

Molecular basis of RNA recognition by the human alternative splicing factor Fox-1

Sigrid D Auweter^{1,5}, Rudi Fasan²,
Luc Reymond³, Jason G Underwood⁴,
Douglas L Black⁴, Stefan Pitsch³
and Frédéric H-T Allain^{1,*}

¹Institute for Molecular Biology and Biophysics, Swiss Federal Institute of Technology (ETH) Zurich, Zurich, Switzerland, ²Institute of Organic Chemistry, University of Zurich, Zurich, Switzerland, ³Laboratory of Nucleic Acid Chemistry LCAN-EPFL, Lausanne, Switzerland and ⁴Department of Microbiology, Immunology and Molecular Genetics and Howard Hughes Medical Institute, University of California Los Angeles, CA, USA

The Fox-1 protein regulates alternative splicing of tissue-specific exons by binding to GCAUG elements. Here, we report the solution structure of the Fox-1 RNA binding domain (RBD) in complex with UGCAUGU. The last three nucleotides, UGU, are recognized in a canonical way by the four-stranded β -sheet of the RBD. In contrast, the first four nucleotides, UGCA, are bound by two loops of the protein in an unprecedented manner. Nucleotides U₁, G₂, and C₃ are wrapped around a single phenylalanine, while G₂ and A₄ form a base-pair. This novel RNA binding site is independent from the β -sheet binding interface. Surface plasmon resonance analyses were used to quantify the energetic contributions of electrostatic and hydrogen bond interactions to complex formation and support our structural findings. These results demonstrate the unusual molecular mechanism of sequence-specific RNA recognition by Fox-1, which is exceptional in its high affinity for a defined but short sequence element.

The EMBO Journal advance online publication, 15 December 2005; doi:10.1038/sj.emboj.7600918

Subject Categories: RNA; structural biology

Keywords: alternative splicing; NMR; protein–nucleic acid recognition; surface plasmon resonance

Introduction

The RNA element UGCAUG has long been known to strongly influence splicing of a variety of alternative exons in mammalian genes, including the *c-src* N1 exon (Black, 1992; Modafferi and Black, 1997), the calcitonin/CGRP exon 4 (Hedjran *et al.*, 1997), the fibronectin exon IIIb (Huh and Hynes, 1994), the fibroblast growth factor receptor 2 exon IIIb (Baraniak *et al.*, 2003), and the nonmuscle myosin II heavy chain B exon N30 (Kawamoto, 1996). The UGCAUG element

is a key feature within the intronic enhancers of these genes, often occurring in multiple copies. A computational study also suggests that UGCAUG is highly over-represented in the downstream introns of neuron- and muscle-specific alternative cassette exons (Brudno *et al.*, 2001).

The Fox-1 (feminizing locus on X) gene was originally identified in *Caenorhabditis elegans*, where it acts as a numerator element in counting the number of X chromosomes relative to ploidy, and determining male or hermaphrodite development (Hodgkin *et al.*, 1994; Skipper *et al.*, 1999). The worm Fox-1 gene product is thought to post-transcriptionally repress the expression of Xol-1, the main switch controlling sex determination (Nicoll *et al.*, 1997; Meyer, 2000). Since several alternatively spliced isoforms of Xol-1 exist while only one of these splice variants is necessary and sufficient as a sex determinant (Rhind *et al.*, 1995), it was speculated that Fox-1 might lead to unproductive splicing of the Xol-1 gene.

In vitro selection experiments identified the sequence GCAUG as the optimal recognition site for the Fox-1 homolog from zebra fish (zFox1; Jin *et al.*, 2003). zFox-1 mRNA was found to be specifically expressed in muscle, while the mouse Fox-1 protein (mFox-1) was found in muscle, heart, and brain tissue. zFox-1 and mFox-1 were shown to repress muscle-specific exons in nonmuscle tissue and to enhance splicing of the Fibronectin exon EIIIB by binding to GCAUG elements. Others showed that tissue-specific isoforms of mouse Fox-1 proteins differ in terms of subcellular localization and activity as splicing regulators (Nakahata and Kawamoto, 2005). Finally, neuronal isoforms of the Fox-1 protein have been shown to mediate splicing activation via UGCAUG elements, and to control inclusion of certain neuron-specific exons (Nakahata and Kawamoto, 2005; Underwood *et al.*, 2005).

In human, there are three genes that encode Fox-1 like proteins. In the Swissprot database, these proteins are referred to as RNA binding motif protein 9 (RBM9), Ataxin 2-binding protein 1 (A2BP1) and Hexaribonucleotide Binding Protein 1 (HRNBP1) (<http://us.expasy.org/sprot/>). In addition, several alternatively spliced variants exist for each of these Fox-1 like proteins. The RNA binding activity of the Fox-1 proteins is believed to reside in a single ~100 amino acid region with homology to the RNA binding domain (RBD, also called RNA recognition motif (RRM) or ribonucleoprotein (RNP) domain). This domain is conserved among the different human Fox-1 homologs and present in nearly all splice variants. In contrast, the flanking N- and C-terminal domains are not as highly conserved and do not show significant similarity to any protein motifs in current data bases. A typical RBD folds into an $\alpha\beta$ -sandwich with a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology in which a four-stranded antiparallel β -sheet is packed against two α -helices. A single RBD generally recognizes three to four nucleotides of single-stranded RNA sequence-specifically using the β -sheet as the primary RNA binding surface (Maris *et al.*, 2005).

*Corresponding author. Institute for Molecular Biology and Biophysics, Swiss Federal Institute of Technology (ETH) Zurich, 8093 Zurich, Switzerland. Tel.: +41 1 633 39 40; Fax: +41 1 633 12 94; E-mail: allain@mol.biol.ethz.ch

⁵PhD Program for Molecular Life Sciences Zurich, Switzerland

Received: 15 August 2005; accepted: 24 November 2005

To understand how the Fox-1 proteins recognize their target RNA sequence, we have determined the solution structure of the RBD of human Fox-1 in complex with the RNA heptamer UGCAUGU using NMR spectroscopy. Surface plasmon resonance (SPR) analyses give further insight into the mechanism underlying RNA recognition by Fox-1 and support our structural findings.

Results

Structure determination

The RBD of Fox-1 adopts a folded structure both in the presence and absence of RNA and gives rise to highly dispersed NMR spectra (Figure 1A). Titration of the RBD of Fox-1 with an RNA 5'-UGCAUGU-3', followed by NMR, shows that saturation is reached at a 1:1 stoichiometric ratio and that the complex is in slow exchange on the NMR time scale. This RNA was chosen because it contains the Fox-1 binding sequence identified by *in vitro* selection experiments (Jin *et al*, 2003) flanked by two uracils as in the downstream control sequence (DCS) of the *c-src* alternative N1 exon (Modafferi and Black, 1997). Addition of this RNA causes large and numerous chemical shift changes in the ^{15}N -labeled HSQC spectrum, indicating that a large number of protein

residues are perturbed by RNA binding (Figure 1A). A comparison of the chemical shifts of the free and the bound form of the protein shows that the perturbed residues are found in the β -strands and in loops $\beta_1\alpha_1$ and $\beta_2\beta_3$ (Figure 1C). Furthermore, all nucleotides of 5'-UGCAUGU-3', from U₁ to U₇, are affected by binding to the protein as indicated by the overlay of the TOCSY spectra of the free and bound RNAs (Figure 1B).

Complete resonance assignments of the protein in complex could be obtained using published methods. Resonance assignment of the RNA was more difficult and required the synthesis of two isotopically labeled RNA oligonucleotides to resolve ambiguities. In one molecule, the sugar moieties of U₁, C₃ and U₅ were ^{13}C -labeled, in the other molecule, the sugar moieties of G₂, A₄, G₆, and U₇ were ^{13}C -labeled. These two partially labeled RNA molecules were essential to unambiguously assign numerous unusual sugar-sugar and intermolecular NOE cross-peaks. In total, 30 conformers of the Fox-1-UGCAUGU complex were calculated from a total of 1460 NOE-derived distance constraints (including 149 intermolecular and 119 intra-RNA distance constraints), six torsion angle constraints and 29 hydrogen bond constraints (see Table I and Materials and methods). The polypeptide backbone of these structures is ordered from P116 to R194 and the

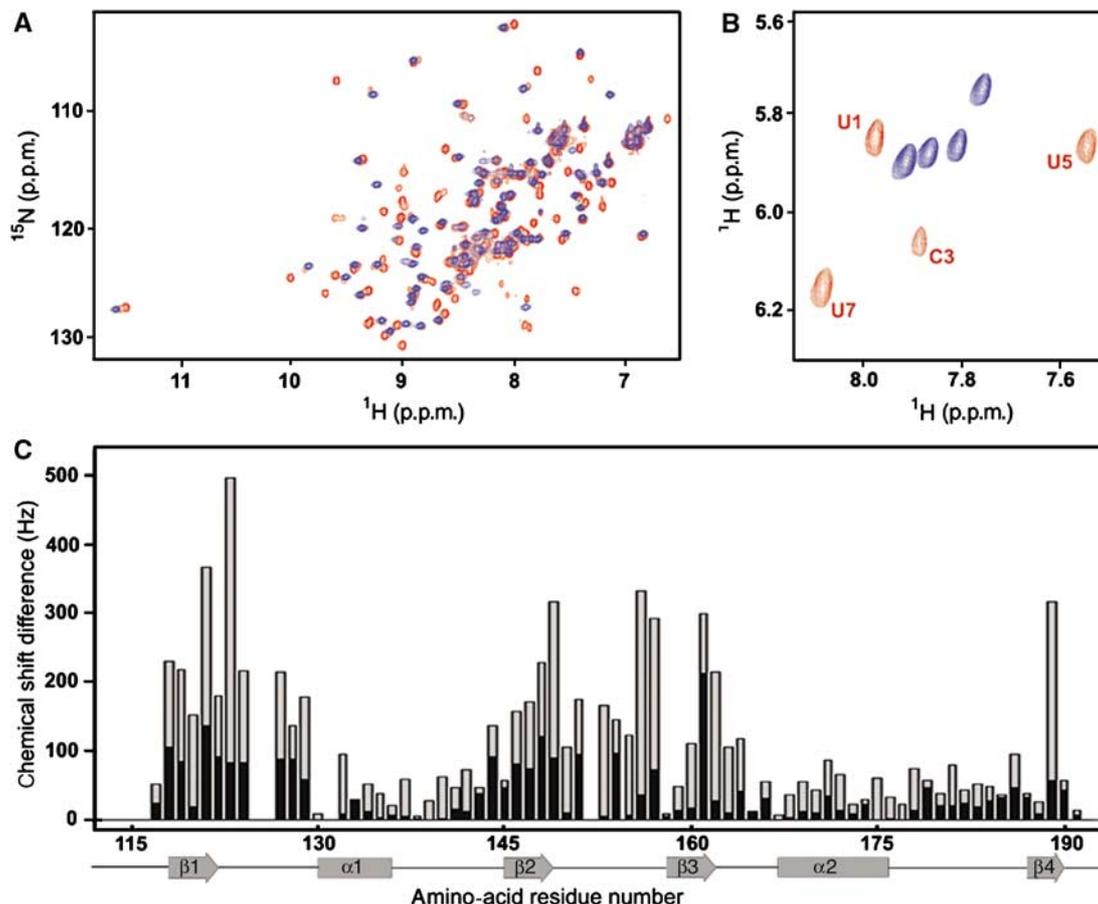


Figure 1 UGCAUGU binds to the RBD of Fox-1 and affects residues in the β -sheet and in loops. (A) ^{15}N -labeled HSQC spectra of ~ 1 mM solutions of the free RBD of Fox-1 (blue) and of the RBD of Fox-1 in the presence of one equivalent of 5'-UGCAUGU-3' (red) at 313 K. (B) Sections of 2D TOCSY spectra showing the H5-H6 correlations of uracil and cytosine of ~ 1 mM solutions of free 5'-UGCAUGU-3' (blue) and of 5'-UGCAUGU-3' in the presence of one equivalent of protein (red). (C) The changes in chemical shift of the backbone amide nitrogen (black) and proton (grey) between free and bound Fox-1 (in Hz, on a 600 MHz spectrometer) are plotted versus the amino-acid residue number. Large chemical shift changes occur in the β -strands as well as in loops $\beta_1\alpha_1$ and $\beta_2\beta_3$ (assignments for residues 125, 126, 131, 152 and after 191 could not be obtained for the free protein).

Table I NMR structure determination statistics

<i>NMR constraints</i>	
Total number of distance constraints	1460
Intermolecular distance constraints	149
RNA intramolecular distance constraints	119
Protein intramolecular distance constraints	1192
Intraresidue	289
Sequential ($ i-j =1$)	295
Medium range ($2 \leq i-j \leq 4$)	170
Long range ($ i-j > 4$)	438
Hydrogen-bond constraints	29
Torsion angle constraints	6
<i>Structure statistics (30 structures)</i>	
Violation statistics	
Average number of NOE violations $>0.2 \text{ \AA}$	4.7 ± 1.7
Maximum NOE violation	0.31 \AA
<i>Energies (kcal mol⁻¹)</i>	
Average distance restraint violation energy	42.7 ± 3.2
Average AMBER energy	-5837.3 ± 16.1
<i>Ramachandran statistics</i>	
Residues in most favored regions	78.6%
Residues in additional allowed regions	18.2%
Residues in generously allowed regions	2.0%
Residues in disallowed regions	1.1%
<i>RMS deviations from ideal covalent geometry</i>	
Bond lengths	0.014 \AA
Bond angles	2.3°
<i>RMS deviations from the mean structure</i>	
Protein backbone: PRO 116–ARG 194	0.42 \AA
Protein heavy atoms: PRO 116–ARG 194	0.95 \AA
RNA heavy atoms: U1–U7	0.55 \AA
<i>Protein + RNA heavy atoms</i>	
PRO 116–ARG 194 and U1–U7	0.90 \AA

conformation of the RNA is precisely defined (Figure 2A). The heavy atoms of the structured part of the entire complex have an RMS deviation of 0.90 \AA (Table I).

Overview of the Fox-1-UGCAUGU complex structure

The protein in the complex adopts the typical $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ fold of an RBD with the two α -helices packed against a four-stranded antiparallel β -sheet (Figure 2). Furthermore, the structure of the protein is characterized by an additional small two-stranded β -sheet located between α_2 and β_4 (Figure 2D).

The RNA, which is unstructured in the free state (data not shown), adopts a bent conformation upon binding to the protein (Figure 2). The RNA bases, rather than the sugar-phosphate backbone, are making most of the contacts to the protein. Three of the seven nucleotides, U₅–U₇, are lying across the canonical binding interface of the RBD, the β -sheet. The remaining four nucleotides (U₁–A₄) are in contact with loops $\beta_1\alpha_1$, $\beta_2\beta_3$ and $\alpha_2\beta_4$. In particular, U₁, G₂ and C₃ are wrapped around a single phenylalanine of the $\beta_1\alpha_1$ -loop, F126. Moreover, G₂ and A₄ form an interesting mismatch base-pair. All the sugars, except for the sugar of U₁, adopt a C2'-endo conformation, and the base of G₆ adopts a *syn* conformation (Figure 2C and D).

Complex formation between the Fox-1 RBD and 5'-UGCAUGU-3' is driven by numerous electrostatic and hydrophobic interactions. Four positively charged side chains, R194, K156, R127, and R184, are in contact with the RNA

phosphate backbone (Figure 2B). Two phenylalanines and one histidine contact the RNA via base stacking. U₁ and G₂ stack on each side of F126, which is part of loop $\beta_1\alpha_1$. U₅ and G₆ stack on H120 and F160, respectively, two residues located on the β -sheet. Additional hydrophobic contacts are seen for the base of C₃ that points its hydrophobic edge towards F126, for the sugars of U₅ and G₆ that pack from both sides against F158, and for the sugar of U₇ that packs against I149 (Figures 2C, D and 3). Hydrophobic interactions equivalent to the ones observed for H120, F158 and F160 were observed in many RBD–RNA complexes (Supplementary Figure S1) (Maris *et al*, 2005). However, the extensive hydrophobic interactions mediated by F126, which contacts U₁, G₂, and C₃ simultaneously, comprise a novel structural feature that is unique to RNA recognition by Fox-1.

Fox-1 is a sequence-specific RNA binding protein

In addition to the numerous hydrophobic and electrostatic interactions that provide affinity, there is a dense network of hydrogen bonds that provide sequence-specificity to the first six nucleotides 5'-UGCAUG-3'. The most important interactions at the protein–RNA interface and within the RNA are described in Supplementary Table SI. Those interactions that are most frequently observed for a certain atom are shown in the middle and right panels of Figure 3. The last nucleotide, U₇, is in contact with the protein as well, and is precisely defined in the structure of the complex (Figure 2A), but its base is not recognized by any specific hydrogen bond.

The recognition of the U₁–C₃ pair is mediated by an intra-RNA hydrogen bond between U₁ O2 and the H42 of C₃ and by two intermolecular hydrogen bonds to the side chains of R127 (U₁) and N151 (C₃) (Figure 3A). However, we do not find a single conformation for this region of the complex (Figure 3A, left panel). In particular, U₁ can be oriented either parallel or perpendicular to F126. This might reflect the physical situation since the NOE intensities of the U₁ H6–H2', H6–H1' and H6–H3' correlations are of similar intensity even at very short mixing times.

G₂ and A₄ form a Trans Watson Crick/Shallow Groove AG base-pair. The guanine is further contacting the backbone carbonyl of I124 with a bifurcated hydrogen bond by its H21 and H1 atoms. Furthermore, the O6 of G₂ is hydrogen bonded to the side chain of R184 (Figure 3B). However, this arginine might also be stacking on G₂, as seen in one third of the structures (Supplementary Table SI).

U₅ is specifically recognized by hydrogen bonds to the backbone amide of T192 and to the backbone carbonyl of N190. Furthermore, in several conformers, the O4 of U₅ is forming a hydrogen bond with the side chain of N189. However, in most structures, the side chain of N189 is hydrogen bonded to H120. Indicated by ¹⁵N HMQC spectra, this histidine is present as the N^{ε2}-H tautomer, and can therefore act as a hydrogen bond acceptor at the N^{δ1} (data not shown, Pelton *et al*, 1993; Drohat *et al*, 1999). The orientation of the N189 side chain is further stabilized by H177, which is also present as the N^{ε2}-H tautomer (Figure 3C).

The base of G₆ is hydrogen bonded to the side chain of R118 and to the backbone carbonyl of T192. R194 is further contacting both the 5'-phosphates of G₆ and U₇. Additionally, two further intra-RNA hydrogen bonds stabilize the RNA structure in the complex: the H22 of G₆ is in contact with

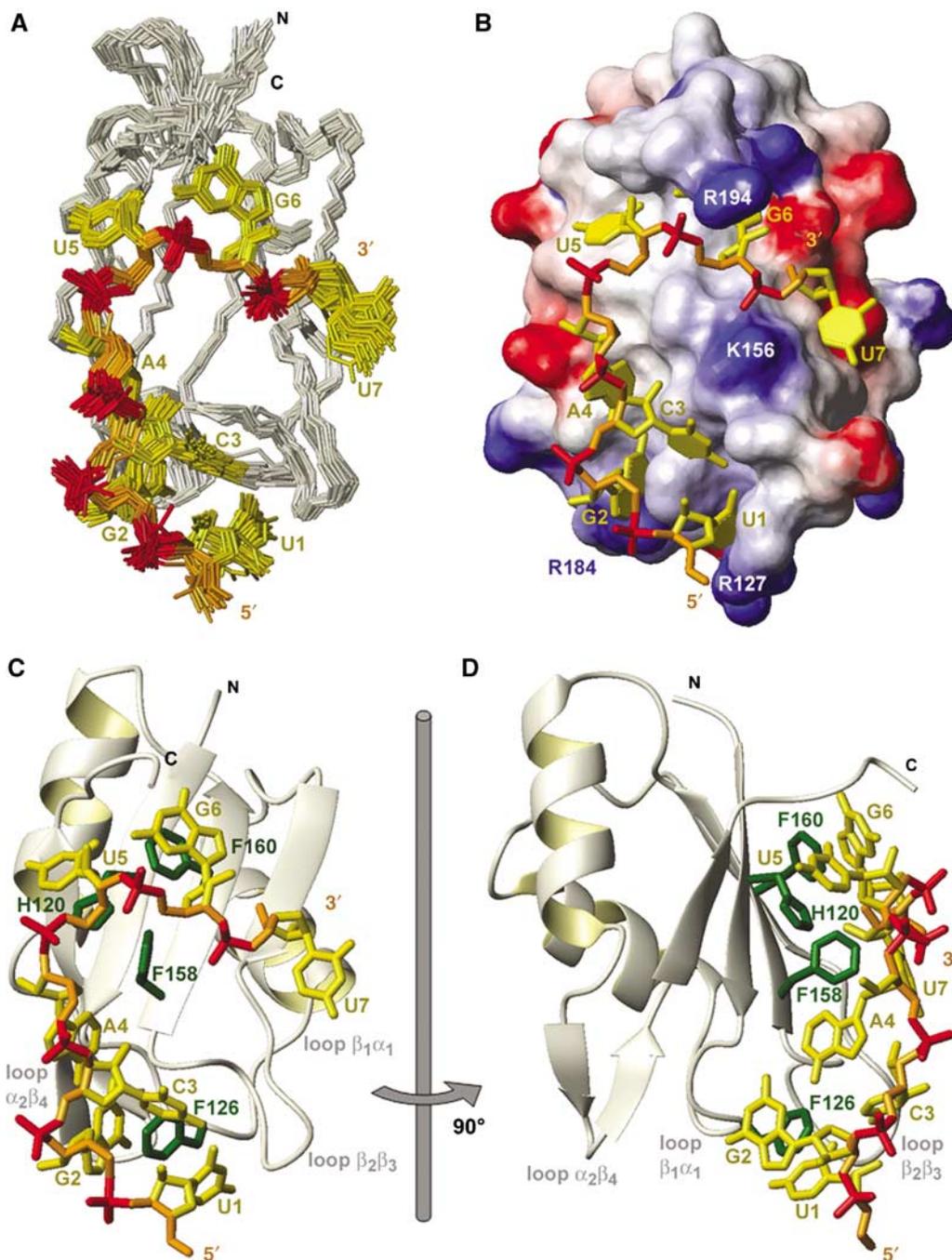


Figure 2 Overview of the solution structure of the RBD of Fox-1 in complex with UGCAUGU. (A) Overlay of the final 30 structures superposed on the heavy atoms of the structured parts of the protein and of the RNA. The protein backbone is gray, the RNA backbone is orange, the phosphate groups are red, and the RNA bases are yellow. Only the ordered region of the protein (residues 116–194) is shown. (B) Surface (heavy atoms of residues 116–194) and stick (heavy atoms of the RNA) representation of the lowest energy structure. The protein surface is painted according to surface potential with red indicating negative charges and blue indicating positive charges. The RNA is colored as in panel (A). (C) The lowest energy structure in ribbon (protein backbone) and stick (RNA) representation. The color scheme is the same as in (A), important protein side chains involved in hydrophobic interactions with the RNA are represented as green sticks. (D) Same as (C) but rotated by 90° around the indicated axis. Figures were generated with MOLMOL (Koradi *et al*, 1996).

the O3' of U₅ and the 2'OH of G₆ is hydrogen bonded to the O3' of U₇ (Figure 3D).

Characterization of the Fox-1–UGCAUGU interactions by surface plasmon resonance

The interaction between Fox-1 and UGCAUGU was further investigated using SPR. In these experiments, an RNA oligonucleotide of the sequence biotin-5'-CUCUGCAUGU-3' was

immobilized on a streptavidin coated chip and binding of Fox-1 to this oligonucleotide was monitored. The affinity of Fox-1 to the immobilized RNA is very high (dissociation constant $K_D = 0.49$ nM at 150 mM NaCl).

To examine in greater detail the electrostatic contribution to Fox-1–RNA complex formation and stability, we performed SPR kinetic analyses at varying salt concentrations. The affinity depends strongly on the salt concentration with

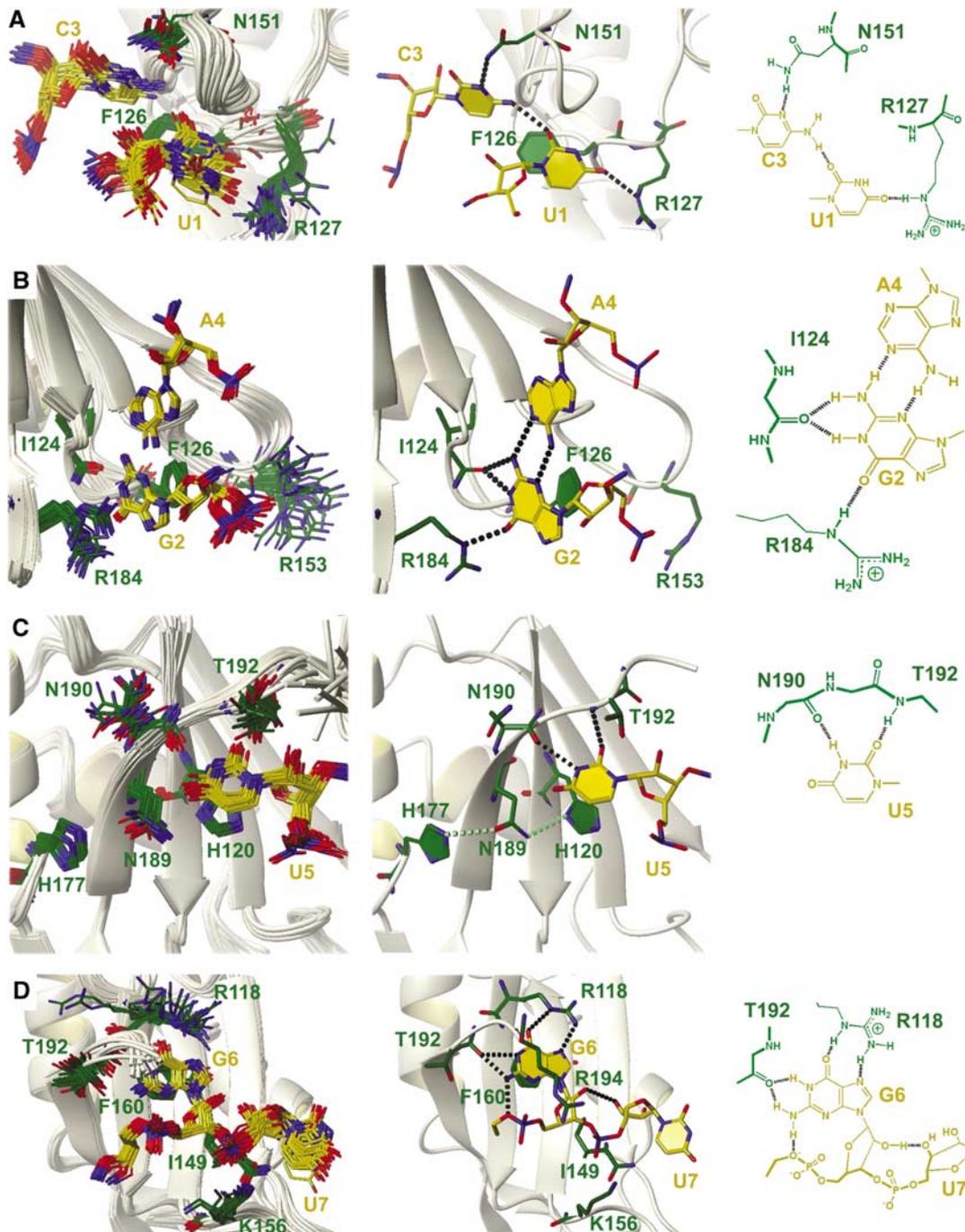


Figure 3 Molecular recognition of UGCAUGU by the RBD of Fox-1. Close-up views of the RNA binding interface of the overlay of the final 30 structures superposed on the heavy atoms of the structured parts of the protein and of the RNA (left), single structures showing the intermolecular and intra-RNA interactions that are most commonly observed in the 30 structures (middle; see Supplementary Table S1) and schematic representations of the hydrogen bonding interactions that are most commonly observed in the 30 structures (right). The ribbon representation of the protein backbone is shown in grey, side chains of the protein are in green and the RNA is in yellow. Recognition of U₁ and C₃ (A), of G₂ and A₄ (B), of U₅ (C), and of G₆ and U₇ (D). Figures were generated with MOLMOL (Koradi *et al*, 1996).

both the association rate constant k_{on} and the dissociation rate constant k_{off} being affected (Figure 4A, Supplementary Table SII). Between 0.075 and 0.6M NaCl, there is a linear relationship between $\log f_{\pm}$ and all three $\log K_{\text{D}}$, $\log k_{\text{off}}$, and $\log k_{\text{on}}$ (Figure 4B) (Debye and Hückel, 1923). This indicates that the activation energy of the rate-limiting step for both association and dissociation is affected by electrostatic interactions, since $\log f_{\pm}$ is proportional to the electrostatic potential of protein and RNA, and $\log k_{\text{on}}$ and $\log k_{\text{off}}$ are

inversely proportional to the activation energy of association and dissociation, respectively.

Even though the structure presented here is very precise (Table I), there are variations in the patterns of hydrogen bonds observed in each structure of the ensemble (Supplementary Table S1). Therefore, it is difficult to tell from the structure alone, which hydrogen bond patterns reflect the physical situation. To characterize important intermolecular and intra-RNA interactions observed in the struc-

ture more precisely, we performed competition experiments with various mutant oligonucleotides. Half maximal inhibitory concentrations (IC_{50} s) were derived for each mutant oligonucleotide. These were compared to the IC_{50} of the immobilized RNA to estimate the energetic contribution of individual interactions to binding (Table II). The binding affinity of the oligonucleotide that was used for structure determination, 5'-UGCAUGU-3', and the immobilized RNA, 5'-CUCUGCAUGU-3', are similar, as can be seen from the nearly identical IC_{50} values. This indicates that the additional three nucleotides of the immobilized RNA do not affect Fox-1 binding. We then tested individual mutations in the competitor RNAs. Replacement of U_1 by either A or C leads to a loss of free binding energy ($\Delta\Delta G$) of 4.0 and 4.5 kJ/mol, respectively. Referring to the structure, each of these mutations results in the loss of one hydrogen bond. Mutating C_3 to U leads to a more dramatic loss of free binding energy ($\Delta\Delta G = 14$ kJ/mol). According to the structure, this mutation

leads to the loss of two hydrogen bonds. The mutations G_2 to A, A_4 to Purine, and A_4 to Inosine lead to $\Delta\Delta G$ values of 15, 5.2 and 13 kJ/mol, respectively. Based on the structure, these replacements should lead to disruption of four, one and two hydrogen bonds, respectively. Replacement of U_5 by C should disrupt one hydrogen bond, and generates a $\Delta\Delta G$ of 3.9 kJ/mol, and replacement of G_6 by A should disrupt four hydrogen bonds, and leads to a binding free energy difference of 19 kJ/mol. The differences in binding energy of the mutant oligonucleotides, compared with the wild-type binding sequence, correlate well with the predicted number of lost hydrogen bonds. The loss of one hydrogen bond gave a $\Delta\Delta G$ between 3.9 and 5.2 kJ/mol (U_1A , U_1C , A_4P , and U_5C). Two predicted lost hydrogen bonds gave a $\Delta\Delta G$ of 13 or 14 kJ/mol (A_4I and C_3U). The G_6A mutation with four predicted hydrogen bonds lost, gave a $\Delta\Delta G$ of 19 kJ/mol. The one inconsistency in the correlation of $\Delta\Delta G$ with lost hydrogen bonds is the G_2A mutation. It leads to a loss of 15 kJ/mol in binding

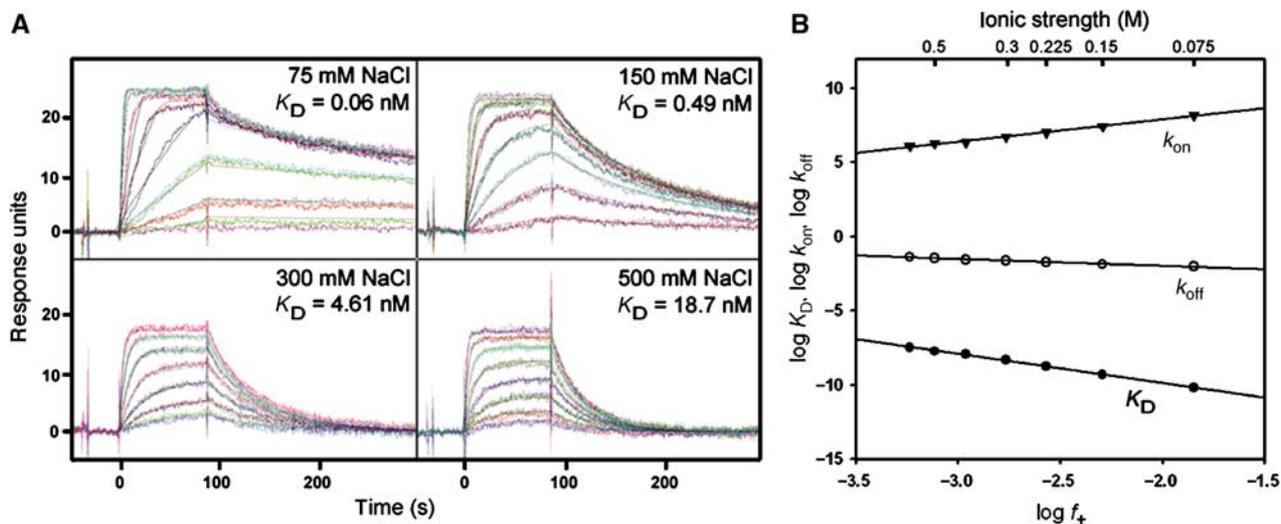


Figure 4 Salt dependence of RNA binding examined by surface plasmon resonance measurements. (A) Representative curves for binding of the RBD of Fox-1 to an immobilized oligonucleotide biotin-5'-CUCUGCAUGU-3' at different salt concentrations. At 75, 150, 300 and 500 mM NaCl, binding curves for 20, 10, 5, 2.5, 1.25, 0.625, 0.312 and 0.156 nM protein, 20, 10, 5, 2.5, 1.25, 0.625 and 0.312 nM protein, 80, 40, 20, 10, 5, 2.5 and 1.25 nM protein, and 320, 160, 80, 40, 20, 10, 5 and 2.5 nM protein are shown, respectively. Curves are fit according to a 1:1 Langmuir interaction model including a correction term for mass transport limitations and are shown as grey lines. (B) Plot of $\log K_D$ (●), $\log k_{on}$ (▲) and $\log k_{off}$ (○) versus $\log f_{\pm}$. f_{\pm} is the electrostatic contribution to the mean rational activity coefficient, which is linked to the ionic strength, see Materials and methods section and Supplementary Table SII. Each data point represents the average of at least three independent measurements.

Table II Surface plasmon resonance studies with mutant oligonucleotides

Oligo	Sequence	IC_{50}^a (nM)	K_D ratio ^b	$\Delta\Delta G$ (kJ/mol) ^c
wt	UGCAUGU	1.60 ± 0.31	0.83	-0.46
CUC-wt	CUCUGCAUGU	1.93 ± 0.21	1	0
U1A	AGCAUGU	9.30 ± 0.60	4.8	4.0
U1C	CGCAUGU	11.8 ± 0.6	6.1	4.5
C3U	UGUAUGU	545 ± 100	280	14
G2A	UACAUGU	677 ± 38	350	15
A4P	UGCPUGU	16.0 ± 1.7	8.3	5.2
A4I	UGCIUGU	392 ± 63	203	13
U5C	UGCACGU	9.40 ± 1.30	4.9	3.9
G6A	UGCAUUAU	3533 ± 586	1830	19

^aValues derived from three independent measurements.

^bThe K_D ratio for binding of the CUC-wt oligonucleotide versus binding of each modified oligonucleotide equals the ratio of IC_{50} s.

^cThe Gibbs free energy difference between a complex of Fox-1-CUC-wt (1) and Fox-1 in complex with the tested oligonucleotide (2) can be calculated as $\Delta\Delta G = -RT \ln K_{D1}/K_{D2} = -RT \ln IC_{50,1}/IC_{50,2}$.

free energy, rather low for the expected loss of four hydrogen bonds. It may be that Arg184 is stacking on G₂ rather than contacting it by a hydrogen bond, as it is seen in about one-third of the NMR structures.

F126 is crucial for the unusual mode of RNA binding by Fox-1

Our structure suggests a critical role for F126 in the unusual binding of the four 5'-nucleotides UGCA. To test the importance of F126, several mutant proteins were prepared, in which F126 was replaced by alanine, histidine, isoleucine, leucine, arginine, tryptophane or tyrosine (Figure 5A). The affinity of UGCAUGU to Fox-1 F126A, Fox-1 F126I and Fox-1 F126R is reduced about 1500-fold ($K_D = 1.62$, 1.62 and 1.58 μM , respectively, at 150 mM NaCl). This effect is comparable to the impact of replacing the RNP consensus residues H120, F158 and F160 by alanine; residues that are known to contribute significantly to RNA binding (Figure 5A). The RNA binding affinity can be almost entirely restored by substituting F126 by a tyrosine and it is only about one order of magnitude less when F126 is substituted by a histidine or a tryptophane, showing that an aromatic residue is critical in this position (Figure 5A). Finally, replacing F126 by leucine gives an intermediate affinity of $3.74 \times 10^{-7} \text{ M}$ (~ 300 fold less), which suggests that hydrophobic packing with a residue that fits sterically can partially substitute for an aromatic side chain.

To further investigate the role of F126 in RNA binding, we recorded a TOCSY spectrum of a 1:1 complex of Fox-1 F126A with UGCAUGU. A comparison of this spectrum with the TOCSY spectrum of a 1:1 complex of wild-type Fox-1 with UGCAUGU shows that the H5-H6 correlations of U5 and U7 are almost in the same position in the spectra of these two complexes, whereas the H5-H6 correlations of U1 and C3 have changed considerably (Figure 5B). This means that U5 and U7 are bound in an analogous way in both complexes, while U1 and C3 are not. Conversely, when removing F160, an aromatic side chain on the β -sheet surface, binding of U1 and C3 is retained, while the H5-H6 crosspeaks of U5 and U7 display very different chemical shifts (Figure 5B).

There are three spectroscopically observable imino protons present in the wild-type Fox-1-UGCAUGU complex, corresponding to the imino of G2, U5 and G6. This indicates that these iminos are engaged in hydrogen bonds (Figure 5C). These iminos give rise to a large number of NOE cross peaks. In the Fox-1 F126A-UGCAUGU complex, the imino groups of U5 and G6 remain observable at almost identical chemical shifts and give rise to the same NOE cross peaks as in the wild-type complex (Figure 5C). In contrast, the imino group of G2 is no longer observable.

Together, these results show that F126 is crucial for the unusual mode of recognition of the four 5'-nucleotides UGCA and that RNA binding by Fox-1 can be divided into two independent parts: a canonical part, mediating the recognition of the 3'-terminal nucleotides via the RNP consensus

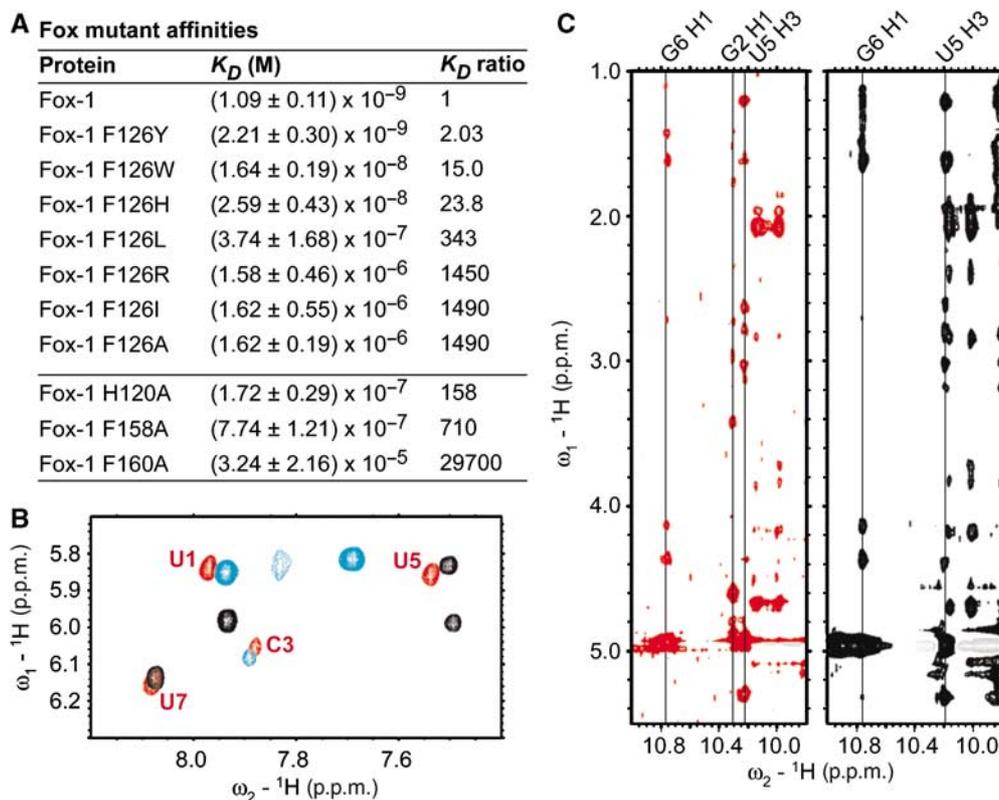


Figure 5 F126 plays a crucial role in RNA binding. (A) Affinities of single amino-acid mutants of Fox-1. Values for K_D s are derived from steady state binding levels at different protein concentrations using surface plasmon resonance. Each measurement was repeated three times at 150 mM NaCl and pH 7.4. (B) Overlay of sections of 2D TOCSY spectra showing the H5-H6 correlations of uracil and cytosine of ~ 1 mM solutions of 5'-UGCAUGU-3' in the presence of one equivalent of Fox-1 (red), Fox-1 F126A (black), and Fox-1 F160A (cyan). (C) Sections of 2D NOESY spectra of a 1:1 complex of Fox-1 (red) or Fox-1 F126A (black) with UGCAUGU showing NOE crosspeaks to the imino protons of G6, G2 and U5.

residues and a novel part that mediates recognition of the 5'-terminal nucleotides and critically depends on F126.

Discussion

The structure shows that the RBD of Fox-1 binds all seven nucleotides of the RNA heptamer UGCAUGU and explains how the first six nucleotides, UGCAUG, are recognized specifically. The structure is in agreement with *in vitro* selection experiments which first identified Fox-1 as a sequence-specific RNA binding protein with specificity to GCAUG (Jin *et al*, 2003). It also confirms the preference for U in the first position of the binding elements seen in studies of its role as an enhancer of alternative splicing (Huh and Hynes, 1994).

The structure of Fox-1 in complex with UGCAUGU demonstrates a novel mode of RNA recognition by the RBD

The structure of the Fox-1-UGCAUGU complex contains several typical attributes. Like all other RBD-RNA or RBD-DNA complexes whose structures have been solved until now (Maris *et al*, 2005), Fox-1 utilizes the β -sheet to bind several nucleotides. In the case of Fox-1, these are U₅, G₆ and U₇ of the UGCAUGU heptamer (Figure 2). Binding affinity to U₅ and G₆ is provided by hydrophobic interactions with three residues within the RNP consensus sequence, F158, F160, and H120. Specificity for U₅ and G₆ is mainly achieved by hydrogen bond interactions between the C-terminus of the domain and functional groups of the bases. These structural features are very similar, for example, to oligonucleotide recognition by the first RBD of hnRNP A1 (Supplementary Figure S1) (Ding *et al*, 1999).

However, there are features that are unique to the mode of recognition of Fox-1. These features mediate the binding of U₁, G₂, C₃ and A₄. Particularly important for RNA binding is the $\beta_1\alpha_1$ loop that contains a phenylalanine, F126 (Figure 2). Three nucleotides, U₁, G₂ and C₃, wrap around this phenylalanine forming a hydrophobic 'cage' around it. The data presented in this study show that this extension of the RNA binding platform of the RBD of Fox-1 is independent from the interactions with the canonical binding site.

A phenylalanine at the position equivalent to F126 of Fox-1 is found in 59 (11%) of the 531 human RBDs published in the Pfam database. In 52 (9.8%) and 23 (4.3%) additional human RBDs, it is exchanged for the similar amino-acids tyrosine and tryptophane, respectively (<http://www.sanger.ac.uk/Software/Pfam/>). Considering the observed amino-acid frequencies in vertebrates, which are 4.0, 3.3 and 1.3% for phenylalanine, tyrosine and tryptophane, respectively, these amino acids are significantly enriched at this position. Another example of a protein that exhibits a phenylalanine at this position of the RBD is the murine mRNA export factor REF2-I. NMR chemical shift mapping experiments have shown that the RBD of this protein contributes to interactions with RNA using loops $\beta_1\alpha_1$ and $\alpha_2\beta_4$, but not via the canonical β -sheet binding interface, with the main RNA binding site located in the flexible N- and C-terminal domains (A Golovanov, G Hautbergue, L-Y Lian, SA Wilson, personal communication, 2005). This implies that this novel feature of RNA recognition is very likely to be shared by many other RBDs. However, a histidine at the equivalent position is found in only 6 (1.1%) human RBDs and is hence under-

represented (average frequency = 2.9%), even though in the case of Fox-1, the F126H mutant has a similar affinity as the F126W mutant.

Another unique feature of RNA recognition by Fox-1 is the unusually high number of intramolecular hydrogen bonds within the bound RNA that are important for sequence specificity and binding affinity. For example, U₁ and C₃ are contacting one another with one hydrogen bond (Figure 3A), G₆ makes hydrogen bond contacts to both U₇ and U₅ (Figure 3D), and most prominently, G₂ and A₄ form a mismatch base pair (Figure 3B) such that A₄ is solely recognized by intra-RNA interactions. Since by NMR analysis, we could observe that the RNA is unstructured in its free form, these interactions are established upon binding to the protein (induced fit) and therefore contribute to complex stability. This is further confirmed by our SPR analysis, where we show that directed disruption of intra-RNA hydrogen bonds leads to a loss of free binding energy. Intra-RNA interactions at the RBD-RNA interface have been observed in other structures of RBD-RNA complexes. However, these were mostly stacking interactions that influence binding affinity but have little impact on sequence specificity (Price *et al*, 1998; Deo *et al*, 1999; Handa *et al*, 1999; Allain *et al*, 2000; Varani *et al*, 2000; Wang and Hall, 2001).

SPR reveals extraordinary affinity, shows the importance of electrostatic interactions for association and confirms the NMR structure

Surface plasmon resonance experiments provided additional insight into the molecular mechanism underlying RNA recognition by Fox-1 and validated the NMR structure (Figure 4, Table II and Supplementary Table SII). The behavior of k_{on} according to the Debye-Hückel theory was shown previously for protein-protein association (Schreiber and Fersht, 1996; Baerga-Ortiz *et al*, 2004) and for an ATPase-ADP/ATP complex (Fedosova *et al*, 2002). Salt dependence of the k_{off} was demonstrated for the N-terminal domain of U1A in complex with the U1 hairpin II (U1hpII) RNA (Katsamba *et al*, 2001). Here, we show that the salt dependence of the k_{off} for a protein-ligand complex follows the Debye-Hückel theory. However, the salt-effect on k_{off} is much weaker than on k_{on} , while k_{on} changes by two orders of magnitude, k_{off} shifts by a factor of about 4 over the concentration range tested. Extrapolation to zero ionic strength, or $\log f_{\pm} = 0$, gives a $k_{on,0}$ of $8.1 \times 10^{10} \text{ M}^{-1} \text{ s}^{-1}$. Since the rate constants for diffusion-limited association for protein-ligand complexes are in the order of 10^5 - $10^6 \text{ M}^{-1} \text{ s}^{-1}$, the rate enhancement due to electrostatic attraction and steering in the Fox-1-UGCAUGU complex at zero ionic strength is about 10^4 - to 10^5 -fold (Berg and vonHippel, 1985). These findings emphasize the role of electrostatic potentials in the initial interaction of Fox-1 with the RNA oligonucleotide. In contrast, the limited effect of the salt concentration on k_{off} suggests that other factors beside short-range electrostatic interactions contribute to the stability of the protein-RNA complex. These findings are in accordance with the net charges of protein (positive) and RNA (negative) and with the structure, where the RNA is engaged in several salt bridges but also in many hydrogen bond and van der Waals contacts with the protein (Supplementary Table SI).

To investigate the energetic contributions of individual intermolecular and intra-RNA hydrogen bonds, mutations

were introduced into the RNA oligonucleotide and binding studies were performed (Table II). These measurements are consistent with the intermolecular and intra-RNA interactions observed in the complex. From our data, it appears that one hydrogen bond will lower the total free energy of the complex by about 4–7 kJ/mol, which is in accordance with the predicted value (Fersht, 1987).

The binding affinity of the Fox-1-UGCAUGU complex with a K_D of 0.49 nM at 150 mM salt is extraordinarily high for a single RBD binding to single-stranded RNA. The N-terminal RBD of U1A was shown to bind with similar affinities to nucleotides exposed in RNA stem-loops. If the U1A binding sequence is present in a single-stranded RNA, the affinity is decreased about 10^4 -fold (Hall, 1994). To achieve nanomolar affinity for single-stranded RNA, most RBD proteins use multiple domains and the high affinity is lost when individual domains are deleted (Zamore *et al*, 1992; Serin *et al*, 1997; Park *et al*, 2000; Sladic *et al*, 2004). In the case of Fox-1, the contacts to three nucleotides provided by a single phenylalanine (F126) of the $\alpha_1\beta_2$ loop in addition to the canonical contacts mediated by the β -sheet surface explain how such a high affinity is reached with a single RBD. Therefore, we asked whether an aromatic residue in the position equivalent to F126 occurs more often in RNA binding proteins containing only one RBD. We analyzed 159 single-RBD proteins published in the Pfam database and found similar frequencies for phenylalanine (11%), tyrosine (11%), tryptophan (3.1%), and histidine (1.9%) as for the full set of 531 human RBDs. Hence, this novel kind of interaction does not seem to be generally employed to substitute for further RBDs. However, evolutionary pressure does not necessarily favor high affinity and multi-RBD proteins could be employed to recognize the distribution of specific binding sites in addition to the sequences themselves.

Implications for alternative splicing regulation

The RBDs of human and *C. elegans* Fox-1 are 75% identical. In zebra fish and mouse, the conservation is even higher (<http://us.expasy.org/sprot/>). Moreover, all the residues that are in contact with the RNA are conserved, including not just the side chains involved in direct stacking, electrostatic and hydrogen bond contacts with the RNA, but the complete side of the protein that is facing the RNA. This suggests that the mode of RNA recognition, and in particular the binding sequence specificity of Fox-1, is conserved from *C. elegans* to human.

The Fox-1 binding sequence, UGCAUGU, is a key element for the regulation of alternative splicing (Huh and Hynes, 1994; Hedjran *et al*, 1997; Modafferi and Black, 1997; Lim and Sharp, 1998; Brudno *et al*, 2001; Baraniak *et al*, 2003; Jin *et al*, 2003). The structure of Fox-1 in complex with UGCAUGU is particularly interesting for understanding this function of the protein. As shown in Figure 2, Fox-1 induces a curvature in the RNA upon binding. Therefore, the binding of Fox-1 to its RNA targets might lead to conformational changes in the RNA that in turn influence splicing regulation. Another possible role for Fox-1 in splicing regulation could be to compete with other splicing factors for the same or overlapping binding sites. The high affinity of Fox-1 determined by SPR indicates that Fox-1 could be an efficient competitor for binding sites on pre-mRNAs. However, tests of Fox-1 activity on model substrates indicate that the protein can activate splicing from a multimerized UGCAUG element. This

activation is independent of other binding elements (Underwood *et al*, 2005). Thus, Fox-1 can apparently activate splicing and not just release an exon from repression by other proteins.

Materials and methods

Protein and RNA preparation

DNA encoding the RBD of Fox-1 (residues 109–208, Swissprot Q9NWB1) was isolated by PCR amplification from a full-length Fox-1 cDNA clone and cloned into pET28a (N-terminal His-tag). The protein was expressed in transformed BL21(DE3) *Escherichia coli* at 37°C in minimal medium M9 containing 1 g l^{-1} $^{15}\text{N-NH}_4\text{Cl}$ and 4 g l^{-1} glucose (for ^{15}N -labeled proteins) or 1 g l^{-1} $^{15}\text{N-NH}_4\text{Cl}$ and 2 g l^{-1} ^{13}C -glucose (for ^{15}N - and ^{13}C -labeled proteins) and 50 mg l^{-1} kanamycin. Cells were grown to $\text{OD}_{600}\approx 0.6$ and induced with 1 mM IPTG. Cells were harvested after 4 h by centrifugation. Cells were resuspended in 20 ml lysis buffer per litre of culture medium (300 mM NaCl, 50 mM Na_2HPO_4 , pH 8.0, 0.002% (v/v) SUPERase RNase inhibitor (Ambion Inc.)) containing 10 mM Imidazole and were lysed by two passages through a cell cracker (Avestin Inc.). The cell lysate was centrifuged at 20 000 g and the supernatant was incubated with NiNTA beads for > 1 h. After washing with lysis buffer, the protein was eluted with a step gradient of imidazole (20–500 mM). The purest fractions as judged by 18% SDS-PAGE were subjected to a second identical NiNTA affinity chromatography. Pure fractions were dialyzed against 51 NMR buffer (20 mM NaCl, 10 mM NaH_2PO_4 , pH 6.5). The protein was concentrated to $\sim 1\text{ mM}$ by centrifugation at 4°C using a 5 kDa molecular mass cutoff membrane. The identity of the protein was confirmed by MALDI MS and N-terminal Edman sequencing. The yield of purified Fox-1 was $\sim 10\text{ mg l}^{-1}$ of culture medium. Protein mutagenesis was carried out following the instructions given by the manufacturer (Quick Change Site-Directed Mutagenesis Kit, Stratagene). All unlabeled RNA oligonucleotides were purchased from Dharmacon Research, deprotected according to the instructions by the manufacturer, desalted using a G-15 size exclusion column (Amersham), lyophilized and resuspended in NMR buffer (20 mM NaCl, 10 mM NaH_2PO_4 , pH 6.5) or water. Oligos of the sequence 5'-UGCAUGU-3' with ^{13}C -labeled sugars of U₁, C₃ and U₅ or ^{13}C -labeled sugars of G₂, A₄, G₆ and U₇ were chemically synthesized by LR and SP (manuscript in preparation).

NMR measurements and resonance assignments

NMR spectra were recorded at 313 K on Bruker DRX-500, DRX-600 and Avance 900 spectrometers. Data was processed with XWINNMR (Bruker) and analyzed with Sparky (<http://www.cgl.ucsf.edu/home/sparky/>). Protein backbone ^1H and ^{15}N resonance assignments for the free protein were obtained using HNCA, HN(CO)CA (Grzesiek and Bax, 1992) and CBCA(CO)NH (Grzesiek and Bax, 1993) spectra acquired on a ^{15}N , ^{13}C labeled protein in 90% H_2O , 10% $^2\text{H}_2\text{O}$. Complete protein backbone ^1H , ^{15}N and ^{13}C resonance assignments of the complex were obtained for residues 116–196 using HNCA and CBCA(CO)NH spectra acquired on a ^{15}N , ^{13}C labeled protein in complex with unlabeled 5'-UGCAUGU-3' in 90% H_2O , 10% $^2\text{H}_2\text{O}$. Aliphatic side chain assignments were obtained from H(C)CH-TOCSY (Bax *et al*, 1990), 3D ^{15}N and ^{13}C NOESY-HSQC ($\tau_m = 150\text{ ms}$) (Talluri and Wagner, 1996; Baur *et al*, 1998) and ^{15}N and ^{13}C HSQC experiments (Susumu Mori *et al*, 1995). Aromatic side chains were assigned using 2D TOCSY ($\tau_m = 50\text{ ms}$) (Bax and Davis, 1985) and 2D NOESY ($\tau_m = 150\text{ ms}$) (Wider *et al*, 1984) spectra. Resonance assignments of the RNA were obtained from 2D NOESY, 2D TOCSY and natural abundance ^{13}C HSQC experiments of ^{15}N -labeled protein in complex with unlabeled RNA and confirmed by ^{13}C HSQC spectra recorded with ^{15}N -labeled protein in complex with 5'-UGCAUGU-3' with ^{13}C -labeled sugars of U₁, C₃ and U₅ or ^{13}C -labeled sugars of G₂, A₄, G₆ and U₇.

Experimental restraints

Seven cycles of CANDID and DYANA (Guntert *et al*, 1997; Herrmann *et al*, 2002) were run to yield a list of automatically assigned intramolecular protein NOE distance constraints. This calculation included peak lists from 3D ^{15}N - and ^{13}C -edited NOESY-HSQC ($\tau_m = 150\text{ ms}$) and 2D NOESY ($\tau_m = 150\text{ ms}$) experiments.

The automatically generated list was reviewed and combined with a list containing manually assigned intra-RNA, intermolecular and additional intra-protein NOE distance restraints. These constraints were derived from 3D ¹⁵N- and ¹³C-edited NOESY-HSQC, 2D NOESY, 2D F1-edited, F2-edited NOESY, 2D F1-filtered, F2-edited NOESY (Peterson *et al*, 2004) and 3D ¹³C F1-filtered, F3-edited NOESY-HSQC (Zwahlen *et al*, 1997) spectra on a complex of ¹³C, ¹⁵N-labeled protein and unlabeled RNA, as well as from 2D F1-filtered, F2-edited NOESY spectra of complexes of ¹⁵N-labeled protein and RNA having ¹³C-labeled sugars of either U₁, C₃, U₅ or G₂, A₄, G₆, U₇. To exclude that critical NOE cross peaks arose from spin diffusion, a 2D NOESY with a short mixing time (30 ms) was recorded and critical NOE restraints were reviewed. NOE cross peaks to the imino protons of G₂, U₅ and G₆ could be observed at 293 K in a 2D NOESY spectrum. Dihedral angle constraints for the sugars of G₂-U₇ ($130^\circ \leq \delta \leq 190^\circ$, i.e., C2'-endo) were added based on high H1'-H2' cross-peak intensities in the 2D TOCSY experiment. In total, 26 intra-protein hydrogen bond constraints were based on slow exchanging amides (¹⁵N-HSQC after ~3 h in ²H₂O at 40°C), typical α shifts and NOE cross-peak patterns typical for secondary structure elements; 3 intermolecular hydrogen bond constraints were based on observable imino protons of G₂, U₅ and G₆ and careful analysis of local NOE cross-peaks. The tautomeric state of His130 and His187 was determined from ¹⁵N HMQC spectra (Pelton *et al*, 1993; Drohat *et al*, 1999). Distance restraints were calibrated using cross-peak intensities corresponding to fixed inter-atomic distances and were assigned upper distance limits of 3.0 (strong), 4.5 (medium) and 6.0 Å (weak) and lower distance limits of 1.8 Å.

Structure calculation

With the final set of constraints, a total of 100 structures of the complex were generated in DYANA (Guntert *et al*, 1997) starting from random structures. The 30 structures with the lowest target function were refined in a restrained simulated annealing run in implicit solvent (generalized Born solvation model (Tsui and Case, 2000)) in the SANDER module of AMBER 7.0 (Pearlman *et al*, 1995) using the Cornell *et al* (1995) force field. The same simulated annealing protocol as described by Padrta *et al* (2002) was used, except that the system was heated to 1500 K and that the time constant for heat bath coupling (TAUTP) was gradually decreased from 0.1 to 0.05 ps during the last picosecond of simulation. The final structures were analyzed with PROCHECK (Laskowski *et al*, 1996).

Surface plasmon resonance

Analyses were carried out using a BIAcore 3000 instrument. All experiments were performed at 25°C using HBS (10 mM HEPES, 150 mM NaCl, 3.4 mM EDTA, 0.005% P20, pH 7.4) as running buffer. When required, NaCl concentration and pH have been adjusted to 75, 125, 225, 300, 400, 500, 600 mM and to 6.0 and 8.5, respectively. For kinetic studies, 5–8 RU of 5' biotinylated CUCUGCAUGU were captured on an SA-chip (BIAcore). Background noise and unspecific binding were corrected using an untreated surface as control surface. Binding studies were carried out injecting serial dilutions of Fox-1 at a flow rate of 70 μ l min⁻¹ for

90 s over the specific and reference surfaces. Protein samples were injected for three times in random order. At the end of each cycle, surfaces were washed with three consecutive 1 min-injections of 1 M NaCl. The reported mean values were derived from at least three independent experiments. Data were globally fit to a simple 1:1 Langmuir interaction model with a correction for mass transport using BIA evaluation software 3.1. Most mutant Fox-1 proteins displayed unfavorable kinetics for SPR kinetic analyses, such that k_{on} and k_{off} could not be reliably determined from curve fits. Therefore, affinity constants of mutant proteins were derived from steady-state binding levels at different protein concentrations using a chip surface coated with ~10 RU of biotinylated RNA and longer association times. For the inhibition assays, 20 RU of biotinylated oligonucleotide were captured on the SA chip. Fox-1 at 2 nM was incubated with different concentrations of mutant oligonucleotides. Solutions were injected for 2 min at a flow rate of 20 μ l/min over the specific and reference surface. Surfaces were regenerated with three 1 min injections of 1 M NaCl. Inhibition curves were obtained by monitoring the decrease of binding response upon increase of oligonucleotide concentration. Values for half-maximal inhibition (IC₅₀) were calculated from fitting curves. Each inhibition assay was carried out in triplicate.

Calculations

According to the Debye and Hückel (1923) theory, $\log f_{\pm}$ is related to the ionic strength as

$$\log f_{\pm} = \frac{A|z_1z_2|\sqrt{I}}{1 + Ba\sqrt{I}}$$

where I is the ionic strength of the solution, $|z_1z_2|$ is the charge product of protein and ligand, $A = 0.512 \text{ M}^{-1/2}$, $B = 0.329 \times 10^8 \text{ M}^{-1/2} \text{ cm}^{-1}$ (Robinson and Stokes, 2002) and a is an adjustable parameter and gave best fits as $a = 5.6 \text{ \AA}$.

Structural data. All restraints used in structure determination and the derived atomic coordinates for the 30 final structures have been deposited at the Protein Data Bank with accession code 2ERR.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online.

Acknowledgements

We would like to thank Dr Richard Stefl for help with computational issues, Goran Malojčić for invaluable discussions and Abdelhamid Benattallah for backbone assignment of the free protein. We would like to thank Professor John A Robinson for helpful support and A Golovanov for communicating unpublished data. We acknowledge the Functional Genomics Center Zurich, especially Mike Scott, for providing explanations on and access to the Biacore instrument. This investigation was supported by grants from the Swiss National Foundation, National Center for Competence in Research Structural Biology to FHTA and SP and by the Roche Research Fund for Biology at the ETHZ to FHTA. FHTA is an EMBO Young Investigator.

References

- Allain F, Bouvet P, Dieckmann T, Feigon J (2000) Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J* **19**: 6870–6881
- Baerga-Ortiz A, Bergqvist S, Mandell J, Komives E (2004) Two different proteins that compete for binding to thrombin have opposite kinetic and thermodynamic profiles. *Protein Sci* **13**: 166–176
- Baraniak A, Lasda E, Wagner E, Garcia-Blanco M (2003) A stem structure in fibroblast growth factor receptor 2 transcripts mediates cell-type-specific splicing by approximating intronic control elements. *Mol Cell Biol* **23**: 9327–9337
- Baur M, Gemmecker G, Kessler H (1998) C-13-NOESY-HSQC with split carbon evolution for increased resolution with uniformly labeled proteins. *J Magn Reson* **132**: 191–196
- Bax A, Clore G, Gronenborn A (1990) H-1-H-1 correlation via isotropic mixing of C-13 magnetization, a new 3-dimensional approach for assigning H-1 and C-13 spectra of C-13 enriched proteins. *J Magn Reson* **88**: 425–431
- Bax A, Davis D (1985) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J Magn Reson* **65**: 355–360
- Berg O, vonHippel P (1985) Diffusion-controlled macromolecular interactions. *Annu Rev Biophys Bio* **14**: 131–160
- Black D (1992) Activation of c-src neuron-specific splicing by an unusual RNA element *in vivo* and *in vitro*. *Cell* **69**: 795–807
- Budno M, Gelfand M, Spengler S, Zorn M, Dubchak I, Conboy J (2001) Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing. *Nucleic Acids Res* **29**: 2338–2348
- Cornell W, Cieplak P, Bayly C, Gould I, Merz K, Ferguson D, Spellmeyer D, Fox T, Caldwell J, Kollman P (1995) A second

- generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* **117**: 5179–5197
- Debye P, Hückel E (1923) Zur Theorie der Elektrolyte. *Phys Z* **24**: 185–206
- Deo R, Bonanno J, Sonenberg N, Burley S (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**: 835–845
- Ding J, Hayashi MK, Zhang Y, Manche L, Krainer AR, Xu RM (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev* **13**: 1102–1115
- Drohac AC, Xiao G, Tordova M, Jagadeesh J, Pankiewicz KW, Watanabe KA, Gilliland GL, Stivers JT (1999) Heteronuclear NMR and crystallographic studies of wild-type and H187Q *Escherichia coli* uracil DNA glycosylase: electrophilic catalysis of uracil expulsion by a neutral histidine 187. *Biochemistry* **38**: 11876–11886
- Fedosova N, Champeil P, Esmann M (2002) Nucleotide binding to Na₄K-ATPase: the role of electrostatic interactions. *Biochemistry* **41**: 1267–1273
- Fersht AR (1987) The hydrogen-bond in molecular recognition. *Trends in Bioc Sci* **12**: 301–304
- Grzesiek S, Bax A (1992) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J Magn Reson* **96**: 432–440
- Grzesiek S, Bax A (1993) Amino-acid type determination in the sequential assignment procedure of uniform C-13/N-15 enriched proteins. *J Biomol NMR* **3**: 185–204
- Guntert P, Mumenthaler C, Wuthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* **273**: 283–298
- Hall K (1994) Interaction of RNA hairpins with the human U1A N-terminal RNA-binding domain. *Biochemistry* **33**: 10076–10088
- Handa N, Nureki O, Kurimoto K, Kim I, Sakamoto H, Shimura Y, Muto Y, Yokoyama S (1999) Structural basis for recognition of the tra mRNA precursor by the sex-lethal protein. *Nature* **398**: 579–585
- Hedjran F, Yeakley J, Huh G, Haynes R, Rosenfeld M (1997) Control of alternative pre-mRNA splicing by distributed pentameric repeats. *Proc Natl Acad Sci USA* **94**: 12343–12347
- Herrmann T, Guntert P, Wuthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* **319**: 209–227
- Hodgkin J, Zellan J, Albertson D (1994) Identification of a candidate primary sex determination locus, fox-1, on the X-chromosome of *Caenorhabditis elegans*. *Development* **120**: 3681–3689
- Huh G, Hynes R (1994) Regulation of alternative pre-messenger-RNA splicing by a novel repeated hexanucleotide element. *Gene Dev* **8**: 1561–1574
- Jin Y, Suzuki H, Maegawa S, Endo H, Sugano S, Hashimoto K, Yasuda K, Inoue K (2003) A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *EMBO J* **22**: 905–912
- Katsamba P, Myszkka D, Laird-Offringa I (2001) Two functionally distinct steps mediate high affinity binding of U1A protein to U1 hairpin II RNA. *J Biol Chem* **276**: 21476–21481
- Kawamoto S (1996) Neuron-specific alternative splicing of non-muscle myosin II heavy chain-B pre-mRNA requires a cis-acting intron sequence. *J Biol Chem* **271**: 17613–17616
- Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**: 51–55
- Laskowski R, Rullmann J, MacArthur M, Kaptein R, Thornton J (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* **8**: 477–486
- Lim L, Sharp P (1998) Alternative splicing of the fibronectin EIIIB exon depends on specific TGCATG repeats. *Mol Cell Biol* **18**: 3900–3906
- Maris C, Dominguez C, Allain FH-T (2005) The RNA recognition motif, a modulable RNA binding platform to regulate posttranscriptional gene expression. *FEBS J* **272**: 2118–2131
- Meyer B (2000) Sex in the worm – counting and compensating X-chromosome dose. *Trends Genet* **16**: 247–253
- Modafferi E, Black D (1997) A complex intronic splicing enhancer from the c-src pre-mRNA activates inclusion of a heterologous exon. *Mol Cell Biol* **17**: 6537–6545
- Mori S, Abeygunawardana C, Johnson MO, van Zijl PCM (1995) Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J Magn Reson Ser B* **108**: 94–98
- Nakahata S, Kawamoto S (2005) Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. *Nucleic Acids Res* **33**: 2078–2089
- Nicoll M, Akerib C, Meyer B (1997) X-chromosome-counting mechanisms that determine nematode sex. *Nature* **388**: 200–204
- Padrta P, Stefl R, Kralik L, Idek L, Sklenar V (2002) Refinement of d(GCGAAGC) hairpin structure using one- and two-bond residual dipolar couplings. *J Biomol NMR* **24**: 1–14
- Park S, Myszkka D, Yu M, Littler S, Laird-Offringa I (2000) HuD RNA recognition motifs play distinct roles in the formation of a stable complex with AU-rich RNA. *Mol Cell Biol* **20**: 4765–4772
- Pearlman D, Case D, Caldwell J, Ross W, Cheatham T, Debolt S, Ferguson D, Seibel G, Kollman P (1995) Amber, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. *Comput Phys Commun* **91**: 1–41
- Pelton JG, Torchia DA, Meadow ND, Roseman S (1993) Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIIGlc, a signal-transducing protein from *Escherichia coli*, using two-dimensional heteronuclear NMR techniques. *Protein Sci* **2**: 543–558
- Peterson RD, Theimer CA, Wu H, Feigon J (2004) New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. *J Biomol NMR* **28**: 59–67
- Price S, Evens P, Nagai K (1998) Crystal structure of the spliceosomal U2B''-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**: 645–650
- Rhind N, Miller L, Kopczynski J, Meyer B (1995) Xol-1 acts as an early switch in the *C.elegans* male hermaphrodite decision. *Cell* **80**: 71–82
- Robinson R, Stokes R (2002) *Electrolyte Solutions*. Mineola, New York, USA: Dover Publications
- Schreiber G, Fersht A (1996) Rapid, electrostatically assisted association of proteins. *Nat Struct Biol* **3**: 427–431
- Serin G, Joseph G, Ghisolfi L, Bauzan M, Erard M, Amalric F, Bouvet P (1997) Two RNA-binding domains determine the RNA-binding specificity of nucleolin. *J Biol Chem* **272**: 13109–13116
- Skipper M, Milne CA, Hodgkin J (1999) Genetic and molecular analysis of fox-1, a numerator element involved in *Caenorhabditis elegans* primary sex determination. *Genetics* **151**: 617–631
- Sladic R, Lagnado C, Bagley C, Goodall G (2004) Human PABP binds AU-rich RNA via RNA-binding domains 3 and 4. *Eur J Biochem* **271**: 450–457
- Talluri S, Wagner G (1996) An optimized 3D NOESY-HSQC. *J Magn Reson Ser B* **112**: 200–205
- Tsui V, Case D (2000) Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* **56**: 275–291
- Underwood JG, Boutz PL, Dougherty JD, Stoilov P, Black DL (2005) Homologues of the *C. elegans* Fox-1 protein are neuronal splicing regulators in mammals. *Mol Cell Biol* **25**: 10005–10016
- Varani L, Gunderson S, Mattaj J, Kay L, Neuhaus D, Varani G (2000) The NMR structure of the 38 kDa U1A protein-PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. *Nat Struct Biol* **7**: 329–335
- Wang X, Hall T (2001) Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat Struct Biol* **8**: 141–145
- Wider G, Macura S, Kumar A, Ernst R, Wüthrich K (1984) Homonuclear two-dimensional H-1-NMR of proteins – experimental procedures. *J Magn Reson* **56**: 207–234
- Zamore P, Patton J, Green M (1992) Cloning and domain-structure of the mammalian splicing factor U2AF. *Nature* **355**: 609–614
- Zwahlen C, Legault P, Vincent S, Greenblatt J, Konrat R, Kay L (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage lambda N-peptide/boxB RNA complex. *J Am Chem Soc* **119**: 6711–6721