Virus Deletion Mutants That Affect a 3' Splice Site in the E3 Transcription Unit of Adenovirus 2

BHEEM M. BHAT, HELEN A. BRADY, AND WILLIAM S. M. WOLD*

Institute for Molecular Virology, St. Louis University Medical School, St. Louis, Missouri 63110

Received 11 April 1985/Accepted 23 May 1985

Five viable virus mutants were constructed with deletions near a 3' splice site located at nucleotide 2157 in the E3 transcription unit of adenovirus 2. The mutants were examined for splicing activity at the 2157 3' splice site in vivo by nuclease-gel analysis of steady-state cytoplasmic mRNA. Splicing was not prevented by an exon deletion (*d719*) that leaves 16 5'-proximal exon nucleotides intact or by intron deletions that leave 34 (*d717*, *d712*) or 18 (*d716*) 3'-proximal intron nucleotides intact. The sequences deleted in one of these intron mutants (*d716*) include the putative branchpoint site used in lariat formation during splicing. Thus, a surrogate branchpoint site apparently can be used for splicing. Another intron mutant (*d714*) has a deletion that leaves 15 3'-proximal intron nucleotides intact; remarkably, this deletion virtually abolished splicing, even though the deletion is only 3 nucleotides closer to the splice site than is the deletion in *d716* which splices normally. The three nucleotides deleted in *d714* that are retained by *d716* are the sequence TGT. The TGT sequence is located on the 5' boundary of the pyrimidine-rich region upstream of the nucleotide 2157 3' splice site. Such pyrimidine-rich regions are ubiquitous at 3' splice sites. Most likely, the TGT is required for splicing at the nucleotide 2157 3' splice site. The TGT may be important because of its specific sequence or because it forms the 5' boundary of the pyrimidine-rich region.

The *cis*-acting signals required for splicing of pre-mRNA in higher eucaryotes have not been fully defined. Splice sites conform to consensus sequences, (CA)AG↓GTRAGT and (Y)_nNYAG↓G, at the 5' and 3' splice sites, respectively (7, 42); this conservation suggests that the consensus sequences are important for splicing. Studies on thalassemias as well as on mutants constructed in vitro have established that sequences at the splice sites are crucial to splicing. Deletion (17, 30, 38, 39, 43, 60) or base substitution (40, 41, 56) of the ↓GT dinucleotide prevent 5' splicing activity. Mutations in the intron-flanking 5' splice consensus sequences eliminate (16, 38, 52) or reduce the efficiency (56) of the splice. In contrast, purine transitions at positions 3 or 4 of the intron-flanking sequence of a cloned rabbit β-globin 5' splice site have no effect on splicing (62). Regarding the 3' splice site, deletion (45) or base substitution (3) of the AG↓ eliminates splicing, as does a deletion into the pyrimidine-rich region that leaves 13 3'-proximal nucleotides (nt) remaining (61). Excluding the splice consensus sequences, other specific intron sequences are not obligatory for splicing, as indicated by studies with deletion mutants (30, 55, 57, 61) and with synthetic splice sites (49). Furthermore, chimeric splices can occur, in which the 5' site is from one gene and the 3' site is from another (11, 25, 29).

Splicing mutations often activate cryptic 5' (17, 22, 39, 41, 56, 62) or 3' (29, 61) splice sites that are not normally used. Cryptic 5' (44) or 3' (8, 20) splice sites can also be activated by mutations that mimic the consensus sequence. In one case, an intron base substitution apparently activated a cryptic 3' splice site located at a distant site (14).

A third intron sequence, in addition to the splice junctions, is required for splicing. Studies on splicing in vitro in adenovirus (Ad) (46) and in human β-globin (51) have indicated that splicing occurs by a lariat mechanism, whereby the GT at the 5' end of the intron is covalently

joined through a 2'-5' phosphodiester bond to an A residue near the 3' splice site. The lariat mechanism also applies to pre-mRNA splicing in vivo in rabbit β-globin (64) and in yeasts (15, 50). The branchpoint site of the lariat has only been observed at A residues in vitro (46, 51). In vivo, branching in heterogeneous nuclear RNA, which presumably reflects lariat formation, occurs largely at A residues (58). Inspection of several genes suggests that the A residue at the branchpoint site lies within a consensus sequence, YNYRAY, located ~20 to 40 nt upstream of the 3' splice site (51, 64). Similar consensus sequences, namely AYTAAT in sea urchins and CTGAC in rats and mice, have been identified by a computer search (28). These consensus sequences are probably the analogs of the yeast TACTAAC box (33, 48), which functions as the splicing lariat branchpoint site (15, 50). The TACTAAC box is strictly required for splicing in yeast pre-mRNA (34). The location of the branchpoint site in regions of metazoan introns that can be deleted without preventing splicing would appear to present a paradox. However, it has been suggested that alternate branchpoint sites can be used when the natural site is missing (46, 51).

In this communication, we describe the genetic analysis of sequences that are required in *cis* for splicing at a 3' splice site. The analysis is done on mRNA synthesized in vivo by viable Ad mutants. Since our mutations are in a complete virus genome, they are in a more natural context than are mutations in expression vectors as reported by some other authors. The 3' splice site that we have analyzed is located at nt 2157 in the E3 transcription unit of Ad2 (see Fig. 1). This 3' splice site, which had been observed previously at an analogous position in Ad5 (12), is used in the formation of E3 mRNA *f* (see Fig. 1). We find that splicing is not prevented by a deletion that leaves 18 3' proximal intron nt intact. The sequences that could be deleted without preventing splicing presumably include the branchpoint site involved in lariat formation. In contrast, splicing is abolished by a deletion

* Corresponding author.

FIG. 1. Schematic representation of the E3 transcription unit of Ad2 and of deletion mutants *dl717*, *dl716*, *dl714*, *dl712*, and *dl719*. The numbers represent nucleotide numbers in the transcription unit (13, 26). Nucleotide +1 is the first major transcription initiation site (4, 12). Open reading frames (ORF) 1 (F1), 2 (F2), and 3 (F3) are shown. Black bars indicate proven proteins, and stippled bars indicate proteins that we propose to exist (13). Proteins pVIII and fiber are late structural proteins coded by the L4 and fiber mRNAs, respectively. Arrows (a through j) represent the spliced structures of the E3A and E3B 3' coterminal families of mRNAs as determined by electron microscopy (10, 31). The RNA 3' end and splice sites have been determined by sequence analysis of cDNA clones of E3 mRNAs (1, 2, 53) or by nuclease-gel analysis (5, 6, 12). Evidence in support of this schematic is discussed by Cladaras and Wold (13) and Cladaras et al. (12).

that leaves 15 3'-proximal nt intact. The splicing defect is probably caused by deletion of a TGT triplet which is located in the 5' boundary of the pyrimidine-rich region at the nt 2157 3' splice site.

MATERIALS AND METHODS

Cells, viruses, and RNA extraction. Procedures for the maintenance of KB cells, Ad preparation, and plaque assays have been described previously (24). For RNA preparation, suspension cultures of KB cells were infected with mutants and maintained in cycloheximide (25 μ g/ml) from 3 to 9 h postinfection (63). Cells were lysed (63), and cytoplasmic RNA was extracted (18).

Isolation of plasmid and virus mutants. Mutations were made in the Ad2 *EcoRI* D fragment cloned in pBR322. For mutants *dl712*, *dl714*, and *dl716*, we used a mutant of pBR322, isolated in our laboratory, which has the sequences between the *ClaI* and *PvuII* sites (54) deleted. For *dl719* we used wild-type pBR322. Deletions were made as described previously (35), by cleaving at the *BclI* (nt 1825), *StuI* (2076), *XmnI* (2106) or *XhoI* (2180) site, deleting with BAL 31 exonuclease, and recircularizing. With *dl712*, a *BamHI* linker, d(CGGGATCCCG) (New England BioLabs, Inc.) was ligated to the plasmid before recircularizing. The deletions were sequenced by the procedure of Maxam and Gilbert (36).

To build the plasmid mutations into the viral genome (Fig. 2), plasmids were cleaved with *EcoRI*, digested with bacterial alkaline phosphatase, and ligated to *EcoRI*-cleaved Ad5 terminal protein-DNA complex. The terminal protein-DNA complex was prepared as described previously (9). The ligation mixture was transfected into KB cells (23) with a glycerol shock at 5 h posttransfection (19). Well-separated plaques were picked and expanded into virus stocks. Virus stocks were screened for the expected Ad5-Ad2 recombinant structure and for the expected deletion by isolating virus 32 P-labeled DNA from the Hirt supernatant (9), cleaving it with *HindIII*, electrophoresing it on 0.8% agarose gels, and autoradiographing the dried gel. *Rec700* was made in the same manner as the virus deletion mutants.

RNAse analysis of cytoplasmic RNA. RNA samples were analyzed by the nuclease-gel procedure (5) with 32 P-labeled RNA probes prepared by in vitro transcription with the SP6 RNA polymerase (37). To prepare the plasmid substrate, the *SmaI-EcoRI* fragment (nt -39 to 2437; see Fig. 1) was excised from the wild-type *EcoRI*-D-pBR322 clone or from the *dl719 EcoRI*-D-pBR322 clone and was cloned between the *SmaI* and *EcoRI* sites in the polylinker of pSP65 (37). The wild-type clone was used to prepare probes for the analysis of mRNA from cells infected by *rec700* and by the intron deletion mutants. The *dl719* clone was used to prepare a probe for the analysis of *dl719* mRNA.

Synthesis of the 32 P-labeled RNA probes was done as

described by Melton et al. (37) except that the reaction contained 80 μ M [α - 32 P]UTP (New England Nuclear Corp.) as label. RNasin, SP6 polymerase, and pSP65 were obtained from Promega Biotech, and RNase-free DNase was obtained from Worthington Biochemicals Corp.

Hybridizations and RNase digestions were performed essentially as described by Melton et al. (37). Cytoplasmic RNA (amounts are given in figure legends) was mixed with 10^6 Cerenkov cpm of probe and hybridized at 40°C for 12 to 16 h. Samples were digested with 40 μ g of RNase A per ml plus 10 U of RNase T1 per ml (both from Sigma Chemical Co.) at 30°C for 1 h in a buffer containing 0.3 M NaCl (37). Samples were electrophoresed on 6% sequencing gels (13). The gels were exposed without using X-ray-intensifying screens.

RESULTS

Construction of Ad deletion mutants. Deletion mutations were originally formed in the cloned Ad2 *Eco*RI D fragment. *Eco*RI-D is located at genomic map positions 76 through 83, and it represents nt -236 to 2437 in the E3 transcription unit (13, 26). Deletions were made at the *Bcl*II (1825), *Stu*I (2076), *Xmn*I (2106), and *Xho*I (2180) sites. The sequences deleted in five mutants are indicated in Fig. 1.

The plasmid deletions were built into virus (Fig. 2). Briefly, the mutated *Eco*RI D fragment was ligated between the *Eco*RI A and B fragments of Ad5 and then transfected into KB cells. Plaques were screened for the expected deletion by *Hind*III digestion of viral 32 P-labeled DNA. All mutants had the expected *Hind*III fragments, including the fragment containing the deletion (data not shown). One virus that we isolated, designated *rec700*, resembles the mutants in having Ad5 sequences in the *Eco*RI A and B fragments but differs by having wild-type Ad2 sequences in the *Eco*RI D fragment. We used *rec700* as the wild-type virus in this study.

A 3' splice site exists at nt 2157 in the E3 transcription unit of Ad2. In a previous study, we used the nuclease-gel procedure with end-labeled probes to map the 5' and 3' ends and 5' and 3' splice sites of the major E3 mRNAs of Ad5 (12). One of the 3' splice sites, which mapped to nt 2201 in the E3 transcription unit of Ad5, is equivalent in sequence context and in location to a potential 3' splice site at nt 2157 in the E3 transcription unit of Ad2 (see reference 13). We determined that the nt 2157 3' splice site is indeed used in the

FIG. 2. Method used to construct virus mutants. Mutations were made in a plasmid, the Ad2 *Eco*RI D fragment (map positions 76 through 83) cloned in pBR322. After DNA sequencing the mutation, the mutated *Eco*RI D fragment was ligated between the Ad5 *Eco*RI A and B fragments and then transfected into KB cells. Plaques were picked, expanded into stocks, and screened for the expected mutation.

FIG. 3. Nuclease gel analysis of RNA from cells infected by *rec700*. The schematic at the bottom depicts the probes used as well as the nuclease-resistant fragments observed in the analysis. The numbers refer to nucleotide length. Heavy roman type indicates nuclease-resistant RNA fragments, italics indicate probes, and light roman type indicates RNA size markers. The E3-specific mRNAs represented by the nuclease-resistant RNA fragments are indicated at the bottom (refer to Fig. 1). The probes were 32 P-labeled RNAs prepared by in vitro transcription with the SP6 RNA polymerase. The probe substrate was the *rec700* *Sma*I-*Eco*RI fragment, nt -39 to 2437 in the E3 transcription unit (Fig. 1), cloned into pSP65. The *Xmn*I and *Stu*I probes were made by cleaving at those sites, respectively, before transcription. The box at the 5' end of the probe represents 9 nt from pSP6. Cytoplasmic RNA was used in 30- μ g portions. Lane a, markers; lane b, *Xmn*I probe versus *rec700* RNA; lane c, *Xmn*I probe versus 30 μ g of tRNA; lane d, *Xmn*I probe; lane e, *Stu*I probe versus *rec700* RNA; lane f, *Stu*I probe versus 30 μ g of tRNA; lane g, *Stu*I probe.

FIG. 4. Nuclease gel analysis of RNA from cells infected by mutants with deletions on the intron side of the nt 2157 3' splice site. The schematic at the bottom depicts the *XmnI* and *StuI* ^{32}P -labeled RNA probes used, as well as the nuclease-resistant fragments observed in the analysis. Triangles at the 5' ends of the fragment arrows indicate the deletions. Cytoplasmic RNA was used in 30- μg portions. See the legend to Fig. 3 for further details. Shown are the *XmnI* probe (lane a) and the *XmnI* probe versus RNA from *rec700* (lane b), *dl717* (lane c), *dl716* (lane d), *dl714* (lane e), and *dl712* (lane f) and versus 30 μg of tRNA (lane n). Also shown are the *StuI* probe (lane g) and the *StuI* probe versus RNA from *rec700* (lane h), *dl717* (lane i), *dl716* (lane j), *dl714* (lane k), and *dl712* (lane l) and versus 30 μg of tRNA (lane m).

Ad2 E3 transcription unit (Fig. 3). (The data shown in Fig. 3 are for *rec700* mRNA, but the nt 2157 3' splice site was also found in wild-type Ad2 mRNA). The probes were RNAs labeled in vitro with $[\alpha\text{-}^{32}\text{P}]\text{UTP}$, using the SP6 RNA polymerase and the Ad *SmaI*-*EcoRI* fragment (nt -39 to 2437; Fig. 1) cloned into pSP65 (37). One probe (the *XmnI* probe) contains 333 nt of viral sequences, extending from the *EcoRI* (nt 2437) to the *XmnI* (2106) sites, plus 9 nt at the 5' end from pSP65 (Fig. 3, lane d and bottom). The other probe (the *StuI* probe) contains 365 nt of viral sequences extending from the *EcoRI* (nt 2437) to the *StuI* (2076) sites, plus 9 nt of pSP65 (Fig. 3, lane g and bottom).

When the *XmnI* probe was hybridized to *rec700* mRNA, two RNase-resistant bands were observed, of 333 and 287 nt (Fig. 3, lane b). The 333-nt band corresponds to E3 mRNAs that are colinear with the probe (mRNAs b, c, and e in Fig. 1). This band is smaller than the probe (Fig. 3, lane d), because the probe contains 9 nt of pSP65 sequences. The 287-nt band represents RNA that extends from the 2157 3' splice site to the *EcoRI* (nt 2437) site. This would be mRNA f in Fig. 1. As expected, the 287-nt band representing mRNA f was also observed with the *StuI* probe (Fig. 3, lane e). The 365-nt band represents mRNAs b, c, and e which are colinear with the *StuI* probe (Fig. 3, lane e). We conclude that a 3' splice site exists at nt 2157 in the E3 transcription unit of Ad2. This splice site is used by mRNA f (Fig. 1).

Analysis of deletions on the intron side of the nt 2157 3' splice site. Four mutants with deletions on the intron side of the 2157 3' splice site were analyzed (Fig. 4). The location and abundance of the splice was unaffected in steady-state cytoplasmic mRNA by the deletions in *dl717* (Fig. 4, lanes c and i) and *dl716* (lanes d and j). This can be seen by comparing the mutant 287-nt bands with those of *rec700* (lanes b and h). (The lower levels of the 287-nt band in lanes i and j were not observed in other experiments.) The deletions in *dl717* and *dl716* leave 34 and 18, respectively, of 3'-proximal intron nucleotides (Fig. 5).

With *dl717*, a major band of 321-nt was observed with both the *XmnI* (Fig. 4, lane c) and *StuI* (lane i) probes. This band represents colinear mRNA extending from the 3' boundary of the deletion in *dl717* (nt 2122) to the *EcoRI* (2437) site (mRNAs b, c, and e in Fig. 1). Similarly, with *dl716*, a major band of 305 nt was observed with the *XmnI* (Fig. 4, lane d) and *StuI* (lane j) probes; this band extends from the deletion (nt 2138) to the *EcoRI* (2437) site.

In contrast to *dl717* and *dl716*, *dl714* was defective in splicing at the nt 2157 site; i.e., no 287-nt band was observed (Fig. 4, lanes e and k). The 302-nt band observed with *dl714* represents colinear mRNA extending from the deletion (nt 2141) to the *EcoRI* (2437) site. The splicing defect is illustrated dramatically in Fig. 6, in which the nt 2157 3' splice signal obtained from 7.5 to 75 μg of *dl714* mRNA (lanes d through i) is compared with the signal from 7.5 μg of *rec700* mRNA (lane a). A faint band which comigrates with the 287-nt band from *rec700* (Fig. 6, lane a) was observed with 50 to 75 μg of *dl714* mRNA (lanes g through i); this could represent a very weak splice at the nt 2157 site, or it could be merely a background band.

The splicing defect in *dl714* is of particular interest because the deletion leaves 15 3'-proximal intron nt, only 3 nt less than the deletion in *dl716* (Fig. 5). This result could be explained by one of the following alternatives. First, the TGT (nt 2139 through 2141) which is present in *dl716* but not in *dl714* is essential for splicing. Second, the sequences at nt 2070 through 2091 (on the 5' side of the deletion in *dl714*) which are present in *dl716* but not in *dl714* are essential for

FIG. 6. Nuclease gel analysis of increasing amounts of the RNA from cells infected by *dI714*. The *StuI* probe from *rec700* DNA was used. See the legends to Fig. 3 and 4 for details. Cytoplasmic RNA was used. Shown are the *StuI* probe versus 7.5 μg of *rec700* RNA (lane a), *StuI* probe alone (lane b), and markers (lane c). Also shown are the *StuI* probe versus 7.5 μg (lane d), 15 μg (lane e), 30 μg (lane f), 45 μg (lane g), 60 μg (lane h), or 75 μg (lane i) of *dI714* RNA and the *StuI* probe versus 30 μg of tRNA (lane j).

splicing. Third, the deletion in *dI714* brings poison sequences on the 5' side of the deletion into proximity with the splice site, which inhibits splicing. The second possibility is excluded by the results with *dI712*; i.e., although nt 1691 through 2122 are deleted in *dI712*, the nt 2157 3' splice site is used efficiently, as indicated by the 287-nt band (Fig. 4, lanes f and l). Indeed, the nt 2157 3' splice site is used more frequently with *dI712* than with *rec700*, *dI717*, or *dI716* (see below). *dI712* also gave a band of 321 nt, which represents colinear mRNA extending from the deletion (nt 2122) to the

FIG. 7. Nuclease gel analysis of RNA from cells infected with *dl719*, a mutant with a deletion on the exon side of the nt 2157 3' splice site. The figure also shows a nuclease gel analysis of *rec700* RNA which establishes that the probe was in excess at 30 μ g of RNA and, therefore, that the data are quantitative. In the *dl719* analysis, the *XmnI* and *StuI* probes were prepared from *dl719* DNA. In the *rec700* analysis, the *StuI* probe was prepared from *rec700* DNA. Cytoplasmic RNA was used. See the legends to Fig. 3 and 4 for details. Lanes a and h, markers; lane b, *dl719 XmnI* probe versus 30 μ g of *dl719* RNA; lane c, *dl719 XmnI* probe versus 30 μ g of tRNA; lane d, *dl719 XmnI* probe; lane e, *dl719 StuI* probe versus 30 μ g of *dl719* RNA; lane f, *dl719* probe versus 30 μ g of tRNA; lane g, *dl719 StuI* probe; lane i, *rec700 StuI* probe; lanes j through l, *rec700 StuI* probe versus 7.5, 15, and 30 μ g of *rec700*, respectively.

EcoRI (2437) site. This band is of the same size as the colinear 321-nt band observed with *dl717* because the 3' boundaries of the deletions in *dl712* and *dl717* are identical (Fig. 5).

Analysis of *dl719* which has a deletion on the exon side of the nt 2157 3' splice site. The deletion in *dl719* leaves 16 5' proximal exon nucleotides at the nt 2157 3' splice site (Fig. 5). The *XmnI* (Fig. 7, lane d) and *StuI* (lane g) probes used to analyze the effects of this deletion were prepared from plasmid clones of *dl719*, and therefore the probes include the *dl719* 65-base-pair deletion. When these probes were hybridized to *dl719* mRNA, a strong band of 222 nt was observed (Fig. 7, lanes b and e). This band corresponds to the 287-nt band in *rec700* mRNA (Fig. 7, lanes j through l), except that it is 65 nt smaller because of the deletion in *dl719*. The 222-nt band representing the nt 2157 splice in *dl719* mRNA was considerably more abundant than the corresponding band from *rec700* mRNA; e.g., compare lanes b and e with lane 1. The intensity of the 287-nt band increased with the amount of *rec700* RNA used (Fig. 7, lanes j through l), which establishes that the probe was in excess and that the results are quantitative. The increased abundance of the nt 2157 3' splice in *dl719* is discussed below. We conclude that the deletion in *dl719* does not adversely affect splicing at the nt 2157 3' splice site.

DISCUSSION

We have shown that a 3' splice site exists at nt 2157 in the E3 transcription unit of Ad2. This site had been detected previously in the E3 transcription unit of Ad5 (12). The nt 2157 3' splice site is used by mRNA *f*, which encodes the putative E3B 10,400-molecular-weight (10.4K) or 7.5K protein or both (Fig. 1). In this paper, we have analyzed viable virus deletion mutants to define the sequences which are required in *cis* for splicing in vivo at the nt 2157 3' splice site. Our results are discussed below.

The branchpoint sequence for lariat formation apparently can be deleted without affecting splicing at the nt 2157 3' splice site. With mutants *dl717*, *dl712*, and *dl716*, deletions that leave 34 or 18 3'-proximal intron nt intact did not prevent splicing at the nt 2157 3' splice site, as assayed in steady-state cytoplasmic mRNA. Splicing occurs by a lariat mechanism, and the consensus sequence for the lariat branchpoint site is of the form YNYRAY (28, 51, 64). Based on this consensus sequence, the Ad2 E3 nt 2157 3' splice branchpoint site is probably located at one of the A's situated 19, 22, 32, 37, 41, or 45 nt upstream from the splice site (Fig. 5). All these A's are deleted in *dl716*, which carries out the nt 2157 3' splice. Therefore, it appears that the natural branchpoint is not obligatory for splicing at the nt 2157 3' splice site. The deletion in *dl716* is such that other A's are brought within 25, 27, 34, 36, and 41 nt of the nt 2157 3' splice site (Fig. 5). Presumably one or more of these A's is used as a surrogate branchpoint site. If so, then there is a lack of stringency in the branchpoint sequence. This would be in accord with the lack of an obvious consensus sequence near 3' splice sites (28, 51, 64). However, it contrasts with the strict conservation of the TACTAAC box in yeast. It is interesting that in yeasts the TACTAAC box but not the AG is essential for 3' splicing (33, 34, 48), whereas in metazoa the AG consensus sequence but not the natural branchpoint site is essential.

Our results are in general accord with earlier genetic studies with viable virus, in which it was concluded that deletions that leave 12 3'-proximal intron nucleotides did not prevent splicing of pre-mRNA for simian virus 40 T antigen

(30, 55, 57). With these mutants, a surrogate branchpoint site may have been used for splicing. However, we note that our results as well as those with simian virus 40 do not establish that surrogate branchpoint sites are used as efficiently as the natural site. That is, if splicing is not the rate-limiting step in mRNA synthesis, then decreased splicing activity might not be reflected in steady-state cytoplasmic mRNA or in the synthesis of T antigen. This also applies to our data on *dI719*, which is deleted on the exon side of the nt 2157 3' splice site. We cannot examine this question for the mutants described in this paper by, e.g., analyzing nuclear RNA for unspliced pre-mRNA. That is, since there is a great abundance of E3 mRNAs (mRNAs *b*, *c*, and *e*) that do not splice at the nt 2157 3' splice site, we would be unable to distinguish unspliced pre-mRNAs from these natural mRNAs.

A deletion that leaves 15 3'-proximal intron nt prevents splicing at the nt 2157 3' splice site. The nt 2157 3' splice is abolished, or almost abolished, by the deletion in *dI714*. This is remarkable because the deletion is only 3 nt closer to the splice site than is the deletion in *dI716*; i.e., 15 nt are remaining in *dI714* versus 18 nt in *dI716* (Fig. 5). One explanation for this result is that the TGT (or a portion thereof) located 18 to 16 nt from the nt 2157 3' splice is crucial for splicing. Most introns contain a G residue flanked by pyrimidines at approximately the same position (see reference 42), and these triplets could be analogs of the TGT.

An alternate explanation is that the TGT represents the 5' boundary of the pyrimidine region upstream of the nt 2157 3' splice site and that the TGT sequence per se may be irrelevant. If so, then deletion of the 5' boundary of the pyrimidine-rich region prevents splicing. Pyrimidine-rich regions are found within ≥ 11 nt of virtually all 3' splice sites (42); therefore, they probably form part of the 3' splice signal. Wieringa et al. (61) have carried out a genetic analysis of the excision of the rabbit β -globin large intron in vivo in an expression vector. They reported that deletions leaving 15 3'-proximal intron nt did not grossly affect splicing, whereas deletions leaving 12 nt reduced splicing 10-fold. A TGT was not required for splicing. Our results are consistent with their results, except that the 2157 site requires 18 to 16 3'-proximal intron nt rather than 15 to 13. With simian virus 40 T antigen pre-mRNA, a deletion that leaves 12 3'-proximal intron nt did not prevent splicing in viable virus mutants (30, 55, 57). Perhaps significantly, two nonoverlapping TGT triplets were retained near the 3' splice site in this simian virus 40 mutant. In any case, it appears that the number of crucial nt in the pyrimidine-rich region can vary somewhat at different 3' splice sites.

A third explanation for the splicing defect in *dI714* is that the deletion brings poison sequences into proximity with the nt 2157 3' splice site, and these sequences inhibit splicing. For example, the AG's at nt 2047, 2052, and 2061 could be poison sequences (Fig. 5). These AG's would be located 38, 33, or 24 nt, respectively, from the nt 2157 3' splice site in *dI714*. AG's are found an average of 41.5 nt upstream of 3' splice sites (42). However, Mount (42) cites five examples where an AG is present only 14 nt upstream of the 3' splice site. This suggests that the splicing defect in *dI714* is probably not caused by AG's that inhibit splicing.

Mutations that destroy splice sites can result in the activation of cryptic splice sites. We have not detected a cryptic 3' splice site in *dI714* mRNA with probes that span nt 1354 to 2437 (data not shown). This is of interest because there are several AG's, e.g., at nt 2047 (Fig. 5), that match the 3' consensus sequence quite well and that are fairly close to the nt 2157 3' splice site. Most likely, all pre-mRNAs in *dI714* are

spliced to the other natural 3' splice sites in the E3 transcription unit, and use of a cryptic 3' splice site is not required.

Deletion of 17 to 81 5' proximal exon sequences does not prevent splicing at the nt 2157 3' splice site. The deletion in *dI719* leaves 16 nt remaining on the exon side of the nt 2157 3' splice site (Fig. 5). This mutation did not prevent splicing at the nt 2157 3' splice site. We conclude that sequences located 17 to 81 nt downstream of the nt 2157 3' splice site are not required for splicing at that site. This is not surprising, because the sequences deleted in *dI719* are in the coding region of mRNA *f* for the putative E3B 10.4K or 7.5K protein or both (Fig. 1). That is, in this E3 exon, as well as in protein-coding exons in general, it would be difficult to conserve simultaneously both the splicing signals as well as the protein sequence.

Enhanced splicing at the nt 2157 3' splice site. mRNA *f* with the nt 2157 3' splice site is severalfold more abundant in mRNA from *dI712* and *dI719* as compared with *rec700* and with the other mutants. In the case of *dI712*, we know that the activity of the 951 5' splice site is enhanced severalfold as measured in steady-state mRNA (S. Deutscher, B. Bhat, M. Pursley, C. Cladaras, and W. Wold, manuscript in preparation). Apparently, the nt 2157 3' splice site is one of the recipients of the enhanced splicing at the nt 951 5' splice site. We do not believe that the enhanced splicing in *dI712* results from an activation effect on the nt 2157 3' splice site because other mutants with deletions similar to *dI712* but with the nt 2157 3' splice site deleted also have enhanced splicing at the nt 951 5' splice site.

In the case of *dI719*, we know that the E3A RNA 3' end signal is destroyed by the deletion (unpublished data). Since all E3 mRNAs must terminate at the E3B site, this provides for increased opportunity to splice at the nt 2157 3' splice site. Again, we have no reason to believe that the deletion in *dI719* actually activates the nt 2157 3' splice site.

The E3A 11.6K protein and the E3B 10.4K or 7.5K protein (or both) are not required in trans for splicing at the nt 2157 3' splice site. The deletion in *dI712* removes the gene for the E3A 11.6K protein. Similarly, the deletion in *dI719* prevents the synthesis of both the putative E3B 10.4K and 7.5K proteins. Therefore, these proteins are not required in trans for splicing at the nt 2157 3' splice site.

Concluding remarks. Despite its complexity, splicing is exquisitely precise. Some of the *cis*-acting components in the splicing reaction are known: the GT and some other nucleotides in the consensus sequence at the 5' splice site; the AG at the 3' splice site; a portion of the pyrimidine-rich region, perhaps including TGT or a similar sequence located ~ 12 to 18 nt upstream from the 3' splice site; and the branchpoint sequence. *Trans*-acting RNA and protein factors probably bind to these sites and carry out the splicing reaction. However, there are many regions in DNA which appear to conform to the consensus sequences as well as do natural splice sites and which have a potential branchpoint site at the appropriate position. Indeed, some of these regions do function as cryptic 5' or 3' splice sites in certain mutants. This observation, and the fact that the branchpoint sequence can be deleted and surrogate branchpoint sites can be used instead argues that the splice consensus sequences and the branchpoint site alone do not determine splicing specificity.

What other *cis*-acting components are involved? There is considerable evidence that specific intron RNA secondary and tertiary structure is crucial for aligning the splice sites in pre-rRNA in nuclei and mitochondria and pre-mRNA in mitochondria of lower eucaryotes (see reference 59). How-

ever, a specific intron RNA tertiary structure cannot be obligatory for splicing of metazoan pre-mRNA, otherwise large deletions or insertions or both in introns would preclude splicing. Similarly, chimeric synthetic splice sites should not splice. Nevertheless, some mutations in introns (27) or exons (21, 30, 47) do reduce or eliminate splicing or activate distant cryptic 3' splice sites (14). In the Ad2 E3 transcription unit, deletions in the region of nt 1691 through 2044 specifically enhance splicing at the nt 951 5' splice site (Deutscher et al., in preparation). These considerations suggest that RNA folding does play a role in splicing, as proposed by Khoury et al. (30) and Kühne et al. (32). RNA folding, acting in concert with *trans*-acting splicing factors, may align the natural 5' and 3' splice sites. There are probably multitudes of possible folding pathways, and the splicing factors fix the RNA in the optimal configuration for splicing. However, other configurations, such as those existing in deletion mutants and in artificial splice constructs, can be sufficient for splicing to occur. Depending upon the pre-mRNA, some mutations may affect the efficiency of the splice or else activate cryptic splice sites. A major challenge is to elucidate the structural relationships between pre-mRNA and splicing factors in the ribonucleoprotein splicing complex.

ACKNOWLEDGMENTS

S. Magie, N. Yacoub, and S. Deutscher isolated the plasmid mutants, and C. Cladaras participated in the planning of the RNase experiments. The virus mutant plaques were developed by N. Takemori.

This work was supported by Public Health Service grants CA24710 and GM31276 from the National Institutes of Health and by Research Career Development Award no. CA00650 (to W.S.M.W.).

LITERATURE CITED

- Ahmed, C. M. I., R. Chanda, N. Stow, and B. S. Zain. 1982. The sequence of 3' termini of mRNAs from early region III of adenovirus 2. *Gene* 19:297-301.
- Ahmed, C. M. I., R. S. Chanda, N. D. Stow, and B. S. Zain. 1982. The nucleotide sequence of mRNA for the M_r 19,000 glycoprotein from early gene block III of adenovirus 2. *Gene* 20:337-344.
- Antonarakis, S. E., S. H. Irkin, T.-C. Cheng, A. F. Scott, J. P. Sexton, S. P. Trusko, S. Charache, and H. H. Kazazian, Jr. 1984. *Proc. Natl. Acad. Sci. U.S.A.* 81:1154-1158.
- Baker, C. C., and E. B. Ziff. 1981. Promoters and heterogeneous 5' termini of the messenger RNAs of adenovirus serotype 2. *J. Mol. Biol.* 149:189-221.
- Berk, A. J., and P. A. Sharp. 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell* 12:721-732.
- Berk, A. J., and P. A. Sharp. 1978. Structure of the adenovirus 2 early mRNAs. *Cell* 14:695-711.
- Breathnach, R., C. Benoist, K. O'Hare, F. Gannon, and P. Chambon. 1978. Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. U.S.A.* 75:4853-4857.
- Busslinger, M., N. Moschonas, and R. A. Flavell. 1981. β^+ Thalassemia: aberrant splicing results from a single point mutation in an intron. *Cell* 27:289-298.
- Chinnadurai, G., S. Chinnadurai, and J. Brusca. 1979. Physical mapping of a large-plaque mutation of adenovirus type 2. *J. Virol.* 32:623-628.
- Chow, L. T., T. R. Broker, and J. B. Lewis. 1979. Complex splicing patterns of RNAs from the early regions of adenovirus-2. *J. Mol. Biol.* 134:265-303.
- Chu, G., and P. A. Sharp. 1981. A gene chimaera of SV40 and mouse β -globin is transcribed and properly spliced. *Nature* (London) 289:378-382.
- Cladaras, C., B. Bhat, and W. S. M. Wold. 1985. Mapping the 5' ends, 3' ends, and splice sites of mRNAs from the early E3 transcription unit of adenovirus 5. *Virology* 140:44-54.
- Cladaras, C., and W. S. M. Wold. 1985. DNA sequence of the E3 transcription unit of adenovirus 5. *Virology* 140:28-43.
- Dobkin, C., R. G. Pergolizzi, P. Bahre, and A. Bank. 1983. Abnormal splice in a mutant human β -globin gene not at the site of a mutation. *Proc. Natl. Acad. Sci. U.S.A.* 80:1184-1188.
- Domdey, H., B. Apostol, R.-J. Lin, A. Newman, E. Brody, and J. Abelson. 1984. Lariat structures are in vivo intermediates in yeast pre-mRNA splicing. *Cell* 39:611-621.
- Esumi, H., Y. Takahashi, S. Sato, S. Nagase, and T. Sugimura. 1983. A seven-base-pair deletion in an intron of the albumin gene of analbuminemic rats. *Proc. Natl. Acad. Sci. U.S.A.* 80:95-99.
- Felber, B. K., S. H. Orkin, and D. H. Hamer. 1982. Abnormal RNA splicing causes one form of α thalassemia. *Cell* 29:895-902.
- Favoloro, J., R. Treisman, and R. Kamen. 1980. Transcription maps of polyoma virus-specific RNA: analysis by two-dimensional nuclease S1 gel mapping. *Methods Enzymol.* 65:718-749.
- Frost, E., and J. Williams. 1978. Mapping temperature-sensitive and host-range mutations of adenovirus type 5 by marker rescue. *Virology* 91:39-50.
- Fukumaki, Y., P. K. Ghosh, E. J. Benz, Jr., V. B. Reddy, P. Lebowitz, B. G. Forget, and S. M. Weissman. 1982. Abnormally spliced messenger RNA in erythroid cells from patients with β^+ thalassemia and monkey cells expressing a cloned β^+ -thalassemic gene. *Cell* 28:585-593.
- Ghosh, P. K., M. Piatak, J. E. Mertz, S. M. Weissman, and P. Lebowitz. 1982. Altered utilization of splice sites and 5' termini in late RNAs produced by leader region mutants of simian virus 40. *J. Virol.* 44:610-624.
- Goldsmith, M. E., R. K. Humphries, T. Ley, A. Cline, J. A. Kantor, and A. W. Nienhuis. 1983. "Silent" nucleotide substitution in a β^+ -thalassemia globin gene activates splice site in coding sequence RNA. *Proc. Natl. Acad. Sci. U.S.A.* 80:2318-2322.
- Graham, F. L., and A. J. van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* 52:456-467.
- Green, M., and W. S. M. Wold. 1979. Preparation of human adenoviruses. *Methods Enzymol.* 58:425-435.
- Hamer, D. H., and P. Leder. 1979. SV40 recombinants carrying a functional RNA splice junction and polyadenylation site from the chromosomal mouse β^{maj} globin gene. *Cell* 17:737-747.
- Hérissé, J., G. Courtois, and F. Galibert. 1980. Nucleotide sequence of the *EcoRI* D fragment of adenovirus 2 genome. *Nucleic Acids Res.* 8:2173-2192.
- Hwang, L.-H. S., J. Park, and G. Gilboa. 1984. Role of intron-contained sequences in formation of Moloney murine leukemia virus *env* mRNA. *Mol. Cell. Biol.* 4:2289-2297.
- Keller, E. B., and W. A. Noon. 1984. Intron splicing: a conserved internal signal in introns of animal pre-mRNAs. *Proc. Natl. Acad. Sci. U.S.A.* 81:7417-7420.
- Khoury, G., J. Alwine, N. Goldman, P. Gruss, and G. Jay. 1980. New chimeric splice junction in adenovirus type 2-simian virus 40 hybrid viral mRNA. *J. Virol.* 36:143-151.
- Khoury, G., P. Gruss, R. Dhar, and C.-J. Lai. 1979. Processing and expression of early SV40 mRNA: a role for RNA conformation in splicing. *Cell* 18:85-92.
- Kitchingman, G. R., and H. Westphal. 1980. The structure of adenovirus 2 early nuclear and cytoplasmic RNAs. *J. Mol. Biol.* 137:23-48.
- Kühne, T., B. Wieringa, J. Reiser, and C. Weissmann. 1983. Evidence against a scanning model of RNA splicing. *EMBO J.* 2:727-733.
- Lanford, C. J., and D. Gallwitz. 1983. Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. *Cell* 33:519-527.
- Lanford, C. J., F.-J. Klinz, C. Donath, and D. Gallwitz. 1984.

- Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal in yeast. *Cell* **36**:645-653.
35. Lee, D. C., R. G. Roeder, and W. S. M. Wold. 1982. DNA sequences affecting specific initiation of transcription *in vitro* from the E111 promoter of adenovirus 2. *Proc. Natl. Acad. Sci. U.S.A.* **79**:41-45.
 36. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**:499-560.
 37. Melton, D. A., P. A. Krieg, M. R. Rebagliati, K. Zinn, and M. R. Green. 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**:7035-7056.
 38. Montell, C., and A. J. Berk. 1984. Elimination of mRNA splicing by a point mutant outside the conserved GU at 5' splice sites. *Nucleic Acids Res.* **12**:3821-3827.
 39. Montell, C., G. Courtois, C. Eng, and A. Berk. 1984. Complete transformation by adenovirus 2 requires both E1A proteins. *Cell* **36**:951-961.
 40. Montell, C., E. F. Fisher, M. H. Caruthers, and A. J. Berk. 1982. Resolving the functions of overlapping viral genes by site-directed mutagenesis at a mRNA splice site. *Nature (London)* **295**:380-384.
 41. Montell, C., E. F. Fisher, M. H. Caruthers, and A. J. Berk. 1984. Control of adenovirus E1B mRNA synthesis by a shift in the activities of RNA splice sites. *Mol. Cell. Biol.* **4**:966-972.
 42. Mount, S. M. 1982. A catalogue of splice junctions. *Nucleic Acids Res.* **10**:459-472.
 43. Orkin, S. H., S. C. Goff, and R. L. Hechtman. 1981. Mutation in an intervening sequence splice junction in man. *Proc. Natl. Acad. Sci. U.S.A.* **78**:5041-5045.
 44. Orkin, S. H., H. H. Kazazian, Jr., S. E. Antonarakis, H. Ostrer, S. C. Goff, and J. P. Sexton. 1982. Abnormal RNA processing due to the exon mutation of β^E -globin gene. *Nature (London)* **300**:768-769.
 45. Orkin, S. H., J. P. Sexton, S. C. Goff, and H. H. Kazazian, Jr. 1983. Inactivation of an acceptor RNA splice site by a short deletion in β -thalassemia. *J. Biol. Chem.* **258**:7249-7251.
 46. Padgett, R. A., M. M. Konarska, P. J. Grabowski, S. F. Hardy, and P. A. Sharp. 1984. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science* **225**:898-903.
 47. Piatak, M., P. K. Ghosh, L. C. Norkin, and S. M. Weissman. 1983. Sequences locating the 5' ends of the major simian virus 40 late mRNA forms. *J. Virol.* **48**:503-520.
 48. Pikielny, C. W., J. L. Teem, and M. Rosbash. 1983. Evidence for the biochemical role of an internal sequence in yeast nuclear mRNA introns: implications for U1 RNA and metazoan mRNA splicing. *Cell* **34**:395-403.
 49. Rautmann, G., H. W. D. Matthes, M. J. Gait, and R. Breathnach. 1984. Synthetic donor and acceptor splice sites function in an RNA polymerase B (II) transcription unit. *EMBO J.* **3**:2021-2028.
 50. Rodriguez, J. R., C. W. Pikielny, and M. Rosbash. 1984. *In vivo* characterization of yeast mRNA processing intermediates. *Cell* **39**:603-610.
 51. Ruskin, B., A. R. Krainer, T. Maniatis, and M. R. Green. 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. *Cell* **38**:317-331.
 52. Solnick, D. 1980. An adenovirus mutant defective in splicing RNA from early region 1A. *Nature (London)* **291**:508-510.
 53. Stalhandske, P., H. Persson, M. Perricaudet, L. Philipson, and U. Pettersson. 1983. Structure of three splice mRNAs from region E3 of adenovirus type 2. *Gene* **22**:157-165.
 54. Sutcliffe, J. G. 1979. Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. *Cold Spring Harbor Symp. Quant. Biol.* **43**:77-90.
 55. Thimmappaya, B., and T. Shenk. 1979. Nucleotide sequence analysis of viable deletion mutants lacking segments of the simian virus 40 genome coding for small t antigen. *J. Virol.* **30**:668-673.
 56. Treisman, R., S. H. Orkin, and T. Maniatis. 1983. Specific transcription and RNA splicing defects in five cloned β -thalassaemia genes. *Nature (London)* **302**:591-596.
 57. Volckaert, G., J. Feunteun, L. V. Crawford, P. Berg, and W. Fiers. 1979. Nucleotide sequence deletions within the coding region for small-t antigen of simian virus 40. *J. Virol.* **30**:674-682.
 58. Wallace, J. C., and M. Edmunds. 1983. Polyadenylated nuclear RNA contains branches. *Proc. Natl. Acad. Sci. U.S.A.* **80**:950-954.
 59. Waring, R. B., J. A. Ray, S. W. Edwards, C. Scazzocchio, and R. W. Davies. 1985. The tetrahymena rRNA intron self-splices in *E. coli*: *in vivo* evidence for the importance of key base-paired regions of RNA for RNA enzyme function. *Cell* **40**:371-380.
 60. White, R. T., P. Berg, and L. P. Villarreal. 1982. Simian virus 40-rabbit β -globin recombinants lacking late mRNA splice sites express cytoplasmic RNAs with altered structures. *J. Virol.* **42**:262-274.
 61. Wieringa, B., E. Hofer, and C. Weissmann. 1984. A minimal intron length but no specific internal sequence is required for splicing the large rabbit β -globin intron. *Cell* **37**:915-925.
 62. Wieringa, B., F. Meyer, J. Reiser, and C. Weissmann. 1983. Unusual splice sites revealed by mutagenic inactivation of an authentic splice site of the rabbit β -globin gene. *Nature (London)* **301**:38-43.
 63. Wold, W. S. M., M. Green, K. H. Brackmann, M. A. Cartas, and C. Devine. 1976. Genome expression and mRNA maturation at late stages of productive adenovirus type 2 infection. *J. Virol.* **20**:465-477.
 64. Zeitlin, S., and A. Efstratiadis. 1984. *In vivo* splicing products of the rabbit β -globin pre-mRNA. *Cell* **39**:589-602.