for which

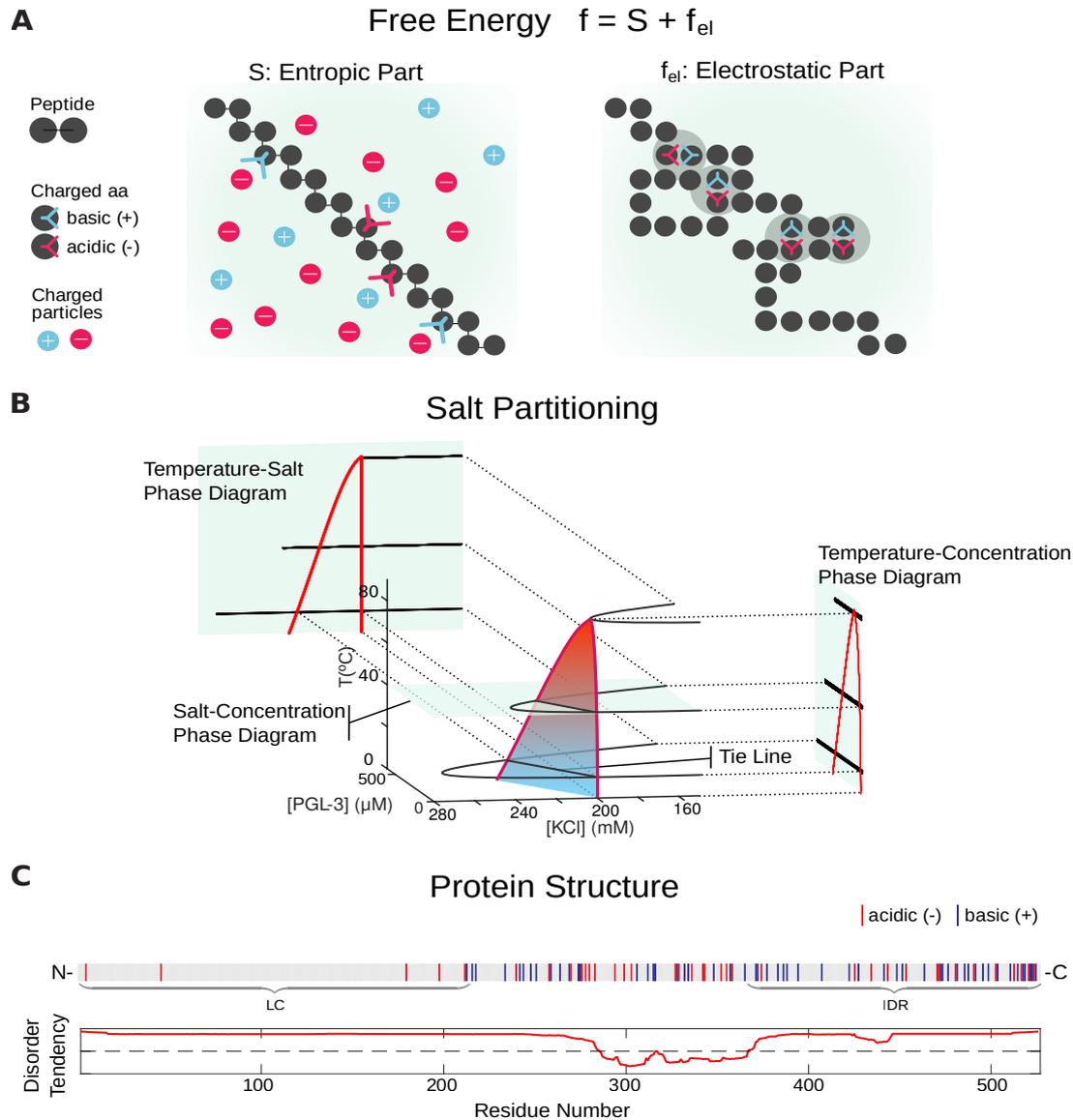


Figure 1: Random Phase Approximation (RPA) to predict the phase behaviour of IDPs.

A: Schematic of short range entropic and long range electrostatic interactions that combine to drive liquid-liquid phase separation in solution, the residues of a peptide chain can be part of several possible multivalent electrostatic interactions. **B:** Our RPA model allows for variable salt concentration and can account for salt partitioning. The corresponding phase diagrams typically show “pointy” shapes. **C:** Identification of domains that may drive LLPS of the protein.

the Random Phase Approximation (RPA) is being applied in order to account for the dependence of the free energy on the protein structure (see Field theoretic approach in the Methods and Materials section). After we determine the complete structure of the protein, we derive the temperature-(protein) concentration and salt-concentration phase diagrams, as sketched in Fig. 1, based on the free energy function f for the proteins (*C. elegans*) PGL-3 and hFUS (see Computation of phase diagrams in Methods and Materials section).

LLPS in PGL-3 is accompanied by salt partitioning We first test the predictions of this model with experimental results for a range of different bulk concentrations of recombinant PGL-3 in an *in vitro* phase-separation assay. The experiments exhibit the generic behaviour, that higher PGL-3 concentrations are necessary at increasing temperatures in order to initiate LLPS at physiological salt (150 mM KCl) concentration. This is quantified at each temperature with a corresponding protein saturation concentration c_{out} (Fig. 2 B,C).

To investigate the role of intrinsically disordered regions (IDRs) of PGL-3, we determine (using MetaDisorder MD2) four small disordered regions (IDRs) and one large IDR (IDR-5: aa 515-693) near the C-terminus, which we denote by PGL-3-C1, see Fig. 2A. It is the longest linear sequence predicted to lack a secondary structure. Indeed, this region has previously been shown to include a set of six C-terminal RGG repeats (aa 633-695), which bind RNA and promote droplet formation Saha et al. [40]. For the derivation of the binodal, specifically the RPA of the electrostatic interactions (f_{el}), we use the longest IDR PGL-3-C1.

For the quantitative comparison, we need to relate the PGL-3 volume fraction ϕ_{aa} and the non-dimensional temperature T^* in terms of the experimentally accessible PGL-3 concentration (c) and temperature (T), respectively. This means, in the case of ϕ_{aa} , we need a factor that turns the number of amino acids per unit volume into a volume fraction. For that purpose, the concentration of the protein is multiplied by the number of amino acids in the protein (693) and divided by the total number of sites in the lattice per unit volume. This latter number is conventionally taken to be 55.5M, 1 over the volume occupied by a water molecule. Hence, for PGL-3 in the presence of KCl

$$\phi_{aa} = \frac{693}{55.5\text{M}}[\text{PGL-3}] \quad \text{and} \quad \phi_s = \frac{1}{55.5\text{M}}[\text{KCl}]. \quad (2.1)$$

For the non-dimensional temperature T^* we obtain

$$T^* = \frac{4\pi\epsilon_0\epsilon_r k_B a}{e^2} T, \quad (2.2)$$

where ϵ_0 is the vacuum permittivity, ϵ_r is the relative permittivity of the medium, k_B is Boltzmann's constant, a the link length and e is the charge of the electron. There are two unknown parameters in Eq. (2.2), a and ϵ_r . We take a to be the C α -C α virtual bond length of 3.8Å and fit ϵ_r , due to the lack of a complete theory to derive its value. The value for the relation between the concentration of the protein and the amino acid volume fraction also needs to be fitted. Not only is the volume of an amino acid clearly different from that of a water molecule, but also the effective volume of the protein is not necessarily a simple linear function of the number of amino acids. We thus obtain a fit for the volume fraction and the non-dimensional temperature using only two scaling parameters through $T = \beta_1 T^*$ and $c = \beta_2 \phi_{aa}$ by using the functional relationship $\phi_{aa} = f(T^*)$ from the phase diagram and a fitting procedure that allows β_1 to vary with the salt concentration on account of the strong dependence of the permittivity on salt concentration. Details of the parameter estimation and fitting procedure are given in the Methods and Materials section.

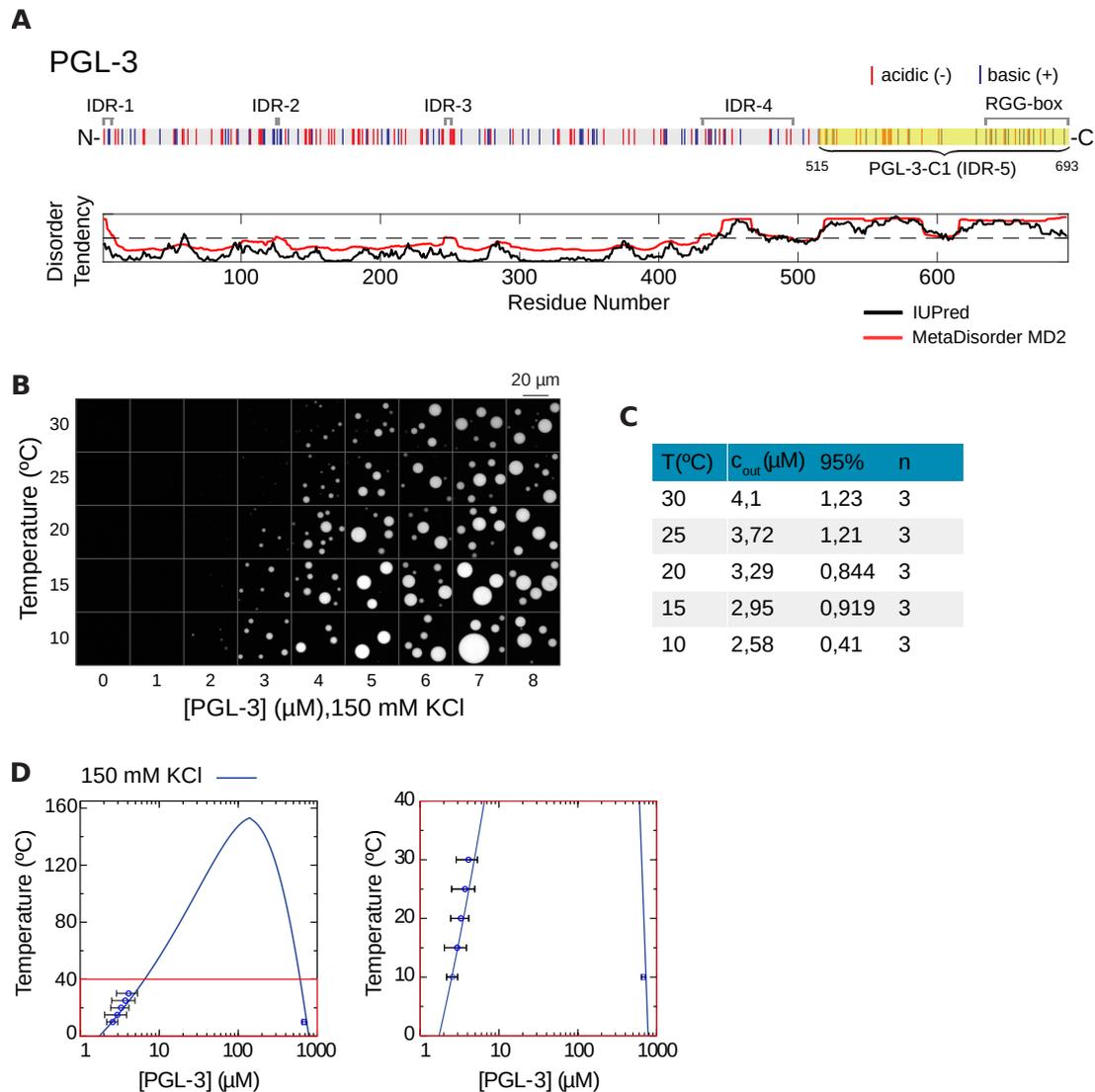


Figure 2: Phase separation of PGL-3

A Sequence of *C. elegans* PGL-3 with negatively (red) and positively (blue) charged amino acids. MetaDisorder MD2 (red) and IUPred (black) identify five disordered regions with IDR-5 (PGL-3-C1) at the C-terminus being the longest and RGG-box as in [40]. PGL-3-C1 (highlighted) is used for the computations of the phase diagrams below. **B** Experimental data show the temperature- and concentration-dependent phase separation of PGL-3 at physiological salt conditions of 150 mM KCl. **C** Quantification of data in **B** to derive c_{out} . n is the number of repetitions of the experiment, 95% corresponds to the confidence interval. **D** Predicted temperature-concentration phase diagram for PGL-3-C1 based on RPA. The phase diagram is computed with $\varepsilon_r = 33.6$ and $\beta_2 = 16.2$ mM, using Eqs. (2)-(7) (see Methods and Materials). To the right is a zoom of the red region. Figure supplements on LLPS for constant salt can be found in the appendix **A**.

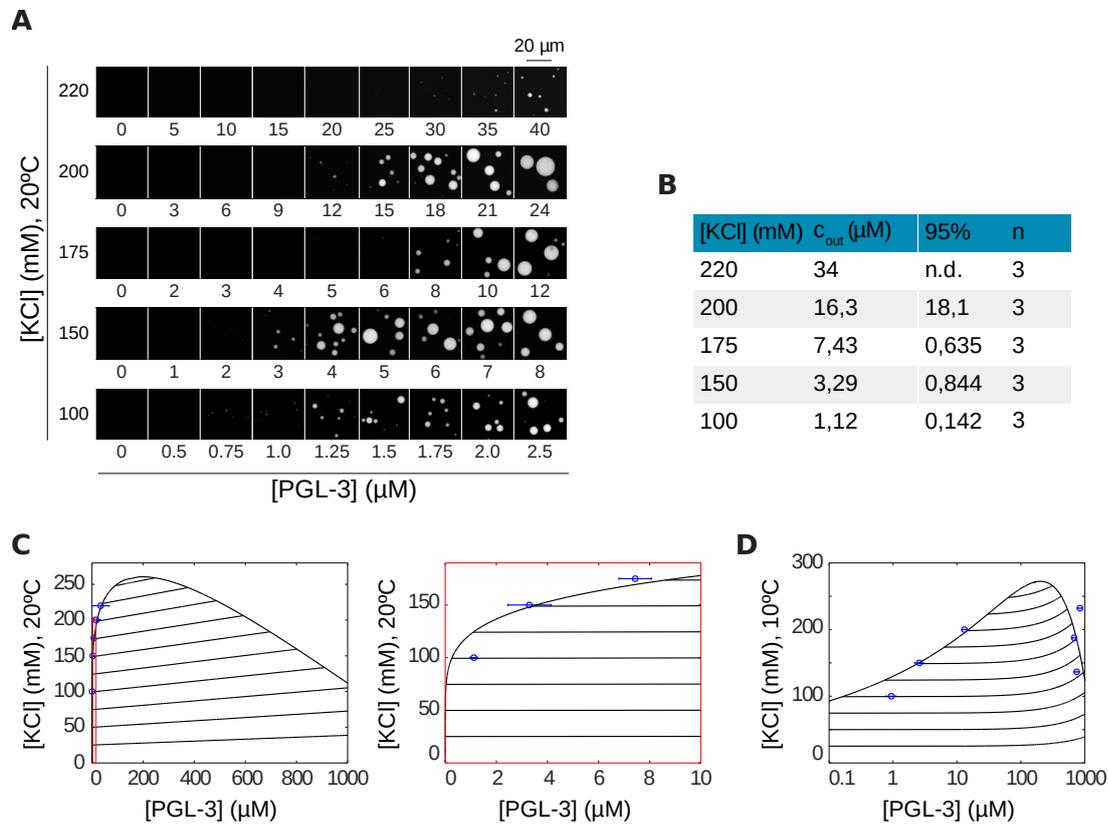


Figure 3: Salt partitioning and the domain PGL-3-C1 capture LLPS for PGL-3.

A Salt- and concentration-dependent phase separation of PGL-3 at 20 °C. They show higher concentration of PGL-3 is needed for higher salt concentrations in order to initiate phase separation. **B** Quantification of data in **A** to derive c_{out} with $n = 3$ repetitions, 95% denotes the corresponding confidence interval. **C** Predicted salt-concentration phase diagram based on PGL-3-C1 computed with the same parameter values and methods as in Fig. 2. Comparison to experimental data are shown for the dilute branch at 20 °C. Panel to the right (red) is a zoom into the left panel. Also shown are the theoretically predicted tie lines showing a slightly positive slope. **D** Predicted salt-concentration phase diagram based on PGL-3-C1 computed as in **C** but at 10 °C and compared to experimental data for both (dilute and condensed) branches. Note, the graph is shown in semi-logarithmic scale, including the tie lines.

The theoretical phase diagram in Fig. 2 D reaches its maximum near 160°C, where the *pointy* feature can be clearly observed. When we limit the temperature to the more meaningful 40°C (Fig 2 D, right panel), we observe that the overlaid experimental points align very well along the dilute branch, particularly at lower temperatures. The value of the concentration for the theoretical condensed branch at 10°C does not fall strictly within the 95% confidence interval, but it is nevertheless very close. Considering that we only have two parameters at our disposal, we can consider this agreement of experiments and theory very good.

Our theoretical model allows for variable salt concentration. As a consequence, the protein-poor and protein-rich phases may have different equilibrium salt concentrations, which implies in turn that the tie lines connecting both equilibria have a non-zero slope. We note that this is responsible for the characteristic *pointy* shape of the temperature phase diagram in Fig. 2 D. It is a slice of the complete 3D phase diagram (shown in the supplemental Fig. S1). It also reflects the experimental procedure, where we start from a solution of a given (fixed) salt content, which implies a constant salt concentration in the protein-poor branch and a varying salt concentration in the protein-rich branch. Starting with the protein-poor branch we analyze how salt affects the saturation concentration c_{out} . Also here we obtain a generic behaviour that for higher salt concentration we need higher protein concentrations for phase separation of PGL-3 to occur (Fig. 3 A). Compared to the dependence we saw for temperature, however, our experimental results suggest that salt seems to have a stronger influence on c_{out} . Specifically, in the range from 100 to 220 mM KCl at 20 °C, c_{out} changes approximately 30-fold from 1.12 to 34 μ M (Fig. 3 B). For the quantitative comparison against experimental results for the salt-concentration phase diagram we use the now already determined scaling parameters to derive the phase diagram from our theoretical model. The resulting theoretical curve at 20°C (Fig. 3 C) has a maximum near 250 mM KCl, a salinity above which we do not expect LLPS. The agreement of the overlaid experimental points with the theoretical dilute branch (Fig. 3 C, right panel) is very good, with the discrepancy with the confidence interval stemming from the fact that the value of the permittivity varies strongly with salt concentration, and we are considering it to be the value fitted for 150 mM KCl. At the lower temperature of 10°C we observe a similar good agreement between theory and experiments in the dilute branch (Fig. 3 D) and on the condensed branch, again, considering we only have two parameters at our disposal.

Note that our theoretical model provides a prediction for the slope of the tie lines and the salt content of the protein-rich branch. As an example, for 10 °C and a salt concentration of 150 mM in the protein-poor branch, we obtain a slope of the tie line equal to 51.9 mol/mol and a salt concentration of 187 mM in the protein-rich branch. We expect to test these results in our future experiments.

FUS-N1 domain is responsible for LLPS We have shown, in the case of PGL-3, how the prediction of its temperature vs. protein concentration or salt vs. protein concentration phase diagrams can be achieved using only minimal experimental input data. However, the actual predicting power of our approach is revealed for structurally more complex proteins such as FUS. In contrast to PGL-3, which has only one long, continuous IDR, the application of the same machine learning tools yields that FUS has two long IDRs. While it is known that FUS contains a prion-like low-complexity domain [24] (LC) (aa 1-214), which overlaps with the IDR near the N terminus (Fig. 4A) (aa 1-285, FUS-N1) and is mostly devoid of charged amino acids, we find an additional IDR near the C terminus (Fig. 4A) (aa 367-526, FUS-C1). The role of these domains in LLPS of FUS is still not known and in particular the role of the LC domain is still debated. In the following, the comparison of the theoretical and experimental results is performed using FUS-N1, which gives the smallest overall error (see Table 1 in Methods and Materials), but below we also explore the consequences of using the other IDRs we have identified. We show that we can not only predict if FUS undergoes LLPS by itself under physiological conditions,

but we can now test whether our approach can determine which domain is responsible for LLPS of FUS.

In a similar fashion as for PGL-3, we first investigate experimentally the temperature-concentration phase diagram in order to fit the parameters of the model. We observe that the influence of temperature on the saturation concentration c_{out} is comparable to the change we see for salt, with a 3.4-fold increase from 0.36 to 1.21 μM between 10 and 25 $^{\circ}\text{C}$. (Fig. 4B,C). In the experiments with FUS we observe a weak salt dependence of c_{out} in a range of 100-200 mM KCl with a 2.5-fold change from 0.5 to 1.26 μM (Fig. 4D, E). Considering FUS-N1 (IDR-N1, highlighted in Fig. 5A), the experimental results show very good agreement with the theory (Fig. 5B). The theoretical phase diagram has a maximum near 150 $^{\circ}\text{C}$ and has an even more striking *pointy* feature. If we zoom in to the region with a biologically meaningful temperature (Fig. 5A, right panel), we can see how well the overlaid experimental points fall upon the theoretical dilute branch. In this case, the experimental point at 10 $^{\circ}\text{C}$ also falls directly upon the condensed branch, thus giving an even better fit than PGL-3, which is manifest in the value of the χ^2 parameter (see Table 1 in Methods and Materials).

As in the PGL-3 case, the experimentally fitted parameter values obtained for the temperature-concentration phase diagram are also used to generate the salt-concentration phase diagram for comparison to the experimental data. The overlay of experimental points and the theory captures the trend of the change of concentration with the salinity in the dilute branch (Fig. 5C). Again the discrepancy in the values of the concentration stems from the fact that we use the value of the permittivity at 150 mM KCl but we know that its value depends strongly on salt concentration. In the condensed branch we obtain a very good agreement between theory and experiments (Fig. 5D), probably on account of the smaller variation of the salinity. For 10 $^{\circ}\text{C}$ we can compute the slope of the tie line for a salt concentration of 150 mM in the dilute branch, giving 25.6 mol/mol, and a predicted salt concentration of 366.4 mM, which we also expect to validate experimentally in the future.

We also investigated the FUS-LC domain and found that the theoretical curve cannot be fitted to allow for comparison with the experimental results, since the scaling factor involves a permittivity smaller than 1 as shown in Table 1 in Methods and Materials. Thus, interestingly, the LC domain should not play a dominant role in LLPS of FUS. Once we discard the FUS-LC region, a further IDR should be considered according to the results from MetaDisorder, FUS-C1 (highlighted in Fig. 5E). In principle one could compare the fits and reason in terms of the goodness of fit (i.e. the χ^2 parameter given in the Methods and Materials section) to decide which region is most likely responsible. But a side by side comparison of the overlay of the experimental data points for FUS-N1 and FUS-C1 shows a much more clear-cut result, which is that the experimental results cannot be fitted at all to the FUS-C1 curve (See Fig. 5F). This is a consequence of our parsimonious fitting approach. The ratio of the protein concentration in the condensed branch with the concentration in the dilute branch obtained from the model is fixed and cannot be adjusted. The approximate maximum value of this ratio (and hence of its logarithm) is marked as $\log(c_{in}/c_{out})_{\text{max}}$ in Fig. 5F, left panel. Thus, when one of the two branches for FUS-C1, in this case the dilute branch (Fig. 5F, left panel) has been fitted, it may leave a difference in the other branch, which here turns out to be one order of magnitude between the theoretical value and the experiments for the condensed branch, i.e. the experimental ratio of c_{in} to c_{out} is one order of magnitude greater than its maximum value. Also, since in our model we allow for salt partitioning, the model does not simply produce a family of curves where the experimental results can always be fitted, but rather a very constrained form of the phase diagrams (Fig. 5F, right panel).

This surprising result is in clear contrast with models where the salinity is considered constant in both phases (as shown in Fig. S1) and thus clearly shows the predictive power of our model, discarding regions for being responsible for LLPS while giving a testable prediction for the salinity of the condensed

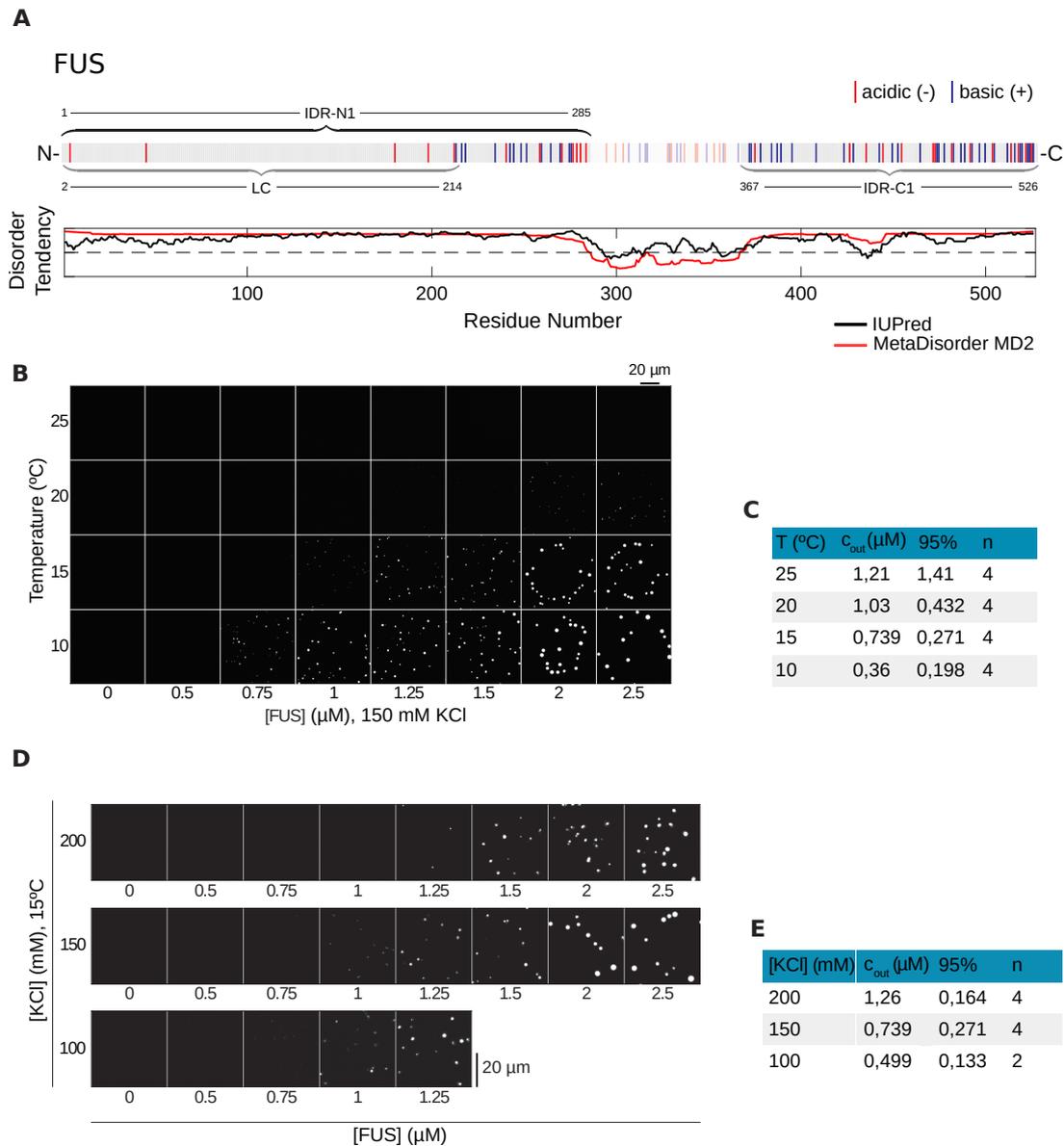


Figure 4: FUS, a protein with several disordered domains where LLPS depends weakly on salt concentration.

A Sequence of hFUS with predicted N-terminal IDR (IDR-N1), which contains the well-described low-complexity (LC) region [44], and a C-terminal IDR (IDR-C1). **B** Temperature- and concentration-dependent LLPS of hFUS-mEGFP at 150 mM KCl. **C** Quantification of data in **B** to derive c_{out} with $n = 4$ repetitions, 95% denotes the corresponding confidence interval. **D** Salt- and concentration-dependent phase separation of hFUS-mEGFP show a weak salt dependence of LLPS. **E** Quantification of data in **D** to obtain c_{out} . Column n corresponds to the number of repetitions and 95% denotes the corresponding confidence interval.

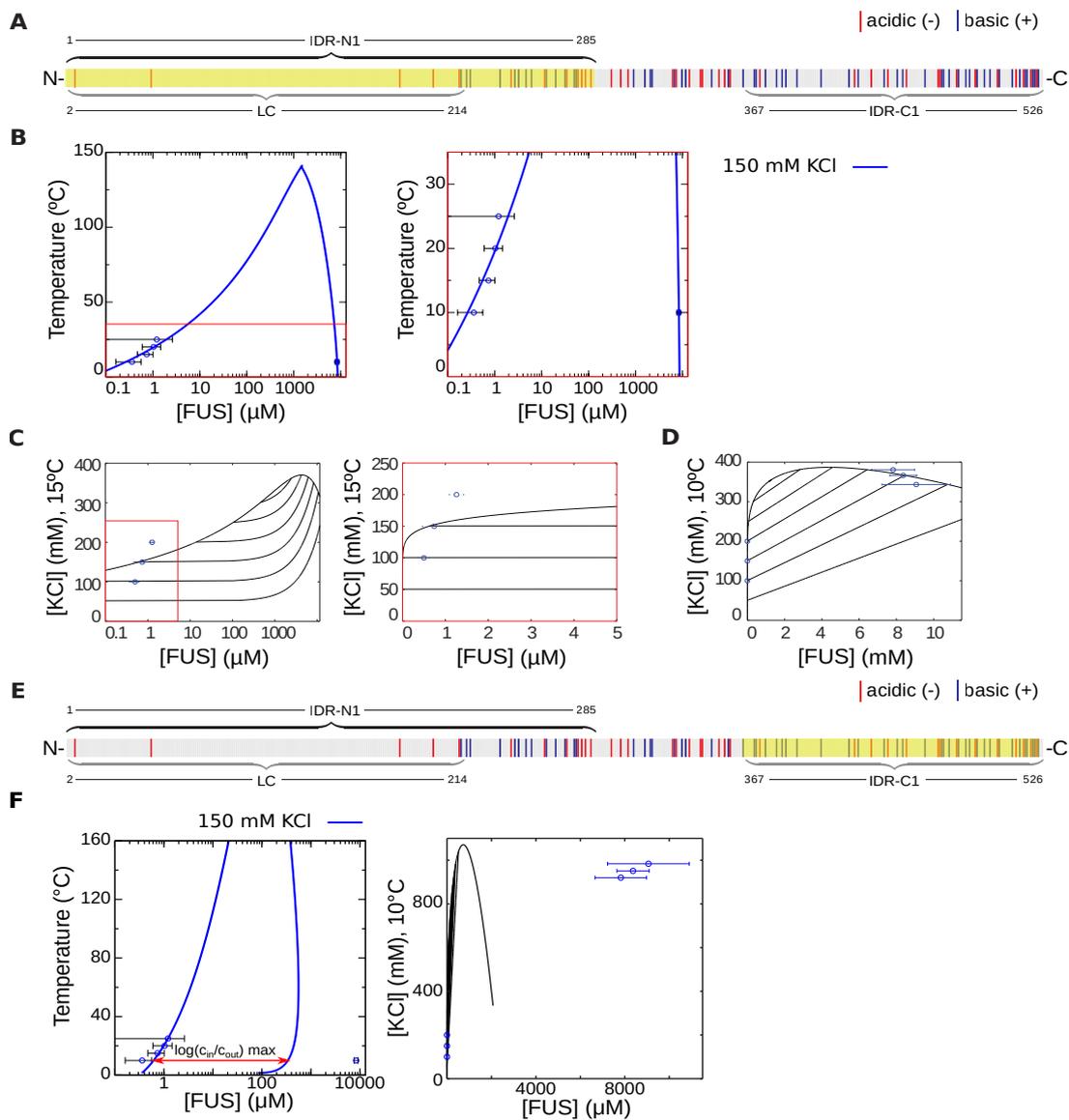


Figure 5: The domain FUS-N1 is responsible for LLPS in FUS.

A Intrinsically disordered region FUS-N1 (IDR-N1, highlighted) **B** Predicted temperature-concentration phase diagram based on RPA analysis using FUS-N1 with parameter values $\varepsilon_r = 8.76$ and $\beta_2 = 207$ mM, for 150 mM KCl. **C** Predicted salt-concentration phase diagram for $T = 20^\circ\text{C}$ obtained using the same parameter values as in **B**. The right panel is a zoom to the red square in the left panel. Experimental data is shown for the dilute branch **D** Predicted salt-concentration phase diagram for $T = 20^\circ\text{C}$, same parameters as **B**. Experimental data is shown for the dilute and condensed branches. **E** Intrinsically disordered region FUS-C1 (IDR-C1, highlighted) **F** Phase diagrams in the FUS-C1 case. Left panel: theoretical temperature-concentration diagram at 150 mM KCl obtained with parameter values $\varepsilon_r = 7.62$ and $\beta_2 = 10.1$ mM. Right panel: salt-concentration diagram obtained with the same parameters. Notice that the experimental points cannot be fitted to the condensed branch of the model, since the ratio of c_{in} to c_{out} is much greater than its maximum value (see discussion in the main text).

phase.

3 Discussion

In this study we investigated the disordered regions of (*C. elegans*) PGL-3 and hFUS regarding their role in LLPS using an RPA for the electrostatic interactions and including variable salt concentrations. By direct comparison of the resulting theoretical predictions of the phase behaviour with *in vitro* experimental results under physiological salt conditions we show that the model is capable to identify specific domains that trigger LLPS.

For FUS, the role in phase separation of the different domains, such as the LC domain, has been controversial, starting with its structural identification. In fact, the LC domain was first identified in a bioinformatics survey for prion-like domains in different proteins Alberti et al. [1], Cushman et al. [12], identifying the LC region to amino acids 1 to 239. Kato et al. [24] identified this region using SEG Wootton and Federhen [45] to be amino acids 2 to 214. We have adopted their definition, as it has become a standard in the subsequent literature, but it has not been possible to reproduce their result using SEG. Kato et al. [24] identified the FUS-LC region as responsible for the protein hydrogel formation by performing experiments with the excised domain, but used a very large concentration, ≈ 2.3 mM. Later, Patel et al. [35] proved that the LC region is necessary for phase separation, since the protein with that region excised would not phase-separate. Interestingly, they also found that a single mutation in the LC domain does not change the phase diagram (even if it has a strong effect on the kinetics of the transformation).

Our results on FUS clearly single out the IDR N1 at the N-terminus (1-285) to lead to phase separation of the full protein in comparison with experiments, while other candidate regions such as the LC or IDR at the C-terminus either requires a permittivity smaller than one or can not be fitted to the experimental data at all. These results are in disagreement with Kato et al. [24] but Wang et al. [44] and Luo et al. [30] support our findings. They showed that the LC domain needs indeed very high concentrations in order to phase-separate on its own, which was the regime tested by Kato et al. [24]. Wang et al. [44] further ascertain that it is the interaction of the LC region and the RNA binding domain what makes the phase transition possible at low protein concentrations. In other recent studies Benayad et al. [5], Dignon et al. [14], coarse grained models in combination with MD simulations have been used, in particular in Dignon et al. [14] it is shown that the longer a sequence containing the LC is, the higher is the propensity towards phase separation. In accordance with our results, Kang et al. [23] demonstrate experimentally that an extended LC domain containing the RGG region next to it phase separated at a much smaller saturation concentration than LC. The domain reported by Kang et al. [23] is very similar to our N1 domain, thus supporting our theoretical results. Moreover, our results show that the phase diagram corresponding to the FUS-C1 region cannot be fitted to the experimental results, thus showing that our model is able to discard IDRs as responsible for phase separation in different ways. These findings suggest further experimental studies on the role of IDR N1, specifically on the role of the amino acids that are lacking in the LC region. It would also require to systematically explore the impact of variations as well as mutations of these regions in our future theoretical studies.

We further note that the connection of the domain structure of proteins with their phase behaviour may be rooted in the underlying model system for polyampholytes, that we used in our analysis. Indeed, Lytle et al. [31] investigated the impact of the blockyness of polyampholytes on their phase behaviour, and Das et al. [13] compared results from Monte Carlo simulations to RPA for specific polyampholyte sequences showing that patterns of larger blocks of charge lead to significantly higher tendency to

phase separate.

We have also shown the impact of salt distribution on the resulting phase diagram and obtained salt partitioning into the condensed phase for both proteins, taking into account the strong salt dependence of the permittivity. The inclusion of salt as a further variable in the model led to phase diagrams with a characteristic *pointy* feature. We showed that this is a generic feature and is rooted in the fact that salt concentration is not imposed to be constant everywhere, but is allowed to vary in the condensed phase. In previous studies this fact has often been neglected but, in fact, it is thermodynamically not consistent to do so.

The experimental corroboration of our predictions on the salt concentration in the condensed phase will be part for our future studies, requiring a detailed discussion of the theoretical model for polyampholyte solutions that we used as a model system for proteins. Here, we remark that the question of salt partitioning during LLPS is also not completely understood, even for polyelectrolytes and polyampholytes. This is the case in the context of Random-Phase-Approximation, Liquid-State-Theory, Monte Carlo simulation and Voorn-Overbeek theory, as well as in experimental studies [26, 32, 33, 36, 37, 42, 46]. Voorn-Overbeek theory predicts an excess of salt in the condensed phase, corresponding to a positive slope of the tie line, while more recent theoretical and experimental studies find that excluded volume effects are responsible for expelling the salt counterions into the dilute phase, thus predicting a negative slope of the tie lines. However, for low overall salt concentration this can be reversed, see for example Li et al. [26].

While RPA is the appropriate tool to address the structural properties of IDPs Romyantsev et al. [39], recent discussion in Das et al. [13], where explicit chain simulation and RPA are compared for a number of polyampholyte sequences, suggest higher order contributions of the functional integral of the partition function in order to address the accuracy of RPA (Borue and Erukhimovich [7], Lin et al. [27], Shen and Wang [41]), specifically in the protein-poor phase.

We also note that further physical interactions also play a role and are still being discovered Berry et al. [6]. Currently, our model does not account for non-specific contacts between positively charged arginine-glycine-glycine (RGG) domains, such as those found in FUS or PGL-3, and negatively charged RNA, which can strengthen the binding affinity of existing RNA binding domains and could provide alternative interaction modes. Also, heterotypic interactions with other regions of the same polypeptide or other proteins are known to drive phase separation Wang et al. [44] but can be included in principle into our framework.

4 Methods and Materials

PGL-3 and PGL3-GFP protein purification PGL-3 was purified from insect cells according to [40]. SF9-ESF cells were infected with baculovirus containing the PGL-3-GFP-6HIS protein under the polyhedrin promoter. Cells were harvested after 3 days of infection by centrifugation at 500 x g for 10 min and then resuspended in lysis buffer (25 mM HEPES 7.25, 300 mM KCl, 10 mM imidazole, 1 mM DTT, 1 protease inhibitor). Cells were lysed by passing the cells 2 times through the LM20 microfluidizer at 15 000 psi. The lysate was then centrifuged at 20 000 rpm for 45 min at 15 °C. The lysate was loaded in a pre-equilibrated Ni-NTA column with lysis buffer at 3 mL/min. The Ni-NTA column was rinsed with 10 C.V of wash buffer (25 mM HEPES 7.25, 300 mM KCl, 20 mM imidazole, 1 mM DTT, 1) and the protein was eluted in 1.5 mL fractions with elution buffer (25 mM HEPES 7.25, 300 mM KCl, 250 mM imidazole, 1 mM DTT). After elution the GFP tagged was cleaved to produce untagged PGL-3. The cleavage was performed using a TEV protease overnight at 4 °C. PGL-3 and PGL-3-GFP proteins

were diluted with Dilution buffer (25 mM Tris pH 8.0, 1 mM DTT) to reach 50 mM KCl before loading the protein in an anion exchange HiTrapQ HP 5 mL column. The HiTrap column was previously equilibrated first with HiTrapQ elution buffer (25 mM Tris pH 8.0, 50 mM KCl, 1 mM DTT) and then with HiTrapQ binding buffer (25 mM Tris pH 8.0, 1 M KCl, 1 mM DTT). The column was mounted in a Äkta Pure FPLC system. After the sample was loaded the column was washed with HiTrapQ binding buffer. The sample was finally eluted with a linear gradient from 0 to 55% of HiTrapQ elution buffer (25 mM Tris pH 8.0, 1 M KCl 1 mM DTT) for 25 C.V. Finally a 100% HiTrap elution buffer step was performed for 5 C.V. The pooled fractions were then loaded in a HiLoad 16/60 Superdex 200 size exclusion chromatography column that was previously equilibrated with superdex buffer (25 mM HEPES 7.25, 300 mM KCl, 1 mM DTT). After size exclusion, the final samples were collected.

FUS protein purification Unlabeled FUS purified from a baculovirus construct containing N-HIS-MBP-FUS-TEV-SNAP. SF9-ESF cells were harvested after three days of infection by centrifugation at 500 x g for 10 min. The cell pellet was resuspended using 50 mL of lysis buffer (50 mM Tris pH 7.4, 500 mM KCl, 5% glycerol, 10 mM imidazole, 1mM PMSF, 1X protease inhibitor) for every 50 mL of cultured cells. The cells were lysed by passing them 2 times through the LM20 microfluidizer at 15 000 psi. The lysate was then centrifuged at 20 000 rpm for 45 min at 15 °C. The supernatant was collected and loaded into a Ni-NTA column that was previously equilibrated with lysis buffer. After loading the sample the column was washed for 10 C.V. with Ni-NTA was buffer (50 mM Tris pH 7.4, 500 mM KCl, 5% glycerol, 20 mM imidazole). The protein was then eluted with Ni-NTA elution buffer (50 mM Tris pH 7.4, 500 mM KCl, 5% glycerol, 300 mM imidazole). The collected fractions where then loaded into a MBPTrap HP column preequilibrated with Ni-NTA elution buffer. The MBP column was washed for 10 C.V with MBP wash buffer (50 mM Tris pH 7.4, 500 mM KCl, 5% glycerol). After washing, the sample was eluted with MBP elution buffer (50 mM Tris pH 7.4, 500 mM KCl, 5% glycerol, 500 mM arginine, 20 mM maltose). The protein was diluted to a concentration of less than 15 μ M using MBP elution buffer. 3C and TEV proteases were then added to cleave the MBP and SNAP tags from the FUS construct. The cleavage reactions were incubated overnight at 18 °C. Finally the protein was loaded in a SepFast GF-HS-L 26 x 600 mm gel filtration to remove the cleaved MBP and SNAP tags and exchange the buffer. The SepFast column was previously equilibrated in storage buffer (50 mM HEPES pH 7.25, 750 mM KCl, 5% glycerol, 1 mM DTT). The sample was concentrated to a final concentration of 15 μ L using 30 kDa Amicon centrifuge filters. FUS-GFP was purified as previously described in [44].

Predicting IDRs In our analysis we mainly use the MetaDisorder predictor by Kozlowski and Bujnicki [25] to identify the disordered and low complexity regions. MetaDisorder queries other predictors and generates a consensus answer, with an algorithm that tests the strength of each method against several datasets. It thus addresses the issue of training-set dependent model predictions. In particular, MetaDisorder includes also widely used IUpred predictors [16].

Field theoretic approach For the purpose of this study, i.e. analysing the impact of different domains of an IDP on its propensity to phase separate, we use field theoretic approaches rooted in statistical mechanics that are able to incorporate detailed structural information of polyampholytes such as proteins. They are obtained from the partition function for ensembles of coarse-grained polyampholytes, which are represented, via the Hubbard-Stratonovich transformation, as multi-dimensional functional integral (path integral) over all possible states or the polymeric system. Field theoretic simulations (FTS) are numerical methods that consider the full functional integral of the associated partition function. However, the integration of the full integral relies on stochastic sampling via Monte-Carlo methods

that have shown considerable convergence problems due to the oscillatory nature of the resulting distribution function [3]. While this approach has recently been used successfully for the tau protein [47], for the development of an appropriate and efficient theory that will allow to characterize a whole class of proteins such as IDPs and proteins that contain intrinsically disordered domains, and can be used to deliver analytical insight into the underlying biophysical principles leading to LLPS. For our analysis we use the Random-Phase Approximation (RPA). As is the case for self-consistent field theory, it can be derived from a saddle-point approximation, taking into account the asymptotically dominant contribution (a single Gaussian distribution) of the functional integral of the partition function, reducing the field theory to a mean-field model where a single Gaussian configuration interacts with an average effective field. It is one of the simplest analytical theories, that can account for small site-specific fluctuations, e.g. of charge patterns and structural features of the protein.

The RPA approach we use here has been set-up previously by Lin et al. [28, 29] for the IDR of Ddx4. For our analysis we use the extended model that includes salt as an additional variable, apart from the protein (here for IDRs of PGL-3 and FUS) (ϕ_{aa}), counterions (ϕ_c), KCl (ϕ_s) and water. The free energy

$$f = -S + f_{el}, \quad (4.1)$$

combines the entropic part S , coming from Flory-Huggins solution theory:

$$-S = \frac{\phi_{aa}}{N} \log \phi_{aa} + \phi_c \log \phi_c + 2\phi_s \log \phi_s + \phi_w \log \phi_w. \quad (4.2)$$

where N denotes the number of amino acids and the solvent fraction can be written as $\phi_w = 1 - \phi_{aa} - \phi_c - 2\phi_s$, with the electrostatic part of the free energy. It is this part for which the Random Phase Approximation has been developed, [7, 8, 11, 17]. It can be expressed in its simplest non-dimensional form as the following integral:

$$f_{el} = \int_0^\infty \frac{\tilde{k}^2}{4\pi^2} \left\{ \log \left[1 + \mathcal{G}(\tilde{k}) \right] - \mathcal{G}(\tilde{k}) \right\} dk, \quad (4.3)$$

with

$$\mathcal{G}(k) = \frac{4\pi}{k^2 (1 + k^2) T^*} \left(2\phi_s + \phi_c + \frac{\phi_{aa}}{N} \sigma_i (G_M)_{ij} \sigma_j \right), \quad (4.4)$$

where T^* is the nondimensionalised temperature which includes the relative permittivity of the medium. The correlation matrix for the chain can be written as

$$(G_M)_{ij} = e^{-k^2|i-j|/6}, \quad (4.5)$$

where the lengths have been nondimensionalized with the characteristic length a of the polymeric link of the associated Gaussian chain, here the domain of interest for protein PGL-3 or FUS. Note, that the neutrality condition implies $\phi_c = |\sigma| \phi_{aa}$, where $\sigma = (1/N) \sum_{i=1}^N \sigma_i$ is the average charge density of the protein, and σ_i is the charge of each of its amino acids. Together, this reduces the system to two quantities to be solved for, ϕ_{aa} and ϕ_s .

Computation of the phase diagrams To compute the phase diagrams including salt dependence, in particular when solving for the tie lines of the model system, we integrate the system using a Gauss-Laguerre quadrature, with the points computed using both our own implementation of the classic method by [20] and the state-of-the-art method by [19]. In order to find the tie lines between two

coexisting points $(\phi_{aa}, \phi_s)|_\alpha$ and $(\phi_{aa}, \phi_s)|_\beta$, we solve the following system, which is equivalent to the equality of the electrochemical potentials and the common tangent construction:

$$\partial_{\phi_{aa}} f(\phi_{aa}, \phi_s)|_\alpha = \partial_{\phi_{aa}} f(\phi_{aa}, \phi_s)|_\beta, \quad (4.6a)$$

$$\partial_{\phi_s} f(\phi_{aa}, \phi_s)|_\alpha = \partial_{\phi_s} f(\phi_{aa}, \phi_s)|_\beta, \quad (4.6b)$$

$$f(\phi_{aa}, \phi_s)|_\alpha - \partial_{\phi_{aa}} f(\phi_{aa}, \phi_s)|_\alpha \phi_{aa}|_\alpha - \partial_{\phi_s} f(\phi_{aa}, \phi_s)|_\alpha \phi_s|_\alpha = \quad (4.6c)$$

$$f(\phi_{aa}, \phi_s)|_\beta - \partial_{\phi_{aa}} f(\phi_{aa}, \phi_s)|_\beta \phi_{aa}|_\beta - \partial_{\phi_s} f(\phi_{aa}, \phi_s)|_\beta \phi_s|_\beta,$$

where the nonlinear system of equations is solved by using the *trust-region-dogleg* algorithm as implemented in Octave.

Note that this is a system of three equations with four unknowns, which means that we can take one of the unknowns as a parameter and solve for the others. In order to solve effectively the system above it is imperative to develop a continuation strategy. Naively, one could vary the unknown taken as a parameter and solve for the others using the prior solution as a guess. This strategy is known as *natural parameter continuation* [21], but the system above presents turning points for the parameter, which requires the development of a *pseudo-arclength continuation* algorithm. The latter implies the definition of a new variable, the pseudo arclength, which is defined as the arclength of the solution curve in the four-dimensional space spanned by the four unknowns in the system above. In practical terms, using this method implies adding an additional equation to the previous system that imposes a fixed increase of the arclength from the prior solution. Thus, we obtain a system of four solutions with four unknowns, that is similarly solved using the *trust-region-dogleg* algorithm.

Parameter estimation and fitting procedure We now express T^* and ϕ_{aa} in terms of experimentally accessible variables, concentrations (c) and temperatures (T). In the case of ϕ_{aa} , we know that it corresponds to the total volume occupied by amino acids over the volume of the total lattice. We take the volume of a lattice site to be that of a single water molecule, which corresponds to $1/55.5M$. We would have then for PGL-3 in the presence of KCl (note that PGL-3 consists of 693 amino acids):

$$\phi_{aa} = \frac{693}{55.5M} [\text{PGL-3}] \quad \text{and} \quad \phi_s = \frac{1}{55.5M} [\text{KCl}]. \quad (4.7)$$

There are two implicit assumptions in this calculation. One, it is assumed that each amino acid takes the same volume as a water molecule, which is clearly not true, and two, that the volume of a protein scales linearly with the number of amino acids. The latter can be argued to be false, since different configurations of the protein will give different volumes. Therefore, there is not a clear and easy way of relating the volume fraction and the protein concentration.

The non-dimensional temperature T^* is related to T by

$$T^* = \frac{4\pi\epsilon_0\epsilon_r k_B a}{e^2} T, \quad (4.8)$$

where ϵ_0 is the vacuum permittivity, ϵ_r is the relative permittivity of the medium, k_B is Boltzmann's constant, a the link length and e is the charge of the electron. The two unknowns in Eq. (4.8) are a and ϵ_r . We take a to be the $C\alpha$ - $C\alpha$ virtual bond length of 3.8\AA and fit ϵ_r , due to the lack of a complete theory to derive its value, we only know that it should be between 2 (typical value for hydrocarbon crystals) and 80 ($\epsilon_{r, H_2O} = 78$).

The impossibility of obtaining accurate theoretical estimates for the relation between volume fraction and the concentration of the protein and for the permittivity of the solution make it necessary to proceed

Table 1: Fitting parameters for the IDRs of FUS and PGL-3 at KCl and temperatures used in experiments. (*) The parameter values for FUS-LC correspond to the constant salinity case, since there is no binodal in the partitioning case for FUS-LC. For any KCl used, the relative permittivity is smaller than 1.

IDR	[KCl] (mM)	Temperature scaling - β_1 (K)	Concentration scaling - β_2 (mM)	χ^2	Relative Permittivity
FUS-C1	100	5349.71	10.1	191.32	8.22
	150	5772.23	10.1	191.32	7.62
	200	6011.68	10.1	191.32	7.31
FUS-N1	100	4476.27	207	3.35	9.82
	150	5019.36	207	3.35	8.76
	200	5355.95	207	3.35	8.21
FUS-LC (*)	100	79019.9	25.3	3.92	0.56
	150	109856	25.3	3.92	0.40
	200	138717	25.3	3.92	0.32
PGL-3-C1	100	1152.38	16.2	23.39	38.16
	150	1307.78	16.2	23.39	33.62
	200	1347.48	16.2	23.39	32.63

with a fitting procedure. There exist two parameters, β_1 and β_2 that allow us to scale the volume fraction and the non-dimensional temperature to fit the data of a T - c phase diagram, i.e. $T = \beta_1 T^*$ and $c = \beta_2 \phi_{aa}$. Once the phase diagram is computed, we obtain a functional relation between T^* and ϕ_{aa} , $\phi_{aa} = f(T^*)$ (note that f has different branches, a dilute and a condensed branch). If we know both parameters we can rescale the previous relation to obtain $c = \beta_2 f(T/\beta_1)$, a predictor for the concentration. We can find both parameters by minimizing the following function:

$$\chi^2 = \sum_{i=1}^N \left(\frac{c_{\text{exp},i} - c_{\text{th},i}}{\sigma} \right)^2, \quad (4.9)$$

where c_{exp} are the experimental values of the concentration, σ is the standard deviation of the experimental points. The sum runs over all experimental points including all branches and all values of the salt concentration. The parameter β_1 is allowed to vary with the salt concentration on account of the strong dependence of the permittivity on salt concentration.

Eq. (4.9) is then minimized using a sequential quadratic programming algorithm as implemented in Octave. The parameter χ^2 will provide then a measure of the goodness of fit for each case. In Table 1 we give a summary of all fitting parameters for all cases we have considered.

5 Acknowledgments

We acknowledge Anthony Hyman for providing all the support needed for the experimental part of this project. EM would like thank the Weierstrass Institute for hosting a research visit. EM and BW would like to thank Andreas Münch for helpful discussions on the theoretical aspects of the research project. For discussions and help with the experimental part of the project we thank, Martine Ruer, Patrick McCall, Tylor Harmon, Jie Wang, and Titius Franzmann. We thank the light microscopy, chromatography,

and protein purification facilities at MPI-CBG for their support. We thank Olympus for providing the CSU-W1 SoRa spinning-disc system based on an IXplore IX83 microscope.

References

- [1] Alberti, S., Halfmann, R., King, O., Kapila, A., and Lindquist, S. (2009). A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell*, 137(1):146–158.
- [2] Alberti, S. and Hyman, A. A. (2021). Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nat Rev Mol Cell Biol*, 22:196–213.
- [3] Baeurle, S. A. (2002). Method of Gaussian Equivalent Representation: A New Technique for Reducing the Sign Problem of Functional Integral Methods. *Physical Review Letters*, 89(8):080602.
- [4] Banani, S. F., Lee, H. O., Hyman, A. A., and Rosen, M. K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nature reviews Molecular cell biology*, 18(5):285–298.
- [5] Benayad, Z., von Bülow, S., Stelzl, L. S., and Hummer, G. (2021). Simulation of fus protein condensates with an adapted coarse-grained model. *Journal of Chemical Theory and Computation*, 17(1):525–537.
- [6] Berry, J., Brangwynne, C., and Haataja, M. P. (2018). Physical principles of intracellular organization via active and passive phase transitions. *Reports on Progress in Physics*.
- [7] Borue, V. Y. and Erukhimovich, I. Y. (1988). A statistical theory of weakly charged polyelectrolytes: fluctuations, equation of state and microphase separation. *Macromolecules*, 21(11):3240–3249.
- [8] Borukhov, I., Andelman, D., and Orland, H. (1998). Random polyelectrolytes and polyampholytes in solution. *The European Physical Journal B - Condensed Matter and Complex Systems*, 5(4):869–880.
- [9] Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., and Hyman, A. A. (2009). Germline p granules are liquid droplets that localize by controlled dissolution/condensation. *Science*, 324(5935):1729–1732.
- [10] Brangwynne, C. P., Tompa, P., and Pappu, R. V. (2015). Polymer physics of intracellular phase transitions. *Nature Physics*, 11(11):899–904.
- [11] Castelnovo, M. and Joanny, J.-F. (2001). Complexation between oppositely charged polyelectrolytes: Beyond the random phase approximation. *The European Physical Journal E*, 6(1):377–386.
- [12] Cushman, M., Johnson, B. S., King, O. D., Gitler, A. D., and Shorter, J. (2010). Prion-like disorders: blurring the divide between transmissibility and infectivity. *Journal of Cell Science*, 123(8):1191–1201.
- [13] Das, S., Eisen, A., Lin, Y.-H., and Chan, H. S. (2018). A lattice model of charge-pattern-dependent polyampholyte phase separation. *The Journal of Physical Chemistry B*, 122(21):5418–5431.
- [14] Dignon, G., Zheng, W., Kim, Y., Best, R., and Mittal, J. (2018). Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Comput Biol*, 14(1):e1005941.

- [15] Dinic, J., Marciel, A. B., and Tirrell, M. V. (2021). Polyampholyte physics: Liquid-liquid phase separation and biological condensates. *Current Opinion in Colloid & Interface Science*, page 101457.
- [16] Dosztányi, Z. (2018). Prediction of protein disorder based on iupred. *Protein Science*, 27(1):331–340.
- [17] Ermoshkin, A. V. and Olvera de la Cruz, M. (2004). Gelation in strongly charged polyelectrolytes. *J. Polym. Sci. Part B*, 42:766–776.
- [18] Flory, P. J. (1942). Thermodynamics of high polymer solutions. *The Journal of chemical physics*, 10(1):51–61.
- [19] Glaser, A., Liu, X., and Rokhlin, V. (2007). A fast algorithm for the calculation of the roots of special functions. *SIAM Journal on Scientific Computing*, 29(4):1420–1438.
- [20] Golub, G. H. and Welsch, J. H. (1969). Calculation of gauss quadrature rules. *Mathematics of computation*, 23(106):221–230.
- [21] Govaerts, W. J. (2000). *Numerical methods for bifurcations of dynamical equilibria*, volume 66. Siam.
- [22] Huggins, M. L. (1942). Some properties of solutions of long-chain compounds. *The Journal of Physical Chemistry*, 46(1):151–158.
- [23] Kang, J., Lim, L., Lu, Y., and Song, J. (2019). A unified mechanism for lpls of als/ftld-causing fus as well as its modulation by atp and oligonucleic acids. *PLoS biology*, 17(6):e3000327.
- [24] Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., Mirzaei, H., Goldsmith, E. J., Longgood, J., Pei, J., et al. (2012). Cell-free formation of rna granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell*, 149(4):753–767.
- [25] Kozłowski, L. P. and Bujnicki, J. M. (2012). MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics*, 13(1):111.
- [26] Li, L., Srivastava, S., Andreev, M., Marciel, A. B., de Pablo, J. J., and Tirrell, M. V. (2018). Phase behavior and salt partitioning in polyelectrolyte complex coacervates. *Macromolecules*, 51(8):2988–2995.
- [27] Lin, Y.-H., Brady, J. P., Chan, H. S., and Ghosh, K. (2020). A unified analytical theory of heteropolymers for sequence-specific phase behaviors of polyelectrolytes and polyampholytes. *The Journal of chemical physics*, 152(4):045102.
- [28] Lin, Y.-H., Forman-Kay, J. D., and Chan, H. S. (2016). Sequence-specific polyampholyte phase separation in membraneless organelles. *Physical review letters*, 117(17):178101.
- [29] Lin, Y.-H., Song, J., Forman-Kay, J. D., and Chan, H. S. (2017). Random-phase-approximation theory for sequence-dependent, biologically functional liquid-liquid phase separation of intrinsically disordered proteins. *Journal of Molecular Liquids*, 228:176–193.
- [30] Luo, F., Gui, X., Zhou, H., Gu, J., Li, Y., Liu, X., Zhao, M., Li, D., Li, X., and Liu, C. (2018). Atomic structures of fus lc domain segments reveal bases for reversible amyloid fibril formation. *Nature structural & molecular biology*, 25:341–346.

- [31] Lytle, T. K., Chang, L.-W., Markiewicz, N., Perry, S. L., and Sing, C. E. (2019). Designing electrostatic interactions via polyelectrolyte monomer sequence. *ACS Central Science*, 5(4):709–718.
- [32] Lytle, T. K. and Sing, C. E. (2017). Transfer matrix theory of polymer complex coacervation. *Soft Matter*, 13:7001–7012.
- [33] Madinya, J. J., Chang, L.-W., Perry, S. L., and Sing, C. E. (2019). Sequence-dependent self-coacervation in high charge-density polyampholytes. *Mol. Syst. Des. Eng.*, pages –.
- [34] Overbeek, J. T. G. and Voorn, M. (1957). Phase separation in polyelectrolyte solutions. theory of complex coacervation. *Journal of Cellular and Comparative Physiology*, 49(S1):7–26.
- [35] Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T. M., et al. (2015). A liquid-to-solid phase transition of the als protein fus accelerated by disease mutation. *Cell*, 162(5):1066–1077.
- [36] Perry, S. L. and Sing, C. E. (2015). Prism-based theory of complex coacervation: Excluded volume versus chain correlation. *Macromolecules*, 48(14):5040–5053.
- [37] Radhakrishna, M., Basu, K., Liu, Y., Shamsi, R., Perry, S. L., and Sing, C. E. (2017). Molecular connectivity and correlation effects on polymer coacervation. *Macromolecules*, 50(7):3030–3037.
- [38] Rauscher, S. and Pomès, R. (2017). The liquid structure of elastin. *Elife*, 6:e26526.
- [39] Rumyantsev, A. M., Jackson, N. E., Yu, B., Ting, J. M., Chen, W., Tirrell, M. V., and de Pablo, J. J. (2019). Controlling complex coacervation via random polyelectrolyte sequences. *ACS Macro Letters*, 8(10):1296–1302.
- [40] Saha, S., Weber, C. A., Nusch, M., Adame-Arana, O., Hoege, C., Hein, M. Y., Osborne-Nishimura, E., Mahamid, J., Jahnel, M., Jawerth, L., et al. (2016). Polar positioning of phase-separated liquid compartments in cells regulated by an mrna competition mechanism. *Cell*, 166(6):1572–1584.
- [41] Shen, K. and Wang, Z.-G. (2017). Electrostatic correlations and the polyelectrolyte self energy. *The Journal of Chemical Physics*, 146:084901.
- [42] Shen, K. and Wang, Z.-G. (2018). Polyelectrolyte chain structure and solution phase behavior. *Macromolecules*, 51(5):1706–1717.
- [43] Shin, Y. and Brangwynne, C. P. (2017). Liquid phase condensation in cell physiology and disease. *Science*, 357(6357):eaaf4382.
- [44] Wang, J., Choi, J.-M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovskiy, A., Drechsel, D., Poser, I., Pappu, R. V., Alberti, S., and Hyman, A. A. (2018). A molecular grammar governing the driving forces for phase separation of prion-like rna binding proteins. *Cell*, 174(3):688 – 699.e16.
- [45] Wootton, J. C. and Federhen, S. (1996). [33] analysis of compositionally biased regions in sequence databases. In *Methods in enzymology*, volume 266, pages 554–571. Elsevier.
- [46] Zhang, P., Shen, K., Alsaifi, N. M., and Wang, Z.-G. (2018). Salt partitioning in complex coacervation of symmetric polyelectrolytes. *Macromolecules*, 51(15):5586–5593.

- [47] Zhang, X., Vigers, M., McCarty, J., Rauch, J. N., Fredrickson, G. H., Wilson, M. Z., Shea, J.-E., Han, S., and Kosik, K. S. (2020). The proline-rich domain promotes Tau liquid–liquid phase separation in cells. *Journal of Cell Biology*, 219(11).

Appendix

Supplemental figure for PGL-3

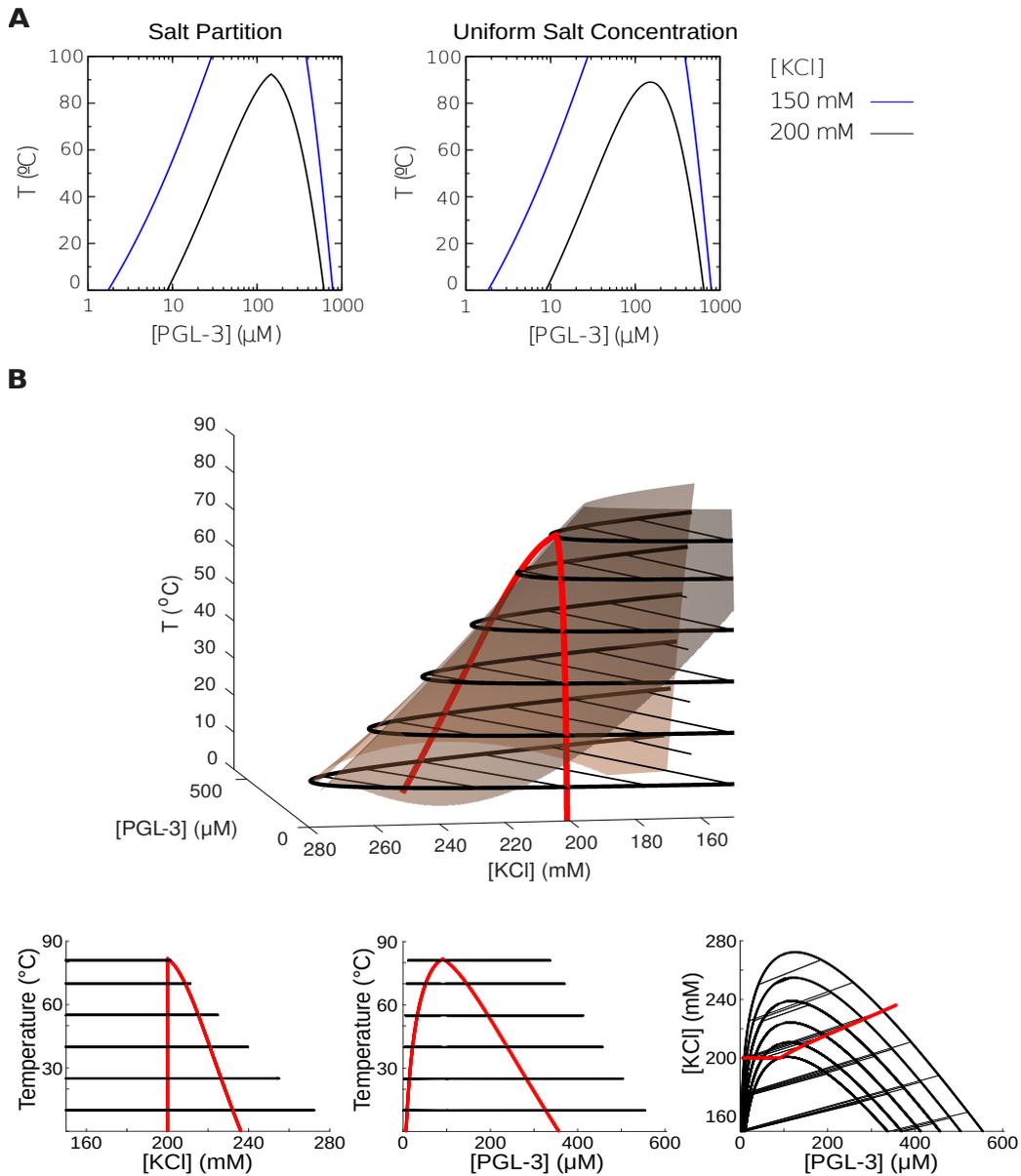


Figure 6:

A Comparison of the temperature-protein concentration phase diagrams for the salt partition (left panel) and constant salt (right panel) cases for PGL-3.
B 3D phase diagram (top panel) and projections on the three planes for PGL-3.