SUPPLEMENTARY INFORMATION

S1 AMINO-ACID CODES AND BIOINFORMATICS DATA

Table I. The 20 naturally occurring amino acids with their one- and three-letter codes, alongside their charges. In simulations with most (but not all) models considered, the charge of histidine is set to half that of the other non-zero charges [see SI Table X below]. Amino acids marked with a '⋆' are aromatic. The last column corresponds to the planar π–π contact interaction frequencies for each residue, extracted from Fig. 1B of Ref. 13, and rescaled to a range between 0 and 1.

Full name	Code		Charge	Freq.
Alanine	Ala	A	0	0.091
Arginine	Arg	R	+	0.552
Asparagine	Asn	N	0	0.353
Aspartate	Asp	D	−	0.195
Cysteine	Cys	C	0	0.127
Glutamine	Gln	Q	0	0.365
Glutamate	Glu	E	−	0.211
Glycine	Gly	G	0	0.220
Histidine	His	H	+	0.668
Isoleucine	Ile	I	0	0.005
Leucine	Leu	L	0	0.021
Lysine	Lys	K	+	0.048
Methionine	Met	M	0	0.073
⋆ Phenylalanine	Phe	F	0	0.712
Proline	Pro	P	0	0.144
Serine	Ser	S	0	0.113
Threonine	Thr	T	0	0.057
⋆ Tryptophan	Trp	W	0	1.000
⋆ Tyrosine	Tyr	Y	0	0.762
Valine	Val	V	0	0.011

S2 AMINO-ACID SEQUENCES OF PROTEINS STUDIED

We give below the amino-acid sequences of all the proteins considered, namely those proteins used in simulations of the radius of gyration; the hnRNPA1 intrinsically disordered region and its 8 variants used for validating phase behaviour; LAF-1 RGG and 2 of its variants; and the FUS IDR, using one-letter codes [Table I] for the amino acids.

S2.1 Sequences of proteins used in radius of gyration calculations

α-synuclein
MDVFM KGLSK AKEGV VAAAE KTKQG VAEAA GKTKE
GVLYV GSKTK EGVVH GVATV AEKTK EQVTN VGGAV
VTGVT AVAQK TVEGA GSIAA ATGFV KKDQL GKNEE
GAPQE GILED MPVDP DNEAY EMPSE EGYQD YEPEA

ACTR
GTQNR PLLRN SLDDL VGPPS NLEGQ SDERA LLDQL
HTLLS NTDAT GLEEI DRALG IPELV NQGQA LEPKQ
D

Ash1
GASAS SSPSP STPTK SGKMR SRSSS PVRPK AYTPS
PRSPN YHRFA LDSPP QSPRR SSNSS ITKKG SRRSS
GSSPT RHTTR VCV

hNHE1cdt
MVPAH KLDSP TMSRA RIGSD PLAYE PKEDL PVITI
DPASP QSPES VDLVN EELKG KVLGL SRDPA KVAEE
DEDDD GGIMM RSKET SSPGT DDVFT PAPSD SPSSQ
RIQRC LSDPG PHPEP GEGEP FFPKG Q

IBB
GCTNE NANTP AARLH RFKNK GKDST EMRRR RIEVN
VELRK AKKDD QMLKR RNVSS FPDDA TSPLQ ENRNN
QGTVN WSVDD IVKGI NSSNV ENQLQ AT

K18
MQTAP VPMPD LKNVK SKIGS TENLK HQPGG GKVQI
INKKL DLSNV QSKCG SKDNI KHVPG GGSVQ IVYKP
VDLSK VTSKC GSLGN IHHKP GGGQV EVKSE KLDFK
DRVQS KIGSL DNITH VPGGG NKKIE

K25
MAEPR QEFEV MEDHA GTYGL GDRKD QGGYT MHQDQ
EGDTD AGLKA EEAGI GDTPS LEDEA AGHVT QARMV
SKSKD GTGSD DKKAK GADGK TKIAT PRGAA PPGQK
GQANA TRIPA KTPPA PKTPP SSGEP PKSGD RSGYS
SPGSP GTPGS RSRTP SLPTP PTREP KKVAV VRTPP
KSPSS AKSRL

N49
GCQTS RGLFG NNNTN NINNS SSGMN NASAG LFGSK
P

N98
GCFNK SFGTP FGGGT GGFGT TSTFG QNTGF GTTSG
GAFGT SAFGS SNNTG GLFGN SQTKP GGLFG TSSFS
QPATS TSTGF GFGTS TGTAN TLFGT ASTGT SLFSS
QNNAF AQNKP TGFGN FGTST SSGGL FGTTN TTSNP
FGSTS GSLFG P

NLS
ACETN KRKRE QISTD NEAKM QIQEE KSPKK KRKKR
SSKAN KPPE

NSP
GCNFN TPQQN KTPFS FGTAN NNSNT TNQNS STGAG
AFGTG QSTFG FNNSA PNNTN NANSS ITPAF GSNNT
GNTAF GNSNP TSNVF GSNNS TTNTF GSNSA GTSLF
GSSSA QQTKS NGTAG GNTFG SSSLF NNSTN SNTTK
PAFGG LNFGG GNNTT PSSTG NANTS NNLFG ATANA
N

NUL
GCGFK GFDTS SSSSN SAASS SFKFG VSSSS SGPSQ
TLTST GNFKF GDQGG FKIGV SSDSG SINPM SEGFK
FSKPI GDFKF GVSSE SKPEE VKKDS KNDNF KFGLS
SGLSN PV

NUS
GCPSA SPAFG ANQTP TFGQS QGASQ PNPPG FGSIS
SSTAL FPTGS QPAPP TFGTV SSSSQ PPVFG QQPSQ
SAFGS GTTPN

P53
MEEPQ SDPSV EPPLS QETFS DLWKL LPENN VLSPL
PSQAM DDLML SPDDI EQWFT EDPGP DEAPR MPEAA
PPVAP APAAP TPAAP APAPS WPL

ProTα
MSDAA VDTSS EITTK DLKEK KEVVE EAENG RDAPA
NGNAE NEENG EQEAD NEVDE EEEEG GEEEE EEEEG
DGEEE DGDED EEAES ATGKR AAEDD EDDDV DTKKQ
KTDED D

SH4-UD
MGSNK SKPKD ASQRR RSLEP AENVH GAGGG AFPAS
QTPSK PASAD GHRGP SAAFA PAAAE PKLFG GFNSS
DTVTS PQRAG PLAGG

Sic1
GSMTP STPPR SRGTR YLAQP SGNTS SSALM QGQKT
PQKPS QNLVP VTPST TKSFK NAPLL APPNS NMGMT
SPFNG LTSPQ RSPFP KSSVK RT

Table III. Experimental radii of gyration for proteins, alongside the experimental salt concentration and the corresponding Debye screening constant (computed using the equation immediately following Eq. (12) of Ref. 71 expressed in SI instead of gaussian units).

Protein	R_g / nm	[salt] / mM	κ / nm^{-1}
α-synuclein [72]	3.31	185	1.40
ACTR [73]	2.51	199	1.45
Ash1 [74]	2.85	150	1.26
hNHE1cdt [73]	3.63	199	1.45
IBB [75]	3.20	162	1.31
K18 [76]	3.80	163	1.31
K25 [76]	4.40	163	1.31
N49 [75]	1.59	162	1.31
N98 [75]	2.86	162	1.31
NLS [75]	2.40	162	1.31
NSP [75]	4.10	162	1.31
NUL [75]	3.00	162	1.31
NUS [75]	2.49	162	1.31
P53 [77]	2.87	208	1.49
ProTα [78, 79]	3.79	155	1.28
SH4-UD [80]	2.90	217	1.52
Sic1 [81]	3.21	162	1.31

S2.2 hnRNPA1 variants

The wild-type hnRNPA1-LCD sequence is shown below.

[residues 186–320 of UniProt sequence P09651-2]

hnRNPA1
MASAS SSQRG RSGSG NFGGG RGGGF GGNDN FGRGG
NFSGR GGFGG SRGGG GYGGS GDGYN GFGND GSNFG
GGGSY NDFGN YNNQS SNFGP MKGGN FGGRS SGPYG
GGGQY FAKPR NQGGY GGSSS SSSYG SGRRF

The sequences of the variants of hnRNPA1 we have considered are shown below, using the nomenclature of Bremer and co-workers [10]. The amino-acid residues different from the wild type are highlighted in red. Estimates of their critical temperatures are given in SI Table IV. For hnRNPA1 variants, we recently reported a validation for the KH model similar to that described in the main text [82].

−3R+3K
MASAS SSQRG KSGSG NFGGG RGGGF GGNDN FGRGG
NFSGR GGFGG SKGGG GYGGS GDGYN GFGND GSNFG
GGGSY NDFGN YNNQS SNFGP MKGGN FGGRS SGGSG
GGGQY FAKPR NQGGY GGSSS SSSYG SGRKF

−4F−2Y
MASAS SSQRG RSGSG NSGGG RGGGF GGNDN FGRGG
NSSGR GGFGG SRGGG GYGGS GDGYN GFGND GSNSG
GGGSS NDFGN YNNQS SNFGP MKGGN FGGRS SGGSG
GGGQY SAKPR NQGGY GGSSS SSSSG SGRRF

−6R+6K
MASAS SSQKG KSGSG NFGGG RGGGF GGNDN FGKGG
NFSGR GGFGG SKGGG GYGGS GDGYN GFGND GSNFG
GGGSY NDFGN YNNQS SNFGP MKGGN FGGKS SGGSG
GGGQY FAKPR NQGGY GGSSS SSSYG SGRKF

+7F−7Y
MASAS SSQRG RSGSG NFGGG RGGGF GGNDN FGRGG
NFSGR GGFGG SRGGG GFGGS GDGFN GFGND GSNFG
GGGSF NDFGN FNNQS SNFGP MKGGN FGGRS SGGSG
GGGQF FAKPR NQGGF GGSSS SSSFG SGRRF

+7K+12D
MASAD SSQRD RDDKG NFGDG RGGGF GGNDN FGRGG
NFSDR GGFGG SRGDG KYGGD GDKYN GFGND GKNFG
GGGSY NDFGN YNNQS SNFDP MKGGN FKDRS SGPYD
KGGQY FAKPR NQGGY GGSSS SKSYG SDRRF

+7R+12D
MASAD SSQRD RDDRG NFGDG RGGGF GGNDN FGRGG
NFSDR GGFGG SRGDG RYGGD GDRYN GFGND GRNFG
GGGSY NDFGN YNNQS SNFDP MKGGN FRDRS SGPYD
RGGQY FAKPR NQGGY GGSSS SRSYG SDRRF

−9F+3Y
MASAS SSQRG RSGSG NFGGG RGGGY GGNDN GGRGG
NYSGR GGFGG SRGGG GYGGS GDGYN GGGND GSNYG
GGGSY NDSGN GNNQS SNFGP MKGGN YGGRS SGGSG
GGGQY GAKPR NQGGY GGSSS SSSYG SGRRS

−12F+12Y
MASAS SSQRG RSGSG NYGGG RGGGY GGNDN YGRGG
NYSGR GGYGG SRGGG GYGGS GDGYN GYGND GSNYG
GGGSY NDYGN YNNQS SNYGP MKGGN YGGRS SGGSG
GGGQY YAKPR NQGGY GGSSS SSSYG SGRRY

We have estimated the upper critical solution temperatures of these hnRNPA1 variants from the experimental phase diagrams given by Bremer and co-workers [10]. It is possible in the first instance to estimate critical temperatures by visual inspection in the light of the law of rectilinear diameter [83], which provides an initial crude estimate. To quantify the data more systematically, we fitted the experimental coexistence data points [10] to

$$T_{\mathrm{coex}} = \alpha \frac{\left(\frac{c_{\mathrm{coex}}}{\beta} - \gamma\right)^2 - \left(\frac{c_{\mathrm{coex}}}{\beta} - \gamma\right)}{(r-1)\left(\frac{c_{\mathrm{coex}}}{\beta} - \gamma\right) + 1}, \tag{S1}$$

where c_{coex} is the concentration at coexistence, and α, β, γ and r are fitting parameters. This is a slightly generalised form of the spinodal curve arising from Flory–Huggins–Staverman theory, chosen here solely because the resulting function has the desired shape. With this approach, we can obtain critical temperatures in a systematic way for all variants considered that have data points reported for both the vapour-like and the liquid-like branch. We show two examples of such fitting in

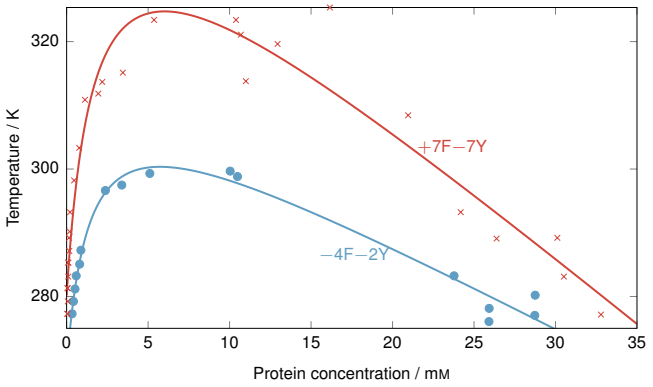

Figure S1. **Estimation of experimental critical temperatures.** We show data points reproduced from the work of Bremer and co-workers [10] alongside fits to Eq. (S1) for two variants of hnRNPA1. The maximum of the fit is taken to correspond to the critical temperature.

Table IV. Experimental upper critical solution temperatures of the hnRNPA1 variants studied, estimated from the phase diagrams reported by Bremer and co-workers [10]. Data points reported to only 3 significant figures are obtained by manually extrapolating the data points and assuming typical binodal behaviour.

Variant	WT	−3R+3K	−4F−2Y	−6R+6K	+7F−7Y	+7K+12D	+7R+12D	−9F+3Y	−12F+12Y
T_c / K	335.9	308.8	300.3	288.0	324.7	333	358	285	334.1

Supplementary Figure S1, and our estimates of the critical temperature in SI Table IV. However, these estimates should be interpreted with a pinch of salt: they are likely to give us the correct ordering, but the error associated with the numerical values is likely to be not insignificant. There were three variants of hnRNPA1 for which insufficient high-density data were reported [10] for the fit to Eq. (S1) to be possible and for which experimental T_c estimates are more approximate [see SI Table IV]; however, even if we remove these variants from further analysis, this does not significantly affect the Pearson coefficients.

S2.3 Additional protein sequences

The sequence of the FUS protein and the variants [11] we have considered is shown below. Changes in sequences are not highlighted for the first three variants since they entail not just single-point mutations, but also additions of residues within the chain that shift the remainder of the sequence.

[UniProt sequence P35637-1]

FUS
MASND YTQQA TQSYG AYPTQ PGQGY SQQSS QPYGQ
QSYSG YSQST DTSGY GQSSY SSYGQ SQNTG YGTQS
TPQGY GSTGG YGSSQ SSQSS YGQQS SYPGY GQQPA
PSSTS GSYGS SSQSS SYGQP QSGSY SQQPS YGGQQ
QSYGQ QQSYN PPQGY GQQNQ YNSSS GGGGG GGGGG
NYGQD QSSMS SGGGS GGGYG NQDQS GGGGS GGYGQ
QDRGG RGRGG SGGGG GGGGG GYNRS SGGYE PRGRG
GGRGG RGGMG GSDRG GFNKF GGPRD QGSRH DSEQD
NSDNN TIFVQ GLGEN VTIES VADYF KQIGI IKTNK
KTGQP MINLY TDRET GKLKG EATVS FDDPP SAKAA
IDWFD GKEFS GNPIK VSFAT RRADF NRGGG NGRGG
RGRGG PMGRG GYGGG GSGGG GRGGF PSGGG GGGGQ
QRAGD WKCPN PTCEN MNFSW RNECN QCKAP KPDGP
GGGPG GSHMG GNYGD DRRGG RGGYD RGGYR GRGGD
RGGFR GGRGG GDRGG FGPGK MDSRG EHRQD RRERP
Y

27R
MASND YTQQA RQSYG AYPTQ PRQGY SQQRS QPYGQ
QSYSG YSQRT DRSGY GQSSY SSYGQ RQNTG YGTQR
TPQGY GSRGG YGSRQ SRQSS YGQQS SYPGY GQQPA
PRSRS GSYGS SRQSS SYGQP QRGSY SQQPS YGGRQ
QSYGQ RQSYN PPQGY GQRNQ YNSSR GRGRG RGRGG
NYGQD QRSMS RGGGR GGGYG NQDQR GGGRS GGYGQ
QASDR GGRGR GGSGG GGGGG GGGYN RSSGG YEPRG
RGGGR GGRGG MGGSD RGGFN KFGGP RDQGS RHDSE
QDNSD NNTIF VQGLG ENVTI ESVAD YFKQI GIIKT
NKKTG QPMIN LYTDR ETGKL KGEAT VSFDD PPSAK
AAIDW FDGKE FSGNP IKVSF ATRRA DFNRG GGNGR
GGRGR GGPMG RGGYG GGGSG GGGRG GFPSG GGGGG
GQQRA GDWKC PNPTC ENMNF SWRNE CNQCK APKPD
GPGGG PGGSH MGGNY GDDRR GGRGG YDRGG YRGRG
GDRGG FRGGR GGGDR GGFGP GKMDS RGEHR QDRRE
R

PLD Y→F
MASND FTQQA TQSFG AFPTQ PGQGF SQQSS QPFGQ
QSFSG FSQST DTSGF GQSSF SSFGQ SQNTG FGTQS
TPQGF GSTGG FGSSQ SSQSS FGQQS SFPGF GQQPA
PSSTS GSFGS SSQSS SFGQP QSGSF SQQPS FGGQQ
QSFGQ QQSFN PPQGF GQQNQ FNSSS GGGGG GGGGG
NFGQD QSSMS SGGGS GGGFG NQDQS GGGGS GGFGQ
QASDR GGRGR GGSGG GGGGG GGGYN RSSGG YEPRG
RGGGR GGRGG MGGSD RGGFN KFGGP RDQGS RHDSE
QDNSD NNTIF VQGLG ENVTI ESVAD YFKQI GIIKT
NKKTG QPMIN LYTDR ETGKL KGEAT VSFDD PPSAK
AAIDW FDGKE FSGNP IKVSF ATRRA DFNRG GGNGR
GGRGR GGPMG RGGYG GGGSG GGGRG GFPSG GGGGG
GQQRA GDWKC PNPTC ENMNF SWRNE CNQCK APKPD
GPGGG PGGSH MGGNY GDDRR GGRGG YDRGG YRGRG
GDRGG FRGGR GGGDR GGFGP GKMDS RGEHR QDRRE
R

RBD R→G
MASND YTQQA TQSYG AYPTQ PGQGY SQQSS QPYGQ
QSYSG YSQST DTSGY GQSSY SSYGQ SQNTG YGTQS
TPQGY GSTGG YGSSQ SSQSS YGQQS SYPGY GQQPA
PSSTS GSYGS SSQSS SYGQP QSGSY SQQPS YGGQQ
QSYGQ QQSYN PPQGY GQQNQ YNSSS GGGGG GGGGG
NYGQD QSSMS SGGGS GGGYG NQDQS GGGGS GGYGQ
QASDG GGGGG GGSGG GGGGG GGGYN RSSGG YEPGG
GGGGG GGGGG MGGSD GGGFN KFGGP RDQGS RHDSE
QDNSD NNTIF VQGLG ENVTI ESVAD YFKQI GIIKT
NKKTG QPMIN LYTDR ETGKL KGEAT VSFDD PPSAK
AAIDW FDGKE FSGNP IKVSF ATRRA DFNGG GGNGG
GGGGG GGPMG GGGYG GGGSG GGGGG GFPSG GGGGG
GQQRA GDWKC PNPTC ENMNF SWRNE CNQCK APKPD
GPGGG PGGSH MGGNY GDDRG GGGGG YDGGG YGGGG
GDGGG FGGGG GGGDG GGFGP GKMDS GGEHR QDRRE
R

PLD 6D
MASND YTQQA TQSYD AYPTQ PGQGY DQQSS QPYDQ
QSYDG YDQST DTSGY DQSSY SSYGQ SQNTG YGTQS
TPQGY GSTGG YGSSQ SSQSS YGQQS SYPGY GQQPA
PSSTS GSYGS SSQSS SYGQP QSGSY SQQPS YGGQQ
QSYGQ QQSYN PPQGY GQQNQ YNSSS GGGGG GGGGG
NYGQD QSSMS SGGGS GGGYG NQDQS GGGGS GGYGQ
QDRGG RGRGG SGGGG GGGGG GYNRS SGGYE PRGRG
GGRGG RGGMG GSDRG GFNKF GGPRD QGSRH DSEQD
NSDNN TIFVQ GLGEN VTIES VADYF KQIGI IKTNK
KTGQP MINLY TDRET GKLKG EATVS FDDPP SAKAA
IDWFD GKEFS GNPIK VSFAT RRADF NRGGG NGRGG
RGRGG PMGRG GYGGG GSGGG GRGGF PSGGG GGGGQ
QRAGD WKCPN PTCEN MNFSW RNECN QCKAP KPDGP

GGGPG GSHMG GNYGD DRRGG RGGYD RGGYR GRGGD
RGGFR GGRGG GDRGG FGPGK MDSRG EHRQD RRERP
Y

The sequence of the FUS PLD region is shown below.

[residues 1–163 of UniProt sequence P35637-1]

FUS PLD MASND YTQQA TQSYG AYPTQ PGQGY SQQSS QPYGQ
QSYSG YSQST DTSGY GQSSY SSYGQ SQNTG YGTQS
TPQGY GSTGG YGSSQ SSQSS YGQQS SYPGY GQQPA
PSSTS GSYGS SSQSS SYGQP QSGSY SQQPS YGGQQ
QSYGQ QQSYN PPQGY GQQNQ YNS

We also considered LAF-1 RGG and two of its variants [84]:

[residues 1–168 of UniProt sequence D0PV95-1]

LAF-1 RGG MESNQ SNNGG SGNAA LNRGG RYVPP HLRGG DGGAA
AAASA GGDDR RGGAG GGGYR RGGGN SGGGG GGGYD
RGYND NRDDR DNRGG SGGYG RDRNY EDRGY NGGGG
GGGNR GYNNN RGGGG GGYNR QDRGD GGSSN FSRGG
YNNRD EGSDN RGSGR SYNND RRDNG GDG

Y→F MESNQ SNNGG SGNAA LNRGG RFVPP HLRGG DGGAA
AAASA GGDDR RGGAG GGGFR RGGGN SGGGG GGGFD
RGFND NRDDR DNRGG SGGFG RDRNW EDRGF NGGGG
GGGNR GFNNN RGGGG GGFNR QDRGD GGSSN FSRGG
FNNRD EGSDN RGSGR SFNND RRDNG GDG

R→K MESNQ SNNGG SGNAA LNKGG KYVPP HLKGG DGGAA
AAASA GGDDK KGGAG GGGYK KGGGN SGGGG GGGYD
KGYND NKDDK DNKGG SGGYG KDKNY EDKGY NGGGG
GGGNK GYNNN KGGGG GGYNK QDKGD GGSSN FSKGG
YNNKD EGSDN KGSGK SYNND KKDNG GDG

Finally, we give the sequences of DDX4-LCD and three of its variants [40]:

[residues 1–236 of UniProt sequence Q9NQI0-1]

DDX4-LCD MGDED WEAEI NPHMS SYVPI FEKDR YSGEN GDNFN
RTPAS SSEMD DGPSR RDHFM KSGFA SGRNF GNRDA
GECNK RDNTS TMGGF GVGKS FGNRG FSNSR FEDGD
SSGFW RESSN DCEDN PTRNR GFSKR GGYRD GNNSE
ASGPY RRGGR GSFRG CRGGF GLGSP NNDLD PDECM
QRTGG LFGSR RPVLS GTGNG DTSQS RSGSG SERGG
YKGLN EEVIT GSGKN SWKSE AEGGE S

CS MGDRD WRAEI NPHMS SYVPI FEKDR YSGEN GRNFN
DTPAS SSEMR DGPSE RDHFM KSGFA SGDNF GNRDA
GKCNE RDNTS TMGGF GVGKS FGNEG FSNSR FERGD
SSGFW RESSN DCRDN PTRND GFSDR GGYEK GNNSE
ASGPY ERGGR GSFDG CRGGF GLGSP NNRLD PRECM
QRTGG LFGSD RPVLS GTGNG DTSQS RSGSG SERGG
YKGLN EKVIT GSGEN SWKSE ARGGE S

F→A MGDED WEAEI NPHMS SYVPI AEKDR YSGEN GDNAN
RTPAS SSEMD DGPSR RDHAM KSGAA SGRNA GNRDA
GECNK RDNTS TMGGA GVGKS AGNRG ASNSR AEDGD
SSGAW RESSN DCEDN PTRNR GASKR GGYRD GNNSE
ASGPY RRGGR GSARG CRGGA GLGSP NNDLD PDECM
QRTGG LAGSR RPVLS GTGNG DTSQS RSGSG SERGG
YKGLN EEVIT GSGKN SWKSE AEGGE S

R→K MGDED WEAEI NPHMS SYVPI FEKDK YSGEN GDNFN
KTPAS SSEMD DGPSK KDHFM KSGFA SGKNF GNKDA
GECNK KDNTS TMGGF GVGKS FGNKG FSNSK FEDGD
SSGFW KESSN DCEDN PTKNK GFSKK GGYKD GNNSE
ASGPY KKGGK GSFKG CKGGF GLGSP NNDLD PDECM
QKTGG LFGSK KPVLS GTGNG DTSQS KSGSG SEKGG
YKGLN EEVIT GSGKN SWKSE AEGGE S

S3 ADDITIONAL PROTEIN POTENTIALS OF MEAN FORCE DATA

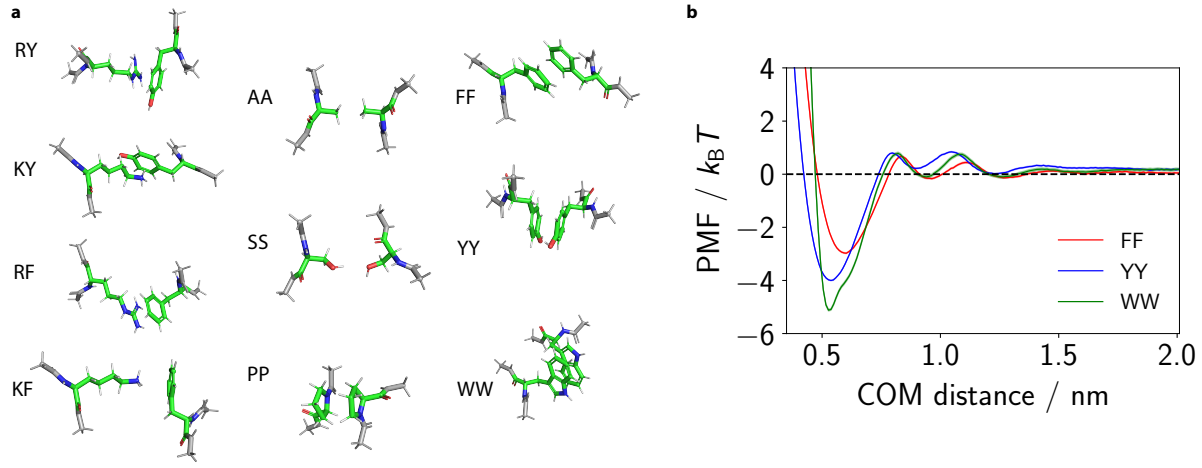

Figure S2. **Additional protein potentials of mean force.** **a** Orientation of amino-acid pairs for PMFs reported in Fig. 2 in the main text. Pairs are labelled using the one-letter codes for the amino acids in question (see SI Table S1). **b** PMFs for amino acids with aromatic side chains, specifically FF, YY and WW. PMFs are computed as described in the main text. Statistical errors (mean±s.d.) are given as error bands; they were computed via Bayesian bootstrapping [85] of 3 independent simulations.

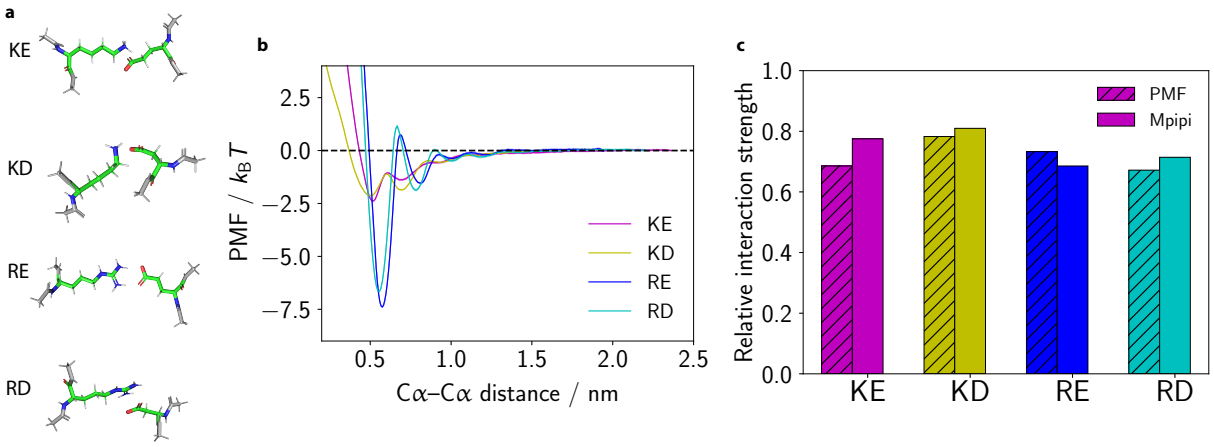

Figure S3. **Comparison of selected attractive charge–charge residue interactions.** **a** Orientation of selected charge–charge amino acid pairs. Pairs are labelled using the one-letter-codes for the amino acids in question (see SI Table S1). **b** PMFs for charge–charge residue pairs, specifically RE, RD, KE and KD. PMFs are computed as described in the main text. Statistical errors (mean±s.d.) are given as error bands; they were computed via Bayesian bootstrapping of 3 independent simulations. **c** Integrated interaction strengths for selected oppositely charged pairs, as implemented in the Mpipi model, normalised based on the RY interaction (as in the main text). The combined potential is fitted so that the ratio of the mean RE/KE and RD/KD strengths closely matches those at the atomistic resolution. When fitting for the Lys-based oppositely charged interactions, we include the first two minima of the PMF since the barrier between them is so small; in the high-temperature limit, these are therefore comparable to the Arg interaction strengths. Differences between Lys and Arg are mainly captured via their short-ranged pairwise contacts and via their homotypic contacts, with Arg parameterised as a stronger sticker and less self-repulsive than Lys (see Fig. 2e in the main text).

S4 RNA POTENTIALS OF MEAN FORCE

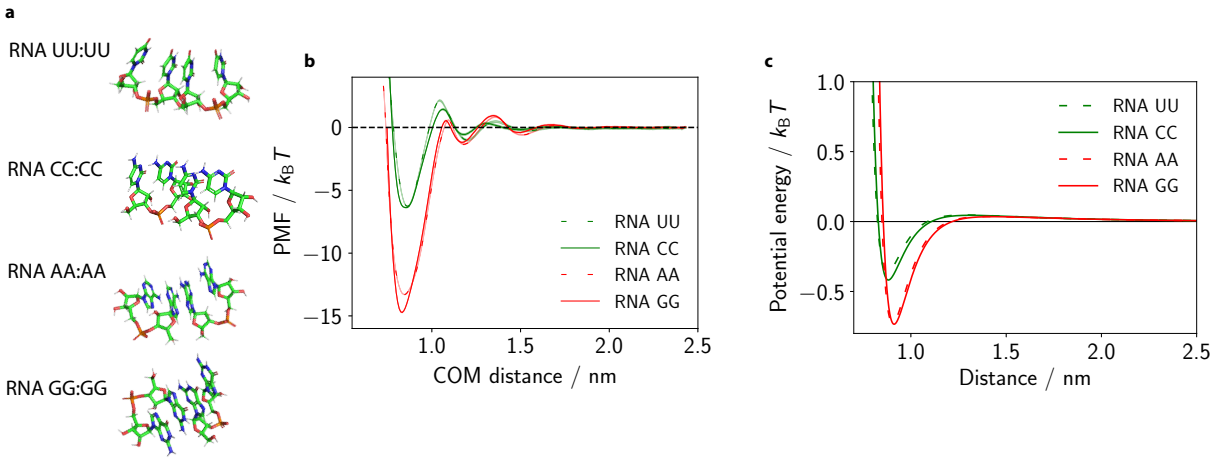

Figure S4. **Comparison of nucleic-acid dimer pair interactions.** **a** Orientation of dimers. Dimer pairs are labelled using the one-letter-codes: adenosine (A), cytidine (C), guanosine (G), uridine (U). **b** PMFs for RNA dimer pairs. PMFs are computed as described in the main text. Statistical errors (mean±s.d.) are given as error bands; they were computed via Bayesian bootstrapping of 3 independent simulations. **c** Pairwise interaction for selected RNA nucleic acid pairs, as implemented in the Mpipi model. Each curve represents the sum of the Wang–Frenkel and Debye–Hückel terms. Typically the atomistic PMF well depths are approximately an order of magnitude greater than the Mpipi model analogues. This disparity is mainly due to the constraints used when computing the PMFs, which allow enhanced sampling of the optimal interaction mode, ignoring other degrees of freedom. The use of explicit solvent in the atomistic calculations versus implicit solvent in the coarse-grained model leads to further differences in the well-depths. Here, the PMF well depths are about 20 times the coarse-grained model, which stems from the aforementioned factors as well as the use of dimer pairs instead of monomer pairs when computing the RNA PMFs.

S5 HPS+CATION–π(i) BENCHMARKS

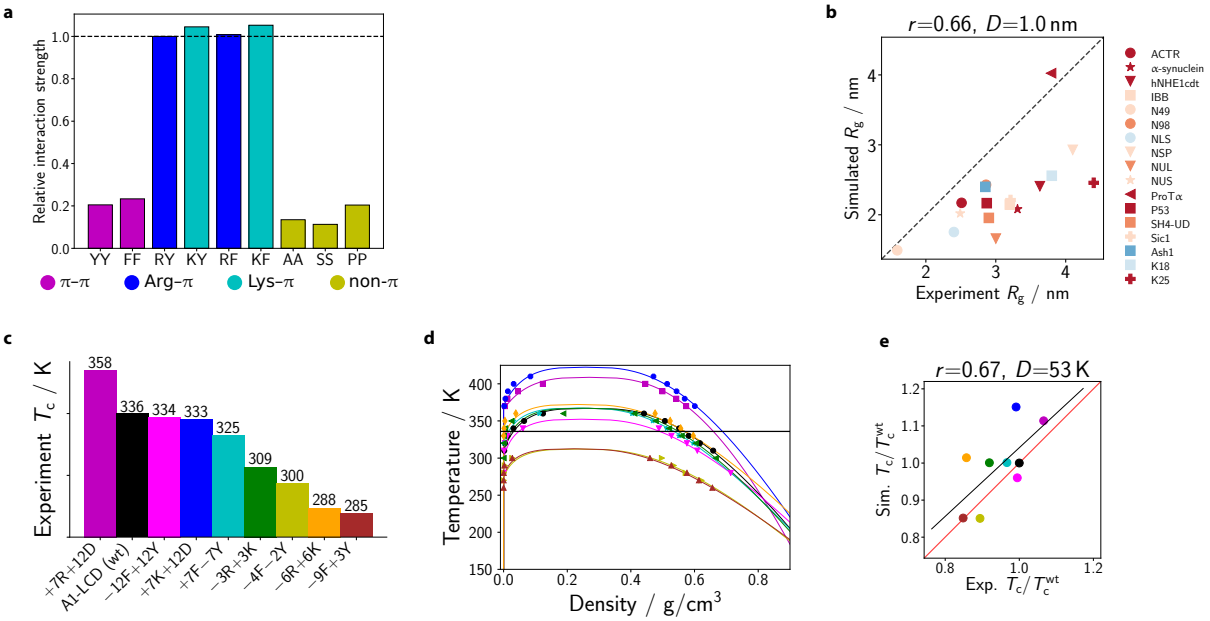

Figure S5. **Data obtained using the HPS+cation–π(i) model.** [40] Descriptions of individual panels are the same as for Supplementary Figure S6 immediately below.

S6 TSCL-M2 BENCHMARKS

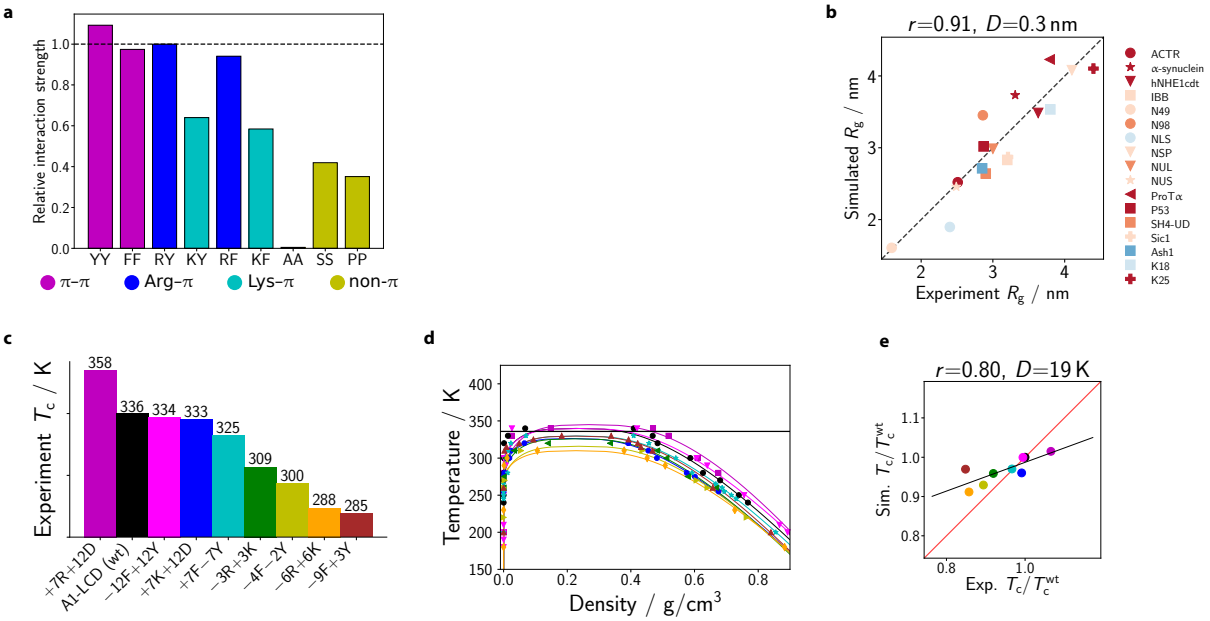

Figure S6. **Data obtained using the M2 parameter set of Tesei et al.** [27] **a** Relative interaction strengths for selected residue pairs, normalised relative to the Arg–Tyr (RY) interaction. A horizontal dashed line at the RY interaction strength is provided for comparison purposes (cf. Fig. 3 of the main text). **b** Comparison of simulated and experimental R_g. Each protein is coloured based on its dominant residue class (as categorised in Fig. 4a of the main text and excluding the 'neutral' class). The dashed line represents a 'perfect fit'. The Pearson correlation coefficient (r) and the root mean squared deviation from the experimental values (D) are reported in the plot title. **c** Experimental critical temperatures of hnRNPA1-LCD variants, reproduced from main text. The colour of each variant used in panel **c** is used in all remaining panels. **d** Phase diagrams for hnRNPA1-LCD variants obtained via direct-coexistence simulations. Estimation of critical points of phase diagrams is described in the main text. Curves are derived from empirical fits of the data to Eqs (6) and (7) of the main text. **e** Simulated critical temperature T_c relative to the critical temperature of the wild type (T_c^{wt}) shown against the experimental analogue. The Pearson correlation coefficient (r) and the root mean squared deviation (D) are provided in the plot title.

S7 LLPS PROPENSITY OF OTHER PROTEINS

For FUS PLD and LAF-1 RGG, experimental critical temperatures are not yet available for direct comparison; however, in vitro studies, including fluorescence microscopy and temperature-dependent turbidity measurements, provide strong evidence for the relative LLPS propensities of these proteins [84, 86]. In addition, the Absinth (self-Assembly of Biomolecules Studied by an Implicit, Novel, Tunable Hamiltonian) [69] potential, which is known to reproduce well experimental conformational ensembles of IDRs, is employed here to obtain estimates of T_c for these proteins. Specifically, using Absinth, we compute the temperature for single-molecule collapse (T_θ), which can be used to infer experimental critical temperatures. For example, the critical temperatures of several proteins are well estimated by their corresponding collapse temperatures computed with Absinth [4, 24]. However, beyond probing single-molecule properties, Absinth is computationally expensive and therefore not applicable to multicomponent LLPS systems.

For the LAF-1 RGG, in vitro studies suggest that the relative ordering of T_c for the WT domain and its mutants is WT>Y→F>R→K [84]. The Mpipi model correctly predicts this trend (Fig. 6a of the main text). Moreover, the critical temperature of LAF-1 RGG (WT) obtained via the Mpipi model (330 K) coincides with the Absinth estimate ($T_\theta \approx 330$ K; see black dotted line in Fig. 6a of the main text). We also employed Mpipi to compute the phase diagram for the FUS PLD (magenta curve in Fig. 6a of the main text) and estimated the temperature for single-molecule collapse via Absinth (magenta dotted line in Fig. 6a of the main text). Here, our critical temperature estimate (340 K) is 8 K higher than T_θ (~332 K). Besides an abundance of Tyr residues, about 65 % of FUS PLD is composed on Ser, Gln and Gly residues. As pointed out in the main text, these residues are commonly classified as spacers. Thus, we speculate that the discrepancy between T_c and T_θ may suggest that the interactions involving these residues may be too strong within the model. We plan to interrogate this point further as more data become available.

We also compute phase diagrams for four variants of the DDX4 NTD (Fig. 6b of the main text). Specifically, we assess the phase behaviour of the WT IDR, the charged-scrambled (CS) variant, a variant where Phe is replaced by Ala (F→A), and one where Arg residues are substituted by Lys (R→K). The LLPS propensities of the DDX4 NTD variants have been thoroughly characterised by Brady and colleagues [14]. They concluded that, although scrambling charges (i.e. the CS variant) reduces the propensity for LLPS of DDX4 NTD, the F→A and R→K mutations result in the IDR not exhibiting phase separation at all the conditions tested [14]. Our computed phase diagrams agree qualitatively with these experimental predictions [14], with T_c decreasing in the order WT>CS≫F→A>R→K. In experiments, the highest temperature at which LLPS is observed for the WT and CS variants differs by approximately 30 K at 100 mM salt, whilst in our model, parameterised at 150 mM salt, the predicted critical temperatures for the WT and CS variants differ by 6 K (Fig. 6b of the main text). Since our coarse-grained beads are isotropic, i.e. they do not explicitly account for orientation-dependent interactions that are likely to be important in charged residues, the effects of charge segregation are less pronounced within our potential. To capture these effects better, it may in the future be useful to include a degree of anisotropic character for some interactions.

In addition to the preceding IDRs, we compute phase diagrams for the full-length FUS protein (FUS WT) and four additional FUS variants whose protein sequences are provided in Sec. S2. Wang *et al.* used fluorescence imaging to determine the saturation concentrations for various FUS mutants [11]. Compared to the WT protein, lower saturation concentrations were reported for the 27R and PLD 6D mutants, suggesting that these mutations enhance LLPS propensities. By contrast, the PLD Y→F and RBD R→G mutants both displayed higher saturation concentrations than the WT protein. We computed the critical temperatures for LLPS for each FUS variant and found that the order of T_c is consistent with the in vitro LLPS propensities, i.e. T_c values decrease in the order 27R>PLD 6D>WT>PLD Y→F>RBD R→G (Fig. 6c of the main text).

S8 SUPPLEMENTARY METHODS: FIGURES AND TABLES

S8.1 Finite-size scaling

Table IX. System sizes used in phase-diagram calculations. The box lengths L_x and L_y are always the same. For hnRNPA1 variants, all simulations were performed with two system sizes.

System	Chains	Beads	L_x / nm	L_z / nm
hnRNPA1 + variants	64	8640	11.3	34.0
	63	8505	10.0	44.0
FUS-LCD	52	8476	12.0	44.0
FUS + variants	24	12 624	13.8	76.9
LAF-1 RGG + variants	100	16 800	14.3	70.8
DDX4 + variants	24	5664	16.3	67.3

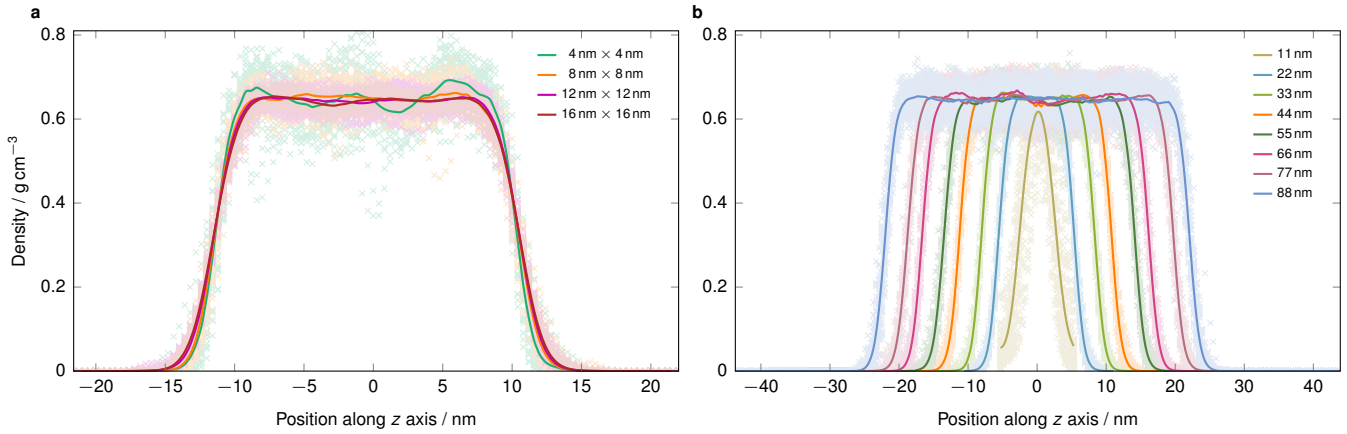

Figure S7. **Finite-size scaling analysis.** Plots of local density against the z-axis co-ordinate for FUS-LCD simulated with the Mpipi potential at 280 K and a constant number density of 1.85 nm^{-3}. The z axis is the longest of the three box dimensions. In panel **a**, the box size is fixed at 44 nm and the cross-sectional area of the interface changes as indicated in the legend. In panel **b**, the cross-sectional area of the interface is fixed at 8 nm × 8 nm, and the length of the box in the z direction changes as indicated in the legend. In each case, we show as pale symbols short-time averages of local densities calculated in 0.4-nm-wide bins along the z axis averaged over 80 000 measurements for each symbol, and repeated across long simulations, resulting in a spread of data points. We also show the mean value of these individual measurements as a solid line. In phase-diagram calculations, an average liquid-like density is then computed across the entire central portion of the density profiles shown.

S8.2 RNA–PolyR–PolyK systems

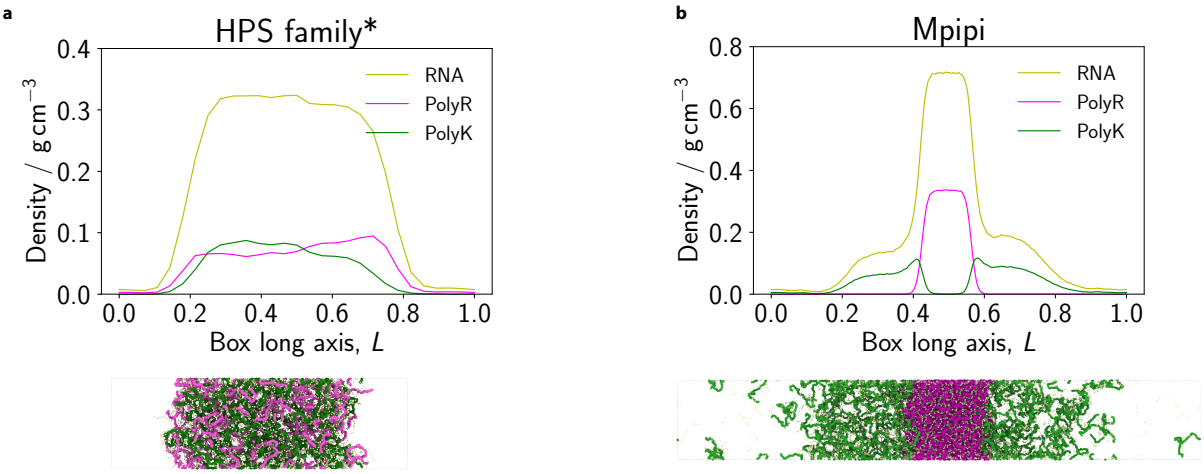

Figure S8. **Comparison of charge-matched RNA–PolyR–PolyK systems.** Plots of local density against the z-axis co-ordinate for RNA–PolyR–PolyK systems simulated with the **a** HPS-KR model and **b** Mpipi potential. **a** We simulate a mixture of PolyK (50 residues; 64 chains), PolyR (50 residues; 64 chains) and RNA (10 residues; 640 chains). For this simulation, we use the parameters for uridine, as proposed by Regy et al. [50]. We dubbed these simulations 'HPS family', which includes HPS-KR, HPS+cation–π(i) and HPS+cation–π(ii), since the Arg, Lys and RNA cross interactions are all the same in these three HPS-based models. **b** We simulate a mixture of PolyK (50 residues; 128 chains), PolyR (50 residues; 128 chains) and RNA (10 residues; 1280 chains). Here, we have extended the Mpipi model to include RNA parameters (see main text and Fig. S4). We also simulate a system containing 64, 64 and 640 chains of PolyK, PolyR and RNA, respectively, via the Mpipi potential and obtain similar multiphasic behaviour (see Fig. S9a). Simulation snapshots of each system are provided below the respective density plots. The colour code in the snapshot is consistent with that used in the density plots. The mixtures are simulated at $T/T_c \approx 0.8$, where T_c is the critical temperature for liquid–vapour phase separation.

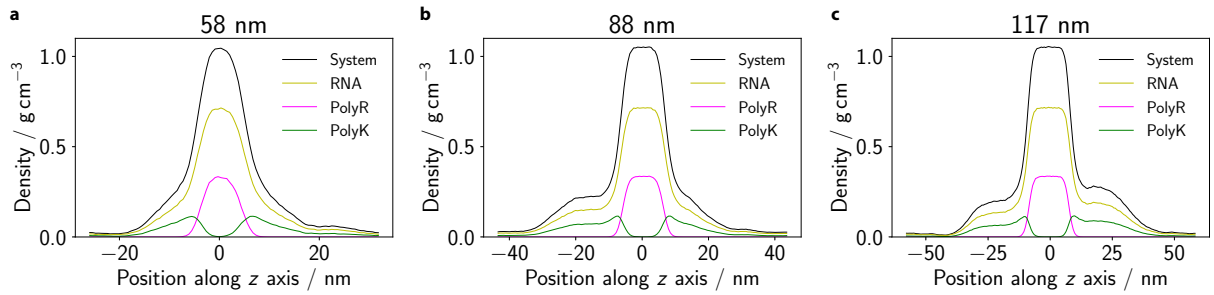

Figure S9. **Finite-size scaling analysis for the multiphase system.** Plots of local density against the longest (z) axis of the simulation box co-ordinate for RNA–PolyR–PolyK systems simulated with the Mpipi potential at $T/T_c \approx 0.8$, where T_c is the critical temperature for liquid–vapour phase separation. The lengths of the RNA, PolyR and PolyR chains are 10, 50, 50 beads, respectively. In **a**, 640 chains of RNA are used and 64 chains each of PolyR and PolyK. These numbers are increased by a factor of **b** 1.5 and **c** 2 in order to assess finite-size effects. The cross-sectional area of the interface in each simulation is fixed at 17.5 nm × 17.5 nm and the longest box dimension changes as indicated in the plot titles.

S8.3 Summary of potential parameters for all models

In SI Table X, we provide a summary of the parameters for each model we have considered using the same units throughout for ease of comparison. The spring constant refers to the harmonic spring constant k in the bonded term, $(k/2)(r-r_{\mathrm{ref}})^2$ (Eq. (2) of the main text), where r_{eff} is the equilibrium bond length for bonded amino-acid residues. Care must be taken not to confuse this quantity with the LAMMPS implementation of harmonic potentials in which the force constant is defined as $K = k/2$. In particular, as in our model, the HPS-KR and KH models [28] use $k = 8.03\,\mathrm{J\,mol^{-1}\,pm^{-2}}$, while the cation–π [40] and FB-HPS [26] reparameterisations use $k = 2.39\,\mathrm{kcal\,mol^{-1}\,Å^{-2}} = 1.0\,\mathrm{J\,mol^{-1}\,pm^{-2}}$ for the spring constant, and the HPS-Urry potential [29] uses $k = 4.0\,\mathrm{J\,mol^{-1}\,pm^{-2}}$. Although the value for the Coulomb cutoff is not reported in all studies, we have used a consistent value of 3.5 nm, except for the FB-HPS and TSCL-M2 models, where Coulomb cutoffs are specified. Although Dannenhoffer-Lafage and Best report a non-bonded potential cutoff of 1.0 nm [26], we presume this holds only for their calculation of the radii of gyration in single-molecule simulations, and we have used a 3σ cutoff, as for the remaining potentials, in order to observe any phase separation.

These parameter values have not always been reported in the literature very clearly, resulting in some confusion regarding both units and the inclusion, or otherwise, of a factor of $1/2$ in the spring constant. As a result, the HPS+cation–π models use the value for the spring constant as reported by Dignon and co-workers [28], although the intended parameterisation of the HPS-KR and KH potentials was as outlined in SI Table X. We have confirmed the latter point with one of the authors of Ref. 28.

Whilst our manuscript was under review, a pre-print with a further model ('TSCL') appeared on the bioRxiv [27]. For completeness, we also include the parameters for this potential in the table below. Interestingly, the dielectric constant in a temperature-dependent way using an empirical relation given by Akerlof and Oshry [87]. This in turn affects the Debye length. We found this to be an interesting and reasonably straightforward addition to this family of models; out of interest, we attempted a series of phase-diagram calculations for the Mpipi model with temperature-dependent dielectric constant and screening length. We have observed no appreciable difference in the behaviour of the A1-LCD variants, which suggests that within the relatively narrow range of temperatures at which we studied our systems, the relatively small reweighting of electrostatic interactions that this results in is not sufficient to change the phase behaviour significantly.

Finally, we remark that when representing non-bonded interactions with Lennard-Jones-based potentials, care must be taken in specifying and reporting the cutoffs used. The choice of the cutoff for non-bonded interactions can greatly affect the phase behaviour predicted by the model. For example, when we originally implemented the TSCL-M2 model, we used a cutoff of 3σ, as had been used for the remaining potentials, but this resulted in behaviour that was significantly different from that reported by the authors [27]. With their kind assistance, we eventually pinned down that the root cause of the discrepancy was this non-bonded cutoff, for which they used a value of 4 nm instead. The larger (4 nm) cutoff results in an increase in T_c of the order of 60 K. Similarly, in our initial implementation of the HPS-Urry potential [29], since no cutoff was reported in the manuscript, we initially used 3σ for the non-bonded cutoff; however, their supporting code suggests a cutoff of 2 nm may have been used instead, which again increases the critical temperatures, in this case by ∼20 K. Since the Wang–Frenkel potential is not truncated and therefore not subject to such abrupt changes with cutoff variations, this serves further to highlight yet another advantage of using it in preference to LJ-based potentials. For some of the LJ-based potentials we benchmark here, cutoff details were not specified in the corresponding manuscripts. Since all were derived from HPS, we have used a cutoff of 3σ in such cases, but we remark that, given the sensitivity of the phase behaviour to the details of the parameterisation, this may be a potential source of inconsistencies between implementations.

A full listing of all interaction parameters of the Mpipi potential is given in SI Tables XI and XII, and a LAMMPS implementation is provided as part of the supporting code.

Table X. Summary of model parameters. WF = Wang–Frenkel potential [19]. AH/LJ = Ashbaugh–Hatch-modified Lennard-Jones potential [88]. The spring constant refers to the harmonic spring constant k in the bonded term, $(k/2)(r - r_{\mathrm{ref}})^2$ (Eq. (2) of the main text), where r_{eff} is the equilibrium bond length for bonded amino-acid residues.

Model	Spring constant k / J mol^{-1} pm^{-2}	r_{eff} / nm	Charge on H / e	Short-ranged potential	Cutoff	Screening κ / nm^{-1}	Debye cutoff / nm
Mpipi	8.03	0.381	0.375	WF	3σ	1.26	3.5
KH[28]	8.03	0.381	0.5	AH/LJ	3σ	1.0	3.5
HPS-KR[28]	8.03	0.381	0.5	AH/LJ	3σ	1.0	3.5
FB-HPS[26]	1.0	0.38	0.5	AH/LJ	3σ*	1.0	3.0
HPS-Urry[29]	4.0	0.382	0	AH/LJ	2.0 nm	1.0	3.5
HPS+cation–π(i)[40]	1.0	0.38	0.5	AH/LJ	3σ	1.0	3.5
HPS+cation–π(ii)[40]	1.0	0.38	0.5	AH/LJ	3σ	1.0	3.5
TSCL-M2[27]	8.03	0.38	variable	AH/LJ	4.0 nm	variable	4.0

Table XI. Wang–Frenkel parameters for the Mpipi potential. For each residue, two lines are provided. The row highlighted in red lists ε/kJ mol^{-1} and the row highlighted in green lists σ/nm. The value of μ is 2 for all entries except the ones highlighted in blue [μ(V–I) = 4, μ(I–I) = 11]. All charged amino acids have $q = \pm 0.75e$, as appropriate, except H, which has $q = 0.375e$, where e is the elementary charge.

	M	G	K	T	R	A	D	E	Y	V	L	Q	W	F	S	H	N	P	C	I
M	0.165536	0.284583	0.136758	0.147114	0.717619	0.186280	0.306152	0.324582	0.959705	0.094437	0.105776	0.502105	1.233991	0.902083	0.211635	0.851406	0.488298	0.247358	0.227764	0.083592
	0.646795	0.557618	0.656778	0.617823	0.664396	0.586850	0.614146	0.631818	0.659004	0.632057	0.648344	0.636519	0.675573	0.653821	0.593894	0.639251	0.618802	0.613446	0.609429	0.649639
G		0.403630	0.255806	0.266161	0.836666	0.305327	0.425199	0.443630	1.078752	0.213484	0.224823	0.621152	1.353039	1.021130	0.330682	0.970454	0.607345	0.366405	0.346812	0.202639
		0.469511	0.567134	0.528442	0.576639	0.498013	0.525925	0.543637	0.571317	0.541239	0.557893	0.548630	0.587924	0.566122	0.505252	0.551541	0.530902	0.524999	0.520872	0.558033
K			0.079986	0.118336	0.510189	0.157502	0.002063	0.002201	0.429701	0.065660	0.076998	0.473328	0.438466	0.482315	0.182858	0.365556	0.459520	0.218581	0.198987	0.054815
			0.667134	0.627940	0.673819	0.596669	0.623699	0.641358	0.668408	0.643354	0.659256	0.645969	0.684969	0.663228	0.603618	0.648660	0.628255	0.623106	0.619115	0.661800
T				0.128691	0.699197	0.167858	0.287729	0.306160	0.941283	0.076015	0.087354	0.483683	1.215565	0.883661	0.193213	0.832984	0.469876	0.228936	0.209342	0.065170
				0.588906	0.635202	0.557795	0.584987	0.602654	0.629806	0.603697	0.619809	0.607329	0.646378	0.624625	0.564801	0.610053	0.589610	0.584320	0.580321	0.621462
R					0.376209	0.738363	0.006368	0.006502	2.543119	0.646520	0.657859	1.054188	2.893893	2.280037	0.763718	0.519808	1.040381	0.799441	0.779847	0.635675
					0.683905	0.604909	0.632986	0.650709	0.678624	0.647780	0.664461	0.655828	0.695255	0.673424	0.612191	0.658837	0.638103	0.631985	0.627828	0.664491
A						0.207024	0.326896	0.345326	0.980449	0.115181	0.126520	0.522849	1.254736	0.922827	0.232379	0.872151	0.509046	0.268102	0.248509	0.104336
						0.527007	0.554510	0.572198	0.599537	0.571467	0.587918	0.576988	0.616122	0.594352	0.534119	0.579777	0.559268	0.553738	0.549687	0.588726
D							0.330938	0.344590	1.100321	0.235049	0.246387	0.642721	1.374603	1.042699	0.352247	0.007355	0.628914	0.387974	0.368380	0.224208
							0.582352	0.600058	0.627647	0.597874	0.614509	0.605004	0.644244	0.622456	0.561730	0.607876	0.587280	0.581455	0.577340	0.614699
E								0.358242	1.118751	0.253479	0.264818	0.661151	1.393034	1.061129	0.370677	0.007494	0.647344	0.406404	0.386811	0.242639
								0.617767	0.645370	0.615504	0.632146	0.622721	0.661969	0.640180	0.579422	0.625600	0.604999	0.599152	0.595036	0.632333
Y									1.753874	0.888606	0.899945	1.296274	2.028156	1.696252	1.005804	1.645576	1.282467	1.041527	1.021934	0.877761
									0.673363	0.642355	0.659039	0.650525	0.690005	0.668159	0.606828	0.653569	0.632798	0.626630	0.622465	0.659055
V										0.023338	0.034677	0.431006	1.162888	0.830984	0.140536	0.780308	0.417199	0.176259	0.156666	0.012493
										0.626600	0.639093	0.619973	0.658912	0.637178	0.578166	0.622610	0.602261	0.597433	0.593567	0.671801
L											0.046016	0.442345	1.174227	0.842323	0.151875	0.791646	0.428538	0.187598	0.168004	0.023832
											0.653407	0.636649	0.675594	0.653860	0.594697	0.639294	0.618936	0.614021	0.610125	0.660772
Q												0.838674	1.570556	1.238652	0.548204	1.187976	0.824867	0.583927	0.564334	0.420161
												0.627785	0.667141	0.645328	0.584255	0.630743	0.610059	0.604035	0.599885	0.636715
W													2.302443	1.970538	1.280086	1.919858	1.556753	1.315814	1.296220	1.152048
													0.706655	0.684798	0.623414	0.670209	0.649416	0.643217	0.639053	0.675603
F														1.638630	0.948182	1.587954	1.224845	0.983905	0.964312	0.820139
														0.662955	0.601640	0.648367	0.627603	0.621441	0.617279	0.653879
S															0.257734	0.897506	0.534397	0.293457	0.273864	0.129691
															0.541267	0.587063	0.566533	0.560932	0.556848	0.595253
H																0.113872	1.174168	0.933229	0.913635	0.769463
																0.633778	0.613018	0.606865	0.602702	0.639315
N																	0.811064	0.570124	0.550531	0.406358
																	0.592335	0.586308	0.582164	0.618999
P																		0.329185	0.309591	0.165419
																		0.580616	0.576515	0.614386
C																			0.289997	0.145825
																			0.572436	0.610587
I																				0.001653
																				0.692168

Table XII. Wang–Frenkel parameters for interactions with RNA bases in the Mpipi potential. For each residue, two lines are provided. The row highlighted in red lists ε/kJ mol^{-1} and the row highlighted in green lists σ/nm. The value of μ is 3 for all entries. All charged residues (both amino acids and nucleic-acid bases) have $q = \pm 0.75e$, as appropriate, except H, which has $q = 0.375e$, where e is the elementary charge. The nucleic-acid codes are shown in blue to distinguish them from corresponding amino-acid codes.

	A	C	G	U
M	1.153872	0.783588	1.162240	0.720828
	0.745398	0.734398	0.748897	0.731898
G	1.272919	0.902635	1.281287	0.839875
	0.656756	0.645756	0.660255	0.643255
K	0.666658	0.444487	0.671678	0.406831
	0.755567	0.744567	0.759067	0.742067
T	1.135450	0.765166	1.143818	0.702406
	0.716453	0.705453	0.719953	0.702953
R	2.518417	1.777849	2.535153	1.652329
	0.763953	0.752953	0.767453	0.750453
A	1.174616	0.804332	1.182984	0.741572
	0.685504	0.674504	0.689004	0.672004
D	1.236573	0.866289	1.244941	0.803529
	0.713176	0.702176	0.716676	0.699676
E	1.250225	0.879941	1.258593	0.817181
	0.730884	0.719884	0.734384	0.717384
Y	1.948041	1.577757	1.956409	1.514997
	0.758682	0.747682	0.762182	0.745182
V	1.082773	0.712489	1.091141	0.649729
	0.735300	0.724300	0.738800	0.721800
L	1.094112	0.723828	1.102480	0.661068
	0.748704	0.737704	0.752204	0.735204
Q	1.490441	1.120157	1.498809	1.057397
	0.735893	0.724893	0.739393	0.722393
W	2.222327	1.852043	2.230695	1.789279
	0.775327	0.764327	0.778827	0.761827
F	1.890419	1.520135	1.898787	1.457375
	0.753478	0.742478	0.756978	0.739978
S	1.199971	0.829687	1.208339	0.766927
	0.692634	0.681634	0.696134	0.679134
H	1.128040	0.757756	1.136408	0.694996
	0.738889	0.727889	0.742389	0.725389
N	1.476634	1.106354	1.485002	1.043590
	0.718168	0.707168	0.721668	0.704668
P	1.235698	0.865414	1.244066	0.802650
	0.712308	0.701308	0.715808	0.698808
C	1.216105	0.845821	1.224473	0.783061
	0.708218	0.697218	0.711718	0.694718
I	1.071932	0.701644	1.080300	0.638884
	0.768084	0.757084	0.771584	0.754584
A	2.142208	1.771924	2.150576	1.709164
	0.844000	0.833000	0.847500	0.830500
C		1.401640	1.780292	1.338880
		0.822000	0.836500	0.819500
G			2.158944	1.717532
			0.851000	0.834000
U				1.276120
				0.817000

S8.4 Estimation of error bars

Each individual simulation has a certain error associated with the density and the temperature; for example, in a typical thermostatted simulation with these models with system sizes we have studied, the mean temperature usually has a standard deviation of the order of 3 K. However, fluctuations about the mean are symmetrical for the temperature and the density, and so may not be reflected in the error associated with the value of the critical temperature. Rather than propagate the error from individual simulation data, we therefore estimate the error in our calculation of the critical temperature by running independent simulations to determine the phase diagram and in turn, using Eqs (6) and (7) of the main text, the upper critical solution temperature. For the hnRNPA1-LCD variants with the Mpipi potential, the standard deviation across 5 independent data points is given for each variant in SI Table XIII. The mean standard deviation is 1.8 K. The largest standard deviation corresponds to −6R+6K variant; the reason for this is that the density profile for −6R+6K exhibits considerable fluctuations within the high-density phase. These fluctuations are primarily caused by interactions between charged residues, and they make the density of the high-density phase more difficult to determine unambiguously. However, all estimates of T_c of this variant correspond to the lowest critical temperature of any variant. Moreover, the error in the critical temperature is not generally larger than the typical fluctuations in temperature for each simulation.

Table XIII. Standard deviation, including Bessel correction, for 5 independent estimates of the critical temperature for each hnRNPA1-LCD variant.

Variant	$\sigma(T_c)$ / K
WT	1.19
−3R+3K	2.58
−4F−2Y	0.74
−6R+6K	5.20
+7F−7Y	1.47
+7K+12D	0.86
+7R+12D	1.09
−9F+3Y	0.36
−12F+12Y	2.30

S8.5 List of algorithms and software

Table XIV. A summary of algorithms and software used in this work.

Method/software	Use
Umbrella sampling [89, 90]	probing interactions between pairs of amino acids and nucleic acids
WHAM [91]	obtaining PMF profiles from umbrella sampling runs
Bayesian bootstrap method [85]	estimating errors in PMF calculations
restrained electrostatic potential (RESP) [92]	refitting sidechain charges from quantum-mechanical calculations for cation–π pairs
Gaussian [93]	quantum-mechanical calculations on cation–π pairs
Absinth [69, 94]	estimating coil-to-globule transition temperature
Gromacs	PMF calculations
lincs [95]	rigid-molecule constraint algorithm in all-atom simulations
particle-mesh Ewald summation [96]	computation of long-ranged electrostatics
PyMol [97]	visualisation of nucleic-acid dimers in Fig. S4
direct-coexistence simulations [20, 98–100]	simulation set-up for computing phase diagrams
Lammps [101]	used to run simulations to obtain R_g values and to compute phase diagrams