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

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

† Corresponding author: FH-T Allain, Institute of Molecular Biology and Biophysics, ETH Zurich, Schafmattstrasse 20, CH-8093 Zürich, Switzerland. Tel.: +41 44 633 3940, Fax: +41 44 633 1294. E-mail: allain@mol.biol.ethz.ch

ABSTRACT

Human hnRNP A1 is a multi-functional protein involved in many aspects of nucleic-acid processing such as alternative splicing, micro-RNA biogenesis, nucleo-cytoplasmic mRNA transport and telomere biogenesis and maintenance. The N-terminal region of hnRNP A1, also named unwinding protein 1 (UP1), is composed of two closely related RNA recognition motifs (RRM), and is followed by a C-terminal glycine rich region. Although crystal structures of UP1 revealed inter-domain interactions between RRM1 and RRM2 in both the free and bound form of UP1, these interactions have never been established in solution. Moreover, the relative orientation of hnRNP A1 RRMs is different in the free and bound crystal structures of UP1, raising the question of the biological significance of this domain movement. In the present study, we have used NMR spectroscopy in combination with segmental isotope labeling techniques to carefully analyze the inter-RRM contacts present in solution and subsequently determine the structure of UP1 in solution. Our data unambiguously demonstrate that hnRNP A1 RRMs interact in solution, and surprisingly, the relative orientation of the two RRMs observed in solution is different from the one found in the crystal structure of free UP1 and rather resembles the one observed in the nucleic-acid bound form of the protein. This strongly supports the idea that the two RRMs of hnRNP A1 have a single defined relative orientation which is the conformation previously observed in the bound form and now observed in solution using NMR. It is likely that the conformation in the crystal structure of the free form is a less stable form induced by crystal contacts. Importantly, the relative orientation of the RRMs in proteins containing multiple-RRMs strongly influences the RNA binding topologies that are practically accessible to these proteins. Indeed, RRM domains are asymmetric binding platforms contacting single-stranded nucleic acids in a single defined orientation. Therefore, the path of the nucleic acid molecule on the multiple RRM domains is strongly dependent on whether the RRMs are interacting with each other. The different nucleic acid recognition modes by multiple-RRM domains are briefly reviewed and analyzed on the basis of the current structural information.

Keywords

Segmental isotope labeling; expressed protein ligation; intein; structural biology; hnRNP A1; UP1; RRM; NMR; inter-domain interaction; ^{13}C-edited half-filter NOESY; domain orientation; RNA recognition modes.

Abbreviations

GB1: B1 domain of streptococcal protein G

hnRNP: heterogeneous nuclear ribonucleoprotein

HSQC: heteronuclear single quantum coherence

MESNA: sodium 2-mercaptoethanesulfonate

Mxe GyrA: gyrA gene from *Mycobacterium xenopi*

NOE: nuclear Overhauser effect

NOESY: nuclear Overhauser effect spectroscopy

PABP: polyadenylate binding protein

r.m.s.d.: root-mean-square deviation

RRM: RNA recognition motif

UP1: unwinding protein 1

Introduction

Eukaryotic mRNAs are transcribed as precursors (pre-mRNAs) containing intervening sequences (introns) that are subsequently removed such that the flanking regions (exons) are spliced together to form mature mRNAs. In addition to constitutive splicing, alternative splicing generates different mRNAs encoding distinct proteins, and hence increases protein diversity (Nilsen and Graveley 2010). For efficient splicing, most introns require a conserved 5' splice site, a branch point sequence followed by a polypyrimidine tract and a 3' splice site. In addition, other signal sequences along alternatively spliced exons, or their flanking introns are targeted by two large families of proteins that finely regulate alternative splicing: the SR-protein family (serine-arginine protein family) and the hnRNP family (heterogeneous nuclear RiboNucleoProtein family).

The hnRNP protein family consists of at least 20 proteins in humans that have been characterized as components of protein complexes bound to pre-mRNA (hnRNP complexes) (Dreyfuss et al. 1993). Most proteins of the hnRNP family contain at least one RRM domain (RNA recognition motif) (Maris et al. 2005), from one RRM in hnRNP C up to four RRMs in hnRNP I (also called Polypyrimidine Tract Binding protein, PTB). They also contain additional domains, like glycine-rich domains or aspartate-glutamate-rich domains that either contribute to RNA recognition and/or mediate protein/protein interaction. Among this family, hnRNP A1 is one of the most abundant and best-characterized components of hnRNP complexes. It is well established that the multi-functional hnRNP A1 protein plays an active role not only in alternative pre-mRNA splicing (Mayeda and Krainer 1992; Cáceres et al. 1994; Yang et al. 1994), but also in the maturation of some micro-RNA precursors (Guil and Cáceres 2007; Michlewski et al. 2008; Michlewski and Cáceres 2010), in nucleo-cytoplasmic mRNA transport (Piñol-Roma and Dreyfuss 1992), in promoting RNA strand annealing (Pontius and Berg 1990), and in telomere biogenesis and maintenance (LaBranche et al. 1998; Zhang et al. 2006; Flynn et al. 2011).

Human hnRNP A1 is a 320-amino-acid protein composed of two closely related RRM domains in its N-terminal region followed by a highly flexible glycine-rich C-terminal region (45% of glycine). The N-terminal region, which includes RRM1 and RRM2 and spans residues 1 to 196, is also known as unwinding protein 1 (UP1). Interestingly, the two RRMs of hnRNP A1 are neither redundant nor functionally equivalent, in spite of their similar sequences and overall structure. Indeed, chimeric protein construction by duplication, deletion or swap of the RRMs differently affects the hnRNP A1 alternative splicing function (Mayeda

et al. 1998). The relative position of the two RRMs is therefore crucial for the alternative splicing activity of hnRNP A1.

To date, several high-resolution crystal structures of the two tandem RRMs of hnRNP A1 have been solved both in their free form and bound to repeats of telomeric DNA fragments. Three structures of free UP1 with resolution ranging from 1.1 Å to 1.9 Å have been refined in the $P2_1$ space group from an identical monoclinic crystal form (pdb accession codes 1UP1, 1HA1 and 1L3K) (Xu et al. 1997; Shamoo et al. 1997; Vitali et al. 2002). These three structures are almost indistinguishable with an average pairwise root-mean-square deviation (r.m.s.d.) for backbone atoms of 0.21 ± 0.04 Å. In these structures, the two RRMs are interacting with one another, mainly *via* two Arg-Asp salt bridges, and hence adopting a single relative orientation. The two RRMs are oriented in an anti-parallel manner, meaning that the two RNA binding surfaces are discontinuous and could bind to RNA strands having opposite 5'-3' polarity. It is important to notice that this fixed relative orientation of the RRMs strongly influences the repertoire of RNA binding topologies that can be formed with hnRNP A1. These different accessible topologies have been discussed previously (Xu et al. 1997; Shamoo et al. 1997). However, since the interdomain interaction surface is relatively modest (~630 $Å^2$) and the residues involved in the Arg-Asp salt bridges are not absolutely conserved in the hnRNP A1-like proteins, the possibility that the association between RRM1 and RRM2 might be the result of crystal packing forces has been pointed out (Shamoo et al. 1997).

In addition to these free structures, eleven structures of UP1 bound to wild-type or diverse mutated repeats of telomeric DNA fragments have been solved with resolution ranging from 1.8 Å to 2.6 Å. All these structures have been refined in the $P4_32_12$ space group from an identical tetragonal crystal form (pdb accession codes 2UP1, 1PGZ, 1PO6 and 1U1K to 1U1R) (Ding et al. 1999; Myers et al. 2003; Myers and Shamoo 2004). These bound structures are almost indistinguishable with a calculated average pairwise r.m.s.d. for protein backbone atoms of 0.19 ± 0.04 Å. Interestingly, the overall interface between RRM1 and RRM2 as observed in this different crystal form is globally conserved as compared with the interface in the free form of the protein. For instance, the two Arg-Asp salt-bridges and other important contacts are similarly present in the structure of the telomeric DNA bound form. However, the relative orientation of the two RRMs is significantly altered and the free and bound structures do not perfectly overlay. Indeed, overall backbone r.m.s.d. between the free and bound UP1 structures is as high as 1.70 Å whereas individual RRMs display a much better agreement (0.32 Å and 0.45 Å for RRM1 and RRM2, respectively). These significant con-

formational changes correspond to a rotation of ~15° of one RRM compare to the other (Ding et al. 1999). This reorientation of the RRMs has been attributed to the binding to the DNA substrate (Ding et al. 1999). However, one could also imagine that this relative movement of the RRMs is due to differences in the protein-protein contacts with neighbouring proteins in the two different crystal lattices (*i.e.* monoclinic and tetragonal crystals). In any case, the interactions at the inter-RRM interface in UP1 are probably quite weak, since the orientation of the two RRMs can be influenced by nucleic acid binding or by contacts with neighbouring molecules in the crystal lattice. Together with the aforementioned doubt brought up with the analysis of the free crystal structure (Shamoo et al. 1997), it directly raises the question whether the contacts between the two RRMs are present at all in solution. In addition, several biochemical studies performed in solution have been favouring a model of two independent RRMs connected by a flexible inter-RRM linker (Casas-Finet et al. 1991; Shamoo et al. 1994; Shamoo et al. 1995). In conclusion, there is to date no direct and unequivocal evidence of the existence, in solution, of an inter-RRM interaction in hnRNP A1.

Importantly, the relative orientation of different RRMs in proteins containing multiple-RRMs strongly influences the modes of RNA binding that are practically accessible to these particular proteins. Indeed, RRMs are asymmetric binding platforms contacting single-stranded nucleic acids in a single defined orientation, namely the 5' extremity towards β-strand 4 and the 3' towards β-strand 2 (Maris et al. 2005) (Supplementary Figure 1). Therefore, RRMs forming a discontinuous and anti-parallel platform may induce a looping in the nucleic acid target, as observed in the case of PTB RRM34 (Oberstrass et al. 2005; Lamichhane et al. 2010); and RRMs interacting to form a continuous binding platform can bind to longer stretches of nucleotides as seen in the structure of the polyA-binding protein (Deo et al. 1999) (See Supplementary Figure 1 for schematic illustrations of these cases). To date, structural information on the spatial organization of multiple RRMs in proteins containing at least two RRMs is still quite limited, since the structures of only a dozen of proteins with multiple RRMs have been solved in their free and/or nucleic acid bound form. These structures revealed that a limited number of distinct situations are actually occurring and exploited to achieve distinct biological functions. These features will be presented and discussed in a latter paragraph of this article, on the basis of the available structures.

In the present study, we have used segmental isotope labeling to determine whether the two RRMs of hnRNP A1 are interacting in solution. Segmental isotope labeling is a very attractive technique to reduce the complexity of NMR spectra with a large number of poten-

tial applications for the study of large and/or multi-domain proteins (Skrisovska et al. 2010). Different methods, known as native chemical ligation (NCL), expressed protein ligation (EPL), and protein trans-splicing (PTS), are available for segmental isotope labeling of proteins (David et al. 2004; Muralidharan and Muir 2006). To date, isotope segmental labeling has not been extensively applied in NMR, although different studies have already demonstrated that segmental labeling is a very elegant and relevant approach to investigate large proteins (Yagi et al. 2004; Minato et al. 2012), to study conformational changes and ligand binding (Anderson et al. 2005), to investigate interdomain interactions within multi-domain proteins (Camarero et al. 2002; Zhang et al. 2007), and also to enable precise protein structure determination of multi-domain proteins (Vitali et al. 2006; Chen and Wang 2011; Chen et al. 2011).

In this work, we have determined the solution structure of the two RRMs of hnRNP A1 using a segmental labeling strategy in order to clearly determine whether these two RRMs are truly interacting in solution. This approach enabled us to unambiguously identify interdomain NOEs between RRM1 and RRM2 of hnRNP A1 and to calculate a precise overall structure. In our solution structure of free UP1, the two Arg-Asp salt bridges are conserved at the interface between RRM1 and RRM2, but surprisingly, the relative orientation of the two RRMs is quite different from the one found in the crystal structure of free UP1 but resembles the one observed in the nucleic-acid bound form of the protein.

Material and methods

Cloning, expression and purification of hnRNP A1 RRM12 (UP1)

The DNA sequence encoding the two RRMs of hnRNP A1 (residues 2-196) (Uniprot entry P09651), were sub-cloned by PCR amplification from pET9d-hnRNPA1 (Mayeda and Krainer 1992) between *BamHI* and *XhoI* cloning sites in *E. coli* expression vector pET28a. The construct contains a N-terminal tag whose sequence MGSSHHHHHHSSGLVPRGSHMENLYFQGG includes a 6 histidine stretch used for protein purification and a TEV-protease cleavage site used for subsequent removal of the purification tag. Proteins were overexpressed in BL21(DE3) codon-plus (RIL) cells in either LB media or M9 minimal media supplemented with $^{15}NH_4Cl$ and ^{13}C-labeled glucose. The cells were grown at 37°C to OD600 ~0.4, cooled down at 30°C and induced at OD600 ~0.6 by adding isopropyl-β-D-thiogalactopyranoside to a final concentration of 0.5 mM. Cells were harvested 15 h after induction by centrifugation. Cell pellets were resuspended in lysis buffer (Tris-HCl pH 8.0 50 mM, NaCl 1 M, EDTA 1mM, DTT 1 mM) and lysed by sonication. Cell lysates were centrifuged 40 min at 45,000 g. Supernatant was loaded on a Ni-NTA column on a ÄKTA Prime purification system (Amersham Biosciences), and the protein of interest was eluted with an imidazole gradient. The fractions containing the protein were pooled, TEV protease was added at a TEV/UP1 ratio of 1/200 (w/w), and specific cleavage of the purification tag was performed at room temperature for 16 h. His-tagged TEV protease and purification tag were further separated from the protein of interest with a Ni-NTA column. The fractions containing the protein were pooled, dialyzed against the NMR Buffer (NaPi pH 6.5 10 mM, DTT 1 mM), and concentrated to ~1.2 mM with a Vivaspin 10,000 MWCO (Sartorius Stedim Biotech).

Cloning, expression and purification of isolated RRM1 and RRM2 of hnRNP A1

The DNA sequence encoding the two individual RRMs of hnRNP A1 (*i.e.* RRM1 from residues 2-97 and RRM2 from residues 95-196), were sub-cloned by PCR amplification from pET9d-hnRNPA1 between *NheI* and *BamHI* cloning sites in *E. coli* expression vector pET28a. The constructs contain a N-terminal tag whose sequence MGSSHHHHHHSSGLVPRGSHMA includes a 6 histidine stretch used for protein purification. Proteins were overexpressed and purified as for the UP1 construct, except that the protein purification tag was not cleaved off by TEV-protease.

Cloning, expression and purification of $^{13}C/^{15}N$-labeled RRM1 construct for protein ligation

The general approach to produce the segmentally labeled RRM12 sample is based on previously published protocols in use in our group (Vitali et al. 2006; Skrisovska and Allain 2008; Michel et al. 2013). The DNA sequence encoding RRM1 (residues 2-94) was sub-cloned from pET9d-hnRNPA1 between *NcoI* and *SapI* cloning sites of *E. coli* expression vector pEM9B (Michel et al. 2013) encoding a C-terminal fusion of the *Mxe* GyrA intein. The *SapI* restriction site naturally present in the pET9d-hnRNPA1 sequence had to be removed with a silent mutation K16K (AAG to AAA) prior to sub-cloning into pEM9B. A minimal sequence modification D94M was introduced into the inter-RRM linker to allow for efficient self-cleavage of the *Mxe* GyrA intein (Southworth et al. 1999). The protein construct was overexpressed with IPTG induction in BL21(DE3) codon-plus (RIL) cells in M9 minimal media supplemented with $^{15}NH_4Cl$ and ^{13}C-labeled glucose at 30 °C for 16 h. The intein fusion construct was purified with a Ni-NTA column on a ÄKTA Prime purification system, dialyzed against inactive reaction buffer (Tris-HCl pH 8.0 50 mM, NaCl 200 mM, EDTA 1 mM, 2-mercaptoethanol 1 mM) and concentrated to 0.25 mM with a Vivaspin 10,000 MWCO.

Cloning, expression and purification of non-labeled RRM2 construct for protein ligation

The DNA sequence encoding RRM2 (residues 95-196) was sub-cloned from pET9d-hnRNPA1 between *NcoI* and *BamHI* cloning sites of *E. coli* expression vector pEM5B (Michel et al. 2013). The required S95C mutation and the preceding TEV-protease cleavage site were introduced with the primers during PCR amplification. The protein construct was overexpressed in LB media, purified using the same procedure as for the RRM1 intein-fusion construct, and concentrated to 0.5 mM in the inactive reaction buffer.

Intein cleavage, protein ligation and ligation product purification

Purified protein samples bearing $^{13}C/^{15}N$-labeled RRM1 and unlabeled RRM2 were mixed with a two times excess of the unlabeled construct, and the reaction was activated by adding 100 mM sodium 2-mercaptoethanesulfonate (MESNA) and TEV-protease at a TEV/protein ratio of 1/200 (w/w). The reaction mixture was incubated at 35 °C for 48 h. The efficiency of the ligation reaction was analysed on SDS-PAGE. Purification and solubilisation tags released by TEV-cleavage and the intein protein released by self-cleavage during the ligation reaction are either retained on a Ni-NTA column or on a chitin column (see (Michel et

al. 2013) for details). The reaction mixture was therefore applied on a Ni-NTA column and thereafter on a chitin column. The flow-through was then loaded on a SP-sepharose column equilibrated with Tris-HCl pH 7.0 25 mM to separate the desired ligated product (RRM12, pI = 7.9) from the unreacted RRMs (*i.e.* RRM1 and RRM2, pI = 7.0 and 8.0, respectively). Bound proteins were eluted with a 0-400 mM NaCl gradient, fractions containing the desired ligated RRM12 were pooled and purity was evaluated by SDS-PAGE. At this stage the sample still contain about 10 % of unligated RRM2 that was further removed by applying the sample at 1 mL/min on a Superdex 26/60 HiLoad Prep Grade column (GE Healthcare) equilibrated with NMR buffer. Purity of the final sample was evaluated by SDS-PAGE to be > 95 %. The sample was concentrated to 1.0 mM in 250 μL with a Vivaspin 10,000 MWCO.

NMR spectroscopy

All NMR spectra were recorded at 303 K on Bruker AVIII-500 MHz, AVIII-600 MHz, AVIII-700 MHz, AVIII-750 MHz and Avance-900 MHz spectrometers (all equipped with a cryoprobe except for AVIII-750). The data were processed using TOPSPIN 2.1 (Bruker) and analyzed with Sparky (Goddard and Kneller 2006). Protein resonances were assigned with 2D (^{1}H,^{15}N)-HSQC, 2D (^{1}H,^{13}C)-HSQC, 3D HNCA, 3D HNCACB, 3D CBCA(CO)NH, 3D HNCO, 3D HN(CA)CO, 3D [^{13}C; ^{15}N; ^{1}H] HCC(CO)NH-TOCSY, 3D [^{1}H; ^{15}N; ^{1}H] HCC(CO)NH-TOCSY, 3D NOESY-(^{1}H,^{15}N)-HSQC and two 3D NOESY-(^{1}H,^{13}C)-HSQC optimized for the observation of protons attached to aliphatic carbons and to aromatic carbons, respectively. In addition, the assignment of aromatic protons was conducted using 2D (^{1}H,^{1}H)-TOCSY and 2D (^{1}H,^{1}H)-NOESY measured in D_2O; histidine protonation and tautomeric form were determined from a long-range (^{1}H,^{15}N)-HSQC spectrum (Pelton et al. 1993). Our assignment of RRM1 agrees for most resonances with previously published assignment (Garrett et al. 1994), yet ours is more complete. We recorded all 3D NOESY spectra with a mixing time of 100 ms and the 2D NOESY spectra with a mixing time of 60 ms. The assignment of interdomain NOEs was based on a 3D ^{13}C F1-edited, F3-filtered NOESY-HSQC spectrum (τ_m = 150 ms) (Zwahlen et al. 1997) measured in D_2O on the RRM12 segmentally labeled sample with only RRM1 ^{13}C-labeled.

We measured NH RDCs from in-phase/anti-phase (^{1}H,^{15}N)-HSQC experiments (Cordier et al. 1999), by comparing the peak positions of the up-field and down-field components measured in isotropic solution and in a dilute liquid crystalline phase. The alignment medium used for RDC measurements contained a mixture of n-dodecyl-penta(ethylene gly-

col) and n-hexanol (3% C12E5/hexanol, r = 0.96) dissolved in 90% H_2O/10% D_2O (Rückert and Otting 2000). Under these conditions, a quadrupolar splitting of 25.9 Hz was observed in the ^2H spectrum.

Protein structure calculation

Automated NOE cross-peak assignments (Herrmann et al. 2002a) and structure calculations with torsion-angle dynamics (Guntert et al. 1997) were performed with the macro noeassign of the software package CYANA 3.0 (Guntert 2004). Unassigned peak lists of the four NOESY spectra were generated as input with the program ATNOS (Herrmann et al. 2002b) and manually cleaned to remove artefact peaks. In addition, a manually assigned peak list corresponding to the 3D ^{13}C-edited half-filter NOESY was also added to account for interdomain NOE measured with the segmentally labeled sample. The interdomain NOE were not manually converted into distance constraints. Instead, peak intensities were automatically calibrated and converted to distance constraints by CYANA with an optimized average-distance-parameter accounting for the presence of only long-range NOE in this particular peak-list. The input also contained 64 hydrogen-bond restraints and 336 backbone dihedral restraints based on the chemical shift information from the program TALOS+ (Shen et al. 2009). Hydrogen bonded amides were identified as slowly exchanging protons in presence of D_2O. Their bonding partner was identified from preliminary structure calculations performed with only NOESY spectra as input. We calculated 100 independent structures that we refined in a water shell with the program CNS 1.3 (Brunger et al. 1998; Brunger 2007) including distance restraints from NOE data, hydrogen-bonds restraints, backbone dihedral restraints from TALOS+ and ^{15}N-^1H RDC restraints as previously described (Barraud et al. 2011). The 20 best energy structures were analyzed with PROCHECK-NMR (Laskowski et al. 1996) and the iCING web server (Doreleijers et al. 2012) (http://nmr.cmbi.ru.nl/icing/). Overall structural statistics of the final water-refined structure are shown in Table 1. Structures were visualized and figures were prepared with program PYMOL (http://www.pymol.org).

Analysis of the structures

Backbone r.m.s.d. between the different UP1 structures and solvent accessible surface areas for the evaluation of the interaction surfaces between interacting RRMs, were calculated with program superpose and areaimol of the CCP4 program suite (Winn et al. 2011). For the analysis of the ^{15}N-^1H RDCs, proton atoms were added to the crystal structures with the program

CNS (Brunger et al. 1998; Brunger 2007). RDC restraints were analysed with the program MODULE (Dosset et al. 2001), and back-calculated after best-fitting the alignment tensor to the different NMR and crystal structures with CYANA 3.0. Quality factors (Q) (Bax et al. 2001) and correlation coefficients (R) were also evaluated with CYANA 3.0 (Guntert 2004).

NMR dynamics

For the NMR dynamics study, ^{15}N T1 and T2 measurements were recorded at 303 K at a ^{1}H frequency of 500 MHz with established methods (Kay et al. 1989; Skelton et al. 1993). ^{15}N T1 values were derived from six (^{1}H,^{15}N)-spectra with different delays: 100, 250, 500, 750, 1500 and 2000 ms. Similarly, ^{15}N T2 values were derived from (^{1}H,^{15}N)-spectra with six different delays: 12.5, 25, 50, 75, 100 and 125 ms. T1 and T2 values were extracted by a curve-fitting subroutine included in the program Sparky (Goddard and Kneller 2006). Overall correlation times (τ_c) were derived from T1/T2 ratio of dispersed and rigid amide resonances, assuming isotropic motion (Fushman et al. 1994).

Results

Initial chemical shift mapping between the individual domains and RRM12

In the context of investigating RNA binding specificity of each individual RRM of hnRNP A1 (manuscript in preparation), we produced ^{15}N-labeled NMR samples of isolated RRM1 and RRM2, as well as RRM12 (UP1). Surprisingly, ^{1}H,^{15}N-HSQC spectrum of UP1 was virtually indistinguishable from the superposition of the two HSQC spectra coming from the isolated RRMs (Figure 1a-d). Indeed, in the case of interacting RRMs within PTB (RRM3 and RRM4), marked differences between the ^{1}H,^{15}N-HSQC footprint of the single RRMs and the one of the double-domain construct RRM34 were observed (Vitali et al. 2006). On the contrary, nearly identical ^{1}H,^{15}N-HSQC spectra have been observed between isolated RRMs and their combined double-domain construct in the context of non-interacting RRMs, for example in the case of Npl3p RRM1 and RRM2 (Skrisovska and Allain 2008) and hnRNP F qRRM1 and qRRM2 (Dominguez and Allain 2006). Based solely on such considerations one might suggest that RRM1 and RRM2 of hnRNP A1 do not interact in solution contrary to what was observed in the crystal structures (Xu et al. 1997; Shamoo et al. 1997). However, after careful inspection of the overlays (Figure 1c-d) we could identify a small number of peaks (*i.e.* one in RRM1 and three in RRM2) with small chemical shift perturbations between 0.1 and 0.15 ppm (Figure 1e-f). Anyhow, puzzled by the very small extent of these chemical shift variations, we decided to determine the structure of UP1 in solution with NMR spectroscopy, and in order to unambiguously assess the existence of interdomain contacts, if any, we prepared a segmentally labeled RRM12 sample with an expressed protein ligation approach.

Segmental isotope labeling of hnRNP A1 RRM12 by expressed protein ligation

Implementation of expressed protein ligation requires a reactive thioester at the C-terminus of RRM1 and a cysteine at the N-terminus of RRM2. Since the interdomain linker contains no natural cysteine, a cysteine was introduced by substituting serine 95 (S95C), and was thus taken as the N-terminus of RRM2. In addition, the preceding aspartate residue (D94) was mutated to methionine (D94M) to allow for efficient self-cleavage of the *Mxe* GyrA intein (Southworth et al. 1999). The resulting amino-acid sequence at the desired ligation site was then $_{91}$SREMCQRP$_{98}$. The reactive C-terminal thioester was obtained with RRM1 fused N-terminally to the *Mxe* GyrA intein and the N-terminal cysteine with a properly engineered TEV-protease cleavage site in front of RRM2 (Figure 2a and material and methods for details). Each construct was expressed separately in *E. coli* allowing for different labeling

scheme for each domain (namely $^{13}C/^{15}N$-labeled for RRM1 and unlabeled for RRM2). Ligation was conducted at 35°C for 48 h with a 2-fold excess of the unlabeled RRM2 in order to increase the ligation efficiency with respect to the $^{13}C/^{15}N$-labeled RRM1 domain. The RRM12 ligated construct was then purified from remaining unligated domains with ion exchange and size-exclusion chromatography (see material and methods and Supplementary Figure 2). We finally obtained ~6 mg of ligated RRM12 (from 2L of culture in M9 medium) and could prepare a concentrated NMR sample (~1.0 mM in 250 µL). This sample was well-folded and we could confirm with 1D ^{1}H NMR that the ligated UP1 adopts the same overall structure as the conventional recombinant UP1 protein (data not shown). This is further supported by the comparison of $^{1}H,^{15}N$-HSQC spectra of ligated UP1 and conventional recombinant UP1, where signals of RRM1 perfectly overlay in each spectra (compare Figure 2b and Figure 1c). This shows that mutations introduced for the ligation reaction in the linker region ($_{94}DS_{95}$ to $_{94}MC_{95}$) do not affect the structure of UP1.

We could therefore measure 3D ^{13}C F1-edited, F3-filtered NOESY-HSQC in order to unambiguously detect interdomain NOE crosspeaks, if true contacts exist between the two RRMs in solution. In this 3D NOESY spectrum, one could clearly see several interdomain NOEs in several cross-sections (Figure 3). For example, clear contacts are seen between the side-chains of Leu13 in RRM1 and Ile164 in RRM2 as well as between those of Met72 in RRM1 and His156 and Asp157 in RRM2. Also, multiple contacts are seen between Lys87 in RRM1 and His156 in RRM2 as well as multiple residues of the interdomain linker (*i.e.* His101, Leu102 and Thr103). In addition, multiple NOEs are seen between the side-chains of Val90 in RRM1 and of several residues in RRM2 (namely, Lys179, Asp160, Val163 and Ile164). We want here to briefly mention that some methyl-methyl contacts would be expected between aliphatic residues for which contacts are clearly seen between a methyl group on one hand and other types of aliphatic protons on the other hand (see for instance Leu13-Hδ1s NOEs to Ile164 Hα, Hβ and Hγ1s on Figure 3a). However, it is difficult to unambiguously observe and assign these methyl-methyl contacts since they are often overlapping with the strong doublets of the diagonal peaks. For this reason, only few methyl-methyl inter domain NOE were confidently identified in the 3D ^{13}C-edited half-filter NOESY. Overall, we could unambiguously assign 64 interdomain NOEs that were thereafter converted into long-range inter-proton distances and used to precisely determine the structure of UP1 in solution. Importantly, this clearly demonstrates that RRM1 and RRM2 are truly interacting in solution.

Structure determination of hnRNP A1 RRM12

In order to precisely analyse the atomic details of the interdomain interface present in solution, and especially to compare this interface with the previously determined crystal structures, we solved the solution structure of hnRNP A1 RRM12 using NMR. A total of 5354 distance constraints were derived from NOESY spectra. This includes 64 interdomain NOE that have been unambiguously assigned with the use of the segmentally labeled UP1 sample (Figure 3). This large number of constraints allowed us to obtain a precise structure with a backbone r.m.s.d. over the entire domain of 0.71 ± 0.16 Å for the ensemble of 20 conformers (Figure 4a and Table 1). Constraints also include 64 hydrogen-bond restraints, backbone dihedral restraints derived from TALOS+ predictions for 168 residues and 101 ^{15}N-^{1}H amide RDC restraints obtained from measurements in a partially oriented sample. Hydrogen bonded amides were identified as slowly exchanging protons in presence of D_2O. Their hydrogen-bond acceptors were identified from preliminary structures. Further details on the assignment and structure calculation procedures are reported in the material and method section. NMR experimental constraints, refinement and structural statistics are presented in Table 1. In addition, as a further analysis of the quality of the structure, we determined the structure of the protein with the same calculation and refinement protocols, but excluding the orientational information obtained from the ^{15}N-^{1}H amide RDCs. We then evaluated the agreement between this NMR ensemble calculated without RDC restraints and the measured RDCs. We calculated an average Q factor (Bax et al. 2001) of 49.3 %, which is in the range of typical Q factors calculated for structures determined only from NOE information (see also Supplementary Table 1 and Supplementary Figure 3).

The structure ensemble together with dynamic information obtained from ^{15}N{^{1}H}-NOE values revealed that UP1 folds as two typical RRM domains separated by a rather flexible interdomain linker. The N- and C-termini are also flexible as well as the long β2-β3 loops in each RRM, this character being more pronounced in RRM2 (see Figure 4a and Supplementary Figure 4). Each individual RRM is very well defined, with average backbone r.m.s.d. of 0.40 Å and 0.58 Å for RRM1 and RRM2, respectively (Table 1). This value is slightly increased for the entire UP1 domain (0.71 ± 0.16 Å), as reflected on the three different overlays of Figure 4a-c. Overall, the structure of each individual RRM is very similar to the previously determined crystallographic structures (Shamoo et al. 1997; Xu et al. 1997) but small local variations exist between our solution NMR structure and crystal structures and will be described in details in the following sections.

Description of the interface between RRM1 and RRM2

The nature of the interface in our solution structure is at first sight similar to the interface observed in previous crystal structures (Shamoo et al. 1997; Xu et al. 1997; Ding et al. 1999). It involves residues from α0, α2 and β4 in RRM1 and residues from α2 in RRM2 (Figure 4e). Note that the regions of interaction perfectly match with the very small chemical shifts differences observed between UP1 and the isolated RRMs (Figure 1e-f). These interactions may be divided in three elements: (*i*) a small hydrophobic cluster involving Leu13, Ile164, Val90 and the aliphatic part of Arg88 side chain on one side of the interface (*top* of Figure 4e); (*ii*) two Arg-Asp salt-bridges, namely Arg88-Asp157 and Arg75-Asp155, in the middle and on the other side of the interface, respectively (*bottom* of Figure 4e); and (*iii*) a central residue, H156, which is sandwiched between the hydrophobic part of Lys87 on one side and the Arg75-Asp155 salt-bridge on the other side and also interacts with Met72 side chain (Figure 4e). Most of these contacts led to a direct spectroscopic evidence in the 3D ^{13}C-edited half-filter NOESY (Figure 3a and b). However, no direct spectroscopic information could be obtained on the two salt-bridges, as the closest observable protons across a salt-bridge of this type, namely Arg-Hδs and Asp-Hβs, are about 7-8 Å apart. The position of these side chains was therefore defined from electrostatic properties and NOEs to surrounding side-chains, and no additional restraints were included to force the formation of these salt-bridges. Nevertheless, each Arg-Asp salt-bridge is present in about 2/3 of the 20 structures of the final NMR ensemble, which strongly support their existence in solution, similarly to what was seen in the crystal structures.

Dynamical study of hnRNP A1 RRMs

In order to bring additional evidence that hnRNP A1 RRM1 and RRM2 are indeed interacting in solution, we wanted to evaluate this aspect using an independent method that would corroborate the experiments obtained with the segmental labeling approach. We therefore performed with NMR a dynamic study for both the single RRM constructs (*i.e.* RRM1 and RRM2) and the RRM12 double-domain construct (UP1). We measured for this three constructs, ^{15}N T1 and T2 relaxation times, as described in the material and methods section. Overall correlation times for each constructs (τ_c) were derived from T1/T2 ratio of dispersed and rigid amide resonances, assuming isotropic motion (Fushman et al. 1994). Overall rotational correlation times of 8.2 ± 0.4 ns and 10.9 ± 1.1 ns were obtained for RRM1 and RRM2

in isolation, respectively, whereas τ_c increases significantly up to 15.8 ± 0.6 ns for the double-domain construct UP1 (Table 2). These values are in good agreement with reported values for domains of these sizes (*i.e.* 10.9 kDa, 11.2 kDa and 22.1 kDa for RRM1, RRM2 and UP1, respectively) (Dayie et al. 1996) and definitely support that RRM1 and RRM2 are interacting in solution.

Comparison of the interdomain interface and of the relative orientation of the individual RRMs between the NMR solution structure and crystal structures of UP1

As reported in a previous paragraph, the different crystal structures of free UP1 are almost indistinguishable and this is also the case for the different crystal structures of UP1 bound to DNA. To simplify our analysis of the differences between our NMR structure and the different crystal structures, we decided to retain only one crystal structure for each form of the protein, free and bound, and to keep the most representative of each class, namely the ones that displayed the lowest backbone r.m.s.d to the other structures of their group. Interestingly, they correspond in both cases to structures with the highest resolution, *i.e.* 1.1 Å for free UP1 (pdb code 1L3K, (Vitali et al. 2002)), and 1.8 Å for bound UP1 (pdb code 1U1R, (Myers and Shamoo 2004)). Similarly, a representative structure of our NMR ensemble was chosen using the similar criterion, *i.e.* the lowest backbone r.m.s.d to the other structures of the ensemble, and remarkably it was also the lowest energy structure of our ensemble. These structures will be thereafter called $UP1_{free}$, $UP1_{bound}$ and $UP1_{NMR}$.

The three structures have a very similar overall organization and can be superimposed on the entire UP1 domains with a relatively good agreement (Figure 5a). However, it has already been mentioned that there is a significant conformational change between $UP1_{free}$ and $UP1_{bound}$ that corresponds to a rotation of ~15° of one RRM in respect to the other (Ding et al. 1999). In order to emphasize this conformational change between $UP1_{free}$ and $UP1_{bound}$, we also superimposed the three structures on RRM1 only and could then visualize better the differences in the orientation of RRM2 (Figure 5b and c). Surprisingly, it appeared very clearly that the relative orientation of the two RRMs in $UP1_{NMR}$, which is a free structure of UP1, more closely resembles the crystal structure of $UP1_{bound}$, and less the one of $UP1_{free}$ (Figure 5b and c). To confirm numerically our visual observation, we calculated pairwise backbone r.m.s.d. between all the different RRMs and UP1s structures. These data are presented in Table 3. It confirmed that $UP1_{NMR}$ is much closer to $UP1_{bound}$ than to $UP1_{free}$ (backbone r.m.s.d of 1.18 Å and 2.18 Å, respectively). Importantly, the comparison of the individual RRMs con-

firmed that these large differences are not due to local differences in the RRM themselves, but really to the different orientations adopted by the RRMs in the different structures (Table 3).

In order to strengthen this observation, we sought at finding unambiguous spectroscopic evidences that would demonstrate that these differences between $UP1_{free}$ and $UP1_{NMR}$ are not due to indirect effects of the structure calculation methodology, but really reflect true differences between solution and crystal structure. We first used the RDC information to evaluate the agreement between the measured NH RDCs and the two different crystal structures. We back-calculated the NH RDCs after best-fitting the alignment tensor to both crystal structures ($UP1_{free}$ and $UP1_{bound}$), and compared them with the set of measured RDCs. Overall, $UP1_{bound}$ agrees better with the measured RDCs than $UP1_{free}$ (Q factors of 47.4 % and 55.4 % for $UP1_{bound}$ and $UP1_{free}$, respectively – Supplementary Table 1). This shows that the measured RDCs could discriminate, independently of any NMR structure determination, the subtle differences associated with the two different relative orientations of the RRMs in the two different crystal structures. This further supports our conclusion regarding the origin of the domain re-orientation between the two crystal structures. In addition, similarly to what we observed for the $UP1_{NMR}$ structure refined with RDCs, the NMR structure calculated without RDC restraints more closely resembles $UP1_{bound}$ than $UP1_{free}$ and gives comparable RDC Q factor (Supplementary Table 1 and Supplementary Figure 5), indicating that the inter-domain NOE data obtained from the segmentally labeled sample would have been sufficient to notice the differences regarding the structural agreement of our NMR structure with $UP1_{free}$ and $UP1_{bound}$. This means that the inter-domain NOE data should contain enough information to discriminate the two conformations. Therefore, we next analysed the inter-domain NOE cross peaks. Due to the geometrical property of this domain rotation, the N-terminal end of helix $\alpha 2$ in RRM2 is barely affected whereas its C-terminal end shows larger amplitude deviations (Figure 5c). Interestingly, the distances between Val90 and Ile164 in $UP1_{free}$ are large (between 8 and 11 Å, Figure 5d) and therefore should not result in any NOE cross-peaks between these side-chains, whereas these two residues are much closer in the structures $UP1_{bound}$ and $UP1_{NMR}$ (Figure 5d) potentially enabling NOE cross peaks to be measured. Such NOE cross peaks were indeed observed between Val90 and Ile164 in the 3D ^{13}C-edited half-filter NOESY (Figure 3 and Supplementary Figure 6), confirming that our NMR measurements are incompatible with the $UP1_{free}$ crystal structure. This further emphasizes the efficiency of the segmental labeling approach for the precise determination of multi-domain protein structures.

Altogether, our NMR structure of UP1 free revealed that RRM1 and RRM2 are truly interacting in solution, and that the relative orientation of these RRMs most closely resembles the one observed in the DNA bound form of UP1. Therefore, it strongly suggests that the two RRMs of hnRNP A1 have in solution a single defined relative orientation that in unchanged upon DNA binding, and that the relative domain movement observed between the two crystal forms of the protein is not a consequence of nucleic acid binding, but rather of differences in the protein-protein contacts between neighbouring proteins in the two different crystal lattices.

Discussion

In this study, we have used segmental isotope labeling to determine the solution structure of the two RRMs of hnRNP A1. This labeling strategy was crucial to unambiguously prove that hnRNP A1 RRMs interact in solution, and this important point could not have been derived solely on the basis of chemical shift difference between the tandem construct and the isolated domains. In addition, the solution structure revealed that the difference in relative orientation of the two RRMs, as observed in the two different crystal forms of the protein (free and bound to DNA), is very likely due to the different crystal lattices rather than due to the binding of nucleic acid. Although the two crystal structures and the NMR solution structures are very similar overall, some local differences exist and their importance will be discussed below. Additionally, this study established unequivocally that in solution, the two RRMs of hnRNP A1 interact with each other in the free form of the protein. This does not represent the most common situation in proteins containing multiple RRMs, and we propose to review hereafter how different arrangement between several RRMs can influence nucleic acid recognition on the basis of the available multiple RRM structures, free and bound to nucleic acid.

Interacting domains and NMR spectroscopy

When we recorded the first NMR spectra of hnRNP A1 RRM12 construct (UP1) and of the isolated RRMs, we were surprised to observe an almost perfect overlay for these two RRMs although several crystal structures showed that the RRMs could interact with each other (Shamoo et al. 1997; Xu et al. 1997). By experience, when ^{1}H,^{15}N-HSQCs overlay almost perfectly, like observed for hnRNP A1 RRMs, one often conclude that the two domains are independent domains (Skrisovska and Allain 2008; Dominguez and Allain 2006; Oberstrass et al. 2005). However, one can validate or not such conclusions using NMR relaxation measurements (^{15}N T1 and T2 relaxation times), and the case of hnRNP A1 prove the need of performing such additional analyses. Indeed, although marked differences between the NMR footprints of isolated and combined domains can be taken as a strong evidence of an interdomain interaction, one should not deduce from a virtually perfect overlay that the domains are independent. In such case, there is a real need for unambiguous methods like NMR relaxation measurements or segmental isotope labeling in combination with interdomain NOE measurements, in order to conclude about possible interdomain interactions. Furthermore, in this particular case of hnRNP A1 RRM12, automated structure calculations performed on the

only basis of regular NOESY spectra measured on fully labeled samples did not converge towards a compact globular domain with interacting RRMs, but showed independent folded RRMs separated by a flexible linker (data not shown). Similarly, structure calculations performed with the RDC information, but without the interdomain NOE data, did not converge towards a single relative position of the two RRMs but led to independent RRMs separated by a flexible linker. In other words, structure calculations performed without the information of the interdomain contacts obtained with the segmentally labeled sample failed to reliably detect and assign interdomain NOEs, demonstrating the need for unambiguous methods and the importance of our segmental labeling strategy. In the case of hnRNP A1 RRMs, we believe that the almost perfect overlays of the NMR footprints (Figure 1) is due to the particular nature of the interface between the RRMs, as compared with PTB RRM3 and RRM4 interface for instance (Oberstrass et al. 2005; Vitali et al. 2006). In PTB RRM34, the interdomain interface involves many hydrophobic side chains forming an important hydrophobic core (Vitali et al. 2006). In hnRNP A1, there is not such an extended hydrophobic core, with an interface composed primarily of a small hydrophobic patch and two Arg-Asp salt-bridges (Figure 4e). In such interface, backbone amide resonances do not seem to be a very sensitive NMR probe. Chemical shifts of the side chains directly involved in the interdomain interface would probably experience larger changes, but it is rather difficult to monitor such changes in initial sample evaluations that are often performed with only ^{15}N-labeled samples. The proper decision on whether domains are interacting or not is crucial as it can latter strongly influence and restrict our understanding of the different mode of nucleic acid binding accessible to a particular multi-domain protein (see Supplementary Figure 1 and paragraph below), and one should therefore pay particular attention to this aspect. This point is of very broad relevance, since among eukaryotic proteins, the presence of multiple RRMs is very common and is estimated to occur in about 44% of the proteins containing at least one RRM (Maris et al. 2005). In addition, the RRM is one of the most abundant protein domains, and proteins with multiple RRMs are estimated to be present in about 1% of human gene products.

Comparison between hnRNP A1 solution and crystal structures

We showed in this study that the relative orientation of the two RRMs of hnRNP A1 in solution closely resembles the one observed in the crystal structure of UP1 bound to DNA (Figure 5), indicating that the two RRMs of hnRNP A1 have a single defined relative orientation in solution, and that the small domain movement observed between the two crystal struc-

tures of the protein (free and bound) is not resulting from nucleic acid binding as previously proposed. The only differences in the side chains interaction that we could find to rationalize the fact that $UP1_{bound}$ would be more stable than $UP1_{free}$ are located in the small hydrophobic cluster formed by Leu13, Ile164 and Arg88. In $UP1_{bound}$, Ile164 is more tightly packed in between the side chains of Leu13 and Arg88, as seen also in $UP1_{NMR}$, whereas in $UP1_{free}$, these side chains are further apart (Supplementary Figure 7). This could explain that $UP1_{bound}$ would be the native conformation of UP1, as observed in solution by NMR, and $UP1_{free}$ a destabilized structure induced by crystallization.

In addition to this difference in domain orientation observed between the free solution structure ($UP1_{NMR}$) and the free crystal structure ($UP1_{free}$), some small local variations exist between these two structures (Supplementary Figure 8). $RRM1_{NMR}$ and $RRM1_{free}$ are almost indistinguishable (backbone r.m.s.d. of 0.55 Å for the entire domain, *i.e.* residues 11-89, Table 3). Even the long β2-β3 loop of $UP1_{NMR}$ agrees quite well with $UP1_{free}$ (Supplementary Figure 8). According to the heteronuclear $^{15}N\{^1H\}$-NOE values (Supplementary Figure 4), residues directly following β2 or directly preceding β3 are relatively rigid and only 4 residues have NOE values < 0.72. Differences between the two structures are more apparent in RRM2 where the β2-β3 and the β1-α1 loops differ significantly (Supplementary Figure 8). These two loops are also regions where the precision of the structure is lower (Figure 4c). This conformational heterogeneity observed in the β2-β3 loop correlates with the heteronuclear $^{15}N\{^1H\}$-NOE values, which are overall lower than the corresponding NOE values in RRM1 β2-β3 loop. The heterogeneity observed in the β1-α1 loop only reflects the lack of observable NMR signals and therefore the lack of chemical shift assignment for this region. This loop is most probably also dynamic, but we do not have clear evidence to support this point. Outside these two loop regions, $RRM2_{NMR}$ and $RRM2_{free}$ overlay quite well (backbone r.m.s.d. of 0.79 Å for residues 105-111, 117-139 and 146-180, Table 3). The only structural difference between rigid residues in $UP1_{NMR}$ and $UP1_{free}$ is located in the loop following helix α2 and involve residues Lys166 and Tyr167 (Supplementary Figure 8). In $UP1_{free}$, Lys166 makes intermolecular contacts with Glu66 of a symmetry related molecule (Shamoo et al. 1997; Xu et al. 1997), leading to a large distortion of the protein backbone for these two residues. Similarly, in the $UP1_{bound}$ structure, Lys166 and Tyr167 makes intermolecular contacts with Asp94 and the carbonyl group of Ile164, respectively, of a symmetry related molecule, leading to comparable distortions of this region (Ding et al. 1999). In some structures of the $UP1_{NMR}$ ensemble, Lys166 makes intramolecular contacts with Glu93 of the interdomain linker. As a

consequence, the long side chains of Lys166 and Tyr167 come in close proximity to the β4 strand which could have some implication for nucleic acid recognition, since this strand forms the main region of sequence specific contacts involved in nucleic acid recognition (Ding et al. 1999). Importantly, this large distortion might be related to the relative domain movement observed between $UP1_{free}$ and $UP1_{bound}$. Indeed, this region directly follows helix α2, which makes most of the interdomain contacts from RRM2, and Lys166 is very close in sequence to Ile164, which experience the largest displacement in this domain rotation of ~15° (Figure 5c). Furthermore, Lys166 and Tyr167 are engaged in different protein-protein contacts in the two different crystal lattices of $UP1_{free}$ and $UP1_{bound}$ structures. However, also the differences in this region might be coupled with the interdomain movement, we cannot exclude that the reorientation would come from additional crystal contacts or from totally different contacts in other regions of the protein.

Finally, two regions of the protein are not seen in the electron density maps of $UP1_{free}$, and become structured upon DNA binding as a consequence of being directly involved in nucleic acid recognition. These two regions correspond to the interdomain linker (residues Arg92 to Leu102), and the C-terminal segment after RRM2 (from residue Ser182). This perfectly correlates with the dynamic information obtained in solution (Supplementary Figure 4), where these residues have $^{15}N\{^{1}H\}$-NOE values reflecting flexible residues. But interestingly, even if these regions are structurally heterogeneous (Figure 4a-c), the secondary structured elements that appear upon DNA binding in these regions – namely a short α helical turn in the interdomain linker, involving residues 93-96, and a C-terminal α-helix after RRM2 involving residues 183-188 – seem to some extent to be already present in the free form. The helical propensity of these two regions is supported by backbone chemical shift and few NOE cross-peaks (data not shown).

Repertoire of nucleic acid recognition modes in proteins containing multiple RRM domains

Since RRM domains are asymmetric binding platforms, the relative orientation of different RRMs in proteins containing multiple-RRMs strongly influences the modes of nucleic acid binding by these particular proteins. In other words, the path of the nucleic acid molecule bound to multiple RRMs is strongly dependent on whether these RRMs are interacting with each other and adopt a single defined relative orientation. A limited number of simplified situations can be used to describe the interplay between two RRMs and a nucleic acid molecule, depending on whether the RRMs are interacting or not in their free state, and whether they

interact or not in the nucleic acid bound state. To date there is no example of RRMs interacting in their free state and not in their bound state, but all the other situations are supported by structural work (Figure 6a-e). Additionally, the dynamic aspects related to nucleic acid recognition are definitely essential for multi-domain proteins and will be briefly mentioned below. However, readers interested in more details on these aspects may also refer to the excellent review by Mackereth and Sattler (2012).

In the situation #1, the RRMs are not interacting in the free form and interact upon nucleic acid binding to form a continuous binding platform (Figure 6a). Many different proteins can be classified in this group, and structural data are very abundant for this particular case, with for instance the different structures of nucleolin (Allain et al. 2000b; Allain et al. 2000a; Johansson et al. 2004; Arumugam et al. 2010), PABP (Deo et al. 1999; Safaee et al. 2012), Hrp1 (Perez-Canadillas 2006; Leeper et al. 2010), Sex-lethal (Handa et al. 1999; Crowder et al. 1999) and HuD (Wang and Tanaka Hall 2001). In these cases, the association of the two RRM platforms allows the continuous recognition of longer nucleic acid stretches (6-10 nucleotides), which often strongly increases the binding affinity compare to isolated domains. Interestingly, in all these different cases, RRM2 binds the 5' end of the RNA and RRM1 the 3' end. Remarkably, the relative orientation of the RRMs and the path of the RNA on the RRM platforms are very similar for different unrelated proteins of this class, namely Sex-lethal, Hrp1 and HuD. In these structures, the interdomain interface is relatively small (~350 $Å^2$) and the interaction is mediated by a limited set of contacts, mainly one single salt-bridge and few hydrogen bonds, and might therefore be quite weak. This might explain why these contacts are induced by nucleic-acid biding and are not present in the free forms of the proteins. In addition, in cases where tandem RRMs could not be crystalized bound to the same RNA molecule (CUGBP1 for example), models have been proposed that have orientation similar to HuD/Sex-lethal/Hrp1 or to PABP (Teplova et al. 2010). However, there is to date no evidence to support one model or the other and it is therefore possible that none of these two models adequately describe CUGBP1 RRM12 binding to RNA. Overall, according to the currently available structural informations, this situation of independent RRMs interacting upon RNA binding seems to be the most common in tandem RRMs although the number of example is still scarce. It is very possible that this particular case might be over-represented in the solved structures since such stable and compact complexes would be more susceptible to crystallize than independent and dynamic RRMs.

In situation #2, the RRMs are interacting in the free form and adopt a single defined

orientation resulting in a discontinuous antiparallel platform that is maintained in the bound form (Figure 6b). This particular topology can induce RNA loops in the bound RNA and have a role for regulating alternative splicing. This looping capacity of interacting tandem RRMs has been demonstrated in the case of PTB RRM34 (Oberstrass et al. 2005; Lamichhane et al. 2010). Such topology with a discontinuous antiparallel platform is also found in the two tandem RRMs of hnRNP A1 suggesting that it might also be able to loop out RNA. However, there is to date no evidence to support such mechanism of action. In these two examples, the surface of interaction between the RRMs is larger than the ones found in the situation #1 (namely ~630 $Å^2$ and ~850 $Å^2$ for hnRNP A1 and PTB RRM34, respectively). These interfaces involve a combination of electrostatic and hydrophobic interactions, with even a large interdomain hydrophobic core in the case of PTB RRM34. It could have been questionable to classify hnRNP A1 in this group if our analysis was only based on the existing crystal structures, since the relative orientation could seem to be altered upon nucleic acid binding. However, our solution structure supports the fact that hnRNP A1 RRMs exist in a single defined orientation, as observed for PTB RRM34.

In situation #3, the RRMs are neither interacting in the free form nor in the bound form (Figure 6c). This case is likely to be quite common considering the high number of cases where tandem RRMs are separated by long and disordered interdomain linkers, but structural data supporting this mode of binding are not very abundant. In most of the studies, the independence of the RRMs is well established in the free form, and is then assumed for the bound form as well, even if this point is not always clearly demonstrated. This mode of binding was proposed for instance for PTB RRM12 (Oberstrass et al. 2005), hnRNP F RRM12 (Dominguez and Allain 2006) and Npl3p RRM12 (Skrisovska and Allain 2008). In these cases, the binding of multiple RRMs to the same nucleic acid molecule increases the overall binding affinity (Shamoo et al. 1995). Since the binding of an individual RRM to RNA can be rather weak (K_D ~ 1 μM), this cooperation is an essential aspect enabling these multi-domain proteins to achieve their function in alternative splicing at low cellular concentration.

In situation #4, the RRMs are interacting in the free form such that one binding platform is occluded by the other RRM. The RRMs are also interacting in their bound form, but in a different relative orientation that forms a continuous binding platform (Figure 6d). This case occurs in the splicing factor U2AF65, where the recognition of a poly-pyrimidine tract RNA is associated with an equilibrium between a closed state and an open state competent for RNA binding (Mackereth et al. 2011). The presence of this closed state results in an autoinhi-

bition of binding that is used to finely tune U2AF binding to various 3' splice sites harbouring different pyrimidine tracts. In both states, the surfaces of interaction between the two RRMs are quite small (~460 Å^2 and ~320 Å^2 for the closed and open states, respectively) and are stabilized by a limited set of electrostatic interactions, namely one potential salt-bridge and few contacts between polar side chains in the closed state of the protein, and a small number of contacts between polar side chains in the open state. This case illustrates the importance of the dynamics and the role of weak interdomain contacts for the binding of multi-RRM domains to RNA, in relation to a subtle regulation of a complex biological mechanism. This equilibrium could not have been deduced from the static crystal structure of U2AF65 (Sickmier et al. 2006), confirming the importance of solution techniques for the analysis of the interaction between multi-domain proteins and nucleic acids.

In situation #5, the RRMs are interacting in the free form such that one binding platform is occluded by the other RRM. This interaction is preserved in the bound form and thus only one RRM can bind to nucleic acid (Figure 6e). This situation has been observed in the transcriptional repressor FIR (Crichlow et al. 2008; Cukier et al. 2010), and could possibly describe as well the situation observed in the splicing factor Prp24 (Bae et al. 2007; Martin-Tumasz et al. 2010). In FIR, the binding platform of RRM2 is occluded by RRM1 through a very large interaction surface involving RRM1 helices (~900 Å^2), leaving only RRM1 binding platform available for DNA binding (Figure 6e). A recent NMR study showed that the occluded RRM2, which cannot bind to DNA, is involved in protein-protein interaction with the transcriptional activator FBP (Cukier et al. 2010), whereas the initial crystal structure favoured a model in which RRM2 would drive the dimerization of FIR (Crichlow et al. 2008). There is a similar arrangement in the three tandem RRMs of Prp24, where RRM2 forms extensive interdomain contacts with RRM1 and RRM3 in the crystal structure of RRM123 (Bae et al. 2007). However, whereas the RRM1-RRM2 interaction is preserved in solution (Bae et al. 2007), RRM2 and RRM3 do not interact in solution, leaving the RRM2 binding platform available for RNA binding of a segment of the U6 snRNA (Martin-Tumasz et al. 2010). The role of the occluded RRM is not as clear as in the case of FIR, but it has been proposed that RRM1 would also interact with U6 snRNA, also with a different surface than the canonical β-sheet platform (Bae et al. 2007). Further data would be needed to generalize these observations, but these two examples suggest that when an RRM binding platform is occluded by a preceding or a following RRM, the occluded RRM could function in protein-protein interaction, or could bind to nucleic acid in a non-canonical manner.

Conclusion

This overview of the structural characterization of tandem RRM proteins and their interaction with nucleic acid illustrates the large repertoire of nucleic acid recognition modes in proteins containing multiple RRM domains. Noticeably, large surfaces of interaction between tandem RRMs (~600-900 $Å^2$) are associated with a static behaviour and a defined relative orientation of the domains (situations #2 and #5), whereas smaller surfaces (~300-500 $Å^2$) correlate with a dynamic behaviour and a re-orientation of the RRMs upon nucleic acid binding (situations #1 and #4). In addition, several examples directly indicate that solution techniques are essential to investigate the interaction of tandem RRMs and more generally multi-domain proteins with nucleic acids. Crystallization can sometimes confine protein domains in non-functional conformations or induce artifactual nucleic-acid topology of binding. To obtain reliable information, in solution, on the relative orientation and on the surface of interaction of different domains of multi-domain proteins, segmental isotope labeling is definitely a method of choice. And considering the progress that have been realized in the last years in developing new methods and/or improving existing protocols for segmentally labeled sample production, we believe that segmental labeling will be a central approach for the future investigations of multi-domain nucleic-acid-binding proteins.

Accession numbers

The chemical shifts of hnRNP A1 RRM12 have been deposited in the BioMagResBank under accession number 18728. The coordinates of the structure have been deposited in the Protein Data Bank under accession code 2LYV.

Acknowledgment

We thank Prof. Adrian R. Krainer for the initial pET9d-hnRNPA1 plasmid and Dr. Erich Michel for vectors and helpful recommendations concerning segmentally labeled sample production. This project was supported by the Swiss National Science Foundation NCCR structural biology. PB was supported by the Postdoctoral ETH Fellowship Program and the Novartis Foundation, formerly Ciba-Geigy Jubilee Foundation.

References

Allain FH, Bouvet P, Dieckmann T, Feigon J (2000a) Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. EMBO J 19 (24):6870-6881.

Allain FH, Gilbert DE, Bouvet P, Feigon J (2000b) Solution structure of the two N-terminal RNA-binding domains of nucleolin and NMR study of the interaction with its RNA target. J Mol Biol 303 (2):227-241.

Anderson LL, Marshall GR, Crocker E, Smith SO, Baranski TJ (2005) Motion of carboxyl terminus of Galpha is restricted upon G protein activation. A solution NMR study using semisynthetic Galpha subunits. J Biol Chem 280 (35):31019-31026.

Arumugam S, Miller MC, Maliekal J, Bates PJ, Trent JO, Lane AN (2010) Solution structure of the RBD1,2 domains from human nucleolin. J Biomol NMR 47 (1):79-83.

Bae E, Reiter NJ, Bingman CA, Kwan SS, Lee D, Phillips GN, Jr., Butcher SE, Brow DA (2007) Structure and interactions of the first three RNA recognition motifs of splicing factor prp24. J Mol Biol 367 (5):1447-1458.

Barraud P, Emmerth S, Shimada Y, Hotz HR, Allain FH, Buhler M (2011) An extended dsRBD with a novel zinc-binding motif mediates nuclear retention of fission yeast Dicer. EMBO J 30 (20):4223-4235.

Bax A, Kontaxis G, Tjandra N (2001) Dipolar couplings in macromolecular structure determination. Methods Enzymol 339:127-174.

Brunger AT (2007) Version 1.2 of the Crystallography and NMR system. Nat Protoc 2 (11):2728-2733.

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54 (Pt 5):905-921.

Cáceres JF, Stamm S, Helfman DM, Krainer AR (1994) Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. Science 265 (5179):1706-1709.

Camarero JA, Shekhtman A, Campbell EA, Chlenov M, Gruber TM, Bryant DA, Darst SA, Cowburn D, Muir TW (2002) Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. Proc Natl Acad Sci U S A 99 (13):8536-8541.

Casas-Finet JR, Karpel RL, Maki AH, Kumar A, Wilson SH (1991) Physical studies of tyrosine and tryptophan residues in mammalian A1 heterogeneous nuclear ribonucleoprotein. Support for a segmented structure. J Mol Biol 221 (2):693-709.

Chen J, Li Q, Wang J (2011) Topology of human apolipoprotein E3 uniquely regulates its diverse biological functions. Proc Natl Acad Sci U S A 108 (36):14813-14818.

Chen J, Wang J (2011) A segmental labeling strategy for unambiguous determination of domain-domain interactions of large multi-domain proteins. J Biomol NMR 50 (4):403-410.

Cordier F, Dingley AJ, Grzesiek S (1999) A doublet-separated sensitivity-enhanced HSQC for the determination of scalar and dipolar one-bond J-couplings. J Biomol NMR 13 (2):175-180.

Crichlow GV, Zhou H, Hsiao HH, Frederick KB, Debrosse M, Yang Y, Folta-Stogniew EJ, Chung HJ, Fan C, De la Cruz EM, Levens D, Lolis E, Braddock D (2008) Dimerization of FIR upon FUSE DNA binding suggests a mechanism of c-myc inhibition. EMBO J 27 (1):277-289.

Crowder SM, Kanaar R, Rio DC, Alber T (1999) Absence of interdomain contacts in the crystal structure of the RNA recognition motifs of Sex-lethal. Proc Natl Acad Sci U S A 96 (9):4892-4897.

Cukier CD, Hollingworth D, Martin SR, Kelly G, Diaz-Moreno I, Ramos A (2010) Molecular basis of FIR-mediated c-myc transcriptional control. Nat Struct Mol Biol 17 (9):1058-1064.

David R, Richter MP, Beck-Sickinger AG (2004) Expressed protein ligation. Method and applications. Eur J Biochem 271 (4):663-677.

Dayie KT, Wagner G, Lefèvre JF (1996) Theory and practice of nuclear spin relaxation in proteins. Annu Rev Phys Chem 47:243-282.

Deo RC, Bonanno JB, Sonenberg N, Burley SK (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell 98 (6):835-845.

Ding J, Hayashi MK, Zhang Y, Manche L, Krainer AR, Xu RM (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. Genes Dev 13 (9):1102-1115.

Dominguez C, Allain FH-T (2006) NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition. Nucleic Acids Res 34 (13):3634-3645.

Doreleijers JF, Sousa da Silva AW, Krieger E, Nabuurs SB, Spronk CAEM, Stevens TJ, Vranken WF, Vriend G, Vuister GW (2012) CING: an integrated residue-based structure validation program suite. J Biomol NMR 54 (3):267-283.

Dosset P, Hus JC, Marion D, Blackledge M (2001) A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. J Biomol NMR 20 (3):223-231.

Dreyfuss G, Matunis MJ, Piñol-Roma S, Burd CG (1993) hnRNP proteins and the biogenesis of mRNA. Annu Rev Biochem 62:289-321.

Flynn RL, Centore RC, O'Sullivan RJ, Rai R, Tse A, Songyang Z, Chang S, Karlseder J, Zou L (2011) TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. Nature 471 (7339):532-536.

Fushman D, Weisemann R, Thüring H, Rüterjans H (1994) Backbone dynamics of ribonuclease T1 and its complex with 2'GMP studied by two-dimensional heteronuclear NMR spectroscopy. J Biomol NMR 4 (1):61-78.

Garrett DS, Lodi PJ, Shamoo Y, Williams KR, Clore GM, Gronenborn AM (1994) Determination of the secondary structure and folding topology of an RNA binding domain of mammalian hnRNP A1 protein using three-dimensional heteronuclear magnetic resonance spectroscopy. Biochemistry 33 (10):2852-2858.

Goddard T, Kneller D (2006) SPARKY 3. University of California, San Francisco.

Guil S, Cáceres JF (2007) The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. Nat Struct Mol Biol 14 (7):591-596.

Guntert P (2004) Automated NMR structure calculation with CYANA. Methods Mol Biol 278:353-378.

Guntert P, Mumenthaler C, Wuthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273 (1):283-298.

Handa N, Nureki O, Kurimoto K, Kim I, Sakamoto H, Shimura Y, Muto Y, Yokoyama S (1999) Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. Nature 398 (6728):579-585.

Herrmann T, Guntert P, Wuthrich K (2002a) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319 (1):209-227.

Herrmann T, Guntert P, Wuthrich K (2002b) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. J Biomol NMR 24 (3):171-189.

Johansson C, Finger LD, Trantirek L, Mueller TD, Kim S, Laird-Offringa IA, Feigon J (2004) Solution structure of the complex formed by the two N-terminal RNA-binding domains of nucleolin and a pre-rRNA target. J Mol Biol 337 (4):799-816.

Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 28 (23):8972-8979.

LaBranche H, Dupuis S, Ben-David Y, Bani MR, Wellinger RJ, Chabot B (1998) Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. Nat Genet 19 (2):199-202.

Lamichhane R, Daubner GM, Thomas-Crusells J, Auweter SD, Manatschal C, Austin KS, Valniuk O, Allain FH-T, Rueda D (2010) RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression. Proc Natl Acad Sci U S A 107 (9):4105-4110.

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8 (4):477-486.

Leeper TC, Qu X, Lu C, Moore C, Varani G (2010) Novel protein-protein contacts facilitate mRNA 3'-processing signal recognition by Rna15 and Hrp1. J Mol Biol 401 (3):334-349.

Mackereth CD, Madl T, Bonnal S, Simon B, Zanier K, Gasch A, Rybin V, Valcarcel J, Sattler M (2011) Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. Nature 475 (7356):408-411.

Mackereth CD, Sattler M (2012) Dynamics in multi-domain protein recognition of RNA. Curr Opin Struct Biol 22 (3):287-296.

Maris C, Dominguez C, Allain FH-T (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J 272 (9):2118-2131.

Martin-Tumasz S, Reiter NJ, Brow DA, Butcher SE (2010) Structure and functional implications of a complex containing a segment of U6 RNA bound by a domain of Prp24. RNA 16 (4):792-804.

Mayeda A, Krainer AR (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. Cell 68 (2):365-375.

Mayeda A, Munroe SH, Xu RM, Krainer AR (1998) Distinct functions of the closely related tandem RNA-recognition motifs of hnRNP A1. RNA 4 (9):1111-1123.

Michel E, Skrisovska L, Wüthrich K, Allain FH (2013) Amino acid-selective segmental isotope labeling of multi-domain proteins for structural biology. Chembiochem, in press.

Michlewski G, Cáceres JF (2010) Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. Nat Struct Mol Biol 17 (8):1011-1018.

Michlewski G, Guil S, Semple CA, Cáceres JF (2008) Posttranscriptional regulation of miRNAs harboring conserved terminal loops. Mol Cell 32 (3):383-393.

Minato Y, Ueda T, Machiyama A, Shimada I, Iwai H (2012) Segmental isotopic labeling of a 140 kDa dimeric multi-domain protein CheA from Escherichia coli by expressed protein ligation and protein trans-splicing. J Biomol NMR 53 (3):191-207.

Muralidharan V, Muir TW (2006) Protein ligation: an enabling technology for the biophysical analysis of proteins. Nat Methods 3 (6):429-438.

Myers JC, Moore SA, Shamoo Y (2003) Structure-based incorporation of 6-methyl-8-(2-deoxy-beta-ribofuranosyl)isoxanthopteridine into the human telomeric repeat DNA as a probe for UP1 binding and destabilization of G-tetrad structures. J Biol Chem 278 (43):42300-42306.

Myers JC, Shamoo Y (2004) Human UP1 as a model for understanding purine recognition in the family of proteins containing the RNA recognition motif (RRM). J Mol Biol 342 (3):743-756.

Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. Nature 463 (7280):457-463.

Oberstrass FC, Auweter SD, Erat M, Hargous Y, Henning A, Wenter P, Reymond L, Amir-Ahmady B, Pitsch S, Black DL, Allain FH-T (2005) Structure of PTB bound to RNA: specific binding and implications for splicing regulation. Science 309 (5743):2054-2057.

Pelton JG, Torchia DA, Meadow ND, Roseman S (1993) Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques. Protein Sci 2 (4):543-558.

Perez-Canadillas JM (2006) Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1. EMBO J 25 (13):3167-3178.

Piñol-Roma S, Dreyfuss G (1992) Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. Nature 355 (6362):730-732.

Pontius BW, Berg P (1990) Renaturation of complementary DNA strands mediated by purified mammalian heterogeneous nuclear ribonucleoprotein A1 protein: implications for a mechanism for rapid molecular assembly. Proc Natl Acad Sci U S A 87 (21):8403-8407.

Rückert M, Otting G (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. Journal of the American Chemical Society 122 (32):7793-7797.

Safaee N, Kozlov G, Noronha AM, Xie J, Wilds CJ, Gehring K (2012) Interdomain Allostery Promotes Assembly of the Poly(A) mRNA Complex with PABP and eIF4G. Mol Cell 48 (3):375-386.

Shamoo Y, Abdul-Manan N, Patten AM, Crawford JK, Pellegrini MC, Williams KR (1994) Both RNA-binding domains in heterogenous nuclear ribonucleoprotein A1 contribute toward single-stranded-RNA binding. Biochemistry 33 (27):8272-8281.

Shamoo Y, Abdul-Manan N, Williams KR (1995) Multiple RNA binding domains (RBDs) just don't add up. Nucleic Acids Res 23 (5):725-728.

Shamoo Y, Krueger U, Rice LM, Williams KR, Steitz TA (1997) Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 A resolution. Nat Struct Biol 4 (3):215-222.

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44 (4):213-223.

Sickmier EA, Frato KE, Shen H, Paranawithana SR, Green MR, Kielkopf CL (2006) Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. Mol Cell 23 (1):49-59.

Skelton N, Palmer A, Akke M, Kordel J, Rance M, Chazin W (1993) Practical Aspects of Two-Dimensional Proton-Detected 15N Spin Relaxation Measurements. Journal of Magnetic Resonance, Series B 102 (3):253-264.

Skrisovska L, Allain FH-T (2008) Improved segmental isotope labeling methods for the NMR study of multidomain or large proteins: application to the RRMs of Npl3p and hnRNP L. J Mol Biol 375 (1):151-164.

Skrisovska L, Schubert M, Allain FH (2010) Recent advances in segmental isotope labeling of proteins: NMR applications to large proteins and glycoproteins. J Biomol NMR 46 (1):51-65.

Southworth MW, Amaya K, Evans TC, Xu MQ, Perler FB (1999) Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein. Biotechniques 27 (1):110-114, 116, 118-120.

Teplova M, Song J, Gaw HY, Teplov A, Patel DJ (2010) Structural insights into RNA recognition by the alternate-splicing regulator CUG-binding protein 1. Structure 18 (10):1364-1377.

Vitali F, Henning A, Oberstrass FC, Hargous Y, Auweter SD, Erat M, Allain FH-T (2006) Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. EMBO J 25 (1):150-162.

Vitali J, Ding J, Jiang J, Zhang Y, Krainer AR, Xu R-M (2002) Correlated alternative side chain conformations in the RNA-recognition motif of heterogeneous nuclear ribonucleoprotein A1. Nucleic Acids Res 30 (7):1531-1538.

Wang X, Tanaka Hall TM (2001) Structural basis for recognition of AU-rich element RNA by the HuD protein. Nat Struct Biol 8 (2):141-145.

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67 (Pt 4):235-242.

Xu RM, Jokhan L, Cheng X, Mayeda A, Krainer AR (1997) Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. Structure 5 (4):559-570.

Yagi H, Tsujimoto T, Yamazaki T, Yoshida M, Akutsu H (2004) Conformational change of H+-ATPase beta monomer revealed on segmental isotope labeling NMR spectroscopy. J Am Chem Soc 126 (50):16632-16638.

Yang X, Bani MR, Lu SJ, Rowan S, Ben-David Y, Chabot B (1994) The A1 and A1B proteins of heterogeneous nuclear ribonucleoparticles modulate 5' splice site selection in vivo. Proc Natl Acad Sci U S A 91 (15):6924-6928.

Zhang Q-S, Manche L, Xu R-M, Krainer AR (2006) hnRNP A1 associates with telomere ends and stimulates telomerase activity. RNA 12 (6):1116-1128.

Zhang Y, Vasudevan S, Sojitrawala R, Zhao W, Cui C, Xu C, Fan D, Newhouse Y, Balestra R, Jerome WG, Weisgraber K, Li Q, Wang J (2007) A monomeric, biologically active, full-length human apolipoprotein E. Biochemistry 46 (37):10722-10732.

Zwahlen C, Legault P, Vincent SJF, Greenblatt J, Konrat R, Kay LE (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage λ N-peptide/boxB RNA complex. Journal of the American Chemical Society 119 (29):6711-6721.

FIGURES LEGENDS

Figure 1

NMR footprint of UP1 and single RRM1 and RRM2 from hnRNP A1

(**a**) (^{15}N,^{1}H)-HSQC spectrum of UP1 (residues 2-196). Positive signals are *in grey*, and negative signals from aliased peaks are *in green*. (**b**) Overlay of two (^{15}N,^{1}H)-HSQC spectra from single RRM1 (residues 2-97) *in deep blue* (positive signals) and *cyan* (negative signals) and from single RRM2 (residues 95-196) *in red* (positive signals) and *yellow* (negative signals). (**c-d**) Overlays of spectra to facilitate chemical shift comparison in UP1 and single RRM domains. Signals have same colours as in panel a and b. (**c**) Overlay of two (^{15}N,^{1}H)-HSQC spectra from single RRM1 (residues 2-97) and from UP1 (residues 2-196). (**d**) Overlay of two (^{15}N,^{1}H)-HSQC spectra from single RRM2 (residues 95-196) and from UP1 (residues 2-196). (**e-f**) Backbone amide chemical shift difference between UP1 (spectrum of panel a) and isolated RRM1 (**e**) and RRM2 (**f**) (spectra of panel c and d). (P) corresponds to proline residues, (-) to residues of the interdomain linker, and (*) to missing amide signals. Secondary structure elements are drawn above the histograms.

Figure 2

Segmentally labeled hnRNP A1 RRM12 for NMR structural investigations

(**a**) Constructs used to generate the C-terminal fragment (RRM2) with an N-terminal cysteine and the N-terminal fragment (RRM1) with a C-terminal reactive thioester. See also the material and method section and Supplementary Figure 2 for details on ligation reaction and product purification. (**b**) (^{15}N,^{1}H)-HSQC spectrum of ligated RRM12 with ^{13}C/^{15}N-labeled RRM1 and unlabeled RRM2. Positive signals are *in deep blue*, and negative signals from aliased peaks are *in cyan*. See also Figure 1 for a comparison of the (^{15}N,^{1}H)-HSQC spectra.

Figure 3

Interdomain contacts as determined with segmentally labeled hnRNP A1 RRM12

(**a**) Series of cross-sections from the 3D ^{13}C-edited half-filtered NOESY measured on the segmentally labeled hnRNP A1 RRM12 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.

Figure 4

NMR solution structure of hnRNP A1 RRM12

(**a-c**) NMR ensemble. Overlay of the 20 final structures with RRM1 *in blue*, RRM2 *in red,* and the interdomain linker *in yellow*. Structures were overlaid over the entire protein (residues 8-188) (**a**), or over each individual RRM, namely residues 8-88 for RRM1 (**b**) and residues 106-188 for RRM2 (**c**). (**d**) Cartoon drawing of a representative structure of the NMR ensemble. Side chains involved in the interdomain interface are shown as sticks *in green*. (**e**) Close-up view of the interdomain interface. The interdomain interface involve residues, shown as sticks *in green*, from α0, α2 and β4 in RRM1 and from α2 in RRM2. In addition, residues from the interdomain linker are also participating in the interface. See also Figure 3b.

Figure 5

Structural comparison of UP1 solution NMR structure with UP1 crystal structures

(**a**) Overlay of $UP1_{free}$ *in blue* (*i.e.* free UP1 crystal structure, pdb code 1L3K), $UP1_{bound}$ *in red* (*i.e.* DNA bound UP1 crystal structure, pdb code 1U1R), and $UP1_{NMR}$ *in yellow* over the entire domain (residues 11-89 and 105-180). (**b-c**) Two different overlays of $UP1_{NMR}$ *in yellow* with $UP1_{bound}$ *in red* (**b**), and with $UP1_{free}$ *in blue* (**c**). The 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 NMR structure overlays better with the $UP1_{bound}$ structure, see for example the large differences in the position of helices α1 and α2 in RRM2. See also Table 3 for a numerical comparison of the structures. (**d**) Close-up view of the large differences in the relative positions of residues Val90 and Ile164 between the different structures. The distances in $UP1_{bound}$ and $UP1_{NMR}$ are compatible with the observation of NOE transfer between these residues, whereas they are too far apart in $UP1_{free}$ for efficient NOE transfer. See also Figure 3 and Supplementary Figure 6 for an illustration of NOE transfers observed between these two residues.

Figure 6

Repertoire of nucleic acid binding modes by proteins containing two RRMs

Schematic representations of the different mode of binding of tandem RRMs to strands of nucleic acid are shown in the *left panels*, and archetypal structures corresponding to each mode

of binding are shown in their bound form in the *right panels*. N-terminal RRMs are *in yellow*, C-terminal RRMs *in red* and nucleic acid molecules *in blue*.

(**a**) (*Situation #1*) The RRMs are not interacting in the free form and interact upon nucleic acid binding to form a continuous binding platform. The RRMs do not have a defined relative orientation in the free form, but have a defined and unique relative orientation in the bound form. This case occurs in diverse proteins, namely PABP, nucleolin, Hrp1, Sex-lethal and HuD. The structure in the right panel represents Hrp1 bound to RNA (pdb code 2CJK) (Perez-Canadillas 2006). (**b**) (*Situation #2*) The RRMs are interacting in the free form to create a discontinuous antiparallel platform and interact in the bound form keeping the same relative orientation. This case occurs in PTB RRM34, and hnRNP A1 RRM12. The structure in the right panel represents PTB RRM3 and RRM4 bound to RNA (pdb code 2ADC) (Oberstrass et al. 2005). The path of the RNA from RRM3 to RRM4 is based on biochemical and biophysical data obtained in solution (Lamichhane et al. 2010). (**c**) (*Situation #3*) The RRMs are neither interacting in the free form nor in the bound form. This case is expected to be quite common, but structural data are not abundant to support this mode of binding. It was proposed for instance for PTB RRM12, hnRNP F RRM12 and Npl3p RRM12. The structure in the right panel corresponds to PTB RRM1 and RRM2 bound to RNA (pdb codes 2AD9 and 2ADB) (Oberstrass et al. 2005). (**d**) (*Situation #4*) The RRMs are interacting in the free form such that one binding platform is occluded by the other RRM. The RRMs are also interacting in their bound form, but in a different relative orientation that forms a continuous binding platform. This case occurs in U2AF65 and its bound form is represented on the right panel (pdb code 2YH1) (Mackereth et al. 2011). (**e**) (*Situation #5*) The RRMs are interacting in the free form such that one binding platform is occluded by the other RRM. The RRMs are also interacting in their bound form and only one RRM can bind to nucleic acid. This case occurs in Prp24 and FIR and the bound form of FIR is represented on the right panel (pdb code 2QFJ) (Crichlow et al. 2008).

TABLES

Table 1

NMR experimental restraints and structural statistics

Distance restraints		
Total NOE		5354
Intra-residue		1108
Sequential		1361
Medium range (|i-j| < 5 residues)		1015
Long range (|i-j| ≥ 5 residues)		1806
Interdomain NOE		64
Hydrogen bonds		64
Distance restraints violations (mean ± s.d.)		
Number of NOE violations > 0.2 Å		3.1 ± 1.0
Maximum NOE violation (Å)		0.27 ± 0.04
TALOS+ derived dihedral restraints		
ϕ		168
ψ		168
Dihedral violations (mean ± s.d.)		
Number of dihedral violations > 5°		0.55 ± 0.51
Maximum dihedral violation (°)		3.5 ± 3.3
RDC restraints		
Number of ^{15}N-^{1}H RDC restraints		101
RRM1		56
RRM2		45
RDC violation (mean ± s.d.)		
Number of RDC violations > 2 Hz		6.0 ± 2.3
Maximum RDC violation (Hz)		3.1 ± 0.3
R.m.s.d. from average structure (Å)		
Backbone[a]		0.71 ± 0.16
Heavy atoms[a]		1.22 ± 0.22
	RRM1[b]	RRM2[c]
Backbone	0.40 ± 0.08	0.58 ± 0.18
Heavy atoms	0.98 ± 0.18	1.17 ± 0.30
Deviation from ideal covalent geometry (mean ± s.d.)		
Bond lengths (Å)		0.0041 ± 0.0001
Bond angles (°)		0.54 ± 0.01
Impropers (°)		1.36 ± 0.05
Ramachandran analysis		
Most favored region		90.1 %
Allowed region		9.8 %
Disallowed region		0.1 %
CING Red/Orange/Green scores		
R/O/G (%)		12/31/57

[a] Protein r.m.s.d. was calculated using residues 11-89, 105-111, 117-139, 146-180 for the ensemble of 20 refined structures ; [b] RRM1 r.m.s.d. was calculated using residues 11-89 ; [c] RRM2 r.m.s.d. was calculated using residues 105-111, 117-139, 146-180.

Table 2

^{15}N T1, T2 and overall correlation time of hnRNP A1 RRM1, RRM2 and RRM12

Protein construct	T1 (ms)	T2 (ms)	T1/T2	τ_c (ns)
RRM1	500 ± 20	89 ± 8	5.6 ± 0.5	8.2 ± 0.4
RRM2	495 ± 30	55 ± 7	9.1 ± 1.8	10.9 ± 1.1
UP1	640 ± 30	36 ± 4	17.5 ± 1.2	15.8 ± 0.6

Table 3

Comparison of $UP1_{free}$, $UP1_{bound}$ and $UP1_{NMR}$ structures

Structures	backbone r.m.s.d. (Å)[a]
UP1	
$UP1_{free}/UP1_{bound}$	1.70
$UP1_{free}/UP1_{NMR}$	2.18
$UP1_{bound}/UP1_{NMR}$	1.18
RRM1	
$RRM1_{free}/RRM1_{bound}$	0.32
$RRM1_{free}/RRM1_{NMR}$	0.55
$RRM1_{bound}/RRM1_{NMR}$	0.57
RRM2	
$RRM2_{free}/RRM2_{bound}$	0.45
$RRM2_{free}/RRM2_{NMR}$	0.79
$RRM2_{bound}/RRM2_{NMR}$	0.89

[a] UP1 r.m.s.d. was calculated using residues 11-89 and 105-111, 117-139, 146-180; RRM1 r.m.s.d. was calculated using residues 11-89; RRM2 r.m.s.d. was calculated using residues 105-111, 117-139, 146-180.