Structural basis of G-tract recognition and encaging by hnRNP F quasi RRMs

Cyril Dominguez[1], Jean-François Fisette[2], Benoit Chabot[2], & Frédéric H.-T. Allain[1]*

[1] Institute of Molecular Biology and Biophysics

ETH Zürich

CH-8093 Zürich,

Switzerland

[2] Département de microbiologie et d'infectiologie

Faculté de médecine et des sciences de la santé

Université de Sherbrooke

Sherbrooke (Québec) J1H 5N4,

Canada

* Correspondence:

Email: allain@mol.biol.ethz.ch

Phone: +41 44 633 39 40

Fax: +41 44 633 12 94

Abstract

The heterogeneous nuclear ribonucleoprotein (hnRNP) F is involved in the regulation of mRNA metabolism by specifically recognizing G-tract RNA sequences. We have determined the solution structures of the three quasi RNA recognition motifs (qRRMs) of hnRNP F in complex with G-tract RNA. These structures show that qRRMs bind RNA in a very unusual manner, the G-tract being "encaged", making the qRRM a novel RNA binding domain. We defined a consensus signature sequence for qRRMs and identified other human qRRM-containing proteins, which also specifically recognize G-tract RNAs. Our structures explain how qRRMs can sequester G-tracts maintaining them in a single-stranded conformation. We also show that isolated qRRMs of hnRNP F are sufficient to regulate the alternative splicing of the Bcl-x pre-mRNA strongly suggesting that hnRNP F would act by remodeling RNA secondary and tertiary structures.

Introduction

Alternative splicing is a universal post-transcriptional mechanism in higher eukaryotes that is important for generating proteomic diversity. It is estimated that >90% of human genes undergo alternative splicing[1]. The process is highly regulated[2] and disruption of the splicing regulatory networks can contribute to several diseases[3]. Regulation of alternative splicing involves cis-acting regulatory elements that are bound by trans-acting factors to enhance or silence splicing of alternative exons. A well-characterized cis-acting element is the G-tract that consists of three or more consecutive guanines. G-tracts are statistically overrepresented near splice sites. When located in introns, G-tracts generally act as splicing enhancers[4] while when located in exons, they mainly act as splicing silencers[5]. G-tracts located in introns were recently shown to be able to buffer genetic variations of 5' splice sites by potentially acting as evolutionary capacitors of splicing changes[6]. G-tracts are also abundant downstream of mammalian polyadenylation signals[7] and in 5' and 3' untranslated regions (UTR)[8]. Some of the trans-acting factors that bind to these G-tracts belong to the heterogeneous nuclear ribonucleoprotein (hnRNP) F/H family of proteins that consists of five members (hnRNP F, hnRNP H, hnRNP H', hnRNP 2H9, and G-Rich Sequence Factor (GRSF) 1)[9,10]. GRSF-1 is mainly involved in translation regulation[11,12] and Internal Ribosome Entry Site (IRES) mediated translation[13], while hnRNP H, H' and F are mostly involved in the regulation of alternative splicing[14-23] and polyadenylation[24-26]. The post-transcriptional regulation by hnRNP F/H proteins is a crucial event since mutations in G-tracts often result in aberrant splicing patterns and can be associated with various diseases[27-32]. Furthermore, expression levels of hnRNP F/H proteins are highly regulated during development[33] and differences in expression levels or post-translational modifications of these proteins are observed upon cellular stress[21,34] and cancer[35].

HnRNP F and H regulate the alternative splicing of the apoptotic regulator B-cell lymphoma/leukemia-2 (Bcl-2) related protein Bcl-x[18]. Bcl-x naturally exists in two forms, Bcl-x_L (233 amino acids) and Bcl-x_S (170 amino acids), which result from alternative 5' splice site utilization, producing a truncated exon 2 for Bcl-x_S[36]. The activity of these isoforms is antagonistic; Bcl-x_L is anti-apoptotic, while Bcl-x_S is pro-apoptotic. In a number of cancer cells, Bcl-x_L is overexpressed, which increases the risk of metastases. The ratio of Bcl-x_S and Bcl-x_L is regulated by hnRNP F and H. These proteins recognize three G-tracts downstream of the Bcl-x_S 5' splice site to encourage the production of the Bcl-x_S isoform[18].

The molecular mechanisms by which hnRNP F/H members regulate post-transcriptional events are still unclear. In some cases, hnRNP F/H proteins regulate splicing by preventing or enhancing the binding of spliceosomal components to the pre-mRNA[17,27,37] or by competing with other trans-acting factors[38]. HnRNP F/H members were also proposed to influence splice site selection through self association that would loop out introns and bring specific pairs of splice sites in closer proximity[22]. Additionally, hnRNP F/H members can help the formation of spliceosomal complexes but not the formation of the ATP-independent pre-spliceosomal early E complex[39]. Furthermore, hnRNP F/H members can regulate splicing by binding G-tracts embedded in RNA stem-loop structures and the structural context of the RNA appears crucial for hnRNP F/H function[19,40]. The regulation of alternative splicing by hnRNP F/H members is therefore complex and these proteins might, depending on the RNA substrate, act at different stages of spliceosome assembly through different mechanisms.

HnRNP F/H family members contain two (2H9) or three (H, H', F and GRSF-1) quasi RNA recognition motifs (qRRMs) and one or two glycine-rich auxiliary domains. These proteins are very similar in sequence (hnRNP F and H share 78% sequence identity)[10,41]. The RNA binding

domains of hnRNP F/H proteins were denoted qRRMs because of their similarity to the classical RRM motif. However, two conserved sequences found in all RRMs (RNP1 and RNP2) involved in RNA binding are poorly conserved in hnRNP F/H family members[41]. Previously, we solved the structures on the three individual qRRMs of hnRNP F and showed that the RNA binding surface is different than the one of the classical RRM[42].

To gain structural insights into the mechanism of G-tract recognition by hnRNP F/H family members, we have determined the solution structures of the three qRRMs of hnRNP F in complex with a G-tract RNA using Nuclear Magnetic Resonance (NMR) spectroscopy. The recognition of G-tract RNA by the three qRRMs is very distinct from the RNA recognition mode of classical RRMs. These new intermolecular contacts are supported by site-directed mutagenesis and isothermal titration calorimetry measurements. Based on these results, we defined a consensus RNA binding signature for qRRMs, identified additional human proteins containing qRRMs and showed that they bind G-tract RNAs. Finally, binding studies and splicing assays of these qRRMs with a natural target, the Bcl-x pre-mRNAs, suggest that hnRNP F/H members exert their function by preventing RNA structure formation.

Results

Structure determination of the qRRM-RNA complexes

We previously solved the structures of each qRRM of hnRNP F free in solution[42]. The binding of each qRRM to G-tract RNA was assessed by NMR chemical shift perturbation experiments, and the RNA binding surface was mapped. All resonances, however, could not be observed in the complex and binding of the third qRRM of hnRNP F with RNA was not detected. Different RNA sequences, temperatures and buffer conditions were therefore tested to optimize the NMR spectral quality of the complexes. The best spectral conditions for the complex were obtained in

20 mM phosphate buffer pH 7.0, 50 mM sodium chloride, at 303K for qRRM1 and qRRM2 and 293K for qRRM3. Although we worked initially with the sequence CGGGAU, the RNA sequence AGGGAU gave better spectra in agreement with the recently determined RNA binding consensus sequence for all hnRNP F/H family members[39]. NMR titration experiments indicate that the three qRRMs of hnRNP F bind this RNA sequence at a ratio of 1:1. Complex formation is in intermediate to slow exchange on the NMR time scale and leads to large chemical shift changes in three loops of the qRRMs (Supplementary Fig. 1). Almost complete resonance assignments were obtained for the three qRRMs in complex with AGGGAU. Assignment of the RNA was greatly facilitated by using two chemically synthesized RNAs with ^{13}C-label sugars[43]. In one RNA, the three guanines were labeled while in the second one only the first, third and fifth nucleotides were labeled. This allowed the unambiguous assignment of the central guanine G3 (Supplementary Fig. 1d). Isothermal titration calorimetry (ITC) measurements were performed at 30 degrees using these NMR conditions and confirmed that each qRRM of hnRNP F binds AGGGAU with a dissociation constant ranging from 0.4 μM to 4.6 μM (Fig. 1).

A total of 1476, 2370, and 1210 nuclear Overhauser effects (nOes) (including 56, 63, 60 intermolecular nOes) for qRRM1-AGGGAU, qRRM2-AGGGAU and qRRM3-AGGGAU, respectively, were converted into distance restraints and used for structure determination (Table 1). Most intermolecular nOes were derived from filtered-edited NOESY experiments using ^{15}N-^{13}C labeled proteins and unlabeled RNA (Supplementary Fig. 2a). In addition, at lower temperature (283K), intermolecular nOes between protons of the protein and imino protons of the guanines could be observed in 2D NOESY spectra (Supplementary Fig. 2b). Additionally, 28 (qRRM1 and qRRM3), and 26 (qRRM2) intra-molecular hydrogen-bonds restraints derived from slowly exchanging amide protons in presence of D_2O were used in the structure determination. A

total of 50 conformers were calculated and the 20 presenting the lowest nOe violations were analyzed. The three ensembles of structures are well defined (Table 1 and Fig. 1). The structure of qRRM2-AGGGAU is best defined due to better spectral quality and therefore a higher number of nOes. The structures of qRRM1-AGGGAU and qRRM3-AGGGAU are less well defined but are sufficiently precise to reveal the molecular basis of G-tract recognition by qRRMs.

Overview of the three qRRM-AGGGAU structures

The three qRRM structures of hnRNP F in complex with RNA display the classical compact $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ RRM fold and are similar to those free in solution[42]. The 5'-AGGGAU-3' RNA adopts a single-stranded conformation upon binding to the qRRMs of hnRNP F. In the structures, all sugars puckers adopt a C2' endo conformation and the bases of A1 and G4 adopt a *syn* conformation.

The structures show that the three qRRMs of hnRNP F bind G-tract RNA in a remarkably similar manner, which is unusual for proteins containing multiple RRMs. In all cases, three loops (loops 1, 3 and 5) are involved in the binding while the canonical binding interface of the classical RRM (the residues located in the two central β-strands) does not interact with the RNA (Fig. 1). The orientation of the RNA relative to the qRRM is very similar in the three structures.

The first two nucleotides, A1 and G2, are involved in base stacking with an aromatic residue of the qRRM located in loop 1 (Trp20 in qRRM1, Phe120 in qRRM2, and Tyr296 in qRRM3). A1 is further stabilized by a positively charged residue located in loop 3 (Arg52, Lys150 and Arg326). The third nucleotide, G3, is involved in stacking interaction with another aromatic residue located in loop 5 (Tyr82, Tyr180 and Tyr356). The fourth nucleotide, G4, is stabilized by an arginine residue located in β1 (Arg16, Arg116 and Arg294). The fifth and sixth nucleotides (A5 and U6) are not involved in stacking interaction but are stabilized by a positively charged

residue located in loop 5 (Arg75, Lys173 and Arg349) and an aromatic residue located in the fourth β-strand (Phe86, Phe184 and Phe360). Although the six nucleotides are well defined in the three structures, only the three guanines are specifically recognized by the qRRMs of hnRNP F. This is consistent with ITC measurements showing that the affinity of hnRNP F qRRM2 and qRRM3 to a GGG trinucleotide only slightly decreased as compared to the affinity to AGGGAU RNA (Table 2c) and that mutations affecting residues contacting A1 (K150A), A5 (K173A) and U6 (F184A) have only moderate effects on the affinity to the RNA (Table 2a).

The G-tract is specifically recognized by hnRNP F qRRMs

The three guanines adopt a compact conformation (resembling an arch), surrounded by three conserved residues stacking each guanine base (two aromatics and one arginine) and forming a molecular cage around the G-tract (Fig. 2a). G4 adopts a *syn* conformation positioning its base above the sugar of G3 and contact G2. The atypical conformation of G4 base is consistent with the unusual upfield chemical shifts observed for G3 sugar protons (Supplementary Fig. 1d). Consistently, mutations of the residues forming the molecular cage (qRRM1 W20A, and qRRM2 R116A, F120A, and Y180A) strongly reduce or abolish binding to the G-tract (Table 2a and 2b).

The central guanine, G3, is specifically recognized by backbone and side chain atoms of the qRRMs (Fig. 2b). G3 H1 is in hydrogen-bond with the backbone carbonyl of the stacked tyrosine. In addition, G3 H21 amino proton can also form a hydrogen-bond with this carbonyl group in the qRRM2-AGGGAU structure. The G3 O6 carbonyl group is specifically recognized in all three structures via one or two intermolecular hydrogen-bonds with a conserved arginine residue (Arg81, Arg179 and Arg355). Additionally, in the qRRM1-AGGGAU and qRRM2-AGGGAU structures, the G3 H22 amino proton forms a hydrogen-bond with the side chain carboxylic group of a conserved glutamic acid (Glu84 and Glu182). The hydrogen-bonds of G3

with these two charged residues are crucial for the binding since mutations qRRM1 E84A, qRRM2 R179A and qRRM2 E182A abolish RNA binding (Table 2a and 2b). In qRRM3 the side chain carboxylic group of the conserved glutamic acid (Glu358) is mainly hydrogen-bonded to A5 H62 and not to G3 H22 (Fig. 2b).

The two flanking guanines, G2 and G4 are also specifically recognized by the qRRMs (Fig. 2c). G2 H1 and H21 are hydrogen-bonded to the backbone carbonyl oxygen of a conserved leucine (Leu18, Leu118, and Leu296). G2 O6 is also recognized through hydrogen-bonds to Tyr298 main-chain amide (qRRM3). G4 adopts a *syn* conformation that allows for the formation of an intra-RNA hydrogen-bond between G2 H22 and G4 O6 in the three complexes.

To assess the importance of the G-tract in qRRM binding, we tested the interaction of qRRM2 and qRRM3 with different RNA sequences having each a G to A mutation (Table 2c). These mutations strongly affect the binding to qRRM2 and qRRM3.

Taken together, the structures explains well how each qRRM of hnRNP F specifically recognizes three consecutive guanines in a remarkably similar fashion and that only a G-tract of three consecutive guanines appears necessary and sufficient for efficient RNA binding to each qRRMs of hnRNP F.

A consensus sequence for the qRRM domain

The structures of the three qRRMs of hnRNP F in complex with the G-tract RNA clearly show that qRRMs bind to RNA in a very distinct manner than the classical RRM (Supplementary Fig. 3). Classical RRMs, such as hnRNP A1 RRMs, bind RNA through their β-sheet surface while qRRMs bind RNA through three loops. A comparison of the qRRM sequences of hnRNP F with other members of the hnRNP H family indicates that the three regions contacting the RNA, especially two amino acids sequences are highly conserved in all

members (Fig. 3a and Supplementary Fig. 4). The first conserved sequence consists of residues R-G-L-P-(W/F/Y) located in loop 1. The second sequence is located in loop 5 and corresponds to (R/K)-(5aa)-R-Y-(V/I/L)-E-(V/I/L)-F. Additionally, the positively charged residue located in loop 3 is also conserved among qRRMs of hnRNP F/H family members. These amino acid sequences are strictly conserved in hnRNP F among vertebrates.

Using these consensus sequences characteristic of the qRRM domain, we investigated if other human proteins defined as RRM-containing proteins could instead contain qRRM motifs. For this purpose, we analyzed all human RRM sequences derived from two different domain databases, Pfam (http://pfam.sanger.ac.uk/) and SMART (http://smart.embl-heidelberg.de/), and searched for consensus sequences corresponding to qRRM domains. In total, 25 unique protein sequences contained the qRRM consensus motifs (Fig. 3b and Supplementary Fig. 4). All these putative qRRM domains are highly conserved in the three regions defining the qRRM RNA binding consensus. Interestingly, we did not find any sequence containing only one of the 2 highly conserved motifs. In all cases, the sequences retrieved based on conservation of one motif, also contained the second one, as well as the positively charged residue located in loop 3 of hnRNP F qRRMs.

Analysis of the proteins containing these sequences identified all hnRNP F/H family members (hnRNP F, hnRNP H, hnRNP H', hnRNP 2H9, and GRSF-1). In addition, we also identified other proteins: RNA Binding Motif (RBM) protein 12 and RBM12B (Swiss-Prot entry: Q9NTZ6 and Q8IXT5), RBM19 (Q9Y4C8), RBM35A and RBM35B (Q6NXG1 and Q9H6T0), and two hypothetical proteins identified by the German cDNA consortium and denoted DKFZp667H197 (Q69YJ7) and DKFZp686A15170 (Q68DG4).

Sinilarly to hnRNP F/H members, the RNP1 and RNP2 sequences typical of the classical RRM are not conserved in these newly identified putative qRRMs (Supplementary Fig. 4). Consequently, it is likely that RBM12, RBM19 and RBM35 contain qRRMs that bind to their target RNA in a similar way as the qRRMs of hnRNP F. Among these proteins, the NMR structure of RBM19 RRM2 has been determined (PDB ID: 2dgw) and shows striking similarities with the structure of hnRNP F qRRM2 (Fig. 3c), suggesting that this qRRM could recognize G-tract.

We therefore tested both RBM19 RRM2 and RBM35A RRM3 for binding to G-tract RNA using ITC. Both domains bind to AGGGAU with μM affinity but do not bind to AGAGAU demonstrating that these qRRMs are indeed specific for G-tract RNA (Fig. 3d).

hnRNP F qRRMs prevent RNA structure formation

Our NMR studies using different G-tract RNAs free in solution indicated that G-tracts are often structured. For instance, the RNA used in our structural study, AGGGAU, forms a tetramolecular G-quadruplex in our NMR conditions (Figs. 4a,b). Circular dichroism spectrum of AGGGAU at 20 degrees is characteristic of a parallel stranded G-quadruplex with a positive peak at 265 nm and a negative peak at 245 nm (Fig. 4a). Furthermore, 1D NMR spectrum of AGGGAU in the same conditions displays three major signals around 11 ppm that are typical for exchangeable imino protons (Fig. 4b, black spectrum), further confirming that the free AGGGAU RNA adopts a G-quadruplex structure. Upon qRRM binding, however, these NMR signals are not present indicating that the G-quadruplex structure is not formed when the RNA is bound to qRRMs (Fig. 4b, red spectrum). This is consistent with our structures showing that AGGGAU adopts a single-stranded conformation when bound to the qRRMs of hnRNP F. We confirmed these results by measuring UV spectra of AGGGAU free or in presence of individual

qRRMs of hnRNP F (Fig. 4c). Both qRRM2 and qRRM3 induce an increase in the absorbance at 260 nm further indicating a disruption of the quadruplex structure. We then investigated if the stability of the G-quadruplex could have an effect of qRRM3 binding. Since potassium ions are known to stabilize G-quadruplex structures, we tested the binding of hnRNP F qRRM3 to AGGGAU in potassium buffer (Supplementary Fig. 5). As expected, the presence of potassium ions increases the melting temperature of the G-quadruplex (Supplementary Figs 5a,b). Next, we tested the binding of hnRNP F qRRM3 to the AGGGAU RNA in both buffers by NMR (Supplementary Fig. 5c). In both buffers, the TOCSY spectrum of the free RNA displays two peaks indicating that the RNA exists in two distinct conformation: G-quadruplex and single-stranded. While, in sodium buffer, the addition of qRRM3 leads to the disappearance of both peaks and the appearance of the bound peak, in potassium buffer, the peak corresponding to the G-quadruplex conformation remains unaffected, suggesting that, in potassium buffer, qRRM3 binds uniquely the single-stranded form of the RNA. We confirmed these results by NMR titration experiments in potassium buffer (Supplementary Fig. 5d) showing that saturation of the complex is achieved at a qRRM3:AGGGAU ratio of ~ 0.1:1. This indicates that, in these conditions, qRRM3 binds only the single-stranded form of the RNA that consists of approximately 10% of the total RNA.

Altogether, our results indicate that hnRNP F qRRMs can prevent G-quadruplex structure formation by binding single-stranded RNA and modifying the single-stranded/G-quadruplex equilibrium and that the binding of qRRMs is dependent on the G-quadruplex stability. Consistently, we previously reported that hnRNP F qRRM3 was not involved in RNA binding because we could not observe by NMR a clear binding to a CUGGGGU RNA sequence derived from the Bcl-x RNA[42]. This can be due to the stability of the G-quadruplex formed *in vitro* in our

NMR conditions, since G-quadruplexes formed by 4 G-tetrads (CUGGGGU RNA) are more stable than G-quadruplexes formed by 3 G-tetrads (AGGGAU RNA).

Next, we investigated whether this property would be true for a natural RNA substrate. We showed previously that hnRNP F and H regulate the splicing of Bcl-x by binding to an RNA sequence located downstream of the 5' splice site of Bcl-x_S[18]. This sequence possesses four G-tracts within 60 nucleotides and could adopt a G-quadruplex structure. 1D NMR spectrum of this RNA displays signals corresponding to imino protons confirming that this RNA is structured in solution at 20 degrees (Fig. 4d). However, the ^{1}H-^{15}N HSQC spectrum shows that this RNA does not form a G-quadruplex structure but adopts a stem-loop structure mainly composed of G-C and G-U base pairs due to the presence of guanine and uracil imino signals (nitrogen frequency between 140 and 150 ppm and around 160 ppm, respectively) (Fig. 4e). A secondary structure prediction of this RNA using mfold[44] suggested the formation of a stem-loop structure (Fig. 4f). This stem-loop structure is highly stable (melting temperature of 57.0˚C) (data not shown). We next tested the effect of hnRNP F qRRM3 on this RNA. As shown in Figures 4d and 4e, the progressive addition of hnRNP F qRRM3 to this RNA results in the total disappearance of the imino signals demonstrating that in the presence of hnRNP F qRRM3, the secondary structure of the RNA is not present anymore.

These results clearly indicate that qRRMs are able to prevent RNA structure formation that can be G-quadruplexes or stem-loops. Analysis and structure predictions of the Bcl-x pre-mRNA suggest that the G-tracts recognized by hnRNP F can potentially be embedded in secondary or tertiary structures (Supplementary Fig. 6). Our results suggest that all these putative structures could be altered by hnRNP F and this could be important for the regulation of Bcl-x alternative splicing.

Single qRRMs of hnRNP F regulate the splicing of Bcl-x

If individual qRRMs of hnRNP F can modulate RNA structures, they might also function as regulator of splicing, hence reproducing the activity of the full-length protein. Using a Bcl-x model pre-mRNA (Fig. 5a), we performed *in vitro* splicing assays in HeLa nuclear extracts in the presence of increasing amounts of the individual qRRMs. It was shown previously that the activity of hnRNP F on Bcl-x splicing required the presence of the G-tracts located downstream of the Bcl-x_S 5' splice site[18]. As shown in Figures 5b, 5d, and 5e, the addition of hnRNP F, qRRM1, or qRRM3 but not qRRM2 modified the Bcl-x_L/Bcl-x_S ratio towards the Bcl-x_S isoform. Interestingly, qRRM1, and qRRM3 have a higher affinity for G-tract RNA than qRRM2 (Fig. 1). In addition, mutants qRRM1 W20A and qRRM1 E84A that do not bind G-tract RNA (Table 2a) had no effect on Bcl-x splicing (Fig. 5c). These results indicate that a single qRRM of hnRNP F is sufficient to impose regulation of Bcl-x splicing and that G-tract binding is necessary for this regulation. Since RBM19 RRM2 also bind G-tract RNAs, we tested its effect on Bcl-x splicing and similar to the qRRMs of hnRNP F, RBM19 RRM2 is able to modify the Bcl-x_L/Bcl-x_S ratio towards the Bcl-x_S isoform (data not shown). The ability of single qRRMs to prevent RNA structure formation suggests that hnRNP F may change the conformation of the RNA in the vicinity of the 5' splice site of Bcl-x_S to facilitate its recognition by the splicing machinery. Moreover, splicing regulation by hnRNP F qRRMs is dependent on the presence of G-tracts since qRRM3 has no effect on Bcl-x splicing when the third G-tract downstream of the Bcl-x_s 5'splice site is mutated (Fig. 5f).

Since a qRRM in isolation has the capacity to regulate the splicing of Bcl-x similarly to the full-length protein, one might ask why three qRRMs have been conserved in the hnRNP F/H protein family? Having multiple qRRM covalently linked in a single chain presents at least two

advantages compared to a single RRM: first, the protein with multiple domains has a higher probability to bind G-tracts than a single domain and second, since most pre-mRNAs targeted by hnRNP F and H contain multiple G-tracts, the multidomain protein will bind with higher affinity since the local concentration of qRRM is increased. This may explain why isolated qRRMs were less active than the full-length protein in our in vitro splicing assays (Fig. 5b).

Discussion

The qRRM is a novel RNA binding domain

The structures of the three qRRMs of hnRNP F in complex with G-tract RNA described here define a novel mechanism for RNA recognition. qRRMs specifically interact with the RNA through three highly conserved loops and not through the β-sheet surface as observed for the classical RRM (Figs. 1 and 2). It has previously been reported that in addition to the β-sheet surface, RRMs can utilize loop residues to bind RNA[45]. In all these cases, however, the RNA is primarily bound by the β-sheet surface and the involvement of loop residues contributes to improve the affinity and/or the specificity towards the RNA. Therefore, the structures of hnRNP F qRRMs in complex with G-tract RNA provide the first examples of protein domains containing a RRM fold that bind RNA through three loops and in which the canonical RNA binding surface is not involved. The different mechanism of RNA binding between qRRM and classical RRM is depicted in Supplementary Figure 3. Both hnRNP F qRRMs and hnRNP A1 RRMs recognize G-rich sequences but the RNA recognition mechanism is very different. The two RRMs of hnRNP A1 bind RNA using the canonical RNP 1 and RNP 2 sequences while hnRNP F qRRMs bind RNA using three loops. The directionality of the RNA is inverted in the two complexes. The mode of guanine recognition is also very different. In hnRNP A1, only the

first guanine is specifically recognized by the RRM and the three guanines are stacked together[46]. In hnRNP F, each of the three guanines is specifically recognized by the qRRM; the three guanines adopt a compact conformation that is stabilized by stacking interaction with qRRM residues that form a molecular cage (Fig. 2).

Based on the amino acid sequences and the structures of hnRNP F qRRMs in complex with a G-tract RNA, we identified five human proteins (RBM12, RBM12B, RBM19, RBM35A and RBM35B) containing a putative qRRM domain (Fig. 3 and Supplementary Fig. 4). RBM12 is a human protein with unknown functions. RBM35 A and B, recently renamed epithelial splicing regulatory protein (ESRP) 1 and 2, are cell-type specific splicing factors[47]. ESRP1 and ESRP2 both contain three RNA binding domains highly similar to hnRNP F qRRMs. RBM19 is an RNA binding protein that is homologous to the yeast Multiple RNA-binding domain-containing protein 1 (Mrdp-1) protein, involved in ribosomal RNA biogenesis[48] through binding to the U3 Small nucleolar RNAs (snoRNP) and the 35S pre-ribosomal RNA.

Our data show that both RBM19 RRM2 and RBM35 RRM3 bind specifically G-tract RNA with affinities similar to hnRNP F qRRMs (Fig. 3d). Furthermore, RBM19 RRM2 can substitute to hnRNP F qRRM1 and qRRM3 in regulating the splicing of Bcl-x (data not shown). These results demonstrate that additional qRRMs can be found outside of the hnRNP F/H family and that they form a subclass of RRMs, which are specific for binding G-tracts. Accordingly, the RNA sequences targeted by RBM19, namely U3 snoRNP and the 35S pre-ribosomal RNA, both contain G-tracts.

Mechanism of hnRNP F/H function in splicing regulation

Why would nature create a new type of RRM for binding G-tract RNA? Guanines are unique in the four RNA bases in that they can stably base-pair with any of the four bases (with at least

two hydrogen-bonds being formed). The qRRM clearly seems to be adapted to neutralize this RNA folding by literally sequestering the G-tract within a protein cage (Fig. 2). The structures of hnRNP A1 and hnRNP D in complex with TAGGG DNA showed that these proteins contain classical RRMs, bind DNA sequences containing G-tracts using the canonical β-sheet surface and prevent G-quadruplex formation[46,49]. However, both hnRNP A1 and hnRNP D specifically recognize a TAG sequence but not a G-tract, leaving two guanines of the G-tract accessible. Therefore, only hnRNP F/H family members are specific for G-tract recognition. This difference in specificity could explain why hnRNP F/H family members contain qRRMs instead of classical RRMs.

The molecular mechanisms by which hnRNP F/H members regulate post-transcriptional gene events are still unclear. Here, we show that a single qRRM of hnRNP F (qRRM1 or qRRM3) or of RBM19 can on its own regulate the splicing of Bcl-x (Fig. 5 and data not shown). To our knowledge, this is the first demonstration of a single RNA binding domain being able to recapitulate the effect of a full-length trans-acting factor in splicing regulation. These results strongly suggest that hnRNP F can act on the pre-mRNA via two possible mechanisms. One mechanism could be that hnRNP F binding prevents the binding of an antagonist splicing factor. The region just upstream of the hnRNP F binding site contains a CUCUUCC sequence (Supplementary Fig. 6b) that could be recognized by the polypyrimidine tract binding protein (PTB). However, a siRNA-mediated reduction in PTB did not affect Bcl-x splicing (data not shown). Although we have not identified additional cis-acting elements surrounding these hnRNP F binding sites on the Bcl-x pre-mRNA, an interference with the binding of other splicing factors cannot be excluded. Based on our findings that qRRMs are able to prevent RNA structure formation (G-quadruplexes or stem-loop), it is tempting to

propose a second mechanism in which the qRRMs would act by disrupting RNA structures present in the pre-mRNA that sequester cis-acting elements. In binding RNA, qRRMs would therefore favor the binding of spliceosomal components or trans-acting factors. There are now growing evidence that RNA secondary structures play an important role in splicing regulation and that the structural context of RNA splicing motifs is part of the "mRNA splicing code"[50,51]. Interestingly, it has been previously reported in at least three occasions that hnRNP F or H regulate splicing through the recognition of G-tracts embedded in stem-loop structures. The RNA surrounding both human immunodeficiency virus (HIV) 1 exon 2 and exon 6D form specific stem-loop structures encompassing the 3' splice site and a G-tract specifically recognized by hnRNP F and H[16,52]. Similarly, the pre-mRNA of the Harvey rat sarcoma viral oncogene homolog (H-ras) contains nine G-tracts located both upstream and downstream of the 5' splice site of the alternative intron D exon (IDX) that are bound by hnRNP F/H and could form a stem-loop structure encompassing the splice site[19]. Here, we show that a portion of the Bcl-x pre-mRNA containing four G-tracts can form a stem-loop structure in vitro (Figs. 4d-f) and that in presence of hnRNP F qRRM3, this RNA sequence becomes unstructured. Our results would support the view that hnRNP F/H can bind to all of the above structured RNAs leading to the opening of the stem-loop structures. In addition to stem-loops, many pre-mRNAs that are regulated by hnRNP F/H, contain multiple consecutive G-tracts that could potentially fold into G-quadruplex[53]. Although the implication of such structures in splicing regulation has not yet been demonstrated, G-quadruplexes have been involved in other important biological processes, such as telomere stabilization, transcription and translation regulation[54]. Our data show that, similarly to their effect on stem-loop RNA structures, hnRNP F qRRMs prevent the

formation of G-quadruplex structures (Figs. 4a-c). In this case, the binding of hnRNP F qRRMs is dependent on the G-quadruplex stability (Supplementary Fig. 5).

The Bcl-x pre-mRNA sequence encompassing the Bcl-x_S and the Bcl-x_L splice sites possesses numerous G-tracts (Supplementary Fig. 6a,b). Three G-tracts immediately downstream of the Bcl-x_S splice site are necessary for hnRNP F binding and splicing regulation (Supplementary Fig. 6b, red)[18] but 6 additional G-tracts are also present (Supplementary Fig. 6b, blue). Interestingly, the Bcl-x_L 5' splice site also contains a G-tract. Using the QGRS mapper software[53], we identified four G-quadruplexes that can potentially form in the Bcl-x pre-mRNA (Supplementary Fig. 6c). In addition, using Mfold[44], we can predict that this pre-mRNA portion can adopt a stable secondary structure (Supplementary Fig. 6d). This potential structure and the associated base-pairs involving the G-tracts are however different than those predicted when considering shorter portions of the Bcl-x pre-mRNA (Fig. 4d and Supplementary Fig. 6e,f,g). Therefore, one cannot easily identify and test which cis-acting element may be sequestered in the Bcl-x pre-mRNA prior to hnRNP F binding. Future experiments would need to be performed to clarify this issue.

Finally, based on their ability to sequester G-tracts on the RNA, hnRNP F/H proteins may also help to stabilize the genome. G-rich sequences and particularly G-tracts elicit the formation of RNA:DNA hybrids on transcribed DNA[55]. The resulting R-loops can block transcription elongation as well as promote mutations and recombination[56]. Overall, the ability of the qRRMs of hnRNP F/H proteins to prevent RNA structure formation and sequester G-tracts may therefore be a key feature that plays a role in a variety of RNA-related processes from transcription to splicing and translation regulation.

Acknowledgements

We would like to thank D. Black for providing the clone of full-length hnRNP F and S. Pitsch for providing sugar-^{13}C-labeled RNAs. We are grateful to F. Oberstrass for his implication in the optimization of the conditions for complex formation, J. Boudet for his help in ITC measurements, L. Shkreta for testing qRRM2 of RBM19 in splicing, S. Jayne and D. Theler for critical reading of the manuscript, and members of the group for helpful discussions. C.D. was supported by postdoctoral fellowships from the Roche Research Foundation for Biology and from the Novartis Research Foundation. Financial support from the Swiss National Science Foundation, the Structural Biology National Center of Competence in Research and the European Alternative Splicing Network (EURASNET) is also acknowledged.

Figure Legends

Figure 1: Overview of the structures and ITC of human hnRNP F qRRM1-AGGGAU, qRRM2-AGGGAU, and qRRM3-AGGGAU.

Left: Isothermal titration calorimetry of hnRNP F qRRMs binding to AGGGAU. Raw calorimetric outputs are shown on top and binding isotherms describing formation of complex between hnRNP F qRRMs and AGGGAU are shown at the bottom. Concentrations of qRRMs and RNA as well as calculated dissociation constants are indicated. Middle: Overlay of the 20 final structures and right: ribbon representation of the lowest energy structure of qRRM1-AGGGAU, qRRM2-AGGGAU, and qRRM3-AGGGAU. Figures were generated with MOLMOL[57].

Figure 2: Specific recognition of G-tract RNA by hnRNP F qRRMs.

(**a**) Close-up view displaying the molecular cage surrounding the G-tract. (**b**) Close-up view displaying the specific recognition of G3 by qRRM1, qRRM2 and qRRM3. (**c**) Close-up view displaying the specific contacts between G2, G4 and backbone atoms of qRRM1, qRRM2 and qRRM3.

Figure 3: A consensus qRRM motif is present in human proteins that are not hnRNP F family members

(**a**) Consensus RNA binding sequence of qRRMs (h = L, V, I or A ; a = F, Y, W or H ; X= any amino acid). (**b**) Sequence alignment of hnRNP F and H qRRMs, and the newly defined qRRMs displaying the three regions of hnRNP F that bind RNA. (**c**) Comparison of the structures of hnRNP F qRRM2 and RBM19 RRM2. Amino acids of hnRNP F qRRM2 involved in G-tract

binding and residues of RBM19 RRM2 located at the corresponding position are displayed. (**d**) Isothermal titration calorimetry of RBM19 RRM2 (left) and RBM35 RRM3 (right) binding to AGGGAU (Black) and AGAGAU (Red).

Figure 4: hnRNP F qRRMs prevent RNA structure formation

(**a**) Circular dichroism spectrum of 25 mM AGGGAU at 20 degrees. (**b**) NMR spectra of the imino region of AGGGAU in absence (black) or presence (red) of hnRNP F qRRM2. (**c**) UV measurement of 20 mM AGGGAU in absence or in presence of one equivalent of qRRM2 or qRRM3 at 20 degrees and of free AGGGAU at 95 degrees. (**d**) 1D and (**e**) (^{15}N-^{1}H)-HSQC spectra displaying the imino region of a portion of the Bcl-x pre-mRNA free, or in presence of increasing amounts of hnRNP F qRRM3. (**f**) Secondary structure prediction of a fragment of Bcl-x pre-mRNA located downstream of the Bcl-x_s splice site. The four G-tracts are colored red.

Figure 5: Activity of the qRRMs of hnRNP F on splice site selection.

(**a**) Diagram of the Bcl-x model pre-mRNA. The position of the competing 5' splice sites and of the primers used for the amplification of the alternative splice products is indicated. (**b**) *In vitro* splicing assays in HeLa nuclear extracts. hnRNP F full length or hnRNP F qRRM1 was added to splicing mixtures at concentrations of 0.5, 1 and 4 μM (lanes 2-7). A control mixture lacking recombinant protein was included (lane 1). The histogram below the gel expresses the percentage of Bcl-xS usage with standard deviations and was compiled from three different experiments. (**c**) qRRM1 mutants (E84A and W20A) that prevent RNA binding were tested. (**d**) and (**e**) Histograms representing the percentage of Bcl-xS usage when increasing amounts of qRRM1, qRRM3 and qRRM2 were used. hnRNP F was used as control in panel (**e**). (**f**) Splicing assays

with the normal (S2.13) and a mutated (S-mut1) pre-mRNA were performed in the presence of qRRM3 at a final concentration of 1 (lane 5), 4 (lanes 2 and 6) and 8 μM (lane 3). S-mut1 contains a GGGG to GCCG mutation in the third G run of the element displayed in Figure 4d that is bound by hnRNP F[18]. (S-mut1 contains a GGGG to GCCG mutation in the last G run of an element bound by hnRNP F[18]). The percentage of Bcl-x_S usage for each lane is 12% (lane 1), 24% (lane 2), 22% (lane 3), 12% (lane 4), 11% (lane 5) and 11% (lane 6).

Table 1: Structural statistics of the hnRNP F qRRM1-AGGGAU, qRRM2-AGGGAU, and qRRM3-AGGGAU complexes.

	qRRM1-AGGGAU	qRRM2-AGGGAU	qRRM3-AGGGAU
NMR distance and dihedral constraints:			
Total	1476	2370	1210
Intermolecular	56	63	60
RNA Intramolecular	80	101	140
Protein Intramolecular	1312	2180	982
Intra-residues	364	410	275
Sequential	331	572	309
Medium range	226	349	142
Long range	391	849	256
Hydrogen bonds (a)	28	26	28
Dihedral angles (b)	6	6	6
Structure Statistics (20 Structures):			
Violations:			
Average number of violations (>0.4Å)	9.8 ± 2.8	13.3 ± 4.4	25.2 ± 2.8
Maximum average violation (Å)	0.56	0.51	0.66
Deviation from ideal covalent geometry:			
Bond length (Å)	0.0041	0.0041	0.0044
Bond Angle(º)	1.82	1.59	1.88
Average r.m.s. deviations to mean structure (Å):			
Protein Backbone (c)	0.83 ± 0.23	0.36 ± 0.06	1.14 ± 0.25
Protein heavy atoms (c)	1.36 ± 0.29	0.72 ± 0.06	1.49 ± 0.22
RNA heavy atoms (d)	1.90 ± 0.47	1.30 ± 0.26	0.75 ± 0.20
Protein + RNA heavy atoms	1.54 ± 0.25	0.90 ± 0.09	1.42 ± 0.20

(a): Hydrogen bonds were defined for the proteins only

(b): dihedral angle restraints were defined for the RNA backbone only

(c): Protein RMSD was calculated using residues 11-98 (qRRM1), 110-194 (qRRM2) and 289-362 (qRRM3) for the ensemble of 20 refined structures

(d): RNA RMSD was calculated using residues 1-6 for the ensemble of 20 refined structures

Table 2: Binding affinities of hnRNP F qRRM1, qRRM2 and mutants to AGGGAU and AGGGAU mutants RNAs.

a) Binding affinity of qRRM2 mutants to AGGGAU RNA

Protein	Kd (M)	Affinity decrease
qRRM2 Wild Type	4.6×10^{-6}	1
qRRM2 R116A	$>500 \times 10^{-6}$	> 100
qRRM2 F120A	$>500 \times 10^{-6}$	> 100
qRRM2 K150A	5.4×10^{-6}	1.2
qRRM2 K173A	11.2×10^{-6}	2.5
qRRM2 R179A	$>500 \times 10^{-6}$	> 100
qRRM2 Y180A	55.8×10^{-6}	12.1
qRRM2 E182A	$>500 \times 10^{-6}$	> 100
qRRM2 F184A	22.6×10^{-6}	4.9

b) Binding affinity of qRRM1 mutants to AGGGAU RNA

Protein	Kd (M)	Affinity decrease
qRRM1 Wild Type	4.2×10^{-7}	1
qRRM1 W20A	$>500 \times 10^{-6}$	> 1000
qRRM1 E84A	$>500 \times 10^{-6}$	> 1000

c) Binding affinity of RNA mutants to qRRM2 and qRRM3

RNA	Binding to qRRM2		Binding to qRRM3	
	Kd (M)	Affinity decrease	Kd (M)	Affinity decrease
AGGGAU	4.6×10^{-6}	1	4.2×10^{-7}	1
GGG	12.5×10^{-6}	2.7	7.8×10^{-7}	2
AAGGAU	440×10^{-6}	95	17.0×10^{-6}	40
AGAGAU	$>500 \times 10^{-6}$	> 100	$>500 \times 10^{-6}$	> 1000
AGGAAU	380×10^{-6}	83	18.7×10^{-6}	45

Material and Methods

Cloning, expression, and purification

qRRM1 (residues 1-102), qRRM2 (residues 103-194), and qRRM3 (residues 277-381) of human hnRNP F (Swiss-Prot entry P52597) were expressed and purified as previously described[42]. cDNAs containing full length RBM19 and RBM35 were purchased from ImaGene. RBM19 RRM2 (residues 218-309) and RBM35 RRM3 (residues 435-535) were subcloned from the cDNA into the Xho1/BamH1 site of the Pet28a(+) vector containing an N-terminal hexa-histidine tag. Expression and purification of RBM19 RRM2 and RBM35 RRM3 were performed using the same protocol as for hnRNP F qRRMs.

Mutagenesis experiments were carried out using the Quickchange Kit (Stratagene) following manufacturer's instructions. All mutant proteins were checked by running (^{15}N-^{1}H)-HSQC spectra to confirm that they were properly folded (data not shown).

AGGGAU oligonucleotides were purchased from Dharmacon Research, deprotected according to manufacturer's instructions, desalted using a G-15 size exclusion column (Amersham), lyophilized and resuspended in NMR buffer. Bcl-x RNA containing 4 G-tracts was produced by *in vitro* run-off transcription with T7 polymerase and purified by anion-exchange high-pressure liquid chromatography under denaturing conditions.

NMR measurements

All NMR measurements were performed in a buffer containing 20mM NaH_2PO_4, 50mM NaCl, 10mM β-mercaptoethanol, pH7 at 303K (qRRM1 and qRRM2) or 293K (qRRM3) using Bruker AVIII-500 MHz equipped with a cryoprobe, AVIII-600 MHz, AVIII-700 MHz equipped with a

cryoprobe and Avance-900 MHz spectrometers. Data were processed using XWINNMR (Bruker) and analyzed with Sparky (http://www.cgl.ucsf.edu/home/sparky/).

Protein sequence-specific backbone and side chain assignments were achieved using 2D (^{15}N-^{1}H)-HSQC, 2D (^{13}C-^{1}H)-HSQC, 3D HNCA, 3D HNCACB, 3D H(C)CH-TOCSY, 3D NOESY-(^{15}N-^{1}H)-HSQC, and 3D NOESY-(^{13}C-^{1}H)-HSQC taking advantage of the backbone and side chains assignments of the free qRRMs. Aromatic proton assignments were performed using 2D-(^{1}H-^{1}H)-TOCSY and 2D-(^{1}H-^{1}H)-NOESY in 100% D_2O.

G-tract RNA resonance assignments in complex with different qRRMs were performed using 2D-(^{1}H-^{1}H)-TOCSY, 2D-(^{1}H-^{1}H)-NOESY and 2D-(^{1}H-^{1}H) double-half filtered NOESY[58] in 100% D_2O for unlabeled RNA, and 3D NOESY-(^{13}C-^{1}H)-HSQC for sugar labeled RNAs.

Intermolecular nOes were obtained using 2D-(^{1}H-^{1}H), F1-^{13}C-filtered, F2-^{13}C-edited NOESY[58] and 2D-(^{1}H-^{1}H)-NOESY in 100% D_2O using ^{15}N-^{13}C labeled qRRMs and unlabeled RNAs. Intermolecular nOes between imino protons of RNA guanines and qRRM protons were obtained using 2D-(^{1}H-^{1}H)-NOESY recorded at 283 and 288K.

All NOESY spectra were recorded with a mixing time of 150 ms, the 3D TOCSY spectra with a mixing time of 23 ms, and the 2D TOCSY with a mixing time of 50 ms.

Structure calculation and refinement

The AtnosCandid software[59,60] was used to generate preliminary structures and a list of automatically assigned nOe distance constraints for each qRRM of hnRNP F in complex with AGGGAU. For each qRRM, peak picking and nOe assignments were performed using 3D NOESY (^{15}N- and ^{13}C- edited) spectra and 2D homonuclear NOESY spectra recorded in D_2O. Additionally, H-bond constraints were added based on hydrogen-deuterium exchange experiments on the amide protons. Seven iterations were performed and 100 independent

structures were calculated at each iteration steps. Structures of the protein-RNA complexes were calculated using CYANA[59] by adding the manually assigned intramolecular RNA and intermolecular distance restraints. For each cyana run, 50 independant structures were calculated. These 50 structures were refined using the SANDER module of AMBER 7.0[61] using the simulated annealing protocol described previously[42]. The 20 best structures based on energy and nOe violations were analyzed using PROCHECK. The ramachandran plot of the qRRM1-AGGGAU, qRRM2-AGGGAU, and qRRM3-AGGGAU complexes indicates that 65.2%, 70.5%, and 70.5% of all residues are in the most favored region, 32.9%, 28.3%, and 28.3% in the allowed region, and 1.9%, 1.2% and 1.2% in the disallowed region.

Isothermal Titration Calorimetry

ITC experiments were performed on a VP-ITC instrument (MicroCal). The calorimeter was calibrated according to the manufacturer's instructions. Samples of protein and RNA were prepared in and thoroughly dialyzed against the NMR buffer. Concentrations of protein and RNA were determined using optical density absorbance at 280 and 260 nm, respectively. The sample cell (1.4 ml) was loaded with 5–10 μM RNA; qRRMs and mutants concentration in the syringe were between 150 and 500 μM. Titration experiments were done at 30°C and consisted of 50 injections, each of 5 μl volume with a 5-min interval between additions. Stirring rate was 307 r.p.m. Raw data were integrated, corrected for nonspecific heats, normalized for the molar concentration and analyzed according to a 1:1 binding model.

Splicing assays

Plasmid S2.13 was cut with XbaI, transcribed and the pre-mRNA was purified as described[18]. One fmole of transcript was incubated at 40°C for 10 min in the presence of wild type or mutated

proteins. HeLa nuclear extract[62] was then added to the mixture under standard splicing conditions[63] and incubated for 2 h at 30°C. Reverse transcription was carried out using the bc2b primer (CGCTCTAGAACTAGTGGATC) and Omniscript (Qiagen). PCR reactions were then performed in the presence of α^{32}P-CTP using Taq DNA polymerase (NEB) and primers bc2b and BX3 (ATGGCAGCAGTAAAGCAAGCG). Amplification products were separated on 4% acrylamide/bisacrylamide (29:1) native gel and quantitated on a PhosphorImager (Molecular Dynamics).

Recombinant His-tagged hnRNP F was produced using baculovirus expression system (BD Biosciences). Wild type and mutated recombinant proteins were dialyzed against buffer D (60 mM HEPES (pH 7.9), 100 mM KCl, 0.2 mM EDTA, 0.5 mM phenylmethylsulphonyl floride, 0.5 mM dithiothreitol and 20% glycerol) prior to incubation in splicing mixtures.

Circular Dichroism

CD measurements were performed on a Jasco J-717 spectropolarimeter using a 0.1 cm quartz cuvette. CD spectra were recorded at a scanning speed of 10 nm/min, recorded every 0.2 nm and accumulated ten times.

UV measurements

UV measurements were performed using a Varian–Cary UV–visible spectrophotometer equipped with a Peltier temperature control device. Standard 1 cm path length quartz cuvettes were used for the measurements.

Accession codes

Chemical shifts of qRRM1-AGGGAU, qRRM2-AGGGAU, and qRRM3-AGGGAU have been deposited to the BioMagResBank under accession numbers 16192, 16193, 16194, respectively.

Coordinates of the three qRRM-AGGGAU structures have been deposited to the Protein Data Bank under accession numbers 2kfy, 2kg0 and 2kg1.

References

1. Wang, E.T. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-6 (2008).
2. Black, D.L. Mechanisms of alternative pre-messenger RNA splicing. *Ann. Rev. Biochem.* **72**, 291-336 (2003).
3. Cooper, T.A., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136**, 777-93 (2009).
4. Yeo, G., Hoon, S., Venkatesh, B. & Burge, C.B. Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc. Natl Acad. Sci. U S A* **101**, 15700-5 (2004).
5. Wang, Z. et al. Systematic identification and analysis of exonic splicing silencers. *Cell* **119**, 831-45 (2004).
6. Xiao, X. et al. Splice site strength-dependent activity and genetic buffering by poly-G runs. *Nat. Struct. Mol. Biol.* **16**, 1094-100 (2009).
7. Zarudnaya, M.I., Kolomiets, I.M., Potyahaylo, A.L. & Hovorun, D.M. Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res.* **31**, 1375-86 (2003).
8. Huppert, J.L., Bugaut, A., Kumari, S. & Balasubramanian, S. G-quadruplexes: the beginning and end of UTRs. *Nuc. Ac. Res.* **36**, 6260-8 (2008).

9. Caputi, M. & Zahler, A.M. Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J. Biol. Chem.* **276**, 43850-9 (2001).
10. Matunis, M.J., Xing, J. & Dreyfuss, G. The hnRNP F protein: unique primary structure, nucleic acid-binding properties, and subcellular localization. *Nuc. Ac. Res.* **22**, 1059-67 (1994).
11. Park, Y.W., Wilusz, J. & Katze, M.G. Regulation of eukaryotic protein synthesis: Selective influenza viral mRNA translation is mediated by the cellular RNA-binding protein GRSF-1. *Proc. Natl Acad. Sci. USA* **96**, 6694-6699 (1999).
12. Ufer, C. et al. Translational regulation of glutathione peroxidase 4 expression through guanine-rich sequence-binding factor 1 is essential for embryonic brain development. *Genes Dev.* **22**, 1838-1850 (2008).
13. Cobbold, L.C. et al. Identification of internal ribosome entry segment (IRES)-trans-acting factors for the Myc family of IRESs. *Mol. Cell. Biol.* **28**, 40-49 (2008).
14. Min, H.S., Chan, R.C. & Black, D.L. The Generally Expressed Hnrnp-F Is Involved in a Neural-Specific Pre-Messenger-Rna Splicing Event. *Genes Dev.* **9**, 2659-2671 (1995).
15. Chen, C.D., Kobayashi, R. & Helfman, D.M. Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. *Genes Dev.* **13**, 593-606 (1999).
16. Jacquenet, S. et al. A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. *J. Biol. Chem.* **276**, 40464-75 (2001).

17. Caputi, M. & Zahler, A.M. SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. *Embo J.* **21**, 845-55 (2002).

18. Garneau, D., Revil, T., Fisette, J.F. & Chabot, B. hnRNP F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J. Biol. Chem.* **280**, 22641-50 (2005).

19. Camats, M., Guil, S., Kokolo, M. & Bach-Elias, M. P68 RNA helicase (DDX5) alters activity of cis- and trans-acting factors of the alternative splicing of H-Ras. *PLoS ONE* **3**, e2926 (2008).

20. Coles, J.L., Hallegger, M. & Smith, C.W.J. A nonsense exon in the Tpm1 gene is silenced by hnRNP H and F. *RNA* **15**, 33-43 (2009).

21. Hai, Y. et al. A G-tract element in apoptotic agents-induced alternative splicing. *Nuc. Ac. Res.* **36**, 3320-3331 (2008).

22. Martinez-Contreras, R. et al. Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PloS Biol.* **4**, 172-185 (2006).

23. Mauger, D.M., Lin, C. & Garcia-Blanco, M.A. HnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 Exon IIIc. *Mol. Cell. Biol.* **28**, 5403-5419 (2008).

24. Qian, Z.W. & Wilusz, J. An RNA-binding protein specifically interacts with a functionally important domain of the downstream element of the simian virus 40 late polyadenylation signal. *Mol. Cell. Biol.* **11**, 5312-20 (1991).

25. Oberg, D., Fay, J., Lambkin, H. & Schwartz, S. A downstream polyadenylation element in human papillomavirus type 16 L2 encodes multiple GGG motifs and interacts with hnRNP H. *J. Virol.* **79**, 9254-9269 (2005).

26. Dalziel, M., Nunes, N.M. & Furger, A. Two G-rich regulatory elements located adjacent to and 440 nucleotides downstream of the core poly(A) site of the intronless melanocortin receptor 1 gene are critical for efficient 3 ' end processing. *Mol. Cell. Biol.* **27**, 1568-1580 (2007).

27. Buratti, E. et al. hnRNP H binding at the 5' splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHbeta genes. *Nuc. Ac. Res.* **32**, 4224-36 (2004).

28. Cogan, J.D. et al. A novel mechanism of aberrant pre-mRNA splicing in humans. *Hum. Mol. Genet.* **6**, 909-12 (1997).

29. Lew, J.M. et al. CDKN1C mutation in Wiedemann-Beckwith syndrome patients reduces RNA splicing efficiency and identifies a splicing enhancer. *Am. J. Med. Genet. A* **127A**, 268-76 (2004).

30. Masuda, A. et al. hnRNP H enhances skipping of a nonfunctional exon P3A in CHRNA1 and a mutation disrupting its binding causes congenital myasthenic syndrome. *Hum. Mol. Genet.* **17**, 4022-35 (2008).

31. Pagani, F., Buratti, E., Stuani, C. & Baralle, F.E. Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. *J. Biol. Chem.* **278**, 26580-8 (2003).

32. Van Laer, L. et al. Nonsyndromic hearing impairment is associated with a mutation in DFNA5. *Nat. Genet.* **20**, 194-7 (1998).

33. Barberan-Soler, S. & Zahler, A.M. Alternative splicing regulation during C. elegans development: splicing factors as regulated targets. *PLoS Genet.* **4**, e1000001 (2008).

34. Chang, L.Y., Ali, A.R., Hassan, S.S. & AbuBakar, S. Human neuronal cell protein responses to Nipah virus infection. *Virol. J.* **4**, 54 (2007).

35. Honoré, B., Baandrup, U. & Vorum, H. Heterogeneous nuclear ribonucleoproteins F and H/H' show differential expression in normal and selected cancer tissues. *Exp. Cell. Res.* **294**, 199-209 (2004).

36. Boise, L.H. et al. bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. *Cell* **74**, 597-608 (1993).

37. Han, K., Yeo, G., An, P., Burge, C.B. & Grabowski, P.J. A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS Biol.* **3**, e158 (2005).

38. Crawford, J.B. & Patton, J.G. Activation of alpha-tropomyosin exon 2 is regulated by the SR protein 9G8 and heterogeneous nuclear ribonucleoproteins H and F. *Mol. Cell. Biol.* **26**, 8791-802 (2006).

39. Schaub, M.C., Lopez, S.R. & Caputi, M. Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. *J. Biol. Chem.* **282**, 13617-13626 (2007).

40. Jablonski, J.A., Buratti, E., Stuani, C. & Caputi, M. The secondary structure of the human immunodeficiency virus type 1 transcript modulates viral splicing and infectivity. *J. Virol.* **82**, 8038-8050 (2008).

41. Honoré, B. et al. Heterogeneous nuclear ribonucleoproteins H, H', and F are members of a ubiquitously expressed subfamily of related but distinct proteins encoded by genes mapping to different chromosomes. *J. Biol. Chem.* **270**, 28780-9 (1995).

42. Dominguez, C. & Allain, F.H. 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. *Nuc. Ac. Res.* **34**, 3634-45 (2006).

43. Wenter, P., Reymond, L., Auweter, S.D., Allain, F.H. & Pitsch, S. Short, synthetic and selectively 13C-labeled RNA sequences for the NMR structure determination of protein-RNA complexes. *Nuc. Ac. Res.* **34**, e79 (2006).

44. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nuc. Ac. Res.* **31**, 3406-15 (2003).

45. Clery, A., Blatter, M. & Allain, F.H.T. RNA recognition motifs: boring? Not quite. *Cur. Opin. Struct. Biol.* **18**, 290-298 (2008).

46. Ding, J.Z. et al. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* **13**, 1102-1115 (1999).

47. Warzecha, C.C., Sato, T.K., Nabet, B., Hogenesch, J.B. & Carstens, R.P. ESRP1 and ESRP2 Are Epithelial Cell-Type-Specific Regulators of FGFR2 Splicing. *Mol. Cell* **33**, 591-601 (2009).

48. Jin, S.B. et al. Mrd1p is required for processing of pre-rRNA and for maintenance of steady-state levels of 40 S ribosomal subunits in yeast. *J. Biol. Chem.* **277**, 18431-9 (2002).

49. Enokizono, Y. et al. Structure of hnRNP D complexed with single-stranded telomere DNA and unfolding of the quadruplex by heterogeneous nuclear ribonucleoprotein D. *J. Biol. Chem.* **280**, 18862-18870 (2005).

50. Hiller, M., Zhang, Z., Backofen, R. & Stamm, S. Pre-mRNA secondary structures influence exon recognition. *PLoS Genet.* **3**, e204 (2007).

51. Warf, M.B. & Berglund, J.A. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem Sci* (2009).

52. Jablonski, J.A. & Caputi, M. Role of Cellular RNA Processing Factors in Human Immunodeficiency Virus Type 1 mRNA Metabolism, Replication, and Infectivity. *J. Virol.* **83**, 981-992 (2009).

53. Kikin, O., D'Antonio, L. & Bagga, P.S. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res* **34**, W676-82 (2006).

54. Patel, D.J., Phan, A.T. & Kuryavyi, V. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nuc. Ac. Res.* **35**, 7429-55 (2007).

55. Roy, D. & Lieber, M.R. G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Mol. Cell. Biol.* **29**, 3124-33 (2009).

56. Aguilera, A. & Gomez-Gonzalez, B. Genome instability: a mechanistic view of its causes and consequences. *Nat. Rev. Genet.* **9**, 204-17 (2008).

57. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *Journal of Molecular Graphics* **14**, 51-5, 29-32 (1996).

58. Peterson, R.D., Theimer, C.A., Wu, H. & Feigon, J. New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. *J. Biomol. NMR* **28**, 59-67 (2004).

59. Herrmann, T., Guntert, P. & Wuthrich, K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR* **24**, 171-89 (2002).

60. Herrmann, T., Guntert, P. & Wuthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209-27 (2002).

61. Case, D.A. et al. The Amber biomolecular simulation programs. *J. Computat. Chem.* **26**, 1668-88 (2005).

62. Dignam, J.D., Lebovitz, R.M. & Roeder, R.G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nuc. Ac. Res.* **11**, 1475-89 (1983).

63. Chabot, B., Blanchette, M., Lapierre, I. & La Branche, H. An intron element modulating 5' splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. *Mol. Cell. Biol.* **17**, 1776-86 (1997).

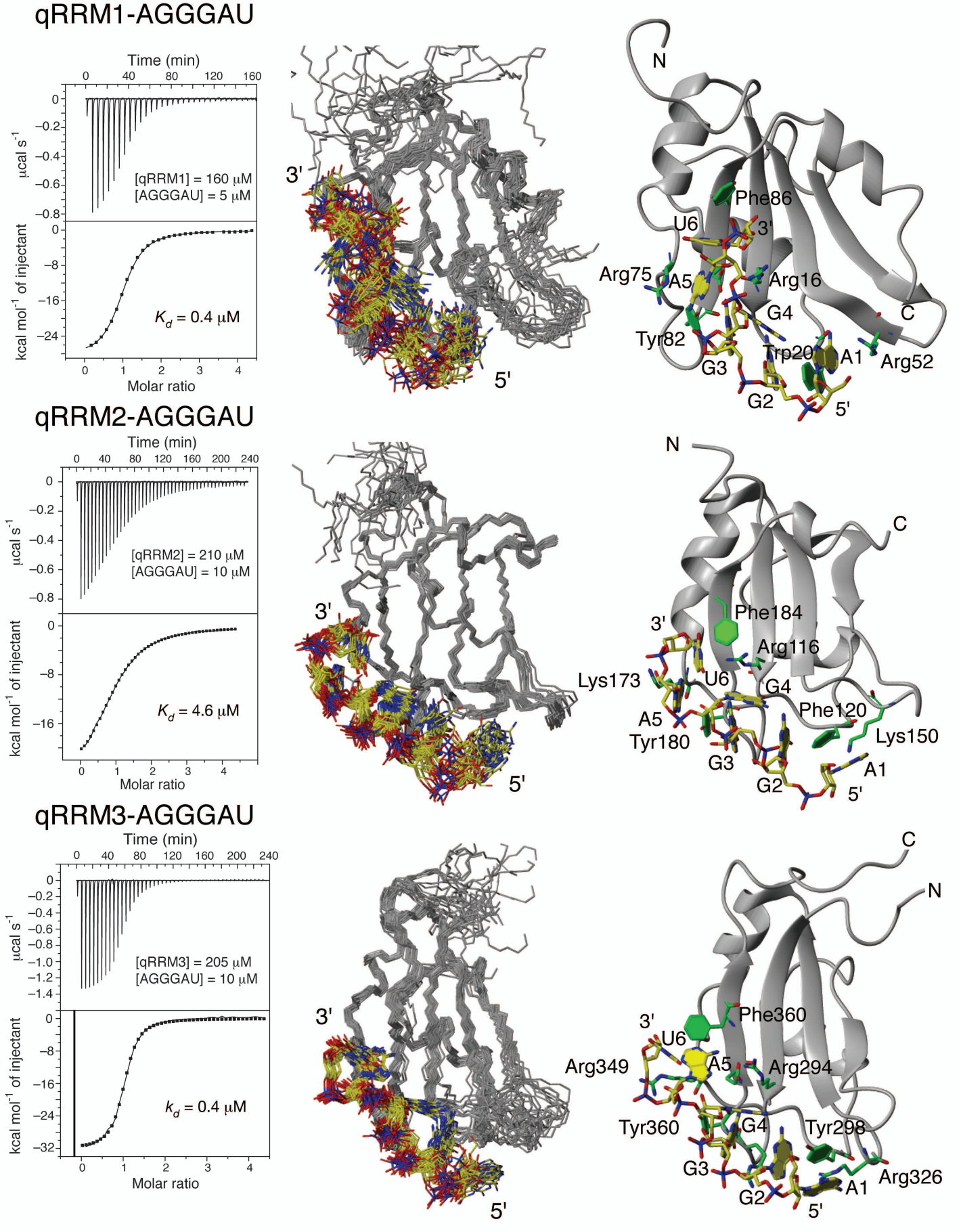


Figure 1 (Allain)

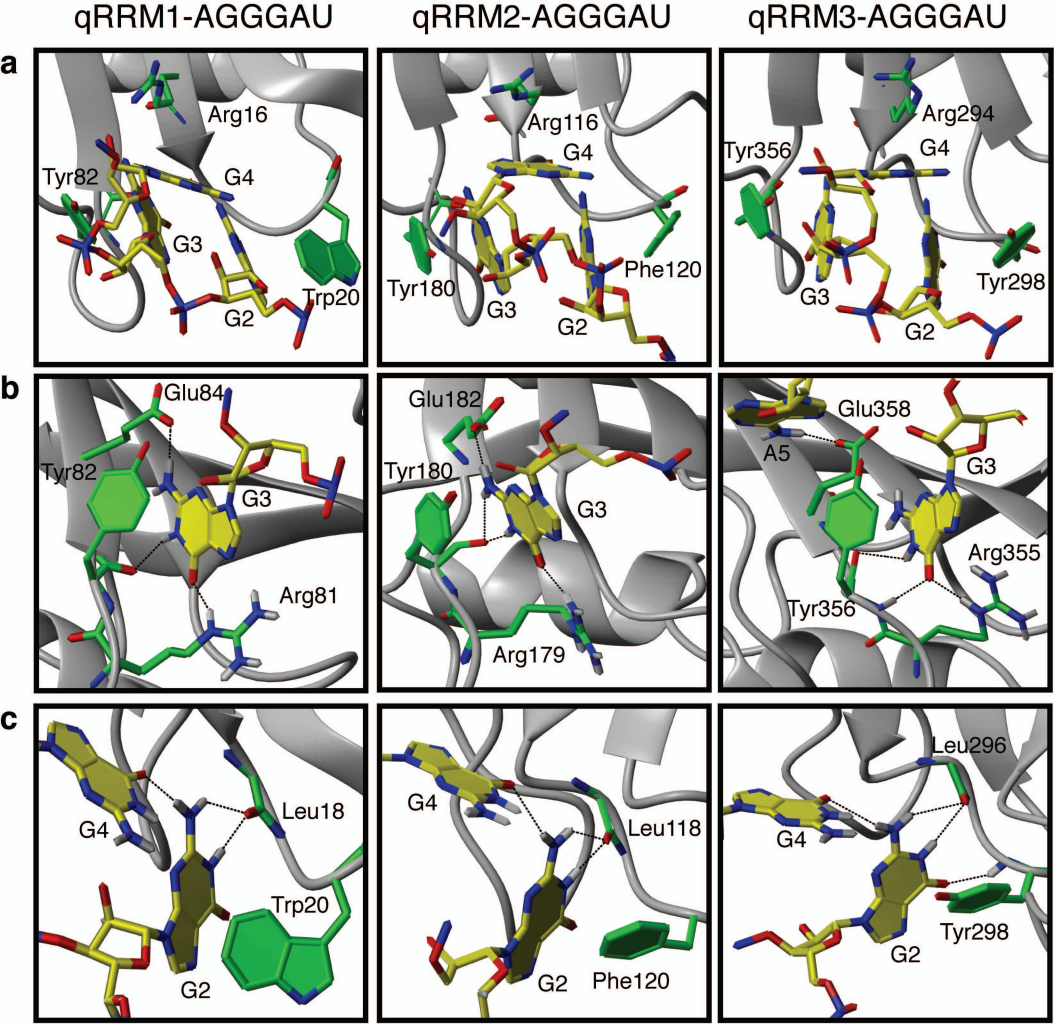


Figure 2 (Allain)

a

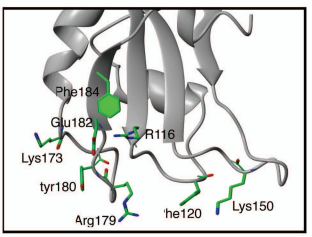
Loop 1 — Loop 3 — Loop 5

(RK)GhPa-X(20-40)-(RK)-X(15-30)-(RK)-X(5)-RYhEhF

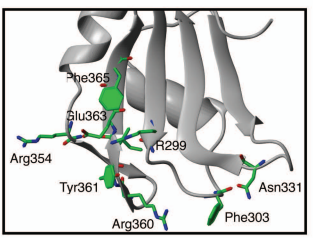
b

HNRNP F	qRRM1	RGLPW-	X_{31}	-R- X_{22}	-RESMGHRYIEVF
HNRNP F	qRRM2	RGLPF-	X_{29}	-K- X_{22}	-KERIGHRYIEVF
HNRNP F	qRRM3	RGLPY-	X_{27}	-R- X_{22}	-RANMQHRYIELF
HNRNP H	qRRM1	RGLPW-	X_{31}	-R- X_{22}	-RETMGHRYVEVF
HNRNP H	qRRM2	RGLPF-	X_{29}	-K- X_{22}	-KERIGHRYIEIF
HNRNP H	qRRM3	RGLPY-	X_{27}	-R- X_{22}	-KANMQHRYVELF
RBM12	RRM2	KGLPF-	X_{29}	-K- X_{22}	-KQYMGNRFIQVH
RBM12B	RRM1	RGLPY-	X_{28}	-R- X_{22}	-RSFMGSRFIEVM
RBM19	RRM2	RGAPF-	X_{27}	-N- X_{22}	-REYMGGRYIEVF
RBM35A	RRM1	RGLPW-	X_{29}	-R- X_{22}	-KHHMGTRYIEVY
RBM35A	RRM2	RGLPF-	X_{32}	-R- X_{22}	-KDLLGKRYIEVF
RBM35A	RRM3	RGLPY-	X_{31}	-R- X_{23}	-KKNMKDRYVEVF
RBM35B	RRM1	RGLPW-	X_{29}	-R- X_{22}	-KHHMGVRYIEVY
RBM35B	RRM2	RGLPY-	X_{31}	-R- X_{23}	-KKVMKERYVEVV
RBM35B	RRM3	RGLPY-	X_{31}	-R- X_{23}	-KKVMKERYVEVV

c

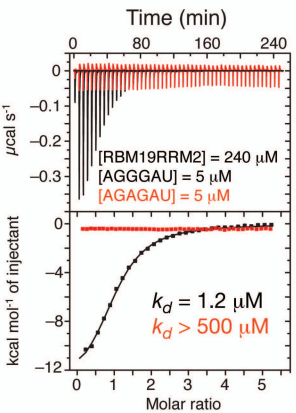


hnRNP F qRRM2



RBM19 RRM2 (pdb:2DGW)

d



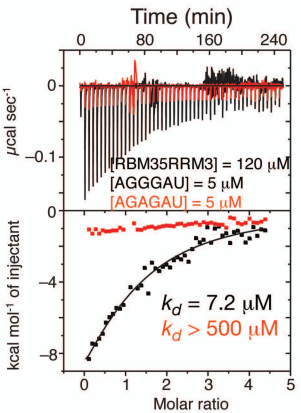


Figure 3 (Allain)

AGGGAU:

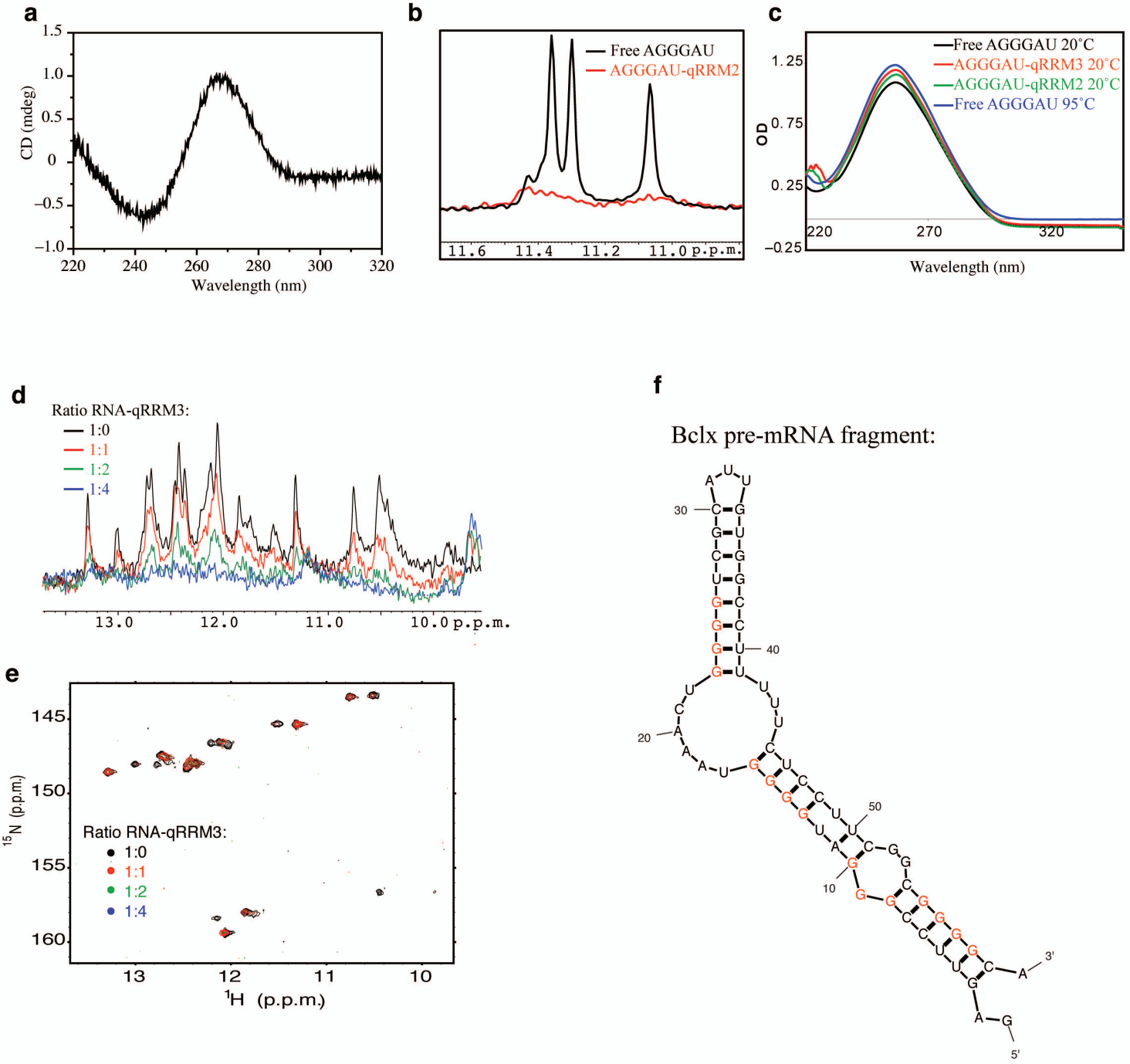


Figure 4 (Allain)

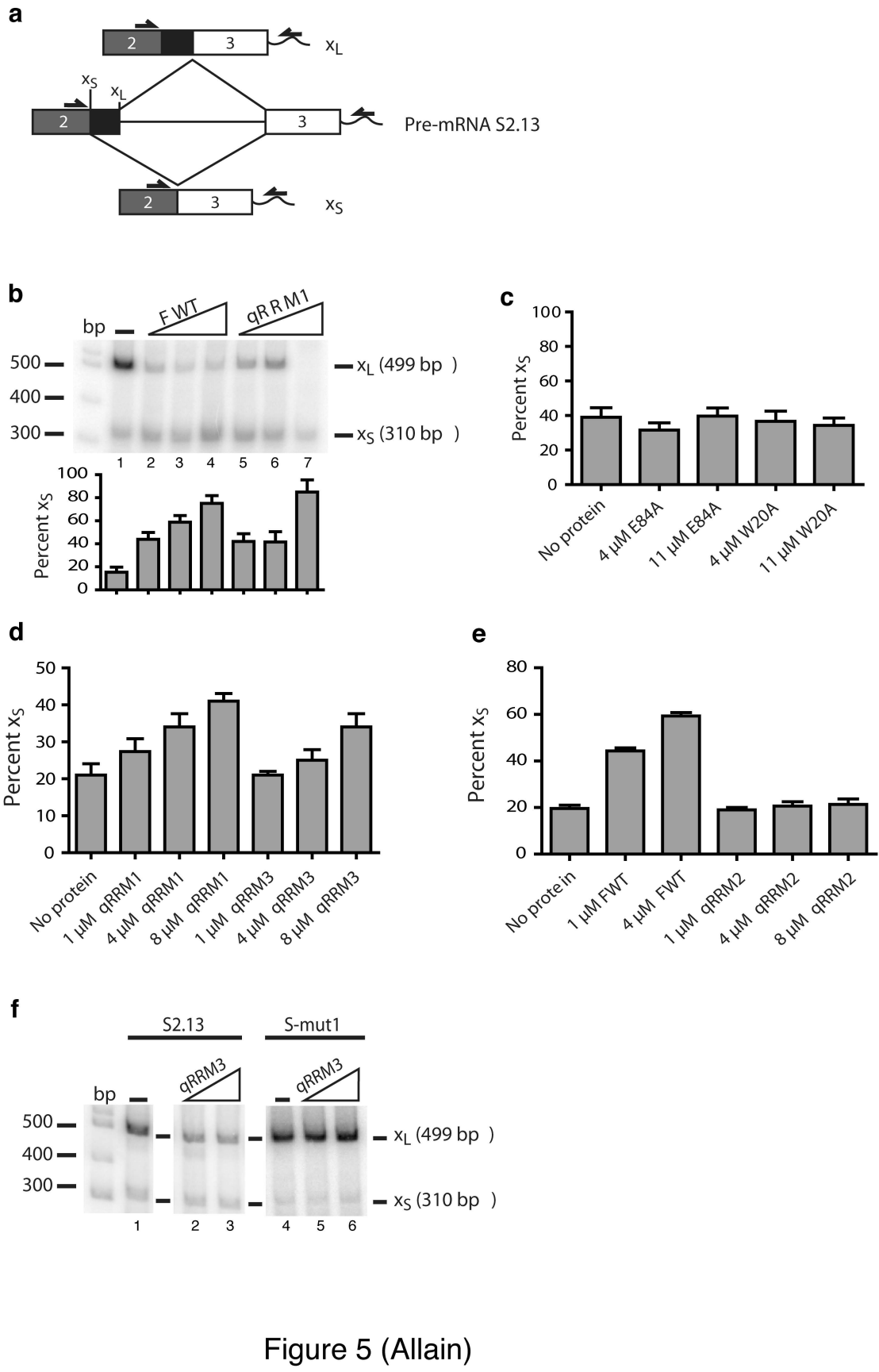


Figure 5 (Allain)