




Comparative roles of charge, π , and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins

Suman Das^a, Yi-Hsuan Lin^{a,b}, Robert M. Vernon^b, Julie D. Forman-Kay^{b,a} , and Hue Sun Chan^{a,1}

^aDepartment of Biochemistry, University of Toronto, Toronto, ON M5S 1A8, Canada; and ^bMolecular Medicine, Hospital for Sick Children, Toronto, ON M5G 0A4, Canada

Edited by José N. Onuchic, Rice University, Houston, TX, and approved October 2, 2020 (received for review April 26, 2020)

Endeavoring toward a transferable, predictive coarse-grained explicit-chain model for biomolecular condensates underlain by liquid–liquid phase separation (LLPS) of proteins, we conducted multiple-chain simulations of the N-terminal intrinsically disordered region (IDR) of DEAD-box helicase Ddx4, as a test case, to assess roles of electrostatic, hydrophobic, cation– π , and aromatic interactions in amino acid sequence-dependent LLPS. We evaluated three different residue–residue interaction schemes with a shared electrostatic potential. Neither a common hydrophobicity scheme nor one augmented with arginine/lysine–aromatic cation– π interactions consistently accounted for available experimental LLPS data on the wild-type, a charge-scrambled, a phenylalanine-to-alanine (FtoA), and an arginine-to-lysine (RtoK) mutant of Ddx4 IDR. In contrast, interactions based on contact statistics among folded globular protein structures reproduce the overall experimental trend, including that the RtoK mutant has a much diminished LLPS propensity. Consistency between simulation and experiment was also found for RtoK mutants of P-granule protein LAF-1, underscoring that, to a degree, important LLPS-driving π -related interactions are embodied in classical statistical potentials. Further elucidation is necessary, however, especially of phenylalanine's role in condensate assembly because experiments on FtoA and tyrosine-to-phenylalanine mutants suggest that LLPS-driving phenylalanine interactions are significantly weaker than posited by common statistical potentials. Protein–protein electrostatic interactions are modulated by relative permittivity, which in general depends on aqueous protein concentration. Analytical theory suggests that this dependence entails enhanced interprotein interactions in the condensed phase but more favorable protein–solvent interactions in the dilute phase. The opposing trends lead to only a modest overall impact on LLPS.

biomolecular condensates | membraneless organelles | phase separation

A preponderance of recent advances demonstrate that liquid–liquid phase separation (LLPS) of intrinsically disordered proteins (IDPs), proteins containing intrinsically disordered regions (IDRs), folded proteins, and nucleic acids is a general biophysical mechanism to achieve functional spatial and temporal organization of biomolecules in both intracellular and extracellular organismal space (1–9). LLPS underpins formation of a variety of biomolecular condensates (10), including intracellular bodies, such as nucleoli and stress granules, that are often referred to as membraneless organelles (4, 11), and precursor of extracellular materials as in the case of sandcastle worm adhesive (12) and elastin in vertebrate tissues (8). These dynamic, phase-separated condensates perform versatile functions, as underscored by their recently elucidated roles in synapse formation and plasticity (7, 13), organization of chromatin (14), regulation of translation (15, 16), B cell response (17), and autophagosome formation (18). The pace of discovery in this very active area of research has been accelerating (19–28).

While experimental progress has been tremendous, theory for the physicochemical basis of biomolecular condensates is still in its infancy. Biomolecular condensates *in vivo* are complex, involving many species of proteins and nucleic acids maintained often by nonequilibrium processes (10, 19, 29–31), rendering atomistic modeling impractical. Facing this challenge, promising initial theoretical steps using coarse-grained approaches were made to tackle simpler *in vitro* LLPS systems as their elucidation is a prerequisite for physical insights into more complex *in vivo* condensates. These recent efforts encompass analytical theories at various levels of approximation (32–42), field theory simulations (43–46), and lattice (47–50) or continuum (51–54) coarse-grained explicit-chain simulations that account for either individual amino acid residues (47, 52, 53, 55) or, at lower structural resolution, groups of residues (56, 57)—including using patchy particle representations (58, 59). The different theoretical/computational approaches are complementary and were applied to address how amino acid composition [number/fraction of hydrophobic (54), aromatic (39, 60), or charged (4) residues] and the sequence pattern of charge (34, 48, 53, 61), hydrophobic (50, 51, 54), or aromatic (60) residues affect LLPS propensity of heteropolymers as well as pertinent impact of temperature

Significance

Mesoscopic condensates of proteins and nucleic acids, including membraneless organelles, serve myriad biological functions, whereas dysregulation of condensates can cause disease. How condensates are governed by genetically encoded biomolecular sequences is of central importance. Critical assessment of how various interactions reproduce experimental phase behaviors indicates that π -interactions' notable facilitation of condensate formation is well reflected in established contact statistics among folded protein structures, although the condensate-driving capability of the large hydrophobic/aromatic residue phenylalanine is weaker than expected. By elucidating physical aspects of the residue–residue contacts in condensates, highlighting these contacts' largely solvent-exposed character with ramifications such as reduced hydrophobic strengths relative to buried nonpolar contacts, fundamental conceptual and quantitative progress is made toward predictive models for biomolecular condensates.

Author contributions: S.D., Y.-H.L., R.M.V., J.D.F.-K., and H.S.C. designed research; S.D., Y.-H.L., R.M.V., and H.S.C. performed research; S.D., Y.-H.L., R.M.V., and H.S.C. analyzed data; and S.D., Y.-H.L., and H.S.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

¹To whom correspondence may be addressed. Email: chan@arrhenius.med.utoronto.ca.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2008122117/-DCSupplemental>.

First published November 2, 2020.

(20, 43, 51, 54), hydrostatic pressure (62–64), salt (41, 46), and osmolyte (27, 63), offering physical insights into the LLPS behaviors of, for example, the DEAD-box RNA helicase Ddx4 (34, 65), RNA-binding protein fused in sarcoma (FUS) (52), prion-like domains (60), and postsynaptic densities (64).

Developing LLPS models with transferable interaction potentials applicable to a wide range of amino acid sequences is essential for advancing fundamental physical understanding of natural biomolecular condensates and engineering of bioinspired materials (66). In this endeavor, the rapidly expanding repertoire of experimental data offers critical assessment of theoretical and computational approaches. Building on aforementioned progress (34, 41, 52, 53), the present study evaluates a variety of interaction schemes for coarse-grained residue-based chain simulations of LLPS of IDPs or IDRs, including but not limited to schemes proposed in the literature (52). We do so by first comparing their sequence-specific predictions against experiment on the RNA helicase Ddx4 for which extensive LLPS data on the wild-type (WT) and three mutant sequences are available to probe the contribution of hydrophobic, electrostatic (4), cation- π , and possibly other π -related (4, 65, 67) interactions. We use these data to benchmark the relative strengths of different types of interaction in our model. Of particular interest are the aromatic (68) and other π -related (67) interactions, which have significant impact on folded protein structure, conformational distribution of IDPs, and LLPS properties (4, 34, 39, 60, 69–73) but are often not adequately accounted for in model potentials (67). Interestingly, a simple statistical potential based upon folded protein structures (74, 75) accounts for the general trend of LLPS properties of the four Ddx4 IDR sequences, including that LLPS is more favored by arginine than lysine despite their essential identical electric charges, but that a model potential that relies solely on hydrophobicity (76) does not. This finding indicates that at the coarse-grained level of residue-residue interactions, IDP/IDR LLPS is governed largely by similar forces—including the π -related ones—that drive protein folding. Analogous agreement between statistical potential model prediction and experiment with respect to the arginine/lysine contrast is also found for the N-terminal RGG domain of P-granule RNA Ddx3 helicase LAF-1 (77). However, experimental data on tyrosine-to-phenylalanine mutants of LAF-1 and FUS (39) indicate that the contribution of the large aromatic residue phenylalanine to LLPS is overestimated by statistical potentials, most likely because the interactions involving phenylalanine in the sequestered hydrophobic core of globular proteins are not sufficiently representative of more solvent-accessible LLPS-driving interactions, pointing to a crucial aspect of LLPS energetics toward which future investigations should be directed. To gain further insights into the electrostatic driving forces for LLPS, we have conducted explicit-water simulation and developed analytical theory which suggests, at variance with previous analyses (35, 37), that the physically expected dependence of effective permittivity on IDR concentration may have a modest instead of drastic impact on LLPS propensity because of a tradeoff between solvent-mediated electrostatic interchain interactions and self-interactions. These findings and their ramifications are discussed below.

Results and Discussion

As described in *Materials and Methods* and *SI Appendix, SI Text*, our coarse-grained protein chain model for IDP LLPS basically follows the approaches in refs. 52, 53, which in turn are based on a recently proposed efficient simulation protocol (78). In the first step of our analysis, we critically assess the models against the experimental data on the Ddx4 IDR (*SI Appendix, Fig. S1*), which indicate that all three Ddx4 IDR mutants—the charge scrambled (CS), phenylalanine-to-alanine (FtoA), and arginine-to-lysine (RtoK) variants—have significantly reduced LLPS propensities

relative to the WT (4, 65, 67). The CS, FtoA, and RtoK variants are useful probes for LLPS energetics. They were constructed specifically to study the experimental effects of sequence charge pattern (the arrangement of charges along CS sequence is less blocky than that in WT while the amino acid composition is unchanged), the relative importance of aromatic/ π -related vs. hydrophobic/nonpolar interactions (all 14 Phe residues in WT Ddx4 IDR are mutated to Ala in FtoA), and the role of Arg vs Lys (all 24 Arg residues in WT IDR are mutated to Lys in RtoK) on the LLPS of Ddx4.

Assessing Biophysical Perspectives Embodied by Different Coarse-Grained Interaction Schemes for Modeling Biomolecular Condensates

We consider the potential functions in the hydrophobicity-scale (HPS) and the Kim-Hummer (KH) models in Dignon et al. (52) as well as the HPS potential with augmented cation- π terms (72), all of which share the same bond energy term, U_{bond} , for chain connectivity and screened electrostatic term, U_{el} , for pairs of charged residues, as described in *SI Appendix, SI Text*. We focus first on the pairwise contact interactions between amino acid residues, which correspond to the U_{aa} energies of either the HPS or KH model (excluding U_{bond} and U_{el}).

The HPS model assumes that the dominant driving force for IDP LLPS is hydrophobicity as characterized by a scale for the 20 amino acid residues. Pairwise contact energy is taken to be the sum of the hydrophobicities of the two individual residues of the pair. The HPS model adopts the scale of Kapcha and Rosky, in which the hydrophobicity of a residue is a composite quantity based on a binary hydrophobicity scale of the atoms in the residue (76).

In contrast, the KH model (79) relies on knowledge-based potentials derived from contact statistics of folded protein structures in the Protein Data Bank (PDB). As such, it assumes that the driving forces for IDP LLPS are essentially identical to those for protein folding at a coarse-grained residue-by-residue level, as obtained by Miyazawa and Jernigan (75), without singling out a priori a particular interaction type as being dominant.

The HPS model has been applied successfully to study the FUS low-complexity domain (80), the RNA-binding protein TDP-43 (81), and the LAF-1 RGG domain as well as its charge shuffled variants (77). A temperature-dependent version of HPS (HSP-T) (51) was also able to rationalize the experimental LLPS properties of artificial designed sequences (82). When both the HPS and KH models were applied to FUS and LAF-1, the predicted phase diagrams were qualitatively similar for a given sequence, although they exhibited significantly different critical temperatures (52), which should be attributable to the difference in effective energy/temperature scale of the two models. Here we conduct a systematic assessment of the two models' underlying biophysical assumptions by evaluating their ability to provide a consistent rationalization of the LLPS properties of a set of IDR sequences.

The scatterplot in Fig. 1A of HPS and KH energies indicates that despite an overall correlation, there are significant outliers. The most conspicuous outliers are interactions involving Arg (red), which are much less favorable in HPS than in KH. By comparison, most of the interactions involving Pro, as depicted by the 16 outlying blue circles as well as the single yellow and single green circles to the left of the main trend, are considerably more favorable in HPS than in KH. Interactions involving Phe (yellow) and Lys (green) essentially follow the main trend. Those involving Phe are favorable to various degrees in both models. However, some interactions involving Lys are attractive in HPS but repulsive in KH. For example, Lys-Lys interaction is attractive at ≈ -0.1 kcal mol⁻¹ for HPS but is repulsive at $\approx +0.2$ kcal mol⁻¹ for KH. Fig. 1B underscores the difference in interaction pattern of the two models for the WT Ddx4 IDR.

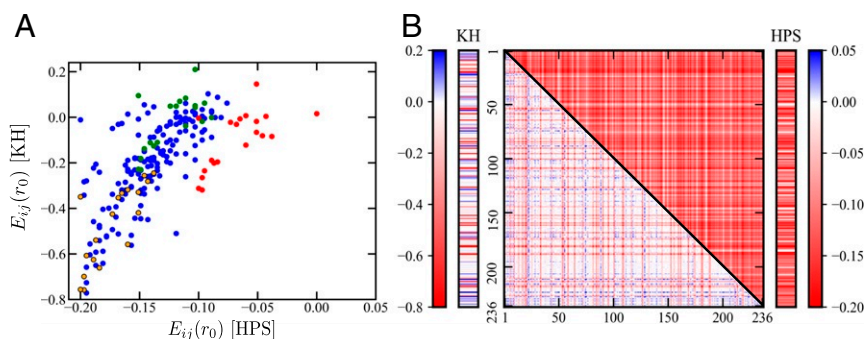


Fig. 1. Comparing two amino acid residue-based coarse-grained potentials. (A) Scatterplot of 210 pairwise contact energies (in units of kcal mol^{-1}) in the HPS (horizontal variable) versus those in the KH (vertical variable) model (52). $E_{ij}(r_0)$ are the pairwise potential energies $U_{aa|HPS}(r)$ or $U_{aa|KH}(r)$ (SI Appendix, SI Text), between two residues of types i, j separated by $r_{ij} = r_0$ where the Lennard-Jones component of the potential is minimum (i, j here stand for labels for the 20 amino acid types). Energies of contacts involving Arg (red circles), Lys (green circles), and Phe (yellow-filled black circles) are colored differently from others (blue circles). (B) Contact energies between residue pairs at positions i, j of the $n = 236$ sequence of WT Ddx4 IDR [Ddx4^{N1} (4)] in the two potentials are color coded by the scales. The vertical and horizontal axes represent residue positions $i, j \leq n$. The $i \neq j$ contact energies in the HPS and KH models are provided in the 2D plot, whereas the $i = j$ contact energies are shown alongside the model potentials' respective color scales.

The KH pattern is clearly more heterogeneous with both attractions and repulsions, whereas the HPS pattern is more uniform with no repulsive interactions. These differences should lead to significantly different predictions in sequence-dependent LLPS properties, as will be explored below.

Because of the importance of cation- π interactions in protein folded structure (69) as well as conformational distribution of IDP and LLPS (4, 39, 72, 73, 77), we study another set of model interaction schemes—in addition to HPS and KH, referred to as HPS + cation- π —that augment the HPS potentials with terms specific for cation- π interactions between Arg or Lys and the aromatic Tyr, Phe, or Trp (Fig. 2). As explained in SI Appendix, SI Text, we consider two alternate scenarios: 1) the cation- π interaction strength is essentially uniform, irrespective of the

cation-aromatic pair (Fig. 2A), as suggested by an earlier analysis (69), and 2) the cation- π interaction strength is significantly stronger for Arg than for Lys (Fig. 2B). The latter scheme is motivated by recent experiments showing that Arg to Lys substitutions reduce LLPS propensity, as in the cases of the RtoK mutant of Ddx4 IDR (67) and variants of FUS (39) and LAF-1 (77), as well as a recent theoretical investigation pointing to different roles of Arg and Lys in cation- π interactions (83). Contact statistics of PDB structures, including those of Miyazawa and Jernigan (74, 75) on which the KH potential is based, may also suggest that Arg- π attractions are stronger than Lys- π interactions. Indeed, among a set of 6,943 high-resolution X-ray protein structures (67), we find that an Arg-aromatic pair is at least 75% more likely than a Lys-aromatic pair to be within a C_α - C_α

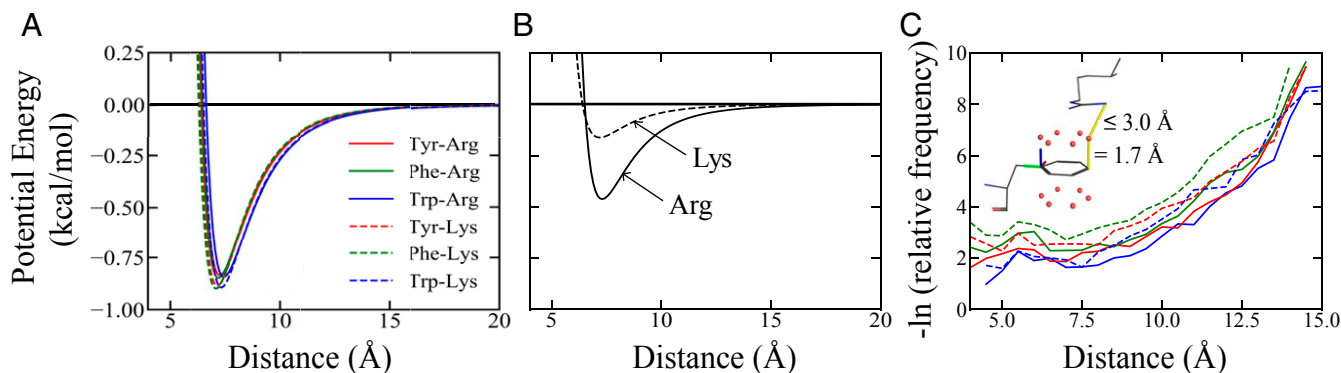


Fig. 2. Possible cation- π interaction potentials. (A) Sum of the coarse-grained HPS potential and a model cation- π interaction with a uniform $(\epsilon_{c\pi})_{ij} = 3.0 \text{ kcal mol}^{-1}$ as a function of residue-residue distance for the residue pairs Arg-Tyr, Arg-Phe, Arg-Trp, Lys-Tyr, Lys-Phe, and Lys-Trp, wherein Tyr, Phe, and Trp are labeled as red, green, and blue, and Arg and Lys are represented by solid and dashed curves. (B) An alternate cation- π potential in which Arg-Tyr/Phe/Trp is significantly more favorable (solid curve; $(\epsilon_{c\pi})_{ij} = 1.85 \text{ kcal mol}^{-1}$) than Lys-Tyr/Phe/Trp (dashed curve; $(\epsilon_{c\pi})_{ij} = 0.65 \text{ kcal mol}^{-1}$). Note that the plotted curves here—unlike those in A—do not contain the HPS potential. (C) Normalized C_α - C_α distance-dependent contact frequencies for the aforementioned six cation- π pairs (color coded as in A) computed using a set of 6,943 high-resolution X-ray protein crystal structures (resolution $\leq 1.8 \text{ \AA}$) from a published nonredundant set (67). Contact pair statistics are collected from residues on different chains in a given structure and residues separated by ≥ 50 amino acids along the same chain. C_α - C_α distances are divided into 0.2- \AA bins. For each bin, the relative frequency is the number of instances of a cation- π -like contact (defined below) divided by the total number of residue pairs with C_α - C_α distances within the narrow range of the bin. Thus, the shown curves quantify the tendency for a given pair of residues to engage in cation- π interaction provided that the pair is spatially separated by a given C_α - C_α distance. Here a cation- π -like contact is recognized if either a Lys NZ or an Arg NH1 nitrogen atom is within 3.0 \AA of any one of the points 1.7 \AA above or below an sp^2 carbon atom along the normal of the aromatic ring in a Tyr, Phe, or Trp residue. This criterion is exemplified by the molecular drawing (inset) of a contact between an Arg (at the top) and a Phe (at the bottom). Colors of the chemical bonds indicate types of atom involved, with carbon in black, oxygen in red, and nitrogen in blue. The red dots here are points on the exterior surfaces of the electronic orbitals farthest from the sp^2 carbons in the aromatic ring. The blue, green, and red lines emanating from a corner of the aromatic ring constitute a local coordinate frame, with the blue line being the normal vector of the plane of the aromatic ring determined from the positions of its first three atoms. The yellow lines mark spatial separations used to define the cation- π -like contact. These yellow lines do not represent chemical bonds.

distance of $\leq 6.5 \text{ \AA}$ (*SI Appendix, Fig. S2*), a separation that is often taken as a criterion for residue–residue contact (74). On top of that, given an Arg–aromatic and a Lys–aromatic pair separated by the same C_{α} – C_{α} distance (Fig. 2C), the Arg–aromatic pair (solid curves) is more likely than the Lys–aromatic pair (dashed curves) to adopt configurations consistent with a cation– π interaction. We emphasize, however, that although a significantly stronger Arg- than Lys-associated cation– π interaction is explored here as an alternate scenario, it is probable, as argued by Gallivan and Dougherty using a comparison between Lys-like ammonium–benzene and Arg-like guanidinium–benzene interactions, that the strengths of the pure cation– π parts of Arg- and Lys–aromatic interactions are similar despite the relative abundance of Arg–aromatic contacts due to other factors (69) such as π – π effects (67).

Hydrophobicity, Electrostatics, and Cation– π Interactions Are Insufficient by Themselves to Rationalize Ddx4 LLPS Data in Their Entirety.

We begin our assessment of models by applying the HPS and HPS + cation– π potentials to simulate the phase diagrams of the four Ddx4 IDRs (*SI Appendix, Fig. S1*), the sequence patterns of which are illustrated in Fig. 3A using a style employed previously, e.g., in refs. 4, 34, 39. The phase diagrams presented here are coexistence curves. When the overall average IDR density lies in between the left and right arms of the coexistence curve at a given temperature, the system phase separates into a dilute phase and a condensed phase with IDR densities given by the low- and high-density values, respectively, of the coexistence curve at the given temperature. When the average IDR density is not in the region underneath the coexistence curve, the system is not phase separated (see, e.g., refs. 20, 35, 84 for introductory discussions). Consistent with experiments (4, 65), the simulated phase diagrams (Fig. 3B) exhibit upper critical solution temperatures, which is a maximum temperature above which phase separation does not occur (corresponding to the local maxima of the coexistence curves at IDR density $\approx 200 \text{ mg/mL}$ in Fig. 3). We emphasize, however, that although the simulated critical temperatures

are assuringly in the same range as those deduced experimentally (65), the model temperature (in K) of our simulated phase diagrams in Figs. 3B and 4 should not be compared directly with experimental temperature. This is not only because of uncertainties about the overall model energy scale but also because the models do not account for the temperature dependence of effective residue–residue interactions (20, 43, 51). For simplicity, our models include only temperature-independent energies as they are purposed mainly for comparing the LLPS propensities of different amino acid sequences on the same footing rather than for highly accurate prediction of LLPS behaviors of any individual sequence.

Fig. 3B, *Left*, provides the HPS phase diagrams at relative permittivity $\epsilon_r = 80$ (approximately equal to that of bulk water, as in ref. 52). The predicted behaviors of the CS and FtoA variants are consistent with experiments in that their LLPS propensities are reduced relative to WT (4, 65), but the predicted enhanced LLPS propensity of RtoK is opposite to the experimental finding of Vernon et al. that the LLPS propensity of RtoK is lower than that of WT (67). This shortcoming of the HPS model is a consequence of its assignment of much less favorable interactions for Arg than for Lys, as noted in Fig. 1A.

Fig. 3B, *Right*, provides the HPS + cation– π phase diagrams. They are computed for $\epsilon_r = 80, 40,$ and 20 to gauge the effect of electrostatic interactions relative to other types of interactions. The ϵ_r -dependent results are also preparatory for our subsequent investigation of the effect of IDR concentration-dependent permittivity (35, 37) on predicted LLPS properties. All of the HPS + cation– π phase diagrams here still disagree with experiment (65, 67) as they all predict a higher LLPS propensity for the RtoK variant than for WT. Apparently, the bias of the HPS potential against Arg interactions is so strong that it cannot be overcome by additional Arg–aromatic interactions that are reasonably more favorable than Lys–aromatic interactions (Fig. 2B). The $\epsilon_r = 80$ results for both uniform and variable cation– π strength exhibit an additional mismatch: contrary to experiments (4, 65), they predict similar LLPS propensities for

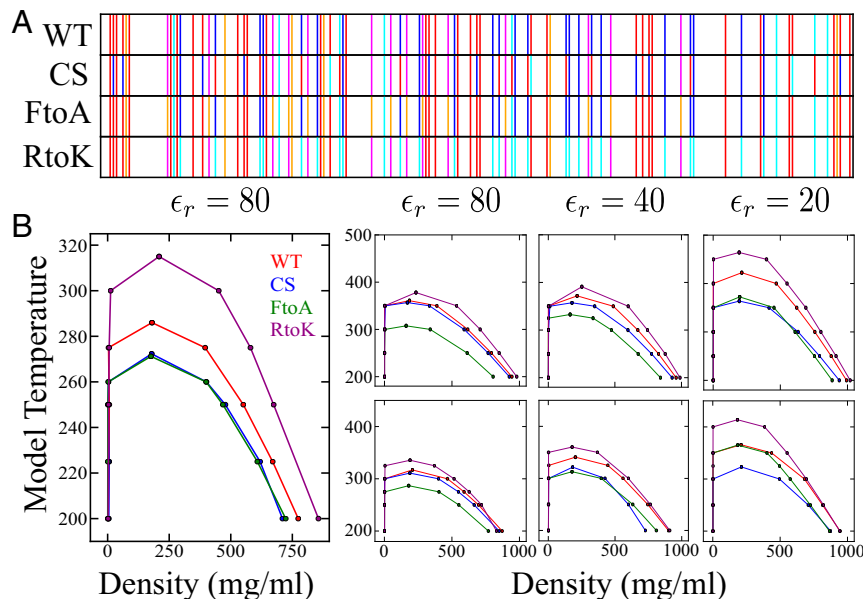


Fig. 3. Simulated phase behaviors of Ddx4 IDR variants in a hydrophobicity-dominant potential augmented by cation– π interactions. (A) Sequence patterns of the WT and its charge-scrambled (CS), Phe to Ala (FtoA) and Arg to Lys (RtoK) variants. Select residue types are highlighted: Ala (orange), Asp and Glu (red), Phe (magenta), Lys (cyan), and Arg (dark blue); other residue types are not distinguished. (B) Simulated phase diagrams of WT, CS, FtoA, and RtoK Ddx4 IDR at various relative permittivities (ϵ_r) as indicated, using the HPS model only (*Left*) and the HPS model augmented with cation– π interactions (*Right*) with either (*Top*) a uniform ($\epsilon_{c\pi}$)_{ij} as described in Fig. 2A or (*Bottom*) different ($\epsilon_{c\pi}$)_{ij} values for Arg and Lys as given in Fig. 2B.

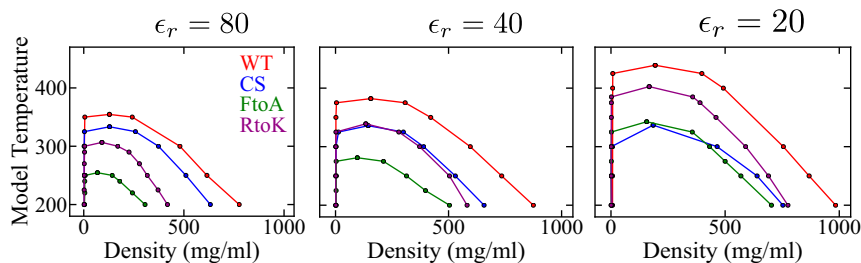


Fig. 4. Simulated phase behaviors of Ddx4 IDR variants using an interaction scheme based largely on PDB-derived statistical potentials. Phase diagrams were computed using the KH model at three different relative permittivities (ϵ_r).

the CS variant and WT, suggesting that under this dielectric condition, electrostatic interactions are unphysically overwhelmed by the presumed cation- π interactions. The $\epsilon_r = 20$ results for variable cation- π also indicate an additional mismatch; in this case they fail to reproduce the experimental trend of a significantly lower LLPS propensity of the FtoA variant relative to that of WT (4), probably because the relatively weak cation- π contribution is overwhelmed by strong electrostatic interactions in this low- ϵ_r situation. Taken together, although a perspective involving only electrostatic and cation- π interactions was adequate to account for sequence-specific LLPS trend of WT and CS (and possibly also FtoA) before the RtoK experiment was performed (34), such a perspective is incomplete when RtoK enters the picture. Fig. 3B shows that the HPS + cation- π model, which takes into account hydrophobic, charged, and cation- π interactions, cannot account for the general trend of available Ddx4 LLPS data. It follows that these interactions—at least when hydrophobicity is accorded by the particular scale (76) adopted by HPS—are insufficient by themselves to account for LLPS of IDRs in general.

Structure-Based Statistical Potentials Provide an Approximate Account of π -Related Driving Forces for Ddx4 LLPS. In contrast to the HPS and HPS + cation- π models, direct application of the KH model—without augmented cation- π terms—leads to an overall trend largely in agreement with experiments (4, 65, 67) for the ϵ_r values tested; i.e., all three Ddx4 IDR variants are predicted by the KH potential to have lower LLPS propensities than WT (Fig. 4). For WT at $\epsilon_r = 80$ and $T = 300$ K, the simulated condensed phase density of ~ 500 mg mL $^{-1}$ is comparable to the experimental value of ~ 400 mg mL $^{-1}$ with [NaCl] = 100 mM at the same temperature (65). Similar KH- and HPS-simulated condensed phase densities were obtained for FUS and LAF-1 IDRs (52). Illustrations of our simulated chain configurations are provided in Fig. 5. Time-dependent mean-square deviation of molecular coordinates has been used to verify liquid-like chain dynamics in the condensed phase of HPS and KH models (52). Examples of similar calculation are provided in *SI Appendix, Figs. S3 and S4*, for the Ddx4 IDRs examined in this work.

This success of the KH model suggests that empirical, knowledge-based statistical potentials derived from the PDB may capture key effects governing both protein folding and IDR LLPS without prejudging the dominance of, or lack thereof, particular types of energetics such as hydrophobicity in the HPS model. In this respect, it would not be surprising that cation- π and other π -related interactions are reflected in these knowledge-based potentials as well. After all, the importance of cation- π interactions in folded protein structure (69) and π - π interactions in IDR LLPS (67) is recognized largely by bioinformatics analyses of the PDB.

As discussed above, a major cause of the shortcoming of HPS in accounting for the LLPS of Ddx4 IDRs (Fig. 3B) is the high degree of unfavorability it ascribes to Arg inter-

actions. Its hydrophobicity scale, based on the atomic partial charges in the OPLS (optimized potentials for liquid simulations) force field, posits that Arg has the least hydrophobicity value of +14.5, and the next-least hydrophobic is Asp with +7.5, whereas Lys has +5.0, and the most hydrophobic is Phe with -4.0 (76). This assignment results in highly unfavorable Arg-associated interactions relative to Lys-associated interactions. In the HPS model, when one of the residues, i , in the pairwise energy $E_{ij}(r_0)$ (Fig. 1A) is Arg, the average of $E_{ij}(r_0)$ over j for all amino acids except the charged residues Arg, Lys, Asp, and Glu is equal to -0.0762 in units of kcal mol $^{-1}$, whereas the corresponding average for Lys is much more favorable at -0.1276 . When the charged residues are included, the trend is the same with the average being -0.0677 for Arg and -0.119

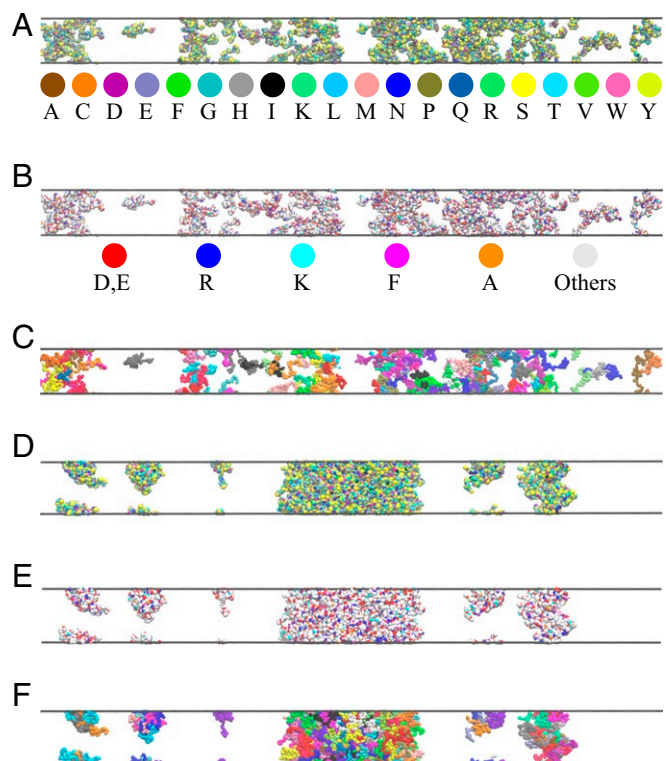


Fig. 5. Illustrative snapshots of Ddx4^{N1} CS phase behaviors simulated using the KH potential for $\epsilon_r = 40$. (A) A non-phase-separated snapshot at model temperature 375 K, wherein the amino acid residues are colored using the default VMD scheme (91, 92) as provided by the key below the snapshot. (B) Same as A except the color scheme (as shown) is essentially identical to that in Fig. 3A. (C) Same as A and B except all residues along the same chain share the same color. Neighboring chains are colored differently to highlight the diversity of conformations in the system. (D–F) A phase-separated snapshot at model temperature 325 K. The color schemes are the same as those in A–C, respectively.

for Lys. In contrast, for the KH model, the trend is opposite with Arg-associated interactions much more favorable: the corresponding average is -0.123 for Arg and -0.041 for Lys when charged residues are excluded in the averaging and -0.0990 for Arg and -0.0161 for Lys when charge residues are included. This trend echoes an earlier eigenvalue analysis of the Miyazawa–Jernigan energies (75) (which underlie the KH potential) indicating that Arg has a significantly larger projection than Lys along the dominant eigenvector (85).

Whereas correlation among hydrophobicity scales inferred from different methods is limited (86–89) with significant variations especially for the nonhydrophobic polar and charged residues (87), the extremely low hydrophobicity assigned by HPS (52, 76) to Arg relative to Lys is unusual. For instance, Lys is substantially less hydrophobic than Arg in two of the three scales tabulated and compared in ref. 89. In a commonly utilized scale based on the free energies of transfer of amino acid derivatives from water to octanol measured by Fauchère and Pliška (90) (the second scale tabulated in ref. 89), Arg is only slightly less hydrophobic ($+5.72 \text{ kJ mol}^{-1}$) than Lys ($+5.61 \text{ kJ mol}^{-1}$), and thus, essentially, Arg and Lys are deemed to possess equally low hydrophobicities. Accordingly, this scale affords a better correlation with the Miyazawa–Jernigan energies (75) (figure 3b of ref. 89) than that exhibited in Fig. 14.

It is reasonable to expect the 210 (or more) residue–residue contact energy parameters in PDB structures-based potentials to contain more comprehensive energetic information than merely the hydrophobicities of the 20 types of amino acid residues. In this regard, it is notable that a higher propensity for Arg than Lys to engage favorably with another residue appears to be a robust feature of PDB statistics. It holds for the cation–aromatic pairs we analyze in *SI Appendix, Fig. S2*, for the KH potential, and also for the original Miyazawa–Jernigan energies put forth in 1985 (74). According to table V of ref. 74, the contact energies e_{ij} between Arg and aromatic or negatively charged residues are -3.54 , -3.56 , -2.75 , -2.07 , and $-1.98 k_B T$ for Arg–Phe, Arg–Trp, Arg–Tyr, Arg–Glu, and Arg–Asp, respectively (k_B is Boltzmann constant, and T is absolute temperature), whereas the corresponding contact energies are weaker for Lys at -2.83 , -2.49 , -2.01 , -1.60 , and $-1.32 k_B T$ for Lys–Phe, Lys–Trp, Lys–Tyr, Lys–Glu, and Lys–Asp, respectively. All 20 Arg interactions are more favorable than the corresponding Lys interactions. The average e_{ij} over all Arg-associated pairs is $-2.22 k_B T$, which is substantially more favorable than the corresponding average of $-1.4795 k_B T$ for the Lys-associated pairs. It is apparent from the present application of KH to the Ddx4 IDRs that this feature is crucial, at least at a coarse-grained level, for an adequate

accounting of the π -related energetics of biomolecular LLPS. Physically, the general preference for Arg– π over Lys– π contacts is underpinned in part by the Arg geometry which allows for favorable guanidinium–aromatic coplanar packing (71), as discussed recently by Wang et al. (ref. 39 and references therein), and that Arg–aromatic is less attenuated by the aqueous dielectric medium because it is less dominated than Lys–aromatic interactions by electrostatics (83). Consistent with this perspective, R to K mutants of FUS (39) and LAF-1 RGG (77) also exhibit substantially reduced LLPS propensity relative to WT.

Phenylalanine Interactions in Liquid Condensates, Expected to Be More Solvent-Exposed, Are Weaker than Statistical Estimates Based on Mostly Core Phenylalanine Contacts in Folded Proteins.

To assess further the generality of the KH model, we apply it to simulate the phase behavior of WT LAF-1 and its RtoK and tyrosine-to-phenylalanine (YtoF) mutants (Fig. 6 and *SI Appendix, Fig. S5*). Simulation studies of LLPS of full-length and the RGG IDR (52) of LAF-1, including the latter's charge shuffled variants (77), have been conducted extensively using the HPS (52, 77) and KH (52) models to gain valuable insights, but phase behaviors of the RtoK and YtoF LAF-1 RGG mutants have not been simulated using these models. Recent experiments indicate that the RtoK mutant does not phase separate under the conditions tested, whereas the LLPS propensity of the YtoF mutant is reduced relative to that of WT (77). As for the case of Ddx4 IDR, the KH-predicted coexistence curve for RtoK LAF-1 RGG in Fig. 6 (purple curve) is consistent with the experimental trend as it has a far lower critical temperature than that of the WT (red curve). However, KH posits a significantly higher LLPS propensity for YtoF (green dashed curve with a much higher critical temperature), which is opposite to experimental observation. On closer inspection, the basic feature that causes this shortcoming is that KH deems Phe interactions more favorable than Tyr interactions. According to KH, $E_{ij}(r_0)$ for $i = \text{Phe}$ averaged over j for all amino acids is $-0.453 \text{ kcal mol}^{-1}$ for Phe, which is significantly more negative than the corresponding average $E_{ij}(r_0)$ of $-0.281 \text{ kcal mol}^{-1}$ for Tyr. For this reason, KH is unlikely to reproduce the experimentally observed reduced LLPS propensities for any other YtoF mutant either, including the YtoF mutants of FUS (39).

This overestimation of the favorability of Phe-related interactions in the LLPS context also causes the KH-predicted LLPS propensity of the FtoA mutant of Ddx4 IDR (Fig. 4, green curves) to be substantially lower than that of the RtoK mutant (purple curves). Experimentally, however, FtoA mutant LLPS is observed at $\sim 350 \text{ mg mL}^{-1}$ protein concentration, but the

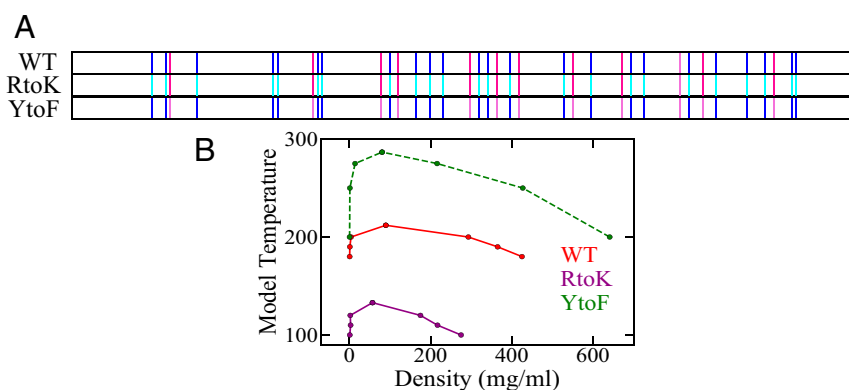


Fig. 6. Simulated phase diagrams of LAF-1 IDRs computed by the KH model at $\epsilon_r = 40$ for the three sequences in *SI Appendix, Fig. S5*. (A) Sequence patterns of the WT and two variants. Highlighted residues are Phe (magenta); Lys (cyan); Arg (dark blue), as in Fig. 3A; and Tyr (pink); other residue types, including Ala, Asp, and Glu which were highlighted in Fig. 3A, are not distinguished here. (B) Phase diagrams for WT and mutant sequences are plotted in different colors as indicated.

RtoK mutant does not phase separate up to 400 mg mL⁻¹ (in comparison, LLPS of WT is observed at 25 mg mL⁻¹) (67). In other words, the KH-predicted rank ordering of LLPS propensities of Ddx4 FtoA and RtoK is opposite to experiment. This is because the KH energy $E_{ij}(r_0)$ for Ala averaged over all 20 residue types, at -0.160 kcal mol⁻¹, is much less favorable than the corresponding -0.453 kcal mol⁻¹ for Phe and the change in average KH $E_{ij}(r_0)$ from Phe to Ala ($-0.160 + 0.453 = +0.293$ kcal mol⁻¹) is far larger than the corresponding change of $-0.0161 + 0.0990 = +0.0829$ kcal mol⁻¹ from Arg to Lys.

Interestingly, while these Ala-, Tyr-, and Phe-related KH energies do not reproduce experimental observations for YtoF and FtoA mutants, the KH energies are largely in line with rank ordering in common hydrophobicity scales. For instance, in the water-to-octanol scale of Fauchère and Pliška (90), the transfer free energies for Ala, Tyr, and Phe are -1.76 , -5.44 , and -10.1 kJ mol⁻¹, respectively. By comparison, the average interaction energies in the HPS model for Ala, Tyr, and Phe differ less, being -0.141 , -0.154 , and -0.168 kcal mol⁻¹, respectively, but Phe-related interactions are still generally more favorable than Tyr-related interactions. Thus, HPS is not only insufficient to account for RtoK experiments (Fig. 3), HPS is not expected to capture the observed YtoF trend either.

The above consideration underscores once again that the energetic contributions of amino acid residues to LLPS are not necessarily dominated by their hydrophobicities. Phe is generally considered to be more hydrophobic than Tyr. However, even for the globular proteins ribonuclease Sa and Sa3, 14 out of 16 YtoF single-site substitutions destabilize the folded state. One of the physical reasons for this behavior is that the Tyr hydroxyl group—which is absent in Phe—makes a hydrogen bond with the N^ε atom of an Arg residue (93). It would be enlightening to investigate whether similar hydrogen bonding effects account to any degree for the preference for Tyr–Arg over Phe–Arg interactions in the LLPS context.

For a relatively large residue that possesses both hydrophobic and aromatic properties such as phenylalanine, the character of its LLPS-driving contributions can also differ significantly from those stabilizing the well-packed core of a folded protein because most residue–residue contacts in the LLPS context are at least partially exposed to solvent even in the condensed phase. This observation offers a perspective to understand KH's inability to capture LLPS behaviors of YtoF and FtoA mutants because KH is derived from statistical potentials which in turn are based upon contact frequencies in folded proteins. Since Phe is one of the most buried residues in folded proteins with on average 0.88 of its surface area inaccessible to solvent upon folding (the corresponding fractions for Tyr, Ala, Lys, and Arg are 0.76, 0.74, 0.52, and 0.64, respectively) (94), the difference in character between Phe's LLPS-driving and folded-state-stabilizing interactions is expected to be more prominent.

Indeed, it has long been recognized that pair interactions in solvent-exposed environments have properties—including thermodynamic signatures (95, 96)—that are distinctly different from those of bulk interactions quantified by common hydrophobicity measurements (97). In this regard, computational evidence suggests that Phe-related interaction in a highly solvent accessible environment—which is more pertinent to LLPS—is not substantially stronger than Ala-related interactions. For instance, using ethane and ethylbenzene as models for Ala and Phe, respectively, explicit-TIP3P (3-point transferable intermolecular potential) water simulations of pairwise potentials of mean force (PMFs) at 298 K by Makowski et al. indicate that the lowest free energies at the ethane–ethane, ethane–ethylbenzene (98), and ethylbenzene–ethylbenzene (99) contact minima are all ≈ -1 kcal mol⁻¹ [Tyr was not considered in these studies (98, 99)]. Thus, in contrast to the vastly different Ala–Ala, Ala–Phe, and Phe–Phe interaction strengths in KH, with $E_{ij}(r_0) =$

-0.142 , -0.425 , and -0.756 kcal mol⁻¹, respectively (52), these explicit-water PMFs suggest a more modest difference in interaction strength between Ala and Phe in the LLPS context, which would be in line with the FtoA experiment. With this knowledge, future investigations should tackle hitherto overlooked features of solvent-accessible contacts among amino acid residues to improve coarse-grained interaction potentials for LLPS.

IDR Concentration Can Significantly Affect the Dielectric Environment of Condensed Droplets but Its Impact on LLPS Propensity Can Be Modest. In recent (52, 53, 55) and the above coarse-grained, implicit-solvent simulations of LLPS of IDRs, electrostatic interactions are assumed, for simplicity, to operate in a uniform dielectric medium with a position-independent ϵ_r . However, the dielectric environment is certainly nonuniform upon LLPS: the electrostatic interaction between two charges is affected to a larger extent by the intervening IDR materials in the condensed phase—where there is a higher IDR concentration—than in the dilute phase. Protein materials have lower ϵ_r s than bulk water (100–102). Analytical treatments with effective medium theories suggest that a decrease in effective ϵ_r with increasing IDR concentration enhances polyampholytes LLPS in a cooperative manner because the formation of condensed phase lowers ϵ_r , and that in turn induces stronger electrostatic attractions that favor the condensed phase (35, 37).

In principle, LLPS of IDR chains in polarizable aqueous media can be directly simulated using explicit-water atomic models wherein partial charges are assigned to appropriate sites of the water and protein molecules, but such simulations are computationally extremely costly because a large number of IDR chains are needed to model LLPS. Here we use explicit-water atomic simulation involving only a few IDR molecules, not to model LLPS but to estimate how the effective ϵ_r depends on IDR concentration. We will then combine this information with analytical formulations to provide a more complete account of the electrostatic driving forces for LLPS. The dielectric properties of folded proteins (100, 101), their solutions (103), and related biomolecular (104) and cellular (105) settings have long been of interest (106). For the current focus on biomolecular condensates, their interior dielectric environments are expected to be of functional significance, e.g., as drivers for various ions and charged molecules to preferentially partition into a condensate (107).

As outlined in *Materials and Methods*, GROMACS (Groningen Machine for Chemical Simulations) (108) and the amber99sb-ildn forcefield (109) are used for our explicit-water simulations using either the TIP3P (110) or the SPC/E (extended simple point charge) (111) model of water in simulation boxes containing WT Ddx4 IDR chains. Relative permittivities are estimated by fluctuations of the system dipole moment (101, 103). Simulations are also performed on artificially constructed Ddx4 (aDdx4) in which the side chain charges of Arg, Lys, Asp, and Glu are neutralized for possible applications when side chain charges are treated separately from that of the background dielectric medium. Methodological details are provided in *SI Appendix, SI Text*. Some of the simulated ϵ_r values are plotted in Fig. 7A to illustrate their dependence on IDR volume fraction ϕ (the $\phi \propto$ concentration relation and an extended set of simulated ϵ_r s are provided in *SI Appendix, SI Text, Fig. S6, and Table S1*). The difference in simulated $\epsilon_r(\phi)$ for Ddx4 and aDdx4 is negligible except at very low IDR concentration (Fig. 7A and *SI Appendix, Fig. S6*), likely because the main contribution to the dielectric effect of IDR in the atomic model is from the partial charges on the protein backbone. Consistent with expectation (35, 37), simulated $\epsilon_r(\phi)$ in Fig. 7A decreases with increasing ϕ for all solvent conditions considered. Permittivity is known to decrease with salt (112, 113). Here this expected effect is observed for TIP3P solution of IDR at low but not at high IDR concentration. Interestingly, the

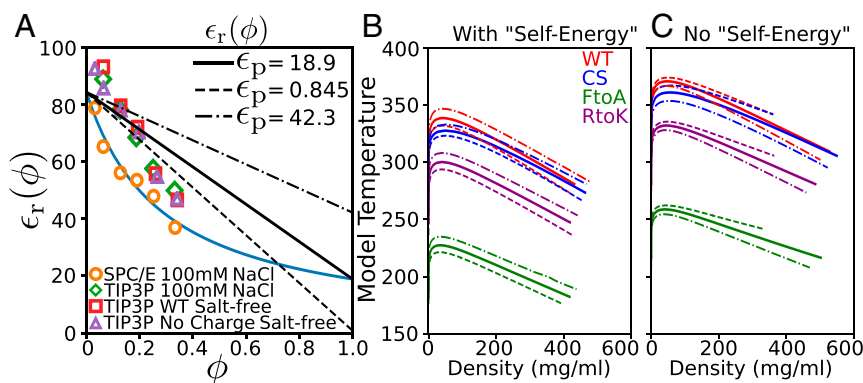


Fig. 7. Effects of IDR concentration-dependent relative permittivity on phase behaviors. (A) Relative permittivity $\epsilon_r(\phi)$ values obtained by atomic simulations (symbols) using various explicit-water models (as indicated at the bottom) are shown as functions of Ddx4 volume fraction ϕ ($\phi = 1$ corresponds to pure Ddx4). The blue curve is a theoretical fit of the SPC/E, [NaCl] = 100 mM explicit-water simulated data based on the Slab [Bragg and Pippard (114)] model (equation 34 of ref. 37), viz., $1/\epsilon_r(\phi) = \phi/\epsilon_p + (1 - \phi)/\epsilon_w$ with the fitted $\epsilon_p = 18.9$ and $\epsilon_w = 84.5$ where ϵ_p and ϵ_w are the relative permittivity of pure protein and pure water, respectively. The black solid, dashed, and dashed-dotted lines are approximate linear models of $\epsilon_r(\phi) = \epsilon_p\phi + \epsilon_w(1 - \phi)$ with the same ϵ_w but different ϵ_p values as indicated (at the top right), resulting in $d\epsilon_r(\phi)/d\phi$ slopes of -65.6 , -83.9 , and -42.2 . (B and C) Theoretical phase diagrams of the four Ddx4 IDR variants were obtained by a RPA theory that incorporates an $\epsilon_r(\phi)$ linear in ϕ . Solid, dashed, and dashed-dotted curves correspond, as in A, to the three different ϵ_p values used in the theory. The electrostatic contribution to the phase behaviors is calculated here using either (B) the expression for f_{el} given in *SI Appendix, SI Text, Eq. S51* [i.e., equation 68 of ref. 35 with its self-interaction term $\mathcal{G}_2(\vec{k})$ excluded] or (C) the full expression for f_{el} (equation 68 of ref. 35 or equivalently *SI Appendix, SI Text, Eq. S2*). Further details are provided in *SI Appendix, SI Text*.

$\epsilon_r(\phi)$ simulated with SPC/E water and 100 mM NaCl exhibits nonlinear decrease with increasing ϕ , which is akin to that predicted by the Bragg–Pippard (114) and Clausius–Mossotti models, but the TIP3P-simulated $\epsilon_r(\phi)$ appears to be linear in ϕ , which is more in line with the Maxwell Garnett and Bruggeman models (37).

We utilize the salient features of the coarse-grained KH chain model for Ddx4 (Fig. 4) and the IDR concentration-dependent permittivities from explicit-water simulations (Fig. 7A) to inform an analytical theory for IDR LLPS, referred to as RPA + FH, that combines a random phase approximation (RPA) of charge sequence-specific electrostatics and Flory–Huggins (FH) mean-field treatment for the other interactions (34, 35). An in-depth analysis of our previous RPA formulation for IDR concentration-dependent ϵ_r (35) indicates that only an $\epsilon_r(\phi)$ linear in ϕ can be consistently treated by RPA (*SI Appendix, SI Text*). In view of this recognition and considering the uncertainties of simulated $\epsilon_r(\phi)$ for different water models (Fig. 7A), three alternative linear forms of $\epsilon_r(\phi)$ (straight lines in Fig. 7A) are used for the present RPA formulation to cover reasonable variations in $\epsilon_r(\phi)$.

The mean-field FH interaction parameters in our RPA + FH models for the four Ddx4 IDRs are obtained from the four sequences' average pairwise nonelectrostatic KH contact energies. For each of the 236-residue sequences, we calculate the average of the $E_{ij}(r_0)$ [KH] quantity (Fig. 14), for a given pair of residue types, over all pairs of residues on the sequence, including a residue with itself ($236 \times 237/2 = 27,966$ pairs total), except those pairs involving two charged residues (Arg–Arg, Arg–Lys, Arg–Asp, Arg–Glu, Lys–Lys, Lys–Asp, Lys–Glu, Asp–Asp, Asp–Glu, and Glu–Glu) because interactions of charged pairs are accounted for by RPA separately. The resulting average energies in units of kcal mol^{-1} , -0.1047 for WT and CS, -0.0689 for FtoA, and -0.0924 for RtoK, are then input with an overall multiplicative scaling factor into RPA + FH theories with ϕ -independent ϵ_r for three different fixed $\epsilon_r = 80, 40$, and 20 . The computed RPA + FH phase diagrams are then fitted to the corresponding phase diagrams simulated by coarse-grained KH chain models in Fig. 4 to determine a single energy scaling factor from the best possible fit (*SI Appendix, Fig. S7*). The product of this factor and the sequence-dependent averages of $E_{ij}(r_0)$ [KH] defined above is now used as the enthalpic FH χ parameters in

the final RPA + FH theories with IDR concentration-dependent $\epsilon_r(\phi)$. Details of unit conversion between our explicit-chain simulation and our analytical RPA + FH formulation are in *SI Appendix, SI Text*.

In this connection, it is instructive to note that the corresponding averages of $E_{ij}(r_0)$ [HPS] for the HPS model are -0.1214 for WT and CS, -0.1179 for FtoA, and -0.1294 for RtoK. In this case, the more favorable (more negative) average energy of RtoK than WT underlies the mismatch of HPS prediction with experiment seen in Fig. 3B.

Fig. 7B and C show the phase diagrams of the four Ddx4 IDRs predicted using RPA + FH theories with three alternative IDR concentration-dependent $\epsilon_r(\phi)$ functions and KH-derived mean-field FH parameters as prescribed above. In all cases considered, the WT sequence (red curves) exhibit a higher propensity to LLPS than the three variants, indicating that this general agreement with the experimental trend seen in Fig. 4 holds up not only under the simplifying assumption of a constant ϵ_r but also when the dielectric effect of the IDRs is taken into account. As discussed in *SI Appendix, SI Text*, we have previously subtracted the self-energy term in the RPA formulation for numerical efficiency because the term has no impact on the predicted phase diagram when ϵ_r is a constant independent of ϕ because the self-energy contribution is identical for the dilute and condensed phases. However, with an IDR concentration-dependent $\epsilon_r(\phi)$, as for the cases considered here, the self-energy—with the short-distance cutoff of Coulomb interaction in the RPA formulation corresponding roughly to a finite Born radius (115)—is physically relevant as it decreases with increasing ϵ_r , and therefore, it affects the predicted LLPS properties as manifested by the difference between Fig. 7B and C. It follows that the self-energy term quantifies a tendency for an individual polyampholyte chain to prefer the dilute phase with a higher ϵ_r —because of its more favorable electrostatic interactions with the more polarizable environment—over the condensed phase with a lower ϵ_r . This tendency disfavors LLPS. At the same time, the lower ϵ_r in the condensed phase entails a stronger interchain attractive electrostatic force that drives the association of polyampholyte chains. Therefore, taken together, relative to the assumption of a constant ϵ_r , the overall impact of an IDR concentration-dependent $\epsilon_r(\phi)$ is expected to be modest because it likely entails a partial tradeoff between these two opposing effects. This consideration

is borne out in Fig. 7B and C. When self-energy is neglected in Fig. 7C, LLPS propensities predicted using $\epsilon_r(\phi)$ s are relatively high (as characterized by the critical temperatures), comparable to those for a fixed $\epsilon_r = 40$ in Fig. 4B. When the physical effect of self-energy is accounted for in Fig. 7B, LLPS propensities predicted using $\epsilon_r(\phi)$ s are significantly lower: overall they are comparable but slightly lower than those for a fixed $\epsilon_r = 80$ in Fig. 4A. Consistent with this physical picture, whereas the $\epsilon_r(\phi)$ with a sharper decrease with increasing ϕ leads to a higher LLPS propensity when self-energy is neglected (dashed curves have higher critical temperatures than dashed-dotted curves of the same color in Fig. 7C), for the physically appropriate formulation with self-energy, a sharper decrease in $\epsilon_r(\phi)$ with increasing ϕ leads to a lower LLPS propensity (dashed curves have lower critical temperatures than dashed-dotted curves of the same color in Fig. 7B).

Conclusion

In summary, we have gained insights into the physical forces that drive the formation of biomolecular condensates by systematically evaluating coarse-grained, residue-based protein chain models embodying different outlooks as to the types of interactions that are important for LLPS of IDRs. By comparing model predictions against experimental data on WT Ddx4 IDR and its three variants as well as WT LAF-1 RGG and its two variants, we acquire essential knowledge from agreement as well as disagreement between simulation and experiment. Aiming to account for all observed relative LLPS propensities of the sequences, we find that hydrophobicity, electrostatic, and cation- π interactions are insufficient by themselves. Rationalization of experiment on arginine-to-lysine variants entails significantly more favorable arginine-associated over lysine-associated contacts, an effect that is most likely underpinned by π - π interactions. This perspective is in line with bioinformatics analysis of LLPS propensities (67) and recent experiments on other IDRs (39, 60). In this regard, it is reassuring that the balance of forces for LLPS of IDRs appears to be partly captured by common PDB-derived statistical potentials developed to study protein folding and binding. However, we found that the condensate-stabilizing contributions of phenylalanine interactions are significantly weaker than that estimated from statistical potentials, most likely because such interactions exist in highly solvent-accessible environments rather than in the sequestered cores or binding interfaces of globular proteins. We have also highlighted the reduced electric permittivity inside con-

densed IDR phases. Although this effect's overall influence on LLPS propensity may be modest because of a tradeoff between its consequences on IDR self-energies and on inter-IDR interactions, the effect of IDR concentration-dependent permittivity by itself should be of functional importance in biology because of its potential impact on biochemical reactions and preferential partition of certain molecules into a given biomolecular condensate. All told, the present study not only serves to clarify the aforementioned issues of general principles, it also represents a useful step toward a transferable coarse-grained model for sequence-specific biomolecular LLPS. Many questions remain to be further investigated nonetheless. These include—and are not limited to—an adequate description of solvent-exposed interactions involving large hydrophobic/aromatic residues such as phenylalanine, a proper balance between attractive and repulsive interactions (53), devising temperature-dependent effective interactions (51), an accurate account of small ion effects (41, 116, 117), and incorporation of nucleic acids into LLPS simulations (43). Progress in these directions will deepen our understanding of fundamental molecular biological processes and will advance the design of novel IDR-like materials as well.

Materials and Methods

Our implicit-water coarse-grained explicit-chain modeling setup used in the first part of the present investigation for multiple IDP molecules follows largely the Langevin dynamics formulations in refs. 52, 53 for IDP LLPS. The simulation protocol features an initial slab-like condensed configuration that allows for efficient equilibration to produce simulated LLPS data (78). As discussed, model energy functions embodying different physical perspectives are considered. The subsequent explicit-water simulations for estimating IDR concentration-dependent relative permittivity are conducted on five WT Ddx4 IDRs using GROMACS (108) in conjunction with the amber99sb-ildn forcefield (109) and with TIP3P (110) or SPC/E (111) waters. Details of our methodology and the development of analytical RPA formulation are provided in *SI Appendix, SI Text*.

Data Availability. All study data are included in the article and *SI Appendix*.

ACKNOWLEDGMENTS. We thank Robert Best, Kingshuk Ghosh, Alex Holehouse, Jeetain Mittal, Rohit Pappu, Jeremy Schmit, and Wenwei Zheng for helpful discussions. This work was supported by Canadian Institutes of Health Research grants MOP-84281 and NJT-155930 to H.S.C., Natural Sciences and Engineering Research Council of Canada Discovery grants RGPIN-2016-06718 to J.D.F.-K. and RGPIN-2018-04351 to H.S.C., and computational resources provided generously by Compute/Calcul Canada.

- C. P. Brangwynne *et al.*, Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009).
- P. Li *et al.*, Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
- M. Kato *et al.*, Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
- T. J. Nott *et al.*, Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **57**, 936–947 (2015).
- A. Mollieix *et al.*, Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).
- Y. Lin, D. S. W. Protter, M. K. Rosen, R. Parker, Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell* **60**, 208–219 (2015).
- M. Zeng *et al.*, Reconstituted postsynaptic density as a molecular platform for understanding synapse formation and plasticity. *Cell* **174**, 1172–1187 (2018).
- L. D. Muiznieks, S. Sharpe, R. Pomès, F. W. Keeley, Role of liquid-liquid phase separation in assembly of elastin and other extracellular matrix proteins. *J. Mol. Biol.* **430**, 4741–4753 (2018).
- N. B. Nedelsky, J. P. Taylor, Bridging biophysics and neurology: Aberrant phase transitions in neurodegenerative disease. *Nat. Rev. Neurol.* **15**, 272–286 (2019).
- S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
- E. Gomes, J. Shorter, The molecular language of membraneless organelles. *J. Biol. Chem.* **294**, 7115–7127 (2019).
- R. J. Stewart, C. S. Wang, I. T. Song, J. P. Jones, The role of coacervation and phase transitions in the sandcastle worm adhesive system. *Adv. Colloid Interface Sci.* **239**, 88–96 (2017).
- X. Chen, X. Wu, H. Wu, M. Zhang, Phase separation at the synapse. *Nat. Neurosci.* **23**, 301–310 (2020).
- B. A. Gibson *et al.*, Organization of chromatin by intrinsic and regulated phase separation. *Cell* **179**, 470–484 (2019).
- B. Tsang *et al.*, Phosphoregulated FMRP phase separation models activity-dependent translation through bidirectional control of mRNA granule formation. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4218–4227 (2019).
- T. H. Kim *et al.*, Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. *Science* **365**, 825–829 (2019).
- L. E. Wong *et al.*, Tripartite phase separation of two signal effectors with vesicles priming B cell responsiveness. *Nat. Commun.* **11**, 848 (2020).
- Y. Fujioka *et al.*, Phase separation organizes the site of autophagosome formation. *Nature* **578**, 301–305 (2020).
- Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).
- Y.-H. Lin, J. D. Forman-Kay, H. S. Chan, Theories for sequence-dependent phase behaviors of biomolecular condensates. *Biochemistry* **57**, 2499–2508 (2018).
- S. Boeynaems *et al.*, Protein phase separation: A new phase in cell biology. *Trends Cell Biol.* **28**, 420–435 (2018).
- R. M. Vernon, J. D. Forman-Kay, First-generation predictors of biological protein phase separation. *Curr. Opin. Struct. Biol.* **58**, 88–96 (2019).
- K. M. Ruff, R. V. Pappu, A. S. Holehouse, Conformational preferences and phase behavior of intrinsically disordered low complexity sequences: Insights from multiscale simulations. *Curr. Opin. Struct. Biol.* **56**, 1–10 (2019).
- S. L. Perry, Phase separation: Bridging polymer physics and biology. *Curr. Opin. Colloid Interface Sci.* **39**, 86–97 (2019).
- G. L. Dignon, W. Zheng, J. Mittal, Simulation methods for liquid-liquid phase separation of disordered proteins. *Curr. Opin. Chem. Eng.* **23**, 92–98 (2019).
- S. Alberti, A. Gladfelter, T. Mittag, Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* **176**, 419–434 (2019).

27. H. Cinar *et al.*, Temperature, hydrostatic pressure, and osmolyte effects on liquid-liquid phase separation in protein condensates: Physical chemistry and biological implications. *Chem. Eur. J.* **25**, 13049–13069 (2019).
28. J.-M. Choi, A. S. Holehouse, R. V. Pappu, Physical principles underlying the complex biology of intracellular phase transitions. *Annu. Rev. Biophys.* **49**, 107–133 (2020).
29. P. A. Chong, J. D. Forman-Kay, Liquid-liquid phase separation in cellular signaling systems. *Curr. Opin. Struct. Biol.* **41**, 180–186 (2016).
30. X.-H. Li, P. L. Chavali, R. Pancsa, S. Chavali, M. M. Babu, Function and regulation of phase-separated biological condensates. *Biochemistry* **57**, 2452–2461 (2018).
31. C. F. Lee, J. D. Wurtz, Novel physics arising from phase transitions in biology. *J. Phys. D Appl. Phys.* **52**, 023001 (2018).
32. A. A. Hyman, C. A. Weber, F. Jülicher, Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
33. C. P. Brangwynne, P. Tompa, R. V. Pappu, Polymer physics of intracellular phase transitions. *Nat. Phys.* **11**, 899–904 (2015).
34. Y.-H. Lin, J. D. Forman-Kay, H. S. Chan, Sequence-specific polyampholyte phase separation in membranless organelles. *Phys. Rev. Lett.* **117**, 178101 (2016).
35. Y.-H. Lin, J. Song, J. D. Forman-Kay, H. S. Chan, Random-phase-approximation theory for sequence-dependent, biologically functional liquid-liquid phase separation of intrinsically disordered proteins. *J. Mol. Liq.* **228**, 176–193 (2017).
36. Y.-H. Lin, H. S. Chan, Phase separation and single-chain compactness of charged disordered proteins are strongly correlated. *Biophys. J.* **112**, 2043–2046 (2017).
37. Y.-H. Lin, J. P. Brady, J. D. Forman-Kay, H. S. Chan, Charge pattern matching as a “fuzzy” mode of molecular recognition for the functional phase separations of intrinsically disordered proteins. *New J. Phys.* **19**, 115003 (2017).
38. L.-W. Chang *et al.*, Sequence and entropy-based control of complex coacervates. *Nat. Commun.* **8**, 1273 (2017).
39. J. Wang *et al.*, A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688–699 (2018).
40. J. D. Schmit, J. J. Bouchard, E. W. Martin, T. Mittag, Protein network structure enables switching between liquid and gel states. *J. Am. Chem. Soc.* **142**, 874–883 (2020).
41. Y.-H. Lin, J. P. Brady, H. S. Chan, K. Ghosh, A unified analytical theory of heteropolymers for sequence-specific phase behaviors of polyelectrolytes and polyampholytes. *J. Chem. Phys.* **152**, 045102 (2020).
42. A. N. Amin, Y.-H. Lin, S. Das, H. S. Chan, Analytical theory for sequence-specific binary fuzzy complexes of charged intrinsically disordered proteins. *J. Phys. Chem. B* **124**, 6709–6720 (2020).
43. Y. Lin *et al.*, Narrow equilibrium window for complex coacervation of tau and RNA under cellular conditions. *eLife* **8**, e42571 (2019).
44. J. McCarty, K. T. Delaney, S. P. O. Danielsen, G. H. Fredrickson, J.-E. Shea, Complete phase diagram for liquid-liquid phase separation of intrinsically disordered proteins. *J. Phys. Chem. Lett.* **10**, 1644–1652 (2019).
45. S. P. O. Danielsen, J. McCarty, J.-E. Shea, K. T. Delaney, G. H. Fredrickson, Molecular design of self-coacervation phenomena in block polyampholytes. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 8224–8232 (2019).
46. S. P. O. Danielsen, J. McCarty, J.-E. Shea, K. T. Delaney, G. H. Fredrickson, Small ion effects on self-coacervation phenomena in block polyampholytes. *J. Chem. Phys.* **151**, 034904 (2019).
47. S. Das, A. Eisen, Y.-H. Lin, H. S. Chan, A lattice model of charge-pattern-dependent polyampholyte phase separation. *J. Phys. Chem. B* **122**, 5418–5431 (2018).
48. N. A. S. Robichaud, I. Saika-Voivod, S. Wallin, Phase behavior of blocky charge lattice polymers: Crystals, liquids, sheets, filaments, and clusters. *Phys. Rev. E* **100**, 052404 (2019).
49. J.-M. Choi, F. Dar, R. V. Pappu, LASSI: A lattice model for simulating phase transitions of multivalent proteins. *PLoS Comput. Biol.* **15**, e1007028 (2019).
50. D. Nilsson, A. Irbäck, Finite-size scaling analysis of protein droplet formation. *Phys. Rev. E* **101**, 022413 (2020).
51. G. L. Dignon, W. Zheng, Y. C. Kim, J. Mittal, Temperature-controlled liquid-liquid phase separation of disordered proteins. *ACS Cent. Sci.* **5**, 821–830 (2019).
52. G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, J. Mittal, Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Comput. Biol.* **14**, e1005941 (2018).
53. S. Das, A. N. Amin, Y.-H. Lin, H. S. Chan, Coarse-grained residue-based models of disordered protein condensates: Utility and limitations of simple charge pattern parameters. *Phys. Chem. Chem. Phys.* **20**, 28558–28574 (2018).
54. A. Statt, H. Casademunt, C. P. Brangwynne, A. Z. Panagiotopoulos, Model for disordered proteins with strongly sequence-dependent liquid phase behavior. *J. Chem. Phys.* **152**, 075101 (2020).
55. G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, J. Mittal, Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9929–9934 (2018).
56. K. M. Ruff, T. S. Harmon, R. V. Pappu, CAMELOT: A machine learning approach for coarse-grained simulations of aggregation of block-copolymeric protein sequences. *J. Chem. Phys.* **143**, 243123 (2015).
57. T. S. Harmon, A. S. Holehouse, R. V. Pappu, Differential solvation of intrinsically disordered linkers drives the formation of spatially organized droplets in ternary systems of linear multivalent proteins. *New J. Phys.* **20**, 045002 (2018).
58. V. Nguemaha, H.-X. Zhou, Liquid-liquid phase separation of patchy particles illuminates diverse effects of regulatory components on protein droplet formation. *Sci. Rep.* **8**, 6728 (2018).
59. A. Ghosh, K. Mazararakos, H.-X. Zhou, Three archetypical classes of macromolecular regulators of protein liquid-liquid phase separation. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 19474–19483 (2019).
60. E. W. Martin *et al.*, Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699 (2020).
61. M. K. Hazra, Y. Levy, Charge pattern affects the structure and dynamics of polyampholyte condensates. *Phys. Chem. Chem. Phys.* **22**, 19368–19375 (2020).
62. H. Cinar, S. Cinar, H. S. Chan, R. Winter, Pressure-induced dissolution and reentrant formation of condensed, liquid-liquid phase-separated elastomeric α -elastin. *Chem. Eur. J.* **24**, 8286–8291 (2018).
63. S. Cinar, H. Cinar, H. S. Chan, R. Winter, Pressure-sensitive and osmolyte-modulated liquid-liquid phase separation of eye-lens γ -crystallins. *J. Am. Chem. Soc.* **141**, 7347–7354 (2019).
64. H. Cinar *et al.*, Pressure sensitivity of SynGAP/PSD-95 condensates as a model for post-synaptic densities and its biophysical and neurological ramifications. *Chem. Eur. J.* **26**, 11024–11031 (2020).
65. J. P. Brady *et al.*, Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E8194–E8203 (2017).
66. S. Roberts *et al.*, Complex microparticle architectures from stimuli-responsive intrinsically disordered proteins. *Nat. Commun.* **11**, 1342 (2020).
67. R. M. Vernon *et al.*, Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* **7**, e31486 (2018).
68. L. M. Salonen, M. Ellerermann, F. Diederich, Aromatic rings in chemical and biological recognition: Energetics and structures. *Angew. Chem. Int. Ed.* **50**, 4808–4842 (2011).
69. J. P. Gallivan, D. A. Dougherty, Cation- π interactions in structural biology. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9459–9464 (1999).
70. J. P. Gallivan, D. A. Dougherty, A computational study of cation- π interactions vs salt bridges in aqueous media: Implications for protein engineering. *J. Am. Chem. Soc.* **122**, 870–874 (2000).
71. P. B. Crowley, A. Golovin, Cation- π interactions in protein-protein interfaces. *Proteins* **59**, 231–239 (2005).
72. J. Song, S. C. Ng, P. Tompa, K. A. W. Lee, H. S. Chan, Polycation- π interactions are a driving force for molecular recognition by an intrinsically disordered oncoprotein family. *PLoS Comput. Biol.* **9**, e1003239 (2013).
73. T. Chen, J. Song, H. S. Chan, Theoretical perspectives on nonnative interactions and intrinsic disorder in protein folding and binding. *Curr. Opin. Struct. Biol.* **30**, 32–42 (2015).
74. S. Miyazawa, R. L. Jernigan, Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. *Macromolecules* **18**, 534–552 (1985).
75. S. Miyazawa, R. L. Jernigan, Residue-residue potentials with a favourable contact pair term and an unfavourable high packing density term, for simulation and threading. *J. Mol. Biol.* **256**, 623–644 (1996).
76. L. H. Kapcha, P. J. Rossky, A simple atomic-level hydrophobicity scale reveals protein interfacial structure. *J. Mol. Biol.* **426**, 484–498 (2014).
77. B. S. Schuster *et al.*, Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11421–11431 (2020).
78. K. S. Sillmore, M. P. Howard, A. Z. Panagiotopoulos, Vapor-liquid phase equilibrium and surface tension of fully flexible Lennard-Jones chains. *Mol. Phys.* **115**, 320–327 (2017).
79. Y. C. Kim, G. Hummer, Coarse-grained models for simulations of multiprotein complexes: Application to ubiquitin binding. *J. Mol. Biol.* **375**, 1416–1433 (2008).
80. A. C. Murthy *et al.*, Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. *Nat. Struct. Mol. Biol.* **26**, 637–648 (2019).
81. A. E. Conicella *et al.*, TDP-43 α -helical structure tunes liquid-liquid phase separation and function. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 5883–5894 (2020).
82. F. G. Quiroz, A. Chilkoti, Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers. *Nat. Mater.* **14**, 1164–1171 (2015).
83. K. Kumar *et al.*, Cation- π interactions in protein-ligand binding: Theory and datamining reveal different roles for lysine and arginine. *Chem. Sci.* **9**, 2655–2665 (2018).
84. R. V. Pappu, X. Wang, A. Vitalis, S. L. Crick, A polymer physics perspective on driving forces and mechanisms for protein aggregation. *Arch. Biochem. Biophys.* **469**, 132–141 (2008).
85. H. S. Chan, Folding alphabets. *Nat. Struct. Mol. Biol.* **6**, 994–996 (1999).
86. D. R. DeVido, J. G. Dorsey, H. S. Chan, K. A. Dill, Oil/water partitioning has a different thermodynamic signature when the oil solvent chains are aligned than when they are amorphous. *J. Phys. Chem. B* **102**, 7272–7279 (1998).
87. P. A. Karplus, Hydrophobicity regained. *Protein Sci.* **6**, 1302–1307 (1997).
88. H. S. Chan, K. A. Dill, Solvation: How to obtain microscopic energies from partitioning and solvation experiments. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 425–459 (1997).
89. H. S. Chan, “Amino acid sidechain hydrophobicity” in *eLS, Encyclopedia of Life Sciences* (John Wiley & Sons, Chichester, United Kingdom, 2002).
90. J. L. Fauchère, V. Pliška, Hydrophobic parameters II of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. *Eur. J. Med. Chem.* **18**, 369–375 (1983).
91. W. Humphrey, A. Dalke, K. Schulten, VMD - visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
92. J. Hsin, A. Arkhipov, Y. Yin, J. E. Stone, K. Schulten, Using VMD: An introductory tutorial. *Curr. Protoc. Bioinform.* **24**, 5.7.1–5.7.48 (2008).
93. C. N. Pace *et al.*, Tyrosine hydrogen bonds make a large contribution to protein stability. *J. Mol. Biol.* **312**, 393–404 (2001).
94. G. D. Rose, A. R. Geselowitz, G. J. Lesser, R. H. Lee, M. H. Zehfus, Hydrophobicity of amino acid residues in globular proteins. *Science* **229**, 834–838 (1985).
95. M. S. Moghaddam, S. Shimizu, H. S. Chan, Temperature dependence of three-body hydrophobic interactions: Potential of mean force, enthalpy, entropy, heat capacity, and nonadditivity. *J. Am. Chem. Soc.* **127**, 303–316 (2005).

96. J. L. MacCallum, M. S. Moghaddam, H. S. Chan, D. P. Tieleman, Hydrophobic association of α -helices, steric dewetting, and enthalpic barriers to protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6206–6210 (2007).
97. R. H. Wood, P. T. Thompson, Differences between pair and bulk hydrophobic interactions. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 946–949 (1990).
98. M. Makowski *et al.*, Simple physics-based analytical formulas for the potentials of mean force for the interaction of amino acid side chains in water. 3. Calculation and parameterization of the potentials of mean force of pairs of identical hydrophobic side chains. *J. Phys. Chem. B* **111**, 2925–2931 (2007).
99. M. Makowski, E. Sobolewski, C. Czaplewski, S. Oldziej, A. Liwo, H. A. Scheraga, Simple physics-based analytical formulas for the potentials of mean force for the interaction of amino acid side chains in water. IV. Pairs of different hydrophobic side chains. *J. Phys. Chem. B* **112**, 11385–11395 (2008).
100. J. J. Dwyer *et al.*, High apparent dielectric constants in the interior of a protein reflect water penetration. *Biophys. J.* **79**, 1610–1620 (2000).
101. J. W. Pitera, M. Falta, W. F. van Gunsteren, Dielectric properties of proteins from simulation: The effects of solvent, ligands, pH, and temperature. *Biophys. J.* **80**, 2546–2555 (2001).
102. C. N. Schutz, A. Warshel, What are the dielectric “constants” of proteins and how to validate electrostatic models? *Proteins* **44**, 400–417 (2001).
103. T. Rudas, C. Schröder, S. Boresch, O. Steinhauser, Simulation studies of the protein-water interface. II. Properties at the mesoscopic resolution. *J. Chem. Phys.* **124**, 234908 (2006).
104. H. Nymeyer, H.-X. Zhou, A method to determine dielectric constants in nonhomogeneous systems: Application to biological membranes. *Biophys. J.* **94**, 1185–1193 (2008).
105. M. Tros *et al.*, Picosecond orientational dynamics of water in living cells. *Nat. Commun.* **8**, 904 (2017).
106. J. Wyman, Jr, Studies on the dielectric constant of protein solutions. I. Zein. *J. Biol. Chem.* **90**, 443–476 (1931).
107. N. Martin, Dynamic synthetic cells based on liquid-liquid phase separation. *ChemBioChem* **20**, 2553–2568 (2019).
108. M. J. Abraham, D. van der Spoel, E. Lindahl, B. Hess, GROMACS development team, GROMACS User Manual version 2016.5 (2018). <http://www.gromacs.org>. Accessed 4 November 2019.
109. K. Lindorff-Larsen *et al.*, Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).
110. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
111. H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).
112. J. B. Hasted, D. M. Ritson, C. H. Collie, Dielectric properties of aqueous ionic solutions. Parts I and II. *J. Chem. Phys.* **16**, 1–21 (1948).
113. A. Levy, D. Andelman, H. Orland, Dielectric constant of ionic solutions: A field-theory approach. *Phys. Rev. Lett.* **108**, 227801 (2012).
114. W. L. Bragg, A. B. Pippard, The form birefringence of macromolecules. *Acta Crystallogr.* **6**, 865–867 (1953).
115. Z.-G. Wang, Fluctuation in electrolyte solutions: The self energy. *Phys. Rev. E* **81**, 021501 (2010).
116. G. Orkoulas, S. K. Kumar, A. Z. Panagiotopoulos, Monte Carlo study of coulombic criticality in polyelectrolytes. *Phys. Rev. Lett.* **90**, 048303 (2003).
117. J. D. Schmit, K. A. Dill, The stabilities of protein crystals. *J. Phys. Chem. B* **114**, 4020–4027 (2010).