

Spliceosomal SL1 RNA binding to U1-70K: the role of the extended RRM

Gopika Gopan^{1,†}, Zhaleh Ghaemi^{1,†}, Caitlin M. Davis^{1,2} and Martin Gruebele^{1,2,3,*}

¹Department of Chemistry, University of Illinois, Urbana, IL 61801, USA, ²Department of Physics, University of Illinois, Urbana, IL 61801, USA and ³Center for Biophysics and Quantitative Biology, University of Illinois, Urbana, IL 61801, USA

Received December 18, 2021; Revised June 19, 2022; Editorial Decision June 21, 2022; Accepted June 29, 2022

ABSTRACT

The RNA recognition motif (RRM) occurs widely in RNA-binding proteins, but does not always by itself support full binding. For example, it is known that binding of SL1 RNA to the protein U1-70K in the U1 spliceosomal particle is reduced when a region flanking the RRM is truncated. How the RRM flanking regions that together with the RRM make up an ‘extended RRM’ (eRRM) contribute to complex stability and structural organization is unknown. We study the U1-70K eRRM bound to SL1 RNA by thermal dissociation and laser temperature jump kinetics; long-time molecular dynamics simulations interpret the experiments with atomistic resolution. Truncation of the helix flanking the RRM on its N-terminal side, ‘N-helix,’ strongly reduces overall binding, which is further weakened under higher salt and temperature conditions. Truncating the disordered region flanking the RRM on the C-terminal side, ‘C-IDR,’ affects the local binding site. Surprisingly, all-atom simulations show that protein truncation enhances base stacking interactions in the binding site and leaves the overall number of hydrogen bonds intact. Instead, the flanking regions of the eRRM act in a distributed fashion via collective interactions with the RNA when external stresses such as temperature or high salt mimicking osmotic imbalance are applied.

INTRODUCTION

RNA molecules inside the cell interact closely with proteins to form ribonucleoprotein (RNP) complexes as a part of cellular machineries such as the ribosome and the spliceosome (1,2). RNA–protein binding is known to be mediated by RNA binding domains (RBDs) (3–5) commonly identified in RNA binding proteins. The 80 residue long RNA recognition motif (RRM), is the earliest to be identified

and the most abundant among these domains (6–8). Until the last decade, the majority of studies, and hence our understanding of RNA–protein binding mechanisms, focused on the minimal globular RBDs. Early reports on the role of linkers connecting two minimal RNA binding motifs in mediating RNA binding (9), on the disordered regions in proteins associated with RNA processing (10,11), and even on RNA-binding proteins that lack a minimal binding domain (12,13), point to the importance of extended RNA-binding domains in RNA–protein interactions. Several ribosomal proteins and splicing factors also exhibit unconventional binding modes involving N- and C-terminal extensions adjacent to the minimal RNA binding domain that form extensive interactions with the RNA (14–17). Extended binding domains are particularly important in the context of RNA processing and storage that involves exchange of RNP components across cell compartments and membraneless organelles (18,19).

The spliceosome is such an RNP, many of whose components assemble and disassemble routinely with the assistance of RRBs (19). The U1 small nuclear ribonucleoprotein (U1 snRNP) particle in particular has an interesting life cycle. It consists of a single RNA folded into four hairpin loops. Stem loop 1 (SL1) binds to the RRM in protein U1-70K (20), and stem loop 2 (SL2) binds to the RRM in protein U1A. Similar to other spliceosomal subunits, U1 snRNP is partly assembled in the cytoplasm with a Smith (Sm) protein heptamer and cycled into the nucleus for further assembly and function (21). The U1 snRNP is involved in the initial recognition event at 5' splice sites of pre-mRNA transcripts (22).

In contrast to U1A and SL2 RNA (23–26), the binding of U1-70K to SL1 RNA has not been studied quantitatively. U1-70K is a 437 residue protein, whose residues 92 through 202, containing the full RRM, result in weakened binding to the SL1 RNA (27). Residues 60 through 216 form the basis of a crystal structure (28), and enable full-strength binding to the SL1 RNA (27). This construct is thus an excellent model system to study the role of an extended RRM (eRRM) in RNA–protein binding. The eRRM contains a

*To whom correspondence should be addressed. Tel: +1 217 244 4245; Email: mgruebel@illinois.edu

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

helix on the N-terminal side of the RRM, the ‘N-helix’ in Figure 1, and an intrinsically disordered region on the C-terminal side of the RRM, the ‘C-IDR’ in Figure 1. The role of these extended regions flanking the minimal U1-70K RRM in RNA binding is not yet understood. In contrast, the remaining mostly disordered regions of U1-70K, residues 1–59 and residues 216–437, are known to be important for binding of the U1 snRNP to accessory proteins such as the Sm and SR proteins (29–33), but there is no experimental evidence that they interact with the SL1 RNA in any way that enhances the RRM binding (27,28), although this does not rule out mediated interactions, e.g. via Sm proteins.

Here we examine how SL1 RNA binding to the RRM is affected by deletion of its flanking N-helix (residues 60–92) making a ‘ Δ N’ sequence, or deletion of C-IDR (residues 203–216) making a ‘ Δ C’ sequence. Specifically, we measure eRRM–RNA interaction when stressed by high temperature or salt concentration. Protein–RNA binding is sensitive to stressors such as elevated temperature or osmotic imbalance, which can lead to changes in binding constants, protein expression levels, or formation of phase separated compartments (34–36) as cells move out of homeostasis. Environmental stresses are known to affect splicing, as well as mature mRNA nuclear export and sequestration in cytoplasmic foci, in order to favor production of heat shock-related RNAs (37). Thus, we hypothesize that the RRM is sufficient for minimal binding under homeostatic conditions, whereas the extra flanking regions of the eRRM could assist binding when U1 snRNP is under stress. To contrast the flanking interactions with RRM binding, we study the wild type RNA and a double mutant that reduces RRM binding affinity of the SL1 RNA and will be suitable for comparison with future in-cell studies that require complex dissociation temperatures $<45^{\circ}\text{C}$.

In this study, equilibrium thermal denaturation of the complex reveals that deletion of the N-helix greatly reduces overall binding as assessed by Förster resonance energy transfer (FRET), whereas deletion of C-IDR affects more the local structure and dynamics of the binding pocket as assessed by tryptophan fluorescence. Analysis of the binding kinetics shows that temperature and high salt stresses affect on- and off-rates in concert to reduce binding affinity, so both search for the binding partner as well as complex stability are hindered by these stresses. The effect of truncations is revealed at the atomistic level by over 20 μs of molecular dynamics (MD) simulations on the eRRM and truncated complexes over a range of temperatures and salt concentrations in accordance with the experimental values. The MD simulation allowed us to rule out loss of stacking in the binding pocket or net loss of hydrogen bonding as the primary causes for the lower observed stability of the Δ N and Δ C complexes. Instead, analysis of the simulation highlights that the C-IDR interacts with the mid-stem of the RNA and supports our experimental finding that the N-helix becomes increasingly important in maintaining complex stability when temperature or mutation stress is applied to the complex. Thus, the N-helix flanking the U1-70K RRM can provide additional binding interactions under heat- or salt-induced stress.

MATERIALS AND METHODS

RNA mutations and RNA handling

The RNA construct designed for the binding studies comprised of 32 bp (nucleotides 17–48, Supplementary Table S1) of human U1 snRNA constituting the stem loop 1 (SL1). Wild type and mutated SL1 RNA oligomers with fluorescent label (Alexa fluor 488 NHS ester coupling at the 3’ end through an amino linker) were obtained from Integrated DNA Technologies (IDT). An alternate version with Alexa Fluor 594 label was also designed for control experiments. Secondary structure and stability of the wild type and mutant constructs were determined using circular dichroism (CD) spectrophotometry. RNA constructs were heat-shocked (3 min at $> 80^{\circ}\text{C}$, followed by 5 min in ice) prior to binding experiments to ensure correct secondary structure.

Plasmid construction, protein expression and purification

DNA sequences coding for all protein constructs were cloned into pDream2.1/MCS vector (Genscript USA Inc) with an N-terminal 6XHis-tag for purification followed by a thrombin cleavage site. All plasmids were chemically transformed into BL21-CodonPlus (DE3)-RIPL competent cells (Agilent) according to the manufacturer protocol. Proteins expression and purification were optimized based on previously reported protocol (38). The cells were grown in 1 L Lennox LB broth (BD Difco) at 37°C for ~ 3 h, and protein expression was induced with 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG, Inalco Pharmaceuticals) at $\text{OD}_{600} = 0.60$. The cultures were grown at 20°C for 12–16 h after induction. The cells were pelleted by centrifugation (Beckman Coulter Avanti J-E, 5000 rpm, 30 min, 10°C) followed by resuspension in lysis buffer prepared by dissolving 2 tablets of protease inhibitor (Pierce™ Protease Inhibitor Mini Tablets, EDTA-free, Thermo Scientific) in 20 ml of 20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4. 20 μl DNase I (New England Biolabs) was added to the cell suspension and lysed by ultrasonication (Qsonica-Q700) with 70% amplitude, 6 s pulse/min and 6 min total process time. The cell lysate was centrifuged (10 000 rpm, 20 min, 10°C) to remove cell debris and supernatant was filtered through a 0.45 μm Millex sterile syringe filter followed by 0.22 μm filter. The filtrate was loaded on a 5 ml HisTrap HP column by ÄKTA pure FPLC (GE Healthcare Life Sciences), and the protein was eluted with elution buffer (20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.4). Purity of eluted protein was assessed using SDS-PAGE and mass spectrometry (MALDI). Protein was dialyzed against storage buffer (10 mM potassium phosphate, 200 mM KCl, pH 7.4), flash frozen in liq. N_2 and stored at -80°C .

Equilibrium thermal dissociation of the eRRM–SL1 RNA complex

Temperature-induced dissociation of the complex was measured on a JASCO FP8300 spectrofluorometer equipped with Peltier temperature control. An equimolar mixture

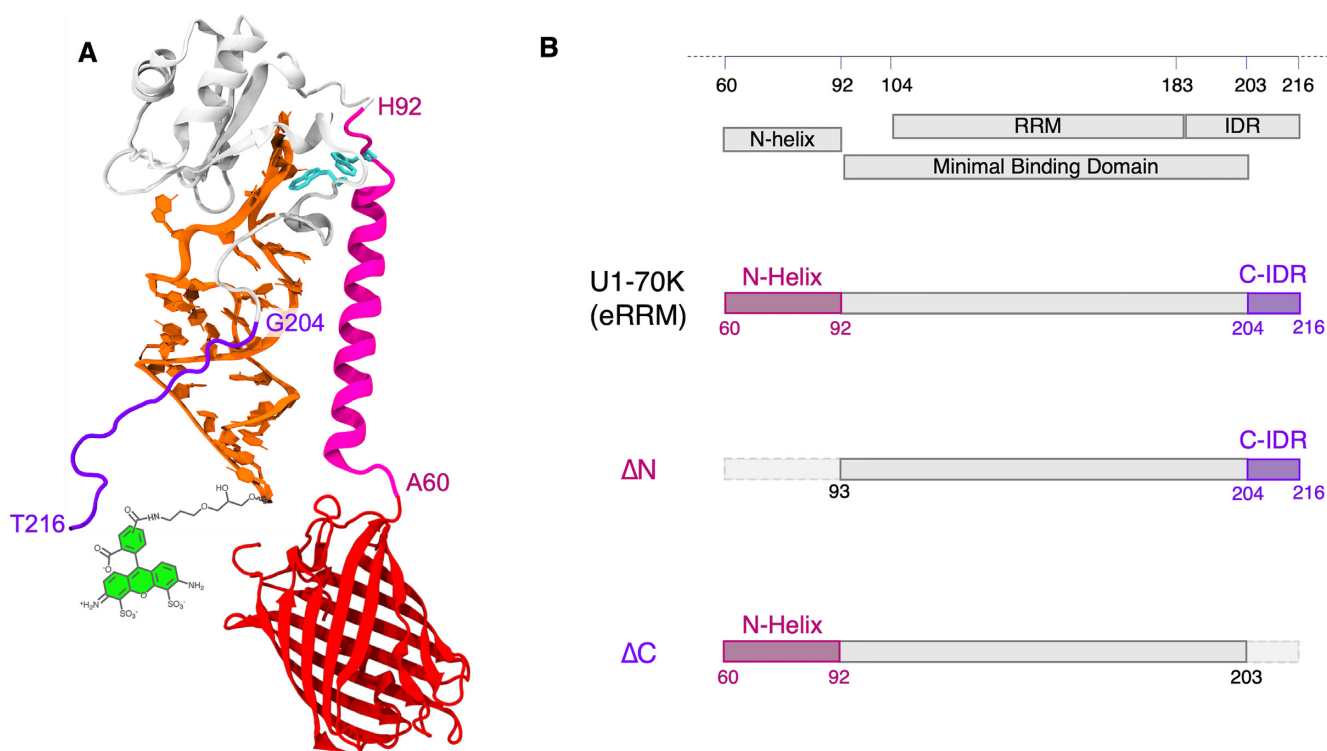


Figure 1. (A) The eRRM-SL1 complex (PDB ID: 4PKD) showing W89 and W187 (cyan licorice) on U1-70K (gray), N-helix (A60 to H92) in magenta, C-IDR (G204 to T216) in violet and FRET labels Alexa Fluor 488 (green) on RNA (orange) and mCherry (red) on U1-70K. (B) Annotated U1-70K sequence in the region of the eRRM (60–216) with domains included in the 3 main protein constructs used for the experiments.

of 1 μ M protein and RNA in 10 mM potassium phosphate buffer (pH 7.4) with appropriate salt concentration (200–500 mM KCl) were added to 150 μ l quartz cuvettes (Starna Cells). For FRET experiments, Alexa Fluor 488 label (FRET donor) on RNA was excited at 488 nm and emission spectra was scanned from 480 to 700 nm at every 3°C from 15–63°C (Supplementary Figure S1A). mCherry on the protein acts as a FRET acceptor and the resulting FRET signal (Supplementary Figure S1B) was used as a probe for complex formation. FRET efficiency was calculated from donor and acceptor emission intensities at 520 nm and 610 nm as follows:

$$E_{\text{FRET}} = \frac{I_{610}}{I_{610} + I_{520}}, \quad (1)$$

where I_x is the fluorescence intensity at x nm.

Tryptophan quenching experiments (39) were performed on the same instrument. The U1-70K eRRM has 2 tryptophan residues (W89, W187) at the RNA binding pocket which could be potentially used as reporters of RNA binding. For thermal dissociation studies, Trp residues were excited at 295 nm and emission was collected from 290 to 450 nm at every 3°C from 15 to 63°C (Supplementary Figure S1C). Tryptophan emission maximum at the highest temperature point was determined from a gaussian fit of the spectrum from 330 to 360 nm. Variation in quenching due to protein–RNA unbinding was reported by the normalized intensity at the emission maximum thus obtained (Supplementary Figure S1D). All normalized tryptophan quenching curves have the unbinding transition set to range from

1 to 2. Top and bottom baselines have negative slopes due to the intrinsic temperature dependence of the tryptophan fluorescence as in case of the FRET labels. The unbinding transitions from unlabeled protein–RNA complexes monitored by tryptophan fluorescence can be compared to their FRET-labeled counterparts to check the effect of labels on dissociation of the protein–RNA complex and protein–RNA binding stability. FRET efficiency is a ‘global’ probe, and Trp quenching probes the local binding pocket.

Temperature-jump relaxation kinetics of protein–RNA binding

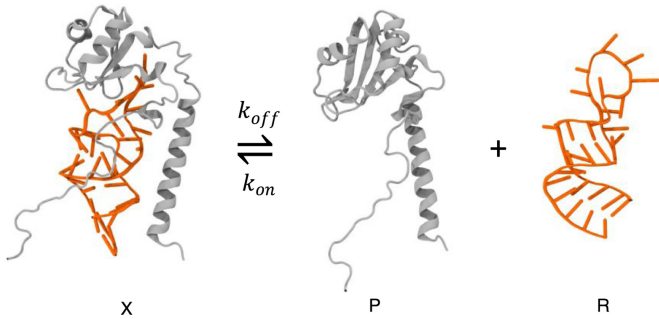
We use two-color fast relaxation imaging (FRFI) to quantify the kinetics and thermodynamics of protein–RNA binding by laser-heating the complex in aqueous buffer on a timescale faster than the dynamics of the system and studying the relaxation kinetics. The detailed instrumental setup for FRFI has been described previously (38). A 2 μ m IR laser is used to heat rapidly and uniformly a sample containing an equimolar protein–RNA mixture in aqueous buffer to perturb its binding equilibrium. The power density of the IR laser is programmed to induce a series of 3–5°C temperature steps, followed by \sim 10 s equilibration time after each step for the system to relax to its new equilibrium state at the higher temperature. The relaxation kinetics of the FRET-labeled sample is imaged onto a camera using an epifluorescence microscope which splits the emission signal into two channels, ‘green’ and ‘red’. The two-color signals are integrated into a signal ‘G’ and ‘R’ and used to compute FRET

efficiency as a function of time as $E(t) = R(t)/[G(t)+R(t)]$. The baseline of the FRET efficiency (from the last 4 s) is averaged to determine the equilibrium FRET value at each temperature, and the decay of the FRET efficiency is used to fit the kinetics as described in the next two sub-sections. The temperature after each step is quantified using the temperature-dependent quantum yield of mCherry as a standard (Supplementary Figure S3).

For sample preparation, microscope slides (VWR, plane selected pre-cleaned, $22 \times 75 \times 1$ mm) were wiped with ethanol to remove dust or other particles. A 13 mm diameter, 120 μm thick spacer (25×25 mm, Grace Biolabs) was stuck to the center of the microscope slide to confine the sample. 40 μl of an equimolar mixture of 1 μM protein and RNA was in the appropriate buffer was loaded into the spacer opening along with 1 μl of 1:2000 dilution of 6 μm green-fluorescent polystyrene beads (10 mg/ml, Degradex), sealed with a cover glass (22×22 mm, Fisherbrand) and the slide is inverted with the cover glass facing down to allow the beads to settle on the cover glass. The microscope slides containing the sample were loaded on the epifluorescence microscope stage in the inverted position. The beads were used to accurately locate the z-position to focus the IR-laser beam onto the sample on the cover glass.

Protein-RNA dissociation equilibrium and kinetics

To analyze all the FReI data, we assumed a bimolecular equilibrium for the complex:



Here, X is the complex concentration, P is the protein concentration, and R is the RNA concentration. The dissociation equilibrium constant is given by $K_d = \exp[-\Delta G(T)/RT] = k_{off} / k_{on} = PR/X = [P_0 - X][R_0 - X]/X$. Here P_0 and R_0 are the total protein and RNA concentrations. ΔG is the free energy of dissociation and was approximated by the linear free energy relationship $\Delta G(T) = G_0 + G_1(T - T_{1/2})$, where G_0 (kJ/mole) = $-0.00831 T \ln[P_0/2]$ and $T_{1/2}$ is the midpoint of the temperature-induced unbinding transition of the protein-RNA complex.

The total signal obtained from equilibrium dissociation experiments can be expressed as

$$S(T) = S_{bound}(T) \left(\frac{X}{P_0} \right) + S_{free}(T) \left(1 - \frac{X}{P_0} \right), \quad (2)$$

where $S(T)$ is the total FRET signal as a function of temperature. $S_{bound}(T)$ and $S_{free}(T)$ are the contributions from the bound and free state of the complex, assumed to be linear functions of temperature due to gradual quantum yield

change of the fluorophores with temperature ($A + BT$ and $C + DT$ respectively, where T is in $^\circ\text{C}$). The fluorescence intensity (quantum yield) of tryptophan as well as the FRET labels decreases linearly in the range of temperature used for the measurements (Supplementary Figure S4), which produces the bound and dissociated complex baselines in Eq. (2) and Figure 2. The baselines can be fitted by linear functions and are useful for determining the temperature in the sample by calibration. Hence, (X/P_0) and $(1 - X/P_0)$ give the fraction of bound and free population of the complex. The equilibrium dissociation constant K_d can be determined by fitting the experimental data to the following expression obtained by modifying and rearranging Eq. (2):

$$S(T) = S_{bound}(T) + [S_{bound}(T) - S_{free}(T)] \left[\frac{P_0 + R_0 + K_d - \sqrt{\{P_0 + R_0 + K_d\}^2 - 4P_0R_0}}{2P_0} \right]. \quad (3)$$

All data were analyzed using IGOR Pro (WaveMetrics) by fitting the parameters G_1 and $T_{1/2}$, and fixing the known parameters P_0 and R_0 .

The two-color data from relaxation kinetics on FReI can be used to determine FRET efficiency as a function of time, which then can be fit to the integrated rate law for the dissociation reaction. Assuming equimolar total protein and RNA concentrations, the differential rate law for the reaction is given by

$$\frac{d[X]}{dt} = -k_{off}[X] + k_{on}[P]^2. \quad (4)$$

Solving the differential rate law for initial conditions, $X = X_0$, $P = P_{tot} - X_0$, where X_0 is the concentration of the complex at the beginning of a temperature step and P_{tot} is the total protein concentration used for the measurement,

$$X(t) = \frac{k_{off} + 2k_{on}P_{tot} - k' \tanh\left[\frac{1}{2}kt + \tanh^{-1}\left[\frac{k_{off} + 2k_{on}(P_{tot} - X_0)}{k'}\right]\right]}{2k_{on}}, \quad (5a)$$

where

$$k' \equiv \sqrt{k_{off}^2 + 4k_{on}P_{tot}}. \quad (5b)$$

The total FRET signal for relaxation kinetics after each temperature will be proportional to the complex concentration and can be fit to the following equation:

$$S(t) = A_0 + A_1 e^{-\frac{t}{\tau}} + A_2 X(t) \quad (6)$$

Here, $\tau = 200$ ms is the known time constant to account for the instrument response (due to relaxation of the fluorescent protein and heating response time), which is the dead-time for the measurements. This equation is fit to the data by fitting the parameter k_{on} only, while using the known values of P_{tot} , K_d at the final temperature of the step (from the thermodynamic fit) and calculating $k_{off} = k_{on}K_d$ as well as X_0 from K_d at the temperature just before the step.

Construction of the computational model system and MD simulations

The structure of the U1-70K protein (residues 60–203) in complex with SL1 RNA was determined by X-ray crystallography (PDB ID: 4PKD) (28). In the eRRM protein,

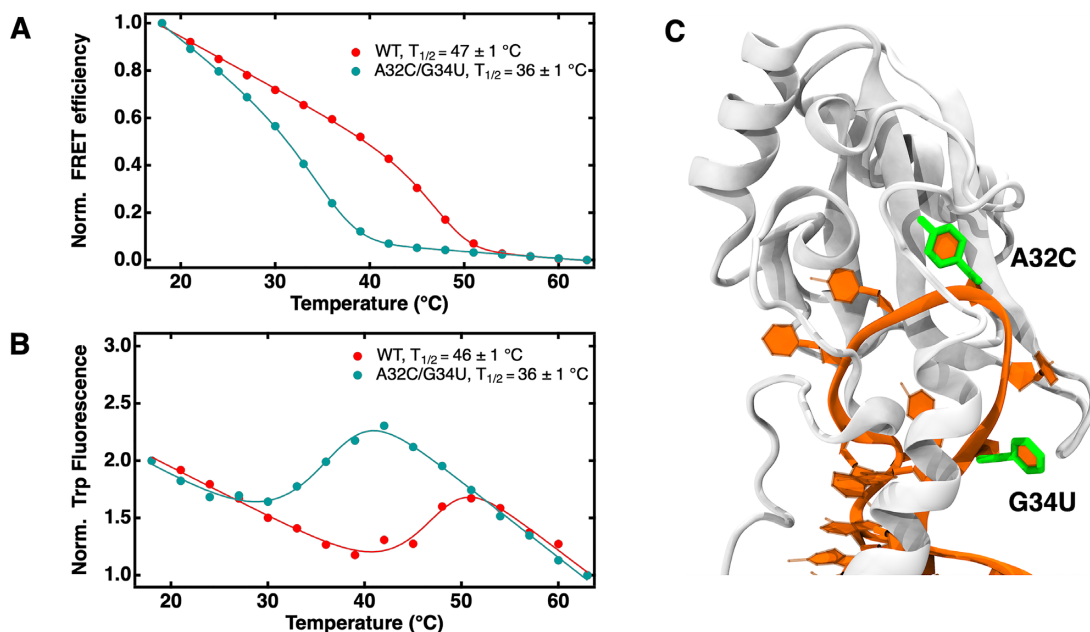


Figure 2. Extended RRM to SL1 binding is destabilized by RNA mutations at the binding site. (A) Normalized FRET efficiency curves of 1 μ M wild type and double mutant SL1 in complex with the U1-70K eRRM in 10 mM potassium phosphate and 300 mM KCl pH 7.4 excited at 488 nm. (B) Normalized tryptophan emission intensity curves under the same conditions. Both FRET efficiency and tryptophan emission curves were fit to eq. 3 in Methods to obtain $T_{1/2}$ shown in the figure legends, and baseline fitting is explained in Eq. (2) in Materials and Methods; (C) eRRM–RNA complex showing mutated sites (A32C, G34U) on SL1 RNA highlighted in green licorice.

13 residues of the C-terminal region were not resolved in the X-ray crystallographic structure. To model the missing residues (residues 204–216 in Figure 1B), trRosetta (40) was used: from the 15 models that were randomly generated, one that is both unstructured and with suitable orientation with respect to the rest of the protein was chosen. The SL1 RNA sequence corresponding to the experimental sequence was obtained by mutating the A32C and G34U nucleotide residues using routine *psfgen* in the NAMD program (41,42). The *Solvate* plugin of the VMD program (43) was used to solvate the simulation box with TIP3P water molecules (44) to provide a distance of 30 Å between the periodic images. After neutralizing the system, an additional 0.3 M KCl and 0.5 M KCl were added to the water box using program *Ionize* to reproduce the low salt and high salt experimental conditions. Because folding equilibria of proteins have been reproduced in extensive tests with CHARMM36 force field, we performed our simulations using CHARMM36m for protein and CHARMM36 for RNA for consistency (45,46); additionally, this force field choice will allow comparison of the U1-70K-SL1 RNA complex dynamics with recent U1A-SL2 RNA complex simulations (24,47). Stepwise minimization and equilibration were performed as follows: (i) 15000 minimization steps with all heavy atoms restrained; (ii) 15000 minimization steps with all heavy atoms restrained except water molecules; (iii) 30000 minimization steps with protein and RNA backbone atoms restrained; (iv) 30000 minimization steps with no positional restraints; (i) 100 ps of equilibration at -173°C with protein and RNA backbone atoms restrained; (ii) 100 ps of equilibration at -73°C with protein and RNA backbone atoms restrained; (iii) unrestrained

equilibration at 25°C . The restraint force constant for all the above simulations was 1 kcal/mol/Å². The equilibrations were performed in the NPT ensemble by using a Langevin thermostat (damping coefficient of 5 ps⁻¹) and a Langevin piston method (48,49) (200 fs piston period and 100 fs piston decay). Periodic boundary condition and NPT ensemble were used for production runs. The long-range electrostatics was computed by particle-mesh Ewald algorithm (50,51), Lennard–Jones interactions were cut off at a distance of 12 Å and the time step was 2 fs. All simulations were performed with NAMD 2.14 and NAMD 3 programs (41,42). We performed a cumulative total of 22.5 μ s of MD production runs. Three replicas of 500 ns were performed for each of the eRRM, Δ N, and Δ C proteins in complex with SL1-RNA. To mimic the experimental condition, each complex was simulated for five different conditions of: (i) 300 mM KCl and 25°C ; (ii) 300 mM KCl and 40°C ; (iii) 300 mM KCl and 50°C ; (iv) 500 mM KCl and 25°C ; (v) 500 mM KCl and 50°C . The root mean square deviation (RMSD) plots of the 45 production runs of 500 ns each (3 complexes \times 3 replicas \times 5 conditions) showed that the complexes reached equilibration in approximately 100 ns (Supplementary Figure S5).

Dynamical network, community, and shortest path analysis

We are interested in which parts of the protein and RNA move together in the MD simulations. Correlated motion shows which part of the protein responds to RNA motion as temperature or salt concentration change. We constructed a network on a set of nodes which are connected by edges. The nodes were defined as the C_{α} and one side chain atoms of

all protein residues, and phosphate and side chain nitrogens of the RNA nucleotides. Edges are placed between a pair of nodes if any heavy atoms from the two corresponding residues are within 4.5 Å of each other for at least 75% of the entire trajectory. Each edge is weighted by the weight $W_{ij} = -\log(|C_{ij}|)$, where C_{ij} are the Pearson correlation value of the two nodes, defined as:

$$C_{ij} = \frac{\langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle}{\sqrt{\langle r_i^2 \rangle \langle r_j^2 \rangle}} \quad (7)$$

where, r_i and r_j are the positional fluctuation vectors of atoms i and j , respectively. On the constructed network we identified ‘communities’ of nodes moving together within the complex. The community detection follows the Girvan–Newman algorithm (52), and they represent the nodes that are densely interconnected with each other. A path length D_{ij} between nodes i and j is defined as the sum of the edge weights between the consecutive nodes (k, l) along the path: $D_{ij} = \sum_{k,l} W_{kl}$. The Floyd–Warshall algorithm (53) was used to find the shortest distance D_{ij} between all pairs of nodes in the network. To ensure network connectivity during the shortest path analysis, the edges between the phosphate groups of the RNA nucleotides were not removed from the final network. The NetworkView plugin of the VMD program was used for analysis and visualization (43,54).

RESULTS

Truncated U1-70K complexes are stable

The extended RRM (eRRM) sequence used in this study is based on the human spliceosomal U1-70K fragment of residues 60–216, used for a crystal structure (28). The missing disordered N-terminal residues (residues 1–59) and C-terminal residues (residues 217–437), important for binding to Sm and SR proteins are not considered here (27,29–31), and are not required for strong SL1 RNA binding (27). Therefore, the 157 amino acid eRRM, shown in Figure 1 (amino acid sequence provided in Supplementary Table S1) is our reference sequence in this study. Thirty-three N-terminal residues constituting the N-helix were deleted to obtain the ΔN truncation. Similarly, ΔC was designed by truncating 13 disordered residues from the C-terminus. These 13 residues were not resolved in the crystal structure. To observe complex formation via FRET, the eRRM was fluorescently labeled by fusing the fluorescent protein mCherry to the N-terminus (Figure 1A) which was chosen as a FRET acceptor from the Alexa Fluor 488 label on the RNA. FRET labeling also enables future direct comparison with in-cell experiments. To avoid labeling the truncated area, the mCherry label of ΔC was placed at the N-terminus, and ΔN was labeled at the C-terminus. Finally, the eRRM was also labeled at the C-terminus as a control. All proteins expressed well in *Escherichia coli* and were purified for RNA binding experiments as described in Materials and Methods. The N-terminal 6x-His tag on U1-70K eRRM was shown to not affect RNA binding (Supplementary Figure S2A). Unless otherwise noted, we will refer to the mCherry-labeled U1-70K extended RRM with

N-terminal 6x-His tag as ‘eRRM,’ the N-terminal truncation as ‘ ΔN ’, and the C-terminal truncation as ‘ ΔC ’. The segments that have been truncated away will be referred to as N-helix and C-IDR respectively.

FRET and tryptophan fluorescence report on complex dissociation

We monitored the eRRM–SL1 RNA complex dissociation by two different fluorescence methods. FRET is a global measure of dissociation, covering the ~ 5 nm range given by the 50% FRET efficiency distance (Förster distance, R_0) of the two fluorophores. For the Alexa 488–mCherry pair, $R_0 = 5.9$ nm (55). In addition to FRET, we measured tryptophan fluorescence because two tryptophan residues, W89 and W187 are located near the RNA binding site (Figure 1A, cyan side chains), and are potentially quenched by the RNA via Dexter transfer (~ 0.5 nm range) (56). Therefore, FRET and tryptophan report on the dissociation reaction in complementary ways: FRET is a global measure of dissociation, whereas tryptophan fluorescence looks at the native-like bound structure near the binding pocket.

Tryptophan fluorescence measurements comparing the eRRM (contains both W89 and W187) and ΔN (missing W89) revealed that W187 is the major contributor ($\sim 80\%$) towards the overall fluorescence signal and hence the major reporter in the binding experiments, although the smaller signal from W89 ($\sim 20\%$ of total Trp fluorescence) was also found to be sensitive to RNA binding (Supplementary Table S2, Figure S2B).

Thermal dissociation of a sample of 1 μM FRET-labeled eRRM-wild type SL1 RNA complex was measured using Alexa Fluor 488 as donor on the RNA and mCherry as acceptor label on the eRRM (Figure 2A). The dissociation reaction yielded an equilibrium dissociation constant as low as 1.7×10^{-10} M in 10 mM potassium phosphate buffer with 200 mM KCl, which closely mimics the in-cell salt concentration (Supplementary Figure S6). In this study, we have used 300 mM KCl, about 100 mM above physiological total ion concentration, as our main reference salt concentration because the lower baselines of unbinding curves of the weaker-binding complexes were dominated by the temperature dependence of the FRET labels in 200 mM KCl. The unbinding transition reported by FRET at $T_{1/2} = 45^\circ\text{C}$ in 300 mM KCl is in good agreement with that reported by loss of tryptophan quenching (Figure 2B), indicating that the progress of the dissociation reaction can be studied by using FRET, useful for future in-cell studies.

We also measured the RNA A32C/G34U double mutant dissociation equilibrium, based on two RNA mutations known to reduce binding affinity (Supplementary Figure S7). The tryptophan intensity of wild type and double mutant RNA is very similar except for a 10°C shift in stability (Figure 2B), consistent with the structure of the binding pocket remaining unchanged near the two tryptophan residues. We measured the two single mutants as well to show that the destabilization of RNA to eRRM binding and hence the decrease in free energy at sites A32 and G34 is additive, a further sign that no major disruption of the binding site occurs (Supplementary Figure S6, Figure S7). The additive effect of the RNA mutations increases the

Table 1. Effect of RNA mutations, salt concentration and protein domain truncations on the stability of RNA binding to the extended RRM and its truncations. $T_{1/2}$ (midpoint temperature for complex dissociation), free energy for complex dissociation at 37°C and dissociation constant K_d are shown. Table shows global parameters derived from FRET measurements unless (Trp) is noted for tryptophan fluorescence measurements. See Supplementary Table S4 for equivalent parameters at 25°C

Protein	RNA	KCl ^a (mM)	$T_{1/2}$ (°C)	ΔG (kJ/mol)	K_d (M)
eRRM	wild type	200	47 ± 1	52 ± 9	1.7×10^{-10}
eRRM	wild type	300	45 ± 1	50 ± 4	4.4×10^{-9}
eRRM	wild type	500	42 ± 1	43 ± 2	4.8×10^{-8}
eRRM	A32C/G34U	200	41 ± 1	44 ± 3	3.7×10^{-8}
eRRM	A32C/G34U	300	35 ± 1	35 ± 1	1.1×10^{-6}
eRRM	A32C/G34U	400	31 ± 1	30 ± 1	7.9×10^{-6}
eRRM	A32C/G34U	500	28 ± 1	27 ± 1	3.3×10^{-5}
ΔN	wild type	300	34 ± 1	34 ± 14	1.7×10^{-6}
ΔN	wild type	500	31 ± 1	27 ± 3	2.4×10^{-5}
ΔN	A32C/G34U	300	28 ± 1	22 ± 6	1.8×10^{-4}
ΔN	A32C/G34U	500	15 ± 1	10 ± 1	2.2×10^{-2}
ΔC	wild type	300	44 ± 1	43 ± 7	5.3×10^{-8}
ΔC	wild type	500	42 ± 1	41 ± 7	1.2×10^{-7}
ΔC	A32C/G34U	200	41 ± 1	42 ± 11	8.1×10^{-8}
ΔC	A32C/G34U	300	34 ± 1	34 ± 4	1.8×10^{-6}
ΔC	A32C/G34U	400	29 ± 1	29 ± 1	1.5×10^{-5}
ΔC	A32C/G34U	500	26 ± 1	25 ± 1	6.2×10^{-5}
ΔC (Trp)	A32C/G34U	300	30 ± 1	29 ± 2	1.2×10^{-5}
ΔC (Trp)	A32C/G34U	500	26 ± 1	21 ± 1	3.2×10^{-4}

^aMeasurements done in 10 mM potassium phosphate with added concentration of KCl as indicated.

dissociation constant K_d by ~ 200 times and reduces the dissociation temperature to 41°C under physiological salt conditions (Table 1).

Temperature-induced dissociation of an alternate FRET construct with mCherry at the C-terminus of the eRRM was also measured in order to rule out the effect of positioning of fluorescent tags on the protein–RNA complex (Supplementary Figure S8). Finally, the unbinding reaction showed good reversibility in the temperature range used for all the studies here (Supplementary Figure S9).

We also checked the effects of physiological levels of Mg^{2+} (0.5–1 mM) (57) on the system. Trp fluorescence was monitored for thermal dissociation of the wild type SL1 and A32C/G34U SL1 with the eRRM in 10 mM potassium phosphate buffers with 200 mM KCl and 0.5–1 mM $MgCl_2$. We did not see any significant effects of up to 1 mM Mg^{2+} on complex stabilities (Supplementary Table S5, Figure S10). Hence all further measurements were carried out in buffers without Mg^{2+} .

Increased salt concentration perturbs electrostatic interactions and increases K_d by three orders of magnitude K_d

In order to investigate the effect of salt (KCl) on stability of the U1-70K eRRM–SL1 complex, we studied the dissociation reaction of an equimolar mixture of FRET-labeled protein and RNA in 10 mM potassium phosphate buffers with KCl concentration varying from a near physiological 200 mM up to 500 mM. Raising the salt concentration by a factor of 2.5 raised the by close to a factor of 900, indicating a significant weakening of protein–RNA binding (Figure 3A,

Table 1). We observed a near-linear dependence of $\log[(M)]$ on $\log[KCl (M)]$ with a slope of 7.4 (Figure 3B) suggesting that at most seven salt bridges are mediating binding between the eRRM and the double mutant RNA (58).

Truncating the N-helix increases K_d by two orders of magnitude and increases complex destabilization under stress

The N-helix flanking the stem of the SL1 RNA contains nine positively charged basic residues that are close to the negatively charged phosphate backbone of SL1 RNA, which is expected to promote the association with SL1 (59). The N-helix does not have any other known binding function within U1 snRNP or to known cofactors. To quantify possible contributions from the N-helix, we studied the ΔN complex. Deletion of the N-helix resulted in a large destabilization of the complex (Figure 4A) shifting the binding free energy by about 13 kJ/mol and increasing the K_d by about a factor of 150 (Table 1). Hence, the N-helix makes a significant contribution to the U1-70K–SL1 complex formation and stabilization. In addition to lowering $T_{1/2}$ by 8°C, the truncation also increases the relative tryptophan fluorescence intensity of bound vs. dissociated state (Figure 4A), indicating increased mobility of the remaining Trp residue (W187) in the truncated complex (60).

The thermal dissociation experiments at higher salt concentration (500 mM KCl) showed that the degree of complex destabilization by salt is higher in the absence of N-helix for the double mutant. RNA–protein binding is thus enhanced in the doubly stressed (mutation and salt) complex (Table 1, Figure 4C and D, Figure S13). Hence in addition to assisting the RRM in complex formation by interacting with the SL1 RNA stem, these results point towards the role of the N-helix in stabilizing the complex under thermal and hyperosmolar stress.

Truncating the C-IDR affects the binding pocket and the overall binding differently

We also quantify the role of the C-IDR (residues 204–216) that follows the minimal binding domain (residues 92–203 of U1-70K) in binding to the RNA (27). Truncation of this region (Figure 1B) yielded ΔC which is predicted to have reduced efficiency in SL1 binding from immunoprecipitation studies (27).

When the more strongly binding wild type SL1 RNA was examined, no significant effect of the C-terminal region was found: in Figure 4B (left), the eRRM and ΔC complexes have very similar dissociation midpoint temperatures $T_{1/2}$ above 44°C. Nearly the same results were obtained by FRET monitoring, sensitive to overall (long-range) binding, and Trp fluorescence, sensitive to the local binding configuration in the binding pocket. Thus, the C-IDR does not appear to be important for maximal binding to wild type RNA, in contrast to the N-helix (magenta bars in Figure 4B, left).

The results are different for the weaker binding SL1 RNA double mutant (Figure 4B, right and Table 1). The ΔC truncation now lowers $T_{1/2}$ slightly and approximately doubles K_d when monitored by FRET (Table 1). However, tryptophan quenching data that monitors the local bind-

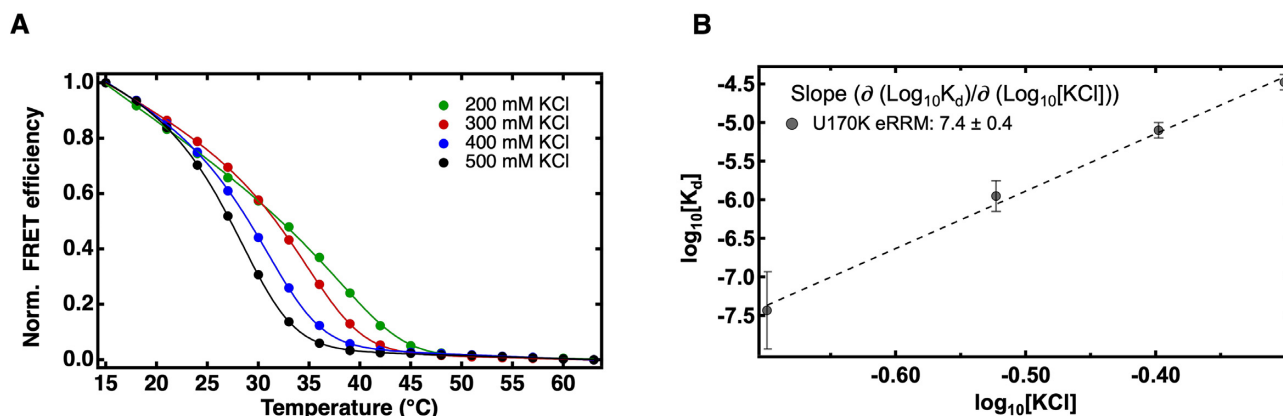


Figure 3. Extended RRM–SL1 interaction shows significant salt dependence. (A) FRET efficiency calculated from fluorescence spectra of equimolar mixture of A32C/G34U SL1 and the U1-70K eRRM in 10 mM potassium phosphate and 200–500 mM KCl pH 7.4 excited at 488 nm collected from 15 to 63°C. Curves were fit to eq. 3 in Methods to obtain thermodynamic parameters; (B) linear correlation between $\log(K_d)$ in M on $\log(\text{KCl})$ in M. See Supplementary Figure S11 to see similar plots for the ΔC complex.

ing pocket shows that C-IDR truncation results in substantial disruption of the local binding pocket, with an order of magnitude increase in K_d (Figure 4A and B, Table 1). To study how high temperature and salt concentration affect the stability of the ΔC complex, we performed thermal dissociation experiments in higher salt (Figure 4C, Figure S13). The results indicate that regardless of the nature of RNA, C-IDR does not stabilize the complex significantly when under temperature and salt stresses (Figure 4D, Figure S13). The full salt dependence of ΔN and ΔC complexes also followed similar overall trend as the eRRM complex (Figure S14).

Kinetics of the eRRM–SL1 interaction shows k_{on} and k_{off} contribute similarly to K_d

To study the kinetics of the protein–RNA interaction, we used the double mutant RNA construct and FRET-labeled protein and subjected the complex to rapid laser-induced temperature steps that perturbed the binding equilibrium (Figure 5A); rapid temperature jumps have been applied previously to nucleic acid–protein dissociation reactions (61). Equilibrium constants derived from the temperature steps matched up well with the equilibrium data obtained earlier (Figure S15B). Each kinetic relaxation was fitted to the bimolecular rate expression Eq. (6) in Materials and Methods, yielding values for k_{on} and k_{off} . A sample fit is shown in Figure 5B, and the full series of fits in Figure S15A.

In a protein–RNA complex dissociation reaction, K_d can increase (i.e. decrease in binding) either because the reactants form the complex more slowly (reduced k_{on}), or because the complex is more weakly bound and falls apart faster (increased k_{off}). Here we found that K_d increases by a factor of about 80 over $\sim 8^\circ\text{C}$ range from 29°C to the physiological temperature at 37°C (Figure 5, Table 2). The off rate is responsible for about a factor of 20, and the on rate for about a factor of 4. Thus, the native complex is significantly disrupted by thermal motion, but the electrostatic search for a binding partner is also hindered at the higher temperature.

Protein truncations stabilize the signature RRM base stacking interaction

We know from the experimental results that truncation of the N-helix destabilizes the complex overall as probed by FRET, and destabilizes the binding pocket interactions as probed by tryptophan fluorescence, with the latter observation also made for the C-IDR truncation. Truncations could affect base-stacking or hydrogen bonding interactions at the RRM binding site or between RNA stem and protein. The most obvious hypothesis based on the data is that truncation removes some of these interactions, but is that simple picture correct? To answer this question, we performed a total of 22.5 μs of molecular dynamics (MD) simulations on the eRRM and truncated complexes (see Materials and Methods) that span a range of temperatures (25, 40 and 50°C) and potassium chloride concentrations (300 and 500 mM) in accordance with the experiments.

In the eRRM–SL1 RNA complex, side chain F106 interacts with nucleotide C31, and F148 with nucleotide C32, via base stacking (Figure 6A). These two base stacking interactions are crucial for the stability of the complex, and a signature of RRM:RNA binding (28). We observed experimentally that the A to C nucleotide substitution at position 32 reduces the dissociation temperature of the complex by 4°C and increases the dissociation constant K_d by an order of magnitude (Supplementary Figure S7, Table S3).

Analysis of the MD trajectories shows that the F106–C31 interaction is unaffected by protein truncations, but surprisingly, the F148–C32 base stacking is stabilized by truncations (Figure 6). To show this, the stability of both base-stacking interactions was monitored by the center-of-mass distance between the interacting residues. For all proteins (eRRM, ΔC and ΔN) at different temperatures and salt concentrations, the stacking interaction F106–C31 is conserved (6E–G). In contrast, the interaction F148–C32 in the eRRM protein shows instabilities even at low temperature and salt concentration and is disrupted completely in all replicas at high temperature and salt concentration (500 mM KCl and 50°C , Figure 6B). However, the ΔC and ΔN truncations stabilize this interaction relative to the eRRM protein (Figure 6C and D).

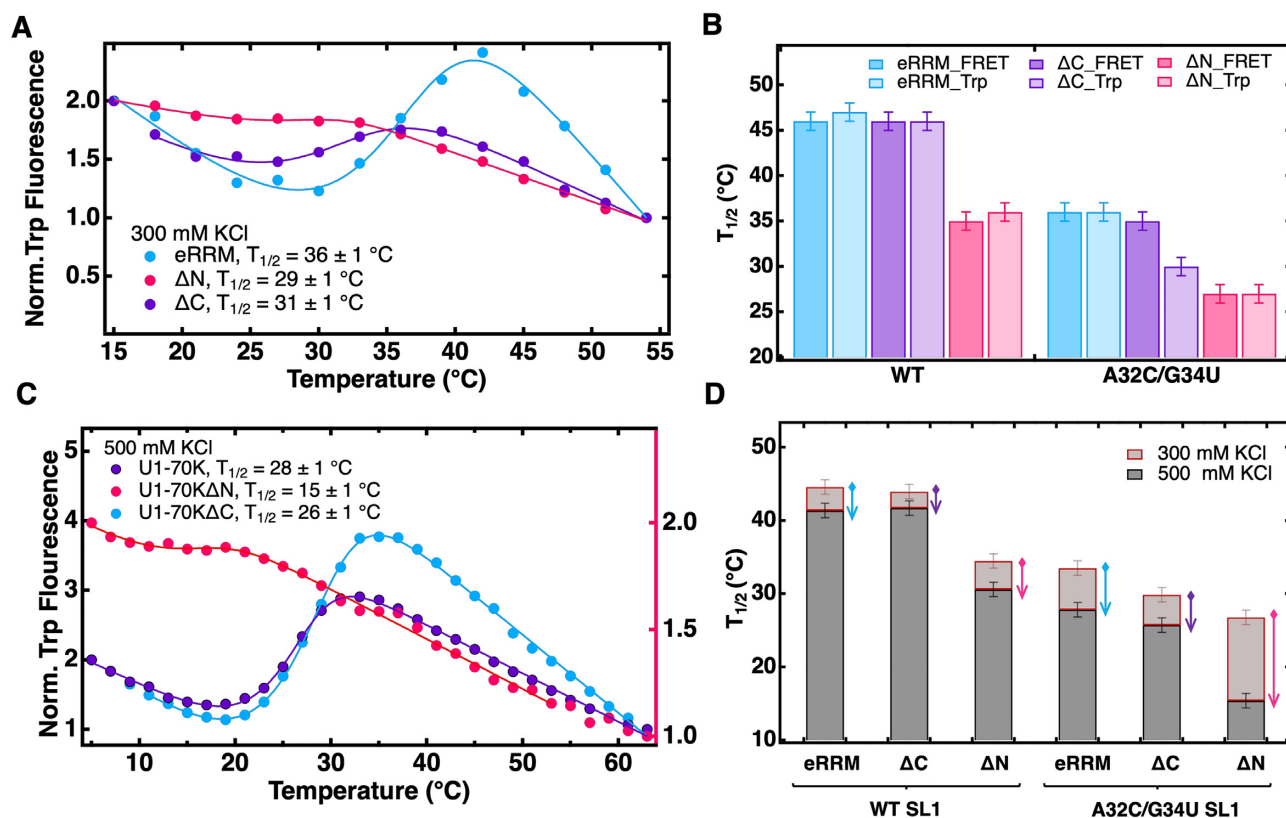


Figure 4. N-helix truncation weakens U1-70K eRRM-SL1 binding whereas the C-IDR truncation only affects the binding pocket of the double mutant. (A) Tryptophan emission intensity curves from fluorescence spectra of equimolar mixture of A32C/G34U SL1 and eRRM (blue), ΔN (magenta) and ΔC (violet) excited at 295 nm collected from 15 to 54°C in 10 mM potassium phosphate and 300 mM KCl pH 7.4. (B) Plot summarizes the effect of N-Helix and C-IDR truncations on binding stabilities of the eRRM in complex with wild type and A32C/G34U SL1; single mutants are shown in Supplementary Figure S12. The effect of N-Helix truncation on binding can be observed clearly in case of all RNA variants, whereas C-IDR truncation only affected the double mutant complex, where it destabilized the binding pocket. FRET was used as a global reporter of binding, whereas Trp quenching reports on the binding pocket. (C) Tryptophan fluorescence curves from thermal dissociation of eRRM, ΔN and ΔC in complex with A32C/G34U SL1 collected from 5 to 63°C in same buffer conditions as B but with higher salt concentration (500 mM KCl). The signal (y-axis) for ΔN complex is plotted on the right hand side so that the unbinding transition is clearly visible. (D) Plot summarizes the extent of high salt destabilization of the protein complexes with wild type and A32C/G34U SL1. The N-helix truncation resulted in a higher degree of destabilization in high salt compared to that when the helix is present.

We can explain the structural origin of this counterintuitive behavior in Figure 7. In the ΔN complex, the linker to the N-helix can sample a wider range of conformations. Particularly, the center-of-mass distance between linker and C32 is on average shorter than the eRRM complex (Figure 7A), and the linker transiently interacts with the RNA repeatedly (Figure 7B). This provides steric hindrance for the nucleotide C32, stabilizing the C32–F148 stacking. In the ΔC complex, the same linker again stabilizes the C32–F148 interaction, but by an entirely different mechanism. In the absence of the disordered C-terminus, the N-helix samples a new conformational ensemble that allows the linker to transiently interact with the unstructured region at the start of the C-terminus (residues 179–203). This new conformation disturbs the interaction between R182 and E179, and E179 is now free to interact with C32 (Figure 7B), contributing to the stability of the C32–F148 interaction.

Total number of hydrogen bonds is unchanged upon ΔN truncation

If protein truncations stabilize the key RRM-RNA stacking interaction, then other effects, such as loss of protein–RNA

hydrogen bonds, could be responsible for the reduced $T_{1/2}$ of the truncated complexes. To investigate this, we looked first at the hydrogen bonds lost when the N-helix is deleted, and next at the effect of the truncations on hydrogen bonding elsewhere between protein and RNA.

We monitored the hydrogen bonding interactions between the N-helix and RNA in the eRRM complex for the duration of a microsecond-long trajectory. The results indicate that at ambient conditions, on average about one hydrogen bond is present between the N-helix and the RNA (Supplementary Figure S17). This finding is superficially in agreement with the reduction of ΔG from ~ 35 to about ~ 28 kJ/mol upon ΔN truncation in Table 1, so one might think that hydrogen bond loss due to helix truncation accounts for the reduced binding in the ΔN complex. However, it is possible that truncating the N-helix also allows the protein to form new hydrogen bonds with the RNA elsewhere. We analyzed this situation by comparing hydrogen bonds in the RRM:RNA binding site upon truncation. Figure 8 shows how the distribution of hydrogen bonds changes with truncations. Specifically, both at ambient and extreme conditions, the mean values of the number of hydrogen bonds are persistently higher for the ΔN complex and statistically

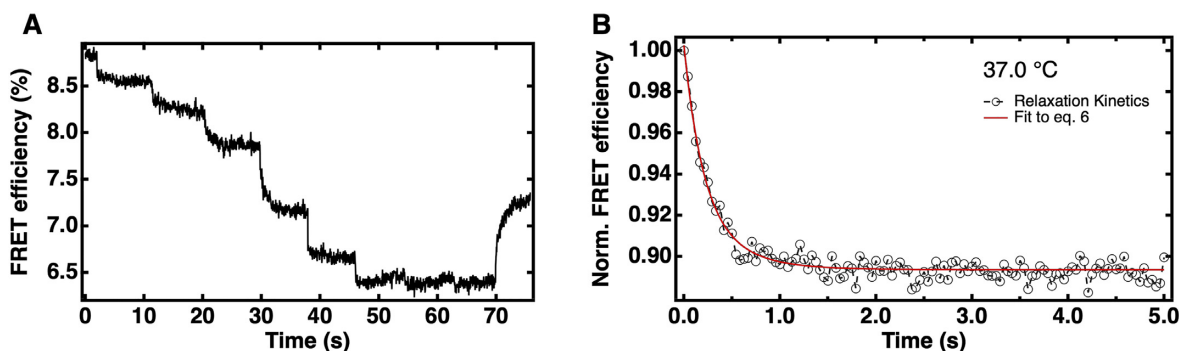


Figure 5. Relaxation kinetics of eRRM-A32C/G34U SL1 interaction from FReI. (A) FRET efficiency of 1 μM complex in response to the rapid temperature jump from 20 to 56°C. (B) Relaxation kinetics of the complex at physiological temperature, 37°C fit to Eq. (6). See Figure S15A for the relaxation kinetics at all temperature jumps from 29°C to 38°C.

Table 2. Kinetic parameters for double RNA mutant complex dissociation in 300 mM KCl. Errors are one standard deviation of the fit. Final temperatures are averages of two replicas, with an uncertainty of $\pm 1.5^\circ\text{C}$

Final temperature	k_{off} (s^{-1})	k_{on} ($\text{M}^{-1} \text{s}^{-1}$)	K_d (M^{-1})
29.2°C	0.05	$4.7 \pm 0.2 \times 10^6$	$1.0 \pm 0.2 \times 10^{-8}$
29.8°C	0.06	$4.3 \pm 0.5 \times 10^6$	$1.4 \pm 0.2 \times 10^{-8}$
33.2°C	0.19	$2.5 \pm 0.2 \times 10^6$	$7.9 \pm 0.2 \times 10^{-8}$
33.5°C	0.27	$2.6 \pm 0.2 \times 10^6$	$8.9 \pm 0.2 \times 10^{-8}$
37.0°C	0.78	$1.5 \pm 0.7 \times 10^6$	$5.2 \pm 0.2 \times 10^{-7}$
38.1°C	1.01	$1.1 \pm 0.4 \times 10^6$	$8.9 \pm 0.2 \times 10^{-7}$

equivalent for the ΔC complex with respect to the eRRM complex. The mean values of the total number of hydrogen bonds for the extended RRM, ΔN , and ΔC complexes at ambient condition are: 3.7 ± 1.3 , 4.5 ± 1.3 , and 3.7 ± 1.3 , respectively, and for the extreme condition are: 3.4 ± 1.3 , 4.2 ± 1.2 , and 3.4 ± 1.2 , respectively. Although truncation of the N-helix loses one direct hydrogen bond to the RNA on average (Supplementary Figure S17), it surprisingly allows about one new hydrogen bond to form in the binding pocket. Thus, hydrogen bond loss cannot really explain the lower stability of the ΔN complex.

The similar ΔG and $T_{1/2}$ in Table 1 for eRRM versus ΔC complex is also consistent with no significant loss or gain of hydrogen bonding following ΔC truncation. ΔC only causes local rearrangements in the binding pocket, as seen in experiment (Figure 4A).

Global conformational change of the truncated complexes

We hypothesize that distributed or allosteric interactions play a role in destabilizing the ΔN complex because neither binding pocket base stacking, nor hydrogen bonding can rationalize the experimentally observed loss of binding affinity upon truncation, at least not within the length of our trajectories. To determine how overall conformation varies between different truncated complexes, we identified the most populated structural state during the simulation trajectories. The identified characteristic structures for the eRRM, ΔN and ΔC complexes are illustrated in Supplementary Figure S19. eRRM and ΔC complexes show similar overall conformations and root mean square deviation of the backbone atoms of the entire complex is in the range of 3.4–4.6 Å under all different conditions. However, in the

ΔN complex the RNA is significantly more flexible, with average root mean square fluctuations of 5.6 Å (the eRRM complex is 3.1 Å). Thus, the N-helix affects not just local binding pocket stacking or hydrogen bonding, but also reduces overall conformational fluctuations. To highlight the effect of truncation on the allosteric paths within the complex we have performed community and shortest path analysis (see Materials and Methods for details) (54).

For the eRRM complex at ambient condition (25°C, 300 mM salt), the RNA is divided into three major communities whose motion is correlated (Figure 9). One contains the end of the stem, another the middle of the RNA stem and part of the C-terminal domain, and the third the connection between the RNA loop and part of the RRM binding pocket (Figure 9A). The motion of the N-helix is decoupled from the RNA (Supplementary Figure S20A), although the elastic network model calculations on the crystal structure of the complex show some degree of correlation between the N-helix and RNA (Supplementary Figure S20B). Thus, an interesting result emerges under the most stabilizing conditions: the disordered C-terminus does move in a correlated fashion with part of the RNA stem. Shortest path analysis was performed to identify the signaling pathways within different parts of the complex. These paths reveal what parts of the complex are connected through the most correlated residues (see Materials and Methods for calculation details). Specifically, the shortest path from the end of the RNA stem to the RRM binding site at residue F148 was calculated. For the eRRM complex at ambient condition, the shortest path entirely passes through the RNA (Figure 9B) However, under the extreme condition this pathway is entirely re-routed through the N-helix (Figure 9C), and the same is seen for the ΔC complex (Supplementary Figure S20C). Thus, the N-helix becomes more important when the eRRM–SL1 binding is disrupted by heat or high salt concentration, as happens during a dissociation experiment such as in Figures 4 and 5.

DISCUSSION

Our results show that the eRRM increases the binding over the well-known minimal RRM, particularly when increased temperature, deleted interactions, or ionic strength stress the RNA–protein binding. It is tempting to assume that RNA–protein complex strength is governed by simple

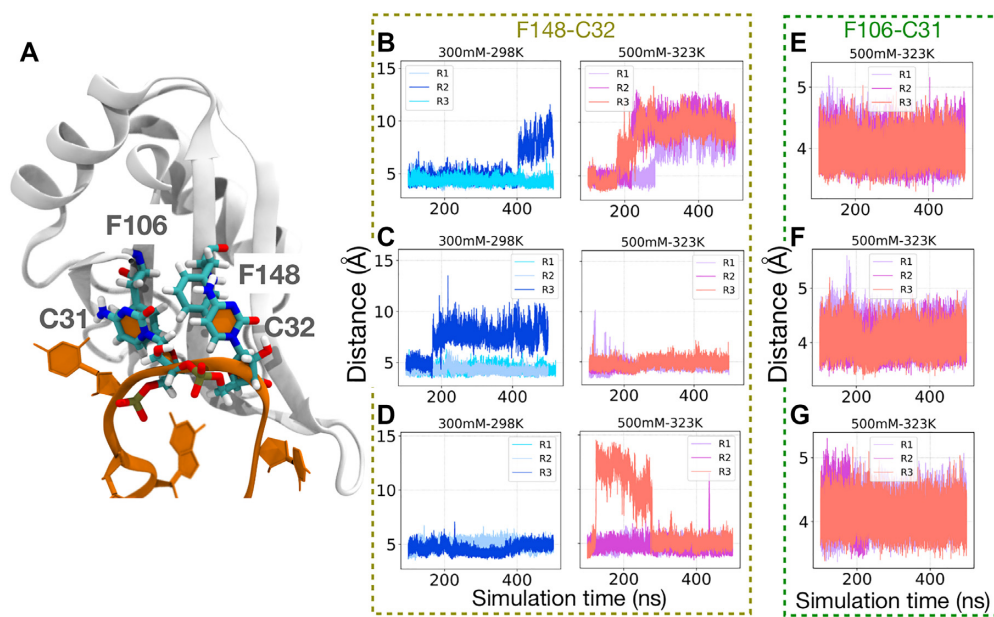


Figure 6. (A) The crucial RRM base-stacking interactions of C31–F106 and C32–F148 within the complex. The distance between the center of mass from C32 to F148 or C31 to F106 in three simulation replicas (R1, R2, and R3) was plotted to monitor their interactions in complexes with different temperatures and salt concentrations for the eRRM complex in (B, E), for ΔC complex in (C, F), and for ΔN complex in (D, G). The C31–F106 base stacking interaction for all protein variations remain fully bound (Supplementary Figure S16), but C32–F148 is least stable in the untruncated complex (see Supplementary Figure S16 for other salt and temperature conditions).

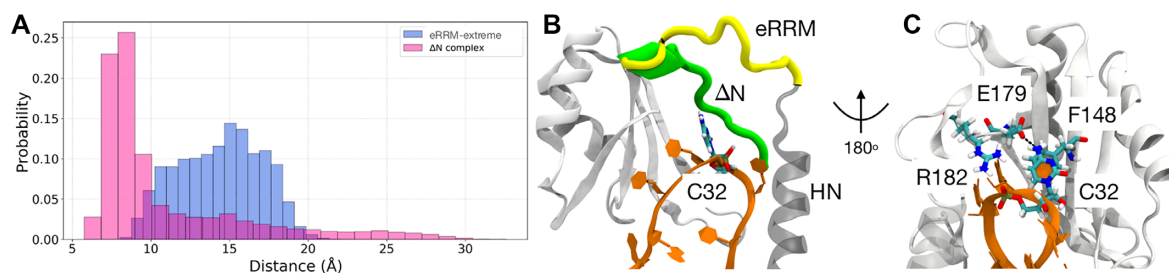


Figure 7. (A) The probability distribution of the distance between the center-of-mass of the linker (connecting active site to N-helix) and C32 backbone for the extreme simulation condition. The same distribution at the ambient condition is shown in SI. (B) The linker (in green) in the ΔN complex at extreme condition can get closer to the RNA than the linker of the eRRM complex (colored in yellow). Therefore, the linker provides steric hindrance for the C32 to increase its distance to F148. (C) The C32–F148 base stacking interaction is stabilized in the ΔC complex by the transient interaction of E179–C32. This interaction results from a conformational change in the linker, which in turn breaks the interaction of R182–E179. Thus, E179 becomes accessible to interact with the base of C32.

rules, such as ‘truncation of the N-helix removes hydrogen bonds and reduces binding.’ In reality, RNA–protein interactions are highly distributed over many types of interactions such as hydrogen bonds, base stacking, and water exclusion from binding sites. Even for RNA binding proteins containing RRM or other classical RNA binding domains, the residues adjacent to these domains can also make a significant contribution towards RNA binding (14–17) thus constituting a fundamental structural unit relevant for biological function. We investigate how the binding of an important RRM, the U1-70K binding site of SL1 RNA, is affected by flanking sequences that make up the eRRM, as was observed in case of the splicing factor U2AF⁶⁵ (61).

In order to study the U1-70K protein–SL1 RNA interactions, we made two types of modifications to the system: truncation of the N-helix or C-IDR flanking the RRM; and a double mutant in the RNA loop to increase the rela-

tive contribution of flanking interactions. The U1-70K sequence we study is highly conserved from humans to yeast (Supplementary Table S6, Figure S21) (28,62) and contains all regions known to stabilize the protein–RNA interaction, which excludes the disordered regions (residues 1–59 and residues 216–437) known only to interact with co-factors of U1 snRNP. One of the RNA mutations (position 32) is part of the RRM binding site, the other (position 34) lies at the beginning of the RNA stem. Based on the MD simulations, the two do not share any protein side chains upon binding, so their additive effect on the stabilization free energy (Table 1 and Supplementary Table S3) is plausible and shows that the binding mechanism is not altered although binding is weakened. The observation that both mutations have similar individual effects immediately highlights that protein–RNA interactions are distributed over the RNA, not just in the RRM binding pocket.

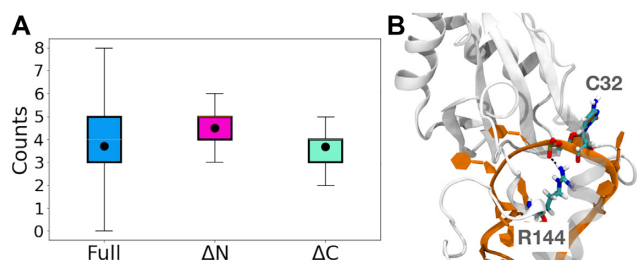


Figure 8. (A) The distribution of the total number of hydrogen bonds defining the RRM:RNA binding site of eRRM:RNA at ambient (25°C, 300 mM salt) conditions for the different protein truncations shown as box plots with the black dot showing the mean values. While the mean values for the eRRM and ΔC complexes are similar, the mean value for the ΔN complex is higher. For other conditions see Supplementary Figure S18. (B) The hydrogen bond between the side chain of R144 and backbone of C32, that is formed in the ΔN complex and contributes to the higher mean value of hydrogen bonds for this complex with respect to the eRRM complex.

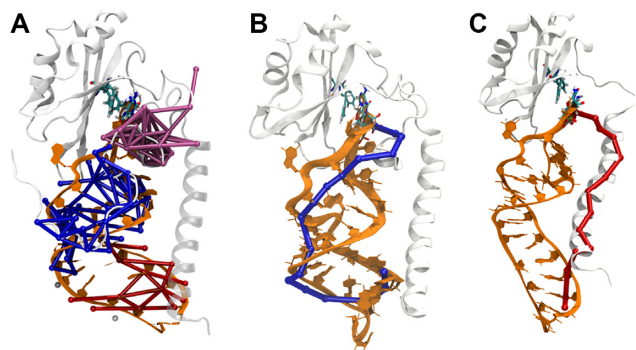


Figure 9. (A) At ambient condition, RNA motion is correlated in three communities: the bottom of the stem (red), the middle of the stem and C-terminal eRRM residues (blue), and the binding loop and binding pocket residues (pink). (B, C) The shortest signaling path connecting the end of the RNA stem to the RRM base stacking interaction reroutes (blue to red) when the temperature and salt concentration is raised in eRRM complex. The same rerouting is observed for the ΔC complex, see Supplementary Figure S20B.

When RNA and protein interact, the dissociation constant can be increased via two kinetic mechanisms: either association is slowed down, or the complex is destabilized and dissociation speeds up. At 300 mM KCl concentration, we saw that a temperature stress increases k_{off} by a factor of 20 and decreases k_{on} by a factor of 4, together increasing K_d . The simulations of the bound complex at higher temperature and salt concentration show higher disorder that could lead to enhanced dissociation under stresses such as increased salt concentration and temperature, as well as truncation of the regions flanking the RRM.

We find that the C-IDR is less important for overall stability than the N-helix. When binding to wild-type RNA, there is no significant change in the dissociation midpoint temperature $T_{1/2}$ with respect to the eRRM, and K_d increases only about three-fold (Table 1) when detected by FRET, a global measure of RNA–protein binding. However, binding studies with the double mutant RNA reveal important contributions from the U1-70K C-IDR to the native structure of the active site: while overall RNA binding is not strongly affected, the effective K_d

monitored by tryptophan fluorescence (a local probe of the binding pocket) increases by two orders of magnitude and $T_{1/2}$ decreases by 6°C. The C-IDR is important when the RNA–protein interaction is compromised (double mutant at high salt). In-cell electrostatics could significantly compromise protein–RNA interaction through non-specific binding competition in a similar way (63), making the C-IDR potentially more important in-cell than *in vitro* for retaining proper binding site structure. In contrast, the N-helix is important for both the wild type RNA and double mutant interaction with U1-70K, and shows no difference between global (FRET) or local (Trp) binding signatures.

The mode of action of the C-IDR and N-helix is revealed at the atomistic level by analysis of the MD simulations. Surprisingly, simulations show that both truncations actually stabilize stacking interaction in the binding pocket, for example by allowing previously blocked protein side chains to interact with the RNA. Likewise, even though N-helix truncation on average deletes one hydrogen bond between the extended RRM and SL1 RNA, the truncation also permits an extra hydrogen bond to be formed in the binding pocket. Thus, we observe no net changes in stacking or hydrogen bonding responsible for the binding pocket stabilization by the C-IDR or for the global stabilization of the complex by the N-helix.

Rather, we find through a community analysis that residues in the C-IDR have dynamics correlated with the center of the RNA stem. The most correlated pathway for the protein–RNA motions—the shortest path—starting at the stem and going to the binding site under ambient conditions goes exclusively through the SL1 RNA, but under stress (high temperature, salt concentration) it proceeds through the N-helix. Additionally, the tryptophan fluorescence quenching experiments clearly suggest a higher degree of complex destabilization under temperature and salt concentration stress when N-helix is truncated. These observations therefore suggest that the N-helix acts like a guardian under stress, but less so when protein and RNA interact under homeostatic conditions. RNA–protein interactions are sensitive to cell stress (34–36), so such extra stabilization of the U1 snRNP could become important in those situations.

The dynamic nature of salt bridges in the complex further supports the ‘distributed interaction’ picture. As expected, long-range electrostatics play an important role in holding together and organizing the complex, as was also observed for the U1A-SL2 complex (24). The dissociation experiments (Figure 3B) indicate about seven salt bridge-equivalent electrostatic interactions between U1-70K and RNA (58). This is consistent with the seven salt bridges present in the crystal. Yet the majority of salt bridges present in the crystal structure are not stable throughout the simulation, which may be due to force field inaccuracy or to the dynamic nature of the complex when not confined in a crystal. Among the salt bridges present in the crystal structure, only R191-C32 at the RRM binding site stays intact persistently in all simulation replicas. Instead, two new salt bridges not seen in the crystal structure form in simulations between the protein and RNA and are consistently present in all three simulation replicas: R200-U28 and R109-U30. We believe that these two bridges make a significant con-

tribution to the net seven observed in our experiments, and on average only five of the seven bridges seen in the static crystal structure contribute, highlighting the dynamic and distributed nature of electrostatic interactions.

In closing, we speculate that interactions with the extended RRM, beyond the main RRM, will be found among other spliceosomal complexes, as well as ribosomal complexes: the reason is that while severe stress can disrupt these cellular machines (37,64), they have been found to be particularly active under mild stress situations or pretreatments which induce increased transcription and translation of certain mRNAs (65,66), and therefore their integrity must be insured. While the N-helix and C-IDR flanking the U1-70K RRM may not be critical when all is well under carefully tailored *in vitro* conditions, they could assist integrity of the spliceosome when cell homeostasis is disrupted.

DATA AVAILABILITY

All protein and RNA sequence data are available in the Supplementary Data, as are numerical results of kinetics and denaturation experiments, and more detailed figures summarizing the simulation analysis.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We acknowledge support from the Beckman Institute Theoretical and Computational Biophysics group, and we thank Kathleen Hall for helpful discussions, in particular on SL1 RNA. We thank the reviewers for careful reading of our manuscript and providing insightful comments suggestions on future prospects of the work.

FUNDING

National Science Foundation (NSF) [MCB 2205665 to M.G.]; C.M.D. was supported in part by a postdoctoral fellowship provided by the PFC: Center for the Physics of Living Cells funded by NSF [PHY 1430124], which were also used to purchase some of the plasmids used in this study; XSEDE [TG-MCB180022 to M.G. and Z.G.], to perform simulations on the Comet XSEDE supercomputer. Funding for open access charge: National Science Foundation. *Conflict of interest statement.* None declared.

REFERENCES

- Dreyfuss,G., Kim,V.N. and Kataoka,N. (2002) Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol.*, **3**, 195–205.
- Glisovic,T., Bachorik,J.L., Yong,J. and Dreyfuss,G. (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.*, **582**, 1977–1986.
- Burd,C.G. and Dreyfuss,G. (1994) Conserved structures and diversity of functions of RNA-Binding proteins. *Science*, **265**, 615–621.
- Kishore,S., Luber,S. and Zavolan,M. (2010) Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression. *Brief. Funct. Genomic.*, **9**, 391–404.
- Lunde,B.M., Moore,C. and Varani,G. (2007) RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.*, **8**, 479–490.
- Keene,J.D. and Query,C.C. (1991) Nuclear RNA-binding proteins. In: Cohn,W.E. and Moldave,K. (eds). *Progress in Nucleic Acid Research and Molecular Biology*. Academic Press, Vol. **41**, pp. 179–202.
- Kenan,D.J., Query,C.C. and Keene,J.D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci.*, **16**, 214–220.
- Cléry,A., Blatter,M. and Allain,F.H.-T. (2008) RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.*, **18**, 290–298.
- Oberstrass,F.C., Auweter,S.D., Erat,M., Hargous,Y., Henning,A., Wenter,P., Reymond,L., Amir-Ahmady,B., Pitsch,S., Black,D.L. *et al.* (2005) Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science*, **309**, 2054–2057.
- Dunker,A.K., Brown,C.J., Lawson,J.D., Iakoucheva,L.M. and Obradović,Z. (2002) Intrinsic disorder and protein function. *Biochemistry*, **41**, 6573–6582.
- Lobley,A., Swindells,M.B., Orenge,C.A. and Jones,D.T. (2007) Inferring function using patterns of native disorder in proteins. *PLoS Comput. Biol.*, **3**, e162.
- Walden,W.E., Selezneva,A.I., Dupuy,J., Volbeda,A., Fontecilla-Camps,J.C., Theil,E.C. and Volz,K. (2006) Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. *Science*, **314**, 1903–1908.
- Tsvetanova,N.G., Klass,D.M., Salzman,J. and Brown,P.O. (2010) Proteome-wide search reveals unexpected RNA-binding proteins in *saccharomyces cerevisiae*. *PLoS ONE*, **5**, e12671.
- Ben-Shem,A., Garreau de Loubresse,N., Melnikov,S., Jenner,L., Yusupova,G. and Yusupov,M. (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science*, **334**, 1524–1529.
- Brodersen,D.E., Clemons,W.M., Carter,A.P., Wimberly,B.T. and Ramakrishnan,V. (2002) Crystal structure of the 30 s ribosomal subunit from *thermus thermophilus*: structure of the proteins and their interactions with 16 s RNA1 *J. Mol. Biol.*, **316**, 725–768.
- Klein,D.J., Moore,P.B. and Steitz,T.A. (2004) The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.*, **340**, 141–177.
- Järvelin,A.I., Noerenberg,M., Davis,I. and Castello,A. (2016) The new (dis)order in RNA regulation. *Cell Commun. Signal.*, **14**, 9.
- Weber,S.C. and Brangwynne,C.P. (2012) Getting RNA and protein in phase. *Cell*, **149**, 1188–1191.
- Matera,A.G. and Wang,Z. (2014) A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.*, **15**, 108–121.
- Patton,J.R., Habets,W., van Venrooij,W.J. and Pederson,T. (1989) U1 small nuclear ribonucleoprotein particle-specific proteins interact with the first and second stem-loops of U1 RNA, with the a protein binding directly to the RNA independently of the 70K and sm proteins. *Mol. Cell Biol.*, **9**, 3360–3368.
- Weber,G., Trowitzsch,S., Kastner,B., Lührmann,R. and Wahl,M.C. (2010) Functional organization of the sm core in the crystal structure of human U1 snRNP. *EMBO J.*, **29**, 4172–4184.
- Larson,J.D. and Hoskins,A.A. (2017) Dynamics and consequences of spliceosome e complex formation. *Elife*, **6**, e27592.
- Law,M.J. (2005) Kinetic analysis of the role of the tyrosine 13, phenylalanine 56 and glutamine 54 network in the U1A/U1 hairpin II interaction. *Nucleic Acids Res.*, **33**, 2917–2928.
- Ghaemi,Z., Guzman,I., Gnuett,D., Luthey-Schulten,Z. and Gruebele,M. (2017) Role of electrostatics in Protein–RNA binding: the global vs the local energy landscape. *J. Phys. Chem. B*, **121**, 8437–8446.
- Kormos,B.L., Baranger,A.M. and Beveridge,D.L. (2006) Do collective atomic fluctuations account for cooperative effects? Molecular dynamics studies of the U1A–RNA complex. *J. Am. Chem. Soc.*, **128**, 8992–8993.
- Law,M.J., Rice,A.J., Lin,P. and Laird-Offringa,I.A. (2006) The role of RNA structure in the interaction of U1A protein with U1 hairpin II RNA. *RNA*, **12**, 1168–1178.
- Query,C.C., Bentley,R.C. and Keene,J.D. (1989) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell*, **57**, 89–101.
- Kondo,Y., Oubridge,C., van Roon,A.-M.M. and Nagai,K. (2015) Crystal structure of human U1 snRNP, a small nuclear

- ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. *life*, **4**, e04986.
29. Stejskalová, E. and Staněk, D. (2014) The splicing factor U1-70K interacts with the SMN complex and is required for nuclear gem integrity. *J. Cell Sci.*, **127**, 3909–3915.
 30. Cho, S., Hoang, A., Sinha, R., Zhong, X.-Y., Fu, X.-D., Krainer, A.R. and Ghosh, G. (2011) Interaction between the RNA binding domains of Ser-arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc. Nat. Acad. Sci. U.S.A.*, **108**, 8233–8238.
 31. Cao, W. and Garcia-Blanco, M.A. (1998) A Serine/Arginine-rich domain in the human U1 70k protein is necessary and sufficient for ASF/SF2 binding *. *J. Biol. Chem.*, **273**, 20629–20635.
 32. Xiao, S.H. and Manley, J.L. (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev.*, **11**, 334–344.
 33. Aubol, B.E., Wozniak, J.M., Fattet, L., Gonzalez, D.J. and Adams, J.A. (2021) CLK1 reorganizes the splicing factor U1-70K for early spliceosomal protein assembly. *Proc. Nat. Acad. Sci. U.S.A.*, **118**, e2018251118.
 34. Kedersha, N.L., Gupta, M., Li, W., Miller, I. and Anderson, P. (1999) RNA-Binding proteins tia-1 and tiar link the phosphorylation of Eif-2 α to the assembly of mammalian stress granules. *J. Cell Biol.*, **147**, 1431–1442.
 35. Guzman, I., Ghaemi, Z., Baranger, A., Luthey-Schulten, Z. and Gruebele, M. (2015) Native conformational dynamics of the spliceosomal U1A protein. *J. Phys. Chem. B*, **119**, 3651–3661.
 36. Nishiyama, H., Itoh, K., Kaneko, Y., Kishishita, M., Yoshida, O. and Fujita, J. (1997) A Glycine-rich RNA-binding protein mediating cold-inducible suppression of mammalian cell growth. *J. Cell Biol.*, **137**, 899–908.
 37. Bond, U. (2006) Stressed out! Effects of environmental stress on mRNA metabolism: effects of environmental stress on mRNA metabolism. *FEMS Yeast Res.*, **6**, 160–170.
 38. Kiskey, L., Miller, K.A., Guin, D., Kong, X., Gruebele, M. and Leckband, D.E. (2017) Direct imaging of protein stability and folding kinetics in hydrogels. *ACS Appl. Mater. Inter.*, **9**, 21606–21617.
 39. Robinson, K.E., Orans, J., Kovach, A.R., Link, T.M. and Brennan, R.G. (2014) Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. *Nucleic Acids Res.*, **42**, 2736–2749.
 40. Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S. and Baker, D. (2020) Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 1496–1503.
 41. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L. and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.
 42. Phillips, J.C., Hardy, D.J., Maia, J.D.C., Stone, J.E., Ribeiro, J.V., Bernardi, R.C., Buch, R., Fiorin, G., Henin, J., Jiang, W. *et al.* (2020) Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.*, **153**, 044130.
 43. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
 44. Jorgensen, W.L. and Madura, J.D. (1983) Quantum and statistical mechanical studies of liquids. 25. Solvation and conformation of methanol in water. *J. Am. Chem. Soc.*, **105**, 1407–1413.
 45. Huang, J. and MacKerell, A.D. (2013) CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.*, **34**, 2135–2145.
 46. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmüller, H. and MacKerell, A.D. (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods*, **14**, 71–73.
 47. Ghaemi, Z., Guzman, I., Baek, J.J., Gruebele, M. and Luthey-Schulten, Z. (2016) Estimation of relative Protein-RNA binding strengths from fluctuations in the bound state. *J. Chem. Theory Comput.*, **12**, 4593–4599.
 48. Martyna, G.J., Tobias, D.J. and Klein, M.L. (1994) Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*, **101**, 4177–4189.
 49. Feller, S.E., Zhang, Y., Pastor, R.W. and Brooks, B.R. (1995) Constant pressure molecular dynamics simulation: the langevin piston method. *J. Chem. Phys.*, **103**, 4613–4621.
 50. Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald: an $\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
 51. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. and Pedersen, L.G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.*, **103**, 8577–8593.
 52. Girvan, M. and Newman, M.E.J. (2002) Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 7821–7826.
 53. Floyd, R.W. (1962) Algorithm 97: shortest path. *Commun. ACM*, **5**, 345.
 54. Sethi, A., Eargle, J., Black, A.A. and Luthey-Schulten, Z. (2009) Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 6620–6625.
 55. Förster, Th. (1948) Zwischenmolekulare energiewanderung und fluoreszenz. *Ann. Phys.*, **437**, 55–75.
 56. Chen, Y. and Barkley, M.D. (1998) Toward understanding tryptophan fluorescence in proteins. *Biochemistry*, **37**, 9976–9982.
 57. Ryschon, T.W., Rosenstein, D.L., Rubinow, D.R., Niemela, J.E., Elin, R.J. and Balaban, R.S. (1996) Relationship between skeletal muscle intracellular ionized magnesium and measurements of blood magnesium. *J. Lab Clin. Med.*, **127**, 207–213.
 58. Record, M.T., Lohman, T.M. and Haseth, deP. (1976) Ion effects on ligand-nucleic acid interactions. *J. Mol. Biol.*, **107**, 145–158.
 59. Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K.W., Li, J. and Nagai, K. (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature*, **458**, 475–480.
 60. Ervin, J., Larios, E., Osváth, S., Schulten, K. and Gruebele, M. (2002) What causes hyperfluorescence: folding intermediates or conformationally flexible native states? *Biophys. J.*, **83**, 473–483.
 61. Sarangi, M.K., Zvoda, V., Holte, M.N., Becker, N.A., Peters, J.P., Maher, L.J. and Ansari, A. (2019) Evidence for a bind-then-bend mechanism for architectural DNA binding protein yNhp6A. *Nucleic Acids Res.*, **47**, 2871–2883.
 62. Fan, T., Zhao, Y.-Z., Yang, J.-F., Liu, Q.-L., Tian, Y., Debatosh, D., Liu, Y.-G., Zhang, J., Chen, C., Chen, M.-X. *et al.* (2021) Phylogenetic comparison and splice site conservation of eukaryotic U1 snRNP-specific U1-70K gene family. *Sci Rep*, **11**, 12760.
 63. Davis, C.M. and Gruebele, M. (2021) Cellular sticking can strongly reduce complex binding by speeding dissociation. *J. Phys. Chem. B*, **125**, 3815–3823.
 64. Yost, H.J. and Lindquist, S. (1991) Heat shock proteins affect RNA processing during the heat shock response of *Saccharomyces cerevisiae*. *Mol Cell Biol*, **11**, 1062–1068.
 65. Tan, S., Wang, W., Tian, C., Niu, D., Zhou, T., Jin, Y., Yang, Y., Gao, D., Dunham, R. and Liu, Z. (2019) Heat stress induced alternative splicing in catfish as determined by transcriptome analysis. *Comp. Biochem. Physiol. Part D: Genomic. Proteomic.*, **29**, 166–172.
 66. Lu, Y., Zhao, P., Zhang, A., Ma, L., Xu, S. and Wang, X. (2020) Alternative splicing diversified the heat response and evolutionary strategy of conserved heat shock protein 90s in hexaploid wheat (*Triticum aestivum* L.). *Front. Genet.*, **11**, 1471.