The different intron 2 species excised *in vivo* from the E2A premRNA of adenovirus-2: an approach to analyse alternative splicing

Phouthone Keohavong, Renata Gattoni, Philippe Schmitt and James Stévenin

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

Received 4 April 1986; Revised and Accepted 10 June 1986

ABSTRACT

In the early period of cellular infection by adenovirus 2, the E2A region gives rise to 2 major mRNA species of 2.0 and 2.3 kilobases, formed by alternative excisions of intron 2 (Gattoni et al., 1986, J. Mol. Biol. 187, 379-307). We have analysed the excision pathways of this intron. Two major intron species of 626 and 337 nucleotides, generated by the use of 2 consensus 3' splicing sites and a minor intron species of 520 nucleotides, generated by the use of another weaker 3' splicing site, are identified, the 3 species sharing a common 5' splicing site. They are detected predominantly in the lariat form. For the 2 major species we analyzed, the branched nucleotides are localized at consensus branching sequences, 26 or 25 nucleotides upstream from the 3' terminal AG. Our results confirm that the first reactions of cleavage at the 5' end of introns and branching occur *in vivo* as described in *in vitro* systems. The second predominant form of intron 2 is the linear segment, whereas the nicked lariat form which is very minor, might not be a genuine product of *in vivo* splicing. All intron 2 molecules show practically intact 5' and 3' terminal sequences, indicating that they are well protected against nuclease attack throughout their life. Therefore, these results indicate that the primary reaction following the excision of the lariat intron is debranching.

In addition, the existence of a potential 5' splicing site contiguous to the major internal 3' splicing site raised the possibility of an elimination of the major 626 nucleotide intron in 2 cycles of excision. However, we demonstrate that intron 2 is systematically excised by a one cycle process, which is likely to represent the general rule for the production of correctly spliced mRNA.

INTRODUCTION

The recent development of *in vitro* cell free systems that accurately splice premRNA has allowed significant progress in the knowledge of the detailed mechanisms of splicing (1-4). It has been shown that the first cleavage reaction occurs at the 5' end of the intron, which is then joined by a 2'-5' phosphodiester bond to an adenosine residue near the 3' end of the intron. A second cleavage at the 3' end of the intron releases the free intron sequence in a lariat form which accumulates in *in vitro* systems (5-7).

It is important to determine whether identical mechanisms are involved in in vivo splicing. However, it has been very difficult to obtain exhaustive data on the complete set of splicing intermediates for a given premRNA due to their complexity and low abundance (for review 8,9-11). Moreover, the primary cleavage occurring at the 5' limit of the intronic sequences separated the premRNA or ligation intermediates into a 5' part, extending from the cap to the cleavage site and a 3' part extending from the cleaved 5' end of the intron to the poly A tail, and we have designated these as 5' and 3' cleavage intermediates, respectively (12). Recently, 5' and (or) 3' cleavage intermediates have been characterized in vivo indicating that the first splicing reaction is the cleavage at the 5' end of intron found in in vitro systems (13,14). However, because such cleavage intermediates are transitory products, they do not systematically accumulate at a sufficient level to allow easy characterization. This was the case for the splicing intermediates of the E2A premRNA of adenovirus (12).

The free introns, which are one of the final splicing products and which should accumulate in the nuclear poly (A)⁻ RNA, have been detected only recently (12,15,16), and the only relatively detailed analysis has been presented by Zeitlin and Efstratiadis (13). Since free introns reflect faithfully the earlier splicing reactions, their direct analysis from the poly(A)⁻ RNA population in conjunction with a study of the poly(A)⁺ splicing intermediates will yield complementary data concerning the intimate splicing mechanisms (12). In addition, such an approach may be helpful for the understanding of alternative splicing as well as determining whether the excision of a given intron occurs in one or several cycles, an excision cycle being defined as the steps of cleavage at the 5' end of intron and branching reaction, followed by the excision of intron and ligation of exons.

The previous analysis of the poly(A)⁺ splicing intermediates of the adenovirus E2A transcription unit indicated that these points are particularly relevant to intron 2 excision (12). For these reasons, we have carried out an exhaustive characterization of the excised intron 2 sequences.

MATERIALS AND METHODS

Cell culture, virus infection and RNA purification

HeLa S3 cells were grown in suspension culture and infected with adenovirus-2 (100 plaques forming units/cell), as described previously (17). Cycloheximide was added at 25 µg/ml 3 hr post-infection and the cells were harvested at 4-5 hr post infection (12,18). In order to follow the RNA during

the first steps of preparation and purification, they were labeled during the last 30 min with 10 μ Ci of [3 H] uridine/ml (10 Ci/mmol). RNAs were prepared from isolated nuclei by the hot phenol/SDS procedure and the contaminating DNA was removed as described by Gattoni et al (12). The nuclear poly(A)⁻ RNA was purified by repeated passage over a poly(U) sepharose column (17). The poly (A)⁻ RNA was denatured in TES (5 mM Tris HCl pH 7.4, 2.5 mM EDTA, SDS 0.5%) at 65°C for 5 min, then fractionated in a 5 to 20% sucrose gradient made in TES buffer by centrifugation at 27 000 rpm for 17 hr at 20°C in a SW27 rotor.

RNA analysis

For Northern analysis, RNAs were electrophoresed on urea polyacrylamide gel as described by Maxam and Gilbert (19) and then transferred electrophoretically onto Z-binding membranes (Schleicher and Schüll) using the electroblotting unit of LKB. The prehybridization and hybridization with labeled DNA probes were as described by Thomas (20), but both mediums contained 1% SDS instead of 0.1%. The probes used were : the nick- translated subcloned BalI-PvuII fragment (positions 3017-2406) containing the entire intron 2 except for the first 28 bp, the 5' end-labeled Z-PvuII fragment (positions 2704-2406, Z is the terminus of a subfragment of the first probe) specific for the 5' half of intron 2, the 5' end-labeled BalI-PstI fragment (positions 3017-2677) specific for the 3' half of this intron. A 5' end- labeled cDNA probe complementary to both leaders and to the terminal portion of the 2.0 kb E2A mRNA body (up to the XhoI site, position 3168), as previously described (12) was also used.

Isolation of specific intron 2 RNA species

In order to further characterize each of the RNA components containing intron 2 sequences detected by Northern blot analysis, they were separated from the 10S-6S gradient fractions on a preparative 5% urea-acrylamide gel, in parallel with labeled DNA markers. Gel slices were excised by reference to the marker positions and the RNA was recovered by electroelution from the gel. Using this procedure it was necessary to excise relatively broad gel slices and consequently, only components migrating with abnormally low mobility (the lariat forms) were expected to be highly purified. The other discrete components should be isolated together with the RNA sequences which migrated as a smear in the gel (see Fig. 2).

Debranching reaction

These assays were performed according to Ruskin and Green (21) by using the S100 fraction obtained from HeLa cells (22). Debranching assays of the

RNA were in 20 mM Hepes pH 8.0, 20% glycerol, 100 mM KCl, 0.5 mM DTT and 20% (v/v) of S100 fraction. Incubation was at 30°C for 30 min. The products were deproteinized (17) and analysed by Northern blot as described above.

Primer extension analysis

RNA species previously isolated were reverse transcribed by using 19-20 residue synthetic oligonucleotides complementary to the expected major 3' end of the introns. RNA was precipitated with an excess of 5' end-labeled oligonucleotides (0.2 to 0.5 ng). The pellet was redissolved in 10 μ l of NPES (250 mM NaCl, 40 mM Pipes pH 6.4, 1 mM EDTA, 0.2% SDS), denatured at 65°C for 3 min and hybridized overnight by slowly cooling to room temperature. After ethanol precipitation, reverse transcription was carried out in 20 μ l of medium containing 50 mM Tris HCl pH 8.0, 8 mM MgCl₂, 80 mM NaCl, 5 mM DTT, 0.5 mM of each of the dNTP and 3 U of reverse transcriptase (BRL) for 1 hr at increasing temperature from 25°C to 42°C. Products were ethanol precipitated and analysed on 15% polyacrylamide sequencing gels.

Labeling of DNA fragments

DNA was nick-translated according to Rigby et al. (23) in the presence of [α^{32} P] dNTP (800 Ci/mmol, Amersham). The 5' end-labeled DNA fragments or oligonucleotides were prepared as described by Maxam and Gilbert (19) in the presence of [γ^{32} P] ATP (500 Ci/mmol, Amersham). The 3' end-labeled DNA fragments were prepared as described by Goodman (24) in the presence of 1 or 2 [α^{32} P] dNTP (800 Ci/mmol). Finally, strand separation of the resulting fragments was according to the method of Maxam and Gilbert (19). After electrophoresis, the bands were electrophoretically eluted.

S1 nuclease mapping

S1 nuclease analysis of RNAs was as described by Weaver and Weissmann (25). RNA samples and 5' or 3' end-labeled DNA probes (5000 to 10000 cpm) were hybridized as described by Casey and Davidson (26). The samples were then treated with 100-150 U of S1 nuclease (BRL) at 20°C for 1 hr. The protected fragments were analysed on a 8% polyacrylamide sequencing gel. In order to precisely localise the S1 cleavage sites, the 5' end-labeled homologous probes or non coding strands of the 3' end-labeled probes were cleaved in Maxam and Gilbert G+A or C+T reactions and co-electrophoresed in the gel.

RESULTS

The splicing events which generate the mRNA species of 2.0 and 2.3 kb from the E2A premRNA have been determined earlier by an exhaustive analysis of the poly(A)⁺ splicing intermediates (12) (see Fig.1). This analysis pro-

Figure 1 : Map of the E2A transcription unit of adenovirus 2.

The transcription of the E2A unit is in the leftward direction, starting from the major initiating residue, designated as residue +1. The exons are noted by boxes, the introns by lines. At the top of the figure are shown the 5' part of the major 2.0 kb and of the minor alternative 2.3 kb mRNAs containing the two leaders (Lea 1 and 2) ligated to the mRNA body. The bottom of the figure represents an enlargement of intron 2. The nucleotide sequences for the alternative 3' splice site (SS) and for the contiguous possible 5' splice site (12), as well as the possible sequences excised from intron 2, are given.

vided evidence that the large intron 1 of 2233 nucleotides (nt) is primarily removed in one excision cycle, despite the presence of more than a dozen potential splicing sites within the intron. In contrast, we showed that intron 2 was subject to alternative splicing events, linking a common 5' splicing site to either of two 3' splicing sites (at positions 3004 and 2715), and removing segments of 626 or 337 nt (Fig. 1). In addition, splicing using 3' site at 2715 generates a potential 5' splicing site at position 2176, making possible the excision of the entire 626 nt segment in 2 cycles : a first excision of the 5' 337 nt segment, followed by a second excision of the remaining 3' 289 nt segment. Therefore, for this analysis we have kept in mind the possible existence of at least 3 different intron segments of 626, 337 and 289 nt.

In this study, we have taken advantage of the dramatic stimulation of transcription of the E2A unit and the high level of the E2A splicing products obtained after a limited treatment of infected HeLa cells with cycloheximide. This treatment results in essentially the same poly(A)⁺ splicing products as in untreated cells (12). These conditions facilitate the analysis of very minor species which may be undetectable using standard conditions.

Figure 2 : Northern blot analysis of the intron 2 - hybridizing RNA species.

Nuclear poly(A)- RNA (Nu) or gradient fractionated RNA of 12S to 6S from infected cells were analysed by probing RNA transfer blots with various ^{32}P -labeled DNA fragments. The scheme at the bottom shows the content of the DNA probes. RNA transfer blots are hybridized :

- In A and B : with the BalI-PvuII entire intron 2 fragment, ^{32}P labeled by nick translation;
- in C : with the ^{32}P end-labeled BalI-PstI fragment, specific for the 3' part of the intron;
- in D : with the ^{32}P end-labeled Z-PvuII fragment, specific for the 5' part of the intron (Z is a cloning site). M : MspI-TaqI fragments of pBR322 DNA as size markers. The apparent size of slower migrating RNA was estimated from 28 and 18S rRNA of HeLa cells. Bd : Band.

Characterization of the intron 2 products by Northern blot experiments.

The nuclear poly(A)⁻ RNA was examined by high resolution Northern analysis using electroblotting from a denaturing 5% acrylamide gel (Fig. 2). With intron 2 specific probes, we detected a series of discrete bands, which were masked by a strong smear of RNAs from 2500 to 500 nt in size (Fig. 2, B-C-D, lane Nu). The smear represents presumably a mixture of growing RNA transcripts and degraded RNA and precludes a clear cut analysis of the discrete intron species. Therefore, the nuclear poly(A)⁻ RNA was fractionated by centrifugation on a sucrose gradient and the fractions between 12S and 6S, where RNA molecules of about 800-300 nt in size are expected to migrate, were examined by Northern blot analysis by using a DNA probe specific for the 626 nt intron 2 (Fig. 2A). With the fractionation procedure we observed a series of 8 RNA bands (designated from 1 to 8) which could be separated into 2 subsets. The RNA bands 1, 2, 4, 5 were concentrated around the 9S fraction, the RNA of band 2, present in traces, being visible in the 9S and 7,5S fractions on the original figure. The RNA bands 3, 6, 7, 8 were concentrated around the 6S fraction. Since the size separation on the sucrose gradient was normal, as indicated by the fractionation of the RNA smear, it was obvious that several components (at least those of bands 1, 2 and 3) had an abnormal mobility in the polyacrylamide gel analysis. This suggests that they have a non linear structure (5,6).

In order to further analyze the different intron 2 products, the 9S and 6S RNA fractions have been examined in parallel with probes specific for the whole, the 5' or the 3' part of the 626 nt intron, the latter overlapping the 5' part on 38 residues (see the bottom of fig. 2). From the 9S fraction, the RNA species of bands 1, 2, 4 and 5, (with an apparent size of 4200, 2900, 850 and 630 nt, respectively) were detected by each of the 3 DNA probes with the same relative efficiency (Fig. 2B, D, C, lanes 9S). This suggests that they contain the totality or a large part of the whole intron 2 of 626 nt.

The other species enriched in the 6S RNA fraction had different hybridization characteristics. The components of bands 3 and 7 (with an apparent size of 1740 and 335 nt), hybridized strongly with the probes specific for the whole, or the 5' part of the 626 nt intron (Fig. 2B, D, lanes 6S) and only weakly with the probe specific for the 3' part (Fig. 2,C). It is likely that these species represent the 337 nt sequences. The two other components from bands 6 and 8 (of apparent size of 440 and 325 nt) were revealed as minor species by the 626 nt intron 2 probe (Fig. 2B, lanes 6S), but gave a stronger signal with the 3' part specific probe (Fig. 2C) and none at all

Table 1

RNA of Band	app. size (nt)	size after debranching (nt)	app. size after heating (nt)	Identity of the RNA
1	4200	630	850	626 nt lariat Int.
2	2900	520	N.D.	520 nt lariat Int.
3	1740	335	425	337 nt lariat Int.
4	850	< 630	< 850	626 nt nicked lariat Int.
5	630	630	< 630	626 nt linear Int.
6	440	440	N.D.	440 nt linear seq.
7	335	335	N.D.	337 nt linear Int.
8	325	325	N.D.	325 nt linear seq.

Summary of data derived from debranching reaction and heat treatment. The isolated RNA from band 1 to 7-8 were treated and analysed as described in fig. 3. A summary of the results is given in table 1.

N.D. : non detectable

with the 5' part specific probe (Fig. 2D). This indicates that they contain primarily sequences of the 3' half of intron 2.

Non linear RNA molecules display an abnormal mobility which is variable in gels of different concentrations of polyacrylamide (5,6). By analysing the 9S and 6S RNA fractions in more concentrated gels (6% and 8% acrylamide instead of 5% as in Fig. 2), we found that RNA of bands 1, 2 and 3, as expected, as well as the RNA of band 4 showed higher apparent sizes (not shown). At the same time, the RNA of bands 5 to 8 showed the same apparent sizes, suggesting that they are linear.

Our previous results concerning the analysis of the splicing intermediates of the E2A premRNA predicted that the excision of intron 2 would generate at least segments of 626 nt or 337 nt (12). Therefore, the data of Fig. 2 were consistent with the possibility that the RNAs of bands 1 and 5 are a non linear form (presumably a lariat) and the linear form of the whole intron 2 of 626 nt, whereas the RNAs of bands 3 and 8 are the lariat and the linear forms of the alternative 337 nt intron 2, respectively (see also Table 1). However, the nature of the other species was less clear and their determination required further experiments.

Conformational analysis of the intron 2 species by debranching reaction and heat treatment

To further elucidate the structure of the intron 2 species, aliquots corresponding to the bands isolated following preparative gel electrophoresis (see Materials and Methods) were submitted to the lariat debranching activity present in HeLa cell extract (21,27) which will transform branched structures

Figure 3 : Northern blot analysis of the isolated RNA species after heat-treatment and debranching reactions.

Isolated RNA species from bands 1 to 7-8 (B1 to B7-8 in the figure) were submitted to the debranching reaction (De) or heat-treated (He) as described in Materials and Methods. Untreated RNA used as control (Co) or treated RNA were electrophoresed on a 5% acrylamide gel, then analysed by probing the RNA transfer blots with the ^{32}P nick-translated BalI-PvuII probe (see Fig. 2). The first lane of each panel (9S or 6S) corresponds to an untreated RNA fraction of 9S or 6S. M : markers from MspI and TaqI cleaved pBR322.

into the corresponding linear sequences (13). In parallel a second aliquot was submitted to a heat-treatment in conditions that permitted at most one nick for each RNA molecule (28) and thus transformed a lariat structure into a nicked lariat form and a nicked lariat structure into a smaller one.

Figure 3 shows the Northern blot analysis of the isolated RNA species from bands 1 to 8 before (control) and after these treatments, and table 1 gives a summary of these results. The RNA of the 9S and 6S gradient fractions was co-electrophoresed as control. The RNA of band 1 (Fig. 3, B1), was partly converted by heat-treatment (lane He) into a discrete component with an apparent size of 850 nt identical to that of band 4, whereas it was quantitatively transformed by the debranching reaction (lane De) into an RNA of 630

nt, a size identical to that of the entire linear intron 2. RNA of band 4 (Fig. 3, B4) was partly converted by heat-treatment (lane He) into a smear of sequences migrating faster than untreated RNA. In contrast, after the debranching reaction (lane De), the RNA was completely transformed into a smear of molecules migrating mainly as RNA segments between 550 and 450 nt, particularly visible in an over exposed autoradiogram. These results, and the conversion of band 1 material into band 4 after heat treatment clearly show that the RNA of band 4 represents the nicked lariat form of the entire intron 2. Moreover, although this form behaves as a very discrete RNA during the gel analysis, the heterogeneity of the debranching products of the RNA of band 4 (Fig.3) indicates that the opening of the 626 nt lariat, occurring in vivo or during the isolation procedure, is at random on the lariat. The RNA of band 4 was also isolated together with a minor discrete RNA with an apparent size of 810 nt (Fig. 3, B4, lane Co). This latter was insensitive to the debranching reaction (lane De), indicating that it represents a linear RNA and it was not further characterized. The RNA of band 5 (Fig. 3, B5, Co), expected to represent the linear form of the entire intron 2, was indeed insensitive to the debranching treatment and was partly transformed into a smear of RNA fragments after heat treatment. Therefore, the excised 626 nt intron 2 is detected in the nuclear poly(A)- RNA in its lariat, nicked lariat, and linear forms (see Table 1) and their significance will be discussed later.

The minor RNA isolated from band 2, which represents between 1 and 2% of the 626 nt lariat intron contains a weak contamination of an RNA migrating as that of band 4 (Fig. 3, B2, lane Co). This contamination is due to a minor but continuous cleavage of the 626 nt lariat RNA of band 1 during the course of electrophoresis, involving the presence of a very small amount of 626 nt nicked lariat in the gel between bands 1 and 4. The heat treatment of the RNA of band 2 does not lead to the appearance of a smaller discrete band, most likely because the amount of this species remains minor relative to the other species (Fig. 3, B2, He). However, the debranching treatment (lane De) completely linearized the species into a discrete 520 nt RNA segment, indicating the existence of an unexpected intron 2 lariat whose terminal sequences will be precised by S1 nuclease mapping below.

The analysis of the species accumulating in the 6S RNA fraction is given in the second part of fig. 3. The species from band 3 (the presumed 337 nt lariat form) was quantitatively transformed after the debranching reaction into a linear form of 335 nt (B3, De), identical to the RNA of band 7. The heat treatment partially transformed RNA of band 3 into a doublet with an

apparent size around 425 nt, which should be the opened lariat form of the 337 nt intron. This form was not, or only very weakly, detected in the 6 S RNA fraction, even with the 5' part specific probe of intron 2 (Fig. 2D). Finally, the 3 remaining RNA of bands 6 to 8 (Fig. 3, B6 and B7-8) all behaved as linear sequences. They were insensitive to the debranching reaction (De), in conditions controlled by the addition of the 626 nt lariat, and the heat treatment did not generate new discrete RNA species.

Therefore, the results of fig. 2 and 3 (as summarized in Table 1) indicate clearly that 6 RNA species from the 8 containing intron 2 sequences represent products of the splicing of intron 2 and they will be further characterized below. The 2 components of bands 6 and 8, other than containing sequences from the 3' part of intron 2 have none of characteristics expected for the lariat or linear forms of the anticipated 289 nt intron. By using a cDNA probe containing exonic sequences surrounding intron 2, Northern analysis of the 6S RNA fraction reveals a series of discrete bands of 600 to 200 nt, the strongest having the same mobility as band 8. Therefore, RNA of bands 6 and 8 are products generated by a non random cleavage of the E2A RNAs presumably non related with splicing events. The fact that they are systematically contaminated by a background of E2A RNA sequences made their further characterization problematic.

Localization of the branch points by primer extension analysis

The structure of the intron 2 species has been analyzed by primer extension, particularly so as to determine their branch sites, since such sites have been mapped only in a limited number of cases for the *in vivo* splicing intermediates in higher eukaryotes (13,29). We have focused our attention on the branch reactions of the major 626 and 337 nt intron 2, using 2 synthetic primers of 19 and 20 residues, starting precisely from the expected 3' ends of the introns (see bottom of fig. 4). The gel analysis of the extended cDNA sequences from both primers is given in fig. 4A and B. In the absence of viral specific RNA, none of the 2 primers were extended (lanes a and f). With the 9S RNA fraction (Fig. 4A, lane b) a 25-26 residue segment, generated from a stop after 6-7 nucleotides, as well as a faint band of 630 residues were observed. The shorter cDNA segments were specifically synthesized using as template the purified lariat form (lane c) and the nicked lariat form (lane d) of the 626 nt intron. Only a 630 residue cDNA was obtained with the 626 nt linear intron (Fig. 4A, lane e). Consequently, the specific block of the cDNA synthesis after 6-7 residues is due to the presence of branching and it maps the branch point at an A (marked with an asterisk) in the following sequence

Figure 4 : Primer-extension analysis of the intron 2 RNA species.

RNA fractions or isolated RNA species were used as templates for primer-extension reactions using the 19 residue primer complementary to the 3' end of the 626 nt intron in (A), or the 20 residue primer complementary to the 3' end of the 337 nt intron in (B). The primers are shown at the bottom of the figure. Extension reactions were in the absence of virus specific RNA (lanes a and f), or in the presence of the 9S RNA fraction (lanes b and g), the 626 nt lariat RNA (lanes c and h), the 626 nt opened lariat RNA (lane d), the 626 nt linear RNA (lanes e and i), the 6S RNA fraction (lane j), the 337 nt lariat RNA (lane k) and the 337 nt linear RNA (lane l). M : markers from *MspI* cleaved pBR322. Pr indicates the position of the primers in the gel. The portion of the gel between 40 and 100 residues has been cut out.

5' GACTG^{*}AC 3' located 26 nucleotides upstream from the 3' end of the intron. The stop at the C residue (see also below) is likely due to termination of the cDNA at the penultimate position.

Similar experiments were carried out with the 20 residue primer (Fig. 4B). Primer extension using the 6S RNA fraction (lane j) generated a block product of 24-25 residues and a cDNA of 340 residues. The latter, which is synthesized by elongation of the primer up to the common 5' splice site of intron 2, was observed with the 9S RNA fraction (lane g), the purified lariat (lane h) and the linear (lane i) forms of the 626 nt intron 2, and with the 337 nt linear species (lane l) (Fig. 4B). In fact, the synthesis of a discrete cDNA band of 340 residues with the 626 nt lariat and the linear forms of both 626 and 337 nt introns indicates strongly that the latter accumulate primarily with intact 5' terminal sequences.

The 24-25 residue segment was specifically obtained with the lariat form of the 337 nt intron 2 (Fig. 4, lane k) and resulted from a synthesis block 4-5 residues upstream from the primer. The 25 residue cDNA maps the branch point at an A residue in the following sequence : 5' CTTCA^{*}AC 3' at position 2690 of the E2A transcription unit, a stop at the penultimate position being also observed.

Therefore, it appears that the branch points occur at 26 and 25 residues upstream from the 3' ends of the 626 and 337 nt introns, respectively. The sequences around these branch points fit the putative consensus sequence PyXPyPyPuAPy (13), except for the first G of the first branching sequence. We were unable to localize the branching site of the 520 nt intron 2 with the low concentration of material available, but potential branching sequences are localized 24 and 42 residues upstream the end of this intron.

S1 nuclease analysis of the intron 2 sequences.

In order to determine the terminal sequences of the isolated intron species at the nucleotide level, we undertook S1 mapping experiments using the probes which are given at the bottom of fig. 5. The 5' termini of the intron 2 species were mapped by the 510 residue FokI-FokI fragment (positions 2090-2600), 5' end-labeled at position 2600 (Fig. 5A). The 626 nt or the 337 nt linear species protected this probe as a sequence of 222 residues (lanes b, e). When compared with the G+A and C+T ladders (lanes G/A, C/T) migrating in parallel, and after taking into account the shift of 1-2 residues in mobility for the sequences chemically cleaved (12,25,30), we map the S1 cleavage precisely at the terminal G residue (position 2379) of the 5' splicing site of intron 2. With the same probe, the 626 nt lariat (lane a),

Figure 5 : S1 mapping analysis of the isolated intron 2 RNA species.

The extent of the DNA probes is given at the bottom of the figure and the star indicates the labeled end of the probe.

(A) : The 5' terminal sequences of the intron 2 RNA species were analysed with the FokI-FokI probe;

(B) : The 3' terminal sequences of the 626 and 520 nt introns analysed with the DdeI-BalI probe;

(C) : The 3' terminal sequences of the 337 nt intron are analysed with the FokI-BalI probe. Positions of the untreated probes (Pr ►) are shown in the figure and an aliquot of the intact probe is electrophoresed in C (lane 1). Lane m represents the S1 nuclease digestion of the probe without addition of RNA. RNAs analysed are : the 626 nt lariat (lanes a, g, h), the 626 nt opened lariat (lanes c, j), the 626 nt linear species (lanes b, i), the 520 nt lariat species (lanes f, k), the 337 nt lariat (lanes d, n) and the 337 nt linear species (lanes e, o). The size of the protected DNA fragments (►) is indicated. Lanes designated as G/A and C/T show the 5' end-labeled probe (in A) or the complementary strand of the probe (in B and C) sequenced by the method of Maxam and Gilbert (1980). M : MspI fragments of pBR322 DNA as size markers.

the 626 nt opened lariat (lane c) and the 337 nt lariat (lane d) species generated a 222-224 residue triplet with the middle band as the most intense (Fig. 5A). This could be due to a steric protection of the DNA probe on one or two additional bases at the proximity of the lariat branch point, by comparison with the linear forms. With the 626 nt linear or opened lariat species and with the 337 nt linear or lariat species (lanes b, c, d, e, respectively), entire DNA probe and (or) faint bands of 320-300 residues were also obtained. These signals are likely generated by some degraded E2A products (see methods) already detected in the isolated RNA in Fig. 3.

The faint 520 nt lariat species (Fig. 2 band 2), whose precise 5' and 3' terminal sequences were unknown, generated with the same probe exactly the same protected sequence of 222-224 residues (Fig. 5, lane f) as the 626 nt lariat species run in parallel (lane g). Consequently, these results show that all of the 6 tested intron 2 RNA species share the same 5' end corresponding to the 5' splice site of intron 2, and that these 5' termini remain intact even when they are not blocked by the branched structure.

To map precisely the 3' end of the same RNA species, we used first the 220 residue DdeI-BalI fragment (positions 2806-3017, extended with 8 bases of pHp 34 linker at the BalI site) 3' end-labeled at the DdeI site (Fig. 5B). The 626 nt lariat, opened lariat and linear species (Fig. 5, lanes h, i, j) protected this probe as a similar set of bands with a 198 residue sequence as the major band. The comparison with the G+A ladder of the complementary probe positioned their 3' end exactly at the terminal G of the 3' splice site of the 626 nt intron 2, at position 3004. The obtention of shorter (197-194 base) fragments of decreasing intensity suggests that some of the 626 nt intron species have lost between 1 and 4 nucleotides at their 3' end (see later).

This probe was also used to determine the 3' end of the minor 520 nt species. As shown in Fig. 5, lane k, it was protected as a 91-94 residue doublet, which localized the 3' end of this minor species at 2 Gs designated by asterisks at positions 2896 and 2899 in the following sequence : 5' (C/T)₅ GAC(G)₄ AAG^{*}/GTGGT^{*} 3'. Therefore, the 520 nt species terminates in the region of a weak consensus-like splice sequence having only 6 pyrimidines upstream from the AG terminal residues, which would account for the infrequent use of this splicing site. As determined from the 5' and 3' termini of this excised intron, this minor species should have a length of 521-518 nt, which fits well with the size of 520 nt as determined after debranching (Fig. 3). However, the apparent cleavage 3 nucleotides downstream from the

terminal AG of the 3' splicing site, as deduced from the 94 residue sequence, remained unclear. Besides the 91-94 base fragments, we also observed the protection of a set of 194-198 residue bands which was expected since the 520 nt lariat species was contaminated by the 626 nt opened lariat species (Fig. 3).

Finally, when the 427 residue FokI-BalI fragment, 3' end-labeled at FokI site, was used to map the 3' terminal sequences of the 337 nt lariat and linear species (Fig. 5C, lanes n, o) a set of bands of decreasing intensity from 118 to 114 residues was observed with both of them. The comparison with the G+A ladder (lane G/A) of the complementary probe resulted in the mapping of the major S1 cleavage at the alternative 3' splicing site at position 2715. Moreover, 117-114 residue bands of decreasing intensity were readily detected, similar to the case of the 626 nt intron species (Fig. 5B, lanes h, i, j). We verified that they were not generated by any initial heterogeneity of the 3' end-labeled probes, resulting from an incomplete repair of the termini of the DNA segments. On the other hand, it was clear that the 3' terminal sequences for the 626 nt and 337 nt intron 2 (-CTATAG/ and ATCTAG/) and the common 5' terminal sequence (/GTATGA-), when hybridized with their complementary probes, should present similar stabilities and therefore similar S1 digestion pattern. Consequently, since only the 3' termini of these RNA species, and not their 5' termini, display some heterogeneity, our finding is consistent with the idea that all these molecules were subjected to a moderate 3' to 5' exonuclease action in the nuclei, leading to the loss of 1 to 4 nucleotides.

Subcellular distribution of free introns.

It was interesting to know whether intact intron forms are predominantly transported towards cytoplasm to be degraded. Hypotonic medium was used to isolate nuclear and cytoplasmic fractions (17) and poly(A)⁻ nuclear and cytoplasmic RNA were examined by Northern analysis as in fig. 2A. When compared to total cellular RNA, we found that the cytoplasmic fraction contained only 10% and 15-20% of the easily analysable 626 nt lariat and linear introns, respectively (data unshown). Since a fraction of these species can leak out of the nuclei upon lysis of the cells, our results are not in favour of an appreciable accumulation of intact introns in cytoplasm, as also shown previously for immunoglobulin intron (15).

DISCUSSION

Structural analysis of the different excised species of intron 2

A detailed analysis of all of the intron 2-specific RNA has provided

clear-cut data concerning the 3 excised intron 2 species of 626, 520 and 337 nt long which are generated by the alternative use of three 3' splicing sites and a common 5' splicing site.

The intron 2 species are mainly present as the lariat form which represents at least 65-75% of all forms. The branch points of the major 626 and 337 nt lariats are localized 26 and 25 residues upstream from the terminal AG, respectively, and such distances are in the range of those determined for authentic branch points for a variety of introns in higher eukaryotes, either in vivo (13, 29) or in vitro (6, 7, 31). We found that the branching sequences of both major intron 2 species are identical or very close to the consensus sequence (13). Therefore, our data, and those of others (13,19) indicate that the first steps of cleavage at the 5' end of introns and branching occur in vivo as described in in vitro systems (6, 7, 31). The results of Ruskin et al (32) on the activation of cryptic points suggest that a mechanism measuring the distance from the 3' splicing site is primarily implicated for branching. However, it is clear that the nucleotides surrounding the branched residue are also of importance. That was demonstrated by Rautmann and Breathnach (33) who analysed the in vivo splicing efficiency of a premRNA containing an intron with or without a consensus branching sequence.

The second major form that we detect is the linear segment of the introns. Such a form is systematically detected in the in vivo studies of higher or lower eukaryotic systems (13, 29, 34), indicating that it represents an ubiquitous product of the in vivo intron excision process. However, the linear intron species is not detected (5, 6, 35, 36) or only as a minor product (27) in in vitro splicing systems, suggesting that the branch point could be protected in vitro (27).

The case for the opened lariat form is less clear cut. We determine that this form is 5-12 times less abundant than the corresponding linear form for the predominant 626 nt intron 2, and it is practically undetectable for the 337 nt intron (quantitation from Fig. 2 and other Northern experiments). Moreover, when RNA was directly isolated from infected cells, the 626 nt opened lariat became a weaker component (no more than 2% of the total 626 nt intron 2, data not shown). Such a low level might be due to an aspecific degradation of the lariat during the preparation steps of the poly(A)-RNA. Therefore, our data suggest that the opened lariat form of the intron is not a genuine product of the in vivo splicing process. In contrast to our results, Zeitlin and Efstratiadis (13) found similar amounts of linear and

opened lariat forms for intron 2 of β globin premRNA from fetal liver. However, it is known that liver is rich in nucleases (37) and a very rapid RNase action during the first steps of RNA extraction cannot be excluded (13).

Concerning the terminal sequences of the introns, we show that the 5' termini of the linear intron forms are as well conserved as those blocked in the branching structure. Since exo or endonucleolytic activities are able to degrade uncapped premRNA sequences in nuclei (38), our results suggest that 5' termini of the linear introns are not over-exposed in the nuclei. In contrast, some 3' termini of the different forms of the intron 2 species are submitted to a moderate nibbling of between 1 and 4 residues. A nuclease (most likely a 3' to 5' exonuclease) which degrades in vitro the 3' tail of a purified lariat RNA has been partially purified from nuclear extracts (27) and it might explain the nibbling that we detect. However, we are unable to detect any lariat form whose 3' linear tail has been further degraded, unlike the case in yeast (34) or in in vitro splicing systems (21). This implies that, in higher eukaryotes, most of the 3' linear tail of the lariat is strongly protected in vivo against any further exonucleolytic activity. Such a specific protection and the protection of all the remaining lariat sequences, are presumably due to the fact that the lariat remains under an RNP form, associated in situ to the splicing complex through its life. Intron lariats have been shown to be also associated to the large splicing complexes in in vitro splicing systems (35, 36).

The intron degradation pathways

Our study gives some indications concerning the fate of the excised introns in in vivo systems. First, the predominant accumulation of the lariat forms strongly indicates that intron lariats remain protected during an appreciable period, whose extent cannot be precised since only steady state RNA is analysed. Our results are consistent with the fact that the intron lariats are primarily debranched after this first protection period. Whether this debranching reaction occurs when lariats are still integrated in the splicing complex or only after their release is unknown. Anyway, linear introns are also protected since they accumulate at an appreciable level. Considering the presence of a 3' to 5' exonuclease activity in the nucleus, one of the major degradation pathways of the linear intron could be a progressive nibbling of the intron sequences. However, we and others (13) do not detect shortened linear intronic segments, after Northern or S1 mapping analysis, thus excluding a dramatic digestion of the termini of the linear intron. Since transport of introns towards cytoplasm does not seem predomi-

nant, we speculate that the linear intron undergoes a stochastic endonucleolytic attack in nuclei. A single cleavage of the linear 626 nt intron could generate a heterogeneous population of smaller sequences, which should be indiscernable from the background of the sequences of the E2A unit, detected in Fig. 2. Obviously a direct analysis on the intron degradation products should be carried out in order to confirm this proposition and to determine whether it represents a general pathway of intron degradation.

Elimination of intron 2 by a one cycle excision

As shown previously, the analysis of *in vivo* splicing intermediates, has revealed the existence of cleavage sites inside intron sequences, especially by using the S1 nuclease protection method and it has been proposed that introns may be excised in a multistep process (9, 10, 11, 39, 40). However, results supporting the elimination of introns in one cycle of excision has been reported recently both in *in vivo* (12, 13, 29) and in *in vitro* systems (5, 6, 7).

The elimination of intron 2 of the E2A premRNA in 2 cycles of excision was theoretically possible since a sequence 5'GUAGUC3' contiguous to the major alternative 3' splicing site at position 2715 represented a good candidate for a cryptic 5' splicing site (12, and Fig. 1). Although missing a G residue at the fifth position when compared to the consensus 5' splicing site (41), this sequence should pair more strongly with the 5' end of the U1 SnRNA than the 5' splicing site (5'GUACUG3') used for the synthesis of the 9S E1A mRNA of adenovirus. However, we are unable to detect the 289 nt lariat species and our data indicate that intron 2 is systematically removed by a one cycle process which is likely to represent the general rule for the production of correctly spliced mRNA molecules (12, 13). Our results suggest also that the presence of 2 potential 5' and 3' splicing sites in contiguous positions inside an intronic sequence does not favour significantly the elimination of the intron in 2 cycles, even if one of them is used frequently. One intrinsic property of the splicing machinery *in vivo* could be that the occurrence of an excision-ligation reaction involves the subsequent protection of the surrounding messenger sequences, resulting in the exclusion of new undesirable excision by the use of cryptic or weaker 5' and 3' splicing sites. There may be other explanations of our observations (42, 43).

Alternative excision of intron 2

We show here that the major 2.0 and 2.3 kb mRNA of the E2A unit, whose poly(A)⁺ splicing intermediates were characterized previously (12), were actually generated by the alternative excision of the 626 and 337 nt intron 2

segments, respectively. In addition, we detected another scarcer alternative excision which occurs by the use of a 3' splicing site somewhat divergent from the consensus sequence. The excision of the 520 nt intron 2 should generate a third mRNA of 2.1 kb which was not previously detected in the poly(A)⁺ nuclear RNA (12). That was not surprising since the 2.1 kb RNA should represent no more than 1-2% of the major 2.0 kb mRNA. The detection of the 520 nt lariat intron 2 clearly shows that very unfrequent splicing events can occur in vivo, the reasons for this being still unclear. We previously proposed that alternative splicing of intron 2 may be favoured by the presence of a large inverted repeated sequence of 106 nt (44), which overlaps the major terminal 3' splicing site, and may mask it in some premRNA molecules (12). The consequence could be an induction of alternative 3' splicing sites. Actually, Solnick has recently demonstrated that RNA secondary structure can induce alternative splicing in vivo and in vitro (45).

ACKNOWLEDGEMENTS

We are grateful to Mr G. Hildwein for expert technical assistance. We thank Mr J.L. Weickert and Mrs M. Gilbert for growing HeLa cells, Mr A. Staub for the synthesis of oligodeoxynucleotide probe and B. Boulay, C. Kutschis and A. Marrel for preparing the manuscript. We also thank Drs G. Richards and R. Breathnach for critically reading the manuscript. This work was partially supported by grants from the CNRS, the INSERM and the Fondation pour la Recherche Médicale.

REFERENCES

1. Kofe, R. and Weissman, S.M. (1982) *Nucl. Acids Res.* **10**, 5429-5445.
2. Padgett, R.A., Hardy, S.F. and Sharp, P.A. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5230-5234.
3. Hernández, N. and Keller, W. (1983) *Cell* **35**, 89-99.
4. Krainer, A.R., Maniatis, T., Ruskin, B. and Green, M.R. (1984) *Cell* **36**, 993-1005.
5. Grabowski, P.J., Padgett, R.A. and Sharp, P.A. (1984) *Cell* **37**, 415-427.
6. Ruskin, B., Krainer, A.R., Maniatis, T. and Green, M.R. (1984) *Cell* **38**, 317-331.
7. Konarska, M.M., Grabowski, P.J., Padgett, R.A. and Sharp, P.A. (1985) *Nature* **313**, 552-557.
8. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* **50**, 349-383.
9. Grosveld, G.C., Koster, A. and Flavell, R.A. (1981) *Cell* **23**, 573-584.
10. Donaldson, D.S., McNab, A.R., Rovera, G. and Curtis, P.J. (1982) *J. Biol. Chem.* **257**, 8655-8660.
11. Mariman, E.C.M., van Beek-Reinders, R.J. and van Venrooij, W.J. (1983) *J. Mol. Biol.* **163**, 239-256.
12. Gattoni, R., Keohavong, P. and Stévenin, J. (1986) *J. Mol. Biol.* **187**, 379-397.
13. Zeitlin, S. and Efstratiadis, A. (1984) *Cell* **39**, 589-602.

14. Sittler, A., Gallinaro, H. and Jacob, M. (1986) *Nucl. Acids Res.* 14, 1187-1207.
15. Coleclough, C. and Wood, D. (1984) *Mol. Cell. Biol.* 4, 2017-2022.
16. Rosenthal, R. and Raskas, H.J. (1985) *Mol. Cell. Biol.* 5, 1084-1092.
17. Stévenin, J., Gattoni, R., Keohavong, P. and Jacob, M. (1982) *J. Mol. Biol.* 155, 185-205.
18. Cross, F.R. and Darnell, J.E. (1983) *J. of Virol.* 45, 683-692.
19. Maxam, A. and Gilbert, W. (1980) *Meth. Enzymol.* 65, 499-560.
20. Thomas, P. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5201-5205.
21. Ruskin, B. and Green, M.R. (1985) *Science* 229, 135-140.
22. Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) *Nucl. Acids Res.* 11, 1475-1489.
23. Rigby, P., Dieckman, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
24. Goodman, H. (1980) *Meth. Enzymol.* 65, 63-64.
25. Weaver, R.F. and Weissmann, C. (1979) *Nucl. Acids Res.* 7, 1175-1193.
26. Casey, J. and Davidson, N. (1977) *Nucl. Acids Res.* 4, 1539-1552.
27. Krämer, A. and Keller, W. (1985) *EMBO J.* 4, 3571-3581.
28. Keith, G., Pixa, G., Fix, C. and Dirheimer, G. (1983) *Biochimie* 65, 661-672.
29. Sittler, A., Gallinaro, H., Kister, L. and Jacob, M. Personal communication.
30. Contreras, R., Gheysen, D., Knowland, J., van de Voorde, A. and Fiers, W. (1982) *Nature* 300, 500-505.
31. Reed, R. and Maniatis, T. (1985) *Cell* 41, 95-105.
32. Ruskin, B., Greene, J.M. and Green, M.R. (1985) *Cell* 41, 833-844.
33. Rautmann, G. and Breathnach, R. (1985) *Nature* 315, 430-432.
34. Domdey, H., Apostol, B., Lin, R.J., Newman, A., Brody, E. and Abelson, J. (1984) *Cell* 39, 611-621.
35. Grabowski, P.J., Seiler, S.R. and Sharp, P.A. (1985) *Cell* 42, 345-353.
36. Frendewey, D. and Keller, W. (1985) *Cell* 42, 355-367.
37. Bartholeyns, J. and Baudhuin, P. (1977) *Biochem. J.* 164, 675-683.
38. Green, M.R., Maniatis, T. and Melton, D.A. (1983) *Cell* 32, 681-694.
39. Berget, S.M. and Sharp, P.A. (1979) *J. Mol. Biol.* 129, 547-565.
40. Avvedimento, V.E., Vogeli, G., Yamada, Y., Maizel, J.V., Pastan, I. and de Crombrughe, B. (1980) *Cell* 21, 689-696.
41. Mount, S.M. (1982) *Nucl. Acids Res.* 10, 459-472.
42. Kühne, T., Wieringa, B., Reiser, J. and Weissmann, C. (1983) *EMBO J.* 2, 727-733.
43. Noble, J.C.S., Prives, C. and Manley, J.L. (1986) *Nucl. Acids Res.* 14, 1219-1235.
44. Munroe, S.H. (1983) *Nucl. Acids Res.* 11, 8891-8900.
45. Solnick, D. (1985) *Cell* 43, 667-676.