Probing the 3′ UTR Structure of U1A mRNA and Footprinting Analysis of Its Complex with U1A Protein†

Sander W. M. Teunissen, Celia W. G. van Gelder, and Walther J. van Venrooij*

Department of Biochemistry, University of Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

Received September 16, 1996; Revised Manuscript Received November 27, 1996⊗

ABSTRACT: The structure of the conserved region of the U1A pre-mRNA (Ag RNA) and its complex with U1A protein was investigated. The previously proposed secondary structure of Ag RNA, derived from enzymatic probing and analysis of the structure and function of mutant mRNAs, is now confirmed by chemical probing data and further refined in the regions where the enzymatic data were not conclusive. The two unpaired nucleotides in the internal loops opposite of the Box sequences as well as the tetraloop could not be cleaved by ribonucleases, but are accessible to chemical probes. Concerning the RNA−protein complex, the protection experiments showed that the Box regions are largely protected when the U1A protein is present. All stem regions in the 5′ part of the structure seem protected against ribonucleases. Unexpectedly, the nucleotides of the tetraloop become accessible to ribonucleases in the RNA−protein complex. This result indicates that the tetraloop undergoes a conformational change upon U1A protein binding. The 3′ part of the Ag RNA sequence, containing the polyadenylation signal in a hairpin structure, showed hardly any protection, a finding that agrees with the fact that U1A does not interfere with the binding of the cleavage polyadenylation specificity factor (CPSF) to the polyadenylation signal.

The removal of introns from premessenger RNA, known as splicing, is an important process in which several small ribonucleoprotein particles (snRNPs) participate. One of them, U1 snRNP, contains a U1 snRNA molecule, at least eight Sm proteins also present in other U snRNPs, and three U1-specific proteins named U1−70K, U1C and U1A [reviewed by Lührmann et al. (1990)].

The U1A protein contains two copies of an RNA binding domain, also referred to as RRM (RNA recognition motif),[1] RNP80 motif, or RNP motif, one at the N-terminal and one at the C-terminal region of the protein. The N-terminal RNP motif binds directly to the second stem−loop of U1 snRNA (Scherly et al., 1989; Lutz-Freyermuth et al., 1990; Nagai et al., 1990; Jessen et al., 1991; Hall & Stump, 1992). The structure of this RNA binding 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 both lying on the same side of the sheet. The loop of the second hairpin of human U1 snRNA contains 10 nucleotides. It has been shown that the first seven of them (with the highly conserved sequence AUUGCAC) are critical for U1A binding, whereas the structural context of this sequence affects binding affinity (Scherly et al., 1989, 1990; Bentley & Keene, 1991; Tsai et al., 1991). Recently, the complex between the N-terminal RNP motif of U1A and the second stem−loop of U1 snRNA has been investigated by NMR (Howe et al., 1994; Hall, 1994) and cross-linking studies (Stump & Hall, 1995). The β-sheet of U1A was shown to form the recognition surface, where the main contacts with the loop of the U1 snRNA hairpin occur. Furthermore, the crystal structure of this RNA−protein complex has been determined (Oubridge et al., 1994), revealing detailed information on the interaction between U1 snRNA and the U1A protein.

It has been shown that the 3′ UTR of the U1A pre-mRNA (Ag RNA) contains a region which has been conserved among vertebrates (Boelens et al., 1993). This region contains two stretches of seven nucleotides (called Box 1 and Box 2) which have a sequence similar to that contained in the second stem−loop of U1 snRNA and which are located in close proximity to the polyadenylation signal. Two human U1A proteins can bind to these two Box regions (Boelens et al., 1993; van Gelder et al., 1993), and experiments *in vitro* and *in vivo* have shown that the binding of two U1A proteins to this region specifically inhibits polyadenylation of the U1A pre-mRNA (Boelens et al., 1993). Thus, the U1A protein regulates the metabolism of its own pre-mRNA. The mechanism of this regulation has been elucidated by *in vitro* studies in which the U1A protein was shown to inhibit both specific and nonspecific polyadenylation by poly(A) polymerase (PAP) via a specific interaction in which the C-termini of both proteins seem to be involved (Gunderson et al., 1994).

Recently the human U1A protein−Ag RNA complex and the relationship between its structure and function in inhibition of polyadenylation *in vitro* were investigated (van Gelder et al., 1993). The secondary structure of the conserved region of the 3′ UTR was determined via a combination of theoretical predictions, phylogenetic sequence alignment, enzymatic structure probing, and analysis of structure and function of mutant mRNAs. It was shown that the integrity of much of this secondary structure is required for both high-affinity binding to U1A protein and specific inhibition of polyadenylation *in vitro* (van Gelder et al., 1993).

† This research was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO).

⊗ Abstract published in *Advance ACS Abstracts,* February 1, 1997.

[1] Abbreviations: RRM, RNA recognition motif; RNP, ribonucleoprotein particle; Ag RNA, a conserved region of the 3′ UTR of U1A pre-mRNA; $U1A_{101}$, the N-terminal 101 amino acids of the U1A protein.

S0006-2960(96)02323-9 CCC: $14.00

Here a more detailed analysis of the Ag RNA and its complex with U1A protein is reported. Chemical structure probing was performed on the Ag RNA both at 20 °C and at 0 °C, which provided a better understanding of some structural features which were not well understood from the enzymatic probing experiments. Both the reactivity of atoms involved in Watson−Crick base pairing and the reactivity of the N7 atoms of purines were investigated. Furthermore, the U1A−Ag RNA complex was probed by Fe^{II}EDTA and several enzymes, yielding not only a footprinting pattern but also structural information on the complex.

MATERIALS AND METHODS

In Vitro Transcription of RNA and Purification of Recombinant U1A Protein. In vitro transcription by T7 RNA polymerase was carried out as described (van Gelder et al., 1993). The conserved region of the 3′ UTR of the U1A mRNA is called Ag RNA and was cloned into the *Eco*RI and *Hind*III sites of pGEM-3Zf(+), resulting in Ag RNA transcripts. The nucleotide sequence of the Ag fragment of U1A extends from VI-842 to VI-951 in the sequence (Nelissen et al., 1991) and includes eight nucleotides at the 5′ end derived from the vector plasmid. Production of recombinant U1A protein and of the mutant protein $U1A_{101}$ (consisting of the first 101 N-terminal amino acids of U1A protein) in *E. coli* was carried out as described (Boelens et al., 1993).

5′- and 3′-End-Labeling. For 5′-end-labeling, the dephosphorylated RNAs or oligodeoxynucleotides were labeled using [γ-^{32}P]ATP and T4 polynucleotide kinase as described previously (van Gelder et al., 1993). For 3′-end-labeling, the RNAs were labeled using [^{32}P]pCp and T4 RNA ligase as described (van Gelder et al., 1994).

Chemical Modification. Concentrations of chemicals were optimized to obtain single hit conditions. Control incubations, in which the reagent was omitted, were always performed in parallel to detect spontaneous pyrimidine-purine breaks, which easily occur in RNA (Krol & Carbon, 1989; Kwakman et al., 1990), and, in the case of the primer extension method, also to detect spontaneous stops of reverse transcriptase. Chemical modifications were performed both on unlabeled (0.3−0.5 μg) and on 3′-end-labeled RNA (3×10^4 cpm), which was always renatured before use. The RNAs were modified under native conditions (N; presence of magnesium), semidenaturing conditions (SD; presence of EDTA), and denaturing conditions (D; high temperature, presence of EDTA). Native and semidenaturing reactions were conducted both at 20 °C and at 0 °C. Modification reactions with DMS (dimethyl sulfate) and CMCT [1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluenesulfonate] were essentially carried out as described (van Gelder et al., 1994). It should be noted that DMS and CMCT as well as the chemicals mentioned below are hazardous and should be handled with proper precautions. Chemically modified nucleotides were detected either by primer extension analysis or by using 3′-end-labeled RNA (in the case of N7-G, N7-A, and N3-C).

In case of probing the N7-G positions with DMS, an additional aniline treatment was carried out to produce strand scission at the site of modification (Krol & Carbon, 1989).

During the DEPC (diethyl pyrocarbonate) treatment, 10−60 μL of DEPC was added to the sample in 200 μL of buffer N or SD [final concentration $(0.3-1.9) \times 10^{-9}$ M; incubation for 1 h at 20 °C]. For D conditions, 3−10 μL of DEPC was added to 200 μL of buffer SD and incubated for 7 min at 90 °C [final concentration $(0.09-0.3) \times 10^{-9}$ M]. Reactions were stopped by ethanol precipitation. An aniline step was performed to produce strand scission at the site of the modification (Krol & Carbon, 1989).

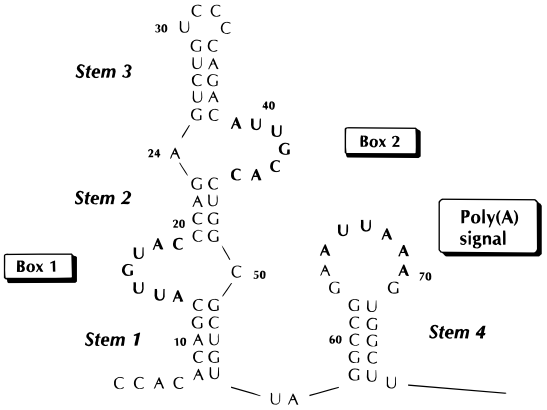

FIGURE 1: Proposed secondary structure and nomenclature of the conserved region of the 3′ UTR of the human U1A pre-mRNA (Ag RNA). The Box sequences, the polyadenylation signal, and stems 1, 2, 3, and 4 are indicated.

Primer Extension Analysis. Primer extension was carried out as described (van Gelder et al., 1994). Oligodeoxynucleotide primers used were 5′-GCTTAACAGCGCCAGG-3′ and 5′-GATTGTGAAAAACCAAACCTC-3′, complementary to nucleotides 45−60 and 81−101 in Ag RNA, respectively. Reverse transcripts were analyzed on 10% denaturing polyacrylamide gels.

Enzymatic Footprinting. After renaturation of the 5′-end-labeled Ag RNA (3×10^4 cpm, final concentration ~6 nM), the specified amount of U1A wt protein (150−300-fold excess) or $U1A_{101}$ (50−300-fold excess) was added (final volume: 20 μL). Buffer conditions were 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, and 50 mM KCl. The complex was allowed to form for 30 min at 20 °C after which the probing reactions were performed with RNase T1 (0.15 unit; U1A wt only), RNase A (1×10^{-5} unit; U1A wt only), RNase T2 (0.005 unit), or RNase V1 (0.06 unit) as described (van Gelder et al., 1993).

Fe^{II}EDTA Footprinting. The reactions were essentially carried out as described (Darsillo & Huber, 1991), with varying concentrations of Fe^{2+}, EDTA, ascorbic acid, and hydrogen peroxide.

RESULTS

Structure Probing of Ag RNA. The proposed secondary structure of the conserved region of the 3′ UTR of the U1A mRNA, called Ag RNA, is shown in Figure 1 and consists of two distinct parts which are separated by only two nucleotides, U56 and A57 (van Gelder et al., 1993). The 5′ part, which has a symmetric structure, contains three stems (numbered 1, 2, and 3), separated by two asymmetrical internal loops containing the Box 1 and 2 sequences, which are required for U1A protein binding (Boelens et al., 1993). The 3′ part of the structure is formed by stem 4 and a 9-nucleotide loop containing the polyadenylation signal (loop 4).

Enzymatic structure probing experiments showed that the central three nucleotides in Box 1 and 2, as well as the

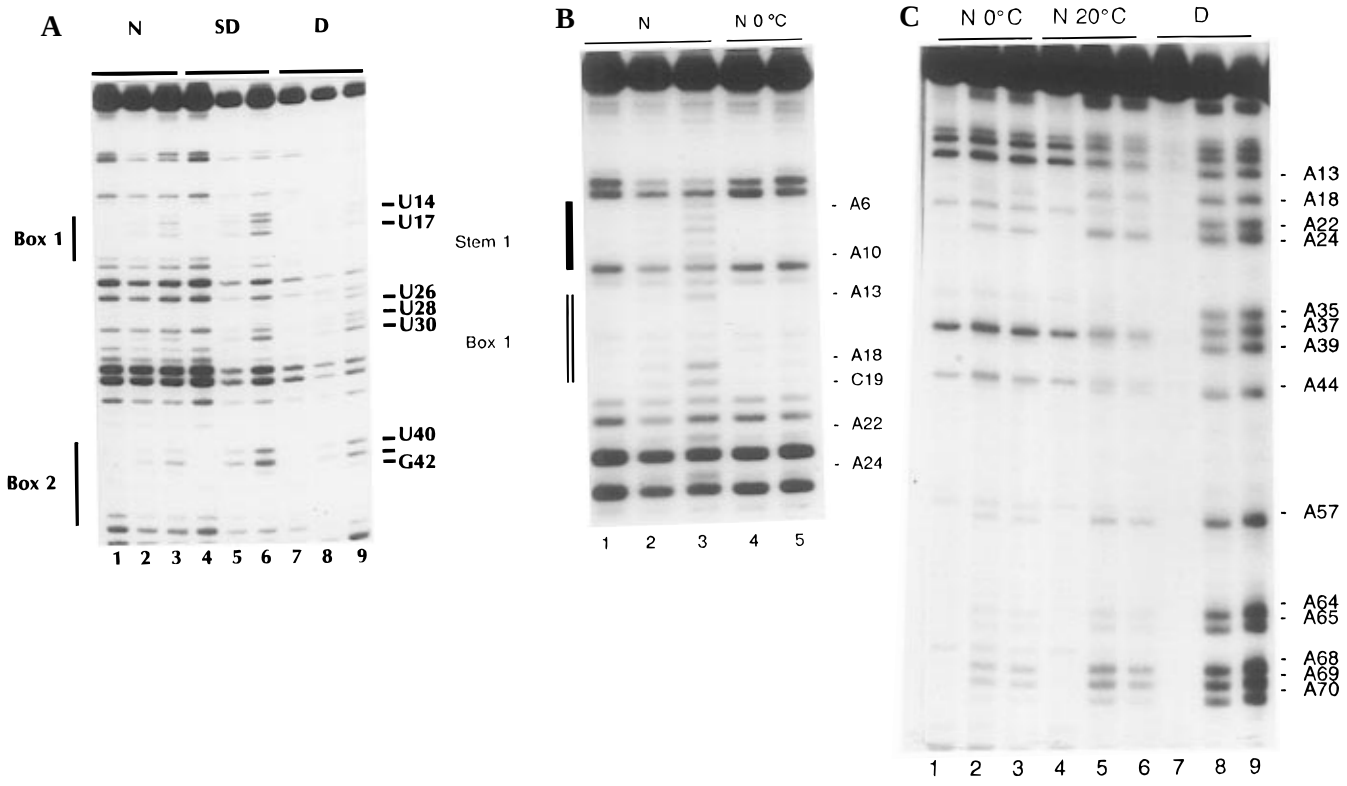

FIGURE 2: Secondary structure probing of Ag RNA. (A) Chemical probing of Ag RNA with CMCT. Detection of modifications was achieved via primer extension. Samples in lanes 1, 4, and 7 are control incubations in which reagent was omitted. Lanes 2, 3, 5, and 6: 50 μL of CMCT (final concentration 0.33 M) incubated at 20 °C for 20, 30, 5, and 10 min, respectively. Lanes 8 and 9: 10 and 25 μL of CMCT (final concentration 0.15 M) incubated for 1 min at 90 °C. (B) Chemical probing of Ag RNA with DMS at 20 °C and at 0 °C. Detection of modifications was achieved via primer extension. Samples in lanes 1 and 4 are control incubations in which reagent was omitted. The reaction conditions are indicated: N (native conditions, 20 °C) and N 0 °C (native conditions, 0 °C). Lane 2: 0.5 μL of DMS incubated for 15 min (final concentration 0.1×10^{-9} M). Lanes 3 and 5: 1.5 μL of DMS (final concentration 0.3×10^{-9} M) incubated for 15 min. (C) Chemical probing of 3′-end-labeled Ag RNA of N7-A positions with DEPC. Samples in lanes 1, 4, and 7 are control incubations in which reagent was omitted. The reaction conditions are indicated above the figure: N (native conditions); D (denaturing conditions). Lanes 2 and 3: 10 and 20 μL of DEPC (final concentration 0.3×10^{-9} and 0.6×10^{-9} M, respectively) incubated for 1 h at 0 °C. Lanes 5 and 6: 10 and 20 μL of DEPC (final concentration 0.3×10^{-9} and 0.6×10^{-9} M, respectively) incubated for 50 min at 20 °C. Lanes 8 and 9: 2 and 4 μL of DEPC (final concentration 0.06×10^{-9} and 0.12×10^{-9} M, respectively) incubated for 4 min at 90 °C.

polyadenylation signal, were single-stranded (van Gelder et al., 1993). The presence of the highly conserved stems 2 and 3, which are needed for U1A protein binding, was established by RNase V1 cleavages and by analyses of the structure and function of mutant mRNAs (van Gelder et al., 1993). However, the behavior of a few other parts of the structure was less easy to understand. Stems 1 and 4 showed cleavage both by RNase V1 and by single-strand-specific ribonucleases. This seemed to indicate that these two stems, which have not been strongly conserved in evolution and which appear not to be important for either U1A protein binding or inhibition of polyadenylation by the U1A protein, are of weak stability or may not exist at all in solution (van Gelder et al., 1993). Furthermore, the tetraloop of stem 3 (nucleotides 30−33) was hardly cleaved by ribonucleases under native conditions, suggesting that its structure might be very compact. A similar behavior was found for the unpaired nucleotides A24 and C50. This could arise either from the possibility that these two nucleotides are located inside the helix or from the possibility that other parts of the Ag RNA sterically hinder the ribonucleases. In contrast, chemical probes are, due to their small size, less sensitive to steric hindrance and therefore could provide more detailed insight in the Ag RNA structure.

The four bases were monitored at their Watson−Crick base-pairing positions by DMS at N1-A and N3-C and by CMCT at N3-U and N1-G. Position N7 of guanine and adenine residues was probed by DMS and DEPC, respectively. The experiments were performed under native, semidenaturing, and denaturing conditions. Tertiary interactions are generally less stable than Watson−Crick interactions and are expected to melt under semidenaturing conditions (Krol & Carbon, 1989). Experiments under such conditions will also give information about the stability of the different double stranded domains. Ag RNA was probed both at 20 °C and at 0 °C. The latter temperature was used to minimize the breathing in this relatively small RNA molecule, a phenomenon observed at 20 °C (see below).

Figure 2A−C shows examples of the chemical probing results for Ag RNA, while Figure 3 summarizes the results of several independent probing experiments for both 20 °C and 0 °C, for the Watson−Crick base-pairing positions (Figure 3A), and for the N7 atom of purines (Figure 3B).

Stem Regions. (A) Stems 2 and 3. At 20 °C, the presence of stems 2 and 3 is clearly supported by the chemical modification data, since many nucleotides are only reactive under denaturing conditions. This is shown, for example, for nucleotides U26 and U28 in stem 3 in Figure 2A (lane 9) for the CMCT reaction. Their counterparts in stem 3, A37 and A35, respectively, are reactive with DMS (data not shown), as is A22 in stem 2 (see Figure 2B, lane 3). This difference in reactivity between adenosines and uridines in

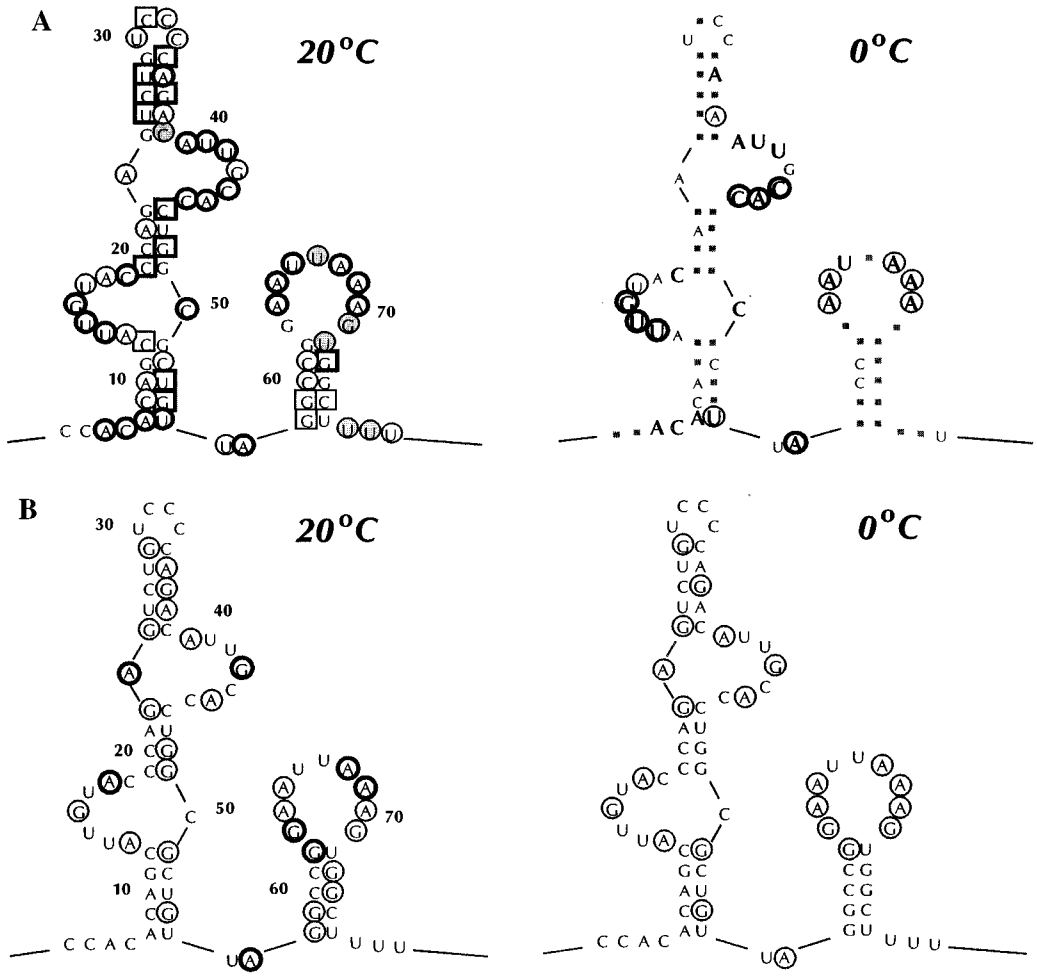

FIGURE 3: Summary of the chemical structure probing of Ag RNA. (A) Reactivity of Watson−Crick positions of nucleotides in Ag RNA toward chemical probes at 20 and 0 °C. Consensus data from several independent experiments using both primer extension and end-labeled detection are shown. Left panel: results from experiments performed at 20 °C. Reactivity under native conditions is indicated by circles. The thickness of the circles represents moderate (thin) or strong (thick) reactivity. Reactivity under semidenaturing conditions is indicated by filled circles. Squares indicate reactivity under denaturing conditions. The thickness of the squares represents moderate (thin) or strong (thick) reactivity. Right panel: results from experiments performed at 0 °C; nucleotides which were not reactive at 20 °C are indicated by shaded boxes; the nucleotides which were reactive at 20 °C are indicated in standard and boldface font (moderate and strong reactivity, respectively). Reactivity at 0 °C is indicated by circles, as in the left panel. Nucleotides for which no reactivity is indicated show RT-stops in the primer extension reaction. (B) Reactivity of N7 positions of purines in Ag RNA toward chemical probes at 20 °C (left panel) and 0 °C (right panel). Consensus data from several independent experiments using 3′-end-labeled Ag RNA are shown. Symbols are identical to those used in Figure 3A. Data are available for nucleotides 11−79.

A-U base pairs has also been observed in helical regions of other RNAs (van Gelder et al., 1994; Krol & Carbon, 1989; Glickman et al., 1988) and is probably related to the fact that DMS (M_r = 126) is smaller than CMCT (M_r = 424).

Concerning the N7 positions of the purines in stems 2 and 3, the guanosines are on the average more reactive toward DMS than the adenosines toward DEPC (see Figure 3B for a summary). This difference occurs because DEPC is larger than DMS, and therefore more sensitive to stacking (Dock-Bregeon et al., 1989; Mougel et al., 1987), which in our case is most clearly visible in the experiments performed at 0 °C. Most nucleotides show reduced accessibility at 0 °C, and two N7 atoms, A35 and A37, cannot be modified at all at this temperature (Figure 2C, lanes 3 and 6). Guanosines 23, 25, 49, and 51 are bordering the two internal loops where they are likely to be more accessible, a behavior also found in other RNA internal loops (van Gelder et al., 1994).

(B) Stems 1 and 4. In agreement with the enzymatic probing (van Gelder et al., 1993), the chemical probing experiments showed that stems 1 and 4 are of weak stability and are breathing at 20 °C. In stem 1, many nucleotides were reactive at their Watson−Crick positions at this temperature (see Figure 2B, lane 3, for nucleotides 6−10). When we lowered the temperature to 0 °C, the Watson−Crick positions of nucleotides 6−10 could no longer be modified by DMS (Figure 2B, lane 5). Only U55 could still be modified by CMCT at this temperature (data not shown), but since this nucleotide is at the end of stem 1 it is likely to be more reactive. The reactivity of the N7 atoms showed a similar temperature-dependent behavior. The N7 atoms of G51 and G54 were reactive at 20 °C, while their reactivity was reduced at 0 °C (data not shown). In stem 4, several nucleotides showed reactivity at 20 °C, both at their Watson−Crick and at their N7 positions (see Figure 3 for a summary), which was reduced when the temperature was lowered. Taken together, these results suggest that stems 1 and 4 indeed can be formed, although they are of weak stability at 20 °C.

Loop and Linker Regions. (A) Box 1 and Box 2 Regions. All nucleotides in the Box 1 and 2 sequences are accessible at their Watson−Crick positions at 20 °C (see Figure 3A for a summary of the data). Figure 2B (lane 3) shows the

accessibility of A13, A18, and C19 in Box 1 to DMS, and Figure 2A (lane 3) shows the accessibility of U14 to U17 in Box 1 and of U40 to G42 in Box 2 to CMCT. At 0 °C, a few nucleotides at the 5′ part of Box 2 become inaccessible at their Watson−Crick positions (Figure 3A, right panel). This is shown for the N1 atoms of A13, A18, and C19 in Figure 2B, lane 5. This behavior is probably due to stacking of the bases which could agree well with the RNase V1 cleavage found at the 5′ parts of both Box sequences (van Gelder et al., 1993). It must be noted, however, that both Box sequences do not behave exactly the same at 0 °C (see Discussion).

Although RNases were unable to cleave the unpaired nucleotides A24 and C50, our chemical probing results show that N3 of C50 is strongly reactive (not shown) and that the N1 atom of A24 (Figure 2B, lane 3) is moderately reactive at 20 °C. For C50, this had to be deduced from reactions with 3′-end-labeled RNA (data not shown) due to the occurrence of a natural stop of reverse transcriptase at C50 in the primer extension reactions. Note that for A24 also the N7 atom is available for modification (Figure 2C, lane 6). At 0 °C, N1 of A24 is no longer accessible (Figure 2B, lane 5), probably due to stacking, but the N7 atom of A24 still can be modified (Figure 2C, lane 3).

(B) Tetraloop. In the tetraloop, U30 is moderately reactive toward CMCT while the reactivity of cytosines 31−33 toward DMS is more difficult to evaluate due to the presence of natural RT-stops, especially at nucleotides 33 and 34. By using 3′-end-labeled RNA, however, it was found that nucleotides 32 and 33 are moderately reactive at their N3 position, while N3 of C31 was only reactive under denaturing conditions (data not shown).

(C) Loop 4 and Linker Region. In full agreement with the enzymatic probing (van Gelder et al., 1993), loop 4 was completely accessible at the Watson−Crick positions under native conditions with the exception of U67, which became accessible to CMCT only under SD conditions (data not shown). The N7 positions of the purines (see Figure 2C, lane 6) are accessible, and A68 through A70 seem more strongly modified than A64 and A65. At 0 °C, most of the nucleotides were still moderately accessible at both Watson−Crick and N7 positions, but the reactivity was clearly reduced compared to 20 °C (shown in Figure 2C, lanes 3 and 6).

The linker region, which connects the 5′ part with the 3′ part of the structure, is formed by the two nucleotides U56 and A57. Both nucleotides are fully accessible at 20 °C (data not shown). At 0 °C, U56 can no longer be modified, but A57 is still accessible, both at N7 (Figure 2C, lanes 2 and 3) and at N1 (see Figure 3A).

Analysis of the Complex of Ag RNA with U1A Protein. To obtain information on the complex of Ag RNA and the U1A protein, footprinting experiments were performed using various ribonucleases and Fe^{II}EDTA. In these experiments, 5′-end-labeled Ag RNA was incubated with an excess of U1A protein. After this, the resulting RNP complexes were probed with RNases A, T1, T2, or V1, or with hydroxyl radicals. Examples of RNase footprinting are shown in Figure 4A, Figure 4C (both RNase T2), and Figure 4B (RNase V1), while Figure 4D summarizes the results obtained by ribonuclease protection.

As might be expected, the Box 1 and 2 regions are almost completely protected by the U1A protein (compare lanes 2 and 3 in Figure 4A). The phosphodiester bond between C43 and A44 is very sensitive in both RNA and RNP, which obscures clear interpretation of the protection pattern at that position. Such intrinsic fragility is a well-known phenomenon in RNA, especially for pyrimidine−adenosine bonds (Mougel et al., 1987). Note that in the experiment shown in Figure 4A, probably some RNase A-like activity was present (Figure 4A, lane 1). A13 in Box 1 becomes a hypersensitive site in the RNP complex (Figure 4A, lane 3). The single-strand-specific RNases also cleave some nucleotides in the stem regions in the naked RNA, for instance nucleotides 26−28 in stem 3 and nucleotides in both stem halves of stem 1. These cleavages are absent or much weaker in the RNP (Figure 4A, compare lanes 2 and 3; see also Discussion).

The protection pattern obtained by RNase V1 in the presence of U1A wt protein (Figure 4B) shows that the stem regions in the 5′ part of the RNA (stems 1, 2, and 3) are protected in the RNA−protein complex, while the 3′ part remains unprotected (compare lanes 3 and 4). One V1 cleavage, between nucleotides U53 and G54, became stronger in the RNP as compared to the naked RNA. Footprinting experiments with RNases V1 and T2 in the presence of $U1A_{101}$ (containing the N-terminal 101 amino acids of U1A) did not show significant differences in the protection pattern (data not shown).

Because nucleotides of the tetraloop in the naked RNA were not cleaved by RNases [see above and van Gelder et al., (1993)], information about protection of this region was not expected to be obtained. Interestingly, however, the tetraloop seems to become accessible in the RNP complex, indicating a structural change in this part of the RNA upon protein binding (see Discussion). This is shown in Figure 4C, lanes 4−6, where RNase T2 cleaves C31 and C32 which are both not cleaved in the naked RNA (Figure 4C, lane 3). Note that this effect seems to be more prominent for $U1A_{101}$ than for U1A (Figure 4C, lanes 4−6, and Figure 4A, lanes 3 and 4).

The 3′ part of the structure (stem 4 and loop 4) does not show much protection (see Figures 4A,B), so this region appears to be largely accessible in the RNP complex. Only the 5′ side of the polyadenylation signal (nucleotides A64 and A65) shows partial protection (Figure 4A, lanes 2 and 4).

The footprinting experiments using Fe^{II}EDTA yielded no distinct protection pattern. Both the free RNA and the U1A-bound RNA showed equal reactivity toward the hydroxyl radicals at all positions (data not shown). Fe^{II}EDTA probing can be used to detect three-dimensional folding of the RNA molecule if riboses are in the interior of the RNA molecule and therefore protected from strand-scission (Celander & Cech, 1990). Our results of the naked RNA showed equal reactivity for all nucleotides, indicating that this RNA has no tertiary interactions.

DISCUSSION

Secondary Structure of Ag RNA. The conserved region of the 3′ UTR of Ag RNA was probed at nucleotide resolution by the utilization of chemical probes. The secondary structure thus obtained further refines the structure predicted previously, which was based upon enzymatic probing and analysis of the structure and function of mutant RNAs (van Gelder et al., 1993). At 20 °C, the highly

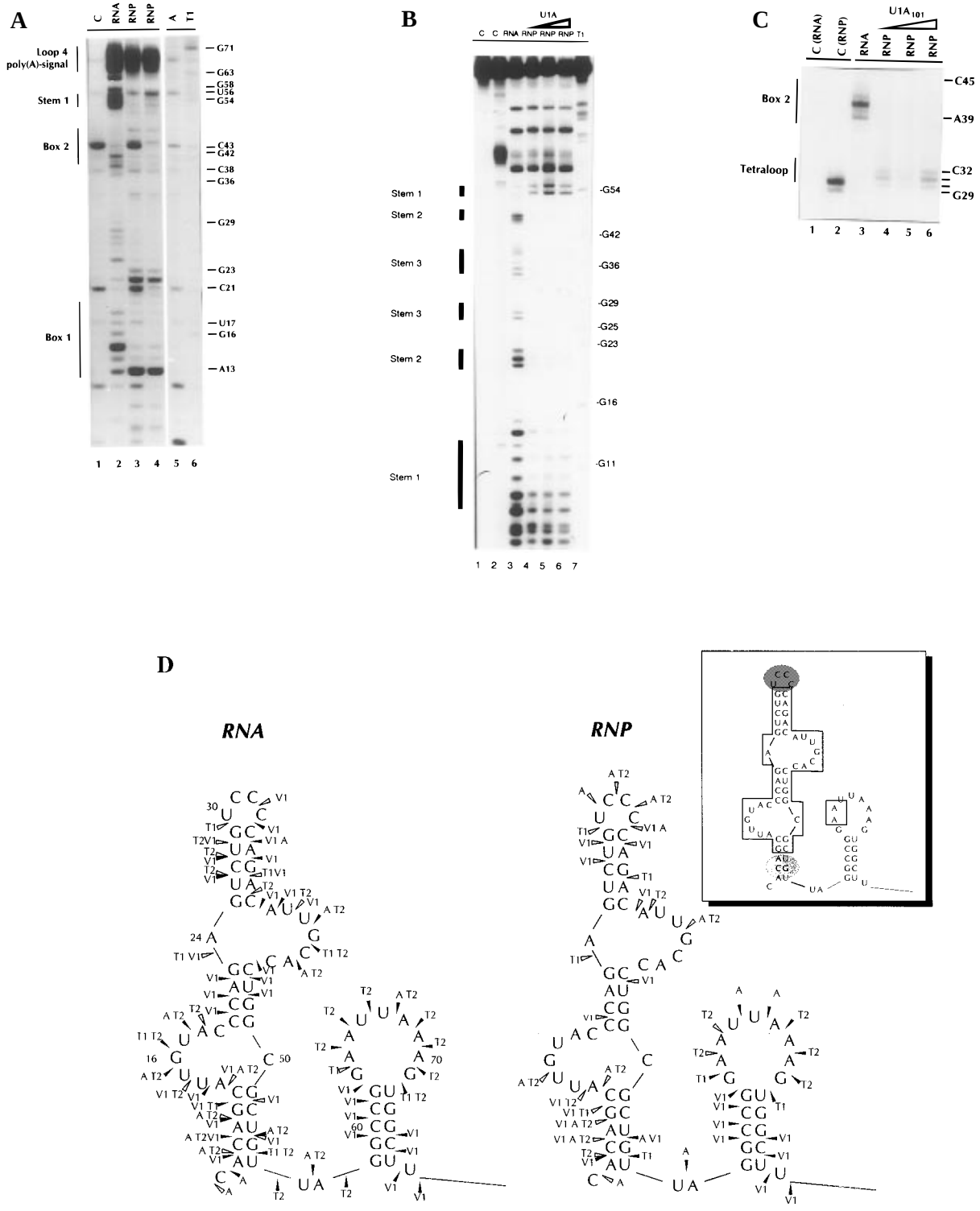

FIGURE 4: Enzymatic probing of the U1A−U1A pre-mRNA complex. (A) Enzymatic footprinting using 5′-end-labeled Ag RNA and single-stranded-specific RNase T2. Lane 1: Control incubation where U1A protein and RNase T2 were omitted. Lane 2: RNA probed at 20 °C for 10 min with RNase T2 (5×10^{-3} unit). Lanes 3 and 4: RNA incubated with respectively 150 and 300 molar excess of U1A protein, probed with RNase T2 (5×10^{-3} unit). Lanes 5 and 6: RNA probed under denaturing conditions with RNases A and T1, to obtain a sequence ladder for U/C and G, respectively. (B) Enzymatic footprinting using 5′-end-labeled Ag RNA and RNase V1. Lane 1: Control incubation where U1A protein and RNase V1 are omitted. Lane 2: Control incubation where both Ag RNA and U1A protein are present, but where RNase V1 is omitted. Lane 3: RNA probed at 20 °C for 10 min with RNase V1 (0.06 unit); Lanes 4, 5, and 6: RNA incubated with respectively 150, 300, and 500 molar excess of U1A protein, probed with RNase V1. Lane 7: RNA probed under denaturing conditions with RNase T1, to obtain a sequence ladder. (C) Enzymatic footprinting of the $U1A_{101}$−Ag RNA complex using T2. The region of the tetraloop is shown. Lanes 1 and 2 are control lanes [incubations of (1) only RNA or (2) RNA and protein] in which the enzyme was omitted. Lane 3 shows the enzymatic probing of the naked RNA, while lanes 4−6 show probing of the complex by increasing amounts of $U1A_{101}$ (50, 100, and 150 times excess, respectively). (D) Summary of RNase data obtained for both the naked Ag RNA and the U1A/$U1A_{101}$−Ag RNA complex. On the left, the digestion pattern of the RNA is shown [adapted from van Gelder et al. (1993)], while on the right the digestion pattern of the U1A/$U1A_{101}$−Ag RNA complex is shown. Strong cleavages are indicated by solid arrows and weak cleavages by open arrows. The enzymes which cut are indicated next to the arrows. The insert shows the protected areas indicated by boxes and the regions of structural changes indicated by shaded circles.

conserved stems 2 and 3 are indeed present while stems 1 and 4 also seem to exist, although not very stable and probably breathing. At 0 °C, all four stems clearly exist.

At 20 °C, all nucleotides in the Box 1 and 2 regions are fully accessible both at their Watson−Crick positions and at the N7 atoms of the purines. This behavior excludes the presence of tertiary interactions between these nucleotides and other parts of the RNA. At 0 °C, several nucleotides in the Box regions are no longer accessible, probably because of stacking. Interestingly, the behavior of the two Box regions at 0 °C is not identical. Box 1 shows more reactivity of both Watson−Crick and N7 positions at the 5′ end of the Box, while Box 2 shows most reactivity at the 3′ end of the Box region. A (somewhat) different structure of the two Boxes could be expected because the two sequences, although almost identical in sequence, have a different structural context in the Ag RNA and also differ in U1A binding affinity. Box 2 forms a much stronger (∼30-fold) binding site for U1A protein than Box 1 (van Gelder et al., 1993).

The two unpaired nucleotides A24 and C50 are clearly accessible at 20 °C. Whether the accessibility of A24 and C50 results from looping out of the helix or from the fact that the structure of the Ag RNA is more open at the internal loops, cannot be deduced from these results. The fact that N7 of G25 can also be modified suggests that A24 is not stacked in the helix.

Cleavage in the tetraloop by RNases was not observed (van Gelder et al., 1993), but chemical probing showed that in the RNA three of the four nucleotides are moderately accessible at 20 °C for the much smaller chemical probes. N3 of C31 was only reactive under denaturing conditions, which could either indicate stacking or an involvement in tertiary interactions under native conditions. At 0 °C, none of the nucleotides in the tetraloop are accessible.

The chemical probing results clearly indicate the presence of stem−loop 4. In the loop containing the AUUAAA polyadenylation signal, it can be seen that all adenosines are reactive at both the N1 and the N7 position, both at 20 °C and at 0 °C.

In conclusion, the probing studies provide a secondary structure for the Ag RNA as shown in Figure 1. The data also suggest that there are no major tertiary interactions within the naked RNA. Almost all purines are accessible at their N7 position. This is supported by our Fe^{II}EDTA experiments which show equal reactivity of all nucleotides in the Ag RNA, indicating the absence of tertiary interactions. A similar behavior toward the Fe^{II}EDTA-generated hydroxyl radicals was found in the 5S rRNA (Celander & Cech, 1990).

The Ag RNA−U1A Protein Complex. Footprinting experiments have been performed on the complex of Ag RNA and the U1A protein using several ribonucleases and Fe^{II}EDTA. Inhibition of reactivity at certain nucleotides can be inferred as a direct protection (and hence contact) of the RNA by the protein at that site. However, reduced reactivity can also be caused by conformational changes in the RNA chain brought about by the addition of the protein, and it is difficult to distinguish between these two modes of reduced reactivity. Furthermore, since RNases are large molecules, steric hindrance may significantly enlarge the protected regions. Footprinting experiments were also conducted with $U1A_{101}$, which contains only the N-terminal RNP motif of U1A, but these experiments yielded the same results.

The protection experiments show that all seven nucleotides contained in the Box 1 and Box 2 regions, as might be expected, are largely protected when the U1A protein is bound. Clearly, these sequences, which have been shown to be important for U1A binding to U1A mRNA (Boelens et al., 1993), are in contact with the U1A protein. Only some nucleotides, located at the 5′ side of both Boxes, can be attacked by the ribonucleases.

All nucleotides in stems 1, 2, and 3 show complete or partial protection against ribonucleases in the presence of U1A protein, and also in the presence of $U1A_{101}$. Around nucleotide 54, RNase V1 cleavage is enhanced when U1A protein is added. This indicates that stem 1 becomes more stable as a result of U1A binding. Alternatively, it could be that stacking of stems 1 and 4 on each other occurs. This is supported by the absence of RNase T2 cleavages for nucleotides G53−G58. The reduction of cleavages by single-strand-specific RNases in stems 1 and 3 could indicate protection of these regions by the U1A protein, but could also be the result of a further stabilization of the stems upon binding of U1A protein. Footprinting of the complex of stem−loop II of U1 snRNA and U1A protein has been performed with both RNase V1 and ethylnitrosourea (Jessen et al., 1991). In that case, only one of the stem halves of stem−loop II appeared to be protected against the probes. This difference in behavior of U1 snRNA (Jessen et al., 1991) as compared to Ag RNA can be due to the difference in size of the RNA substrates or to the presence of two rather bulky U1A proteins in the latter case, instead of one in the case of U1 snRNA. The environment of the recognition sequence also differs between the two RNA substrates. In the case of U1 snRNA, the seven nucleotides are flanked by three loop nucleotides with a flexible conformation, whose only function is to allow the seven nucleotide recognition sequence to adopt the conformation necessary for protein binding (Williams & Hall, 1996). In the Ag RNA, the seven nucleotides are flanked by stem structures and a bulged nucleotide. Obviously, these nucleotides are less flexible and therefore might induce a different structure of the RNA, leading to a different protection pattern.

Interestingly, in the RNP complex, the nucleotides in the tetraloop (U30−C33) seem to become accessible. This may indicate that this loop undergoes a considerable conformational change upon U1A protein binding. It appears that the loop opens up, with its nucleotides becoming available to the probes. Although we cannot completely exclude that the bands appear as a result of secondary enzyme cuts in the RNA molecule, this seems not very likely. The observed protection pattern in Box 2 indicates that the U1A protein is still able to bind the Ag RNA (Figure 4C, compare lanes 3 and 4−6), suggesting that the structure of the complex is still intact. If the bands in the tetraloop are secondary enzyme cuts, the complex would probably be disrupted. Similar behavior of a tetraloop is also found in bacteriophage R17 where the tetraloop structure is altered upon R17 coat protein binding (Varani, 1995). Our results are seemingly in contrast with NMR data recently published (Gubser & Varani, 1996), in which it was shown that the tetraloop structure remained unperturbed upon protein binding. However, in those experiments, a modified Ag RNA was used which contained only Box 2 flanked by two stem structures and hence, able to bind only one U1A protein. More importantly, stem 3 was shortened by one base pair, and the

tetraloop had a different sequence as in the original Ag RNA used by us, in order to stabilize stem 3. The recently described mechanism for complex formation does not describe a possible conformational change in the tetraloop (Allain et al., 1996), but in this study, the same modified RNA was used as in the study mentioned before (Gubser & Varani, 1996).

The 3′ part of the structure is formed by stem−loop 4, and this part shows, as expected, no protection, except for some limited changes at the 5′ side of the loop. This means that this region is largely accessible in the RNP complex. This is in agreement with the finding that U1A does not interfere with the binding of the cleavage polyadenylation specificity factor (CPSF) to the polyadenylation signal during polyadenylation of the U1A pre-mRNA (Gunderson et al., 1994).

The crystal structure of the U1A−U1 snRNA complex (Oubridge et al., 1994) revealed that the recognition surface of the protein binds the RNA in a "stem-downwards" manner (Nagai et al., 1990). Since the recognition sequence on the Ag RNA is the same, the most likely binding mode of the U1A protein to this RNA will be similar. A theoretical model has been proposed for the complex of two binding domains of U1A and the Ag RNA (excluding stem−loop 4) using the crystallographic data combined with molecular modeling (Jovine et al., 1996). Our experimental data show that stem 2 is protected, both by the U1Awt as well as by the $U1A_{101}$ protein. This indicates that the N-terminal RNP motif is sufficient for protecting stem 2 and possibly involved in the protein−protein interactions. Our experimental data are in complete agreement with the theoretical model which predicts protein−protein interactions between Lys96 and Asp24, as well as between Lys60 and Gln39 (Jovine et al., 1996).

The model for the U1A−3′UTR complex could also explain the results of the Fe^{II}EDTA experiments. Since the protein−protein interactions occur via the major groove of stem 2, this leaves the minor groove exposed to the solvent. Furthermore, the minor grooves of stem 1 and stem 3 are exposed to the solvent. Fe^{II}EDTA-generated hydroxyl radicals attack the riboses of the RNA in the minor groove, which therefore would not result in any protection pattern.

In conclusion, we have obtained detailed experimental information concerning the structure of Ag RNA and its complex with U1A protein. These data can aid in elucidating the process of autoregulation of the U1A protein.

ACKNOWLEDGMENT

We want to thank Drs. Pascale Romby and Chantal Ehresmann for advice with the chemical probing reactions and Dr. Ger Pruijn for critical reading of the manuscript.

REFERENCES

Allain, F. H., Gubser, C. C., Howe, P. W., Nagai, K., Neuhaus, D., & Varani, G. (1996) *Nature 380*, 646−650.
Bentley, R. C., & Keene, J. D. (1991) *Mol. Cell Biol. 11*, 1829−1839.
Boelens, W. C., Jansen, E. J., van Venrooij, W. J., Stripecke, R., Mattaj, I. W., & Gunderson, S. I. (1993) *Cell 72*, 881−892.
Celander, D. W., & Cech, T. R. (1990) *Biochemistry 29*, 1355−1361.
Darsillo, P., & Huber, P. W. (1991) *J. Biol. Chem. 266*, 21075−21082.
Dock Bregeon, A. C., Westhof, E., Giege, R., & Moras, D. (1989) *J. Mol. Biol. 206*, 707−722.
Glickman, J. N., Howe, J. G., & Steitz, J. A. (1988) *J. Virol. 62*, 902−911.
Gubser, C. C., & Varani, G. (1996) *Biochemistry 35*, 2253−2267.
Gunderson, S. I., Beyer, K., Martin, G., Keller, W., Boelens, W. C., & Mattaj, L. W. (1994) *Cell 76*, 531−541.
Hall, K. B. (1994) *Biochemistry 33*, 10076−10088.
Hall, K. B., & Stump, W. T. (1992) *Nucleic Acids Res. 20*, 4283−4290.
Hoffman, D. W., Query, C. C., Golden, B. L., White, S. W., & Keene, J. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2495−2499.
Howe, P. W., Nagai, K., Neuhaus, D., & Varani, G. (1994) *EMBO J. 13*, 3873−3881.
Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C., & Nagai, K. (1991) *EMBO J. 10*, 3447−3456.
Jovine, L., Oubridge, C., Avis, J. M., & Nagai, K. (1996) *Structure 4*, 621−631.
Krol, A., & Carbon, P. (1989) *Methods Enzymol. 180*, 212−227.
Kwakman, J. H., Konings, D. A., Hogeweg, P., Pel, H. J., & Grivell, L. A. (1990) *J. Biomol. Struct. Dyn. 8*, 413−430.
Lührmann, R., Kastner, B., & Bach, M. (1990) *Biochim. Biophys. Acta 1087*, 265−292.
Lutz Freyermuth, C., Query, C. C., & Keene, J. D. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6393−6397.
Mougel, M., Eyermann, F., Westhof, E., Romby, P., Expert Bezancon, A., Ebel, J. P., Ehresmann, B., & Ehresmann, C. (1987) *J. Mol. Biol. 198*, 91−107.
Nagai, K., Oubridge, C., Jessen, T. H., Li, J., & Evans, P. R. (1990) *Nature 348*, 515−520.
Nelissen, R. L., Sillekens, P. T., Beijer, R. P., Geurts van Kessel, A. H., & van Venrooij, W. J. (1991) *Gene 102*, 189−196.
Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., & Nagai, K. (1994) *Nature 372*, 432−438.
Scherly, D., Boelens, W., van Venrooij, W. J., Dathan, N. A., Hamm, J., & Mattaj, I. W. (1989) *EMBO J. 8*, 4163−4170.
Scherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J., & Mattaj, I. W. (1990) *Nature 345*, 502−506.
Stump, W. T., & Hall, K. B. (1995) *RNA 1*, 55−63.
Tsai, D. E., Harper, D. S., & Keene, J. D. (1991) *Nucleic Acids Res. 19*, 4931−4936.
van Gelder, C. W., Gunderson, S. I., Jansen, E. J., Boelens, W. C., Polycarpou Schwarz, M., Mattaj, I. W., & van Venrooij, W. J. (1993) *EMBO J. 12*, 5191−5200.
van Gelder, C. W., Thijssen, J. P., Klaassen, E. C., Sturchler, C., Krol, A., van Venrooij, W. J., & Pruijn, G. J. (1994) *Nucleic Acids Res. 22*, 2498−2506.
Varani, G. (1995) *Annu. Rev. Biophys. Biomol. Struct. 24*, 379−404.
Williams, D. J., & Hall, K. B. (1996) *J. Mol. Biol. 257*, 265−275.

BI9623237