Two structurally different RNA molecules are bound by the spliceosomal protein U1A using the same recognition strategy

Luca Jovine, Chris Oubridge, Johanna M Avis† and Kiyoshi Nagai*

Background: Human U1A protein binds to hairpin II of U1 small nuclear RNA (snRNA) and, together with other proteins, forms the U1 snRNP essential in pre-mRNA splicing. U1A protein also binds to the 3′ untranslated region (3′UTR) of its own pre-mRNA, inhibiting polyadenylation of the 3′ end and thereby downregulating its own expression. The 3′UTR folds into an evolutionarily conserved secondary structure with two internal loops; one loop contains the sequence AUUGCAC and the other its variant AUUG*U*AC. The sequence AUUGCAC is also found in hairpin II of U1 snRNA; hence, U1A protein recognizes the same heptanucleotide sequence in two different structural contexts. In order to better understand the control mechanism of the polyadenylation process, we have built a model of the U1A protein–3′UTR complex based on the crystal structure of the U1A protein–hairpin II RNA complex which we determined previously.

Results: In the crystal structure of the U1A protein–hairpin II RNA complex the AUUGCAC sequence fits tightly into a groove on the surface of U1A protein. The conservation of the heptanucleotide in the 3′UTR strongly suggests that U1A protein forms identical sequence-specific contacts with the heptanucleotide sequence when complexed with the 3′UTR. The crystal structure of the hairpin II complex and the twofold symmetry in the 3′UTR RNA provide sufficient information to restrict the conformation of the 3′UTR RNA and have enabled us to build a model of the 3′UTR complex .

Conclusions: In the U1A–3′UTR complex, sequence-specific interactions are made entirely by the conserved heptanucleotide and the last base pair (C:G) of the stem. The structure is stabilized by protein–protein contacts and by electrostatic interactions between basic amino acids of the protein and the phosphate backbone of the RNA stem regions. The formation of a protein dimer necessary for the inhibition of poly(A) polymerase requires a conformational change of the C termini of the proteins upon RNA binding. This mechanism could prevent the inhibition of poly(A) polymerase by free U1A protein. The model is consistent with biochemical data, and the protein–protein interactions within the 3′UTR complex account for the cooperativity of U1A protein binding to the 3′UTR. The model also serves as an important structural guide for designing further experiments to understand the interaction between the U1A–3′UTR complex and poly(A) polymerase.

Address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

†Present address: Department of Biochemistry and Applied Molecular Biology, UMIST, Manchester M60 1QD, UK.

*Corresponding author.
E-mail: kn@mrc-lmb.cam.ac.uk

Key words: molecular modelling, polyadenylation, post-transcriptional regulation, RNA–protein recognition, U1A spliceosomal protein, 3′ untranslated region

Received: 22 Jan 1996
Revisions requested: 19 Feb 1996
Revisions received: 4 Mar 1996
Accepted: 5 Mar 1996

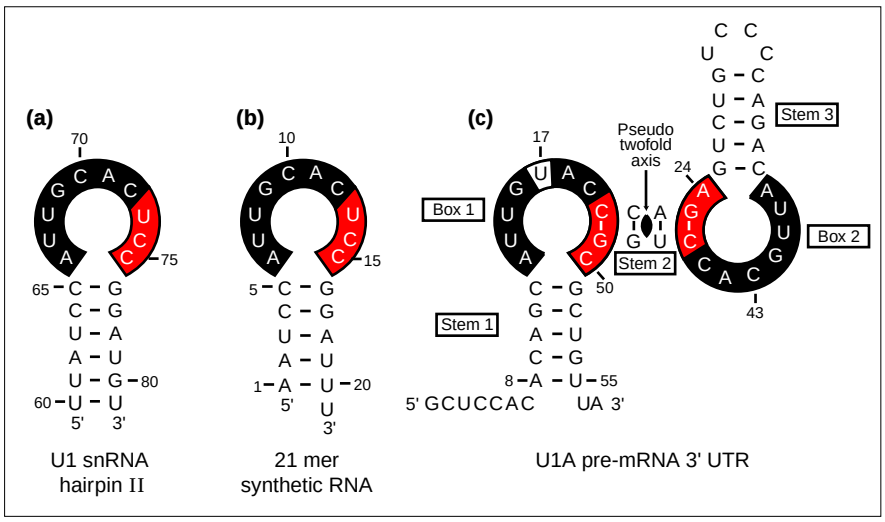
Structure 15 May 1996, 4:621–631

 ISSN 0969-2126

Introduction

U1A spliceosomal protein exists in the nucleus as a component of U1 snRNP (small nuclear ribonucleoprotein), a large RNA–protein complex involved in pre-mRNA splicing [1]. U1A protein binds to hairpin II of U1 snRNA, which has a ten nucleotide AUUGCACUCC loop (Fig. 1a) [2–4]. U1A protein consists of 282 residues and contains two RNP domains, one at the N terminus and one at the C terminus, which are connected by a protease sensitive linker [5]. The RNP domain is a commonly occurring RNA-binding domain of 80–90 amino acids [6,7]. Scherly *et al.* [8] showed that a fragment of U1A protein containing the first 102 residues, which contains one of the two RNP domains, binds to hairpin II of U1 snRNA with the same affinity as the full length protein. The structure of a fragment of U1A protein containing residues 2–95 was solved by X-ray crystallography [9]. This showed that the RNP motif is folded into a globular domain consisting of a four stranded β sheet flanked on one side by two α helices. A subsequent NMR study, using a longer fragment of U1A protein (residues 2–102), showed that an additional α helix (helix C) is present around residue 95 but did not establish its position in relation to the RNP domain [10].

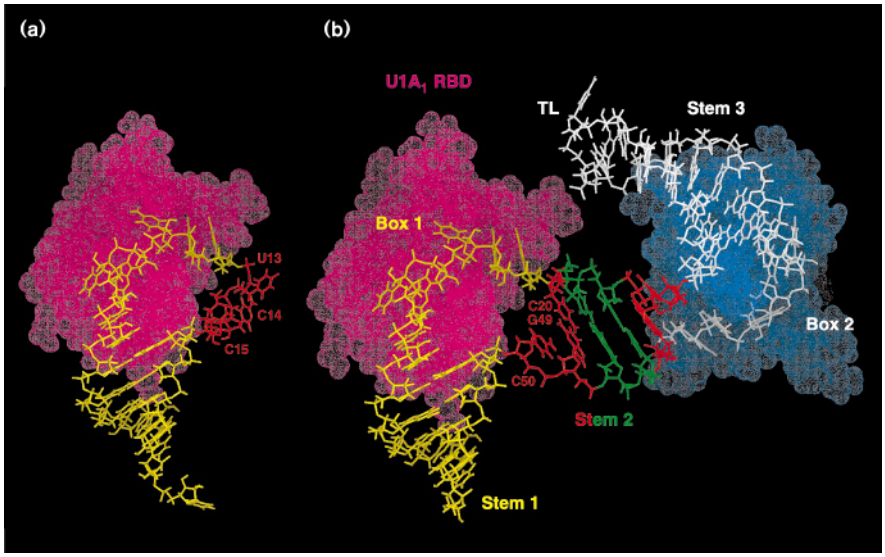
Figure 1



Diagrammatic representations of the sequence and secondary structure of U1A binding sites. In each case the conserved seven nucleotide sequence is shown as white on black, and the three nucleotides which make no specific contacts with U1A are shown as white on red. (a) Hairpin II of U1 snRNA. (b) The synthetic 21 nucleotide hairpin used in co-crystallization with U1A RBD. The last three nucleotides, UCC, in the loops of hairpins (a,b) make no contacts with the protein in the co-crystal structure and are replaced in (c) the 3′UTR of U1A pre-mRNA, by the terminal base pairs of Stem 2 and the unpaired nucleotides A24 and C50. The variant nucleotide in the loop of Box 1, U17, is shown boxed as black on white. The pseudo twofold axis of secondary structure symmetry is indicated.

The precise stereochemistry of the interaction with U1 snRNA hairpin II was revealed by the crystal structure at 1.92 Å resolution of the human U1A protein RNA-binding domain (residues 2–98, referred to as RBD throughout this paper) complexed with a 21 nucleotide RNA representing hairpin II of U1 snRNA (Figs 1b,2a) [11]. The AUUGCAC heptanucleotide of hairpin II fits into a groove formed between loop 3 (situated between β2 and β3 strands) and the C-terminal region of the RNP domain. The heptanucleotide, together with the C:G base pair which closes the loop, forms an extensive direct and water-mediated hydrogen bond network with amino acid residues on the surface of the β sheet. The RNA bases stack onto either an adjacent base, a protein side chain, or both, showing the importance of base stacking for the stabilization of RNA structure and in RNA–protein interactions. In contrast, the bases of the last three nucleotides of the loop do not interact with the protein and are less well ordered [11]. The crystal structure of the U1A RBD–RNA hairpin II complex shows that helix C extends parallel to the β2 strand and that its position is stabilized by hydrophobic interactions between Ile58, Ile93 and Ile94 and by a hydrogen bond between Ser91 and Thr11. We have recently carried out an NMR study of a fragment of U1A protein (containing residues 2–117) and found that helix C extends across the β sheet [12]. This shows that RNA binding is accompanied by a large movement of helix C.

Figure 2



Structures of the U1A protein RBD (amino acid residues 2–97) shown as van der Waals surface representations with stick representations of RNA. (a) U1A RBD complexed with 21 nucleotide RNA representing U1 snRNA hairpin II [11]. The protein surface is shown in fuchsia and RNA is coloured yellow, except for the last three nucleotides in the loop which are coloured red, as in Figure 1b. (b) Energy minimized starting model of U1A RBD/U1A pre-mRNA 3′UTR complex with $U1A_1$ RBD, Box 1 and Stem 1 in the same orientation as the U1A RBD/RNA hairpin in (a). The two protein molecules $U1A_1$ RBD and $U1A_2$ RBD are shown in fuchsia and blue respectively. RNA nucleotide coordinates derived from the crystal structure are coloured yellow (Stem 1) and white (Stem 3), Stem 2 nucleotides are coloured red in correspondence with Figure 1c and the remaining Stem 2 nucleotides are shown in green. The tetraloop, labelled TL, is derived from the UUCG tetraloop NMR coordinates [32] and coloured white.

U1A protein also binds to the 3′ untranslated region (3′UTR) of its own pre-mRNA and regulates its expression through an unprecedented mechanism [13]. U1A is transported into the nucleus by virtue of its nuclear localization signal and incorporated into U1 snRNP [14]. Nuclear levels of U1A protein are tightly regulated, presumably because a large excess of U1A protein has a deleterious effect on pre-mRNA splicing. U1A protein downregulates its own expression by binding to the 3′UTR of its pre-mRNA and thereby inhibiting the polyadenylation of the pre-mRNA [13] through a direct interaction with poly(A) polymerase [15]. A sequence comparison of the 3′UTR of the U1A protein gene from human, *Xenopus* and mouse showed that the pre-mRNA can be folded into an evolutionarily conserved secondary structure with two internal loops, one with the AUUGCAC heptanucleotide (identical to that found in hairpin II of U1 snRNA) and the other with one base change, AUUG*U*AC [16]. Based on the crystal structure of the U1A RBD–RNA hairpin complex we previously proposed that the interaction between the U1A RBD and the AUUGCAC heptanucleotide together with the C5:G16 loop-closing base pair (Fig. 1b) is conserved in each of the two internal loops in the 3′UTR complex. We also proposed that the heptanucleotide and the middle stem (Stem 2) can be linked by one base pair of Stem 2 (C20:G49; C46:G23) and one unpaired nucleotide (C50; A24) (Fig. 1c) [11].

We have carried out molecular modelling of a complex consisting of two molecules of the U1A RBD bound to the 3′UTR of U1A pre-mRNA. A set of stereochemically plausible models can be constructed on the basis of the crystal structure of the hairpin II complex. This study yielded one structure which is consistent with the enzymatic protection and mutagenesis experiments reported by van Gelder *et al.* [16] and allows a hypothesis to be formed which explains the molecular basis of cooperativity of U1A protein binding to the 3′UTR RNA.

The 3′UTR of mRNA is often used as a binding site for proteins that regulate translation, degradation and localization of specific mRNAs [17–19]. These proteins are therefore important determinants of temporal and spatial expression of developmental genes [19–21]. Our study of the interaction between U1A protein and the 3′UTR of its mRNA could provide important insight into the molecular mechanism of these essential biological processes.

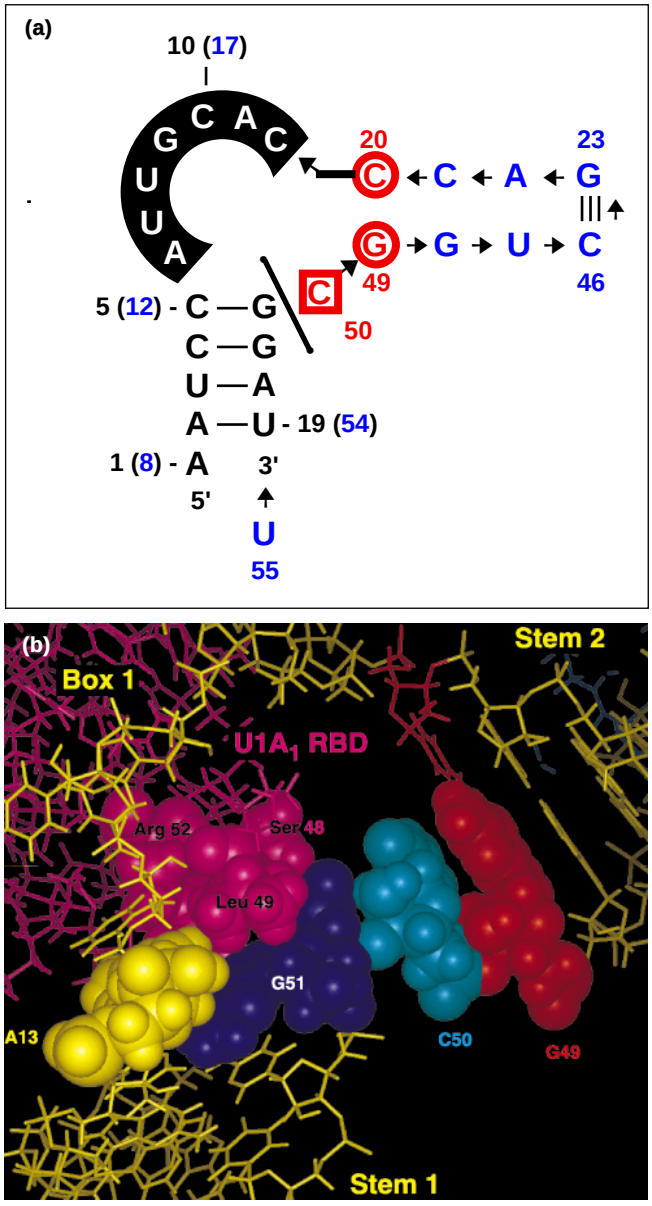
Results

Initial model building

The crystal structure of the U1A RBD–RNA hairpin II complex [11] shows that the U1A RBD forms an extensive network of hydrogen bonds and stacking interactions with the AUUGCAC heptanucleotide and the C5:G16 base pair which closes the stem. Considering the conservation of the heptanucleotide sequence between hairpin II and the 3′UTR sequences [15], it is almost certain that these interactions are conserved in the U1A–3′UTR complex (Fig. 1a,c). The conformation of these conserved nucleotides (Fig. 2a, represented in yellow) and the protein structure was therefore kept unaltered for the 3′UTR model. In the crystal structure of the hairpin II complex the bases of the last three loop nucleotides, UCC (shown in red in Fig. 2a), do not make any sequence specific contact with the protein and merely act as a linker between the heptanucleotide and the stem [11]. This is consistent with the RNA selection experiments by Tsai *et al.* [22] which showed that RNAs selected from a random pool, which are bound by U1A protein with high affinity, share the constant heptanucleotide sequence followed by three variable nucleotides. We have therefore proposed that these three nucleotides could be replaced by one base pair and one nucleotide in the 3′UTR complex [11], as shown schematically in Figure 2b. Our proposal is further supported by a recent paper which showed that these three nucleotides can be replaced by an ethylene glycol linker without affecting U1A protein binding [23].

The initial model of the U1A–3′UTR complex was constructed using the crystal structure coordinates of the U1A RBD–RNA hairpin II complex (PDB ID code: 1URN), excluding the last three nucleotides of the loop (UCC), and the program MC-SYM [24–26] as explained in Figure 3a. This program scans a nucleotide conformational database, derived from RNA structures determined by X-ray crystallography and NMR, to generate all the stereochemically possible RNA conformers that are consistent with a set of user-defined constraints. A series of binding assays conducted by van Gelder *et al.* [16] showed that, in order for two molecules of U1A protein to bind the 3′UTR of their pre-mRNA and to inhibit its polyadenylation, the base pairing ability of the two strands of Stem 2 must be maintained. An NMR study of an unbound synthetic RNA resembling half of the 3′UTR showed that the nucleotides corresponding to Stem 2 adopt A-form geometry [27]. The canonical A-form structure of duplex RNA was therefore used as a starting model for Stem 2 (shown in blue in Fig. 3a). We attempted to append Stem 2 nucleotides (C20–G23 and C46–G49) as A-form helix (MC-SYM conformational set type_A) to the last residue of the loop heptanucleotide, C12 in the crystal structure RNA (Fig. 1b) or C19 in the 3′UTR sequence (Fig. 1c), leaving nucleotide C50 (shown in red in Fig. 3a) free to adopt any possible conformation (full_A′ set: 36 alternative conformations derived from a connection sampling; represented by a square in Fig. 3a). This did not yield any solution, since the phosphate group of the unpaired nucleotide C50 could not be brought close enough to the O3′ of G51 to form a phosphodiester bond and close the loop. We therefore relaxed the A-form constraints of nucleotides C20:G49, the first base pair of Stem

Figure 3



(a) Modelling of RNA Stem 2, unpaired base C50 and 3′ terminal base U55 onto the 21 nucleotide RNA structure from the U1A RBD–RNA hairpin II complex. The conformation of RNA derived from the crystal structure, including the conserved heptanucleotide and vertical stem, is kept unaltered (shown in black). Nucleotides (shown in red and blue) corresponding to Stem 2 (Fig. 1c) and unpaired C50, are built onto the RNA derived from the crystal structure. The direction in which every nucleotide was appended to the chain is indicated by arrows. The red circles around C20 and G49 and the red box around C50 represent the MC-SYM conformational sets sample_A and full_A′, that were respectively assigned to these nucleotides (see the Results section). Three vertical bars indicate the Watson–Crick base pair between G23 and C46, as implemented in the MC-SYM script. The bar between C50 and G51 symbolizes a loop closure constraint of 2.0 Å. Numbers in black refer to those in the hairpin II complex and numbers in blue and red to those in the 3′UTR (see Fig. 1b,c). The second half of the 3′UTR binding site is generated by a symmetry operation around the pseudo twofold axis. (b) Details of C50 stacking on G49. Nucleotides A13, G49, C50, G51 and amino acid residues Ser48, Leu49 and Arg52 are shown as space-filling models; all other bonds are shown in stick representation. C50 is prevented from stacking on G51 by protein side chains lying above it, particularly Leu49, and by the purine base of A13.

2, by assigning to them the conformational set sample_A (17 alternative conformations for C3′-endo, antinucleotides; indicated by a circle in Fig. 3a). In both cases, this led to a localized distortion of the RNA helix, with a loss of base pairing between C20 and G49. We chose the conformer in which the two bases were closest to having base pair geometry and manually modified their mutual orientation with the program O [28]. The ‘refine zone’ option of the program was used to automatically refine the geometry of the entire Stem 2, so that complete base pairing between its two strands was established. These manipulations resulted in a slight deviation of the Stem 2 helix from the canonical A form.

The only unknown parameter remaining was the position of the unpaired base of nucleotide C50, which connects the conserved part of the complex (one U1A RBD [$U1A_1$], Stem 1 and the loop heptanucleotide) with the middle stem. There were three possibilities: the C50 base could stack onto base G49, it could stack onto base G51, or it could stack on neither. The last possibility is energetically unfavourable as it would expose a large hydrophobic surface of base C50 to solvent. This position would be possible if C50 is involved in a specific interaction with the protein, similar to that observed in the structures of the glutaminyl- and aspartyl-tRNA synthetase-tRNA complexes, where tRNA anticodon bases are specifically recognized by the protein [29–31]. However, the identity of base 50 in *Xenopus laevis* pre-mRNA 3′UTR (where Box 1 and Box 2 sequences (Fig. 2c) are swapped), is changed from C to A, while the corresponding nucleotide in Box 2, nucleotide 24, is still A [16]; this argues against a specific interaction of the unpaired bases 24 and 50 with protein. We then tested, both by manual building and with MC-SYM, whether it was possible to stack base C50 onto the base of nucleotide G51. In all cases it was not possible to obtain this type of stacking without bringing C50 too far from the preceding base G49 for them to be connected. It is also not possible to stack C50 onto G51 as the latter is involved in many interactions with the protein and also partially stacked against base A13 (Fig. 3b); stacking of base C50 on G51 was therefore sterically hindered unless these interactions were broken. These considerations led us to conclude that C50 must stack onto base G49. Exploration of this possibility with MC-SYM yielded two solutions, both of which introduced a sharp turn in the RNA sugar–phosphate backbone between nucleotides C50 and G51. Only one of these solutions showed satisfactory stacking base geometry and was therefore chosen for subsequent analysis. The 3′ terminal nucleotide of Stem 1, U55, corresponding to hairpin nucleotide U20 in the 21-mer synthetic RNA, which was not well defined in the crystallographic structure, was added in A form to the RNA chain.

As can be seen in Figure 1c, there is an evident secondary structure symmetry between the two parts of the 3′UTR

RNA. The symmetry can be defined by a pseudo twofold axis passing through the middle of Stem 2, not taking into account the nucleotides upstream or downstream of Stem 1 or the tetraloop which closes Stem 3 (Fig. 1c). The structure of the other half of the 3′UTR complex, including the second U1A RBD and Stem 3, was generated by symmetry operation around the pseudo twofold axis.

The UUCG tetraloop structure, determined by NMR [32], was then added to the newly created Stem 3 (TL, Fig. 2b). Finally, the RNA sequence was changed to the natural human U1A pre-mRNA 3′UTR sequence with the Insight II program (BIOSYM Technologies, Inc., San Diego, CA). U1A RBD in the hairpin II complex crystal contains two mutations, Tyr31→His and Gln36→Arg, which were reverted to the wild type residues in the model. As the crystal structure did not completely determine all amino acid side-chain atom positions, the modelling enabled us to complete these side chains while retaining the coordinates defined by the crystal structure.

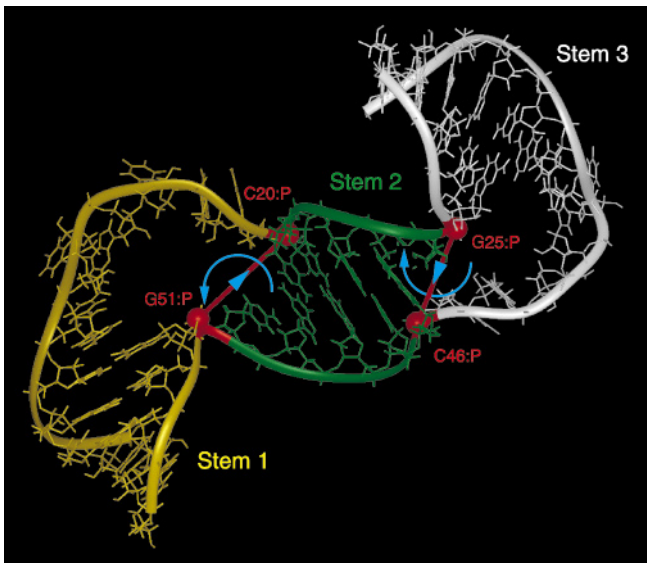
Refinement of initial model

The initial model was refined by energy minimization. The application of differential constraints to different regions of the model, as described in the Materials and methods, was crucial in order to smoothly integrate the *de novo* generated parts of the structure with those derived from experimental data. This resulted in excellent overall geometry of the model, as evaluated by both PROLSQ and PROCHECK [33]. Moreover, it permitted us to relax the variant Box 1 base (U17) and the protein side chains in its vicinity in order to improve their geometry. The minimized starting model is depicted in Figure 2b. Its root mean square (rms) deviation from the unminimized structure (excluding hydrogens) is 0.170 Å, the largest local changes being in the unpaired nucleotides A24 and C50. Movement of these nucleotides was expected since their backbone conformation was strained in the unminimized model to allow the closing of the internal loops.

Generation of alternative conformers of the starting model

In order to determine whether other conformations were energetically and geometrically feasible, the model was searched for regions of RNA with potential flexibility. The model consists of three domains with little intrinsic flexibility: the two conserved domains, derived from the crystal structure of the U1A RBD–hairpin II complex (Fig. 2b: yellow/fuchsia and white/blue, respectively), the middle stem (Stem 2) and the two unpaired nucleotides (shown in red and green in Fig. 2b). Four phosphodiester bonds which join these three domains are sites of potential flexibility, as is evident from Figure 4. These bonds have rotational freedom, and two hinges could be defined across the Stem 2 helix axis by the two pairs of phosphorus atoms G51:P–C20:P and G25:P–C46:P. We then performed a series of simultaneous rotations of the two

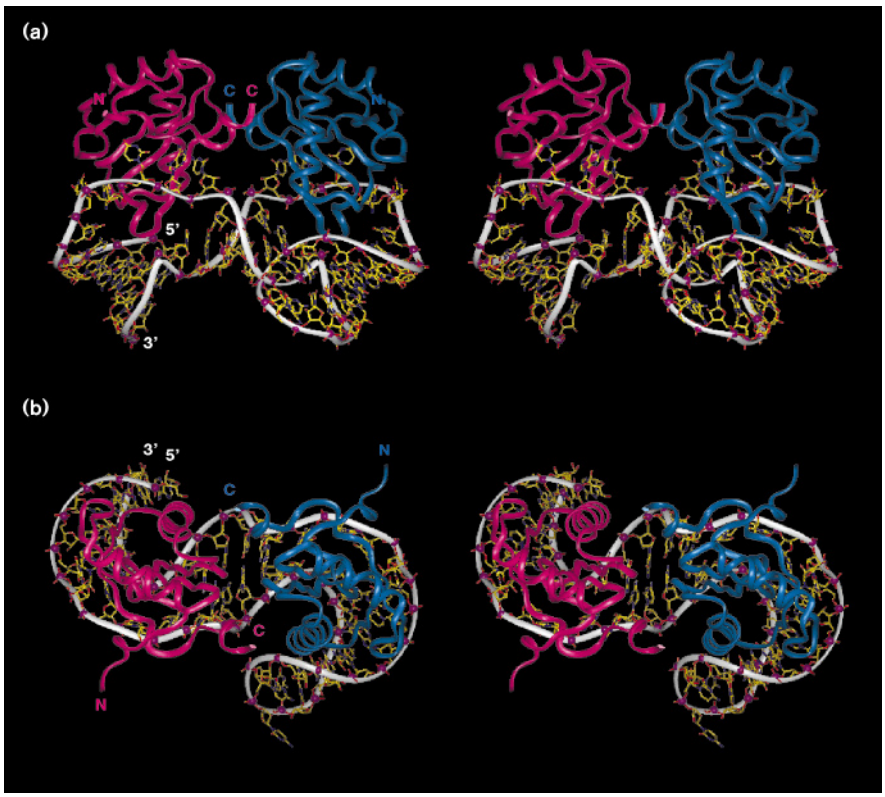
Figure 4



RNA coordinates from the energy minimized starting model, excluding the tetraloop, with a ribbon representation traced along the phosphate groups. Bonds are shown in stick representation. The hinges about which rotations were made, to generate a series of different conformers, are shown as thick red lines between the hinge phosphorus atoms, which are shown as large red spheres. Blue arrow heads on the hinges point away from the origins of the rotational axes. Semicircular blue arrows show the direction of two anticlockwise (positive) rotations. The RNA is colour coded as in Figure 2b, except for Stem 2, which is shown all in green. Protein molecules have been omitted for clarity.

conserved domains of the model around the two hinges with increments of 5° to generate possible alternative conformers. The rotations were executed in both clockwise (negative) and anticlockwise (positive) pairs from the starting positions so that the overall pseudo twofold symmetry of the model was maintained. The removal of all the conformers that showed steric clashes left structures which had been rotated within the range –45° to +130° from the starting structure. All models, including the starting structure, were then subjected to further energy minimization. Stem 2, the two unpaired nucleotides and their adjacent O3′ atoms (the region between the two 'hinges') were left unconstrained to allow movements in the *de novo* modelled RNA region; this was to correct poor geometry resulting from the rotations, as described in Materials and methods. All charged protein side chains were unconstrained to allow optimal electrostatic interactions to occur at the RNA–protein interface. Tyr31 and Gln36 side chains built *de novo* were also unconstrained. Final total geometric R factors and total energies were computed for each minimized structure (data not shown). The conformers with rotations of –40° to +75° were geometrically and energetically favourable after energy minimization.

Figure 5



Side (a) and top (b) stereo views of the UTR–37.5 model structure. RNA is plotted in stick representation with a white ribbon traced along the phosphate groups, which are shown as purple spheres, carbon atoms are shown in yellow, nitrogens in cyan, oxygens in red. Proteins $U1A_1$ RBD and $U1A_2$ RBD are represented in fuchsia and blue respectively by a ribbon traced along the α-carbons. The N and C termini of the proteins and the 5′ and 3′ ends of the RNA are marked.

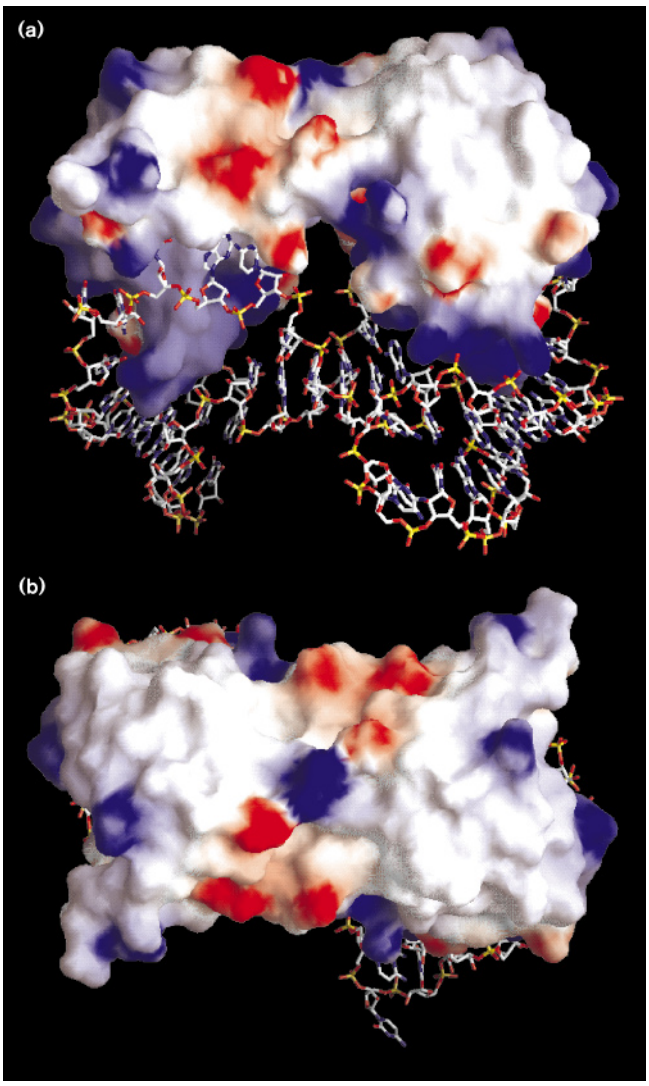
The 5° stepwise search found the –35° structure to have the lowest energy. The step size was then reduced to search around this structure. The conformer with the lowest total energy, 44.3 kcal below the mean energy of the geometrically and energetically favourable subset, is produced by a –37.5° pair of rotations (this structure will be referred to henceforth as UTR-37.5).

Evaluation of UTR-37.5

Top and side views of the lowest energy conformer (the UTR–37.5 structure) are shown in Figures 5 and 6. The two U1A RBDs have been brought into close proximity and the amino acid residues on loop 2 (between helix A and β2), loop 4 (between β3 and helix B) and helix C of the two subunits form an extensive interface. Further rotation of the conserved domains about the hinges results in protein–protein clashes, UTR–37.5 therefore lies at one extreme of acceptable hinge angles. Extensive protein contacts are made to bases of the conserved heptanucleotide sequences and the phosphate backbone of the rest of the RNA molecule. In Figure 6, U1A RBDs are shown as surface representations and coloured according to electrostatic potential. The RNA is folded into an S-shaped conformation, when viewed along the pseudo twofold axis. The major groove side of each arm of the S shape (Stem 1 and Stem 3) of the RNA embraces one U1A RBD. The RNA backbone of Stem 1 and Stem 3 thus wraps around the protein molecules to form a cup-like structure. The phosphate backbones of these stems form the cups' rims, in which the most basic regions of the proteins sit. The interactions between the protein and RNA in the Stem 1 and Stem 3 regions are left unaltered from those in the U1A RBD–hairpin II complex structure. In the case of $U1A_1$ RBD, Lys20 and Lys22 interact with the phosphate groups of C9–C12. The same two Lys residues in $U1A_2$ RBD interact with the phosphate groups of C34–A37. Stem 2 phosphates can make long range electrostatic interactions with Lys23 and Lys27.

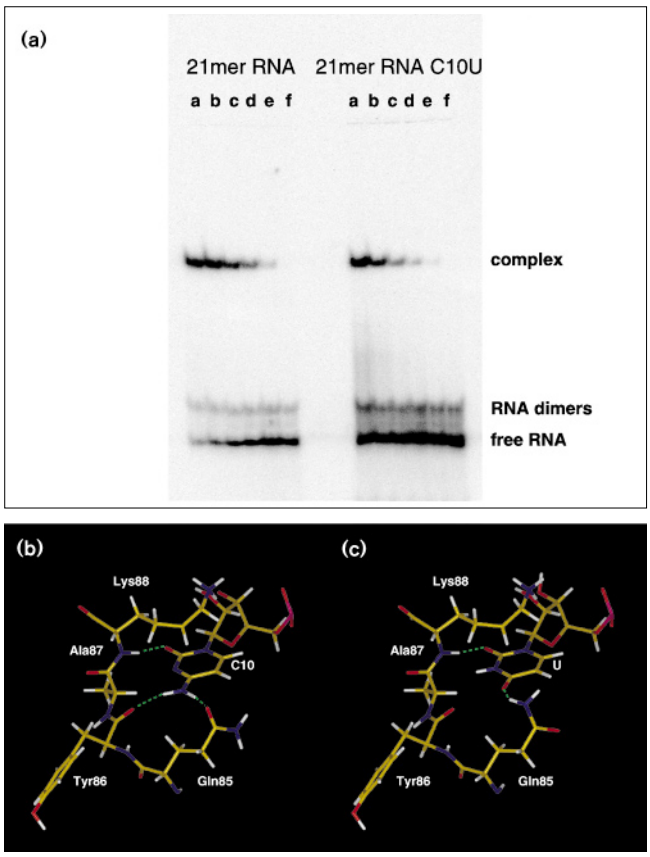
The overall complementarity of the surfaces of the two protein molecules is excellent, as can be seen in Figure 6. Since the only degree of freedom in the model approximates to one dimensional rotation around the two hinge regions, the protein surfaces which contact each other in the UTR–37.5 model cannot be significantly altered unless the Stem 2 helix is severely distorted from its position in UTR–37.5. Although it is extremely difficult to predict residue–residue contacts at the protein interface, our UTR–37.5 model exhibits favourable polar interactions. These contacts are related by the pseudo twofold axis and are between Lys96 and Asp24 (~5.2 Å apart) and between Lys60 and Gln39 (at a favourable hydrogen

Figure 6



Side (a) and top (b) views of the UTR–37.5 model structure, oriented as in Figure 5. The proteins are shown as electrostatic surface representations with positively and negatively charged regions in blue and red respectively. RNA is shown in stick representation with carbons in white, nitrogens in blue, oxygens in red and phosphorous in yellow.

Figure 7



(a) Gel mobility shift assay of 21 nucleotide RNA hairpins with U1A RBD at concentrations a=128nM, b=32nM, c=8nM, d=2nM, e=0.5nM, f=no protein. A tendency for dimers to form in short RNA hairpins was noted, presumably due to base pairing between the complementary hairpin stems. These dimers are not significantly bound by protein as compared with the hairpin monomers (data not shown). 50% binding of the 21 nucleotide RNA used for co-crystallization with U1A RBD (for sequence see Fig. 1b) was observed between 2nM and 8nM protein concentration. The 21 nucleotide RNA with a C10→U replacement, representing the variant sequence in Box 1 of the 3′UTR (see Fig. 1c), was 50% bound at a protein concentration of between 32nM and 128nM. The C10→U replacement is therefore bound with at least fourfold less affinity. (b) C10 forms hydrogen bonds (green dotted lines) with Ala87 amine, Tyr86 carbonyl and Gln85 amide oxygen in the U1A RBD–RNA complex crystal structure [11]. (c) There is a definite loss of the hydrogen bond with Tyr86 carbonyl when the cytosine N4 is replaced with the O4 of uracil. The bond with the Gln85 amide oxygen is also lost, but can be replaced via a ~180° χ_3 rotation of the amide group of the side chain of Gln85, which allows a hydrogen bond to form between the Gln85 amine and the uracil O4.

bonding distance [~2.9 Å]). Further optimization of the subunit interactions would be possible by altering side chain conformation at the interface, but no such attempt was made here as high constraints were imposed on the side chain atoms of uncharged amino acids.

The binding of the two protein molecules to the 3′UTR binding sites is cooperative: the affinities of U1A for Box 1 or Box 2 alone are ~80-fold and ~threefold lower respectively than the overall affinity for the intact 3′UTR region [16]. The simplest explanation for this cooperativity is that there is direct interaction between the protein molecules bound at the two sites, and the protein–protein contacts in the UTR–37.5 model strongly support this hypothesis. It is also possible that the binding of the first protein molecule contributes to cooperativity by stabilizing the RNA in a conformation for which the second protein has a higher affinity. In the crystal structure of free U1A protein two molecules of U1A protein interact through a hydrophobic surface on the opposite side of the RNA-binding surface [9]. However, the dimer of U1A proteins

in our 3′UTR model is distinct from that observed in the free protein crystal. Biochemical [4] and NMR [12] experiments have shown that U1A protein exists predominantly as a monomer in solution, even at high concentration. This raises the question as to how protein–protein interactions can account for the cooperativity of U1A protein binding to the 3′UTR RNA. It must be noted that, in our model, helix C is involved in the formation of the protein dimer and the position of helix C is different in free and bound protein. Avis *et al.* [12] showed that the binding of RNA to U1A protein induces a large movement of helix C. Hence the protein dimer interface involving helix C can be stabilised only when complexed with the RNA.

Differences in binding the variant Box 1 heptanucleotide sequence

Box 1 of the 3′UTR (which contains the C→U substitution, Fig. 1c) has been reported to have a lower affinity (~27-fold) for U1A than Box 2 [16]. A 21 nucleotide RNA hairpin (Fig. 1b) and its variant with a C10→U substitution were compared for protein binding affinity in a gel mobility assay with U1A RBD (Fig. 7a). The substitution of this base in the context of the U1 snRNA hairpin II sequence leads to at least a fourfold decrease in affinity.

Figure 7b shows the crystal structure of the U1A RBD–hairpin II complex in the vicinity of the cytosine at position 10. We have attempted to find a possible structural reason for the lower affinity of U1A for the C→U variant by substituting C for U in the crystal structure. This substitution resulted in the loss of a hydrogen bond with the main chain carbonyl oxygen of Tyr86 when the N4 of cytosine is replaced by the O4 of uracil (Fig. 7b,c). A second hydrogen bond with the side chain amide oxygen of Gln85 may be replaced by a hydrogen bond with the side chain carbonyl group, as the glutamine side chain is free to rotate around its χ_3 angle (Fig. 7c). This replacement, therefore, results in the loss of at least one hydrogen bond and can reasonably account for the lower affinity of U1A protein for this sequence.

Discussion

The model presented in this paper was built systematically from, for the most part, experimentally determined structures. *De novo* modelled RNA was based on a database of known nucleotide conformations. Modelling was therefore restricted to varying torsion angles at two defined regions of flexibility. The definition of flexible and rigid regions within protein and RNA molecules, and the subsequent use of these definitions to limit the exploration of conformational space during molecular dynamics calculations, have been described [34–37].

The construction of this model depended upon the assumption that U1A RBD makes the same contacts with the 3′UTR internal loops as it does with the U1 snRNA hairpin II heptanucleotide and the C:G base pair closing the loop. The complex of a U1A protein RBD and an RNA resembling half of the 3′UTR has recently been investigated by NMR [38] and the results are entirely consistent with our assumption.

In our model both protein molecules lie on the same side of the 3′UTR RNA. This could be predicted on the basis of RNA geometry, and was suggested by van Gelder *et al.* [16], although further details of the relationship between the two proteins could not be foreseen. Two U1A protein molecules must bind to the 3′UTR in order to downregulate polyadenylation *in vitro* [13,16]. U1A protein also has to directly interact with poly(A) polymerase for downregulation to occur [15]. This suggests that both U1A molecules directly contact poly(A) polymerase to inhibit its activity. A fragment of U1A protein containing the N-terminal 101 residues is sufficient for the cooperative binding to the 3′UTR but sequences C-terminal to the U1A RBD (amino acids 99–138) are required for the interaction with poly(A) polymerase (IW Mattaj, personal communication). As shown in Figure 5, the C termini of both U1A RBDs are on the face of the protein dimer opposite to that involved in the interaction with the RNA. This arrangement allows the protein residues following the N-terminal RNP domains to lie on top of the structure shown in Figure 6a. Therefore, although it cannot predict specific features of the U1A–poly(A) polymerase interaction, the model is consistent with the above data in showing that the parts of the two U1A molecules required for interaction with poly(A) polymerase could be presented to its surface while the U1A molecules were still bound to the RNA. A recent NMR study [12] showed that, upon binding RNA, a large conformational change occurs in U1A residues 95–117, which are C-terminal to the N-terminal RBD. This conformational change has two very important consequences which have implications for the interaction of poly(A) polymerase with this region. Firstly, helix C can promote protein–protein interactions only in the RNA-bound form and hence the conformational change contributes to cooperative protein binding to the 3′UTR sequence. Secondly, the conformational change causes the C-terminal extensions (residues 99–138) of both bound N-terminal RBDs which are necessary for inhibition of poly(A) polymerase to be displayed on the same side of the protein dimer so they could simultaneously interact with poly(A) polymerase. Hence U1A protein is able to inhibit poly(A) polymerase activity only when bound to 3′UTR RNA, as has been previously described [15]. A system of high specificity and sensitivity can be envisaged in which the protein conformation required to downregulate polyadenylation is achieved by two protein molecules cooperatively binding to specific RNA-binding sites in close proximity.

It has been shown that changing the conserved length of Stem 2 has a detrimental effect on both the binding of

U1A protein and the inhibition of polyadenylation (IW Mattaj, personal communication). As is obvious from the helical nature of Stem 2, lengthening of this stem moves the two proteins further apart and also causes rotation of one protein with respect to the other (Fig. 5). The interaction of the two protein molecules suggested by our model would be completely prevented by these two effects (separation and rotation).

Our model proposes a number of interactions between U1A RBDs and regions of the phosphate backbone of the 3′UTR. We predict that U1A protein binding to the 3′UTR will protect the phosphates of Stem 2 and of the unpaired nucleotides A24 and C50 from chemical ethylation, for example by ethylnitrosourea. Stems 1 and 3 are expected to have patterns of protection similar to that of the stem in the U1A–hairpin II complex [4]. We also expect that amino acid replacements at our proposed protein–protein interface would affect the cooperativity of protein binding.

Biological implications.

In the majority of DNA–protein complexes DNA recognition occurs via a protein α helix. This so-called 'recognition' α helix fits into the wide and shallow major groove of duplex DNA and makes sequence-specific contacts. In contrast, the major groove of double-stranded RNA is narrow and deep and rarely used as a sequence-specific protein binding site. Hairpins, bulges and internal loops are common RNA secondary structural elements, in which the major groove is widened and bases are more exposed. These secondary structural elements in RNA are often used as binding sites for proteins. Therefore, while in the case of DNA, which is always duplex, only the sequence must be evolutionary conserved, conservation of binding sites on RNA molecules involves both sequence and secondary structure context.

Human U1A spliceosomal protein binds to hairpin II of U1 small nuclear RNA (snRNA) and, together with other proteins, forms U1 snRNP, a large protein–RNA complex essential for pre-mRNA splicing. U1A protein also binds to the 3′ untranslated region (3′UTR) of its own pre-mRNA and inhibits polyadenylation at the 3′ end. Unpolyadenylated pre-mRNA is rapidly degraded and production of U1A protein is prevented. Hence, the nuclear level of U1A protein is autoregulated by a unique mechanism. The U1A protein binding site in the 3′UTR is a double-stranded region which contains two internal loops. How is U1A protein able to recognize these two different binding sites?

Hairpin II of U1 snRNA has a ten nucleotide loop, containing the heptanucleotide sequence AUUGCAC, and the C:G base pair which closes the loop. The crystal structure of the U1A protein–hairpin II complex has previously been determined and shows that the AUUGCAC heptanucleotide sequence and a stem containing a C:G base pair interact extensively with U1A protein. These interactions include stacking of RNA bases and aromatic protein side chains, and direct and water-mediated hydrogen bonds between RNA bases and hydrogen bond donors and acceptors of the protein. The RNA stem also interacts with the basic region of the protein.

One of the internal loops of the 3′UTR of U1A pre-mRNA contains the same AUUGCAC heptanucleotide found in hairpin II; the other internal loop contains its variant sequence AUUG*U*AC. The C:G base pair, which makes sequence specific contacts, is also conserved at the equivalent positions in the 3′UTR. Hence, the heptanucleotide and C:G base pair of the 3′UTR are likely to interact with U1A protein in the same manner as they do in the hairpin II complex. On the basis of this premise and the twofold symmetry inherent in the 3′UTR secondary structure, we have used the crystal structure of the U1A–hairpin II complex to construct a model of the U1A–3′UTR complex .

Although the structures of the two binding sites of U1A protein appear quite different, their complexes with U1A are formed using exactly the same sequence-specific interactions between the heptanucleotide and the protein and the same electrostatic interactions between the RNA stem and protein. The U1A–3′UTR complex is further stabilized by protein–protein interactions and electrostatic interactions between a short middle stem situated between the two internal loops, and the protein. This can explain the previously observed cooperative protein binding to the 3′UTR RNA. The two U1A protein molecules in the complex lie on the same side of the 3′UTR, allowing direct, simultaneous interaction with poly(A) polymerase, the enzyme responsible for polyadenylation. The model is consistent with existing biochemical data and provides a structural framework for understanding the interaction between the U1A–3′UTR complex and poly(A) polymerase. Two U1A protein binding sites arose which allow the same RNA–protein contacts to occur within different secondary structures. This is an intriguing example of molecular evolution.

The 3′UTR of mRNA is often used as a binding site for proteins that regulate translation, degradation and localization of mRNA. Our model provides the first example of one such interaction at the molecular level.

Materials and methods

Model building

Stem 2, C50 and U55 were generated by the RNA modelling program MC-SYM, version 1.3, using a conformational database derived from all the nucleic acid structures determined so far by X-ray crystallography and NMR [24,25]. The same program was used to stack the unpaired nucleotide C50 onto base G49. The program O [28] was used for the

joining of RNA Stem 1, Box 1, Stem 2 and protein $U1A_1$ RBD with the RNA Stem 3, Box 3 and protein $U1A_2$ RBD. O was also used for the addition of the UUCG tetraloop [32] to RNA Stem 3, the last two base pairs of which were also derived from the tetraloop NMR structure, as described in the Results. The 'refine zone' option of O was used to improve RNA geometry following these steps. RNA sequences were mutated to the human U1A 3′UTR sequence using the 'Nucleotide Replace' feature of the Biopolymer module of Insight II 2.3.0 (BIOSYM Technologies, Inc., San Diego, CA). The 'Residue Replace' option of Biopolymer was used to mutate His31 to Tyr and Arg36 to Gln on both protein molecules and to complete partially defined side chains Lys20 and Lys96, which were not represented by clear electron density in the crystal structure.

The starting model was optimized by energy minimization using the program Discover, version 2.97 (BIOSYM Technologies, Inc., San Diego, CA), interfaced to the AMBER forcefield [39–41] and driven by the program AutoDiscover 2.0 (written by ourselves). A distance-dependent dielectric constant set to e(r) = 4r was used to simulate the screening effects of solvent. The 1–4 non-bonded interactions were scaled by a factor of 0.5. A non-bonded cut-off of 12.0 Å with a buffer width of 3.5 Å and a switching distance of 1.5 Å were imposed. Energy minimization was performed using steepest descent algorithm (SD) for 500 iterations when the maximum derivative was ~20 kcal mol^{-1} $Å^{-1}$; conjugate gradient algorithm (CG) was then applied until the maximum derivative was less than 0.001 kcal mol^{-1} $Å^{-1}$. Differential constraints were applied to different parts of the model by adapting to the RNA–protein complex a set of constraints suitable for proteins (N Taylor, personal communication). RNA nucleotides derived from the crystal structure were highly restrained during minimization using 100 kcal $Å^{-1}$ with a maximum force of 200 kcal mol^{-1} $Å^{-1}$. Tetraloop nucleotides and the last two base pairs of Stem 3 were restrained using 10 kcal $Å^{-1}$ with a maximum force of 100 kcal mol^{-1} $Å^{-1}$, while modelled nucleotides together with the variant Box 1 nucleotide U17 were left relatively unrestrained (1 kcal $Å^{-1}$, maximum force 10 kcal mol^{-1} $Å^{-1}$). Protein $U1A_2$ RBD, which binds to the consensus Box 2 sequence, was highly restrained using 100 kcal $Å^{-1}$ with a maximum force of 200 kcal mol^{-1} $Å^{-1}$ for main chain atoms and 10 kcal $Å^{-1}$ with a maximum force of 100 kcal mol^{-1} $Å^{-1}$ for side chain atoms. Residues of protein $U1A_1$ RBD were restrained using the same parameters as for $U1A_2$ RBD, except for residues Glu5, Tyr13, Phe56 and Gln85–Thr89. These residues are in close proximity to nucleotide U17, and were restrained with 10 kcal $Å^{-1}$ with a maximum force of 100 kcal mol^{-1} $Å^{-1}$ for main chain atoms and 1 kcal $Å^{-1}$ with a maximum force of 10 kcal mol^{-1} $Å^{-1}$ for side chain atoms. All water molecules from the crystal structure were retained and relatively unrestrained with 10 kcal $Å^{-1}$ with a maximum force of 100 kcal mol^{-1} $Å^{-1}$.

Rotamers of the energy minimized model were created in Insight II 2.3.0 as specified in the Results. The geometry of the resulting structures was evaluated with PROLSQ 2.14 [33] using a dictionary adapted for RNA–protein complexes (PR Evans, personal communication) and their energies were calculated with Insight II 2.3.0. These structures were further energy minimized using the same general parameters as for the original model, except that following 500 iterations of SD the CG algorithm was applied until the maximum derivative was less than 1.0 kcal mol^{-1} $Å^{-1}$. RNA restraints used during minimization were 100 kcal $Å^{-1}$ with a maximum force of 200 kcal mol^{-1} $Å^{-1}$, except for the Stem 2 region, the two strands of which were defined as going from the O3′ atom of C19 to the phosphate group of G25, and from the O3′ atom of C45 to the phosphate group of G51. This part of the RNA was subjected to a restraint of 1 kcal $Å^{-1}$ with a maximum force 10 kcal mol^{-1} $Å^{-1}$. Both protein molecules were highly restrained, as specified for the $U1A_2$ RBD protein in the first round of minimization, except for all charged side chains and *de novo* built side chains of residues Lys20, Tyr31, Gln36 and Lys96, which were not subjected to any restraint (see Results). Main chain atoms of these residues were restrained using 10 kcal $Å^{-1}$ with a maximum force of 100 kcal mol^{-1} $Å^{-1}$. Water molecules were included and restrained as described previously.

The rotamer UTR–37.5 was further minimized until the maximum derivative was less than 0.001 kcal mol^{-1} $Å^{-1}$ and its final coordinates have been deposited in the Brookhaven Protein Data Bank (ID code 3UTR).

RNA helical parameters were evaluated using the 'Nucleotide Measure' feature of the Biopolymer module of Insight II 2.3.0, which is based on RE Dickerson's NEWHEL91 program suite. Coordinate format conversions were performed with the program PDBInsight2PDB 1.0 written by L Jovine.

XRNA 5.0 (B Weiser and HF Noller, personal communication), O 5.10.3 and Insight II 2.3.0 were used for the visualization of structures. Figures were drawn using the latter program, except for Figure 6, which was made with GRASP 1.1 [42]. All the modelling was performed on a Silicon Graphics Indigo 2 Extreme 64 Mb; the molecular mechanics computations of the initial model were performed on a Silicon Graphics POWER Challenge 256 Mb.

Gel mobility shift assay

Binding reactions were carried out in a total volume of 20μl for 20 min at room temperature. The reactions contained 10mM HEPES.NaOH, pH 7.4, 50mM KCl, 1mM $MgCl_2$, 30mM NaCl, 60μg ml^{-1} sonicated calf thymus DNA, <0.1nM 5′ ^{32}P-labelled RNA probe, ~6.6μM *Escherichia coli* mixed tRNA as cold competitor (66 000-fold molar excess). Protein dilutions were in 50mM NaCl, 100μg ml^{-1} sonicated calf thymus DNA. RNA probes were 5′ ^{32}P-labelled using T4 polynucleotide kinase and gel purified. Free and bound RNA were separated on a 12% 19:1 acrylamide:*bis*-acrylamide gel containing 90mM Tris.borate, pH 8.0 (1×TB) and 0.1% Triton X-100 and run in 1×TB buffer at 220V. Gels were dried against DE81 paper before autoradiography or exposure against an imaging plate. The amount of complexed and free RNA in gel bands was quantified from the imaging plate or film images. The proportion of complex to total RNA (excluding dimers) was plotted and best fits to σ plots were calculated, essentially as described in [43]. Apparent K_ds were taken as the protein concentration necessary for 50% RNA binding, and these were used to judge relative affinities.

Acknowledgements

We would like to thank Neil Taylor, Tadeusz Skarzynski and Alan Wonacott of Glaxo Wellcome Research and Development UK for constructive discussions and help with energy minimization. Thanks to Ignacio Tinoco for tetraloop coordinates, Phil Evans for help with PROLSQ, Jong Park for insights and precious machine time, Eric Westhof for comments, James Gilbert for film scanning, Joanna Westmoreland for Figure 1, Christian Kambach, Robert Young, Gianna Panetta and Rachel Marshall for helpful criticism. LJ is supported by Glaxo Wellcome Research and Development UK. This project was funded by the Medical Research Council and Human Frontier Science Program.

References

1. Lührmann, R., Kastner, B. & Bach, M. (1990). Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochim. Biophys. Acta* 1087, 265–292.
2. Scherly, D., Boelens, W., van Venrooij, W.J., Dathan, N.A., Hamm, J. & Mattaj, I.W. (1989). Identification of the RNA binding segment of human U1A protein and definition of its binding site on U1snRNA. *EMBO J.* 8, 4163–4170 .
3. Lutz-Freyermuth, C., Query, C.C. & Keene, J.D. (1990). Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA. *Proc. Natl. Acad. Sci. USA.* 87, 6393–6397.
4. Jessen, T.H., Oubridge, C., Teo, C.H., Pritchard, C. & Nagai, K. (1991). Identification of molecular contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA. *EMBO J.* 10, 3447–3456.
5. Sillekens, P.T.G., Habets, W.J., Beijer, R.P. & van Venrooij, W.J. (1987). cDNA cloning of the human U1 snRNA associated A protein: extensive homology between U1 and U2 snRNP-specific proteins. *EMBO J.* 6, 3841–3848.
6. Nagai, K., Oubridge, C., Ito, N., Avis, J. & Evans, P. (1995). The RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA. *Trends Biochem. Sci.* 20, 235–240.

7. Birney, E., Kumar, S. & Krainer, A.R. (1993). Analysis of the RNA recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res*. **21**, 5803–5816.
8. Scherly, D., Boelens, W.C., Dathan, N.A., van Venrooij, W.J. & Mattaj, I.W. (1990). Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B′′ and their cognate RNAs. *Nature* **345**, 502–506.
9. Nagai, K., Oubridge, C.J., Jessen, T.H., Li, J. & Evans, P.R. (1990). Crystal structure of the RNA binding domain of the small nuclear ribonucleoprotein A. *Nature* **348**, 515–520.
10. Howe, P., Nagai, K., Neuhaus, D. & Varani, G. (1994). NMR-studies of U1 snRNA recognition by the N-terminal RNP domain of the human U1A protein. *EMBO J.* **13**, 3873–3881.
11. Oubridge, C., Ito, N., Evans, P.R., Teo, C.H. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432–438.
12. Avis, J.M., Allain, F.H.-T., Howe, P., Varani, G., Nagai, K. & Neuhaus, D. (1996). Solution structure of the N-terminal RNP domain of U1A protein: the role of the C-terminal residues in structure stability and RNA binding. *J. Mol. Biol.* **257**, 398–411.
13. Boelens, W.C., Jansen, E.J., van Venrooij, W.J., Stripecke, R., Mattaj, I.W. & Gunderson, S.I. (1993). The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell* **72**, 881–892.
14. Kambach, C. & Mattaj, I.W. (1992). Intracellular distribution of the U1A protein depends on active transport and nuclear binding to U1 snRNA. *J. Cell Biol.* **118**, 11–21.
15. Gunderson, S.I., *et al*., & Mattaj, I.W. (1994). The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell* **76**, 531–541.
16. van Gelder, C.W.G., *et al.*, & van Venrooij, W.J. (1993). A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A is required for regulation of polyadenylation. *EMBO J.* **12**, 5191–5200.
17. Standart, N. & Jackson, R.J. (1994). Regulation of translation by specific protein/mRNA interactions. *Biochimie* **76**, 867–879.
18. Singer, R.H. (1993). RNA zipcode for cytoplasmic addresses. *Curr. Biol.* **3**, 719–721.
19. St Johnston, D. (1995). The intracellular localization of messenger RNAs. *Cell* **81**, 161–170.
20. Dubnau, J. & Struhl, G. (1996). RNA recognition and translational regulation by a homeodomain protein. *Nature* **379**, 694–699.
21. Rivera-Pomar, R., Niessing, D., Schmidt-Ott, U., Gehring, W.J. & Jäckle, H. (1996). RNA binding and translational suppression by bicoid. *Nature* **379**, 746–749.
22. Tsai, D.E., Harper, D.S. & Keene, J.D. (1993). U1-snRNP-A protein selects a ten nucleotide consensus sequence from a degenerate RNA pool presented in various structural contexts. *Nucleic Acids Res*. **19**, 4931–4936.
23. Williams, D.J. & Hall, K.H. (1996). RNA hairpins with non-nucleotide spacers bind efficiently to the human U1A protein. *J. Mol. Biol.* **257**, 265–275.
24. Major, F., Turcotte, M., Gautheret, D., Lapalme, G., Fillion, E. & Cedergren, R. (1991). The combination of symbolic and numerical computation for three-dimensional modeling of RNA. *Science* **253**, 1255–1260.
25. Gautheret, D., Major, F. & Cedergren, R. (1993). Modeling the three-dimensional structure of RNA using discrete nucleotide conformational sets. *J. Mol. Biol*. **229**, 1049–1064.
26. Major, F., Gautheret, D. & Cedergren, R. (1993). Reproducing the three-dimensional structure of a tRNA molecule from structural constraints. *Proc. Natl. Acad. Sci. USA.* **90**, 9408–9412.
27. Gubser, C.C. & Varani, G. (1996). Structure of the polyadenylation regulatory element of the human U1A pre-mRNA 3′-untranslated region and interaction with the U1A protein. *Biochemistry* **35**, 2253–2267.
28. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110–119.
29. Rould, M.A., Perona, J.J., Söll, D. & Steitz, T.A. (1989). Structure of *E. coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2.8 Å resolution. *Science* **246**, 1135–1142.
30. Rould, M.A., Perona, J.J. & Steitz, T.A. (1991). Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature* **352**, 213–218.
31. Cavarelli, J., Rees, B., Ruff, M., Thierry, J.C. & Moras, D. (1993). Yeast tRNA(Asp) recognition by its cognate class II aminoacyl-tRNA synthetase. *Nature* **362**, 181–184.
32. Cheong, C., Varani, G. & Tinoco, I., Jr. (1990). Solution structure of an unusually stable RNA hairpin, 5′GGAC(UUCG)GUCC. *Nature* **346**, 680–682.
33. Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* **50**, 760–776.
34. Harvey, S.C. & McCammon, J.A. (1981). Intramolecular flexibility in phenylalanine transfer RNA. *Nature* **294**, 286–287.
35. Tung, C.S., Harvey, S.C. & McCammon, J.A. (1984). Large-amplitude bending motions in phenylalanine transfer RNA. *Biopolymers* **23**, 2173–2193.
36. Amadei, A., Linssen, A.B.M. & Berendsen, H.J.C. (1993). Essential dynamics of proteins. *Proteins* **17**, 412–425.
37. van Aalten, D.M.F., Amadei, A., Linssen, A.B.M., Eijsink, V.G.H., Vriend, G. & Berendsen, H.J.C. (1995). The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. *Proteins* **22**, 45–54.
38. Allain, F.H.-T., Gubser, C.C., Howe, P.W.A., Nagai, K., Neuhaus, D. & Varani, G. (1996). Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature* **380**, 646–650.
39. Weiner, P.K. & Kollman, P.A. (1981). AMBER: Assisted Model Building with Energy Refinement. A general program for modeling molecules and their interactions. *J. Comput. Chem.* **2**, 287–303.
40. Weiner, S.J., *et al*., & Winer, P. (1984). A new forcefield for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc*. **106**, 765–784.
41. Weiner, S.J., Kollman, P.A., Nguyen, D.T. & Case, D.A. (1986). An all atom forcefield for simulations of proteins and nucleic acids. *J. Comput. Chem*. **7**, 230–252.
42. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 282–296.
43. Burd, C.G. & Dreyfuss, G. (1994). RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* **13**, 1197–1204.