Secondary structure of splice sites in adenovirus mRNA precursors

Stephen H. Munroe

Department of Biology, Marquette University, Milwaukee, WI 53233, USA

Received 24 July 1984; Revised and Accepted 22 October 1984

ABSTRACT

In order to investigate the possible role of RNA secondary structure in determining the efficiency and specificity of mRNA splicing, the structures of sequences at three acceptor splice sites in adenovirus were studied. Transcripts spanning intron-exon junctions were synthesized using SP6 RNA polymerase and analyzed using single and double-strand specific nucleases. Distinctive patterns of nuclease cleavage were observed for each of the 3 sites examined. At both sites in the E2a region sequences adjacent to the splice sites were particularly susceptible to digestion with T1 and S1 nucleases. In contrast, a splice site for hexon mRNA was largely resistant to these nucleases. The results obtained suggest that the conformation of the RNA at some, but not all, acceptor sites may enhance the accessibility of these sites to factors involved in splicing nuclear RNA and confirm the presence of a large, previously predicted hairpin structure centered on the acceptor site at 67 map units.

INTRODUCTION

The mechanism by which splice sites are recognized in nuclear precursors of mRNA is poorly understood despite extensive information concerning nucleotide sequences (1) and mutations (2-4) which affect splicing. A role for RNA secondary structure has been suggested in order to account for the effects of mutations on both the efficiency and specificity of splicing (2, 5). Furthermore, since the short consensus sequences found at the donor (5') and acceptor (3') splice sites are not sufficient in themselves to unambiguously determine the site of splicing (6,7), it is possible that additional information is provided by the higher order structure of the region bounding the sites. While much attention has focused on the possible role of intermolecular base pairing of U1 small nuclear RNA with intron sequences in a manner that facilitates splicing of adjacent exons (6) intramolecular interactions within the mRNA precursor may also be important in the efficient recognition and utilization of splice sites (7,8).

In order to investigate local secondary structure of splice sites found in adenovirus mRNA precursors present in vivo at low concentrations I have

used SP6 RNA polymerase (9,10) to synthesize in quantity a series of RNA molecules overlapping several well-characterized acceptor sites. The conformation of these molecules was probed by partial digestion with single- or double-strand specific nucleases. The size of the in vitro transcripts spanning one site was systematically varied in order to confirm the location of interacting sequences. The results of the detailed characterization of this site confirm the presence of a previously predicted (7) large hairpin structure at the acceptor site for the body of the adenovirus DNA-binding protein mRNA (DBP mRNA) located at 67 map units on the viral genome and provides evidence for additional stable intramolecular interactions. Two other acceptor sites were also characterized including another in DBP mRNA (11), and one for hexon mRNA located at 52 map units (12). The results presented here represent the first experimental characterization of secondary structure at splice sites in nuclear precursors of mRNA molecules.

METHODS

Plasmids

pSP62-PL, a transcription vector containing a SP6 phage promoter upstream of a multiple cloning site (9,10), was generously provided by Dr. D.A. Melton. Three recombinant plasmids were constructed by inserting fragments of adenovirus 2 DNA into the Hind III or Pst I site in the polylinker region. The adenovirus fragments inserted in the vector are described as follows, giving the position on the viral genome: 1.) a Pst I fragment from nucleotides 23,338 to 24,416 (64.9 to 67.4 map units--one map unit equals 359.4 nucleotides) inserted in the 1 orientation relative to the SP6 promoter (pSPP80-L); 2.) a Pst I fragment from nucleotide 24,417 to 25,451 (67.9 to 70.8) inserted in the 1 orientation (pSPP35-L); and 3.) a Hind III fragment from nucleotide 18,318 to 26,370 (50.1 to 73.4 map units) inserted in the r orientation (pH3A-R).

In Vitro Transcription

Plasmids purified by banding on CsCl (13) were digested with restriction enzymes (Pharmacia P-L Biochemicals or Bethesda Research Laboratories) according to the manufacturer's recommendation. Digested plasmids were extracted once with buffered phenol, twice with chloroform, precipitated with ethanol and resuspended in TE buffer (0.01M Tris-HCl, pH 8.0, and 0.001M EDTA) at 0.5 mg/ml. The linearized plasmids were transcribed using fraction V of SP6 RNA polymerase, prepared as previously described (9) from SP6 infected Salmonella typhimurium his 2558. Phage and bacteria were the generous gift

of Dr. Melton. Transcription reactions were carried out as described by Green *et al.* (10), except that reactions were scaled up to a volume of 0.5 or 1.0 ml. The standard reaction contained 10–20 $\mu\text{Ci/ml}$ ^3H -CTP (ICN, 22Ci/mmol) and 0.5mM each of ATP, GTP, CTP and UTP. Reactions were initiated by the addition of 140u/ml of SP6 polymerase and incubated for 45 to 60 min at 37°C. DNase I (Worthington, DPRF) was then added, the incubation continued for 10 min to digest the template and 0.1 volume of 0.2M EDTA, 0.5 volumes distilled phenol, 0.5 volumes chloroform with 1% isoamyl alcohol and 0.2% sodium dodecyl sulfate (BDH, specially pure) were added. Following extraction with phenol/chloroform, chloroform alone, and precipitation of ethanol, RNA was resuspended in 0.15 ml TE buffer and passed through G-100 Sephadex. Void volume fractions were pooled (0.6 ml) and ethanol precipitated. Yields of 10 to 100 μg of RNA were obtained from 1 ml transcription reactions depending on the size and sequence of the RNA. Transcripts uniformly labeled with ^{32}P -GTP (New England Nuclear, 600Ci/mmol) were sized by electrophoresis on acrylamide or agarose gels following denaturation with glyoxal (14).

End-labeling and Partial Digestion

RNA samples were routinely dephosphorylated before end-labeling in 0.1 ml 0.025 M Tris-HCl, pH 8.3, with calf intestinal alkaline phosphatase (Boehringer Mannheim, 0.01u/ml) for 30 min at 50°C. 5–15 μg of RNA was end-labeled with 100–150 μCi ^{32}P -cytidine bis-phosphate (pCp, New England Nuclear, 2900Ci/mmol). End-labeling was carried out in a total volume of 25 μl as described previously (15). Twelve units RNA ligase (Pharmacia P-L Biochemicals) were added and the sample was incubated at 4°C for 45 to 60 min and terminated by the addition of 0.1 volume 0.2M EDTA. Ten μg of yeast tRNA (Sigma) was added as carrier prior to extraction with phenol/chloroform and chloroform. Following precipitation with ethanol RNA was resuspended in 20 μl of water and an equal volume of loading buffer (10M urea, 20mM EDTA, 0.1% bromophenol blue and 0.1% xylene cyanol) and the sample was electrophoresed on a 5% acrylamide gel containing 8.3M urea and 0.5x TBE buffer (1x TBE is 0.08M Tris, pH 8.3, 0.088 boric acid, and 0.0025M EDTA). Bands containing full length end-labeled RNA were identified by autoradiography, cut out, and electroeluted. The RNA was extracted with phenol/chloroform, and ethanol precipitated after adding carrier tRNA. Samples were pelleted, rinsed twice with 70% ethanol and dissolved in water to a final concentration of sample RNA of 0.1 to 0.3 mg/ml with tRNA carrier added to 1 mg/ml.

All nucleases used for partial digestion of RNA were obtained from Pharmacia P-L Biochemicals. Sequence-specific partial digestion with T1, U2

and Phy M and B cerus ribonucleases (sequence grade) were carried out in buffers provided by the manufacturer at 60°C. Structure-specific digestions were carried out under conditions described by Vournakis *et al.* (16). Specifically, ribonuclease T1 digestion was carried out in 0.01M sodium citrate, pH 5.0, 0.2M NaCl, 0.01M MgCl₂. S1 digestion was in 0.04M sodium acetate, pH 4.5, 0.01M ZnSO₄, 0.2M NaCl, and ribonuclease V1 digestion was in 0.1M Tris-HCl, pH 7.5, 0.002M MgCl₂, and 0.2M NaCl. Each digestion contained 1 µg of RNA in a total volume of 2 µl. Unless specified otherwise, these digestions were incubated at 37°C for 5 min. Ribonuclease T1 digestions were also carried out at 60°C for 8 min in Buffer I (0.017M sodium citrate, 0.85mM EDTA, 0.02% xylene cyanol and 0.04% bromophenol blue) with 0-8M urea. Alkaline hydrolysis was carried out in 0.1M NaHCO₃/Na₂CO₃ buffer, pH 9.2, and 0.033M EDTA at 90°C for 4 to 6 min. All reactions were stopped by the addition of 2 µl of loading buffer, frozen on dry ice and thawed for 15-20 sec at 90°C immediately before loading.

Gel electrophoresis was carried out on thin 0.04 x 40 cm 8% or 12% polyacrylamide sequencing gels in 0.5x TBE with 8.3M urea with an acrylamide:methylene bisacrylamide ratio of 20:1. Electrophoresis was at 40 to 60W. Gels were autoradiographed with an intensifying screen at -70°C. Each set of reactions analyzed included an undigested RNA sample as a control for ribonuclease contamination.

Computer Analysis of RNA Secondary Structure

The results of experiments probing RNA structure were analyzed using an efficient interactive RNA folding program (RNAFLD) generously provided by Dr. M. Zuker and the National Research Council of Canada (17,18). The revised base stacking energies of Tinoco *et al.* (19,20) were used. This program was run on a VAX/VMS-780 system which folded sequences of 1.0 kilobases in 2 h cpu time. Additional structures in subfragments of these sequences were examined interactively. Program features permit the incorporation of nuclease digestion data. Sequences analyzed included the L3 region (12) and E2a region (11, R.J. Roberts, personal communication) of adenovirus 2.

RESULTS

Plasmids and Transcripts

As summarized in Table 1 and Fig. 1, three plasmids spanning three known acceptor splice sites in adenovirus were used to synthesize six different RNA molecules analyzed in this study. The sizes of the *in vitro* transcripts

Table 1: In Vitro Transcripts

Adenovirus Transcription Unit	Template	Runoff Site	Transcript Length (nucleotides)	Distance from 3' end to Splice Site (nucleotides)
E2a	pSPP80-L	Xho I	542	163
	pSPP80-L	Ava I	429	50
	pSPP80-L	Hinf I	354	(-24)
	pSPP35-L	EcoR V	783	72
L3	pH3A-R	Ava I	601	83
	pH3A-R	Sma I	634	115

were determined by glyoxal gel electrophoresis (14) and corresponded exactly to those calculated in Table 1 from the sequences of the SP6 promoter (M.R. Green, personal communication), polylinker insertion site and adenovirus 2 fragments. End-label analysis of two of these transcripts indicates that transcription terminates within one nucleotide of the 5' end of the coding strand. In this paper specific transcripts are described in terms of the template plasmid and the restriction enzyme used to generate the 3' runoff site, e.g., the pSPP80-L/XhoI transcript is a RNA molecule synthesized on pSPP80-L template truncated at the Xho I site.

Secondary Structure of Acceptor Site at 67 Map Units

Earlier modeling studies suggested that the acceptor site at 67 map units in the E2a region of adenovirus might be located at the apex of an exceptionally large hairpin structure (7). In order to determine if this structure shown in Fig. 2 is in fact present, 3' end-labeled transcripts

FIGURE 1: Position of splice sites and in vitro transcripts within E2a region of adenovirus. Bold lines represent the 3 exons and thin lines the 2 introns of early DBP mRNA (11). Position of E2a transcripts listed in Table 1 are shown.

FIGURE 2: Proposed RNA hairpin centered on splice site at 67 map units in adenovirus 2 DBP pre-mRNA. Nucleotides 24,146 to 24,041 are shown for the 1-strand E2a transcript (11, R.J. Roberts, personal communication). Major cleavage sites shown in Fig. 3 are indicated. (→) primary T1 or S1 (single-strand specific) cut sites. (↗) secondary T1 cut sites. (▲) very strong and (▴) moderately strong V1 (double-strand specific) sites. Bold, curved arrow indicates site of splicing (11).

spanning this site were analyzed by partial digestion with strand-specific nucleases. Ribonuclease T1 and S1 nuclease were used to identify single-stranded regions of RNA. Ribonuclease T1 cuts after G residues while S1 cuts in a relatively non base-specific fashion (16). Ribonuclease V1 was used to identify double-stranded regions of RNA (16). This enzyme lacks base-specificity, although some sites in double-stranded stems are strongly preferred for reasons which are poorly understood (16,21,22). T1 ribonuclease digestions were also carried out at 60°C in 0 to 8M urea in order to define regions of the secondary structure affected by increasingly stringent denaturing conditions. Base-specific digestions with U2, Phy M and B. cerus ribonucleases provided sequence-specific markers.

T1 ribonuclease digestion of pSPP80/AvaI.

Initial studies of the acceptor site at 67 map units were carried out with the pSPP80/AvaI transcript (Fig. 1) which has its 3' end directly at the base of the predicted hairpin structure (Fig. 2). Since the 106 nucleotide hairpin contains 32 G residues distributed throughout its length, the cuts generated by T1 ribonuclease shown in lanes 6-17 of Fig. 3A and lane 1 of Fig. 3B are especially informative. The 32 G residues in this structure can be grouped into 4 classes on the basis of their relative intensity under different conditions of T1 digestion. These classes correspond to specific loop and stem regions within the proposed structure. In the following discussion positions of residues are numbered in relation to the splice site with +1 being a cut on the 3' side of the first exon nucleotide and -1 a cut on the 3' side of the last intron nucleotide.

The first class of T1 cut sites includes 4 G residues cleaved at the lowest T1 concentration used. As shown in lanes 15-17 of Fig. 3A the 3' region of pSPP80/AvaI is cut at positions -8, -19, +37 and +35 (in order of

decreasing intensity) after digestion for 2 to 20 min with 0.005 units of T1. These positions correspond to 3 or 4 different loops in the proposed structure. G (-8), the strongest cut site, is found in the apical loop, G (-19) and G (+37) are in small internal loops, and G (+35) is present either in a single base bulge loop (as shown) or at the 5' side of an expanded interior loop including G (+37). No other G residues within this region are cleaved with 0.005 units of T1 ribonuclease (faint bands at positions -5, -4 and +14 represent a contaminating nuclease and were not seen in duplicate experiments). Cuts at these 4 positions may represent primary cut sites within the pSPP80-L/AvaI transcript, since they are observed under digestion conditions which cleave a fraction of the total RNA sample. The remaining G residues within the proposed hairpin stem are cleaved by ribonuclease T1 only under more extreme conditions as detailed below.

The second class of G residues consists of 5 sites which are cleaved very strongly at higher enzyme concentrations or at elevated temperatures. As shown in lanes 6-14 of Fig. 3A these sites at positions -1, +1, +3, +6 and +7 all lie on the 3' side of an 8 or 9 base-pair stem which comprises the apical portion of the large hairpin. While no cleavages are observed in this region at low concentrations of enzyme (lanes 15-17, Fig. 3A) at higher concentrations of T1 (lane 13, Fig. 3A or lane 1, Fig. 3B) or at 60°C in the presence of urea (lanes 6-12, Fig. 3A) these positions are cut more intensely than some of the class 1 sites described above. This behavior strongly suggests that class two sites represent secondary cutting in a region made very accessible to ribonuclease T1 by upstream primary cuts, presumably those at positions -8 or -19. Compare lanes 16, 14, and 13 in Fig. 3A for digestion after 5 min at increasing T1 concentrations. Digestion at 60°C, with urea (Fig. 3A, lanes 6 to 12) or without urea (results not shown) also produces strong cuts at these 5 positions, showing that digestion under mildly denaturing conditions also exposes these residues. This result is consistent with the relatively weak stacking expected for the 7AU and GU pairs in this stem. At 60°C the apical region appears to open sufficiently for relatively uniform cutting within the stem as indicated by the similar intensity of these 5 bands. The position of the class one (primary cut site) and class two (strong secondary cut sites) is summarized in Fig. 2.

The third and fourth classes of G residues are those residues cleaved weakly by ribonuclease T1 in the presence of urea at 60°C. The third set consists of 11 G residues on the 3' side of the hairpin stem at positions

12-33), which are cut much more weakly than the class two residues at 60°C, as seen in lanes 6 to 12, Fig. 3A. The fourth class consists of 12 G residues located on the opposite 5' side (at positions -58 to -27) of the same stem region as the third class. G residues in this fourth class are cut only at 60°C in concentrations of urea equal to or greater than 4M and appear as very faint bands (Fig. 3A, lanes 9-12). On some gels, each of the 12 residues on the 5' side was clearly resolved as a pattern of very faint, fine bands. The difference in the intensity of cut sites on opposite sides of the hairpin stem probably reflects their position relative to the strong apical cut sites and the 3' labeled end, not an intrinsic difference in the accessibility of these 23 residues found in this GC-rich stem (Fig. 2).

FIGURE 3: Partial digestion of pSPP80-L/AvaI transcript of E2a splice site at 67 map units. Autoradiographs of 12% (A) and 8% (B) polyacrylamide gels. A. Lane 1 control, no nuclease; lanes 2, 18 and 28 alkaline hydrolysis; lane 3 5' phosphorylation following digestion with 0.025u T1 in Buffer I; lane 4 nonphosphorylated control, same as lane 3 but no kinase was added; lane 5 0.007u U2; lane 6 3.75M urea, 0.07u T1; lanes 7 and 8 2M urea, 0.07 and 0.025u T1; lanes 9 and 10 4M urea, 0.07 and 0.025u T1; lanes 11 and 12 6M urea, 0.07 and 0.025u T1; lanes 13 and 14 T1 0.05 and 0.017u T1 (no urea); lanes 15 to 17 0.005u T1 for 2, 5 and 20 min; lanes 19 and 20 0.25 and 0.1u S1; lanes 21 and 22 0.05u S1 for 5 and 20 min; lanes 23 and 24 0.028 and 0.009u V1; lanes 25 to 27 0.014 u V1, 2, 5 and 20 min. B. Lanes 1 to 4 same as lane 14 and lanes 18 to 20 in A; lane 5 0.05u S1, 2 min; lanes 6 to 8 same as lanes 21 to 23 in A. Bold arrow marks splice site.

S1 nuclease and V1 ribonuclease digestion of pSPP80/AvaI. Partial digestions with S1 nuclease and ribonuclease V1 also support the proposed hairpin structure. Since both of these enzymes cut on the 3' side of phosphate groups leaving a 5' phosphate on the 3' end-labeled fragments it is necessary to determine the mobility of labeled fragments obtained with T1 ribonuclease which leaves a free 5' hydroxyl group. 3' labeled fragments were first cleaved with T1 ribonuclease and one-half of this partial digest was phosphorylated at the 5' OH end with polynucleotide kinase (16), using unlabeled ATP. This was run in parallel with the other half of the same digestion sample which had not been phosphorylated. As shown in Fig. 3A (lanes 3 and 4) phosphorylated residues run one-half step slower than control T1 cut fragments throughout the length of the gel. This result is paradoxical in that the extra phosphate group increases the charge density on the fragments. Although it has been suggested that this shift, observed previously with fragments obtained from S1 nuclease digestions (16,23), is due to a phosphatase activity contaminating the nuclease (23), such contamination could not account for the T4 polynucleotide kinase dependent shift demonstrated here.

A series of 4-6 major S1 nuclease cuts are seen at increasing levels of digestion (Fig. 3B, lanes 3 to 7) correspond to cuts between residues -7 to -2. These cuts are centered within the apical hairpin loops as shown in Fig. 2. While no other strong S1 cut sites are seen within 175 nucleotides from the 3' end (Fig. 3A, lane 19-22) series of 3 faint bands at positions -21 to -19 mark the 5' side of a small internal loop (Fig. 3B, lanes 3, 4 and 7) and a cluster of 4-5 additional faint S1 cuts in an A-rich region at positions -66 to -59 mark the 5' end of the hairpin.

The most prominent V1 cleavage sites are 6 consecutive cuts at positions -14 to -9 on the 5' side of the apical stem (Fig. 3A, lanes 23-27, Fig. 3B, lane 8). These cuts extend to the 3' side of the U (-9) probably representing 3' overcutting by V1 (22). Several other moderately strong V1 cleavage sites are observed on the 5' side of the apical stem region. These cuts confirm the double-stranded structure of this region, which is also defined as a preferred site of secondary cutting with T1. Two weaker V1 cut sites are consistently observed at positions -34 and -29 (Fig. 3A, lane 23), on the 5' side of G residues present in single base bulge loops. The positions of these sites of preferred V1 cutting are summarized in Fig. 2. The lack of other strong V1 cut sites on the helical stem may reflect a preference for cutting by ribonuclease V1 at the end of helical regions.

FIGURE 4: Transcripts overlapping the hairpin structure at 67 map units in region E2a. The additional 113 nucleotides present at end of pSP80-L/XhoI transcript may displace all or part of hairpin structure found in pSP80-L/AvaI. The shorter pSP80-L/HinfI transcript cannot form hairpin shown. 51 nucleotides transcribed from pSP62-PL are indicated at right.

Partial digestion of other transcripts overlapping the inverted repeat at 67 map units. It seemed possible that the formation of the hairpin structure observed in the pSP80-L/AvaI transcript is strongly influenced by the selection of a 3' runoff site directly adjacent to the base of this stem. Sequences 3' of this site, for example, can potentially pair with GC-rich regions in the 3' side of the stem. It is also possible, though not likely, that the digestion patterns interpreted above in terms of the large hairpin shown, may be generated by a number of smaller hairpin structures. Therefore, both shorter (pSP80-L/HinfI) and longer (pSP80-L/XhoI) transcripts overlapping sequences within the model structure were analyzed as outlined schematically in Fig. 4.

Analysis of both transcripts supports the proposed secondary structure. As expected, the shorter pSP80-L/HinfI transcript displayed a very different cleavage pattern than the Ava I truncated transcript in the 33 base region overlapping the hairpin stem. For example, strong cut sites are observed with T1 and/or S1 nuclease at positions -46 to -50 consistent with a short hairpin loop in the 3' terminal region (results not shown). On the other hand, the cleavage pattern of the longer Xho I truncated transcripts in this 106 nucleotide region was almost identical to that observed in the pSP80-L/AvaI transcript. Essentially every cleavage fragment observed in the region +48 to -58 in Fig. 3 was present in comparable digests of the two transcript with the same relative intensity. These include cuts at the 4 class 1 primary sites (with G -8 being most intense), 5 intense cuts corresponding to class 2 sites and 11 fainter class 3 cuts. Cleavages of pSP80-L/XhoI with S1 nuclease and ribonuclease V1 were also nearly identical to those

described above for pSPP80/AvaI (results not shown). Additional cleavage sites within the 3' region of the pSPP80-L/XhoI transcript were also used to model secondary structure in a larger region bordering this splice site as shown in Fig. 7A below.

Secondary Structure of Two Other Acceptor Splice Sites in Adenovirus

A striking feature of the digestion patterns shown in Fig. 3 is that all 3 enzymes tested cleave preferentially within a small region immediately upstream of the splice acceptor site. This observation suggests that secondary structure may enhance the accessibility of sites proximal to the splice site, including the upstream pyrimidine tract characteristic of all 3' splice sites in nuclear mRNA. In order to determine whether or not the preferential cleavage sites observed in this region is a general feature of acceptor splice sites, transcripts from two other regions of adenovirus were studied.

The pSPP35-L/EcoRV transcript (Fig. 1, Table 1) includes 660 nucleotides of the first intron of E2a and almost all of the short second exon of DBP mRNA (Fig. 1). Enzymatic digestion of this transcript with ribonuclease T1 or S1 nuclease reveals 2 extended regions resistant to single-strand specific nuclease. One of these regions extends from the 3' end (+72) to +22 (not shown), and the other one lies between residues -26 to -78 upstream of the splice point (Fig. 5, lanes 12 to 17). Since these regions are also cut frequently with ribonuclease V1, they are presumably double-stranded in nature. The region extending about 25 nucleotides on either side of this splice site is highly accessible to single strand specific cutting including both primary and secondary cut sites. The major primary S1 nuclease cleavage site at residue -24 (see lane 14) marks the 3' boundary of this region. Four T1 sites (-8, -4, +2, +4) observed at low enzyme concentrations (Fig. 5, lanes 7 and 8) bracket the splice site. Additional sites of S1 and T1 cleavage (e.g., positions +13 and +15, lanes 5 and 6, or positions -13 to -6, lanes 12 and 13 in Fig. 5) probably represent secondary cut sites since their intensity varies sharply with increased digestion and overlap short regions of consecutive V1 cut sites. A model for the secondary structure of this region is discussed below (Fig. 7B).

While partial digestion analysis of each of the two acceptor splice sites in region E2a indicates that the regions adjacent to the splice point are preferentially cut with single strand specific nucleases including both primary and secondary cuts, this contrasts with the third site studied. The acceptor splice for hexon mRNA site at 52 map units was probed in two

FIGURE 5: Partial digestion of pSPP35-L/EcoRV transcript of E2a splice site at 69 map units. Autoradiograph of 8% polyacrylamide gel. Lane 1 control, no nuclease; lanes 2 and 11 alkaline hydrolysis; lane 3 0.007u U2; lane 4 3.75M urea, 0.007u T1; lanes 5 and 6 0.003 and 0.0017u T1; lanes 7 to 10 0.0005u T1, 2, 5, 15 and 45 min; lanes 12 and 13 0.25 and 0.1u S1; lanes 14 to 17 0.05u S1, 2, 5, 15 and 45 min; lanes 18 and 19 0.028 and 0.009u V1; lanes 20 to 22 0.014u V1, 2, 5, 15 min. Bold arrow splice site.

transcripts extending 115 (pH3A-R/SmaI) and 83 (pH3A-R/AvaI) nucleotides downstream of the acceptor site. In both transcripts the secondary structure at the acceptor site is essentially identical. This site is located in the center of a 40 nucleotide region which is cut weakly by T1 and S1 nucleases (Fig. 6) even at relatively high enzyme concentrations (Fig. 6, lanes 4 and 8) and under partially denaturing conditions (results not shown). These observations suggest that this region is largely double-stranded, consistent with the many V1 cleavage sites also present in this region (Fig. 6, lane 1).

FIGURE 6: Partial digestion of pH3A-R/AvaI transcript of L3 splice site at 52 map units. Autoradiograph of 8% polyacrylamide gel. Lane 1 0.07u V1; lanes 2 to 4 0.05, 0.1 and 0.25u S1; lane 5 alkaline digestion; lanes 6 to 8 0.0004, 0.0017 and 0.05u T1. Bold arrow marks splice site.

Modeling of Secondary Structure in E2a Transcripts

Secondary structures for in vitro transcripts were modeled using an efficient RNA folding program, RNAFLD (17,18). Fig. 7 shows possible secondary structures for sequences of about 250 nucleotides spanning the E2a acceptor sites at 67 and 69 map units. These structures show a good fit to digestion data and represent, with only minor alterations, the most stable structures possible for these sequences. The structure predicted for the 67 map unit site (Fig. 7A) is exceptionally stable in terms of its overall standard free energy (-174 kcal/mole or -0.70 kcal/mole/nucleotide). This stabilization energy reflects both the GC-rich base composition and the presence of several unusually large regions of complementary sequences. The structure shown consists of 3 hairpin stems including the large hairpin shown in Fig. 2. This latter structure and the apical region of the middle

FIGURE 7: Proposed secondary structure for acceptor splice sites in DBP mRNA. Structure of sites at (A) 67 map units (nucleotides 24,173 to 23,926) and (B) 69 map units (24,961 to 24,713). Major cleavage sites for S1 ($\rightarrow$), T1 ($\rightarrow$) and V1 ($\rightarrow$) nucleases are shown. ($\rightarrow$) mark 3' ends of transcripts, ($\rightarrow$) marks splice sites. Some uncertainty (≤ 3 nucleotides) exist in position of cleavage sites between +110 to +80 in (A) and -65 to -85 in (B) and several V1 sites upstream of -90 are omitted in (B).

hairpin were found in all large subfragments of E2a examined which overlap this region. The 3' hairpin, however, and the base of the middle hairpin shown are one of several relatively stable possible conformations.

Fig. 7B shows a possible secondary structure for the splice site at 69 map units. Although this structure is substantially less stable than the one shown in Fig. 7A, with an overall calculated free energy of -111 kcal/mole (-0.45 kcal/mole/nucleotide), it also is found within all 1 kb subfragments examined which span this region. The structure shown fits nuclease data obtained for pSPP35/EcoRV in most respects. The region from -26 to +20, for example, contains many sites of primary and secondary cleavage by enzymes specific for single-stranded regions. In the model this region spanning the splice site is represented by a series of 3 short branched hairpin structures. Adjacent sequences include predominantly double-stranded segments resistant to S1 and T1 nuclease and susceptible to V1 nuclease. An interesting feature of this structure is that the donor site for the second exon falls near the 3' end of a large stem. Model structures of sequences which include this exon and adjacent intron sequences support the possibility that this donor site lies at the boundary of the compactly folded domain shown.

Although models of secondary structure of relatively long sequences such as these must be regarded as tentative due to uncertainties in the calculations and data used in their derivation (19,24) and difficulties in interpreting nuclease digestion data (16), the structures shown in Fig. 7 are useful as working models for evaluating the role of secondary structure in RNA splicing and in organization of nuclear ribonucleoprotein complexes (25).

DISCUSSION

In this study single and double-strand specific nucleases have been used to characterize secondary structure of RNA sequences spanning 3 acceptor splice sites in adenovirus mRNA. The results obtained indicate that each site, although functionally homologous with respect to RNA processing, displays a distinctive pattern of single and double-stranded regions. The direct structural analysis of relatively small in vitro transcripts as used in this study provides a general method for the analysis of higher order structure in large RNA molecules, especially when complemented with other, indirect methods of analysis (26).

The secondary structure of splice site sequences and, more generally,

the accessibility of these regions to proteins or other factors involved in splicing may be important for splice site function. Intron sequences directly upstream of acceptor splice sites are required for RNA splicing. Acceptor sites in all known nucleus-encoded mRNA precursors have a pyrimidine tract directly upstream of an invariant AG dinucleotide (1,27). In yeast another upstream sequence, called the TACTAAC box, is required for acceptor site recognition (28,29). Recent work has identified a possibly analogous upstream sequence involved in splicing globin (30) and adenovirus (31) mRNA. Sites of preferential nuclease cleavage mapped in this study may be indicative of the accessibility of residues within splice site proximal sequences. At the 67 map unit site studied sequences directly upstream of the splice site (residues -1 to -8) are exceptionally accessible to digestion with single-strand specific nucleases (Fig. 3). Adjacent nucleotides within a short double-stranded stem (-9 to -14) are especially susceptible to a double-strand specific nuclease (Fig. 3). Thus the region upstream of this splice site is accessible to nucleases of different specificity. The major sites of primary cleavage by single-strand nucleases at two other sites examined lie 22-24 and 40-50 nucleotides upstream of the acceptor sites at 69 and 52 map units respectively. Both acceptor sites in DBP mRNA are similar in that the regions bounding the splice sites are readily cleaved by high T1 or S1 nuclease concentrations or under mild denaturation conditions, suggesting that these regions are readily opened for single-strand specific nuclease attack. In contrast, sequences immediately adjacent to the acceptor site for hexon mRNA are relatively resistant to single-strand specific cleavage under all conditions tested. These differences in susceptibility to nuclease digestion may reflect variations in the location and accessibility of sequences required for splicing.

The secondary structure of the acceptor site for DBP mRNA at 67 map units was analyzed in detail since this site is located at the center of an unusual, large inverted repeat sequence. Digestion data obtained from 3 transcripts overlapping this site were fully consistent with a predicted hairpin stem and loop structure. Detailed analysis of this data for the pSP80/AvaI transcript provide strong evidence for this structure (Fig. 2). Other model structures (Fig. 7) proposed for large regions bounding the two acceptor sites in DBP mRNA are also in generally good agreement with partial digestion data (Figs. 5 and 6). These structures, however, should be regarded primarily as working models, given the potential complexity of RNA folding in sequences of this length and difficulties in

the interpretation of nuclease data (16).

The major question left unanswered by this study is whether or not secondary structure of splice sites, such as those described here, plays a role in splice site function. The conformation of these three sites in vivo may conceivably be quite different from that studied in model transcripts in vitro. Proteins bound to RNA transcripts may influence secondary structure or higher order folding of RNA molecules (32). The extensive base pairing interactions between intron and exon sequences, proposed in Figs. 2 and 7, must eventually be disrupted in order to separate the excised introns from the processed mRNA.

A number of lines of evidence, however, suggest that local secondary structure, such as that described here, may play a role in splice site function. While the minimal information necessary to define functional splice sites appears to reside in rather short sequences (33,34), the context in which these sequences occur is of critical importance (4,34). The diverse effects of various mutations (3) and rearrangements (4,35) are not readily explained by a straightforward linear scanning mechanism. The efficient splicing of chimeric genes constructed from two different genes (36) demonstrates that sequences in adjacent exons or at the ends of a single intron are not uniquely related. Such results point to a general role for secondary or higher order structure in determining the sites utilized in RNA splicing. Finally, the detailed analysis of the splicing of other types of RNA molecules, tRNA (37,38), rRNA (20,39) and mitochondrial mRNA (40, 41), demonstrates the importance of RNA secondary structure in the processing of these molecules. Although splicing of these molecules clearly differs in many ways from that of mRNA in the nucleus, it is possible that similarities in mechanism or principle of splice site recognition underlie splicing of different RNA species.

ACKNOWLEDGEMENTS

I wish to thank D.A. Melton and M. Rebagliati for advice and stocks used in the preparation of SP6 polymerase, M.R. Green and R.J. Roberts for sequence data prior to publication, M. Zuker for the RNA folding programs, J.P. Schaub for assistance with computer programs, G.L. Waring for comments on this manuscript, and B. DeNoyer for expert typing. In particular, I wish to acknowledge the skillful technical assistance of R.S. Duthie. This work was supported by grants from the N.S.F. (PCM8302825), American Cancer Society (Milwaukee Division) and N.I.H. (RR07196).

REFERENCES

1. Mount, S.M. (1982) Nucl. Acids Res. 10, 459-472.
2. Khoury, G., Gruss, P., Dhar, R. and Lai, C.-J. (1979) Cell 18, 85-92.
3. Treisman, R., Orkin, S.H. and Maniatis, T. (1983) Nature 302, 591-596.
4. Wieringa, B., Meyer, F., Reiser, J. and Weissmann, C. (1983) Nature 301, 38-43.
5. Kuhne, T., Wieringa, R., Reiser, J. and Weissmann, C. (1983) EMBO J. 2, 727-733.
6. Mount, S.M. and Steitz, J.A. (1983) in Methods of DNA and RNA Sequencing, Weissman, S.M., Ed., pp. 399-436, Praeger Publishers, New York.
7. Munroe, S.H. (1983) Nucl. Acids Res. 11, 8891-8900.
8. Burnett, L. (1982) J. Theoret. Biol. 97, 351-366.
9. Butler, E.T. and Chamberlain, M.J. (1982) J. Biol. Chem. 257, 5772-5778.
10. Green, M.R., Maniatis, T. and Melton, D.A. (1983) Cell 32, 681-694.
11. Kruijer, W., van Schaik, F.M.A. and Sussenbach, J.S. (1982) Nucl. Acids Res. 10, 4493-4500.
12. Akusjarvi, G. and Persson, H. (1981) J. Virol. 38, 469-482.
13. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor.
14. McMasters, G.K. and Carmichael, G.C. (1977) Proc. Natl. Acad. Sci., U.S.A. 74, 4835-4838.
15. England, T.E. and Uhlenbeck, O.C. (1978) Nature 275, 560-561.
16. Vournakis, J.N., Celantano, J., Finn, M., Lockard, R.E., Mitra, T., Pavlakis, G., Troutt, A., van den Berg, M. and Wurst, R. (1981) in Gene Amplification and Analysis, Chirikjian, J.D. and Papas, T.S., Eds., Vol. 2, pp. 268-298. Elsevier/North-Holland, New York.
17. Zuker, M. and Stiegler, P. (1981) Nucl. Acids Res. 9, 5125-5140.
18. Jacobson, A.B., Good, L., Simonetti, J. and Zuker, M. (1984) Nucl. Acids Res. 12, 45-52.
19. Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40-41.
20. Cech, T.R., Tanner, N.K., Tinoco, I., Wier, B.R., Zuker, M. and Pearlman, P.S. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 3903-3907.
21. Lockard, R.E. and Kumar, A. (1981) Nucl. Acids Res. 9, 5125-5140.
22. Auron, P.E., Weber, L.D. and Rich, A. (1982) Biochem. 21, 4700-4706.
23. Garrett, R.A. and Olesen, S.O. (1982) Biochem. 21, 4823-4830.
24. Woese, C.R., Gutell, R.R., Gupta, R. and Noller, H.F. (1983) Microbiol. Rev. 47, 621-669.
25. Munroe, S.H. (1982) J. Mol. Biol. 162, 585-606.
26. Qu, L.H., Michot, B. and Bachellerie, J.-P. (1983) Nucl. Acids Res. 11, 5903-5920.
27. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
28. Langford, C.J. and Gallwitz, D. (1983) Cell 33, 519-527.
29. Piklielny, C.W., Teem, J.L. and Rosbach, M. (1983) Cell 34, 395-403.
30. Ruskin, B., Krainer, A.R., Maniatis, T. and Green, M.R. (1984) Cell 38, 317-331.
31. Padgett, R.A., Konarska, M.M., Grabowski, P.J., Hardy, S.F. and Sharp, P.A. (1984) Science 225, 898-903.
32. Thomas, J.O., Raziuddin, Sobuta, A., Boublik, M. and Szer, W. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 2888-2892.
33. Wieringa, B., Hofer, E. and Weissmann, C. (1984) Cell 37, 915-925.
34. Rautmann, G., Matthes, H.W.D., Gait, M.J. and Breathnach, R. (1984) EMBO J. 3, 2021-2028.

35. Lang, K.M. and Spritz, R.A. (1983) *Science* 220, 1351-1355.
36. Chu, G. and Sharp, P.A. (1981) *Nature* 289, 378-382.
37. Swerdlow, H. and Guthrie, C. (1984) *J. Biol. Chem.* 259, 5197-5207.
38. Baldi, M.I., Mattoccia, E. and Tocchini-Valentini, G.P. (1983) *Cell* 35, 109-115.
39. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) *Cell* 31, 147-157.
40. Davies, R.W., Waring, R.B., Ray, J.A., Brown, T.A. and Scazzocchio, C. (1982) *Nature* 300, 719-724.
41. Schmelzer, C., Schmidt, C., May, K. and Schweyen, R.J. (1983) *EMBO J.* 2, 2047-2052.