Improved Predictions of Phase Behaviour of Intrinsically Disordered Proteins by Tuning the Interaction Range

Giulio Tesei[1,*] and Kresten Lindorff-Larsen[1,*]

1 Structural Biology and NMR Laboratory & the Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, Copenhagen, Denmark

* giulio.tesei@bio.ku.dk, lindorff@bio.ku.dk

Abstract

The formation and viscoelastic properties of condensates of intrinsically disordered proteins (IDPs) is dictated by amino acid sequence and solution conditions. Because of the involvement of biomolecular condensates in cell physiology and disease, furthering our understanding of the relationship between protein sequence and phase separation (PS) may have important implications in the formulation of new therapeutic hypotheses. Here, we present CALVADOS 2, a coarse-grained model of IDPs that accurately predicts conformational properties and propensity to undergo PS for diverse sequences and solution conditions. In particular, we systematically study the effect of varying the range of nonionic interactions and use our findings to improve the temperature scale of the model. We further optimize the residue-specific model parameters against experimental data on the conformational properties of 55 proteins, while also leveraging 70 hydrophobicity scales from the literature to avoid overfitting the training data. Extensive testing shows that the model accurately predicts chain compaction and PS propensity for sequences of diverse length and charge patterning, as well as at different temperatures and salt concentrations.

1 Introduction

Biomolecular condensates may form via phase separation (PS) into coexisting solvent-rich and macromolecule-rich phases. PS is driven by multiple transient interactions which are in many cases engendered by intrinsically disordered proteins (IDPs) and low-complexity domains (LCDs) of multi-domain proteins [1, 2, 3, 4, 5]. The propensity of IDPs to phase separate and the viscoelastic properties of the condensates are dictated by the amino acid sequence of the constituent IDPs. Moreover, condensates of some IDPs reconstituted in vitro tend to undergo a transition to a dynamically-arrested state, wherein oligomeric species can nucleate and ultimately aggregate into fibrils [6, 7, 2, 3, 8, 9, 10, 11, 12, 13, 14]. As accumulating evidence suggests that these processes may be involved in neurodegeneration and cancer [15, 16, 17, 18], understanding how PS and rheological properties of condensates depend on protein sequence is a current research focus. Due to the transient nature of the protein-protein interactions underpinning PS, quantitative characterization of biomolecular condensates via biophysical experimental methods is challenging, and hence molecular simulations have played an important role in aiding the interpretation of experimental data on condensates reconstituted in vitro [19]. Molecular simulations of the PS of IDPs require a minimal system size of $\sim$100 chains and long simulation times to sample the equilibrium properties of the two-phase system. Therefore, to enhance computational efficiency, it is useful to reduce the complexity of the system by modeling the solvent as a continuum while coarse-graining the atomistic representation of the protein to fewer interaction sites.

A widely used class of coarse-grained models of IDPs describes each residue as a single site centered at the Cα atom. Charged residues interact via salt-screened electrostatic interactions whereas the remaining nonionic nonbonded interactions are incorporated in a single short-range potential characterized by a set of "stickiness" parameters. The

"stickiness" parameters are specific to either the single amino acid or pairs of residues and were originally derived from a hydrophobicity scale [20]. Other models, based on the lattice simulation engines LaSSI [21] and PIMMS [22], classify the amino acids into a reduced number of residue types with distinct "stickiness", ranging from binary categorizations into stickers and spacers [4, 23, 24] to more detailed descriptions and parameterizations [25, 26]. Recently, the accuracy of the "stickiness" parameters has been considerably improved. This has been achieved by leveraging (i) experimental data on single-chain properties, (ii) statistical analyses of protein structures, and (iii) residue-residue free energy profiles calculated from all-atom simulations [27, 28, 29, 30, 31, 26]. In particular, we and others have proposed an automated procedure to optimize the "stickiness" parameters so as to maximize the agreement with experimental small-angle X-ray scattering (SAXS) and paramagnetic relaxation enhancement (PRE) NMR data for a large set of IDPs [28, 29, 32, 26]. To ensure the transferability of the model across sequence space, we employed a Bayesian regularization approach [32, 33]. As the regularization term, we defined the prior knowledge on the "stickiness" parameters in terms of 87 hydrophobicity scales reported in the literature. The resulting M1 parameters capture the relative propensities to phase separate of a wide range of IDP sequences. However, we also observed that a systematic increase in simulation temperature of about 30 °C was needed to quantitatively reproduce the experimental concentration of the dilute phase coexisting with the condensate on an absolute scale. Herein, we refer to this model as the first version of the CALVADOS (Coarse-graining Approach to Liquid-liquid phase separation Via an Automated Data-driven Optimisation Scheme) model (CALVADOS 1).

In this class of coarse-grained models of IDPs, nonionic interactions are modeled via a Lennard-Jones-like potential, which decays to zero only at infinite residue-residue distances. For computational efficiency, the potential is typically calculated up to a cutoff distance, r_c, and interactions between particles that are farther apart are ignored. Although this truncation may introduce severe artifacts, in the different implementations of the models, the value of r_c has varied considerably between 1 and 4 nm [29, 20, 34, 31, 30, 35, 32]. Here, we systematically investigate the effect of the cutoff of nonionic interactions on single-chain compaction and PS propensity. We find that decreasing the cutoff from 4 to 2 nm results in a small increase in the radius of gyration whereas the PS propensity significantly decreases. We exploit this effect to improve the temperature-dependence of the CALVADOS 1 model by tuning the cutoff of the nonionic potential. Further, we perform a Bayesian optimization of the "stickiness" parameters using a cutoff of 2.4 nm and an augmented training set. We show that the updated model (CALVADOS 2) has improved predictive accuracy.

2 Results and Discussion

When applying a cutoff scheme, we neglect the interactions of residues separated by a distance, r, larger than the cutoff, r_c. For the most strongly interacting residue pair (between two tryptophans), the nonionic potential of the CALVADOS 1 model at $r_c = 2$ nm takes the value of -5 J mol^{-1}, that is, only a small fraction of the thermal energy (Fig. 1*A*). However, the Lennard-Jones potential falls off slowly whereas the number of interacting partners increases quadratically with increasing r. Therefore, in a simulation of a protein-rich phase, decreasing the cutoff from 4 to 2 nm (Fig. 1*A*) can imply ignoring a total interaction energy per protein of several times the thermal energy.

We first look into the effect of the choice of cutoff on the conformational ensembles of isolated proteins. We simulated single IDPs of different sequence length, $N = 71$–441, and average hydropathy, $\langle\lambda\rangle = 0.33$–$0.63$. The average radii of gyration, $\langle R_g\rangle$, calculated from simulation trajectories are systematically larger when we use $r_c = 2$ nm, compared to the values obtained using $r_c = 4$ nm. CALVADOS 1 was optimized using the longer r_c and estimating the ensemble average R_g values as the root-mean-square R_g, $\sqrt{\langle R_g^2\rangle}$. Since $\sqrt{\langle R_g^2\rangle}$ is systematically larger than $\langle R_g\rangle$, decreasing r_c to 2 nm results in a slight improvement of the agreement between the calculated $\langle R_g\rangle$ values and the experimental data (Fig. 1*B*).

To gain further insight into the effect of the cutoff, we performed simulations of single chains of α-Synuclein, hnRNPA1 LCD, PNt and Tau 2N4R (Table S1 and S2) using $r_c = 2$, 2.5, 3 and 4 nm. Irrespective of the sequence, $\langle R_g\rangle$

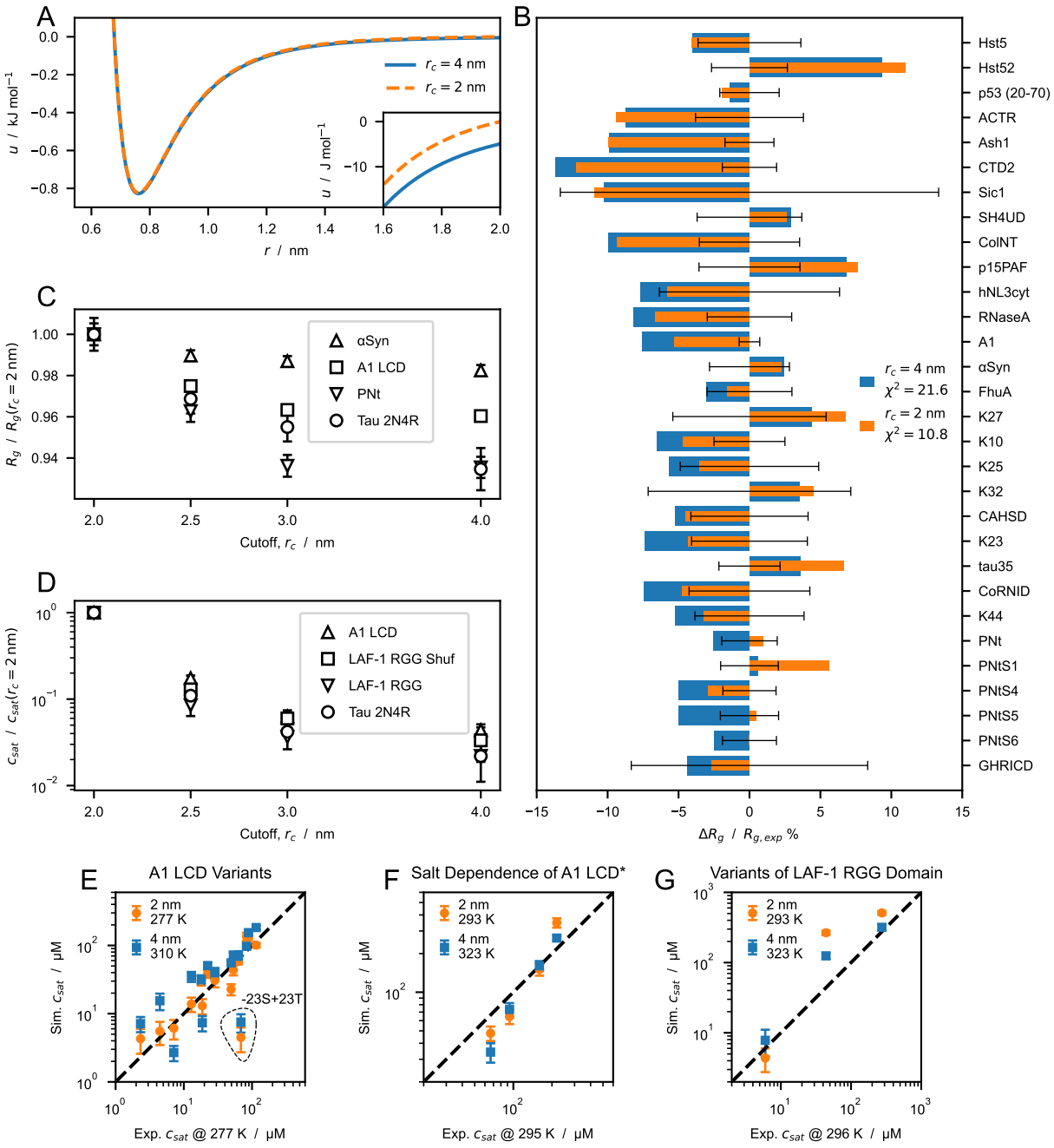

Figure 1: Effect of cutoff size on predictions of radii of gyration, R_g, and saturation concentration, c_{sat}, from simulations performed using the CALVADOS 1 parameters. (*A*) Nonionic Ashbaugh-Hatch potentials between two W residues with cutoff, r_c, of 4 (blue solid line) and 2 nm (orange dashed line). The inset highlights differences between the potentials for $r_s \leq r \leq r_c$. (*B*) Relative difference between experimental and predicted radii of gyration, $\langle R_g \rangle$, for $r_c = 4$ (blue) and 2 nm (orange). χ^2_r values reported in the legend are calculated for all the sequences in Table S1. (*C*) $\langle R_g \rangle$ of α-Synuclein, hnRNPA1 LCD, PNt and human full-length tau (Table S1 and S2) from simulations performed with increasing cutoff size, r_c, and normalized by the value at $r_c = 2$ nm. (*D*) Saturation concentration, c_{sat}, for hnRNPA1 LCD, the randomly shuffled sequence of LAF-1 RGG domain, LAF-1 RGG domain and human full-length tau for increasing values of r_c and normalized by the c_{sat} at $r_c = 2$ nm. (*E–G*) Correlation between c_{sat} from simulations and experiments for (*E*) A1 LCD variants, (*F*) A1 LCD* WT at $[\mathrm{NaCl}] = 0.15$, 0.2, 0.3 and 0.5 M and (*G*) variants of LAF-1 RGG domain (Table S4).

decreases monotonically with increasing r_c. However, the effect on compaction appears to be larger for long sequences and high content of hydrophobic residues, both of which result in an increased number of shorter intramolecular distances. For example, upon increasing the r_c from 2 to 4 nm, the $\langle R_g \rangle$ of α-Synuclein ($N = 140$, $\langle \lambda \rangle = 0.33$) decreases by 2.3% whereas the effect is more pronounced for hnRNPA1 LCD ($N = 137$, $\langle \lambda \rangle = 0.61$) and Tau 2N4R ($N = 441$, $\langle \lambda \rangle = 0.38$), with a decrease in $\langle R_g \rangle$ of 4.0% and 7.7%, respectively.

To investigate the effect of the cutoff distance on PS propensity, we performed direct-coexistence simulations of 100 chains of hnRNPA1 LCD, LAF-1 RGG domain (WT and shuffled sequence with higher charge segregation), and Tau 2N4R (Table S4). From the simulation trajectories of the two-phase system at equilibrium, we calculate c_{sat}, i.e. the protein concentration in the dilute phase coexisting with the condensate. The higher the c_{sat} value, the lower the propensity of the IDP to undergo PS. As expected from the increased contact density in the condensed phase, the choice of cutoff has a considerably larger impact on c_{sat} than on chain compaction: decreasing r_c from 4 to 2 nm results in an increase in c_{sat} of over one order of magnitude. In contrast to what we observed for the $\langle R_g \rangle$, the decrease in c_{sat} does not show a clear dependence on sequence length and average hydropathy. From the multi-chain trajectories of hnRNPA1 LCD, LAF-1 RGG domain (WT and shuffled sequence) and Tau 2N4R obtained using $r_c = 4$ nm, we estimate that the increase in nonionic energy per protein upon decreasing the cutoff from 4 to 2 nm is 13 ± 1 kJ mol^{-1} (mean$\pm$standard deviation), respectively (Fig. S1*A*). Assuming that the number of interactions neglected by the shorter cutoff is proportional to the sequence length and the amino acid concentration in the condensate, the small variance in the energy increase across the different IDPs finds explanation in the fact that the simulated systems display similar values of $N^2 \times c_{con}$ (Fig. S1*A*), where c_{con} is the protein concentration in the condensate. The ratio $U(r_c = 2\ \mathrm{nm})/U(r_c = 4\ \mathrm{nm})$ of the nonionic energies for $r_c = 2$ and 4 nm is also largely system independent (Fig. S1*B*). Moreover, decreasing the temperature by $\sim$30 K in the range between 310 and 323 K has a rather small impact on the relative strength of the electrostatic interactions with respect to the thermal energy, due to the decrease in the dielectric constant of water with increasing temperature (Fig. S1*C*). Therefore, we speculate that the effect of decreasing r_c can be compensated by simulating the system at a lower temperature (Fig. S1*B*).

With these considerations in mind, we use the CALVADOS 1 model with $r_c = 2$ nm to run direct-coexistence simulations of IDPs for which c_{sat} has been measured experimentally (Table S4), i.e. variants of hnRNPA1 LCD, hnRNPA1 LCD* at various salt concentrations, and LAF-1 RGG domain variants. As we have shown that decreasing the range of the nonionic interactions disfavours PS, we perform these simulations at the experimental temperatures, which are lower by $\sim$30 K than those required to reproduce the experimental c_{sat} values when the model is simulated with $r_c = 4$ nm (Fig. 1*E–G*). The two-fold decrease in r_c enables the model to quantitatively recapitulate the experimental c_{sat} data at the temperature at which the experiments were conducted. Notably, we show this for diverse sequences, across a wide range of ionic strengths, and for variants with different charge patterning and numbers of aromatic and charged residues. These results suggest that the range of interaction of the Lennard-Jones potential may be too large [36]. While the r^{-6} dependence is strictly correct for dispersion interactions between atoms, the nonionic potential of our model incorporates a variety of effective nonbonded interactions between residues, and hence the Lennard-Jones potential is not expected to capture the correct interaction range [31].

Since CALVADOS 1 was developed using $r_c = 4$ nm, we examined whether reoptimizing the model with the shorter cutoff could result in a comparably accurate model. As detailed in the Methods Section, we performed a Bayesian parameter-learning procedure [32] using an improved algorithm, an expanded training set (Table S1), and $r_c = 2$ nm. Fig. S2 shows that the new model tends to underestimate the c_{sat} values of the most PS-prone sequences. We hypothesize that during the optimization the reduction of attractive forces due to the shorter cutoff is overcompensated by an overall increase in λ. We tested this hypothesis by performing the optimization with increasing values of r_c, in the range between 2.0 and 2.5 nm, and found that the c_{sat} values predicted from simulations performed with $r_c = 2.0$ nm increased monotonically with the r_c used for the optimization (Fig. S3). Using $r_c = 2.4$ nm for the optimization resulted in a model with improved accuracy compared to CALVADOS 1 (Fig. 2), especially for the PS of LAF-1 RGG domain and the $-$23S$+$23T variant of A1 LCD. To test the robustness of the approach, the optimization was carried out starting from $\lambda_0 = 0.5$ for all the amino acids (Fig. 2) and from λ_0 =M1 (Fig. S4 and S5). The difference between the resulting sets of optimal

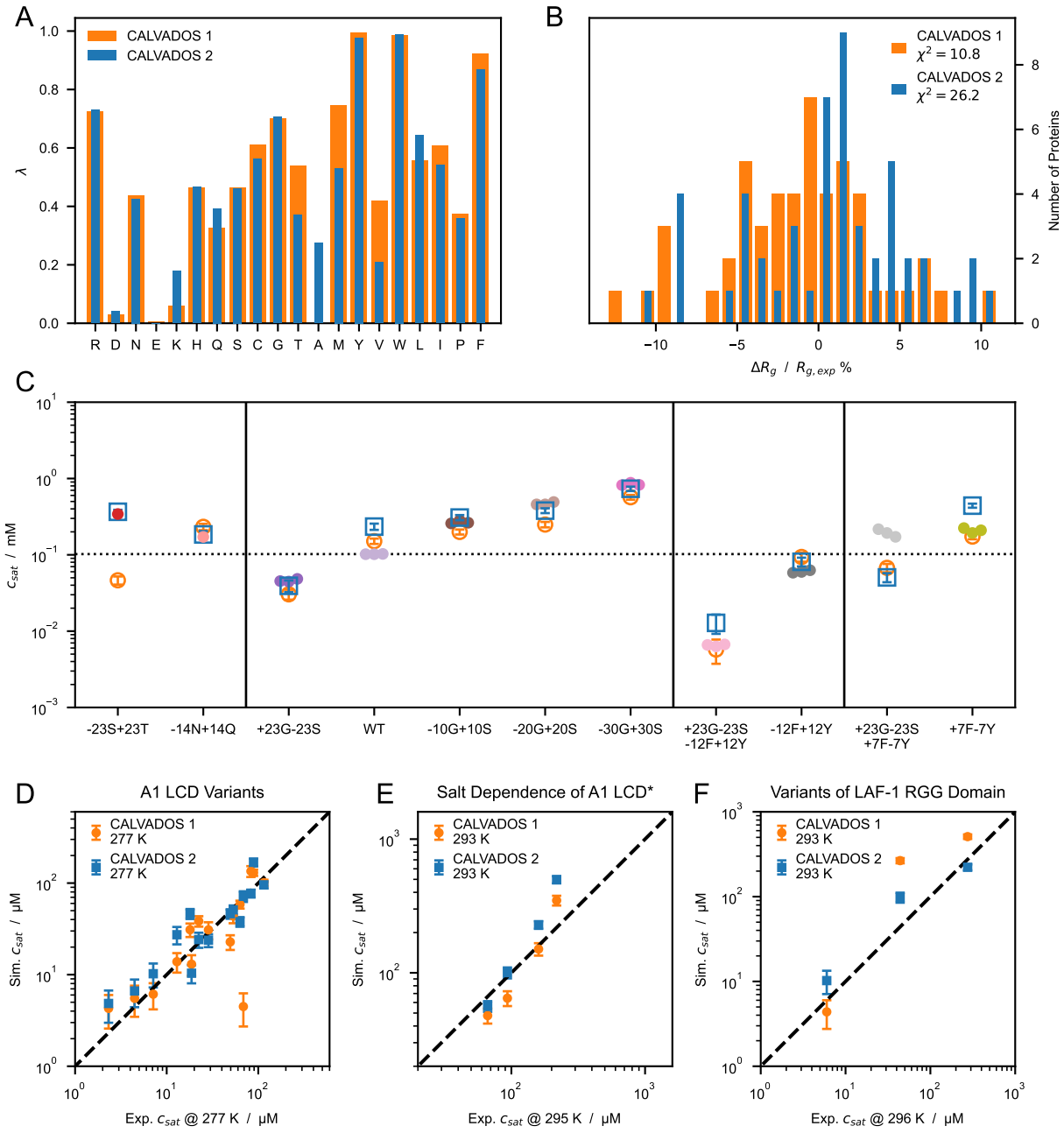

Figure 2: (*A*) Comparison between $\boldsymbol{\lambda}$ sets of CALVADOS 1 (orange) and CALVADOS 2 (blue). (*B*) Distribution of the relative difference between experimental (Table S1) and predicted radii of gyration, $\langle R_g \rangle$, for CALVADOS 1 (orange) and CALVADOS 2 (blue). (*C*) Comparison between saturation concentrations, c_{sat}, at 293 K of variants of hnRNPA1 LCD measured by Bremer, Farag, Borcherds et al. [37] (closed circles) and corresponding predictions of CALVADOS 1 (open orange circles) and CALVADOS 2 (open blue squares). (*D–F*) Correlation between c_{sat} from simulations and experiments for (*D*) A1 LCD variants, (*E*) A1 LCD* WT at $[\mathrm{NaCl}] = 0.15$, 0.2, 0.3 and 0.5 M and (*F*) variants of LAF-1 RGG domain (Table S4).

λ values (Fig. S5*A*) is lower than 0.08 for all the residues and exceeds 0.05 only for S, T and A. The model obtained starting from $\lambda_0 = 0.5$ is more accurate at predicting PS propensities and will be referred to as CALVADOS 2 hereafter. The λ values of CALVADOS 1 and 2 differ mostly for K, T, A, M, and V, whereas the smaller deviations ($|\Delta\lambda| < 0.09$) observed for Q, L, I, and F (Fig. 2*A*) are within the range of reproducibility of the method (Fig. S5*A*). Although CALVADOS 1 was optimized using $r_c = 4$ nm, predictions of single-chain compaction from simulations performed using $r_c = 2$ nm are more accurate for CALVADOS 1 than for CALVADOS 2. This result can be explained by the opposing effects of decreasing the cutoff and calculating R_g values as $\langle R_g \rangle$ insteead of $\sqrt{\langle R_g^2 \rangle}$. In fact, the $\sqrt{\langle R_g^2 \rangle}$ values predicted by

CALVADOS 1 are strikingly similar to the $\langle R_g \rangle$ values predicted by CALVADOS 2 (Fig. S6).

The correlation between experimental and predicted R_g values for the 67 proteins of Table S1 and S2 is excellent for both CALVADOS 1 and 2 (Fig. S7*A*). On the other hand, CALVADOS 2 is more accurate than CALVADOS 1 at predicting PS propensities, as evidenced by Pearson's correlation coefficients of 0.93 and 0.82, respectively, for the experimental and predicted c_{sat} values of the 26 sequences of Table S4 (Fig. S7*B*).

Capturing the interplay between short-range nonionic and long-range ionic interactions is essential for accurately modeling the PS of IDPs [38, 39, 37, 40]. Our results show that the decrease in the range of the nonionic potential reported in this work does not significantly perturb the balance between ionic and nonionic forces. In fact, CALVADOS 1 and 2 accurately predict the PS propensities of A1 LCD* at various salt concentrations, as well as the c_{sat} of variants of A1 LCD and LAF-1 RGG domain with different charge patterning (Fig. 1*E–G* and 2*D–F*). Moreover, CALVADOS 1 and 2 recapitulate the effect of salt concentration and charge patterning on the chain compaction of A1 LCD* [41] and p27-C constructs [42], respectively (Fig. S8).

As additional test systems, we considered constructs of the 1–80 N-terminal fragment of yeast Lge1, which have been recently investigated using turbidity measurements [43]. CALVADOS 2 correctly predicts that the WT Lge1$_{1-80}$ construct undergoes PS at the experimental conditions, albeit with a hundred times larger c_{sat} (50 ± 6 μM at $c_s = 100$ mM) compared to experiments (< 1 μM). In agreement with experiments, CALVADOS 2 predicts that mutating all the 11 R residues to K increases c_{sat} by over one order of magnitude whereas mutating the 14 Y residues of the 1–80 fragment to A abrogates PS (Fig. S9).

3 Conclusions

In the context of a previously developed Cα-based IDP model (CALVADOS), we show that neglecting the long range of attractive Lennard-Jones interactions has a small impact on the compaction of a single chain while strongly disfavouring PS. The effect can be explained by the smaller number of neglected pair interactions for a residues in an isolated chain compared to the dense environment of a condensate. Moreover, we find that the effect of reducing the range of interaction by a factor of two is relatively insensitive to sequence length and composition. Therefore, decreasing the cutoff of the Lennard-Jones potential of the Cα-based model engenders a similar generic effect on chain compaction and PS as a corresponding increase in temperature. We take advantage of this finding to solve the temperature mismatch of the CALVADOS model. Namely, we decrease the cutoff of the nonionic interactions from 4 to 2 nm and obtain accurate c_{sat} predictions at the experimental conditions, whereas simulations at temperatures higher by 30 °C were required in the original implementation. Finally, we used the shorter cutoff to reoptimize the "stickiness" parameters of the model against experimental data reporting on single-chain compaction. The small expansion of the chain conformations is overcompensated by an overall increase in "stickiness" so that the resulting model tends to underestimate the experimental c_{sat} values. By systematically increasing the cutoff used in the development of the "stickiness" scale, we find that performing the optimization using $r_c = 2.4$ nm results in a model (CALVADOS 2) which yields accurate predictions from simulations run using $r_c = 2$ nm at the experimental conditions. We present CALVADOS 2 as an improvement of our previous model by testing on sets of experimental R_g and c_{sat} data comprising 16 and 36 systems, respectively, which were not used in the parameterization of the model.

4 Methods

4.1 Molecular Simulations

Molecular dynamics simulations are conducted in the NVT ensemble using the Langevin integrator with a time step of 10 fs and friction coefficient of 0.01 ps^{-1}. Non-bonded interactions between residues separated by one bond are excluded from the energy calculations. Functional forms and parameters for bonded and nonbonded interactions are reported in the "Bonded and Nonbonded Interactions" Subsection. Single chains of N residues are simulated using HOOMD-blue v2.9.3 [44] in a cubic box of side length $0.38 \times (N-1) + 4$ nm under periodic boundary conditions, starting from the fully extended linear conformation. Each chain is simulated in ten replicas for $\sim 6 \times 0.3 \times N^2$ ps if $N > 100$ and for 18 ns otherwise. The initial 100 frames of each replica are discarded, so as to sample 5,000 weakly correlated conformations for each protein (Fig. S10). Direct-coexistence simulations are performed using openMM v7.5 [45] in a cuboidal box of side lengths $[L_x, L_y, L_z] = [25, 25, 300]$, $[17, 17, 300]$ and $[15, 15, 150]$ nm for Tau 2N4R, Ddx4 LCD, and for the remainder of the proteins, respectively. In the starting configuration, 100 chains are aligned along the z-axis and with their middle beads placed in the xy-plane at random (x, y) positions which are more than 0.7 nm apart. Multi-chain simulations are carried out for at least 2 µs, saving frames every 0.5 ns (Fig. S11, S12, and S13). After discarding the initial 0.6 µs, the slab is centered in the box at each frame as previously described [32] and the equilibrium density profile, $\rho(z)$, is calculated by averaging over the trajectory of the system at equilibrium. The densities of the dilute and protein-rich phases are estimated as the average densities in the regions $|z| < z_{DS} - t/2$ and $|z| > z_{DS} + 6t$ nm, where z_{DS} and t are the position of the dividing surface and the thickness of the interface, respectively. z_{DS} and t are obtained by fitting the semi-profiles in $z > 0$ and $z < 0$ to $\rho(z) = (\rho_a + \rho_b)/2 + (\rho_b - \rho_a)/2 \times \tanh\left[(|z| - z_{DS})/t\right]$, where ρ_a and ρ_b are the densities of the protein-rich and dilute phases, respectively. The uncertainty of the density values is estimated as the standard error obtained from the blocking approach [46] implemented in the BLOCKING software (github.com/fpesceKU/BLOCKING).

4.2 Bonded and Nonbonded Interactions

In this study, we used the following truncated and shifted Ashbaugh-Hatch potential [47],

$$u_{AH}^{SP}(r) = \begin{cases} u_{LJ}(r) - \lambda u_{LJ}(r_c) + \epsilon(1-\lambda), & r \le 2^{1/6}\sigma \\ \lambda\left[u_{LJ}(r) - u_{LJ}(r_c)\right], & 2^{1/6}\sigma < r < r_c \\ 0, & r > r_c, \end{cases} \tag{1}$$

where $\epsilon = 0.8368$ kJ mol^{-1}, $r_c = 2$ or 4 nm, and u_{LJ} is the Lennard-Jones potential:

$$u_{LJ}(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{2}$$

σ and λ are arithmetic averages of amino acid specific parameters quantifying size and hydropathy, respectively. For σ, we use the values calculated from van der Waals volumes by Kim and Hummer [48] whereas, for λ, we use the recently proposed M1 parameters [32] and the values optimized in this work.

Salt-screened electrostatic interactions are modeled via the Debye-Hückel potential,

$$u_{DH}(r) = \frac{q_i q_j e^2}{4\pi\epsilon_0\epsilon_r}\frac{\exp\left(-r/D\right)}{r} \tag{3}$$

where q is the average amino acid charge number, e is the elementary charge, $D = \sqrt{1/(8\pi B c_s)}$ is the Debye length of an electrolyte solution of ionic strength c_s and $B(\epsilon_r)$ is the Bjerrum length. Electrostatic interactions are truncated and shifted at the cutoff distance $r_c = 4$ nm, irrespective of the value of r_c used for Eq. 1. We use the following empirical relationship [49]

$$\epsilon_r(T) = \frac{5321}{T} + 233.76 - 0.9297 \times T + 1.417 \times 10^{-3} \times T^2 - 8.292 \times 10^{-7} \times T^3, \tag{4}$$

to model the temperature-dependent dielectric constant of the implicit aqueous solution. The Henderson–Hasselbalch equation is used to estimate the average charge of the histidine residues, assuming a pK_a value of 6 [50].
The amino acid beads are connected by harmonic potentials,

$$u_{bond}(r) = \frac{1}{2}k(r - r_0)^2, \tag{5}$$

of force constant $k = 8033$ kJ mol^{-1} nm^{-2} and equilibrium distance $r_0 = 0.38$ nm.

4.3 Optimization of the "Stickiness" Scale

The optimization of the "stickiness" parameters, $\boldsymbol{\lambda}$, is carried out to minimize the cost function $\mathcal{L}(\boldsymbol{\lambda}) = \langle \chi^2_{R_g}(\boldsymbol{\lambda}) \rangle + \eta \langle \chi^2_{PRE}(\boldsymbol{\lambda}) \rangle - \theta \ln [P(\boldsymbol{\lambda})]$ using an algorithm which is analogous to the one we previously described [32]. $\chi^2_{R_g}$ and χ^2_{PRE} quantify the discrepancy between model predictions and experimental data, and are defined as $\chi^2_{R_g} = [(R_g^{exp} - R_g^{calc})/\sigma^{exp}]^2$ and $\chi^2_{PRE} = \frac{1}{N_{labels}N} \sum_j^{N_{labels}} \sum_i^{N_{res}} [(Y_{ij}^{exp} - Y_{ij}^{calc})/\sigma_{ij}^{exp}]^2$, respectively, where σ^{exp} is the error on the experimental values, Y is either PRE rates or intensity ratios and N_{labels} is the number of spin-labeled mutants used for the NMR PRE data. In the expression for the cost function, the coefficients are set to $\eta = 0.1$ and $\theta = 0.05$. The prior is the distribution of λ, $P(\boldsymbol{\lambda})$, derived from a subset of the hydrophobicity scales reported in Table 3 and 4 of Simm et al. [51]. Specifically, only the 70 scales that are unique after min-max normalization (Fig. S14) are used for the calculation of $P(\lambda)$, namely Wimley, BULDG reverse, MANP780101, VHEG790101, JANIN, JANJ790102, WOLR790101, PONP800101–6, Wilson, FAUCH, ENGEL, ROSEM, JACWH, CowanWhittacker, ROSM880101 reverse, ROSM880102 reverse, COWR900101, BLAS910101, CASSI, CIDH920101, CIDH920105, CIDBB, CIDA+, CIDAB, PONG1–3, WILM950101–2, WILM950104, Bishop reverse, NADH010101–7, ZIMJ680101, NOZY710101, JONES, LEVIT, KYTJ820101, SWER830101, SWEET, EISEN, ROSEF, GUYFE, COHEN, NNEIG, MDK0, MDK1, JURD980101, SET1–3, CHOTA, CHOTH, Sweet & Eisenberg, KIDER, ROSEB, Welling reverse, Rao & Argos, GIBRA, and WOLR810101 reverse. $P(\boldsymbol{\lambda})$ is obtained via multivariate kernel density estimation, as implemented in scikit-learn [52], using a Gaussian kernel with bandwidth of 0.05. This prior is 20-dimensional and contains information on the λ-distribution of the single amino acid (Fig. S15) as well as on the covariance matrix (Fig. S16) inferred from our selection of 70 hydrophobicity scales.

In the first step of the optimization procedure, the λ values for all the amino acids are set to 0.5, $\lambda_0 = 0.5$, and these parameters are used to simulate the proteins of the training set (Table S1). We proceed with the first optimization cycle, wherein, at each k-th iteration, the λ values of a random selection of five amino acids are nudged by random numbers picked from a normal distribution of standard deviation 0.05 to generate a trial $\boldsymbol{\lambda_k}$ set. For each ith frame, we calculate the Boltzmann weight as $w_i = \exp\{-[U(\boldsymbol{r_i}, \boldsymbol{\lambda_k}) - U(\boldsymbol{r_i}, \boldsymbol{\lambda_0})]/k_BT\}$, where U is the nonionic potential. The trial $\boldsymbol{\lambda_k}$ is discarded if the effective fraction of frames, $\phi_{eff} = \exp\left[-\sum_i^{N_{frames}} w_i \log(w_i \times N_{frames})\right]$, is lower than 60%. Otherwise, the acceptance probability follows the Metropolis criterion, $\min\left\{1, \exp\left[\frac{\mathcal{L}(\boldsymbol{\lambda}_{k-1}) - \mathcal{L}(\boldsymbol{\lambda}_k)}{\xi_k}\right]\right\}$, where ξ_k is a unitless control parameter. Each optimization cycle is divided into ten micro-cycles, wherein the control parameter, ξ, is initially set to $\xi_0 = 0.1$ and scaled down by 1% at each iteration until $\xi < 10^{-8}$. From the complete optimization cycle, we select the $\boldsymbol{\lambda}$ set yielding the lowest estimate of $\mathcal{L}$. Consecutive optimization cycles are performed from simulations runs carried out with the intermediate optimal $\boldsymbol{\lambda}$ set. To show that the procedure is reproducible and that the final $\boldsymbol{\lambda}$ set is relatively independent of the initial conditions, we performed an additional optimization procedure starting from the M1 model, λ_0 =M1 [32]. The optimization performed in this work differs from our previous implementation [32] also for the following details: (i) nine additional sequences have been included in the training set (Table S1 and S3); (ii) single chains are simulated as detailed in the "Molecular Simulations" Subsection; (v) the average radius of gyration is calculated as $\langle R_g \rangle$ instead of $\sqrt{\langle R_g^2 \rangle}$.

Code and Data Availability

Scripts and parameters to perform single-chain and direct-coexistence simulations using CALVADOS 1 and 2 are available at github.com/KULL-Centre/CALVADOS and archived on Zenodo at doi.org/10.5281/zenodo.6815068. All code and data to reproduce the results presented in this work are available at github.com/KULL-Centre/papers/tree/main/2022/CG-cutoffs-Tesei-et-al.

Acknowledgments

We thank Anna Ida Trolle for her help in setting up the protocol for single-chain simulations. We thank Rosana Collepardo, Jerelle A. Joseph, and Aleks Reinhardt for useful discussions that led us to explore the effects of cutoffs. We acknowledge access to computational resources from the ROBUST Resource for Biomolecular Simulations (supported by the Novo Nordisk Foundation; NNF18OC0032608) and Biocomputing Core Facility at the Department of Biology, University of Copenhagen. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101025063.

Supporting Figures and Tables

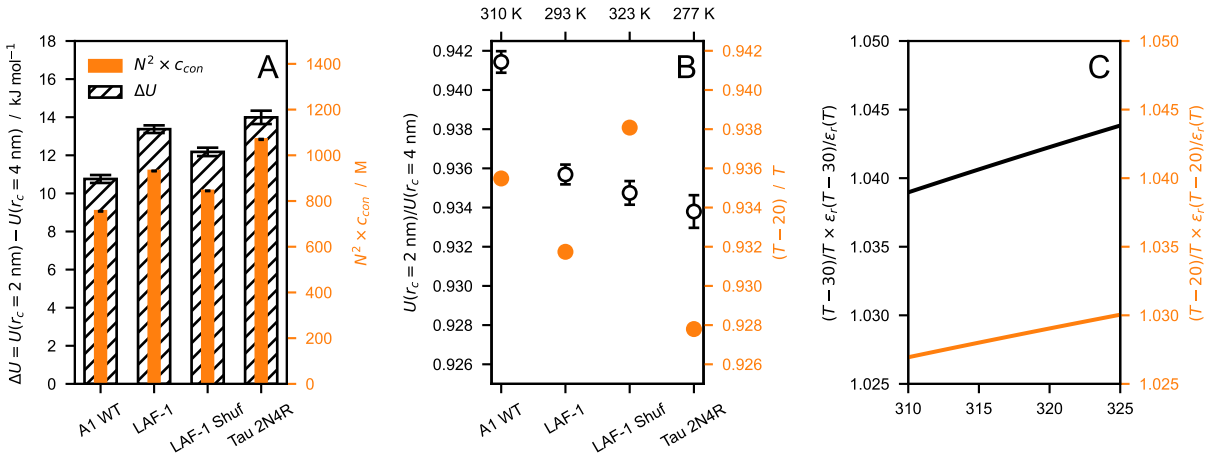

Figure S1: (*A*) Comparison between nonionic energy difference per protein ($\Delta U = U(r_c = 2\ \mathrm{nm}) - U(r_c = 4\ \mathrm{nm})$, hatched) and $N^2 \times c_{con}$ (orange), where N is the sequence length and c_{con} is the molar protein concentration in the condensate. (*B*) Ratio between nonionic energies calculated with $r_c = 2$ and 4 nm (open circles) compared to the ratio of the thermal energy at $T' = T - 20$ K and at T (orange). (*C*) Increase in electrostatic energy relative to the thermal energy upon decreasing the temperature by 30 (black) and 20 K (orange). The data shown in this figure are obtained from simulations of hnRNPA1 LCD, LAF-1 RGG domain (WT and shuffled sequence) and Tau 2N4R performed at $T = 310$, 293, 323, and 277 K, respectively, and using $r_c = 4$ nm. Error bars are standard deviations over trajectories of the systems at equilibrium.

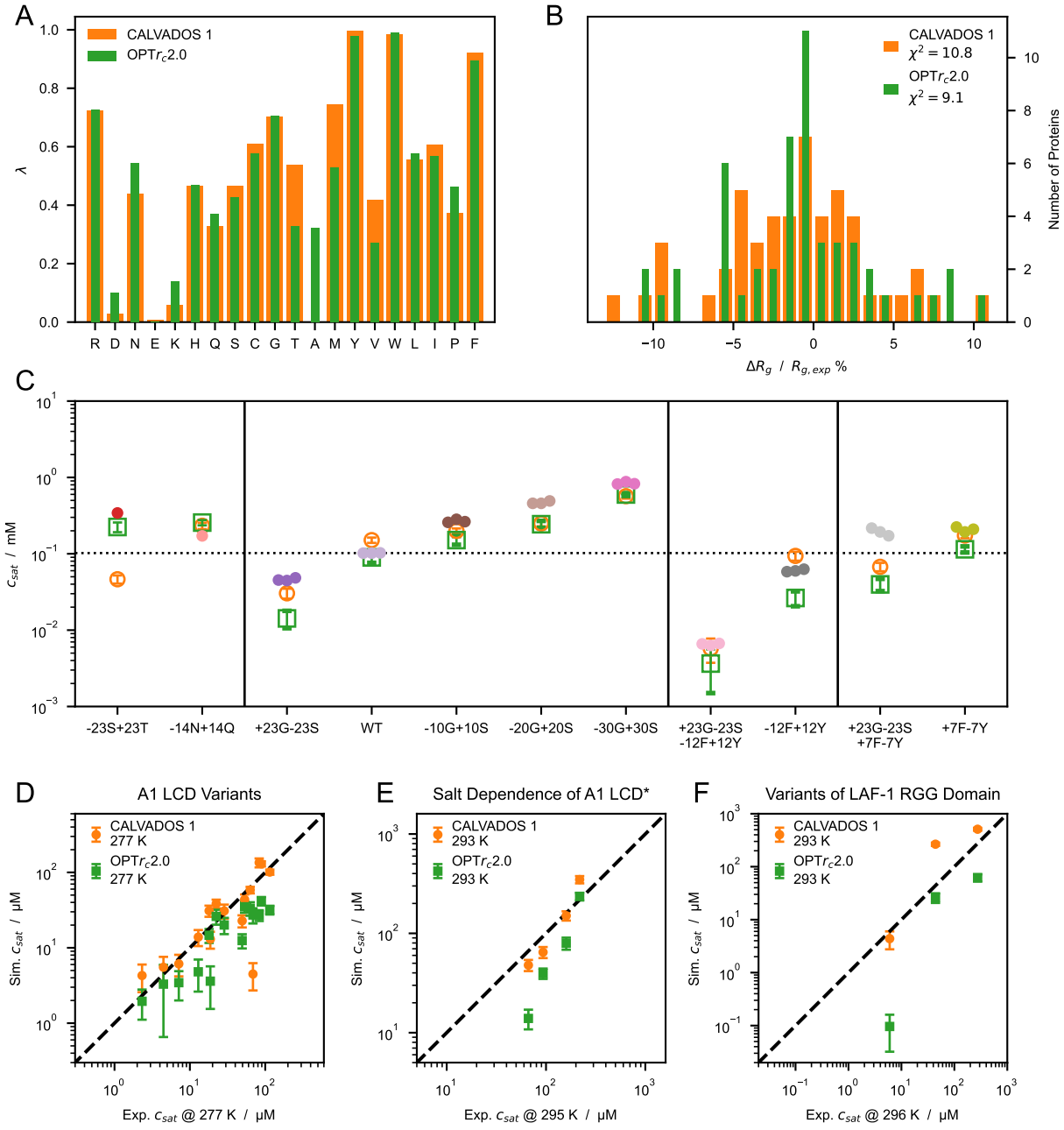

Figure S2: (*A*) Comparison between $\boldsymbol{\lambda}$ sets of CALVADOS 1 (orange) and the model resulting from the optimization with $r_c = 2.0$ nm (OPTr_c2.0, green). (*B*) Distribution of the relative difference between experimental (Table S1) and predicted radii of gyration, $\langle R_g \rangle$, for CALVADOS 1 (orange) and OPTr_c2.0 (blue). (*C*) Comparison between saturation concentrations, c_{sat}, at 293 K of variants of hnRNPA1 LCD measured by Bremer, Farag, Borcherds et al. [37] (closed circles) and corresponding predictions of CALVADOS 1 (open orange circles) and OPTr_c2.0 (open green squares). (*D–F*) Correlation between c_{sat} from simulations and experiments for (*D*) A1 LCD variants, (*E*) A1 LCD* WT at $[\mathrm{NaCl}] = 0.15$, 0.2, 0.3 and 0.5 M and (*F*) variants of LAF-1 RGG domain (Table S4).

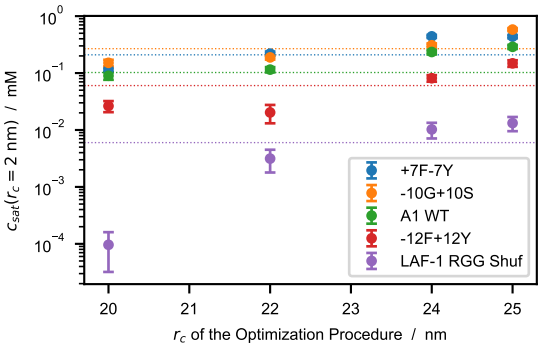

Figure S3: Saturation concentrations, c_{sat}, as a function of the cutoff used to optimize the model. c_{sat} values are calculated from simulations performed using $r_c = 2.0$ nm whereas the models are optimized using $r_c = 2.0$, 2.2, 2.4, and 2.5 nm. Horizontal dotted lines represent experimental c_{sat} values from the references reported in Table S4.

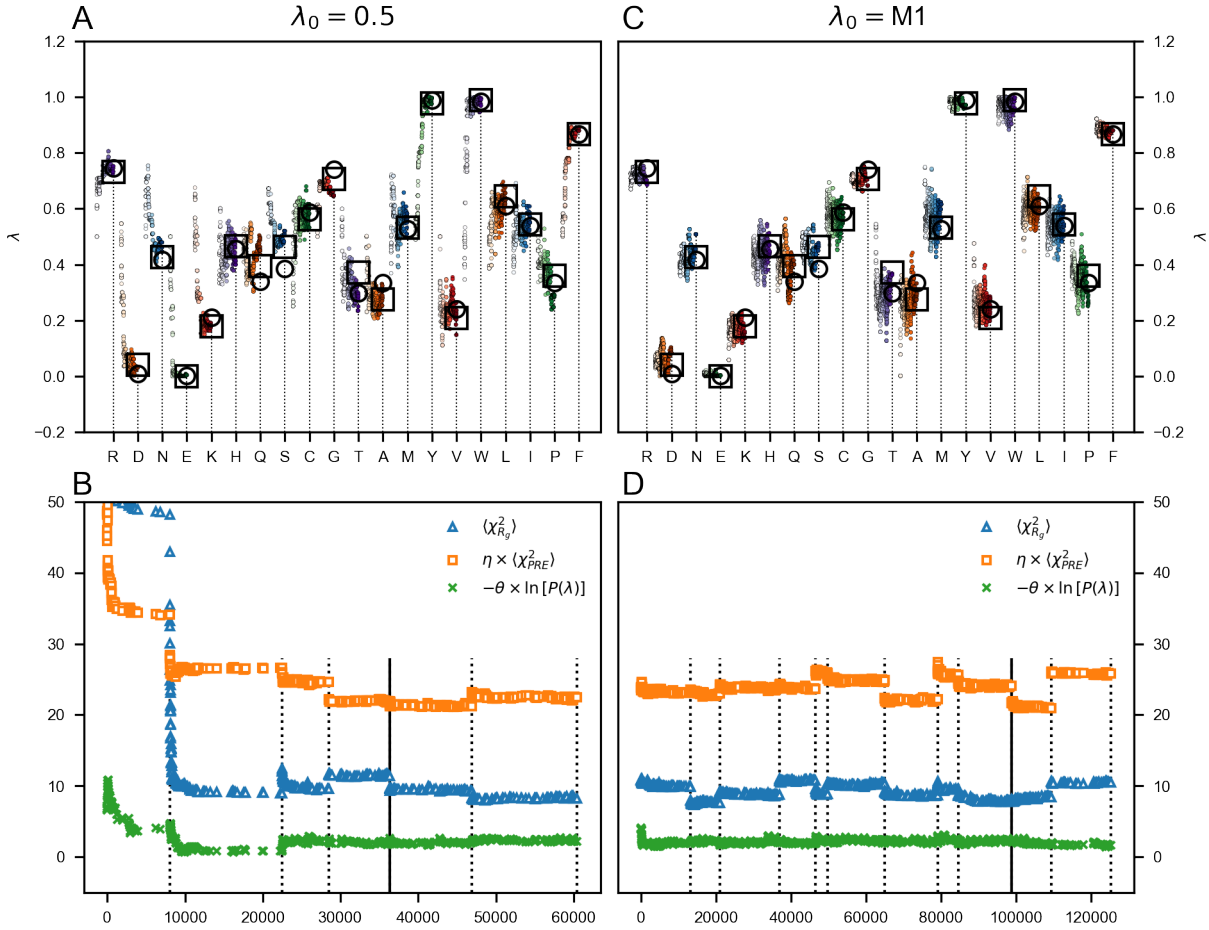

Figure S4: Optimization of the λ parameters starting from $\lambda_0 = 0.5$ (*A* and *B*) and λ_0 =M1 (*C* and *D*). (*A* and *B*) Evolution of the λ parameters during three consecutive optimization cycles. The color gradient from light to dark shade indicates increasing number of iterations. Open squares and circles show optimal λ sets obtained from indipendent optimizations starting from $\lambda_0 = 0.5$ and λ_0 =M1, respectively. (*C* and *D*) Evolution of $\chi 2_{R_g}$ (blue triangles), $0.1 \times \chi 2_{PRE}$ (orange squares), and the regularization term $0.05 \times \ln[P(\boldsymbol{\lambda})]$ (green circles). Dotted vertical lines indicate updated sampling by molecular simulations, whereas the remaining points are estimated from reweighted ensembles. Solid vertical lines indicate the optimal λ set corresponding to the lowest total cost function, $\mathcal{L}$.

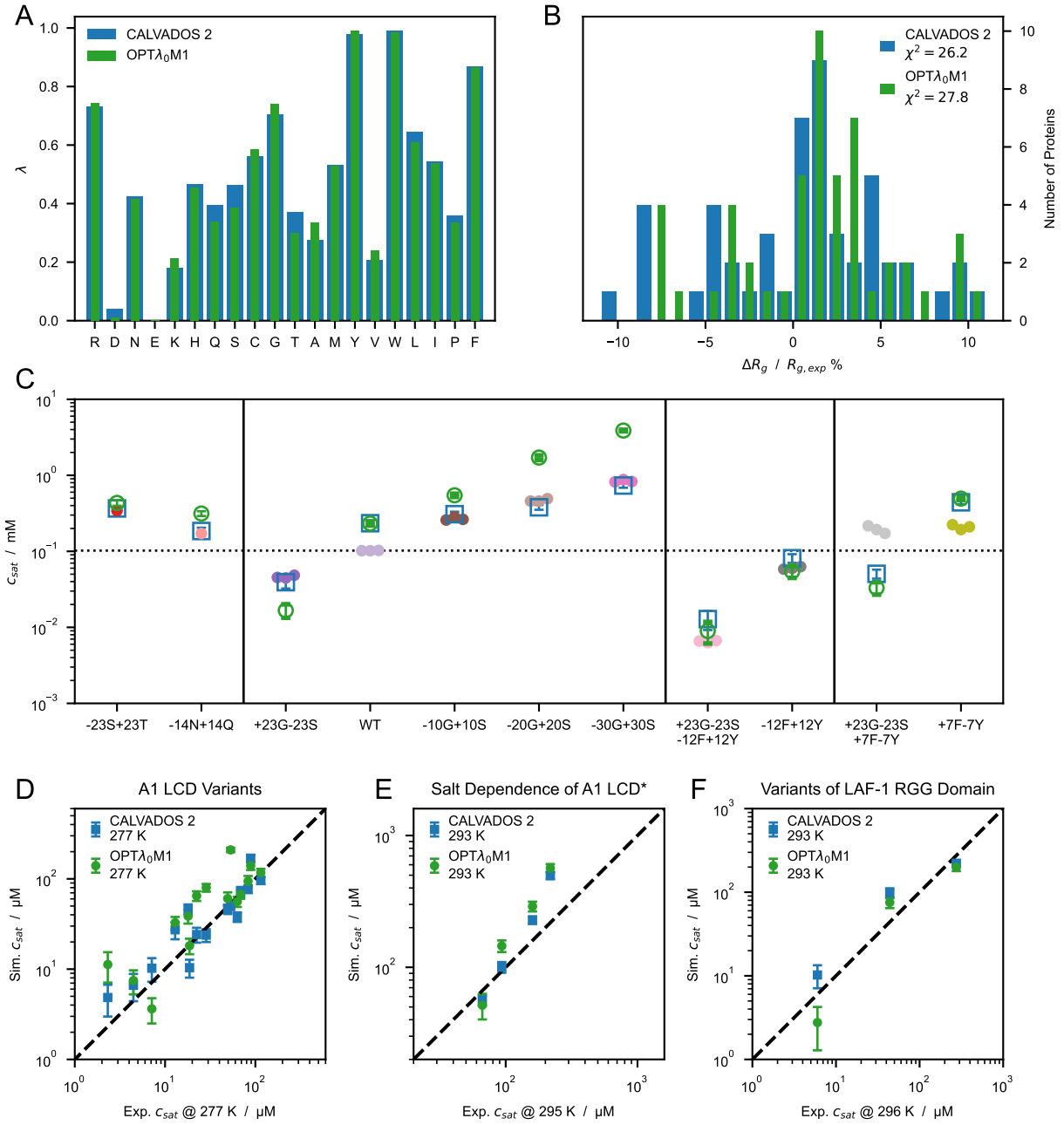

Figure S5: (*A*) Comparison between $\boldsymbol{\lambda}$ sets optimized starting from $\lambda_0 = 0.5$ (CALVADOS 2, blue) and λ_0 =M1 (OPTλ_0M1, green) using $r_c = 2.4$ nm. (*B*) Distribution of the relative difference between experimental (Table S1) and predicted radii of gyration, $\langle R_g \rangle$, for CALVADOS 2 (blue) and OPTλ_0M1 (green). (*C*) Comparison between saturation concentrations, c_{sat}, at 293 K of variants of hnRNPA1 LCD measured by Bremer, Farag, Borcherds et al. [37] (closed circles) and corresponding predictions of CALVADOS 2 (open blue squares) and OPTλ_0M1 (open green circles). (*D–F*) Correlation between c_{sat} from simulations and experiments for (*D*) A1 LCD variants, (*E*) A1 LCD* WT at $[\mathrm{NaCl}] = 0.15$, 0.2, 0.3 and 0.5 M and (*F*) variants of LAF-1 RGG domain (Table S4).

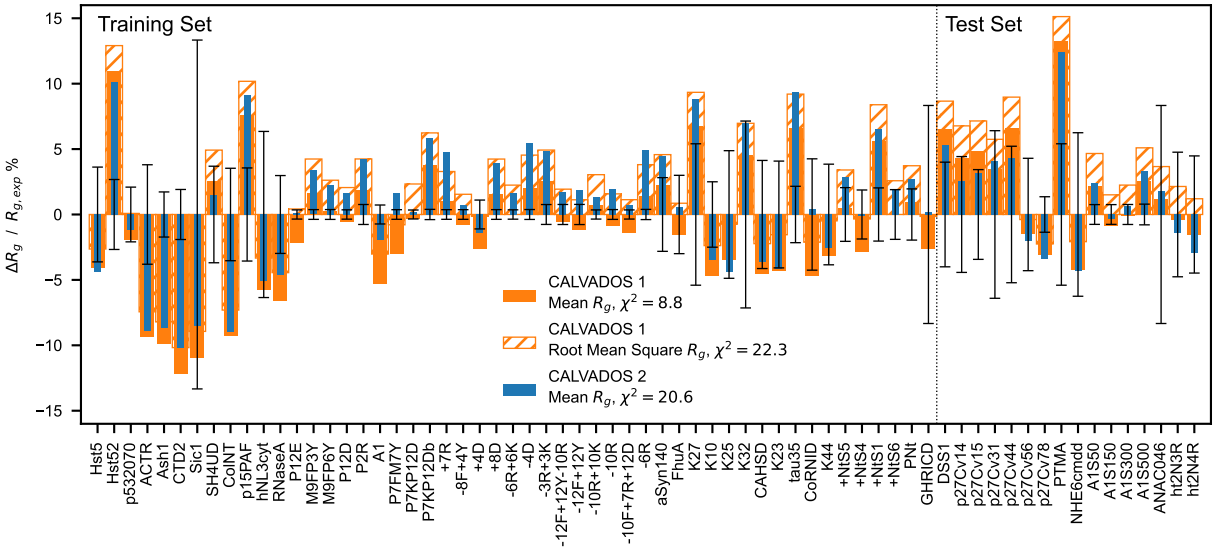

Figure S6: Relative difference between experimental and predicted radii of gyration for CALVADOS 1 (orange) and CALVADOS 2 (blue). Full and hatched bars show $(R_g^{calc} - R_g^{exp})/R_g^{exp}$ where R_g^{calc} is calculated as the mean $\langle R_g^2 \rangle$ or the root mean square $\sqrt{\langle R_g^2 \rangle}$, respectively. The vertical dashed line split the plot into the 51 sequences of the training set (Table S1) and the 16 sequences of the test set (Table S2).

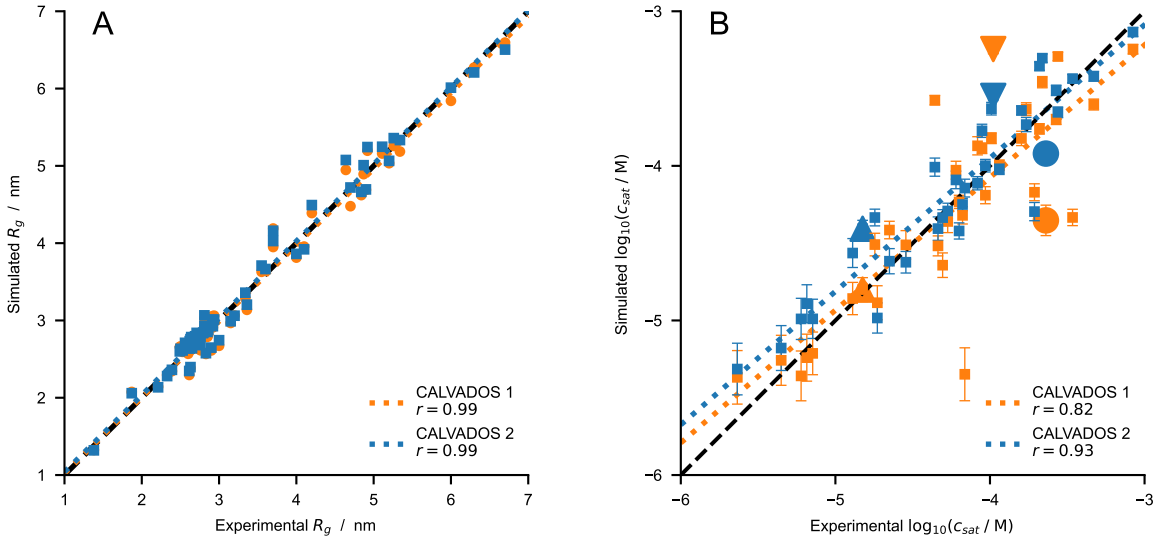

Figure S7: Comparison between experimental and predicted (*A*) R_g (Table S1 and S2) and (*B*) c_{sat} values for CALVADOS 1 (orange) and CALVADOS 2 (blue). Pearson's r coefficients are reported in the legend. Small squares in *B* show the same data as in Fig. 2*C–F* whereas the large upward triangle, downward triangle, and circle show values for A2 LCD, FUS LCD, and Ddx4 LCD, respectively, at the conditions reported in Table S4.

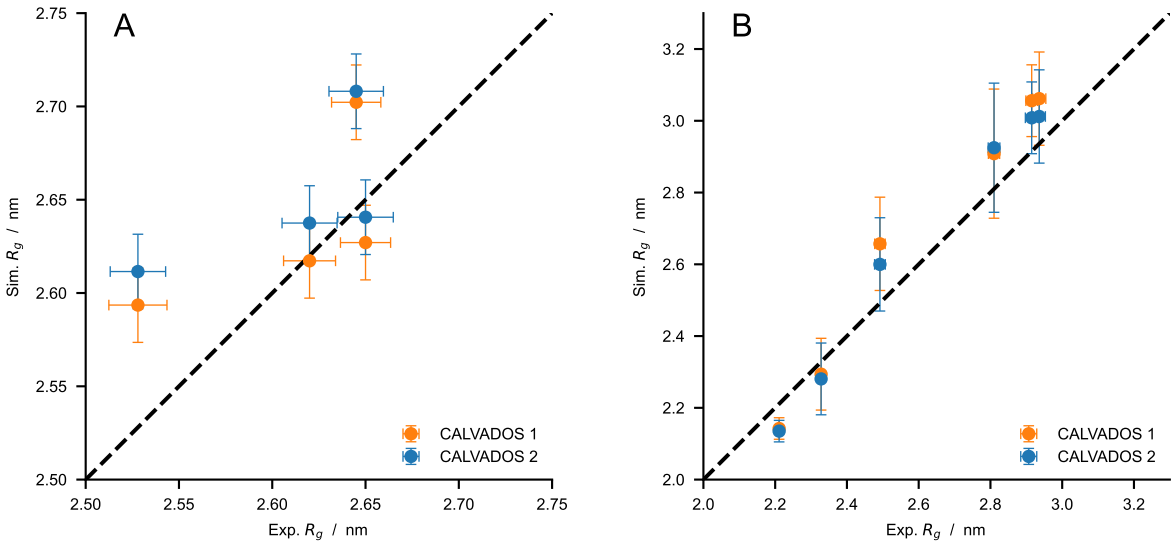

Figure S8: Comparison between experimental R_g values and predictions of CALVADOS 1 (orange) and CALVADOS 2 (blue) for (*A*) A1 LCD* at different salt concentrations ($50\ \mathrm{mM} < c_s < 500\ \mathrm{mM}$) and (*B*) p27-C constructs of different charge patterning ($0.1 < \kappa < 0.8$). Experimental conditions and references are reported in Table S2.

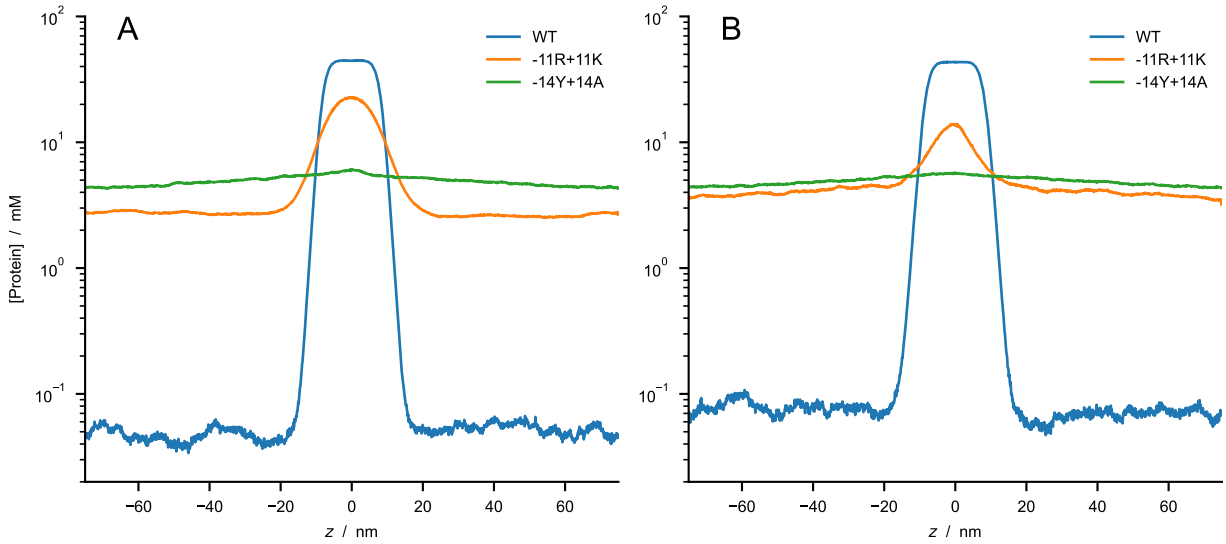

Figure S9: Predictions of CALVADOS 2 direct-coexistence simulations of the PS of constructs of the 1–80 N-terminal fragment of yeast Lge1 simulated at (*A*) $c_s = 100$ mM and (*B*) 500 mM. Protein concentration profiles are shown as a function of the long side of the simulation cell for WT (blue), -11R+11K variant (orange), and -14Y+14A variant (green).

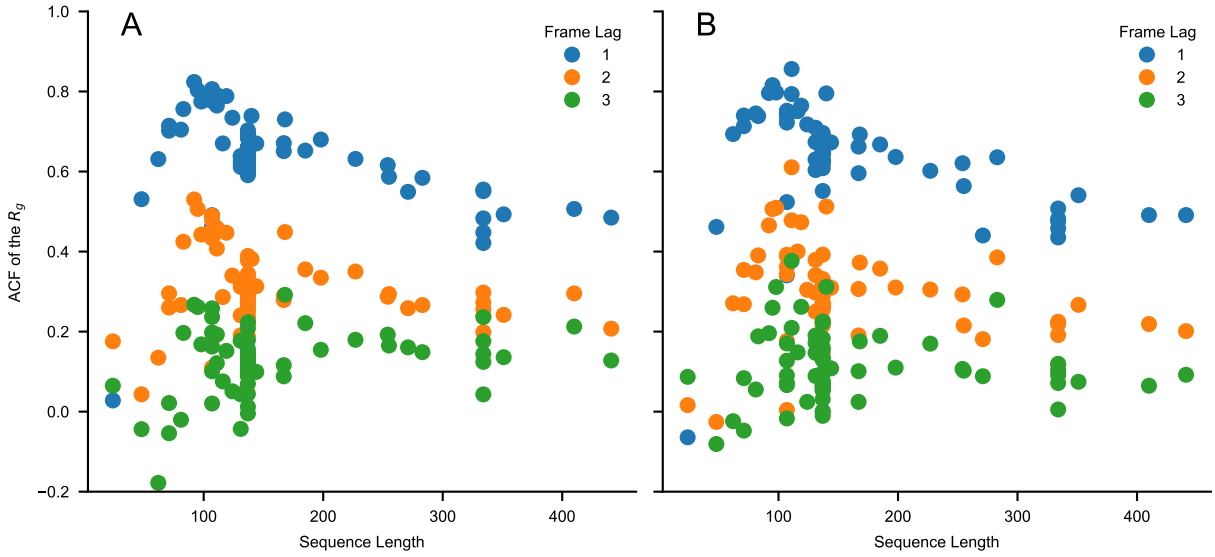

Figure S10: Values of the autocorrelation function of the R_g for a lag time of one, two and three frames as a function of sequence length, N. The autocorrelation is calculated for the sequences of Table S1 and S2 simulated using (*A*) CALVADOS 1 and (*B*) CALVADOS 2 for $\sim 6 \times 0.3 \times N^2$ ps if $N > 100$ and for 18 ns otherwise. 600 simulation frames are saved every $\sim 0.003 \times N^2$ ps if $N > 100$ and every 30 ps otherwise. The initial 100 frames are discarded from each simulation.

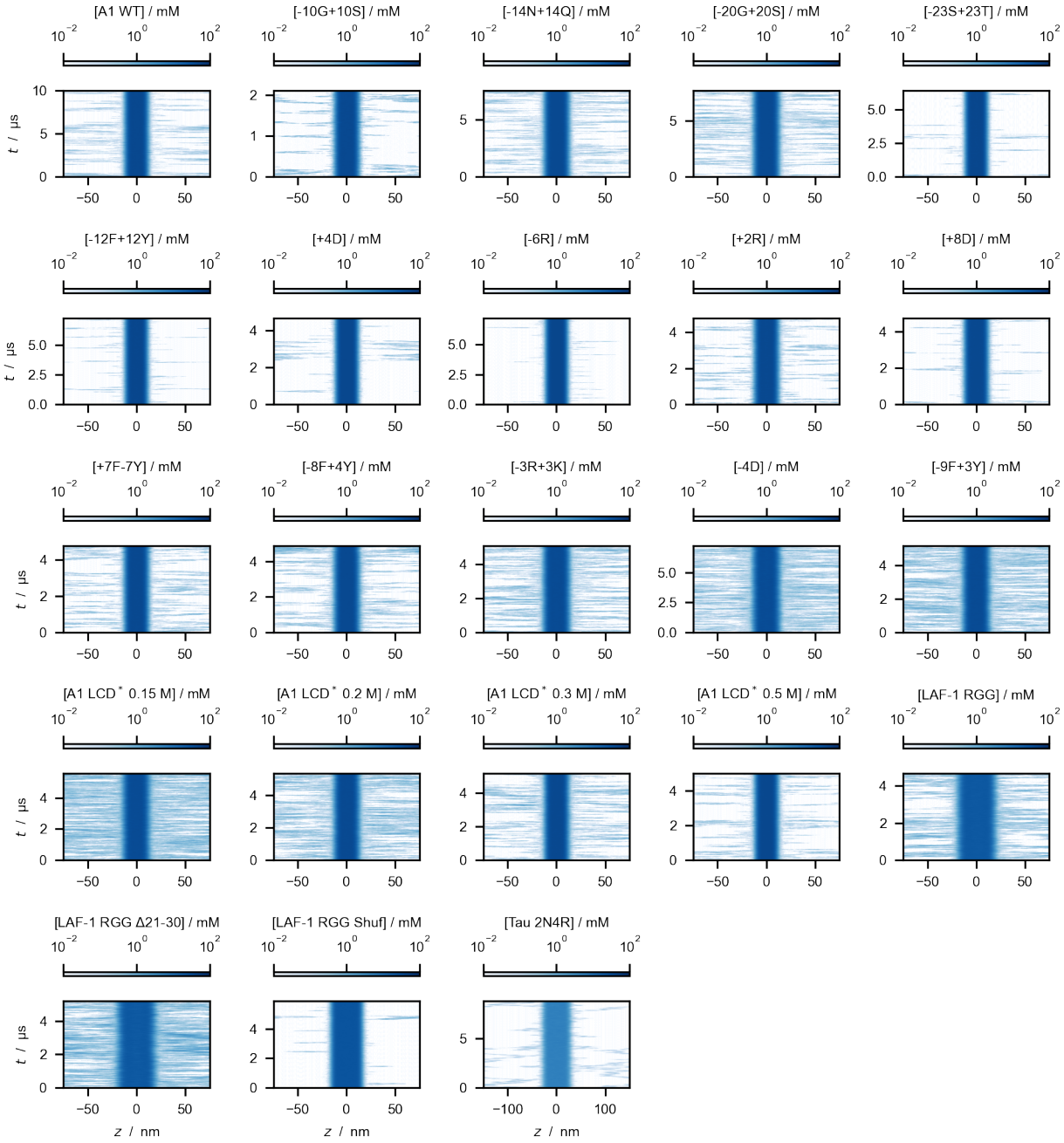

Figure S11: Time evolution of the protein concentration along the z-axis of the simulation cell, as obtained from direct-coexistence simulations performed with the CALVADOS 1 model and $r_c = 4$ nm.

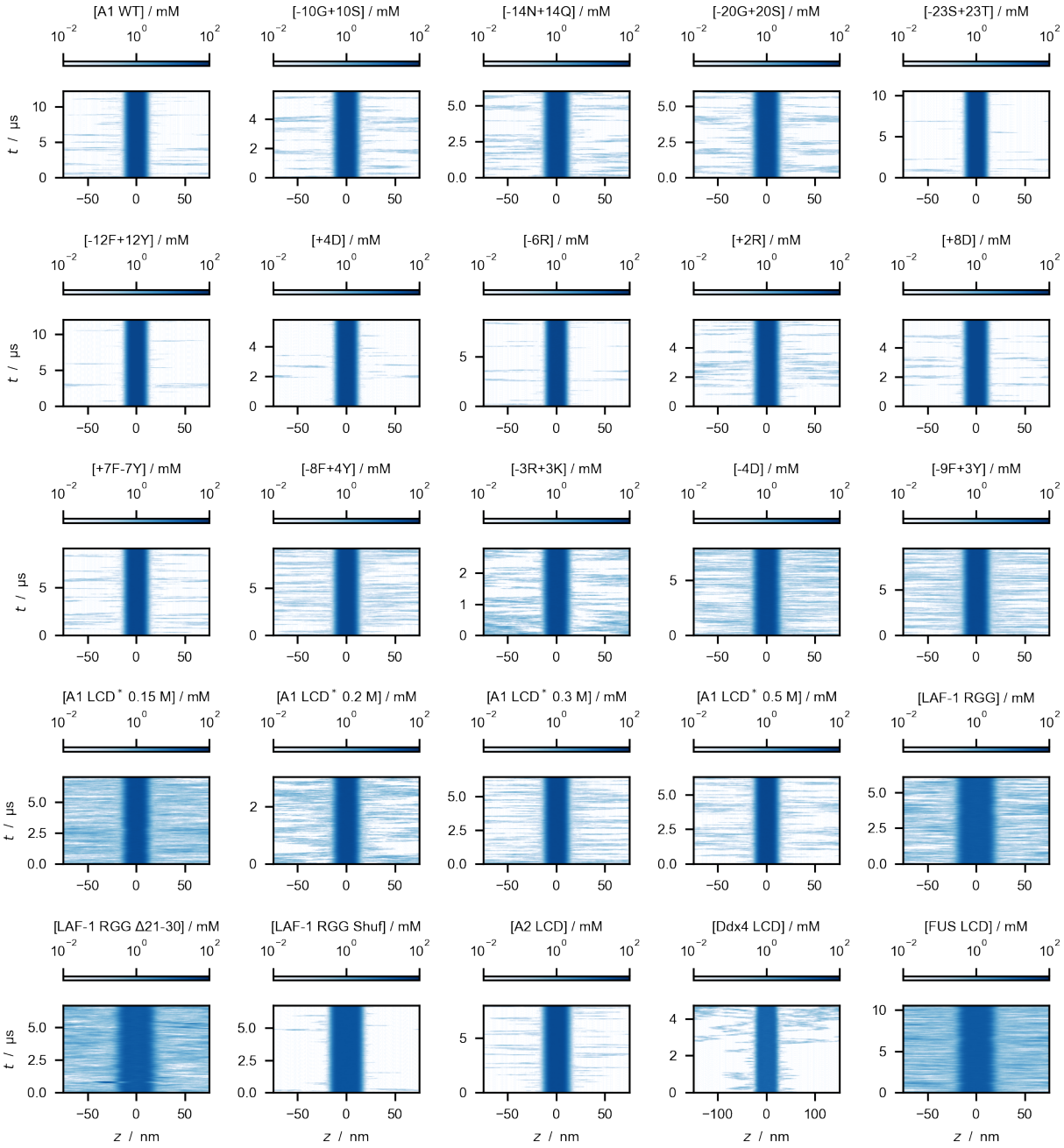

Figure S12: Time evolution of protein density along the z-axis of the simulation cell, as obtained from direct-coexistence simulations performed with the CALVADOS 1 model and $r_c = 2$ nm.

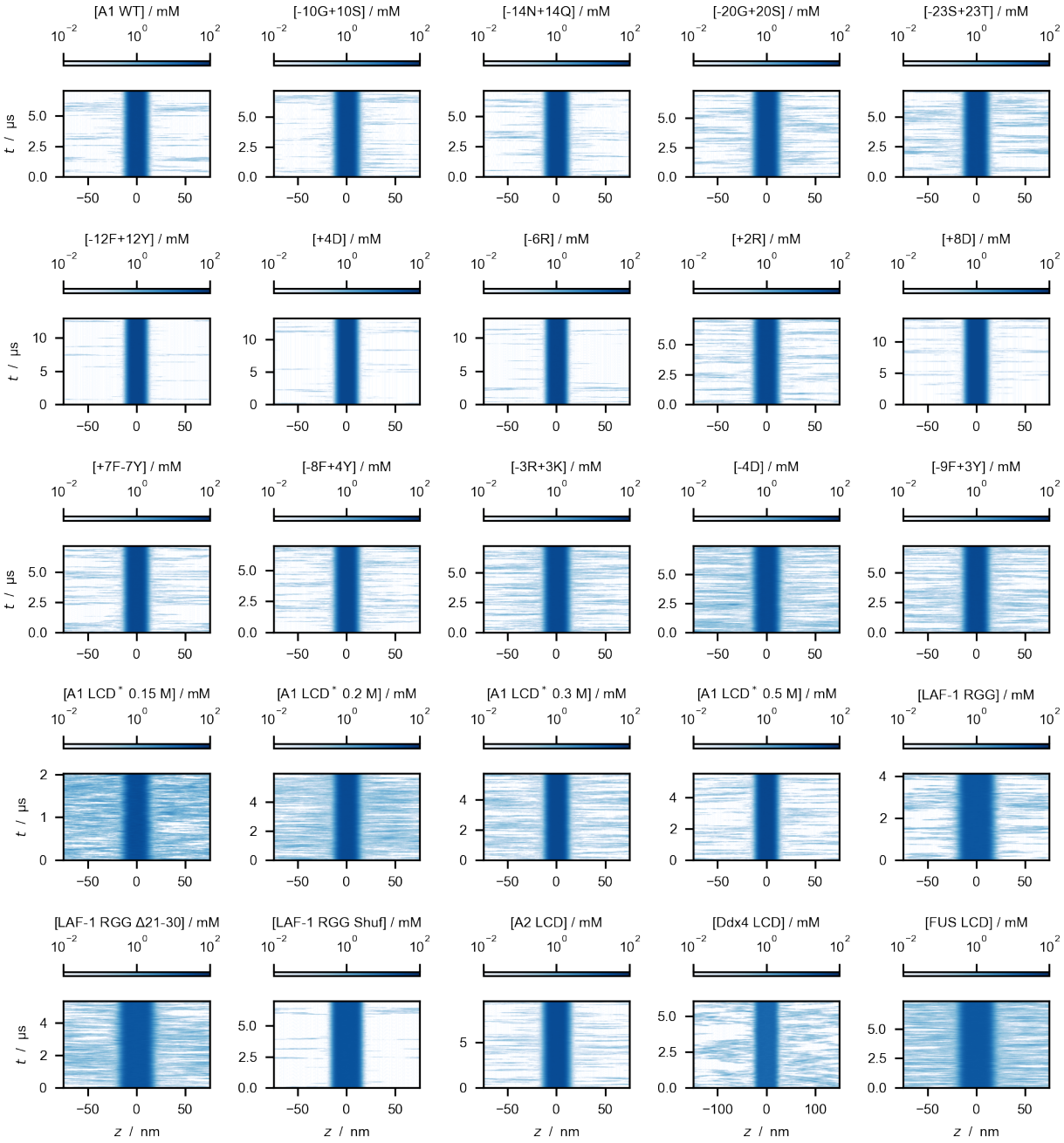

Figure S13: Time evolution of protein density along the z-axis of the simulation cell, as obtained from direct-coexistence simulations performed with the CALVADOS 2 model and $r_c = 2$ nm.

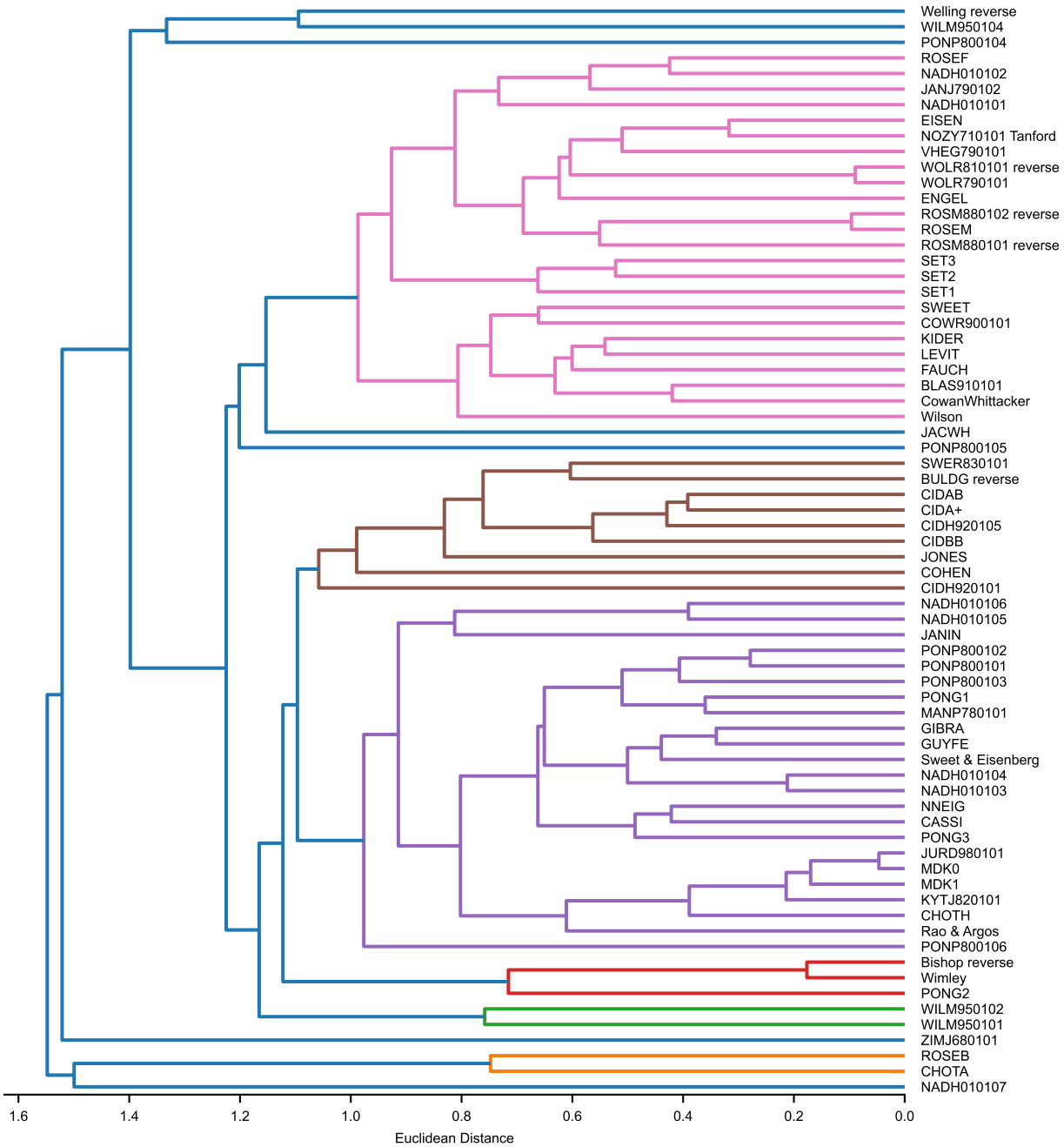

Figure S14: Hierarchical clustering dendrogram of 70 min-max normalized hydrophobicity scales selected from the set by Simm et al. [51]. Agglomerative clustering is performed using Euclidean distances and the average linkage method as implemented in the Python scikit-learn package [52].

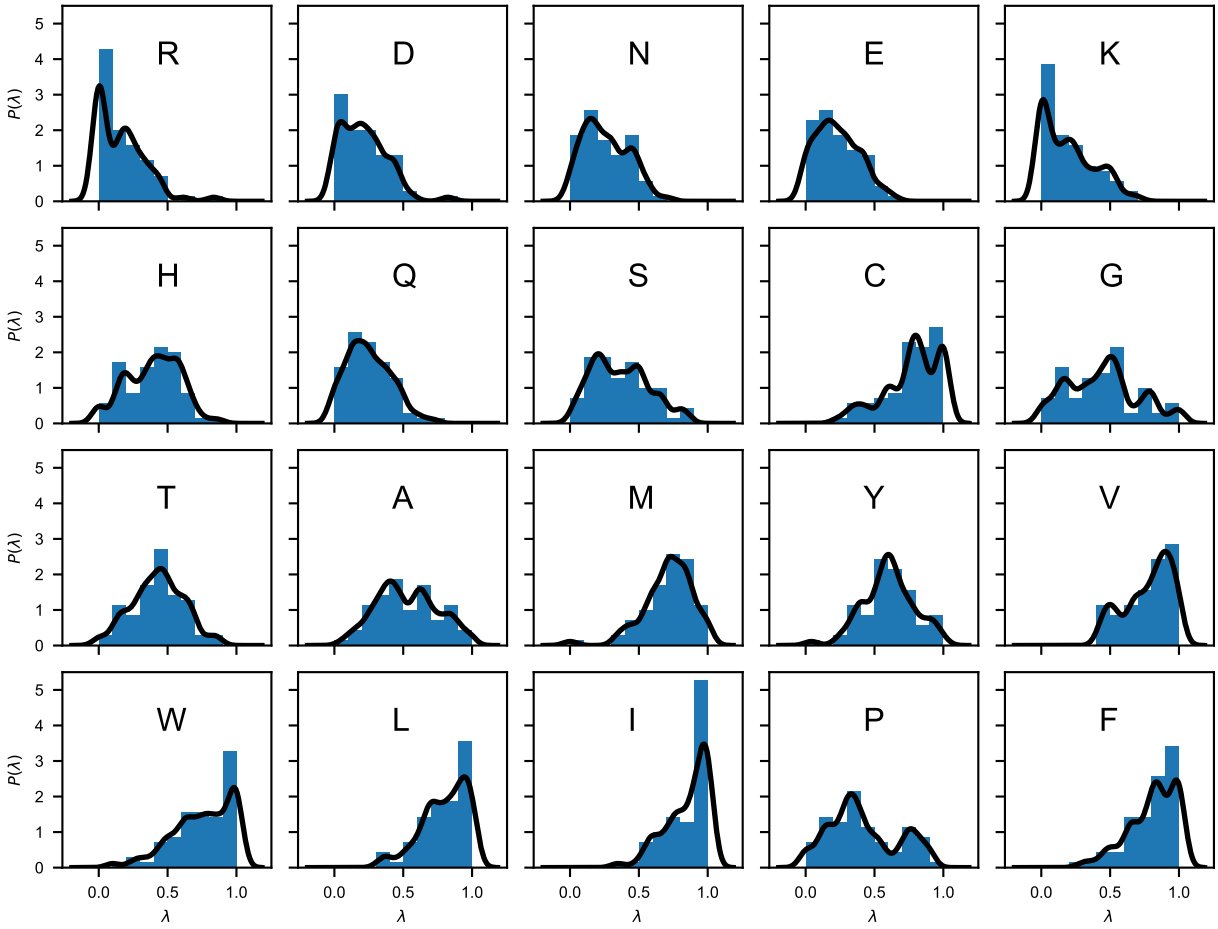

Figure S15: Probability distributions of the "stickiness" parameters, $P(\lambda)$, obtained from 70 min-max normalized hydrophobicity scales selected from the set by Simm et al. [51]. Blue bars are histograms with bin width of 0.1. Black lines are obtained as 1D projections of the multivariate kernel density estimation implemented in scikit-learn [52], using a Gaussian kernel with bandwidth of 0.05.

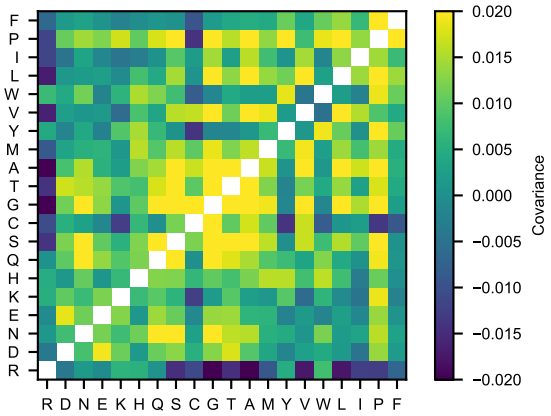

Figure S16: Covariance matrix of the 70 min-max normalized hydrophobicity scales selected from the set by Simm et al. [51]. The upper triangle of the matrix shows the covariance calculated directly from the 70 min-max normalized hydrophobicity scales whereas the lower triangle of the matrix shows the covariance calculated from the multivariate kernel density estimation averaging over the 70 min-max normalized hydrophobicity scales.

Table S1: Solution conditions and experimental radii of gyration of proteins included in the training set for the Bayesian parameter-learning procedure.

Protein	N	R_g (nm)	T (K)	c_s (M)	pH	Ref.
Hst5	24	1.38 ± 0.05	293	0.15	7.5	[53]
$(Hst5)_2$	48	1.87 ± 0.05	298	0.15	7.0	[54]
p53 (20-70)	62	2.39 ± 0.05	277	0.1	7.0	[55]
ACTR	71	2.6 ± 0.1	278	0.2	7.4	[56]
Ash1	81	2.9 ± 0.05	293	0.15	7.5	[57, 58]
CTD2	83	2.61 ± 0.05	293	0.12	7.5	[59, 58]
Sic1	92	3.0 ± 0.4	293	0.2	7.5	[60]
SH4UD	95	2.7 ± 0.1	293	0.2	8.0	[61]
ColNT	98	2.8 ± 0.1	277	0.4	7.6	[62]
p15PAF	111	2.8 ± 0.1	298	0.15	7.0	[63]
hNL3cyt	119	3.2 ± 0.2	293	0.3	8.5	[64]
RNaseA	124	3.4 ± 0.1	298	0.15	7.5	[65]
A1	137	2.76 ± 0.02	298	0.15	7.0	[37]
-10R	137	2.67 ± 0.01	298	0.15	7.0	[37]
-6R	137	2.57 ± 0.01	298	0.15	7.0	[37]
+2R	137	2.62 ± 0.02	298	0.15	7.0	[37]
+7R	137	2.71 ± 0.01	298	0.15	7.0	[37]
-3R+3K	137	2.63 ± 0.02	298	0.15	7.0	[37]
-6R+6K	137	2.79 ± 0.01	298	0.15	7.0	[37]
-10R+10K	137	2.85 ± 0.01	298	0.15	7.0	[37]
+12D	137	2.80 ± 0.01	298	0.15	7.0	[37]
+4D	137	2.72 ± 0.03	298	0.15	7.0	[37]
+8D	137	2.69 ± 0.01	298	0.15	7.0	[37]
-9F+3Y	137	2.68 ± 0.01	298	0.15	7.0	[37]
+12E	137	2.85 ± 0.01	298	0.15	7.0	[37]
+7K+12D	137	2.92 ± 0.01	298	0.15	7.0	[37]
+7K+12D blocky	137	2.56 ± 0.01	298	0.15	7.0	[37]
-4D	137	2.64 ± 0.01	298	0.15	7.0	[37]
-8F+4Y	137	2.71 ± 0.01	298	0.15	7.0	[37]
-10F+7R+12D	137	2.86 ± 0.01	298	0.15	7.0	[37]
+7F-7Y	137	2.72 ± 0.01	298	0.15	7.0	[37]
-12F+12Y	137	2.60 ± 0.02	298	0.15	7.0	[37]
-12F+12Y-10R	137	2.61 ± 0.02	298	0.15	7.0	[37]
-9F+6Y	137	2.65 ± 0.01	298	0.15	7.0	[37]
αSyn	140	3.55 ± 0.1	293	0.2	7.4	[66]
FhuA	144	3.34 ± 0.1	298	0.15	7.5	[65]
K27	167	3.70 ± 0.2	288	0.15	7.4	[67]
K10	168	4.00 ± 0.1	288	0.15	7.4	[67]
K25	185	4.10 ± 0.2	288	0.15	7.4	[67]
K32	198	4.20 ± 0.3	288	0.15	7.4	[67]
CAHSD	227	4.8 ± 0.2	293	0.07	7.0	[68]
K23	254	4.9 ± 0.2	288	0.15	7.4	[67]
Tau35	255	4.7 ± 0.1	298	0.15	7.4	[69]
CoRNID	271	4.7 ± 0.2	293	0.2	7.5	[70]
K44	283	5.2 ± 0.2	288	0.15	7.4	[67]
PNt	334	5.1 ± 0.1	298	0.15	7.5	[65, 71]
PNt Swap1	334	4.9 ± 0.1	298	0.15	7.5	[71]
PNt Swap4	334	5.3 ± 0.1	298	0.15	7.5	[71]
PNt Swap5	334	4.9 ± 0.1	298	0.15	7.5	[71]
PNt Swap6	334	5.3 ± 0.1	298	0.15	7.5	[71]
GHRICD	351	6.0 ± 0.5	298	0.35	7.3	[72, 73]

Table S2: Solution conditions and experimental radii of gyration of proteins simulated in this study but not included in the training set for the Bayesian parameter-learning procedure.

Protein	N	R_g (nm)	T (K)	c_s (M)	pH	Ref.
DSS1	71	2.5 ± 0.1	288	0.17	7.4	[73]
p27Cv14	107	2.936 ± 0.13	293	0.095	7.2	[42]
p27Cv15	107	2.915 ± 0.10	293	0.095	7.2	[42]
p27Cv31	107	2.81 ± 0.18	293	0.095	7.2	[42]
p27Cv44	107	2.492 ± 0.13	293	0.095	7.2	[42]
p27Cv56	107	2.328 ± 0.10	293	0.095	7.2	[42]
p27Cv78	107	2.211 ± 0.03	293	0.095	7.2	[42]
PTMA	111	3.7 ± 0.2	288	0.16	7.4	[73]
NHE6cmdd	116	3.2 ± 0.2	288	0.17	7.4	[73]
A1 LCD*	131	2.645 ± 0.02	293	0.05	7.5	[41]
A1 LCD*	131	2.65 ± 0.02	293	0.15	7.5	[41]
A1 LCD*	131	2.62 ± 0.02	293	0.3	7.5	[41]
A1 LCD*	131	2.528 ± 0.02	293	0.5	7.5	[41]
ANAC046	167	3.6 ± 0.3	298	0.14	7.0	[73]
Tau 2N3R	410	6.3 ± 0.3	298	0.15	7.4	[69]
Tau 2N4R	441	6.7 ± 0.3	298	0.15	7.4	[69]

Table S3: Experimental conditions for the intramolecular PRE data included in the training set.

Protein	N	N_{labels}	$\omega_I/2\pi$ (MHz)	T (K)	c_s (M)	pH	Ref.
FUS	163	3	850	298	0.15	5.5	[2]
FUS12E	164	3	850	298	0.15	5.5	[2]
OPN	220	10	800	298	0.15	6.5	[74]
αSyn	140	5	700	283	0.2	7.4	[75]
A2	155	2	850	298	0.005	5.5	[3]

Table S4: Conditions used for the direct-coexistence simulations performed in this study and references to the experimental data. Shaded rows highlight systems which are not included in the correlation plot of Fig. S7*B* because of lack of experimental c_{sat} values.

Protein	N	c_s (mM)	pH	Ref.	T (K)		
					4 nm	2 nm	Fig. 1*C*
6His-TEV-Lge1$_{1-80}$-StrepII WT	114	500	7.5	[43]	-	293	-
6His-TEV-Lge1$_{1-80}$-StrepII -11R+11K	114	500	7.5	[43]	-	293	-
6His-TEV-Lge1$_{1-80}$-StrepII -14Y+14A	114	500	7.5	[43]	-	293	-
A1 LCD WT	137	150	7.0	[4, 37]	310 & 323	277 & 293	310
A1 LCD +7F-7Y	137	150	7.0	[37]	310 & 323	277 & 293	-
A1 LCD -12F+12Y	137	150	7.0	[37]	310 & 323	277 & 293	-
A1 LCD -23S+23T	137	150	7.0	[37]	310 & 323	277 & 293	-
A1 LCD -14N+14Q	137	150	7.0	[37]	310 & 323	277 & 293	-
A1 LCD -10G+10S	137	150	7.0	[37]	310 & 323	277 & 293	-
A1 LCD -20G+20S	137	150	7.0	[37]	310 & 323	277 & 293	-
A1 LCD -30G+30S	137	150	7.0	[37]	323	293	-
A1 LCD +23G-23S	137	150	7.0	[37]	323	293	-
A1 LCD +23G-23S+7F-7Y	137	150	7.0	[37]	323	293	-
A1 LCD +23G-23S-12F+12Y	137	150	7.0	[37]	323	293	-
A1 LCD -9F+3Y	137	150	7.0	[37]	310	277	-
A1 LCD -8F+4Y	137	150	7.0	[37]	310	277	-
A1 LCD -3R+3K	137	150	7.0	[37]	310	277	-
A1 LCD -6R	137	150	7.0	[37]	310	277	-
A1 LCD -4D	137	150	7.0	[37]	310	277	-
A1 LCD +4D	137	150	7.0	[37]	310	277	-
A1 LCD +8D	137	150	7.0	[37]	310	277	-
A1 LCD +2R	137	150	7.0	[37]	310	277	-
A1 LCD* WT	131	150	7.0	[76]	323	293	-
A1 LCD* WT	131	200	7.0	[76]	323	293	-
A1 LCD* WT	131	300	7.0	[76]	323	293	-
A1 LCD* WT	131	500	7.0	[76]	323	293	-
LAF-1 RGG Domain	176	150	7.5	[77]	323	293	293
LAF-1 RGG Domain Shuffled	176	150	7.5	[77]	323	293	323
LAF-1 RGG Domain Δ21-30	166	150	7.5	[77]	323	293	-
A2 LCD	155	10	5.5	[78]	-	297	-
FUS LCD	163	150	7.4	[79]	-	297	-
Ddx4 LCD	236	130	6.5	[80]	-	297	-
Human Full-Length Tau (2N4R)	441	70	7.4	-	-	-	277

References

[1] Jie Wang, Jeong-Mo Choi, Alex S. Holehouse, Hyun O. Lee, Xiaojie Zhang, Marcus Jahnel, Shovamayee Maharana, Régis Lemaitre, Andrei Pozniakovsky, David Drechsel, Ina Poser, Rohit V. Pappu, Simon Alberti, and Anthony A. Hyman. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell*, 174(3):688–699.e16, July 2018. doi:10.1016/j.cell.2018.06.006.

[2] Zachary Monahan, Veronica H Ryan, Abigail M Janke, Kathleen A Burke, Shannon N Rhoads, Gül H Zerze, Robert O'Meally, Gregory L Dignon, Alexander E Conicella, Wenwei Zheng, Robert B Best, Robert N Cole, Jeetain Mittal, Frank Shewmaker, and Nicolas L Fawzi. Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *The EMBO Journal*, 36(20):2951–2967, August 2017. doi:10.15252/embj.201696394.

[3] Veronica H. Ryan, Gregory L. Dignon, Gül H. Zerze, Charlene V. Chabata, Rute Silva, Alexander E. Conicella, Joshua Amaya, Kathleen A. Burke, Jeetain Mittal, and Nicolas L. Fawzi. Mechanistic view of hnRNPA2 low-complexity domain structure, interactions, and phase separation altered by mutation and arginine methylation. *Molecular Cell*, 69(3):465–479.e7, February 2018. doi:10.1016/j.molcel.2017.12.022.

[4] Erik W. Martin, Alex S. Holehouse, Ivan Peran, Mina Farag, J. Jeremias Incicco, Anne Bremer, Christy R. Grace, Andrea Soranno, Rohit V. Pappu, and Tanja Mittag. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science*, 367(6478):694–699, February 2020. doi:10.1126/science.aaw8653.

[5] Tanja Mittag and Rohit V. Pappu. A conceptual framework for understanding phase separation and addressing open questions and challenges. *Molecular Cell*, 82(12):2201–2214, June 2022. URL: https://doi.org/10.1016%2Fj.molcel.2022.05.018, doi:10.1016/j.molcel.2022.05.018.

[6] Avinash Patel, Hyun O. Lee, Louise Jawerth, Shovamayee Maharana, Marcus Jahnel, Marco Y. Hein, Stoyno Stoynov, Julia Mahamid, Shambaditya Saha, Titus M. Franzmann, Andrej Pozniakovski, Ina Poser, Nicola Maghelli, Loic A. Royer, Martin Weigert, Eugene W. Myers, Stephan Grill, David Drechsel, Anthony A. Hyman, and Simon Alberti. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell*, 162(5):1066–1077, August 2015. doi:10.1016/j.cell.2015.07.047.

[7] Tetsuro Murakami, Seema Qamar, Julie Qiaojin Lin, Gabriele S. Kaminski Schierle, Eric Rees, Akinori Miyashita, Ana R. Costa, Roger B. Dodd, Fiona T.S. Chan, Claire H. Michel, Deborah Kronenberg-Versteeg, Yi Li, Seung-Pil Yang, Yosuke Wakutani, William Meadows, Rodylyn Rose Ferry, Liang Dong, Gian Gaetano Tartaglia, Giorgio Favrin, Wen-Lang Lin, Dennis W. Dickson, Mei Zhen, David Ron, Gerold Schmitt-Ulms, Paul E. Fraser, Neil A. Shneider, Christine Holt, Michele Vendruscolo, Clemens F. Kaminski, and Peter St George-Hyslop. ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. *Neuron*, 88(4):678–690, November 2015. doi:10.1016/j.neuron.2015.10.030.

[8] Nicholas M. Kanaan, Chelsey Hamel, Tessa Grabinski, and Benjamin Combs. Liquid-liquid phase separation induces pathogenic tau conformations in vitro. *Nature Communications*, 11(1):2809, June 2020. doi:10.1038/s41467-020-16580-3.

[9] Susanne Wegmann, Bahareh Eftekharzadeh, Katharina Tepper, Katarzyna M Zoltowska, Rachel E Bennett, Simon Dujardin, Pawel R Laskowski, Danny MacKenzie, Tarun Kamath, Caitlin Commins, Charles Vanderburg, Allyson D

Roe, Zhanyun Fan, Amandine M Molliex, Amayra Hernandez-Vega, Daniel Muller, Anthony A Hyman, Eckhard Mandelkow, J Paul Taylor, and Bradley T Hyman. Tau protein liquid–liquid phase separation can initiate tau aggregation. *The EMBO Journal*, 37(7):e98049, February 2018. doi:10.15252/embj.201798049.

[10] Thomas R. Peskett, Frédérique Rau, Jonathan O'Driscoll, Rickie Patani, Alan R. Lowe, and Helen R. Saibil. A liquid to solid phase transition underlying pathological huntingtin exon1 aggregation. *Molecular Cell*, 70(4):588–601.e6, May 2018. doi:10.1016/j.molcel.2018.04.007.

[11] Soumik Ray, Nitu Singh, Rakesh Kumar, Komal Patel, Satyaprakash Pandey, Debalina Datta, Jaladhar Mahato, Rajlaxmi Panigrahi, Ambuja Navalkar, Surabhi Mehra, Laxmikant Gadhe, Debdeep Chatterjee, Ajay Singh Sawner, Siddhartha Maiti, Sandhya Bhatia, Juan Atilio Gerez, Arindam Chowdhury, Ashutosh Kumar, Ranjith Padinhateeri, Roland Riek, G. Krishnamoorthy, and Samir K. Maji. α-synuclein aggregation nucleates through liquid–liquid phase separation. *Nature Chemistry*, 12(8):705–716, June 2020. doi:10.1038/s41557-020-0465-9.

[12] Maarten C Hardenberg, Tessa Sinnige, Sam Casford, Samuel Dada, Chetan Poudel, Elizabeth A Robinson, Monika Fuxreiter, Clemens Kaminksi, Gabriele S Kaminski-Schierle, Ellen A A Nollen, Christopher M Dobson, and Michele Vendruscolo. Observation of an α-synuclein liquid droplet state and its maturation into lewy body-like assemblies. *Journal of Molecular Cell Biology*, January 2021. doi:10.1093/jmcb/mjaa075.

[13] Jitao Wen, Liu Hong, Georg Krainer, Qiong-Qiong Yao, Tuomas P. J. Knowles, Si Wu, and Sarah Perrett. Conformational expansion of tau in condensates promotes irreversible aggregation. *Journal of the American Chemical Society*, 143(33):13056–13064, 2021. doi:10.1021/jacs.1c03078.

[14] Samuel T. Dada, Maarten C. Hardenberg, Lena K. Mrugalla, Mollie O. McKeon, Ewa Klimont, Thomas C. T. Michaels, and Michele Vendruscolo. Spontaneous nucleation and fast aggregate-dependent proliferation of α-synuclein aggregates within liquid condensates at physiological ph. *bioRxiv*, 2021. doi:10.1101/2021.09.26.461836.

[15] Bin Wang, Lei Zhang, Tong Dai, Ziran Qin, Huasong Lu, Long Zhang, and Fangfang Zhou. Liquid-liquid phase separation in human health and diseases. *Signal Transduction and Targeted Therapy*, 6(1), August 2021. doi:10.1038/s41392-021-00678-1.

[16] Jiahua Lu, Junjie Qian, Zhentian Xu, Shengyong Yin, Lin Zhou, Shusen Zheng, and Wu Zhang. Emerging roles of liquid-liquid phase separation in cancer: From protein aggregation to immune-associated signaling. *Frontiers in Cell and Developmental Biology*, 9, June 2021. doi:10.3389/fcell.2021.631486.

[17] Jeong Hyun Ahn, Eric S. Davis, Timothy A. Daugird, Shuai Zhao, Ivana Yoseli Quiroga, Hidetaka Uryu, Jie Li, Aaron J. Storey, Yi-Hsuan Tsai, Daniel P. Keeley, Samuel G. Mackintosh, Ricky D. Edmondson, Stephanie D. Byrum, Ling Cai, Alan J. Tackett, Deyou Zheng, Wesley R. Legant, Douglas H. Phanstiel, and Gang Greg Wang. Phase separation drives aberrant chromatin looping and cancer development. *Nature*, 595(7868):591–595, June 2021. doi:10.1038/s41586-021-03662-5.

[18] Salman F. Banani, Lena K. Afeyan, Susana W. Hawken, Jonathan E. Henninger, Alessandra Dall'Agnese, Victoria E. Clark, Jesse M. Platt, Ozgur Oksuz, Nancy M. Hannett, Ido Sagi, Tong Ihn Lee, and Richard A. Young. Genetic variation associated with condensate dysregulation in disease. *Developmental Cell*, 2022. doi:https://doi.org/10.1016/j.devcel.2022.06.010.

[19] Nicolas L. Fawzi, Sapun H. Parekh, and Jeetain Mittal. Biophysical studies of phase separation integrating experimental and computational methods. *Current Opinion in Structural Biology*, 70:78–86, October 2021. doi:10.1016/j.sbi.2021.04.004.

[20] Gregory L. Dignon, Wenwei Zheng, Young C. Kim, Robert B. Best, and Jeetain Mittal. Sequence determinants of protein phase behavior from a coarse-grained model. *PLOS Computational Biology*, 14(1):e1005941, January 2018. doi:10.1371/journal.pcbi.1005941.

[21] Jeong-Mo Choi, Furqan Dar, and Rohit V. Pappu. LASSI: A lattice model for simulating phase transitions of multivalent proteins. *PLOS Computational Biology*, 15(10):e1007028, October 2019. doi:10.1371/journal.pcbi.1007028.

[22] Alex S. Holehouse and Rohit V. Pappu. Pimms (0.24 pre-beta), December 2019. doi:10.5281/zenodo.3588456.

[23] Alex S. Holehouse, Garrett M. Ginell, Daniel Griffith, and Elvan Böke. Clustering of aromatic residues in prion-like domains can tune the formation, state, and organization of biomolecular condensates. *Biochemistry*, 60(47):3566–3581, November 2021. doi:10.1021/acs.biochem.1c00465.

[24] Mrityunjoy Kar, Furqan Dar, Timothy J. Welsh, Laura T. Vogel, Ralf Kühnemuth, Anupa Majumdar, Georg Krainer, Titus M. Franzmann, Simon Alberti, Claus A. M. Seidel, Tuomas P. J. Knowles, Anthony A. Hyman, and Rohit V. Pappu. Phase-separating RNA-binding proteins form heterogeneous distributions of clusters in subsaturated solutions. *Proceedings of the National Academy of Sciences*, 119(28), July 2022. doi:10.1073/pnas.2202222119.

[25] Anne Bremer, Mina Farag, Wade M. Borcherds, Ivan Peran, Erik W. Martin, Rohit V. Pappu, and Tanja Mittag. Deciphering how naturally occurring sequence features impact the phase behaviors of disordered prion-like domains. *bioRxiv*, 2021. doi:10.1101/2021.01.01.425046.

[26] Mina Farag, Samuel R. Cohen, Wade M. Borcherds, Anne Bremer, Tanja Mittag, and Rohit V. Pappu. Condensates of disordered proteins have small-world network structures and interfaces defined by expanded conformations. *bioRxiv*, 2022. doi:10.1101/2022.05.21.492916.

[27] Suman Das, Yi-Hsuan Lin, Robert M. Vernon, Julie D. Forman-Kay, and Hue Sun Chan. Comparative roles of charge, π, and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences*, 117(46):28795–28805, November 2020. doi:10.1073/pnas.2008122117.

[28] Andrew P. Latham and Bin Zhang. Consistent force field captures homologue-resolved hp1 phase separation. *Journal of Chemical Theory and Computation*, 17(5):3134–3144, 2021. doi:10.1021/acs.jctc.0c01220.

[29] Thomas Dannenhoffer-Lafage and Robert B. Best. A data-driven hydrophobicity scale for predicting liquid–liquid phase separation of proteins. *The Journal of Physical Chemistry B*, 125(16):4046–4056, April 2021. doi:10.1021/acs.jpcb.0c11479.

[30] Roshan Mammen Regy, Jacob Thompson, Young C. Kim, and Jeetain Mittal. Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins. *Protein Science*, n/a(n/a), May 2021. doi:10.1002/pro.4094.

[31] Jerelle A. Joseph, Aleks Reinhardt, Anne Aguirre, Pin Yu Chew, Kieran O. Russell, Jorge R. Espinosa, Adiran Garaizar, and Rosana Collepardo-Guevara. Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nature Computational Science*, 1(11):732–743, November 2021. doi:10.1038/s43588-021-00155-3.

[32] Giulio Tesei, Thea K. Schulze, Ramon Crehuet, and Kresten Lindorff-Larsen. Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proceedings of the National Academy of Sciences*, 118(44), 2021. doi:10.1073/pnas.2111696118.

[33] Anders B. Norgaard, Jesper Ferkinghoff-Borg, and Kresten Lindorff-Larsen. Experimental parameterization of an energy function for the simulation of unfolded proteins. *Biophysical Journal*, 94(1):182–192, January 2008. doi:10.1529/biophysj.107.108241.

[34] Andrés R. Tejedor, Adiran Garaizar, Jorge Ramírez, and Jorge R. Espinosa. 'rna modulation of transport properties and stability in phase-separated condensates. *Biophysical Journal*, 120(23):5169–5186, 2021. doi:https://doi.org/10.1016/j.bpj.2021.11.003.

[35] Suman Das, Alan N. Amin, Yi-Hsuan Lin, and Hue Sun Chan. Coarse-grained residue-based models of disordered protein condensates: utility and limitations of simple charge pattern parameters. *Phys. Chem. Chem. Phys.*, 20:28558–28574, 2018. doi:10.1039/C8CP05095C.

[36] Xipeng Wang, Simón Ramírez-Hinestrosa, Jure Dobnikar, and Daan Frenkel. The lennard-jones potential: when (not) to use it. *Physical Chemistry Chemical Physics*, 22(19):10624–10633, 2020. doi:10.1039/c9cp05445f.

[37] Anne Bremer, Mina Farag, Wade M. Borcherds, Ivan Peran, Erik W. Martin, Rohit V. Pappu, and Tanja Mittag. Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nature Chemistry*, December 2021. doi:10.1038/s41557-021-00840-w.

[38] Davide Mercadante, Johannes A. Wagner, Iker V. Aramburu, Edward A. Lemke, and Frauke Gräter. Sampling long- versus short-range interactions defines the ability of force fields to reproduce the dynamics of intrinsically disordered proteins. *Journal of Chemical Theory and Computation*, 13(9):3964–3974, August 2017. URL: https://doi.org/10.1021%2Facs.jctc.7b00143, doi:10.1021/acs.jctc.7b00143.

[39] Ibraheem Alshareedah, Taranpreet Kaur, Jason Ngo, Hannah Seppala, Liz-Audrey Djomnang Kounatse, Wei Wang, Mahdi Muhammad Moosa, and Priya R. Banerjee. Interplay between short-range attraction and long-range repulsion controls reentrant liquid condensation of ribonucleoprotein–RNA complexes. *Journal of the American Chemical Society*, 141(37):14593–14602, August 2019. doi:10.1021/jacs.9b03689.

[40] Milan Kumar Hazra and Yaakov Levy. Biophysics of phase separation of disordered proteins is governed by balance between short- and long-range interactions. *The Journal of Physical Chemistry B*, 125(9):2202–2211, February 2021. doi:10.1021/acs.jpcb.0c09975.

[41] Erik W Martin, F Emil Thomasen, Nicole M Milkovic, Matthew J Cuneo, Christy R Grace, Amanda Nourse, Kresten Lindorff-Larsen, and Tanja Mittag. Interplay of folded domains and the disordered low-complexity domain in mediating hnRNPA1 phase separation. *Nucleic Acids Research*, 49(5):2931–2945, February 2021. doi:10.1093/nar/gkab063.

[42] Rahul K. Das, Yongqi Huang, Aaron H. Phillips, Richard W. Kriwacki, and Rohit V. Pappu. Cryptic sequence features within the disordered protein p27^{Kip1} regulate cell cycle signaling. *Proceedings of the National Academy of Sciences*, 113(20):5616–5621, May 2016. doi:10.1073/pnas.1516277113.

[43] Anton A. Polyansky, Laura D. Gallego, Roman G. Efremov, Alwin Köhler, and Bojan Zagrovic. Protein compactness and interaction valency define the architecture of a biomolecular condensate across scales. *bioRxiv*, 2022. doi:10.1101/2022.02.18.481017.

[44] Joshua A. Anderson, Jens Glaser, and Sharon C. Glotzer. HOOMD-blue: A python package for high-performance molecular dynamics and hard particle monte carlo simulations. *Computational Materials Science*, 173:109363, February 2020. doi:10.1016/j.commatsci.2019.109363.

[45] Peter Eastman, Jason Swails, John D. Chodera, Robert T. McGibbon, Yutong Zhao, Kyle A. Beauchamp, Lee-Ping Wang, Andrew C. Simmonett, Matthew P. Harrigan, Chaya D. Stern, Rafal P. Wiewiora, Bernard R. Brooks, and Vijay S. Pande. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology*, 13(7):e1005659, July 2017. doi:10.1371/journal.pcbi.1005659.

[46] H. Flyvbjerg and H. G. Petersen. Error estimates on averages of correlated data. *The Journal of Chemical Physics*, 91(1):461–466, July 1989. doi:10.1063/1.457480.

[47] Henry S. Ashbaugh and Harold W. Hatch. Natively unfolded protein stability as a coil-to-globule transition in charge/hydropathy space. *Journal of the American Chemical Society*, 130(29):9536–9542, July 2008. doi:10.1021/ja802124e.

[48] Young C. Kim and Gerhard Hummer. Coarse-grained models for simulations of multiprotein complexes: Application to ubiquitin binding. *Journal of Molecular Biology*, 375(5):1416–1433, February 2008. doi:10.1016/j.jmb.2007.11.063.

[49] G. C. Akerlof and H. I. Oshry. The dielectric constant of water at high temperatures and in equilibrium with its vapor. *Journal of the American Chemical Society*, 72(7):2844–2847, July 1950. doi:10.1021/ja01163a006.

[50] Hidetada Nagai, Ko Kuwabara, and Giorgio Carta. Temperature dependence of the dissociation constants of several amino acids. *Journal of Chemical & Engineering Data*, 53(3):619–627, March 2008. doi:10.1021/je700067a.

[51] Stefan Simm, Jens Einloft, Oliver Mirus, and Enrico Schleiff. 50 years of amino acid hydrophobicity scales: revisiting the capacity for peptide classification. *Biological Research*, 49(1), July 2016. doi:10.1186/s40659-016-0092-5.

[52] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[53] S. Jephthah, L. Staby, B. B. Kragelund, and M. Skepö. Temperature dependence of intrinsically disordered proteins in simulations: What are we missing? *Journal of Chemical Theory and Computation*, 15(4):2672–2683, March 2019. doi:10.1021/acs.jctc.8b01281.

[54] Eric Fagerberg, Linda K. Månsson, Samuel Lenton, and Marie Skepö. The effects of chain length on the structural properties of intrinsically disordered proteins in concentrated solutions. *The Journal of Physical Chemistry B*, 124(52):11843–11853, December 2020. doi:10.1021/acs.jpcb.0c09635.

[55] Jing Zhao, Alan Blayney, Xiaorong Liu, Lauren Gandy, Weihua Jin, Lufeng Yan, Jeung-Hoi Ha, Ashley J. Canning, Michael Connelly, Chao Yang, Xinyue Liu, Yuanyuan Xiao, Michael S. Cosgrove, Sozanne R. Solmaz, Yingkai Zhang, David Ban, Jianhan Chen, Stewart N. Loh, and Chunyu Wang. EGCG binds intrinsically disordered n-terminal domain of p53 and disrupts p53-MDM2 interaction. *Nature Communications*, 12(1), February 2021. doi:10.1038/s41467-021-21258-5.

[56] Magnus Kjaergaard, Ann–Beth Nørholm, Ruth Hendus–Altenburger, Stine F. Pedersen, Flemming M. Poulsen, and Birthe B. Kragelund. Temperature-dependent structural changes in intrinsically disordered proteins: Formation of α-helices or loss of polyproline II? *Protein Science*, 19(8):1555–1564, 2010. doi:https://doi.org/10.1002/pro.435.

[57] Erik W. Martin, Alex S. Holehouse, Christy R. Grace, Alex Hughes, Rohit V. Pappu, and Tanja Mittag. Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation. *Journal of the American Chemical Society*, 138(47):15323–15335, 2016. doi:10.1021/jacs.6b10272.

[58] Fan Jin and Frauke Gräter. How multisite phosphorylation impacts the conformations of intrinsically disordered proteins. *PLOS Computational Biology*, 17(5):e1008939, May 2021. doi:10.1371/journal.pcbi.1008939.

[59] Eric B. Gibbs, Feiyue Lu, Bede Portz, Michael J. Fisher, Brenda P. Medellin, Tatiana N. Laremore, Yan Jessie Zhang, David S. Gilmour, and Scott A. Showalter. Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II c-terminal domain. *Nature Communications*, 8(1), May 2017. doi:10.1038/ncomms15233.

[60] Gregory-Neal W. Gomes, Mickaël Krzeminski, Ashley Namini, Erik W. Martin, Tanja Mittag, Teresa Head-Gordon, Julie D. Forman-Kay, and Claudiu C. Gradinaru. Conformational ensembles of an intrinsically disordered protein consistent with NMR, SAXS, and single-molecule FRET. *Journal of the American Chemical Society*, 142(37):15697–15710, August 2020. doi:10.1021/jacs.0c02088.

[61] Utsab R. Shrestha, Puneet Juneja, Qiu Zhang, Viswanathan Gurumoorthy, Jose M. Borreguero, Volker Urban, Xiaolin Cheng, Sai Venkatesh Pingali, Jeremy C. Smith, Hugh M. O'Neill, and Loukas Petridis. Generation of the configurational ensemble of an intrinsically disordered protein from unbiased molecular dynamics simulation. *Proceedings of the National Academy of Sciences*, 116(41):20446–20452, September 2019. doi:10.1073/pnas.1907251116.

[62] Christopher L. Johnson, Alexandra S. Solovyova, Olli Hecht, Colin Macdonald, Helen Waller, J. Günter Grossmann, Geoffrey R. Moore, and Jeremy H. Lakey. The two-state prehensile tail of the antibacterial toxin colicin n. *Biophysical Journal*, 113(8):1673–1684, October 2017. doi:10.1016/j.bpj.2017.08.030.

[63] Alfredo De Biasio, Alain Ibáñez de Opakua, Tiago N. Cordeiro, Maider Villate, Nekane Merino, Nathalie Sibille, Moreno Lelli, Tammo Diercks, Pau Bernadó, and Francisco J. Blanco. p15paf is an intrinsically disordered protein with nonrandom structural preferences at sites of interaction with other proteins. *Biophysical Journal*, 106(4):865–874, February 2014. doi:10.1016/j.bpj.2013.12.046.

[64] Aviv Paz, Tzviya Zeev-Ben-Mordehai, Martin Lundqvist, Eilon Sherman, Efstratios Mylonas, Lev Weiner, Gilad Haran, Dmitri I. Svergun, Frans A.A. Mulder, Joel L. Sussman, and Israel Silman. Biophysical characterization of the unstructured cytoplasmic domain of the human neuronal adhesion protein neuroligin 3. *Biophysical Journal*, 95(4):1928–1944, August 2008. doi:10.1529/biophysj.107.126995.

[65] Joshua A. Riback, Micayla A. Bowman, Adam M. Zmyslowski, Catherine R. Knoverek, John M. Jumper, James R. Hinshaw, Emily B. Kaye, Karl F. Freed, Patricia L. Clark, and Tobin R. Sosnick. Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science*, 358(6360):238–241, October 2017. doi:10.1126/science.aan5774.

[66] Mustapha Carab Ahmed, Line K. Skaanning, Alexander Jussupow, Estella A. Newcombe, Birthe B. Kragelund, Carlo Camilloni, Annette E. Langkilde, and Kresten Lindorff-Larsen. Refinement of α-synuclein ensembles against SAXS data: Comparison of force fields and methods. *Frontiers in Molecular Biosciences*, 8, April 2021. doi:10.3389/fmolb.2021.654333.

[67] Efstratios Mylonas, Antje Hascher, Pau Bernadó, Martin Blackledge, Eckhard Mandelkow, and Dmitri I. Svergun. Domain conformation of tau protein studied by solution small-angle x-ray scattering. *Biochemistry*, 47(39):10345–10353, September 2008. doi:10.1021/bi800900d.

[68] Cherie Hesgrove, Kenny H. Nguyen, Sourav Biswas, Charles A. Childs, KC Shraddha, Bryan X. Medina, Vladimir Alvarado, Feng Yu, Shahar Sukenik, Marco Malferrari, Francesco Francia, Giovanni Venturoli, Erik W. Martin, Alex S. Holehouse, and Thomas C. Boothby. Tardigrade cahs proteins act as molecular swiss army knives to mediate desiccation tolerance through multiple mechanisms. *bioRxiv*, 2021. doi:10.1101/2021.08.16.456555.

[69] Chen Lyu, Stefano Da Vela, Youssra Al-Hilaly, Karen E. Marshall, Richard Thorogate, Dmitri Svergun, Louise C. Serpell, Annalisa Pastore, and Diane P. Hanger. The disease associated tau35 fragment has an increased propensity to aggregate compared to full-length tau. *Frontiers in Molecular Biosciences*, 8, 2021. doi:10.3389/fmolb.2021.779240.

[70] Tiago N. Cordeiro, Nathalie Sibille, Pierre Germain, Philippe Barthe, Abdelhay Boulahtouf, Fréderic Allemand, Rémy Bailly, Valérie Vivat, Christine Ebel, Alessandro Barducci, William Bourguet, Albane le Maire, and Pau Bernadó. Interplay of protein disorder in retinoic acid receptor heterodimer and its corepressor regulates gene expression. *Structure*, 27(8):1270–1285.e6, August 2019. doi:10.1016/j.str.2019.05.001.

[71] Micayla A. Bowman, Joshua A. Riback, Anabel Rodriguez, Hongyu Guo, Jun Li, Tobin R. Sosnick, and Patricia L. Clark. Properties of protein unfolded states suggest broad selection for expanded conformational ensembles. *Proceedings of the National Academy of Sciences*, 117(38):23356–23364, September 2020. doi:10.1073/pnas.2003773117.

[72] Pernille Seiffert, Katrine Bugge, Mads Nygaard, Gitte W Haxholm, Jacob H Martinsen, Martin N Pedersen, Lise Arleth, Wouter Boomsma, and Birthe B Kragelund. Orchestration of signaling by structural disorder in class 1 cytokine receptors. *Cell Communication and Signaling*, 18(1):1–30, 2020.

[73] Francesco Pesce, Estella A. Newcombe, Pernille Seiffert, Emil E. Tranchant, Johan G. Olsen, Birthe B. Kragelund, and Kresten Lindorff-Larsen. Assessment of models for calculating the hydrodynamic radius of intrinsically disordered proteins. *bioRxiv*, 2022. doi:10.1101/2022.06.11.495732.

[74] D. Kurzbach, A. Vanas, A. G. Flamm, N. Tarnoczi, G. Kontaxis, N. Maltar-Strmečki, K. Widder, D. Hinderberger, and R. Konrat. Detection of correlated conformational fluctuations in intrinsically disordered proteins through paramagnetic relaxation interference. *Physical Chemistry Chemical Physics*, 18(8):5753–5758, 2016. doi:10.1039/c5cp04858c.

[75] Matthew M. Dedmon, Kresten Lindorff-Larsen, John Christodoulou, Michele Vendruscolo, and Christopher M. Dobson. Mapping long-range interactions in α-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *Journal of the American Chemical Society*, 127(2):476–477, January 2005. doi:10.1021/ja044834j.

[76] Erik W. Martin, Tyler S. Harmon, Jesse B. Hopkins, Srinivas Chakravarthy, J. Jeremías Incicco, Peter Schuck, Andrea Soranno, and Tanja Mittag. A multi-step nucleation process determines the kinetics of prion-like domain phase separation. *Nature Communications*, 12(1), July 2021. doi:10.1038/s41467-021-24727-z.

[77] Benjamin S. Schuster, Gregory L. Dignon, Wai Shing Tang, Fleurie M. Kelley, Aishwarya Kanchi Ranganath, Craig N. Jahnke, Alison G. Simpkins, Roshan Mammen Regy, Daniel A. Hammer, Matthew C. Good, and Jeetain Mittal. Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. *Proceedings of the National Academy of Sciences*, 117(21):11421–11431, May 2020. doi:10.1073/pnas.2000223117.

[78] Veronica H Ryan, Theodora M Perdikari, Mandar T Naik, Camillo F Saueressig, Jeremy Lins, Gregory L Dignon, Jeetain Mittal, Anne C Hart, and Nicolas L Fawzi. Tyrosine phosphorylation regulates hnRNPA2 granule protein partitioning and reduces neurodegeneration. *The EMBO Journal*, 40(3), December 2020. doi:10.15252/embj.2020105001.

[79] Anastasia C. Murthy, Gregory L. Dignon, Yelena Kan, Gül H. Zerze, Sapun H. Parekh, Jeetain Mittal, and Nicolas L. Fawzi. Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. *Nature Structural & Molecular Biology*, 26(7):637–648, July 2019. doi:10.1038/s41594-019-0250-x.

[80] Jacob P. Brady, Patrick J. Farber, Ashok Sekhar, Yi-Hsuan Lin, Rui Huang, Alaji Bah, Timothy J. Nott, Hue Sun Chan, Andrew J. Baldwin, Julie D. Forman-Kay, and Lewis E. Kay. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings of the National Academy of Sciences*, 114(39):E8194–E8203, September 2017. doi:10.1073/pnas.1706197114.