The quantitation and distribution of splicing intermediates in HeLa cells and adenovirus RNAs

J.David Reilly, John C.Wallace¹ and Mary Edmonds

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA

Received June 1, 1987; Revised and Accepted August 4, 1987

ABSTRACT

The steady state level of splicing intermediates in HeLa cells and in adenovirus RNA made late in the infectious cycle has been measured by a branch point analysis. About one in ten poly A(+) nuclear RNAs contained a branch point, but only 1/3 as many adenovirus RNAs were branched. Fewer branches were found in the poly A(-) RNAs of the nucleus and of late adenovirus transcripts suggesting that excised lariat introns do not accumulate *in vivo*. Branched RNAs were found in the poly A(+) RNAs from a nuclear matrix fraction, but several experiments failed to show an enrichment in these splicing intermediates in this matrix fraction. Branches were found in all size classes of poly A(+) nuclear RNA and were not exclusively associated with either the 3' or 5' regions, but were randomly distributed within RNA molecules. These results as well as the base and sequence data on branch points (1,18) are consistent with the conclusion that branched poly A(+) RNAs are splicing intermediates.

INTRODUCTION

The branched RNA molecules first detected as ribonuclease resistant oligonucleotides in digests of polyadenylated nuclear RNA (1) are now known to be intermediates in RNA splicing (2,3). A branch is created during RNA splicing once the phosphodiester bond at the 5' splice site is cleaved, leaving a phosphate on the GMP at the 5' end of the intron that is then esterified to the 2' hydroxyl group of an adenosine residue about 20 to 40 nucleotides upstream from the 3' splice site (Figure 1). This 2'5' phosphodiester bond creates a branch that encloses most of the intron within a circle, leaving any remaining downstream sequences in a single stranded tail. The first of these circular structures, called lariats, to accumulate in cell extracts during splicing retains the downstream exon(s) and is called an intron-exon lariat. Subsequent exon splicing releases the branched intron as a lariat (Figure 1). It was the unusual electrophoretic mobility of these lariats that allowed them to be separated from other linear RNA molecules in splicing extracts. Similar lariat forms were soon detected in β -globin pre-mRNA of

Figure 1 - Pathway for pre-mRNA splicing (2,3).

rabbit liver (4) and in yeast that had been transfected with plasmids bearing genes for actin (5).

The base composition and sequence of the individual branch points analyzed thus far from lariats synthesized *in vitro* (6) are strikingly similar to those branch points recovered from digests of the total polyadenylated RNA of the cell nucleus (1). This indicates that the branched RNAs first detected in cells are or once were lariats, presumably of the intron-exon type since they retain a poly A tail. Isolation of all of the branched RNAs from the cell is essentially precluded by the enormous complexity of the pre-mRNA population. However, we have found that isolated branch points can be quantitated quite reproducibly in RNA from cells heavily labeled with ³²P-phosphate. Although we regularly detect one branch point in about 40,000 nucleotides of polyadenylated nuclear RNA, none are found in the cytoplasmic mRNA of exponentially dividing HeLa cells. It was this disparity that led us to predict that branched RNAs were intermediates from which branch points were removed during the processing of pre-mRNA to mature mRNA (1).

Using the branch point analysis technique we have extended these observations in several directions that provide a more quantitative view of the distribution of splicing intermediates in the cell nucleus. The branched RNA content of adenovirus specific RNAs has been compared with the mRNA precursor populations in non-infected cells. The number of splicing intermediates in RNAs isolated from subfractions of the nucleus such as the matrix has been measured, as has the number of branch points in different RNA size classes and at different sites within RNA molecules.

EXPERIMENTAL PROCEDURES

Cell Growth and Labeling

HeLa Cells: Approximately 8×10^7 cells in late logarithmic growth were labeled with 800 $\mu\text{Ci}/\text{ml}$ of ^{32}P -phosphate at a concentration of 2×10^6 cells/ml for 4 hrs at 37°C . Cells were harvested, washed, and fractionated into cytoplasm and nuclei as described (1).

Adenovirus Infected Cells: About 8×10^7 cells at a concentration of 1.6×10^6 cells/ml were infected with adenovirus type 5 at 100 plaque forming units per cell. One hour later the cells were diluted 4-fold with Eagle's minimal essential medium containing 2% calf serum. Fourteen hours later 50 mCi ^{32}P -phosphate was added and labeling was continued for 4 hours. Infected cells were harvested, washed and fractionated as described below.

Cell Fractionation:

Isolation of Nuclei: Nuclei were prepared from uninfected and adenovirus infected cells by the method of Penman except that the DNase step was omitted (7).

Isolation of the Nuclear Matrix: Matrix fractions were prepared from uninfected cells by a modification of the method of Ciejak et al. (8) except that all procedures were done at 0°C and nuclei were not sedimented through dense sucrose prior to matrix isolation. Glycerol was omitted from buffer A (10 mM Tris-HCl, pH 7.5; 10 mM NaCl; 1.5 mM MgCl_2). Non-detergent washed nuclei from ^{32}P -labeled cells were washed in 4 volumes of buffer A containing 0.5% Triton X-100 followed by 4 volumes of buffer A lacking Triton X-100. The pellet recovered from each wash by centrifugation at $450 \times g$ for 5 minutes was then resuspended in a volume of buffer A equal to twice the volume of the pellet before adding an equal volume of DNAase I (1 mg/ml of DPFF grade from Cooper Biochemicals, further purified from contaminating RNase by the method of Maxwell, et al. (9)). Incubation continued at 0°C for 20 minutes with occasional stirring. The suspension was diluted with 3 volumes of buffer A and the mixture was centrifuged at $4000 \times g$ for 10 minutes. The pellet was resuspended in 2 volumes of buffer A. An equal volume of 1M NaCl in buffer A was then added dropwise with gentle mixing. The suspension was then kept at 0°C for 15 minutes. The pellet recovered after centrifugation at $4000 \times g$ for 10 minutes was designated the nuclear matrix fraction. The supernatant fractions from the DNase I digestion and the high salt extraction were pooled and designated as the nuclear supernatant.

RNA Isolation and Fractionation

Total RNA: RNA was extracted from each fraction by the hot phenol/SDS method (10).

Poly A(+) RNA: Poly A(+) was isolated by three cycles of binding of RNA to oligo dT cellulose as described (11), except for a 2 min heating at 60° before each binding to minimize contamination with non-polyadenylated RNA. The first unbound RNA fraction is designated as the poly A(-) RNA.

Adenovirus RNA: Adenovirus poly A(+) RNA was separated from host poly A(+) nuclear RNA by hybridization to nitrocellulose filters (type BA85, 45µ, 11 mm diameter, Schleicher and Schuell). About 10 µg of alkali denatured adenovirus DNA type 5 was bound to each filter as described (12). Prior to hybridization three DNA filters in a 5 ml glass vial were incubated with shaking for 3 hrs at 50°C in 150 µl containing NaCl (600 mM), Tris-HCl pH 7.4 (10 mM), EDTA (2 mM), 0.1% SDS, 50% deionized formamide and tRNA at 30 µg/ml. After removal of hybridization buffer the same solution containing the ³²P-labeled poly A(+) RNA was added in 1/3 aliquots to 3 of these vials so that RNA from 1.4×10^7 infected cells was exposed to 90 µg of adenovirus DNA. After 72 hrs of gentle shaking at 50°C, filters were extensively washed at 65°C with wash buffer (hybridization buffer lacking formamide and tRNA). Filters were rinsed once with wash buffer that lacked SDS. Bound RNA was released with 150 µl H₂O at 100°C for 10 min. Adenovirus RNA bound to filters accounted for 35-40% of the nuclear poly A(+). Less than 1% of the nuclear RNA of uninfected cells bound to the same adenovirus DNA filters.

Size Fractionation of Poly A(+) RNA: Nuclear polyadenylated RNA (about 7 µg plus 40 µg yeast tRNA carrier) was dissolved in 60 µl ETS (0.01M EDTA, 0.01M Tris-HCl pH 7.4, 0.1% SDS), 140 µl DMSO was added before the solution was heated at 63°C for 3 min. 200 µl of NETS (ETS plus 0.1M NaCl) was then added, and the solution layered on top of a 12 ml 15-30% sucrose gradient in NETS and centrifuged for 15 hrs at 23,500 rpm in a Beckman SW-40 rotor. 0.44 ml fractions were collected, and 3 µl aliquots were counted for radioactivity. Appropriate fractions were pooled and RNA was precipitated with ethanol. When needed, yeast tRNA was added to a concentration of 10 µg/ml before ethanol addition.

Alkali Nicked Poly A(+) RNA: 5-6 µg of RNA containing 40 µg of purified tRNA carrier was added to 0.5 ml of 0.2M NaOH at 0°C for 2-3 min. The solution was neutralized by the addition of 0.5 ml of 0.2M HCl, 0.1M Tris-HCl pH 7.4, 0.2% SDS, 0.02M EDTA. Oligo (dT)-cellulose slurry was added and the poly (A) containing and lacking molecules were separated as described above.

Isolation and Purification of Branch Points

RNAse Digestions/Poly A(+) RNA: RNA (15 μ g RNA that included 10 μ g of purified *E. coli* tRNA as carrier) in 200 μ l of a solution containing 5 mM Tris-HCl, pH 7.4, 5 mM Na EDTA pH 7.4, 0.3 μ g of RNase A (Sigma Type XII-A), 50 units of RNase T1 (Calbiochem) and 20 units of RNase T2 (Sigma) was incubated for at least 4 hrs at 37°C before the pH was reduced to 4.5 by the addition of sodium acetate to a final concentration of 20 mM. An additional 20 units of RNase T2 was added and incubation was continued for at least 4 hrs. Prolonging the incubation to 18 hrs did not affect branch recovery. This digest is referred to in the text as the RNase T2 digest.

Poly A(-) Nuclear RNA: About 100 μ g of the nuclear RNA not bound to oligo dT cellulose was digested as described above for the poly A(+) RNA, except that after lowering the pH to 4.5, 100 units of RNase T2 were added. 18 hrs later another 100 units were added and incubation was continued for 4 hrs at 37°C.

Removal of 5' Caps From RNase Digests: Acetylated dihydroxyboronyl cellulose, DBAE, (Collaborative Research; no longer commercially available) was prepared for use as follows. The powder was suspended in buffer A (1M NaCl, 0.05M NaAcetate pH 5.0) and allowed to swell at 4°C for at least 8 hrs. It was then pelleted at low centrifugal force and resuspended in sorbitol buffer (0.5M morpholine-HCl pH 8.7, 0.01M MgCl₂, 0.2M NaCl, 1.0M sorbitol) at 4°C. The sorbitol buffer was removed by centrifugation, the cellulose was washed four times with buffer B (1.0M NaCl, 0.05M sodium (2-N-morpholine)ethanesulfonate pH 5.5)), before resuspending in buffer A plus 1 mM NaN₃ for storage at 4°C. A suitable replacement for DBAE cellulose can be prepared by linking m-aminophenylboronic acid to CM-sepharose by the method of Pace and Pace (13). Digests of poly A(+) RNA can be processed and applied directly to DBAE columns as described (1). However, a preliminary purification that reduces the vast excess of ³²P labeled mononucleotide can simplify subsequent DBAE chromatography. This extra step was always used for the poly A(-) RNA digests which were diluted 20-fold with 50 mM morpholine pH 8.0 before passing over a 2.5 x 1 cm diameter DEAE column equilibrated with 50 mM morpholine-HCl pH 8.0. The bulk of the ³²P labeled mononucleotides (> 90%) were recovered in a 20 ml wash of the column. Bound material was eluted with 300 mM NaCl in 50 mM morpholine pH 8.0. The eluate was adjusted to the composition of buffer C (500 mM NaCl, 100 mM MgCl₂, 50 mM morpholine-HCl pH 7.0, 20% ethanol) before it was applied to a 7 x 0.7 cm column of DBAE cellulose equilibrated with buffer C at 4°C. After 10-15 minutes to allow binding, buffer C was

passed over the column at 4°C until 50 ml fractions were collected. Bound material that would include cap 1 was eluted from the column with 50 mM Na-acetate pH 4.5 at 23°C. Quantitation of caps requires additional purification on DEAE cellulose (see below).

DEAE Cellulose Chromatography: The ^{32}P -labeled fractions not bound to DBAE cellulose were pooled and diluted with 11 volumes of 7 M urea in 25 mM Tris-HCl pH 7.4. Afterwards, ^3H -labeled oligonucleotides with net charges ranging from -2 to -7 were added. The digest was then pumped over a 25 x 1 cm DEAE Sephacel column that had been equilibrated with 50 mM NaCl in 7 M urea, 25 mM Tris pH 7.4. The column was washed with 40 ml of this buffer before it was developed with a 250 ml linear gradient of 50 mM to 225 mM NaCl in this same urea-Tris buffer. In contrast to RNase digests of poly A(+) RNAs (1) and Figure 3, the ^{32}P -elution profile in the region of the -6 charge marker for the poly A(-) RNA was heterodisperse. The fractions pooled from this region were, therefore, diluted to 50 mM NaCl with the 7 M urea-Tris pH 7.4 buffer containing fresh ^3H -oligonucleotide markers before re-application to another identical DEAE column that was developed as before. Although more homogeneous, the ^{32}P eluting with the -6 charge marker did not elute as the symmetrical peak observed for the branch points from a poly A(+) RNA digest. To quantitate branch points with the accuracy required for these studies, the ^{32}P -labeled material was de-salted (14), dried at low temperature in a Bio Dryer (Savant).

Characterization of Branch Points

Polyacrylamide Gel Electrophoresis: Branch points were separated from RNase digestion products of similar charge by electrophoresis in denaturing polyacrylamide gels that separate on the basis of size. Samples dissolved in 90% formamide (v/v) containing bromphenol blue and xylene cyanol tracking dyes were heated at 65°C for 1 min before loading on a 20% polyacrylamide: 7M urea vertical slab gel (15). Electrophoresis was done at 18 watts. All of the ^{32}P in the gel from the branch points purified from the poly A(+) RNA digest migrated as a single band, while about 33% of the ^{32}P with a -6 charge purified from the poly A(-) RNA migrated in the same region. The ^{32}P eluted from each gel band was desalted on a small DEAE column before drying in vacuo. Nuclease P1 digestion released approximately 50% of the ^{32}P as inorganic phosphate from each sample.

High Voltage Paper Electrophoresis: Branch points and products released by nuclease P1 digestion were analyzed by high voltage paper electrophoresis (16) as reported previously (1).

Nuclease P1 Digestion: Purified branch points taken up in 50 μ l of 10 mM Na Acetate pH 5.0 containing 25 μ g of purified tRNA carrier and 20 μ g of nuclease P1 (Sigma Biochemicals) were incubated for 2 hrs at 37°C.

RESULTS

Isolation and Quantitation of Branch Points in RNA Populations

We have developed a method for quantitating the number of branches in an RNA population that depends on the release by ribonuclease of a branched oligonucleotide from 32 P-labeled RNA (Figure 2). The method is based on the following experimental considerations: (1) RNA labeled with 32 P to a high specific activity is completely digested to 3' nucleoside monophosphates with RNase T2, a ribonuclease lacking base specificity; (2) Any 5' CAPS released along with the branch points are removed from the digests by complexing the cis-diol groups of the m^7 GMP moiety characteristic of CAPS to borate groups attached to cellulose. This eliminates any overlap during subsequent purification of branches with nuclear type 1 CAPS (m^7 GpppN^mpNp) with a charge of -5 as well as any contaminating type 2 CAPS found only in cytoplasm (m^7 GpppN^m-pN^mpNp) that have the same net charge as branch points (-6); (3) 32 P-labeled material in the digest that lacks cis-diols is then fractionated by high resolution anion exchange chromatography. The 32 P elution profile from a polyadenylated nuclear RNA ribonuclease digest is shown in Figure 3A. Material recovered from HeLa RNA in the peak moving slightly behind an oligo-

Figure 2 - The branched structure isolated from a RNase T2 digestion of polyadenylated nuclear RNA of HeLa Cells (1).

Figure 3 - DEAE cellulose chromatography of ribonuclease digests of poly A(+) RNA. A. HeLa cell nuclei. B. Adenovirus specific RNA synthesized in late infection.

nucleotide marker with a -6 charge has been shown by a detailed secondary analysis to have the structure of Figure 2 (1).

Since the amount of ^{32}P -labeled material in branch points available in these experiments was seldom sufficient for full-scale secondary analysis, we have adopted simplified criteria for branch identification. Adherence to these criteria is necessary for the quantitation of branch points in RNA samples that have not been subjected to secondary analysis. The crucial feature of the analysis involves the endonuclease from penicillium, nuclease P1, that not only hydrolyzes RNA to 5' nucleotides, but also removes 3', but not 2' or 5' monoesterified phosphate from nucleotides and oligonucleotides (17). Thus it converts 50% of the phosphate of branch points of the type shown in Figure 2 to inorganic phosphate. It was the unexpected magnitude of this conversion that first drew our attention to the novel features of branch points. Other structures with similar net negative charge that could comigrate with branch points during ion exchange chromatography at pH 7.5, such as an RNase resistant pentanucleotide or a nucleoside tetrphosphate released from an initiated, but uncapped 5' end of RNA would give substantially less inorganic phosphate on nuclease P1 treatment, i.e., 20 and 25 percent respectively.

Figure 4 - High voltage electrophoresis at pH 3.5 of RNase T2 resistant structures with net charge of -6. A. Poly A(+) nuclear RNA of HeLa cells. B. Nuclease P1 digest of A. C. Poly A(+) adenovirus specific RNA made late in infection of HeLa cells. D. Nuclease P1 digest of C.

Should materials lacking cis-diols and migrating with a charge of -6 fail to release the expected quantity of inorganic phosphate on nuclease P1 treatment then additional purification would be necessary. One such purification route applied to the poly A lacking nuclear RNA of HeLa cells is described in the Methods section.

The criteria outlined above have been met in the case of adenovirus specific RNAs obtained during late infection of HeLa cells. Figure 3 compares the ^{32}P profiles of cellular and adenovirus RNA digests on anion exchange columns. The ^{32}P migrating as a distinct peak slightly ahead of a -6 charge marker seen in both digests signals the presence of branch points. Digestion of the material with nuclease P1 released the expected quantity of ^{32}P from each RNA (Figure 4). The ^{32}P -profile of branch point migration seen after high voltage paper electrophoresis at pH 3.5 reveals another property of these branch points, i.e., their heterogeneity. HeLa cell nuclear RNA branch points display 2 or 3 major components (Figure 4) at pH 3.5 that remain as more slowly migrating branch core structures after removal of the 3' phosphates. These structures represent branch points of different base composition. Base and sequence analysis of HeLa branch points showed base heterogeneity although the composition was not random (1). AMP served as the

branching nucleotide most of the time (75%), CMP was used as well (17%) while GMP and UMP were apparently rarely used (1,18). GMP accounted for 50% of the bases in branching nucleotides N2 plus N3 (Figure 2), while UMP and CMP, but not AMP, accounted for the rest. These compositions have been found for branch points of specific introns formed in vitro (2,3,6). Thus far only adenosine has been reported as the branching nucleotide, although a CMP branch has been observed in the first intron of human growth hormone pre-mRNA (K. Hartmuth and A. Barta, personal communication). The 50% GMP content of N2 + N3 is of course the result of the splicing mechanism in which the GMP at the 5' end of all introns becomes esterified to the 2' hydroxyl of the branching nucleotide. This leaves either UMP or CMP linked to the branch via the usual 3'5' phosphodiester bond that may account for much of the heterogeneity of branches seen here. It is interesting to note in Figure 4 that adenovirus branch points differ from those of HeLa cells in their electrophoretic mobility at pH 3.5. A component that co-migrates with UMP after nuclease P1 digestion is not seen in the HeLa branch population. This would suggest that adenovirus RNA transcribed late in infection may have intron branches of unusual composition. A UMP branch point with a second UMP at the N3 position (Figure 2) might account for the rapidly migrating branch core seen in Figure 4.

Branches in HeLa Cell and Adenovirus RNA

Quantitative data on the number and location of branch points in RNA molecules in the nucleus and in subnuclear fractions will be useful not only for understanding pre-mRNA processing in vivo, but for devising methods for the localization of splicing events in the nucleus. Branch specific antibodies, for example, could be used to locate lariat intermediates in situ by immunofluorescence analysis. We have used the branch point analysis technique described above to establish the number of branches in the poly A(+) and poly A(-) RNAs in the nucleus of HeLa cells and in adenovirus transcripts recovered during late infection of these cells. To standardize the quantitation of branch points for all RNA populations, we have defined a measure of branch abundance which is the total number of RNA nucleotides that contain a single branch point. As reported previously (1) and as seen in the last column of Table 1, the total poly A(+) nuclear RNA of HeLa cells contains about one branch for every 30 - 40,000 nucleotides. Branches are only one third as abundant in adenovirus specific poly A(+) RNA made late in infection. In both populations, the poly A(+) RNA contains more total branch points than the poly A(-) RNA. If all or most branched RNA molecules are

TABLE 1. Branches in HeLa Cell and Adenovirus Nuclear RNA

RNA	Mononucleotides cpm x 10 ⁶	Branches cpm	Nucleotides per branch
<u>HeLa Nucleus</u>			
poly A(+)	24.4	2990	33,000
poly A(-)	407	1900	154,000*
<u>Late Adenovirus</u>			
poly A(+)	90	2970	121,000
poly A(-)	150	1385	434,800

*Based on nucleotides estimated for RNA Polymerase II transcripts in poly A(-) RNA (see text)

generated during splicing reactions as our branch point composition data suggests, we can assume that the poly A(+) branched RNAs are true splicing intermediates in which the lariat intron remains attached to the adjacent exon that in turn is attached to a processed 3' end. The poly A(-) branched RNAs should contain all of the excised lariat introns as well as any lariat intermediates that have yet to be polyadenylated. This fraction might also include a lariat intermediate that has been temporarily detached from the 3' end of the message. This might occur if two introns could be removed from the same RNA essentially simultaneously. In such a case an upstream intron-exon lariat intermediate could be temporarily detached from the 3' end of the RNA during the removal of a second downstream intron. The poly A(-) RNAs are mostly RNA Polymerase I and III transcripts. An estimate of the RNA Polymerase II transcripts can be made using values for the 5'm⁷GpppN cap content measured in nuclear RNAs larger than 750 nucleotides in Chinese hamster cells (19). In this case 75% of such caps were in poly A(-) and 25% in poly A(+) RNAs. When these values are applied to the data of Table 1, about 18% of the poly A(-) RNAs are Pol II transcripts which yields one branch in 154,000 nucleotides. The lower levels of branches in the poly A(-) relative to the poly A(+) RNAs shows that lariat introns do not accumulate *in vivo* either in HeLa cells or adenovirus splicing reactions. The fate of the intron sequences cannot, however, be deduced from these data since debranching appears to be the result of a highly specific enzymatic cleavage of the A₂,p₅,G bond at the branch point (20). The product of the debranching reaction in splicing extracts is a linearized intron of identical sequence. The lower abundance of splicing intermediates, as well as lariat

introns, in the late adenovirus RNA population may reflect a more rapid rate of splicing for this relatively restricted set of RNAs as compared with the overall splicing rate for the far more complex population of cellular RNAs.

Branches in the Nuclear Matrix

A large fraction of the hnRNA of the nucleus is associated with a subnuclear fraction that remains after detergent treated nuclei are digested with DNase and extracted with high salt (21,22). When such preparations are viewed by electron microscopy, a fibrillar network made up largely of proteins is seen to fill a space that retains the shape of the nucleus (23). This preparation is often referred to as the nuclear matrix. Precursor mRNAs as well as pre-ribosomal RNAs have been reported to be preferentially bound to this nuclear subfraction while mature RNA species are found in both matrix and washes recovered during matrix isolation (8). Small amounts of DNA also remain tightly bound to the matrix (22). Use of specific exon and intron gene probes indicated that mRNA precursors for ovalbumin and ovomucoid were preferentially associated with matrix preparations (8). It has also been reported that genes transcriptionally activated by specific hormones tend to be bound to the matrix while inactive genes are not (24). These and other observations have led to the view of the matrix fraction as a site for transcription and for RNA processing as well, although there has been little evidence for the latter. It has recently been found that RNA splicing is associated with a multicomponent particle called a spliceosome that assembles when pre-mRNAs are added to nuclear extracts (25,26,27). One of the proteins tightly associated with hnRNP particles (the C protein) has recently been found in such spliceosome preparations (28). The relationship of these particles to each other and to the nuclear matrix preparation is as yet unclear. A recent report on the splicing of β -globin pre-mRNAs in HeLa cells transfected with β -globin genes, links splicing to components of the nuclear matrix (29). In this case, β -globin mRNA splicing intermediates, i.e. lariat structures as well as β -globin pre-mRNA were found to be preferentially associated with the matrix fraction, while the salt washes recovered during matrix preparation were enriched in mature β -globin mRNA and excised linear β -globin introns. This matrix preparation by itself was unable to carry out splicing in vitro, but when added to an RNA depleted nuclear extract, β -globin splicing intermediates were detected very early in the reaction. To explain this abolition of the lag period, that characterizes splicing of exogenously added pre-mRNA to such nuclear extracts, it was suggested that a pre-assembled spliceosome was present in matrix preparations. Splicing

TABLE 2. Association of Branches with the Nuclear Matrix

Exp	Poly A(+) RNA	Nucleotides/Branch	
	% on Matrix	Matrix	Supernatant
1	80	35,000	42,000
2	51	27,000	42,000
3	36	32,600	41,900

did not occur in vitro however, until soluble factors removed from the matrix during its preparation were restored by the addition of a nuclear extract.

We have used a branch point analysis to seek evidence for an enrichment of splicing intermediates in the matrix fraction. A matrix fraction was prepared by a method similar to that used in the β -globin experiments cited above (29). The polyadenylated RNAs that were isolated from the matrix pellet and from the combined salt washes recovered during its preparation were analyzed for branch points. In these RNAs, branches should be present only as lariat-exon splicing intermediates, since excised lariat introns are transferred to the poly A(-) RNA fraction after oligo dT cellulose treatment.

Although most poly A(+) nuclear RNAs are not branched (only about one in ten), a preferential association of lariats with a matrix fraction should be detected as an increased branch abundance, i.e. a decrease in the number of nucleotides needed to contain one branch. A concomitant increase in this value should be seen in the supernatant poly A(+) RNA. The data of Table 2 show that this value did not differ significantly in these two RNA fractions in any of three separate experiments. All values fell within the expected range of branch point concentrations found in the total poly A(+) RNA of the unfractionated nucleus, i.e. one branch per 32,000 - 48,000 nucleotides (a range observed in 7 separate experiments). The slight enrichment in branches in the matrix fraction seen in each experiment thus cannot be considered as significant.

It should be noted that the fraction of the total poly A(+) RNA associated with the matrix differed in each of the three separate experiments. Although we expected that RNA degradation might be a variable difficult to control during the isolation of the matrix fraction, we were unable to detect any major differences in size distributions of the poly A(+) RNAs obtained from these three experiments. Nor were the RNA sizes noticeably different from

Figure 5 - Electrophoretic mobility of ^{32}P -labeled poly A(+) and poly A(-) RNAs from nuclear matrix and nuclear supernatant fractions of HeLa cells on 2.2M formaldehyde-1% Agarose gels (38).
1. Poly A(-) from cytoplasm. 2. Poly A(-) from nuclear supernatant fraction. 3. Poly A(-) nuclear RNA from the matrix fraction. 4. Poly A(+) nuclear RNA in the matrix fraction. 5. Poly A(+) nuclear RNA in the nuclear supernatant. 6. Poly A(+) RNA in the cytoplasm.

those of the total poly A(+) RNAs recovered from unfractionated HeLa nuclei (compare with sedimentation profile of Figure 6). Figure 5 shows one such experiment in which these RNAs were resolved by electrophoresis in a denaturing polyacrylamide gel. The poly A(+) RNAs from matrix and non-matrix fractions were much larger on average than the poly A(+) mRNA from the cytoplasm. The latter showed a typical size distribution concentrated around the 18S rRNA marker. The poly A lacking nuclear RNA of the matrix does show the expected enrichment in ribosomal precursor RNAs (32S) while the supernatant fraction did not.

In spite of the lack of a plausible explanation for the variability in quantities of poly A(+) RNA associated with the matrix in different experiments, it should be noted that enrichment in splicing intermediates was not found regardless of the fraction of the total poly A(+) RNA bound to the matrix. If splicing intermediates are more tightly bound to the matrix than are the unbranched RNAs which represent more than 90% of the poly A(+) RNAs of the nucleus, then experiment 3 should have shown this most clearly where

TABLE 3. Recovery of Branches from Nicked Poly A(+) Nuclear RNA

Experiment	Nucleotides ^a x 10 ⁶		Branches		CAPS	
	A(+)	A(-)	A(+)	A(-)	A(+)	pA(-)
No. 1						
cpm	11.0	13.5	920	815	2440	3650
number ^b	11.0	13.5	230	204	415	627
% ^c	45	55	53	47	40	60
No. 2						
cpm	3.46	7.10	405	540	137	2300
number	3.46	7.10	101	135	24	397
%	33	67	43	57	6	94

^aA(+) and A(-) indicate the poly A containing and lacking molecules respectively

^bCalculated as described in reference (1)

^cPercent distribution between A(+) and A(-) molecules

more of the total RNA and consequently more unbranched RNA was recovered in the supernatant fraction.

Thus our experiments failed to show an enrichment of branched splicing intermediates in a matrix fraction, although such intermediates were recovered from this fraction. We can only conclude either that there is no preferential binding of polyadenylated lariats to matrix components or none that is stable enough to survive the matrix isolation procedures used here.

Location of Branch Points Within RNA Molecules

Once it was understood that branches form within introns undergoing splicing, it was obvious that branch points would be uniquely located in individual pre-mRNAs, but should be randomly distributed over the total RNA population given the wide variations in the location of introns in different genes. This prediction was confirmed for the poly A(+) nuclear RNA population of HeLa cells by introducing random nicks in these RNAs by a controlled exposure to alkali. The 3' portion of the nicked molecules were then separated from non-polyadenylated fragments by oligo dT cellulose chromatography. More than 98% of the poly A sequences in these experiments were recovered in the bound fraction (data not shown). Table 3 summarizes the results of branch point analyses carried out on the bound and unbound fragments.

The first significant observation is seen in experiment 1 where slightly nicked RNAs (45% of the ³²P remained covalently associated with poly (A)) had lost a significant fraction of the total branches (47%). Had all

TABLE 4. Relative Distribution of Branches in Size Fractionated RNA

RNA Size Class	Average ^a Length	NMP Branch	Molecules ^b Branch
I	1468	16,400	11.2
II	3323	40,300	11.6
III	5073	72,900	13.7
IV	7869	93,400	11.2

^aWeight average length, in nucleotides

^bThe number of molecules in each size class is calculated as follows. The length, in nucleotides, of the molecules in each fraction of the sucrose gradient (Figure 6) is calculated according to the formula, adapted from reference (37).

$$\text{length} = 41.9 \times (\text{fraction number})^{1.7}$$

The percentage of RNA in each sucrose gradient fraction within a given size class is then calculated by dividing the cpm in that fraction by the sum of the cpm of the fractions in its size class. The cpm-equivalent number of molecules in each fraction can then be calculated as

$$\text{cpm-equivalent no. of molecules} = \frac{(\% \text{ RNA in a fraction}) \times (\text{total cpm in a size class})}{\text{length}}$$

The number in each fraction is then summed over each size class to determine the total cpm-equivalent number of molecules in each size class. For example, the average size in nucleotides of the molecules in fraction number 18 was

$$41.9 \times (18)^{1.7} = 5704$$

The aliquot taken for fraction 18 in Figure 6 contained 14,900 cpm, and the total cpm in the aliquots of size class III in Figure 6 (fractions 16-19) was 62,600. Thus fraction 18 represented $14,900/62,600 = 23.8\%$ of the radioactivity in size class III. Size class III contained a total of 8.31×10^6 cpm, so fraction 18 had $.238 \times 8.31 \times 10^6$ cpm in total. Dividing this figure by the length of the molecule, calculated above,

$$\frac{1.98 \times 10^6}{5704} = 347$$

indicates that there are 347 cpm-equivalent molecules in fraction 18.

Performing these calculations gives the number of cpm-equivalent molecules in each sucrose gradient fraction, which can then be summed to the number in each size class.

Due to the very poor resolution at the top of the sucrose gradient, fractions 1-4 were not used in these calculations.

Figure 6 - Sucrose density gradient sedimentation (preparative) of nuclear polyadenylated RNA. The RNA was divided into the four size classes shown, which were analyzed for components as shown in Table 4. Marker RNA was sedimented in a parallel gradient.

branches been very close to the 3' ends of RNA, such a large fraction should not have been detached by this treatment. It should be noted however, that many molecules escaped nicking in this experiment since a substantial portion of them retained their 5' capped ends. Thus it was still possible that branches were preferentially associated with 5' regions of RNA and most of them were in unnicked RNAs. When nicking was increased (experiment 2) so that only 33% of the nucleotides remained attached to poly A, a large fraction of the branches (43%), but very few caps were still associated with 3' polyadenylated fragments. From these results it is apparent that branches are not preferentially associated with 3' or 5' halves of pre-mRNAs although a slight preference for the 3' fragments was seen in both experiments. This could be due either to an actual tendency for more introns to be close to the 3' ends of pre-mRNAs, or it may merely reflect the tendency for branches to be present in smaller RNAs (noted in Table 4) which present a smaller size target for nicking and therefore are more likely to remain undegraded. In any case this variable location of branch points within the poly A(+) pre-mRNA population is consistent with the wide variation in arrangements of introns within gene sequences.

The Size of Branched RNA Molecules

Splicing intermediates can be isolated from nuclear RNA in a wide range of sizes as seen in Table 4. Four size classes of polyadenylated nuclear RNA were obtained by pooling fractions from a sucrose gradient in which the total poly A(+) nuclear RNA had been sedimented (Figure 6). The average length of RNA in each size class ranged from about 1500 nucleotides in the smallest (pool I) up to 8000 in the largest. Branch analysis of each pool showed branch points to be most abundant in the smallest size class since only 16,000 nucleotides were needed to contain a single branch, while more than 90,000 were needed for the largest RNAs. When total nucleotides are converted to the average number of RNA molecules in each size class as described in the legend to Figure 6 the branch abundance can be related to the fraction of the total molecules that contain a branch point. Table 4 shows that in all size classes about one in eleven molecules are branched. This means that almost one in ten poly A(+) nuclear RNAs is a splicing intermediate. That the molar concentration of intermediates is about the same in all size classes of nuclear RNA was somewhat surprising. A deficiency of branches might have been expected in smaller RNAs particularly those in pool I that overlap the average size of mature mRNAs. A significant pool of mature mRNAs in the nucleus would dilute out any branched intermediates in this case. Also smaller RNAs would be expected to contain fewer introns that need to be spliced. The actual preponderance of branches in smaller RNAs is apparently not the result of degradation since most of the RNAs in this preparation had capped 5' ends (data not shown). An excess of branch points in small polyadenylated RNAs could result from a splicing mechanism that removes introns sequentially, starting from the 3' end. Such a mechanism would produce lariats that contain only a single exon. The available data on the direction and temporal order of intron removal in vitro and in vivo does not support such a obligatory mechanism. Both in vitro (30) and in vivo evidence (4) show for example that the intron at the 5' end of the β -globin pre-mRNA, IVS-1, is usually removed before IVS-2.

DISCUSSION

According to the current model of pre-mRNA splicing, the branch points present in RNase digests of poly A(+) RNA of the cell nucleus were once in the lariat introns of splicing intermediates, while those in the poly A(-) nuclear RNA were derived from the lariat introns excised during splicing. Support for this conclusion is found in the close match of the base

sequences from branches in total cellular RNA with those of branch points from lariat introns made in vitro (6). The data reported here on branch distribution in different size RNAs and in different regions of RNA molecules, as well as our earlier observation that cytoplasmic RNA is devoid of branches (1) are consistent with the conclusion that most, if not all, branches in poly A(+) are generated within introns during splicing. This claim cannot be made with similar confidence for branches in the poly A(-) RNA since their composition and distribution have not been studied. The possibility that branches are present on 2'hydroxyl groups of RNA molecules other than mRNA precursors must be considered. However, excised lariat introns of β -globin pre-mRNA have been identified in rabbit liver (4) and in actin introns from yeast transfected with a yeast actin gene (5), making it likely that the branch points found in the poly A(-) nuclear RNA of HeLa cells are in fact derived from excised lariats.

The finding that nearly one in ten poly A(+) nuclear RNAs is branched provides the first quantitative data on the steady state level of a key splicing intermediate in exponentially growing cells. The fact that the number of branches in the poly A(-) nuclear RNA presumably in excised introns, was actually somewhat lower (30% in Table 1), shows that excised lariat introns do not accumulate to any extent in vivo. This was also the case for branches in these same two adenovirus RNA populations although branches were far less abundant than in HeLa cells. This near equivalence of branches in splicing intermediates and excised lariat introns, suggests that splicing and debranching, although undoubtedly separate activities, are coordinated, presumably to prevent the accumulation of nucleotides in branch points. We presume here that the debranching activity first described in nuclear extracts (20) and recently purified from cytoplasmic extracts of HeLa cells (31) that specifically cleaves the 2'5' phosphodiester bond of lariats is responsible for branch removal in vivo. Support for debranching as a first step in intron degradation is found in the transfection experiments cited above where small amounts of linear introns were detected along with the excised lariat introns (4,5). A different degradation sequence is suggested by the detection in splicing extracts of lariat introns with shortened 3' tails (31) and also by the presence of a 3' exonuclease activity in these extracts that may be specific for lariat structures (32).

Our failure to find a preferential association of branched intermediates with the nuclear matrix fraction was somewhat unexpected in view of the reports of others. We cannot now ascribe our results to any obvious RNA

degradation or to any preferential loss of splicing intermediates during matrix isolation. In fact branch abundance in both matrix and non-matrix poly A(+) RNAs was well within the rather limited range of values routinely obtained for unfractionated RNA. Until the properties of this subnuclear fraction are defined more precisely and until the effects of the variables involved in matrix preparation on such properties are understood, quantitative studies of the type we report here may be premature.

In general our experiments show the feasibility of quantitatively analyzing splicing intermediates in vivo using a branch point analysis. Feasibility will depend of course on the specific activity of the ^{32}P -labeled RNA that can be generated in vivo and on the levels of RNase digestion products with net negative charges similar to branch points. The low levels of such contaminants simplified branch point purification from the poly A(+) RNAs of HeLa cells and of adenovirus. Additional procedures were needed, however, to remove unidentified contaminants from the poly A(-) nuclear RNA digests of HeLa cells (see Methods). Branched RNAs have not been reported in other types of cultured cells as yet, nor have fluctuations in the levels of splicing intermediates been measured for cells in different growth states. An antibody highly specific for the branch point has been prepared that may provide an immunoassay that could simplify such analyses (Munns, T.W., Reilly, J.D., and Edmonds, M., unpublished experiments). The branch point analysis described here should be readily adapted to cells growing in suspension cultures. A comparison of the number of branch points in mouse myeloma cells with HeLa cells might be of interest in view of the much more rapid rates of in vitro splicing observed with nuclear extracts of these cells (MOPC 315) as compared with similar extracts from HeLa cells (33), although the differences seen in vitro may well have a trivial explanation.

Branch point analysis could provide crucial evidence for a trans splicing mechanism recently proposed to account for an identical 35 nucleotide sequence found at the 5' end of all the mRNAs of trypanosomes (34). This spliced leader sequence is part of a 137 nucleotide sequence transcribed from a small exon that is part of a 1.35 kb gene that is repeated 200 times in the trypanosome genome. The fact that transcription of this leader sequence precursor is unlinked from pre mRNA transcription has led to a model of transplicing in which a linear branched intron-exon, rather than a lariat intron-exon is an intermediate. According to the model, joining of the 35 nucleotide leader sequence to an mRNA exon is accompanied by excision of this

branched intron. Evidence for a branched intermediate in trans splicing has recently been reported (35,36).

ACKNOWLEDGMENT

We thank Drs. Jim Williams and Brian Karger of Carnegie-Mellon University for adenovirus type 5 and for their advice on methods for infecting and for harvesting the virus from HeLa Cells.

This research was supported by NIH Grant GM 32585.

¹Present address: Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907, USA

REFERENCES

1. Wallace, J.C. and Edmonds, M. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 950-954.
2. Padgett, R.A., Konarska, M.M., Grabowski, P.J., Hardy, S.F. and Sharp, P.A. (1984) *Science* **225**, 898-903.
3. Ruskin, B., Krainer, A.R., Maniatis, T. and Green, M.R. (1984) *Cell* **38**, 317-331.
4. Zeitlin, S. and Efstratiadis, A. (1984) *Cell* **39**, 589-602.
5. Domdey, H., Apostol, B., Lin, R.J., Newman, A., Brody, E. and Abelson, J. (1984) *Cell* **39**, 611-621.
6. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) *Ann. Rev. Biochem.* **55**, 1119-1150.
7. Penman, S. (1966) *J. Mol. Biol.* **17**, 117-130.
8. Ciejak, E.M., Nordstrom, J.L., Tsai, M.J. and O'Malley, B.W. (1982) *Biochemistry* **21**, 4945-4953.
9. Maxwell, I.H., Maxwell, F. and Hahn, W.E. (1977) *Nucl. Acids Res.* **4**, 241-243.
10. Girard, M. (1967) *Meth. in Enzymol.* **12**, 581-588.
11. Nakazato, H. and Edmonds, M. (1974) *Meth. in Enzymol.* **29**, 431-433.
12. Kafatos, F.C., Jones, C.W. and Efstratiadis, A. (1979) *Nucl. Acids Res.* **7**, 1541-1552.
13. Pace, B. and Pace, N.R. (1980) *Anal. Biochem.* **107**, 128-135.
14. Tener, G. (1968) *Meth. in Enzymol.* **12**, 398-404.
15. Sanger, F. and Coulson, A.R. (1978) *FEBS Lett.* **87**, 107-115.
16. Salzman, N.P. and Sebring, E.D. (1964) *Anal. Biochem.* **8**, 126-134.
17. Fujimoto, M., Kuninaka, A. and Yoshino, H. (1974) *Agric. Biol. Chem.* **38**, 785-790.
18. Wallace, J.C. (1983) Ph.D. Thesis, University of Pittsburgh.
19. Salditt-Georgiev, M. and Darnell, J.E. (1982) *Mol. Cell. Biol.* **2**, 701-707.
20. Ruskin, B. and Green, M.R. (1985) *Science* **229**, 135-140.
21. Herman, R., Weymouth, L. and Penman, S. (1978) *J. Cell Biol.* **78**, 663-674.
22. Miller, T.E., Huang, C.-Y. and Pogo, A.D. (1978) *J. Cell Biol.* **76**, 675-691.
23. Berezney, R. and Coffey, D.S. (1974) *Biochem. Biophys. Res. Commun.* **60**, 1410-1417.

24. Ciejak, E.M., Tsai, M.J. and O'Malley, B.W. (1983) *Nature* 306, 607-609.
25. Grabowski, P.J., Seiler, S.R. and Sharp, P.A. (1985) *Cell* 42, 345-353.
26. Brody, E. and Abelson, J. (1985) *Science* 228, 963-967.
27. Frendeway, D. and Keller, W. (1985) *Cell* 42, 355-367.
28. Chou, Y.D., Grabowski, P.J., Sharp, P.A. and Dreyfuss, G. (1986) *Science* 231, 1534-1538.
29. Zeitlin, S., Parent, A., Silverstein, S. and Efstratiadis, A. (1987) *Mol. Cell Biol.* 7, 111-120.
30. Padgett, R.A., Konarska, M.M., Aebi, M., Hornig, H., Weissmann, C. and Sharp, P.A. (1985) *Proc. Natl. Acad. Sci. U.S.A.* 82, 8349-8353.
31. Arenas, J. and Hurwitz, J. (1987) *J. Biol. Chem.* 262, 4274-4279.
32. Krämer, A. and Keller, W. (1985) *EMBO J.* 4, 3571-3581.
33. Goldenberg, C. (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81, 4707-4711.
34. Borst, P. (1986) *Ann. Rev. Biochem.* 55, 701-732.
35. Sutton, R.E. and Boothroyd, J.C. (1986) *Cell* 47, 527-535.
36. Murphy, W.J., Watkins, K.P. and Agabian, N. (1986) *Cell* 47, 517-525.
37. Derman, E., Goldberg, S. and Darnell, J.E. (1976) *Cell* 9, 465-472.
38. Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H. (1977) *Biochemistry* 16, 4743-4751.