Supporting Information

Dignon et al. 10.1073/pnas.1804177115

1. Supporting Methods

Sequences of the proteins used in this work. (Mutated or permuted residues shown in red except for shuffled LAF-1 sequence.)

FUS LC WT

```
MASNDYTQQA TQSYGAYPTQ PGQGYSQQSS QPYGQQSYSG YSQSTDTSGY GQSSYSSYGQ SQNTGYGTQS TPQGYGSTGG YGSSQSSQSS YGQQSSYPGY
GQQPAPSSTS GSYGSSSQSS SYGQPQSGSY SQQPSYGGQQ QSYGQQQSYN PPQGYGQQNQ YNS
```

FUS 6E

```
MASNDYTQQA TQSYGAYPTQ PGQGYEQQSE QPYGQQSYSG YSQSTDTSGY GQSSYSSYGQ SQNTGYGEQS TPQGYGSTGG YGSEQSEQSS YGQQSSYPGY
GQQPAPSSTS GSYGSSEQSS SYGQPQSGSY SQQPSYGGQQ QSYGQQQSYN PPQGYGQQNQ YNS
```

FUS 6E'

```
MASNDYTQQA TQSYGAYPEQ PGQGYEQQSE QPYGQQSYSG YEQSTDTSGY GQSSYSSYGQ EQNTGYGTQS TPQGYGSTGG YGSEQSSQSS YGQQSSYPGY
GQQPAPSSTS GSYGSSSQSS SYGQPQSGSY SQQPSYGGQQ QSYGQQQSYN PPQGYGQQNQ YNS
```

FUS 6E*

```
MASNDYEQQA TQSYGAYPTQ PGQGYEQQSS QPYGQQSYSG YSQSTDTSGY GQSSYSSYGQ SQNTGYGTQS TPQGYGSTGG YGSEQSEQSS YGQQSSYPGY
GQQPAPSSTS GSYGSSEQSS SYGQPQSGSY EQQPSYGGQQ QSYGQQQSYN PPQGYGQQNQ YNS
```

FUS 12E

```
MASNDYEQQA EQSYGAYPEQ PGQGYEQQSE QPYGQQSYSG YEQSTDTSGY GQSSYSSYGQ EQNTGYGEQS TPQGYGSTGG YGSEQSEQSS YGQQSSYPGY
GQQPAPSSTS GSYGSSEQSS SYGQPQSGSY EQQPSYGGQQ QSYGQQQSYN PPQGYGQQNQ YNS
```

FUS40

```
MASNDYTQQA TQSYGAYPTQ PGQGYSQQSS QPYGQQSYSG
```

FUS YtoF

```
MASNDFTQQA TQSFGAFPTQ PGQGFSQQSS QPFGQQSFSG FSQSTDTSGF GQSSFSSFGQ SQNTGFGTQS TPQGFGSTGG FGSSQSSQSS FGQQSSFPGF
GQQPAPSSTS GSFGSSSQSS SFGQPQSGSF SQQPSFGGQQ QSFGQQQSFN PPQGFGQQNQ FNS
```

hnRNPA2 CTD WT

```
GRGGNFGFGD SRGGGGNFGP GPGSNFRGGS DGYGSGRGFG DGYNGYGGGP GGGNFGGSPG YGGGRGGYGG GGPGYGNQGG GYGGGYDNYG GGNYGSGNYN
DFGNYNQQPS NYGPMKSGNF GGSRNMGGPY GGGNYGPGGS GGSGGYGGRS RY
```

hnRNPA2 CTD D290V

```
GRGGNFGFGD SRGGGGNFGP GPGSNFRGGS DGYGSGRGFG DGYNGYGGGP GGGNFGGSPG YGGGRGGYGG GGPGYGNQGG GYGGGYDNYG GGNYGSGNYN
VFGNYNQQPS NYGPMKSGNF GGSRNMGGPY GGGNYGPGGS GGSGGYGGRS RY
```

hnRNPA2 CTD P298L

```
GRGGNFGFGD SRGGGGNFGP GPGSNFRGGS DGYGSGRGFG DGYNGYGGGP GGGNFGGSPG YGGGRGGYGG GGPGYGNQGG GYGGGYDNYG GGNYGSGNYN
DFGNYNQQLS NYGPMKSGNF GGSRNMGGPY GGGNYGPGGS GGSGGYGGRS RY
```

LAF-1 IDR WT

```
MESNQSNNGG SGNAALNRGG RYVPPHLRGG DGGAAAAASA GGDDRRGGAG GGGYRRGGGN SGGGGGGGYD RGYNDNRDDR DNRGGSGGYG RDRNYEDRGY
NGGGGGGGNR GYNNNRGGGG GGYNRQDRGD GGSSNFSRGG YNNRDEGSDN RGSGRSYNND RRDNGGDG
```

LAF-1 IDR P24G/P25G

```
MESNQSNNGG SGNAALNRGG RYVGGHLRGG DGGAAAAASA GGDDRRGGAG GGGYRRGGGN SGGGGGGGYD RGYNDNRDDR DNRGGSGGYG RDRNYEDRGY
NGGGGGGGNR GYNNNRGGGG GGYNRQDRGD GGSSNFSRGG YNNRDEGSDN RGSGRSYNND RRDNGGDG
```

LAF-1 IDR (scramble 21-28)

```
RMESNQSNNG GSGNAALNRG GYGGDGGAAA AASAGGDDRR GGVAGGGGYR RGGGNSGGGG GGGYDRPGYN DNRDDRDNRG GSGGYGRDRN YEDRPGYNGG
GGGGGNRGYN NNRGGGGGGH YNRQDRGDGG SSNFSRGGYN NRLDEGSDNR GSGRSYNNDR RDNGGRDG
```

LAF-1 Shuffle

```
AGLNYGSDGG YNGDNAHGGN GRNGGNGRDR YYRRNRYRGG GGGERNRGDN GGNGNPGRGG RNGAGSSRGG NGSGQEAGGA YGGDVRGDDY GFGDGNNNDY
QGASRGRGDR SGNGGGRDGG SARGGRRNGD PGDSGNYSAG GRRNREDSGL GASDYGDDRG MYSGNNGN
```

TDP-43 CTD WT

```
GRFGGNPGGF GNQGGFGNSR GGGAGLGNNQ GSNMGGGMNF GAFSINPAMM AAAQAALQSS WGMMGMLASQ QNQSGPSGNN QNQGNMQREP NQAFGSGNNS
YSGSNSGAAI GWGSASNAGS GSGFNGGFGS SMDSKSSGWG M
```

SV series(1, 2):

```
sv1:  EKEKEKEKEK EKEKEKEKEK EKEKEKEKEK EKEKEKEKEK EKEKEKEKEK
sv2:  EEEKKKEEEK KKEEEKKKEE EKKKEEEKKK EEEKKKEEEK KKEEEKKKEK
sv3:  KEKKKEKKEE KKEEKEKEKE KEEKKKEEKE KEKEKKKEEK EKEEKKEEEE
sv4:  KEKEKKEEKE KKEEEKKEKE KEKKKEEKKK EEKEEKKEEK KKEEKEEEKE
sv5:  KEKEEKEKKK EEEEKEKKKK EEKEKEKEKE EKKEEKKKKE EKEEKEKEKE
sv6:  EEEKKEKKEE KEEKKEKKEK EEEKKKEKEE KKEEEKKKEK EEEEKKKKEK
sv7:  EEEEKKKKEE EEKKKKEEEE KKKKEEEEKK KKEEEEKKKK EEEEKKKKEK
sv8:  KKKKEEEEKK KKEEEEKKKK EEEEKKKKEE EEKKKKEEEE KKKKEEEEKE
sv9:  EEKKEEEKEK EKEEEEEKKE KKEKKEKKKE EKEKEKKKEK KKKEKEEEKE
sv10: EKKKKKKEEK KKEEEEEKKK EEEKKKEKKE EKEKEEKEKK EKKEEKEEEE
sv11: EKEKKKKKEE EKKEKEEEEK EEEEKKKKKE KEEEKEEKKE EKEKKKEEKK
sv12: EKKEEEEEEK EKKEEEEKEK EKKEKEEKEK KEKKKEKKEE EKEKKKKEKK
sv13: KEKKKEKEKK EKKKEEEKKK EEEKEKKKEE KKEKKEKKEE EEEEEKEEKE
sv14: EKKEKEEKEE EEKKKKKEEK EKKEKKKKEK KKKKEEEEEE KEEKEKEKEE
sv15: KKEKKEKKKE KKEKKEEEKE KEKKEKKKKE KEKKEEEEEE EEKEEKKEEE
sv16: EKEKEEKKKE EKKKKEKKEK EEKKEKEKEK KEEEEEEEEE KEKKEKKKKE
sv17: EKEKKKKKKE KEKKKKEKEK KEKKEKEEEK EEKEKEKKEE KKEEEEEEEE
sv18: KEEKKEEEEE EEKEEKKKKK EKKKEKKEEE KKKEEKKKEE EEEEKKKKEK
sv19: EEEEEKKKKK EEEEEKKKKK EEEEEKKKKK EEEEEKKKKK EEEEEKKKKK
sv20: EEKEEEEEEK EEEKEEKKEE EKEKKEKKEK EEKKEKKKKK KKKKKKEEEE
sv21: EEEEEEEEEK EKKKKKEKEE KKKKKKEKKE KKKKEKKEEE EEEKEEEKKK
sv22: KEEEEKEEKE EKKKKEKEEK EKKKKKKKKK KKKEKKEEEE EEEEKEKEEE
sv23: EEEEEKEEEE EEEEEEEKEE KEKKKKKKEK KKKKKKEKEK KKKEKKEEKK
sv24: EEEEKEEEEE KEEEEEEEEE EEEKKEEKK KKKEKKKKKK KEKKKKKKKK
sv25: EEEEEEEEEE EKEEEEKEEK EEKEKKKKKK KKKKKKKKKK KEEEKKEEKE
sv26: KEEEEEEEKE EKEEEEEEEE EKEEEEKEEK KKKKKKKKKK KKKKKKKKKE
sv27: KKEKKKEKKE EEEEEEEEEE EEEEEEEEEK EEKKKKKKKK KKKKKKKEKK
sv28: EKKKKKKKKK KKKKKKKKK KKEEEEEEEE EEEEEEEEEE KKEEEEEKEK
sv29: KEEEEKEEEE EEEEEEEEEE EEEEEEEKKK KKKKKKKKKK KKKKKKKKKK
sv30: EEEEEEEEEE EEEEEEEEEE EEEEEKKKKK KKKKKKKKKK KKKKKKKKKK
```

Coarse-Grained Model. We employ our recently developed C_α-based model, where proteins are represented as flexible chains, and each amino acid residue is considered as a single particle. Bonds are modeled using harmonic springs with a spring constant of 10 kcal/(mol Å^2) and a bond length of 3.8 Å. Long-range electrostatics are modeled using a Coulombic term with Debye-Hückel electrostatic screening(3), having the functional form:

$$E_{ij}(r) = \frac{q_i q_j}{4\pi D r} \exp(-\kappa r), \qquad [1]$$

in which $\kappa^{-1} = 10$, the Debye screening length corresponding to approximately 100 mM salt at room temperature, and $D = 80$, the dielectric constant of water. Nonbonded pairwise interactions are modeled using one of the two knowledge-based potentials we have previously applied to these systems(4).

The first pairwise interaction model, the hydrophobicity scale (HPS) model is based on amino acid residue hydrophobicity from Kapcha and Rossky(5), and applied to a Lennard-Jones-like functional form which can be used to scale the strength of interactions based on hydrophobicity(6):

$$\Phi(r) = \begin{cases} \Phi_{LJ} + (1-\lambda)\epsilon, & \text{if } r \leq 2^{1/6}\sigma \\ \lambda \Phi_{LJ}, & \text{otherwise} \end{cases} \qquad [2]$$

in which Φ_{LJ} is the standard Lennard-Jones potential and λ represents hydrophobicity. ϵ is set equal to 0.2 kcal/mol in order to minimize deviation of R_g from multiple FRET and SAXS experimental measurements of unfolded proteins(4).

The second model used is the Kim-Hummer (KH) model which was derived from the Miyazawa-Jernigan pair potential(7) for use with weakly binding folded proteins(8). The KH model can be expressed as:

$$\Phi(r) = \begin{cases} \Phi_{LJ} + 2\epsilon, & \text{if } \epsilon > 0 \text{ and } r < 2^{1/6}\sigma \\ -\Phi_{LJ}, & \text{otherwise} \end{cases} \qquad [3]$$

where positive values of ϵ will result in a fully repulsive potential. The model was parameterized by the experimental osmotic second virial coefficient of lysozyme and the binding affinity of the ubiquitinCUE complex(8).

Simulation Methods. Slab configurations were initially generated by conducting 100 ns simulations at constant temperature and pressure, starting from a dispersed phase of protein chains with periodic boundary conditions at 150 K and 1 bar, maintained by a Langevin thermostat and a Parrinello-Rahman barostat(9). The x- and y- dimensions were set to ~ 15 nm which is sufficient to prevent chains from interacting with their periodic images. The z-dimension of the box is then extended to > 200 nm. Production simulations were conducted for ~ 5 μs at constant temperature and volume. The first 1 μs of simulation was discarded as equilibration, and the remainder is used to calculate the density profile, the phase diagram and T_c. All slab simulations were conducted using HOOMD-Blue v2.1.5 (10). The errors of the T_c were estimated by using a block average with 5 blocks.

In order to obtain T_θ, single-chain simulations were conducted at a range of temperatures using replica exchange molecular dynamics (REMD)(11), with a temperature list of 150.0, 170.1, 193.0, 218.9, 248.3, 281.7, 319.5, 362.4, 411.1, 466.3, 529.0, and 600.0 K. For the 9 polyampholyte sequences where T_θ falls outside this range, we ran additional simulations with an extended

temperature range. Simulations were conducted in cubic boxes with periodic boundaries, large enough that a protein chain will not encounter its periodic image, and temperature was maintained using a Langevin thermostat. All single-chain simulations were conducted using LAMMPS(12). For each temperature we estimated ν by fitting to:

$$R_{ij} = b|i-j|^{\nu}. \quad [4]$$

An alternative way of obtaining ν by using only the radius of gyration is through equation (13–15)

$$R_g^2 = \sqrt{\frac{\gamma(\gamma+1)}{2(\gamma+2\nu)(\gamma+2\nu+1)}}bN^{\nu} \quad [5]$$

in which $\gamma \approx 1.1615$(16). We further estimated T_θ by interpolating the temperature at which ν=0.5. The errors of T_θ were estimated by using a block average method and dividing the entire trajectory into 5 blocks.

In order to obtain T_B, first the potential of mean force (PMF) of two protein chains was calculated via Monte Carlo (MC) method using an umbrella sampling strategy. An harmonic biasing potential was applied to center of mass distance, d, between the two proteins with a spring constant of 0.1 kcal/(mol Å^2). The center of the distance, d_0, for umbrella sampling varied from 0 Å to 102.9 Å with an interval of 3.4 Å for $d_0 < 40$ Å and an interval of 6.9 Å for $d_0 > 40$ Å so that the density of umbrella windows is doubled for the distances at which the two IDPs are in close contact. The weighted histogram analysis method was then used to merge the umbrella sampling data and compute the PMF(17). The corresponding radial distribution function $g(r)$ was calculated from PMF and B_{22} is obtained from that using the following equation:

$$B_{22} = 2\pi \int_0^{\infty} [1-g(r)]\, r^2 dr. \quad [6]$$

The errors of B_{22} were estimated by using a block average with 5 blocks. In order to determine T_B considering the errors of B_{22}, we follow a bootstrapping strategy: by first generating 1000 sets of B_{22} data at the temperatures simulated taking into account the errors of B_{22}; second linearly interpolating the temperature at which B_{22}=0 (if multiple temperatures are obtained with B_{22}=0, we pick the one in the middle); and at last obtain T_B and the errors from the mean and standard deviation of the 1000 trials.

Slab method comparison with Monte Carlo methods. The use of slab method can also be justified against other methods of sampling phase coexistence, such as the agreement between results of LJ liquids from Sheng et al.(18) who use an iterative approach involving Monte Carlo simulations of flexible polymers and calculation of chemical potentials in the two phases, and from Silmore et al.(19), who utilize molecular dynamics simulations using slab method. We have plotted their data together to show they are in good agreement (See Fig. S1).

Fitting scheme for T_c. We have described the fitting scheme for obtaining T_c from the density profile in our previous work (4) and will briefly discuss here using FUS WT with KH model as an example. The critical temperature T_c can be obtained by fitting

$$\rho_H - \rho_L = A(T_c - T)^{\beta} \quad [7]$$

where β=0.325 is the critical exponent (20), and ρ_H and ρ_L are the concentrations of the high- and low-density phases, respectively. A is a protein-specific fitting parameter. Since we only have a rough estimate of the critical temperature for a specific IDP sequence based on their molecular properties in isolation, and their sequence composition, we always run simulations at more temperatures than usually necessary. The minimum fitting temperature ($T_{\min}$) is selected as the lowest temperature at which ρ_L is nonzero in the simulation, whereas the maximum fitting temperature ($T_{\max}$) is determined by checking the fitting errors (4). However, we find that the fitting of T_c is largely insensitive to the number and location of temperatures used for fitting (Fig. S4).

Formation of a slab. To further elaborate on the validity of the extrapolated T_c, we present simulation snapshots at few time points for several temperatures in the vicinity of the computed T_c (Fig. S5). It is quite clear from these snapshots that the system tends to form a single phase above the T_c and remains in a two-phase coexistence below T_c, as one would expect if the computed T_c value was accurate. Moreover, a system initiated from fully dispersed protein chains at a temperature below T_c forms a dense protein phase (slab) though the process itself may take a long time thereby making it more efficient to start the simulations from a slab configuration (Fig. S5, Movie S3). The final results though will be independent of the starting configuration as we have previously shown(4).

2. Supporting Figures

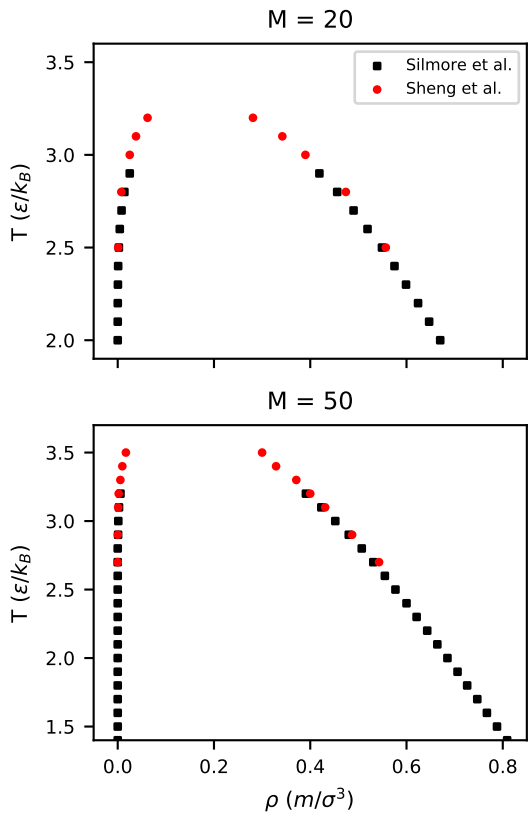

Fig. S1. Phase diagrams calculated using different methods for fully flexible Lennard Jones chains give very similar results. Sheng et al.(18) use grand canonical Monte Carlo simulations of a small assembly of polymers and calculate chemical potential using chain increment method as an iterative approach to determine phase coexistence densities, while Silmore et al.(19) utilize molecular dynamics simulations with slab geometry similar our procedure in this work.

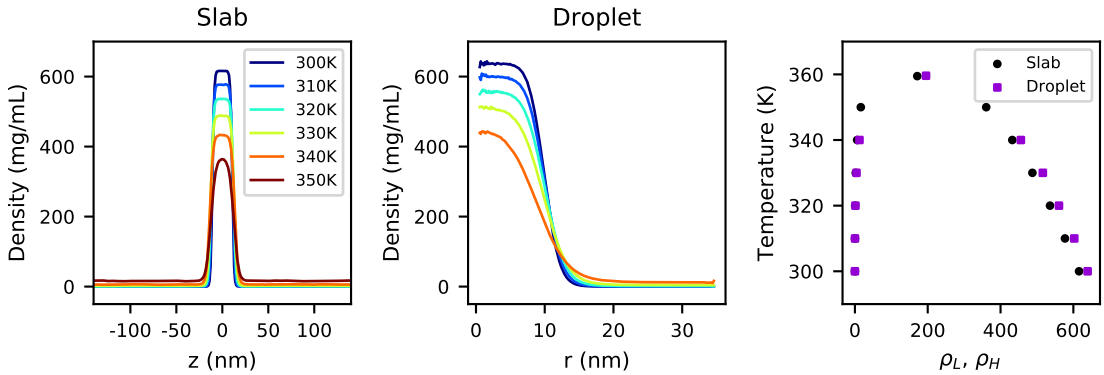

Fig. S2. Density profiles of simulations of 100 chains of FUS using Slab and Droplet geometry, and comparison of their phase diagrams.

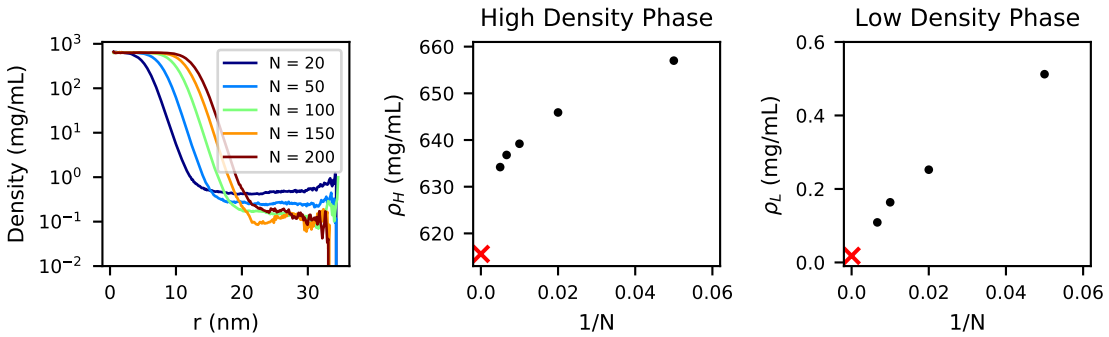

Fig. S3. Density profiles of FUS chains at 300K using different system sizes. Slab densities are indicated by red "X", and have been placed at 1/N = 0 as the slab geometry is expected to minimize finite-size effects. Densities in both phases tend toward the results of the slab simulation with increasing system size. For N=200, the low density phase does not converge due to the small box size, so that point has been omitted in the third subplot.

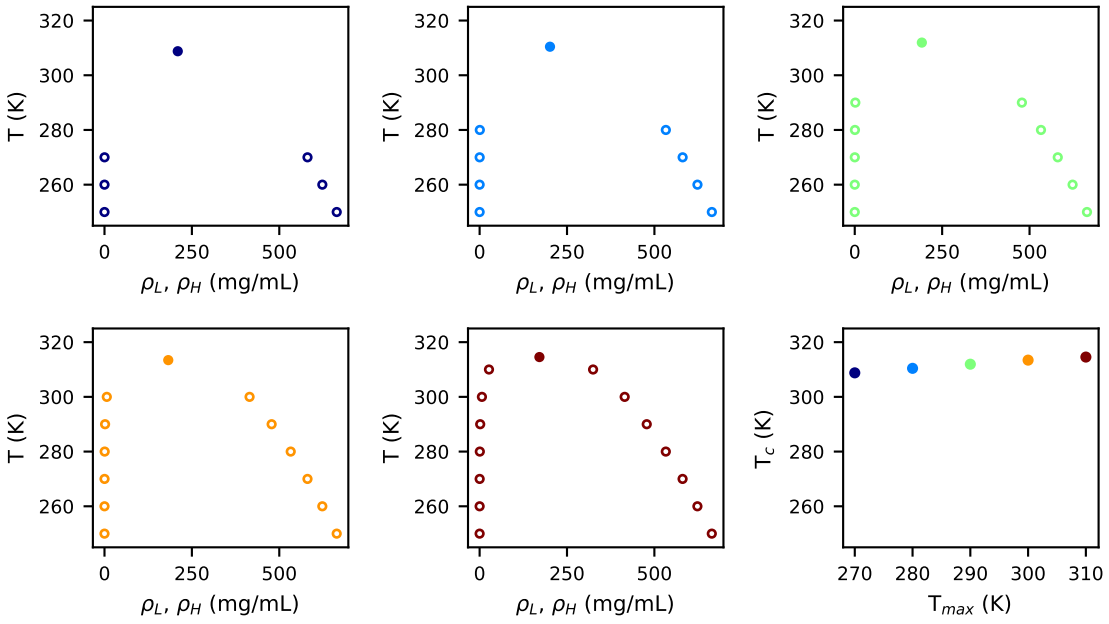

Fig. S4. The number and location of data points has a small influence on the accuracy of the extrapolated T_c value. As the highest temperature used for fitting ($T_{\max}$) gets closer to T_c, the extrapolation is generally better.

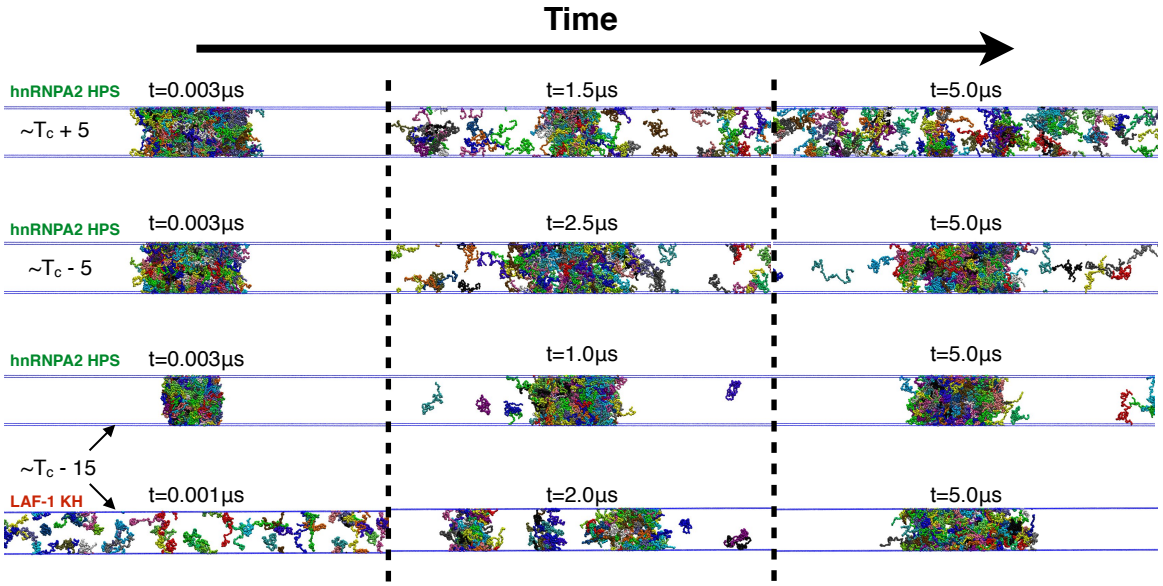

Fig. S5. Time evolution of simulations starting from slab configuration at 300, 310 and 320K for hnRNPA2, and starting from continuous dispersed phase of LAF-1 at 210K. The slab breaks up at temperatures above T_c, while it remains phase separated at temperatures below T_c. When starting from a dispersed phase, the system eventually relaxes to a slab at temperatures below T_c after sufficient time.

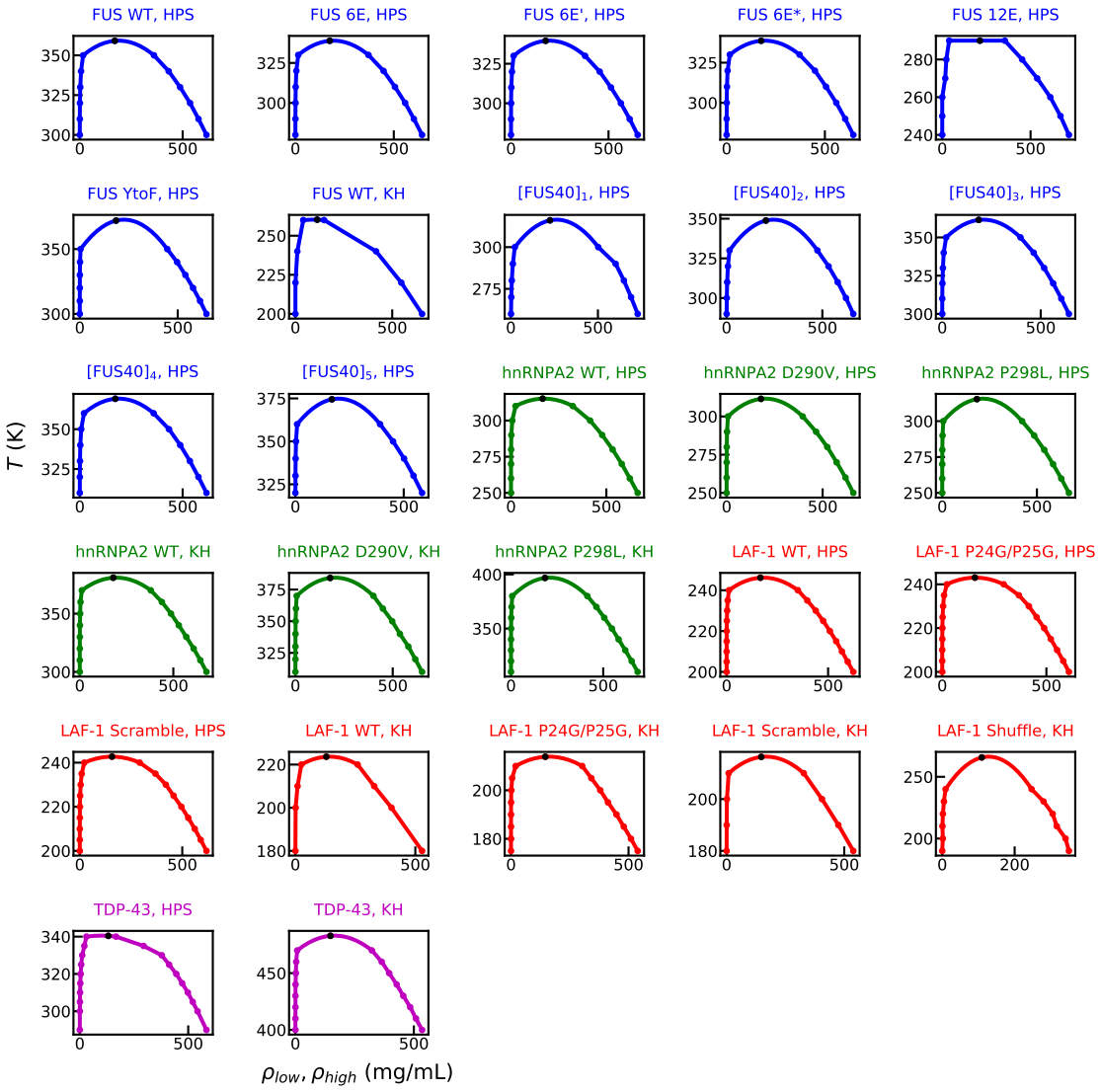

Fig. S6. Phase diagrams of all the protein sequences we have simulated. The black dots show the critical temperature we determined from the phase diagram.

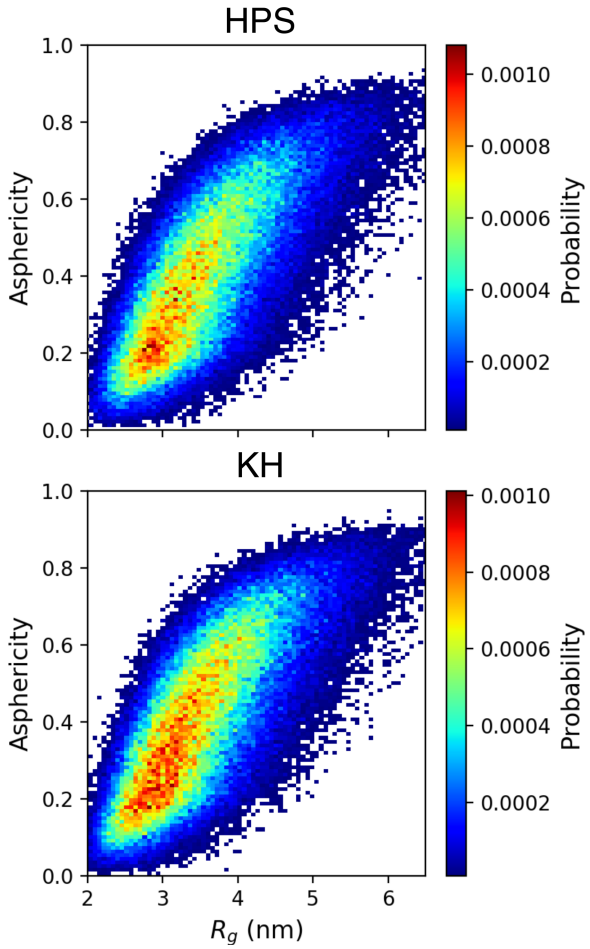

Fig. S7. Comparison of chain dimensions of LAF-1 IDR WT at 300K between our models (i.e. HPS at the top and KH at the bottom) and ABSINTH model(21) shown by Fig. 3a in Wei et al.(22). This shows LAF-1 is sampling both collapsed and extended conformations.

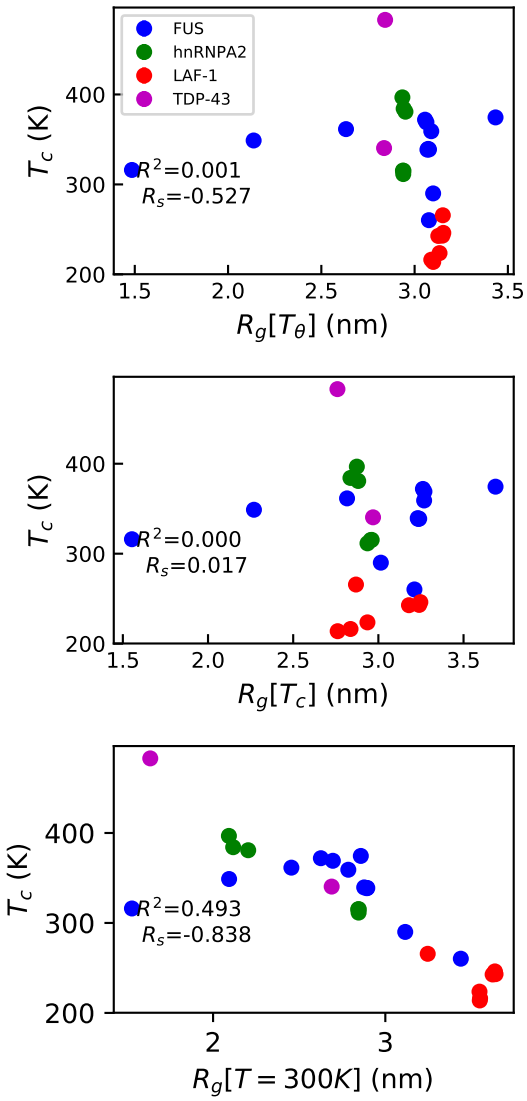

Fig. S8. Comparison between T_c and R_g at T_θ (top), at T_c (middle) or at 300K (bottom). R^2 shows the square of Pearson correlation coefficient, and R_s the Spearman correlation coefficient. The poor correlation of T_c with R_g is likely due to the wide range of different chain lengths for the FUS variants studied here.

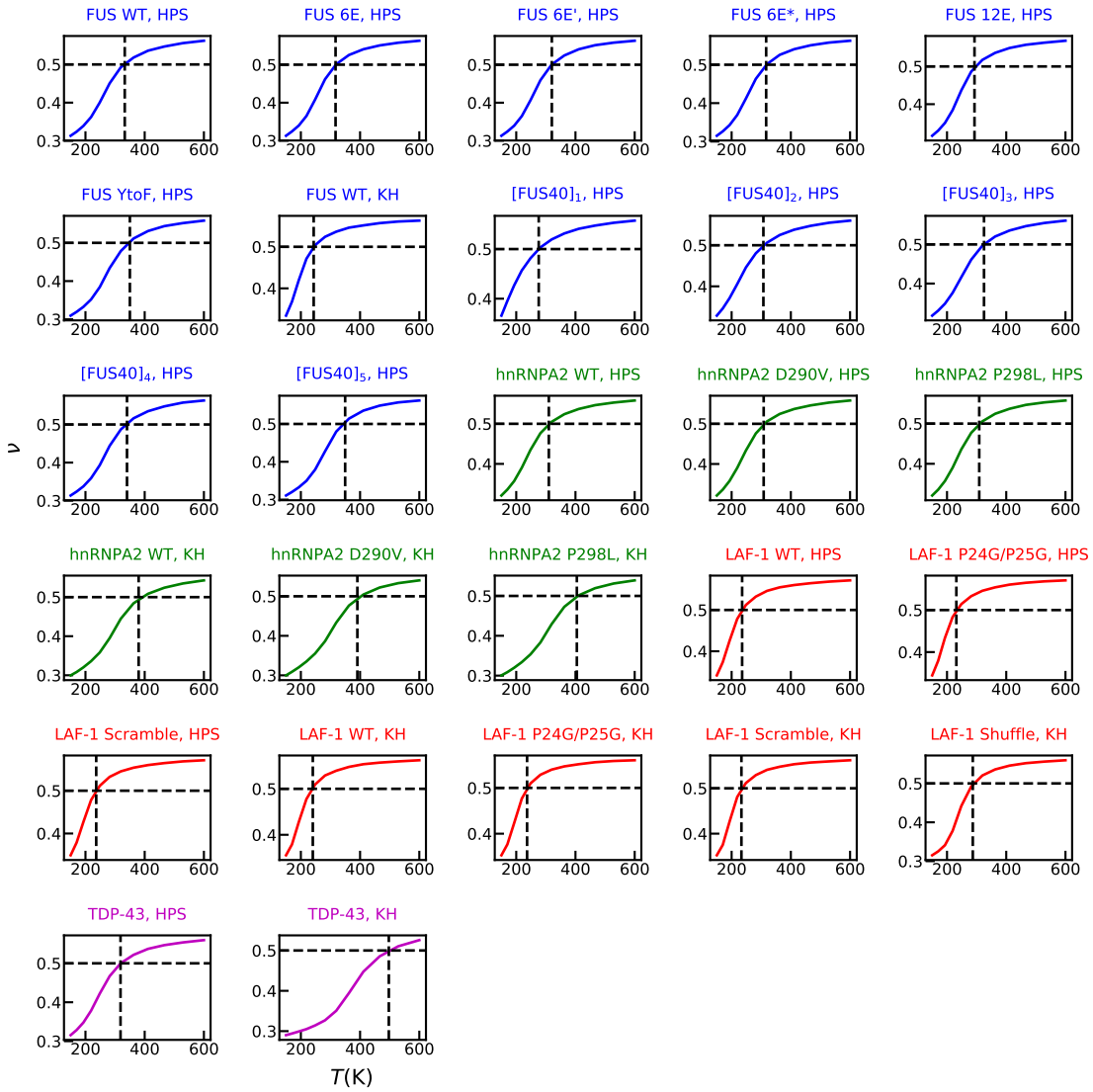

Fig. S9. Polymer scaling exponent ν as a function of the temperature for each protein sequences used in the comparison with T_c. The black dashed lines show when $\nu = 0.5$ in the theta solvent condition whereas the black dots show T_θ determined by obtaining the crossing point with the line $\nu = 0.5$.

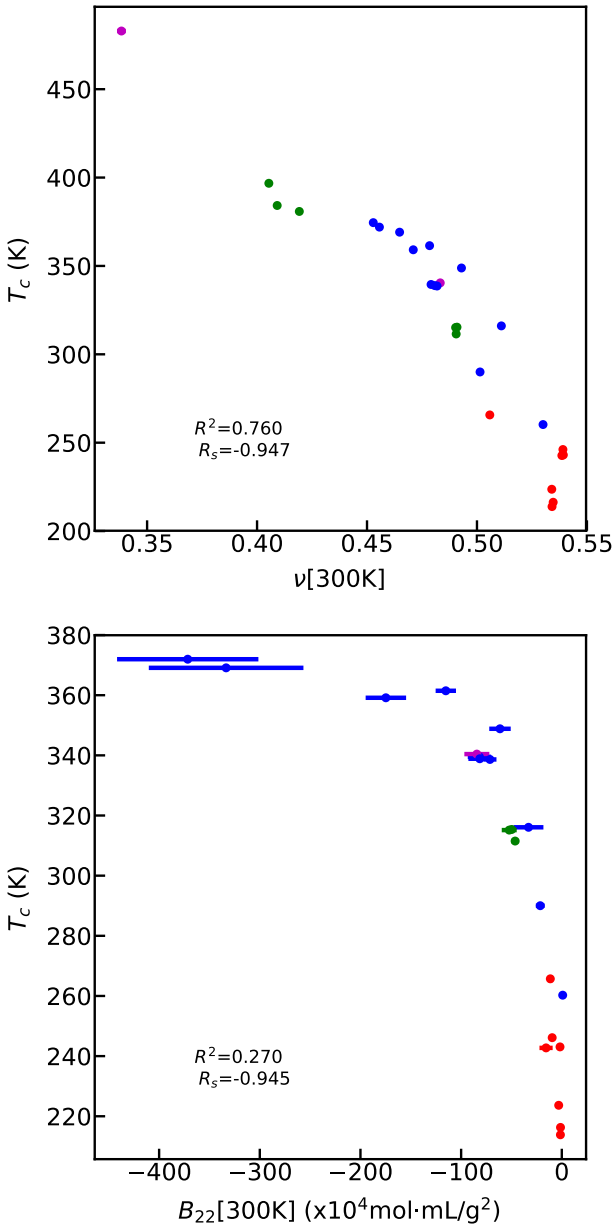

Fig. S10. Comparison between ν or B_{22} at 300K, and T_c. R^2 shows the square of Pearson correlation coefficient, and R_s the Spearman correlation coefficient.

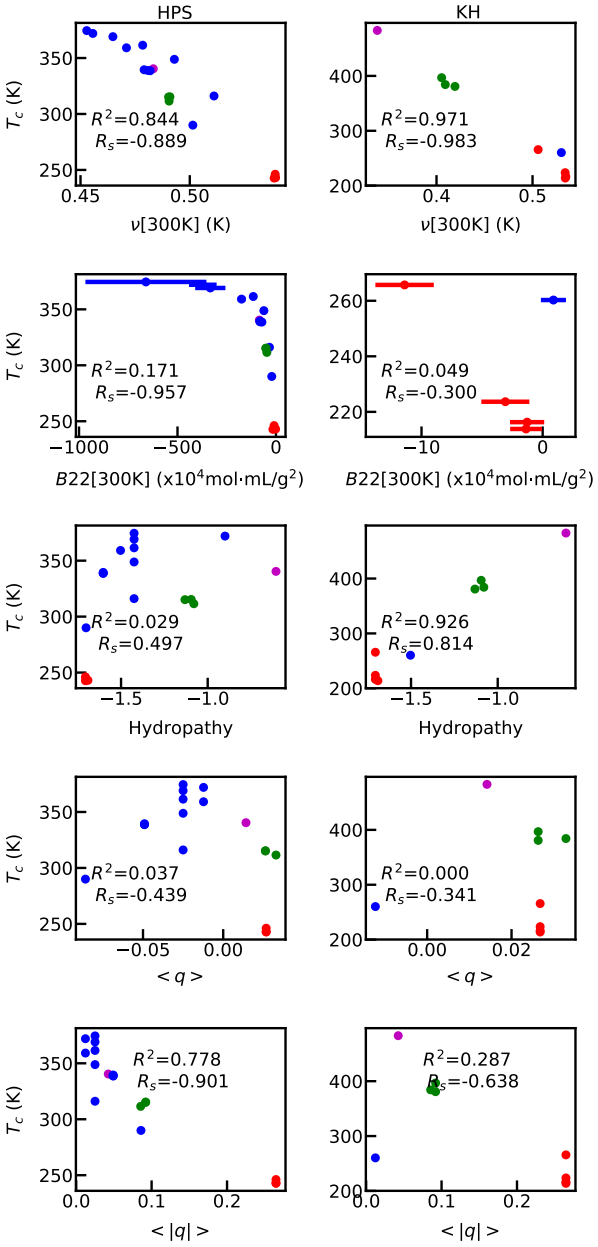

Fig. S11. Comparison between different metrics shown in Table S1 and T_c using two different coarse-grained potentials. R^2 shows the square of Pearson correlation coefficient, and R_s the Spearman correlation coefficient. Hydropathy is the mean Kyte-Doolittle score (23) of residues within the sequence, $\langle q \rangle$ is the mean net charge per residue, and $\langle |q| \rangle$ is the mean absolute charge per residue.

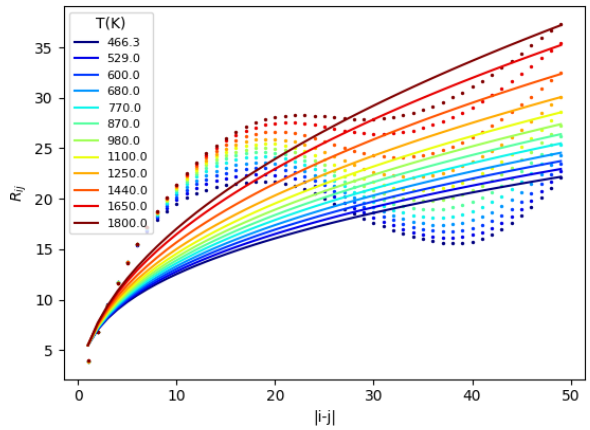

Fig. S12. For highly charge-segregated polyampholyte sequence, sv23(1), the intra-chain distances do not fit the polymer scaling law well, even at T_θ.

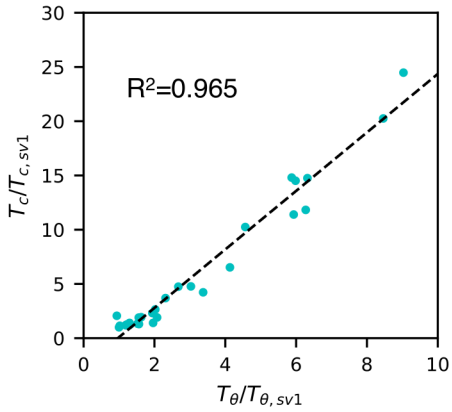

Fig. S13. Calculating T_θ from scaling exponent determined using an analytical equation between R_g and ν as described by Zheng et al.(15) instead of fitting to the average distances as a function of the sequence separation (Fig. S12) also gives very good correlation with T_c results from Lin and Chan(2).

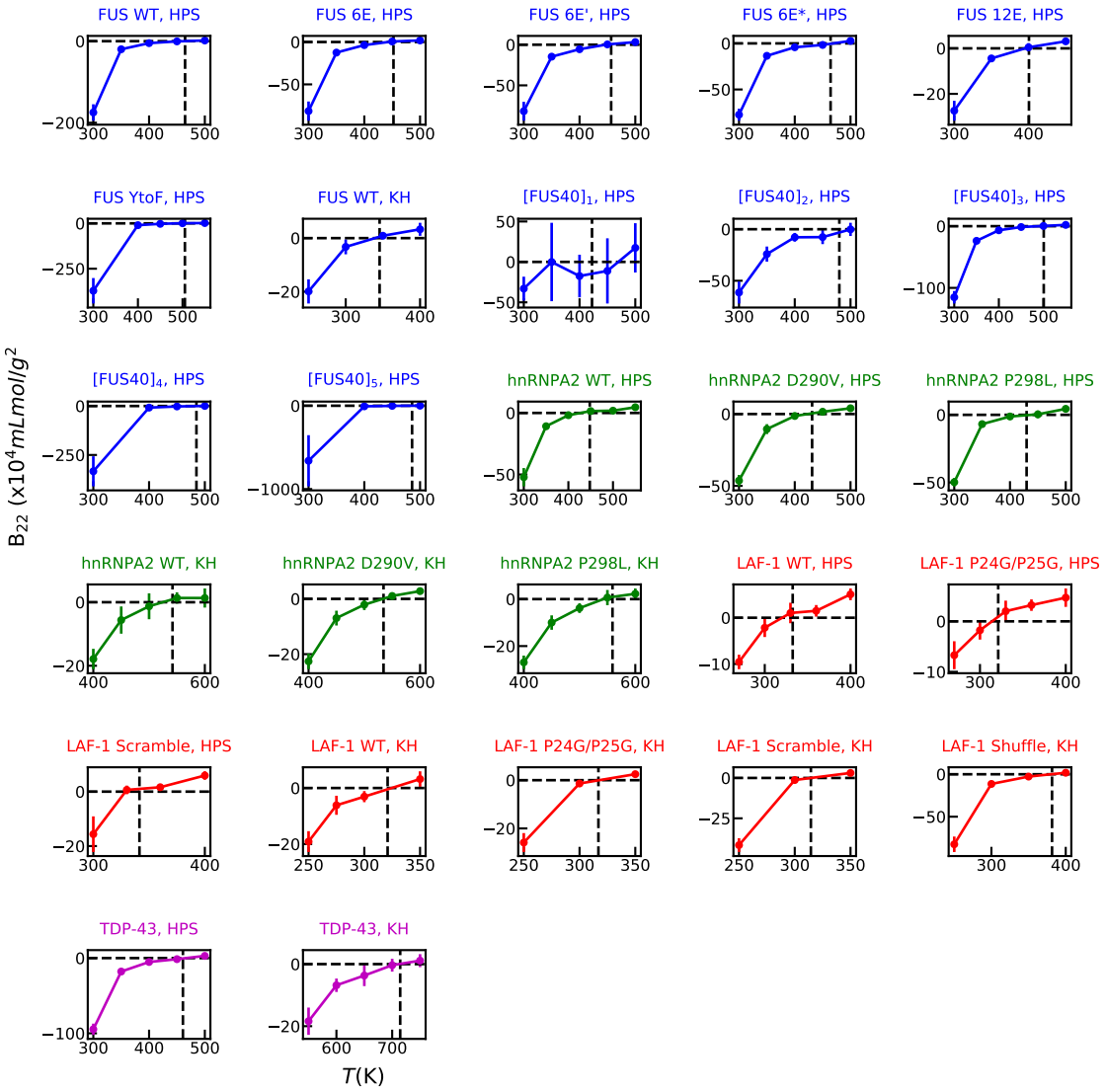

Fig. S14. B_{22} of all the protein sequences we have simulated. The black dashed lines show the Boyle temperature we determined from the B_{22} plot as a function of T.

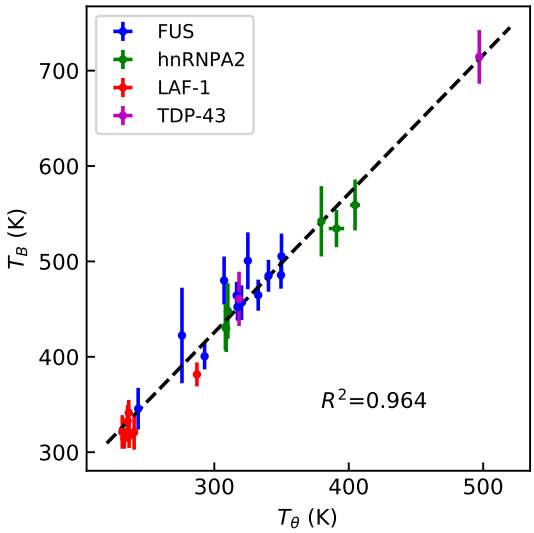

Fig. S15. Comparison between T_B and T_θ.

3. Supporting Tables

Table S1. List of intrinsically disordered or unfolded proteins and simulation model combinations used in this study where simulation models are hydrophobicity scaling (HPS) and Kim-Hummer (KH). Average hydropathy is calculated from the Kyte-Doolittle scale(23), q_{tot} is the net charge, and FCR is the fraction of charged residues.

ID	Protein	Model	Length	T_c (K)	T_θ (K)	T_B (K)	Hydropathy (Kyte-Doolittle)	q_{tot}	FCR
	FUS WT	HPS	163	359.1	332.7	464.6	-1.5030	-2	0.0123
	FUS 6E	HPS	163	338.9	317.2	452.3	-1.6029	-8	0.0491
	FUS 6E'	HPS	163	339.5	320.2	456.7	-1.6029	-8	0.0491
	FUS 6E*	HPS	163	338.6	316.6	464.3	-1.6029	-8	0.0491
	FUS12E	HPS	163	290.0	292.8	400.5	-1.7019	-14	0.0859
	$[\mathrm{FUS40}]_1$	HPS	40	316.1	275.9	422.4	-1.4247	-1	0.0250
	$[\mathrm{FUS40}]_2$	HPS	80	348.8	307.3	480.0	-1.4247	-2	0.0250
	$[\mathrm{FUS40}]_3$	HPS	120	361.5	324.9	500.7	-1.4247	-3	0.0250
	$[\mathrm{FUS40}]_4$	HPS	160	369.1	340.2	484.9	-1.4247	-4	0.0250
	$[\mathrm{FUS40}]_5$	HPS	200	374.5	349.6	485.7	-1.4247	-5	0.0250
	FUS YtoF	HPS	163	372.0	350.0	505.3	-0.9000	-2	0.0123
	hnRNPA2 WT	HPS	152	315.2	310.1	448.0	-1.1313	+4	0.0921
	hnRNPA2 D290V	HPS	152	311.5	308.2	431.1	-1.0800	+5	0.0987
	hnRNPA2 P298L	HPS	152	315.4	308.8	429.8	-1.0953	+4	0.0921
	LAF-1 IDR WT	HPS	168	246.1	235.5	332.6	-1.7055	+4.5	0.2648
	LAF-1 IDR P24G/P25G	HPS	168	243.1	231.5	321.3	-1.6911	+4.5	0.2648
	LAF-1 IDR scramble(21-28)	HPS	168	242.7	236.3	341.4	-1.7055	+4.5	0.2648
	TDP-43 CTD	HPS	141	340.4	318.4	460.7	-0.6066	+2	0.0426
	FUS WT	KH	163	260.3	243.4	345.6	-1.5030	-2	0.0123
	hnRNPA2 WT	KH	152	380.8	379.6	542.1	-1.1313	+4	0.0921
	hnRNPA2 D290V	KH	152	384.2	390.9	534.5	-1.0800	+5	0.0987
	hnRNPA2 P298L	KH	152	396.8	404.7	559.1	-1.0953	+4	0.0921
	LAF-1 IDR WT	KH	168	223.6	240.4	320.9	-1.7055	+4.5	0.2648
	LAF-1 IDR P24G/P25G	KH	168	213.8	236.6	316.9	-1.6911	+4.5	0.2648
	LAF-1 IDR scramble(21-28)	KH	168	216.3	233.1	314.5	-1.7055	+4.5	0.2648
	LAF-1 Shuffle	KH	163	265.7	287.1	381.6	-1.7055	+4.5	0.2648
	TDP-43 CTD	KH	141	482.9	497.1	714.4	-0.6066	+2	0.0426

4. Supporting Movies

Movie S1 Slab simulation of FUS WT at 270K.

Movie S2 Simulation of phase coexistence using droplet geometry for 150 chains of FUS WT at 300K.

Movie S3 Slab simulation starting from dispersed configuration relaxes to single assembly over time.

Movie S4 Single chain of hnRNPA2 WT from REMD simulation at 170K, below T_θ.

Movie S5 Double chain simulation of hnRNPA2 WT from umbrella sampling simulation at 300K with center-of-mass distance set to 0.

1. Das RK, Pappu RV (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U.S.A.* 110:13392–13397.
2. Lin YH, Chan HS (2017) Phase separation and single-chain compactness of charged disordered proteins are strongly correlated. *Biophys. J.* 112(10):2043–2046.
3. Debye P, Hückel E (1923) De la theorie des electrolytes. i. abaissement du point de congelation et phenomenes associes. *Physikalische Zeitschrift* 24(9):185–206.
4. Dignon GL, Zheng W, Kim YC, Best RB, Mittal J (2018) Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Comput. Biol.* 14(1):e1005941.
5. Kapcha LH, Rossky PJ (2014) A simple atomic-level hydrophobicity scale reveals protein interfacial structure. *J. Mol. Biol.* 426(2):484–498.
6. Ashbaugh HS, Hatch HW (2008) Natively unfolded protein stability as a coil-to-globule transition in charge/hydropathy space. *J. Am. Chem. Soc.* 130(29):9536–9542.
7. Miyazawa S, Jernigan RL (1996) Residue-residue potentials with a favourable contact pair term and an unfavourable high packing density term, for simulation and threading. *J. Mol. Biol.* 256:623–644.
8. Kim YC, Hummer G (2008) Coarse-grained models for simulation of multiprotein complexes: application to ubiquitin binding. *J. Mol. Biol.* 375:1416–1433.
9. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190.
10. Anderson JA, Lorenz CD, Travesset A (2008) General purpose molecular dynamics simulations fully implemented on graphics processing units. *J. Comput. Phys.* 227(10):5342–5359.
11. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics methods for protein folding. *Chem. Phys. Lett.* 314:141–151.
12. Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* 117(1):1–19.
13. Flory PJ (1949) The configuration of real polymer chains. *J. Chem. Phys.* 17:303–310.
14. Witten Jr T, Schafer L (1978) Two critical ratios in polymer solutions. *J. Phys A-Math. Gen.* 11(9):1843.
15. Zheng W, et al. (2018) Inferring properties of disordered chains from fret transfer efficiencies. *J. Chem. Phys.* 148(12):123329.
16. Le Guillou J (1977) Jc le guillou and j. zinn-justin, phys. rev. lett. 39, 95 (1977). *Phys. Rev. Lett.* 39:95.
17. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13(8):1011–1021.
18. Sheng YJ, Panagiotopoulos AZ, Kumar SK, Szleifer I (1994) Monte carlo calculation of phase equilibria for a bead-spring polymeric model. *Macromolecules* 27(2):400–406.
19. Silmore KS, Howard MP, Panagiotopoulos AZ (2017) Vapour–liquid phase equilibrium and surface tension of fully flexible lennard–jones chains. *Mol. Phys.* 115(3):320–327.
20. Rowlinson JS, Widom B (2013) *Molecular theory of capillarity*. (Courier Corporation).
21. Vitalis A, Pappu RV (2008) ABSINTH: A new continuum solvent model for simulations of polypeptides in aqueous solutions. *J. Comput. Chem.* 30:673–699.
22. Wei MT, et al. (2017) Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nat. Chem.* 9(11):1118.
23. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157(1):105–132.