Volume 9 Number 23 1981 Nucleic Acids Research

Sequence of U1 RNA from Drosophila melanogaster: implications for U1 secondary structure and possible involvement in splicing

Stephen M.Mount and Joan A.Steitz

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06510, USA

Received 25 September 1981

ABSTRACT

U1 RNA from cultured Drosophila melanogaster cells (K_c) was identified by its ability to be recognized, as an RNP, by anti-(U1)RNP antibodies from human lupus patients. Its sequence was deduced largely from direct analysis of the RNA molecule and then confirmed by DNA sequence determinations on a genomic clone isolated by hybridization to Drosophila U1 RNA. The Drosophila U1 RNA sequence exhibits 72% agreement with human U1 RNA. Nucleotides 3-11, which are complementary to the entire consensus sequence for donor (5′) splice junctions in hnRNA, and to part of the acceptor (3′) consensus, are exactly conserved. However, nucleotides 14-21, postulated to interact only with acceptor junctions, differ. Comparison of the Drosophila U1 sequence with vertebrate U1 sequences allows a particular secondary structure model to be preferred over others. These results are consistent with the hypothesis that U1 snRNPs are involved in splicing, but suggest specific modifications of the model detailing molecular interactions between U1 RNA and hnRNA during the splicing reaction.

INTRODUCTION

U1 RNA is an abundant small nuclear RNA (snRNA) which was first observed in vertebrate cell nuclei over a decade ago (1). The first U1 RNA sequence (from rat) was published in 1974 (2). More recently, U1 RNAs from humans and chickens have been sequenced, and the original rat sequence has been corrected (3,4). In cell extracts, U1 RNA exists as a protein complex (snRNP) (5), which is sometimes found associated with heterogeneous nuclear RNA-protein complexes (hnRNP) (6,7). SnRNPs containing U1 are recognized by two classes of antisera from patients with the autoimmune disease systemic lupus erythematosus (5). One of these antibodies, known as anti-Sm, also binds snRNPs containing the small nuclear RNAs U2, U4, U5 or U6. A second class of antisera recognizes only U1 snRNPs; the historic name of this specificity is anti-RNP, but we will use the name anti-(U1)RNP in order to avoid confusion.

The 5′ terminus of vertebrate U1 RNA is m_3GpppAmUmACΨΨACCUGGCAGGGGAGA...

The underlined portion of this sequence is extensively complementary to the consensus $_{A}^{C}AG/GT_{G}^{A}AGT$ (7; S. Mount, in preparation) for sequences occurring at the 5′ ends of intervening sequences in nuclear genes specifying mRNAs. The overlined region of the U1 sequence is complementary to the consensus for the 3′ ends of intervening sequences, which is $(^{T}_{C})_{n}N^{C}_{T}AG/G$ (where n is usually between 15 and 30). These observations, together with information about the abundance, subcellular location and antigenic conservation of U1 snRNPs, led us (7) and others (8) to speculate that U1 snRNPs might play a role in splicing. More recently, the demonstration that anti-(U1)RNP antibodies inhibit splicing in isolated nuclei has provided experimental support for this idea (9). Invertebrates (10), protozoans (11) and plants (12) splice at least some of their nuclear-encoded mRNAs, and the splice junctions in these organisms conform nicely to the above consensus sequences (which are derived primarily from vertebrate genes).

If U1 snRNPs are indeed involved in the recognition of conserved splice junction sequences, then the 5′ terminus of U1 RNA should likewise be highly conserved. Here we present confirmation of that prediction in the case of the invertebrate Drosophila melanogaster. Sequences within U1 postulated to interact with donor splice junctions, and with the region of acceptor splice junctions immediately surrounding the splice point, are exactly conserved. However, the region of U1 RNA postulated to interact with the pyrimidine stretch preceding acceptor splice junctions is not conserved.

MATERIALS AND METHODS

Cell lines, maintenance, and labeling

The K_C cell line, supplied by Dr. Michael W. Young, Rockefeller University, was maintained in D22 medium as described (13). K_C cells were radioactively labeled in incomplete D22 (lacking yeast extract) for 10-15 hours in the presence of 100 μCi/ml $^{32}PO_4$. Friend erythroleukemia cells (mouse), from Dr. A. Sartorelli, Yale University, were maintained on RPMI 1640 (GIBCO), supplemented with 5% heat-inactivated bobby calf serum (GIBCO), 60 μg/ml penicillin, and 100 μg/ml streptomycin. Friend cells were radioactively labeled at 2 x 10^5 cells/ml in phosphate-free minimal essential medium (MEM-GIBCO) for 15-20 hours in the presence of 10 μCi/ml $^{32}PO_4$.

Sera

Sera from patients with lupus or related autoimmune disorders were provided by Dr. J. Hardin, Yale University, Dr. E. Tan, U. of Colorado, and

Dr. M. Reichlin, U. of Buffalo. Before use, sera were precipitated three times with 40% ammonium sulfate, dialyzed against cold 17.5 mM sodium phosphate, and cleared of precipitate (14), the preparations were then made 0.15 M in NaCl, 0.03 M in Tris pH 7.4, and stored in aliquots containing 8-20 $O.D._{280}$ units/ml of protein. The specificity of all sera was ascertained by analysis of immunoprecipitated RNAs as described (15) to insure that they were both monospecific with respect to anti-RNP activity and free of anti-DNA antibodies (see ref. 16).

RNA fractionation and analysis

Immune complexes were precipitated using Pansorbin (Calbiochem) and the extracted RNAs were fractionated in one dimension on polyacrylamide gels as previously described (15) except that the concentration of Tris borate in the gels and buffer was raised to 90 mM. RNAs were eluted by the crush-and-soak method (17).

T1 RNase and RNase A fingerprints were prepared as described by Barrell (18) using thin layer homochromatography on Cel PEI 300 (Brinkmann) in the second dimension (5). Oligonucleotides were subsequently eluted and analyzed by digestion with a second nuclease followed by separation either on DEAE paper at pH 3.5 (18), in the two-dimensional system used for fingerprint analysis, or in the chromatographic systems designed by Nishimura (19) for the analysis of modified nucleotides.

Unlabeled RNA was immunoprecipitated and labeled at the 5′ end using tobacco acid pyrophosphatase and polynucleotide kinase (20) or at the 3′ end using T4 RNA ligase and 5′-[^{32}P]-cytidine 5′,3′ bis phosphate (21). Enzymatic sequencing was as described by Donis-Keller et al. (22) and elaborated by Vournakis et al. (23). Wandering spot analysis was performed using the limited alkaline digestion conditions described for enzymatic sequencing (22,23) followed by fractionation in the two-dimensional system of Barrell (18).

Computer analysis of RNA secondary structure

Potential secondary structure interactions were identified using the dyad symmetry program of Queen and Korn (24). Thermodynamic stabilities were estimated using the stability numbers of Tinoco et al. (25) as detailed by Salser (26).

Cloning and Sequencing of Dm U1 DNA

A library of Canton S strain D. melanogaster DNA in the vector λ Charon 4A was kindly supplied by Dr. S. Artavanis-Tsakonas, Yale University. Plaques were screened using immunoprecipitated dU1 which had been partially

digested with T1 RNase (Calbiochem) in a reaction containing 300 ng of U1 RNA and approximately 150 pg of T1 RNase for 5 minutes on ice, and subsequently labeled with γ-^{32}P-ATP using T4 polynucleotide kinase (20). All cloning and sequencing procedures have been previously described (27,28). Filter hybridization, used to identify DNA fragments containing U1 sequences, was performed using the RNA probe detailed above as described (29), except for the omission of Denhardt's solution from the hybridization buffer.

RESULTS

Lupus antibodies cross-react with Drosophila RNPs

Lupus antibodies with anti-Sm or anti-(U1)RNP specificity represent a useful tool for identifying and purifying U1 RNA from various eukaryotic species. Their applicability in the case of D. melanogaster requires that the antigenic region of the fly U1 snRNP be sufficiently conserved for recognition by the human autoantibody. This was expected as both anti-Sm and anti-(U1)RNP had been shown to recognize snRNPs from Lepidopteran (fall army worm) cells (7).

Figure 1 shows that anti-Sm and anti-(U1)RNP do precipitate ^{32}P-labeled RNAs from K_c cells; the anti-Sm lane (lane 3) reveals 5 predominant RNAs, whereas anti-(U1)RNP precipitates one of these, called dU1 (lane 7). The particular preparation shown is atypical in that the material running with tRNAs is usually not visible in an anti-(U1)RNP precipitate (lane 7); rather, two U1 fragments (which will be discussed in relation to RNA sequencing) not seen here were frequently present. A U1 RNA fragment occasionally immunoprecipitated from mammalian cell extracts by either anti-(U1)RNP or anti-Sm sera (7,30), U1*, has never been observed in immunoprecipitates of extracts of K_c cells. Finally, note also that the Drosophila (U1)RNP antigenic determinant seems to be less cross-reactive with the human sera than the Drosophila Sm antigenic determinant (compare the darkness of the dU1 band in lanes 5 and 7).

The five Drosophila snRNAs shown in lanes 3 and 5 are designated dU2, dU1, dU4, dU5 and dU6 by analogy with their counterparts in vertebrate cells. Correspondence has been convincingly demonstrated not only for dU1 RNA (this paper), but also for dU6, which is sufficiently conserved to exhibit striking similarity with mammalian U6 (5) at the level of a T1 RNase fingerprint (not shown). In contrast to Sm and (U1)RNP, the lupus antigens called Ro and La (which are also found on small RNPs in mammalian cells (15,31)) are not detected in Drosophila cells (lanes 13-16). The antigens Jo

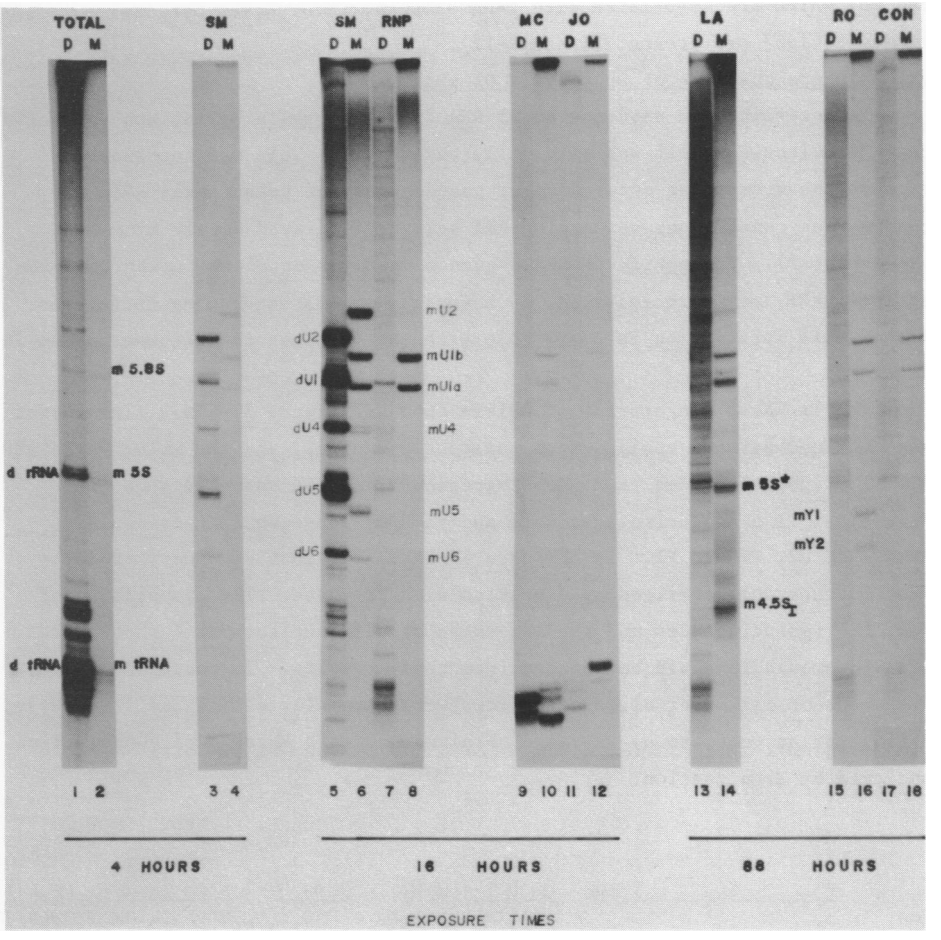

Figure 1. Immune precipitates from ^{32}P-labeled Drosophila and mouse cells
Small RNAs included in antibody precipitates from extracts of ^{32}P-labeled Drosophila K_c (D; odd number lanes) and mouse Friend erytholeukemia (M, even numbered lanes) cells were fractionated on a 10% polyacrylamide gel as described in Materials and Methods. Each lane represents precipitation from the same amount of extract (with the exception of lanes 1 and 2, which represent one tenth that amount). The time of exposure of the gel indicates the relative abundance of the RNA shown (or the relative reactivity of the serum used). The antibody used is designated at the top of each pair of lanes; individual RNAs have been previously identified (31) by fingerprint analysis as indicated. Lanes 17 and 18 were produced using non-immune sera.

and Mc, which are associated with tRNA subsets, (M.D. Rosa, personal communication) are present (lanes 9-12).

RNA sequence analysis of Drosophila U1 RNA

To determine the sequence of U1 RNA from Drosophila cells, analysis of the RNA molecule itself was originally undertaken. This was necessary because the widespread occurrence of pseudogenes for human small RNAs compromises conclusions concerning RNA sequences derived solely from DNA data (27). Drosophila U1 RNA sequence data was obtained, using RNA from anti-(U1)RNP immunoprecipitates, by a variety of RNA sequencing techniques. All of this information is compatible with and supports the sequence shown in Figure 2.

The T1 RNase fingerprint of dU1 is shown in Figure 3. This fingerprint was obtained without variation on numerous occasions, and the molar yield of all T1 oligonucleotides is in good agreement with the sequence shown in Figure 2. T1 oligonucleotides were analyzed by subsequent digestion with RNases A, U2, T2 and P1. In the case of RNase U2, most secondary products were further characterized by digestion with T2 RNase. The composition of all T1 oligonucleotides and the sequences of oligonucleotides 1 through 10 and 14 through 16 could be deduced from these results. In addition, the sequences of all other oligonucleotides were partially delineated. An RNase A fingerprint (not shown) was also obtained and each RNase A oligonucleotide analyzed by T1 digestion.

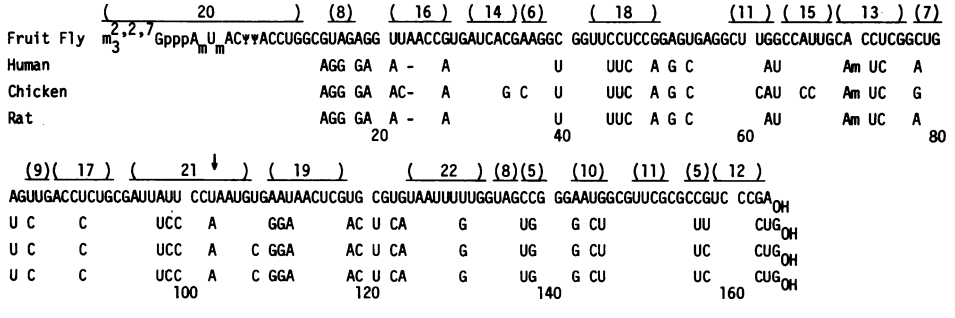

Figure 2. The sequence of U1 RNA from several species.
The fruit fly sequence, shown on the top line, was deduced as described in the text. Numbers in brackets indicate the positions of T1 RNase oligonucleotides shown in Fig. 3. The human, chicken and rat sequences are from reference no. 3, and only differences from the Drosophila sequence are indicated. An arrow indicates the position of a cleavage which generated two frequently observed dU1 fragments (see text).

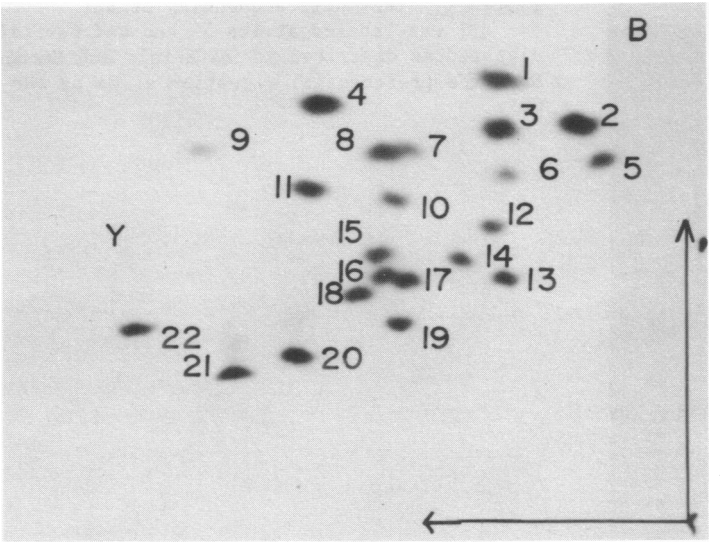

Figure 3. T1 RNase fingerprint of dU1.
This fingerprint was prepared as described in Materials and Methods. B and Y indicate the positions of the blue and yellow dyes in the first (horizontal) and second (vertical) dimensions.

Frequently, two fragments of dU1 were found in anti-(U1)RNP immunoprecipitates (not shown). T1 and RNase A fingerprints (together with confirming secondary digests) showed that these fragments were 5′ and 3′ pieces. They were therefore useful in localizing many oligonucleotides. The precise site of the cleavage that generates these two fragments is indicated by an arrow in Figure 2 and was deduced from the high yield of oligonucleotide 10 (AAUG) in a T1 RNase fingerprint of the 3′ fragment.

A great deal of sequence information was provided by enzymatic sequencing of end-labeled dU1 RNA (see Materials and Methods). Use of the enzymes CL3 and B. cereus ribonuclease enabled pyrimidine discrimination in most cases (see Figure 4). Gels could be read accurately up to 130 nucleotides from the labeled end. However, distortion of band spacing and inefficient cutting, presumably due to secondary structure (see Fig. 8), rendered nucleotides 137-164 unreadable whether the label was on the 3′ or the 5′ end. Additionally, pyrimidine discrimination based on the RNA sequencing gels alone, though usually proven later to be correct by the DNA sequence, was never convincing.

The sequence at the 3′ terminus of dU1 RNA was analyzed using antibody-

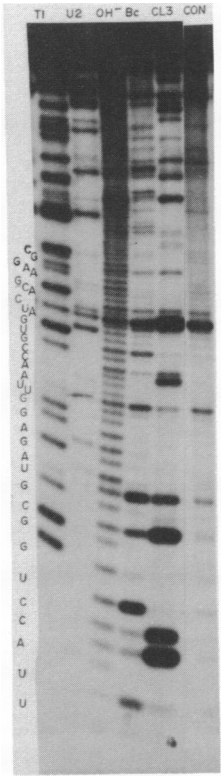

Figure 4. Enzymatic sequencing of dU1
dU1 was labeled at its 5′ end and partially digested as described in Materials and Methods. Note the preferential digestion at Cs by the enzyme CL3.

precipitated molecules labeled with 5′-[32P]-cytidine 5′,3′ bis phosphate by T4 RNA ligase. Minor heterogeneity in the RNA sequence seems to exist near, but not at, this end. Upon extended electrophoresis in a polyacrylamide gel, two, three or four closely spaced dU1 bands of comparable intensity were routinely seen. When these bands were individually digested to completion with alkali, most yielded Ap as the only (>90%) radioactive product, and complete digestion with RNase T1 reproducibly yielded ApCp. (Sometimes a band yielded Cp in the alkali digest and a trinucleotide of uncertain composition in the T1 digest.) These results suggest that most dU1 molecules end GA_{OH} (which is converted to $G_pA_p{}^*C_p$ upon ligation). If two of the closely spaced 3′ end-labeled dU1 bands were sequenced by enzymatic methods side by side on the same gel, identical sequences could be read from the two species; but corresponding bands in the region 5′ to nucleotide 137 were shifted in mobility by a single nucleotide. Two dimensional wandering spot analysis of the individual dU1 species invariably

produced a pattern like that shown in Figure 5, indicating a single sequence for the 3′ end of dU1 RNA. The pattern in this figure is clearly consistent with the DNA sequence (which was used to identify the shifts). Moreover, there is no indication of heterogeneity in any of the RNase T1 or A fingerprints produced from dU1. All the above findings suggest that heterogeneity in the dU1 sequence of K_c cells does exist, but that it is minor, is a single base deletion or insertion between nucleotides 137 and 158, and occurs in a position that has no effect on RNase A or T1 oligonucleotides. Alternatively, it is conceivable (but unlikely) that these results reflect the presence of stable conformational isomers of the dU1 molecule.

dU1 was scanned for modified or unusual nucleotides by two dimensional chromatography of uniformly labeled RNA digested to completion with either P1

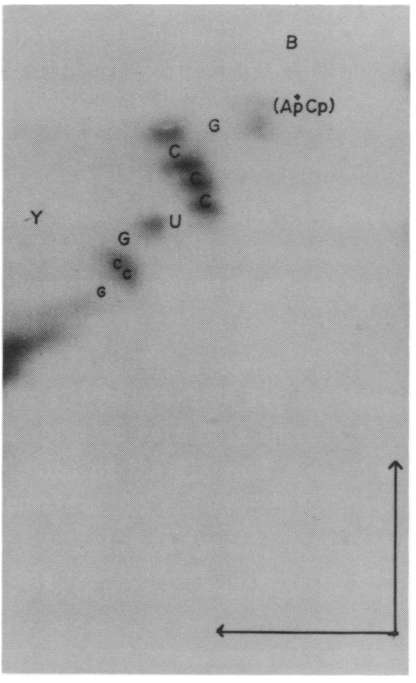

Figure 5. Wandering spot pattern of 3′ end-labeled dU1
dU1 labeled at its 3′ end with 5′-[^{32}P]-cytidine 5′,3′ bis phosphate was partially digested by alkali and the products were separated as described in Materials and Methods. Electrophoresis was from right to left and homochromatography was from bottom to top. B and Y indicate the positions of the blue and yellow dyes.

or T2 RNase. Figure 6 shows the results after T2 digestion of the entire molecule. Individual T1 oligonucleotides were also analyzed in this way (data not shown), revealing that only oligonucleotide 20, which contains the 5′ end of dU1, yields modified residues. Oligonucleotide 20 was therefore digested with U2, the products were separated by the two dimensional method used in fingerprint preparation, and each of the U2 oligonucleotides was separately analyzed for modified bases following digestion with P1 nuclease. From all these analyses the following could be concluded: 1) The ribose methylation of adenosine at position 70 (which is present in vertebrate U1) is absent in dU1 (see Figure 6); 2) all of the pseudouridine is located in positions 5 and 6; 3) these positions are completely modified (free of uridine); and 4) ribose methylated uridine is present at position 2. The ribose methylation in position 2 was confirmed by mobility shift analysis of kinased RNA, which revealed no shift corresponding to the loss of the A in position 3 (data not shown).

The 5′ cap of dU1 was shown to be 2,2,7trimethyl guanosine (m3G) by comparison with rodent U2, which is known to contain a m3G cap (32) (see Figure 7). When dU1 and mouse U2 were digested with tobacco acid pyrophosphatase, each released two radiolabeled products. Upon chromatography in the isobutyric acid system described by Nishimura (19), one

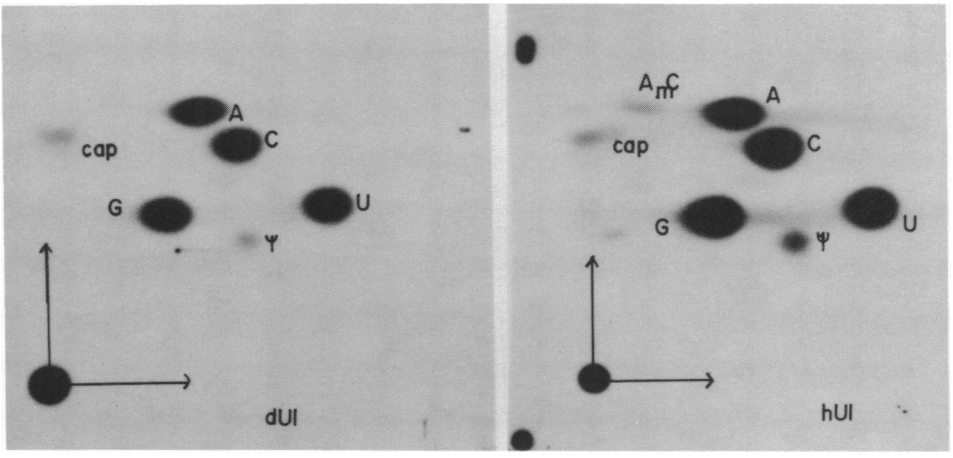

Figure 6. Modified base analysis of dU1
In vivo labeled, antibody precipitated, K_c cell U1 (designated dU1) or gel purified HeLa cell U1 (designated hU1) were digested with T2 RNase and separated as described (19). Note the absence of A_mC in the Drosophila pattern. "Cap" refers to the T2 resistant structure $m_3GpppAmUmAp$.

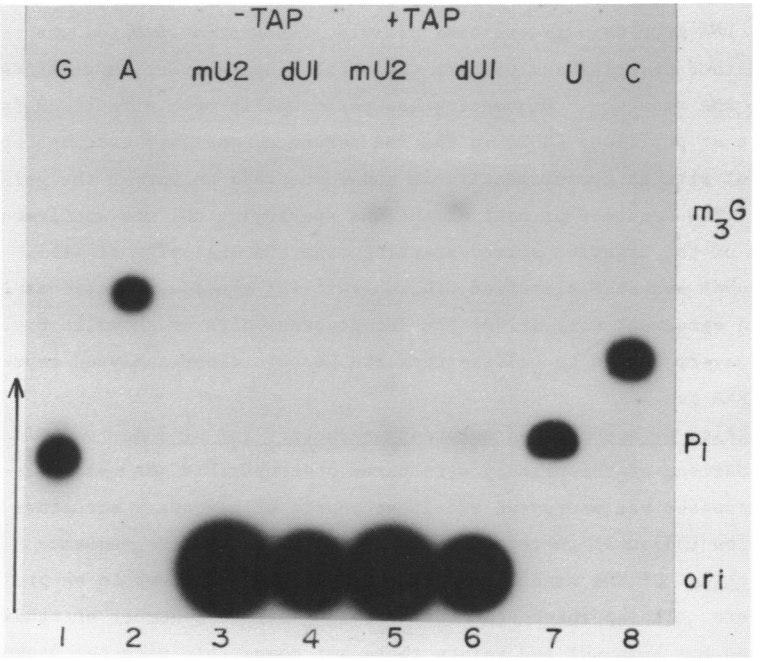

Figure 7. Identification of trimethyl guanosine
Chromatography was on Avicel plates in the isobutyric acid:ammonia system of Nishimura (19). Mouse U2 RNA (lanes 3 and 5) or antibody precipitated Drosophila U1 RNA (lanes 4 and 6) were treated with tobacco acid pyrophosphatase (B.R.L.) as described (20) (lanes 5 and 6) or tested in a mock reaction without enzyme (lanes 3 and 4). GMP, AMP, UMP and CMP markers are shown in lanes 1,2,7 and 8 respectively.

of these exhibits a mobility typical of orthophosphate; the other migrates ahead of the AMP marker, which itself is known to move faster than both 7 methyl G and 2,2 dimethyl G in this system. Because the dU1 cap migrates identically to m3G, but significantly differently from other nucleotides, it is reasonable to conclude that the dU1 cap is m3G.

Selection and sequencing of a genomic clone

Because uncertainties in a few regions of the dU1 sequence were not easily resolved by repeated analyses of the RNA itself, a genomic clone was isolated and sequenced. As detailed in Materials and Methods, a library of D. melanogaster (strain Canton S) DNA cloned into λ Charon 4A was screened by hybridization to ^{32}P-labeled dU1. A hybridizing recombinant phage was isolated, and the dU1 region was localized to a single 4.3 kilobase EcoR1

restriction fragment, which was then sublconed into the unique EcoRl site of pBR325. DNA from this recombinant plasmid, designated pDmUl.4, was sequenced by the method of Maxam and Gilbert (17) utilizing restriction sites deduced from the RNA sequence. Sequencing was performed in both directions from the HpaI site at positions 20-25 in the RNA sequence; analysis extended upstream to an AluI site at approximately -12 and downstream as far as the gels could be read. The sequence of most of the DNA specifying dUl was confirmed by analysis of the negative strand starting from the ThaI site at +153.

The DNA sequence clarified C/U ambiguities, elucidated residues 137-164, and is in agreement with all of the RNA sequence data as shown in Fig. 2. There is every reason to believe that the genomic clone analyzed represents a true Ul RNA gene.

Determination of a conserved secondary structure for Ul RNAs

Comparison of the primary structures of comparable RNA molecules from diverse species has proven an excellent source of secondary structure information (33). It therefore seemed reasonable that the sequence of Drosophila Ul RNA might allow certain RNA conformations to be preferred over others. The approach used here was to generate a number of candidate structures for human Ul and reject those not compatible with the Drosophila sequence. First, a list of possible secondary structure interactions was obtained using the dyad symmetry program of Queen and Korn (24). These were then combined in various ways to form seven candidate structures, each having a calculated stabilization energy ($-\Delta G$) of greater than 50 Kcal per mole (25). All but one of these human Ul RNA structures (and most of the secondary structure interactions from which the structures were derived) are inaccessible to Drosophila Ul RNA. The one structure available to Ul RNA from both species is shown in Figure 8. Note that the few differences in the rat and chicken Ul sequences do not prevent these RNAs from being drawn in this way.

DISCUSSION

Ul RNA has been postulated to perform a specific function in the processing of eukaryotic mRNA precursors by utilizing base pairing to align splice junctions in hnRNA (7,8). Our finding that the sequence of Drosophila Ul is completely conserved in the region complementary to the donor (5′) splice junction consensus sequence for all eukaryotic species adds further support to this hypothesis. Moreover, both the lack of conservation of the portion of the Ul sequence postulated to interact with the pyrimidine-

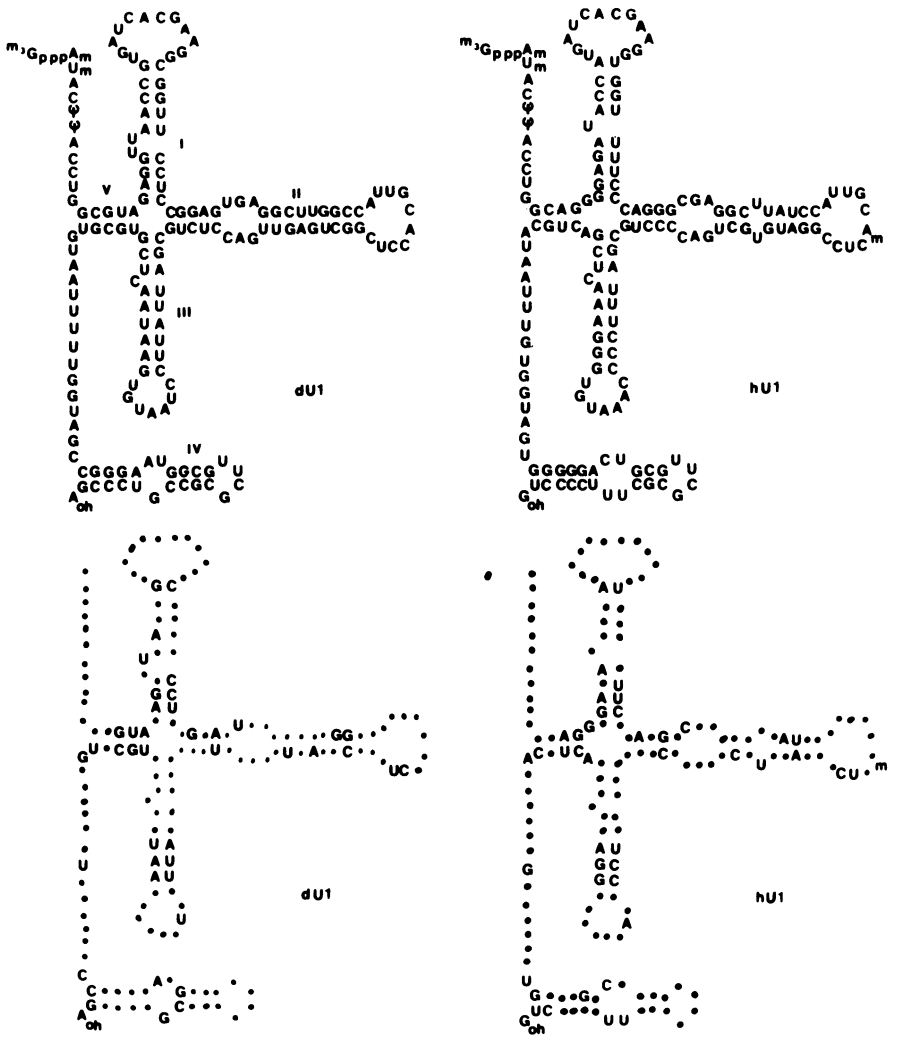

Figure 8. Potential secondary structure of U1 RNAs
Drosophila U1 (dU1) and human U1 (hU1) are illustrated in the same secondary structure. Drawings at the bottom of the figure show only those bases which differ between the two RNAs. Stems are numbered as in Branlant et al. (39).

rich region adjacent to acceptor (3′) splice junctions and the secondary structure we propose for U1 RNAs have implications for the exact role U1 might play in the splicing reaction. Finally, the distribution of conserved nucleotides in U1 RNAs focuses attention on other regions of the molecule

which might be functionally important.

The role of U1 RNA in splicing

In the case of Drosophila, only five donor splice junction sequences have been published, but even these few are sufficient for the consensus sequence $^{C}_{A}AG/GT^{A}_{G}AGT$, derived primarily from vertebrate sequences, to be discerned. (The individual sequences are ATG/GTGCGT, CAG/GTGCGT (34), AAG/GTGAGT (35), AAG/GTAACT and GCG/GTAAGT (36)). Thus, nucleotides 3-11 at the 5′ end of U1 RNA (to which these splice junction sequences are complementary) should be conserved between Drosophila and vertebrates. In fact, the first thirteen nucleotides are unchanged, and the first eleven are in a single-stranded region of the secondary structure model proposed in Fig. 8, making them potentially available for interaction with other RNA molecules.

The original proposals that a small RNA might be involved in splicing all suggested a cross-over model in which the region base pairing with donor junctions was adjacent to a region base pairing with acceptor junctions (7,8,37). In the case of U1 RNA, this model was especially appealing because acceptor splice sites were known to be pyrimidine rich, and the sequence of vertebrate U1 RNA between nucleotides 14-21 is exclusively purines. We can now enumerate four difficulties with the idea that nucleotides 14-21 directly interact with acceptor splice junctions in hnRNA. 1) The pyrimidine rich stretch adjacent to acceptor splice sites has no base per base consensus as donor sites do, but merely consists of a long region (typically 20 or 30 residues) rich in pyrimidines (especially U) and devoid of the dinucleotide AG (38). If this stretch were designed to be recognized by an RNA molecule, one would expect a compilation of large numbers of acceptor splice junctions to reveal some preference for particular nucleotides in particular positions. In fact the reverse is true - all positions in the pyrimidine rich regions of acceptor sites show essentially the same frequency of occurrence of the four bases (S. Mount, in preparation). 2) The pyrimidine rich region of acceptors is frequently much longer than the purine rich region of U1. 3) Only three of the eight nucleotides in U1 positions 14 through 21 are conserved, and the region is not exclusively purines in the Drosophila sequence. Here, it is important to point out that sequenced Drosophila acceptor sites are not particularly complementary to the sequence from 14 to 21 in dU1, but do fit the general acceptor consensus. (These sequences are: CTTTCCATTGCAG/CT, TGTTATCCTGCAG/GC, CTGTCCTGTTCAG/GT (34) AAATCCATTGCAG/AT (35) TATTCAATCCTAG/AA and ATAACACCTTTAG/AA (36)).

4) Finally, the nucleotides between positions 14 and 21 are likely to be involved in secondary structure interactions within the U1 molecule (Fig. 8).

It therefore seems likely that the U1 RNA molecule itself is involved in the recognition of donor splice junctions and perhaps also of that portion of acceptor splice junctions directly at the splice site. However, the question of what might recognize the conserved pyrimidine rich sequence preceding acceptor splice junctions must be reopened. One possibility is that this recognition is handled primarily by the protein portion of the U1 snRNP. Another is that some other snRNP or simply a protein fulfills this function. A biochemical approach is clearly needed to settle this mechanistic question.

U1 RNA secondary structure

The secondary structure we have derived for U1 RNA is very similar to that proposed by Branlant et al. (39) on the basis of nuclease digestion studies of the pure RNA. The similarity between structures derived from such different types of data makes them all the more compelling. The major difference between our structure and theirs is the pairing of nucleotides 12-16 with nucleotides 119-123 to create an additional stem (labeled V in Fig. 8); this requires a rearrangement of the top of stem III. In addition, stems I and II are slightly altered to make them compatible with the dU1 sequence. Note that neither model shows a pairing between nucleotides 6-9 and 133-136 as proposed by Ohshima et al. (40); such an interaction is marginally stable by thermodynamic calculations, but contradicts nuclease digestion data on both snRNPs (41; our unpublished data) and on the RNA itself (39). The overall thermodynamic stability for our model is -56.9 Kcal per mole for dU1 and -62.1 Kcal per mole for hU1, whereas the model of Branlant et al. corresponds to a stabilization energy of -51.8 Kcal per mole for hU1. Although calculated stabilization energies (25) are not always reliable (O. Uhlenbeck, personal communication), these numbers certainly support the secondary structure we present. In any case, it is the basic similarity of the two models, not their differences, which should be stressed.

Conserved sequences in U1 RNAs

The sequence of dU1 from K_C cells agrees with the human U1 sequence in 119 of 165 positions. (In this computation the cap nucleotide was ignored and the ribose methylation of A_{70} in the human sequence was omitted; the one nucleotide insertion at position 23 and the final extra nucleotide of the human sequence were included.) This high degree of homology, 72%, is similar

to the 78% conservation of 5S ribosomal RNA sequences between the same two species (42), suggesting a comparable degree of functional constraint in the two molecules.

As has been found with other conserved RNAs (33), it is the single-stranded regions of the different U1 RNAs which are most highly conserved (see Fig. 8). Interestingly, there are six positions (22,34,36,61,66 and 67) in which the Drosophila, human and rat sequences agree, but the chicken sequence differs (Fig. 2), and five of these changes are of unpaired bases. (Curiously five of them are changes to a C in the chicken sequence). Thus, the long term evolutionary behavior (between vertebrates and invertebrates) of RNA sequences may differ from the short term evolutionary behavior (between mammals and birds) of RNA sequences. An alteration in a single-stranded region, while less acceptable, may nevertheless be more frequent, since mutations in stems usually require two complementary changes.

Inspection of the dU1 sequence also underscores another property of U1 RNA pointed out by Krol et al. (43). It was their observation that U1,U4 and U5 contain a homologous single-stranded region, which can be written $YAAU_nG$, 5′ to a region of base-pairing. Indeed, UAAUUUGUGGUAG occurs in HeLa cell U1, CAAUUUUUGACAG in HeLa cell U4, and UAAUUUUUUGAG in Hela cell U5. The sequence UAAUUUUUGGUAG, occurring at positions 125 through 137 in dU1, fits into this scheme nicely, as the length of this U stretch is similar to those found in human U4 and U5. The possible role of this sequence in U RNA function remains to be elucidated. Possibly meaningful is the the resemblance that these U1 and U4 sequences bear to acceptor splice sites. It should also be remembered that U1, U2, U4, U5 and U6 RNAs bind common proteins (5).

The DNA sequence

Extensive sequences flanking the region specifying U1 RNA on pDmU1.4 are yet to be determined. However, several nucleotides on each side are known. The 5′ flanking sequence, AGGAAAGC, is strikingly similar to the sequence immediately preceding the chicken U1 gene analyzed by Roop et al. (4), AGCAAAGC. This degree of conservation in noncoding DNA is exceptional and tantalizing.

The results of additional sequence work should be revealing, particularly with respect to the still open question of which polymerase synthesizes U1 RNA. Also interesting is the now approachable question of the exact number of U1 RNA genes and their organization in the Drosophila genome.

ACKNOWLEDGEMENTS

We thank Dr. Michael W. Young for providing K_C cells, Dr. Spyros Artavanis-Tsakonas for providing the library of recombinant phage, Martha Krikeles for providing superlative technical expertise, Drs. Eng Tan, Morris Reichlin and John Hardin for providing sera and Joe Hendrick for preparing sera. We are grateful to M.D. Rosa, R. Reed, J. Hendrick, R. Denison and others on the hall for expert advise and criticism. This research was supported by an N.I.H. grant to J.A.S. and a N.S.F. Fellowship to S.M.M.

REFERENCES

1. Weinberg, R. and Penman, S. (1968) J. Mol. Biol. 38, 289-304.
2. Reddy, R., Ro-Choi T.S., Henning, D., and Busch, H. (1974) J. Biol. Chem. 249, 6486-6494.
3. Branlant, C., Krol, A., Ebel, J.P., Lazar, E., Gallinaro, H., Jacob, M., Sri-Widada, J. and Jeanteur, P. (1980) Nucleic Acids Res. 8, 4143-4154.
4. Roop, D.R., Kristo, P., Stumph, W.E., Tsai, M.J. and O'Malley, B.W. (1981) Cell 23, 671-680.
5. Lerner, M.R. and Steitz, J.A. (1979) Proc. Natl. Acad. Sci. USA 76, 5495-5499.
6. Zieve, G. and Penman, S. (1980) J. Mol. Biol. 144, .
7. Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L. and Steitz, J.A. (1980) Nature 283, 220-224.
8. Rogers, J. and Wall, R. (1980) Proc. Natl. Acad. Sci. USA 77, 1877-1879.
9. Yang, V.W., Lerner, M.R., Steitz, J.A., Flint, S.J. (1981) Proc. Natl. Acad. Sci. USA 78, 1371-1375.
10. Tsujimoto, Y. and Suzuki, Y. (1979) Cell 18, 591-600.
11. Gallwitz, D. and Sures, I. (1980) Proc. Natl. Acad. Sci. USA 77, 2546-2550.
12. Sun, S.M., Slightom, J.L. and Hall, T.C. (1981) Nature 289, 37-41.
13. Rubin, G.M. and Hogness, D.S. (1975) Cell, 6 207-213.
14. Garvey, J.S., Cremer, N.E. and Suesdork, D.H. (1977) p. 189-210 in Methods in Immunology, 3rd edition.
15. Lerner, M.R., Boyle, J.A., Hardin, J.A. and Steitz, J.A. (1981) Science 211, 400-402.
16. Lerner, E.A., Lerner, M.R., Janeway, C.A. and Steitz, J.A. (1981) Proc. Natl. Acad. Sci. USA 78, 2737-2741.
17. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
18. Barrell, B.G. (1971) in Procedures in Nucleic Acid Research, eds. Cantoni, G.L. and Davies, D.R. (Harper, New York), Vol. 2, 751-779.
19. Nishimura, S. (1972) Prog. Nucl. Acid Res. Mol. Biol. 12, 49-85.
20. Lockard, R.E., Rieser, L. and Vournakis, J.N. in Gene Amplification and Analysis, Volume II: Analysis of Nucleic Acids by Enzymatic Methods, eds. Chirijian, J.G. and Papas, T.S. (Elsevier-North Holland), 1981.
21. England, T.E. and Uhlenbeck, O.C. (1978) Nature 275, 550-561.
22. Donis-Keller, H., Maxam, A. and Gilbert, W. (1977) Nucl. Acids Res. 4, 2527-2538.
23. Vournakis, J.N., Celentano, J., Finn, M., Lockard, R.E., Mitra, T., Troutt, A., Pavlakis, G., van den Berg, M. and Wurst, B. in Gene Amplification and Analysis, Volume II: Analysis of Nucleic Acids by Enzymatic Methods, eds. Chirijian, J.G. and Papas, T.S. (Elsevier-North Holland), 1981.
24. Queen, C.L. and Korn, L.J. (1980) Methods in Enzymology 65, 595-609.

25. Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biology 246, 40-41.
26. Salser, W. (1977) Cold Spring Harbor Symp. Quant. Biol. 42, 985-1002.
27. Denison, R.A., VanArsdell, S.W., Bernstein, L.B. and Weiner, A.M. (1981) Proc. Natl. Acad. Sci. USA 78, 810-814.
28. Reed, R.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3428-3432.
29. Wise, J.A. and Weiner, A.M. (1980) Cell 22, 109-118.
30. Brunel, C., Sri-Wada, J., Lelay, M.-N., Jeanteur, P. and Liautard, J.-P. (1981) Nucleic Acids Res. 9, 815-830.
31. Hendrick, J.P., Wolin, S.L., Rinke, J., Lerner, M.R. and Steitz, J.A. Molec. Cell Biol., in press.
32. Shibata, H., Ro-Choi, T.S., Reddy, R., Choi, Y.C., Henning, D. and Busch, H. (1975) J. Biol. Chem. 250, 3903-3920.
33. Glotz, C., Zweib, C., Brimacombe, R., Edwards, K. and Kossel, H. (1981) Nucleic Acids Res. 9, 3287-3306.
34. Fyrberg, E.A., Bond, B.J., Hershey, N.D., Mixter, K.S. and Davidson, N. (1981) Cell 24, 107-116.
35. Holmgren, R., Corces, V., Morimoto, R., Blackman, R. and Meselson, M. (1981) Proc. Natl. Acad. Sci. USA 78, 3775-3778.
36. Benyajati, C., Place, A.R., Powers, D.A. and Sofer, W. (1981) Proc. Natl. Acad. Sci. USA 78, 2717-2721.
37. Murray, V. and Holliday, R. FEBS Lett. 106, 5-7 (1979).
38. Seif, I., Khoury, G. and Dhar, R. (1979) Nucl. Acids Res. 6, 3387-3389.
39. Branlant, C., Krol, A., Ebel, J.-P., Gallinaro, H., Lazar, E. and Jacob, M. (1981) Nucleic Acids Res. 9, 841-858.
40. Ohshima, Y., Izoh, M., Okada, N. and Miyata, T. (1981) Proc. Natl. Acad. Sci. USA 78, 4471-4474.
41. Epstein, P., Reddy, R. and Busch, H. (1981) Proc. Natl. Acad. Sci. USA 78, 1562-1566.
42. Erdmann, V.A. (1978) Nucleic Acids Res. Special Supplement.
43. Krol, A., Branlant, C., Lazar, E., Gallinaro, H. and Jacob, M. (1981) Nucleic Acids Res. 9, 2699-2716.