How to cite: *Angew. Chem. Int. Ed.* **2022,** *61,* e202204311
International Edition: doi.org/10.1002/anie.202204311
German Edition: doi.org/10.1002/ange.202204311

Biophysics

Phase Separation of Heterogeneous Nuclear Ribonucleoprotein A1 upon Specific RNA-Binding Observed by Magnetic Resonance**

*Irina Ritsch[+], Elisabeth Lehmann[+], Leonidas Emmanouilidis, Maxim Yulikov, Frédéric Allain, and Gunnar Jeschke**

Abstract: Interaction of heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) with specific single-stranded RNA and its relation to liquid-liquid phase separation (LLPS) were studied in vitro by magnetic resonance based on site-directed spin labelling. An ensemble model of dispersed hnRNP A1 in the absence of RNA was derived from distance distributions between spin labelled sites and small angle X-ray scattering. This model revealed a compact state of the low-complexity domain and its interaction with the RNA recognition motifs. Paramagnetic relaxation enhancement NMR spectroscopy confirmed this interaction. Addition of RNA to dispersed hnRNP A1 induced liquid-droplet formation. Such LLPS depended on RNA concentration and sequence, with continuous wave EPR spectroscopy showing an influence of RNA point mutations on local protein dynamics. We propose that an interplay of sequence-specific RNA binding and LLPS contributes to regulation of specific RNA segregation during stress response.

Introduction

Liquid-liquid phase separation (LLPS) is a reversible demixing process that underlies the formation of membraneless organelles in cells.[1–3] These multi-component compartments play an important role in response to stress conditions[4] and can be studied in vitro via the formation of liquid droplets (LDs). The main cellular mediators of LLPS are proteins equipped with multivalent, highly flexible protein domains, or inter-domain linkers.[5,6] Here, we focus on the human heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1).[7] Among multiple functions of hnRNP A1 are alternative splicing regulation[8] and mRNA export from the nucleus to the cytoplasm.[9,10] In its central role for gene regulation, hnRNP A1 has implications in many diseases, for example as a target for viral infections,[11,12] cancer,[13] or in neurodegenerative diseases, where point mutations in hnRNP A1 have been linked to a variant of inclusion body myopathy as well as amyotrophic lateral sclerosis.[14] HnRNP A1 consists of two homologous but functionally non-equivalent RNA-recognition motif (RRM) domains that bind a large variety of RNA targets with some sequence specificity,[15,16] and a carboxy-terminal glycine-rich domain (residues 197–320). The protein lacking the G-rich domain, but containing both RRMs (residues 1–196), is known as unwinding protein 1 (UP1). The G-rich domain does not adopt a fixed conformation in solution[5] and is thus considered an intrinsically disordered domain (IDD). The LLPS properties of purified hnRNP A1 and other members of the hnRNP protein family were characterized in detail in vitro,[5,17] where phase separation can lead to the formation of LDs, hydrogels, and insoluble aggregates.[5,14] The IDD is necessary and sufficient to induce LLPS, and recently the importance of the regular spacing of aromatic residues was reported.[18] In contrast, UP1 alone does not phase separate. However, structural characterization in the dispersed and phase separated state revealed the importance for LLPS of largely electrostatic interactions involving sites both in the IDD and the RRMs.[19] We here focus on interactions that drive LLPS in solutions of hnRNP A1 in the presence of RNA. Understanding such mixed protein/oligonucleotide phase-separated solutions is receiving increased attention. For example it has been reported that RNA can modify LD morphology and viscosity,[6,20–22] but little is known at the molecular level. We here study the interplay of RNA binding and phase separation of hnRNP A1, which is potentially relevant for regulation of splicing. To this end, we use specific short single-stranded RNAs derived from human intronic splicing silencer N1 from SMN2 gene transcript (SMN2-ISS-N1), a natural target sequence of hnRNP A1 involved in survival of motor neuron exon 7 splicing.[7,23] RNA binding by UP1 has been studied in detail, and high-resolution structures are available.[23–25] Here, we address RNA binding to full-length hnRNP A1 and the ensuing RNA-induced LLPS by complementary in vitro biophysical

[*] I. Ritsch,[+] M. Yulikov, G. Jeschke
Laboratory of Physical Chemistry, Department of Chemistry and Applied Bioscience, ETH Zurich
Vladimir-Prelog-Weg 2, 8093 Zürich (Switzerland)
E-mail: gjeschke@ethz.ch
E. Lehmann,[+] L. Emmanouilidis, F. Allain
Institute of Biochemistry, Department of Biology, ETH Zurich
Hönggerbergring 64, 8093 Zürich (Switzerland)
[+] These authors contributed equally to this work.

[**] A previous version of this manuscript has been deposited on a preprint server (https://doi.org/10.1101/2022.03.21.485092).

methods. In particular, we use site-directed spin labelling (SDSL) in combination with NMR spectroscopy, as well as pulsed and continuous wave (CW) electron paramagnetic resonance (EPR) spectroscopy. We localize the IDD in the dispersed state in the absence of RNA by paramagnetic relaxation enhancement (PRE) measurements, double electron-electron resonance (DEER) spectroscopy, and small-angle X-ray scattering (SAXS). The resulting solution state model serves as a basis for interpreting RNA sequence dependence of LLPS upon addition of SMN2-ISS-N1-derived RNAs to dispersed hnRNP A1.

Results and Discussion

For our study of RNA-induced LLPS we recombinantly over-expressed and purified protein constructs from *E. coli.* cultures using a slightly modified version of the affinity-tag purification protocol with a cleavable His-tag reported previously (for details see Supporting Information).[26] To understand the role of the IDD in RNA-binding, we compared full-length hnRNP A1 (most abundant isoform P09651-2, 320 residues) to "RRM-only" UP1 (residues 1–196), see schematic representation in Figure 1A. First, we characterized the structure of full-length hnRNP A1 in the absence of RNA in the dispersed state (< 100 μM hnRNP A1 in 50 mM sodium phosphate, pH 6.5, 100 mM L-arginine, 100 mM L-glutamate) by applying a recently developed spin labelling approach for modelling of IDDs.[27] We found that spin labelling of the two partially buried native Cys in the RRMs with commonly available nitroxide spin labels led to rapid, irreversible protein aggregation, most likely due to disrupting the native fold. Therefore, we designed variants with both native Cys mutated to non-reactive residues (C43S, C175A) as background in all SDSL experiments. Following established design criteria, we engineered solvent-accessible spin labelling sites that are remote from known RNA binding sites. These labelling sites in structurally well characterized UP1 serve as spatially well separated reference points (beacons) in the folded domains: 52 in the β2-β3 loop of RRM1, 144 in the equivalent β2-β3 loop of RRM2 (β7-β8 in full-length hnRNP A1), and 182 in the C-terminal

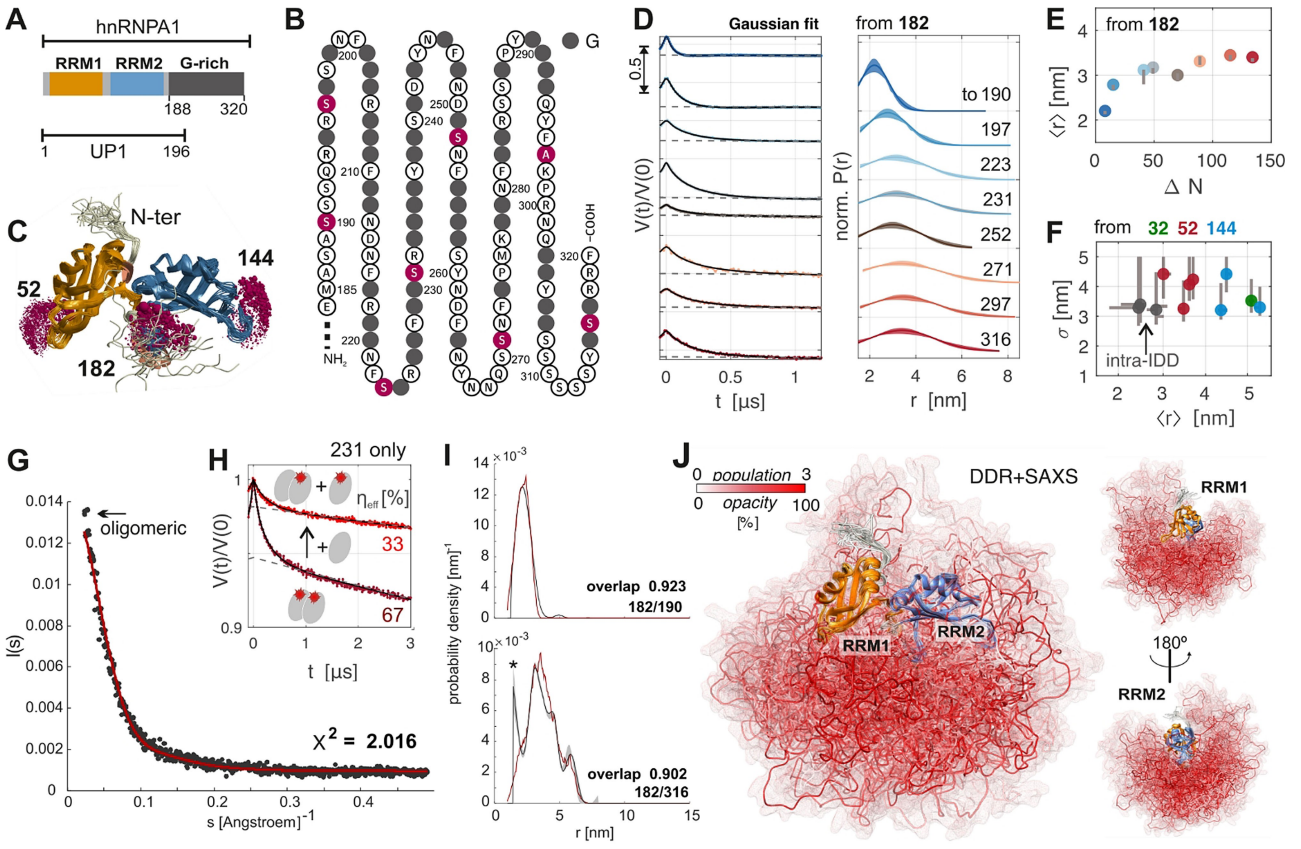

Figure 1. Ensemble model of hnRNP A1. A) Domain organization of hnRNP A1 and UP1; B) IDD (197 to 320), grey filled circles represent glycine, spin labelling sites are marked purple; C) solution structure of UP1 [pdb: 2LYV] with spin-label rotamer simulation (purple spheres) at the major beacon sites; D) primary data (left, arrow indicates modulation depth) and single Gaussian fits (right) of dipolar distance restraints measured from beacon 182 to sites in the IDD; E) mean distance $\langle r \rangle$ as a function of primary sequence separation ΔN of fits in (D); F) width σ vs. $\langle r \rangle$ of remaining distance restraints, error bars from validation with DeerLab (https://jeschkelab.github.io/DeerLab); G) ensemble SAXS curve and CRYSOL fit with DDR + SAXS model of hnRNP A1, the first data points were omitted in fitting to exclude contribution of oligomers; H) primary DEER data (red), background fit (dashed grey), and fit (black) for hnRNP A1 spin labelled at site 231 before and after mixing (1 : 1) with unlabelled protein (experimentally determined total effective labelling efficiency $\eta = c_{spin}/c_{protein}$ indicated); I) exemplary DDR fulfillment of the DDR + SAXS model; J) conformer population-weighted visualization of an ensemble model obtained with DDR + SAXS restraints.

helix of RRM2 (see rotamer simulations in Figure 1C). We also selected a total of eight "reporter" sites for sampling the IDD (190, 197, 223, 231, 252, 271, 297, and 316; all in this series except A297 are serines). This selection aimed at sampling approximately uniform sequence intervals (see Figure 1B). All mutants could be labeled well at approximately 25 μM hnRNP A1 and ten-fold molar excess of methanethiosulfonate spin label (MTSL). Doubly spin labelled mutants were used for distance distribution measurements by DEER spectroscopy. These measurements either involved one beacon in an RRM and one reporter in the IDD ("inter-domain restraints"), or both reporters within the IDD ("intra-domain restraints"). A selection of primary data and fits are shown in Figure 1D. For full experimental data (19 distance distributions) and fits see Figure S1 and S2, and Table S1. As quality control for the beacon sites we also measured distance distributions between pairs of beacons, which showed distributions that match the expectation from the known structural model of the RRMs (Figure S3). The shape of most distance distributions involving at least one site in the IDD was approximated well by a two-parameter single-Gaussian distance distribution (mean distance $\langle r \rangle$ and distribution width σ, detailed data analysis protocol can be found in the Supporting Information). Our previous methodological study revealed that deviations from a Gaussian distribution are insignificant compared to other uncertainties in ensemble modelling.[28] Hence, we can discuss the general trends in terms of $\langle r \rangle$ and σ. All distance distributions are very broad, confirming the expected large flexibility of the IDD. Based on this distance information we tested whether the IDD behaves like a random polymer chain. We focused our analysis on beacon 182, which is located close to the end of RRM2 (residue 184) and the beginning of the IDD (here defined as residue 188). All DEER measurements between beacon 182 and sites in the IDD thus give distance information on protein segments that are expected to be predominantly disordered. This allows us to analyze the distance distributions in terms of the primary sequence separation ΔN between the IDD reporter site and beacon 182. For a true random polymer coil we expect that the effective end-to-end length (and thus the fitted $\langle r \rangle$ and σ) of segments up to each reporter site increases monotonously with ΔN. As seen in Figure 1E, the increase of $\langle r \rangle$ conforms to this expectation only for $\Delta N < 50$. For distance distributions with $\Delta N > 89$ $\langle r \rangle$ and σ are approximately constant, and unexpectedly the distance with $\Delta N = 70$ (182/252) is slightly shorter than with $\Delta N = 49$ (182/231). In combination the observation contradicts the pure random coil model, but the deviation is small given the large σ values. Weak structural preferences could also be identified by correlating the fitted $\langle r \rangle$ and σ values for the remaining distance measurements (Figure 1F). Again the fitted large σ values confirm high overall structural flexibility, but we found that, on average, the IDD reporter sites have shorter mean distances to the loop of RRM1 (represented by beacon 52) than to the equivalent loop in RRM2 (beacon 144), which suggest a slight preference for an asymmetric placement with respect to the RRMs. One distance to an auxiliary beacon was measured as well (32/231), which identifies residue 32 as a more remote site in RRM1. In addition, we analyzed measurements of intra-domain segments 231/271, 231/316, and 271/316 in the same manner (grey in Figure 1F). Despite variation of ΔN between 40 and 85 in these pairs, all fitted mean distances are similar, and compared to expectations, relatively short and narrow. This indicates that individual IDD conformers preferentially sample conformations that are more compact than a well solvated random coil.

For a combined and more detailed analysis, we used the distance distribution restraints (DDRs) for construction of an atomistic structural ensemble model of full length hnRNP A1 using a previously reported force-field free algorithm.[27] The individual conformers are built by in silico growing the domain from a chosen anchor residue at the C-terminal end of the solution model of UP1 (pdb 2LYV).[26] As the C-terminal section of UP1 (188–196) is only weakly experimentally restrained and strongly diverging, we constructed conformers for residues 188–320. A large stochastic raw ensemble of individual conformers was generated using the 19 DDRs as Gaussian restraints, and taking into account the primary sequence by sampling from residue-specific Ramachandran-statistics. This ensemble was refined and contracted using the EnsembleFit module of MMMx (https://github.com/gjeschke/MMMx), additionally taking into account an experimental SAXS curve as shown in Figure 1G (for experimental details see Supporting Information). This step fits conformer populations and discards lowly populated conformers, while aiming for a balanced fit to the DDRs and SAXS curve. For SAXS fitting with CRYSOL,[29] the very short angle ($s < 0.02195$ Å^{-1}) region was disregarded, since it is highly sensitive to distortions from weak contributions of aggregates or transient oligomers (marked "oligomeric" in Figure 1G). For population fitting, we specified DDRs by model-free distance distributions obtained by Tikhonov regularization. Including jack-knife resampling and a second round of population re-weighting,[28] we arrived at the final structural ensemble of full length hnRNP A1 consisting of 138 conformers visualized in Figure 1J. The opacity of the conformers in the visualization reflects their population. Figure 1I shows by two examples that the input DDRs are well reproduced. The model confirms that the IDD is compacted around the RRMs, but is nonetheless able to populate diverse molecular conformations. It occupies space close to the linker region between the RRMs, distant from the N-terminal residues and close to the RNA-binding interface. The unexpectedly short distances (given the considerable length of the IDD) measured between the RRMs and the C-terminal region close to reporter site 316 can only be fulfilled by the IDD looping back towards the RRMs. We checked that an ensemble based on only Ramachandran statistics and the SAXS curve does not significantly sample such compact conformations (see Figure S4). This may explain why a previous study by SAXS and coarse grained molecular dynamics-based simulations approach[19] did not reveal the localization of the IDD that we find here. However, the broad distribution of conformers in our model suggests that structural preferences are weak, which implies an only slightly rugged energy hypersurface,

and is in agreement with the transient interaction between the RRMs and the flexible IDD found in the previous study.[19]

As mentioned above, the SAXS curve indicates presence of a minor fraction of oligomers. We used the DEER experiment to independently monitor their presence by analyzing the "background" inter-spin distance contribution.[30] In Figure 1H we show DEER data obtained for hnRNP A1 singly spin-labelled at site 231. As expected, the bulk of the data is fitted well by a slow decay (dashed grey line), which accounts for dispersed, homogeneously distributed protein at the low micro-molar concentration used in our experiments. Additionally, and in agreement with SAXS, we observed a fast-decaying component that confirms that a small fraction of protein molecules is in close contact. The amplitude of a non-background contribution in DEER is typically referred to as the modulation depth Δ, and the observed value of <0.05 indicates that $<10\,\%$ of hnRNP A1 molecules contribute to this fraction (approximating that fully labelled, fully dimerized molecules would result in $\Delta = 0.5$[30]). Additional experiments with more labelling sites, and distance analysis of the intermolecular DEER experiments can be found in Figure S6. Interestingly, when admixing unlabeled protein, we observed substantial suppression of the steep contribution (Figure 1H). Hence, exchange between oligomers and the admixed unlabeled protein must have occurred, implying that oligomerization is at least partially reversible. The singly spin-labelled samples also allowed to validate our results by paramagnetic relaxation enhancement (PRE) NMR experiments. In Figure 2A we show normalized PRE intensity ratios obtained from hnRNP A1 that is both spin-labelled and ^{15}N-isotope-labelled (NMR active/EPR active). MTSL was placed at one of the three sites 231, 271, or 316 (for experimental and analysis details see Supporting Information). In these all-interaction experiments, we cannot distinguish *a priori* between PRE arising from inter- and intramolecular proximity of the label to a given residue. However, intermolecular-only control experiments confirmed that intermolecular PRE is weak (Figure S7), in agreement with the SAXS and DEER results. We can thus interpret the PRE as predominantly intramolecular contributions. By mapping the PRE

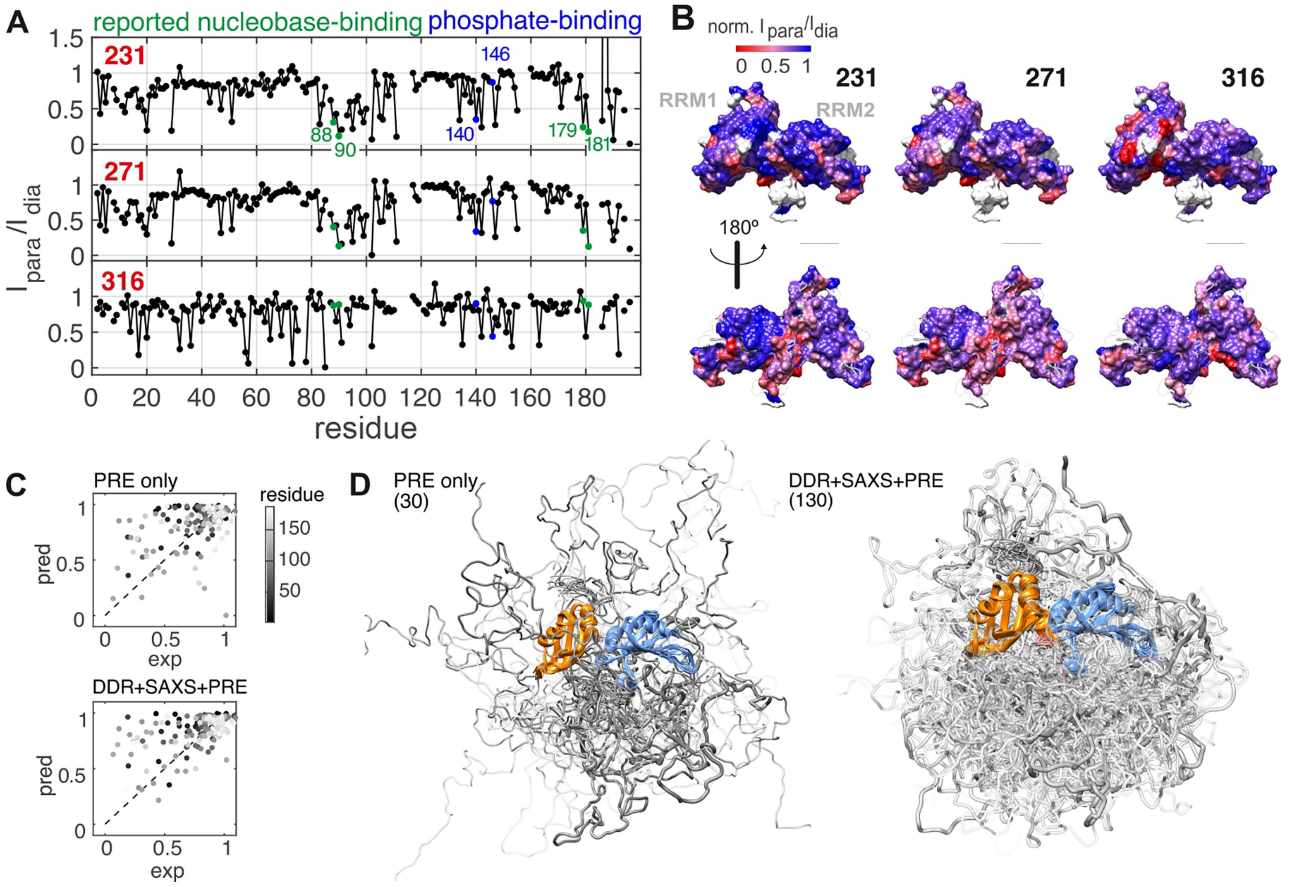

Figure 2. Validation of the ensemble model and determination of RRM interaction sites by PRE A) relative cross-peak volume ratio I_{para}/I_{dia} in ^{1}H-^{15}N-HSQC obtained with 50 μM hnRNP A1 singly spin labelled at sites 231, 271, or 316; data normalized to ratio at (buried) residue (117); residues that can coordinate RNA bases (green) or RNA backbone (blue) via interaction with the protein backbone are highlighted following Beusch et al.[23] B) mapping of PRE to the surface of UP1; C) experimental PRE vs PRE predictions for spin label at site 231 using a previously unrefined model ("PRE only"), or the fully refined ensemble model ("DDR + SAXS + PRE"); D) ensemble models used for PRE prediction (number of conformers after population reweighting is indicated); RRM1 (orange) and RRM2 (blue) shown as ribbon; IDD shown as worm model where transparency and width reflect fitted conformer population.

ratios to the solution-state model of UP1 (Figure 2B), we found moderate to strong signal reduction for a large fraction of the RRM surface. This is in agreement with the IDD chain sampling many remote surface interaction sites in our ensemble model. A similar, slightly more localized PRE pattern was observed with the spin labels at sites 231 and 271, whereas a non-specific pattern was observed with the label at site 316. Using a recently published PRE-prediction algorithm,[31] we find reasonable agreement of our "DDR + SAXS" ensemble model with the experimental PREs (r.m.s.d.$_{PRE}$ = 0.0822, Figure S4 and S5, Table S2). To improve our model, we included the PRE data at the ensemble refinement stage. Thereby, PRE fit quality could be improved (r.m.s.d.$_{PRE}$ = 0.0446) while simultaneously maintaining high fit quality for DEER and SAXS data (Figure 2C,D). We attribute the still mediocre PRE fit quality to the influence of backbone dynamics on PRE that we neglected, and potentially to slight inconsistency between ambient temperature PRE data and DEER data measured with flash-frozen samples. Interestingly, an ensemble fitted to only PRE restraints from a raw ensemble considering only residue-specific Ramachandran statistics resulted in a slightly worse PRE fit (r.m.s.d.$_{PRE}$ = 0.0581). Comparison of this "PRE-only" ensemble with the "DDR + SAXS + PRE" ensemble (Figure 2D) reveals that using the same number of starting conformers before population refitting, the DDR-based conformational sampling found a significantly larger number of conformers (total 130, approximately uniformly populated) that are consistent with the PRE data than the "PRE only" approach, which only found few significant chains (total 30, strong differences in population). Additionally, the "DDR + SAXS + PRE" conformers are simultaneously consistent with the DEER and SAXS datasets, which is not the case for the "PRE only" ensemble (Figure S5).

Looking more in depth at the PRE sites with the spin label at site 231, we observed clusters of residues in the N- and C-terminal flanking regions of RRM1, as well as in the β2-β3 loop and C-terminal helix of RRM2. Both regions constitute major sites for RNA-recognition. Strikingly, significant PRE effects were observed for the backbone nitrogen peaks of R88 and V90 in RRM, and R179 and R181 in RRM2. The same residues have been reported previously to directly recognize the "AG" motifs in SMN2-ISS-N1-related RNA variants by interaction via the protein backbone.[23] In that model proposed earlier by Beusch et al.,[23] RNA with the wild-type SMN2-ISS-N1 sequence (Figure 3A) can bind to UP1 in a 1:1 ratio, with RRM1 binding a 3′-AG motif, and RRM2 binding a 5′-AG motif. In this complex, the RNA wraps around both RRMs and the linker region, covering a large fraction of the positively charged surface of UP1 (see Figure 3B). Thus sites that form the RNA-binding interface coincide with sites that interact transiently with the IDD around site 231 and 271. We were intrigued by this overlap, because it implies that RNA binding requires at least partial displacement of the IDD from the RRM surface. With our ensemble model in hand, we investigated such a potential role of the IDD in RNA binding. In addition to SMN2-ISS-N1 based "wild type" (wt) RNA (also denoted as "RNA*gg*"), we used three additional RNA variants where either the 5′ AG motif ("RNA*ag*") or the 3′ AG motif ("RNA*ga*"), or both ("RNA*aa*") have been mutated. This choice is motivated by the earlier finding that already a single point mutation of these two core 5′-AG-′3 binding motifs (AG→AA) can strongly affect splicing repression.[23] Note that in this purine-to-purine point mutation, adenine as the substituting nucleobase has similar overall size and aromatic properties as the wild-type guanine, yet guanine has a higher hydrogen bonding capacity.

RNA binding to UP1 in our buffer conditions was confirmed by non-denaturing electrophoretic mobility shift assays (EMSA, see Figure 3C). With RNA-staining by toluidine, we observed a resolved band ("1") with reduced mobility compared to free RNA ("f") upon 1:1 stoichiometric incubation of UP1 with either wild-type RNA. The same band was observed with RNA*aa*, showing that even the doubly mutated RNA can still bind to the RRMs. Because significant amounts of free RNA were observed, both at stoichiometric ratio and at RNA excess, we concluded that the overall binding affinity is rather low with both wt and mutated RNA. A faint second resolved band was observed with wt RNA ("2"), which is likely related to an RNA/protein complex of different stoichiometry. The absence of this second band with mutated RNA might be related to different dynamics of binding and unbinding, which could lead to broadening of the band below the detection limit. In contrast to the mobile bands observed with UP1, EMSA with the same RNAs and hnRNP A1 led to the accumulation of immobile material in the gel pockets ("i")). A very faint band of a slowly migrating complex is visible with RNA*aa* ("1a"), but not with wt RNA. As is the case with UP1, significant amounts of free RNA are also observed with hnRNP A1. Thus, while some RNA is trapped in the aggregates, a substantial amount remains unbound.

In order to characterize the immobile component in EMSA, we used confocal microscopy and optical turbidity measurements at 600 nm, which are well established methods to monitor aggregation and LLPS.[5] Samples were stabilized by low melting agarose gel (1 % w/v), which prevents LD sedimentation and allowed stable observation of the samples for several hours. We observed LLPS upon adding RNA to 25 μM hnRNP A1 (shown at 1:1 ratio) as the formation of many small LDs (≈1 μm) seen by confocal microscopy (Figure 3F). In addition to LDs, recognized by their spherical shape and occasional fusion events, we sometimes observed non-spherical particles and insoluble aggregates (see Figure S8). All four RNA variants were able to induce LLPS with hnRNP A1 and none of them with UP1 (Figure S9). The size and appearance of the agarose-stabilized LDs depended on absolute protein concentration and RNA-to-protein ratio. Larger LDs (up to 10 μm) were formed upon addition of wt RNA or the doubly mutated RNA*aa* at a 0.5:1 (RNA:protein) molar ratio to 100 μM hnRNP A1, as seen by fluorescence microscopy when incubating with GelRed® RNA-stain (Biotium, USA) fluorescent dye (Figure 3G). The dye appears to be distributed homogeneously within each droplet, as can be seen in side-

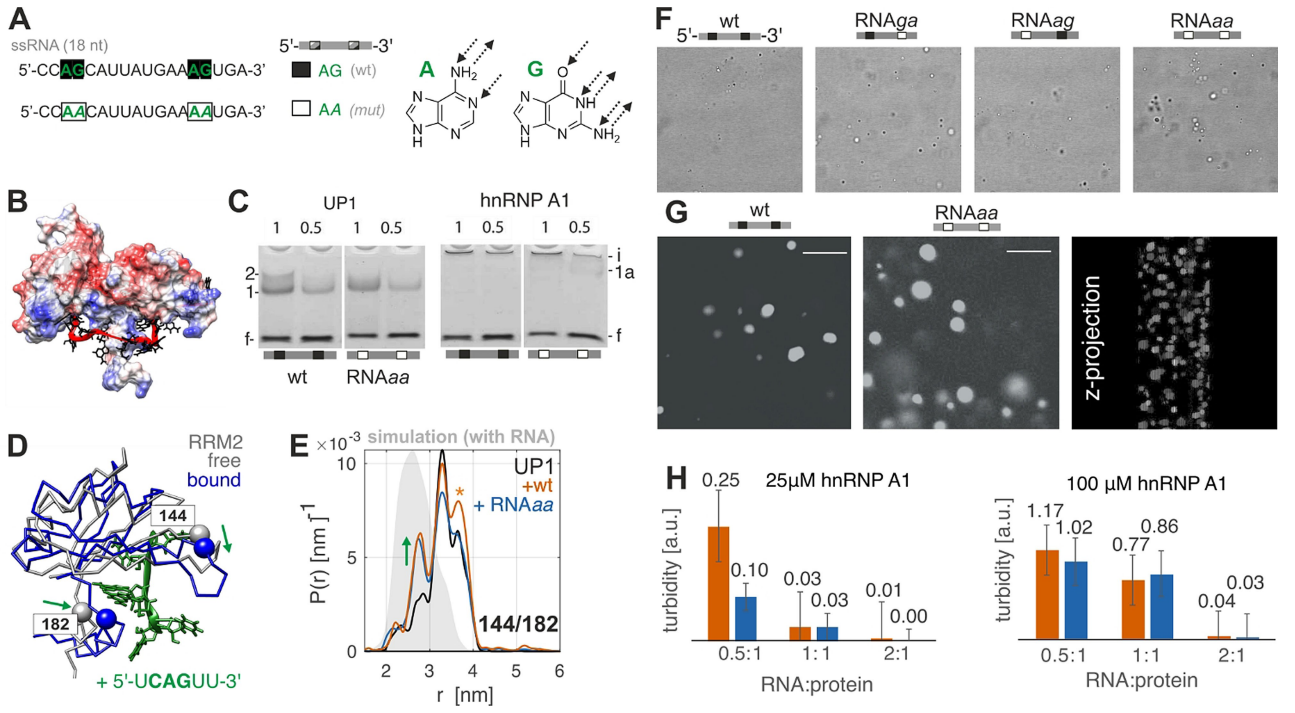

Figure 3. Effects of RNA binding to hnRNP A1 A) SMN2-ISS-N1 wt RNA sequence, and RNA variant with two G→A point mutations in the core AG motifs ("RNA*aa*"); B) Coulombic colored model of RNA-bound UP1 from Beusch et al.;[23] negatively charged RNA backbone shown in red C) toluidine-stained EMSA with wt RNA or RNA*aa* incubated with UP1 or hnRNP A1 at two stoichiometric ratios (protein:RNA); RNA concentration was fixed at 50 µM; distinct complex bands (1, 2 with UP1 and 1a with hnRNP A1), free RNA (f), and immobile (i) material in the gel pocket are marked; D) expected RRM2 conformational change upon RNA binding; overlay of RRM2 in free UP1 (grey, PDB: 2LYV) and RNA-bound RRM2 (blue, PDB: 5MPL) E) distance distribution between beacons 144/182 of free UP1 (black), UP1 + 1 : 1 wt RNA (orange), and UP1 + 1 : 1 RNA*aa* (blue) overlayed on the MMM simulation with free UP1 (grey area); F) confocal transmission microscopy images of 25 µM hnRNP A1 after mixing with RNA variants (1 : 1 ratio); RNAga and RNAag are variants with only one of the two AG motifs mutated (scale bar 10 µm); G) *left, center:* fluorescence microscopy images of 25 µM hnRNP A1 with 0.5 : 1 (RNA:protein) stained with GelRed® RNA stain. *right:* projection of fluorescent image stack in third dimension obtained with RNA*aa* H) turbidity (by OD_{600}) of hnRNP A1 after mixing with wt RNA (orange) or RNA*aa* (blue); protein concentration was fixed at low or high concentration as indicated.

view images from reconstructing focal plane stack, suggesting a homogeneous distribution of RNA in the droplets. The turbidities for the conditions used for imaging as well as for additional RNA:protein ratios are shown in Figure 3H. Surprisingly, at low hnRNP A1 concentration (25 µM), we observed an RNA-sequence dependent LLPS effect, where RNA-induced LLPS was weaker for the mutated variants than for the wt RNA (Figure 3H). Specificity was expected to occur via UP1, so to understand this finding, we screened for conformational changes upon binding by different RNA variants to UP1, where we know that no LLPS occurs. We selected the pair of beacons 144 and 182 in RRM2 for this test, because comparison of the conformations of RRM2 in the absence of RNA and in complex with short RNA[23] suggests that, upon binding the C-terminal helix moves slightly towards the RRM2 β2-β3 loop region (Figure 3D). Indeed, we observe a partial shift to shorter distances for a similar fraction of the distribution with both wt and doubly mutated RNA. The shift is in line with simulations with the NMR structures of free UP1 and of the RNA-RRM2 complex (Figure 3E and S9), but already the modelling of the free state is not in perfect agreement (Figures S3), which limits our interpretation of this weak effect. In the case of wt RNA, we observed a slightly enhanced dipolar modulation amplitude which we attribute to RNA-enhanced protein-protein contacts, in agreement with EMSA (band "2" in Figure 3C). The upper distance shoulder is also slightly increased with wt RNA (asterisk in Figure 3E), which is likely to be an artefact of the enhanced spin interactions due to multi-spin effects.[32] Overall, the DEER binding analysis did not provide a clear explanation of RNA-sequence dependent LLPS at low protein concentration. Unfortunately, the protein and thus spin concentration were too high for interpretation of analogous DEER experiments in the LDs that are induced by RNA binding to full length hnRNP A1 (Figure S10). Instead, we turned to ambient-temperature CW EPR spectroscopy to investigate changes in dynamics of the IDD upon RNA binding and LLPS. We expected that local molecular crowding would lead to reduced backbone and spin label side chain mobility. Such reduced mobility is observable by line broadening in CW EPR spectra, which reports on dynamics in the high pico- to low nanosecond range. The results of this dynamics study for reporters 231, 271 and 316 in the IDD are presented in Figure 4. Additional results with labelling sites in the RRMs can be found in Figure S11. We first characterized the

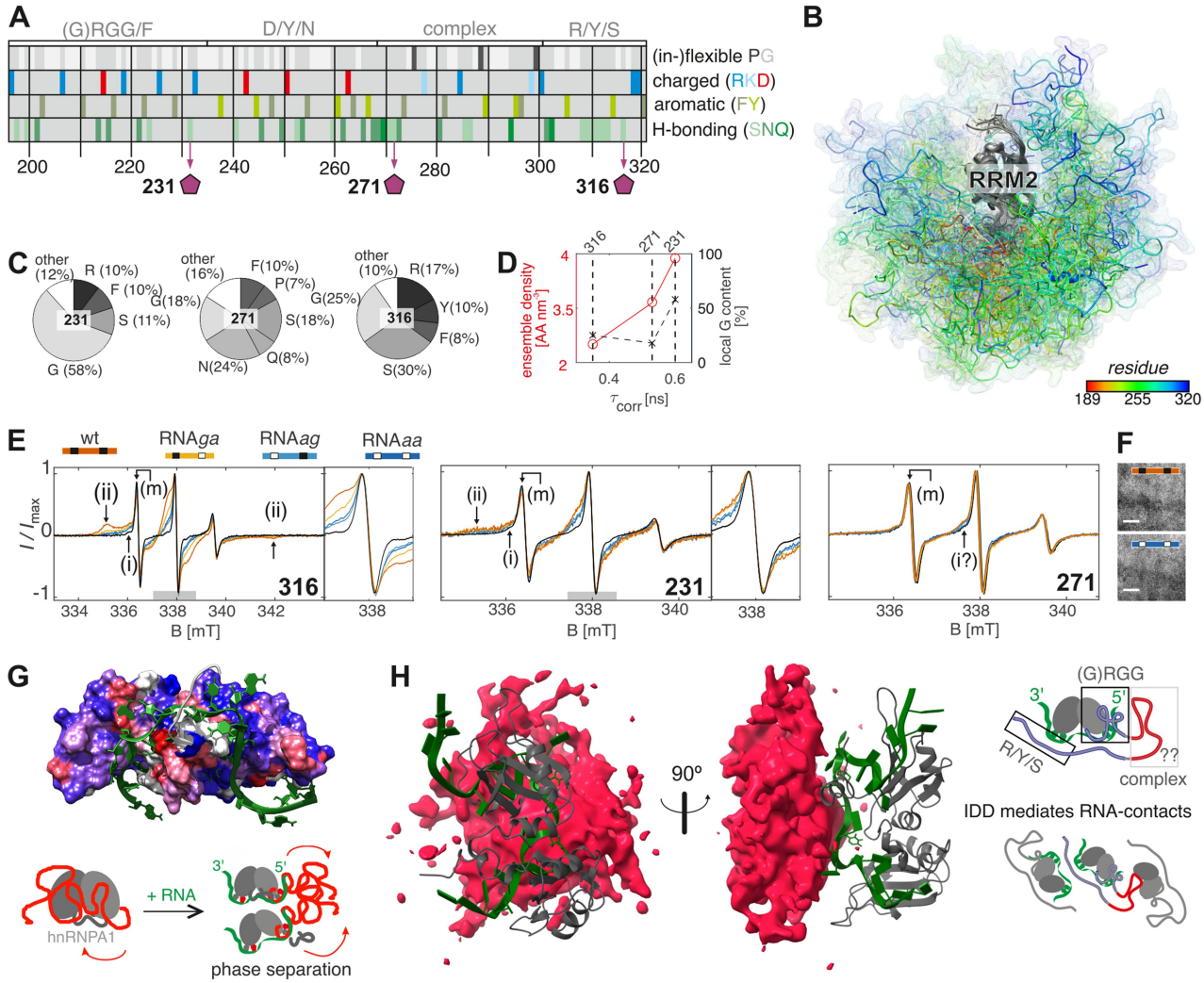

Figure 4. Relation between RNA binding and LLPS. A) sequence breakdown of G-rich IDD of hnRNP A1; B) side view looking at RRM2 of full length hnRNP A1 ensemble model "DDR + SAXS + PRE"; IDD chains are color-coded by residue C) local amino acid type statistics in radius of 1 nm around reporter sites in model shown in (B); D) fitted τ_{corr} of nitroxide spin label at reporter sites 231, 271 and 316 spectra vs. local amino acid composition ("o"), or vs. local ensemble density ("x"); E) overview and zoom on central line of CW EPR of 25 μM hnRNP A1 reporter sites, normalized to maximum intensity; black: free protein (no RNA); color-coded: +1:1 molar ratio of RNA variants F) confocal image of mutant labelled at site 271 with 1:1 RNA, scale bar 10 μm; G) RNA mapped on surface of RRMs color-coded by PRE from site 231 (stronger PRE shown as red, see Figure 2(B)) and scheme of proposed RNA vs IDD competition mechanism for interaction with the RRMs; H) RNA mapped on RRMs in ensemble model of hnRNP A1 with IDD (red) displayed as average occupied volume, schematic drawing of the proposed cooperation mechanism (recognition of RRM-bound RNA).

dynamics in the RNA-free, dispersed state (agarose-stabilized buffer). As expected, the three spectra with free hnRNP A1 (black lines in Figure 4E) show very fast spin label mobility. The lineshapes could be fitted well using the EasySpin[33] toolbox, assuming a single isotropically tumbling component with rotational correlation time τ_{corr} (see also Figure S12, and fitting procedure described in Supporting Information). Site 316, close to the C-terminus, has the shortest correlation time and thus highest mobility (τ_{corr} = 0.35 ns), followed by site 271 (τ_{corr} = 0.53 ns) and site 231 (τ_{corr} = 0.60 ns). We expect that intrinsic sidechain mobility is governed by an interplay of local density and compositional bias that affects backbone dynamics and thus investigated whether it correlates with local sequence context and average density at the label site. Such analysis in terms of local amino acid composition is related to the proximity to "stickers" (interacting) and "spacers" (stretches of non-interacting).[17] Selected amino acid categories are schematically represented in Figure 4A. Glycine and proline are highlighted in the top row and are expected to be key regulators of local high or low backbone flexibility, respectively. Below, we indicate different types of "stickers": in the second row charged residues; in the third row aromatic residues; and in the last row hydrogen-bond donor and acceptor amino acids. With our ensemble model of hnRNP A1 at hand, we included spatial aspects into local compositional analysis. To this end we calculated the average amino acid density and composition in a one nanometer spherical

zone around the reporters from the model shown in Figure 4B (DDR + SAXS + PRE ensemble model, for computation details see Supporting Information). The resulting amino acid (AA) statistics are shown in Figure 4C for amino acids more abundant than 5 %. The abundance of glycine, as well as the local average ensemble density are plotted against the reporter mobility in Figure 4D. Site 231 with the lowest sidechain mobility somewhat unexpectedly is located in a region with high local excess of flexible over inflexible residues (58 % G, 0 % P), but in contrast is located in the on average densest region of the IDD (3.78 AA nm^{-3}). Similarly, site 316 with the lowest density (3.03 AA nm^{-3}, 25 % G, 0 % P) has the highest experimental mobility. Site 271 falls in the intermediate mobility and density range, and is located at a site of comparatively higher sequence complexity, in particular close to proline (3.29 AA nm^{-3}, 18 % G, 7 % P). On this small dataset the average ensemble density thus appears to be suitable for predicting local dynamics, and is somewhat more robust than individual side chain statistics.

Having established spin label dynamics at the reporter sites for the free state, we studied changes upon mixing 25 μM spin-labelled hnRNP A1 with wt RNA and the RNA variants. From turbidity measurements and microscopy, we know that mixing induces LLPS, and we observed strongly broadened spectral components at site 231, indicating strongly reduced label mobility.

Co-existence of multiple resolved spectral components implies that different populations of protein (e.g. dispersed vs RNA-bound and/or phase separated) co-exist in the sample, which interchange on a time-scale slower than the EPR experimental resolution of about 1 μs. In our case, dispersed protein with high sidechain mobility appears to coexist with phase-separated protein with lower mobility. The ratio of the immobile to the mobile component initially increases with the ratio of added RNA:protein, but plateaus after a ratio 1:1 (see Figure S13). We annotate the rapid rotational tumbling also observed in the free state as the mobile (m) component, moderate spectral broadening as the intermediate (i) component, and the contribution where the low-field minimum of the spectrum can clearly be identified as the immobile (ii) component.

Broadened spectral components were observed with all four RNA variants for sites 231 and 316, but not for site 271. The latter appears to be a special case, since placing the spin label at this site substantially suppressed RNA-induced LLPS, as observed by the absence of LDs (Figure 4F), and by reduced turbidity compared to wild-type hnRNP A1 (0.036 ± 0.059 with wt RNA, and 0.004 ± 0.004 with RNA*aa*, at 0.5:1 RNA: protein ratio (for wild-type see Figure 3H). Note that the RRM domains in all three constructs are the same, therefore the observed perturbation must arise from positioning the spin label at different sites in the IDD. The LLPS suppression by perturbing site 271 is unexpected, since early work[5] showed that LLPS of RNA-free hnRNP A1 did not require the presence of a "steric-zipper" motif, located just N-terminal of site 271. In contrast, our study suggests that, in the context of RNA-induced LLPS, this region of the IDD is critical for promoting phase separation. Site 271 is located in a region with low G content and does not contain RGG motifs known to be involved in RNA binding, yet is enriched in amino acids with high H-bonding capacity (50 % = N/S/Q).

The CW-EPR spectra with labels at sites 231 and 316 agree with the previously discussed RNA-sequence dependent LLPS effect at low hnRNP A1 concentration. At site 231, the 5′ AG motif (present in RNA*ga*, and wt, but not in RNA*ag* and RNA*aa*) was found to be necessary and sufficient to observe an immobile component (ii). This implies that LLPS in our set of RNA variants is governed not only by the number of G nucleotides or AG dinucleotide motifs; it appears to be sequence-specific to some extent. With RNA*ag* and RNA*aa*, an immobile component (ii) was not observed and the contribution of the mobile component (m) remained higher. That RNA-binding affects mobility at site 231 might be expected even in the absence of LLPS, since it is located close to an arginine residue and just C-terminal of several (G)RGG motifs. The mechanism by which RGG is reported to interact with RNA[34] is unspecific to RNA sequence, with the exception of being able to recognize structural motifs, such as G-quadruplexes.[35,36]

An RNA-sequence specific LLPS effect is even more strikingly reported by the label at site 316, close to the C-terminus in an S (30 %)/Y (10 %)/R (17 %)-rich region. There are no RGG motifs in proximity to 316, yet the RNA sequence dependence is even stronger than at site 231. An immobile component (ii) can be clearly discerned in the presence of wt RNA and RNA*ga*, in addition to the broadened component (i), which is observed with all RNAs. As the immobile component is well resolved at site 316, mobility at this site must change by at least an order of magnitude (τ_{corr}(wt RNA) > 10 ns) between the free state and the state interacting with wt RNA (for additional spectral fitting see Supporting Information). Furthermore, the mobile component (m) is almost fully suppressed at site 316 with wt RNA.

For explanation of the unexpected RNA sequence-dependent LLPS and mobility changes, we consider the CW EPR observations in the context of the PRE measurements and of our ensemble model. For RNA-free, dispersed hnRNP A1, the IDD is mostly proximal to RNA binding sites in the RRMs (e.g. V90). This is illustrated by mapping the RNA from the 1:1 model by Beusch et al.[23] onto UP1 color-coded by PRE intensity for spin labelling site 231. The sites of strong PRE on the RRMs overlap with sites that recognize nucleobases in the RNA (Figure 4G). The aromatic and H-bonding residues in the IDD may thus act as intramolecular competitors for RNA binding. Our ensemble model implies displacement of the IDD from the RRMs upon RNA binding. This would in turn expose the sticky IDD to an aqueous environment, triggering enhanced protein-protein interactions and thus LLPS due to unsatisfied binding preferences of this domain. The mutated RNA variants are expected to be less efficient at displacing the IDD from the RRM surface than wt RNA. In addition to intermolecular contacts after IDD displacement, the IDD is also positioned close to the RRM-bound RNA, possibly facilitating additional direct interactions, which we speculate

might increase specificity in the binding (Figure 4H). This may explain the strong effect at site 316, which is surrounded by a large fraction of amino acids with hydrogen bonding capacity.

Selectivity due to cooperative binding of the RRMs and the IDD to RNA could act as a possible regulatory mechanism. Weak differences in stability of binding of the RNA to the RRM can be amplified by additional interactions with the IDD, leading to tighter binding and stronger immobilization of target sequences, which reduces promiscuity of full hnRNP A1 compared to UP1. Note that this is consistent with observations that additional regulation of biomolecular condensate formation can be achieved by post-translational modification of sites in the IDD. Phosphorylation of serine, a major regulator of hnRNP A1 cellular localization,[4] introduces local negative charge in a previously uncharged region, and changes hydrogen bonding preferences, which might trigger the release of the bound RNA and dissolution of the condensed phase.

Conclusion

We report that specific RNA-binding by hnRNP A1 is tightly linked to LLPS. Already sub-stoichiometric addition of specific RNAs, which carry two, one or no AG dinucleotide motifs embedded in the natural SMN2-ISS-N1 target sequence, induces LLPS. While at high protein concentrations (100 μM) LDs appear with any of the tested RNA variants, RNA-sequence dependent LLPS was observed at lower protein concentrations (25 μM). DEER experiments on singly-labeled hnRNP A1 as well as SAXS and intermolecular PRE experiments revealed a minor fraction of oligomeric hnRNP A1 even at low protein concentration and in the absence of RNA. Both PRE data and an ensemble model of hnRNP A1 revealed a compact conformation of the IDD and its interaction with the RNA-binding face of the RRMs in the dispersed state. CW EPR experiments on singly spin-labelled mutants detected lower sidechain mobility in the presence of RNA. This effect is slightly more prominent with the wild-type RNA than with the mutated RNA and depends on the placement of the spin label in the IDD. Our findings demonstrate that single point mutations in the specifically recognized RNA significantly impact this process. We speculate that other proteins consisting of coupled RNA-binding and disordered domains may also exhibit such behavior, as it would provide some specificity to the complex interaction networks of phase-separating proteins that are promiscuous RNA binders.

Acknowledgements

We would like to thank Emil Dedic and Georg Dorn for help with the SAXS acquisition and data processing. SAXS was measured with an instrument that was kindly provided by the Mezzenga lab at ETH Zurich. Confocal microscopy images were acquired at the ScopeM facility at ETH Zurich. We also want to thank Irene Beusch, Fred Damberger, Laura Esteban-Hofer, Christoph Gmeiner, and Daniel Klose for useful discussions. We are grateful to Tanja Mittag and co-workers for discussions about the PRE experiments. Open Access funding provided by Eidgenössische Technische Hochschule Zürich.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Ensemble models obtained with DEER, SAXS, and PRE restraints (PED00212), with DEER and SAXS restraints (PED00213), with only DEER restraints (PED00214), and with only PRE restraints (PED00215) were deposited to the Protein Ensemble Database. Experimental data, MMMx restraint and ensemble analysis files (.mcx), restraint data, raw ensembles, and ensemble lists with populations (.ens) are available at Zenodo.[37]

Keywords: biophysics · EPR spectroscopy · intrinsically disordered domain · RNA · NMR spectroscopy

[1] S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 285–298.
[2] S. Boeynaems, S. Alberti, N. L. Fawzi, T. Mittag, M. Polymenidou, F. Rousseau, J. Schymkowitz, J. Shorter, B. Wolozin, L. V. D. Bosch, P. Tompa, M. Fuxreiter, *Trends Cell Biol.* **2018**, *28*, 420.
[3] A. A. Hyman, C. A. Weber, F. Jülicher, *Annu. Rev. Cell Dev. Biol.* **2014**, *30*, 39–58.
[4] E. Allemand, S. Guil, M. Myers, J. Moscat, J. F. Cáceres, A. R. Krainer, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3605–3610.
[5] A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag, J. P. Taylor, *Cell* **2015**, *163*, 123–133.
[6] Y. Lin, D. S. W. Protter, M. K. Rosen, R. Parker, *Mol. Cell* **2015**, *60*, 208–219.
[7] J. Jean-Philippe, S. Paz, M. Caputi, *Int. J. Mol. Sci.* **2013**, *14*, 18999–19024.
[8] H. L. Okunola, A. R. Krainer, *Mol. Cell. Biol.* **2009**, *29*, 5620–5631.
[9] H. Siomi, G. Dreyfuss, *J. Cell Biol.* **1995**, *129*, 551–560.
[10] B. J. Lee, A. E. Cansizoglu, K. E. Süel, T. H. Louis, Z. Zhang, Y. M. Chook, *Cell* **2006**, *126*, 543–558.
[11] M. Tolbert, C. E. Morgan, M. Pollum, C. E. Crespo-Hernández, M.-L. Li, G. Brewer, B. S. Tolbert, *J. Mol. Biol.* **2017**, *429*, 2841–2858.
[12] N. Jain, C. E. Morgan, B. D. Rife, M. Salemi, B. S. Tolbert, *J. Biol. Chem.* **2016**, *291*, 2331–2344.
[13] C. Yu, J. Guo, Y. Liu, J. Jia, R. Jia, M. Fan, *J. Cell. Physiol.* **2015**, *230*, 2252–2261.
[14] H. J. Kim, N. C. Kim, Y.-D. Wang, E. A. Scarborough, J. Moore, Z. Diaz, K. S. MacLea, B. Freibaum, S. Li, A. Molliex, A. P. Kanagaraj, R. Carter, K. B. Boylan, A. M. Wojtas, R. Rademakers, J. L. Pinkus, S. A. Greenberg, J. Q. Trojanowski, B. J. Traynor, B. N. Smith, S. Topp, A.-S. Gkazi, J. Miller, C. E. Shaw, M. Kottlors, J. Kirschner, A. Pestronk, Y. R. Li, A. F. Ford, A. D. Gitler, M. Benatar, O. D. King, V. E.

Kimonis, E. D. Ross, C. C. Weihl, J. Shorter, J. P. Taylor, *Nature* **2013**, *495*, 467–473.
[15] N. Jain, H.-C. Lin, C. E. Morgan, M. E. Harris, B. S. Tolbert, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2206–2211.
[16] A. Mayeda, S. H. Munroe, R. M. Xu, A. R. Krainer, *RNA* **1998**, *4*, 1111–1123.
[17] J. Wang, J.-M. Choi, A. S. Holehouse, H. O. Lee, X. Zhang, M. Jahnel, S. Maharana, R. Lemaitre, A. Pozniakovsky, D. Drechsel, I. Poser, R. V. Pappu, S. Alberti, A. A. Hyman, *Cell* **2018**, *174*, 688–699.
[18] E. W. Martin, A. S. Holehouse, I. Peran, M. Farag, J. J. Incicco, A. Bremer, C. R. Grace, A. Soranno, R. V. Pappu, T. Mittag, *Science* **2020**, *367*, 694–699.
[19] E. W. Martin, F. E. Thomasen, N. M. Milkovic, M. J. Cuneo, C. R. Grace, A. Nourse, K. Lindorff-Larsen, T. Mittag, *Nucleic Acids Res.* **2021**, *49*, 2931–2945.
[20] V. H. Ryan, S. Watters, J. Amaya, B. Khatiwada, V. Venditti, M. T. Naik, N. L. Fawzi, *Nucleic Acids Res.* **2020**, *48*, 10542–10554.
[21] S. Boeynaems, A. S. Holehouse, V. Weinhardt, D. Kovacs, J. V. Lindt, C. Larabell, L. V. D. Bosch, R. Das, P. S. Tompa, R. V. Pappu, A. D. Gitler, *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 7889–7898.
[22] M. Garcia-Jove Navarro, S. Kashida, R. Chouaib, S. Souquere, G. Pierron, D. Weil, Z. Gueroui, *Nat. Commun.* **2019**, *10*, 3230.
[23] I. Beusch, P. Barraud, A. Moursy, A. Cléry, F. H.-T. Allain, *eLife* **2017**, *6*, e25736.
[24] J. C. Myers, Y. Shamoo, *J. Mol. Biol.* **2004**, *342*, 743–756.
[25] H. Kooshapur, N. R. Choudhury, B. Simon, M. Mühlbauer, A. Jussupow, N. Fernandez, A. N. Jones, A. Dallmann, F. Gabel, C. Camilloni, G. Michlewski, J. F. Caceres, M. Sattler, *Nat. Commun.* **2018**, *9*, 2479.
[26] P. Barraud, F. H.-T. Allain, *J. Biomol. NMR* **2013**, *55*, 119–138.
[27] G. Jeschke, *Proteins Struct. Funct. Bioinf.* **2016**, *84*, 544–560.
[28] I. Ritsch, L. Esteban-Hofer, E. Lehmann, L. Emmanouilidis, M. Yulikov, F. H.-T. Allain, G. Jeschke, *Front. Mol. Biosci.* **2021**, *8*, 148.
[29] D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries, D. I. Svergun, *J. Appl. Crystallogr.* **2017**, *50*, 1212–1225.
[30] G. Jeschke, *Annu. Rev. Phys. Chem.* **2012**, *63*, 419–446.
[31] G. Tesei, J. M. Martins, M. B. A. Kunze, Y. Wang, R. Crehuet, K. Lindorff-Larsen, *PLoS Comput. Biol.* **2021**, *17*, e1008551.
[32] T. von Hagens, Y. Polyhach, M. Sajid, A. Godt, G. Jeschke, *Phys. Chem. Chem. Phys.* **2013**, *15*, 5854–5866.
[33] S. Stoll, A. Schweiger, *J. Magn. Reson.* **2006**, *178*, 42–55.
[34] P. A. Chong, R. M. Vernon, J. D. Forman-Kay, *J. Mol. Biol.* **2018**, *430*, 4650–4665.
[35] X. Liu, T. Ishizuka, H.-L. Bao, K. Wada, Y. Takeda, K. Iida, K. Nagasawa, D. Yang, Y. Xu, *J. Am. Chem. Soc.* **2017**, *139*, 7533–7539.
[36] M. Ghosh, M. Singh, *Nucleic Acids Res.* **2018**, *46*, 10246–10261.
[37] [dataset] I. Ritsch, E. Lehmann, L. Emmanouilidis, M. Yulikov, F. Allain, G. Jeschke, 2022, Phase separation of hnRNP A1 upon specific RNA-binding observed by magnetic resonance, Zenodo, Version 1.0.0, https://doi.org/10.5281/zenodo.6384003.

Manuscript received: March 26, 2022
Accepted manuscript online: July 22, 2022
Version of record online: August 31, 2022