scientific report

The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction

Lenka Skrisovska[1], Cyril F. Bourgeois[2,3], Richard Stefl[1], Sushma-Nagaraja Grellscheid[4], Liliane Kister[2,3], Philipp Wenter[5], David J. Elliott[4], James Stevenin[2,3+] & Frédéric H.-T. Allain[1++]

[1]Institute of Molecular Biology and Biophysics, ETH Zurich, Swiss Federal Institute of Technology, Zurich, Switzerland, [2]Institut de Génétique et de Biologie Moléculaire et Cellulaire, Inserm, U596 & CNRS UMR7104, Illkirch, France, [3]Université L. Pasteur, Strasbourg, France, [4]Institute of Human Genetics, University of Newcastle upon Tyne, Newcastle upon Tyne, UK, and [5]Laboratory of Nucleic Acid Chemistry-EPFL, Lausanne, Switzerland

The RBMY (RNA-binding motif gene on Y chromosome) protein encoded by the human Y chromosome is important for normal sperm development. Although its precise molecular RNA targets are unknown at present, it is suggested that human RBMY (hRBMY) participates in splicing in the testis. Using systematic evolution of ligands by exponential enrichment, we found that RNA stem–loops capped by a $C^A/_UCAA$ pentaloop are high-affinity binding targets for hRBMY. Subsequent nuclear magnetic resonance structural determination of the hRBMY RNA recognition motif (RRM) in complex with a high-affinity target showed two distinct modes of RNA recognition. First, the RRM β-sheet surface binds to the RNA loop in a sequence-specific fashion. Second, the β2–β3 loop of the hRBMY inserts into the major groove of the RNA stem. The first binding mode might be conserved in the paralogous protein heterogeneous nuclear RNP G, whereas the second mode of binding is found only in hRBMY. This structural difference could be at the origin of the function of RBMY in spermatogenesis.
Keywords: alternative splicing; NMR; protein–nucleic acid recognition; spermatogenesis; SELEX

INTRODUCTION

Testes have a highly specialized gene expression program, which is necessary to carry out the complex differentiation of

spermatogonia into mature spermatozoa. In particular, alternative splicing of specific pre-mRNAs is prevalent in testis, although little is known about how these events are regulated at the molecular level (Venables, 2002). Among the potential specific regulators of alternative splicing in testis, the family of human *RBMY* (*hRBMY*; RNA-binding motif gene on Y chromosome) genes was identified as a candidate for the azoospermia factor (AZF; Elliott, 2004). Human RBMY is expressed specifically in the nuclei of adult male germ cells throughout all transcriptionally active stages of spermatogenesis, and deletion of the functional copies of *RBMY* is associated with an arrest of meiotic division I during spermatogenesis (Elliott *et al*, 1997). Human RBMY has an amino-terminal RNA recognition motif (RRM) and a carboxy-terminal domain composed of four repetitions of a Ser-Arg-Gly-Tyr tetrapeptide motif (SRGY box; Ma *et al*, 1993). *RBMY* is found on the Y chromosome of all mammals (Mahadevaiah *et al*, 1998). The mouse RBMY (mRBMY) contains an RRM with 74% similarity to hRBMY, followed by only one SRGY box. RBMY-deficient mice do not show the same phenotype as in humans; they have abnormal sperm development but are not sterile (Mahadevaiah *et al*, 1998).

Human and mouse *RBMY* have a chromosome X-located paralogue (*RBMX*), which encodes the widely expressed heterogeneous nuclear ribonucleoprotein (RNP) G (Delbridge *et al*, 1999). Human hnRNP G contains an N-terminal RRM with 88% similarity to hRBMY, followed by only one SRGY box. There is also a third human retrogene (*hnRNP G-T*) that belongs to the same family, which is expressed only in the testis. hnRNP G-T contains an N-terminal RRM with 84% similarity to hRBMY but no SRGY box (Elliott *et al*, 2000b).

Although nothing is known about the RNA targets of these proteins, several studies suggest a role in the regulation of RNA processing (Venables *et al*, 1999, 2000; Elliott *et al*, 2000a). Human RBMY was shown to interact with Sam68 and the closely related T-STAR protein (signal transduction and RNA binding), which are considered to be molecular transducers between cell signalling and splicing regulation (Elliott, 2004). Furthermore, hRBMY can affect splicing through its interaction with SRp20

[1]Institute of Molecular Biology and Biophysics, ETH Zurich, Swiss Federal Institute of Technology, 8093 Zurich, Switzerland
[2]Institut de Génétique et de Biologie Moléculaire et Cellulaire, and Inserm, U596 and CNRS UMR7104, 67400 Illkirch, France
[3]Université L. Pasteur, 67000 Strasbourg, France
[4]Institute of Human Genetics, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 3BZ, UK
[5]Laboratory of Nucleic Acid Chemistry-EPFL, 1015 Lausanne, Switzerland
[+]Corresponding author. Tel.: +33 (0)3 88 65 33 61; Fax: +33 (0)3 88 65 32 01; E-mail: stevenin@titus.u-strasbg.fr
[++]Corresponding author. Tel.: +41 1 633 3940; Fax: +41 1 633 1294; E-mail: allain@mol.biol.ethz.ch
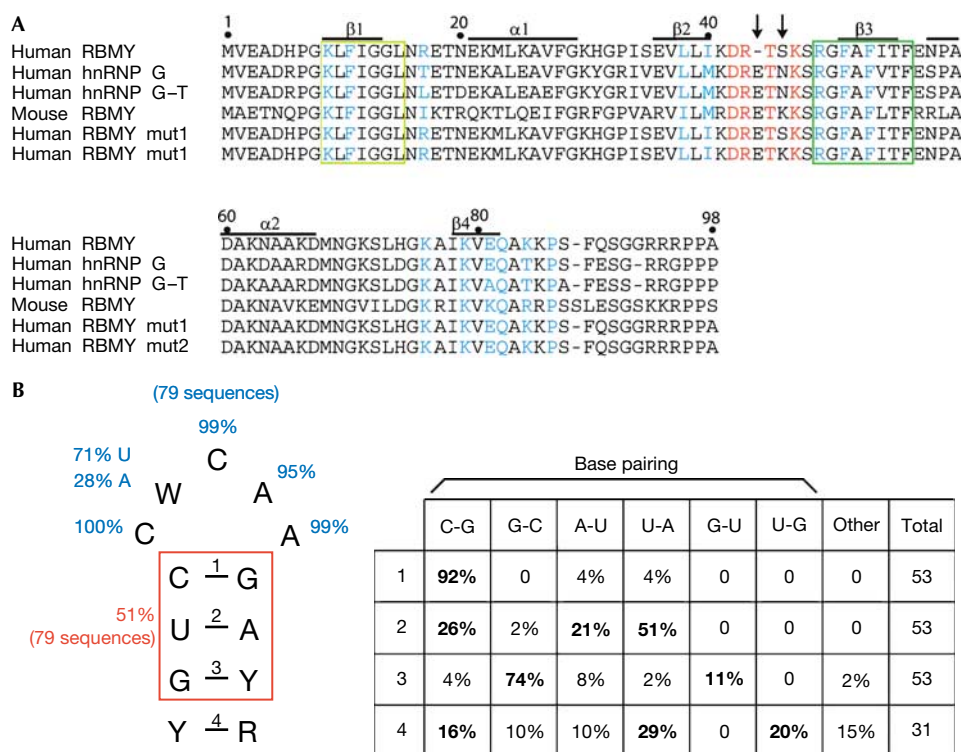
A

| | | β1 | | 20 | α1 | | β2 40 | ↓ ↓ | β3 | |

Human RBMY   MVEADHPGKLFIGGLNRETNEKMLKAVFGKHGPISEVLLIKDR-TSKSRGFAFITFENPA
Human hnRNP G   MVEADRPGKLFIGGLNTETNEKALEAFGKYGRIVEVLLMKDRETNKSRGFAFVTFESPA
Human hnRNP G-T MVEADRPGKLFIGGLNLETDEKALEAEFGKYGRIVEVLLMKDRETNKSRGFAFVTFESPA
Mouse RBMY   MAETNQPGKIFIGGLNIKTRQKTLQEIFGRFGPVARVILMRDRETKKSRGFAFLTFRRLA
Human RBMY mut1 MVEADHPGKLFIGGLNRETNEKMLKAVFGKHGPISEVLLIKDRETSKSRGFAFITFENPA
Human RBMY mut1 MVEADHPGKLFIGGLNRETNEKMLKAVFGKHGPISEVLLIKDRETKKSRGFAFITFENPA

| | 60 | α2 | | β4 80 | | 98 | |

Human RBMY   DAKNAAKDMNGKSLHGKAIKVEQAKKPS-FQSGGRRRPPA
Human hnRNP G   DAKDAARDMNGKSLDGKAIKVEQATKPS-FESG-RRGPPP
Human hnRNP G-T DAKAAARDMNGKSLDGKAIKVAQATKPA-FESS-RRGPPP
Mouse RBMY   DAKNAVKEMNGVILDGKRIKVKQARRPSSLESGSKKRPPS
Human RBMY mut1 DAKNAAKDMNGKSLHGKAIKVEQAKKPS-FQSGGRRRPPA
Human RBMY mut2 DAKNAAKDMNGKSLHGKAIKVEQAKKPS-FQSGGRRRPPA

B

	C-G	G-C	A-U	U-A	G-U	U-G	Other	Total
1	**92%**	0	4%	4%	0	0	0	53
2	**26%**	2%	**21%**	**51%**	0	0	0	53
3	4%	**74%**	8%	2%	**11%**	0	2%	53
4	**16%**	10%	10%	**29%**	0	**20%**	15%	31

Fig 1 | Amino-acid sequence of the heterogeneous nuclear RNP G protein family and RNA-binding SELEX consensus sequence obtained for human RBMY. (**A**) Sequence alignment of human RBMY (hRBMY) RRM, and related orthologues and paralogues. The RNP2 and RNP1 motifs are indicated by light green and dark green boxes, respectively. Amino acids of hRBMY found in contact with the RNA loop and stem are shown in blue and red, respectively. The arrows show the two non-conserved mutated amino acids of β2–β3 loop. (**B**) Consensus sequence obtained after SELEX. All analysed sequences (supplementary Fig S1 online) can fold into a stable stem–loop structure with a highly conserved CWCAA loop. The table shows the statistical predominance for specific base-pairing in the stem, including non-canonical base pairs (G-U or U-G). The red box shows the core consensus sequence of the stem that we observed in 51% of sequences. RBMY, RNA-binding motif gene on Y chromosome; SELEX, systematic evolution of ligands by exponential enrichment; Y, C or U; W, U or A; R, G or C.

or Tra2β (Elliott *et al*, 2000a, Venables *et al*, 2000), which belong to the family of SR-rich pre-mRNA splicing regulators (Bourgeois *et al*, 2004).

To understand better the biological function of this family of proteins, we carried out systematic evolution of the ligand by exponential enrichment (SELEX) with hRBMY RRM. We found that the RRM binds with high specificity to stem–loop RNAs containing a $C^{A}/_{U}CAA$ consensus sequence in the loop. We determined the solution structure of hRBMY RRM in complex with one of its RNA targets. The structure shows that the hRBMY RRM not only recognizes the loop in a sequence-specific manner, but also the shape of the RNA as the β2–β3 loop of the RRM is inserted into the major groove of the RNA stem.

RESULTS

Identification of human RBMY RNA targets by SELEX

Human RBMY contains an RRM with unknown properties; therefore, we used the SELEX approach to determine its RNA-binding specificity (Cavaloc *et al*, 1999). The glutathione *S*-transferase (GST) fusion protein included the 108 N-terminal residues of hRBMY containing the RRM. After six cycles of selection, we obtained fully coherent results as the same consensus motif was obtained with two distinct 20-nucleotide

(nt) randomized matrices (supplementary Fig S1 online). Interestingly, the selected motifs (Fig 1) consist of a hairpin, including an invariable pentaloop ($C^{A}/_{U}CAA$) and a stem of variable length (between 4 and 11 perfect base pairs) in which a C-G base-pair is predominantly adjacent to the loop. Altogether, 51% of the selected clones had a GUC–loop–GAY structure. However, the selection of other stem sequences indicates that hRBMY can recognize stems with different base pair compositions. Electrophoretic mobility shift assays (EMSA) confirmed that the interaction occurs with high affinity ($K_d \approx 10^{-9}$ M; see below).

Structure of RBMY RRM in complex with S1A RNA

To understand the molecular basis of the recognition by hRBMY RRM for the selected RNA stem–loops, we investigated the structure of one complex using nuclear magnetic resonance (NMR; supplementary Fig S2 online). The chosen RNA ('S1A') was derived from the S1 sequence. S1A is 21 nt-long and contains eight base pairs capped by a CACAA loop (supplementary Fig S2A online). The stem contains the most common SELEX sequence, GUC–loop–GAC.

The structure of the complex was determined with very high precision because a structure calculation was used with a very high number of nuclear Overhauser effect (NOE) constraints
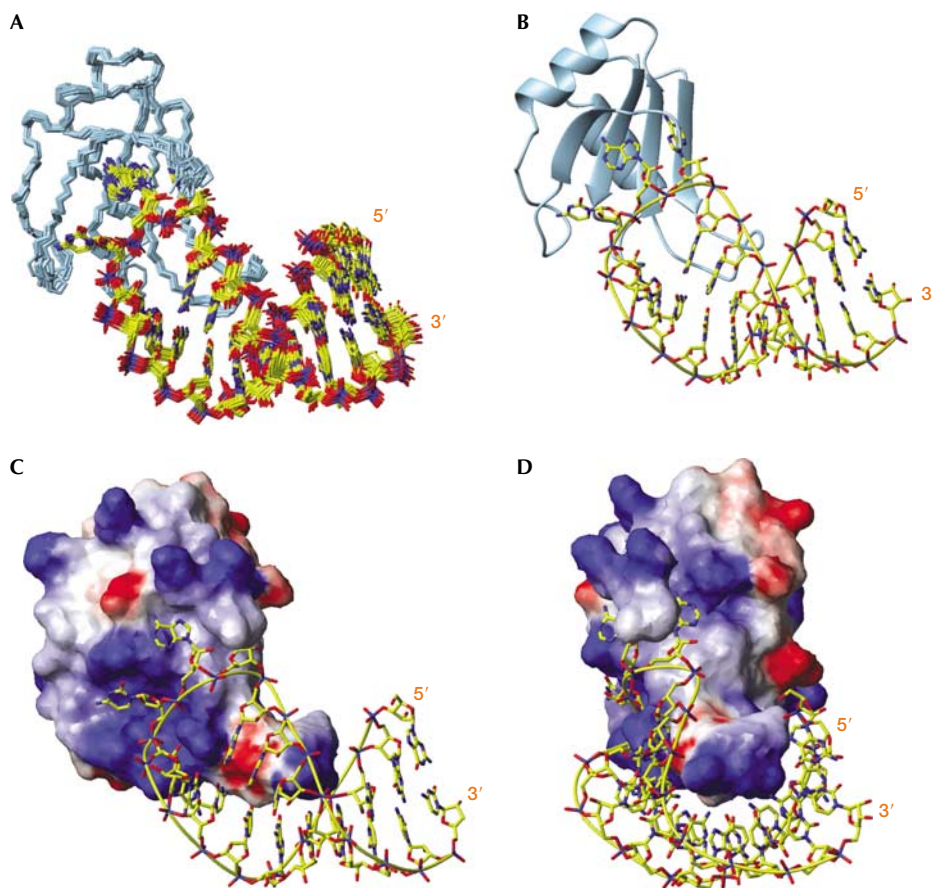
Fig 2 | Overall view of the complex between the human RBMY RRM and S1A RNA. (**A**) The 17 lowest energy structures of the complex are superimposed. (**B**) View of the lowest energy structure of the ensemble. (**C**) Surface representation of the protein in the complex and (**D**) rotated by 90°. RBMY, RNA-binding motif gene on Y chromosome; RRM, RNA recognition motif.

(1,879 including 124 intermolecular), supplemented by 54 angle constraints from residual dipolar couplings in both the RNA and the protein (supplementary Table SI online; Fig 2A). Human RBMY RRM adopts the expected βαββαβ topology, with two α-helices packed against one side of the four-stranded β-sheet (Maris *et al*, 2005).

Human RBMY RRM specifically recognizes a CAA triplet

C11, *A12* and *A13* (nucleotides are written in italics to distinguish them from amino acids) protrude from the loop and are spread on the RRM β-sheet surface (Fig 2B). The three nucleotides are stabilized by contact with the protein main chain and several side chains but not by intra-RNA interactions (Fig 3). *A12* adopts an unusual *syn* conformation and the sugar puckers of *C11* and *A13* are both C2'-*endo*. There are specific contacts with *A12* and *A13*, which discriminate for adenines at these positions. *A13* is recognized by K84 main-chain oxygen and by K9 (β1) and L38 (β2) side chains (Fig 3C). Similarly, *A12* is specifically recognized by the K84 main-chain amide (Fig 3B), the Q82 (β4) main-chain oxygen and the E81 (β4) side chain (K84 amide experiences a large chemical shift change on complex formation; supplementary Fig S2B online). *C11* is recognized by a contact with the K79 (β4)

side chain (Fig 3A). From the structure, it appears that a *U* could also be accommodated at this position.

The β2–β3 loop inserts into the RNA major groove

An unexpected structural feature of the complex is the interaction of the RRM β2–β3 loop with the major groove of the RNA helix from the second base pair to the last base pair of the stem (Figs 2C,D,4A). The seven amino acids of the β2–β3 loop (D42–R48) form a β-hairpin that inserts itself into the deep major groove of the RNA helix. Most of the intermolecular interactions are non-sequence-specific in nature and involve side-chain and main-chain contact with phosphate groups of the RNA (Fig 4A). All five side chains from D42 to D46 are involved in the recognition of the RNA major groove. The R43 side chain forms salt bridges with *A3* and *C4* phosphate, whereas its main-chain amide forms contacts with *A15* phosphate on the opposite strand of the helix. The K46 side chain forms a salt bridge with *C16* phosphate (Fig 4A). The R43 and K46 side chains, located on opposite strands of the β2–β3 loop, cross each other to contact on the phosphate oxygen with the different strands of the helix (Fig 4B). Contacts by T44 to *G2* phosphate, and possibly by S45, further stabilize the interaction. Finally, the side chain of D42 forms two hydrogen
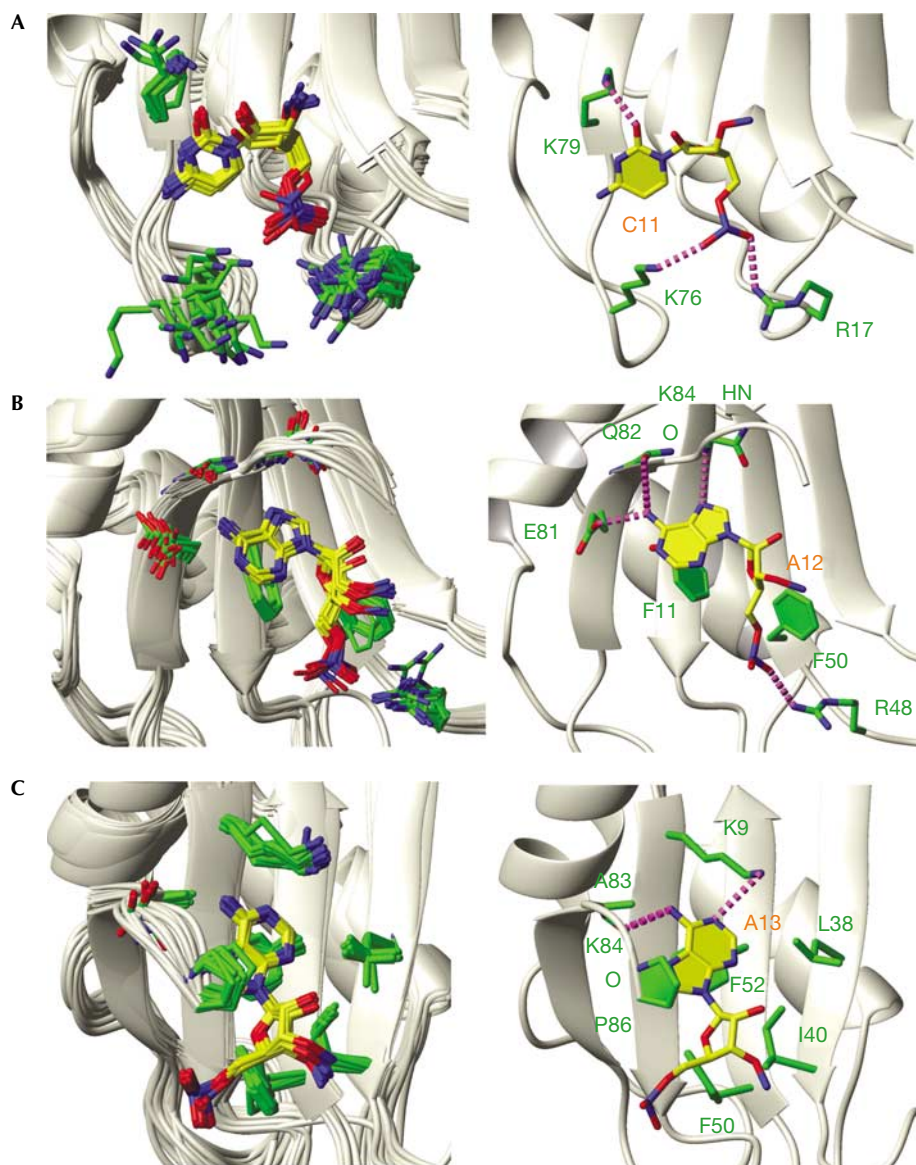
Fig 3 | Close-up of the interactions between the RRM β-sheet and *C11*, *A12* and *A13* of the RNA. (A) The ensemble of complex structures are superimposed around *C11* (left) and a view of the most representative structure (right). **(B)** As in **(A)**, but around *A12*. **(C)** As in **(A)**, but around *A13*. Protein side chains are shown in green and RNA residues are shown in yellow. Potential hydrogen bonds are represented by purple dotted lines. RRM, RNA recognition motif.

bonds, one with the main-chain amide of R48 and the other with *A15* amino (Fig 4A).

RBMY RRM stabilizes the 5′ end of the pentaloop
C9 and *A10* extend the RNA helical stack, as *C9* stacks over *C8*, and *A10* over *C9*. In addition, *C9* forms contacts with *C8* phosphate (Fig 4C) and the R17 side chain (β1–α1 loop) stacks over *A10* and forms contacts with *C11* phosphate. *C11* phosphate is also in contact with K76 (α2–β4 loop, Fig 4C). Human RBMY binding selectivity for *C9* and *A10* seems to be indirect; *C9* might be preferred to other nucleotides to prevent the formation of a base pair with *A13* in the free RNA and *A10* might be preferred to

a *G* as a *G* amino would clash sterically with the protein backbone. Finally, R48 further stabilizes the complex by forming contacts with all three RNA elements (Fig 4C): the stem (*G14*), the CAA triplet (phosphate of *A12*), and *C9* or *A10*.

Mutagenesis studies confirm the recognition mode
We carried out EMSA experiments using RNA sequences that are representative of the two sets of sequences identified by SELEX. Sequences 2, 4 and 6 from set I (data not shown), S1 and S2 from set II, as well as S1A, all showed efficient binding with the GST-fused hRBMY RRM (Fig 5A). The apparent K_d was between 0.6 and 0.9 nM for S1 and S2, respectively. Mutating *A12* or *A13* to a
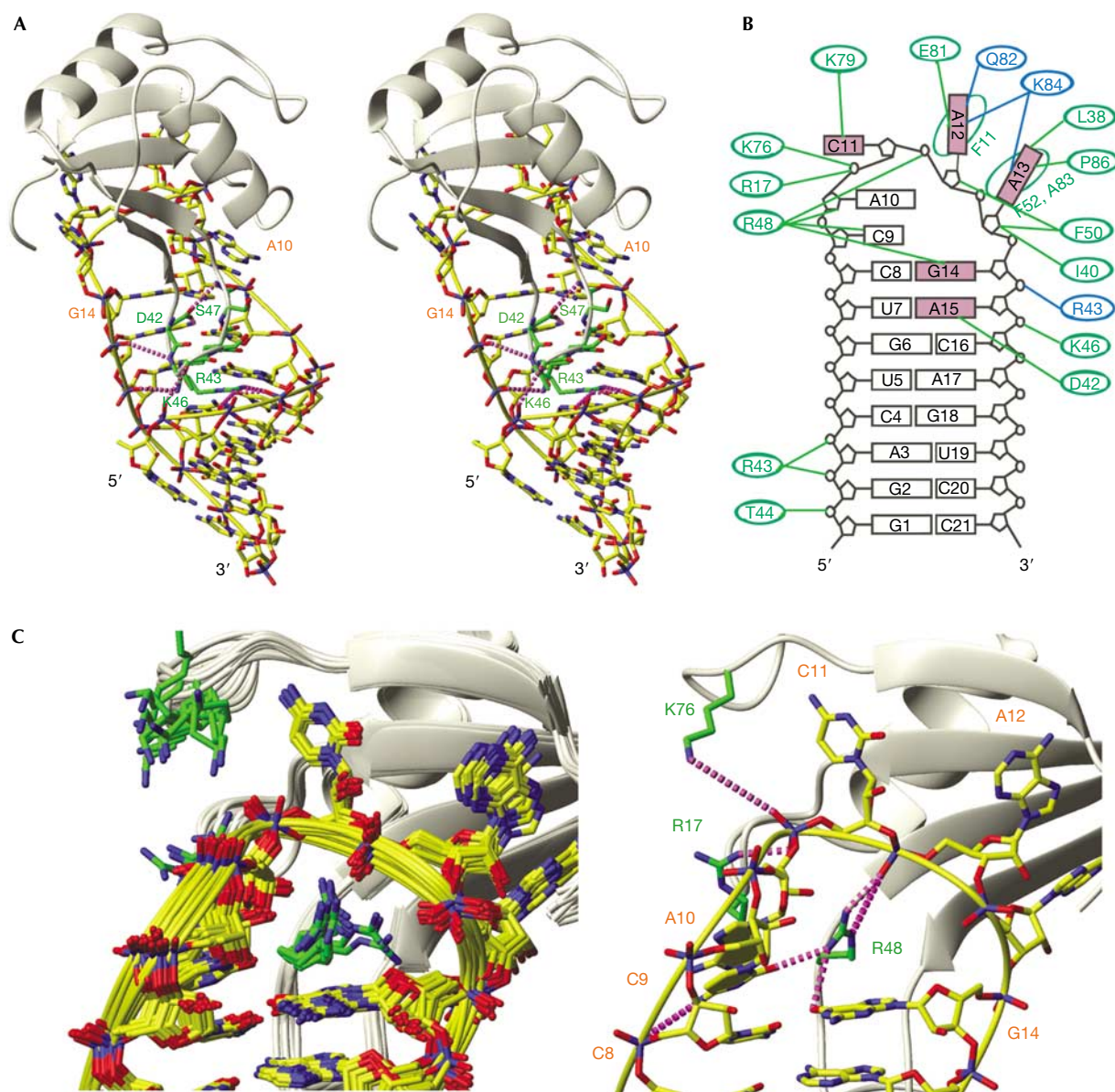
Fig 4 | View of the protein–RNA interactions between RRM loops and the RNA, and schematic diagram of the interactions. (**A**) Stereoviews of the contact between the RRM β2–β3 loop and the RNA stem. (**B**) Schematic diagram of interactions. The amino-acid residues in contact with the RNA side chain and main chain are shown in green and blue, respectively. The nucleotides shown in purple are sequence specifically recognized by human RBMY RRM. (**C**) The ensemble of complex structures are superimposed around C9, A10 and C11 (left), and a view of the most representative structure. RBMY, RNA-binding motif gene on Y chromosome; RRM, RNA recognition motif.

G resulted in a complete loss of binding, showing that the identity of these residues is crucial for high affinity (Fig 5B). By contrast, the replacement of *C11* by a *U* (*C11U*) was better tolerated as it resulted in an approximately fourfold decrease in affinity. This agrees well with the structure, as the identity of *C11* is not recognized as strongly as that of *A12* or *A13*. Changing the size of the pentaloop also has a strong effect, as the removal of the A/U residue (*A10*) resulted in very weak binding. Similarly, the insertion of an additional *A* (*+A13*) resulted in a complete loss of binding (Fig 5B). These experiments confirm that a loop size of five nucleotides is optimal for hRBMY binding. Finally, we tested

whether the hydrogen bond between D42 and *A15* was crucial by replacing *U7-A15* by a *C7-G15* base pair (Fig 5B). Surprisingly, this change in the sequence did not alter the affinity. It is possible that a slight rearrangement could take place, allowing D42 to interact with *C7* in the mutant RNA instead of *A15* in the wild type. This mutation indicates that the contact mediated by D42 is not sequence specific. Thus, the interactions mediated by the β2–β3 loop would constitute a shape-specific recognition of the RNA major groove.

The structure shows that the β2–β3 loop fits into the major groove of the RNA, and that there are steric constraints associated
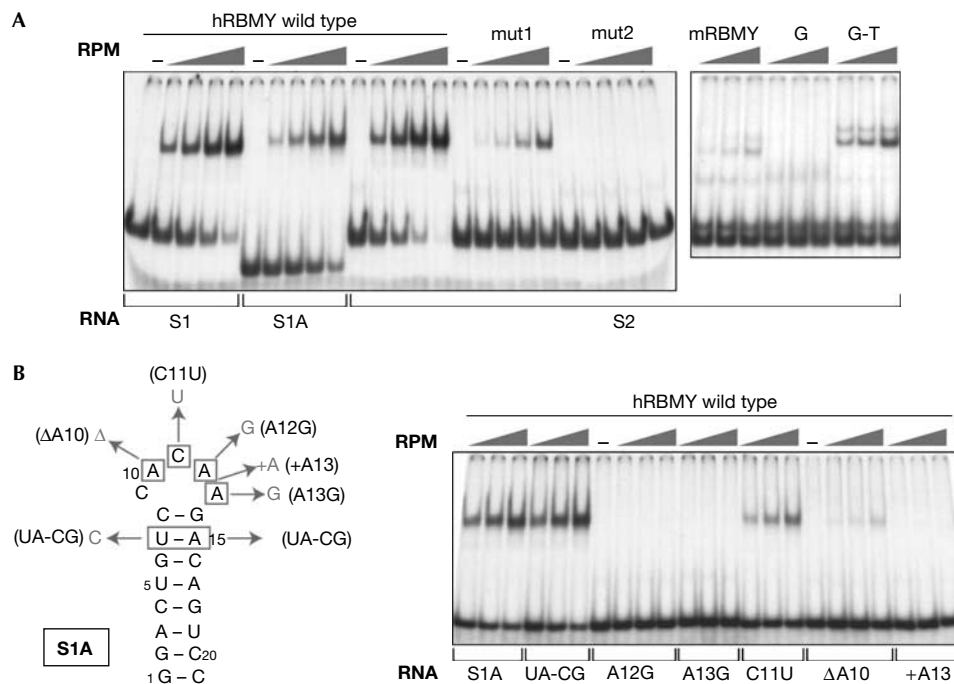
Fig 5 | SELEX RNA sequences are recognized specifically by human RBMY RRM. (**A**) The β2–β3 loop of human RBMY (hRBMY) RRM defines the specificity of interaction with its RNA targets. Increasing concentrations of the different RRMs (0.625, 1.25, 2.5 and 5 nM for wild-type hRBMY, 1.25, 2.5, 5 and 10 nM for mutant hRBMY, 2.5, 5 and 10 nM for mRBMY and heteronuclear RNP G and G-T) were incubated with RNA and the complexes were separated by EMSA. (**B**) The sequence of the RNA pentaloop is a crucial determinant of the interaction with the hRBMY RRM. The different mutations introduced in the S1A RNA are shown in grey boxes. For EMSA, the same concentrations of hRBMY RRM (2.5, 5 and 10 nM) were added to a constant concentration of wild-type or mutant RNA, and resulting complexes were analysed as in (**A**). EMSA, electrophoretic mobility shift assay; RBMY, RNA-binding motif gene on Y chromosome; RRM, RNA recognition motif; SELEX, systematic evolution of ligands by exponential enrichment.

with this insertion. Interestingly, the RRM of the mRBMY—as well as those of the human hnRNP G and G-T—shows two changes in the β2–β3 loop compared with that of hRBMY (Fig 1A): the three RRMs contain an additional E between R43 and T44, and S45 of hRBMY is replaced by K in mRBMY or N in hnRNP G and G-T. To study the role of these residues, two hRBMY RRM mutants (mut1 and mut2; Fig 1A) were made and their affinity was compared with the wild-type RRM (Fig 5A, left panel). Strikingly, the simple insertion of an E in the β2–β3 loop resulted in more than a tenfold decrease in binding. Furthermore, when the additional S to K mutation was carried out—to make a mouse-like β2–β3 loop—no detectable interaction was observed (Fig 5A). This confirms the crucial role of the β2–β3 loop of the hRBMY RRM for binding RNA stem–loops. We also showed that the three RRMs of mRBMY, human hnRNP G and G-T could not efficiently bind to the S2 stem–loop (Fig 5A, right panel).

DISCUSSION
A dual mode of RNA recognition by human RBMY
Using SELEX, we have identified an unusual RNA binding consensus sequence for hRBMY RRM. The high-affinity sites are RNA stem–loops with a $C^{A/_U}CAA$ loop and a GUC–loop–GAY consensus in the last three base pairs of the stem (Fig 1B). The structure of the RRM complexed with a stem–loop containing a CACAA loop showed that the recognition of the RNA is both

sequence- and shape-specific (Stefl *et al*, 2005). The structure explains how *C9, A12, A13* and *G14* are sequence-specifically recognized by the RRM (Figs 3,4), confirming and explaining the SELEX consensus in the pentaloop and at the first base pair. However, from a structural point of view, it remains unclear why an $^A/_U$ and a C are preferred in the second and third positions, respectively of the loop in the SELEX sequence. It is possible that other base types in these two positions would modify the accessibility of the last pentaloop triplet or induce a folding of RNA that might prevent protein binding or lower the binding affinity. More surprisingly, we found that the RNA stem is recognized by the RRM β2–β3 loop, which forms a β-hairpin and is inserted into the major groove of the RNA helix. This recognition is shape specific, as it is the complementary shape and charge between the β2–β3 loop and the RNA major groove that dictate this intermolecular interaction (Fig 2). When this β2–β3 loop is elongated or mutated, RNA binding is weakened or lost, confirming the importance of this interaction for complex formation. These two modes of binding result from an 'induced fit' of one of the binding partners. In the sequence-specific mode, the unstructured RNA pentaloop folds after binding on the rigid β-sheet and becomes ordered. In the shape-recognition mode, the flexible β2–β3 loop inserts into the major groove of the rigid RNA stem and becomes ordered (L.S. & F.A., unpublished data). This mode of recognition is unprecedented among RRMs, once again

RNA recognition by other hnRNP-G family members

We have shown that the RRMs of human hnRNP G, G-T or mRBMY are unable to bind with high affinity to the hRBMY-specific hairpin structure. This is in good agreement with our structural data, when considering the role of the β2–β3 loop in strengthening the interaction between hRBMY and the RNA (Figs 4,5). However, as K9, L38 and E81 are conserved between hnRNP G and hRBMY, it is possible that hnRNP G RRM can similarly recognize CAA in a sequence-specific manner. Interestingly, one RNA sequence identified as a potential target for hnRNP G contains a CAA triplet (Nasim et al, 2003). By contrast, the replacement of E81 in hnRNP G-T and mRBMY (by A or K, respectively) is likely to impair binding to CAA. So far, little data are available concerning the RNA-recognition properties of RBMY and related hnRNP proteins. A previous study indicated that hRBMY and hnRNP G are nonspecific RNA-binding proteins (Hofmann & Wirth, 2002). We show here that this might not be the case, but further analyses are necessary to understand better the properties of all these related RRMs.

Finally, the particular RNA-binding properties of hRBMY could have an important role in the function of the protein in human testes. The infertility caused by deletions in this RRM-encoding gene (Elliott et al, 1997) indicate that crucial RNA processing pathways are disrupted. Protein–protein interactions and sub-nuclear localization experiments suggest a role for RBMY in splicing, although its precise function is still not clear. Testis is one of the tissues in which alternative splicing is largely used (Xu et al, 2002; Yeo et al, 2004), and is needed to establish the exclusive pattern of gene expression that occurs throughout the different stages of spermatogenesis (Venables, 2002). A recent study carried out with 52 different tissues and more than 10,000 genes showed that testis has the highest rate of divergence in alternative splicing events between human and mouse (Kan et al, 2005). Our findings might explain why the mRBMY seems to have a function different from that of the hRBMY (Szot et al, 2003), as both proteins might have different RNA targets. However, we cannot rule out that the natural hRBMY target sequences differ significantly from the stem–loop motifs we identified by SELEX. As a first step towards the identification of the biological RNA targets of hRBMY in vivo, we screened an alternative exon database using an algorithm based on the conservation of the stem–loop structure we characterized by SELEX (supplementary information online). This screening showed that putative hRBMY targets indeed exist within or in the vicinity of exons alternatively spliced in the testis (supplementary Table S2 online). Further work is necessary to analyse whether these RNA sequences are evolutionarily conserved in mammals and whether they are functionally relevant.

METHODS

SELEX and EMSA. SELEX was carried out as described previously (Cavaloc et al, 1999), with only minor modifications using the GST-fused RRM of hRBMY (amino acids 1–108; for details see the supplementary information online).

For EMSA, we used the GST-fused RRMs of hRBMY, mRBMY, hnRNP G or hnRNP G-T, as well as the two mutants of hRBMY. [^{32}P]RNA was transcribed in vitro with T7 RNA polymerase and incubated with the different proteins as described previously (Cavaloc et al, 1999).

Cloning, expression and purification of hRBMY RRM for NMR. The N-terminal RRM (amino acids 1–108) of the hRBMY was subcloned in pET30a+ (Invitrogen Corp., Carlsbad, CA, USA). The construction included a two-residue linker (L-E) between the RRM and the 6 × His tag. For ^{15}N and ^{15}N,^{13}C-labelling, Escherichia coli BL21(DE3)pLysS was used. Expression was carried out in M9 media containing [^{15}N]NH$_4$Cl and [^{13}C]glucose. Human RBMY RRM was purified by Ni affinity and cation exchange chromatography. The protein solution was concentrated to 1 mM, as measured by UV spectroscopy at 205 nm.

RNA transcription and complex formation for NMR analysis. S1A RNA was prepared in vitro using T7 RNA polymerase and purified by anion exchange chromatography. RNA samples were dissolved in 25 mM NaH$_2$PO$_4$/NaOH at pH 7.0.

The hRBMY RRM–S1A complex was studied at 1 mM at a 1:1 ratio of protein and RNA in 25 mM NaH$_2$PO$_4$/NaOH and 25 mM NaCl buffer (pH 7.0; see the supplementary information online for the resonance assignment strategy).

Structure determination. In total, 124 intermolecular NOEs between the RRM and S1A were assigned. The RDCs of hRBMY RRM–S1A were obtained using the Pf1 phages as an aligning medium.

The structure determination was carried out as described previously (Oberstrass et al, 2006). The 17 final conformers with the lowest total energy or with the lowest alignment tensor energy were selected to form the final ensemble of conformers.

Structural data. All restraints used in structure determination and the 17 final structures have been deposited at the Protein Data Bank under the accession code 2FY1.

Supplementary information is available at EMBO reports online (http://www.emboreports.org).

REFERENCES
Bourgeois CF, Lejeune F, Stevenin J (2004) Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. Prog Nucleic Acid Res Mol Biol 78: 37–88
Cavaloc Y, Bourgeois CF, Kister L, Stevenin J (1999) The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. RNA 5: 468–483
Delbridge ML, Lingenfelter PA, Disteche CM, Graves JA (1999) The candidate spermatogenesis gene RBMY has a homologue on the human X chromosome. Nat Genet 22: 223–224
Elliott DJ (2004) The role of potential splicing factors including RBMY, RBMX, hnRNPG-T and STAR proteins in spermatogenesis. Int J Androl 27: 328–334
Elliott DJ, Bourgeois CF, Klink A, Stevenin J, Cooke HJ (2000a) A mammalian germ cell-specific RNA-binding protein interacts with ubiquitously expressed proteins involved in splice site selection. Proc Natl Acad Sci USA 97: 5717–5722

Elliott DJ, Millar MR, Oghene K, Ross A, Kiesewetter F, Pryor J, McIntyre M, Hargreave TB, Saunders PTK, Vogt PH, Chandley AC, Cooke H (1997) Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. *Proc Natl Acad Sci USA* **94:** 3848–3853

Elliott DJ, Venables JP, Newton CS, Lawson D, Boyle S, Eperon IC, Cooke HJ (2000b) An evolutionarily conserved germ cell-specific hnRNP is encoded by a retrotransposed gene. *Hum Mol Genet* **9:** 2117–2124

Hofmann Y, Wirth B (2002) hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-β1. *Hum Mol Genet* **11:** 2037–2049

Kan Z, Garrett-Engele PW, Johnson JM, Castle JC (2005) Evolutionarily conserved and diverged alternative splicing events show different expression and functional profiles. *Nucleic Acids Res* **33:** 5659–5666

Ma K *et al* (1993) A Y chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis. *Cell* **75:** 1287–1295

Mahadevaiah SK *et al* (1998) Mouse homologues of the human AZF candidate gene RBM are expressed in spermatogonia and spermatids, and map to a Y chromosome deletion interval associated with a high incidence of sperm abnormalities. *Hum Mol Genet* **7:** 715–727

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

Nasim MT, Chernova TK, Chowdhury HM, Yue B-G, Eperon IC (2003) HnRNP G and Tra2β: opposite effects on splicing matched by antagonism in RNA binding. *Hum Mol Genet* **12:** 1337–1348

Oberstrass FC, Lee A, Stefl R, Janis M, Chanfreau G, Allain FH (2006) Shape-specific recognition in the structure of the Vts1p SAM domain with RNA. *Nat Struct Mol Biol* **13:** 160–167

Stefl R, Skrisovska L, Allain FHT (2005) RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. *EMBO Rep* **6:** 33–38

Szot M, Grigoriev V, Mahadevaiah SK, Ojarikre OA, Tour A, Von Glasenapp E, Rattigan A, Turner JMA, Elliott DJ, Burgoyne PS (2003) Does Rbmy have a role in sperm development in mice? *Cytogenetic and Genome Research* **103:** 330–336

Venables JP (2002) Alternative splicing in the testes. *Curr Opin Genet Dev* **12:** 615–619

Venables JP, Elliott DJ, Makarova OV, Makarov EM, Cooke HJ, Eperon IC (2000) RBMY, a probable human spermatogene. *Hum Mol Genet* **9:** 685–694

Venables JP, Vernet C, Chew SL, Elliott DJ, Cowmeadow RB, Wu J, Cooke HJ, Artzt K, Eperon IC (1999) T-STAR/ETOILE: a novel relative of SAM68 that interacts with an RNA-binding protein implicated in spermatogenesis. *Hum Mol Genet* **8:** 959–969

Xu Q, Modrek B, Lee C (2002) Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res* **30:** 3754–3766

Yeo G, Holste D, Kreiman G, Burge CB (2004) Variation in alternative splicing across human tissues. *Genome Biol* **5:** R74