PNAS
www.pnas.org

Supplementary Information

An autoinhibitory intramolecular interaction proof-reads RNA recognition by the essential splicing factor U2AF2

Hyun-Seo Kang[a,b,1], Carolina Sánchez-Rico[a,b,1], Stefanie Ebersberger[c], F. X. Reymond Sutandy[c], Anke Busch[c], Thomas Welte[d], Ralf Stehle[b], Clara Hipp[b], Laura Schulz[c], Andreas Buchbender[c], Kathi Zarnack[e], Julian König[c,2], and Michael Sattler[a,b,2]

[2]Correspondence may be addressed to:
Email: j.koenig@imb-mainz.de or Email: sattler@helmholtz-muenchen.de

This PDF file includes:

Supplementary text
Tables S1 to S2
Figures S1 to S12
SI References

Supplementary text

Delineating domain boundaries of RRM1,2

We re-evaluated the N-terminal boundary of RRM1,2 by comparing four RRM1,2 constructs with varying size (residues 88-342, 124-342, 140-342, 148-342) using ITC and NMR (**Table S1, Figure S2**). The N-terminal extension by eight additional residues (construct 140-342) of U2AF2 RRM1,2 leads to virtually identical chemical shifts for amide signals as seen in a larger construct comprising residues 88-342 (**Figure S2**), suggesting comparable structural and dynamical features. Correspondingly, the construct 140-342, which is referred to as RRM1,2 from now on, shows RNA binding affinity for U9 RNA equivalent to that of construct 88-342 by ITC (**Figure S2**). When comparing NMR spectra of various U2AF2 RRM1,2 constructs, we observed significant chemical shift differences on the resonances that mainly affect RRM2, i.e. comparison of RRM1,2 (residues 140-342) with RRM1 (residues 140-237) and RRM2 (residues 258-342) (**Figure 1D**). Intriguingly, the residues in RRM2 which show strongest chemical shift differences are clustered on the β-sheet adjacent and partially overlapping with the RNA binding interface (**Figure 1E**), suggesting a potential role of the linker for RNA binding. However, no obvious chemical shift differences are observed in the case of RRM1. This indicates the absence of significant interactions between RRM1 and the linker. Comparison of NMR spectra of RRM1-linker (residues 140-261) and linker-RRM2 (residues 231-342) indicate that the N-terminal region of the linker, especially residues 237-243, is independent of RRM1 and RRM2 domains as no amide chemical shift perturbations are observed in the absence of one of the domains (**Figure S3**).

Comparison with previously reported structures of U2AF2 RRM1,2

Figure S4D shows a superposition of the new *apo* structure presented in this work with the two previously published crystal structures of RRM1,2, when bound to strong poly-U RNA ligands (1, 2). Note that the solution structures of the free and RNA-bound RRM1,2 previously reported by Mackereth et al (3) are not included in the comparison as they were based on the domain structures or RRM1 and RRM2 reported in Sickmier et al (2), and are thus basically identical considering the individual domain structures. Therefore, as in Sickmier et al (2), in which the RRM1,2 linker had been deleted in the protein crystallized, no structural information for the linker was available or considered by Mackereth et al either(3).

Key observations reported by Mackereth et al (3) were: (i) The tandem RRM1,2 domain arrangements bound to RNA, that had been reported by Sickmier et al (2) do not report on the solution conformation and rather result from crystal packing artifacts. The crystal structure neither reflects the solution conformation when bound to RNA nor the *apo* RRM1,2 domain arrangement. In fact, the *apo* structure has to be represented by an ensemble of structures that comprises closed, open and “detached” domain arrangements, in which the RRM domains are not in close spatial proximity (4). (ii) The U2AF2 RRM1,2 tandem domain arrangements reflect a conformational equilibrium between closed (*apo*) and open (bound to strong Py-tract RNA) arrangements of the RRM1 and RRM2 domains, which is shifted depending on the “strength” (i.e. RNA binding affinity) of the RNA ligand. Further evidence for this equilibrium was reported by us using FRET experiments (5). The structure of the RRM1,2 complex bound to a high-affinity Py-tract RNA reported by Agrawal et al (1) comprises extended RRM1 and RRM2 domain constructs that we have also used in the present work. The structure by Agrawal et al confirms the open domain arrangement that we have reported previously (3). It is important to note that none of the crystallographic structures revealed the functionally important dynamics that involves the open/closed equilibrium of domain arrangements. This also precluded the role of the linker, which we

show in the current manuscript to be a dynamic structural element that plays a key role in defining the RNA binding specificity of U2AF2.

Effect of linker mutations on the RRM1,2 conformation and RNA binding

RRM1 alone binds to Py-tract RNA significantly weaker (K_D > 300 μM) compared to RRM2 (K_D > 5 μM) (3). However, as a tandem, RRM1,2 cooperatively bind to RNA with a significantly enhanced binding affinity to strong Py tracts. In the RNA complex, both domain adopt a side-by-side arrangement, which is weakly populated but selected from a large conformational space that the two domains adopt in the absence of RNA (4).

Next, we examined how mutations in the linker, which affect the conformational dynamics of the domains, affect the RNA binding of RRM1,2. First, we introduced various single point mutations in the linker to identify the key residues for disrupting the linker/RRM2 interaction, while minimizing the mutations in the linker sequence. NMR amide chemical shifts of RRM1,2 V254G resemble those of RRM2 lacking the linker. This means that key residues that are involved in interactions with the linker in RRM1,2-WT, now exhibit chemical shifts similar to those of the isolated RRM2 or RRM1,2-GS. Furthermore, when we introduced a paramagnetic spin label at the residue 315 in RRM2 of the RRM1,2 V254G, the signature inter-domain PREs (paramagnetic relaxation enhancement) in RRM1, as described in our previous study (4), are significantly weakened due to the removal of the linker/RRM2 interaction, similar to what is observed with RRM1,2-GS. However, no significant changes in the RNA binding properties are found upon the V254G mutation. Taken together, these data indicate that, while V254 is a hotspot establishing numerous contacts with RRM2, the complete C-terminal region of the linker is important to modulate the RNA binding affinity of RRM1,2.

A recent study (1) also proposed a role of the U2AF2 RRM1,2 domain linker. This was based on an initial series of linker-only mutations in combination with fluorescence anisotropy-based binding measurements. Using the *AdML* Py-tract (5'-CCCUUUUUUUUCC-3') that mimics the strong Py-tract U9 RNA used in the present study, the authors found a 2-3-fold reduction in binding affinity upon mutating 12 linker-residues to glycine and V254 to proline. In our study, we do not observe an obvious change in binding affinity for the strong Py-tract U9 RNA, even when we introduced GS mutations for the 25 linker residues, including V254P. Therefore, it is less plausible that the linker-to-RNA contacts, seen in the crystal structure, play a significant role in stabilizing or reinforcing RNA binding. Notably, when bound to U9 RNA, NMR chemical shifts of the V254 amide signal do not exhibit any significant chemical shift perturbation, arguing against strong contacts in this region. Further, the previous study observed the largest reduction in binding affinity only when a single linker mutation (V254P) was combined with two mutations that affect the additional N-terminal helix α0 (V147A) and a proximal residue (R227A in the C-terminal end of RRM1 adjacent to the new helix), showing comparable binding affinity (5.6 μM) to that of their previous shorter RRM1,2 construct missing the N- and C-terminal helices (3.6 μM). In light of our NMR data and the structural and biochemical analyses, the reduction in affinity is likely explained by reduced contributions of the flanking helices and rather than the RRM1,2 linker. Most likely, the linker/RRM1 interaction that stabilizes binding of strong Py-tract RNAs and the dynamic linker/RRM2 autoinhibition mediate the binding preference of U2AF2 for long continuous Py-tracts, by aiding the juxtaposition of the two domains on the RNA and preventing interactions with weaker, non-cognate RNA sites, respectively.

Binding kinetics in NMR titration experiments

The spectral changes of individual NMR signals during the titration experiments reflect the binding kinetics (i.e. off-rates) of the interactions (**Figure 2C; Figure S6, S7**). When comparing titrations of

RRM1,2-WT and -GS with weak Py-tracts, the titration behavior of RNA-binding residues changes from a fast/intermediate regime for RRM1,2-WT to an intermediate/slow regime for RRM1,2-GS, specifically for RRM2 (**Figure 2C; Figure S6A**). In contrast, the titration behavior and binding kinetics for signals in the region of RRM1 are comparable for RRM1,2-WT and -GS (**Figure S7**). Together, this suggests that absence of the linker/RRM2 interaction results in slower off-rates and an increased binding strength of RRM2 for weak Py-tracts, consistent with our switchSENSE and ITC data.

Supplementary Tables

Table S1 - Isothermal titration calorimetry

U2AF2	RNA	N	K_D (μM)	# of repeats
RRM1,2 short	**U9**	1.0 ± 0.1	1.23 ± 0.02	3
RRM1,2	**U4A8U4**	0.8 ± 0.3	4.048 ± 0.825	3
RRM1,2	**U9**	0.9 ± 0.2	0.185 ± 0.035	10
RRM1,2 (GS)	**U4A8U4**	0.7 ± 0.3	1.235 ± 0.376	4
RRM1,2 (GS)	**U9**	0.7 ± 0.2	0.170 ± 0.038	3
231-342	**U4A8U4**	0.9 ± 0.4	94.150 ± 1.344*	2
231-342	**U9**	0.6 ± 0.2	16.400 ± 4.770	3
258-342	**U4A8U4**	0.8 ± 0.3	9.633 ± 3.357	3
258-342	**U9**	0.5 ± 0.2	4.447 ± 0.351	3
88-342	**U4A8U4**	0.6 ± 0.3	3.485 ± 0.035	2
88-342	**U9**	1.0 ± 0.2	0.237 ± 0.054	3
88-342 (GS)	**U4A8U4**	0.7 ± 0.1	1.330 ± 0.105	1
88-342 (GS)	**U9**	0.7 ± 0.2	0.207 ± 0.054	2

Table S1. Binding affinities and stoichiometries determined in ITC experiments for various U2AF2 constructs with RNA ligands indicated are shown. "*" indicates significantly weaker binding with difficulty in reaching the saturation point.

Table S2 – Structural statistics

Experimental restraints			
Distance restraints			
Total NOE			3638
Total NOE RRM1/RRM2			1675/1650
Short-range	$\lvert i-j \rvert \leq 1$		1870
Medium-range	$1 < \lvert i-j \rvert < 5$		610
Long-range	$\lvert i-j \rvert \geq 5$		1149
Dihedral angle restraints (from TALOS)			
ϕ			140
ψ			140
Structure statistics			
r.m.s.d. from experimental restraints (mean and s.d.)			
Distance restraints (Å), no violation >0.5 Å			0.018 ± 0.001
Dihedral angle restraints(°, no violation >0.5°)			0.41 ± 0.05
Deviations from idealized geometry			
Bond lengths (Å)			0.00351 ± 0.00009
Bond angles (°)			0.47 ± 0.01
Impropers (°)			1.35 ± 0.06
Average pairwise r.m.s.d.** (Å)			
		backbone	heavy atoms
RRM1		0.45 ± 0.09	1.08 ± 0.11
RRM2		0.42 ± 0.06	1.07 ± 0.08

Table S2. Pairwise r.m.s.d. was calculated for secondary structure regions in the 10 lowest energy structures of U2AF2 (RRM1: 150-155, 162-178,186-191,197-201, 206-214, 218-227; RRM2: 260-264, 272-281, 286-292, 301-305, 308-320, 330-334) after water refinement using MolMol.
Ramachandran plot: 81.7%, 18.2% and 0.1% of residues are found in the most favored, additionally allowed and generously allowed regions, respectively. No residues are found in disallowed regions.

Supplementary Figures

Figure S1

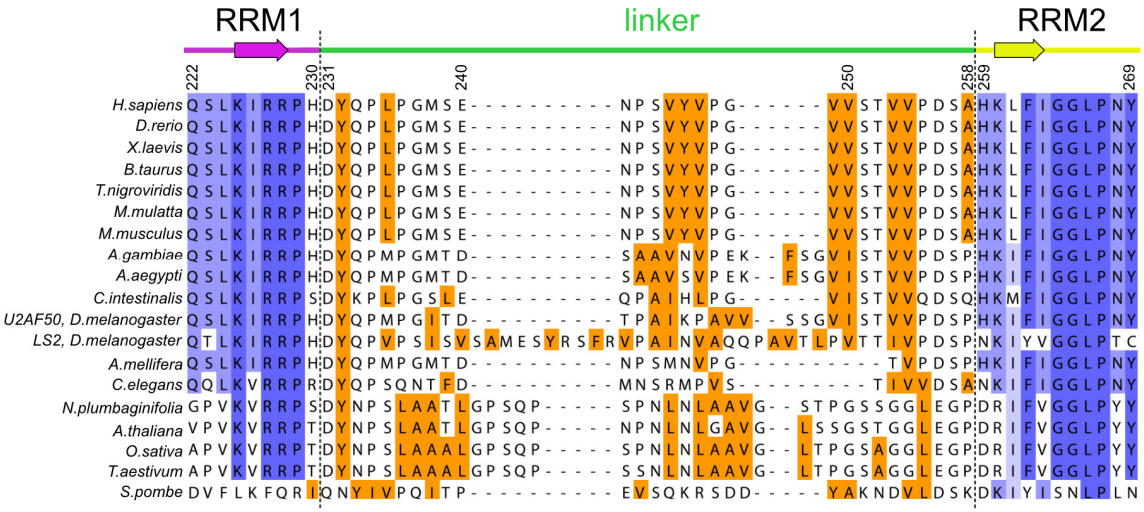


Figure S1. Multiple sequence alignment of U2AF2 orthologs. The length and the presence of multiple hydrophobic aliphatic and aromatic residues (orange) in the RRM1-RRM2 linker (residues 231-258) are conserved in U2AF2 orthologs from fungi to human. Residue numbers correspond to human U2AF2 (see Methods for the details).

Figure S2

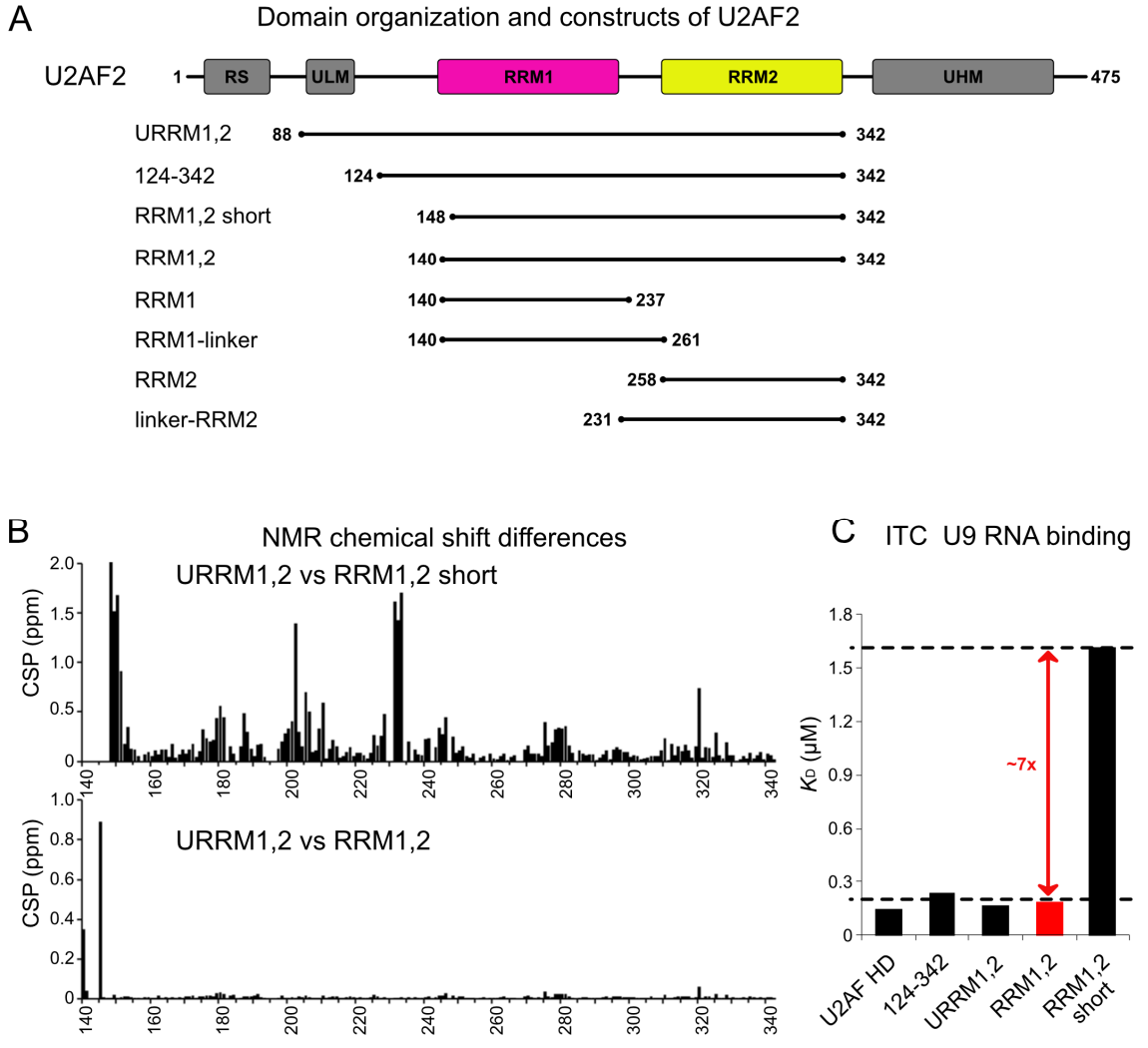


Figure S2. The new RRM1,2 region (residues 140-342) captures the complete RNA binding affinity of U2AF2. Comparison of U9 Py-tract RNA binding to various constructs of U2AF2. **(A)** Various U2AF2 constructs comprising the RRM1 and/or RRM2 domains, the connecting linker, and/or the U2AF1-binding ULM region (in URRM1,2) as indicated. **(B)** Chemical shift differences seen in NMR titrations comparing URRM1,2 with the previously used construct RRM1,2 short (top) and RRM1,2 (bottom). **(C)** ITC experiments for the binding of U2AF constructs to U9 RNA. HD, minimal heterodimer, comprised of URRM1,2 and the UHM domain of U2AF1.

Figure S3

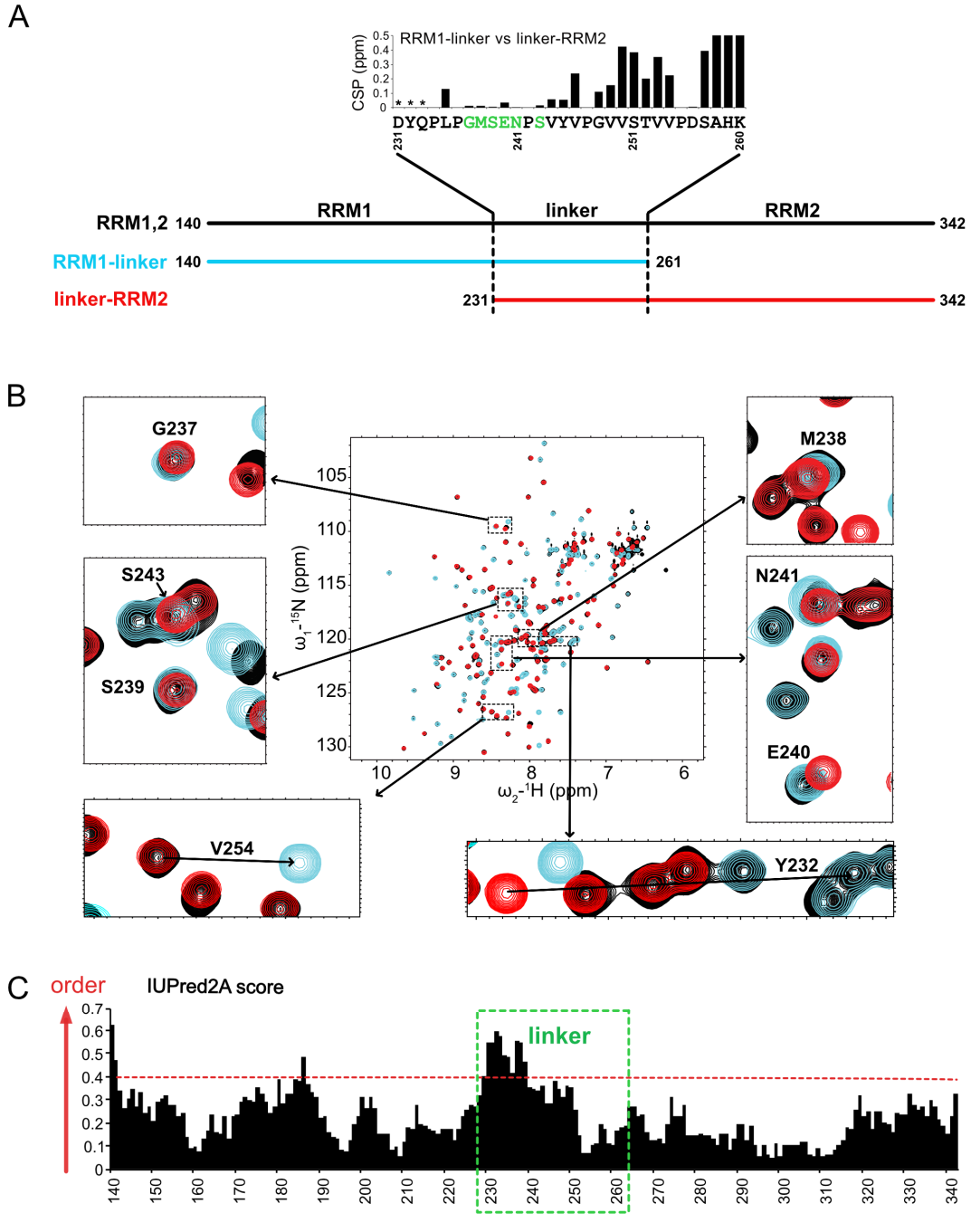


Figure S3. Identification of disordered residues in the N-terminal region of the U2AF2 RRM1-RRM2 linker. (A) Various regions of U2AF2 (RRM1,2, RRM1-linker, linker-RRM2; see **Figure S2**) were studied. Comparison of NMR spectra indicates intrinsically disordered residues (green font) in the N-terminal region of the RRM1-RRM2 linker, based on the fact that the corresponding NMR signals are superimposable in the three spectra. **(B)** Significant NMR chemical shift changes of RRM1- or RRM2-interacting residues, Y232 and V254, are shown in comparison to disordered residues, where only very small spectral differences are observed. **(C)** The region of residues 230-240 in the RRM1-RRM2 linker (green box) is predicted to be disordered (IUPred2A score > 0.4) based on IUPred2A (6).

Figure S4

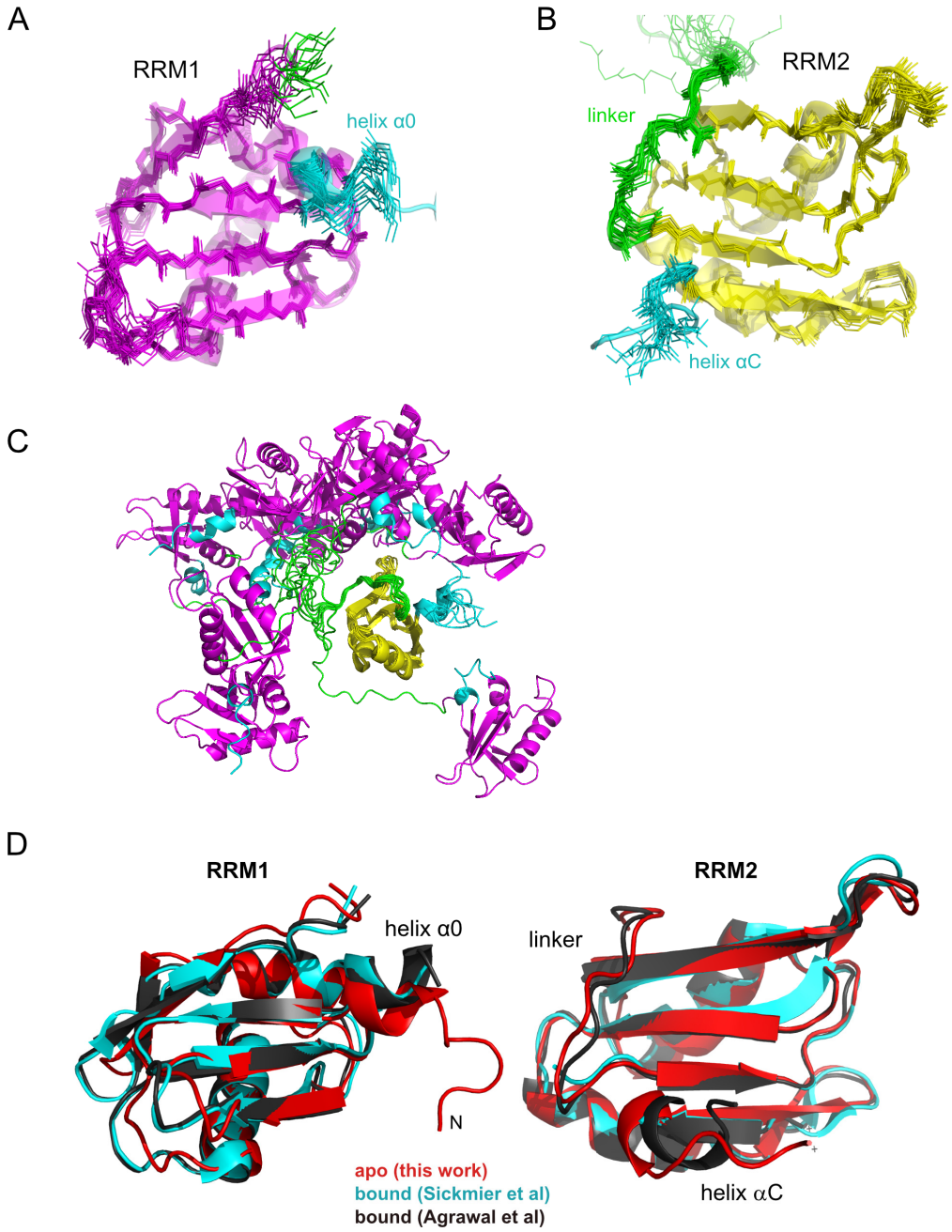


Figure S4. NMR ensemble resulting from the structure calculation of U2AF2 RRM1,2 (residues 140-342). **(A,B)** Structural ensembles individually superimposed for RRM1 (A) and RRM2 (B) are well converged. The N-terminal helix α0 of RRM1 and the C-terminal helix αC of RRM2 are shown in cyan. The RRM1-RRM2 linker is shown in green. **(C)** The relative domain arrangement of RRM1 (magenta) and RRM2 (yellow) are not defined relative to each other. The RRM1-RRM2 linker is shown in green. **(D)** Structural comparisons of RRM1 and RRM2 with previously reported structures of U2AF2, including free RRM1,2 (red, this study, 2tr0.pdb), short RNA-bound RRM1,2 (cyan, 2g4b.pdb; ref. (2)), and extended RNA-bound RRM1,2 (black, 5ev1.pdb, ref. (1))

Figure S5

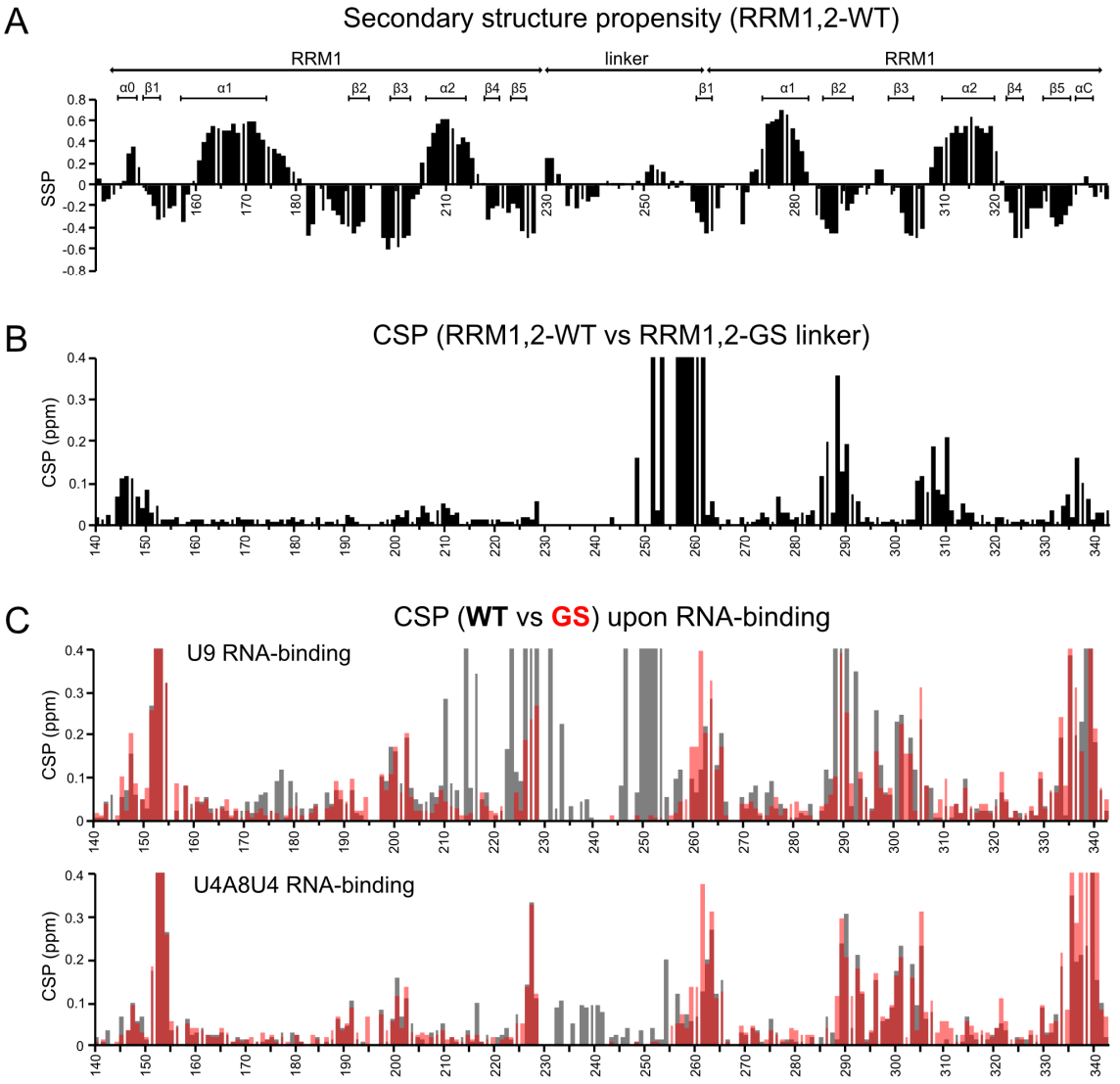


Figure S5. NMR derived secondary structure and chemical shift changes associated with linker mutation and RNA. (A) Backbone chemical shift-derived secondary structure of free RRM1,2 confirms the additional flanking helical elements RRM1 helix α0 and RRM2 helix αC. **(B)** Amide chemical shift differences comparing unbound RRM1,2-WT against the RRM1,2-GS highlight the regions in RRM2 that are affected by the wildtype linker. **(C)** Amide chemical shift perturbation (CSP) in the presence of two-fold molar excess of strong (U9, above) or weak (U4A8U4, below) Py-tract RNAs to RRM1,2-WT (black) or the Gly-Ser-linker variant (RRM1,2-GS, red). Much stronger CSPs are observed for the C-terminal region of the linker upon binding to U9 RNA, as the U9 RNA but not the weaker U4A8U4 RNA ligand can effectively compete with the wildtype linker interaction, see also **Figure 3D**. Canonical binding sites at the RRM1 and RRM2 β-sheets are comparable.

Figure S6

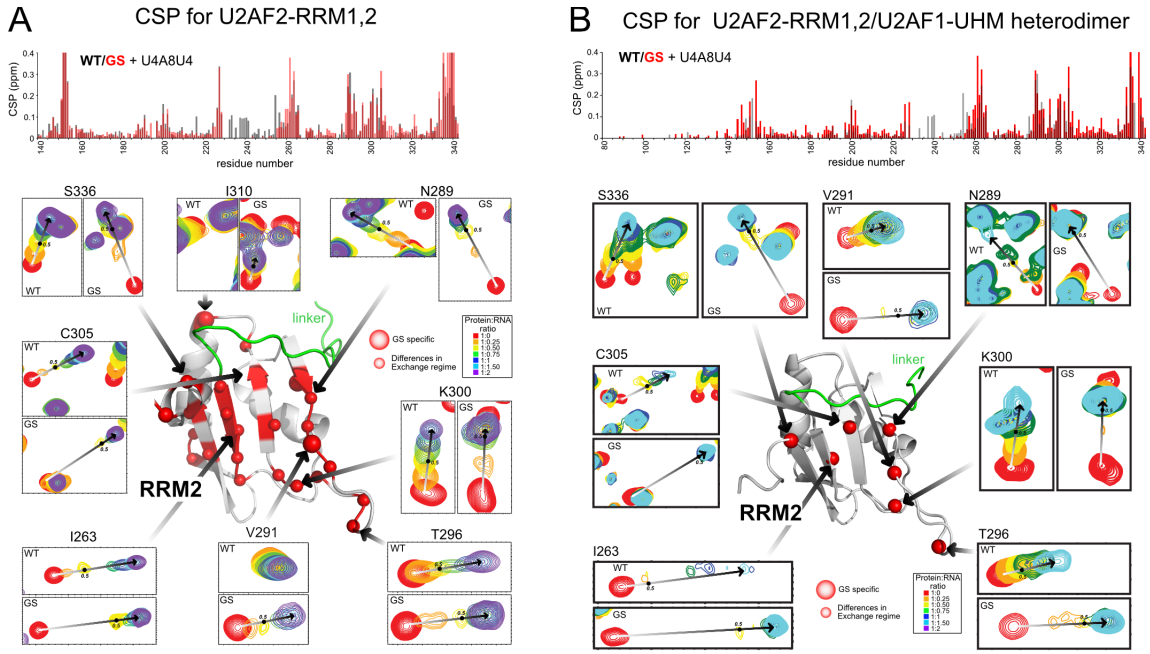


Figure S6. Comparison of NMR titrations involving RRM1,2-WT and -GS with the weak Py-tract RNA (U4A8U4). (A) Comparison of NMR titrations of RRM1,2-WT and -GS with the weak Py-tract RNA U4A8U4. For residues in RRM2, reduced off-rates (higher binding affinity) are indicated by the exchange regime of the NMR titrations, which changes from fast/intermediate towards intermediate/slow exchange behavior in the GS mutant. **(B)** Comparison of NMR titrations of the minimal U2AF heterodimer comprising U2AF2 RRM1,2-WT and -GS and U2AF1-UHM (residues 39-152) with the weak Py-tract RNA (U4A8U4). Chemical shift perturbation plots (above) of U2AF2-RRM1,2 and U2AF minimal heterodimer with WT (black) or GS (red) linker upon adding two-fold excess of the weak Py-tract RNA U4A8U4.

Figure S7

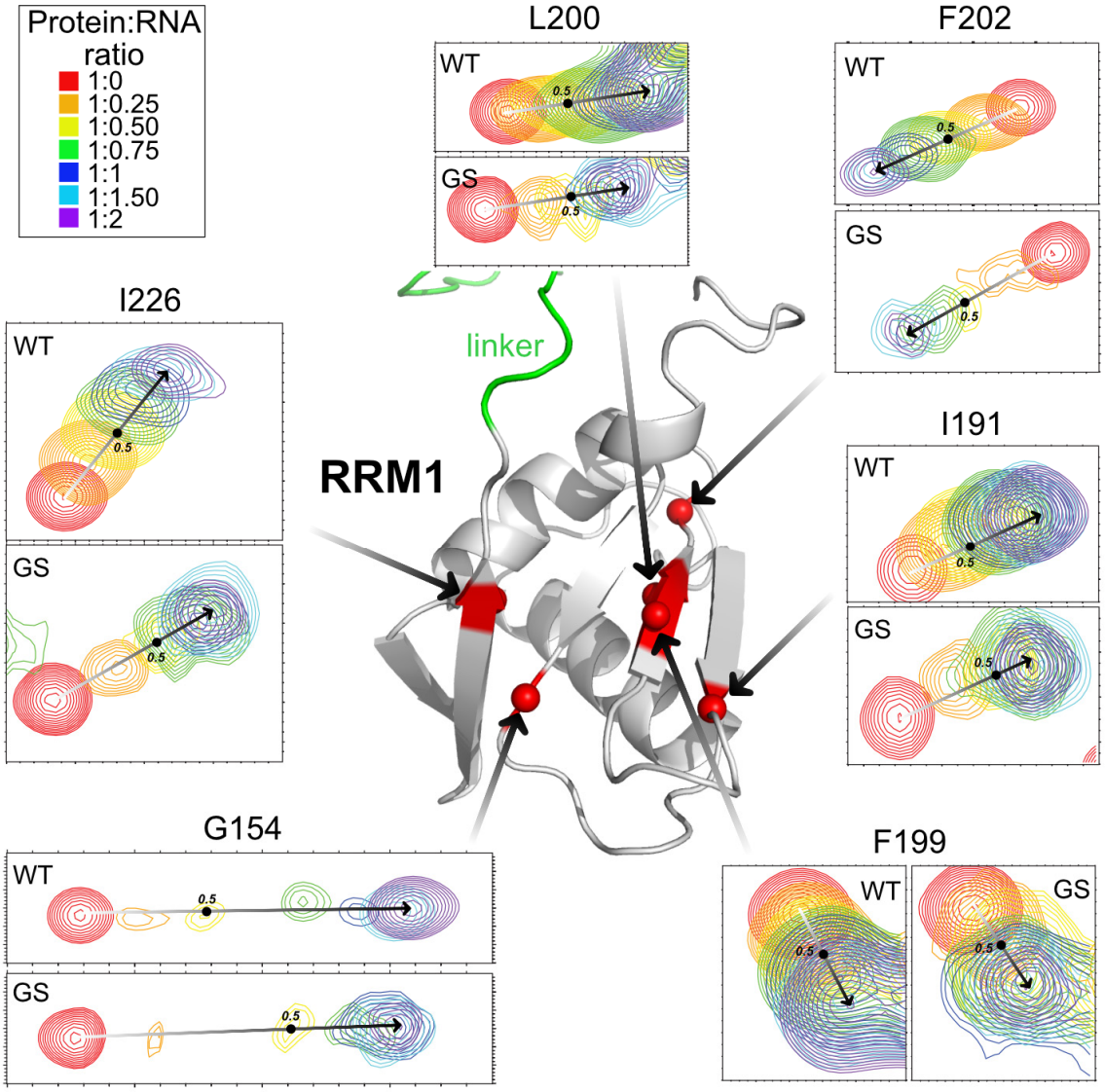


Figure S7. Spectral changes and binding kinetics upon binding of RRM1,2-WT or -GS to weak Py-tract RNA (U4A8U4) are similar for residues in RRM1, which are not affected by the linker.

Figure S8

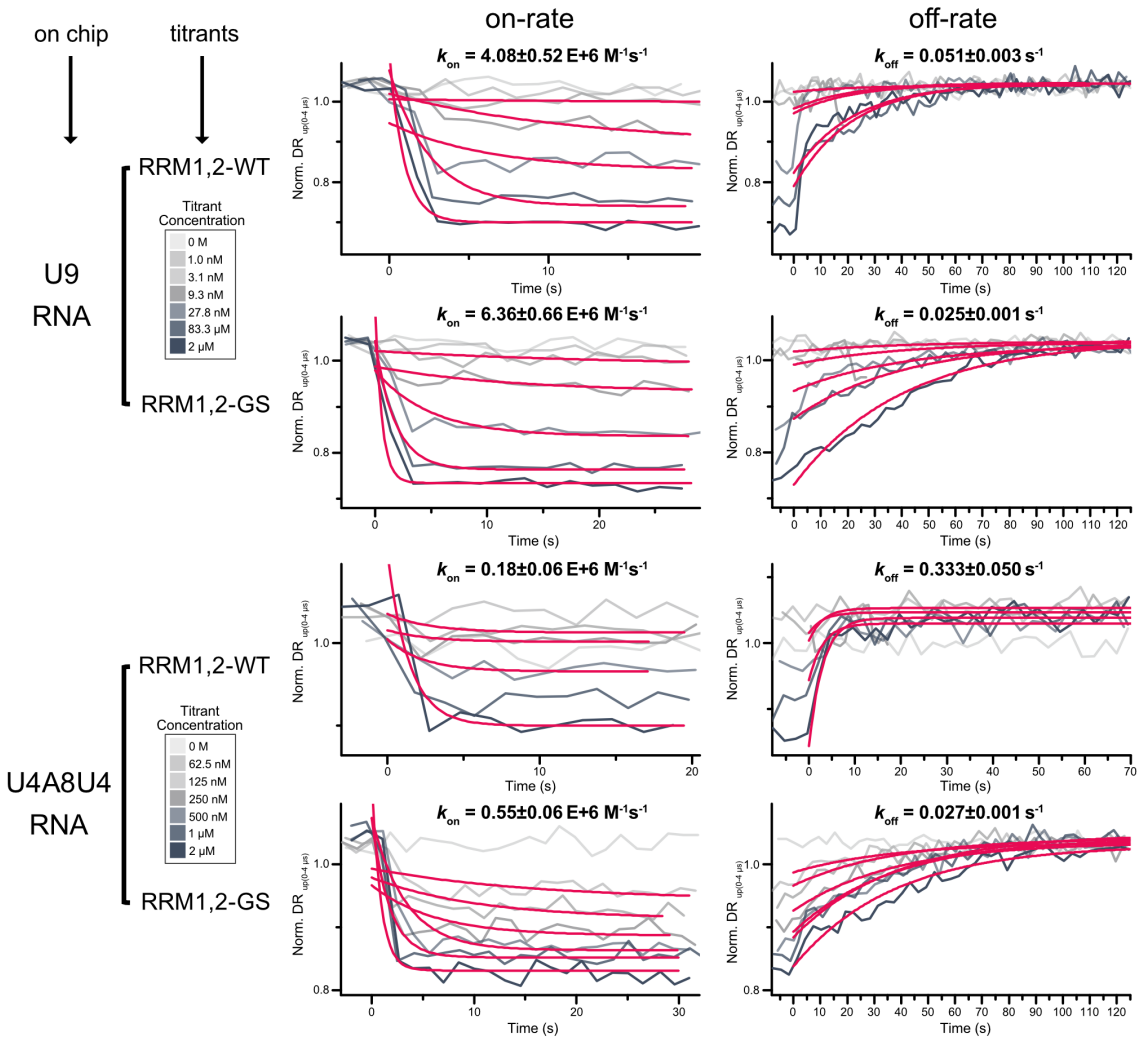
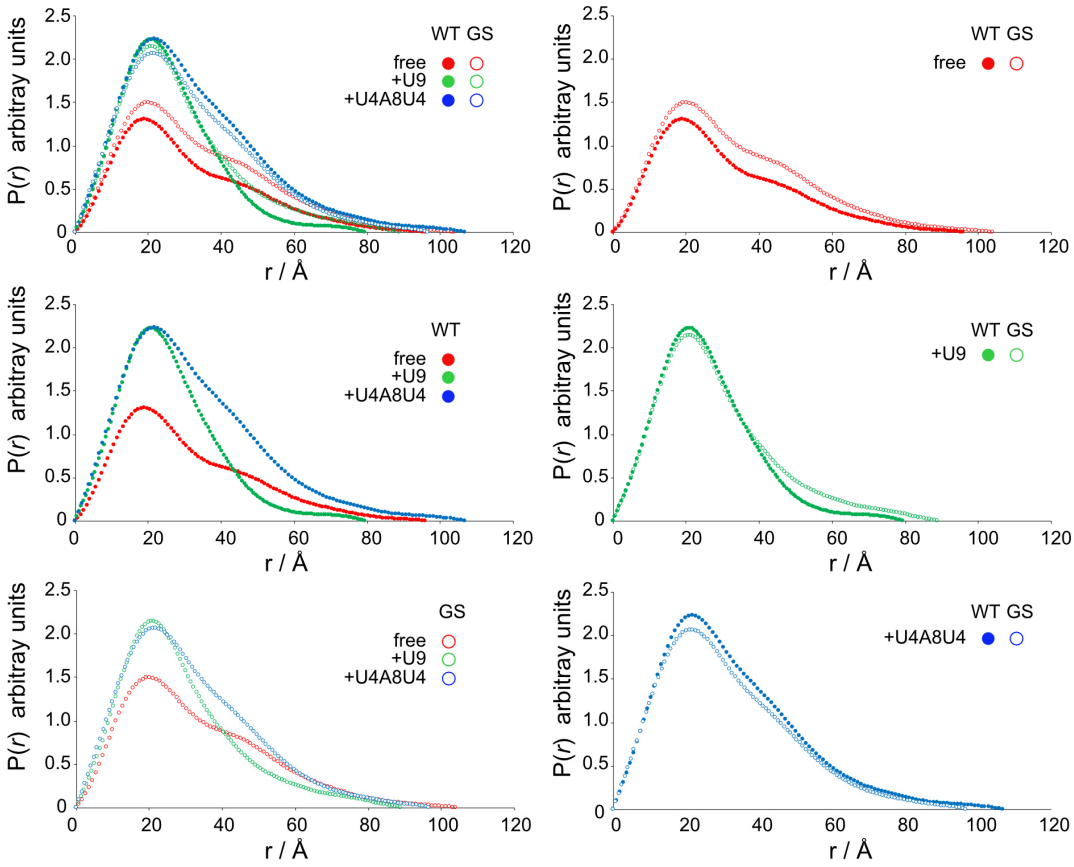


Figure S8. Experimental on- and off-rates determined by SwitchSENSE.
Association (k_{on}) and dissociation (k_{off}) rates for the binding of RRM1,2-WT and -GS to strong (U9) and weak (U4A8U4) Py-tract RNAs. RNAs are attached to the DNA nanolever on the chip through a complementary DNA sequence in the DNA/RNA hybrid. The dynamic response (DR) of the nanolever is recorded with an increasing protein concentration.

Figure S9

A



B

	unbound		+U9		+U4A8U4	
	WT	GS	WT	GS	WT	GS
D_{max} /Å	96.0	104.0	79.4	88.8	106.7	96.7
R_g /Å	24.8	26.8	20.2	22.9	26.1	25.2

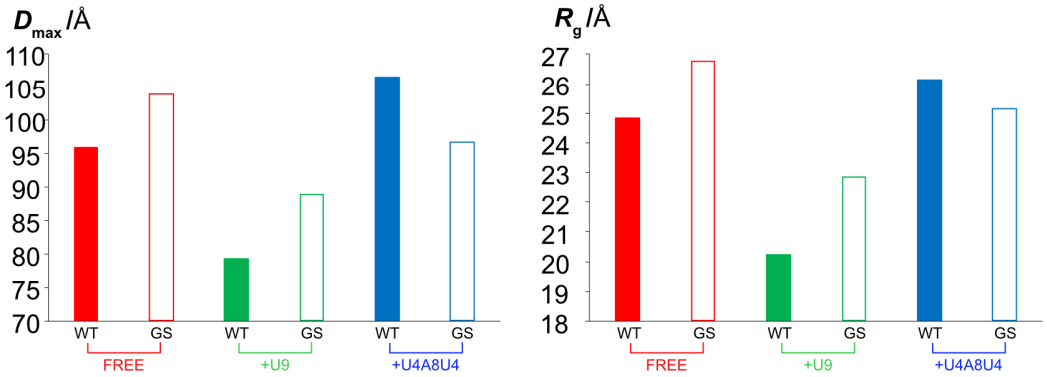


Figure S9. SAXS analysis of RRM1,2-WT and RRM1,2-GS with different RNA ligands. (A) Intensity (I_0) at the zero scattering angle, pairwise distance distribution P(r), and **(B)** maximum distance (D_{max}) and radius of gyration (R_g) for RRM1,2-WT and RRM1,2-GS free, and bound to strong (U9) and weak (U4A8U4) Py-tract RNAs.

Figure S10

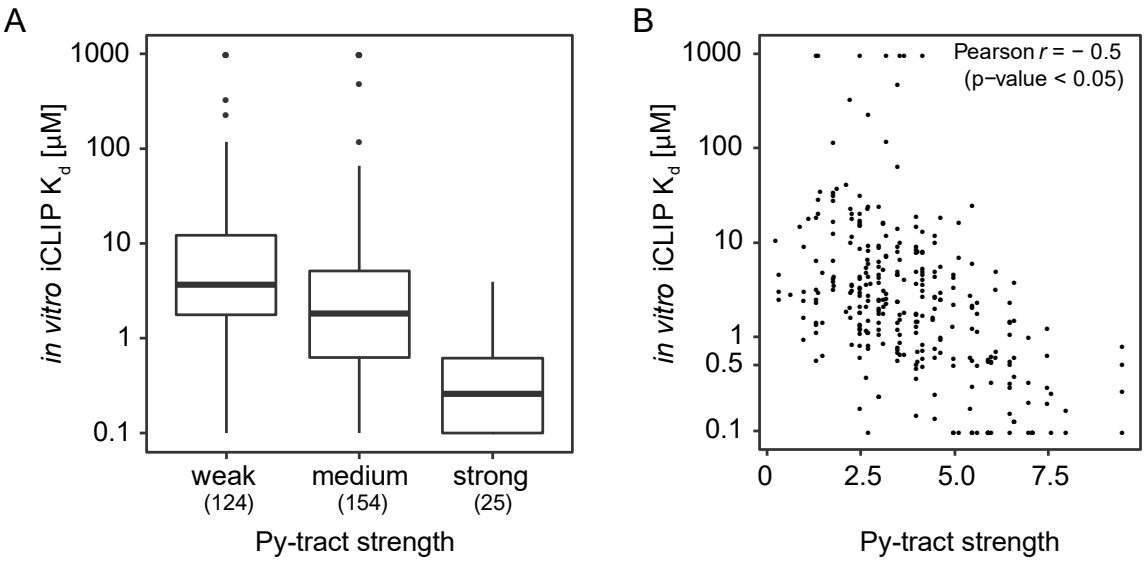


Figure S10. Py-tract strength directly translates into U2AF2 affinity. (A) Boxplot shows distribution of K_d values for U2AF2 binding sites stratified by the strength of the associated Py-tracts (weak/medium/strong). K_d values were taken from a previous *in vitro* iCLIP study (7) which covered 303 out of 424 binding sites measured in this study. Boxes represent quartiles, centre lines denote 50th percentile, and whiskers extend to most extreme values within 1.5x interquartile range. **(B)** Scatterplot compares K_d values against Py-tract strength. Pearson correlation coefficient and associated *P* value are given.

Figure S11

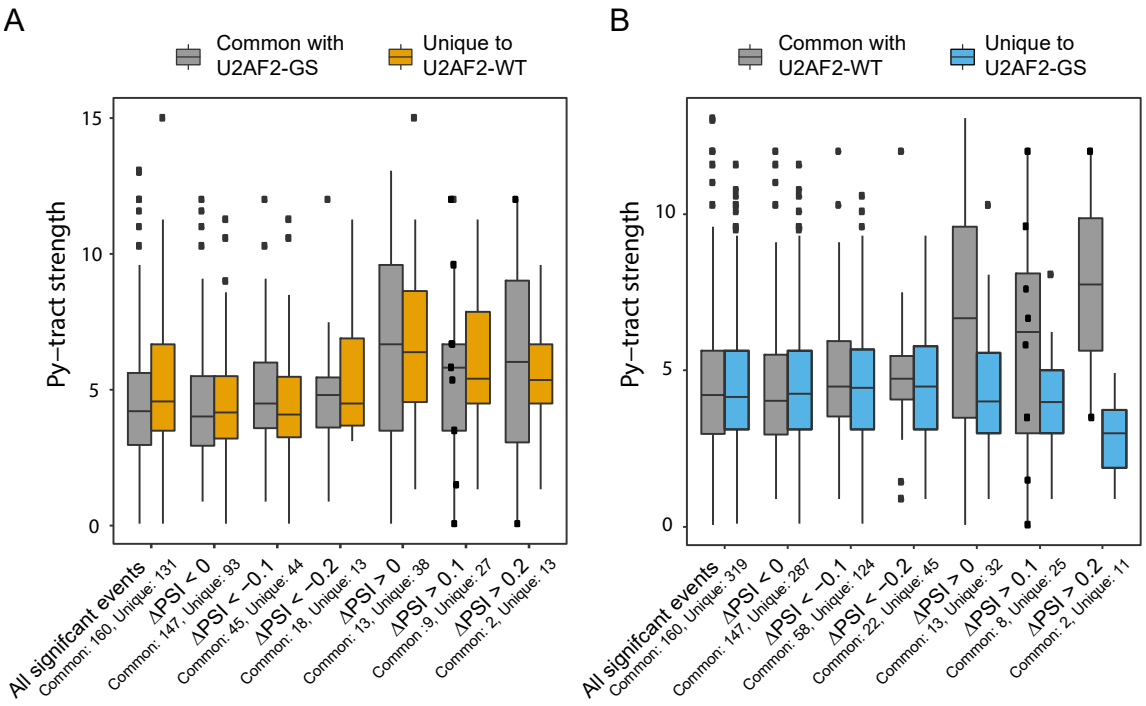


Figure S11. Exons that strongly increase only upon U2AF2-GS overexpression display weaker Py-tracts. Boxplots display the Py-tract strengths for significantly regulated exons (false discovery rate [FDR] < 0.05) that are either shared between RRM1,2-WT and RRM1,2-GS (grey, 'Common') or selectively regulated ('Unique') upon overexpression of **(A)** RRM1,2-WT (orange) or **(B)** RRM1,2-GS (blue). Exons were selected for increasing effect sizes (|ΔPSI| > 0 or 0.1 or 0.2). A subset of the categories is shown in **Figure 5D**. Boxes represent quartiles, centre lines denote 50th percentile, and whiskers extend to most extreme values within 1.5x interquartile range. The number of unique and common exons in each category is given below.

Figure S12

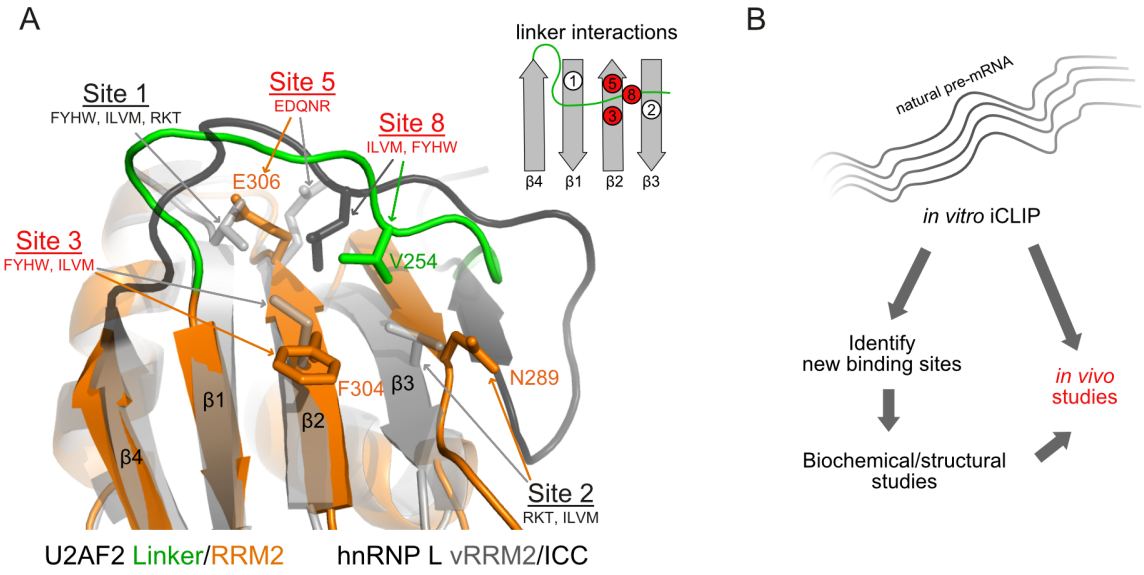


Figure S12. Comparison of RRM structures and interactions with linker regions. (A) Superposition of the U2AF2 linker-RRM2 structure (green/orange) to the structure of the vRRM2/C-term coil region (grey/black) of hnRNPL (2mqm.pdb). Key residues that interact with linker regions flanking the core RRM fold (see schematic figure at the right top) are shown in sticks for both structures and residues are annotated for U2AF2 (see discussion for details). The sites conserved in U2AF2 are highlighted in red. Residue types found at the corresponding positions in RRMs that show linker interactions are indicated. **(B)** Flowchart of our approach combining large-scale protein-RNA interaction mapping and high-resolution structural biology to study the roles of flanking regions of the core RRM as employed in the current study.

SI References

1. A. A. Agrawal *et al.*, An extended U2AF(65)-RNA-binding domain recognizes the 3' splice site signal. *Nat Commun* **7**, 10950 (2016).
2. E. A. Sickmier *et al.*, Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol Cell* **23**, 49-59 (2006).
3. C. D. Mackereth *et al.*, Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* **475**, 408-411 (2011).
4. J. R. Huang *et al.*, Transient electrostatic interactions dominate the conformational equilibrium sampled by multidomain splicing factor U2AF65: a combined NMR and SAXS study. *J Am Chem Soc* **136**, 7068-7076 (2014).
5. L. Voith von Voithenberg *et al.*, Recognition of the 3' splice site RNA by the U2AF heterodimer involves a dynamic population shift. *Proc Natl Acad Sci U S A* **113**, E7169-E7175 (2016).
6. B. Meszaros, G. Erdos, Z. Dosztanyi, IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res* **46**, W329-W337 (2018).
7. F. X. R. Sutandy *et al.*, In vitro iCLIP-based modeling uncovers how the splicing factor U2AF2 relies on regulation by cofactors. *Genome research* **28**, 699-713 (2018).