Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors

Maria M. Konarska*, Paula J. Grabowski, Richard A. Padgett & Phillip A. Sharp

Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The branch site of lariat RNAs produced during the splicing of the first two late leader exons of adenovirus-2 has a structure of ... $A^{2'p5'G...}_{3'p5'U...}$. There is a distinct complementarity between sequences preceding the adenosine at the branch site and the 5′ terminus of the intervening sequence.

EUKARYOTIC genes frequently contain intervening sequences that are removed from precursor RNAs by splicing. Introns occur in many genes coding for proteins, transfer RNAs and ribosomal RNAs in nuclear, mitochondrial and chloroplast DNAs[1-3], yet no single general mechanism seems to be involved in this splicing. Detailed studies of *in vitro* splicing of precursor tRNAs revealed that two ligation pathways operate in plant and animal systems[4-7]. The plant pathway involves ligation of two tRNA exons with 5′-phosphate and 2′-3′-cyclic phosphate ends, leading to the formation of a $N^{2'}_{3'}$(p)5′N bond[4,5]. The cyclic phosphate is displaced to the 2′ position in the above structure. In precursor (pre)-tRNA splicing in HeLa cell extracts, however, two exons with a 5′ hydroxyl group and 2′,3′-cyclic phosphate ends are joined in a normal phosphodiester bond[6,7]. A completely different mechanism operates in the splicing of rRNA precursors in the protozoan *Tetrahymena*[8]. A series of transesterification reactions initiated by an attack of the 3′ hydroxyl group of a guanosine cofactor is catalysed by the RNA molecule itself[9,10]. The products consist of a normal phosphodiester bond linking the two exons and the excised intervening sequence that is subsequently circularized.

The recent developments of *in vitro* systems capable of splicing precursors of messenger RNA allows the mechanism of the reaction to be elucidated[11-15]. Using a pre-RNA containing sequences from the adenovirus-2 (Ad-2) major late transcription unit and a whole-cell extract of HeLa cells, we have shown previously that splicing requires Mg^{2+} and ATP (ref. 14) and is inhibited by antisera against small nuclear ribonucleoprotein particles containing U1 RNA[16]. In this system, a 5′-terminal cap structure on the precursor RNA is required for splicing[17]. The reaction results in the formation of the spliced exons and the intervening sequence excised as a lariat RNA[18,19]. Early in the reaction a probable intermediate, consisting of the lariat form of the intervening sequence joined to the 3′ exon, was detected[18,19]. We have shown previously[19] that both lariat RNAs contain an unusual nuclease-resistant structure, an oligonucleo-

* Permanent address: Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02-532 Warsaw, Poland.

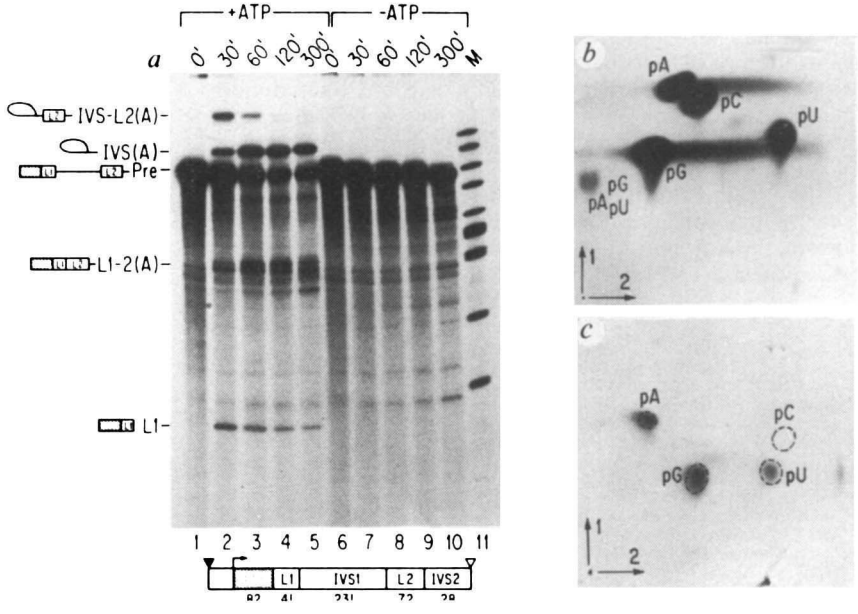


Fig. 1 *a*, Time course of the appearance in a HeLa nuclear extract of the products of splicing using the SP6 RNA polymerase-generated pre-RNA. Pre-RNA labelled with all four [α-^{32}P]NTPs was incubated in a HeLa cell extract with (lanes 1-5) or without (lanes 6-10) 1.5 mM ATP for the times indicated above the lanes. The RNA products were analysed on a 10% polyacrylamide 8 M urea gel; lane 11, ^{32}P-DNA markers from *Hin*cII digest of ΦX174 DNA. Left, structures of the various RNAs produced in the reaction; below, the structure of the DNA transcription template. L1 and L2 are the first and second leader exons of the Ad-2 major late transcription unit; IVS-1 is a shortened form of the first intervening sequence; IVS-2 a portion of the second intervening sequence; the hatched box represents sequences upstream of the *in vivo* RNA start site present in the RNA made with SP6 RNA polymerase. *b*, Analysis of the nuclease P_1-digestion products of purified IVS-A by two-dimensional TLC using solvent A in the first dimension and solvent B in the second dimension. *c*, The branched oligonucleotide was eluted from the plate in *b* and digested with snake venom phosphodiesterase. The products were analysed by two-dimensional TLC using solvent A in the first dimension and solvent C in the second dimension.

Methods. Pre-RNA (460 nucleotides) was transcribed from a *Bgl*I-cleaved pRSP-1-Δ IVS DNA template with SP6 RNA polymerase using a G(5′)ppp(5′)G dinucleotide primer[17]. Nuclear extract was prepared from uninfected HeLa cells as described previously[22]. Standard splicing reactions (25 μl) contained: 50,000 c.p.m. ^{32}P-RNA precursor, 1.5 mM ATP, 5 mM creatine phosphate, 2.5 mM $MgCl_2$, 44 mM KCl, 8.8 mM HEPES *p*H 7.6, 0.2 mM EDTA, 0.9 mM dithiothreitol, 7.5% glycerol and 11 μl nuclear extract. After 2 h of incubation at 30 °C, RNA was extracted and analysed on a 10% polyacrylamide 8 M urea gel[18]. Preparative reactions contained 5-25 × 10^6 c.p.m. ^{32}P-RNA and were carried out in a total volume of 100 μl. IVS-A RNA from a preparative splicing reaction was purified by polyacrylamide gel electrophoresis. Nuclease P_1 digestions were carried out in 10 μl of 50 mM ammonium acetate *p*H 5.3 with 1 μg enzyme for 1 h at 50 °C. Snake venom phosphodiesterase digestions were carried out in 10 μl of 50 mM Tris-HCl *p*H 7.8 with 1 μg enzyme for 1 h at 37 °C. Products of the digestions were separated by two-dimensional TLC on cellulose plates (Merck) using solvent A (isobutyric acid/concentrated NH_4OH:H_2O; 577:38:385), solvent B (*t*-butanol/concentrated HCl/H_2O; 14:3:3) or solvent C (saturated $(NH_4)_2SO_4$/1 M sodium acetate/isopropanol; 80:18:2). In the experiment described here, the ratio of incorporation of [α-^{32}P]-NTPs into pre-RNA was as follows: A/1.0, G/2.0, U/0.7, C/ − 1.5.

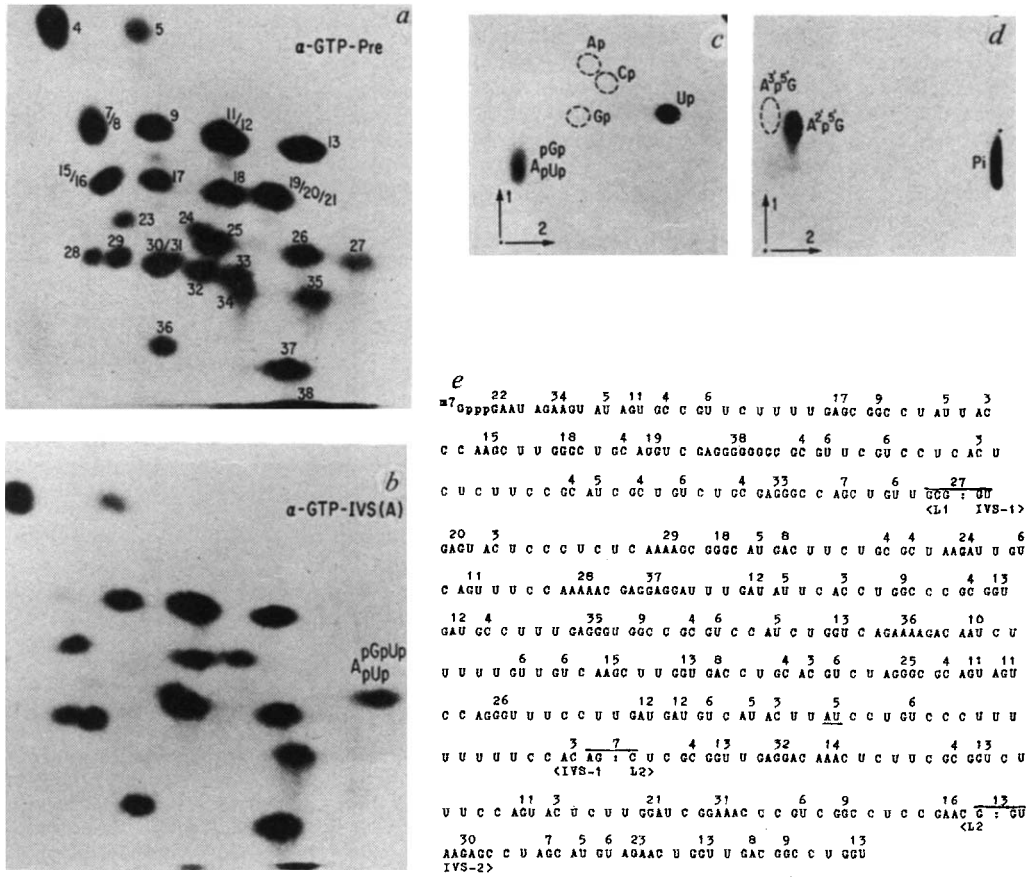


Fig. 2 *a*, Two-dimensional RNA fingerprint analysis of an RNase A digestion of purified pre-RNA labelled with [α-^{32}P]GTP. *b*, Two-dimensional RNA fingerprint analysis of RNase A digestion of purified IVS-A RNA labelled with [α-^{32}P]-GTP. *c*, The branch-containing oligonucleotide from *b* was digested with RNase T_1 and analysed by TLC in solvent A in the first dimension and solvent B in the second dimension. *d*, The branch-containing oligonucleotide from *b* was treated with phosphatase followed by periodate cleavage and β-elimination followed by phosphatase. The products were analysed by TLC as in *c*. The positions of the unlabelled A 2′p5′G and A 3′p5′G markers (Sigma) are indicated. *e*, Sequence of pre-RNA with all RNase A oligonucleotides numbered. Sequences corresponding to exons and introns are indicated and the AU dinucleotide at the branch site is underlined.

Methods. RNA fingerprint analysis. RNA was dissolved in 20 μl 10 mM Tris-HCl (*p*H 7.5), 1 mM EDTA and digested with 1.25 μg RNase A for 2 h at 37 °C. The sample was lyophilized, resuspended in 3 μl of H_2O and the oligonucleotides separated by electrophoresis on cellulose acetate from left to right followed by homochromatography on DEAE-cellulose plates from bottom to top as described previously[32]. The branch-containing oligonucleotide was isolated from the DEAE-cellulose plate by elution with 1 M triethylamine bicarbonate, then digested with RNase T_1 (Calbiochem) (2 U in 50 mM Tris-HCl, *p*H 7.7) for 3 h at 37 °C or with calf alkaline phosphatase (Boehringer) (1 U in 50 mM Tris *p*H 7.7) for 1.5 h at 37 °C. Phosphatase-treated branched oligonucleotide was subjected to periodate oxidation in 10 μl 0.1 M $NaIO_4$, 0.1 M sodium acetate *p*H 5.3 for 2 h at 20 °C in the dark followed directly by analine treatment in 25 μl of 0.5 M aniline 0.2 M sodium acetate, *p*H 5.3, for 3 h at 20 °C in the dark[33]. The sample was extracted with phenol/chloroform (1:1), lyophilized and digested with phosphatase as above.

tide which is similar to the branch trinucleotide described previously[20], containing 2′–5′ and 3′–5′ phosphodiester bonds joined to a single adenosine residue[19]. The *in vitro* splicing of human β-globin pre-mRNA is similar[21]. In neither study was a complete analysis of a branch oligonucleotide reported.

We present here a detailed characterization of the branch oligonucleotide. Specifically we show that: (1) the branch trinucleotide has a structure of $A^{2'p5'G}_{3'p5'U}$; (2) the 5′-terminal G of the intervening sequence is joined directly to the adenosine residue at the branch site; (3) the phosphate group from the 5′ splice site is incorporated into the 2′–5′ phosphodiester bond; (4) the branch site is 24 nucleotides upstream of the 3′ splice site; and (5) the probable intermediate in the reaction, the 5′ exon, terminates in a 3′ hydroxyl group.

Substrate and reaction conditions

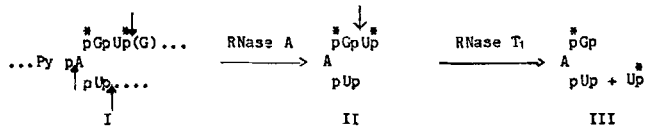
As a model substrate in our studies, we have used a precursor RNA containing the first two leader exons (L1 and L2) of the Ad-2 major late transcription unit separated by a deleted form of the intervening sequence[18]. Pre-RNA was transcribed by SP6 RNA polymerase from a DNA template containing an SP6 promoter cloned upstream of the Ad-2 sequences[17]. Transcription of this substrate was primed using a G(5′)ppp(5′)G dinucleotide[17] and in most cases, the pre-RNA processed by incubation in a HeLa nuclear extract[22]. All the characteristic products and intermediates described previously for the whole-cell extract reaction were also detected after incubation in the nuclear extract; specifically, the splicing reaction displayed a similar time course and dependence on Mg^{2+} and ATP (Fig. 1*a*). Moreover, as has been reported previously[15,21], the efficiency of RNA processing was significantly higher in the nuclear extract than in the whole-cell extract (data not shown).

Oligonucleotide branch structure

To analyse the base composition of the nuclease-resistant structure in the lariat RNAs, the excised intervening sequence, IVS-A RNA, was isolated from a splicing reaction in which the pre-RNA was labelled with all four [α-^{32}P]-labelled nucleoside triphosphates (NTPs) (Fig. 1*a*). IVS-A RNA was digested to completion with nuclease P_1 and the products separated by two-dimensional TLC (Fig. 1*b*). In addition to the four anticipated nucleotides, a nuclease-resistant oligonucleotide was resolved, purified and then digested with snake venom phosphodiesterase, yielding radiolabelled pA, pG and pU but no pC (Fig. 1*c*). Thus, the nuclease P_1-resistant oligonucleotide is composed of A, G and U residues in approximately equal molar quantities. (The variation in radioactivity between the three mononucleotides in Fig. 1*c* is due to differences in specific activity.) We have reported previously that β-elimination of the P_1 nuclease-resistant oligonucleotide yielded a 2′,3′,5′ ATP ($5'pA^{2'p}_{3'p}$) residue[19], suggesting that the G and U are linked to the 2′ and 3′ positions at the RNA branch point (see below).

The lariat structure of the intervening sequence is formed by the joining of sequences at the 5′ splice site to a site near the 3′ end of the molecule[19]. Neither the earlier study nor the above data, however, established that the G at the 5′ end of the IVS was retained in the formation of the branch. Definitive identification of this G as part of the branch was provided by an analysis of [α-^{32}P]GTP labelled RNA. Pre-RNA and IVS-A RNA were digested in parallel with RNase A and analysed by two-

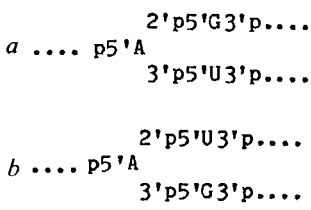
dimensional fingerprint chromatography (Fig. 2*a*,*b*). A unique branch-containing oligonucleotide was identified in the IVS-A RNA digest that was absent from the pre-RNA digest. Subsequent RNase T_1 digestion of this purified oligonucleotide released a radiolabelled Up and a product migrating in the two-dimensional TLC system with the same mobility as the previously-described RNase T_2-resistant branch trinucleotide (Fig. 2*c* and ref. 19). These results are consistent with the following scheme:



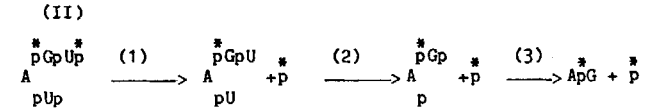
(* Denote ^{32}P-phosphate residues, ↓ and ↓ indicate cleavage sites of RNases A and T_1, respectively.) The net charge of structures II and III was determined by DEAE TLC; consistent with the above scheme, the RNase A product (II) had a net charge of −7 and treatment with RNAse T_1, yielded products with net charges of −6 and −2 (data not shown).

The sequence GUG, established by these data as one arm of the branch, is expected if the 5′ splice site is retained precisely in the reaction. As the radioactive phosphate in oligonucleotide III is found in a 2′-5′ phosphodiester bond (see below), the phosphate at the 5′ splice site must be conserved in branch formation.

Two alternative branch structures are in agreement with the data presented so far, which differ by the distribution of G and U at the 2′ or 3′ position:



To distinguish between these two possibilities, the RNase A oligonucleotide containing the branch site (structure II) was isolated from [α-^{32}P]GTP-labelled IVS-A RNA (Fig. 2*b*) and treated consecutively with phosphatase, periodate/analine (β-elimination), then phosphatase. This sequence of reactions should yield the following set of labelled products:



(* Denote ^{32}P-phosphate residues.) Structure *a* would yield the dinucleotide A2′p5′G, whereas structure *b* would yield A3′p5′G, with a ^{32}P-phosphate released in both cases. As shown in Fig. 2*d*, the product of the reaction co-migrates in two-dimensional TLC with the marker A2′p5′G and not with the marker A3′p5′G, thus the 5′ terminal G residue of the intervening sequence is joined via a 2′-5′ phosphodiester bond to an A residue.

Location of branch site

The branch site is in the RNase T_1-resistant oligonucleotide UCAUACUUAUCCUG, within the intervening sequence from −19 to −32 nucleotides from the 3′ splice site[19]. This oligonucleotide contains two AU dinucleotides (underlined in the sequence above) consistent with the established structure of the branch. To discriminate between the two alternative branch sites, we took advantage of the fact that one of the above AU sequences

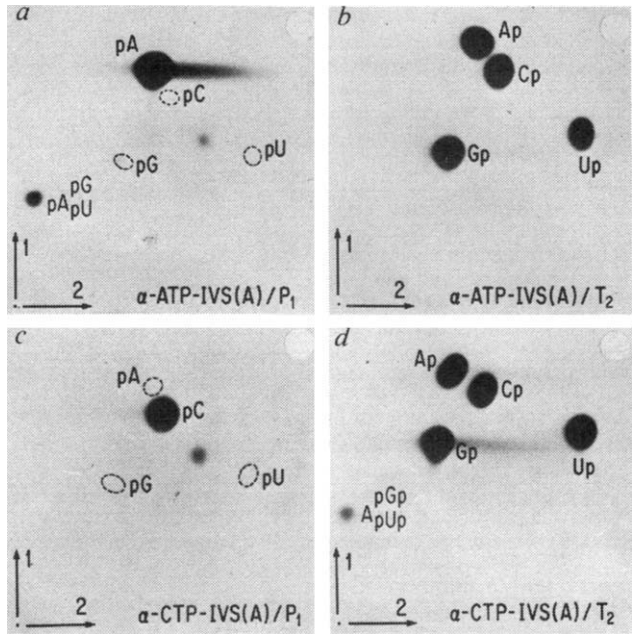


Fig. 3 Localization of the branch site by nearest-neighbour analysis. IVS-A RNAs labelled with either [α-^{32}P]-ATP (*a*, *b*) or [α-^{32}P]-CTP (*c*, *d*) were digested with nuclease P_1 (*a*, *c*) or RNase T_2 (*b*, *d*). Products of the reactions were resolved by TLC on cellulose plates in solvent A in the first dimension and solvent B in the second dimension. Positions of nucleoside 5′ or 3′ monophosphate markers and the branched trinucleotide are indicated. The additional spot observed near the centre of plates *a* and *c* is cytosine 5′ phosphate, 2′,3′-cyclic phosphate, originating probably from the nonspecific breakdown of RNA (data not shown). Digestions with RNase T_2 (Calbiochem, 1 U in 50 mM ammonium acetate *p*H 5.3) were for 1 h at 50 °C. Digestions with nuclease P_1 were as in Fig. 1 legend.

is followed by an A whereas the other is followed by a C residue. A phosphate transfer analysis was thus performed by digesting IVS-A RNA labelled with [α-^{32}P]ATP or [α-^{32}P]CTP in parallel with RNase T_2 (Fig. 3). As the branched trinucleotide was labelled with [α-^{32}P]CTP (Fig. 3d) but not with [α-^{32}P]-ATP (Fig. 3*b*), the downstream AU sequence is the exclusive site of branch formation. Nuclease P_1 digestion was performed as a control for specific labelling of the IVS-A RNAs and to demonstrate the presence of a branch trinucleotide in the [α-^{32}P]ATP-labelled RNA (Fig. 3*a*,*c*).

As a further demonstration of branch location, the branch containing RNase T_1 oligonucleotide was isolated from IVS-A RNA labelled with both [α-^{32}P]CTP and [α-^{32}P]UTP and digested subsequently with the A-specific RNase U_2. RNase U_2 cleavage requires a 2′-hydroxyl group and therefore will not cleave after the A residue at the branch site. RNase U_2 digestion of the branch containing RNase T_1 oligonucleotide produced UCA, UA and a large branch-containing oligonucleotide $CUUA^{G}_{UCCUG}$ (Fig. 4*a*). A control digestion of the equivalent RNase T_1 oligonucleotide isolated from pre-RNA gave the expected products UCA, UA, CUUA and UCCUG (Fig. 4*b*). This and the above analyses demonstrate that the branch containing RNase T_1 oligonucleotide has the structure $UCAUACUUA^{G}_{UCCUC}$.

Structure of the first exon intermediate

We have shown previously that the IVS-L2-A RNA is a lariat RNA species composed of the intervening sequence and the L2 exon[18,19]. This RNA appears early in the splicing reaction and is probably part of an intermediate (ref. 18 and Fig. 1*a*). The IVS-L2-A RNA results from a cleavage at the 5′ splice site of precursor RNA[18,19]; the same event should produce the L1 RNA, recoverable as a free 5′ exon. An RNA species of the appropriate length (~120 nucleotides) appeared early in the reaction in an

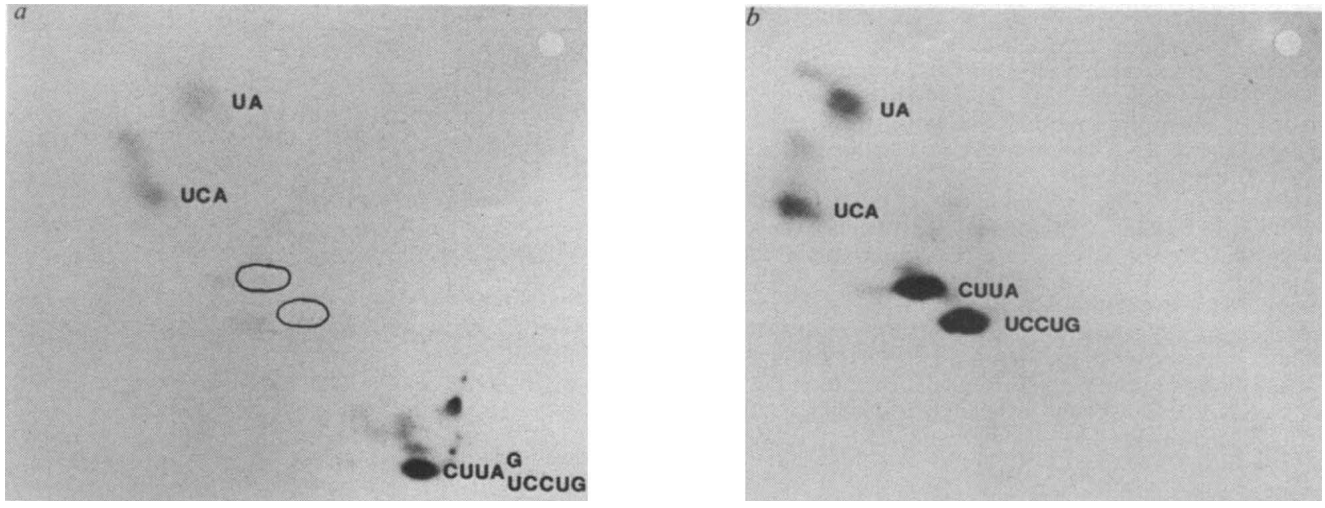


Fig. 4 Localization of the branch site by RNase U_2 analysis. Two-dimensional RNA fingerprint analysis of RNase U_2 digestions of: *a*, the branch containing RNase T_1 oligonucleotide from IVS-A RNA, and *b*, the analogous, unmodified RNase T_1 oligonucleotide from pre-RNA. IVS-A RNA and pre-RNA were labelled with both [α-^{32}P]-CTP and [α-^{32}P]-UTP. The RNase T_1 oligonucleotides were isolated from a 15% polyacrylamide 8 M urea gel[19] and digested with 0.01 U RNase U_2 (Calbiochem) in 20 mM sodium citrate *p*H 3.5 for 30 min at 30 °C. Two-dimensional fingerprint chromatography was as in Fig. 2.

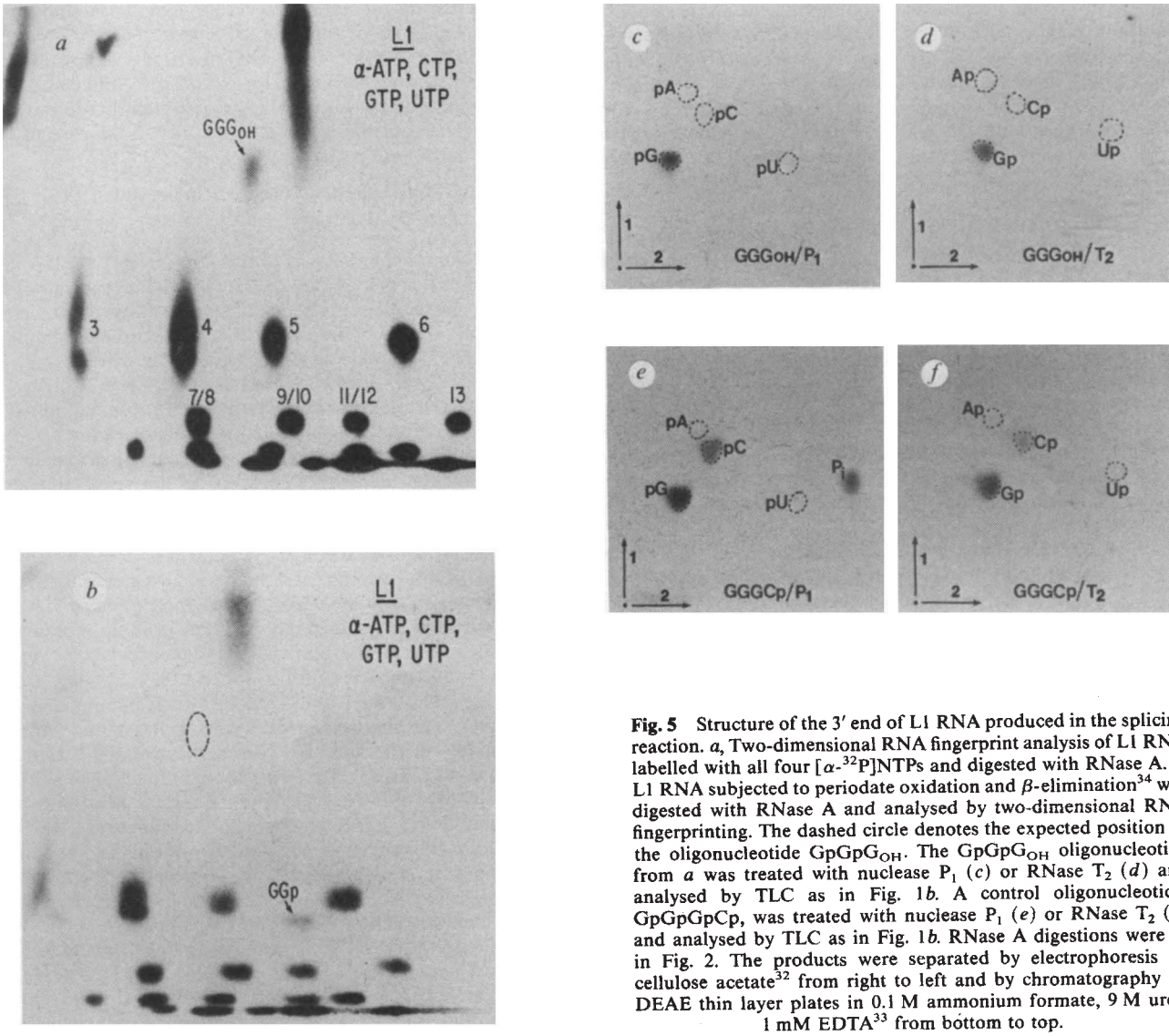


Fig. 5 Structure of the 3′ end of L1 RNA produced in the splicing reaction. *a*, Two-dimensional RNA fingerprint analysis of L1 RNA labelled with all four [α-^{32}P]NTPs and digested with RNase A. *b*, L1 RNA subjected to periodate oxidation and β-elimination[34] was digested with RNase A and analysed by two-dimensional RNA fingerprinting. The dashed circle denotes the expected position of the oligonucleotide $GpGpG_{OH}$. The $GpGpG_{OH}$ oligonucleotide from *a* was treated with nuclease P_1 (*c*) or RNase T_2 (*d*) and analysed by TLC as in Fig. 1*b*. A control oligonucleotide, GpGpGpCp, was treated with nuclease P_1 (*e*) or RNase T_2 (*f*) and analysed by TLC as in Fig. 1*b*. RNase A digestions were as in Fig. 2. The products were separated by electrophoresis on cellulose acetate[32] from right to left and by chromatography on DEAE thin layer plates in 0.1 M ammonium formate, 9 M urea, 1 mM EDTA[33] from bottom to top.

ATP-dependent fashion (Fig. 1*a*). This RNA contains L1 sequences, shown by RNase A fingerprint analysis and hybridization to single-stranded M13 DNAs containing sequences complementary to the 5′ portion of the pre-RNA (data not shown).

As the 3′ end of the L1 RNA is involved potentially in the formation of the spliced product, we sought to determine its structure. The L1 RNA cleaved precisely at the 5′ splice site should have a 3′ sequence of . . . UGGG (see Fig. 2*e*). As shown above, the phosphate moiety from the 5′ splice site is contained in the 2′–5′ phosphodiester bond of the branch. Thus, it is probable that the L1 RNA has a 3′ hydroxyl terminus. To investigate this, an RNase A digestion of the purified L1 RNA was analysed by a two-dimensional fingerprint. A unique oligonucleotide with the anticipated mobility of $GpGpG_{OH}$ is present in L1 RNA labelled with four [α-^{32}P]NTPs (Fig. 5*a*) or [α-^{32}P]GTP (data not shown). This oligonucleotide is not present in L1 RNA labelled with [α-^{32}P]CTP. It is also absent from both L1–2 RNA and pre-RNA labelled with all four [α-^{32}P]NTPs (data not shown). If the L1 RNA terminates with 2′ and 3′ hydroxyl groups, then periodate oxidation and β-elimination should yield GpGp as a product of RNase A digestion. As anticipated, the $GpGpG_{OH}$ oligonucleotide (net charge −2) was absent and a new oligonucleotide migrating at the expected position of GpGp (net charge −3) appeared after this treatment (Fig. 5*b*), consistent with the location of $GpGpG_{OH}$ at the 3′ terminus of L1 RNA. To confirm the identity of the $GpGpG_{OH}$ oligonucleotide, we determined its base composition by elution from the fingerprint and digestion with RNase T_2 or nuclease P_1. Only the anticipated products Gp and pG were observed (Fig. 5*c*,*d*); control digestions on oligonucleotide GpGpGpCp yielded the expected products (Fig. 5*e*,*f*). Note that nuclease P_1 digestion did not release phosphate from oligonucleotide $GpGpG_{OH}$, consistent with the presence of a 3′ terminal hydroxyl group. The assignment of the 3′ hydroxyl group in the L1 exon agrees with the previous finding that the phosphate moiety at the splice junction of the two exons is derived from the 3′ splice site[19].

Mechanism of mRNA precursor splicing

On the basis of the work reported here and other recent findings[18,19,21], we propose a scheme for the splicing of mRNA precursors (Fig. 6). A multicomponent precursor mRNA–ribonucleoprotein complex with a poorly defined structure is formed first; it probably contains a small nuclear U1 ribonucleoprotein particle bound to the 5′ splice site[16,23,24]. An involvement of other small nuclear ribonucleoproteins in mRNA splicing is also possible[25,26]. The lag in the time course observed *in vitro*[13–15] and the recognition of the 5′ cap structure at an early stage of the reaction[17] may reflect an ATP-dependent organization of precursor into a complex which properly positions the two exons.

The first characterized step in the reaction is cleavage of the RNA precursor at the 5′ splice site, producing a 5′ exon terminating in a 3′ hydroxyl group and a lariat structure containing the remainder of the RNA. The branch in the lariat results from the joining of the 5′ terminal G of the intervening sequence to the 2′ hydroxyl group of an A. The branch site in this particular example is positioned 24 nucleotides upstream of the 3′ splice site. Interestingly, there is a distinct complementarity between the sequences at the 5′ terminus of the intervening sequence and the nucleotides preceding the A at the branch site (see Fig. 6). The close proximity of the 2′ hydroxyl of the A at the branch site to the 5′ splice site in a secondary structure could facilitate the cleavage/ligation reaction. It is not clear whether cleavage at the 5′ splice site and formation of the 2′–5′ phosphodiester bond occurs through a concerted transesterification reaction[8] or through a series of enzymatic cleavage/ligation reactions. The phosphate group from the 5′ splice site (circled in Fig. 6) is retained in the 2′–5′ phosphodiester bond of the branch structure, consistent with the released 5′ exon terminating in a 3′ hydroxyl group. A second cleavage/ligation reaction joins the two exons by a 3′–5′ phosphodiester bond and releases the lariat form of

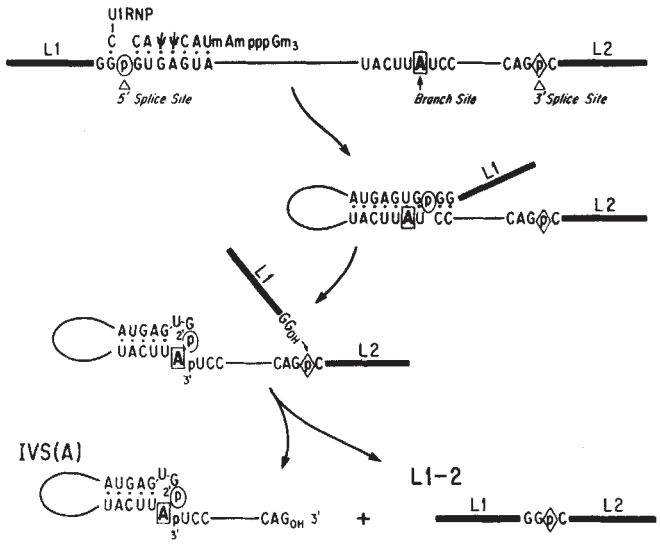


Fig. 6 Proposed scheme of pre-mRNA splicing.

the intervening sequence with a 3′ hydroxyl group[19]. In this reaction, the phosphate group at the 3′ splice site (enclosed by a diamond in Fig. 6) is joined to the 3′ hydroxyl group of the 5′ exon. As in the previous step, this reaction could either proceed as one concerted transesterification reaction, or could consist of a series of separate reactions.

The degree of complementarity (10 base pairs (bp) including G:U pairing) between sequences at the 5′ splice site and the branch site of IVS-1 is probably greater than that found in other introns. The equivalent complementarity in the case of human β-globin is 9 base pairs[21].

$^{3'}$ACUAUGGUUG:GA$^{5'}$ 5′ splice site of IVS-1

$^{5'}$AGGCACUGAC UC$^{3'}$ branch site of IVS-1

Some complementarity is found also in yeast mRNA introns, where the conserved sequence UACUAAC is required upstream of all 3′ splice sites; the sequence at the 5′ splice site is almost invariably G:GUAUGU[27,28], these two sequences being complementary[28]. Single base changes in the yeast intron UACUAAC sequence reduce the efficiency of splicing[29]; the UACUACC variant is totally defective in splicing, not surprisingly as the mutated A is probably the base of the branch site.

Sequences encompassing potential branch sites in mammalian introns are not as highly conserved as the yeast equivalent. A search of mammalian intron sequences using the yeast paradigm yielded a consensus sequence of CTGAC[26]. In the rabbit β-globin large intron, deletion of sequences containing the best example of this consensus sequence, almost certainly the normal branch site, did not abolish accurate splicing *in vivo*[30]. Although alternative branch sites might be used in these mutants, these data suggest that the branch site sequence is not the major determinant for selection of 3′ splice sites.

The sequences near the 5′ splice site are complementary to both the 5′ terminal sequences of U1 RNA and branch-site sequences (see Fig. 6). Base-pairing between the 5′ terminus of U1 RNA and the 5′ splice site is thought to mediate recognition for splicing; following selection of the 5′ splice site, the U1 RNA could be displaced by branch-site sequences. As described above, the efficiencies of splicing may not be highly dependent on the degree of complementarity between the 5′ splice site and the branch site. Accurate splicing *in vivo* has been often observed with chimaeric genes, where the 5′ splice site and the 3′ splice site (including the branch site) originated from two different and unrelated genes (see, for example, ref. 31). The degree of complementarity between the branch site and the 5′ splice site may regulate the rate of splicing by stabilizing their interaction. For example, it might determine the order of excision of multiple

introns in a single transcript. In cases where alternative splice sites are used, the relative strengths of the interactions may determine the ratios of splice site use. Until the branch sites used in such RNAs are determined, these possibilities are difficult to evaluate.

The formation of a branch seems to be a unique feature of the splicing of pre-mRNAs. This unconventional structure might be simply a result of the splicing mechanism. It is also possible, however, that the reactive 5′ end of the intervening sequence is prevented in this way from recombining with the 5′ exon, driving the reaction in the forward direction[9]. Finally, the branch structure might serve as a positive signal for rapid degradation of intervening sequences *in vivo*.

The pathway of splicing described here may not be restricted necessarily to nuclear protein-coding genes. A recent finding that the excised, circular intervening sequence produced during splicing of the transcripts of the yeast mitochondrial gene for subunit I of cytochrome *c* oxidase may be a lariat RNA containing a branch structure (L. A. Grivell, personal communication) suggests that the splicing mechanism operating in these two classes of genes is similar. Moreover, the fact that nuclear introns contain internal sequences important to splicing (branch sites) relates this class of introns to both mitochondrial mRNA introns and nuclear and mitochondrial rRNA introns (for review see ref. 3). Further studies are necessary to determine whether the structural similarities of these introns are accompanied by a corresponding resemblance in the mechanisms of splicing.

We thank Claire Moore for discussions and for preparation of the nuclear extract. P.J.G. acknowledges Peter Gegenheimer for suggestions on RNA fingerprint analysis. M.M.K. holds a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund, P.J.G. from Helen Hay Whitney Foundation and R.A.P. from the Myron Bantrell Fellowship. This work was supported by NIH grants RO1-GM32467 and PO1-CA26717 and NSF grant PCM-8200309 to P.A.S. and partially by NIH core grant PO1-CA14051.

Received 18 October; accepted 6 December 1984.

1. Abelson, J. *A. Rev. Biochem.* **48**, 1035–1063 (1979).
2. Sharp, P. A. *Cell* **23**, 643–646 (1981).
3. Cech, T. *Cell* **34**, 713–716 (1983).
4. Konarska, M., Filipowicz, W. & Gross, H. J. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1474–1478 (1982).
5. Greer, C. L., Peebles, C. L., Gegenheimer, P. & Abelson, J. *Cell* **32**, 537–546 (1983).
6. Filipowcz, W. & Shatkin, A. J. *Cell* **32**, 547–557.
7. Laski, F. A., Fire, A. Z., RajBhandary, U. L. & Sharp, P. A. *J. biol. Chem.* **258**, 11974–11980 (1983).
8. Cech, T. R., Zaug, A. J. & Grabowski, P. J. *Cell* **27**, 487–496 (1981).
9. Kruger, K. *et al. Cell* **31**, 147–157 (1982).
10. Zaug, A. J., Grabowski, P. J. & Cech, T. R. *Nature* **301**, 578–583 (1983).
11. Kole, R. & Weissman, S. M. *Nucleic Acids Res.* **10**, 5429–5445 (1982).
12. Padgett, R. A., Hardy, S. F. & Sharp, P. A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5230–5234 (1983).
13. Hernandez, N. & Keller, W. *Cell* **35**, 89–99 (1983).
14. Hardy, S. F., Grabowski, P. J., Padgett, R. A. & Sharp, P. A. *Nature* **308**, 375–377 (1984).
15. Krainer, A. R., Maniatis, T., Ruskin, B. & Green, M. R. *Cell* **36**, 993–1005 (1984).
16. Padgett, R. A., Mount, S. M., Steitz, J. A. & Sharp, P. A. *Cell* **35**, 101–107 (1983).
17. Konarska, M. M., Padgett, R. A. & Sharp, P. A. *Cell* **38**, 731–736 (1984).
18. Grabowski, P. J., Padgett, R. A. & Sharp, P. A. *Cell* **37**, 415–427 (1984).
19. Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F. & Sharp, P. A. *Science* **225**, 898–903 (1984).
20. Wallace, J. C. & Edmonds, M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 950–954 (1983).
21. Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. *Cell* **38**, 317–331 (1984).
22. Dignam, J. D., Lebowitz, R. M. & Roeder, R. G. *Nucleic Acids. Res.* **11**, 1475–1489 (1983).
23. Mount, S. M., Pettersson, I., Hinterberger, M., Karmer, A. & Steitz, J. A. *Cell* **33**, 509–518 (1983).
24. Krämer, A., Keller, W., Appel, B. & Lührmann, R. *Cell* **38**, 299–307 (1984).
25. Ohshima, Y., Itoh, M., Okada, N. & Miyata, T. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4471–4474 (1981).
26. Keller, E. B. & Noon, W. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7417–7420 (1984).
27. Langford, C. J. & Gallwitz, D. *Cell* **33**, 519–527 (1983).
28. Pikielny, C. W., Teem, J. L. & Rosbash, M. *Cell* **34**, 395–403 (1983).
29. Langford, C. J., Klinz, F.-J., Donath, C. & Gallwitz, D. *Cell* **36**, 645–653 (1984).
30. Wieringa, B., Hofer, E. & Weissmann, C. *Cell* **37**, 915–925 (1984).
31. Chu, G. & Sharp, P. A. *Nature* **289**, 378–382 (1981).
32. Silberklang, M., Gillum, A. M. & RajBhandary, U. L. *Meth. Enzym.* **59**, 58–109 (1979).
33. Konarska, M., Filipowicz, W., Domdey, H. & Gross, H. J. *Nature* **293**, 112–116 (1981).
34. Rose, J. K. & Lodish, H. F. *Nature* **262**, 32–37 (1976).

LETTERS TO NATURE

Pavo XD-10, an X-ray QSO with extended optical structure

R. Gilmozzi*†, J. V. Wall†, P. G. Murdin†, P. R. Jorden†, D. J. Thorne†, I. G. van Breda† & J. A. Peacock‡

* ESA Satellite Tracking Station, Villafranca del Castillo, PO Box 54065, Madrid, Spain
† Royal Greenwich Observatory, Herstmonceux Castle, Hailsham, East Sussex BN27 1RP, UK
‡ Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

Deep charge-coupled device (CCD) images of the quasi-stellar object (QSO) identified with the Einstein X-ray source Pavo XD-10 show an optical structure that is dramatically extended. Here we discuss the morphology, the photometry from the CCD data, and the limited spectral information which indicates a redshift (z) of 0.72 for the system. We show that if the extended structure represents a galaxy hosting the $z = 0.72$ QSO, the galaxy luminosity is 2 mag brighter than the host galaxies detected for QSOs at lower redshifts. Instead, the system may be the result of a galaxy–galaxy interaction. A third and more speculative possibility is that it is the result of gravitational lensing, with much of the 'underlying' structure representing the lens galaxy in the line-of-sight.

Pavo XD-10 was identified as an X-ray QSO in a detailed investigation[1] of the Pavo field in the Einstein Observatory deep X-ray survey[2]. To extend this investigation, we took broad-band (*BVRI*) exposures of several areas in the Pavo field at the prime focus of the 3.9-m Anglo-Australian Telescope (AAT) in September 1981, using a CCD camera built at the Royal Greenwich Observatory (RGO)[3]. We noted then the extension to QSO

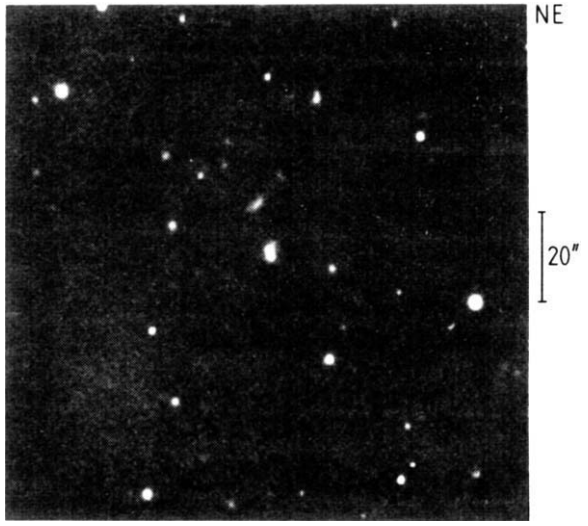


Fig. 1 A photographic representation of the region about Pavo XD-10 (1950.0 $\alpha = 21$ h 11 min 32.6 s, $\delta = -67°\ 54'\ 26''$), obtained from the 3,600-s *I*-band image taken with a CCD camera at the Cassegrain focus of the Danish 1.5-m telescope at ESO/La Silla. The QSO identification is the image elongated north-south in the centre of the frame, object 'a' of ref. 1.

XD-10. Much better observing conditions (photometric, with 1 arc s seeing) were obtained during the commissioning of a second RGO CCD camera at the *f*/8 Cassegrain focus of the Danish 1.5-m telescope at the European Southern Observatory (ESO)/La Silla in August 1982. Both CCD cameras were based on RCA 53612 detectors cooled to 150 K in N_2 cryostats. The