Splice site consensus sequences are preferentially accessible to nucleases in isolated adenovirus RNA

Stephen H. Munroe* and R. Scott Duthie⁺

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

Received 22 July 1986; Revised and Accepted 26 September 1986

ABSTRACT

The conformation of RNA sequences spanning five 3' splice sites and two 5' splice sites in adenovirus mRNA was probed by partial digestion with single-strand specific nucleases. Although cleavage of nucleotides near both 3' and 5' splice sites was observed, most striking was the preferential digestion of sequences near the 3' splice site. At each 3' splice site a region of very strong cleavage is observed at low concentrations of enzyme near the splice site consensus sequence or the upstream branch point consensus sequence. Additional sites of moderately strong cutting near the branch point consensus sequence were observed in those sequences where the splice site was the preferred target. Since recognition of the 3' splice site and branch site appear to be early events in mRNA splicing these observations may indicate that the local conformation of the splice site sequences may play a direct or indirect role in enhancing the accessibility of sequences important for splicing.

INTRODUCTION

mRNA splicing is a remarkably precise process obligatory for the expression of most known mRNA molecules in higher eukaryotes. This precision is evident in the large size (>50 kilobases) or large number of introns (>50) present in some mRNA precursors (pre-mRNAs) which undergo efficient and accurate splicing (1,2). Despite extensive investigation of sequences (3), intermediates (4,5) and factors (6-8) required for mRNA splicing, the mechanism by which splice sites are recognized is poorly understood. Although it is clear that consensus sequences present at the 5' and 3' splice sites are essential for splicing (9), these sequences are, as presently defined, too short or too loosely conserved to determine unambiguously the site of splicing (3,10). For example, 3' splice sites in pre-mRNAs are characterized by 3 sequence elements: First, a tightly conserved AG dinucleotide immediately upstream of the splice site (3); second, a polypyrimidine tract of variable length and composition adjacent to the AG dinucleotide in the intron (3); and third, a branch point consensus sequence (YURAY where Y = U or C and R = A or G) which includes an adenylate residue

located approximately 20-40 nucleotides upstream of the splice site where the 5' end of the excised intron is ligated to a 2' OH group to form a lariat structure (4,11,12). However, this combination of sequences occurs in most pre-mRNAs much more frequently than do functional 3' splice sites. Thus, further information is necessary to specify precisely the site of splicing.

Information necessary for specifying the exact position of splice sites in pre-mRNAs may take the form of either additional sequences or elements of higher order structure present in the 3-dimensional structure of the RNA molecule in vivo. With respect to the latter possibility it is interesting to note that in all other types of intron containing RNA molecules, including tRNA precursors (13), and class I (14-16) and class II (17-19) introns, the conformation of the RNA substrate clearly plays a determinative role in splicing. This role is especially evident in the case of those class I and class II transcripts which have been shown to undergo autocatalytic splicing in the absence of proteins (14-19). In each self-splicing RNA molecule a conserved pattern of RNA secondary structure located near the splice junction and within internal regions of the introns is essential for splicing (15,16,20). A number of mechanistic similarities, as well as differences, have been noted between class I and class II introns and those of nuclear pre-mRNAs (21). Class II introns, in particular, resemble nuclear pre-mRNAs in forming lariat intermediates with branch points near the 3' splice site (17-19). This branch point site appears to fall within a highly conserved region of secondary structure (18,19).

In contrast to the other intron-containing molecules, however, there is no evidence for the conservation of specific base pairing in nuclear pre-mRNA. Although a number of models have been proposed for specific base pairing between opposite ends of the intron (22,23) or between adjacent exons (24,25), the significance of these proposals has been questioned since transcripts containing heterologous exons or mutations which disrupt the proposed secondary structure seem to be spliced efficiently (26,27). It remains possible, however, that RNA conformation may play a more general role in the recognition of splice sites or modulation of splice site activity in a way which does not require specific long-range secondary structure. For example, local features of the higher order structure of the pre-mRNA might directly affect the accessibility of splice site regions to factors involved in splicing or might mediate the cooperative assembly of the active splicing complex in less direct ways.

In a previous study we observed that two of three 3' splice sites

examined were preferentially accessible to digestion with single-strand specific nucleases (28). In this paper, we report that sequences near the branch point consensus sequence at each of five 3' splice sites examined are cleaved by single-strand specific nucleases. Furthermore, at all of these sites sequences near the branch point consensus sequences or the 3' splice site itself are the preferred sites of cleavage within a region of 150-250 nucleotides surrounding the splice site. These observations indicate that the local conformation of splice site proximal sequences may enhance the accessibility of short single-stranded regions near 3' splice sites in pre-mRNAs.

MATERIALS AND METHODS

Plasmids and Oligonucleotides

Construction of pSP80L, pSP35L and pH3AR has been described (28). pE1a, containing the entire E1a region of adenovirus 2 (ad2), was obtained by inserting the *Tha*I fragment (1.2-5.8 map units) into the *Sma*I site of pSP65 (29). pHBA was constructed by inserting the ad2 *Hind*III/*Bam*HI fragment (73.4-60.1 map units) into pSP62 (29) cleaved with the same enzymes. pXBC was constructed by subcloning an adenovirus 12 (ad12) *Xho*I/*Bam*HI fragment [65-58 map units; obtained from a plasmid (30) generously provided by Dr. W. Doerfler, Cologne] into pSP64 (29) cleaved with *Sal*I and *Bam*HI. pXBL, containing the entire E2a region of ad2 was constructed from sequences extending from the *Bss*HII site at 75.4, near the early cap site, to the *Bam*HI site at 60.1 map units. This fragment was inserted into pSP64 at the *Xba*I and *Bam*HI sites after ligating the *Bss*HII cut end to *Xba*I linkers (31).

Oligodeoxyribonucleotides were synthesized (32) with the generous assistance of Dr. J.L. Fox and B. Powell (Abbott Laboratories). Three oligonucleotides complementary to transcripts of the E2a region of ad2 were prepared: 67.1 (nucleotides 23,920-23,941, ref. 33), 67.2 (24,045-24,077) and 69.3 (24,667-24,686).

RNA Synthesis and Partial Nuclease Digestion

RNA was transcribed in vitro with SP6 polymerase from linearized plasmid templates as described (28,29). Transcripts are designated according to the plasmid template and the restriction endonuclease cleaving at the runoff site (28). DNA templates and transcription products are described in Table 1. Transcripts were end-labeled with ³²P-cytidine bisphosphate (New England Nuclear, 2900 Ci/mmol) and purified by electrophoresis exactly as described (28). Nuclease digestions were carried out for 5 min at 37°C in either

high salt buffers containing 0.2 M NaCl as previously described (28,34) or (for some T1 ribonuclease digestions and all T2 ribonuclease digestions) in low salt buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 2 mM MgCl₂). Units of enzyme activity cited are nominal units of the manufacturer (Pharmacia P-L Biochemicals). Duplicate samples of digested transcripts were analyzed on 8% and 12% acrylamide urea gels run for different periods of time in order to resolve sequences distant from the 3' end. Each run included minus nuclease controls to test for the presence of contaminating nuclease activity. In all experiments with end-labeled RNAs these control lanes were essentially free of degraded RNA. Each gel also included at least 3 different markers: an alkaline digest, and T1 and U2 ribonuclease digestions carried out under denaturing conditions as described previously (28). These markers facilitated the alignment of cleavage sites with the known nucleotide sequence (33). RNA secondary structure was modeled with an interactive computer program, RNAFLD (35), provided by Dr. M. Zuker (National Research Council of Canada, Ottawa), using free energy values given previously (15).

Primer Extension Mapping

Primer extension mapping of nuclease cleavage sites was carried out using 5' end-labeled oligodeoxyribonucleotides as primers. Primers were annealed to RNA in 4 to 8 fold molar excess. Annealing was accomplished by heating for 5 min at 90°C followed by slow cooling in steps to room temperature over the period of 1 h in 50 mM Tris-HCl, pH 8.3, 60 mM NaCl, 10 mM dithiothreitol. A volume of 4 µl contained 0.05–0.1 pmoles RNA. Primer extension was carried out in 50 mM Tris-HCl, pH 8.3, 60 mM NaCl (Figs. 3A,B) or 80 mM NaCl (Fig. 3C), 6 mM MgCl₂ (Figs. 3A,B) or 4 mM MgCl₂ (Fig. 3C), 140 mM KCl, 10 mM dithiothreitol and 0.2 mM each of dATP, dGTP, dCTP and dTTP (36) with 4 units of reverse transcriptase (Life Sciences) in a total volume of 10 µl. Dideoxyribonucleotides were used to obtain parallel sequencing ladders from undigested RNA templates.

RESULTS

Accessibility of 3' Splice Sites to Partial Nuclease Digestion

Transcripts spanning 3' splice sites in ad2 and ad12 (Table 1) were labeled at their 3' ends and analyzed by partial nuclease digestion in order to determine the susceptibility of splice site proximal sequences to single-strand specific nucleases. In each transcript prominent, very intense cut sites are evident at low concentrations of T1, T2 and/or S1 nucleases within a region extending from 10 nucleotides downstream to 50 nucleotides

Table 1: In vitro Transcripts.

<u>Region</u>	<u>Template/runoff site</u>	<u>Length</u>	<u>Distance from 3' end to splice site⁺</u>
E1a (Ad2)	pE1a/XbaI	902	112 (3.4)
L3 (Ad2)	pH3AR/AvaI	601	83 (52.3)
E2a (Ad12)	pXBC/DdeI	304	59 (67.0)
E2a (Ad2)	pSPP80/XhoI	542	163 (67.0)
E2a (Ad2)	pSPP80/AvaI	429	50 (67.0)
E2a (Ad2)	pSPP35/EcoRV	783	72 (69.0)
E2a (Ad2)	pXBL/HaeIII	166	48 (75.2)
E2a (Ad2)	pHBA/DraI	3922	1596 (67.0, 68.8, 69.0)

⁺Positions of splice sites probed in map units are shown in parentheses (33).

upstream of the splice site. In transcripts of the E2a 67 map unit splice sites of ad2 and ad12 the sites of most intense cutting with S1 nuclease and T2 ribonuclease are located within 10 nucleotides of the 3' splice site (Fig. 1, lanes 1 and 10, S1 nuclease results for ad2 and T2 ribonuclease results for ad12 not shown). At the other three sites examined, in transcripts of the E1a, L3 and E2a 69 map unit regions, the most intense cutting at low enzyme concentrations occurs near the branch point consensus sequences indicated in Fig. 1 (lanes 3, 6, 13, 15 and 16). Autoradiographs shown in Fig. 1 provide relatively low resolution analysis of the splice site region on 12% polyacrylamide gels in order to display cutting upstream and downstream of the splice site. The exact positions of major cut sites near the splice site and branch point sequences, summarized in Fig. 2, were determined by parallel, higher resolution runs of identical samples on 8% polyacrylamide gels. All of the bands apparent in Fig. 1 represent digestion with the indicated nucleases as judged by comparison with minus enzyme controls (not shown; cf. ref. 28). While other strong cleavage sites are seen at low enzyme concentrations, the splice site or branch point proximal cuts represent preferred cleavage sites within a region of 150-250 nucleotides as shown in Fig. 1.

In addition to intense cut sites near the branch points in the E1a, L3 and 69 map unit E2a regions, moderately strong cleavage sites are observed at the branch point consensus sequence in the 67 map unit E2a sites of both ad2 and ad12 (Fig. 1, lanes 1-11 and Fig. 2). Cut sites are also observed

Figure 1. Partial digestion of 3' splice site sequences with single-strand specific nucleases. 3' end-labeled transcripts were partially digested with 2 or 3 single-strand specific nucleases. Arrows mark position of 3' splice sites (3' ss) and putative branch point sequences (*). Numbers indicate positions relative to 3' splice site (-1) where negative numbers represent introns and positive numbers exons. Lanes 1-2 pSPP80/XhoI transcripts cleaved with 0.0002 units (u) T2 ribonuclease and 0.007 u T1, in low salt buffer; lanes 3-8 pSPP35/EcoRV cleaved with 0.05 u and 0.25 u S1 nuclease, 0.0001 u, 0.0005 u and 0.001 u T2 ribonuclease, and 0.0025 u T1 ribonuclease; lanes 9-11 pXBC/DdeI cleaved with 0.1 u and 0.05 u S1 and 0.005 u T1; lanes 12-15 pE1a/XbaI cleaved with 0.0005 u and 0.0001 u T2 and 0.01 u and 0.0025 T1; and lanes 16-17 pH3AR/AvaI cleaved with 0.05 u S1 and 0.004 u T1. Reactions shown in lanes 1 and 2 were carried out in the presence of 50 ng 67.1 oligomer in addition to 800 ng tRNA, present in all reactions. Under these conditions the oligomer had no effect on RNA digestion.

Figure 2. Location of nuclease cleavage sites near 3' splice sites in adenovirus RNA. The position of strong cleavage sites observed at low concentrations of enzymes are indicated for S1 nuclease (↑), T2 ribonuclease (↓) and T1 ribonuclease (▽). The size of the arrows represents approximate relative intensities of the corresponding bands. Nucleotides shown in boxes represent either actual or possible branch point sites. The latter are the best fit found to the consensus matrix found by Keller and Noon in mammalian sequences (12) and the percent fits shown are calculated as described (12). The 3' splice site is indicated by gap at right. Positions shown are accurate to at least ± 1 nucleotide.

directly at the branch point consensus sequence at the E1a site (Fig. 1, lane 13, and Fig. 2) as well as the very intense cleavage sites 7-13 nucleotides away (Fig. 1, lane 13 and Fig. 2). At higher concentrations of nuclease many other bands are observed, some of which are as intense as those representing sites cleaved preferentially at lower enzyme concentrations (compare lane 3 to lane 4 and lane 12 to 13 in Fig. 1). Since such bands apparently represent sites of secondary cleavage which become accessible following nicking at primary cut sites (28,34), they are not included in Fig. 2. Lane 11 of Fig. 1 shows a number of moderately intense cut sites which are relatively faint at lower concentrations of T1 ribonuclease. S1 nuclease and T2 ribonuclease cut sites show a high degree of overlap since both of these enzymes cut single-stranded regions of RNA relatively nonspecifically with respect to base sequence. T1 ribonuclease cleaves less frequently since it is specific for guanosine residues in single-stranded regions of RNA (34,37).

Figure 3. Splice site proximal nuclease partial digestion sites in a 4 kb E2a transcript. Cleavage sites near 67 map units (A) and 69 map units (B and C) splice sites were mapped in pHBA/DraI using end-labeled primers 67.1 and 69.3 respectively. One pmole of RNA was digested as follows. Panels A and B: lane 1 control, no nuclease; lane 2 0.17 u V1 ribonuclease; lanes 3-5 0.0001 u, 0.0005 u and 0.0025 u T2 ribonuclease; lanes 6-8 0.005 u, 0.05 u and 0.5 u T1 ribonuclease (low salt buffer). Panel C: lanes 1-3 0.5 u, 0.05 u and 0.005 u T1 ribonuclease (low salt buffer); lanes 4-6 0.0025 u, 0.0005 u and 0.0001 u T2 ribonuclease; lane 7 0.17 u V1 ribonuclease; lane 8 control, no nuclease. V1 ribonuclease is specific for double-stranded RNA (34).

While the location of the E1a branch point site in the nearly identical (99% homologous) transcript of adenovirus 5 and the E2a 67 map unit site in ad2 have been determined (11,38), the locations of the branch points at the other three sites given here are based on sequence analysis. The consensus sequences indicated in Figs. 1 and 2 represent best fits within a 40 nucleotide region (-15 to -55) to both the consensus sequence of Keller and Noon (12) and a smaller sample of experimentally determined sequences (11). Four of the five consensus sequences shown fall within a region (-25 to -30) characteristic of most known branch points (11). Although the branch point consensus sequence for L3 hexon mRNA splicing lies slightly further upstream than the most distant known branch point (4), this sequence represents a substantially better fit (89%) to the consensus sequence (12) than another candidate site closer to the splice site (58%).

Since partial digestion studies shown in Figs. 1 and 2 were carried out with transcripts truncated only 59-163 nucleotides downstream of the 3' splice sites (Table 1), it is important to ascertain whether the choice of runoff site affects the secondary structure immediately adjacent to the

splice site. In order to determine if local secondary structure within relatively short transcripts pSPP80/XhoI and pSPP35/EcoRV is displaced in the larger E2a pre-mRNA, a much longer overlapping RNA, pHBA/DraI (Table 1) which spans most of the E2a region was examined by partial nuclease digestion. Primer extension (36) of 2 oligomers (67.1 and 67.2) which hybridize 148 and 12 nucleotides, respectively, downstream of the 67 map unit splice site and one oligomer (69.3) which hybridizes 105 nucleotides downstream of the 69 map unit site was used to map cleavage sites near these 3' splice sites in the 3.9 kb pHBA/DraI transcript. Results of primer extension mapping are easily compared to results from 3' end-labeled transcript since the 5' labeled end of the complementary primer provides a reference point analogous to that of the 3' end label on the transcript. Most of the bands evident in Fig. 1, lanes 1 and 2, are seen in lanes 3 and 7, respectively, of Fig. 3A. Major cleavage sites at +82, +51, +37, -6 and -8 are apparent in both Figs. 1 and 3A at the 67 map unit site in the E2a region of ad2. Cut sites between positions +10 and -50 were accurately mapped by primer extension of the 67.2 oligomer (data not shown). Cut sites near +43, -26 and -80 are apparent at the 69 map unit region in both 3' end-labeled transcripts (Fig. 1, lanes 7 and 8) and in primer extension analysis of pHBA/DraI (Fig. 3B, lanes 3 and 7). T1 and T2 ribonuclease cut sites apparent in the shorter transcript near the splice site at +1 and +3 are, however, much less apparent in the longer transcript. This may reflect either differences in local splice site conformation and accessibility or variations in the level of digestion between the two experiments. The close overall match of the digestion patterns at both the 67 and 69 map unit sites shown in Figs. 1 and 3, however, indicates that the conformation of these splice sites is quite similar in the short and long RNA molecules overlapping these two regions. The numerous additional bands observed in reverse transcription of undigested RNA (Figs. 3A and B, lane 1) represent either premature termination or lack of intact RNA template. Since the relative intensity of these bands is very sensitive to variations in monovalent and divalent cations and relatively insensitive to conditions of preparation and incubation of the RNA it is likely that most of these bands represent premature termination.

Accessibility of 5' Splice Sites to Nuclease Digestion

Primer extension mapping of the 69 map unit site shown in Figs. 3B and 3C also provides information on the accessibility of sequences at the 5' splice site. Previously we suggested on the basis of analysis of the pSPP35/EcoRV transcript and secondary structure modeling that the 5' splice

site might fall at the boundary of a large RNA folding domain, and may therefore also be relatively exposed to nuclease attack (28). As shown in Figs. 3B and 3C the GG dinucleotide spanning the 5' splice site is cut strongly with T1 ribonuclease (Fig. 3B, lane 7; Fig. 3C, lanes 2 and 3) while other nucleotides within the 5' splice site consensus sequence are cut preferentially with ribonuclease T2 (e.g., Fig. 3C, lane 5). The nuclease sensitivity of the 5' splice site of exon 1 at 75 map units was also examined (Fig. 4). At this site sequences directly downstream of the splice site are also preferentially cleaved with T1 ribonuclease (lanes 2-4) and to a lesser extent with S1 nuclease (lanes 5, 6). These results from 5' splice sites in the E2a region suggest that 5' as well as 3' branch sites may be present in single-stranded regions in pre-mRNA. However, nuclease cleavage of regions near these two 5' splice sites does not appear as selective as that observed at the 3' splice sites examined.

Secondary Structure of E2a Transcripts from Adenovirus Serotypes 2 and 12

It was of interest to compare the secondary structure of two homologous splice sites in distantly related human adenoviruses (39) in order to determine whether an exceptionally large hairpin structure previously demonstrated to be present at the 67 map unit site in isolated ad2 transcripts (28) is present in ad12. Although the ad2 (33) and ad12 (40) sequences are identical at 60% of their positions in this region, very few of the 42 base pairs found in the ad2 transcript (28) are possible in ad12, as indicated in the sequence alignment and ad2 secondary structure shown (Fig. 5A). Since there are no reciprocal substitutions of nucleotides such as commonly found in homologous rRNA molecules, which typically maintain similar secondary structures while diverging in sequence (41), it is evident that this specific secondary structure is not conserved. On the other hand, some general similarities can be seen in the nuclease cleavage patterns and secondary structure models obtained for ad2 and ad12 (Fig. 5). First, at both sites the strongest cleavage occurs near the splice site itself, with weaker cutting at the branch point consensus sequence. Second, in both sequences the strong splice-site proximal cutting appears to correspond to cutting in the apical loop of a large hairpin structure approximately centered on the splice site. Third, both hairpins, although quite different in overall structure, are stabilized primarily by base-pairing of GC-rich regions at the base of the hairpin stem about 50 nucleotides from the splice site. Similar complementary regions are also found at these positions in adenoviruses 4 and 7 which represent 2 additional subgenera (42 and

Figure 4. Partial digestion of sequences at the 5' splice site at 75 map units in the E2a region. 3' end-labeled pXBL/HaeIII was digested with 0.05 u, 0.017 u and 0.005 u T1 ribonuclease (lanes 2-4) and 0.25 and 0.05 u S1 nuclease (lanes 5 and 6) in high salt buffers for 5 min at 37°C. Markers in lane 1 are from a T1 ribonuclease digestion carried out under partially denaturing conditions as described (28).

Figure 5. Comparison of sequence and structure at 3' splice sites at 67 map units in adenovirus types 2 and 12. A. Alignment of the sequence of ad2 and ad12 shown relative to folded hairpin in ad2 (28). Bases are shown for ad12 only where different. Dashes mark gaps in alignment or base pairing. Asterisks indicate regions where 2 or more consecutive base pairs are conserved in ad12 sequence. Bold arrow marks splice site and small arrows mark probable branch point residues. Vertical dashes indicate paired bases in ad2 sequence while dotted lines show base pairs possible in only the ad12 sequence. B. Model secondary structure obtained by folding pXBC/DdeI sequence using RNAFLD program (35). Sites of cleavage with S1 nuclease

8459

Figure 6. Titration of hairpin stem at 67 map units by complementary oligodeoxyribonucleotide. Increasing concentrations of 67.2 33-mer were used to disrupt secondary structure as described in text. 67.2 oligomer and 3' end-labeled pSP80/AvaI RNA were heated at 90°C for 3 min in water then transferred to 65°C. Low salt buffer was added and after 10 min the hybrids were slowly cooled to room temperature. Molar ratios of DNA to RNA are indicated at top of figure. Lanes 1 and 10 contain markers obtained by T1 digestion of partially denatured RNA and alkaline hydrolysis, respectively; hybrids and mock-annealed RNA were digested with 0.1 u T1 ribonuclease (lanes 2, 4, 6 and 8), with 0.01 u T1 ribonuclease (lanes 3, 5, 7 and 9) and with 0.005 u T2 ribonuclease (lanes 11-14).

unpublished results of G.R. Kitchingman and C.O. Quinn). The secondary structure model for the ad12 site shown in Fig. 4B represents a very good fit to the digestion data. Although the stability of the secondary structure in ad12 is less than that found for ad2, the stabilization energy for the structure shown here (-111 kcal/mole) is close to that of the optimal secondary structure (35) for this region (-113 kcal/mole).

In order to characterize further the secondary structure previously described (10,28) for the 67 map unit 3' splice site in ad2 increasing concentrations of a 33 nucleotide oligomer (67.1) complementary to residues 12 to 44 (Fig. 5A) were used to displace intramolecular secondary structure. The 33-mer was annealed to 3' end-labeled pSP80/AvaI transcript in 7:1, 33:1 and 330:1 molar excess over RNA (0.4-20 μ M DNA) and the DNA:RNA hybrids formed were digested with T1 (Fig. 6, lanes 2-7) and T2 (lanes 11-13) ribonucleases. At 33-330 fold molar excess of DNA the hairpin stem opens up as is most clearly seen at the highest concentration (Fig. 6, lanes 2 and 3). The intensity of cleavage with T1 of apical position -8 decreases while cleavage at -19 increases. A shift in the major T2 cut sites from residues -4 to -7 to a site centered on residue -18 is also seen in lane 11. Other positions in the upstream stem also become more nuclease sensitive (Fig. 6, lane 2). Since the 33-mer does not hybridize directly to sequences in the apical loop, this result demonstrates that strong cutting seen in this region in the native structure is directly related to the conformation of the RNA and not to the nucleotide sequence of the site per se and is consistent with unfolding of the secondary structure previously proposed for this region (28).

DISCUSSION

Our results demonstrate that sequences near either the 3' splice site or the upstream branch point consensus sequences of several adenovirus pre-mRNAs are preferentially digested with low concentrations of single-strand specific nucleases. The selectivity of cutting within these regions appears substantial. Bands representing these cleavage sites are the most intense ones observed within a region of at least 150-250 nucleotides, while the size of the target region is approximately 40 nucleotides in length. The sites of strongest cleavage are typically short AU-rich sequences (Fig. 2). At each of the 3' splice sites examined strong or moderately strong cutting was also observed within 3 nucleotides of the branch point sequence. Two 5' splice sites were also accessible to nuclease digestion. The sensitivity of sites to low levels of nuclease almost certainly reflects the local conformation of the RNA molecule since denaturing conditions (28) and titration of RNA secondary structure with competing oligodeoxyribonucleotides (Fig. 6) drastically alter the pattern of nuclease cleavage.

Partial nuclease digestion patterns also provide information concerning the presence of specific elements of RNA secondary structure. Tentative

models consistent with nuclease digestion data have been obtained by computer analysis of sequences at 3 of the splice sites studied as shown in Fig. 4 above and a previous report (28). Consideration of these model structures suggests that local secondary structure may play a general rather than specific role in mRNA splicing since different splice sites and branch points in vitro assume a variety of diverse conformations. For example, at the 69 map units site, the branch point occurs near the apical end of a short hairpin stem entirely within the intron (-36 to -14). In contrast, at the 67 map units site in ad2, the branch point falls within a short internal loop near the middle of a larger stem structure (28). Moreover, unlike ribosomal structural RNAs (41), very little specific base pairing is conserved at 67 map unit sites in adenoviruses representing two distinct subgenera (39). Certain general features relating to the nuclease accessibility of splice-site proximal residues in these transcripts are conserved, however, as described above. It is worth noting that models for RNA secondary structure are by themselves insufficient to predict the relative susceptibility to nuclease cleavage of residues within a given region. Partial digestion with structure-specific nucleases thus provides information concerning the accessibility of specific regions not obtainable from secondary structure models alone.

Branch point sequences and 3' splice junctions have been implicated in a number of early steps in RNA splicing (7,8,43), including the binding of snRNP complexes (7,8) and the stepwise assembly of active spliceosomes (44-46). Since certain of these steps take place very quickly upon the addition of isolated pre-mRNA substrates to the splicing system in vitro, and precede the ATP-dependent appearance of covalently altered intermediates (8,39,43-46), it seems plausible that the higher order structure of the RNA substrate plays a significant role in mediating RNA splicing. The accessible single-stranded regions of pre-mRNA molecules described here might, for example, serve as entry points for diffusible RNA-binding splicing factors, which upon initial binding make relatively non-specific contacts with the RNA before forming more extensive ones characteristic of site-specific binding.

Two questions are important for evaluating the relevance of our observations to the mechanism of mRNA splice site selection in vivo. First, is the higher order structure of isolated RNA molecules probed here related to RNA structure in the nucleus? Second, does this structure in fact influence the outcome of splicing? Since pre-mRNA molecules in vivo are extensively bound with abundant hnRNP proteins (47-50), it is likely that the RNA conformation

in the nucleus differs from that of the isolated RNA transcripts studied here in vitro. The in vivo structure, however, probably represents a balance between potentially stable intramolecular interactions and intermolecular ones involving both bound proteins and snRNP complexes in active processing complexes (6-8,44-46). Recent work has shown that the introduction of large inverted repeats into pre-mRNA influences the outcome of splicing (51). These experiments appear to demonstrate that both long-range and relatively local secondary structure interactions between perfectly complementary regions in vivo can block splicing at normally functional sites (51). Local features of RNA conformation in naturally occurring sequences near conserved splicing signals may therefore also play a role in modulating splice site recognition and utilization.

ACKNOWLEDGEMENTS

We thank Dr. W. Doerfler for the plasmid containing the ad12 fragment, Dr. G.R. Kitchingman for unpublished sequence data, Dr. J.L. Fox and B. Powell for assistance with oligodeoxyribonucleotide synthesis, Dr. G.L. Waring and Dr. T. Maniatis for their comments, and B. DeNoyer for her expert assistance in preparing this manuscript. D.J. Van Den Berg constructed the plasmid pEla. This work was supported by awards from the NSF (PCM-8302825) and the NIH Biomedical Research Support Grant to Marquette University (RR07196).

*To whom correspondence should be addressed

+Present address: Pharmacia P-L Biochemicals, Inc., 2202 North Bartlett Avenue, Milwaukee, WI 53202, USA

REFERENCES

1. Schneuwly, S., Kuroiwa, A., Baumgartner, P. and Gehring, W.J. (1986) EMBO J. 5, 733-740.
2. Yamada, R., Liau, G., Mudryj, M., Obici, S. and deCrombrughe, B. (1984) Nature 310, 333-337.
3. Mount, S.M. (1982) Nucl. Acids Res. 10, 459-472.
4. Ruskin, B., Krainer, A., Maniatis, T. and Green, M.R. (1984) Cell 38, 317-331.
5. Padgett, R.A., Konarska, M.M., Grabowski, P.J., Hardy, S.F. and Sharp, P.A. (1984) Science 225, 898-903.
6. Kramer, A., Keller, W., Appel, B. and Luhrmann, R. (1984) Cell 38, 299-307.
7. Black, D.L., Chabot, B. and Steitz, J.A. (1985) Cell 42, 737-750.
8. Chabot, B., Black, D.L., LeMaster, D.M. and Steitz, J.A. (1985) Science 230, 1344-1354.

9. Wieringa, B., Meyer, F., Reiser, J. and Weissmann, C. (1983) Nature 301, 38-43.
10. Munroe, S.H. (1983) Nucl. Acids Res. 11, 8891-8900.
11. Reed, R. and Maniatis, T. (1985) Cell 41, 95-105.
12. Keller, E.B. and Noon, W. (1984) Proc. Natl. Acad. Sci. 81, 7417-7420.
13. Baldi, M.I., Mattoccia, E. and Tochini-Valentini, G.P. (1983) Cell 35, 109-115.
14. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Cell 31, 147-157.
15. Cech, T.R., Tanner, N.K., Tinoco, I., Weir, B.R., Zuker, M. and Perlman, P.S. (1983) Proc. Natl. Acad. Sci. 80, 3903-3907.
16. Waring, R.B., Ray, J.A., Edwards, S.W., Scazzocchio, C. and Davies, R.W. (1985) Cell 40, 371-380.
17. Peebles, C.L., Perlman, P.S., Mecklenburg, K.L., Petrillo, M.L., Tabor, J.H., Jarrell, K.A. and Cheng, H.-L. (1986) Cell 44, 213-223.
18. Van der Veen, R., Arnberg, A.C., van der Horst, G., Bonen, L., Tabak, H.F. and Grivell, L.A. (1986) Cell 44, 225-234.
19. Schmelzer, C. and Schweyen, R.J. (1986) Cell 46, 557-565.
20. Michel, F. and DuJon, B. (1983) EMBO J. 2, 33-38.
21. Cech, T.R. (1986) Cell 44, 207-210.
22. Pikielny, C.W., Teem, J.L. and Rosbash, M. (1983) Cell 34, 395-403.
23. Konarska, M.M., Grabowski, P.J., Padgett, R.A. and Sharp, P.A. (1985) Nature 313, 552-557.
24. Tonegawa, S., Maxam, A.M., Tizard, R., Bernard, O. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. 75, 1485-1489.
25. Trapnell, B.C., Tolstoshev, P. and Crystal, R.G. (1980) Nucl. Acids Res. 8, 3659-3672.
26. Chu, G. and Sharp, P.A. (1981) Nature 289, 378-382.
27. Jacquier, A., Rodriguez, J.R. and Rosbash, M. (1985) Cell 43, 423-430.
28. Munroe, S.H. (1984) Nucl. Acids Res. 12, 8437-8456.
29. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) Nucl. Acids Res. 12, 7035-7056.
30. Vogel, S., Brotz, M., Kruczek, I., Neuman, R., Eick, D., Winterhoff, U. and Doerfler, W. (1981) Gene 15, 273-278.
31. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
32. Caruthers, M.H. (1983) In Weisman, S.M. (ed.), Methods of DNA and RNA Sequencing, Praeger Publishers, New York, pp. 1-68.
33. Roberts, R.J., Akusjarvi, G., Alestrom, P., Gelinas, R.E., Gingeras, T.R., Sciaky, D. and Pettersson, U. (1986) In Doerfler, W. (ed.), Adenovirus DNA, Martinus Nijhoff Publishing, Boston, pp. 1-51.
34. 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 Chirikjian, J.D. and Papas, T.S. (eds.), Gene Amplification and Analysis, Elsevier/North Holland, New York, Vol. 2, pp. 268-298.
35. Zuker, M. and Stiegler, P. (1981) Nucl. Acids Res. 9, 5125-5140.
36. Shelness, G.S. and Williams, D.L. (1985) J. Biol. Chem. 260, 8637-8646.
37. Cary, C.P.H. and Vournakis, J.N. (1984) Nucl. Acids Res. 12, 6763-6778.
38. Keohavong, P., Gattoni, R., Schmitt, P. and Stevenin, J. (1986) Nucl. Acids Res. 14, 5207-5227.
39. Green, M., Mackey, J.K., Wold, W.S.M. and Rigden, P. (1979) Virology 93, 481-492.
40. Kruijer, W., van Schaik, F.M.A., Speijer, J.G. and Sussenbach, J.S. (1983) J. Virology 128, 140-153.

41. Woese, C.R., Gutell, R., Gupta, R. and Noller, H.F. (1983) Microbiol. Rev. 47, 621-669.
42. Kitchingman, G.R. (1985) Virology 146, 90-101.
43. Ruskin, B. and Green, M.R. (1985) Cell 43, 131-142.
44. Brody, E. and Abelson, J. (1985) Science 228, 963-967.
45. Friendewey, D. and Keller, W. (1985) Cell 42, 355-367.
46. Grabowski, P.J., Seiler, S.R. and Sharp, P.A. (1985) Cell 42, 345-353.
47. Beyer, A.L., Christensen, M.E., Walker, B.W. and LeStourgeon, W.M. (1977) Cell 11, 127-138.
48. Mayrand, S., Setyono, B., Greenberg, J.R. and Pederson, T. (1981) J. Cell Biol. 90, 380-384.
49. Jacob, M., Devilliers, J.-P., Fuchs, J.-P., Gallinaro, H., Gattoni, R., Judes, C. and Stevenin, J. (1981) In Busch, H. (ed.), The Cell Nucleus, Academic Press, New York, Vol. 8, pp. 194-246.
50. Munroe, S.H. (1982) J. Mol. Biol. 162, 585-606.
51. Solnick, D. (1985) Cell 43, 667-676.