Biochemistry and Chemical Biology

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

**Reviewed Preprint**
**v1** • February 28, 2025
Not revised

**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



**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 **solid**, 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.104950.1.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 folded protein structures, 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 both from the possibility of tyrosine forming hydrogen bonds and the lower free energy of transfer of tyrosine into the condensate. The high level of hydration of polypeptides in dense phases drives this effect. Notably, as the dielectric constant of the surrounding environment approaches that of an apolar solvent, the trend reverses, with phenylalanine becoming the stronger sticker. These findings highlight the role of hydration in modulating aromatic interactions and provide a clear explanation for the crossover in sticker strength between tyrosine and phenylalanine in different media.

## Introduction

For decades, cellular organization was primarily understood through the lens of membranebound compartmentalization. However, it has become increasingly clear that membrane-less organelles – such as stress granules, nuclear speckles, and Cajal bodies– are also widespread within cells. [1] At least in some cases, these organelles form by phase separation of their components,[2] primarily disordered proteins and nucleic acids, which partition into dense and dilute phases. The condensates thus formed retain properties of liquids, like fusion, dripping and wetting.[3] This phenomenon has generated significant interest in the physical mechanisms governing the phase behavior of biomolecular mixtures.[4]

Many proteins undergoing phase separation share common characteristics, including deviations from typical compositions of folding proteins and simple sequences with multiple amino acid repeats. [5] 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. Borrowing language from the theory of associative polymers,[6] these units of sequence have been termed “stickers”, in the case of aromatics and positively charged residues, and “spacers”, for the polar residue repeats.[7–9] Spacers act as linkers that lend flexibility to the polypeptide mesh in the protein-dense phase. In contrast, stickers play a crucial 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.[10] Recent experiments have quantified the influence of different types of stickers in the polymer properties and phase behaviour. [8] This raises a crucial question: what is the fundamental origin of the relative strengths of different stickers?

This question has recently been addressed for the cationic amino acids lysine (Lys) and arginine (Arg).[11–15] Despite having the same net charge, these residues turn out not to be interchangeable. Mutagenesis experiments on LAF-1 have shown that substituting arginine with lysine completely suppresses phase separation. [12] This distinction is functionally relevant, as Arg to Lys substitutions affect speckle formation. [14] Additionally, experiments on the intrinsically disordered region (IDR) of Ddx4 indicate that LLPS is favoured by Arg relative to Lys.[11,12,16] 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.[11] These models are often parametrized using hydrophobicity scales that capture the “stickiness” of different amino acid residues.[17–20] However, by replacing the hydrophobicity scales in the coarse-grained model with interaction energies from amino acid contact matrices – derived from statistical analysis of the PDB– the correct trends were captured. [11] A key to the greater propensity for LLPS in the case of Arg may derive from the pseudo-aromaticity of this residue.[10,21,22]

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.[23] 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 double-mutant cycle that Tyr-Tyr and Phe-Phe pairs make nearly identical contributions to protein stability.[24] However, experimental evidence on various proteins suggests that Tyr is a stronger driver of condensation.[8,10,12,25] This is demonstrated by the reduced propensity to phase separate of Tyr-to-Phe mutants of FUS and LAF-1[10,12,25] and the opposite effect in Phe-to-Tyr mutants in hnRNPA1.[8,10] 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. [8]

The properties of Phe and Tyr may give important insights on the distinct behaviour as stickers in biomolecular condensates. Hydrophobicity scales typically rank Phe as the most hydrophobic residue,[20,26] consistent with the greater hydration free energy of Tyr relative to Phe[27,28] (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. [11] 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**).[29] We note that a different hydrophobicity scale based on peptides undergoing inverse temperature transitions (the Urry scale, [30] where Tyr is more hydrophobic) can account for the correct rank order in saturation concentration. [19] 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. [31,32] 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**). [33] This has led to the suggestion that tyrosine-tyrosine hydrogen bonds are stronger than those formed by tyrosine and water.[34]

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 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. [11,25] 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 molecular condensates. This peptide family has been characterized extensively in the past, both from experiment[35] and simulation.[36] We have built structures for the peptides using Ambertools[37] and used the Amber 99sb⋆ force field, including corrections for backbone torsions,[38] and TIP3P[39] water. 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. [40] For the simulations with different organic solvents, we used the GAFF parameters derived by Mobley and co-workers.[41]

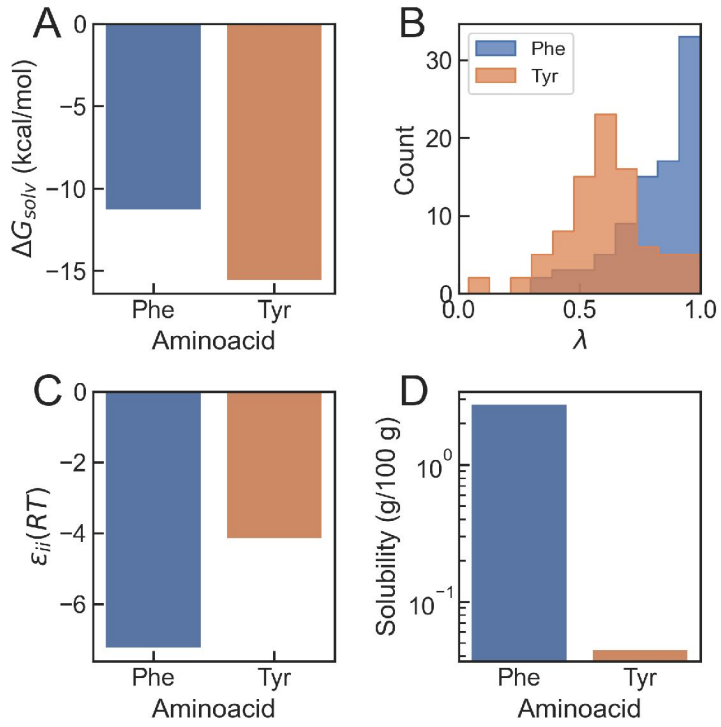


**Figure 1.**

Bibliographic values for different properties of phenylalanine and tyrosine. (A) Solvation free energy ($\Delta G_{solv}$).[28] (B) Probability distributions of min-maxed normalized hydropathy values λ from bibliographic hydrophobicity scales.[20] (C) Self-interaction energy ($\varepsilon_{ij}$) from the Miyazawa-Jernigan contact matrix. [29] (D) Solubility in water at 25°C.[33]

## 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.[42] Specifically, the thermodynamic cycle we construct involves four different states for the GGXGG peptide, with X=F (λ = 0) or X=Y (λ = 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 ( $\Delta G^{\text{F}}_{\text{Transfer}}$ and $\Delta G^{\text{Y}}_{\text{Transfer}}$ ), while the horizontal branches involve an alchemical transformation ($\Delta G_{\text{F}\to\text{Y}}(\text{H}_2\text{O})$ and $\Delta G_{\text{F}\to\text{Y}}(\text{Cond})$). Using simulations, we can easily calculate the free energies corresponding to the horizontal branches. This enables us to estimate the desired quantity as

$$\Delta\Delta G_{\text{Transfer}} = \Delta G^{\text{Y}}_{\text{Transfer}} - \Delta G^{\text{F}}_{\text{Transfer}} = \Delta G_{\text{F}\to\text{Y}}(\text{Cond}) - \Delta G_{\text{F}\to\text{Y}}(\text{H}_2\text{O}) \qquad (1)$$

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 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 model condensate, 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[43] and velocity rescaling[44] thermostats, respectively, to set the temperature at 298 K, and the Berendsen barostat [43] to set the pressure at 1 bar on the latter. Finally, we performed equilibrium simulations of 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.[45]

The duration of the equilibrium runs was 1-2 $\mu$s for the condensate simulations and 500 ns for all other solvents. 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 (λ = 0) or Tyr (λ = 1). For the alchemical transitions, we use a non-equilibrium switching method[46–48] often used for the calculation of differences in binding affinities[49,50] or free energy changes upon mutation.[51,52] We run short (50 ps) simulation trajectories where the λ-state was switched forward (λ = 0 → 1) or backward (λ = 1 → 0). We have run additional simulations of the solvents in the absence of the solute using the same protocols. To study interactions in the biomolecular condensates, we run simulations of equimolar GSY and GSF mixtures in a slab configuration, which we later sliced to keep only the dense phase. This was then briefly re-equilibrated as described above. Production simulations of the condensates were run for 1 $\mu$s.

For all simulations, we have used the Gromacs software package[53] (v.2024). To generate hybrid topologies for the different λ-states, we have used the PMX software[46–48] (v. 3.0).

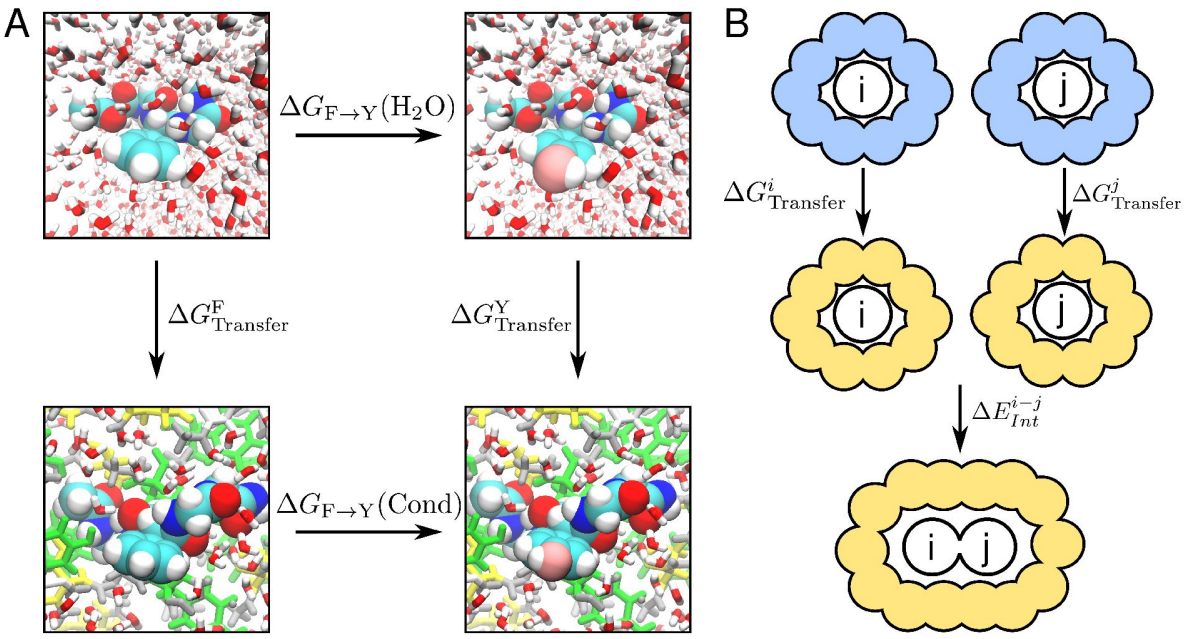


**Figure 2.**

(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.

### Analysis

We have analyzed the non-equilibrium switching simulations using the PMX package tools.[46–48] 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[54] (for more details, we refer to the PMX papers[46–48]). For our free energy estimates, we report bootstrap errors.

The simulations have been analyzed using a combination of analysis programs in the Gromacs suite[53] and in-house Python scripts that make extensive use of the MDTraj library. [55] Specifically, we used gmx dipoles to calculate the dielectric constant $\varepsilon$ for condensates and different solvents from the fluctuations in the dipole moment of the simulation box, **M**, which is defined as[56]

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

In this expression, $\varepsilon_0$ is the vacuum permittivity, $k_B$ is Boltzmann's constant and $V$ is the volume of the simulation box. Error analysis has been performed by blocking.

## 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 $\omega$B97XD functional[57] and Pople's 6-311++G(d,p) basis set[58] 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.[59] The interaction energy in a given solvent was calculated according to

$$\Delta E_{\text{Int}}^{ij} = E^{ij} - \left(E^i + E^j\right) \qquad (3)$$

where *i* and *j* specify the interacting monomers. We consider p-cresol to represent the sidechain of tyrosine (Y, hereafter), toluene to represent the sidechain 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[60] to calculate accurate solvation-free energies of the monomers in different solvents, using B3LYP/6-31+G(d,p) level of theory. [61,62] Thus, the transfer free energy from water to a given solvent *s* is calculated according to

$$\Delta G_{\text{Transfer}}^{i} = \Delta G^{i}(s) - \Delta G^{i}(\text{H}_2\text{O}) \qquad (4)$$

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 in biomolecular condensates. 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. [63–65] As a computationally efficient alternative, we use a simple model peptide with sequence GGXGG, with X being either F or Y, which is transferred from water into the dense mixtures of short peptides we have recently characterized.[40] Peptide condensates are useful model systems as they retain many properties of protein condensates but allow for much greater computational efficiency. [40,63,66,67]

Free energy estimates for the horizontal branches in the thermodynamic cycle in **Figure 2** were obtained from alchemical transformations, which we performed using a state-of-the-art non-equilibrium switching method[46–48] (see Methodology). The first step in this procedure involves running long equilibrium simulations at the initial λ-states (i.e. λ = 0 and λ = 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 must remain buried within the condensate for all the snapshots that we use as initial frames, so that we do not average the work in the dilute and dense phases. In **Figure 3A** we show the solvent accessible surface area (*SASA*) of all protein residues in a GSY condensate with the model peptide submerged. After an initial increase in *SASA*, at around 100 ns the values stabilize and remain flat for the remainder of the simulation, suggesting that equilibrium between the dense and dilute phases has been reached. We also show the density profiles for all the amino acid residues in the condensate and for the GGXGG peptide alone (**Fig. 3B**). For both λ values, we find that during the whole duration of the simulation, the GGXGG peptide remains buried within the condensate (see a representative snapshot in **Fig. 3C**). The protein density in the GSY condensate is comparable to the values reported before with a different force field (i.e. ∼ 1000 mg/mL protein and ∼200 mg/mL water),[40] and slightly larger than in full-length IDP condensates. [64]

After preparing our system, we selected initial configurations for non-equilibrium switching from evenly spaced snapshots of the λ = 0 and λ = 1 trajectories starting at 100 ns. We ran 200 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 inside the condensate and in water. For the GSY condensate, we find that $\Delta\Delta G_{\mathrm{Transfer}}$ is small and negative (−1.91 ± 0.53 kJ/mol), indicating that transfer into a peptide condensate is slightly more favourable for the Tyr-containing peptide than for the Phe variant. This result is consistent with mounting experimental evidence that points to a greater propensity to form condensates for Tyr rather than Phe in variants of proteins like FUS, LAF1 and hnRNPA.

## Similar protein-protein interactions in Tyr and Phe condensates

One possible explanation for this result is the formation of preferential sticker-sticker interactions between the GGYGG peptide and the GSY condensate. While the hydroxyl group in tyrosine in the pentapeptide may form hydrogen bonds with other tyrosines in GSY, these should be depleted for the GGFGG variant. Conversely, if the condensate were composed of Phe instead of Tyr (i.e. the GSF condensate), there would be no reason for a preference for GGYGG. Additionally, from hydrophobicity arguments alone, transfer to the GSF condensate should be favoured for Phe. To

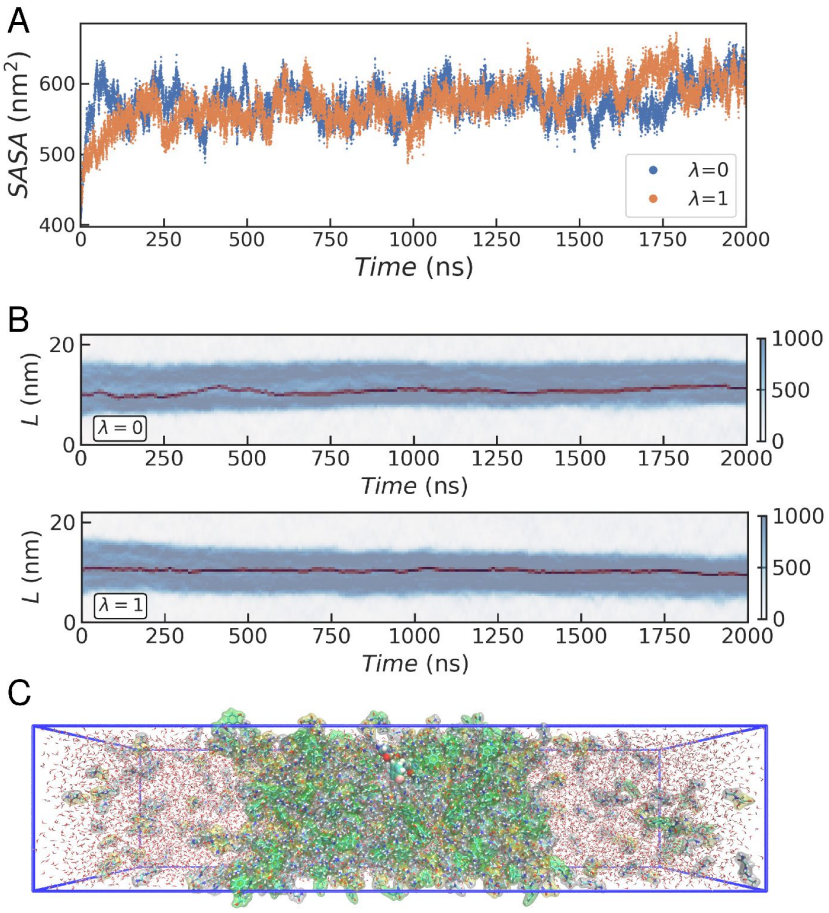


**Figure 3.**

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

test this hypothesis, we have run the same procedure for GSF instead of GSY. We show the time series data for *SASA* and protein density for the GSF simulations in the Supporting Information Figure S1. We find that $\Delta\Delta G_{\text{Transfer}}$ = −2.75 ± 0.58 kJ/mol, hence also favourable to Tyr and of magnitude comparable to the result for transfer to GSY.

To gain further insight into the dominant interactions within condensates formed by different stickers, we have run additional simulations of isolated dense phases for GSY and GSF condensates (see **Figure 4A**). First we look at the relative number of 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 **Fig. 4B**). This result recapitulates observations in atomistic simulations of full-length IDRs[64] and is consistent with the scaffold-client relationship between aromatic and polar residues.[68] To focus on $\pi - \pi$ interactions between pairs of aromatic side chains, we calculated the mean inter-side chain distances and the angles between vectors normal to the aromatic rings. **Figure 4C** shows their distribution, which is similar in both condensates. Even though aromatic $\pi - \pi$ stacking is feasible, it does not play a dominant role (see the regions with low inter-residue distance and angles close to 0 or $\pi$). We also examine $sp^2 - \pi$ contacts formed between aromatic and peptide bond groups, [69] which are much more common. This is consistent with previous simulation work[64,70,71] and also with a statistical analysis of experimental PDB structures which showed that these interactions are particularly prevalent in disordered regions of folded proteins.[69] Interestingly, also for these $sp^2 - \pi$ interactions, no significant differences were found between GSY and GSF condensates.

These findings suggest that the preference for Tyr over Phe to form condensates, as captured by $\Delta\Delta G_{\text{Transfer}}$, is not primarily driven by differences in protein-protein interaction patterns. Although interactions may be slightly more favourable for Tyr than for Phe, as suggested by atomistic potentials of mean force, [31,32] these residues have similar contact statistics both for sticker-sticker and for sticker-spacer contacts, and both exhibit a preference for $sp^2 - \pi$ rather than $\pi - \pi$ interactions within the condensate.

## Condensates as hydrated media

An alternative explanation could involve the solvent effects on the Phe/Tyr peptides. To investigate this, we begin by analysing the distribution of water molecules around the distinct side chains in the GSF and GSY dense phase simulations, revealing notable differences. **Figure 5A** presents the radial distribution function ($g(r)$) for water oxygen around the $C^{\zeta}$ in Phe and Tyr side chains. In the case of Tyr, a prominent peak at short distances suggests the presence of water molecules involved in hydrogen bonds, which are enthalpically favourable and entropically unfavourable due to the water structuring around the Tyr hydroxyl group. [25] These thermodynamic contributions are intertwined in the calculated $\Delta\Delta G_{\text{Transfer}}$ and will vary depending on whether transfer of the peptide from water occurs into a condensate, into the core of a folded protein or an altogether different solvent, capable or not of hydrogen bonding.

To investigate this dependence, we performed the same type of alchemical calculations using solvents such as octanol or cyclohexane –used in the past as models for the interior of folded proteins[33,73–76]– as well as shorter alcohols or acetone. The results of these calculations are shown in **Figure 5B**. In the case of a purely non-polar solvent like cyclohexane, the difference in free energy from Phe to Tyr is large and positive ($\Delta\Delta G_{\text{Transfer}}$ = 13.29 ± 0.20 kJ/mol), indicating that Phe is favoured. This is also the case for aromatic solvents like benzene and toluene. This result is consistent with the stronger interaction strength of Phe-Phe interactions in the Miyazawa and Jernigan potential[29] (see **Fig. 1C**), derived from contacts in the apolar cores of proteins where hydrophobicity dominates. The difference in free energy is smaller for alcohols and decreases gradually as we reduce the length of the aliphatic chain. For acetone, Tyr is slightly more favourable ($\Delta\Delta G_{\text{Transfer}}$ = −0.91 ± 0.25 kJ/mol). Clearly, the $\Delta\Delta G_{\text{Transfer}}$ values have a dependence on the dielectric constant ($\varepsilon$), which we have also calculated for the GSY and GSF condensates (see

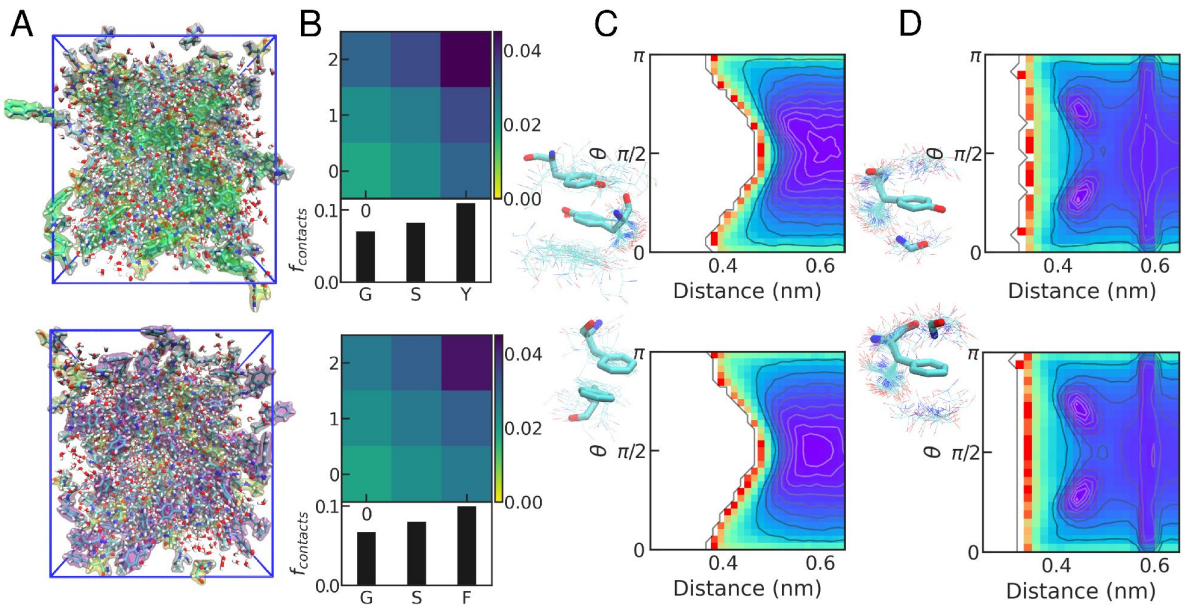


**Figure 4.**

(A) Representative snapshots from the simulations of dense phases. (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 $\theta$ between the vectors normal to the rings.[72] (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 $\theta$ 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.

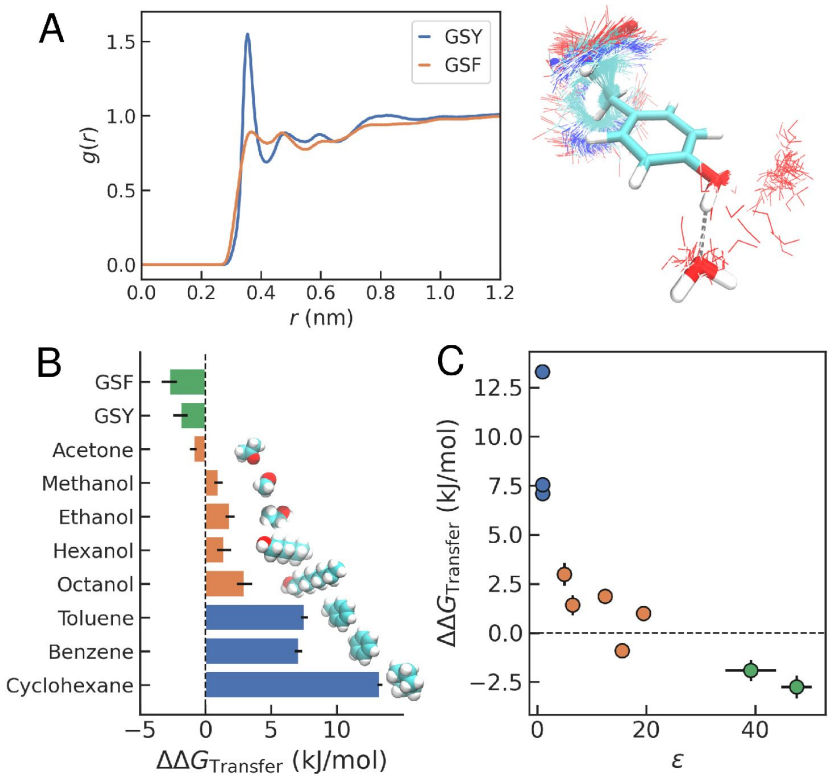


**Figure 5.**

(A) Radial distribution function for water oxygen around the $C^{\zeta}$ 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, $\varepsilon$.

**Figure 5C**). Our $\varepsilon$ values for the condensates (39 ± 5 for GSY and 47 ± 3 for GSF) are surprisingly close to that derived from experiments on Ddx condensates (45±13),[77] and appear at the rightmost part of the diagram. In summary, for non-polar solvents with low $\varepsilon$ Phe is preferred to Tyr, while for polar solvents, including the condensates, the trend is reversed. At intermediate values of $\varepsilon$, there is a crossover between a regime dominated by hydrophobicity and one favoured by hydrogen bonding by Tyr.

## Quantum calculations confirm a crossover in interaction strength

To characterize 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.[72,78] We consider different modes of interaction between the sidechains 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 *cross* and *antiparallel* configurations, whereas in the case of Y-Y, *H-bonded* and *antiparallel* are most favoured (see **Fig. 6A** 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. [79,80] Other structures, like parallel or T-shaped, are less stable for toluene, although in the benzene dimer, T-shaped conformers can be favored. [81] In our optimizations, 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, [72] 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 sidechains interacting with an amide group representing a peptide bond in the prevalent $sp^2$-$\pi$ interactions[69] (see **Fig. 6C**). 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.

To explore the environmental dependence in contact formation for amino acid pairs, we have calculated values of the interaction energies $(\Delta E_{\mathrm{Int}}^{ij})$ considering a set of protic and non-protic solvents (see **Fig. 6A**). 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 $\pi$-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 $\pi$-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, [29,82] 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 $\Delta E_{\mathrm{Int}}^{ij}$ (see **Figure 2B**). Hence, for a given solvent, the total energy can be calculated from the sum of these two contributions as

$$\Delta E_{\mathrm{Contact}}^{ij} = \Delta E_{\mathrm{Int}}^{ij} + \Delta G_{\mathrm{Transfer}}^{i} + \Delta G_{\mathrm{Transfer}}^{j} \qquad (5)$$

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.[60] Notice that when the solvent is a low dielectric non-protic solvent such as cyclohexane, $\Delta G_{\mathrm{Transfer}}^{i}$ recapitulates the results from a standard hydrophobicity scale.[26]

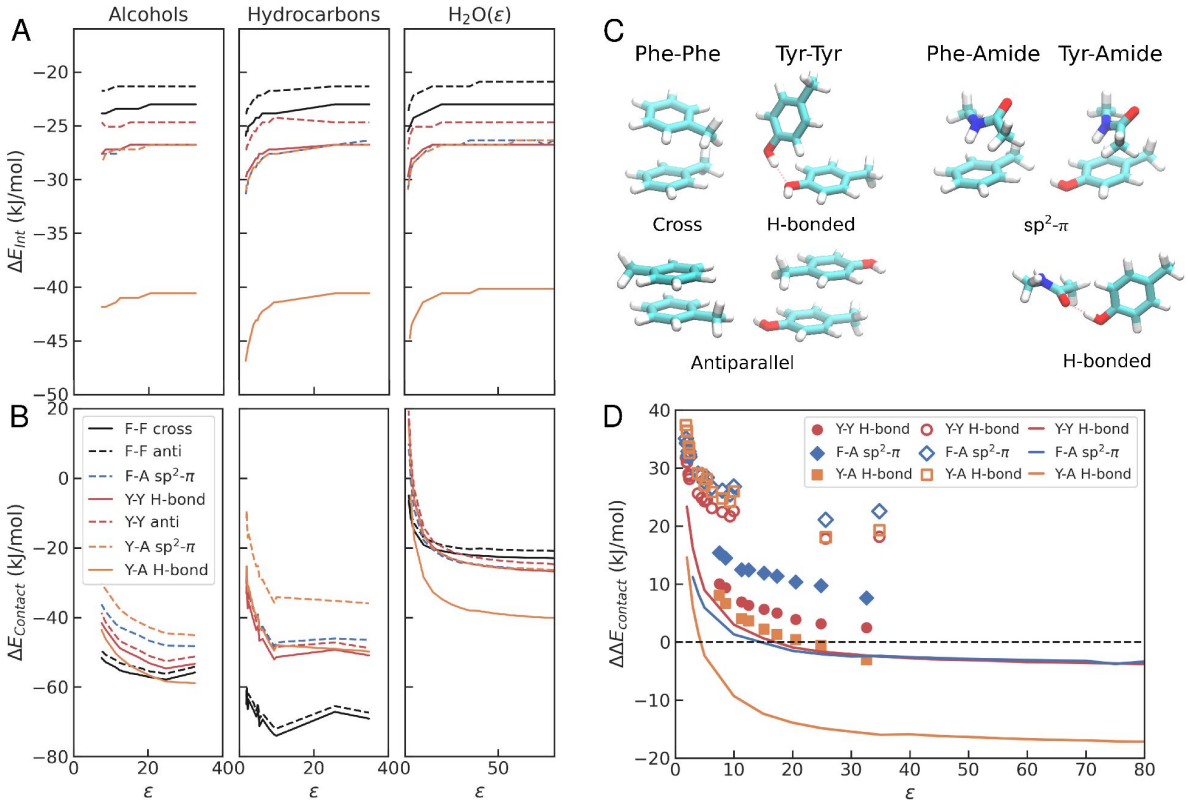


**Figure 6.**

(A) $\Delta E^{ij}_{Int}$ 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.

When both contributions are considered, the picture changes dramatically (see **Fig. 6B**). 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 $\varepsilon$ and F-F pairs at low $\varepsilon$. 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 **Fig. 6C**). Also, we find a similar tendency in this case, but now the crossover is obtained at slightly lower values of $\varepsilon$. In **Figure 6D**, 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[27,28] and stronger hydrophobicity of Phe[20,26] and the greater contact energies assigned to Phe-Phe interactions in proteins[29] with the higher sticker strength of Tyr in biomolecular condensates. [8,10,12,25] In the past, the different propensities of positively charged stickers to form condensates had been explained from the fundamental properties of Lys and Arg.[10,11,21,22] However, an explanation for the different strengths of Phe and Tyr remained elusive. [11] 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.[83] This type of environment can be modelled using aliphatic solvents, an approach followed extensively in the past.[33,73–76] In this low-dielectric regime, Phe-Phe interactions are the most favourable, as evidenced by both our classical transfer free energies (**Fig. 5C**) and the estimated quantum $\Delta\Delta E^{ij}_{\text{Contact}}$ values (**Fig. 6**). 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[29] and are known to be dominated by hydrophobicity effects.[84] At the other extreme, in environments with high dielectric constants such as aqueous or polar solvent conditions, Tyr-Tyr hydrogen bonds become more favourable. Biomolecular condensates, due to their high water content (~ 200 mg/mL in our simulations, but up to ~600 mg/mL in experiments[64]), are best modelled as a solvent with an intermediate dielectric. Hence, they operate in the second regime, although the influence of the dielectric is subtle, due to the possibility of crossover as a function of the dielectric.

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 influence of multivalency in protein IDRs.[7] 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-$\pi$ interactions that have been characterized by others.[11,16,49,72,85] Also, results may vary somewhat when considering different classical force fields. For example, our condensates are more protein-dense than found in experiments, possibly due to the propensity of the TIP3P water model to induce compact states.[86] However 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.

## 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.

## Additional files

***Supporting Information***

# References


(1) Banani S. F., Lee H. O., Hyman A. A., Rosen M. K. 2017) **Biomolecular condensates: organizers of cellular biochemistry** *Nat. Rev. Mol. Cell. Biol* **18**:285–298

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

(3) Brangwynne C. P., Eckmann C. R., Courson D. S., Rybarska A., Hoege C., Gharakhani J., Jülicher F., Hyman A. A. 2009) **Germline P granules are liquid droplets that localize by controlled dissolution/condensation** *Science* **324**:1729–1732

(4) Choi J.-M., Holehouse A. S., Pappu R. V. 2020) **Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions** *Annu. Rev. Biophys* **49**:107–133

(5) Martin E. W., Mittag T. 2018) **Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions** *Biochemistry* **57**:2478–2487

(6) Semenov A. N., Rubinstein M. 1998) **Thermoreversible gelation in solutions of associative polymers. 1. Statics** *Macromolecules* **31**:1373–1385

(7) Martin E. W., Holehouse A. S., Peran I., Farag M., Incicco J. J., Bremer A., Grace C. R., Soranno A., Pappu R. V., Mittag T. 2020) **Valence and patterning of aromatic residues determine the phase behavior of prion-like domains** *Science* **367**:694–699

(8) Bremer A., Farag M., Borcherds W. M., Peran I., Martin E. W., Pappu R. V., Mittag T. 2022) **Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains** *Nat. Chem* :196–207

(9) Mittag T., Pappu R. V. 2022) **A conceptual framework for understanding phase separation and addressing open questions and challenges** *Mol. Cell* **82**:2201–2214

(10) Wang J., Choi J.-M., Holehouse A. S., Lee H. O., Zhang X., Jahnel M., Maha-rana S., Lemaitre R., Pozniakovsky A., Drechsel D., Poser I., Pappu R. V., Alberti S., Hyman A. A. 2018) **A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins** *Cell* **174**:688–699

(11) Das S., Lin Y.-H., Vernon R. M., Forman-Kay J. D., Chan H. S. 2020) **Comparative roles of charge, π, and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins** *Proc. Natl. Acad. Sci. U.S.A* **117**:28795–28805

(12) Schuster B. S., Dignon G. L., Tang W. S., Kelley F. M., Ranganath A. K., Jahnke C. N., Simpkins A. G., Regy R. M., Hammer D. A., Good M. C., Mittal J. 2020) **Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior** *Proc. Natl. Acad. Sci. U.S.A* **117**:11421–11431

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

(14) Greig J. A., Nguyen T. A., Lee M., Holehouse A. S., Posey A. E., Pappu R. V., Jedd G. 2020) **Arginine-enriched mixed-charge domains provide cohesion for nuclear speckle condensation** *Mol. Cell* **77**:1237–1250

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

(16) Brady J. P., Farber P. J., Sekhar A., Lin Y.-H., Huang R., Bah A., Nott T. J., Chan H. S., Baldwin A. J., Forman-Kay J. D. 2017) **others Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation** *Proc. Natl. Acad. Sci. U.S.A* **114**:E8194–E8203

(17) Dignon G. L., Zheng W., Kim Y. C., Best R. B., Mittal J. 2018) **Sequence determinants of protein phase behavior from a coarse-grained model** *PLoS Comput. Biol* **14**:1–23

(18) Dignon G. L., Zheng W., Best R. B., Kim Y. C., Mittal J. 2018) **Relation between single-molecule properties and phase behavior of intrinsically disordered proteins** *Proc. Natl. Acad. Sci. U.S.A* **115**:9929–9934

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

(20) Tesei G., Schulze T. K., Crehuet R., Lindorff-Larsen K. 2021) **Accurate model of liquidliquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties** *Proc. Natl. Acad. Sci. U. S. A* **118**:e2111696118

(21) 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

(22) Hong Y., Najafi S., Casey T., Shea J.-E., Han S.-I., Hwang D. S. 2022) **Hydrophobicity of arginine leads to reentrant liquid-liquid phase separation behaviors of arginine-rich proteins** *Nat. Commun* **13**:7326

(23) Maraldo A., Rnjak-Kovacina J., Marquis C. 2024) **Tyrosine–a structural glue for hierarchi-cal protein assembly** *Trends Biochem. Sci*

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

(25) Lin Y., Currie S. L., Rosen M. K. 2017) **Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs** *J. Biol. Chem* **292**:19110–19120

(26) Kyte J., Doolittle R. F. 1982) **A simple method for displaying the hydropathic character of a protein** *J. Mol. Biol* **157**:105–132

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

(28) Chang J., Lenhoff A. M., Sandler S. I. 2007) **Solvation free energy of amino acids and side-chain analogues** *J. Phys. Chem. B* **111**:2098–2106

(29) Miyazawa S., Jernigan R. L. 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

(30) Urry D. W., Gowda D. C., Parker T. M., Luan C.-H., Reid M. C., Harris C. M., Pattanaik A., Harris R. D. 1992) **Hydrophobicity scale for proteins based on inverse temperature transitions** *Biopolymers* **32**:1243–1250

(31) Chelli R., Gervasio F. L., Procacci P., Schettino V. 2002) **Stacking and T-shape competition in aromaticaromatic amino acid interactions** *J. Am. Chem. Soc* **124**:6133–6143

(32) Joseph J. A., Reinhardt A., Aguirre A., Chew P. Y., Russell K. O., Espinosa J. R., 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

(33) 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

(34) Pace C. N., Horn G., Hebert E. J., Bechert J., Shaw K., Urbanikova L., Scholtz J. M., Sevcik J. 2001) **Tyrosine hydrogen bonds make a large contribution to protein stability** *J. Mol. Biol* **312**:393–404

(35) Plaxco K. W., Morton C. J., Grimshaw S. B., Jones J. A., Pitkeathly M., Campbell I. D., Dobson C. M. 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

(36) Workman R. J., Pettitt B. M. 2021) **Thermodynamic Compensation in Peptides Following Liquid–Liquid Phase Separation** *J. Phys. Chem. B* **125**:6431–6439

(37) Salomon-Ferrer R., Case D. A., Walker R. C. 2013) **An overview of the Amber biomolecular simulation package** *Wiley Interdiscip. Rev. Comput. Mol. Sci* **3**:198–210

(38) Lindorff-Larsen K., Piana S., Palmo K., Maragakis P., Klepeis J. L., Dror R. O., Shaw D. E. 2010) **Improved side-chain torsion potentials for the Amber ff99SB protein force field** *Proteins* **78**:1950–1958

(39) Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W., Klein M. L. 1983) **Comparison of Simple Potential Functions for Simulating Liquid Water** *J. Chem. Phys* **79**:926–935

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

(41) Bannan C. C., Calabró G., Kyu D. Y., Mobley D. L. 2016) **Calculating Partition Coefficients of Small Molecules in Octanol/Water and Cyclohexane/Water** *J. Chem. Theory Comput* **12**:4015–4024

(42) Mey A. S. J. S., Allen B. K., Bruce McDonald H. E., Chodera J. D., Hahn D. F., Kuhn M., Michel J., Mobley D. L., Naden L. N., Prasad S., Rizzi A., Scheen J., Shirts M. R., Tresadern G., Xu H. 2020) **Best Practices for Alchemical Free Energy Calculations [Article v1.0]** *Living J. Comp. Mol. Sci* **2**:18378

(43) Berendsen H. J., Postma J. v., Van Gunsteren W. F., DiNola A., Haak J. R. 1984) **Molecular dynamics with coupling to an external bath** *J. Chem. Phys* **81**:3684–3690

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

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

(46) Seeliger D., de Groot B. L. 2010) **Protein Thermostability Calculations Using Alchemical Free Energy Simulations** *Biophys. J* **98**:2309–2316

(47) Gapsys V., Michielssens S., Seeliger D., de Groot B. L. 2015) **pmx: Automated protein structure and topology generation for alchemical perturbations** *J. Comput. Chem* **36**:348–354

(48) Aldeghi M., de Groot B. L., Gapsys V., Sikosek T. 2019) **Accurate Calculation of Free Energy Changes upon Amino Acid Mutation** *Computational Methods in Protein Evolution* New York, NY: Springer New York :19–47

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

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

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

(52) Martinez-Martin I., Crousilles A., Herrero-Galan E., Velazquez-Carreras D., Mortensen S. A., Ochoa J. P., Garcia-Pavia P., de Sancho D., Wilmanns M., Alegre-Cebollada J. 2023) **Structural basis of domain destabilization by dilated-cardiomyopathy-causing missense mutations in titin** *Biophys. J* **122**:172a

(53) Abraham M. J., Murtola T., Schulz R., Páll S., Smith J. C., Hess B., Lindahl E. 2015) **GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers** *SoftwareX* **1-2**:19–25

(54) Bennett C. H. 1976) **Efficient estimation of free energy differences from Monte Carlo data** *J. Comput. Phys* **22**:245–268

(55) McGibbon R. T., Beauchamp K. A., Harrigan M. P., Klein C., Swails J. M., ndez C. X., Schwantes C. R., Wang L.-P., Lane T. J., Pande V. S. 2015) **MD-Traj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories** *Biophys. J* **109**:1528–1532

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

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

(58) Hehre W. J., Ditchfield K., Pople J. A. 1972) **Self-consistent molecular orbital methods. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules** *J. Chem. Phys* **56**:2257–2261

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

(60) Marenich A. V., Cramer C. J., Truhlar D. G. 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

(61) Becke A. D. 1993) **A new mixing of Hartree-Fock and local density-functional theories** *J. Chem. Phys* **98**:1372–1377

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

(63) 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

(64) Zheng W., Dignon G. L., Jovic N., Xu X., Regy R. M., Fawzi N. L., Kim Y. C., Best R. B., Mittal J. 2020) **Molecular Details of Protein Condensates Probed by Microsecond Long Atomistic Simulations** *J. Phys. Chem. B* **124**:11671–11679

(65) Galvanetto N., Ivanović M. T., Chowdhury A., Sottini A., Nüesch M. F., Nettels D., Best R. B., Schuler B. 2023) **Extreme dynamics in a biomolecular condensate** *Nature* **619**:876–883

(66) 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

(67) Brown W. H., Potoyan D. A. 2024) **Phase separation of multicomponent peptide mixtures into dehydrated clusters with hydrophilic cores** *Biophys. J* **123**:349–360

(68) Dignon G. L., Best R. B., Mittal J. 2020) **Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties** *Annu. Rev. Phys. Chem* **71**:53–75

(69) Vernon R. M., Chong P. A., Tsang B., Kim T. H., Bah A., Farber P., Lin H., Forman-Kay J. D. 2018) **Pi-Pi contacts are an overlooked protein feature relevant to phase separation** *eLife* **7**:e31486

(70) Murthy A. C., Tang W. S., Jovic N., Janke A. M., Seo D. H., Perdikari T. M., Mittal J., Fawzi N. L. 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

(71) Rekhi S., Garcia C. G., Barai M., Rizuan A., Schuster B. S., Kiick K. L., Mit-tal J. 2024) **Expanding the molecular language of protein liquid–liquid phase separation** *Nat. Chem* :1–12

(72) 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*

(73) Kauzmann W. 1959) **Adv. Protein Chem** Elsevier :1–63

(74) 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

(75) Wimley W. C., Creamer T. P., White S. H. 1996) **Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides** *Biochemistry* **35**:5109–5124

(76) Pace C. N., Fu H., Fryar K. L., Landua J., Trevino S. R., Shirley B. A., Hendricks M. M., Iimura S., Gajiwala K., Scholtz J. M., Grimsley G. R. 2011) **Contribution of Hydrophobic Interactions to Protein Stability** *J.Mol. Biol* **408**:514–528

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

(78) Carter-Fenk K., Liu M., Pujal L., Loipersberger M., Tsanai M., Vernon R. M., Forman-Kay J. D., 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

(79) 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**

(80) Rogers D. M., Hirst J. D., Lee E. P., Wright T. G. 2006) **Ab initio study of the toluene dimer** *Chem. Phys. Lett* **427**:410–413

(81) Chipot C., Jaffe R., Maigret B., Pearlman D. A., Kollman P. A. 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

(82) Thomas P. D., Dill K. A. 1996) **An iterative method for extracting energy-like quantities from protein structures** *Proc. Natl. Acad. Sci. U.S.A* **93**:11628–11633

(83) Pitera J. W., Falta M., van Gunsteren W. F. 2001) **Dielectric properties of proteins from simulation: the effects of solvent, ligands, pH, and temperature** *Biophys. J* **80**:2546–2555

(84) Pokarowski P., Kloczkowski A., Jernigan R. L., Kothari N. S., Pokarowska M., Kolinski A. 2005) **Inferring ideal amino acid interaction forms from statistical protein contact potentials** *Proteins* **59**:49–57

(85) Hazra M. K., 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

(86) Piana S., Donchev A. G., Robustelli P., Shaw D. E. 2015) **Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States** *J. Phys. Chem. B* **119**:5113–5123

## 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

**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

**For correspondence:** xabier.lopez@ehu.eus

## Editors

Reviewing Editor
**Rosana Collepardo**
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 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.

(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.

(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.

(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 ε = 45{plus minus}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 (ε = 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.

(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.

(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 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.

(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.

https://doi.org/10.7554/eLife.104950.1.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.

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? 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. 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? 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.

https://doi.org/10.7554/eLife.104950.1.sa1

**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.

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?

https://doi.org/10.7554/eLife.104950.1.sa0