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

David De Sancho[1*] **and Xabier López**[1*]

***For correspondence:**
david.desancho@ehu.eus (DS);
xabier.lopez@ehu.eus (XL)

[1]Polimero eta Material Aurreratuak: Fisika, Kimika eta Teknologia, Kimika Fakultatea, UPV/EHU & Donostia International Physics Center (DIPC), PK 1072, 20018 Donostia-San Sebastian, Euskadi, Spain

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 devia-

tions 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

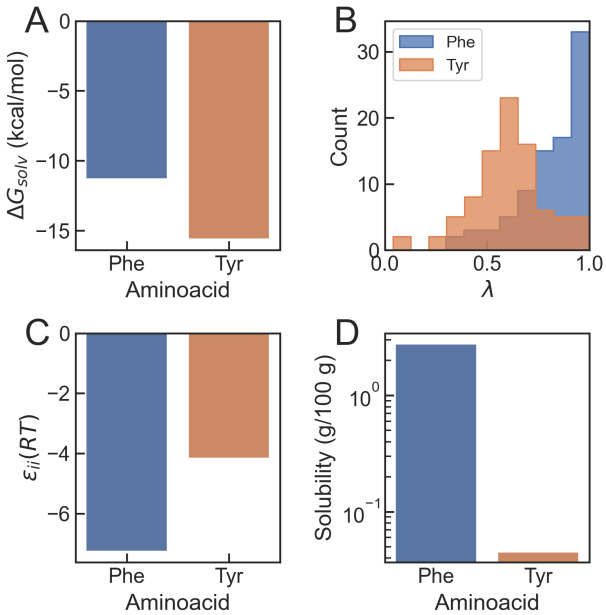


Figure 1. Bibliographic values for different properties of phenylalanine and tyrosine. (A) Solvation free energy (ΔG_{solv}) (***Chang et al., 2007***). (B) Probability distributions of min-maxed normalized hydropathy values λ from bibliographic hydrophobicity scales (***Tesei et al., 2021***). (C) Self-interaction energy (ε_{ii}) from the Miyazawa-Jernigan contact matrix (***Miyazawa and Jernigan, 1996***). (D) Solubility in water at 25°C (***Nozaki and Tanford, 1971***).

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***).

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***; ***Das et al., 2020***). 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***) and simulation (***Workman and Pettitt, 2021***). We have built structures for the peptides using Ambertools (***Salomon-Ferrer et al., 2013***). Our condensates are formed by an equimolar mixture of Gly, Ser and either Tyr or Phe "dipeptides" (in fact, terminally capped amino acids), as before (***De Sancho, 2022***). 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***; ***Lindorff-Larsen et al., 2010***), and TIP3P water (***Jorgensen et al., 1983***). 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***). For the simulations with different organic solvents, we used the GAFF parameters derived by Mobley and co-workers (***Bannan et al., 2016***). 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 ($\Delta\Delta G_{\mathrm{Transfer}}$) 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 ($\Delta G^{\mathrm{F}}_{\mathrm{Transfer}}$ and $\Delta G^{\mathrm{Y}}_{\mathrm{Transfer}}$), while the horizontal branches involve the alchemical transformation ($\Delta G_{\mathrm{F}\to\mathrm{Y}}(\mathrm{H_2O})$ and $\Delta G_{\mathrm{F}\to\mathrm{Y}}(\mathrm{Cond})$). Using simulations, we can easily calculate the free energies corresponding to the horizontal branches. This enables us to estimate the desired quantity as

$$\Delta\Delta G_{\mathrm{Transfer}} = \Delta G^{\mathrm{Y}}_{\mathrm{Transfer}} - \Delta G^{\mathrm{F}}_{\mathrm{Transfer}} = \Delta G_{\mathrm{F}\to\mathrm{Y}}(\mathrm{Cond}) - \Delta G_{\mathrm{F}\to\mathrm{Y}}(\mathrm{H_2O}) \tag{1}$$

Negative $\Delta\Delta G_{\mathrm{Transfer}}$ 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

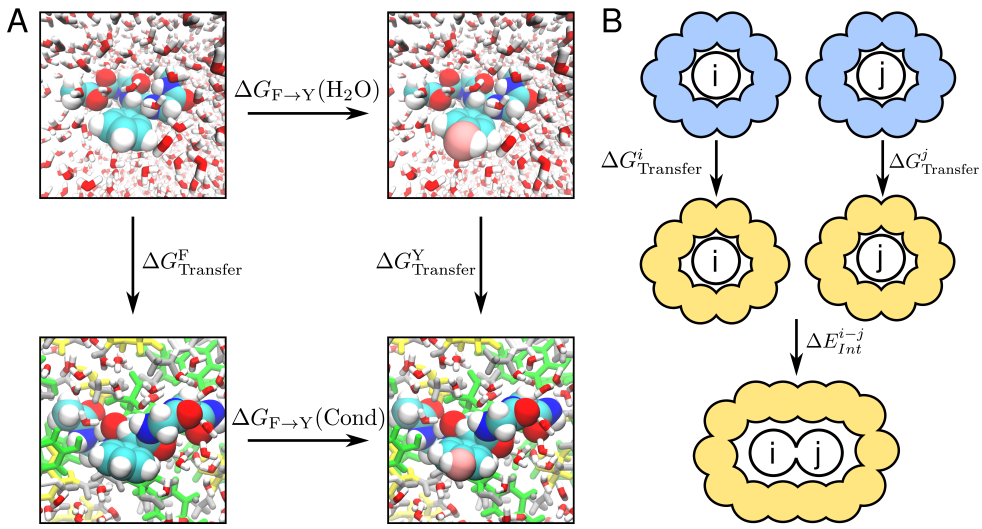


Figure 2. (a) Thermodynamic cycle used in this study for the estimation of free energy differences upon mutation for the insertion of a peptide in a molecular condensate. (B) Schematic of the process of contact formation between two molecules i and j used in the quantum chemical calculations. We consider that contact formation involves the transfer from water (blue) to a different medium (orange) and the interaction in this medium between the entities involved.

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 μs 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 ($\lambda = 0 \to 1$) or backward ($\lambda = 1 \to 0$).

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 μs.

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 (i.e. $W = \int_{\lambda=0}^{\lambda=1} \frac{\partial H}{\partial \lambda} d\lambda$) 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

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

where x_{DS} and t are, respectively, the position and width of the dividing surface between dense and dilute phases (***Tesei et al., 2021***). 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

$$\rho_H - \rho_L = A(T_c - T)^{\beta} \tag{3}$$

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

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

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***)

$$\varepsilon = 1 + \frac{\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2}{3\varepsilon_0 k_B T \langle V \rangle}. \tag{5}$$

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 ωB97XD functional (***Chai and Head-Gordon, 2008***) and Pople's 6-311++G(d,p) basis set (***Hehre et al., 1972***) 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***). The interaction energy in a given solvent was calculated according to

$$\Delta E_{\text{Int}}^{ij} = E^{ij} - \left(E^{i} + E^{j}\right) \tag{6}$$

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

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

As solvents, we have considered three different groups:

1. Alcohols: methanol, ethanol, 1-propanol, 1-butanol, 1-poentanol, 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, nromobenzene, 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***).

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 3B). 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

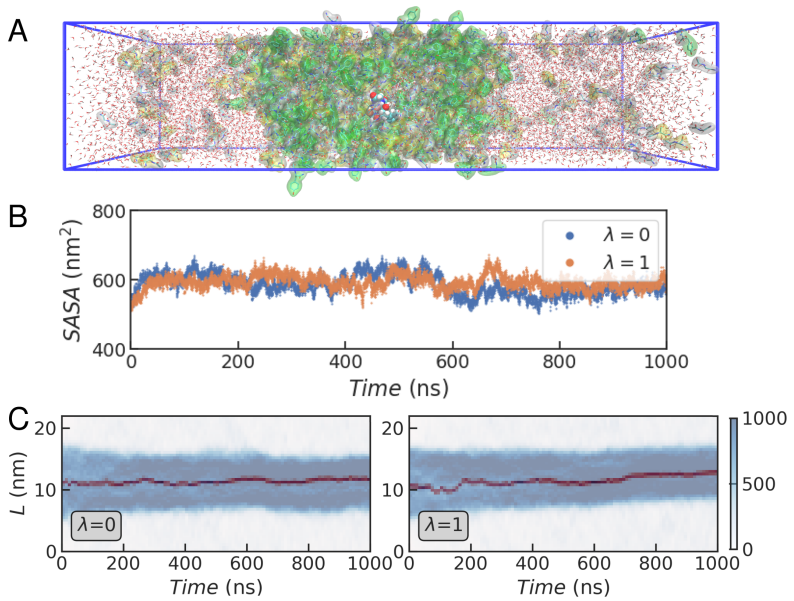


Figure 3. (A) Representative simulation box with a fully solvated GSY condensate in slab geometry including a GGXGG peptide (spheres) and the capped amino acid mixture (G: white, S: yellow and Y: green). (B) Time series for the solvent accessible surface area ($SASA$) in a representative trajectory of 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.

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

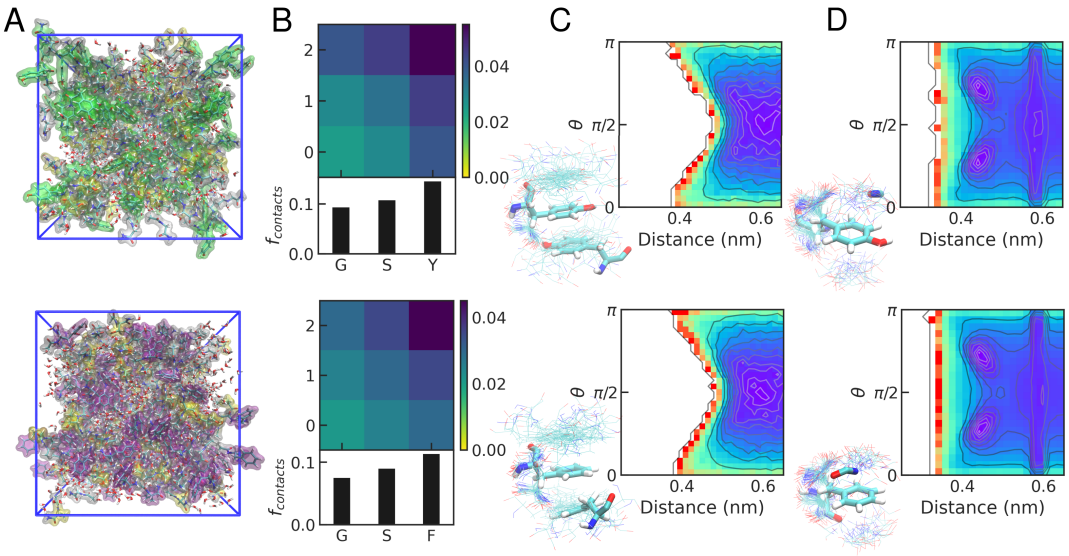


Figure 4. (A) Representative snapshots from the simulations of dense phases. G/S/F/Y dipeptides are shown as transparent surfaces (G: white, S: yellow, Y: green, F: purple). (B) Interaction matrix for the normalized number of contacts between different pairs of amino acid residues in the condensates (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-π 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.

those from the simulations in water, we obtain a value for the transfer free energy difference of $\Delta\Delta G_{\text{Transfer}}$ =-2.5±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***).

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.

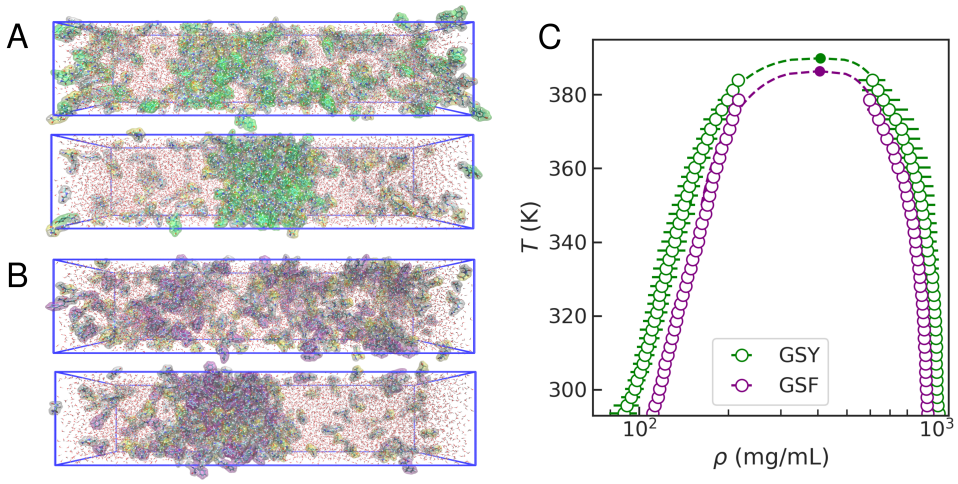


Figure 5. (A) Representative snapshots of REMD simulations of the GSY condensate at high (top) and low (bottom) temperatures. Colour code as in Figure 3. (B) Same for the GSF condensate. (C) Phase diagrams for GSY (green) and GSF (purple). Empty circles correspond to simulations and filled circles correspond to fitted critical temperatures (T_c) and densities (ρ_c).

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}).

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 strenghts 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^{ζ} 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

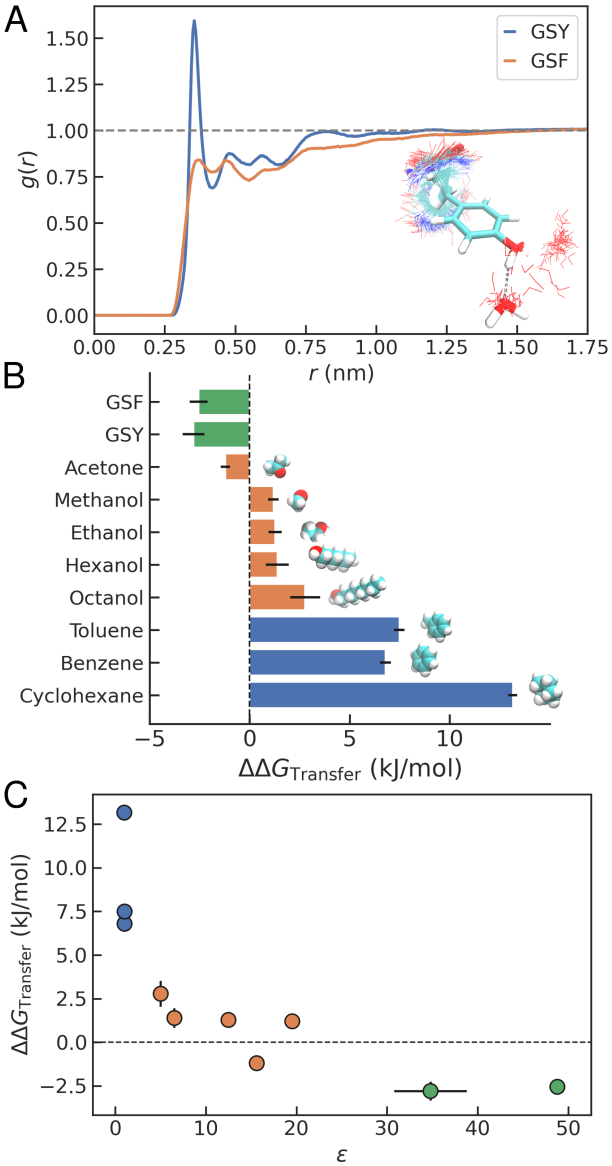


Figure 6. (A) Radial distribution function for water oxygen around the C^{ζ} in Phe/Tyr for GSF and GSY condensates. We show a representative overlay of simulation snapshots where water molecules are hydrogen-bonded to the Tyr side chain. (B) Transfer free energy differences from water to a different 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.

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 thermodynamic 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.

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±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±0.2 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 ϕ =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

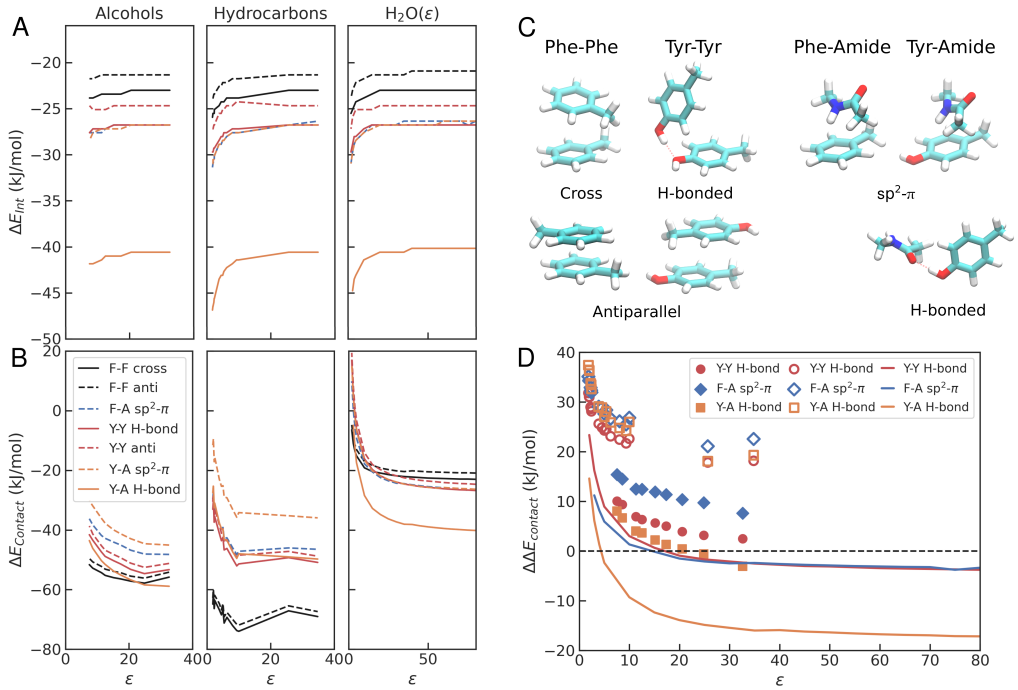


Figure 7. (A) ΔE^{ij}_{Int} and (B) $\Delta E^{ij}_{contact}$ for different interaction pairs ij and different solvents as a function of the dielectric constant, ε. (C) Optimized geometries from quantum chemical calculations for Phe and Tyr interaction pairs. (D) $\Delta\Delta E^{\text{Solvent}}_{\text{Contact}}$ 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.

forms a hydrogen bond with the amide through the alcohol group of p-cresol.

To explore the environmental dependence in contact formation for amino acid pairs, we have calculated values of the interaction energies ($\Delta E^{ij}_{\text{Int}}$) 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-π interactions, namely, Y-Y H-bond, Y-A sp^2-π and F-A sp^2-π 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 $\Delta E^{ij}_{\text{Int}}$ (see Figure 2B). Hence, for a given solvent, the total energy can be calculated from the sum of these two contributions as

$$\Delta E^{ij}_{\text{Contact}} = \Delta E^{ij}_{\text{Int}} + \Delta G^{i}_{\text{Transfer}} + \Delta G^{j}_{\text{Transfer}} \tag{8}$$

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, $\Delta G^{i}_{\text{Transfer}}$ 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, $\Delta E^{ij}_{\text{Contact}}$ 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 $\Delta\Delta E^{ij}_{\text{Contact}}$ 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 $\Delta\Delta E^{ij}_{\text{Contact}}$ 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 $\Delta\Delta E^{ij}_{\text{Contact}}$ 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***) and are known to be dominated by hydrophobicity effects (***Pokarowski et al., 2005***). 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***; ***Joseph et al., 2021***). 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***)), 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***).

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***) or the influence of multivalency in protein IDRs (***Martin et al., 2020***). 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***; ***Das et al., 2020***; ***Aldeghi et al., 2018***; ***Hazra and Levy, 2023***; ***Calinsky and Levy, 2024***). 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***; ***Brown and Potoyan, 2024***; ***Emelianova et al., 2025***). 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***). 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 (ε =100±5, (***Braun et al., 2014***)). However, using the TIP4P-Ew water model (with ε =64±1, (***Horn et al., 2004***)), 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***). The effect 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.

Acknowledgments

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).

References

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015; 1-2:19 – 25. doi: 10.1016/j.softx.2015.06.001.

Aldeghi M, Gapsys V, de Groot BL. Accurate estimation of ligand binding affinity changes upon protein mutation. ACS Cent Sci. 2018; 4(12):1708–1718.

Aldeghi M, de Groot BL, Gapsys V. In: Sikosek T, editor. Accurate Calculation of Free Energy Changes upon Amino Acid Mutation New York, NY: Springer New York; 2019. p. 19–47. doi: 10.1007/978-1-4939-8736-8_2.

Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular biochemistry. Nat Rev Mol Cell Biol. 2017; 18(5):285–298. doi: 10.1038/nrm.2017.7.

Bannan CC, Calabró G, Kyu DY, Mobley DL. Calculating Partition Coefficients of Small Molecules in Octanol/Water and Cyclohexane/Water. J Chem Theory Comput. 2016; 12(8):4015–4024. doi: 10.1021/acs.jctc.6b00449.

Becke AD. A new mixing of Hartree-Fock and local density-functional theories. J Chem Phys. 1993; 98(2):1372–1377. doi: 10.1063/1.464304.

Bennett CH. Efficient estimation of free energy differences from Monte Carlo data. J Comput Phys. 1976; 22(2):245–268.

Berendsen HJ, Postma Jv, Van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984; 81(8):3684–3690.

Best RB, Hummer G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. J Phys Chem B. 2009; 113(26):9004–9015. doi: 10.1021/jp901540t.

Brady JP, Farber PJ, Sekhar A, Lin YH, Huang R, Bah A, Nott TJ, Chan HS, Baldwin AJ, Forman-Kay JD, et al. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. Proc Natl Acad Sci USA. 2017; 114(39):E8194–E8203.

Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, Jülicher F, Hyman AA. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science. 2009; 324(5935):1729–1732.

Braun D, Boresch S, Steinhauser O. Transport and dielectric properties of water and the influence of coarse-graining: Comparing BMW, SPC/E, and TIP3P models. The Journal of chemical physics. 2014; 140(6).

Bremer A, Farag M, Borcherds WM, Peran I, Martin EW, Pappu RV, Mittag T. Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. Nat Chem. 2022; (14):196–207. doi: 10.1038/s41557-021-00840-w.

Brown WH, Potoyan DA. Phase separation of multicomponent peptide mixtures into dehydrated clusters with hydrophilic cores. Biophys J. 2024; 123(3):349–360.

Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. 2007; 126(1):014101.

Calinsky R, Levy Y. Aromatic Residues in Proteins: Re-Evaluating the Geometry and Energetics of π–π, Cation-π, and CH- π Interactions. J Phys Chem B. 2024; .

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. The Energetic Origins of Pi–Pi Contacts in Proteins. J Am Chem Soc. 2023; 145(45):24836–24851.

Chai JD, Head-Gordon M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. Phys Chem Chem Phys. 2008; 10(44):6615–6620.

Chang J, Lenhoff AM, Sandler SI. Solvation free energy of amino acids and side-chain analogues. J Phys Chem B. 2007; 111(8):2098–2106.

Chelli R, Gervasio FL, Procacci P, Schettino V. Stacking and T-shape competition in aromatic- aromatic amino acid interactions. J Am Chem Soc. 2002; 124(21):6133–6143.

Chengteh L, Yang W, Parr RG. Development of the Colic-Salvetti correlation-energy formula into a functional of the electron density. Phys Rev B. 1988; 37(2):785–789.

Chipot C, Jaffe R, Maigret B, Pearlman DA, Kollman PA. 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. 1996; 118(45):11217–11224. doi: 10.1021/ja961379l.

Choi JM, Holehouse AS, Pappu RV. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. Annu Rev Biophys. 2020; 49(1):107–133. doi: 10.1146/annurev-biophys-121219-081629.

Das S, Lin YH, Vernon RM, Forman-Kay JD, Chan HS. Comparative roles of charge, π, and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins. Proc Natl Acad Sci USA. 2020; 117(46):28795–28805.

De Sancho D. Phase separation in amino acid mixtures is governed by composition. Biophys J. 2022; 121(21):4119–4127.

Dignon GL, Best RB, Mittal J. Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties. Annu Rev Phys Chem. 2020; 71(1):53–75. doi: 10.1146/annurev-physchem-071819-113553.

Dignon GL, Zheng W, Kim YC, Best RB, Mittal J. Sequence determinants of protein phase behavior from a coarse-grained model. PLoS Comput Biol. 2018 01; 14(1):1–23. doi: 10.1371/journal.pcbi.1005941.

Emelianova A, Garcia PL, Tan D, Joseph JA. Prediction of small-molecule partitioning into biomolecular condensates from simulation. bioRxiv. 2025; p. 2025–03.

Fisher RS, Elbaum-Garfinkle S. Tunable multiphase dynamics of arginine and lysine liquid condensates. Nat Commun. 2020; 11(1):4628.

Galvanetto N, Ivanović MT, Chowdhury A, Sottini A, Nüesch MF, Nettels D, Best RB, Schuler B. Extreme dynamics in a biomolecular condensate. Nature. 2023; 619(7971):876–883.

Gapsys V, Michielssens S, Seeliger D, de Groot BL. pmx: Automated protein structure and topology generation for alchemical perturbations. J Comput Chem. 2015; 36(5):348–354. doi: https://doi.org/10.1002/jcc.23804.

Gapsys V, Michielssens S, Seeliger D, de Groot BL. Accurate and rigorous prediction of the changes in protein free energies in a large-scale mutation scan. Angew Chem Int Ed. 2016; 55(26):7364–7368.

Gapsys V, Pérez-Benito L, Aldeghi M, Seeliger D, Van Vlijmen H, Tresadern G, De Groot BL. Large scale relative protein ligand binding affinities using non-equilibrium alchemy. Chem Sci. 2020; 11(4):1140–1152.

Gobbi A, Frenking G. Y-conjugated compounds: the equilibrium geometries and electronic structures of guanidine, guanidinium cation, urea, and 1, 1-diaminoethylene. J Am Chem Soc. 1993; 115(6):2362–2372.

Greig JA, Nguyen TA, Lee M, Holehouse AS, Posey AE, Pappu RV, Jedd G. Arginine-enriched mixed-charge domains provide cohesion for nuclear speckle condensation. Mol Cell. 2020; 77(6):1237–1250.

Hazra MK, Levy Y. Cross-talk of cation- π interactions with electrostatic and aromatic interactions: A salt-dependent trade-off in biomolecular condensates. J Phys Chem Lett. 2023; 14(38):8460–8469.

Hehre WJ, Ditchfield K, Pople JA. 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. 1972; 56(5):2257–2261. doi: 10.1063/1.1677527.

Hong Y, Najafi S, Casey T, Shea JE, Han SI, Hwang DS. Hydrophobicity of arginine leads to reentrant liquid-liquid phase separation behaviors of arginine-rich proteins. Nat Commun. 2022; 13(1):7326.

Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J Chem Phys. 2004; 120(20):9665–9678.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. J Chem Phys. 1983; 79(2):926–935.

Joseph JA, Reinhardt A, Aguirre A, Chew PY, Russell KO, Espinosa JR, Garaizar A, Collepardo-Guevara R. Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. Nat Comput Sci. 2021; 1(11):732–743. doi: 10.1038/s43588-021-00155-3.

Kauzmann W. Some factors in the interpretation of protein denaturation. In: *Adv. Protein Chem.*, vol. 14 Elsevier; 1959.p. 1–63.

Kaygisiz K, Sementa D, Athiyarath V, Chen X, Ulijn RV. Context dependence in assembly code for supramolecular peptide materials and systems. Nature Reviews Materials. 2025; p. 1–24.

Kim YC, Hummer G. Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. J Mol Biol. 2008; 375(5):1416–1433.

Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982; 157(1):105–132.

Lin Y, Currie SL, Rosen MK. Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. J Biol Chem. 2017; 292(46):19110–19120.

Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins. 2010; 78(8):1950–1958. doi: 10.1002/prot.22711.

Lobo S, Najafi S, Shell MS, Shea JE. Hydrophobicity in Intrinsically Disordered Protein Force Fields: Implications for Conformational Ensembles and Protein–Protein Interactions. J Phys Chem B. 2025; .

Maraldo A, Rnjak-Kovacina J, Marquis C. Tyrosine–a structural glue for hierarchical protein assembly. Trends Biochem Sci. 2024; .

Marenich AV, Cramer CJ, Truhlar DG. 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. 2009; 113(18):6378–6396.

Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, Grace CR, Soranno A, Pappu RV, Mittag T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. Science. 2020; 367(6478):694–699. doi: 10.1126/science.aaw8653.

Martin EW, Mittag T. Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. Biochemistry. 2018; 57(17):2478–2487. doi: 10.1021/acs.biochem.8b00008.

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. Titin domains with reduced core hydrophobicity cause dilated cardiomyopathy. Cell Rep. 2023; 42(12).

McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, Schwantes CR, Wang LP, Lane TJ, Pande VS. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophys J. 2015; 109(8):1528 – 1532. doi: 10.1016/j.bpj.2015.08.015.

Mennucci B. Polarizable continuum model. Wiley Interdiscip Rev Comput Mol Sci. 2012; 2(3):386–404.

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. Best Practices for Alchemical Free Energy Calculations [Article v1.0]. Living J Comp Mol Sci. 2020; 2(1):18378. doi: 10.33011/livecoms.2.1.18378.

Miller MS, Lay WK, Elcock AH. Osmotic pressure simulations of amino acids and peptides highlight potential routes to protein force field parameterization. J Phys Chem B. 2016; 120(33):8217–8229.

Mittag T, Pappu RV. A conceptual framework for understanding phase separation and addressing open questions and challenges. Mol Cell. 2022; 82(12):2201–2214. doi: https://doi.org/10.1016/j.molcel.2022.05.018.

Miyazawa S, Jernigan RL. Residue–residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J Mol Biol. 1996; 256(3):623–644.

Murthy AC, Tang WS, Jovic N, Janke AM, Seo DH, Perdikari TM, Mittal J, Fawzi NL. Molecular interactions contributing to FUS SYGQ LC-RGG phase separation and co-partitioning with RNA polymerase II heptads. Nat Struct Mol Biol. 2021; 28(11):923–935.

Nerenberg PS, Head-Gordon T. Optimizing protein- solvent force fields to reproduce intrinsic conformational preferences of model peptides. J Chem Theory Comput. 2011; 7(4):1220–1230.

Neumann M. Dipole moment fluctuation formulas in computer simulations of polar systems. Mol Phys. 1983; 50(4):841–858.

Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowietz A, Craggs TD, Bazett-Jones DP, Pawson T, Forman-Kay JD, Baldwin AJ. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. Mol Cell. 2015; 57(5):936–947.

Nozaki Y, Tanford C. The Solubility of Amino Acids and Two Glycine Peptides in Aqueous Ethanol and Dioxane Solutions: establishment of a hydrophobicity scale. J Biol Chem. 1971; 246(7):2211–2217. doi: 10.1016/S0021-9258(19)77210-X.

Pace CN, Fu H, Fryar KL, Landua J, Trevino SR, Shirley BA, Hendricks MM, Iimura S, Gajiwala K, Scholtz JM, Grimsley GR. Contribution of Hydrophobic Interactions to Protein Stability. JMol Biol. 2011; 408(3):514–528. doi: https://doi.org/10.1016/j.jmb.2011.02.053.

Pace CN, Horn G, Hebert EJ, Bechert J, Shaw K, Urbanikova L, Scholtz JM, Sevcik J. Tyrosine hydrogen bonds make a large contribution to protein stability. J Mol Biol. 2001; 312(2):393–404.

Paloni M, Bailly R, Ciandrini L, Barducci A. Unraveling Molecular Interactions in Liquid–Liquid Phase Separation of Disordered Proteins by Atomistic Simulations. J Phys Chem B. 2020; 124(41):9009–9016. doi: 10.1021/acs.jpcb.0c06288.

Paloni M, Bussi G, Barducci A. Arginine multivalency stabilizes protein/RNA condensates. Protein Sci. 2021; 30(7):1418–1426.

Parrinello M, Rahman A. Crystal structure and pair potentials: A molecular-dynamics study. Phys Rev Lett. 1980; 45(14):1196.

Piana S, Donchev AG, Robustelli P, Shaw DE. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. J Phys Chem B. 2015; 119(16):5113–5123. doi: 10.1021/jp508971m.

Pitera JW, Falta M, van Gunsteren WF. Dielectric properties of proteins from simulation: the effects of solvent, ligands, pH, and temperature. Biophys J. 2001; 80(6):2546–2555.

Plaxco KW, Morton CJ, Grimshaw SB, Jones JA, Pitkeathly M, Campbell ID, Dobson CM. The effects of guanidine hydrochloride on the'random coil 'conformations and NMR chemical shifts of the peptide series GGXGG. J Biomol NMR. 1997; 10:221–230.

Pokarowski P, Kloczkowski A, Jernigan RL, Kothari NS, Pokarowska M, Kolinski A. Inferring ideal amino acid interaction forms from statistical protein contact potentials. Proteins. 2005; 59(1):49–57.

Radzicka A, Wolfenden R. Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. Biochemistry. 1988; 27(5):1664–1670.

Regy RM, Thompson J, Kim YC, Mittal J. Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins. Protein Sci. 2021; 30(7):1371–1379. doi: 10.1002/pro.4094.

Rekhi S, Garcia CG, Barai M, Rizuan A, Schuster BS, Kiick KL, Mittal J. Expanding the molecular language of protein liquid–liquid phase separation. Nat Chem. 2024; p. 1–12.

Rekhi S, Mittal J. Amino acid transfer free energies reveal thermodynamic driving forces in biomolecular condensate formation. Proc Natl Acad Sci USA. 2025; 122(30):e2425422122. doi: 10.1073/pnas.2425422122.

Rogers DM, Hirst JD, Lee EPF, Wright TG. Ab initio study of the toluene dimer. Chem Phys Lett. 2006; 427(4-6):410–413. doi: 10.1016/j.cplett.2006.07.022.

Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. Wiley Interdiscip Rev Comput Mol Sci. 2013; 3(2):198–210.

Schuster BS, Dignon GL, Tang WS, Kelley FM, Ranganath AK, Jahnke CN, Simpkins AG, Regy RM, Hammer DA, Good MC, Mittal J. Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. Proc Natl Acad Sci USA. 2020; 117(21):11421–11431. doi: 10.1073/pnas.2000223117.

Seeliger D, de Groot BL. Protein Thermostability Calculations Using Alchemical Free Energy Simulations. Biophys J. 2010; 98(10):2309–2316. doi: 10.1016/j.bpj.2010.01.051.

Semenov AN, Rubinstein M. Thermoreversible gelation in solutions of associative polymers. 1. Statics. Macromolecules. 1998; 31(4):1373–1385.

Serrano L, Bycroft M, Fersht AR. Aromatic-aromatic interactions and protein stability: investigation by double-mutant cycles. J Mol Biol. 1991; 218(2):465–475.

Shimizu S, Chan HS. Statistical mechanics of solvophobic aggregation: Additive and cooperative effects. J Chem Phys. 2001; 115(7):3424–3431.

Shimizu S, Chan HS. Origins of protein denatured state compactness and hydrophobic clustering in aqueous urea: inferences from nonpolar potentials of mean force. Proteins: Structure, Function, and Bioinformatics. 2002; 49(4):560–566.

Shin Y, Brangwynne CP. Liquid phase condensation in cell physiology and disease. Science. 2017; 357(6357):eaaf4382. doi: 10.1126/science.aaf4382.

Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett. 1999; 314(1-2):141–151.

Tang Y, Bera S, Yao Y, Zeng J, Lao Z, Dong X, Gazit E, Wei G. Prediction and characterization of liquid-liquid phase separation of minimalistic peptides. Cell Rep Phys Sci. 2021; 2(9):100579. doi: 10.1016/j.xcrp.2021.100579.

Tesei G, Schulze TK, Crehuet R, Lindorff-Larsen K. Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. Proc Natl Acad Sci U S A. 2021; 118(44):e2111696118. doi: 10.1073/pnas.2111696118.

Thomas PD, Dill KA. An iterative method for extracting energy-like quantities from protein structures. Proc Natl Acad Sci USA. 1996; 93(21):11628–11633.

Tsuzuki S, Honda K, Uchimaru T, Mikami M. Ab initio calculations of structures and interaction energies of toluene dimers including CCSD(T) level electron correlation correction. J Chem Phys. 2005; 122(14). doi: 10.1063/1.1876092.

Urry DW, Gowda DC, Parker TM, Luan CH, Reid MC, Harris CM, Pattanaik A, Harris RD. Hydrophobicity scale for proteins based on inverse temperature transitions. Biopolymers. 1992; 32(9):1243–1250.

Vernon RM, Chong PA, Tsang B, Kim TH, Bah A, Farber P, Lin H, Forman-Kay JD. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. eLife. 2018; 7:e31486.

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. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. Cell. 2018; 174(3):688–699.e16.

Wimley WC, Creamer TP, White SH. Solvation energies of amino acid side chains and backbone in a family of host- guest pentapeptides. Biochemistry. 1996; 35(16):5109–5124.

Wolfenden R, Andersson L, Cullis P, Southgate C. Affinities of amino acid side chains for solvent water. Biochemistry. 1981; 20(4):849–855.

Wood RH, Thompson PT. Differences between pair and bulk hydrophobic interactions. Proc Natl Acad Sci U S A. 1990; 87(3):946–949.

Workman RJ, Huang CJ, Lynch GC, Pettitt BM. Peptide diffusion in biomolecular condensates. Biophys J. 2024; 123(12):1668–1675.

Workman RJ, Pettitt BM. Thermodynamic Compensation in Peptides Following Liquid–Liquid Phase Separation. J Phys Chem B. 2021; 125(24):6431–6439.

Zheng W, Dignon GL, Jovic N, Xu X, Regy RM, Fawzi NL, Kim YC, Best RB, Mittal J. Molecular Details of Protein Condensates Probed by Microsecond Long Atomistic Simulations. J Phys Chem B. 2020; 124(51):11671–11679. doi: 10.1021/acs.jpcb.0c10489.

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

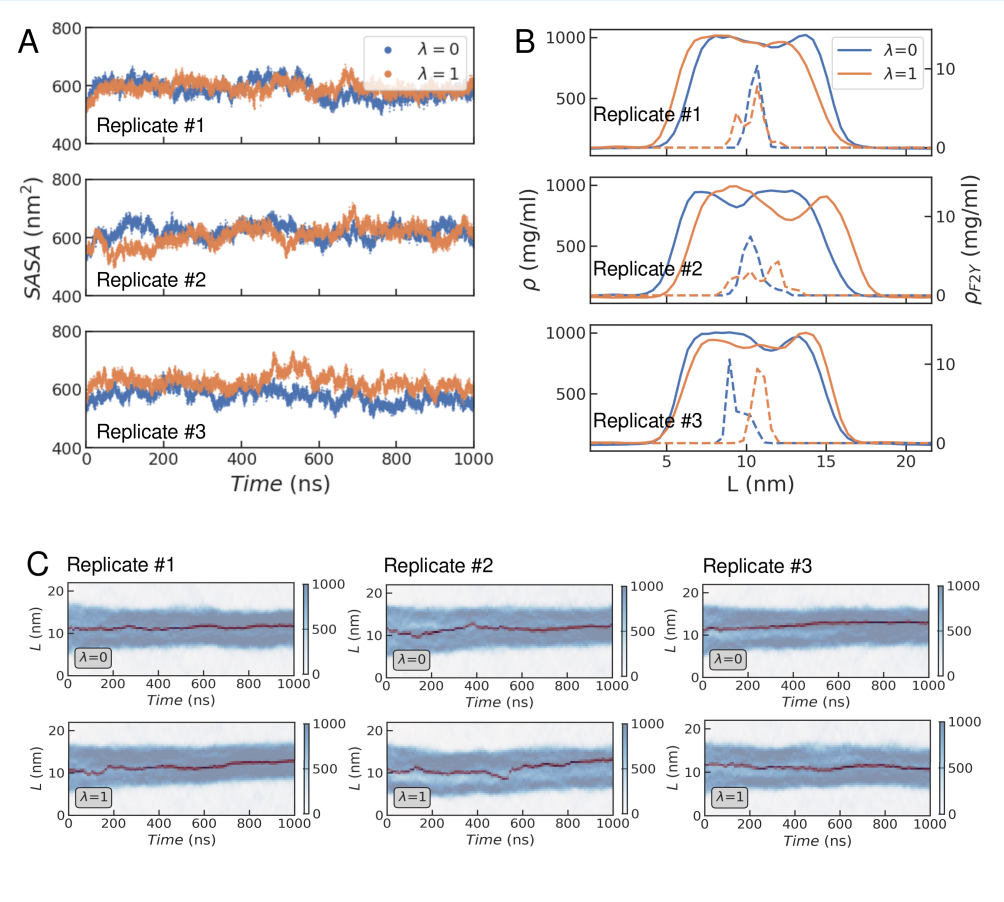
Appendix 1—table 2. Number of molecules used in pure solvent simulations for the determination of dielectric constants.

System	Simulation Time (μs)
GSY slab	3×1
GSF slab	3×1
GSY slab TIP4P-Ew	1
GSY slab REMD	48×1
GSF slab REMD	48×1
Solvents	0.5

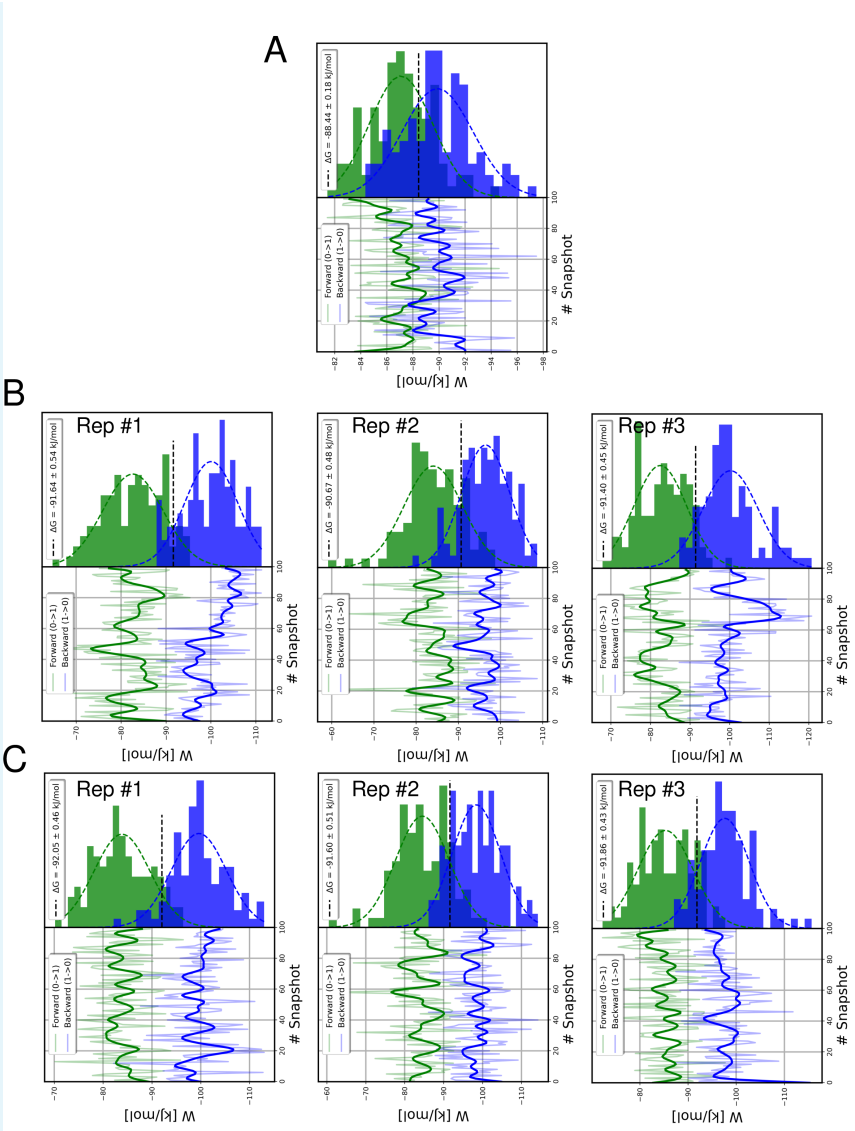
Appendix 1—table 3. Simulation times for each of the families of systems considered in this study. For GSY and GSF condensates, we run triplicates. For the simulations in pure solvents (including water), a single simulation run was performed for each system.

Appendix 2

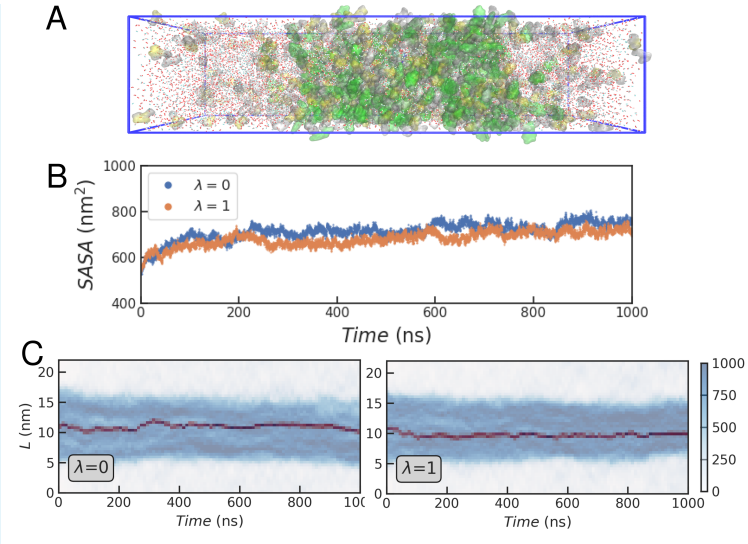
Supporting Figures



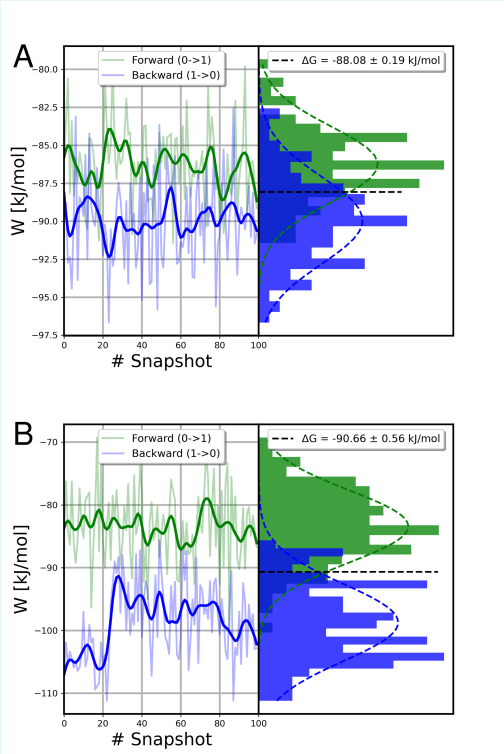
Appendix 2—figure S1. Replicates of the GGXGG peptide simulations in the GSY condensate. (A) Time series for the accessible surface area for different values of λ. (B) Density profiles of peptide (solid lines) and F/Y residue (right Y-axis) averaged over the simulation runs. (C) Time series for the density profiles of all peptides (blue) and the F2Y residue (red).



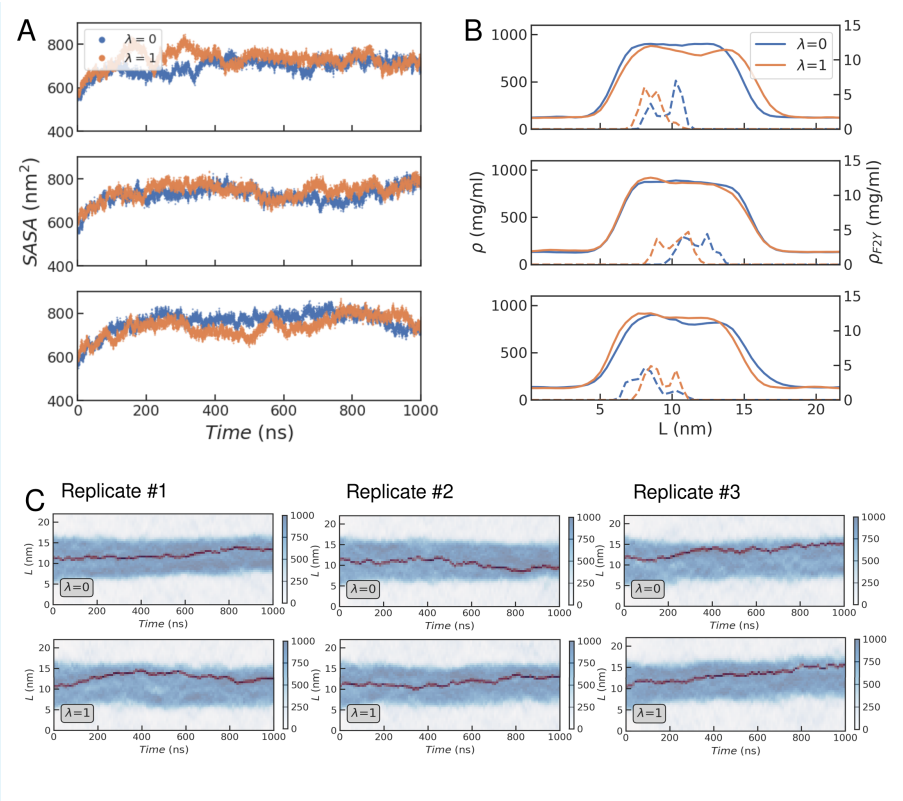
Appendix 2—figure S2. Work values for the forward and reverse alchemical transitions for different initial snapshots (bottom) and their distributions (top). The BAR estimator is shown in the legend. Each panel corresponds to a different solvent. (A) TIP3P water. (B) Replicates of GSY condensates. (C) Replicates of GSF condensates.



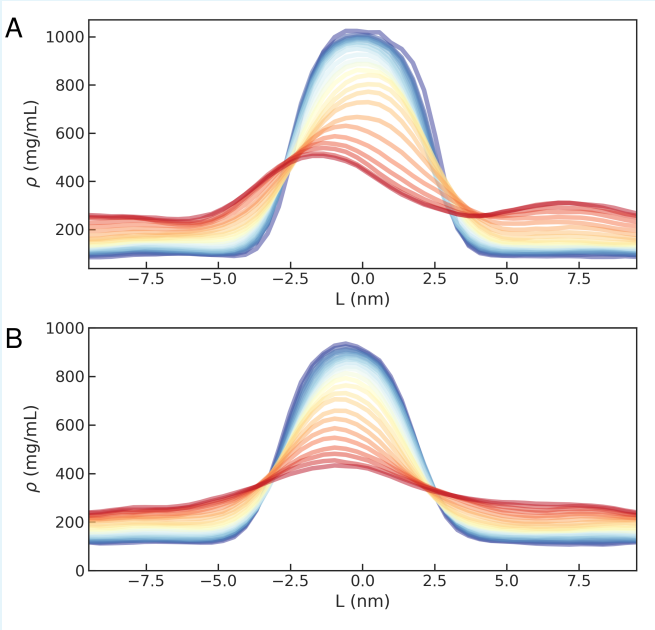
Appendix 2—figure S3. Results for the GSY condensate in the TIP4P-Ew water model. (A) Representative simulation box with a fully solvated GSY condensate in slab geometry including a GGXGG peptide (spheres) and the capped amino acid mixture (G: white, S: yellow and Y: green). (B) Time series for the solvent 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.



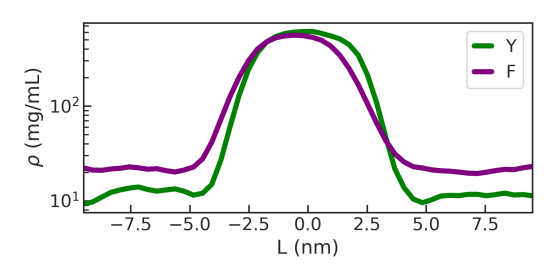
Appendix 2—figure S4. Results for the GSY condensate in the TIP4P-Ew water model. Work values and their distribution for forward and backward transitions within water (A) and the GSY condensate (B).



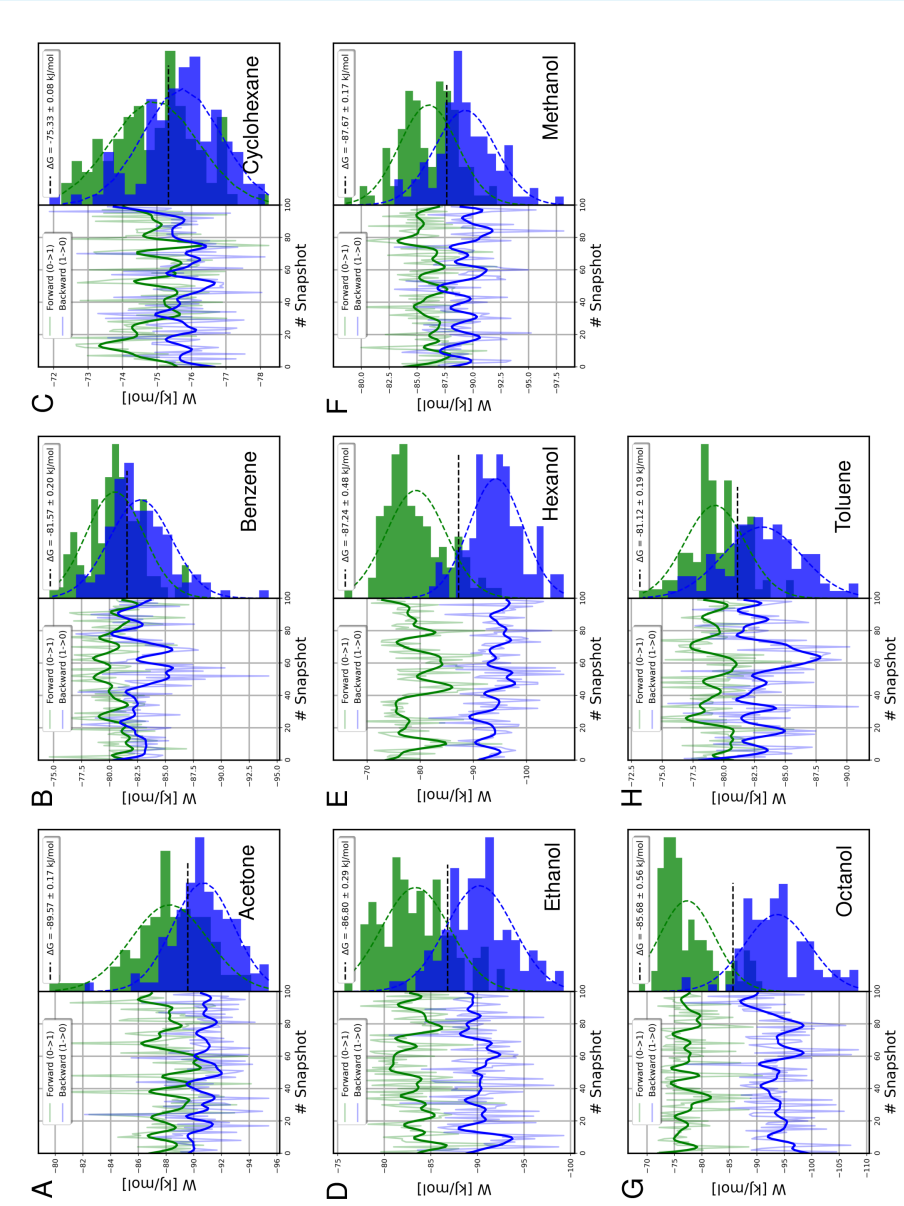
Appendix 2—figure S5. Replicates of the GGXGG peptide simulations in the GSF condensate. (A) Time series for the accessible surface area for different values of λ. (B) Density profiles of peptide (solid lines) and F/Y residue (right Y-axis) averaged over the simulation runs. (C) Time series for the density 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.