Systems NMR: single-sample quantification of RNA, proteins and metabolites for biomolecular network analysis

Journal Article

Author(s):
Nikolaev, Yaroslav (iD); Ripin, Nina; Soste, Martin; Picotti, Paola; Iber, Dagmar; Allain, Frédéric H.-T.

Corrected: Publisher Correction

Systems NMR: single-sample quantification of RNA, proteins and metabolites for biomolecular network analysis

Yaroslav Nikolaev [1]*, Nina Ripin [1], Martin Soste[2], Paola Picotti[2], Dagmar Iber [3] and Frédéric H.-T. Allain [1]*

Cellular behavior is controlled by the interplay of diverse biomolecules. Most experimental methods, however, can only monitor a single molecule class or reaction type at a time. We developed an in vitro nuclear magnetic resonance spectroscopy (NMR) approach, which permitted dynamic quantification of an entire 'heterotypic' network—simultaneously monitoring three distinct molecule classes (metabolites, proteins and RNA) and all elementary reaction types (bimolecular interactions, catalysis, unimolecular changes). Focusing on an eight-reaction co-transcriptional RNA folding network, in a single sample we recorded over 35 time points with over 170 observables each, and accurately determined five core reaction constants in multiplex. This reconstruction revealed unexpected cross-talk between the different reactions. We further observed dynamic phase-separation in a system of five distinct RNA-binding domains in the course of the RNA transcription reaction. Our Systems NMR approach provides a deeper understanding of biological network dynamics by combining the dynamic resolution of biochemical assays and the multiplexing ability of 'omics'.

The regulation of cellular behavior is complex, and emerges from the dynamic interplay of diverse biomolecules, including proteins, RNA and metabolites. Most experimental methods, however, can only monitor a single molecule class or reaction type at a time (catalysis, bimolecular interactions and unimolecular state changes), limiting our ability to measure complex cellular dynamics.

Furthermore, studies of networks often face a choice between biochemical methods—measuring dynamic (for example, time-resolved) data only for a few network components, and 'omics' methods—measuring a large number of components, but usually with little dynamic information in a single sample[1]. This lack of dynamic data covering multiple network components is among the main limitations[2–4] in developing validated mechanistic, mathematical models for cellular networks, which are key to understanding the underlying logic of these networks[5].

To address the above challenges, we sought to devise a nuclear magnetic resonance- (NMR)-based approach that would (1) allow us to monitor 'heterotypic' networks and pathways—involving different molecule or reaction types—entirely, in a single in vitro sample, and (2) provide quantitative dynamic data for modeling of the network mechanisms. With certain limitations on molecule size and concentration ($\lesssim 50$–$100\,kDa$, $\gtrsim 10$–$50\,\mu M$)[6,7], solution NMR can monitor any reaction type or molecular class in a wide range of conditions, including unfractionated cell extracts and living cells[8]. The use of NMR to monitor reactions is common in chemistry[9], and in recent years, NMR has also been used to follow the dynamics of small-scale reaction networks in biology. However, those studies focused on individual molecule classes; that is, metabolites[10–14], proteins[15,16] or RNA[17–19].

We sought to monitor a more complex network by NMR, that comprises a wide range of different molecule and reaction types.

Co-transcriptional RNA folding is an important cellular process that simultaneously involves RNA, proteins and metabolites, and is still poorly understood. RNA molecules of the same sequence may form distinct folded structures, with distinct functions and fates, depending on the effectors present during RNA transcription[20,21]. Insights are still limited on how the final RNA structures are influenced by co-transcriptional interactions of the transcribing RNA. The core reactions of the underlying network are RNA synthesis from metabolites, RNA folding and protein–RNA interactions (Fig. 1a). Our aims were (1) to design an assay to monitor all main components of the network simultaneously by NMR spectroscopy, using specific signatures of different molecules in NMR spectra; (2) to establish a mathematical model explaining our observations and (3) to perturb the network with proteins and drug molecules to gain system-level insight into its dynamics. This assay revealed competitive weakening of specific hnRNP A1 protein–RNA interactions by unspecific nucleotide-bearing molecules, and exposed the dynamic phase-separation of proteins in the course of RNA transcription. We termed this methodological approach 'Systems NMR'—a potential generic name for NMR-driven reconstruction of biomolecular reaction networks.

Results

Monitoring a co-transcriptional RNA folding network. We first sought to investigate whether the RNA-binding protein UP1 (a natural fragment of hnRNP A1) would perturb the co-transcriptional folding of three cognate RNA hairpins of this protein: SMN1 and SMN2, two hairpins present in exon 7 of human *SMN1* and *SMN2* genes, respectively[22], and the stem loop 2 (EV2) of the internal ribosome entry site of Enterovirus 71 (ref. [23]) (Fig. 1b). To distinguish the RNA-specific UP1 perturbations from other changes in the

[1]Department of Biology, Institute of Molecular Biology & Biophysics, ETH Zurich, Zurich, Switzerland. [2]Department of Biology, Institute of Biochemistry, ETH Zurich, Zurich, Switzerland. [3]Department of Biosystems Science and Engineering, ETH Zurich, Zurich, Switzerland. *e-mail: yaroslav.v.nikolaev@gmail.com; allain@mol.biol.ethz.ch
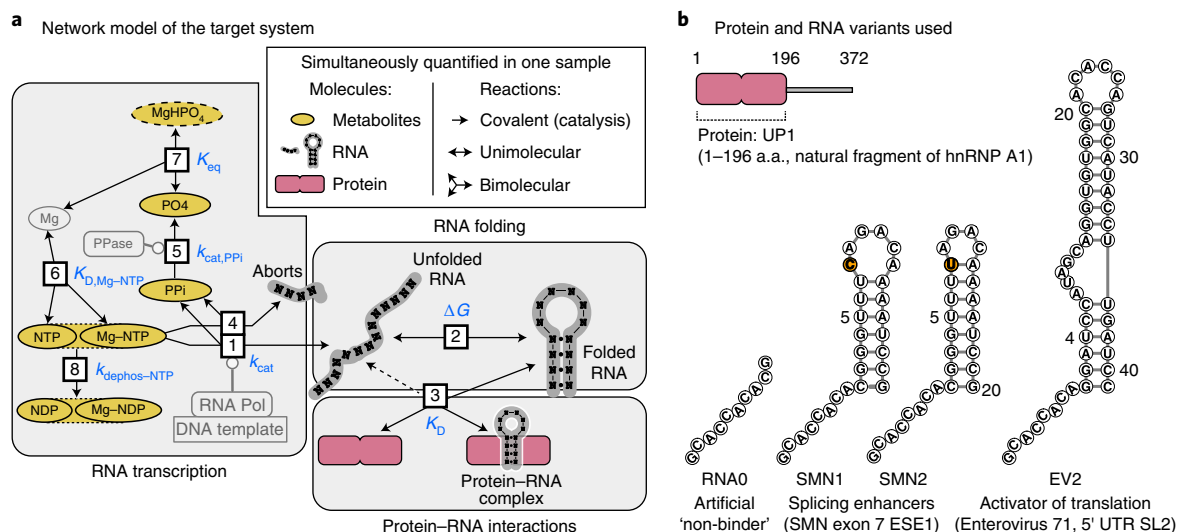
Fig. 1 | Molecules and reactions of the target co-transcriptional network. a, Organization of reactions in the reconstructed network: 1, RNA transcription; 2, RNA folding; 3, binding of regulatory protein to RNA (dashed line indicates that protein can also bind to the unfolded RNA); 4, formation of RNA aborts; 5, hydrolysis of pyrophosphate; 6, formation of Mg–NTP (nucleotide triphosphate) complexes; 7, formation of $MgHPO_4$ salt aggregates and 8, dephosphorylation of NTPs to NDPs (merged ovals indicate that NMR observables of free and Mg-bound nucleotide-phosphates; for example, NTP and MgNTP, cannot be separately discriminated). PO4 refers ambiguously to multiple states of phosphate ion, including PO_4^{3-} and its protonated forms. Indirectly observed $MgHPO_4$ species is shown with a dashed contour. Non-observable species—Mg, pyrophosphatase (PPase), RNA polymerase (RNA Pol) and DNA template—are shaded. Metabolites, proteins and RNA are shown in ocher, rose and gray, respectively. **b**, Protein and RNA variants used in Systems NMR experiments. To make the products of the abortive RNA transcription uniform in all constructs (**a**, reaction 4), the hairpin RNAs include a non-native single-stranded 5' overhang matching the sequence of the control RNA0. The hairpins include two non-native closing GC pairs to offset the instability caused by the 5' single-stranded overhang. The single C > U mutation between SMN1 and SMN2 RNAs is highlighted in orange. ESE1, exonic splicing enhancer 1; UTR, untranslated region; SL2, stem loop 2; a.a., amino acids.

network dynamics (pH, nucleotide triphosphate concentrations), a fourth, 'non-binding' RNA0 (RNA zero) was designed and tested as a control (Fig. 1b and Methods).

In our experiments, the DNA template, the nucleotide triphosphates (NTPs), $MgCl_2$, the pyrophosphatase and the RNA-binding protein are initially mixed in an NMR tube (Fig. 2a), and then transcription is triggered by addition of the T7 RNA polymerase. The reaction network (Fig. 1a) is subsequently monitored for ~20 h by repeating several NMR experiments (Fig. 2b–d and Supplementary Video 1): 1D (one-dimensional) ^{31}P, to monitor the levels of metabolites and RNA; 1D ^1H, to monitor RNA folding; and 2D (two-dimensional) ^1H-^{15}N, to monitor protein interactions. Each set of measurements takes ~30 min to record, yielding an overall dataset of ~120–160 NMR spectra with ~40 time points for each individual spectrum type. The combined number of resolved quantifiable NMR signals at each time point exceeded 170: 8 in the ^{31}P spectrum (Fig. 2b), more than 20 in the ^1H spectrum (Fig. 2c) and over 150 protein backbone amide signals in the 2D ^1H-^{15}N spectrum (Fig. 2d). For quantitative modeling of the target network, ten signals were used (Supplementary Table 1): the ^{31}P signals of inorganic phosphate (PO4 (a placeholder for multiple PO4 species), referring to PO_4^{3-} and its protonated forms), RNA phosphate, α, β and γ-phosphate of the NTP, α and β-phosphate of nucleotide diphosphate (NDP) (Fig. 2b,e); the ^1H imino signals of RNA uracils U5 (SMN1 and SMN2) or U4 (EV2) (Fig. 2f) and the ^1H-^{15}N signals of the selected UP1 residues reporting on RNA binding, His33 and Arg75 (Fig. 2g). In this study, these ten signals were sufficient to quantify the key parameters of the target eight-reaction network. The data from remaining signals can still be used in subsequent studies to investigate the system in more detail. For example, to analyze individual conversion rates of four NTPs, or to analyze RNA and protein perturbations not just via overall reaction constants, but with residue-level resolution.

Among the key features of NMR is the intrinsically quantitative nature of the observed signals, which permits direct determination of certain physico-chemical molecular properties with few or no calibrations. To quantify the metabolite and RNA concentrations, we measured the integrals of corresponding signals in ^{31}P spectra (Fig. 2h). Linewidths of NMR signals combine information about the molecule size (tumbling rate) and dynamics (lifetime) of molecular states, we therefore measured the linewidths of the well-separated imino signals of the folded RNA in ^1H spectra to quantify the RNA stability (Ura5 and Ura4, see Fig. 2i). The positions of NMR signals report on the chemical environment of the corresponding atoms. Therefore, to quantify the RNA binding to the protein, we measured the shifts in the positions of selected protein 'reporter' signals, which shifted systematically in the ^1H-^{15}N spectra between the free and bound protein states as the more RNA was bound (His33 and Arg75, Fig. 2j).

In summary, a quantitative NMR assay was established with dedicated reporter signals to monitor metabolite, RNA and protein dynamics in one sample (Supplementary Table 1 and Supplementary Video 1).

Network model from NMR data. To integrate the measured data and evaluate our understanding of the network dynamics, a mathematical model combining ordinary differential equations (ODE) was formulated (Methods). The initial model consisted of three reactions: RNA synthesis, RNA folding and protein–RNA binding (Fig. 1a, reactions 1–3). Unexpectedly, a reduction in the total integral of ^{31}P-containing species was observed over time (T_1-relaxation-corrected), sometimes followed by sharp drops in the concentration of free PO4 species at the end of transcription (Fig. 2h, first panel, blue trace). Further analysis revealed that the designed assay could also sense the formation of soluble $MgHPO_4$ aggregates, which are not directly visible in solution NMR. Extension of the network
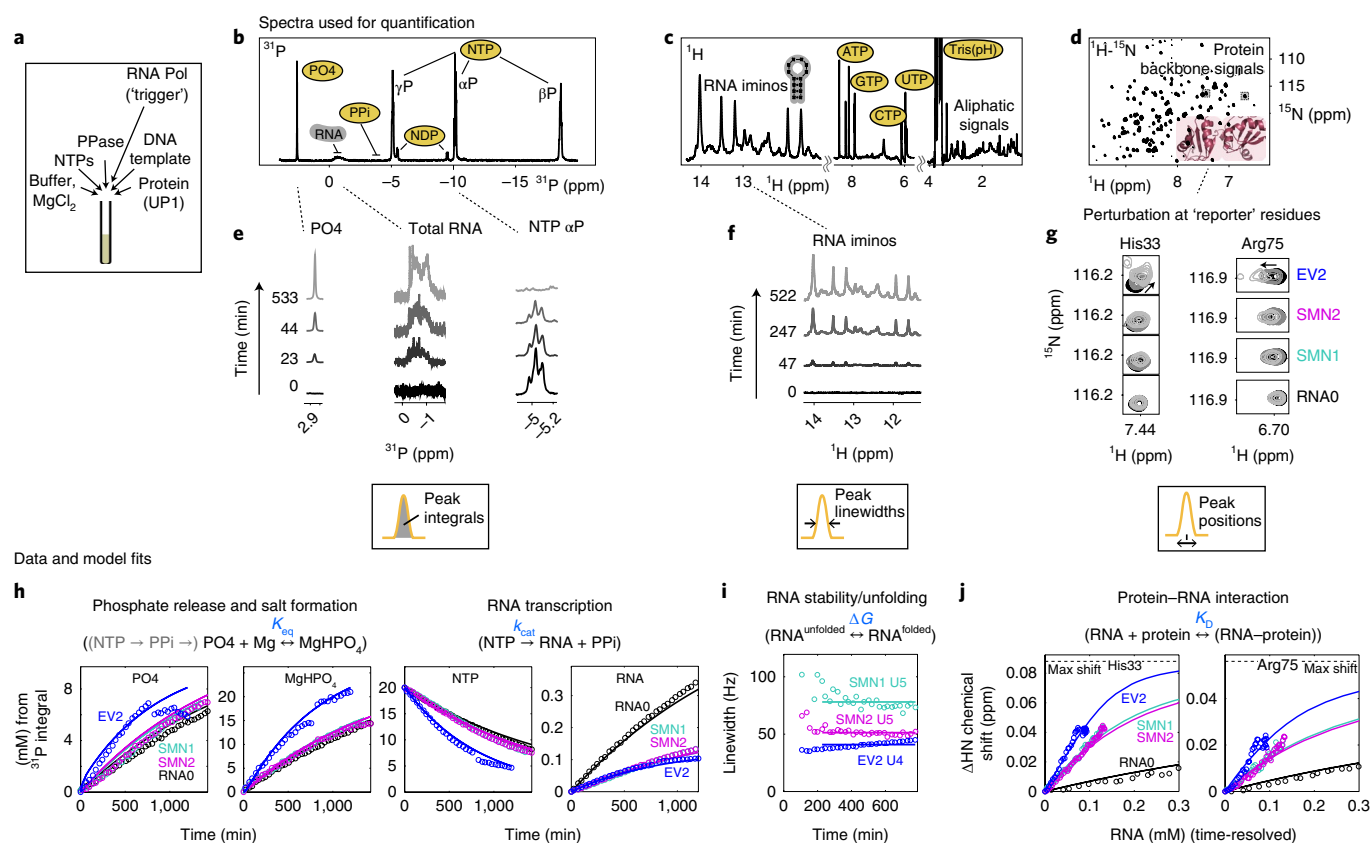
Fig. 2 | NMR observables, data and mathematical model fits used for network parameterization. a, Sample composition, 450 µl NMR tube with 150 µM UP1 protein, 33 nM T7 RNA polymerase and 20 mM starting NTPs, from which 0.1–0.4 mM RNA is synthesized by the end of transcription. **b–d**, Full spectra recorded in the NMR assay. **e–g**, Selected, time-resolved, spectra regions used for network quantification. From 1D ^{31}P spectra (**b,e**) the PO4, RNA, PPi, NDPs and NTPs species are quantified using integrals of the corresponding signals. From 1D ^1H spectra (**c,f**) the folded RNA is quantified using linewidths of imino proton signals. The 2D ^1H-^{15}N spectra (**d,g**) show a 'fingerprint' of all amino acid ^1H-^{15}N moieties in the protein, and the positions of selected reporter residues at the RNA-binding interface are used to quantify protein–RNA interaction. Experiments were repeated at least three times independently, using different batches of protein and/or DNA template, with similar results ($n=3$ (RNA0, SMN1), $n=4$ (SMN2, EV2)). **h–j**, Time-resolved quantified observables used for network parameterization (circles), and the resulting model fits (solid lines). **h**, Integrals from ^{31}P spectra converted into mM concentrations using the starting 20 mM NTPs as calibration reference. Abrupt intensity jumps in free PO4 and MgHPO4 are discussed in the main text. **i**, Linewidths of U5 imino signals in SMN RNAs and U4 in EV2. **j**, Chemical shift perturbations (CSP) of His33 and Arg75 residues of UP1 protein plotted against time-resolved RNA concentration. Time-resolved animation of exemplary data is shown in Supplementary Video 1.

model with the relevant reactions (Fig. 1a, reactions 4–8) allowed to quantify the MgHPO4 solubility, which matched the literature data (see below).

Correlating time-resolved concentrations of the synthesized RNA with the shifts of protein reporter signals within the same mathematical model, we could see that the established assay can sense the differences in UP1 protein affinity to the four tested RNAs. In particular, the smallest UP1 perturbations were observed in presence of the 'non-binder' RNA0, intermediate perturbations observed with the 'moderate'-affinity SMN RNAs, and strongest perturbations with 'high'-affinity EV2 RNA (Fig. 2g,j).

Validation of Systems NMR-derived reaction constants. The behavior of a reaction network can be predicted at any concentrations of reactants if the constants (fundamental parameters) of all reactions are known. Deriving these fundamental constants from experimental data is one of the main goals of mathematical modeling of reaction networks. From a single NMR assay, we could determine the constants of five out of the eight network reactions (numbers 1–3, 7 and 8, Fig. 1a), while the constants of the other reactions (numbers 4–6) were fixed (Methods and Supplementary Table 2). The five unconstrained constants were determined by fitting the mathematical model to the time-resolved NMR

observables. For each RNA (RNA0, SMN1, SMN2, EV2) at least three NMR assay replicates were recorded and fitted (Supplementary Fig. 1). For validation, four out of five multiplex-derived Systems NMR constants were compared with the constants derived by classical approaches, when a single reaction was perturbed at a time (the catalytic rate constant k_{cat}, the free energy ΔG, the affinity constant K_D and the equilibrium constant $K_{eq,MgHPO4}$). Remarkably, all tested constants were in agreement with classical methods.

The equilibrium constant for the formation of soluble MgHPO4 aggregates ($K_{eq,MgHPO4}$) showed an average value of 1.31 ± 0.06 mM, closely matching the 0.97 ± 0.05 mM value reported in the literature[24] (Fig. 3a).

The expected catalytic rate constant $k_{cat}=0.26\pm0.07$ nucleotides per second (nt s^{-1}) of T7 RNA polymerase[25] closely matched the average of 0.4 ± 0.12 nt s^{-1} for the three short RNAs in Systems NMR datasets (Fig. 3b). Both the literature reference and the NMR k_{cat} constants are averaging the initiation and elongation phases of transcription. This is manifested in the roughly two-fold increase in the overall $k_{cat}=0.73\pm0.06$ nt s^{-1} for the longer EV2 RNA (Fig. 3b), for which the polymerase spends more time in the faster elongation phase.

Based on the measured ultraviolet spectroscopy (UV) melting experiments (Supplementary Note 1), the free energy ΔG of folding of the two RNA hairpins (SMN1 and SMN2, differing by a single
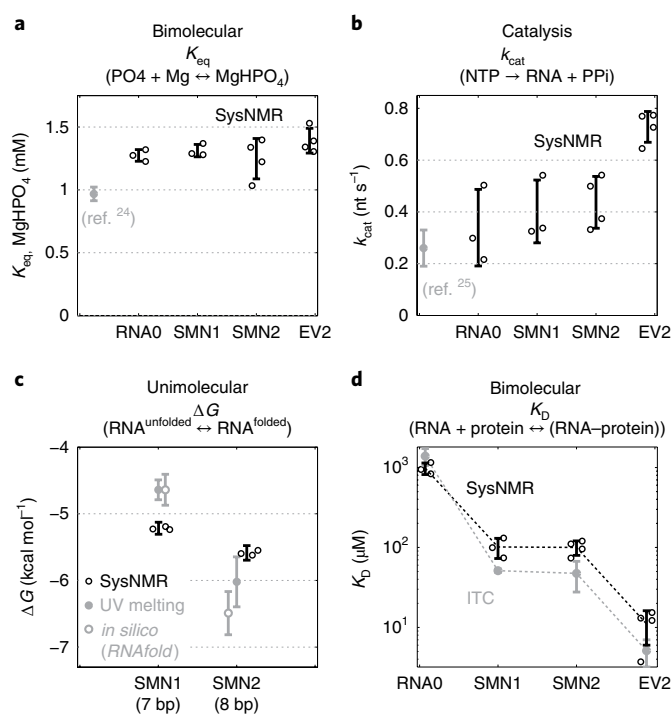
Fig. 3 | Validation of Systems NMR-derived constants. Systems NMR data shown in black, validation data in gray. **a**, Equilibrium constant for the formation of MgHPO$_4$ salt aggregate (from the decrease of total ^{31}P integral) compared to literature data[24]. **b**, RNA synthesis rate constant (from the decrease of NTP signals) compared to literature data[25]. **c**, Unimolecular free energy of RNA folding (from the linewidths of ^{1}H imino signals) compared to experimental ultraviolet (UV)-melting data (filled gray circles) and theoretical predictions by RNAfold (empty gray circles). **d**, Dissociation constant for protein–RNA complex formation (from the chemical shifts of protein ^{1}H-^{15}N signals) compared to ITC measurements. In Systems NMR data, the experimental parameter values (K_{eq}, k_{cat}, ΔG, K_D) and error bars correspond to the means ± s.d. from independent network model fits using independent experimental dataset replicates ($n = 3$ (RNA0, SMN1), $n = 4$ (SMN2, EV2), with different batches of DNA template and/or protein). In the validation experiments, the constants (UV-ΔG, ITC-K_D) correspond to the optimized parameter ± s.d. uncertainty of data fits for the melting curve (UV-ΔG) and binding isotherm (ITC-K_D). Individual data points, where available, are shown as circles next to the error bars.

base pair) are expected to be -4.64 ± 0.15 and -6.02 ± 0.37 kcal mol^{-1} (Fig. 3c). Systems NMR measured -5.2 ± 0.1 and -5.6 ± 0.1 kcal mol^{-1} for the respective constants (Fig. 3c). SMN2 stabilities thus matched within the standard deviation limits, while SMN1 stability was over-estimated by ~0.3 kcal mol^{-1} in NMR compared to ultraviolet data. Analysis of the SMN1 U5 imino signal peak shapes revealed partial peak doubling (Supplementary Fig. 2), suggesting that a more complex than a two-state model would be required for accurate analysis of SMN1 stability. RNA0 does not form hairpins, and the EV2 hairpin is too stable for the accurate ultraviolet-melting ΔG determination; therefore, their stabilities were not evaluated.

Isothermal titration calorimetry (ITC) was used to validate the affinity constants (K_D) of UP1 with the four different RNAs: one control (RNA0), the two hairpins of moderate-affinity (SMN1, SMN2) and the high-affinity EV2 RNA. The ITC affinity constants were $1,391 \pm 331$ μM for RNA0, 51.3 ± 2.5 μM for SMN1, 47.4 ± 19.7 μM for SMN2 and 5.1 ± 1.9 μM for EV2 (Fig. 3d and Supplementary Fig. 3). The corresponding constants from Systems NMR for the same four RNAs were 978 ± 162, 101 ± 29, 100 ± 20 and 11.1 ± 5.1 μM,

respectively, based on 3–4 replicate measurements for each RNA (Fig. 3d). Systems NMR appeared therefore accurate for the K_D measurement of the unspecific RNA0 control, but showed systematically weaker binding for the three specific RNA targets. This level of weakening could originate from unspecific UP1 interactions with RNA aborts (2–8 nt in length) and with free NTPs. The ability of UP1 to bind RNA aborts is evident from its $1,391$ μM ITC-derived affinity to the non-binding RNA0, whose sequence matches the sequence of RNA aborts in all four RNAs. Affinity of UP1 to free NTPs was also experimentally confirmed by standard NMR titrations, measuring an overall $K_{D,UP1-NTPs}$ of $16,200 \pm 2,100$ μM (Supplementary Fig. 4). RNA aborts increase from 0 to ~2,000 μM, and NTPs decrease from 20,000 to 5,000 μM during the transcription reaction (Supplementary Fig. 1), thus both of these can weaken the affinity of UP1 to specific RNAs under these conditions.

In summary, Systems NMR accurately quantified all core reaction constants of the target network in multiplex. All validated reaction constants matched the reference values with <2.5-fold difference.

RNA perturbations by proteins and small molecules. The reconstructed network was then perturbed by proteins and drug candidate molecules to gain insight into the network dynamics.

To probe the effect of protein on the folded RNA, the assays were performed under two conditions: 'co-transcriptionally', when the UP1 protein was added from the start and present during the entire period of RNA synthesis, and 'post-transcriptionally', when UP1 protein was added only near the end of transcription. The experiments showed that UP1 appears to (1) at least partially unwind the SMN2 hairpin and (2) forms a 2:1 complex with the EV2 RNA when UP1 is added post-transcriptionally, and only a 1:1 complex with EV2 when UP1 is present co-transcriptionally (Supplementary Fig. 2).

To probe the effect of small molecules, the reactions with SMN2 RNA were performed in presence of drug candidate molecules, recently developed to correct the aberrant splicing of exon 7 from the *SMN2* gene[26]. The experiments with SMN2 ESE1 suggested that under given co-transcriptional conditions one of the molecules may influence RNA folding, and another one reduces RNA transcription rate (Supplementary Fig. 5).

Multiplexed monitoring of protein perturbations during RNA transcription. RNA-binding proteins often synergize or compete for binding to the same RNA. For example, the splicing of the SMN2 exon 7 is regulated by hnRNP A1, SRSF1, hnRNP G and Tra2-β1 (ref. [27]). To facilitate the multiplexed analysis of interactions in this system of several RNA-binding proteins, we devised two labeling schemes that visualized the protein–RNA interaction interfaces in the RNA recognition motifs (RRM) of these proteins in one sample at the same time (Fig. 4a–c and Supplementary Note 2). RNA transcription was then performed in the presence of five ^{15}N-valine-labeled protein constructs mixed together (two independent RRMs in the case of SRSF1), monitoring all valines and quantifying their perturbations in real time (Fig. 4d,e).

During the reaction, all observed constructs except SRSF1–RRM2 showed an unexpected bi-modal response (Fig. 4e), with valine signals first decreasing and then increasing their intensity, many without substantial change in the signal positions. The samples showed evidence of liquid–liquid phase separation[28], which was confirmed by microscopy (Fig. 4f and Supplementary Fig. 6). Notably, none of the protein constructs included disordered regions. The number of phase-separated droplets decreased more than ten-fold when the transcription was performed in presence of individual proteins at the same total concentration as in the mixture of five (Fig. 4g and Supplementary Fig. 6). This suggests that this RNA-dependent phase separation is not driven here simply by high protein concentrations, but involves interactions between specific proteins.
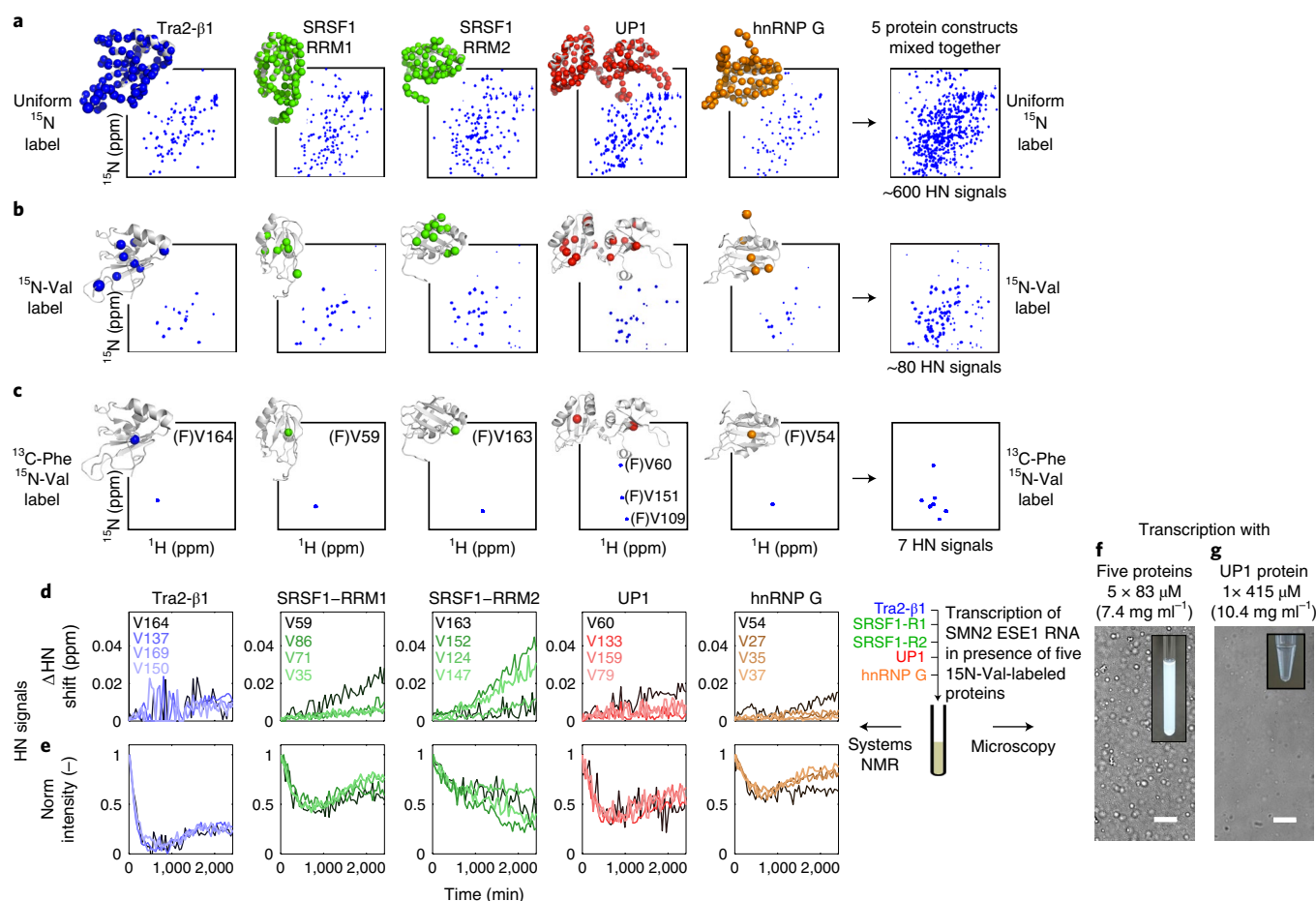
Fig. 4 | Monitoring multiple proteins in one NMR sample employing selective labeling. a–c, NMR spectra of five protein constructs (in columns) with three different labeling schemes: uniform ^{15}N (**a**), selective ^{15}N-Val (**b**) and double-selective ^{13}C-Phe,^{15}N-Val (**c**). Protein structures with observable NMR signals as colored spheres are shown next to all the spectra. **a,b**, show 2D HN spectra and **c** shows 2D HN(CO) spectra. Uniform ^{15}N labeling reveals HN signals for all residues except prolines in the 2D HN experiment; ^{15}N-Val labeling, only HN signals of valines; ^{13}C-Phe,^{15}N-Val labeling, HN signals of valines preceded by phenylalanines. The last column shows the spectral overlap on combining five proteins in one sample. **d,e**, Chemical shift perturbation (**d**) and intensity (**e**) changes of HN signals in a mixture of five proteins during transcription of SMN2 ESE1 RNA. Traces of four valine residues, including the most strongly perturbed ones, for each protein are shown. **f,g**, Microscopy images of the RNA transcription performed with (**f**) five protein constructs and (**g**) only UP1 protein at identical 415 μM total protein concentration (in mass units 7.4 mg ml^{-1} for the five-protein sample and 10.4 mg ml^{-1} for pure UP1). Scale bars (**f,g**), 10 μm. Five-protein transcription reactions were repeated twice ($n = 2$ independent experiments) with similar results.

Decrease followed by increase of protein NMR signal intensities in the absence of substantial change of the NMR signal positions suggests that the proteins phase separate into larger assemblies at the start of transcription reaction when the RNA/protein ratio is low, and are partially re-dissolved later when the RNA concentration increases. This matches the recently reported RNA-dependent phase-separation of hnRNP A1, TDP43 and FUS proteins in vitro and in vivo[29].

To evaluate whether the presented NMR setup could be implemented under physiological conditions, we measured nuclear concentrations of the UP1/hnRNP A1, SRSF1, hnRNP G and Tra2-β1 proteins in human embryonic kidney (HEK)293 cells, and found those to be 30, 7.6, 3.3 and 5.2 μM, respectively (Supplementary Fig. 7). This is close to the current NMR sensitivity limits (≥10–50 μM), suggesting that our setup, at least for hnRNP A1, can be tuned for observations under near-physiological conditions.

Discussion

The derivation of individual catalytic[30], unimolecular[31] and bimolecular[32] reaction constants by NMR is not uncommon, but the

elementary reaction types and main biomolecule classes in a single sample. Due to the non-destructive nature of NMR, each sample yields not just a snapshot of the network, but reveals its dynamics over time or another variable condition, thereby giving deeper insight into the network logic.

The different reaction constants determined from individual multiplexed NMR assays in our study appear accurate, showing <2.5-fold difference with validation values (Fig. 3). The differences between network-based and single-reaction-based assays can reveal unaccounted cross-talk reactions, such as the unspecific interactions of UP1 protein with the abortive RNAs and free NTPs detected here. Our results correlate with the recent UP1 specificity screens[33] and suggest that in vivo UP1/hnRNP A1 protein affinity to specific RNA targets will likely be ~1,000-fold weaker than the nM-range affinities anticipated from single-reaction in vitro assays[23].

Another emergent behavior we detected was the RNA-driven in vitro phase-separation in a system of five protein domains (Fig. 4d), which was largely absent for individual domains under the same conditions. This observation suggests that Systems NMR could be used to probe structural perturbations of proteins in phase-separated droplets and membraneless organelles[29], an emerging

research area with connections to various age-related disorders[28]. The method can resolve residue-level signals of multiple proteins at once, does not require chemical modifications of proteins and allows monitoring of enzymatic activities within the same assay.

NMR assay limits. One specific requirement of the assay developed here is the need to design a ~8–10 nt-long 5′ overhang RNA sequence that minimizes interference of short abortive RNAs with specific protein–RNA interactions and RNA folding. This sequence is designed algorithmically and can be used as a separate control to identify specific RNA effects from the other network perturbations.

More broadly, for a generic reaction network, present-day solution NMR permits the direct observation of rigid molecules below ~50–100 kDa in size[6,34] and at minimal concentration of ~10–50 μM[7,35]. Under certain conditions, observation of 1 MDa complexes can be achieved[36], and in combination with dissolution dynamic nuclear polarization technology, molecules at sub-μM concentrations can be transiently observed[37,38].

For catalytic reactions, NMR permits quantification of kinetic (non-equilibrium) processes on the time-scales going from seconds to hours and days[39]. For unimolecular reactions, many NMR techniques are available[31,40], potentially allowing quantification of low-populated molecular states down to fractions of percent from the main species (~5–10 kcal mol^{-1} in free energy difference). For bimolecular interactions, NMR currently permits direct quantification of dissociation constants in the low-μM to medium-mM range[32]. And by monitoring competitive displacement of weak-affinity ligands, also low-nM dissociation constants can be quantified[41].

As suggested by the selective labeling experiments shown here (Fig. 4) and the recent multiplexed NMR kinase assays[42], at least a few dozen of protein-focused reactions should be observable by NMR in one sample in parallel. The same multiplexing is also feasible for metabolites[43,44], but may be challenging for RNAs due to the higher degeneracy of their NMR signals[45].

While small-molecule NMR signals can mostly be interpreted ab initio, the assignment of observed signals to specific molecular epitopes in macromolecules requires time. Nevertheless, the NMR signal assignments from ~7,000 unique protein and ~600 unique RNA NMR structures are already available in the Protein Data Bank (pdb.org), providing an already vast starting ground for NMR network reconstructions.

Mathematical ODE models of reaction networks can be easily formulated using rule-based modeling[46], and computational methods exist to efficiently estimate network parameters and perform model selection[47–49], with virtually no limitations for moderately-large networks expected in NMR assays.

Applications. The generalized workflow in Systems Biology consists of four steps: experiment, modeling, prediction and testing of predictions, often repeated iteratively[50]. By uniquely providing both multiplexed and dynamic data from single samples at the first experimental stage, Systems NMR can accelerate the downstream development of accurate mathematical models, the understanding of network dynamics and the resulting predictions. Because NMR can dynamically monitor molecules in complex environments including living cells[10,12,13], the determination of true rates and constants for cellular networks in their natural context can generate reusable data for modeling and prediction of network dynamics.

Another advantage is that in vitro Systems NMR reconstructions provide an experimental ground of intermediate complexity, between simplified single-reaction in vitro assays and often very complex in vivo networks. Such moderate complexity may already reveal emergent network properties, such as phase-separation of RNA-binding domains observed here.

Considering specific applications, Systems NMR can advance studies of 'heterotypic' networks involving different molecule and/or reaction types. For example, concurrent quantification of perturbations in different parts of a biochemical network such as RNA transcription, folding and protein interactions observed here; or simultaneous quantification of catalysis and allosteric interactions in synthetic biology networks[51], or monitoring cross-talk between metabolic and signaling pathways[52,53].

In conclusion, combining the dynamic resolution of biochemical assays and the multiplexing ability of omics, we expect Systems NMR to pave the way to a deeper systems-level understanding of biological network dynamics both in fundamental and applied contexts.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41592-019-0495-7.

References

1. Szallasi, Z. *System Modeling in Cellular Biology* Ch. 10 (MIT Press, 2006); https://doi.org/10.7551/mitpress/9780262195485.001.0001
2. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
3. Howsmon, D. P. & Hahn, J. Regularization techniques to overcome over-parameterization of complex biochemical reaction networks. *IEEE Life Sci. Lett.* **2**, 31–34 (2016).
4. Yugi, K. & Kuroda, S. Metabolism as a signal generator across trans-omic networks at distinct time scales. *Curr. Opin. Syst. Biol.* **8**, 59–66 (2017).
5. Hart, Y. & Alon, U. Review the utility of paradoxical components in biological circuits. *Mol. Cell* **49**, 213–221 (2013).
6. Frueh, D. P., Goodrich, A. C., Mishra, S. H. & Nichols, S. R. NMR methods for structural studies of large monomeric and multimeric proteins. *Curr. Opin. Struct. Biol.* **23**, 734–739 (2013).
7. Kwan, A. H., Mobli, M., Gooley, P. R., King, G. F. & Mackay, J. P. Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS J.* **278**, 687–703 (2011).
8. Freedberg, D. I. & Selenko, P. Live cell NMR. *Annu. Rev. Biophys.* **43**, 171–192 (2014).
9. Bernstein, M. A. Reaction monitoring using NMR. *Magn. Reson. Chem.* **54**, 422 (2016).
10. Wolak, J. et al. in *Handbook of Metabolomics* (eds Whei-Mei Fan T., Lane A. N. & Higashi R. M.) 321–392 (Springer, 2012).
11. Eicher, J. J., Snoep, J. L. & Rohwer, J. M. Determining enzyme kinetics for systems biology with nuclear magnetic resonance spectroscopy. *Metabolites* **2**, 818–843 (2012).
12. Lerche, M. H., Jensen, P. R., Karlsson, M. & Meier, S. NMR insights into the inner workings of living cells. *Anal. Chem.* **87**, 119–132 (2015).
13. Smith, M. J. et al. Real-time NMR monitoring of biological activities in complex physiological environments. *Curr. Opin. Struct. Biol.* **32**, 39–47 (2015).
14. Fan, T. W. & Lane, A. N. Applications of NMR spectroscopy to systems biochemistry. *Prog. Nucl. Magn. Reson. Spectrosc.* **92-93**, 18–53 (2016).
15. Cordier, F. et al. Ordered phosphorylation events in two independent cascades of the PTEN C-tail revealed by NMR. *J. Am. Chem. Soc.* **134**, 20533–20543 (2012).
16. Theillet, F.-X. et al. Cell signaling, post-translational protein modifications and NMR spectroscopy. *J. Biomol. NMR* **54**, 217–236 (2012).
17. Lee, M.-K., Gal, M., Frydman, L. & Varani, G. Real-time multidimensional NMR follows RNA folding with second resolution. *Proc. Natl Acad. Sci. USA* **107**, 9192–9197 (2010).
18. Wacker, A., Buck, J., Richter, C., Schwalbe, H. & Wöhnert, J. Mechanisms for differentiation between cognate and near-cognate ligands by purine riboswitches. *RNA Biol.* **9**, 672–680 (2012).
19. Steinert, H. et al. Pausing guides RNA folding to populate transiently stable RNA structures for riboswitch-based transcription regulation. *eLife* **6**, 1–18 (2017).
20. Breaker, R. R. Prospects for riboswitch discovery and analysis. *Mol. Cell* **43**, 867–879 (2011).
21. Zhang, J. & Landick, R. A two-way street: regulatory interplay between RNA polymerase and nascent RNA structure. *Trends Biochem. Sci.* **41**, 293–310 (2016).
22. Singh, R. N. Evolving concepts on human SMN pre-mRNA splicing. *RNA Biol* **4**, 7–10 (2007).

23. Levengood, J. D., Tolbert, M., Li, M. & Tolbert, B. High-affinity interaction of hnRNP A1 with conserved RNA structural elements is required for translation and replication of enterovirus 71. *RNA Biol.* **10**, 1136–1145 (2013).

24. Racz, G. J. & Soper, R. J. Solubility of dimagnesium phosphate trihydrate and trimagnesium phosphate. *Can. J. Soil Sci.* **48**, 265–269 (1968).

25. Huang, Y., Beaudry, A., Mcswiggen, J. & Sousa, R. Determinants of ribose specificity in RNA polymerization: effects of Mn^{2+} and deoxynucleoside monophosphate incorporation into transcripts. *Biochemistry* **36**, 13718–13728 (1997).

26. Sivaramakrishnan, M. et al. Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers. *Nat. Commun.* **8**, 1476 (2017).

27. Singh, N. N. et al. An intronic structure enabled by a long-distance interaction serves as a novel target for splicing correction in spinal muscular atrophy. *Nucleic Acids Res.* **41**, 8144–8165 (2013).

28. Boeynaems, S. et al. Protein phase separation: a new phase in cell biology. *Trends Cell Biol.* **28**, 420–435 (2018).

29. Maharana, S. et al. RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918–921 (2018).

30. Bowen, S. & Hilty, C. Time-resolved dynamic nuclear polarization enhanced NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **47**, 5235–5237 (2008).

31. Bothe, J. R. et al. Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *Nat. Methods* **8**, 919–931 (2011).

32. Fielding, L. NMR methods for the determination of protein-ligand dissociation constants. *Prog. Nucl. Magn. Reson. Spectrosc.* **51**, 219–242 (2007).

33. Jain, N., Lin, H.-C., Morgan, C. E., Harris, M. E. & Tolbert, B. S. Rules of RNA specificity of hnRNP A1 revealed by global and quantitative analysis of its affinity distribution. *Proc. Natl Acad. Sci. USA* **114**, 2206–2211 (2017).

34. Takeuchi, K. & Wagner, G. NMR studies of protein interactions. *Curr. Opin. Struct. Biol.* **16**, 109–117 (2006).

35. Bieri, M. et al. Macromolecular NMR spectroscopy for the non-spectroscopist: beyond macromolecular solution structure determination. *FEBS J.* **278**, 704–715 (2011).

36. Wiesner, S. & Sprangers, R. Methyl groups as NMR probes for biomolecular interactions. *Curr. Opin. Struct. Biol.* **35**, 60–67 (2015).

37. Frydman, L. & Blazina, D. Ultrafast two-dimensional nuclear magnetic resonance spectroscopy of hyperpolarized solutions. *Nat. Phys.* **3**, 415–419 (2007).

38. Kim, Y. & Hilty, C. Affinity screening using competitive binding with fluorine-19 hyperpolarized ligands. *Angew. Chemie Int. Ed.* **54**, 1–5 (2015).

39. Gal, M., Schanda, P., Brutscher, B. & Frydman, L. UltraSOFAST HMQC NMR and the repetitive acquisition of 2D protein spectra at Hz rates. *J. Am. Chem. Soc.* **129**, 1372–1377 (2007).

40. Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* **5**, 808–814 (2009).

41. Dalvit, C., Fagerness, P. E., Hadden, D. Ta, Sarver, R. W. & Stockman, B. J. Fluorine-NMR experiments for high-throughput screening: Theoretical aspects, practical considerations, and range of applicability. *J. Am. Chem. Soc.* **125**, 7696–7703 (2003).

42. Thongwichian, R. et al. A multiplexed NMR-reporter approach to measure cellular kinase and phosphatase activities in real-time. *J. Am. Chem. Soc.* **137**, 6468–6471 (2015).

43. Soininen, P. et al. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst* **134**, 1781–1785 (2009).

44. Lindon, J. C., London, I. C. & Kensington, S. Biofluids studied by NMR spectroscopy. *Encycl. Spectrosc. Spectrom.* **1**, 128–141 (2010).

45. Lu, K., Miyazaki, Y. & Summers, M. F. Isotope labeling strategies for NMR studies of RNA. *J. Biomol. NMR* **46**, 113–125 (2010).

46. Chylek, La et al. Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **6**, 13–36 (2014).

47. Geier, F., Fengos, G., Felizzi, F. & Iber, D. in *Computational Modeling of Signaling Networks* (eds Liu, X. & Betterton, M. D.) 23–39 (Humana Press, 2012).

48. Maiwald, T. & Timmer, J. Dynamical modeling and multi-experiment fitting with PottersWheel. *Bioinformatics* **24**, 2037–2043 (2008).

49. Raue, A. et al. Data2Dynamics: a modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics* **31**, 3558–3560 (2015).

50. Vanlier, J., Tiemann, C. A., Hilbers, P. A. J. & van Riel, N. A. W. Parameter uncertainty in biochemical models described by ordinary differential equations. *Math. Biosci.* **246**, 305–314 (2013).

51. Chen, Z., Rappert, S. & Zeng, A. Rational design of allosteric regulation of homoserine dehydrogenase by a nonnatural inhibitor. *ACS Synth. Biol.* **4**, 126–131 (2015).

52. Smith, M. J. & Ikura, M. Integrated RAS signaling defined by parallel NMR detection of effectors and regulators. *Nat. Chem. Biol.* **10**, 223–230 (2014).

53. Tripodi, F., Nicastro, R., Reghellin, V. & Coccetti, P. Post-translational modifications on yeast carbon metabolism: Regulatory mechanisms beyond transcriptional control. *Biochim. Biophys. Acta - Gen. Subj.* **1850**, 620–627 (2015).

Acknowledgements

Author contributions

Y.N. conceived the idea, obtained initial funding, performed experiments, data analysis and network modeling. F.H.-T.A. supervised the project, provided infrastructure and financial support. D.I. supervised the network modeling part of the project. N.R. expressed and purified selectively labeled proteins and contributed to the transcription-NMR experiments. P.P. and M.S. supervised and provided technical and analytical support on mass spectrometry analyses. Y.N. wrote the manuscript with input from all of the authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-019-0495-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to Y.N. or F.H.-T.A.

Peer review information: Allison Doerr was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

A Supplementary Protocol describing how to set up and analyze data using Systems NMR for the presented here network is available at Protocol Exchange (https://doi.org/10.21203/rs.2.9160/v1) and the most recent version can be found at github.com/systemsnmr/ivtnmr.

RNA construct design. The sequence of the control RNA0 was designed algorithmically, using custom-built MATLAB scripts (github.com/systemsnmr/ivtnmr), from all possible sequences using four requirements: starts with G; contains no purine pairs that are recognized specifically by UP1 protein; contains ≥30% purines, to reduce RNA polymerase dissociation/abortion at the initiation stage; does not form stable dimers or hairpins with itself or target SMN and EV2 RNA sequences. This resulted in seven variants, of which (5′-GCACCACACG-3′) was chosen, as it showed fewest unspecific signals in the NMR imino region during transcription. The hairpin RNAs included non-native single-stranded 5′ overhang matching the sequence of the control RNA0 to make the abortive RNA products uniform in all constructs, and contained two non-native closing GC pairs to offset the instability caused by the 5′ single-stranded overhang.

DNA templates. For RNA transcription, corresponding sequences were cloned into pTX1 vector[54] at SapI sites, using double-stranded DNAs from commercial (Microsynth AG) single-stranded oligos: RNA0 (ATAGCACCACACG, TCACGTGTGGTGC), SMN1 (ATAGCACCACACGGGGTTTCAGACAAAATCCG, TCACGGATTTTGTCTGAAACCCGTGTGGTGC), SMN2 (ATAGCACCACACGGGGTTTTAGACAAAATCCG, TCACGGATTTTGTCTAAAACCCGTGTGGTGC), EV2 (ATAGCACCACAG-GATCAATAGCAGGTGTGGCACACCAGTCATACCTTGATCC, TCAGGATCAAGGTATGACTGGTGTGCCACACCTGCTATTGATCCTGTGGTGC). Plasmids were purified using the Nucleobond Xtra Midi kit (Macherey Nagel) and the final pellets washed three times with 70% ethanol, dried and linearized by BsaI (NEB) enzyme for 15 h at 50 °C in NEB3.1 buffer.

Proteins. All constructs and purification procedures were described earlier: UP1 (ref. [55]), SRSF1-RRM1 and RRM2 (ref. [56]), Tra2-β1 (ref. [57]) and hnRNP G[58]. After purification, proteins were transferred into transcription-NMR buffer (40 mM Tris-HCl, 0.01% Triton-X100, 5 mM dithiothreitol (DTT), pH 7.7) by dialysis, then flash-frozen and stored at −20 °C.

In vitro transcription in NMR tube. Reactions were performed at 30 °C, 40 mM Tris-HCl, 0.01% Triton-X100, 5 mM DTT, pH 7.7 supplemented with 5 mM of each nucleotide triphosphate (AppliChem), 24 mM MgCl₂, 1 U per ml inorganic pyrophosphatase from baker's yeast (Sigma), 5% D₂O, 50 μM 4,4-dimethyl-4-silapentane-1-sulfonic acid, 280 nM T7 RNA polymerase and 33 nM DNA template. Protein concentrations were 150 μM in single-protein reactions, and 83 μM of each protein was used for multi-protein reactions. In multi-protein experiments, 50 mM L-Arg and 50 mM L-Glu (AppliChem) were added to reduce protein aggregation, which have also likely reduced the systems' propensity for phase separation. T7 RNA polymerase was purified at 4 °C using Ni–NTA HisTrap chromatography (GE Healthcare), and stored at 70 μM concentration in 25 mM Tris-HCl pH 8, 50 mM NaCl, 0.5% β-mercaptoethanol (β-ME) and 50% w/v glycerol at −20 °C.

NMR experiments. Experiments were measured on Bruker AVIII-600 MHz with a CPQCI cryoprobe, and consisted of repeating series of 1D ¹H-watergate (spectral width (SW) 22 ppm; acquisition time (AQ) 0.62 s; D1 (interscan delay) 1 s; number of scans (NS) 128), 2D ¹H-TOCSY (SW 10/9 ppm; AQ 0.17/0.018 s; D1 0.5 s; NS 4), 1D ³¹P (SW 50 ppm; AQ 0.66 s; NS 256; carrier −8.22 ppm; D1 0.8 s), 1D ¹H-SOFAST (SW 24 ppm; AQ 0.053 s; D1 0.1 s; NS 1,536; ¹H excitation with Pc9 pulse, 5 ppm wide, centered at 12.9 ppm) and 2D ¹H¹⁵N-SOFAST-HMQC (H/N: SW 16/23.5 ppm; AQ 0.106/0.035 s; D1 0.2 s; NS 24; ¹⁵N carrier 117.8; ¹H excitation with Pc9 pulse, 4 ppm wide, centered at 7.95 ppm). BEST–TROSY 2D HN(CO) for analysis of selectively labeled proteins was provided by F. Lohr (BMRZ, Goethe Universitat Frankfurt) and measured with (H/N: SW 12/16 ppm; AQ 0.107/0.090 s; D1 0.2 s; NS 16; ¹⁵N carrier 117.8; ¹H excitation with Pc9 pulse, 4.2 ppm wide, centered at 8.5 ppm). NMR spectra were sorted, processed and analyzed using TopSpin 3.x (Bruker), custom-built Python and MATLAB scripts and the rbnmr routine (N. Nyberg, RBNMR, MATLAB Central File Exchange no. 40332, 2013). Chemical shifts of protein residues in 2D HN spectra were traced in CARA (cara.nmr.ch). SOFAST, selective optimized flip-angle short-transient; HMQC, heteronuclear multiple quantum coherence; BEST-TROSY, band-selective excitation short-transient transverse relaxation optimized spectroscopy.

NMR observables. For final network modeling, ten signals were used: ³¹P spectra: (1) PO4 (PO₄³⁻ and its protonated forms), (2) RNA, (3–5) αNTP, βNTP, γNTP, (6–7) αNDP, βNDP; (8) ¹H spectra: U5 (SMN1/2) or U4 (EV2) imino signals; (9–10) 2D HN spectra—His33 and Arg75 residues of UP1. Populations of phosphate-containing species were calculated stoichiometrically from ³¹P integrals using αNTP integral at time = 0 as 20 mM internal calibration. Each ³¹P integral

was T₁-relaxation-weighted using the ratios of corresponding integrals measured in the reference ³¹P spectra with 30 and 0.8 s interscan delays at the end of the transcription reaction, that is, ($I_{31P, corrected} = I_{31P} \times (I_{31P,ref,d1=30s}/I_{31P,ref,d1=0.8s})$). NTP and NDP populations were quantified from αP integrals, MgHPO₄ was calculated from decrease of the total T1-weighted ³¹P integral of all species. For long RNAs ($\gtrsim$20 nt), the 1D ³¹P signals became too broad, preventing accurate quantification at a reasonable time-resolution, so RNA concentrations were calculated from the decay of NTP signals. Due to the NMR signal degeneracy, the fractions of RNA and aborts could not be quantified within the NMR assay, and were fixed to 30 and 70%, by nucleotide mass, based on quantitative ultraviolet (260 nm) high-performance liquid chromatography (HPLC) analysis of reaction end-products.

The fraction of bound UP1 protein was derived from the chemical shift perturbation (CSP) of HN signals of His33 and Arg75. These residues were chosen as reporters for two reasons. First, they appeared to sense the same molecular epitope in all four protein–RNA complexes, as they all displayed signals moving in the same direction during transcription (Fig. 2g). Based on the existing UP1-RNA/DNA structures (PDB: 4YOE, 2UP1), these two residues are located near the RNA-binding pocket, but do not directly interact with the RNA, which explains how they could sense the same epitope changes independent of the RNA sequence. Second, these signals were sensitive to the differences in affinity of the RNA binding, as the magnitude of the ¹H-¹⁵N CSP varied for four different protein–RNA complexes (Fig. 2g,j). In assays with 150 μM UP1 concentration, both signals appeared predominantly in fast exchange with respect to the NMR time scale, and so the fast exchange assumption[59] was used during modeling. The experimental HN CSPs were calculated using $\Delta HN = \sqrt{(\Delta H^2 + (\Delta N \times 0.2)^2)/2)}$. The effect of pH change on histidine signal position was assumed negligible because the transcription buffer pH = 7.7 is far from the histidine p$Ka \approx 6$, and the chemical shifts of other surface-exposed UP1 histidines did not exhibit the same perturbation effects. Calculating the fraction of the bound protein under the fast exchange regime requires information on protein signal positions in the free and fully bound states. The shifts of the bound state are usually estimated as one of the parameters of the K_D fitting procedure, as the asymptote of the protein saturation curve. The 150 μM UP1 protein data did not approach this saturation under the assay conditions, because of the low RNA/protein ratio, since the final concentrations of specific RNAs reached only ~120 μM, giving only ~0.8:1 RNA/protein ratio. Addition of pure RNA to saturate the protein under these conditions consistently led to protein precipitation, which correlates with UP1's ability to phase-separate and aggregate in presence of RNA. Therefore, to estimate the HN signal positions of His33 and Arg75 in the saturated protein, an additional set of transcription experiments was recorded using 20 and 30 μM UP1 and the 'high-affinity' EV2 RNA. This allowed to increase protein saturation by reaching ~6:1 and 4:1 RNA/protein ratios (~120/20 and 120/30 μM, respectively) at the end of the reaction. Due to the poor NMR sensitivity at 20–30 μM protein concentrations, the 2D HN spectra in these experiments required ~10 h acquisition time and could only be recorded as non-time-resolved spectra at end of transcription reaction, when the system reached equilibrium. The combined CSP data from the datasets with high (150 μM) and low (20–30 μM) protein concentration gave an imperfect fit with the single-site binding model, suggesting that UP1 is already close to saturation in the assays with 150 μM protein concentration (Supplementary Fig. 8). This is likely due to additional weak UP1 binding sites in the EV2 RNA, as suggested by Arg75 peak splitting in assays with 20–30 μM UP1 concentration. The chemical shifts of His33 and Arg75 residues in the fully saturated protein state for the final ODE modeling of the four main datasets (RNA0, SMN1, SMN2, EV2 at 150 μM UP1) were taken as the shifts giving best fit when simultaneously fitting the data from EV2 RNA datasets with 150, 30 and 20 μM UP1 protein concentration (Supplementary Fig. 8).

All above ³¹P and HN data was used for global parameterization of the ODE model, and ¹H imino signals were used for line shape analysis and RNA ΔG derivation.

Mathematical modeling. The model was built using the BioNetGen language[60] and resulted in eight rate equations and nine differential equations:

Rate equations

$$v_1 (\text{synthesis of RNA}) = k_{cat,RNA} \times NTP$$
$$v_{2F, 2R} (\text{RNA folding}) = \text{analytical solution, see below} \tag{1}$$

$$v_{3F} (\text{binding of protein to RNA}) = k_{on} \times \text{protein} \times \text{RNA} \tag{2}$$

$$v_{3R} (\text{dissociation of protein−RNA complex}) = k_{off} \times [\text{protein−RNA complex}] \tag{3}$$

$$v_4 (\text{synthesis of aborts}) = k_{cat, aborts} \times NTP \tag{4}$$

$$v_5 (\text{hydrolysis of pyrophosphate}) = k_{catPPi} \times PPi$$
$$v_{6F,6R} (\text{formation of Mg−NTP complexes}) = \text{fixed constant, see below} \tag{5}$$

$$v_{7F} (\text{formation of MgHPO}_4 \text{ salt}) = k_{precip} \times PO_4 \tag{6}$$

$$v_{7R}(\text{dissociation of MgHPO}_4 \text{ salt}) = k_{\text{dissolve}} \times \text{MgHPO}_4 \qquad (7)$$

$$v_8(\text{dephosphorylation of NTPs}) = k_{\text{dephos.NTP}} \times \text{NTP} \qquad (8)$$

Differential equations

$$\frac{d\text{NTP}}{dt} = -v_1 - v_4 - v_8 \qquad (9)$$

$$\frac{d\text{RNA}}{dt} = +\frac{v_1}{\text{RNA length}} - v_{3F} + v_{3R} \qquad (10)$$

$$\frac{d\text{PO4}}{dt} = +2 \times v_5 - v_{7F} + v_{7R} + v_8 \qquad (11)$$

$$\frac{d\text{PPi}}{dt} = +v_1 \times \frac{\text{RNA length} - 1}{\text{RNA length}} + v_4 \times \frac{\text{Aborts length} - 1}{\text{Aborts length}} - v_5 \qquad (12)$$

$$\frac{d\text{MgHPO}_4}{dt} = +v_{7F} - v_{7R} \qquad (13)$$

$$\frac{d\text{NDP}}{dt} = +v_8 \qquad (14)$$

$$\frac{d\text{Aborts}}{dt} = +\frac{v_4}{\text{Aborts length}} \qquad (15)$$

$$\frac{d\text{Protein}}{dt} = -v_{3F} + v_{3R} \qquad (16)$$

$$\frac{d[\text{RNA-protein complex}]}{dt} = +v_{3F} - v_{3R} \qquad (17)$$

At the given hairpin stability ($\approx -5.5 \text{ kcal mol}^{-1}$) the SMN and EV2 RNAs are predominantly in a single state (99,99% folded). The expected rate of hairpin folding ($v_{2F} = 8 - 51 \times 10^3 \text{ s}^{-1}$ for ~30 nt hairpin)[61] is three orders of magnitude faster than protein–RNA encounters at the unbiased diffusion rate (~$10^5 \text{ M}^{-1}\text{s}^{-1}$)[62]. For these reasons the final ODE model treats the folded and unfolded RNA as a single species, and the stability of RNA is derived independently from imino signal line shape analysis at each reaction time point (see below). Because of the sequence identity with the abortive products, RNA0 length is used as a weighted sum of 10 nt full length RNA0 and the 4 nt average length of corresponding abortive products: $0.3 \times 10 + (1 - 0.3) \times 4 = 5.8 \text{ nt}$. The final ODE model was deposited in BioModels[63] (MODEL1812270001).

Model fitting. The ODE model was fitted to the NMR-based observables employing the gradient based Trust-Region method, using a custom-built set of MATLAB routines (github.com/systemsnmr/ivtnmr) based on earlier code[47]. Model parameters, their boundaries and initial optimization conditions are given in Supplementary Table 2. The experimental errors were assumed to follow a Gaussian distribution and the standard deviations of individual data points were calculated from spectral noise for ^{31}P integrals, and from the variance of protein HN chemical shifts at the end of the reaction (when the network is approaching equilibrium state). The values and standard deviations of the derived network parameters were obtained with two alternative approaches. In the first method, we calculated the standard deviation between estimates that were obtained in independent ODE model fits of several NMR replicate datasets, varying batches of the protein and/or DNA template preparations (Fig. 3 and Supplementary Fig. 1). In the second method, parameter uncertainties were estimated with bootstrap analysis. Here, we generated 200 model fits using resampling of the full data vector with replacement. The confidence intervals obtained by bootstrapping (Supplementary Fig. 9) were narrower than the variability between the 3–4 experimental replicates, and therefore the values from the replicate analysis were chosen as the ones more realistically reflecting the parameter variance.

Derivation of final constants. The specific constants used in the rate equations of the ODE model were converted to more general constants—for comparison with other experiments and literature data.

The T7 RNA polymerase enzyme (280 nM) is assumed to be saturated by the substrate (4–20 mM NTPs) during the reaction, so the K_M can be ignored and the concentration of the active enzyme can be defined by the limiting concentration of the DNA template (33 nM).

$$k_{\text{cat}} = \frac{(k_{\text{cat,RNA}} + k_{\text{cat,Aborts}})}{33 \text{ nM}(\text{concentration of DNA template})} \qquad (18)$$

Because reactions 3 and 7 (Fig. 1a, protein binding and aggregation of MgHPO$_4$) equilibrate faster than the time-resolution of the current setup, only the equilibrium constants (K_{eq} and K_D, ratios of kinetic rate constants) for these reactions could be parameterized.

$$K_{\text{eq}} = K_D = \frac{k_{\text{off}}}{k_{\text{on}}} \qquad (19)$$

Pyrophosphate hydrolysis can be quantified from the specific pyrophosphate (PPi) observable, but this rate was too rapid for meaningful quantification. Therefore, this constant was fixed based on the enzyme activity provided by the manufacturer (Sigma).

$$k_{\text{catPPi}} = 1 \text{ mM min}^{-1} \qquad (20)$$

The formation of soluble Mg–PO$_4$ aggregates depends on the concentration of the free Mg. Apart from Mg–PO$_4$, the free Mg can also participate in Mg–NTP and Mg–RNA complexes. The fraction of the free Mg was estimated from the published $K_{D,\text{Mg-NTP}} = 0.3 \text{ mM}$ (ref. [64]) using the general bimolecular binding isotherm[65] denoted with f in the equation below. For the calculation the concentration of free Mg available for Mg–PO$_4$ salt formation was assumed constant during the reaction.

$$\text{Mg}_{\text{free}} = f(\text{Mg}_{\text{total}}, \text{NTP}_{\text{total}}, K_{D,\text{Mg.NTP}}) = f(24, 20, 0.3) = 5.1 \text{ mM} \qquad (21)$$

$$K_{\text{eq,MgHPO}_4} = \frac{k_{\text{precip}}}{k_{\text{dissolve}}} \times \frac{1}{\text{Mg}_{\text{free}}} \qquad (22)$$

RNA folding ΔG from imino signal line shape analysis. For fitting, 1D ^{1}H-SOFAST spectra were Fourier-transformed with no apodization function. Imino signals were fitted to a single-Lorentzian function using the lorentzfit routine (J. Wells, Lorentzfit, MATLAB Central File Exchange no. 33775, 2015). The fits used 0.2 ppm fitting window and assumed a baseline fixed at zero signal intensity. In the single-Lorentzian fit, the intrinsic broadening of a signal cannot be distinguished from broadening due to overlapping signals, thus accurate quantification requires well-resolved signals. The fitted linewidth parameter (full width at half maximum) was used to derive the unfolding-driven imino-exchange rate ($k_{\text{ex,unfolding}}$).

$$k_{\text{ex}} = \text{linewidth} \times \pi \qquad (23)$$

$$k_{\text{ex}} = k_{\text{ex,unfolding}} + k_{\text{ex,base-flipping}} + R_2 \qquad (24)$$

The contribution of B0 field inhomogeneity ($R_{2(B0)}$) is considered negligible. The combined contribution of base-flipping ($k_{\text{ex,base-flipping}}$)[31] and transverse relaxation rate (R_2) was determined from linewidths of imino signals in purified SMN2 hairpin additionally stabilized by terminal GCs (13-base pair, GGCGGGUUUUGGC-AGAC-GCCAAAAUCCGCC). In this stabilized RNA, the exchange by global unfolding is suppressed ($\Delta G = -23 \text{ kcal mol}^{-1}$), which was confirmed by the negligible dependence of its imino integrals on temperature (not shown). The combined ($k_{\text{ex,base-flipping}} + R_2$) value for imino signals in an UA pair flanked by GU and UA pairs, under transcription buffer conditions, was determined to be 61.4 s^{-1}.

Imino linewidths depended on pH (the concentration of imino-exchange catalyst), indicating that the system is under the bimolecular exchange regime (EX2)[66], and hence the measured k_{ex} reports on the equilibrium constant of RNA unfolding/opening ($K_{\text{eq,unfolding}} = K_{\text{op}}$). The intrinsic exchange rate $k_{\text{ex,intrinsic}}$ (same as the exchange from the open state, $k_{\text{ex,open}}$) in the transcription buffer was measured to be ~10^6 s^{-1} for both UTP and GTP, using a protocol described elsewhere[67]. The final free energy of folding was determined using:

$$K_{\text{eq, unfolding}} = \frac{k_{\text{ex,unfolding}}}{k_{\text{ex,intrinsic}}} \qquad (25)$$

$$\Delta G_{\text{folding}} = -\Delta G_{\text{unfolding}} = R \times T \times \log(K_{\text{eq,unfolding}}) \qquad (26)$$

where R is the gas constant and T is the absolute temperature.

RNA purification. RNAs for ITC and ultraviolet-melting experiments were purified by anion-exchange HPLC under denaturing 6 M urea, 80 °C conditions, followed by n-butanol extraction, snap-cooling and lyophilization.

Ultraviolet temperature melting. For the melting experiments the RNA hairpins were produced without the single-stranded 5′ overhang to eliminate ultraviolet baseline distortions caused by this single-stranded region. The experiments used 2 μM RNAs in 10 mM sodium-cacodylate, pH 7.35, 5 mM MgCl$_2$ and 25 mM L-Arg/L-Glu buffer. Details of the analysis are shown in Supplementary Note 1.

ITC. Experiments used conditions approximating those at the end of transcription-NMR reaction: 40 mM Tris-HCl, 0.01% Triton-X100, 2.5 mM β-ME, pH 7.5, 37 mM NaPO$_4$, 2.6 mM NTPs, 24 mM MgCl$_2$ and 303 K. DTT was replaced with β-ME due to its instability and background heat changes. For each RNA, an RNA-to-buffer

titration was performed and subtracted from the RNA-to-protein data. Data from the first injection point was discarded. To better represent the pool of unspecific RNAs in the transcription reaction, RNA0 was purified as a combined pool of 2–10 nt RNAs from transcription reaction. The concentration of this RNA0 + aborts pool was normalized by the average size of 5.8 nt based on HPLC-ultraviolet-weighted composition of the corresponding transcription mixture (30% full length RNA × 10 nt + 70% aborts × 4 nt average). For RNA0, SMN1 and SMN2 RNAs the ITC data was fitted with the one-site interaction model using MicroCal Origin (Supplementary Fig. 3a–c). A one-site K_D was needed for all RNAs to compare it with the one-site binding model used in the ODE analysis. The standard fitting protocol of MicroCal Origin encountered local minima when fitting the one-site model to the EV2-UP1 ITC data (likely due to the known bi-modal binding of UP1 to this RNA[23]). Therefore for EV2-UP1, the apparent one-site K_D constant was derived as a mean of three different fitting protocols (Supplementary Fig. 3d–g): (1) high-affinity K_D from two-site model in MicroCal Origin, (2) one-site K_D fitted using a stoichiometric 1:1 equilibrium model in Affinimeter software (affinimeter.com) and (3) one-site K_D fitted using a general ligand:target equilibrium binding isotherm[65], which only parameterizes the K_D constant (without considering ΔH) and assumes $N = 1$. All three fitting procedures yielded comparable K_D for EV2-UP1 binding (4.8, 7.1 and 3.3 μM, respectively).

Perturbation experiments. Post-transcriptional perturbations by UP1 protein were performed by adding, 13 h after transcription start, 67.5 nmol of ^{15}N-UP1. Protein was ~1,500 μM concentration, to achieve ≤10% dilution of the mixture. In small-molecule perturbation experiments, the SMN2 ESE1 transcription mixture was spiked co-transcriptionally with 200 μM of one of three molecules (NVS-SM1, smn-C5 and smn-C7 (ref. [26]) in 1% DMSO) or 1% DMSO alone. In these experiments, SMN2 DNA template used an earlier version of the 5′ overhang sequence (5′-GCGCCGUA-3′), before it was optimized at three positions (3,7,8) to reduce its self-complementarity.

Imino signal broadening on UP1 protein binding. Imino signal linewidths could be primarily influenced by the (1) changes in k_{ex} exchange rate of iminos due to unfolding of the stem, (2) k_{ex} changes due to base-flipping and (3) line broadening due to enhanced transverse relaxation (R_2) and B0 field inhomogeneity ($R_{2(B0)}$). Thus, overall linewidth equals $\Delta\nu_{1/2} = (k_{ex,unfolding} + k_{ex,base-flipping} + R_2 + R_{2(B0)})/\pi$. Contribution of B0 field inhomogeneity is considered negligibly small. The R_2 relaxation increase on formation of a 1 to1 RNA–UP1 protein complex was estimated to be $41/\pi$ Hz, given 22.25 kDa UP1 mass, at 303 K, in phosphate buffer saline (0.001 kg m^{-1} s^{-1}), assuming spherical shape of the protein with $r_w = 2.4$ Å (hydration layer) and $R_2 = 5 \times \tau_c$ (correlation time)[68].

Cell culture and nuclear extracts. Extracts were prepared using published procedures[69] from HEK293 cells grown to confluence.

Mass spectrometry for quantification of specific protein concentrations in cells. Pure recombinant uniformly ^{15}N-labeled proteins were spiked into nuclear extracts at a 0.15–0.2 μM concentration. Resulting extracts were reduced, alkylated and digested using trypsin before peptide desalting and purification as previously described[70]. Selected reaction monitoring (SRM) on a triple-quadrupole mass spectrometer was used for targeted proteomic measurements. SRM assays were generated as previously described[71] by selecting the 4–5 most intense transitions from samples with pure ^{15}N-labeled recombinant proteins digested with the same protocol. Sum peak areas of transitions for each peptide were used to calculate the intensity ratio between ^{15}N reference and ^{14}N endogenous peptide signals. The mean and standard deviation of all peptides for each protein were used to find the concentration of endogenous proteins in nuclear extracts.

Selective labeling. Proteins were expressed in minimal M9 medium supplemented with ^{15}N-Val and ^{13}C-Phe, and with all amino acids and nucleosides in unlabeled form (Supplementary Table 3). A limiting amount of unlabeled Phe/Val was added for transaminase suppression. The ^{13}C/^{15}N-labeled amino acids were added only 10 min before induction. Cells were harvested 3–5 h post-induction.

Statistics. Statistical analyses and experiment replicate numbers, where applicable, are described in the corresponding figure legends and method sections. Unless otherwise indicated the derived values and error bars correspond to the mean ± s.d.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Raw NMR data from the main experiments was deposited in Zenodo (10.5281/zenodo.2554066). Source data for the main figures are included in the online article. SRM-MS-Proteomics data was deposited in PeptideAtlas (PASS01365). The ODE network model was deposited in BioModels[54] (MODEL1812270001). RNA-encoding plasmids were deposited in AddGene (ID nos. 126040, 126041, 126042, 126043). Further data and code are available from the corresponding authors upon request.

Code availability
Main code with examples and a protocol for the Systems NMR setup is available at github.com/systemsnmr/ivtnmr.

References
54. Michel, E., Duss, O. & Allain, F. H.-T. in *Bacterial Regulatory RNA: Methods and Protocols* (eds Arluison, V. & Valverde, C.) 177–195 (Springer, 2018).
55. Barraud, P. & Allain, F. H.-T. Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRMs influences the nucleic acid binding topology. *J. Biomol. NMR* **55**, 119–138 (2012).
56. Cléry, A. et al. Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proc. Natl Acad. Sci. USA* **110**, E2802–E2811 (2013).
57. Cléry, A. et al. Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-β1. *Nat. Struct. & Mol. Biol.* **18**, 443 (2011).
58. Moursy, A., Allain, F. H.-T. & Cléry, A. Characterization of the RNA recognition mode of hnRNP G extends its role in SMN2 splicing regulation. *Nucleic Acids Res.* **42**, 6659–6672 (2014).
59. Williamson, M. P. Progress in nuclear magnetic resonance spectroscopy using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).
60. Harris, L. A. et al. BioNetGen 2.2: advances in rule-based modeling. *Bioinformatics* **32**, 3366–3368 (2016).
61. Hyeon, C. & Thirumalai, D. Chain length determines the folding rates of RNA. *Biophys. J.* **102**, L11–L13 (2012).
62. Qin, S., Pang, X. & Zhou, H.-X. Automated prediction of protein association rate constants. *Structure* **19**, 1744–1751 (2011).
63. Chelliah, V. et al. BioModels: ten-year anniversary. *Nucleic Acids Res.* **43**, D542–D548 (2015).
64. Donghi, D., Pechlaner, M., Finazzo, C., Knobloch, B. & Sigel, R. K. O. The structural stabilization of the κ three-way junction by Mg(II) represents the first step in the folding of a group II intron. *Nucleic Acids Res.* **41**, 2489–2504 (2013).
65. Pollard, T. D. A guide to simple and informative binding assays. *Mol. Biol. Cell* **21**, 4061–4067 (2010).
66. Russu, B. I. M. Probing site-specific energetics in proteins and nucleic acids by hydrogen exchange and nuclear magnetic resonance spectroscopy. *Methods* **379**, 152–175 (2004).
67. Rinnenthal, J., Klinkert, B., Narberhaus, F. & Schwalbe, H. Direct observation of the temperature-induced melting process of the *Salmonella* fourU RNA thermometer at base-pair resolution. *Nucleic Acids Res.* **38**, 3834–3847 (2010).
68. Cavanagh et al. *Protein NMR Spectroscopy: Principles and Practice* Ch. 1.4 (Acad. Press, 1996).
69. Kislinger, T. et al. PRISM, a generic large scale proteomic investigation strategy for mammals. *Mol. Cell. Proteomics* **2**, 96–106 (2003).
70. Soste, M. et al. A sentinel protein assay for simultaneously quantifying cellular processes. *Nat. Methods* **11**, 1045–1048 (2014).
71. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4**, 222 (2008).

naturereasearch

Corresponding author(s):   Yaroslav Nikolaev, Frederic Allain

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	TopSpin 3.2 & 4.0.3
Data analysis	TopSpin 3.2 & 4.0.3 BioNetGen 2.17 Rulebender 2.0r382 Matlab R2011a, R2016 Python 2.7.14 ImageJ v2.0.0 Affinimeter v2.1710 Prism v5.0a Cara 1.8.4 Skyline v3.7.0.10940 All important parameters used to analyze the data are described in methods section. Main Python and Matlab scripts implementing data processing and analyses are available at github.com/systemsnmr/ivtnmr, as stated in Methods section.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about <u>availability of data</u>

All manuscripts must include a <u>data availability statement</u>. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw NMR data from the main experiments was deposited in Zenodo (10.5281/zenodo.2554066). Source data for the main figures are included in the online article. SRM-MS-Proteomics data was deposited in PeptideAtlas (PASS01365). The ODE network model was deposited in BioModels (MODEL1812270001). RNA-encoding plasmids were deposited in AddGene (ID #126040, #126041, #126042, #126043). Main code with examples and a protocol for the Systems NMR setup is available at github.com/systemsnmr/ivtnmr. Further data and code are available from the corresponding authors upon request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see <u>nature.com/authors/policies/ReportingSummary-flat.pdf</u>

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Where statistical analyses were required, Systems NMR experiments were performed 2-4 times (biological replicates, varying DNA template batch and/or protein batch). Specific numbers of replicates are stated in corresponding figure captions.
Data exclusions	- To make the time dimension uniform for all transcription-NMR datasets, for ODE model fits only the data below 24 hours was used, even if more time-points were recorded. - In calculations of imino linewidths - the initial points of transcription, before imino signal becomes visible, were not used in quantification of RNA stability because there is no signal to quantify. - In ITC analyses, following the standard procedures, the first injection points, in which the ligand is contaminated with target, were not used in fitting.
Replication	All attempts at replication were successful. The aberrations in 31P integrals due to apparently stochastic MgHPO4 aggregation are discussed in main text.
Randomization	Randomization was not relevant to our study.
Blinding	Blinding was not relevant to our study.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about <u>availability of materials</u>

Obtaining unique materials	RNA-encoding plasmids were deposited in AddGene (ID #126040, #126041, #126042, #126043). Further plasmids (e.g. for protein production) are available upon request.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T (human embryonic kidney) cells were obtained from the European Collection of Cell Cultures (ECACC No. 85120602).
Authentication	Cell lines were not authenticated.
Mycoplasma contamination	Cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	Based on ICLAC register, no commonly misidentified lines were used.