***In vitro* splicing of adenovirus E1A transcripts: characterization of novel reactions and of multiple branch points abnormally far from the 3' splice site**

Renata Gattoni, Philippe Schmitt and James Stevenin

Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Institut de Chimie Biologique, Faculté de Médecine, 11 rue Humann, 67085 Strasbourg Cédex, France

Received February 4, 1988; Accepted February 29, 1988

ABSTRACT

During the analysis of the *in vitro* alternative splicing of the natural E1A transcript of adenovirus, other minor reactions were detected (Schmitt et al., 1987, Cell 50,31-39). We report here their characterization. The first reaction concerns the excision of a 216 nucleotide intron delineated by the 9S 5' splice site and a 3' splice site 216 nucleotides downstream. It can occur on the premRNA transcript and the 13S and 12S mRNA species. Strikingly, the reaction uses one of 3 branch points located 51, 55 or 59 residues upstream of the 3' splice site, a distance which is unusually long since all the branch points mapped up to now are located between 18-37 nucleotides of the 3' splice site. The dramatic accumulation of the corresponding lariat intermediates, likely related to this long spacing indicates that the second splicing step is relatively unefficient. The second kind of reaction analysed is a cryptic splicing which uses a 3' splice site generated by the junction of the 13S mRNA exons, and leads to the formation of ψ 12S and ψ 9S mRNAs. *In vitro*, this reaction occurs only from a 13S mRNA transcript, and not from the 13S mRNA newly formed in the splicing assay, consistent with what has been observed *in vivo*. Thus, both the well known alternative and the minor reactions occurring *in vivo* from E1A premRNA and mRNAs are detected *in vitro*, implying that most of the alternative splicing machinery is reconstituted in the *in vitro* system.

INTRODUCTION

Development of faithful and efficient *in vitro* splicing systems has greatly facilitated the general comprehension of the premRNA splicing reaction in higher eukaryotes (1-3). A two-step mechanism has been proposed for premRNA splicing. The first step is a cleavage at the 5' splice site and a concomitant formation of a lariat by joining the 5' end of the IVS to an A residue, some 18-37 nucleotides upstream of the 3' splice site (4-6). The second step involves the release of the intron as a lariat by cleavage at the 3' splice site and the ligation of the exons. The splicing of premRNA occurs in a large multicomponent complex (7,8) termed spliceosome (9). Small nuclear ribonucleoproteins particles (snRNPs) are required for premRNA splicing and they are important components of the spliceosome (for reviews, 10,11).

In spite of recent progress, all the details of the mechanism for recognition of the bona fide splice sites: the 5' and 3' sites at the exon-intron junctions and the branch sites, are not still completely established. For example, all the parameters important for the recognition, then the selection of the conventional splice sites, versus the cryptic splice sites (used only in abnormal situations) are not known. From this perspective, analysis of alternative splicing is interesting because it involves the selective use of several splice sites present on the premRNA chain. Different parameters, including the primary sequence of the splice sites (12) and the size of the introns (13) can be easily assessed. Artificial premRNAs containing additional splice sites or inverted sequences are useful for the analysis of effect of the relative position of the splice sites (14-17) or of the premRNA secondary structure (18) on splice site selection. However, the results of such structural studies cannot be interpreted too generally. In vivo, the observations that splice site selection may be controlled in a temporal or tissue specific manner (for reviews see 11,19,20), suggest that factors, in addition to premRNA structure, control alternative splicing. There are many instances where trans-acting factors can be directly implicated in changes in splicing patterns (21-23). The characterization of such trans-acting factors is essential for a complete understanding of the mechanisms controlling alternative splicing.

With these observations in mind, we previously developed an in vitro system capable of catalysing all of the classical alternative splicing reactions of the natural E1A transcript of adenovirus observed in vivo (24). Unexpectedly, we detected additional unidentified products indicating that other reactions had occurred (24). In order to characterize completely our in vitro system, we have analyzed these additional splicing reactions. One of them is a new type of splicing reaction involving the excision of a 216 nucleotide intron, delineated by the 9S 5' splice site and a novel 3' splice site 216 nucleotides downstream. This reaction was first pointed out from in vitro studies (24), and subsequently, the mRNA product was observed in infected cells (25,26). We demonstrate here that this reaction is unusual because the branching reaction occurs at one of three adenosine residues located an abnormally long distance from the 3' end of the intron and the consequences on the two steps of splicing are analysed. Another cryptic reaction, which was detected in HeLa cells transfected with the 13S cDNA (27,28), occurs in vitro, but this reaction takes place only under conditions which mimic the in vivo situation. Therefore, we conclude that

the in vitro system we developed is able to catalyse, under appropriate ionic conditions, alternative or cryptic splicing reactions identical to those occurring in vivo.

MATERIALS AND METHODS

Sp6 transcription and in vitro splicing

The DNA templates used for transcription were inserted between the SmaI and the XbaI sites of the polylinker of the pSP65 vector (29). The Sp4 clone contains E1A DNA sequence from position 533 to position 1341 of the genome (24). In the 13S cDNA clone, the genomic DNA sequence of the Sp4 clone was replaced by the 13S cDNA sequence. In vitro transcription of the DNA templates linearized with HindIII was performed as described by Krainer et al. (3) except that transcription reaction included 500 μ M of cap dinucleotide GpppG (30) and 10 mM NaCl. The RNAs were labeled by incorporation of [α^{32} P] CTP. The RNA was purified by RNAase-free DNAase I treatment and subsequent deproteinization, as described by Green et al. (31). The occurrence of a long U stretch ($U_9A_2U_5$) in the 13S IVS of the E1A gene induces first an unfaithful transcription generating a transcript with additional residues and second, a premature termination which results in the formation of a minor RNA of 680 nucleotides long (24). Since the presence of the latter did not hinder the analysis of the splicing reactions, the Sp4 substrate was not further purified.

In vitro splicing reactions were essentially performed as described by Krainer et al. (3), except that variable salt concentrations (indicated in the text) were used (24). The splicing reactions contained, in a volume of 25 μ l, 13 mM HEPES (pH 7.9), 39 to 60 mM KCl, 1.6 to 3.2 mM $MgCl_2$, 120 μ M EDTA, 0.5 mM ATP, 20 mM creatine phosphate, 0.5 mM dithiothreitol (DTT), 2.6% (w/v) polyvinyl alcohol, 200 U/ml of RNasin, 40 μ g/ml tRNA, 14% (v/v) glycerol, 10 μ l of nuclear extract and ~ 10000 Cerenkov cpm of [32 P] RNA, corresponding to 0.02–0.05 pmol per reaction. Splicing reactions were incubated at 30°C for 20 to 120 minutes as indicated and the processed RNAs were purified as described previously (24).

Acrylamide gel analysis of RNAs

RNA splicing products were analysed on 5% acrylamide/8M urea sequencing gels (32) with a 30:1 ratio of acrylamide : bisacrylamide as in Figure 1, or with a 19:1 ratio as previously (24). For preparative purposes, RNA products were fractionated on thick 5% acrylamide/8M urea gels. Radioactive bands were located by autoradiography, excised and electroeluted.

Direct RNA characterization

Isolated RNA species representing presumptive lariats were submitted to the debranching treatment (33) then deproteinized. These RNAs were analysed on sequencing gels, or used as substrates for primer extension experiments. For T1 RNAase digestion experiments, RNAs were digested for 3h at 40°C with 5U of RNAase T1 in 20 µl of 10 mM Tris-HCl (pH 7.5) and 1 mM EDTA containing 5 µg of carrier tRNA. The digestion products were fractionated on a 15% acrylamide sequencing gel. The RNAase protection experiments were essentially performed according to Keohavong et al. (34).

Primer extension analysis

Primer extension was performed using either a synthetic oligodeoxynucleotide : primer B of 20 residues (positions 1092-1111) or oligonucleotides purified from genomic restriction fragments : primer A of 26 residues from the AvaI-AvaII fragment (positions 858-883) and primer C of 28 residues from the AvaII-AvaI fragment (positions 1229-1256). The 5' end labeled primers were hybridized to isolated RNA species, then extended in the presence of AMV reverse transcriptase essentially as previously described (24). In order to obtain a partial sequencing of the analysed species, some assays have been done in the presence of dideoxynucleoside triphosphates. The cDNA products were analysed on 8% acrylamide sequencing gels.

RESULTS

To analyse the E1A premRNA splicing, we used the Sp4 transcript described previously (24), which is derived from the Ad2 genomic sequence and contains the 9S, the 12S and 13S 5' splice sites and the common 3' splice site (Fig. 1C). A derivative of the original Sp4 clone, which lacks the 13S intervening sequences, was used for the synthesis of a 13S transcript (Fig. 1C). Labelled substrates were incubated with a nuclear extract of non infected HeLa cells (35) under ionic conditions appropriate for alternative splicing into the 13S, 12S and 9S mRNAs (24). The RNA species expected from alternative splicing (four RNA products for each of the 3 splicing reactions) were previously characterized (24) and only the clearly resolved products are indicated on the left in Fig. 1A. Additional RNA species were also detected, indicating that other splicing reactions occur concomitantly with the alternative splicing reactions. The comparison of the splicing patterns obtained from the Sp4 and 13S RNA transcript (Fig. 1A,B), shows the presence of several similar splicing products, suggesting

that the newly spliced 13S mRNA product generated from the Sp4 transcript can serve as a substrate for a second splicing reaction.

In order to analyse the new splicing reactions, we have characterized the previously uncharacterized splicing products (24) detected in Fig. 1A and B, as follows. Individual species were isolated from preparative gels and their structure was determined by several assays: a debranching assay (Fig. 2), primer extension (Fig. 3), and RNAase protection mapping, or one dimensional RNAase T1 oligonucleotide mapping (Fig. 4). For clarity, we have numbered the newly characterized splicing products from 1 to 12 and their schematic representation is given in Fig. 1.

THE 216 nt INTRON SPLICING REACTION

Characterization of the 524 nt mRNA and 638 nt mRNA species

This minor reaction involves the excision of an intron sequence of 216 nucleotides lying between the 9S 5' splice site at position 637 and a 3' splice site terminating at position 852, whose sequence: (Py)₄A(Py)₈GGCAG is very close to the consensus acceptor sequence (36) (see Fig. 1C for the structure of mRNAs). Our interest was focused on this reaction upon analysis of the well-represented species 5, a linear RNA of ~ 520 nucleotides which accumulates at the end of incubation of both Sp4 and 13S transcripts (Fig. 1A, B). RNAase protection experiments using two 13S cDNA probes of different size, indicated that this species was homologous to the 13S mRNA sequence for a length of ~ 380 nucleotides, roughly from position 820-850 to a position close to 1340 (not shown). This also suggested that a non-colinear 5' sequence of ~ 120 nucleotides (likely exon 1) was joined to the 380 nucleotide fragment to form a mRNA species. This assumption was confirmed by primer extension analysis using primer A (positions 858-883) (Fig. 3 I). The comparison of the cDNA products extended up to T and C residues from the Sp4 transcript (lanes b and c) or from the 524 nt mRNA (species 5) (lanes e and f) showed a sequence divergence between both RNAs, from the AG dinucleotide at positions 851-852. This partial sequencing showed that the non-colinear sequence of the 524 nt mRNA was that of the 9S exon E1, which also starts by an AG dinucleotide and it was recognized up to at least positions 556-560. Moreover, the size of the cDNA extended without dideoxynucleotides (lane d), indicated that this RNA species starts at or near the first residue of the Sp4 transcript. Thus, the 524 nt mRNA (species 5) was formed from the 13S mRNA molecules, by the additional excision of a 216 nt intron. That this excision could also directly occur on the Sp4 transcript

Figure 1: Analysis of the minor splicing reactions occurring on Sp4 (A) and 13S mRNA (B) transcripts. The processing was in the presence of 2.1 mM $MgCl_2$ and 39 mM KCl and the RNA analysed on a denaturing 5% acrylamide gel. The species produced by the "classical" alternative splicing are represented at the left of panel A. The species specific to the 216 nt intron or to the cryptic splicing reaction are indicated between panels A and B or on the right of panel B, respectively. In B, the 655 nt RNA could originate from the use of a 5' splice site located at 1023 (12) and of the cryptic 3' splice site. The schematic drawings of the lariat species are along side while the mRNA products are represented in C. The boxes correspond to the exons and the dotted lined (A, B) to the introns. The markers are TaqI fragments (M_1) or MspI fragments (M_2) of pBR322 DNA.

Figure 2: Debranching analysis of the lariat species involved in the minor reactions. Every lariat RNA species was subjected (+) or not (-) to the debranching reaction, then analysed either in a 5% acrylamide gel of 30:1 acrylamide: bisacrylamide (species 7 and 8), or in a 19:1 cross-linked 5% acrylamide gel (species 1, 2 and 3) or 6% acrylamide gel (species 6). The sizes of the debranched species are indicated in nucleotides. Because of the low cross-linking of the preparative gels used for their isolation, species 7 and 8 are contaminated with the initial 13S transcript. The triplets of bands produced in the 216 nt intron reaction (species 1, 2, 3 and 6) are indicated by a, b and c. The additional broad band around 390 nucleotides for species 6 (+ and -) corresponds to background RNA, co-isolated with species 6 from the preparative gel. DNA markers M_1 and M_2 are as indicated in Fig. 1.

to generate a 638 nt mRNA was supported by the detection of a minor linear RNA of ~ 640 nucleotides (species 4), which accumulates after 1h of incubation (Fig. 1A). The 5' part of this species, analysed by primer extension using primer A as above, exhibited an RNA sequence identical to that of species 5 (not shown).

Analysis of the lariat intermediates

The exon E1 intermediate, formed by a primary cleavage at position 636, was expected to be indiscernible from the 9S exon E1 of 129 nucleotides, previously characterized (24). The IVS-E2 generated from the Sp4

Figure 3: Primer extension analysis of the splicing products. The Sp4 and 13S transcripts (Tr.) or the isolated species were subjected to primer extension analysis in the presence, or absence of dideoxynucleotides. In some cases, the debranched RNA from lariat species were also analysed, as indicated at the top of the lanes. The RNA sequence complementary to that of the extended cDNA is indicated along side. The size of the complete extended cDNA from the mRNA species or the debranched species is given in

nucleotides. The limits between exon 2 and IVS or exon 2 and exon 1 are indicated along side by a set of triangles.

I) Analysis of the Sp4 transcript and the 524 nt mRNA. The presence of a background of cleaved Sp4 transcripts, isolated with the 524 nt mRNA, was indicated by the detection of a minor cDNA of 380 nucleotides (lane d) and by minor stops in the presence of ddCTP (lane f), identical to those obtained with the Sp4 transcript (lane c).

II) Analysis of the 12S and ϕ 12S mRNAs and the 9S and ϕ 9S mRNAs. The small arrow-heads in lanes i and m show the shift of one nucleotide between the 12S and ϕ 12S and the 9S and ϕ 9S sequences, respectively.

III) Analysis of the IVS-E2 intermediates related to the 216 nt intron excision (species 1, 2 and 3) and the Sp4 transcript. With all the tested RNAs, including the Sp4 transcript, a band of 228 nucleotides was detectable but it was likely artefactual.

IV) Analysis of the 13S transcript, the ϕ 9S intron (species 8) and the ϕ 12S intron (species 12).

V) Location of the primers (Pr. A,B or C) and of the branching points (B.P.) on the E1A sequence. The adjacent boxes represent the additional nucleotides transcribed from the pSP65 DNA sequence. The different 5' and 3' splice sites are underlined by a dashed line and the branch points are marked with arrow-heads.

transcript (the 638 nt mRNA - IVS-E2) was anticipated to be a branched RNA with a sequence content of 725 nucleotides, from position 637 to the 3' end of the Sp4 transcript. The IVS-E2 generated from 13S mRNA (the 524 nt mRNA - IVS-E2) should have the same termini, but lacks the 13S intervening sequence. The group of 3 bands constituting species 1, detectable after 0.5h of incubation (Fig. 1A) was a good candidate for the 638 nt mRNA - IVS-E2, since the debranching reaction transformed it into a discrete linear segment of 730 nucleotides (Fig. 2). Similarly, the group of 3 bands constituting species 2, detectable after 1h of incubation of the Sp4 transcript (Fig. 1A) or after 0.5h for the 13S transcript (Fig. 1B), was debranched into a discrete fragment of 620 nucleotides (Fig. 2), a size close to that expected for the 524 nt mRNA - IVS-E2. Another minor triplet of bands (RNA species 3) was masked by the initial Sp4 transcript (Fig. 1A) but could be detected and isolated after analysis on a 5% acrylamide gel with an acrylamide: bisacrylamide ratio of 20:1 (24) instead of 30:1 as in Fig. 1. Debranching of species 3 gave a discrete RNA of 480 nucleotides (Fig. 2), whose size is close to that expected for the IVS-E2 formed during the 216 nt intron excision from the 12S mRNA. During splicing of the 13S transcript, another triplet of bands (species 9) was detected at the end of incubation (Fig. 1B). Species 9 has roughly the same characteristics as species 3 (not shown), but is formed from the ϕ 12S mRNA, instead of the 12S mRNA (see below).

It is noteworthy that species 1, 2, 3 and 9 display in gel analyses, the same pattern of 3 closely spaced bands a, b and c, as indicated in Fig. 2. This heterogeneity is a consequence of the use of the 3 different branch points, which were mapped by reverse transcription experiments, using primer A located downstream of the 216 nt intron (see Fig. 3 V). A first gel electrophoresis of the extended cDNA products, where the primer A of 26 nucleotides migrated to the bottom of the gel, showed that strong stops of transcription corresponding to putative branch points were mapped at a long distance (>50 nucleotides) from the terminal AG of the 216 nt IVS, (not shown). After a longer electrophoretic run (Fig. 3 III), an identical set of 3 stops was detected for each of the species 1, 2 and 3 (lanes r, s, u). A comparison with the transcription products of the initial Sp4 RNA in the presence of ddCTP or ddTTP (lanes p, q) located them precisely, at a nucleotide just preceding the A residues (*) at positions 794, 798 and 802, in the sequence : GATTGA*CTTA*TTCA*CT. That the 3 stops are due to a set of branch points was demonstrated by analysing species 1, 2 and 3, after debranching. With debranched species 2 (lane t) or with the others (not shown), the 3 strong stops of transcription disappeared and a predominant band of 250 nucleotides was observed. Its size accurately corresponded to that expected for the cDNA that extends from primer A to the 9S 5' splice site (position 637). Since the primer extension of a branched RNA systematically stops at the residue just downstream of the branch point (37), these results demonstrate that the 216 nt intron excision takes place via a branching reaction occurring at one of 3 A residues, located 51, 55 or 59 residues upstream of the 3' splice site junction.

Taken together, the results suggest that the triplets of bands a, b and c for each of the species 1, 2 and 3, correspond to lariats branched at 51, 55 or 59 residues of the 3' end of the 216nt intron, respectively. We confirmed this by analysing as above species 1, isolated either after 45 min of splicing reaction, when band a is predominant, or after 2 h of incubation, when the 3 bands a, b and c are roughly equivalent (Fig. 1A). We observed a strong stop at the proximal branch point early in the splicing reaction (not shown), whereas at the end of the incubation (2 h), the stops at the two more distal branch points predominated, as in Fig. 3 III (lanes r to u).

Characterization of the excised 216 nt intron

Due to the existence of 3 branch points, 3 forms of free 216 nt intron should be detected at late times of incubation. On 5 % acrylamide gels, a

set of 2 bands and more rarely a faint third band, were detected with an apparent size close to 390 nucleotides (Fig. 1A). In a 8 % acrylamide gel, these branched species migrated more slowly than the initial Sp4 transcript and a triplet of bands, constituting species 6, was systematically observed, in a 6:3:1 ratio for the three forms, designated as a, b and c, respectively (not shown). When the species 6a-b was debranched, a discrete segment of 217 nucleotides was obtained (Fig. 2). Its structure was confirmed by analysing RNAase T1 resistant products. For the Sp4 transcript, the T1 RNAase products should be, in decreasing order : a series of 22-31 nt fragments due to the inaccurate transcription of this sequence (24), three fragments of 19, 18, 17 nucleotides, all located in the 216 nt IVS and shorter fragments. The 3 potential branch points are all carried by the 17 nt fragment (positions 794-810) which has not to be detected for branched RNAs. Figure 4 shows the analysis on a 15 % acrylamide gel of the RNAase T1 digestion products of Sp4 RNA (lane a), of Sp4 RNA stopping at the SmaI site at position 1008 (lane b) and of Sp2 RNA, starting at position 815 instead of 533 (lane c). As expected, the 17, 18 and 19 nt fragments are detected with the 2 first RNAs but not with the third one. With the RNA species 6a and 6b, only the 2 first oligonucleotides of 19 and 18 residues were detected (lanes d,f). After debranching, the 17 nt fragment became visible (lane e), confirming that the species 6a and 6b are effectively 2 forms of the branched 216 nt intron. In Fig. 4, unexpected resistant fragments are displayed on the gels. Two of them, with an apparent size of 75 and 25 nucleotides, are present with species 6a (lane d). They include at least the sequence of 17 nucleotides and the branched G residue since they disappear after debranching reaction (lane e).

In summary, the 216 nt intron splicing reaction can take place on the Sp4 transcript directly, generating the 638 nt mRNA, or on the 13S mRNA to generate the 524 nt mRNA. This reaction can also occur on the 12S and ϕ 12S mRNAs, since the corresponding IVS-E2 (species 3 and 9, respectively) were detected. However, the final mRNA products (the 386 nt or the 385 nt mRNA, respectively) were poorly and irregularly detected, thus excluding their further characterization.

SPLICING REACTION USING A CRYPTIC 3' SPLICE SITE AT THE JUNCTION OF THE 13S mRNA EXONS

In the 13S mRNA, the junction of exon E1 terminating by a pyrimidine rich stretch with exon E2 starting at a G residue (Fig. 3 V), generates the sequence $(Py)_3G(Py)_4ACAG$ which is very similar to a consensus acceptor

Figure 4: RNAase T1 digestion analysis of the splicing products. The labeled transcripts Sp4, Sp4 terminating at 1008, Sp2 (starting at position 815) and the splicing products were digested with RNAase T1, with or without prior enzymatic debranching. The digestion products were fractionated in a denaturing 15% acrylamide gel. The sizes of the expected RNA fragments are indicated and roughly correspond to those deduced from the DNA markers. The size markers are the MspI DNA fragments of pBR322 (M2).

sequence (36). When a plasmid containing 13S cDNA is transfected into human cells, a splicing reaction occurs which utilizes this cryptic 3' splice site and the 9S or 12S 5' splice site, leading to the production of mRNAs which exhibit a deletion of a G residue at position 1226, relative to the genuine sequence of 9S and 12S mRNA (27,18). We will designate these RNAs as pseudo (ψ) 12S and 9S mRNA sequences. Since this cryptic reaction was not detected in cells infected with adenovirus or transfected with the E1A gene (27), it was interesting to determine whether the cryptic reaction takes place in vitro, and under which conditions. Accordingly, we have analyzed the splicing products of the 13S transcript. Under salt conditions optimal for alternative splicing of E1A premRNA, as in Fig. 1A, we detected a set of splicing products (Fig. 1B). Some of them (species 2, 5, 10, 11) have an electrophoretic migration identical or similar to that of the species formed from the Sp4 transcript. Species 10 and 11 (Fig. 1B), which have a migration indistinguishable from that of the 12S and 9S mRNA (Fig. 1A) were isolated. They were analysed in parallel with authentic 12S and 9S mRNAs by reverse transcription of primer C in the presence or absence of

ddATP (Fig. 3 II). Species 10 and 11 (lanes j,n) have a sequence identical to that of the genuine 12S and 9S mRNA (lanes h,l) respectively, except for the systematic shift of one residue in the ladder of the stops at the U residues, indicating a deletion of one residue close to, or exactly at the exon junction. Therefore, species 10 and 11 have characteristics identical to ψ 12S and the ψ 9S mRNA species.

Among the other products of these cryptic splicing reactions, the ψ 9S IVS-E2 was expected to contain the 13S mRNA sequence between positions 637 and 1341 and the ψ 9S IVS the same sequence between positions 637 and 1111 plus the G residue at position 1226 (see Fig. 1C). Species 7 and 8, detected in the early and late stages of the processing, respectively (Fig. 1B) were transformed after debranching into linear RNAs of 620 and 485 nucleotides, respectively (Fig. 2), a size close to that expected for the ψ 9S IVS-E2 (611 nucleotides) and the ψ 9S intron (476 nucleotides). For the ψ 12S splicing reaction, no species which could correspond to the IVS-E2 was detected, but species 12, a lariat RNA with an apparent size of 185 nucleotides accumulated at the end of incubation (Fig. 1B). Debranching of this species gives a linear segment of 140 nucleotides (not shown), very close to the size expected for the ψ 12S intron. By reverse transcription of species 8 and 12, the putative ψ 9S and ψ 12S excised introns, using primer B (positions 1091-1110) complementary to the 3' tail of the cryptic introns, a strong stop was detected just preceding an A residue in the Sp4 sequence (Fig. 3 IV, lanes x,y). It locates the branch point 29 residues upstream of the 3' end of the cryptic IVS, at the last A residue of the sequence GCTATA_xT, which weakly resembles the branch-point consensus sequence (10,11). Reverse transcription of debranched species 12 (Fig. 3 IV, lane z), in the presence of ddTTP showed that the RNA sequence upstream of the branch point was that of the Sp4 sequence, and obtention of a cDNA band of 138 nucleotides confirmed that species 12 is the ψ 12S IVS.

The comparison of the splicing products generated from Sp4 and 13S mRNA transcripts (Fig. 1A,B) clearly shows that the reactions at cryptic splice sites do not occur on the newly spliced 13S mRNA generated from the Sp4 transcript. However, this is not due to the inability of the 13S mRNA to serve as a substrate for a second splicing reaction, since the 216 nt intron is spliced with good efficiency. We also verified that the absence of cryptic reactions was not the result of a chemical modification of the newly spliced 13S mRNA; after its isolation and purification, this RNA was

TABLE 1: Efficiency of the splicing under various salt concentrations.

KCl concentration (mM) :	60	48	39	39
MgCl ₂ concentration (mM) :	3.2	2.6	2.1	1.6
Nature of the reaction				
13S mRNA	++++	+++++	++	+
12S mRNA	0	+	+++++	+++++
9S mRNA	0	+	+++++	+++++
216nt IVS	0 to ++	+++++	+++	+++
ψ12S mRNA	0 to +	++	+++++	+++++
ψ9S mRNA	0	+	+++++	+++++

The efficiency of the splicing reaction was scored semi quantitatively by determining the relative amounts of the IVS-E2 or the final products of the different splicing reactions on a scale from 0(0) to 5(+++++). Data from figure 2 from Schmitt et al (24) and from other experiments allowed to compare all the reactions.

a good substrate for the cryptic reaction in a second splicing assay (not shown).

MODALITIES OF THE MINOR SPLICING REACTIONS OF THE E1A TRANSCRIPTS

We have previously demonstrated that the efficiency of the use of the alternative 9S, 12S or 13S 5' splice sites is strongly dependent on the ionic conditions of the reaction (24). In the standard ionic conditions (60 mM KCl and 3.2 mM MgCl₂) as defined by Krainer et al. (3), only the 13S mRNA was formed, whereas the 12S and 9S reactions were optimum in 39 mM KCl and 1.6–2.1 mM MgCl₂, the latter concentrations being routinely used here. To analyse further the molecular basis of these salt requirements, we roughly estimated the efficiencies of the minor reactions and we compared them to those of the 3 alternative major reactions [data from Fig. 2 of Schmitt et al., (24) and from other specific experiments]. As summarized in Table 1, we observed that the cryptic reactions generating the ψ12S and ψ9S mRNA from the 13S transcript were optimum at the lower salt concentrations tested (39 mM in KCl, 1.6–2.1 mM in MgCl₂) and this salt requirement was very similar to that of the corresponding 9S and 12S mRNA reactions. This indicates that the differential behaviour of the 9S–12S 5' splice sites and of the 13S 5' splice site during splicing of the Sp4 transcript is not primarily due to mutual competition, since the 13S site is absent from the 13S transcript. In contrast, the excision of the 216 nt intron exhibits a salt requirement which is intermediate between that of the 13S and that of the

TABLE 2: Relative amounts of the splicing products from different reactions.

Initial substrate	Sp4 transcript			13S mRNA transcript	
Splicing reaction	9S mRNA	638 nt mRNA	524 nt mRNA	524 nt mRNA	ψ9S mRNA
Splicing products					
IVS-E2	4	69	164.4	85	3.8
excised IVS	60.8	93.2		25	45.6
mRNA	46.8	10	60.8	26	45
IVS-E2/IVS ratio	0.07	2.50		3.40	0.08

The amounts of the different species were determined by scanning the autoradiograms shown in Figure 1 and integrating the corresponding peaks. The values for each species (in arbitrary units) were normalized relatively to the sequence content to allow a direct comparison of the splicing products for each reaction.

12 or 9S mRNA reactions. This reaction is optimal at 48 mM KCl and 2.6 mM MgCl₂, but can show a certain variability according to the nuclear extract preparation.

Finally, since we detected some anomalies in the 216 nt intron splicing reaction, we have analysed in detail the accumulation of individual products of different splicing reactions. Consequently, we compared the amounts of the IVS-E2 and of the final products of the 638 nt mRNA and 524 nt mRNA reactions, accumulated late in vitro, to those of the 9S or ψ 9S mRNA reactions (Table 2). For the 9S and ψ9S mRNA reactions, we determined that the IVS-E2 intermediate represented only 7-8% of the accumulated mRNA or of the excised intron, as expected, since this intermediate is a transitory product. In contrast, we observed that the 638 nt and 524 nt mRNA - IVS-E2 intermediates generated from the Sp4 transcript, accumulated strongly (compare amounts of species 1-2 and 9S IVS-E2 in Fig. 1A). The 638 and 524 nt mRNA - IVS-E2, taken together represented 250-300% of the 216 nt intron or of the two mRNAs (Table 2). A plausible explanation for these results is that the second step of the splicing reaction, which follows the branching, is much less efficient than the first step of the reaction, the consequence being a strong accumulation of the IVS-E2 intermediates during processing.

Another interesting aspect of these reactions is that the relative

distribution of the branched forms a, b and c of the IVS-E2 intermediates (species 1 and 2) is clearly different at the early and late times of incubation (Fig. 1A). After 30 min of incubation, form a of the 638 nt mRNA-IVS-E2 is predominant whereas both 638 nt and 524 nt mRNA-IVS-E2, branched at the most distal positions (forms b and c), accumulate at the end of incubation. Consistent with this result, only form a of the 216 nt IVS accumulate predominantly, indicating that the second step of splicing is more efficient when the branch point is located at position -51, relative to positions -55 and -59. Therefore these results suggest that the efficiency of the second step of splicing is highly dependent on the distance between the 3' splice site and the branch point.

DISCUSSION

The development of in vitro systems able to reconstitute the alternative splicing reactions observed in vivo is vital for the subsequent characterization of the trans-acting factors modulating alternative splicing reactions. Our previous analysis of an in vitro system, reconstituted from a nuclear extract of uninfected HeLa cells and which, under appropriate ionic conditions, catalysed alternative splicing of E1A transcripts similar to that observed in vivo during early times of infection, represented the primary step for the development of such systems (24). Several new features of the system, described here, reinforce this statement. First, the splicing reaction involving the 216 nt intron excision that we analysed in detail has been recently detected in infected cells (25,26). Therefore, this reaction represents a genuine pathway of the alternative processing of the E1A transcript. In vivo, the 13S and 12S mRNAs are used as substrates for the 216 nt intron excision, thus generating 2 minor mRNAs designated as 11S and 10S mRNA respectively, (25,26). In vitro, the 13S and 12S mRNA, can serve as substrates, as well as the initial Sp4 transcript, indicating that the 216 nt intron excision can precede the major 13S mRNA reaction. Moreover, both in vivo and in vitro, this reaction is minor, presumably due to the peculiarities of the branching reaction.

The second minor reaction that we have characterized is the cryptic reaction generating the ψ 12S and ψ 9S mRNA from the 13S transcript. In contrast, the newly formed 13S mRNA in the in vitro system derived from the Sp4 transcript cannot serve as a substrate for the cryptic reaction, although this same 13S mRNA does undergo the 216 nt intron excision which occurs ~ 300 nucleotides upstream of the 13S exon junction. Therefore, the

inhibition of the cryptic reaction is a direct consequence of the first 13S mRNA reaction. Interestingly, these results are reminiscent of those obtained in vivo where the ϕ 12S and ϕ 9S mRNA are formed in cells transfected with the 13S cDNA plasmid, but not in infected cells or in cells transfected with the E1A gene (27). However, the mechanisms involved in the inhibition of the cryptic reaction are unknown, and it is unclear whether they are the same in vivo and in vitro. It is conceivable that some splicing factors remain associated with the exon junction after the splicing reaction, thus blocking any further processing around this junction.

The differential salt requirement observed for the alternative splicing into the 13S, 12S and 9S mRNAs (24) is also observed for the minor reactions. Our previous data excluded the possibility that the salt variations are primarily required to favour optimal binding of snRNPs with the three E1A 5' splice sites which have different nucleotide sequences (24). Accordingly, the 216 nt intron excision and the 9S and ϕ 9S mRNA reactions, exhibit different salt optima, in spite of the fact that the same 9S 5' splice junction is involved. In fact, the parameters engaged for the determination of optimal salt conditions are unknown. It was previously noted that the excision of larger introns required lower salt concentrations (24). However, we have shown that the splicing of introns of 126 and 573 nucleotides of the human β globin premRNA, which are very similar in size to the 13S and 9S IVS, does not exhibit differential salt requirements (unpublished data).

It is noteworthy that, under appropriate ionic conditions, all the alternative or cryptic reactions occurring on E1A transcripts or mRNAs in infected or transfected HeLa cells are observed in vitro. This indicates that most of the splicing parameters engaged in vivo are reconstituted in vitro under these conditions. However, this does not necessarily mean that our or similar systems splice any premRNA efficiently, particularly those exhibiting unusual characteristics. For instance, it has been shown that splicing of genuine SV40 small t mRNA is highly inefficient in vitro (38,39) and that in vitro processing of human growth hormone transcripts raises difficulties (40).

Concerning the excision of the 216 nt intron, we demonstrate that this reaction occurs via a branching at one of the 3 adenosine residues, located 51, 55 and 59 residues upstream of the 3' splice site. This distance is abnormally long if we consider that all the authentic or cryptic branch points have been mapped, in higher eukaryotes, between 18 and 37 nucleo-

tides upstream of the 3' terminal AG of the introns (for reviews, 10,11). This is the first example of branch points so far from the 3' splice site. In addition, the authentic branch site of β globin small intron, when moved from 37 to 49 nucleotides upstream of the 3' end of intron, was completely inactivated suggesting the existence of a strong distance constraint for the branch site selection (41). In the 216 nt IVS, an A residue surrounded by 3 G residues (GGAG) is present 23 nucleotides upstream of the 3' splice site. However, we have detected neither the corresponding lariat intermediate, nor a branching at this A residue by primer extension analysis of the total spliced RNA (not shown). The GGAG sequence was not recognized either as a 3' splice site. The absence of adenosines between residues -24 and -50 might explain why branching occurs so far from the 3' splice site. This is in agreement with the notion that the choice of an adenosine is a very strong requirement for the selection of branching site, since only one exception has been observed for splicing of genuine premRNA in vivo (Hartmuth and Barta, personal communication). That this requirement is not absolute in vitro, is also shown by using modified premRNAs, which are mutated in the branch sequence (42).

Another peculiarity of the 216 nt intron excision reaction is the strong accumulation of the IVS-E2 intermediates during the course of incubation, indicating that the second step of splicing is less efficient than the first one. Such accumulation has been noted by other groups, under conditions where the second step of splicing is inferred to be inhibited or strongly decreased (37,42-45)). In addition, we show that, the farther the branch point is from the 3' end of the intron, the greater is the accumulation of the corresponding lariat intermediate. Therefore, we suggest that the distance between the branch point and the 3' splice site is more critical for the second step than for the first step of splicing, a fact which was previously unknown for higher eukaryotes. Similar conclusions have been reached for a yeast system (43), although the 2 steps of splicing in higher and lower eukaryotes do not have exactly the same requirements (46,47).

Taken together, our results and another set of data demonstrate more generally that the second step of splicing has stricter requirements with respect to branch point position and branch structure than the branch reaction itself. For example, it was shown that, although alteration of the GT residues at the 5' end of intron, or mutation of the A branch residue does not have a dramatic effect on lariat formation, these modifications completely inhibit intron release (42,48,49). Therefore, the branch structure

remains a very important element in the second step of splicing, although the chemical reactions are theoretically restricted to the 3' splice junction for its cleavage and to the extremities of both exons for their ligation.

A last unusual aspect of the 216 nt intron reaction is the existence of 3 potential branch points used with an efficiency varying from 1 to 10 approximately. The first analyses of branching reaction showed the existence of a single site (5,50,51). However, more recent studies often suggest the use of additional minor sites (24,41,52,53). It has been also shown that the use of multiple and differential branch points is engaged for the alternative splicing of SV40 early premRNA to large T and small t mRNAs (54). However, differential branching should not necessarily represent a general feature of premRNA in which alternative 5' splice sites are spliced to a common 3' splice site, as proposed (54), since this is not the case for the classical alternative E1A premRNA splicing (24).

The excision of the 216 nt intron can be considered as an alternative pathway since this reaction is optional, and its poor efficiency might be primarily related to the unusual characteristics of the branching reaction. By using this splicing reaction as a model, we are currently analysing how the parameters of the primary structure and of the spacing between branch point and 3' splice site are involved, first in the selection of the branching sites and second in the efficiency of the 2 steps of the splicing reaction.

ACKNOWLEDGMENTS

We are grateful to Mr. G. Hildwein for competent technical assistance. We thank Mr. J.L. Weickert and Mrs. M. Gilbert for growing HeLa cells, and Mr. A. Staub for the synthesis of oligodeoxynucleotide primer. We also thank Drs. M. Jacob and J. White for critically reading the manuscript. We thank B. Boulay and C. Werlé for the illustrations, and the secretarial staff for typing the manuscript.

This work was supported by grants from CNRS, INSERM, the Association pour la Recherche sur le Cancer and the Fondation pour la Recherche Médicale.

REFERENCES

1. Padgett, R.A., Hardy, S.F. and Sharp, P.A. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5230-5234.
2. Hernandez, N. and Keller, W. (1983) *Cell*, **35**, 89-99.
3. Krainer, A.R., Maniatis, T., Ruskin, B. and Green, M.R. (1984) *Cell*, **36**, 993-1005.
4. Grabowski, P.J., Padgett, R.A. and Sharp, P.A. (1984) *Cell*, **37**, 415-427.
5. Ruskin, B., Krainer, A.R., Maniatis, T. and Green, M.R. (1984) *Cell*, **38**, 317-331.

6. Padgett, R.A., Konarska, M.M., Grabowski, P.J., Hardy, S.F., Sharp, P.A. (1984) *Science*, **225**, 898-903.
7. Frendewey, D. and Keller, W. (1985) *Cell*, **42**, 355-367.
8. Grabowski, P.J., Seiler, S.R. and Sharp, P.A. (1985) *Cell*, **42**, 345-353.
9. Brody, E. and Abelson, J. (1985) *Science*, **228**, 963-967.
10. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) *Ann. Rev. Biochem.* **55**, 1119-1150.
11. Green, M.R. (1986) *Ann. Rev. Genet.*, **20**, 671-708.
12. Zhuang, Y. and Weiner, A.M. (1986) *Cell*, **46**, 827-835.
13. Tsurushita, N. and Korn, L.J. (1987) *Mol. Cell. Biol.*, **7**, 2602-2605.
14. Kühne, T., Wieringa, B., Reiser, J. and Weissmann, C. (1983) *EMBO J.*, **2**, 727-733.
15. Lang, K.M. and Spritz, R.A. (1983) *Science*, **220**, 1351-1355.
16. Reed, R. and Maniatis, T. (1986) *Cell*, **46**, 681-690.
17. Eperon, L.P., Estibeiro, J.P. and Eperon, I.C. (1986) *Nature*, **324**, 280-282.
18. Solnick, D. (1985) *Cell*, **43**, 667-676.
19. Leff, S.E., Rosenfeld, M.G. and Evans, R.M. (1986) *Ann. Rev. Biochem.*, **55**, 1091-1117.
20. Breitbart, R.E., Andreadis, A. and Nadal-Ginard, B. (1987) *Ann. Rev. Biochem.*, **56**, 467-495.
21. Rozek, C.E., Davidson, N. (1983) *Cell*, **32**, 23-34.
22. Leff, S.E., Evans, R.M. and Rosenfeld, M.G. (1987) *Cell*, **48**, 517-524.
23. Breitbart, R.E. and Nadal-Ginard, B. (1987) *Cell*, **49**, 793-803.
24. Schmitt, P., Gattoni, R., Keohavong, P. and Stévenin, J. (1987) *Cell*, **50**, 31-39.
25. Ulfendahl, P.J., Linder, S., Kreivi, J.P., Nordqvist, K., Svensson, C., Hultberg, H. and Akusjärvi, G. (1987) *EMBO J.*, **6**, 2037-2044.
26. Stephens, C. and Harlow, E. (1987) *EMBO J.*, **6**, 2027-2035.
27. Svensson, C., Petersson, U. and Akusjärvi, G. (1983) *J. Mol. Biol.*, **165**, 475-499.
28. Hurwitz, D.R. and Chinnadurai, G. (1985) *J. Virol.*, **54**, 358-363.
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. Konarska, M.M., Padgett, R.A. and Sharp, P.A. (1984) *Cell*, **38**, 731-736.
31. Green, M.R., Maniatis, T. and Melton, D.A. (1983) *Cell*, **32**, 681-694.
32. Maxam, A. and Gilbert, W. (1980) *Meth. Enzymol.*, **65**, 499-560.
33. Ruskin, B. and Green, M.R. (1985) *Science*, **229**, 135-140.
34. Keohavong, P., Gattoni, R., LeMoullec, J.M., Jacob, M. and Stévenin, J. (1982) *Nucl. Acids Res.*, **10**, 1215-1229.
35. Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) *Nucl. Acids Res.*, **11**, 1475-1489.
36. Mount, S.M. (1982) *Nucl. Acids Res.*, **10**, 459-472.
37. Jacquier, A. and Rosbash, M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5835-5839.
38. Noble, J.C.S., Prives, C. and Manley, J.L. (1986) *Nucl. Acids Res.*, **14**, 1219-1235.
39. van Santen, V.L. and Spritz, R.A. (1986) *Nucl. Acids Res.*, **14**, 9911-9926.
40. Hartmuth, K. and Barta, A. (1987) *Nucl. Acids Res.*, **15**, 7005-7025.
41. Ruskin, B., Greene, J.M. and Green, M.R. (1985) *Cell*, **41**, 833-844.
42. Hornig, H., Aebi, M. and Weissmann, C. (1986) *Nature*, **324**, 589-591.
43. Cellini, A., Felder, E. and Rossi, J.J. (1986) *EMBO J.*, **5**, 1023-1030.
44. Chabot, B. and Steitz, J.A. (1987) *Mol. Cell. Biol.*, **7**, 281-293.

45. Parent, A., Zeitlin, S. and Efstratiadis, A. (1987) J. Biol. Chem. 262, 11284-11291.
46. Rymond, B.C. and Rosbash, M. (1985) Nature, 317, 735-737.
47. Rymond, B.C., Torrey, D.D. and Rosbash, M. (1987) Genes and Development, 1, 238-246.
48. Aebi, M., Hornig, H., Padgett, R.A., Reiser, J. and Weissmann, C. (1986) Cell, 47, 555-565.
49. Freyer, G.A., Arenas, J., Perkins, K.K., Furneaux, H.M., Pick, L., Young, B., Roberts, R.J. and Hurwitz, J. (1987) J. Biol. Chem., 262, 4267-4273.
50. Zeitlin, S. and Efstratiadis, A. (1984) Cell, 39, 589-602.
51. Konarska, M.M., Grabowski, P.J., Padgett, R.A. and Sharp, P.A. (1985) Nature, 313, 552-557.
52. Reed, R. and Maniatis, T. (1985) Cell, 41, 95-105.
53. Keohavong, P., Gattoni, R., Schmitt, P. and Stévenin, J. (1986) Nucl. Acids. Res., 14, 5207-5227.
54. Noble, J.C.S., Pan, Z.Q., Prives, C. and Manley, J.L. (1987) Cell, 50, 227-236.