Nucleic Acids Research, Vol. 20, No. 6 1215–1221

An additional long-range interaction in human U1 snRNA

Christine Sturchler, Philippe Carbon and Alain Krol*
Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue René Descartes, 67084 Strasbourg Cedex, France

Received January 30, 1992; Revised and Accepted February 25, 1992

ABSTRACT

We present evidence for the existence of an additional long-range interaction in vertebrate U1 snRNAs. By submitting human U1 snRNP, HeLa nuclear extracts, authentic human or X. laevis in vitro transcribed U1 snRNAs to RNase V1, a nuclease specific for double-stranded regions, cleavages occurred in the sequence ΨΨACC (positions 5–9) residing in the 5′ terminal region of the RNA. The RNase V1 sensitive region is insensitive to single-stranded probes, something unexpected knowing that it was considered single-stranded in order to base-pair to pre-mRNA 5′ splice site. We have identified the sequence GGUAG (positions 132–136) as the only possible 3′ partner. Mutants, either abolishing or restoring the interaction between the partners, coupled to an RNase V1 assay, served to substantiate this base-pairing model. The presence of this additional helix, even detected in nuclear extracts under in vitro splicing conditions, implies that a conformational change must occur to release a free U1 snRNA 5′ end.

INTRODUCTION

The removal of introns from mRNA precursors is an essential step in the expression of eukaryotic genes. This process is achieved by a multicomponent ribonucleoprotein complex called the spliceosome. An assembled spliceosome consists of four separate small ribonucleoprotein particles, the U1, U2, U5, U4/U6 snRNPs, and auxiliary factors that may act independently of snRNPs. These snRNPs consist of one (U1, U2, U5) or two (U4/U6) snRNAs associated with at least seven common proteins (the Sm proteins) and a variable number of proteins unique to the particular snRNP (for a review, see ref. 1–3).

The U1 snRNP particle is essential for the splicing of pre-mRNAs in vitro. In higher eukaryotes, it is composed of at least 10 proteins: the B, B′, D, D′, E, F and G proteins constitute the common Sm proteins whereas proteins 70 K, A and C are specific for U1 (2). One function performed by the U1 snRNP takes place in the first step of the splicing event and involves its binding to the 5′ splice site of the mRNA precursor through RNA-RNA interactions with the 5′ end of U1 snRNA (4–6). The recognition of the 5′ splice site by the 5′ terminal sequence of U1 snRNA required that this particular domain be single-stranded and readily accessible to this splice junction in the U1 snRNP. This has been shown to be the case by various approaches, including derivation of secondary structure models for U1 snRNAs and oligonucleotide-targeted degradation of U1 snRNA by RNase H (7–10). An additional role for the U1 snRNP has been described in mammalian spliceosome assembly which consists in promoting the stable binding of the U2 snRNP to the pre-mRNA branch site (11). Although the U1 snRNA–5′ splice site interaction is necessary for recognition of the 5′ splice site, it is not by itself sufficient. In an earlier report, it has been suggested that the proteins of the U1 snRNP also play a role in 5′ splice site recognition (12). More recently, it was shown that the U1 snRNP-specific C protein, which is needed for the binding of U1 snRNP to the 5′ splice site, might potentiate this base pairing (13). In addition, factors acting in trans could also aid the RNA duplex to form. Indeed, it has been proposed that two soluble factors participate in the recognition of the 5′ splice site (14).

In the study presented in this report, we addressed the question as to how is the 5′ terminal region of human U1 snRNA organized. In other words, we asked whether this functional domain possesses a large number of degrees of freedom as would be anticipated for a single-stranded region or rather displays a constrained conformation resulting from tertiary bonds. Data collected from the work described here revealed a surprise. We present evidence that the sequence ΨΨACC (positions 5–9) contained in the supposedly single-stranded 5′ terminal region of U1 snRNA forms a long range interaction with the GGUAG complementary partner located immediately downstream of the Sm binding site.

MATERIALS AND METHODS

Site-directed mutagenesis and DNA constructs

The starting template for engineering X. laevis U1 snRNA mutants was the X. laevis T7XlU1wt gene cloned downstream of a T7 promoter described in (15, 16) in which we removed two of the three Gs that immediately follow the T7 promoter. Construction of the mutants was done by using the Amersham in vitro mutagenesis kit. M13 RF DNAs were then subcloned into pUC18 or pUC19 vectors. All constructions were verified by DNA sequence analysis.

* To whom correspondence should be addressed

In vitro transcription of X. laevis wild-type or mutant U1 snRNAs

In vitro transcription under the control of the T7 promoter was performed in 50 μl of a mixture containing 5 μg of the linearized pUC templates (cut at BamHI or XhoI or elsewhere when indicated), 40mM Tris-HCl pH 8, 20 mM $MgCl_2$, 5 mM DTT, 1 mM spermidine, 0.5 mg/ml BSA, 40 units RNasin (Promega), 4 mM each NTP and 140 units of T7 RNA polymerase prepared from the T7 gene 1 (provided by Dr. F. William Studier). RNA transcripts were purified on 6% polyacrylamide gels and electroeluted in the cold using the Biotrap system (Schleicher and Schüll).

Nuclear extracts

Nuclear extracts from HeLa cells (the cells were kindly provided by the cell culture group at the LGME-Strasbourg) were prepared according to (17). The protein concentration was measured by the micro-Bradford assay.

Structure probing

The substrates for solution structure probing were either authentic or in vitro transcribed U1 snRNAs, human U1 snRNPs (kindly provided by Dr. R. Lührmann and prepared according to ref. 18) or the U1 snRNP contained in the HeLa nuclear extracts. Details concerning experimental conditions relevant to cleavage with RNases T2 or V1 were previously described (15, 19). Standard conditions were as follows unless otherwise stated in legends to Figures. RNase V1 cleavage occurred at 2.10^{-2} – 5.10^{-2} unit/μg RNA (tRNA was added except for probing nuclear extracts) in 10 mM Tris-HCl pH 7.5, 10mM $MgCl_2$, 100 mM KCl. Incubation was carried out at either 0°C, 20°C or 30°C. RNase T2 digestion was carried out at $2{,}5.10^{-3}$–5.10^{-2} unit/ μg RNA in similar conditions. Mapping of the cleaved bonds was performed by primer extension of a 5'-^{32}P labeled primer complementary to positions 124 –137.

RESULTS

The 5' end of U1 snRNA exhibits an unexpected sensitivity towards structural probes

In a work describing the solution structure of U1 snRNA, we reported that positions Ψ5, C8 and C9 were protected against the action of chemical probes (15; see also Figure 1 for their location in the U1 secondary structure). Protection of these bases implied either that the region in which they lie is not fully single-stranded or, if it is, that these positions might participate in tertiary bonding. Whatever the possibility, the peculiar behaviour of these three positions was puzzling since they were previously reported to be included in a perfectly single-stranded area (7–10) which base-pairs to the pre-mRNA 5' splice site (6). This incited us to undertake a detailed structural analysis of this 5' portion of the U1 snRNA molecule.

In preliminary experiments not described in (15), we observed a susceptibility to RNase V1 of a portion of U1 snRNA comprised between Ψ5 and C9. As this enzyme senses double-stranded RNA regions (15, 19 and ref. therein), there could be a correlation between sensitivity to RNase V1 and the chemical protection of Ψ5, C8 and C9. Therefore, the susceptibility to RNase V1 of the U1 snRNA 5' end was examined under various conditions. In the human U1 snRNP, cleavage occurs between positions C4 and C9 (Figure 2A, lanes 2–4). Digestion performed in nuclear extracts leads to a profile similar to that obtained with the isolated U1 snRNP (Figure 2B, lanes 5–7). Again, and more surprisingly, digestion performed in nuclear extracts but in conditions used for obtaining in vitro splicing, i.e. 2mM $MgCl_2$, 50 mM KCl and incubation at 30°C (4), yields also an RNase V1 digestion pattern (Figure 2B, lanes 2 and 3). The same experiment, performed this time on phenol-extracted U1 snRNA shows that the 5' end of the naked RNA is also digested by RNase V1 (Figure 2C, lanes 2–5), ensuring that the cutting positions were not simply the result of a conformation induced by the presence of the U1 snRNP proteins. There is slight variations in position and cleavage intensities when looking at U1 snRNAs on the one hand or U1 snRNP alone or in nuclear extracts, on the other. At this point, it is difficult to explain these differences which might originate from the different substrates employed.

Lastly, T7XlU1wt which is an X. laevis U1 snRNA produced by in vitro transcription under the control of a T7 promoter also yields an RNase V1 profile in the C4-U10 area. (Figure 2D, lanes 2–4). A moderate variation is observable between the cleavage patterns of the transcribed and authentic U1 snRNAs, in particular for both the shifting of the cleavage positions and the weakness of the U6/A7 cut in the transcribed RNA. This was reproducible in at least 90% of the experiments but we cannot provide any explanation yet. However, obtention of cleavages with an in vitro transcript indicates that neither the trimethylguanosine cap structure nor the few modified bases contained in an authentic U1 snRNA, but absent in the T7 transcript, have a profound influence on the production of the cuts. Interestingly, in Figure 2D and to a lesser extent in Figure 2C, the intensity of the C8/C9 RNase V1 cleavage is as strong as that observed for the A19-A21 positions which are located in a double helical piece of RNA, a preferred target for RNase V1.

Although there is a modest variation between the RNase V1 profiles presented in the above data, the important point to make is to underline that this enzyme does cleave. Thus, from (15) and the above data we conclude that based on chemical and enzymatic probing, the 5' end of U1 snRNA does not respond

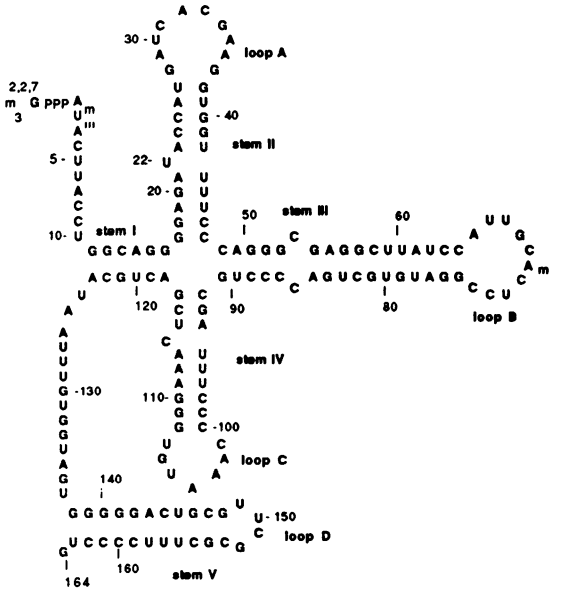

Figure 1. The secondary structure model of human U1 snRNA derived from (7, 15).

as expected from a single-stranded RNA. We could therefore propose that part of this region is base-paired with a partner, even in splicing extracts. However, knowing that RNase V1 can also cleave single-stranded RNA that adopts an helical conformation (20), the following experiments were carried out in order to distinguish between these possibilities.

The 3′ region between 129–164 is required for the acquisition of a peculiar structure at the 5′ end

If RNase V1 can cleave between Ψ5 and C9 simply because this segment would adopt an helical conformation, then it should also be cleaved by nucleases specific for single-stranded regions. The U1 snRNP and splicing extracts were treated by RNase T2 in buffer conditions identical to those used for probing with RNase V1. In the U1 snRNP, the Ψ5-C9 region is not cleaved by RNase T2 (Figure 2A, lanes 6 and 7). We verified that absence of cleavage in this region was not due to the inactivity of the enzyme by monitoring its activity for other regions of the molecule; we also verified that absence of cleavage was not an effect of base preferences of RNase T2 by employing S1 nuclease at neutral pH which did not cut either (data not shown). The RNase T2 digestion profile of the U1 snRNP contained in HeLa nuclear

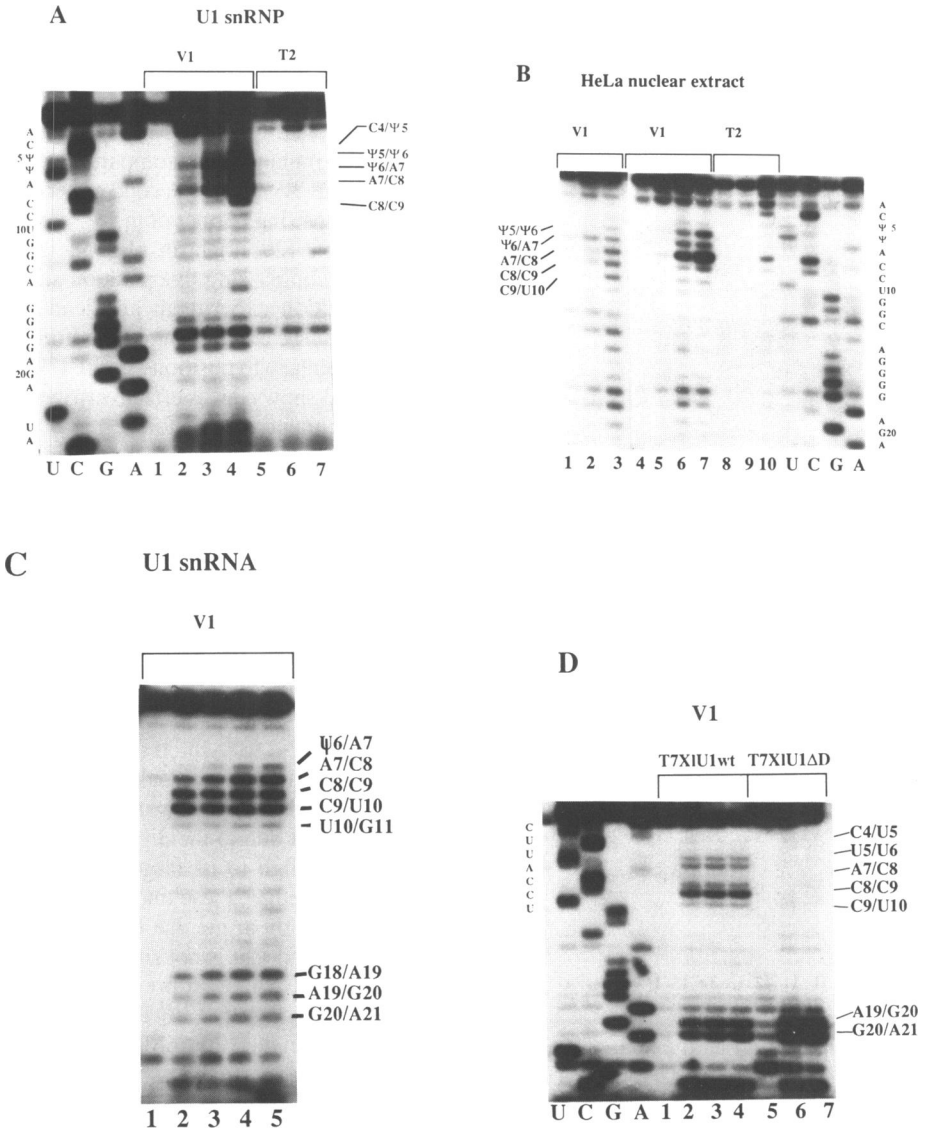

Figure 2. The 5′ end of U1 snRNA exhibits an unexpected sensitivity towards structural probes. **(A)** Human U1 snRNP (approximately 1 μg of protein equivalent per time-point) was digested by RNase V1 for 5, 10 and 15 min (lanes 2, 3, 4, respectively) or by RNase T2 for 5 and 10 min (lanes 6 and 7, respectively, according to Materials and Methods). Lanes 1 and 5 are control experiments with no enzyme added and incubated for the longest time-point. The sequence (obtained by the chain termination method using the same labeled primer and reverse transcriptase) is indicated on the left, positions cleaved by RNase V1 on the right. **(B)** HeLa nuclear extracts (5 μl at 7.5 mg/ml protein concentration per time-point) were treated by RNase V1 (0.06 unit per time-point) for 5 min (lane 2) and 10 min (lane 3) under in vitro splicing conditions (4, 5) or for 5, 10, 15 min (lanes 5–7) under standard conditions. Lanes 1 and 4 are control experiments without enzyme. RNase T2 digestion was performed under in vitro splicing conditions with 0.04 unit per time-point, for 5 min and 10 min (lanes 9 and 10). Lane 8 is the control without enzyme. Lanes 1–3 and 4–10 result from separate experiments. **(C)** Phenol-extracted human U1 snRNA (250 ng per time- point) treated by RNase V1 for 5, 10, 15, 20 min (lanes 2–5, respectively) or incubated without enzyme for 20 min (lane 1). **(D)** RNase V1 treatment of in vitro transcribed X. laevis wild-type U1 snRNA (T7XlU1wt) or of the truncated version of it (T7XlU1ΔD) explicited in the text. Digestion (1 μg of transcribed RNA per time-point) was for 5, 10, 15 min (lanes 2–4), 5 and 10 min (lanes 6, 7). Lanes 1 and 5 represent treatment without enzyme.

extracts is shown in Figure 2B, lanes 9 and 10. While cleavage occurs between positions A3/C4 and A7/C8, actually overlapping RNase V1 cleavage for this latter position, phosphodiester bonds comprised between C4 and A7 are barely, if not at all, cut.

The RNase T2 and S1 nuclease insensitivities of a portion of the U1 snRNA 5′ end suggest that it is not single-stranded, at least along the whole A1-U10 sequence. To test this hypothesis and to seek the existence of a putative pairing partner, we pursued our investigations further by taking advantage of in vitro transcribed U1 snRNAs. The T7XlU1wt DNA was linearized at naturally occurring restrictions sites, yielding truncated U1 snRNAs ending at either C31 (Sau3AI) or C46 (MnlI) or G117 (TaqI) which were submitted to RNase V1 treatment. Construct T7XlU1ΔD (16), which carries a XhoI site substituting AUAAUU to CUCGAG (positions 124−129), was linearized at the XhoI site, thereby generating an RNA ending at position 128. The combined use of all the truncated mutant U1 snRNAs shows that the sole amputation of the region comprised between

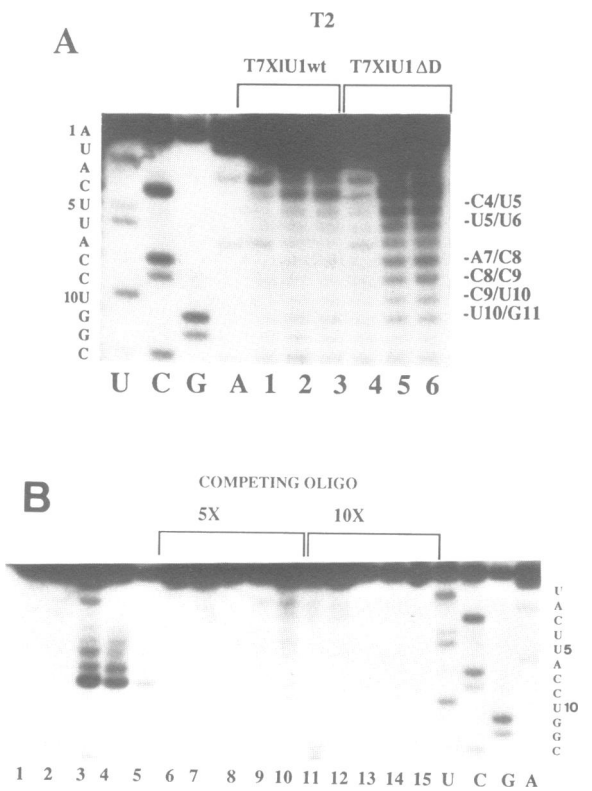

Figure 3. The 3′ region between 129−164 is required for the acquisition of a peculiar structure at the 5′ end. **(A)** RNase T2 treatment of the in vitro transcribed X. laevis wild-type (T7XlU1wt) or truncated (T7XlU1ΔD) U1 snRNAs. Digestion was for 5 min (lanes 2 and 5) and 10 min (lanes 3 and 6) under conditions described in Figure 2B and D. In lanes 1 and 4, the enzyme was omitted. **(B)** RNase V1 digestion experiment after incubation of the T7XlU1wt RNA with a competing oligo. The RNA (1 μg) contained in the RNase V1 buffer was incubated with increasing 2× (not shown), 5× and 10× molar concentrations of the competing oligodeoxynucleotide (complementary to U1 positions 124−137). After a 20 min incubation at room temperature, the RNA-DNA duplex was transferred to ice and submitted to RNase V1 digestion under standard conditions. T7XlU1wt RNA was digested with RNase V1 in absence of the oligo for 5, 10, 15 min (lanes 3−5. In the latter, part of the material was lost). Lanes 1 and 2 are controls without enzyme. In lanes 8−10 and 13−15, digestion was performed in the presence of increasing molar amounts (5×, 10×) of the oligo for the same times as in lanes 3−5. Lanes 6, 11 and 7, 12 represent controls without enzyme but containing the oligo in lanes 7 and 12.

129 and the 3′ end was sufficient to abolish RNase V1 cleavages at the 5′ end (data not shown and Figure 2D, lanes 5−7). This effect cannot be attributed to the XhoI sequence since a construct carrying this substitution, but not cleaved by XhoI thereby generating a full length transcript, still enables RNase V1 cleavages to occur (not shown).

The above data eliminate the possibility that the 5′ end adopts a single-stranded conformation that would account for its being responsive to RNase V1. This responsiveness is rather due to the existence of a long-range interaction. Two sets of experiments substantiate this finding. First, if our assumption that the 5′ end is engaged in a base-pairing interaction with a sequence residing between 129 and 164, then submitting the truncated T7XlU1ΔD RNA to RNase T2 digestion should release cleavage at the 5′ end since this RNA would be exempt of this long-range interaction. Comparison of the T7XlU1wt and T7XlU1ΔD RNA digestion patterns reveals that this is effectively what happens since the wild-type does not exhibit RNase T2 cuts while the truncated mutant does (compare lanes 2 and 3 with lanes 5 and 6 in Figure 3A). In the second experiment, we exploited the fact that if the pairing partner of the 5′ sequence is located downstream of position 129, then an oligonucleotide complementary to a region encompassing this position should disrupt the binding in a competition experiment and lead to abolition of RNase V1 cleavages. The experiment was performed with an oligonucleotide complementary to positions U124-U137 of the T7XlU1wt RNA. Figure 3B shows this is effectively the case, regardless of whether the experiment was performed at 0°C or 20°C. At a five-fold molar excess of the oligo, there is a residual amount of RNase V1 cleavage (lane 10) which disappears at a 10-fold molar excess. A control experiment employing an oligonucleotide not complementary to any U1 sequence as the competitor was performed and in this case RNase V1 cleavage still occurs (not shown). The same competition experiment could not be realized with U1 snRNP or nuclear extracts. The reason is that the proteins attached to the Sm site undoubtedly mask the sequence complementary to the oligonucleotide. We also tried to perform the same experiment with a 5 mer oligo complementary to positions 132−136 which should be available for binding in the U1 snRNP. In both the naked RNA and the particle, the chase of RNase V1 cleavage was not reproducibly obtained. This is in major part due to the difficulty in obtaining stable binding of the 5 mer oligo to its target site on the RNA.

Taken together, these findings and the chemical data described in (15) strongly argue that RNase V1 senses a region of the U1 snRNA 5′ end that is not single-stranded but rather base-paired, in contrast to what was previously established (7−10). GGUA-G (positions 132 to 136) is the only sequence complementary to ΨΨACC in the 124 to 137 area, location selected by the above competition experiment which itself eliminated the possibility of a partner located in the 138−164 region, as deduced from using the truncated T7XlU1ΔD transcript. GGUAG is therefore a likely candidate for participating in an interaction with the ΨΨACC sequence (Figure 1).

A long-range interaction involving the ΨΨACC (positions 5 to 9) and GGUAG (positions 132 to 136) sequences

To establish whether ΨΨACC and GGUAG are the partners of a long-range interaction, we introduced clustered point mutations abolishing or restoring the proposed pairing scheme and their capabilities to do so were assayed by their RNase V1 sensitivities. As we are using here in vitro transcribed RNAs lacking modified

bases, we will refer to U5/U6 instead of Y5/Y6. Mutants S1 and S2 (Figure 4C) substitute either UAG (positions 134−136) to CCC in mutant S1 or UUA (positions 5−7) to GGG in mutant S2. They carry mutations in either of the pairing partner and were therefore designed to abolish the base-pairing. The double mutant S3 (Figure 4C) results from the combination of S1 and S2 and restores a five base-pair complementarity by compensatory base changes. The results of the RNase V1 assay (diagrammed in Figure 4C) show that when the complementarity for only two CG base-pairs is afforded, like in mutant S1, no RNase V1 cleavage occurs (compare lanes 6, 7 with lanes 2−4 in Figure 4A). In mutant S2 in which the complementarity is extended by one base-pair, only a fraction of the wild-type RNase V1 cleavages is restored, with the cuts occurring between C8/C9 and C9/U10 (compare lanes 9−11 with lanes 2−4 in Figure 4A). When the 5 base-pair complementarity is furnished to mutant S3, cuts are produced between G5 and C8 in addition to those between C8/C9 and C9/U10 also present in mutant S2. This might be related to the fact that the complementarity introduced by the three GC compensatory base changes can actually restore the 5 base-pair long interaction in vitro.

To confirm this first indication, another series of mutants was created which changed more profoundly the sequence of each of the partners (S4 to S7 shown in Figure 4C. The results are

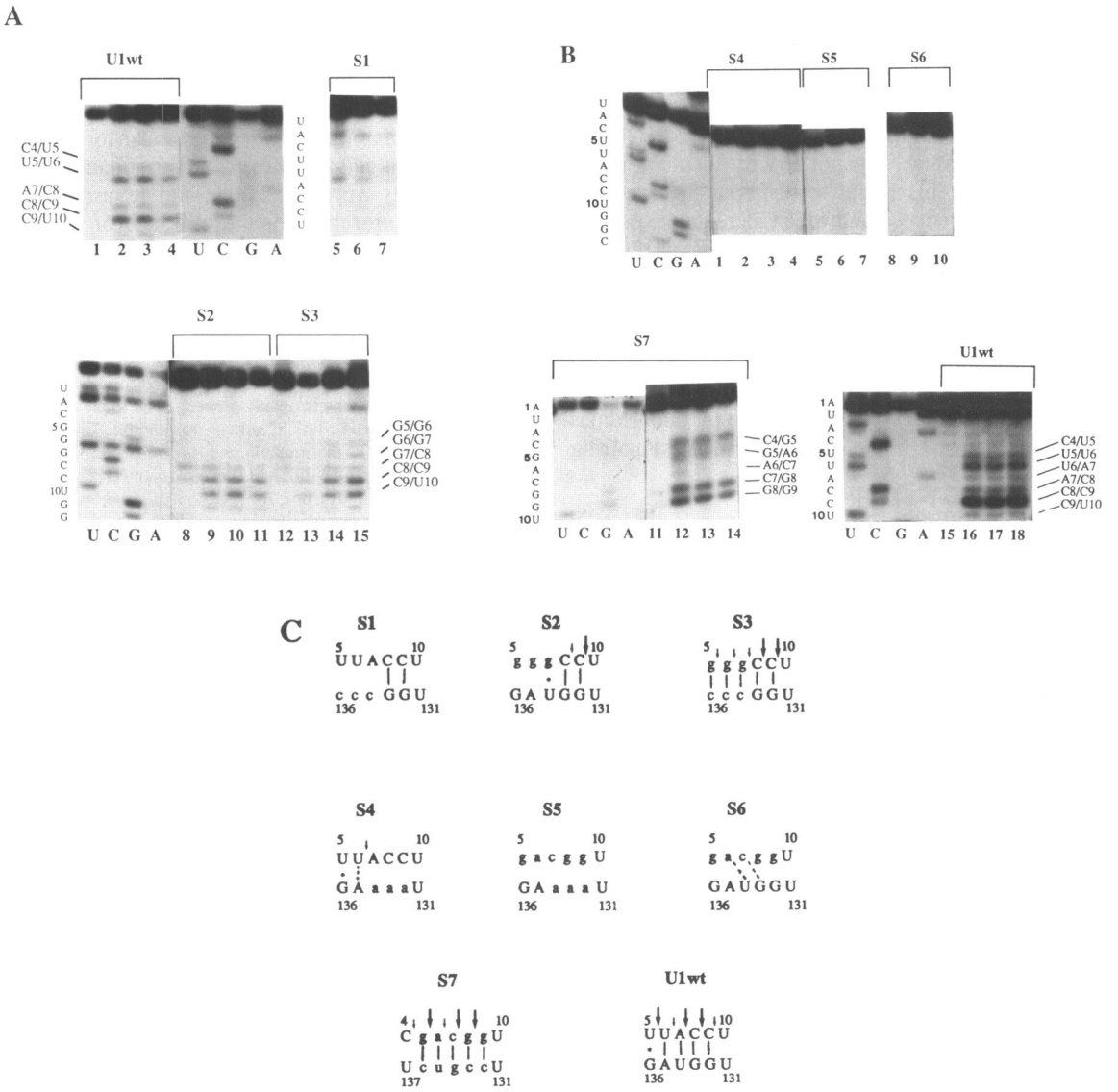

Figure 4. RNase V1 probing of various site-directed mutants of the proposed long-range interaction. **(A)** The S1, S2 and S3 mutants whose sequence is shown in Figure 4C were treated by RNase V1 for 5 min (lanes 6, 9 and 13), 10 min (lanes 7, 10 and 14) and 15 min (lanes 11 and 15). Lanes 5, 8 and 12 are controls without enzyme. The in vitro transcript T7XlU1wt (U1wt) was also treated by RNase V1 as a control experiment for 5, 10, 15 min (lanes 2, 3, 4 respectively). Lane 1: no enzyme added. Different migrations are assembled in a mountaged rearrangement. **(B)** RNase V1 digestion of the S4, S5, S6 and S7 mutants shown in Figure 4C. Digestion was for 5 min (lanes 2, 6, 9 and 12), 10 min (lanes 3, 7, 10, 13) and 15 min (lane 14). RNase V1 digestion of the T7XlU1wt (U1wt) for 5 min, 10 min and 15 min (lanes 16, 17, 18, respectively). In lanes 1, 5, 8, 11 and 15, no enzyme was added. That in lanes 2−4 and 6−10 the enzyme actually worked was attested by its capability to cleave the AGA (positions 18−21) region (not displayed in this Figure). Autoradiographs for S4, S5 and S6 mutants arise from different sets of experiments. **(C)** Sequence of the S1 to S7 mutants. Mutations that alter the wild-type sequence (U1 wt) are in lower-case letters. Dotted lines are for tentative base-pairing. Arrows point to the RNase V1 cleavages deduced from Figure 4A and B. The length of the arrows stand for the intensity of the cuts.

summarized in the same Figure). Substitution of sequence GGU (positions 132–134) by AAA in mutant S4 leads to an almost complete disappearance of the RNase V1 cuts (Figure 4B, lanes 2–4). To ascertain that no residual base-pairing in mutant S4 could account for the very faint band observed between U6 and A7 (Figure 4B, lanes 2–4), the sequence of the 5′ partner of mutant S4 was also substituted (UUACC to GACGG). This yields mutant S5 harboring substitutions in both pairing partners between which no sequence complementarity can be found. It has a more profound effect than S4 since no RNase V1 cleavage is detected (lanes 6, 7 in Figure 4B). Mutant S6 carries the same 5′ substitution as mutant S5 but retains the wild-type 3′ partner sequence. Such a mutation does not allow the formation of any stable interaction and, again, this is corroborated by the complete absence of cleavage (Figure 4B, lanes 9, 10).

Remarkably, when in mutant S7 the 5 base-pair complementarity is restored with the use of compensatory base changes introduced into the 3′ partner, then the RNase V1 cleavage pattern is recovered (compare lanes 12–14 with lanes 16–18 in Figure 4B). In both the S3 and S7 mutants, the RNase V1 cleavage profile is similar, but not strictly identical to the wild-type profile. In S3, this might result from the sequence changes that have been introduced and which increase the number of GC pairs with respect to the wild-type. These can provoke a stacking of the guanosine residues in the helix.

The above mutants unambiguously establish that the UUACC and GGUAG sequences base-pair in the in vitro transcribed U1 snRNA molecule, as shown in Figure 5. Although the experiments have not been performed with an authentic U1 snRNA molecule, it is very unlikely that the two pseudouridines which are found instead of U5 and U6 can change the pairing scheme.

DISCUSSION

We have determined experimentally the existence of an additional long-range interaction in human and Xenopus U1 snRNAs which involves a base-pairing between ΨΨACC (positions 5–9) and GGUAG (positions 132–136). It extends the previously found, phylogenetically conserved, long-range interaction (stem I in Figure 5) closing the 5′ and 3′ ends of the U1 snRNA core structure (8, 21). The presence of the additional helix was detected whether the RNA is naked in solution or complexed in the U1 particle, isolated or in nuclear extracts under conditions for splicing in vitro. As observed in the Results section, RNase V1 cleavages extended more towards the 3′ end in the U1 snRNA than in the U1 particle. This can be interpreted to mean that the Sm proteins whose binding site is lying just upstream of the sequence GGUAG might hinder access to the nuclease, which is not the case in the naked U1 snRNA.

The data presented here are not in agreement with our findings described in (7) in which we have shown that the 5′ end was single-stranded. However, at that time this region was sensed by using S1 nuclease at its normal acidic pH (pH ca. 4.5). This buffer condition certainly disrupted the long-range base-pairing explaining why we could not detect it. In this report, RNase T2 and S1 nuclease digestions were performed at a neutral pH. Also, in the work described in (7), we did not observe RNase V1 cuts at the 5′ end. This might originate from the fact that we were using a 3′ terminally labeled molecule, about 160 nt downstream from the area of interest which is too far for obtaining good RNA resolution on gels. However, we are unable to explain why S1 nuclease cuts were observable and RNase V1 cuts were not.

We have searched for the presence of a similar interaction between homologous regions of U1 snRNAs from a variety of phyla (Table 1). A sample of U1 snRNA species, from plants

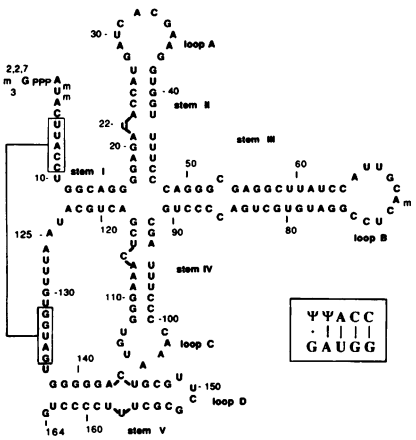

Figure 5. The secondary structure of the human U1 snRNA. The additional long-range interaction formed by the boxed sequences is blown-up in the inset.

Table 1. Proposed additional long-range base-pairings in various U1 snRNAs. Sequences are from the literature cited in parenthesis. Nucleotide modifications were inserted when experimentally proven. Solid triangles denote positions where Ψ were found instead of U in human, rat and chicken U1. U1 snRNAs from P. polycephalum, T. thermophila and S. pombe are shorter than the others at the 5′ end, explaining the different coordinates of certain nucleotides. In K. lactis, S. cerevisiae and S. uvarum, the location of the 3′ partner is different due to the insertion of the yeast specific domains. However, the sequence selected resides in a region homologous to that found in the short U1 snRNAs.

5 ▲▲ 9 UUACC GAUGG 136 132	5 9 UUACC UAUGC 140 136	1 5 AUACC GAUAG 125 121
Human (22), rat (22), mouse (22), chicken (22) X. laevis (22), X. tropicalis (our unpublished work), A. mexicanum (23), sea urchin (24), Drosophila (8), bean (22), soybean (22)	C. elegans (25)	P. polycephalum (26)
2 7 CUUACC GGAUCG 134 129	5 9 UUACC AGUCG 133 129	3 6 ACUU UGAC 133 130
T. thermophila (27)	U1.1	U1.36
	Molecular variants in P. sativum (28)	
5 9 ΨmUmACC G ACGG 131 127	1 5 ACUUA UGAGU 129 125	1 7 AUACUUA UAUGAGU 525 519
Chlorella (22)	S. pombe (29)	K. lactis (21)
5 9 UUACC AGUUU 560 556		
S. cerevisiae S. uvarum (21)		

to mammals, follows the experimentally tested base-pairing model due to strict sequence conservation. In the other species examined, the coordinates of the 3′ partner sequence are not strictly conserved but the overall location is. The extreme cases found in Saccharomyces and K. lactis large U1 snRNAs must be considered separately given the insertion of yeast specific domains. In T. thermophila, pea U1.36, S. pombe, and K. lactis, the base-pairing extends more towards the 5′ end of the molecule, the latter presenting an exception with a 7 bp complementarity. In P. polycephalum, the base complementarity is weak. It must be pointed out, however, that it is the only example known so far in which an A replaces a U in the phylogenetically conserved UU doublet (positions 5/6 relative to human U1 snRNA numbering). Wherever variations in the strength of the base-pairing occur, they are caused by mutations in the 3′ partner only. This arises from the fact that the evolutionary pressure prevents the drift of the 5′ partner sequence for obvious functional reasons and therefore does not enable the occurrence of compensatory base changes. The survey shown in Table 1 establishes that, eventhough the interaction is not stricto sensu phylogenetically supported since it can vary in the number of base-pairs or in register, it is nevertheless feasible in all U1 species examined. Conceivably, its stability could be increased by coaxial stacking onto the 6 bp helix I.

At first glance, the interaction presented here might appear provocative with regard to the known properties of the U1 snRNA 5′ end. Specifically, these establish that it must be single-stranded to allow base-pairing with pre-mRNA 5′ splice site and, as a corollary, that in vitro oligonucleotide targeted degradation of this region by RNase H abolishes splicing in vitro (4, 5, 9, 10, 30). The additional interaction identified in U1 snRNA implies it should be disrupted to release a free U1 snRNA 5′ end. It can be retorted that opening of the 5bp helix will not require the consumption of a large amount of energy and refolding of RNA components of the spliceosome is not unprecedented with the cases of the U4/U6 and U2/U6 base-pairings (3, 31, 32). Second, the site-directed degradation assay with RNase H employed to abolish splicing in vitro was also used to demonstrate that the U1 snRNA 5′ end was single stranded in the U1 snRNP. This is not inconsistent with our findings since these experiments were performed in the presence of either a long oligo (9) or a large excess of it (4, 5,10, 30), those conditions which are optimal for disrupting a preexisting base-pairing and allowing subsequent binding of the oligo.

What might be the necessity for masking a functional region in U1 snRNA? At present, explanations can only be speculative. The more straightforward one is that such an interaction might protect the region which base-pairs to pre-mRNA. A consequence of the interaction portrayed in Figure 5 is that the single-stranded Sm binding site, positions 125−131, becomes excluded as an internal loop between both short helices. It therefore loses degrees of freedom, is constrained and highly exposed, gaining a geometry favorable for optimal interactions with the Sm proteins.

ACKNOWLEDGEMENTS

We are grateful to R. Lührmann, M. Acker and the cell culture group at the LGME-Strasbourg, F. William Studier, J. Hamm and I. Mattaj for their generous gifts of human U1 snRNP, HeLa cells, T7 gene 1, T7XlU1wt and T7XlU1ΔD DNAs, respectively. A. Theobald is thanked for her unvaluable help in purifying T7 RNA polymerase, A. Lescure, E. Myslinski and E. Westhof for critical readings and stimulating discussions and A. Hoeft for oligodeoxynucleotide synthesis.

REFERENCES

1. Steitz, J.A., Black, D.L., Gerke, V., Parker, K.A., Krämer, A., Frendewey, D. and Keller, W. (1988) In Birnstiel, M.L. (ed.). *Structure and function of major and minor small nuclear ribonucleoprotein particles*. Springer-Verlag, Heidelberg, pp. 115−154.
2. Lührmann, R., Kastner, B. and Bach, M. (1990) *Biochim. Biophys. Acta.* **1087**, 265−292.
3. Guthrie, C. (1991) *Science* **253**, 157−163.
4. Black, D.L., Chabot, B. and Steitz, J.A. (1985) *Cell* **42**, 737−750.
5. Krämer, A., Keller, W., Appel, B. and Lührmann, R. (1984) *Cell* **38**, 299−307.
6. Zhuang, Y. and Weiner, A.M. (1986) *Cell* **46**, 827−835.
7. Branlant, C., Krol, A., Ebel, J.P., Gallinaro, H., Lazar, E. and Jacob, M. (1981) *Nucl. Acids Res.* **9**, 841−858.
8. Mount, S.M. and Steitz, J.A. (1981) *Nucl. Acids Res.* **9**, 6351−6368.
9. Lazar, E., Jacob, M., Krol, A. and Branlant, C. (1982) *Nucl. Acids Res.* **10**, 1193−1201.
10. Rinke, J., Appel, B., Blöcker, H., Frank, R. and Lührmann, R. (1984) *Nucl. Acids Res.* **12**, 4111−4126.
11. Barabino, S.M.L., Blencowe, B.J., Ryder, U., Sproat, B.S. and Lamond, A.I. (1990) *Cell* **63**, 293−302.
12. Mount, S.M., Petterson, I., Hinterberger, M., Karmas, A. and Steitz, J.A. (1983) *Cell* **33**, 509−518.
13. Heinrichs, V., Bach, M., Winkelmann, G. and Lührmann, R. (1990) *Science*, **247**, 69−72.
14. Stolow, D.T. and Berget, S.M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 320−324.
15. Krol, A., Westhof, E., Bach, M., Lührmann, R., Ebel, J.P. and Carbon, P. (1990) *Nucl. Acids Res.* **18**, 3803−3811.
16. Hamm, J., Kazmaier, M. and Mattaj, I.W. (1987) *EMBO J.* **6**, 3479−3485.
17. Digman, J.D., Lebovitz, R. and Roeder, R.G. (1983) *Nucl. Acids Res.* **11**, 1475−1489.
18. Bach, M., Bringmann, P. and Lührmann, R. (1990) *Methods in Enzymology* **181**, 232−257.
19. Krol, A. and Carbon, P. (1989) *Methods in Enzymology* **180**, 212−227.
20. Lowman, H.B. and Draper, D.E. (1986) *J. Biol. Chem.* **261**, 5396−5403.
21. Kretzner, L., Krol, A. and Rosbach, M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 851−855.
22. Guthrie, C. and Patterson, B. (1988) *Annu. Rev. Genet.* **22**, 387−419.
23. Murgo, S., Krol, A. and Carbon, P. (1991) *Gene* **99**, 163−170.
24. Brown, D.T., Morris, G.F., Chodchoy, N., Sprecher, C. and Marzluff, W.F. (1985) *Nucl. Acids Res.* **13**, 537−556.
25. Thomas, J., Lea, K., Zucker-Aprison, E. and Blumenthal, T. (1990) *Nucl. Acids Res.* **18**, 2633−2642.
26. Myslinski, E., Wilhelm, F.X. and Branlant, C. (1989) *Nucl. Acids Res.* **17**, 1019−1034.
27. Orum, H., Nielsen, H. and Engberg, J. (1991) *J. Mol. Biol.* **222**, 219−232.
28. Hanley, B.A. and Schuler, M.A. (1991) *Nucl. Acids Res.* **19**, 1861−1869.
29. Porter, G., Brennwald, P. and Wise, J.A. (1990) *Mol. Cell. Biol.* **10**, 2874−2881.
30. Krainer, A.R. and Maniatis, T. (1985) *Cell* **42**, 725−736.
31. Wu, J. and Manley, J.L. (1991) *Nature* **352**, 818−821.
32. Datta, B. and Weiner, A.M. (1991) *Nature* **352**, 821−824.