The EMBO Journal vol.12 no.13 pp.5191-5200, 1993

A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A protein is required for regulation of polyadenylation

Celia W.G.van Gelder, Samuel I.Gunderson[1], Eric J.R.Jansen, Wilbert C.Boelens[1], Maria Polycarpou-Schwarz[1], Iain W.Mattaj[1] and Walther J.van Venrooij

University of Nijmegen, Department of Biochemistry, PO Box 9101, 6500 HB Nijmegen, The Netherlands and [1]European Molecular Biology Laboratory, Gene Expression Programme, Meyerhofstrasse 1, D-69117 Heidelberg, Germany

Communicated by I.W.Mattaj

The human U1A protein-U1A pre-mRNA complex and the relationship between its structure and function in inhibition of polyadenylation *in vitro* were investigated. Two molecules of U1A protein were shown to bind to a conserved region in the 3′ untranslated region of U1A pre-mRNA. The secondary structure of this region was determined by a combination of theoretical prediction, phylogenetic sequence alignment, enzymatic structure probing and molecular genetics. The U1A binding sites form (part of) a complex secondary structure which is significantly different from the binding site of U1A protein on U1 snRNA. Studies with mutant pre-mRNAs showed that the integrity of much of this structure is required for both high affinity binding to U1A protein and specific inhibition of polyadenylation *in vitro*. In particular, binding of a single molecule of U1A protein to U1A pre-mRNA is not sufficient to produce efficient inhibition of polyadenylation.

Key words: polyadenylation/RNA-protein interaction/RNA structure/U1 snRNP/U1A protein

Introduction

The removal of introns from pre-messenger RNA, known as splicing, is an important process in which several small ribonucleoprotein particles (snRNPs) participate. One of them, U1 snRNP, interacts with the pre-mRNA by a mechanism that includes pairing between bases at the 5′ end of U1 snRNA and sequences located at the 5′ splice site. U1 snRNPs contain at least eight proteins (B′, B, D1, D2, D3, E, F and G), which also occur in other U snRNPs, and three U1-specific proteins named 70K, U1C and U1A (Lührmann *et al.*, 1990). The U1A protein binds directly to the second stem-loop of U1 snRNA (Scherly *et al.*, 1989; Lutz-Freyermuth *et al.*, 1990). The protein contains two RNP motifs, of which the N-terminal copy is responsible for binding to U1 snRNA (Scherly *et al.*, 1989; Lutz-Freyermuth *et al.*, 1990; Nagai *et al.*, 1990; Jessen *et al.*, 1991; Hall and Stump, 1992). The structure of this domain of the U1A protein has been determined by X-ray crystallography and NMR studies (Nagai *et al.*, 1990; Hoffman *et al.*, 1991) and consists of a four-stranded antiparallel β-sheet with two α-helices lying on the same side of the sheet.

The loop of the hairpin to which U1A binds has the sequence AUUGCACUCC. It has been shown that the first seven nucleotides, AUUGCAC, which are highly conserved between U1 snRNAs from various species, are critical for specific U1A protein binding, while the structural context of this sequence affects binding affinity (Scherly *et al.*, 1989, 1990; Bentley and Keene, 1991; Tsai *et al.*, 1991). If the loop sequence of stem-loop II of U1 snRNA is present in the absence of a stable stem, the affinity for the U1A protein drops (Scherly *et al.*, 1989; Tsai *et al.*, 1991). Quantitative mobility shift assays of the loop sequence of stem-loop II, present either in a linear structural context or in a hairpin structure with a loop larger than that found in U1 snRNA, showed an ~100-fold reduction in binding affinity for the U1A protein relative to the wild type stem-loop II (Tsai *et al.*, 1991). RNase protection experiments on U1 snRNP particles showed that both the loop sequence and ~5 bp of the stem are protected by the U1A protein (Bach *et al.*, 1990). Bound U1A protein also protects several 5′ stem phosphates, as well as some loop phosphates, against ethylation by ethylnitrosourea (Jessen *et al.*, 1991).

Recently it has been shown that the 3′ untranslated region (UTR) of the U1A pre-mRNA contains a region which has been conserved between vertebrate species (Boelens *et al.*, 1993). This region contains two stretches of seven nucleotides, one of which is identical to the seven nucleotides of the U1 snRNA loop mentioned above, while the other is the same in six out of seven positions. These sequences will be referred to as Box 1 and Box 2 respectively in this paper, with Box 1 being the more 5′ of the two. Boxes 1 and 2 are located in close proximity to the cleavage and polyadenylation signal. The distance between the two boxes is conserved, as is the distance from Box 2 to the polyadenylation signal.

It was demonstrated that binding of U1A protein to this region of the U1A pre-mRNA, which depends upon these U1 snRNA-like sequences, causes inhibition of polyadenylation of the U1A pre-mRNA (Boelens *et al.*, 1993). Although it was not determined how many molecules of U1A protein were bound to each pre-mRNA, the number was shown to be greater than one.

In the experiments reported here, the structure of the U1A-binding region of the pre-mRNA was investigated by a variety of techniques and the number of protein molecules bound was determined. Further, the structural characteristics of the U1A protein-U1A pre-mRNA complex were examined in relation to its function in inhibition of polyadenylation. The U1 snRNA-like sequences are shown to form parts of two asymmetric internal loops present in a complex secondary structure. This part of the U1A pre-mRNA is compared with stem-loop II of U1 snRNA, the other RNA structure to which U1A protein is known to bind specifically.

Results

Two molecules of U1A protein bind to the pre-mRNA

It was previously shown (Boelens *et al.*, 1993) that more than one molecule of U1A protein is able to bind to each U1A pre-mRNA. To determine the exact number of bound protein molecules, we adapted an assay often used to examine DNA–protein complexes (Hope and Struhl, 1987). Two differently sized U1A protein derivatives that bind to U1 snRNA with similar affinity (Lutz-Freyermuth *et al.*, 1990; Nagai *et al.*, 1990) were produced in, and purified from, *Escherichia coli*. These were full-length U1A (Awt) protein and a fragment of U1A containing the N-terminal 101 amino acids (A101). The two proteins were allowed to bind to a region of the U1A pre-mRNA (the Ag fragment) shown to be necessary and sufficient for U1A binding (Boelens *et al.*, 1993) and the resultant complexes analyzed by native gel electrophoresis. The Ag fragment contains the human U1A pre-mRNA sequences shown in Figure 2A plus 33 nt of 3′ flanking sequence from the U1A gene and 8 nt of 5′ flanking sequence derived from the cloning vector.

The position of unbound Ag RNA after native gel electrophoresis is shown in lane 1 of Figure 1. Addition of either A101 or Awt protein (represented by empty and filled squares respectively) results in the appearance of two retarded complexes (lanes 2 and 4) suggestive of binding of either one or two proteins to the RNA. The differential requirement for Awt and A101 proteins in complex formation was probably due to the fact that much of the Awt protein in this particular preparation was not competent in RNA binding, since other preparations of Awt exhibited greater RNA binding capacity (data not shown). No additional intermediate complexes were seen when less protein was added (data not shown) while increasing the amount of either protein resulted in disappearance of both free RNA and the lower of the two RNA–protein complexes with a concomitant increase in the upper complex (lanes 3 and 5). Next, the two U1A derivatives were mixed before RNA binding (lanes 6 and 7). In addition to the two previously detected, slowly migrating complexes (cf. lanes 3 and 5) a single additional complex of intermediate mobility was seen. The lack of additional intermediate complexes indicates that two, and not more, molecules of U1A bind to each RNA.

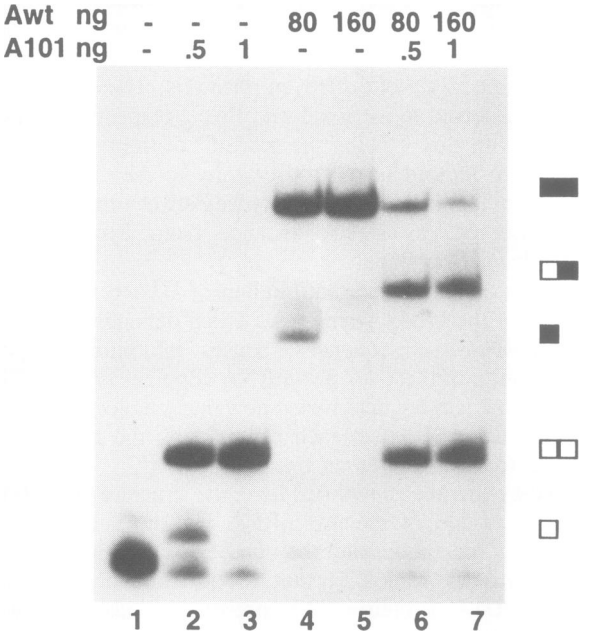
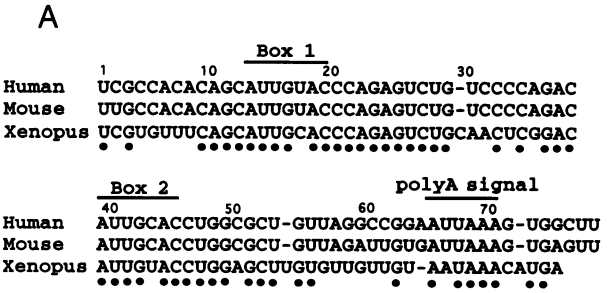


Fig. 1. Two molecules of U1A protein bind to each pre-mRNA. ^{32}P-labelled Agwt RNA was incubated without U1A protein (lane 1), with A101 protein (lanes 2–3), with U1Awt protein (lanes 4–5) or with a mixture of A101 and U1Awt (lanes 6–7). The amount of protein added is indicated above the lanes. The boxes on the right represent the protein components of the complexes. Filled boxes are U1Awt and empty boxes A101.

Sequence alignment and structure prediction

While the entire 3′ UTRs of the human and mouse U1A mRNAs are very similar (79% identical), the only region of high conservation of both with the *Xenopus* U1A mRNA sequence starts ~55 nt upstream of the A(A/U)UAAA cleavage and polyadenylation signals (Figure 2A) (Sillekens *et al.*, 1987; Scherly *et al.*, 1991; M.Bennett and J.Craft, personal communication). The entire region encompassing Boxes 1 and 2 and the cleavage and polyadenylation signal (Figure 2A) is 73% identical between human and *Xenopus* and 93% identical between human and mouse. The spacing between Box 2 and the polyadenylation signal is also identical for the three sequences. This localized sequence conservation and the fact that the Ag fragment has an affinity for U1A protein indistinguishable from that of the entire U1A pre-



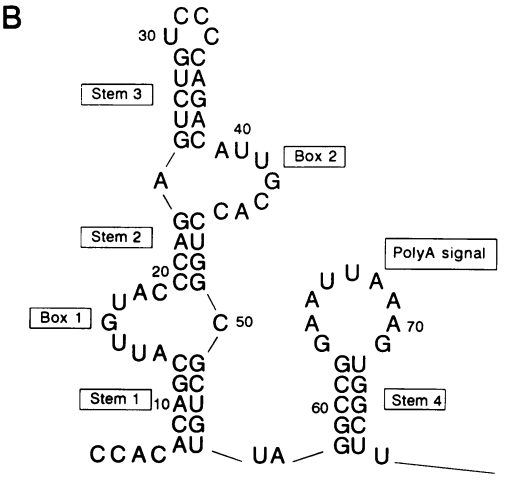


Fig. 2. (**A**) Sequence alignment of the conserved part of the 3′ UTR sequences of human, mouse and *Xenopus laevis* U1A pre-mRNAs. Nucleotides that are identical in all three sequences are marked and the Box 1, Box 2 and polyadenylation sequences are indicated. No sequences 3′ to those shown are available from either U1A cDNAs or the U1A gene of *Xenopus*. (**B**) Proposed secondary structure of the 3′UTR of the human U1A pre-mRNA. The Box 1 and 2 sequences, the cleavage and polyadenylation signal and stems 1, 2, 3 and 4 are indicated.

mRNA (see below), led us to expect that the Ag fragment would fold similarly either alone or in the context of the complete pre-mRNA.

Optimal and suboptimal foldings were calculated for the complete human and *Xenopus* U1A pre-mRNAs, for the 3′ UTR sequences of the cDNAs and for segments of these 3′ UTRs, using the FOLD and MFOLD programs (Jaeger *et al.*, 1990; Zuker *et al.*, 1991). In the majority of the predicted low-energy structures the Box 1 and 2 sequences are partially or completely single-stranded and are separated by a phylogenetically conserved stem−loop. Several possible structures exist for the sequences flanking the boxes, especially for the region which contains the cleavage and polyadenylation signal.

Secondary structure models were derived by combining phylogenetic and free energy data. The version in Figure 2B is that for the human U1A mRNA. The model consists of two distinct parts. The 5′ part contains three stems (numbered 1, 2 and 3), separated by two asymmetric internal loops containing the Box 1 and 2 sequences. A single unpaired nucleotide is present on the strand opposite each box sequence.

Comparison with the *Xenopus* U1A mRNA sequence provides support for the 5′ part of the model between A10 and U53 (Figure 3B; the conserved region lies between the large arrows). All non-conserved nucleotides in this region are at unpaired positions with the exception of the A35 to G change, but this difference replaces an A-U pair with G-U. The extra nucleotide that is inserted in the *Xenopus* sequence is located in the terminal loop. Outside of this region phylogenetic support for the model is weak. Stem 1 in the *Xenopus* sequence is only 3 bp long. Further, although it is possible to draw hairpin structures in which the cleavage and polyadenylation signals are in loops, these

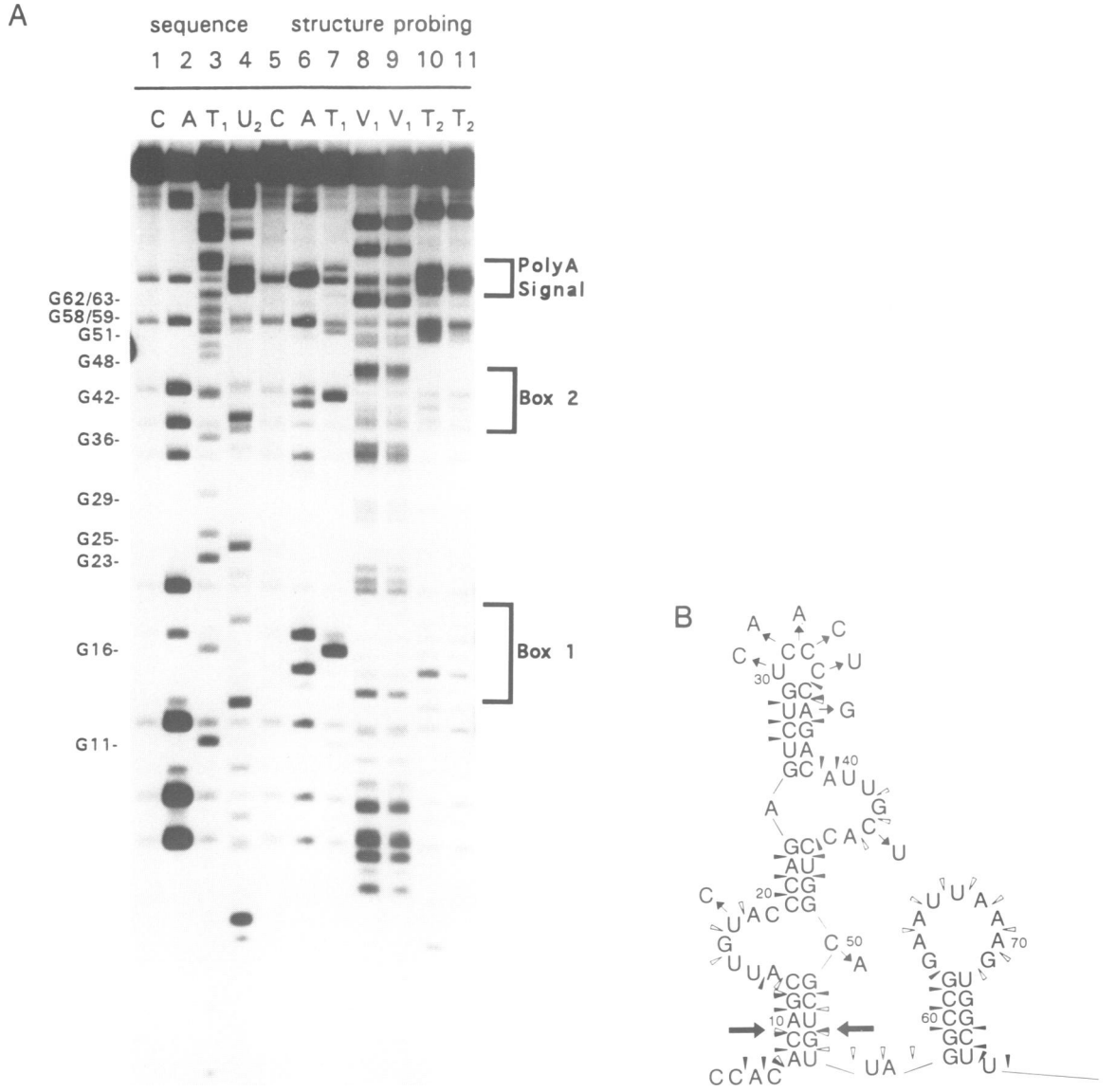


Fig. 3. (**A**) Enzymatic digestions of the Ag RNA under denaturing (sequence) and native (structure probing) conditions. RNA samples were treated as described in Materials and methods. The samples in lanes 1 and 5 are control reactions, to which no enzyme was added. Lanes 2−4 contain reactions under denaturing conditions (the enzymes used are indicated) while lanes 6−11 designate reactions under native conditions (two concentrations were used for RNases V1 and T2). The positions of guanosines cleaved by RNase T1 under denaturing conditions are indicated on the left. (**B**) The secondary structure of the human Ag RNA sequence. Consensus data from several independent experiments (only strong cuts) are shown. RNase V1 cleavage is indicated by closed triangles, RNase A/T1/T2 cleavage with open triangles. For the most conserved part of the structure (nt 10−53, indicated with large arrows), the nucleotide changes in the corresponding *Xenopus* RNA are indicated (small arrows).

are not well conserved with respect to the human structure in either the *Xenopus* or mouse U1A pre-mRNAs. One interesting aspect of the *Xenopus* sequence is that the non-conserved nucleotides in the two boxes are reversed in position (Figure 3B).

Enzymatic structure probing

To test the proposed structure, RNase digestions of 5′ end-labelled Ag fragments were carried out under native and denaturing conditions using RNases A, U2, T1, T2 and V1. A typical example of the results is shown in Figure 3A, while Figure 3B summarizes the results of several independent experiments. It can be seen that the central three nucleotides of the Box 1 and 2 sequences [nt 15–17 (UGU) in Box 1 and nt 41–43 (UGC) in Box 2] are cleaved efficiently by the enzymes T1, A and T2, which are known to cut 3′ of nucleotides present in single-stranded regions. In some experiments RNase T2 also appears to cut between other nucleotides in Boxes 1 and 2, but these cleavages were less reproducible. The terminal loop (nt 30–33) was almost never cut under native conditions, suggesting that its structure is very compact.

RNase V1 cuts, which indicate double-stranded or stacked bases, were clearly seen in the regions of stems 2 and 3 and, less reproducibly, in stem 1. In the latter stem, some positions were also cut by RNases A and T1. Therefore stem 1, if it exists, does not seem to be very stable under these conditions. RNase V1 cuts were found in the 5′ part of Box 2, which could point to some base stacking. The bulged nucleotides A24 and C50 were never cut under native conditions and are therefore probably located inside the helix. The polyadenylation signal is clearly single-stranded (RNase T2 cuts), flanked by double-stranded regions (RNase V1 cuts). The region between the 5′ and 3′ parts of the structure was efficiently cleaved by RNase T2, although in a few experiments weak V1 cuts were also found.

RNA mutants

To test the 5′ part of the structure more thoroughly and to obtain more information on the less conserved regions we next constructed mutations in the human Ag fragment. Single

Table I. Binding assays of 3′ UTR mutants

A. Stem mutants

	Sequences	Direct/indirect assay (150 mM)	Direct/indirect assay (500 mM)	Inhibition of polyadenylation
Agwt		+	+	+
3A	$_{25}$GUCUG GUCUC$_{34}$	+	–	–
3AB	$_{25}$CAGAG GUCUC$_{34}$	+	+	+
3CTD	see text	N.D.	+	+
2A	$_{20}$CCAG CCAG$_{46}$	+	–	–
2B	$_{20}$GGUC GGUC$_{51}$	+	–	–
2AB	$_{20}$GGUC CCAG$_{46}$	+	+	+
1A	$_{8}$ACAGC UCAGC$_{51}$	+	–	–
1B	$_{8}$AGUCG UGUCG$_{51}$	+	+	+
1AB	$_{8}$AGUCG UCAGC$_{51}$	+	+	+
4A	$_{58}$GAGAG UCGGU$_{72}$	+	N.D.	+
4B	$_{58}$GGCCG UUCU$_{72}$	+*	N.D.	+
4AB	$_{58}$GAGAG UUCUU$_{72}$	+*	N.D.	+

B. Loop mutants

	Sequences	Direct assay (150 mM)	Direct assay (500 mM)	Indirect assay (150 or 500 mM)	Inhibition of polyadenylation
Agwt		+	+	+	+
ΔB2	$_{39}$GGAUCCC$_{45}$	+	–	–	–
ΔB1	$_{13}$GGAUCCC$_{19}$	+	+	–	–
ΔB1/B2	both	–	–	–	–

N.D., not determined

*Binding properties of stem 4 mutants were established by using bandshift assays.

mutants (called 1A, 1B, 2A, 2B and 3A) were designed to disrupt each of the three 5′ stem structures by mutating individual strands of each putative helix (see Table I for mutant sequences). In the double mutants (1AB, 2AB and 3AB), which were designed to maintain complementarity and the putative structure, the sequences of both strands of each stem were interchanged. Further, mutant 3CTD was constructed, in which stem 3 and the terminal loop were replaced by CGGCGCUUCGGCGCCG. This sequence is predicted to form a stem composed of six GC base pairs with a highly stable tetraloop (Tuerk *et al.*, 1988). If the model were correct, this mutation should not disturb the U1A binding sites.

Enzymatic digestions, as described above, were performed on most of these mutant RNAs. The single mutants 2A and 3A clearly showed a distortion of the structure in the mutated region while the double mutants 2AB and 3AB had digestion patterns similar to the Agwt fragment (data not shown). The results with the stem 1 mutants were less easily interpretable; there was no clear difference between mutants 1B and 1AB, and, as with Agwt, both V1 and single-strand-specific enzymes cut in the stem 1 region of both these mutants.

Two single mutants and one double one were prepared in stem 4. In 4A and 4B the individual strands of the stem were mutated singly to disrupt the potential pairing and in 4AB the mutations were combined to restore pairing (see Table I). Nuclease digestion of mutant 4A suggested that this mutation disrupted stem 4. However, in the double mutant 4AB, in which stem 4 should reform, V1 cleavage was only partially restored (data not shown).

From these experiments we conclude that much of the proposed structure is likely to be correct although there is doubt about the existence, or at least the stability, of stem 1. The mutants could therefore be used to test the structural requirements for U1A protein binding and inhibition of polyadenylation. In addition to the mutants described above we also used the ΔB1, ΔB2 and ΔB1/2 mutants in which the sequences of Boxes 1 and 2 were altered individually or in combination (Boelens *et al.*, 1993; see Table IB).

Binding of U1A protein to the mutants

Two assays that can detect the binding of U1A protein to an RNA have been described previously (Boelens *et al.*, 1993). In the direct assay ^{35}S-labelled U1A protein is incubated with biotinylated RNA. Proteins that bind the RNA can be recovered via precipitation by Streptavidin–agarose and analysed by SDS–PAGE. In the indirect assay, which gives positive results only when at least two molecules of U1A protein are bound to each RNA, the ^{35}S-labelled U1A is precipitated via non-radioactive biotinylated U1A protein.

We tested the mutant Ag RNAs in these assays. As reported (Boelens *et al.*, 1993), both ΔB1 and ΔB2 can still bind U1A in the direct assay in the presence of 150 mM KCl, while ΔB1/2 cannot (Figure 4A, left panel). However, if the KCl concentration is increased to 500 mM, binding to the ΔB2 mutant, which only retains Box 1 and thus an imperfect match to the U1 snRNA sequence, is undetectable (Figure 4A, right panel). Previously, it was shown that ΔB1 and ΔB2 bind maximally one molecule of U1A protein (Boelens *et al.*, 1993).

To characterize further the binding to these mutants, and to define better the reduction in affinity of the ΔB2 mutant, the dissociation constants (K_D) of their binding to U1A protein were determined. First, the K_D of the complex

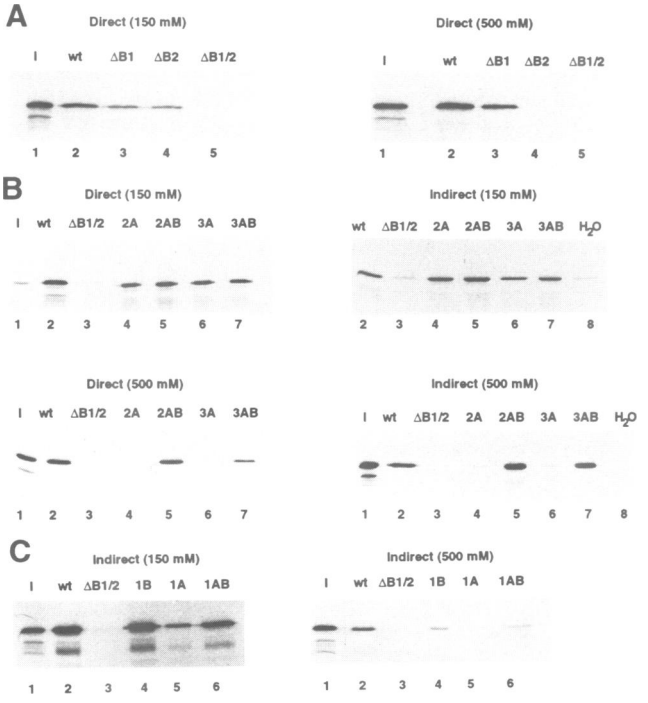


Fig. 4. (**A**) Binding of ^{35}S-labelled U1A protein (lanes 1, 10% of the input protein per assay) to various RNA substrates at 150 mM KCl (left panel) and 500 mM KCl (right panel). The RNAs used were Ag wt RNA (lane 2), mutant ΔB1 (lane 3), mutant ΔB2 (lane 4) and mutant ΔB1/B2 (lane 5). The mutants are from Boelens *et al.* (1993). (**B**) Binding assays for stem 2 and stem 3 mutants. Left panels: direct binding assay as described in panel A at either 150 mM (upper) or 500 mM (lower) KCl. Right panels: indirect binding assay. Precipitation of ^{35}S-labelled U1A protein via biotinylated U1A protein in the presence of various RNA substrates, as indicated above the lanes. (**C**) Binding assays for stem 1 mutants. Indirect binding assays carried out at 150 mM (left panel) or 500 mM (right panel) KCl.

Table II. Dissociation constants of various RNA–U1A protein complexes

RNA	K_D (M)	n
U1 RNA	$5(\pm 3) \times 10^{-11}$	4
U1A pre-mRNA	$10(\pm 6) \times 10^{-11}$	4
Ag	6×10^{-11}	1
ΔB1	$30(\pm 10) \times 10^{-11}$	3
ΔB2	$800(\pm 100) \times 10^{-11}$	2

n, number of independent determinations.

between U1 snRNA and the U1A protein was established. Under the conditions used (see Materials and methods) the K_D of this complex was $5(\pm 3) \times 10^{-11}$ M (Table II). This value is very similar to that (2×10^{-11} M) determined by Hall and Stump (1992), who also assayed binding with a nitrocellulose filter binding assay, but used a much shorter RNA substrate and different buffer conditions. Both of these values are considerably ($\sim 10^3$-fold) lower (i.e. indicative of tighter binding) than K_Ds determined for similar complexes measured by native gel electrophoresis (Lutz-Freyermuth *et al.*, 1990; Jessen *et al.*, 1991).

In our assay the human U1A pre-mRNA–U1A protein complex has a K_D of $10(\pm 6) \times 10^{-11}$ M. Taking into account the measured variation in the K_D values, U1 snRNA and the U1A pre-mRNA therefore exhibit comparable binding affinity. Note, however, that the K_D measured for the U1A pre-mRNA is complex since two U1A protein molecules bind to the pre-mRNA, and, in this assay, only one molecule has to be bound to score positive. In the single experiment carried out with the Ag fragment, the K_D was indistinguishable from those of either the pre-mRNA or U1 snRNA (Table II). The ΔB1 mutant showed an ~3-fold lower binding affinity [$K_D = 30(\pm 10) \times 10^{-11}$ M] than the wild type (wt) pre-mRNA. In the case of the ΔB2 mutant, which only contains the imperfect Box 1 binding sequence, the binding affinity decreased by a factor of ~80 [$K_D = 800(\pm 100) \times 10^{-11}$ M]. The affinity of the wt pre-mRNA is higher than the additive affinities of the two single-site mutants, indicating that there might be some cooperativity in the binding of the two U1A protein molecules. This conclusion was supported by electrophoretic mobility shift assays where, at protein concentrations at which low binding site saturation was achieved, the amount of U1A required to occupy both sites on an RNA was 2- to 4-fold greater than that required to occupy a single site (data not shown). The K_Ds of the two individual sites (Table II) would predict that, without cooperativity, ~30-fold more protein should be required.

Mutants affecting stems 1–3 of the structural model (Figure 2B) were next tested in the direct and indirect assays. Unexpectedly, mutants 2A, 2B and 3A, in which stems 2 or 3 were disrupted, could still bind U1A protein at 150 mM KCl in both the direct and indirect assays (Figure 4B, upper panels, lanes 4 and 6 and data not shown). Thus, disruption of either of the stems did not prevent interaction with U1A protein. When the assays were carried out at 500 mM KCl, however, it was evident that the affinity of the single mutants for U1A protein was reduced. Mutants 2A, 2B and 3A were incapable of interaction with even one molecule of U1A protein in these conditions (Figure 4B, lower panels, lanes 4 and 6 and data not shown). Restoration of stems 2 and 3 in the 2AB and 3AB double mutants restored U1A protein binding in both assays (Figure 4B, lanes 5 and 7). The 3CTD mutant, which contains a more stable terminal stem–loop, showed U1A protein binding comparable to that of wt U1A pre-mRNA (data not shown), providing further support for the presence of stem 3.

In the case of the stem 1 mutants, a less clear-cut result was obtained. At high, but not at low, salt concentration one of the single mutants, 1A, failed to bind U1A protein (Figure 4C, left and right panels, lanes 5), suggesting that stem 1 might be needed for protein binding. Mutant 1A showed wt behaviour in the direct assay at 150 mM salt, but did not detectably bind U1A protein in this assay at 500 mM salt (data not shown). The other single mutant (1B), on the other hand, as well as the double mutant (1AB), both showed behaviour comparable to that of wt pre-mRNA (Figure 4C, lanes 2, 4 and 6). One explanation for this behaviour might be that some of the base positions which are mutated in mutant 1A (nt 51–54) are necessary for U1A protein binding in the absence of a stem structure. It is, however, also possible that the 1A mutation causes changes in the structure to occur in high salt and thus affects U1A protein binding in a less direct way.

Both the single and double mutants of stem 4 bound U1A protein like the wt RNA (data not shown), indicating that this part of the structure is not necessary for U1A protein binding. Taken together, these results support the structural data summarized earlier, since they indicate that the highly conserved and stable stems 2 and 3 are important for high affinity U1A binding. The less conserved and less stable stem 1 is not required for U1A binding, as shown by mutant 1B, but the phenotype of the 1A mutant suggests that the stem might stabilize binding in some circumstances. Stem 4 is not needed for U1A protein binding.

Inhibition of polyadenylation

One functional consequence of U1A protein binding is inhibition of U1A pre-mRNA polyadenylation (Boelens *et al.*, 1993). The effects of the various mutations were therefore tested in an *in vitro* cleavage and polyadenylation assay. U1A protein addition to these assays results in specific inhibition of polyadenylation of the U1A wt substrate (Figure 5A, left panel). Considerably more U1A protein is required to inhibit polyadenylation of the double mutant ΔB1/B2, which cannot bind U1A specifically (Figure 5A, right panel). Polyadenylation of the two single mutants, ΔB1 and ΔB2 (Figure 5A, middle panels), is inhibited at a level of U1A protein only ~4-fold lower than that required for non-specific inhibition. This indicates that for efficient inhibition of the cleavage and polyadenylation reaction, it is crucial that two molecules of U1A protein can bind to the U1A pre-mRNA substrate.

The behaviour of the stem mutants in the polyadenylation inhibition assay closely mirrored their ability to bind U1A protein in 500 mM KCl in the binding assays described above. In the case of stems 2 and 3, the 2A and 3A single mutants, which are defective in U1A binding at high salt, behaved similarly to the ΔB1/2 double mutant (Figure 5B) while the 2AB and 3AB double mutants, in which stems 2 and 3 are restored, behaved similarly to the wt pre-mRNA. The 3CTD mutant also showed wt behaviour in polyadenylation inhibition (data not shown). The behaviour of the stem 1 mutants was also in agreement with the results

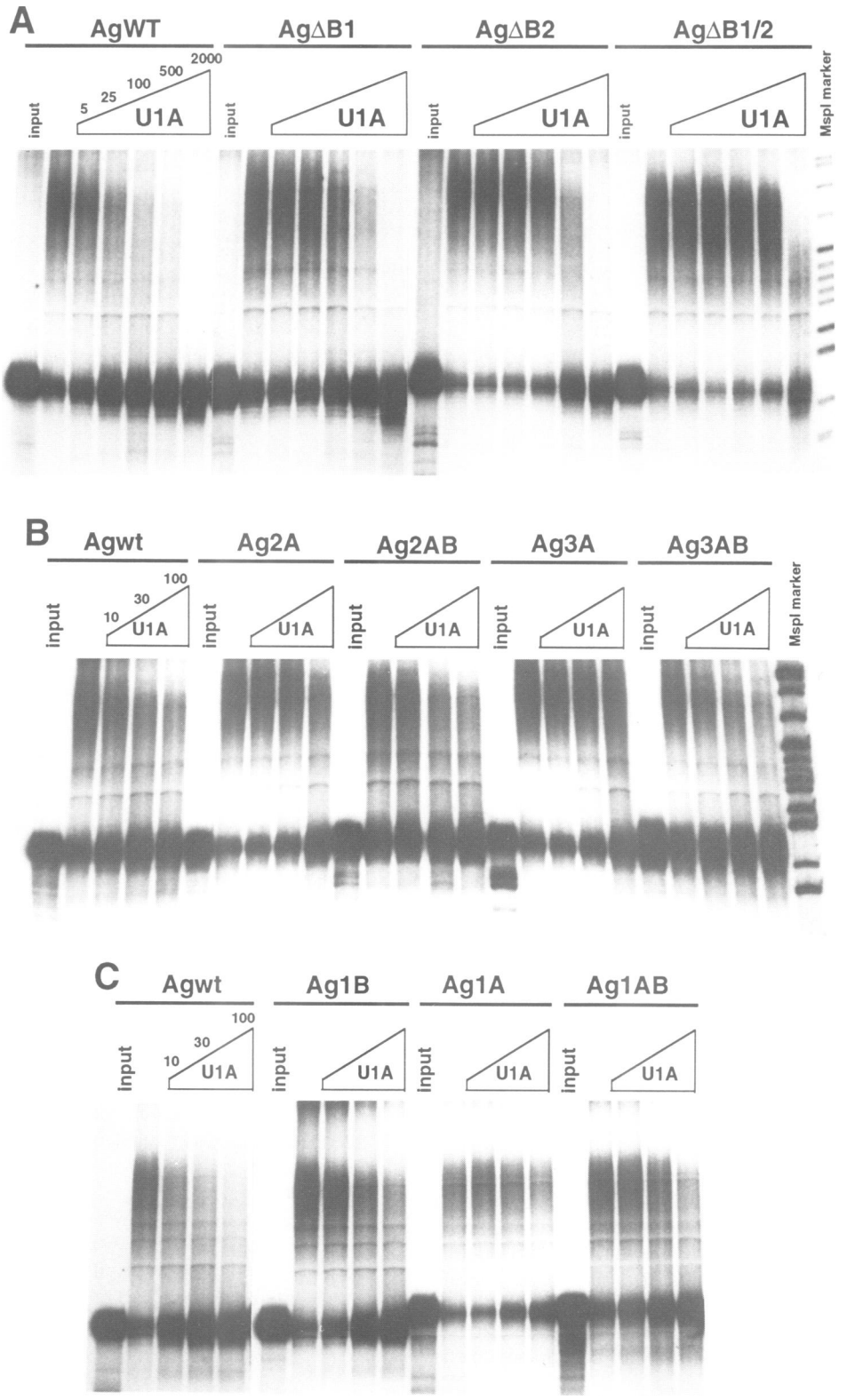


Fig. 5. Effect of U1A protein on *in vitro* polyadenylation of Agwt RNA and the 3′ UTR mutants. (**A**) Loop mutants. Recombinant, highly purified U1A protein was preincubated with the labelled RNA substrate for 5 min at room temperature. The 3′ processing reaction was initiated by addition of the reaction buffers and nuclear extract. The labelled RNA assayed is indicated above each panel. The first lane of each panel is the input precursor RNA in the absence of nuclear extract or U1A protein. The second lane of each panel is polyadenylation in the absence of exogenously added U1A protein. The remaining lanes of each panel show the effect of addition of increasing amounts of exogenous U1A protein with the amounts indicated in nanograms above each lane. The lane on the extreme right is a ^{32}P end-labelled *Msp*I digest of pBR322. (**B**) Stem 2 and 3 mutants. The type of labelled RNA used is indicated above each panel. The lanes of each panel are the same as described in panel A, except that the amounts of exogenously added U1A protein are different (ranging from 10 to 100 ng) as indicated above the panel. The lane on the extreme right is a ^{32}P end-labelled *Msp*I digest of pBR322. (**C**) Stem 1 mutants. The labelled RNA assayed is indicated above each panel. The lanes of each panel are the same as described in panel A, except that the amounts of exogenously added U1A protein are different (ranging from 10 to 100 ng) as indicated above the panel.

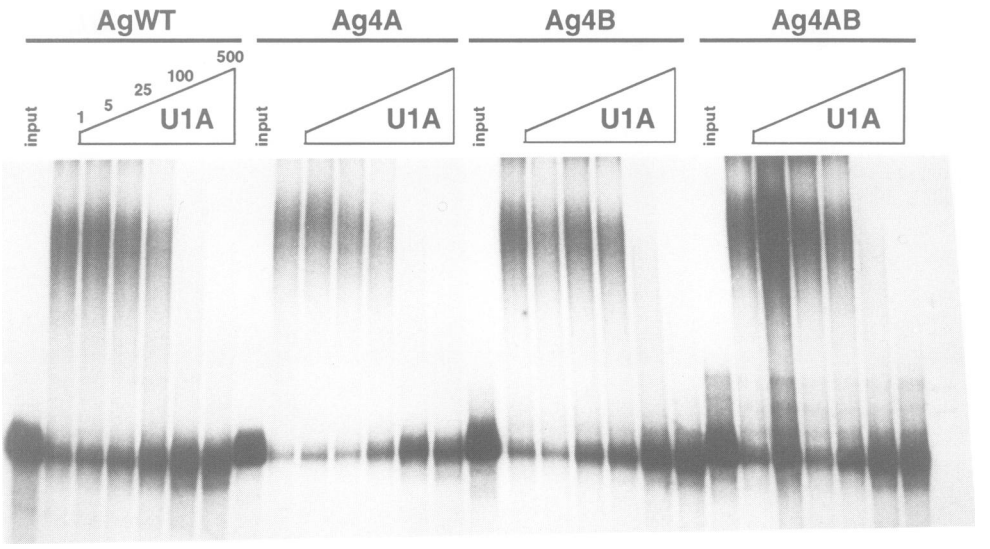


Fig. 6. Effect of the U1A protein on the *in vitro* polyadenylation of the Agwt RNA and the stem 4 mutants. The labelled RNA assayed is indicated above each panel. The lanes of each panel are the same as described in Figure 5A, except that the amounts of exogenously added U1A protein are different (ranging from 1 to 500 ng) as indicated above the panel. The lane on the extreme right is a ^{32}P end-labelled *Msp*I digest of pBR322.

of U1A protein binding at high salt. Mutants 1B and 1AB showed inhibition of polyadenylation comparable to the wt pre-mRNA (Figure 5C), while the 1A mutant showed no inhibition of polyadenylation.

As mentioned above, the existence of stem 4 in the human U1A pre-mRNA is supported by the nuclease digestion data but the stem has not been strongly conserved in evolution. To examine directly a possible functional role for this structure we tested the three mutants 4A, 4B and 4AB. All three mutants behaved similarly to the wt pre-mRNA in the polyadenylation inhibition assay (Figure 6). Thus, even if stem 4 does form, its existence does not seem to be important for the inhibition of polyadenylation by U1A protein.

Discussion

Structure of the 3′ UTR of U1A (pre-)mRNA

The structure of the region of the human U1A pre-mRNA responsible for binding to the U1A protein and thus for mediating autoregulatory inhibition of polyadenylation has been examined. Various lines of evidence suggest that the structure is complex. From top to bottom it starts with a tetraloop bounded by a stem of 5 bp (stem 3). Stem 3 is followed by an asymmetric internal loop, containing on one strand a 7 nt sequence required for U1A protein binding. On the other strand a single unpaired nucleotide is found, which probably stacks into the helix since it is inaccessible to nucleases. A second stem of four base pairs (stem 2) separates this internal loop from a second asymmetric internal loop, similar to the first, which may, or may not, be bounded by a further short helix (stem 1).

Apart from stem 1, all the secondary structure elements in the 5′ part of the structure were shown to be required for optimal binding to U1A protein and for function in polyadenylation inhibition. In the case of stem 1, the evolutionary conservation of the potential to form at least a short stem at this position suggests that the stem, though metastable, may exist. The effects of mutations in this putative stem on U1A protein binding were diverse. The results obtained with mutant 1B, however, established that the potential to form stem 1 is not essential for U1A protein binding.

The second structural element in the conserved region of human U1A pre-mRNA is a stem–loop with the AUUAAA cleavage and polyadenylation signal forming most of the loop. This structural feature is unnecessary for U1A protein binding and for inhibition of polyadenylation and, in addition, is not well conserved in evolution. Thus, even if this part of the structure does form *in vivo* it is unlikely to have any relevance for autoregulation.

What might be the reason for the complexity of the proven part of the structure, the region to which U1A protein binds? First, the data presented indicate that efficient inhibition of polyadenylation is only possible when two molecules of U1A protein can bind to the pre-mRNA. Second, the binding studies show that the two protein molecules bind cooperatively. An attractive aspect of the structure from this point of view is that the two asymmetric internal loops are spaced approximately half a helical turn apart (if standard RNA geometry is applicable). Although it can be assumed that the internal loops will induce a distortion or a kink in the helix (Chastain and Tinoco, 1991), the U1A binding sites may therefore lie side by side on one face of the helix, favouring interaction between the two protein molecules during binding. Note, however, that we do not know whether the observed cooperativity of binding is due to protein–protein interaction or to changes induced in the pre-mRNA structure on binding the first molecule of U1A protein.

Comparison of two U1A binding sites

The K_Ds of the two physiologically relevant U1A protein–RNA complexes studied to date, those involving U1 snRNA and U1A pre-mRNA, are very similar and indicative of very high affinity binding. The tight binding to U1 snRNA is perhaps explicable since U1A protein in the free state would turn off its own production via autoregulation and, presumably, U1 snRNA without U1A might be non-functional (Hamm *et al.*, 1990; but see Liao *et al.*, 1993). There seems not to be an obvious rationale

for such a strong interaction between U1A protein and its pre-mRNA.

Given the high affinity of both RNAs for U1A protein it is interesting to compare them. The U1A binding site on U1 snRNA is stem−loop II or B (Scherly *et al.*, 1989; Lutz-Freyermuth *et al.*, 1990). Parts of the 10 nt loop sequence and the presence of a stable stem, but apparently not the detailed structure of the stem, are critical for tight binding (Scherly *et al.*, 1989, 1990; Lutz-Freyermuth *et al.*, 1990; Bentley and Keene, 1991; Jessen *et al.*, 1991; Tsai *et al.*, 1991; Hall and Stump, 1992). A model for the U1A−U1 snRNA interaction has been proposed (Jessen *et al.*, 1991) in which most of the protein−RNA contacts are with the phosphates of the RNA backbone and the loop sequence is proposed to be mainly required to generate the correct backbone conformation.

The structural context of the most U1 snRNA-like sequence in the U1A pre-mRNA (Box 2) as a 7 nt unpaired strand in an asymmetric loop sandwiched between two stems, would appear to be rather different from its context in U1 stem−loop II. Given the conformational flexibility of RNA it is premature to say that the structure of the two tight binding sites will be different, but further examination of the role of the single-stranded bases in protein binding as well as high resolution studies of the two RNAs to reveal similarities and differences in their structures would be particularly interesting areas of study.

Inhibition of polyadenylation

The major conclusions of this study with regard to polyadenylation inhibition are that structural changes in U1A pre-mRNA that result either in a reduction in affinity for U1A protein or in the loss of the capacity to bind two molecules of U1A protein alleviate the inhibitory effects of U1A protein on cleavage and polyadenylation reactions *in vitro*.

The requirement for two bound protein molecules for inhibition might be most easily compatible with a simple model in which U1A protein sterically hinders interaction of one of the multiple cleavage and polyadenylation factors (see Wahle and Keller, 1992 for a review) with the U1A pre-mRNA. However, more complex models involving specific interaction between U1A protein or a particular structure in U1A pre-mRNA induced by U1A binding and one or more of the processing factors are not ruled out. These possibilities can now be tested.

Materials and methods

Sequence alignment and secondary structure prediction

The alignment of the three U1A sequences was made with the program PILEUP, which is part of the University of Wisconsin GCG package v.7.0 (Devereux *et al.*, 1984), and was adjusted manually. The programs FOLD and MFOLD (Zuker *et al.*, 1991) were used to generate optimal and suboptimal foldings of different regions of the three RNA sequences.

Enzymatic structure probing

The Ag and mutant RNAs used in this study were dephosphorylated at their 5′ ends and then radioactively labelled using [γ-^{32}P]ATP and T4 polynucleotide kinase according to Ehresmann *et al.* (1987). The labelled RNAs were purified by electrophoresis on a 10% polyacrylamide−urea denaturing gel. The full-length RNA products were cut out of the gel and eluted overnight at 4°C in a buffer containing 0.5 M NH_4Ac (pH 6.5), 10 mM $MgCl_2$ and 0.1% SDS (Krol and Carbon, 1989). The RNA was precipitated with ethanol and resuspended in water.

Labelled RNA ($2-3 \times 10^4$ c.p.m.) was supplemented with 4 μg of total yeast RNA as carrier. Digestion with RNase T1 (0.01 U), T2 (0.005 U), U2 (0.2 U; only in buffer D), A (1×10^{-6} U) or V1 (0.06 U; only in buffer N) were performed at room temperature for 10 min in buffer N or at 50°C for 5 min in buffer D. Buffer N (native conditions) contained 10 mM Tris pH 7.5, 10 mM $MgCl_2$ and 50 mM KCl. Buffer D (denaturing conditions) contained 7 M urea, 1 mM EDTA and 25 mM sodium acetate.

Preparation of mutants

The Ag sequence was inserted as an *Eco*RI−*Hin*dIII fragment into the pGEM-3z(+) vector. Single-stranded DNA was produced with the helper phage M13 K07 and mutations were introduced using the oligonucleotide-directed mutagenesis kit from Amersham. All mutants were checked by DNA sequencing.

Binding and polyadenylation assays

RNA and biotinylated RNA transcription by T7 RNA polymerase, production of ^{35}S-labelled U1A protein in wheat germ extract, production of recombinant U1A protein from *E.coli*, its biotinylation, the direct and indirect RNA−protein binding assays and *in vitro* polyadenylation reactions were all carried out as described by Boelens *et al.* (1993). The nucleotide sequence of the Ag fragment of U1A extends from position 842 to position 951 in the sequence (Nelissen *et al.*, 1991) and includes 8 nt at the 5′ end derived from the vector plasmid. Since U1A protein loses polyadenylation inhibition activity when stored, the amount required to inhibit polyadenylation of the wt pre-mRNA was determined empirically for each experiment.

For the electrophoretic mobility shift experiment, ^{32}P-labelled RNA was heated at 95°C for 3 min and quenched on ice for 1 min. 2×10^4 c.p.m. were added to the protein in a 10 μl reaction containing 10 mM Na-HEPES (pH 7.4), 50 mM KCl, 1 mM $MgCl_2$ and 200 ng of competitor tRNA at room temperature. The reaction was immediately loaded on a 7% native acrylamide gel (60:1 acrylamide:bisacrylamide), containing 10 mM Tris−borate pH 8.3, 1 mM EDTA and 0.1% Triton X-100. The gel was autoradiographed for 2−12 h at −80°C.

Filter binding assay

To determine the dissociation constants for the interaction between U1A protein and RNA substrates a nitrocellulose filter binding assay was used. A constant concentration of U1A protein in 10 μl buffer 1 containing 100 mM KCl, 2 mM $MgCl_2$, 20 mM HEPES−KOH pH 7.9, 5% glycerol, 0.5 mM DTE and 0.5 mg/ml BSA was mixed with 90 μl buffer 2 containing 10 mM Tris−HCl pH 7.5, 100 mM KCl, 2 mM $MgCl_2$, 0.1 mM EGTA, 0.5 mM DTE, 0.1 μg/μl tRNA and varying concentrations of ^{32}P-labelled RNA substrates. After equilibration at 20°C for 120 min, samples were filtered through pre-soaked Schleicher and Schuell BA85 0.45 μm nitrocellulose filters using a dot blot manifold (Schleicher and Schuell SRC96). The samples were subsequently washed twice with 200 μl buffer 2 without tRNA. The filters were dried and the amount of ^{32}P-labelled RNA bound to the filter was quantified by scintillation counting. The K_Ds were determined by Scatchard plot analysis.

Acknowledgements

Use of the services and facilities of the Dutch CAOS CAMM Center, under grant numbers SON 326-052 and STW NCH99.1751, is gratefully acknowledged. We thank M.Bennett and J.Craft for making the mouse U1A cDNA sequence available prior to publication. The research was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO), EMBO, EMBL and a postdoctoral fellowship (to S.G.) from the Alexander von Humboldt Foundation.

References

Bach,M., Krol,A. and Lührmann,R. (1990) *Nucleic Acids Res.*, **18**, 449−457.

Bentley,R.C. and Keene,J.D. (1991) *Mol. Cell. Biol.*, **11**, 1829−1839.

Boelens,W.C., Jansen,E.J.R., Van Venrooij,W.J., Stripecke,R., Mattaj,I.W. and Gunderson,S.I. (1993) *Cell*, **72**, 881−892.

Chastain,M. and Tinoco,I.,Jr (1991) *Prog. Nucleic Acid Res. Mol. Biol.*, **41**, 131−177.

Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387−395.

Ehresmann,C., Baudin,F., Mougel,M., Romby,P., Ebel,J.-P. and Ehresmann,B. (1987) *Nucleic Acids Res.*, **15**, 9109−9129.

Hall,K.B. and Stump,W.T. (1992) *Nucleic Acids Res.*, **20**, 4283−4290.

Hamm,J., Dathan,N.A., Scherly,D. and Mattaj,I.W. (1990) *EMBO J.*, **9**, 1237−1244.

Hoffman,D.W., Query,C.C., Golden,B.L., White,S.W. and Keene,J.D. (1991) *Proc. Natl Acad. Sci., USA*, **88**, 2495–2499.
Hope,I.A. and Struhl,K. (1987) *EMBO J.*, **6**, 2781–2784.
Jaeger,J.A., Turner,D.H. and Zuker,M. (1990) *Methods Enzymol.*, **183**, 281–303.
Jessen,T.H., Oubridge,C., Teo,C.H., Pritchard,C. and Nagai,K. (1991) *EMBO J.*, **10**, 3447–3456.
Krol,A. and Carbon,P. (1989) *Methods Enzymol.*, **180**, 212–227.
Liao,X.C., Tang,J. and Rosbash,M. (1993) *Genes Dev.*, **7**, 419–428.
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**, 6393–6397.
Nagai,K., Oubridge,C., Jessen,T.H., Li,J. and Evans,P.R. (1990) *Nature*, **348**, 515–520.
Nelissen,R.L.H., Sillekens,P.T.G., Beijer,R.P., Van Kessel,A.H.M.G. and Van Venrooij,W.J. (1991) *Gene*, **102**, 189–196.
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. (1990) *Nature*, **345**, 502–506.
Scherly,D., Kambach,C., Boelens,W., Van Venrooij,W.J. and Mattaj,I.W. (1991) *J. Mol. Biol.*, **219**, 577–584.
Sillekens,P.T., Habets,W.J., Beijer,R.P. and Van Venrooij,W.J. (1987) *EMBO J.*, **6**, 3841–3848.
Tsai,D.E., Harper,D.S. and Keene,J.D. (1991) *Nucleic Acids Res.*, **18**, 4931–4936.
Tuerk,C. *et al.* (1988) *Proc. Natl Acad. Sci. USA*, **85**, 1364–1368.
Wahle,E. and Keller,W. (1992) *Annu. Rev. Biochem.*, **61**, 419–440.
Zuker,M., Jaeger,J.A. and Turner,D.H. (1991) *Nucleic Acids Res.*, **19**, 2707–2714.

Received on July 26, 1993; revised on September 9, 1993