Phase behaviour of hnRNPA1 low-complexity domain mutants described by different sequence-dependent models

Alejandro Feito,[1] Ignacio Sanchez-Burgos,[2] Ignacio Tejero,[1] Eduardo Sanz,[1] Antonio Rey,[1] Rosana Collepardo-Guevara,[3,4] Andres R. Tejedor,[1,3,a)] and Jorge R. Espinosa[1,3,b)]

[1)] *Department of Physical-Chemistry Universidad Complutense de Madrid Av. Complutense s/n, Madrid 28040, Spain*

[2)] *Maxwell Centre, Cavendish Laboratory, Department of Physics, University of Cambridge, J J Thomson Avenue, Cambridge CB3 0HE, United Kingdom.*

[3)] *Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK*

[4)] *Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK*

(Dated: 28 August 2024)

Intracellular liquid-liquid phase separation (LLPS) of proteins and nucleic acids represents a fundamental mechanism by which cells compartmentalise their material and fulfill biological function. Molecular simulations have proved to be invaluable for gaining microscopic insights of the physicochemical processes underlying this phenomenon. Here, we systematically compare several state-of-the-art sequence-dependent residue-resolution models to test their performance in reproducing the phase diagram and the condensate material properties of seven different variants of the low-complexity domain (LCD) of hnRNPA1 protein (A1-LCD)—a protein related to the pathological fibrillisation of stress granules. By testing the performance of 5 different coarse-grained force fields—HPS, HPS-cation-π, CALVADOS2, Mpipi, and Mpipi-Recharged—in describing the condensate saturation concentration, critical temperature for phase-separation, and condensate viscosity of the different variants, we unveil the key intermolecular contacts controlling the phase behaviour of this protein and its mutants. We find that the dominant interactions enabling LLPS of A1-LCD condensates are mostly cation-π contacts such as arginine-tyrosine and arginine-phenylalanine as well as π-π interactions mediated by tyrosines and phenylalanines. Interestingly, we also find a correlation between condensate stability and viscoelasticity for A1-LCD mutants showing how higher critical temperatures for phase-separation imply larger condensate viscosities. Overall, our results serve as a benchmark for different coarse-grained models to study LLPS of protein prion-like domains, and further establish a direct link between condensate phase behaviour and individual intermolecular interactions.

Keywords: Liquid-liquid phase separation, intrinsically disordered proteins, molecular dynamics simulations, coarse-grained models

I. INTRODUCTION

The formation of intracellular membraneless organelles (MLOs) represents an essential mechanism enabling the spatiotemporal organisation and functional regulation of the cell material[1–3]. MLOs—comprised of a wide variety of biomolecules including intrinsincally disordered proteins (IDPs), multi-domain proteins, and DNA or RNA strands—are formed through the spontaneous demixing of biomolecules by means of liquid-liquid phase separation (LLPS). These membraneless compartments are fundamental for accomplishing diverse biological functions such as cell signaling[4–6], buffering protein concentrations[7,8], compartmentalisation[7–10], genome silencing[11,12], or noise buffering[13–15], among others[16,17]. Archetypal proteins enabling the formation of phase-separated condensates—such as the widely studied *fused in sarcoma* (FUS)[18–20], the TAR DNA-binding protein of 43kDa (TDP-43)[20–22], or the heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1)[23,24]—are normally characterised by presenting a low-complexity domain (LCD) in their sequence. Such type of domains prevail intrinsically disordered, and they can establish multiple transient intermolecular interactions which have been consistently proved to be crucial in promoting condensate formation via LLPS[25–28]. Moreover, beyond condensate formation, LCDs can also trigger the progressive rigidification of liquid-like biomolecular condensates into solid-like states[29–31], which has been related to the onset of several neurodegenerative pathologies such as amyotrophic lateral sclerosis (ALS)[32–34], frontotemporal dementia (FTD)[35,36], Alzheimer[37,38], or even some types of cancer[17,39,40]. Thus, unravelling the conditions, factors, and interactions regulating protein self-assembly, and their subsequent potential liquid-to-solid transition into aberrant solid states represents an urgent challenge in cell biology[25,34].

The precise aspects determining the phase separation behaviour of biomolecules is a complex problem that requires a combination of both experimental and theoretical approaches[2,41,42]. Specifically, computational methods can crucially contribute to exploring the underlying mechanisms and parameters inducing LLPS due to the

[a)] Electronic mail: at949@cam.ac.uk

[b)] Electronic mail: jorgerene@ucm.es

fact that they provide detailed information about the forces that bridge biomolecules together[41–43]. Computational efforts can widely vary in resolution and performance spanning from atomistic simulations[44–47]—where only a few protein replicas or protein segments can be studied—to low-resolution models[48–51]—where the interactions are simplified to improve computational efficiency and hundreds of molecules can be included in the system. The simulation approach that lies at the interface between optimising computational performance, exploring substantial size and timescales, and considering the chemical make-up of molecules, utilizes coarse-grained (CG) models with sequence-dependent interactions at one-bead per amino acid resolution[52]. These CG models have gained much attention in the past few years and many of them have been proposed[41,52–61], mostly inspired by the pioneering work of Dignon *et al.*[52] for studying IDPs' LLPS. These models have been outstanding in providing molecular explanations to the phase behaviour of many different proteins[62–68], as well as the impact of other biomolecules—e.g. RNA or DNA— in controlling condensate architecture, stability, or transport properties[47,69,70]. Through iterative refinement and validation against experimental benchmarks, CG models continue evolving, and can offer meaningful insights into the complex interplay between protein sequence, structure, dynamics, and the resulting collective phase behaviour[42,52].

Experimental measurements are one of the fundamental baselines to guide the development of sequence-dependent models for studying protein LLPS[52,53]. The initial parameterisations of CG models were mostly based on reproducing *in vitro* measurements of single-protein radius of gyration (R_g)[71–75] aiming to reproduce sets of experimental R_g values of IDPs[41,56,59,76]. Recently, the development of these models has drifted towards reproducing phase diagrams of proteins with near-quantitative accuracy[41,53,54,61,77]. Novel approaches for model parameterisation encompass machine-learning algorithms[41,55,57], bioinformatic analysis[53], atomistic simulations[53,61], and direct validation against *in vitro* data[53,61,77]. Therefore, experimental efforts providing temperature-dependent coexistence lines, single-protein radius of gyration, or viscosity values are enormously valuable, aside of their own interest, to assess the predictive capability of CG models[77,78].

In this work, we compare the performance of five different CG residue-resolution models in describing LCD protein phase behaviour, including the HPS[43], HPS-cation-π[56], CALVADOS2[41], Mpipi[79] and Mpipi-Recharged[61] models. More precisely, we focus on the phase equilibrium of several hnRNPA1-LCD (referred to as A1-LCD in what follows) mutants, and we compare modelling predictions against *in vitro* results[77,78], shedding light on the sensitivity of different CG models to capture sequence modifications. We establish a direct comparison between the experimental saturation concentration of the different mutants and the corresponding predicted values. Furthermore, we investigate condensate viscosity, which is not trivial to capture because despite being intimately related to the specific parameterisation of the intermolecular interactions, it is not a target property in model parameterization optimization. Once established which model most reliably describes the experimental behaviour of A1-LCD[77,78], we analyse the protein intermolecular contact maps and the most frequent residue-residue interactions within condensates for the studied mutants. We aim to elucidate the nature of the intermolecular forces controlling LLPS in A1-LCD due to its central role in the formation of stress granules, the so-called 'crucibles' in ALS and FTD neuropathologies[24,25,34]. Overall, our findings not only contribute to the refinement of CG modelling, but also provide valuable insights into the molecular mechanisms underlying A1-LCD phase separation and its link between sequence composition and condensate material properties.

II. PREDICTION OF A1-LCD COEXISTENCE LINES USING DIFFERENT SEQUENCE-DEPENDENT MODELS

The determination of the phase diagram establishes the thermodynamic conditions to investigate the formation and stability of biomolecular condensates[80,81]. Numerous experiments have consistently shown how aromatic residues represent key 'stickers' to sustain protein phase separation[23,77,78,82,83]. Therefore, to evaluate the performance of different models such as the HPS[52], HPS-cation-π[56], CALVADOS2[55], Mpipi[53] and Mpipi-Recharged[61] in reproducing condensate behaviour (see section SI in the Supplementary Material (SM) for technical details about the models), we consider the different mutants of A1-LCD reported in Ref.[78] that display multiple variations in the type of aromatic residues along their sequence (see Table S2 in SM for the sequences studied). These variants are the following: (1) WT+NLS which corresponds to the A1-LCD wild-type sequence, which features a nuclear localization signal[78], and contains 8 tyrosines (Y) and 12 phenylalanines (F); (2) allF with 19F; (3) allY with 19Y; (4) allW with 19 tryptophans (W); (5) YtoW with 7W and 12F; (6) FtoW with 7Y and 12W; and (7) W^- with 13W. Importantly, the mutations in all variants are performed on the aromatic residues maintaining the patterning and just changing the chemical identity of the residues[78].

We compute the temperature-density phase diagrams of the aforementioned sequences by means of Direct Coexistence (DC)[84,85] simulations (see Fig. 1A and section SII in the SM for further methodological details). In the DC method, the two coexisting phases are simulated by preparing periodically replicated slabs of the two phases (the condensed and the diluted phase) in the same simulation box. Once the system is equilibrated, a density profile along the long axis of the box can be extracted to compute the density (or concentration) of the two coexisting phases (Fig. 1A). Importantly, we note that

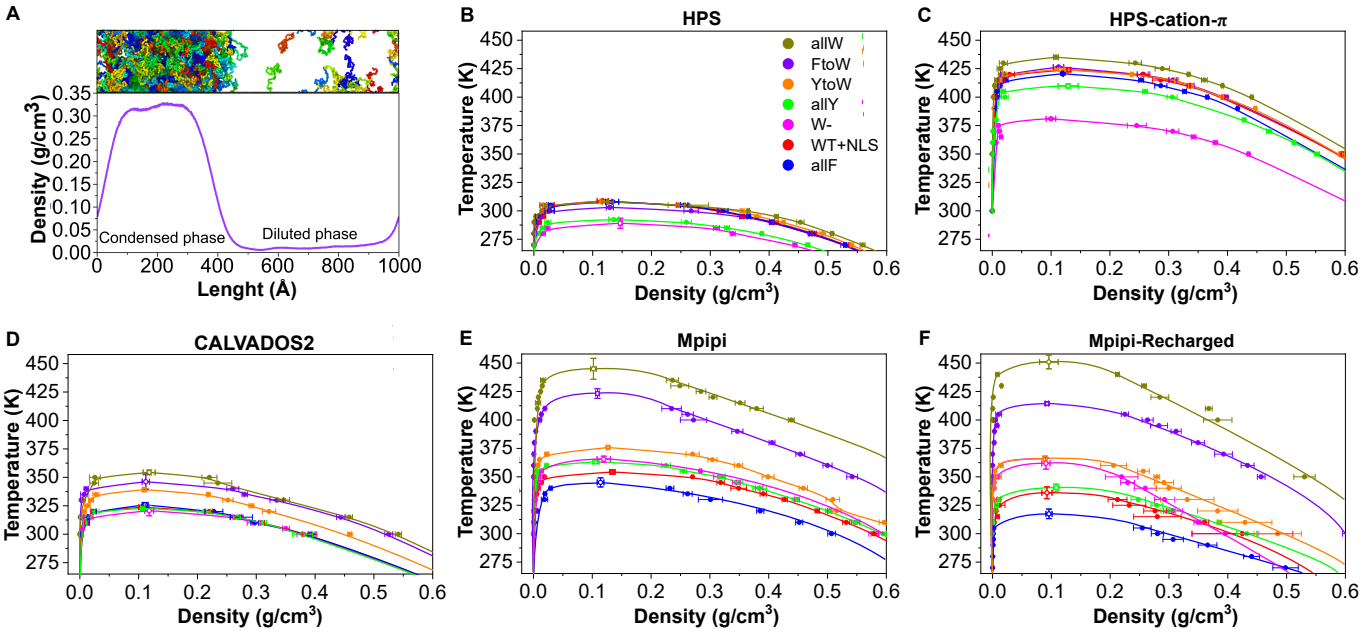

FIG. 1. (**A**) Snapshot of a DC simulation used to calculate the phase diagram and critical temperature of A1-LCD (WT+NLS) protein (Top). Different protein replicas are depicted by different colours. The average density profile along the long axis of the DC simulation to determine the coexistence densities is shown below. Temperature–density phase diagrams of A1-LCD variants calculated via DC simulations employing the HPS (**B**), HPS-cation-π (**C**), CALVADOS2 (**D**), Mpipi (**E**) and Mpipi-Recharged (**F**) models. Critical temperatures are represented by empty circles while filled circles depict coexistence densities obtained from DC simulations. The lines serve as a guide to the eye. The colour coding is preserved throughout all the panels for all the variants as indicated in the legend of panel B.

none of the 5 employed models includes explicit solvent, hence, effectively in our DC simulations the protein diluted phase is a vapour phase. The phase diagrams shown in Fig. 1B-F clearly illustrate the sensitivity of the different models to capture precise residue changes across the protein sequence. The Mpipi and Mpipi-Recharged (Fig. 1E and F, respectively) show the largest variation in the critical temperature (T_c) as a function of the aromatic residue identity (the T_c spans more than 100 K depending on the sequence). In contrast, the HPS model exhibits the lowest responsiveness to different mutations by only varying its T_c approximately 30 K from the lowest to the highest T_c variant (Fig. 1B).

Since the temperature–concentration phase diagram of three of these sequences (WT+NLS, allY, and allF) has been recently measured *in vitro*[77,78], we now establish a direct comparison between modelling and experiments so we can assess the ability of the models to predict A1-LCD phase behaviour. As depicted in Fig. 2A, the critical temperature of WT+NLS, allF, and allY variants predicted by the HPS model—initially devised to determine single-protein radius of gyration through short-ranged interactions based on residue contact statistics of folded proteins[76]—is consistently underestimated in all cases. If we plot the predicted critical temperature from simulations (orange symbols in Fig. 2F) *vs.* the corresponding experimental value (in which the perfect fit is represented by a black line of slope 1 and 0 intercept; Fig. 2F), we observe that the phase-separation trend due to the sequence specificity is not reproduced by the HPS model. On the other hand, the HPS-cation-π model[56]—a reparameterisation of the HPS model to better account for cation–π interactions—overestimates the critical temperature of the considered variants (Fig. 2B and 2F). In contrast, the CALVADOS2 model[55]—a further reparameterisation of the hydrophobicity scale of HPS-KR via a Bayesian parameter-learning procedure—provides simulation results that closely match the experimental T_c's and coexistence lines (Fig. 2C). Nevertheless, a closer look on this comparison for the CALVADOS2 (pink data in Fig. 2F) evidences that the model does not capture the trend followed by the experimental variants. The Mpipi[53]—especially designed for amino acid pairs involving π-π interactions—does not quantitatively capture the critical experimental temperature of these variants (Fig. 2D) showing a moderate overestimation of $\sim$20 K in the predicted T_c's. However, it predicts the correct sensitivity to changes in the amino acid sequence (cyan symbols in Fig. 2F). Importantly, while the π-π interactions are crucial for the phase behaviour of these variants, the Mpipi model slightly underestimates the contributions of electrostatic interactions due to the use of scaled charges[61]. In that respect, the Mpipi-Recharged—which improves charged amino acid interactions by applying specific non-symmetrical electrostatic interactions among charged residues through a Yukawa potential based on

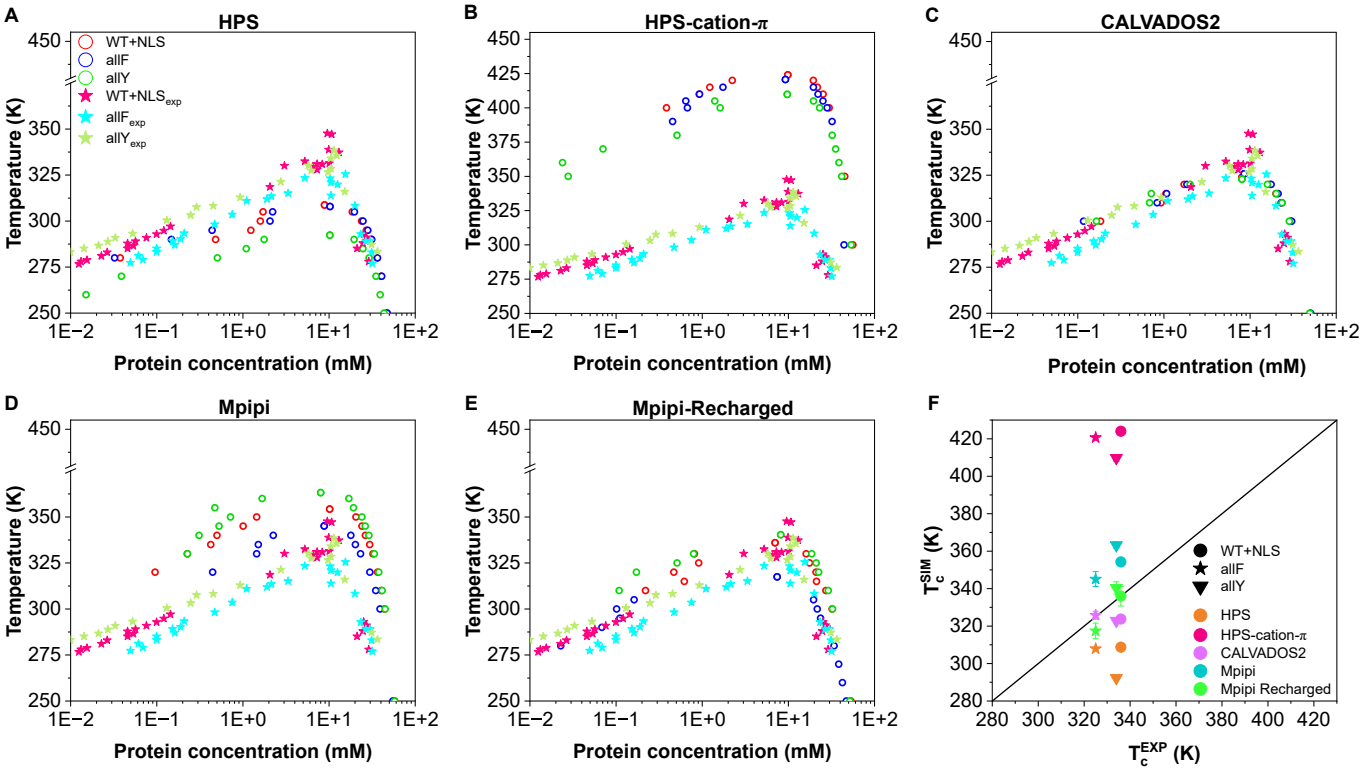

FIG. 2. Experimental (solid stars) *vs.* simulated (empty circles) phase diagrams for the WT+NLS, allF, and allY variants of A1-LCD using different models: HPS (**A**), HPS-cation-π (**B**), CALVADOS 2 (**C**), Mpipi (**D**), and Mpipi-Recharged (**E**). (**F**) Comparison of the calculated *vs.* experimental values of T_c for the aforementioned variants[78]. The black line depicts the perfect fit between experiments and simulations.

atomistic simulations with explicit water and ions[61]—shows the most accurate agreement with the experimental values capturing not only the trend but also the absolute value of the variants' critical temperature (green symbols in Fig. 2F). Overall, from this analysis, we conclude that CALVADOS2, Mpipi, and Mpipi-Recharged provide a very reasonable description of the condensate phase behaviour of A1-LCD (WT+NLS) and its two additional aromatic mutants.

III. COMPARISON OF THE LLPS SATURATION CONCENTRATION BETWEEN EXPERIMENTS AND SIMULATIONS

The concentration beyond which condensate formation starts (C_{sat}) is one of the critical quantities controlling at intracellular level when proteins and nucleic acids give rise to dynamic MLOs[7,86]. In that sense, *in vitro* studies are extremely useful to characterise such threshold, dictating the border between a single homogeneous phase and the emergence of phase separated condensates[77,87–89]. Most of these studies have focused on physiological conditions of salt, pH, and temperature[22,24,87,89–94]. However, recent studies have also reported how C_{sat} varies with temperature and the limit beyond which LLPS is not achievable[77,78]. The saturation concentration quantifies the ability of a given protein to phase separate, thus, the higher the value of C_{sat}, the lower propensity to form condensates[22,24,87,91]. Similarly, the critical temperature evaluated through DC simulations serves as a direct indicator of the ability of a protein to undergo LLPS[69]. Indeed, *in vitro* C_{sat}'s have been shown to inversely correlate with the critical temperature[77]. A higher critical temperature is directly related to a stronger phase separation driving force. Thus, we can correlate the critical temperature obtained from simulations to the experimental saturation concentration at a given constant temperature (e.g., 298 K) to obtain a qualitative overview of the models sensitivity for describing protein LLPS as a function of specific sequence mutations[61,69].

In Fig. 3, we show the simulated critical temperature (normalised by the critical temperature of WT+NLS) against the experimental saturation concentration[78] measured at 298 K for the different A1-LCD mutants. These results reveal the sensitivity of the studied models to describe the impact of mutations on A1-LCD condensate's stability (Fig. 3A-E). Notably, the HPS and HPS-cation-π models do not show a significant variation in the critical temperature as a function of the applied mutations (Fig. 3A and 3B, respectively) despite of the fact that C_{sat} spans over more than 2 orders of magnitude with the performed mutations[78]. In contrast, the Mpipi and

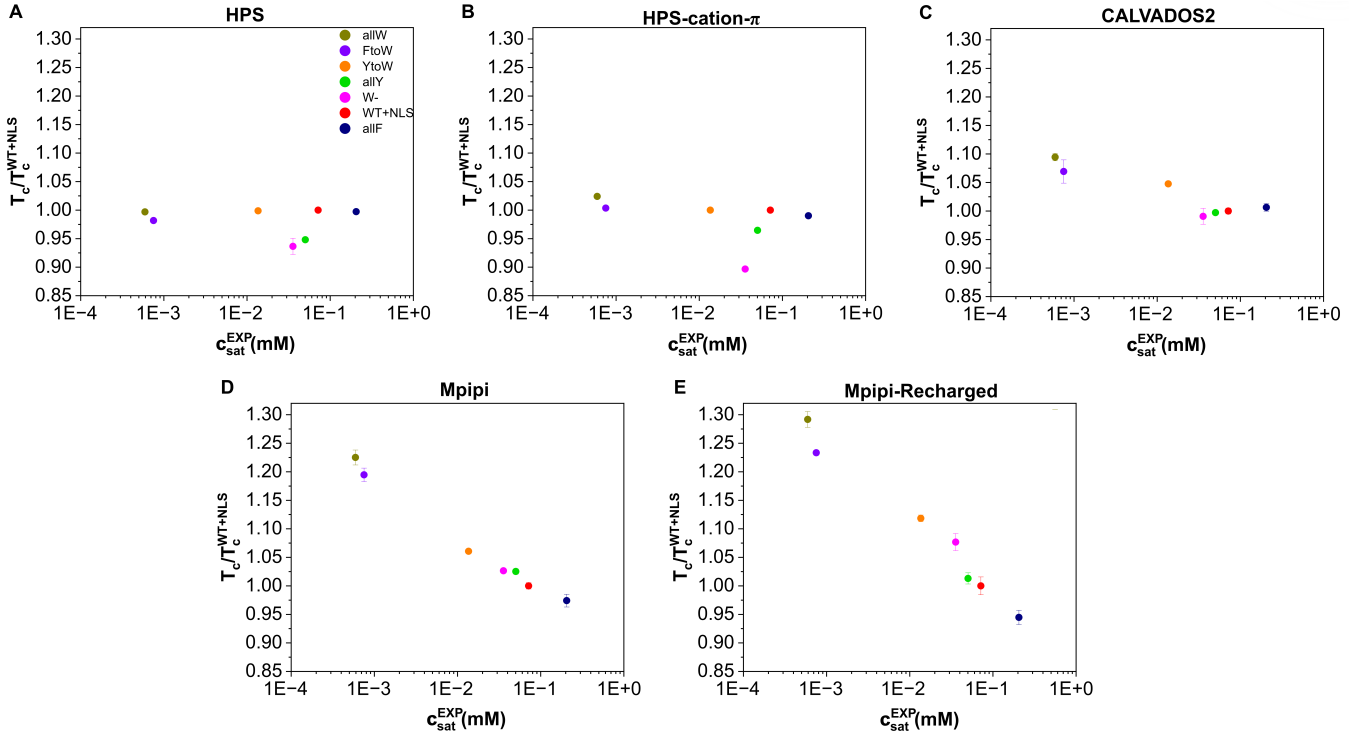

FIG. 3. Predicted critical temperature (normalised by the critical temperature of the WT+NLS sequence) for the different models HPS (**A**), HPS-cation-π (**B**), CALVADOS2 (**C**), Mpipi (**D**), and Mpipi-Recharged (**E**) *vs.* the experimental saturation concentration at 298 K reported in Ref.[78] for the different A1-LCD mutants.

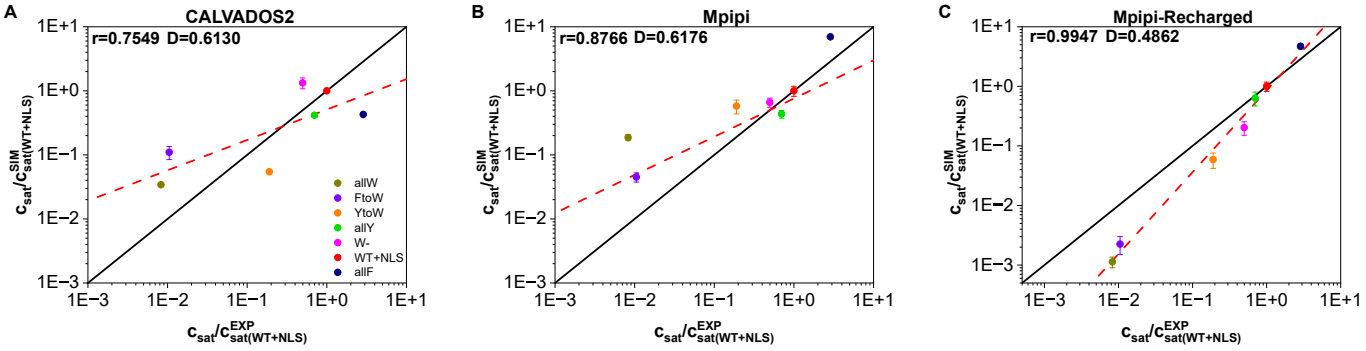

FIG. 4. Simulated *vs.* experimental saturation concentration (C_{sat}), normalised by the WT+NLS C_{sat}[78] for the different variants. The Pearson correlation coefficient (r) and the root mean square deviation from the experimental values (D) are displayed for each set of modelling data. The black lines indicate a perfect match between experimental and computational values, while the red dashed lines depict the linear regression for each set of data.

Mpipi-Recharged models exhibit a considerable dependence of the critical temperature with the protein sequence (Fig. 3D and 3E, respectively), which is fully compatible with the relative experimental propensity of the studied sequences to undergo LLPS. Finally, the CALVADOS2 shows an intermediate behaviour between HPS and Mpipi's models with a moderate decrease of T_c with C_{sat} (Fig. 3C). Although for a few sequences the change in C_{sat} does not correlate with T_c, as for those variants with higher saturation concentration, the CALVADOS2 yields an overall fair qualitative description of the experimental propensity for phase-separation.

In view of these results, we now consider the latter three models to further investigate A1-LCD mutants LLPS behaviour. We explicitly compute the saturation concentration at 298 K by extensively sampling the equilibrium protein concentration of the diluted phase in coexistence with the condensates using DC simulations. This quantity is computationally expensive—especially at low temperatures—due to the difficulty to sample low protein concentrations[68]. However, it enables a direct comparison of the predicted C_{sat} with the experimental

values reported in Ref.[78]. In Fig. 4, we plot the predicted saturation concentration (normalised by that of the WT+NLS sequence) for the CALVADOS2, Mpipi and Mpipi-Recharged models against the experimental C_{sat} (also normalised) of the different studied mutants. We quantify the deviation of the modelling predictions *vs.* the experimental values by computing the mean square root deviation of the normalised simulated C_{sat}'s with respect to the normalised experimental values (D, for further details see section SVII in the SM). Lower values of D imply lower deviation with respect to the perfect fit between simulations and experiments (diagonal black lines in Fig. 4). Moreover, we compute the Pearson correlation coefficient (r) to quantify the linearity of the modelling data sets, and to determine if, at least qualitatively, the models reproduce the experimental trend. We find that the Mpipi-Recharged provides the lowest deviation (D) with respect to experimental C_{sat}'s, and the highest Pearson correlation coefficient compared to both Mpipi and CALVADOS2 models. These two models, nevertheless, show an acceptable qualitative description of C_{sat} despite the fact that they moderately deviate from the experimental perfect fit. Taken together from this stringent test, we can conclude that all three models perform remarkably well, especially the Mpipi-Recharged, considering the huge difficulty for a coarse-grained model to quantitatively match experimental coexistence lines.

IV. CONDENSATE VISCOELASTIC BEHAVIOUR OF A1-LCD MUTANTS BY DIFFERENT MODELS

Condensate viscoelastic properties are strongly related to health and disease[34,95,96]. Whereas liquid-like states in condensates are associated to biological function, their progressive transition into solid and gel-like states has been linked to the onset of multiple neurological disorders[24]. Examples of protein condensates displaying hardening over time have been reported for RNA-binding proteins such as FUS[97,98], TDP-43[99], or hnRNPA1[100] among others[30]. Hence, it is crucial to determine whether condensates behave as a liquid, as a gel, or if they consistently increase their viscosity over time. Moreover, characterising which precise interactions and domains control their time-dependent transport properties[97,101–103] is central to understand the molecular onset between functional and dysfunctional behaviour.

Computationally, the generalised Green-Kubo relation[104,105] (see section SVI in the SM) is one of the most efficient methods for evaluating viscosity of protein condensates using CG models[31,69,106]. Calculating the shear stress relaxation function ($G(t)$) using the Green-Kubo approach provides information of both inter- and intra-molecular protein interactions, which can critically rely on the chosen CG model[69,102,106]. Since sequence-dependent models are often parameterised to reproduce experimental single-protein radius of gyration or phase diagrams[41,43,53,54,59], evaluating condensate viscosity serves as an excellent benchmark for testing their predictive capability of material properties. To validate the predicted viscosities, we employ the experimental *in vitro* measurements at $T \sim 298\ K$ from Ref.[78]. We only use the CALVADOS2, Mpipi and Mpipi-Recharged models to calculate the shear stress relaxation function of the different A1-LCD variants under bulk conditions at 298K (see Fig. 5A and simulation details in section SII of the SM), since both HPS and HPS-Cation-π failed in describing the LLPS sensitivity of A1-LCD to specific sequence mutations (Fig. 3A and 3B, respectively). The determination of $G(t)$ for the WT+NLS sequence using the three models (Fig. 5B) clearly demonstrates the critical influence of the model parameterisation on the resulting viscoelastic properties of the condensates. While $G(t)$ at short timescales is rather similar for all models, indicating that bond, angle, and intramolecular conformational relaxation is similar for all of them; the slope and terminal decay of $G(t)$, which strongly determines the condensate viscosity, is significantly different among all of them. Nevertheless, such variability is much more pronounced when extending the calculation for all the different A1-LCD mutants using a given model, as shown by the Mpipi-Recharged in Fig. 5C. Such drastic variation in condensate viscosity (of approximately 2 orders of magnitude) depicts the strong influence that specific sequence mutations have on regulating the transport properties of condensates, as previously found for other proteins such as FUS[107], TDP-43[99], or Tau[108,109].

In Fig. 5, we represent the estimated viscosity (η) for all A1-LCD sequences from simulations against the corresponding experimental values[78], both normalised by the WT+NLS η, for the CALVADOS2 (D), Mpipi (E), and Mpipi-Recharged (F) models. The calculation of η from $G(t)$ merely consists in integrating the shear stress relaxation modulus (G) as a function of time (for further details see section SVI in the SM). Normalisation by the WT+NLS viscosity is due to the fact that the predicted viscosity by CG models differs from typical experimental values by a few orders of magnitude[69] due to the coarse-graining nature of the model which considers the solvent to be a continuum medium. We compare the relative difference among the sequences with respect to the WT+NLS sequence, and between experiments and simulations by normalising both sets of data (Fig. 5D-F). As expected, a higher sensitivity to sequence mutations by the model to the LLPS critical temperature (see Fig. 3C-E) is also reflected in a larger variation in the condensate viscosity as a function of the sequence variation (Fig. 5D-F). The Mpipi-Recharged exhibits the widest range of viscosity values, spanning almost 2 orders of magnitude depending on the protein variant. The linear correlation coefficient (r) and the root mean squared deviation (D) from the experimental data[78] evidences the contrast between Mpipi, CALVADOS2, and Mpipi-Recharged models, suggesting the latter as the best fit for A1-LCD condensate viscoelastic behaviour.

We also investigate whether a correlation between the

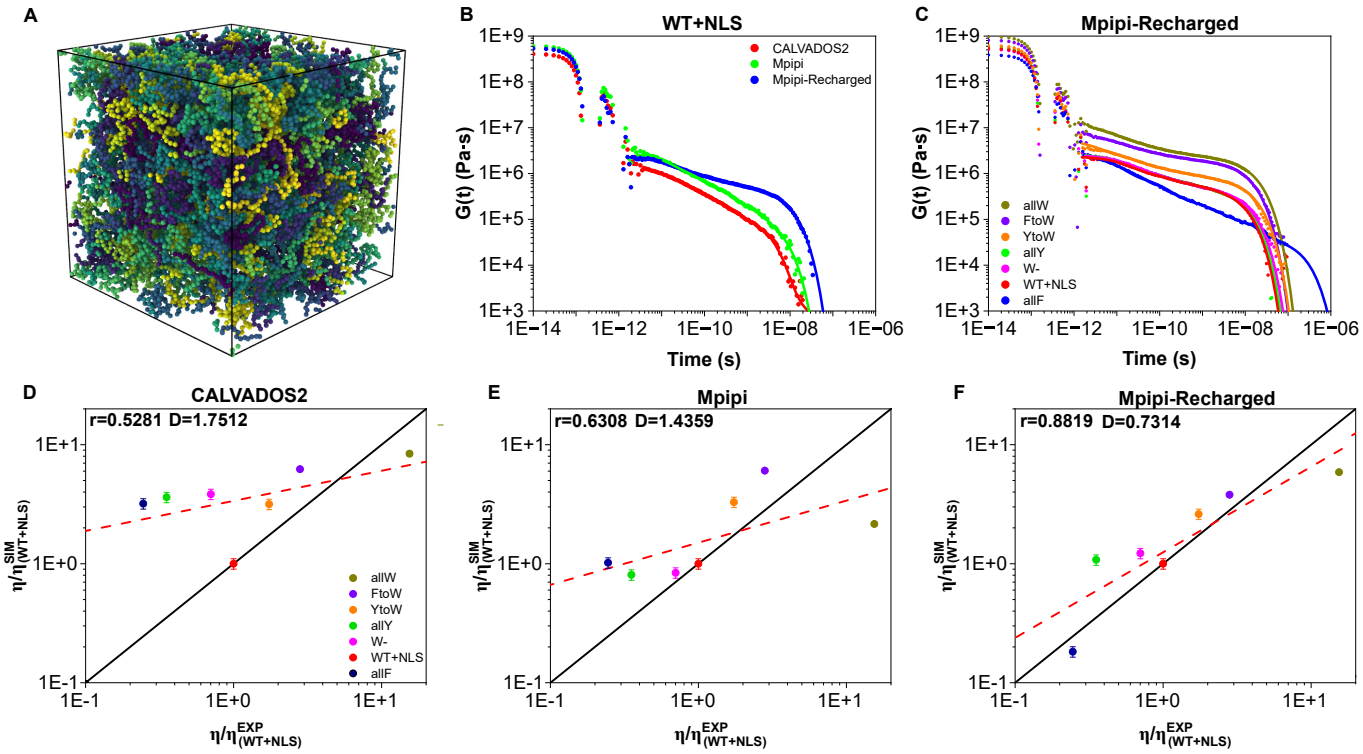

FIG. 5. (**A**) Snapshot of a bulk NVT simulation with 200 A1-LCD (WT+NLS) protein replicas (each of them depicted by a different tone of colour) employed to compute the condensate viscosity. (**B**) Shear stress relaxation modulus ($G(t)$) for the WT+NLS sequence at 298 K evaluated by different models. (**C**) $G(t)$ for different A1-LCD mutants using the Mpipi-Recharged model. In both **B** and **C** panels, $G(t)$ raw data obtained during the simulations are represented by filled symbols, while Maxwell's mode fits to the data (as described in section SVI of the SM) are plotted with solid lines. **D**, **E** and **F** predicted versus *in vitro* viscosity (both reduced by the corresponding WT+NLS value) for all variants under study and the CALVADOS, Mpipi and Mpipi-Recharged models respectively. . The meaning of r, D, and the red dashed and black solid lines is the same as in Fig. 4.

critical temperature and condensate viscosity can be established for the different studied variants. A correlation between these two quantities is expected since intermolecular interaction strength, which favours condensation (i.e., increases T_c), should also diminish molecular mobility (increases η). In Fig. 6, we plot the critical temperature (normalised by the T_c of the WT+NLS sequence) *vs.* condensate viscosity (also normalised by the η of the WT+NLS sequence) for the different studied A1-LCD mutants as predicted by CALVADOS2 (A), Mpipi (B), and Mpipi-Recharged (C) models. Moreover, experimental *in vitro* data from Ref.[78] for the WT+NLS, allF and allY sequences are included (empty squares). As it can be seen, the Mpipi-Recharged predicts the correlation experimentally found for these three variants (depicted by a grey band linearly extrapolated). The full phase diagram (including T_c) for the rest of the variants was not reported in Ref.[78], nevertheless, it is expected that a similar correlation as that predicted by the Mpipi-Recharged model holds since it reasonably predicts the experimental critical temperature, saturation concentration, and condensate viscosity of all these variants, as shown in Figs. 2, 4, and 5, respectively. The CALVADOS2 also shows a correlation between the critical temperature and viscosity (Fig. 6A). However, the direct comparison with experimental data is not satisfactory given that measurements and model predictions do not span over the same range of values for viscosity (Fig. 5D). For the case of Mpipi, a weak correlation between these two quantities is observed (Fig. 6B). Confirming this correlation between T_c and η in condensates for other protein sequences would be extremely relevant since both quantities are challenging to measure, and only in rare occasions are both reported under the same conditions[77,78]. Hence, inferring one from the other, at least qualitatively, can provide complementary highly relevant information regarding the stability and/or transport properties of biomolecular condensates.

V. KEY INTERMOLECULAR INTERACTIONS IN A1-LCD CONDENSATES

Computational approaches can contribute to a deeper understanding of biomolecular phase separation by evaluating the underlying molecular information that is hardly accessible in experiments[69,74,110]. A powerful example is the calculation of intra and inter-molecular contact

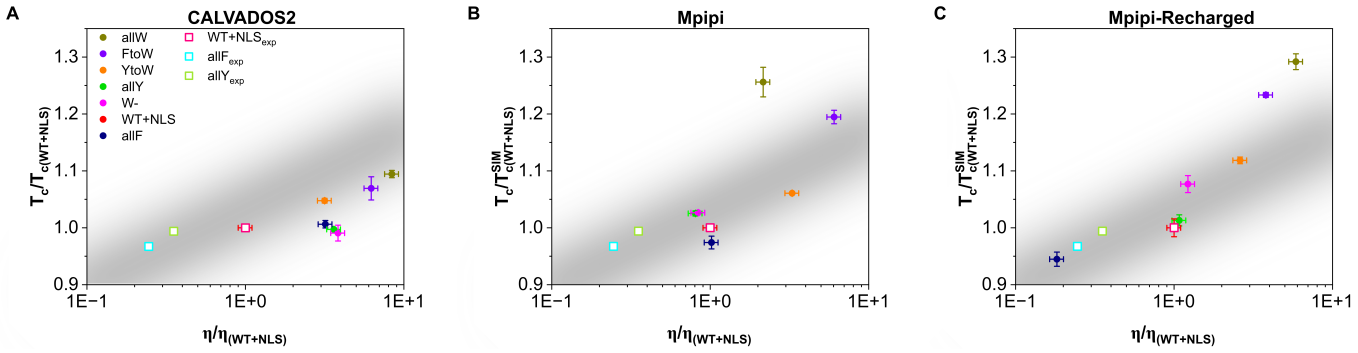

FIG. 6. Correlation between the critical temperature (normalised by the T_c of the WT+NLS sequence) and the condensate viscosity (also normalised by the η of WT+NLS) for the different studied A1-LCD mutants as predicted by the CALVADOS2 (A), Mpipi (B), and Mpipi-Recharged (C) models. Experimental data from Ref.[78] for the WT+NLS, allF and allY sequences have been also included as empty squares (see legend). A linear trend to the experimental correlation between these two quantities for the three sequences measured *in vitro*[78] is shown with a grey band.

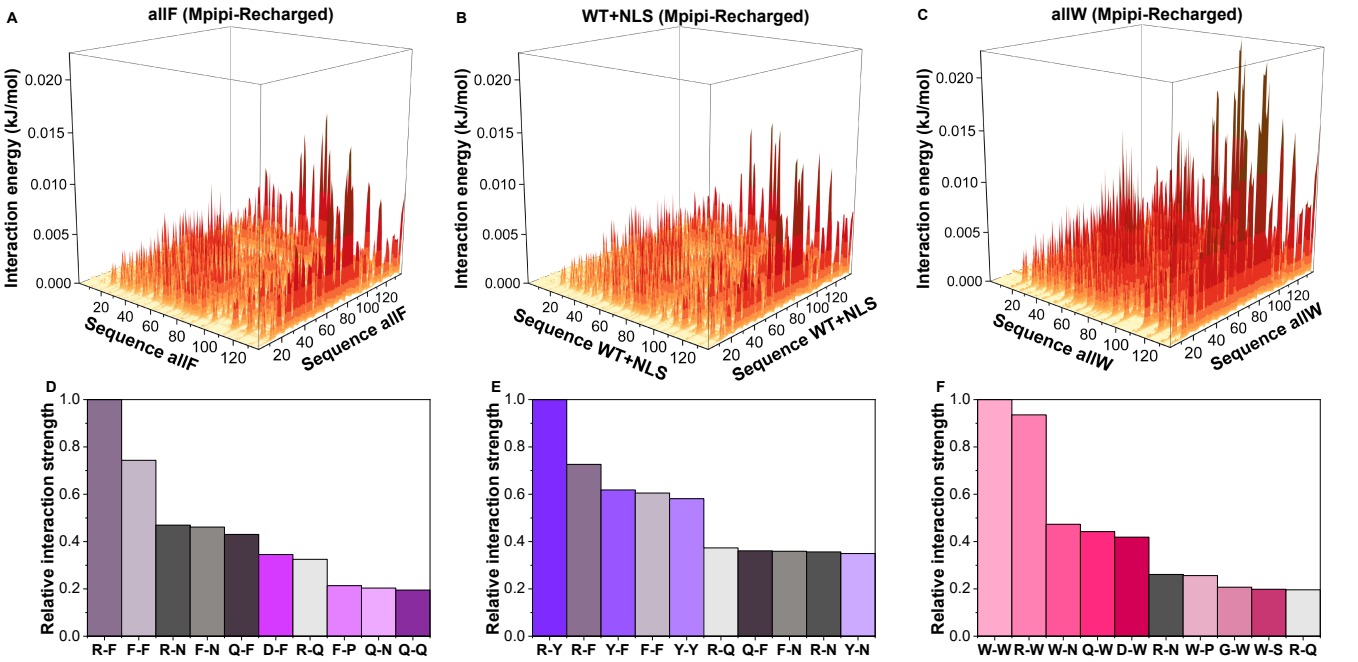

FIG. 7. Contact maps of pairwise intermolecular interaction energy for allF (**A**), WT+NLS (**B**), and allW (**C**) condensates at T = 0.95 T_c as predicted by the Mpipi-Recharged model. The most frequent intermolecular amino acid pairwise interactions sustaining the condensates of allF (**D**), WT+NLS (**E**), and allW (**F**) are also displayed. Pairwise contacts have been both normalised by the highest value of residue-residue pairwise interactions and their abundance across their sequence (see section SV in the SM for further details on these calculations).

maps[52,69] which provide microscopic insights about the key amino acids and protein domains enabling biomolecular self-assembly. Nevertheless, objectively, the outcome of this type of analysis is intimately related to the specific model characteristics and its interaction parameters. Therefore, different models may provide widely different contact domain probabilities for the same protein condensate.

Since the Mpipi-Recharged model has consistently shown to be the most accurate force field of those implemented in this study for reproducing A1-LCD phase behaviour (Figs. 2-6), we now employ it to elucidate the intermolecular interaction potential landscape of the different residues across the sequence when forming phase-separated condensates. By using a distance criterion between pairs of amino acids belonging to different protein replicas and evaluating the energy contribution of the potential terms (e.g., Wang-Frenkel and Yukawa for the Mpipi-Recharged[61]) at the interacting distance (see section SV in the SM), we obtain the contact energy (CE) maps sustaining the liquid network within the different condensates. In Fig. 7, we show the CE maps for con-

densates of the WT+NLS, allF, and allW sequences at 0.95 T_c, referring T_c to the corresponding critical temperature of each variant (Fig. 7A-C). Strikingly, the CE maps of allF and WT+NLS condensates look remarkably similar to each other (Fig. 7A and 7B). Nonetheless, that of allW condensates (Fig. 7C) show significantly larger contact interactions among residues located at the end of the sequence (i.e., from the 90th residue onwards). As a consequence of such increase of intermolecular contacts at the end of the sequence, it also shows stronger interaction energies all along the sequence which indicates a cooperative interplay in sustaining the condensate assembly (Fig. 7C). This increase of intermolecular interactions in the allW mutant explains the large increment in viscosity shown (both computationally and experimentally) in Fig. 5 for this variant with respect to the WT+NLS and allF sequences.

When comparing the CE maps for WT+NLS condensates predicted by the different models (Fig. E1), we find significantly different landscapes of potential interactions enabling LLPS depending on each parameterisation. Interestingly, we find that both HPS and CALVADOS2 present more homogeneous maps of intermolecular interactions than the Mpipi, HPS-cation-π, and Mpipi-Recharged. That is probably a consequence of both models predicting that π-π interactions are stronger than cation-π contacts (Fig. E2). However, while in the CALVADOS2 such relative difference is moderate[55], in the HPS original parameterisation such difference is much more evident[56]. On the other hand, the HPS-cation-π, Mpipi, and Mpipi-Recharged models predict that LLPS is mainly enabled by some specific prion-like sub-domains. Whilst all models show a strong interacting region from the 105th to 115th residue, the HPS-cation-π, Mpipi, and Mpipi-Recharged also present a highly interacting subdomain from the 130th amino acid to the 135th residue. In addition, the HPS and HPS-cation-π, but more prominently the Mpipi, and Mpipi-Recharged models show two sub-domains between the 75th and 85th residues and between the 15th and 45th (although the latter in lower extent) that also significantly contribute to A1-LCD LLPS (Fig. E1).

We now extract from Fig. 7A-C using the Mpipi-Recharged model the most frequent amino acid pairs contributing to stabilise the condensed phase (Fig. 7D-F). The relative interaction strength of each pair has been normalised by that of the contact pair with highest energetic contribution. As expected, the key intermolecular interactions in A1-LCD (WT+NLS) condensates are mostly cation-π (e.g., RY and RF) and π-π (e.g., FF, YF, YY) interactions. In allF condensates, both cation-π (i.e., RF) and π-π (YF and FF) contacts on top of interactions between asparagine (N) with arginine (R) and phenylalanine (F) are the most common contacts. For allW condensates, WW and WR interactions are by far the two most predominant contacts followed by interactions of tryptophan with N, glutamine (Q), and aspartic acid (D). These results are compatible with the framework of stickers and spacers for protein LLPS in which multivalent proteins are considered as heteropolymers made of stickers (i.e. binding sites for associative interactions) and spacers (regions in between stickers)[111,112]. Within such framework, as proposed in Ref.[77] for A1-LCD condensates, tyrosine, phenylalanine, and tryptophan should act as main stickers, arginine as a context-dependent sticker, and the rest of amino acids as spacers. Our simulations confirm that these interactions are abundant across the different studied mutants with varying extents and probabilities depending on the precise sequence composition (Fig. 7D-F.)

We now focus on the most frequent interactions in WT+NLS condensates predicted by the Mpipi-Recharged (Fig. 7E) and the CALVADOS2 model (Fig. E2). While the Mpipi-Recharged suggests that cation-π interactions such as RY and RF are stronger contributors than YY, FF, and YF to promote LLPS, CALVADOS2 proposes the opposite, being π-π contacts more frequent/energetic than cation-π interactions (Fig. E2C). Similarly to the Mpipi-Recharged, the five most energetic contacts enabling LLPS according to the Mpipi model are RY, RF, YF, FF, and YY (Fig. E2D). In contrast, the HPS model does not predict any type of cation-π interaction as main contributors for A1-LCD phase-separation (Fig. E2A). It suggests as key contacts π-π interactions as YF, FF, and YY, as well as contacts between methionine and aromatic residues, which in the model parameterisation have a considerable interaction strength[76]. Such underestimation of cation-π interactions in describing protein LLPS was corrected in the HPS-cation-π reparameterisation[56] by adding an extra potential term to the model. Nevertheless, as shown in Fig. E2B, the HPS-cation-π reparameterisation significantly overestimates cation-π contacts predicting that RY, RF, KY and KF contribute four times more than aromatic interactions such as FF and YF. Such overestimation of cation-π interactions by this model—which leads to extremely high critical temperatures for LLPS as shown in Fig. 2(B)—has been previously discussed in Refs.[31,53,56,69] for other RNA-binding proteins such as FUS, TDP-43, or DDX4, among others. Overall, from this analysis we find that both cation-π and π-π interactions are key LLPS-stabilising contacts in A1-LCD condensates. However, our simulations suggest that pairwise contacts such as RY and RF may play a greater role than purely aromatic contacts such as YY, YF, and FF in A1-LCD phase-separation.

VI. CONCLUSIONS

Computer simulations are a powerful tool to complement *in vitro* and *in vivo* observations in biomolecular condensates[42,113]. In this study we have compared the predictive accuracy of five different sequence-dependent residue-resolution models for A1-LCD protein phase separation. A1-LCD is the low-complexity domain of hn-

RNPA1 protein which enables condensate formation via LLPS[77]. We have systematically evaluated several observables relevant to condensate stability, including the critical temperature for LLPS and the saturation concentration. The direct comparison between the simulated phase diagrams and the experimental coexistence lines has demonstrated a considerable improvement in the most recent models[53,55,61] with respect to the pioneering works that established this coarse-grained computational approach to understand protein LLPS[52,56]. Our results show that the Mpipi-Recharged model displays quantitative accuracy in predicting the critical temperature and condensate saturation concentration of the different A1-LCD mutants (Figs. 2 and 4) reported in Ref.[77,78].

Furthermore, the evaluation of condensate viscosity provides a strong test on the reliability of these models to predict the transport properties of biomolecular condensates. We find that the CALVADOS2, Mpipi, and specially the Mpipi-Recharged model, reproduce the relative viscoelastic variation of protein condensates as a function of specific sequence mutations (Fig. 5). This demonstrates how coarse-grained models can powerfully predict the impact of sequence variations on condensate material properties. Importantly, such characterisation is fundamental to distinguish between liquid-like and solid-like behaviour, which is deeply connected with biological function and pathological malfunction[24,34,42,95,96]. Interestingly, we find a clear correlation—both experimentally and computationally (Fig. 6)—between condensate stability (i.e., T_c) and viscoelasticity; implying that higher critical temperatures for phase-separation are linked with larger condensate viscosities. Such correlation is associated to the strength of inter-molecular interactions: stronger interactions enhance condensation and, in turn, hinder molecular diffusivity[31,97].

Finally, we have evaluated the intermolecular contact energy maps of A1-LCD condensates displayed by the different models studied here, paying particular attention to the Mpipi-Recharged force field (Fig. 7). Our analysis provides the pairwise residue-residue interactions that may potentially sustain A1-LCD condensates. We find as key intermolecular interactions cation-π (e.g., RY and RF) and π-π (e.g., FF, YF, YY) contacts. Besides these interactions, also aromatic (Y and F) and arginine contacts with asparagine and glutamine are abundant. These results are compatible with the proposed framework of stickers and spacers for protein LLPS[77,85,111]. Within such framework, tyrosine, phenylalanine, and tryptophan would act as main stickers, arginine as a context-dependent sticker, and the rest of amino acids as spacers[77]. All these interactions were present across the different studied mutants with varying probabilities depending on their sequence composition as shown in Fig. 7D-F. In contrast, the HPS model does not predict any type of cation-π interaction as main contributors for phase-separation (Fig. E2). Such underestimation of cation-π interactions in describing protein LLPS was later addressed in the HPS-cation-π reparameterisation. However, the HPS-cation-π significantly overestimates the impact of cation-π contacts in A1-LCD condensates leading to extremely high critical temperatures for LLPS, as shown in Fig. 2.

Overall, recent models have significantly improved the prediction power in describing protein phase behaviour, which is built upon the influential work of their predecessors[41,52,56]. We note that more sequence-dependent models such as the HPS-Urry[54] or the FB-HPS[58] have not been tested in this work due to computational limitations in terms of the number of models that are feasible to simulate for all these different quantities. However, and as shown in our previous work[53], we expect these models to perform similarly well as the Mpipi in capturing the phase behaviour of A1-LCD condensates and its mutants. Ultimately, with this study, we aim to demonstrate the continuous improvement that coarse-grained modelling is achieving in quantitatively describing the phase behaviour of biocondensates, and in providing complementary microscopic information that is often hardly accessible through standard experimental setups.

VII. ACKNOWLEDGEMENTS

A. F. acknowledges funding from the Ramon y Cajal fellowship (RYC2021-030937-I) and Spanish National Grant (PID2022-136919NA-C33) I. S.-B. acknowledges funding from the Derek Brewer scholarship of Emmanuel College and EPSRC Doctoral Training Programme studentship, number EP/T517847/1. R.C.-G. acknowledges funding from the European Research Council (ERC) under the European Union Horizon 2020 research and innovation programme (grant agreement 803326). A. T. is funded by European Research Council (ERC) under the European Union Horizon 2020 research and innovation programme (grant agreement 803326) and Ramon y Cajal fellowship (RYC2021-030937-I). J. R. E. also acknowledges funding from the Roger Ekins Research Fellowship of Emmanuel College, the Ramon y Cajal fellowship (RYC2021-030937-I) and the Spanish National Agency for Research (PID2022-136919NA-C33). A.R also acknowledges funding from PID2023-147156NB-I00 of the Spanish Ministry for Science, Innovation and Universities. This work has been performed using resources provided by the Cambridge Tier-2 system operated by the University of Cambridge Research Computing Service (http://www.hpc.cam.ac.uk) funded by EPSRC Tier-2 capital grant EP/P020259/1.

VIII. DATA AVAILABILITY STATEMENT

We provide Supplementary Material and a Github repository that contains the LAMMPS input scripts for every studied model as well as the configuration file for the considered A1-LCD mutants.

IX. REFERENCES

[1] A. A. Hyman and K. Simons, Science **337**, 1047 (2012).
[2] A. A. Hyman, C. A. Weber, and F. Jülicher, Annual review of cell and developmental biology **30**, 39 (2014).
[3] C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Jülicher, and A. A. Hyman, Science **324**, 1729 (2009).
[4] X. Su, J. A. Ditlev, E. Hui, W. Xing, S. Banjade, J. Okrut, D. S. King, J. Taunton, M. K. Rosen, and R. D. Vale, Science **352**, 595 (2016).
[5] Q. Su, S. Mehta, and J. Zhang, Molecular cell **81**, 4137 (2021).
[6] Q. Xiao, C. K. McAtee, and X. Su, Nature Reviews Immunology **22**, 188 (2022).
[7] S. Boeynaems, S. Alberti, N. L. Fawzi, T. Mittag, M. Polymenidou, F. Rousseau, J. Schymkowitz, J. Shorter, B. Wolozin, L. Van Den Bosch, *et al.*, Trends in cell biology **28**, 420 (2018).
[8] H. Zhou, Z. Song, S. Zhong, L. Zuo, Z. Qi, L.-J. Qu, and L. Lai, Angewandte Chemie **131**, 4912 (2019).
[9] J. A. Lemkul and D. R. Bevan, The Journal of Physical Chemistry B **114**, 1652 (2010).
[10] A. Musacchio, The EMBO journal **41**, e109952 (2022).
[11] A. R. Strom, A. V. Emelyanov, M. Mir, D. V. Fyodorov, X. Darzacq, and G. H. Karpen, Nature **547**, 241 (2017).
[12] G. J. Narlikar, Journal of Biosciences **45**, 5 (2020).
[13] D. Deviri and S. A. Safran, Proceedings of the National Academy of Sciences **118**, e2100099118 (2021).
[14] J. A. Riback and C. P. Brangwynne, Science **367**, 364 (2020).
[15] A. Klosin, F. Oltsch, T. Harmon, A. Honigmann, F. Jülicher, A. A. Hyman, and C. Zechner, Science **367**, 464 (2020).
[16] J. Sheu-Gruttadauria and I. J. MacRae, Cell **173**, 946 (2018).
[17] J. J. Bouchard, J. H. Otero, D. C. Scott, E. Szulc, E. W. Martin, N. Sabri, D. Granata, M. R. Marzahn, K. Lindorff-Larsen, X. Salvatella, *et al.*, Molecular cell **72**, 19 (2018).
[18] S. Qamar, G. Wang, S. J. Randle, F. S. Ruggeri, J. A. Varela, J. Q. Lin, E. C. Phillips, A. Miyashita, D. Williams, F. Ströhl, *et al.*, Cell **173**, 720 (2018).
[19] A. C. Murthy, G. L. Dignon, Y. Kan, G. H. Zerze, S. H. Parekh, J. Mittal, and N. L. Fawzi, Nature structural & molecular biology **26**, 637 (2019).
[20] J. L. Carey and L. Guo, Frontiers in molecular biosciences **9**, 826719 (2022).
[21] H.-R. Li, W.-C. Chiang, P.-C. Chou, W.-J. Wang, and J.-r. Huang, Journal of Biological Chemistry **293**, 6090 (2018).
[22] L. McGurk, E. Gomes, L. Guo, J. Mojsilovic-Petrovic, V. Tran, R. G. Kalb, J. Shorter, and N. M. Bonini, Molecular cell **71**, 703 (2018).
[23] E. W. Martin, F. E. Thomasen, N. M. Milkovic, M. J. Cuneo, C. R. Grace, A. Nourse, K. Lindorff-Larsen, and T. Mittag, Nucleic acids research **49**, 2931 (2021).
[24] A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag, and J. P. Taylor, Cell **163**, 123 (2015).
[25] L. Guo and J. Shorter, Molecular cell **60**, 189 (2015).
[26] X. Wang, J. C. Schwartz, and T. R. Cech, Nucleic acids research **43**, 7535 (2015).
[27] M. Farag, S. R. Cohen, W. M. Borcherds, A. Bremer, T. Mittag, and R. V. Pappu, Nature communications **13**, 7722 (2022).
[28] W. Borcherds, A. Bremer, M. B. Borgia, and T. Mittag, Current opinion in structural biology **67**, 41 (2021).
[29] M. P. Hughes, M. R. Sawaya, D. R. Boyer, L. Goldschmidt, J. A. Rodriguez, D. Cascio, L. Chong, T. Gonen, and D. S. Eisenberg, Science **359**, 698 (2018).
[30] S. Ray, N. Singh, R. Kumar, K. Patel, S. Pandey, D. Datta, J. Mahato, R. Panigrahi, A. Navalkar, S. Mehra, *et al.*, Nature chemistry **12**, 705 (2020).
[31] A. R. Tejedor, I. Sanchez-Burgos, M. Estevez-Espinosa, A. Garaizar, R. Collepardo-Guevara, J. Ramirez, and J. R. Espinosa, Nature communications **13**, 1 (2022).
[32] P. P. Gopal, J. J. Nirschl, E. Klinman, and E. L. Holzbaur, Proceedings of the National Academy of Sciences **114**, E2466 (2017).
[33] P. St George-Hyslop, J. Q. Lin, A. Miyashita, E. C. Phillips, S. Qamar, S. J. Randle, and G. Wang, Brain research **1693**, 11 (2018).
[34] A. Patel, H. O. Lee, L. Jawerth, S. Maharana, M. Jahnel, M. Y. Hein, S. Stoynov, J. Mahamid, S. Saha, T. M. Franzmann, *et al.*, Cell **162**, 1066 (2015).
[35] I. R. Mackenzie, A. M. Nicholson, M. Sarkar, J. Messing, M. D. Purice, C. Pottier, K. Annu, M. Baker, R. B. Perkerson, A. Kurti, *et al.*, Neuron **95**, 808 (2017).
[36] Y. Lu, L. Lim, and J. Song, Scientific Reports **7**, 1 (2017).
[37] S. Ambadipudi, J. Biernat, D. Riedel, E. Mandelkow, and M. Zweckstetter, Nature communications **8**, 1 (2017).
[38] I. Bishof, E. B. Dammer, D. M. Duong, S. R. Kundinger, M. Gearing, J. J. Lah, A. I. Levey, and N. T. Seyfried, Journal of Biological Chemistry **293**, 11047 (2018).
[39] S. Spannl, M. Tereshchenko, G. J. Mastromarco, S. J. Ihn, and H. O. Lee, Traffic **20**, 890 (2019).
[40] S. Mehta and J. Zhang, Nature Reviews Cancer **22**, 239 (2022).
[41] G. Tesei, T. K. Schulze, R. Crehuet, and K. Lindorff-Larsen, Proceedings of the National Academy of Sciences **118**, e2111696118 (2021).
[42] B. Szała-Mendyk, T. M. Phan, P. Mohanty, and J. Mittal, Current Opinion in Chemical Biology **75**, 102333 (2023).
[43] G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal, PLoS computational biology **14**, e1005941 (2018).
[44] M. Paloni, R. Bailly, L. Ciandrini, and A. Barducci, The Journal of Physical Chemistry B **124**, 9009 (2020).
[45] W. Zheng, G. L. Dignon, N. Jovic, X. Xu, R. M. Regy, N. L. Fawzi, Y. C. Kim, R. B. Best, and J. Mittal, The Journal of Physical Chemistry B **124**, 11671 (2020).
[46] J. Sponer, M. Krepl, P. Banas, P. Kuhrova, M. Zgarbova, P. Jurecka, M. Havrila, and M. Otyepka, Wiley Interdisciplinary Reviews: RNA **8**, e1405 (2017).
[47] S. Torrino, W. Oldham, A. R. Tejedor, I. Sanchez-Burgos, N. Rachedi, K. Fraissard, C. Chauvet, C. Sbai, B. P. O'Hara, S. Abelanet, *et al.*, "Mechano-dependent sorbitol accumulation supports biomolecular condensate," (2023).
[48] A. Garaizar, T. Higginbotham, I. Sanchez-Burgos, A. R. Tejedor, E. Sanz, and J. R. Espinosa, The Journal of Chemical Physics **157** (2022).
[49] J. R. Espinosa, A. Garaizar, C. Vega, D. Frenkel, and R. Collepardo-Guevara, The Journal of chemical physics **150**, 224510 (2019).
[50] V. Nguemaha and H.-X. Zhou, Scientific reports **8**, 1 (2018).
[51] I. Sanchez-Burgos, J. R. Espinosa, J. A. Joseph, and R. Collepardo-Guevara, Biomolecules **11**, 278 (2021).
[52] G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal, PLoS computational biology **14**, e1005941 (2018).
[53] J. A. Joseph, A. Reinhardt, A. Aguirre, P. Y. Chew, K. O. Russell, J. R. Espinosa, A. Garaizar, and R. Collepardo-Guevara, Nature Computational Science **1**, 732 (2021).
[54] R. M. Regy, J. Thompson, Y. C. Kim, and J. Mittal, Protein Science **30**, 1371 (2021).
[55] G. Tesei and K. Lindorff-Larsen, Open Research Europe **2**, 94 (2023).
[56] S. Das, Y.-H. Lin, R. M. Vernon, J. D. Forman-Kay, and H. S. Chan, Proceedings of the National Academy of Sciences **117**, 28795 (2020).
[57] J. M. Lotthammer, G. M. Ginell, D. Griffith, R. Emenecker, and A. S. Holehouse, Biophysical Journal **123**, 43a (2024).
[58] T. Dannenhoffer-Lafage and R. B. Best, The Journal of Physical Chemistry B **125**, 4046 (2021).
[59] A. Garaizar and J. R. Espinosa, The Journal of Chemical Physics **155**, 125103 (2021).
[60] A. Statt, H. Casademunt, C. P. Brangwynne, and A. Z. Panagiotopoulos, The Journal of Chemical Physics **152**, 075101 (2020).

[61] A. R. Tejedor, A. Aguirre Gonzalez, M. J. Maristany, P. Y. Chew, K. Rusell, J. Ramirez, J. R. Espinosa, and R. Collepardo-Guevara, bioRxiv (2024), 10.1101/2024.07.26.605370.

[62] A. C. Murthy, G. L. Dignon, Y. Kan, G. H. Zerze, S. H. Parekh, J. Mittal, and N. L. Fawzi, Nature structural & molecular biology **26**, 637 (2019).

[63] G. L. Dignon, R. B. Best, and J. Mittal, Annual Review of Physical Chemistry **71**, 53 (2020).

[64] G. L. Dignon, W. Zheng, and J. Mittal, Current opinion in chemical engineering **23**, 92 (2019).

[65] B. S. Schuster, G. L. Dignon, W. S. Tang, F. M. Kelley, A. K. Ranganath, C. N. Jahnke, A. G. Simpkins, R. M. Regy, D. A. Hammer, M. C. Good, *et al.*, Proceedings of the National Academy of Sciences **117**, 11421 (2020).

[66] V. H. Ryan, G. L. Dignon, G. H. Zerze, C. V. Chabata, R. Silva, A. E. Conicella, J. Amaya, K. A. Burke, J. Mittal, and N. L. Fawzi, Molecular cell **69**, 465 (2018).

[67] A. Abyzov, M. Blackledge, and M. Zweckstetter, Chemical Reviews **122**, 6719 (2022).

[68] M. J. Maristany, A. A. Gonzalez, J. R. Espinosa, J. Huertas, R. Collepardo-Guevara, and J. A. Joseph, bioRxiv , 2023 (2023).

[69] A. R. Tejedor, A. Garaizar, J. Ramírez, and J. R. Espinosa, Biophysical Journal **120**, 5169 (2021).

[70] I. Sanchez-Burgos, J. R. Espinosa, J. A. Joseph, and R. Collepardo-Guevara, PLoS computational biology **18**, e1009810 (2022).

[71] R. Funari, N. Bhalla, and L. Gentile, ACS Measurement Science Au **2**, 547 (2022).

[72] X. Huang and R. Powers, Journal of the American Chemical Society **123**, 3834 (2001).

[73] K. Araki, N. Yagi, R. Nakatani, H. Sekiguchi, M. So, H. Yagi, N. Ohta, Y. Nagai, Y. Goto, and H. Mochizuki, Scientific reports **6**, 30473 (2016).

[74] J. A. Riback, M. A. Bowman, A. M. Zmyslowski, C. R. Knoverek, J. M. Jumper, J. R. Hinshaw, E. B. Kaye, K. F. Freed, P. L. Clark, and T. R. Sosnick, Science **358**, 238 (2017).

[75] G.-N. W. Gomes, M. Krzeminski, A. Namini, E. W. Martin, T. Mittag, T. Head-Gordon, J. D. Forman-Kay, and C. C. Gradinaru, Journal of the American Chemical Society **142**, 15697 (2020).

[76] G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, and J. Mittal, Proceedings of the National Academy of Sciences **115**, 9929 (2018).

[77] A. Bremer, M. Farag, W. M. Borcherds, I. Peran, E. W. Martin, R. V. Pappu, and T. Mittag, Nature Chemistry **14**, 196 (2022).

[78] I. Alshareedah, W. M. Borcherds, S. R. Cohen, M. Farag, A. Singh, A. Bremer, R. V. Pappu, T. Mittag, and P. R. Banerjee, "Sequence-encoded grammars determine material properties and physical aging of protein condensates," (2023).

[79] J. A. Joseph, A. Reinhardt, A. Aguirre, P. Y. Chew, K. O. Russell, J. R. Espinosa, A. Garaizar, and R. Collepardo-Guevara, Nature Computational Science **1**, 732 (2021).

[80] P. Li, S. Banjade, H.-C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, *et al.*, Nature **483**, 336 (2012).

[81] S. Boeynaems, S. Alberti, N. L. Fawzi, T. Mittag, M. Polymenidou, F. Rousseau, J. Schymkowitz, J. Shorter, B. Wolozin, L. Van Den Bosch, *et al.*, Trends in cell biology **28**, 420 (2018).

[82] R. M. Vernon, P. A. Chong, B. Tsang, T. H. Kim, A. Bah, P. Farber, H. Lin, and J. D. Forman-Kay, elife **7**, e31486 (2018).

[83] E. W. Martin, A. S. Holehouse, I. Peran, M. Farag, J. J. Incicco, A. Bremer, C. R. Grace, A. Soranno, R. V. Pappu, and T. Mittag, Science **367**, 694 (2020).

[84] A. Ladd and L. Woodcock, Chemical Physics Letters **51**, 155 (1977).

[85] J. R. Espinosa, J. A. Joseph, I. Sanchez-Burgos, A. Garaizar, D. Frenkel, and R. Collepardo-Guevara, Proceedings of the National Academy of Sciences **117**, 13238 (2020).

[86] S. Alberti and D. Dormann, Annu. Rev. Genet **53**, 171 (2019).

[87] M.-T. Wei, S. Elbaum-Garfinkle, A. S. Holehouse, C. C.-H. Chen, M. Feric, C. B. Arnold, R. D. Priestley, R. V. Pappu, and C. P. Brangwynne, Nature Chemistry **9**, 1118 (2017).

[88] A. Wang, A. E. Conicella, H. B. Schmidt, E. W. Martin, S. N. Rhoads, A. N. Reeb, A. Nourse, D. Ramirez Montero, V. H. Ryan, R. Rohatgi, *et al.*, The EMBO journal **37**, e97452 (2018).

[89] S. Maharana, J. Wang, D. K. Papadopoulos, D. Richter, A. Pozniakovsky, I. Poser, M. Bickle, S. Rizk, J. Guillen-Boixet, T. M. Franzmann, M. Jahnel, L. Marrone, Y.-T. Chang, J. Sterneckert, P. Tomancak, A. A. Hyman, and S. Alberti, Science **360**, 918 (2018).

[90] J. Wang, J.-M. Choi, A. S. Holehouse, H. O. Lee, X. Zhang, M. Jahnel, S. Maharana, R. Lemaitre, A. Pozniakovsky, D. Drechsel, *et al.*, Cell **174**, 688 (2018).

[91] C. N. Johnson, K. A. Sojitra, E. J. Sohn, A. K. Moreno-Romero, A. Baudin, X. Xu, J. Mittal, and D. S. Libich, Journal of the American Chemical Society **146**, 8071 (2024).

[92] W. M. Babinchak, R. Haider, B. K. Dumm, P. Sarkar, K. Surewicz, J.-K. Choi, and W. K. Surewicz, Journal of Biological Chemistry **294**, 6306 (2019).

[93] M. Hofweber, S. Hutten, B. Bourgeois, E. Spreitzer, A. Niedner-Boblenz, M. Schifferer, M.-D. Ruepp, M. Simons, D. Niessing, T. Madl, *et al.*, Cell **173**, 706 (2018).

[94] G. Krainer, T. J. Welsh, J. A. Joseph, J. R. Espinosa, S. Wittmann, E. de Csillery, A. Sridhar, Z. Toprakcioglu, G. Gudiškytė, M. A. Czekalska, *et al.*, Nature Communications **12**, 1 (2021).

[95] S. Alberti and A. A. Hyman, Nature reviews Molecular cell biology **22**, 196 (2021).

[96] B. Portz, B. L. Lee, and J. Shorter, Trends in biochemical sciences **46**, 550 (2021).

[97] L. Jawerth, E. Fischer-Friedrich, S. Saha, J. Wang, T. Franzmann, X. Zhang, J. Sachweh, M. Ruer, M. Ijavi, S. Saha, *et al.*, Science **370**, 1317 (2020).

[98] S. Chatterjee, Y. Kan, M. Brzezinski, K. Koynov, R. M. Regy, A. C. Murthy, K. A. Burke, J. J. Michels, J. Mittal, N. L. Fawzi, *et al.*, Advanced Science **9**, 2104247 (2022).

[99] E. L. Guenther, Q. Cao, H. Trinh, J. Lu, M. R. Sawaya, D. Cascio, D. R. Boyer, J. A. Rodriguez, M. P. Hughes, and D. S. Eisenberg, Nature structural & molecular biology **25**, 463 (2018).

[100] M. Linsenmeier, L. Faltova, C. Morelli, U. Capasso Palmiero, C. Seiffert, A. M. Küffner, D. Pinotsi, J. Zhou, R. Mezzenga, and P. Arosio, Nature chemistry **15**, 1340 (2023).

[101] R. S. Fisher and S. Elbaum-Garfinkle, Nature communications **11**, 1 (2020).

[102] D. Sundaravadivelu Devarajan, J. Wang, B. Szała-Mendyk, S. Rekhi, A. Nikoubashman, Y. C. Kim, and J. Mittal, Nature Communications **15**, 1912 (2024).

[103] S. Boeynaems, A. S. Holehouse, V. Weinhardt, D. Kovacs, J. Van Lindt, C. Larabell, L. Van Den Bosch, R. Das, P. S. Tompa, R. V. Pappu, *et al.*, Proceedings of the National Academy of Sciences **116**, 7889 (2019).

[104] J. Ramirez, S. K. Sukumaran, B. Vorselaars, and A. E. Likhtman, The Journal of chemical physics **133**, 154103 (2010).

[105] A. R. Tejedor, J. R. Tejedor, and J. Ramírez, The Journal of Chemical Physics **157** (2022).

[106] A. R. Tejedor, R. Collepardo-Guevara, J. Ramirez, and J. R. Espinosa, The Journal of Physical Chemistry B **127**, 4441 (2023).

[107] H. Deng, K. Gao, and J. Jankovic, Nature Reviews Neurology **10**, 337 (2014).

[108] S. Wegmann, B. Eftekharzadeh, K. Tepper, K. M. Zoltowska, R. E. Bennett, S. Dujardin, P. R. Laskowski, D. MacKenzie, T. Kamath, C. Commins, *et al.*, The EMBO journal **37**, e98049 (2018).

[109] S. K. Rai, A. Savastano, P. Singh, S. Mukhopadhyay, and M. Zweckstetter, Protein Science **30**, 1294 (2021).

[110] N. Galvanetto, M. T. Ivanović, A. Chowdhury, A. Sottini, M. F. Nüesch, D. Nettels, R. B. Best, and B. Schuler, Nature **619**, 876 (2023).
[111] J.-M. Choi and R. V. Pappu, Biophysical Journal **118**, 492a (2020).
[112] U. B. Choi, H. Sanabria, T. Smirnova, M. E. Bowen, and K. R. Weninger, Biomolecules **9**, 114 (2019).
[113] H.-X. Zhou, D. Kota, S. Qin, and R. Prasad, Chemical Reviews (2024).

X. EXTENDED FIGURES

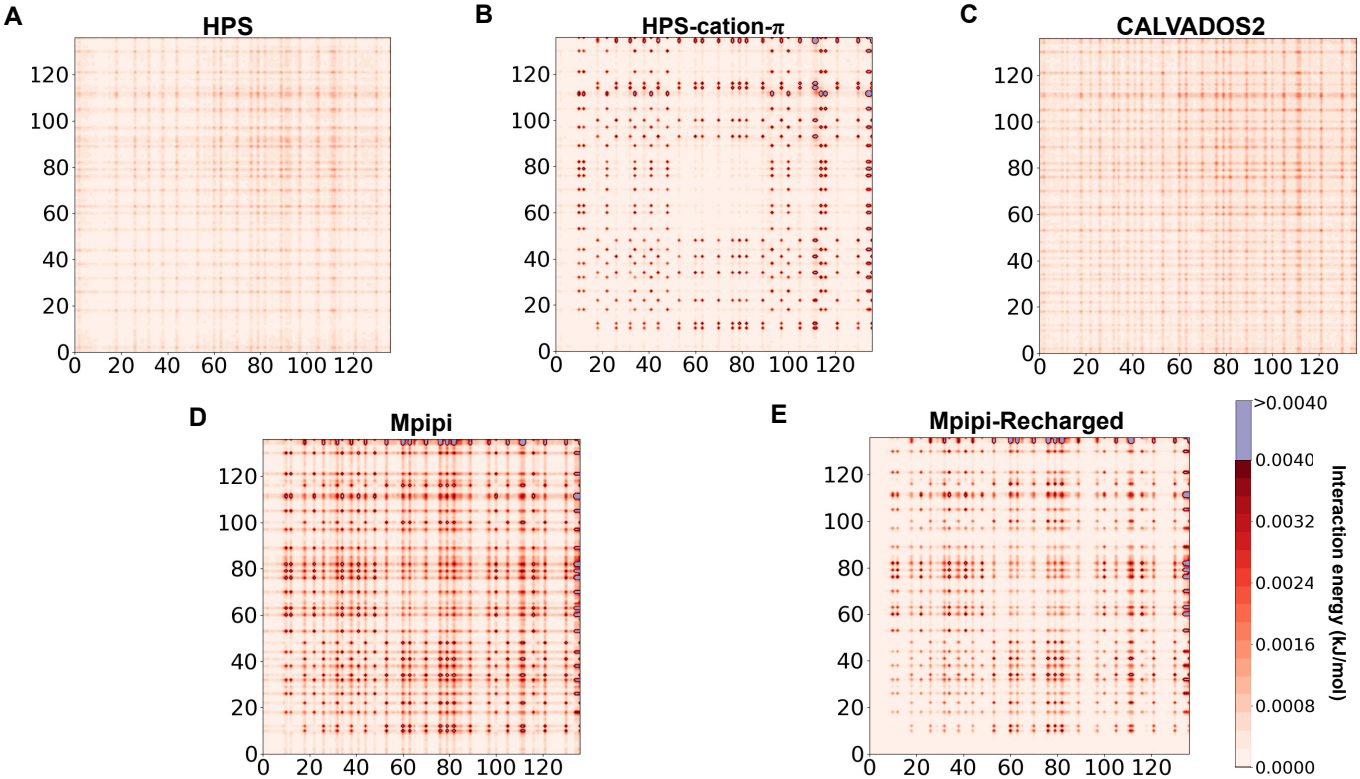

FIG. E1. Contact maps of energy interaction for the A1-LCD (WT+NLS) sequence predicted by the HPS (**A**), HPS-cation-π (**B**), CALVADOS2 (**C**), Mpipi (**D**), and Mpipi-Recharged (**E**) models at T = 0.95 T_c (where T_c refers to the critical temperature of the WT+NLS sequence of each model) at the condensate equilibrium density corresponding to such temperature. The associated energy corresponding to a given specific interaction is depicted by the side bar. Details on these calculations are provided in section SV of the SM.

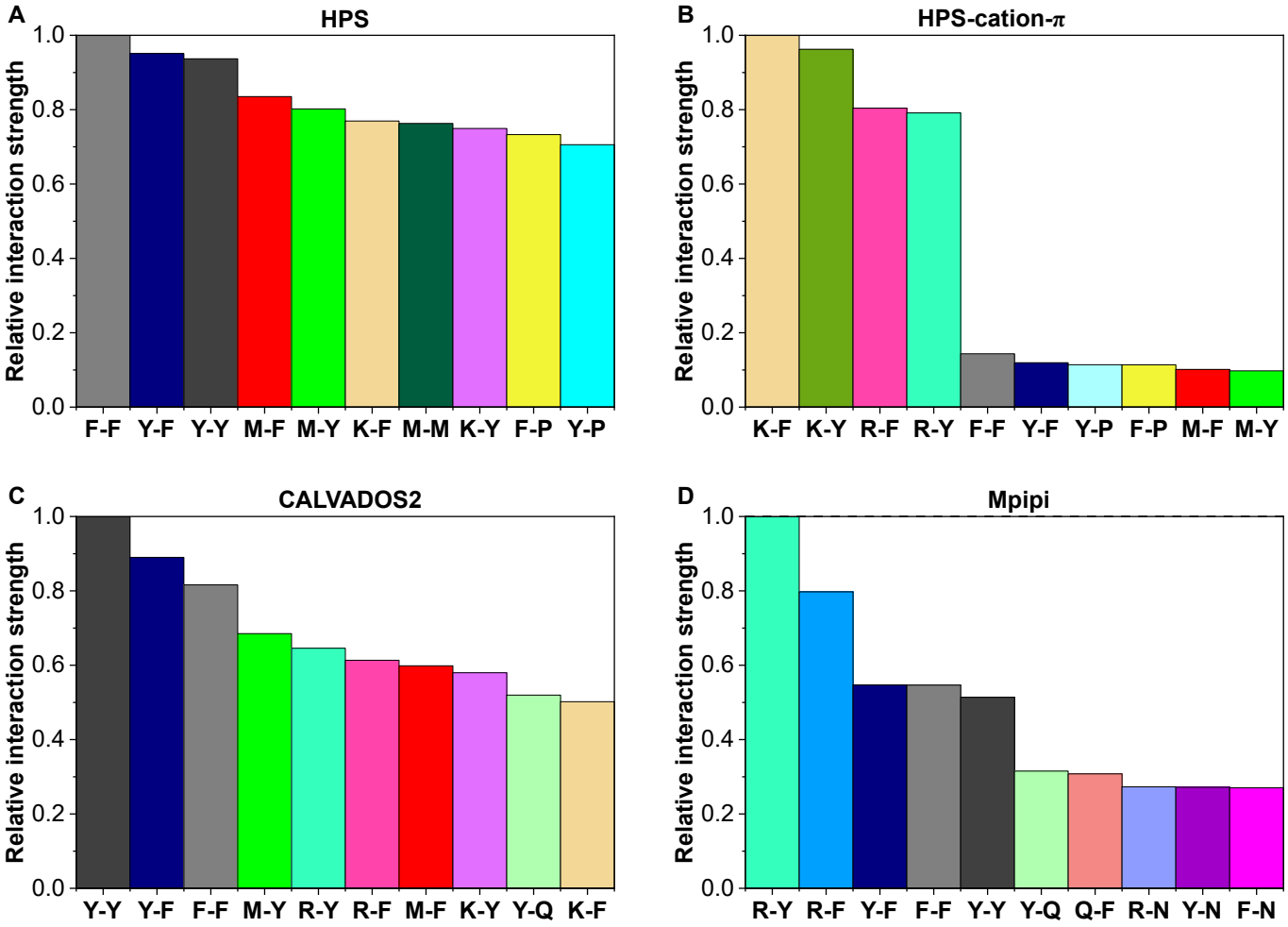

FIG. E2. Predominant intermolecular interactions contributing to LLPS of A1-LCD (WT+NLS) protein as predicted by the HPS (A), HPS-cation-π (B), CALVADOS2 (C), and Mpipi (D) models. The presented contribution by each residue-residue pair has been normalised by the highest contact pair. Also normalisation by the pair residue-residue abundance across the sequence has been applied.