A new U2 RNA secondary structure provided by phylogenetic analysis of trypanosomatid U2 RNAs

Toinette Hartshorne and Nina Agabian

Intercampus Program in Molecular Parasitology, School of Pharmacy, University of California, San Francisco, California 94143-1204 USA

A new model of U2 RNA secondary structure has been developed by comparing U2 RNA sequences from distantly related trypanosomatids, which process their RNAs by trans-splicing, and U2 RNAs from cis-splicing organisms. The trypanosomatid U2 RNA retains structural similarity in important functional domains of cis-splicing U2 RNAs yet differs from previous consensus models in that only two helices, rather than three, can form in the stem–loop II region. This alteration eliminates the capacity for pseudoknot formation and produces a single-stranded region 3' to stem–loop II, which may be accessible for snRNP protein binding.

Trypanosomatid U2 RNAs differ in the putative branchpoint recognition sequences, which completely diverge from the conserved GUAGUA consensus of cis-splicing organisms.

[Key Words: U2 RNA; secondary structure; trypanosomatids]

Received August 13, 1990; revised version accepted October 9, 1990.

The maturation of nuclear messenger RNAs (mRNAs) by trans splicing occurs in several kinetoplastid protozoan and nematode species (Laird 1989; Nilsen 1989) by a process that is related to the removal of introns from pre-mRNAs by cis splicing. Trans splicing was first described for the African parasite Trypanosoma brucei, in which all mRNA transcripts receive at their 5' ends an identical 39-nucleotide spliced leader (SL) sequence (Parsons et al. 1984) that possesses a unique “cap 4’” structure (Ferry et al. 1987). This SL is donated by a 139-nucleotide precursor transcript, the SL RNA (for review, see Laird 1989). The SL RNA and acceptor pre-mRNAs contain 5'- and 3'-splice site consensus sequences, respectively. During trans splicing, a branched Y-structure intermediate, analogous to the lariat structure of cis-spliced introns [Konarska and Sharp 1988; Black and Pinto 1989], is formed by these two RNA substrates (Murphy et al. 1986, Sutton and Boothroyd 1986).

In cis-splicing, trans-acting U small nuclear ribonucleoprotein (snRNP) particles and auxiliary factors assemble together with pre-mRNAs in a spliceosomal complex where splicing occurs (for review, see Guthrie and Patterson 1988; Steitz et al. 1988). Trypanosomatids possess only a subset of the major snRNPs that are found in cis spliceosomes. T. brucei homologs of U2, U4, and U6 RNAs have been characterized (Tschudi et al. 1986, 1988; Mottram et al. 1989) and shown to be required in trans splicing by site-directed degradation of each in permeabilized T. brucei cells [Tschudi and Ullu 1990]. By analogy with cis-splicing systems, the trypanosomatid U2 snRNP may be expected to interact with branchpoint sequences. Likewise, the striking evolutionary conservation of the trypanosomatid U4/U6 snRNP with those of other organisms [Mottram et al. 1989] argues for a central and analogous role of this complex in both cis and trans splicing. Whereas the U4/U6 snRNP appears to associate with the U5 snRNP near the 3' splice site of cis introns (Konarska and Sharp 1988; Black and Pinto 1989), assembly of the trypanosomatid U4/U6 snRNP with pre-mRNAs seems to occur in the absence of a U5 snRNP equivalent. The requirement for a U1 snRNP, which interacts by intermolecular base-pairing with the 5'-splice site, may have been supplanted by the SL RNA in trypanosomatid trans splicing. The SL RNA can form a stem–loop structure within which the 5'-splice site sequences are base-paired with upstream SL [exon] sequences [Bruzik et al. 1988]. Furthermore, the SL RNA is found as an RNP particle in vivo in T. brucei [Michaeli et al. 1990], indicating that it not only functions as an exon donor, but also as a component of the splicing machinery.

The U2 snRNP has pivotal roles in spliceosome formation and branchpoint selection in cis-splicing systems. A consensus U2 secondary structure model from all known U2 RNA sequences was proposed in which the U2 RNA folds into four stem–loop domains with two single-stranded regions [Guthrie and Patterson 1988; Ares and Igel 1989]. The T. brucei U2 RNA conforms reasonably well to this model [Guthrie and Patterson 1988; Ares and Igel 1989], implying similarities in function. However, it diverges from other U2 RNAs in three major regions: (1) The putative branchpoint recognition region does not contain the GUAGUA sequence,
which is perfectly conserved in all other U2 RNAs (Guthrie and Patterson 1988) and which base-pairs with branchpoint sequences in yeast and mammals (Parker et al. 1987; Wu and Manley 1989; Zhuang and Weiner 1989); [2] the single-stranded region 3′ to stem–loop II that contains the conserved Sm antigen-binding site required for 2,2,7-trimethylguanosine cap formation and nuclear accumulation of U2 snRNPs in Xenopus (Mattaj 1988) is absent; [3] stem–loop III is also missing, thus producing a structure in which stem–loop II adjoins stem–loop IV.

In the present study, the structure of trypanosomatid U2 RNAs has been examined further by comparative analysis (Noller 1984). U2 RNA genes from distantly related trypanosomatids were isolated, sequenced, and compared. These were the insect parasite, Leptomonas collosoma, which diverged from mammalian-infective trypanosomes ~260 million years ago, and the South American stercorarian trypanosome, Trypanosoma cruzi, which diverged from the African salivarian trypanosome T. brucei ~80 million years ago (Lake et al. 1988). Comparisons between trypanosomatid and other known U2 RNA sequences provided phylogenetic evidence for a new U2 RNA secondary structure that differs from the model proposed previously (Ares and Igel 1989) in that only two helices, in contrast to three, can form in the region of stem–loop II.

Results

L. collosoma and T. cruzi U2 RNA genes are multicopy

Although U2 RNA is encoded by multiple genes in most eukaryotes, the U2 RNA gene of T. brucei is found in single copy (Tschudi et al. 1986; Mottram et al. 1989). The copy number of U2 RNA genes in L. collosoma and T. cruzi was determined by Southern analysis of the genomic DNA of each organism using as probe an Sp6-de¬
rived RNA transcript containing sequences complementary to T. brucei U2 RNA (cTbU2) (Fig. 1). In contrast to T. brucei, three classes of U2 RNA genes were distinguished in both L. collosoma and T. cruzi, and members of each class were isolated from genomic libraries (see Materials and methods). Restriction endonuclease digestion of L. collosoma genomic DNA with enzymes recognizing 4 bp produced two Taql fragments of ~690 and 490 bp, and two HpaII fragments of ~350 and 650 bp that hybridized with the cTbU2 probe. The 690-bp Taql and 350-bp HpaII fragment signals were more intense than the others [Fig. 1A]. Six L. collosoma U2 RNA genes isolated fell into three classes, with two members each, as characterized by Southern analysis: Class I U2 RNA genes have a 690-bp Taql and a 350-bp HpaII fragment, class II genes have a 490-bp Taql and a 350-bp HpaII fragment, and class III genes have a 690-bp Taql and a 650-bp HpaII fragment [Fig. 1A]. Digestion of genomic L. collosoma DNAs using several restriction enzymes that recognize a 6-bp motif [Fig. 1B and data not shown], produced only single cTbU2-hybridizing bands ranging in size from 6 to 15 kb, suggesting that the multiple U2 RNA genes might be closely associated. However, since each class was isolated independently on recombinant phage containing inserts of ~16 kb in size, it is more likely that the U2 genes are located in similar but remote genomic contexts. Consistent with the latter interpretation, a U2 RNA gene of the class I variety was isolated from a size-selected genomic library containing 7- to 10-kb HindIII fragments [data not shown] corresponding to a single 8-kb HindIII band detected by Southern analysis [Fig. 1B].

T. cruzi U2 RNA genes were identified and isolated in a similar fashion. Digestion of T. cruzi genomic DNA with Taql yielded three fragments of ~510, 380, and 270 bp in size, while HpaII digestions showed two fragments of 200 and 360 bp that hybridized with the cTbU2 probe [Fig. 1A]. Eleven independent recombinants isolated from a T. cruzi genomic library fell into three classes by Southern analysis: Class I consisted of six isolates with 380-bp Taql and 360-bp HpaII fragments, class II consisted of four isolates with 510-bp Taql and 360-bp HpaII fragments, and class III had only one member with 270-bp Taql and 200-bp HpaII fragments [Fig. 1A]. Southern analysis of T. cruzi genomic DNAs digested with BamHI and EcoRI produced three fragments that hybridized with cTbU2, while only two fragments were detected in HindIII and SalI digests and five fragments were detected in PsiI digests [Fig. 1B]; various double digestions showed two to five bands. The three EcoRI and three of the PsiI fragments were assigned to one member of each class of isolated genes; class I and class II both contained the 5.2-kb SalI fragment and class III contained the 3.6-kb SalI fragment [data not shown]. Thus, the three classes of T. cruzi U2 RNA genes identified and isolated appear to be representative of all possible gene classes by restriction fragment analysis, yet the complexity of the cTbU2 Southern hybridization pattern suggests that additional copies of these U2 RNA genes may exist.

DNA sequence of L. collosoma and T. cruzi U2 RNA genes

Taql and HpaII fragments from one isolate of each class of L. collosoma and T. cruzi U2 RNA genes were cloned for DNA sequence analysis. The sequencing strategy and deduced restriction map for each U2 RNA gene is shown in Figure 2; the location of the encoded U2 RNA gene sequences within the fragments is indicated. Figure 3 shows the DNA sequences of each L. collosoma and T. cruzi U2 RNA gene.

Comparisons between the classes of L. collosoma U2 RNA genes reveal that each encodes an identical U2 RNA sequence of ~152 nucleotides, as estimated from Northern analysis and by alignment with the T. brucei U2 RNA sequence [see below]. The three classes are distinguished by the sequences of their flanking regions, which diverge at 11 of 124 nucleotide positions in the 5′ region and at 8 of 63 nucleotides in the 3′-noncoding sequences. In contrast, each of the three T. cruzi U2 RNA genes encodes distinct U2 RNA sequences of ~148 nucleotides in length. Class I and class II U2 RNA genes (abbreviated U2I and U2II) are closely related, diverging only at positions 36 and 39 within the coding sequence.
Figure 1. Southern analysis of trypanosomatid U2 RNA sequences. (A) Fifteen micrograms of genomic DNA [GEN] from T. cruzi [lanes 1 and 2] and L. collosoma [lanes 9 and 10] and 15 ng of recombinant λ DNAs containing T. cruzi [lanes 3–8] and L. collosoma [lanes 11–16] U2 RNA genes were digested with TaqI [T, odd-numbered lanes] or HpaII [H, even-numbered lanes], separated on a composite 0.8% agarose/3.5% acrylamide gel, transferred to Nytran membrane and probed with cTbU2. T. cruzi or L. collosoma U2I clones are in lanes 3 and 4 and 12 and 13, U2II in lanes 5 and 6 and 14 and 15, and U2III in lanes 7 and 8 and 15 and 16, respectively. Molecular markers [given in bp], are from pBR322 plasmid DNA digested with MspI. The highest molecular weight fragment in lane 9 was not seen in other experiments and is due to partial digestion of genomic DNA. (B) Six micrograms of genomic DNA from T. cruzi [lanes 1–10] and L. collosoma [lanes 11–13] was digested with restriction enzymes, separated on a 0.8% agarose gel, transferred to Nytran membranes, and probed with cTbU2. Enzymes used are BamHI [B], HindIII [H], EcoRI [E], PstI [P], and SalI [S]. Molecular weight markers [given in kb] are from λ DNA digested with HindIII and EcoRI.

The regions flanking U2I and U2II are also similar, deviating at only five positions in 242 nucleotides of 5’ sequence and at three positions in 285 nucleotides of 3’ sequence. The class III U2 RNA gene [U2III] is considerably different from those of the other classes, varying within the coding sequence at eight positions and six positions from U2I and U2II, respectively. The flanking regions of U2III also deviate significantly from those of the other two classes; 10 and 11 of 65 nucleotides 5’ to the coding sequence and 32 and 33 of 77 nucleotides of 3’-noncoding sequence differ from U2I and U2II RNA genes, respectively. The 5’-flanking region from -65 to -90 nucleotides is completely divergent from U2I and U2II. The sequence differences in the flanking regions raise the possibility that U2III RNA gene expression might be regulated in a different manner than that of the U2I and U2II RNA genes.

T. cruzi and L. collosoma U2 RNA gene expression

Only a single-sized U2 RNA species is revealed by Northern analysis in each organism [Fig. 4]; this is as expected for L. collosoma, where all of the U2 genes encode an identical RNA species. The site of transcription initiation was determined by primer extension sequencing of the U2 RNAs of each organism; this analysis additionally ensured that the sequenced U2 RNA genes corresponded with those that were expressed. The first residue of L. collosoma and T. cruzi U2 RNAs was mapped to an A residue in the DNA sequence (position +1 in Figs. 2 and 3; data not shown). Residues 2–130 of the L. collosoma U2 RNA sequence were identical to those predicted from DNA sequence analysis (data not shown). T. cruzi epimastigote U2 RNAs were sequenced using two oligonucleotide primers complementary to all three classes of U2 RNA genes [see Materials and methods]. Residues 2–65 were obtained from the TcU2-M primer (Fig. 5) and residues 2–123 were obtained from the TcU2-3’ primer [data not shown]. Although there are at least three different U2 RNA gene classes in T. cruzi, the U2 RNA population yielded only a single class of sequence corresponding to that of the U2III RNA gene. Seven of the eight positions at which the U2 RNA genes differ were examined in these experiments, and expression of neither U2I nor U2II RNA genes was detected. Thus, U2I and U2II do not appear to be expressed in T.
Hartshorne and Agabian

A new U2 RNA secondary structure

were examined for conserved base-paired regions, using this structure was generated by the method of comparing C. bean sequences in the same region. The T. cruzi, L. collosoma, and T. cruzi U2 RNA sequences are shown aligned in Figure 6, together with their consensus secondary structure. The compiled sequences differ by 47-base differences and insertions and retain 68% identity. The greatest degree of similarity (92%) is found within the 5' 66 residues of the U2 RNAs, the sequences becoming increasingly divergent thereafter. This region of the trypanosomatid U2 RNAs shares 68% identity with the analogous regions of mammalian, bean, and fly sequences and slightly less similarity with C. elegans and S. ceevisiae U2 RNAs. For comparison, mammalian U2 RNAs share 80% identity with C. elegans and S. cerevisiae U2 RNAs and 95% with fly and bean sequences in the same region.

A new consensus secondary structure for trypanosomatid U2 RNAs is shown in Figure 7 [see also Fig. 6]. This structure was generated by the method of comparative analysis [Noller 1984]; trypanosomatid sequences were examined for conserved base-paired regions, using computer-generated secondary structures and the consensus U2 RNA model [Guthrie and Patterson 1988; Ares and Igel 1989] as guidelines. Although not presented in Figure 7, the RNA sequences deduced from T. cruzi U2I and U2II RNA genes also conform to this structure. In the previous consensus U2 RNA model, metazoan U2 RNAs are predicted to fold into four stem-loop domains wherein stem-loops I and II and stem-loops II and III are separated by single-stranded regions. The previously proposed T. brucei U2 RNA structure conforms to this model except that the third stem-loop and adjacent single-stranded region are deleted, producing a structure in which stem-loop II abuts stem-loop IV [Guthrie and Patterson 1988; Shuster and Guthrie 1988; Ares and Igel 1989]. The newly proposed trypanosomatid U2 RNA secondary structure model alters the context of the stem-loop II region, only two helices can form in this region in all trypanosomatid U2 RNAs in contrast to the three helices predicted previously. In addition, a single-stranded region now forms between stem-loops II and IV, analogous to the core protein-binding region of all other U2 RNAs. When the modification is applied to the yeast and metazoan U2 RNAs, all U2 RNAs become very similar in structure, except that the stem-loop III region varies in yeast and is missing in trypanosomes.

In the trypanosomatid U2 RNA model, the first 6 residues are single-stranded and the next 20 residues form stem-loop I. The stem residues are identical in all organisms except for the 14/19 base pair adjacent to the loop, the existence of which is supported by the presence of U/A base pairs in trypanosomatids and C. elegans, in place of the C/G base pairs found otherwise. Interestingly, this pairing is disrupted in T. cruzi U2I and U2II, where residue 14 is a C instead of a U and cannot form a Watson–Crick base pair with A19.

Following stem-loop I, single-stranded region I extends from position 27 to 46. Although this is the most highly conserved portion of metazoan and yeast U2 RNAs, 8 of these 20 positions diverge between trypanosomatid U2 RNAs and other eukaryotes; 3 of these positions also vary between trypanosomatid U2 RNAs (see Fig. 6). In yeast and metazoans, this region contains the perfectly conserved GUAGUA sequence, which is involved in branchpoint recognition, forming Watson–Crick base pairs with conserved intron sequences [Parker et al. 1987; Wu and Manley 1989, Zhuang and Weiner 1989]. The GUAGUA sequence diverges among trypanosomatid U2 RNAs to UAUUAA in T. brucei, UUUUAA in L. collosoma and T. cruzi [UUUCAA in T. cruzi U2I], and also UAAUAA in Leishmania enriettii [Miller and Wirth 1988]. Thus, the GUAGUA sequence is not only missing in trypanosomatid U2 RNAs, but the trypanosomatid sequences in this region are also not conserved.

The structure of the stem-loop II region of trypanosomatid U2 RNAs is considerably different in the current and former models of U2 RNA secondary structure. Previously, it had been suggested that the highly conserved region between residues 47 and 95 in human U2 could potentially form three helices; simultaneous formation of all three helices would result in a pseudoknot struc-
Figure 3. DNA Sequences of *L. collosoma* (A) and *T. cruzi* (B) U2 RNA genes. The noncoding strand of DNA is shown. The position of the first nucleotide of the U2 RNA-coding sequence is denoted as +1. The putative 3' ends of the coding sequences are approximately +153 for *L. collosoma* and +148 for *T. cruzi* U2 RNAs. Dashes in U2I and U2II sequences indicate identity with the U2I sequence. Dots indicate a gap in the alignment of sequences.
Northern analysis of trypanosomatid U2 RNAs. Ten micrograms of total RNA from T. brucei, T. cruzi, and L. collosoma were loaded in each lane of a 6% polyacrylamide/8 M urea gel, transferred to Nytran membranes, and hybridized with a \(^{32}\)P-labeled cTbU2 SP6-transcript probe. The faint band of lower molecular weight in the T. cruzi lane is presumably due to degradation of the U2 RNA, since it accumulated in the RNA preparation with time.

Fig. 4

Our revised trypanosomatid U2 RNA secondary structure model alters the context of residues 81–103 (T. brucei numbering), placing these in a single-stranded domain that was double-stranded in the former model. This region is variable between trypanosomatid U2 RNAs (only 11 of 26 residues are conserved), maintaining a character that is purine-rich in the first half and pyrimidine-rich in the second half. By analogy to metazoan and yeast U2 snRNPs, this single-stranded region may be available for binding of Sm-like antigens (for review, see Mattaj 1988). However, only sequences...
Trypanosomatid U2 RNAs

|       | I       | I'       |single-stranded |
|-------|---------|---------|---------------|
| T. brucei | 1 UUCUC UCUCGGCU AUUU AGCUUAGA UCAAGUUUAACUGUCUCU 46 |
| T. cruzi | 1 UUCUC UCUCGGCU AUUU AGCUUAGA UCAAGUUUAACUGUCUCU 46 |
| L. collosoma | 1 UUCUC UCUCGGCU AUUU AGCUUAGA UCAAGUUUAACUGUCUCU 46 |
| CONSENSUS | AUAUC UCUCGGCU AUUU AGCUUAGA UCAAGUUUAACUGUCUCU |

Mammalian U2 RNA sequences are aligned with a consensus sequence. The T. brucei U2 RNA sequence has been taken from Tschudi and Ullu [1986]. Only the expressed form of the T. cruzi U2 RNA, U2III RNA, is represented. The dots within trypanosomatid sequences indicate gaps in the alignment. In the consensus sequence, uppercase letters indicate identity between all three sequences, lowercase letters indicate identity between two sequences, and a dot indicates a difference among all three sequences. The secondary structure of the U2 RNA is indicated above the sequences [see Fig. 7]. Stem–loop structures are denoted by dashed lines with gaps where mismatched bulges occur. The five dots over the region 104–143 suggest the potential to base-pair with their counterpart SL RNAs [Milhausen et al. 1984]. Whereas 10 contiguous residues are complementary between T. brucei RNAs and L. collosoma RNAs in T. cruzi, the very divergent AACCGCUAGA in L. collosoma, and the very divergent AACCGCUAGA in T. cruzi.

As recognized previously, an equivalent of the metazoan U2 RNA stem–loop III is missing entirely in T. brucei U2 RNA [Guthrie and Patterson 1988; Ares and Igel 1989] as it is in L. collosoma and T. cruzi U2 RNAs. The 3' end of all U2 RNAs forms a stem–loop IV structure that varies in primary sequence [Guthrie and Patterson 1988; Ares and Igel 1989]. The trypanosomatid stem–loop IV structure, which is formed by residues 105–143 in T. brucei, is well supported by multiple compensatory differences between species [see Fig. 7].

Phylogenetic comparison upholds the potential for an SL–U2 RNA interaction

For trans-RNA splicing in trypanosomatids to occur, the SL RNA must somehow recognize and interact with the acceptor RNA, perhaps through association with the U2 snRNP. It has been suggested that the U2 and SL RNAs might interact via RNA base-pairing since the T. brucei RNAs show regions of sequence complementarity [Tschudi et al. 1986; Mottram et al. 1989]. The U2 RNA sequences from L. collosoma and T. cruzi were analyzed for the potential to base-pair with their counterpart SL RNAs [Milhausen et al. 1984]. Whereas 10 contiguous residues are complementary between T. brucei RNAs and L. collosoma RNAs in T. cruzi, the very divergent AACCGCUAGA in L. collosoma, and the very divergent AACCGCUAGA in T. cruzi.

Discussion

The new secondary structure proposed for T. brucei U2 RNA was derived through a combination of phylogenetic analysis of trypanosomatid U2 RNA sequences, their comparison with metazoan and yeast U2 RNA sequences, and computer-assisted RNA folding methods. The secondary structure model has been modified to exclude the possi-
Hartshorne and Agabian

Figure 7. Secondary structure of trypanosomatid U2 RNAs. The T. brucei U2 RNA sequence (Tschudi et al. 1986) is shown. Differences between U2 RNAs are shown in shaded boxes for T. cruzi U2III RNA and unshaded boxes for L. collosoma U2 RNA; deletions are denoted as a circle within the box; insertions are denoted by lines extending from boxed residues to their insertion positions. Stem-loops II.2 and IV of T. cruzi and L. collosoma U2 RNAs are redrawn for clarity beside the structure. The stem-loop IV bulge residue is a G in the T. brucei sequence reported by Mottram et al. (1989). Putative branchpoint recognition and core protein-binding regions are indicated in single-stranded regions I and II, respectively. The final trypanosomatid U2 RNA structures have ΔG° values, estimated from Zuker and Steigler parameters (Zuker and Steigler 1981), of ~40.7 kcal/mole for T. brucei, ~41.0 for T. cruzi, and ~40.3 for L. collosoma; additional energetic stability contributed by the CUUCGG loop (Tuerk et al. 1988) of II.2 of L. collosoma U2 RNA was not included in the calculation.

bility of a pseudoknot arrangement or alternate hairpin conformations within the stem–loop II region, since only two of the three possible helices predicted to form in other U2 RNAs (Ares and Igel 1989) can form in the trypanosomatid U2 RNAs. We propose that this new model is applicable to the U2 RNAs of all organisms, whether these utilize cis- or trans-RNA splicing or a combination of both. This proposal is supported by data obtained by M. Ares, who has arrived independently at the same model by in vitro chemical probing and in vivo mutational analysis of the yeast U2 RNA [M. Ares, pers. comm.]. This conservation of U2 RNA structure argues for central and analogous roles of the U2 snRNP in trans- and cis-splicing processes, such as branch or lariat formation, splice site juxtaposition, and interactions with other conserved factors, i.e., the U4/U6 snRNPs. It is possible that the significant differences in primary sequence that remain between trypanosomatid U2 RNAs and those of cis-splicing organisms, as well as disparities in other components of their U2 snRNPs, impart properties required to mediate distinctive aspects of trypanosomatid trans splicing.

The proposed U2 RNA secondary structure results in expansion of the single-stranded region downstream of stem–loop II to include residues formerly associated with helical structure. This region contains the consensus sequence for binding of the core Sm antigens in metazoan and yeast U2 RNAs [Mattaj 1988]. In known metazoan U2 RNA sequences [from Guthrie and Patterson 1988], the region between stem–loops II and III increases in length from 16 to 27 residues. In trypanosomatids, the change is even more dramatic, expanding the analogous single-stranded region from 0 to 8 residues in earlier T. brucei U2 RNA models (Guthrie and Patterson 1988; Shuster and Guthrie 1988; Ares and Igel 1989) to 24–25 residues. This alteration creates a potential domain in trypanosomatid U2 RNAs for interaction with core proteins, which is consistent with the finding that this region is protected from oligonucleotide probing in the T. brucei U2 snRNA yet is accessible for probing in the deproteinated U2 RNA (J. Dungan, unpubl.). The sequences with the closest match to the consensus RAU_{2-4}GR Sm antigen-binding site [Mattaj 1988] in trypanosomatid U2 RNAs have the pattern AAYY(G/U)YUG(A/C). A related sequence, AAGUUUGC, is found in the counterpart region of the T. brucei U4 RNA [Mottram et al. 1989]; also, the SL RNAs from various trypanosomatids contain sequences that resemble this Sm-like consensus (Bruzik et al. 1988). An examination of these trypanosomatid U and SL sequences does not reveal a clear motif for protein recognition, moreover, the trypanosomatid sequence deviations exceed the latitude in Sm antigen recognition disclosed by mutational analysis of the Sm site in yeast [Jones and Guthrie 1990]. However, it is possible that the trypanosomatid SL and U RNP particles contain common core proteins that have a less stringent requirement for binding than the Sm antigens of yeast and metazoan U RNAs.
By the new model, all U2 RNAs are similar in secondary structure within the 5' 100–120 residues that include stem–loops I and II and the single-stranded regions that contain residues involved in branchpoint recognition and core protein binding. Downstream regions diverge between species and are apparently unnecessary in cis splicing, as shown by deletion analysis in yeast and *Xenopus* (Igel and Ares 1988; Shuster and Guthrie 1988; Hamm et al. 1989; McPheeters et al. 1989; Pan and Prives 1989). Given the similarities in the new structures, the functions of the trypanosomatid U2 RNA can now be assessed by comparison to other U2 RNAs. The 5' 15 nucleotides and a stretch of residues containing the branchpoint recognition region of human and *Xenopus* U2 RNAs have been implicated in pre-scarification and conversion to a functional spliceosome complex [Frendewey et al. 1987; Steitz et al. 1988; Barabino et al. 1989]. These 5' 15 nucleotides are well conserved between trypanosomatid and metazoan U2 RNAs; their destruction abolishes trans-splicing activity [Tschaudt and Ullu 1990], suggesting that they are involved in pre-scarification and conversion, perhaps at the level of 3'-splice site recognition, as indicated for metazoans (Steitz et al. 1988).

The branchpoint recognition region of U2 RNAs from trypanosomatids and cis-splicing organisms may be expected to function in analogous manners. However, this region, which contains the invariant GUAGUA sequence in all metazoan and yeast U2 RNAs [Guthrie and Patterson 1988], not only diverges in a single trypanosomatid U2 RNA but also between trypanosomatid U2 RNAs, having the consensus U(U/A)(U/A)UA. In yeast, branchpoint recognition depends on complementarity between the GUAGUA sequence of U2 RNA and the conserved UACUAAC sequence of introns [Guthrie and Patterson 1988]. The branchpoint sequence is poorly conserved in mammalian systems; its recognition requires a number of protein factors [Guthrie and Patterson 1988], which interact with conserved intron sequences between the branchpoint and the 3'-splice site. In *T. brucei*, the branchpoint sequences of α- and β-tubulin genes have been mapped in vivo. They are A residues but lack homology to the consensus branchpoint sequences of yeast and mammalian cells and do not even resemble each other [Patztelt et al. 1989]. These intron sequences do not share recognizable sequence complementarity with the putative branchpoint recognition region of the U2 RNA. Thus, it is probable that branchpoint selection by the U2 snRNP in trypanosomes will be facilitated by protein factors that recognize 3'-splice site sequences, similarly to mammalian systems. That the GUAGUA sequence is perfectly conserved among all eukaryotes examined except trypanosomatids is curious; perhaps the GUAGUA sequence is absolutely required for cis-splicing-related functions and is unimportant for trans splicing. This argument is supported by the heterogeneity in this region among trypanosomatid U2 RNAs. Alternatively, the less conserved trypanosomatid sequence might have diverged for trans-splicing-specific functions. It should also be noted that the flanking regions of the metazoan and yeast U2 RNA GUAGUA sequence are well conserved in all U2 RNAs, including trypanosomatids [this paper; Guthrie and Patterson 1988]. The possible functions of this region have not been analyzed specifically, but it could possibly aid branchpoint selection and utilization or be available for interaction with other U RNAs.

**Materials and methods**

**Southern and Northern analysis**

Genomic DNA from *L. collosoma* was isolated as described previously for *T. brucei* [Mottram et al. 1989]. *T. cruzi* (Y strain) genomic DNA was kindly donated by J.L. Rosales-Encina and V. Nussenzweig. Total RNA was acquired from epimastigote stage *T. cruzi* (Y strain) cells (also a gift of J.L. Rosales-Encina and V. Nussenzweig) and *L. collosoma* cells by the guani­dinium–hot phenol RNA method as described by Maniatis et al. [1982]. Nytran membranes, prepared by Southern or Northern methods [details in figure legends; Maniatis et al. 1982], were hybridized with ^32P-labeled antisense, *T. brucei* U2 RNAs (cTbU2) transcribed from EcoR1-linearized pG-U2C [Michaels et al. 1990]. For detection of heterologous U2 genes and RNAs, hybridizations were carried out routinely in a solution of 5x SSPE, 5x Denhardt’s solution, 0.1% SDS, 50 µg/ml E. coli tRNA, and 10^² cpm of probe/ml at 50°C for 16 hr, followed by four 20-min washes in 2x SSPE, 0.1% SDS, at 55°C.

**Screening of genomic DNA libraries**

A genomic library of *L. collosoma* DNA was constructed by partial digestion of genomic DNA with *Sau3A* enzyme, size selection of 9–23 kb fragments on 0.7% preparative agarose gel, recovery of DNA by electrophoresis, and ligation of fragments into BamHI-digested bacteriophage EMBL3 arms. Phage were packaged in vitro using BRL Gigapack Gold extracts and grown on *E. coli* Q358. A *T. cruzi* (Y strain) genomic library was kindly provided by A. Gonzalez. Recombinant λ DNAs were transferred to nitrocellulose membranes [Maniatis et al. 1982], and clones containing U2 sequences were identified by hybridization with cTbU2 as described for Southern analysis. U2-hybridizing plaques occurred at a frequency of ~1/5000 in *L. collosoma* and *T. cruzi* genomic libraries.

**DNA subcloning and sequencing**

*Taql* and *Hpall* DNA fragments from each class of recombinant bacteriophage were subcloned into M13 replicative form DNA by standard methods. Both ^32P-labeled cTbU2 transcript and TbU2-5' oligonucleotide probes were used to identify recombinant phage-containing U2 sequences. Libraries were probed as described for Southern analysis except that when the TbU2-5' oligonucleotide probe was used, hybridizations and washes were done at 37°C. DNA sequences were determined by the dideoxy chain termination method [Sanger et al. 1977]. Both universal sequencing primers and custom-made primers (see below) were used. IntelliGenetics programs were useful for compiling and analyzing sequence data.

**AMV reverse transcriptase sequencing of RNAs**

The sequence of expressed *T. cruzi* and *L. collosoma* U2 RNAs were derived by primer extension sequencing with avian mye­loblastosis virus (AMV) reverse transcriptase. Reactions were...
done as described previously (Patzelt et al. 1989), with the following exceptions. Oligonucleotide primers were gel-purified from 20% polyacrylamide/8 M urea gels following labeling with [γ-32P]ATP. Eight micrograms of DNAse I-treated T. cruzi or L. collosoma RNA was hybridized with 5 pmoles (≈10^6 cpm) of oligonucleotide primers TcU2-M or LeU2-3', respectively, for use in sequencing reactions. For sequence analysis of T. cruzi U2 RNA using the TeU2-3' oligonucleotide primer, size-fractionated RNA was used because multiple RNAs present in total RNA preparations apparently hybridized to TeU2-3'. One hundred fifty micrograms of T. cruzi RNA was electrophoresed on a 6% polyacrylamide/8 M urea gel, and RNAs of 140-160 nucleotides in size were excised, eluted from the gel, and ethanol-purified (Maniatis et al. 1982). One-half of the recovered material was hybridized with 5 pmoles of TeU2-3' primer per sequence analysis.

DNA oligonucleotide primers used in DNA and RNA sequence analysis and U2 subclone selection

The following oligonucleotides, complementary to the indicated regions of T. brucei (Tb), T. cruzi (Tc), and L. collosoma (Lc) U2 RNAs, were used in RNA and/or DNA sequence analysis.

TeU2-M 5'-TCCCTGGACAGCAG-3' nucleotides 83-69 of Tc U2 RNA
TeU2-3' 5'-AACGCTCCATC-3' nucleotides 144-130 of Tc U2 RNA
LeU2-3' 5'-CCAGCTCTCGGCTG-3' nucleotides 147-134 of Le U2 RNA

The oligonucleotide TbU2-5', the DNA copy of the first 14 nucleotides of U2 RNA, was used in hybridizations to U2-containing M13 libraries of T. cruzi and L. collosoma DNAs and in DNA sequence analysis.

TMD-5' 5'-ATATCCTTCTCGGCT-3'

Secondary structure modeling

The Zuker RNA FOLD program utilizing the parameters of Zuker and Steigler (Zuker and Steigler 1981) and Salser (Salser 1977) was used to determine energetically optimal RNA secondary structures of U2 RNA sequences under study. Phylogenetic comparisons of the derived structures enabled deduction of a consensus U2 secondary structure model.

Acknowledgments

We are grateful to J.L. Rosales-Encina and V. Nussenzweig for their gifts of T. cruzi cells and genomic DNA and to A. Gonzalez for his T. cruzi cDNA library used in this study. We thank M. Ares and J. Dungan for sharing their results prior to publication and to D. Cuss and C. Cass for technical assistance. We also thank members of the Agabian laboratory, K. Perry McNally, and G. Newport for their comments on the manuscript. This work was supported by grants to N.A. from the John D. and Catherine T. MacArthur Foundation, and from National Institutes of Health grant AI21975. N.A. is a Burroughs Wellcome Scholar of Molecular Parasitology.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

References

Ares, M., Jr. and A.H. Igel. 1989. Phylogenetic comparison of U2 small nuclear RNA sequences suggests a pseudoknotted structure. UCLA Symp. Mol. Cell. Biol. 94: 13-23.

Barabino, S.M., B.S. Sproat, U. Ryder, J. Blencowe, and A.I. Lamon. 1989. Mapping U2 snRNP-pre-mRNA interactions using biotinylated oligonucleotides made of 2'-OMe RNA. EMBO J. 8: 4171-4178.

Black, D.L. and A.L. Pinto. 1989. U5 small nuclear ribonucleoprotein: RNA structure analysis and ATP-dependent interaction with U4/U6. Mol. Cell. Biol. 9: 3350-3359.

Bruzik, J.P., K. van Doren, D. Hirsh, and J.A. Steitz. 1988. Trans splicing involves a novel form of small nuclear ribonucleoprotein particles. Nature 335: 559-561.

Frendewey, D., A. Kramer, and W. Keller. 1987. Different small nuclear ribonucleoprotein particles are involved in different steps of splicing complex formation. Cold Spring Harbor Symp. Quant. Biol. 52: 287-298.

Guthrie, C. and B. Patterson. 1988. Spliceosomal snRNAs. Annu. Rev. Genet. 20: 671-708.

Hamm, J., N.A. Daishan, and I.W. Mattaj. 1989. Functional analysis of mutant Xenopus U2 snRNAs. Cell 59: 159-169.

Igel, A.H. and M. Ares Jr. 1988. Internal sequences that distinguish yeast from metazoan U2 snRNA are unnecessary for pre-mRNA splicing. Nature 334: 450-453.

Jones, M.H. and C. Guthrie. 1990. Unexpected flexibility in an evolutionarily conserved protein: RNA interaction: Genetic analysis of the Sm binding site. EMBO J. 8: 2555-2561.

Konarska, M.M. and P.A. Sharp. 1988. Association of U2, U4, U5, and U6 small nuclear ribonucleoproteins in a spliceosome-type complex in absence of precursor RNA. Proc. Natl. Acad. Sci. 85: 5459-5462.

Laird, P.W. 1989. Trans splicing in trypanosomes—Archaism or adaption? Trends Genet. 5: 204-208.

Lake, J.A., V.F. de la Cruz, P.C.G. Ferreira, C. Morel, and L. Simpson. 1988. Evolution of parasitism: Kinetoplastid protozoan history reconstructed from mitochondrial RNA gene sequences. Proc. Natl. Acad. Sci. 85: 4779-4783.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Mattaj, I.W. 1988. Use of RNP assembly and transport. In Structure and function of major and minor small nuclear ribonucleoprotein particles [ed. M.L. Birnstiel], pp. 100-114. Springer-Verlag, Heidelberg.

McPheeters, D.S., P. Fabrizio, and J. Abelson. 1989. In vitro reconstitution of functional yeast U2 snRNPs. Genes Dev. 3: 2124-2136.

Michaeli, S., T.G. Roberts, K.P. Watkins, and N. Agabian. 1990. Isolation of distinct small ribonucleoprotein particles containing the spliced leader and U2 RNAs of Trypanosoma brucei. J. Biol. Chem. 265: 10582-10588.

Milhausen, M.R.G., S. Nelson, S. Sather, M. Selkirk, and N. Agabian. 1984. Identification of a small RNA containing the trypanosome spliced leader: A donor of shared 5' sequences of trypanosomatid mRNAs? Cell 38: 721-729.

Miller, S.I. and D.F. Wirth. 1988. Trans splicing in Leishmania enriettii and identification of ribonucleoprotein complexes containing the spliced leader and U2 equivalent RNAs. Mol. Cell. Biol. 8: 2597-2603.

Mottram, J., K.L. Perry, P.M. Lizardi, R. Luhrmann, and N. Agabian. 1989. Isolation and sequence of four U snRNA genes of Trypanosoma brucei brucei: Identification of the trypanosome U2, U4, and U6 RNA analogues. Mol. Cell. Biol. 9: 1212-1223.

Murphy, W.J., K.P. Watkins, and N. Agabian. 1986. Identification of a novel branch structure as an intermediate in trypanosome mRNA processing: Evidence for trans splicing.
Nilsen, T.W. 1989. Mini Review: Trans-splicing in nematodes. Exp. Parasit. 69: 413–416.

Noller, H.F. 1984. Structure of ribosomal rRNA. Annu. Rev. Biochem. 53: 119–162.

Pan, Z. and C. Prives. 1989. U2 snRNA sequences that bind U2-specific proteins are dispensable for the function of U2 snRNP in splicing. Genes Dev. 3: 1887–1898.

Parker, R., P. Siliciano, and C. Guthrie. 1987. Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. Cell 49: 220–239.

Parsons, M., R.G. Nelson, K. Watkins, and N. Agabian. 1984. Trypanosome mRNAs share a common 5’ spliced leader sequence. Cell 38: 309–316.

Parzelt, E., K.L. Perry, and N. Agabian. 1989. Mapping of branch site in trans-spliced pre-mRNAs of Trypanosoma brucei. Mol. Cell. Biol. 9: 4291–4297.

Perry, K., K.P. Watkins, and N. Agabian. 1987. Trypanosome mRNAs have unusual “cap 4” structures acquired by addition of a spliced leader. Proc. Natl. Acad. Sci. 84: 8190–8194.

Salser, W. 1977. Globin messenger-RNA sequences. Analysis of base-pairing and evolutionary implications. Cold Spring Harbor Symp. Quant. Biol. 42: 985–1002.

Sanger, F., S. Nicklen, and A.R. Coulsen. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. 74: 5463–5467.

Shuster, E.O. and C. Guthrie. 1988. Two conserved domains of yeast U2 snRNA are separated by 945 nonessential nucleotides. Cell 55: 41–48.

Steitz, J.A., D.L. Black, V. Gerke, K.A. Parker, A. Kramer, D. Frendewey, and W. Keller. 1988. Functions of the abundant U snRNPs. In Structure and function of major and minor small nuclear ribonucleoprotein particles (ed. M.L. Birnstiel), pp. 115–154. Springer-Verlag, Heidelberg.

Sutton, R.E. and J.C. Boothroyd. 1986. Evidence for trans splicing in trypanosomes. Cell 47: 527–535.

Tschudi, C. and E. Ullu. 1990. Destruction of U2, U4, or U6 small nuclear RNA blocks trans splicing in trypanosome cells. Cell 61: 459–466.

Tschudi, C., F.F. Richards, and E. Ullu. 1986. The U2 RNA of Trypanosoma brucei gambiense: Implications for a splicing mechanism in trypanosomes. Nucleic Acids Res. 14: 8893–8903.

Tschudi, C., A.R. Krainer, and E. Ullu. 1988. The U6 small nuclear RNA from Trypanosoma brucei. Nucleic Acids Res. 16: 11375.

Tuerk, C., P. Gauss, C. Thermes, D.R. Groebe, M. Gayle, N. Guild, G. Stormo, Y. d’Aubenton-Carafa, O.C. Uhlenbeck, I. Tinoco, E.N. Brody, and L. Gold. 1988. CUUCCG hairpins: Extraordinarily stable RNA secondary structures associated with various biochemical processes. Proc. Natl. Acad. Sci. 85: 1364–1368.

Wu, J. and J.L. Manley. 1989. Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes Dev. 3: 1553–1561.

Zhuang, Y. and A.M. Weiner. 1989. A compensatory base change in human U2 snRNA can suppress a branch site mutation. Genes Dev. 3: 1545–1552.

Zuker, M. and P. Steigler. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. 9: 133–148.
A new U2 RNA secondary structure provided by phylogenetic analysis of trypanosomatid U2 RNAs.

T Hartshorne and N Agabian

*Genes Dev.* 1990, 4:
Access the most recent version at doi:10.1101/gad.4.12a.2121

---

**References**
This article cites 36 articles, 15 of which can be accessed free at: http://genesdev.cshlp.org/content/4/12a/2121.full.html#ref-list-1

**License**

**Email Alerting Service**
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.