

Crossover in Aromatic Amino Acid Interaction Strength: Tyrosine vs. Phenylalanine in Biomolecular Condensates

Reviewed Preprint

v2 • September 24, 2025

Revised by authors

Reviewed Preprint

v1 • February 28, 2025

David De Sancho , Xabier López 

Polimero eta Material Aurreratuak: Fisika, Kimika eta Teknologia, Kimika Fakultatea, UPV/EHU & Donostia International Physics Center (DIPC), Donostia / San Sebastian, Spain

 https://en.wikipedia.org/wiki/Open_access
 Copyright information

eLife Assessment

This **important** study uses advanced computational methods to elucidate how environmental dielectric properties influence the interaction strengths of tyrosine and phenylalanine in biomolecular condensates. The evidence supporting the claims of the authors is **convincing**, as the simulations are performed rigorously providing mechanistic insights into the origin of the differences between the two aromatic amino acids considered. This study will be of broad interest to researchers studying biomolecular phase separation.

<https://doi.org/10.7554/eLife.104950.2.sa3>

Abstract

Biomolecular condensates often form through the self-assembly of disordered proteins with low-complexity sequences. In these polypeptides, the aromatic amino acids phenylalanine and tyrosine act as key “sticker” residues, driving the cohesion of dense phases. Recent studies on condensates suggest a hierarchy in sticker strength, with tyrosine being more adhesive than phenylalanine. This hierarchy aligns with experimental data on amino acid solubilities and potentials of mean force derived from atomistic simulations. However, it contradicts conventional chemical intuition based on hydrophobicity scales and pairwise contact statistics from experimental structures of proteins, which suggest phenylalanine should be the stronger sticker. In this work, we use molecular dynamics simulations and quantum chemistry calculations to resolve this apparent discrepancy. Using simple model peptides and side-chain analogues, we demonstrate that the experimentally observed hierarchy arises from the lower free energy of transfer of tyrosine into the condensate, mediated by both stronger protein-protein interactions and solvation effects in the condensate environment. Notably, as the dielectric constant of the media surrounding the stickers approaches that of an apolar solvent, the trend reverses, and phenylalanine becomes the stronger sticker. These findings highlight the role of the chemical environment in modulating protein-protein interactions, providing a clear explanation for the crossover in sticker strength between tyrosine and phenylalanine in different media.

Introduction

For decades, cellular organisation was primarily understood through the lens of membrane-bound compartmentalisation. However, it has become increasingly clear that membrane-less organelles—such as stress granules, nuclear speckles, and Cajal bodies—are also widespread within cells (Banani et al., 2017 [↗](#)). At least in some cases, these organelles form by phase separation of their components (Shin and Brangwynne, 2017 [↗](#)), primarily disordered regions of proteins and nucleic acids, which partition into dense and dilute phases. The condensates thus formed retain properties of liquids, like fusion, dripping and wetting (Brangwynne et al., 2009 [↗](#)). This phenomenon has generated significant interest in the physical mechanisms governing the phase behaviour of biomolecular mixtures (Choi et al., 2020 [↗](#)).

Many proteins undergoing phase separation share common characteristics, including deviations from typical compositions of folding proteins and simple sequences with multiple amino acid repeats. Specifically, in the sequences of proteins that experience upper critical solution temperature (UCST) transitions, like the low-complexity regions of FUS, hnRNPA1, Ddx4, LAF-1 or atGRP7, we observe long stretches rich in polar residues interspersed with aromatics or positively charged residues (Martin and Mittag, 2018 [↗](#)). Borrowing language from the theory of associative polymers (Semenov and Rubinstein, 1998 [↗](#)), these units of sequence have been termed “stickers”, in the case of aromatics and positively charged residues, and “spacers”, for the polar residue repeats (Martin et al., 2020 [↗](#); Bremer et al., 2022 [↗](#); Mittag and Pappu, 2022 [↗](#)). Spacers act as linkers that lend flexibility to the polypeptide mesh in the protein-dense phase. In contrast, stickers play a key role in determining both the single-chain properties of the polymer and the phase behaviour of the condensate through interactions involving their aromatic or charged groups (Wang et al., 2018 [↗](#)). Recent experiments have quantified the influence of different types of stickers in the polymer properties and phase behaviour (Bremer et al., 2022 [↗](#)). This raises a crucial question: what is the fundamental origin of the relative strengths of different stickers?

This matter has recently been investigated in the context of the cationic amino acids lysine (Lys) and arginine (Arg) (Wang et al., 2018 [↗](#); Das et al., 2020 [↗](#); Schuster et al., 2020 [↗](#); Fisher and Elbaum-Garfinkle, 2020 [↗](#); Greig et al., 2020 [↗](#); Paloni et al., 2021 [↗](#); Hong et al., 2022 [↗](#)). Despite having the same net charge, these residues turn out not to be interchangeable. Mutagenesis experiments on LAF-1 have shown that substituting Arg with Lys completely suppresses phase separation (Schuster et al., 2020 [↗](#)). This distinction is functionally relevant, as Arg to Lys substitutions affect speckle formation (Greig et al., 2020 [↗](#)). Additionally, experiments on the intrinsically disordered region (IDR) of Ddx4 indicate that phase separation is favoured by Arg relative to Lys (Brady et al., 2017 [↗](#); Das et al., 2020 [↗](#); Schuster et al., 2020 [↗](#)). Das and co-workers attempted to explain arginine’s greater propensity to phase separate in Ddx4 variants using coarse-grained simulations with two different energy functions (Das et al., 2020 [↗](#)). The model was first parametrized using a hydrophobicity scale, aimed to capture the “stickiness” of different amino acids (Dignon et al., 2018 [↗](#)), but this did not recapitulate the correct rank order in the stability of the simulated condensates (Das et al., 2020 [↗](#)). By replacing the hydrophobicity scale with interaction energies from amino acid contact matrices—derived from a statistical analysis of the PDB (Dignon et al., 2018 [↗](#); Miyazawa and Jernigan, 1996 [↗](#); Kim and Hummer, 2008 [↗](#))—they recovered the correct trends (Das et al., 2020 [↗](#)). A key to the greater propensity to phase separate in the case of Arg may derive from the pseudo-aromaticity of this residue, which results in a greater stabilization relative to the more purely cationic character of Lys (Gobbi and Frenking, 1993 [↗](#); Wang et al., 2018 [↗](#); Hong et al., 2022 [↗](#)).

Here we focus on the distinct roles of the main aromatic residues acting as stickers—tyrosine (Tyr) and phenylalanine (Phe)—excluding tryptophan due to its much lower abundance (Maraldo et al., 2024 [↗](#)). Given that they only differ in a hydroxyl group, one could expect Tyr and Phe to be

equally relevant to phase separation, especially considering the finding from a double-mutant cycle that Tyr-Tyr and Phe-Phe pairs make nearly identical contributions to protein stability (Serrano et al., 1991). However, experimental evidence on various proteins suggests that Tyr is a stronger driver of condensation (Lin et al., 2017; Wang et al., 2018; Schuster et al., 2020; Bremer et al., 2022). This is demonstrated by the reduced propensity to phase separate of Tyr-to-Phe mutants of FUS and LAF-1 (Lin et al., 2017; Wang et al., 2018; Schuster et al., 2020) and the opposite effect in Phe-to-Tyr mutants of hnRNPA1 (Wang et al., 2018; Bremer et al., 2022). Understanding the origin of the different sticker strengths of Phe and Tyr is important due to its functional relevance. In a large sample of hnRNPA1 variants, the fractions of Phe and Tyr have been found to co-vary, suggesting that evolutionary control of composition fine-tunes the properties of condensates (Bremer et al., 2022).

The molecular properties of Phe and Tyr may give important insights about their distinct behaviour as stickers in condensates. Hydrophobicity scales typically rank Phe as the most hydrophobic residue (Kyte and Doolittle, 1982; Tesei et al., 2021), consistent with the greater hydration free energy of Tyr relative to Phe (Wolfenden et al., 1981; Chang et al., 2007) (see **Figure 1A-B**). As in the case of Arg and Lys, using hydrophobicity as a proxy for interaction energy in bead simulation models proved insufficient to explain the relative strengths of Phe and Tyr as stickers (Das et al., 2020). However, in this case, statistical contact matrices also could not capture the correct order of stickiness. In the Miyazawa and Jernigan statistical potential, Tyr-Tyr contacts are weaker than Phe-Phe (with energies of -4.17 and -7.26 in RT units, respectively; see **Figure 1C**) (Miyazawa and Jernigan, 1996). We note that a different hydrophobicity scale based on peptides undergoing inverse temperature transitions—i.e. the Urry scale (Urry et al., 1992), where Tyr is more hydrophobic—can account for the correct rank order in saturation concentration (Regy et al., 2021). On the other hand, the potentials of mean force between amino acids calculated with atomistic force fields have a deeper free energy well for the Tyr-Tyr pair than for Phe-Phe (Chelli et al., 2002; Joseph et al., 2021). A final data point of interest is the extremely low solubility of Tyr, over an order of magnitude smaller than that of Phe (0.045 and 2.79 g/100 g of water, respectively; see **Figure 1D**) (Nozaki and Tanford, 1971). This low solubility has led to the suggestion that tyrosine-tyrosine hydrogen bonds are stronger than those formed by tyrosine and water (Pace et al., 2001).

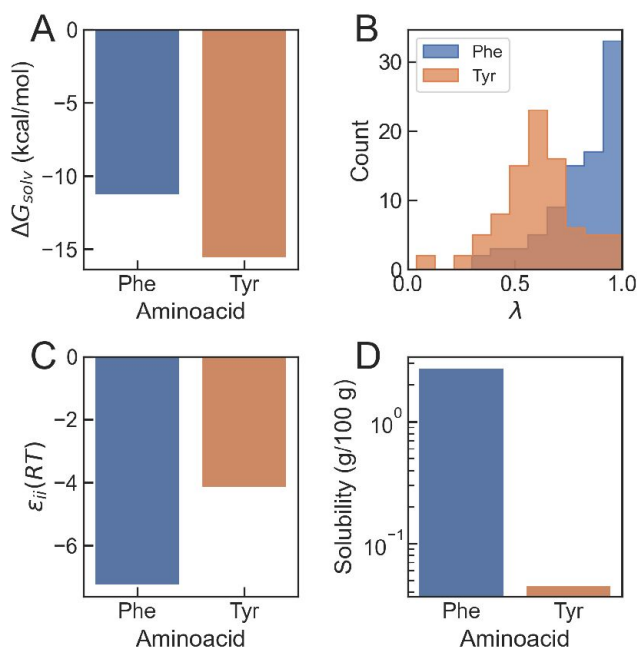


Figure 1.

Bibliographic values for different properties of phenylalanine and tyrosine.

(A) Solvation free energy (ΔG_{solv}) (Chang et al., 2007 [DOI](#)). (B) Probability distributions of min-maxed normalized hydrophathy values λ from bibliographic hydrophobicity scales (Tesei et al., 2021 [DOI](#)). (C) Self-interaction energy (ϵ_{ii}) from the Miyazawa-Jernigan contact matrix (Miyazawa and Jernigan, 1996 [DOI](#)). (D) Solubility in water at 25°C (Nozaki and Tanford, 1971 [DOI](#)).

In summary, experimental results on condensates containing Phe/Tyr variants do not seem to align with either solvation free energies, most hydrophobicity scales or statistical contact potentials, but are consistent with calculations with atomistic force fields in solution and solubilities in water. One possible explanation for these conflicting findings is that, due to their level of hydration, molecular condensates may differ significantly from the tightly-packed cores of folded protein structures (Lin et al., 2017 [DOI](#); Das et al., 2020 [DOI](#)). Here we address this paradox using a combination of classical molecular dynamics (MD) simulations and quantum chemical calculations. First, we estimate transfer free energies of peptides including these aromatic residues into model peptide condensates and find that they are more favourable for Tyr than for Phe, an effect that is reversed when we perform the same transformation in apolar media. DFT calculations confirm that the interaction energies in Tyr-Tyr pairs are stronger than those between Phe residues. However, the transfer free energy contribution dominates at sufficiently low dielectric constants, making Phe-Phe pairs more favourable. These findings recapitulate the right rank order of interaction strengths of aromatic stickers in biomolecular condensates but also their crossover in low dielectric media like the hydrophobic cores of folded proteins.

Materials and Methods

Classical molecular dynamics simulations

Molecular models

In this work, we report simulations of terminally-capped GGXGG peptides with X=F/Y in different media, including water, organic solvents and peptide condensates. This family of peptides has been characterised extensively in the past, both from experiment (Plaxco et al., 1997 [DOI](#)) and simulation (Workman and Pettitt, 2021 [DOI](#)). We have built structures for the peptides using Ambertools (Salomon-Ferrer et al., 2013 [DOI](#)). Our condensates are formed by an equimolar mixture of Gly, Ser and either Tyr or Phe “di-peptides” (in fact, terminally capped amino acids), as before (De Sancho, 2022 [DOI](#)). For most of our simulations we have used the Amber ff99SB*-ILDN force field, including corrections for backbone and side chain torsions (Best and Hummer, 2009 [DOI](#); Lindorff-Larsen et al., 2010 [DOI](#)), and TIP3P water (Jorgensen et al., 1983 [DOI](#)). To evaluate the impact of the water model in our results, we have also prepared simulations with the TIP4P-Ew water model (Horn et al., 2004 [DOI](#)). For the simulations with different organic solvents, we used the GAFF parameters derived by Mobley and co-workers (Bannan et al., 2016 [DOI](#)). Tables with a comprehensive description of the simulation systems considered in this work can be found in Appendix 1.

Thermodynamic cycle

To estimate differences in the transfer free energy ($\Delta\Delta G_{\text{Transfer}}$) of our short peptides into a different medium—either a condensate or a different solvent—we use an alchemical (i.e. non-physical) transformation (Mey et al., 2020). Specifically, the thermodynamic cycle we construct involves four different states for the GGXGG peptide, with X=F ($\lambda = 0$) or X=Y ($\lambda = 1$), which is inserted either in water or in the alternative medium. We illustrate this thermodynamic cycle in the case of transfer into a condensate in **Figure 2A**. The vertical branches of the thermodynamic cycle are associated to the free energy of transfer of the model peptides from water into the condensate $\Delta G_{\text{Transfer}}^{\text{F}}$ and $\Delta G_{\text{Transfer}}^{\text{Y}}$, while the horizontal branches involve the alchemical transformation ($\Delta G_{\text{F} \rightarrow \text{Y}}(\text{H}_2\text{O})$ and $\Delta G_{\text{F} \rightarrow \text{Y}}(\text{Cond})$). Using simulations, we can easily calculate the free energies corresponding to the horizontal branches. This enables us to estimate the desired quantity as

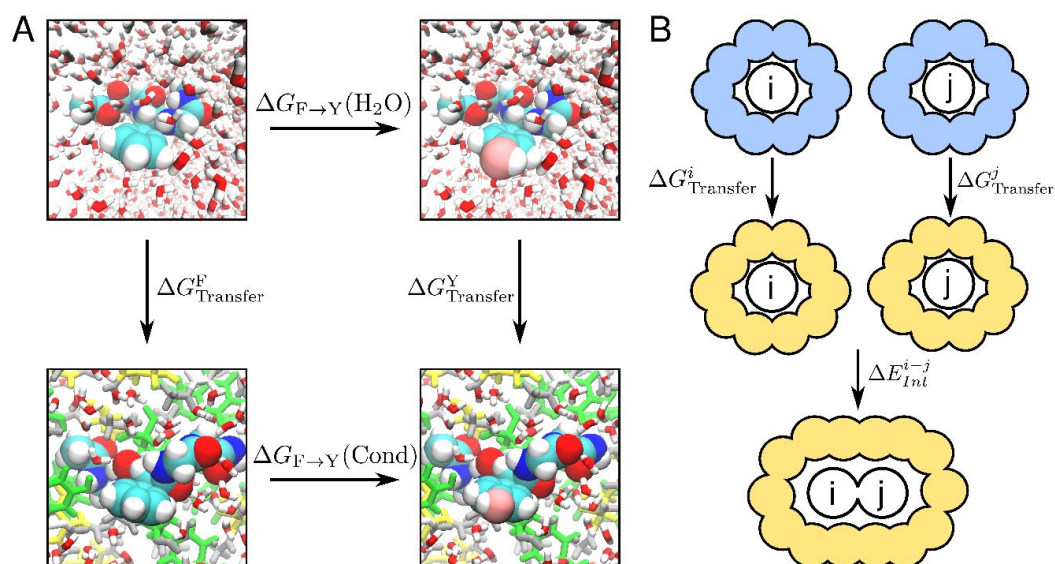


Figure 2.

(a) Thermodynamic cycle used in this study for the estimation of free energy differences upon mutation for the insertion of a peptide in a molecular condensate. (B) Schematic of the process of contact formation between two molecules *i* and *j* used in the quantum chemical calculations. We consider that contact formation involves the transfer from water (blue) to a different medium (orange) and the interaction in this medium between the entities involved.

$$\Delta\Delta G_{\text{Transfer}} = \Delta G_{\text{Transfer}}^{\text{Y}} - \Delta G_{\text{Transfer}}^{\text{F}} = \Delta G_{\text{F} \rightarrow \text{Y}}(\text{Cond}) - \Delta G_{\text{F} \rightarrow \text{Y}}(\text{H}_2\text{O}) \quad (1)$$

Negative $\Delta\Delta G_{\text{Transfer}}$ values indicate that transferring the Tyr peptide into the condensate is more favourable than transferring the Phe peptide, whereas positive values indicate the opposite.

Simulation setup

For the simulation runs in pure solvents, we simply inserted the GGXGG peptide in an octahedral box, leaving at least 1 nm of distance to each wall. Then the boxes were filled with the corresponding solvent molecules. For the simulations in the ternary model condensates, we first inserted the peptide at the central position of a rectangular box with dimensions 8.5 nm × 4.5 nm × 4.5 nm, containing an equimolar mixture of Gly, Ser and either Tyr or Phe dipeptides, which we

term GSY and GSF, respectively. Then, the box was expanded up to 20 nm and filled with water molecules. This results in a simulation box where the peptide of interest is immersed in a protein-dense slab in contact with a dilute phase.

The equilibration protocol is identical in all cases. First, we ran an energy minimization using a steepest descent algorithm, followed by short simulations in the NVT and NPT ensembles, using the Berendsen (Berendsen et al., 1984) and velocity rescaling (Bussi et al., 2007) thermostats, respectively, to set the temperature at 298 K, and the Berendsen barostat (Berendsen et al., 1984) to set the pressure at 1 bar on the latter. Finally, we performed equilibrium simulations in the NPT ensemble using a leap-frog stochastic dynamics integrator with a time-step of 2 fs at 298 K and setting the pressure to 1 bar with the Parrinello-Rahman barostat (Parrinello and Rahman, 1980). The duration of the equilibrium runs was 1 μ s for the runs with a condensate slab configuration and 500 ns for all other solvents (see Appendix 1). We run triplicates of the GSY and GSF simulations. Note that for each system-solvent combination, two different sets of simulations had to be performed, one for each of the alchemical (λ) states corresponding to the central peptide residue being Phe ($\lambda = 0$) or Tyr ($\lambda = 1$).

For the alchemical transitions, we use a non-equilibrium switching method (Seeliger and de Groot, 2010; Gapsys et al., 2015; Aldeghi et al., 2019) often used for the calculation of differences in binding affinities (Aldeghi et al., 2018; Gapsys et al., 2020) or free energy changes upon mutation (Gapsys et al., 2016; Aldeghi et al., 2019; Martinez-Martin et al., 2023). After discarding the first 100-200 ns from the long equilibrium runs, we selected 100 evenly spaced snapshots as initial states. Then, we run short (50 ps) simulation trajectories where the λ -state was switched forward ($\lambda = 0 \rightarrow 1$) or backward ($\lambda = 1 \rightarrow 0$).

We have run additional simulations of the GSY and GSF condensates in the absence of the longer peptide using the temperature replica exchange method (REMD) (Sugita and Okamoto, 1999). For these simulations, we first set up systems using the protocol described above for the slab simulations. Then, we ran a 50 ns, high-temperature simulation at 500 K in the NVT ensemble. From this trajectory, we dumped 48 randomly selected snapshots that were used as initial conformations for the different replicas. We ran REMD for 500 ns at temperatures ranging between

293.5 and 411.6 K. Replica swaps were attempted every 2 ps. The first 200 ns for each replica were discarded from the analysis. To study interactions in the biomolecular condensates, we isolated the dense phase in the slab from the REMD replica at room temperature. This was then briefly re-equilibrated as before, and production simulations of the condensates were run for 1 μ s.

For all simulations, we have used the Gromacs software package (Abraham et al., 2015) (v.2024). To generate hybrid topologies for the different λ -states, we have used the PMX software (Seeliger and de Groot, 2010; Gapsys et al., 2015; Aldeghi et al., 2019) (v. 3.0).

Analysis

The simulations have been analysed using a combination of analysis programs in the PMX and Gromacs packages (Abraham et al., 2015) and in-house Python scripts that make extensive use of the MDTraj library (McGibbon et al., 2015). For our free energy estimates, we have analysed the non-equilibrium switching simulations using the PMX tools (Seeliger and de Groot, 2010; Gapsys et al., 2015; Aldeghi et al., 2019). Specifically, we calculate the work values from the forward and backward transitions $W = \int_{\lambda=0}^{\lambda=1} \frac{\partial H}{\partial \lambda} d\lambda$ and then estimate free energy differences between the initial and final states using Bennet's Acceptance Ratio (Bennett, 1976). For more details, we refer to the PMX papers (Seeliger and de Groot, 2010; Gapsys et al., 2015; Aldeghi et al., 2019). Errors were estimated from bootstrapping, except in the case of the GSY and GSF condensates, where we estimate the standard error of the mean from the three replicates.

To obtain the phase diagrams, we calculate the peptide densities of the dilute and dense phases (ρ_L and ρ_H , respectively) from the REMD trajectories. These were obtained after fitting the density profile to the expression

$$\rho(x) = \frac{\rho_L + \rho_H}{2} + \frac{\rho_L - \rho_H}{2} \times \tanh\left(\frac{|x| - x_{DS}}{t}\right) \quad (2)$$

where x_{DS} and t are, respectively, the position and width of the dividing surface between dense and dilute phases (Tesei et al., 2021 [DOI](#)). Then, using data only for temperatures where we could clearly distinguish between dense and dilute phases, we obtained the critical temperature (T_c) from a fit to the equation

$$\rho_H - \rho_L = A(T_c - T)^\beta \quad (3)$$

where $\beta = 0.325$ is the critical exponent (Dignon et al., 2018 [DOI](#)). The critical density (ρ_c) was derived from the law of rectilinear diameters

$$\frac{\rho_H + \rho_L}{2} = \rho_c + c(T_c - T). \quad (4)$$

We used gmx dipoles to calculate the dielectric constant (ϵ) for different condensates and solvents from the fluctuations in the dipole moment of the simulation box, $\mathbf{M}$, which is defined as (Neumann, 1983 [DOI](#))

$$\epsilon = 1 + \frac{\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2}{3\epsilon_0 k_B T \langle V \rangle}. \quad (5)$$

In this expression, ϵ_0 is the vacuum permittivity, k_B is Boltzmann's constant, and V is the volume of the simulation box. For the calculation of the dielectric constant of condensates, we used the simulations of isolated dense phases mentioned above. Error analysis has been performed by blocking or determined as the standard error of the mean when derived from multiple trajectories.

Quantum chemical calculations

We performed quantum mechanical calculations using the Gaussian 16 program to study the interaction energy of the different complexes. All calculations were made at the Density Functional Theory (DFT) level, using the ω B97XD functional (Chai and Head-Gordon, 2008 [DOI](#)) and Pople's 6-311++G(d,p) basis set (Hehre et al., 1972 [DOI](#)) for geometry optimization. Energetics were refined with the 6-311++G(3df,2p) basis set. Geometry optimizations were done in different solvents, using the Polarizable Continuum Model approach (Mennucci, 2012 [DOI](#)). The interaction energy in a given solvent was calculated according to

$$\Delta E_{\text{int}}^{ij} = E^{ij} - (E^i + E^j) \quad (6)$$

where i and j specify the interacting monomers. We consider p-cresol to represent the side chain of tyrosine (Y, hereafter), toluene to represent the side chain of phenylalanine (F hereafter) and N-methyl-acetamide to represent an amide bond group (A, hereafter). Dimers ij include different combinations of Y, F, and A in various orientations.

To calculate the free energy transfer of a monomer from water to a given solvent, we used the SMD solvation model (Marenich et al., 2009 [↗](#)) to calculate accurate solvation-free energies of the monomers in different solvents, using B3LYP/6-31+G(d,p) level of theory (Becke, 1993 [↗](#); Chengteh et al., 1988 [↗](#)). Thus, the transfer free energy from water to a given solvent *s* is calculated according to

$$\Delta G_{\text{Transfer}}^i = \Delta G^i(s) - \Delta G^i(\text{H}_2\text{O}) \quad (7)$$

As solvents, we have considered three different groups:

1. Alcohols: methanol, ethanol, 1-propanol, 1-butanol, 1-pentanol, 1-hexanol, 1-nonanol and 1-decanol.
2. Aliphatic solvents: 2-nitropropane, 2-bromopropane, 1-bromopropane, 1-bromopentane, 1-fluorooctane, n-pentane, n-decane nitrobenzene, o-dichlorobenzene, bromobenzene, iodobenzene, ethylbenzene, benzene, toluene and cyclohexane.
3. Water with varying dielectric values.

These solvents span a broad range of values of the dielectric constant and can be viewed to represent different chemical environments where contact formation can take place.

Results and Discussion

Transfer free energies into a condensate slab

Our first goal is to find whether differences in transfer free energy for Phe and Tyr variants capture the hierarchy in sticker strength derived from experiments. Ideally one would calculate the transfer free energies for the protein of interest inside a realistic protein condensate. However, this would involve simulations of multiple copies of an IDR with tens or hundreds of amino acids, resulting in a system with very slow conformational dynamics (Paloni et al., 2020 [↗](#); Zheng et al., 2020 [↗](#); Galvanetto et al., 2023 [↗](#)). As a computationally efficient alternative, we use a simple model peptide with sequence GGXGG, with X being either F or Y, which is transferred into the dense mixtures of short peptides we have recently characterized (De Sancho, 2022 [↗](#)) (see **Figure 3A** [↗](#)). Peptide condensates are useful model systems as they retain many properties of protein condensates but allow for much greater computational efficiency (Paloni et al., 2020 [↗](#); Tang et al., 2021 [↗](#); De Sancho, 2022 [↗](#); Brown and Potoyan, 2024 [↗](#)).

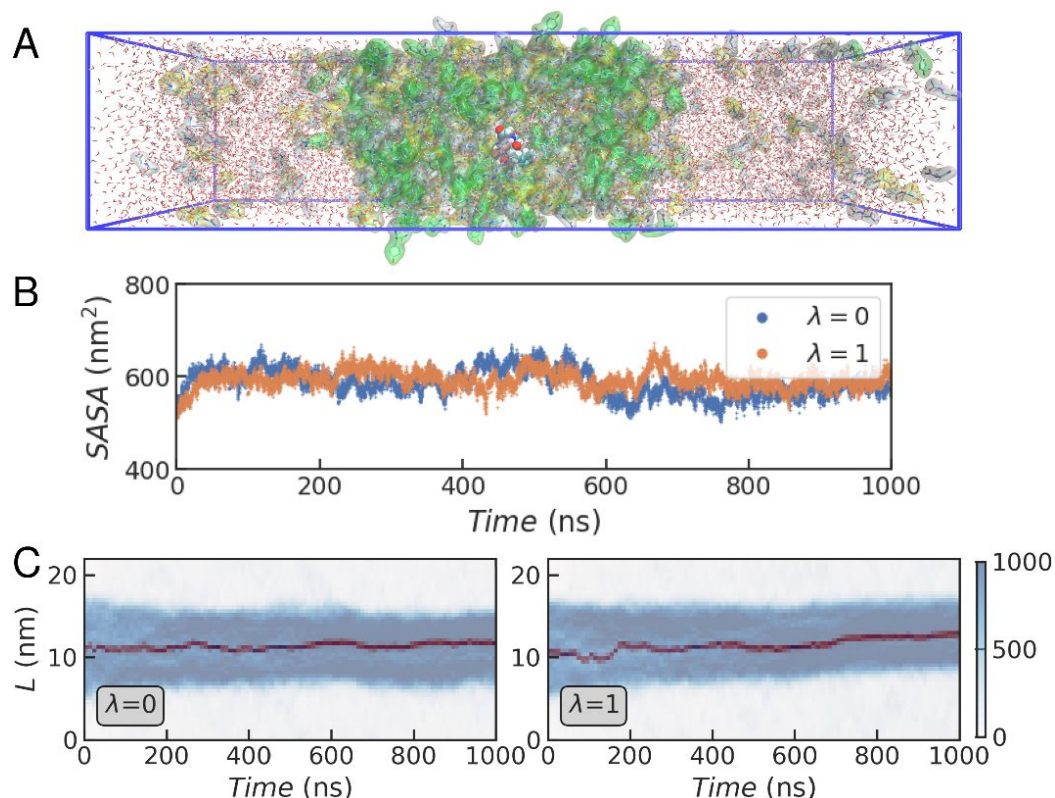


Figure 3.

(A) Representative simulation box with a fully solvated GSY condensate in slab geometry including a GGXGG peptide (spheres) and the capped amino acid mixture (G: white, S: yellow and Y: green). (B) Time series for the solvent accessible surface area (SASA) in a representative trajectory of the GGXGG peptide within the GSY condensate for different values of λ . (C) Time evolution of the density profiles calculated across the longest dimension of the simulation box (L) in the coexistence simulations. In blue we show the density in mg/mL of all the peptides, and in dark red that of the F/Y residue in the GGXGG peptide.

We define the thermodynamic cycle in [Figure 2](#) and estimate free energies associated to its horizontal branches using alchemical transformations, which we performed using a state-of-the-art non-equilibrium switching method ([Seeliger and de Groot, 2010](#); [Gapsys et al., 2015](#); [Aldeghi et al., 2019](#)) (see Methods). The first step in this procedure involves running long equilibrium simulations of the peptide of interest at the initial λ -states (i.e. where $X=F$ for $\lambda = 0$ and $X=Y$ for $\lambda = 1$). While these calculations are straightforward for a peptide in solution, within the condensate, we must first ensure that the condensate slab has converged to a stable density. Also, the peptide that experiences the transformation, which is not restrained, must remain buried within the condensate for all the snapshots that we use as initial frames, to avoid averaging the work in the dilute and dense phases. In [Figure 3B](#), we show the solvent accessible surface area (SASA) of all protein residues in a representative trajectory of the GSY condensate with the model peptide submerged. After an initial increase in SASA, the values stabilise and remain flat for the remainder of the simulation, which is indicative of stable dense and dilute phases. We also show the density profiles for all the amino acid residues in the condensate and for the mutated residue ([Figure 3C](#)). For both λ values, we find that during the whole duration of the simulation, the GGXGG peptide remains buried within the condensate. Very similar results were obtained in three independent replicates started from different initial configurations of the peptide slab (see [Figure S1](#) in Appendix 2).

We discarded the initial 100 ns of these long trajectories and selected initial configurations for non-equilibrium switching from evenly spaced snapshots of the $\lambda = 0$ and $\lambda = 1$ datasets. For each of the replicates, we ran 100 independent simulations forward and backwards, collecting the values of $(\partial H/\partial \lambda)$. Alchemical transformations were also performed in the same way for the peptide in TIP3P water (see Methods). Then we estimate transfer free energy differences by combining data from the transformations within the condensate and in water (see Appendix 2, **Figure S2A** and **B**). We find that $\Delta\Delta G_{\text{Transfer}}$ is small and negative (-2.9 ± 0.5 kJ/mol), indicating that transfer into the peptide condensate consisting of Gly, Ser and Tyr is more favourable for the Tyr-containing peptide than for the Phe variant. This result is consistent with mounting experimental evidence that suggests a greater propensity for Tyr to form condensates compared to Phe in variants of proteins such as FUS, LAF1, and hnRNPA1.

A limitation of our model peptide condensates is their high protein densities, which are larger than those of full-length IDRs (Zheng et al., 2020). This may be influenced by the propensity of the TIP3P water model to induce compact states (Piana et al., 2015). We have repeated the same procedure using the TIP4P-Ew water model (Horn et al., 2004), to estimate $\Delta\Delta G_{\text{Transfer}}$ for the GGXGG peptide into a GSY condensate (see **Figure S3** in Appendix 2). In combination with force fields from the Amber ff99SB family, this water model resulted in improved properties for peptides in dilute (Nerenberg and Head-Gordon, 2011) and dense peptide solutions (Miller et al., 2016). With this force field-water model combination, we have obtained a value of -2.6 ± 0.6 kJ/mol (see **Figure S4**), which is in good agreement with the result in TIP3P water.

One possible explanation for these results is the formation of preferential sticker-sticker interactions between the GGYGG peptide and the Tyr residues in the GSY condensate. While the hydroxyl group of Tyr can form hydrogen bonds, such interactions are expected to be depleted for the GGFGG variant. Conversely, if the condensate were composed of Phe instead of Tyr, there would be no reason to favour GGYGG. Moreover, both hydrophobicity considerations and interaction energies from contact matrices suggest that transfer into a GSF condensate should be more favourable for the GGFGG peptide than for GGYGG. To test this hypothesis, we have run the same procedure in a Gly/Ser/Phe (GSF) peptide condensate. We show the time series data for SASA and protein density for the GSF simulations in **Figure S5** of Appendix 2. Combining these results with those from the simulations in water, we obtain a value for the transfer free energy difference of $\Delta\Delta G_{\text{Transfer}} = -2.5 \pm 0.5$ kJ/mol (see Appendix 2, **Figure S2C**). Hence, irrespective of the aromatic within the condensate, transfer is more favourable for the Tyr-containing peptide.

Similar protein-protein interaction patterns in Tyr and Phe condensates

To gain further insight into the dominant interactions, we have run additional simulations of isolated dense phases for GSY and GSF condensates, without the GGXGG peptide (see **Figure 4A**). First, we look at the statistics for the different interaction pairs inside the condensates. We find that the interaction patterns are very similar in both cases, with a greater number of contacts formed by the aromatic stickers and smaller contributions from Gly and Ser (see **Figure 4B**). This result recapitulates our previous observations (De Sancho, 2022) and also those from atomistic simulations of low complexity regions of FUS and Ddx4 (Zheng et al., 2020). The greater contact probability of Tyr/Phe relative to Gly/Ser is consistent with the scaffold-client relationship between aromatic and polar residues (Dignon et al., 2020).

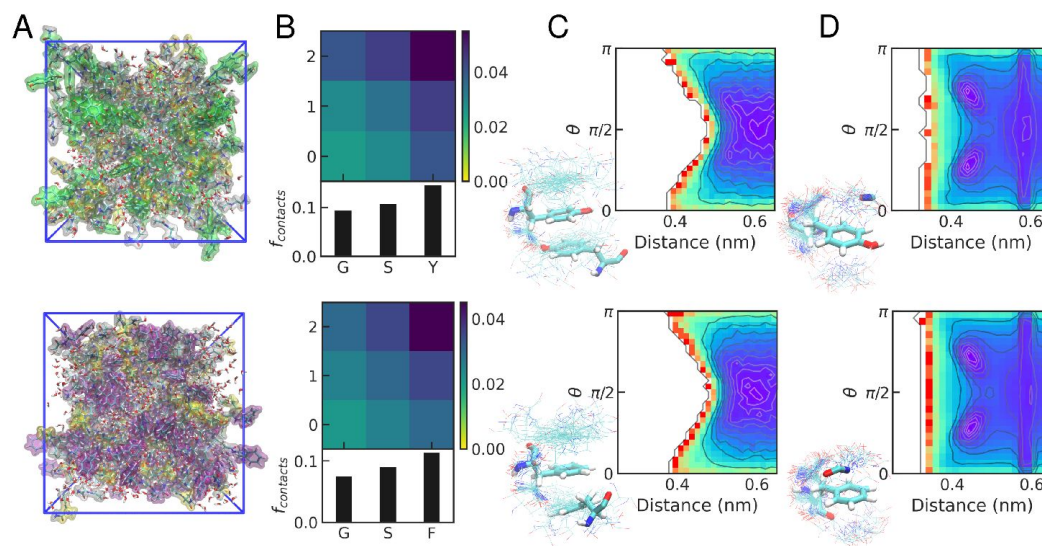


Figure 4.

(A) Representative snapshots from the simulations of dense phases. G/S/F/Y dipeptides are shown as transparent surfaces (G: white, S: yellow, Y: green, F: purple). (B) Interaction matrix for the normalized number of contacts between different pairs of amino acid residues in the condensates (top) and for each type of amino acid (bottom). (C) Density plots for side chain interactions between aromatic side chains, as characterised by the mean inter-residue distance and the angle θ between the vectors normal to the rings (Calinsky and Levy, 2024). (D) Density plots for $sp^2-\pi$ interactions between amide bonds and aromatic side chains, as characterised by the mean inter-group distance and the angle θ between the vectors normal to the peptide bond and ring planes. In all panels, results for GSY and GSF condensates are shown on the top and bottom, respectively. Representative snapshots of relevant interactions for each type of pair are shown.

To focus on $\pi-\pi$ interactions between pairs of aromatic side chains, we have also calculated the mean inter-side chain distances and the angles between vectors normal to the aromatic rings (θ). In **Figure 4C**, we show their distribution, which is very similar in the GSY and GSF condensates. Even though planar $\pi-\pi$ stacking is feasible, it does not play a dominant role (see the regions with low inter-residue distance and θ angles close to 0 or π). We also examine $sp^2-\pi$ contacts formed between aromatic and peptide bond groups, which are common in our peptide dense phases. This is consistent with previous simulation work (Zheng et al., 2020; Murthy et al., 2021; Rekhi et al., 2024) and also with a statistical analysis of experimental structures from the PDB, which showed that these interactions are particularly prevalent in disordered regions of folded proteins (Vernon et al., 2018). Interestingly, also for these $sp^2-\pi$ interactions, we find no significant differences in interaction patterns between GSY and GSF condensates.

Same interaction patterns, different stabilities

Our results indicate that Tyr's greater ability to promote condensate formation, relative to Phe, is not primarily due to differences in protein-protein interaction patterns between GSY and GSF condensates. Importantly, this similarity does not imply that the two systems have an equal propensity for phase separation, as the energetics of the interactions may differ. To interrogate this point, we have calculated the phase diagrams for the GSY and GSF condensates. We have used REMD simulations to estimate the peptide density of the dilute and dense phases at the temperatures where these can be resolved (see snapshots at high and low temperatures in **Figure 5A-B** and density profiles in **Figure S6** of Appendix 2). In **Figure 5C**, we show the calculated phase diagrams for both systems, which have remarkably similar shapes. However, the

saturation density (ρ_{sat}), i.e. the amount of peptide required for the system to phase separate, is slightly lower in the GSY condensate than in GSF. Additionally, GSY has a slightly greater critical temperature (T_c).

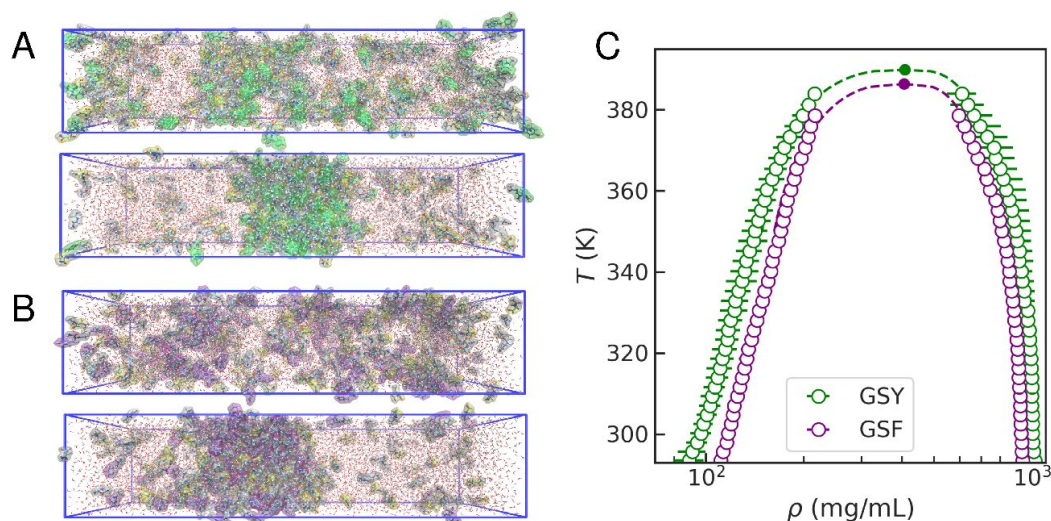


Figure 5.

(A) Representative snapshots of REMD simulations of the GSY condensate at high (top) and low (bottom) temperatures. Colour code as in Figure 3. (B) Same for the GSF condensate. (C) Phase diagrams for GSY (green) and GSF (purple). Empty circles correspond to simulations and filled circles correspond to fitted critical temperatures (T_c) and densities (ρ_c).

For simplicity, we have calculated the densities of all the peptides in the mixture as if they were partitioning uniformly. However, a more rigorous analysis would involve treating stickers and spacers separately, as we did in our more systematic study for mixtures with different concentrations of stickers and spacers (De Sancho, 2022). Given the role as a “scaffold” of the aromatics in the dense phases, we expect them to be overrepresented in the condensate relative to the spacers (“clients”). In Figure S7 we show the density profiles of aromatics at room temperature, where we find that the saturation density of Tyr is lower than that of Phe. This overemphasises, that, even if Phe and Tyr play similar roles as scaffolds in peptide condensates, Tyr has a stronger propensity to phase separate, in agreement with the hierarchy of sticker strengths observed experimentally.

Condensates as hydrated peptide solvents

We have thus far found that condensate formation by Tyr and Phe is mediated by similar contact patterns and interaction modes, dominated by contacts between aromatic residue pairs, which are more favourable for Tyr. An important driver of the observed behaviour may involve differences in protein-solvent interactions by the Phe/Tyr peptides. To investigate this, we begin by analysing the distribution of water molecules around the aromatic side chains in the dense phase simulations of GSF and GSY (see Fig. 4A), revealing notable differences. In Figure 6B, we show the radial distribution function ($g(r)$) for water oxygen around the C^γ in the aromatic side chains. In the case of Tyr, there is a prominent peak in $g(r)$ at short distances due to the presence of water molecules involved in hydrogen bonds. These interactions are enthalpically favourable and entropically unfavourable due to the water structuring around the Tyr hydroxyl group (Lin et al., 2017). The different thermo-dynamic contributions are intertwined in the calculated

$\Delta\Delta G_{\text{Transfer}}$. The net balance will depend on whether the peptide is transferred from water into a condensate, into the core of a folded protein or an altogether different solvent, capable or not of hydrogen bonding.

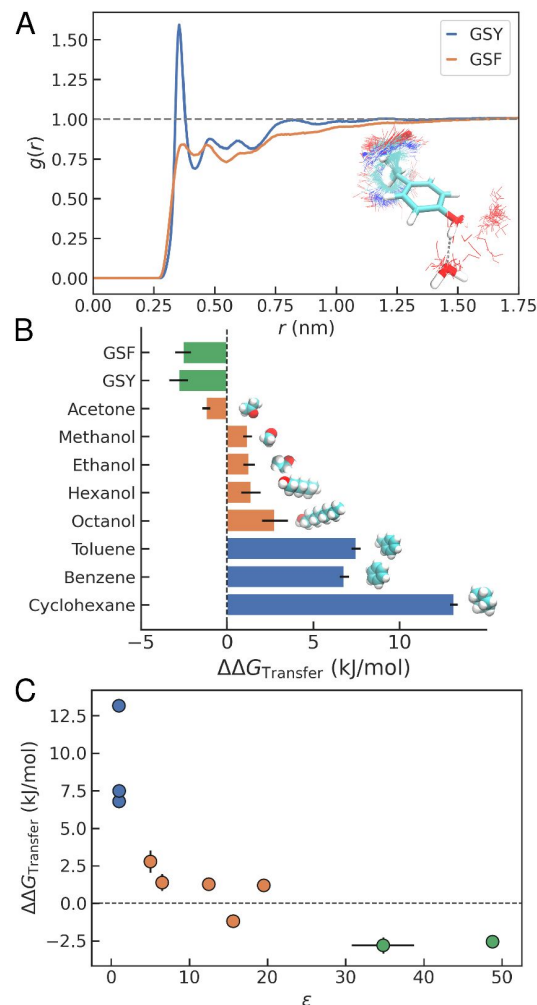


Figure 6.

(A) Radial distribution function for water oxygen around the C ζ in Phe/Tyr for GSF and GSY condensates. We show a representative overlay of simulation snapshots where water molecules are hydrogen-bonded to the Tyr side chain. (B) Transfer free energy differences from water to a different medium between Tyr and Phe. We consider condensates (green), polar solvents (orange) and apolar solvents (blue). (C) Same as a function of calculated dielectric constant, ϵ .

The picture that emerges from our transfer free energy calculations is one where condensates act as media capable of interacting with the peptide via different types of interactions, including $\pi - \pi$ and $sp_2 - \pi$ contacts. To investigate the effect of different environments, we have performed the same type of calculations using a battery of solvents with diverse properties. These include cyclohexane or octanol—used in the past as models for the interior of folded proteins (Kauzmann, 1959; Nozaki and Tanford, 1971; Radzicka and Wolfenden, 1988; Wimley et al., 1996; Pace et al., 2011)—and also shorter alcohols or acetone. The results of these calculations are shown, together with those of the different peptide condensates and solutions, in **Figure 6B**. In the case of a purely non-polar solvent like cyclohexane, the difference in free energy from Phe to Tyr is large and positive ($\Delta\Delta G_{\text{Transfer}} = 13.2 \pm 0.3$ kJ/mol), indicating that Phe is favoured. This is also the

case for aromatic solvents like benzene and toluene (with $\Delta\Delta G_{\text{Transfer}}$ of 6.8 ± 0.3 and 7.5 ± 0.3 kJ/mol, respectively). Lacking partners for the interactions that favour Tyr in the target solvent, the greater hydrophobicity of Phe dominates in the net $\Delta\Delta G_{\text{Transfer}}$. This is consistent with the stronger interaction strength of Phe-Phe interactions in the Miyazawa and Jernigan potential (Miyazawa and Jernigan, 1996) (see **Figure 1C**), derived from contacts in the apolar cores of proteins. The difference in free energy is smaller for alcohols and decreases gradually as the length of the aliphatic chain decreases. For acetone, transfer of Tyr is slightly more favourable ($\Delta\Delta G_{\text{Transfer}} = -1.2\pm0.2$ kJ/mol).

A parameter that we can determine from our MD trajectories characterising the properties of the different solvents, and also the condensates, is the dielectric constant (ϵ , see Methods). Clearly, the $\Delta\Delta G_{\text{Transfer}}$ values have a dependence on the value of ϵ (see **Figure 6C**). For the GSY and GSF condensates, ϵ values range between 30 and 50, surprisingly close to that derived from experiments on Ddx condensates using Flory-Huggins theory (45 ± 13) (Nott et al., 2015) and from atomistic simulations of Ddx4 (~ 35 – 50 at a volume fraction of $\Phi = 0.4$) (Das et al., 2020). In summary, for non-polar solvents with low ϵ , Phe is preferred to Tyr, while for polar solvents, including the condensates, the trend is reversed. At intermediate values of ϵ , there is a crossover between a regime dominated by hydrophobicity and one favoured by the stronger interactions established by Tyr.

Quantum calculations confirm a crossover in interaction strength

To characterise these subtle effects in greater detail, we resort to quantum chemical calculations, which can more accurately determine interaction strengths for specific configurations involving aromatic residues (Carter-Fenk et al., 2023; Calinsky and Levy, 2024). We consider different modes of interaction between the side chains of Phe and Tyr, represented by toluene (F) and p-cresol (Y), respectively. We considered various orientations in the optimization process and selected the two most stable structures for each case. For F-F, the most stable structures correspond to *textitcross* and *antiparallel* configurations, whereas in the case of Y-Y, *H-bonded* and *antiparallel* are most favoured (see **Figure 7A** and **C**). Our results are consistent with previous work on the toluene dimer, where the energy minima also corresponded to the cross and antiparallel structures, with only minor differences in their energies (Tsuzuki et al., 2005; Rogers et al., 2006). Other structures, like parallel or T-shaped, are less stable for toluene, although in the benzene dimer, T-shaped conformers can be favoured (Chipot et al., 1996). In our optimisations, T-shaped F-F structures tended to collapse to less stable parallel structures, and hence we do not include them here. The interaction energy values we obtained are qualitatively similar to those in the literature (Calinsky and Levy, 2024), though slightly more negative, and they follow the same rank order (i.e. Y-Y H-bonded < antiparallel Y-Y < cross F-F). We also calculate the energies for Phe and Tyr side chains interacting with an amide group (A) representing a peptide bond in the prevalent sp^2 - π interactions (Vernon et al., 2018) (see **Figure 7C**). In the case of Tyr, we have also analyzed the configuration where it forms a hydrogen bond with the amide through the alcohol group of p-cresol.

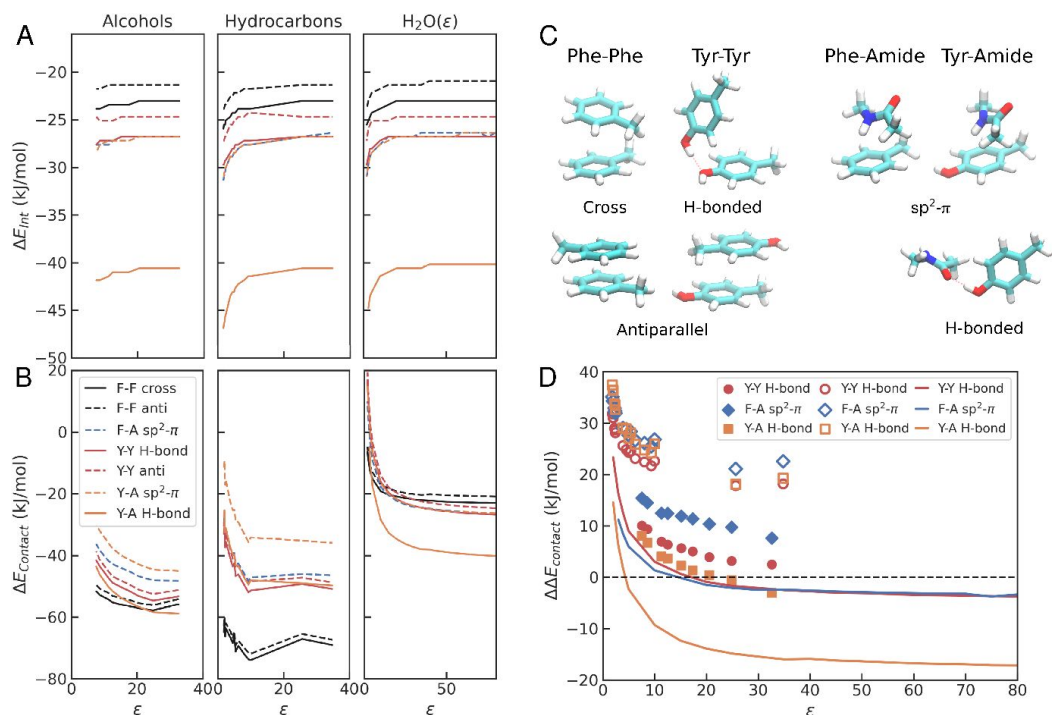


Figure 7.

(A) ΔE_{int}^{ij} and (B) $\Delta E_{contact}^{ij}$ for different interaction pairs ij and different solvents as a function of the dielectric constant, ϵ . (C) Optimized geometries from quantum chemical calculations for Phe and Tyr interaction pairs. (D) $\Delta\Delta E_{Contact}^{Solvent}$ with respect to the cross-F-F interaction for different solvents as a function of the dielectric. Different symbols correspond to the different solvent types. Filled: alcohols; empty: alkanes/benzenes; lines: water with different dielectric.

To explore the environmental dependence in contact formation for amino acid pairs, we have calculated values of the interaction energies (ΔE_{int}^{ij}) considering a set of protic and non-protic solvents (see **Figure 7A**). Additionally, we have run the same calculations using a modified water with the dielectric constant as a free parameter. The overall trends in interaction energies are very similar across interaction types. In all cases, the interaction energy tends to become stronger as the value of the dielectric decreases. Also, the relative strength of interactions is consistent across solvents. The strongest interaction corresponds to the Y-A hydrogen-bonded interaction. Next comes a series of partial π -stacking configurations with contributions of hydrogen bonds and $sp^2-\pi$ interactions, namely, Y-Y H-bond, Y-A $sp^2-\pi$ and F-A $sp^2-\pi$ structures, which we have found to be particularly abundant in our MD runs. Finally, the π -stacked F-F cross is the weakest interaction among those included in this analysis.

Therefore, if we consider these interaction energies as the primary driving force for condensate formation, Tyr should be more adhesive than Phe independently of the environment. However, as realized by Miyazawa and Jernigan (Miyazawa and Jernigan, 1996; Thomas and Dill, 1996), contact formation, either within the core of folded proteins or in the case of molecular condensates, involves two processes: first, the transfer of these groups from water to an

environment with a lower dielectric; and second, the interaction between side chain groups (or with the backbone), which is captured by $\Delta E_{\text{Int}}^{ij}$ (see **Figure 2B**). Hence, for a given solvent, the total energy can be calculated from the sum of these two contributions as

$$\Delta E_{\text{Contact}}^{ij} = \Delta E_{\text{Int}}^{ij} + \Delta G_{\text{Transfer}}^i + \Delta G_{\text{Transfer}}^j \quad (8)$$

The last two terms on the right-hand side of this expression capture the propensity of a monomer to be transferred from water to the solvent, which can be modelled using polarizable continuum models such as SMD (Marenich et al., 2009). Notice that when the solvent is a low dielectric non-protic solvent such as cyclohexane, $\Delta G_{\text{Transfer}}^i$ recapitulates the results from a standard hydrophobicity scale (Kyte and Doolittle, 1982).

When both contributions are considered, the picture changes dramatically (see **Figure 7B**). In non-protic aliphatic solvents, such as alkanes and benzene derivatives, interaction for F-F pairs becomes more favourable than in Y-Y pairs, irrespective of the dielectric constant. On the other hand, in alcohols, $\Delta E_{\text{Contact}}^{ij}$ tends to favour Y-A interactions at high ϵ and F-F pairs at low ϵ . In other words, at intermediate dielectrics, we find a crossover between F and Y in the propensity to form pairwise contacts in solvents. In our calculations, we have also considered a water-like solvent where the dielectric constant can be tuned as a free parameter (see **Figure 7C**). Also, we find a similar tendency in this case, but now the crossover is obtained at slightly lower values of ϵ . In **Figure 7D**, we show $\Delta\Delta E_{\text{Contact}}^{ij}$ values, where we are using as reference the energy of the F-F cross configuration, for three selected interactions: Y-Y H-bond, Y-A H-bond, and F-A, for the three types of environments. Negative values of $\Delta\Delta E_{\text{Contact}}^{ij}$ indicate a preference of the corresponding interaction relative to F-F cross, while positive values indicate that cross-F-F is stronger. Clearly, there is a crossover between a regime where toluene forms stronger interactions and one where p-cresol is preferred, which is highly influenced by the type of environment. In a non-protic solvent, where hydrogen bonding is not possible, or in a protic solvent with low dielectric constant, we find that F-F interactions are favoured. In contrast, in high dielectric environments, Y-Y and Y-A interactions dominate.

Conclusions

In this work, we have aimed to reconcile the apparent contradictions between the lower solvation-free energy (Wolfenden et al., 1981; Chang et al., 2007) and stronger hydrophobicity of Phe (Kyte and Doolittle, 1982; Tesei et al., 2021) and the greater contact energies assigned to Phe-Phe interactions in proteins (Miyazawa and Jernigan, 1996) with the higher sticker strength of Tyr in biomolecular condensates (Lin et al., 2017; Wang et al., 2018; Schuster et al., 2020; Bremer et al., 2022). In the past, the different propensities of positively charged stickers to form condensates had been explained from the fundamental properties of Lys and Arg (Wang et al., 2018; Das et al., 2020; Gobbi and Frenking, 1993; Hong et al., 2022). However, an explanation for the different strengths of Phe and Tyr remained elusive (Das et al., 2020).

Through a combination of classical MD simulations and quantum mechanical calculations, we have found that, in addition to the distinct types of interactions formed by Phe and Tyr, a crucial factor in their overall propensity for contact formation across different environments is the transfer free energy. At one extreme, we find environments such as the well-packed core of a protein, where the dielectric constant can be as low as 2-4 (Pitera et al., 2001). This type of environment can be modelled using aliphatic solvents, an approach followed extensively in the past (Kauzmann, 1959; Nozaki and Tanford, 1971; Radzicka and Wolfenden, 1988; Wimley et al., 1996; Pace et al., 2011). In this low-dielectric regime, Phe-Phe interactions are the most favourable, as evidenced by both our classical transfer free energies (**Figure 6C**) and the estimated quantum $\Delta\Delta E_{\text{Contact}}^{ij}$ values (**Figure 7**).

This explains why Phe forms the strongest interactions in contact matrices like that by Miyazawa and Jernigan, which were derived from statistics from 3D structures of proteins (Miyazawa and Jernigan, 1996 [DOI](#)) and are known to be dominated by hydrophobicity effects (Pokarowski et al., 2005 [DOI](#)). At the other extreme, in environments with high dielectric constants such as aqueous or polar solvent conditions, Tyr-Tyr interactions become more favourable, as inferred from potentials of mean force derived from atomistic molecular simulations (Chelli et al., 2002 [DOI](#); Joseph et al., 2021 [DOI](#)). Biomolecular condensates, due to their high water content (~200-300 mg/mL in our simulations, but up to ~600 mg/mL in experiments, (Zheng et al., 2020 [DOI](#))), are best modelled as a solvent with an intermediate dielectric and favourable interactions between stickers. Hence, they operate in the second regime, although the influence of the dielectric is subtle, due to the possibility of crossover. We note that, although we have not included in our analysis positively charged residues that form cation- π interactions with aromatics, the observed crossover will also be relevant to Arg/Lys contacts with Phe and Tyr. Following the rationale of our findings, within condensates, cation-Tyr interactions are expected to promote phase separation more strongly than cation-Phe pairs. Our findings align with a recent report from Rekhi et al. using the equilibrium alchemical techniques that emphasizes the role of transfer free energy in biomolecular condensates (Rekhi and Mittal, 2025 [DOI](#)).

The conclusions from our work are necessarily limited due to our reductionist approach. First, in our calculations, we are considering very short peptides including a single sticker, and hence we cannot account for the constraints imposed by chain connectivity (Shimizu and Chan, 2001 [DOI](#)) or the influence of multivalency in protein IDRs (Martin et al., 2020 [DOI](#)). Also, we have used a minimal alphabet of possible interaction pairs limited to two types of amino acid residues. We do not consider effects relative to cation- π interactions that have been characterised by others (Brady et al., 2017 [DOI](#); Das et al., 2020 [DOI](#); Aldeghi et al., 2018 [DOI](#); Hazra and Levy, 2023 [DOI](#); Calinsky and Levy, 2024 [DOI](#)). In our calculations, the minimal peptides result in high protein densities in the dense phases, which are nonetheless consistent with those found by other authors in peptides with greater variation in length and sequence space (Workman et al., 2024 [DOI](#); Brown and Potoyan, 2024 [DOI](#); Emelianova et al., 2025 [DOI](#)). Additionally, results may vary somewhat when considering different classical force fields. Recent work points to a substantial variation in hydrophobicity across force fields (Lobo et al., 2025 [DOI](#)). Another potential limitation may arise from the choice of water model, as the dielectric constant—which plays an important role in our observations—is known to vary significantly between water models. In the case of the TIP3P model used in most of our calculations, it exceeds the experimental value ($\epsilon = 100 \pm 5$, (Braun et al., 2014 [DOI](#))). However, using the TIP4P-Ew water model (with $\epsilon = 64 \pm 1$, (Horn et al., 2004 [DOI](#))), we have found consistent results. Careful benchmarking of force fields against properties of peptide condensates offers a promising avenue for future efforts in force-field calibration.

The methodology proposed in the present work can be easily extended to other types of interactions. The overall qualitative conclusions on the crossover in interaction strengths as a function of the environment are likely to be maintained if different quantum approaches are used for contact calculations or different force fields are used for the classical simulations. The context dependence of interaction strengths that we observe is related to a growing body of work in the study of peptide self-assembly (Kaygisiz et al., 2025 [DOI](#)). The effect has deep roots in the concepts of “pair” and “bulk” hydrophobicity (Wood and Thompson, 1990 [DOI](#); Shimizu and Chan, 2002 [DOI](#)). While pair hydrophobicity is connected to dimerisation equilibria (i.e. the second step in **Figure 2B** [DOI](#)), bulk hydrophobicity is related to transfer processes (the first step). Our work stresses the importance of considering both the pair contribution that dominates at high solvation, and the transfer free energy contribution, which overwhelms the interaction strength at low dielectrics.

Acknowledgements

The authors gratefully acknowledge conversations with Priya R. Banerjee, Rohit Pappu and Tanja Mittag, which inspired this research. The work has been financed by the Spanish Ministry of Science and Innovation through project PID2021-127907NB-I00, and the Basque Government (Project IT1584-22). The authors also thank the IZO-SGI SGiker (UPV/EHU/ERDF,EU) and DIPC for technical and human support and for the allocation of computational resources. The authors thankfully acknowledge RES resources provided by the Barcelona Supercomputing Center in MareNostrum5 (BCV-2025-2-0002).

Appendix 1

Supporting Tables

System	Number of G/S/F-Y residues
GSY	300 / 300 / 300
GSF	300 / 300 / 300
GSY REMD	145 / 145 / 145
GSF REMD	145 / 145 / 145

Appendix 1—table 1.

Copies of each of the terminally capped amino acid residues in the condensate simulations.

For GSY and GSF condensates, we use an equimolar mixture of all three components.

System	Number of molecules
Cyclohexane	500
Benzene	500
Toluene	500
Methanol	1000
Ethanol	1000
Hexanol	500
Octanol	500
Acetone	1000

Appendix 1—table 2.

Number of molecules used in pure solvent simulations for the determination of dielectric constants.

System	Simulation Time (μ s)
GSY slab	3 $\times$ 1
GSF slab	3 $\times$ 1
GSY slab TIP4P-Ew	1
GSY slab REMD	48 $\times$ 1
GSF slab REMD	48 $\times$ 1
Solvents	0.5

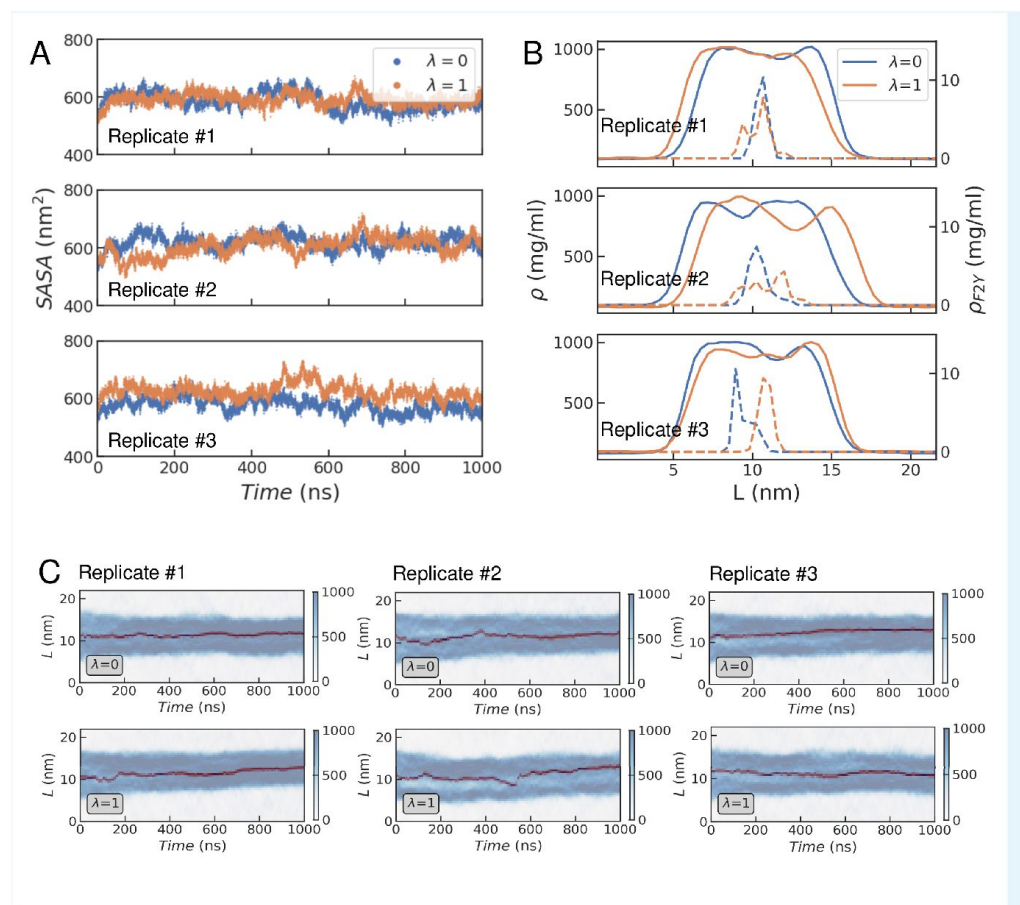
Appendix 1—table 3.

Simulation times for each of the families of systems considered in this study.

For GSY and GSF condensates, we run triplicates. For the simulations in pure solvents (including water), a single simulation run was performed for each system.

Appendix 2

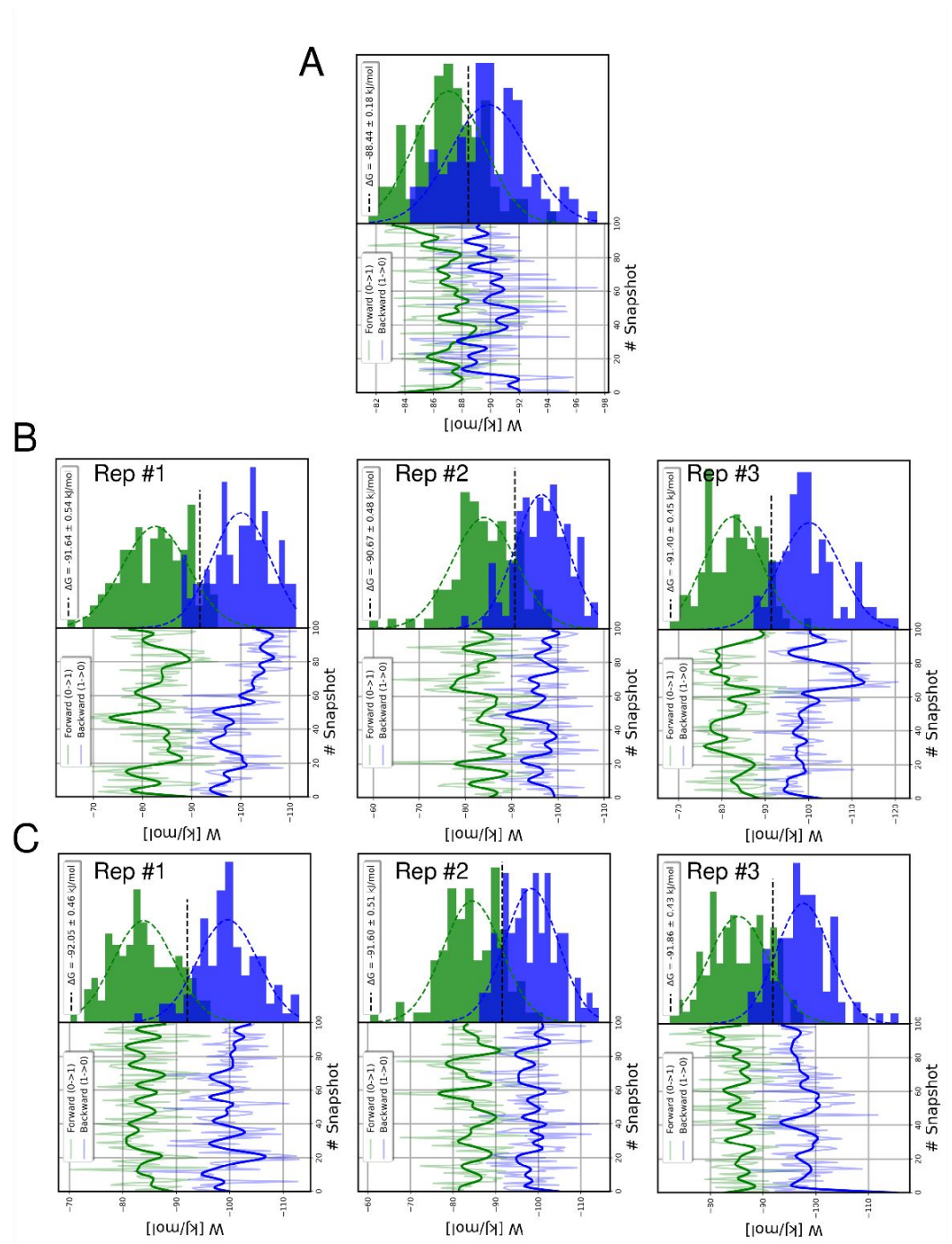
Supporting Figures



Appendix 2—figure S1.

Replicates of the GGXGG peptide simulations in the GSY condensate.

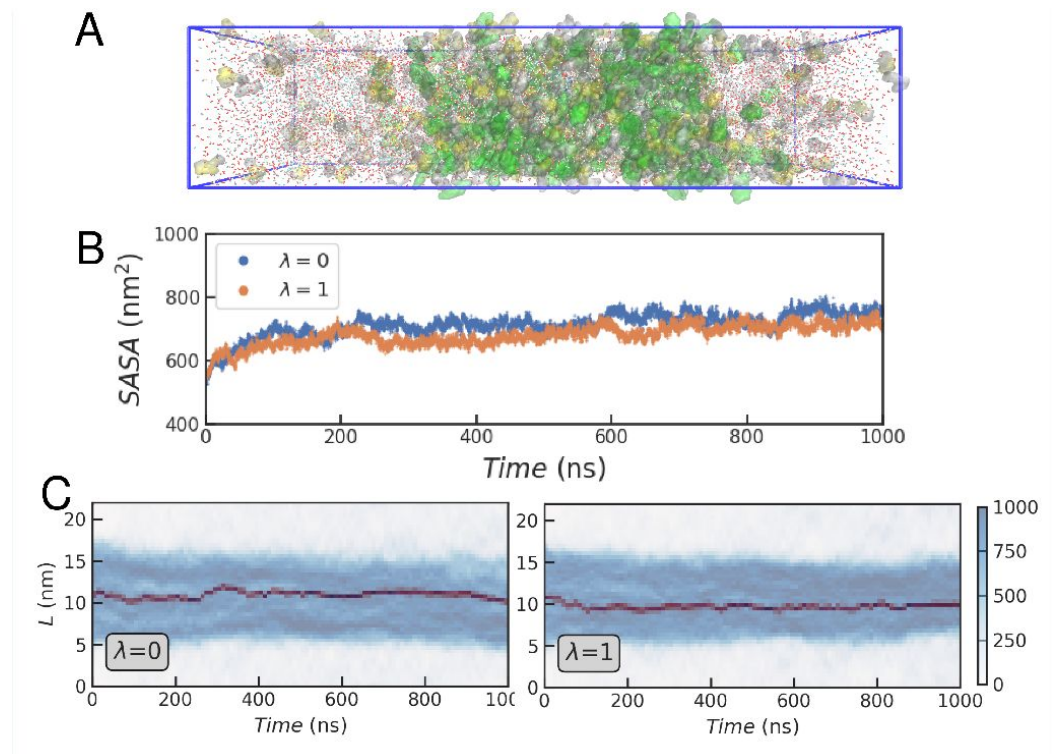
(A) Time series for the accessible surface area for different values of λ . (B) Density profiles of peptide (solid lines) and F/Y residue (right Y-axis) averaged over the simulation runs. (C) Time series for the density profiles of all peptides (blue) and the F2Y residue (red).



Appendix 2—figure S2.

Work values for the forward and reverse alchemical transitions for different initial snapshots (bottom) and their distributions (top).

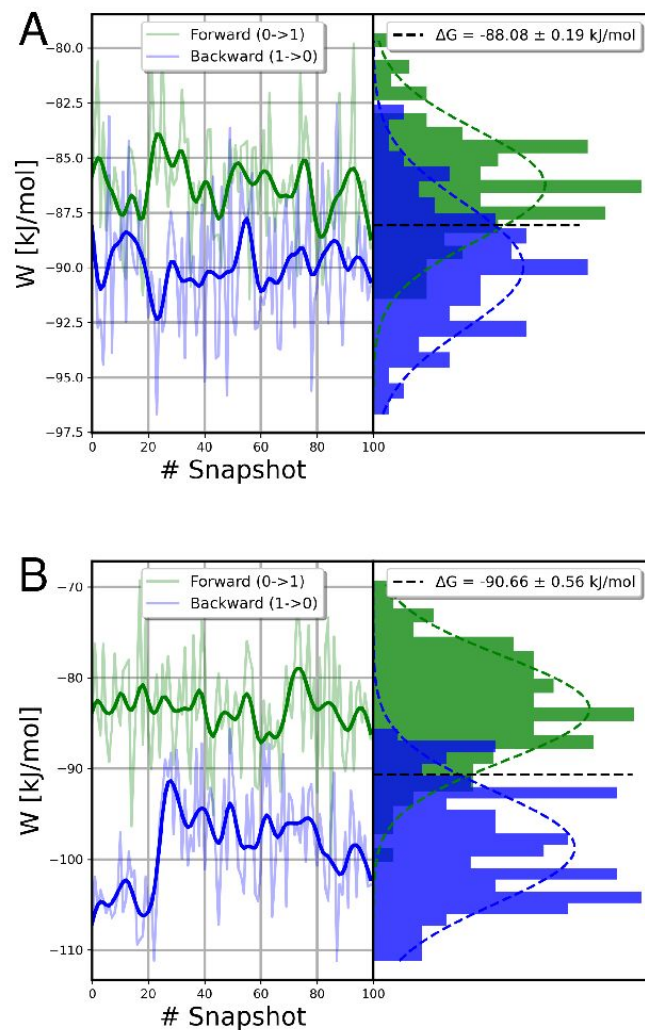
The BAR estimator is shown in the legend. Each panel corresponds to a different solvent. (A) TIP3P water. (B) Replicates of GSY condensates. (C) Replicates of GSF condensates.



Appendix 2—figure S3.

Results for the GSY condensate in the TIP4P-Ew water model.

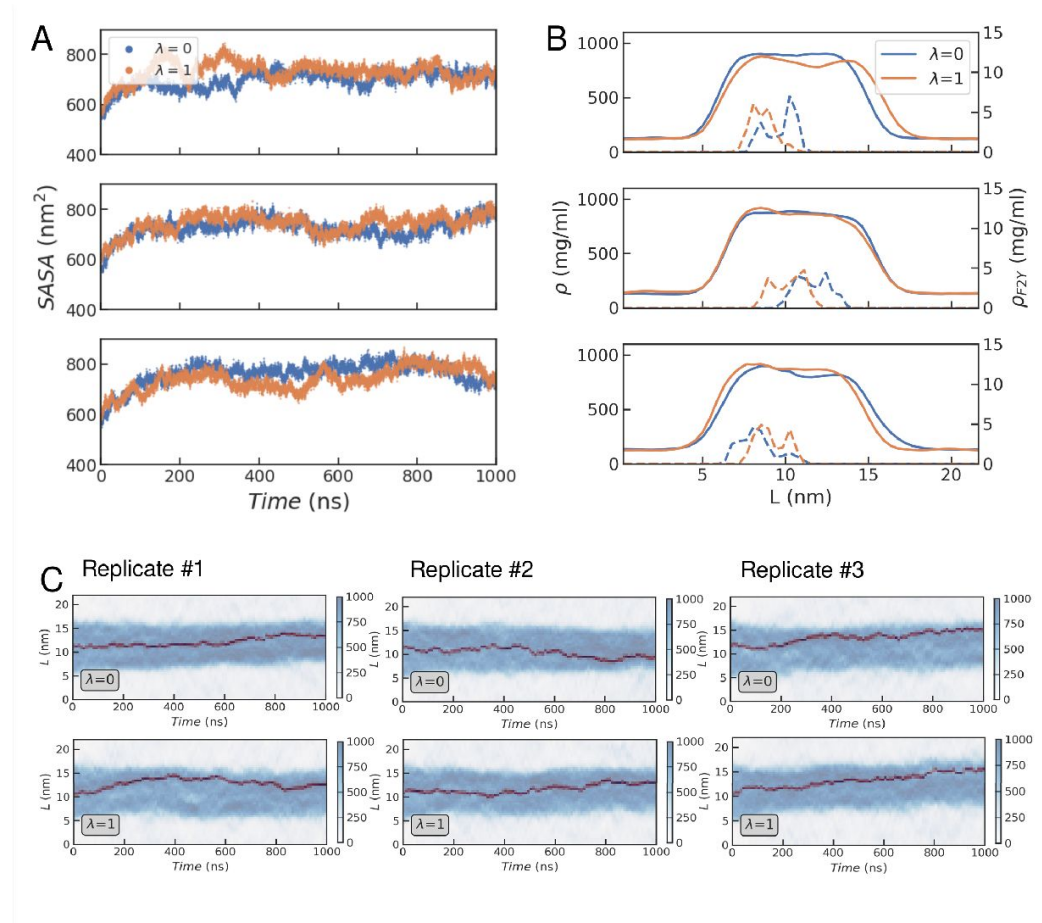
(A) Representative simulation box with a fully solvated GSY condensate in slab geometry including a GGXGG peptide (spheres) and the capped amino acid mixture (G: white, S: yellow and Y: green). (B) Time series for the solvent accessible surface area (SASA) for different values of λ . (C) Time evolution of the density profiles calculated across the longest dimension of the simulation box (L) in the coexistence simulations.



Appendix 2—figure S4.

Results for the GSY condensate in the TIP4P-Ew water model.

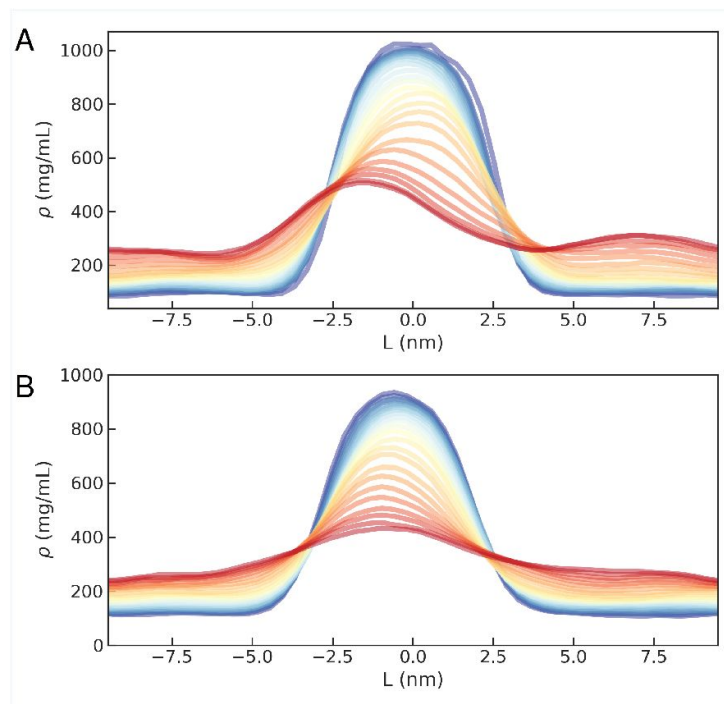
Work values and their distribution for forward and backward transitions within water (A) and the GSY condensate (B).



Appendix 2—figure S5.

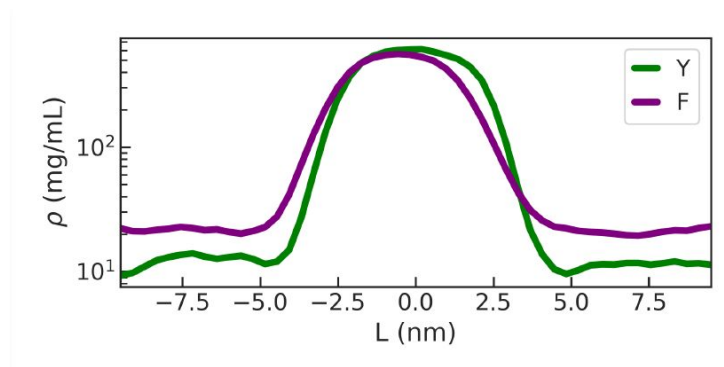
Replicates of the GGXGG peptide simulations in the GSF condensate.

(A) Time series for the accessible surface area for different values of λ . (B) Density profiles of peptide (solid lines) and F/Y residue (right Y-axis) averaged over the simulation runs. (C) Time series for the density profiles of all peptides (blue) and the F2Y residue (red).



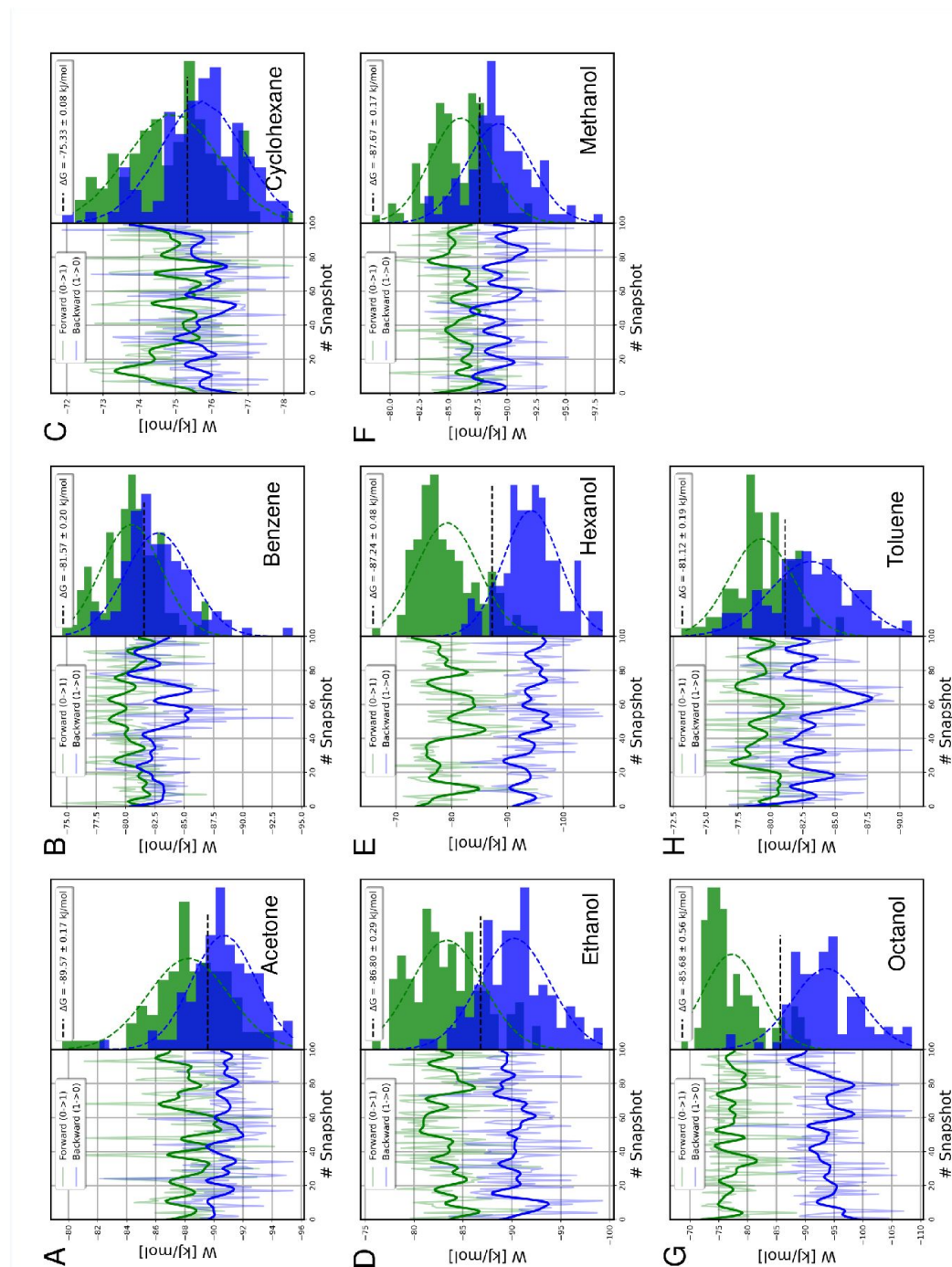
Appendix 2—figure S6.

Density profiles at different temperatures for the GSY (A) and GSF (B) condensates.



Appendix 2—figure S7.

Density profiles for sticker residues across the simulation box from the REMD simulations at room temperature.



Appendix 2—figure S8.

Work values for the forward and reverse alchemical transitions for different initial snapshots (left) and their distributions (right).

The BAR estimator is shown in the legend. Each panel corresponds to a different solvent: (A) Acetone. (B) Benzene. (C) Cyclohexane. (D) Ethanol. (E) Hexanol. (F) Methanol. (G) Octanol. (H) Toluene.

References

- Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. (2015) **GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers** *SoftwareX* **1-2**:19–25 <https://doi.org/10.1016/j.softx.2015.06.001> | Google Scholar
- Aldeghi M, Gapsys V, de Groot BL (2018) **Accurate estimation of ligand binding affinity changes upon protein mutation** *ACS Cent Sci* **4**:1708–1718 [Google Scholar](#)
- Aldeghi M, de Groot BL, Gapsys V. (2019) **Accurate Calculation of Free Energy Changes upon Amino Acid Mutation** In: Sikosek T, editors. *Computational Methods in Protein Evolution* New York: Springer pp. 19–47 https://doi.org/10.1007/978-1-4939-8736-8_2 | Google Scholar
- Banani SF, Lee HO, Hyman AA, Rosen MK (2017) **Biomolecular condensates: organizers of cellular biochemistry** *Nat Rev Mol Cell Biol* **18**:285–298 <https://doi.org/10.1038/nrm.2017.7> | Google Scholar
- Bannan CC, Calabró G, Kyu DY, Mobley DL (2016) **Calculating Partition Coefficients of Small Molecules in Octanol/Water and Cyclohexane/Water** *J Chem Theory Comput* **12**:4015–4024 <https://doi.org/10.1021/acs.jctc.6b00449> | Google Scholar
- Becke AD (1993) **A new mixing of Hartree-Fock and local density-functional theories** *J Chem Phys* **98**:1372–1377 <https://doi.org/10.1063/1.464304> | Google Scholar
- Bennett CH (1976) **Efficient estimation of free energy differences from Monte Carlo data** *J Comput Phys* **22**:245–268 [Google Scholar](#)
- Berendsen HJ, Postma Jv, Van Gunsteren WF, DiNola A, Haak JR (1984) **Molecular dynamics with coupling to an external bath** *J Chem Phys* **81**:3684–3690 [Google Scholar](#)
- Best RB, Hummer G. (2009) **Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides** *J Phys Chem B* **113**:9004–9015 <https://doi.org/10.1021/jp901540t> | Google Scholar
- Brady JP, Farber PJ, Sekhar A, Lin YH, Huang R, Bah A, Nott TJ, Chan HS, Baldwin AJ, Forman-Kay JD, et al. (2017) **Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation** *Proc Natl Acad Sci USA* **114**:E8194–E8203 [Google Scholar](#)
- Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, Jülicher F, Hyman AA (2009) **Germline P granules are liquid droplets that localize by controlled dissolution/condensation** *Science* **324**:1729–1732 [Google Scholar](#)
- Braun D, Boresch S, Steinhauser O. (2014) **Transport and dielectric properties of water and the influence of coarse-graining: Comparing BMW, SPC/E, and TIP3P models** *The Journal of chemical physics* **140** [Google Scholar](#)
- Bremer A, Farag M, Borchers WM, Peran I, Martin EW, Pappu RV, Mittag T. (2022) **Deciphering how naturally occurring sequence features impact the phase behaviours of disordered**

prion-like domains *Nat Chem* :196–207 <https://doi.org/10.1038/s41557-021-00840-w> | [Google Scholar](#)

Brown WH, Potoyan DA (2024) **Phase separation of multicomponent peptide mixtures into dehydrated clusters with hydrophilic cores** *Biophys J* **123**:349–360 [Google Scholar](#)

Bussi G, Donadio D, Parrinello M. (2007) **Canonical sampling through velocity rescaling** *J Chem Phys* **126**:014101 [Google Scholar](#)

Calinsky R, Levy Y. (2024) **Aromatic Residues in Proteins: Re-Evaluating the Geometry and Energetics of π - π , Cation- π , and CH- π Interactions** *J Phys Chem B* [Google Scholar](#)

Carter-Fenk K, Liu M, Pujal L, Loipersberger M, Tsanai M, Vernon RM, Forman-Kay JD, Head-Gordon M, Heidar-Zadeh F, Head-Gordon T. (2023) **The Energetic Origins of Pi-Pi Contacts in Proteins** *J Am Chem Soc* **145**:24836–24851 [Google Scholar](#)

Chai JD, Head-Gordon M. (2008) **Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections** *Phys Chem Chem Phys* **10**:6615–6620 [Google Scholar](#)

Chang J, Lenhoff AM, Sandler SI (2007) **Solvation free energy of amino acids and side-chain analogues** *J Phys Chem B* **111**:2098–2106 [Google Scholar](#)

Chelli R, Gervasio FL, Procacci P, Schettino V. (2002) **Stacking and T-shape competition in aromatic-aromatic amino acid interactions** *J Am Chem Soc* **124**:6133–6143 [Google Scholar](#)

Chengteh L, Yang W, Parr RG (1988) **Development of the Colic-Salvetti correlation-energy formula into a functional of the electron density** *Phys Rev B* **37**:785–789 [Google Scholar](#)

Chipot C, Jaffe R, Maigret B, Pearlman DA, Kollman PA (1996) **Benzene dimer: A good model for π - π interactions in proteins? A comparison between the benzene and the toluene dimers in the gas phase and in an aqueous solution** *J Am Chem Soc* **118**:11217–11224 <https://doi.org/10.1021/ja961379l> | [Google Scholar](#)

Choi JM, Holehouse AS, Pappu RV (2020) **Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions** *Annu Rev Biophys* **49**:107–133 <https://doi.org/10.1146/annurev-biophys-121219-081629> | [Google Scholar](#)

Das S, Lin YH, Vernon RM, Forman-Kay JD, Chan HS (2020) **Comparative roles of charge, π , and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins** *Proc Natl Acad Sci USA* **117**:28795–28805 [Google Scholar](#)

De Sancho D. (2022) **Phase separation in amino acid mixtures is governed by composition** *Biophys J* **121**:4119–4127 [Google Scholar](#)

Dignon GL, Best RB, Mittal J. (2020) **Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties** *Annu Rev Phys Chem* **71**:53–75 <https://doi.org/10.1146/annurev-physchem-071819-113553> | [Google Scholar](#)

Dignon GL, Zheng W, Kim YC, Best RB, Mittal J. (2018) **Sequence determinants of protein phase behavior from a coarse-grained model** *PLoS Comput Biol* **14**:1–23 <https://doi.org/10.1371/journal.pcbi.1005941> | [Google Scholar](#)

- Emelianova A, Garcia PL, Tan D, Joseph JA (2025) **Prediction of small-molecule partitioning into biomolecular condensates from simulation** *bioRxiv* [Google Scholar](#)
- Fisher RS, Elbaum-Garfinkle S. (2020) **Tunable multiphase dynamics of arginine and lysine liquid condensates** *Nat Commun* **11**:4628 [Google Scholar](#)
- Galvanetto N, Ivanović MT, Chowdhury A, Sottini A, Nüesch MF, Nettels D, Best RB, Schuler B. (2023) **Extreme dynamics in a biomolecular condensate** *Nature* **619**:876–883 [Google Scholar](#)
- Gapsys V, Michielssens S, Seeliger D, de Groot BL (2015) **pmx: Automated protein structure and topology generation for alchemical perturbations** *J Comput Chem* **36**:348–354 <https://doi.org/10.1002/jcc.23804> | [Google Scholar](#)
- Gapsys V, Michielssens S, Seeliger D, de Groot BL (2016) **Accurate and rigorous prediction of the changes in protein free energies in a large-scale mutation scan** *Angew Chem Int* **55**:7364–7368 [Google Scholar](#)
- Gapsys V, Pérez-Benito L, Aldeghi M, Seeliger D, Van Vlijmen H, Tresadern G, De Groot BL (2020) **Large scale relative protein ligand binding affinities using non-equilibrium alchemy** *Chem Sci* **11**:1140–1152 [Google Scholar](#)
- Gobbi A, Frenking G. (1993) **Y-conjugated compounds: the equilibrium geometries and electronic structures of guanidine, guanidinium cation, urea, and 1, 1-diaminoethylene** *J Am Chem Soc* **115**:2362–2372 [Google Scholar](#)
- Greig JA, Nguyen TA, Lee M, Holehouse AS, Posey AE, Pappu RV, Jedd G. (2020) **Arginine-enriched mixed-charge domains provide cohesion for nuclear speckle condensation** *Mol Cell* **77**:1237–1250 [Google Scholar](#)
- Hazra MK, Levy Y. (2023) **Cross-talk of cation- π interactions with electrostatic and aromatic interactions: A salt-dependent trade-off in biomolecular condensates** *J Phys Chem Lett* **14**:8460–8469 [Google Scholar](#)
- Hehre WJ, Ditchfield K, Pople JA (1972) **Self-consistent molecular orbital methods. XII. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules** *J Chem Phys* **56**:2257–2261 <https://doi.org/10.1063/1.1677527> | [Google Scholar](#)
- Hong Y, Najafi S, Casey T, Shea JE, Han SI, Hwang DS (2022) **Hydrophobicity of arginine leads to reentrant liquid-liquid phase separation behaviors of arginine-rich proteins** *Nat Commun* **13**:7326 [Google Scholar](#)
- Horn HW, Swope WC, Pitner JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. (2004) **Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew** *J Chem Phys* **120**:9665–9678 [Google Scholar](#)
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) **Comparison of Simple Potential Functions for Simulating Liquid Water** *J Chem Phys* **79**:926–935 [Google Scholar](#)
- Joseph JA, Reinhardt A, Aguirre A, Chew PY, Russell KO, Espinosa JR, Garaizar A, Collepardo-Guevara R. (2021) **Physics-driven coarse-grained model for biomolecular phase separation**

with near-quantitative accuracy *Nat Comput Sci* **1**:732–743 <https://doi.org/10.1038/s43588-021-00155-3> | [Google Scholar](#)

Kauzmann W. (1959) **Some factors in the interpretation of protein denaturation** In: Anfinsen CB, Anson ML, Bailey K, Edsall JT, editors. *Adv. Protein Chem* Elsevier pp. 1–63 [Google Scholar](#)

Kaygisiz K, Sementa D, Athiyarath V, Chen X, Ulijn RV (2025) **Context dependence in assembly code for supramolecular peptide materials and systems** *Nature Reviews Materials* :1–24 [Google Scholar](#)

Kim YC, Hummer G. (2008) **Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding** *J Mol Biol* **375**:1416–1433 [Google Scholar](#)

Kyte J, Doolittle RF (1982) **A simple method for displaying the hydropathic character of a protein** *J Mol Biol* **157**:105–132 [Google Scholar](#)

Lin Y, Currie SL, Rosen MK (2017) **Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs** *J Biol Chem* **292**:19110–19120 [Google Scholar](#)

Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE (2010) **Improved side-chain torsion potentials for the Amber ff99SB protein force field** *Proteins* **78**:1950–1958 <https://doi.org/10.1002/prot.22711> | [Google Scholar](#)

Lobo S, Najafi S, Shell MS, Shea JE (2025) **Hydrophobicity in Intrinsically Disordered Protein Force Fields: Implications for Conformational Ensembles and Protein–Protein Interactions** *J Phys Chem B* [Google Scholar](#)

Maraldo A, Rnjak-Kovacina J, Marquis C. (2024) **Tyrosine—a structural glue for hierarchical protein assembly** *Trends Biochem Sci* [Google Scholar](#)

Marenich AV, Cramer CJ, Truhlar DG (2009) **Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions** *J Phys Chem B* **113**:6378–6396 [Google Scholar](#)

Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, Grace CR, Soranno A, Pappu RV, Mittag T. (2020) **Valence and patterning of aromatic residues determine the phase behavior of prion-like domains** *Science* **367**:694–699 <https://doi.org/10.1126/science.aaw8653> | [Google Scholar](#)

Martin EW, Mittag T. (2018) **Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions** *Biochemistry* **57**:2478–2487 <https://doi.org/10.1021/acs.biochem.8b00008> | [Google Scholar](#)

Martinez-Martin I, Crousilles A, Ochoa JP, Velazquez-Carreras D, Mortensen SA, Herrero-Galan E, Delgado J, Dominguez F, Garcia-Pavia P, de Sancho D, et al. (2023) **Titin domains with reduced core hydrophobicity cause dilated cardiomyopathy** *Cell Rep* **42** [Google Scholar](#)

McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, Schwantes CR, Wang LP, Lane TJ, Pande VS (2015) **MDTraj: A Modern Open Library for the Analysis of**

Molecular Dynamics Trajectories *Biophys J* **109**:1528–1532 <https://doi.org/10.1016/j.bpj.2015.08.015> | [Google Scholar](#)

Mennucci B. (2012) **Polarizable continuum model** *Wiley Interdiscip Rev Comput Mol Sci* **2**:386–404 [Google Scholar](#)

Mey ASJS, Allen BK, Bruce McDonald HE, Chodera JD, Hahn DF, Kuhn M, Michel J, Mobley DL, Naden LN, Prasad S, Rizzi A, Scheen J, Shirts MR, Tresadern G, Xu H. (2020) **Best Practices for Alchemical Free Energy Calculations [Article v1.0]** *Living J Comp Mol Sci* **2**:18378 <https://doi.org/10.33011/livecoms.2.1.18378> | [Google Scholar](#)

Miller MS, Lay WK, Elcock AH (2016) **Osmotic pressure simulations of amino acids and peptides highlight potential routes to protein force field parameterization** *J Phys Chem B* **120**:8217–8229 [Google Scholar](#)

Mittag T, Pappu RV (2022) **A conceptual framework for understanding phase separation and addressing open questions and challenges** *Mol Cell* **82**:2201–2214 <https://doi.org/10.1016/j.molcel.2022.05.018> | [Google Scholar](#)

Miyazawa S, Jernigan RL (1996) **Residue–residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading** *J Mol Biol* **256**:623–644 [Google Scholar](#)

Murthy AC, Tang WS, Jovic N, Janke AM, Seo DH, Perdikari TM, Mittal J, Fawzi NL (2021) **Molecular interactions contributing to FUS SYGQ LC-RGG phase separation and co-partitioning with RNA polymerase II heptads** *Nat Struct Mol Biol* **28**:923–935 [Google Scholar](#)

Nerenberg PS, Head-Gordon T. (2011) **Optimizing protein-solvent force fields to reproduce intrinsic conformational preferences of model peptides** *J Chem Theory Comput* **7**:1220–1230 [Google Scholar](#)

Neumann M. (1983) **Dipole moment fluctuation formulas in computer simulations of polar systems** *Mol Phys* **50**:841–858 [Google Scholar](#)

Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowietz A, Craggs TD, Bazett-Jones DP, Pawson T, Forman-Kay JD, Baldwin AJ (2015) **Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles** *Mol Cell* **57**:936–947 [Google Scholar](#)

Nozaki Y, Tanford C. (1971) **The Solubility of Amino Acids and Two Glycine Peptides in Aqueous Ethanol and Dioxane Solutions: establishment of a hydrophobicity scale** *J Biol Chem* **246**:2211–2217 [https://doi.org/10.1016/S0021-9258\(19\)77210-X](https://doi.org/10.1016/S0021-9258(19)77210-X) | [Google Scholar](#)

Pace CN, Fu H, Fryar KL, Landua J, Trevino SR, Shirley BA, Hendricks MM, Iimura S, Gajiwala K, Scholtz JM, Grimsley GR (2011) **Contribution of Hydrophobic Interactions to Protein Stability** *J Mol Biol* **408**:514–528 <https://doi.org/10.1016/j.jmb.2011.02.053> | [Google Scholar](#)

Pace CN, Horn G, Hebert EJ, Bechert J, Shaw K, Urbanikova L, Scholtz JM, Sevcik J. (2001) **Tyrosine hydrogen bonds make a large contribution to protein stability** *J Mol Biol* **312**:393–404 [Google Scholar](#)

Paloni M, Bailly R, Ciandrini L, Barducci A. (2020) **Unraveling Molecular Interactions in Liquid–Liquid Phase Separation of Disordered Proteins by Atomistic Simulations** *J Phys Chem B* **124**:9009–9016 <https://doi.org/10.1021/acs.jpcc.0c06288> | [Google Scholar](#)

- Paloni M, Bussi G, Barducci A. (2021) **Arginine multivalency stabilizes protein/RNA condensates** *Protein Sci* **30**:1418–1426 [Google Scholar](#)
- Parrinello M, Rahman A. (1980) **Crystal structure and pair potentials: A molecular-dynamics study** *Phys Rev Lett* **45**:1196 [Google Scholar](#)
- Piana S, Donchev AG, Robustelli P, Shaw DE (2015) **Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States** *J Phys Chem B* **119**:5113–5123 <https://doi.org/10.1021/jp508971m> | [Google Scholar](#)
- Pitera JW, Falta M, van Gunsteren WF (2001) **Dielectric properties of proteins from simulation: the effects of solvent, ligands, pH, and temperature** *Biophys J* **80**:2546–2555 [Google Scholar](#)
- Plaxco KW, Morton CJ, Grimshaw SB, Jones JA, Pitkeathly M, Campbell ID, Dobson CM (1997) **The effects of guanidine hydrochloride on the 'random coil' conformations and NMR chemical shifts of the peptide series GGXGG** *J Biomol NMR* **10**:221–230 [Google Scholar](#)
- Pokarowski P, Kloczkowski A, Jernigan RL, Kothari NS, Pokarowska M, Kolinski A. (2005) **Inferring ideal amino acid interaction forms from statistical protein contact potentials** *Proteins* **59**:49–57 [Google Scholar](#)
- Radzicka A, Wolfenden R. (1988) **Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution** *Biochemistry* **27**:1664–1670 [Google Scholar](#)
- Regy RM, Thompson J, Kim YC, Mittal J. (2021) **Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins** *Protein Sci* **30**:1371–1379 <https://doi.org/10.1002/pro.4094> | [Google Scholar](#)
- Rekhi S, Garcia CG, Barai M, Rizuan A, Schuster BS, Kiick KL, Mittal J. (2024) **Expanding the molecular language of protein liquid–liquid phase separation** *Nat Chem* :1–12 [Google Scholar](#)
- Rekhi S, Mittal J. (2025) **Amino acid transfer free energies reveal thermodynamic driving forces in biomolecular condensate formation** *Proc Natl Acad Sci USA* **122**:e2425422122 <https://doi.org/10.1073/pnas.2425422122> | [Google Scholar](#)
- Rogers DM, Hirst JD, Lee EPF, Wright TG (2006) **Ab initio study of the toluene dimer** *Chem Phys Lett* **427**:410–413 <https://doi.org/10.1016/j.cplett.2006.07.022> | [Google Scholar](#)
- Salomon-Ferrer R, Case DA, Walker RC (2013) **An overview of the Amber biomolecular simulation package** *Wiley Interdiscip Rev Comput Mol Sci* **3**:198–210 [Google Scholar](#)
- Schuster BS, Dignon GL, Tang WS, Kelley FM, Ranganath AK, Jahnke CN, Simpkins AG, Regy RM, Hammer DA, Good MC, Mittal J. (2020) **Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior** *Proc Natl Acad Sci USA* **117**:11421–11431 <https://doi.org/10.1073/pnas.2000223117> | [Google Scholar](#)
- Seeliger D, de Groot BL (2010) **Protein Thermostability Calculations Using Alchemical Free Energy Simulations** *Bio-phys J* **98**:2309–2316 <https://doi.org/10.1016/j.bpj.2010.01.051> | [Google Scholar](#)

- Semenov AN, Rubinstein M. (1998) **Thermoreversible gelation in solutions of associative polymers. 1. Statics** *Macro-molecules* **31**:1373–1385 [Google Scholar](#)
- Serrano L, Bycroft M, Fersht AR (1991) **Aromatic-aromatic interactions and protein stability: investigation by doublemutant cycles** *J Mol Biol* **218**:465–475 [Google Scholar](#)
- Shimizu S, Chan HS (2001) **Statistical mechanics of solvophobic aggregation: Additive and cooperative effects** *J Chem Phys* **115**:3424–3431 [Google Scholar](#)
- Shimizu S, Chan HS (2002) **Origins of protein denatured state compactness and hydrophobic clustering in aqueous urea: inferences from nonpolar potentials of mean force** *Proteins: Structure, Function, and Bioinformatics* **49**:560–566 [Google Scholar](#)
- Shin Y, Brangwynne CP (2017) **Liquid phase condensation in cell physiology and disease** *Science* **357**:eaaf4382 <https://doi.org/10.1126/science.aaf4382> | [Google Scholar](#)
- Sugita Y, Okamoto Y. (1999) **Replica-exchange molecular dynamics method for protein folding** *Chem Phys Lett* **314**:141–151 [Google Scholar](#)
- Tang Y, Bera S, Yao Y, Zeng J, Lao Z, Dong X, Gazit E, Wei G. (2021) **Prediction and characterization of liquid-liquid phase separation of minimalistic peptides** *Cell Rep Phys Sci* **2**:100579 <https://doi.org/10.1016/j.xcrp.2021.100579> | [Google Scholar](#)
- Tesei G, Schulze TK, Crehuet R, Lindorff-Larsen K. (2021) **Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties** *Proc Natl Acad Sci U S A* **118**:e2111696118 <https://doi.org/10.1073/pnas.2111696118> | [Google Scholar](#)
- Thomas PD, Dill KA (1996) **An iterative method for extracting energy-like quantities from protein structures** *Proc Natl Acad Sci USA* **93**:11628–11633 [Google Scholar](#)
- Tsuzuki S, Honda K, Uchimaru T, Mikami M. (2005) **Ab initio calculations of structures and interaction energies of toluene dimers including CCSD(T) level electron correlation correction** *J Chem Phys* **122** <https://doi.org/10.1063/1.1876092> | [Google Scholar](#)
- Urry DW, Gowda DC, Parker TM, Luan CH, Reid MC, Harris CM, Pattanaik A, Harris RD (1992) **Hydrophobicity scale for proteins based on inverse temperature transitions** *Biopolymers* **32**:1243–1250 [Google Scholar](#)
- Vernon RM, Chong PA, Tsang B, Kim TH, Bah A, Farber P, Lin H, Forman-Kay JD (2018) **Pi-Pi contacts are an overlooked protein feature relevant to phase separation** *eLife* **7**:e31486 <https://doi.org/10.7554/eLife.31486> | [Google Scholar](#)
- Wang J, Choi JM, Holehouse AS, Lee HO, Zhang X, Jahnel M, Maharana S, Lemaitre R, Pozniakovsky A, Drechsel D, Poser I, Pappu RV, Alberti S, Hyman AA (2018) **A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins** *Cell* **174**:688–699 [Google Scholar](#)
- Wimley WC, Creamer TP, White SH (1996) **Solvation energies of amino acid side chains and backbone in a family of hostguest pentapeptides** *Biochemistry* **35**:5109–5124 [Google Scholar](#)

Wolfenden R, Andersson L, Cullis P, Southgate C. (1981) **Affinities of amino acid side chains for solvent water** *Bio-chemistry* **20**:849–855 [Google Scholar](#)

Wood RH, Thompson PT (1990) **Differences between pair and bulk hydrophobic interactions** *Proc Natl Acad Sci U S A* **87**:946–949 [Google Scholar](#)

Workman RJ, Huang CJ, Lynch GC, Pettitt BM (2024) **Peptide diffusion in biomolecular condensates** *Biophys J* **123**:1668–1675 [Google Scholar](#)

Workman RJ, Pettitt BM (2021) **Thermodynamic Compensation in Peptides Following Liquid–Liquid Phase Separation** *J Phys Chem B* **125**:6431–6439 [Google Scholar](#)

Zheng W, Dignon GL, Jovic N, Xu X, Regy RM, Fawzi NL, Kim YC, Best RB, Mittal J. (2020) **Molecular Details of Protein Condensates Probed by Microsecond Long Atomistic Simulations** *J Phys Chem B* **124**:11671–11679 <https://doi.org/10.1021/acs.jpcb.0c10489> | [Google Scholar](#)

Author information

David De Sancho

Polimero eta Material Aurreratuak: Fisika, Kimika eta Teknologia, Kimika Fakultatea, UPV/EHU & Donostia International Physics Center (DIPC), Donostia / San Sebastian, Spain
ORCID iD: [0000-0002-8985-2685](https://orcid.org/0000-0002-8985-2685)

For correspondence: david.desancho@ehu.eus

Xabier López

Polimero eta Material Aurreratuak: Fisika, Kimika eta Teknologia, Kimika Fakultatea, UPV/EHU & Donostia International Physics Center (DIPC), Donostia / San Sebastian, Spain
ORCID iD: [0000-0002-2711-3588](https://orcid.org/0000-0002-2711-3588)

For correspondence: xabier.lopez@ehu.eus

Editors

Reviewing Editor

Rosana Colleparado

University of Cambridge, Cambridge, United Kingdom

Senior Editor

Qiang Cui

Boston University, Boston, United States of America

Reviewer #1 (Public review):

This is an interesting and timely computational study using molecular dynamics simulation as well as quantum mechanical calculation to address why tyrosine (Y), as part of an intrinsically disordered protein (IDP) sequence, has been observed experimentally to be stronger than phenylalanine (F) as a promoter for biomolecular phase separation. Notably, the authors identified the aqueous nature of the condensate environment and the

corresponding dielectric and hydrogen bonding effects as a key to understand the experimentally observed difference. This principle is illustrated by the difference in computed transfer free energy of Y- and F-containing pentapeptides into solvent with various degrees of polarity. The elucidation offered by this work is important. The computation appears to be carefully executed, the results are valuable, and the discussion is generally insightful. However, there is room for improvement in some parts of the presentation in terms of accuracy and clarity, including, e.g., the logic of the narrative should be clarified with additional information (and possibly additional computation), and the current effort should be better placed in the context of prior relevant theoretical and experimental works on cation- π interactions in biomolecules and dielectric properties of biomolecular condensates. Accordingly, this manuscript should be revised to address the following, with added discussion as well as inclusion of references mentioned below.

(1) Page 2, line 61: "Coarse-grained simulation models have failed to account for the greater propensity of arginine to promote phase separation in Ddx4 variants with Arg to Lys mutations (Das et al., 2020)". As it stands, this statement is not accurate, because the cited reference to Das et al. showed that although some coarse-grained model, namely the HPS model of Dignon et al., 2018 PLoS Comput did not capture the Arg to Lys trend, the KH model described in the same Dignon et al. paper was demonstrated by Das et al. (2020) to be capable of mimicking the greater propensity of Arg to promote phase separation than Lys. Accordingly, a possible minimal change that would correct the inaccuracy of this statement in the manuscript would be to add the word "Some" in front of "coarse-grained simulation models ...", i.e., it should read "Some coarse-grained simulation models have failed ...". In fact, a subsequent work [Wessén et al., J Phys Chem B 126: 9222-9245 (2022)] that applied the Mpipi interaction parameters (Joseph et al., 2021, already cited in the manuscript) showed that Mpipi is capable of capturing the rank ordering of phase separation propensity of Ddx4 variants, including a charge scrambled variant as well as both the Arg to Lys and the Phe to Ala variants (see Fig.11a of the above-cited Wessén et al. 2022 reference). The authors may wish to qualify their statements in the introduction to take note of these prior results. For example, they may consider adding a note immediately after the next sentence in the manuscript "However, by replacing the hydrophobicity scales ... (Das et al., 2020)" to refer to these subsequent findings in 2021-2022.

(2) Page 8, lines 285-290 (as well as the preceding discussion under the same subheading & Fig.4): "These findings suggest that ... is not primarily driven by differences in protein-protein interaction patterns ..." The authors' logic in terms of physical explanation is somewhat problematic here. In this regard, "Protein-protein interaction patterns" appears to be a straw man, so to speak. Indeed, who (reference?) has argued that the difference in the capability of Y and F in promoting phase separation should be reflected in the pairwise amino acid interaction pattern in a condensate that contains either only Y (and G, S) and only F (and G, S) but not both Y and F? Also, this paragraph in the manuscript seems to suggest that the authors' observation of similar contact patterns in the GSY and GSF condensates is "counterintuitive" given the difference in Y-Y and F-F potentials of mean force (Joseph et al., 2021); but there is nothing particularly counterintuitive about that. The two sets of observations are not mutually exclusive. For instance, consider two different homopolymers, one with a significantly stronger monomer-monomer attraction than the other. The condensates for the two different homopolymers will have essentially the same contact pattern but very different stabilities (different critical temperatures), and there is nothing surprising about it. In other words, phase separation propensity is not "driven" by contact pattern in general, it's driven by interaction (free) energy. The relevant issue here is total interaction energy or critical point of the phase separation. If it is computationally feasible, the authors should attempt to determine the critical temperatures for the GSY condensate versus the GSF condensate to verify that the GSY condensate has a higher critical temperature than the GSF condensate. That would be the most relevant piece of information for the question at hand.

(3) Page 9, lines 315-316: "...Our ϵ [relative permittivity] values ... are surprisingly close to that derived from experiment on Ddx4 condensates (45 ± 13) (Nott et al., 2015)". For accuracy, it should be noted here that the relative permittivity provided in the supplementary information of Nott et al. was not a direct experimental measurement but based on a fit using Flory-Huggins (FH), but FH is not the most appropriate theory for polymer with long-spatial-range Coulomb interactions. To this reviewer's knowledge, no direct measurement of relative permittivity in biomolecular condensates has been made to date. Explicit-water simulation suggests that relative permittivity of Ddx4 condensate with protein volume fraction ≈ 0.4 can have relative permittivity ≈ 35 -50 (Das et al., PNAS 2020, Fig.7A), which happens to agree with the $\epsilon = 45 \pm 13$ estimate. This information should be useful to include in the authors' manuscript.

(4) As for the dielectric environment within biomolecular condensates, coarse-grained simulation has suggested that whereas condensates formed by essentially electric neutral polymers (as in the authors' model systems) have relative permittivities intermediate between that of bulk water and that of pure protein ($\epsilon = 2$ -4, or at most 15), condensates formed by highly charge polymers can have relative permittivity higher than that of bulk water [Wessén et al., J Phys Chem B 125:4337-4358 (2021), Fig.14 of this reference]. In view of the role of aromatic residues (mainly Y and F) in the phase separation of IDPs such as A1-LCD and LAF-1 that contain positively and negatively charged residues (Martin et al., 2020; Schuster et al., 2020, already cited in the manuscript), it should be useful to address briefly how the relationship between the relative phase-separation promotion strength of Y vs F and dielectric environment of the condensate may or may not be change with higher relative permittivities.

(5) The authors applied the dipole moment fluctuation formula (Eq.2 in the manuscript) to calculate relative permittivity in their model condensates. Does this formula apply only to an isotropic environment? The authors' model condensates were obtained from a "slab" approach (p.4) and thus the simulation box has a rectangular geometry. Did the authors apply their Eq.2 to the entire simulation box or only to the central part of the box with the condensate (see, e.g., Fig.3C in the manuscript). If the latter is the case, is it necessary to use a different dipole moment formula that distinguishes between the "parallel" and "perpendicular" components of the dipole moment (see, e.g., Eq.16 in the above-cited Wessén et al. 2021 paper). A brief added comments will be useful.

(6) With regard to the general role of Y and F in the phase separation of biomolecules containing positively charged Arg and Lys residues, the relative strength of cation- π interactions (cation-Y vs cation-F) should be addressed (in view of the generality implied by the title of the manuscript), or at least discussed briefly in the authors' manuscript if a detailed study is beyond the scope of their current effort. It has long been known that in the biomolecular context, cation-Y is slightly stronger than cation-F, whereas cation-tryptophan (W) is significantly stronger than either cation-Y and cation-F [Wu & McMahon, JACS 130:12554-12555 (2008)]. Experimental data from a study of EWS (Ewing sarcoma) transactivation domains indicated that Y is a slightly stronger promoter than F for transcription, whereas W is significantly stronger than either Y or F [Song et al., PLoS Comput Biol 9:e1003239 (2013)]. In view of the subsequent general recognition that "transcription factors activate genes through the phase-separation capacity of their activation domain" [Boija et al., Cell 175:1842-1855.e16 (2018)] which is applicable to EWS in particular [Johnson et al., JACS 146:8071-8085 (2024)], the experimental data in Song et al. 2013 (see Fig.3A of this reference) suggests that cation-Y interactions are stronger than cation-F interactions in promoting phase separation, thus generalizing the authors' observations (which focus primarily on Y-Y, Y-F and F-F interactions) to most situations in which cation-Y and cation-F interactions are relevant to biomolecular condensation.

(7) Page 9: The observation of a weaker effective F-F (and a few other nonpolar-nonpolar) interaction in a largely aqueous environment (as in an IDP condensate) than in a nonpolar environment (as in the core of a folded protein) is intimately related to (and expected from) the long-recognized distinction between "bulk" and "pair" as well as size dependence of hydrophobic effects that have been addressed in the context of protein folding [Wood & Thompson, PNAS 87:8921-8927 (1990); Shimizu & Chan, JACS 123:2083-2084 (2001); Proteins 49:560-566 (2002)]. It will be useful to add a brief pointer in the current manuscript to this body of relevant resource in protein science.

Comments on revisions:

The authors have largely addressed my previous concerns and the manuscript has been substantially improved. Nonetheless, it will benefit the readers more if the authors had included more of the relevant references provided in my previous review so as to afford a broader and more accurate context to the authors' effort. This deficiency is particularly pertinent for point number 6 in my previous report about cation- π interactions. The authors have now added a brief discussion but with no references on the rank ordering of Y, F, and W interactions. I cannot see how providing additional information about a few related works could hurt. Quite the contrary, having the references will help readers establish scientific connections and contribute to conceptual advance.

<https://doi.org/10.7554/eLife.104950.2.sa2>

Reviewer #2 (Public review):

Summary:

In this preprint, De Sancho and López use alchemical molecular dynamics simulations and quantum mechanical calculations to elucidate the origin of the observed preference of Tyr over Phe in phase separation. The paper is well written, and the simulations conducted are rigorous and provide good insight into the origin of the differences between the two aromatic amino acids considered.

Strengths:

The study addresses a fundamental discrepancy in the field of phase separation where the predicted ranking of aromatic amino acids observed experimentally is different from their anticipated rankings when considering contact statistics of folded proteins. While the hypothesis that the difference in the microenvironment of the condensed phase and hydrophobic core of folded proteins underlies the different observations, this study provides a quantification of this effect. Further, the demonstration of the crossover between Phe and Tyr as a function of the dielectric is interesting and provides further support for the hypothesis that the differing microenvironments within the condensed phase and the core of folded proteins is the origin of the difference between contact statistics and experimental observations in phase separation literature. The simulations performed in this work systematically investigate several possible explanations and therefore provide depth to the paper.

<https://doi.org/10.7554/eLife.104950.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Reviewer #1 (Public review):

This is an interesting and timely computational study using molecular dynamics simulation as well as quantum mechanical calculation to address why tyrosine (Y), as part of an intrinsically disordered protein (IDP) sequence, has been observed experimentally to be stronger than phenylalanine (F) as a promoter for biomolecular phase separation. Notably, the authors identified the aqueous nature of the condensate environment and the corresponding dielectric and hydrogen bonding effects as a key to understanding the experimentally observed difference. This principle is illustrated by the difference in computed transfer free energy of Y- and F-containing pentapeptides into a solvent with various degrees of polarity. The elucidation offered by this work is important. The computation appears to be carefully executed, the results are valuable, and the discussion is generally insightful. However, there is room for improvement in some parts of the presentation in terms of accuracy and clarity, including, e.g., the logic of the narrative should be clarified with additional information (and possibly additional computation), and the current effort should be better placed in the context of prior relevant theoretical and experimental works on cation- π interactions in biomolecules and dielectric properties of biomolecular condensates. Accordingly, this manuscript should be revised to address the following, with added discussion as well as inclusion of references mentioned below.

We are grateful for the referee's assessment of our work and insightful suggestions, which we address point by point below.

(1) Page 2, line 61: "Coarse-grained simulation models have failed to account for the greater propensity of arginine to promote phase separation in Ddx4 variants with Arg to Lys mutations (Das et al., 2020)". As it stands, this statement is not accurate, because the cited reference to Das et al. showed that although some coarse-grained models, namely the HPS model of Dignon et al., 2018 PLoS Comput did not capture the Arg to Lys trend, the KH model described in the same Dignon et al. paper was demonstrated by Das et al. (2020) to be capable of mimicking the greater propensity of Arg to promote phase separation than Lys. Accordingly, a possible minimal change that would correct the inaccuracy of this statement in the manuscript would be to add the word "Some" in front of "coarse-grained simulation models ...", i.e., it should read "Some coarse-grained simulation models have failed ...". In fact, a subsequent work [Wessén et al., J Phys Chem B 126: 9222-9245 (2022)] that applied the Mpipi interaction parameters (Joseph et al., 2021, already cited in the manuscript) showed that Mpipi is capable of capturing the rank ordering of phase separation propensity of Ddx4 variants, including a charge scrambled variant as well as both the Arg to Lys and the Phe to Ala variants (see Figure 11a of the above-cited Wessén et al. 2022 reference). The authors may wish to qualify their statements in the introduction to take note of these prior results. For example, they may consider adding a note immediately after the next sentence in the manuscript "However, by replacing the hydrophobicity scales ... (Das et al., 2020)" to refer to these subsequent findings in 2021-2022.

We agree with the referee that the wording used in the original version was inaccurate. We did not want to expand too much on the previous results on Lys/Arg, to avoid overwhelming our readers with background information that was not directly relevant to the aromatic residues Phe and Tyr. We have now introduced some of the missing details in the hope that this will provide a more accurate account of what has been achieved with different versions of coarse-grained models. In the revised version, we say the following:

Das and co-workers attempted to explain arginine's greater propensity to phase separate in Ddx4 variants using coarse-grained simulations with two different energy functions (Das et

al., 2020). The model was first parametrized using a hydrophobicity scale, aimed to capture the “stickiness” of different amino acids (Dignon et al., 2018), but this did not recapitulate the correct rank order in the stability of the simulated condensates (Das et al., 2020). By replacing the hydrophobicity scale with interaction energies from amino acid contact matrices — derived from a statistical analysis of the PDB (Dignon et al., 2018; Miyazawa and Jernigan, 1996; Kim and Hummer, 2008)— they recovered the correct trends (Das et al., 2020). A key to the greater propensity for LLPS in the case of Arg may derive from the pseudo-aromaticity of this residue, which results in a greater stabilization relative to the more purely cationic character of Lys (Gobbi and Frenking, 1993; Wang et al., 2018; Hong et al., 2022).

(2) Page 8, lines 285-290 (as well as the preceding discussion under the same subheading & Figure 4): "These findings suggest that ... is not primarily driven by differences in protein-protein interaction patterns ..." The authors' logic in terms of physical explanation is somewhat problematic here. In this regard, "Protein-protein interaction patterns" appear to be a straw man, so to speak. Indeed, who (reference?) has argued that the difference in the capability of Y and F in promoting phase separation should be reflected in the pairwise amino acid interaction pattern in a condensate that contains either only Y (and G, S) and only F (and G, S) but not both Y and F? Also, this paragraph in the manuscript seems to suggest that the authors' observation of similar contact patterns in the GSY and GSF condensates is "counterintuitive" given the difference in Y-Y and F-F potentials of mean force (Joseph et al., 2021); but there is nothing particularly counterintuitive about that. The two sets of observations are not mutually exclusive. For instance, consider two different homopolymers, one with a significantly stronger monomer-monomer attraction than the other. The condensates for the two different homopolymers will have essentially the same contact pattern but very different stabilities (different critical temperatures), and there is nothing surprising about it. In other words, phase separation propensity is not "driven" by contact pattern in general, it's driven by interaction (free) energy. The relevant issue here is total interaction energy or the critical point of the phase separation. If it is computationally feasible, the authors should attempt to determine the critical temperatures for the GSY condensate versus the GSF condensate to verify that the GSY condensate has a higher critical temperature than the GSF condensate. That would be the most relevant piece of information for the question at hand.

We are grateful for this very insightful comment by the referee. We have followed this suggestion to address whether, despite similar interaction patterns in GSY and GSF condensates, their stabilities are different. As in our previous work (De Sancho, 2022), we have run replica exchange MD simulations for both condensates and derived their phase diagrams. Our results, shown in the new Figure 5 and supplementary Figs. S6-S7, clearly indicate that the GSY condensate has a lower saturation density than the GSF condensate. This result is consistent with the trends observed in experiments on mutants of the low-complexity domain of hnRNPA1, where the relative amounts of F and Y determine the saturation concentration (Bremer et al., 2022).

(3) Page 9, lines 315-316: "...Our ϵ [relative permittivity] values ... are surprisingly close to that derived from experiment on Ddx4 condensates (45{plus minus}13) (Nott et al., 2015)". For accuracy, it should be noted here that the relative permittivity provided in the supplementary information of Nott et al. was not a direct experimental measurement but based on a fit using Flory-Huggins (FH), but FH is not the most appropriate theory for a polymer with long-spatial-range Coulomb interactions. To this reviewer's knowledge, no direct measurement of relative permittivity in biomolecular condensates has been made to date. Explicit-water simulation suggests that the relative permittivity of Ddx4 condensate with protein volume fraction ≈ 0.4 can have a relative permittivity ≈ 35 -50

(Das et al., PNAS 2020, Fig.7A), which happens to agree with the $\epsilon = 45$ {plus minus}13 estimate. This information should be useful to include in the authors' manuscript.

We thank the referee for this useful comment. We are aware that the estimate we mentioned is not direct. We have now clarified this point and added the additional estimate from Das et al. In the new version of the manuscript, we say:

Our ϵ values for the condensates (39 ± 5 for GSY and 47 ± 3 for GSF) are surprisingly close to that derived from experiments on Ddx condensates using Flory-Huggins theory (45 ± 13) (Nott et al., 2015) and from atomistic simulations of Ddx4 (~ 35 – 50 at a volume fraction of $\phi = 0.4$) (Das et al., 2020).

(4) As for the dielectric environment within biomolecular condensates, coarse-grained simulation has suggested that whereas condensates formed by essentially electric neutral polymers (as in the authors' model systems) have relative permittivities intermediate between that of bulk water and that of pure protein ($\epsilon=2$ -4, or at most 15), condensates formed by highly charged polymers can have relative permittivity higher than that of bulk water [Wessén et al., J Phys Chem B 125:4337-4358 (2021), Fig.14 of this reference]. In view of the role of aromatic residues (mainly Y and F) in the phase separation of IDPs such as A1-LCD and LAF-1 that contain positively and negatively charged residues (Martin et al., 2020; Schuster et al., 2020, already cited in the manuscript), it should be useful to address briefly how the relationship between the relative phase-separation promotion strength of Y vs F and dielectric environment of the condensate may or may not be change with higher relative permittivities.

We thank the referee for their comment regarding highly charged polymers. However, we have chosen not to address these systems in our manuscript, as they are significantly different from the GSY/GSF peptide condensates under investigation. In polyelectrolyte systems, condensate formation is primarily driven by electrostatic interactions and counterion release, while we highlight the role of transfer free energies. At high dielectric constants (and dielectrics even higher than that of water), the strength of electrostatic interactions will be greatly reduced. In our approach to estimate differences between Y and F, the transfer free energy should plateau at a value of $\Delta\Delta G=0$ in water. At greater values of $\epsilon > 80$, it becomes difficult to predict whether additional effects might become relevant. As this lies beyond the scope of our current study, we prefer not to speculate further.

(5) The authors applied the dipole moment fluctuation formula (Eq.2 in the manuscript) to calculate relative permittivity in their model condensates. Does this formula apply only to an isotropic environment? The authors' model condensates were obtained from a "slab" approach (page 4 and thus the simulation box has a rectangular geometry. Did the authors apply Equation 2 to the entire simulation box or only to the central part of the box with the condensate (see, e.g., Figure 3C in the manuscript). If the latter is the case, is it necessary to use a different dipole moment formula that distinguishes between the "parallel" and "perpendicular" components of the dipole moment (see, e.g., Equation 16 in the above-cited Wessén et al. 2021 paper). A brief added comment will be useful.

We have calculated the relative permittivity from dense phases only. These dense phases were sliced from the slab geometry and then re-equilibrated. Long simulations were then run to converge the calculation of the dielectric constant. We have clarified this in the Methods section of the paper. We say:

For the calculation of the dielectric constant of condensates, we used the simulations of isolated dense phases mentioned above.

(6) Concerning the general role of Y and F in the phase separation of biomolecules containing positively charged Arg and Lys residues, the relative strength of cation- π interactions (cation-Y vs cation-F) should be addressed (in view of the generality implied by the title of the manuscript), or at least discussed briefly in the authors' manuscript if a detailed study is beyond the scope of their current effort. It has long been known that in the biomolecular context, cation-Y is slightly stronger than cation-F, whereas cation-tryptophan (W) is significantly stronger than either cation-Y and cation-F [Wu & McMahon, JACS 130:12554-12555 (2008)]. Experimental data from a study of EWS (Ewing sarcoma) transactivation domains indicated that Y is a slightly stronger promoter than F for transcription, whereas W is significantly stronger than either Y or F [Song et al., PLoS Comput Biol 9:e1003239 (2013)]. In view of the subsequent general recognition that "transcription factors activate genes through the phase-separation capacity of their activation domain" [Boija et al., Cell 175:1842-1855.e16 (2018)] which is applicable to EWS in particular [Johnson et al., JACS 146:8071-8085 (2024)], the experimental data in Song et al. 2013 (see Figure 3A of this reference) suggests that cation-Y interactions are stronger than cation-F interactions in promoting phase separation, thus generalizing the authors' observations (which focus primarily on Y-Y, Y-F and F-F interactions) to most situations in which cation-Y and cation-F interactions are relevant to biomolecular condensation.

We thank our referee for this insightful comment. While we restrict our analysis to aromatic pairs in this work, the observed crossover will certainly affect other pairs where tyrosine or phenylalanine are involved. We now comment on this point in the discussions section of the revised manuscript. This topic will be explored in detail in a follow-up manuscript we are currently completing. We say:

We note that, although we have not included in our analysis positively charged residues that form cation- π interactions with aromatics, the observed crossover will also be relevant to Arg/Lys contacts with Phe and Tyr. Following the rationale of our findings, within condensates, cation-Tyr interactions are expected to promote phase separation more strongly than cation-Phe pairs.

(7) Page 9: The observation of weaker effective F-F (and a few other nonpolar-nonpolar) interactions in a largely aqueous environment (as in an IDP condensate) than in a nonpolar environment (as in the core of a folded protein) is intimately related to (and expected from) the long-recognized distinction between "bulk" and "pair" as well as size dependence of hydrophobic effects that have been addressed in the context of protein folding [Wood & Thompson, PNAS 87:8921-8927 (1990); Shimizu & Chan, JACS 123:2083-2084 (2001); Proteins 49:560-566 (2002)]. It will be useful to add a brief pointer in the current manuscript to this body of relevant resources in protein science.

We thank the referee for bringing this body of work to our attention. In the revised version of our work, we briefly mention how it relates to our results. We also note that the suggested references have pointed to another of the limitations of our study, that of chain connectivity, addressed in the work by Shimizu and Chan. While we were well aware of these limitations, we had not mentioned them in our manuscript. Concerning the distinction between pair and bulk hydrophobicities, we include the following in the concluding lines of our work:

The observed context dependence has deep roots in the concepts of "pair" and "bulk" hydrophobicity (Wood and Thompson, 1990; Shimizu and Chan, 2002). While pair hydrophobicity is connected to dimerisation equilibria (i.e. the second step in Figure 2B), bulk hydrophobicity is related to transfer processes (the first step). Our work stresses the importance of considering both the pair contribution that dominates at high solvation, and

the transfer free energy contribution, which overwhelms the interaction strength at low dielectrics.

Reviewer #2 (Public review):

Summary:

In this preprint, De Sancho and López use alchemical molecular dynamics simulations and quantum mechanical calculations to elucidate the origin of the observed preference of Tyr over Phe in phase separation. The paper is well written, and the simulations conducted are rigorous and provide good insight into the origin of the differences between the two aromatic amino acids considered.

We thank the referee for his/her positive assessment of our work. Below, we address all the questions raised one by one.

Strengths:

The study addresses a fundamental discrepancy in the field of phase separation where the predicted ranking of aromatic amino acids observed experimentally is different from their anticipated rankings when considering contact statistics of folded proteins. While the hypothesis that the difference in the microenvironment of the condensed phase and hydrophobic core of folded proteins underlies the different observations, this study provides a quantification of this effect. Further, the demonstration of the crossover between Phe and Tyr as a function of the dielectric is interesting and provides further support for the hypothesis that the differing microenvironments within the condensed phase and the core of folded proteins is the origin of the difference between contact statistics and experimental observations in phase separation literature. The simulations performed in this work systematically investigate several possible explanations and therefore provide depth to the paper.

Weaknesses:

While the study is quite comprehensive and the paper well written, there are a few instances that would benefit from additional details. In the methods section, it is unclear as to whether the GGXGG peptides upon which the alchemical transforms are conducted are positioned restrained within the condensed/dilute phase or not. If they are not, how would the position of the peptides within the condensate alter the calculated free energies reported?

The peptides are not restrained in our simulations and can therefore diffuse out of the condensate given sufficient time. Although the GGXGG peptide can, given sufficient time, leave the peptide condensate, we did not observe any escape event in the trajectories we used to generate starting points for switching. Hence, the peptide environment captured in our calculations reflects, on average, the protein-protein and protein-solvent interactions inside the model condensate. We believe this is the right way of performing the calculation of transfer free energy differences into the condensate. We have clarified this point when we describe the equilibrium simulation results in the revised manuscript. We say:

Also, the peptide that experiences the transformation, which is not restrained, must remain buried within the condensate for all the snapshots that we use as initial frames, to avoid averaging the work in the dilute and dense phases.

On the referee's second point of whether there would be differences if the peptide visited the dilute phase, the answer is that, indeed, we would. We expect that the behaviour of the peptide would approach $\Delta\Delta G=0$, considering the low protein concentration in the dilute

phase. For mixed trajectories with sampling in both dilute and dense phases, our expectation would be a bimodal distribution in the free energy estimates from switching (see e.g. Fig. 8 in DOI:10.1021/acs.jpcc.0c10263). Because we are exclusively interested in the transfer free energies into the condensate, we do not pursue such calculations in this work.

It would also be interesting to see what the variation in the transfer of free energy is across multiple independent replicates of the transform to assess the convergence of the simulations.

Upon submission of our manuscript, we were confident that the results we had obtained would pass the test of statistical significance. We had, after all, done many more simulations than those reported, plus the comparable values of $\Delta G_{\text{Transfer}}$ for both GSY and GSF pointed in the right direction. However, we acknowledge that the more thorough test of running replicates recommended by the referee is important, considering the slow diffusion within the Tyr peptide condensates due to its stickiness. Also, the non-equilibrium switching method had not been tested before for dense phases like the ones considered here.

We have hence followed our referee's suggestion and done three different replicates, 1 μ s each, of the equilibrium runs starting from independent slab configurations, for both the GSY and GSF condensates (see the new supporting figures Fig. S1, S2 and S5). We now report the errors from the three replicates as the standard error of the mean (bootstrapping errors remain for the rest of the solvents). Our results are entirely consistent with the values reported originally, confirming the validity of our estimates.

Additionally, since the authors use a slab for the calculation of these free energies, are the transfer free energies from the dilute phase to the interface significantly different from those calculated from the dilute phase to the interior of the condensate?

We thank the referee for this valuable comment, as it has pointed us in the direction of a rapidly increasing body of work on condensate interfaces, for example, as mediators of aggregation, that we may consider for future study with the same methodology. However, as discussed above, we have not considered this possibility in our work, as we decided to focus on the condensate environment, rather than its interface.

The authors mention that the contact statistics of Phe and Tyr do not show significant difference and thereby conclude that the more favorable transfer of Tyr primarily originates from the dielectric of the condensate. However, the calculation of contacts neglects the differences in the strength of interactions involving Phe vs. Tyr. Though the authors consider the calculation of energy contact formation later in the manuscript, the scope of these interactions are quite limited (Phe-Phe, Tyr-Tyr, Tyr-Amide, Phe-Amide) which is not sufficient to make a universal conclusion regarding the underlying driving forces. A more appropriate statement would be that in the context of the minimal peptide investigated the driving force seems to be the difference in dielectric. However, it is worth mentioning that the authors do a good job of mentioning some of these caveats in the discussion section.

We thank the referee for this important comment. Indeed, the similar contact statistics and interaction patterns that we reported originally do not necessarily imply identical interaction energies. In other words, similar statistics and patterns can still result in different stabilities for the Phe and Tyr condensates if the energetics are different. Hence, we cannot conclude that the GSF and GSY condensate environments are equivalent.

To address this point, we have run new simulations for the revised version of our paper, using the temperature-replica exchange method, as before. From the new datasets, we derive the phase diagrams for both the GSF and GSY condensates (see the new Fig. 5). We find that

the tyrosine-containing condensate is more stable than that of phenylalanine, as can be inferred from the lower saturation density in the low-density branch of the phase diagram. In consequence, despite the similar contact statistics, the energetics differ, making the saturation density of the GSY slightly lower than that of GSF. This result is consistent with experimental data by Bremer et al (Nat. Chem. 2022).

Reviewer #3 (Public review):

Summary:

In this study, the authors address the paradox of how tyrosine can act as a stronger sticker for phase separation than phenylalanine, despite phenylalanine being higher on the hydrophobicity scale and exhibiting more prominent pairwise contact statistics in folded protein structures compared to tyrosine.

We are grateful for the referee's favourable opinion on the paper. Below, we address all of the issues raised.

Strengths:

This is a fascinating problem for the protein science community with special relevance for the biophysical condensate community. Using atomistic simulations of simple model peptides and condensates as well as quantum calculations, the authors provide an explanation that relies on the dielectric constant of the medium and the hydration level that either tyrosine or phenylalanine can achieve in highly hydrophobic vs. hydrophilic media. The authors find that as the dielectric constant decreases, phenylalanine becomes a stronger sticker than tyrosine. The conclusions of the paper seem to be solid, it is well-written and it also recognises the limitations of the study. Overall, the paper represents an important contribution to the field.

Weaknesses:

How can the authors ensure that a condensate of GSY or GSF peptides is a representative environment of a protein condensate? First, the composition in terms of amino acids is highly limited, second the effect of peptide/protein length compared to real protein sequences is also an issue, and third, the water concentration within these condensates is really low as compared to real experimental condensates. Hence, how can we rely on the extracted conclusions from these condensates to be representative for real protein sequences with a much more complex composition and structural behaviour?

We agree with the main weakness identified by the referee. In fact, all these limitations had already been stated in our original submission. Our ternary peptide condensates are just a minimal model system that bears reasonable analogies with condensates, but definitely is not identical to true LCR condensates. The analogies between peptide and protein condensates are, however, worth restating:

- (1) The limited composition of the peptide condensates is inspired by LCR sequences (see Fig. 4 in Martin & Mittag, 2018).
- (2) The equilibrium phase diagram, showing a UCST, is consistent with that of LCRs from Ddx4 or hnRNPA1.
- (3) The dynamical behaviour is intermediate between liquid and solid (De Sancho, 2022).
- (4) The contact patterns are comparable to those observed for FUS and LAF1 (Zheng et al, 2020).

The third issue pointed out by the referee requires particular attention. Indeed, the water content in the model condensates is low (~ 200 mg/mL for GSY) relative to the experiment (e.g. ~ 600 mg/mL for FUS and LAF-1 from simulations). Considering that both interaction patterns and solvation contribute to the favorability of Tyr relative to Phe, we speculate that a greater degree of solvation in the true protein condensates will further reinforce the trends we observe.

In any case, in the revised version of the manuscript, we have made an effort to insist on the limitations of our results, some of which we plan to address in future work.

Reviewer #3 (Recommendations for the authors):

(1) The fact that protein density is so high within GSY or GSF peptide condensates may significantly alter the conclusions of the paper. Can the authors show that for condensates in which the protein density is ~ 0.2 - 0.3 g/cm³, the same conclusions hold? Could the authors use a different peptide sequence that establishes a more realistic protein concentration/density inside the condensate?

Unfortunately, recent work with a variety of peptide sequences suggests that finding peptides in the density range proposed by the referee may be very challenging. For example, Pettit and his co-workers have extensively studied the behaviour of GGXGG peptides. In a recent work, using the CHARMM36m force field and TIP3P water, they report densities of ~ 1.2 - 1.3 g/mL for capped pentapeptide condensates (Workman et al, Biophys. J. 2024; DOI: 10.1016/j.bpj.2024.05.009). Brown and Potoyan have recently run simulations of zwitterionic GXG tripeptides with the Amber99sb-ILDN force field and TIP3P water, starting with a homogenous distribution in cubic simulation boxes (Biophys. J. 2024, DOI: 10.1016/j.bpj.2023.12.027). In a box with an initial concentration of 0.25 g/mL, upon phase separation, the peptide ends up occupying what would seem to be $\sim 1/3$ of the box, although we could not find exact numbers. This would imply densities of ~ 0.75 g/mL in the dense phase, with the additional problem of many charges. Finally, Joseph and her co-workers have recently simulated a set of hexapeptide condensates with varied compositions using a combination of atomistic and coarse-grained simulations. For the atomistic simulations, the Amber03ws force field and TIP4P water were used (see BioRxiv reference 10.1101/2025.03.04.641530). They have found values of the protein density in the dense phase ranging between 0.8 and 1.2 g/mL. The consistency in the range of densities reported in these studies suggests that short peptides, at least up to 7-residues long, tend to form quite dense condensates, akin to those investigated in our work. While the examples mentioned do not comprehensively span the full range of peptide lengths, sequences, and force fields, they nonetheless support the general behaviour we observe. A systematic exploration of all these variables would require an extensive search in parameter space, which we believe falls outside the scope of the present study.

(2) Do the conclusions hold for phase-separating systems that mostly rely on electrostatic interactions to undergo LLPS, like protein-RNA complex coacervates? In other words, could the authors try the same calculations for a binary mixture composed of polyR-polyE, or polyK-polyE?

This is an excellent idea that we may attempt in future work, but the remit of the current work is aromatic amino acids Phe and Tyr only. Hence, we do not include calculations or discussion on polyR-polyE systems in our revised manuscript.

(3) One of the major approximations made by the authors is the length of the peptides within the condensates, which is not realistic, or their density. Specifically, could they

double or triple the length of these peptides while maintaining their composition so it can be quantified the impact of sequence length in the transfer of free energies?

We thank the referee for this comment and agree with the main point, which was stated as a limitation in our original submission. The suggested calculations anticipate research that we are planning but will not include in the current work. One of the advantages of our model systems is that the small size of the peptides allows for small simulation boxes and relatively rapid sampling. Longer peptide sequences would require conformational sampling beyond our current capabilities, if done systematically. An example of these limitations is the amount of data that we had to discard from the new simulations we report, which amounts to up to 200 ns of our replica exchange runs in smaller simulation boxes (i.e. >19 μ s in total for the 48 replicas of the two condensates!). As stated in the answer to point 1, we have found in the literature work on peptides in the range of 1-7 residues with consistent densities. Additionally, a recent report using alchemical transformations using equilibrium techniques with tetrapeptide condensates, pointing to the role of transfer free energy as driving force for condensate formation, further supports the observations from our work.

Minor issues:

(1) The caption of Figure 3B is not clear. It can only be understood what is depicted there once you read the main text a couple of times. I encourage the authors to clarify the caption.

We have rewritten the caption for greater clarity. Now it reads as follows:

Time evolution of the density profiles calculated across the longest dimension of the simulation box (L) in the coexistence simulations. In blue we show the density of all the peptides, and in dark red that of the F/Y residue in the GGXGG peptide.

(2) Why was the RDF from Figure 5A cut at such a short distance? Can the authors expand the figure to clearly show that it has converged?

In the updated Figure 5 (now Fig. 6), we have extended the $g(r)$ up to $r=1.75$ nm so that it clearly plateaus at a value of 1.

<https://doi.org/10.7554/eLife.104950.2.sa0>