The EMBO Journal vol.10 no.11 pp.3447-3456, 1991

Identification of molecular contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA

Timm-H.Jessen, Chris Oubridge, C.Hiang Teo, Clare Pritchard and Kiyoshi Nagai

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

Communicated by M.Perutz

We recently determined the crystal structure of the RNP domain of the U1 small nuclear ribonucleoprotein A and identified Arg and Lys residues involved in U1 RNA binding. These residues are clustered around the two highly conserved segments, RNP1 and RNP2, located in the central two β strands. We have now studied the U1 RNA binding of mutants where potentially hydrogen bonding residues on the RNA binding surface were replaced by non-hydrogen bonding residues. In the RNP2 segment, the Thr11→Val and Asn15→Val mutations completely abolished, and the Tyr13→Phe and Asn16→Val mutations substantially reduced the U1 RNA binding, suggesting that these residues form hydrogen bonds with the RNA. In the RNP1 segment Arg52→Gln abolished, but Arg52→Lys only slightly affected U1 RNA binding, suggesting that Arg52 may form a salt bridge with phosphates of U1 RNA. Ethylation protection experiments of U1 RNA show that the backbone phosphates of the 3′ two-thirds of loop II and the 5′ stem are in contact with the U1 A protein. The U1 A protein−U1 RNA binding constant is substantially reduced by A→G and G→A replacements in loop II, but not by C→U or U→C replacements. Based on these biochemical data we propose a structure for the complex between the U1 A ribonucleoprotein and U1 RNA.

Key words: mutagenesis/RNA binding protein/stem−loop structure/U1 A protein/U1 RNA/U1 snRNP

Introduction

Several small nuclear ribonucleoprotein particles (snRNPs) take part in the excision of introns from pre-messenger RNA (pre-mRNA). U1 and U2 snRNPs recognize the 5′ splice site and the branch point of pre-mRNA respectively (reviewed in Sharp, 1987; Maniatis and Reed, 1987; Steitz *et al.*, 1988; Lührmann *et al.*, 1990; Mattaj, 1990; Ruby and Abelson, 1991). U1 snRNP consists of one 165 nucleotide-long RNA molecule, three specific protein components called U1 A, U1 70K and U1 C, and several additional protein components that are also found in other snRNPs. U1 RNA can be folded in such a way that it forms four loops with double-stranded stems; the U1 70K and U1 A proteins bind to stem−loops I and II respectively (Hamm *et al.*, 1988; Patton and Pederson, 1988; Scherly *et al.*, 1989; Query *et al.*, 1989; Bach *et al.*, 1990; Hamm *et al.*, 1990). The U1 A protein contains two copies of the RNP domain (Sillekens *et al.*, 1987) which consists of ~80 residues and is used as an RNA binding module in many other RNA binding proteins (Dreyfuss *et al.*, 1988; Bandziulis *et al.*, 1989; Query *et al.*, 1989). Scherly *et al.* (1989) showed that the N-terminal 101 residues, containing a single RNP domain, are sufficient for specific binding to stem−loop II of U1 RNA. Similarly a 111 residue fragment of the U1 70K protein containing a single copy of the RNP domain confers full binding specificity to stem−loop I of U1 RNA (Query *et al.*, 1989).

We have recently solved the crystal structure of the N-terminal RNP domain of the U1 A protein at 2.8 Å resolution (Nagai *et al.*, 1990). It consists of a four-stranded anti-parallel β sheet, flanked on one side by two α helices. The RNP1 and RNP2 segments, which are the two most conserved regions in the RNP domain, lie side by side in the central two β strands. We have also studied the U1 RNA binding properties of the U1 A protein mutants in which the Lys and Arg residues are replaced in turn by Gln. Replacement of one of the RNP2 residues, Arg52, with Gln completely abolished the U1 RNA binding. The Lys22→Gln, Lys27→Gln, Lys50→Gln, Lys80→Gln and Lys98→Gln mutations also significantly reduce U1 RNA binding (see Figure 4 in Nagai *et al.*, 1990). Apart from Lys98, the position of which is not known, most of these residues are clustered around the RNP1 and RNP2 segments. This shows that the U1 RNA binds to the surface of the four-stranded β sheet and to loops at one edge of the β sheet.

Crystallographic studies of DNA−protein complexes have revealed hydrogen bonds between bases and phosphates of the DNA and side chains of the protein that determine the binding specificity (Steitz, 1990). These hydrogen bonds are often formed by asparagines or glutamines, because their side chains can act as both donors and acceptors. Asn, Thr, Ser and Gln residues are on the RNA binding surface of the U1 A protein. We have now mutated these to residues of similar size, but which are unable to donate or accept hydrogen bonds. U1 RNA binding studies of these mutant proteins show that some of them do indeed form crucial contacts. We have also carried out ethylation protection experiments on the U1 A protein−U1 RNA complex to identify phosphate groups in contact with the protein. Scherly *et al.* (1989, 1990a) showed that the nucleotide sequence of the loop II of U1 RNA is the major determinant of the binding specificity; therefore bases in this region of U1 RNA are likely to form specific hydrogen bonds to the protein. In order to identify these contacts we have introduced transition mutations (U⟷C or G⟷A) at every base in the loop II of U1 RNA in turn and studied its binding to the U1 A protein.

The U2 B″ protein found in U2 snRNP is similar to the U1 A protein and contains two RNP domains, but they are linked by a shorter polypeptide (Habets *et al.*, 1987). The N-terminal and C-terminal RNP domains of the U1 A protein show strong homology to the corresponding domains of the U2 B″ protein. The homology between N-terminal and C-terminal domains within each of these proteins is much lower

© Oxford University Press

(Sillekens *et al.*, 1987). The U1 A protein alone binds to stem−loop II of U1 RNA (Scherly *et al.*, 1989; Bach *et al.*, 1990), but the U2 B″ protein binds to stem−loop IV of U2 snRNA only in the presence of the U2 A′ protein. Scherly *et al.* (1990a, 1990b) showed that the N-terminal domain of the U2 B″ protein containing residues 1−98, retains the binding sites for U2 RNA and the U2 A′ protein (Scherly *et al.*,1990a). In this N-terminal region the amino acid sequences of the U1 A and U2 B″ proteins differ at only 20 sites, and some subset of these replacements must therefore account for the differences in RNA binding specificity between the U1 A and U2 B″ proteins and the binding of the U2 A′ protein to the U2 B″ protein. RNA binding studies of chimeric proteins between the U1A and U2 B″ proteins showed that the β2 strand and the β2−β3 loop are predominantly responsible for the discrimination between the U1 and U2 RNA. We have introduced some of the residues found in the U2 B″ protein into the U1 A protein and have estimated the contribution of these residues in distinguishing U1 from U2 RNAs.

The RNP domain is used as an RNA binding module in many RNA binding proteins, including the U1 70K protein (Query *et al.*, 1989), the poly(A) binding protein (Sachs *et al.*, 1987; Swanson *et al.*, 1987) and the hnRNA binding proteins (Dreyfuss *et al.*, 1989). Our results provide some insight into protein−RNA contacts in other proteins of this family.

Results

U1 RNA binding studies of mutant U1 A proteins

A fragment of the U1 A protein containing residues 1−102 (A102) has full sequence specific binding activity to U1 RNA; we have therefore studied the RNA binding properties of mutant A102 proteins. The U1 RNA stem−loop II (nucleotide 50−92) was synthesized by *in vitro* transcription with T3 RNA polymerase (Morris *et al.*, 1986; Milligan *et al.*, 1987) and this RNA was used for both mobility-shift and filter binding assays to estimate the RNA−protein binding constants. The quantitative results of the filter binding assay were more reproducible than those of the mobility-shift experiment, but the latter allowed us to distinguish between specific complex formation and non-specific binding. The two kinds of assay gave consistent results, even though in the mobility-shift assay free and protein-bound RNA molecules are not in true equilibrium. Figure 1 shows the RNA binding surface of the N-terminal RNP domain of the U1 A protein based on our X-ray crystallographic structure (Nagai *et al.*, 1990). In this region we have mutated only surface residues which could make contacts with RNA. By examining the crystal structure we were able to choose substitutions in such a way that they caused minimal instability and structural disturbance in the protein. The mutants listed in Table I were all expressed as soluble protein in *Escherichia coli* and purified without any indication of instability above that of the wild type protein. Figure 2 shows a typical mobility-shift experiment and is representative of the different binding characteristics observed in mutant proteins. The dissociation constant of the complex (K_D) for the wild type A102 is estimated to be 10 nM by mobility-shift assay, in good agreement with that obtained by the filter binding assay. The Gln54→Phe mutation increases the dissociation constant >100-fold, but

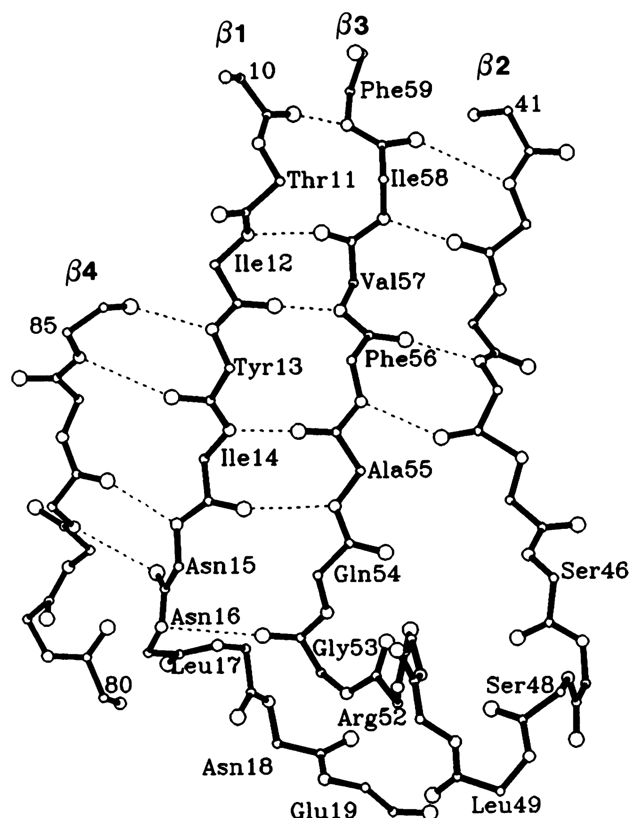


Fig. 1. The RNA binding surface of the U1 A protein consisting of a four stranded antiparallel β sheet. RNP2 (residues 11−17) and RNP1 (residues 52−59) lie side by side in the central two β-strands, β1 and β3.

Table I. Relative binding affinities of mutant A102 proteins to U1 RNA stem−loop II

Protein	K_D(nM)	Residue in U2 B″ protein	Location within structure
A102 wt	10		
A96	60		
N 9 V	40	N	N-terminal loop
H 10 N	10	H	″
T 11 V	−	T	β1 (RNP2)
Y 13 F	250	Y	″
N 15 V	−	N	″
N 16 V	150	N	″
N 18 A	50	N	Loop β1−helix A
E 19 D	20	D	″
Q 39 H	10	H	Loop helix A − β2
S 46 A	20	L	β2
S 46 L	90	L	″
S 48 A	20	T	Loop β2−β3
S 48 T	10	T	″
S46LS48T	80	L,T	β2, loop β2−β3
L 49 A	20	M	Loop β2−β3
R 52 Q	−	R	″ (RNP1)
R 52 K	30	R	″
Q 54 F	1400	Q	β3 (RNP1)
F 56 Y	30	F	″ ″
D 79 V	30	G	Loop helix B−β4
Q 85 V	20	Q	β4

The dissociation constant K_D of the U1 RNA stem−loop II with various mutant proteins are shown in comparison with the wild-type (wt) A102 protein (amino acids 1−102 of the U1 A protein). Total loss of specific binding is indicated by −; K_D was reproducible within ±20%.

formation of a specific complex can be seen with a large protein excess. In contrast, no specific complex formation is observed for the Thr11→Val mutant, and addition of a large protein excess merely produces a smear due to non-specific binding. K_D increases up to 10-fold for the Ser46→Leu and Arg52→Lys mutants. Table I summarizes the U1 RNA binding properties of all the mutants estimated by mobility-shift experiments and filter binding assays.

Amino acid residues involved in RNA binding

The RNP2 segment. The conserved RNP2 segment comprises the $\beta 1$ strand and the $\beta 1-A$ loop region. In this region Ile12 and Ile14 residues are internal residues which form part of the protein core, but Thr11, Tyr13, Asn15 and Asn16 are on the surface and could interact with the RNA. In the crystal structure, the hydroxyl group of Thr11 forms no hydrogen bond with other residues and the Thr11→Val mutation is therefore unlikely to cause any significant structural changes. However, this mutation completely abolishes the U1 RNA binding, indicating that this residue makes crucial RNA contacts. Thr11 was not included in RNP2 by Bandziulis *et al.* (1989), but in many RNA binding proteins this position is occupied by Thr, Ser, Asn, Arg, or Lys, all of which can form hydrogen bonds and are probably involved in RNA binding. We therefore propose that Thr11 is a part of RNP2.

The side chain of Asn15 is also free in solution (Nagai *et al.*, 1990), and the replacement of this residue with Val also completely abolishes the U1 RNA binding. The Asn16→Val mutation substantially reduces RNA binding, but the Asn18→Ala mutation has only a moderate effect. The binding of U1 RNA is also reduced substantially by the Tyr13→Phe mutation, although a mobility shift experiment shows that a specific complex is formed at high protein concentration. The phenol oxygen of Tyr13 forms a hydrogen bond with the side chain of Gln54. This suggests that the phenol oxygen may either form a hydrogen bond directly with RNA or orientate the Gln54 side chain to permit formation of a hydrogen bond with RNA.

The RNP1 segment. The RNP1 segment comprises the $\beta 3$ strand and its preceding loop ($\beta 2-\beta 3$ loop). We have shown that the Arg52→Gln mutation completely abolishes the specific U1 RNA binding (Nagai *et al.*, 1990). The Arg52 side chain is disordered in the electron density map, hence it is unlikely that the inability of the Arg52→Gln mutant to bind U1 RNA is due to a structural disturbance. This side chain is therefore likely to form either a salt bridge with a phosphate or a hydrogen bond with a base of RNA. The same position is occupied by either Arg or Lys in most proteins in this family (Bandziulis *et al.*, 1989; Query *et al.*, 1989) and the RNA binding is indeed only slightly reduced by the Arg52→Lys mutation in the A102 protein. In most proteins of this family the position equivalent to Gln54 is occupied by Phe or Tyr (Bandziulis *et al.*, 1989) which form a cluster of aromatic residues on the surface of the β-sheet with Tyr or Phe at positions corresponding to residues 13 and 56 (Nagai *et al.*, 1990). In the U1 A protein structure Gln54 forms a hydrogen bond with Tyr13 which could replace the usual van der Waals interaction between aromatic side chains at these positions. This may merely stabilize the protein; but alternatively either the Gln54 or the Tyr13 side chain may form a hydrogen bond with a base or backbone phosphate of RNA. To test this hypothesis we replaced Gln54 with Phe which is found in many proteins at this position; this replacement caused a substantial loss of binding energy. The replacement of Phe56 by Tyr has only a small effect.

The β2 strand and following loop. Scherly *et al.* (1990a, 1990b) replaced the region of the U1 A protein (amino acids 1-101) containing residues 40-49 with the corresponding region of U2 B'' protein; this hybrid protein is no longer capable of binding U1 RNA. Ser46 and Ser48 are on the surface of the β sheet where they may form specific hydrogen bonds to U1 RNA. The Ser46→Leu and Ser48→Thr replacements in the U2 B'' protein may therefore be responsible for its inability to bind U1 RNA. In order to test this idea, we introduced the Ser46→Leu, Ser46→Ala, Ser48→Ala mutations separately and together, but they had only small effects. Other residues in this region are mutated to smaller residues in the U2 B'' protein: Leu42→Val, Leu44→Val and Val45→Ala. All these replacements may contribute to the rejection of U1 RNA by the U2 B'' protein. Scherly *et al.* (1990b) showed that the binding of U1 RNA to one of their chimeric proteins was weakened by the Leu17→Met and Glu19→Asp mutations, but we found that the Glu19→Asp mutation alone does not affect the binding of the U1 A protein to the U1 RNA significantly (Table I).

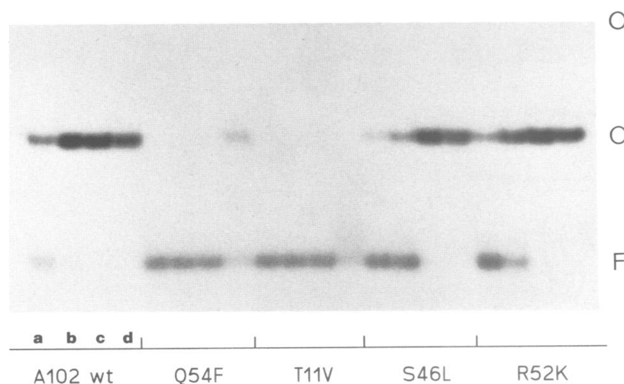
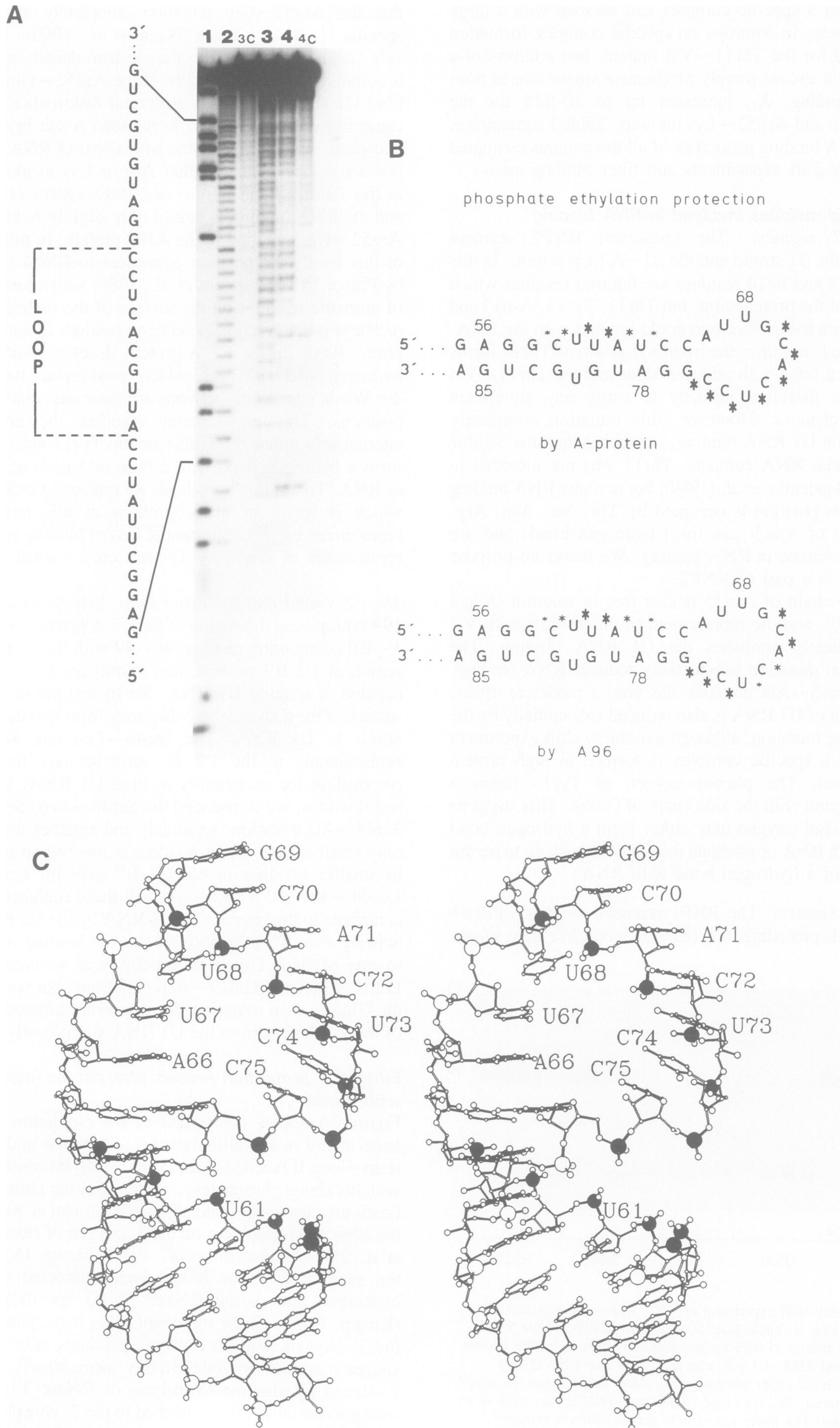


Fig. 2. Mobility-shift experiment of RNA with A102 mutants. U1 RNA stem-loop II (nucleotides 50-92) was labelled at the 5' end with ^{32}P and incubated with various concentrations of mutant protein. The complexed RNA (C) was separated from free RNA (F) by polyacrylamide gel electrophoresis on a native gel; (O) denotes origin of electrophoresis. The wild-type (wt) protein (A102) was used as an internal control. The lower case letters denote different protein concentrations: (a), 25 nM; (b), 100 nM; (c), 400 nM; (d), 1600 nM.

Ethylation protection reveals phosphates interacting with protein

Figure 3A shows the results of the ethylation protection experiments with full-length U1 A protein and U1 RNA stem-loop II labelled at the 5' end. Ethylnitrosourea reacts with backbone phosphates, and the resulting phosphotriester bonds are cleaved by mild alkaline treatment at 50°C leaving the ethylated phosphates on the 3' oxygen of ribose (Romby et al., 1985; Ehresmann *et al.*, 1987). Figure 3A also shows the same RNA after it had been subjected to alkaline hydrolysis and partial RNase T1 (G specific) digestion (Knapp, 1989). These treatments also leave phosphates on the 3' oxygen, but the cleavage products after ethylnitrosourea treatment migrate slightly more slowly than those produced by alkaline hydrolysis or RNase T1 digestion, owing to the ethyl group attached to the 3' phosphate. Bands arising from the ethylnitrosourea and subsequent alkaline



A
3´
1 2 3C 3 4 4C
G
U
C
G
U
G
U
A
G
G
C
C
U
C
A
C
G
U
U
A
C
C
U
A
U
U
C
G
G
A
G
5´
LOOP
B
phosphate ethylation protection
56
68
5´... G A G G C U U A U C C A U G C
3´... A G U C G U G U A G G C A C U C
85
78
by A-protein
56
68
5´... G A G G C U U A U C C A U G C
3´... A G U C G U G U A G G C A C U C
85
78
by A96
C
G69
C70
A71
U68
C72
U67
U73
C74
A66
C75
U61
G69
C70
A71
U68
C72
U67
U73
C74
A66
C75
U61

treatment can be assigned unambiguously using these marker lanes. In the absence of protein, the phosphate groups of U1 RNA are uniformly modified by ethylnitrosourea, both under native (lane 3) and denaturing (data not shown) conditions. In contrast to ethylation of tRNA (Romby *et al.*, 1985), ethylation of U1 RNA stem–loop II is insensitive to its secondary structure; on the other hand in the presence of the full-length U1 A protein two regions of U1 RNA are clearly protected from ethylation. The same experiment has also been carried out with the truncated A96 (U1 A protein, residue 1–96) and A102 proteins. Phosphodiester bonds protected from ethylation by the full-length A protein and A96 protein are shown with asterisks in Figure 3B. In both cases clear protection of the 5′ stem is seen from the phosphate 5′ to C59 and extends to the phosphate 5′ to C64. Clear protection in the loop starts at the phosphate 5′ to C70 and extends to the phosphate 5′ to G76. Protection of the phosphates 5′ to C72 and U73 is weaker in the truncated A96 protein, while the A102 protein provides intermediate protection. The A96 protein is the shortest polypeptide of the U1 A protein which specifically binds to U1 RNA (Lutz-Freyermuth *et al.*, 1990), but our mobility-shift experiment shows that its binding is weaker than that of the A102 protein. A lysine at position 96 is not essential for binding since the Lys96→Gln mutation in the A102 protein (Nagai *et al.*, 1990) only slightly weakens U1 RNA binding; however, removal of both amino acids Lys96 and Lys98 leads to a complete loss of specific binding.

Krol *et al.* (1990) proposed a three-dimensional model of U1 snRNA based on chemical and enzymatic footprint experiments. Figure 3C shows a model of U1 RNA stem–loop II drawn with the coordinates kindly provided by Drs E.Westhof and A.Krol. Phosphate groups protected from ethylation in the protein–RNA complex are shown with filled balls in Figure 3C. These all lie on one side of the molecule.

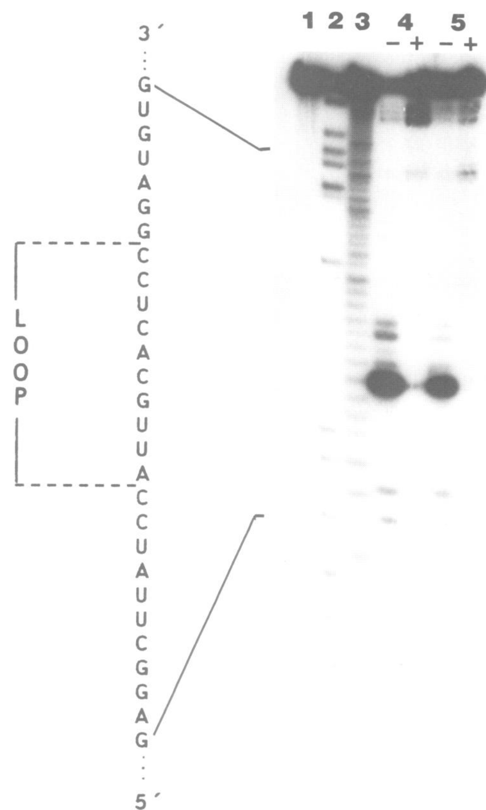


Fig. 4. RNase V1 digest of U1 RNA in the absence and presence of U1 A protein. Autoradiograph of 5′-labelled U1 RNA stem–loop II (nucleotides 50–92) analysed by polyacrylamide gel electrophoresis on a denaturing 20% gel after various treatments: lane 1, untreated RNA; lane 2, RNase T1 digest giving 'G-ladder'; lane 3, alkaline hydrolysis; lane 4, 5, RNase V1 digest in the absence (–) and presence (+) of U1 A protein; Lane 4 and 5 differ in buffer conditions, for details see Materials and methods. Note that the bands in lane 4 and 5 are shifted one nucleotide upwards since RNase V1 leaves 3′-OH on its cleavage products whereas RNase T1 digestion and alkaline hydrolysis produce 3′-phosphate ends (Krupp and Gross, 1979; Lockard and Kumar, 1981).

RNase V1 footprinting experiment

In order to gain more insight into the structure of the complex, we have carried out footprinting experiments using various RNases. A double strand-specific RNase V1 (Knapp *et al.*, 1989) cleaves the stem region of this RNA at several sites; the phosphodiester bond between U60 and U61 is preferentially cleaved, as reported by Krol *et al.* (1990). The U1 A protein completely protects these cleavage sites, including the hypersensitive site in the 5′ stem, which shows that the stem is in contact with the protein (Figure 4). The 3′ strand of the stem is cleaved around U79 more preferentially in the presence of the U1 A protein. This may be due to a structural change in the stem upon complex formation.

Effect of mutations in the RNA loop

In the crystal structures of DNA–protein complexes, hydrogen bonds are found between bases of the DNA and both the side chain and main chain atoms of the proteins (Steitz, 1990). Mutations in protein or DNA which disrupt such hydrogen bonds reduce the binding energy, but replacement of bases which do not form hydrogen bonds could still affect binding through conformational changes of the DNA. In the DNA duplex some hydrogen bond donors and acceptors are used for base pairing and are not available for interaction with proteins, but in the loop region of the RNA all hydrogen bond donor and acceptor positions of bases may form hydrogen bonds with the protein, as in the tRNA anti-codon loop in the glutaminyl tRNA–synthetase complex (Rould *et al.*, 1990). Scherly *et al.* (1990) showed that the nucleotide sequence of the U1 RNA loop II predominantly determines binding specificity. In order to probe this interaction further, we have synthesized a 22mer oligoribonucleotide corresponding to the U1 RNA stem–loop

Fig. 3. Ethylation protection of RNA phosphates by U1 A protein. **(A)** Autoradiograph of 5′-labelled U1 RNA stem–loop II (nucleotides 50–92) analysed by polyacrylamide gel electrophoresis on a denaturing 20% gel after various treatments: lane 1; RNase T1 digest giving 'G-ladder'; lane 2, alkaline hydrolysis; lane 3, ethylnitrosourea-treated free RNA with subsequent mild alkaline cleavage; lane 3C, control experiment identical to previous one only lacking ethylnitrosourea; lane 4, ethylnitrosourea treated complexed RNA (incubated with U1 A protein prior to treatment) with subsequent mild alkaline cleavage; lane 4C, control experiment identical to previous one only lacking ethylnitrosourea. **(B)** Schematic representation of the results obtained by the ethylation protection experiments. The sequence of the U1 RNA stem–loop II is shown, A96 represents amino acids 1–96 of the U1 A protein. Asterisks mark phosphates which are protected from ethylation in the presence of the respective protein. Large asterisks refer to complete protection, small ones represent weak protection. **(C)** Stereo view of the U1 RNA stem–loop II (nucleotides 57–84) according to the model proposed by Krol *et al.* (1990); phosphates protected by the U1 A protein are marked by filled balls.

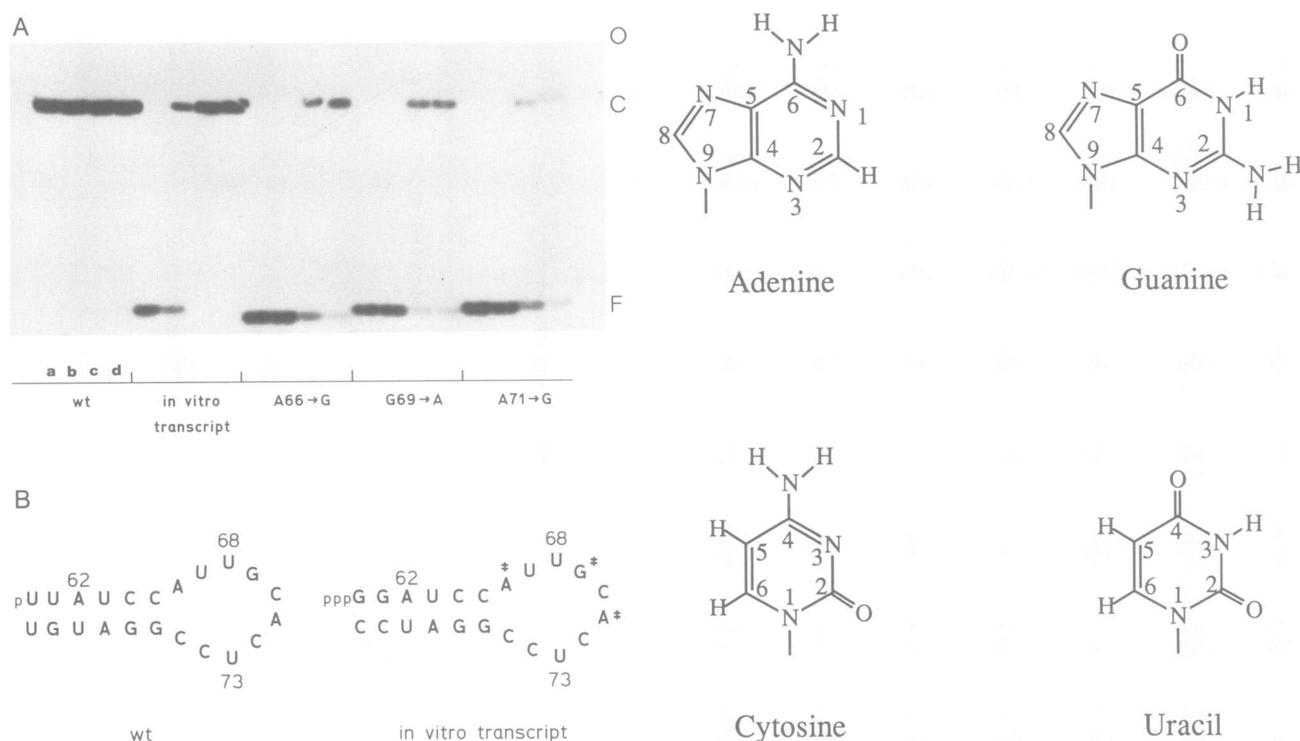


Fig. 5. Effect of U1 RNA loop II transition mutants on binding strength to A102 protein. **(A)** Mobility-shifts (see Figure 2) of ^{32}P-labelled mutant U1 RNA stem–loop II (nucleotides 60–81) after incubation with various amounts of A102 protein. The complexed RNA is marked by C, free RNA by F, origin by O; the difference between wild-type (wt) RNA and the *in vitro* transcript is explained in (B); small letters denote different A102 protein concentrations: (a), 0.1 μM; (b), 0.5 μM; (c), 2 μM; (d), 5 μM. **(B)** The sequence of the chemically synthesized wild-type (wt) U1 RNA stem–loop II (nucleotides 60–81) is shown in comparison with the *in vitro* transcript. Transition mutants in the loop which led to reduced binding strength to the A102 protein are marked with ‡. **(C)** Chemical structure of the four bases in the RNA.

II using T3 RNA polymerase (Morris *et al.*, 1986; Milligan *et al.*, 1987) and introduced transition mutations at every position of the loop in turn (Figure 5). The stem of these RNA molecules is shorter than the region protected from ethylation by A102 protein and has a triphosphate attached to its 5′ end (Figure 5B). Furthermore the stem sequence is not authentic, because it contains a perfect duplex of a GGAUCC sequence. Binding of this 22mer RNA with a wild-type loop sequence to the A102 protein is significantly weaker than that of a chemically synthesized RNA having the authentic stem sequence with a 5′ monophosphate (Figure 5B). This shows that the stem sequence also contributes to protein binding. None of the C↔U mutations in the loop significantly affect protein binding but the binding is ~10-fold reduced by purine↔purine substitutions at A66, G69 and A71. These purine bases are therefore likely either to form hydrogen bonds with the U1 A protein or to affect the binding through changes in the RNA backbone conformation. In the loop IV of U2 RNA, the C corresponding to position 72 is replaced by G; but we observed that introduction of the C72→G into U1 RNA had only a small effect on the A102 protein binding (data not shown), which confirms the observation by Scherly *et al.* (1990a).

Discussion

Our mutagenesis and ethylation protection experiments have identified important contacts between the U1 A protein and U1 RNA. Mutations of residues in the RNP1 and RNP2 segments reduce the binding energy to U1 RNA, showing that some of them may form critical contacts. In particular, replacement of Thr11 and Asn15 with non-hydrogen bonding residues of the same size and similar shape completely abolished the binding to U1 RNA. Considering the critical role of Thr11 in the RNA binding and the conserved nature of amino acids at the equivalent position in other RNA binding proteins, Thr11 should be included in RNP2 (Bandziulis *et al.*, 1989). Gln54, Tyr13 and Arg52 apparently also form critical RNA contacts.

The RNP1 and RNP2 segments are conserved between the U1 A and U2 B″ proteins and are therefore likely to interact with sequences common to both the loop II of U1 RNA and the loop IV of U2 RNA. The nucleotide sequences of the 5′ halves of the two loops are identical but, in the 3′ half of the stem–loop IV of U2 RNA, C72 is replaced by G and an A is inserted between U73 and C74. Therefore the 5′ halves of the two loops (A66–U67–U68–G69) are likely to be the ones that interact with RNP1 and RNP2. The loop II of U1 RNA and the loop IV of U2 RNA contain 10 and 13 nucleotides respectively; however the first and last nucleotides of the latter loop are uridines which are likely to form a U–U base pair. Scherly *et al.* (1990a) showed that residues in the β2 strand and the β2–β3 loop are major determinants of specificity that discriminate between U1 RNA and U2 RNA. Hence this region probably interacts with the 3′ half of the loop where the sequences of the loop II of U1 RNA and the loop IV of U2 RNA differ. These considerations permit two possibilities for positioning the U1 RNA stem–loop II on the surface of the β sheet. If the side of the loop that forms the continuation of the major groove faces the four-stranded β sheet then the stem will point downwards as shown in Figure 6A. If the minor groove side

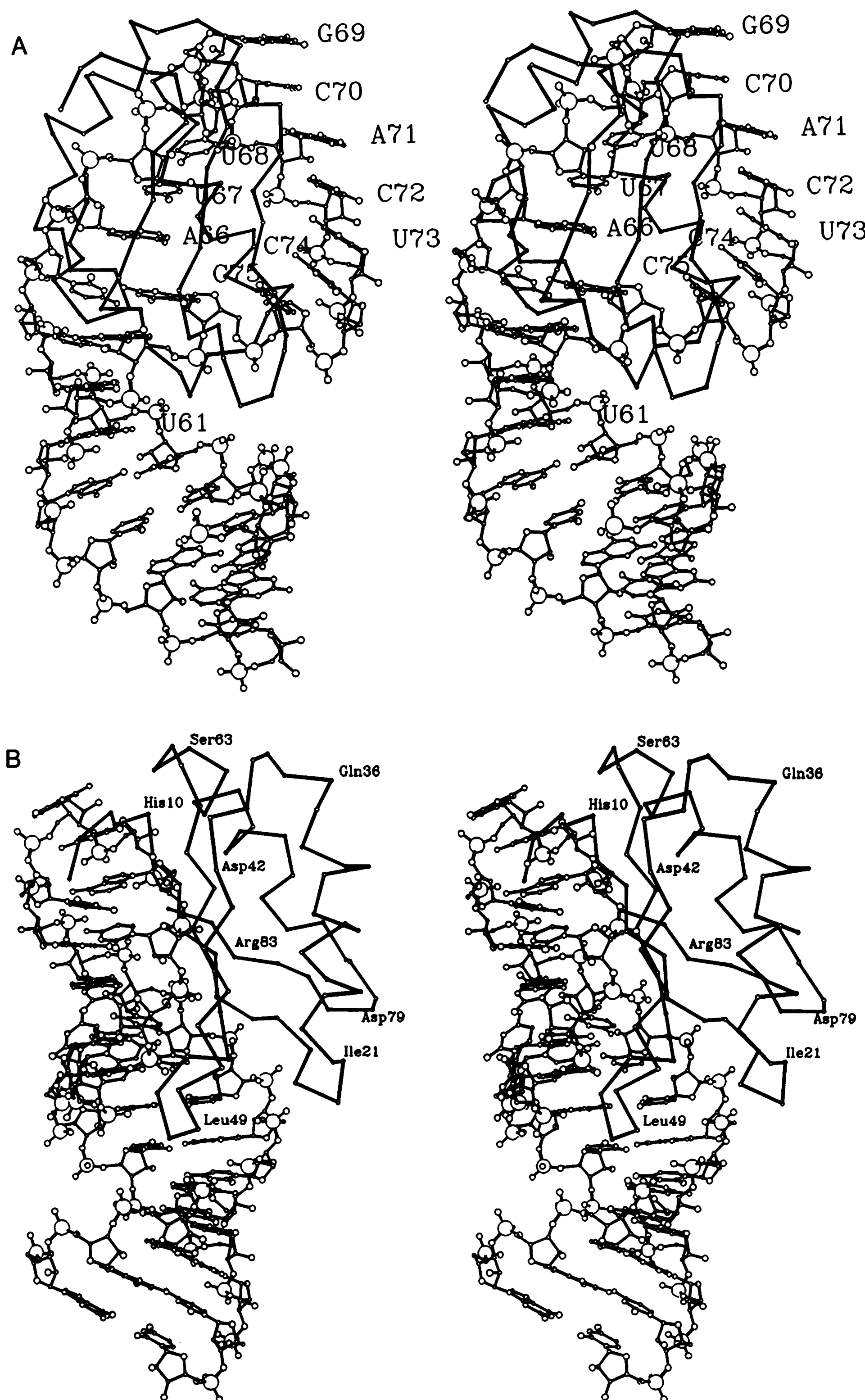


Fig. 6. Model of the complex of the N-terminal RNP domain of the U1 A protein and its cognate RNA in stereo. **(A)** The Cα-backbone of the RNP domain (Nagai *et al.*, 1990) is represented as a solid line. All atoms of the U1 RNA stem–loop II (nucleotides 57–84) are shown according to the model proposed by Krol *et al.* (1990); the bases in the loop as well as U61 in the 5′-stem are labelled. The view focuses on the binding of the RNA loop to the surface of the four-stranded β-sheet with the RNA stem pointing downward. **(B)** A view showing the interface between the four-stranded β sheet and the U1 RNA loop II. Some Cα atoms of the RNP domain are labelled. The 5′ stem of the U1 RNA loop II is embedded between a pair of basic jaws formed by the β1-A helix and β2–β3 loops. The experiments by Scherly *et al.* (1990b) show that the U2 A′ protein binding surface on the U2 B″ protein mainly consists of the A helix and β2 strand which is fully accessible in this model.

of the loop faces the β sheet, the stem would point upwards in Figure 6A (not shown). The ethylation protection experiment shows that the 3′ two-thirds of the loop and the 5′ strand of the stem are in contact with the protein (Figure 3) and the V1 nuclease digestion (Figure 4) of the complex indicates that the 5′ stem is not accessible to the nuclease.

Using the model of the U1 RNA stem−loop II generated by Krol *et al.* (1990) we tried to see which of the two alternative models can explain our ethylation protection data. The anticodon loop of tRNAs has seven unpaired bases, and its structure is stabilized predominantly by base stacking (Robertus *et al.*, 1974; Kim *et al.*, 1974; Moras *et al.*, 1980). The phosphate backbone makes a sharp turn between the second and third nucleotides of the loop, and bases before and after this turn are stacked on top of each other. The U1 RNA stem−loop II has ten unpaired bases and base stacking is expected to play a major role in the stabilization of the loop structure. The model of the U1 RNA stem−loop II proposed by Krol *et al.* (1990) assumes a sharp turn in the phosphate backbone between U68 and G69, and bases are stacked on top of each other before and after this turn. If the loop is placed on the surface of the β sheet in the upward position, the stem sticks out from the protein and the minor groove side of the loop faces the β-sheet. In this model the phosphates of the 5′ stem face away from the protein and hence the protein is unlikely to protect them from ethylation.

Figure 6A shows how the stem−loop II may interact with the surface of the four-stranded β sheet in the downward model. In this model the surfaces of the loop and the β sheet show surprisingly good complementarity and the 5′ side of the stem is wedged between the upper and lower jaws formed by the $\beta2-\beta3$ and $\beta1-$A helix loops (Figure 6B). The sharp turn between U68 and G69 twists the phosphate backbone and brings all phosphates on the 3′ side of this turn to the surface of the β sheet. In the loop region the phosphate groups on the 5′ side of this turn are not protected from ethylation, and the first clear protection starts at the sharp turn of the backbone at the phosphate 5′ to C70 (Figure 3). This clear boundary of ethylation protection is consistent with the model of the U1 RNA stem−loop II proposed by Krol *et al.* (1990). This model accounts for the ethylation protection of phosphates in the 3′ two thirds of the loop and the 5′ stem. In the 3′ two thirds of the loop, phosphates point towards the surface of the β sheet where they could interact with the $\beta2$ strand and the loop between the $\beta2$ and $\beta3$ strands. Carter and Kraut (1974) and Church *et al.* (1977) pointed out that a repeating unit of DNA or RNA phosphate backbone nearly coincides with that of an antiparallel β strand. The main chain amide groups of the $\beta2$ strand may form hydrogen bonds with the phosphates in the 3′ half of the RNA loop. Hydrogen bonds between main chain amide groups and phosphates of DNA have been observed in the 434 phage repressor−DNA complex (Aggarwal *et al.*, 1988). Ethylation protection of the phosphate backbone in this region could be accounted for by those hydrogen bonds. In the 5′ stem, phosphates lying between 5′ to C59 and 5′ to C64 are protected from ethylation, although the protection of the first and last phosphates is very weak. In our model the phosphate backbone in this region is wedged between the upper and lower loops and the protection from ethylation of this region is therefore well accounted for. In this region, the Lys22→Gln, Lys23→Gln, Arg47→Gln, Lys50→Gln and Arg52→Gln mutations reduce the U1 RNA binding to varying extents (Nagai *et al.*, 1990), suggesting that these residues hold the 5′ stem between the jaws. Mutations of Lys80 and Arg83 also reduce the binding to U1 RNA. This result is consistent with our model in which these residues are also near the phosphate backbone of the 5′ stem and could stabilize the RNA binding by electrostatic interaction. In the crystal structure of A95, residues beyond 91 are disordered and cannot be located. Our mutagenesis experiments as well as the deletion experiments by Scherly *et al.* (1989) and Lutz-Freyermuth *et al.* (1990) show that Lys96 and Lys98 are important and that they may also increase the RNA binding by electrostatic interactions.

Considering the surface and charge complementarity of the U1 A protein and U1 RNA and the results of our ethylation protection experiments, our model seems plausible (Figure 6A and B). Are there any hydrogen bonds between the bases in the loop and the protein side chains? The complete inhibition of binding by the Thr11→Val, Asn15→Val and Arg52→Gln mutations and its partial inhibition by the Tyr13→Phe and Gln54→Phe mutations suggests that these residues may form critical hydrogen bonds with U1 RNA. What do they interact with? If they interact with specific RNA bases, then replacement of these bases should equally inhibit the complex formation. We found, however, that none of the C⟷U replacements in the loop II affected the binding to the U1 A protein significantly and none of the G⟷A replacements in the loop led to complete loss of binding. Since C and U share the same hydrogen bonding group at only the 2 keto-oxygen and A and G at only the N7 and N3 positions (Figure 5C), these results suggest that either these bases do not form hydrogen bonds with the protein, or if they do, the base replacements still preserve hydrogen bonds at the conserved positions which could prevent total loss of specific binding. Based on our experiments and model, we suggest that Thr11 may form a hydrogen bond with the phosphate group 5′ to C70. This phosphate is at the sharp turn, and a hydrogen bond to this phosphate may be critical for the conformation of the RNA loop. Based on UV cross-linking experiments, Merrill *et al.* (1988) proposed ring stacking interactions between bases of RNA and the conserved aromatic residues of RNP1 and RNP2; the fluorescence quenching experiment by Sachs *et al.* (1987) is consistent with this proposal. Here U68 and U67 may interact with Phe56 and Tyr13 by ring stacking.

In some DNA−protein complexes, hydrogen bonds form between N7 and the NH_2 group of glutamine, and between the N6 amine of adenine and the carbonyl group of the glutamine side chain. In the λ and 434 phage repressors a Gln side chain that is hydrogen bonded to adenine also forms a hydrogen bond with another Gln side chain (Jordan and Pabo, 1988; Aggarwal *et al.*, 1988). A similar hydrogen bonding network may exist in the U1 A protein, where a hydrogen bond between Tyr13 and Gln54 could stabilize the hydrogen bonding between A66 and Gln54. When A66 is replaced by G, the hydrogen bond to N7 can still form but a hydrogen bond is no longer possible to the O6 position, and this change may account for the modest (10-fold) reduction in binding. In the 3′ two thirds of the loop, where phosphates are in contact with the β sheet, the RNA bases are probably not in direct contact with the protein. For example if C72 were in contact with the protein, the large change in volume accompanying the C72→G replacement would be expected to have a larger effect on protein binding. The strong protection of phosphates at C72 and A71 from ethylation indicates that the phosphate backbone rather than the bases are in contact with the protein at these positions. A reduction of ~10-fold in binding upon C72→G, G69→A and A71→G could be accounted for by changes in the phosphate backbone conformation.

How does U2 RNA bind to the U2 B′′ protein? U2 RNA binds to the U2 B′′ protein only in the presence of the U2 A′ protein. When residues 40–49 of the U1 A protein were replaced by the corresponding segment of the U2 B′′ protein, the U2 A′ protein bound weakly to the mutated U1 A protein, and two additional mutations, Asp24→Glu and Lys28→Arg enhanced the binding of the U2 A′ protein. Based on these results (Scherly *et al.*, 1990b), we proposed that the U2 A′ protein interacts with the β2 strand and the A helix of the U1 A protein (Nagai *et al.*, 1990). As shown in Figure 6B, this U2 A′ protein binding site is fully accessible in our model of the complex. U2 RNA has an insertion of A between U73 and C74, and this insertion will disrupt the interaction between the 3′ half of the loop and the β2 strand, so that this RNA can no longer bind the U1 A protein (Scherly *et al.*, 1990a). The binding of the U1 A′ protein to the β2 strand and A helix of the U2 B′′ protein would probably have two effects: it may induce structural changes of the U2 B′′ protein or it may introduce additional interactions between the U2 A′ protein and the 3′ half of the loop and 5′ stem of the RNA, so that the binding of the U2 RNA to the U2 B′′ protein could be stabilized.

Our crystal of the A95 protein (residues 1–95) contains two molecules in each asymmetric unit, and the interface between the two molecules related by the non-crystallographic dyad is tightly packed with hydrophobic amino acids; we therefore suspected that these dimers may be stable enough to exist in solution. However, gel filtration experiments show that the A102 protein is monomeric in solution (unpublished results). In the model of our complex with U1 RNA, this dimer interface is free; it may possibly be used for interactions with other proteins, such as the C-terminal RNP domain of the U2 B′′, in spliceosomal assembly.

Our mutagenesis and ethylation protection experiments led us to a model of the U1 A protein–U1 RNA complex which accounts for most biochemical properties of the U1 A and U2 B′′ proteins. Upon binding to aminoacyl tRNA synthetase, the anti-codon loop of glutaminyl tRNA undergoes a substantial structural change, and bases are splayed out to form extensive hydrogen bonds with the protein (Rould *et al.*, 1990). The stem–loop II of U1 RNA and the U1 A protein may also undergo conformational changes upon complex formation, but our model outlines at least the gross features of the likely interactions between the U1 A protein and U1 RNA stem–loop II. The base replacement experiment in the loop II suggests that only few base-specific hydrogen bonds may link the RNA to the protein, and that interactions with the backbone phosphates, ring stacking interactions between bases and aromatic amino acid residues and van der Waals interactions between the β sheet and RNA loop may determine specificity. The detailed interactions between the U1 A protein and U1 RNA remain to be determined by X-ray analysis of a crystalline complex.

Materials and methods

RNA synthesis

The 46mer of the U1 RNA stem–loop II containing nucleotides 50–92 with three additional guanines at the 5′ end was synthesized by *in vitro* transcription with T3 RNA polymerase (Morris *et al.*, 1986) from a linearized plasmid carrying the corresponding cDNA. The RNA was labelled at the 5′ end using polynucleotide kinase and purified on a denaturing polyacrylamide gel.

The 22mers of the U1 RNA stem–loop II containing nucleotides 62–79 plus two additional GC basepairs at its end were obtained by the method of Milligan (1987) using [α-^{32}P]UTP and T3 RNA polymerase (Morris *et al.*, 1986); the transcription products were purified on a denaturing polyacrylamide gel.

The 22mer RNA with wild-type sequence (nucleotides 60–81) was chemically synthesized by the phosphoramidite method (Beaucage and Caruthers, 1981) with tertbutyldimethylsilyl-2′-hydroxyl protection (Usman *et al.*, 1987) (monomers from Milligen/Biosearch, synthesizer: ABI 380B). Cleavage from the controlled pore glass support and base deprotection were achieved with ethanolic ammonia solution. Deprotection of the 2′-OH group was performed with a 1 M solution of tetrabutylammonium fluoride in tetrahydrofuran (Aldrich Chemical Co.) for 24 h in the dark, the reaction was quenched with 0.1 M triethylammoniumacetate pH 7.0 (Gait *et al.*, 1991). The desalted RNA was lyophilized and resuspended in 10 mM Tris–HCl, 0.1 mM EDTA, pH 8.0. An aliquot was labelled at the 5′-end using [γ-^{32}P]ATP and polynucleotide kinase before purification on a denaturing polyacrylamide gel.

Preparation of the U1 A protein and its mutants

Various mutants of human U1 A protein were produced in *Escherichia coli* using the T7 RNA polymerase expression vector (Studier *et al.*, 1990) and purified on a CM–Sepharose column as described previously (Nagai *et al.*, 1990).

RNA–protein binding assays

For a mobility-shift experiment ^{32}P-labelled RNA (~5 nM) was incubated with various concentrations of protein in 10 mM Na-HEPES (pH 7.4), 50 mM KCl and 1 mM $MgCl_2$ at room temperature for 20 min. The complex was separated from unbound RNA on a native 12% polyacrylamide gel containing 100 mM Tris–borate (pH 8.3), 1 mM EDTA and 0.1% Triton X-100; the wild-type protein (A102) was always used as an internal control. The gel was autoradiographed without any further treatment for 4–16 h at −70°C, and the dissociation constants (K_D) were estimated from the proportion of complexed and free RNA.

A filter binding assay was performed using the protocol kindly provided by Dr W.Boelens. 90 μl buffer B (10 mM Tris–HCl pH 7.4, 100 mM KCl, 2 mM $MgCl_2$, 0.1 mM EDTA, 0.5 mM dithioerythritol) containing 10 μg tRNA and 5′-^{32}P-labelled U1 RNA stem–loop II (activity: ~10^4 Cherenkov c.p.m.) were mixed with 10 μl buffer A (20 mM K-HEPES pH 7.9, 100 mM KCl, 2 mM $MgCl_2$, 5% glycerol, 0.5 mM dithioerythritol, 0.5 mg/ml bovine serum albumin) containing various concentrations of A102 protein (between 0.25 and 16 μM) and incubated for 90 min at room temperature. The mixture was passed through a dot-blot manifold (Schleicher & Schüll, SRC 96) containing a nitrocellulose filter (Schleicher & Schüll, BA 85) equilibrated with buffer B. The samples were washed twice with 200 μl buffer B, the membrane was dried at room temperature and the radioactivity of the single dots containing bound RNA was quantified in a scintillator as Cherenkov counts. K_D was calculated from the protein concentration necessary for binding 50% of the RNA.

Ethylation protection experiments and RNase V1 digest

The ethylation protection experiments were performed essentially as described by Romby *et al.* (1985). 5′-labelled U1 RNA stem–loop II (nucleotides 50–92) was incubated with truncated and full-length A protein in 45 μl alkylation buffer (150 mM sodium cacodylate–HCl pH 8.0, 5 mM $MgCl_2$, 0.3 mM EDTA). After 20 min at room temperature the sample was split into two 22.5 μl aliquots; 2.5 μl freshly prepared saturated solution of ethylnitrosourea in 100% ethanol was added to the first tube and 2.5 μl of 100% ethanol to the second (control experiment). The mixtures were incubated at room temperature in the dark. A 200-fold excess of protein over RNA was used in order to dilute out any chemical modifications of the protein. After 3 h the protein-containing samples were phenol extracted, 10 μg of tRNA was added to each sample and the RNA precipitated twice by ethanol. Cleavage of the phosphotriester bonds was achieved by incubation in 100 mM Tris–HCl pH 9.0 at 50°C for 5 min, the resulting RNA fragments were analysed on a 20% denaturing polyacrylamide gel. The assignment of the bands was performed by comparison with partial RNase T1 digest and alkaline hydrolysis of the U1 RNA stem–loop II.

In a typical RNase V1 digest experiment the U1 RNA stem–loop II (nucleotides 50–92, ~10 nM) labelled at the 5′ end was treated with 0.05 units of the enzyme (Pharmacia) in 6 μl 25 mM Tris–HCl pH 7.4, 200 mM NaCl, and 10 mM $MgCl_2$ in the presence of 1 μg tRNA at 37°C for 10 min. The digest was stopped by adding 4 μl formamide dye and freezing of the sample. In the case of protection experiments the RNA was incubated with the U1 A protein (2 μM end concentration) for 20 min at room temperature (UK) prior to addition of the RNase. The resulting RNA fragments were analysed as described above. To ensure proper binding and

to exclude any buffer effects the same digest was carried out in 10 mM Na-HEPES, 50 mM KCl and 10 mM $MgCl_2$, hence conditions almost identical to that used in the mobility-shift assay.

Acknowledgements

We thank Drs M.F.Perutz, P.R.Evans, W.Sundquist, J.Karn, R.Turner, S.Phillips and B.Luisi for critical reading of the manuscript, Drs E.Westhof and A.Krol for the coordinates of U1 RNA stem−loop II, Dr W.Boelens for the filter binding assay protocol, and J.Fogg and T.Smith for oligonucleotides. This work was supported by the Medical Research Council, NIH and HFSP grant. Timm-H.Jessen was supported by the Butenandt fellowship of the Max-Planck Gesellschaft.

References

Aggarwal,A.K., Rodgers,D.W., Drottar,M., Ptashne,M. and Harrison,S.C. (1988) *Science*, **242**, 899−907.
Bach,M., Krol,A. and Lührmann,R. (1990) *Nucleic Acids Res.*, **18**, 449−457.
Bandziulis,R.J., Swanson,M.S. and Dreyfuss,G. (1989) *Genes Dev.*, **3**, 431−437.
Beaucage,S.L. and Caruthers,M.H. (1981) *Tetrahedron Lett.*, **22**, 1859−1862.
Carter,C.W. and Kraut,J. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 283−287.
Church,G.M., Sussman,J.L. and Kim,S.H. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 1458−1462.
Dreyfuss,G., Swanson,M.S. and Pinol-Roma,S. (1988) *Trends Biochem. Sci.*, **13**, 86−91.
Ehresmann,C., Baudin,F., Mougel,M., Romby,P., Ebel,J.-P. and Ehresmann,B. (1987) *Nucleic Acids Res.*, **15**, 9109−9128.
Gait,M.J., Pritchard,C.E. and Slim,G.C. (1991) In Eckstein,F. (ed.) *Oligonucleotides and Analogues: A Practical Approach*. IRL Press at Oxford University Press, in press.
Habets,W.J., Sillekens,P.T.G., Hoet,M.H., Schalken,J.A., Roebroek,A.J.M., Leunissen, J.A.M., van de Ven,W.J.M. and van Venrooij,W.J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 2421−2425.
Hamm,J., Kazmaier,M. and Mattaj,I.W. (1987) *EMBO J.*, **6**, 3479−3485.
Hamm,J., Dathan,N.A., Scherly,D. and Mattaj,I.W. (1990) *EMBO J.*, **9**, 1237−1244.
Jordan,R.S. and Pabo,C.O. (1988) *Science*, **242**, 893−899.
Kim,S.H., Suddath, F.L., Quigley, G.J., McPherson,A., Sussman,J.L., Wang, A.H.J., Seeman,N.C. and Rich,A. (1974) *Science*, **185**, 435−440.
Knapp,G. (1989) *Methods Enzymol.*, **180**, 192−212.
Krol,A., Westhof,E., Bach,M., Lührmann,R., Ebel,J.-P. and Carbon,P. (1990) *Nucleic Acids Res.*, **18**, 3803−3811.
Krupp,G. and Gross,H.J. (1979) *Nucleic Acids Res.*, **6**, 3481−3490.
Lockard,R.E. and Kumar,A. (1981) *Nucleic Acids Res.*, **9**, 5125−5140.
Lührmann,R., Kastner,B. and Bach,M. (1990) *Biochim. Biophys. Acta*, **1087**, 265−292.
Lutz-Freyermuth,C., Query,C.C. and Keene,J.D. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 6292−6397.
Mattaj, I.W. (1990) *Curr. Opin. Cell. Biol.* **2**, 528−538.
Maniatis,T. and Reed,R. (1987) *Nature*, **325**, 673−678.
Merrill,B.M., Stone,K.L., Cobianchi,F., Wilson,S.H. and Williams,K.R. (1988) *J. Biol. Chem.*, **263**, 3307−3313.
Milligan,J.F., Groebe,D.R., Witherell,G.W. and Uhlenbeck,O.C. (1987) *Nucleic Acids Res.*, **15**, 8783−8798.
Moras,D., Comarmond,M.B., Fischer,J., Weiss,R., Thierry,J.C., Ebel,J.P. and Giege, R. (1980) *Nature*, **288**, 669−674.
Morris, C.E., Klement, J.F. and McAllister, W.T. (1986) *Gene*, **41**, 193−200.
Nagai,K., Oubridge,C., Jessen,T.-H., Li,J. and Evans,P.R. (1990) *Nature*, **348**, 515−520.
Patton,J.R. and Pederson,T. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 747−751.
Query,C.C., Bentley,R.C. and Keene,J.D. (1989) *Cell*, **57**, 89−101.
Robertus,J.D., Ladner,J.E., Finch,J.T., Rhodes,D., Brown,R.S., Clark,B.F.C. and Klug,A. (1974) *Nature*, **250**, 546−551.
Romby,P., Moras,D., Bergdoll,M., Dumas,P., Vlassov,V.V., Westhof,E., Ebel,J.P. and Giege,R. (1985) *J. Mol. Biol.*, **184**, 455−471.
Rould,M.A., Perona,J.J., Söll,D. and Steitz,T.A. (1989) *Science*, **246**, 1135−1142.
Ruby, S.W. and Abelson, J. (1991) *Trends Genet.* **7**, 79−85.
Sachs,A.B., Davis,R.W. and Kornberg,R. (1987) *Mol. Cell. Biol.*, **7**, 3268−3276.
Scherly,D., Boelens,W., van Venrooij,W.J., Dathan,N.A., Hamm,J. and Mattaj,I.W. (1989) *EMBO J.*, **8**, 4163−4170.
Scherly,D., Boelens,W., Dathan,N.A., van Venrooij,W.J. and Mattaj,I.W. (1990a) *Nature*, **345**, 502−506.
Scherly,D., Dathan,N.A., Boelens,W., van Venrooij,W.J. and Mattaj,I.W. (1990b) *EMBO J.*, **9**, 3675−3681.
Sharp,P. (1987) *Science*, **235**, 766−771.
Sillekens,P.T.G., Habets,W.J., Beijer,R.P. and van Venrooij,W.J. (1987) *EMBO J.*, **6**, 3841−3848.
Steitz,J.A., Black,D.L., Gerke,V., Parker,K.A., Kramer,A., Frendewey,D. and Keller,W. (1988) In Birnstiel,M.L. (ed.) *Structure and Function of Major and Minor snRNPs*. Springer Verlag, New York. pp. 115−154.
Steitz,T.A. (1990) *Quart. Rev. Biophys.*, **23**, 205−280.
Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.*, **185**, 60−89.
Swanson,M.S., Nakagawa,T.Y., LeVan,K. and Dreyfuss,G. (1987) *Mol. Cell. Biol.*, **7**, 1731−1739.
Usman,N., Ogilvie,K.K., Jiang,M.-Y. and Cedergreen,R.J. (1987) *J. Am. Chem. Soc.*, **109**, 7845−7854.

Received on May 21, 1991; revised on July 5, 1991