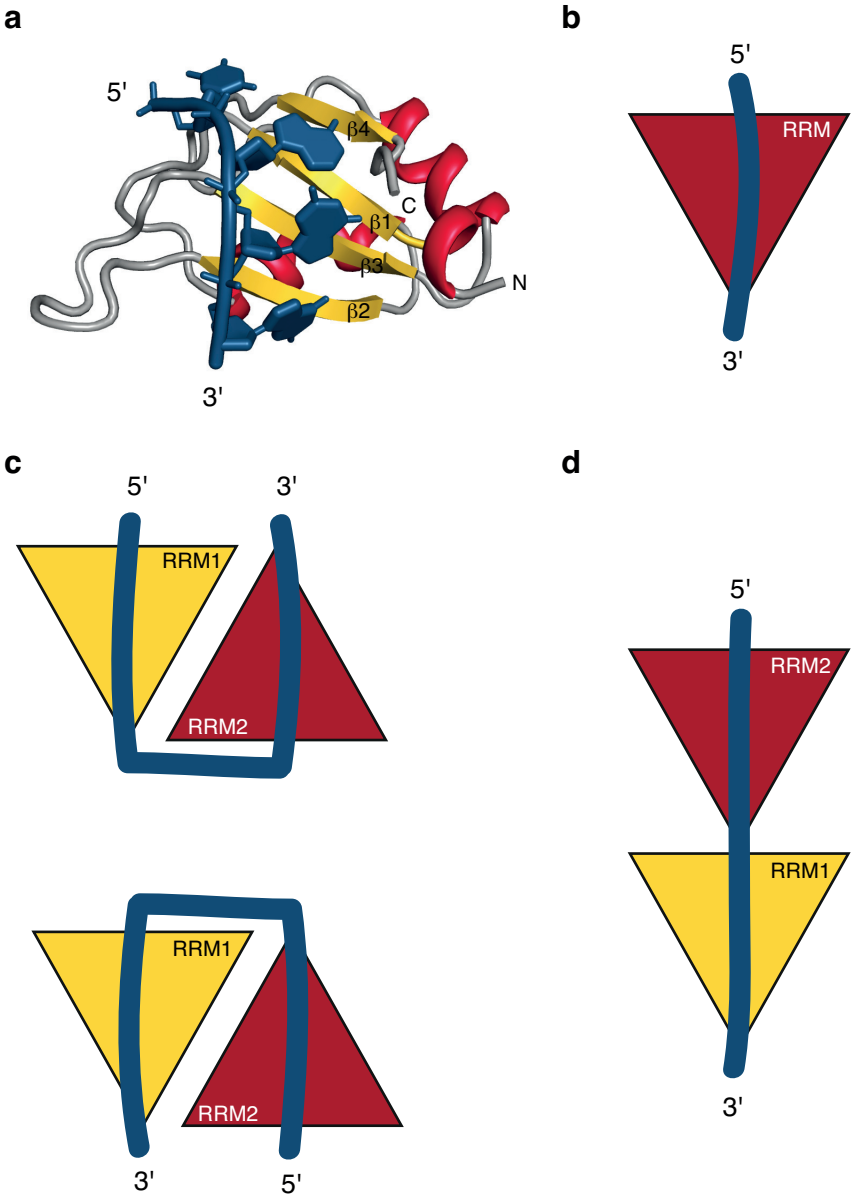
Supplementary Material for

Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRMs influences the nucleic acid binding topology

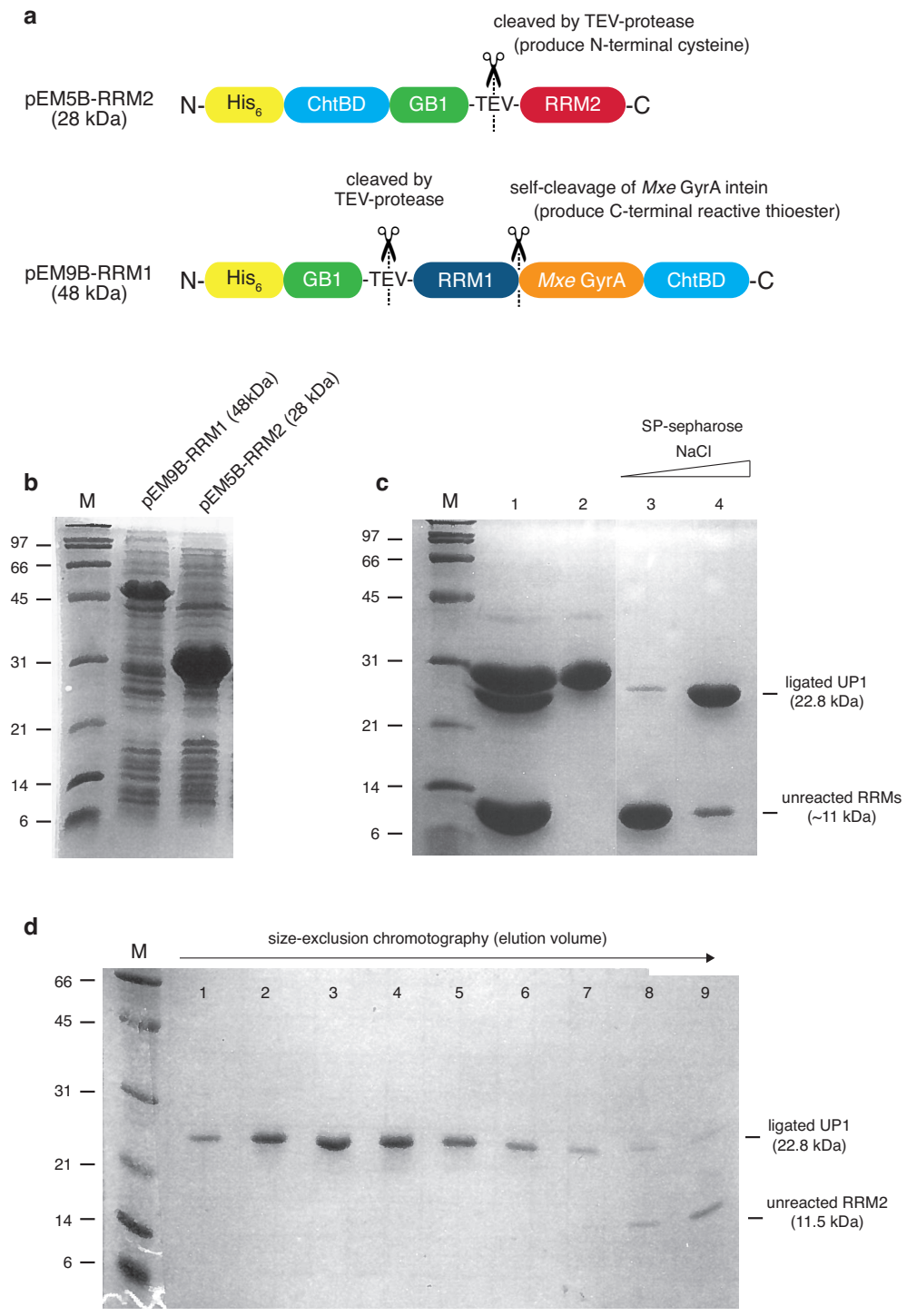
Pierre Barraud & Frédéric H.-T. Allain
Institute of Molecular Biology and Biophysics, ETH Zurich, CH-8093 Zürich, Switzerland

Supplementary Figures 1-8
Supplementary Table 1

Supplementary Figures and Tables



Supplementary Figure 1:
RRM domains are asymmetric binding platforms
(a) Structure of an archetypal RRM bound to nucleic acid (hnRNP A1 RRM1 bound to telomeric DNA, pdb code 2UP1). The β-sheet surface formed by β-strands 1-4 (*in yellow*) is an asymmetric surface that binds single-stranded nucleic acids (*in blue*) in a single defined orientation, namely the 5' extremity towards β-strand 4 and the 3' towards β-strand 2. **(b)** Schematic representation of an RRM domain bound to ssRNA or ssDNA. The asymmetric β-sheet surface is schematically represented as a triangle (*in red*), imposing a 5'-3' orientation to the nucleic acid strand (*in blue*). **(c-d)** Schematic representations illustrating how the relative orientation of different RRMs in proteins containing multiple-RRMs can influence the modes of RNA binding. RRMs forming a discontinuous and anti-parallel platform can induce a looping in the nucleic acid target (for example PTB RRM34) **(panel c)**; and RRMs interacting to form a continuous binding platform can bind to longer stretches of nucleotides (for example the polyA-binding protein) **(panel d)**.

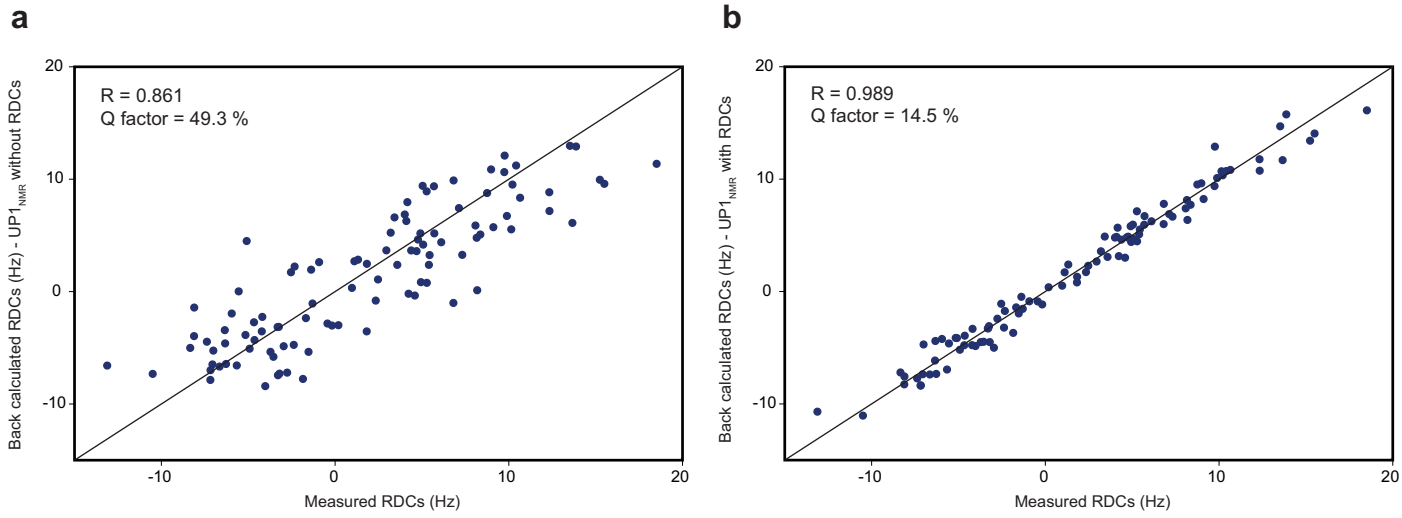


Supplementary Figure 2:
Expressed protein ligation: expression, intein cleavage, ligation and purification.
(a) Constructs used for expressed protein ligation. pEM5B-RRM2 is used for generation of the C-terminal fragment with an N-terminal cysteine; pEM9B-RRM1 is used for generation of the N-terminal fragment *via* self-cleavage of *Mxe* GyrA intein that generates a C-terminal reactive thioester. **(b)** Total *E. coli* soluble proteins after 16h induction with pEM9B-RRM1 (line #1) and pEM5B-RRM2 (line #2). **(c)** Ligation product purification: (M) protein ladder, size in kDa. (#1) Ligation mixture after removal of purification and solubilisation tags on Ni-NTA column; bands (from *top* to *bottom*) correspond to *Mxe* GyrA-ChtBD (28 kDa), ligated UP1 (22.8 kDa), and unreacted RRMs (~11 kDa). (#2) *Mxe* GyrA-ChtBD (28 kDa) retained on chitin column. (#3 and #4) separation of ligated UP1 and unreacted RRMs *via* ion exchange chromatography (SP-sepharose). **(d)** ligated UP1 is further purified from unreacted RRM2 via size-exclusion chromatography. Fractions 1-6 were pooled and concentrated up to 1.0 mM for NMR measurements.

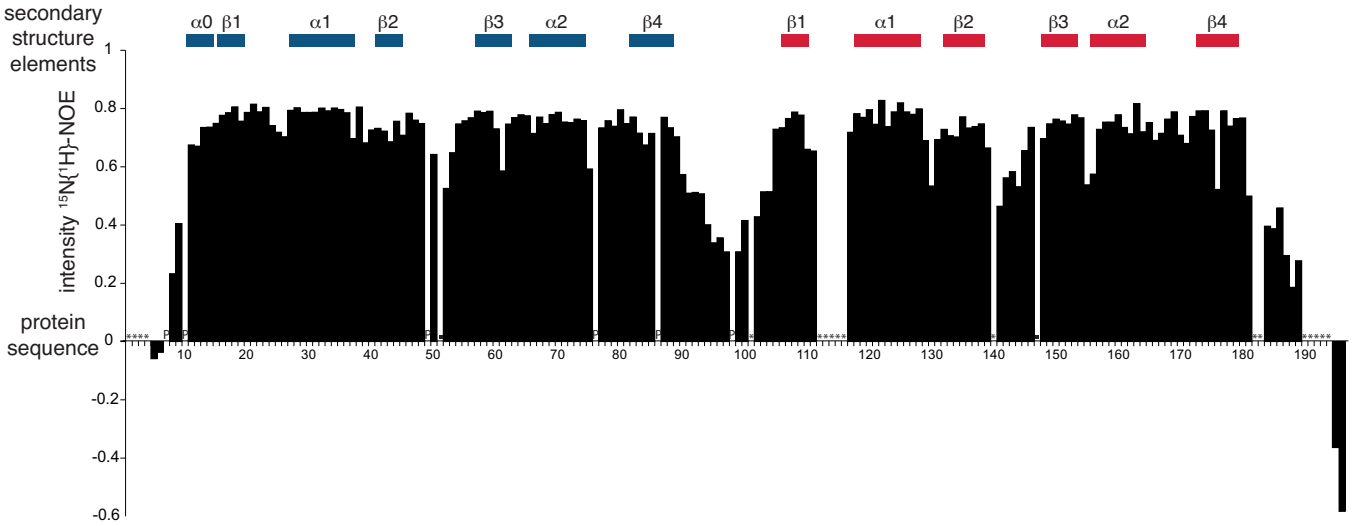
Supplementary Table 1
Agreement between experimental UP1 NH RDCs and back calculated RDCs.

Structure	Number of RDCs (*a*)	Correlation coefficient (R)	Q factor (%) (*b*)	Number of RDC violation > 5 Hz (*c*)
$UP1_{NMR}$ (*d*)	101	0.989 ± 0.001	14.5 ± 0.5	0
$UP1_{NMR}$ (*e*)	101	0.861 ± 0.014	49.3 ± 2.4	5
$UP1_{free}$	98	0.824	55.4	8
$UP1_{bound}$	98	0.874	47.4	7

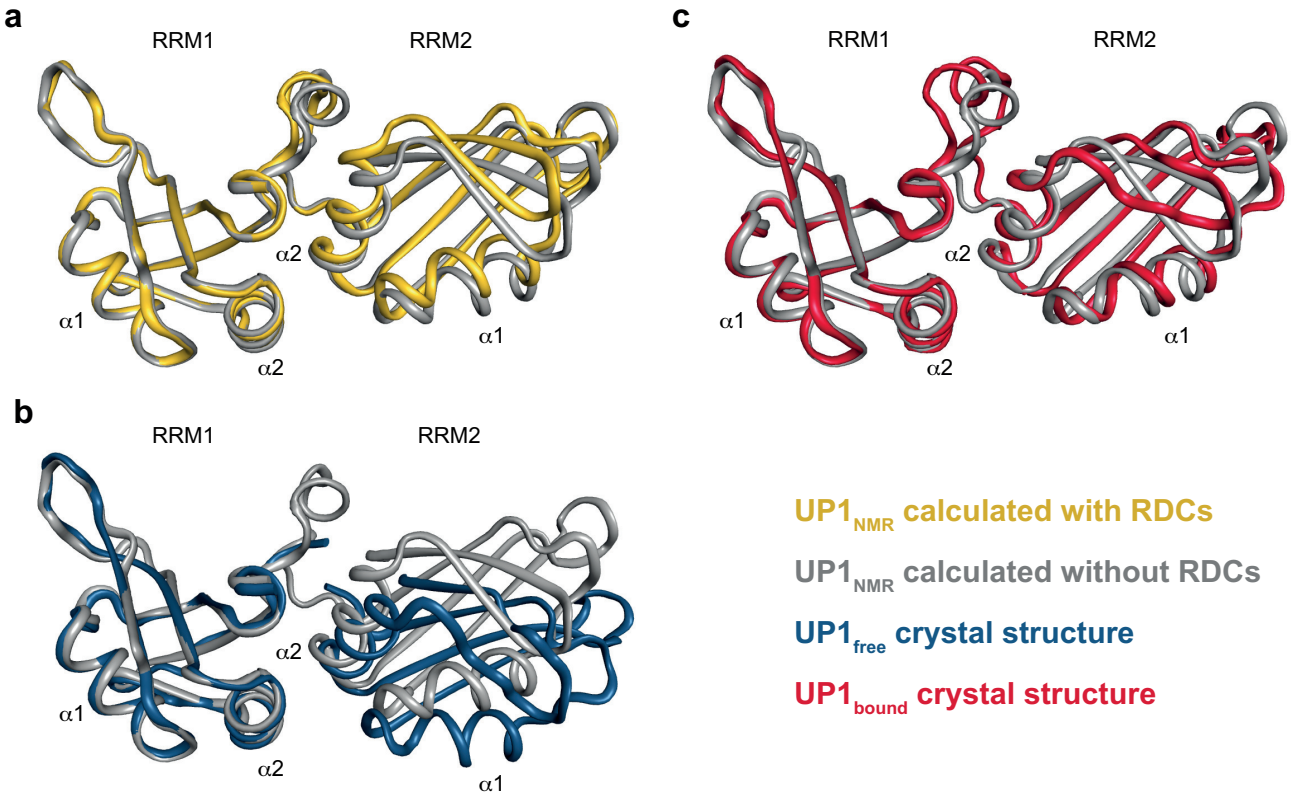
(*a*) The entire set of 101 NH RDCs is used in the case of solution NMR structures, whereas RDCs from residues Gln165, Lys166 and Tyr167 are omitted for the analysis of crystal structures (98 RDCs in total), since this region present an obvious backbone distortion leading to large RDC violations (See Supplementary Figure 8).
(*b*) Quality factor Q defined as $Q = rms(D^{obs}-D^{calc})/rms(D^{obs})$ (Bax A, Kontaxis G, Tjandra N (2001) Methods Enzymol 339:127–174).
(*c*) For NMR structures, the number of consistent RDC violations are listed, *i.e.* the number of RDC violation > 5 Hz that are found in at least 6 structures of the bundle.
(*d*) UP1 NMR structure refined using the set of measured RDCs.
(*e*) UP1 NMR structure calculated and refined without RDC restraints.



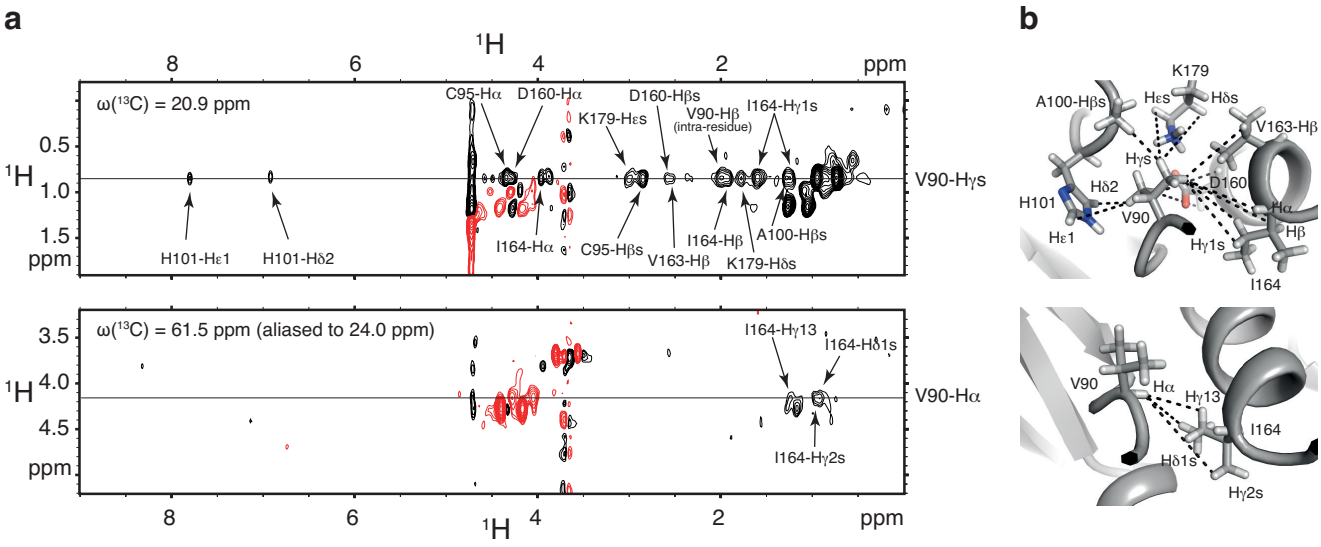
Supplementary Figure 3:
Correlation plots between experimental and back-calculated RDCs for the $UP1_{NMR}$ structures.
Plot of experimental UP1 NH RDCs versus values calculated after best-fitting the alignment tensor to (**a**) the $UP1_{NMR}$ structure calculated and refined without RDC restraints, and (**b**) the $UP1_{NMR}$ structure refined including 101 NH RDC restraints. The plots are drawn for a representative structure of the bundle chosen as the structure with the lowest backbone r.m.s.d to the other structures of the ensemble. The values for R and Q factors refer to the average value over the 20 structures of the ensemble (See also Supplementary Table 1).



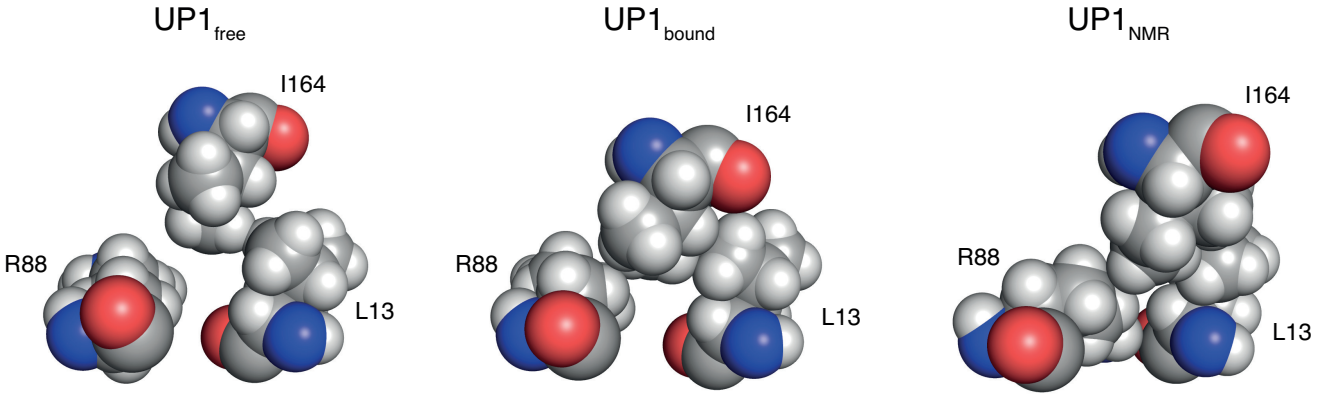
Supplementary Figure 4:
Heteronuclear $^{15}N\{^{1}H\}$-NOE values plotted against UP1 sequence
Values for residues marked with black squares could not be determined due to resonance overlap. Residues marked with a P correspond to prolines. Residues marked with an asterisk have a missing amide assignment. The secondary structure elements are represented as rectangles above the diagram with RRM1 *in blue* and RRM2 *in red*.



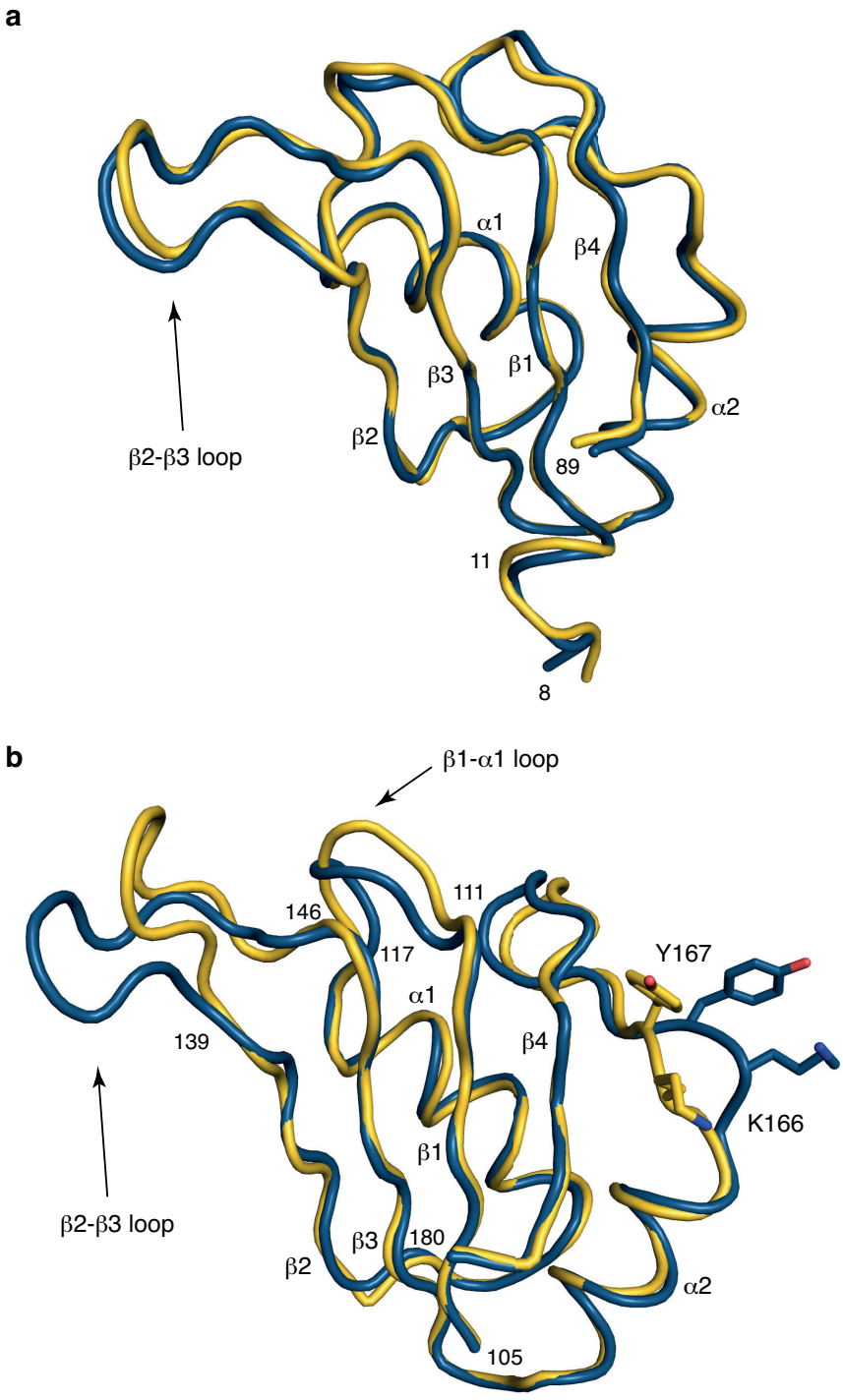
Supplementary Figure 5:
Structural comparison of UP1 NMR structure calculated without RDCs with NMR and crystal structures.
(a-c) Different overlays of $UP1_{NMR}$ calculated without RDCs *in grey* with $UP1_{NMR}$ calculated with RDCs *in yellow* (panel **a**), with $UP1_{free}$ *in blue* (panel **b**), and with $UP1_{bound}$ *in red* (panel **c**). For the NMR structures, a representative structure of the bundle was chosen as the structure with the lowest backbone r.m.s.d to the other structures of the ensemble. All structures are superimposed onto RRM1 (residues 11-89) to emphasize the differences in the relative orientation of the two RRMs in the different structures. The two NMR structures (calculated with and without RDCs) overlay better with the $UP1_{bound}$ structure, see for example the large differences in the position of helices α1 and α2 in RRM2 for the $UP1_{free}$ structure (panel **b**).



Supplementary Figure 6:
Interdomain contacts between Val90 and Ile164 as determined with segmentally labeled UP1
(**a**) Series of cross-sections from the 3D ^{13}C-edited half-filtered NOESY measured on the segmentally labeled hnRNP A1 RRM12 sample with RRM1 ^{13}C/^{15}N labeled and RRM2 unlabeled. Assignment of interdomain NOEs are reported on the spectra and illustrated as dashed lines on the solution NMR structure (see panel **b**). ^{13}C chemical shifts are indicated for each cross-section. Positive signals are *in black* and negative ones *in red*. Strips of signals at 4.72 ppm and 3.68 ppm correspond to residual water and buffer signals, respectively. (**b**) Illustration of interdomain NOEs of each cross-section from panel **a** on the solution NMR structure. NOEs are represented as dashed lines between the corresponding protons or groups of protons.



Supplementary Figure 7:
Differences in the hydrophobic cluster formed by Ile164, Arg88 and Leu13 between UP1$_{free}$, UP1$_{bound}$ and UP1$_{NMR}$ structures
Ile164 side chain is more tightly packed in between Leu13 and Arg88 in UP1$_{bound}$ and UP1$_{NMR}$ structures (*middle and right panel*) than in UP1$_{free}$ structure (*left panel*). This support the fact that UP1$_{bound}$ would be the native conformation of UP1 and UP1$_{free}$ a destabilized structure induced by crystallization.



Supplementary Figure 8:
Structural comparison of RRM1 and RRM2 between the free solution structure and the free crystal structure of UP1
(a) Superimposition of RRM1 from free UP1 crystal structure ($UP1_{free}$, *in blue*) and UP1 solution NMR structure ($UP1_{NMR}$, *in yellow*). Backbone r.m.s.d. for residues 11-89 = 0.55 Å (see Table 3).
(b) Superimposition of $UP1_{free}$ (*in blue*) and $UP1_{NMR}$ (*in yellow*) on RRM2. Backbone r.m.s.d. for residues 105-111, 117-139, 146-180 = 0.79 Å (see Table 3). The main structural differences, apart from flexible loop regions, are located after helix α2 and involve residues Lys166 and Tyr167. In $UP1_{free}$, Lys166 makes intermolecular contacts with Glu66 of a symmetry related molecule, whereas for some structures of the $UP1_{NMR}$ ensemble, it makes intramolecular contacts with Glu93 of the interdomain linker.