

Published in final edited form as:

Nat Methods. 2019 July 09; 16(8): 743–749. doi:10.1038/s41592-019-0495-7.

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

Yaroslav Nikolaev^{1,*}, Nina Ripin¹, Martin Soste², Paola Picotti², Dagmar Iber³, Frédéric H.-T. Allain^{1,*}

¹Department of Biology, Institute of Molecular Biology & Biophysics, ETH Zürich, Zürich, Switzerland ²Department of Biology, Institute of Biochemistry, ETH Zürich, Zürich, Switzerland

³Department of Biosystems Science and Engineering, ETH Zürich, Zürich, Switzerland

Abstract

Cellular behavior is controlled by the interplay of diverse biomolecules. Most experimental methods, however, can monitor only 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, RNA) and all elementary reaction types (bimolecular interactions, catalysis, unimolecular changes). Focusing on an 8-reaction co-transcriptional RNA folding network, in a single sample we recorded over 35 time-points with over 170 observables each, and accurately determined 5 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”.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence should be addressed to yaroslav.v.nikolaev@gmail.com or allain@mol.biol.ethz.ch.

Reporting Summary

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

Code availability

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

Data availability

Raw NMR data from the main experiments was deposited in Zenodo ([10.5281/zenodo.2554066](https://doi.org/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 ⁶³ (MODEL1812270001). RNA-encoding plasmids were deposited in AddGene (ID #126040, #126041, #126042, #126043). Further data and code are available from the corresponding authors upon request.

Author Contributions

Y.N. conceived the idea, obtained initial funding, performed experiments, data analysis and network modeling. F.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 Financial Interests

The authors declare no competing financial interests.

Introduction

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 monitor only a single molecule class or reaction type at a time (catalysis, bimolecular interactions, unimolecular state changes), limiting our ability to measure complex cellular dynamics.

Furthermore, studies of networks often face a choice between biochemical methods – measuring dynamic (e.g. 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 sample¹. 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 understand the underlying logic of these networks⁵.

To address the above challenges, we sought to devise a Nuclear Magnetic Resonance spectroscopy (NMR)-based approach which would (i) allow to monitor "heterotypic" networks and pathways – involving different molecule or reaction types – entirely, in a single in vitro sample; and (ii) provide quantitative dynamic data for modeling of the network mechanisms. With certain limitations on molecule size and concentration (≤ 50 -100 kDa, ≥ 10 -50 μ 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 cells⁸. The use of NMR to monitor reactions is common in chemistry⁹, 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, i.e. metabolites^{10–14}, proteins^{13,15,16}, 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 which 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 (i) 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; (ii) to establish a mathematical model explaining our observations; and (iii) 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 the RNA transcription. We termed this methodological approach "Systems NMR" – a potential generic name for NMR-driven reconstruction of biomolecular reaction networks.

Monitoring a co-transcriptional RNA folding network

We first sought to investigate if 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, SMN2 – two hairpins present in exon 7 of human *SMN1* and *SMN2* genes respectively²², and the stem loop II (EV2) of the IRES of Enterovirus 71²³ (Fig. 1b). To distinguish the RNA-specific UP1 perturbations from other changes in the network dynamics (pH, nucleotide tri-phosphate 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 tri-phosphates (NTPs), MgCl₂, 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 hours by repeating several NMR experiments (Fig. 2b,c,d and Supplementary Video 1): 1D (one-dimensional) ³¹P – to monitor the levels of metabolites and RNA, 1D ¹H – to monitor RNA folding, and 2D (two-dimensional) ¹H-¹⁵N – to monitor protein interactions. Each set of measurements takes ~30 minutes 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 ³¹P spectrum (Fig. 2b), more than 20 in the ¹H spectrum (Fig. 2c), and over 150 protein backbone amide signals in the 2D ¹H-¹⁵N spectrum (Fig. 2d). For quantitative modelling of the target network 10 signals were used (Supplementary Table 1): the ³¹P signals of inorganic phosphate (PO₄ – referring to PO₄³⁻ and its protonated forms), RNA phosphate, α, β and γ-phosphate of the NTP, α and β-phosphate of NDP (Fig. 2b,e); the ¹H imino signals of RNA uracils U5 (SMN1 and SMN2) or U4 (EV2) (Fig. 2f); and the ¹H-¹⁵N signals of the selected UP1 residues reporting on RNA binding – His33 and Arg75 (Fig. 2g). In this study, these 10 signals were sufficient to quantify the key parameters of the target 8-reaction network. The data from remaining signals can still be used in subsequent studies to investigate the system in more detail. E.g. to analyse individual conversion rates of four NTPs, or to analyse 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 little or no calibrations. To quantify the metabolite and RNA concentrations, we measured the integrals of corresponding signals in ³¹P 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, Ura4, 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-¹⁵N spectra between the free and bound protein states as the more RNA was bound (His33, Arg75, Fig. 2j).

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

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 3 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 PO_4 at the end of the 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 model with the relevant reactions (Fig. 1a, reactions 4-8) allowed to quantify the MgHPO_4 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 5 out of the 8 network reactions (# 1-3, 7-8, Fig. 1a), while the constants of the other reactions (# 4,5,6) 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 – k_{cat} , G , K_D , $K_{\text{eq,MgHPO}_4}$ – were compared with the constants derived by classical approaches, when a single reaction is perturbed at a time. Remarkably, all tested constants were in agreement with classical methods.

The equilibrium constant for the formation of soluble MgHPO_4 aggregates ($K_{\text{eq,MgHPO}_4}$) showed an average value of 1.31 ± 0.06 mM, closely matching the 0.97 ± 0.05 mM value reported in the literature²⁴ (Fig. 3a).

The expected catalytic rate constant $k_{\text{cat}} = 0.26 \pm 0.07$ nt s^{-1} of T7 RNA Polymerase²⁵ 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 2-fold

increase in the overall $k_{\text{cat}} = 0.73 \pm 0.06 \text{ nt s}^{-1}$ for the longer EV2 RNA (Fig. 3b), when the polymerase spends more time in the faster elongation phase.

Based on the measured UV-melting experiments (Supplementary Note 1), the free energy (ΔG) of folding of the two RNA hairpins (SMN1 and SMN2, differing by single base-pair) are expected to be -4.64 ± 0.15 and -6.02 ± 0.37 kcal/mol (Fig. 3c). Systems NMR measured -5.2 ± 0.1 and -5.6 ± 0.1 kcal/mol for the respective constants (Fig. 3c). SMN2 stabilities thus matched within the standard deviation limits, while SMN1 stability was overestimated by ~ 0.3 kcal/mol in NMR compared to UV 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 UV-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 $1,391 \pm 331 \mu\text{M}$ for RNA0, $51.3 \pm 2.5 \mu\text{M}$ for SMN1, $47.4 \pm 19.7 \mu\text{M}$ for SMN2 and $5.1 \pm 1.9 \mu\text{M}$ 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 $11.1 \pm 5.1 \mu\text{M}$, 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 $1,391 \mu\text{M}$ 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 $K_{D, \text{UP1-NTPs}}$ of $16,200 \pm 2,100 \mu\text{M}$ (Supplementary Fig. 4). RNA aborts increase from 0 to $\sim 2,000 \mu\text{M}$, and NTPs decrease from 20,000 to $5,000 \mu\text{M}$ 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 candidates molecules, recently developed to correct the aberrant splicing of exon 7 from SMN2 gene²⁶. 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- β 1²⁷. 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 ¹⁵N-valine-labeled protein constructs mixed together (two independent RRM in 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 bimodal response (Fig. 4e) – with valine signals first decreasing and then increasing their intensity, many without significant change in the signal positions. The samples showed evidence of Liquid-Liquid-Phase-Separation²⁸, 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 10-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 driven here not simply by high protein concentrations, but involves interactions between specific proteins.

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 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- β 1 proteins in HEK293 cells, and found those to be 30, 7.6, 3.3 and 5.2 μ M respectively (Supplementary Fig. 7). This is close to the current NMR sensitivity limits (10-50 μ M), suggesting that our setup, at least for hnRNP A1, can be tuned for observations under near-physiological conditions.

Discussion

The derivation of individual catalytic³⁰, unimolecular³¹ and bimolecular³² reaction constants by NMR is not uncommon, but Systems NMR approach enables us to quantify a network with all 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 screens³³ and suggest that *in vivo* UP1/hnRNP A1 protein affinity to specific RNA targets will likely be ~1000-fold weaker than the nM-range affinities anticipated from single-reaction *in vitro* assays²³.

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 organelles²⁹, an emerging research area with connections to various age-related disorders²⁸. 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 nucleotide-long 5' overhang RNA sequence which 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 achieved³⁶, and in combination with hyperpolarization methods, 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 days³⁹. 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 in free energy difference). For bimolecular interactions, NMR currently permits direct quantification of dissociation constants in the low- μM to medium-mM range³². 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 ~7000 unique protein and ~600 unique RNA NMR structures are 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 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, like phase-separation of RNA binding domains observed here.

Considering specific applications, Systems NMR can give strong advance to the studies of "heterotypic" networks – involving different molecule and/or reaction types. For example, concurrent quantification of perturbations in different parts of a biochemical network like 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 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 (doi: [10.21203/rs.2.9160/v1](https://doi.org/10.21203/rs.2.9160/v1)), and the most recent version 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 – which are recognized specifically by UP1 protein; contains 30% purines – to reduce RNA Polymerase dissociation/abortion at initiation stage; does not form stable dimers or hairpins with itself or target SMN and EV2 RNA sequences. This resulted in 7 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 vector⁵⁴ at SapI sites, using dsDNAs from commercial (Microsynth AG) single-stranded oligos: RNA0 (ATAGCACCACACG, TCACGTGTGGTGC), SMN1 (ATAGCACCACACGGGTTTCAGACAAAATCCG, TCACGGATTTTGTCTGAAACCCGTGTGGTGC), SMN2 (ATAGCACCACACGGGTTTTAGACAAAATCCG, TCACGGATTTTGTCTAAAACCCGTGTGGTGC), EV2 (ATAGCACCACAGGATCAATAGCAGGTGTGGCACACCAGTCATACCTTGATCC, TCAGGATCAAGGTATGACTGGTGTGCCACACCTGCTATTGATCCTGTGGTGC). Plasmids were purified using Nucleobond Xtra Midi kit (Macherey Nagel), final pellets washed three times with 70% ethanol, dried and linearized by BsaI (NEB) enzyme for 15 hours at 50°C in NEB3.1 buffer.

Proteins

All constructs and purification procedures were described earlier: UP1⁵⁵, SRSF1-RRM1 and RRM2⁵⁶, Tra2- β 1⁵⁷, hnRNP G⁵⁸. 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, 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/ml Inorganic Pyrophosphatase from baker's yeast (Sigma), 5% D₂O, 50 μ M 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS), 280 nM T7 RNA polymerase and 33 nM DNA template. Proteins were 150 μ M – in single-protein reactions, and 83 μ M of each

protein – 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 likely also 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 μ M concentration in 25 mM Tris-HCl pH 8, 50 mM NaCl, 0.5% β -mercaptoethanol (β -ME), 50% w/v glycerol at –20°C.

NMR experiments

Experiments were measured on Bruker AVIII-600 MHz with CPQCI cryoprobe, and consisted of repeating series of 1D 1 H-watergate (spectral width (SW) 22 ppm; acquisition time (AQ) 0.62 s; D1 (interscan delay) 1 s; number of scans (NS) 128), 2D 1 H-TOCSY (SW 10 / 9 ppm; AQ 0.17 / 0.018 s; D1 0.5 s; NS 4), 1D 31 P (SW 50 ppm; AQ 0.66 s; NS 256; carrier –8.22 ppm; D1 0.8 s), 1D 1 H-SOFAST (SW 24 ppm; AQ 0.053 s; D1 0.1 s; NS 1536; 1 H excitation with Pc9 pulse, 5 ppm wide, centered at 12.9 ppm) and 2D 1 H 15 N-SOFAST-HMQC (H/N: SW 16 / 23.5 ppm; AQ 0.106 / 0.035 s; D1 0.2 s; NS 24; 15 N carrier 117.8; 1 H 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 Frank 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; 15 N carrier 117.8; 1 H 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 (Nyberg N., RBNMR, MATLAB Central File Exchange #40332, (2013)). Chemical shifts of protein residues in 2D HN spectra were traced in CARA (cara.nmr.ch).

NMR observables

For final network modelling 10 signals were used: 31 P spectra – 1) PO₄, 2) RNA, 3-5) α NTP, β NTP, γ NTP, 6-7) α NDP, β NDP; 8) 1 H 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 31 P integrals using α NTP integral at time=0 as 20 mM internal calibration. Each 31 P integral was T₁-relaxation-weighted using the ratios of corresponding integrals measured in the reference 31 P spectra with 30 s and 0.8 s interscan delay at the end of the transcription reaction, i.e. $[I_{31P, corrected} = I_{31P} \times (I_{31P, ref, d1=30s} / I_{31P, ref, d1=0.8s})]$. NTP and NDP populations were quantified from α P integrals, MgHPO₄ was calculated from decrease of the total T₁-weighted 31 P integral of all species. For long RNAs (> \approx 20 nt), the 1D 31 P 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 UV(260 nm)-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. Firstly, 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 are not directly interacting with the RNA, which explains how they could sense the same epitope changes independent of the RNA sequence. Secondly, these signals were sensitive to the differences in affinity of the RNA binding, as the magnitude of the ^1H - ^{15}N chemical shift perturbation varied for four different protein:RNA complexes (Fig. 2g,j). In assays with $150\ \mu\text{M}$ UP1 concentration, both signals appeared predominantly in fast exchange with respect to the NMR time scale, and so the fast exchange assumption⁵⁹ was used during modeling. The experimental HN CSPs were calculated using $\Delta\text{HN} = \sqrt{(\Delta\text{H}^2 + (\Delta\text{N} \times 0.2)^2)/2}$. The effect of pH change on histidine signal position was assumed negligible because the transcription buffer pH = 7.7 is far from the histidine pKa ≈ 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 $150\ \mu\text{M}$ UP1 protein data did not approach this saturation under the assay conditions, because of the low protein:RNA ratio, since the final concentrations of specific RNAs reached only $\sim 120\ \mu\text{M}$, giving only $\sim 0.8:1$ 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 $30\ \mu\text{M}$ UP1 and the "high-affinity" EV2 RNA. This allowed to increase protein saturation by reaching $\sim 6:1$ and $4:1$ RNA:protein ratios ($\sim 120:20$ and $120:30\ \mu\text{M}$ respectively) at the end of the reaction. Due to the poor NMR sensitivity at 20 - $30\ \mu\text{M}$ protein concentrations, the 2D HN spectra in these experiments required ~ 10 hour 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 ($150\ \mu\text{M}$) and low (20 - $30\ \mu\text{M}$) protein concentrations gave an imperfect fit with the single-site binding model, suggesting that UP1 is already close to saturation in the assays with $150\ \mu\text{M}$ 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 - $30\ \mu\text{M}$ 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 $150\ \mu\text{M}$ UP1) were taken as the shifts giving best fit when simultaneously fitting the data from EV2 RNA datasets with 150 , 30 and $20\ \mu\text{M}$ UP1 protein concentration (Supplementary Fig. 8).

All above ^{31}P and HN data was used for global parameterization of the ODE model, and ^1H -imino signals were used for lineshape analysis and RNA G derivation.

Mathematical modeling

The model was built using the BioNetGen language⁶⁰ and resulted in 8 rate equations and 9 differential equations:

Rate equations

$$v_1(\text{synthesis of RNA}) = k_{cat, RNA} \cdot NTP$$

$$v_{2F, 2R}(\text{RNA folding}) = \text{analytical solution, see below} \quad (1)$$

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

$$v_{3R}(\text{dissociation of protein} \bullet \text{RNA complex}) = k_{off} \cdot [\text{protein} \bullet \text{RNA}] \quad (3)$$

$$v_4(\text{synthesis of aborts}) = k_{cat, Aborts} \cdot NTP \quad (4)$$

$$v_5(\text{hydrolysis of pyrophosphate}) = k_{catPPi} \cdot PPi$$

$$v_{6F, 6R}(\text{formation of Mg} \bullet \text{NTP complexes}) = \text{fixed constant, see below} \quad (5)$$

$$v_{7F}(\text{formation of MgHPO4 salt}) = k_{precip} \cdot PO4 \quad (6)$$

$$v_{7R}(\text{dissociation of MgHPO4 salt}) = k_{dissolve} \cdot \text{MgHPO4} \quad (7)$$

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

Differential equations

$$\frac{dNTP}{dt} = -v_1 - v_4 - v_8 \quad (9)$$

$$\frac{dRNA}{dt} = + \frac{v_1}{rna\ length} - v_{3F} + v_{3R} \quad (10)$$

$$\frac{dPO4}{dt} = + 2 \cdot v_5 - v_{7F} + v_{7R} + v_8 \quad (11)$$

$$\frac{dPPi}{dt} = + v_1 \cdot \frac{rna\ length - 1}{rna\ length} + v_4 \cdot \frac{aborts\ length - 1}{aborts\ length} - v_5 \quad (12)$$

$$\frac{dMgHPO4}{dt} = + v_{7F} - v_{7R} \quad (13)$$

$$\frac{dNDP}{dt} = +v_8 \quad (14)$$

$$\frac{dAborts}{dt} = +\frac{v_4}{aborts\ length} \quad (15)$$

$$\frac{dProtein}{dt} = -v_{3F} + v_{3R} \quad (16)$$

$$\frac{d[RNA.Protein]}{dt} = +v_{3F} - v_{3R} \quad (17)$$

At the given hairpin stability (≈ -5.5 kcal/mol) the SMN and EV2 RNAs are predominantly in a single state (99,99% folded). The expected rate of hairpin folding ($v_{2F} = 8.51 * 10^3 \text{ s}^{-1}$ for ~ 30 nt hairpin)⁶¹ is three orders of magnitude faster than protein-RNA encounters at the unbiased diffusion rate ($\sim 10^5 \text{ M}^{-1} \text{ s}^{-1}$)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 lineshape analysis at each reaction time point (see below). Because of the sequence identity with 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: $0.3 * 10 + (1-0.3) * 4 = 5.8$ nt. The final ODE model was deposited in BioModels⁶³ (MODEL1812270001).

Model fitting

The ODE model was fitted to the seven 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 code⁴⁷. 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 enzyme (280 nM T7 RNAP) 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).

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

Because reactions # 3 and #7 (Fig. 1a, protein binding and aggregation of MgHPO₄) equilibrate faster than the time-resolution of the current setup, only the equilibrium constants (K_{eq} , K_D – ratios of kinetic rates) for these reactions could be parameterized.

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

Pyrophosphate hydrolysis can be quantified from the specific 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).

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

The formation of soluble Mg-PO₄ aggregates depends on the concentration of the free Mg. Apart from Mg-PO₄, the free Mg can also participate in Mg-NTP and Mg-RNA complexes. The fraction of the free Mg was estimated from the published $K_{D, Mg.NTP} = 0.3 \text{ mM}$ ⁶⁴, using the general bimolecular binding isotherm⁶⁷ denoted with f in the equation below. For the calculation the concentration of free Mg available for Mg-PO₄ salt formation was assumed constant during the reaction.

$$M_{g, free} = f(Mg_{total}, NTP_{total}, K_{D, Mg.NTP}) = f(24, 20, 0.3) = 5.1 \text{ mM} \quad (21)$$

$$K_{eq, MgHPO4} = \frac{k_{precip}}{k_{dissolve}} \cdot \frac{1}{Mg_{free}} \quad (22)$$

RNA folding G from imino signal lineshape analysis

For fitting, 1D ¹H-SOFAST spectra were Fourier-transformed with no apodization function. Imino signals were fit to a single-lorentzian function using the lorentzfit routine (Wells J., Lorentzfit, MATLAB Central File Exchange #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 ($k_{ex, unfolding}$).

$$k_{ex} = linewidth \cdot \pi \quad (23)$$

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

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

Imino linewidths depended on pH (concentration of imino-exchange catalyst), indicating that the system is under the bimolecular exchange regime (“EX2”)⁶⁵, and hence the measured k_{ex} reports on the equilibrium constant of RNA unfolding/opening ($K_{eq,unfolding} = K_{op}$). The intrinsic exchange rate $k_{ex,intrinsic}$ (same as the exchange from the open state, $k_{ex,open}$) in the transcription buffer was measured to be $\sim 10^6 \text{ s}^{-1}$ for both UTP and GTP, using a protocol described elsewhere⁶⁶. The final free energy of folding was determined using:

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

$$\Delta G_{folding} = - \Delta G_{unfolding} = R \cdot T \cdot \log(K_{eq,unfolding}) \quad (26)$$

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

RNA purification

RNAs for ITC and UV-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.

UV temperature-melting

For the melting experiments the RNA hairpins were produced without the single-stranded 5' overhang to eliminate UV 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 , 25 mM L-Arg/L-Glu buffer. Details of the analysis 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 , 303K. 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 nts based on HPLC-UV-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 RNA²³). 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 stoichiometric 1:1 equilibrium model in Affinimeter software (affinimeter.com), and (3) one-site K_D fitted using general ligand:target equilibrium binding isotherm⁶⁷, which parameterizes only 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).

Perturbation experiments

Post-transcriptional perturbations by UP1 protein were performed by adding, 13 hours after transcription start, 67.5 nmol of ¹⁵N-UP1. Protein was ~1500 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-C7²⁶ 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 upon 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 $\nu_{1/2} = (k_{ex, unfolding} + k_{ex, base-flipping} + R_2 + R_{2(B0)})/\pi$. Contribution of B0 field inhomogeneity is considered negligibly small. The R_2 relaxation increase upon formation of a 1:1 RNA:UP1 protein complex was estimated to be $41/\pi$ Hz, given 22.25 kDa UP1 mass, at 303K, in phosphate buffer saline ($0.001 \text{ kg m}^{-1} \text{ s}^{-1}$), assuming spherical shape of the protein with $r_w=2.4 \text{ \AA}$ hydration layer, and $R_2 = 5*\tau_c$ (correlation time)⁶⁹.

Cell culture and nuclear extracts

Extracts were prepared using published procedures⁶⁸ 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 0.15-0.2 μM concentration. Resulting extracts were reduced, alkylated and digested using trypsin prior to 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. $^{13}\text{C}/^{15}\text{N}$ -labeled amino acids were added only 10 minutes before induction. Cells were harvested 3-5 hours 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 \pm s.d.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank J. Vollmer and G. Fengos for the help with network modeling. We acknowledge G. Wider and all members of the ETH BNSP platform for excellent maintenance of the NMR infrastructure, S. Heinrich for the help with microscopy, and E. Lehmann for providing T7 polymerase. We thank all members of the Allain Lab, in particular F. Damberger, and the Parpan retreat participants for helpful discussions. This work was supported by the Promedica Stiftung, Chur (Grant 1300/M to Y.N.), Novartis Foundation and Krebsliga Zurich (Y.N.), NCCR RNA and Disease by the Swiss National Science Foundation (F.A.).

References

1. Szallasi, Z. System Modeling in Cellular Biology Ch. 10. The MIT Press; Cambridge: 2006.
2. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet.* 2004; 5:101–113. [PubMed: 14735121]
3. Howsmon DP, Hahn J. Regularization Techniques to Overcome Over-Parameterization of Complex Biochemical Reaction Networks. *IEEE Life Sci Lett.* 2016; 2:31–34. [PubMed: 29104899]
4. Yugi K, Kuroda S. Metabolism as a signal generator across trans-omic networks at distinct time scales. *Curr Opin Syst Biol.* 2017; 8:59–66.
5. Hart Y, Alon U. Review The Utility of Paradoxical Components in Biological Circuits. *Mol Cell.* 2013; 49:213–221. [PubMed: 23352242]

6. Frueh DP, Goodrich AC, Mishra SH, Nichols SR. NMR methods for structural studies of large monomeric and multimeric proteins. *Curr Opin Struct Biol.* 2013; 23:734–739. [PubMed: 23850141]
7. Kwan AH, Mobli M, Gooley PR, King GF, Mackay JP. Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS J.* 2011; 278:687–703. [PubMed: 21214860]
8. Freedberg DI, Selenko P. Live cell NMR. *Annu Rev Biophys.* 2014; 43:171–192. [PubMed: 24895852]
9. Bernstein MA. Reaction monitoring using NMR. *Magn Reson Chem.* 2016; 54:422. [PubMed: 27038257]
10. Wolak J, et al. Chapter 11 Noninvasive Fluxomics in Mammals by Nuclear Magnetic Resonance Spectroscopy. *Handb Metabolomics.* 2012; :321–392. DOI: 10.1007/978-1-61779-618-0
11. Eicher JJ, Snoep JL, Rohwer JM. Determining Enzyme Kinetics for Systems Biology with Nuclear Magnetic Resonance Spectroscopy. *Metabolites.* 2012; 2:818–843. [PubMed: 24957764]
12. Lerche MH, Jensen PR, Karlsson M, Meier S. NMR Insights into the Inner Workings of Living Cells. *Anal Chem.* 2015; 87:119–132. [PubMed: 25084065]
13. Smith MJ, et al. Real-time NMR monitoring of biological activities in complex physiological environments. *Curr Opin Struct Biol.* 2015; 32:39–47. [PubMed: 25727665]
14. Fan TW, Lane AN. Applications of NMR spectroscopy to systems biochemistry. *Prog Nucl Magn Reson Spectrosc.* 2016; 92–93:18–53.
15. Cordier F, et al. Ordered Phosphorylation Events in Two Independent Cascades of the PTEN C-tail Revealed by NMR. *J Am Chem Soc.* 2012; 134:20533–20543. [PubMed: 23171049]
16. Theillet F-X, et al. Cell signaling, post-translational protein modifications and NMR spectroscopy. *J Biomol NMR.* 2012; 54:217–236. [PubMed: 23011410]
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.* 2010; 107:9192–9197. [PubMed: 20439766]
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.* 2012; 9:672–680. [PubMed: 22647526]
19. Steinert H, et al. Pausing guides RNA folding to populate transiently stable RNA structures for riboswitch-based transcription regulation. *Elife.* 2017; 6:1–18.
20. Breaker RR. Prospects for Riboswitch Discovery and Analysis. *Mol Cell.* 2011; 43:867–879. [PubMed: 21925376]
21. Zhang J, Landick RA. Two-Way Street : Regulatory Interplay between RNA Polymerase and Nascent RNA Structure. *Trends Biochem Sci.* 2016; 41:293–310. [PubMed: 26822487]
22. Singh RN. Evolving concepts on human SMN pre-mRNA splicing. *RNA Biol.* 2007; 4:7–10. [PubMed: 17592254]
23. Levensgood JD, 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.* 2013; 10:1136–1145. [PubMed: 23727900]
24. Racz GJ, Soper RJ. Solubility of Dimagnesium Phosphate Trihydrate and Trimagnesium Phosphate. *Can J Soil Sci.* 1968; 48:265–269.
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.* 1997; 36:13718–13728. [PubMed: 9354643]
26. Sivaramakrishnan M, et al. Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers. *Nat Commun.* 2017; 8:1476. [PubMed: 29133793]
27. Singh NN, 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.* 2013; 41:8144–8165. [PubMed: 23861442]
28. Boeynaems S, et al. Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* 2018; 28:420–435. [PubMed: 29602697]
29. Maharana S, et al. RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science (80-.).* 2018; 360:918–921.

30. Bowen S, Hilty C. Time-resolved dynamic nuclear polarization enhanced NMR spectroscopy. *Angew Chem Int Ed Engl.* 2008; 47:5235–5237. [PubMed: 18512824]
31. Bothe JR, et al. Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *Nat Methods.* 2011; 8:919–931. [PubMed: 22036746]
32. Fielding L. NMR methods for the determination of protein-ligand dissociation constants. *Prog Nucl Magn Reson Spectrosc.* 2007; 51:219–242.
33. Jain N, Lin H-C, Morgan CE, Harris ME, Tolbert BS. Rules of RNA specificity of hnRNP A1 revealed by global and quantitative analysis of its affinity distribution. *Proc Natl Acad Sci.* 2017; 114:2206–2211. [PubMed: 28193894]
34. Takeuchi K, Wagner G. NMR studies of protein interactions. *Curr Opin Struct Biol.* 2006; 16:109–117. [PubMed: 16427776]
35. Bieri M, et al. Macromolecular NMR spectroscopy for the non-spectroscopist: beyond macromolecular solution structure determination. *FEBS J.* 2011; 278:704–715. [PubMed: 21214861]
36. Wiesner S, Sprangers R. Methyl groups as NMR probes for biomolecular interactions. *Curr Opin Struct Biol.* 2015; 35:60–67. [PubMed: 26407236]
37. Frydman L, Blazina D. Ultrafast two-dimensional nuclear magnetic resonance spectroscopy of hyperpolarized solutions. *Nat Phys.* 2007; 3:415–419.
38. Kim Y, Hilty C. Affinity Screening Using Competitive Binding with Fluorine-19 Hyperpolarized Ligands. *Angew Chemie Int Ed.* 2015; 54:1–5.
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.* 2007; 129:1372–1377. [PubMed: 17263421]
40. Baldwin AJ, Kay LE. NMR spectroscopy brings invisible protein states into focus. *Nat Chem Biol.* 2009; 5:808–814. [PubMed: 19841630]
41. Dalvit C, Fagerness PE, Hadden DTa, Sarver RW, Stockman BJ. Fluorine-NMR experiments for high-throughput screening: Theoretical aspects, practical considerations, and range of applicability. *J Am Chem Soc.* 2003; 125:7696–7703. [PubMed: 12812511]
42. Thongwichian R, et al. A Multiplexed NMR-Reporter Approach to Measure Cellular Kinase and Phosphatase Activities in Real-Time. *J Am Chem Soc.* 2015; 137:6468–6471. [PubMed: 25963544]
43. Soininen P, et al. High-throughput serum NMR metabolomics for cost-effective holistic studies on systemic metabolism. *Analyst.* 2009; 134:1781–1785. [PubMed: 19684899]
44. Lindon JC, London IC, Kensington S. Biofluids Studied by NMR Spectroscopy. *Encycl Spectrosc Spectrom.* 2010; 1:128–141.
45. Lu K, Miyazaki Y, Summers MF. Isotope labeling strategies for NMR studies of RNA. *J Biomol NMR.* 2010; 46:113–125. [PubMed: 19789981]
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.* 2014; 6:13–36. [PubMed: 24123887]
47. Geier, F, Fengos, G, Felizzi, F, Iber, D. Computational Modeling of Signaling Networks. Liu, X, Betterton, MD, editors. Humana Press; 2012. 23–39.
48. Maiwald T, Timmer J. Dynamical modeling and multi-experiment fitting with PottersWheel. *Bioinformatics.* 2008; 24:2037–2043. [PubMed: 18614583]
49. Raue A, et al. Data2Dynamics: A modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics.* 2015; 31:3558–3560. [PubMed: 26142188]
50. Vanlier J, Tiemann CA, Hilbers PAJ, van Riel NAW. Parameter uncertainty in biochemical models described by ordinary differential equations. *Math Biosci.* 2013; 246:305–314. [PubMed: 23535194]
51. Chen Z, Rappert S, Zeng A. Rational Design of Allosteric Regulation of Homoserine Dehydrogenase by a Nonnatural Inhibitor. *ACS Synth Biol.* 2015; 4:126–131. [PubMed: 24344690]

52. Smith MJ, Ikura M. Integrated RAS signaling defined by parallel NMR detection of effectors and regulators. *Nat Chem Biol.* 2014; 10:223–230. [PubMed: 24441586]
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.* 2015; 1850:620–627.
54. Michel, E, Duss, O, Allain, FH-T. *Bacterial Regulatory RNA: Methods and Protocols.* Arluisson, V, Valverde, C, editors. Springer; New York: 2018. 177–195.
55. Barraud P, Allain FH-T. Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRM motifs influences the nucleic acid binding topology. *J Biomol NMR.* 2012; 55:119–138. [PubMed: 23247503]
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 U S A.* 2013; 110:E2802–11. [PubMed: 23836656]
57. Cléry A, et al. Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2- β 1. *Nat Struct & Mol Biol.* 2011; 18:443. [PubMed: 21399644]
58. Moursy A, Allain FH-T, Cléry A. Characterization of the RNA recognition mode of hnRNP G extends its role in SMN2 splicing regulation. *Nucleic Acids Res.* 2014; 42:6659–6672. [PubMed: 24692659]
59. Williamson MP. Progress in Nuclear Magnetic Resonance Spectroscopy Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc.* 2013; 73:1–16. [PubMed: 23962882]
60. Harris LA, et al. BioNetGen 2.2: advances in rule-based modeling. *Bioinformatics.* 2016; 32:3366–3368. [PubMed: 27402907]
61. Hyeon C, Thirumalai D. Chain length determines the folding rates of RNA. *Biophys J.* 2012; 102:L11–L13. [PubMed: 22325296]
62. Qin S, Pang X, Zhou H-X. Automated prediction of protein association rate constants. *Structure.* 2011; 19:1744–1751. [PubMed: 22153497]
63. Chelliah V, et al. BioModels: Ten-year anniversary. *Nucleic Acids Res.* 2015; 43:D542–D548. [PubMed: 25414348]
64. Donghi D, Pechlaner M, Finazzo C, Knobloch B, Sigel RKO. 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.* 2013; 41:2489–2504. [PubMed: 23275550]
65. Russu BIM. Probing Site-Specific Energetics in Proteins and Nucleic Acids by Hydrogen Exchange and Nuclear Magnetic Resonance Spectroscopy. *Methods.* 2004; 379:152–175.
66. 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.* 2010; 38:3834–3847. [PubMed: 20211842]
67. Pollard TD. A Guide to Simple and Informative Binding Assays. *Mol Biol Cell.* 2010; 21:4061–4067. [PubMed: 21115850]
68. Kislinger T, et al. PRISM, a generic large scale proteomic investigation strategy for mammals. *Mol Cell Proteomics.* 2003; 2:96–106. [PubMed: 12644571]
69. Cavanagh, et al. *Protein NMR Spectroscopy: Principles and Practice.* Ch 1.4. New York: Academic Press; 1996.
70. Soste M, et al. A sentinel protein assay for simultaneously quantifying cellular processes. *Nat Methods.* 2014; 11:1045–1048. [PubMed: 25194849]
71. Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol.* 2008; 4:222. [PubMed: 18854821]

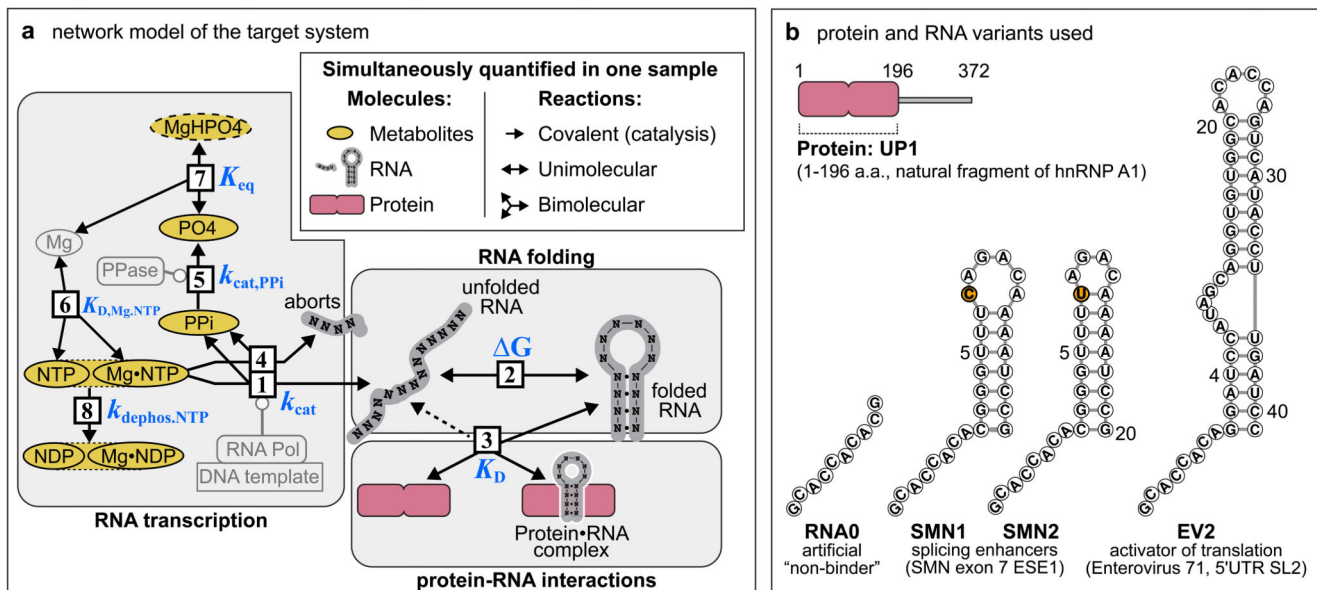


Figure 1. Molecules and reactions of the target co-transcriptional network.

(a) Organization of reactions in the reconstructed network: [1] RNA transcription, [2] RNA folding, [3] binding of regulator protein to RNA (dashed line indicates that protein can also bind to the unfolded RNA), [4] formation of RNA aborts, [5] hydrolysis of pyrophosphate (PPi), [6] formation of Mg NTP (nucleotide triphosphate) complexes, [7] formation of MgHPO₄ salt aggregates, [8] dephosphorylation of NTPs to NDPs (nucleotide diphosphates). merged ovals indicate that NMR observables of free and Mg-bound nucleotide-phosphates, e.g. NTP and MgNTP, cannot be separately discriminated). PO₄ refers ambiguously to multiple states of phosphate ion, including PO₄³⁻ and its protonated forms. Indirectly observed MgHPO₄ species is shown with dashed contour. Non-observable species – Mg, pyrophosphatase (PPase), RNA Polymerase (abbreviated with RNA Pol), DNA template – are shaded. Metabolites, proteins and RNA are shown in ochre, rose and grey respectively.

(b) Protein and RNA variants used in Systems NMR experiments. To make the products of the abortive RNA transcription uniform in all constructs (panel (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. Abbreviations, ESE1 – exonic splicing enhancer 1, UTR - untranslated region, SL2 – stem-loop 2.

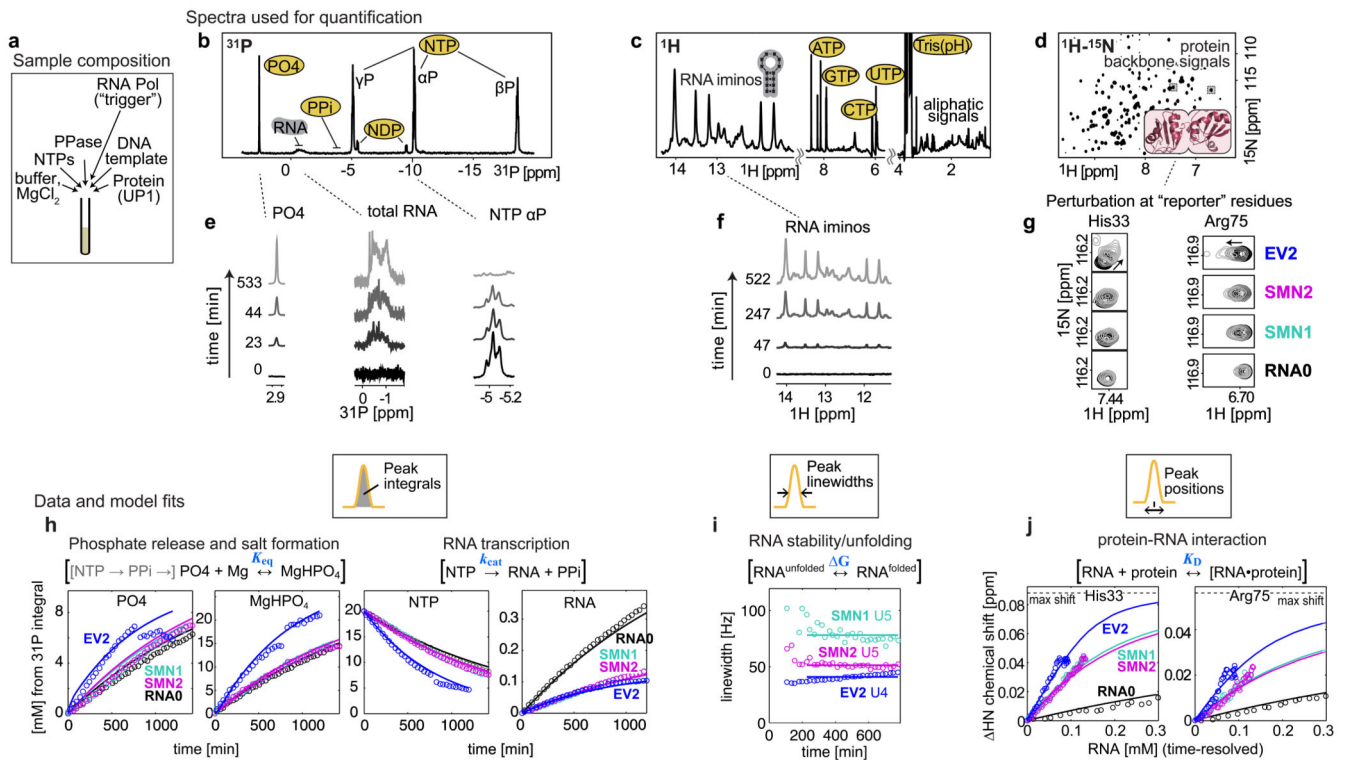


Figure 2. NMR observables, data and mathematical model fits used for network parameterization.

(a) Sample composition, 450 μ L NMR tube with 150 μ M 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,c,d) Full spectra recorded in the NMR assay. (e,f,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. 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,i,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 main text. (i) Linewidths of U5 imino signals in SMN RNAs and U4 in EV2. (j) Chemical shift perturbations 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.

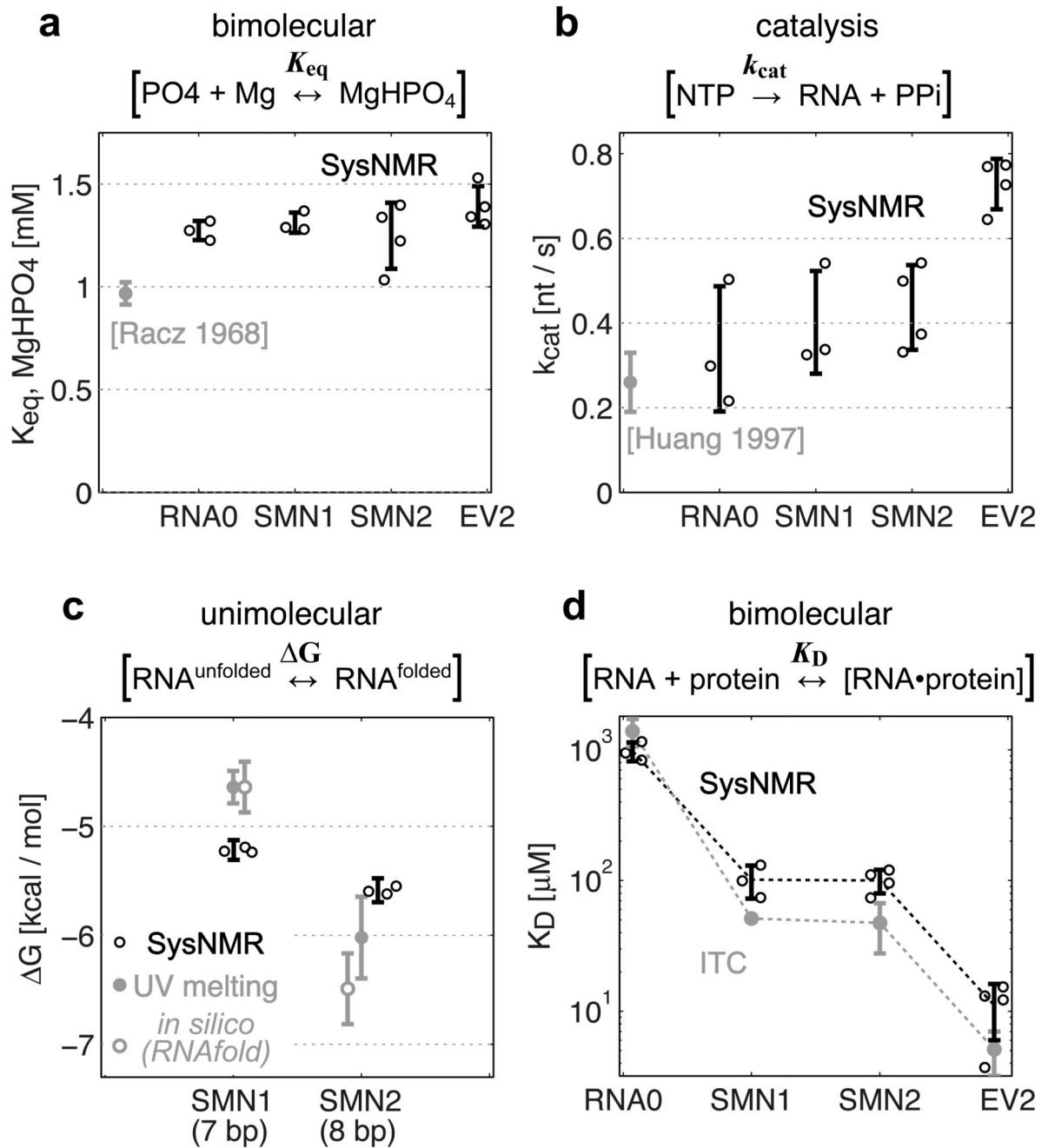


Figure 3. Validation of Systems-NMR-derived constants.

Systems NMR data shown in black, validation data in grey. (a) Equilibrium formation of MgHPO_4 salt aggregate constant (from the decrease of total ^{31}P integral), and (b) RNA synthesis rate constant (from the decrease of NTP signals), compared to literature data. (c) Unimolecular free energy of RNA folding (from the linewidths of ^1H imino signals) compared to experimental UV-melting data (filled grey circles) and theoretical predictions by RNAfold (empty grey circles). (d) Dissociation constant for protein-RNA complex

formation (from the chemical shifts of protein ^1H - ^{15}N signals) compared to ITC measurements.

In Systems NMR data, the experimental parameter values (K_{eq} , k_{cat} , G , K_{D}) and error bars correspond to the means \pm 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- G , ITC- K_{D}) correspond to the optimized parameter \pm s.d. uncertainty of data fits for the melting curve (UV- G) and binding isotherm (ITC- K_{D}). Individual data points, where available, are shown as circles next to the error bars.

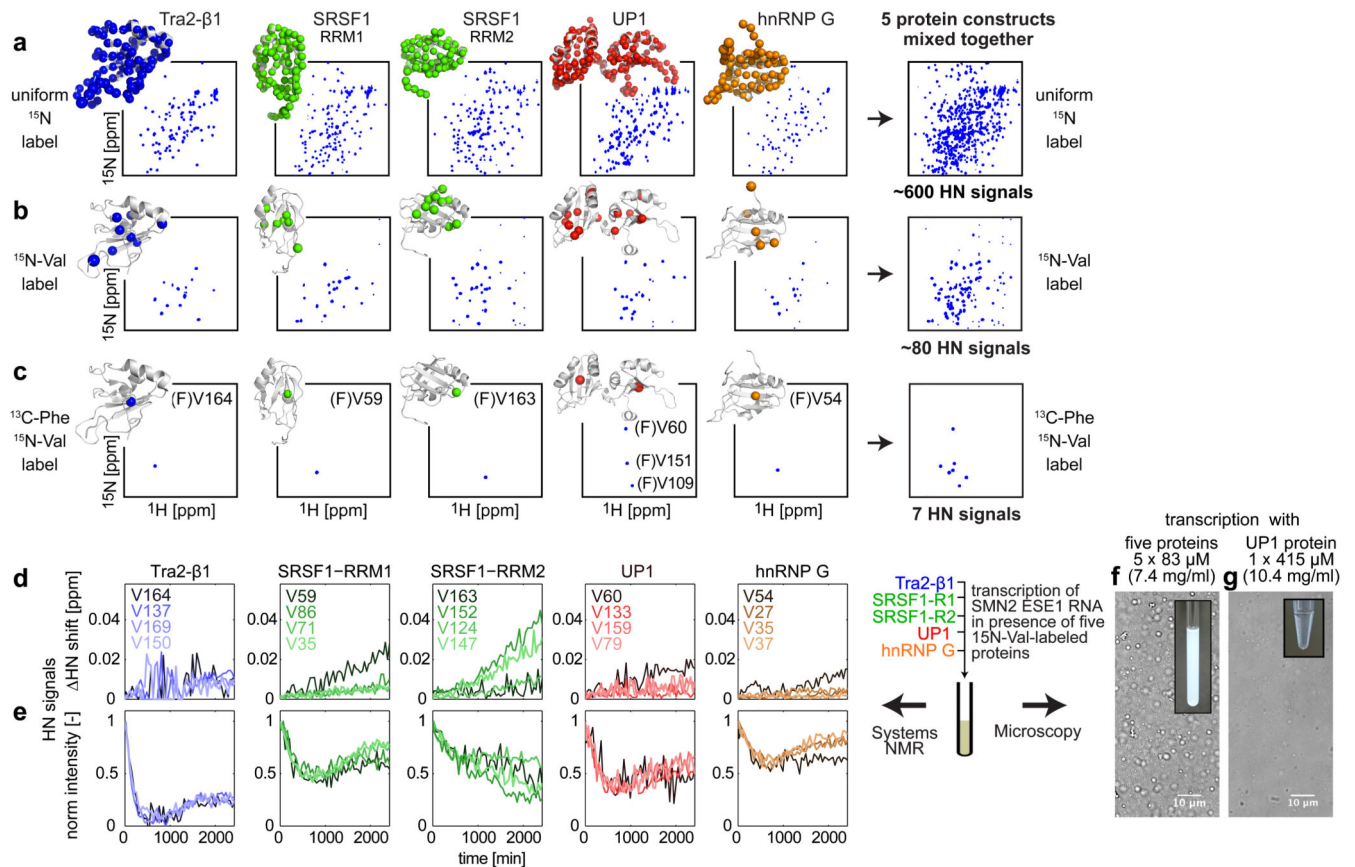


Figure 4. Monitoring multiple proteins in one NMR sample employing selective labeling. NMR spectra of five protein constructs (in columns) with three different labeling schemes – uniform ^{15}N (a), selective ^{15}N -Val (b) and double-selective [^{13}C -Phe, ^{15}N -Val] (c). Protein structures with observable NMR signals as colored spheres are shown next to all the spectra. (a, b) show 2D HN spectra, (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. Last column shows the spectral overlap upon combining five proteins in one sample.

Chemical Shift Perturbation (CSP) (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.

Microscope images of the RNA transcription performed with (f) 5 protein constructs and (g) only UP1 protein at identical 415 μM total protein concentration (in mass units 7.4 mg/ml for 5-protein sample, and 10.4 mg/ml for pure UP1). Five-protein transcription reactions were repeated twice ($n = 2$ independent experiments) with similar results.